Zooming in:
from multi-omics to single function in hyperthermophilic archaeon *Sulfolobus solfataricus*

Pawel Sierocinski

Thesis committee

Promotor

Prof. Dr J. van der Oost
Personal chair at the Laboratory of Microbiology

Co-promotor
Prof. Dr W. de Vos
Distinguished Professor Microbiology

Other members

Prof. Dr W van Berkel Wageningen University & Research
Dr SJJ Brouns UHD Delft University of Technology
Dr D Swarts UD Wageningen University & Research
Dr Nico Claassens, PostDoc Max Planck Institute of Molecular Plant Physiology, Potsdam

The research was conducted under the auspices of the Graduate School VLAG (Advanced studies in Food Technology, Agrobiotechnology, Nutrition and Health Sciences).

Zooming in: from multi-omics to single function in hyperthermophilic archaeon *Sulfolobus solfataricus*

Pawel Sierocinski

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Monday 26th August 2019
at 11 a.m. in the Aula

Pawel Sierocinski
Zooming in: from multi-omics to single function in hyperthermophilic archaeon *Sulfolobus solfataricus*, 210 pages

PhD thesis, Wageningen University, Wageningen, the Netherlands (2017) With references, with summary in English

ISBN 978-94-6343-974-9
DOI https://doi.org/10.18174/476997

Table of contents

Chapter 1

Introduction and Outlook

Pawel Sierocinski

Till the 1960s life as we knew it occupied a comfortable niche that overlapped with the human temperature range. Even though the earliest reports of life at temperatures above 80°C were published back in 1897 (Davis 1897) they were generally discarded as artefacts. Temperatures reaching over 80°C were considered too high for any living creature to survive, let alone to thrive. In 1963 Kempner speculated, based on the analysis of hot springs in Yellowstone National Park that 73°C is the upper temperature limit for life (Kempner 1963). This paradigm shifted soon thereafter when Brock and Freeze managed to isolate and cultivate *Thermus aquaticus,* a bacterium with a temperature range from 40°C to 79°C (Brock & Freeze 1969). This opened the doors for further investigation of environments previously assumed hostile for life, resulting in the discovery of a great diversity of thermophiles and hyperthermophiles, both marine and terrestrial.

Hyperthermophiles, defined as organisms that thrive at elevated temperatures with optimal growth at or above 80°C (Stetter 2006), occupy diverse sets of environments – from the submarine black smokers, though terrestrial and marine hot springs to high temperature compost heaps. This diversity of ecosystems allows for multiple life strategies. Hyperthermophiles include both aerobic and anaerobic life forms. Most are autotrophic, using hydrogen as the electron donor and a range of electron acceptors, including CO_2, sulphur, and nitrate. Their autotrophy is not obligatory and a majority has been classified

as opportunistic heterotrophs capable of metabolising a wide range of organic compounds either by aerobic respiration or fermentation.

One key feature that all the hyperthermophiles share is the presence of the reverse gyrase enzyme (Forterre *et al.* 1995). Although it can be occasionally found in regular thermophiles as well, it has never been observed in mesophiles. The reverse gyrase enzyme is present in bacterial hyperthermophiles but it is of archaeal origin, suggesting it was evolved in HT Archaea and subsequently transferred to HT Bacteria through a horizontal gene transfer event shortly after the two domains split. The enzyme is responsible for positive supercoiling of DNA and knock out strains of hyperthermophiles lacking it are viable, but thermosensitive (Atomi *et al.* 2004), most likely causing deficient strains to lose competition with the gyrase possessing organism in the hyperthermophilic conditions.

Archaea

Even though hyperthermophiles share multiple similarities, they span two groups separated by the oldest rift in phylogenetic history of life, i.e. the split between the Bacteria and the Archaea. Their similarities, from both morphological and physiological perspective, caused that initially they were all classified as bacteria. For example, *Sulfolobus solfataricus*, the subject of this thesis, has been initially classified as an atypical member of the genus *Pseudomonas*. It was only in 1977 when Carl Woese established a novel method of determining phylogeny that was based on similarities of conserved regions of the ribosomal genes (Woese & Fox 1977). His results suggested that, although morphologically identical, some microorganisms show a genetic divergence indicating an ancient split from the rest of the prokaryotes.

Initially it was the methanogens that did not fit in the prokaryotic puzzle, but soon after it turned out that the majority of thermophiles cluster closer to the methanogens than to bacteria. The idea did not catch quickly in the morphology dominated field of phylogenetics, but in 1990 it was proposed, again by Woese, that life should be reorganised into three domains (Woese *et al.* 1990) and since then this classification has become a new paradigm on how the life on earth has evolved, and should be organised.

Archaea are distinct from the rest of the nucleus-free life not only due to their 16S RNA sequences. They have a unique composition of their cell membrane, consisting of ether-linked isoprenoid lipids (Kates 1977) a trait that allows them to thrive in environments where other microorganisms fail (Gliozzi *et al.* 2002). They are the only group of organisms that can perform metabolic processes that are key to the nutrient cycling on our planet. Main example is methanogenesis, a key process in anaerobic conditions that allows removal of acetate, CO_2 and hydrogen thus protecting the microbial communities from accumulation of harmful by-products of fermentation. Annually 500 billion tons of methane are produced by methanogens making it a truly planetary scale process fully facilitated by Archaea (Conrad 2009). Archaea are also responsible for recently discovered processes of anaerobic methane oxidation (Raghoebarsing *et al.* 2006) and anaerobic ammonia oxidation (Schmidt *et al.* 2002) that play a key role in the stability of nutrient cycles of the planet.

One of the most interesting features of Archaea is their DNA processing machinery. Even though they are similar to prokaryotes in terms of the metabolism, their processing of DNA resembles the one of eukaryotes. They have similar regulatory proteins and sequences, similar tRNA genes and, at

least in some cases, their replication starts, unlike bacterial one, from multiple origins of replication. This has led to Archaea becoming a model system for preliminary studies of eukaryote replication, transcription and translation, combining a relatively homologous mechanism and the ease of growing and manipulating the genetics in comparison with the eukaryotes. This similarity resulted with a novel concept of the origin of Eukaryotic cell as a fusion between an archaeon and a prokaryote, changing the tree of life into a ring-like structure (Rivera & Lake 2004). Archaea are also a key element of studying the origins of life on the planet, with multiple hypotheses suggesting that it might have required hot environments for the first cellular replicators to kick off.

Archaea were originally divided into two major kingdoms: *Euryarchaeota*, containing halophilic Archaea, methanogens and some of the (hyper)thermophiles and *Crenarchaeota*, harbouring most of the known (hyper)thermo (acido)philes. This has been challenged by the more recent discoveries of multiple novel groups of Archaea, including *Nanoarchaeota, Thaumarchaeota, Lokiarchaeota* and *Korarchaeota*, which makes the current phylogeny of Archaea a work in progress (Huber *et al.* 2003; Brochier-Armanet *et al.* 2008; Petitjean *et al.* 2015; Spang *et al.* 2015). The new discoveries, greatly facilitated by the cheaper sequencing technology, also put a dent in the long held belief that Archaea are mainly involved in the extreme environments. Archaeal sequences are ubiquitous in all the sampled environments, and make up a significant part of mesophilic strata. The initial abundance of extremophiles was probably an artefact; the extreme environments where Archaea are predominant were disproportionally sampled, while mesophilic Archaea were too rare to be readily discovered and cultivated. Yet since typical mesophilic environments are vastly bigger than

the extreme ones, a large diversity of Archaea from those environments still outnumbers the extremophiles.

Thermophiles and *Sulfolobus solfataricus*

That said, thermophiles are still one of the hallmarks of Archaea and are among the best-studied organisms in the domain. They are a window towards the limits of life on the planet and an excellent model for ecological studies due to the relative simplicity of the communities they occupy. Thermophiles have been of great use in science, providing key tools for the genetic engineering revolution of the 1980s and 1990s like the DNA polymerase enzymes required for the Polymerase Chain Reaction obtained from a variety of thermophiles and hyperthermophiles, *T. aquaticus* (Taq) and *P. furiosus* (Pfu) being notable examples. Also the industry embraced thermostable enzymes using their unique properties in the processes where high temperature is easily achievably at low cost, due to high capability of heat recycling or the process already being run at high temperatures. Starch hydrolysis, where thermophilic enzymes complement a high-temperature industrial process to increase its efficiency and reduce costs is an example of such use.

One of the early terrestrial isolates was *Sulfolobus solfataricus* . The representatives of the genus *Sulfolobus*, first discovered and described by Brock in 1972 (Brock *et al.* 1972), have been found in several locations worldwide in muddy, aerobic hot springs characterised by low pH (1.5-3.5) and high temperature (76–90°C). The representatives of the genus were a predominant group in examined environments, allowing for a direct identification by microscopy straight from the environmental samples.

Sulfolobus discoveries

After its discovery, *Sulfolobus* quickly became a model for studying thermophilic and acidophilic Archaea. As usual with the models, *Sulfolobus* possessed qualities that made it a promising subject of studies. It is easy to grow in the lab setup (Brock *et al.* 1972), it is growing relatively well in a broad range of conditions and on a broad range of substrates, suggesting interesting regulatory features (Grogan 1989). Furthermore quickly it became feasible to grow it in larger volumes, showing promise to use it as an industrially relevant strain (Park & Lee 1997; Schiraldi *et al.* 1999). This paired up with its very interesting biology. *Sulfolobus* uses different central carbon metabolism compared to bacteria, and furthermore it has two competing CCM pathways (Danson 1989). Equally important, *Sulfolobus* has a whole host of viral parasites with unique set of features, capable of enduring thermoacidophilic conditions (Prangishvili *et al.* 2001; Lipps 2006). Research into mobile genetic elements of *Sulfolobus* has led to discovery of unique virus and plasmid families (Zillig *et al.* 1998; Greve *et al.* 2004), unknown in other species, and has provided tools for genetic engineering.

In addition, it is worth mentioning the interesting early history of archaeal research. After Woese showed that Archaea are a different domain of life from Bacteria (Woese & Fox 1977; Woese *et al.* 1990), other researchers speculated that a fusion of archaeal and bacterial genomes was the ancestor of eukaryotic cells (Cavalier-Smith 1987) based on the discoveries related to the publication of the first thermophilic archaeon genome sequencing project on *Methanococcus jannaschii* (Bult *et al.* 1996a). Later, this model resulted in a variation, the ring of life hypothesis, again with a bacterial-archaeal fusion

as the origin of eukaryotes (Rivera & Lake 2004). This was corroborated by the unique features of Archaea in terms of their transcription and translation mechanisms, much more similar to eukaryotes than to their bacterial counterparts (Reiter *et al.* 1990; Qureshi *et al.* 1997; Blombach *et al.* 2015). This hypothesis was also supported by the differences in bacterial and archaeal cell cycle, in particular in the presence of multiple origins of replication, characteristic for eukaryotes but not bacteria (Lundgren *et al.* 2004; Robinson *et al.* 2004a). Recently, metagenome analysis of hydrothermal vents has revealed the existence of new archaeal phyla *Lokiarchaeota* in which the genes were present that encoded many typical eukaryotic features - this has created major excitement, as this may be the missing link corresponding to the ancestor of the first eukaryotic cell (Spang *et al.* 2015).

The early adaptation of *Sulfolobus* as the model for studying both archaea and thermophiles led to one more important milestone. *Sulfolobus* was one of the first thermophilic archaea to be fully sequenced in 2001 (She *et al.* 2001b) after *Methanococcus jannaschii* (Bult *et al.* 1996b) and *Archaeoglobus fulgidus* (Klenk *et al.* 1997) . The knowledge of genetic background of the unique features of *Sulfolobus* lead to increased interest in the organism and new discoveries related to its transcriptome (Lundgren *et al.* 2004; Snijders *et al.* 2006), genome regulation (Brinkman *et al.* 2002; Peeters *et al.* 2004) or proteomics (Chong & Wright 2005; Barry *et al.* 2006).

The rise of those new approaches has led to a better description of *Sulfolobus* physiology, including its pentose metabolism (Brouns *et al.* 2006), central carbon metabolism (Lamble *et al.* 2004; Ettema *et al.* 2008), or impact of stressful conditions like UV (Fröls *et al.* 2007) or heat shock (Tachdjian & Kelly

2006b). *Sulfolobus* has been also at the forefront of one of the most important discoveries in recent years – the CRISPR-Cas system (Peng *et al.* 2003).

But most of the research done before the start of the research described in this thesis (2007) has focused on individual aspect of the cell – be it genome, transcriptome or proteome. There was a lack of multi-omics approaches that would let us consolidate those findings and use all the mentioned techniques in order to test old hypotheses and use the results to test new ones. And to use such approach to make predictive models and search for new potential questions related to *S. solfataricus*.

An opportunity of combining expertise of multiple research groups became possible in the course of the Systems biology in MicroOrganisms (SysMO) initiative. The philosophy of SysMO projects was to explore scientific questions relevant to basic biology of microorganisms and at the same time to develop standard research tools for the organisms used. Key element was enhancing the cooperation between wet lab researchers and modellers thus establishing community standards improving the systems biology research. Systems biology has been anecdotally prone to miscommunication between lab and *in silico* researchers leading to incomplete or confusing models or models based on input data that was not fit for purpose. The ambition of SysMO was to construct a platform allowing researchers sharing data used for the models in such form that the shortcomings of mixing multiple disciplines can be overcome – including a database for such data that adheres to the data sharing standards. One of the involved consortia, SulfoSys included 11 institutions in 6 European countries, and has been set up in order to combine the lab work and modelling.

SulfoSys project has been set up as an attempt to use a high throughput methodology and employ it in order to produce a detailed metabolic model of the *S. solfataricus* Central Carbon Metabolism (CCM) at varying temperatures. Looking at different temperatures, was aimed at elucidating the roles of branched Entner–Doudoroff) (ED) catabolic pathway and the gluconeogenic Embden–Meyerhof–Parnas (EMP) pathway to test how *S. solfataricus* can be so extremely robust in its viable condition range. Its maximum growth rate varies by only factor of 2 between its optimum (78°C) and minimum growth temperature of 65°C (Grogan 1989). Furthermore, there are very small differences in its growth rate over a wide range of pH values – only 30% difference between pH 2.0 and pH of 6.0. Such robustness suggests regulatory mechanisms that compensate between different conditions and allow switching between pathways and their branches according to the needs. This is most likely an adaptation to life in steep condition gradients. Moreover, in the conditions in which *S. solfataricus* grows, some reactions to occur spontaneously in a part of its temperature range, while requiring enzymes in others. Example of such reaction can be found in the pentose oxidation pathway, where the step of converting D-arabinonolactone to D-arabinonic acid can occur spontaneously, while it does require enzymatic conversion in the mesophilic organisms.

Sequencing projects of various thermophiles, coupled with biochemical research allowed us to gain some insight in their lifestyle, however the uniqueness of their proteins along with particularities of the metabolic pathways used has made the research a significant challenge. One of the solutions proposed to tackle this bottleneck was the use of high throughput methods for analysis of the proteome and transcriptome of those organisms. The analysis of the whole network of transcripts within organisms allows not

only a unique insight in the cell's physiology during growth. Co-transcription of certain genes also greatly helps in identification of missing links in the cell metabolism and can pinpoint possible functional homologies as well as regulatory mechanisms within a genome.

Before the advent of cheap next generation sequencing (NGS) the choice in techniques available and organisms suitable for such an approach was limited. One of the pioneering methods to analyse transcriptomics on a full genome scale was the microarray technology. The origins of the technology lie in the mid-1970s, when Grunstein and Hogness developed a colony hybridisation technique allowing the detection of targeted DNA by hybridising it with a radioactively labelled probe complementary to the sequence of interest . This forefather of microarrays was relatively primitive compared to the later designs but worked under the similar principle: hybridisation of a probe and target DNA where one was labelled and detecting the signal. The technology that allowed mass printing of DNA oligonucleotides on glass chips allowed construction of first full genome arrays containing full set of genes from a previously sequenced organism. The use of microarray technology has been quickly adapted within thermophiles (see chapter 2) and yielded further insight into the cellular mechanisms as well as metabolism of hyperthermophiles. First hyper-thermophile microarray experiments have been conducted on *P. furiosus*, an anaerobic deep-sea archaeon in 2001. *S. solfataricus* full genome array followed soon after with experiments that confirmed that unlike bacteria, Archaea have multiple origins or replication, putting their DNA processing closer to that of Eukaryotes. Further experiments using the same system provided insight in the unique pathways governing the metabolism of arabinose in *S. solfataricus.*

The data obtained using high throughput methods does not only shed light on the biology of the cell and influence of external conditions on it. It also is a very good input source for the biological modelling, due to large and uniform datasets that allow making very accurate and predictive models. The models allow discovering biological factors that otherwise might escape detection using traditional methods. The discrepancies between the model and results of in vitro or in situ experimentation point towards a factor unaccounted before, that plays a key role in the studied process. Parameter fitting procedures combined with wet lab experimentation can lead to the detection of such factors . But findings still lean on the human factor. It is key to have the idea and to avoid the trap of post-data collection hypothesizing in order to distinguish between the valid findings and pure artefacts, which are very likely when analysing the massive datasets provided by the high-throughput datasets.

Outline of this Thesis

In order to develop hypotheses presented in the introduction, we have undertaken several experiments, which we describe in the following chapters. The chapters have been published or submitted for publication, however for the purposes of this thesis, some of them have been adapted to make sure the thesis is a coherent stand-alone publication. Some of the chapters have been furthermore updated in order to keep them in accordance with most recent findings in the field.

Chapter 2 shows the history of transcriptomics research in thermophiles and hyperthermophiles. We look at a wide range of research questions that

transcriptome allows answering and showcase some of the most influential research on thermophile transcriptomics in last decades.

Chapter 3 describes the multi-omics toolbox we developed for *S. solfataricus* research. Standardisation of methodology is a key aspect when it comes to combining experimental work and modelling, we propose a complete collection of methods together with results showing the *S. solfataricus* Central Carbon Metabolism in shifting temperatures.

Chapter 4 focuses on the poorly studied and potentially important membrane proteome of *S. solfataricus*. We show that better methods allow us elucidating the composition of membrane proteome, and from there, its function. We detect and increased the number of membrane peptides and show a differential protein pattern after cultivation at optimal and at suboptimal temperatures.

Chapter 5 addresses the transcription regulation within the *S. solfataricus* genome. Based on the results from previous chapters we find a putative regulatory sequence responsive to the temperature change and confirm the transcription patterns using RT-qPCR.

Chapter 6 looks at experimental evolution as a tool for confirming adaptive traits in *S. solfataricus*. Looking at sub and super-optimal growth temperatures we try to elucidate whether the selection in fluctuating conditions is selecting for a more generalist growth pattern as opposed to constant selective pressure, which should result in selection for specialists in a given condition.

Chapter 7 presents the general conclusions of this thesis focusing on multi approach strategy of looking at biological systems. We try to show how combination of complementary techniques driven by an overreaching hypothesis can aid in finding answers unattainable otherwise. We focus on the links between the previous chapters and propose how findings from one experiment can drive further research.

Chapter 2

Hot transcriptomics

Pawel Sierocinski*, Jasper Walther*, John van der Oost

*Authors contributed to this chapter equally. This is a modified version of the 2011 “Hot Transcriptomics” review by Walther, Sierocinski and van der Oost published in Archaea. This version was updated by the findings between 2011 and 2018 in the field of hypothermic transcriptomics.

Abstract

DNA microarray technology allows for a quick and easy comparison of complete transcriptomes, resulting in improved molecular insight in fluctuations of gene expression. After emergence of the microarray technology about a decade ago, the technique has now matured and has become routine in many molecular biology laboratories. Numerous studies have been performed that have provided global transcription patterns of many organisms under a wide range of conditions. Initially, implementation of this high-throughput technology has led to high expectations for groundbreaking discoveries. Here an evaluation is performed of the insight that transcriptome analysis has brought about in the field of hyperthermophilic archaea. The examples that will be discussed have been selected on the basis of their impact, in terms of either biological insight or technological progress.

Thermophiles

Forty years ago it was generally accepted that life was not possible at temperatures higher than 60°C. In 1969, however, Brock and Freeze discovered that the upper temperature limit goes as high as 75°C when

microorganisms were isolated from thermal springs in Yellowstone National Park . The pioneering work of Brock set the stage for further exploration of a wide range of thermal ecosystems. Numerous microorganisms defined as thermophiles have since been found to thrive optimally between 50 and 80°C, but also many appeared to have their optimal temperature for growth from 80°C to well above 100°C, the hyperthermophiles. Recently it has been shown that some archaea can endure temperatures as high as 122°C and even proliferate in such conditions. Although there are several bacterial representatives in the group as well, most of the known hyperthermophiles belong to the archaea.

Thermophilic organisms can be found in water-containing geothermally heated environments. These volcanic ecosystems are mainly situated along terrestrial and submarine fracture zones where tectonic plates are converging or diverging. The terrestrial biotopes of (hyper)thermophiles are mainly aerobic, sulphur containing solfataric fields with temperature as high as 100°C (depending on the altitude) and the pH in a dual range: either acidic (values from below zero to 4.0) or neutral to slightly alkali (7.0–9.0). The marine biotopes for (hyper)thermophiles consist of different hydrothermal systems ranging from shallow to abyssal depths. Temperatures in those anaerobic environments can range up to 400°C and the pH is usually in the range of 5.0 to 8.5.

Progress in culturing thermophilic archaea and in the revolution of DNA sequencing technology has resulted in a rapidly increasing amount of (meta)genomic data on these extreme microorganisms. This has not only led to the discovery of robust biocatalysts but also to fundamental insight into (i) physiology: including unique metabolic enzymes, pathways, and regulation (ii)

biochemistry: the molecular basis of thermostability of biomolecules and (iii) phylogeny: theories on the evolution of the eukaryotic cell .

The first complete genome analysis of an archaeon, *Methanocaldococcus jannaschii ,* was a big step towards confirmation of the monophyletic position of the archaea, with respect to the bacteria and the eukaryotes. In addition, archaea appeared to possess a bacterial-like compact chromosomal organization with clustering of genes as polycistronic units (operons), and with only few interrupted genes (introns). Moreover, the archaeal systems that drive the flow of genetic information (transcription, translation, replication, DNA repair) generally correspond to the core of the eukaryal counterparts. These initial observations of bacterial-like “information storage” and eukaryal-like “information processing” have been confirmed by the analyses of subsequently sequenced hyperthermophilic model archaea: the euryarchaea *Pyrococcus* spp. (*P. furiosus, P. abyssi, P. horikoshii*) as well as the crenarchaea *Sulfolobus* spp. (*S. solfataricus, S. tokodaii, S. acidocaldarius*) (Makarova & Koonin 2003). The comparative analysis of the genome of the hyperthermophilic bacterium *Thermotoga maritima* to *Pyrococcus furiosus* (both isolated from shallow thermal vents at the same beach (Volcano, Italy)) led to the conclusion that horizontal (or lateral) gene transfer substantially contributes to the apparent high degree of genome flexibility . In addition, the comparison of closely related species (*P. furiosus, P. abyssi, P. horikoshii*) revealed a high degree of genome plasticity. It was also proposed that the lateral gain as well as the loss of genes is a modular event . Horizontal gene transfer has also been proposed to explain the relatively high degree of homology between genomic loci of the euryarchaeon *Thermoplasma acidophilum* and the crenarchaeon *S. solfataricus*, phylogenetically distant archaea, that inhabit the same environmental niche (65–85°C, pH 2.0). The

Sulfolobus-like genes in the *T. acidophilum* genome are clustered into at least five discrete regions, again indicating modular recombination of larger DNA fragments.

After establishing a genome sequence, comparative genomics analyses are performed to assign potential functions for the identified open reading frames. In the majority of the studied prokaryotic genomes, the fraction of hypothetical and conserved hypothetical genes amounts to 40–60% of the coding regions . Hence, one of the main challenges of the postgenome era still is to improve the functional annotation of genes by integrating classical approaches (physiology, biochemistry, and molecular genetics) with genomics-based high-throughput approaches (comparative, functional, and structural genomics). Obvious targets of comparative and functional analysis of archaeal genomes are the numerous missing links in metabolic pathways as well as the largely unknown regulatory systems with either eukaryal or bacterial characteristics [7],[8].

Archaeal Transcriptomics

DNA microarrays have initially been established as high-throughput functional genomics tools to study eukaryotic and bacterial model systems. Initial assumptions suggested that microarray can be used as a general research tool ; however after more than a decade of experience it should be concluded that the application of microarray has its pros and cons. The choice of possible microarray approaches ranges from rather simple layouts comparing two states, to relatively complicated multistate experimental hybridization schemes. The development of appropriate analytical methods has appeared to be a crucial requirement to enable analysis of the more complicated

experimental designs and to allow drawing conclusions from relatively small differences in expression profiles. Consequently, high-quality microarray analyses not only require careful experimentation (cultivation, nucleic acid analysis, hybridization) but also state-of-the-art data processing. This has allowed for the high-resolution analysis of time course experiments and of multi-condition experiments . In most recent studies, the majority of DNA microarrays are used either (i) as a pilot experiment that should provide leads for further investigations , (ii) as a refinement tool to confirm previous gene expression studies , or (iii) as one of many high-throughput methods to be integrated in a systems biology analysis . Below, selected examples of transcriptome analyses of (hyper)thermophilic archaea are described in more detail. Selection is has been based on technological and/or scientific impact. An overview of archaeal transcriptome studies can be seen in Table 1.

Table 1. A list of different archaeal transcriptome publications. This table shows that transcriptome studies are mostly done to elucidate metabolic processes or the behaviour of different Archaea in stress situations. The publications are sorted by subject. Per subject the publications are sorted by year of publication. We included some environmental studies because they give a crucial insight in the ecological function of archaeal species. We excluded some of these publications because in our view they focused more on non-archaeal species, which is a subject not related to this article. The studies referring to thermophiles are in bold. The studies described in this paper in more detail are marked with an asterisk next to the reference.

Species	Experiment aim	Reference
Haloferax volcanii	**Central carbon metabolism**	(Schut *et al.* 2001)*****

Pyrococcus furiosus	Central carbon metabolism	(Baliga *et al.* 2002)
Halobacterium salinarum NRC-1	Anaerobic respiration	(Zaigler *et al.* 2003)
Methanosarcina mazei	Metabolism of methanogenic substrates	(Schut *et al.* 2003)
Sulfolobus solfataricus	Central carbon metabolism	(Müller & DasSarma 2005)
Sulfolobus solfataricus	Pentose metabolism	(Hovey *et al.* 2005)
Methanosarcina barkeri	Methanogen metabolism/methods	(Snijders *et al.* 2006)
Methanosarcina mazei	Nitrogen metabolism and regulation	(Brouns *et al.* 2006)
Pyrococcus furiosus	Starch metabolism	(Culley *et al.* 2006)
Pyrococcus furiosus	Metabolism of elemental sulfur	(Veit *et al.* 2006)
Halobacterium salinarum R1	Adaptation to phototrophy	(Lee *et al.* 2006)
Methanosarcina acitovorans	Acetate and methanol metabolism	(Schut *et al.* 2007)
Environmental array	Ammonium oxidation	(Twellmeyer *et al.* 2007)
Metallosphaera sedula	Electron transport chain	(Li *et al.* 2007)
Methanosarcina	Methanogenesis	(Rich *et al.* 2008)
Pyrobaculum aerophilum	Terminal electron acceptor studies	(Auernik & Kelly 2008)
Thermoproteus tenax	Central carbohydrate metabolism	(Ferry & Lessner 2008)
Halobacterium salinarum R1	Phosphate-dependent behaviour	(Cozen *et al.* 2009)

Halobacterium salinarum NRC-1	Global response to nutrient availability	(Zaparty *et al.* 2008)
Haloferax volcanii	D-Xylose metabolism	(Wende *et al.* 2009)
Methanosarcina mazei	Response to nitrogen availability	(Schmid *et al.* 2009)
Metallosphaera sedula	Auto- hetero- and mixotrophic growth	(Johnsen *et al.* 2009)
Metallosphaera sedula	Bioleaching	(Jäger *et al.* 2009)
Environmental transcriptomics	Cellulolysis and methanogenesis	(Xia *et al.* 2014)
Environmental transcriptomics	Scavenging organic compounds	(Li *et al.* 2015)
Sulfolobus solfataricus	Adaptation to low pH	(McCarthy *et al.* 2015)*
Thermococcus onnurineus	H_2 production	(Lee *et al.* 2016)
	Stress	
Pyrococcus furiosus	Heat shock response	(Shockley *et al.* 2003)
Pyrococcus furiosus	Cold shock response	(Weinberg *et al.* 2005)
Halobacterium salinarum NRC-1	UV irradiation	(McCready *et al.* 2005)
Methanocaldococcus janaschii	Heat and cold shock	(Boonyaratanakornkit *et al.* 2005)
Methanosarcina barkeri	Heat shock and air exposure	(Zhang *et al.* 2006)
Methanocaldococcus janaschii	Pressure stress	(Boonyaratanakornkit *et al.* 2006)
Pyrococcus furiosus	Response to gamma irradiation	(Williams *et al.* 2007)

Methanosarcina mazei	Salt adaptation	(Pflüger *et al.* 2007)
Methanococcus maripaludis	H-limitation and growth rate	(Hendrickson *et al.* 2007)
Halobacterium salinarum NRC-1	Response to change in temperature and salinity	(Coker *et al.* 2007)
Sulfolobus solfataricus	UV irradiation	(Fröls *et al.* 2007)*
Sulfolobus solfataricus; S. acidocaldarius	UV irradiation	(Dorazi *et al.* 2007)
Sulfolobus solfataricus	Heat Shock Response	(Tachdjian & Kelly 2006a)
Halobacterium salinarumNRC-1	UV irradiation	(Baliga *et al.* 2002)
Sulfolobus solfataricus	Oxygen stress	(Simon *et al.* 2009)
Methanococcoides burtonii	Heat stress	(Campanaro *et al.* 2011)
Thermococcus kodakaraensis	Heat stress	(Kanai *et al.* 2010)
Pyrococcus furiosus	Heat stress	(Keese *et al.* 2010)
Sulfolobus solfataricus	Heat stress	(Cooper *et al.* 2009)
Pyrococcus furiosus	Oxidative stress	(Strand *et al.* 2010)
Methanohalophilus portucalensis	Hypo- and Hyper-salt stress	(Shih & Lai 2010)
Thermoanaerobacter tengcongensis MB4	Cold shock	(Liu *et al.* 2014)
Pyrococcus yayanosii	Pressure shock	(Michoud & Jebbar 2016)
Metallosphaera sedula	Heavy metal shock	(Wheaton *et al.* 2016)
	Replication	
Sulfolobus solfataricus; S. acidocaldarius	Origin of replication	(Robinson *et al.* 2004; Duggin *et al.* 2008)
Halobacterium salinarum NRC-1	Cell cycle regulation	(Baumann *et al.* 2007)
Pyrococcus abyssi	Origin of replication	(Matsunaga *et al.* 2007)

Sulfolobus acidocaldarius	Cell cycle	(Lundgren & Bernander 2007)*
	Various	
		(Stralis- Pavese *et al.* 2004)
Environmental array	Methanotroph diversity in landfills	(Hamilton-Brehm *et al.* 2005)
Pyrococci	Genomic DNA hybridization	(Andersson *et al.* 2006)
Sulfolobus solfataricus; S. acidocaldarius	RNA decay	(Xia *et al.* 2006)
Methanococcus maripaludis	Mutant studies	(Lange *et al.* 2007)
Haloferax volcanii	Promoter studies	(Kanai *et al.* 2007)
Thermococcus kodakaraensis	Promotor studies	(Santangelo *et al.* 2008)
Thermococcus kodakaraensis	Archaeal operon prediction	(Dambeck & Soppa 2008)
Haloferax volcanii	Deletion mutant analysis	(Garrido *et al.* 2008)
Environmental array	Detection of acidophilic activity	(Ortmann *et al.* 2008)
Sulfolobus solfataricus	Viral infection	(Grogan *et al.* 2008)
Sulfolobus	Genomic hybridizations	(Andersson *et al.* 2010)
Sulfolobus	Transcription bias near OriC	(Wurtzel *et al.* 2010)
Sulfolobus solfataricus	Single base resolution map of the genome	(Yergeau *et al.* 2009)
Environmental array	Antarctic soil community	(Reichlen *et al.* 2010)
Methanosarcina acetivorans	Regulation of genes	(Schwaiger *et al.* 2010)
Halobacterium salinarum R1	Control of multiple genes by regulatory proteins	(Facciotti *et al.* 2010)

Haloacterium salinarum NRC-1	Physiological readjustments during growth	(Goberna *et al.* 2010)
Environmental array	Methanogens in cattle excreta	(Parnell *et al.* 2010)
Environmental array	Gene transfer	
Environmental metatranscriptomics	Ammonia metabolism in hydrothermal plume	(Baker *et al.* 2012)
Sulfolobus ssp.	Formation of biofilm	(Koerdt *et al.* 2011)
Sulfolobus solfataricus	Prevalence of circular RNA	(Danan *et al.* 2012)
Sulfolobus solfataricus	Viral infection	(Ren *et al.* 2013)
Metalosphaera sedula	Copper and arsenic resistance	(McCarthy *et al.* 2014)
Thermus thermophillus	Effects of a gene KO on transcriptome	(Swarts *et al.* 2015)

Sulfur Metabolism

The first microarray analysis reported on either a hyperthermophilic archaeon was a pilot study on *P. furiosus* that focused on a subset of 271 metabolic genes . This analysis focused on a new sulfur-reducing enzyme complex from *P. furiosus*. The experiment showed at least a twofold change in signal intensity for about 50 ORFs that were represented on the array. Subsequently, this initial study was followed by the analyses of a complete genome array using the same strategy. For most genes the complete ORFs were printed on the array as PCR-amplified fragments. These studies addressed the adaptation of *P. furiosus* cells to the availability of sulfur, different carbon sources, and cold shock.

Heat Shock Response

Although hyperthermophiles have a temperature optimum above 80°C, they still can experience heat stress. As in other severe stress conditions, a heat shock will result in retardation or even complete arrest of growth of the organism. This is a consequence of dropping rates of transcription ; under such conditions protein synthesis appears to be limited to a subset of proteins that play a crucial role in dealing with the stress factor to allow survival. When a heat shock is experienced by the cell, two of the biggest threats are the denaturation of proteins and the increased fluidity of the membrane. In order to cope with these problems, hyperthermophilic archaea have developed their own strategies to cope with such conditions. The hyperthermophilic heat shock responses of two distinct hyperthermophilic archaea, *P. furiosus* and

S. solfataricus (Figure 1), were investigated using transcriptomics. Both organisms seem to react to the same kind of stress differently.

The heat shock experiment using *P. furiosus* was conducted by growing the cells on a mixture of tryptone and yeast extract at a suboptimal temperature of 90°C and then shifting the temperature to 105°C . Cells were harvested after 60 minutes and compared to cells grown at 90°C. *P. furiosus* seems to react in several ways: (i) the compatible solutes di-*myo*-inositol-1,1′-phosphate (DIP) and trehalose seem to be produced in order to stabilize its proteins ; (ii) proteins were further stabilized by the upregulation of several chaperonin-related genes such as the Hsp60-like thermosome, the Hsp20-like small heat shock protein, and two other proteins (VAT) that are predicted to be involved in both protein unfolding (for proteolyses) and refolding processes; (iii) several genes encoding glycoside hydrolases were upregulated, either as a general stress response or as a directed adaptation to heat stress that may enhance the production of sugar-based compatible solutes.

The heat shock experiment conducted with *S. solfataricus* was set up differently . The cells were grown at an optimal temperature of 80°C and then shifted to 90°C. Samples were taken 10 minutes before heat shock, 5, 30, and 60 minutes after heat shock allowing for the elucidation of temporal transcriptome changes. This approach showed that about one-third of the genome (~1000 genes) was differentially regulated in the first 5 minutes. Surprisingly, around 200 of the upregulated genes were IS elements, showing that almost all of these selfish elements of *S. solfataricus* are activated when the cells encounter (temperature) stress; it may well be that the transposition by itself also contributes to part of the modulated expression of other genes.

In contrast to the findings with *P. furiosus,* no evidence was found of induced expression of enzymes involved in compatible solute production. It has been observed that genes that encode different subunits of the RNA polymerase are downregulated, suggesting that transcription is going down. Furthermore, the gene encoding the DNA polymerase II is down, while several DNA repair-related genes have a higher expression. The expression of several transporter genes (e.g., Iron, Cobalt, Phosphate, Sulfate, Amino Acids, Arabinose, Glucose, Maltose) went down. Interestingly, also many transcriptional regulators were differentially expressed, namely, TetR, and the GntR-like repressors. Furthermore the gene encoding the *γ*-subunit of the thermosome was downregulated, while the genes encoding the *α*- and *β*-subunits were unaffected, which was consistent with the previous findings of a change in composition of the thermosome from $1\alpha:1\beta:1\gamma$ to $2\alpha:1\beta:0\gamma$. In conclusion, this experiment showed that in *S. solfataricus* the transcriptional response to a heat shock is instantaneous, but apparently not at the level of compatible solutes. The DNA polymerase II gene is downregulated and a decrease in growth rate is observed. Furthermore the transcription of different subunits of the RNA polymerase is reduced suggesting a global transcription reduction. Many transcriptional regulators appear to play a role in coping with a heat shock in *S. solfataricus,* and it would be very interesting to establish their specific function, that is, their target promoters. The difficulty in comparing these two studies is mainly caused by the different sampling approach. In case of *S. solfataricus* the shift has been made from the temperature at which the growth is the fastest; in case of *Pyrococcus* there might be additional variation in the results related to the suboptimal temperature at the beginning of the experiment.

Viral Infections and Microorganism Interactions

In most environments viral particles significantly outnumber microbial cells, indicating that viral infection is a common threat to the majority of organisms. Hyperthermophiles are not an exception to this rule. Here we discuss two viral infection studies of *S. solfataricus*, both of which have been conducted by using DNA microarrays that contained oligonucleotides corresponding to genes of both *S. solfataricus* as well as genes from selected *S. solfataricus* viruses and plasmids. One study described infection by the lytic virus STIV (*Sulfolobus Turreted Icosahedral Virus)* that usually only kills part of the *S. solfataricus* population in its life cycle , whereas comparable analyses have been performed on the well-studied lysogenic SSV1 virus (*Sulfolobus shibatae* Virus 1) .

The study of STIV conducted by Ortmann et al. comprises of the isolation of a *S. solfataricus* mutant that is hypersensitive to the studied virus with almost all cells of a culture being killed in the lytic cycle. STIV is a dsDNA virus with a circular genome of 17 kb, containing 37 predicted ORFs. Analysis of the viral transcriptome showed the upregulation of 47 of the 52 viral microarray probes, which cover the viral genes and some intergenic regions in both directions. Transcription of viral genes was first detected at 8 hpi (hours post infection), whereas at 16 hpi most viral genes are expressed. At 24 hpi a shift takes place from virus replication to preparation for lysis and around this time point most viral genes are expressed; general cell lysis occurs at 32 hpi. Although the expression starts at different time points, no real temporal expression has been observed in this experiment; however, one cannot rule out that this is a resolution issue due to suboptimal synchronization of the infection cycle. At the early stage of viral gene expression (8 hpi) there are four transcripts and an intergenic region that are being expressed. These

genes are most probably responsible for initiation of the early infection process. Expression of most structural viral genes is found at 16 hpi and thereafter. Of the 177 host genes that were differentially regulated (more than 2-fold), of which 124 were upregulated, most are associated with either DNA replication and repair or genes of unknown function, suggesting that STIV uses host proteins to aid the replication of its own DNA. An important upregulated protein concerns the ESCRTIII homolog, which has recently been reported to be essential for the cell division in *Sulfolobales* ; the upregulation may suggest involvement in the recently discovered release system for both STIV and SirV that involves unique pyramid-like structures (Figure 2) . All of the downregulated host genes were regulated just before cell lysis at 32 hpi and were associated with metabolism.

An infection study of SSV1 with *S. solfataricus* as a host has been conducted in order to find out more about the transcriptome fluctuations of this lysogenic virus and its host . Initially infection by SSV1 seems not to affect the growth rate of the infected cells; at least partly, the SSV1 genome is integrated at a specific site in the host chromosome ; however, as soon as SSV1 starts to produce and release viral particles, the cell growth is significantly retarded. Viral production can be greatly stimulated after UV induction. The first viral transcripts can already be found at 1 hpi, while most viral genes are active at 8.5 hpi. The viral genes are clustered as 9 operons, comprising both regulatory genes and structural genes. The regulatory genes are the first ones to be transcribed, and the genes coding for the coat protein of the virus are produced at a later stage.

There are more differences between the two studies, and only few similarities. Comparison of the two datasets is not straightforward, mainly because it

compares infection by two distinct types of viruses (lytic versus lysogenic); in addition there are some methodological differences like the different time points involved, number of time points taken into account, and so forth. One of the main differences concerns the fact that STIV seems to have a larger impact on the host due to a more profound regulation of host genes (177 instead of 55); this may correlate with its lytic live-cycle. However, to deduce general patterns it will be necessary to compare the transcription profiles during a synchronized infection of additional viruses. A recent study on the infection of the closely related *S. islandicus* with the lytic virus SirV revealed a dramatic degradation of the host chromosome upon viral assembly and proliferation ; no transcriptome analysis of host genes after infection of this system has yet been reported.

The microarray technique can be used to observe the interactions between two distinct species. One such attempt has been done on a bacteria, *Thermotoga maritima*, which has been grown alone as well as in a coculture with a archaea, a methanogenic thermophile, *Methanocaldococcus janaschii* . This experiment yielded an interesting view on the importance of the H_2 transfer in hot environment. The experiment focused on a shift from the mid logarithmic growth phase to the early stationary. It has been observed that the growth of *T. maritima* has been boosted 3- to 5-fold due to removal of inhibiting H_2. Also the methane production of *M. jannaschii* has been increased twofold compared with pure culture. The transcriptome analysis of the 2 samples from the early stationary phase showed that in the pure culture of *T. maritima*, 127 genes have been significantly upregulated in comparison with the coculture. Half of those were associated with the central carbon metabolism. At the same time, in the coculture of the 113 genes upregulated, the main groups present were ABC transporters and carbohydrate

hydrolases. This suggests that the pure culture conditions support the main metabolic pathways while the coculture conditions seem to boost the scavenging. The scavenging strategy may be boosted by the exopolysaccharide (EPS) produced by the co-culture cells that form aggregates to enhance the hydrogen transfer . Another, less obvious conclusion from the experiment was the confirmation that in this case, a microarray platform designed to analyze one species can be successfully used to analyze a co-culture condition.

Genome Replication and the Cell Cycle

Up until 2004 it was assumed that genome replication with multiple origins of replication was a typical Eukaryotic-like feature . In 2004, different groups independently discovered that *Sulfolobus* spp. has multiple origins of replication . Using 2D DNA gels, two origins of replication could be demonstrated in *S. solfataricus,* while a microarray approach (quantification of genomic DNA by hybridizing it with a DNA microarray) was used to prove that *Sulfolobus* spp. has actually three origins of replication (Figure 3). In the latter study *Sulfolobus* cells were treated with acetic acid in order to synchronize the initiation of replication. After removal of the acetic acid inhibition, the cells were harvested at different time points and genomic DNA was extracted and hybridized on a microarray. It was revealed that all three *cdc6*-like genes in both *S. acidocaldarius* and *S. solfataricus* were functional. Although this was a major breakthrough in the field of prokaryotic genome replication, it should be stressed that other archaea (incl. *P. abyssi*) have a single origin of replication . Together with the fact that none of the known bacterial chromosomes possess multiple origins, this strongly suggests that

multiple origins are an archaeal invention, and that the last universal common ancestor (LUCA) most likely possessed a single origin of replication .

