The Response of *Haloferax volcanii* to Salt and Temperature Stress: A Proteome Study by Label-Free Mass Spectrometry

Živojin Jevtić, Britta Stoll, Friedhelm Pfeiffer, Kundan Sharma, Henning Urlaub, Anita Marchfelder,* and Christof Lenz*

In-depth proteome analysis of the haloarchaeal model organism *Haloferax volcanii* has been performed under standard, low/high salt, and low/high temperature conditions using label-free mass spectrometry. Qualitative analysis of protein identification data from high-pH/reversed-phase fractionated samples indicates 61.1% proteome coverage (2509 proteins), which is close to the maximum recorded values in archaea. Identified proteins match to the predicted proteome in their physicochemical properties, with only a small bias against low-molecular-weight and membrane-associated proteins. Cells grown under low and high salt stress as well as low and high temperature stress are quantitatively compared to standard cultures by sequential window acquisition of all theoretical mass spectra (SWATH-MS). A total of 2244 proteins, or 54.7% of the predicted proteome, are quantified across all conditions at high reproducibility, which allowed for global analysis of protein expression changes under these stresses. Of these, 2034 are significantly regulated under at least one stress condition. KEGG pathway enrichment analysis shows that several major cellular pathways are part of *H. volcanii*’s universal stress response. In addition, specific pathways (purine, cobalamin, and tryptophan) are affected by temperature stress. The most strongly downregulated proteins under all stress conditions, zinc finger protein HVO\_2753 and ribosomal protein S14, are found oppositely regulated to their immediate genetic neighbors from the same operon.

1. Introduction

Like bacteria, archaea often face rapid and thorough changes of environmental conditions to which they have to respond quickly in order to survive. Although bacterial stress responses have already been under investigation for some time, comparatively little is known about how archaeal organisms respond to stresses. Archaea not only live in sites with extreme living conditions but are ubiquitous or even dominant in many environments. Despite their abundance they are still understudied, even though they are useful model systems for processes in both eukarya and bacteria and especially interesting for their specific archaeal features.[1] To learn more about the archaeal stress response, we comprehensively investigated the response of the halophilic archaeon *Haloferax volcanii* to salt and temperature stress on the proteome level. *Haloferax volcanii* is a halophilic euryarchaeon first isolated from the shores of the Dead Sea.[2] It grows best at around 45 °C, requires a salinity of approximately 2.5 M NaCl (≈15%) and maintains an equally high intracellular salt concentration.[2,3] *Haloferax volcanii* is a model organism for the haloarchaea, it is easy to cultivate under laboratory conditions, its genome is sequenced,[4] and genetic techniques are available.[1] In contrast to other haloarchaea, *H. volcanii* grows on a wide range of salt concentrations[2,5] suggesting that its osmoregulatory responses are very efficient.

More than 30 years ago Daniels et al. performed the first analyses of protein changes induced by heat shock and low salt concentrations in different haloarchaea including *H. volcanii.[6] They
observed the enhanced expression of several proteins under stress conditions, only one of which was mapped at the gene level.\textsuperscript{[7,8]} This was supported by Mojica et al. who confirmed these observations and showed that general stress response proteins might be important for adaptation to low and high salt concentrations.\textsuperscript{[9]} In both studies, both transient and permanent expression changes were detected during pulse/chase experiments by SDS-PAGE and 2D gel electrophoresis (2DE), however without identification of the involved proteins. A more global approach was taken 11 years later using 2DE and MS. Here, the method first had to be adjusted to the low isoelectric points of the Haloferax proteins.\textsuperscript{[10]} Forty-four proteins were detected in another study as being differentially expressed in response to changing salt concentrations; 18 of these were identified by MS analysis,\textsuperscript{[11]} among them the stress response protein PspA. Similarly, a 2DE study identified heat shock proteins in Halobacterium.\textsuperscript{[12]} Recently, several large scale quantitative proteome analyses of archaeal organisms have been published, for example, in Halobacterium salinarum.\textsuperscript{[13–15]} In H. volcanii, a shotgun proteomics approach using multiple separation and enrichment strategies was published 8 years ago, identifying 1296 of the 4063 annotated proteins.\textsuperscript{[4,16]} Recently, spectral counting,\textsuperscript{[17,18]} differential isotopic labelling strategies\textsuperscript{[19,20]} as well as label-free quantitation\textsuperscript{[21]} were employed to study oxidative stress, the impact of individual mutations, or the exponential-stationary growth phase transition, respectively, in Haloferax. Additional publications cover topics at the intersection of proteomics and the stress response for other archaea.\textsuperscript{[22–27]}

Here, we aimed at comprehensively profiling proteome changes in H. volcanii grown continuously under temperature and salt stress conditions. To this end, we used the recently established SWATH-MS method (SWATH-MS: Sequential Window Acquisition of All Theoretical Mass Spectra),\textsuperscript{[28–30]} which allows for global label-free quantitation of multiple proteomes in a highly parallel fashion, without the need to establish, for example, autotrophic strains or chemical labelling workflows. As a peptide-centric approach, SWATH-MS is best performed using a priori, context-specific peptide and protein identification data to extract quantitative information at acceptable false-positive rates.\textsuperscript{[31]} To this end, we prepared an in-depth MS/MS spectral library from both standard and stress condition samples, which presents the most comprehensive experimental proteome map of H. volcanii to date. SWATH-MS has previously been employed for in-depth profiling of bacterial and archaean proteomes.\textsuperscript{[15,32]} In this study, it provided highly reproducible and detailed protein expression profiles under salt and temperature stress.

