2017

Novel Cell Surface Anchoring Mechanism Of Prokaryotic Secreted Protein

Mohd Farid Abdul Halim

University of Pennsylvania, farid_halim872002@yahoo.com

Follow this and additional works at: https://repository.upenn.edu/edissertations

Part of the Genetics Commons, Microbiology Commons, and the Molecular Biology Commons

Recommended Citation


https://repository.upenn.edu/edissertations/2155

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/edissertations/2155

For more information, please contact repository@pobox.upenn.edu.
Novel Cell Surface Anchoring Mechanism Of Prokaryotic Secreted Protein

Abstract
The microbial cell surface is decorated with a variety of protein structures that play important roles in key cellular processes such as providing cell stability, facilitating interactions between cells, and interacting with the environment. One important feature of the biosynthesis of these structures is the proper anchoring of proteins to the cell surface. In silico work recently predicted a novel protein anchoring mechanism for a subset of surface proteins that contain a conserved C-terminal tripartite architecture, which consists of a conserved motif, followed by a hydrophobic (H) domain, and positively charged amino acids. Using the well-studied model archaeon Haloferax volcanii S-layer glycoprotein (SLG), previously thought to be anchored in the membrane via its C-terminal hydrophobic domain, I have shown that its H-domain instead is processed and that this processing is dependent on a novel enzyme, the archaeosortase (ArtA). The characterization of a strain lacking ArtA is also described, demonstrating its substrates proper anchoring is important for maintaining cell stability, morphology, motility, and mating efficiency. I also have demonstrated that the SLG is anchored to the membrane via a covalent lipid anchor, the first C-terminal lipid-anchoring mechanism described for any prokaryote. Furthermore, I have shown that this process is also ArtA-dependent and that the substrate conserved PGF motif is critical for processing and lipid anchoring. This novel anchoring mechanism is not limited to SLG, as I have demonstrated Hvo_0405, is also processed in an ArtA-dependent manner and showed that the conserved tripartite architecture does not necessarily need to lie at the C-terminus. Interestingly, I was able to show that Hvo_0405, unlike SLG, which is a Sec substrate, is transported across the H. volcanii membrane via the Twin Arginine (Tat) pathway demonstrating that Sec and Tat substrates can be C-terminally processed in an ArtA-dependent manner. Considering that ArtA homologs are conserved among diverse organisms of both prokaryotic domains, data obtained from my work will also serves as foundation for future studies of cell surface anchoring in other archaea and bacteria.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Biology

First Advisor
Mechthild Pohlschroder

Keywords
Archaea, Cell surface anchoring, Secreted proteins, S-layer glycoprotein

Subject Categories
Genetics | Microbiology | Molecular Biology

This dissertation is available at ScholarlyCommons: https://repository.upenn.edu/edissertations/2155
NOVEL CELL SURFACE ANCHORING MECHANISM OF PROKARYOTIC SECRETED PROTEIN

MOHD FARID ABDUL HALIM

A DISSERTATION

in

Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2017

Supervisor of Dissertation

________________________
Mechthild Pohlschroder, PhD
Professor of Biology and Undergraduate Chair in Biology

Graduate Group Chairperson

________________________
Micheal Lampson, PhD
Associate Professor of Biology

Dissertation Committee

Fevzi Daldal, PhD Philip A. Rea, PhD
Professor of Biology Professor of Biology

Mark Goulian, PhD Benjamin Garcia, PhD
Professor of Biology Presidential Professor of Biochemistry and Biophysics
NOVEL CELL SURFACE ANCHORING MECHANISM OF PROKARYOTIC SECRETED PROTEIN

COPYRIGHT

2017

Mohd Farid Abdul Halim

This work is licensed under the
Creative Commons Attribution-
NonCommercial-ShareAlike 3.0
License

To view a copy of this license, visit

https://creativecommons.org/licenses/by-nc-sa/3.0/us/
"In the name of God, most Gracious, most Merciful"

"Dengan nama Allah, Maha Pemurah, Lagi Maha Mengasihani"
ACKNOWLEDGMENT

This thesis and the accompanying PhD would not have been possible without the help of many people throughout my life and academic career. First, thank you to my thesis committee chair- Dr. Fevzi Daldal, other committee members: Dr. Mark Goulian, Dr. Philip Rea, and Dr. Ben Garcia for their questions, suggestions, and guidance allowed me to move quickly and efficiently towards graduation. Also, thank you to my collaborators in the Pacific Northwest National Lab (PNLL), Garcia lab, Kelly Karch, and in the Hippler lab, Stefan Schulze without whom none of the mass spec in this thesis would have been possible. Next, thank you to the former and current members of the Pohlschroder lab. Andrew, Manuela, Zhicong, James, Yitian, Rianne, Georgio, Evan, Jonathan, and Ron for their help in training me when I first got in the lab and countless favors whom I owed in getting my project done in time. Getting to have you guys around makes working in the lab is both comforting and helpful. A special thank you goes to Andrew, who trained me in the lab, entertained all my questions about *Haloferax volcanii*, and getting my project up and running. A special thanks to Rianne, have been my labmate and friend through for most part of my years as PhD student. Special thanks to Georgio and Evan as well for their countless contribution in discussion, figuring out stuff in the lab, reading the grants proposal, submitted journals and my thesis dissertation. Mostly, I owe numerous thanks and grateful to my mentor, Mecky Pohlschroder. She has always been there for me and I would not be completing my PhD without her opening her arms accepting me as a student in the lab. With her guidance, I have matured as a scientist and feel ready to take on the new challenges in my next career. Switching to the more personal thank yous, thank you
to my family who always has supported my endeavors even from halfway across the globe.

Thank you to my friends at Drexel University who always keeping my sanity while doing the PhD here at Penn.
ABSTRACT

NOVEL CELL SURFACE ANCHORING MECHANISM OF PROKARYOTIC SECRETED PROTEIN

Mohd Farid Abdul Halim
Dr. Mechthild Pohlschroder

The microbial cell surface is decorated with a variety of protein structures that play important roles in key cellular processes such as providing cell stability, facilitating interactions between cells, and interacting with the environment. One important feature of the biosynthesis of these structures is the proper anchoring of proteins to the cell surface. *In silico* work recently predicted a novel protein anchoring mechanism for a subset of surface proteins that contain a conserved C-terminal tripartite architecture, which consists of a conserved motif, followed by a hydrophobic (H) domain, and positively charged amino acids. Using the well-studied model archaeon *Haloferax volcanii* S-layer glycoprotein (SLG), previously thought to be anchored in the membrane via its C-terminal hydrophobic domain, I have shown that its H-domain instead is processed and that this processing is dependent on a novel enzyme, the archaeosortase (ArtA). The characterization of a strain lacking ArtA is also described, demonstrating its substrates proper anchoring is important for maintaining cell stability, morphology, motility, and mating efficiency. I also have demonstrated that the SLG is anchored to the membrane via a covalent lipid anchor, the first C-terminal lipid-anchoring mechanism described for any prokaryote. Furthermore, I have shown that this process is also ArtA-dependent and that the substrate conserved PGF...
motif is critical for processing and lipid anchoring. This novel anchoring mechanism is not limited to SLG, as I have demonstrated Hvo_0405, is also processed in an ArtA-dependent manner and showed that the conserved tripartite architecture does not necessarily need to lie at the C-terminus. Interestingly, I was able to show that Hvo_0405, unlike SLG, which is a Sec substrate, is transported across the H. volcanii membrane via the Twin Arginine (Tat) pathway demonstrating that Sec and Tat substrates can be C-terminally processed in an ArtA-dependent manner. Considering that ArtA homologs are conserved among diverse organisms of both prokaryotic domains, data obtained from my work will also serves as foundation for future studies of cell surface anchoring in other archaea and bacteria.
# TABLE OF CONTENTS

ACKNOWLEDGMENT ........................................................................................................ iv  

ABSTRACT ......................................................................................................................... vi  

LIST OF TABLES ............................................................................................................... xi  

LIST OF FIGURES ............................................................................................................ xii  

CHAPTER 1 ......................................................................................................................... 1  

INTRODUCTION .................................................................................................................. 1  

1.1 Archaea ....................................................................................................................... 1  

1.2 Archaeal Membrane .................................................................................................... 5  

1.3 Archaeal Protein Transport ........................................................................................ 9  

1.4 Protein Anchoring Mechanism .................................................................................. 20  

1.5 *Haloferax volcanii* as a model archaeon ..................................................................... 34  

CHAPTER 2 ......................................................................................................................... 37  

*Haloferax volcanii* archaeosortase is required for motility, mating, and C-terminal processing of the S-layer glycoprotein .................................................................................. 37  

2.2 Introduction .................................................................................................................. 38  

2.3 Materials and methods .............................................................................................. 42  

2.4 Result .......................................................................................................................... 50  

2.5 Discussion .................................................................................................................. 57  

2.6 Tables ......................................................................................................................... 63
Permuting the PGF-CTERM signature motif blocks both archaeosortase-dependent C-terminal cleavage and prenyl lipid attachment for the *Haloferax volcanii* S-layer glycoprotein ........................................................................................................................................ 70

CHAPTER 3 ......................................................................................................................................... 70

3.1 Abstract ......................................................................................................................................... 70
3.2 Importance ...................................................................................................................................... 71
3.3 Introduction ................................................................................................................................... 72
3.4 Materials and methods .................................................................................................................. 74
3.5 Result and discussion ..................................................................................................................... 81
3.6 Concluding remark ........................................................................................................................ 91
3.7 Figures ........................................................................................................................................... 93

CHAPTER 4 ......................................................................................................................................... 102

ArtA-dependent processing of a Tat substrate containing a conserved tripartite structure that is not localized at the C-terminus ........................................................................................................ 102

4.1 Abstract ......................................................................................................................................... 102
4.2 Introduction ................................................................................................................................... 103
4.3 Materials and methods .................................................................................................................. 107
4.4 Results .......................................................................................................................................... 116
4.5 Concluding Remarks ..................................................................................................................... 125
4.6 Figures ........................................................................................................................................... 128
LIST OF TABLES

Table 2.1: *H. volcanii* predicted ArtA substrates

Table 4.1: Peptide identifications in strains expressing Hvo_0405RR or Hvo_0405KK
LIST OF FIGURES

Figure 2.1: Depletion of ArtA results in a growth defect
Figure 2.2: High-resolution mass spectrometry analysis of S-layer glycoprotein
Figure 2.3: Phase contrast microscopy of H. volcanii H53 and ΔartA in liquid culture
Figure 2.4: Transmission electron microscopy
Figure 2.5: The ΔartA strain exhibits a motility defect
Figure 2.S1: Confirmation of the artA gene deletion
Figure 3.1: The conserved SLG PGF-motif is critical for ArtA-dependent processing
Figure 3.2: ArtA and conserved C-termina PGF motif are required for H. volcanii SLG lipid-modification
Figure 3.3: The ΔartA and slg$^{G796F, F797G}$ exhibit distinct cell morphology compared to the wild-type
Figure 3.4: SLG$^{G796F, F797G}$ forms thicker S-layer compared to wild-type
Figure 3.S1: The PCR screening confirmed the trpA gene insertion for identification of C-terminal PGF motif replacement mutants
Figure 3.S2: The SLG-PFG exhibits faster migration on LDS-PAGE gel compared to wild-type SLG
Figure 3.S3: H. volcanii SLG lipid-modification is ArtA-dependent
Figure 3.S4: The mobility shift of SLG is not attributed to lack of AglB-dependent N-glycosylation
Figure 3.S5: C-terminally unprocessed SLG-PFG causes an H. volcanii growth defect similar to the deletion of artA
Figure 4.1: The conserved SLG PGF-motif is critical for ArtA-dependent processing
Figure 4.2: ArtA and conserved C-terminal PGF motif are required for H. volcanii SLG lipid-modification
Figure 4.3: The ΔartA and slg$^{G796F, F797G}$ exhibit distinct cell morphology compared to the wild-type
Figure 4.4 SLG$^{G796F, F797G}$ forms thicker S-layer compared to wild-type
Figure 4.S1: Confirmation of hvo_0405 deletion mutant by PCR
Figure A.1: Confirmation of the pssA gene deletion
Figure A.2: Deletion of pssA results in a slight growth defect
Figure A.3: The pssA in trans expression restored the ΔpssA strain motility defect
Figure A.4: The ΔpssA exhibit irregular cell morphology compared to the wild-type
Figure A.5: Deletion of pssA gene result in the SLG (arrowhead) LDS-PAGE migration shift similarly observed in ΔartA and sigG796F,F797G
Figure A.6: H. volcanii SLG lipid-modification is significantly reduced upon pssA deletion
Figure A.7: The deletion of pssA results in the release of processed Hvo_0405 N-terminal fragment into the supernatant instead of anchored to the membrane
INTRODUCTION

1.1 Archaea

The archaea are classified as prokaryotes, organisms that bear no nucleus structure. This is one of many features they share with bacteria along with the presence of a circularized chromosome, operon system, and translation initiation sequence (Londei 2005). However, analysis of their universally conserved small subunit ribosomal RNA (SSU rRNA) sequences led to the proposal that bacteria and archaea are two distinct domains (Woese and Fox 1977). Later, this culminated in the reclassification of all living organisms into three distinct domains consisting of eukaryotes, bacteria, and archaea (Woese, Kandler et al. 1990). Prior to its classification into the third domain of life, certain members of the archaea had long been known to carry distinct cellular properties compared to other prokaryotes such as resistance to many antibiotics, such as beta-lactams, and the presence of highly distinct cytoplasmic membranes, which at the time were thought to result from convergent evolution in these highly diverse prokaryotes (Dridi, Fardeau et al. 2011, Braun 2014).

Many members of the domain archaea thrive in extremely harsh environments ranging from those with extremely high temperatures (hot springs/geysers) to low temperatures (ice lake glaciers), high salinity (salt lake/saltern pool) to acidic/alkaline water, or even high pressure (deep sea hydrothermal vent) (Hartzell, Millstein et al. 1999,
Rothschild and Mancinelli 2001, Cavicchioli 2006, Chaban, Ng et al. 2006, Pikuta, Marsic et al. 2007, Bowers and Wiegel 2011, Reed, Lewis et al. 2013). While it was previously known that methanogenic archaea are present in the human and animal gut microbiome, metagenomic analysis reveals that the range of habitats where archaea can be found is expansive including in the soil, oceans, and other parts of the human microbiome such as dental cavities (Karner, DeLong et al. 2001, Chaban, Ng et al. 2006, Ye, Liu et al. 2009, Matarazzo, Ribeiro et al. 2011, Gaci, Borrel et al. 2014, Bang and Schmitz 2015, Horz, Robertz et al. 2015). While the study of archaeal impact on human health is still in its nascent stage, recent research underscores the importance of archaea as part of a healthy gut microbiome (Nkamga, Henrissat et al. 2016, Miragoli, Federici et al. 2017), while others point out the potential roles of archaea as pathogens such as their involvement in causing gastrointestinal tract disorder (Lecours, Marsolais et al. 2014, Ishaq, Moses et al. 2016).

Besides lacking the nucleus structure, the similarities between archaea and bacteria are also extended towards their morphology and cellular structure. The archaeal cell size resembles that of bacteria and both cells lack any membrane-bound organelles such as mitochondria or an endoplasmic reticulum (ER). Hence, a majority of cellular processes for these two domains such as energy generation and protein transport occurs at the cytoplasmic membrane (Braun 2014, Jaehme and Slotboom 2015). The archaeal and bacterial cytoplasmic membrane also play a crucial role in providing the scaffold for surface structures that provide cellular stability such as the cell wall. However, the
component(s) that make up the cell wall structure in archaea and bacteria can be different. While the bacterial cell wall is made up of a peptidoglycan layer, many archaeal cell walls, also referred to as the S-layer, consist of paracrystalline arrays of proteinaceous subunits known as S-layer glycoprotein (SLG) (Sára and Sleytr 2000, Vollmer and Höltje 2004). The halobacterial SLG was the first prokaryotic protein discovered to be glycosylated and is thought to be one of the most highly abundant proteins in the biosphere (Sára and Sleytr 2000). Alternatively, some archaeal cell walls can be made up of other structures such as pseudomurein, glutaminylglycan, heterosaccharide, or methanochondroitin, which resembles the bacterial peptidoglycan but is different in chemical structure (Meyer and Albers 2001, Visweswaran, Dijkstra et al. 2011, Klingl 2014, Perras, Wanner et al. 2015). Nonetheless, some archaeal species, such as Ignococcus hospitalis have been discovered to possess a periplasmic space, and a distinct outer sheath structure that mimics the double membrane features of gram-negative bacteria (Rachel, Wyschkony et al. 2002).

Interestingly, the archaea also share many similarities with eukaryotes regarding cellular machinery such as informational genes involved in DNA replication (Olsen and Woese 1996), regulation of transcription with histones, chromatin-associated proteins (Bell, Botting et al. 2002), and having unique ribosomal protein families for protein translation (Rivera, Jain et al. 1998, Lecompte, Ripp et al. 2002). Although not necessarily providing direct evidence, these similarities lend credence to the previous proposal of eukaryote emergence from archaea, previously known as the eocyte hypothesis (Archibald 2008). Based on ribosomal structure analysis, the eocyte hypothesis proposed that the
eukaryotic lineage recently emerged from within the deep branches of the archaea domain, specifically from the Crenarchaeota phylum, resulting in only two primary domains of life; bacteria and archaea (Lake, Henderson et al. 1984). This hypothesis was eventually eclipsed by the then widely-known Woese's “three-domains of life” dogma that put eukaryotes, bacteria, and archaea as distinct primary domains (Woese, Kandler et al. 1990).

The current concept of the tree of life began to be called into question as different methods used for phylogenetic analysis of the same set of conserved core genes led to different outcomes in which some supported the three-domain of life dogma (Ciccarelli, Doerks et al. 2006) while others indeed supported the eocyte hypothesis (Katoh, Kuma et al. 2001). Recent expanded data sets and improvements in phylogenetic analysis methods, which have taken into account the variation in the evolutionary rate of different protein sequence sites, have given a fresh breath to the eocyte hypothesis (Yang and Roberts 1995, Tourasse and Gouy 1999, Cox, Foster et al. 2008). This provides compelling evidence for the re-classification of the tree of life into only two primary domains, suggesting the archaeal links in the origin of eukaryotes and perhaps explaining their shared homologs in cellular machinery. The recent discovery of a novel archaeal phylum, Lokiarchaeota, further supports the eocyte hypothesis. This is based on the discovery of many proteins encoded by genes in this phylum that share homology to the eukaryotic signature protein (ESP) such as those involved in membrane remodeling, protein trafficking and cytoskeletal remodeling (Spang, Saw et al. 2015). Additionally, recent findings of the ‘Asgard’ superphylum indicate that the presence of genes encoding ESP in these archaea are widely distributed
(Zaremba-Niedzwiedzka, Caceres et al. 2017), underscoring the importance of studying this domain to perhaps answer the longstanding conundrum on the evolution of eukaryotes.

1.2 Archaeal Membrane

Archaea belong to a unique domain in which while the complexity of its cellular machinery is close to that found in eukaryotes, the cell is still morphologically similar to bacteria, resulting in chimeric characteristics found in eukaryotes and bacteria. Nonetheless, deserving to be classified into its own domain, archaea also carry unique traits, such as the component that makes up their cytoplasmic membrane. First, cytoplasmic membrane radiolabeling in the haloarchaeon, *Halobacterium salinarum*, reveal that the chirality of the archaeal glycerol backbone consists of the enantiomeric glycerol-1-phosphate (G-1-P) compared to the glycerol-3-phosphate (G-3-P) found in bacteria and eukaryotes (Kates, Wassef et al. 1970, Ventosa and Oren 1996). Despite the presence of G-3-P specific enzyme in this organism (specifically, the glycerolphosphate dehydrogenase and glycerol kinase) (Wassef, Sarner et al. 1970), it was postulated that following the ether lipid chain alkylation, the G-3-P glycerol lipid backbone was later inverted into G-1-P (Kakinuma, Yamagishi et al. 1988). However, later identification of G-1-P specific enzymes in *Methanobacterium thermoautotrophicum* (Nishihara and Koga 1995, Nishihara and Koga 1997) and other various archaea cell-free homogenate indicate that the alkylated G-1-P lipid backbone can be directly synthesized without stereochemical inversion from G-3-P (Zhang, Daniels et al. 1990, Nishihara, Yamazaki et al. 1999).
Secondly, another distinguishing feature between bacterial and eukaryotic versus archaeal membranes is the lipid chain that makes up the hydrophobic region of the cytoplasmic membrane. While bacterial and eukaryotic lipid chains consist of fatty acids (Raetz and Dowhan 1990), archaeal lipid chains consist of isoprenoid chains (De Rosa, Gambacorta et al. 1982, De Rosa, Gambacorta et al. 1983, De Rosa, Gambacorta et al. 1986, Schouten, Hopmans et al. 2000). In order to obtain optimal membrane fluidity, especially in response to temperature stress, bacterial adaptation is limited to the alteration of the length and saturation of their fatty acid composition (Gaughran 1947, Zhang and Rock 2008). However, in addition to altering the chain length, archaea can also transform their isoprenoid chain structure to further increase their membrane stability in response to the extreme environments they inhabit (see below) (Russell and Fukunaga 1990).

Thirdly, instead of being connected by an ester linkage, as found in bacteria and eukaryotes, *Halobacterium salinarum* membrane lipid radiolabeling revealed that the archaeal isoprenoid lipid chain is connected to the glycerol backbone moiety via an ether linkage (Kakinuma, Yamagishi et al. 1990, Zhang and Rock 2008). Once the unsaturated isoprenoid chain is synthesized through the mevalonate pathway (see below) (De Rosa, Gambacorta et al. 1986), it is linked to G-1-P moiety by a cytoplasmic prenyltransferease known as geranylgeranylgluceryl diphosphate (GGGP) synthase (Moldoveanu and Kates 1988), followed by its reduction into saturated isoprenoid chains (Poulter, Aoki et al. 1988). A study in *Methanothermobacter marburgensis* revealed that the second ether linkage formation of the isoprenoid lipid chain and the glycerol backbone is catalyzed by a

Another unique archaeal cytoplasmic membrane characteristic, predominantly found among thermophilic archaea such as *Thermoplasma acidophilum* and *Sulfolobus acidocaldarius*, is the presence of tetraether lipids that span the entire width of the cytoplasmic membrane forming an atypical lipid monolayer instead of the conventional lipid bilayer (Langworthy 1977, Gliozi, Rolandi et al. 1983, Elferink, de Wit et al. 1994, Komatsu and Chong 1998). Other thermophilic archaea can also have their membrane lipid isoprenoid chains cross-linked forming an H-shaped tetraether lipid (Jacquemet, Barbeau et al. 2009) or forming cyclopentane/cyclohexane-ringed structures within the isoprenoid chain (De Rosa, Esposito et al. 1980, Damsté, Schouten et al. 2002). These strategies are believed to increase the membrane packing, reduce the rotational flexibility of the isoprenoid chain, and decrease the overall membrane fluidity for a stable cytoplasmic membrane at higher temperatures (Shinoda, Shinoda et al. 2005, Benvegnu, Lemiègre et al. 2008).

The basic building block for isoprenoid synthesis in all domains of life is based on a single isoprene unit known as isopentenyl pyrophosphate (IPP) and its isomer, dimethylallyl pyrophosphate (DMAPP) (Mizioroko 2011). This five carbon isoprene unit can be synthesized either via the classical mevalonate (MVA) pathway or the mevalonate independent pathway 1-deoxy-D-xylulose 5-phosphate [DOXP/MEP]) pathway (Goldstein and Brown 1990, Rohmer 1999). In archaea, studies in *Thermoplasma*
acidophilum, S. acidocalcarous, and Halobacterium sp. revealed that the synthesis of IPP and DMAPP, is exclusively achieved via the mevalonate (MVA) pathway. (Langworthy, Mayberry et al. 1974, Mayberry-Carson, Langworthy et al. 1974, Kates 1977, Lange, Rujan et al. 2000). The archaeal MVA pathway involves a series of sequential biochemical reactions that begin with the conversion of acetate to acetyl-CoA molecules (Tachibana, Tanaka et al. 1996, De Cima, Rúa et al. 2005). The condensation reaction of two acetyl-CoA molecules produces acetoacetyl-CoA, which is converted to hydroxymethyl-glytaryl-CoA (HMG-CoA), and further converted to mevalonate (Caforio and Driessen 2016). Once phosphorylated, the mevalonate-5-phosphate is converted into isopentenyl phosphate (IP) via a decarboxylation reaction, which is followed by phosphorylation to produce IPP or converted into its isomeric form, DMAPP (Matsumi, Atomi et al. 2011, Dellas, Thomas et al. 2013). While the 5-carbon unit IPP and DMAPP can be condensed into varying isoprenoid lipid chain lengths (Wang and Oh numa 1999, Liang, Ko et al. 2002, Vandermoten, Haubr uge et al. 2009), the IPP and DMAPP are typically condensed into 20-carbon (geranylgeranyl) or 25-carbon (farnesylgeranyl) isoprenoid chains for the synthesis of archaeal cytoplasmic membrane (Ohnuma, Suzuki et al. 1994, Tachibana, Yano et al. 2000, Hemmi, Ikejiri et al. 2002).
1.3 Archaeal Protein Transport

While some proteins synthesized by the cells remain within the cytoplasm, some membrane proteins and partially hydrophobic proteins need to be inserted or transported across the cytoplasmic membrane. These integral membrane proteins or secreted proteins plays various critical roles for maintaining cellular function such as nutrient uptake, signal transduction, and degrading polymers. In order to ensure these proteins’ proper localization, the cells utilize several cytoplasmic membrane transport pathways such as the YidC-dependent membrane insertion pathway, the Sec-mediated transport pathway, and the Tat-mediated transport pathway. The process mediated by these pathways occurs at the cytoplasmic membrane of archaea and bacteria, while in eukaryotes, it takes place at the ER membrane.

1.3.1 YidC-dependent membrane protein insertion

A subset of membrane proteins is integrated into the cytoplasmic membrane using a specialized membrane insertion system known as the YidC insertase pathway (Dalbey, Kuhn et al. 2014). The prototypical *E. coli* YidC, represent a superfamily of membrane insertases of which homologs were found in bacteria (YidC), mitochondria (Alb3), and chloroplasts (OxaI) (Moore, Harrison et al. 2000, Serek, Bauer-Manz et al. 2004). The YidC/Alb3/OxaI insertase pathway mediates the membrane protein insertion by allowing the binding of its substrates to its inner leaflet membrane amphiphilic groove, reducing the substrates energy requirement for translocation across the cytoplasmic membrane (Dalbey,
Studies of *E. coli* YidC indicate that several of its membrane protein substrates include M13, the Pf3 coat protein, and the subunit c of the *E. coli* F\(_1\) F\(_0\) ATP synthase (Wolfe, Rice et al. 1985, Chen, Samuelson et al. 2002, Serek, Bauer-Manz et al. 2004, Yi, Celebi et al. 2004, Kuhn and Kiefer 2017).

A previous direct homology search of the YidC/Oxa1/Alb3 protein family in the archaea domain discovered limited homologs in the euryarchaeota phylum as their function as membrane insertases remained undetermined (Yen, Harley et al. 2001). However, a recent finding and structural analysis of the archaeal YidC-like protein, DUF106, in *Methanocaldococcus janaschii* (Mj0480) revealed that the archaeal homolog has limited sequence identity and different architecture compared to *E. coli* YidC (Borowska, Dominik et al. 2015). While the *E. coli* YidC is a large protein with a molecular weight of 60 kDa consisting of 6 transmembrane (TM) segments including a large periplasmic domain (P1), the archaeal YidC (Mj0480) has a smaller protein size of 23 kDa with only 3 TM segments and lacking the sizeable P1 domain (Borowska, Dominik et al. 2015). Nevertheless, the archaeal YidC still encodes the conserved TM domains that form the hydrophilic groove found in the bacterial YidC (Kumazaki, Chiba et al. 2014, Kumazaki, Kishimoto et al. 2014) and the P1 domain has been shown to be expendable for *E. coli* YidC activity (Jiang, Chen et al. 2003, Ravaud, Stjepanovic et al. 2008, Imhof, Kuhn et al. 2011, Kumazaki, Kishimoto et al. 2014, Borowska, Dominik et al. 2015). Despite the endogenous archaeal YidC substrates having yet to be identified, the putative archaeal YidC (Mj0480) is able to bind the subunit c of the F\(_0\)F\(_1\) ATPase, an *E. coli* YidC substrate, via direct interaction of
its hydrophilic region and the TM segment of the substrate nascent chain (Yen, Harley et al. 2001, Jiang, Chen et al. 2003, Borowska, Dominik et al. 2015). This suggests that Mj0480 is indeed a bona fide archaeal YidC insertase. A revised homology search in the archaea domain using Mj0480 found its homologs in the crenarchoeta and korarchaeota phyla, indicating that this protein is widely present in all domains of life (Kuhn and Kiefer 2017).

While YidC has been shown to function on its own, this protein has also been reported to integrate with the SecYEG complex forming the Sec holotranslocon complex which mediates the membrane insertion of the Sec substrates (see below) (Komar, Botte et al. 2015, Komar, Alvira et al. 2016). While the YidC substrates consist of small proteins with a small soluble domain and either one or two TM domains (Kuhn 1988, Wickström, Wagner et al. 2011), it was postulated that the YidC-Sec holotranslocon substrates includes larger and more complex membrane proteins (Yi, Jiang et al. 2003, Celebi, Yi et al. 2006, Zhu, Klenner et al. 2012).

1.3.2 Sec-mediated protein transport pathway

For all domains of life, proteins that are targeted to be translocated across the cytoplasmic membrane often possess a conserved N-terminal amino acid tripartite structure known as the signal sequence or signal peptide, which directs them to the membrane translocation system known as the Sec pathway (Pugsley, 1993; Driessen, 1994; Izard and Kendall, 1994; den Blaauwen and Driessen, 1996). The tripartite structure typically consists
of 18–26 amino acids beginning with positively charged residues (N-domain) followed by a hydrophobic stretch (H-domain) and polar residues (C-domain), in which resides the proteolytic cleavage site for the signal peptidase (SPase) (von Heijne, 1985; Jones et al., 1990; Izard and Kendall, 1994).

