TECHNISCHE UNIVERSITÄT MÜNCHEN
Max-Planck-Institut für Biochemie
Abteilung für Molekulare Strukturbiologie

Proteomics and transcriptomics analysis of *Thermoplasma acidophilum*

Na Sun

Vollständiger Abdruck der von der Fakultät für Chemie
der Technischen Universität München
zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften
genehmigten Dissertation.

Vorsitzender:		Univ.-Prof. Dr. S. Weinkauf
Prüfer der Dissertation:	1.	Hon.-Prof. Dr. W. Baumeister
	2.	Univ.-Prof. Dr. J. Buchner

Die Dissertation wurde am 18.03.2009 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 03.06.2009 angenommen.

Table of contents

Zusammenfassung **1**
Abstract **4**
1 Introduction **7**
1.1 *Thermoplasma acidophilum* 7
1.2 Transcriptomics and proteomics 7
1.2.1 Transcriptomics 8
1.2.2 Proteomics 10
2 Aim of the study **16**
3 Materials and methods **17**
3.1 Project 1: 2DE based proteomics analysis of *T. acidophilum* 17
3.2 Project 2: Protein complexes of *T. acidophilum* 21
3.3 Project 3: Quantitative proteomics and transcriptomics analysis of *T. acidophilum* cultured under aerobic and anaerobic conditions 25
4 Results and discussion **31**
4.1 Project 1: 2DE based proteomics analysis of *T. acidophilum* 31
4.1.1 Results 31
4.1.2 Discussion 41
4.2 Project 2: Protein complexes of *T. acidophilum* 44
4.2.1 Results 44
4.2.2 Discussion 58
4.3 Project 3: Quantitative proteomics and transcriptomics analysis of *T. acidophilum* cultured under aerobic and anaerobic conditions 62
4.3.1 Results 62
4.3.2 Discussion 92
5 Conclusions and perspectives **102**
6 Abbreviations **105**
7 References **107**
8 Acknowledgements **118**
9 Appendix **120**

Zusammenfassung

Aufgrund seiner geringen zellulären Komplexität, des Fehlens einer Zellwand und des bereits sequenzierten und kleinen Genoms ist *Thermoplasma acidophilum* ein attraktiver Modellorganismus für das *'Visual Proteomics'* Projekt, dessen Ziel es ist, einen umfassenden Atlas der makromolekularen Proteinkomplexe mittels Kryo-Elektronentomographie (Kryo-ET) und Mustererkennungsmethoden zu erstellen. *T. acidophilum* ist ein thermoacidophiles Archaeon, das optimal um 59 °C und bei pH 1-2 wachsen kann. Es kann sowohl unter (mikro)aeroben Bedingungen, dabei dient Sauerstoff als terminaler Elektronen-Akzeptor, als auch unter anaeroben Bedingungen durch Schwefelatmung wachsen. Diese physiologische Vielseitigkeit, gepaart mit der zellulären Einfachheit bietet eine Plattform, die Unterschiede von zellulären Protein- und mRNA-Konzentrationen mit Proteomik- und Transkriptomik-Methoden zu identifizieren und zu quantifizieren. Die Ergebnisse werden ein wichitiger Meilenstein für die Validierung der Kryo-ET-Daten von *T. acidophilum* sein.

In einem ersten Schritt wurde zweidimensionale Gelelektrophorese (2DE) in Kombination mit Matrix-unterstützter Laser Desorption/Ionisation Time-of-Flight Massenspektrometrie (MALDI-TOF-MS) verwendet, um einen globalen Überblick über das zytosolische Proteom von *T. acidophilum*, kultiviert unter aeroben Bedingungen, zu erhalten. Mehr als 900 Spots wurden mit der 2DE aufgelöst, von denen 271 Proteine identifiziert werden konnten. Coomassie-Blau-Färbung der 2D-Gele erlaubte eine ungefähre Quantifizierung der Proteinspots. Daraus konnten vorläufige Schlussfolgerungen über die am häufigsten exprimierten Proteine gezogen werden. Parallel dazu, mit einer stärkeren Fokussierung auf das *'Visual Proteomics'* Projekt, wurde eine Glycerol-Dichtegradientenultrazentrifugation in Kombination mit 2DE-MALDI-TOF-MS-Analyse eingesetzt, um Untereinheiten makromolekularer Komplexe aus *T. acidophilum* zu identifizieren. In den hochmolekularen Fraktionen des Dichtegradienten wurden neben den bereits bekannten (ribosomale Untereinheiten, Translationsinitiationsfaktor EIF-6, Elongationsfaktor 1 und DNA-abhängige RNA-Polymerase) zehn weitere Proteine als potentielle Bausteine von Komplexen neu identifiziert. Aufgrund des limitierten dynamischen Bereichs der 2DE und des eingeschränkten Identifizierungsvermögens der MALDI-TOF-MS, konnte nur eine begrenzte Anzahl von Proteinen entdeckt werden.

Um die Identifikationsrate der Proteine zu erhöhen und damit die Inventarisierung der Protein-Komplexe zu verfeinern, wurde der zytosolische Zellextrakt von *T. acidophilum* mittels einer Kombination aus Molekularsieb-Chromatographie (MSC) und Flüssig-chromatographie-Tandem-Massenspektrometrie (LC-MS/MS), die im Vergleich zu MALDI-TOF-MS einen höheren Dynamikumfang aufweist, aufgetrennt und analysiert. Mit dieser Methode wurden einerseits die Komplexität der Probe reduziert und andererseits die Untereinheiten von makromolekularen Proteinkomplexen identifiziert. Im Vergleich zu der 2DE-MALDI-TOF-MS Methode erhöhte sich bei der MSC-LC-MS/MS die Anzahl der identifizierten Proteine deutlich. Insgesamt wurden 466 Proteine identifiziert, die 40% des annotierten cytosolischen Proteoms abdecken. Davon besaßen 187 Proteine die Fähigkeit zu makromolekularen Einheiten mit einer Masse von über 300 kDa zu assemblieren. Alle Komplexe, die im Glycerol-Dichtegradienten kombiniert mit 2DE-MALDI-TOF-MS identifiziert wurden, wurden hier gefunden. Die anschließende bioinformatische Analyse zeigte, dass 111 Proteine zu 35 bereits charakterisierten Proteinkomplexen gehörten, von denen 10 Proteine das Molekulargewicht ihrer funktionellen Homologen aus anderen Organismen übertrafen. Dies konkretisierte die Idee, dass Proteinkomplexe verschiedener Arten, trotz starker Sequenzhomologie der entsprechenden Untereinheiten, unterschiedliche Strukturen und Stöchiometrien aufweisen können. Infolgedessen müssen deren Strukturen individuell erforscht werden.

Basierend auf den bisherigen Ergebnissen entschieden wir uns für eine weitergehende quantitative Proteom- und Transkriptom-Analyse von *T. acidophilum* unter aeroben und anaeroben Bedingungen. Hierfür wurden moderne hochauflösende massenspektrometrische Technologien (eine Kombination aus ein-dimensionaler Gelelektrophorese (1D-SDS-PAGE) und linear-Quadrupol-Ion-Trap Fouriertransformations-Ionenzyklotronresonanz Massen-spektrometrie (LTQ-FTICR-MS)) verwendet, was die Identifikationsrate deutlich erhöhte. Es wurden 1025 Proteine identifiziert, die 88% des annotierten cytosolischen Proteoms abdecken. Basierend auf der labelfreien quantitativen Proteomuntersuchung zeigten 263 Proteine deutliche Regulation bei Abwesenheit von Sauerstoff. Mit dieser Methode wurden insgesamt 39 makromolekulare Proteinkomplexe (die bereits mittels der MSC-LC-MS/MS Methode identifizierten 35 Proteinkomplexe und 4 zusätzliche putative makromolekulare Komplexe) identifiziert, von denen 28 Komplexe quantifiziert wurden. Fünfzehn dieser

Komplexe wurden unter anaeroben Bedingungen erheblich reguliert. Parallel dazu wurden cDNA-Microarrays des gesamten Genoms von *T. acidophilum* erstellt, um Transkriptom-untersuchungen durchzuführen. Die Analyse ergab, dass 445 Gene unter anaeroben Bedingungen reguliert wurden. Mehr als 40% der Membranprotein-kodierende Gene (145 von 335 ORFs) wurden reguliert, welche hauptsächlich an extrazellulärer Polypeptid-degradation und am Ionen- und Aminosäuretransport beteiligt sind. Der Vergleich zwischen Proteom und Transkriptom zeigte eine schwache positive Korrelation zwischen mRNA-Transkription und Proteinexpression, was auf umfangreiche posttranskriptionale Regulierungsmechanismen in *T. acidophilum* hindeutet.

Insgesamt lieferten die umfassenden quantitativen Proteom- und Transcriptom-Untersuchungen an *T. acidophilum*, kultiviert unter aeroben und anaeroben Bedingungen, einen detaillierten Überblick über die Konzentrationen der exprimierten Proteine und der transkribierten mRNAs und Informationen über die biochemische Anpassung der Zelle an anaerobe Bedingungen. Im Fokus des langfristig angelegten *'Visual Proteomics'* Projekts sind die Identifikation und Quantifikation der makromolekularen Proteinkomplexe in *T. acidophilum* von grosser Bedeutung. Die Ergebnisse bieten eine Plattform für weitere Studien, die darauf abzielen, die erhaltenen Quantifikationsdaten mit Kryo-ET-Daten von *T. acidophilum* Zellen zu korrelieren. Desweiteren können die gewonnenen Erkenntnisse als Ausgangspunkt für weitere Experimente, zur biochemischen und strukturellen Charakterisierung von bislang nicht oder nur teilweise untersuchte Komplexen, dienen.

Abstract

Features, like low cellular complexity, lack of cell wall and relatively small (and sequenced) genome, make *Thermoplasma acidophilum* an attractive model organism for 'visual proteomics' studies aiming to provide a comprehensive cellular atlas of macromolecular complexes using cryo-electron tomography (cryo-ET) and pattern recognition methods. *T. acidophilum* is a thermoacidophilic archaeon that grows optimally at 59 °C and pH 1-2. When it grows (micro)aerobically, it uses O_2 as a terminal electron acceptor, but when it is cultured anaerobically, S^0 is the terminal electron acceptor. This physiological versatility together with the simplicity of the cell provides a good platform to detect and compare changes in protein and mRNA expression levels with more established proteomics and microarray methods. These data can be used as a cross-reference to validate the cryo-ET pattern recognition results obtained from *T. acidophilum* cells.

As a starting point, two-dimensional gel electrophoresis (2DE) and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) were used to obtain a global overview of the cytoplasmic proteome of *T. acidophilum* cultured under aerobic conditions. Over 900 spots were resolved by 2DE, from which 271 proteins were identified. We could quantify the most abundant proteins as Coomassie Blue-stained 2D gels allow protein quantitation albeit with limited accuracy. In parallel, with a stronger focus on our group's 'visual proteomics' project, glycerol gradient ultracentrifugation coupled to 2DE-MALDI-TOF-MS analysis was used to identify subunits of macromolecular complexes. In the glycerol gradient high density fractions, ten as yet uncharacterized proteins (besides the well-known ribosomal subunits, translation initiation factor eIF-6 related protein, elongation factor 1 and DNA-dependent RNA polymerase) were identified as putative building blocks of protein complexes. Due to the limitations in dynamic range of 2DE and in protein identification capability of MALDI-TOF-MS, only a limited number of proteins were identified.

In order to increase the identification rate and refine the inventory of protein complexes, the combination of molecular sieve chromatography (MSC) and liquid chromatography tandem mass spectrometry (LC-MS/MS) were employed. These techniques when combined have a

higher dynamic range and are suitable for the processing of complex samples when compared to MALDI-TOF-MS. This approach reduces the sample complexity allowing identification of proteins involved in the formation of macromolecular assemblies. When compared to the 2DE-MALDI-TOF-MS approach, the number of the proteins identified increased to 466, covering 40% of the annotated cytosolic proteome, and 187 of them could be classified as probable subunits of macromolecular complexes with sizes over 300 kDa. Each of the complexes identified by the glycerol gradient coupled 2DE-MALDI-TOF-MS was found here. Subsequent bioinformatics analysis revealed that 111 of these proteins belong to 35 well-studied complexes, of which 10 exceeded the respective sizes of their functional homologues from other organisms, consistent with the theory that composition and quarternary structure of many complexes varies substantially between species (despite the high sequence homology of the corresponding subunits). As a consequence, these structures need to be studied individually.

Based on the previous findings, we conducted a more thorough quantitative proteomics and transcriptomics analysis of *T. acidophilum* cultured under aerobic and anaerobic conditions. Here, state-of-the-art mass spectrometric technologies (a combination of one-dimensional gel electrophoresis (1D-SDS-PAGE) and linear quadrupole ion trap-fourier transform ion cyclotron resonance mass spectrometry (LTQ-FTICR-MS)) were used, which increased the protein identification efficiency tremendously. 1025 proteins were identified, covering 88% of the cytosolic proteome. Based on label-free quantitative proteomics analysis, 263 proteins showed significant up- or down-regulation in the absence of oxygen. From this approach, 39 protein complexes were identified (including the 35 complexes identified with the MSC-LC-MS/MS approach, with 4 additional putative macromolecular complexes). 28 complexes could be quantified, and 15 out of these complexes were regulated under anaerobiosis. In parallel, a genome-wide cDNA microarray analysis was carried out to study the transcriptome of *T. acidophilum*, which revealed that 445 genes changed expression levels under anaerobic conditions. More than 40% of the membrane protein-encoding genes (145 out of 335 ORFs) were regulated at mRNA level, and functionally most are associated with extracellular polypeptide degradation, or ion and amino acid transport. The comparison of transcriptomes and proteomes showed a weak positive correlation between mRNA and protein expression ratios, suggesting extensive posttranscriptional regulatory mechanisms in *T. acidophilum*.

Taken together, comprehensive quantitative transcriptomics and proteomics studies on the physiological state of *T. acidophilum* under both aerobic and anaerobic conditions provided us with a detailed overview of the changes in expression levels of mRNAs and proteins. We learned a great deal about specific physiological adaptions of the cells to anaerobiosis. These results can serve as a platform for further studies that will target a cross-correlation of the proteomics data with the cryo-ET data originating from cellular tomograms of aerobically and anaerobically grown *T. acidophilum* cells. In addition, the identification and quantitation of macromolecular complexes supplies a basis for further experiments focused on the isolation and biochemical and/or structural characterization of previously unknown or partially characterized complexes.

1 Introduction

1.1 *Thermoplasma acidophilum*

The archaeon *Thermoplasma acidophilum* which thrives at 55-60°C and pH 0.5-4 is a member of the *Euryarchaeota* lineage of the Archaea. *T. acidophilum* can grow aerobically using oxygen as a terminal electron acceptor, but when it is cultured anaerobically, S^0 is needed for optimal growth [1]. *T. acidophilum* lacks a rigid cell wall and is pleomorphous, with cell sizes varying between 0.2 and 2 µm [2, 3]. The cells can retain their structural integrity under these rather extreme conditions; at elevated pH values, the plasma membrane is destroyed, causing immediate cell lysis and death [4]. *T. acidophilum* contains a 1.5 Mbp chromosome that has been sequenced; it comprises 1507 open reading frames (ORF's) and 1482 protein-encoding genes, amongst which 29% are similar to proteins of unknown function, and 16% have no significant similarity to any described protein [5].

T. acidophilum is a fascinating microorganism, the study of which can give new insights into how extremophiles can live at high temperatures and low pH without the structural protection of a conventional cell wall. *T. acidophilum* contains macromolecular complexes involved in protein folding, degradation and turnover that look like simple versions of related structures in eukaryotic cells, and many of these macromolecular assemblies have helped to elucidate the structure and function of their more complex eukaryotic homologues [6]. The ability of *T. acidophilum* to adapt to anaerobiosis opens avenues for investigating, comparing and quantifying protein and mRNA expression levels of different growth conditions with proteomics and transcriptomics methods to validate/correlate to cryo-electron tomography (cryo-ET) data.

1.2 Transcriptomics and proteomics

Genomics provides an overall description of the complete set of genetic instructions contained within the genome that are available to a cell, i.e. the 'blueprint' of a cell [7]. Nowadays more than 1000 organisms have been sequenced (see Genomes OnLine Databases at http://wit.integratedgenomics.com). The rapid pace of genome sequencing has resulted in many newly discovered genes that have been ascribed no function or a function that at best has been poorly described. This impetus to understand the function of newly discovered genes

is leading biologists toward the systematic analysis of the expression levels of the components that constitute a biological system [7]. Recent advances in high-throughput technologies enable quantitative monitoring of the abundance of various biological molecules and allow the determination of their variation between biological states on a genomic scale. Two typical platforms are DNA microarrays that measure messenger RNA (mRNA) transcript levels, and mass spectrometry based proteomics that quantify protein abundances [8]. The mRNA levels are immensely informative concerning cell state and the activity of genes, as for most genes changes in mRNA abundance are related to changes in protein abundance. The essence of proteomics is that the final expression product rather than an intermediate is measured. In addition, proteomics enables the detection of post-translational protein modifications, protein complexes, and protein localization, which are unable to be obtained directly by measurements of mRNA [9]. In this respect, protein and mRNA based measurements are complementary. Integration of transcriptomics and proteomics data can generate coherent hypotheses for given biological situations and discover new emergent properties that arise from the systemic view [10].

1.2.1 Transcriptomics

The task of transcriptomics is to compare the mRNA profiles using complementary DNA (cDNA) microarrays [7]. The fundamental basis of cDNA microarrays is the process of hybridization. Two strands of nucleic acid, cDNA or RNA, hybridize if they are complementary to each other. This principle is exploited to measure the unknown quantity of one RNA or cDNA molecule (target) on the basis of the amount of a complementary sequence (probe) that has hybridized to the target. The level of hybridization is quantified by measuring the level of a detectable chemical label, used to mark the target or the probe sequence in the experiment [11].

In the microarray technique, the probe sequences are immobilized on a solid surface by covalent bonds formed to a chemical matrix (via epoxy-silane, amino-silane, lysine, polyacrylamide or others). The solid surface can be glass or a silicon chip [11]. The sample is usually labelled with a fluorescent dye that can be detected by a light scanner that scans the surface of the chip. Each probe matches a particular nucleotide sequence present in the sample material, a preparation of RNA or cDNA. After the hybridization of the sample

material on the chip, the signal detected on the spots upon light scanning, measures the abundance of each specific nucleotide sequence present in the sample. Observing all the microarray spots at the same time gives the profile of a sample; in the most common application of cDNA microarrays, the target sample is mRNA and the total microarray image represents the transcriptional profile of the sample [11]. Therefore, cDNA microarrays can provide a genome-wide portrait of the transcriptome, disclosing differential expression of genes at whole genome levels.

Generally, DNA microarrays can be classified as two-channel and one-channel arrays [11]:

Two-channel microarrays are typically hybridized with cDNA prepared from two samples to be compared and that are labelled with two different fluorophores [12]. Fluorescent dyes commonly used for cDNA labelling include indocarbocyanin (Cy3), which has a fluorescence emission wavelength of 570 nm (corresponding to the green part of the light spectrum), and indodicarbocyanin (Cy5) with a fluorescence emission wavelength of 670 nm (corresponding to the red part of the light spectrum). The two carbocyanin (Cy)-labelled cDNA samples are mixed and hybridized to a single microarray that is then scanned in a microarray scanner to visualize fluorescence of the two fluorophores after excitation with a laser beam of a defined wavelength. Relative intensities of each fluorophore may then be used in ratio-based analysis to identify up-regulated and down-regulated genes [13].

In one-channel microarrays, the arrays are designed to give estimations of the absolute levels of gene expression. Therefore the comparison of two conditions requires two separate single-dye hybridizations. As only a single dye is used, the data collected represent absolute values of gene expression. These may be compared to other genes within a sample or to reference "normalizing" probes used to calibrate data across the entire array and across multiple arrays. One strength of the single-dye system lies in the fact that an aberrant sample cannot affect the raw data derived from other samples, because each array chip is exposed to only one sample (as opposed to a two-channel system in which a single low-quality sample may drastically impinge on overall data precision even if the other sample was of high quality). Another benefit is that data are more easily compared to arrays from different experiments; the absolute values of gene expression may be compared between studies conducted months or

years apart. A drawback to the one-color system is that, when compared to the two-color system, twice as many microarrays are needed to compare samples within an experiment [11].

1.2.2 Proteomics

1.2.2.1 General workflow of proteomics

Proteomics is the global analysis of the protein complement of a defined biological specimen. It is a scientific discipline that promises to bridge the gap between the understanding of genome sequence and cellular behaviour [7]. In recent years, proteome analysis has been accelerated by the extensive use of mass spectrometers in protein identification.

Figure 1.1 describes the steps of a typical mass spectrometry (MS) based proteomics experiment. Briefly, protein samples are fractionated by different means to reduce their complexity, e.g. by one-dimensional gel electrophoresis (1D-SDS-PAGE) or two-dimensional gel electrophoresis (2DE). Following visualization by staining with Coomassie Brilliant Blue, fluorescent dyes, or silver, proteins of interest are excised, and subjected to in-gel digestion. The generated peptide mixture is separated on- or off-line using single or multiple dimensions of peptide separation. Peptides are then ionized by Electrospray or Matrix-Assisted Laser Desorption and can be analysed by various different mass spectrometer detectors. Various types of sources, analysers and detectors can be mixed and matched in various configurations but some combinations are more convenient, such as matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF), electrospray ionization (ESI)-MS and tandem MS-MS. Finally, the peptide-sequence data that are obtained from the mass spectra are searched against protein databases using one of several database-searching programs.

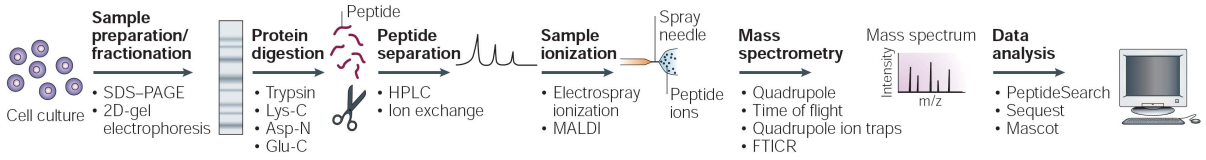


Figure 1.1: A flow chart of proteome analysis methods [14]. Proteins are separated by 2DE or 1D-SDS-PAGE. Following visualization by staining with Coomassie Brilliant Blue, fluorescent dyes, or silver, proteins of interest are excised, and subjected to in-gel digestion. The generated peptide mixture is separated on- or off-line using single or multiple dimensions of peptide separation. Peptides are then ionized by Electrospray or Matrix-Assisted Laser Desorption and can be analysed by various different

mass spectrometer detectors. Finally, the peptide-sequencing data that are obtained from the mass spectra are searched against protein databases using one of several database-searching programs.

1.2.2.2 Mass spectrometric principles and instrumentation

Mass spectrometric measurements are carried out in the gas phase on ionized analytes. By definition, a mass spectrometer consists of an ion source, a mass analyser that measures the mass-to-charge ratio (*m*/*z*) of the ionized analytes, and a detector that registers the number of ions at each *m*/*z* value [15].

Ion source

ESI [16] and MALDI [17] are the two techniques most commonly used to volatize and ionize the proteins or peptides for mass spectrometric analysis. ESI ionizes the analytes out of a solution and is therefore readily coupled to liquid-based (for example, chromatographic and electrophoretic) separation tools (Figure 1.2a). MALDI sublimates and ionizes the samples out of a dry, crystalline matrix via laser pulses (Figure 1.2b). MALDI-MS is normally used to analyse relatively simple peptide mixtures, whereas integrated liquid-chromatography ESI-MS systems (LC-MS) are preferred for the analysis of complex samples [15].

a b

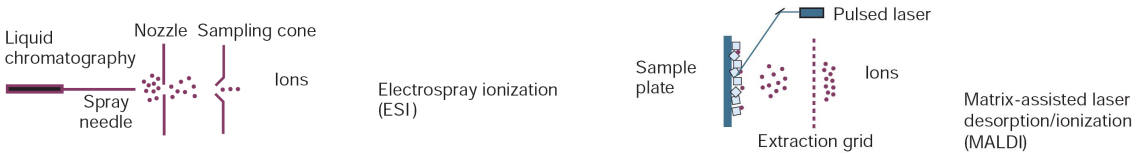


Figure 1.2: Ionisation methods for mass spectrometry. a) Electrospray ionization (ESI), b) Matrix-assisted laser desorption/ionization (MALDI) [15].

Mass analyser

The mass analyser is central to the technology. In the context of proteomics, its key parameters are sensitivity, resolution, mass accuracy and the ability to generate information-rich ion mass spectra from peptide fragments (tandem mass or MS/MS spectra). There are five basic types of mass analysers currently used in proteomics research. These are the time-of-flight (TOF), ion trap, quadrupole, fourier transform ion cyclotron resonance (FTICR)-MS

and Orbitrap analysers [18] (Figure 1.3) [15]. They are very different in design and performance, each with its own strengths and weaknesses. These analysers can stand alone or, in some cases, be put together in tandem to take advantage of the strengths of each [15].

The TOF analyser (Figure 1.3a) is a pulsed analyser. It is usually combined with MALDI. The travel-time of the ion is related to the *m/z* ratio; the "lighter" ions travel quicker. To allow the fragmentation, MALDI ion sources are coupled to two types of TOF instruments (TOF-TOF, Figure 1.3b). Ions are selected in a first mass analyser (TOF1), fragmented in a collision cell and the fragment ion masses are identified by the TOF2 analyser. These instruments have high sensitivity, resolution, mass accuracy and high throughput, but they are only suitable for relatively simple peptide mixtures. Therefore, MALDI-TOF is commonly used in conjunction with prior protein fractionation using either one- or two-dimensional gel electrophoresis.

In ion-trap analysers (Figure 1.3c-e), the ions are first captured for a certain time interval and are then subjected to MS or MS/MS analysis. Ion traps are robust, sensitive and relatively inexpensive. The disadvantage is relatively low mass accuracy, due to the accumulation of ions in the trap chamber.

The FTICR-MS (Figure 1.3f) captures the ions under high vacuum in a high magnetic field. It has high sensitivity, mass accuracy, resolution and dynamic range. But the expense and operation complexity has limited their routine use in proteomics research.

The Orbitrap traps (Figure 1.3g) moves ions in an electrostatic field [18]. It has inherited all the advantages of FTICR, i.e. high mass accuracy, resolution and dynamic range while having increased sensitivity, being a much smaller instrument and not needing a superconducting magnet.

Hybrid mass spectrometers are commonly used in proteomics studies. For instance, the linear quadrupole ion trap (LTQ)-FTICR [19] and LTQ-Orbitrap [18] combine the advantages of fast scanning rate and high sensitivity of the LTQ with the advantages of high mass accuracy, high resolution, and high dynamic range of the FTICR and Orbitrap.

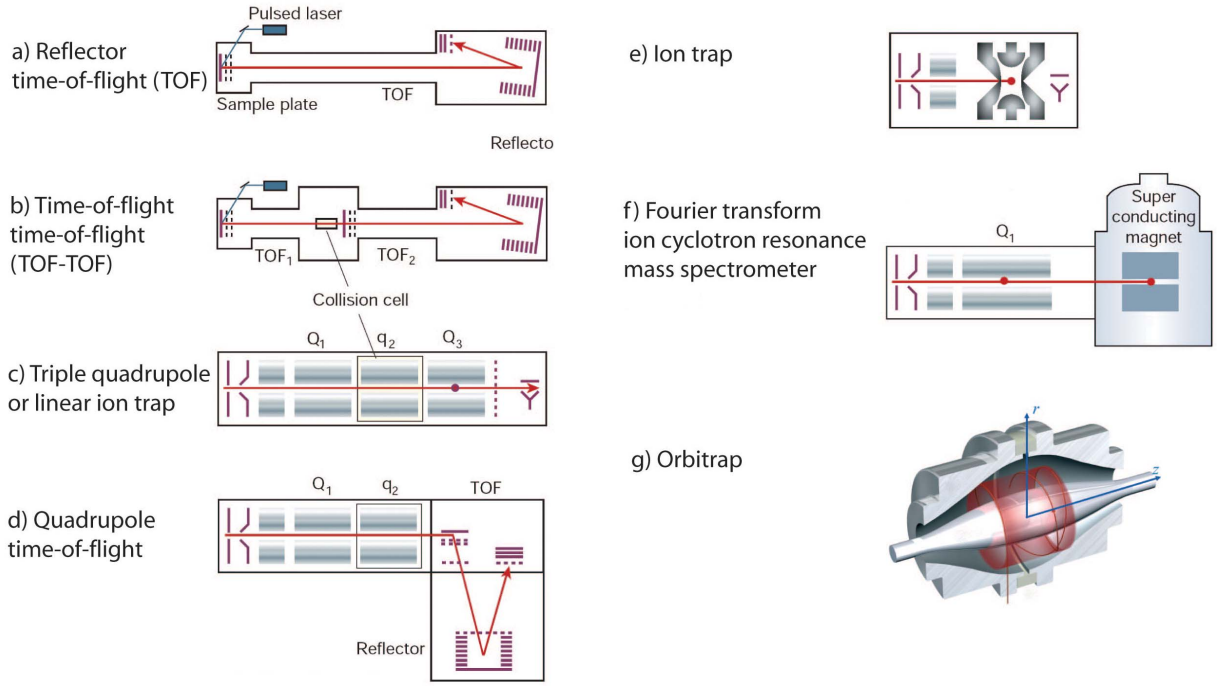


Figure 1.3: Different types of mass spectrometers. a) Time-of-flight (TOF) instruments, b) TOF-TOF instrument, c) Quadrupole mass spectrometer, d) quadrupole TOF instrument, e) Three-dimensional ion trap mass spectrometer, f) FTICR-MS instrument, g) Orbitrap [15, 18].

Detector

The final element of the mass spectrometer is the detector. The detector records the charge induced when an ion passes by or contacts a surface. If a scan is conducted in the mass analyser, the charge induced in the detector during the course of the scan will produce a mass spectrum, a record of the mass to charge ratios (*m/z*'s) at which ions are present [20, 21].

1.2.2.3 Quantitative proteomics

Biological research often requires the knowledge of protein amounts and their changes under different conditions. To meet these needs, various quantitation methods have been developed in the past few years.

Stable-isotope labelling quantitation

Most of the relative quantitative proteomics studies use metabolic labelling using heavy salts or amino acids (e.g. stable isotope labelling by amino acids in cell culture (SILAC), ^{15}N) [22,23], chemical stable-isotope labelling (such as, isotope-coded affinity tags (ICAT), isotope coded

protein label (ICPL)) [24, 25] or enzyme-catalyzed incorporation of ^{18}O [26, 27]. This method makes use of the facts that pairs of chemically identical analytes of different stable-isotope composition can be differentiated in a mass spectrometer owing to their mass difference, and that the ratio of signal intensities for such analyte pairs accurately indicates the abundance ratio for the two analytes.

Figure 1.4 represents different methods of stable-isotope protein labelling for quantitative proteomics. Proteins/peptides are labelled via metabolic, chemical or enzyme reactions. Then, the labelled proteins or peptides are combined and analysed by mass spectrometry. The relative quantitation can be performed by comparing the peak intensities of heavy and light isotope labelled peptide pairs.

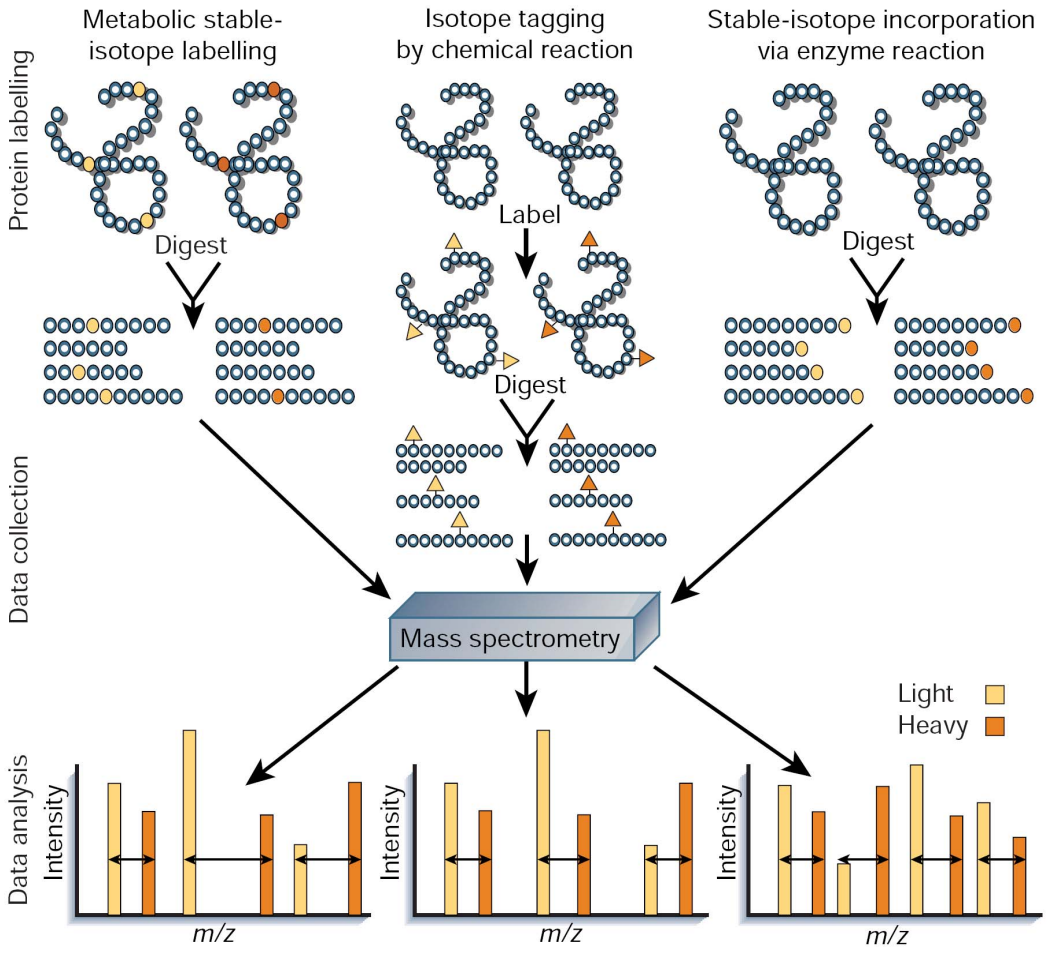


Figure 1.4: Methods for stable-isotope protein labelling for quantitative proteomics [15].

Label-free quantitation

Although providing great accuracy and usability, these stable-isotope labelling quantitation methods are limited by costs, complex labelling procedures, incomplete labeling and complicated purification steps. Recently, the less expensive, straightforward and reproducible label-free quantitation approaches, which are based on comparisons of the ion intensities of identical proteolytic peptides to accurately determine relative ratios of particular proteins, have become more and more widespread [28-31].

Absolute quantitation

In addition to the relative quantitation, knowledge about the absolute abundance of protein components is an important prerequisite for building quantitative predictive models of cellular behaviour. Traditional mass spectrometric strategies have been reported by introducing internal standards [32], typically by mixing isotopically labelled internal standard peptides [33]. However, this has been limited on a proteomics scale due to the expense and difficulty of synthesizing numerous external reference peptides. In this work, a novel label-free approach was used for absolute quantitation. Theoretically, the intensity of each peptide is already a measurement of protein abundance [34]. However, peptides with different amino acid sequences represent different signal intensities in the MS measurement. Thus, to estimate the absolute protein abundance, the ratio between the total intensity of all identified peptides for a protein and the protein sequence length is calculated, thereby averaging out intensity variations caused by protein size and sequence variations.

2 Aim of the study

The proteomics and trancriptomics studies form part of a project aimed at visualizing the proteome of *T. acidophilum* and thereby providing a comprehensive cellular atlas of macromolecular complexes. The idea of this 'visual proteomics' approach is based on a multi-step procedure that comprises the proteomics analysis, the creation of a template library, the acquisition of three-dimensional (3D) cellular tomograms, interpretation of the tomograms by pattern recognition procedures, and finally, the generation of a cellular atlas [35].

In this sub-project, three proteomics approaches and a microarray analysis were carried out to make an inventory of the expressed cytosolic proteins of *T. acidophilum* with special regard on protein complexes. In particular, 2DE combined with MALDI-TOF-MS was applied to provide a preliminary overview of the expressed cytosolic proteins in *T. acidophilum*. Next, glycerol gradient untracentifugation coupled with 2DE-MALDI-TOF-MS and molecular sieve chromatography (MSC) combined with LC-MS/MS was performed to provide identification data for macromolecular complexes with Mw over 300 kDa, and which can thus serve as templates for further structural characterizations by EM. Furthermore, 1D-SDS-PAGE combined with the highly sensitive LTQ-FTICR-MS was applied for more comprehensive studies of expressed proteins. Finally, MS-based, label-free proteomics studies and NimbelGen cDNA microarrays were used to perform quantitative studies of expressed proteins and transcribed mRNAs of *T. acidophilum* cultured under aerobic and anaerobic conditions to provide information concerning physiological adaptations to anaerobiosis. The transcriptomics and proteomics data were combined and compared to generate coherent hypotheses for given biological situations, and the quantitative data can be used as basis for validation of cryo-ET data.

3 Materials and methods

3.1 Project 1: 2DE based proteomics analysis of *T. acidophilum*

Growth of *T. acidophilum* DSM 1728 and cell extract preparation

T. acidophilum was cultured aerobically using the procedure of Robb *et al.* [36] with minor modifications as described earlier [37]. The cells from 500 ml culture (A_{540}=1.0) were harvested by centrifugation at 4000 g for 10 minutes at 4 °C. The cell pellet was washed with MilliQ water and stored at –80 °C overnight if it was not used immediately. Then, the pellets were resuspended in 500 µl MilliQ water containing protease inhibitors. The pH of the cell suspension was elevated with 1 M non-buffered Tris to pH 8, after which 0.5 mg DNaseI was added, and the mixture was incubated on ice until cell lysis occurred. The lysate was centrifuged at 34,800 g for 30 min at 4 °C in a Beckman ultracentrifuge to remove cell debris. The supernatant was used immediately or stored at –20 °C until further use. For 2DE, the cell extracts containing 50 µg (silver staining) or 500 µg (Coomassie Blue staining) protein were combined with 2DE sample buffer according to the GE Healthcare 2DE protocol (2-D Electrophoresis using immobilized pH gradients – Principles and Methods).

Glycerol density gradient ultracentrifugation

Extracts (4 mg in 200-500 µl buffer) from four independent *T. acidophilum* cultures were loaded directly onto 35 ml 10-50% glycerol gradients in 50 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$ and 1 mM adenosine triphosphate (ATP). A BIOCOMP Model 117 Gradient Mate (BioComp Instruments) was set to perform a 3 min run at 15 revolutions per minute (rpm) speed and 80° angle to form a continuous glycerol gradient from the overlayed 10% and 50% glycerol solutions. Centrifugation was carried out for 22 h at 25,000 rpm in a Beckman SW 28 swing-out rotor maintained at 4 °C. After centrifugation, the gradients were fractionated into 35 fractions of 1 ml each by injection of 86% glycerol at the bottom of the tube through a density gradient fractionator (Instrumentation Specialties Company). The fractions - collected from top to bottom - were loaded both on 1D-SDS-PAGE gels to check protein distribution and on native gradient gels to check protein size. Fractions of interest (12-13, 17-19, 22-24, and 26-28, over the size of the 20S proteasome (700 kDa) – were selected, pooled and

precipitated to obtain higher protein concentration. After precipitation with 13% trichloroacetic acid (TCA), protein samples were washed and dissolved in 2DE sample buffer.

2DE

2DE was done according to the GE Healthcare 2DE protocol using 3-11 NL 18 cm long immobilized pH gradient (IPG) strips (GE Healthcare) for first dimension separation, and 11% polyacrylamide (PA) gel in Laemmli buffer system [38] for second dimension separation. Isoelectric focusing (IEF) was performed with a Multiphor II unit (GE Healthcare) at 20 °C, employing the following voltage profile: linear increase from 0 to 500 V in 1 min, linear increase from 500 to 3500 V in 1.5 h, 3500 V for 14 h. Total focusing time was about 49,000 Vh. The IPG strips were stored at –80 °C if they were not used immediately. Separation in the second dimension was achieved using the Protean II xi cell (BIO-RAD).

After completion, the 2DE proteins were fixed and stained. Colloidal Coomassie Brilliant Blue G-250 staining was carried out according to Neuhoff *et al.* [39]. Briefly**,** PA gels were fixed in 12% TCA for 15 minutes followed by washing steps with several changes of water. The gels were stained with Coomassie Blue G250 staining solution overnight at room temperature and the excess of stain was washed out with distilled water. Silver staining was carried out according to Mortz *et al.* [40] with minor modifications. Gels were fixed twice in 50% (v/v) methanol and 12% (v/v) acetic acid for 30 minutes, washed three times with 50% (v/v) ethanol for 20 minutes and soaked for 1 minute in 200 mg/l $Na_2S_2O_3$. After washing twice with MilliQ H_2O for 1 minute, gels were stained for 20 minutes in a freshly made solution containing 2 g/l $AgNO_3$ and 375 µl/l formaldehyde, washed briefly with MilliQ H_2O and developed in a solution including 60 g/l Na_2CO_3, 5 mg/l $Na_2S_2O_3$ and 250 µl/l formaldehyde until the desired contrast was obtained. This usually took 1-5 minutes depending on the temperature. Then the staining solution was drained, gels were washed with MilliQ H_2O, and the staining was terminated with 5% (v/v) acetic acid for 10 minutes. The gels were soaked in 1% (v/v) acetic acid and stored at 4 °C.