### 2. Experimental Section

#### 2.1. Strains and Growth Conditions

*Haloferax volcanii* strain H119 was used for all experiments. H119 is a triple mutant (ΔpyrE2, ΔtrpA, and ΔleuB) of strain DS70,\textsuperscript{[13]} which is derived from DS2 by curing of the small plasmid pHV2.\textsuperscript{[14]} It is frequently used as a standard laboratory strain because the three deleted genes function as selection markers for genetic manipulations. H119 was grown aerobically with shaking (200 rpm) in Hv-YPC medium\textsuperscript{[13]} to exponential phase (0.55 OD\textsubscript{650} at 45 °C with an optimal concentration of 18% saltwater (2.46 m NaCl, 88 mM MgCl\textsubscript{2}, 85 mM MgSO\textsubscript{4}, 56 mM KCl, 12 mM Tris-HCl, and pH 7.5). For low and high salt stress, salt concentration was adjusted to 15% and 23%, respectively, without modifying the ion ratios. Standard conditions (18% saltwater) correspond to 15% NaCl, low salt (15% saltwater) corresponds to 10.8% NaCl, and high salt (23% saltwater) corresponds to 19.2% NaCl. For low and high temperature stress, cells were grown at 30 and 53 °C, respectively. Cell growth was monitored by measurement of the OD at 650 nm wavelength (OD\textsubscript{650}). For each condition, three biological replicates were prepared.

#### 2.2. Protein Isolation and In-Solution Digestion

*Haloferax* cells were harvested by centrifugation, resuspended in 18% saltwater, and sonicated. Cells were subsequently incubated with sodium taurodeoxycholate (0.006% final concentration), and nonsoluble parts were removed by centrifugation (100 000 × g for 1 h at 4 °C). The supernatant was incubated with DNase I, exonuclease III, and RNase A to digest nucleic acids. Aliquots of 0.5 mL were frozen in liquid nitrogen and stored at −80 °C. Aliquots were thawed and proteins were precipitated by adding 4.5 mL 100% cold acetone and incubation at −20 °C overnight. Proteins were pelleted by centrifugation at 6000 × g for 60 min at 4 °C. Pellets were washed four to five times with 80% acetone and in a final washing step with 5 mL 80% ethanol. Pellets were air-dried and frozen in liquid nitrogen until further use.

For LC/MS/MS analysis, equivalents of 50 μg of protein pellets were resuspended using 1% Rapigest (Waters) cleavable surfactant dissolved in 25 mM ammonium bicarbonate (Ambic).\textsuperscript{[15]} Reduction of disulfide bonds and alkylation of free
cysteine residues was performed by incubation with 50 mM dithiothreitol in 25 mM Ambic (56 °C, 1 h) and 100 mM iodoacetamide in 25 mM Ambic (37 °C, 1 h), respectively. Protein digestion was carried out overnight at 37 °C using sequencing grade porcine trypsin (Promega) at a 1:20 enzyme to substrate ratio (w:w). Cleavage of the surfactant was carried out by acidification with 5% TFA, and the resulting fatty acids were removed by centrifugation (13,000 rpm, 30 min, RT). The supernatants containing tryptic peptides were transferred to new tubes, dried in a SpeedVac, and stored at −20 °C.

For generating a spectral library by data-dependent mass spectrometry acquisition (DDA), two pools were generated from peptide aliquots of a) cultures grown in standard conditions and b) cultures from the examined stress conditions. Samples were dried in a SpeedVac and redissolved in 0.1% TFA. Pools a and b were each separated into eight fractions of increasing hydrophobicity using high-pH/reversed-phase separation in a spin column format (Pierce High pH Peptide Fractionation Kit, Thermo Fisher Scientific). Separations were performed in duplicate, and the corresponding fraction duplicate pooled, dried in a SpeedVac and stored at −20 °C until further analysis.

2.3. LC/MS/MS Acquisition

Proteolytic digests and reversed-phase peptide fractions were dissolved in 2% ACN, 0.1% formic acid (FA) in water to a nominal concentration of 0.2 μg μL⁻¹. All samples were analyzed on a nanoflow chromatography system (Eksigent nanoLC 425, Sciex) coupled to a hybrid quadrupole-orthogonal time of flight mass spectrometer (TripletOF 5600+, Sciex). A total of 1 μg protein equivalent was preconcentrated on a self-packed reversed-phase C18 precolumn (40 mm × 0.15 mm Reprosil C18-AQ 120 Å, particle size 5 μm; Dr. Maisch) and separated on a self-made capillary column (250 mm × 0.075 mm Reprosil C18-AQ 120 Å, particle size 3 μm) using a linear gradient (5% to 35% ACN, 0.1% FA over 90 min, 300 nL min⁻¹, 50 °C).

To generate a qualitative spectral library, reversed-phase peptide fractions from both pools were analyzed in duplicate in DDA mode, with an MS survey scan of m/z 350–1250 accumulated for 250 ms at a resolution of 35,000 Full Width Half Maximum (FWHM). Up to 25 candidate precursors with a charge state of 2+ to 4+ and a threshold intensity of 125 cps were isolated for MS/MS, then dynamically excluded for 20 s after one occurrence. MS/MS spectra were accumulated for 100 ms within a range of m/z 180–1750, at a resolution of 17,500 FWHM. Fragmentation was achieved using default rolling collision energy settings and nitrogen as collision gas.