The cell cycle of the *Sulfolobus* spp. is relatively well studied and, although some archaeal species show modifications to this model , it is currently used as archetype of the archaeal cell cycle. An important mechanistic difference, however, concerns the involvement of the ESCRT-III-based system in crenarchaea, versus the FtsZ-based, tubulin-directed system in euryarchaea . *S. solfataricus*, interestingly, possesses both the ESCRT-III encoding genes as well as a gene hypothesized to be an FtsZ paralog . In 2007, Lundgren and Bernander used a microarray approach to analyze a time series of synchronized cells of *S. acidocaldarius* to show that a cyclic induction of genes is involved in the cell cycle . The cell growth was arrested in the G2 phase by addition of acetic acid (dissipates membrane potential and inhibits overall metabolic activity at low pH); after resuspending the cells in fresh medium, the synchronized cells started to grow again after 30 minutes. Cells were analyzed at 8 different time points allowing a good overview of global gene expression patterns starting at the G2 phase (0–30 minutes) going all the way through the cycle until the cells are again in the G2 phase (about 200 minutes later). In a parallel study, using a distinct manner of synchronization in which cells are captured at low temperature right after cell division (the baby machine), Samson et al. presented a cell cycle-dependent transcription of ESCRT-III system components and a Vps4 homolog in *S. acidocaldarius* . Interestingly, though not annotated as ESCRT/Vps4, similar expression profiles of these genes were described in the parallel study mentioned above . The observed activity of ESCRT-III system in Crenarchaeal cell cycle suggests a common ancestry of cell division mechanisms in archaea and eukarya.

Apart from shedding light on the cell division mechanisms, microarray analysis allowed observing a cyclic expression of different kinases, at least seven transcription factors, as well as the three *cdc6* genes. These findings suggest that the cell cycle is regulated at different levels. Of the three *cdc6* genes, *cd6-1* is the first to be highly expressed, slightly before the G_1/S transition. Shortly after the induction of the first *cdc6* gene, the *cdc6-3* gene is induced, confirming its secondary role to the *cdc6-1* gene. The gradual induction of the *cdc6-2* gene slightly before the cells approach the G_2 phase suggests a negative regulatory role in chromosome regulation as suggested in earlier studies . On the other hand, the data from Duggin et al. implies that the Cdc6 protein levels during the cell cycle synchronized using the baby machine remain unchanged. The discrepancy between the results is hypothesized to be an effect of two different synchronization methods rather than from the cell cycle itself. Acetate can induce stress in the cells and influence transcription of some stress response-related genes. It can also be a result of differential levels of transcript levels and protein; however this possibility is undermined by the fact that other studies showed a correlation between protein and transcript level in case of this gene .

Pentose Metabolism in Archaea

Most genomes consist of considerable fractions of hypothetical genes for which a function cannot accurately be predicted. These genes are either too distantly related to well-established orthologs to be recognized as such; alternatively, they may encode novel types of proteins, either involved in unique processes/bioconversions or playing a role in a known process but being the result of a non-orthologous gene displacement . Microarrays can help elucidating the function of these hypothetical genes, by comparing the

transcriptomes in condition where a given process/pathway is expected to be active or not. As such, appropriate transcription profiles could serve as leads for further research.

A good example of a successful microarray-based discovery in archaeal metabolism concerns the elucidation of a pentose-converting pathway in *S. solfataricus*. Unlike many other bacteria and eukaryotes, Archaea do not seem to have the classical oxidative pentose phosphate pathway to produce pentose precursors. In addition, until recently the mechanism of the catabolic process of many pentoses in Archaea was not understood in great detail . The analysis of Brouns et al. helped to understand how D-arabinose is metabolized by *S. solfataricus*; moreover, insight was gained in the composition of some general pentose oxidation pathways in both Archaea and bacteria . In this study, the microarray technology has been used as an initial step of pathway elucidation and allowed for composing a short list of potential candidate enzymes. Comparison between cells grown on D-arabinose and D-glucose revealed that 16 genes were significantly upregulated in the first condition. These included the genes encoding the 4 subunits of a previously identified arabinose ABC transporter, a putative sugar permease, and 5 hypothetical enzymes. Comparing the sequences of the intergenic regions revealed the presence of a conserved palindromic motif in promoter regions of 5 of the upregulated genes: the arabinose ABC transporter operon, and 4 of the hypothetical genes. Production and characterization of the 4 corresponding enzymes has resulted in unraveling the arabinose-degrading pathway.

A further *in silico* investigation of the genes resulted in the finding of different but very similar degradation pathways for several C_5 (D- and L-arabinose, D-

xylose, hydroxyl-proline) and C_6 (D-glucaric acid, D-galactaric acid) substrates , used by different organisms. Interestingly, all proposed pathways converge at 2,5-dioxopentanoic acid, which is converted to the citric acid cycle intermediate 2-oxoglutaric acid (*α*-ketoglutarate). This is yet another example of the metabolic tinkering during the evolution of metabolic pathways . As biochemical pathways of archaea can be very different from their bacterial/eukaryotic counterparts, DNA microarrays in combination with the currently established gene disruption techniques for *Sulfolobus spp.* and *Thermococcus kodakaraensis* may provide a solid basis for subsequent analyses.

RNA-seq era

The current transcriptomics approach relies on high throughput RNA-seq techniques, where RNA is used to generate complementary DNA (cDNA) that will then be sequenced . A major practical advantage is that this procedure is based on general, species-independent protocols, which allows transcriptomics of organisms with no known or annotated genome . It was also used to culture unculturable species by linking their transcriptome to their nutritional needs, which allowed creation of a custom made medium. Moreover, it allows for comparison of multiple species in co-culture by simultaneous analysis using the same platform, without a need of designing very specific microchips. Because of these features, this technology is frequently used the transcriptomics analysis of environmental samples.

A disadvantage of this approach for analysis of prokaryotic transcriptomes is the overabundance of the rRNA-species, compared to the mRNA-species (only <5% of the total cellular RNA consists of mRNA). This overabundance

of non-mRNA species in the sequenced sample results in a high-noise factor and also could result in not detecting mRNA that is present in only low amounts. Therefore many protocols rely on the specific removal of rRNA before actual sequencing . Most of them are based on techniques that fish out mRNA by using the poly-A tail, which eukarial mRNA posses, but prokaryotes do not. Despite these practical challenges, Wurtzel et al. have successfully analyzed the transcriptome of *S. solfataricus* by deep sequencing, without the removal of the rRNA . They have grown the organism on glucose, cellobiose, and cellulose and sequenced the cDNA using the Illumina Genome Analyzer (Solexa). Of the originally proposed set of 3300 genes , the deep-sequencing study managed to correct the annotation of 162 genes, define 80 new ORFs, predict 80 noncoding RNA's, predict a possible hypersensitive RNA cleavage site, and determine the operon structures of more than 1000 transcriptional units. Moreover, they have found that at least 80 of the *S. solfataricus* operons have overlapping antisense transcripts, a relatively high number (8%) in prokaryotes. These *cis*-encoding transcripts most likely play a role in control of gene expression at either transcriptional or translational level .
Soon after multiple studies examining the transcriptome of hyperthermophiles followed. Most of the transcriptomic experiments focused on a-typical transcript that eluded the microarray based experiments beforehand . This is linked with a boost in discovery of the roles of small RNA particles, for example in CRISPR/Cas system .

RNA-Seq enabled novel experimental approaches to be tested in hyperthermophiles. One example is evolutionary adaptation of *Sulfolobus solfataricus* to more acidic conditions and analysis of its transcriptome before and after the adaptation . Cell have been grown at pH of 3.00 and gradually transferred to lower pH conditions over three years until they were capable of

growing at pH of 0.80. The initial strain was not viable at pH values of 1.60. After that, the transcriptomes of original strain, strain recovered from pH 1.50 and 1.00 were analysed. This showed major changes of the regulation in genes involved in oxidative stress, leading to upregulation of TCA cycle. This is consistent with the fact that although *Sulfolobus* thrives in low pH, its cytoplasmic pH is neutral. Keeping the pH gradient in more acidic environment requires more energy to be spent on actively pumping protons out of the cell. Another finding was a major change in the regulation of genes encoding membrane and proteins involved in lipid metabolism. This suggests a much higher turnover of membrane lipids in the more acidic environment. These findings were consistent with the reduced growth rates in lower pH. The upkeep cost of cells exposed to more stressful conditions causes diversion of energy towards those functions leaving less substrate that can be directed towards growth. This also explains why in nature *Sulfolobus* is not found growing in lower pH values, as the energy available in the natural habitats is not sufficient to allow them successfully pumping the protons, repairing constantly damaged membrane and have enough surplus to sustain growth.

Standardized Procedures

High-throughput functional genomics approaches are frequently combined in systems biology approaches aiming at modeling the physiology of microbial cells. A very good example of such a systems approach in mesophilic archaea is a study by Bonneau et al. , in which transcriptome analysis was part of an integrated analysis aiming at the reconstruction of a gene networks in the halophilic archaeon Halobacterium sp. By using different transcription regulators, genetic modification, and high-throughput methods, a model has

been generated that describes the behavior of this network in a range of conditions. Such a systems approach combined with modeling allows picturing the interactions of an organism and predicting its behavior in the natural environment. The difficulty of such an approach lies in synchronizing a large research project and having a uniform biomaterial to start with.

An example of such a systems biology approach in thermophilic archaea concerns the SulfoSYS project , which is part of the European SysMO consortium. A major goal of the latter consortium is to establish well-integrated systems biology projects on selected model organisms. A major goal of the SYSMO projects is to perform a multidisciplinary, functional genomics approach that should be highly reproducible because of the implementation of well-described, standard protocols. In the SulfoSYS project the model organism S. solfataricus is cultivated in a very controlled way. The obtained cells are then distributed among the different researchers to perform transcriptomics, proteomics, metabolomics, as well as biochemical analyses; eventually the data are included in an integrated metabolic model. The stringency of cultivation and sampling has been important also due to a comparison of cells from different temperature values. As the half-lives of some mRNA particles can be as low as 2 minutes , a slight difference in sampling may lead to a large difference in the transcript level. The impact of the careful preparation of biological samples in functional genomics analyses, including DNA microarray experiments, has not always been appreciated; on the other hand it is generally accepted that this may significantly affect the reproducibility of this approach. The SulfoSYS project puts much weight on careful sample preparation and on verifying the quality of the obtained cell material before performing actual experiments ; this has resulted in a combined dataset with microarray and deep sequencing data that are in very

good agreement.The SysMO consortium puts extra weight on giving an unrestricted and easy access to the generated data . As far as the datasets of respective microarrays are usually freely available, the multitude of standards, methods, and platforms severely impedes the possibilities of comparing two datasets with each other. Applying the deposition standards, as Minimum Information About a Microarray Experiment (MIAME) , certainly helps to validate the quality of the data; however, a simplified standard for results storage could be proposed to allow quick and efficient analysis of deposited datasets.

Conclusions and Outlook

DNA microarrays have been very successful during the first decade of the 21st century, as a high-throughput research tool that has led to important scientific discoveries, including important findings on cell biological/metabolic features of hyperthermophilic archaea, as outlined above. The most frequently used DNA microarrays (based on oligonucleotides) have restrictions because the probe design is based on previously made assumptions with respect to predicted genes; this implies that small ORFs and noncoding RNAs are generally not included on microarrays. In addition, the commonly used technology only allows for relatively limited numbers of spots that can be printed on one slide. The problem of an incomplete set of probes is solved by using tiled DNA microarrays, which are composed of overlapping oligonucleotides. The used probe lengths and the degree of tiling between overlapping probes determine the resolution that can be achieved; typically 2–4 × 10^5 probes are printed per slide, with probe size ranging between 50 and 75 nucleotides. Tiled arrays cover the two complete strands of the target chromosomes .

At present new ways of obtaining global transcriptomic data are predominantly used. Sequencing cDNA (RNA-seq), from the very advent of the technology seemed to be very promising and delivered this promise expanding analytical scope of transcriptomics. In eukaryotes ORF prediction is not as easy as in prokaryotes and this has often led to the development of cDNA libraries for the production of microarrays. RNA-seq, although frequently used in eukaryotic transcriptomics, become a standard tool in microbial analysis thanks to advances in sequencing power and excluding rRNA reads. The sequencing approach has the advantage that the same platform can be used for different species, resulting in a better interspecies comparison by omitting the cross-platform bias. This opens up the door for environmental transcriptome profiles, allowing for the monitoring of metagenome-based gene expression in the environment, as opposed to the artificial conditions that are generally imposed on them in a laboratory setting. A further advantage might be that RNA-seq is less prone to signal loss due to mutations that arise during cultivation. Although this technique is not yet readily accessible for most labs, the anticipated reduction of sequencing costs in the near future might make this a very attractive general technique for transcriptome analysis for both eukaryotes and prokaryotes. A decrease in the use of the DNA microarray as a research tool and an increase of using sequencing-related techniques in this field may be expected , with some predicting even that the technology is going to go extinct altogether .

RNA-seq might turn out to be quintessential in examining environmental samples where not all of the components have been known beforehand. For instance, they might greatly help to increase our understanding of phage pressure on the potential hosts that takes place in situ by finding more viral transcripts and watching the response of the thermophiles to multiple viruses

present in the environment. One can assume that hyperthermophilic environments are a very good target for early attempts of metatranscriptomic analyses as the ecology of such niches is generally less complex than that of aquatic or soil ecosystems, making it easier to deal with big dataset covering many organisms.

But microarrays did not become obsolete. Instead there might be a renaissance of the technology that ironically is driven by the advances in sequencing field. Due to improvements in dye quality and oligo-nucleotide printing microarrays used as a fast diagnostic tool, with services provided by outside labs are coming back
strongly, especially in the industrial applications where analysing known unknowns very rapidly and at a lower price can be a successful strategy. These arrays allow analysing a full genome at a much lower price, as they let the researchers process up to 384 samples in one run with instant result collection. Microarrays have thus stopped to be tools of discovery to the extent they were in their early days and started to be diagnostic tools that lack the indepth of de-novo sequencing but beat them in speed and cost. That did not yet translate to the field of thermophiles but certainly with industrial partners it is possible in the future to design a microarray system custom made to analyse the transcriptome of thermophilic methanogenic communities or ones designed to look at the ecosystem of hydrothermal vents and other hyperthermophilic environments to look for microbial functionalities useful in applied fields.

Chapter 3
"Hot standards" for the thermoacidophilic archaeon *Sulfolobus solfataricus*

Melanie Zaparty[1*] , Dominik Esser[1*], Susanne Gertig[2*], Patrick Haferkamp[1*], Theresa Kouril[1*], Andrea Manica[3*], Trong K. Pham[4*], Julia Reimann[5], Kerstin Schreiber[2], Pawel Sierocinski[1*,6], Daniela Teichmann[3], Marleen van Wolferen[8], Mathias von Jan[7*], Patricia Wieloch[2*], Sonja V. Albers[5], Arnold J. M. Driessen[8], Hans-Peter Klenk[7], Christa Schleper[3,9], Dietmar Schomburg[2], John van der Oost[6],Phillip C. Wright[4] and Bettina Siebers[1]

(1) Faculty of Chemistry, Biofilm Centre, Molecular Enzyme Technology and Biochemistry, University of Duisburg-Essen, Lotharstraße, 47057 Duisburg, Germany; (2) Department of Bioinformatics and Biochemistry, Technische Universität Braunschweig, Langer Kamp 19b, 38106 Braunschweig, Germany (3) Department of Genetics in Ecology, University of Vienna, Althanstraße 14, 1090 Vienna, Austria (4) Biological and Environmental Systems Group, ChELSI, Department of Chemical and Process Engineering, University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK (5) Molecular Biology of Archaea, Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Straße, 35043 Marburg, Germany (6) Laboratory of Microbiology, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, The Netherlands (7) e.gene Biotechnologie GmbH, Poeckinger Fussweg 7a, 82340 Feldafing, Germany (8) Department of Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands (9) Department of Biology, University of Bergen, Jahnebakken 5, 5020 Bergen, Norway

*M. Zaparty, D. Esser, S. Gertig, P. Haferkamp, T. Kouril, T. K. Pham, P. Sierocinski, M. von Jan and P. Wieloch contributed equally to this project

Abstract

Within the archaea, the thermoacidophilic crenarchaeote *Sulfolobus solfataricus* has become an important model organism for physiology and biochemistry, comparative and functional genomics, as well as, more recently also for systems biology approaches. Within the *Sulfolobus* Systems Biology "SulfoSYS" project the effect of changing growth temperatures on a metabolic network is investigated at the systems level by integrating genomic, transcriptomic, proteomic, metabolomic and enzymatic information for production of a silicon cell-model. The network under investigation is the central carbohydrate metabolism. The generation of high-quality quantitative data, which is critical for the investigation of biological systems and the successful integration of the different datasets, derived for example from high-throughput approaches (e.g., transcriptome or proteome analyses), requires the application and compliance of uniform standard protocols, e.g., for growth and handling of the organism as well as the "–omics" approaches. Here, we report on the establishment and implementation of standard operating procedures for the different wet-lab and in silico techniques that are applied within the SulfoSYS-project and that we believe can be useful for future projects on *Sulfolobus* or (hyper)thermophiles in general. Beside established techniques, it includes new methodologies like strain surveillance, the improved identification of membrane proteins and the application of crenarchaeal metabolomics.

Electronic supplementary material

The online version of this article (doi:10.1007/s00792-009-0280-0) contains supplementary material, which is available to authorized users.

Abbreviations

CCM: Central carbohydrate metabolism

ED: Entner–Doudoroff
EMP: Embden–Meyerhof–Parnas
SOP: Standard operating procedure
SulfoSYS: *Sulfolobus* Systems Biology

Introduction

The thermoacidophilic archaeon *S. solfataricus* represents one of the best studied members of the (hyper)thermophilic organisms within the phylum crenarchaeota, and thus represents a most suitable archaeal representative for “Hot Systems Biology”.
Systems Biology represents a relatively young scientific area that is applied at various levels of living systems, i.e., a metabolic network, cells or interacting organisms. Systems Biology aims to systematically decipher the communication between parts and modules or complex biological systems and how these lead to functioning of these systems (Snoep & Westerhoff 2005). Furthermore, Systems Biology enables the potential to realize a quantitative view on, for instance, metabolic processes of an organism including the regulatory mechanisms.

S. solfataricus optimally grows at 80°C (60–92°C) and pH 2–4. The *S. solfataricus* strain P2 (DSM 1617) was originally isolated from Pisciarelli, Italy (Zillig *et al.* 1980), but closely related strains reside in high numbers in virtually all acidic hot springs around the globe. The organism is a strict aerobe and grows heterotrophically on a variety of organic compounds as carbon and energy source such as sugars (e.g., glucose, galactose, arabinose, sucrose), amino acids or peptides (Grogan 1989), thus, *S. solfataricus* can be easily maintained in the laboratory with relatively little special equipment. The

complete genome sequence is available (She *et al.* 2001a), functional genomics approaches have been applied to study this organism, including transcriptomics, proteomics and comparative genomics (Verhees *et al.* 2003; Snijders *et al.* 2006). Furthermore, several in vitro assay systems to analyse aspects of information processing in (hyper)thermophiles, such as replication, transcription or translation, have been established for *S. solfataricus* (Ruggero *et al.* 1993; Bell & Jackson 2001; Kelman & White 2005; Barry & Bell 2006) and many of its proteins have been crystallized. The development of genetic tools for *S. solfataricus* has been a major breakthrough that allows for the study of gene functions and the potential to perturb the system (Jonuscheit *et al.* 2003; Worthington *et al.* 2003; Albers & Driessen 2008; Wagner *et al.* 2009).

The *Sulfolobus* systems biology ("SulfoSYS")-project (Albers *et al.* 2009) represented the first (hyper-)thermophilic Systems Biology project, funded within the European trans-national research initiative "Systems Biology of Microorganisms" (SysMO; http://www.sysmo.net/). Within the SulfoSYS-project, focus lies on studying the effect of temperature variation on the central carbohydrate metabolism (CCM) of *S. solfataricus* (Albers *et al.* 2009) that is characterized by the branched Entner–Doudoroff (ED)-like pathway for sugar (glucose, galactose) degradation (Lamble *et al.* 2003, 2005; Ahmed *et al.* 2005; Kim & LEE 2005; Kim & Lee 2006) and the Embden–Meyerhof–Parnas (EMP)-like pathway, which is employed during gluconeogenesis (Snijders *et al.* 2006) for review see (Van der Oost & Siebers 2007; Zaparty *et al.* 2008).

The effect of temperature changes on the CCM network of *S. solfataricus* is analyzed by the tight integration of bioinformatics, genome, transcriptome, proteome, metabolome, and enzymatic data, with all –omic and biochemical

data being produced from identical batches of biomass. Beside providing experimental data, one main part of this highly integrative project is the in silico analysis of the CCM network, including the design of a sufficiently precise model according to the silicon cell type model (http://www.siliconcell.net, (Olivier & Snoep 2004)). This model will allow for the computation of the *S. solfataricus* CCM, and in particular to investigate its robustness to changes in temperature at the system level.

Prerequisites for reproducibility and reliability of the produced datasets and the successful integration of the different data are the establishment and application of uniform standards, e.g., for the handling of the organism as well as the realization of the coordinated experiments. A basic necessity for the project was the evaluation of a suitable *S. solfataricus* strain and control of its genomic stability, followed by the optimization and standardization of growth conditions, handling of glycerol stocks and biomass production. First pilot experiments have been performed with *S. solfataricus* grown at 80°C (optimal growth temperature) compared to 70°C in order to improve and implement the SOPs, as well as establish the new methodologies applied to *S. solfataricus*.

Here, we report on the establishment and application of standard operating procedures (SOPs) regarding genomic, transcriptomic, proteomic, metabolomic as well as biochemical techniques applied for a comprehensive analysis of the CCM of the thermoacidophile *S. solfataricus* in the course of the SulfoSYS-project. Within the scientific archaeal community, this project represents the first effort to prepare common standards. Furthermore, new methodologies like the iTRAQ method for membrane proteome analysis have been established and applied successfully. Moreover, to our knowledge, this is the first report on metabolome analyses performed with a crenarchaeon.

In general, working with (hyper)thermophilic organisms (Bacteria or Archaea) or (hyper)thermophilic enzymes, is not always favorable due to the sometimes substantial technical challenges. However, it also harbors several experimental advantages, for example recombinant (hyper)thermophilic proteins can be easily purified from mesophilic hosts via heat precipitation, and because of their high rigidity they tend to crystallize easier. With our work we want to further contribute to establish *S. solfataricus* and also other (hyper)thermophiles as model organisms.

The *S. solfataricus* "Hot standards" will be updated on a regular basis and will be available, together with additional information (e.g., workflows), at the SulfoSYS homepage http://www.sulfosys.com/.

Strain evaluation and test for genomic stability of *S. solfataricus* strains P1 and P2

A special feature of the *S. solfataricus* genome is the presence of about 20 different types of mobile transposable elements (IS-elements) that occur at 10–25 copies each in the genome and that have been demonstrated to actively move or multiply (Schleper *et al.* 1994; Martusewitsch *et al.* 2000; Redder *et al.* 2001; She *et al.* 2001a). Therefore, a particularly strict control of the genomic integrity of the organism is required over the course of the experiments. To avoid accumulation of mutations, it is common practice in most laboratories working with *Sulfolobus*, to prepare a large number of stocks from a primary culture obtained from DSMZ, from which experiments are started freshly, but the effectiveness of this procedure has not been examined.

In order to evaluate this maintenance procedure and to select a suitable strain for a Systems Biology project, seven different stocks of the *S. solfataricus* strains P1 and P2 (DSM 1616 and 1617) were compared. They were collected from the partners within the consortium as well as from the German Collection of Microorganisms and Cell Cultures (DSMZ), where stocks had been deposited about 15 years ago.

Cells from each stock were grown in parallel under identical conditions and chromosomal DNA was prepared (SOP_SSO_080901). Probes targeting four different IS elements (ISC1058, ISC1217, ISC1439 and ISC1359), were used in Southern hybridizations to produce characteristic footprints of the genomic DNA (Fig. 1). Three out of three tested *S. solfataricus* P1 stocks showed highly similar patterns in these hybridizations, as did four out of five different stocks from *S. solfataricus* P2. Only one stock that had been subcultured for several months in the laboratory showed major changes in the chromosomal footprints with all four probes tested (two of these are shown in Fig. 1, stock 2). All other stocks stemmed from laboratories in which cultures were routinely discarded after three to four passages in order to avoid the accumulation of spontaneous mutations. This analysis showed for the first time, that the maintenance of the strains as performed in most laboratories is indeed quite effective. The stock of *S. solfataricus* P2 (DSM1617) deposited at DSMZ was selected to be used in the SulfoSYS-project, in order to allow comparability to studies from other laboratories and because the complete genome of this strain is available (She *et al.* 2001a). The strain has not undergone major genomic rearrangements during its maintenance at the DSMZ, since its chromosomal patterns were mostly identical to the four other stable stocks, including one that stems from the W. Zillig's laboratory and has not been touched over the last 15 years (lane 2, Fig. 1).

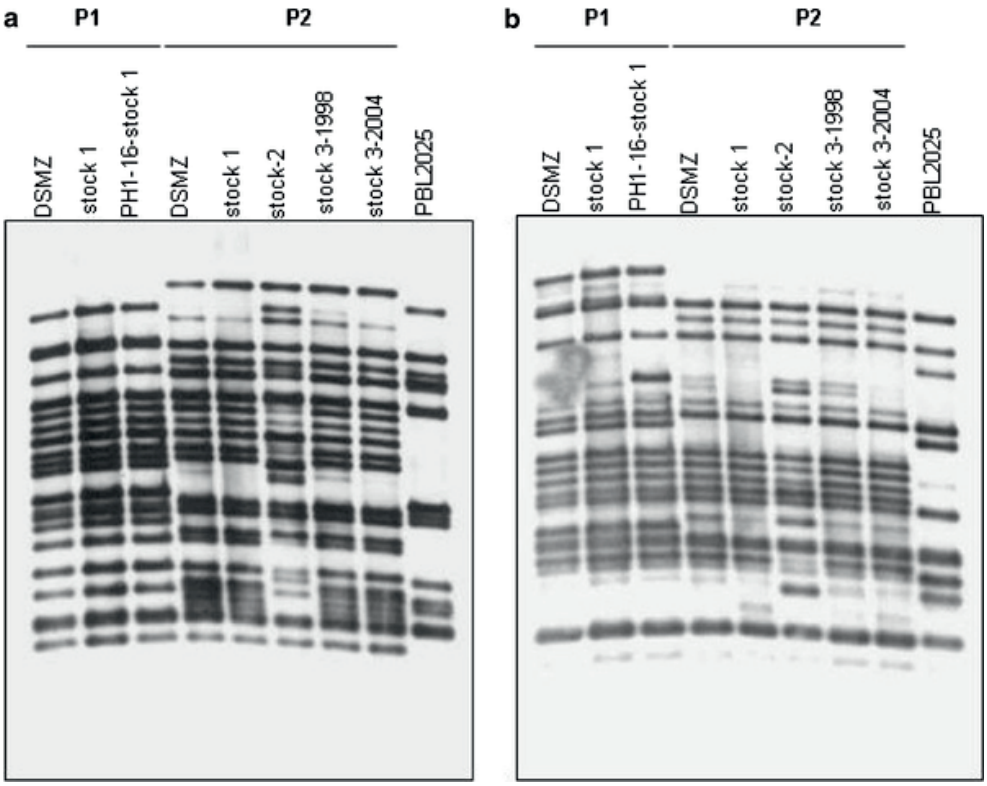

Fig. 1: Southern hybridization of *Afl*III-cut chromosomal DNAs hybridized with DIG-DNA probes of IS-element ISC1439 (**a**) and ISC1058 (**b**), respectively. *Lanes 1–3* Strain *S. solfataricus* P1 (DSM 1616), *lanes 4–8* strain P2 (DSM1617), *lane 9* strain PBL2025 (used for constructions of knockout mutants (Worthington *et al.* 2003). *DSMZ* stock obtained freshly from DSMZ, *stock 1–3* obtained from three different laboratories of this consortium, in which *S. solfataricus* is regularly grown. Stocks 3/1999 and 3/2004 were kept in the same laboratory, but were obtained in two different years

A detailed SOP procedure has been established for the production of glycerol stocks (SOP_SSO_080906a, b; for details see supplement S1) and for the

evaluation of genomic integrity of the strain after fermentations in the SulfoSYS project (SOP_SSO_080901). For each fermentation, cells were grown from stock cultures to avoid the accumulation of mutations. In addition, Southern hybridizations are used to make sure that the stocks have not been contaminated by the virus SSV1 or its derivatives that are routinely used in the laboratories for genetic manipulations (SOP_SSO_080901).

Test for genomic stability (SOP_SSO_080901)

The different *S. solfataricus* strains are grown at 78°C and pH 3 in Brock's basal salt medium supplemented with 0.2% D-arabinose and 0.1% tryptone. Pyrimidine-auxotrophic mutants (PH1-16) are grown in media supplemented with 10 µg/ml uracil. For the isolation of chromosomal DNA 10 ml of an exponentially grown liquid culture (A_{600nm} = 0.25–0.4) are precooled on ice and centrifuged for 10 min at 4,000 rpm and 4°C. The cells are resuspended in 500 µl TEN solution (20 mM Tris/HCl, 1 mM EDTA, 100 mM NaCl) and 500 µl TEN solution supplemented with 1.6% *N*-laurylsarcosine and 0.12% Triton X-100. After an incubation of 30 min at room temperature, the chromosomal DNA is extracted with phenol:chloroform:isoamylalcohol (25:24:1) twice and two times with chloroform, finally the DNA is precipitated with ethanol. For southern hybridizations, 3 µg of chromosomal DNA are incubated with *Afl*III and separated on a 0.7% agarose gel. The DNA is blotted on nylon membranes and hybridized with digoxigenin-labeled double stranded DNA probes (approx. 1,000 bp) specific for each of the four IS-elements used in the analysis or the virus SSV1, respectively.

Standardized fermentation of ***S. solfataricus*** P2

S. solfataricus is an obligate aerobe and a chemo-organo-heterotroph, growing on various carbon sources, such as yeast extract, tryptone or various sugars, amino acids and peptides (Grogan 1989). The thermoacidophilic organism optimally grows at 80°C (60–92°C) and pH 2–4. Cultivation of the organism under well-defined conditions represents one of the most important prerequisites for reproducibility and reliability of the produced data derived from the different technologies as well as subsequent data integration. Determination of the optimal growth conditions and the fermenter set-up, have been performed at the optimal growth temperature of 80°C (Fig. 2; SOP_SSO_080903).

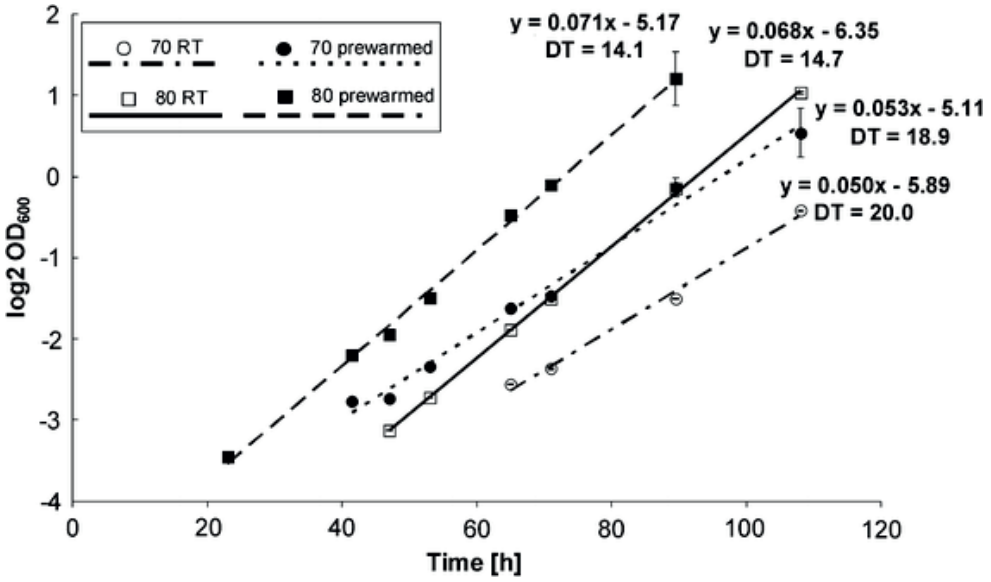

Fig. 2: Log phase of *S. solfataricus* growth at 70 and 80°C (log2 scale). Inoculation of the medium preheated to desired temperature (*filled circle*, *filled square*), inoculation at room temperature (RT) and subsequently heated to desired temperature (*open circle*, *open square*). Growth at 70°C (*filled circle*, *open circle*) and growth at 80°C (*filled square*, *open square*) is shown. *Lines*

represent trend lines for given conditions with equation and doubling time (DT) (h), R^2 values are in all cases >0.98

Minimal medium (SOP_SSO_080902)

The minimal medium according to (Brock *et al.* 1972) contains (amount per litre): 1.3 g $(NH_4)_2SO_4$, 0.28 g KH_2PO_4, 0.25 g $MgCl_2 \times 7H_2O$, 0.07 g $CaCl_2 \times 2H_2O$, 0.02 g $FeCl_2 \times 4H_2O$, 1.8 mg $MnCl_2 \times 4H_2O$, 4.5 mg $Na_2B_4O_7 \times 10H_2O$, 0.22 mg $ZnSO_4 \times 7H_2O$, 0.06 mg $CuCl_2 \times 2H_2O$, 0.03 mg $Na_2MoO_4 \times 2H_2O$, 0.03 mg $VOSO_4 \times 2H_2O$ and 0.01 mg $CoCl_2 \times 6H_2O$. Demineralized water with a value of resistivity not lower than 18.2 MΩ cm at 25°C is used for all solutions. Thus, the medium is uniform, independent from geography or used demineralization technique. Prior to autoclaving, the pH of the medium is set to 3.5 using H_2SO_4 The sterile filtered iron solution is kept in the dark at RT and added to the medium just before inoculation. The filter sterilized carbon sources such as glucose (30%) are added just before inoculation to reach a final of concentration of 0.3%.

Batch fermentation in flasks (SOP_SSO_080903)

The aerobic cultivation of *S. solfataricus* is carried out in 25–100 ml batch cultures in long-neck Erlenmeyer flasks (50–500 ml) at 70 and 80°C in minimal medium containing 0.3% glucose as carbon source (for exometabolome analysis only 0.15% glucose are used, SOP_SSO_080912) according to SOP_SSO_080902. An optimal oxygen supply is given by shaking (160 rpm) using a Thermotron shaker. Prewarmed medium (70 or 80°C, respectively) is inoculated with 200 µl glycerol stock (working stock; SOP_SSO_080906b, supplement S1) and growth is monitored spectrophotometrically at 600 nm. Afterwards, cells are chilled on ice and

harvested by centrifugation (6,000×*g*, 15 min, 4°C) in the exponential growth phase (OD_{600} = 0.8–1) approximately after 96 h of growth and either directly used for analysis or stored at −80°C. For subsequent metabolome analysis cells are harvested by centrifugation (4,629×*g*, 5 min, 25°C), cell pellet is resuspended in 20 ml 0.9% NaCl (w/v) at RT and washed twice (4,629×*g*, 3 min, 25°C; 5810 R) (SOP_SSO_080912a).

Fermenter set-up and fermentation (SOP_SSO_080904)

Fermentation of *S. solfataricus* is performed in a 1.5 l fermenter (Applikon) with controlled temperature and pH settings. Also, oxygen dissolution (dO_2 [%]) is algorithm controlled. Cells are aerated using air.

The organism is grown at respective temperatures and a pH of 3.5 in the minimal medium according to ((Brock *et al.* 1972) SOP_SSO_080902). The temperature of the medium (without glucose and the iron solution) is pre-set 1 day before fermentation start. Calibration of the pH and dO_2 is completed, when the temperature in the fermenter is stable for 16 h.

The buffers used to calibrate the pH electrode for the fermenter (pH 7.0: 0.12 g NaH_2PO_4 in 90 ml H_2O, set pH to 7.15, adjust to 100 ml; pH 3.0: 0.156 g NaH_2PO_4 in 90 ml H_2O, adjust pH to 2.85, adjust volume to 100 ml) are pre-warmed to the respective growth temperature. The oxygen electrode is pre-calibrated prior to fermentation at the respective temperature. At 80°C experimentally determined dO_2 = 80% is the optimal value for *S. solfataricus* for the used setup. As it relates to 3.5 mg/l of dissolved oxygen, this value is used for lower temperatures. The algorithm used to grow *S. solfataricus* P2 cells (for details see supplement S2) is designed to keep the dissolved oxygen

at a level as close as possible to 80%. It is based on regulating stirrer speed and aeration intensity, and taking the growth phase estimate into account (for details see supplement S2).

For the SulfoSYS-experiments cells have been grown on 0.3% glucose as carbon source. Optical densities of liquid cultures are monitored at 600 nm (OD_{600}). The fermenter is inoculated with 0.05 l of a pre-culture OD_{600} = 1.0 (±0.2). Pre-cultures are prepared using −80°C glycerol stocks to inoculate pre-heated medium (respective growth temperature) as it is shown in Fig. 2 to significantly reduce the lag phase of growth.

Cell harvest (SOP_SSO_080905)

When the culture reaches an OD_{600} = 0.85 (±0.15), the cells are sampled in aliquots of 20 ml (for transcriptomics and proteomics), 50 ml (for enzyme assays) or custom amounts dependant on OD_{600} (for the metabolomics). Further samples are taken for strain integrity evaluation. Cells are quickly cooled down to 4°C by dipping the collected cells in centrifugation tubes in liquid nitrogen for 30 s and finishing the cooling down in iced water to prevent sample freezing. Subsequently, cells are collected by centrifugation (3,500×*g*, 12 min, 4°C), catalogued and stored at −80°C in cell samples stock.

Preparation ***S. solfataricus*** glycerol stocks (SOP_SSO_080906a,b)

Beside the development of standard fermentation procedure, uniform handling has been established to prepare *S. solfataricus* glycerol stock solutions. The *S. solfataricus* strain 1617 has been acquired from the DMSZ and a master stock has been prepared (SOP_SSO_080906a, for details see

supplement S1). Based on this master stock, the working stocks are prepared (SOP_SSO_080906b; for details see supplement S1), which are used for inoculation of fermentations.

The master stock is obtained after limited amount of transfers from the DMSZ stock, thus, guaranteeing genetic stability. Part of the master stock has been re-inoculated to create a bulk quantity of working stock used in the experiments. In case of the working stock running out, it can be recreated using the master stock (for details see supplement S1).

Glucose uptake measurements in ***S. solfataricus***

The genome of *S. solfataricus* harbours several primary and secondary transporters (She *et al.* 2001b), but as in all Archaea with only a few exceptions e.g., *Thermofilum pendens*, (Anderson *et al.* 2008) the organism lacks the phosphoenolpyruvate-dependent phosphotransferase system (PTS). Some of the primary active transporters represent sugar binding-protein-dependent ATP-binding cassette (ABC) transporters, and systems have been identified for the uptake of glucose, arabinose, trehalose, cellobiose, maltose and maltotriose (Albers *et al.* 1999, 2000, 2001, 2004; Elferink *et al.* 2001). Recently, the pH-dependent uptake of glucose via a high affinity ABC transporter has been characterized (Albers *et al.* 1999; Elferink *et al.* 2001). Compared to other sugars, glucose has been shown to be most effectively transported.

Preparation of cells (SOP_SSO_080907a)

S. solfataricus P2 cells are grown in 50 ml of Brock medium according to the SOP (SOP_SSO_080902) except containing 0.4% glucose at 80°C until an OD_{600} of 0.3–0.4. Cells are collected by centrifugation (3,000×*g*, 15 min, 4°C) and resuspended in 50 ml of minimal Brock medium (SOP_SSO_080903). This procedure is repeated three times, and cells are finally resuspended to 1/10 of the starting volume at a protein concentration of about 10 mg/ml. Protein concentrations are determined by the BioRad Protein Assay ((Bradford 1976)) with BSA as the standard.

Glucose uptake measurements (SOP_SS_080907b)

Uptake measurements using (^{14}C-) labeled glucose (291 mCi/mmole, GE Healthcare) are performed at 60, 65 and 70°C (Table 1) using a previously described filter based assay (Albers *et al.* 1999). The concentrated cell suspension (10 µl) is added to 90 µl of minimal Brock medium and the solution is pre-warmed for 2 min at 60°C. Next 1 µl of the labelled glucose solution that is diluted with unlabeled glucose to the desired concentration is added yielding a final glucose concentration of 0.1–20 µM. After 10 s, the reaction is stopped by the addition of 2 ml of ice-cold 0.1 M LiCl and the mixture is rapidly filtered through a nitrocellulose filter (0.45 µm pore size, BA 85 nitrocellulose, Schleicher & Schuell). Filters are washed with 2 ml of 0.1 M LiCl and dissolved in 2 ml of scintillation fluid (Emulsifier Scientillator Plus, Perkin Elmer) and counted with a liquid scintillation analyzer 1600CA (Perkin Elmer).

Table 1: Results for glucose uptake in *S. solfataricus* cells grown at 65 and 70°C

Growth temperature (°C)	Uptake temperature (°C)	OD_{600}	Protein concentration (mg/ml)	K_m(μM)	V_{max}(nmol min^{-1} (mg protein)$^{-1}$)
65	65	0.368	15.43	0.44	0.45
65	70	0.368	15.43	0.56	0.62
70	65	0.298	6.29	0.12	0.61
70	70	0.298	6.29	0.23	0.85

Results

The in vitro uptake assay system for glucose has previously been established (Albers *et al.* 1999), [Fig. S1 in the supplemental material] and the apparent K_m for glucose uptake at 60°C and a pH 3.5 has been determined to be 1.9 μM with a V_{max} value of 0.9 nmol min^{-1} (mg protein)$^{-1}$. The assay has been established and performed at 65 and 70°C (Table 1). The assay is currently optimized for use at higher temperatures around 80°C, at which metabolism occurs so fast that label is evaporating as CO_2 very rapidly. The measurements will be tried with only 5 and 2.5 s incubation time.

Reconstruction of the central carbohydrate metabolism (CCM) network by comparative genomics

On the basis of the genome sequence information (She *et al.* 2001a) and previous bioinformatic and experimental studies (Verhees *et al.* 2003; Ahmed *et al.* 2005; Snijders *et al.* 2006; Van der Oost & Siebers 2007) the respective pathways of the CCM of *S. solfataricus* have been reconstructed (Albers *et al.* 2009). CCM reconstruction revealed the presence of:

(i) The branched Entner–Doudoroff (ED) pathway that is promiscuous for glucose and galactose degradation (Lamble *et al.* 2003, 2005; Ahmed *et al.* 2005; Kim & LEE 2005; Kim & Lee 2006). The pathway is characterized by two different branches, a non- and a semiphosphorylative one:

(ii) The Embden-Meyerhof-Parnas (EMP) pathway that is employed during gluconeogenesis.
(iii) An oxidative TCA cycle (including glyoxylate shunt), which is responsible for the complete oxidation of glucose to carbon dioxide by using oxygen as terminal electron acceptor.
(iv) The reverse ribulose-monophosphate (RuMP) pathway, which is utilized in pentose phosphate metabolism.
(v) Finally, pathways for the synthesis and degradation of the storage compound glycogen (Skorko *et al.* 1989) as well as the disaccharide trehalose, which is known as compatible solute involved in stress response, are present.