While there are several classes of signal peptidases, signal peptidase I (SPaseI) is universally conserved and its processing activity would completely remove the Sec signal peptide from the precursor protein (Date 1983, Cregg, Wilding et al. 1996, Zhang, Greenberg et al. 1997, Rose, Brüser et al. 2002, Dilks, Giménez et al. 2005, Tuteja 2005, Giménez, Dilks et al. 2007, De Castro, Ruiz et al. 2008, Uthandi, Saad et al. 2010, Abdul Halim, Stoltzfus et al. 2017). The SPaseI is a serine protease, in which the conserved serine and lysine residues form a catalytic dyad that is critical for its catalytic activity (Dalbey and von Heijne 1992, Sung and Dalbey 1992, Klenotic, Carlos et al. 2000). Upon Sec signal peptide cleavage by the SPaseI, the mature proteins are either released into the periplasm/extracellular space or remain associated with the cell through other surface anchoring mechanisms, such as carboxy-terminal (C-terminal) membrane anchoring.

While the Sec signal peptide was originally discovered in eukaryotes (Blobel and Sabatini 1971, Milstein, Brownlee et al. 1972), the components that are critical for the Sec-transport pathway were later identified through genetic and molecular biological studies in bacteria and eukaryotes (Emr, Hanley-Way et al. 1981, Oliver and Beckwith 1981, Denks, Vogt et al. 2014). In order to successfully transport proteins across the cytoplasmic membrane barrier, the Sec pathway is composed of multiple subunits that make up the Sec
translocon machinery. In bacteria, the Sec machinery that forms the protein-conducting channel complex consists of heterotrimeric SecY, SecE, and SecG proteins (Ito, Wittekind et al. 1983, Riggs, Derman et al. 1988, Nishiyama, Hanada et al. 1994). This protein channel contains a hydrophilic interior allowing for efficient translocation of Sec substrates across the hydrophobic barrier imposed by the cytoplasmic membrane (Simon and Blobel 1991, Crowley, Liao et al. 1994). The SecY protein (Sec61α in eukaryotes) is highly conserved across the three domains of life (Arndt 1992, Hartmann, Sommer et al. 1994, Rensing and Maier 1994, Kath and Schäfer 1995, Hanein, Matlack et al. 1996). For example, although the archaeal SecY/Sec61α homolog is more similar to the eukaryotic Sec61α, expression of this archaeal pore component from the methanogen, *Methanococcus vanielli*, in *E. coli* could complement the bacterial deletion phenotype (Auer, Spicker et al. 1991). Moreover, the conservation of the genomic localization of secE in prokaryotic domains has enabled the identification of the archaeal SecE/Sec61γ homolog in *M. genitalium* and *M. jannaschii*, which is actually more closely related to the eukaryotic Sec61γ (Hartmann, Sommer et al. 1994). This not only indicates the conservation of SecE/Sec61γ across domains of life, but also has provided a homology link between bacterial SecE and eukaryotic Sec61γ, which would not have been noted due to a lack of direct sequence homology between these two proteins (Hartmann, Sommer et al. 1994). However, while all domains have a third Sec channel component, it is not conserved across all life domains. While bacterial SecG and eukaryotic Sec61β lack any homology to each other, analysis of archaeal genomes revealed homologs of eukaryotic Sec61β, reflecting

In addition to the SecYEG/Sec61αβγ channel, there are also several auxiliary proteins that play crucial roles as part of the Sec translocon machinery. In bacteria, the SecA protein has been identified as an ATPase that drives the protein translocation through the SecYEG channel (Oliver and Beckwith 1981, Osborne, Clemons et al. 2004). However, there are no SecA homologs that have been identified in eukaryotes. Instead, the Sec protein translocation is driven by BiP, which is localized in the ER lumen (Nguyen, Law et al. 1991). Interestingly, despite the archaeal Sec channel components’ close homology to the eukaryotic Sec61p channel, no BiP homolog has been identified in archaea, which is perhaps due to the lack of ATP outside the cell (Pohlschröder, Prinz et al. 1997). Additionally, the SecA homolog is also not detected in archaea (Calo and Eichler 2011) despite their genome encoding functional SecD and SecF proteins (Pogliano and Beckwith 1994, Hand, Klein et al. 2006), which are suggested to be involved in stabilizing the membrane cycling of SecA according to studies of bacterial homologs (Economou, Pogliano et al. 1995, Duong and Wickner 1997). The lack of SecA or BiP ATPase homologs for the archaeal Sec-transport pathway has been suggested to be due to all of the Sec substrate membrane translocation in this domain occurring co-translationally, which does not require additional ATPase machinery (Pohlschröder, Prinz et al. 1997).
1.3.3 Signal Recognition Particles (SRP)

In all domains of life, the co-translational transport of Sec substrates requires a ribonucleoprotein complex known as the signal recognition particle (SRP), which binds to the N-terminal signal sequence and guides the nascent peptide chain to the Sec translocon (Halic, Becker et al. 2004, Grudnik, Bange et al. 2009). While the number of components that make up the SRP complex varies between domains of life, some of the components are conserved throughout all domains such as the homologs of SRP subunit SRP54, which play the primary role of binding and guiding the nascent peptide chain to the membrane (Moll, Schmidtke et al. 1999, Maeshima, Okuno et al. 2001, Pakhomova, Deep et al. 2002). The eukaryotes have the most complex SRP with 6 protein components; SRP68, SRP72, SRP54, SRP19, SRP14, and SRP9 including a 7S RNA (Keenan, Freymann et al. 2001). On the other hand, the bacteria have the simplest SRP with only a single protein, an SRP54 homolog (Ffh), and a 4.5S RNA (Buskiewicz, Kubarenko et al. 2005). Meanwhile, the archaeal SRP is composed of two protein subunits; SRP54 and SRP19 proteins and a 7S RNA, which are highly homologous to their eukaryotic counterparts (Zwieb and Eichler 2002). Nonetheless, while the SRP19 homolog is present in both the eukaryote and archaea domains, this protein is expendable for archaeal SRP function, reflecting the flexible and unique feature of SRP in the archaea domain (Diener and Wilson 2000, Yurist, Dahan et al. 2007).

Analysis of H. volcanii and Archaeoglobulus fulgidus SRP components confirm the interaction between the SRP19 and SRP54 proteins (Rose and Pohlschröder 2002),
which result in SRP-RNA conformational changes and its binding to the SRP54 (Bhuiyan, Gowda et al. 2000, Hainzl, Huang et al. 2002, Oubridge, Kuglstatter et al. 2002, Tozik, Huang et al. 2002, Egea, Napetschnig et al. 2008). However, it was proposed that the SRP54 may already have intrinsic affinity towards the SRP-RNA (Diener and Wilson 2000), which is consistent with the observation that the SRP19 deletion does not seem to impair protein transport as opposed to SRP54 deletion, which has deleterious effects in the cells (Rose and Pohlschröder 2002, Yurist, Dahan et al. 2007).

Besides the SRP, another important component of the Sec co-translational transport is the SRP receptor (SR), which is located at the cytoplasmic membrane where the Sec channel is localized (Egea, Shan et al. 2004). While the eukaryotic SR is composed of two proteins, SRα and SRβ proteins, both bacterial and archaeal SR receptors consist of a single protein, FtsY, a homolog of the eukaryotic SRα (Miller, Bernstein et al. 1994). The gene encoding the archaeal SR, FtsY, is essential (Haddad, Rose et al. 2005) since the SRP–ribosome nascent polypeptide complex (RNC) relies on the FtsY to guide the Sec substrate nascent polypeptide chain to the cytoplasmic membrane (Lichi, Ring et al. 2004, Egea, Tsuruta et al. 2008, Hainzl and Sauer-Eriksson 2015). The interaction between SRP and SR is regulated by GTP hydrolysis, in which the GTP bound FtsY-SRP complex is stable but dissociates upon GTP hydrolysis, allowing the delivery of the nascent peptide chain to the Sec channel for its translocation across the cytoplasmic membrane (Egea, Shan et al. 2004, Focia, Shepotinovskaya et al. 2004, Wild, Bange et al. 2016).
Sec substrate translocation across the cytoplasmic membrane does not always occur co-translationally. In prokaryotes, the protein translocation across the cytoplasmic membrane via the Sec channel can also involve those that have been fully synthesized (post-translational) (Ortenberg and Mevarech 2000, Irihimovitch and Eichler 2003). However, the size of the Sec channel pore only permits the translocation of proteins in an unfolded conformation (Park, Ménétret et al. 2014). Hence, these fully synthesized proteins require cytosolic chaperones such as SecB to keep them in a translocation-competent state (Josefsson and Randall 1981, Dale and Krebs 1999, Ortenberg and Mevarech 2000, Irihimovitch and Eichler 2003, Van den Berg, Clemons et al. 2004).

1.3.4 Twin arginine translocation (Tat)-mediated protein transport

An alternate protein transport pathway that also permits post-translational protein export across the cytoplasmic membrane and has been discovered in prokaryotes, chloroplast, and mitochondria is known as the Twin-arginine translocation (Tat) pathway (Burger, Gray et al. 2013, Cledon and Cline 2013, Berks 2015). However, in stark contrast to the Sec transport pathway, the Tat transport pathway is dedicated to cytoplasmic membrane translocation of proteins that are already in their folded conformation (Berks 2015). Additionally, while the Tat transport pathway is also an active transport process similar to the Sec transport pathway, the driving force for its substrates’ cytoplasmic membrane translocation is dependent on the proton gradient instead of ATP (Mould and Robinson 1991, Yahr and Wickner 2001).
The proteins that are exported across the cytoplasmic membrane via the Tat transport pathway, also known as Tat substrates, carry an N-terminal tripartite structure of a signal sequence that is reminiscent of the Sec signal peptide, albeit with a slightly longer signal sequence, a more positively charged N-domain, and a slightly less hydrophobic H-domain (Cristóbal, de Gier et al. 1999, Tjalsma, Bolhuis et al. 2000). Nonetheless, the most distinguishing feature of the Tat signal sequence is that it contains two highly conserved consecutive arginine residues between the charged N-domain and H-domain, specifically represented by the conserved (S/T)-R-R-x-F-L-K motif where x is a polar amino acid or glycine (Berks 1996). Mutation of these highly conserved arginine residues (RR) within this motif, even to amino acids of similar charge such as lysine, is sufficient to impede the Tat substrate translocation across the cytoplasmic membrane (Stanley, Palmer et al. 2000, Alami, Trescher et al. 2002, Giménez, Dilks et al. 2007). However, a single arginine residue mutation still permits the cytoplasmic membrane translocation of the Tat substrate (Hinsley, Stanley et al. 2001, Ize, Gérard et al. 2002).

The Tat-transport pathway involves several components that make up its translocon complex, which mainly involve two membrane protein families, TatA and TatC. They are distinguished by their structural features in that the TatA protein family is membrane-anchored by a single C-terminal TM domain while the TatC protein is an integral multi-TM-anchored membrane protein (Berks 2015). In the well-established Tat-transport pathway models of *E. coli* and plant chloroplasts, the Tat translocon complex is composed of the TatA, TatB and TatC protein subunits (Rodriguez, Rouse et al. 2013, Zhang, Wang
et al. 2014). The TatB protein belongs to the TatA protein family but has developed a distinct function in these organisms. Nonetheless, in plant thylakoids, the homology between TatA and TatB extended up half of the protein sequence and a suppressor mutation in *E. coli* TatA may substitute the function of its TatB (Blaudeck, Kreutzenbeck et al. 2005). This is consistent with the archaeal and some gram-positive bacteria’s Tat translocon complex assembly which only consist of TatA and TatC proteins (Yen, Tseng et al. 2002, Dilks, Rose et al. 2003, Goosens and van Dijl 2016). Despite this difference, paralogs of TatA and TatC have been identified in these organisms such as *E. coli* TatE, *B. subtilis* TatAd-TatCd, and *H. volcanii* TatAo-TatCo (Yen, Tseng et al. 2002, Dilks, Rose et al. 2003, Goosens and van Dijl 2016).

Despite it being suggested that the Tat transport pathway is utilized for nearly half of soluble protein secretion in haloarchaeal species, the translocation mechanism of Tat-transport pathway in this domain have not been fully elucidated (Bolhuis 2002, Rose, Brüser et al. 2002, Dilks, Giménez et al. 2005, Giménez, Dilks et al. 2007). However, studies in bacteria such as *E. coli* and chloroplast have provided some frameworks for the mechanism of the Tat-transport pathway in general. Summarily, the TatB and TatC proteins are preassembled at the cytoplasmic membrane, serving as acceptor proteins that recognize and interact with the Tat substrate signal sequence (Alami, Lüke et al. 2003). Upon the Tat substrates’ binding, the TatBC complex triggers the oligomerization of TatA proteins, the main component of the Tat translocon, allowing the translocation of the Tat substrates across the cytoplasmic membrane (Sargent, Gohlke et al. 2001). Once the Tat
substrate translocation is completed and processed by signal peptidase (see above), the TatA oligomer is disassembled from the TatABC complex (Mori and Cline 2002). Despite this molecular elucidation, the actual mechanism in which the Tat substrate is transported across the cytoplasmic membrane remains ambiguous. While some suggest that oligomerized TatA may form protein channels (Gohlke, Pullan et al. 2005), others propose the role of TatA in local destabilization of the membrane bilayer (Brüser and Sanders 2003, Rodriguez, Rouse et al. 2013).

1.4 Protein Anchoring Mechanism

1.4.1 Transmembrane anchoring

While some proteins that are translocated across the cytoplasmic membrane are released into the extracellular environment, others stay attached to the cell surface by a stretch of hydrophobic residues known as the transmembrane (TM) domain. Studies of TM anchoring in bacteria, reveal that there are a number of Tat substrates that encode either a C- or N-terminal hydrophobic domain allowing these proteins to be TM-anchored to the cytoplasmic membrane (Sargent, Berks et al. 2002, Hatzixanthis, Palmer et al. 2003). For example, the deletion of the C-terminal hydrophobic domain of the E. coli [NiFe] hydrogenase subunit HybO protein, leads to its release into the periplasm (Hatzixanthis, Palmer et al. 2003) while fusing this hydrophobic domain to the soluble periplasmic proteins, TorA and SufI, resulted in the membrane anchoring of these proteins instead of their secretion into the periplasm. (Hatzixanthis, Palmer et al. 2003). On the other hand,
the Tat substrates can also be N-terminally TM-anchored via their non-processed signal peptide. Both *Paracoccus denitrificans* Rieske iron sulfur protein (ISP) and *Salmonella enterica* TtrA protein carry the canonical Tat signal sequence, but lack the signal peptidase cleavage site (Bachmann, Bauer et al. 2006, De Buck, Vranckx et al. 2007, James, Coulthurst et al. 2013). Since their N-terminal signal peptides remained unprocessed upon cytoplasmic membrane translocation, it serves as the protein membrane anchor via the hydrophobic region. The fusion of the TtrA N-terminal signal peptide and soluble periplasmic *E. coli* AmiA protein also results in membrane anchoring without abolishing its catalytic activity.

The TM anchoring of proteins in archaea has also been studied previously such as the *H. volcanii* Tat substrates halocyanin 2 and halocyanin 3 proteins, which encode a C-terminal hydrophobic domain that is predicted to serve as their TM anchor (Giménez, Dilks et al. 2007). In order to confirm the role of this hydrophobic domain, truncated gene constructs were made of which the sequence encoding the halocyanin C-terminal TM domain was deleted and subsequently expressed. The Tat signal peptide processing of the truncated halocyanin mutants indicated that the C-terminal hydrophobic domain is not required for their cytoplasmic membrane translocation. However, while the wild-type halocyanin were found to be localized at the cytoplasmic membrane, the truncated halocyanin mutants were released into the extracellular environment indicating that the C-terminal hydrophobic domain is indeed critical for their membrane anchoring (Giménez, Dilks et al. 2007). Another protein that depends on the C-terminal TM domain for
membrane anchoring is the SlaB, a subunit of *Sulfolobus acidocaldarius* and *Sulfolobus solfataricus* S-layer (Grogan 1996, Veith, Klingl et al. 2009, Albers and Meyer 2011, Sleytr, Schuster et al. 2014). Protein modelling revealed that the SlaB protein structure consists of multiple beta sandwich domains, followed by a coiled-coil domain, and a C-terminal TM domain (Grogan 1996, Veith, Klingl et al. 2009). Based on these structure, it was proposed that the SlaB protein serves as a membrane anchor for another S-layer subunit, SlaA. The SlaB coiled-coil domain forms a rigid structure that protrudes outward and becomes the base structure for SlaA protein binding, hence anchoring both S-layer subunits to the cell surface (Veith, Klingl et al. 2009).

1.4.2 *N*-terminal covalently lipid anchoring

Another mechanism of protein anchoring to the cytoplasmic membrane is via the covalent anchoring of the protein N-terminus with a lipid moiety known as a lipoprotein. Across the bacterial and archaeal domains, lipoproteins are identified by their signal peptide carrying a conserved motif known as a lipobox, which contains canonical leucine-alanine-glycine-cysteine (LAGC) residues at the -3 to +1 position with the +1 cysteine conserved in all lipoproteins (Giménez, Dilks et al. 2007, Thompson, Widdick et al. 2010). In bacteria, upon translocation across the cytoplasmic membrane, either by the Sec transport pathway or the Tat transport pathway, the +1 cysteine sulphydryl group side chain forms a thioether linkage with a diacylglycerol moiety tethering the lipoprotein into the cytoplasmic membrane with two acyl chains (Gan, Gupta et al. 1993, Sankaran and Wu
1994, Qi, Sankaran et al. 1995). This reaction is catalyzed by an acyltransferase known as prolipoprotein diacylglycerol transferase (Lgt) (Mao, Zhao et al. 2016). Subsequently, the lipoprotein undergoes signal peptide processing by the signal peptidase II (SPaseII) LspA, which cleaves off the signal peptide immediately upstream of the diacylated cysteine residue. In gram-negative and some gram-positive bacteria, the processed diacylated lipoprotein undergoes another lipid modification with the third acyl group via amide linkage to the amino group of the +1 cysteine residue. (Gupta and Wu 1991, Sutcliffe and Harrington 2004, Babu, Priya et al. 2006, Rezwan, Grau et al. 2007). This reaction is catalyzed by apolipoprotein N-acyltransferase (Lnt) resulting in the mature lipoprotein being N-terminally lipid anchored by three acyl chain covalent linkage (N-acyl S-diacylglycerol Cys) (Narita and Tokuda 2016).

In gram-negative bacteria, some lipoproteins need to be localized to the outer membrane (OM). These OM lipoproteins are transported across the periplasmic region via the lipoprotein localization (Lol) transport pathway. The Lol transport pathway in *E. coli* consists of several components located at the inner membrane (LolC, LolD and LolE), which form an ABC-transporter complex to release lipoproteins from the inner membrane; at the periplasm (LolA), which serve as chaperone proteins; and at the outer membrane (LolB), which functions as a Lol receptor protein (Collin, Guilvout et al. 2011). The lipoprotein targeted to the OM lipoproteins carries a specific Lol-sorting signal within its N-terminal signal sequence (Yamaguchi, Yu et al. 1988, Schulze and Zückert 2006).
Among the earliest lipoproteins identified in archaea was the haloalkophilic archaeon, *Natronobacterium pharaohnis*’s, halocyanin, a blue copper protein involved in cellular electron transfer reactions (Mattar, Scharf et al. 1994). While the analysis of the gene sequence encoding the halocyanin predicted that the expected protein molecular weight is 17, 223 kDa, mass spectrometry analysis of purified halocyanin yielded a lower molecular weight of 15, 456 kDa, suggesting a post-translational processing. Since the halocyanin peptide sequence comparison with other lipoproteins is consistent with it carrying the N-terminal lipobox motif and its N-terminus is inaccessible for protein sequencing, it was suggested that the halocyanin precursor is N-terminally modified with a diphytanyl glycerol diether and undergoes signal peptide processing. Indeed, the calculated peptide mass by considering these post-translational modifications is consistent with the peptide mass value detected by the mass spectrometry of the purified halocyanin, indicating that this protein is an archaeal lipoprotein.

As the genomic sequences of various archaea became available in later years, many putative lipoproteins containing the signal peptide with a putative lipobox motif were identified in the euryarchaea phylum, especially among halophilic archaea, despite being predicted to be very limited or absent among the crenarchaea phylum. (Bolhuis 2002, Falb, Pfeiffer et al. 2005, Giménez, Dilks et al. 2007, Storf, Pfeiffer et al. 2010). The *in vivo* studies of the *H. volcanii* Tat transport pathway indicate that a majority of haloarchaeal lipoprotein are secreted via this pathway based on the importance of their highly conserved +1 cysteine residue for proper processing of their N-terminal signal peptide and protein
stability (Giménez, Dilks et al. 2007, Storf, Pfeiffer et al. 2010). While some of these cysteine replacement lipoprotein mutants remain cell-associated due to the anchoring via the hydrophobic stretch of the unprocessed signal peptide, other H. volcanii lipoproteins, such as DsbA and Mbp, have been reported to be released into the supernatant upon cysteine substitution (Giménez, Dilks et al. 2007). Hence, despite the conserved function of the lipobox motif, some components of lipid modification and signal peptide processing of archaeal lipoproteins may differ slightly from bacterial lipoprotein anchoring machineries. Since the archaeal homologs of Lgt, Lsp (SPaseII), and Lnt have not been identified in this domain, it will be interesting to identify the archaeal lipoprotein biosynthesis enzymes, which may reflect a novel lipoprotein biosynthesis due to this domain’s unique membrane structure (Szabo and Pohlschroder 2012).

In contrast to the SPaseI, which homologs are present across the domains of (Tokunaga, Loranger et al. 1983, Yamagata, Daishima et al. 1983), the SPaseII Lsp homologs have thus far only been identified in bacteria (Storf, Pfeiffer et al. 2010, Buddelmeijer 2015). Deletion of SPaseII in B. subtilis reveals that the gene is non-essential but is critical for cell growth at low (15°C) and high temperatures (48°C) (Tjalsma, Kontinen et al. 1999). While SPaseII is similar to the SPaseI in which both signal peptidase are membrane bound proteases, site directed mutagenesis of several residues that made up the SPaseII catalytic site revealed its proteolytic properties as aspartic acid protease (Tjalsma, Zanen et al. 1999).
1.4.3 Protein-protein interaction

An alternative strategy for the proteins to remain cell-associated is by binding or associating with other proteins that are already anchored to the cytoplasmic membrane. This protein-protein interaction strategy is the basis for the assembly of archaeal filamentous surface structures such as pili, flagella, and in some cases, its cell wall. The archaeal pili is assembled from the polymerization of type IV pilin protein subunits, which carry a distinct N-terminal tripartite structure in the signal peptide with the motif G/A/S at the -1 position and a charged amino acid at the -2 position (Szabo, Stahl et al. 2007, Imam, Chen et al. 2011). Upon translocation across the cytoplasmic membrane via the Sec-transport pathway, archaeal type IV pilins undergo signal peptide processing by a third class of signal peptidases known as the prepilin peptidase/signal peptidase III (SPaseIII) (Georgiadou, Pelicic et al. 2014, Makarova, Koonin et al. 2016). In contrast to the SPaseI and SpaseII signal peptide processing, the prepilin peptidase/SPaseIII processing targets a cleavage site preceding the signal peptide H-domain (Strom and Lory 1993, Bardy and Jarrell 2002, Albers, Szabó et al. 2003, Pohlschroder and Esquivel 2015). This results in the signal peptide H-domain being retained as part of the mature type IV pilin with its hydrophobicity serving as the interface for the protein-protein interactions between other type IV pilins forming the central hydrophobic core during pilus polymerization (Craig, Volkmann et al. 2006, Hansen and Forest 2006, Albers and Pohlschröder 2009). The importance of this hydrophobic domain is underscored by the discovery of a subset of *H. volcanii* type-IV pilins, PilA, which encode a highly conserved 30-amino acid sequence
within its signal peptide h-domain (Esquivel, Xu et al. 2013). While swapping the pilin PilA h-domain with the archaeal flagellin signal peptide h-domain does not affect the stability and signal peptide processing of the pilin, the pilins are no longer able to polymerize to form the pilus structure (Esquivel, Xu et al. 2013). In addition to the pilins, the archaeal type-IV pilus assembly is also dependent on the PilB ATPase, and PilC, a transmembrane protein that is hypothesized to serve as the cytoplasmic membrane anchor for the pilus structure (Pellicic 2008, Lassak, Ghosh et al. 2012, Takhar, Kemp et al. 2013).


Originally involved in the type-IV pili biosynthesis, SPaseIII was later discovered to be involved in the processing of a broad range of proteins that carry the prepilin-like signal sequence, hence allowing assembly of type IV pili-like surface structures that can serve various functions such as mediating surface adhesion, twitching motility through pili retraction, extending surface area for nutrient uptake (bindosome), DNA binding, and type-II protein secretion (Pugsley and Dupuy 1992, Strom and Lory 1993, Lory and Strom 1997,
Another example of this anchoring strategy for surface proteins is the assembly of archaeal flagella, which is assembled from protein-protein interactions between its flagellin protein subunits. Analogous to the pilin protein, the flagellin also carries a signal peptide that is recognized and processed by the prepilin peptidase/SPaseIII (Kalmokoff and Jarrell 1991, Bardy, Mori et al. 2002, Bardy, Eichler et al. 2003, Tripepi, Imam et al. 2010). Analysis of the archaeal flagellin genomic locus also revealed several key components of flagella biosynthesis machineries that share homology to that of the type-IV pili assembly components. For example, the FlaI protein, which has been shown to exhibit ATPase activity (Ghosh, Hartung et al. 2011, Reindl, Ghosh et al. 2013), is homologous to the PilB ATPase, and the FlaJ protein, is proposed to serve as a membrane anchor protein similar to the PilC protein (Bayley and Jarrell 1998, Peabody, Chung et al. 2003). It was proposed that FlaI and FlaJ proteins form a complex that is responsible for the flagellar rotation (Reindl, Ghosh et al. 2013). Indeed, importance of these proteins for flagella biosynthesis is demonstrated when the deletion of the genes encoding the SPaseIII (PibD/FlaK), FlaI, and FlaJ proteins result in non-flagellated cells and defects in flagella assembly (Bardy and Jarrell 2002, Thomas, Mueller et al. 2002, Chaban, Ng et al. 2007, Tripepi, Imam et al. 2010). Recently, one of the critical machineries in Sulfolobus acidocaldarius flagella
biosynthesis, FlaF, has been shown to bind directly to its cell wall, the S-layer (Banerjee, Tsai et al. 2015).

The assembly of *S. acidocaldarius* and other *Sulfolobales* species’ cell wall, the S-layer, is also dependent on the protein-protein interactions between its two subunits, SlaA and SlaB proteins (Veith, Klingl et al. 2009). The SlaA protein remains associated to the cell surface by binding to the SlaB protein, which is anchored to the cytoplasmic membrane. While the exact molecular interaction between SlaA and SlaB proteins has not been completely elucidated, two hypothetical mechanisms of protein-protein interaction were proposed; the first model proposed that the SlaA protein binds to the N-terminal domain of the SlaB protein, and the second model suggests that the SlaA forms a mushroom-headed rivet structure by binding to the extended SlaB stalk domain (Veith, Klingl et al. 2009).

1.4.4 Sortase-mediated cell wall anchoring

In addition to the membrane anchoring and interaction with other cell-associated proteins, the surface proteins can also remain cell-associated by covalently-anchoring to the peptidoglycan layer of bacterial cell wall (Schneewind, Model et al. 1992). The mechanism of protein anchoring to the cell wall is catalyzed by a transpeptidase known as sortase A (SrtA), named after its role in ‘sorting’ its substrates to the cell wall (Mazmanian, Liu et al. 1999, Ton-That, Liu et al. 1999). While SrtA homologs were initially found in *S. aureus*, further genomic analyses of this protein homolog revealed that it is present in
almost all gram-positive bacteria such as *Actinomyces naeslundii, Bacillus subtilis, Enterococcus faecalis, and Streptococcus pneumoniae*, and anchors a diverse group of substrates as few as 2 in *Clostridium acetobutylicum* to as many as 43 in *L. monocytogenes* (SJÖDAHL 1977, Eliasson, Olsson et al. 1988, Schneewind, Mihaylova-Petkov et al. 1993, Pallen, Lam et al. 2001, Palmqvist, Foster et al. 2002, Boekhorst, de Been et al. 2005).

The *Staphylococcus aureus* SrtA is the prototypical ‘housekeeping sortase’, which recognizes and anchors substrate proteins carrying a distinct C-terminal tripartite structure consisting of a highly conserved LPXTG motif, a hydrophobic domain, and positively-charged residues, also known as the cell wall sorting signal (CWSS) (Schneewind, Mihaylova-Petkov et al. 1993). In addition to the CWSS, the SrtA substrate also carries the N-terminal Sec signal peptide, which is cleaved by SPaseI upon translocation across the cytoplasmic membrane via the Sec translocon. The SrtA substrates also undergo C-terminal proteolytic cleavage by SrtA, specifically between the threonine (T) and glycine (G) residue of the LPXTG motif, after its cytoplasmic membrane translocation (Navarre and Schneewind 1994). This is followed by the formation of an acyl-enzyme intermediate between the SrtA and its substrate via the SrtA’s catalytic cysteine residue and the substrate’s carboxyl terminus of threonine, forming a thioester bond linkage (Ton-That, Liu et al. 1999). Next, the amino group of the pentaglycine within Lipid II, a peptidoglycan precursor, proceeds to break the thioester bond linkage of the acyl-enzyme intermediate, and transfer the cleaved SrtA substrate to the Lipid II moiety. The Lipid II with the covalently linked SrtA substrate is later incorporated into the growing peptidoglycan chain,
which ultimately anchors the SrtA substrate to the peptidoglycan cell wall (Spirig, Weiner et al. 2011, Siegel, Reardon et al. 2016).