For quantitative analysis, the nonpooled, nonfractionated biological replicates for each sample were analyzed in technical triplicate using SWATH acquisition with 60 variable size Q1 isolation windows. Precursor isolation window widths were calculated using the SWATH Variable Window Calculator v1.0 (Sciex). A TOF MS survey scan was conducted for 250 ms across the m/z 350–1250 range, followed by product ion scans accumulated for 65 ms using rolling collision energy settings for 2+ precursors, a collision energy spread of ±5 V and a product ion m/z range of 350–1500.

2.4. LC/MS/MS Data Processing

Protein identification was performed using the Paragon search engine v5.0.0.0 implemented in ProteinPilot v5.0 build 4769 (Sciex) against an in-house H. volcanii protein sequence database (v19.11.2015, 4105 entries, 79 additional noncoding [spurious] ORFs, File S1. Supporting Information) complemented with 51 common lab contaminants. This version includes reannotations that were made in the context of whole-genome TSS analyses.[38] Replica data from DDA analysis of the peptide fractions (1x “standard” pool and 2x “stress” pool) were combined in a single search using “thorough” settings. Protein and peptide identifications were trimmed to an estimated false discovery rate (FDR) of 1% using a forward/reverse decoy strategy. An MS/MS spectral library was generated using the SWATH 2.0 Microapp integrated in PeakView v2.1 build 11041 (Sciex). SWATH data for standard, salt stress, and temperature stress conditions were processed together for subsequent analysis of differential protein expression.[39] Peak extraction used the following parameters: up to ten peptides per protein; up to six transitions per peptide; exclusion of modified peptides; peptide confidence threshold 95%; and 1% FDR threshold.[39] Peptide retention times were aligned by linear regression using a set of evenly distributed endogenous peptides detected in all samples. After alignment, retention time windows of 8 min were used for peak grouping. Extracted transition peak areas were summed up to the respective peptide and protein areas, which were exported for further analysis. MS raw data, protein identification, and protein quantitation results were deposited in the ProteomeXchange Consortium PRIDE[30] partner repository under dataset identifier PXD011056.

2.5. Data Analysis

Coverage values were analyzed by calculating the distribution of major physicochemical properties (isoelectric point and molecular weight) of identified, quantified, and differentially expressed proteins, and comparing them to the complete proteome. Membrane proteome coverage was similarly assessed using TMHMM server v2.0 (www.cbs.dtu.dk/services/TMHMM/).[40] KEGG mapper (Kyoto Encyclopedia of Genes and Genomes, http://www.genome.jp/kegg/pathway.html for pathway reconstruction; http://www.genome.jp/kegg/pathway2.html for pathway coloring) was used to visualize the coverage of metabolic pathways.[41,42] Exported protein areas were normalized by total area sums. Normalized values were imported into Perseus v1.5.6.0[43] and transformed to log2 scale. A nondirected principal component analysis (PCA) was carried out to examine the reproducibility of biological and technical replicates. Protein peak areas of all stress conditions were compared pairwise to standard conditions using Student’s t-test (p < 0.05) with Benjamini–Hochberg correction. For quantitative analyses and discussion, the original results matrix (2244 quantified proteins) was reduced to values showing significant up- or downregulating for at least one stress condition, without any further filtering for a minimum fold change (2034 proteins). At moderate stringency, this allowed for a global view on protein abundance changes with regard to cellular processes. The protein quantitation matrix was
Further normalized by z-Scoring and analyzed by Hierarchical Clustering using Euclidean Distance and k-means preprocessing. Selected protein clusters indicative of either universal or stress-specific trends were analyzed for enrichment of KEGG pathways using the DAVID Bioinformatics suite at medium stringency settings ([https://david.ncifcrf.gov/](https://david.ncifcrf.gov/)). Additional functional classification was performed using the Archaeal Clusters of Orthologous Genes (arCOGs) system. In short, arCOG protein sequences assigned to “Haloferax volcanii DS2” were extracted from the arCOG protein sequence file, identity-matched to sequences from the *H. volcanii* annotated proteome, and the remaining sequences further curated manually including BLASTp analysis.

In order to determine the proteins with the highest expression changes either in clusters from hierarchical clustering or for a given stress condition, log2 fold change (logFC) values were converted to fold change (FC) values using custom PERL scripts. Only values representing significant regulation under a given stress condition were considered. For clusters assigned to a single stress-specific trend, proteins were ranked directly by their FC value. For universal stress conditions and for clusters assigned to multiple stress-specific trends, the maximal FC value was used for ranking. The 20 most differentially expressed proteins were inspected for biological relevance.

### 3. Results and Discussion

*Haloferax volcanii* cells were grown continuously under standard or stress conditions in order to analyze for long-term adaptation at the protein level. OD650 monitoring indicated a strong impact of temperature stress on cell growth, whereas changes in salinity caused relatively minor changes (Figures S1 and S2, Supporting Information). We first analyzed cultures by LC/MS/MS in order to generate a sequence-annotated MS/MS spectral library as a basis for further quantitation. This also allowed to check which parts of the proteome are accessible by our straightforward, single-shot methodology. DDA data obtained from pooled samples fractionated by high-pH/reversed-phase were used for protein identification by peptide-to-spectrum matching (Figure 1). At a respective FDR of 1% on both the peptide and the protein level, 2509 protein identifications were obtained (Table S1, Supporting Information). This corresponds to 61.1% of the predicted *H. volcanii* proteome, and represents the most comprehensive proteome characterization of this organism to date. This is in keeping with comprehensive proteome studies in other archaea, which suggests that a coverage of ≈65% of the predicted proteome is the maximum that may be expected under a defined set of growth conditions, even when a much greater analytical effort is made. The molecular weight and isoelectric point...
distributions of the identified proteins reflected those of the overall proteome. We did, however, observe an under-representation of small proteins, which is probably a side effect of the acetone precipitation used in our protocol, combined with a reduced number of MS-compatible peptides per protein (Figures S3 and S4, Supporting Information). Analysis of TMHMM-predicted transmembrane domains in the protein identifications and predicted proteome showed that we identified 378 out of 1000 predicted membrane proteins, which reflects some under-representation as TM protein recovery is 15% for identified proteins, whereas this ratio is 24% for the complete proteome. This demonstrates that our straightforward, nonfractionating sample preparation protocol does not bias strongly against the detection of membrane proteins either. Finally, annotation of the identified proteins onto the H. volcanii KEGG map of metabolically active proteins showed high coverage (395/453, or 87.2%). Taken together, our data show that the simple but straightforward workflow consisting of lysis, protein precipitation, detergent-assisted solubilization, trypsinization, and single-shot LC/MS/MS analysis provides a comprehensive, low-bias view of the expressed proteome of H. volcanii under the examined conditions.