Reconstruction of the CCM network (SOP_SSO_080908)

The genome sequence information of *S. solfataricus* and other organisms as well as additional bioinformatic data have been derived from the UCSC Archaeal Genome Browser (http://archaea.ucsc.edu/). Blast search analyses are performed by using the nucleotide and protein blast tools (e.g., blastn, blastp, psi-blast) from the National Center for Biotechnology Information (NCBI; http://blast.ncbi.nlm.nih.gov/Blast.cgi). For genomic context analyses the STRING database (http://string.embl.de/) and for comparative genomics the respective tools from IMG (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi?page=home) and from the LBMGE Genomics ToolBox (http://www-archbac.u-psud.fr/genomics/GenomicsToolBox.html) are applied. For pathway reconstruction the KEGG PATHWAY tool from the Kyoto Encyclopedia of Genes and Genomes (KEGG; http://www.genome.jp/kegg/) and for gaining detailed enzymatic information (e.g., enzyme reactions, specificities or enzymatic parameters) the BRENDA database (http://www.brenda-enzymes.org/) is used. The network reconstruction and annotations are regularly updated by using the above described methods and tools.

Results

A total of 97 genes have been identified that encode homologs with either a confirmed or a predicted function in the CCM network of *S. solfataricus* (Fig. 3; (Albers *et al.* 2009)). For several of these identified candidate genes, different functions are predicted, thus, their physiological function needs to be verified. To confirm the gene assignments the enzymatic activities of the recombinant gene products are analyzed (see SOPs_SSO_080913).

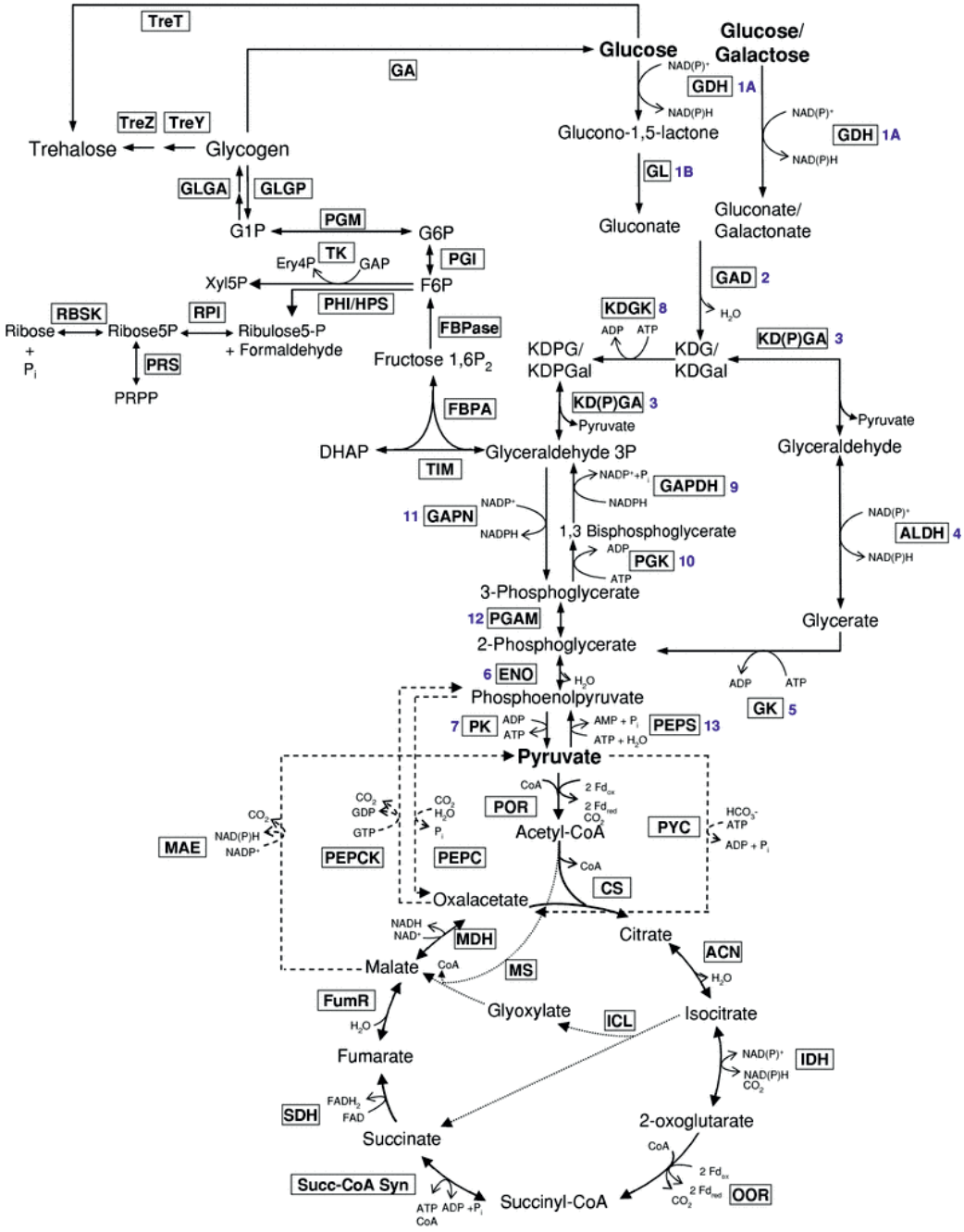
TreT
Glucose
Glucose/
Galactose
GA
NAD(P)+
GDH 1A
NAD(P)H
TreZ TreY
Trehalose
Glycogen
Glucono-1,5-lactone
NAD(P)+
GDH 1A
NAD(P)H
GL 1B
GLGA GLGP
G1P
PGM
G6P
Gluconate
Gluconate/
Galactonate
Ery4P TK GAP
PGI
Xyl5P
F6P
GAD 2
H2O
PHI/HPS
KDGK 8
ADP ATP
RBSK
RPI
Ribose
+
Pi
Ribose5P
Ribulose5-P
+ Formaldehyde
FBPase
KDPG/
KDPGal
KDG/
KDGal
KD(P)GA 3
PRS
PRPP
Fructose 1,6P2
FBPA
KD(P)GA 3
Pyruvate
Pyruvate
Glyceraldehyde
DHAP
TIM
Glyceraldehyde 3P
NADP+ Pi
NADPH
GAPDH 9
11 GAPN
NADP+
NADPH
1,3 Bisphosphoglycerate
NAD(P)+
ALDH 4
NAD(P)H
ADP
PGK 10
ATP
3-Phosphoglycerate
12 PGAM
2-Phosphoglycerate
Glycerate
6 ENO
H2O
Phosphoenolpyruvate
ADP ATP
GK 5
7 PK
ADP
ATP
AMP + Pi
ATP + H2O
PEPS 13
Pyruvate
CoA
2 Fdox
2 Fdred
CO2
POR
Acetyl-CoA
CO2
GDP
GTP
CO2
H2O
Pi
PYC
HCO3-
ATP
ADP + Pi
CO2
NAD(P)H
NADP+
MAE
PEPCK
PEPC
CoA
CS
Oxalacetate
NADH
NAD+
MDH
Citrate
ACN
H2O
Malate
CoA
MS
FumR
H2O
Glyoxylate
ICL
Isocitrate
Fumarate
NAD(P)+
NAD(P)H
CO2
IDH
FADH2
FAD
SDH
2-oxoglutarate
Succinate
CoA
2 Fdox
2 Fdred
CO2
OOR
Succ-CoA Syn
ATP ADP +Pi
CoA
Succinyl-CoA

Fig. 3: Reconstructed CCM of *S. solfataricus*. Identified CCM reactions (enzyme abbreviations *boxed*) involved in the branched ED and the EMP pathway [reactions numbered, corresponding to Table 3)], the citric acid cycle including the glyoxylate shunt (*dotted arrow*) the reversed ribulose monophosphate pathway, C3/C4 conversions (*dashed arrow*) as well as glycogen and trehalose metabolism. Intermediates: *DHAP*dihydroxy acetonephosphate, *Ery4P* erythrose 4-phosphate, *F6P* fructose 6-phosphate, fructose 1,6P_2, fructose 1,6-bisphosphate, *GAP* glyceraldehyde 3-phosphate, *G6P* glucose 6-phosphate, *KD(P)G* 2-Keto-3-deoxy-6-(phospho) gluconate, *KD(P)Gal* 2-Keto-3-deoxy-6-(phospho) galactonate. Enzymes (including EC number): *ACN* aconitase (EC 4.2.1.3), *CS* citrate synthase (EC 2.3.3.1), *ENO* enolase (6; EC 4.2.1.11), *FBPA* fructose-1,6-bisphosphate aldolase (EC 4.1.2.13), *FBPase* fructose-1,6-bisphosphatase (EC 3.1.3.11), *FumR* fumarate hydratase (EC 4.2.1.2), *GA* glucan-1,4-α-glucosidase (EC 3.2.1.3), *GAD*gluconate dehydratase (2; EC 4.2.1.39), *GADH* glyceraldehyde dehydrogenase (4; EC 1.2.1.3), *GAPDH*glyceraldehyde-3-phosphate dehydrogenase (9; EC 1.2.1.12/13), *GAPN* non-phosphorylating GAP dehydrogenase (11; EC 1.2.1.9), *GDH* glucose dehydrogenase (1A; EC 1.1.47), *GK* glycerate kinase (5; EC 2.7.1-), *GL* gluconolactonase (1B; EC 3.1.17), *GLGA* glycogen synthase (EC 2.4.1.11), *GLGP*glycogen phosphorylase (EC 2.4.1.1), *ICL* isocitrate lyase (EC 4.1.3.1), *IDH* isocitrate dehydrogenase (EC 1.1.1.41), *KD(P)GA* KD(P)G aldolase (3; active on KDG as well as KDPG; EC 4.1.2.-), *KDGK* KDG kinase (8; EC 2.7.1.45), *MAE* malic enzyme (EC 1.1.1.38), *MDH* malate dehydrogenase (EC 1.1.1.37), *MS* malate synthase (EC 2.3.3.9), *OOR* α-oxoglutarate ferredoxin oxidoreductase (EC 1.2.7.3), *PEPC* PEP carboxylase (EC 4.1.1.31), *PEPCK* PEP carboxykinase (EC 4.1.1.32), *PEPS* phosphoenolpyruvate synthetase (13; EC 2.7.9.2), *PGAM* phosphoglycerate mutase (12; EC 5.4.2.1), *PGI* glucose-

6-phosphate isomerase (EC 5.3.1.9), *PGK* phosphoglycerate kinase (10; EC 2.7.2.3), *PGM* phosphoglucomutase (EC 5.4.2.2), *PHI/HPS* 3-hexulose-6-phosphate isomerase/3-hexulose-6-phosphate synthase (EC 5.-.-.-/4.1.2.-), *PK* pyruvate kinase (7; EC 2.7.1.40), *POR* pyruvate synthase (EC 1.2.7.1), *PRS* ribose phosphate pyrophosphokinase (EC 2.7.6.1), *PYC* pyruvate carboxylase (EC 6.4.1.1), *RBSK* ribokinase (EC 2.7.1.15),*RPI* ribose-5-phosphate isomerase (EC 5.3.1.6), *SDH* succinate dehydrogenase (EC1.3.99.1), *Succ-CoASyn* succinyl-cenzymA synthetase (EC 6.2.1.5), *TIM* triosephosphate isomerase (EC 5.3.1.1), *TK*transketolase (EC 2.2.1.1), *TreT* trehalose glycosyltransferring synthase (2.4.1.B2), *TreY*maltooligosyltrehalose synthase (EC 5.4.99.15), *TreZ* trehalose hydrolase (EC 3.2.1.141)

Comparative genomics

A comparative genomics approach is used to identify potential transcription factors (TFs) involved in the regulation of the CCM of *S. solfataricus* P2. This analysis basically followed a two-step strategy: first, all putative TFs in the genome of *S. solfataricus* P2 were identified globally. Subsequently, potential CCM regulators were selected by a genomic context scan.

Global identification of putative TFs (SOP_SSO_080909a)

The global identification of putative TFs included different approaches. One source of information was the genome annotation, which was accessed via IMG (Markowitz *et al.* 2008); http:// img.jgi. doe.gov/ and revealed a total of 51 predicted TFs in the genome of *S. solfataricus* P2. In addition to the annotation, two online databases ArchaeaTF (Wu *et al.* 2008); http://bioinformatics.zj.cn/archaeatf/ and DBD (Wilson *et al.* 2008);

www.transcriptionfactor.org/, which both are specialized for the prediction of TFs, were analyzed to receive a more reliable and comprehensive set of predicted TFs. Following this SOP (additional information available at http://www.sulfosys.com), the predicted TFs of the three online databases IMG, ArchaeaTF and DBD were compared and united to a total set of 138 (Fig. 4).

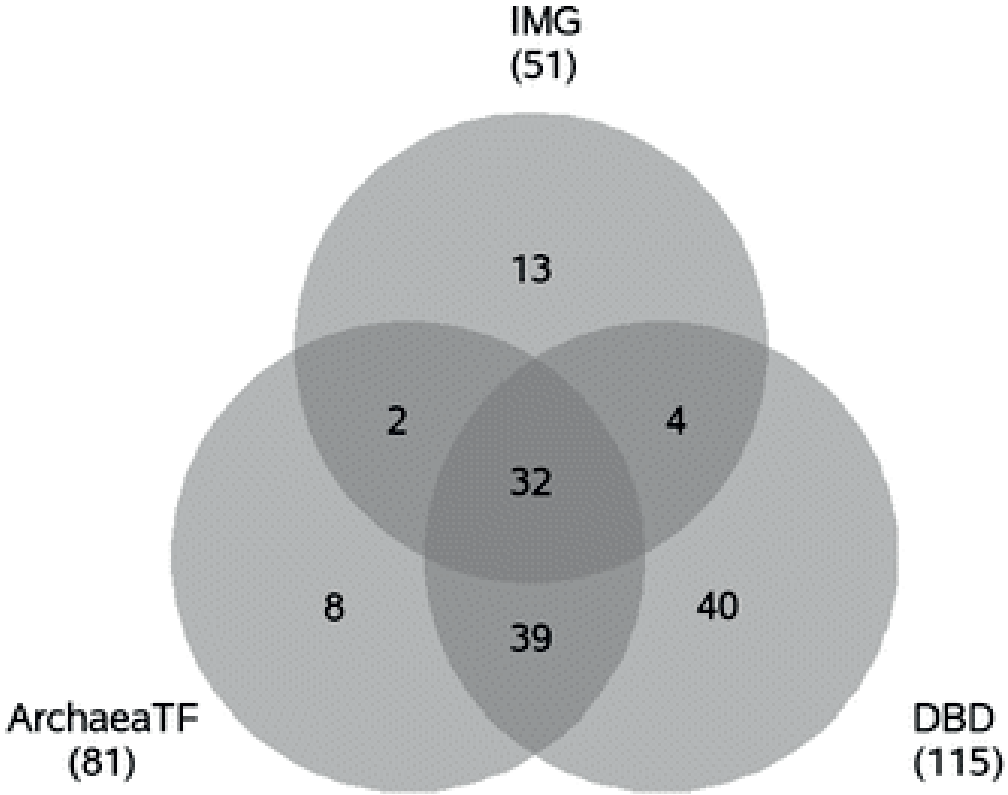

Fig. 4: Venn diagram depicting the overlaps between the predicted sets of TFs in the genome of *S. solfataricus*P2, according to three different online databases. The numbers of predicted TFs in IMG, ArchaeaTF and DBD are 51, 81 and 115, respectively. The total amount of all three databases results in 138 different putative TFs

Identification of putative TFs by psi-BLAST-based approach (SOP_SSO_080909b)

Like in all other prokaryotes with sequenced genomes, not all protein functions of *S. solfataricus* P2 are known. Within the total of 3,048 protein-coding genes, 1,487 (i.e., 49%) are without or with uncertain function prediction, according to the annotation of IMG. In order to identify putative TFs in this fraction of genes, a psi-BLAST-based (Altschul *et al.* 1997) approach was performed. Following this procedure (SOP_SSO_080909b; details available at http://www.sulfosys.com), weak sequence similarities between proteins of unknown function and proteins of reported function in transcriptional regulation could be detected very sensitively.

Context-based approach for identifying putative TFs of the CCM (SOP_SSO_080909c)

The resulting set of 696 psiBLAST predicted TF candidates was examined by a genomic context scan, together with the total of 138 additional TFs which were predicted following SOP_SSO_080909a (see above and supplemental material S4). Here, the genomic neighborhoods of 57 of the identified CCM genes (see SOP_SSO_080908) were searched for the presence of the predicted TF candidates. The results were then manually examined, to determine if the corresponding pair of CCM-gene and TF candidate is likely to be co-transcribed in an operon or co-regulated bidirectionally. This resulted in a set of 81 candidate transcriptional regulators of the CCM, 34 of those are considered to be „strong candidates" for one of the following reasons: (1) the *e* value of a hit between candidate TF and a known transcription factor in the

psi-BLAST-report is smaller than 1e-15, or (2) the candidate TF was predicted by (at least) one of the online databases IMG, ArchaeaTF or DBD.

The psi-BLAST approach detected four genes as candidate TFs, which also belong to the reported CCM-genes: SSO0286, SSO2281, SSO3041 and SSO3226; the latter three are considered to be strong candidates for TFs. These genes possibly have both functions (moonlighting), CCM-gene and TF. One of these four moonlighting candidates, SSO2281 is a glucose-6-phosphate-isomerase and another one SSO3226 is a fructose-1,6-bisphosphate aldolase. For these proteins, moonlighting functions have been reported in Eukaryotes (Jeffery *et al.* 2000; Sherawat *et al.* 2008) }. Although these two proteins are likely to have multiple functions, a role as TF has not been described so far, nor has a DNA-binding property been reported. Experimental verification and available corresponding protein structures, structural comparisons with transcription factors or DNA-binding proteins might give further insight. The other two moonlighting candidates are SSO0286, a fructose-1,6-bisphosphate phosphatase, and SSO3041, a putative gluconolactonase. For these proteins, no further evidence for moonlighting functions was found in the present literature.

Transcriptome analyses

In order to investigate temperature adaptation strategies on the transcriptional level, different methods, i.e., DNA microarray analyses and real-time reverse transcription qPCR are used. The qPCR experiments mainly serve to verify the results obtained from the microarray analyses and a protocol will be available for download from the SulfoSYS homepage (http://www.sulfosys.com).

Microarray analyses

The 70-mer oligonucleotide DNA microarray has been designed and constructed in the group of John van der Oost (Wageningen University, NL, USA) by using the OligoWiz 2.0 (Wernersson & Nielsen 2005) software for oligonucleotide prediction. The array harbors a total of 8,860 spots, including probes for roughly 3,500 *S. solfatricus* genes, which are spotted in duplicate on the array, as well as those of viruses and plasmids of *Sulfolobus*. As negative controls 32 human sequences and 268 targets from *Arabidopsis thaliana* are comprised on the microarray in duplicate. In former studies, the RNA and cDNA preparation techniques had been optimized (Snijders *et al.* 2006; Fröls *et al.* 2007) revealing good and reproducible results with this oligoarray.

Preparation of mRNA from ***S. solfataricus*** cells (SOP_SSO_080910a)

Total RNA is extracted from *S. solfatricus* cells that have been rapidly frozen in liquid nitrogen as described in fermentation protocols (SOP_SSO_080902-5).

For the isolation of *S. solfataricus* mRNA, the MirVana miRNA Isolation Kit (AMBION) according to the instructions of the manufacturer with slight modifications of the protocol is used. Cell pellets harvested from 20 ml of culture at OD_{600} = 0.85(±0.15) are taken from the sample stock. For optimal results all reagents in the initial steps of the protocol are used in double amounts. The samples are separated in two tubes during the acid phenol:chloroform:IAA (125:24:1, Ambion) extraction and proceeded

according to manufacturers protocol. Finally, bound RNA is eluted by using 50 µl of pre-heated (95°C) H_2O instead of 100 µl as recommended by the manufacturer [detailed protocol in supplementary materials (S3)]. RNA concentration is determined by using a Nanodrop RNA protocol (Thermo). The concentration of the prepared mRNAshould be at least 1.3 µg/µl.

cDNA synthesis and labeling by reverse transcription (SOP_SSO_080910b)

Reverse transcription has been performed using a mix of standard nucleotides, with a 1:4 mixture of dTTP and aminoallyl dUTP (Ambion). The 50x aadUTP + dNTP mixture is prepared by dissolving 10 µl each of 100 mM dATP, dGTP, dCTP, 16 µl 50 mM aminoallyl-dUTP (AMBION–AM8439) and 2 µl 100 mM dTTP in 0.1 M KPO_4 (pH 8.0). Single stranded cDNA is generated out of 20 µg total RNA by using a standard protocol for Superscript III (Invitrogen). The reaction is stopped with 4.5 µl 0.1 M EDTA pH 8.0. By the addition of 3 µl 1 M NaOH, followed by further incubation at 70°C for 15 min, the RNA template is degraded. The sample is neutralized by adding 3 µl of 1 M HCl.

The samples are purified by using the Cleanup–MinElute Kit (Qiagen) according to the manufacturer's instructions, except slight modifications: 80% ethanol is used for the wash steps and elution is performed by the addition of $NaHCO_3$ pH 8.6.

For the following labeling reaction using the Alexa dyes 647 and 555 (Invitrogen), cDNA concentration should be at least 80 ng/µl. Quantification is performed using a Nanodrop. For the labeling, add 18.4 µl of the cDNA

sample to 3 µl of appropriate dye dissolved in DMSO and incubate for 1.5 h at RT in darkness.

For purification using the Cleanup–MinElute Kit (Qiagen), combine samples to be co-hybridized. All subsequent steps are performed according to the manufacturer's instructions. The concentration of the pooled and labeled cDNA should be at least 120 ng/µl, as verified by Nanodrop and microarray measurements. In both cases the dye concentrations should be >0.7 pmol/µl.

Hybridization (SOP_SSO_080910c)

Prior to hybridization of the labeled cDNA to the microarrays, the slides are pre-hybridized in pre-warmed 5 × SSC containing 0.1% SDS and 10 µg/ml BSA, at 42°C for 40 min. Afterwards, the slides are washed thoroughly (30 s steps) in three Coplin jars with A.bidest. followed by briefly dipping them in isopropanol. Finally, the slides are dried in Microarray High-Speed Centrifuge (MHC, Arrayit; 2,000×*g*, 30 s, RT) and used for hybridization within 1 h.

For hybridization, 17.4 µl of the labeled cDNA is mixed with 1 µl tRNA (10 µg/µl), 1 µl herring sperm DNA (10 µg/µl) and 42.6 µl hybridization mixture containing 27 µl deionized formamide, 15 µl 20 ×SSC and 0.75 µl SDS (10%). The sample is incubated for 2 min at 95°C and subsequently cooled on ice for 1 min.

After quick-spin (10,000×*g*, 10 s, RT) the sample is applied on a slide (under a lifterslip). A.bidest (15 µl) is added to appropriate wells in the hybridization chamber to prevent evaporation. The slides are sealed for incubation at 42°C in darkness for 16–20 h. Afterwards, the slides are incubated in 2 ×SSC, 0.1%

SDS for 5 min and in 0.1 ×SSC, 0.1% SDS for 20 min (both steps performed in the dark at 42°C). Later slides are washed 5× in Coplin jars containing 0.1 ×SSC and finally dried by centrifugation in MHC (2,000×*g*, 30 s, RT).

Scanning, extraction features, normalization and data analyses (SOP_SSO_080910d)

Each hybridization experiment using the 70-mer oligonucleotide DNA array has been performed as a dye swap, which provides a mean to exclude spots, where hybridization errors occur. Scans are performed with the GenePix Pro 4000B scanner (Axon). In a first scan of each array, 60% of laser intensity and in a second scan only 10% of laser intensity have been used, in order to be able to determine the proper ratios in spots saturated at 60%.

Features are extracted with GenePixPro 6.0 software (Axon) and flagged bad if intensities are below 3 times of the background in case of both dyes.
A feature is also excluded from further analysis, if the R^2 of the spot is <0.6, which indicates lack of homogeneity of the spot. Results acquired in the form of *.gpr file are converted to *.mev and normalized using Midas software (TIGR). The main normalization tool is Lowess (Quackenbush 2002; Yang *et al.* 2002) and log mean centering. By this means, extracted and normalized data can be transferred to Microsoft Excel sheets that allow for quick analysis and annotation of the data. Since the main interest is in up- and down-regulated genes, which corresponds to log2 ratio values >1 and <−1, respectively, the initial confirmation of statistical soundness of the data can be performed using *Z* test, testing if population of results with a given standard deviation is higher or lower than input value. By setting the input values at 1 and −1 we can statistically assess significance of the up-regulation of a given

gene (for value >1, *z* value ≤ 0.05; for value <1, *z* value ≥ 0.95). Further analysis can be performed using SAM analysis in MeV program (Tusher *et al.* 2001).

Results

The pilot experiment involving transcriptomics has been performed by comparing cells grown in batch fermenter cultures at 80 and 70°C. Two biological samples have been used and a total of four microarrays have been hybridized. It has been assumed that log2 ratios higher than 1 and lower than −1 indicate significant fluctuation of the gene expression of the gene. Upregulation has been assessed using the *Z* test with 95% confidence level. Apart from the set of regulated genes, all genes involved in CCM have been compared.

In total, 24 genes are significantly up-regulated at 80°C and 43 genes are down-regulated. The up-regulated genes include a superoxide dismutase, indicating higher presence of reactive oxygen intermediates at higher temperature. Furthermore, *nadA* gene was overexpressed, suggesting higher rate of NAD synthesis. Other annotated genes include those coding for a large subunit of the replication factor C (RFC), a transcription activator in the thiamine synthesis pathway (*tenA*-2) and a small heat shock protein from hsp20 family. Four genes up-regulated are involved in amino acid synthesis, transport and proteolysis, suggesting scavenging of the dead cell material from the culture.

Surprisingly, the biggest group of down-regulated genes at 80°C consists of small and large subunit ribosomal genes (Table 2). A total of ten ribosome-

related genes are down-regulated. This may indicate that in suboptimal conditions protein synthesis is one of the limiting factors for the population growth. It has to be noted here that nine of them are found in a large operon, which tend to have lower stability. It has been shown (Andersson *et al.* 2006) that all of these transcripts have a half life of no longer than 3 min. Another interesting finding is the down-regulation of the γ subunit of the thermosome (Table 2), which is consistent with findings of (Kagawa *et al.* 2003). Other genes include two subunits of the cytochrome c complex, two putative RNA helicases related to deaD family (Table 2) There are also six genes coding for putative ABC transporter binding proteins, which are downregulated at 80°C (Table 2). This might indicate scavenging debris from cells that die due to cold shock, as two of the transporters are binding sugars not present in the medium, in which cells have been grown (arabinose and maltose) and other two bind dipeptides. The remaining two transporters have not yet been assigned a function, but based on sequence similarity they might play a role in oligosaccharide uptake. Other candidates have no assigned function or are distantly related to proteins from other species.

Table 2: Significantly regulated genes comparing growth at 80 versus 70°C revealed from transcriptomic analysis

Gene ID	Annotation	80 versus 70°C log2 ratio (±SD)
SSO0068	SSU ribosomal protein S9AB (rps9AB)	−1.29 (±0.38)
SSO0489	Phosphate binding periplasmic protein precursor (pstS)	−1.91 (±0.25)
SSO0697	LSU ribosomal protein L30AB (rpl30AB)	−1.85 (±0.84)
SSO0698	SSU ribosomal protein S5AB (rps5AB)	−2.07 (±0.70)
SSO0700	LSU ribosomal protein L19E (rpl19E)	−1.73 (±0.67)
SSO0704	LSU ribosomal protein L5AB (rpl5AB)	−1.44 (±0.35)
SSO0707	LSU ribosomal protein L24AB (rpl24AB)	−1.60 (±0.60)
SSO0716	LSU ribosomal protein L2AB (rpl2AB)	−1.73 (±0.72)

SSO0718	LSU ribosomal protein L4AE (rpl4AE)	−1.25 (±0.29)
SSO1274	Oligo/dipeptide transport, permease protein (dppB-1)	−1.80 (±0.74)
SSO1275	Oligo/dipeptide transport, permease protein (dppC-1)	−1.19 (±0.27)
SSO1889	ATP-dependent RNA helicase	−1.74 (±0.73)
SSO2036	ATP-dependent RNA helicase	−1.26 (±0.24)
SSO3000	Thermosome gamma subunit	−2.11 (±0.60)
SSO3043	ABC transporter, binding protein	−2.05 (±0.99)
SSO3047	ABC transporter, permease	−1.37 (±0.55)
SSO3053	Maltose ABC transporter, maltose binding protein	−2.29 (±0.85)
SSO3066	Arabinose ABC transporter, arabinose binding protein	−1.51 (±0.61)
SSO3120	Metabolite transport protein, putative	−1.69 (±0.94)
SSO3198	Muconate cycloisomerase related protein	−1.28 (±0.49)
SSO6391	SSU ribosomal protein S14AB (rps14AB)	−1.44 (±0.53)
SSO6401	LSU ribosomal protein L23AB (rpl23AB)	−1.85 (±0.64)
SSO2088	Peptidase, putative	1.12 (±0.12)
SSO0316	Superoxide dismutase [Fe] (sod)	1.17 (±0.20)
SSO2603	Small heat shock protein hsp20 family	1.33 (±0.52)
SSO2598	Transcriptional activator (tenA-2)	1.35 (±0.52)
SSO0998	Quinolinate synthetase (nadA)	1.99 (±0.27)
SSO2549	Amino acid transporter, putative	2.27 (±0.45)
SSO0769	Activator 1, replication factor C (RFC) large subunit (rfcL)	2.56 (±0.89)

A log2 ratio >1 indicates up-regulation at 80°C, log2 < −1 indicates down-regulation at 80°C. For all genes *Z* test values ≤0.05

SD standard deviation

Of the 97 genes hypothesized to be involved in the CCM network, 91 have been found using the transcriptome analysis. Most genes do not show statistically significant differential expression. The genes of the branched ED pathway (Fig. 3) also do not show differential expression between the two conditions with the exception of SSO3198 coding for gluconate dehydratase and SSO3194 encoding the non-phosphorylating glyceraldehyde 3-phosphate dehydrogenase (GAPN) (Table 3). The encoding genes are twofold down-regulated at 80°C. They are located in the ED operon SSO3198-3197-3195-3194; (Ahmed *et al.* 2005), and the other genes from

the same cluster indicate a similar regulation (with the exception of SSO3195 KDG kinase; Table 3). Also the proteomic data (SOPs_SSO_080911) show no significant differences except for the GAPN, which is in accordance to the transcriptomic data, downregulated at 80°C at the proteomic level (Table 3). These first results suggest that the regulation of the CCM in *S. solfataricus* is placed on different regulatory levels.

Table 3: Results of the initial transcriptomic and proteomic analyses of the glycolytic, branched ED pathway of *S. solfataricus* in response to growth at 80 versus 70°C

Gene ID	Reaction no. Fig. 3	Gene product	EC no.	Transcriptomics 80 versus 70°C log2 ratio (±SD)	Proteomics 80 versus 70°C log2 ratio (±SD)
SSO3003	1A	Glucose-1-dehydrogenase (GDH)[a]	1.1.1.47	−0.34 (±0.11)	NF
SSO2705	1B	Gluconolactonase (GL)	3.1.1.17	−0.16 (±0.20)	0.34 (±0.06)
SSO3041	1B	Gluconolactonase (GL)	3.1.1.17	−0.42 (±0.32)	NF
SSO3198	2	Gluconate dehydratase (GAD)[b]	4.2.1.39	−1.28 (±0.49)	−0.44 (±0.06)
SSO3197	3	2-keto-3-deoxy-(6-phospho)-gluconate/galactonate aldolase (KD(P)GA)[b]	4.1.2.-	−0.78 (±0.15)	−0.27 (±0.60)
SSO2636	4	Aldehyde ferredoxin oxidoreductase, β-subunit (AOR)	1.2.7.-	−0.54 (±0.23)	0.29 (±0.04)
SSO2637	4	Aldehyde ferredoxin oxidoreductase, γ-subunit (AOR)	1.2.7.-	−1.12 (±0.53)	0.36 (±0.17)
SSO2639	4	Aldehyde ferredoxin oxidoreductase, α-subunit (AOR)	1.2.7.-	−1.28 (±0.88)	−0.05 (±0.10)
SSO0666	5	Glycerate kinase (GK)	2.7.1.-	−0.45 (±0.21)	−0.40 (±0.14)

SSO0913	6	Enolase (ENO)	4.2.1.11	0.02 (±0.09)	−0.25 (±0.21)
SSO0981	7	Pyruvate kinase (PK)	2.7.1.40	0.63 (±0.43)	0.07 (±0.13)
SSO3195	8	2-keto-3-deoxy-gluconate/galactonate kinase (KDGK)[b]	2.7.1.45	−0.09 (±0.21)	NF[b]
SSO0528	9	Glyceraldehyde-3-phosphate (GAP) dehydrogenase (GAPDH)	1.2.1.12/13	−0.12 (±0.32)	0.62 (±0.13)
SSO0527	10	Phosphoglycerate kinase (PGK)	2.7.2.3	−0.50 (±0.44)	0.45 (±0.16)
SSO3194	11	Non-phosphorylating GAP dehydrogenase (GAPN)[c]	1.2.1.9	−1.18 (±0.44)	−1.47 (±0.65)
SSO0417	12	Phosphoglycerate mutase (PGMA)	5.4.2.1	−0.51 (±0.36)	−1.36 (±0.47)
SSO0883	13	Phosphoenolpyruvatesynthetase (PEPS)	2.7.9.2	−0.65 (±0.37)	−0.40 (±0.20)

A log2 ratio >1 indicates up-regulation at 80°C, log2 < −1 indicates down-regulation at 80°C. For all genes*Z* test reaveld values ≤0.05

SD standard deviation, *NF* not found [a](Lamble *et al.* 2003) [b](Ahmed *et al.* 2005) [c](Ettema *et al.* 2008)

Proteome analyses

In course of the SulfoSYS-project one goal is to quantitatively measure and understand protein expression changes, protein interaction networks, non-covalent interactions and post-translational modifications of the CCM proteins of *S. solfataricus* in response to temperature changes.

Different approaches for protein quantitation for membrane proteomes are applied within this project, since membrane proteins play most important roles

during cell life. The iTRAQ method is used for global expression profiling, to compare up to eight fully adapted cell states.

Cellular extraction (SOP_0809011a)

Frozen cells are firstly washed twice with ice-cold water, then they are centrifuged at 6,000×*g* before being resuspended in 1 mL of extraction buffer, which contains 43 mM NaCl, 81 mM $MgSO_4$ and 27 mM KCl (Bisle *et al.* 2006). Protein extraction is carried out using an ultra sonicator (Sonifier 450, Branson) 4 times (alternatively 1 min of sonication and 1 min on ice) at 70% duty cycle. Samples are then centrifuged at 3,000×*g* for 5 min at ×4°C to discard unbroken cells and debris, the supernatant is collected before centrifugation again at 100,000×*g* for 90 min 4°C using a sucrose gradient detailed as elsewhere (Bisle *et al.* 2006). The pellets are collected as enriched membrane fractions. These membrane fractions are then delipidated using chloroform/methanol as detailed by (Wessel & Flügge 1984) with some modifications. Briefly, the membrane is resuspended in 400 µl of methanol, vortexed at 1,500 rpm for 30 s and centrifuged at 9,000×*g* for 20 s at room temperature. The pellet is collected by discarding the supernatant, then resuspended in 100 µl of chloroform and 1,500 rpm for 30 s, and centrifuged at 9,000×*g* for 20 s room temperature. The recovery of membrane is performed using phase separation, where 300 µl of water is added to the sample, followed by 1,500 rpm for 30 s and centrifugation at 9,000×*g* for 90 s. While the upper phase is discarded carefully, 300 µl of methanol are added to the interphase (containing precipitated proteins) and lower phase. This sample is mixed by vortexing at 1,500 rpm for 1 min, followed by centrifugation at 9,000×*g* for 2 min to pellet membrane proteins. The pellet is collected by discarding the supernatant and then drying in a vacuum concentrator before

being resuspended in 100 µl of 0.5 M TEAB pH 8.5 buffer containing 0.095% SDS. The sample is dissolved totally by sonicating for 5 min before the total protein concentration is determined using the RC-DC Protein Quantification Assay (Bio-Rad, UK). This sample is then ready for the iTRAQ labeling step. For soluble protein analysis, cells are resuspended in 0.5 M TEAB pH 8.5 before being extracted as detailed above.

iTRAQ labeling (SOP_0809011b)

A total of 100 µg protein of each phenotype is used for iTRAQ analysis. Protein samples are reduced, alkylated, digested and labeled with iTRAQ reagents according to the manufacturer's protocol (Applied Biosystems, USA). Briefly, samples are reduced by adding 2 µl of 50 mM tris-(2-carboxyethyl) phosphine (TCEP) and incubating at 60°C for 1 h; then cysteines are alkylated with 1 µl of 200 mM methyl methanethiosulfonate (MMTS) for 10 min at room temperature. The digestion step at 37°C overnight is carried out using trypsin MS grade (Promega, UK) with the ratio of trypsin:proteins 1:20. Then these samples were labeled with iTRAQ reagents in isopropanol (or ethanol). After incubation at room temperature for 4 h, labeled samples were combined before being dried in a vacuum concentrator.

In the case of the combination of both, trypsin and chymotrypsin, for the digestion step, samples are firstly digested with trypsin on the first day (at a ratio of 1:40) and then a mixture of chymotrypsin and trypsin (ratio enzyme: protein = 1:40 for each) on the second day. After digestion by trypsin, the partially digested sample is centrifuged at 13,000×*g* for 1 h at room temperature to pellet undigested proteins, then, while supernatant was collected and transferred to a new tube, the pellet is resuspended again in

methanol before a mixture of trypsin and chymotrypsin is added (refer to (Fischer *et al.* 2006) for chymotrypsin digestion details). The sample is then incubated overnight at 37°C. After digestion, this sample is centrifuged again at 13,000×*g* to pellet undigested proteins, the supernatant is collected and mixed with the previous trypsin digested supernatant. The mixture of digested peptides is then dried in a vacuum concentrator before being resuspended in 30 µl of 0.5 M TEAB pH8.5 for the iTRAQ labeling step. To enhance the protein digestion step for the membrane fractions, the use of sodium deoxycholate (SDC) with a final concentration of 0.007% has also been applied (see (Masuda *et al.* 2008) for more detail).

Strong cation exchange (SCX; SOP_0809011c)

The dried iTRAQ samples are resuspended in buffer A (details below) and then fractionated using a SCX technique on a BioLC HPLC system (Dionex, UK) to clean the sample, as well as reduce its complexity. The SCX fractionation is carried out using a PolySulfoethyl A column (PolyLC, USA) 5 µm particle size in a length of 20 cm × 2.1 mm in diameter, 200 Å pore size. The system is operated at a flow rate of 0.2 ml/min, and with an injection volume of 120 µl. The mobile phase is used consisting of buffers A and B. While buffer A contains 10 mM KH_2PO_4, 25% acetonitrile, pH3, buffer B consists of 10 mM KH_2PO_4, 25% acetonitrile and 500 mM KCl, pH3. A gradient of 60 min is used, 5 min at 100% buffer A, followed by ramping from 5 to 30% buffer B for 40 min, 30–100% B over 5 min and finally 100% A for 5 min. A UV detector UVD170U and Chromeleon Software (Dionex, The Netherlands) are used to record the chromatogram. Labeled peptide fractions are collected every minute, subsequently each fraction is dried in a vacuum concentrator.

Mass spectrometry analysis (SOP_0809011d)

Selected dried labeled peptides samples are redissolved in 50 µl of buffer A consisting of 0.1% formic acid and 3% acetonitrile, and then MS analysis is performed on a QStar XL Hybrid ESI Quadrupole time-of-flight tandem mass spectrometer, ESI-qQ-TOF–MS/MS (Applied Biosystems, Canada), coupled with a nano-LC system comprising a combination of a LC Packings Ultimate 3000 (Dionex, UK). An injection of 15 µl of sample is submitted to the nano-LC–MS/MS system. The LC gradient is operated at a flow rate of 300 µl/min, consisting of 5% buffer B (0.1% formic acid and 97% acetonitrile) to 30% buffer B over 85 min, followed by a 5 min ramp to 95% buffer B, and then 10 min at 5% buffer B. The ESI–MS detector mass range is set at 350–1800 m/z. The MS data acquisition is performed in the positive ion mode. During the scan, peptides with a +2, +3, or +4 charge state are selected for fragmentation, and the time for summation of MS/MS events is set up at 3 s.

Data searching (SOP_0809011e)

MS/MS data are analyzed using Phenyx software v.2.6 (Geneva Bioinformatics, Switzerland) with the *S. solfataricus* P2 protein database (2977 ORFs) downloaded June 2007 from NCBI (http://www.ncbi.nlm.nih.gov/). The search parameters for peptides and MS/MS tolerance are as follows: 0.2 Da peptide tolerance, default parent charge were +2, +3 and +4 with trust parent charge: yes. Acceptance parameters are set as following: minimum peptide length, peptides *z* score, maximum *P* value and AC score were 5, 5, 10^{-5} and 5, respectively. Fixed modifications of MMTS, cys_CAM, iTRAQ_K, iTRAQ_Ntermi are used, and enzymes used for searching are trypsin alone or a combination of trypsin and chymotrypsin (in Experiment 3) with one missed cleavage for both. The results are exported to Excel

(Microsoft 2008, USA) for further analyses. Although Phenyx software is used for searching and exporting data, the data analysis is carried out as suggested by the Protein Pilot v2.0 software documentation (Applied Biosystems, USA), since Phenyx does not automatically calculate iTRAQ quantitation. All peptides are converted to $\log_{10}$ space before the calculation of the protein ratio is applied, as per the equation adapted from the Protein Pilot software documentation. Subsequently, the correcting of the bias median ratio of each protein is also applied. Moreover, the estimation of false determination rate is also carried using spectra derived from a decoy databases (generated from *S. solfataricus* reversed sequences) as described by (Elias & Gygi 2007). We adjusted parameters for MS/MS searching to get the false determination rate (for each experiment) less than 0.2%.

Results

Protein identification for quantitative membrane proteomic analysis of ***S. solfataricus***

In this investigation, three different iTRAQ-8plex experiments have been analyzed for enriched membrane fractions, including one experiment carried out as suggested by the original protocol (Experiment 1), and two experiments for modified protocols (Experiment 2 for trypsin and chymotrypsin, Experiment 3 trypsin and chymotrypsin with the presence of SDC). Cells grown at 80°C have been used as the controls and labeled with iTRAQ reagents 118, 119 and 121 (119 and 121 used as an independent biological replicate whilst 118 and 119 used as technical replicate), and samples at 70°C were labeled with reagents 115, 116 and 117 (115 and 116 used as an independent biological replicate, 116 and 117 used as a technical replicate).

As a result, the numbers of proteins detected for three different iTRAQ experiments are shown in Fig. 5. It is clear that more proteins were detected for Experiments 2 and 3 as a result, more membrane proteins and trans-membrane proteins were also detected for Experiments 2 and 3 compared to Experiment 1 (for more details see Fig. 5). These data agree with a previous study, since more membrane proteins were found with the presence of SDC (Masuda *et al.* 2008). There also seems to be more membrane and transmembrane proteins being found in Experiment 3 compared to Experiment 2 (for more details see Fig. 6). Moreover, in term of cell localization, the highest number of integral membrane proteins was identified for Experiment 3.

