

9 Functional Genomics of the Thermo-Acidophilic Archaeon *Sulfolobus solfataricus*

John van der Oost¹, Jasper Walther¹, Stan J J Brouns¹, Harmen J G van de Werken¹, Ambrosius P L Snijders², Phillip C Wright², Anders Andersson³, Rolf Bernander⁴ and Willem M de Vos¹

¹ Laboratory of Microbiology, Wageningen University, The Netherlands; ² Biological and Environmental Systems Group, Department of Chemical and Process Engineering, Sheffield University, UK;

³ Department of Biotechnology, KTH – Royal Institute of Technology, Stockholm, Sweden;

⁴ Department of Molecular Evolution, Evolutionary Biology Center, Uppsala University, Sweden



CONTENTS

Introduction
Thermophiles
Thermophile genomics
Transcriptome analysis
Proteome analysis
Concluding remarks

◆◆◆◆◆ INTRODUCTION

Archaea and bacteria that optimally grow at temperatures above 60°C and 80°C are referred to as thermophiles and hyperthermophiles, respectively (Stetter, 1996). Since their discovery in the late 1960s (Brock and Freeze, 1969), attempts were made to reveal the secrets of the thermal resistance of these microorganisms, initially by physiological, biochemical and genetic analysis (Allers and Mevarech, 2005). In addition, the sequencing of the genomes of many thermophiles during the last decade has allowed for a series of genome-based research lines. Comparative genomics is the *in silico* analysis of genome data that aims to predict the metabolic potential of an organism, including the inter-conversions of metabolites and the regulation thereof (Ettema *et al.*, 2005; Makarova and Koonin, 2003). Functional genomics is the experimental analysis at the level of RNA (transcriptomics), protein (proteomics), as well as metabolites (metabolomics). In general, such holistic

studies aim at addressing the phenotypic response of an organism either to different cultivation conditions, or to genotypic variations. Here we review recent developments of functional genomics of thermophiles in general, and of the thermo-acidophilic archaeon *Sulfolobus solfataricus* in particular.

◆◆◆◆◆ THERMOPHILES

Forty years ago it was generally accepted that life was not possible at temperatures higher than 60°C. In 1969, however, Thomas Brock and co-workers extended the upper temperature limit to 75°C when a microorganism was isolated from thermal springs in Yellowstone National Park (Brock and Freeze, 1969). This organism was characterized as a Gram-negative bacterium that was named *Thermus aquaticus*. Subsequently, Brock's group isolated a remarkable thermo-acidophilic microorganism that they named *Sulfolobus acidocaldarius* from a number of both thermal acid soils and acid hot springs in Yellowstone National Park where there is an abundant source of sulfur, low pH (less than 3.0), and high temperature (65 to 90°C) (Brock *et al.*, 1972). Based on morphological analysis, *Sulfolobus* was initially classified as a thermo-acidophilic bacterium with some remarkable properties (Shivers and Brock, 1973). However, the molecular classification introduced by Carl Woese and colleagues indicated that *Sulfolobus* does not belong to the Bacteria, but rather to a newly-discovered prokaryotic domain: the Archaea (for review see Woese, 1987; Pace, 1997).

The pioneering work of Brock set the stage for further exploration of a wide range of thermal ecosystems worldwide. This has resulted in an ever-growing collection of thermophiles (both Archaea and Bacteria; optimal growth 60–80°C) and hyperthermophiles (mainly Archaea and some Bacteria; optimal growth >80°C) that have been isolated from terrestrial solfataric fields (e.g. the crenarchaea *Sulfolobus solfataricus*, *Sulfolobus acidocaldarius*, *Sulfolobus tokodaii*, isolated from Italy, USA, Iceland, Japan, New Zealand), but also from marine ecosystems ranging from shallow vents (e.g. the bacterium *Thermotoga maritima*, and the euryarchaea *Pyrococcus furiosus*, *Thermococcus kodakaraensis* from Italy and Japan) to deep sea smokers (Pacific, Atlantic Ocean; e.g. the euryarchaea *Methanocaldococcus jannaschii*, *Pyrococcus abyssi*, *Pyrococcus horikoshii*) (reviewed by Fukui *et al.*, 2005; Huber *et al.*, 2000; Rothschild and Mancinelli, 2001; Stetter, 1996, 1999). The most robust hyperthermophiles described to date are crenarchaea of the family Pyrodictiaceae: *Pyrolobus fumarii* (optimum at 106°C, growth up to 113°C; Blochl *et al.*, 1997) and a related iron-reducing isolate (optimum 106°C, growth between 113 and 121°C; Kashefi and Lovley, 2003). Remarkably, vegetative cultures of *Pyrolobus* and *Pyrodictium* have been demonstrated to survive autoclaving (Kashefi and Lovley, 2003; Stetter, 1999).

◆◆◆◆◆ THERMOPHILE GENOMICS

The last decade several thermophilic Archaea have been selected for complete genomic sequencing (Table 9.1), not only for the expected discovery of robust biocatalysts, but also because of the anticipated insight into: (i) thermophile physiology (unique metabolic enzymes and pathways, especially in Archaea; Ettema *et al.*, 2005); Makarova and Koonin, 2003, (ii) the molecular basis of thermostability of bio-molecules (enhanced numbers of charged residues at surface and subunit/domain interfaces; Cambillau and Claverie, 2000), and (iii) the evolution of the eukaryotic cell (fusion of archaeal and bacterial cell; Rivera and Lake, 2004).

The first complete genome analysis of an archaeon, *Methanocaldococcus jannaschii* (Bult *et al.*, 1996), confirmed 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 organisation 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) 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*) (reviewed by Makarova and Koonin, 2003). The comparative analysis of the genome of the hyperthermophilic bacterium *Thermotoga maritima* with that of *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 (Koonin *et al.*, 2001; Nelson *et al.*, 1999). In addition, the comparison of closely related species (*P. furiosus*, *P. abyssi*, *P. horikoshii*) revealed a high degree of genome plasticity; moreover, it was proposed that the lateral gain as well as the loss of genes is a modular event (Ettema *et al.*, 2001). Horizontal gene transfer has also been proposed to explain the relatively high degree of homology between genome fragments of the euryarchaeon *Thermoplasma acidophilum* and the crenarchaeon *Sulfolobus solfataricus*, phylogenetically distant Archaea but inhabiting the same environment (65–85°C, pH 2). The *Sulfolobus*-like genes in the *T. acidophilum* genome are clustered into at least five discrete regions, again indicating recombination of larger DNA fragments (Frickey and Lupas, 2004; Ruepp *et al.*, 2000).

After the genome sequence of thermophiles was established, comparative genomics analyses have been performed to assign potential functions for the identified open reading frames. As in all studied genomes, many unique and conserved hypothetical genes (typically half of the total number of genes) were found for which a function could not reliably be predicted. Hence, the main challenge of the

Table 9.I Thermophilic Archaea and Bacteria that are mentioned in the text

Name	Lifestyle	Genome (Mbp)	Reference
Bacteria			
<i>Thermotoga maritima</i>	Anaerobic chemorganotroph	1.9	Nelson <i>et al.</i> (1999)
Archaea – Crenarchaea			
<i>Sulfolobus solfataricus</i>	Aerobic thermo-acidophile, chemorganotroph, sulphur oxidizer	3.0	She <i>et al.</i> (2001)
<i>Sulfolobus acidocaldarius</i>	See <i>S. solfataricus</i>	2.3	Chen <i>et al.</i> (2005)
<i>Sulfolobus tokodaii</i>	See <i>S. solfataricus</i>	2.7	Kawarabayasi <i>et al.</i> (2001)
Archaea – Euryarchaea			
<i>Methanocaldococcus jannaschii</i>	Anaerobic thermophilic chemo-lithoautotrophic methanogen	1.7	Bult <i>et al.</i> (1996)
<i>Pyrococcus furiosus</i>	Anaerobic heterotrophic thermophile, grows on peptides and sugars	1.9	Robb <i>et al.</i> (2001)
<i>P. abyssi</i>	Anaerobic heterotrophic thermophile, grows on peptides	1.8	Cohen <i>et al.</i> (2003)
<i>Pyrococcus horikoshii</i>	See <i>P. abyssi</i>	1.7	Kawarabayasi <i>et al.</i> (1998)
<i>Thermoplasma acidophilum</i>	Facultative anaerobic chemorganotrophic thermoacidophile, lacks cell-wall	1.6	Ruepp <i>et al.</i> (2000)
<i>Thermoplasma volcanium</i>	See <i>T. acidophilum</i>	1.6	Kawashima <i>et al.</i> (2000)
Archaea – Nanoarchaea			
<i>Nanoarchaeum equitans</i>	Obligate archaeal parasite, depends on host for metabolites, extremely reduced genome	0.5	Waters <i>et al.</i> (2003)

post-genome era 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 thermophile genomes are the numerous missing links in metabolic pathways as well as the largely unknown regulatory systems (Ettema *et al.*, 2005; Makarova and Koonin, 2003).

The holistic, and often high-throughput analysis of genome structure and function relations is referred to as Functional Genomics. This genomics-based experimental analysis is sub-divided on the basis of the level of analysis: RNA (transcriptomics), protein (proteomics), and metabolites (metabolomics). We here review recent developments of the different types of functional genomics analyses of thermophilic model organisms, with a main focus on the thermo-acidophilic archaeon *Sulfolobus solfataricus*.