3.1. Protein Quantitation in H. Volcanii under Standard, Temperature, and Salt Stress Conditions

We constructed an annotated Peptide MS/MS Spectral Library from the protein identification results, and used it to extract quantitation data out of replicate single-shot SWATH-MS acquisitions for standard, low and high, and low and high temperature conditions. Single-shot acquisition allowed us to use high biological and technical replication (3 × 3). Quantitative data were extracted for 2244 proteins (89.4% of the identified, or 54.7% of the complete proteome) at 1% FDR of peak integration. We compared the identified and quantified sets of proteins to the predicted proteome following functional annotation using the arCOGs classification approach.[46] With few exceptions, the identified and quantified proteomes closely reflect the predicted one, indicating that our experimental approach provides a relatively unbiased view of cellular function (Table S3, Supporting Information). Cumulative CV analysis showed 41–59% of proteins with CVs of 20% or better, and 76–83% of proteins with CVs of 40% or better, with median CVs between 15% and 23% for the five series. The observed protein peak areas spanned a dynamic range of 4.6 orders of magnitude (Figure S5, Supporting Information). Consequently, pairwise t-testing of stress conditions against the standard sample resulted in high numbers of statistically significant fold changes (Table S1, Supporting Information), which enables the detection of relatively moderate expression changes. Nondirected PCA showed clear separation in the major components by biological state rather than by biological or technical replicate, again indicating high reproducibility (Figure 2). Proteome profiles under low and high temperature stresses clustered further away from standard conditions than the two salt stress conditions, indicating a more pronounced impact of temperature changes on protein expression, a finding corroborated by hierarchical clustering analysis.

3.2. The Search for a “Universal Stress Signature”

Our first question aimed at identifying common responses to different stresses. To this end, we performed a hierarchical clustering analysis of the 2034 proteins which showed significant up- or downregulation (p < 0.05, Benjamini–Hochberg corrected) under at least one stress condition relative to standard conditions (Figure 3). All stress conditions showed pronounced protein regulation patterns, with high temperature stress exhibiting the greatest distance to standard conditions, and low salt and low temperature stresses showing highest similarity of patterns.

Clustering analysis of protein regulation profiles showed three clusters that consistently exhibited up- or downregulation under all stress conditions (Figure 3 and Figure S6, Supporting Information): cluster 1976 consisting of 245 consistently and significantly downregulated proteins, and clusters 1977 and 1981 containing 134 consistently and significantly upregulated proteins (Table 1, Supporting Information). Again we performed arCOGs annotation of these “universal stress response proteins.” Surprisingly, this annotation yielded a pattern that closely reflects that of the predicted proteome, rendering this approach largely uninformative (Figure S8 and Table S3, Supporting Information). We therefore interrogated the clusters for functional enrichment based on their KEGG pathway annotations. Although KEGG currently only contains functional annotation for 921 out the 4105 predicted H. volcanii proteins, its pathway assignments are nonetheless more stringent and indicative of biological processes than, for example, Gene Ontology.

Several KEGG pathways were significantly enriched in clusters 1976, 1977, and 1981, which showed consistent up- or down-regulation of proteins under all stress conditions (“universal stress response”; Table 1). Three of the nine enriched pathways summarize general metabolic charts (metabolic pathways, carbon metabolism, and biosynthesis of amino acids) rather than specific pathways, which may indicate global metabolic fine-tuning as a universal stress response. In addition, four more specific pathways were slightly enriched (glycolysis; citrate cycle;
pyrimidine metabolism; glycine, serine, and threonine metabolism) as well as two maps from Genetic Information Processing (RNA polymerase and nucleotide excision repair).

All enriched KEGG maps were visually inspected, as well as other pathways that came into focus upon analysis of more defined proteome changes. In several cases, we additionally implemented annotation enhancements based on the current state-of-the-art,[37] which are not yet reflected in KEGG (see Text, Supporting Information). In addition to overall enriched biological pathways, we also observed a number of more defined proteome changes associated with multiple stress conditions by inspecting the most strongly regulated members across all conditions (Table 2, Supporting Information). The most strongly downregulated protein (HVO_2753, 230- to 670-fold) is a small protein (59 residues). It is one of the 72 proteins that belong to the “small CPxCG related zinc finger proteins.”[47] The gene for HVO_2753 is encoded in an operon together with HVO_2752, which codes for the beta subunit of translation elongation factor aEF-1. However, HVO_2753 and HVO_2752 are oppositely regulated. HVO_2753 is generally downregulated, whereas HVO_2752 is strongly upregulated at low and high temperatures (up to 26-fold) but is downregulated at high salt. The next highly downregulated protein under all stress conditions was ribosomal protein S14 (HVO_2550; 78- to 217-fold). Ribosomal protein S14