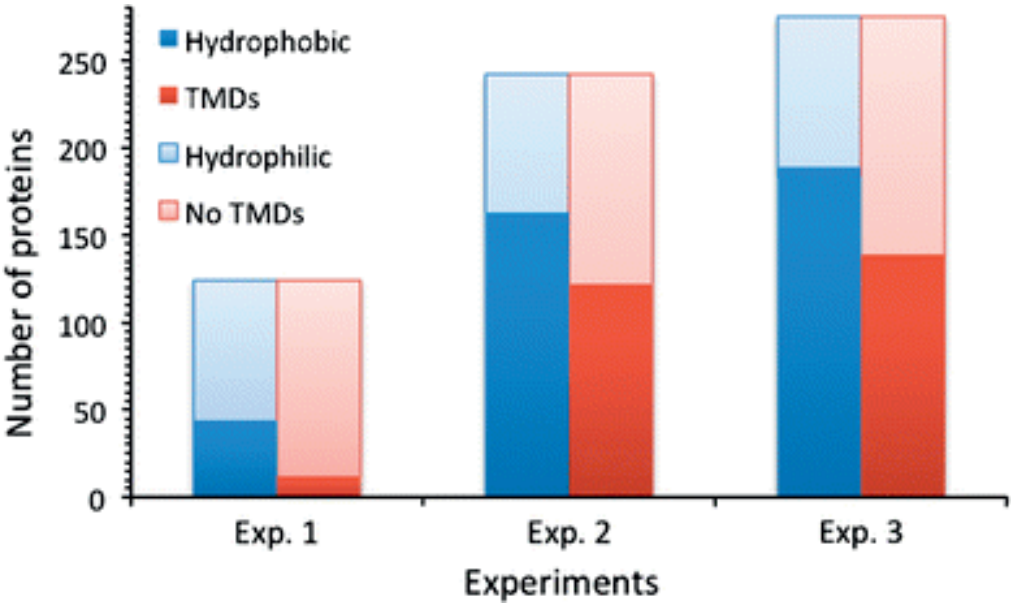

Fig. 5: Number of proteins detected in the three different iTRAQ experiments. The identification of these proteins' membrane properties based on hydrophobic (*dark blue*) and transmembrane domains (TMDs, *dark red*) found, are shown

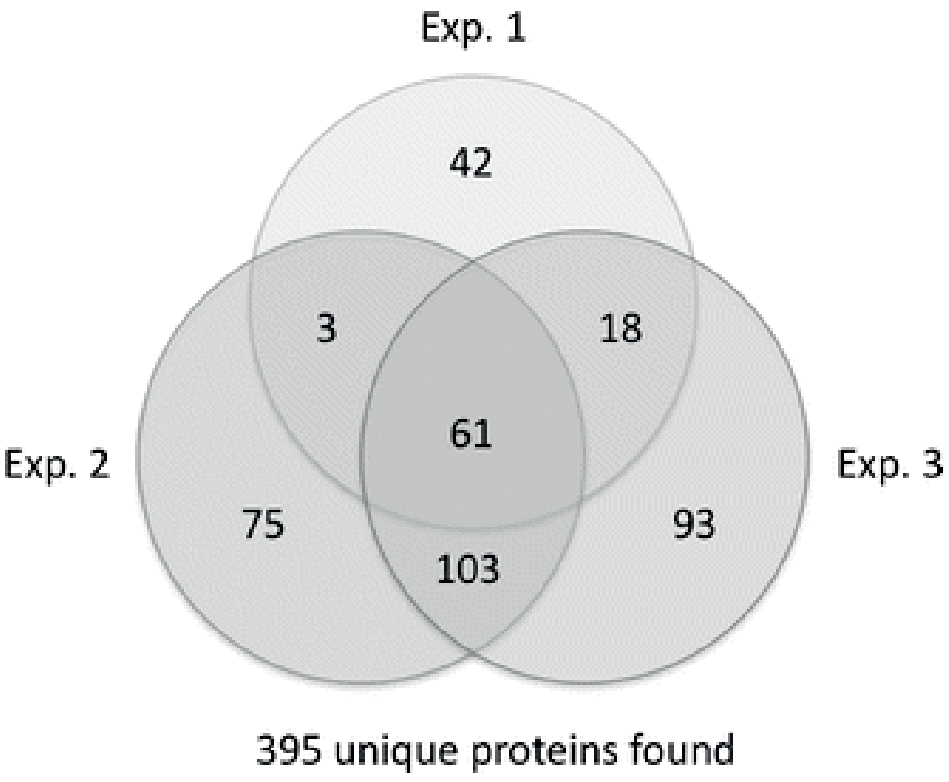

Fig. 6: Total numbers of proteins detected for enriched membrane fractions from three different iTRAQ experiments. Peptide detection

Therefore, we can assert that the combination of both SDC and chymotrypsin for trypsin digestion is suitable for *S. solfataricus* integral membrane proteins. A slightly increased total number of detected proteins are also found in Experiment 3, because more peptides are released during the digestion step, when using a combination of trypsin and chymotrypsin with a presence of SDC.

By combining proteins detected in all three different iTRAQ experiments for enriched membrane fractions 395 proteins were found as shown in Fig. 6.

For bottom-up proteomic analysis, the identification and quantitation of protein are based on peptide-level assignments; therefore, it is necessary to discuss this issue here. The numbers of distinct peptides detected for each experiment are 749, 1374 and 1635 for Experiments 1, 2 and 3, respectively.

Since SDS and SDC are applied in this study, and these compounds are known to be unfriendly compounds for mass spectrometry, and excess amounts of these compounds affect the labeling step. Therefore, we evaluated the affect of these chemicals to the iTRAQ labeling step, as well as nano-LC MS/MS operation via the efficiency of iTRAQ labeling, where the evaluation was calculated based on the percentage of labeled peptides compared to the total number of detected peptides (labeled and unlabeled peptides). However, we could not detect any difference within these experiments, since there were a small percentage of unlabeled peptides being detected; actually only two unlabeled peptides were solely identified in Experiment 3. Therefore, we can conclude that the SDC concentration used in this study was acceptable for the iTRAQ labelling step.

Membrane proteins

As discussed above, more peptides than proteins are detected for enriched membrane fractions in Experiments 2 and 3. To ensure that all proteins detected here contained membrane properties, these proteins were examined based on membrane properties including hydrophobic (Gravy score), TMDs found (TMHMM, http://www.cbs.dtu.dk/services/TMHMM/) and cell localization (http://www-archbac.u-psud.fr/projects/sulfolobus/). As a result, of 395 merged proteins (from all 3 experiments), 373 proteins were found to be

membrane proteins, where 233 were proteins observed with more than two different membrane properties.

In summary, we have applied successfully iTRAQ for *S. solfataricus* (P2) quantitative membrane proteomic analysis (Fig. 7), since of 284 proteins detected, 246 proteins were found as membrane proteins. A merged data from all different iTRAQ data led to 395 unique proteins were detected, in which 373 were found as membrane proteins. All merged proteins from iTRAQ experiments and more details about membrane proteins' regulations can be found in "*Quantitative Proteomic Analysis of Sulfolobus solfataricus Membrane Proteins*" (Pham *et al.* 2010).

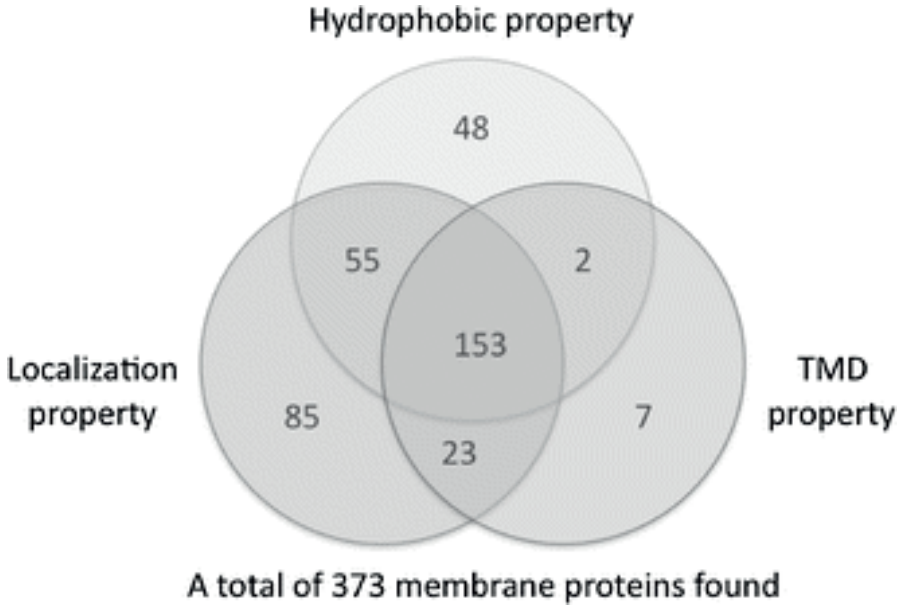

Fig. 7: Classification of merged proteins base on membrane properties

Metabolome analyses

The metabolic composition reflects the set of metabolites within a cell at a certain timepoint. Metabolites take part in regulatory mechanisms, directly in allosteric regulation of enzyme activities but also indirectly by influencing transcriptional and translational control. Therefore, the integration of metabolome data (relative metabolite concentrations) can (i) highlight regulatory mechanisms taking place due to the temperature change, (ii) help to complete functional gene annotations by identification of missing enzymatic activities, (iii) being used in order to identify and analyze specific metabolic pathways and, (iv) provide data for the computational cell simulations.

First quantitative analysis of changes of metabolite concentrations due to temperature changes comparing 80 versus 70°C have been performed with cell mass derived from batch flask fermentation (SOP_SSO080903; Tables 4 and 5). In addition, exometabolome analyses have been performed, comprehending all metabolites that are excreted into the growth medium and therefore depict a picture of the metabolome during a period of metabolic and biological activity prior to sampling.

Table 4: Ratios of detected metabolites in samples derived from cells grown at 80 versus 70°C (CCM compounds and metabolites of amino acid and nucleic acid metabolism as well as of glycosylated protein and lipid biosynthesis. Higher metabolite concentrations at 70°C are indicated in bold fonts and lower concentrations at 70°C are itaclicized. Others represent no significant changes).

Metabolites	Ratio
CCM metabolism	
KDG/KDGal	**0.11**
Glyceraldehyde	**0.58**
Citrate	*3.13*
3-Phosphoglycerate	*2.86*
Succinate	1.75
Glycerate	1.56
Glucose 6-phosphate	1.51
Trehalose	1.45
Glucose	1.33
Fructose 6-phosphate	1.25
Malate	1.18
Fumarate	1.11
Galactose	0.09
Pyruvate	NF
2-Oxoglutarate	NF
Glucono-1,5-lactone	NF
Glucose-1-phosphate	NF
Dihydroxyacetonphosphate	NF
2-Phosphoglycerate	NF
Phosphoenolpyruvate	NF
Fructose 1,6-bisphosphate	NF
1,3 Bisphosphoglycerate	NF
Glyceraldehyde 3-phosphate	NF
Isocitrate	NF
Oxaloacetate	NF
KDPG/KDPGal	Not available

Table 5: Ratios of detected metabolites in samples derived from cells grown at 80 versus 70°C

Metabolites	Ratio

Metabolites	Pathway	Ratio
Other metabolites		
Valine	Amino acid metabolism	0.12
Isoleucine	Amino acid metabolism	0.10
Glucosamine	Precursor of glycosylated proteins and lipids	0.16
Leucine	Amino acid metabolism	0.19
Spermidine	Nucleic acid and protein synthesis	0.21
Alanine	Amino acid metabolism	0.31
Thymine	Pyrimidine metabolism	0.35
Putrescine	Amino acid metabolism	0.39
Glutamic acid	Amino acid metabolism	0.40
Lysine	Amino acid metabolism	0.42
Threonine	Amino acid metabolism	0.57
Aspartic acid	Amino acid metabolism	0.62
Beta-Alanine	Amino acid metabolism	2.50
Glycine	Amino acid metabolism	1.61
Serine	Amino acid metabolism	2.32
Phenylalanine	Amino acid metabolism	3.70

As one important prerequisite for the set-up of the protocols for *S. solfataricus* metabolome analysis, cell growth and handling of the organism have been performed according to the developed SOPs (SOP_SSO080902-4).

However, a special protocol for cell treatment directly after harvest by centrifugation had to be established (SOP_SSO_080912a).

Sample preparation (SOP_SSO_080912a)

Cell mass is obtained from batch fermentation (SOP_SSO_080903). 20 mg cell dry weight (that is equivalent to $38/OD_{600}$ nm = x ml *S. solfataricus* culture) is harvested by centrifugation (4,629×*g*, 5 min, 25°C; 5810 R, Eppendorf). After harvesting, the cell pellet is resuspended (by shaking) in 20 ml 0.9% NaCl (w/v) at RT and washed twice (4,629×*g*, 3 min, 25°C; 5810 R, Eppendorf).

Subsequently, cells are resuspended in 1.5 ml methanol (containing 60 µl ribitol (c = 0.2 g l^{-1}) and lyzed in an ultrasonic bath for 15 min at 70°C. Afterwards, the sample is incubated on ice for 2 min, 1.5 ml of deionized water is added and the sample is vortexed. For extraction of metabolites 1 ml chloroform is added and the sample is mixed by vortexing. After centrifugation (4,629×*g*, 5 min, 4°C; 5810 R, Eppendorf) the upper, polar phase is transferred into a fresh tube (2 ml) and dried in a vacuum concentrator (SpeedVac, Eppendorf) for 1 h with rotation and overnight without rotation. Final step is the derivatization of the metabolites for subsequent GC–MS analysis: Hereunto, 20 µl pyridine, containing 20 mg ml^{-1} methoxyamine hydrochloride are added to the dried sample (vortex for 1 min). After incubation in a thermomixer (600 rpm, 90 min, 30°C; Thermomixer *comfort*, Eppendorf) 32 µl *N*-methyl-*N*-trimethylsilyltrifluoroacetamide (MSTFA) is added (vortex for 1 min). Samples are incubated again for 30 min at 37°C (shaking speed 600 rpm) followed by 120 min at 25°C (shaking speed 600 rpm). After subsequent centrifugation (18,400×*g*, 5 min, RT; 5424, Eppendorf)

50 µl of the sample are transferred in a glass vial containing a micro cartridge for GC–MS analysis.

For exometabolome analysis cells of a *S. solfataricus* batch culture are grown on 0.15% glucose (instead of 0.3%) and harvested in the exponential growth phase by centrifugation (4,629 × *g*, 5 min, 25°C, 5810 R, Eppendorf). The supernatant is collected and 40 µl ribitol (c = 0.2 g l^{-1}) as internal standard are added to 500 µl of culture supernatant. Subsequently, the sample is transferred in a 2 ml eppendorf tube and dried in a vacuum centrifuge (SpeedVac, Eppendorf) for 1 h with rotation and overnight without rotation. Afterwards metabolites are derivatized for GC/MS analysis (SOP_SSO_080912a) that is performed following SOP_SSO_080912b.

GC–MS analysis (SOP_SSO_080912b)

The system consists of a TRACE mass spectrometer coupled to a TRACE gas chromatograph with an AS 3000 autosampler (all devices from Thermo Finnigan GmbH, Egelsbach, Germany). The system operates under the Xcalibur software (version 1.2, Thermo Finnigan GmbH, Egelsbach, Germany). Positive electron ionization (EI +) mode at 70 eV is used for ionization. Tuning is done according to the operating manual using perfluorotri-N-butylamine (Fluorochem Ltd., Derbys, UK) as reference gas. Full scan mass spectra are acquired from 40 to 800 m/z with a scan rate of 2/s and a solvent delay time of 6 min. The chromatography was performed using a 30 m, 0.25 mm, 0.25 µm film thickness, DB-5MS column (J&W Scientific, Folsom, USA) with a helium flow of 1 ml min^{-1}. For measurements a derivatized sample volume of 2 µl was injected in split mode (25:1) at 70°C and the solvent was evaporated in 0.2 min. Injections were made using a

programmed temperature vaporizer (PTV) injector supplied with a 12 × 2 mm glass liner manually filled with glass wool (Restek GmbH, Bad Homburg, Germany). For sample transfer the temperature was increased to 280°C at a rate of 14°C s^{-1} followed by an additional constant temperature period at 280°C for 2 min. The oven temperature is increased at 1°C min^{-1} to 76°C and then with 6°C min^{-1} to 325°C, after 10 min isothermal cool-down to 70°C.

Results

A total of 70 metabolites from widely different metabolic pathways can be detected in the exponential growth phase for *S. solfataricus* (Table S1, supplemental material). Derived data have been compared to available bacterial metabolome data. The most obvious difference is that *S. solfataricus* shows a much smaller number of metabolites compared to Bacteria, such as *Corynebacterium glutamicum* (Strelkov *et al.* 2004) or *Pseudomonas aeruginosa* (Frimmersdorf et al., unpublished). These data are of special interest, because to our knowledge this is the first metabolome analysis for a thermoacidophilic organism.

Some of the detected metabolites in samples derived from cells grown at 80°C (optimal growth temperature) and 70°C show differences in relative concentrations (Tables 4 and 5). Especially some amino acids have considerably increased concentrations at the lower growth temperature (70°C). Valine, leucine, isoleucine, alanine, aspartic acid, lysine, threonine and glutamic acid have been detected in higher concentrations at 70°C. In accordance with this finding, an up-regulation of genes and proteins involved in amino acid biosynthesis at lower cultivation temperatures than 80°C has been observed by the transcriptomic and proteomic analyses (70°C) and has

been reported previously for the hyperthermophilic euryarchaeon *Pyrococcus furiosus* (Weinberg *et al.* 2005).

Interestingly, the polyamines putrescine and spermidine are detected in high concentrations in *S. solfataricus* and it has previously been shown that polyamines play an important role in stabilizing DNA and RNA at high temperatures in the hyperthermophilic bacterium *Thermus thermophilus* (Cava *et al.* 2009). However, from the comparison of *S. solfataricus* cells grown at 80 versus 70°C putrescine and spermidine are detected in higher amounts in cells grown at 70°C.

In contrast, the CCM metabolism shows only small differences in metabolite concentrations comparing growth at 80 versus 70°C. Citrate and 3-phosphoglycerate are present in lower concentrations, whereas glyceraldehyde and 2-keto-3-deoxy gluconate (KDG) are detected in higher concentrations at 70°C.

The exometabolome analysis revealed only a small number of detectable compounds (only a few peaks identified in the GC–MS analysis). The identified metabolites are glucose, glycerol, erythritol and inositol. The detected glycerol probably comes from the glycerolstock that has been used for inoculation and glucose has been used as carbon source (0.15%). The sugar alcohols erythritol and inositol are found in high concentrations in the supernatant as well as in the cell. The accumulation of these known compatible solutes is discussed as a thermoprotective trait in the extremely hyperthermophilic *Pyrolobus fumarii* (Gonçalves *et al.* 2008) and therefore, a role as compatible solutes can also be assumed for *S. solfataricus*.

Biochemistry of the CCM enzymes

Goals of the biochemical analyses are to identify and confirm the key players of the CCM network of *S. solfataricus* suggested from the genomic reconstruction (SOP_080908; Fig. 3) and particularly, to provide detailed enzymatic and biochemical information of the recombinant CCM enzymes in order to study the behaviour and regulation of the network under temperature change. Focus lies on providing detailed information on substrate specificity, kinetic information (V_{max}-, K_m-, K_{cat}-values) as well as regulatory properties of key enzymes predicted by modelling.

A prerequisite for the biochemical and enzymatic analyses is the availability of recombinant proteins. Therefore, the respective CCM candidate genes are cloned and heterologously expressed in *Escherichia coli*, which is performed according to standard protocols (SOP_SSO_080913a). However, if the recombinant expression in *E. coli* fails, i.e., expression in an insoluble form (inclusion bodies formation) or no expression at all, the respective candidates are expressed in *S. solfataricus* by using the recently developed virus vector based expression system in *S. solfataricus* (SOP_SSO_080913b; (Albers *et al.* 2006). Moreover, homologous expression is used to identify post-translational modifications or to unravel protein–protein interactions, which have not been identified yet. In addition, the constructed over-expression strains (perturbation experiments) will be further analyzed to challenge and improve the established models via transcriptome, proteome as well as the metabolome analyses.

The obtained recombinant proteins from *E. coli* or *S. solfataricus*, respectively, are purified to homogeneity by standard purification methods, like heat

precipitation, ion exchange or hydrophobic interaction chromatography, gelfiltration, and subsequently characterized according to their biochemical, kinetic and regulatory properties (for examples see SOP_SSO_080913c and SOP_SSO_080913d).

The effect of temperature variation at the enzyme level is also studied by determining enzyme activities in crude extracts of *S. solfataricus* grown at different temperatures (SOP_0809012e). Assays for the respective enzymes involved in the branched ED pathway, which is the initial focus of the project (Albers *et al.* 2009), have been established at high temperature. The cell mass of *S. solfataricus* grown at the optimal growth temperature of 80°C has been obtained from the central fermentation unit. The derived data (V_{max} values) play an important role for the parameterization of the constructed models of the CCM network (Drengstig *et al.* 2008; Ni *et al.* 2009).

Cloning and heterologous expression in ***E. coli***(SOP_SSO_080913a)

In order to prove the gene assignments of the identified CCM candidates, the respective genes are cloned into the vector pBlueScript (Novagen) via PCR mutagenesis. The *E. coli* strain K12 DH5α (Hanahan 1983) is used for cloning, storage and preparation of the recombinant plasmid-DNA. For heterologous expression of recombinant *S. solfataricus* proteins the genes are cloned via PCR-mutagenesis (oligonucleotide primers are purchased from Invitrogen) into the pET vector system (Novagen; Table 6) and the strains *E. coli* BL21(DE3), BL21(DE3) pLysS (Studier & Moffatt 1986), BL21-CodonPlus(DE3)-RIL Stratagene; (Carstens & Waesche 1999) and Rosetta (DE3) pRIL (Novagen) are used for the production of the recombinant proteins. The BL21-CodonPlus(DE3)-pRIL and the Rosetta (DE3) pRIL

strains contain plasmids encoding (*argU*, *ileY*, *leuW* and*argU, argW, glyT, lleX, leuW, proL*, respectively) and therefore, these hosts allow for the expression of genes encoding tRNAs for the rare argenine (AGA, AGG, CGA), glycine (GGA), isoleucine (AUA), leucine (CUA), and proline (CCC) codons.

Table 6: Plasmids and their application

Vector	Resistance	Application	Source of supply, reference
pET15b & pET11c	Amp^r	Heterologous expression of *S. solfataricus*proteins in *E. coli*	Novagen, Merck Biosciences
pET24a & pET24d	Kan^r	Heterologous expression of *S. solfataricus*proteins in *E. coli*	Novagen, Merck Biosciences
pMZ1	Amp^r	Cloning of *S. solfataricus* genes for homologous expression contains C-terminal tandem (strep-his)-tag	(Zolghadr *et al.* 2007)
SSV1		*S. solfataricus* shuttle vector	(Jonuscheit *et al.* 2003; Albers *et al.* 2006)
pLysS	Cam^r	Heterologous expression of T7 lysozyme in *E. coli*	Novagen, Merck Biosciences
pRIL	Cam^r	Expression of rare tRNA genes (*argU*, *ileY*, *leuW*)	Stratagene, La Jolla (USA)

The aerobic cultivation of the different *E. coli* strain is carried out in 3–400 ml batch cultures in test glasses or Erlenmeyer flasks at 37°C in Luria–Bertani (LB) medium (1% tryptone, 0.5% yeast extract, 0.5% NaCl (w/v), pH 7) or on solid medium plates (LB medium containing 1.5% (w/v) agar–agar). An optimal oxygen supply of the smaller liquid cultures (3–400 ml) is given by vigorously shaking (220 rpm; Thermotron). Mass cultures of the expression strains are grown at 37°C in a 4 l fermenter [Minifors, Infors AG Bottmingen (CH)] in LB medium. Antibiotics are added according to the plasmid-encoded antibiotic resistance in the following concentrations: ampicillin 100 µg/ml, kanamycin 50 µg/ml and chloramphenicol 34 µg/ml. Liquid LB medium

containing the appropriate antibiotic is inoculated with a preculture (1% (v/v)) and growth is monitored spectrophotometrically at 578 nm. Recombinant protein expression is induced at an OD_{578} of 0.6–0.8 by the addition of 1 mM isopropyl-β-D-thiogalactopyranosid (IPTG) and cultivation is continued for 3–4 h. Afterwards, cells are chilled on ice, harvested by centrifugation (6,000×*g*, 15 min, 4°C) and stored at −80°C.

Cloning and homologous expression in ***S. solfataricus*** (SOP_SSO_080913b)

This virus vector based expression system relies on the complementation of uracil auxotrophic mutants of the *S. solfataricus* strain PH1-16 with the selectable marker genes *pyrEF* (Albers *et al.* 2006). Many efforts failed to heterologously express, for example gluconate dehydratase (GAD, SSO3198) in an active, soluble form in *E. coli*. Therefore, SSO3198 was one of the first candidates cloned into the entry vector pMZ1 (via *Nco*I/*Bam*HI), which contains a C-terminal tandem-tag (Strep-His-tag) and the *araS*promoter (arabinose inducible promoter).

After the transfer of the expression cassette containing the SSO3198 gene into the virus shuttle vector pMJ05 (via *Bln*I/*Eag*I; (Jonuscheit *et al.* 2003; Albers *et al.* 2006), the resulting plasmid (pSVA124) was used to transform the *S. solfataricus* expression strain PH1-16 via electroporation (25 µF, 2.5 kV, 400 Ω; time constant should be between 4–5.2 ms) as described previously (Schleper *et al.* 1992). Positive transformants have been selected, growth has been performed in Brock medium (SOP_SSO_080902, lacking uracil) containing 0.1% NZ-amine at 80°C and expression is induced by the addition of 0.2% D-arabinose at $OD_{600}$of ~0.3. Cultivation is continued until an OD_{600} of 0.8–0.9. Afterwards, cells are chilled, harvested by centrifugation

(7,000×*g*, 15 min, 4°C) and stored at −80°C. For enzyme preparation a 40 l fermenter has been performed.

Preparation of recombinant enzymes (SOP_SSO_080913c)

Recombinant *E. coli* cells are resuspended (1:3) in chilled lysis buffer: 0.1 M HEPES/KOH buffer, pH 7 at room temperature. Recombinant *S. solfataricus* cells are resuspended (1:3) in chilled 50 mM HEPES/KOH, pH 8.5, 100 mM KCl, containing 250 µl complete Protease Inhibitor (7x, Roche). Cell disruption is carried out by sonication (4 times: 2 min pulse/1 min cooling). After centrifugation (45 min, 16,000×*g*, 4°C) the supernatant is decanted and for determination of protein concentration the BioRad Protein Assay based on the Bradford protein quantitation method (Bradford 1976) is used.

Preparation of ***S. solfataricus*** crude extracts (SOP_SSO_080913d)

Resuspension of 0.5 g (wet weight) cells in 1.5 ml 0.1 M HEPES/KOH buffer, pH 7 at room temperature, containing 5 mM DTT and 250 µl complete Protease Inhibitor (7×, Roche). Cell disruption is carried out by sonication (4×, 2 min pulse/1 min cooling). After centrifugation (45 min, 16,000×*g*, 4°C) the supernatant is dialyzed overnight against 0.1 M HEPES/KOH pH 7 at room temperature. For determination of protein concentration the BioRad Protein Assay based on the Bradford protein quantitation method (Bradford 1976) is used. Between 0.25–1 mg total protein is used for the different enzyme assays using crude extracts.

Non-phosphorylating glyceraldehyde-3-phosphate (GAP) dehydrogenase (GAPN; E.C. 1.2.1.9) and gluconate dehydratase (GAD; EC 4.2.1.39) activity in cell-free extracts (Table 7; SOP_SSO_080913e, f) GAPN activity is determined in a continuous enzyme assay at 70°C and 80°C (Table 7). The assay is performed in 0.1 M HEPES/KOH (pH 6.5 is set at 80°C assay temperature) containing 5 mM $NADP^+$ and 300 µg of crude extract in a total volume of 0.5 ml. Reactions are started by the addition of GAP (final concentration 10 mM). Enzymatic activity is measured by monitoring the formation of NADPH and the increase of absorbance at 340 nm by using a specord 210 photometer (Analytik Jena). For each assay three independent measurements are performed.

Table 7: Enzymatic activities of GAPN (SSO3194) and GAD (SSO3198) assayed at 80 and 70°C in cell-free extracts of *S. solfataricus* grown at 80 and 70°C

Growth temperature:	80°C		70°C	
Assay temperature:	**80°C**	**70°C**	**80°C**	**70°C**
E: GAD (U/mg)	0.167	0.127	0.114	0.092
S: gluconate (U/mg)	±0.0108	±0.0001	±0.012	±0.0047
E: GAD (U/mg)	0.077	0.052	0.043	0.029
S: galactonate (U/mg)	± 0.0005	±0.0024	±0.0029	±0.0024
E: GAPN (U/mg)	0.036	0.021	0.054	0.021
S: GAP (U/mg)	±0.0014	±0.0003	±0.004	±0.0014

GAD activity in crude extracts (350 µg crude extract) is measured in a discontinuous enzyme assay at 70 and 80°C (Table 7). The assay is performed in 0.1 M HEPES/KOH (pH 6.5 at the respective assay temperature

(70 or 80°C) containing 10 mM $MgCl_2$ and 10 mM galactonate or 15 mM gluconate, respectively. Reactions are started by the addition of substrate. The sample is incubated in a thermoblock, after 0, 2.5, 5, 7.5 and 10 min of incubation, 25 µl sample is withdrawn on ice and the reaction is stopped by the addition of 2.5 µl of 12% (w/v) trichloroacetic acid.

Enzymatic activity is determined using the TBA assay (modified from (Buchanan *et al.* 1999)): Precipitated proteins are removed by centrifugation (16,000×*g*, 15 min at 4°C) and 20 µl of the supernatants are oxidized by the addition of 125 µl of 25 mM periodic acid/0.25 M H_2SO_4 and incubated at RT for 20 min. Oxidation is terminated by the addition of 250 µl of 2% (w/v) sodium arsenite in 0.5 M HCl. 1 ml of 0.3% (w/v). Subsequently, TBA is added and the chromophore is developed by heating at 100°C for 10 min. Subsequently, a sample (0.5 ml) of the solution is then removed and the color is intensified by adding to an equal volume of DMSO. The change in absorbance is followed at 549 nm ($\varepsilon_{chromophore}$ = 67.8 × 10^3 M^{-1} cm^{-1}). For each assay three independent measurements are performed.

Western blotting and detection of the recombinant ***S. solfataricus*** proteins (SOP_SSO_080913g)

Electrophoretically separated tagged proteins are transferred from the PAA gel to a hydrophobic membrane (PVDF-(ProBlott) or Nylon-membrane (Roth)) by wet electroblotting.

The transfer is carried out using a tank blot system (Biometra). Therefore, after the electrophoresis run, the gel and two Whatman paper (Schleicher & Schuell) are equilibrated in transfer buffer (50 mM Tris, 380 mM Glycin, 0.1%

SDS, 20% methanol) for 15 min. The membrane is briefly moistened with 100% (v/v) methanol and afterwards also equilibrated in transfer buffer. The blot assembly is performed as recommended by the blot system manufacturer (Biometra). The transfer is carried out with 12 V over night (~20 h) at 4°C and after blotting the membrane is air dried. Blotting efficiency is controlled by the transfer of the applied pre-stained protein marker (PageRuler, Fermentas) on the PAA gel.

For immunodetection the membrane is incubated for 5 min in 100% (v/v) methanol, washed three times for 5 min with PBST-buffer (1× PBS (63.2 mM Na_2HPO_4, 11.7 mM KH_2PO_4, 68 mM NaCl pH ~7.3) + 0.3% Tween-20) at RT on a rotary shaker, blocked for 1 h at RT by either using PBST-buffer containing 5% skim milk (his-Tag detection) or PBST-buffer containing 0.2% I-Block (Applied Biosystems;*StrepII*-tag detection). After three times washing for 5 min using PBST-buffer either containing 2.5% skim milk or 0.1% I-Block, 1:2,000 Anti-His antibody AP conjugate (rabbit; Abcam) or 1:4,000 *Strep*-Tactin AP conjugate (IBA BioTAGnology) are added to the respective PBST-buffer. Incubation is carried out for at least 1 h 30 min at RT on a rotary shaker. Afterwards, the membrane is washed six times for 5 min at RT using PBST-buffer either containing 2.5% skim milk or 0.1% I-Block. Finally, the membrane is washed two times for 10 min in A.bidest. and incubated for 15 min at 37°C in 9 ml pre-warmed A.bidest., containing 1 ml CDP-Star (Invitrogen). Chemiluminescence is detected by using the VersaDoc System (BioRad).

<u>Purification of obtained recombinant GAPN (SSO3194; Fig. 8) and the GAD (SSO3198; Fig. 9) (SOP SSO 0809013c, d)</u>

For enrichment of the recombinant GAPN, the resulting *E. coli* crude extract is diluted 1:1 with 0.1 M HEPES/KOH buffer, pH 7 at RT and subjected to a heat precipitation for 20 min at 70°C. After heat precipitation, the samples are cleared by centrifugation (16,000×*g* for 30 min at 4°C). The supernatant is dialyzed overnight against 20 mM HEPES/KOH (pH 6.5, 70°C), containing 5 mM dithiothreitol, subjected to ion exchange chromatography on UNO Q-12 (Bio-Rad Laboratories) pre-equilibrated by using the respective buffer, and eluted with a salt gradient from 0 to 1 M NaCl. Fractions containing the GAPN (checked by SDS–PAGE) are pooled and concentrated via centrifugal concentrators (Vivaspin6, Sartorius Stedim Biotech). Afterwards, the sample is dialyzed overnight against 50 mM HEPES/KOH (pH 6.5, 70°C), containing 5 mM dithiothreitol, 300 mM NaCl, and subjected to gelfiltration on HiLoad 26/60 Superdex 200 prep grade (Amersham Biosciences) preequilibrated in the respective buffer (Fig. 8).

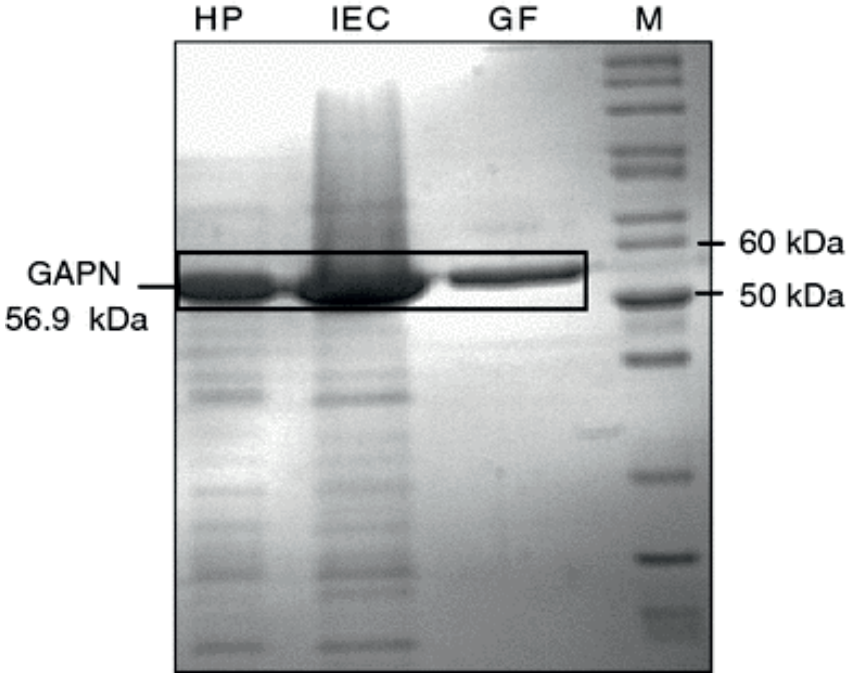

Fig. 8: Purification of the heterologously expressed GAPN from *S. solfataricus* by using the *E. coli* pET expression system. *HP* Heat precipitation at 70°C, *IEC* ion exchange chromatography, *GF* gelfiltration, *M* protein ladder (Page ruler™, fermentas)

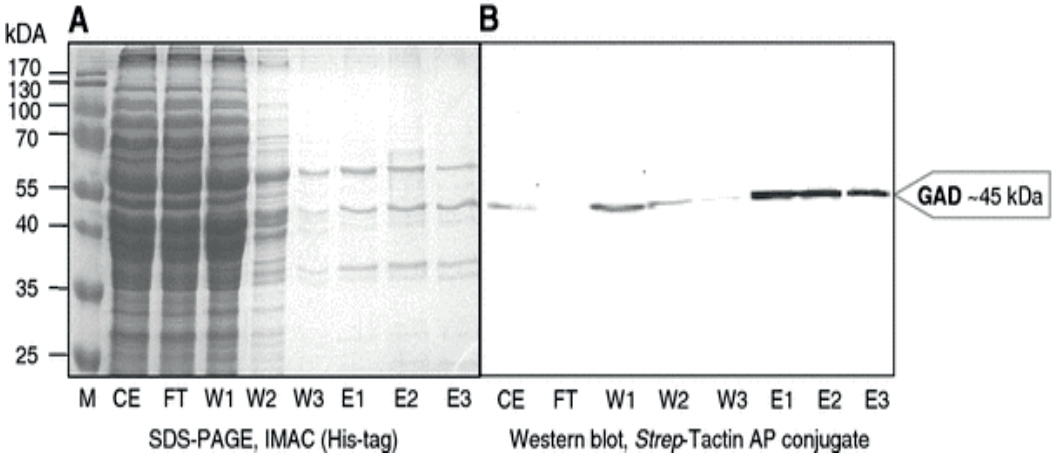

Fig. 9: SDS PAGE gel (**a**) and western blot (**b**) showing homologous expression and purification of the *S. solfataricus* GAD (SSO3198). **a** Coomassie stained 12.5% PAA gel of His tag-specific affinity chromatography fractions. **b** Detection of the blotted *S. solfataricus* GAD using S*trep*-Tactin, revealing a protein of about 49 kDa (including tandem tag). *M* Protein standard, *CE* crude extract, *FT* flow through,*W1-3* washing fractions, *E1-3* elution fractions.

The homologously expressed recombinant GAD from *S. solfataricus* is isolated via the attached His-tag by Immobilized Metal Affinity Chromatography (IMAC) using a His-Select column (Qiagen, Hilden) and HIS-Select® Nickel Affinity Gel (Sigma). Hereunto, the resulting *S. solfataricus* crude extract is applied onto nickel-nitrilotriacetic acid (Ni–NTA) affinity columns (5 ml volume, Qiagen) equilibrated with 50 mM HEPES/KOH, pH 8.5 containing 100 mM KCl (buffer 1). The column is washed three times with 2× column volume buffer 1 containing 25 mM imidazole. Bound GAD is eluted in three steps with buffer 1 containing 250 mM imidazole. After monitoring purification by SDS–PAGE, the protein has been blotted and stained with Strep-Tactin (streptavidine

analogue; IBA; Fig. 9).

Activity of the recombinant GAPN (EC 1.2.1.9; SOP_SSO_0809013e)

GAPN activity is determined in a continuous enzyme assay at 80 and 70°C (Table 8). The standard assay is performed in 0.1 M HEPES/KOH (pH 6.5 is set at the respective assay temperature (70 or 80°C) containing 2 mM $NADP^+$ and 5 µg of purified protein in a total volume of 0.5 ml. Reactions are started by the addition of 3 mM D,L-GAP. Enzymatic activity is measured by monitoring the change in absorbance due to the increase of NADPH at 340 nm (εNADPH, 70°C = 5.71 $mM^{-1}(cm^{-1})$. For each assay three independent measurements are performed.

Table 8: Kinetic parameters of the GAPN (SSO3194) assayed at 80 and 70°C

D,L-GAP (mM)	NADP (mM)	Assay temp (°C)	V_{max}(U/mg)	K_m(mM)	K_{cat} (min^{-1}) (s^{-1})	K_{cat}/K_m(mM^{-1} s^{-1}]
3	2	80	10.58	0.95	544.97 9.08	9.51
3	2	70	7.46	1.51	384.17 6.40	4.25

The kinetic parameters (V_{max} and K_m) are calculated by iterative curve-fitting (Hanes) using the program Origin (Microcal Software, Northampton, MA, USA).

Activity of the recombinant GAD (EC 4.2.1.39; SOP_SSO_0809013f)

Recombinant GAD activity has been confirmed via the modified thiobarbituric acid (TBA)-assay (Buchanan *et al.* 1999) by using 7.5 µg of the purified protein (enriched elution fraction). Activity is determined in a discontinous enzyme assay at 80°C. The assay is performed in 0.1 M HEPES/KOH (pH 6.5 is set at the respective assay temperature 80°C) containing 10 mM $MgCl_2$

and 10 mM gluconate or 10 mM galactonate, respectively. Reactions are started by the addition of substrate.

For initial enzymatic analysis the sample is incubated at 80°C and after 0 and 10 min, 25 µl of the sample is transferred on ice. The reaction is stopped by the addition of 2.5 µl of 12% (w/v) trichloroacetic acid. Precipitated protein is removed by centrifugation (16,000×*g*, 15 min, 4°C). Enzymatic activity is determined by using a modified thiobarbituric acid (TBA)-assay (Buchanan *et al.* 1999).

Acknowledgments

The authors thank the Federal Ministry of Education and Resarch (BMBF), Germany, the Netherlands Organization for Scientific Research (NWO), the Research Council of Norway (RCN), and the Biotechnology, Biological Research Council (BBSRC), United Kingdom, as well as the partner universities (University of Bergen (Norway), University of Duisburg-Essen (Germany), Wageningen University and University of Groningen (The Netherlands), University of Sheffield and the University of Manchester (The United Kingdom), Free University Amsterdam (The Netherlands) for financial support of the SulfoSYS-project (SysMo P–N-01-09-23).

Open Access

Chapter 4

Quantitative Proteomic Analysis of Sulfolobus solfataricus Membrane Proteins

Trong Khoa Pham† Pawel Sierocinski‡ John van der Oost‡ and Phillip C. Wright*†

† ChELSI Institute, Biological and Environmental Systems Group, Department of Chemical and Process Engineering, The University of Sheffield, Mappin Street, Sheffield, S1 3JD, United Kingdom,
‡ Laboratory of Microbiology, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, The Netherlands

Abstract

A quantitative proteomic analysis of the membrane of the archaeon *Sulfolobus solfataricus* P2 using iTRAQ was successfully demonstrated in this technical note. The estimated number of membrane proteins of this organism is 883 (predicted based on Gravy score), corresponding to 30% of the total number of proteins. Using a modified iTRAQ protocol for membrane protein analysis, of the 284 proteins detected, 246 proteins were identified as membrane proteins, while using an original iTRAQ protocol, 147 proteins were detected with only 133 proteins being identified as membrane proteins. Furthermore, 97.2% of proteins identified in the modified protocol contained more than 2 distinct peptides compared to the original workflow. The successful application of this modified protocol offers a potential technique for quantitatively analyzing membrane-associated proteomes of organisms in the archaeal kingdom. The combination of 3 different iTRAQ experiments resulted in the detection of 395 proteins (g2 distinct peptides) of which 373 had

predicted membrane properties. Approximately 20% of the quantified proteins were observed to exhibit g1.5-fold differential expression at temperatures well below the optimum for growth.