◆◆◆◆◆ TRANSCRIPTOME ANALYSIS

A transcriptome is the complete set of RNA transcripts detectable in an organism of interest at a certain moment in time. A typical transcriptome experiment is the analysis either of a single genotype (e.g. wild type) of the selected organism that is grown under distinct cultivation conditions, or of different genotypes (e.g. wild-type and mutant) that are grown under similar conditions. Analysis of the complete set of transcripts may be performed by using a DNA microarray, comprising DNA fragments that correspond to all genes. These fragments can be (i) a genomic fragment (plasmid library), (ii) a PCR-amplified (full-length/partial) copy of the gene, or (iii) one or more oligonucleotides (length either 25-, 50-, 60-, or 70-mer), the sequences of which correspond to gene fragments (Schena, 1996). The DNA fragments are either spotted on a nitrocellulose filter (macroarray), or printed on a glass slide (microarray); alternatively, oligonucleotides (typically 25-mers) may be synthesized on the microarray directly (Affymetrix). When dedicated arrays are available, the experimental procedures are as follows (experimental details are provided below): (i) isolation of total RNA at different time points from a single culture, or simultaneously from two separate cultures, (ii) synthesis of cDNA from each batch of RNA, and incorporation of aminoallyl-dUTP, (iii) labelling of cDNA with distinct variants of a fluorescent dye (e.g. Cy-3 and Cy-5; Amersham Biosciences), (iv) hybridization of a specific DNA microarray with the labelled cDNA, and (v) data acquisition (scanning and image analysis), processing (filtering and normalization) and analysis (averaging, statistical analysis). The resulting ratios of fluorescent signals reflect the relative levels and possibly the fluctuation of expression of all genes included on the DNA microarray.

DNA Microarrays of Thermophiles

DNA microarrays have initially been established as high-throughput functional genomics tools to study eukaryal and bacterial model systems. The first microarray analysis reported on either a hyperthermophile or an archaeon was a pilot study on *Pyrococcus furiosus* that focussed on 271 PCR products of (potential) metabolic genes (Schut *et al.*, 2001; Table 9.2). This initial study was followed by analyses on a complete genome array (Schut *et al.*, 2003; Weinberg *et al.*, 2005; Table 9.2). These studies addressed the adaptation of *Pyrococcus* cells to the availability of sulfur, to different carbon sources, and to cold shock. A similar approach has been taken for the bacterium *Thermotoga maritima*: 269 PCR products of genes predicted to be responsible for the potential of the organism to grow on a wide variety of carbohydrates and (poly)peptides were spotted on glass slides, and hybridized with differentially labelled cDNA derived from RNA from cultures grown with glucose and tryptone (Chhabra *et al.*, 2003; Table 9.2). A complete array of *Thermotoga maritima* has recently been established at TIGR (Nguyen *et al.*, 2004); using chemostat cultures on different sugars catabolite repression has been studied in this thermophilic bacterium.

DNA Microarray Analysis of *Sulfolobus*

DNA microarrays have recently also been developed for *Sulfolobus solfataricus* and *S. acidocaldarius* (Andersson *et al.*, 2005). Based on the genome sequences of *S. solfataricus* (3.0 Mbp) (She *et al.*, 2001) and *S. acidocaldarius* (2.3 Mbp) (R. Garrett, personal communication), internal fragments of genes (100–800 bp) were PCR amplified, checked by agarose gel electrophoresis, and printed on glass slides. Software (DualPrime; <http://www.biotech.kth.se/molbio/microarray/>) has been developed that allowed for reducing the number of primers required to PCR amplify related (often orthologous) genes or fragments thereof: 2488 genes from *Sulfolobus solfataricus* and 1960 from *Sulfolobus acidocaldarius* (Lundgren *et al.*, 2004). The excellent quality of this *Sulfolobus* microarray has been demonstrated by an experiment in which hybridization with genomic DNA from synchronized cells in different phases of the cell cycle revealed the three origins of replication of the *Sulfolobus* chromosome (Lundgren *et al.*, 2004). A wide range of transcriptomics experiments has been performed with these arrays (Bernander *et al.*, unpublished), including recent experiments in which *S. solfataricus* was grown on sugars (e.g. glucose, arabinose) or on peptides (e.g. tryptone) (Brouns *et al.*, unpublished; Snijders *et al.*, 2006). In both studies, comparative and different functional genomics methods have been integrated. Below, experimental details of a typical transcriptome analysis experiment of *S. solfataricus* are provided.

Table 9.2 Thermophile microarrays – comparison of relevant details; for details (on *Sulfolobus solfataricus*; <http://www.biotech.kth.se/molbio/microarray/index.html>) see text

#	(1) <i>Pyrococcus furiosus</i>	(2) <i>Pyrococcus furiosus</i>	(3) <i>Pyrococcus furiosus</i>	(4) <i>Thermotoga maritima</i>	(5) <i>Sulfolobus solfataricus</i>
Reference	Schut <i>et al.</i> (2001)	Schut <i>et al.</i> (2003) Weinberg <i>et al.</i> (2005)	Shockley <i>et al.</i> , (2003)	Chhabra <i>et al.</i> (2003), Johnson <i>et al.</i> (2005)	Brouns <i>et al.</i> (unpublished results), Snijders <i>et al.</i> (2006)
Genome ORFs (estimated)	2065	2065	2065	1928	2950
Array ORFs	271	2065	201	269	2488
Array	Aminosilane-coated glass slides	Aminosilane-coated glass slides	CMT-GAPS aminosilane- coated slides (Corning)	As (3)	Aminosilane-coated glass slides
Synth/print	Slide printer (Omnigrid)	As (1)	DNA was attached to the substrate by UV cross-linking in at 250 mJ and baking at 75°C for 2 h	As (3)	
DNA	PCR Full length, or 1000 bp	As (1)	PCR 400–700 bp	PCR Size not indicated	PCR
Cultivation	Batch	Batch	Batch	Batch	Batch

(continued)

Table 9.2 Continued

#	(1) <i>Pyrococcus furiosus</i>	(2) <i>Pyrococcus furiosus</i>	(3) <i>Pyrococcus furiosus</i>	(4) <i>Thermotoga maritima</i>	(5) <i>Sulfolobus solfataricus</i>
Lysis				Cell pellet frozen on dry ice; frozen cells were disrupted by using a mortar and pestle and 12 volumes of lysis/binding solution (Ambion)	Peletted cells dissolved in Lysis Buffer (guanidinium isothiocyanate, citrate, sarcosyl, beta-mercaptoethanol; see text)
Total RNA isolation	Acid phenol extraction	As (1)	RNAqueous kit (Ambion).	As (3)	Acid phenol extraction
RT enzyme	Stratascript RT (Stratagene), purified using a QIAquick PCR purification kit (Qiagen)	As (1)	Stratascript RT (Stratagene), purified using GFX columns (Amersham)	Stratascript RT (Stratagene), purified using GFX columns (Amersham)	Superscript-II RT (Invitrogen), purified using MinElute columns (Qiagen)

RT mix	1 mM dATP, 1 mM dCTP, 1 mM dGTP, 0.3 mM dTTP, 0.5 mM aminoallyl dUTP Random 9-mers (Stratagene)	ARES DNA labelling kit (Molecular Probes)	0.5 mM dATP, 0.5 mM dCTP, 0.5 mM dGTP, 0.3 mM dTTP, 0.2 mM aminoallyl dUTP random 6-mers (Invitrogen)	0.5 mM dATP, 0.5 mM dCTP, 0.5 mM dGTP, 0.3 mM dTTP, 0.2 mM aminoallyl dUTP random 6-mers (Invitrogen)	0.5 mM dATP, 0.5 mM dCTP, 0.5 mM dGTP, mM dTTP, (Amersham Biotech) 0.4 mM aminoallyl dUTP (Sigma) random 6-mers/9-mers (Qiagen)
Dyes	Alexa 488 Alexa 594 (Molecular Probes)	Alexa dyes 488, 546, 594, or 647 (Molecular Probes)	Cy-3, Cy-5 (Amersham Biotech)	Cy-3, Cy-5 (Amersham Biotech)	Cy-3, Cy-5 (Amersham Biotech)
Pre-hybridization	Not indicated	Not indicated	5x SSC, 0.1% SDS, and 1% BSA at 42°C for 45 min	5x SSC, 0.1% SDS, and 1% BSA at 42°C for 45 min	5x SSC, 0.1% SDS, and 1% BSA at 42°C for 40 min
Hybridization probe	Pooled differentially- labelled cDNA	Pooled differentially- labelled cDNA	Pooled differentially- labelled cDNA, with 20 ug competitor DNA (COT1); denatured 3 min at 95°C	Pooled differentially- labelled cDNA, with 40 µg competitor DNA (COT1/polyA); denatured 3 min at 95°C	Pooled differentially- labelled cDNA, with 20 µg competitor DNA (tRNA, hering sperm DNA); denatured 2 min at 95°C

(continued)

Table 9.2 Continued

#	(1) <i>Pyrococcus furiosus</i>	(2) <i>Pyrococcus furiosus</i>	(3) <i>Pyrococcus furiosus</i>	(4) <i>Thermotoga maritima</i>	(5) <i>Sulfolobus solfataricus</i>
Hybridization	65°C in a humidity chamber (Arrayit), 10–15 h	65°C (?) in Genetac hybridization station (Genomic Solutions), 10–15 h	Buffer: 50% formamide, 5x SSC, 0.1% SDS; 42°C in hybridization chamber (Corning), 16–20 h	Buffer: 50% formamide, 5x SSC, 0.1% SDS; 42°C in hybridization chamber (Corning), 16–20 h	Buffer: 50% formamide, 5x SSC, 0.1% SDS; 42°C in hybridization chamber, at least 20 h
Wash	Washed twice for 5 min in each of 2x SSC-0.1% SDS (SDS), and 0.2x SSC-0.1% SDS and then rinsed in distilled water and blown dry with compressed air	Washed for 20 in each of 2x SSC-0.1% Tween 20, 0.2x SSC-0.1% Tween 20, 0.2x SSC, and finally rinsed in distilled water and blown dry with compressed air	Washed for 4 min in 1x SSC-0.2% SDS (42°C), for 4 min in 0.1x SSC-0.2% SDS (20°C), and for 4 min in 0.1x SSC (20°C) and allowed to air dry.	Washed for 4 min in 1x SSC-0.2% SDS (42°C), for 4 min in 0.1x SSC-0.2% SDS (20°C), and for 4 min in 0.1x SSC (20°C) and allowed to air dry	Washed for 5 min in 2x SSC-0.1% SDS (42°C), for 10 min in 0.1x SSC-0.1% SDS (20°C), and for 5 min in 0.1x SSC (20°C) and blown dry with compressed air