![Figure 3. Hierarchical clustering of proteins significantly regulated in at least one stress condition. Major clusters showing consistent up- (red) or downregulation (blue) relative to standard condition in either single or all conditions are highlighted on the left. Software-generated cluster identifiers are used. Exemplifying a common biological response, clusters 1977 and 1981 are treated as one group throughout this manuscript.](image-url)
is also oppositely regulated to its genomic neighbor, ribosomal protein L5 (HVO\_2551, upregulated four to eightfold under all stress conditions; Figure S8, Supporting Information). The two genes are encoded within the longest polycistronic operon (spc operon) consisting of 25 ribosomal protein genes and the two genes have a 4-base gene overlap, indicating translational coupling. Opposite regulation of these ribosomal proteins might be related to an additional, extra-ribosomal function. It should be noted that a strong but oppositely oriented regulation has also been observed for a small subset of *Escherichia coli* ribosomal proteins.\[^{48}\] Notably, *E. coli* L5 was upregulated under high temperature stress, whereas three other ribosomal proteins (S2, S4, and S11) were strongly downregulated under the same conditions.

The most strongly generally upregulated protein is the plant-type ferredoxin Fdx (HVO\_2995, 20- to 43-fold), which has a redox potential similar to NAD.\[^{49}\] This ferredoxin is involved in oxidative decarboxylation of pyruvate to acetyl-CoA and of alphaketoglutarate to succinyl-CoA as shown for *Halobacterium*.\[^{49-51}\] These enzymes of pyruvate metabolism and the citric acid cycle are also generally upregulated and contribute to the identification of these KEGG pathways as enriched. This may indicate an enhanced energy demand under such stress conditions. Fdx is also involved in nitrate assimilation\[^{52}\] due to its probable interaction with nitrite reductase (HVO\_1788, substrate may also be sulfite). This enzyme is, however, downregulated at high temperature. For additional defined proteome changes assigned to a general stress response see Text S1, Supporting Information.

### 3.3. Specific Temperature Stress Responses

Following the identification of common, universal stress signatures among the examined salt and temperature stresses, we next set out to investigate more stress-specific proteome alterations. Both PCA and z-scored hierarchical clustering indicated that high and low temperature stress had the strongest impact on *H. volcanii* protein expression. Clusters 1979 (376 significant proteins), 1978 (148 significant proteins) and 1983 (117 significant proteins) contain proteins specifically up- or downregulated upon temperature stress.

The heat shock response has been well studied in all three domains of life, leading to the identification of several heat shock protein families.\[^{53-55}\] As only few members of these families were actually found to be upregulated upon continuous growth at high temperature, we systematically inspected all *Halofex* proteins classified into one of the known heat shock protein families. This confirmed that most of these are not upregulated upon continuous growth at high temperature (for full details see Text S5, Supporting Information). This is not surprising in itself as the heat shock response is rapid but transient.\[^{22,53,55,56}\]

### Table 1. KEGG pathway enrichment analysis

Analysis was performed using proteins either significantly altered in at least three out of four stress conditions (Universal Stress Response) or in a specific stress condition (high and low temperature, high salt) compared to standard condition.