Systems Biology of Microorganisms, SysMO, is an EU transnational project with the aim to discover and predict the dynamics of molecular processes via computerized mathematical modeling (www.sysmo.net). *Sulfolobus solfataricus* P2 has been selected as one of the model organisms. More information on the overall goals and team involved with *S. solfataricus* are presented elsewhere(Albers *et al.* 2009). *S. solfataricus* is an archaeon which was first isolated from sulfur-rich hot springs in a solfataric field near Naples (Italy), and grows optimally at 80°C and pH 3-4 (Zillig *et al.* 1980). Life at very high temperature requires special cellular strategies in order to survive. However, how the cells can manage to thrive at high temperature and respond to changes in temperature, at least from the proteomic viewpoint, is not yet clear.

Membrane proteins play important functions in many processes including nutrient transport, signal transduction, and energy conversion. Moreover, they generally represent one-third of all cellular proteins (from bacterial, archaeal, and eukaryotic organisms) (Stevens & Arkin 2000; Wallin & Heijne 2008) Although they are very functionally important parts of living cells, their quantitative proteomics analysis is still relatively rare, mainly due to technical difficulties. Many studies have been performed to improve techniques for identification, as well as quantification of membrane proteins from the archaeal domain of life (Bisle *et al.* 2006; Assiddiq *et al.* 2008; Palmieri *et al.* 2009). Although diverse techniques have been applied for quantitative proteomics in the archaea (Bunai & Yamane 2005; Klein *et al.* 2005; Bisle *et*

al. 2006), including published quantitative proteomic analyses of *S. solfataricus* , these studies have mostly focused on cytosolic proteins or whole cell lysates, with few investigations focusing on membrane proteins (Snijders *et al.* 2005, 2006; Chong *et al.* 2007b, a). Furthermore, application of iTRAQ solely for membrane proteomics analysis has not been reported yet, despite the fact that this technique offers many advantages (Khoa Pham & Wright 2007). Therefore, the purpose of this technical study is to develop and assess a modified method for iTRAQ-based quantitative membrane proteomics using *S. solfataricus* P2 grown at different temperatures (65, 70, and 80°C). The purpose of this paper is not to discuss the reliability of the biological and technical replicates for each iTRAQ experiment, since these issues were discussed in detail by Chong et al. (Chong *et al.* 2006) and Gan et al. (Gan *et al.* 2007) for this organism previously. Here, we will discuss the reliability of this technique in terms of membrane protein expression across 3 different iTRAQ experiments.

Data from this study (together with metabolomic, transcriptomic and classical biochemical data) will be used for future modeling purposes to generate an in silico network systems biology model for this archaeon. To date, initial models of the response of *S. solfataricus* to different temperatures have been built. More details can be seen at http://bioinfo.ux.uis.no/ sulfosys/ and http://jjj.biochem.sun.ac.za/sysmo/projects/ Sulfo-Sys/index.html. Although much progress has been made in improving biochemical data analysis, until now, many predicted central carbon metabolism (CCM) proteins have not been kinetically investigated because of technical limitations. Therefore, a combination of quantified protein data and quantified metabolomic data (metabolic fluxes) will offer a wealth of behavioral information.

Materials and Methods

Cell Growth Condition, Protein Extraction, Membrane Protein Isolation, and iTRAQ Labeling.

Cells were obtained from aerobic batch fermentation on minimal medium with 0.3% glucose, harvested during exponential growth (OD600 0.85 (0.15); details of media composition and growth conditions are listed and detailed in Zaparty et al. (Zaparty *et al.* 2009) The archaeon *S. solfataricus* P2 was grown at different temperatures (65, 70, 80°°C) to investigate the proteome responses of this organism to temperatures reduced from the 80°C optimum, especially for most of the predicted central carbohydrate metabolism (CCM) candidates. Samples were collected in late exponential phase, then cells were extracted subsequently; membrane and soluble proteins were separated before being applied to a quantitative proteomic analysis as shown in Figure S1 (supplementary materials 2). Since this was the first time iTRAQ was applied solely for enriched membrane fractions, three different iTRAQ experiments with different treatments were done to evaluate this technique and choose the best method. While the first iTRAQ experiment (Exp. 1) was performed using the original protocol from Applied Biosystems, the second (Exp. 2) and the third (Exp. 3) experiment were performed according to adjusted protocols (Figure S1). The comparison of protein expressions at various temperatures was performed based on analysis of iTRAQ labelled peptides from each of the assessed temperatures, allowing an evaluation of the efficiency of the modified protocols compared to the iTRAQ original protocol. Duplicate independent biological replicate samples, as well as technical replicates, were examined (as shown in Figure S2 (supplementary materials 2)) to ensure that the changes in protein expressions of interesting

proteins reflect significant cellular process changes in response to temperature.

The buffer used for protein extraction was devoid of detergents, so that the isolation of the insoluble fraction (defined as the fraction not dissolved in high salt contents buffers used for extractions) could be achieved successfully. The iTRAQ labeling step was performed as detailed elsewhere (Zaparty *et al.* 2009) briefly, 100 µg of protein from each phenotype was used for iTRAQ analysis. Protein samples were reduced, alkylated, digested and labeled with iTRAQ reagents. Details of protein digestions using either trypsin or combination of trypsin and chymotrypsin can be found in the literature.18 In the case of trypsin and chymotrypsin, the proteins were digested with trypsin at a 1:40 (w/w) ratio overnight, and then a mixture of both trypsin and chymotrypsin with a 1:40 enzyme/protein ratio for the second day. Finally, digested peptides were combined prior to iTRAQ labelling (Zaparty *et al.* 2009).

Here, 3 different iTRAQ experiments were performed for enriched membrane fractions and an independent biological replicate for each phenotype was also utilized and labeled with iTRAQ regents 113 and 114 for samples at 65°C, labels 115 and 116 for 70°C, and labels 119 and 121 for 80°C. Furthermore, technical replicates were also carried out where iTRAQ reagents 116 and 117 were used for samples at 70°C, and 118 and 119 used for samples at 80°C (see Figure S2 for more details). Exp. 2 was carried out in absence of SDC (sodium deoxycholate) with both trypsin and chymotrypsin present; Exp. 3 was performed with this surfactant and both trypsin and chymotrypsin present. Details of buffers used as well as the membrane isolation, delipidation, protein

digestions and iTRAQ labelling are described in detail elsewhere.(Zaparty *et al.* 2009)

Strong Cation Exchange and Mass Spectrometry Analysis, Data Searching, and Data Analyses.

Since the presence of residual iTRAQ reagents and surfactants (e.g., SDS and SDC) negatively impacts on downstream proteomic analysis (MS/ MS analysis), strong cation exchange (SCX) chromatography was applied to clean as well as fractionate samples prior to MS analyses. Detailed procedures for this step can be found elsewhere.18 Briefly, buffers containing 10 mM KH2PO4, 25% acetonitrile, pH 3, and 10 mM KH2PO4, 25% acetonitrile and 500 mM KCl, pH 3, were used as mobile phases. iTRAQ fractionated peptides were collected every minute. SCX fractions with high intensities (detected at 214 nm) were subjected to MS/MS analyses.

The MS/MS analyses were performed on a QStar XL Hybrid ESI Quadrupole time-of-flight tandem mass spectrometer, ESIqQ-TOF-MS/MS (Applied Biosystems/MDS Sciex), coupled to a LC Packings Ultimate 3000 (Dionex, U.K.) nano-LC system. Details of MS/MS operating parameters are described elsewhere (Zaparty *et al.* 2009).

iTRAQ MS/MS data were subsequently analyzed using Phenyx software v.2.6 (Geneva Bioinformatics, Switzerland) with the *S. solfataricus* P2 protein database downloaded June 2007 from NCBI. This microorganism was fully sequenced in 2001 with a G + C content of 35.8% and a genome size of approximately of 3.0 Mb encoding 2977 open reading frames (ORFs) (She *et al.* 2001a). (Furthermore, the ORFs have recently been confirmed by

transcriptome sequencing and 80 new transcribed ORFs have been found (Wurtzel *et al.* 2010)) Details on the parameters can be found elsewhere(Zaparty *et al.* 2009). Briefly, MS tolerance was 0.6 and MS/MS tolerance were set at peptide tolerance 0.2 Da, charge +2 and +3 +4, min peptide length, z-score, max p-value and AC score were 5, 5, 10-5, and 5, respectively, and enzymes used for searching were trypsin alone or a combination of trypsin and chymotrypsin (in Exp. 2 and 3) with one missed cleavage permitted for both cases. The results were then exported to Excel (Microsoft 2008) for further analyses.

iTRAQ uses the peak areas (or intensities) of reporter ions for comparisons of peptide expression ratios, which are then averaged per protein to yield protein expression ratios. In terms of proteomic analyses, it is important to meet both requirements: minimize false positive detections and avoid false negative identifications. Therefore, all MS/MS data were searched against two different databases: one from standard protein sequences and another from reversed database (protein sequences were written from C to N) (see (Elias & Gygi 2007) for more details). All parameters for Phenyx searching were then adjusted to get the false positive rate <0.2% (parameters were briefly mentioned above, see (Zaparty *et al.* 2009) for more details). As a result, the full lists of peptides observed in all 3 different iTRAQ experiments are summarized in sheets 1-3 in supplementary materials 1, as well as the number of peptides from a decoy database (sheet 4 in the same supplementary materials file).

Since Phenyx V2.6 provides only lists detected peptides together with their iTRAQ reporter ion intensities, calculations of (peptides) protein expressions were carried out manually based on the instructions in the Protein Pilot v.2.0

documentation (Applied Biosystems). Briefly, peptide reporter ion intensities for each phenotype were averaged (see column S, T, U in the sheets Exp. 1, 2, and 3 in supplementary materials 1 for details) prior to subsequent comparisons (as shown in Figure S2) by matching pairs (65°C compared to 80 and 70°C compared to 80°C, see columns V and W in the sheets Exp. 1, 2, and 3 in supplementary materials 1 for details) and then these ratios were transformed to log form (base 10) (see columns X and Y in the sheets Exp. 1, 2, and 3 in supplementary materials 1 for details) before an average value was taken for each protein ratio. Finally, the inverse logarithm of these average values (in log form) was calculated to give the final ratios. Subsequently, a median value for each pair comparison was made, and then a final protein expression was calculated by dividing each pair value by the median value (for each pair of comparison) (see columns I and K in the sheets Exp. 1, 2, and 3 in supplementary materials 2 for the final results). Furthermore, the error factors (EF) were also considered and calculated for each pair comparison. These values were obtained by taking the inverse logarithm of an average value (in $\log_{10}$ form) of the standard deviation that was taken from all peptide ratios (for each pair) contributed for each protein. The results are shown in columns J and L in the sheets Exp. 1, 2, and 3 in supplementary materials 2. The final list of proteins with ratios and EF values is presented in the sheets for Exp. 1-3, in supplementary materials 2.

By investigating the biological replicates for this archaeon, 1.5-fold was used (as recommended by Chong et al.15 for this microorganism) as a cutoff for differential expression by considering the differences in protein expressions within these biological comparisons. Thus, proteins that exhibited expression changes (up or down) of greater than 1.5-fold (with consideration of the error factor (EF) value <2.0, as per Applied Biosystems' iTRAQ instructions) were

considered to be differentially expressed under lower temperatures (65 and 70°C) compared to the optimum temperature (80°C). These proteins were used for further evaluation and discussion of biological implications. To estimate the biological and technical replicates for each iTRAQ experiment, a t test was carried out for each phenotype (65, 70, 80°C).

Bioinformatics Analysis.

Generally, membrane proteomes consist of all proteins associated to the membrane which are formed by (i) spanning the lipid bilayer with transmembrane domain (TMD) known as integral membrane, (ii) having a covalent bound lipid layer, (iii) being a subunit of a protein complex, or (iv) having electrostatic interactions with the integral membrane proteins or the lipid bilayer.7(Santoni *et al.* 2000). The identification of membrane proteins in theory is mostly based on bioinformatic algorithms such as Gravy score for determination of hydrophobic properties (http://www.bioinformatics.org/sms2/protein_gravy.html) (Kyte & Doolittle 1982), TMHMM for determination of TMDs v.2.0 (http://www.cbs.dtu.dk/services/TMHMM/) (Krogh *et al.* 2001), Psortb v.2.0.4 for determination of protein localization (http://www.psort.org/psortb/) (Gardy *et al.* 2005), and Prosite motif for determination of lipid anchors (http://www.expasy.ch/prosite/). The membrane associated proteins (complex subunit) can be only found in gene annotation for *S. solfataricus* (http://www-archbac. u-psud.fr/projects/sulfolobus/). To ensure that all proteins detected in this study contain membrane properties, we examined these proteins using all these bioinformatics tools detailed above apart from the Prosite motif, since no lipobox has been found for this organism (Albers & Driessen 2002).

Results and Discussion

Number of Distinct Peptides and Unique Proteins Detected.

Since the identification and quantification of proteins with iTRAQ were performed at the peptide level, it is therefore necessary to evaluate the effectiveness of such an approach. The first aspect examined here is the number of distinct peptides detected for each of the three iTRAQ experiments, since this would affect the number of detected proteins as well as the quality of quantification. The numbers of detected distinct peptides (with more than 2 distinct peptides per protein) from all iTRAQ experiments are presented in Figure 1A. It is clear that more distinct peptides were found in Exp's 2 and 3 (trypsin and chymotrypsin) compared to Exp. 1 (trypsin alone). More distinct peptides were also observed in Exp. 3 (with SDC) compared to Exp. 2 (without SDC). Our result confirms that the presence of SDC enhanced membrane protein digestion, and this surfactant, was suitable for not only the identification of proteins (as reported by (Masuda *et al.* 2008)) but also for quantification.

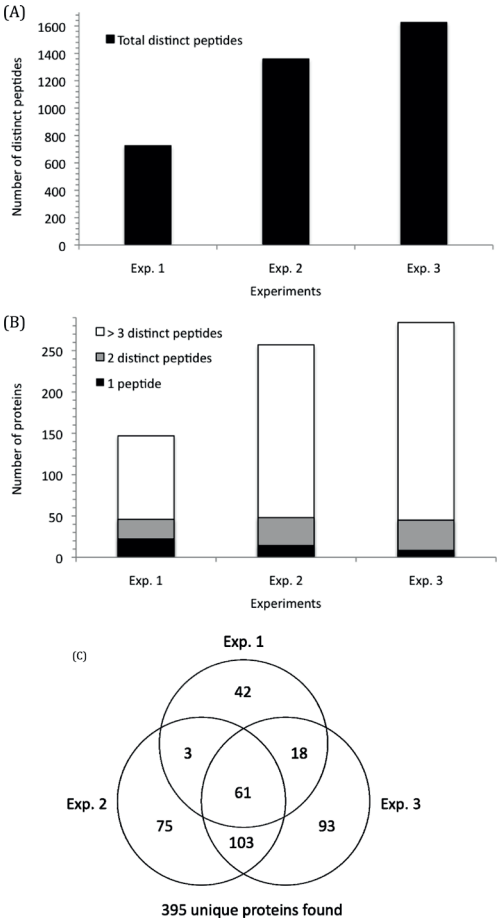

Figure 1: Numbers of detected distinct peptides (A) and then the number of distinct peptides vs the number of proteins (B), as well as a combination of proteins measured in the 3 different iTRAQ experiments (C). Panels B and C show only proteins detected with more than 2 distinct peptides.

The higher number of detected distinct peptides (1626 peptides in Exp. 3 vs 726 peptides in Exp. 1) led to an increased number of identified proteins. While only 147 unique proteins were found in Exp. 1, the number of proteins increased significantly to 257 for Exp. 2 and 284 for Exp. 3. Therefore, increments of 75% and 93% in the number of proteins detected were

observed for Exp's. 2 and 3 compared to Exp. 1, respectively. To investigate the contributions of the numbers of distinct proteins to the overall number of proteins detected, the relationship between the number of distinct peptides contributed for each protein was also assessed, and results are shown in Figure 1B. An interesting finding here was that the number of proteins detected by single peptide was very low, especially for Exp. 3, where only 9 proteins detected by single peptides were observed (compared to 15 and 23 proteins for Exp. 2 and Exp. 1, respectively). As a consequence, percentages of 15.6, 5.8 and 3.2% of proteins detected by single protein were calculated, and the results show that the modified protocol works well for quantitative analysis of the *S. solfataricus* membrane proteome.

Since 3 different iTRAQ workflows were assessed in this study, it was observed that some proteins overlapped within these experiments. Therefore, the data was merged to generate a larger list, and to assess iTRAQ technical reproducibility. The combination of these 3 iTRAQ experiments resulted in 395 proteins being detected overall (with more than 2 distinct peptides) in the enriched membrane fractions (see Figure 1C).

The Efficiency of Protein Digestion and Peptide Labeling.

One of the problems when using iTRAQ in the present study with the original protocol was the detection of a low number of quantifiable peptides. Therefore, to overcome this issue, the analyses of this technique using different digestion protocols were investigated here. The testing with different protocols increased the number of quantifiable peptides, as well as increased the numbers of quantified distinct peptides when a combination was applied. It is clear that the use of both trypsin and chymotrypsin, with the enhancement

of SDC, improved iTRAQ analysis of membrane proteins. For iTRAQ, the success of this technique is mostly based on the tryptic digestion and labeling steps; for this reason, the use of trypsin alone for complete digestion of membrane proteins seems to be difficult to achieve, especially for integral membrane proteins, since TMDs are very difficult to cleave by trypsin(Eichacker *et al.* 2004). For that reason, here, we used detergent pretreatment (SDS/SDC) and digestion by a cocktail of trypsin and chymotrypsin, leading to a significantly increased number of labeled peptides detected using this combination, as shown in Figure 1. Moreover, a significant increase of detected integral membrane proteins (Figure 2B) was observed.

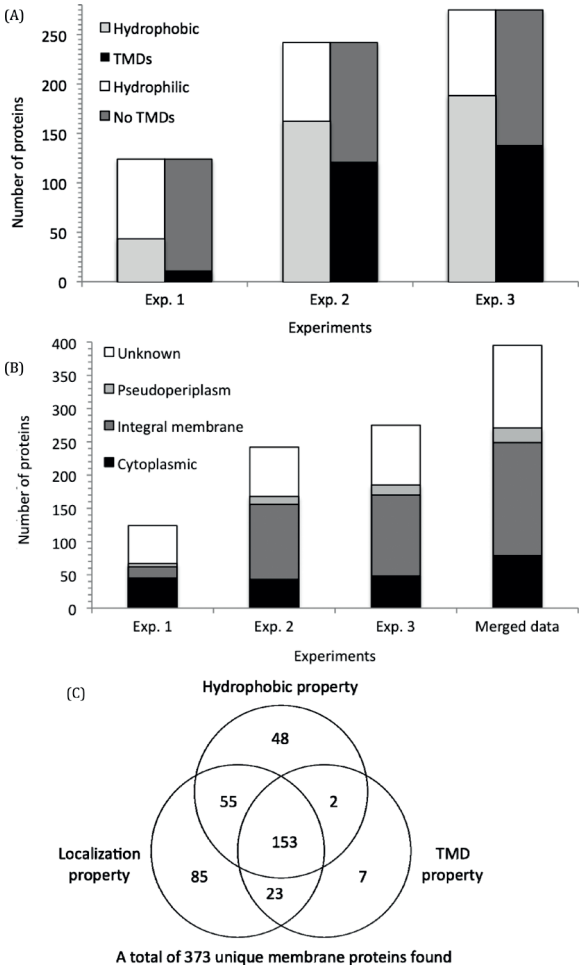

Figure 2. An examination of membrane proteins' properties. (A) The classification of proteins based on hydrophobic properties (http://www.bioinformatics.org/sms2/protein_gravy.html) and TMDs (http://www.cbs.dtu.dk/services/TMHMM/), (B) based on localization (http://www.psort.org/psortb/). (C) The classifications of proteins from merged data. All proteins were characterized based on 3 main membrane properties including hydrophobic, TMDs found and localization. As a result, 373 membrane proteins were found (a membrane protein was considered if it contained at least 1 membrane property).

The sizes of peptides digested by trypsin increases with the number of TMDs (for each protein). This means that if more TMDs were observed for a protein, when trypsin is used for digestion, peptide lengths will be longer compared to those when trypsin is used for nonmembrane proteins (Bisle *et al.* 2006). Therefore, to increase the identification and quantification of membrane proteins (especially integral membrane proteins), the use of both trypsin and chymotrypsin combined with an enhancing agent (SDS) for digestion was applied (Exp. 3). As a result, 122 integral membrane proteins were detected for Exp. 3. The average peptide-to-protein ratios presented for membrane proteins are shown in Figure 1A.

Table 1. t-Test for Biological Duplicate and technical replicates for ech iTRAQ Experiment

	No. of distinct		t test for 65°C	t test for 70°C		t test for 80°C	
Experiment	peptides	Unique proteins	Biological duplicate	Biological duplicate	Technical replicate	Biological duplicate	Technical replicate
Exp. 1	749	147	0.02	0.27	0.24	0.03	0.02
Exp. 2	1374	257	0.03	0.37	0.13	0.16	0.22
Exp. 3	1635	284	0.04	0.38	0.23	0.32	0.18

Table 2. The illustration of selected protein ratios found as overlapping in 3 different iTRAQ Experiments.

ORF	Proteins	Distinct peptides	Exp. 1	Distinct peptides	Exp. 2	Distinct peptides	Exp. 3	Average	SD
SSO0176	AAA family ATPase	9	0.83	13	0.88	14	0.84	0.85	0.03
SSO7114	SSU ribosomal protein S27E	2	1.62	2	1.51	2	1.61	1.58	0.06
SSO2984	Hypothetical protein		2.00		2.12	2	1.81	1.97	0.22

Since an excess amount of SDS (and probably SDC) could affect the labeling step, we thought it prudent to examine this aspect here. Peptides with missing reporter ions were considered as unlabeled peptides, and used for estimating the efficiency of the peptide labeling step. The efficiency of the labeling step was calculated based on the ratio of unlabeled peptides to the total number of detected peptides for each iTRAQ experiment. However, it appeared that the use of these chemicals here was suitable for iTRAQ, since a very small percentage of unlabeled peptides was observed in all 3 iTRAQ experiments (see the sheets Exp. 1, 2, and 3 in supplementary materials 1); actually only 2 unlabeled peptides were found in Exp. 3 (2/4973-0.04%).

Characterization, Classification, and Localization of Detected Membrane Proteins.

The full list of identified and quantified proteins is shown in Exp. 1-3 in supplementary materials 2. The modified protocol seems to be a suitable method for identification and quantitation of integral membrane proteins, since 122 integral membrane proteins were detected with more than 2 distinct peptides (Exp. 3), compared to 17 integral membrane proteins found in Exp. 1 (Figure 2B). Since archaea generally possess a single membrane (i.e., cytoplasmic membrane), they lack outer membrane or periplasm proteins (Albers *et al.* 2004). However, archaea do contain proteins somehow attached to the outside the cytoplasmic membrane, a region referred to as the pseudoperiplasm (Bardy *et al.* 2003). In this investigation, we were also able

to detect up to 21 pseudoperiplasm proteins, however, much less than the obtained number of integral membrane proteins (see supplementary materials 2 and Figure 2B). From Figure 2A we also can see that more hydrophobic proteins were detected in Exp. 3 compared to other experiments, and the same trend was observed for proteins detected with the presence of TMDs. As expected, most detected proteins were hypothetical proteins, since these proteins are predicted to comprise 40.3% of the whole *S. solfataricus* proteome. The next largest groups were proteins involved in translation and transport. While most of the proteins observed in translation group were ribosomal proteins, most proteins in the transport group were ABC transporters or related to ABC transporters.

To ensure that proteins detected in this study fully have membrane-associated features, the 395 merged proteins were then characterized in terms of hydrophobicity, TMDs found and localization properties. The results are shown in Figure 2C, where 373 membrane proteins were characterized, and of these, 153 contain 3 different membrane properties, and 80 proteins contain 2 different membrane properties. Therefore, we believe that at least 233 proteins identified and quantified here are true membrane proteins, while 124 proteins are retained as uncharacterized (as annotated in the genome). These proteins could be either true-membrane proteins or cytoplasmic contaminants, and the characterization of these proteins should be undertaken in future work.

Table 2 illustrates the reliabilities of some membrane proteins detected overlapping in all iTRAQ experiments. Using the (1.5-fold cutoff criterion, the number of up- and downregulated proteins from the merged iTRAQ data are shown in Table 3 (supplementary materials 2), while the full list of these

proteins appears in the sheet “merged data” in supplementary materials 2. Only regulated proteins identified with an error factor (EF) value less than 2.0 (see (Shilov *et al.* 2007)) are taken forward for biological discussions to ensure that all these regulated protein ratios reflect distinct biological changes in the proteome. Furthermore, to meet the guideline for publication of proteomics data recommended in the Paris Consensus published details in the Molecular and Cellular Proteomics journal (http://www .mcponline.org/misc/ParisReport_Final.dtl), only regulated proteins detected with more than two distinct peptides are used for biological discussions.

Both up- and down-regulation of proteins at reduced temperatures compared to the optimum temperature were assessed. To ensure that regulations reflect true biological changes in cell, we investigated MS/MS data based on peptide level via three criteria: (i) number of distinct quantified peptides detected for each protein (supplementary materials 2), (ii) measurements of the variation of biological replicates for each phenotype (65°C, 70 and 80°C) via a t test performed for each iTRAQ experiment (Table 1), and (iii) the variation of each protein ratio from all 3 iTRAQ experiments (some proteins can be found in at least 2 iTRAQ experiments, some found in only a single iTRAQ experiment) (standard variation). The protein ratio variation was slightly lower when the protein ratio was found in both iTRAQ experiments rather than 3 iTRAQ experiments. The reproducibility of the quantitation was found to be sufficiently high (see Table 2 for illustration).

Despite a dramatic shift in temperatures (10 and 15°C, from 80 to 70 and 65°C), only 72 unique membrane proteins (19.3% of quantified membrane proteins) showed significant regulations more than (1.5-fold. At 65°C compared to 80°C, 39 proteins were up-regulated, and at 70°C, 50 were

uniquely up-regulated (29 proteins overlapping with those at 65°C). At 65°C, 19 were down-regulated, and at 70°C, 21 were down-regulated (9 proteins overlapping with those at 65°C). The highest number of regulated proteins at reduced temperatures belonged to the hypothetical group (Figure 3). The second largest groups were translation (22 proteins) and IS elements (22 proteins). The third group was the transport group. Interestingly, all regulated proteins relating to IS elements process and most proteins involved in transport processes were up-regulated under reduced temperatures. The highest numbers of proteins in the transportation group was ABC transporters. Twenty-eight proteins relating to ABC transportation were observed, where 4 and only 1 ABC transporter were up- and down-regulated, respectively, at reduced temperature. Moreover, 20 ribosomal proteins (translation group) were also detected, and of these, 11 and 6 proteins were observed as being up- and down-regulated, respectively. The up- regulation of both these transportation and ribosomal proteins may reflect the fact that the membrane association of ribosomes at reduced temperatures might enhance membrane protein biosynthesis,30 rather than quantify the amount of ribosomal proteins. A large number of hypothetical proteins were also found, and by definition the functions of these proteins have not been clarified yet; therefore, characterization of these proteins needs to be done in the future.

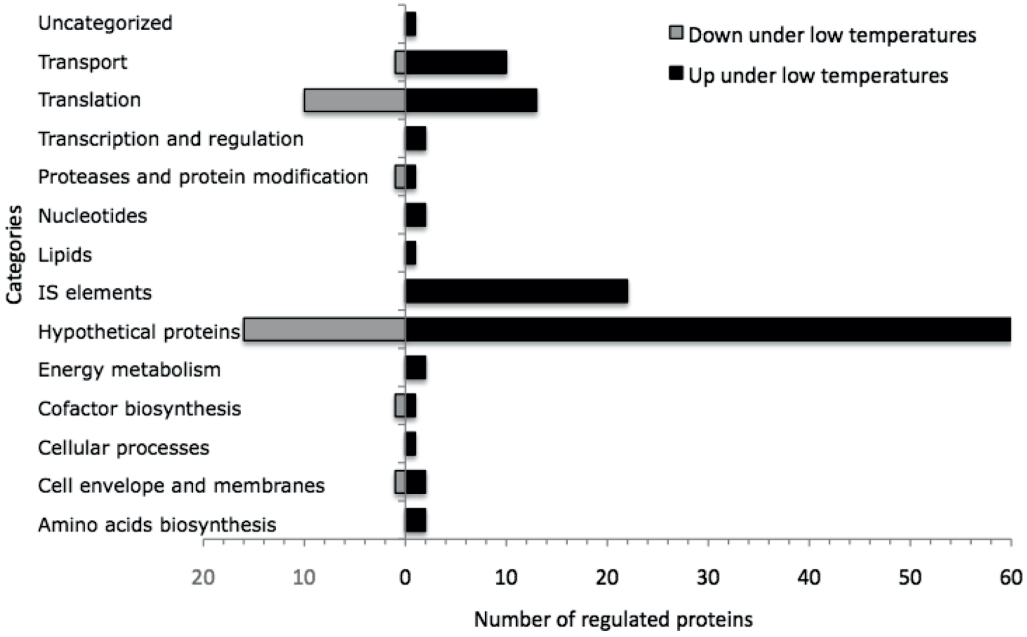

Figure 3. Protein classification of regulated proteins as annotated in the whole genome

Several studies have focused on membrane proteins in the archaea, but few membrane studies have been carried out for *S. solfataricus*. Recently, an attempt was made to analyze secreted membrane vesicles from 3 different Sulfolobus species, including *S. acidocaldarius*, *S. solfataricus* and *S. tokodaii*. 33 SDS-PAGE was used for protein separation and 29, 48, and 29 proteins were identified from *S. acidocaldarius*, *S. solfataricus* and *S. tokodaii*, respectively(Ellen *et al.* 2009). From that study 15 membrane proteins from secreted membrane vesicles in *S. solfataricus* were in common with proteins identified here. However, those results were limited to identification only, and unlike here, no quantitative information was obtained. In other studies of archaeal membrane proteomes, for example in the analysis of the extreme halophilic archaeon *Halobacterium sp*. NRC-1 using an ion trap, 426 proteins were identified (not quantified), and of these, only 165 proteins were expected to be membranerelated (Goo *et al.* 2003). In a less common quantitative study of membrane proteins in *Halobacterium salinarum*, 155 membrane proteins were identified and quantified using DIGE and 16-BAC/SDSPAGE techniques.7

2-DE cannot (usually) detect proteins with molecular weights >200 kDa or <10 kDa (Graham *et al.* 2007). Moreover, this technique may not be suitable for membrane protein analysis since there is a restriction in the use of buffers for the isoelectric focusing (IEF) step (for solubilization, membrane proteins often require certain detergents; however, these detergents may be incompatible with IEF). In terms of the quantitative proteomic analysis of *S. solfataricus* , we observed that most analyses are gel-based, and it is widely recognized that this approach can take longer than shotgun proteomics work- flows (Chong & Wright 2005). In recent years, quantitative proteomic analysis based on shotgun workflows (e.g., iTRAQ or TMT) have been applied widely

to applications in human cells, rat, *Saccharomyces cerevisiae* and mesophilic bacteria such as Escherichia coli, but application to *S. solfataricus* is still rare. Only a few studies, mostly carried out by our group, have been published to date (Snijders *et al.* 2005, 2006; Chong *et al.* 2006, 2007b, a). A common theme among these limited iTRAQ-based studies, though, is the numbers of proteins (from whole cell extraction including both soluble and insoluble protein) identified and quantified from each experiment has been limited to <300, including both soluble and membrane proteins (Chong *et al.* 2007a, b). In this current investigation, we measure up to 246 membrane proteins with a single analysis or up to 373 using different analysis combinations.

As mentioned above, many quantitative proteomic analyses of *S. solfataricus* based on gels or shotgun techniques have been published. However, none of these studies focused on global quantitative membrane proteomics, despite the important role membrane proteins play in cellular process including energy transduction (e.g., ATP generation via oxidative phosphorylation), signal transduction (e.g., nutrient sensing), and transport (e.g., import of sugars/peptides). This is of special relevance for archaea such as *S. solfataricus* , since relatively limited information is available on proteins and mechanisms of these processes. In *S. solfataricus*, eukaryal-like protein kinases and phosphatases and few target proteins of regulatory phosphorylation-dephosphorylation have been identified, but the involved mechanisms are still unknown.

Here, a new iTRAQ workflow was developed for *S. solfataricus* membrane protein analysis where a proteome comparison between cells growth under optimum temperature (80°C) and lower temperatures (70 and 65°C) was made to gain an understanding of this microorganism under reduced

temperatures. The application and combination of data from 3 different iTRAQ experiments provides validation for quantitation data.

Data here, when combined with soluble protein results (data not shown), results in a total quantitative measure of approximately 1000 proteins. This combination provides rich information for understanding of *S. solfataricus* at reduced temperatures, especially for carbon central metabolism (CCM) where 43 proteins were differentially quantified (data unpublished) from 57 predicted CCM proteins. These data, when combined with other -omics-level data of others in the SulfoSYS consortium (www.sulfosys.com/) will provide sufficient information to begin to construct and build up an in silico model of this archaeon. A large number of quantified membrane proteins (395 proteins) here represent a large subset of the *S. solfataricus* membrane proteome (883 proteins predicted based on Gravy score) for cells grown under reduced temperatures (65 and 70°C) compared to optimum conditions (80°C). In this study, we also were able to measure up to 53 ribosomal proteins. Of these, 17 were differentially regulated. This might have resulted from contamination during membrane preparation steps. However, it has been observed that up to 50% of the ribosomes in archaea are found to be attached to the membrane by specific interactions (Ring & Eichler 2004)

The high abundance of some ribosomal proteins might have resulted from an increase of membranes attached to the ribosomes (enhanced production of membrane proteins, pseudoperiplasm proteins, and extracellular proteins), rather than an increased concentration of the ribosomal proteins themselves.7 Deeper biological discussion of regulated proteins and their wider context is the subject of future work.

Conclusions

iTRAQ was successfully applied for quantitative membrane protein analysis of *S. solfataricus* P2 using a modified protocol. In this case, 284 proteins were detected (where 275 proteins were found with g2 distinct peptides) of which 246 were membrane proteins. The modified protocol was suitable for analyzing membrane proteins especially integral membrane proteins, offering potential application of this technique for quantifying membrane proteins in other organisms. Combining all 3 different iTRAQ experiments resulted in 395 proteins being quantified (with more than 2 distinct peptides) of which 373 were membrane proteins. Although over 80% of the quantified proteins remained unchanged in expressions when temperatures were reduced compared to the 80°C optimum, the numbers of regulated proteins provided sufficient information to begin to understand the temperature response of this archaeon. Many processes such as IS elements, amino acid biosynthesis, nucleotides, lipids and transportation were induced at reduced temperatures.

Acknowledgments. We would like to thank the BBSRC (Grant Reference BBF0034201; SysMO - SulfoSYS) and the EPSRC for funding (GR/S84347/01 and EP/E036252/1). We also would like to thank Prof. Bettina Siebers for her careful and critical reading of the manuscript. Supporting Information Available: All protein and peptide data, as well as additional methodological figures/data are available as spreadsheets. This material is available free of charge via the Internet at http://pubs.acs.org

Chapter 5

Temperature promoter motif regulates gene expression in *S. solfataricus*.

Pawel Sierocinski, John van der Oost

Abstract

Transcription regulation in Archaea is not yet studied in great detail. Transcription regulation is a key level of prokaryotic homeostasis as it allows reacting rapidly to a changing environment. This strategy appears to be especially important in environments characterised by steep gradients of chemical and physical conditions, such as hyperthermophilic ecosystems. Here present we a microarray analysis of the thermo-acidophilic archaeon *Sulfolobus solfataricus* and identify a putative regulatory motif that may be involved in transcriptional regulation upon temperature shifts. The motif is strongly conserved across phylogenetically related Archaea. The potential use of this regulatory system in biotechnological applications is discussed.

Introduction

The transcription machinery of Archaea has to be seen through the evolutionary history of the domain. Archaea are uniquely positioned on the phylogenetic tree of life (Bell & Jackson 1998) as, at the same time, an early split from bacteria-like ancestor, and the putative ancestral model of eukaryote transcription machinery. The core of the archaeal RNA polymerase distantly resembles the simpler, bacterial one. However, the archaeal RNAP

complex has multiple additional subunits very similar to its eukaryotic RNAP-II counterpart (Kwapisz *et al.* 2008). This agrees well with the fact that the transcription process in Archaea proceeds similarly to that of Eukaryotes, with a TATA-box-containing promoter region that is being recognized by a TATA-binding protein (TBP) and a transcription factor B (TFB) that allows for recruitment of the RNAP tot the promoter. On the other hand, transcriptional regulation in Archaea has been demonstrated, at least in some cases, to be more related to the bacterial than to the eukaryotic system (Peeters & Charlier 2010; Gindner *et al.* 2014). The split between the archaeal and the bacterial domains occurred 3.5-3.8 billion years ago (Weiss *et al.* 2016) which could have largely obscured any genetic similarities between the particular regulators, making comparative analysis difficult. This means that in order to discover most of the cryptic archaeal regulatory mechanisms, phylogenetically unbiased approaches are required such as transcriptomic analysis.

Yet there have been some discoveries of highly regulated transcription in Archaea in general, and in *Sulfolobus solfataricus* in particular. Studies involving arabinose metabolism have shown a conserved regulatory sequences just upstream of the TATA-box of the genes involved in pentose metabolism in *S. solfataricus* (Brouns *et al.* 2006). The same is true for a well-described palindromic sequence in the genes regulated by the regulator ss-LrpB, which regulates its own transcription as well as that of other genes, including pyruvate ferredoxin oxidoreductase (Peeters & Charlier 2010). The common trait of these regulatory downstream sequences is their palindromic (or semi-palindromic) character, either reflecting association with a dimeric regulator or a secondary structure in case it would be transcribed as RNA. In addition, the few available archaeal transcriptional regulators appear to block

transcription rather than activating it, due to the position of their binding site upstream the TATA box.

Even though Archaea have been studied for decades and some of these studies have focused on transcriptomics (Walther *et al.* 2010), very few transcriptional regulators and corresponding regulatory sequences have been identified and described in literature to date. A possible explanation for having bacteria-like regulation of transcription of the eukaryal-like archaeal system is the different half-lives of RNA transcripts in organisms with and without a nucleus. While prokaryotic mRNAs generally have half-lives of minutes, several orders of magnitude lower than the stability of their proteins; eukaryotic mRNAs can survive from hours to days, i.e. only 5-fold less than their proteins (Schwanhäusser *et al.* 2011; Pérez-Ortín *et al.* 2013).

This is partially linked with the length of the cell cycle and variability of the environment prokaryotes occupy. For example *Sulfolobus solfataricus* typically occurs in terrestrial acidic hot springs, characterised by very steep temperature gradients (Brock *et al.* 1972). This has forced the organism to evolve a very robust metabolic network. It does grow almost equally well around its optimum (80°C) as well as at the thresholds of its viable temperature range (65-92°C) (Grogan 1989). The adaptations to such a lifestyle occur at different levels: protein regulation (through transcriptome modulation, translation regulation, etc.), enzyme activity tuning (allosteric regulation), and metabolic pathway versatility, by having multiple solutions present in the genome. It is the combination of these regulatory mechanisms that allows the organism to occupy such a rapidly changing niche, but since the proteome is much more stable, it is the transcriptome that is responsible for rapid response in the timescale of minutes.

Our previous research (Zaparty *et al.* 2009), focused on examining all the layers of regulation (eg. transcriptomics, proteomics, metabolomics and enzymatic regulation) in *S. solfataricus* during growth under optimal conditions and at the low-end of its temperature range. The results of that study revealed a number of genes that were significantly up/down- regulated when comparing the growth in the optimal 80°C as opposed to the 70°C (Table 1), which is close to the upper temperature limit of *Sulfolobus* growth (Grogan 1989). Most of the differentially expressed genes found were not assigned a function, but that is to be expected as sub-optimal growth conditions in thermophiles are, at best, poorly studied.

Results

Comparing gene expression of *S. solfataricus* cultivated in controlled fermenters at 80°C to growth at 70°C, revealed five genes up-regulated and five genes down-regulated by a factor of at least four (Table 1). Using this dataset, we screened the promoter regions of these genes for significantly overabundant motifs using RSAT Tools (Medina-Rivera *et al.* 2015). The search showed that two up/down-regulated genes (SSO0503 and SSO3098, paralogs with 50% nucleotide and 15% amino acid identity) both have a very strong palindromic motif ATTACCCSNNGGGTAAT located in their promoter region, just upstream of their (predicted) TATA-box (Table 2).

Table 1: Significantly regulated genes, 80°C vs. 70°C growth conditions.

Gene ID	Log_2 ratio 80°C vs. 70°C	p-value	Annotation
SSO3053	-2.29	0.00001	Maltose ABC transporter, maltose binding protein
SSO2797	-2.16	0.00015	Conserved hypothetical protein
SSO3000	-2.11	0.00000	Thermosome gamma subunit (thermophilic factor 55) (ring complex gamma subunit)(chaperonin gamma subunit) (thsC)
SSO0698	-2.07	0.00001	SSU ribosomal protein S5AB (rps5AB)
SSO3043	-2.05	0.00135	ABC transporter, binding protein
SSO0998	1.99	0.00000	Quinolinate synthetase (nadA)
SSO2549	2.27	0.00000	Amino acid transporter, putative
SSO0769	2.56	0.00000	Activator 1, replication factor C (RFC) large subunit (rfcL)
SSO0816	3.80	0.00000	Hypothetical protein
SSO0503	4.14	0.00000	Conserved hypothetical protein
SSO3098	4.16	0.00000	Conserved hypothetical protein

Strikingly, the motif was GC rich (41%), considering that *S. solfataricus* promoter areas (up to 400 bp from the ATG codon, not counting sequences overlapping with other genes) have an even lower GC content (30.9%) than the average for the genome. The fact that the motif was present in the promoter region of two paralog genes may suggest it is the result of a duplication rather than a conserved, functionally important motif. However the part of the promoter close to the TATA-box, where motif is located has 57% identity – higher that the identity of the paralog gene itself, while parts further upstream show almost no identity (10%). The latter is expected as promoter sequences are generally poorly conserved in Archaea. The same relative

difference in expression levels of the genes applies to the comparison between 65°C and 80°C and 70°C and 80°C. There was no difference in transcription between 65°C and 70°C in the microarray analysis pointing towards a sharp switch-like regulation of the analysed genes above 70°C.

To check if the sequence similarity was due to chance alone, we further looked at the promoter regions of the orthologs of these genes in related hyperthermophiles. We found the promoter motif is highly conserved upstream of the orthologs, both of the *Sulfolobus* genus and of closely related genera (Table 2). Most Single-Nucleotide Polymorphisms (SNPs) we found were located in the variable three-nucleotide region in the middle of the palindromic sequence (23 of the 36 SNPs found), see Figure 1. The phylogenetic distances in terms of the motif sequence suggest that at least to some extent, the promoter region of the examined genes is not following the phylogenetic relationship in the same way that the 16s rDNA genes do, eg. *Metallosphaera sedula*, even though closely related to *M. cuprina*, clusters better with *S. tokodaii*, while M. sedula has the motif sequence much more related to other *Sulfolobales*. On the other hand, more distantly related species, such as members of *Volcanisaeta ssp*. Are clearly different from other examined species. This points to either convergent evolution or horizontal gene transfer in closely related species, while divergent evolution shaped the motif region in further related genera.