Protein digestion and sample preparation

For spot picking, protein digestion and sample preparation, we used the procedures and equipment of the Proteomics Service Facility of the MPI for Biochemistry, following the

methods published by Tebbe *et al.* [41]. Briefly, Protein spots of eleven two-dimensional (2D) gels were visually detected, and then automatically excised using Bruker's Proteineer SP spot-picking robot, which transfers the 1.5 mm gel pieces automatically into 96-well filter microtitre plates. The use of filter microtitre plates has the advantage that all incubation media could be separated from the gel pieces by centrifugation, enabling the eluate to be collected in a microtitre plate placed underneath. Protein bands of one-dimensional (1D)-gels were excised from the gels using a sterile scalpel and cut into1 mm^3 cubes. Excised spots of silver-stained gels were destained in 30 µl working solution and washed 3 times with 100 µl MilliQ H_2O for 5 min each. For Coomassie-stained gels, this step was not needed. Gel pieces were shrunk with 50% v/v acetonitrile (ACN) for 15 min and rehydrated in 50 mM NH_4HCO_3 for 15 min. This treatment was repeated twice. For 1D gels, two additional steps were needed: firstly, 10 mM dithiothreitol (DTT) in 50 mM NH_4HCO_3 was added to the gel pieces and incubated at 56°C for 45 min to break the -S-S- linkages; secondly, 55 mM iodoacetamide in 50 mM NH_4HCO_3 was added and incubated for 30 min to alkylate thiol groups of proteins. Protein digestion was performed with 0.1 µg/spot modified trypsin at 37 °C, overnight. The supernatant was collected and combined with the eluates of the following three elution steps: (i) water; (ii) 50% v/v ACN; and (iii) 50% v/v ACN, 0.1% v/v trifluoroacetic acid (TFA). The combined eluates were frozen in liquid nitrogen and dried in a SpeedVac centrifuge. To remove remaining ammonium bicarbonate, the pellet was resuspended in water, frozen and lyophilised. This procedure was repeated once. Each dried sample was finally dissolved in 10 µl 33% v/v ACN, 0.1% v/v TFA. Using Bruker's MAP pipetting robot, 0.5 µl of each sample was mixed automatically with the same amount of α-cyano-4-hydroxycinnamic acid (matrix for MALDI-TOF-MS) in 40% v/v ACN, 0.1% v/v TFA on a ground steel MALDI target.

Protein identification by MALDI-TOF-MS

MALDI-TOF-MS spectra were recorded automatically using a Reflex III spectrometer (Bruker Daltonics) with an accelerating voltage of 20 kV. The spectra were externally calibrated with an in-house optimized mixture of eight peptides ranging from 1046.54 Da to 3494.65 Da. Spectra were automatically annotated using the vendor's "Xmass" program package (version 5.1.16), which returns monoisotopic masses. Parameters used for the macro

operation were the following: peakpicking method “SNAP”, maxpks 40, PC 2.2, goodness 30, peakdig 3. An average mass accuracy of 120 parts-per-million (ppm) was obtained.

Biotools software (Bruker Daltonics) integrating the locally installed Mascot search engine (Matrix Science, London, UK) [42] was used to search the database comprising all ORF’s of *T. acidophilum* (http://pedant.gsf.de). Search parameters were as follows: 1 missed cleavage sites, +/- 200 ppm tolerance (to account for reduced mass accuracy of larger peptides), carbamidomethylation of cysteines as fixed modification, and oxidation of methionine as variable modification. Proteins were considered to be identified if the Mascot program returned a MOWSE score of at least 44, indicating a probability of less than 5% that the observed match is a random event.

3.2 Project 2: Protein complexes of *T. acidophilum*

Growth of *T. acidophilum* DSM 1728 and cell extract preparation

The sample preparation was same as described in § 3.1 with minor modifications: to keep the native structures of proteins, 5mM $MgCl_2$, 1 mM ATP, 1 mM DTT and 20 % glycerol were added to the cell lysate.

Molecular sieve chromatography

To separate proteins on the base of their size, 2 mg protein mixture in a maximum volume of 200 µl protein extraction buffer was loaded onto a Superose 6 10/300 GL column (GE Healthcare). The column was connected to a high-performance liquid chromatography system (ÄKTA Purifier 10; GE Healthcare), which was operated at room temperature. The protein separation buffer contained 25 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, and 1 mM DTT. The flow rate was adjusted to 0.4 ml/min and 25 protein-containing fractions, each of 0.6 ml volume, were collected. The protein elution profile was monitored with a UV detector operating at 280 nm.

The column was calibrated with molecular sieve molecular mass (Mw) markers (Sigma): Blue Dextran (2000 kDa), thyroglobulin (669 kDa), apoferritin (443 kDa), β-amylase (200 kDa), alcohol dehydrogenase (150 kDa), albumin (66 kDa), and carbonic anhydrase (29 kDa).

Protein digestion

One hundred and fifty µl of each fraction from the MSC was dried by centrifugation under vacuum, and reconstituted in 100mM Tris-HCl (pH 8.5) containing 8M urea. The samples were reduced with 10 mM DTT at 37 °C for 30 min, followed by alkylation with 50 mM iodoacetamide in the dark for 1h. The urea concentration was reduced with water to 1M, and the proteins were digested with sequencing grade modified trypsin (Promega) at 37°C overnight using a protease:protein ratio of 1:50 (w/w). The resulting peptides were desalted with MonoTip C18 (GL Science Inc., Tokyo, Japan) pipette tips and eluted with 0.1% TFA in 60% acetonitrile. Eluted peptides were dried and dissolved in 100 µl 2% acetonitrile containing 0.1% TFA.

LC-MS/MS (LCQ-DECA XP plus ion trap MS) analysis and database searches

Tryptic peptides (10 µl) were separated by reverse-phase chromatography using a 0.2 mm inner diameter (i.d.) x 15 cm EX-Nano MonoCap capillary column (GL Science) connected to a MAGIC 2002 LC system (Michrom BioResouces Inc., Auburn, CA, USA). Peptides were automatically injected onto a Peptide CapTrap column (0.5 mm i.d. x 2.0 mm, Michrom BioResources) equilibrated with solvent A (2% acetonitrile, 0.1% formic acid) for concentration. After concentration, the sample was loaded onto a reversed-phase column and a 180 min linear gradient from 5% to 30% of solvent B (90% acetonitrile, 0.1% formic acid) at a flow rate of 120 µl/min was applied. The eluate was split by a MAGIC Splitter (Michrom BioResouces) to approximately 1.0 µl/min and introduced directly into an LCQ-DECA XP plus ion trap mass spectrometer (Thermo Fisher Scientific Inc., Waltham, MA, USA) equipped with a nanoelectrospray ion (NSI) source (AMR Inc., Tokyo, Japan). The NSI needle (FortisTip capillary needle, 20 µm i.d., AMR) attached directly to the reversed-phase column was used as the NSI interface. The mass spectrometer was operated in positive ion mode at spray voltage 1.8 kV, and a capillary temperature of 250 °C. No sheath or auxiliary gas was used. The data were acquired in a data-dependent mode by alternating a MS scan survey over the range *m/z* 400-2000 and a MS/MS scan in an exclusion dynamic mode under the automated control of Xcalibur (version 1.3) software (Thermo Fisher Scientific). The most intense ion was selected and excluded for further selection for a duration of 0.5 min. MS/MS data were acquired using a 2 *m/z* ion isolation width and the normalized collision energy was set at 35%.

All tandem MS spectra were searched against an in-house prepared decoy protein sequence database comprising direct and reverse sequences (to facilitate false positive rate estimation) of 1482 entries derived from the *T. acidophilum* DSM 1728 genome database (Genbank entry: AL139299) in FASTA format and analysed with the TurboSequest algorithm in the Bioworks 3.2 software package (Thermo Fisher Scientific). The mass tolerance of the intact precursor and fragment ions was set at 2 and 1 Da, respectively. The MS/MS data files were searched allowing up to 2 internal missed tryptic cleavages. The following modifications were permitted: carbamidomethylated cysteine (+57 Da) and oxidized methionine (+16 Da). The identified peptides were further evaluated using the following filters: the number of top

matches (=5), the ScoreFinal (Sf, ≥0.85) and the peptide probability (≤1e-003). ScoreFinal (Sf) combines five SEQUEST scores (Sp, RSp, Ions, XCorr, and ΔCn) using a neural network to reflect the strength of peptide assignment on a scale of 0.0 to 1.0 [43, 44]. With these criteria, no peptides were found matching to more than one protein database entry, the false positive rate was below 1%, as estimated by reverse database searching [45], and proteins having at least two unique peptides of distinct sequences identified were accepted. Proteins that were represented only by one peptide were eliminated from the list, even if the same peptide was found in several fractions/experiments; however, we regarded the presence of such a protein (represented only with one peptide in one experiment) as confirmed if another unique peptide of the same protein was found in other runs.

To check the correctness of the assumption that MSC lowers the sample complexity, and therefore that the number of identified proteins can be increased, two more LC-MS/MS measurements were carried out. Fractions of the MSC separation were pooled by five (11-15, 16-20, 21-25, 26-30, 31-35) and analysed, after which the complete crude extract was subjected to LC-MS/MS analysis by a single injection. In these experiments, the same sample preparation methods were used, but the LC separation of the tryptic peptides was modified. A 0.2 mm i.d. x 75 cm EX-Nano MonoCap capillary column (GL Science) was used with the following set up: 390 min linear gradient elution from 5% to 30% of solvent B (90% acetonitrile, 0.1% formic acid) at a flow rate of 120 µl/min.

Bioinformatic tools used for the characterization of protein complexes

Database searches were carried out using the Protein-protein Basic Local Alignment Search Tool (BLAST) of the National Center for Biotechnology Information (NCBI) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) [46, 47], and the Conserved Domain Database (CDD, http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) was used to obtain general information about the biological function of the identified proteins [48]. In order to assign the proteins to biochemical pathways, the Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/tool/search_pathway.html) database was searched [49]. Using the munich information center for protein sequences (MIPS) database the proteins were classified into functional categories (http://mips.gsf.de). The Protein Extraction, Description and ANalysis Tool (PEDANT) genome database provided exhaustive automatic analysis of

genomic sequences of *T. acidophilum* (http://pedant.gsf.de). Protein structure models ('templates') to be used for pattern recognition–based searching in cryo-ETs of *T. acidophilum* were generated from high resolution atomic structure data deposited in the Protein Data Bank (PDB, http://www.pdb.org/pdb/home/home.do [50]). A 3D density map was created using the mass and coordinates of each atom; it was then filtered to match the resolution of the tomogram for the template matching procedure. The resulting maps were displayed in a gallery using Amira (TGS Mercury Labs).

3.3 Project 3: Quantitative proteomics and transcriptomics analysis of *T. acidophilum* cultured under aerobic and anaerobic conditions

Preparing protein extracts of *T. acidophilum* cultured under aerobic and anaerobic conditions

The growth of aerobic cells was same as described in § 3.1.

For anaerobic cultures, cells were grown in medium, which was modified as follows: the pH was adjusted to 2.0 with H_2SO_4 (96%), 0.4% (w/v) S^0 was added, the medium was heated at 100 °C for 30 min to sterilize the sulfur, and oxygen was reduced by adding 1.2 g/l $Na_2S{\cdot}xH_2O$ (x=7-9). Cell cultures were placed into a plastic beaker – to prevent corrosion – that was positioned in a Parr bomb 1,2 l metal container. The gas phase was filled with 80% N_2/20% CO_2 (pressure 2.5 kp/cm^2) and the culture was agitated continuously at 300 rpm with a magnetic stirrer. Cells were grown at 59 °C for 4 days untill the late-exponential phase [36]. Before harvest, the culture was shaken briefly to loosen the cells attached to S^0, in the following, S^0 was removed by using a glass filter with a pore size of 16-40μm. From the filtrate, cells were harvested by centrifugation at 4000 g for 10 minutes at 4 °C, and then washed with MilliQ water (pH 4) and stored at –80 °C before lysis.

The preparation of the cell extract was same as described in project 1 (see § 3.1).

Three independent biological replicates were prepared for both aerobic and anaerobic conditions to obtain robust quantitation statistics and to increase the identification rate of less abundant proteins.

1D-SDS-PAGE and in-gel digest

Protein concentrations were measured using the Bio-Rad Protein Assay (BIO-RAD). Protein samples (20 μg of total protein) of each aerobic and anaerobic extract were separated on a 4-12% NuPage Novex Bis-Tris gel (Invitrogen) and stained using the Colloidal Blue Staining Kit (Invitrogen). The in-gel protein digestion by trypsin was performed according to Shevchenko *et al.* (2006) [51] with minor modifications. Briefly, each of the gel lanes was cut

into 15 slices. To facilitate protein identification, the intensely stained protein bands were separated from the weakly stained ones. Each excised slice was chopped into 1 mm^3 cubes, washed twice with 50 mM ammonium bicarbonate for 20 min at 25 ^oC, dehydrated in 100% ethanol for 10 min at 25 ^oC, and dried in a SpeedVac for 5 min. Protein reduction was performed by incubating the gel pieces in a solution of 10 mM DTT, 50 mM ammonium bicarbonate for 60 minutes at 56 °C, followed by alkylation in the dark with 55 mM iodoacetamide, 50 mM ammonium bicarbonate for 45 min at 25^oC. After washing twice with 50 mM ammonium bicarbonate, the gel pieces were dehydrated with 100% ethanol and dried in a SpeedVac. The gel pieces were rehydrated in 12.5 ng/µL trypsin in 50 mM ammonium bicarbonate and incubated for 16 hours at 37°C for protein digestion. The digestion was stopped by adding 2 µL 100% TFA. The supernatant was transferred to a fresh tube. Peptides were extracted twice with 30% ACN containing 3% TFA, and then dehydrated twice with 100% ACN. The extracts were concentrated in a SpeedVac to 10-20% of the original volume to remove ACN, and then purified using C_{18} StageTips [52]. Peptides were eluted three times from StageTips using 15 µl 80% acetonitrile containing 0.5% acetic acid. The samples were dried in a SpeedVac to 5 µl and acidified with 5 µl of 2% acetonitrile containing 1% trifluoroacetic acid. Afterwards, 5 µl samples were analysed with an LTQ-FTICR mass spectrometer.

LC-MS/MS

The LC-MS/MS analysis was performed as described by Graumann *et al.* (2007) [53]. The peptide mixture was separated by nanoscale C_{18} reverse-phase liquid chromatography (Proxeon Biosystems, Odense, Denmark) coupled on-line to a 7-T LTQ-FTICR mass spectrometer (Thermo Electron, Bremen, Germany) [54]. Briefly, each sample was eluted with 0-40% solvent (80% MeCN in water containing 0.5% acetic acid) over 90 min and injected into the mass spectrometer via a nanoelectrospray ion source (Proxeon Biosystems, Odense, Denmark). The mass spectrometer was operated in positive ion mode and employed a data-dependent automatic switch between MS and MS/MS acquisition modes. After accumulating a target value of 5,000,000 ions in LTQ, a full scan was acquired in the FTICR analyser with a resolution of 100,000 at m/z 400. The five most intense ions from the range 300-1800 m/z were sequentially accumulated and fragmented in the LTQ, with a target value of 5,000 for

each ion species. The MS/MS fragmentation was performed by collision induced fragmentation (CID). Total cycle time was approximately 3 s. Former target ions selected for MS/MS were dynamically excluded for 30 s.

The general parameters were: spray voltage, 2.2 kV; no sheath and auxiliary gas flow; ion transfer tube temperature, 175ºC; normalized collision energy using wide-band activation mode, 35% for MS^2. Ion selection thresholds were 500 counts for MS^2. An activation q of 0.25 and activation time of 30 ms was applied in MS^2 acquisitions. Data were acquired using the Xcalibur software.

Mass spectrometry data analysis

Mass spectra were analysed using the in-house developed software MaxQuant (version 1.0.11.5), which performs peak list generation, estimation of false positive rate and computational algorithms based on Cox *et al.* [55]. The data were searched against the protein sequence database comprising direct and reverse sequences of 1482 entries derived from the *T. acidophilum* DSM 1728 genome database using Mascot Daemon (Version 2.1.0, Matrix Science [56]). Enzyme specificity was set to trypsin, allowing for cleavage N-terminal to proline and between aspartic acid and proline [54]. Carbamidomethyl cysteine was set as fixed modification, whereas oxidation of methionine, and N-acetylation were set as variable modifications. The false positive rate was set to 1% at peptide and protein levels. The minimum peptide length was set to 6. This acceptance criterion was subsequently strengthened to a minimum of two unique identified peptides per protein for compilation of the final list.

Relative label-free quantitation

The MaxQuant software was used for label-free quantitation analysis, which contains algorithms for retention time alignment, transferring identifications between runs, normalization of intensities, and protein quantitation [34, 57]. Figure 3.1 shows the workflow of the label-free quantitation. Two-fold change was used to define biological regulation. The coefficient of variation (CV) and a two-sided Student's t-test were used to perform statistical evaluations. The CV is defined as the ratio of the standard deviation to the mean, which

represents the inter-experimental variations of experimental replicates. The Student's t-test was used to calculate the significances in expression level changes of the regulated proteins.

The statistical criteria defined for the label-free quantitation in this study are the following:

For the proteins within 2-fold change ($0.5<Ratio_{anaerob/aerob}<2$), CV tests were used to judge the experimental variations. CVs of normalized intensities for proteins detected in 3 aerobic and anaerobic biological replicates were calculated, respectively. Proteins with $0.5<Ratio_{anaerob/aerob}<2$ and CVs smaller than 50% both under aerobic and anaerobic conditions are clustered as unchanged proteins.

For the proteins with more than 2-fold change ($Ratio_{anaerob/aerob} \geq 2$ or $Ratio_{anaerob/aerob} \leq 0.5$), the p-value of the Student's t-test was set with a cutoff of 0.05 to give significant quantitation data with 95% confidence. Proteins with $Ratio_{anaerob/aerob} \geq 2$ or $Ratio_{anaerob/aerob} \leq 0.5$ and $p \leq 0.05$ are defined as regulated proteins.

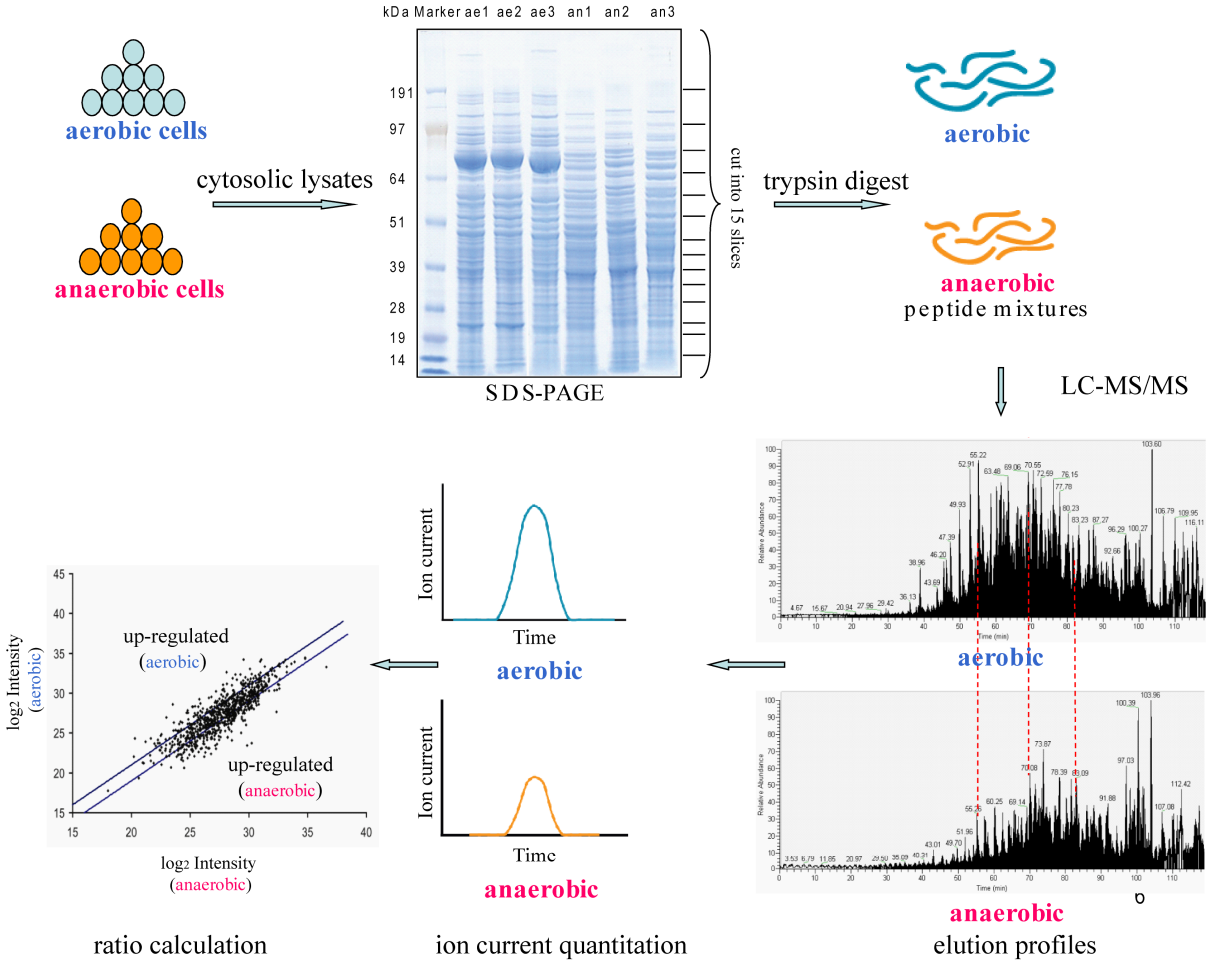


Figure 3.1: Workflow of the label-free quantitation. Proteins of two different states are digested and analysed by MS separately. The quantitative ratios of the proteins are calculated based on the ion current intensities of the common peptides.

Protein abundance estimation

The abundance of each peptide was estimated from the area under the extracted ion current peak, and the abundance of protein expressed was then calculated based on the total intensity of identified peptides divided by the protein sequence length.

Microarray analysis

Total RNA of *T. acidophilum* was isolated with the RNeasy Protect Bacteria Kit (Qiagen). The transcriptomics analysis was performed on TI273075 60mer chips of Roche NimbleGen microarrays (NimbleGen Systems of Iceland, LLC). Probes were selected for all protein sequences (1482) and labelled with Cy3. The median number of probes per sequence is 20,

and each probe is replicated 5 times on the chip. The probes are randomly distributed over the surface of the array. Unused features are filled with randomly generated probes of comparable GC content. ArrayStar v2.0 software (DNASTAR, Inc.) was used for the data analysis. Three independent biological replicates were processed for aerobic and anaerobic conditions, respectively. The statistic analysis was based on Student's t-test.

Bioinformatics tools

Additional to the bioinformatic tools, BLAST [46, 47], CDD [48], KEGG [49], MIPS and PEDANT that were described in § 3.2, Cytoscape along with its Plug-in Bingo 2.0 was used to analyse the distribution of experimental datasets among various protein groups and to identify significantly overrepresented biological functions of the proteins [58]. The Gene Ontology (GO) annotations of proteins were compared with the ones from a reference proteome (e.g. identified proteins vs. the entire protein database, or a subset of the identified proteins vs. the overall identified proteins). To assign corresponding GO identifiers to each protein entry, the PEDANT database (http://pedant.gsf.de) was performed. The hypergeometric test and the Benjamini & Hochberg False Discovery Rate correction were performed to derive overrepresented functions [58], and a probability value of 0.05 was considered significant.

4 Results and discussion

4.1 Proje ct 1: 2DE based proteomics analysis of *T. acidophilum*

4.1.1 Results

2DE reference map of *T. acidophilum* cytoplasmic proteins

Most of the *T. acidophilum* proteins were found in the pH range 4-7, although the theoretical isoelectric point (p*I*) values predicted a protein accumulation around pH 8 as well (data not shown). Comparing the number of the detected spots, we found that despite an 8-fold increase in protein concentration loaded on Coomassie-stained gels (600 spots), the silver staining visualized more protein spots (933 spots). The greater sensitivity of the silver staining did not result in an increase in the proportion of proteins identified through MALDI-TOF-MS, with only 442 (34%) out of the 1317 excised spots being identified. In contrast, Coomassie Blue G250 stain gave much better MS results, 1235 (75%) out of the 1640 excised spots being identified. Altogether, 271 proteins were identified of which 178 were found 3 times or more, 29 were found twice, and 64 were found once.

The majority of the spots contained only single proteins but in several cases the MS analysis indicated a mixture of proteins. When second or third hits could not be verified in consecutive experiments, these were omitted from the protein list.

Protein distribution over multiple spots seems to be common for *T. acidophilum* (Figure 4.1.1). The isoforms of ribonucleotide reductase (Ta1475), thermosome subunits (Ta0980 and Ta1276), Fe-superoxide dismutase (SOD) (Ta0013), translation elongation factor aEF-1 alpha chain (Ta0444), succinyl-CoA synthetase alpha subunit (Ta1331), peroxiredoxin 1 (Ta0152), VAT ATPase (Ta0840), glutamine synthetase (Ta1498), a conserved hypothetical protein (Ta0085), S-adenosylmethionine (AdoMet) synthetase (Ta0059) and proteasome alpha subunit (Ta1288) are indicated. The reason for this isoform distribution is unknown but nevertheless, the theoretical and experimental protein distribution in 2D gels correlated well, thus corroborating the analysis technique.

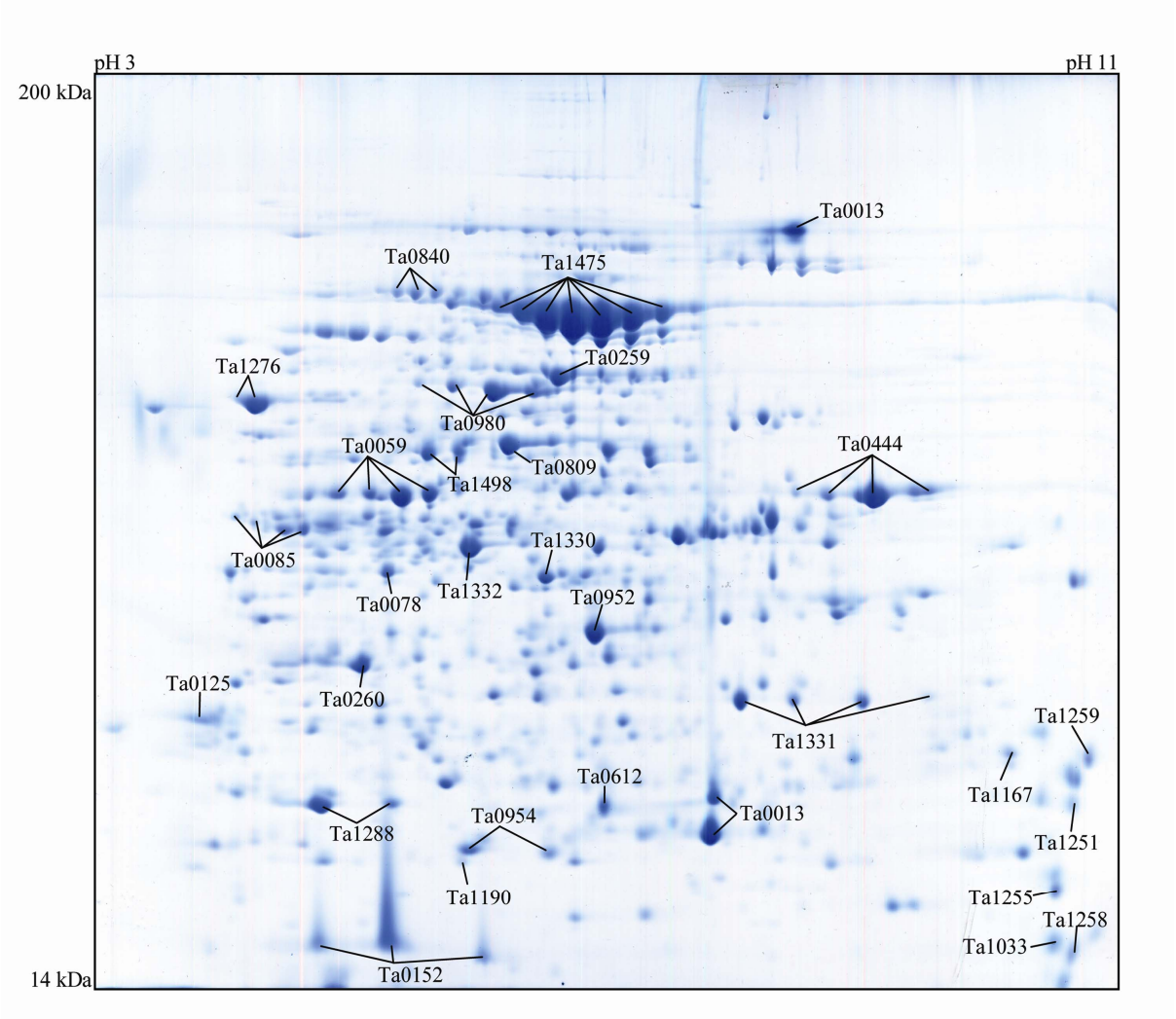


Figure 4.1.1: 2DE reference map of cytosolic proteins of T. acidophilum. Protein isoforms were detected by either IEF (Ta0444, Ta1475 etc.) or by size separation (Ta0013). Undissociated protein complexes were also detected (Ta0013). Only the most prominent proteins are annotated on this map.

Functional categories of the identified proteins

An overview of the distribution of the identified proteins in functional categories (based on the MIPS database, http://mips.gsf.de), including a more detailed description of the most prominent proteins is provided. The complex-forming ability of eighteen of these proteins is summarized in Table 4.1.1. Proteins of Coomassie-stained gels were quantified using PDQuest (BioRad) software; however, a precise quantitation is difficult due to the presence of protein isoforms – which requires that each of the spots is analysed and summed – and to highly expressed proteins which exceeded the dynamic range of the staining method (data not shown).

A large proportion of the identified proteins can be assigned a metabolic function (43.0%). These include groups of amino acid (15.6%), nucleotide (9.6%), C-compound and carbohydrate (11%), lipid and fatty acid (5.9%) and vitamin and cofactor (5.1%) metabolism. Proteins involved in energy metabolism were the second most abundant group (16.5%), followed by the categories of protein synthesis (10.2%) and protein fate (5.5%). Three groups, 'classification still preliminary' (13.9%), the 'unclassified proteins' (11%) and proteins of cellular transport (1.5%) were underrepresented compared to the genome annotation data (Table 4.1.2).

Table 4.1.1 List of highly abundant proteins of T. acidophilum. The gene number (ORF), annotation, complex formation, and the Mw of the complex are indicated. Complex formation of the protein in other organisms is also considered (Ta, T. acidophilum; Tt, Thermus thermophilus; Mj, Methanococcus jannaschii; Pf, Pyrococcus furiosus; Psp, Pyrococcus sp. strain KOD1; Ap, Aeropyrum pernix K1).

ORF	Annotation	Complex-formation	*M*w of complex (kDa)	Organism	Reference
14 ribosomal proteins	ribosome	70S complex	2400	Tt	[59]
Ta0980, Ta1276	thermosome	two eight membered rings	900	Ta	[60]
Ta1288, Ta0612	20S proteasome	four rings	700	Ta	[61]
Ta1498	glutamine synthetase	dodecamer	637	Psp	[62, 63]
Ta1331, Ta1332	succinyl-CoA synthetase	tetramer or dimer	500	Ta	unpublished data
Ta0840	VAT ATPase	hexamer	500	Ta	[64]
Ta0152	peroxiredoxin	hexadecamer	~490	Ap	[65]
Ta1330	ornithine carbamoyltransferase	dodecamer (four trimers)	420	Pf	[66, 67]
Ta0259, Ta0260	pyruvate ferredoxin oxidoreductase	yes	300	Ta	unpublished data
Ta1475	ribonucleotide reductase	monomer	87	Ta	[68]
Ta0059	S-adenosylmethionine synthetase	dimer	86	Mj	[69]
Ta0013	Fe-superoxide dismutase	-	-	-	-
Ta0078	phosphoribosyl- transferase	-	-	-	-
Ta0085	conserved hypothetical protein	-	-	-	-
Ta0125	alkyl hydroperoxide reductase subunit F	-	-	-	-
Ta0444	translation elongation factor aEF-1 alpha chain	free monomer or in association with the ribosome	-	-	[70]
Ta0809	aldehyde dehydrogenase	-	-	-	-
Ta0952	malate dehydrogenase	-	-	-	-

- : information not available.

Table 4.1.2 Functional distribution of T. acidophilum cytosolic proteins identified from 2D SDS-PA gels. The MIPS database server was used for automated protein categorisation. Both proteome (271 identified proteins) and genome data (1482 ORF's) were submitted to the server, and the results are provided in absolute (match) and relative (%) numbers.

Functional Category	Proteome (271)		Genome (1482)	
	match	%	match	%
METABOLISM	117	43.1	291	19.3
Amino acid metabolism	41	15.1	101	6.7
Nitrogen and sulfur metabolism	9	3.3	16	1.0
Nucleotide metabolism	26	9.6	52	3.4
Phosphate metabolism	1	0.4	1	0.1
C-compound and carbohydrate metabolism	29	10.7	66	4.4
Lipid, fatty acid and isoprenoid biosynthesis	16	5.9	35	2.3
Vitamins, cofactors, prosthetic group	14	5.2	42	2.8
Secondary metabolism	4	1.5	20	1.3
ENERGY	45	16.6	99	6.6
Glycolysis and gluconeogenesis	8	3.0	12	0.8
Pentose phosphate pathway	5	1.8	10	0.7
Tricarboxylic acid pathway	18	6.6	23	1.5
Electron transport	4	1.5	22	1.4
Respiration	6	2.2	21	1.4
CELL CYCLE AND DNA PROCESSING	4	1.5	36	2.7
TRANSCRIPTION	9	3.3	41	2.7
PROTEIN SYNTHESIS	28	10.3	90	5.9
PROTEIN FATE	15	5.5	43	2.8
Protein folding	5	1.8	13	0.9
Protein degradation	8	3.0	19	1.3
CELLULAR TRANSPORT	4	1.5	109	7.2
CELL RESCUE AND DEFENCE	11	4.0	27	1.8
BIOGENESIS OF CELL. COMP.	3	1.1	11	0.7
CLASSIFICATION NOT YET CLEAR-CUT	38	14.0	368	24.4
UNCLASSIFIED PROTEINS	30	11.0	468	31.0

Transcription, translation and protein fate

We found a large set of proteins belonging to the complexes of exosome, ribosome, and DNA-dependent RNA polymerase that, together with the presence of a great number of proteins taking part in amino acid (65) or nucleotide (34) metabolism, indicates that *T. acidophilum* maintains highly active biochemical pathways to balance between protein, DNA and RNA degradation and biosynthesis.

Proteins involved in RNA biosynthesis and degradation were well represented; three subunits of the archaeal DNA-dependent RNA polymerase (Ta0391, Ta0392 and Ta1030) were found as well as the TATA-box binding protein (Ta0199). The DNA-directed RNA polymerase consists of 11 subunits, and the Mw of the complex is 500 kDa [71]. Subunits (Ta1291, Ta1292,

Ta1294, Ta0613 and Ta0929) of the putative exosome, a complex of RNases, RNA-binding proteins and helicases which mediates processing and 3′→5′ degradation of a variety of RNA species [72] were also identified.

Proteins needed for protein translation were represented by 14 ribosomal subunits, translation factor Ta0302, translation elongation factor aEF-1 α (Ta0444), GTP-binding protein (Ta1237), translation initiation factors (Ta1212, Ta0322) and aminoacyl tRNA synthetases for valine (Ta0040), serine (Ta0468), asparagine (Ta0946), lysine (Ta1163), histidine (Ta0099) and alanine (Ta0499). One of the most abundant proteins was the translation elongation factor aEF-1 alpha chain (Ta0444) that can be present in the cytosol either as a free monomer or in a complex with the ribosome [70].

Proteins that belong to the functional category protein fate (determining protein folding, modification, destination and degradation) were also abundant. At least one member of each chaperone class found in *T. acidophilum* [73] was expressed (Ta0125, glutaredoxin related protein; Ta0471, small heat-shock protein (hsp20); Ta0840, VAT ATPase; Ta1175, VAT-2 protein; Ta0866, thioredoxin related protein; Ta0980, thermosome α-chain; Ta1276, thermosome β-chain; Ta1011, peptidyl-prolyl *cis-trans* isomerase related protein; and Ta1087, probable DnaK-type molecular chaperone). The identified representatives of the cytosolic protease clans were as follows: Ta1490, tricorn core protease; Ta0301, tricorn cofactor F2; Ta1288, proteasome α subunit; Ta0612, proteasome β subunit; Ta1439, methionine aminopeptidase I; Ta1037, proline dipeptidase; and Ta0465, Pfpl-related endopeptidase.

The most abundant proteins of protein fate were the thermosome, VAT ATPase and proteasome. The thermosome is a molecular chaperone composed of two subunits, α and β, that are arranged in two stacked eight-membered rings (900 kDa) with a central cavity that provides a sequestered environment for *in vivo* protein folding [74]. VAT ATPase Ta0840 is a hexameric (~500 kDa, 15.5 nm) archaeal member of the Cdc48/p97 family of AAA ATPases (ATPases associated with a variety of cellular activities) [64]. It has two ATPase domains and a 185 residue amino-terminal substrate-recognition domain, VAT-N. VAT shows activity in protein folding and unfolding and thus shares the common function of these ATPases in disassembly and/or degradation of protein complexes [64]. The 20S proteasome is a macromolecular assembly (700 kDa) designed to confine proteolytic activity to an inner

cavity. Access to the central proteolytic nanocompartment is restricted to unfolded proteins, which necessitates a functional coupling of the 20S proteasome to a substrate-recognition and unfolding machinery [61]. Most of the peptides generated by the proteasome are degraded further to single amino acids that can be used in cell metabolism and for the synthesis of new proteins. In *T. acidophilum*, the giant tricorn protease (Ta1490) and its interacting factors work downstream of the proteasome, processing the peptides into amino acids [75]. However, we found that the quantity of expressed tricorn protease was much lower than that of the proteasome levels (data not shown).

Abundant proteins of amino acid metabolism

Glutamine synthetase (Ta1498) produces glutamine from glutamic acid and ammonia. It forms a dodecameric complex in bacteria, which is comprised of two face-to-face hexameric rings forming a cylindrical aqueous channel [62]. It was also shown to be functional in a dodecameric form (637 kDa) in the hyperthermophilic archaeon *Pyrococcus* sp. strain KOD1 [63], and we therefore anticipate that it has a similar architecture in *T. acidophilum*.

Anabolic ornithine carbamoyltransferase (OTCase) (Ta1330) catalyzes the carbamoylation of L-ornithine to form citrulline in the sixth step of the arginine biosynthesis pathway. Most anabolic OTCases form trimers of about 105 kDa. The anabolic OTCase of the hyperthermophilic archaeon *Pyrococcus furiosus* is a dodecamer composed of four catalytic trimers (420 kDa) [66, 67], which is therefore suitable for cryo-ET analysis.

Vitamin, cofactor, and prosthetic group metabolism

Interestingly, proteins that catalyze the biosynthesis of coenzymes like folate (Ta0079, Ta1022), vitamin B_{12} (Ta0078, Ta0571, Ta0652-Ta0660) and B_6 (pyridoxine biosynthesis pyroA protein, Ta0522) were highly expressed. Ta0078 is a conserved hypothetical protein that shows homology to CobT. This putative phosphoribosyltransferase plays a central role in the synthesis of alpha-ribazole-5'-phosphate, an intermediate for the lower ligand of cobalamin that is essential for *de novo* cobalamin (vitamin B_{12}) synthesis in bacteria [76]. Strikingly, the highly abundant enzyme ribonucleotide reductase (Ta1475) that catalyzes the conversion of both purine and pyrimidine nucleotides to deoxynucleotides in all organisms

and provides all the monomeric precursors essential for both DNA replication and repair [68] needs B_{12} to carry out its catalytic functions [77].

The synthesis of numerous biological compounds and the regulation of many metabolic processes require the addition or removal of one-carbon units. Ta0060, Ta0898, Ta1476 and Ta1478 are identified enzymes that take part in the one-carbon pool by means of folate. Tetrahydrofolate (THF) coenzymes mediate C1 transfer reactions that are involved in several major cellular processes including the synthesis of purines and thymidylate, amino acid metabolism, pantothenate synthesis, and methionine (Met) synthesis. Met is the direct precursor of S-AdoMet, which in turn is the source of methyl units for the synthesis of a myriad of molecules [78]. S-AdoMet synthetase (Ta0059) catalyzes the only known route of AdoMet biosynthesis [79], and therefore plays a central role in the metabolism of all cells. The biological roles of AdoMet include acting as the primary methyl group donor for DNA and protein methylation, as precursor to the polyamines, and as a progenitor of a 5'-deoxyadenosyl radical [80-82].

Cell rescue and defense proteins

We identified 11 out of the annotated 27 cell defense proteins. The most prominent highly expressed members of this group were the anti-oxidative enzymes SOD (Ta0013), peroxiredoxins (Ta0152, Ta0473 and Ta0954) and the alkyl hydroperoxide reductase subunit F (Ta0125). SOD is a metalloenzyme playing a central role in the defensive system of organisms towards the toxicity of superoxide radicals [83]. Peroxiredoxins contain a reactive Cys residue in the conserved region near the N-terminus, and this catalytic Cys residue forms cysteine-sulfenic acid as a reaction intermediate during the reduction of peroxide [84]. Ta0125 exhibits homology to alkyl hydroperoxide reductase subunit F (AhpF) that together with AhpC (a peroxiredoxin homologue) catalyzes the NADH-dependent reduction of organic hydroperoxides (or hydrogen peroxide) to their corresponding alcohols and water [85]. For the structure, mechanism and regulation of peroxiredoxins, we refer the reader to a recent review [86].

Putative complex-forming proteins identified by glycerol density gradient ultracentrifugation coupled to 2DE-MALDI-TOF-MS

Glycerol density gradient ultracentrifugation coupled to 2DE-MALDI-TOF-MS was used for the identification of proteins that most probably are building blocks of macromolecular complexes over the size of 1 MDa. First, aliquots of the glycerol gradient fractions were loaded on 1D gels to localize subunits of known macromolecules (proteasome and thermosome) using MS, followed by pooling fractions 12-13 (A), 17-19 (B), 22-24 (C), and 26-28 (D), respectively, and analysis by 2DE (Figure 4.1.2, A-D).

The appearance or increasing intensity of a protein spot in gels B, C and D indicated that the protein sedimented as a subunit of a larger complex, while we judged that the faded appearance of a protein in these gels indicated a complex of lower Mw (i.e. proteasome, thermosome, peroxiredoxin Ta0152). Putative complex-forming proteins are listed in Table 4.1.3.