In addition to the housekeeping SrtA protein, many bacterial species encode additional sortase paralogs that are grouped into distinct classes that recognize different types of substrates carrying distinct CWSS motifs (Comfort and Clubb 2004, Spirig, Weiner et al. 2011). For example, the class B sortase recognizes the substrate carrying the C-terminal tripartite structure with a NPQTN motif (Schneewind, Mihaylova-Petkov et al. 1993, Zong, Mazmanian et al. 2004). The sortases that belong to this class are those that share homology to the S. aureus Sortase B (SrtB), the second sortase to be characterized (Maresso, Chapa et al. 2006). The genomic localization of the gene encoding SrtB and its substrates are located in the iron-regulated surface determinants (isd) gene locus, which is important for iron binding and uptake by the cells (Mazmanian, Ton-That et al. 2002). The studies in S. aureus and Bacillus anthracis indicate that of the proteins anchored by the SrtB, such as the surface protein IsdC, most are critical for extracellular iron-binding and uptake by the cells (Mazmanian, Ton-That et al. 2002, Maresso, Chapa et al. 2006). As suggested by the locus where the genes encoding SrtB and its substrates were found, their expression is regulated in response to iron availability in the extracellular environment (Mazmanian, Ton-That et al. 2002). However, it has been reported that the Streptococcus pyogenes SrtB, despite having sequence similarities to the class B sortase, is involved in pilin polymerization, which is commonly associated with class C sortases (Kang, Coulibaly et al. 2011). The class C sortases are unique since they are not only involved in anchoring
their substrates to the cell wall, but also mediate pilin polymerization for assembly of the pilus structure. The role of sortase in this class was first discovered in *Corynebacterium diphtheria*, which released its pili upon enzymatic digestion of its peptidoglycan cell wall with murein hydrolase (Ton-That and Schneewind 2003). This suggests that its pili are anchored to the cell wall in an analogous manner to the SrtA substrates (Ton-That and Schneewind 2003). The well-characterized sortase-catalyzed *C. diphtheria* pilus, consists of three sortase substrates, SpaA, SpaB, and SpaC. The SpaC pilin makes up the tip of the pilus, followed by the SpaA pilin, the major structural component of the pilus, and SpaB pilin, the base of the pilus. In additional to the N-terminal signal peptide and CWSS, containing a LPLTG motif, these pilin proteins also contain a conserved N-terminal pilin motif denoted by ‘WXXXVXVYPK’ in which the conserved lysine (K) residue is critical for the sortase-catalyzed pilin polymerization (Ton-That and Schneewind 2003, Ton-That and Schneewind 2004). The class D sortase is predominantly found in *Bacillus spp.*, and is involved in cell wall anchoring of surface proteins that are important, but not critical for spore formation under oxygen stress conditions (Marraffini and Schneewind 2006).

Finally, class E and F sortases have not been fully characterized. However, the *C. diptheriae* class F sortase may carry out a similar role to the class C sortase in anchoring the polymerized pilin to the cell wall (Ton-That and Schneewind 2003). Further genomic analysis also found sortase homologs in gram-negative bacteria, such as *S. putrefasciens*, with putative sortase substrates containing domains that bind to the extracellular matrix glycosaminoglycan hyaluronic acid and another substrate containing a von Willebrand
factor type A (vwa) domain (often associated with ligand binding in Eukarya) (Pallen, Lam et al. 2001).

The in silico analyses have identified an analogous sortase-like anchoring system, the exosortase/archaeosortase system, in gram-negative bacteria and archaea, respectively (Haft, Paulsen et al. 2006, Haft, Payne et al. 2012). While sortase homologs have not been identified in the organisms that encodes these anchoring systems, their genomes do encode putative substrates with a C-terminal tripartite architecture that is reminiscent of the sortase substrate CWSS. For example, in gram-negative bacteria, several proteins carry a C-terminal PEP motif followed by a hydrophobic stretch, and a positively charged residue (PEP-Cterm) that resembles the C-terminal of SrtA substrate LPXTG motif. By using the Partial Phylogenetic Profiling method to analyze the genomes that encode these PEP-Cterm proteins, an enzyme known as exosortase was identified and proposed to be responsible for processing and anchoring these PEP-Cterm proteins. Additionally, a distant archaeal homolog of exosortase, archaeosortase A, was also identified in the euryarchaeota phylum (Haft, Payne et al. 2012). Interestingly, the genomes that encode ArtA also encode proteins with C-terminal tripartite structure albeit with a distinct PGF motif (PGF-Cterm), which include the S-layer glycoproteins that were previously thought to be anchored via its C-terminal TM domain (Haft, Payne et al. 2012). Despite lacking homology to the sortase, the exo/archaeosortase is proposed to carry out similar proteolytic cleavage reaction to that of sortase based on the arrangement of amino acid residues that made up their putative the catalytic triad (Ton-That, Mazmanian et al. 2002, Haft, Payne et al. 2012). Hence, it was
hypothesized that the exosortase/PEP-C-term and ArtA/PGF-C-term represent the anchoring system analogous to the sortase/LPXTG in gram-negative bacteria and archaea, respectively (Haft, Paulsen et al. 2006, Haft, Payne et al. 2012). Nonetheless, due to the lack of peptidoglycan cell wall in the gram-negative and archaea, the exosortase and archaeosortase anchoring system may anchor their cleaved substrates differently from the sortase-anchored proteins. A similar in silico analysis also predicted a rhomboid family of serine proteases (rhombosortase) that recognize and process proteins carrying a distinct C-terminal tripartite structure (GlyGly-C-term) (Haft and Varghese 2011). The rhombosortase/GlyGly-C-term is found in many important bacterial genera such as *Shewanella*, *Vibrio*, *Acinetobacter*, and *Ralstonia* spp. Nonetheless, these findings remain as in silico predictions in which in vivo and in vitro studies will be crucial in confirming these predictions as well as elucidating the functions of this anchoring system.

1.5 *Haloferax volcanii* as a model archaeon

*Haloferax volcanii* is one of the model organisms that has been widely used to investigate various cellular processes in archaea such as protein transport (Rose and Pohlschröder 2002, Pohlschröder, Dilks et al. 2004, Dilks, Giménez et al. 2005), protein glycosylation (Konrad and Eichler 2002, Calo, Kaminski et al. 2010), lipid modification (Konrad and Eichler 2002, Abdul Halim, Karch et al. 2015), DNA repair, recombination, and replication (McCready 1996, Norais, Hawkins et al. 2007, Delmas, Shunburne et al. 2009). This organism belongs to a group of archaea known as haloarchaea, which
encompass those that thrive in high salt environments. Originally isolated from the Dead Sea, *Haloferax volcanii* is categorized as a moderate haloarchaeon since it requires “only” 2-3 molar salt as opposed to extreme haloarchaea, such as *Haloquadrartum walsbyi* that can survive at salt concentrations near saturation (Mullakhanbhai and Larsen 1975, Bolhuis, Palm et al. 2006). In order to survive the high salt environment, *Haloferax volcanii* uses a salt-in strategy, in which it accumulates KCl in the cytoplasm to counteract the osmotic pressure posed by the extracellular NaCl. At the molecular level, haloarchaeal proteins are more acidic than non-haloarchaeal proteins providing a hydration shell to protect the proteins from denaturation in a high salt environment (Oren 2008).

In addition to its moderate salt requirements for growth, other properties that make *Haloferax volcanii* suitable to be studied in the lab are that it can grow aerobically, having an optimum growth temperature of 45 °C, with a doubling time of about 2 hours. Moreover, compared to other haloarchaea, such as *Halobacterium salinarum, H. volcanii* has a stable genome with a low number of insertion sequence (IS) elements that have the potential to alter the genome composition (Pfeifer and Blaseio 1990, Hartman, Norais et al. 2010). The *Haloferax volcanii* genome also encodes biosynthesis genes for a majority of its required amino acids, making it possible to be grown in minimal media and creating auxotrophs for genetic selections (Allers and Ngo 2003, Allers, Ngo et al. 2004). Many genetic tools have been developed for this haloarchaeon such as the introduction of suicide vectors with selectable markers for making gene deletions based on the ‘pop-in pop-out’ strategy, development of overexpression vectors with an inducible promoter for controlled gene
complementation (Allers, Ngo et al. 2004, Allers, Barak et al. 2010), and the recent development of transposon mutagenesis (Kiljunen, Pajunen et al. 2014). The biochemical protocols have also been well-developed (Cline, Lam et al. 1989, Cline, Schalkwyk et al. 1989, Dyall-Smith 2004) with protein purification under native or denaturing conditions (Dilks, Giménez et al. 2005, Humbard, Miranda et al. 2010) and mass spectrometry analysis capabilities (Esquivel and Pohlschroder 2014). Electron microscopy techniques to analyze the cell surface structure has also been well-developed (Tripepi, Esquivel et al. 2013)
CHAPTER 2

*Haloferax volcanii* archaeosortase is required for motility, mating, and C-terminal processing of the S-layer glycoprotein

This chapter has been published in:


2.1 Abstract

Cell surfaces are decorated by a variety of proteins that facilitate interactions with their environments and support cell stability. These secreted proteins are anchored to the cell by mechanisms that are diverse, and, in archaea, poorly understood. Recently published *in silico* data suggest that in some species a subset of secreted euryarchaeal proteins, which includes the S-layer glycoprotein, is processed and covalently linked to the cell membrane by enzymes referred to as archaeosortases. *In silico* work led to the proposal that an independent, sortase-like system for proteolysis-coupled, carboxy-terminal lipid modification exists in bacteria (exosortase) and archaea (archaeosortase). Here, we provide the first *in vivo* characterization of an archaeosortase in the haloarchaeal model organism *Haloferax volcanii*. Deletion of the artA gene (HVO_0915) resulted in multiple biological phenotypes: (a) poor growth, especially under low-salt conditions, (b) alterations in cell shape and the S-layer, (c) impaired motility, suppressors of which still exhibit poor growth, and (d) impaired conjugation. We studied one of the ArtA substrates, the S-layer
glycoprotein, using detailed proteomic analysis. While the carboxy-terminal region of S-layer glycoproteins, consisting of a threonine-rich O-glycosylated region followed by a hydrophobic transmembrane helix, has been notoriously resistant to any proteomic peptide identification, we were able to identify two overlapping peptides from the transmembrane domain present in the ΔartA strain but not in the wild-type strain. This clearly shows that ArtA is involved in carboxy-terminal posttranslational processing of the S-layer glycoprotein. As it is known from previous studies that a lipid is covalently attached to the carboxy-terminal region of the S-layer glycoprotein, our data strongly support the conclusion that archaeosortase functions analogously to sortase, mediating proteolysis-coupled, covalent cell surface attachment.

2.2 Introduction

Prokaryotes secrete a wide variety of proteins across the cytoplasmic membrane. Some secreted proteins, such as many toxins and polymer-degrading enzymes, are released into the extracellular environment; however, most remain attached to the cells, by intercalating a terminal transmembrane segment into the cell membrane; through a covalent link between an amino-terminal cysteine and the lipid bilayer; or by associating with other cell-bound proteins (Szabo and Pohlschroder 2012). Recent in silico data have suggested that archaea also anchor a portion of membrane-bound proteins through a covalent association of their carboxy-termini with the lipid bilayer (Haft, Payne et al. 2012).

The genomes of some euryarchaeal species contain a gene that encodes a predicted transpeptidase with multiple transmembrane segments, archaeosortase A (Haft, Payne et
This archaeosortase is distantly related to a predicted bacterial exosortase, which, despite a lack of sequence similarity, is believed to show functional similarities to sortase (Haft, Paulsen et al. 2006). All three proteins – sortase, exosortase, and archaeosortase – are predicted to process target proteins, hereafter called their substrates, with removal of a short, hydrophobic segment. As was first shown for sortase A from *Staphylococcus aureus* (Mazmanian, Liu et al. 1999), sortases recognize a tripartite structure at the C-terminus consisting of a signature motif, a transmembrane alpha helix domain, and a cluster of basic residues.

Sortase A recognizes the specific signature motif (LPXTG), which is located immediately upstream of a transmembrane alpha helix domain in the substrate (Mazmanian, Liu et al. 1999). To date, the majority of Gram-positive bacterial species examined encode homologs of this sortase, which are now known to play a critical role in several basic biological processes (Spirig et al., 2011).

Sortase A substrates have amino-terminal signal peptides that are recognized by the machinery of the Sec protein transport pathway, which carries proteins to the cytoplasmic membrane where they exit the cytoplasm through the Sec pore (Kline, Kau et al. 2009). The hydrophobic stretch of a sortase substrate inserts laterally into the lipid bilayer, temporarily anchoring the protein to the exterior of the cytoplasmic membrane, where the sortase recognizes the conserved signature motif and transfers its substrate to the peptide portion of the peptidoglycan precursor lipid II (Perry, Ton-That et al. 2002). This transfer involves proteolytic cleavage of the substrate protein near the carboxy-terminus and the
formation of a covalent intermediate with the sortase via a thioester bond (Ton-That, Liu et al. 1999). This intermediate is subject to nucleophilic attack by the free amino group of lipid II, releasing the substrate protein from the sortase and covalently linking it to lipid II (Ton-That and Schneewind 1999). The lipid II-linked sortase substrate is incorporated into the cell wall, anchoring it to the cell surface (Spirig, Weiner et al. 2011).

Additional types of sortase, which recognize substrates carrying distinct signature motifs as part of the tripartite structure, have since been identified in many Gram-positive bacterial species (Schneewind and Missiakas 2012). These minor sortases may not act exclusively as transpeptidases to peptidoglycan, but also cross-link pilin precursors. Sortases appear largely limited to the Gram-positive bacteria, although a dedicated system (one sortase for one substrate) occurs in a number of marine proteobacterial species (Comfort and Clubb 2004, Haft, Selengut et al. 2013). However, Haft et al. identified a number of Gram-negative bacteria that have families of Sec substrates sharing a homology domain structurally similar to the tripartite structure of known sortase substrates (Haft, Paulsen et al. 2006). Each genome encoding putative substrates always encodes at least two to over fifty substrates always encode at least one member of a well-defined protein family which henceforth is referred to as exosortase. Exosortases, which are multiple membrane-spanning proteins, lack any homology to known sortases, but contain conserved cysteine, arginine, and histidine residues, which are also found in the catalytic triad of the sortase family. Based on this observation, it was proposed that the predicted exosortases may perform functions similar to those of sortases, while recognizing a different motif, Pro-
Glu-Pro (PEP-CTERM). Members of the exosortase family, such as EpsH from the methanolan biosynthesis cassette of *Methylobacillus sp* strain 12S, frequently are found within a locus of exopolysaccharide production genes, (hence the designation “exosortase”), despite not being involved in exopolysaccharide biosynthesis per se (Yoshida et al., 2003).

Although *in silico* analyses have confirmed that the patterns of the distribution of exosortases and its putative substrates in Gram-negative species are identical, exosortase processing of substrates containing the PEP-CTERM motif has not been confirmed *in vivo*.

A recent study by Haft *et al.* has revealed that many euryarchaeal genomes encode distant homologs of the bacterial exosortase (Haft, Payne et al. 2012). Consistent with this finding, these archaecal genomes also encode potential substrates having the typical (exo)sortase tripartite structure, which includes the signature motif, the transmembrane domain, and positively charged residues at the carboxy-terminus (Haft, Payne et al. 2012). These homologs of exosortases, which were termed archaeosortases, were classified into distinct subfamilies, and for each subfamily, substrates with a specific signature motif were proposed by genome analysis. Archaeosortase A (ArtA), which is found in many euryarchaea, probably recognizes a Pro-Gly-Phe motif (PGF-CTERM). Surprisingly, one of these predicted ArtA substrates is the S-layer glycoprotein. The S-layer glycoprotein forms a paracrystalline lattice that functions as the euryarchaeal cell wall. This protein was previously believed to be anchored to the cell membrane by the intercalation of a transmembrane segment (Sumper 1987, Sumper, Berg et al. 1990). However, previous
mass spectrometric analyses of S-layer proteins have failed to identify peptides at the carboxy-terminus of the protein, which was taken as evidence that the transmembrane segment may not be present in the mature protein (Haft, Payne et al. 2012). Moreover, at least a portion of the *H. volcanii* S-layer glycoprotein has been found to be lipid-modified (Kikuchi, Sagami et al. 1999, Konrad and Eichler 2002, Kandiba, Guan et al. 2012). This may indicate that archaeosortases, and thus possibly also exosortases, process the carboxy-terminus of its substrates by proteolytic cleavage and transpeptidation to lipids, mirroring the proteolytic cleavage and transpeptidation to peptidoglycan catalyzed by sortase A. To confirm that the archaeosortase plays a role in anchoring proteins, such as the S-layer glycoprotein, to the cell surface, we characterized the archaeosortase in the model euryarchaeon *Haloferax volcanii*. The created ∆artA mutant, although viable, is impaired in growth, motility, and conjugation, and possesses an altered cell morphology. This points to the biological importance of the archaeosortase ArtA. We provide compelling initial evidence that the altered cell morphology can be attributed to incomplete maturation of the S-layer glycoprotein.

2.3 Materials and methods

*Reagents.* All enzymes used for standard molecular biology procedures were purchased from New England BioLabs, except for iProof High-Fidelity DNA polymerase, which was purchased from Bio-Rad. The MF membrane filters (0.025 µm) and Ultracel-3K membrane were purchased from Millipore. DNA and plasmid purification kits were
purchased from Qiagen. NuPAGE gels, buffers, and reagents were purchased from Life Technologies (Invitrogen). Difco agar and Bacto yeast extract were purchased from Becton, Dickinson, and Company. Peptone was purchased from Oxoid. 5-Fluoroorotic acid (5-FOA) was purchased from Toronto Research Biochemicals. All other chemicals and reagents were purchased from either Fisher or Sigma.

**Strains and growth conditions.** The plasmids and strains used in this study are listed in Table 2. *H. volcanii* strains were grown at 45°C in liquid or on solid complex (MGM) (Dyall-Smith 2004) or semi-defined (CA) (Dyall-Smith 2004) medium. Plates contained 1.5% agar unless mentioned otherwise. To ensure equal concentrations of agar in all plates, the agar was completely dissolved prior to autoclaving, and the autoclaved medium was stirred before plates were poured. *H. volcanii* strain H53 (Allers, Ngo et al. 2004) was grown in MGM medium without any supplements. *H. volcanii* strain H98 (Allers, Ngo et al. 2004) was grown in MGM medium supplemented with thymidine (40 µg/ml final concentration). For the selection of the ΔartA (HVO_0915) mutant (see below), 5-FOA was added at a final concentration of 150 µg/ml in CA medium, and uracil was added at one-fifth its normal concentration (i.e., a 10 µg/ml final concentration) during 5-FOA selection. *H. volcanii* strain H53 and the ΔartA mutants transformed with pTA963 or the recombinant derived from it were grown in CA medium supplemented with tryptophan (50 µg/ml), while the transformed *H. volcanii* strain H98 and ΔartA were grown in CA medium supplemented with thymidine (40 µg/ml), hypoxanthine (40 µg/ml), and tryptophan (50
Escherichia coli DH5α and DL739 strains were grown at 37°C in NZCYM medium supplemented with ampicillin (200 µg/ml), when necessary (Blattner, Williams et al. 1977).

Generation of a chromosomal artA deletion in H53 and H98. Chromosomal deletions were generated by using a homologous recombination (pop-in pop-out) method previously described (Allers, Ngo et al. 2004). Plasmid constructs for use in the pop-in pop-out knockout process were generated by using overlap PCR as described previously (Tripepi, Imam et al. 2010) with the following modification: approximately 685 nucleotides flanking the artA gene were PCR amplified and cloned into the haloarchaeal suicide vector pTA131. The final construct, pAF3, contained upstream and downstream artA flanking region inserts (primers used are listed in Table S1 in the supplemental material). The insertion of the correct DNA fragment into the cloning site of the recombinant plasmid was verified by sequencing using the same primers. To confirm the chromosomal replacement event at the proper location on the chromosome, colonies derived from these techniques were screened by PCR using primers listed in Table S1. The artA deletion mutant generated in strains H53 and H98 were designated AF103 and AF110, respectively.

Construction of expression vectors encoding the ArtA protein. The artA gene coding region was amplified by PCR and cloned into the low-copy expression vector pTA963 under the control of an inducible tryptophanase promoter (ptna) to induce the expression of this gene. The recombinant pTA963 carrying the artA gene was designated pAF9. pAF9 was isolated from E. coli DH5α and transformed into E. coli DL739 (Table 2). Using the
standard polyethylene glycol (PEG) protocol (Dyall-Smith, 2004), nonmethylated plasmid DNA isolated from *E. coli* DL739 was used to transform *H. volcanii*. To complement Δ*a*rta strains, plasmid pAF9 was transformed into AF103 (derived from H53) to result in AF109, and into AF110 (derived from H98) to result in JZ102. The parent strains H53 and H98 were also transformed with pAF9 to generate strains JZ101 and JZ103, respectively, which thus had two copies of the *artA* gene under control of distinct promoters.

**Motility assay.** The motility assay was performed on motility plates (0.3% agar in CA medium). For the defined CA agar media, tryptophan, thymidine, and hypoxanthine were added at final concentrations of 50 µg/ml, 40 µg/ml, and 40 µg/ml respectively. A toothpick was used to stab inoculate the agar, followed by incubation at 45°C. Halo sizes around the stab-inoculation site were measured after 3 to 5 days of incubation.

**Isolation of Δ*artA* motility suppressor strain.** Ten colonies of the Δ*a*rta strain were stab-inoculated on the motility plate and incubated for 10 days at 45°C. The colonies that exhibited swimming motility after the incubation period were re-stabbed on the new motility plate to ensure the consistency of the motility suppressor. The halo size of the Δ*a*rta motility suppressor mutants was measured after 4-6 days and compared to the wild type. This Δ*a*rta motility suppressor strain will hereafter be referred to as Δ*artA*.

**Conjugation assay.** The conjugation rate was assayed using a modified version of a protocol previously described (Mevarech and Werczberger 1985). In brief, equal volumes (3 to 5 ml) from cultures of two different auxotrophic *H. volcanii* strains (H53, Δ*trpA*, requiring tryptophan and H98, Δ*hdrB*, requiring thymidine) at OD$_{600}$ of 0.3 to 0.5 were
mixed. The mixture was filtered through Millipore Swinnex 25 mm filter units equipped with corresponding filters (pore size 0.45 µm, type HA 25-mm filters) using a 5-ml syringe. The filter unit was then disassembled and filter discs were placed cell side up on solid MGM medium supplemented with thymidine, and then incubated at 45°C overnight. After incubation, the filters were removed from the medium, placed in 2-ml Eppendorf tubes containing 1 ml of 18% salt water (Dyall-Smith 2004), and shaken on a rotator for 1 h. Fifty µL of an undiluted sample and samples that were diluted 1:100 were plated on CA agar medium supplemented with uracil, but lacking tryptophan and thymidine. This prevented the growth of the auxotrophic strains H53 or H98, respectively, and selected for conjugants that contained both the trpA and hdrB genes. Twenty µL of samples that were diluted 1:100,000 were also plated on MGM agar medium supplemented with thymidine for viable cell counts.

**Growth curves.** Growth curves were generated using a Biotek Power Wave X2 microplate spectrophotometer. *H. volcanii* parent strain H53, the ΔartA mutant, and the ΔartA complementation strain were first incubated in 5 ml liquid cultures in CA medium supplemented with tryptophan (final concentration 50 µg/ml), with continuous shaking at 45°C, until suitable OD₆₀₀ values (0.2 to 0.5) were reached. Approximately 6 µl of each culture (adjusted slightly for OD₆₀₀ differences) were then transferred into 194 µl of fresh CA medium with normal (18%), low (14%), or high (23%) salt concentration and supplemented with tryptophan (final concentration 50 µg/ml), and grown to the stationary phase, with OD₆₀₀ recordings taken every 30 min.
**Purification of the S-layer glycoprotein.** The isolation of the S-layer glycoprotein was performed by taking advantage of the surface structure purification technique as described previously (Fedorov, Pyatibratov et al. 1994), with modifications described below. H53 wild-type cells or ΔartA cells were inoculated into 5 ml MGM liquid medium. Two liters of MGM medium were inoculated with two 5 ml cultures each and the cultures were harvested at an OD$_{600}$ of approximately 0.3 by centrifugation at 8,700 rpm (JA-10 rotor; Beckman) for 30 min. The supernatant was centrifuged again (8,700 rpm for 30 min) and incubated at room temperature with 4% (wt/vol) PEG 6000 for 1 h. The PEG-precipitated proteins were then centrifuged at 16,000 rpm (JLA-16.250 rotor; Beckman) for 50 min at 4°C, and the S-layer purified by cesium chloride (CsCl) density gradient centrifugation (overnight centrifugation at 50,000 rpm) (VTI-65.1 rotor; Beckman). CsCl was dissolved in a 3 M NaCl saline solution to a final density of 1.37 g/cm. The CsCl gradient fractions were dialyzed against water and concentrated using an Ultracele-3K membrane (Millipore). The electrophoresis of the proteins in the fractions was performed with 7% NuPAGE Tris-Acetate (TA) gels (Life Technologies) under denaturing conditions using Tris-Acetate buffer at pH 8.1 to separate the S-layer glycoprotein from other proteins in the fraction. The proteins on the 7% NuPAGE Tris-Acetate gels were stained with Coomassie Brilliant Blue and the protein band that corresponded to the expected S-layer glycoprotein size was cut from the gel and submitted for mass spectrometry analysis.

**Mass spectrometry.** The S-layer glycoprotein band from the Coomassie-stained bands from the 7% NuPAGE Tris-Acetate gels was tryptically digested for LC-MS/MS
analysis according to published procedures, except that samples were not alkylated (Shevchenko, Tomas et al. 2006). The previously described capillary reversed-phase liquid chromatography (RPLC) system for peptide separations (Livesay, Tang et al. 2008) was slightly modified with regards to the pumping system. Briefly, the High-Performance Liquid Chromatography (HPLC) system consisted of a custom configuration of Agilent 1200 nanoLC pumps (Agilent Inc., Santa Clara, CA), 2-position Valco valves (Valco Instruments Co., Houston, TX), and a PAL autosampler (Leap Technologies, Carrboro, NC), allowing for fully automated sample analysis across four separate HPLC columns (3-µm Jupiter C18 stationary phase, Phenomenex, Torrence, CA). The mobile phases consisted of 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B). The HPLC system was equilibrated with 100% mobile phase A for 2 min, followed by a 100 minute RPLC gradient (0-12% buffer B over 10 min, 12-35% buffer B over 55 min, 35%-95% buffer B over 10 min, 95% buffer B for 5 min and 0% buffer A for 20 min). The flow rate during the run was set at 400 nL/min. ESI using an etched fused-silica tip (Livesay, Tang et al. 2008) was employed to interface the RPLC separation to a LTQ Orbitrap Velos mass spectrometer (Thermo Scientific, San Jose, CA). Precursor ion mass spectra (AGC $1 \times 10^6$) were collected for 400–2000 m/z range at a resolution of 60 K, followed by data dependent HCD MS/MS (normalized collision energy 35%, AGC $3 \times 10^4$) of the eight most abundant ions at a resolution of 15 K. A dynamic exclusion time of 60 sec was used to discriminate against previously analyzed ions.
**Mass spectrometry data analysis.** HCD MS/MS data were processed using SEQUEST (Eng 1994) and a *H. volcanii* proteome database downloaded from Genbank ([http://www.ncbi.nlm.nih.gov/protein](http://www.ncbi.nlm.nih.gov/protein)) (Hartman, Norais et al. 2010). The database search against the decoy database, using both forward and reverse sequences, was performed with the following settings: no enzyme rule, methionine oxidation as a variable modification, and precursor mass limit at 5000 Da. A peptide mass tolerance of 3 Da and a fragment mass tolerance of 0.5 Da were allowed. First hits from the SEQUEST analysis were rescored using a MS-Generating function (MSGF) (Elias and Gygi 2007) for spectral probabilities. Peptide identifications were filtered based on a 1% false discovery rate against the decoy database (Elias and Gygi 2007, Kim, Gupta et al. 2008).

Peptide quantitation was performed using the identified peptide signal intensity in precursor spectra eliminating redundant peptides with weak signal intensity. ICR2LS software ([http://omics.pnl.gov](http://omics.pnl.gov)) was used to determine the most abundant isotope peaks of the expected peptides. Matched peaks were filtered using 10 ppm mass accuracy and 0.2 tolerance LC window in normalized elution time. Maximum signal intensity was extracted from each dataset (no normalization). Results of this analysis are presented in Table S2.

**Light microscopy and cell measurements.** The *H. volcanii* H53 wild type, ΔartA, ΔartA*, ΔartA + artA, and wt + artA strains were grown in 5 ml liquid cultures to an OD₆₀₀ of ~ 0.1. One ml of cells was concentrated by centrifuging at 6800 rpm for 5 minutes and pellets were resuspended in 20 µL of MGM liquid media. The concentrated cells were transferred on to a microscope slide and observed under the light microscope. Light
microscopy was performed using an Eclipse TE2000-U inverted microscope (Nikon) equipped with a Plan Apo 100×1.0 NA objective and Cascade 512B CCD camera (Photometrics) driven by Metamorph imaging software (Molecular Devices). ImageJ software was used to measure the lengths and widths of each cell, and to calculate the average cell perimeter. At least 20 cells were used to measure each parameter. Cells were counted using a Petroff-Hauser Counting Chamber.

*Electron microscopy.* The *H. volcanii* H53 wild-type and ΔartA cells were cultured in 5 ml liquid MGM until they reached OD$_{600}$ of 0.3. Cells were concentrated by centrifugation and drawn into 200 µm diameter Spectrapor® microdialysis tubing (Spectrum Laboratories). Segments of 2.5 mm tubing containing cells in culture media were frozen at high pressure (4350 psi) using an Abra model HPM010. Cellular water was replaced with acetone, 2% OsO$_4$, and 0.1% UA (uranyl acetate) at -90°C for 72 h. Cells were washed three times in acetone at room temperature and incubated overnight in 1:1 acetone:EPON followed by two exchanges of EPON, before polymerization at 60°C for 48 h. Thin sections were cut at 60 nm thickness. Sections were counterstained in 2% UA and Reynolds lead citrate solution before observation at 80 KeV on an FEI Tecnai12 electron microscope. Images were recorded digitally using a Gatan Ultrascan 1000 4-megapixel camera at the indicated magnifications.