Scan	Scan Array 5000 spectrometer (Packard, Meriden, Conn.) with the appropriate laser and filter settings and analyzed by using Quantarray (Packard)	Scan Array 5000 slide reader (Perkin-Elmer); identified and quantitated by using the Gleams software package (Nutec, Houston, Tex.).	Scanarray 4000 scanner and Quantarray software, respectively (GSI Lumonics, Billerica, Mass.). Local background intensity was subtracted from each spot signal, and spotting buffer (50% dimethyl sulfoxide (DMSO)) was used to subtract global background. The signal from the Cy-3 channel was normalized to the signal from the Cy-5 channel based on total signal intensity.	Slides were scanned using a Scanarray 4000 scanner (GSI Lumonics and Billerica); Signal intensity data were obtained using Quantarray (GSI Lumonics).	Hybridised slides are scanned using the high-throughput DNA Microarray scanner model G2565BA from Agilent Technologies (see text)
Statistics	Paired <i>t</i> test	Individual <i>t</i> -test; Holm's step-down <i>p</i> value adjustment procedure was performed to give modified <i>p</i> values	Paired <i>t</i> test		Student's <i>t</i> -test (SAM); see text.

Gene amplification and printing of *Sulfolobus* microarray

As mentioned above, different types of DNA microarrays exist: Affymetrix-type with short oligonucleotides (25-mers) synthesized on the array (Schna, 1996), glass slides printed with oligonucleotides (50/75-mers), or glass slides printed with either larger DNA fragments (PCR products) were approximately 1000 bp or, in case of smaller coding regions, full length; Table 9.1). The first developed array for *Sulfolobus* consists of PCR-amplified fragments of the majority of the coding regions of two model *Sulfolobus* species. The amplification was done with specific primers (Andersson *et al.*, 2005), using a touchdown PCR protocol, allowing for melting temperature differences between primer pairs. After a quality test of the PCR-products on agarose gel, purification has been performed using multiscreen 384-SEQ filter plates (Millipore). After washing with double demineralized water (ddH₂O), the DNA was then dissolved in 30 µl ddH₂O, and DMSO was added to a final concentration of 50%. The PCR fragments were spotted in duplicate on amino(propyl)-silane-coated glass slides using a robotic (Qarray; Genetix) slide printer (Lundgren *et al.*, 2004). Apart from this frequently used technology, slides with alternative coatings are available: (i) epoxy-silane, for attaching amine-modified DNA, (ii) aldehyde silane, also for attaching amine-modified DNA, and (iii) poly-L-lysine coating, that forms a dense layer of amine groups (like aminosilane) for initial ionic attachment of the DNA-phosphate groups, which can subsequently be attached covalently to the slide by either baking or by UV irradiation (<http://www.eriemicroarray.com/substrates/epoxy.aspx>).

Sulfolobus growth and harvest

Sulfolobus solfataricus P2 (DSM1617) was grown on sugars (0.3% (w/v) D-glucose, D-galactose or D-arabinose) or on peptides (0.3% (w/v) tryptone) using the defined medium (Table 9.3), slightly adjusted from Brock's basal salt medium (Brock *et al.*, 1972). To obtain exponentially growing cells, 1000 ml medium is inoculated with appropriate volume of an exponentially growing culture, and grown for 10 generations to an OD₆₀₀ of 0.1–0.3 (approximately 60 h in baffled Erlenmeyer flasks (2 liter) at 80°C in a rotary shaker). Flow cytometry is a convenient method to examine the physiological status of the cells (DNA and cell integrity), and to monitor cell cycle progression (DNA content distribution; relative number of cells in different cell cycle stages) (Bernander and Poplawski, 1997). At the appropriate density, (part of) the cell culture is quickly poured into 50 ml Greiner tubes and cooled on icewater. The cold cell suspension is centrifuged (15 min; 5000 × g; 4°C), after which the supernatant is carefully removed, and tubes are stored at –80°C.

Sulfolobus lysis and RNA isolation

All solutions described in this section are routinely made RNase-free by double autoclaving (for practical suggestions see: http://www.ambion.com/techlib/tb/tb_180.html). In addition, RNase-free water is ddH₂O,

Table 9.3 *Sulfolobus* Defined Medium, and Rich Medium (DSM)

Defined medium used for the different carbon sources; adjust pH at room temperature to 4.0 with H₂SO₄; autoclave and store at room temperature.

(NH ₄) ₂ SO ₄	2.5 g l ⁻¹
KH ₂ PO ₄	3.1 g l ⁻¹
MgCl ₂ · 6H ₂ O	203.3 mg l ⁻¹
Ca(NO ₃) ₂ · 4H ₂ O	70.8 mg l ⁻¹
FeSO ₄ · 7H ₂ O	2.0 mg l ⁻¹
MnCl ₂ · 4H ₂ O	1.8 mg l ⁻¹
Na ₂ B ₄ O ₇ · 2H ₂ O	4.5 mg l ⁻¹
ZnSO ₄ · 7H ₂ O	0.22 mg l ⁻¹
CuCl ₂ · 2H ₂ O	0.06 mg l ⁻¹
Na ₂ MoO ₄ · 2H ₂ O	0.03 mg l ⁻¹
VOSO ₄ · 2H ₂ O	0.03 mg l ⁻¹
CoCl ₂ · 6H ₂ O	0.01 mg l ⁻¹

Carbon source

0.2–0.4% (w/v) monosaccharide, or tryptone

Wolin vitamin stock (2000×); adjust to pH 4.0 with H₂SO₄, filter sterilize and store at RT; add before use of defined medium:

Biotin	2 mg l ⁻¹
Folic acid	2 mg l ⁻¹
Pyridoxine-HCl	10 mg l ⁻¹
Riboflavin	10 mg l ⁻¹
Thiamine-HCl	5 mg l ⁻¹
Nicotinic acid	5 mg l ⁻¹
DL-Ca-Pantothenate	5 mg l ⁻¹
Vitamin B12	0.1 mg l ⁻¹
p-Aminobenzoic acid	5 mg l ⁻¹
Lipoic acid	5 mg l ⁻¹

Rich medium (Zillig *et al.*, 1980); Medium for *Sulfolobus solfataricus* strain DSM No. 1617; <http://www.dsmz.de/media/med182.htm>

Yeast extract (Difco)	2.00 g
KH ₂ PO ₄	3.10 g
(NH ₄) ₂ SO ₄	2.50 g
MgSO ₄ · 7H ₂ O	0.20 g
CaCl ₂ · 2H ₂ O	0.25 g
Distilled water	1000 ml

Adjust pH of the medium at room temperature to 3.5 with 10 N H₂SO₄ prior to autoclaving.

and incubated overnight with 0.1% (w/v) diethyl pyrocarbonate (DEPC, commonly used to irreversibly inactivate RNase), after which the solution is autoclaved to remove remaining DEPC. The Lysis Buffer contains 4 M guanidinium isothiocyanate, 25 mM sodium citrate (pH 7.0), 0.5% sarcosyl and 0.07 M β -mercaptoethanol. To minimize handling of the hazardous guanidinium isothiocyanate, it is recommended to prepare a stock solution (stable for at least 3 months when stored in the dark at room temperature): 50 g guanidinium isothiocyanate is dissolved into 58.6 ml ddH₂O, after which 3.52 ml 0.75 M sodium citrate (pH 7.0) is added. This solution is heated to 65°C and supplemented with 5.28 ml of sarcosyl (pre-heated to 65°C to liquefy). Preparation of Lysis Buffer is completed by adding 5.2 μ l β -mercaptoethanol to 1.0 ml of the latter stock solution. Cell lysis is accomplished as outlined in the box below.

S. solfataricus – Procedure for cell lysis

- Use 1 ml Lysis Buffer to dissolve the pelleted cells (in the 50 ml Greiner tubes), divided over two Eppendorf tubes, and incubated on ice with regular mixing.
- After 15 min add the following *cold* solutions to each tube:
 - 100 μ l 2 M Na-acetate (4°C; pH 5.2)
 - 500 μ l water-saturated phenol (4°C; pH 4.5).
 - 100 μ l 24:1 (v/v) chloroform-isoamyl alcohol mixture (4°C).
- Mix quickly and thoroughly and put on ice for 15 min.
- Next, the samples are centrifuged (15 min; 15 000 \times g; 4°C), resulting in distinct phases: RNA is in the top aqueous phase and DNA and proteins are primarily in the phenol or intermediate phase.
- The aqueous phase is carefully transferred to a new Eppendorf tube, and 500 μ l cold chloroform-isoamyl alcohol (24:1 (v/v); 4°C) is added and quickly and vigorously mixed.
- After an incubation of 15 min on ice, the solutions are spun down (30 min; 15 000 \times g; 4°C). The top (aqueous) phase is transferred to a new tube, and supplemented with an equal volume of cold isopropanol. This is mixed and incubated overnight at –20°C.
- Next, the tubes are centrifuged (15 min; 15 000 \times g; 4°C) and the supernatant is removed. The supernatant is dried and resuspended in 174 μ l DEPC-treated water. Remaining traces of DNA are subsequently removed by addition of:
 - 5 μ l (45 U) RNase-free DNase I (Ambion),
 - 20 μ l 10x DNase buffer (Ambion),
 - 1 μ l (40 U) RNAGuard RNase inhibitor (Promega).
- This solution is mixed and incubated for 30 min at 37°C, and the RNA is purified by adding the following *cold* solutions:
 - 20 μ l 2M Na acetate (4°C; pH 5.2),
 - 100 μ l water saturated phenol (4°C; pH 4.5),
 - 20 μ l 24:1 (v/v) chloroform-isoamyl alcohol mixture (4°C).