The function and structure of several of them, like the ribosome, translation initiation factor eIF-6 related protein, elongation factor 1, the DNA-dependent RNA polymerase, and the pyruvate dehydrogenase complex are well studied. The rest of the proteins belong to the categories of 'hypothetical' or 'conserved hypothetical proteins'. We found by BLAST search that Ta0316 is a twitching motility (PilT) related protein. Reportedly, PilT is a large hexameric ATPase that is required for pilin retraction, disassembly, and degradation [87]. However, Ta0316 might have other functions, as a more detailed sequence analysis using the SMART server revealed three conserved domains (http://smart.embl-heidelberg.de). The N-terminal sequence encodes for PilT N terminus (PINc) that is predicted to play role in nucleotide binding and potentially being found in RNases. The central part of the protein shows homology to AAA ATPase sequences and on the C-terminus there is a ribonucleoprotein K-homology type 1 domain – a motif found in nucleic acid (mostly RNA) binding proteins. Ta0522 shows homology to the pyridoxine biosynthesis PyroA protein. PyroA proteins form a highly conserved protein family with members in the Archaea, Bacteria, and Eukarya, and the *pyroA* gene product of *Aspergillus nidulans* is required for the biosynthesis of pyridoxine and resistance to photosensitizers such as methylene blue [88]. The other proteins like Ta0078 (CobT homologue), Ta0341 (glycogen debranching enzyme 1

related protein), Ta0890 (predicted transcriptional regulator), Ta1155 (probable ribonuclease Z) and Ta1201 (probable 3-isopropylmalate dehydrogenase) also exhibit homology to known sequences to some extent, but their 3D structure, quaternary structure, subunit composition, and biological function remains to be investigated.

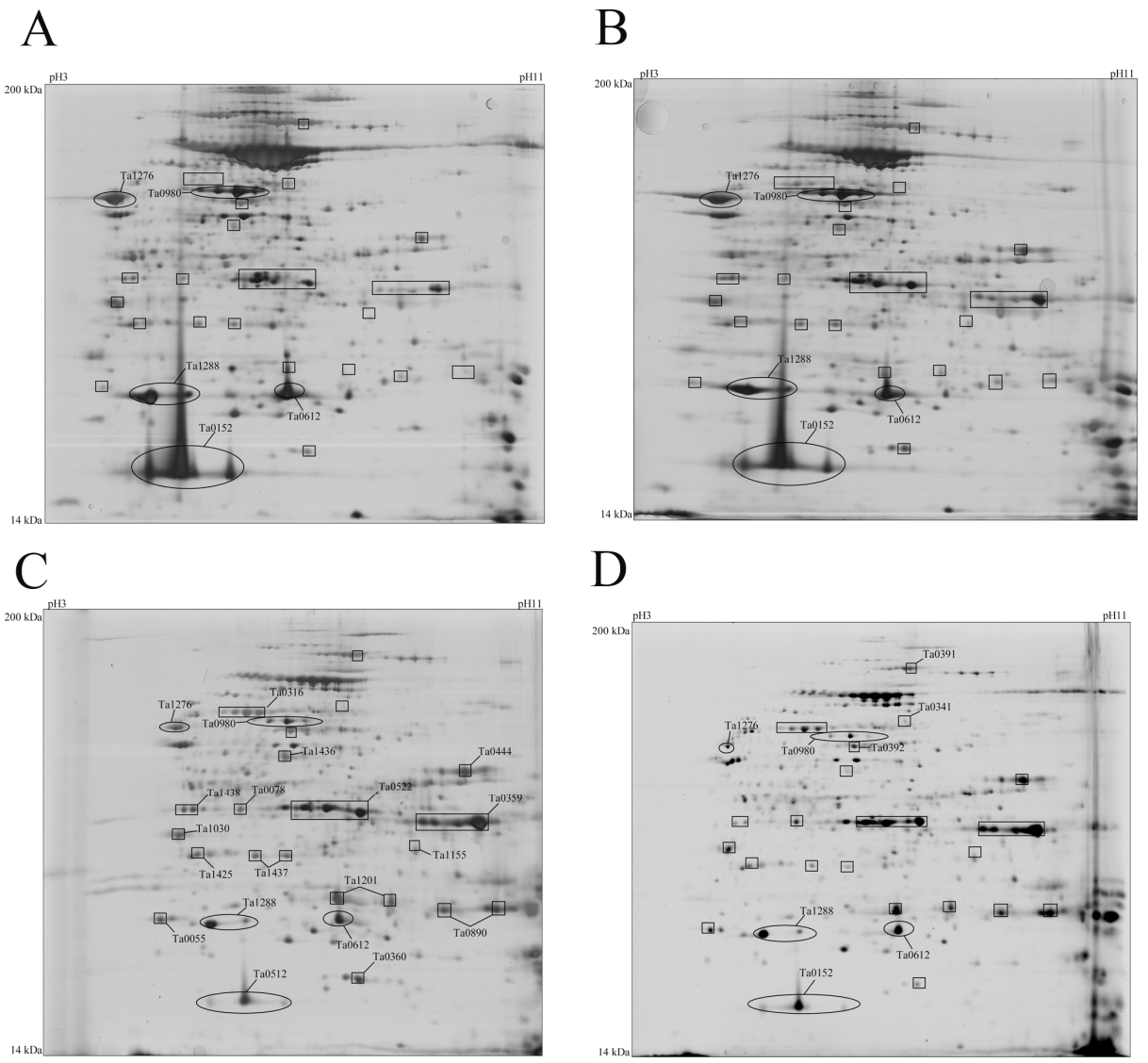


Figure 4.1.2: 2D display of cytosolic proteins of T. acidophilum from fractions obtained by glycerol gradient ultracentrifugation. Fractions 12-13 (A), 17-19 (B), 22-24 (C) and 26-28 (D) were pooled, respectively, and analysed by 2DE. The appearance or increasing intensity of a protein spot in gels B, C, and D indicates that the given protein sediments as a subunit of a large complex (marked with rectangles). Putative complex forming proteins are annotated on gels C and D. Decreasing spot intensity indicates complexes of lower Mw and as representatives of this category peroxiredoxin (Ta0152), the proteasomal (Ta0612, Ta1288) and the thermosomal (Ta0980, Ta1276) subunits are shown (marked with ovals).

Table 4.1.3 List of proteins identified from the heavy glycerol gradient ultracentrifugation fractions, which exhibited increasing intensity in four consecutive gels. The gene number (ORF), protein annotation, function and comments are indicated.

ORF	Annotation	Function	Comments
Ta0359	acidic ribosomal protein P0 related protein	protein translation factor	
Ta0055	translation initiation factor eIF-6-related protein	protein translation factor	Ta0055 and Ta0444 can be adhered to the
Ta0444	elongation factor 1	protein translation factor	ribosome.
Ta0391	probable DNA-directed RNA polymerase, chain A	DNA-directed RNA polymerase II largest subunit	Ta0391, Ta0392 and Ta1030 together with 8
Ta0392	probable DNA-directed RNA polymerase, chain A	DNA-directed RNA polymerase III largest subunit	more subunits form the active complex.
Ta1030	DNA-dependent RNA polymerase, subunit D-related protein	-	
Ta1436	probable lipoamide acyltransferase	transfer of the acetyl group to CoA	Ta1436, Ta1437 and Ta1438 together with
Ta1437	probable 3-methyl-2-oxobutanoate dehydrogenase chain E1-beta	oxidative decarboxylation of pyruvate	Ta1435 might form a complex that resembles
Ta1438	pyruvate dehydrogenase E1 component, alpha subunit	oxidative decarboxylation of pyruvate	pyruvate dehydrogenase.
Ta0078	conserved hypothetical protein (phosphoribosyltransferase)	essential for cobalamin (vitamin B_{12}) production	-
Ta0316	conserved hypothetical protein	twitching mobility (PilT) related protein	-
Ta0341	glycogen debranching enzyme isoform 1 (agl) related protein	-	-
Ta0522	probable pyridoxine biosynthesis PyroA protein	pyridoxine (vitamin B_6) synthesis	-
Ta0890	conserved hypothetical protein	predicted transcriptional regulator	-
Ta1155	conserved hypothetical protein	ribonuclease Z, metal-dependent hydrolases of the beta-lactamase superfamily III	-
Ta1201	conserved hypothetical protein	probable 3-isopropylmalate dehydratase - *Archaeoglobus fulgidus*	-

-: information not available.

4.1.2 Discussion

The work presented here gives an overview of the expressed cytosolic proteins and of the macromolecular complexes of *T. acidophilum* cultured under aerobic growth conditions at 59 °C, pH 1.5-1.8. We used the 2DE-based protein separation and MALDI-TOF-MS protein identification method and Coomassie Blue or silver staining to visualize proteins displayed in denaturing gels. Based on database search, we established a list of proteins that form complexes and for which a 3D structure has been solved, therefore allowing such complexes to serve as templates in cryo-ET pattern recognition. Additionally, we coupled 2DE-MALDI-TOF-MS to protein separation by glycerol gradient ultracentrifugation to identify proteins that are constituents of larger complexes.

The protein separation by 2DE in combination with MALDI-TOF-MS protein identification is a common method to investigate the proteome of organisms whose genome sequence is known. It was especially suitable for *T. acidophilum* as it has a relatively small genome with 1482 protein-coding ORF's and the number of protein spots that can be analysed in a 2D gel lies in the range of several thousand. Despite the distribution of many proteins in multiple spots (isoforms), the protein resolution with the chosen IEF strip and second dimensional gel size was high enough to display single proteins (over 900 spots were resolved of which 271 proteins were identified) and to obtain a satisfactory separation of acidic and basic proteins as well. The two staining methods used in our experiments were complementary to each other, as with Coomassie G250 we could detect less spots but the protein identification ratio was much higher than when working with silver. Coomassie-stained gels were analysed to distinguish highly and poorly expressed proteins.

The large majority of the proteins identified participate in fundamental biochemical pathways like energy metabolism, energy production, amino acid metabolism, purine and pyrimidine biosynthesis, replication, transcription, translation, RNA degradation, protein degradation, cell membrane biosynthesis, fatty acid metabolism and cofactor biosynthesis. A proteomics approach on a natural acid mine drainage biofilm community consisting of *Leptospirillum* and *Ferroplasma* species [89] showed similar results. Ribosomal proteins (13%), chaperones (11%), thioredoxins (9%) and proteins involved in defense against reactive radical species (8%) were also highly abundant, indicating a life style of permanent struggle against oxidative stress.

Additionally, proteins involved in amino acid metabolism, translation, and energy production and conversion, cell envelope biogenesis, coenzyme metabolism and protein folding and modification were also abundant.

The high expression level of proteasomes, chaperones (thermosome, VAT, DnaK), elongation factors, translation initiation factors, aminoacyl tRNA synthases and ribosomes in *T. acidophilum* cells indicates a high protein turnover rate. This can be due to the production of large amounts of reactive oxygen species and peroxide that can oxidize or otherwise damage cell constituents, the latter of which are mostly proteins, as these are present in the cell in the highest amounts [90]. There is an active detoxifying process in the *T. acidophilum* cells, marked with large quantities of SOD, alkyl hydroperoxide reductase and three peroxiredoxins, but it is likely that their activity is insufficient and this results in protein, RNA and DNA damage and consequently high macromolecular turnover. Supporting evidence for the fast RNA and DNA turnover can be the extremely high amount of ribonucleotide reductase that catalyzes the production of desoxyribonucleotides from ribonucleotides. This enzyme needs vitamin B_{12} for its activity [77] and we found 10 proteins of the B_{12} biosynthesis pathway indicating that most probably this vitamin is produced *de novo*.

Koonin *et al.* [72] described a superoperon of exosomal genes in *Archaea* that in addition to the predicted exosome components, encodes the catalytic subunits of the proteasome, two ribosomal proteins and a DNA-directed RNA polymerase subunit. These observations suggest that in *Archaea*, a tight functional coupling exists between translation, RNA processing and degradation, apparently mediated by the predicted exosome, and protein degradation, mediated by the proteasome. Although the RNase P subunits are missing in *T. acidophilum*, we suppose that the remaining expressed exosomal proteins are functional. It will be interesting to study their complex-forming ability and to compare these to the recently solved structure of exosome RNase PH core complex of *Sulfolobus solfataricus* [91]. In contrast to findings concerning the *S. solfataricus* exosome, we could not confirm co-sedimentation of the *T. acidophilum* exosomal counterpart with the ribosome, as there were no detectable exosomal subunits in the heavy glycerol gradient fractions.

In addition to the proteins participating in central biochemical processes, we found that 14% of the proteins belonged to the group of hypothetical/conserved hypothetical proteins. The

function of these proteins in *T. acidophilum* remains elusive. We found several proteins such as a β-galactosidase homologue (Ta1323) or Ta1060 exhibiting similarity to the bacterial atrazine-degrading chlorohydrolase that might have industrial applicability. Other proteins like Ta0247 (a homologue of carboxysome-forming proteins) and Ta0881 (a carbonate dehydratase homologue that is associated with the carboxysome) were also found. Carboxysomes are polyhedral inclusion bodies present in CO_2-fixing microorganisms [92, 93]; they have a size of 120 nm and serve to protect ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) [94]. There is no evidence for the presence of carboxysome or carboxysome-related structures in *T. acidophilum*; the function(s) of these proteins needs further investigation. We found evidence that a putative glycogen-debranching enzyme was also expressed under the given conditions, and the KI reaction indicated the presence of a starch-like polymer in the crude extract (data not shown), although we could not find a homologue of the glycogen initiation peptide. Glycogen has a globular shape and it has an average size of 40 nm [95]. The EM structure of glycogen is available [95], and it can serve as a template for our cryo-ET analyses to verify these observations. In the search for candidates for archaeal cytoskeleton, we found that the MreB homologue (Ta0583) and Ta1488 were expressed. The formation of filaments by these proteins has yet to be demonstrated *in vitro* and *in vivo*.

In conclusion, the 2DE-MALDI-TOF-MS proteomics approach provided information on macromolecular complexes of *T. acidophilum*. Using the cytoplasmic proteome analysis, we identified abundant complex-forming proteins, the structure, subunit composition and biological function of which are mostly well studied in *T. acidophilum* or in other archaea, whilst the glycerol gradient protein prefractionation resulted in the identification of higher Mw complexes expressed at low level that have not been studied in *T. acidophilum* previously.

4.2 Project 2: Protein complexes of *T. acidophilum*

4.2.1 Results

Prior to protein sample separation, the Superose 6 column was calibrated with molecular sieve Mw markers in the range of 29 kDa to 2000 kDa (Figure 4.2.1). Based on the calibration curve, fraction 22 was calculated to correspond to an apparent Mw of 300 kDa, an arbitrary threshold for protein complexes that can be analysed by cryo-ET [96]. The first proteins eluted of *T. acidophilum* cell extract appeared in fraction 11 and the last ones were in fraction 35. As we were interested in macromolecular complexes, the emphasis for the data analysis fell within fractions 11-22, although each of the protein-containing fractions (11-35) was analysed by LC-MS/MS.

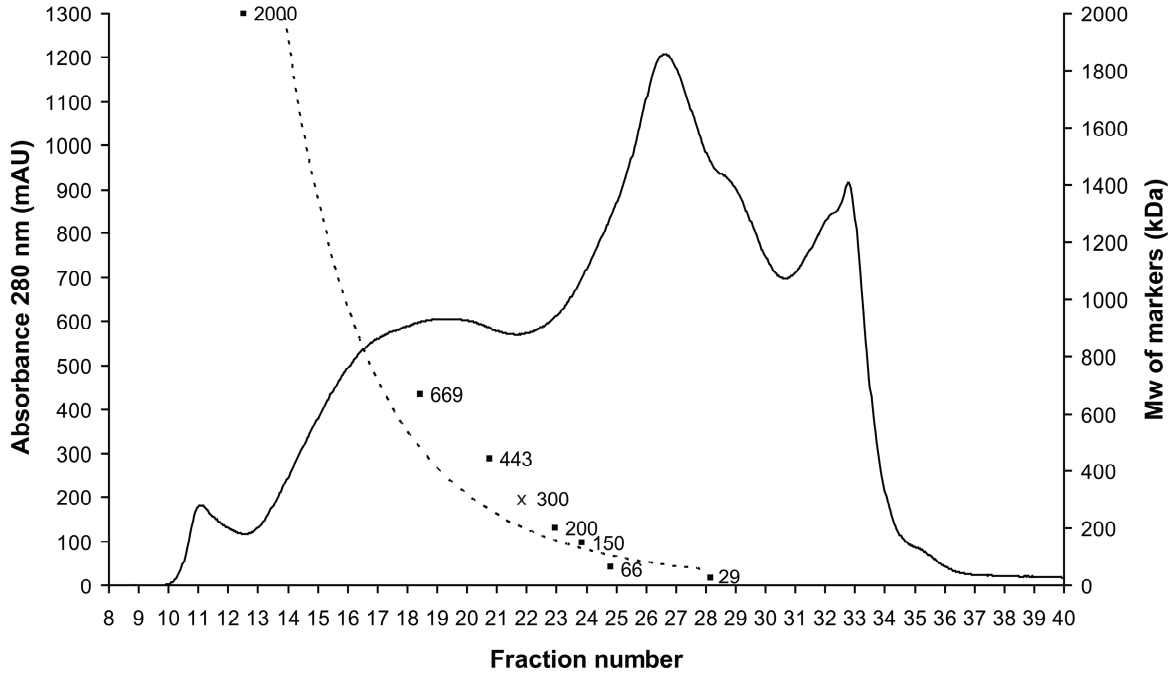


Figure 4.2.1: Elution profile of the T. acidophilum cell extract (2 mg in 200 µl) fractionated on a Superose 6 column (solid line). The dotted line indicates the standard curve calculated on the base of the elution time of the molecular sieve Mw markers. The protein elution begins at Fr11 and the last protein-containing fraction is 35. The positions of Mw markers in the fractions are: Fr11, void volume; Fr13, Blue dextran 2000 kDa; Fr19, 669 kDa; Fr21, 443 kDa; Fr23, 200 kDa; Fr24, 150 kDa; Fr25, 66 kDa; Fr29, 29 kDa. Fr22 was calculated where proteins of Mw 300 kDa elute.

We used 180 and 120 min elution gradients in preliminary peptide separation experiments. The longer gradient resulted in more identified proteins (411 vs. 175); therefore, the protein list was created based on these data, but the results of the shorter elution time experiment were also used to confirm protein identifications. In total, 521 proteins were identified from the MSC fractions. Elimination of the proteins represented by only a single peptide reduced that number to 411. These numbers for the pooled fractions were 484/359 and for the single injection 269/170, respectively. These ratios indicate the strength of size fractionation: more than twice the number of proteins were identified compared to the single injection. The pooled fractions and single injection methods were also useful in the identification of several lowly abundant proteins, and we also used these data for confirming the presence of proteins which were represented by a single peptide only. We regarded 466 proteins as identified, 464 of which are cytosolic proteins, covering 40 % of the cytosolic proteome of *T. acidophilum* (the genome of *T. acidophilum* encodes 1482 proteins, 77% of which (1146) are located in the cytosol). Forty percent of the proteins (193) appeared in the fractions over the size of 150 kDa, and 80% of these are subunits of complexes larger than 669 kDa. Most of them are ribosomal and ribosome-associated proteins that dominate fractions 13-19. Ninety proteins were identified to have Mw between 66-150 kDa reflecting a tendency to form smaller complexes (dimers, trimers, tetramers etc.). In contrast to these data, the theoretical molecular weights of the identified proteins are lower than 150 kDa, and the majority (92%) even has a Mw smaller than 66 kDa (Figure 4.2.2). For 2 proteins the Mw estimation was doubtful, due to the sole appearance in single injection experiments.

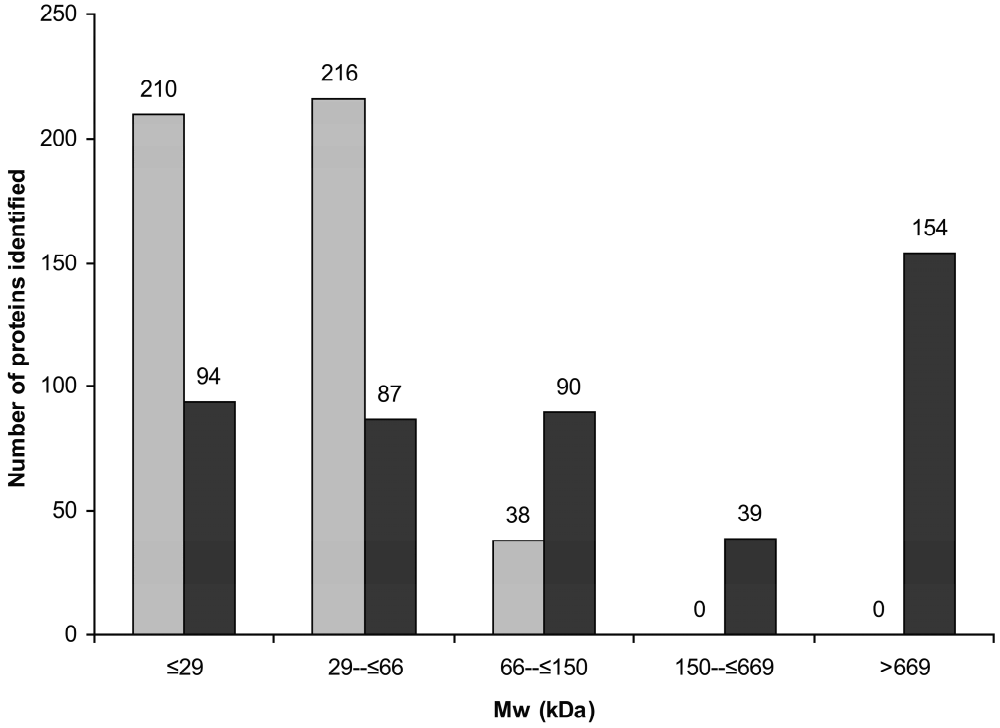


Figure 4.2.2: Comparison of the theoretical and experimental Mw distribution of the 466 identified proteins. Grey bars show the theoretical values; black bars show the experimental values.

Functional categories of the proteins identified

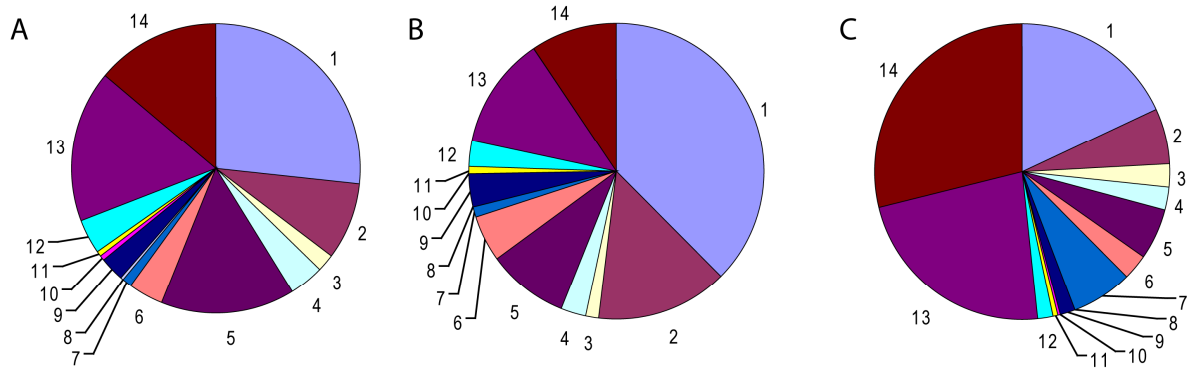
The data sets of 2DE-MALDI-TOF-MS [37] and of MSC-LC-MS/MS proteomics approaches that consist of 271 and 466 proteins, respectively, overlap by 232 proteins. The more sensitive MSC-LC-MS/MS method provided 234 new entries, but it failed to identify 39 proteins that were detected by the 2DE-MALDI-TOF-MS approach.

For comparison, the protein list of the MSC-LC-MS/MS experiment was subjected to "functional category" analysis using the MIPS database server (http://mips.gsf.de), which provides an automated protein categorization, and compared with the results of the 2DE-MALDI-TOF-MS experiment [37] and to the whole proteome [73] of *T. acidophilum* (Figure 4.2.3).

The protein functional distribution patterns of the two approaches show significant differences. A higher percentage of the proteins from the 2DE-MALDI-TOF-MS approach

falls into the functional categories of metabolism (43.10/30.40), energy (16.60/9.85), protein fate (5.53/4.71), cell rescue, defense, and virulence (4.05/2.99), when compared to MSC-LC-MS/MS experiment. In the case of transcription (3.32/4.49), protein synthesis (10.30/16.90), subcellular localization (2.95/4.28), classification not yet unambiguous (14.00/19.40), and unclassified proteins (11.00/15.80) the ratio is reversed (Figure 4.2.3).

In both cases, proteins belonging to the MIPS categories "classification not yet unambiguous" and "unclassified proteins" are underrepresented compared to the genome annotation data, while categories of the metabolism, energy, and protein synthesis are strongly over-represented (Figure 4.2.3).



	Functional categories	MSC-LC-MS/MS (466 identified proteins)		2DE-MALDI-TOF-MS (271 identified proteins)		Proteome data (1482 annotated proteins)	
		Absolute number	Relative match (%)	Absolute number	Relative match (%)	Absolute number	Relative match (%)
1	Metabolism	142	30.40	117	43.10	291	19.30
2	Energy	46	9.85	45	16.60	99	6.56
3	Cell cycle and DNA processing	10	2.14	4	1.47	40	2.65
4	Transcription	21	4.49	9	3.32	41	2.72
5	Protein synthesis	79	16.90	28	10.30	90	5.97
6	Protein fate (folding, modification, destination)	22	4.71	15	5.53	43	2.85
7	Cellular transport	6	1.28	4	1.47	109	7.23
8	Cellular communication/signal transduction mechanism	1	0.21	0	0.00	2	0.13
9	Cell rescue, defense and virulence	14	2.99	11	4.05	27	1.79
10	Cell fate	3	0.64	0	0.00	3	0.19
11	Biogenesis of cellular components	4	0.85	3	1.10	11	0.72
12	Subcellular localization	20	4.28	8	2.95	22	1.65
13	Classification not yet unambiguous	91	19.40	38	14.00	368	24.40
14	Unclassified proteins	74	15.80	30	11.00	468	31.00

Figure 4.2.3: Distribution of proteins belonging to three data sets in MIPS functional categories. A) MSC-LC-MS/MS data set (466 identified proteins), B) 2DE-MALDI-TOF-MS data set (271 identified proteins), C) proteome data set of T. acidophilum (1482 annotated proteins).

Complex-forming proteins of *T. acidophilum* over the size of 300 kDa

Aside from the classical proteomics approaches, our group has a special interest in large protein complexes of *T. acidophilum*. In the high Mw fractions (11-22), 187 proteins were found. These proteins were classified on the basis of their subunit composition or subunit

distribution in the Superose 6 fractions into the following four groups: 1) known protein complexes, which are furthermore grouped into three subgroups: 1.1 protein complexes with similar Mw to existing references, 1.2 protein complexes with different Mw to existing references, 1.3 protein complexes with unknown structures; 2) proteins that might adhere to macromolecules (DNA/RNA, protein complexes, and storage material); 3) random aggregates and/or dissociated complexes; 4) hypothetical proteins from which only limited information is available (Table S1).

1) Subunit homologues of known protein complexe

The high Mw Superose 6 fractions contained 187 proteins, and 111 of these were identified as subunits of 35 complexes. Most of these complexes have been studied in *T. acidophilum* or in other organisms with respect to their function and structure; therefore, a figure was created that shows the 3D structure of 13 complexes (Figure 4.2.4). The DNA-directed RNA polymerase complexes, exosome, ferredoxin oxidoreductase (Ta0259, Ta0260), glutamine synthetase, ornithine carbamoyltransferase (OTCase), peroxiredoxin (Ta0152), proteasome, pyruvate dehydrogenase/2-oxoacid oxidoreductase (Ta1435-1438), ribosome, succinyl-CoA synthetase, thermosome, tricorn protease, and VAT ATPase have been found and discussed previously [37]; therefore, only the newly found complexes are discussed here. For a few complexes, size determination was difficult due to rapid dissociation and the appearance of the subunits in smaller fractions.

1.1) Protein complexes with similar Mw to existing references

In addition to the well studied peroxiredoxin Ta0152 (600 kDa), three other peroxiredoxin homologues, Ta0473 (150 kDa), Ta0954 (300 kDa), and Ta1368 (subunit size), were found. As the sizes of Ta0152 and Ta0954 are in the range that can be studied by cryo-ET, this might open ways for structural analysis of their interactions with their regenerating partners, which will be very informative in combination with studies on substrate specificities and expression levels at different growth conditions to reveal their biological functions.

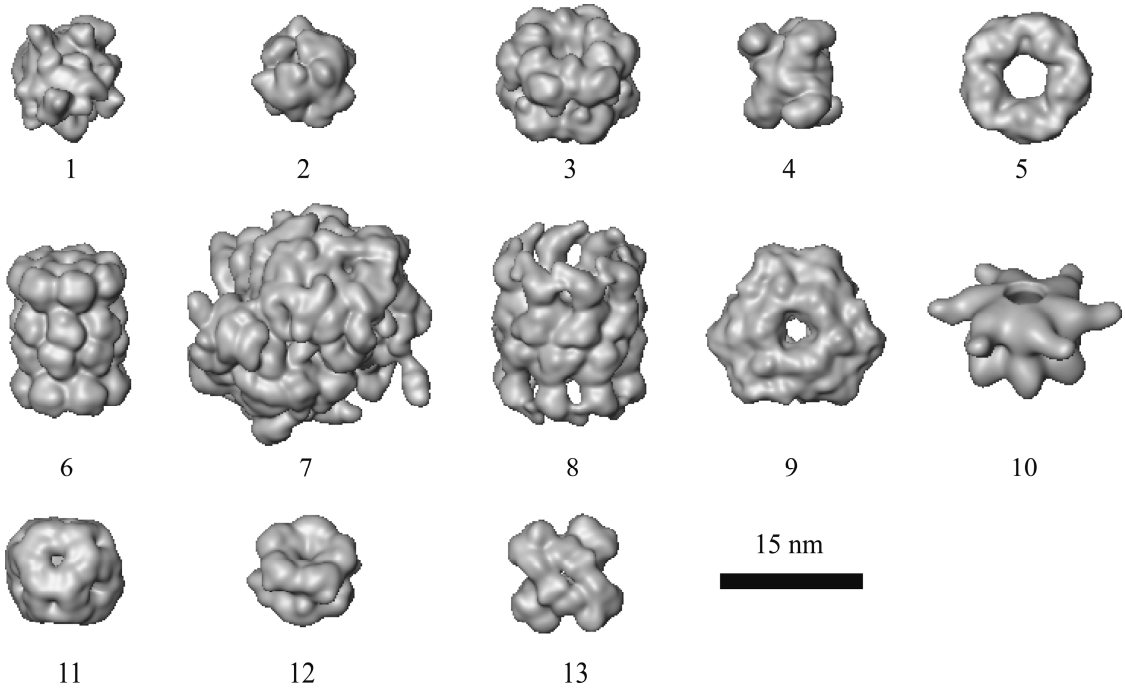


Figure 4.2.4: 3D structures of protein complexes. 1. DNA-directed RNA polymerase complexes (Thermus aquaticus [97]), 2. exosome (Sulfolobus solfataricus [98]), 3. glutamine synthetase (Salmonella typhimurium [99]), 4. ornithine carbamoyltransferase (Pyrococcus furiosus[66]), 5. peroxiredoxin (Aeropyrum pernix K1 [100]), 6. proteasome (T. acidophilum [61]), 7. ribosome (Thermus thermophilus), 8. thermosome (T. acidophilum [74]), 9. tricorn protease (T. acidophilum [101]), 10. VAT ATPase (T. acidophilum [64]), 11. small heat shock protein (Methanococcus jannaschii [102]), 12. transaldolase (Ta0616) (Thermotoga maritima), 13. glutamate dehydrogenase (Ta0635) (Thermococcus litoralis [103]).

The genes of Ta0157 and Ta0158 overlap by 25 bp, substantiating transcriptionally coupled gene expression. The two gene products were identified in the same fractions (>1 MDa), and they presumably correspond to a complex. Basic Local Alignment Search Tool (BLAST) searches revealed significant homologies to the complex-forming SbcC and SbcD proteins of *E. coli*. These have a wide variety of DNA repair and maintenance functions, including homologous recombination and non-homologous end joining. The sedimentation equilibrium analysis of the SbcC/SbcD complex detected two species with respective molecular masses of 720 (in the absence of Mn^{2+}) and 1210 kDa (in the presence of Mn^{2+}). The 720 kDa species represents a hexameric SbcC complex, whereas the size of the 1210 kDa complex is close to the predicted $SbcC_6$ $SbcD_{12}$ (1249 kDa) subunit composition [104]. The EM image of the SbcCD complex shows a structure consisting of two globular domains linked by a long filamentous rod [105]. The size of the respective *E. coli* and *T. acidophilum* complexes are very similar. Therefore, we assume that Ta0157 and Ta0158 have a function and quaternary structure

similar to SbcCD. Ta0787 (chromosome segregation protein related protein) was also found in the same fractions as SbcCD homologues, and it shows high sequence similarity to Ta0157 on the N- and C-terminus. Further experiments are needed to deduce whether the co-occurrence of the three proteins is due to specific protein-protein interactions or just coincidence.

The small heat shock protein (sHSP) Ta0471 assembles into a 600 kDa complex. It is typical for sHSP's to have a relatively small subunit size and to form large oligomeric assemblies of 200-800 kDa. Most of them act as molecular chaperones, having the ability to prevent aggregation of denatured proteins under stress conditions [102, 106]. The elements of the Hsp70 chaperone system (Ta1087 (DnaK), Ta1088 (DnaJ), and Ta1086 (GrpE)) were also found. This system has many cellular functions and we refer the reader to the review by W. A. Houry [107] for more details. DnaK usually occurs as a monomer, and sometimes as a dimer, while DnaJ and GrpE exist mostly in homodimeric forms. In the MSC fractions Ta1087, Ta1088 and Ta1086 appeared at approximate sizes of 600 kDa, >1 MDa, and 150 kDa, respectively, and these data did not indicate interactions among the three proteins. The huge size of Ta1088 might be due to its adhesion to the ribosome via the nascent target protein, but there is currently no definite explanation for the sizes of Ta1087 and Ta1086.

Ta0616 exhibited significant homology to transaldolases, and its calculated size was approx. 300 kDa, similar to the native transaldolase of *Methanocaldococcus jannaschii,* which is a decamer with a molecular mass of 271 +/- 27 kDa [108]. Transaldolases participate in the non-oxidative branch of the pentose phosphate pathway, catalyzing the reversible transfer of dihydroxyacetone (derived from fructose-6-phosphate) to erythrose-4-phosphate to yield sedoheptulose-7-phosphate and glyceraldehyde-3-phosphate [109].

Glutamate dehydrogenases occupy a pivotal position in the metabolism of most organisms as they provide a link between carbohydrate and nitrogen metabolisms. The calculated size of the *T. acidophilum* counterpart Ta0635 is 300 kDa. This is in good agreement with the size of glutamate dehydrogenase (270 kDa) of the hyperthermophile *Thermococcus litoralis,* which was shown to be a hexamer of 46 kDa subunits [103]. The other glutamate dehydrogenase homologue Ta0776 (79% identical to Ta0635) was present only in fraction 24, indicating a slightly smaller size than Ta0635, and also in agreement with a hexameric conformation.

1.2) Protein complexes with Mw different from existing references

We classified those complexes into this group, that exhibited (generally much bigger) Mw different from existing references. These alterations might origin from real structural differences, but we can't exclude severe aggregation processes as the cause of these large sizes.

Ta0522 was found in fractions 12-20, indicating Mw of >1 MDa, and it exhibited extensive homology to the Pdx1 subunit of pyridoxal 5´-phosphate (PLP) synthase. PLP synthase catalyzes the deoxyxylulose 5-phosphate-independent *de novo* biosynthesis of vitamin B6 and it is built up of Pdx1 (synthetase) and Pdx2 (glutaminase) subunits [110]. Twelve Pdx1 subunits of *Bacillus subtilis* form a double hexameric ring having a Mw of 370 ± 11 kDa, to which 12 Pdx2 glutaminase subunits attach themselves like the cogs of a cogwheel, interacting only with the Pdx1 subunits but not with each other [111]. The size of the complex is 576 ± 20 kDa, and the interaction of Pdx1 and Pdx2 is believed to be transient in nature [112]. On its own, Pdx1 is present in solution as two species with molecular masses of 186 ± 8 and 370 ± 11 kDa, corresponding to the hexameric and dodecameric forms, respectively, the dodecamer being the predominant species. Ta0009 shows homology to Pdx2, but this protein was detected only in low Mw fractions and could not be shown to be related to Ta0522. This observation does not exclude the possibility that these two proteins build a complex, since the interaction can be weak, and complex-stabilizing additives (like glutamine in case of the Pdx1Pdx2 complex) should be used in the separation buffer to preserve the intact structure. Nevertheless, we suppose that the quaternary architecture of Ta0522 is different from the known structure of the Pdx1Pdx2 complex as without Ta0009 it exhibited a larger size than the Pdx1Pdx2 complex of *B. subtilis*.

Ta0574 and Ta0575 are the respective putative homologues of the regulatory and catalytic subunits of the aspartate carbamoyltransferase that catalyzes the first step of pyrimidine biosynthesis - the condensation of carbamoyl phosphate and aspartate to form *N*-carbamoyl-L-aspartate and inorganic phosphate. The holoenzymes in *E. coli* and in *Sulfolobus acidocaldarius* are composed of six catalytic chains (Mw 34 kDa each), grouped into two trimeric catalytic units, and six regulatory chains (Mw 17 kDa each) forming three homodimers, which add up to a 310 kDa complex [113-116]. Ta0574 and Ta0575 were found in

both high (>1 MDa) and low Mw (subunit size) fractions, that might reflect a different and unstable subunit architecture that partially dissociates during purification.

Ta0756 (muconate cycloisomerase related protein) shows homology to muconate lactonizing enzyme from *Pseudomonas putida*. Muconate lactonizing enzyme is the second enzyme involved in the degradation of catechol to beta-ketoadipate and catalyzes the cycloisomerization of *cis,cis*-muconic acid to muconolactone[117]. In *Pseudomonas putida* it is an octamer with a Mw of 320 kDa [118]. Ta0756 was identified as a complex larger than 1 MDa indicating a different structure compared to its homologue.

Ta1066, Ta1318 and Ta1044 are homologous to the PurL, PurQ and PurS subunits of formylglycinamide ribonucleotide amidotransferase (FGAR-AT), respectively. This complex catalyzes the conversion of FGAR, ATP, and glutamine to formylglycinamidine ribonucleotide (FGAM), ADP, Pi, and glutamate in the fourth step of the purine biosynthetic pathway. The size of the PurLQS complex in *Bacillus subtilis* [119] is much smaller (2PurS/1PurL/1PurQ, 153 kDa) compared to Ta1066, Ta1318 and Ta1044 (~1MDa); therefore, the subunit composition and structure of the *T. acidophilum* enzyme must be determined experimentally.

Ta1073 probably belongs to one of the biggest complexes of *T. acidophilum* with Mw over 1 MDa. It shows high similarity to uridine 5'-diphosphate-sulfoquinovose synthase **(**SQD1) from *Arabidopsis thaliana*. The SQD1 protein is the key enzyme involved in the formation of the sulfolipid head group precursor, uridine 5'-diphosphate (UDP)-sulfoquinovose, from UDP-glucose and sulfite. It is proposed that SQD1 forms a 250 kDa complex *in vivo* with additional accessory proteins [120, 121], the nature of which is still unknown. The Ta1073-containing complex is much bigger in size than its homologue; therefore the structure, the possible complex forming partners and the biochemical function(s) need to be determined.

Ta1164 is an orotate phosphoribosyltransferase (OPRTase) homologue that takes part in the pyrimidine metabolism pathway catalyzing the magnesium-dependent conversion of alpha-D-phosphoribosylpyrophosphate (PRPP) and orotate to orotidine 5'-monophosphate (OMP) and pyrophosphate. Its Mw was determined as being larger than 1 MDa. OPRTase and orotidine 5'-monophosphate decarboxylase (OMPDC) of *Plasmodium falciparum* form a multi enzyme

complex of 140 kDa [122]. The OPRTase from *T. acidophilum* seems not to form a stable complex with the OMPDC counterpart Ta0073, as the latter was found only in low Mw fractions.

Ta1194 was found as a high Mw (>1 MDa) complex in the molecular sieve chromatograpy experiment. The protein encodes for the phosphopantetheine adenylyltransferase (PPAT) domain on the N-terminus and for a conserved domain with unknown function (DUF84) on the C-terminus. PPAT is an essential enzyme that catalyzes a rate-limiting step of coenzymeA (CoA) biosynthesis [123], and its quaternary structure varies from organism to organism: it can be a 18 kDa monomer (*Pyrococcus abyssi)* [123], a 108 kDa trimer (*Brevibacterium ammoniagenes*) [124] or a 108 kDa hexamer (*E. coli and Thermus thermophilus* HB8) [125]. The structure and function of Ta1194 most probably differs from these proteins as it encodes for the DUF84 domain as well; therefore, functional and structural investigations on the purified complex by biochemical methods and by single particle analysis are crucial.

The amino acid (AA) sequence of Ta1210 is highly homologous to trehalose-6-phosphate synthase (OtsA) of *E. coli*. OtsA is a UDP-glucose (UDP-Glc)-dependent glycosyl-transferase, which catalyzes the synthesis of trehalose-6-phosphate using UDP-Glc as donor and the alpha anomer of glucose-6-phosphate as the acceptor. Trehalose-6-phosphate is subsequently dephosphorylated by trehalose-6-phosphate phosphatase (OtsB) to yield trehalose. In *E. coli*, OtsA and OtsB exist as two separate enzymes, and OtsA is present in an equilibrium of dimers and tetramers (~200 kDa) [126]. Interestingly, in *Saccharomyces cerevisae*, trehalose is synthesized by a large multisubunit complex (600-800 kDa) comprising the two catalytic activities trehalose-6-phosphate synthase (TPS) and trehalose-6-phosphate phosphatase (TPP) [127]. Four genes have been identified that appear to encode components of the trehalose synthase complex: TPS1, TPS2, TPS3, and TSL1. TPS1 is homologous to Ta1210 over its entire sequence. For TPS2, the C-terminal part shows homology to Ta1209, and the N-terminal part shows homology to Ta1210. Both TPS3 and TSL1 show homology to Ta1210 in the middle part of the sequences. We found that Ta1210 of *T. acidophilum* appears as a complex with a Mw>1 MDa, but it remains unclear if Ta1209 – an OtsB homologue – is part of the assembly, as we were not able to identify it among the expressed proteins.