**2.4 Result**

The *H. volcanii* ΔartA strain has a growth defect. Haft *et al.* reported the identification of eight predicted ArtA substrates, including the S-layer glycoprotein ((Haft,
51

Payne et al. 2012) and Table 1). In order to examine the role of H. volcanii ArtA in vivo, we used homologous recombination to construct deletion strains that lack artA. We verified the deletion of artA by PCR, using primers designed to amplify the coding region as well as the DNA regions upstream and downstream of artA. A PCR product for the coding region was obtained for the parent strains (H53 and H98) but not for the ΔartA strain. The PCR product obtained with DNA from the deletion strain was significantly smaller than the product with DNA from the parent strain (Fig. S1). Most of the experiments were performed with strain H53 and derivatives thereof. H98 and its ΔartA derivative was only used for conjugation assays.

The H53 ΔartA strain was viable, but displayed a growth defect in media containing a low (14%), standard (18%), or high (23%) salt concentration (Fig. 1). The most notable growth defect was observed in 14% salt water concentration, while the artA deletion strain exhibited the least severe defect at 23%. These results are consistent with the cell wall of ΔartA being defective, because H. volcanii cells, which maintain a high internal concentration of KCl to counter the high external NaCl concentration, are exposed to the highest turgor pressure in low salt environments. The artA deletion was complemented by cloning the artA gene on a low-copy plasmid under the control of tryptophan-inducible promoter. When complementation strains were grown in medium containing tryptophan, growth was very similar to that of H53 (Fig. 1).

The S-layer glycoprotein is carboxy-terminally processed and ArtA is required for this posttranslational modification. To confirm that it is indeed processed at its carboxy-
terminus, the S-layer glycoprotein was isolated and purified from the supernatant fractions of cultures from ΔartA and its parent H53. We observed that significantly more protein was released into the supernatant in the ΔartA strain than in the H53 strain (data not shown), which may indicate reduced stability of the S-layer. The glycoprotein was in-gel digested with trypsin and peptides were characterized using LC-MS/MS. Peptide identification of the samples revealed 119 distinct peptides at a false discovery rate below 1% (Fig. 2A). The abundance of peptides was determined using the AMT (accurate mass and time tag) approach (Pasa-Tolic, Masselon et al. 2004). For nearly all of the peptides (117 of 119), the signal intensity was very similar for samples from the ΔartA strain and its parent. However, the identification of two overlapping peptides near the extreme C-terminus (VALVGAALLALR and VGAALLALR) was restricted to ΔartA samples. These peptides were found in all ΔartA samples but completely absent from the H53 samples (Fig. 2B). To confirm this finding, the peptide/spectrum matches for both these peptides were further verified manually. The precursor ion identified in survey scans had a perfect isotopic correlation with the theoretical and < 5ppm mass error (Fig. 2C). The fragmentation spectra for these precursors showed strong b/y ladder representation with low ppm error (Fig. 2D). Therefore, the identification of the two carboxy-terminal peptides in ΔartA samples was highly reliable. No tandem mass spectra from these two peptides were identified in H53 samples. Moreover, the AMT quantification strategy did not find any parent ion features that matched these peptides. As extra validation, we generated an extracted ion chromatogram (XIC) for the m/z values of the peptides discovered in the ΔartA samples.
These XICs showed zero intensity for the expected LC retention window, confirming that the carboxy-terminal peptides VALVGAALLALR and VGAALLALR were not present in the H53 samples. Thus, the wild-type S-layer glycoprotein was indeed proteolytically processed and the extreme carboxy-terminal region was not present in the mature protein. Moreover, the overlapping carboxy-terminal peptides identified in the ΔartA samples were the first to be identified in the carboxy-terminal S-layer glycoprotein region by any mass spectrometric analysis, strongly suggesting that carboxy-terminal processing of this target occurs in the wild-type and is prevented in the ΔartA strain.

We noted that the abundance of the two carboxy-terminal peptides VALVGAALLALR and VGAALLALR (5.80 and 5.61 respectively) agreed with the bulk of the peptide intensities for the ΔartA samples (mean 5.82, standard deviation 0.70). This suggested an even stoichiometry for all the identified peptides, meaning that there was likely only one form of the S-layer glycoprotein identified in the ΔartA samples: an isoform that was incompletely processed.

The only other peptide exclusively identified in ΔartA samples was DSAIGDGSLPSGPSNGATLNDLTGYLDTLDQNNN. While this low-abundance peptide was never selected for fragmentation in H53 samples, it was detected in these samples with a maximum signal intensity ratio of 0.85, consistent with the observed mean 0.91 and standard deviation 0.07 for the other 116 peptides (Table S2).

**H. volcanii ΔartA cells are significantly shorter than those of the parent strain.** Considering that the S-layer is thought to be critical for the determination and maintenance
of the cell shape, we examined samples taken from H53 and ΔartA strain cultures for changes in cell shape and noticeable defects in ΔartA cell morphology using phase contrast microscopy (Fig. 3). Although H. volcanii H53 cells do exhibit a degree of morphological variation, the majority of early log phase parent and ΔartA cells were rod-shaped. However, while the cells of the parent strain had an elongated appearance (mean length: 4.7 µm; SD: 0.9 µm), cells from the ΔartA cultures were significantly shorter (mean length: 2.6 µm; SD: 0.6 µm). wt + artA cells were slightly more elongated (mean length: 5.1 µm; SD: 1.1 µm) than the parent strain, and ΔartA +artA cells possessed the longest rod-like shape (mean length: 5.8 µm; SD: 0.5 µm). The cell width of all the strains measured was comparable. Viability counts revealed no significant differences among the strains.

To directly analyze the structure of the S-layer, the cell walls of wild-type and ΔartA cells were examined by transmission electron microscopy (Fig. 4). Cells were prepared by high pressure freezing followed by low temperature acetone, 2% OsO₄, and 0.1% uranyl acetate substitution, before being embedded into EPON resin. The most obvious difference between wild-type and ΔartA cells was their cell morphologies. The cytoplasm of the ΔartA strain also stained darker. Although the assembly of an S-layer surface structure was found in H53 and the ΔartA strains, distinct differences were evident at higher magnification. In general, the ΔartA strain S-layer had a less compacted appearance and, with an S-layer width average that ranged from 11 nm to 15 nm, was wider than the H53 S-layer, which had a width that averaged from 6 nm to 10 nm. The measurements of the S-layer thickness were performed for five wild-type and mutant cells,
and they revealed very consistent results. These observations are consistent with the S-layer glycoprotein processing defect identified in the ΔartA strain by MS analysis.

_Deletion of artA impairs motility and conjugation._ Next to the S-layer glycoprotein, seven additional predicted _H. volcanii_ Sec substrates contain a similar tripartite C-terminal region (Table 1) suggesting the involvement of ArtA in cellular processes other than cell wall biosynthesis. Interestingly, when stab-inoculating H53, ΔartA, as well as the ΔartA complementation strains into (0.3%) motility agar, H53 cells produced a growth halo between 3 to 5 days after stab-inoculation, which indicated swimming motility. No growth halo was observed in the ΔartA strain during the same incubation period. However, the ΔartA complementation strains expressing the tryptophan-inducible artA in-trans not only restored the ΔartA swimming motility, it exhibited a greater motility than H53. Hypermotility of the complemented ΔartA strain was evident from the larger growth halo 5 days after inoculation (Fig. 5). The empty expression plasmid pTA963 itself had no effect.

Notably, when colonies of the ΔartA strains were stab-inoculated on the motility plate and incubated for an extended period (>10 days), all of the colonies ultimately suppressed the motility defect and exhibited swimming motility. Upon restabbing, the motile ΔartA cells started moving after 4-6 days (Fig. 5). Using PCR analysis, we confirmed that the motility was not due to reversion of the ΔartA genotype (Supp. Fig. 1). Also, while the motility defect was partially suppressed in these strains, the growth defect remained (Fig. 1). This, as well as the fact that the cells had a size similar to that of the
non-motile ΔartA strain (Mean length: 2.7 µm; SD: 0.7 µm) (Fig. 3), suggested that the suppression was independent of the S-layer.

*H. volcanii* conjugation occurs independently of the presence of either type IV pili or flagella (Tripepi, Imam et al. 2010). Several predicted ArtA substrates contain cadherin and fibronectin-like domains (Table 1), which are known to facilitate interactions with other cells (Brasch, Harrison et al. 2012, Chagnot, Listrat et al. 2012). *H. volcanii* conjugates by forming cytoplasmic bridges (Rosenshine, Tchelet et al. 1989). Therefore, we tested whether the *artA* deletion had an influence on conjugation by modifying a published conjugation assay (Mevarech and Werczberger 1985). In brief, we co-cultured *H. volcanii* tryptophan (H53, ΔartA) and thymidine (H98, ΔartA) auxotrophs, each harbouring an *artA* deletion, in a selective medium that lacked both tryptophan and thymidine. In the absence of these compounds, independent growth of either auxotroph was not possible. We also plated diluted cultures on Modified Growth Medium (MGM) supplemented with thymidine in order to obtain viable counts. It should be noted that both H98 wild-type and mutant strains exhibited comparable growth to the wild-type and mutant H53 background strains (data not shown).

The conjugation frequency for the parent strains was approximately \(4 \times 10^{-3}\), while the average conjugation frequency for the ΔartA derivatives was only \(3.7 \times 10^{-4}\). Thus, we observed that conjugation rates between ΔartA strains were an order of magnitude lower than those of the parent strains when the strains were co-cultured on a filter placed on a complex medium (MGM) plate and subsequently co-cultured in semi-defined Casamino
Acids (CA) medium lacking tryptophan and thymidine. In this experiment, artA was deleted in both strains, because surface adhesin-mediated contact might have been possible even if only one strain expressed the adhesin.

In summary, we showed that various biological phenomena like growth, cell shape, motility, and conjugation were critically dependent on an intact artA gene, even though this gene proved not to be essential.

2.5 Discussion

Following up on in silico data, we obtained the first in vivo evidence of the function of archaeosortase. Our data strongly support the functional similarity of archaeosortases with bacterial sortases, namely, proteolytic removal of a carboxy-terminal transmembrane segment as a late step in protein processing, and not by hydrolysis, but by a transpeptidation that anchors proteins covalently to the cell surface. The signature motif for sortase recognition, LPXTG, sits four or five residues away from the transmembrane segment, while the analogous PGF motif in archaeosortase targets lies flush to the transmembrane segment. This contrast follows the model that archaeosortase attaches target proteins covalently with a lipid found in the membrane, while sortase attaches proteins to the growing cell wall.

Our in vivo data also support the in silico prediction that archaeosortases share similar functionality with the distantly related bacterial homolog, exosortase, for which in vivo data have – to our knowledge – not yet been obtained. Thus, prokaryotes
independently evolved at least twice a similar mechanism to recognize, process, and anchor to the cell surface proteins that are transported across the cytoplasmic membrane and contain a conserved carboxy-terminal tripartite structure.

We showed that while it was possible to make an artA deletion strain, H. volcanii lacking this gene exhibited several phenotypes: a growth defect, significantly smaller cells that are more fragile, and impaired motility and conjugation. The effects on cell shape and fragility are consistent with a destabilization of the S-layer for H. volcanii lacking ArtA. Our data strongly support the hypothesis that this is due to the involvement of ArtA in carboxy-terminal processing of the S-layer glycoprotein. Concerning the motility and conjugation phenotypes, these may be indirect effects of an altered S-layer or may be attributed to other predicted ArtA substrates.

Direct evidence that ArtA is involved in the processing of the S-layer carboxy-terminus was obtained by LC-MS/MS analysis of this glycoprotein purified from wild-type and ΔartA cells. We readily identified two overlapping peptides from the carboxy-terminal TM domain of S-layer glycoprotein purified from the ΔartA strain. Conversely, we failed to identify this same peptide from the wild-type, strongly suggesting that the wild-type carboxy-terminus is indeed processed. The failure to identify the peptide in the wild-type strain cannot be attributed to technical problems, as our peptide quantification data show that the other 117 identified peptides from the S-layer glycoprotein are found in similar quantities whether purified from the wild-type or from the ΔartA strain. In addition, the two overlapping peptides from the carboxy-terminal TM domain were found in a quantity
similar to that of the other 117 peptides from that same strain. This data clearly shows that the *H. volcanii* S-layer glycoprotein is processed and that ArtA is involved in this processing mechanism.

Similar to the sortase substrates, it can be assumed that the S-layer glycoprotein is subject to two types of posttranslational modification at its carboxy-terminus: lipid attachment and proteolytic cleavage. While lipid attachment in the carboxy-terminal region has been previously demonstrated (Kikuchi, Sagami et al. 1999, Konrad and Eichler 2002, Kandiba, Guan et al. 2012), carboxy-terminal proteolytic cleavage had previously not been confirmed. In contrast, it had been proposed that the S-layer glycoprotein is anchored by its carboxy-terminal transmembrane domain (Sumper 1987, Sumper, Berg et al. 1990). Interestingly, a recent paper suggested that two S-layer glycoprotein forms co-exist in wild-type *H. volcanii* (Kandiba, Guan et al. 2012). In addition to a membrane-associated fraction that requires detergent for removal from the membrane, Kandiba *et al.* identified a more weakly membrane-associated fraction that can be released by EDTA treatment. Moreover, mass spectrometry analysis of the base-treated, weakly interacting fraction identified a lipid, which was covalently linked in an alkaline-sensitive manner. As lipid attachment should strengthen and not weaken the membrane interaction, this data supports our experimental results, which show that the S-layer glycoprotein undergoes carboxy-terminal proteolytic processing, a coupling reminiscent of the coupling performed by sortases.

At this stage, we do not have a positive identification of the processed carboxy-terminal peptide from the S-layer glycoprotein in the wild-type strain, despite more
extensive attempts for its identification. We searched for peptides that carry a modification of uncertain mass using the USTags approach (data not shown) but did not find any promising leads in yet unidentified mass peaks, and failed to identify the corresponding peptide. Any proteolytic peptide from this region is supposedly a glyco-lipo-peptide, and it is known that identifying such peptides are associated with extreme experimental challenges. Consistent with this assumption, not a single peptide from this Thr-rich region could be identified, neither in this current study nor in any other mass spec analysis of an archaeal S-layer glycoprotein tested (Haft, Payne et al. 2012). Although the underlying technical problems have yet prohibited the positive identification of the carboxy-terminally processed peptide, we are certain that the data already presented unambiguously show that ArtA is involved in carboxy-terminal posttranslational modification of the S-layer glycoprotein and strongly support that proteolytic cleavage of the carboxy-terminal transmembrane domain occurs.

While we primarily concentrated on the effect of the ΔartA mutant on the S-layer glycoprotein, the additional phenotypic effects described in the current manuscript point to the biological importance of the archaeosortase. The reduced growth, impaired motility, and reduced conjugation efficiency of the ΔartA mutant may be an indirect effect due to a weakened S-layer or a direct effect due to the absence of a distinct ArtA substrate.

The observed growth defect may be due to poor cell wall stability. The most severe growth defect is observed when cells are grown in media containing low (14%) rather than moderate (18%) or high (23%) salt concentration (Fig. 1), probably due to an increased
cell turgor under these growth conditions. We observed some variation in growth in these experiments. Occasionally the $\Delta artA + artA$ strain grew to higher densities than the parent strain, and sometimes the parent strain $wt + artA$ strain grew to higher densities. There seems to be a narrow range of $artA$ expression that is optimal for growth under the conditions tested.

The effect on cell shape can be directly attributed to the S-layer glycoprotein. The growth defect may also be due to reduced cell stability resulting from an impaired S-layer. However, it should also be noted that under certain growth conditions, an S-layer independent ArtA substrate may add to the growth defect. For example, one of the substrates is a cobalamin uptake protein that may be critical when cells are grown on defined media.

It is currently enigmatic if the effect on motility is a secondary effect also mediated by an impaired S-layer or if another ArtA substrate is responsible for this effect. A motor must be anchored to the cell membrane (Ghosh et al., 2011) and interaction with the S-layer may participate in anchoring. However, as the molecular details of the archaeal motor are still largely unresolved, a distinction between direct and indirect effects is currently not possible. However, it is unlikely that the motility defect is solely due to S-layer instability, as suppressors have been obtained that are motile, albeit not at wild-type levels, yet have a severe growth defect. In this respect, it will also be interesting to analyze why a probable $artA$ overexpressor (the complementation strain with $artA$ transcription from Trp-inducible pAF9 plasmid) is hypermotile.
Finally, it is currently not clear why the $\Delta artA$ mutant demonstrates impaired conjugation. In fact, a destabilized S-layer would be expected to facilitate rather than prevent DNA uptake. Thus, other predicted ArtA substrates, some of which contain domains reminiscent of eukaryotic cell attachment molecules, may be responsible for the observed effect.

In summary, we have shown experimentally that ArtA is important for several biological phenomena. The comparison of the S-layer glycoprotein from wild-type and the $\Delta artA$ mutant strains strongly supports the hypothesis that the S-layer glycoprotein is not attached via a carboxy-terminal transmembrane region, but rather that the carboxy-terminal region is cleaved off. The S-layer glycoprotein is then anchored to the lipid bilayer. It is highly likely that ArtA is directly involved in this posttranslational processing step. The availability of the mutant will allow a deeper experimental analysis of the archaeosortase activity. It needs to be analyzed if all archaeosortase targets carry a lipid covalently attached to the carboxy-terminal region. Moreover, the multiple effects of the deletion of $artA$ indicates that analysis of this enzyme and its substrates will yield a better understanding of processes ranging from DNA transfer to motility and nutrient uptake.
### 2.6 Tables

**Table 2.1: Predicted *H. volcanii* ArtA substrates**

<table>
<thead>
<tr>
<th>Protein</th>
<th>Thr-rich Region</th>
<th>C-term Domain</th>
<th>Annotation, Conserved Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hvo_0405</td>
<td>TEPLVNTPTPTPSGNETAASASETSTSTSTGD</td>
<td>PGFGLGGAAGGAVLVRVTGSPG</td>
<td>NifU- and LVIVD domains</td>
</tr>
<tr>
<td>Hvo_1095</td>
<td>ENSTLTPDRDAAATATTTAATADGSGIV</td>
<td>PGFGAGGVAALLVAVFVAARGL</td>
<td>hypothetical</td>
</tr>
<tr>
<td>Hvo_1110</td>
<td>TTTACDATATATATETTEATTTESETTDT</td>
<td>PGFVVALAAGAFLLLRRR</td>
<td>Cobalamin-binding protein</td>
</tr>
<tr>
<td>Hvo_2006</td>
<td>DMTDEPTETETDEPETETMDEMETTESTDA</td>
<td>PGFGLVVALVAALVVAR</td>
<td>hypothetical</td>
</tr>
<tr>
<td>Hvo_2072</td>
<td>TTTTGGPTETTTTTEPETETTTETTTT</td>
<td>PGFGIAVALVGAVALLAR</td>
<td>S-layer glycoprotein</td>
</tr>
<tr>
<td>Hvo_2160</td>
<td>TTTTTPDGTTTTGTAGFDDETSTSAASGGTTDDPV</td>
<td>PGFMAALAVALLAVVAR</td>
<td>CARDB' and Big3' domains</td>
</tr>
<tr>
<td>Hvo_2533</td>
<td>REDVELRVEDFENGDGAPREDAYETETTSSA</td>
<td>PGFGLVAVVAGLVVVAR</td>
<td>COG4885</td>
</tr>
<tr>
<td>Hvo_A0263</td>
<td>DTTITDESAATDSITDITDDEPATTASDS</td>
<td>PGFGVVLTVVAMLVGTTFVRRARCR</td>
<td>GluG-domain, cadherin-repeat</td>
</tr>
<tr>
<td>Hvo_B0206</td>
<td>AGDGDGSGDSDGGTPETETTEDNTPATTTTDTPV</td>
<td>PGFVSVALAALVIGVLLARRR</td>
<td>Collagen triple helix domain</td>
</tr>
</tbody>
</table>

1. **PGF**-motif in bold; hydrophobic stretches are underlined, positive charges are in italics; 2. the C-terminal 160 amino acids, including NifU-domain, are not included; 3. Cell adhesion related domain found in bacteria; 4. Bacterial Ig-like domain
2.7 Figures

Figure 2.1: Depletion of ArtA results in a growth defect. *H. volcanii* H53 (wt), ΔartA, and the motility suppressor (ΔartA*), all transformed with an empty pTA963 plasmid, as well as H53 and ΔartA expressing artA expressed from pTA963, (wt + artA and ΔartA + artA, respectively) were inoculated in liquid semi-defined media CA supplemented with tryptophan (final concentration 50 µg/ml) at low (14%), normal (18%) and high (23%) salt concentration.
Figure 2.2 High-resolution mass spectrometry analysis of S-layer glycoprotein. (A) Protein sequence coverage from the identified peptides excised from in-gel tryptic digestion. Red typeset is used to indicate regions covered by identified peptides from H53 strain samples; underline regions covered from ΔartA samples. Green underline is used for the overlapping TM peptides that are found exclusively in ΔartA samples. The PGF motif is highlighted yellow. The signal peptide is marked grey. Incomplete sequence coverage is possibly due to an uncharacteristically sparse distribution of cleavage residues Lys and Arg in the protein sequence and to post-translational modifications. (B) Ratio of signal intensity from H53 and ΔartA samples for 119 peptides covering the S-layer glycoprotein; two carboxy-terminal peptides (green spots) were observed and identified exclusively in the ΔartA sample. The blue spot indicates a low-intensity peptide of 34 amino acids, which was found in both samples but was subjected to MS/MS analysis only in 1 of 8 analyzed samples, the one being from ΔartA. For the carboxy-terminal 19 of this 34 amino acid peptide, no other overlapping identification was obtained and thus this region is underlined.
and in black typeset in panel A. (C) The theoretical and observed isotope correlation for peptide VALVGAALLALR. Theoretical values are shown in red, and the observed mass spectrum in this region is shown in black. Both the m/z values and the intensity of the identified isotopic envelope match the peptide sequence. (D) The fragments from tandem mass spectra matching the proposed peptide. Dots indicate that a fragment peak was found (e.g. b3) and the color of the dot indicates the mass accuracy.

Figure 2.3: Phase contrast microscopy of *H. volcanii* H53 and ΔartA in liquid culture. The majority of *H. volcanii* parent and mutant strains are rod-shaped. However, cells of the parent strain (wt) are nearly twice as long as cells of ΔartA cultures. Morphology of a ΔartA motility suppressor (ΔartA*) is comparable to that of the non-motile ΔartA. Both the wt + artA and ΔartA + artA exhibited more elongated rod-like shapes than did the wild-type, with the ΔartA + artA displaying the largest average length of all strains observed. Size bar: 10 µm.
**Figure 2.4: Transmission electron microscopy:** *H. volcanii* H53 (wt) or Δ*artA* cells were preserved using a high pressure freezing technique and sliced into thin sections. (A) H53 preparations reveal mostly rod-shaped wild-type cells, while Δ*artA* cells are significantly thinner and have irregular shapes. Size bar: 50 nm. (B) While H53 and Δ*artA* strains assemble an S-layer (arrows), the Δ*artA* strain S-layer has a less compacted appearance and is wider (11 to 15 nm) than the H53 S-layer (6 to 10 nm). Size bars 50 nm.
Figure 2.5: The ΔartA strain exhibits a motility defect. *H. volcanii* H53 (wt), ΔartA, and the motility suppressor (ΔartA*), all transformed with an empty pTA963 plasmid, as well as H53 and ΔartA expressing artA from pTA963 (wt + artA and ΔartA + artA, respectively), were stab inoculated on motility plate (0.3% agar). While H53 produced a growth halo after 3 to 5 days, no growth halo was observed for the ΔartA strain during the same incubation period. Complementing the deletion strain with artA from a plasmid not only restored motility but resulted in hypermotility compared to the H53 strain.
Figure 2.S1. Confirmation of the artA gene deletion. (A) The genomic region containing the artA gene in H. volcanii. Primers used to amplify the artA coding region (inner primer, IP), as well as the flanking region (ca 700 bp upstream and downstream of artA, (outer primer, OP)) are indicated. PCR was performed using genomic DNA templates isolated from the H53 wild-type, ΔartA, or a ΔartA-motility suppressor (ΔartA*) strains. (B) PCR products obtained with the inner primers IP. No amplicons were obtained from ΔartA or ΔartA*. (C) PCR products obtained with the outer primers OP. The amplicon sizes are approximately 900 nucleotides longer for the parent strain H53 than for the strains ΔartA and ΔartA*. 
CHAPTER 3

Permuting the PGF-CTERM signature motif blocks both archaeosortase-dependent C-terminal cleavage and prenyl lipid attachment for the *Haloferax volcanii* S-layer glycoprotein

This chapter has been published in:


3.1 Abstract

For years, the S-layer glycoprotein (SLG), the sole component of many archaeal cell walls, was thought to be anchored to the cell surface by a C-terminal transmembrane segment. Recently, however, we demonstrated that the *Haloferax volcanii* SLG C-terminus is removed by an archaeosortase (ArtA), a novel peptidase. SLG, which was previously shown to be lipid modified, contains a C-terminal tripartite structure, including a highly conserved proline-glycine-phenylalanine (PGF) motif. Here, we demonstrate that ArtA does not process an SLG variant where the PGF motif is replaced with a PFG motif (*slg*<sub>G796F, F797G</sub>). Furthermore, using radiolabeling, we show that SLG lipid-modification requires the PGF motif and is ArtA-dependent, lending confirmation to the use of a novel C-terminal lipid-mediated protein anchoring mechanism by prokaryotes. Similar to the *ΔartA* strain, the growth, cellular morphology, and cell wall of the *slg*<sub>G796F, F797G</sub> strain in which modifications of additional *H. volcanii* ArtA substrates should not be altered, are
adversely affected, demonstrating the importance of these post-translational SLG modifications. Our data suggest that ArtA is either directly or indirectly involved in a novel proteolysis-coupled, covalent lipid-mediated anchoring mechanism. Given that archaeosortase homologs are encoded by a broad range of prokaryotes, it is likely that this anchoring mechanism is widely conserved.

3.2 Importance

Prokaryotic proteins bound to cell surfaces through intercalation, covalent attachment or protein-protein interactions play critical roles in essential cellular processes. Unfortunately, the molecular mechanisms that anchor proteins to archaean cell surfaces remain poorly characterized. Here, using the archaeon *H. volcanii* as a model system, we report the first *in vivo* studies of a novel protein anchoring pathway involving lipid-modification of a peptidase processed C-terminus. Our findings not only yield important insights into poorly understood aspects of archaean biology, but also have important implications for key bacterial species including those of the human microbiome. Additionally, insights may facilitate industrial applications, given that photosynthetic cyanobacteria encode uncharacterized homologs of this evolutionarily conserved enzyme, or may spur development of unique drug delivery systems.
3.3 Introduction

A variety of protein complexes decorate cell surfaces where they support cell stability and facilitate interactions between the cell and the extracellular environment (Albers and Meyer 2011, Thanassi, Bliska et al. 2012, Foster, Geoghegan et al. 2014). Determining the mechanisms that allow these proteins to remain associated with the cell envelope is critical for developing a comprehensive understanding of the biosynthesis of cell surface structures. Protein anchoring to the cell membrane was likely first accomplished by means of transmembrane (TM) segments (Renthal 2010). However, as cell membranes became densely packed with protein complexes, protein anchoring mechanisms that allowed more economic use of membrane surface area emerged (Pohlschröder, Hartmann et al. 2005). One such mechanism anchors proteins through covalent attachment of their N-terminal cysteine to the lipid bilayer (Hutchings, Palmer et al. 2009, Zückert 2014, Buddelmeijer 2015). In many gram-positive bacteria, a subset of proteins are processed and anchored to the cell wall via the carboxy-terminus (C-terminus) by an enzyme known as sortase, a transpeptidase that recognizes a conserved tripartite structure at the C-terminus consisting of a signature motif, a TM alpha helical domain, and a cluster of basic residues (Schneewind and Missiakas 2014). Following proteolytic cleavage of the substrate near its C-terminus and formation of a covalent intermediate of the substrate with the sortase via a thioester bond, the sortase then transfers the substrate to a cell wall peptidoglycan precursor, resulting in covalent attachment of the substrate to the cell wall (Ton-That, Liu et al. 1999, Perry, Ton-That et al. 2002).
In silico analyses by Haft et al. (2012) identified several species of gram-negative bacteria and archaea that encode a set of Sec substrates, that share a C-terminal domain structurally similar to the tripartite structure of the sortase substrates (Haft, Paulsen et al. 2006, Haft, Payne et al. 2012). While species encoding these putative substrates lack sortase homologs, genome comparisons revealed that each genome encoding such substrates also encodes at least one member of a well-defined protein family, the exosortases and archaeosortases in bacteria and archaea, respectively (Haft, Paulsen et al. 2006, Haft, Payne et al. 2012).

The first in vivo characterization of an archaeosortase was carried out in the model archaeon, Haloferax volcanii (Abdul Halim, Pfeiffer et al. 2013), which has a cell wall consisting of a single protein, the S-layer glycoprotein (Sumper, Berg et al. 1990). This glycoprotein was previously thought to be anchored to the cell membrane by the intercalation of a C-terminal TM segment (Sumper, Berg et al. 1990). However, we have demonstrated that this TM segment, which is part of a conserved tripartite structure, is processed in an archaeosortase A (ArtA)-dependent manner (Abdul Halim, Pfeiffer et al. 2013). In light of previous results demonstrating that this extensively studied cell wall subunit is lipid modified with arachetidic acid (2,3-di-O-phytanlyl-sn-glycerylphosphate) (Kikuchi, Sagami et al. 1999, Konrad and Eichler 2002, Nishimura and Eguchi 2006, Kandiba, Guan et al. 2013), we hypothesized that ArtA also mediates proteolysis-coupled lipid-modification. Cells lacking ArtA exhibit a growth defect, are relatively small, and have a severe motility defect (Abdul Halim, Pfeiffer et al. 2013). While some of these
phenotypes are consistent with defects that might be caused by an unstable cell wall due to the improper surface anchoring of the unprocessed SLG, *H. volcanii* is predicted to contain eight additional ArtA substrates, so the failure to accurately process these other substrates might also contribute to the observed phenotypes (Haft, Payne et al. 2012).