Table 2: Prevalence of the discovered motif. Start and End indicate the position of the motif in relation to the translation starting site.

Gene ID	Start	End	Sequence found	Species	Ortholog of SSO0503	Ortholog of SSO3098
Ahos_0440	-45	-31	TAAGGGGTACCCTAA	*A. hospitalis*	Yes	No
Cmaq_0860	-20	-4	TTAGGGTAACCCGAA	*C. macquilensis*	Yes	No

Mcup_0755	-21	-5	TTTGGGTAACCCTAA	*M. cuprina*	Yes	No
Msed_1464	-19	-3	TTAGGGTTACCCTAA	*M. sedula*	Yes	No
Saci_1798	-30	-14	TTAGGGTATCCCAAA	*S. acidocaldarius*	Yes	No
SiH_1580	-28	-12	TTAGGGTAACCCTAA	*S. islandicus HVE10 4*	Yes	No
LD85_1859	-28	-12	TTAGGGTAACCCTAA	*S. islandicus L D 8 5*	Yes	No
LS215_1730	-28	-12	TTAGGGTAACCCTAA	*S. islandicus LS 215*	Yes	No
M1425_1604	-28	-12	TTAGGGGAACCCTAA	*S. islandicus M14 25*	Yes	No
SSO0503	-29	-13	TTAGGGCTACCCTAA	*S. solfataricus*	Yes	No
ST2474	-18	-4	TTCGGGCTACCCTAA	*S. tokodai*	Yes	No
Vdis_1512	-15	1	TTAGGGTAACCCTAA	*V. distributa*	Yes	No
VMUT_0103	-45	-29	TTAGGGTAGCCCTAA	*V. moutinovskia*	Yes	No
VMUT_2197	-56	-40	TTAGGGTTACCCTAA	V. moutinovskia	No	Yes
SiH_2207	-21	-7	TTAGGGTTACCCTAA	S. islandicus HVE10 4	No	Yes
LD85_2538	-22	-6	TTAGGGTTACCCTAA	S. islandicus L D 8 5	No	Yes
Vdis_1388	-118	-104	TTAGGGTAGCCCTAA	V. distributa	No	Yes
Vdis_1388	-82	-68	TTAGGGTAACCCTAA	V. distributa	No	Yes
VMUT_2197	-20	-4	TTAGGGCTACCCTAA	V. moutinovskia	No	Yes
Ahos_0044	-43	-27	TTAGGGTTACCCTTA	A. hospitalis	No	No
Mcup_1220	-19	-3	TTAGGGTAAACCTAA	M. cuprina	No	No
Cmaq_1292	-21	-5	TTAGGGTAACCCGAA	C. macquilensis	No	No
Vdis_1220	-20	-6	TTAGGGTAGCCCTAA	V. distributa	No	No
VMUT_2067	-21	-5	TTAGGGTAACCCAAA	V. Moutinovskia	No	No

Ahos_1040	-33	-17	TTAGGCTAACCCTAA	A. hospitalis	No	No
Msed_0409	-34	-18	TTAGGTTAACCCTAA	M. sedula	No	No
SiH_0968	-75	-61	TAAGGGTTACCCTAA	S. islandicus HVE10 4	No	No
SiH_0968	-51	-37	ATAGGGTAACCCTAA	S. islandicus HVE10 4	No	No
ST0796	-25	-9	TTAGGGTTACCCTTA	S. tokodai	No	No

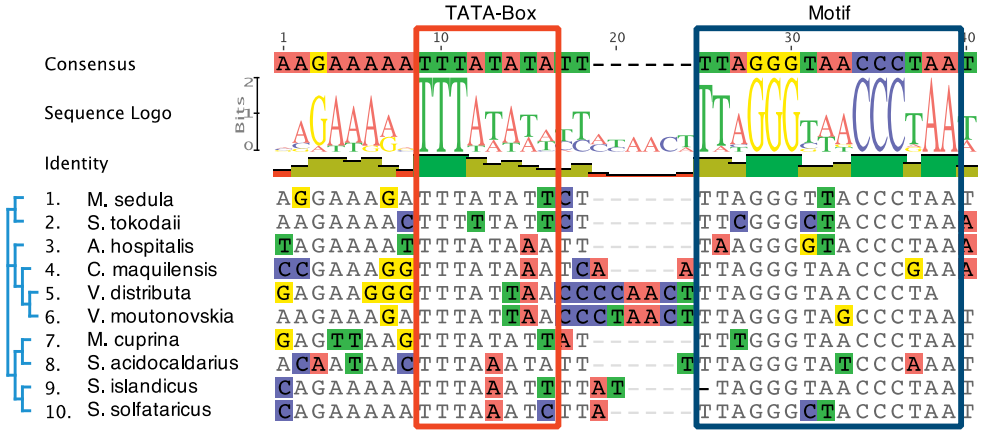

Figure 1: Web Logo and upstream sequences of the promoter region of the analysed motif across hyperthermophilic Archaea. The red frame indicates the motif and the blue one, TATA-Box. See Table 2 for gene annotations.

Moreover, the significance of that finding is confirmed by the conservation of the position of the motif in relation to the transcription start site (TSS, as deduced from predicted transcription factor-B Recognition Element (BRE) and TATA box; (Figure 2) as well as ribosome binding site (RBS) and translation

initiation site (TIS). In most cases the sequence can be found 4-10 nucleotides (nt) downstream the TATA-box, just upstream the TTS (located 25 nt downstream the TATA box) and 20-35 nt upstream the TIS. The positioning is consistent with other known transcription regulators in Archaea, suggesting that the transcription can be blocked by a regulator binding to the motif through blocking the TATA-binding protein from binding to the TATA-box, thus hampering recruitment of the RNA polymerase. An alternative way of regulation would be to have multiple transcriptional start sites for a single gene, where the motif would be incorporated in one of the alternative transcripts and cause the transcription to terminate prematurely during unfavourable conditions. We examined the second possibility by looking at the transcriptome map of *S. solfataricus* (Wurtzel *et al.* 2010) and found that there is an alternative transcript present that encompasses the motif sequence seen predominantly when cells were grown on cellulose.

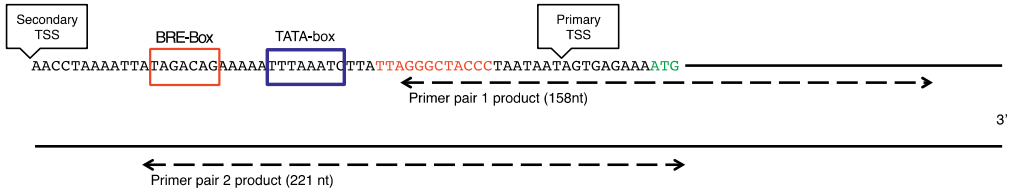

Figure 2: Architecture of the SSO0503 transcript, with two Transcription Start Sites (TSS) indicated and the targets for the primers used in the experiment indicated. Promoter sequence is specified, gene sequence is symbolised by the black line and not to scale with the promoter sequence.

In an attempt to reveal how the regulatory sequence works, we used an RT-qPCR approach. The microarray analysis conducted initially had a relatively low temperature resolution. In order to determine the regulation tipping point, we tested the cells grown on glucose or on a mixture of glucose and cellulose between 65°C and 80°C, in 3°C steps. Two primer-pairs were used, one targeting mid-gene section of SSO0503 and the other one targeting the region between the canonical TSS and the motif of the same gene, in order to account for the differences between both transcript quantities (Figure 2).

In both feeding regimes, the gene was overexpressed at the upper temperatures of the 65-80°C range (Fig. 3). Importantly, this difference was bigger when cells were grown in the presence of cellulose, rather than on glucose alone, in agreement with expression patterns reported previously (Wurtzel *et al.* 2010). The ratios for both primer pairs correlate significantly (Spearman $\rho = 0.88$, $p = 0.001$) Indicating either that both transcript are showing changes at a similar level, or that the shorter transcript is absent. The transcription rate was the highest at 74°C, 77°C and 80°C suggesting a temperature-dependant response with the gene being transcribed in a dose-dependent fashion rather than by a sharp switch in transcription.

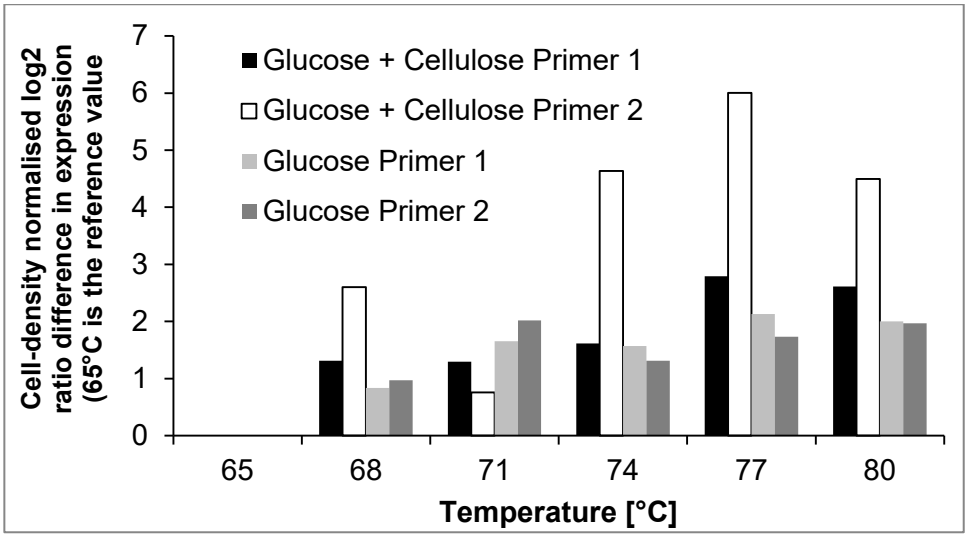

Figure 3: Expression patterns of SSO0503 measured with RT-qPCR under different temperature regimes dependant on the diet. Expression values in $\log_2$. Primer 1 refers to the primer pair targeting the canonical transcript, Primer 2 refers to the primer pair targeting the alternative transcript.

Discussion

Transcriptional regulation in microorganisms is a key regulatory mechanism allowing a fast response to changing conditions. As *Sulfolobus* lives in hot mud springs where temperature can change rapidly over a steep gradient, it would make sense if at least some of its transcription regulation would respond to temperature shifts. As proteins can survive for hours while mRNA only lives for minutes, there is a big difference in the abundance of transcripts and the products of translation. There are thousand times more proteins than mRNA particles in a mammalian cell, with this difference only getting more pronounced in unicellular organisms.

The discrepancy between the transcription rate and the level of proteins present in the cell can be explained by multiple factors, a major one being the cell cycle length and variability of environment with the lack of homeostasis mechanisms akin to the ones of higher organisms. This should in theory promote active transcription switches. In the situation where transcribing a gene causes its product to stay in the cell for more than a doubling time of the organism, a switch that shuts transcription can stop accumulating unnecessary proteins until favourable conditions arise. In order to elucidate the molecular mechanism controlling the up-regulated expression of a gene at elevated temperatures, we analysed the upstream promoter regions of the genes in question. We looked at conserved motifs, and because of the character of the previously described regulatory sequences, we put particular weight on the palindromic sequences.

Reanalysing previously obtained data, we were able to pinpoint an interesting palindromic motif upstream genes that were highly regulated by changes in temperature. We furthermore confirmed that indeed there is a significant up-regulation of the gene controlled by this motif under the tested conditions. This gives us a potential tool in biotechnological applications, where lowering the temperature could switch off a process when needed and allow for detoxification of the environment. We must, however stress that if the regulation is at the protein level, this would limit the tool to *Sulfolobaceae* and related genera that do possess the motif. This would not be a severe limitation as there are not many organisms that could thrive in the conditions where the regulation is viable, but is worth noting.

SSO0503, is the gene we focus on in this study. It codes for a putative membrane protein of unknown function. It has six predicted trans-membrane

helices. Unfortunately bioinformatics analysis yielded no information on the function of the gene, which would be interesting to determine the importance of its regulation in different temperatures. The gene has two types of transcripts, as shown by (Wurtzel *et al.* 2010), see Figure 2 for details. An interesting follow-up of this experiment would be to knock-out the gene and check the viability of such knock out in different temperature conditions.

A most likely explanation for the observed difference in transcription is a regulatory protein binding to the motif site, but other explanations are possible. One of them is that the strong palindromic motif causes a hairpin structure in the transcript at low temperatures preventing translation. This possibility is less likely, however, as we see a similar pattern of temperature dependent increase of transcription with the longer transcript that contains the motif and the shorter one devoid of it. Our results suggest either (i) that both of them are transcribed, but independently of the type, they react exactly in the same way to the condition changes in terms of temperature, or (ii) that the longer transcript, that encompasses the regulatory motif, is so dominant in numbers, that the shorter transcript levels are of little importance. Both are possible but the second option appears much more likely, as it is more parsimonious and fits with the general pattern of promoter regulation of the gene.

All the data point towards the regulation on the level of the promoter motif. This is further strengthened by the fact that the motif itself is strongly conserved across the genera we looked at (see Table 2). Although the low identity level on both the nucleotide/amino acid level of the two *S. solfataricus* genes/proteins (SSO0503 and SSO3098) indicates that they diverged as

paralogs relatively early, the promoter region in which the motif is localised is much more conserved than the gene itself.

Future work is required to elucidate the mechanism: identify the potential regulator that binds the motif, or show that the motif also functions autonomously as a thermometer riboswitch in distantly related thermophiles. The latter possibility would make this system a useful tool in genetic manipulation.

Materials and Methods:

Microarray data:
The microarray data used in this paper has been previously published in Zaparty et al 2010, where a more detailed experimental description is available. The cells grown at 70 and 80°C in a fermenter were pelleted, rapidly cooled and later used to extract the RNA fraction using mirVANA kit (Ambion). RNA was converted to cDNA and labelled with Alexa Dyes (Alexa 647, Alexa 555; Invitrogen). The labelled cDNA was hybridised with a custom-made oligonucleotide microarray and scanned (GenePix Pro 4000B, Axon). The data was further normalised to account for transcript level differences between the samples and analysed using MIDAS software (TIGR).

Search for the regulatory motif
The genes that were most up- and down-regulated in the microarray data (more than 4-fold difference between both tested conditions) were tested using Regulatory Sequence Tools (RSAT). Their promoter sequences (from the predicted translation initiation codon down to -400 bp downstream of the gene or down to the neighbouring gene, whichever is closer) were input in the

tool and searched for repeating sequences and repeating palindromic sequence (Motif Search and Dyad Search). The analysis yielded two significant GC-rich dyads (GGGNNNCCC) in the promoter regions of SSO0503 and SSO3098 genes. Further analysis showed that the motifs are located at the same position in relation to the TSS and can be expanded to ATTAGGGNNNCCCTAAT. In order to assess whether the motif was purely coincidental or on contrary, well conserved, we searched for its presence in other bacteria. We used BLAST to identify the motif in other bacterial and archaeal species and found numerous hyperthermophiles containing it. A re-analysis using a multi-species option in the RSAT, showed that it is present in the same position as found in *S. solfataricus* in relation to the TSS in other hyperthermophilic genera (See Table 2) and is highly conserved. In order to check if there are no alternative transcripts for the genes from S. solfataricus, we looked through the supplementary material (see Wurtzel et al. 2010) and found that apart from the canonical TSS, there is a second type of the transcript present that encompasses the motif.

Cells and growth:

The experiments have been performed using the *S. solfataricus* P2 type-strain, grown on a chemically defined medium (Brock *et al.* 1972; Zaparty *et al.* 2009) with 0.3% glucose as the carbon source with the addition of 0.1% cellulose in the second experiment. The cells used in the experiment were grown in a 400 ml fermenter at 80°C and pH of 3.00. The fermenter was mixed by aeration with sterile air. After reaching the OD of 0.5 the fermenters have been gradually cooled down by 3°C at a time to reach the temperature of 65°C. Each drop in temperature was sustained for 2h in order to make sure that the organism can change its transcriptome in response to the conditions.

Two 20 ml samples were taken at each time point for RNA analysis, cooled down in liquid nitrogen and spun down at 4°C (15 min, 3500g).

RNA isolation and RT-qPCR

RNA was isolated from cells grown at a range of temperatures (65°C to 80°C, every 3°C) using Trizol extraction (Chomczynski & Sacchi 1987). After the isolation the RNA was quantified using Nanodrop 2000 (Thermo Fisher Scientific). In order to eliminate any DNA contamination in the sample, they were diluted 100 fold and aliquots of 30 µl were DNAse treated using DNA free DNA Removal Kit (Ambion). We used two different primers targeting the gene (Supplementary Material Table 1) – one for the mid-gene region, one targeting the alternative transcript previously found. In order to normalise the tested RNA for the differences in cell density, we also looked at the quantity of the 16S rRNA transcript to normalise for cell numbers. We used The RT-qPCR was run using Brilliant III Ultra-Fast SYBR® Green QRT-PCR Master Mix (Agilent Technologies) on StepOne Real Time PCR system. After the 10 minute RT step at 50°C, 3 minutes at 95°C, we run a 40-cycle programme (5 seconds 95°C, 10 seconds 60°C) and a full melting curve. Data has been analysed using LinReg.

Chapter 6

Evolution of *S. solfataricus* in fluctuating temperature conditions.

Pawel Sierocinski[1,2], Jesica Soria Pascual[2], John van der Oost[1], Chris Bryan[2,3]

1 Laboratory of Microbiology, University of Wageningen
2 Environment and Sustainability Institute, University of Exeter
3 Geomicrobiology and Environmental Monitoring Unit, French Geological Survey

Abstract

Evolution in extreme conditions is vastly understudied. Here we look at experimental evolution of *Sulfolobus solfataricus,* a thermophilic archaeon that grows optimally at 80°C. We set out to analyse its phenotypical response to fluctuating and stable suboptimal conditions (65°C and 84°C). In particular we tested whether fluctuation of conditions can select for more robust generalists that are capable to outcompete specialist cells evolved in stable conditions. We found that evolution under temperature fluctuation conditions promotes the ability of the cells to thrive in the hotter than optimal temperature range, however not at the optimal temperature. Furthermore, the cells adapted to cold temperature have shown hindered growth both in optimal and higher than optimal temperature range. These results suggest that cold shock conditions may play an important role in the generation of *S. solfataricus* genetic variation. In fluctuating conditions this leads to increased fitness, while in cells not exposed to high temperatures, deleterious mutations accumulate leading to a decreased fitness.

Introduction

S. solfataricus is a model organism for hyperthermophilic aerobic Archaea. Its metabolism and adaptation to a thermophilic lifestyle was examined using multiple techniques ranging from cultivation-based methods (Grogan 1989), enzymatic assays (Ettema *et al.* 2008; Zaparty *et al.* 2009), and -omics approaches (Tachdjian & Kelly 2006b; Zaparty *et al.* 2009). Surprisingly, very little has been done to assess *S. solfataricus* using an experimental evolution approach (McCarthy *et al.* 2015).

Evolution experiments have been used with success (Kawecki *et al.* 2012) to look at adaptation to novel environments (Riley *et al.* 2001; Elena & Lenski 2003), host-pathogen interactions (Buckling & Rainey 2002; Hall *et al.* 2011), determining function of unknown genes (Velicer *et al.* 2006) or major transitions in evolution (Blount *et al.* 2008; Ratcliff *et al.* 2012). Experimental evolution was also used to look at temperature adaptation in *E. coli*, showing the potential of this approach in determining the key players responsible for being able to cope with temperature ranges out of optimum (Bennett & Lenski 1993).

Sulfolobus solfataricus has over 40% of its genes annotated as “hypothetical” and “conserved hypothetical”. For the majority of the other genes functions are assigned based only on the similarity with mesophilic genes and proteins. Given the vastly different growth conditions of thermophiles, and extremophiles in general, this does not necessarily translate to having same function. On the other hand, analysis of genomic adaptations might reveal genes either that are redundant and are selected against in different conditions. Reproducible patterns of mutations can pinpoint traits beneficial in

a particular condition. Especially reproducible patterns of selected beneficial mutations can be helpful in elucidating the importance of a given function (Blount *et al.* 2008; Beaumont *et al.* 2009).

We test the organism in both stable and fluctuating conditions in order to look at the possibility that fluctuating conditions provide stronger selective pressure that gives rise to higher fitness (Niinemets & Valladares 2008). In our case the hypothesis is that generalist cells evolved in an environment fluctuating between the two extreme conditions will have a higher fitness than the specialist cells evolved either in a cold or in a hot environment. This has been shown to be the case in microbial communities in a mesophilic ecosystem (Ketola *et al.* 2013), but *S. solfataricu*s has unique features that add depth to such analysis.

S. solfataricus has a genome with an unprecedented number of IS elements (covering approx 10% of its genome (Brügger *et al.* 2004) that are speculated to be the main driver of its evolution (Martusewitsch *et al.* 2000; Blount & Grogan 2005). The IS elements have been shown to be active at both ends of the viable temperature spectrum of *S. solfataricus,* suggesting that they play an important role in the evolution of the species. Indeed the comparison between different isolates (Brügger *et al.* 2004) or even looking at strains used in different labs (Zaparty *et al.* 2009) shows a pattern of IS element shuffling that is an important driver of evolution within this particular species. This might be an adaptation to the lifestyle of *S. solfataricus* as a planktonic organism (as evidenced by its weak biofilm formation) in the hot springs that is a subject to rapid changes and has to cope with them equally rapidly. This is opposed to *S. acidocaldarius* that forms strong biofilms, therefore being capable of occupying a much more stable niche (Koerdt *et al.* 2011). Moreover, living in

extreme conditions limits evolutionary capabilities of the organisms. Simply the possible number of solutions to a given protein sequence is lower than in mesophilic organisms due to the fact it has to simultaneously fulfil a function and be capable to withstand the adverse conditions, enforcing the density of hydrophobic domains not needed in weaker thermophiles and mesophiles (Szilágyi & Závodszky 2000). This creates an evolutionary trade-off: generation of variation within the populations will inadvertently create deleterious mutations at a higher rate than in the mesophiles, thus hindering evolvability potential of the species.

In order to test these predictions we grew *S. solfataricus* in six fermenters: two at constant 65°C, two with constant 84°C and two shifting between 65°C and 84°C every transfer. Each time we transferred 13.5% of the culture to a fresh medium, to optimise generation and propagation of new variants during transfers (Wahl & Gerrish 2001). After 8 transfers, the experiment was terminated and samples were tested for their fitness.

Materials and methods:

The strain used was *Sulfolobus solfataricus* P2. Cells were grown in the standard Sulfolobus medium as described earlier (Brock *et al.* 1972): 1.3 g $(NH_4)_2SO_4$, 0.28 g KH_2PO_4, 0.25 g $MgCl_2 \times 7H_2O$, 0.07 g $CaCl_2 \times 2H_2O$, 0.02 g $FeCl_2 \times 4H_2O$, 1.8 mg $MnCl_2 \times 4H_2O$, 4.5 mg $Na_2B_4O_7 \times 10H_2O$, 0.22 mg $ZnSO_4 \times 7H_2O$, 0.06 mg $CuCl_2 \times 2H_2O$, 0.03 mg $Na_2MoO_4 \times 2H_2O$, 0.03 mg $VOSO_4 \times 2H_2O$ and 0.01 mg $CoCl_2 \times 6H_2O$. Demineralized milliQ water was used to prepare all the solutions used. Prior to autoclaving, the pH of the medium was set to 3.5 using H_2SO_4. The sterile iron solution was kept in the dark at RT and added to the medium before inoculation.

The cells were grown in air-lift fermenters with gas addition as a way of mixing the medium. Initially, each fermenter had 487.5 mL medium added, and before inoculation 7.5 ml of filter sterilised 20% glucose solution set at pH 3.5 was added to each of them along with 5 ml of Sulfolobus culture at OD_{600}=1.0. The cells in the fermenters were grown until the OD of 1.0 was reached in all fermenters, after which 13.5% of each fermenter (67.5 mL was transferred to 425 mL fresh, pre-warmed medium with 7.5 mL 20% glucose solution. The 13.5% transfers were chosen based on the findings of Wahl et al, showing this value as an optimal one for maximising the arising of new mutations and at the same time making sure that they will be as well transferred rather than lost due to dilution.

The fermenters were grown at 65°C and 84°C – temperatures at the low and high end of Sulfolobus tolerance range. Two fermenters were grown continuously on low temperature, two on high and two fermenters were grown in shifting conditions where temperature was switched from high to low and vice versa after each transfer. The experiments concluded after eight transfers. Cells from each fermenter were harvested and stocked for fitness assays and cell paste collected for DNA extraction. The DNA was isolated using MolBio Soil kit standard protocol and sent for sequencing.

After eight transfers we regrew the harvested cells in order to measure the growth rate and carrying capacity in the conditions used in the experiment as well as in the optimal conditions to which the ancestral strain was adapted (80°C). Fitness assays were done in shaking incubators at 65°C and 80°C and using the fermenters at 84°C (due to technical constraints this test was impossible to conduct in a shaking incubator). The two values we looked at

were the growth rate and the carrying capacity (as the OD_{600} after 7 days of cultivation).

Results:

The results show that there are significant differences between the evolution under different conditions. The main finding (Fig. 1) is that cells cultivated at low temperature stop performing well in the optimum and heat stress conditions, being outperformed by the cells grown both at constant heat as well as in the shifting conditions. Adaptation to the cold temperature in this time scale does not yield any measurable advantage over the cells grown in the higher temperatures or in the shifting conditions. While cells grown in the shifting conditions and at 84°C do equally well at the optimal growth temperature of 80°C, *Sulfolobus* evolved under the continuous heat stress grows slower at 84°C than the cells exposed to shifting conditions.

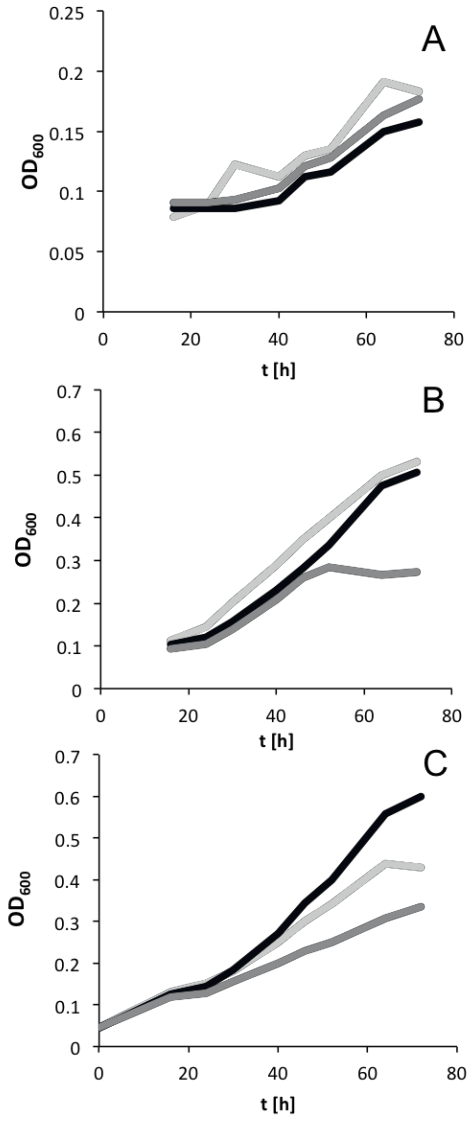

Fig. 1: Growth of evolved *S. solfataricus* cells at 65°C (A), 80°C (B) and 84°C (C). Cells were evolved at 65°C (grey line), 84°C (light grey line) and in shifting conditions (dark grey).

We have furthermore tested the carrying capacity (maximum sustainable cell density) of the cells when grown at 65°C and 80°C (Fig. 2), with results that confirm those of the growth curves. The cells evolved in shifting conditions and at 84°C performed significantly better than the cells evolved at the low temperature values when tested in optimal 80°C, while there are no significant differences when comparing the treatments at 65°C.

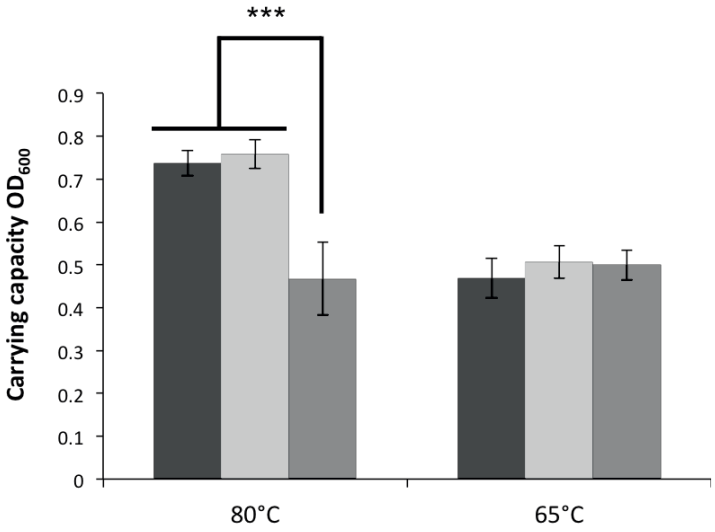

Figure 2: Carrying capacity of cells evolved in shifting conditions (dark grey), 84°C (light grey) and 65°C (grey). The carrying capacity of cells evolved at 65°C is significantly lower at 80°C (one-way ANOVA, $p < 0.01$, $F=34.4$).

Discussion:

We observed that cells evolved under strong cold shock selective pressure, lose their ability to grow robustly in the heat shock conditions, and more

importantly, even in the optimal temperature range. There are several possible explanations of that phenomenon. The most parsimonious one is that the *S. solfataricus* cells need far more adaptations to the hyperthermophilic lifestyle than to the survival in the colder conditions. Constant selection under cold conditions will lead to deleterious mutations in key genes involved in heat resistance capability, which will be detrimental when the cells are subsequently put in the hyperthermophilic conditions. In normal circumstances such deletions might take a long time to accumulate and the timescales tested in this experiment can be insufficient to measure such a drastic shift in performance. But in case of *S. solfataricus* the abundant IS elements can knock out multiple genes fairly rapidly if they are activated. The IS elements are speculated to be active outside of the optimal growth conditions (Brügger *et al.* 2004), therefore big genomic changes at the low end of temperature tolerance are to be expected.

Normal physiological conditions for *Sulfolobus* lives are defined by steep temperature gradients. There, such low temperature conditions are not likely to last long. The cells are either moved to much colder environment where their metabolism stops completely or back to optimal conditions where negative mutations will be likely purged. However our experimental setup allowed them to persist for multiple rounds of selection in such suboptimal environment. If the IS elements were active in the low temperature treatment, mutations that are very costly in high temperatures but incur no cost at 65°C can quickly accumulate, hampering the growth at 80°C and 84°C.

This observation is strengthened by another result: the cells grown under shifting conditions did not lose their ability to grow robustly at any the temperatures tested. Fluctuating selection pressure reduces the chance of

fixation of detrimental mutations in any conditions by periodically purging them before they can start to proliferate and dominate the population. Furthermore, it gives the chance for variants performing better in hotter temperature range to get selected during the 84°C stage of temperature fluctuation, leading to selection of phenotypes that do well both in 65°C and 84°C.

This is confirmed by our results. Cells grown in shifting conditions were slightly more robust at 84°C than the ones grown continuously in these conditions. This suggests, that genetic diversity IS-element shuffling, or any other mechanism, was generated in low temperature and selected for in the 84°C. The two important steps: generation f diversity and selection for successful variants occurred in separate environments. Such a mechanism is unsurprising for an organism that, like *S. solfataricus*, lives in rapidly variable conditions.

Another observation from this experiment is that while cells evolved at 65°C lose their fitness at higher temperatures, the reverse was not observed. The possible explanation for this phenomenon is that heat shock gradually degrades cell machinery, cold shock does not. Cold shock kills cells due to a sudden loss of cell wall functionality, which means that as long as the cell wall remains functional, other cellular mechanisms remain intact. The cellular machinery will work slower as enzymes lose their activity with the temperature drop, but they do not cease to operate, like they would in heat shock conditions. *S. solfataricus* cells evolved in high temperature do not lose their ability to grow at the lower border of their viability, because changes leading to it would also damage their performance in the high temperature range. At the same time the cold conditions knocking out genes that allow *S. solfataricus* to cope with heat stress in does not incur a fitness cost.

This relatively simple experiment shows how fluctuating selective pressure impacts short-term evolution of *Sulfolobus solfataricus*, but it has several limitations. To fully answer how adaptable hyperthermophiles are to stable conditions within their viability range, a longer experiment would be needed. These initial findings, along with longer term experiments (McCarthy *et al.* 2015), show that *Sulfolobus* is a potentially interesting model to study evolution. That is especially true in the conditions that, through IS-element mobility, allow rapid generation of variability in population.

Our initial findings show that cold shock conditions are potentially a target for such cold-shock boosted evolution. This can be further expanded to study the co-evolution of *Sulfolobus* and its parasites (viruses), adaptation to novel substrates (sugars), or using a combination of evolution and sequencing techniques to aid annotating of hypothetical proteins in *S. solfataricus*. These techniques can be harnessed in industrial applications where hyperthermophiles are potentially useful production organisms due to their robustness that enables efficient catalysis of chemical reactions under harsh conditions. But aside from applied aspects, with the affordability of high-throughput DNA analysis, *Sulfolobus* can be a useful model to address outstanding questions related to both microbial evolution and ecology.

Chapter 7

Summary, discussion and general conclusions

Pawel Sierocinski

This thesis describes a multi-approach analysis of the thermophilic archaeon *Sulfolobus solfataricus*, aiming to gain insight in its capacity to survive a wide range of temperatures (65-88°C) in its natural habitats. Coping with major fluctuations of temperatures requires a robust cell structure and metabolism, as well as a wide array of strategies that encompass regulation on transcription and translation levels. In addition, the activity and stability of enzymes that constitute the metabolic pathways should match with this temperature range. This creates a complex network of dependencies and implies that elucidating the actual mechanisms to adapt to fluctuating temperatures is challenging. I used the analysis of the transcriptome, proteome and looked at the enzyme activities of key metabolic pathways in order to try to look into the processes that allow *Sulfolobus* surviving between 65°C and 88°C with minimal impact on its growth rate. I also looked at the evolutionary mechanisms that drive *Sulfolobus* adaptation to the upper and lower limits of its niche.

While looking at a model organism in the laboratory setting, we should never lose sight of its ecology and its evolutionary history. *Sulfolobus* typically thrives in extreme terrestrial environments with elevated temperatures (60-90°C) and, high acidity (pH 1-3). The big issue that the organism has to cope with are steep gradients of temperature. Wafer thin margins separate hospitable zone from temperatures either too hot or too cold for the organism to thrive or survive. Moreover, with changes within the viable ranges, the biochemistry

available to the organism change as well with some metabolites undergoing spontaneous conversions in high temperatures, or needing enzymatic treatment in others.

The field of -omics analysis of microorganisms is a challenging one. It requires a good research question, thoughtful experimental design, mastering multiple techniques involved in the –omics analysis and the ability to distil signal from very noisy datasets. Any error at any stage will result in a bias in the final conclusion that might cause a fatal flaw to the whole interpretation. This thesis was a quest to approach all the mentioned issues in a systematic manner. Starting from asking a relevant question (Chapter 1) through looking the analysis of the available data (Chapter 2), setting the uniform set of methods that would allow reproducible testing of the model system from multiple angles (Chapter 3) towards using the acquired data to disentangle rules governing the proteome and the transcriptome of *Sulfolobus solfataricus* and putting it in the context of evolutionary adaptation to the variable and ruthless environment it lives in. We later wanted to fill in some unknowns with a more specific question. In order to achieve it, we looked at the improvement on the analysis of the membrane proteome of *S. solfataricus* (Chapter 4). This aimed at exposing any patterns in the expression of the first line of cellular defence and possibly linking transporters in the membrane with the metabolism at each tested temperature. The next step was looking in finer detail on the results from Chapter 3 by analysing the upstream promoter regions, and linking them with expression patterns. In Chapter 5, we analysed a conserved promoter motif present in multiple members of thermoacidophilic Archaea and linked it to the temperature-dependent transcription regulation. In the last Chapter (Chapter 6) the evolution of *S. solfataricus* is described. Monitoring evolutionary change is a strong tool for analysing whether either play an

important role in a certain condition, or are insignificant then, therefore being under a negative selection pressure.

All these methods combined give us only a glimpse of the complexity behind the 4 billion years of evolution that selected the hyperthermophiles and allowed them to thrive in conditions that are lethal to other forms of life. Each chapter should give the reader a different perspective on the processes that let *Sulfolobus* survive, adapt and adjust to the unfavourable conditions it lives in.

Chapter 1 gives a general overview of the history of discovery of hyperthermophiles. I try to look at the questions and applications that drove this field in the past. By doing that we specify the characteristic traits that hyperthermophiles possess, like their unique Central Carbon Metabolism, and thus look at potential research lines that would be interesting to pursue in relation to the hyperthermophiles. It also focuses on the developments that were made in the field and ways in which modern techniques allowed us to answer more and more questions related to the biology of thermophilic adaptation. I also introduce the history of transcriptome analysis of the thermophiles, which is the main focus of Chapter 2.

Chapter 2 is a review of the transcriptomic analysis of the hyperthermophiles from the advent of the technology, to the latest developments. I looked at particular cases that are the most relevant to this thesis. This chapter was planned as a way to look into the types of questions that can be answered using the transcriptomic analysis approach. Microarrays were adapted very early in the study of thermophiles, perhaps reflecting the exploratory drive of researchers involved in the study of such organisms. I show how a well-placed

question can greatly improve the capability of the technology to answer questions asked.

For example – without the microarray data it would be very difficult to elucidate the key players of the pentose pathway in *Sulfolobus solfataricus*. The array data not only helped pinpointing the key enzymes of the pathway, but also allowed the researchers to find the regulatory sequence responsible for the transcription regulation of the genes. Moreover, the absence of a key enzyme present in mesophiles let them test and conclude that parts of the pathway are based on spontaneous reactions.

I also looked at the discovery of the multiple origins of replication in Archaea, also made possible by careful analysis of the transcriptome, and combining transcriptomics with a very careful experimental design. Without a good way of synchronising the cells, transcriptomics would never yield intelligible results, showing the importance of careful experimental design.

I tried to speculate on the possible future of transcriptome analysis in thermophiles, showing the early impact of the sequencing techniques. The new methods differ greatly from the microarray technology. They can, in contrary to arrays, show evolution at the same time as they show transcriptome regulation. Moreover, they can look at the unknown species and multi-species communities and combine the study of thermophiles with ecology rather than focusing on the species mono-cultures. This is a key development, as the relative simplicity of thermophilic communities, as compared to soil microbial communities, or those related to the gut, allows testing simple ecological predictions in controlled environment, something that is technically difficult using the complex mesophilic species assemblies.

Transcriptomics of hyperthermophiles moved on since the inception of this thesis, but as the techniques changed, the general experimental approach stayed similar. Microarrays are no longer a viable technique of discovery, quickly made obsolete by the more efficient RNAseq technique (Marguerat & Bähler 2010). But the general design of the studies is still the same. Our approach of analysing *S. solfataricus* metabolism using the multi-omics approach has been successfully replicated with RNAseq replacing microarray as the transcriptomics analysis tool. This has shown how *Sulfolobus* manages to utilise fucose, by comparing growth on L-fucose and D-glucose (Wolf *et al.* 2016). The results of that paper show that the approach used in this thesis can be successfully applied to reveal many elusive features at the level of gene expression,

Transcriptomics used to be a standalone technique when the experiments related to this thesis were planned, but over time it has become more of a integrated, complementary analysis. Recent papers on hyperthermophiles use transcript analysis as an additional tool to confirm results obtained using other methods. One example of such study is experimental evolution of acid tolerance of *S. solfataricus* (McCarthy *et al.* 2016). Linking the evolutionary change within the genome and transcriptomics changes is a potent tool of identifying key features for tested conditions. Results from chapter 3 of this thesis would gain a lot of insight if we would have combined the evolution experiments with the other analyses used.

Another example of recent transcriptome analysis in *S. solfataricus* was an analysis in which the shift in transcripts during the different growth phases was investigated (Wang n.d.). Significant changes were observed in the

transcription pattern when looking at the log and stationary phase, providing insight in the signalling pathways responsible for the changes between different stages of the *Sulfolobus* population life cycles.

Even though it is rare, transcript analysis can still be done using microarray technique. In a well-defined experimental system, microarray analysis can be a very fast and efficient method. Recently it has been used in *S. solfataricus* to look at the CRISPR response during a infection of SSV1 and SSV2 viruses (Fusco *et al.* 2015). This research showed that SSV2 but not SSV1 virus caused the activation of the CRISPR associated genes and lead to decrease in the viral load. It is worth noting that Sulfolobus was an organism at the forefront of CRISPR discoveries (Tang *et al.* 2005), therefore CRISPR loci were used in the design of the microarrays used in this thesis.

Looking at the previous and recent research on the transcriptomics of *Sulfolobus*, one pattern emerges. Transcriptomics are an invaluable tool for the analysis of microbes, but they yield much more decisive results when analysing zero-one problems, like comparing two different substrates (Brouns *et al.* 2006; Wolf *et al.* 2016) or comparing viral infection and control (Fusco *et al.* 2015). When looking at more subtle interactions where differences lay in subtle interactions between multiple gene expression patterns, transcriptomic analysis may not be the best individual method of analysis. This may change when more recently developed machine learning approaches would be applied to transcriptomic datasets (Piles *et al.* 2019). However, it suggests that transcript analysis as of now is not the best tool for looking at small shifts in temperature, as in such cases one would expect a more incremental shift, not easy to disentangle from the noise using the statistical tools available.

Chapter 3 of this thesis combines the results of our analysis of the Central Carbon Metabolism with a full methodological toolbox for working with *Sulfolobus solfataricus*. The latter is a key resource for future researchers – allows them to access well described and tested methods, that in theory should make it easier for anyone to study *S. solfataricus*. The methods were developed not only for biologists, but also for modellers. The methods focus on the high reproducibility aspect of research. The clear standard operating procedures (SOPs) are a key feature that allows good feedback between the modeller and the lab-based biologist, enhancing good practice and good exchange of information.

Methods are also designed in such way as to encourage people not familiar with the biology of thermophiles to enter the field, making the adaptation to a new model organism as easy as possible. This has been a successful approach, as multiple groups used our methods to conduct their research in *S. solfataricus* (Blombach *et al.* 2015; Stark *et al.* 2017) and other related species, where some of the methods were transferrable (Wagner *et al.* 2012; Jiang *et al.* 2014).

Chapter 3 combines a method paper with a research paper. The results encompass genomics, enzymatic assays, bioinformatics, transcriptomics, proteomics and metabolomics. It is unsurprising that using such a broad approach paints a complicated picture. There is little agreement between the proteomic data and transcriptomics, suggesting that transcriptional regulation and protein stability are complementary, e.g. proteins that are more stable in higher temperatures have less pronounced regulation patterns or less stable transcripts. On the other hand unstable proteins might be regulated heavily at

the transcription stage, which is suggested by the lack of the reads for proteins produced based on the most differentially transcribed RNA.