Ribulose-phosphate 3-epimerase (RPE) (Ta1315) catalyzes the inter-conversion of ribulose-5-phosphate to xylulose-5-phosphate in the Calvin cycle and in the oxidative pentose phosphate pathway. The RPEs from *Oryza sativa* (48 kDa) [128] and *Plasmodium falciparum* (51 kDa) [129] are dimers. The RPEs from *Solanum tuberosum* (148 kDa) [130] and *Synechocystis* (150kDa) [131] are described by a conserved hexameric quaternary structure. The size of the RPE homologue of *T. acidophilum* (Ta1315) is over 1 MDa, which seems unique compared to the structures known.

Ta1476 and Ta1478 were found in the same fractions, having an approximate Mw of 600 kDa. Both enzymes participate in the one-carbon pool by folate biochemical pathway; Ta1476 is annotated as glutamate formiminotransferase and Ta1478 as formate-tetrahydrofolate ligase. Ta1478 shows high sequence similarity to N^{10}-formyltetrahydrofolate synthetases (FTHFS) from *Clostridium cylindrosporum*, which catalyze the ATP-dependent activation of the formate ion via its addition to the N^{10} position of tetrahydrofolate. FTHFS is a highly expressed key enzyme in both the Wood-Ljungdahl pathway of autotrophic CO_2 fixation (acetogenesis) and the glycine synthase/reductase pathways of purinolysis. In *Clostridium cylindrosporum* FTHFS has a homotetrameric structure with a native mass of about 240 kDa [132, 133]. Whether Ta1476 and Ta1478 interact with each other and associate with Ta1477 (a formiminotransferase cyclodeaminase-related protein which was present in smaller fractions, and whose gene overlaps with Ta1478) remains elusive. A BLAST search with the NP_990234 formiminotransferase cyclodeaminase sequence of *Gallus gallus* against the *T. acidophilum* protein database clearly indicates that its N-terminal sequence is highly similar with Ta1476, and the C-terminal sequence is highly homologous to Ta1477.

1.3) Protein complexes with unknown structures

Ta0326 (FixC), Ta0327 (FixX), Ta0328 (FixA), and Ta0329 (FixB) display significant amino acid sequence homology to polypeptides encoded by the *fixABCX* operon of *Rhodospirillum rubrum*, which encodes a putative membrane complex participating in an electron transfer chain by shuttling electrons to nitrogenase [134]. The *fixAB* gene products have high sequence similarity to the mammalian electron transfer flavoprotein subunits beta and alpha, respectively, suggesting a role for these proteins in redox reactions [135, 136]. The size distribution of the four subunits from *T. acidophilum* is not typical. The estimated size of the

complex is 550 kDa, but probably due to protein precipitation and complex dissociation, the subunits are present in both high and low Mw fractions. We did not find evidence that the complex is membrane-bound; therefore, it would be an ideal target for structural and functional studies.

The size distribution of Ta0825 was similar to that of the FixABCX complex. This protein shows homology to glucokinases and to proteins of the ROK (Rho kinase) family, and to NagC, a transcriptional regulator. To predict from the size distribution of Ta0825 whether it acts as a regulatory protein or as a glucokinase is difficult, but there is a chance that it has DNA-binding properties as well.

The calculated size of Ta0882 is approximately 300 kDa. This protein shows extensive homology to enolases, which are homodimeric enzymes that catalyze the reversible dehydration of 2-phospho-D-glycerate to phosphoenolpyruvate as part of the glycolytic and gluconeogenesis pathways. It remains elusive why the size of Ta0882 complex is four times that of the supposed dimeric size.

The formate dehydrogenase related protein Ta0425 respectively encodes for one 2Fe-2S and one 4Fe-4S iron-sulfur cluster binding domain. Although the central and C-terminal sequences of the protein are highly homologous to formate dehydrogenase H that catalyzes the reversible oxidation of formate to CO_2 with the release of a proton and two electrons, the N-terminal sequence more closely resembles the Fe-S protein subunit of succinate dehydrogenases/fumarate reductases. The whole length protein also showed significant homology to uncharacterized anaerobic dehydrogenases (COG3383) [48]. This protein was detected together with Ta0424, corresponding to a Mw of 440 kDa. The complex might contain Ta0424 subunits; on the gene level, the C-terminus of Ta0425 and the N-terminus of Ta0424 overlap by four bases. Ta0424 overlaps with Ta0423 (an uncharacterized protein required for formate dehydrogenase activity) on the C-terminus, indicating an operon structure, but the latter protein escaped detection if it was expressed at all.

Ta1413 was found in nearly the same fractions as Ta1412 among proteins having a size around 1 MDa. Ta1413 encodes for an N-terminal pyridine nucleotide-disulphide oxidoreductase domain within a larger NAD(P)H-nitrite reductase domain and it is annotated

as sulfide dehydrogenase [73], while Ta1412 is a hypothetical protein. It might well be accidental that these two proteins were found in similar fractions, but it cannot be excluded that they belong to a single complex.

2) Proteins that might adhere to macromolecules

A plethora of proteins were found among large complexes, the structures of which would not correspond to known sizes unless they adhered to macromolecules like DNA/RNA, ribosomes, or storage material(s).

Eight of these proteins can sediment together with the ribosome, being ribosome-binding proteins taking part in the ribosomal structure processing or in protein translation (Table S1). We found a large set of nucleotide-binding proteins (29) which probably do not form supramolecular assemblies; rather, they are bound to their target DNA or RNA sequences, which results in an increased apparent Mw.

Two (Ta0339, Ta0341) out of three neighboring gene products were detected in the void volume, indicating a huge size, while the third protein (Ta0342) had a much lower Mw. The KEGG pathway database places Ta0339 and Ta0342 in the starch and sucrose metabolism pathway as an alpha-amylase related protein, producing dextrin from starch and/or glycogen, and a putative glucoamylase producing α-D-glucose from dextrin or glycogen, respectively. Ta0341 shows homology to GDE_C amylo-1,6-glucosidase and GDB1 glycogen de-branching enzymes. It is unclear whether Ta0339 and Ta0341 form a complex and/or if they adhere to yet unspecified storage material(s). The other genes in their vicinity encode for a glycosyl asparaginase-related protein (Ta0338, Fr27, 28), a homologue of RfaG glycosyltransferase (Ta0340, not found), a glycoamylase/glycosyl hydrolase (Ta0342, Fr21, 22), and a WcaG nucleoside-diposphate-sugar epimerase (Ta0343, Fr24, 25, 26), respectively, indicating an active but yet uncharacterized biochemical pathway.

3) Random aggregates and/or dissociated complexes

After analyzing all of the fractions, we found 18 proteins which were present in both void volume fractions (11-12) and in low Mw fractions (25-27). The reasons for this protein distribution pattern can be diverse: 1) labile complexes dissociate during separation; 2) there

are low Mw proteins that randomly aggregate and form protein clumps; or 3) a group of proteins might be present in the cytosol in association with large complexes, or in free form.

The degradation of complexes and random aggregation of proteins were proved experimentally. Portions of the cell extract were kept at room temperature for 16 and 40 hours, respectively, and then subjected to Superose 6 fractionation. By comparing the three elution profiles (0, 16 and 40 hours), we found that the amount of proteins in the void volume and in the low Mw fractions were increasing with time, and that the size of the intermediate peaks were decreasing, indicating both random aggregation and protein dissociation (Figure 4.2.5).

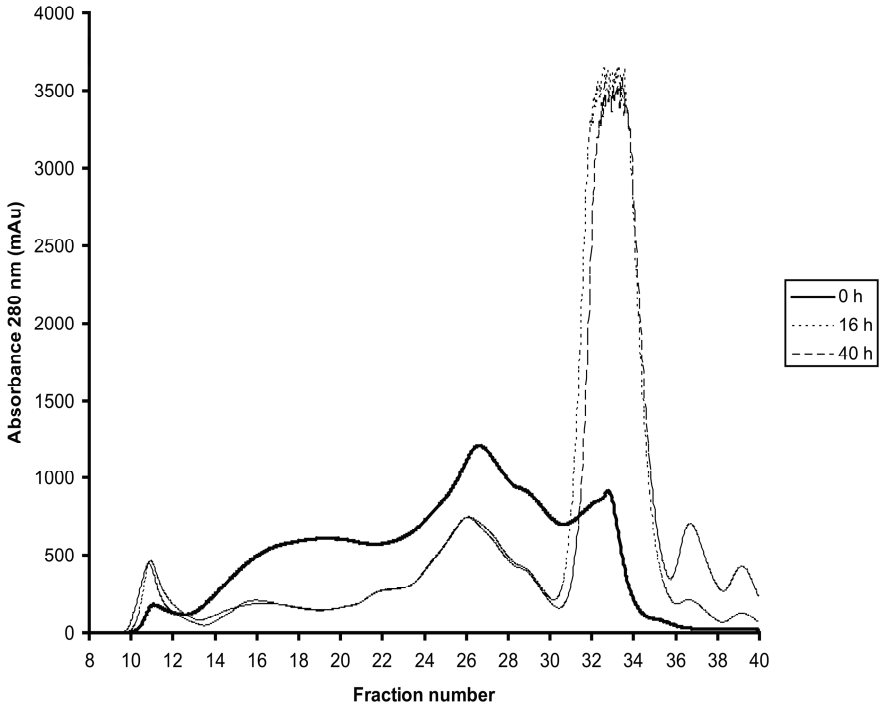


Figure 4.2.5: Changes in the elution profile of T. acidophilum cell extract (4 mg in 200 µl) on a Superose 6 column: 0 h (—), after 16 h (....), and 40 h (- - -) incubation time at RT. The increase in peak height in the void volume (8 ml) indicates the formation of random aggregates. The decrease in the intermediate Mw fraction and increase in the low Mw fraction indicates the dissociation of protein complexes over time.

4) Hypothetical proteins

The fourth group contains 18 hypothetical proteins. The results of the homology searches are listed in Table S1. Several of these proteins encode DNA or RNA binding/processing

domain(s), and some of them have protein binding functions. A significant number have no putative homologues, and thus their 3D structure, quaternary structure, subunit composition and biological function remain elusive.

Ta1412 and Ta0424 are small proteins that belong to the cluster of orthologous group (COG)2427, encoding for the 40 AA uncharacterized conserved motif DUF1641 (Ta1125 also has the same motif, but it is not expressed). Ta0424 probably forms a complex with Ta0425, and Ta1412 with Ta1413, as discussed earlier. To substantiate the roles of these proteins within the respective complexes, detailed structural and functional analyses are needed.

We were also able to show that Ta0159 is expressed and that it has an ATPase domain, but to some extent it also exhibits homology to the DNA segregation protein FtsK. Whether it forms a complex with Ta0157 and Ta0158 (SbcCD homologues that have ATP-dependent double-stranded DNA exonuclease activity) remains elusive, as it was present in smaller fractions compared to the other two proteins.

4.2.2 Discussion

In a previous study, we attempted to carry out a detailed analysis of the protein complexes of *T. acidophilum* [37], but due to the low expression level of large Mw complexes and the relative insensitivity of the 2DE-MALDI-TOF-MS analysis, only a limited number of complexes were found. In this study, the cytosolic proteins of *T. acidophilum* were fractionated by molecular sieve chromatography and the protein-containing fractions were analysed by LC-MS/MS. The number of the proteins identified was increased by 195 with this method, and especially the identification efficiency of low Mw was increased (Figure 4.2.6). The protein number increased in almost each of the functional categories, but the most dramatic changes were in the groups of protein synthesis (ribosomal proteins) and unclassified proteins.

Our MSC-LC-MS/MS experiments unambiguously identified two proteins (Ta0494 and Ta1171) that were present among the originally annotated *T. acidophilum* proteins, but eliminated from the list by a recent curation (http://pedant.gsf.de). These findings urge the curation of the *T. acidophilum* database with respect to proteomics data.

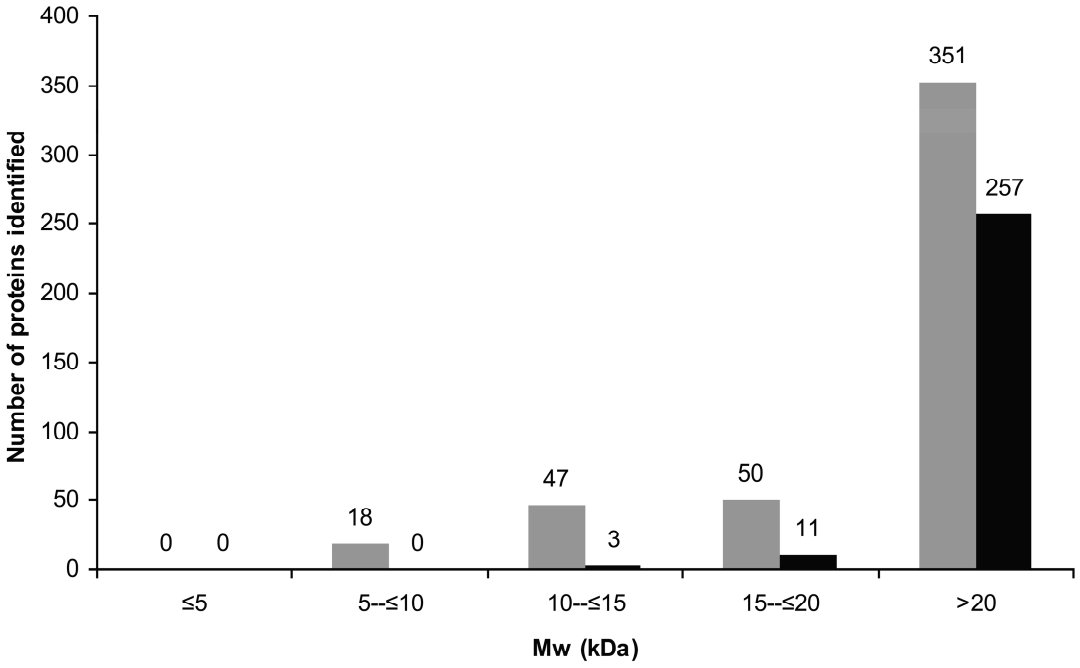


Figure 4.2.6: Comparison of the theoretical Mw distribution of the 466 and 271 proteins identified by MSC-LC-MS/MS (grey bars) and 2DE-MALDI-TOF-MS (black bars), respectively. The MSC-LC-MS/MS approach has better identification rate in the low Mw range (5-20 kDa) than the 2DE-MALDI-TOF-MS method.

Interestingly, there were 39 proteins (expressed at low levels) detected by the 2DE-MALDI-TOF-MS analysis that escaped detection by LC-MS/MS. The reason for this can be that in 2D gels the proteins are concentrated in distinct spots and minor proteins can also be detected if the sample complexity is low. On the other hand, the Superose 6 separation distributes the proteins in several fractions, diluting the concentration down to a range of which they can't be detected.

Additionally, these poorly expressed proteins can be masked by highly abundant proteins during LC separation. A good example for this group of proteins is Ta0583, an actin homologue that we have investigated in detail [137]. It is a minor protein, constituting only 0.035% of the cytosolic proteins as determined by immunostaining.

Another problem of MSC is the appearance of highly abundant proteins in multiple fractions. The best example for this is Ta1475 (the most abundant protein expressed under aerobic conditions) which was found in each of the investigated fractions, making size determination

difficult. The second effect that influences protein size determination is that not only the protein Mw but also the protein shape influences the migration/elution from the MSC. For instance, Ta1330, Ta0085 and Ta0481 eluted in fractions 23-25, indicating Mw 200 kDa, but the Mws of these complexes are around 400-500 kDa (unpublished data). These points of our approach can be criticized, but there are few methods available to separate native proteins on the basis of their size.

We observed that a number of proteins were distributed both in large and small but not in the intermediate fractions. This phenomenon may be due to slow dissociation of weak complexes, where the remnants of the complexes are present in the large fractions, whereas the subunits appear in small fractions. Proteins that temporarily bind to macromolecules can exhibit a similar distribution pattern that is difficult to recognize in the case of hypothetical proteins. Therefore, individual investigations of these proteins and biochemical pathways will be needed. The third process that causes similar protein distribution is the random aggregation of monomers and dimers to produce large, heterogeneous clumps that appear in the void volume. To maintain the integrity of complexes and to avoid random aggregation of proteins, the addition of reagents like glycerol, glycine, mannitol, ectoine, salts, metal ions etc. will be needed, and the ideal composition of the buffer must be determined experimentally. There might be no single buffer composition that works for each of the complexes and most probably there will be complexes that need to be treated individually.

The MSC-LC-MS/MS can be an alternative method to study/confirm the subunit composition of known complexes. A good example for this is the exosome. In comparison to the native exosome complex of *Sulfolobus solfataricus* which is constituted from the DnaG-like protein besides the Rrp4, Rrp41, Rrp42, and Csl4 subunits [138], it is obvious that the *T. acidophilum* exosome is only comprised of homologues of the four latter proteins, Ta1291, Ta1293, Ta1294 and Ta0929, respectively. The DnaG-like homologue Ta0097 is present in much higher Mw fractions; therefore, we can not substantiate that this protein represents a core exosome protein in *T. acidophilum*.

We further strengthen the hypothesis that, in archaea, tight functional coupling exists between translation, RNA processing and degradation (apparently mediated by the predicted exosome), and protein degradation (mediated by the proteasome) [72], and we also pointed this

out previously [37]. We analysed the gene organization downstream of the putative exosome subunits (Ta1292-1294). There are six open reading frames in this region, which encode for proteins taking part in the above processes/complexes. Ta1295 and Ta1297 are ribosomal proteins (L37Ae and L21e homologues, respectively), Ta1296 (not expressed) is a predicted pseudouridylate synthase with strong partial homology to human dyskerin. Ta1298 shows homology to the small subunit of the DNA-dependent RNA polymerase II, Rpb4 (it was present in fractions 21 and 28, and therefore it is difficult to determine its size, or its complex formation with the RNA polymerase complex). Ta1299 is a predicted RNA-binding protein (Fr31-35), while Ta1300 (not expressed) shows homology to KsgA (dimethyladenosine transferase) that takes part in rRNA methylation. As Ta1298 is a putative homologue of the Rpb4 subunit of DNA-dependent RNA polymerase II and is encoded in the mega operon, we can extend the theory of the tight functional coupling of translation, RNA processing and degradation, and protein degradation with transcription in *T. acidophilum*.

This approach gave a good overview regarding protein complexes of *T. acidophilum* that can serve as templates in cryo-ET experiments. We have identified 35 complexes, among which the structure and function of many have been studied in other organisms, but their structure and function should also be determined in *T. acidophilum*. Besides these complexes, we identified 18 hypothetical or putative proteins that need further experimental investigations.

4.3 Project 3: Quantitative proteomics and transcriptomics analysis of *T. acidophilum* cultured under aerobic and anaerobic conditions

4.3.1 Results

4.3.1.1 Evaluation of proteomics data

In this study, the cytosolic proteome of *T. acidophilum* was investigated by a combination of one-dimensional gel electrophoresis and on-line electrospray tandem mass spectrometry (GeLCMS). A label-free mass spectrometric approach comparing the ion intensities of identical tryptic peptides [34] was used for proteomics quantitation of *T. acidophilum* cultured under aerobic and anaerobic conditions.

Identification of proteins expressed under aerobic and anaerobic conditions

In this proteomics approach, a total of 1025 proteins were identified, 1004 of which are cytosolic proteins, covering 88% of the *T. acidophilum* cytosolic proteome (1146 ORFs). One hundred and three of the 142 undetected cytosolic proteins are hypothetical and/or hypothetical proteins with unknown functions, while the remaining 39 proteins are predominantly involved in metabolism (14 proteins), energy (5 proteins), cell cycle and DNA processing (5 proteins). The reason why these proteins escaped detection could be their small size (23 of them have Mw less than 10 kDa, e.g. ribosomal proteins Ta1241, Ta1257 and Ta1317a) and/or their low expression level (in the microarray experiment, the intensity values of 38 transcripts are between 7 and 9 on a scale of 7-16).

The functional distribution of the cytosolic proteins identified in comparison to the genome-predicted cytosolic proteome is shown in Figure 4.3.1. These results demonstrate that the experimental proteome analysis is comprehensive, with no evident bias for or against specific protein classes.

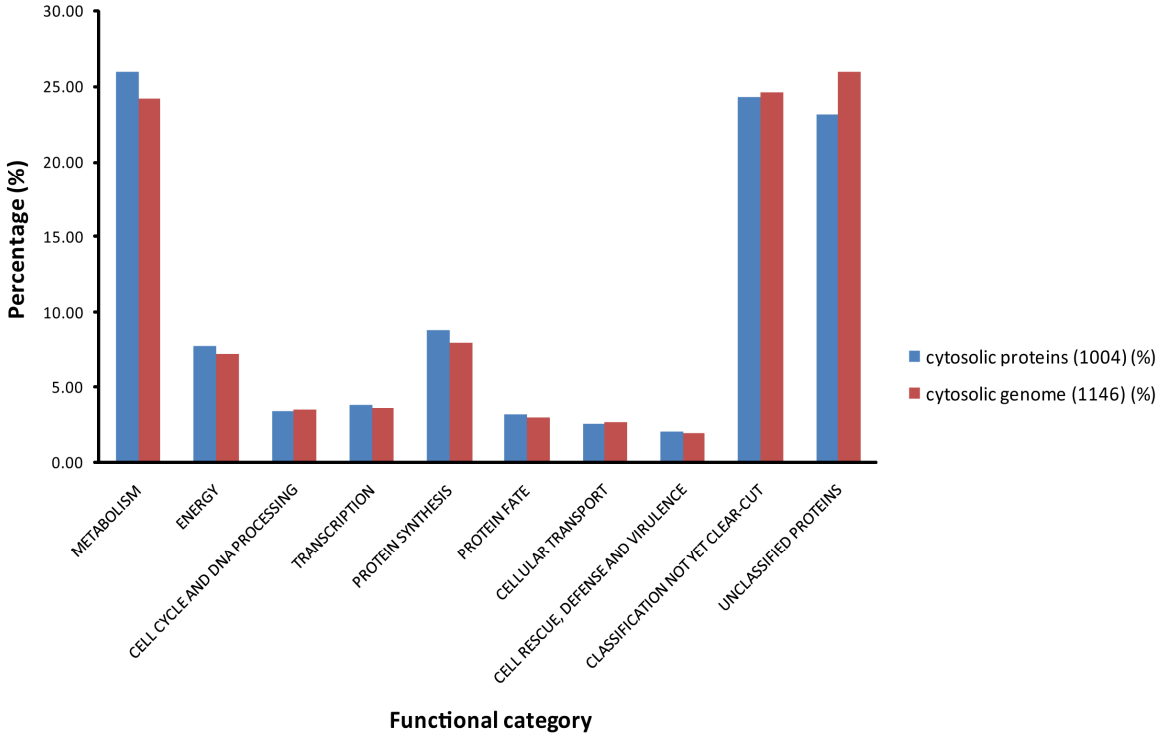


Figure 4.3.1: Functional distribution of identified cytosolic proteins vs. theoretical cytosolic genome. The MIPS database server was used for automated categorization.

In spite of the tremendous increase in the number of proteins identified using the GeLC-MS method compared to the 2DE-MALDI-TOF-MS approach (1025 vs 271 proteins) [37], the same chromosomal gap was found after matching the proteins identified to the theoretical ORFs, i.e. 40 proteins at the chromosome region starting from Ta0697 to Ta0750 escaped identification if they were expressed at all, suggesting that these proteins were not needed for the growth of *T. acidophilum* either with or without oxygen. The corresponding mRNA expression levels of these genes were also at the lower detection limit. The protein database shows that these proteins are mostly hypothetical proteins. However, recent BLAST homology searches revealed that many of these proteins contain known domains characteristic of transposases, glycosyl transferases, ATP/GTP/DNA binding proteins, and ABC transporters. We hypothesize that these genes might be induced or expressed in response to specific compounds or under specific conditions.

Comparing the aerobic and anaerobic proteome of *T. acidophilum*, we found that 99% (1014) of the detected proteins were present for both culture conditions, and only 9 proteins were unique to aerobic and 2 to anaerobic cells. Whether these proteins are really unique to one

culture conditions or escaped identification in the other one due to low abundances is disputable. These 11 proteins were represented at low expression level with absolute abundances in the range of 10^1 to 10^4 arbitrary units (AU), and most of them are annotated as hypothetical proteins with unknown functions; however, recent BLAST searches showed that proteins exclusively detected in aerobically grown cells contain conserved domains characteristic for vitamin B_{12}-transporters, ATP-dependent RNA helicases, NAD(P)H:FMN oxidoreductases, radical SAM enzymes and thermopsins.

Reproducibility and accuracy of label-free quantitation data

The CV value, which expresses the standard deviation in percentage of the sample mean, was used to check the reproducibility of label-free quantitation data. Proteins with at least two intensity values in three experimental replicates were used for CV evaluations. Figure 4.3.2 displays the histogram of CVs of normalized intensities for proteins identified in cells grown under aerobic and anaerobic conditions, respectively. The majority of the data (CV<70%) fit a Gaussian distribution. The median and mean of the CV values were calculated to evaluate the precision of the experiments: the median is described as the number separating the higher half of a population from the lower half; the mean is the arithmetic average of a set of values. For our dataset, the median and mean CV values are 29.0% and 35.8% for aerobic and 35.1% and 41.9% for anaerobic samples, respectively, indicating a reasonable degree of precision of label-free quantitation experiments.

Logarithmic plots of protein ratio vs. intensity (Figure 4.3.3) of three aerobic and anaerobic replicates are distributed around ratio one, indicating good reproducibility of the measurements. The plot migrates closer to ratio one at high protein intensity range (>10^8), which indicates higher precision of quantitation for abundant proteins. This result shows good agreement with the label-free quantitative study of cell lines and tissues by Cox *et al.* [34], but due to the technical limitation, the accuracy of label-free quantitation is lower (by a factor of 4-5 on average) compared to SILAC-based quantitation [34]. The outlier spots with large experimental variances can be eliminated using a CV cutoff of 50% (Figure 4.3.3).

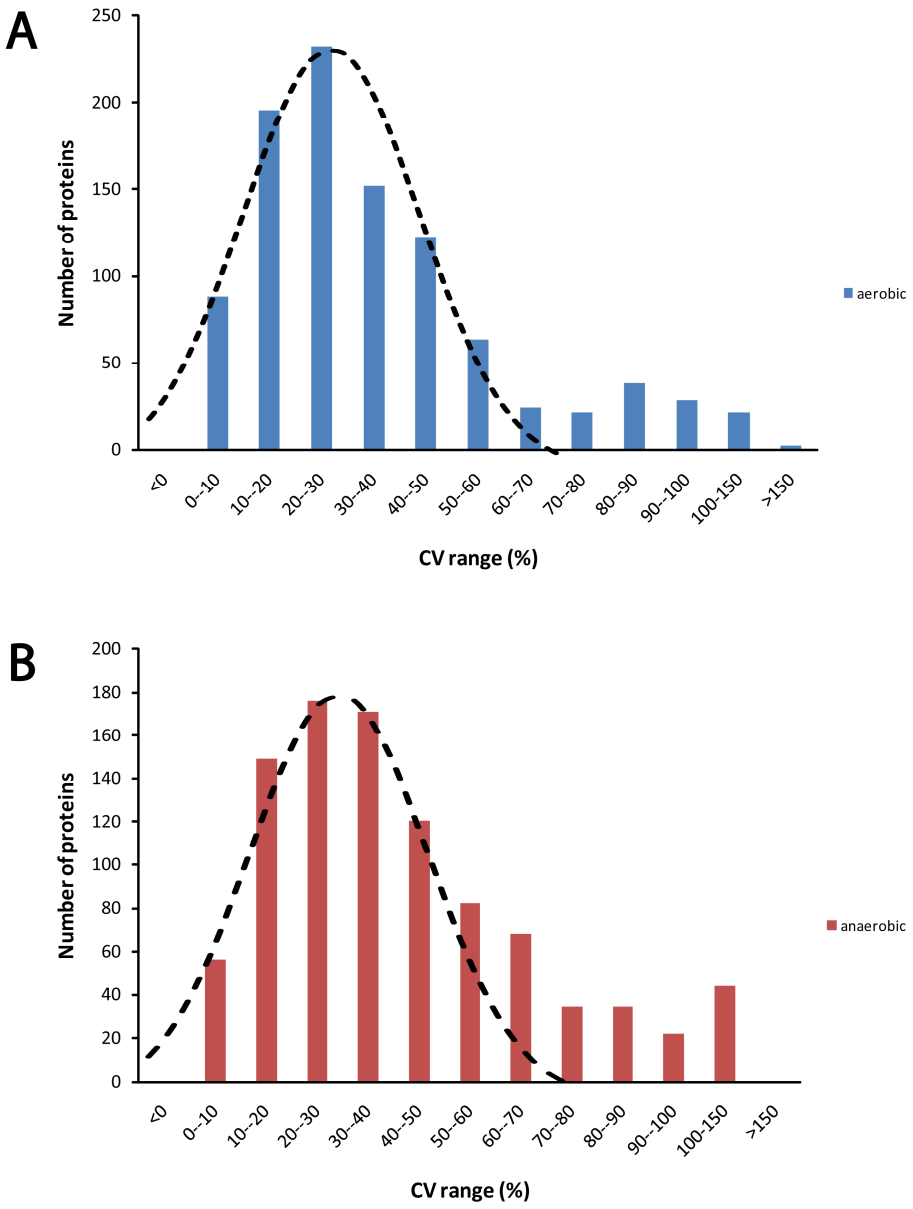


Figure 4.3.2: Histograms of CVs of normalized intensities for proteins identified in three aerobic (A) and three anaerobic (B) biological replicates, respectively. The majority of the data (CV<70%) fit a Gaussian distribution, as indicated by the black dotted line.

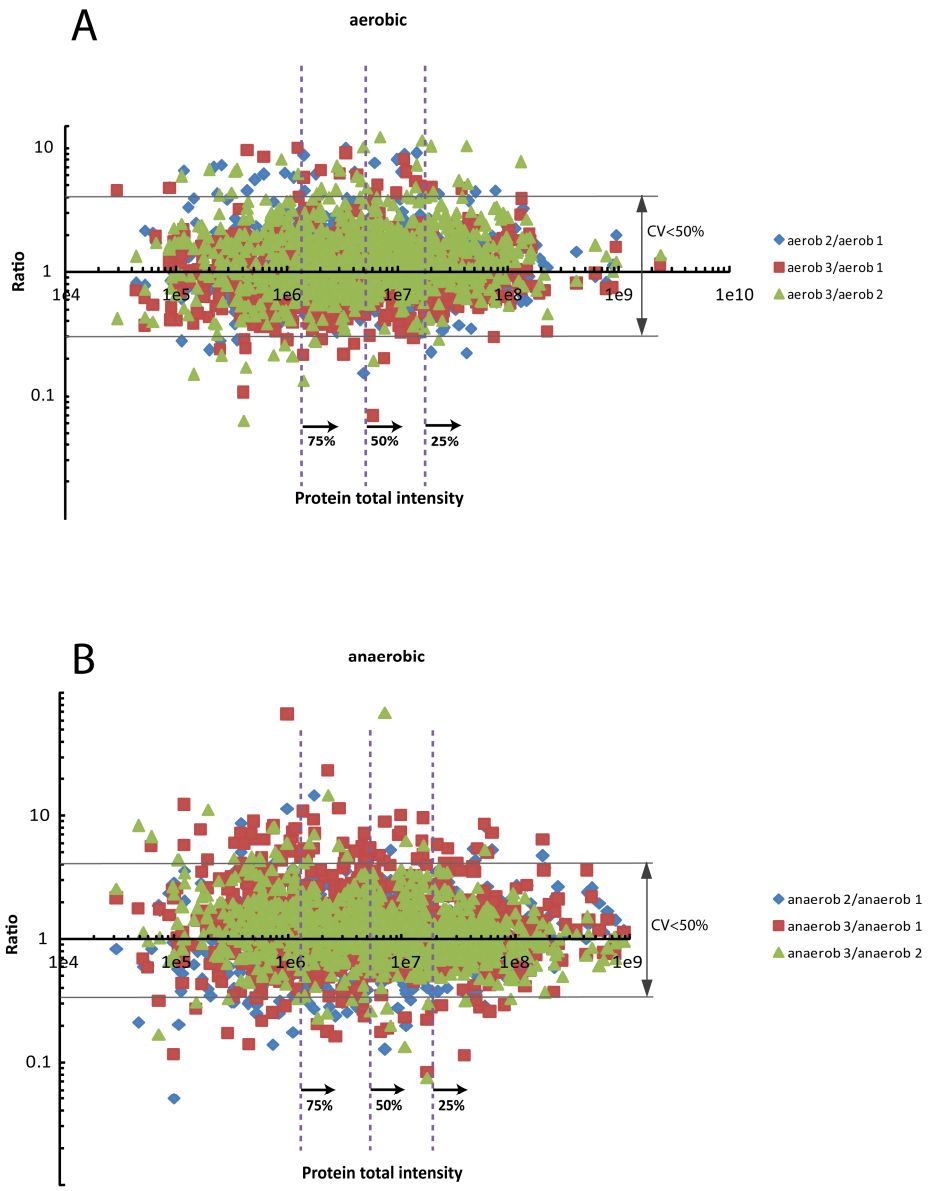


Figure 4.3.3: Logarithmic plots of protein ratios vs. intensities. The vertical lines indicate the quartiles of the intensity distribution. Blue, red and green spots represent protein ratios of the three aerobic (A) and the three anaerobic (B) replicates.

Quantitation of proteins expressed under the investigated conditions

For the quantitation data analysis, the CV and a two-sided Student's t-test were used to perform statistical evaluations, and a two-fold change in protein abundance was set to define biological regulation.

The CV represents the inter-experimental variations of experimental replicates. The histogram of the CV values of normalized intensities for identified proteins shows a Gaussian distribution (Figure 4.3.2), with median values of 29.0% (aerobic)/35.1% (anaerobic) and standard deviations of 25.0% (aerobic)/27.4% (anaerobic). To substantiate these results, the histogram of CVs of the sub-proteome of macromolecular complexes (Mw≥300kDa) was analysed. It shows a similar Gaussian distribution as the total proteome, with median values of 25.9% (aerobic)/35.4% (anaerobic) and standard deviations of 22.1% (aerobic)/27.2% (anaerobic). Thus, 50% indicates one standard deviation from the mean, which can be considered as a reasonable cutoff for experimental variation.

The Student's t-test was used to calculate the significances in expression level changes of the regulated proteins. It was performed with a p-value cutoff of 0.05 to give significant quantitation data with 95% confidence.

Thorough statistical analyses of the mass spectrometric data enabled us to establish a robust quantitation dataset consisting of 604 quantified proteins, of which 341 remained unchanged at criteria: $0.5<Ratio_{anae/ae}<2$ and CV≤50%, 263 cytosolic proteins were regulated of which 146 were down-regulated (criteria: $Ratio_{anae/ae}\leq 0.5$ and $p\leq 0.05$) and 117 were up-regulated under anaerobic conditions (criteria: $Ratio_{anae/ae}\geq 2$ and $p\leq 0.05$) (Table S2). The remaining identified proteins could not be quantified due to their low abundances and/or large experimental variances. Remarkably, 89 of the regulated proteins are hypothetical proteins.

Comparative protein abundance analysis based on the three proteomics experiments

To compare the effectiveness in protein identification rate of the three proteomics approaches described in this work, a protein abundance analysis was carried out. Protein abundance values from the GeLCMS experiment were used (Figure 4.3.4). Arbitrarily, we set the value≥5.5 as a cutoff for high abundant, 4.6-5.5 for intermediate, 3.7-4.6 for low abundant, and ≤3.7 for very low abundant proteins. With the 2DE-MALDI-TOF-MS and MSC-LC-MS/MS methods, we identified mostly intermediate and high abundant proteins, while with the highly sensitive GeLCMS method, the identification rate of low/very low abundant proteins was extended dramatically.

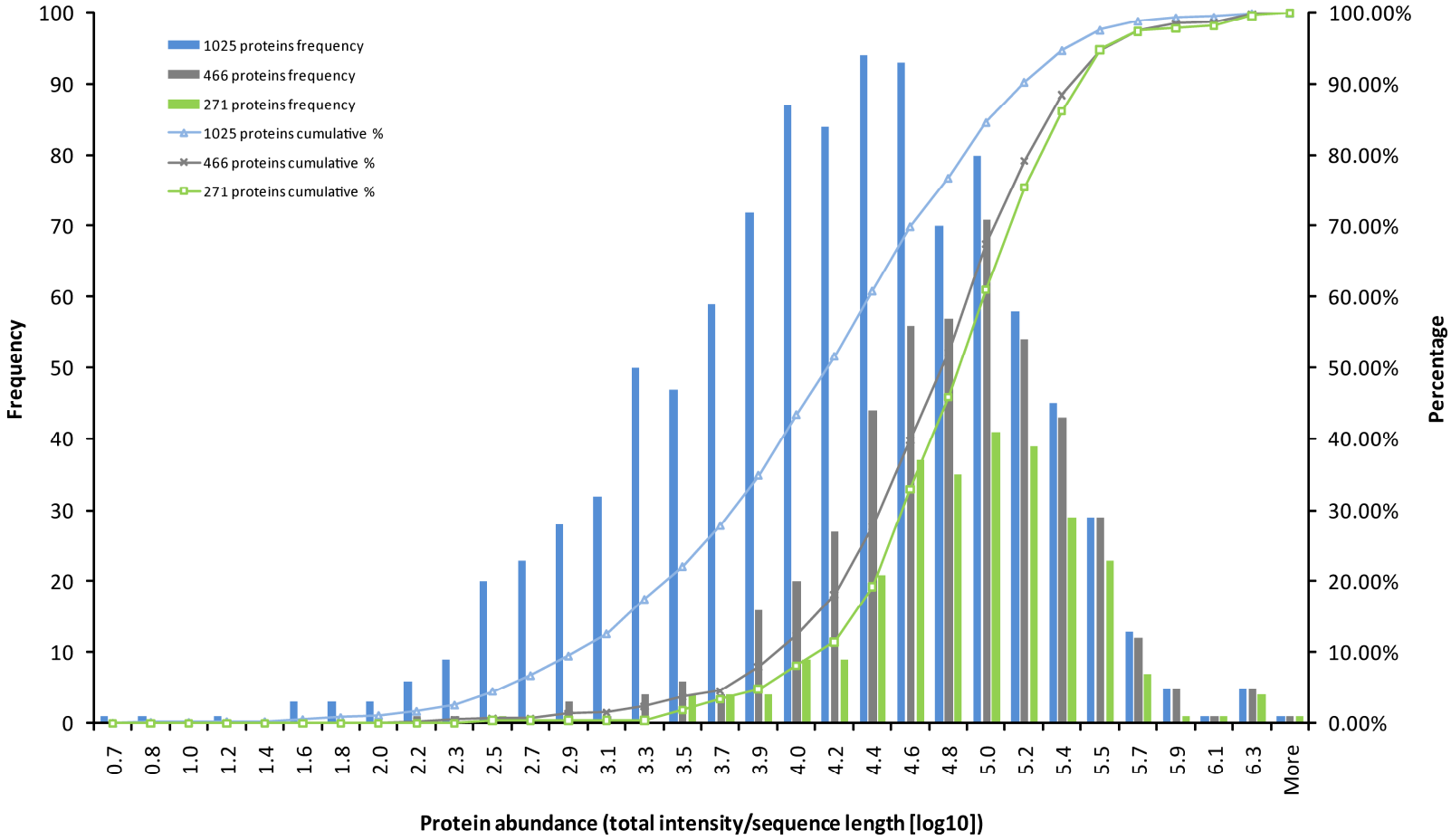
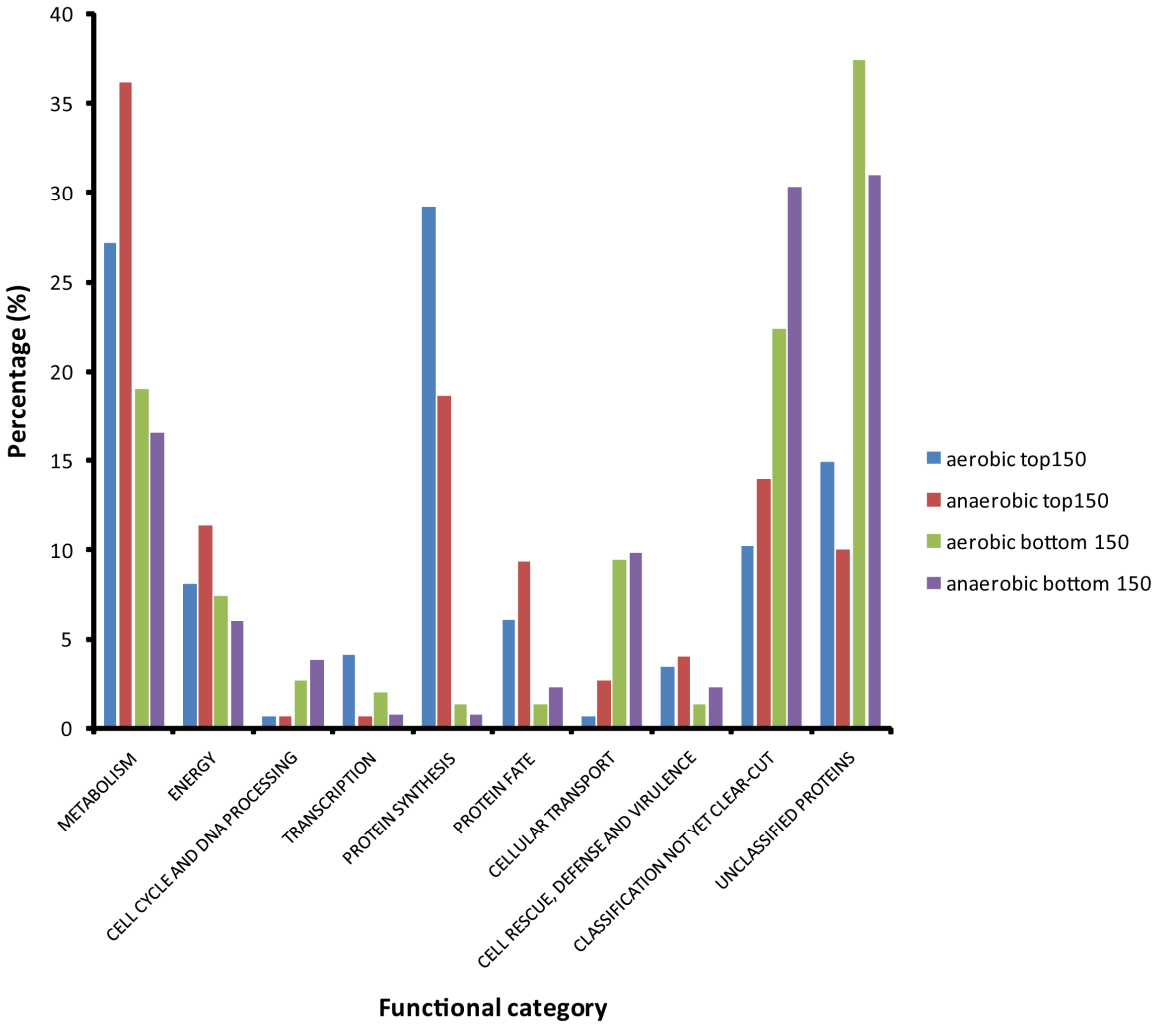


Figure 4.3.4: Distributions of protein abundances (log_{10}) under aerobic conditions. Blue bar and blue curve indicate frequency and cumulative percentage of the 1025 proteins identified with the GeLCMS method. Grey bar and grey curve show frequency and cumulative percentage of the 466 proteins identified with the MSC-LC-MS/MS method. Green bar and green curve represent frequency and cumulative percentage of the 271 proteins identified with the 2DE-MALDI-TOF-MS method.