In this study, we present data confirming that the S-layer glycoprotein is lipid-modified in an ArtA-dependent manner. Moreover, a replacement mutant of the SLG in which the conserved tripeptide proline, glycine, phenylalanine (PGF) has been permutated to PFG, is not processed, providing the first experimental confirmation of the importance of the PGF motif for ArtA processing. This mutant also allowed the identification of phenotypes that are, at least in part, due to an improperly anchored S-layer.

### 3.4 Materials and methods

**Reagents.** All enzymes used for standard molecular biology procedures were purchased from New England BioLabs. The ECL Prime Western blotting system detection and horseradish peroxidase linked goat anti-rabbit IgG antibodies were purchased from Amersham Biosciences. The polyvinylidene difluoride membrane, MF membrane filters (0.025 µm), and Ultrace-3K membrane were purchased from Millipore. The DNA and plasmid purification kits were purchased from Qiagen. NuPAGE gels, buffers, reagents, and the Pro-Q Emerald 300 glycoprotein stain kit were purchased from Invitrogen. Difco agar and Bacto yeast extract were purchased from Becton, Dickinson, and Company. Peptone was purchased from Oxoid. 5-Fluoroorotic acid (5-FOA) was purchased from
Toronto Research Biochemicals. $^{14}$C mevalonic acid was a gift from Henry M. Miziorko, University of Missouri. All other chemicals and reagents were purchased from either Fisher or Sigma.

**Strains and growth conditions.** The plasmids and strains used in this study are listed in Table 1. *H. volcanii* strains were grown at 45 °C in solid agar (1.5% w/v) with semi-defined casamino acid (CA) medium supplemented with tryptophan and uracil (50 µg/ml final concentration) (Dyall-Smith 2004). *H. volcanii* strains transformed with pTA963 or its derivatives were grown in CA medium supplemented with tryptophan (50 µg/ml final concentration). For selection of the substitution mutants (see below), cells were grown at 45°C in CA medium supplemented with 5-FOA (150 µg/ml final concentration) and uracil (10 µg/ml final concentration). *Escherichia coli* strains were grown at 37°C in NZCYM medium supplemented with ampicillin (200 µg/ml) (Blattner, Williams et al. 1977).

**Generation of chromosomal substitutions.** Chromosomal substitutions were generated using a homologous recombination (pop-in pop-out) method previously described by Allers and Ngo (Allers, Ngo et al. 2004). To facilitate the recovery of mutant recombinants, the construct used for homologous recombination also contained the *trpA* gene (encoding gb|AAA72864.1) between the *slg* (encoding gb|ADE05144.1) (400 nucleotides downstream of SLG stop codon) and the downstream *hvo_2073* (encoding WP_004041957.1), to minimize possible polar effects. Plasmid constructs were generated using overlap polymerase chain reaction (PCR) as described previously (Tripepi, Imam et
Upon homologous recombination, cells having trpA inserted into the chromosome were selected by growing them on CA agar plates lacking tryptophan. Recombinants were sequenced to identify the cell colonies containing an slg gene with the PFG codon mutation, slg\textsuperscript{G796F,F797G} (Fig. S1). An isogenic strain, FH37, containing the trpA gene insertion, but having an unaltered slg PGF motif, was constructed and used as wild-type control.

**Plasmid transformation:** The wild-type FH37, ΔartA and slg\textsuperscript{G796F,F797G} strains were transformed with non-methylated plasmid pTA963, or pTA963 expressing ArtA (WP_004044024.1) (pAF9) (Abdul Halim, Pfeiffer et al. 2013), using the standard polyethylene glycol (PEG) transformation protocol (Dyall-Smith 2004). The pTA963 is an overexpression vector derived from pBluescript II with the *H. volcanii* pHV2 replication origin in addition to both *pyrE2* (encoding WP_004044590.1) and *hdrB* (WP_004044748.1) markers. The plasmid also contains 6X His-tag at the Mutiple Cloning Site (MCS) by the insertion of 26-bp fragment of (CAC)\textsuperscript{6} tract at the NdeI and PciI site (Allers, Barak et al. 2010).

**CsCl purification of S-layer glycoprotein.** The purification of SLG was performed by cesium chloride (CsCl) gradient centrifugation as described previously (Tripepi, You et al. 2012). Briefly, colonies from a solid-agar plate were inoculated into 5 ml CA liquid medium. Two liters of CA medium were inoculated with this 5-ml culture, and the cultures were harvested at optical density (OD) of approximately 0.3 at 600 nm [OD\textsubscript{600}] by centrifugation at 13358 \times g (JA-10 rotor; Beckman) for 30 min. The supernatant
was centrifuged again (13358 × g for 30 min) and incubated at room temperature with 4% w/v polyethylene glycol (PEG) 6000 for 1 hour. The PEG-precipitated proteins were then centrifuged at 39175 × g (JLA 16.250) for 50 min at 4°C, and the SLG was purified by CsCl density gradient centrifugation (overnight centrifugation at 237083 × g) (VTI-65.1 rotor; Beckman). CsCl was dissolved in a 3 M NaCl saline solution to a final density of 1.37 g cm⁻¹.

**Protein extraction, LDS-PAGE and Coomassie staining.** Protein from cell pellets, CsCl purified samples, and trichloroacetic acid (TCA)-precipitations of *H. volcanii* strains were prepared, separated by lithium dodecyl sulfate polyacrylamide gel electrophoresis (LDS-PAGE) and stained with Coomassie Brilliant Blue as previously described (Abdul Halim, Pfeiffer et al. 2013). Briefly, the liquid cultures were grown until the mid-log phase ([OD₆₀₀] of ~0.5). Subsequently, cells were collected by centrifugation at 4,300 × g for 10 min. Cell pellets were resuspended and lysed in 1% LDS supplemented with 100 mM dithiothreitol (DTT). The electrophoresis of the protein samples was performed with 7% tris-acetate (TA) LDS-PAGE gels under denaturing conditions by using TA running buffer. The Coomassie staining were carried out by incubating the gels with Coomassie Brilliant Blue G250 dye for 2 hours followed by washes in De-staining #1 solution (40% methanol, 7% acetic acid) for 2 hours and De-staining #2 solution (2% methanol, 8% acetic acid) for at least 6 hours.

**Mass spectrometry analysis.** The band corresponding to SLG and SLG^{G796F, F797G} from the CsCl purified samples of wild-type FH37 and slg^{G796F, F797G} strains, respectively,
were excised from Coomassie-stained 7% TA LDS-PAGE gels followed by in-gel digestion with trypsin, as previously described (Shevchenko, Tomas et al. 2006, Lin and Garcia 2012). Upon centrifugation of the digested sample, the resulting supernatants were dried in a vacuum centrifuge and reconstituted in 0.1% acetic acid for Mass Spectrometry (MS) analysis. SLG samples were desalted as previously described (Shevchenko, Tomas et al. 2006).

A fused silica microcapillary column packed with Reprosil-pur C18 resin (3 μm, Dr. Maisch GmbH) was fitted with a fused silica emitter (New Objective; 10 μm tip). After using an Eksigent NanoLC AS-2 autosampler to load a sample, an Eksigent NanoLC 2D Plus HPLC system was used to deliver a 2-phase gradient: 0.1% formic acid in water (A) or acetonitrile (B). The HPLC was coupled to a linear ion trap-Orbitrap mass spectrometer (LTQ-Orbitrap Elite; Thermo Scientific).

SLG peptides were separated over an 80 minute gradient (250 μL/min constant flow): 2%B for 2 minutes, 12%B for 10 minutes, 35%B for 55 minutes, and 95%B for 10 minutes, and 95%B for 5 minutes. A full MS spectrum was obtained at 60,000 resolution followed by eight data-dependent MS/MS acquisitions of the eight most abundant ions using HCD fragmentation.

SLG data were analyzed by searching against an *H. volcanii* proteome database (Uniprot: http://www.uniprot.org/taxonomy/309800) using SEQUEST (precursor mass tolerance: 10pm; fragment mass tolerance: 0.5 Da, max missed cleavages: 2, false discovery rate: 0.1). Methionine oxidation was set as a variable modification, and cysteine
carbamidomethylation was set as a fixed modification. Xcalibur Qual software was used for manual visualization and analysis of raw data. Abundance values were obtained by integrating the area under the XIC. To correct for differences in protein loading between samples, abundance values were normalized to the average abundance of three internal peptides that were identified in both SLG and SLG<sup>G796F, F797G</sup> (AA 453-467: SGDGSSILSLTGTYR, m/z: 757.379; AA 579-588: VTAHILSVGR, m/z: 526.814; AA 487-504: SLTSEFTSGVSSNSIR; m/z: 930.453), which should have the same relative abundance in both samples. Normalization was completed by dividing the raw abundance value of the peptide of interest by the average of the raw abundances of the three internal peptides.

**Lipid radiolabeling.** The wild-type FH37, ΔartA, and slg<sup>G796F, F797G</sup> cells were grown in 5 ml liquid CA. Upon reaching mid-log phase ([OD<sub>600</sub>] ~0.5), 20 µl of each culture was transferred into 1 ml fresh liquid CA in which <sup>14</sup>C mevalonic acid, resuspended in ethanol, was added to a final concentration of 1 µCi/ml. The cultures were harvested after 24 hours for the wild type FH37, 96 hours for the ΔartA strain, and 72 hours for the slg<sup>G796F, F797G</sup> strain when they had reached the [OD<sub>600</sub>] of 0.48, 0.41 and 0.40, respectively. Similar radiolabeling methods were also carried out for the wild type + pTA963, ΔartA + pTA963 and ΔartA + artA strains (FH37 + pTA963, AF103 + pTA963, and AF109, respectively). Proteins were precipitated from 1 ml cultures with 10% TCA followed by delipidation as described previously (Kikuchi, Sagami et al. 1999, Konrad and Eichler 2002). The delipidated proteins were separated by 7% TA LDS-PAGE. For analysis of the
samples, the gel was dried onto blotting paper using a Gel Dryer (Bio-Rad Model 583), and exposed to a Phosphor Screen (Molecular Dynamics) for 3 weeks, and analyzed using a Typhoon Imager (Amersham Biosciences).

**Glycosylation staining.** The cell lysates of the wild-type FH37, \( \Delta artA \), \( slg^{G796F} \), \( F797G \), and \( \Delta aglB \) strain cultures grown to mid-log phase (\([OD_{600}] \sim 0.5\) ), were separated by 7% TA LDS-PAGE. The protein glycosylation was visualized using the Pro-Q Emerald 300 glycoprotein gel staining kit (Molecular Probes) as described previously (Steinberg, Pretty On Top et al. 2001).

**Electron Microscopy.** The transmission electron microscopy (TEM) was carried as previously described (Abdul Halim, Pfeiffer et al. 2013) with the following modifications; the wild-type + pTA963, \( \Delta artA \) + pTA963, and \( slg^{G796F, F797G} \) + pTA963 cells were grown in 5 ml liquid semi-defined CA media until they reached \([OD_{600}] \) of 0.363, 0.330 and 0.468, respectively. Next, the cells were concentrated by centrifugation at 6800 \( \times g \) for 1 minute, washed with 18% salt water, and examined using JEOL10 transmission electron microscope.

**Growth Curve.** Cultures for the wild-type FH37, \( \Delta artA \), and \( slg^{G796F, F797G} \) strains were inoculated into 5 ml liquid CA media from colonies and grown to \( OD_{600} \sim 0.55 \). Subsequently, 10 \( \mu l \) of each culture was transferred to 5 ml liquid CA media and grown to stationary phase, with \( OD_{600} \) readings taken every 3-4 hours. The average \([OD_{600}] \) readings of 10 replicates were used to construct the growth curve.
**Light microscopy.** The wild-type FH37, \( \Delta artA \), and \( slg^{G796F, F797G} \) strains were grown in 5 ml CA media until they reached an \([OD_{600}]\) of 0.4-0.5. Subsequently, 2 ml of each culture was concentrated by centrifugation at \( 4,911 \times g \) for 1 min and pellets were resuspended in 10 \( \mu l \) of liquid CA media. Then, 1 \( \mu l \) of the concentrated cells was transferred to a microscope slide and observed using the light microscope as described previously (Abdul Halim, Pfeiffer et al. 2013).

### 3.5 Result and discussion

**The PGF motif is critical for ArtA-dependent SLG C-terminal processing.** To determine whether the conserved PGF-motif is indeed critical for ArtA-dependent C-terminal processing, we engineered an SLG substitution mutant, in which the codon encoding the conserved PGF motif was mutated to a PFG motif (Fig. 1A). *In silico* analysis revealed that the glycine and phenylalanine of this highly conserved motif, found at the start of the conserved C-terminal tripartite structure of predicted ArtA substrates, are invariant, suggesting that the motif is crucial for ArtA-recognition (Haft, Payne et al. 2012). Using homologous recombination, the \( slg^{G796F, F797G} \) strain was constructed, in which a chromosomal copy of \( slg \) was replaced with an \( slg^{G796F, F797G} \) mutation and selected for by a \( trpA \) gene inserted downstream (Fig. S1). Additionally, an isogenic strain encoding unaltered SLG C-terminal PGF motif with an exactly corresponding \( trpA \) insertion, FH37, was also constructed. This strain, which exhibits wild-type morphology and growth phenotypes, was used as the wild-type control in our experiments.
The SLG and SLG^{G796F, F797G} variant protein constructs were purified from the supernatants of the respective strain cultures, using CsCl density gradient centrifugation. Proteins in CsCl gradient fractions containing the SLG or SLG^{G796F, F797G} were separated using LDS-PAGE. Purified SLG^{G796F, F797G}, which, similar to the SLG purified from ΔartA supernatants, migrated slightly faster than SLG purified from the wild-type supernatants (Fig. S2 and (Abdul Halim, Pfeiffer et al. 2013)), was trypsin digested in the gel, and subjected to mass-spectrometry analysis (Lin and Garcia 2012). SEQUEST analysis of these trypsin-digested peptides revealed that both the SLG and SLG^{G796F, F797G} mutant construct yield similar peptides (SLG coverage: 12.8%; SLG^{G796F, F797G} coverage: 20.6%), including the N-terminal peptide of SLG upon signal peptidase I processing. Total extracted ion chromatograms (XICs) revealed that this N-terminal peptide (AA 35-48) has very similar abundance (same order of magnitude) and elutes simultaneously in the SLG and SLG^{G796F, F797G} samples (SLG: 12.2; SLG^{G796F, F797G}: 35.1 normalized abundance) (Fig. 1B, D). Consistent with previous results, mass spectrometric analysis of the trypsin-digested wild-type SLG detected statistically insignificant levels of peptides, essentially equivalent to background noise, that were derived from the C-terminal end of the protein subsequent to the PGF motif (Fig, 1C, D). This result is in accordance with the C-terminal region being processed by ArtA (Abdul Halim, Pfeiffer et al. 2013). Conversely, similar to SLG isolated from the ΔartA strain, the C-terminal peptides AA 813-824: VALVGAALLALR (Fig. 1C, D) was readily detected for digested SLG^{G796F, F797G} (3 orders of magnitude higher abundance than wild-type SLG). MS/MS fragmentation of the C-
terminal peptide yielded near complete coverage of the possible fragment ions, lending strong support that it was correctly identified (Fig. 1E). These results, in combination with the finding that the N-terminal peptide is found in roughly equal abundance, lends strong support that the C-terminal peptide is not processed in the SLG$^{G796F, F797G}$ and that ArtA-processing requires that a substrate contain the conserved PGF motif. It is possible that the retention of C-terminal domain results in faster migration of the SLG$^{G796F, F797G}$ compared to the SLG on the LDS-PAGE gels due to altered detergent binding to the hydrophobic stretch (Rath, Glibowicka et al. 2009).

At this stage, the mass spectrometric analysis that we carried out on the digested SLG was unsuccessful in identifying either the ArtA processing site or the lipid moiety that modifies this glycoprotein. The inability to identify these post-translational modifications is due to challenges in determining the C-terminal peptide mass of the processed SLG as this region contains a stretch of threonine amino acid residues, which are potential targets for O-glycosylation, a poorly understood post-translational modification in archaea (Sumper, Berg et al. 1990).

**H. volcanii SLG is lipid-modified in an ArtA-dependent manner.** Radiolabeled mevalonic acid, a natural prenyl group precursor in archaeal lipid biosynthesis pathways, was previously shown to be incorporated into the mature SLG (Kikuchi, Sagami et al. 1999, Konrad and Eichler 2002). Recently, this lipid covalent linkage to the SLG has been determined to specifically involve the archaetidic acid (2,3-di-O-phytanyl-sn-glycerylphosphate) (Nishimura and Eguchi 2006, Kandiba, Guan et al. 2013). However,
the molecular machineries involved in creating this covalent lipid linkage to the SLG had not yet been elucidated. To determine whether this SLG lipid-modification is ArtA-dependent, we carried out the $^{14}$C mevalonic acid radiolabeling on the wild-type as well as $\Delta artA$ strains. After proteins in delipidated radiolabeled cell extracts were separated by LDS-PAGE, fluorography of the gels detected a significant band at the size corresponding to the SLG in protein extracts isolated from wild-type cells while no signal was detected in that region of the gel for cell extracts of strains lacking ArtA (Fig. 2). It is unlikely that the failure to detect SLG-labeling in the $\Delta artA$ strain extracts was due to a defective uptake or incorporation of the radiolabeled mevalonic acid into the cytoplasmic membrane, as additional lower molecular weight radiolabeled bands were reproducibly observed for protein in cell extracts from both the wild-type and $\Delta artA$ cells. These low molecular weight radiolabeled proteins, perhaps represent the ArtA-independent N-terminally lipidated lipoproteins. Lipobox-containing $H. volcanii$ proteins, in which an invariant cysteine near the N-terminus of the mature protein is essential for processing the signal peptide and is required for cell-surface anchoring in a subset of proteins, have previously been characterized (Storf, Pfeiffer et al. 2010). These results, suggesting that SLG lipid-modification is ArtA-dependent, were further supported by the finding that SLG radiolabeling was restored when the chromosomal $artA$ deletion is complemented by $artA$ overexpression in trans (Fig. S3). ArtA-dependent radiolabeled protein bands other than the SLG were not observed in this fluorography, which might be expressed at low concentrations or may not be expressed at all under the conditions tested. Conversely,
Coomassie staining of proteins separated by LDS-PAGE suggest that SLG, as indicated by its corresponding protein band previously defined by mass spectrometry and Western blot analysis (Fig. 1, S2, and (Abdul Halim, Pfeiffer et al. 2013)), is highly abundant in both wild-type and \( \Delta artA \) cell extracts, underscoring the fact that the lack of a signal in the fluorograph was not due to low protein abundance (Fig. 2).

\textbf{SLG}^{G796F, F797G} \textbf{mutant protein is not lipid-modified.} To determine the importance of the conserved PGF motif for lipid modification, we also carried out the \(^{14}\)C mevalonic acid labeling on \textit{H. volcanii} slg\textsuperscript{G796F, F797G} mutant, as described above. Similar to results that were obtained for the SLG isolated from the \( \Delta artA \) strain, fluorographic analysis failed to detect the radiolabeled SLG\textsuperscript{G796F, F797G} (Fig. 2). This finding strongly suggests that the SLG lipid modification is not only ArtA-dependent, it also requires the conserved C-terminus PGF motif. Since neither the SLG expressed in the \( \Delta artA \) strain nor SLG\textsuperscript{G796F, F797G} are covalently lipid-modified, it is also possible that the absence of this lipid moiety contributes to the unusual migration behavior of these proteins on the LDS-PAGE gel (Konrad and Eichler 2002).

Alternatively, it is also possible that the lack of other post-translational modifications, such as glycosylations, could contribute to the migration shifts of the SLG isolated from the \( \Delta artA \) strain and the SLG\textsuperscript{G796F, F797G}. However, the LDS-PAGE glycostaining demonstrated that both the SLG of the \( \Delta artA \) strain and SLG\textsuperscript{G796F, F797G} are glycosylated (Fig. S4). The SLG obtained from a strain lacking the oligosaccharyltransferase AgIB (WP\_013035363.1), which is required for N-glycosylation
under optimal salt concentrations, exhibits migration behavior similar to that of wild-type, suggesting that the lack of AglB-dependent N-glycosylation of the SLG is insufficient to cause such a migration shift (Fig. S4) (Abu-Qarn, Yurist-Doutsch et al. 2007). However, since N-glycosylation of the SLG can involve at least one additional, poorly understood, N-glycosylation pathway (Kaminski, Guan et al. 2013), as well as at least one O-glycosylation pathway (Sumper, Berg et al. 1990) whose components to date remain elusive, determining whether either of these pathways may be responsible for the SLG migration shift presents challenges and is beyond the scope of this paper.

The $slg^{G796F,F797G}$ strain growth and cell morphology defects are only slightly less severe than those of the $\Delta artA$ strain. We previously reported that the $\Delta artA$ strain exhibits severe growth and morphology defects, phenotypes that can be complemented by expressing ArtA in trans (Abdul Halim, Pfeiffer et al. 2013). These phenotypic defects of the $\Delta artA$ strain may be due solely to improper anchoring of the SLG to the cell membrane or could be due at least in part to the lack of processing of the other ArtA substrates (Abdul Halim, Pfeiffer et al. 2013). In the $slg^{G796F,F797G}$ mutant strain, presumably, all of the other ArtA substrates are processed properly, so observing phenotypic defects matching those of the $\Delta artA$ strain would indicate that those defects are specifically due to disrupted processing of the S-layer glycoprotein. Therefore, we first determined the growth curve for the $slg^{G796F,F797G}$ strain in semi-defined casamino acid (CA) media, and found that the $slg^{G796F,F797G}$ strain displays a growth defect similar to that of the $\Delta artA$ strain (Fig. S5). It should be noted that transformation of $\Delta artA$ but not $slg^{G796F,F797G}$ strain with the $pyrE2$-
bearing expression plasmid pTA963 (from which we derived the construct used to overexpress ArtA (pAF9) for the complementation studies—see Fig. S3 and (Abdul Halim, Pfeiffer et al. 2013)) in CA medium lacking uracil to select for the transformants, causes a more severe growth defect than is displayed by the non-transformed mutant strains grown in uracil-supplemented CA media (Fig. S5). It is possible that the addition of uracil to non-transformed culture aids growth, possibly, since the pTA963 pyrE2 gene required for uracil biosynthesis was used as its selective marker for the transformants (Allers, Barak et al. 2010). The transformants may have had a growth disadvantage because they rely on the pTA963 pyrE2 gene product for the uracil biosynthesis, in contrast to the untransformed cells where uracil was readily available in the media. The pTA963 also contains a replication origin derived from the H. volcanii natural plasmid, pHV2, which has yet to be fully characterized, and its replication may incur additional metabolic cost to the cells compared to the untransformed cells (Rosenshine, Tchelet et al. 1989, Allers, Barak et al. 2010).

Since the presence of pTA963 affected the growth rate of the ΔartA strain, cell morphology analyses were performed for both sets of cultures. Using light microscopy to examine cells from cultures of untransformed cells grown in liquid CA media, we determined that wild-type cells, inoculated from colonies grown on agar plates, are predominantly pleomorphic disk-shaped cells (~300 cells per panel), while approximately 20% of the cells are rod-shaped with an average cell length of approximately 3.9 µm (SD: 1.5 µm). A similar portion of the ΔartA strain cultures are rod-shaped, however, these rods
are shorter (2.2 µm; SD: 1.1 µm) than the wild-type rods. While there appear to be fewer disk-shaped cells in the ΔartA strain cultures (~200 cells per panel) compared to the wild-type sample, the highly irregular shapes and large size of these cells, compared to the wild-type cells, suggests that the ΔartA cells may be forming aggregates or are enlarged due to abnormal cell division (Duggin, Aylett et al. 2015). The slg<sup>G796F, F797G</sup> cultures also contain irregular disk-shaped cells (~250 cells per panel), as well as rod-shaped cells of shorter average length (2.3 µm; SD: 0.3 µm); however, the percentage of rods is reduced to approximately 2% (Fig. 3).

The presence of the pTA963 plasmid has little or no effect on the average length of rods observed among either the wild-type or slg<sup>G796F, F797G</sup> cells, with an average length of 3.9 µm (SD: 1.3 µm) and 2.9 µm (SD: 1.2 µm), respectively. In contrast, the ΔartA rod-shaped transformants are slightly shorter with an average length of 1.8 µm (SD: 0.3 µm).

Moreover, in cultures of cells containing the plasmid, approximately 90% of wild-type and slg<sup>G796F, F797G</sup> transformants, and 100% of ΔartA transformants, are rods (Fig. 3).

Interestingly, while serial transfers of the liquid wild-type culture to new liquid media results in more than 90% disk-shaped cells, regardless of whether the cells contain the plasmid, the morphology phenotypes of the mutants vary significantly from the wild-type cells upon extended liquid-to-liquid transfer. For both the slg<sup>G796F, F797G</sup> and ΔartA strains, after two transfers in liquid media, untransformed cells are highly pleomorphic and appear to form clumps of highly deformed disk-shaped cells. Moreover, approximately 75% of slg<sup>G796F, F797G</sup> cells transformed with pTA963 are rod-shaped, upon initial liquid-
to-liquid transfer (data not shown), with 53% remaining rod-shaped after a second transfer. Most unexpectedly, ΔartA cells transformed with pTA963 remained exclusively rod-shaped even after two liquid-to-liquid transfers (Fig. 3).

The biological significance of, and the underlying mechanisms involved in, regulating archaeal cell morphology, has only recently begun to be fully appreciated (Duggin, Aylett et al. 2015). Here we report that the lack of proper processing and anchoring of surface proteins can affect the ability of cells to alter their cell morphology. Based on our results, it appears that one or more ArtA substrates play a critical role in this process. Based on the inability of ΔartA cells, and, to a lesser extent, cells expressing the SLG replacement mutant, but not wild-type cells, to morph from rod-shaped to disk-shaped cells, additional analyses and characterizations of the ArtA substrates may allow us to identify specific proteins involved in regulating cell morphology, and may ultimately lead to the discovery of the molecular mechanisms that allow cells to modify their shapes as conditions demand.

Moreover, while the more severe growth and morphology phenotypes exhibited by the mutant transformants suggest that the ArtA-mediated surface protein processing and anchoring is not limited to the SLG, the phenotypes of the slgG796F, F797G strain also underscore the importance of ArtA-mediated processing and anchoring of the SLG for growth and maintenance of cell morphology (Mescher and Strominger 1976).

The S-layer of slgG796F, F797G cells is thicker than that of the wild-type cells. TEM of thin sections of wild-type and ΔartA cells previously revealed that the mutant strain has
a thicker cell wall (Abdul Halim, Pfeiffer et al. 2013). Considering that morphological differences may also affect cell wall thickness, we used similar technique to compare the S-layer thicknesses of $slg^{G796F,F797G}$ cells to those of $\Delta artA$ and wild-type cells, using cells transformed with pTA963, as the rod-shape morphology is predominant among these cells upon inoculation from colonies grown on solid agar plates into semi-defined CA liquid media (Fig. 3).

Despite having an unprocessed C-terminus and lacking lipid anchoring, the mutant SLG of an $slg^{G796F,F797G}$ strain transformed with pTA963 can be assembled into an S-layer, similar to the SLG in $\Delta artA$ cells. While the average thickness of S-layers of the $slg^{G796F,F797G}$ transformants is approximately 16 nm (SD: 1.6 nm), S-layers of wild-type and $\Delta artA$ transformants have similar, average thicknesses (12 nm) (SD: 2.6 nm and 2.4 nm, respectively) (Fig. 4). Unlike wild-type or $slg^{G796F,F797G}$ transformants, $\Delta artA$ cells are surrounded by substantial amounts of additional loose material. Although the origin and composition of this material is unclear, based on its proximity to the cell wall, it is possible that it is S-layer that has been shed. Supporting this notion, we observed that while the majority of wild-type cells are devoid of this material, the few cells surrounded by loose material exhibit a reduced S-layer width, averaging 9.5 nm. It is possible that under the growth conditions used, the presence of pTA963, or the use of semi-defined CA growth media (cells examined in (Abdul Halim, Pfeiffer et al. 2013) were grown on complex media), may impose greater stress on the $\Delta artA$ strain, causing a large fraction of its unstable S-layer to be shed into its surrounding environment. It is also possible that the
unstable S-layer is extremely sensitive to mechanical stress exerted by the centrifugal force applied to the cultures, even at a low speed, for concentrating and washing the cells prior to high-pressure freezing preservation. While the material is also present in the $slg^{G796F, F797G}$ cultures, it is less abundant than in the $\Delta artA$ cultures, suggesting that in the $slg^{G796F, F797G}$ strain, proper processing and anchoring of ArtA substrates other than the SLG may sufficiently stabilize the S-layer to prevent mass shedding.

Alternatively, this material might be cytoplasmic content released into the extracellular environment by lysed cells, which would indirectly suggest that the $\Delta artA$ cells face a severe cell stability issue under the conditions tested. This is consistent with the more severe growth defect exhibited by $\Delta artA$ transformants, particularly when grown in low salt media, in which defects in cell wall anchoring and processing of additional ArtA substrates may leave the fragile cells more prone to lysis (Abdul Halim, Pfeiffer et al. 2013). In either case, this underscores the importance of proper anchoring of the SLG, as well as the other ArtA substrates for maintaining cell stability in the high salinity environment.

### 3.6 Concluding remark

Anchoring mechanisms involving either N- or C-terminal transmembrane segments have been well characterized in bacteria and archaea (Giménez, Dilks et al. 2007, Albers and Meyer 2011, Dalbey, Wang et al. 2011, Luirink, Yu et al. 2012). However, while anchoring mechanisms involving N-terminal covalently attached lipids have been studied
in detail in bacteria (Zückert 2014, Buddelmeijer 2015), and archaeal proteins requiring the invariant lipobox cysteine for proper surface attachment have been identified (Storf, Pfeiffer et al. 2010), this is the first report to present evidence consistent with protein anchoring of secreted prokaryotic proteins that is mediated through the covalent attachment of a cell membrane lipid to the C-terminus. Our findings provide a sound foundation for future biochemical and molecular biological studies that will allow us to fully understand the mechanisms underpinning the ArtA-dependent processing and lipidation of the C-terminus of the SLG.