- This is mixed thoroughly and put on ice for 15 min.
- After centrifugation (30 min; 15 000 × g; 4°C), the aqueous phase is transferred to another tube and mixed with an equal amount of cold isopropanol, and incubated for at least 2 h, or rather overnight, at −20°C.
- After centrifugation (15 min; 15 000 × g; 4°C), the pellet is washed by adding 300 µl cold 75% ethanol.
- The RNA sample is centrifuged again (10 min, 10 000 × g, at 4°C), after which the supernatant is carefully removed. The remaining RNA pellet is briefly dried in the Speedvac (to avoid problems with dissolving the pellets, the pellets should not be dehydrated too much).
- The RNA pellet is dissolved in an appropriate solution (e.g. RNase-free ddH₂O) at a concentration of 3–4 µg µl⁻¹, and stored at −80°C.

A typical RNA yield is 10–20 µg per 50 ml cell culture harvested at OD₆₀₀ of 0.1–0.3. The concentration of RNA is routinely determined using a spectrophotometer (e.g. NanoDrop ND-1000; <http://www.nanodrop.com>). An alternative method of RNA isolation, which has previously been used for transcription analysis of *Sulfolobus* (Brinkman *et al.*, 2002), concerns the straight-forward RNeasy system (Qiagen RNeasy manual; http://www1.qiagen.com/literature/handbooks/PDF/RNASTabilizationAndPurification/FromAnimalAndPlantTissuesBacteriaYeastAndFungi/RNY_Midi_Maxi/1017849HBRNY_MidiMaxi0601WW.pdf). A possible drawback of the latter system is that part of the RNA molecules, especially those smaller than 200 nucleotides, is eluted during the washing steps, resulting in an underestimation of small RNAs (such as 5S rRNA, tRNAs); obviously the home made solutions for the phenol/chloroform extraction are less expensive than the RNeasy kit. On the other hand, the RNeasy method has practical advantages in that it is a quick and easy protocol.

***Sulfolobus* cDNA synthesis and labelling**

The synthesis of cDNA and the subsequent labelling are performed as outlined in the box below.

S. solfataricus – Procedure for cDNA synthesis and labelling

- A purified RNA sample (5–20 µg) is mixed with 5 µg random primer (Qiagen; <http://www1.qiagen.com>; 2.8 µl 1 mM random hexamer (MW 1791.7 g mol⁻¹), and/or 1.8 µl 1 mM random nonamer (MW is 2718.55 g mol⁻¹)) is added, and the total volume is adjusted to 11.6 µl with RNase-free water.
- Incubate for 10 min at 70°C, and then transfer to ice for at least 2 min.
- Add 8.4 µl Reverse Transcriptase (RT)-mix (Table 94), mix, and incubate for 2 h at 42°C to allow synthesis of the complementary strand, and incorporation of aminoallyl-dUTP (site for covalent attachment of fluorescent label, see below).

- Add 2 μl 200 mM EDTA and 3 μl 1 M NaOH, mix, and incubate for 15 min at 70°C (this will stop the reaction, and degrade the RNA).
- Add 3 μl 1 M HCl to neutralize the mixture.
- cDNA is purified by MinElute columns (Qiagen; alternative DNA purification systems can be used as well), and eventually elute in 10 μl 0.1 M NaHCO₃ (pH 9.0).

Table 9.4 *Sulfolobus* cDNA Synthesis – Solutions

Reverse Transcriptase (RT)-mixture:	
5x RT-buffer (Invitrogen)	4.0 μl
50x aadNTP mix (see below)	0.4 μl
DTT (0.1 M)	2.0 μl
SuperscriptII (200 U μl^{-1} ; Invitrogen)	2.0 μl
50x aadNTP (4:1 aadUTP:dTTP):	
10 μl each of 100 mM dATP, dGTP, dCTP (Amersham Pharmacia Biotech)	
8 μl 100 mM aminoallyl-dUTP (Sigma)	
2 μl 100 mM dTTP (Amersham Pharmacia Biotech)	
Dissolve all components in 0.1 M KPO ₄ buffer (pH 8.0)	

The Cy3 and Cy5 dyes (Amersham Biosciences; <http://www4.amershambiosciences.com>) are each dissolved in 55 μl DMSO and distributed over 10 Eppendorf tubes, dried by Speedvac, and stored at 4°C, preferably in a desiccator. For subsequent labelling with fluorescent Cy3 and Cy5 labels (Amersham Biosciences; alternative labels are available (Table 9.1)), the eluted cDNA solution (10 μl) is transferred to a Cy-dye aliquot tube (Amersham Biosciences), mixed, and incubated for 1.5 h at room temperature, in the dark. The reactions to be co-hybridized are mixed, and purified by a MinElute column (Qiagen; alternative DNA purification systems can be used as well (Table 9.1)). After elution from the spin column, the labelled cDNA is ready for microarray hybridization.

***Sulfolobus* microarray hybridization**

Each microarray is hybridized with two differently labelled cDNA samples: Cy3 or Cy5 (Amersham Biosciences). To enhance the reliability of the analysis, both technical replicates (including “dye swap”) as well as biological replicates (using RNA from different cultures) are routinely performed. Replicates obviously increase the accuracy of the microarray analysis (Yang and Speed, 2002). Although proper experimental design to some extent can reduce the number of technical replicates, a minimum of three biological replicates is recommended (Lee *et al.*, 2000).

The slides are transferred to a 50 ml Falcon tube with 50 ml of the pre-heated pre-hybridization solution (1% (w/v) BSA, 5x SSC (from 20x SSC stock: 3 M NaCl, 300 mM tri-Na citrate, pH 7.0), and 0.1% (w/v) SDS;

the BSA is mixed with the preheated SSC/SDS solution at 42°C until dissolved), and incubated for 40 min at 42°C. Subsequently, slides are washed three times with ddH₂O, and once with isopropanol. Slides are dried either by blowing with compressed air or by using a slide centrifuge. Hybridization should be started within one hour; details are described in the box below.

S. solfataricus – Procedure for microarray hybridization

- To 20 µl labelled DNA, the following solutions are added:
 - 1 µl tRNA (10 µg µl⁻¹),
 - 1 µl herring-sperm DNA (10 µg µl⁻¹),
 - 58 µl hybridization mix (final concentration: 50% (v/v) formamide, 5x SSC, 0.1% w/v SDS).
- The probe-solution is incubated for 2 min at 95°C, after which it is cooled on ice for at least 1 min, and spun down.
- A coverslip is applied to the slide and the probe-solution is injected. After transfer to the hybridization chamber (slide container), the array is hybridized at 42°C for at least 20 h.
- Washing consists routinely of the following steps:
 - 5 min in 42°C pre-warmed 2× SSC, 0.1% SDS,
 - 10 min in room temp 0.1× SSC, 0.1% SDS,
 - 5 × 1 min in room temp 0.1× SSC.
- Then slides are dried carefully, either by centrifugation or by blowing with compressed air.

Scanning, data extraction and normalization

After hybridization, the microarrays are scanned, using two different lasers, usually at a final resolution of at least 10% of the spot size using the optimal laser and filter settings (Leung and Cavalieri, 2003). The scanning settings determine the signal-to-noise ratio; optimal settings can be determined by either performing a quick pre-scan while manually adjusting the settings, or by using specific automatic optimization protocols that are linked to some scanners. During scanning, two different images are created that are analyzed by programs such as GenePix Pro (http://www.axon.com/gn_GenePix_File_Formats.html). The availability of open source microarray software has recently been reviewed (Dudoit *et al.*, 2003). Usually, a data filtering is required to exclude low quality spots. Usually, spots excluded from subsequent analysis include (i) spots with very low intensity (i.e. signal below the background plus 2 times the standard deviation (SD)), (ii) spots with unevenly distributed intensity (i.e. the ratio of medians deviates more than 20% from the regression ratio), and (iii) spots with saturated intensity. Before normalization and calculation of ratios, the background can be subtracted from the total intensities of the different spots. Although subtraction of

background can improve the detection of some differentially expressed genes, it should be kept in mind that it may result in an increase of the overall variance and thus, in a decrease of the sensitivity of the measurement (Qin and Kerr, 2004). There are many different ways of normalizing data; usually too much normalization of data results in overall reduced ratios, which gives rise to a decreased sensitivity. For that reason, it is suggested to use software that indicates the consequences of distinct normalization steps. If this option is not available, a safe choice would be to restrict to a global lowess (LOcally WEighted Scatterplot Smoothing) normalization for each slide (Cleveland and Devlin, 1988). To determine differentially expressed genes, different statistical tests are available. Differential expression is usually determined via the classical Student's *t*-test or variations thereof, including SAM (Significance Analysis of Microarrays) and B-test (Lönstedt and Speed, 2002). The latter variants are especially well suited for analysis of data with relatively few replicates; these methods tend to reduce the number of false positives, and as such are generally more suitable for DNA microarray experiments. The calculated *p* value is the statistical probability that the difference in gene expression occurs by change, the norm being that a gene is considered differentially expressed at $p < 0.05$ (Cui and Churchill, 2003).