Functional categories of the most and least abundant proteins

The absolute quantitation showed that the intensities of proteins identified varied from 10^0 to 10^6 AU, indicating a broad dynamic range of protein abundances. Interestingly, a large proportion (>60%) of the most (top 150) and least (bottom 150) abundant proteins were common between aerobic and anaerobic conditions. The functional category annotations of the 150 most and least abundant proteins were performed based on the MIPS database. In general, the abundant proteins were predominantly involved in metabolism (aerobic 27%, anaerobic 36%), protein synthesis (aerobic 29%, anaerobic 19%), energy (aerobic 8%, anaerobic 11%), and protein fate (aerobic 6%, anaerobic 9%), whereas in the low abundance group, significantly fewer proteins were associated with the aforementioned processes. Instead, the majority of the low abundant proteins (~60%) fell into unknown functional categories (Figure 4.3.5). However, we found that 25 (for aerobic)/27 (for anaerobic) hypothetical proteins were among the abundant ones, which suggests that they are important for the cell and that their functions should be studied.



Functional category	Aerobic top 150 (%)	Anaerobic top 150 (%)	Aerobic bottom 150 (%)	Anaerobic bottom 150 (%)
METABOLISM	27.2	36.2	19.0	16.6
ENERGY	8.2	11.4	7.5	6.1
CELL CYCLE AND DNA PROCESSING	0.7	0.7	2.7	3.8
TRANSCRIPTION	4.1	0.7	2.0	0.8
PROTEIN SYNTHESIS	29.2	18.7	1.4	0.8
PROTEIN FATE	6.1	9.4	1.4	2.3
CELLULAR TRANSPORT	0.7	2.7	9.5	9.8
CELL RESCUE, DEFENSE AND VIRULENCE	3.4	4.0	1.4	2.3
CLASSIFICATION NOT YET CLEAR-CUT	10.2	14.0	22.4	30.3
UNCLASSIFIED PROTEINS	14.9	10.0	37.40	31.0

Figure 4.3.5: The distribution of the 150 most and least abundant proteins of aerobic and anaerobic conditions in functional categories. The annotation is based on the MIPS database.

4.3.1.2 Evaluation of transcriptomics data

As a complement to the proteomics studies, a cDNA microarray-assisted transcriptomics analysis was performed to provide a genome-wide portrait of the transcriptome, thus giving insights into differential expression of genes at genome level, including both cytosolic and membrane protein-encoding genes.

Reproducibility of transcriptomics data

The hierarchical clustering of the microarray data representing the expression level of 1482 genes is displayed in Figure 4.3.6. The high similarities of the heat map views of the 3 aerobic and the 3 anaerobic measurements, respectively, indicate excellent reproducibility of the microarray experiment. The histograms of the data show that, under aerobic conditions, most of the genes have intermediate or high transcription levels, while under anaerobic conditions more genes are present with low transcription level.

Identification and quantitation results of transcriptomics analysis

The transcriptomics analysis revealed that the expression levels of all 1482 genes were determined with high reliability (mean, median and standard deviation of CVs from three biological replications being 6.7%, 5.7% and 4.7%, respectively). Similar to the proteomics dataset, a two-fold cutoff was set as an indicator of significant biological regulation. The Student's t-test p-value of 0.05 was set as a cutoff for significant quantitations. The results revealed that 445 genes exhibited more than 2-fold regulation, 205 of which are annotated as hypothetical proteins. Remarkably, 145 out of the 445 regulated genes encode for membrane proteins, covering 43% of the membrane genome (335 ORFs). One hundred and thirty of these membrane protein-encoding genes were down-regulated under anaerobic and 15 under aerobic conditions, while the distribution of the remaining 300 regulated mRNAs that are translated to cytosolic proteins were 90 up-regulated and 210 down-regulated under anaerobic conditions (Table S3).

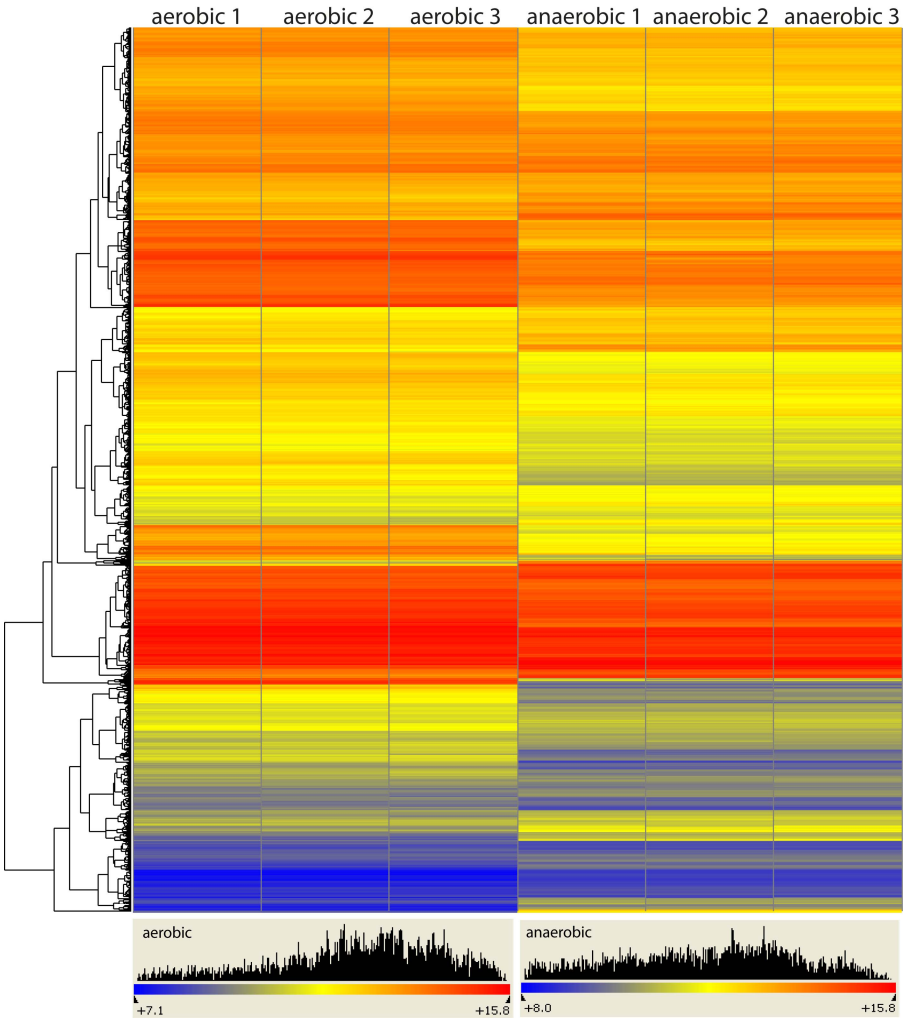


Figure 4.3.6: Hierarchical clustering of microarray data. The transcription level of 1482 genes from three aerobic and three anaerobic replicates are colour-coded with a heat map, with blue representing low, yellow intermediate and red high transcription levels, as indicated in the histogram inset. Each row in the grid corresponds to one gene, and each column represents a single experiment. The gene expression profiles are clustered using the hierarchical method, which groups data points by agglomerating them one-by-one into ever-growing groups. This grouping is done by first finding the shortest link among all of the data points, and then combining those two points into a group. The algorithm then finds the next shortest distance, including the distance from other points to this new group, and groups even further. Once all of the data points have been grouped, the resulting clusters are displayed in the heat map view with the associated experiment tree and gene tree on the horizontal and vertical axes, respectively. By default, the Euclidean distance metric and the centroid linkage method are used in hierarchical clustering (ArrayStar v2.0, DNASTAR, Inc.) [139].

4.3.1.3 Correlation of transcriptomics and proteomics datasets

A quantitative comparison between protein and mRNA expression ratios was performed. Figure 4.3.7.A illustrates a scatter plot comparing mRNA and protein expression ratios for the 604 common proteins/mRNAs, which can be assigned to both proteomics and transcriptomics quantitation data. The Pearson correlation coefficient of these two datasets is 0.57, indicating a weak positive correlation between mRNA and protein expression ratios. The slope of the regression curve for transcript ratios versus protein ratios is 0.32, showing that the general changes at transcript level are suppressed 3-fold compared to the changes at the proteome level. For example, Ta1475 is one of the most down-regulated proteins under anaerobic conditions (more than 10-fold change, see Figure 4.3.8); however, the mRNA level change of this protein is only 3-fold. More examples of these findings are Ta0068, Ta0329, Ta0635, Ta0773, Ta0897, Ta0952 and Ta1153. They are 7- to 50-fold up-regulated at protein level under anaerobic conditions, whereas they show only a 2- to 5-fold up-regulation at transcript level.

We next evaluated only those genes that showed significant regulation at mRNA level or at protein level (Figure 4.3.7.B). After filtering out the none-regulated signals, the overall correlation between transcript and protein increased to 0.7. A Venn diagram of regulated cytosolic mRNA versus regulated proteins (Figure 4.3.7.C) shows an overlap of ~30% (97 genes). It indicates that a large number of genes are solely regulated at either mRNA or protein level. The ratio of down and up-regulated proteins is close to one (146:117), while the ratio between down and up-regulated mRNAs is around two (210:90). One explanation for this difference could be that due to the experimental variation, large numbers of proteins were not quantified, whereas all mRNAs (604 vs. 1482) were quantified.

Notably, there were 6 genes showing opposite regulation tendencies, including Ta0194m (predicted hydrolase), Ta0298 (alpha-glucosidase related protein) and Ta1129 (sulfide-quinone reductase related protein) which were 5- to 10-fold up-regulated at protein level, while their mRNA level showed 2- to 7-fold down-regulation; Ta0684 (deoxyribose-phosphate aldolase), Ta0842 (hypothetical protein) and Ta1121 (cob(I)alamin adenosyltransferase related protein) which were 2- to -5 fold down-regulated at protein level but 2- to 3-fold up-regulated at mRNA level. No specific biological relationships could be

figured out for these proteins. Thus, further investigations of their functions and posttranscriptional regulation processes are needed.

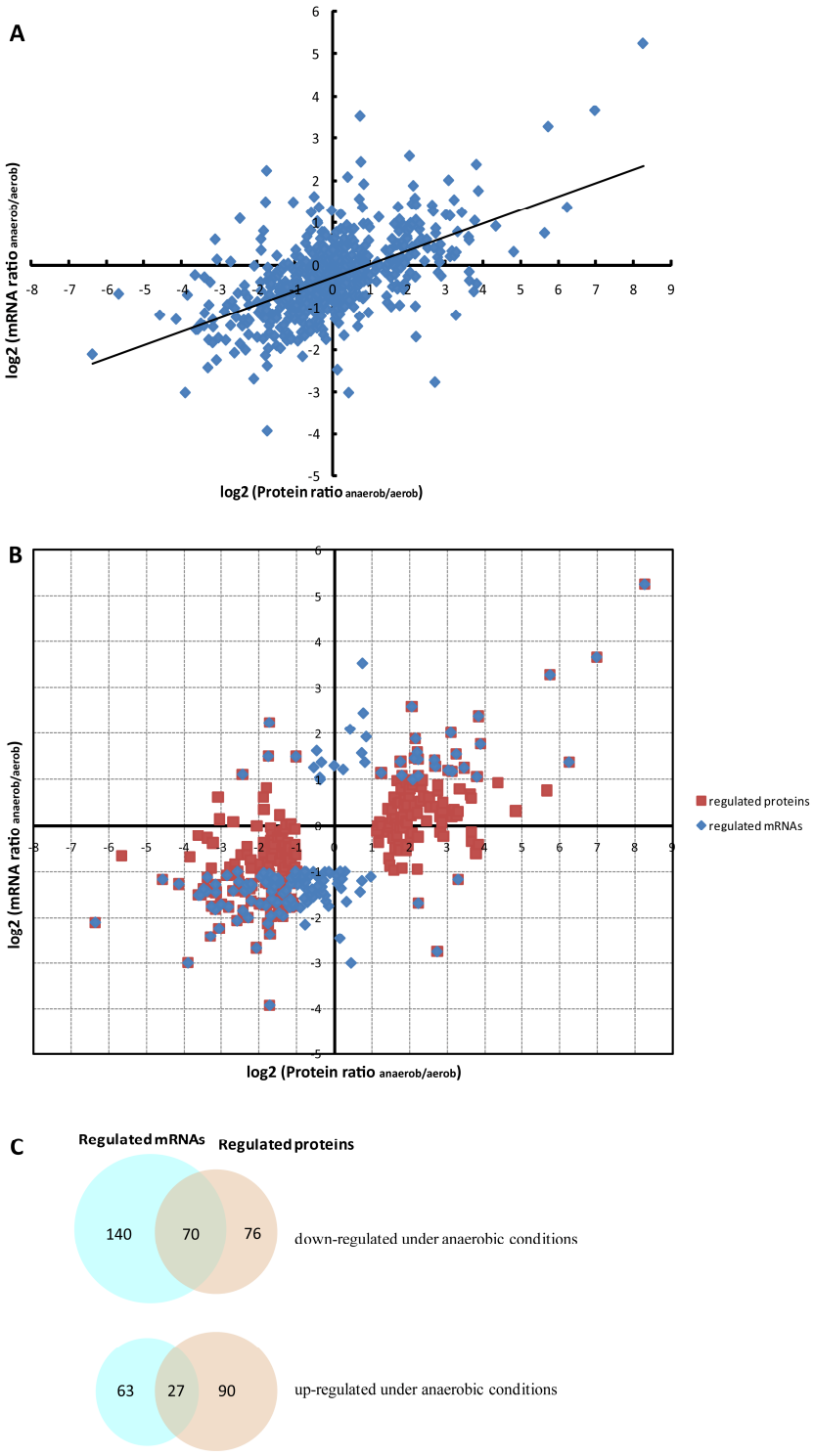


Figure 4.3.7: Scatter plot representation of the correlation between mRNA and protein expression ratios. The log_2 values of mRNA and protein expression ratios were calculated and plotted. The linear

regression curve is shown as a solid line. A) Correlation analysis of the 604 common genes for which both protein ratios and mRNA ratios were determined. The plot indicates low correlation with a Pearson correlation coefficient of 0.57. B) Correlation plot of regulated genes at transcript level and protein level. After filtering out the none-regulated signals, the Pearson correlation between transcript and protein increased to 0.7 (blue spot: only regulated at mRNA level; red spot: only regulated at protein level; blue spots with red border: regulated genes both at mRNA level and protein level). C) Regulated cytosolic mRNAs vs. proteins. Overall analysis of regulated cytosolic genes at protein and mRNA level showed an overlap of ~30%.

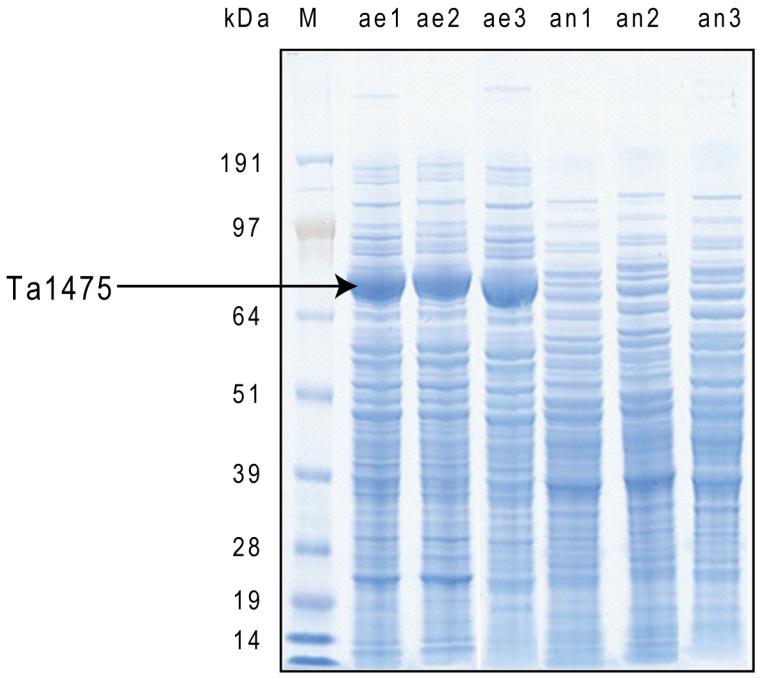


Figure 4.3.8: 1D-SDS-PAGE of the protein samples of three aerobic and three anaerobic replicates. Protein samples (20 µg of total protein) of each aerobic and anaerobic extract were separated on a 4-12% NuPage Novex Bis-Tris gel (Invitrogen) and stained with Colloidal Blue Staining Kit (Invitrogen). Mw marker is shown on the left hand side. The arrow indicates the band of Ta1475, which is one of the most down-regulated proteins under anaerobic conditions; however, its change at mRNA level is damped more than 3-fold compared to the change at protein level.

4.3.1.4 Analysis of the expression level of macromolecular complexes with Mw over 300 kDa

In Project 2, 35 macromolecular complexes with a Mw over 300 kDa were identified in cytosolic extracts of *T. acidophilum* by using the MSC-LC-MS/MS approach [140]. Here, bioinformatic analysis (BLAST and PDB database searches) of the proteins, which were more than two-fold regulated in the absence of oxygen, identified 4 additional putative macromolecular complexes, i.e. alpha-glucosidase (Ta0298, 480 kDa [141]), galactonate

dehydratase (Ta0085m, 350 kDa [142]), a homologue of glutamate synthase beta subunit (Ta0414, 800 kDa [143, 144]) and a homologue of protoporphyrin IX magnesium chelatase subunits D and I (Ta0576, 300 kDa, [145, 146]).

Alpha-glucosidase (Ta0298) showed 14-fold up-regulation under anaerobic conditions. The crystal structure of alpha-glucosidase from the hyperthermophilic archaeon *Sulfolobus solfataricus* has been determined as a hexamer of 480 kDa. Alpha-glucosidases contributing to starch and glycogen metabolisms and galactose metabolism are a group of exo-acting enzymes that play essential roles in carbohydrate metabolism and in glycoprotein processing and quality control [141]. The 14-fold up-regulation demonstrates its importance under anaerobic conditions. Interestingly, in contrast with the high up-regulation of Ta0298 at the protein level, its mRNA level was 2-fold down-regulated.

Ta0085m was 3-fold up-regulated at protein level in anaerobic conditions. A BLAST search indicated that Ta0085m has a high homology to gluconate dehydratase in *Sulfolobus solfataricus*, which exhibited a size of 350 kDa in a gel filtration experiment [142].

Glutamate synthases are NADPH-dependent enzymes, and they are complex iron-sulfur flavoproteins essential in ammonia assimilation processes to form various amino acids. They catalyze the NADPH-dependent conversion of α-ketoglutarate and L-glutamine to glutamate using either reduced pyridine nucleotides (NADH or NADPH) or reduced ferredoxin as the physiological electron donor [147]. Ta0414 shows homology to the glutamate synthase beta subunit. It was more than 8-fold up-regulated in anaerobic condition at protein level but unchanged at mRNA level. Glutamate synthases of *E. coli* and *Azospirillum brasilense* were reported to contain α and β subunits forming a heterotetramer $(\alpha\ \beta)_4$ with a Mw ~800 kDa [143, 144]. In *T. acidophilum* no homologue of the alpha subunit could be found.

We found that Ta0576, which was 2-fold down-regulated under anaerobic conditions, shows high homology to protoporphyrin IX magnesium chelatase (Mg chelatase) subunits D and I of *Synechocystis*, which can aggregate in an ATP-dependent manner, changing from a dimeric to an octameric structure with a Mw around 300 kDa [145, 146].

Quantitation data of 39 macromolecular complexes with Mw over 300kDa

The calculation of the absolute abundance of a macromolecular complex was done by summing up the average abundances of the subunits identified coming from three independent measurements and divided by the number of the subunits included. Using the same criteria for protein abundance as described above (≥5.5 as a cutoff for high abundant, 4.6-5.5 for intermediate, 3.7-4.6 for low abundant, and ≤3.7 for very low abundant proteins), we concluded that 49% of the macromolecular complexes had intermediate expression levels, while 28% were low/very low abundant, and 23% were high abundant in the cell (Figure 4.3.9). The transcriptomics and proteomics quantitation data of these 39 macromolecular complexes are represented in Table 4.3.1. The proteomics quantitation data showed that 12 complexes were up-regulated, 3 were down-regulated, 11 showed no significant changes under anaerobic conditions and 11 complexes were not quantified due to large experimental variances. For unchanged and down-regulated complexes, the proteomics data are in good correlation with tanscriptomics data, which means that consistent regulation tendencies were found at both protein and transcript level. In contrast, 10 out of the 12 up-regulated complexes remained unchanged or they were even slightly down-regulated at mRNA level, and only Fix ABCX and glutamate dehydrogenase showed the same up-regulation tendencies at both protein and mRNA level.

Generally, quantified subunits belonging to the same complex showed consistent up/down regulations, except for the chromosome segregation protein complex (Ta0157, Ta0158) and several ribosomal proteins.

Ta0157 and Ta0158 are predicted to form a chromosome segregation protein complex by the molecular sieve chromatography study [140]; however, their proteomics quantitation data showed different regulation tendencies in the absence of oxygen. Ta0157 was unchanged whereas Ta0158 was 5-fold down-regulated. Their mRNA levels were both 2-fold down-regulated. The significantly different expression levels of Ta0157 and Ta0158 in the absence of oxygen might go together with structural variances of the complex, which needs to be proven by further experiments.

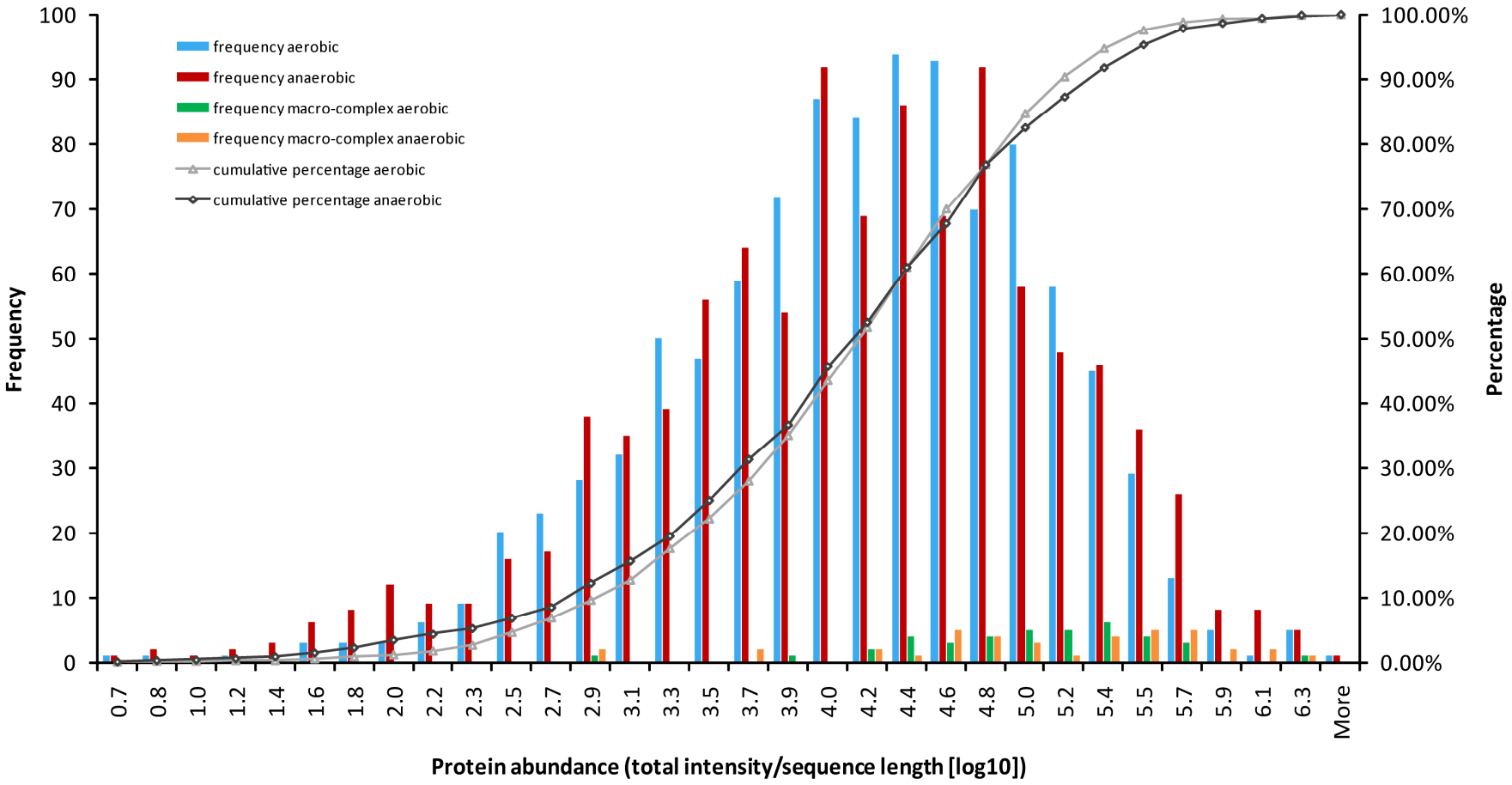


Figure 4.3.9: Histogram of log_{10} protein abundances based on absolute quantitation. Blue bar and grey curve indicate frequency and cumulative percentage of aerobic samples. Red bar and black curve show frequency and cumulative percentage of anaerobic samples. Green and yellow bars represent the frequencies of 39 macromolecular complexes under aerobic and anaerobic conditions, respectively.

According to the genome annotation, there are 57 ribosomal proteins in *T. acidophilum*. The majority of them were identified in our proteomics study, except for Ta1257, Ta1241 and Ta1317a, probably due to their low Mw (~6 kDa). Most of the ribosomal proteins remained unregulated under anaerobic conditions. Interestingly, the 50S ribosomal proteins L10 (Ta1057m) and L31 (Ta0054) showed 10-fold down-regulation, in contrast, L11 (Ta0361, 2-fold up) was 2-fold up-regulated under anaerobic conditions. In contrast to the proteomics data, the mRNA levels of all ribosomal subunits were very consistent, showing no significant changes in the absence of oxygen.

Table 4.3.1: Quantitative proteomics and transcriptomics data of macromolecular complexes. Protein annotation, Mw of the complexes, changes in protein expression level (protein Ratio$_{anae/ae}$), changes in mRNA expression level (mRNA Ratio$_{anae/ae}$) and absolute intensities in protein expression level are listed.

Protein complexes	***Mw***	***Protein Ratio$_{anae/ae}$[1]***	***mRNA Ratio$_{anae/ae}$[1]***	***Absolute intensity (log_{10})***	
				aerobic[1]	**anaerobic[1]**
Down-regulated under anaerobic conditions					
formate dehydrogenase (Ta0424, 0425)	440 kDa	0.09	0.36	5.26	4.66
heat shock protein DnaJ (Ta1088)	>1 MDa	0.22	0.52	3.79	2.87
protoporphyrin IX magnesium chelatase (Ta0576)	300 kDa	0.49	0.47	4.05	3.53
Up-regulated under anaerobic conditions					
alpha-glucosidase (Ta0298)	480 kDa	9.82	0.44	2.75	2.75
enolase (Ta0882)	300 kDa	3.62	1.36	4.60	4.86
ferredoxin oxidoreductase (Ta0259m, 0260)	300 kDa	5.20	0.81	4.93	5.68
fixABCX (Ta0326-0329)	550 kDa	10.43	2.28	4.48	5.51
galactonate dehydratase (Ta0085m)	350 kDa	2.95	0.82	5.12	5.51
glutamate dehydrogenase (Ta0635)	~300 kDa	14.85	3.41	5.02	5.98
glutamate synthase [NADPH] small chain (Ta0414)	800 kDa	9.54	1.17	4.40	4.97
succinyl-CoA synthetase (Ta1331-1332m)	500 kDa	3.79	1.14	5.21	5.77
sulfide dehydrogenase (Ta1412-1413)	1 MDa	13.15	1.62	5.17	6.05
thermosome (Ta0980, Ta1276)	900 kDa	2.20	0.94	6.23	6.14
tricorn protease (Ta1490)	~720 kDa	2.78	0.54	4.66	4.57
VAT ATPase (Ta0840)	~500 kDa	2.72	0.97	5.23	5.30
Unchanged expression level					
alpha, alpha-trehalose-phosphate synthase (Ta1210, Ta1209)	>1 MDa	1.15	0.75	4.19	4.08
exosome (Ta0929, Ta1292-1294m)	500 kDa	0.94	0.77	5.06	4.92
glucose kinase (Ta0825)	550 kDa	1.18	1.21	5.00	5.20
glutamine synthetase (Ta1498)	~600 kDa	0.57	0.69	5.54	5.48
ornithine carbamoyltransferase (Ta1330)	~420 kDa	1.18	1.06	5.56	5.58
orotate phosphoribosyltransferase (Ta1164)	>1 MDa	1.82	1.07	5.50	5.38
peroxiredoxin (Ta0954m)	300 kDa	1.90	0.85	5.51	5.60
phosphopantetheine adenylyltransferase (Ta1194)	>1 MDa	0.51	0.70	4.80	4.52
pyruvate dehydrogenase (Ta1435-1438)	>1 MDa	1.01	0.63	4.60	4.50
ribulose-phosphate 3-epimerase (Ta1315)	>1 MDa	1.10	0.59	4.37	4.35
sulfolipid biosynthesis (Ta1073)	>1 MDa	0.75	1.11	4.92	4.47
Subunits with different ratios					
chromosome segregation protein (Ta0157, 0158)	>1 MDa	1.22 (Ta0157), 0.18 (Ta0158)	0.57	4.27	3.63
ribosome*	2.4 MDa	*	0.89	5.18	5.17

Not quantified complexes (or not every of the subunits were quantified)					
glutamate formiminotransferase (Ta1476, 1478)	600 kDa	-	1.59	5.00	5.29
aspartate carbamoyltransferase (Ta0574, 0575)	>1 MDa	-	0.69	5.00	4.75
DNA-directed RNA polymerase	500 kDa	-	0.76	4.88	4.72
phosphoribosylformylglycinamidine synthase (Ta1066, 1318m, 1044m)	1 MDa	-	0.93/0.64/3.09	4.63	4.52
proteasome (Ta0612, Ta1288)	700 kDa	-	0.79	5.62	5.75
DnaK (Ta1087)	600 kDa	-	0.82	4.71	5.19
muconate cycloisomerase (Ta0756)	>1 MDa	-	3.92	4.29	4.70
peroxiredoxin (Ta0152)	~600 kDa	-	0.08	5.68	5.49
pyridoxal 5'-phosphate synthase (Ta0522, Ta0009)	>1 MDa	-	1.61	5.23	5.60
small heat shock protein (Ta0471)	600 kDa	-	0.86	4.88	5.36
transaldolase (Ta0616)	300 kDa	-	0.87	5.41	5.59

[1]: for complexes, which consist of more than one subunit, average values of the subunits are shown in the table.
*: ribosomal proteins show different ratios, for details see Table S2.
-: not quantified due to large experimental variances.

4.3.1.5 Categorization of the regulated genes on the basis of their biochemical functions

Bioinformatic analyses were performed using the Gene Ontology, KEGG pathway and MIPS online servers for biochemical functional categorization of the up-/down-regulated genes.

Regulated genes encoding for membrane proteins

Among the 445 regulated transcripts, 145 encodes for membrane proteins, covering 43% of the membrane genome (335 ORFs). The majority of these genes (90%, 130) were more than 2-fold down-regulated under anaerobic conditions, and functionally they are associated with ion transport, amino acid transport, ATPase and endopeptidase activities.

Regulated genes encoding for cytosolic proteins

1) GO analysis regarding general molecular functions and biological processes

Despite the fact that only 97 genes out of the regulated 300 mRNAs and 254 proteins are overlapping, the GO analysis of the regulated cytosolic mRNAs and the regulated proteins shows a similar overrepresentation of the biological functions such as carbohydrate metabolism, TCA intermediate metabolism, generation of precursor metabolites and energy under anaerobic, and cobalamin (vitamin B_{12}) biosynthesis under aerobic conditions, respectively.

2) Analysis of selected biochemical pathways

2.1) Schematic representation of the main proteolytic pathways, proton transport pathways and glucose metabolism

Three central biochemical pathways, the main proteolytic pathways, proton transport pathways, and the glucose metabolism including the non-phosphorylated ED pathway, glycolysis/gluconeogenesis pathway (EMP) and TCA cycle, were analysed in more detail (Figure 4.3.10), as in total 14 macromolecular complexes are involved in these processes, therefore, they can serve as good templates for visual proteomics studies with respect to complete pathways.

The main proteolytic pathways

The main proteolytic pathways in *T. acidophilum* are the uptake and degradation of exogenous polypeptides, the quality control of damaged proteins and the degradation of regulatory proteins [5].

a) The uptake and degradation of exogenous polypeptides

T. acidophilum is a thermoacidophilic scavenger, which can not grow without exogenous oligopeptides and proteins provided by yeast extract under laboratory conditions [148]. The large number of genes (291) encoding for proteins that take part in the protein metabolism pathways reflects this life style, which is more pronounced by the up-regulation (2- to 10-fold) of a set of extracellular polypeptide uptake-transport systems (12 proteins) and 5 membrane anchored extracellular proteases at transcript level under aerobic conditions (where *T. acidophilum* grows faster). Additionally, the BLAST sequence analysis of 72 hypothetical membrane protein encoding genes, which were significantly up-regulated (up to 100-fold) under aerobic conditions at mRNA level, showed that a large number (20) of them have predicted functions related to protein, oligopeptide and amino acid transporter and protease activities. Hence, 15 hypothetical membrane transporters and 3 membrane proteases, which were newly identified by the BLAST analysis, are also shown in Figure 4.3.10.A. The proteomics data of these membrane proteins are missing because only cytosolic proteins were analysed in our project.

b) The quality control of damaged proteins and degradation of regulatory proteins

Chaperones and proteases are important enzymes for the quality control and degradation of damaged proteins in the cell. A large portion of chaperones and proteases were 2- to 4-fold up-regulated under anaerobiosis at protein level, including tricorn, tricorn cofactors, VAT, thermosome and proteasome (only one subunit was quantified), whereas their transcript expression levels remained unchanged.

Proton transport pathways

Figure 4.3.10.B displays the hypothetical proton transport pathways in *T. acidophilum* adapted from the proteomics study of *Thermoplasma volcanium* [149].

a) Aerobic proton transport

Most of the proteins involved in aerobic proton transport [149] failed to be quantified being membrane-associated proteins. The mRNA levels of the majority of these genes were significantly increased under aerobic conditions, except for NADH dehydrogenase. NADH dehydrogenase is predicted to be part of the aerobic respiratory chain, which transfers electrons from NADH to oxygen [150]. Unexpectedly, instead of an increase in aerobic conditions, most subunits of NADH dehydrogenase remained unchanged and the α and β subunits (Ta0969m, Ta0970m) were even 2- to 3-fold up-regulated under anaerobic conditions. NADH dehydrogenase is a large complex with Mw 880 kDa, consisting of 12 subunits encoded by one operon [150]. The transcription of the NADH dehydrogenase genes starts with Ta0970m and ends with Ta0959. The position of the 12 genes on the chromosome of *T. acidophilum* shows that the C-terminus of Ta0970m and the N-terminus of Ta0969m overlap by 12 bases, and they are most probably transcriptionally coupled. However, Ta0969m is not overlapping with Ta0968 which might explain the different transcription level. The significant up-regulation of the NADH dehydrogenase α and β subunits is due to unknown regulatory processes and might result in structural changes of NADH dehydrogenase complex under anaerobic conditions.

The other aerobic electron transport chain members (membrane bound proteins) like succinate dehydrogenase (Ta1001-Ta1004) and the putative cytochrome C oxidase homologue

(Ta0435) were not changed, but the cytochrome bd complex encoding genes (Ta0992, Ta0993) were highly up-regulated under aerobic conditions, classifying them into the group of genes that changed the most at mRNA level (up to 20-fold). The genes of the ATPase complex (Ta0001-Ta0008, Ta0001z) were just slightly up-regulated under aerobic conditions. This regulation pattern might indicate that part of the aerobic electron (proton) transport chain (NADH dehydrogenase and succinate dehydrogenase) remain active during anaerobic respiration (together with the ATPase), while the final step of O_2 reduction is inactivated.

b) Anaerobic proton transport

Most of the enzymes participating in the anaerobic proton transport are cytosolic, and their expression levels were dramatically up-regulated (up to 300-fold) under anaerobic conditions. However, we did not find any highly up-regulated candidate that could catalyze S^0 reduction under anaerobic conditions: Ta0046 and Ta0047 are homologues of the sulfhydrogenase subunits of *T. volcanium* [149] which reduce elemental sulphur to H_2S, but their expression level was unchanged both at the protein and mRNA levels. Instead, we found that Ta0837, a homologue of the recently characterized NAD(P)H sulfure oxidoreductase (NSR, PF1186) which catalyzes the S^0 reduction in *Pyrococcus furiosus* [151], was ~4 times up-regulated under anaerobiosis. Whether this enzyme is responsible for S^0 reduction in *T. acidophilum* needs to be proven. It might be that the enzyme activity of Ta0837 is much higher under anaerobic conditions and it can fulfill its function to reduce S^0 to H_2S by ~4 times up-regulation.

The fixABCX complex might be exceptionally important in this anaerobic proton transport as it encodes for the key ferredoxin. Interestingly, its close homologue from *Rhodospirillum rubrum* is membrane bound and shuttles protons to the anaerobically active and also membrane bound nitrogenase complex that converts N_2 to NH_3 [134]. In *T. acidophilum*, both the fixABCX and the putative S^0 reductase (Ta0837) are cytosolic enzymes.

The up-regulation and probable interaction of five macromolecular complexes (Ta0259-260, Ta0326-329, Ta0414, Ta0635 and Ta1435-1438) involved in the putative anaerobic proton transport provide good opportunities to investigate their interactions by cryo-EM.

Glucose metabolism

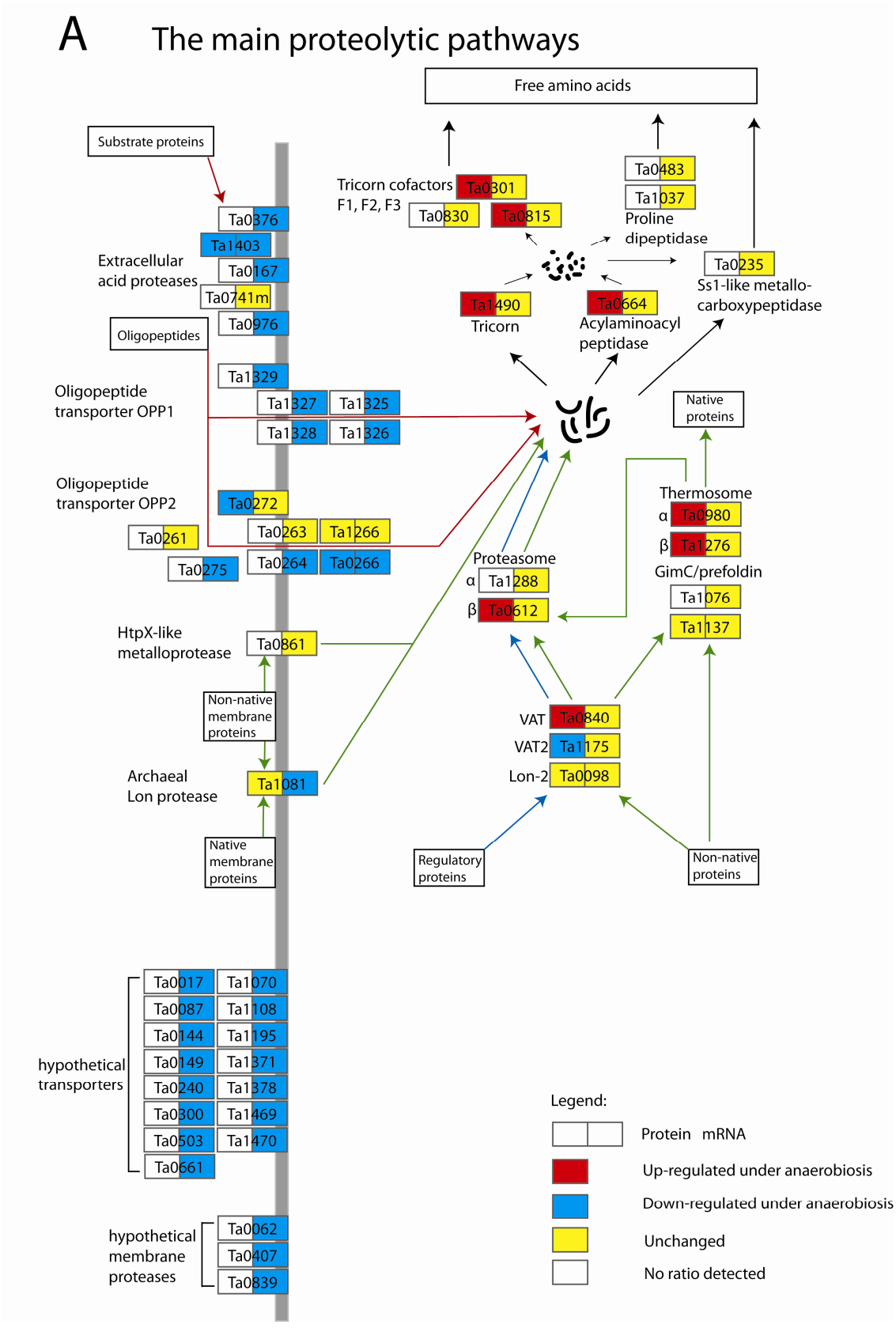
The non-phosphorylated ED pathway, the EMP pathway, and the TCA cycle are members of the central metabolic pathways of *T. acidophilum* [5, 152].