The research described here also provides a basis on which to initiate studies aimed at identifying the functions of bacterial exosortases and their substrates, which, although present in a highly diverse set of bacteria, have not yet been studied in vivo. Future investigations of lipid-anchored proteins, as well as comparisons of the organisms that use these lipid-mediated protein-anchoring strategies to those that use other anchoring strategies, will clarify the benefits and limitations of this protein-anchoring mechanism. These comparisons, combined with complementary in vivo studies, will be crucial to gaining a greater understanding of prokaryotic cell surface biogenesis.
3.7 Figures

Figure 3.1: The conserved SLG PGF-motif is critical for ArtA-dependent processing.

The SLG C-terminal region (AA: 766-827) consists of a threonine-rich stretch followed by the conserved PGF motif (bold), a hydrophobic stretch (underline) and positively-charged residues (*end of protein) (A). XICs of gel-purified SLG and SLG\textsuperscript{G796F,F797G} peptides from AA: 35-48 (B); or AA: 813-824 (C). The masses and normalized abundances (NA) of the chromatographic peak are indicated. Graphical representation of the relative abundance of the N-terminal peptide (AA 35-48) and C-terminal peptide (AA813-824) in WT SLG

93
(purple) and SLG$^{G796F, F797G}$ (blue) cells (D) MS/MS fragment ions from C-terminal peptide AA 813-824 ([M+2H]$^{2+}$ = 583.885 m/z) (E). The lines indicate which fragment ions were detected (b ions above the sequence; y ions below). Identification of a large number of the expected fragment ions, lends strong support that the peptide was correctly identified.

Figure 3.2: ArtA and conserved C-terminal PGF motif are required for *H. volcanii* SLG lipid-modification. Fluorography of protein extracts isolated from wild-type FH37 (wt), ΔartA, and slg$^{G796F, F797G}$ cells grown in the presence of 1µCi/ml $^{14}$C mevalonic acid (left). Labeled SLG (arrow) is detected only in the wt extract. Additional labeled lower molecular weight proteins likely are ArtA-independent secreted lipobox-containing proteins (star) (Storf, Pfeiffer et al. 2010). Coomassie stain of protein extracts from wt,
ΔartA, and slg^{G796F, F797G} cells (right). The migration of molecular mass standards is indicated on the left (in kDa).

Figure 3.3: The ΔartA and slg^{G796F, F797G} exhibit distinct cell morphology compared to the wild-type. Phase contrast microscopy of wild-type FH37 (wt), ΔartA, and slg^{G796F, F797G} without or with pTA963 grown to mid-log phase in liquid semi-defined CA media supplemented with tryptophan and uracil or with tryptophan only, respectively. Liquid media were inoculated with cells from colonies grown on solid CA agar plates (top row), or cells that had been grown by serial transfer in liquid media to mid-log phase (bottom row). Size bar: 10 μm.
**Figure 3.4: SLG$^{G796F, F797G}$ forms thicker S-layer compared to wild-type.** Wild-type FH37 (wt), ΔartA, and slg$^{G796F, F797G}$ cells transformed with pTA963 were preserved with high pressure freezing techniques followed by thin slice section microscopy. The S-layer thickness was determined by measuring the region between the cell membrane (white arrowhead) and the outer edge of the S-layer (black arrow). The slg$^{G796F, F797G}$ cells contains wider S-layer (16 nm) compared to the wt (12 nm) and contains material around the cells. While displaying similar S-layer width to the wt, the ΔartA cells generally contained significant amounts of material surrounding the cell compared to the other strains. Size bar: 100 nm.
Figure 3.S1: The PCR screening confirmed the trpA gene insertion for identification of C-terminal PGF motif replacement mutants. (A) Schematic representation of genomic replacement in slg-pfg mutant. Replacement construct includes trp gene to allow selection of recombinants. A strain containing the trpA gene insertion without pgf-codon replacement (FH37) was constructed to exclude effects of trpA on slg-pfg phenotypes. (B) PCR was performed using genomic DNA templates isolated from the H53, FH37, or slg(G796F, F797G) strains, using flanking primers P1 and P6. The amplicon sizes are approximately 1000 nucleotides longer for the FH37 and slg(G796F, F797G) strains than
for the H53 consistent with \textit{trpA} gene insertion. The \textit{pfg}-codon mutation (*) in the \textit{slg-pfg} was confirmed by DNA sequencing.

\textbf{Figure 3.S2: The SLG-PFG exhibits faster migration on LDS-PAGE gel compared to wild-type SLG.} Western blot with anti-SLG of the CsCl purified SLG from the wild-type FH37 (wt) and \textit{slg}^{G796F,F797G} confirmed their expression (left). The coomassie stain of the CsCl-purified SLG and SLG^{G796F,F797G} (right). The migration of molecular mass standards is indicated on the left (in kDa).
Figure 3.S3: *H. volcanii* SLG lipid-modification is ArtA-dependent. Fluorography of protein extracts isolated from wild-type FH37 (wt), ΔartA and ΔartA + artA cells grown in the presence of 1µCi/ml ¹⁴C mevalonic acid (left). Labeled SLG (arrow) is detected in the wt and artA complementation (ΔartA + artA) extract but not in the artA deletion background (ΔartA). Coomassie stain of protein extracts from wt, ΔartA, ΔartA + artA cells (right). The migration of molecular mass standards is indicated on the left (in kDa).
Figure 3.S4: The mobility shift of SLG is not attributed to lack of AglB-dependent N-glycosylation. The cell lysate of the wild-type FH37 (wt), DartA, slg<sup>G796F,F797G</sup>, and DaglB samples stained with Pro-Q Emerald 300 glycoprotein stain indicating the presence of major glycosylated protein bands corresponding to SLG (left). Total protein coomasie stain of this gel reveal staining with Pro-Q Emerald 300 is not due to nonspecific binding to the high abundance proteins (right). The SLG mobility shift observed in DartA and slg<sup>G796F,F797G</sup> strains is not due to lack of AglB-dependent N-glycosylation as the DaglB strain, lacking the oligosaccharyltransferase component known to be involved in <i>H. volcanii</i> SLG N-glycosylation, does not exhibit similar mobility shift. The migration of molecular mass standards is indicated on the left (in kDa).
Figure 3.S5. C-terminally unprocessed SLG-PFG causes an *H. volcanii* growth defect similar to the deletion of *artA*. Growth curves of wild-type FH37 (wt), Δ*artA*, and *slg*<sup>G796F, F797G</sup> strains without (A) or with pTA963 (B) inoculated in 5 ml liquid semi-defined CA media supplemented with tryptophan and uracil or tryptophan only, respectively. While a similar growth defect was observed between the Δ*artA* and *slg*<sup>G796F, F797G</sup>, transformation with pTA963 resulted in a slightly greater growth defect in the Δ*artA* strain.
CHAPTER 4

ArtA-dependent processing of a Tat substrate containing a conserved tripartite structure that is not localized at the C-terminus

This chapter has been published in:

4.1 Abstract

Most prokaryotic secreted proteins are transported to the cell surface using either the general secretion (Sec) or twin arginine translocation (Tat) pathway. A majority of secreted proteins are anchored to the cell surface while the remainder are released into the extracellular environment. The anchored surface proteins play a variety of important roles in cellular processes, ranging from facilitating interactions between cells to maintaining cell stability. The extensively studied S-layer glycoprotein (SLG) of Haloferax volcanii, previously thought to be anchored via C-terminal intercalation into the membrane, was recently shown to be lipidated and to have its C-terminal segment removed in processes dependent upon archaeosortase A (ArtA), a recently discovered enzyme. While SLG is a Sec substrate, in silico analyses presented here reveal that, of eight additional ArtA substrates predicted, two also contain predicted Tat signal peptides, including Hvo_0405, which has a highly conserved tripartite structure that lies closer to the center of the protein than to its C-terminus, unlike other predicted ArtA substrates identified to date. We demonstrate that, even given its atypical location, this tripartite structure, which likely
resulted from the fusion of genes encoding an ArtA substrate and another encoding a cytoplasmic protein, is processed in an ArtA-dependent manner. Using an Hvo_0405 mutant lacking the conserved “twin” arginines of the predicted Tat signal peptide, we show that Hvo_0405 is indeed a Tat substrate and that ArtA substrates include both Sec and Tat substrates. Finally, we confirmed the Tat-dependent localization and SPaseI cleavage site of Hvo_0405 using mass spectrometry.

4.2 Introduction

Prokaryotic cell surfaces are decorated with a diverse set of proteins that perform functions that are crucial to a variety of cellular processes, ranging from cell-cell communication to stress response and cell stability. Most surface proteins are transported across the cytoplasmic membrane via either the Sec or the Twin arginine transport (Tat) pathway (Denks, Vogt et al. 2014, Berks 2015, Goosens and van Dijl 2016). The Sec pathway, which is found in all species of eukaryotes as well as those of both prokaryotic domains, transports proteins across the cytoplasmic membrane in an unfolded conformation. In most organisms, this pathway typically transports the major portion of secreted proteins across the cytoplasmic membrane (Dilks, Rose et al. 2003). However, the haloarchaea, such as *Haloferax volcanii*, are predicted to use the Tat pathway, which transports proteins in a folded conformation, for a large portion of their secreted proteins (Dilks, Rose et al. 2003, Dilks, Giménez et al. 2005). The reason why many haloarchaeal proteins use the Tat pathway is perhaps because proteins transported in an unfolded conformation do not fold
properly in highly saline extracytoplasmic environments lacking ATPase-driven chaperones (Rose, Brüser et al. 2002, Dilks, Rose et al. 2003, Pohlschroder, Dilks et al. 2004). For example, nearly half of all H. volcanii secreted proteins are predicted to be transported to the cell surface via the Tat pathway (Dilks, Giménez et al. 2005, Giménez, Dilks et al. 2007). Once on the cell surface, a subset of secreted proteins are released into the extracellular environment. However, the vast majority are anchored to the cell surface: i.e., to the cytoplasmic membrane of archaea and gram-positive bacteria, to the inner (cytoplasmic) membrane or the outer (surface) membrane of gram negative-bacteria and a subset of archaea, or to the prokaryotic cell wall (Pohlschroder, Dilks et al. 2004, Saier 2006, Nass, Poll et al. 2014, Schneewind and Missiakas 2014). A diverse array of protein anchoring pathways have evolved to facilitate retention of these secreted proteins at the cell surface, including a recently described mechanism that facilitates C-terminal lipid-mediated anchoring of prokaryotic proteins (Haft, Payne et al. 2012, Abdul Halim, Pfeiffer et al. 2013, Abdul Halim, Karch et al. 2015).

The H. volcanii S-layer glycoprotein (SLG), which is the sole component of the cell wall in this halophilic archaeon (Sleytr, Schuster et al. 2014), contains a C-terminal tripartite structure consisting of a PGF motif, followed by a hydrophobic stretch and a positively charged region (Haft, Payne et al. 2012). C-terminal processing and lipid-modification of this protein is dependent on a novel enzyme known as the archaeosortase, which is homologous to the exosortase in gram-negative bacteria (Haft, Payne et al. 2012, Abdul Halim, Pfeiffer et al. 2013, Abdul Halim, Karch et al. 2015). The predicted
substrates of this enzyme contain a highly conserved C-terminal tripartite structure reminiscent of the substrates of a family of acyltransferases found in many gram-positive bacterial species known as sortases (Haft, Paulsen et al. 2006, Haft, Payne et al. 2012, Schneewind and Missiakas 2014, Siegel, Liu et al. 2016). The sortases process and subsequently covalently link their substrates to the bacterial cell wall. Although substrates of both sortase and archaeosortase contain similar highly conserved tripartite structures, and the enzymes each contain similar putative catalytic amino acid triads, no sequence or structural homology has been detected between the sortases and archaeosortases (Haft, Payne et al. 2012).

Recently, we determined that SLG processing and lipidation both depend on the presence of the PGF motif in this ArtA substrate (Abdul Halim, Karch et al. 2015). While only the highly abundant SLG in *H. volcanii* has been confirmed to be processed and anchored in an ArtA-dependent manner, *in silico* analyses predicted that many archaeal and bacterial genomes encode an archaeosortase or exosortase, respectively. These analyses have also shown that any genome that encodes either of these homologous enzymes also encodes secreted proteins that contain the highly conserved tripartite structure (Haft, Payne et al. 2012). In *H. volcanii, in silico* analyses have revealed that, in addition to the SLG, seven proteins carrying the signal peptide as predicted by SignalP and/or PRED-SIGNAL program, contain the highly conserved C-terminal tripartite structure of a predicted ArtA substrate (Bagos, Tsirigos et al. 2009, Haft, Payne et al. 2012).
In addition to the previously confirmed ArtA substrate, SLG, the study described here revealed a second ArtA substrate, Hvo_0405, a predicted Tat substrate containing the conserved tripartite structure of the ArtA substrates but closer to the center of the protein than to the C-terminus. This provides an excellent opportunity to determine whether ArtA can recognize and process proteins in which the conserved tripartite structure is centrally located as oppose to C-terminally located, and whether a subset of Tat substrates can be processed and anchored to the cell surface in an ArtA-dependent manner.

Our *in silico* analyses have revealed that the more central location of the conserved tripartite structure in Hvo_0405 may be the result of a fusion of two genes, one encoding an ArtA substrate and the second encoding a protein containing a NifU-like domain that is commonly found in the Fe-S cluster scaffold protein, NifU of nitrogen-fixing bacteria and rhodobacterial species; the function(s) of these proteins have not yet been determined (Marchler-Bauer, Derbyshire et al. 2015). While Hvo_0405 is not expressed under standard laboratory conditions, we have shown that, when expressed *in trans* in *H. volcanii*, Hvo_0405 can be processed in an ArtA-dependent manner. Moreover, site-directed mutagenesis of the conserved “twin” arginines in the Tat signal peptide prevents processing of this protein. These data strongly suggest that this, and other, Tat substrates, in addition to a subset of Sec substrates that contain the conserved tripartite structure of ArtA substrates, can be processed in an ArtA-dependent manner.
4.3 Materials and methods

**Strains and growth conditions.** The plasmids and strains used in this study are listed in Table 1. *H. volcanii* strain H53 and its derivatives were grown at 45 °C in liquid (orbital shaker at 250 rpm) or on solid semi-defined casamino acid (CA) medium supplemented with tryptophan and uracil (both at 50 µg ml\(^{-1}\) final concentration) (Dyall-Smith 2004). Solid medium plates contained 1.5% (w/v) agar. To ensure equal agar concentrations in all plates, agar was completely dissolved in the medium prior to autoclaving, and autoclaved medium was stirred before plates were poured. Strains transformed with pTA963 were grown on CA medium supplemented with tryptophan (50 µg ml\(^{-1}\) final concentration) (Allers, Barak et al. 2010). For selection of the deletion mutant, 5-fluoroorotic acid (5-FOA) (Toronto Research Biochemicals) (150 µg ml\(^{-1}\) final concentration) and uracil (Sigma) (10 µg ml\(^{-1}\) final concentration) were added to CA medium. *Escherichia coli* strains were grown at 37 °C in NZCYM (Fisher Scientific) medium supplemented with ampicillin (Sigma) (100 µg ml\(^{-1}\) final concentration) (Blattner, Williams et al. 1977).

**Nucleotide and amino acid sequence analysis.** Protein and nucleotide alignments of *Haloferax* spp. Hvo_0405 homologs were performed with Clustal Omega using the sequences: *H. alexandrinus* (Gene locus tag: RS08345 and RS08350), *H. denitrificans* (Gene locus tag: RS09235 and RS09240), *H. elongans* (Gene locus tag: RS04170 and RS04175), *H. gibbonsii* (Gene locus tag: ABY42_01850 and ABY42_01855), *H. larsenii* (Gene locus tag: RS15950 and RS15955), *H. lucentense* (Gene locus tag: RS15835 and
RS15830), *H. mediterranei* (Gene locus tag: AFK18107 and AFK18108), *H. mucosum* (Gene locus tag: RS07215 and RS07210), *H. prahovense* (Gene locus tag: RS02190 and RS02195), and *H. sulfurifontis* (Gene locus tag: RS09845 and RS09840). Codon sequences were also determined by Geneious version 6.1.8 (Biomatters Ltd.). Signal sequence prediction was performed using TatFind Server (http://signalfind.org/tatfind.html) (Rose, Brüser et al. 2002), SignalP 4.1 Server (http://www.cbs.dtu.dk/services/SignalP/) (Petersen, Brunak et al. 2011), and TatP 1.0 Server (http://www.cbs.dtu.dk/services/TatP/) (Bendtsen, Nielsen et al. 2005).

**Plasmid construction.** The pH24 plasmid, used in generating the Δ*hvo_0405* strain, was constructed by first PCR amplifying 700 bp 5’ of *hvo_0405* using primers FW_0405_XhoI (OP 1) and RV_0405_Up-DW as well as 700 bp 3’ of *hvo_0405* using primers FW_0405_Up-DW and RV_0405_XbaI (OP 2). Subsequently, these two fragments were joined into a single amplicon by overlap-extension PCR and cloned into pTA131 using XhoI and XbaI. Primers FW_0405_XhoI (OP 1) and RV_0405_XbaI (OP 2) were used to verify the insert by DNA sequencing.

The pH25 plasmid used for complementation experiments was constructed by PCR amplifying the 1,971 bp *hvo_0405* coding sequence with a 3’ GGP linker and 6xHis tag using primers FW0405NdeI (IP 1) and RV0405EcoRIHIS (IP 2). This amplicon was cloned into pTA963 using NdeI and EcoRI. Primers FW0405NdeI (IP 1) and RV0405EcoRIHIS (IP 2) were used to verify the insert by DNA sequencing.
The pJS143 plasmid used for hvo_0405 expression in E. coli BL21(DE3) strain was constructed by PCR amplifying 1,281 bp of hvo_0405 (bp 64-1344) using primers Hvo-0405-1F and Hvo-0405-1R. This amplicon was cloned into pET-22b(+) using NdeI and HindIII. The insert was sequenced using primers T7 and T7term.

The pJS151 plasmid containing the hvo_0405 coding sequence with codons for the twin arginine residues at the N-terminus mutated to two lysine residues, was constructed by PCR amplification using primers Hvo0405-3F and RV0405EcoRIHIS (IP 2). This amplicon was cloned into pTA963 using NdeI and EcoRI. The insert was sequenced using primers pTA963-1F and pTA963-1R.

Polymerase, ligase, and restriction enzymes were purchased from New England Biolabs. Plasmids were initially transformed into E. coli DH5α cells and grown on NZCYM 1.5% agar plates or in liquid media supplemented with ampicillin (100 µg ml⁻¹ final concentration). Plasmid preparations were performed using QIAprep Miniprep (Qiagen) or PureLink Plasmid Prep (Invitrogen) kits. Prior to H. volcanii transformation, plasmids were passaged through the dam− E. coli strain DL739 (Blyn, Braaten et al. 1990). All primers are listed in Table S1.

**H. volcanii transformation and chromosomal substitution.** H. volcanii was transformed using the polyethylene glycol (PEG) method (Dyall-Smith 2004). Deletion of the hvo_0405 coding sequence in both the H53 and ΔartA backgrounds was completed using the pFH24 plasmid, and the homologous recombination (pop-in pop-out) method and 5-
FOA (Allers, Ngo et al. 2004). To confirm the deletion of the \textit{hvo\_0405} coding sequence in the resulting FH24 and FH35 strains, PCR amplification of the \textit{hvo\_0405} region was performed using the FW\_0405\_XhoI (OP 1) and RV\_0405\_XbaI (OP 2) primers in both the H53 and $\Delta artA$ as well as FH24 and FH35 deletion strains.

\textbf{Subcellular fractionation of \textit{H. volcanii} cultures.} \textit{H. volcanii} transformants were cultured in CA liquid media supplemented with tryptophan (50 µg ml$^{-1}$ final concentration) to an OD$_{600}$ of approximately 0.5. Cells were separated from the supernatant by centrifugation at 6,000 x g for 5 minutes; subsequently, supernatants were spun a second time and frozen at -80 °C along with the cell pellets. Cell pellets were fractionated further by first lysing the cells in phosphate buffered saline (PBS) with 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF; 1 mM) (Sigma), phenylmethylsulfonyl fluoride (PMSF; 1 mM) (Sigma), and ethylenediaminetetraacetic acid (EDTA; 10 mM) (Invitrogen) protease inhibitors. The cell lysate was centrifuged at 6,000 x g for 10 minutes to pellet unlysed cells and the cleared cell lysate was then centrifuged at approximately 303,500 x g for 30 min in an Optima™ MAX-TL Ultracentrifuge (Beckman). Subsequently, the membrane-rich pellet was resuspended in PBS. Both the membrane and cytoplasm fractions were centrifuged again at 303,500 x g for 30 min. Proteins in both the cytoplasm and culture supernatant fractions were precipitated using 10% trichloroacetic acid (ThermoFisher Scientific). Protein pellets from membrane, cytoplasm, and culture
supernatant fractions were resuspended in NuPAGE lithium dodecylsulfate (LDS) sample buffer (Invitrogen) in volumes proportional to the original culture.

**Anti-Hvo_0405-N-term antibody.** The anti-Hvo_0405-N-term antibody was generated by first amplifying a 1,281 bp fragment of *hvo_0405* (bp 64-1344) that was 3’ of the predicted signal peptide and 5’ of predicted glycosylation sites, and cloned into pET-22b(+) using *Nde*I and *Hind*III restriction sites to create pJS143. The pJS143 plasmid was transformed into the *E. coli* BL21(DE3) strain (Studier and Moffatt 1986) and the *hvo_0405* fragment, under control of a T7 promoter, was expressed via genome-encoded T7 polymerase induction with isopropyl β-D-1-thiogalactopyranoside (IPTG) (1 mM) (Sigma) in NZCYM supplemented with ampicillin (100 µg ml$^{-1}$ final concentration) at 37 °C. After 4-5 hours of induction, cells were chilled in ice water and pelleted at 4,000 x g for 15 min at 4 °C. Cell pellets were frozen at -80 °C.

Subsequently, the C-terminally 6xHis-tagged N-terminal fragment of Hvo_0405 was purified by Nickel (Ni)-affinity chromatography. Using Ni-NTA Superflow beads (Qiagen) and denaturing conditions (8 M urea, 100 mM NaH$_2$PO$_4$, 10 mM Tris Base), the purified N-terminal protein fragment of ArtA processed Hvo_0405 was eluted using the manufacturer’s instructions. Purified protein was then concentrated using Amicon ultra centrifugal filters with a 3 kDa molecular weight cut-off (Millipore), and PBS added twice in order to reduce the urea concentration to ~0.4 M. The protein concentration was then determined using a NanoDrop 1000 spectrophotometer (Thermo-Fisher Scientific). Polyclonal antibodies were generated in two rabbits using Complete Freund’s Adjuvant
over a 90-day production schedule (Cocalico Biologicals, Inc., Reamstown, PA). Both rabbits showed nearly identical immune responses. Specificity of the anti-Hvo_0405-N-term antibodies for Hvo_0405 was verified by immunoblotting against purified recombinant Hvo_0405.

**Immunoblotting.** Liquid cultures were grown until the mid-log phase (OD$_{600} 0.4 - 0.5$) and the cells were harvested by centrifugation at 4,300 x g for 10 min at 4 °C. Cell pellets were resuspended and lysed in 1% (vol/vol) NuPAGE LDS supplemented with 100 mM dithiothreitol (DTT) (Sigma) and stored at -20 °C. *H. volcanii* cell pellets or subcellular fractions were electrophoresed on either 10% or 12% Bis-Tris NuPAGE polyacrylamide gels (Invitrogen) with NuPAGE 3-(N-morpholino)propanesulfonic acid (MOPS) sodium dodecyl sulfate (SDS) running buffer (Invitrogen). Proteins were then transferred to a polyvinylidene difluoride (PVDF) membrane (Millipore) using a semi-dry transfer apparatus at 15 V for 30 minutes (BioRad). Next, the membrane was washed twice in PBS, blocked for one hour in 3% bovine serum albumin (BSA) in PBS, and washed twice in PBS with 1% Tween-20 and once in PBS. For primary antibodies, either the mouse α-Penta-His antibody (Qiagen; catalog # 34660; 200 µg ml$^{-1}$) was used at 1:1,000 in 3% BSA in PBS with sodium azide, or the rabbit α-Hvo_0405-N-term serum was used at 1:10,000 in 3 % BSA in PBS with sodium azide. Subsequently, the membrane was washed twice in PBS with 1% Tween-20 and once in PBS. For secondary antibodies, either the Amersham ECL α-mouse IgG, horseradish peroxidase (HRP)-conjugated, from sheep (GE) was used at 1:20,000 in 10% non-fat milk in PBS, or the Amersham ECL α-rabbit
IgG, HRP-conjugated, from donkey (GE) was used at 1:60,000 in 10% non-fat milk in PBS. The membrane was then washed two to four times in PBS with 1% Tween-20. HRP activity was assessed using the Amersham ECL Prime Western Blotting Detection Reagent (GE) and autoradiography. At least three independent biological replicates were performed for each sample before one representative gel was selected for the immunoblotting result.

**Growth Curve.** *H. volcanii* liquid cultures were inoculated from colonies into 5 ml liquid CA media supplemented with tryptophan (50 µg ml\(^{-1}\) final concentration), with continuous shaking at 45°C until mid-log phase (OD\(_{600}\) ~0.4-0.5). Subsequently, approximately 6 µl of each culture (adjusted slightly for OD\(_{600}\) differences) were transferred into a well of a sterile 96-well plate containing 194 µl of fresh liquid medium (eight replicates of two biological replicates). Cultures were grown to stationary phase with continuous shaking, with OD\(_{600}\) recordings taken every 30 min using the Biotek Power Wave X2 microplate spectrophotometer as described previously (Abdul Halim, Pfeiffer et al. 2013).

**Light microscopy.** One ml of a mid-log phase culture (OD\(_{600}\) 0.4 - 0.5) was concentrated by centrifugation at 4,911 \(\times\) g for 1 min and pellets were resuspended in 10 µl of liquid CA media. 1 µl of the concentrated cells was transferred to a microscope slide and observed using the Eclipse TE2000-U inverted epifluorescence microscope (Nikon USA, Inc.) light microscope as described previously (Abdul Halim, Pfeiffer et al. 2013).
Adhesion assay. *H. volcanii* surface adhesion assay was carried out based on the modified air-liquid interface (Ali) assay as described previously (Esquivel, Xu et al. 2013). Three ml of an *H. volcanii* mid-log phase culture (OD$_{600}$ 0.4-0.5) was transferred into each well of a sterile 12-well plate. Plastic coverslips (22 × 22 mm; 0.19 - 0.25 mm thick) were vertically inserted into each well, followed by overnight incubation at 45°C without shaking. The next day, the coverslips were gently retrieved from the well and the adhered cells were fixed by incubating the coverslips in 2 % acetic acid for 3 minutes. Subsequently, the coverslips were stained in 0.1 % w/v crystal violet solution for 10 minutes. After washing the excess stain with distilled water and air-drying, the coverslips were examined with light microscopy.

LC-MS analysis. Proteins from membrane fractions were solubilized with 2% SDS in 10 mM Tris-HCl pH 7.6. For all samples the protein concentration was determined by bicinchoninic acid assay (BCA Protein Assay Kit by Thermo Scientific Pierce) before loading 50 µg protein onto Amicon ultra centrifugal filters (0.5 mL, 30 kDa MWCO, Millipore) for tryptic digestion using the filter-aided sample preparation method (Wisniewski, Zougman et al. 2009) as described previously (Esquivel, Schulze et al. 2016). Peptides were dried in a vacuum centrifuge and stored at -20 °C. Two biological replicates were processed and analyzed by LC-MS.

After reconstitution of the peptides in 2% (v/v) acetonitrile, 0.1% (v/v) formic acid in ultrapure water, chromatographic separation was performed on an Ultimate 3000 RSLCnano System (Thermo Scientific). The sample was desalted on a trap column (C18
PepMap 100, 300 µm x 5 mm, 5 mm particle size, 100 Å pore size) (Thermo Scientific) for 5 min using 0.05% (v/v) TFA, 2% (v/v) acetonitrile in ultrapure water with a flow rate of 10 µl min⁻¹. The trap column was then switched in line with the separation column (Acclaim PepMap100 C18, 75 mm x 15 cm., 2 mm particle size, 100 Å pore size) (Thermo Scientific). For chromatographic separation the mobile phases were composed of 0.1% (v/v) formic acid in ultrapure water (A) and 80% acetonitrile, 0.08% formic acid in ultrapure water (B) and the following gradient was employed using a flow rate of 300 nl min⁻¹: 5 min 2.5% B, increase to 7.5% B over 4 min, increase to 40% B over 26 min, increase to 99% B over 1 min, 10 min 99% B. The column was re-equilibrated with 2.5% B for 25 min.

The LC system was coupled via a nanospray source to a Q Exactive Plus mass spectrometer (Thermo Scientific) operating in positive ion mode. MS data were acquired from 375 m/z to 2000 m/z at a resolution of 70,000 for MS1. AGC target was set to 10⁶ and maximum injection time to 100 ms. The 12 most abundant precursor ions were triggered for fragmentation by higher-energy C-trap dissociation with a normalized collision energy of 27. MS2 spectra were acquired in a dynamic scan range with a fixed first mass of 150 m/z at a resolution of 17,500 with an AGC target of 10⁵ and a maximum injection time of 120 ms. Precursor ions with an unassigned charge state as well as charge 1 or greater than 5 were rejected. Fragmented ions were dynamically excluded for 15 s.