◆◆◆◆◆ PROTEOME ANALYSIS

A proteome is the complete set of proteins present in a given organism under specific conditions at a certain moment. The analysis of proteomes requires (i) separating complex protein mixtures into discrete protein components, (ii) measuring their relative abundances, and (iii) identifying the individual protein components. Two-dimensional gel electrophoresis (2DE) is the classical method to separate proteins on the basis of their charge (isoelectric focusing, IEF) and of their size (sodium dodecyl sulfate polyacrylamide gel electrophoresis, SDS-PAGE). Proteins on a 2D gel can be visualized and subsequently quantified using several methods, the most frequently used are staining by: (i) Coomassie Brilliant Blue (reproducible quantification over linear dynamic range 50–1000 ng; CBB G250 (Sigma)), (ii) acidic silver nitrate (extremely sensitive (range 1–50 ng), although a drawback appears to be inhomogeneous protein staining), (iii) non-covalent fluorescent dyes (range 1–1000 ng; SYPRO Ruby (Molecular Probes) has been reported to be quite reproducible with similar binding characteristics as CBB) (reviewed by Barry *et al.*, 2003).

Separation and identification of complex protein and peptide mixtures are essential steps to understand the function and roles of proteins in the cell. Although new methods using multidimensional liquid chromatography have recently been developed (e.g. Multidimensional Protein Identification Technology (MudPIT) (McDonald and Yates, 2002; Washburn, 2004; Washburn *et al.*, 2002), protein separation by 2DE and subsequent identification by mass spectrometry is still the most

frequently used strategy in proteomics. After 2DE, stained or labelled proteins are extracted from individual spots and trypsin digested. Protein identification is routinely performed by either Matrix-Assisted Laser Desorption Ionization Time-of-Flight mass spectrometry (MALDI-TOF MS) or electrospray ionisation tandem mass spectrometry (ESI MS/MS). In the case of MALDI-TOF MS, proteins are usually identified through a peptide mass fingerprint (PMF) type of search. Each digested protein provides a specific fingerprint consisting of tryptic peptide masses. Software algorithms are then used to identify the protein from a database. The popularity of this approach appears to decline, since the explosively increased sequence database and because of the fact that sample impurities can complicate the identification process. ESI MS/MS provides much greater specificity towards database searching and can easily be integrated with LC-based methods for peptide separation. Therefore, MS/MS-based techniques have become the predominant tool for peptide identification. More recently, tandem mass spectrometers with a MALDI interface have become available (MALDI-TOF-TOF) that will help to increase the throughput of MALDI-based identification of 2D gel spots, and offers complementarity in types of peptides identified. In particular, on-line micro capillary reversed phase liquid chromatography interfaced to a tandem mass spectrometer (or to a spotting robot in the case of the MALDI-TOF-TOF) has made the MS/MS analysis more comprehensive, and has allowed for a higher throughput (Lim *et al.*, 2003; Link *et al.*, 1999).

Relative quantification of proteins on 2D gels can be obtained by imaging the intensity of (CBB, silver) stained proteins or fluorescent dye (SYPRO Ruby)-labelled proteins. When two protein samples are labelled with distinct fluorescent probes, the differential analysis can be performed using a single 2D gel. This approach has been termed "Difference in Gel Electrophoresis" (DiGE). More accurate and reproducible methods to measure relative expression of proteins labelled with stable isotopes have recently been established. These methods generate mass-over-charge (m/z) differences for each of the homologous peptides and proteins in the proteome, and expression ratios are measured by comparing peak areas for the protein, or peptide ions, measured in the mass spectrometer. Stable isotopes can be incorporated into proteins by distinct labelling approaches: (1) *chemical*: after cultivation the protein samples are digested with proteases, after which light (hydrogen- or ^{12}C -containing) and heavy (deuterium- or ^{13}C -containing) versions of Isotope-Coded Affinity Tags (ICAT) are covalently linked to cysteinyl residues of peptides, after which the labelled peptides are purified by affinity chromatography (no 2DE required, but 1D gels often help in pre-fractionation) (Gygi *et al.*, 1999; Washburn *et al.*, 2002; Zhou *et al.*, 2002); (2) *enzymatic*: protease-catalyzed hydrolysis of proteins in the presence of normal and heavy water, resulting in peptides with either ^{16}O or ^{18}O at the C-terminus (Oda *et al.*, 1999); and (3) *metabolic labelling*: cultivation with labelled substrates (e.g. ^{12}C and ^{13}C -glucose, or ^{14}N and ^{15}N -ammonium sulfate) and separation of the differently labelled protein samples by 2DE or liquid chromatography (Snijders *et al.*, 2005a).

Proteomics of thermophiles

In a pilot analysis of the proteome of *Methanocaldococcus jannaschii* (Giometti *et al.*, 2001), significant changes in the abundance of a subset of predominant proteins has been observed in response to culture conditions and phase of the growth curve (exponential compared with stationary). Interestingly, several proteins were found to exist in multiple forms with different isoelectric points and molecular weights (see discussion below, *Sulfolobus*); the relative abundance of these protein variants appeared to change with growth conditions. Although variation due to sample treatment cannot be ruled out, these data might reflect post-translational modifications. Although the identity of the modifications remains to be identified, it is tempting to assume that this reflects a means of functional regulation at protein level. In a subsequent study (Giometti *et al.*, 2002), 170 of the most abundant proteins have been identified in total lysates of *M. jannaschii*. To optimize the number of proteins detected, two different protein stains (Coomassie Blue R250 or silver nitrate) and two different first-dimensional separation methods (isoelectric focusing or non-equilibrium pH gradient electrophoresis) were used. Again, evidence of post-translational modification of numerous *M. jannaschii* proteins has been reported, as well as indications of incomplete dissociation of protein-protein complexes.

Lim *et al.* (2003) have recently performed a comparative analysis of protein identification for a total of 162 protein spots separated by two-dimensional gel electrophoresis from *M. jannaschii* and *Pyrococcus furiosus*, using MALDI-TOF peptide mass mapping and LC-MS/MS. 100% of the gel spots analyzed were successfully matched to the predicted proteins in the two corresponding open reading frame databases by LC-MS/MS while 97% of them were identified by MALDI-TOF mapping. The high success rate from the peptide mass mapping partly correlated with careful sample treatment (desalting/concentrating), but also with optimization of the search parameters, e.g. by incorporating amino acid sequence modifications into database searches. The obtained high sequence coverage in combination with digestion with several proteolytic enzymatic of different specificity is proposed as a method for future analysis of post-translational modifications (Lim *et al.*, 2003).

High-throughput Multidimensional Protein Identification Technology based on microcapillary LC/LC/MS/MS has recently been used to identify 963 proteins of the proteome of *M. jannaschii*, corresponding to as much as 54% of the whole genome (Zhu *et al.*, 2004). Almost half of the identified proteins have an unknown function, being annotated either as "conserved hypothetical" or as "hypothetical" proteins. The majority of the proteins predicted to be involved in distinct metabolic pathways were among the identified proteins. In addition, predicted intein peptides were detected, as well as peptides created by protein splicing. High peptide number, spectrum count, and sequence coverage have been used as indicators of high expression levels (Zhu *et al.*, 2004).

As a means to release all experimental data of proteomics studies, to allow comparison of different proteomics analyses, and to enable integration with other data sets in a systems biology setting, the publicly available GELBANK database has recently been developed for the display of protein profiles generated by two-dimensional gel electrophoresis (<http://gelbank.anl.gov>). GELBANK is a database of two-dimensional gel electrophoresis (2DE) gel patterns of proteomes from organisms with known genome information, with relevant technical information. It includes the completed, mostly microbial proteomes available from the National Center for Biotechnology Information (Babnigg and Giometti, 2004).

Proteomics of *Sulfolobus*

In two recent studies, the proteome of *Sulfolobus solfataricus* has been analysed (Snijders *et al.*, 2005a, 2006). Proteins corresponding to 349 ORFs were separated and identified using 2DE followed by LC-ESI-MS/MS and database searching (Mass Spectrometry protein sequence DataBase (MSDB); <http://csc-fserve.hh.med.ic.ac.uk/msdb.html>). Moreover, it was shown that ^{15}N and ^{13}C metabolic labelling for peptide quantification in Archaea has significant advantages compared to traditional gel-based quantification methods (Snijders *et al.*, 2005a, 2005b).

Cultivation and metabolic labelling

S. solfataricus is generally grown aerobically on defined medium in 250-ml flasks. Each flask contains 50 ml basic medium, 25 μl Wolfe's vitamins and glucose as the carbon source with a final concentration of 0.3–0.4% (w/v) (Table 9.3). Cells are routinely inoculated at an optical density of 0.2 (OD_{530}). Cells are harvested at the appropriate phase of the growth curve, and washed twice with basic medium and once with 10 mM Tris/HCl buffer (pH 7.0). The wash steps are necessary to remove salts and contaminants that may interfere with the 2DE protocol, in particular with the iso-electric focussing. After this, cells can be stored at -20°C until further processed. During the process described below, considerable care was taken to ensure that culture-to-culture variation is minimal, and cultures should be prepared in at least triplicate; to further enhance the reproducibility, cultivation in well-controlled fermenters (batch, continuous culture) is currently being developed.

In the case of the ^{15}N or ^{13}C labelling experiments, the same protocol for cell growth may be used, but the nitrogen and carbon sources are replaced by $(^{15}\text{NH}_4)_2\text{SO}_4$ (Sigma) or ^{13}C -glucose, respectively. The isotope-labelled nutrients are readily incorporated into cell material and have no observable effect on the growth characteristics of *S. solfataricus* (Snijders, unpublished data). In order to ensure full incorporation of the isotope label into all proteins, at least 8 doubling times are required (approximately 48 h). Once full labelling is achieved, the cells can be stored as glycerol stocks until further use.

For a quantitative experiment based on metabolic labelling, two separate growth experiments are set up. In one case, the organism is grown in normal “unlabelled” medium. In the second case, the same organism is grown in the ^{15}N (or ^{13}C)-containing medium. In this way a “light” and a “heavy” proteome are created. Therefore, batch cultures are grown in parallel starting at an OD_{530} of 0.2 with either ^{14}N - or ^{15}N -ammonium sulphate. When the OD_{530} reaches a value of 0.5, the differentially labelled cultures are mixed. To ensure that equal amounts of biomass are mixed, slight corrections in volume should be made to correct for slight deviations with respect to the cultures’ optical densities. Subsequently, cell harvest, preparation of cell extracts, 2-DE and protein identification are performed exactly the same way for the labelled/unlabelled cells as for the unlabelled cells (Snijders *et al.*, 2006). In this case, peptide identification and quantification occur at the last (MS) stage of the protocol as described below.