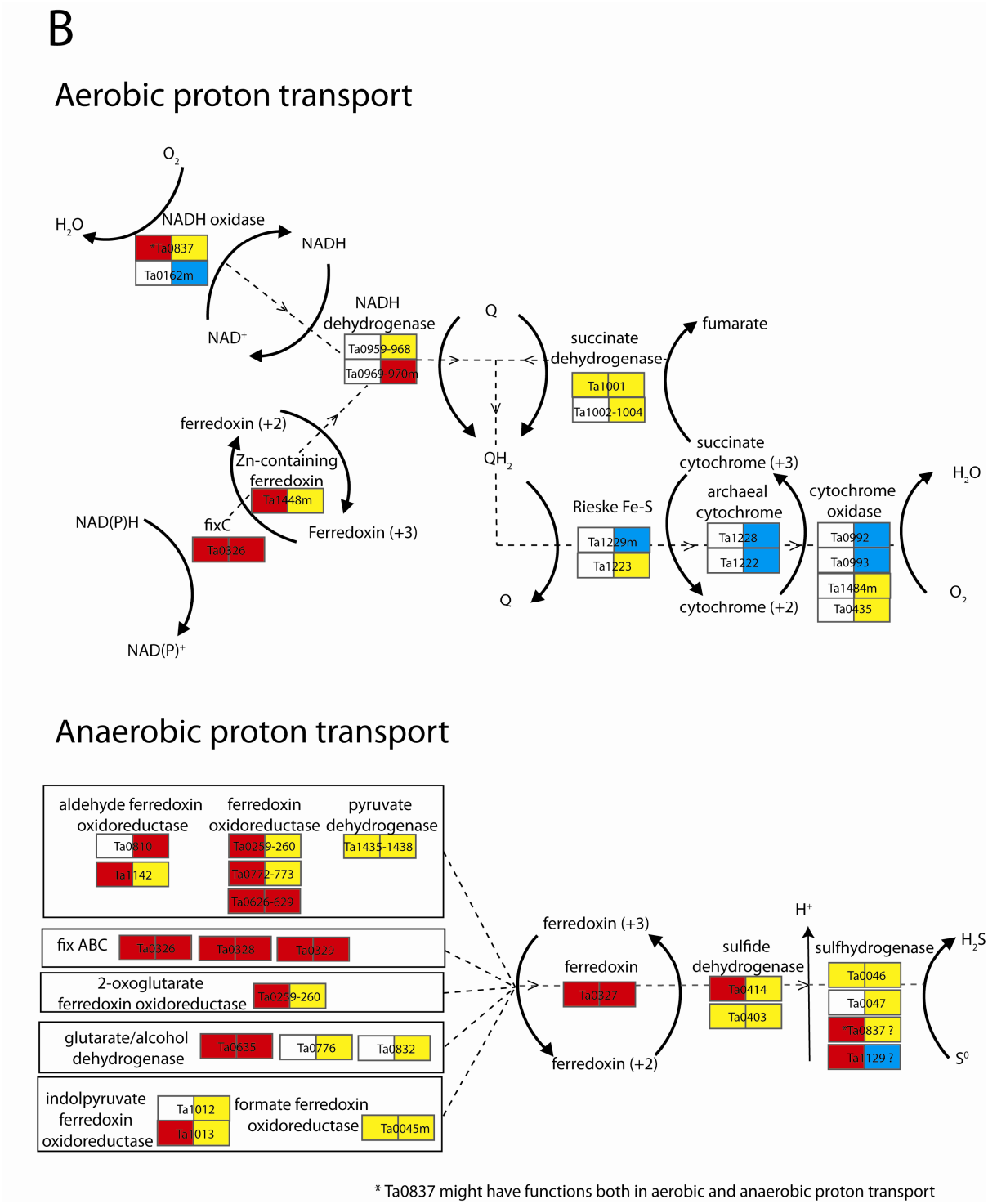
Generally, the non-phosphorylated ED pathway, which carries out glucose degradation in *T. acidophilum*, showed more than 3-fold up-regulation under anaerobic conditions based on the proteomics data, except for Ta0619 and Ta0453m whose ratios were unchanged. The gene encoding for gluconate dehydratase, which catalyzes the dehydration of gluconate to 2-keto-3-deoxygluconate, has not been annotated in *T. acidophilum* [5]; however, we found by BLAST search that Ta0085m shows high homology (identity: 42%) to SSO3198, which was characterized as gluconate dehydratase involved in the non-phosphorylated ED pathway in *Sulfolobus solfataricus* [142]. Similar to the other proteins of the non-phosphorylated ED pathway, Ta0085m was also ~3-fold up-regulated at protein level in anaerobic conditions, and it might function as gluconate dehydratase in *T. acidophilum*. Ta0809 was characterized as glyceraldehyde dehydrogenase catalyzing the oxidation of glyceraldehyde to glycerate in *T. acidophilum* [153], while the KEGG pathway database annotates Ta0810 and Ta0834 as the enzymes, which convert glyceraldehyde to glycerate. Ta0834 failed to be identified in the proteomics study. Under anaerobiosis the expression level of Ta0809 remained unchanged, whereas Ta0810 showed 14-fold up-regulation. These results might indicate that Ta0810 takes over the glyceraldehyde dehydrogenase function of Ta0809 under anaerobic conditions, which needs further investigations.

The complete set of TCA cycle enzymes was identified in our proteomics study and they showed 2- to 9-fold up-regulation under anaerobic conditions, except for succinate dehydrogenase (membrane bound protein) for which we have no quantitation data.

Remarkably, the lower branch [154] of the EMP pathway enzymes, converting glyceraldehyde 3-phosphate to pyruvate, showed up-regulation in the absence of oxygen, whereas the enzymes catalyzing the six-carbon compound transformations in the upper branch [154] remained unchanged.

Compared to the strong up-regulation of the enzymes involved in the non-phosphorylated ED pathway, the EMP pathway lower branch, and the TCA cycle at protein level, the transcription level of these genes changed just slightly, if at all (Figure 4.3.10.C).


A
The main proteolytic pathways
Free amino acids
Substrate proteins
Tricorn cofactors F1, F2, F3
Ta0301
Ta0830
Ta0815
Ta0483
Ta1037
Proline dipeptidase
Ta0235
Ss1-like metallo-carboxypeptidase
Ta1490
Tricorn
Ta0664
Acylaminoacyl peptidase
Extracellular acid proteases
Ta0376
Ta1403
Ta0167
Ta0741m
Ta0976
Oligopeptides
Ta1329
Oligopeptide transporter OPP1
Ta1327
Ta1325
Ta1328
Ta1326
Native proteins
Oligopeptide transporter OPP2
Ta0272
Ta0261
Ta0263
Ta1266
Ta0275
Ta0264
Ta0266
Thermosome
α Ta0980
β Ta1276
Proteasome
α Ta1288
β Ta0612
GimC/prefoldin
Ta1076
Ta1137
HtpX-like metalloprotease
Ta0861
Non-native membrane proteins
VAT Ta0840
VAT2 Ta1175
Lon-2 Ta0098
Archaeal Lon protease
Ta1081
Native membrane proteins
Regulatory proteins
Non-native proteins
hypothetical transporters
Ta0017
Ta0087
Ta0144
Ta0149
Ta0240
Ta0300
Ta0503
Ta0661
Ta1070
Ta1108
Ta1195
Ta1371
Ta1378
Ta1469
Ta1470
hypothetical membrane proteases
Ta0062
Ta0407
Ta0839
Legend:
Protein mRNA
Up-regulated under anaerobiosis
Down-regulated under anaerobiosis
Unchanged
No ratio detected



* Ta0837 might have functions both in aerobic and anaerobic proton transport

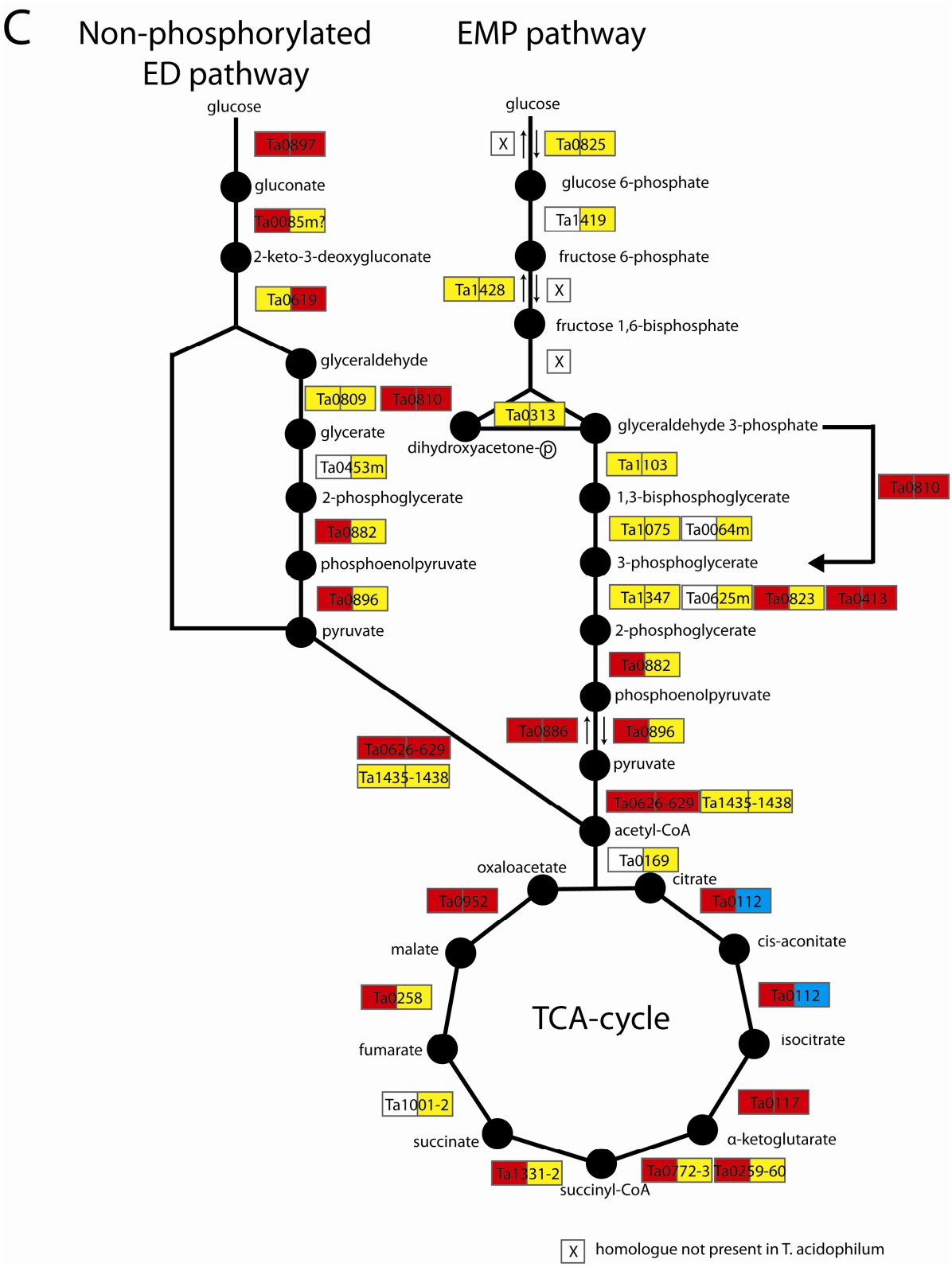

C
Non-phosphorylated ED pathway
EMP pathway
glucose
Ta0897
gluconate
Ta0085m?
2-keto-3-deoxygluconate
Ta0619
glyceraldehyde
Ta0809
Ta0810
glycerate
Ta0453m
2-phosphoglycerate
Ta0882
phosphoenolpyruvate
Ta0896
pyruvate
glucose
X
Ta0825
glucose 6-phosphate
Ta1419
fructose 6-phosphate
Ta1428
X
fructose 1,6-bisphosphate
X
Ta0313
dihydroxyacetone-p
glyceraldehyde 3-phosphate
Ta1103
1,3-bisphosphoglycerate
Ta1075
Ta0064m
3-phosphoglycerate
Ta1347
Ta0625m
Ta0823
Ta0413
2-phosphoglycerate
Ta0882
phosphoenolpyruvate
Ta0886
Ta0896
pyruvate
Ta0810
Ta0626-629
Ta1435-1438
Ta0626-629
Ta1435-1438
acetyl-CoA
Ta0169
oxaloacetate
Ta0952
citrate
Ta0112
cis-aconitate
Ta0112
isocitrate
Ta0117
α-ketoglutarate
Ta0772-3
Ta0259-60
succinyl-CoA
Ta1331-2
succinate
Ta1001-2
fumarate
Ta0258
malate
TCA-cycle
X
homologue not present in T. acidophilum

Figure 4.3.10: Schematic representations of proteolytic pathways, proton transport pathways, and glucose metabolism. Quantitative transcriptomics and proteomics data of enzymes involved in these pathways are presented. The left half of the boxes represents the protein expression ratio; the right half represents the mRNA expression ratio. Protein/mRNA levels increased under anaerobiosis are indicated in red; unchanged are indicated in yellow, and decreased under anaerobiosis are shown in blue; a colorless box means the proteins were either not detected or did not satisfy the quantitation restrictions. A) The main proteolytic pathways in T. acidophilum (modified from Ruepp et al. [5]). Red arrows indicate the metabolic pathway, green arrows the quality control pathway of damaged proteins and blue arrows the regulatory pathway. B) Aerobic and anaerobic proton transports were modified from the proton transport pathways of Thermoplasma volcanium [149]. The enzyme homologues in T. acidophilum were matched by BLAST search and displayed in the figure. C) Non-phosphorylated ED pathway, EMP pathway, and TCA cycle of T. acidophilum [5]. ' X ' means a homologue could not be identified in the genome.

2.2) Tryptophan biosynthesis

Transcriptomics data showed that tryptophan biosynthesis is one of the most highly up-regulated processes under aerobic conditions. Enzymes (Ta0803m-Ta0808), which catalyze the transformation of chrorismate to tryptophan, were more than 50-fold up-regulated under aerobic conditions at transcript level. The gene organization of Ta0803m-Ta0808 shows an operon structure [155]. The Secondary Structure Prediction analysis (http://bp.nuap.nagoya-u.ac.jp/sosui/) indicates that all these proteins are soluble proteins; however, their identification in the proteomics study failed (only Ta0808 was identified with 4 detected peptides). One explanation for their escape of identification in the cytoplasmic extract is that the proteins encoded by Ta0803m-Ta0808 form a complex and that they are attached/anchored to the membrane via an unknown process or that there is a still unknown membrane-associating subunit in the complex.

2.3) Exosomal superoperon

Koonin *et al.* [72] described a superoperon of exosomal genes in *Archaea* that in addition to the predicted exosome components, encodes the catalytic subunits of the proteasome, two ribosomal proteins and a DNA-directed RNA polymerase subunit. All these 17 predicted exosomal superoperon-encoded proteins were identified in our study. The proteomics and transcriptomics quantitation data showed that the majority of these genes were not regulated in the absence of oxygen. This result indicates a constant network of coregulation and functional and physical interactions in a striking range of central cellular functions in

T. acidophilum, including translation and cotranslational protein folding, RNA processing, degradation, modification, and transcription.

2.4) One carbon pool by folate

Five out of the 8 proteins that take part in one-carbon folate pool are 2- to 5-fold up-regulated under anaerobic conditions, while the transcription level of these genes shows only slight up-regulation (1.5- to 2-fold) under anaerobic conditions. This might suggest enhanced metabolic activity under anaerobiosis because the one-carbon folate pool can be used to gain reducing equivalents by various catabolic reactions and to provide C1 compounds for nucleotide, methionine, and panthotenate biosynthesis [156].

2.5) Coping with oxygen stress

The antioxidative enzymes in *T. acidophilum* are superoxide dismutase (Ta0013), peroxiredoxins (Ta0152, Ta0473, and Ta0954m) and the alkyl hydroperoxide reductase (Ta0125). The transcript level of Ta0013, Ta0125 and Ta0152 was 2- to 10-fold up-regulated under aerobic conditions which is in accordance with the high oxygen stress. However, their expression at protein level remained unchanged. At the moment we do not have good explanation for this phenomenon.

3) Regulated proteins belonging to the same family

In addition to the biochemical pathway clustering, regulated proteins were also analysed in terms of protein families, to provide comprehensive concepts on protein groups. At protein level, putative cytoskeletal proteins and oxidoreductases were highly up-regulated under anaerobic conditions; however, at mRNA level, no significant changes or less magnitude of regulations were shown.

Putative cytoskeletal proteins

T. acidophilum has been reported to contain a well-developed cytoskeleton as an adaptation for respiration upon elemental sulfur, which is insoluble in acidic conditions and can be metabolized only through direct contact [157, 158]. However, our proteomics and transcriptomics studies indicate that *T. acidophilum* might be able to solubilise the elemental sulphur through

an unknown process and perform cytosolic S^0 reduction, as we didn't find highly up-regulated membrane proteins encoding transcripts which can be assigned to S^0 respiration, in contrast, a cytosolic protein Ta0837, a homologue of the recently characterized NAD(P)H sulfure oxidoreductase (NSR, PF1186) which catalyzes the S^0 reduction in *Pyrococcus furiosus* [151], was 4 times up-regulated at anaerobic level and might be responsible for the cytosolic S^0 reduction in *T. acidophilum*. Therefore, it is not likely that a cytoskeleton is a prerequisite for S^0 respiration.

The putative intermediate filament protein homologue (Ta1488) was relatively abundant under both growth conditions (absolute intensity ~10^5 AU) and its expression level didn't show any change in the absence of oxygen, while the other potential cytoskeleton protein, the predicted actin homologue Ta0583 [137], showed significant up-regulation (3.4-fold) under anaerobiosis, which suggests that Ta0583 could be functionally related to anaerobic S^0 respiration. Ta0583 is lowly expressed (0.035% of the cytosolic proteins), therefore it is not likely that it is visible in cryo-EM investigations.

Oxidoreductases

The proteomics data showed that 43 of the anaerobically up-regulated proteins possess oxidoreductase activities, including 2-oxoacid-oxidizing enzymes with an iron-sulfur protein as electron acceptor, disulfide oxidoreductases with a disulfide as acceptor, and oxidoreductases acting on a sulfur group of donors. These proteins can not be assigned to a common biochemical pathway, whereas they are distributed throughout 33 different metabolic pathways, including amino acid, carbohydrate, energy, lipid, nucleotide, cofactor and vitamin metabolism. The mRNA level of these oxidoreductases remained unchanged or was up-regulated only to a very small extent.

Transcriptional regulatory proteins

The majority of the 18 identified transcriptional regulatory proteins, whose biological functions are unknown, were unchanged both at protein and mRNA levels. The exceptions are: Ta0868m and Ta0550m (predicted transcription regulators) with more than 10-fold down-regulation under anaerobic conditions at protein level, Ta1362m (Fe^{2+} uptake regulation protein) with more than 4-fold up-regulation under anaerobic conditions at protein

level, Ta0608 (transcription regulator related protein), Ta1339 (cell-cycle regulation histidine triad (HIT) related protein) and Ta1394 (Nta operon transcriptional regulator related protein) with over 2-fold up-regulation under anaerobic conditions at transcript level and Ta1110 (transcription regulator ArsR related protein) with over 2-fold down-regulation under anaerobic conditions at transcript level. These regulatory proteins might play crucial roles in the adaptation to changing O_2 concentrations in the environment, therefore detailed investigations are suggested.

4.3.2 Discussion

Here we present the first integrated quantitative proteomics and transcriptomics studies of *T. acidophilum* cultured under aerobic and anaerobic conditions. We chose these conditions, as in preliminary 2D-SDS-PAGE experiments we found significantly different protein expression patterns, and many of the differentially expressed proteins were subunits of protein complexes. We decided to investigate and quantitate both protein expression and mRNA transcription levels to correlate the two data sets. We chose a label-free quantitation method developed by Cox *et al*. [34] to quantitate the protein expression, and the NimbleGen cDNA microarrays to quantitate mRNA transcription levels, respectively.

The label-free quantitation approach

Generally, label-free quantitation is based on the comparison of the signal intensities of unlabelled peptides. Because the unlabelled samples need to be processed separately, any upstream sample-processing steps, like separation of proteins or peptides, potentially can cause systematic errors and thus influence the accuracy of the quantitation results. Therefore, a label-free quantitation method requires a very precise and accurate performance of the whole proteomics workflow. Compared to label-free quantitation, stable-isotope labelling quantitation based methods tolerate the signal fluctuations in HPLC and in mass spectrometers better and provide more accurate results. However, they have limitations:

While metabolic labelling can be applied to studies in higher animals, this method is not suitable for every organism [22,23]. We tried to label the *T. acidophlium* proteome by culturing the cells in the presence of SILAC labelled yeast extract (Lysine8-^{13}C6 ^{15}N2). However, the MS analysis showed that only 78% of the lysine-containing peptides were labelled (data not

shown). The incomplete lysine labelling could be due to the degradation of lysine8 and the synthesis of lysine via unknown biochemical pathways. In further experiments, to avoid this incomplete incorporation of modified amino acids during *in vivo* labelling of cells, a lysine-auxotrophic strain would be needed and tested instead of using the wild type *T. acidophilum*.

Chemical labelling is universally applicable [24, 25, 26, 27], but there are several factors that can affect the final quantitation accuracy. First, the labels are introduced relatively late in the sample processing, leaving room for the accumulation of systematic errors until the time of labelling. Then, the accuracy of a chemical labelling method is also influenced by the efficiency of chemical reactions, complex labelling procedures, and complicated purification steps. Furthermore, quantitation softwares for chemical labelling techniques are still in its infancy and thus hinders also the usage of this method.

Here, we have used a newly developed label-free method, in which sophisticated algorithmic steps and normalization procedures for data analyses were performed [34]. Therefore, the two key ingredients for the success of SILAC based quantitation – high peptide mass accuracies and high identification rates – were carried over and the errors introduced during the upstream sample preparation can mainly be eliminated. Hence, this label-free quantitation method can provide precise protein ratios on a proteome-wide scale.

Proteomics data

Compared to previous proteomics approaches, 2DE-MALDI-TOF-MS and MSC-LC-MS/MS [37, 140], the use of state-of-the-art mass spectrometric technology (1D-SDS-PAGE combined with LTQ-FTICR-MS) increased the number of identified proteins tremendously, resulting in 88% coverage of the *T. acidophilum* cytosolic proteome. The label-free quantitation approach – comparing the ion intensities of tryptic peptides – is relatively straightforward, avoids the usage of costly labelling reagents and is reproducible within acceptable limits. Using this approach, 604 proteins were quantified, the expression levels of 263 proteins of which were found to be affected by anaerobic conditions. The proteins identified and quantified represent a considerable proportion of the proteome, which is a great benefit for global functional studies.

Twelve percent of the annotated cytosolic proteins, being mostly hypothetical proteins with unknown functions, escaped detection. The reason for that could be their low Mw or their low expression level if they were expressed at all. The MS-based proteomics analysis has limitations in sensitivity and dynamic range. For proteins with low Mw and/or low abundances, much lower amounts of tryptic peptides can be generated and other highly abundant peptides can thus mask their presence, and therefore prevent their identification. Another possible reason for a protein not to be identified could be that it might adhere to the membrane or membrane-anchored proteins and is thus not present in the cytoplasmic sample.

The protein sets identified under aerobic and anaerobic conditions are highly overlapping (99%); however, with respect to the quantitation results, anaerobiosis caused significant alterations at protein expression level. These results strongly indicate that for the maintenance of normal cell growth and homeostasis, almost all proteins are expressed at a certain level. This prepares the cell for all the situations it has to cope with, which means in case of a sudden change in the cell's environment, that the cell can respond to it immediately without being damaged. Also during the adaptation period it has enough time to adjust the expression level of the proteins required to re-establish normal homeostasis.

Microarray data

A highly reproducible cDNA microarray-assisted transcriptomics analysis was performed and used to validate the proteomics data. Since membrane proteins are highly hydrophobic, it is difficult to study them by proteomics methods. Complementary to the proteomics studies, the transcriptomics analysis provides accurate, genome-wide information on regulatory responses, thus giving insights of differential expressions including both cytosolic and membrane gene transcripts.

A large proportion of the regulated genes are encoding for hypothetical proteins

Remarkably, 34% of the 263 regulated proteins and 46% of the 445 regulated mRNAs are annotated as hypothetical proteins. Such a high proportion of regulated hypothetical proteins affected by anaerobic culture conditions stresses the necessity to investigate their function(s) in the cell**.**

Low correlation between regulated proteins and mRNAs

A thorough quantitation analysis shows that the anaerobic growth conditions cause extensive gene regulation (~30%) both at protein and mRNA levels. It is commonly assumed that mRNA abundance is a surrogate for protein amounts. However, our results indicate that this assumption is often not valid, as we only found a low correlation (0.57) between mRNA and protein level changes, and the overlap of the regulated proteins and mRNAs is only 30%. Similar results were reported recently in mouse ESCs [53], *Drosophila* [159] and yeast [160, 161], which confirm our results, indicating that the poor agreements between transcriptome and proteome data are generally present. One reason of this could be, that the amount of a protein within a cell does not only depend on its transcription rate, but also on several posttranscriptional factors such as the stability of the mRNA, its translation efficiency and, finally, the turnover rate of the protein itself [159]. Another possible reason would be that the amount of the RNA transcribed and protein translated does not reflect the cumulative cellular activity of a protein well. Enzyme activities can be significantly influenced by cofactors, specific molecules, ions or other regulators, locally altered pH and other means [150]. We can hypothesize that a number of enzymes might be solely active or show enhanced activity under aerobiosis or anaerobiosis. Therefore, the change of enzyme activity in response to culture conditions could be an important factor to be investigated. To prove this theory, enzyme activities should be measured under *in vivo* conditions. Alternatively, a metabolomic approach which studies the quantity of all metabolites present in the cell at a given condition appropriate might be useful.

Furthermore, several groups of proteins might be unstable in the presence of oxygen. They could be damaged under oxidative stress which in turn may cause the difference between mRNA and protein amounts. For example, proteins belonging to the oxidoreductase families exhibited significant up-regulation in the absence of oxygen, while the mRNA amounts of these genes remained unchanged or were just slightly up-regulated. The dominant molecular functions connected to these up-regulated oxidoreductases are related to electron-transfer processes through sulfur-related groups. Previous studies have shown that the sulfur cluster-containing genes, e.g. thiol/disulfide redox balance and iron-sulfur clusters, are specifically sensitive to oxygen and inactivated by superoxides [162, 163]. Oxygen can convert the exposed

iron-sulfur clusters to unstable forms that decompose rapidly [163], whilst the structures of these iron-sulfur clusters have substantial inherent stability under anaerobic conditions. These sulfur-related electron transfers might play important roles in the anaerobic redox reactions and sulfur metabolism. Further detailed investigations of functions of this group of proteins and their stabilities under oxidative stress are needed.

Cobalamin (vitamin B_{12}) biosynthesis

Despite the low correlation between the transcriptomics and proteomics data, the increase of cobalamin (vitamin B_{12}) biosynthesis in aerobic condition was proven by both datasets. Vitamin B_{12} is normally involved in DNA synthesis/regulation, fatty acid synthesis and energy production. Moreover, the most abundant protein in aerobic cultures, ribonucleotide reductase (Ta1475), which catalyzes the conversion of purine and pyrimidine nucleotides to deoxynucleotides and provides monomeric precursors essential for DNA replication and repair [68], needs vitamin B_{12} to carry out its catalytic functions. The high up-regulation of vitamin B_{12} biosynthesis under aerobic conditions might indicate a higher turnover of DNA replication and repair when compared to anaerobiosis; however, most of the DNA repair proteins remained unchanged both at protein and mRNA level which is in conflict with this presumption. Further studies are needed to explain the phenomena of the up-regulation of Ta1475 and cobalamin biosynthesis in aerobic cells.

Tryptophan biosynthesis

Tryptophan biosynthetic genes exhibited high up-regulation under aerobic condition at the transcript level. NAD^+, NADP and a number of important biogenic amines and quinone compounds are derived from the aromatic amino acids or from their precursors. NAD^+/NADP is essential for electron transfer reactions. The increase of tryptophan biosynthesis under aerobic conditions indicates highly NAD^+/NADP dependent redox processes under aerobic conditions, whereas under anaerobic conditions, the electron transfers are mostly performed through ferredoxin, which has a lower redox potential than NAD^+/NADP [164].

Glucose metabolism

Only the lower branch of the EMP pathway enzymes, converting glyceraldehyde 3-phosphate to pyruvate, showed up-regulation in the absence of oxygen; whereas the enzymes catalyzing the six-carbon compound transformations in the upper branch of the EMP pathway remained unchanged. The lower branch of the EMP pathway is well conserved and prevalent among the three domains of life, which is unsurprising, as it is the key connecting point for pathways involved in the amino acid, pentose-phosphate and purine precursor generation for the further synthesis of essential biomolecules [154, 165]. In contrast, the enzymes of the upper part vary greatly, especially in archaea and bacteria [154], i.e. *T. acidophilum* lacks homologues for phosphofructokinase and fructose bisphosphate aldolase, thus, the presence of a working EMP pathway in *T. acidophilum* could not be confirmed [5]. The different expression levels of upper and lower parts of the EMP pathway suggest that *T. acidophilum* might utilize alternative enzymes for the six-carbon compound transformations in the EMP pathway.

In *Saccharomyces cerevisiae*, a eukaryotic system, posttranscriptional regulation of glycolysis during anaerobic fermentation has been reported [161] and glycolytic mRNAs were shown to be poor indicators of glycolytic flux [166]. Previous studies of central carbon metabolism of *S. cerevisiae* have shown that in the absence of oxygen, where respiration can no longer occur, substrate-level phosphorylation in glycolysis is essential to sustain cellular growth. The TCA-cycle has been shown not to work as a cycle under anaerobic conditions in *S. cerevisiae* [167]. The net-gain of ATP during fermentation is only two molecules, while in aerobic respiration the net-gain is 36 molecules of ATP. To compensate for the low ATP yield, the glycolytic flux needs to be accelerated in *S. cerevisiae* during anaerobic fermentation [166].

However, there is no fermentation and substrate-level phosphorylation in *T. acidophilum*. Instead, there is S^0 respiration and complete glucose degradation through the ED pathway and energy generation via TCA-cycle under anaerobiosis. We found that the enzymes involved in glucose degradation were up-regulated anaerobically at protein level but not at transcript level, which indicated that modulations at posttranscriptional level play a major role in glucose metabolism in *T. acidophilum*. Our study stresses the necessity to investigate the molecular mechanisms involved in the regulation of translation and degradation of glycolytic proteins in *T. acidophilum*. Under anaerobic conditions, *T. acidophilum* uses S^0 as a terminal

electron acceptor, which has larger redox potentials than O_2, meaning that anaerobic S^0 respiration ($\Delta G^{0'} = -333$ kJ) is less energy-efficient than aerobic respiration ($\Delta G^{0'} = -2844$ kJ) [168]. Similar to the accelerated glycolytic flux during fermentation in *S. cerevisiae*, the increased glucose degradation in anaerobically cultured *T. acidophilum* could be a response to compensate for the low energy yield of anaerobic S^0 respiration.

The complete protein set of the TCA-cycle has been quantified in our proteomics study. The TCA pathway is important for energy release. It was highly up-regulated under anaerobic conditions as well, which could be due to the increased consumption of nutrients to compensate for the low energy yield during anaerobic S^0 respiration. Moreover, the TCA pathway is also essential for intermediary metabolism, and many intermediates are drawn out of the cycle to be used as precursors in a variety of biosynthetic pathways. The GO analysis of the regulated genes showed up-regulation of TCA intermediate metabolism, generation of precursor metabolites and energy, under anaerobiosis. Therefore, the increased expression of TCA cycle and conjunct intermediate metabolism proteins might suggest a higher intermediary metabolic activity under anaerobic conditions.

When the TCA cycle works in the reverse direction, it is called reductive TCA-cycle. The reductive TCA-cycle is associated with anaerobic energy production and electron transport through Fe^{3+} reduction in *Ferroplasma acidarmanus* [169]. Whether the dramatic up-regulation of TCA-cycle in *T. acidophilum* is related to the anaerobic energy production and electron transport through S^0 is still an open question.

On the other hand, the reductive TCA cycle can act as a mechanism of autotrophic CO_2 fixation as it has been found in certain nonphototrophic hyperthermophiles, including *Sulfolobus*, *Thermoproteus* and *Aquifex* [168]. However, citrate lyase, a key enzyme of CO_2 fixation pathway that cleaves citrate into acetyl-CoA and oxaloacetate, was not identified in *T. acidophilum*. The *in vitro* enzyme assay of citrate lyase [170] of *T. acidophilum* cytosol didn't show any activity (data not shown). It might be that the citrate lyase activity is sensitive to oxygen, therefore anaerobic or *in vivo* measurement of citrate lyase activity is needed.

The uptake and degradation of exogenous polypeptides

Remarkably, the transcription levels of more than 40% of the membrane protein encoding genes showed significant changes. The functional clustering of these regulated genes shows up-regulation of transporters and membrane-anchored proteases under aerobic conditions, which results in increased ion transport and increased uptake of extracellular polypeptides and amino acids. This finding, together with the up-regulated glucose metabolism in anaerobically grown cells, can lead to the hypothesis that there is substrate preference for cells growing aerobically or anaerobically. Aerobic cells might use more peptides, oligopeptides and proteins than anaerobically growing cells, while the latter ones use more glucose to compensate for low energy yield.

Protein biosynthesis and protein fate

Most of the ribosomal proteins and proteins involved in translation initiation/elongation factors remained unchanged under anaerobic conditions both at protein and mRNA level. Moreover, genes encoded by an exosomal superoperon (Koonin *et al.* [72]) were also not regulated under anaerobiosis. These observations suggest that the core translational machinery and RNA processing and degradation processes are stably maintained and the overall translation rate is in fact close to equal under aerobic and anaerobic conditions. Surprisingly, key enzymes of protein folding, modification and degradation processes, like tricorn, thermosome, proteosome, VAT and molecular chaperon proteins, were found to be highly up-regulated under anaerobic conditions. In transcriptomics and proteomics studies of yeast, up-regulation of proteasome and several chaperones were also detected under anaerobiosis. [161] The reason for the enhanced expression of these proteins belonging to the functional category of 'protein fate' is not known. One factor to consider might be that, energetically, the anaerobic cells are less effective and to maintain an effective protein quality control saves energy.

Proton transport pathways

The proton transport pathways in *T. acidophilum* are not studied yet; however, homologues of all enzymes involved in putative proton transport pathways of *T. volcanium* [149] are present in *T. acidophilum*. Based on the study of *T. volcanium,* aerobic proton transports are membrane-

associated processes, catalyzed by membrane proteins, whereas the putative anaerobic proton transport pathway seems mainly consisted of cytoplasmic proteins. Our proteomics and transcriptomics studies indicate that *T. acidophilum* might be able to solubilise the elemental sulphur through an unknown process and perform cytosolic S^0 reduction, as we didn't find highly up-regulated membrane proteins which could be assigned to S^0 respiration. Instead, a predicted cytosolic sulphur respiration enzyme, the sulfide-quinone reductase homologue (Ta1129) [5], with an 6.7-fold increase at protein level under anaerobic conditions was identified. Additionally, Schut *et al.* reported a novel S^0-reducing system involving NAD(P)H sulfure oxidoreductase (NSR, PF1186) and a membrane bound oxidoreductase (MBX, PF1441-1453) in the hyperthermophilic archaeon *Pyrococcus furiosus* [151]. MBX and NSR are proposed to be the key enzymes responsible for the reoxidation of ferredoxin and NAD(P)H [151]. In *T. acidophilum*, Ta0837, a homologue to PF1186, was 4-fold up-regulated under anaerobic conditions. Whether *T. acidophilum* uses Ta0837 and Ta1129 as replacements of Ta0046-47 in sulfur metabolism remains unclear. There is no MBX complex in *T. acidophilum*, but 9 of the NADH dehydrogenase subunits (Ta0959, Ta0960, Ta0961, Ta0962, Ta0965, Ta0966, Ta0967m, Ta0968 and Ta0969m) show sequence similarities with MBX subunits (PF1441-1448). Together with the detected up-regulation of two NADH dehydrogenase subunits (Ta0969m and Ta0970m), under anaerobic conditions, we might speculate that anaerobic NADH dehydrogenase in *T. acidophilum* has similar functions like MBX, and is involved in the S^0 reduction process. Further research targeting the precise role and structure of NADH dehydrogenase and mechanisms of S^0 reduction is needed.

Macromolecular complexes

We quantified 28 macromolecular complexes (Mw over 300 kDa), the data of which can be used to correlate to and validate cryo-ET data.

The regulation patterns of those complexes whose genes are organized in operons were consistent both at protein and mRNA level, e.g. FIX Ta0326-328, Ta0259-260, Ta1412-1413, Ta0423-425, Ta0574-575, Ta0390-392 and exosome (Ta1292-1294), proving the accuracy of the proteomics and trancriptomics quantitation. On the other hand, the quantitation data of proteins and mRNAs can be used to confirm the subunit composition of complexes, and thus, serve as evidence for the stoichiometry of respective putative complexes.

The subunits of chromosome segregation protein complex (Ta0157 and Ta0158), ribosome and NADH dehydrogenase (Ta0959-Ta0979) showed different quantitation ratios under different conditions. Therefore, the structure of these complexes should be investigated experimentally, to prove whether they possess different structures adapted to aerobic or anaerobic conditions. To investigate this further, ribosomes were purified from aerobically and anaerobically cultured *T. acidophilum* cells and quantified with the same label-free quantitation method used in the project (data not shown). The results couldn't prove different expression levels of ribosomal proteins. In contrast, all ribosomal proteins quantified showed ratios around 1, suggesting that the ribosome maintains a constant composition in both aerobic and anaerobic cells. Previous studies have also shown up- or down-regulations of several ribosomal proteins following biological treatments [161, 171]. The different expression levels of 50S ribosomal proteins L10 (Ta1057m, decreased by 10-fold), L31 (Ta0054, decreased by 10-fold) and L11 (Ta0361, increased by 2-fold) might be caused by properties in addition to those associated with ribosomal structure and function, such as protein synthesis, intracellular trafficking, protein transport and apoptosis [172]. For example, in a eukaryotic system, L10 is supposed to be involved in the translational control of gene expression [172], L11 plays a central role on ribosomal conformational changes during translocation [173], and L31 shows functional significance in cancer studies [174].

Altogether, comprehensive quantitative proteomics and transcriptomics analyses of *T. acidophilum* cultured under aerobic and anaerobic conditions enabled us to investigate the cellular response to a changing environment. Several important biological responses to anaerobiosis were found. The quantitation data of transcriptomics and proteomics analyses showed low correlation, which demonstrates the necessity to investigate posttranscriptional regulations in *T. acidophilum*. Future work should include studies of the membrane proteins, especially the 15 hypothetical membrane transporters and 3 proteases, posttranscriptional regulation of the glucose metabolism, and anaerobic S^0 respiration.

5 Conclusions and perspectives

The cytosolic proteome of *T. acidophilum* was investigated by three MS-based proteomics approaches. In the first attempt, 2DE was coupled to MALDI-TOF-MS to obtain an initial catalogue of the expressed proteins. We extended this project by using glycerol gradient ultracentrifugation to separate protein complexes. Protein complex-containing fractions were further analysed by 2DE-MALDI-TOF-MS. Due to the limited resolution of the 2D gels and the relatively low identification ability of MALDI-TOF-MS, only 271 proteins and 16 putative macromolecular complex-forming proteins were identified. After these experiments, we concluded that many of these complexes with Mw above 1 MDa are present in the cells at low abundance and that the number of the proteins identified did not cover the complete aerobic proteome of *T. acidophilum*.

Therefore, in the next experiment, we chose a more robust and sensitive method to improve the protein identification rate. The cytoplasmic extract of *T. acidophilum* was fractionated by MSC to lessen sample complexity, and the fractions were analysed by LC-MS/MS. This method enabled us to identify 466 proteins, of which 187 were found in the high Mw fractions. A large proportion of the proteins in *T. acidophilum* are organized into complexes (dimers, trimers, tetramers, etc.), indicating a trend to avoid monomeric forms. We identified 35 macromolecular complexes with sizes over 300 kDa, of which 10 exceeded their functional homologues from other organisms in terms of size, supporting the idea that composition and structure of many complexes can vary among species despite the high sequence homology of their subunits, and therefore need to be investigated individually. We encountered protein precipitation and protein complex dissociation during sample separation that caused uncertainties in determining the subunit composition of several complexes. To our knowledge, there is no universal method which would be effective to keep all complexes intact. Therefore, the buffer conditions should be determined experimentally on an individual basis, regardless of the protein separation method (gel filtration, antibody-based purification, affinity chromatography etc.) used to isolate them.

As a further means of refining the protein inventory, we used state-of-the-art mass spectrometric technologies (a combination of 1D-SDS-PAGE and high sensitive LTQ-FTICR-MS) which increased the identification to more than 1000 proteins, covering almost the whole (88%) cytoplasmic proteome of *T. acidophilum*.