Peptide spectrum matches (PSMs) were identified with the Python framework Ursgal (v 0.4.0) (Kremer, Leufken et al. 2016) using the database search engines
X!Tandem (Craig and Beavis 2003), MS-GF+ (Kim and Pevzner 2014) and OMSSA (Geer, Markey et al. 2004). The database consisted of the UniProt H. volcanii proteome (proteome ID UP000008243, download date June 10, 2016) merged with sequences from the Common Repository of Adventitious Proteins (http://www.thegpm.org/crap/) resulting in a total database of 4,103 proteins. Decoy sequences were generated by peptide shuffling and included for all proteins. The Ursgal profile ‘QExactive+’ was used including a precursor ion tolerance of 5 ppm and a fragment ion tolerance of 20 ppm. Additionally, carbamidomethylation of cysteine was set as a fixed modification and oxidation of methionine as well as N-terminal acetylation as potential modifications. Three missed cleavages and semi-enzymatic peptides were allowed. Statistical post-processing of unified results was performed with Percolator (Kall, Canterbury et al. 2007). Since the error probability for each PSM should be controlled in order to validate the presence of certain peptides, filtering by posterior error probability (PEP ≤ 0.01) was chosen instead of a global false discovery rate. Results from all search engines for both replicates were combined.

4.4 Results

i. Sequence analysis of Hvo_0405.

In previous work, the H. volcanii genome was predicted to encode eight ArtA substrates (Haft, Payne et al. 2012). Subsequent analysis has revealed an additional H. volcanii polypeptide, Hvo_0405, containing a putative ArtA recognition motif that is not located at the C-terminus, as it is in other predicted substrates. Instead, the PGF motif,
followed by a hydrophobic stretch of amino acids (aa) and a positively charged region, is located 171 aa N-terminal to the stop codon of Hvo_0405 (Figure 1A).

The Hvo_0405 region N-terminal to the PGF motif contains a predicted Tat signal peptide and contains an LVIVD repeat domain (Figure 1A), which is predicted to form a four β-stranded β-propeller domain, which is often found in archaeal and bacterial proteins that localize to the cell surface (Adindla, Inampudi et al. 2007, Marchler-Bauer, Derbyshire et al. 2015). The Hvo_0405 region C-terminal to the PGF motif contains a conserved NifU-like domain; originally discovered in proteins involved in nitrogen fixation, proteins containing an NifU domain are found in many species and typically bind a 4Fe-4S cluster (Figure 1A) (Ouzounis, Bork et al. 1994, Py, Gerez et al. 2012, Marchler-Bauer, Derbyshire et al. 2015).

To determine whether the non-canonical ArtA substrate structure found in Hvo_0405 is conserved in other *Halofex* spp., we performed BLAST searches for homologs and created a protein alignment using all *Halofex* spp. sequenced to date. Surprisingly, *H. volcanii* Hvo_0405 demonstrated sequence homology to two polypeptides in other *Halofex* species, including an LVIVD repeat domain- and a NifU domain containing protein (Figure 1A and data not shown). The predicted protein containing the LVIVD repeat domain also contains an N-terminal predicted Tat signal peptide, in addition to the conserved ArtA recognition tripartite structure at the C-terminus. Therefore, the most parsimonious explanation for the ArtA recognition motif not being localized to the C-terminus in Hvo_0405 is that at some point during its evolutionary history a mutation in
the *H. volcanii* genome resulted in the fusion of two polypeptides that had previously been separately translated, with the polypeptide that became the N-terminal portion of the nascent fusion containing an ArtA recognition motif.

Since a fusion of two genes to form a gene encoding a single polypeptide requires a mutation or genome rearrangement, we sought to determine the precise genomic changes in *H. volcanii* that resulted in the fusion of Hvo_0405. We performed nucleotide alignments of the corresponding genomic regions for each of the eleven *Haloferax* spp. (Figure 1B and data not shown). We found that the nucleotide sequences are highly conserved between *H. volcanii* and the other *Haloferax* spp., but discovered a single base pair deletion in the *H. volcanii hvo_0405* region encoding the charged domain of the tripartite structure. The deletion of this single cytosine base causes a frameshift that leaves the stop codon out of frame and an intergenic region that is translated without interruption, while preserving the frame of the downstream sequence encoding the protein containing the NifU-domain, thereby resulting in a single *H. volcanii* protein containing both the LVIVD repeat domain and NifU domain. Interestingly, we also found that in *Haloferax mediterranei*, the otherwise conserved PGF motif is mutated to TGF. While we previously showed that a PFG motif is required for ArtA-dependent processing, this finding raises the question of whether a motif having a threonine rather than a proline at position +1, can be recognized by ArtA, potentially expanding the range of predicted ArtA substrates significantly (Haft, Payne et al. 2012, Abdul Halim, Karch et al. 2015).

*ii. Construction and characterization of an H. volcanii Δhvo_0405 strain.*
To determine whether this fusion protein is functional and what role it might play in the cell, we created a Δhvo_0405 strain and sought to characterize its defects. The Δhvo_0405 strain (Figure S1) does not exhibit a discernible growth defect in comparison to the H53 background strain in a semi-defined CA medium (Figure S2A). Additionally, when grown in CA liquid culture, inoculated either with a colony or after liquid-to-liquid transfers, the cells of the Δhvo_0405 strain display morphologies similar to those of the parental strain (Figure S2B). Moreover, the deletion of hvo_0405, which is predicted to encode a cell surface protein containing a low-complexity LVIVD repeat domain, does not result in any discernible surface adhesion phenotype under standard conditions (Figure S2B). To date, we have been unable to reliably ascribe any phenotype to the loss of hvo_0405.


Previous in silico studies have described only putative archaeosortase and exosortase substrates containing a tripartite structure located at the C-terminus (Haft, Paulsen et al. 2006, Haft, Payne et al. 2012). Since the Hvo_0405 tripartite structure is atypically located, we asked whether H. volcanii ArtA has as a strict requirement that the tripartite structure of a substrate be located at the C-terminus. To investigate this possibility, we constructed a plasmid containing the hvo_0405 coding sequence fused to a C-terminal polyhistidine tag under the control of a tryptophan-inducible promoter. Additionally, we created an anti-Hvo_0405-N-term antibody that only detects the N-terminal portion of Hvo_0405, which contains the LVIVD repeat domain. When we
induced expression of Hvo_0405 from a plasmid in cells containing ArtA, we observed both the full length Hvo_0405 polypeptide, consisting of both an N-terminus containing an LVIVD-repeat domain and a C-terminus containing an NifU-like domain, and processed forms of Hvo_0405 detected by immunoblot using either α-Hvo_0405-N-term or α-Penta-His antibodies (Figure 2A-B). Strikingly, upon immunoblot analysis of protein isolated from cells lacking ArtA, we only observed a band corresponding to the full-length (not yet processed by ArtA) form of Hvo_0405 (Figure 2A-B), strongly suggesting that Hvo_0405 is processed in an ArtA-dependent manner. To date, we have been unable to detect Hvo_0405 in non-transformed H53 parental strain cells, grown under a variety of conditions (semi-defined versus complex media or growth in low salt, cells grown in biofilms), via immunoblot with the anti-Hvo_0405-N-term antibody Fig. 2A-B and data not shown). This result, which is consistent with mass spectrometric analyses that failed to identify Hvo_0405 peptides in H53 (data not shown) and microarray analysis not identifying significant hvo_0405 transcript (Charles Daniels, unpublished data), may explain the absence of an observable phenotype for the Δhvo_0405 strain.

Canonical ArtA substrates, such as the S-layer glycoprotein, are predicted to localize to the cell membrane. Since our data suggest that Hvo_0405 is an atypical ArtA substrate, we sought to determine the subcellular localization of the processed N- and C-terminal portions of Hvo_0405. To this end, we purified cytoplasmic, membrane, and supernatant fractions from both cells that express ArtA and cells that lack ArtA, each induced to express the Hvo_0405 construct. We found that, although some full-length protein was observed
in membrane fractions, full-length Hvo_0405 predominantly localizes to the cytoplasm, suggesting that this protein is post-translationally inserted into the membrane, consistent with our *in silico* prediction of it being a Tat substrate (Figure 1A and Figure 3A-B). Again, only full-length Hvo_0405 was observed in ArtA-deficient cells (Figure 3A-B), supporting our hypothesis that Hvo_0405 is processed in an ArtA-dependent manner. The processed N-terminal fragment containing the LVIVD repeat domain, which was detected with a signal intensity significantly greater than that of the full-length protein, localizes exclusively to the membrane fraction, suggesting that, like SLG, this ArtA-processed protein is membrane anchored (Figure 3A). Unfortunately, attempts to detect lipid-modified Hvo_0405 were unsuccessful (data not shown). Using mass spectrometric analysis to resolve this matter was not possible due to the limited information available on the lipid modification site. While it is possible that Hvo_0405 is not lipid-anchored, we cannot exclude the possibility that the Hvo_0405 protein concentration was too low to be detected using ¹⁴C-labeled mevalonic acid, which we had previously used to confirm the ArtA-dependent lipidation of the highly abundant SLG (Abdul Halim, Karch et al. 2015).

Interestingly, the processed C-terminal fragment containing the NifU-like domain is also exclusively localized to the membrane, but, consistent with results from whole-cell fractions, it is found in much lower abundance than the full-length form, possibly due to the degradation of this C-terminal region (Figure 3B). While membrane association of the C-terminal region is likely due to the N-terminal transmembrane segment remaining after ArtA processing, in *Haloferax* species in which this NifU domain-containing protein is not
fused to an ArtA substrate, it is likely a cytoplasmic protein. Therefore, this membrane-associated protein may be unstable because it is not adapted to function as a membrane protein.

*iv. Hvo_0405 is a Tat substrate.*

As noted above, identification of full-length Hvo_0405 in the cytoplasm could be explained by the fact that, once translated, this protein, as predicted, is transported via the Tat pathway. We then sought to experimentally determine whether Hvo_0405 is indeed a Tat substrate. Since both Sec and Tat transport pathways are essential in *H. volcanii*, deleting the genes involved in these pathways to determine which transport pathway is used by Hvo_0405 is not feasible (Dilks, Giménez et al. 2005). Therefore, we generated a plasmid expressing a gene, encoding an Hvo_0405 construct, in which the nucleotides encoding the highly conserved two arginine residues at the N-terminus of Hvo_0405 (Figure 1A), predicted to be crucial for Tat substrate recognition, were mutated to encode two lysine residues (Hvo_0405KK). We then expressed this tryptophan-inducible Hvo_0405KK mutant in *trans* in the Δhvo_0405 strain to determine whether the mutated protein can still be transported properly. By immunoblot analysis of proteins isolated from cell membrane and cytoplasm fractions, we found that the twin arginine residues are required for Hvo_0405 processing, as the processed forms of Hvo_0405 in the membrane fractions were only observed in strains expressing Hvo_0405 that contains the twin arginine residues, while the strain expressing the Hvo_0405KK mutant contained only full-length protein (Figure 4A-B).
Interestingly, we found that a fraction of the unprocessed Hvo_0405KK continued to be localized to the membrane (Figure 4A-B). While it is possible that the presence of this small fraction of unprocessed Hvo_0405KK is due to contamination from the cytoplasm, this would further support our hypothesis that Hvo_0405 is indeed a Tat substrate as the Hvo_0405KK is no longer targeted to the membrane. However, we hypothesize that the likely reason for the Hvo_0405KK membrane localization is that the Sec machinery recognizes the hydrophobic region in the tripartite structure of the Hvo_0405KK mutant and targets the protein to the cytoplasmic membrane. However, we further suggest that, due to its, at least partially folded state, the Hvo_0405KK mutant protein would be unable to properly translocate across the cytoplasmic membrane through the Sec pore, resulting in the full-length Hvo_0405KK mutant protein localizing to the membrane but preventing ArtA-dependent processing. In this scenario, it would be unlikely that the evolutionarily conserved SPaseI, the peptidase that processes Sec and Tat substrate signal peptides, would be able to process Hvo_0405KK. To test this hypothesis, we performed mass spectrometric analysis of trypsin-digested proteins isolated from the membrane and cytoplasmic fractions of the Δhvo_0405 strain expressing either the wild-type Hvo_0405RR or the mutated Hvo_0405KK. While multiple tryptic peptides were detected in the trypsin-treated protein isolated from membrane fractions of strains expressing either Hvo_0405RR or Hvo_0405KK, a semi-tryptic peptide, HPGFPEPLGR, in protein from the wild-type strain expressing Hvo_0405RR (Table 2). This semi-tryptic peptide likely represents the most N-terminal peptide of the mature Hvo_0405 protein and
its N-terminal sequence indicates the position of the SPaseI cleavage site, which is consistent with the predicted signal peptide cleavage site identified by SignalP (Petersen, Brunak et al. 2011). In contrast, while we did not detect any semi-tryptic fragments in the membrane fractions from the mutated strain expressing Hvo_0405KK, a tryptic peptide including the SPaseI processing site, PATAHPGFEPILGR, was identified in both replicates, suggesting that only the unprocessed signal peptide was present (Table 2). It should be noted that, despite the generally accepted “Keil rules” (Keil 1992), tryptic cleavage before proline is possible (Rodriguez, Gupta et al. 2008). These results provide additional support for Hvo_0405 as a bona fide Tat substrate, as well as providing one of only a handful of in vivo confirmed SPaseI processing sites for archaeal or bacterial Tat substrates.

A second semi-tryptic peptide, SDESESLKER, was identified exclusively in protein isolated from the cytoplasmic fractions of strains expressing either Hvo_0405RR and Hvo_0405KK (Table 2). Interestingly, in all other Haloferax spp, this semi-tryptic fragment corresponds to the penultimate position of the region containing the start codon of the protein containing NifU-domain (Figure 1). The removal of the N-terminal methionine is consistent with methionine aminopeptidase processing observed in many prokaryotic species (Falb, Aivaliotis et al. 2006). This observation supports our hypothesis that the DNA lying between the regions encoding the Hvo_0405 N-terminal region, which contains the LVIVD repeat domain, and C-terminal region containing NifU-like domain,
contains an internal promoter that continues to drive expression of the protein containing the NifU-like domain.

Using TatFind and TatP programs, we also determined whether any of the other putative *H. volcanii* ArtA substrates are predicted to be Tat substrates (Dilks, Giménez et al. 2005, Giménez, Dilks et al. 2007). Despite containing twin arginines in the charged region of the signal peptides of Hvo_2160 and Hvo_B0206, neither of these two predicted ArtA substrates, were predicted by TatFind to be Tat substrates. Alternatively, sequence analysis with TatP predicted that Hvo_2160 is indeed carrying a Tat signal peptide. The discrepancies between these Tat substrate prediction programs indicate that it is imperative to carry out additional *in vivo* studies like the one presented here to obtain a larger training set for optimizing the subcellular localization prediction programs.

**4.5 Concluding Remarks**

In this work, we demonstrate that the predicted ArtA substrate Hvo_0405 is likely a fusion protein of an ArtA substrate and a cytoplasmic protein, resulting in a tripartite structure no longer located at the C-terminus. Despite the atypical localization of this putative ArtA recognition motif, we demonstrate that Hvo_0405 is indeed processed in an ArtA-dependent manner. Interestingly, there are fewer positively charged residues following the PGF motif and hydrophobic stretch in Hvo_0405 than in other *Haloferax* spp. (one versus five to six, respectively, Figure 1). Since this paucity of positively charged residues apparently does not interfere with ArtA processing, we can conclude that either this is not
a requirement for ArtA substrates, consistent with the putative ArtA substrate Hvo_1095 that also contains only a single positively charged residue at the C-terminus, or that the role of these residues in anchoring the C-terminus to the plasma membrane is subsumed by the large C-terminal NifU-containing domain of Hvo_0405. Together with our observation that the conserved PGF motif is TGF in the *H. mediterranei* homolog of LVIVD domain-containing N-terminal region of Hvo_0405, these finding suggest that ArtA has a broader substrate repertoire than previously thought (Figure 1).

Our data also suggest additional mechanisms by which proteins can become associated with the cell membrane. The LVIVD domain-containing N-terminal region of Hvo_0405 is likely membrane anchored via lipidation, similar to other ArtA substrates. It is intriguing that the NifU domain-containing C-terminal region is also membrane associated, likely due to the hydrophobic region that was originally part of the tripartite structure but now remains attached to the N-terminus after ArtA-dependent processing. This is likely a novel cellular location for this NifU domain containing protein, as it is almost certainly a cytoplasmic protein in other *Halofex* spp., consistent with the identification of the second semi-tryptic peptide, exclusively in the cytoplasm (Table 2).

Using site-directed mutagenesis, we demonstrate that Hvo_0405 is a Tat substrate and that its processing is dependent upon the Tat pathway by both immunoblotting and mass spectrometry. Additionally, our mass spectrometry data confirms the SPaseI processing site *in vivo*. Prediction of SPI processing sites in Tat substrates has been challenging due to a paucity of confirmed SPI processing sites, evidenced by the failure of
TatP to accurately predict the Hvo_0405 processing site. Studies like these will almost certainly provide better training sets for prediction programs, not only for accurate prediction of Sec and Tat substrates in halophilic archaea, but also in the prediction of ArtA substrates in archaea and exosortase substrates in gram-negative bacteria.
4.6 Figures

Figure 4.1: The conserved SLG PGF-motif is critical for ArtA-dependent processing. The SLG C-terminal region (AA: 766-827) consists of a threonine-rich stretch followed by the conserved PGF motif (bold), a hydrophobic stretch (underline) and positively-charged residues (*end of protein) (A). XICs of gel-purified SLG and SLG<sup>G796F, F797G</sup> peptides from AA: 35-48 (B); or AA: 813-824 (C). The masses and normalized abundances (NA) of the chromatographic peak are indicated. Graphical representation of the relative abundance of the N-terminal peptide (AA 35-48) and C-terminal peptide (AA813-824) in WT SLG (purple) and SLG<sup>G796F, F797G</sup> (blue) cells (D). MS/MS fragment ions from C-terminal peptide AA 813-824 ([M+2H]<sup>2+</sup> = 583.885 m/z) (E). The lines indicate which fragment ions were
detected (b ions above the sequence; y ions below). Identification of a large number of the expected fragment ions, lends strong support that the peptide was correctly identified.

Figure 4.2: ArtA and conserved C-terminal PGF motif are required for *H. volcanii* SLG lipid-modification. Fluorography of protein extracts isolated from wild-type FH37 (wt), ΔartA, and *slg*<sup>G796F, F797G</sup> cells grown in the presence of 1µCi/ml <sup>14</sup>C mevalonic acid (left). Labeled SLG (arrow) is detected only in the wt extract. Additional labeled lower molecular weight proteins likely are ArtA-independent secreted lipobox-containing proteins (star) (Storf, Pfeiffer et al. 2010). Coomassie stain of protein extracts from wt, ΔartA, and *slg*<sup>G796F, F797G</sup> cells (right). The migration of molecular mass standards is indicated on the left (in kDa).
Figure 4.3: The ΔartA and slg$^{G796F,F797G}$ exhibit distinct cell morphology compared to the wild-type. Phase contrast microscopy of wild-type FH37 (wt), ΔartA, and slg$^{G796F,F797G}$ without or with pTA963 grown to mid-log phase in liquid semi-defined CA media supplemented with tryptophan and uracil or with tryptophan only, respectively. Liquid media were inoculated with cells from colonies grown on solid CA agar plates (top row), or cells that had been grown by serial transfer in liquid media to mid-log phase (bottom row). Size bar: 10 μm.
**Figure 4.4 SLG$^{G796F, F797G}$ forms thicker S-layer compared to wild-type.** Wild-type FH37 (wt), ΔartA, and slg$^{G796F, F797G}$ cells transformed with pTA963 were preserved with high pressure freezing techniques followed by thin slice section microscopy. The S-layer thickness was determined by measuring the region between the cell membrane (white arrowhead) and the outer edge of the S-layer (black arrow). The slg$^{G796F, F797G}$ cells contain wider S-layer (16 nm) compared to the wt (12 nm) and contains material around the cells. While displaying similar S-layer width to the wt, the ΔartA cells generally contained significant amounts of material surrounding the cell compared to the other strains. Size bar: 100 nm.
Table 4.1: Peptide identifications in strains expressing Hvo_0405RR or Hvo_0405KK

<table>
<thead>
<tr>
<th>Peptide Sequence</th>
<th>Semi-tryptic</th>
<th>Membrane</th>
<th>Cytosol</th>
<th>Membrane</th>
<th>Cytosol</th>
</tr>
</thead>
<tbody>
<tr>
<td>PATAHPGPFEPLGR</td>
<td>no</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>HPGPFEPGLGR</td>
<td>yes</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>EAVVSADGETVFVAATSGYAVVIDAPER</td>
<td>no</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>QELLAER</td>
<td>no</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RDPLSDR</td>
<td>no</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RDPLSDREDGPFR</td>
<td>no</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>DPLSDREDGPFR</td>
<td>no</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>AYLTAINDGRNPVLVDIDSGDELGR</td>
<td>no</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SVHDVWWHDR</td>
<td>no</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>VAHVALWDAGTWLVDLSDPASPWVLGA</td>
<td>no</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>VSPGDPEIEAALTPR</td>
<td>no</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>STIEPPASPDPTLGGVWTTAHNFDFR</td>
<td>no</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>STIEPPASPDPTLGGVWTTAHNFDFRDR</td>
<td>no</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>STIEPPASPDPTLGGVWTTAHNFDFRDR</td>
<td>no</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>HDVSDPTDPVELAWWR</td>
<td>no</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>HDVSDPTDPVELAWWRDPER</td>
<td>no</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ASFWTAYAQYPFADEGVFVASSR</td>
<td>no</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GVGDASPALYTFPDHAGDQR</td>
<td>no</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>GVGDASPALYTFPDHAGDQRDPPTLR</td>
<td>no</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>TTEIAFLSPAGASGEPFHSPWGSK</td>
<td>no</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SMSDESESLK</td>
<td>no</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>SMSDESESLKER</td>
<td>no</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SDESESLKER</td>
<td>yes</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>VVPSPLNSR</td>
<td>no</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>VPSSGDHGSSTVEGGR</td>
<td>no</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
Figure 4.S1: Confirmation of *hvo_0405* deletion mutant by PCR.

(A) Schematic representation of genomic location of *hvo_0405* and primers used to screen for the ∆*hvo_0405* strain. (B) PCR was performed with genomic DNA templates isolated from the H53 and ∆*hvo_0405* strains, using flanking primers FW_0405_XhoI (OP1) and RV_0405_Xbal (OP2). The amplicon is approximately 2,000 nucleotides shorter for the ∆*hvo_0405* strain than for the H53 strain, consistent with an *hvo_0405* gene deletion. The PCR performed to amplify the *hvo_0405* gene, using the FW0405NdeI (IP 1) and RV0405EcoRIHIS (IP 2) primers, confirmed that the *hvo_0405* gene is present in the H53 strain but absent in the ∆*hvo_0405* strain.
CHAPTER 5

Conclusions and Future Directions

The characterization of ArtA and its substrates, presented in this thesis has provided valuable data to initiate the elucidation of a novel cell surface protein anchoring mechanism in prokaryotes. The findings from my studies will lead to a better understanding many crucial archaeal biological processes, since as the intermediary between the cells and its environment, proper processing and anchoring of surface proteins is important as these proteins mediate various cellular processes such as motility, adhesion, signal transduction, biofilm formation, and maintaining cellular integrity. This work will also likely lead to valuable insights into a homologous protein anchoring mechanism in many gram-negative bacterial species that encode archaeosortase homologs, exosortases, which have yet been characterized in vivo. Key bacterial species containing exosortases have applications in sustainable industry and medicine, including microbial electrogenesis and healthy human microbiome, respectively (Haft, Paulsen et al. 2006, Haft, Payne et al. 2012). Finally, understanding the mechanism that mediates the lipid-anchoring of these substrates, facilitated by the archaeosortase/exosortase, could revolutionize the pharmaceutical industry, particular since archaeal SLG is already being used to stabilize liposomes as a novel drug delivery system, thus our findings in SLG lipid anchoring may lead to an
improved liposome stability for a better drug delivery efficiency (Ucisik, Küpcü et al. 2013).

In the 4 chapters presented in this work, I describe in vivo characterization of the ArtA-dependent surface protein anchoring to the archaeal cytoplasmic membrane. I have started dissecting the critical components as well as the molecular mechanism of this anchoring system. First, I showed that the ArtA-dependent processing and anchoring of its substrates is critical for various cellular processes as the absence of ArtA leads to H. volcanii cells exhibiting various phenotypic defects, including a reduced growth rate, impaired swimming motility, abnormal cellular morphology, and inefficient mating.

The in silico prediction that a subset of H. volcanii surface proteins are processed in an ArtA-dependent manner was confirmed by showing that at least two of its predicted substrates, the SLG and Hvo_0405, are indeed only C-terminally processed in the presence of ArtA. The discovery of SLG as a bona fide ArtA substrate not only provided the first evidence of archaeal C-terminally processed protein, but also introduce a new paradigm of the SLG anchoring mechanism. In contrast to the previously long-held belief, H. volcanii SLG is not lipid-anchored by a TM domain but rather is membrane-anchored via a covalent lipid linkage at the C-terminus, a posttranslational modification that I showed to be dependent on ArtA. The C-terminal processing and lipid anchoring of ArtA substrates might be a cellular strategy to minimize the risk of overcrowding the cytoplasmic membrane space, especially since prokaryotes have limited cytoplasmic membrane space. As some of the surface proteins are anchored to the cytoplasmic membrane by a
transmembrane domain, this type of anchoring can easily lead to the overcrowding of the cytoplasmic membrane space. By lipid-anchoring a subset of its surface proteins on the cytoplasmic membrane, this will free up the lipid bilayer, which might be better used by for other protein complexes that facilitate various crucial cellular processes, such as nutrient uptake, protein transport, osmoregulation, or energy production.

ArtA-dependent C-terminal processing and anchoring of its substrates requires a highly conserved C-terminal PGF motif. I showed that the permutation of the ArtA substrate PGF to PFG motif is sufficient to block its C-terminal processing, indicating that this motif plays a critical role in its recognition as ArtA substrate. Additionally, the unprocessed ArtA substrate also lacks the lipid modification, suggesting that either the substrate C-terminal processing precede the covalent lipid modification or these posttranslational modifications occur simultaneously. While the sequential steps of these posttranslational modifications have yet to be elucidated, preliminary data support the former hypothesis of C-terminal processing occurring prior to the covalent lipid modification (see appendix chapter).

We also have shown that the PGF motif does not needs to be located at the C-terminus of the protein. The bona fide ArtA substrate, Hvo_0405, encodes its tripartite structure closer to the middle of the protein. Additionally, while the original prediction of ArtA substrates was limited to those secreted via the Sec transport pathway, analysis of Hvo_0405 ArtA-dependent processing also reveal that the ArtA processing can involve proteins that are translocated across the cytoplasmic membrane via the Tat transport
pathway. It is worth noting that in both cases, the ArtA deletion and PGF mutation of ArtA substrates does not affect their SPaseI signal peptide processing, indicating that the proteins are fully translocated across the cytoplasmic membrane prior to the C-terminal processing. Hence, based on the genetic and biochemical evidence gathered thus far, we proposed the molecular mechanism of ArtA-dependent surface protein anchoring as follows: once translocated across the cytoplasmic membrane, the ArtA substrate is temporarily TM anchored at the C-terminus via the TM domain of the tripartite architecture, allowing proper orientation of the PGF motif to be recognized by the ArtA. The ArtA predicted catalytic site are positioned to cleave the substrate, which may already have undergone several post-translational modifications, such as O-glycosylation at a threonine-rich region present immediately N-terminal to the PGF-motif (Table 1.1). Analogous to the sortase transpeptidation reaction, the proposed molecular mechanism of ArtA substrates C-terminal cleavage is based on the arrangement of the amino acid residues cysteine (Cys), arginine (Arg) and histidine (His) that made up the putative ArtA catalytic triad (Ton-That, Mazmanian et al. 2002, Haft, Payne et al. 2012). Upon ArtA substrate C-terminal cleavage, its N-terminus domain is transiently attached to the conserved Cys of the ArtA catalytic triad, followed by transfer onto a diphytanylglyceryl phosphate moiety, anchoring the protein to the cytoplasmic membrane via a covalent lipid linkage.

The forthcoming studies should focus on confirming the role of this catalytic triad in being required for ArtA processing activity. Intriguingly, at least two of the predicted triad residues, cysteine and arginine, resides near or within the ends of transmembrane
helices, with the hydrophobic residues flanking the cysteine are highly conserved among the Haloferax spp. ArtA. Generation of ArtA antibodies, as well as Western blot detection of ArtA-his, has been unsuccessful. However, an artA-GFP expressed in trans expression in a ΔartA strain, can complement the phenotypes associated with the chromosomal deletion of artA. Hence, overexpression of similar constructs in which one of the predicted catalytic triad amino acids is replaced with an alanine should allow us to determine whether these amino acids are indeed critical for ArtA function, provided they are stable. If the phenotypic defects of a ΔartA strain are not complemented by the expression of a replacement mutant, or substrates, such as Hvo_0405, are not processed, these residues are likely required for ArtA catalytic activity. Ultimately, to show that ArtA is directly involved in processing the substrate and test the importance of the catalytic residues, an in vitro assay will be needed to test the cleavage activity of this enzyme. An analogous approach to the previously reported sortase in vitro processing can be carried out (Ton-That, Mazmanian et al. 2000). Briefly, a short peptide bearing the PGF-Cterm motif flanked by a fluorophore and quencher can be synthesized and used as a reporter substrate. The increase in its fluorescent activity is dependent upon the substrate processing as the quencher molecule is released away from the fluorophore. The in vitro assay would involve incubating the PGF-Cterm peptide with the purified ArtA, and a control setup in which the ArtA is left out from the mix. If the ArtA is directly involved in the substrate processing, increase in the peptide fluorescent activity should be observed in an ArtA-dependent manner.
Furthermore, the expression of an ArtA replacement mutant will provide a glimpse at the molecular mechanism of the ArtA substrates lipid modification. By testing the SLG lipidation in the presence of the ArtA replacement mutant, it may reveal if different residue within the catalytic triad is important for the substrate processing or for the lipid modification. Additionally, while we showed that SLG lipidation is ArtA-dependent and our preliminary data suggests another enzyme, PssA, is also involved in this posttranslational modification, it is still unclear which protein is directly responsible for adding the lipid to the C-terminus of processed substrates. The PssA is known for its catalytic activity in transferring the lipid moiety to a serine residue, of which two serine residues are identified near the SLG PGF-Cterm motif. By testing the lipidation of the SLG serine replacement mutants, it would not only allow the identification of the target residue for SLG lipid modification, but also determine the direct involvement of PssA in ArtA substrates lipid modification.