| KEGG term                        | KEGG pathway                  | N/total | Fold enrichment | p-Value | Corrected |
|----------------------------------|--------------------------------|---------|-----------------|---------|-----------|
| Universal Stress Response (Clusters 1976, 1977, and 1981) | hvo01100 Metabolic pathways 84/499 2.3 1.93 × 10^{-5} 1.66 × 10^{-6} | 1.3 2.49 × 10^{-4} 1.71 × 10^{-2} |                   |
|                                  | hvo00240 Pyrimidine metabolism 14/48 2.2 5.74 × 10^{-1} 1.80 × 10^{-1} | 2.2 5.74 × 10^{-1} 1.80 × 10^{-1} |                   |
|                                  | hvo01200 Carbon metabolism 23/104 1.7 9.70 × 10^{-3} 2.01 × 10^{-1} | 1.7 9.70 × 10^{-3} 2.01 × 10^{-1} |                   |
|                                  | hvo03020 RNA polymerase 6/13 3.5 1.88 × 10^{-2} 2.79 × 10^{-1} | 3.5 1.88 × 10^{-2} 2.79 × 10^{-1} |                   |
|                                  | hvo00260 Glycine/serine/treonine metabolism 10/37 2.1 4.13 × 10^{-2} 4.41 × 10^{-1} | 2.1 4.13 × 10^{-2} 4.41 × 10^{-1} |                   |
|                                  | hvo00020 Citrate cycle (TCA cycle) 8/27 2.3 5.09 × 10^{-2} 4.52 × 10^{-1} | 2.3 5.09 × 10^{-2} 4.52 × 10^{-1} |                   |
|                                  | hvo00010 Glycolysis/glucogenogenesis 10/39 1.4 7.53 × 10^{-2} 4.91 × 10^{-1} | 1.4 7.53 × 10^{-2} 4.91 × 10^{-1} |                   |
|                                  | hvo01230 Biosynthesis of amino acids 21/112 1.2 5.34 × 10^{-2} 6.11 × 10^{-1} | 1.2 5.34 × 10^{-2} 6.11 × 10^{-1} |                   |
|                                  | hvo03420 Nucleotide excision repair 5/13 2.9 7.71 × 10^{-2} 4.39 × 10^{-1} | 2.9 7.71 × 10^{-2} 4.39 × 10^{-1} |                   |
| Temperature-specific response (Clusters 1978, 1979, and 1983) | hvo03010 Ribosome 37/62 2.3 1.93 × 10^{-5} 1.66 × 10^{-6} | 2.3 1.93 × 10^{-5} 1.66 × 10^{-6} |                   |
|                                  | hvo00860 Porphyrin and chlorophyll metabolism 16/31 2.0 4.64 × 10^{-3} 1.81 × 10^{-1} | 2.0 4.64 × 10^{-3} 1.81 × 10^{-1} |                   |
|                                  | hvo00220 Arginine biosynthesis 10/20 1.9 4.58 × 10^{-2} 7.39 × 10^{-1} | 1.9 4.58 × 10^{-2} 7.39 × 10^{-1} |                   |
|                                  | hvo04122 Sulfur relay system 6/9 2.6 5.33 × 10^{-2} 6.92 × 10^{-1} | 2.6 5.33 × 10^{-2} 6.92 × 10^{-1} |                   |
|                                  | hvo01100 Biosynthesis of sec. metabolites 69/227 1.2 5.34 × 10^{-2} 6.11 × 10^{-1} | 1.2 5.34 × 10^{-2} 6.11 × 10^{-1} |                   |
|                                  | hvo03030 DNA replication 10/21 1.8 6.28 × 10^{-2} 6.05 × 10^{-1} | 1.8 6.28 × 10^{-2} 6.05 × 10^{-1} |                   |
| High salt-specific response (Cluster 1980) | hvo01100 Metabolic pathways 24/499 1.5 2.03 × 10^{-3} 6.30 × 10^{-2} | 1.5 2.03 × 10^{-3} 6.30 × 10^{-2} |                   |
|                                  | hvo00790 Folate biosynthesis 4/13 9.7 6.04 × 10^{-3} 9.51 × 10^{-2} | 9.7 6.04 × 10^{-3} 9.51 × 10^{-2} |                   |
|                                  | hvo03020 RNA polymerase 4/13 7.3 5.75 × 10^{-2} 4.79 × 10^{-1} | 7.3 5.75 × 10^{-2} 4.79 × 10^{-1} |                   |
|                                  | hvo00770 Pantothenate and CoA biosynthesis 3/17 5.5 9.31 × 10^{-2} 5.53 × 10^{-1} | 5.5 9.31 × 10^{-2} 5.53 × 10^{-1} |                   |
at high temperature but even more strongly upregulated under low salt stress conditions. All three subunits of the thermo-some are upregulated at high temperature, consistent with previous observations in Haloflexax and other archaea.[8,13,24,37–39] A family of proteins that we found to be specifically impacted by high temperature is the UspA protein family. Haloflexax volcanii has 37 members of this family, of which 24 were significantly upregulated at high temperature and only three were significantly regulated under other stress conditions but not by high temperature. The prototype for this large protein family is E. coli UspA (universal stress protein A); the six UspA paralogs have been partially characterized, and some of these are required for survival under different stress conditions.[40] One UspA protein was also found to be temperature-induced in Halobacterium.[41]

The clusters with proteins that are up- or downregulated upon temperature stress were subjected to enrichment analysis, which led to the identification of several significantly enriched KEGG pathways (Table 1). One of the five enriched pathways summarizes general metabolic charts (biosynthesis of secondary metabolites). Enrichment analysis puts the focus on DNA replication, where six core replication proteins are downregulated by high temperature (Figure S7, Supporting Information). The sulfur relay system is enriched but does not show a coherent regulation pattern. However, also in focus are ribosomal proteins, many of which (24) were downregulated at high temperature (Figure S8, Supporting Information). However, several ribosomal proteins deviated from this paradigm (see above).

Two biosynthetic pathways are enriched. To allow for a more in-depth analysis, we decided to color KEGG maps, with each of the six major clusters having assigned a specific color (plus gray for the remainder of significantly regulated proteins). Inspection of colored KEGG maps is highly efficient and thus we extended our analysis to cover the complete set of metabolism-related KEGG maps. In total, we detected four pathways with coordinate regulation due to temperature stress. Coordinated regulation of a majority of proteins that belong to the same pathway has, to the best of our knowledge, been rarely reported for archaea. De novo arginine biosynthesis enzymes decrease in relative amount with increasing temperature. We found five enzymes upregulated at low temperature (cluster 1983) and three enzymes downregulated at high temperature (cluster 1979) (Figure S9, Supporting Information). For the pathway porphyrin and chlorophyll metabolism, overall enrichment was only slight. The complete pathway of de novo cobalamin biosynthesis, however, was downregulated at high temperature (Figure 4A; see Figure S10, Supporting Information for the original KEGG map). The slightness of enrichment reported by statistical analysis is probably due to the many other pathways (HEME and chlorophyll) being present on the same KEGG map. Starting at sirohydrochlorin, the branch point between heme and cobalamin biosynthesis in Haloflexax, seven of the 11 enzymes that generate cobyrinate a,c-diamide belong to cluster 1979. The cobalochelatase reaction was recolored based on a curated annotation of this biosynthetic pathway (for details see Text S2, Supporting Information). Two no-colored reactions (EC 2.1.1.195 and 1.3.1.106) reflect pathway gaps. Candidates for closure of these pathway gaps are genes with unassigned function, which are encoded within the cobalamino cluster (e.g., HVO_B0052 and HVO_B0055). These were similarly downregulated.