Following the identification of the protein complexes, thorough quantitative analyses were carried out. The cytosolic proteome and the genome-wide transcriptome of aerobically and anaerobically cultured *T. acidophilum* cells were compared. We found that most of the cytosolic proteins were expressed at a certain level under both conditions, which might mean that *T. acidophilum* is always prepared for coping with the environmental challenges it faces within the shortest possible time interval. The anaerobic growth conditions caused significant gene regulations (~30%) both at protein and mRNA levels. Remarkably, the expression level of many of the hypothetical proteins changed dramatically under anaerobiosis, indicating their crucial role in adaptation to environmental challenges. Membrane proteins play vital roles in the communication between the cell and its environment. Mostly those genes which are functionally associated with ATPase activity, endopeptidase activity, ion and amino acid transports were down-regulated under anaerobic conditions at the mRNA level. Moreover, the non-phosphorylated ED-pathway, the lower branch of the EMP pathway and the TCA cycle were up-regulated significantly, indicating accelerated glucose degradation under anaerobic conditions, which might compensate for the low energy yield during anaerobic S^0 respiration. Many of the putative anaerobic proton transport proteins were shown to be cytosolic. Most of the enzymes participating in this pathway form complexes, five of which are over 300 kDa. Therefore, they can serve as templates for visual proteomics studies with respect to complete pathways. The transcriptome and proteome data show only weak positive correlation, which indicates extensive posttranscriptional regulation mechanisms. On the other hand, a low correlation between the two datasets can originate from the fact that, under aerobiosis or anaerobiosis, proteins can behave differently. Therefore the measurement of *in vivo* enzyme activities would be an important factor in drawing final conclusions on a system biology level. Integration of transcriptomics, proteomics and metabolomics studies, together with the measurement of *in vivo* enzyme activities would help us to generate coherent hypotheses and discover new emergent properties that arise from the systemic view.

Further studies on the membrane proteome would provide valuable information on the adaptation capability of *T. acidophilum* to changing environments, as more than 40% of the membrane protein-encoding genes were regulated at mRNA level in response to anaerobiosis. Another promising field to investigate is the possible posttranscriptional regulation mechanisms which might help us to explain the low correlation between transcriptomics and

proteomics data. Furthermore, an extensive research on the hypothetical and partially characterized proteins and protein complexes will facilitate the creation of a template library which can be used in the 'visual proteomics' approach to map the macromolecular complexes in *T. acidophilum*.

6 Abbreviations

1D	one-dimensional
1D-SDS-PAGE	one-dimensional gel electrophoresis
2D	two-dimensional
2DE	two-dimensional gel electrophoresis
3D	three-dimensional
AAA ATPases	ATPases associated with a variety of cellular activities
ACN	acetonitrile
AdoMet	adenosylmethionine
AhpC	alkyl hydroperoxide reductase subunit C
AhpF	alkyl hydroperoxide reductase subunit F
ATP	adenosine triphosphate
AU	arbitrary units
BLAST	basic local alignment search tool
CDD	Conserved Domain Database
cDNA	complementary DNA
CID	collision induced fragmentation
cMDH	cytosolic malate dehydrogenase
COG	clusters of orthologous groups
cryo-ET	cryo-electron tomography
CV	coefficient of variation
Cy	carbocyanin
Cy3	indocarbocyanin
Cy5	indodicarbocyanin
DTT	dithiothreitol
ED pathway	Entner-Doudoroff pathway
EM	electron microscopy
EMP pathway	Embden-Meyerhof-Parnas pathway
ESI	electrospray ionization
ET	electron tomography
FTICR	fourier transform ion cyclotron resonance
GeLCMS	a combination of one-dimensional gel electrophoresis and on-line electrospray tandem mass spectrometry
GO	Gene Ontology
ICAT	isotope-coded affinity tags
ICPL	isotope coded protein label
IEF	isoelectric focusing
IPG	immobilized pH gradient
KEGG	Kyoto Encyclopedia of Genes and Genomes
KH	ribonucleoprotein K-homology
LC	liquid chromatography

LC-MS/MS	liquid chromatography tandem mass spectrometry
LTQ	linear quadrupole ion trap
m/z	mass to charge ratio
MALDI	matrix-assisted laser desorption/ionization
Mbp	mega base pairs
MDa	mega dalton
MIPS	munich information center for protein sequences
mRNA	messenger RNA
MS	mass spectrometry
MSC	molecular sieve chromatography
Mw	molecular mass
NCBI	National Center for Biotechnology Information
NL	non linear
NSI	nanoelectrospray ion
ORF	open reading frame
OTCase	ornithine carbamoyltransferase
PA	polyacrylamide
PDB	Protein Data Bank
PEDANT	Protein Extraction, Description and ANalysis Tool
p*I*	isoelectric point
ppm	parts-per-million
rpm	revolutions per minute
RuBisCO	ribulose-1,5-bisphosphate carboxylase/oxygenase
Sf	ScoreFinal
SILAC	stable isotope labelling by amino acids in cell culture
SOD	Fe-superoxide dismutase
TCA	trichloroacetic acid
TCA cycle	tricarboxylic acid cycle
TFA	trifluoroacetic acid
THF	tetrahydrofolate
TOF	time-of-flight
VAT ATPase	valosine-containing protein-like ATPase from *Thermoplasma acidophilum*

7 References

1. Segerer, A., Langworthy, T.A. & Stetter, K.O. Thermoplasma-acidophilum and Thermoplasma-volcanium sp-nov from sulfatara fields. *Syst. Appl. Microbiol.* **10**, 161-171 (1988).
2. Schleper, C. et al. Picrophilus gen-nov, fam-nov - a novel aerobic, heterotrophic, thermoacidophilic genus and family comprising archaea capable of growth around pH-0. *J. Bacteriol.* **177**, 7050-7059 (1995).
3. Darland, G., Brock, T.D., Samsonof.W & Conti, S.F. Thermophilic, acidophilic mycoplasma isolated from a coal refuse pile. *Science* **170**, 1416-1418 (1970).
4. Smith, P.F., Langwort.Ta, Mayberry, W.R. & Hougland, A.E. Characterization of membranes of *Thermoplasma acidophilum*. *J. Bacteriol.* **116**, 1019-1028 (1973).
5. Ruepp, A. et al. The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature* **407**, 508-513 (2000).
6. Cowan, D. Genomics - Use your neighbour's genes. *Nature* **407**, 466-467 (2000).
7. Simpson, R.J. Proteins and proteomics: a laboratory manual. (Cold Spring Harbor, New York; 2002).
8. Nie, L., Wu, G., Culley, D.E., Scholten, J.C. & Zhang, W. Integrative analysis of transcriptomic and proteomic data: challenges, solutions and applications. *Critical reviews in biotechnology* **27**, 63-75 (2007).
9. Lockhart, D.J. & Winzeler, E.A. Genomics, gene expression and DNA arrays. *Nature* **405**, 827-836 (2000).
10. Rho, S., You, S., Kim, Y. & Hwang, D. From proteomics toward systems biology: integration of different types of proteomics data into network models. *BMB reports* **41**, 184-193 (2008).
11. Rinaldis, E. & Lahm, A. DNA microarrays current applications. (Horizon bioscience, Norfolk NR18 0JA; 2007).
12. Shalon, D., Smith, S.J. & Brown, P.O. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res* **6**, 639-645 (1996).
13. Tang, T. et al. Expression ratio evaluation in two-colour microarray experiments is significantly improved by correcting image misalignment. *Bioinformatics (Oxford, England)* **23**, 2686-2691 (2007).
14. Steen, H. & Mann, M. The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.*

5, 699-711 (2004).

15. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198-207 (2003).
16. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. & Whitehouse, C.M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71 (1989).
17. Karas, M. & Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **60**, 2299-2301 (1988).
18. Scigelova, M. & Makarov, A. Orbitrap mass analyzer--overview and applications in proteomics. *Proteomics* **6 Suppl 2**, 16-21 (2006).

19. Syka, J.E. et al. Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications. *Journal of proteome research* **3**, 621-626 (2004).
20. Dubois, F. et al. A comparison between ion-to-photon and microchannel plate detectors. *Rapid Communications in Mass Spectrometry* **13**, 786-791 (1999).
21. Park, M.A., Callahan, J.H. & Vertes, A. An inductive detector for time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* **8**, 317–322 (1994).
22. Ong, S.E. et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376-386 (2002).
23. Washburn, M.P., Ulaszek, R., Deciu, C., Schieltz, D.M. & Yates, J.R., 3rd Analysis of quantitative proteomic data generated via multidimensional protein identification technology. *Analytical chemistry* **74**, 1650-1657 (2002).
24. Gygi, S.P. et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature biotechnology* **17**, 994-999 (1999).
25. Schmidt, A., Kellermann, J. & Lottspeich, F. A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* **5**, 4-15 (2005).
26. Shevchenko, A. et al. Rapid 'de novo' peptide sequencing by a combination of nanoelectrospray, isotopic labeling and a quadrupole/time-of-flight mass spectrometer. *Rapid Commun Mass Spectrom* **11**, 1015-1024 (1997).
27. Yao, X., Freas, A., Ramirez, J., Demirev, P.A. & Fenselau, C. Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Analytical chemistry* **73**, 2836-2842 (2001).
28. Fang, R. et al. Differential label-free quantitative proteomic analysis of Shewanella oneidensis cultured under aerobic and suboxic conditions by accurate mass and time tag approach. *Mol Cell Proteomics* **5**, 714-725 (2006).
29. Ono, M. et al. Label-free quantitative proteomics using large peptide data sets generated by nanoflow liquid chromatography and mass spectrometry. *Mol Cell Proteomics* **5**, 1338-1347 (2006).
30. Cutillas, P.R. & Vanhaesebroeck, B. Quantitative profile of five murine core proteomes using label-free functional proteomics. *Mol Cell Proteomics* **6**, 1560-1573 (2007).
31. Old, W.M. et al. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics* **4**, 1487-1502 (2005).
32. Silva, J.C., Gorenstein, M.V., Li, G.Z., Vissers, J.P. & Geromanos, S.J. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol Cell Proteomics* **5**, 144-156 (2006).
33. Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W. & Gygi, S.P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 6940-6945 (2003).
34. Cox, J., Luber, C., Nagaraj, N. & Mann, M. Novel normalization and protein quantitation techniques for comprehensive and accurate ralative label-fee quantitative proteomics of cell lines and tissues. *in preparation* (2008).

35. Nickell, S., Kofler, C., Leis, A.P. & Baumeister, W. A visual approach to proteomics. *Nat Rev Mol Cell Biol* **7**, 225-230 (2006).
36. Robb, F.T. & Place, A.R. Archaea: A Laboratory Manual, Thermophiles. (Cold Spring Harbor Laboratory Press, 1995).
37. Sun, N. et al. Proteomics Analysis of Thermoplasma acidophilum with a Focus on Protein Complexes. *Mol Cell Proteomics* **6**, 492-502 (2007).
38. Laemmli, U.K. Cleavage of structural proteins during assembly of head of bacteriophage-T4. *Nature* **227**, 680-685 (1970).
39. Neuhoff, V., Arold, N., Taube, D. & Ehrhardt, W. Improved staining of proteins in polyacrylamide gels including isoelectric-focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* **9**, 255-262 (1988).
40. Mortz, E., Krogh, T.N., Vorum, H. & Gorg, A. Improved silver staining protocols for high sensitivity protein identification using matrix-assisted laser desorption/ionization-time of flight analysis. *Proteomics* **1**, 1359-1363 (2001).
41. Tebbe, A. et al. Analysis of the cytosolic proteome of *Halobacterium salinarum* and its implication for genome annotation. *Proteomics* **5**, 168-179 (2005).
42. Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data *Electrophoresis* **20**, 3551-3567 (1999).
43. Tang, H.Y. et al. A novel four-dimensional strategy combining protein and peptide separation methods enables detection of low-abundance proteins in human plasma and serum proteomes. *Proteomics* **5**, 3329-3342 (2005).
44. Omenn, G.S. et al. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* **5**, 3226-3245 (2005).
45. Peng, J., Elias, J.E., Thoreen, C.C., Licklider, L.J. & Gygi, S.P. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *Journal of proteome research* **2**, 43-50 (2003).
46. Altschul, S.F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* **25**, 3389-3402 (1997).
47. Schaffer, A.A. et al. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic acids research* **29**, 2994-3005 (2001).
48. Marchler-Bauer, A. et al. CDD: a conserved domain database for interactive domain family analysis. *Nucleic acids research* **35**, D237-240 (2007).
49. Kanehisa, M. et al. From genomics to chemical genomics: new developments in KEGG. *Nucleic acids research* **34**, D354-357 (2006).
50. Berman, H.M. et al. The Protein Data Bank. *Nucleic acids research* **28**, 235-242 (2000).

51. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature protocols* **1**, 2856-2860 (2006).

52. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Analytical chemistry* **75**, 663-670 (2003).

53. Graumann, J. et al. SILAC-labeling and proteome quantitation of mouse embryonic stem cells to a depth of 5111 proteins. *Mol Cell Proteomics* (2007).

54. Olsen, J.V., Ong, S.E. & Mann, M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics* **3**, 608-614 (2004).

55. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* **26**, 1367-1372 (2008).

56. Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567 (1999).

57. Cox, J., C., L. & Mann, M. Proteome-wide accurate label-free quantitation of human cells. *HUPO 2008, Amsterdam* (2008).

58. Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics (Oxford, England)* **21**, 3448-3449 (2005).

59. Yusupov, M.M. et al. Crystal structure of the ribosome at 5.5 angstrom resolution. *Science* **292**, 883-896 (2001).

60. Schoehn, G., Hayes, M., Cliff, M., Clarke, A.R. & Saibil, H.R. Domain rotations between open, closed and bullet-shaped forms of the thermosome, an archaeal chaperonin. *J. Mol. Biol.* **301**, 323-332 (2000).

61. Baumeister, W., Walz, J., Zuhl, F. & Seemuller, E. The proteasome: paradigm of a self-compartmentalizing protease. *Cell* **92**, 367-380 (1998).

62. Pearson, J.T., Dabrowski, M.J., Kung, I. & Atkins, W.M. The central loop of *Escherichia coli* glutamine synthetase is flexible and functionally passive. *Arch. Biochem. Biophys.* **436**, 397-405 (2005).

63. Rahman, R., Jongsareejit, B., Fujiwara, S. & Imanaka, T. Characterization of recombinant glutamine synthetase from the hyperthermophilic archaeon *Pyrococcus* sp. strain KOD1. *Appl. Environ. Microbiol.* **63**, 2472-2476 (1997).

64. Coles, M. et al. The solution structure of VAT-N reveals a 'missing link' in the evolution of complex enzymes from a simple beta alpha beta beta element. *Curr. Biol.* **9**, 1158-1168 (1999).

65. Jeon, S.J. & Ishikawa, K. Characterization of novel hexadecameric thioredoxin peroxidase from *Aeropyrum pernix* K1. *J. Biol. Chem.* **278**, 24174-24180 (2003).

66. Massant, J., Wouters, J. & Glansdorff, N. Refined structure of *Pyrococcus furiosus* ornithine carbamoyltransferase at 1.87 A. *Acta Crystallogr Sect D-Biol Crystallogr* **59**, 2140-2149 (2003).

67. Legrain, C. et al. Biochemical characterisation of ornithine carbamoyltransferase from *Pyrococcus furiosus*. *Eur. J. Biochem.* **247**, 1046-1055 (1997).
68. Stubbe, J., Ge, J. & Yee, C.S. The evolution of ribonucleotide reduction revisited. *Trends Biochem. Sci.* **26**, 93-99 (2001).
69. Graham, D.E., Bock, C.L., Schalk-Hihi, C., Lu, Z.C.J. & Markham, G.D. Identification of a highly diverged class of S-adenosylmethionine synthetases in the archaea. *J. Biol. Chem.* **275**, 4055-4059 (2000).
70. Stark, H. et al. Visualization of elongation factor Tu on the *Escherichia coli* ribosome. *Nature* **389**, 403-406 (1997).
71. Prangishvilli, D., Zillig, W., Gierl, A., Biesert, L. & Holz, I. DNA-dependent RNA-polymerases of thermoacidophilic archaebacteria. *Eur. J. Biochem.* **122**, 471-477 (1982).
72. Koonin, E.V., Wolf, Y.I. & Aravind, L. Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach. *Genome Res.* **11**, 240-252 (2001).
73. Ruepp, A. et al. The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature* **407**, 508-513 (2000).
74. Gutsche, I., Essen, L.O. & Baumeister, F. Group II chaperonins: new TRiC(k)s and turns of a protein folding machine. *J. Mol. Biol.* **293**, 295-312 (1999).
75. Tamura, N., Lottspeich, F., Baumeister, W. & Tamura, T. The role of tricorn protease and its aminopeptidase-interacting factors in cellular protein degradation. *Cell* **95**, 637-648 (1998).
76. Roth, J.R., Lawrence, J.G., Rubenfield, M., Kiefferhiggins, S. & Church, G.M. Characterization of the cobalamin (vitamin-B_{12}) biosynthetic genes of *Salmonella typhimurium*. *J. Bacteriol.* **175**, 3303-3316 (1993).
77. Larsson, K.M. et al. Structural mechanism of allosteric substrate specificity regulation in a ribonucleotide reductase. *Nat Struct Mol Biol* **11**, 1142-1149 (2004).
78. Jabrin, S., Ravanel, S., Gambonnet, B., Douce, R. & Rebeille, F. One-carbon metabolism in plants. Regulation of tetrahydrofolate synthesis during germination and seedling development. *Plant Physiol.* **131**, 1431-1439 (2003).
79. Lu, Z.C.J. & Markham, G.D. Enzymatic properties of S-Adenosylmethionine synthetase from the archaeon *Methanococcus jannaschii*. *J. Biol. Chem.* **277**, 16624-16631 (2002).
80. Tabor, C.W. & Tabor, H. Polyamines. *Annu. Rev. Biochem.* **53**, 749-790 (1984).
81. Cantoni, G.L. Biological methylation - selected aspects. *Annu. Rev. Biochem.* **44**, 435-451 (1975).
82. Frey, P.A. Radical mechanisms of enzymatic catalysis. *Annu. Rev. Biochem.* **70**, 121-148 (2001).
83. Knapp, S. et al. Refined crystal structure of a superoxide dismutase from the hyperthermophilic archaeon *Sulfolobus acidocaldarius* at 2.2 angstrom resolution. *J. Mol. Biol.* **285**, 689-702 (1999).

84. Claiborne, A., Mallett, T.C., Yeh, J.I., Luba, J. & Parsonage, D. Structural, redox, and mechanistic parameters for cysteine-sulfenic acid function in catalysis and regulation. *Advances In Protein Chemistry, Vol 58* **58**, 215-276 (2001).

85. Wood, Z.A., Poole, L.B. & Karplus, P.A. Structure of intact AhpF reveals a mirrored thioredoxin-like active site and implies large domain rotations during catalysis. *Biochemistry* **40**, 3900-3911 (2001).

86. Wood, Z.A., Schroder, E., Harris, J.R. & Poole, L.B. Structure, mechanism and regulation of peroxiredoxins. *Trends Biochem. Sci.* **28**, 32-40 (2003).

87. Forest, K.T., Satyshur, K.A., Worzalla, G.A., Hansen, J.K. & Herdendorf, T.J. The pilus-retraction protein the biological assembly PilT: ultrastructure of the biological assembly. *Acta Crystallogr Sect D-Biol Crystallogr* **60**, 978-982 (2004).

88. Osmani, A.H., May, G.S. & Osmani, S.A. The extremely conserved pyroA gene of *Aspergillus nidulans* is required for pyridoxine synthesis and is required indirectly for resistance to photosensitizers. *J. Biol. Chem.* **274**, 23565-23569 (1999).

89. Ram, R.J. et al. Community proteomics of a natural microbial biofilm. *Science* **308**, 1915-1920 (2005).

90. Davies, M.J. The oxidative environment and protein damage. *BBA-Proteins Proteomics* **1703**, 93-109 (2005).

91. Evguenieva-Hackenberg, E., Walter, P., Hochleitner, E., Lottspeich, F. & Klug, G. An exosome-like complex in *Sulfolobus solfataricus*. *EMBO Rep.* **4**, 889-893 (2003).

92. Price, G.D. & Badger, M.R. Evidence for the role of carboxysomes in the cyanobacterial CO_2-concentrating mechanism. *Can J Bot-Rev Can Bot* **69**, 963-973 (1991).

93. Orus, M.I., Rodriguez-Buey, M.L., Marco, E. & Fernandez-Valiente, E. Changes in carboxysome structure and grouping and in photosynthetic affinity for inorganic carbon in *Anabaena* strain PCC 7119 (cyanophyta) in response to modification of CO_2 and Na^+ supply. *Plant Cell Physiol.* **42**, 46-53 (2001).

94. Shively, J.M., van Keulen, G. & Meijer, W.G. Something from almost nothing: carbon dioxide fixation in chemoautotrophs. *Annu. Rev. Microbiol.* **52**, 191-230 (1998).

95. Wanson, J.C. & Drochman.P Rabbit skeletal muscle glycogen - a morphological and biochemical study of glycogen beta-particles isolated by precipitation-centrifugation method. *J. Cell Biol.* **38**, 130-150 (1968).

96. Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quarterly reviews of biophysics* **28**, 171-193 (1995).

97. Murakami, K.S., Masuda, S. & Darst, S.A. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. *Science* **296**, 1280-1284 (2002).

98. Lorentzen, E., Dziembowski, A., Lindner, D., Seraphin, B. & Conti, E. RNA channelling by the archaeal exosome. *EMBO Rep* (2007).

99. Yamashita, M.M., Almassy, R.J., Janson, C.A., Cascio, D. & Eisenberg, D. Refined atomic model of glutamine synthetase at 3.5 A resolution. *The Journal of biological chemistry* **264**, 17681-17690 (1989).

100. Mizohata, E. et al. Crystal structure of an archaeal peroxiredoxin from the aerobic hyperthermophilic crenarchaeon Aeropyrum pernix K1. *Journal of molecular biology* **354**, 317-329 (2005).

101. Brandstetter, H., Kim, J.S., Groll, M. & Huber, R. Crystal structure of the tricorn protease reveals a protein disassembly line. *Nature* **414**, 466-470 (2001).

102. Kim, K.K., Kim, R. & Kim, S.H. Crystal structure of a small heat-shock protein. *Nature* **394**, 595-599 (1998).

103. Britton, K.L. et al. Structure determination of the glutamate dehydrogenase from the hyperthermophile Thermococcus litoralis and its comparison with that from Pyrococcus furiosus. *Journal of molecular biology* **293**, 1121-1132 (1999).

104. Connelly, J.C., de Leau, E.S., Okely, E.A. & Leach, D.R. Overexpression, purification, and characterization of the SbcCD protein from Escherichia coli. *The Journal of biological chemistry* **272**, 19819-19826 (1997).

105. Connelly, J.C., Kirkham, L.A. & Leach, D.R. The SbcCD nuclease of Escherichia coli is a structural maintenance of chromosomes (SMC) family protein that cleaves hairpin DNA. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 7969-7974 (1998).

106. Laksanalamai, P., Jiemjit, A., Bu, Z., Maeder, D.L. & Robb, F.T. Multi-subunit assembly of the Pyrococcus furiosus small heat shock protein is essential for cellular protection at high temperature. *Extremophiles* **7**, 79-83 (2003).

107. Houry, W.A. Chaperone-assisted protein folding in the cell cytoplasm. *Current protein & peptide science* **2**, 227-244 (2001).

108. Soderberg, T. & Alver, R.C. Transaldolase of Methanocaldococcus jannaschii. *Archaea (Vancouver, B.C* **1**, 255-262 (2004).

109. Jia, J. et al. Crystal structure of transaldolase B from Escherichia coli suggests a circular permutation of the alpha/beta barrel within the class I aldolase family. *Structure* **4**, 715-724 (1996).

110. Tambasco-Studart, M. et al. Vitamin B6 biosynthesis in higher plants. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13687-13692 (2005).

111. Strohmeier, M. et al. Structure of a bacterial pyridoxal 5'-phosphate synthase complex. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 19284-19289 (2006).

112. Raschle, T., Amrhein, N. & Fitzpatrick, T.B. On the two components of pyridoxal 5'-phosphate synthase from Bacillus subtilis. *The Journal of biological chemistry* **280**, 32291-32300 (2005).

113. Weber, K. New structural model of E. coli aspartate transcarbamylase and the amino-acid sequence of the regulatory polypeptide chain. *Nature* **218**, 1116-1119 (1968).

114. Honzatko, R.B. et al. Crystal and molecular structures of native and CTP-liganded aspartate carbamoyltransferase from Escherichia coli. *Journal of molecular biology* **160**, 219-263 (1982).

115. Gerhart, J.C. & Schachman, H.K. Distinct subunits for the regulation and catalytic activity of aspartate transcarbamylase. *Biochemistry* **4**, 1054-1062 (1965).

116. De Vos, D. et al. Crystal structure of T state aspartate carbamoyltransferase of the hyperthermophilic archaeon Sulfolobus acidocaldarius. *Journal of molecular biology* **339**, 887-900 (2004).

117. Yeh, W.K. & Ornston, L.N. Origins of metabolic diversity: substitution of homologous sequences into genes for enzymes with different catalytic activities. *Proceedings of the National Academy of Sciences of the United States of America* **77**, 5365-5369 (1980).

118. Goldman, A., Ollis, D., Ngai, K.L. & Steitz, T.A. Crystal structure of muconate lactonizing enzyme at 6.5 A resolution. *Journal of molecular biology* **182**, 353-355 (1985).

119. Anand, R. et al. A model for the Bacillus subtilis formylglycinamide ribonucleotide amidotransferase multiprotein complex. *Biochemistry* **43**, 10343-10352 (2004).

120. Shimojima, M. & Benning, C. Native uridine 5'-diphosphate-sulfoquinovose synthase, SQD1, from spinach purifies as a 250-kDa complex. *Archives of biochemistry and biophysics* **413**, 123-130 (2003).

121. Mulichak, A.M., Theisen, M.J., Essigmann, B., Benning, C. & Garavito, R.M. Crystal structure of SQD1, an enzyme involved in the biosynthesis of the plant sulfolipid headgroup donor UDP-sulfoquinovose. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 13097-13102 (1999).

122. Krungkrai, S.R., Prapunwattana, P., Horii, T. & Krungkrai, J. Orotate phosphoribosyltransferase and orotidine 5'-monophosphate decarboxylase exist as multienzyme complex in human malaria parasite Plasmodium falciparum. *Biochemical and biophysical research communications* **318**, 1012-1018 (2004).

123. Nalezkova, M., de Groot, A., Graf, M., Gans, P. & Blanchard, L. Overexpression and purification of Pyrococcus abyssi phosphopantetheine adenylyltransferase from an optimized synthetic gene for NMR studies. *Protein expression and purification* **39**, 296-306 (2005).

124. Martin, D.P. & Drueckhammer, D.G. Separate enzymes catalyze the final two steps of coenzyme A biosynthesis in Brevibacterium ammoniagenes: purification of pantetheine phosphate adenylyltransferase. *Biochemical and biophysical research communications* **192**, 1155-1161 (1993).

125. Takahashi, H. et al. Structure and implications for the thermal stability of phosphopantetheine adenylyltransferase from Thermus thermophilus. *Acta crystallographica* **60**, 97-104 (2004).

126. Gibson, R.P., Turkenburg, J.P., Charnock, S.J., Lloyd, R. & Davies, G.J. Insights into trehalose synthesis provided by the structure of the retaining glucosyltransferase OtsA. *Chemistry & biology* **9**, 1337-1346 (2002).

127. Bell, W. et al. Composition and functional analysis of the Saccharomyces cerevisiae trehalose synthase complex. *The Journal of biological chemistry* **273**, 33311-33319 (1998).

128. Jelakovic, S., Kopriva, S., Suss, K.H. & Schulz, G.E. Structure and catalytic mechanism of the cytosolic D-ribulose-5-phosphate 3-epimerase from rice. *Journal of molecular biology* **326**, 127-135 (2003).

129. Caruthers, J. et al. Structure of a ribulose 5-phosphate 3-epimerase from Plasmodium falciparum. *Proteins* **62**, 338-342 (2006).

130. Kopp, J., Kopriva, S., Suss, K.H. & Schulz, G.E. Structure and mechanism of the amphibolic enzyme D-ribulose-5-phosphate 3-epimerase from potato chloroplasts. *Journal of molecular biology* **287**, 761-771 (1999).

131. Wise, E.L., Akana, J., Gerlt, J.A. & Rayment, I. Structure of D-ribulose 5-phosphate 3-epimerase from Synechocystis to 1.6 A resolution. *Acta crystallographica* **60**, 1687-1690 (2004).

132. MacKenzie, R.E. & Rabinowitz, J.C. Cation-dependent reassociation of subunits of N10-formyltetrahydrofolate synthetase from Clostridium acidi-urici and Clostridium cylindrosporum. *The Journal of biological chemistry* **246**, 3731-3736 (1971).

133. Himes, R.H. & Rabinowitz, J.C. Formyltetrahydrofolate synthetase. III. Studies on the mechanism of the reaction. *The Journal of biological chemistry* **237**, 2915-2925 (1962).

134. Edgren, T. & Nordlund, S. The fixABCX genes in Rhodospirillum rubrum encode a putative membrane complex participating in electron transfer to nitrogenase. *Journal of bacteriology* **186**, 2052-2060 (2004).

135. Eichler, K., Buchet, A., Bourgis, F., Kleber, H.P. & Mandrandberthelot, M.A. The fix Escherichia coli region contains four genes related to carnitine metabolism. *J. Basic Microbiol.* **35**, 217-227 (1995).

136. Arigoni, F., Kaminski, P.A., Hennecke, H. & Elmerich, C. Nucleotide sequence of the fixABC region of Azorhizobium caulinodans ORS571: similarity of the fixB product with eukaryotic flavoproteins, characterization of fixX, and identification of nifW. *Mol Gen Genet* **225**, 514-520 (1991).

137. Roeben, A. et al. Crystal structure of an archaeal actin homolog. *Journal of molecular biology* **358**, 145-156 (2006).

138. Walter, P. et al. Characterization of native and reconstituted exosome complexes from the hyperthermophilic archaeon Sulfolobus solfataricus. *Mol Microbiol* **62**, 1076-1089 (2006).

139. Draghici, S. Data analysis tools for DNA microarrays. (Chapman & Hall/CRC, Florida; 2003).

140. Sun, N. et al. Protein complexes of *Thermoplasma acidophilum*: a proteomics approach. *Submitted*.

141. Ernst, H.A. et al. Structure of the Sulfolobus solfataricus alpha-glucosidase: implications for domain conservation and substrate recognition in GH31. *Journal of molecular biology* **358**, 1106-1124 (2006).

142. Kim, S. & Lee, S.B. Identification and characterization of Sulfolobus solfataricus D-gluconate dehydratase: a key enzyme in the non-phosphorylated Entner-Doudoroff pathway. *The Biochemical journal* **387**, 271-280 (2005).

143. Miller, R.E. & Stadtman, E.R. Glutamate synthase from Escherichia coli. An iron-sulfide flavoprotein. *The Journal of biological chemistry* **247**, 7407-7419 (1972).

144. Petoukhov, M.V. et al. Quaternary structure of Azospirillum brasilense NADPH-dependent glutamate synthase in solution as revealed by synchrotron radiation x-ray scattering. *The Journal of biological chemistry* **278**, 29933-29939 (2003).

145. Gibson, L.C., Willows, R.D., Kannangara, C.G., von Wettstein, D. & Hunter, C.N. Magnesium-protoporphyrin chelatase of Rhodobacter sphaeroides: reconstitution of activity by combining the products of the bchH, -I, and -D genes expressed in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 1941-1944 (1995).

146. Jensen, P.E., Gibson, L.C. & Hunter, C.N. Determinants of catalytic activity with the use of purified I, D and H subunits of the magnesium protoporphyrin IX chelatase from Synechocystis PCC6803. *The Biochemical journal* **334 (Pt 2)**, 335-344 (1998).

147. Vanoni, M.A. & Curti, B. Glutamate synthase: a complex iron-sulfur flavoprotein. *Cell Mol Life Sci* **55**, 617-638 (1999).

148. Smith, P.F., Langworthy, T.A. & Smith, M.R. Polypeptide nature of growth requirement in yeast extract for Thermoplasma acidophilum. *J Bacteriol* **124**, 884-892 (1975).

149. Kawashima, T., Yokoyama, K., Higuchi, S. & Suzuki, M. Identification of proteins present in the archaeon *Thermoplasma volcanium* cultured in aerobic or anaerobic conditions. *Proc. Japan Acad.* **81**, 204-219 (2005).

150. Berg, J., Tymoczko, J. & Stryer, L. Biochemistry, Edn. 5th. (W. H. Freeman and Co., New York; 2002).

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

152. Budgen, N. & Danson, M.J. Metabolism of glucose via a modified Entner-Doudoroff pathway in the thermoacidophilic archaebacterium *Thermoplasma acidophilum*. *FEBS Lett.* **196**, 207-210 (1986).

153. Reher, M. & Schoenheit, P. Glyceraldehyde dehydrogenases from the thermoacidophilic euryarchaeota *Picrophilus torridus* and *Thermoplasma acidophilum*, key enzymes of the non-phosphorylative Entner-Doudoroff pathway, constitute a novel enzyme family within the aldehyde dehydrogenase superfamily. *FEBS Lett* **580**, 1198-1204 (2006).

154. Ronimus, R.S. & Morgan, H.W. Distribution and phylogenies of enzymes of the Embden-Meyerhof-Parnas pathway from archaea and hyperthermophilic bacteria support a gluconeogenic origin of metabolism. *Archaea (Vancouver, B.C* **1**, 199-221 (2003).

155. Xie, G., Keyhani, N.O., Bonner, C.A. & Jensen, R.A. Ancient origin of the tryptophan operon and the dynamics of evolutionary change. *Microbiol Mol Biol Rev* **67**, 303-342, table of contents (2003).

156. Futterer, O. et al. Genome sequence of Picrophilus torridus and its implications for life around pH 0. *Proc Natl Acad Sci U S A* **101**, 9091-9096 (2004).

157. Searcy, D.G. & Hixon, W.G. Cytoskeletal origins in sulfur-metabolizing archaebacteria. *Bio Systems* **25**, 1-11 (1991).

158. Hixon, W.G. & Searcy, D.G. Cytoskeleton in the archaebacterium *Thermoplasma acidophilum* - viscosity increase in soluble extracts. *Bio Systems* **29**, 151-160 (1993).

159. Bonaldi, T. et al. Combined use of RNAi and quantitative proteomics to study gene function in Drosophila. *Molecular cell* **31**, 762-772 (2008).

160. de Godoy, L.M. et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455**, 1251-1254 (2008).

161. de Groot, M.J. et al. Quantitative proteomics and transcriptomics of anaerobic and aerobic yeast cultures reveals post-transcriptional regulation of key cellular processes. *Microbiology (Reading, England)* **153**, 3864-3878 (2007).

162. Rocha, E.R., Tzianabos, A.O. & Smith, C.J. Thioredoxin reductase is essential for thiol/disulfide redox control and oxidative stress survival of the anaerobe Bacteroides fragilis. *Journal of bacteriology* **189**, 8015-8023 (2007).

163. Gu, S.Y., Yan, X.X. & Liang, D.C. Crystal structure of Tflp: a ferredoxin-like metallo-beta-lactamase superfamily protein from Thermoanaerobacter tengcongensis. *Proteins* **72**, 531-536 (2008).

164. Kerscher, L. & Oesterhelt, D. Pyruvate:ferredoxin oxidoreductase-new findings on an ancient enzyme. *Trends Biochem.Sci.* **7**, 371-374 (1982).

165. Galperin, M.Y. & Koonin, E.V. Functional genomics and enzyme evolution. Homologous and analogous enzymes encoded in microbial genomes. *Genetica* **106**, 159-170 (1999).

166. Daran-Lapujade, P. et al. Role of transcriptional regulation in controlling fluxes in central carbon metabolism of Saccharomyces cerevisiae. A chemostat culture study. *The Journal of biological chemistry* **279**, 9125-9138 (2004).

167. Camarasa, C., Grivet, J.P. & Dequin, S. Investigation by 13C-NMR and tricarboxylic acid (TCA) deletion mutant analysis of pathways for succinate formation in Saccharomyces cerevisiae during anaerobic fermentation. *Microbiology (Reading, England)* **149**, 2669-2678 (2003).

168. Madigan, M., martinko, J. & Parker, J. Biology of Microorganisms. (Sourthern Illinois University Carbondale, 2000).

169. Dopson, M., Baker-Austin, C. & Bond, P. Towards determining details of anaerobic growth coupled to ferric iron reduction by the acidophilic archaeon 'Ferroplasma acidarmanus' Fer1. *Extremophiles* **11**, 159-168 (2007).

170. Hugler, M., Huber, H., Stetter, K.O. & Fuchs, G. Autotrophic CO_2 fixation pathways in archaea (Crenarchaeota). *Archives of microbiology* **179**, 160-173 (2003).

171. Zhu, Z., Boobis, A.R. & Edwards, R.J. Identification of estrogen-responsive proteins in MCF-7 human breast cancer cells using label-free quantitative proteomics. *Proteomics* **8**, 1987-2005 (2008).

172. Dresios, J., Panopoulos, P. & Synetos, D. Eukaryotic ribosomal proteins lacking a eubacterial counterpart: important players in ribosomal function. *Molecular microbiology* **59**, 1651-1663 (2006).

173. Garcia-Marcos, A. et al. In vivo assembling of bacterial ribosomal protein L11 into yeast ribosomes makes the particles sensitive to the prokaryotic specific antibiotic thiostrepton. *Nucleic acids research* **35**, 7109-7117 (2007).

174. Chester, K.A. et al. Identification of a human ribosomal protein mRNA with increased expression in colorectal tumours. *Biochimica et biophysica acta* **1009**, 297-300 (1989).

8 Acknowledgements

This work was performed in the Department of Molecular Structural Biology, Max-Planck-Institute of Biochemistry, München.

I would like to express profound gratitude to Professor Dr. Wolfgang Baumeister, head of the Department of Molecular Structural Biology, for giving me the opportunity to pursue my Ph.D. thesis in his department, and for his generous support and supervision.

I would like to express my deepest gratitude to my supervisor Dr. István Nagy not only for his supervision and encouragement, but also for his critical corrections of all the manuscripts and my Ph.D. thesis. His preciseness, optimism, diligence, modesty and inspiring attitude towards science have set a good example and will continue to have a good influence on me in the future.

I am highly thankful to Dr. Stephan Nickell for his invaluable support, suggestion, encouragement and continuous guidance enabled me to complete my work successfully.

I am grateful to all my colleagues for creating a pleasant working atmosphere and giving me helping hands whenever it was necessary. Thanks to Roland Knispel, Marius Boicu, Dr. Christine Kofler, Dr. Eri Sakata, Dr. Jasuo Mitani, Dr. Julio Ortiz, Oana Mihalache, Stefan Bohn, Manuela Gruska, Ulrike Maurer and Dr. Jürgen Peters who helped me with the experiments, EM analysis, and writing this thesis. Thomas Hrabe, Florian Beck, Inga Wolf, Thomas Haller and Andreas Korinek are thanked for the continuous support in computed evaluation of the results. I would like to express my deep gratitude to Dr. István Nagy, Professor Dr. Wolfgang Baumeister, Roland Knispel, Dr. Stephan Nickell, Dr. Jürgen Peters, Dr. Valerio Matias, Dr. Andrew Leis and Dr. Dennis Thomas for critical reading and correction of my thesis among their busiest time of work. I am also highly thankful to our secretaries, Birgit Book and Sabina von Polenz, for their always patient and kind help and supports with all the official and personal things.

I would like to take the opportunity to thank the Department of Proteomics and Signaltransduction, Max-Planck-Institute of Biochemistry, especially Prof. Dr. Matthias Mann and Dr. Cuiping Pan for their generous supports, discussion and collaboration;

Dr. Jürgen Cox, Dr. Johannes Graumann, Peter Bandilla, Dr. Jesper Olsen, Mario Oroshi, Dr. Chanchal Kumar, Dr. Nagarjuna Nagaraj, Dr. Aiping Lu, Dr. Chuong Nguyen, Shubin Ren, Dr. Yong Zhang and Alexandre Zougman for helping me with the MS instrument, sharing their knowledge in bioinformatics and statistics.

I would like to thank Prof. Dr. Tomohiro Tamura and Dr. Noriko Tamura, and all the lab members of the National Institute of Advanced Industrial Science and Technology, Japan, not only for their great help and supports for my work, but also for their friendships during my stay in Japan. The collaboration with them was a sheer joy.

I would like also to thank the Department of Membrane Biochemistry, Max-Planck-Institute of Biochemistry, for MALDI-MS technical support, particularly Dr. Frank Siedler and Beatrix Scheffer for introduction and hand-by-hand teaching of mass spectrometry; Dr. Birgit Bisle, Sigrid Bauer, Dr. Andreas Tebbe, Dr. Kosta Konstantinidis and Barbara Fischer for their nice help and advices by the MS problems; Katarina Furtwängler and Rita Schwaiger for their kind help and suggestions for assessing the quality of the RNA samples.

Thanks to Florian Fröhlich from the Organelle Architecture and Dynamics group, Max-Planck-Institute of Biochemistry, for his kind help with the SILAC-labelled *T. acidophilum* experiment.

Last but not least, to my beloved family and boyfriend Chao. Thanks for your deep understanding and constant supports. Without these, I would not have been achieved so much.