The ArtA substrates C-terminal PGF motif is also located directly adjacent to a stretch of amino acids rich in threonines, which may be O-glycosylated as suggested by the mass spectrometry analysis, which components involved remain elusive in archaea. This suggest that the ArtA processing and O-glycosylation of these substrates might be linked. Thus, the development of protein co-purification methods for ArtA affinity chromatography may identify other uncharacterized proteins, such as components of archaeal O-glycosylation.
While the work presented here shows that *H. volcanii* SLG and Hvo_0405 is processed in an ArtA-dependent manner, we have not yet defined the ArtA processing site within the substrate PGF-CTERM motif. The attempt to sequence the C-terminus of processed SLG has not been successful, possibly due to this region being post-translational lipid-modified and possibly O-glycosylated (Sumper, Berg et al. 1990). However, as noted in the appendix chapter section, Hvo_0405 can be expressed and processed in the background strain with the archaetidylserine synthase (*pssA*) gene deleted. Hence, mass spectrometry of the processed non-lipidated Hvo_0405 N-terminal peptide isolated from the supernatant fraction may provide us with the identity of the processing site.

To date, with the exception of SLG, the functions of predicted *H. volcanii* ArtA substrates remain undetermined. Consistent with the ΔartA strains exhibiting various severe phenotypes, one or more of the ArtA substrates are likely critical for the cellular processes. The deletion of Hvo_0405 result in the cells exhibiting swimming motility defect. As described in Chapter 4, Hvo_0405 is a natural fusion protein, which in the other closely related *Halofex* spp., is expressed as two independent proteins. The Hvo_0405 N-terminal fragment contains the LVIVD-repeat domain found in many prokaryotic cell surface proteins, while its C-terminal fragment contains a NifU-like domain found in nitrogen-fixing bacteria and rhodobacterial species. Separate expression of each Hvo_0405 protein fragment in Δhvo_0405 may determine which fragment plays a critical role for the *H. volcanii* motility. Preliminary data also demonstrated that deletion mutant lacking another predicted ArtA substrate, Hvo_2160, exhibits a severe mating defect. This protein
contains domains reminiscent of the cell attachment domains in bacteria, which may be involved in mating (Table 2.1). Since mechanism that mediate the archaeanal mating process has yet to be elucidated, this would provide an opportunity to begin dissecting this cellular process. (Mevarech and Werczberger 1985, Rosenshine, Tchelet et al. 1989).

As mentioned previously, the characterization of ArtA also serves as foundation to understand its bacterial homolog, exosortase, which is present in a diverse group of gram-negative bacteria, including many species that play important roles in biotechnology and bioremediation (Haft, Paulsen et al. 2006, Haft, Payne et al. 2012). Mirroring a similar correlation between ArtA and its substrates, the exosortase gene, xrtA, is found only in bacterial genomes that also encode proteins containing a conserved C-terminal tripartite architecture, albeit carrying a different motif, PEP-Cterm, consisting of a highly conserved C-terminal Proline (P), Glutamate (E), and Proline (P) residues, suggesting that the proteins containing a PEP-Cterm domain are XrtA substrates (Haft, Paulsen et al. 2006). The genetically-amenable Geobacter sulfurreducens, one of the bacterial species that encodes XrtA, can be used as model organism to elucidate this anchoring mechanism as the biochemical and molecular biological tools are readily available for this organism. This anaerobe plays invaluable and versatile roles in a variety of technological applications, ranging from groundwater bioremediation to microbial fuel cell electrogensis (Rooney-Varga, Anderson et al. 1999, Coppi, Leang et al. 2001, Holmes, Finneran et al. 2002, Bond and Lovley 2003, Methe, Nelson et al. 2003). As these processes are enhanced in a biofilm setting, and a previous study suggested that one of the putative XrtA substrates, GSU1994,
is important for effective biofilm formation on fuel cell electrodes, elucidating the exosortase-mediated anchoring pathway may have implications that go beyond simply furthering our basic understanding of the cell biology of this organism (Franks, Nevin et al. 2010). Subcellular localization studies of XrtA substrates in this gram-negative bacterium should also be intriguing. Similar to \textit{H. volcanii}, the substrates may be anchored to the inner membrane. However, they might also be released into the periplasmic space, transported across the periplasm and anchored to the inner or outer leaflet of the outer membrane, or released into the external environment. Interestingly, a second gene, conserved in all genomes encoding XrtA and its predicted substrates, encodes a unique periplasmic protein, EpsI, which perhaps is part of a novel molecular machinery that transports Xrt substrates across the periplasm. Hence, elucidating the exo/archaeosortase anchoring mechanism does not only have the potential to confirm that this novel anchoring mechanism is widely conserved across prokaryotes, it also may lay the groundwork for future studies that might reveal an entirely novel protein secretion pathway.
The archaetidylserine transferase role in prenyl lipid modification of the *Haloferax volcanii* S-layer glycoprotein

A1.1 Introduction

*H. volcanii* SLG was long thought to be membrane anchored via its C-terminal TM domain (Sumper, Berg et al. 1990). However, recent findings that SLG undergoes ArtA-dependent C-terminal processing, combined with previous data demonstrating covalent lipid modification of SLG after its membrane translocation (Konrad and Eichler 2002), it is now hypothesized that SLG cytoplasmic membrane anchoring is achieved via the covalent lipid linkage at its C-terminus instead of through the TM domain. However, the biosynthesis machinery and biochemical pathway for this posttranslational modification remain uncharacterized. While *H. volcanii* SLG covalent lipid modification is discovered to be ArtA-dependent, it is still not known whether the ArtA is directly responsible for this posttranslational modification. Interestingly, in silico analysis of several methanogenic archaea, such as *Methanosarcina acetivorans* C2A and *Methanosarcina mazei* Go1, revealed that their *artA* genomic location is adjacent to the gene encoding the putative achaetidylserine synthase (PssA) enzyme (Morii and Koga 2003, Haft, Payne et al. 2012).
The bacterial PssA homolog has been characterized to be involved in the biosynthesis of the cytoplasmic membrane phospholipids, specifically by catalyzing the transfer of diacylglycerol phospholipid moiety onto a serine residue to synthesize the phosphatidylserine (PS) (Moser, Aktas et al. 2014). The *in vitro* analysis of archaeal PssA in methanogenic archaeon *Methanothermobacter thermautotrophicus* has also confirmed its enzymatic function of transferring archaeol lipid onto a free serine residue (Morii and Koga 2003). Sequence analysis of the archaeal PssA homologs and its substrate specificity revealed its homology to the gram-positive bacteria PssA such as encoded in *Bacillus subtilis*, which belong to subclass II pssA compared to *E.coli* subclass I pssA (Matsumoto 1997, Morii and Koga 2003). Studies of the cellular localization of *B. subtilis* PssA revealed that the protein is localized at the cytoplasmic membrane, especially at the septum and this septal localization is dependent on the FtsZ protein, a key component in the cell division (Nishibori, Kusaka et al. 2005). The *H. volcanii* PssA is also predicted as a cytoplasmic membrane protein, which according to the membrane topology prediction program, Phobius, containing multiple TM domains (Käll, Krogh et al. 2004). Hence, I hypothesized that besides its role in the phospholipids biosynthesis, the PssA may work in tandem with the ArtA at the cytoplasmic membrane in transferring the archaeol lipid moiety onto the serine residue at the C-terminus of substrates upon ArtA processing. Additionally, the protein sequence analysis revealed that 7 out of 8 ArtA substrates encode at least one serine residue near the PGF-Cterm motif (Table 2.1). However, despite the ability of PssA in transferring lipid moiety to a free serine residue, its catalytic activity in
transferring the lipid moiety to serine residue within a peptide chain has never been investigated.

A1.2 Result

The ΔpssA strain exhibiting growth defect, impaired motility, and irregular cell morphology. In order to test its potential role in the ArtA substrates lipid modification, *H. volcanii* PssA homolog (Hvo_1143) was identified and the gene encoding the protein is deleted from the chromosome (Figure 1). Preliminary phenotypic analysis of the ΔpssA strain revealed a mild growth defect compared to the wild-type strain (Figure 2) as well as a motility defect (Figure 3). However, both phenotypes can be restored to wild-type levels upon expression of pssA in trans, eliminating the polar effect of downstream genes in contributing to the ΔpssA phenotypes. Since cells lacking ArtA exhibited irregular cell morphology, we also investigated the ΔpssA morphology. Preliminary light microscopy analysis of cultures either inoculated from colonies grown on agar plate or inoculated from serial liquid transfer revealed that the ΔpssA exhibited a distinct cell morphology compared to the wild-type and complementation strains. For the liquid cultures inoculated from colonies on the plate, both the wild-type and pssA complementation strain cells are predominantly disc-shaped cells with few rod-shaped cells (Figure 4). In contrast, the ΔpssA strain cells are predominantly rod-shaped with a few deformed disc-shaped cells. The morphological distinction between the ΔpssA and the wild type strain is more apparent when the extended serial transfer of liquid cultures was examined. While the rod-shaped
cells are no longer present in all cultures examined, the ΔppsA strain exhibited a highly pleomorphic cell shape with deformed clumps compared to the wild-type and complementation strains that have a consistent disc-shaped cells (Figure 4). However, the phenotypic defect exhibited by the ΔpssA strain is not as severe as the ΔartA strain, which no longer able to switch from the rods to disc-shaped cells after extended serial liquid sub-inoculation. Instead, morphological defect exhibited by ΔpssA seems to closely resembles the slg$^{G786F, F797G}$ strain morphological phenotype, either in the culture inoculated from colonies on a plate or the culture from the serial liquid sub-inoculation (Figure 4).

The ΔpssA SLG lipid modification is greatly impaired but not completely abolished. Previously, the improperly anchored SLG isolated from the ΔartA strain exhibited a migration shift upon protein separation on an LDS-PAGE gel. While the exact factor(s) contributing to this phenomenon have yet to be determined, we postulated that perhaps this might be due to lack of additional post-translational modifications, such as substrate lipidation or glycosylation. The cell extracts of the wild-type, ΔartA, slg$^{G786F}$, $^{F797G}$, and ΔpssA strains were separated on the LDS-PAGE gel and the band corresponding to SLGs were examined. Interestingly, the ΔpssA SLG exhibits a similar migration pattern to the ΔartA and slg$^{G786F, F797G}$ SLGs, suggesting that the pssA deletion may interfere with proper SLG anchoring or its posttranslational modification(s) (Figure 5). Based on the hypothesized role of the PssA in transferring a lipid moiety to the ArtA substrate, we determined SLG lipid modification in the ΔpssA strain. The fluorograph of radiolabeled

146
cultures using the archaeal lipid precursor, $^{14}$C mevalonate, indicate that, unlike the wild-type and pssA complementation strain, the SLG only shows weak radiolabeling of a band whose size corresponds to SLG (Figure 6). While the role of PssA in SLG lipidation is still inconclusive, this result suggests that it does play a significant role in this posttranslational process. Perhaps, the ArtA-dependent lipid modification involved additional components, allowing the lipid modification of a minor subset of SLG in the ΔpssA strain.

The pssA deletion does not inhibit the ArtA-dependent C-terminal processing. Since the lack of PssA results in a severe defect in the ArtA substrate lipid modification, we next wanted to see whether the ArtA substrate processing is also compromised. To address this question, we employed Hvo_0405 as a reporter substrate, since its ArtA-dependent processing can be easily determined using western blot analysis as described in Chapter 4. Using the antibody that recognized the N-terminal portion of Hvo_0405, western blot analysis of cell lysate extract of ΔpssA expressing Hvo_0405 initially suggest that the ArtA processing is impaired as only the full length Hvo_0405 unprocessed precursor band were detected (data not shown). However, when the Western blot analysis against the supernatant fraction were carried out with the same antibody, we detected the band consistent with the size of processed N-terminus of Hvo_0405 in supernatant fraction of the ΔpssA culture as oppose to its membrane anchored in the wild-type culture (Figure 7). This result suggests that the ArtA substrate processing is not affected by the pssA deletion. However, it further supports the hypothesis that PssA is required for covalent
lipid anchoring of ArtA substrates to the membrane. We also noted the lower molecular weight band that likely represents the product of degraded Hvo_0405 N-terminus.

A1.3 Discussion

The in silico work by Haft et al., suggested that the ArtA substrates are lipid-anchored to the cytoplasmic membrane upon its C-terminal processing (Haft, Payne et al. 2012). While previously I have shown that the substrate lipid modification is ArtA-dependent, it was proposed that this posttranslational modification is catalyzed by a different protein, archaetidylserine transferase (PssA), which is a homolog of bacterial phosphatidylserine transferase. The phenotypic defects exhibited by the ΔpssA strain are not as severe as the ΔartA strain. However, the cell morphology of ΔpssA is closely resembles the slg<sup>G786F, F797G</sup> strain, suggesting the anchoring SLG may be affected in the cells lacking PssA. However, it is also possible that observed phenotypic defect of the ΔpssA mutant strain might be due to unstable cytoplasmic membrane, which lacking the PS phospholipid contributing its altered lipid composition. Nonetheless, previous lipid analysis of H. volcanii cytoplasmic membrane suggest that PS is not a major component of its membrane (Sprott, Larocque et al. 2003). The H. volcanii membrane lipid analysis is perhaps required to determine if the lack of PssA cause a significant alteration in its lipid composition. However, if that PssA is indeed involved in the ArtA substrates anchoring, it might be limited to only a subset of substrates, such as SLG, as suggested by the ΔpssA phenotypes and the ΔpssA SLG radiolabeling.
While the lipid radiolabeling results suggest that PssA is involved in the SLG lipid modification, we still have yet to show if the PssA is indeed directly involved in this SLG posttranslational modifications. Additionally, since the deletion of \textit{pssA} does not completely abolished the SLG lipid modification as previously reported in the \textit{ΔartA} or \textit{slg}^{G786F, F797G} SLG, additional components may be involved or perhaps PssA only involve in the lipid modification of a subset of SLG. Nonetheless, the \textit{ΔpssA} lipid radiolabeling result reported here is still preliminary, which requires additional replicates to confirmed the consistency of the findings.

On the other hand, the Western blot analysis of Hvo\textunderscore0405 indeed suggest that its membrane anchoring is dependent on the PssA. Since the ArtA processing is not affected in the \textit{ΔpssA} strain, the absence of this protein cause the released of the processed Hvo\textunderscore0405 N-terminus into the extracellular environment. Further mass spec analysis of the SLG may still be needed to confirm that the status of ArtA processing in the \textit{ΔpssA} strain. Additionally, LDS-PAGE migration shift exhibited by the \textit{ΔpssA} SLG seems to suggest that it may lack some post-translational modifications, which hopefully can also be elucidated by the mass spectrometry analysis.
A1.4 Figures

**Figure A.1: Confirmation of the pssA gene deletion.** The genomic region containing the \( pssA \) (A) gene in \( H. volcanii \). Primers used to amplify the flanking region (700 bp upstream and downstream of \( pssA \) (outer primer, OP) and the \( pssA \) coding region (inner primer, IP) are indicated. PCR was performed using genomic DNA templates isolated from the H53 wild-type, and \( \Delta pssA \). For the PCR products obtained with the outer primers OP, the amplicon sizes are approximately 600 nucleotides longer for the parent strain H53 than for the \( \Delta pssA \) strain. Using the inner primers IP, no amplicons were obtained from \( \Delta pssA \) strain.
Figure A.2: Deletion of *pssA* results in a slight growth defect. *H. volcanii* H53 (wt) and Δ*pssA*, transformed with an empty pTA963 plasmid, as well as Δ*pssA* complementation were inoculated in liquid semi-defined media CA supplemented with tryptophan (final concentration 50 µg/ml).
The *H. volcanii* H53 (wt) and Δ*pssA* transformed with an empty pTA963 plasmid, as well as Δ*pssA* expressing *pssA* from pTA963 (Δ*pssA + pssA*), were stab-inoculated on motility plate (0.3% agar). While H53 and Δ*pssA* complementation strain produced comparable growth halo size after 3 to 5 days, only a small growth halo was observed for the Δ*pssA* strain during the same incubation period.

**Figure A.3:** The *pssA* in trans expression restored the Δ*pssA* strain motility defect.
Figure A.4: The ΔpssA exhibit irregular cell morphology compared to the wild-type. Phase contrast microscopy of wild-type (wt), ΔpssA, ΔartA, and slg(G796F,F797G) with empty vector pTA963, and ΔpssA complemented with pTA963 expressing pssA grown to mid-log phase in liquid semi-defined CA media supplemented with tryptophan. Liquid media were inoculated with cells from colonies grown on solid CA agar plates (top row), or cells that had been grown by serial transfer in liquid media to mid-log phase (bottom row). Size bar: 10 µm.
Figure A.5: Deletion of *pssA* gene result in the SLG (arrowhead) LDS-PAGE migration shift similarly observed in Δ*artA* and Δ*slg*<sup>G796F,F797G</sup>. Coomassie blue stain of protein extracts from H53 wild-type (wt), Δ*artA*, and Δ*slg*<sup>G796F,F797G</sup>, and Δ*pssA* strain. The migration of molecular mass standards is indicated on the left (in kDa).
Figure A.6: *H. volcanii* SLG lipid-modification is significantly reduced upon *pssA* deletion. Fluorography of protein extracts isolated from wild-type H53 (wt), Δ*pssA* and Δ*pssA* + *pssA* cells grown in the presence of 1μCi/ml 14C mevalonic acid (left). Labeled SLG (arrow) is detected in the wt and *pssA* complementation (Δ*pssA* + *pssA*) extract but a very weak signal in the *pssA* deletion background (Δ*pssA*). The SLG mobility shift in Δ*pssA* is restored upon complementation by *pssA* expression *in trans*. The migration of molecular mass standards is indicated on the left (in kDa).
Figure A.7: The deletion of pssA results in the release of processed Hvo_0405 N-terminal fragment into the supernatant instead of anchored to the membrane. Western blot analysis of membrane (m), cytoplasmic (c), and supernatant (s) fractions for protein isolated from mid-log-phase liquid cultures of H53 wild-type (WT) or the ΔpssA mutant, transformed with pTA963 expressing Hvo_0405 with a 6X His tag, using anti-Hvo_0405-N-term antibodies recognizing the N terminus of Hvo_0405. Numbers indicate molecular masses in kilodaltons, and the arrowhead indicates the processed Hvo_0405 N-terminus. Asterisk (*) represent degradation product of Hvo_0405 N-terminus.
Table 1. Plasmids and strains

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Column1</th>
<th>Relevant characteristic (s)</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pET-22b(+)</td>
<td>AmpR; expression vector: T7 promoter, C-term 6XHis tag</td>
<td>Novagen</td>
<td></td>
</tr>
<tr>
<td>pTA131</td>
<td>Amp’; pBluescript II with BamHI-XbaI fragments from pGB70 harboring pfdx-pyrE2</td>
<td>Allers et al., 2004</td>
<td></td>
</tr>
<tr>
<td>pTA963</td>
<td>Amp’, pyrE2 and hdrB markers, Trp-inducible (ptna) promoter</td>
<td>Allers et al., 2010</td>
<td></td>
</tr>
<tr>
<td>pTA231</td>
<td>AmpR; pBluescript II with pHV2 replication origin harboring pfdx-trpA</td>
<td>Allers et al., 2004</td>
<td></td>
</tr>
<tr>
<td>pAF3</td>
<td>pTA131 carrying chromosomal flanking regions of artA</td>
<td>Abdul Halim et al., 2013</td>
<td></td>
</tr>
<tr>
<td>pAF9</td>
<td>pTA963 carrying artA</td>
<td>Abdul Halim et al., 2013</td>
<td></td>
</tr>
<tr>
<td>pFH24</td>
<td>pTA131 with insert: XhoI::700 bp 5' of hvo_0405::700bp 3' of hvo_0405::XbaI</td>
<td>Abdul Halim et al., 2017</td>
<td></td>
</tr>
<tr>
<td>pFH25</td>
<td>pTA963 with insert: p.tnaA::NdeI::hvo_0405 cds::GGP linker::6x His::stop::EcoRI</td>
<td>Abdul Halim et al., 2017</td>
<td></td>
</tr>
<tr>
<td>pFH26</td>
<td>pTA131 carrying fragment containing nucleotide encoding PFG with 400 bp flanking the intergenic region of slg and hvo_2073. The trpA gene cloned from pTA231 was inserted into the middle of the flanking fragments.</td>
<td>Abdul Halim et al., 2015</td>
<td></td>
</tr>
<tr>
<td>pJS143</td>
<td>pET-22b(+) with insert: T7 promoter::lacO::ribosomal binding site::NdeI::hvo_0405 (bp 64-1344)::KL(HindIII)AAAAL linker::6x His tag::stop</td>
<td>Abdul Halim et al., 2017</td>
<td></td>
</tr>
<tr>
<td>pJS151</td>
<td>pTA963 with insert: p.tnaA::Hvo_0405 cds with coding bases 7-12 mutated from CGCCGC to AAGAAG (changes RR to LL)::GGP linker::6XHis tag::stop</td>
<td>Abdul Halim et al., 2017</td>
<td></td>
</tr>
<tr>
<td>Strain</td>
<td>Description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>E. coli</strong></td>
<td>DH5α</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>E. coli</em> F− ϕ80lacZΔM15 (lacZYA-argF)U169 recA1endA1 hsdR17(rK− mK−) phoA supE44 thi-1 gyrA96 relA1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Invitrogen</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DL739</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>E. coli</em>, MC-4100 recA dam-13::Tn9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Blyn <em>et al.</em>, 1990</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BL21(DE3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>E. coli</em> CmR; strain B F− ompT gal dcm lon hsdSB(rB− mB−) gal dcm λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λS) pLysS[T7p20 orip15A]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Miroux <em>et al.</em>, 1996</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>H. volcanii</strong></td>
<td>H53</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H. volcanii, ΔpyrE2 ΔtrpA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Allers <em>et al.</em>, 2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AF103</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H53 ΔartA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abdul Halim <em>et al.</em>, 2013</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AF104</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H53 ΔartA* motility suppressor mutants</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abdul Halim <em>et al.</em>, 2013</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AF109</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AF109 containing pAF9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abdul Halim <em>et al.</em>, 2013</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AF110</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H98 ΔartA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abdul Halim <em>et al.</em>, 2013</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ΔaglB</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H53 DaglB</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abu-Qarn <em>et al.</em>, 2007</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FH24</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H53 Δhvo_0405</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abdul Halim <em>et al.</em>, 2017</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FH35</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H53 ΔartA Δhvo_0405</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abdul Halim <em>et al.</em>, 2017</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FH37</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H53 ΔpyrE2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abdul Halim <em>et al.</em>, 2015</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FH38</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H53 slgG796F, F797G ΔpyrE2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abdul Halim <em>et al.</em>, 2015</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FH41</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FH35 containing pFH25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abdul Halim <em>et al.</em>, 2017</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FH59</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FH24 containing pJS151</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abdul Halim <em>et al.</em>, 2017</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H98</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H. volcanii, ΔpyrE2 ΔhdrB</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Allers <em>et al.</em>, 2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>JS1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H53 containing pTA963</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abdul Halim <em>et al.</em>, 2017</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>JS13</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H53 containing pFH25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abdul Halim <em>et al.</em>, 2017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Code</td>
<td>Description</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------------------------------</td>
<td>--------------------------------</td>
<td></td>
</tr>
<tr>
<td>JS16</td>
<td>FH24 containing pTA963</td>
<td>Abdul Halim et al., 2017</td>
<td></td>
</tr>
<tr>
<td>JS17</td>
<td>FH24 containing pFH25</td>
<td>Abdul Halim et al., 2017</td>
<td></td>
</tr>
<tr>
<td>JS23</td>
<td>FH35 containing pTA963</td>
<td>Abdul Halim et al., 2017</td>
<td></td>
</tr>
<tr>
<td>JZ101</td>
<td>H53 containing pAF9</td>
<td>Abdul Halim et al., 2013</td>
<td></td>
</tr>
<tr>
<td>JZ102</td>
<td>AF110 containing pAF9</td>
<td>Abdul Halim et al., 2013</td>
<td></td>
</tr>
<tr>
<td>JZ103</td>
<td>H98 containing pAF9</td>
<td>Abdul Halim et al., 2013</td>
<td></td>
</tr>
</tbody>
</table>
Table S2: Primers used for PCR amplification

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5’-3’)</th>
<th>Target sequences/Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>artA KO FW BamHI</td>
<td>AAGCTAGGATCCGAACTCGAATCCTCC</td>
<td>685 bp upstream of <em>artA</em> start codon, extension towards <em>artA</em> gene</td>
</tr>
<tr>
<td>artA KO RV XbaI</td>
<td>GAACCTCTAGACCTACCACGATACTG</td>
<td>688 bp downstream of <em>artA</em> stop codon, extension towards <em>artA</em> gene</td>
</tr>
<tr>
<td>artA KO FW up-down</td>
<td>CCGCATCCAGTACGTTCAGCGCTCGGAGT</td>
<td>1 bp downstream of <em>artA</em> stop codon, extension away from <em>artA</em> gene</td>
</tr>
<tr>
<td>artA KO RV up-down</td>
<td>ACTCCGACGGCGTAAACGTACTGGATCGG</td>
<td>1 bp upstream of <em>artA</em> stop codon, extension away from <em>artA</em> gene</td>
</tr>
<tr>
<td>artA OE FW</td>
<td>GCTATCCCATATGCCGCTTCCTC</td>
<td>1 bp upstream of <em>artA</em> start codon, extension towards <em>artA</em> gene</td>
</tr>
<tr>
<td>artA OE RV</td>
<td>TATATTGAATTCAGTCGAGCCTAAACTC</td>
<td>1 bp downstream of <em>artA</em> stop codon, extension towards <em>artA</em> gene</td>
</tr>
<tr>
<td>FW_Hvo_2072 PFG XhoI</td>
<td>CTCGAGGAGCCGACTGAGGAGACCCCCCGTCCTCCAACGCCCTTCGGCGTGATCGCT</td>
<td>400 bp upstream of the trpA insertion site with primers encoding the PFG codon (highlighted), extension towards the insertion site</td>
</tr>
<tr>
<td>p.fdx-trpA upstream RV</td>
<td>CCGGAACGCACATAAGTCCCTCTCGGTTTCG</td>
<td>1 bp upstream of the trpA insertion site, extension away from insertion site. This primer was also used to sequence the slg C-term to confirm the PFG motif encoding chromosomal replacement.</td>
</tr>
<tr>
<td>p.fdx-trpA FW</td>
<td>CAGAACGCAGAGGACTTATGTGCGTTCCGG</td>
<td>1 bp upstream of p.fdx-trpA in pTA231, extension towards trpA gene</td>
</tr>
<tr>
<td>p.fdx-trpA RV</td>
<td>CTGTTTTCGAGTACGCTACGCAGCTCGTGGATACCG</td>
<td>1 bp downstream of p.fdx-trpA stop codon in pTA231, extension towards trpA gene</td>
</tr>
<tr>
<td>p.fdx-trpA downstream FW</td>
<td>CGAGGGGTGTTTATCCACCGGATCGTGCGTACTCGAAGTGGCA</td>
<td>1 bp downstream of the trpA insertion site, extension away from insertion site</td>
</tr>
<tr>
<td>Gene References</td>
<td>Primer Sequences</td>
<td>Notes</td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------</td>
<td>-------</td>
</tr>
<tr>
<td>RV_Hvo20_73_XbaI</td>
<td>ATCTTATCTAGAGAAGCGCGAGT GATCTACGG</td>
<td>400 bp downstream of the <em>trpA</em> insertion site, extension towards the insertion site</td>
</tr>
<tr>
<td>FW_0405_XhoI</td>
<td>AATAAAAACTCGAGAACTGGGCCA CTGGA</td>
<td>700 bp upstream of <em>hvo_0405</em>, extension towards <em>hvo_0405</em></td>
</tr>
<tr>
<td>RV_0405_XbaI</td>
<td>ATCTTATCTAGAGTTCCGCCCCGC TGGA</td>
<td>700 bp downstream of <em>hvo_0405</em>, extension towards <em>hvo_0405</em></td>
</tr>
<tr>
<td>FW_0405_Up-DW</td>
<td>ACCCGCAGATTCATCCCACCTT TTTTCGC</td>
<td>1 bp downstream of <em>hvo_0405</em> stop codon, extension away from <em>hvo_0405</em> gene</td>
</tr>
<tr>
<td>RV_0405_Up-DW</td>
<td>GCGAAAAAAGGTGGGATGAATC TGCAGGT</td>
<td>1 bp upstream of <em>hvo_0405</em> start codon, extension away from <em>hvo_0405</em> gene</td>
</tr>
<tr>
<td>FW_0405_KO_flank</td>
<td>CGCCGCTTATCATCTACGTC</td>
<td>649 bp upstream of <em>hvo_0405</em>, extension towards <em>hvo_0405</em></td>
</tr>
<tr>
<td>RV_0405_KO_flank</td>
<td>CGAGGGAACGGTGGAACT</td>
<td>601 bp downstream of <em>hvo_0405</em>, extension towards <em>hvo_0405</em></td>
</tr>
<tr>
<td>FW0405Nd_el</td>
<td>ATTAATCATATGGACCGCCGCAA GTTC</td>
<td><em>hvo_0405</em> start codon, extension towards <em>hvo_0405</em></td>
</tr>
<tr>
<td>RV0405Ec_oRIHIS</td>
<td>TATATTGAATTCTCAGTGATGGT GATGTTGATGCCGGCCGCGGAAG TGCCCGGGTCC</td>
<td><em>hvo_0405</em> stop codon, extension towards <em>hvo_0405</em>, contains codon encoding 6X-His tag sequence</td>
</tr>
</tbody>
</table>


Klingl, A. (2014). "S-layer and cytoplasmic membrane—exceptions from the typical archaeal cell wall with a focus on double membranes."


aureus at the LPXTG motif." Proceedings of the National Academy of Sciences 96(22): 12424-12429.


191