Upon closer inspection, two KEGG maps showed regulation of specific pathways that are embedded in a wider context, thus precluding their identification as statistically enriched. Seven of the 11 enzymes of de novo purine biosynthesis that convert phosphoribosyl-pyrophosphate (PRPP) into IMP are downregulated at high temperature (cluster 1979) (Figure 4B; see Figure S11, Supporting Information for the original KEGG map). For details on the remaining enzymes see Text S3, Supporting Information. The other is tryptophan biosynthesis (Figure 4C, see Figure S12, Supporting Information for the original KEGG map). Here, all enzymes for conversion of chorismate into tryptophan are classified into cluster 1978 (orange, high-temperature up, low-temperature down). This is only evident from the recolored KEGG map based on detailed curation (see Text S4, Supporting Information).

Looking for detail beyond pathway enrichment, we inspected the 20 most strongly up- or downregulated proteins assigned to high-temperature stress clusters (Table S2, Supporting Information). The protein showing the highest upregulation is of unknown function (HVO_0575). The second-highest upregulation (HVO_1599) is a very short (49 aa) conserved hypothetical protein, again with no indication on its function. These observations highlight that many H. volcanii proteins are still of unknown function, with the “small proteome” being severely understudied. Both facts together currently hamper the comprehensive analyses of proteome and transcriptome data.

Among the most strongly downregulated proteins at high temperature is the S-layer glycoprotein (eightfold). Further studies are required to investigate a) if the cell surface is altered under these conditions, replacing the 2D protein array by a distinct surface structure, b) if posttranslational modifications like altered N-glycosylation patterns are responsible for this result, or c) if technical reasons (e.g., increased protein shedding) are responsible for this effect. It should be noted in this context that AglD (HVO_0798), which adds the terminal hexose to the canonical N-glycan of the cell surface glycoprotein, is also strongly downregulated at high temperature.

The most strongly upregulated protein at low temperature (20-fold) is archaellin A1 (HVO_1210, ArlA1, previously FlgA1). ArlA2 (FlgA2, HVO_1211) was not detected in our dataset, potentially due to a high degree of similarity to ArlA1. Also, strongly upregulated at low temperature (14-fold) was tryptophanase (TnaA, HVO_0009), which metabolizes Trp and thus reduces its concentration. This is consistent with the downregulation of Trp biosynthesis at low temperature (see above).

3.4. Specific Salt Stress Responses

We next looked at specific salt stress responses. Although low salinity did not result in unique stress response patterns beyond the universal stress response on the protein expression level, high salinity produced a set of 145 significantly downregulated proteins (Figure S6, Supporting Information, Cluster 1980). Again, we employed enrichment analysis to identify biological processes specifically involved (Table 1). Surprisingly,
neither KEGG enrichment analysis nor manual inspection of the cluster’s proteins produced a clear picture on high-salinity adjustments. Significant enrichment was observed for a general metabolic chart (metabolic pathways) and the RNA polymerase pathway (see Figure S13, Supporting Information). Among the other two enriched pathways, enzymes assigned to pantothenate and CoA biosynthesis were scattered. In the other, folate biosynthesis, the assigned enzymes were actually unrelated to the named metabolic pathway but are involved in queuosine biosynthesis, a tRNA modification that may be reduced under high salt conditions. In addition, three strongly upregulated proteins at high salt (two of which are uniquely regulated) belong to the histidine utilization cluster (hutUGIH, HVO_A0559 to HVO_A0562), which converts histidine into glutamate. Contrary to expectations, high salinity did not result in marked expression changes of proteins associated with the outer cell membrane, indicating that H. volcanii is generally already well adjusted to changing salinity levels in this regard.

3.5. General Observations

Table S2, Supporting Information lists the most strongly up- and downregulated proteins under each of the four stress conditions (high and low temperature, high and low salt). Taken together with the results of the “global view” taken by pathway enrichment analysis, several of general observations can be drawn from these data: i) most of the listed proteins are highly regulated under more than one condition. This is consistent with the large size of the universal stress response clusters (up: 1977/1981, 134 significant proteins; down: 1976, 245 significant proteins), indicating that responses to different stresses are overall highly similar, or even universal. The only notable exceptions were cluster 1978 (148 significant proteins) that contains proteins that show a continuous increase in protein concentration with increasing temperature, and cluster 1980 (145 significant proteins) that contains proteins with reduced expression under high salinity. ii) Only very few strongly regulated proteins are uniquely regulated, not being significantly regulated under any of the other three conditions.
stress conditions. This may, however, be a consequence of the extreme depth of the analysis (2034 proteins significantly regulated under at least one stress condition) and the very high reproducibility of SWATH-MS quantification. iii) A significant fraction of the strongly regulated proteins is currently only generally annotated, that is, their molecular activity is yet unresolved, or they are still considered “hypothetical” or “conserved hypothetical.” This “gray proteome” is not so much an indication of insufficient annotation[17] but rather of a demand for additional biological experimentation, which significantly lags behind progress in genome sequencing. As a consequence, functional enrichment analyses, for example, by KEGG pathways are currently not yet able to fully leverage the depth of global proteome profiling datasets for moderately characterized organisms such as H. volcanii. Less stringent “global” annotation approaches such as arCOGs, however, were not informative. iv) In several cases, our observations pointed to global pathway annotations by KEGG that were not fully consistent with the current knowledge on the function of individual proteins in the pathway, which again calls for further experimentation on archaean biochemistry.