The main finding of this chapter is a clear difference in the intracellular amino acid presence between 70°C and 80°C. This finding is backed by both proteomics and transcriptomics – more transcripts encoding proteins connected with amino acid metabolism are present at lower temperatures, and the same goes for the proteins themselves. This suggests that protein synthesis is enhanced at lower temperature. Lower temperatures might require more enzymes to run processes at the same speed. Other interesting insight from the metabolomics analysis is the low number of metabolites found. Intracellularly, it was only 70 compounds, while out of the cells only a handful of chemicals were identified. Apart from the components of the medium, inositol and erithritol were the only two compounds that produced GC/MS peaks. That shows two things. Firstly, *Sulfolobus* is under a strong pressure to use all the compounds it produces internally – either because of the resource scarcity or due to the fact that pores or transporters could weaken its membrane, jeopardising temperature tolerance. Secondly, these two sugars must play an important role to be an exception. Indeed, in other hyperthermophiles it was shown that they act as thermo-protective compounds.

The transcriptome shows an up-regulation of genes related to translation and ribosomal proteins, which is in accordance with the results of the proteomic analysis. The main conclusion from these results is *Sulfolobus* at 70°C grows at a similar rate as at 80°C but in order to achieve it, it must boost its metabolic rate. Most differentially regulated genes both in terms of proteomics and metabolomics are found at the lower temperatures. The causes for that may

be higher stress or compensation for slower metabolism by overproduction of relevant enzymes. The fact that multiple transporters are overrepresented at lower temperatures does suggest scavenging processes are occurring at the lower spectrum of temperature, but the alternative hypothesis could not be excluded based on our data.

This Chapter also shed some light on possible regulation of the Central Carbon Metabolism (CCM) of *Sulfolobus*. We first identified a set of 97 putative CCM genes and a set of 138 transcription factors (TFs). Importantly 4 of the genes that are putative TFs are genes coding for the CCM enzymes. All 4 of them are also catalysing reactions of the 6-carbon compounds, suggesting that possible regulation is taking place before the CCM pathways branch into the Entner-Doudoroff (ED) or Embden-Meyerhoff-Parnas (EMP) pathways. Unfortunately neither the transcriptome or the proteome showed major differences, with only one gene/protein significantly overexpressed at lower temperature for both techniques, strongly suggesting that the regulation of the CCM is on a different level of organisation.

The results might be slightly underwhelming but the methods selection has a lasting effect on the field. Using the findings from this paper and from the SysMO project, our colleagues found how the two branches of ED pathway are regulated (Kouril *et al.* 2013). Furthermore, they looked at an important metabolic switch between EMP and ED pathways (Haferkamp *et al.* 2019). This offers a glimpse into a novel regulatory network related to the uniqueness of the Central Carbon Metabolism in hyperthermophiles and shows how they can cope with the instability of their environment using a very intricate regulatory mechanisms.

A similar approach of combining multiple -omics was used to study the capacity of *S. solfataricus* to grow on L-fucose (Wolf *et al.* 2016). As mentioned before, transcriptomics was a key part in elucidating the fucose metabolism but apart from that, researchers looked at the enzyme activities, metabolomics, proteomics and applied modelling to solve this problem.

A uniform toolbox and multiomics approach in analysing *S. solfataricus* is also a tool that might allow for the use of this organism as an industrial workhorse. Its catabolic potential, resistance to contamination and the established genetic toolbox make it a very good candidate for such role (Quehenberger *et al.* 2017). One of the issues that stop it from becoming more popular is the limited understanding of metabolic regulation and this chapter and the research that has removed some of the obstacles to achieving it. *S. solfataricus* has multiple advantages over mesophilic organisms living at neutral pH: it can survive in the hostile conditions used for plant biomass hydrolysis with little modification. It also lacks catabolic suppression, and thus can utilise multiple carbon sources simultaneously. This would vastly improve performance of and therefore reduce costs of such industrial process. Looking at the metabolism of *S. solfataricus* in such mixed media and finding potential valuable chemicals one could obtain using it should be the priority in next multi-omics analyses.

Chapter 4 of this thesis focuses on the improvements in the extraction and measurements of the membrane proteins using iTRAQ method. Standard protocols used in proteomics were at the time relatively weak in terms of measuring the membrane proteome. Which is detrimental in the examination of changes related to the temperature shifts. Membrane, as the first line of defence from the environmental factors and the entry point of metabolites in the cells is potentially a good spot for differential presence of proteins. The

under-representation of membrane proteome in total proteome extracted can vastly limit the analytical power of any analysis.

The modifications of the standard iTRAQ protocols using an additional chymotrypsin and trypsin digestion and surfactant to improve the yield of membrane related proteins. The modifications turned to be successful. The modified protocols yielded 75% more protein signals without the surfactant and 93% more when surfactant was added. Moreover, the vast majority of the proteins identified exclusively by the modified protocols were either proteins identified as membrane, or hypothesised to be membrane related. For example standard protocol showed 17 proteins annotated as integral membrane proteins, while the modified protocol with digestion and surfactant added – 122 proteins, an increase of over 7-fold.

The results obtained from this experiment showed that majority of proteins were up-regulated in lower temperatures, a somewhat unexpected find. Curiously, the protein related to *Sulfolobus* insertion sequence (IS) elements were up-regulated in low temperatures, showing that low temperature conditions can stimulate IS element mobility, which is hypothesised to be one of the most potent methods of generating genetic diversity in the population of *S. solfataricus* (Martusewitsch *et al.* 2000; Blount & Grogan 2005). *S. solfataricus* has an unprecedented number of IS elements in its genome and previous results indicate their importance in the evolution of the species, however the majority of the IS elements were inactive in the experiments conducted before. This might be as most of these elements are degraded and incapable of further mobility, but alternative hypothesis is that the conditions in which the IS elements were tested were not right. The proteomic analysis from this chapter suggests that low temperatures may be a promising direction

to explore the role of IS elements in *S. solfataricus.* Other groups were also up-regulated in low temperatures, including a large number of transporters, along with few proteins related to amino acid and nucleotide metabolism, cell envelope and transcription and central metabolism.

The field of proteome analysis has steadily developed since the publishing of Chapter 4. The use of iTRAQ technique has, for example, let us understand the role of protein phosphorylation as a regulatory mechanism in *S. solfataricus* (Esser *et al.* 2012). Regulatory role of phosphorylation was not studied in this thesis, and judging by the results, this might have been an interesting aspect to focus on in relation to the temperature shift. Phosphoproteins are not only abundant in *S. solfataricus*, they also show very distinct patterns when cells are grown on glucose and tryptone. Previous analyses comparing *S. solfataricus* cells grown on these two substrates showed few differences in gene expression and protein expression (Snijders *et al.* 2006), suggesting that maybe looking at protein phosphorylation in addition to these two techniques is a better strategy to find distinct differences.

Proteomics were also essential in shedding light on another role of phosphate in *S. solfataricus* cells. Polyphosphates were shown to be a key element of *Sulfolobus* resilience to toxic conditions and a knockout mutant unable to accumulate polyphosphate was more susceptible to stress conditions, including copper stress (Soto *et al.* 2019). This research also shows the resilience of *Sulfolobus* genome – with one mechanism of copper resistance absent, other CopA mediated is upregulated in the early stages of growth and thus partially compensates for the lost function.

Membrane proteomics can be of particular importance when looking at the interactions of organisms with their parasites and the environment. In case of *Sulfolobus*, there has been a particular focus on the proteome and virus infections. Proteome analysis showed the role of membrane proteins and membrane vesicle formation for the spread of STIV viruses (Maaty *et al.* 2012) and other research shows the importance of membrane proteins identified in this chapter during the infection of SIRV2 virus (Deng *et al.* 2014).

Chapter 5 looks at the promoter regions of *S. solfataricus* genome trying to find general patterns and potential regulatory sequences. Based on the transcriptomic data, we selected groups of differentially regulated genes in search of regulatory motifs upstream of the gene. This search yielded a finding of a palindromic motif highly conserved across hyper-thermophilic Archaea. Although found fully only in front of two genes coding for hypothetical proteins in *S. solfataricus*, we decided to examine the strength of the motif, as these two genes were the most up-regulated at high temperature. This showed the potential of the motif to be temperature responsive. The fact that the motif was so well conserved between different hyperthermophilic species added strength to our prediction, as the upstream motifs in hyperthermophiles are usually very variable.

Our analysis confirmed that one of the genes containing the putative motif was indeed significantly regulated across the temperature gradient, in accordance with the microarray data from the earlier experiments. Unfortunately we were not able to elucidate the function of the proteins encoded by the genes regulated by the motif, as they show little resemblance to annotated proteins, however it is most likely a trans-membrane protein due to the seven trans-membrane domains we identified using bioinformatics tools.

The motif has potential practical applications. The use of the temperature-induced transcription can be employed in dual-phase fermentations, when upon accumulation of unfavourable by-products changing the conditions switches off the production of the enzyme responsible for the process allowing detoxification at a lower temperature.

This chapter also shows how high throughput dataset analysis using bioinformatics tools can be used to fish for specific findings that otherwise would be unlikely to be pinpointed. In this case limiting the number of genes looked at by including only the significantly regulated ones has let us find the motif, that would not be otherwise numerous sufficiently to pass significance thresholds used in upstream sequence analysis. Another possible approach for looking for archaeal regulatory sequences, that shows promise based on our results, is comparative genomics on upstream regions of orthologs from further related species. Highly conserved and/or palindromic sequences are potential good targets for further analyses.

Regulatory sequences have been a longstanding focus of the Sulfolobus research. This has partially to do with the fact that the archaeal transcription system is more similar to the eukaryotic than to the bacterial one. Large part of research on regulatory motif binding has been conducted on the Leucin-responsive regulatory Protein (Lrp) family of regulators (Napoli *et al.* 1999; Brinkman *et al.* 2002; Peeters *et al.* 2004) and this field has significantly expanded in recent years. This has included an attempt to look at in vitro binding of the motif and the regulator, showing large disproportion between being able to bind to the regulatory sequence in vitro and not being able to observe such binding in vivo (Nguyen-Duc *et al.* 2013). This is a key finding

in general studies of regulators, where *in vitro* analysis of binding has been a crucial part of determining the roles of the regulators. The authors show that *in vitro* binding does not necessarily reflect the *in vivo* situation. The study also identified multiple potential binding sites, suggesting a much subtler regulation network present. This would allow more accurate fine tuning of the transcriptome than we can currently detect, leaving a large gap in our understanding of transcription regulation to be filled. More research since showed the same discrepancy using BarR regulator in closely related *S. acidocaldarius* (Liu *et al.* 2016), indicating the need for both *in vitro* fishing for new motifs and confirmation of their functionality *in vivo*. It also shows that multiple regulators can share their regulons, again pointing to more subtle interactions in genomic regulation.

Another recent important development in transcription regulation research in *Sulfolobales* was the discovery of the role FadR regulators play an important role in regulation of fatty acid metabolism (Wang *et al.* 2019). This approach shows that understanding the function of the regulated genes greatly improves the chance of success of *in vitro* testing the regulators. Unfortunately, this is something that is missing in this thesis, as we were not able to find the function of the gene regulated by the palindromic motif we tested.

The transcription of the *Sulfolobus* genome can be also regulated, as recently shown, by the presence of small RNA (Orell *et al.* 2018). The authors were able to change the phenotype of *Sulfolobus acidocaldarius* of forming biofilm by deleting the gene of one such small non-coding RNA molecule (RrrR) on the genome. This resulted in upregulated transcription of several genes, and decreased capability of forming biofilms. The authors speculate that such

RNA particles can act as a sponge for other transcription regulators by binding them and reducing the amount of them that can bind regulatory sequences in the genome. This may suggest that strains deficient in such RNA particles may be good models for testing regulatory sequences.

In order to find additional key mechanisms of *S. solfataricus* temperature adaptation, we decided to look at the impact evolution has on it. In Chapter 6 we describe an experimental evolution experiment where we grow cells in stable conditions at temperature below and above the optimum and in a variable condition, where temperature is changed after every transfer. Sub-optimal conditions are speculated to induce mutagenesis caused by the mobility of IS elements. Our own results from Chapter 4 suggest that IS mobility might be active at low temperatures while previous data shows activity of IS elements at high temperatures (Tachdjian & Kelly 2006b). We used stable conditions that should promote IS mobility in order to look whether it will lead to disabling genes redundant in a given condition and selecting these variants. The variable temperature conditions have been shown to select for better adaptation (Ketola *et al.* 2013) and shifts from one temperature to another should purge mutations that are deleterious in either condition, while selecting for deletions favourable in the medium, removing the temperature effect. *Sulfolobus solfataricus* has been hypothesised to be a planktonic hyperthermophile, due to its lower capacity to adhere to surfaces, as opposed to anecdotally sessile *S. acidocaldarius* for example. It also has many IS elements, while *S. acidocaldarius* does not. If the IS mobility in fluctuating environment allows generalism, this would explain why there is such a big difference in IS element prevalence between the two species.

The results show that evolution in the 65°C in stable conditions results in phenotypes that have reduced fitness under higher temperatures but it is not the same for the opposite temperature. This suggests that while adaptation to colder conditions is not dependant on the genes useful in heat shock conditions, adaptation to heat shock requires functionalities necessary for optimal growth in the colder environment. This means that while adaptation to 65°C does not incur penalties on loss of genes essential in heat shock. Therefore, when evolved under stable conditions *S. solfataricus* gradually loses ability to perform optimally under more demanding heat stress. The rapid loss of those functionalities may be caused by rapid generation of genetic variation due to IS element shuffling at low temperatures.

At the same time, when applying a fluctuating selective pressure, *S. solfataricus* does not lose fitness in either of the conditions, suggesting that fluctuating selective pressure at the frequency used in the experiment purges the unfavourable mutations. Furthermore, the cells grown under fluctuating conditions show higher fitness at 84°C than the populations selected only at 84°C, which can mean that growth at lower temperatures can help generating diversity, which is later selected under the heat shock conditions.

Evolutionary approaches in *S. solfataricus* have been absent from the literature until recently, but it seems that the field if starting to bud. One example is the study of adaptation of *Sulfolobus* to extreme acidophily (McCarthy *et al.* 2015), where an experimental evolution approach was combined with genome resequencing and transcriptomics to look at the potential mechanisms responsible for the new traits. The strains were able to grow at pH values lower by over 2.0 than the optimum of the wild type strain, and several mutations that enable this adaptation were identified, leading to

new lines of research. Another example of *Sulfolobus* adaptive evolution was a study of spontaneous mutant strain PBL2025 (Qiu *et al.* 2017), which contains a large deletion of 46 genes, 6 of which are considered crucial for the CCM. This study used the fact that the rare evolutionary event already occurred, and looked on how is it possible that such a significant deletion does not incapacitate the strain, and even gives it advantage over the wild type.

Evolution experiments are potentially powerful to identify functions of hypothetical genes or pathways. It might also be a good approach to look at the traits more complex than presence or absence of one compound in the medium, like for example the coexistence of multiple species of thermophiles. Combining ecology and evolution of thermophiles and their viral parasites might reveal interesting insights. So far *Sulfolobus* has been mainly analysed out of the community context but its genome evolved in the community context. It would be interesting to see if species of hyperthermophiles living together cooperate or compete and how such interactions may change their phenotypes and performance.

Those between species interactions may also be important drivers of non-genetic inheritance of *S. solfataricus* (Payne *et al.* 2018; Johnson *et al.* 2019). This line of research showed that aside from mutational changes, *S. solfataricus* can transmit traits in a non-genetic fashion, something that has not been shown before. This has been demonstrated using the strains adapted to high acid conditions described above, showing that experimental evolution approach can yield surprising insights. Some of the acid tolerant lines achieved this feat without any mutations in the genome. Furthermore, the presumed homologous recombination of the genes involved did reduce their acid tolerance even though the sequence was exactly the same, strongly

pointing at an epigenetic-like mechanism on the chromatin level. This finding complicates the analyses of experimental evolution in Sulfolobus, but on the other hand it gives us a very accessible model organism to study epigenetic inheritance as soon as the full mechanism is elucidated.

This thesis' aim was to shed a light on the temperature adaptations of *S. solfataricus*. We used multiple approaches in order to achieve it. From high throughput methods to focus on single gene, from molecular biology to letting evolution take its course, we hopefully show a multifaceted approach towards tackling the same question from completely different angles, yet leading to inter-connected and cohesive conclusions. *Sulfolobus*, even though relatively well studied, for an extremophile, still has many grey areas and it is one of our regrets not to be able to solve some of the problems encountered during the course of this work.

Even though we found a regulatory motif and showed that it does work, the regulator for it remained elusive. The transcriptome under different conditions proved to be very stable. Here a possible error was our will to control all the factors, which resulted in a defined medium with a single carbon source. Perhaps using a more complex mixture of sugars and proteins would require a use of a broader array of genes, thus leading to more pronounced differences on the transcriptomic level and bigger range of membrane proteins utilised in different conditions that would let us explore more threads in search for *Sulfolobus* temperature regulation.

The experimental evolution approach could have been also used more frequently in this model. There are surprisingly few evolution experiments in hyperthermophiles, and *S. solfataricus* in particular, given the interesting

biology behind it and relative ease of cultivation. Experimental evolution is of particular interest here due to the unique features, including robust performance in varying conditions and extremely high IS element content of the genome. But these regrets will hopefully be addressed in the future. This thesis aimed at providing sound conclusions based on solid methodology and as such, I hope it achieved its goal and added a small brick to the magnificent temple science is.

Appendices:

References

About the author

Acknowledgements

List of publications

Overview of completed training activities

References:

1.Ahmed, H., Ettema, T.J., Tjaden, B., Geerling, A.C., van der Oost, J. & Siebers, B. (2005). The semi-phosphorylative Entner-Doudoroff pathway in hyperthermophilic archaea: a re-evaluation. *Biochem J*, 390, 529–540

2.Albers, S.-V., Birkeland, N.-K., Driessen, A.J.M., Gertig, S., Haferkamp, P., Klenk, H.-P., *et al.* (2009). SulfoSYS (Sulfolobus Systems Biology): towards a silicon cell model for the central carbohydrate metabolism of the archaeon Sulfolobus solfataricus under temperature variation

3.Albers, S.-V. & Driessen, A.J.M. (2008). Conditions for gene disruption by homologous recombination of exogenous DNA into the Sulfolobus solfataricus genome. *Archaea*, 2, 145–149

4.Albers, S.-V., Elferink, M.G.L., Charlebois, R.L., Sensen, C.W., Driessen, A.J.M. & Konings, W.N. (1999). Glucose transport in the extremely thermoacidophilic Sulfolobus solfataricus involves a high-affinity membrane-integrated binding protein. *J. Bacteriol.*, 181, 4285–4291

5.Albers, S.-V., Jonuscheit, M., Dinkelaker, S., Urich, T., Kletzin, A., Tampe, R., *et al.* (2006). Production of recombinant and tagged proteins in the hyperthermophilic archaeon Sulfolobus solfataricus. *Appl. Environ. Microbiol.*, 72, 102–111

6.Albers, S.-V., Vossenberg, J.L., Driessen, A.J. & Konings, W.N. (2001). Bioenergetics and solute uptake under extreme conditions. *Extremophiles*, 5, 285–294

7.Albers, S.-V., Van de Vossenberg, J.L., Driessen, A.J. & Konings, W.N. (2000). Adaptations of the archaeal cell membrane to heat stress. *Front Biosci*, 5, D813-20

8.Albers, S.S.-V. & Driessen, A.A.J. (2002). Signal peptides of secreted proteins of the archaeon Sulfolobus solfataricus: a genomic survey. *Arch. Microbiol.*, 177, 209–216

9.Albers, S.S.-V., Koning, S.M.S., Konings, W.N.W. & Driessen, A.J.M. (2004). Insights into ABC transport in archaea. *J. Bioenerg. Biomembr.*, 36, 5–15

10.Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25, 3389–3402

11.Anderson, I., Rodriguez, J., Susanti, D., Porat, I., Reich, C., Ulrich, L.E., *et al.* (2008). Genome sequence of Thermofilum pendens reveals an exceptional loss of biosynthetic pathways without genome reduction. *J. Bacteriol.*, 190, 2957–2965

12.Andersson, A.F., Lundgren, M., Eriksson, S., Rosenlund, M., Bernander, R. & Nilsson, P. (2006). Global analysis of mRNA stability in the archaeon Sulfolobus. *Genome Biol.*, 7

13.Andersson, A.F., Pelve, E.A., Lindeberg, S., Lundgren, M., Nilsson, P. & Bernander, R. (2010). Replication-biased genome organisation in the crenarchaeon Sulfolobus. *BMC Genomics*, 11, 454

14.Assiddiq, B.B.F., Snijders, A.A.P.L., Chong, P.K.P., Wright, P.C. & Dickman, M.J. (2008). Identification and characterization of Sulfolobus solfataricus P2 proteome using multidimensional liquid phase protein separations. *J. Proteome Res.*, 7, 2253–2261

15.Atomi, H., Matsumi, R. & Imanaka, T. (2004). Reverse gyrase is not a prerequisite for hyperthermophilic life. *J. Bacteriol.*, 186, 4829–4833

16.Auernik, K.S. & Kelly, R.M. (2008). Identification of components of electron transport chains in the extremely thermoacidophilic crenarchaeon Metallosphaera sedula through iron and sulfur compound oxidation transcriptomes. *Appl. Environ. Microbiol.*, 74, 7723–7732

17.Baker, B.J., Lesniewski, R.A. & Dick, G.J. (2012). Genome-enabled transcriptomics reveals archaeal populations that drive nitrification in a deep-sea hydrothermal plume. *ISME J.*, 6, 2269–2279

18.Baliga, N.S., Pan, M., Goo, Y.A., Eugene, C.Y., Goodlett, D.R., Dimitrov, K., *et al.* (2002). Coordinate regulation of energy transduction modules in Halobacterium sp. analyzed by a global systems approach. *Proc. Natl. Acad. Sci.*, 99, 14913–14918

19.Bardy, S.S.L., Eichler, J. & Jarrell, K.F.K. (2003). Archaeal signal peptides—a comparative survey at the genome level. *Protein Sci.*, 12, 1833–1843

20.Barry, E.R. & Bell, S.D. (2006). DNA replication in the archaea. *Microbiol. Mol. Biol. Rev.*, 70, 876–887

21.Barry, R.C., Young, M.J., Stedman, K.M. & Dratz, E.A. (2006). Proteomic mapping of the hyperthermophilic and acidophilic archaeon Sulfolobus solfataricus P2. *Electrophoresis*, 27, 2970–2983

22.Baumann, A., Lange, C. & Soppa, J. (2007). Transcriptome changes and cAMP oscillations in an archaeal cell cycle. *BMC Cell Biol.*, 8

23.Beaumont, H.J.E., Gallie, J., Kost, C., Ferguson, G.C. & Rainey, P.B. (2009). Experimental evolution of bet hedging. *Nature*, 462, 90–93

24.Bell, S.D. & Jackson, S.P. (1998). Transcription and translation in Archaea: A mosaic of eukaryal and bacterial features. *Trends Microbiol.*

25.Bell, S.D. & Jackson, S.P. (2001). Mechanism and regulation of transcription in archaea. *Curr. Opin. Microbiol.*, 4, 208–213

26.Bennett, A.F. & Lenski, R.E. (1993). Evolutionary Adaptation to Temperature II. Thermal Niches of Experimental Lines of Escherichia coli. *Evolution (N. Y).*, 47, 1

27.Bisle, B., Schmidt, A., Scheibe, B., Klein, C., Tebbe, A., Kellermann, J., *et al.* (2006). Quantitative profiling of the membrane proteome in a halophilic archaeon. *Mol. Cell. Proteomics*, 5, 1543–1558

28.Blombach, F., Salvadori, E., Fouqueau, T., Yan, J., Reimann, J., Sheppard, C., *et al.* (2015). Archaeal TFEα/β is a hybrid of TFIIE and the RNA polymerase III subcomplex hRPC62/39. *Elife*, 4

29.Blount, Z.D., Borland, C.Z. & Lenski, R.E. (2008). Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. *Proc. Natl. Acad. Sci.*, 105, 7899–7906

30.Blount, Z.D. & Grogan, D.W. (2005). New insertion sequences of Sulfolobus: Functional properties and implications for genome evolution in hyperthermophilic archaea. *Mol. Microbiol.*, 55, 312–325

31.Boonyaratanakornkit, B., Córdova, J., Park, C.B. & Clark, D.S. (2006). Pressure affects transcription profiles of Methanocaldococcus jannaschii despite the absence of barophilic growth under gas‐transfer limitation. *Environ. Microbiol.*, 8, 2031–2035

32.Boonyaratanakornkit, B.B., Simpson, A.J., Whitehead, T.A., Fraser, C.M., El-Sayed, N.M.A. & Clark, D.S. (2005). Transcriptional profiling of the hyperthermophilic methanarchaeon Methanococcus jannaschii in response to lethal heat and non-lethal cold shock. *Environ. Microbiol.*, 7, 789–797

33.Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.*, 72, 248–254

34.Brinkman, A.B., Bell, S.D., Lebbinkt, R.J., De Vos, W.M. & Der Van Oost, J. (2002). The Sulfolobus solfataricus Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability. *J. Biol. Chem.*, 277, 29537–29549

35.Brochier-Armanet, C., Boussau, B. & Gribaldo, S. (2008). Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat. Rev.*

36.Brock, T.D., Brock, K.M., Belly, R.T. & Weiss, R.L. (1972). Sulfolobus: A new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch. Mikrobiol.*, 84, 54–68

37.Brock, T.D. & Freeze, H. (1969). Thermus aquaticus gen. n. and sp. n., a nonsporulating extreme thermophile. *J. Bacteriol.*, 98, 289–297

38.Brouns, S.J.J., Walther, J., Snijders, A.P.L., van de Werken, H.J.G., Willemen, H.L.D.M., Worm, P., *et al.* (2006). Identification of the Missing Links in Prokaryotic Pentose Oxidation Pathways evidence for enzyme recruitment. *J. Biol. Chem.*, 281, 27378–27388

39.Brügger, K., Torarinsson, E., Redder, P., Chen, L. & Garrett, R.A. (2004). Shuffling of Sulfolobus genomes by autonomous and non-autonomous mobile elements. *Biochem. Soc. Trans.*, 32, 179–83

40.Buchanan, C.L., Connaris, H., Danson, M.J., Reeve, C.D. & Hough, D.W. (1999). An extremely thermostable aldolase from Sulfolobus solfataricus with specificity for non-phosphorylated substrates. *Biochem. J.*, 343, 563–570

41.Buckling, A. & Rainey, P.B. (2002). Antagonistic coevolution between a bacterium and a bacteriophage. *Proc. R. Soc. B Biol. Sci.*, 269, 931–936

42.Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., *et al.* (1996a). Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. *Science (80-.).*, 273, 1058–1073

43.Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., *et al.* (1996b). Complete genome sequence of the Methanogenic archaeon, Methanococcus jannaschii. *Science (80-.).*, 273, 1058–1073

44.Bunai, K. & Yamane, K. (2005). Effectiveness and limitation of two-dimensional gel electrophoresis in bacterial membrane protein proteomics and perspectives. *J. Chromatogr. B*, 815, 227–236

45.Campanaro, S., Williams, T.J., Burg, D.W., De Francisci, D., Treu, L., Lauro, F.M., *et al.* (2011). Temperature- dependent global gene expression in the Antarctic archaeon Methanococcoides burtonii. *Environ. Microbiol.*, 13, 2018–2038

46.Carstens, C.-P. & Waesche, A. (1999). Codon bias-adjusted BL21 derivatives for protein expression. *Homo*, 11, 11.3

47.Cava, F., Hidalgo, A. & Berenguer, J. (2009). Thermus thermophilus as biological model. *Extremophiles*, 13, 213–231

48.Cavalier-Smith, T. (1987). The Origin of Eukaryote and Archaebacterial Cells. *Ann. N. Y. Acad. Sci.*, 503, 17–54

49.Chomczynski, P. & Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.*, 162, 156–159

50.Chong, P.K. & Wright, P.C. (2005). Identification and characterization of the Sulfolobus solfataricus P2 proteome. *J. Proteome Res.*, 4, 1789–1798

51.Chong, P.K.P., Burja, A.M.A., Radianingtyas, H., Fazeli, A. & Wright, P.C. (2007a). Proteome analysis of Sulfolobus solfataricus P2 propanol metabolism. *J. Proteome Res.*, 6, 1430–1439

52.Chong, P.K.P., Burja, A.M.A., Radianingtyas, H., Fazeli, A. & Wright, P.C. (2007b). Proteome and transcriptional analysis of ethanol-grown Sulfolobus solfataricus P2 reveals ADH2, a potential alcohol dehydrogenase. *J. Proteome Res.*, 6, 3985–3994

53.Chong, P.K.P., Gan, C.S.C., Pham, T.K.T. & Wright, P.C. (2006). Isobaric tags for relative and absolute quantitation (iTRAQ) reproducibility: Implication of multiple injections. *J. Proteome Res.*, 5, 1232–1240

54.Coker, J.A., DasSarma, P., Kumar, J., Müller, J.A. & DasSarma, S. (2007). Transcriptional profiling of the model Archaeon Halobacterium sp. NRC-I: Responses to changes in salinity and temperature. *Saline Systems*, 3, 1–17

55.Conrad, R. (2009). The global methane cycle: Recent advances in understanding the microbial processes involved. *Environ. Microbiol. Rep.*

56.Cooper, C.R., Daugherty, A.J., Tachdjian, S., Blum, P.H. & Kelly, R.M. (2009). Role of vapBC toxin-antitoxin loci in the thermal stress response of Sulfolobus solfataricus. *Biochem. Soc. Trans.*, 37, 123–126

57.Cozen, A.E., Weirauch, M.T., Pollard, K.S., Bernick, D.L., Stuart, J.M. & Lowe, T.M. (2009). Transcriptional map of respiratory versatility in the hyperthermophilic crenarchaeon Pyrobaculum aerophilum. *J. Bacteriol.*, 191, 782–794

58.Culley, D.E., Kovacik Jr, W.P., Brockman, F.J. & Zhang, W. (2006). Optimization of RNA isolation from the archaebacterium< i> Methanosarcina barkeri</i> and validation for oligonucleotide microarray analysis. *J. Microbiol. Methods*, 67, 36–43

59.Dambeck, M. & Soppa, J. (2008). Characterization of a Haloferax volcanii member of the enolase superfamily: Deletion mutant construction, expression analysis, and transcriptome comparison. *Arch. Microbiol.*, 190, 341–353

60.Danan, M., Schwartz, S., Edelheit, S. & Sorek, R. (2012). Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Res.*, 40, 3131–3142

61.Danson, M.J. (1989). Central metabolism of the archaebacteria: an overview. *Can. J. Microbiol.*, 35, 58–64

62.Davis, B.M. (1897). the Vegetation of the Hot Springs of Yellowstone Park. *Science*, 6, 145–157

63.Deng, L., He, F., Bhoobalan-Chitty, Y., Martinez-Alvarez, L., Guo, Y. & Peng, X. (2014). Unveiling Cell Surface and Type IV Secretion Proteins Responsible for Archaeal Rudivirus Entry. *J. Virol.*, 88, 10264–10268

64.Dorazi, R., Götz, D., Munro, S., Bernander, R. & White, M.F. (2007). Equal rates of repair of DNA photoproducts in transcribed and non-transcribed strands in Sulfolobus solfataricus. *Mol. Microbiol.*, 63, 521–529

65.Drengstig, T., Ueda, H.R. & Ruoff, P. (2008). Predicting perfect adaptation motifs in reaction kinetic networks. *J. Phys. Chem. B*, 112, 16752–16758

66.Duggin, I.G., McCallum, S.A. & Bell, S.D. (2008). Chromosome replication dynamics in the archaeon Sulfolobus acidocaldarius. *Proc. Natl. Acad. Sci. U. S. A.*, 105, 16737–16742

67.Eichacker, L.L.A., Granvogl, B., Mirus, O., Müller, B.C.B., Miess, C. & Schleiff, E. (2004). Hiding behind hydrophobicity transmembrane segments in mass spectrometry. *J. Biol. Chem.*, 279, 50915–50922

68.Elena, S.F. & Lenski, R.E. (2003). Microbial genetics: Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat. Rev. Genet.*, 4, 457–469

69.Elferink, M.G.L., Albers, S., Konings, W.N. & Driessen, A.J.M. (2001). Sugar transport in Sulfolobus solfataricus is mediated by two families of binding protein-dependent ABC transporters. *Mol. Microbiol.*, 39, 1494–1503

70.Elias, J.J.E. & Gygi, S.P.S. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods*, 4, 207–214

71.Ellen, A.F., Albers, S.V., Huibers, W., Pitcher, A., Hobel, C.F.V., Schwarz, H., *et al.* (2009). Proteomic analysis of secreted membrane vesicles of archaeal Sulfolobus species reveals the presence of endosome sorting complex components. *Extremophiles*, 13, 67–79

72.Esser, D., Pham, T.K., Reimann, J., Albers, S. V., Siebers, B. & Wright, P.C.

(2012). Change of Carbon Source Causes Dramatic Effects in the Phospho-Proteome of the Archaeon *Sulfolobus solfataricus*. *J. Proteome Res.*, 11, 4823–4833

73.Ettema, T.J.G., de Vos, W.M. & van der Oost, J. (2005). Discovering novel biology by in silico archaeology. *Nat. Rev. Microbiol.*, 3, 859–869

74.Ettema, T.J.G.G., Ahmed, H., Geerling, A.C.M.M., van der Oost, J. & Siebers, B. (2008). The non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase (GAPN) of Sulfolobus solfataricus: a key-enzyme of the semi-phosphorylative branch of the Entner–Doudoroff pathway. *Extremophiles*, 12, 75–88

75.Facciotti, M.T., Pang, W.L., Lo, F., Whitehead, K., Koide, T., Masumura, K., *et al.* (2010). Large scale physiological readjustment during growth enables rapid, comprehensive and inexpensive systems analysis. *BMC Syst. Biol.*, 4

76.Ferry, J.G. & Lessner, D.J. (2008). Methanogenesis in marine sediments. *Ann. N. Y. Acad. Sci.*

77.Fischer, F., Wolters, D., Rögner, M. & Poetsch, A. (2006). Toward the Complete Membrane Proteome High Coverage of Integral Membrane Proteins Through Transmembrane Peptide Detection. *Mol. Cell. Proteomics*, 5, 444–453

78.Forterre, P., Confalonier, F., Charbonnier, F. & Duguet, M. (1995). Speculations on the origin of life and thermophily: Review of available information on reverse gyrase suggests that hyperthermophilic procaryotes are not so primitive. *Orig. Life Evol. Biosph.*, 25, 235–249

79.Fröls, S., Gordon, P.M.K., Panlilio, M.A., Duggin, I.G., Bell, S.D., Sensen, C.W., *et al.* (2007). Response of the Hyperthermophilic Archaeon Sulfolobus solfataricus to UV Damage. *J. Bacteriol.*, 189, 8708–8718

80.Fusco, S., Liguori, R., Limauro, D., Bartolucci, S., She, Q. & Contursi, P. (2015). Transcriptome analysis of Sulfolobus solfataricus infected with two related fuselloviruses reveals novel insights into the regulation of CRISPR-Cas system. *Biochimie*, 118, 322–332

81.Gan, C., Chong, P. & Pham, T. (2007). Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ). *J. proteome*

82.Gardy, J.L.J., Laird, M.M.R., Chen, F., Rey, S., Walsh, C.J., Ester, M., *et al.* (2005). PSORTb v. 2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics*, 21, 617–623

83.Garrido, P., González-Toril, E., García-Moyano, A., Moreno-Paz, M., Amils, R. & Parro, V. (2008). An oligonucleotide prokaryotic acidophile microarray: Its validation and its use to monitor seasonal variations in extreme acidic environments with total environmental RNA. *Environ. Microbiol.*, 10, 836–850

84.Gindner, A., Hausner, W. & Thomm, M. (2014). The TrmB family: a versatile group of transcriptional regulators in Archaea. *Extremophiles*

85.Gliozzi, A., Relini, A. & Chong, P.L.G. (2002). Structure and permeability properties of biomimetic membranes of bolaform archaeal tetraether lipids. *J. Memb. Sci.*, 206, 131–147

86.Goberna, M., Gadermaier, M., García, C., Wett, B. & Insam, H. (2010). Adaptation of methanogenic communities to the cofermentation of cattle excreta and olive mill wastes at 37 C and 55 C. *Appl. Environ. Microbiol.*, 76, 6564–6571

87.Gonçalves, L.G., Lamosa, P., Huber, R. & Santos, H. (2008). Di-myo-inositol phosphate and novel UDP-sugars accumulate in the extreme hyperthermophile Pyrolobus fumarii. *Extremophiles*, 12, 383–389

88.Goo, Y.Y.A., Eugene, C.Y., Baliga, N.S.N., Tao, W.A.W., Pan, M., Aebersold, R., *et al.* (2003). Proteomic analysis of an extreme halophilic archaeon, Halobacterium sp. NRC-1. *Mol. Cell. Proteomics*, 2, 506–524

89.Graham, R.R.L.J., Graham, C. & McMullan, G. (2007). Microbial proteomics: a mass spectrometry primer for biologists. *Microb. Cell Fact.*, 6, 26

90.Greve, B.O., Jensen, S., Brügger, K., Zillig, W. & Garrett, R.A. (2004). Genomic comparison of archaeal conjugative plasmids from Sulfolobus. *Archaea*, 1, 231–239

91.Grogan, D.W. (1989). Phenotypic characterization of the archaebacterial genus Sulfolobus: comparison of five wild-type strains. *J. Bacteriol.*, 171, 6710–6719

92.Grogan, D.W., Ozarzak, M.A. & Bernander, R. (2008). Variation in gene content among geographically diverse Sulfolobus isolates. *Environ. Microbiol.*, 10, 137–146

93.Haferkamp, P., Tjaden, B., Shen, L., Bräsen, C., Kouril, T. & Siebers, B. (2019). The Carbon Switch at the Level of Pyruvate and Phosphoenolpyruvate in Sulfolobus solfataricus P2. *Front. Microbiol.*, 10, 757

94.Hall, A.R., Scanlan, P.D. & Buckling, A. (2011). Bacteria - Phage Coevolution and the Emergence of Generalist Pathogens. *Am. Nat.*, 177, 44–53

95.Hamilton-Brehm, S.D., Schut, G.J. & Adams, M.W.W. (2005). Metabolic and evolutionary relationships among Pyrococcus species: Genetic exchange within a hydrothermal vent environment. *J. Bacteriol.*, 187, 7492–7499

96.Hanahan, D. (1983). Studies on transformation of< i> Escherichia coli</i> with plasmids. *J. Mol. Biol.*, 166, 557–580

97.Hendrickson, E.L., Haydock, A.K., Moore, B.C., Whitman, W.B. & Leigh, J.A. (2007). Functionally distinct genes regulated by hydrogen limitation and growth rate in methanogenic Archaea. *Proc. Natl. Acad. Sci. U. S. A.*, 104, 8930–8934

98.Hovey, R., Lentes, S., Ehrenreich, A., Salmon, K., Saba, K., Gottschalk, G., *et al.* (2005). DNA microarray analysis of Methanosarcina mazei Gö1 reveals adaptation to different methanogenic substrates. *Mol. Genet. Genomics*, 273, 225–239

99.Huber, H., Hohn, M.J., Stetter, K.O. & Rachel, R. (2003). The phylum Nanoarchaeota: Present knowledge and future perspectives of a unique form of life. *Res. Microbiol.*

100.Jäger, D., Sharma, C.M., Thomsen, J., Ehlers, C., Vogel, J. & Schmitz, R.A. (2009). Deep sequencing analysis of the Methanosarcina mazei Gö 1 transcriptome in response to nitrogen availability. *Proc. Natl. Acad. Sci. U. S. A.*, 106, 21878–21882

101.Jeffery, C.J., Bahnson, B.J., Chien, W., Ringe, D. & Petsko, G.A. (2000). Crystal structure of rabbit phosphoglucose isomerase, a glycolytic enzyme that moonlights as neuroleukin, autocrine motility factor, and differentiation mediator. *Biochemistry*, 39, 955–964

102.Jiang, C.-Y., Liu, L.-J., Guo, X., You, X.-Y., Liu, S.-J. & Poetsch, A. (2014). Resolution of carbon metabolism and sulfur-oxidation pathways of Metallosphaera cuprina Ar-4 via comparative proteomics. *J. Proteomics*, 109, 276–289

103.Johnsen, U., Dambeck, M., Zaiss, H., Fuhrer, T., Soppa, J., Sauer, U., *et al.* (2009). D-xylose degradation pathway in the halophilic archaeon Haloferax volcanii. *J. Biol. Chem.*, 284, 27290–27303

104.Johnson, T., Payne, S., Grove, R., McCarthy, S., Oeltjen, E., Mach, C., *et al.* (2019). Methylation deficiency of chromatin proteins is a non-mutational and epigenetic-like trait in evolved lines of the archaeon Sulfolobus solfataricus. *J. Biol. Chem.*, 294, 7821–7832

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

106.Kagawa, H.K., Yaoi, T., Brocchieri, L., McMillan, R.A., Alton, T. & Trent, J.D. (2003). The composition, structure and stability of a group II chaperonin are temperature regulated in a hyperthermophilic archaeon. *Mol. Microbiol.*, 48, 143–

156

107.Kanai, T., Akerboom, J., Takedomi, S., Van De Werken, H.J.G., Blombach, F., Van Der Oost, J., *et al.* (2007). A global transcriptional regulator in Thermococcus kodakaraensis controls the expression levels of both glycolytic and gluconeogenic enzyme-encoding genes. *J. Biol. Chem.*, 282, 33659–33670

108.Kanai, T., Takedomi, S., Fujiwara, S., Atomi, H. & Imanaka, T. (2010). Identification of the Phr-dependent heat shock regulon in the hyperthermophilic archaeon, Thermococcus kodakaraensis. *J. Biochem.*, 147, 361–370

109.Kates, M. (1977). The phytanyl ether-linked polar lipids and isoprenoid neutral lipids of extremely halophilic bacteria. *Prog. Chem. Fats Other Lipids*

110.Kawecki, T.J., Lenski, R.E., Ebert, D., Hollis, B., Olivieri, I. & Whitlock, M.C. (2012). Experimental evolution. *Trends Ecol. Evol.*

111.Keese, A.M., Schut, G.J., Ouhammouch, M., Adams, M.W.W. & Thomm, M. (2010). Genome-wide identification of targets for the archaeal heat shock regulator Phr by cell-free transcription of genomic DNA. *J. Bacteriol.*, 192, 1292–1298

112.Kelman, Z. & White, M.F. (2005). Archaeal DNA replication and repair. *Curr. Opin. Microbiol.*, 8, 669–676

113.Kempner, E.S. (1963). Upper temperature limit of life. *Sci. (New York, NY)*, 142, 1318–1319

114.Ketola, T., Mikonranta, L., Zhang, J., Saarinen, K., Örmälä, A.M., Friman, V.P., *et al.* (2013). Fluctuating Temperature Leads To Evolution Of Thermal Generalism And Preadaptation To Novel Environments. *Evolution (N. Y).*, 67, 2936–2944

115.Khoa Pham, T. & Wright, P.C. (2007). Proteomic analysis of Saccharomyces cerevisiae

116.Kim, S. & LEE, S. (2005). Identification and characterization of Sulfolobus solfataricus D-gluconate dehydratase: a key enzyme in the non-phosphorylated Entner-Doudoroff pathway. *Biochem. J*, 387, 271–280

117.Kim, S. & Lee, S.B. (2006). Characterization of Sulfolobus solfataricus 2-keto-3-deoxy-D-gluconate kinase in the modified Entner-Doudoroff pathway. *Biosci. Biotechnol. Biochem.*, 70, 1308–1316

118.Klein, C., Garcia‐Rizo, C., Bisle, B., Scheffer, B., Zischka, H., Pfeiffer, F., *et al.* (2005). The membrane proteome of Halobacterium salinarum. *Proteomics*, 5, 180–197