Recently, a useful extension of the stable isotope labelling approach was introduced (Snijders *et al.*, 2005a). In this study, dual labelling with ^{15}N -ammonium sulphate and ^{13}C -glucose was used for both quantification and identification of peptides in *S. solfataricus*. This variation of the metabolic labelling method is not discussed here further, but clearly offers some strong advantages as three phenotypes can be simultaneously examined.

Cell lysis

The -20°C frozen cells are thawed and immediately resuspended in 1.5 ml of 10 mM Tris-HCl buffer (pH 7.0), and 25 μl of a protease-inhibitor cocktail (Protease Inhibitor Cocktail for use with bacterial cell extracts, Sigma) is added to the cell suspension. Cells are disrupted by sonication for 10 min on ice (Soniprep 150, Sanyo). Insoluble cell material is removed by centrifugation (15 min; $5000 \times g$; 4°C). The protein concentration of the supernatant is then determined using the Bradford Protein Assay (Sigma). At this stage, the supernatant can be stored at -80°C .

2-D electrophoresis, protein visualization and image analysis

Below, the details are provided for (i) a recently developed protocol of 2-DE of *S. solfataricus* cell-free extracts (Snijders *et al.*, 2006), and (ii) a summary of a study in which a systematic optimisation of some critical steps in 2-DE of *S. solfataricus* lysates has been evaluated (Barry *et al.*, 2003).

S. solfataricus – Procedure for 2-DE

- A sample mix is prepared by mixing the cell-free extract (prepared as described above) with a Rehydration Buffer to yield a final concentration of 50 mM DTT, 8 M Urea, 2% CHAPS (Sigma), 0.2% (w/v) Pharmalyte ampholytes (e.g. pH 3–10) (Fluka) and a trace of Bromophenol Blue.
- Each IPG strip (e.g. pH 3–10; Bio-Rad) is rehydrated overnight with 300 μl (400 μg of protein) of the sample mix.

- Isoelectric focussing (IEF) is performed using a 3-step protocol at 20°C using a Protean IEF cell (Bio-Rad).
- Subsequently, the voltage is linearly ramped to 250 V over 30 min to desalt the strips.
- Next, the voltage is linearly ramped to 1000 V over 2.5 h.
- Finally, the voltage is rapidly ramped to 10 000 V for 40 000 V* h to complete the focussing. At this stage, the strips can be stored overnight at -20°C.
- Focussed strips are incubated for 15 min in a solution containing 6 M urea, 2% SDS, 0.375 M Tris-HCl (pH 8.8), 20% glycerol, and 2% (w/v) DTT.
- The solution is discarded and the strips are incubated in a solution containing 6 M urea, 2% SDS, 0.375 M Tris-HCl (pH 8.8), 20% glycerol, and 4% iodoacetamide.
- After equilibration, proteins are separated in the second dimension using SDS-PAGE performed using a Protean II Multicell (Bio-Rad) apparatus on 10% T (concentration total polymer: % (w/v) acrylamide + % (w/v) N, N'-methylenebisacrylamide), 2.6% C gels (concentration crosslinker: % (w/v) N, N'-methylenebisacrylamide; gel dimensions: 17 cm × 17 cm × 1 mm).
- Electrophoresis is carried out with a constant current of 16 mA/gel for 30 min.
- Subsequently the current is increased to 24 mA/gel for another 7 h.
- Gels are stained using Coomassie Brilliant Blue G250 (Sigma).
- Gels are scanned using a GS-800 densitometer (Bio-Rad) at 100 microns resolution.

All spot detection and quantification is performed with PDQUEST 7.1.0 (Bio-Rad). With this method, Snijders *et al.* (2006) created a 2D reference map on which approximately 500 spots are visualized. In the case of the metabolic labelling experiments, the gel image was matched to the reference map and protein spots of interest were selected for MS analysis and quantitation.

Using sonicated cell extracts of *S. solfataricus* cultures, Barry *et al.* (2003) have performed an extensive comparative study on 2D sample application; it should be noted that in this study the obtained lysates were not subjected to any subsequent purification or fractionation (e.g. removal of membrane fragments and lipids by centrifugation; see below) because the goal was to obtain a single gel system to resolve the complete proteome, including both soluble and membrane proteins. When comparing three different methods for applying the protein samples on the IEF strips (cup-loading, active rehydration and passive rehydration) it was concluded that for basic proteins (using IPG 6-11 strips (Amersham-Pharmacia) on the Protean IEF Cell (Bio-Rad; <http://www.bio-rad.com>) and its IEF tray) sample application by cup-loading is by far superior over both rehydration methods (greatest number of detectable spots, best gel-to-gel reproducibility, lowest spot quantity variations); in the case of acidic proteins (using IPG 4-7 strips) active rehydration and cup-loading appeared to give the best results (Barry *et al.*, 2003). In this study, gels were stained with SYPRO Ruby (non-covalent fluorescent dye; Molecular Probes; <http://probes.invitrogen.com>), and analysis was performed with

the PDQuest package (Bio-Rad). Experimental details are summarized in Table 9.3.

Protein isolation and identification by MS

Spots of interest are manually excised from the CBB-stained 2D gels, and destained with 200 mM ammonium bicarbonate with 40% acetonitrile at 37°C (twice for 30 min). The gel pieces are incubated overnight in trypsin solution (0.4 µg trypsin (Sigma) in 50 µl of 40 mM ammonium bicarbonate in 9% acetonitrile). Subsequently, peptides are extracted in three subsequent extraction steps using 5 µl of 25 mM ammonium bicarbonate (10 min, room temperature), 30 µl acetonitrile (15 min, 37°C), 50 µl of 5% formic acid (15 min, 37°C) and finally with 30 µl acetonitrile (15 min, 37°C). Using this three-step protocol, the peptide recovery was significantly higher compared to simplified extraction methods. Therefore, the quality of MS spectra was improved and protein sequence coverage was significantly increased (data not shown). All extracts are pooled and dried in a vacuum centrifuge, then stored at -20°C.

The lyophilized peptide mixture is resuspended in 10 µl 0.1% formic acid in 3% acetonitrile. This mixture is separated on a PepMap C-18 RP capillary column (LC Packings), and eluted in a 30-min gradient *via* a LC Packings Ultimate nanoLC directly onto the mass spectrometer. Peptides are analysed using a QStarXL electrospray ionization quadrupole time-of-flight tandem MS (ESI qQ-TOF; Applied Biosystems; <http://www.appliedbiosystems.com>). The data acquisition on the MS is performed in the positive ion mode using Information Dependent Acquisition (IDA). Peptides with charge states 2 and 3 are selected for tandem mass spectrometry. IDA data were submitted to Mascot for database searching in a sequence query type of search (www.matrixscience.com). The settings are as follows: peptide tolerance 2.0 Da; MS/MS tolerance 0.8 Da; carbamidomethyl modification of cysteine is set as a fixed modification; methionine oxidation is set as a variable modification: maximal 1 missed cleavage site by trypsin was allowed. The search is performed against the Mass Spectrometry protein sequence DataBase (MSDB). MOWSE scores greater than 50 are considered significant (Snijders *et al.*, 2006).

Peptide quantification

Heavy and light versions of a protein or peptide have the same physicochemical properties. Therefore, no distinction can be made between the two versions during cell growth, protein extraction and separation. Only in the MS stage of the protocol, a difference can be observed. Labelled and unlabelled peptides appear as doublets in the MS spectrum. The difference in mass between the unlabelled and labelled peaks corresponds to the number of nitrogen (or carbon) atoms present in the peptide. In the metabolic labelling experiments, peptide identification of the light, ¹⁴N version of the peptide is performed as described above. The heavy ¹⁵N version of the peptide is identified by changing the isotope abundance of ¹⁵N nitrogen to 100% in the Analyst data dictionary. Next, the peak areas of both versions of the same peptide are integrated

over time using LC-MS reconstruction tool in the Analyst software (Applied Biosystems). In addition, an extracted ion chromatogram (XIC) is constructed for each peptide. The XIC is an ion chromatogram that shows the intensity values of a single mass (peptide) over a range of scans. This tool is used to check for chromatographic shifts between heavy and light versions of the same peptide. Generally, metabolic labelling with either ^{15}N or ^{13}C does not cause chromatographic shifts.

Protein identification by MS

Using the described proteomics method, 325 unique soluble *S. solfataricus* proteins have been identified in 255 spots from the 2D reference map. The MS technique employed is very sensitive, with reliable MS identification even of the faintest spots visualized by CBB staining. In many cases, multiple proteins per spot were found and preliminary use of pH 4-7 and pH 5-8 zoom gels still usually yielded multiple proteins per Coomassie-stained spot (Snijders, personal communication). Significant MOWSE scores (>51) were found for all 255 spots analysed. Generally, one peptide (intact mass and MS/MS ion spectrum) was sufficient for confident identification of *S. solfataricus* proteins against the MSDB. In most cases, however, multiple peptides of the same protein were found. On an average, the sequence coverage was 30%. Complete sequence coverage was never achieved. There was no relation between the sequence coverage and the protein's size. MOWSE scores greater than 800 were only achieved for proteins larger than 48 kDa.