4. Concluding Remarks
In this study, we employed label-free profiling by SWATH-MS to observe global proteome changes in H. volcanii under low and high salinity, and low and high temperature stress conditions. Using a simple, but highly reproducible single-shot protocol, we obtained high proteome coverage with regard to protein identification, quantitation, and differential expression. Functional analysis for enrichment of KEGG pathway annotations as well as manual biochemical annotation revealed that the examined stresses triggered largely similar expression changes that affect proteins participating in many cellular pathways. The number of proteins affected by only a single, specific stress response was low by comparison, indicating that these rather reflect cellular fine-tuning than fundamentally different adjustments. Functional analysis was hampered by a lack of functional annotation for a significant number of differentially expressed proteins. More data about this “gray proteome” are required to allow the comprehensive evaluation of proteome and transcriptome data. Following progress in the definition or functional annotation of genes, it will be worthwhile to re-interrogate our global “digital map” of the H. volcanii stress proteome in the light of new information.

Supporting Information
Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements
Z.J. and B.S. contributed equally to this work. A.M. designed and supervised the overall study. B.S. grew the cells and isolated the proteins. C.L. designed and evaluated all mass spectrometry analyses. Z.J. and K.S. performed mass spectrometry experiments. C.L. and Z.J. performed statistical evaluation. F.P., A.M., and C.L. participated in data analysis and annotation. C.L., F.P., and A.M. wrote the manuscript. H.U. critically revised the manuscript. C.L. would like to thank Monika Raabe and Annika Kleinert (MPIbpc, Göttingen) for expert technical assistance. Work in AM’s laboratory was funded by the DFG (MA1538/24-1) in the framework of SPP2002.

Conflict of Interest
The authors declare no conflict of interest.

Keywords
archaea, mass spectrometry, proteome, stress response, SWATH

Received: December 31, 2018
Revised: August 27, 2019
Published online: October 1, 2019

[1] J. A. Leigh, S. V. Albers, H. Atomi, T. Allers, FEMS Microbiol. Rev. 2011, 35, 577.
[2] M. F. Mullakhanbhui, H. Larsen, Arch. Microbiol. 1975, 104, 207.
[3] M. Mevarech, R. Werczerberger, J. Bacterial. 1985, 162, 461.
[4] A. L. Hartman, C. Norais, J. H. Badger, S. Delmas, S. Haldenby, R. Madupu, J. Robinson, H. Khouri, Q. Ren, T. M. Lowe, J. A. Maupin-Furlow, M. Pohlschroeder, C. Daniels, F. Pfeiffer, T. Allers, J. A. Eisen, PloS One 2010, 5, e9605.
[5] M. Torreblanca, F. Rodriguez-Valera, G. Juez, A. Ventoza, M. Kamemura, M. Kates, Syst. Appl. Microbiol. 1986, 8, 89.
[6] C. J. Daniels, A. H. Z. McKee, W. F. Doolittle, EMBO J. 1984, 3, 745.
[7] C. Ferrer, F. J. Mojica, G. Juez, F. Rodriguez-Valera, J. Bacterial. 1996, 178, 309.
[8] Y. P. Kuo, D. K. Thompson, A. S. Jean, R. L. Charlebois, C. J. Daniels, J. Bacterial. 1997, 179, 6318.
[9] F. J. Mojica, E. Cisneros, C. Ferrer, F. Rodriguez-Valera, G. Juez, J. Bacterial. 1997, 179, 5471.
[10] I. M. Karadzic, J. A. Maupin-Furlow, Proteomics 2005, 5, 354.
[11] K. A. Bidle, P. A. Kirkland, J. L. Nannen, J. A. Maupin-Furlow, Microbiology 2008, 154, 1436.
[12] H. D. Shukla, Proteome Sci. 2006, 4, 6.
[13] R. R. Weng, H. W. Shu, S. W. Chin, Y. Kao, T. W. Chen, C. C. Liao, Y. G. Tsay, W. V. Ng, OMICS: J. Integr. Biol. 2014, 18, 65.
[14] P. T. Van, A. K. Schmid, N. L. King, A. Kaur, M. Pan, K. Whitehead, T. Koide, M. T. Facchiotti, Y. A. Goo, E. W. Deutsch, D. J. Reiss, P. Mallick, N. S. Baliga, J. Proteome Res. 2008, 7, 3755.
[15] G. Losensky, K. Jung, H. Urlaub, F. Pfeifer, S. Fröls, C. Lenz, Proteomics 2016, 17, 1600111.
[16] P. A. Kirkland, M. A. Humbard, C. J. Daniels, J. A. Maupin-Furlow, J. Proteome Res. 2008, 7, 5033.
[17] M. Cerletti, R. A. Paggi, C. R. Guevara, A. Poetsch, R. E. De Castro, J. Proteomics 2015, 121, 1.
[18] M. I. Costa, M. Cerletti, R. A. Paggi, C. Trötschel, R. E. De Castro, A. Poetsch, M. I. Giménez, J. Proteome Res. 2018, 17, 961.
[19] L. J. McMillan, S. Hwang, R. E. Farah, J. Koh, S. Chen, J. A. Maupin-Furlow, Environ. Microbiol. 2018, 20, 385.
[20] M. Cerletti, R. Paggi, C. Trötschel, M. C. Ferrari, C. R. Guevara, S. Albaum, A. Poetsch, R. De Castro, J. Proteome Res. 2018, 17, 1158.
[21] M. Cerletti, M. I. Giménez, C. Trötschel, C. D’Alessandro, A. Poet- sch, R. E. Castro, A. R. Paggi, Proteomics 2018, 18, 1800116.
[22] C. Liu, L. Mao, X. Zheng, J. Yuan, B. Hu, Y. Cai, H. Xie, X. Peng, X. Ding, MicrobiolOpen 2019, 8, e00715.