119.Klenk, H.P., Clayton, R.A., Tomb, J.F., White, O., Nelson, K.E., Ketchum, K.A., *et al.* (1997). The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus. *Nature*, 390, 364–370

120.Koerdt, A., Orell, A., Pham, T.K., Mukherjee, J., Wlodkowski, A., Karunakaran, E., *et al.* (2011). Macromolecular Fingerprinting of *Sulfolobus* Species in Biofilm: A Transcriptomic and Proteomic Approach Combined with Spectroscopic Analysis. *J. Proteome Res.*, 10, 4105–4119

121.Kouril, T., Wieloch, P., Reimann, J., Wagner, M., Zaparty, M., Albers, S.V., *et al.* (2013). Unraveling the function of the two Entner-Doudoroff branches in the thermoacidophilic Crenarchaeon Sulfolobus solfataricus P2. *FEBS J.*, 280, 1126–1138

122.Krogh, A., Larsson, B.B., Von Heijne, G., Sonnhammer, E.L.L. & Heijne, G. Von. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.*, 305, 567–580

123.Kwapisz, M., Beckouët, F. & Thuriaux, P. (2008). Early evolution of eukaryotic DNA-dependent RNA polymerases. *Trends Genet.*

124.Kyte, J. & Doolittle, R.F.R. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, 157, 105–132

125.Lamble, H.J., Heyer, N.I., Bull, S.D., Hough, D.W. & Danson, M.J. (2003). Metabolic pathway promiscuity in the archaeon Sulfolobus solfataricus revealed by studies on glucose dehydrogenase and 2-keto-3-deoxygluconate aldolase. *J. Biol. Chem.*, 278, 34066–34072

126.Lamble, H.J., Milburn, C.C., Taylor, G.L., Hough, D.W. & Danson, M.J. (2004). Gluconate dehydratase from the promiscuous Entner-Doudoroff pathway in Sulfolobus solfataricus. *FEBS Lett.*, 576, 133–136

127.Lamble, H.J., Theodossis, A., Milburn, C.C., Taylor, G.L., Bull, S.D., Hough, D.W., *et al.* (2005). Promiscuity in the part-phosphorylative Entner–Doudoroff pathway of the archaeon< i> Sulfolobus solfataricus</i>. *FEBS Lett.*, 579, 6865–6869

128.Lange, C., Zaigler, A., Hammelmann, M., Twellmeyer, J., Raddatz, G., Schuster, S.C., *et al.* (2007). Genome-wide analysis of growth phase-dependent translational and transcriptional regulation in halophilic archaea. *BMC Genomics*, 8, 415

129.Lee, H.-S., Shockley, K.R., Schut, G.J., Conners, S.B., Montero, C.I., Johnson, M.R., *et al.* (2006). Transcriptional and biochemical analysis of starch metabolism in the hyperthermophilic archaeon Pyrococcus furiosus. *J. Bacteriol.*, 188, 2115–2125

130.Lee, S.H., Kim, M.S., Lee, J.H., Kim, T.W., Bae, S.S., Lee, S.M., *et al.* (2016). Adaptive engineering of a hyperthermophilic archaeon on CO and discovering the underlying mechanism by multi-omics analysis. *Sci. Rep.*, 6, 22896

131.Li, L., Li, Q., Rohlin, L., Kim, U., Salmon, K., Rejtar, T., *et al.* (2007). Quantitative Proteomic and Microarray Analysis of the Archaeon M ethanosarcina acetivorans Grown with Acetate versus Methanol. *J. Proteome Res.*, 6, 759–771

132.Li, M., Baker, B.J., Anantharaman, K., Jain, S., Breier, J.A. & Dick, G.J. (2015). Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. *Nat. Commun.*, 6, 8933

133.Lipps, G. (2006). Plasmids and viruses of the thermoacidophilic crenarchaeote Sulfolobus. *Extremophiles*

134.Liu, B., Zhang, Y. & Zhang, W. (2014). RNA-seq-based analysis of cold shock response in Thermoanaerobacter tengcongensis, a bacterium harboring a single cold shock protein encoding gene. *PLoS One*, 9, e93289

135.Liu, H., Wang, K., Lindås, A.C. & Peeters, E. (2016). The genome-scale DNA-binding profile of BarR, a β-alanine responsive transcription factor in the archaeon Sulfolobus acidocaldarius. *BMC Genomics*, 17, 569

136.Lundgren, M., Andersson, A., Chen, L., Nilsson, P. & Bernander, R. (2004). Three replication origins in Sulfolobus species: Synchronous initiation of chromosome replication and asynchronous termination. *Proc. Natl. Acad. Sci. U. S. A.*, 101, 7046–7051

137.Lundgren, M. & Bernander, R. (2007). Genome-wide transcription map of an archaeal cell cycle. *Proc. Natl. Acad. Sci.*, 104, 2939–2944

138.Maaty, W.S., Selvig, K., Ryder, S., Tarlykov, P., Hilmer, J.K., Heinemann, J., *et al.* (2012). Proteomic analysis of sulfolobus solfataricus during sulfolobus turreted icosahedral virus infection. *J. Proteome Res.*, 11, 1420–1432

139.Makarova, K.S. & Koonin, E. V. (2003). Comparative genomics of Archaea: how much have we learned in six years, and what's next. *Genome Biol*, 4, 115

140.Marguerat, S. & Bähler, J. (2010). RNA-seq: From technology to biology. *Cell. Mol. Life Sci.*

141.Markowitz, V.M., Szeto, E., Palaniappan, K., Grechkin, Y., Chu, K., Chen, I.-M.A., *et al.* (2008). The integrated microbial genomes (IMG) system in 2007: data content and analysis tool extensions. *Nucleic Acids Res.*, 36, D528–D533

142.Martusewitsch, E., Sensen, C.W. & Schleper, C. (2000). High spontaneous mutation rate in the hyperthermophilic archaeon Sulfolobus solfataricus is mediated by transposable elements. *J. Bacteriol.*, 182, 2574–2581

143.Masuda, T., Tomita, M. & Ishihama, Y. (2008). Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *J. Proteome Res.*, 7, 731–740

144.Matsunaga, F., Glatigny, A., Mucchielli-Giorgi, M.E., Agier, N., Delacroix, H., Marisa, L., *et al.* (2007). Genomewide and biochemical analyses of DNA-binding activity of Cdc6/Orc1 and Mcm proteins in Pyrococcus sp. *Nucleic Acids Res.*, 35, 3214–3222

145.McCarthy, S., Ai, C., Wheaton, G., Tevatia, R., Eckrich, V., Kelly, R., *et al.* (2014). Role of an archaeal PitA transporter in the copper and arsenic resistance of Metallosphaera sedula, an extreme thermoacidophile. *J. Bacteriol.*, 196, 3562–3570

146.McCarthy, S., Johnson, T., Pavlik, B.J., Payne, S., Schackwitz, W., Martin, J., *et al.* (2015). Expanding the limits of thermoacidophily in the archaeon Sulfolobus solfataricus by adaptive evolution. *Appl. Environ. Microbiol.*, 82, 857–867

147.McCarthy, S., Johnson, T., Pavlik, B.J., Payne, S., Schackwitz, W., Martin, J., *et al.* (2016). Expanding the Limits of Thermoacidophily in the Archaeon Sulfolobus solfataricus by Adaptive Evolution. *Appl. Environ. Microbiol.*, 82, 857–867

148.McCready, S., Müller, J.A., Boubriak, I., Berquist, B.R., Ng, W.L. & DasSarma, S. (2005). UV irradiation induces homologous recombination genes in the model archaeon, Halobacterium sp. NRC-1. *Saline Systems*, 1

149.Medina-Rivera, A., Defrance, M., Sand, O., Herrmann, C., Castro-Mondragon, J.A., Delerce, J., *et al.* (2015). RSAT 2015: Regulatory sequence analysis tools. *Nucleic Acids Res.*, 43, W50–W56

150.Michoud, G. & Jebbar, M. (2016). High hydrostatic pressure adaptive strategies in an obligate piezophile Pyrococcus yayanosii. *Sci. Rep.*, 6, 27289

151.Müller, J.A. & DasSarma, S. (2005). Genomic analysis of anaerobic respiration in the archaeon Halobacterium sp. strain NRC-1: dimethyl sulfoxide and trimethylamine N-oxide as terminal electron acceptors. *J. Bacteriol.*, 187, 1659–1667

152.Napoli, A., Van Der Oost, J., Sensen, C.W., Charlebois, R.L., Rossi, M. & Ciaramella, M. (1999). An Lrp-like protein of the hyperthermophilic archaeon Sulfolobus solfataricus which binds to its own promoter. *J. Bacteriol.*, 181, 1474–1480

153.Nguyen-Duc, T., van Oeffelen, L., Song, N., Hassanzadeh-Ghassabeh, G.,

Muyldermans, S., Charlier, D., *et al.* (2013). The genome-wide binding profile of the Sulfolobus solfataricus transcription factor Ss-LrpB shows binding events beyond direct transcription regulation. *BMC Genomics*, 14, 828

154.Ni, X.Y., Drengstig, T. & Ruoff, P. (2009). The control of the controller: molecular mechanisms for robust perfect adaptation and temperature compensation. *Biophys. J.*, 97, 1244–1253

155.Niinemets, Ü. & Valladares, F. (2008). Environmental Tolerance. In: *Encyclopedia of Ecology*. University of Chicago Press, pp. 1370–1376

156.Olivier, B.G. & Snoep, J.L. (2004). Web-based kinetic modelling using JWS Online. *Bioinformatics*, 20, 2143–2144

157.Van der Oost, J. & Siebers, B. (2007). The glycolytic pathways of Archaea: evolution by tinkering. *Archaea Evol. Physiol. Mol. Biol.*, 22, 247–260

158.Orell, A., Tripp, V., Aliaga-Tobar, V., Albers, S.V., Maracaja-Coutinho, V. & Randau, L. (2018). A regulatory RNA is involved in RNA duplex formation and biofilm regulation in sulfolobus acidocaldarius. *Nucleic Acids Res.*, 46, 4794–4806

159.Ortmann, A.C., Brumfield, S.K., Walther, J., McInnerney, K., Brouns, S.J.J., Van De Werken, H.J.G., *et al.* (2008). Transcriptome analysis of infection of the archaeon Sulfolobus solfataricus with Sulfolobus turreted icosahedral virus. *J. Virol.*, 82, 4874–4883

160.Palmieri, G., Cannio, R., Fiume, I., Rossi, M. & Pocsfalvi, G. (2009). Outside the unusual cell wall of the hyperthermophilic archaeon Aeropyrum pernix K1. *Mol. Cell. Proteomics*, 8, 2570–2581

161.Park, C.B. & Lee, S.B. (1997). Constant-volume fed-batch operation for high density cultivation of hyperthermophilic aerobes. *Biotechnol Tech.*, 11, 277–281

162.Parnell, J.J., Rompato, G., Latta IV, L.C., Pfrender, M.E., Van Nostrand, J.D., He, Z., *et al.* (2010). Functional biogeography as evidence of gene transfer in hypersaline microbial communities. *PLoS One*, 5, e12919

163.Payne, S., McCarthy, S., Johnson, T., North, E. & Blum, P. (2018). Nonmutational mechanism of inheritance in the Archaeon Sulfolobus solfataricus. *Proc. Natl. Acad. Sci.*, 115, 12271–12276

164.Peeters, E. & Charlier, D. (2010). The Lrp family of transcription regulators in archaea. *Archaea*

165.Peeters, E., Thia-Toong, T.L., Gigot, D., Maes, D. & Charlier, D. (2004). Ss-LrpB, a novel Lrp-like regulator of Sulfolobus solfataricus P2, binds cooperatively to

three conserved targets in its own control region. *Mol. Microbiol.*, 54, 321–336

166.Peng, X., Brügger, K., Shen, B., Chen, L., She, Q. & Garrett, R.A. (2003). Genus-specific protein binding to the large clusters of DNA repeats (short regularly spaced repeats) present in Sulfolobus genomes. *J. Bacteriol.*, 185, 2410–2417

167.Pérez-Ortín, J.E., Alepuz, P., Chávez, S. & Choder, M. (2013). Eukaryotic mRNA decay: Methodologies, pathways, and links to other stages of gene expression. *J. Mol. Biol.*, 425, 3750–3775

168.Petitjean, C., Deschamps, P., López-García, P., Moreira, D. & Brochier-Armanet, C. (2015). Extending the conserved phylogenetic core of archaea disentangles the evolution of the third domain of life. *Mol. Biol. Evol.*, 32, 1242–1254

169.Pflüger, K., Ehrenreich, A., Salmon, K., Gunsalus, R.P., Deppenmeier, U., Gottschalk, G., *et al.* (2007). Identification of genes involved in salt adaptation in the archaeon Methanosarcina mazei Gö1 using genome-wide gene expression profiling. *FEMS Microbiol. Lett.*, 277, 79–89

170.Pham, T.K., Sierocinski, P., van der Oost, J. & Wright, P.C. (2010). Quantitative proteomic analysis of Sulfolobus solfataricus membrane proteins. *J. Proteome Res.*, 9, 1165–1172

171.Piles, M., Fernandez-Lozano, C., Velasco-Galilea, M., González-Rodríguez, O., Sánchez, J.P., Torrallardona, D., *et al.* (2019). Machine learning applied to transcriptomic data to identify genes associated with feed efficiency in pigs. *Genet. Sel. Evol.*, 51, 10

172.Prangishvili, D., Stedman, K. & Zillig, W. (2001). Viruses of the extremely thermophilic archaeon Sulfolobus. *Trends Microbiol.*

173.Qiu, W., Pham, T.K., Zou, X., Ow, S.Y. & Wright, P.C. (2017). Natural Mutagenesis-Enabled Global Proteomic Study of Metabolic and Carbon Source Implications in Mutant Thermoacidophillic Archaeon *Sulfolobus solfataricus* PBL2025. *J. Proteome Res.*, 16, 2370–2383

174.Quackenbush, J. (2002). Microarray data normalization and transformation. *Nat. Genet.*, 32, 496–501

175.Quehenberger, J., Shen, L., Albers, S.V., Siebers, B. & Spadiut, O. (2017). Sulfolobus - A potential key organism in future biotechnology. *Front. Microbiol.*

176.Qureshi, S.A., Bell, S.D. & Jackson, S.P. (1997). Factor requirements for transcription in the Archaeon Sulfolobus shibatae. *EMBO J.*, 16, 2927–2936

177.Raghoebarsing, A., Pol, A., van de Pas-Schoonen, K., Smolders, A., Ettwig, K.,

Rijpstra, W., *et al.* (2006). A microbial consortium couples anaerobic methane oxidation to denitrification RID B-9395-2011 RID F-5114-2011 RID B-8834-2011. *Nature*, 440, 918–921

178.Ratcliff, W.C., Denison, R.F., Borrello, M. & Travisano, M. (2012). Experimental evolution of multicellularity. *Proc. Natl. Acad. Sci.*, 109, 1595–1600

179.Redder, P., She, Q. & Garrett, R.A. (2001). Non-autonomous mobile elements in the crenarchaeon< i> Sulfolobus solfataricus</i>. *J. Mol. Biol.*, 306, 1–6

180.Reichlen, M.J., Murakami, K.S. & Ferry, J.G. (2010). Functional analysis of the three TATA binding protein homologs in Methanosarcina acetivorans. *J. Bacteriol.*, 192, 1511–1517

181.Reiter, W.D., Hudepohl, U. & Zillig, W. (1990). Mutational analysis of an archaebacterial promoter: essential role of a TATA box for transcription efficiency and start-site selection in vitro. *Proc. Natl. Acad. Sci.*, 87, 9509–9513

182.Ren, Y., She, Q. & Huang, L. (2013). Transcriptomic analysis of the SSV2 infection of Sulfolobus solfataricus with and without the integrative plasmid pSSVi. *Virology*, 441, 126–134

183.Rich, J.J., Dale, O.R., Song, B. & Ward, B.B. (2008). Anaerobic ammonium oxidation (anammox) in Chesapeake Bay sediments. *Microb. Ecol.*, 55, 311–320

184.Riley, M.S., Cooper, V.S., Lenski, R.E., Forney, L.J. & Marsh, T.L. (2001). Rapid phenotypic change and diversification of a soil bacterium during 1000 generations of experimental evolution. *Microbiology*, 147, 995–1006

185.Ring, G. & Eichler, J. (2004). In the archaea Haloferax volcanii, membrane protein biogenesis and protein synthesis rates are affected by decreased ribosomal binding to the translocon. *J. Biol. Chem.*, 279, 53160–53166

186.Rivera, M.C. & Lake, J.A. (2004). The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature*, 431, 152–155

187.Robinson, N.P., Dionne, I., Lundgren, M., Marsh, V.L., Bernander, R. & Bell, S.D. (2004a). Identification of Two Origins of Replication in the Single Chromosome of the Archaeon Sulfolobus solfataricus. *Cell*, 116, 25–38

188.Robinson, N.P., Dionne, I., Lundgren, M., Marsh, V.L., Bernander, R. & Bell, S.D. (2004b). Identification of Two Origins of Replication in the Single Chromosome of the Archaeon Sulfolobus solfataricus. *Cell*, 116, 25–38

189.Ruggero, D., Creti, R. & Londei, P. (1993). In vitro translation of archaeal natural mRNAs at high temperature. *FEMS Microbiol. Lett.*, 107, 89–94

190.Santangelo, T.J., Čuboňová, L., Matsumi, R., Atomi, H., Imanaka, T. & Reeve, J.N. (2008). Polarity in archaeal operon transcription in Thermococcus kodakaraensis. *J. Bacteriol.*, 190, 2244–2248

191.Santoni, V., Molloy, M. & Rabilloud, T. (2000). Membrane proteins and proteomics: un amour impossible? *Electrophoresis*, 21, 1054–1070

192.Schiraldi, C., Marulli, F., Lernia, I. Di, Martino, A. & De Rosa, M. (1999). A microfiltration bioreactor to achieve high cell density in Sulfolobus solfataricus fermentation. *Extremophiles*, 3, 199–204

193.Schleper, C., Kubo, K. & Zillig, W. (1992). The particle SSV1 from the extremely thermophilic archaeon Sulfolobus is a virus: Demonstration of infectivity and of transfection with viral DNA. *Proc. Natl. Acad. Sci. U. S. A.*, 89, 7645–7649

194.Schleper, C., Röder, R., Singer, T. & Zillig, W. (1994). An insertion element of the extremely thermophilic archaeon Sulfolobus solfataricus transposes into the endogenous β-galactosidase gene. *Mol. Gen. Genet. MGG*, 243, 91–96

195.Schmid, A.K., Reiss, D.J., Pan, M., Koide, T. & Baliga, N.S. (2009). A single transcription factor regulates evolutionarily diverse but functionally linked metabolic pathways in response to nutrient availability. *Mol. Syst. Biol.*, 5

196.Schmidt, I., Hermelink, C., Van de Pas-Schoonen, K., Strous, M., Op den Camp, H.J., Kuenen, J.G., *et al.* (2002). Anaerobic ammonia oxidation in the presence of nitrogen oxides (NOx) by two different lithotrophs. *Appl. Environ. Microbiol.*, 68, 5351–5357

197.Schut, G.J., Brehm, S.D., Datta, S. & Adams, M.W.W. (2003). Whole-genome DNA microarray analysis of a hyperthermophile and an archaeon: Pyrococcus furiosus grown on carbohydrates or peptides. *J. Bacteriol.*, 185, 3935–3947

198.Schut, G.J., Bridger, S.L. & Adams, M.W.W. (2007). Insights into the metabolism of elemental sulfur by the hyperthermophilic archaeon Pyrococcus furiosus: characterization of a coenzyme A-dependent NAD (P) H sulfur oxidoreductase. *J. Bacteriol.*, 189, 4431–4441

199.Schut, G.J., Zhou, J. & Adams, M.W.W. (2001). DNA microarray analysis of the hyperthermophilic archaeon Pyrococcus furiosus: evidence for a new type of sulfur-reducing enzyme complex. *J. Bacteriol.*, 183, 7027–7036

200.Schwaiger, R., Schwarz, C., Furtwängler, K., Tarasov, V., Wende, A. & Oesterhelt, D. (2010). Transcriptional control by two leucine-responsive regulatory proteins in Halobacterium salinarum R1. *BMC Mol. Biol.*, 11

201.Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., *et al.* (2011). Global quantification of mammalian gene expression control. *Nature*, 473, 337–342

202.She, Q., Singh, R.K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M.J., *et al.* (2001a). The complete genome of the crenarchaeon Sulfolobus solfataricus P2. *Proc. Natl. Acad. Sci.*, 98, 7835–7840

203.She, Q., Singh, R.K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M.J., *et al.* (2001b). The complete genome of the crenarchaeon Sulfolobus solfataricus P2. *Proc. Natl. Acad. Sci.*, 98, 7835–7840

204.Sherawat, M., Tolan, D.R. & Allen, K.N. (2008). Structure of a rabbit muscle fructose-1, 6-bisphosphate aldolase A dimer variant. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 64, 543–550

205.Shih, C.J. & Lai, M.C. (2010). Differentially expressed genes after hyper- and hypo-salt stress in the halophilic archaeon Methanohalophilus portucalensis. *Can. J. Microbiol.*, 56, 295–307

206.Shilov, I.V.I., Seymour, S.L.S., Patel, A.A.A., Loboda, A., Tang, W.H., Keating, S.P., *et al.* (2007). The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol. Cell. Proteomics*, 6, 1638–1655

207.Shockley, K.R., Ward, D.E., Chhabra, S.R., Conners, S.B., Montero, C.I. & Kelly, R.M. (2003). Heat shock response by the hyperthermophilic archaeon Pyrococcus furiosus. *Appl. Environ. Microbiol.*, 69, 2365–2371

208.Simon, G., Walther, J., Zabeti, N., Combet-Blanc, Y., Auria, R., Van Der Oost, J., *et al.* (2009). Effect of O2 concentrations on Sulfolobus solfataricus P2. *FEMS Microbiol. Lett.*, 299, 255–260

209.Skorko, R., Osipiuk, J. & Stetter, K.O. (1989). Glycogen-bound polyphosphate kinase from the archaebacterium Sulfolobus acidocaldarius. *J. Bacteriol.*, 171, 5162–5164

210.Snijders, A.A.P.L., Vos, M. de, de Vos, M.G.J. & Wright, P.C. (2005). Novel approach for peptide quantitation and sequencing based on 15N and 13C metabolic labeling. *J. Proteome Res.*, 4, 578–585

211.Snijders, A.P.L., Walther, J., Peter, S., Kinnman, I., De Vos, M.G.J., Van de Werken, H.J.G., *et al.* (2006). Reconstruction of central carbon metabolism in Sulfolobus solfataricus using a two-dimensional gel electrophoresis map, stable isotope labelling and DNA microarray analysis. *Proteomics*, 6, 1518–1529

212.Snoep, J.L. & Westerhoff, H. V. (2005). From isolation to integration, a systems biology approach for building the Silicon Cell. In: *Systems Biology*. Springer, pp. 13–30

213.Soto, D.F., Recalde, A., Orell, A., Albers, S.V., Paradela, A., Navarro, C.A., *et al.* (2019). Global effect of the lack of inorganic polyphosphate in the extremophilic archaeon Sulfolobus solfataricus: A proteomic approach. *J. Proteomics*, 191, 143–152

214.Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., *et al.* (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*, 521, 173–179

215.Stark, H., Wolf, J., Albersmeier, A., Pham, T.K., Hofmann, J.D., Siebers, B., *et al.* (2017). Oxidative Stickland reactions in an obligate aerobic organism – amino acid catabolism in the Crenarchaeon Sulfolobus solfataricus. *FEBS J.*, 284, 2078–2095

216.Stetter, K.O. (2006). *Hyperthermophiles in the history of life*

217.Stevens, T.J. & Arkin, I.T. (2000). Do more complex organisms have a greater proportion of membrane proteins in their genomes? *Proteins*, 39, 417–20

218.Stralis-Pavese, N., Sessitsch, A., Weilharter, A., Reichenauer, T., Riesing, J., Csontos, J., *et al.* (2004). Optimization of diagnostic microarray for application in analysing landfill methanotroph communities under different plant covers. *Environ. Microbiol.*, 6, 347–363

219.Strand, K.R., Sun, C., Li, T., Jenney Jr, F.E., Schut, G.J. & Adams, M.W.W. (2010). Oxidative stress protection and the repair response to hydrogen peroxide in the hyperthermophilic archaeon Pyrococcus furiosus and in related species. *Arch. Microbiol.*, 192, 447–459

220.Strelkov, S., von Elstermann, M. & Schomburg, D. (2004). Comprehensive analysis of metabolites in Corynebacterium glutamicum by gas chromatography/mass spectrometry. *Biol. Chem.*, 385, 853–861

221.Studier, F.W. & Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.*, 189, 113–130

222.Swarts, D.D.C., Koehorst, J.J., Westra, E.R.E., Schaap, P.J., van der Oost, J., Peters, L., *et al.* (2015). Effects of Argonaute on Gene Expression in Thermus thermophilus. *PLoS One*, 10, e0124880

223.Szilágyi, A. & Závodszky, P. (2000). Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: Results of a

comprehensive survey. *Structure*, 8, 493–504

224.Tachdjian, S. & Kelly, R.M. (2006a). Dynamic metabolic adjustments and genome plasticity are implicated in the heat shock response of the extremely thermoacidophilic archaeon Sulfolobus solfataricus. *J. Bacteriol.*, 188, 4553–4559

225.Tachdjian, S. & Kelly, R.M. (2006b). Dynamic metabolic adjustments and genome plasticity are implicated in the heat shock response of the extremely thermoacidophilic archaeon Sulfolobus solfataricus. *J. Bacteriol.*, 188, 4553–4559

226.Tang, T.H., Polacek, N., Zywicki, M., Huber, H., Brugger, K., Garrett, R., *et al.* (2005). Identification of novel non-coding RNAs as potential antisense regulators in the archaeon Sulfolobus solfataricus. *Mol. Microbiol.*, 55, 469–481

227.Tusher, V.G., Tibshirani, R. & Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci.*, 98, 5116–5121

228.Twellmeyer, J., Wende, A., Wolfertz, J., Pfeiffer, F., Panhuysen, M., Zaigler, A., *et al.* (2007). Microarray analysis in the archaeon Halobacterium salinarum strain R1. *PLoS One*, 2, e1064

229.Veit, K., Ehlers, C., Ehrenreich, A., Salmon, K., Hovey, R., Gunsalus, R.P., *et al.* (2006). Global transcriptional analysis of Methanosarcina mazei strain Gö1 under different nitrogen availabilities. *Mol. Genet. Genomics*, 276, 41–55

230.Velicer, G.J., Raddatz, G., Keller, H., Deiss, S., Lanz, C., Dinkelacker, I., *et al.* (2006). Comprehensive mutation identification in an evolved bacterial cooperator and its cheating ancestor. *Proc. Natl. Acad. Sci. U. S. A.*, 103, 8107–8112

231.Verhees, C.C., Kengen, S., Tuininga, J.J., Schut, G., Adams, M., de VOS, W., *et al.* (2003). The unique features of glycolytic pathways in Archaea. *Biochem. J*, 375, 231–246

232.Wagner, M., Berkner, S., Ajon, M., Driessen, A.M., Lipps, G. & Albers, S.-V. (2009). Expanding and understanding the genetic toolbox of the hyperthermophilic genus Sulfolobus. *Biochem. Soc. Trans.*, 37, 97

233.Wagner, M., van Wolferen, M., Wagner, A., Lassak, K., Meyer, B.H., Reimann, J., *et al.* (2012). Versatile Genetic Tool Box for the Crenarchaeote Sulfolobus acidocaldarius. *Front. Microbiol.*, 3, 214

234.Wahl, L.M. & Gerrish, P.J. (2001). The Probability That Beneficial Mutations Are Lost in Populations with Periodic Bottlenecks. *Evolution (N. Y).*, 55, 2606–2610

235.Wallin, E. & Heijne, G. Von. (2008). Genome-wide analysis of integral

membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.*, 7, 1029–1038

236.Walther, J., Sierocinski, P. & Van Der Oost, J. (2010). Hot transcriptomics. *Archaea*, 2010

237.Wang, K. (n.d.). Transcription regulation and growth phase transition in hyperthermophilic archea. Stockholm University

238.Wang, K., Sybers, D., Maklad, H.R., Lemmens, L., Lewyllie, C., Zhou, X., *et al.* (2019). A TetR-family transcription factor regulates fatty acid metabolism in the archaeal model organism Sulfolobus acidocaldarius. *Nat. Commun.*, 10, 1542

239.Weinberg, M. V, Schut, G.J., Brehm, S., Datta, S. & Adams, M.W.W. (2005). Cold shock of a hyperthermophilic archaeon: Pyrococcus furiosus exhibits multiple responses to a suboptimal growth temperature with a key role for membrane-bound glycoproteins. *J. Bacteriol.*, 187, 336–348

240.Weiss, M.C., Sousa, F.L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-Sathi, S., *et al.* (2016). The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.*, 1, 16116

241.Wende, A., Furtwängler, K. & Oesterhelt, D. (2009). Phosphate-dependent behavior of the archaeon Halobacterium salinarum strain R1. *J. Bacteriol.*, 191, 3852–3860

242.Wernersson, R. & Nielsen, H.B. (2005). OligoWiz 2.0—integrating sequence feature annotation into the design of microarray probes. *Nucleic Acids Res.*, 33, W611–W615

243.Wessel, D. & Flügge, U.-I. (1984). A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.*, 138, 141–143

244.Wheaton, G.H., Mukherjee, A. & Kelly, R.M. (2016). Transcriptomes of the extremely thermoacidophilic archaeon Metallosphaera Sedula exposed to metal “shock” reveal generic and specific metal responses. *Appl. Environ. Microbiol.*, 82, 4613–4627

245.Williams, E., Lowe, T.M., Savas, J. & DiRuggiero, J. (2007). Microarray analysis of the hyperthermophilic archaeon Pyrococcus furiosus exposed to gamma irradiation. *Extremophiles*, 11, 19–29

246.Wilson, D., Charoensawan, V., Kummerfeld, S.K. & Teichmann, S.A. (2008). DBD—taxonomically broad transcription factor predictions: new content and functionality. *Nucleic Acids Res.*, 36, D88–D92

247.Woese, C.R. & Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci.*, 74, 5088–5090

248.Woese, C.R., Kandler, O. & Wheelis, M.L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci.*, 87, 4576–4579

249.Wolf, J., Stark, H., Fafenrot, K., Albersmeier, A., Pham, T.K., Müller, K.B., *et al.* (2016). A systems biology approach reveals major metabolic changes in the thermoacidophilic archaeon Sulfolobus solfataricus in response to the carbon source L-fucose versus D-glucose. *Mol. Microbiol.*, 102, 882–908

250.Worthington, P., Hoang, V., Perez-Pomares, F. & Blum, P. (2003). Targeted disruption of the α-amylase gene in the hyperthermophilic archaeon Sulfolobus solfataricus. *J. Bacteriol.*, 185, 482–488

251.Wu, J., Wang, S., Bai, J., Shi, L., Li, D., Xu, Z., *et al.* (2008). ArchaeaTF: an integrated database of putative transcription factors in Archaea. *Genomics*, 91, 102–107

252.Wurtzel, O., Sapra, R., Chen, F., Zhu, Y., Simmons, B.A. & Sorek, R. (2010). A single-base resolution map of an archaeal transcriptome. *Genome Res.*, 20, 133–141

253.Xia, Q., Hendrickson, E.L., Zhang, Y., Wang, T., Taub, F., Moore, B.C., *et al.* (2006). Quantitative proteomics of the archaeon Methanococcus maripaludis validated by microarray analysis and real time PCR. *Mol. Cell. Proteomics*, 5, 868–881

254.Xia, Y., Wang, Y., Fang, H.H.P., Jin, T., Zhong, H. & Zhang, T. (2014). Thermophilic microbial cellulose decomposition and methanogenesis pathways recharacterized by metatranscriptomic and metagenomic analysis. *Sci. Rep.*, 4, 6708

255.Yang, I. V, Chen, E., Hasseman, J.P., Liang, W., Frank, B.C., Wang, S., *et al.* (2002). Within the fold: assessing differential expression measures and reproducibility in microarray assays. *Genome Biol*, 3, 1-0062.12

256.Yergeau, E., Schoondermark-Stolk, S.A., Brodie, E.L., Déjean, S., DeSantis, T.Z., Gonçalves, O., *et al.* (2009). Environmental microarray analyses of Antarctic soil microbial communities. *ISME J.*, 3, 340–351

257.Zaigler, A., Schuster, S.C. & Soppa, J. (2003). Construction and usage of a onefold-coverage shotgun DNA microarray to characterize the metabolism of the archaeon Haloferax volcanii. *Mol. Microbiol.*, 48, 1089–1105

258.Zaparty, M., Esser, D., Gertig, S., Haferkamp, P., Kouril, T., Manica, A., *et al.* (2009). “Hot standards” for the thermoacidophilic archaeon Sulfolobus solfataricus. *Extremophiles*, 14, 119–142

259.Zaparty, M., Zaigler, A., Stamme, C., Soppa, J., Hensel, R. & Siebers, B. (2008). DNA microarray analysis of central carbohydrate metabolism: glycolytic/gluconeogenic carbon switch in the hyperthermophilic crenarchaeum Thermoproteus tenax. *J. Bacteriol.*, 190, 2231–2238

260.Zhang, W., Culley, D.E., Nie, L. & Brockman, F.J. (2006). DNA microarray analysis of anaerobic Methanosarcina barkeri reveals responses to heat shock and air exposure. *J. Ind. Microbiol. Biotechnol.*, 33, 784–790

261.Zillig, W., Arnold, H.P., Holz, I., Prangishvili, D., Schweier, A., Stedman, K., *et al.* (1998). Genetic elements in the extremely thermophilic archaeon Sulfolobus. *Extremophiles*

262.Zillig, W., Stetter, K.O., Wunderl, S., Schulz, W., Priess, H. & Scholz, I. (1980). The Sulfolobus-“Caldariella” group: taxonomy on the basis of the structure of DNA-dependent RNA polymerases. *Arch. Microbiol.*, 125, 259–269

263.Zolghadr, B., Weber, S., Szabó, Z., Driessen, A.J.M. & Albers, S. (2007). Identification of a system required for the functional surface localization of sugar binding proteins with class III signal peptides in Sulfolobus solfataricus. *Mol. Microbiol.*, 64, 795–806

About the author

Pawel Sierocinski was born on the 21st December 1981 in Warsaw, Poland. In 2001 he started studying biotechnology in Warsaw University of Life Sciences, in 2005 he went to do Erasmus Exchange at Wageningen University where he transferred and finished his MSc in 2007. Same year he started his PhD in Microbial Genetics in Wageningen University Laboratory of Microbiology, working on multi-omics of *Sulfolobus solfataricus*. Currently he is a Research Fellow at the University of Exeter researching ecology and evolution of methane producing communities.

Acknowledgements

Over the time it took to complete my thesis many people have contributed to its completion. I would like to thank them for the support, help and motivation they kindly provided me with.

First of all my thanks go to John. Before you became my promotor, you were the person who infected me with the love of microbiology. When I first came to Wageningen I was not in awe of all things microbial and it was your course that changed that status and quite efficiently. You were also the person that gave me the chance of making a career of my newly found passion, for which I am eternally grateful. You were always there when I needed your advice or support, and never stopped believing in me even at times when I had doubts about myself. It is your unshaken belief in me and unconditional support that made it all possible. My thanks go also to Willem, who always managed to compress a life time of scientific wisdom into a five minute conversation and with a single glance identify any problems with my research.

Secondly I would like to thank my paranymphs. Corné, from the guy telling me off when we first met till the guy telling me off just today, you have been my external voice of reason. You soldiered on by my side through times that must have been frustrating for you – and that is something that very few people would have patience to do. You always listened and then converted what you heard into a perfect advice, even though it took me some time to understand some of it. Mark, from the moment we met I knew it was special relationship. You are one of a kind (but you know it, don't you?) and as such you were an important catalyser of changing me into the person I am today. Car pooling with you was probably the most intellectually challenging experience of my life and it shaped my way of thinking a lot. In some cases even for better. You helped me by being my scientific mentor, by helping when I needed it and by not helping when I needed that exactly.

I would like to thank all by colleagues from the office. Wally for taking me under his wing and teaching me all I needed to venture into the magical world of transcriptomics. Stan for giving me the best advice possible and expanding my way of thinking. Fab for always being in the middle of things with his sober evaluation. Jaapie for adding this rock and roll aspect to looking at research. Matthias for showing that stoic does not mean apathetic. Harmen for being polemic at all cost and using Socratic method to improve my way of thinking.

Thanks to Nicolas, Pierpaolo – you were always there to listen to my ramblings for which I am very sorry. Edze, Bart, Mark M., Mark L., Marco, – office football would not be the same without you guys. Rie, Magnus – sorry for the office football. Thanks to Daan, John R., Tom, Sinan, Sjon, Raymond Bram, Tessa, Bas, Detmer, Susanne and all the other BacGen people over the years.

I would also like to thank all the people that taught me during my time in Wageningen. Gosse, Serve, Fons, Caroline, Hauke – you were essential in making me interested in microbiology.

No science can realistically be done without collaborations. Bettina, Christa, Dietmar, Phil, Sonja, Melanie, Patrick, Andrea K., Andrea M., Benham, Dominic, Patricia, Khoa, it was a pleasure working together and thanks for all the help.

I would like to thank also to my new colleagues in Exeter. Firstly to my boss, Angus, for freedom he granted me over the years. Thanks to Mick for getting me on track with the change of topic and mentoring me through this major change in my life. Thanks to Uli, Peter, Elze, Siobhan, Adela, Ville, Pedro, Floh and Jesica for creating a one-in-a-million work atmosphere.

I would like to thank my family – my mom Dorota and my father Piotr for equipping me all through my childhood with skills I needed throughout my life. For years of dedication and support but most importantly for making sure that I know they will stand by me whatever. With such support you can achieve all you dream of. It has been a true blessing. I would like to thank my brother Adam for being my exact polar opposite at the same time as being so similar to me. Thanks bro – that always made my understanding of the world richer. Having such a close contact with someone that processes the world in a different way makes one much more humble and I really appreciate that lesson in humility you have always given me. Voglio anche ringraziare la mia nuova famiglia. Grazie Assunta e Ciccio! Per tutto l'aiuto che ho ricevuto da voi, per tutto il supporto, per credere in me.

Last, but definitely not least I would like to thank my wife and my daughter. Daniela, baby, you have been my rock through all this time. And I think that we use the rock analogy in a much too narrow meaning to describe others. Rock is not only a solid support, firm ground we walk on. Rock gave us tools to build the whole civilisation, rock is something we use for sheltering through rough times, rock gave us fire to nourish and to protect us, rock is the constant

as time passes and other things change. You are all this and much more to my one person civilisation and that is more than anyone could ask. None of this would not be possible without you, nor would it matter if you were not there. On top of all that support you gifted me with the greatest treasure of all: Clara. The ultimate reason for being thankful every day. Clara, now that you can read it – daddy is very grateful for all the brave help you have given him every day that you were here. Love you!

List of publications

Sierocinski P, Bayer F, Yvon-Durocher G, Burdon M, Großkopf T, Alston M, Swarbreck D, Hobbs PJ, Soyer OS, Buckling A, (2018). Biodiversity-function relationships in methanogenic communities. *Molecular Ecology 27*, 4641-4651.

Sierocinski P, Milferstedt K, Bayer F, Soyer O, Buckling A. (2017). A single community dominates structure and function of a mixture of multiple methanogenic communities. *Current Biology* 27, 3390-3395.

Friman VP, Soanes-Brown D, **Sierocinski P**, Molin S, Johansen HK, Merabishvili M, Pirnay JP, De Vos D, Buckling A (2016). Pre-adapting parasitic phages to a pathogen leads to increased pathogen clearance and lowered resistance evolution with *Pseudomonas aeruginosa* cystic fibrosis bacterial isolates. *Journal of Evolutionary Biology*, 29, 188-198.

McLeman A, **Sierocinski P**, Hesse E, Buckling A, Perron G, Hülter N, Johnsen PJ, Vos M (2016). No effect of natural transformation on the evolution of resistance to bacteriophages in the *Acinetobacter baylyi* model system, *Scientific Reports,* 6, 37144.

Esser D, Kouril T, Zaparty M, **Sierocinski P**, Chan PP, Lowe T, Van der Oost J, Albers S-V, Schomburg D, Makarova KS, Siebers B (2011), Functional curation of the *Sulfolobus solfataricus* P2 and S. acidocaldarius 98-3 complete genome sequences. *Extremophiles,* 15, 6, 711.

Walther J, **Sierocinski P**, van der Oost J (2011). Hot transcriptomics. *Archaea,* 2010.

Pham TK, **Sierocinski P**, van der Oost J, Wright PC (2010). Quantitative proteomic analysis of *Sulfolobus solfataricus* membrane proteins. *Journal of proteome research* 9, 2, 1165-1172.

Zaparty M, Esser D, Gertig S, Haferkamp P, Kouril T, Manica A, Pham TK, Reimann J, Schreiber K, **Sierocinski P**, Teichmann D, van Wolferen M, von Jan M, Wieloch P, Albers S-V, Driessen AJM, Klenk H-P, Schleper C, Schomburg D, van der Oost J, Wright PC, Siebers B (2010). “Hot standards” for the thermoacidophilic archaeon *Sulfolobus solfataricus*. *Extremophiles,* 14, 1, 119.

Albers S-V, Birkeland N-K, Driessen AJM, Gertig S, Haferkamp P, Klenk H-P, Kouril T, Manica A, Pham TK, Ruoff P, Schleper C, Schomburg D, Sharkey KJ, Siebers B, **Sierocinski P**, Steuer R, van der Oost J, Westerhoff HV, Wieloch P, Wright PC, Zaparty M (2009). SulfoSYS (Sulfolobus Systems Biology): towards a silicon cell model for the central carbohydrate metabolism of the archaeon *Sulfolobus solfataricus* under temperature variation. *Biochemical Society Transactions,* 37, 1, 58-64.

Overview of completed training activities

Conferences:

SYSMO kick-off meeting 2007
Thermophiles conference, Bergen 2007
SYSMO meeting in Vienna 2008
SYSMO meeting in Bad Honnef 2008
SYSMO PAL meeting, Heidelberg 2008
Sulfosys meeting in Essen 2008
Ruby on rails workshop, Heidelberg 2008
SYSMO PAL meeting in Amsterdam 2009
Spring Sympoosium NVvM in Arnhem 2010
Microbiology conference, Pappendal 2010
CRISPR Conference in Wageningen 2010
Eco-genomics course, Wageningen 2010
Aquatic Virus workshop, Texel 2011
Exp. Evolution & Community Dynamics, Helsinki 2016
ESEB conference in Groningen 2017
ISME conference in Leipzig 2018
Metagenomics Workshop, Penryn 2018
EE&CD, Helsinki 2018

Courses:

Radiation Expert, Wageningen 2007
Scientific Writing, Wageningen 2008
MICRO excursion 2009
VLAG PhD week 2009
Presentation skills, Penryn 2012
Academic CV writing, Penryn 2013
Data visualisation, Truro 2013
Social Networking in research, Penryn 2014
Preparing for leadership, Bristol 2015
Finding funding in STEM, Penryn 2015

The research described in this thesis was conducted in the Laboratory of Microbiology at Wageningen University and financially supported by SysMO

Cover design by Edyta Wocial

Printed by ProefschriftMaken / Digiforce