Apart from looking at abundance levels, proteomics studies are currently ongoing that focus on the attenuation of enzyme activity by protein post-translational modification. Another important observation is that a number of proteins occur in more than 1 spot. Interestingly, this was true for a large number of proteins involved in the TCA cycle (e.g. 2-oxoglutarate oxidoreductase (SSO2815) was found in eight different spots). There are a number of explanations for this, including (1) the protein might exist in multiple forms in the cell, e.g. post-translationally modified *versus* non modified, (2) The protein was modified during protein extraction or during 2-DE (e.g. methionine oxidation), and (3) the protein does not resolve well on the gel and therefore "smears" out over a large pH or mass range. Future proteomics studies are expected to provide additional clues that will reveal the details of quantitative and qualitative modulations of proteins as means of regulating the metabolism of *S. solfataricus*.

◆◆◆◆◆ CONCLUDING REMARKS

The intention of this overview is to describe ongoing developments at the level of functional genomics of archaeal and bacterial thermophiles, with particular emphasis on Archaea of the genus *Sulfolobus*. Obviously, *Sulfolobus* would not have gained the status of "model archaeon" without a long history of classical studies (physiology, biochemistry and

molecular genetics) that have been performed over the last three decades in numerous research laboratories. This has resulted in considerable insight in fundamental principles of the archaeal cell, including central metabolic pathways (Snijders *et al.*, 2006, Verhees *et al.*, 2003), replication and cell cycle (Lundgren *et al.*, 2004; Margolin and Bernander, 2004; Robinson *et al.*, 2004), transcription (Bell and Jackson, 2001; Thomm, 1996) and translation (Tumbala *et al.*, 1999). As such, classical research has provided the basis on which modern genomics-related approaches have been built (comparative, functional and structural genomics). At present, important developments in biochemical and molecular genetics analyses are under way, allowing novel approaches for analysis *in vitro* (heterologous expression, directed and random mutagenesis, study of protein/protein and protein/DNA interactions), and *in vivo* (chromosomal knockouts, *in trans* overexpression, phenotype characterization) (current progress reviewed by Allers and Mevarech, 2005; Baliga *et al.*, 2004).

Recent developments of transcriptomics and proteomics tools for thermophiles in general, and *Sulfolobus* in particular have been reviewed. It should be noted, however, that developments in thermophile genomics proceed at an impressive pace: additional hyperthermophile genomes are released on a regular basis, comparative tools become more sophisticated and functional predictions become more reliable, DNA microarrays are available, and recent breakthroughs illustrate the role of mass spectrometry-based proteomics and metabolomics as an indispensable tool for molecular and cellular biology, and for the emerging field of systems biology.

Now the stage is set for yet another challenge in the study of thermophiles, i.e. the integration of classical and modern technologies to acquire knowledge in the functioning of metabolic networks and of the regulatory circuits of these systems. Many breakthroughs are expected in the near future, resulting in a gain of insight into the evolution and functioning of these intriguing thermophile systems.

Acknowledgements

This work was supported by a grant from the European Union in the framework of the SCREEN project (contract QLK3-CT-2000-00649), and by the Netherlands Organisation of Scientific Research (NWO) BioMolecular Informatics Programme, grant 050.50.206. We thank the EPSRC for funding, and Advanced Fellowship support. In addition, we acknowledge The Swedish Research Council as well as The Swedish Foundation for Strategic Research.

References

- Allers, T. and Mevarech, M. (2005). Archaeal genetics – the third way. *Nature Rev. Genet.* **6**, 58–73.

- Andersson, A., Bernander, R. and Nilsson, P. (2005). Dual-genome primer design for construction of DNA microarrays. *Bioinformatics* **21**, 325–332.
- Babnigg, G. and Giometti, C. S. (2004). GELBANK: a database of annotated two-dimensional gel electrophoresis patterns of biological systems with completed genomes. *Nucleic Acids Res. (Database issue)*, **32**, D582–D585.
- Baliga, N. S., Bjork, S. J., Bonneau, R., Pan, M., Iloanusi, C., Kottemann, M. C., Hood, L. and DiRuggiero, J. (2004). Systems level insights into the stress response to UV radiation in the halophilic archaeon *Halobacterium* NRC-1. *Genome Res.* **14**, 1025–1035.
- Barry, R. C., Alsaker, B. L., Robison-Cox, J. F. and Dratz, E. A. (2003). Quantitative evaluation of sample application methods for semipreparative separations of basic proteins by two-dimensional gel electrophoresis. *Electrophoresis* **24**, 3390–3404.
- Bell, S. D. and Jackson, S. P. (2001). Mechanism and regulation of transcription in archaea. *Curr. Opin. Microbiol.* **4**, 208–213.
- Bernander, R. and Poplawski, A. (1997). Cell cycle characteristics of the thermophilic archaea. *J. Bacteriol.* **179**, 4963–4969.
- Blöchl, E., Rachel, R., Burggraf, S., Hafenbradl, D., Jannasch, H. W. and Stetter, K. O. (1997). *Pyrolobus fumarii*, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113°C. *Extremophiles* **1**, 14–21.
- Brinkman, A. B., Bell, S. D., Lebbink, R. J., de Vos, W. M. and van der Oost, J. (2002). The *Sulfolobus solfataricus* Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability. *J. Biol. Chem.* **277**, 29537–29549.
- Brock, T. D. and Freeze, H. (1969). *Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile. *J. Bacteriol.* **98**, 289–297.
- Brock, T. D., Brock, K. M., Belly, R. T. and Weiss, R. L. (1972). *Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch. Mikrobiol.* **84**, 54–68.
- Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. and Venter, J. C. (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**, 1058–1073.
- Cambillau, C. and Claverie, J. M. (2000). Structural and genomic correlates of hyperthermostability. *J. Biol. Chem.* **275**, 32383–32386.
- Chen, L., Brugger, K., Skovgaard, M., Redder, P., She, Q., Torarinsson, E., Greve, B., Awayez, M., Zibat, A., Klenk, H. P. and Garrett, R. A. (2005). The genome of *Sulfolobus acidocaldarius*, a model organism of the Crenarchaeota. *J. Bacteriol.* **187**, 4992–4999.
- Chhabra, S. R., Shockley, K. R., Connors, S. B., Scott, K. L., Wolfinger, R. D. and Kelly, R. M. (2003). Carbohydrate-induced differential gene expression patterns in the hyperthermophilic bacterium *Thermotoga maritima*. *J. Biol. Chem.* **278**, 7540–7552.
- Cleveland, W. and Devlin, S. (1988). Locally weighted regression – an approach to regression-analysis by local fitting. *J. Amer. Statist. Assoc.* **83**, 596–610.
- Cohen, G. N., Barbe, V., Flament, D., Galperin, M., Heilig, R., Lecompte, O., Poch, O., Prieur, D., Querellou, J., Ripp, R., Thierry, J. C., van der Oost, J.,

- Weissenbach, J., Zivanovic, Y. and Forterre, P. (2003). An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*. *Mol. Microbiol.* **47**, 1495–1512.
- Cui, X. and Churchill, G. A. (2003). Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol.* **4**, 210.
- Dudoit, S., Gentleman, R. C. and Quackenbush, J. (2003). Open source software for the analysis of microarray data. *Biotechniques* March 2003, Suppl. 45–51.
- Ettema, T., van der Oost, J. and Huynen, M. (2001). Modularity in the gain and loss of genes: applications for function prediction. *Trends Genet.* **17**, 485–487.
- Ettema, T., de Vos, W. M. and van der Oost, J. (2005). Discovering new biology by *in silico* archaeology. *Nature Rev. Microbiol.*, **3**, 859–869.
- Frickey, T. and Lupas, A. N. (2004). PhyloGenie: automated phylome generation and analysis. *Nucleic Acids Res.* **32**, 5231–5238.
- Fukui, T., Atomi, H., Kanai, T., Matsumi, R., Fujiwara, S. and Imanaka, T. (2005). Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. *Genome Res.* **15**, 352–363.
- Giometti, C. S., Reich, C. I., Tollaksen, S. L., Babnigg, G., Lim, H., Yates, J. R. 3rd and Olsen, G. J. (2001). Structural modifications of *Methanocaldococcus jannaschii* flagellin proteins revealed by proteome analysis. *Proteomics* **1**, 1033–1042.
- Giometti, C. S., Reich, C., Tollaksen, S., Babnigg, G., Lim, H., Zhu, W., Yates, J. and Olsen, G. (2002). Global analysis of a “simple” proteome: *Methanocaldococcus jannaschii*. *J. Chromatogr. B* **782**, 227–243.
- Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H. and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnol.* **17**, 994–999.
- Huber, R., Huber, H. and Stetter, K. O. (2000). Towards the ecology of hyperthermophiles: biotopes, new isolation strategies and novel metabolic properties. *FEMS Microbiol. Rev.* **24**, 615–623.
- Johnson, M. R., Montero, C. I., Conners, S. B., Shockley, K. R., Bridger, S. L. and Kelly, R. M. (2005). Population density-dependent regulation of exopolysaccharide formation in the hyperthermophilic bacterium *Thermotoga maritima*. *Mol. Microbiol.* **55**, 664–674.
- Kashefi, K. and Lovley, D. R. (2003). Extending the upper temperature limit for life. *Science* **301**, 934.
- Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y., Sakai, M., Ogura, K., Otsuka, R., Nakazawa, H., Takamiya, M., Ohfuku, Y., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K. and Kikuchi, H. (1998). Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3. *DNA Res.* **5**, 55–76.
- Kawashima, T., Amano, N., Koike, H., Makino, S., Higuchi, S., Kawashima-Ohya, Y., Watanabe, K., Yamazaki, M., Kanehori, K., Kawamoto, T., Nunoshiro, T., Yamamoto, Y., Aramaki, H., Makino, K. and Suzuki, M. (2000). Archaeal adaptation to higher temperatures revealed by genomic sequence of *Thermoplasma volcanium*. *Proc. Natl. Acad. Sci. USA* **97**, 14257–14262.
- Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H., Hosoyama, A., Fukui, S.,

- Nagai, Y., Nishijima, K., Otsuka, R., Nakazawa, H., Takamiya, M., Kato, Y., Yoshizawa, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Masuda, S., Yanagii, M., Nishimura, M., Yamagishi, A., Oshima, T. and Kikuchi, H. (2001). Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain 7. *DNA Res.* **8**, 123–140.
- Koonin, E. V., Makarova, K. S. and Aravind, L. (2001). Horizontal gene transfer in prokaryotes: quantification and classification. *Annu. Rev. Microbiol.* **55**, 709–742.
- Lee, M. L., Kuo, F. C., Whitmore, G. A. and Sklar, J. (2000). Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc. Natl. Acad. Sci. USA* **97**, 9834–9839.
- Leung, Y. F. and Cavalieri, D. (2003). Fundamentals of cDNA microarray data analysis. *Trends Genet.* **19**, 649–659.
- Lim, H., Eng, J., Yates, J. R. 3rd, Tollaksen, S. L., Giometti, C. S., Holden, J. F., Adams, M. W., Reich, C. I., Olsen, G. J. and Hays, L. G. (2003). Identification of 2D-gel proteins: a comparison of MALDI/TOF peptide mass mapping to mu LC-ESI tandem mass spectrometry. *J. Amer. Soc. Mass Spectrom.* **14**, 957–970.
- Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., Yates and J. R. 3rd. (1999). Direct analysis of protein complexes using mass spectrometry. *Nature Biotechnol.* **17**, 676–682.
- Lönnstedt, I. and Speed, T. (2002). Replicated microarray data. *Statistica Sinica* **12**, 31–46.
- Lundgren, M., Andersson, A., Chen, L., Nilsson, P. and Bernander, R. (2004). Three replication origins in *Sulfolobus* species: synchronous initiation of chromosome replication and asynchronous termination. *Proc. Natl. Acad. Sci. USA* **101**, 7046–7051.
- Makarova, K. S. and Koonin, E. V. (2003). Comparative genomics of archaea: how much have we learned in six years, and what's next? *Genome Biol.* **4**, 115.
- Margolin, W. and Bernander, R. (2004). How do prokaryotic cells cycle? *Curr. Biol.* **14**, R768–R770.
- McDonald, W. H., Yates and J. R. 3rd. (2002). Shotgun proteomics and biomarker discovery. *Disease Markers* **18**, 99–105.
- Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Nelson, W. C., Ketchum, K. A., McDonald, L., Utterback, T. R., Malek, J. A., Linher, K. D., Garrett, M. M., Stewart, A. M., Cotton, M. D., Pratt, M. S., Phillips, C. A., Richardson, D., Heidelberg, J., Sutton, G. G., Fleischmann, R. D., Eisen, J. A., White, O., Salzberg, S. L., Smith, H. O., Venter, J. C. and Fraser, C. M. (1999). Evidence for lateral gene transfer between archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323–329.
- Nguyen, T. N., Ejaz, A. D., Brancieri, M. A., Mikula, A. M., Nelson, K. E., Gill, S. R. and Noll, K. M. (2004). Whole-genome expression profiling of *Thermotoga maritima* in response to growth on sugars in a chemostat. *J. Bacteriol.* **186**, 4824–4828.
- Oda, Y., Huang, K., Cross, F. R., Cowburn, D. and Chait, B. T. (1999). Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. USA* **96**, 6591–6596.
- Pace, N. R. (1997). A molecular view of microbial diversity and the biosphere. *Science* **276**, 734–740.

- Qin, L. X. and Kerr, K. F. (2004). Empirical evaluation of data transformations and ranking statistics for microarray analysis. *Nucleic Acids Res.* **32**, 5471–5479.
- Rivera, M. C. and Lake, J. A. (2004). The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* **431**, 152–155.
- Robb, F. T., Maeder, D. L., Brown, J. R., DiRuggiero, J., Stump, M. D., Yeh, R. K., Weiss, R. B. and Dunn, D. M. (2001). Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: implications for physiology and enzymology. *Meth. Enzymol.* **330**, 134–157.
- Robinson, N. P., Dionne, I., Lundgren, M., Marsh, V. L., Bernander, R. and Bell, S. D. (2004). Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell* **116**, 25–38.
- Rothschild, L. J. and Mancinelli, R. L. (2001). Life in extreme environments. *Nature* **409**, 1092–1101.
- Ruepp, A., Graml, W., Santos-Martinez, M. L., Koretke, K. K., Volker, C., Mewes, H. W., Frishman, D., Stocker, S., Lupas, A. N. and Baumeister, W. (2000). The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature* **407**, 508–513.
- Schena, M. (1996). Genome analysis with gene expression microarrays. *Bioessays* **18**, 427–431.
- Schut, G. J., Zhou, J. and Adams, M. 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.
- Schut, G. J., Brehm, S. D., Datta, S. and Adams, M. 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.
- She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C., Clausen, I. G., Curtis, B. A., de Moors, A., Erauso, G., Fletcher, C., Gordon, P. M., Heikamp-de Jong, I., Jeffries, A. C., Kozera, C. J., Medina, N., Peng, X., Thi-Ngoc, H. P., Redder, P., Schenk, M. E., Theriault, C., Tolstrup, N., Charlebois, R. L., Doolittle, W. F., Duguet, M., Gaasterland, T., Garrett, R. A., Ragan, M. A., Sensen, C. W. and van der Oost, J. (2001). The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. USA* **98**, 7835–7840.
- Shivvers, D. W. and Brock, T. D. (1973). Oxidation of elemental sulfur by *Sulfolobus acidocaldarius*. *J. Bacteriol.* **114**, 706–710.
- Shockley, K. R., Ward, D. E., Chhabra, S. R., Connors, S. B., Montero, C. I. and Kelly, R. M. (2003). Heat shock response by the hyperthermophilic archaeon *Pyrococcus furiosus*. *Appl. Environ. Microbiol.* **69**, 2365–2371.
- Snijders, A. P. L., de Vos, M. G. J. and Wright, P. C. (2005a). Novel approach for peptide quantitation and sequencing based on ¹⁵N and ¹³C metabolic labeling. *J. Proteome Res.* **4**, 578–585.
- Snijders, A. P. L., de Vos, M. G., de Koning, B. and Wright, P. C. (2005b). A fast method for quantitative proteomics based on a combination between two-dimensional electrophoresis and ¹⁵N-metabolic labelling. *Electrophoresis* **26**, 3191–3199.
- Snijders, A. P. L., Walther, J., Peter, S., Kinnman, I., de Vos, M. G. J., van de Werken, H. J. G., Brouns, S. J. J., van der Oost, J. and Wright, P. C. (2006). Reconstruction of central carbon metabolism in *Sulfolobus solfataricus* using a two-dimensional gel electrophoresis map, stable isotope

- labelling and DNA microarray analysis. *Proteomics*, doi: 10.1002/pmic.200402070.
- Stetter, K. O. (1996). Hyperthermophilic prokaryotes. *FEMS Microbiol. Rev.* **18**, 149–158.
- Stetter, K. O. (1999). Extremophiles and their adaptation to hot environments. *FEBS Lett.* **452**, 22–25.
- Thomm, M. (1996). Archaeal transcription factors and their role in transcription initiation. *FEMS Microbiol. Rev.* **18**, 159–171.
- Tumbula, D., Vothknecht, U. C., Kim, H. S., Ibba, M., Min, B., Li, T., Pelaschier, J., Stathopoulos, C., Becker, H. and Söll, D. (1999). Archaeal aminoacyl-tRNA synthesis: diversity replaces dogma. *Genetics* **152**, 1269–1276.
- Verhees, C. H., Kengen, S. W., Tuininga, J. E., Schut, G. J., Adams, M. W., de Vos, W. M. and van der Oost, J. (2003). The unique features of glycolytic pathways in archaea. *Biochem. J.* **375**, 231–246.
- Washburn, M. P. (2004). Utilisation of proteomics datasets generated via multidimensional protein identification technology (MudPIT). *Brief. Funct. Genomic. Proteomic* **3**, 280–286.
- Washburn, M. P., Ulaszek, R., Deciu, C., Schieltz, D. M., Yates and J. R. 3rd. (2002). Analysis of quantitative proteomic data generated via multidimensional protein identification technology. *Anal. Chem.* **74**, 1650–1657.
- Waters, E., Hohn, M. J., Ahel, I., Graham, D. E., Adams, M. D., Barnstead, M., Beeson, K. Y., Bibbs, L., Bolanos, R., Keller, M., Kretz, K., Lin, X., Mathur, E., Ni, J., Podar, M., Richardson, T., Sutton, G. G., Simon, M., Söll, D., Stetter, K. O., Short, J. M. and Noordevier, M. (2003). The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc. Natl. Acad. Sci. USA* **100**, 12984–12988.
- Weinberg, M. P., Schut, G. J., Brehm, S., Datta, S. and Adams, M. 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.
- Woese, C. R. (1987). Bacterial evolution. *Microbiol. Rev.* **51**, 221–271.
- Yang, Y. H. and Speed, T. (2002). Design issues for cDNA microarray experiments. *Nature Rev. Genet.* **3**, 579–588.
- Zhou, H., Ranish, J. A., Watts, J. D. and Aebersold, R. (2002). Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. *Nature Biotechnol.* **20**, 512–515.
- Zhu, W., Reich, C. I., Olsen, G. J., Giometti, C. S., Yates and J. R. 3rd. (2004). Shotgun proteomics of *Methanocaldococcus jannaschii* and insights into methanogenesis. *J. Proteome Res.* **3**, 538–548.
- Zillig, W., Stetter, K. O., Wunderl, S., Schulz, W., Priess, H. and Scholz, I. (1980). The *Sulfolobus-Caldariella* group, taxonomy on the basis of the structure of DNA-dependent RNA polymerases. *Arch. Microbiol.* **125**, 259–269.