9 Appendix

Table S1: Bioinformatic analysis of the 187 putative macromolecular complexes forming proteins. The protein IDs, annotation of the complexes, Mw of the complexes according to references, Mw of the complexes according to the MSC experiment and protein annotations are indicated. The colours and abbreviations used in the table are:

	Subunit(s) of complexes
	Subunits of one specific complex
	Proteins that can be adhered to large complexes, and/but they can be present in the cytosol in free form
	Complexes that can dissociate to subunits, or proteins that can form random aggregates
	Hypothetical proteins
S	Subunits of one specific complex identified with one peptide but was found in several fractions same as the other complex subunits
P	Proteins identified in the pooled samples

	KEGG pathways
PATH: tac00010	Glycolysis / Gluconeogenesis
PATH: tac00020	Citrate cycle (TCA cycle)
PATH: tac00030	Pentose phosphate pathway
PATH: tac00040	Pentose and glucuronate interconversions
PATH: tac00052	Galactose metabolism
PATH: tac00071	Fatty acid metabolism
PATH: tac00190	Oxidative phosphorylation
PATH: tac00220	Urea cycle and metabolism of amino groups
PATH: tac00230	Purine metabolism
PATH: tac00240	Pyrimidine metabolism
PATH: tac00251	Glutamate metabolism
PATH: tac00252	Alanine and aspartate metabolism
PATH: tac00260	Glycine, serine and threonine metabolism
PATH: tac00271	Methionine metabolism
PATH: tac00280	Valine, leucine and isoleucine degradation
PATH: tac00290	Valine, leucine and isoleucine biosynthesis
PATH: tac00330	Arginine and proline metabolism
PATH: tac00450	Selenoamino acid metabolism
PATH: tac00500	Starch and sucrose metabolism
PATH: tac00550	Peptidoglycan biosynthesis
PATH: tac00620	Pyruvate metabolism
PATH: tac00630	Glyoxylate and dicarboxylate metabolism
PATH: tac00640	Propanoate metabolism
PATH: tac00650	Butanoate metabolism
PATH: tac00660	C5-Branched dibasic acid metabolism
PATH: tac00670	One carbon pool by folate
PATH: tac00680	Methane metabolism
PATH: tac00710	Carbon fixation in photosynthetic organisms
PATH: tac00720	Reductive carboxylate cycle (CO_2 fixation)
PATH: tac00750	Vitamin B6 metabolism
PATH: tac00770	Pantothenate and CoA biosynthesis
PATH: tac00860	Porphyrin and chlorophyll metabolism
PATH: tac03010	Ribosome
PATH: tac03020	RNA polymerase

	Protein ID	***Complex annotation***	***Mw of the complex according to references (kDa)***	***Mw of the complex according to the MSC experiment (kDa)***	***Protein annotation***
Cell cycle and DNA processing					
Ribosome biogenesis					
PATH: tac03010	Ta0050	ribosome	2400	>1000	30S ribosomal protein S19e
	Ta0054				50S ribosomal protein L31
	Ta0091				30S ribosomal protein S12
	Ta0092				30S ribosomal protein S7
	Ta0323				30S ribosomal protein S6
	Ta0358				50S ribosomal protein L12
	Ta0359				50S ribosomal protein L10
	Ta0360				50S ribosomal protein L1
	Ta0361				50S ribosomal protein L11
	Ta0432				30S ribosomal protein S9
	Ta0433				50S ribosomal protein L13
	Ta0434				probable 50S ribosomal protein L18
	Ta0445				30S ribosomal protein S10
	Ta1031				30S ribosomal protein S11
	Ta1032				30S ribosomal protein S4
	Ta1033				30S ribosomal protein S13
	Ta1040				30S ribosomal protein S8
	Ta1057				50S ribosomal protein L10
	Ta1092				30S ribosomal protein S24e
S	Ta1093				30S ribosomal protein S27ae
	Ta1114				50S ribosomal protein L24
	Ta1115				30S ribosomal protein S28e
	Ta1116				50S ribosomal protein L7Ae
	Ta1131				30S ribosomal protein S15
	Ta1167				30S ribosomal protein S3a
	Ta1190				30S ribosomal protein S2
	Ta1204				30S ribosomal protein S27
	Ta1249				50S ribosomal protein L15
	Ta1250				50S ribosomal protein L30
	Ta1251				30S ribosomal protein S5P
	Ta1252				50S ribosomal protein L18
	Ta1253				50S ribosomal protein L19e
	Ta1254				50S ribosomal protein L32e
	Ta1255				50S ribosomal protein L6
	Ta1256				30S ribosomal protein S8
	Ta1257				30S ribosomal protein S14
	Ta1258				50S ribosomal protein L5
	Ta1259				30S ribosomal protein S4
	Ta1261				50S ribosomal protein L14
S	Ta1262				30S ribosomal protein S17
	Ta1265				30S ribosomal protein S3
	Ta1266				50S ribosomal protein L22
	Ta1267				30S ribosomal protein S19P
	Ta1268				50S ribosomal protein L2
	Ta1269				50S ribosomal protein L23
	Ta1270				50S ribosomal protein L4

	Ta1271				50S ribosomal protein L3
	Ta1286				50S ribosomal protein L15
	Ta1295				50S ribosomal protein L37
	Ta1297				50S ribosomal protein L21
	Ta1124				ribosome biogenesis protein
Translation					
	Ta0055				translation initiation factor IF-6
	Ta0322				translation initiation factor IF-2 gamma subunit
	Ta1112				translation initiation factor IF-2
P	Ta1203				translation initiation factor IF-2 alpha subunit
	Ta1212				translation initiation factor IF-2B subunit alpha
	Ta0444				elongation factor Tu
	Ta0446				elongation factor EF-2
DNA/RNA BINDING PROTEINS					
Aminoacyl-tRNA-synthetases					
	Ta0551				phenylalanyl-tRNA synthetase beta subunit
	Ta0609				asparaginyl-tRNA synthetase
	Ta0849				alanyl-tRNA synthetase
tRNA modification					
	Ta1493				queuine tRNA-ribosyltransferase related protein
Transcription					
PATH: tac00230, tac00240, tac03020	Ta0390	DNA-directed RNA polymerase	500	500	DNA-directed RNA polymerase subunit beta
PATH: tac00230, tac00240, tac03020	Ta0391				DNA-directed RNA polymerase subunit alpha
PATH: tac00230, tac00240, tac03020	Ta0392				DNA-directed RNA polymerase subunit A
S	Ta0431				DNA-directed RNA polymerase subunit N
PATH: tac00230, tac00240, tac03020					
PATH: tac00230, tac00240, tac03020	Ta1030				DNA-directed RNA polymerase subunit D
PATH: tac00230, tac00240, tac03020	Ta1089				DNA-directed RNA polymerase subunit E'
PATH: tac00230, tac00240, tac03020	Ta1160				DNA-directed RNA polymerase subunit K
PATH: tac00230, tac00240	Ta1298?				hypothetical protein COG1460 DNA dept RNA polymerase subunit F
PATH: tac00230, tac00240, tac03020	Ta1416				DNA-directed RNA polymerase subunit L
	Ta0940				transcription initiation factor IIB
	Ta0945				transcription initiation factor IIB
Regulatory proteins					
	Ta0115				hypothetical protein helix_turn_helix, Arsenical Resistance Operon Repressor
	Ta0134				hypothetical protein Sugar-specific transcriptional regulator TrmB
	Ta0466				hypothetical protein PhoU Phosphate uptake regulator
	Ta0601				hypothetical protein Predicted transcriptional regulator
	Ta0608				transcription regulator related protein

	Ta0685				hypothetical protein COG1378 Predicted transcriptional regulators TrmB
	Ta0890				hypothetical protein Predicted transcriptional regulators TrmB
	Ta1064				hypothetical protein Transcriptional regulator of a riboflavin/FAD biosynthetic operon
Cell cyle and DNA processing					
	Ta0063				DNA topoisomerase I
	Ta0093				DNA-binding protein HTa
	Ta1104				DNA repair and recombination protein RadA
	Ta1054				DNA gyrase subunit A
	Ta0157	SbcC/SbcD complex	1210	>1000	chromosome segregation protein, SbcC
S	Ta0158				hypothetical protein: SbcD
	Ta0159				hypothetical protein: COG0433, predicted ATPase
	Ta0787				chromosome segregation protein related protein
	Ta0912				chromatin protein
	Ta0052				hypothetical protein: DNA-binding protein
	Ta0069				hypothetical protein: Archaeal holliday junction resolvase (hjc).
	Ta1149				hypothetical protein: Single-stranded DNA-binding replication protein A
Predicted DNA/RNA binding proteins					
	Ta0333				hypothetical protein: YSH1, Predicted exonuclease of the beta-lactamase fold involved in RNA processing
	Ta0613				predicted metal-dependent RNase
	Ta1418				predicted RNA-binding protein homologous to eukaryotic snRNP
	Ta1492				probable RNase L inhibitor
	Ta1457				hypothetical protein: nucleic-acid-binding protein containing a Zn-ribbon
Exosome					
	Ta1292	exosome	300	500	exosome complex RNA-binding protein 1
	Ta1293				exosome complex exonuclease 1
	Ta1294				exosome complex exonuclease 2
	Ta0929				hypothetical protein: Csl4 homolog
	Ta0097				DNA primase DnaG homolog
METABOLISM					
Lipid, fatty acid and isoprenoid metabolism					
PATH: tac00071	Ta0582				acetyl-CoA acetyltransferase
	Ta1073	uridine 5'-diphosphate-sulfoquinovose synthase homolog	250	>1000	probable sulfolipid biosynthesis protein
CO2 fixation, AA metabolism					
	Ta0765				aspartate transaminase related protein
PATH: tac00030, tac00040, tac00710	Ta1315	ribulose-phosphate 3-epimerase	50 (dimer); 150 (hexamer)	>1000	ribulose-phosphate 3-epimerase
AA metabolism					
PATH: tac00271, tac00450	Ta0059				S-adenosylmethionine synthetase
	Ta0756	muconate cycloisomerase related protein	320	>1000	muconate cycloisomerase related protein
PATH: tac00251, tac00330	Ta0635	glutamate dehydrogenase	270	300	probable glutamate dehydrogenase
PATH: tac00220, tac00330	Ta1330	ornithine carbamoyltransferase	420	300	ornithine carbamoyltransferase
PATH: tac00251,	Ta1498	glutamine synthetase	637	600	probable glutamine synthetase

tac00550					
Purin and pyrimidine metabolism					
	Ta1044	formylglycinamide ribonucleotide amidotransferase	153	1000	phosphoribosylformylglycinamidine synthase
PATH: tac00230	Ta1066				phosphoribosylformylglycinamidine synthase subunit II
PATH: tac00230	Ta1318				phosphoribosylformylglycinamidine synthase subunit I
PATH: tac00230	Ta1107				phosphoribosylaminoimidazole carboxylase
PATH: tac00240, tac00252	Ta0574	aspartate carbamoyltransferase	310	>1000	aspartate carbamoyltransferase regulatory subunit
PATH: tac00240, tac00252	Ta0575				aspartate carbamoyltransferase catalytic subunit
PATH: tac00240	Ta1164	orotate phosphoribosyltransferase	140	>1000	orotate phosphoribosyltransferase
Starch and sucrose metabolism					
	Ta0339				Alpha-amylase
	Ta0341				glycogen debranching enzyme isoform 1 (agl) related protein
	Ta0342				hypothetical protein: Glucoamylase and related glycosyl hydrolases
P PATH: tac00500	Ta1210	trehalose-6-phosphate synthase	200	>1000	alpha, alpha-trehalose-phosphate synthase (UDP-forming) related protein
Vitamin, cofactor and prosthetic groups					
PATH: tac00750	Ta0522	pyridoxal 5'-phosphate synthase	576	>1000	pyridoxine biosynthesis protein
PATH: tac00860	Ta0078				hypothetical protein NaMN:DMB phosphoribosyltransferase CobT
PATH: tac00860	Ta1434				hypothetical protein Cobalamin adenosyltransferase
PATH: tac00860	Ta0652				hypothetical protein Siroheme synthase (precorrin-2 oxidase/ferrochelatase domain
PATH: tac00860	Ta0659				probable precorrin-4 methylase
PATH: tac00670	Ta0010				Aminomethyltransferase (glycine cleavage system T protein)
PATH: tac00670	Ta1476	glutamate formiminotransferase	unknown	600	probable glutamate formiminotransferase
PATH: tac00630, tac00670	Ta1478	formate-tetrahydrofolate ligase	240	600	probable formate-tetrahydrofolate ligase
	Ta1412				hypothetical protein: COG2427
	Ta1413	sulfide dehydrogenase	unknown	1000	sulfide dehydrogenase related protein
Pantothenate and CoA biosynthesis					
PATH: tac00770	Ta1194	phosphopantetheine adenylyltransferase	108	>1000	phosphopantetheine adenylyltransferase
ENERGY					
TCA					
	Ta0112				aconitate hydratase
	Ta0117				probable isocitrate dehydrogenase
	Ta0258				probable fumarase
PATH: tac00020, tac00720	Ta0259	ferredoxin oxidoreductase (together with Ta0260)	300	300	2-oxoacid--ferredoxin oxidoreductase, alpha subunit
	Ta0952				malate dehydrogenase
	Ta1001				succinate dehydrogenase (membrane bound, complex formation with Ta1002, Ta1003 and Ta1004)
PATH: tac00020 , tac00640 , tac00660, tac00720	Ta1331	succinyl-CoA synthetase	500	500	succinyl-CoA synthetase alpha subunit
PATH: tac00020 ,	Ta1332				succinyl-CoA synthetase subunit beta

tac00640 , tac00660, tac00720					
ED					
PATH: tac00010, tac00020, tac00260, tac00280, tac00620	Ta1435	pyruvate dehydrogenase	9000	>1000	dihydrolipoamide dehydrogenase
PATH: tac00010, tac00020, tac00620	Ta1436				dihydrolipoamide acetyltransferase
PATH: tac00010, tac00020, tac00290, tac00620, tac00650	Ta1437				probable 3-methyl-2-oxobutanoate dehydrogenase chain E1-beta
PATH: tac00010, tac00020, tac00290, tac00620, tac00650	Ta1438				probable 3-methyl-2-oxobutanoate dehydrogenase alpha chain precursor
EMP					
PATH: tac00010, tac00052, tac00500	Ta0825	glucose kinase related protein	unknown	550	glucose kinase related protein
PATH: tac00010	Ta0882	enolase	80	300	phosphopyruvate hydratase
	Ta1103				glyceraldehyde-3-phosphate dehydrogenase
Pentose phosphate pathway					
PATH: tac00030	Ta0616	transaldolase	300	300	transaldolase
Oxidative phosphorylation (phosphate metabolism)					
PATH: tac00190	Ta0399				inorganic pyrophosphatase
Respiration					
	Ta0326	fixABCX	unknown	550	fixC protein related
S	Ta0327				hypothetical protein
	Ta0328				FIXA (related to carnitine metabolism) related protein
	Ta0329				probable electron transfer flavoprotein, alpha subunit
Respiration					
PATH: tac00630, tac00680	Ta0424	formate dehydrogenase	unknown	440	hypothetical protein: COG2427
PATH: tac00630, tac00680	Ta0425				formate dehydrogenase related protein
CELL RESCUE, DEFENSE AND VIRULENCE					
	Ta0125				alkyl hydroperoxide reductase subunit f related protein
	Ta0152	peroxiredoxin	490	600	probable peroxiredoxin
	Ta0954	peroxiredoxin	490	300	putative peroxiredoxin
PROTEIN FATE					
Protein/peptide degradation					
	Ta0612	proteasome	700	700	proteasome, beta chain
	Ta1288				proteasome subunit alpha
	Ta1490	tricorn	720	700	tricorn protease
	Ta1037				proline dipeptidase related protein
Protein folding and stability					
	Ta0840	VAT ATPase	500	500	VAT ATPase (VCP-like ATPase)
	Ta0980	thermosome	900	900	thermosome, alpha chain
	Ta1276				thermosome beta chain
	Ta0471	hsp20	200-800	600	small heat shock protein (hsp20) related protein

	Ta1011				peptidyl-prolyl cis-trans isomerase related protein
	Ta1087	DnaK	200-800	600	molecular chaperone DnaK (for the activity it needs the interaction of Ta1086 and Ta1088)
P	Ta1088	DnaJ	200-800	>1000	heat shock protein DnaJ related protein
Hypothetical proteins					
	Ta0105				hypothetical protein Zn-dependent hydrolases, including glyoxylases GloB
	Ta0182				hypothetical protein No putative conserved domains
	Ta0316				hypothetical protein PilT
	Ta0518				hypothetical protein Reovirus viral attachment protein sigma 1
	Ta0547				hypothetical protein No putative conserved domains
	Ta0561				hypothetical protein Tetraticopeptide motif
	Ta0759				hypothetical protein No putative homology
	Ta0798				hypothetical protein predicted coiled coil, no homology
	Ta0812				hypothetical protein CBS domain
PATH: tac00770	Ta0889				hypothetical protein Phosphopantothenoylcysteine synthetase/decarboxylase
	Ta1127				hypothetical protein TadD PilF pilus assembly
	Ta1159				hypothetical protein COG1860
	Ta1201				hypothetical protein (ATP-grasp superfamily General function prediction only COG2047
	Ta1217				hypothetical protein No putative conserved domains
P	Ta1307				hypothetical protein USP (Uneiversal Stress Protein) related
	Ta1352				hypothetical protein hydrolysis of nucleoside diphosphates linked to other moieties
	Ta1441				hypothetical protein ATP-grasp superfamily COG193
	Ta1488				hypothetical protein COG coiled coil, SMC, DNA polymeraseIII S. aureus

Table S2: Quantitative proteomics data of the 341 unchanged proteins, 146 proteins which were down-regulated, 117 proteins which were up-regulated anaerobically, and 421 unquantified proteins which were not satisfy the quantitation restrictions. The protein IDs, annotations, changes in protein expression level (anaerobic/aerobic), p-values of the Student's t-test, CV values of normalized intensities of the three aerobic replicates, CV values of normalized intensities of the three anaerobic replicates, absolute intensities in protein expression level (aerobic) and absolute intensities in protein expression level (anaerobic) were listed. Abbreviations: AN, detected anaerobically; AE, detected aerobically; NA, no available value.

Protein ID	***Annotation***	***Ratio (anae/ae)***	***p-value***	***CV_{ae}***	***CV_{anae}***	***Absolute intensity (aerob)***	***Absolute intensity (anaerob)***
Unchanged							
Ta0001	V-type ATP synthase subunit E	0.64	1.03E-01	42.98	44.69	2.48E+04	7.96E+03
Ta0004	V-type ATP synthase subunit A	0.58	5.06E-03	18.98	1.29	6.74E+04	4.10E+04
Ta0010m	Aminomethyltransferase (glycine cleavage system T protein)	1.18	1.85E-01	5.09	28.17	1.84E+05	1.48E+05
Ta0016	hypothetical protein	0.90	2.20E-01	42.21	34.22	1.39E+03	6.34E+02
Ta0020	hypothetical protein	1.05	2.21E-01	21.59	39.00	8.47E+03	5.68E+03
Ta0021m	hypothetical protein	0.60	8.39E-02	43.41	29.30	6.28E+03	2.04E+03
Ta0027	alcohol dehydrogenase related protein	0.51	8.35E-03	22.39	20.88	9.59E+04	6.71E+04
Ta0030	acetyl-CoA acetyltransferase	0.52	8.09E-03	21.50	23.74	5.83E+04	2.19E+04
Ta0031	hypothetical protein	0.82	3.55E-02	7.22	43.60	2.53E+04	2.71E+04
Ta0033	hypothetical protein	1.14	1.88E-01	21.60	19.25	5.76E+03	3.93E+03
Ta0034	hypothetical protein	0.53	9.89E-03	20.01	38.58	1.60E+05	8.73E+04
Ta0035	hypothetical protein	1.53	9.27E-01	35.98	34.80	2.64E+04	3.63E+04
Ta0039	hypothetical protein	1.04	1.87E-01	12.16	44.56	5.98E+03	7.07E+03
Ta0045m	Formate dehydrogenase, alpha subunit	0.51	8.27E-02	47.58	37.00	2.99E+03	2.16E+03
Ta0046	anaerobic sulfite reduction protein A related protein	0.95	2.71E-01	43.88	33.69	3.29E+02	3.01E+02
Ta0050	30S ribosomal protein S19e	1.71	5.73E-01	4.47	35.39	9.99E+04	1.16E+05
Ta0055m	translation initiation factor IF-6	0.60	2.71E-02	23.82	46.51	5.36E+04	2.06E+04
Ta0056	orotate phosphoribosyltransferase	0.60	1.10E-02	12.86	46.92	1.34E+05	5.89E+04
Ta0057m	Rad3-related DNA helicase	0.57	8.31E-03	19.95	25.30	4.02E+03	2.52E+03
Ta0063	DNA topoisomerase I	1.07	3.47E-01	45.53	12.83	2.21E+04	8.20E+03
Ta0072	cell division protein FtsZ	1.80	1.74E-01	5.61	17.18	4.44E+04	4.00E+04
Ta0078	hypothetical protein	0.68	1.19E-03	7.47	19.23	3.60E+05	1.62E+05
Ta0079	dihydropteroate synthase related protein	0.82	7.42E-03	4.35	27.41	1.21E+05	8.88E+04
Ta0080	cystathionine beta-lyase	0.56	3.02E-02	30.43	33.27	9.66E+04	3.84E+04
Ta0081	thymidylate kinase	0.82	1.55E-01	44.39	6.09	8.31E+03	3.75E+03
Ta0082	probable phosphoribosylglycinamide formyltransferase	0.88	1.37E-01	26.89	46.95	3.30E+04	1.14E+04
Ta0092	30S ribosomal protein S7	1.67	6.23E-01	31.43	20.57	1.62E+05	1.80E+05
Ta0096	hypothetical protein	1.78	5.49E-01	29.72	34.52	2.94E+04	1.56E+04
Ta0102	isopentenyl pyrophosphate isomerase	0.55	1.71E-03	4.22	37.84	9.21E+04	3.74E+04
Ta0103	gamma-glutamyl kinase related protein	0.77	7.77E-02	35.21	6.86	4.72E+03	2.51E+03
Ta0104m	Acetyltransferase	0.74	1.99E-02	10.24	42.19	3.15E+04	1.48E+04
Ta0105	hypothetical protein	1.04	3.30E-01	33.51	47.23	5.81E+04	6.81E+04
Ta0110	hypothetical protein	0.95	1.13E-01	9.87	45.85	2.37E+04	3.58E+04
Ta0113	threonine dehydratase	1.68	6.68E-01	7.72	40.56	2.48E+05	2.15E+05
Ta0116	3-dehydroquinate dehydratase	0.88	2.53E-02	20.05	7.51	1.76E+04	9.98E+03
Ta0119m	ribose-phosphate pyrophosphokinase	0.75	1.76E-02	19.23	22.03	1.37E+05	8.77E+04
Ta0122	xylulose kinase related protein	1.03	1.77E-01	26.21	27.66	1.33E+05	1.37E+05
Ta0134	hypothetical protein	1.07	1.47E-01	26.03	11.49	6.37E+04	1.16E+05
Ta0138	neopullulanase related protein	0.75	7.89E-02	36.71	7.59	1.44E+03	6.14E+03
Ta0150	hypothetical protein	1.12	4.24E-01	34.83	44.91	8.16E+03	4.52E+03
Ta0157	chromosome segregation protein	1.22	6.09E-01	47.68	38.83	2.63E+04	7.84E+03
Ta0175	phosphoglycolate phosphatase	1.27	1.65E-01	15.04	1.99	2.44E+04	1.88E+04
Ta0177	hypothetical protein	1.11	3.63E-01	40.38	22.85	4.69E+04	2.18E+04

Ta0179	hypothetical protein	1.32	6.43E-01	13.87	41.25	9.35E+03	1.49E+04
Ta0201	hypothetical protein	0.62	1.10E-01	45.90	46.16	3.41E+03	1.19E+03
Ta0206	glycyl-tRNA synthetase	0.79	1.06E-01	33.30	38.99	7.78E+04	7.61E+04
Ta0207	hypothetical protein	1.57	8.69E-01	19.25	43.08	4.58E+04	4.64E+04
Ta0215	hypothetical protein	1.00	1.86E-01	29.21	30.25	9.27E+03	8.59E+03
Ta0217	hypothetical protein	1.18	5.38E-01	48.68	26.84	9.83E+03	1.10E+04
Ta0218	hypothetical protein	1.86	3.04E-01	20.23	24.35	5.39E+03	1.62E+04
Ta0231	hypothetical protein	1.62	7.84E-01	22.61	42.42	7.17E+03	1.18E+04
Ta0233	gliding motility protein related	1.18	1.35E-01	17.91	9.32	5.13E+03	5.72E+03
Ta0234	hypothetical protein	1.02	1.14E-01	8.15	38.02	4.58E+04	5.53E+04
Ta0239m	urocanate hydratase	1.30	5.54E-01	32.24	11.81	2.10E+05	7.49E+04
Ta0248	hypothetical protein	1.06	1.52E-01	25.32	18.40	1.55E+03	1.06E+03
Ta0249	probable N-acylamino acid racemase	1.47	9.47E-01	16.77	39.61	2.41E+04	3.60E+04
Ta0251	3-ketoacyl-(acyl-carrier-protein) reductase	1.98	4.51E-01	19.65	48.69	1.50E+05	3.74E+05
Ta0262	hypothetical protein	1.54	9.40E-01	46.43	44.88	1.19E+04	1.21E+04
Ta0277	hypothetical protein	1.68	6.70E-01	33.03	29.55	4.36E+03	6.36E+03
Ta0278	hypothetical protein	1.11	1.52E-01	16.70	25.86	5.89E+05	6.35E+05
Ta0279	hypothetical protein	1.01	2.68E-01	40.47	24.91	1.35E+04	1.36E+04
Ta0281	arginyl-tRNA synthetase	1.12	2.54E-01	21.53	32.85	2.98E+05	1.28E+05
Ta0284	shikimate 5-dehydrogenase	1.28	5.67E-01	33.90	23.18	1.55E+03	2.46E+03
Ta0285	3-dehydroquinate synthase	0.63	6.49E-02	39.50	1.09	1.79E+04	1.12E+04
Ta0286m	hypothetical protein	0.53	2.76E-02	31.07	29.92	5.44E+04	4.16E+04
Ta0287m	Predicted oxidoreductase	1.65	6.55E-01	38.20	8.34	9.98E+04	1.47E+05
Ta0289	hypothetical protein	1.21	3.53E-01	15.18	32.91	6.97E+04	8.51E+04
Ta0293	hypothetical protein	1.45	8.55E-01	20.01	18.46	4.32E+04	4.05E+04
Ta0302	sua5 related protein	1.08	2.86E-01	29.04	36.18	2.57E+04	1.61E+04
Ta0304	hypothetical protein	1.21	4.75E-01	36.53	23.27	2.58E+04	4.54E+04
Ta0310	uroporphyrinogen decarboxylase	1.34	6.96E-01	25.10	38.22	1.87E+04	1.96E+04
Ta0311	ferrochelatase	0.56	9.87E-03	18.25	39.79	2.68E+04	1.10E+04
Ta0313	triosephosphate isomerase	1.53	9.32E-01	24.15	42.48	3.80E+04	4.49E+04
Ta0315	asparagine synthetase related protein	0.78	3.95E-02	21.94	31.17	5.00E+03	2.44E+03
Ta0316	hypothetical protein	0.97	1.26E-01	22.81	33.65	3.76E+04	4.13E+04
Ta0320m	Predicted transcription regulator	1.66	7.30E-01	14.40	46.18	1.40E+04	1.85E+04
Ta0321	hypothetical protein	0.96	5.41E-02	22.33	6.67	6.43E+03	3.72E+03
Ta0323m	30S ribosomal protein S6	1.15	3.38E-01	29.79	27.73	1.52E+05	1.86E+05
Ta0325	HAM1 protein related	1.47	9.69E-01	35.84	39.14	2.53E+03	7.13E+03
Ta0330	threonyl-tRNA synthetase	1.75	1.94E-01	11.99	13.13	4.35E+04	4.58E+04
Ta0338	L-asparaginase related protein	1.85	3.44E-01	14.73	29.07	6.51E+04	1.42E+05
Ta0341	glycogen debranching enzyme isoform 1 (agl) related protein	0.76	1.67E-01	48.28	30.28	1.52E+03	2.01E+03
Ta0342	hypothetical protein	0.98	1.33E-01	26.73	25.18	2.08E+04	2.35E+04
Ta0344m	Predicted kinase related to galactokinase and mevalonate kinase	0.71	7.64E-02	30.74	48.14	9.54E+04	4.36E+04
Ta0347	hypothetical protein	0.85	3.02E-02	15.33	29.76	8.53E+03	5.52E+03
Ta0350	hypothetical protein	0.67	1.92E-02	16.82	41.51	1.38E+03	1.38E+03
Ta0353	hypothetical protein	1.49	9.93E-01	33.33	36.94	1.50E+04	9.16E+03
Ta0356	putative deoxyhypusine synthase	1.79	2.46E-01	18.93	13.80	7.38E+04	1.94E+05
Ta0358	50S ribosomal protein L12	0.56	8.94E-02	47.89	20.50	3.91E+05	3.17E+05
Ta0359	50S ribosomal protein L10	1.75	1.87E-01	9.31	13.72	1.58E+05	2.91E+05
Ta0360	50S ribosomal protein L1	1.13	1.37E-01	10.60	26.80	2.81E+05	2.55E+05
Ta0362	hypothetical protein	0.85	1.14E-01	35.23	17.77	2.53E+03	4.42E+03
Ta0388	hypothetical protein	1.19	4.76E-01	39.11	24.78	9.67E+03	1.67E+04
Ta0391	DNA-directed RNA polymerase subunit alpha	1.11	5.15E-02	12.77	12.94	1.03E+05	5.69E+04
Ta0392	DNA-directed RNA polymerase subunit A	0.94	2.05E-01	36.74	33.64	1.96E+05	5.21E+04
Ta0393m	transcription elongation factor NusA	0.73	5.35E-02	26.95	36.95	3.44E+05	2.22E+05
Ta0396m	predicted serine/threonine protein kinase	0.64	3.20E-02	29.90	14.67	1.30E+04	6.20E+03
Ta0398	acyl-CoA dehydrogenase, short-chain specific, related protein	1.42	8.37E-01	7.55	41.82	1.57E+05	1.79E+05
Ta0399	inorganic pyrophosphatase	1.84	2.23E-01	9.77	21.11	2.26E+05	2.63E+05

Ta0403	dihydroorotate dehydrogenase electron transfer subunit	1.26	4.41E-01	17.55	31.31	2.56E+04	2.52E+04
Ta0406m	Hydroxymethylglutaryl-CoA reductase	0.94	1.94E-01	39.96	13.74	3.26E+04	1.72E+04
Ta0419	hypothetical protein	0.72	8.55E-02	38.07	20.23	3.07E+03	1.49E+03
Ta0432	30S ribosomal protein S9	1.04	8.62E-02	19.30	17.86	9.98E+04	2.52E+05
Ta0436	bifunctional short chain isoprenyl diphosphate synthase related protein	1.01	2.27E-01	38.37	14.06	1.46E+04	1.85E+04
Ta0444	elongation factor Tu	0.94	3.29E-02	10.72	27.04	1.45E+06	1.38E+06
Ta0445	30S ribosomal protein S10	0.53	2.98E-02	30.18	44.07	1.82E+05	3.67E+04
Ta0446	elongation factor EF-2	1.24	4.23E-01	26.64	23.03	5.72E+05	3.28E+05
Ta0451m	ATPase (AAA superfamily)	0.54	5.11E-03	9.86	47.08	5.75E+03	4.68E+03
Ta0456m	Malate oxidoreductase (malic enzyme)	1.76	3.78E-01	9.04	25.23	2.30E+04	5.39E+04
Ta0457	hypothetical protein	0.73	1.66E-02	16.28	31.76	1.20E+04	1.00E+04
Ta0458	hypothetical protein	0.50	6.74E-03	20.04	29.04	8.63E+03	1.99E+03
Ta0459	hypothetical protein	1.41	8.51E-01	24.98	40.12	8.33E+03	1.31E+04
Ta0460	endonuclease V related protein	0.55	5.04E-03	13.51	38.37	4.25E+03	4.96E+03
Ta0461	hypothetical protein	1.06	2.28E-01	33.68	13.85	2.90E+02	5.66E+02
Ta0462	probable methylmalonyl-CoA mutase, alpha subunit, N-terminus	0.73	5.61E-02	25.62	42.12	1.06E+04	3.81E+03
Ta0464	hypothetical protein	0.90	1.39E-01	33.85	26.65	5.66E+03	1.09E+04
Ta0466	hypothetical protein	1.50	9.69E-01	24.71	24.79	3.13E+05	7.14E+05
Ta0468	seryl-tRNA synthetase	1.12	3.11E-01	29.00	30.76	6.95E+04	5.57E+04
Ta0472	hypothetical protein	1.23	5.03E-01	19.06	44.30	1.24E+03	9.94E+02
Ta0477	hypothetical protein	1.86	5.26E-01	26.16	45.19	6.49E+03	9.82E+03
Ta0479	glucosamine-fructose-6-phosphate aminotransferase related protein	0.56	2.63E-02	29.69	28.10	1.57E+03	1.05E+03
Ta0486	uridylate kinase	1.80	4.80E-01	24.06	31.97	1.50E+04	2.42E+04
Ta0487	competence damage-inducible protein A	1.16	3.46E-01	18.97	38.69	1.13E+04	1.78E+04
Ta0493	molybdopterin biosynthesis protein MoaB related protein	1.63	7.80E-01	29.58	42.34	4.30E+04	6.11E+04
Ta0499	hypothetical protein	1.80	2.17E-01	14.72	15.98	2.85E+04	4.88E+04
Ta0506m	thiamine biosynthesis protein ThiI	0.66	2.15E-02	26.02	7.46	4.55E+03	2.81E+03
Ta0508m	cell division protein FtsZ	1.94	3.96E-01	14.87	40.43	3.39E+04	3.01E+04
Ta0515	signal recognition particle protein Srp54	0.54	2.54E-02	30.31	25.41	2.07E+04	8.18E+03
Ta0516m	Predicted geranylgeranyl reductase	0.77	1.31E-01	39.50	38.44	2.28E+05	1.36E+05
Ta0519	asparaginyl-tRNA synthetase	0.68	1.14E-01	43.76	34.33	6.00E+04	3.47E+04
Ta0525	hypothetical protein	1.02	3.67E-01	49.10	30.76	2.33E+03	2.21E+03
Ta0526	thiamine monophosphate kinase	0.87	7.85E-02	16.97	43.92	1.64E+04	1.31E+04
Ta0528	hypothetical protein	0.93	1.17E-01	25.93	32.97	1.54E+04	1.48E+04
Ta0535	cysteine synthase related protein	1.16	4.54E-01	38.69	33.24	2.59E+03	3.21E+03
Ta0540	heterodisulfide reductase related protein	0.89	4.83E-02	5.41	40.52	3.82E+04	3.07E+04
Ta0541	heterodisulfide reductase related protein	0.52	7.70E-02	46.20	30.63	8.11E+03	3.37E+03
Ta0542	probable D-lactate dehydrogenase	1.27	6.47E-01	43.37	34.58	1.92E+04	2.95E+04
Ta0571	glutamate-1-semialdehyde aminotransferase	1.46	9.36E-01	23.64	33.86	5.65E+04	5.42E+04
Ta0572	porphobilinogen deaminase	1.88	5.43E-01	31.99	46.99	1.85E+04	3.63E+04
Ta0575	aspartate carbamoyltransferase catalytic subunit	0.82	1.24E-02	13.68	21.85	1.41E+05	8.19E+04
Ta0587	hypothetical protein	1.29	6.78E-01	47.27	27.27	9.36E+03	2.38E+03
Ta0595m	phosphoribosylamine--glycine ligase	0.90	1.60E-01	28.47	45.73	5.16E+04	5.49E+04
Ta0598m	deoxycytidine triphosphate deaminase	0.68	2.12E-02	18.38	39.33	7.25E+03	2.94E+03
Ta0599	acyl-CoA hydrolase related protein	1.13	4.52E-01	42.09	37.11	2.12E+04	1.45E+05
Ta0601	hypothetical protein	0.79	2.39E-02	17.57	28.22	4.45E+05	4.46E+05
Ta0609m	asparaginyl-tRNA synthetase	0.76	2.03E-03	4.41	21.49	1.62E+05	6.35E+04
Ta0613m	Predicted metal-dependent RNase	0.68	3.75E-03	15.03	8.74	2.53E+04	1.39E+04
Ta0617	probable transketolase	0.94	1.05E-02	6.01	20.32	6.93E+04	5.33E+04
Ta0618	probable transketolase	1.12	2.77E-01	30.52	21.25	4.00E+04	8.21E+04
Ta0619	probable 2-keto-3-deoxy gluconate aldolase	1.00	1.72E-01	25.52	35.18	8.84E+03	9.07E+03
Ta0621	hypothetical protein	1.64	7.11E-01	19.72	35.16	2.00E+04	3.38E+04
Ta0622	inorganic polyphosphate/ATP-NAD kinase	1.85	4.88E-01	18.93	41.47	1.21E+04	2.12E+04
Ta0637	tetrahydrofolylpolyglutamate synthase related protein	1.27	2.09E-01	11.35	15.08	6.32E+03	8.52E+03

Ta0639	phenylalanyl-tRNA synthetase alpha subunit	0.83	7.41E-04	4.40	12.39	2.23E+04	1.30E+04
Ta0646	acyl-CoA dehydrogenase related protein	0.87	1.49E-02	16.59	10.55	1.15E+04	4.50E+03
Ta0652	hypothetical protein	1.13	3.52E-01	31.62	31.10	3.70E+05	1.66E+05
Ta0655	probable cobalamin biosynthesis precorrin-3 methylase	0.56	8.22E-03	15.29	43.01	7.34E+04	3.31E+04
Ta0656	cobalt-precorrin-6A synthase	0.57	6.07E-03	11.20	43.73	2.94E+04	1.26E+04
Ta0657	precorrin-8W decarboxylase related protein	0.68	3.67E-02	27.17	31.19	6.69E+04	2.26E+04
Ta0658	precorrin-2 methyltransferase related protein	1.61	8.15E-01	46.49	26.07	2.02E+04	1.72E+04
Ta0659	probable precorrin-4 methylase	0.59	7.25E-04	9.82	13.73	1.57E+05	4.26E+04
Ta0678	hypothetical protein	1.97	2.97E-01	34.66	23.34	5.44E+02	6.07E+02
Ta0685	hypothetical protein	1.50	9.63E-01	4.42	17.21	1.44E+04	1.75E+04
Ta0694	hypothetical protein	0.68	1.34E-02	20.71	17.57	3.24E+03	2.45E+03
Ta0724	hypothetical protein	0.51	2.85E-02	32.28	30.18	3.41E+03	7.41E+02
Ta0739	hypothetical protein	0.85	1.46E-01	34.94	38.22	2.49E+03	1.41E+03
Ta0753	mannonate dehydratase related protein	1.47	9.41E-01	35.50	13.48	2.27E+04	5.61E+04
Ta0762	2-phosphoglycerate kinase related protein	0.68	5.34E-02	27.48	46.80	3.22E+04	2.33E+04
Ta0778	pyruvoyl-dependent arginine decarboxylase	1.34	6.07E-01	24.74	22.92	8.09E+04	1.96E+05
Ta0781	hypothetical protein	1.02	1.53E-01	27.67	21.54	1.27E+04	5.84E+03
Ta0784	hypothetical protein	1.02	1.76E-01	27.52	27.35	7.64E+04	1.17E+05
Ta0787	chromosome segregation protein related ptotein	1.73	3.01E-01	9.33	18.81	1.85E+04	4.96E+03
Ta0788m	Diaminopimelate decarboxylase	1.80	5.43E-01	49.72	16.82	1.42E+04	1.63E+04
Ta0799	DNA replication licensing factor MCM related protein	0.85	1.02E-01	33.71	19.05	2.94E+03	3.33E+03
Ta0809	probable aldehyde dehydrogenase	1.39	6.43E-01	21.68	8.64	1.56E+06	1.81E+06
Ta0812	hypothetical protein	2.00	3.35E-01	23.39	36.19	2.64E+05	3.44E+05
Ta0813	hypothetical protein	1.18	4.29E-01	36.12	26.41	2.38E+04	2.41E+04
Ta0814	hypothetical protein	0.90	2.71E-01	47.13	44.11	3.23E+04	3.16E+04
Ta0821	hypothetical protein	0.79	3.28E-02	23.83	15.90	1.40E+03	9.39E+02
Ta0824	chorismate synthase	1.00	2.99E-01	43.07	29.49	1.58E+04	6.99E+03
Ta0825	glucose kinase related protein	1.18	4.37E-01	24.75	43.71	9.93E+04	1.57E+05
Ta0828	hypothetical protein	0.72	1.69E-02	20.76	18.83	7.00E+04	1.57E+05
Ta0835	ski2-like helicase	1.12	3.86E-02	6.79	16.37	9.93E+03	4.28E+04
Ta0836	hypothetical protein	0.61	6.58E-04	10.46	4.27	5.78E+03	5.04E+03
Ta0838	Large helicase related protein	1.37	7.82E-01	39.55	28.75	3.41E+03	1.28E+03
Ta0852	hypothetical protein	0.95	2.71E-01	46.18	28.69	9.00E+03	7.44E+03
Ta0867	hypothetical protein	0.83	1.31E-01	32.82	43.42	7.23E+03	4.28E+03
Ta0869	ABC transporter related protein	0.91	5.32E-03	10.73	10.02	8.95E+03	4.31E+03
Ta0871	hypothetical protein	0.65	4.97E-02	31.81	35.26	9.91E+04	4.61E+04
Ta0880	ribokinase related protein	0.76	2.47E-02	17.15	32.96	2.30E+04	1.95E+04
Ta0881m	Predicted acetyltransferase (isoleucine patch superfamily)	1.79	3.83E-01	28.39	17.23	8.65E+04	1.33E+05
Ta0893	hypothetical protein	1.26	3.75E-01	12.90	28.53	5.25E+03	8.62E+03
Ta0899	NAD(+) synthetase	0.88	6.31E-02	25.33	19.92	7.13E+04	7.16E+04
Ta0905	GTP-binding protein Obg related protein	1.20	4.55E-01	23.73	40.41	6.94E+03	1.21E+04
Ta0907	DNA polymerase II	1.63	6.94E-01	32.62	18.16	2.41E+04	1.39E+04
Ta0912	chromatin protein	1.11	2.89E-01	25.82	34.97	7.99E+05	1.14E+06
Ta0917	DNA polymerase III subunit beta	0.90	6.86E-02	23.32	25.25	1.66E+05	2.21E+05
Ta0929	hypothetical protein	1.59	7.58E-01	29.75	16.15	3.42E+04	3.86E+04
Ta0930	hypothetical protein	1.94	3.79E-01	24.23	36.34	9.54E+03	5.89E+04
Ta0937	prolyl-tRNA synthetase	0.70	8.05E-02	35.55	36.33	1.17E+05	7.54E+04
Ta0946	aspartyl-tRNA synthetase	1.19	1.25E-01	15.55	11.95	3.83E+04	4.96E+04
Ta0949	fibrillarin	0.69	1.14E-01	41.97	41.69	3.58E+03	3.94E+03
Ta0953m	Predicted ATPase (RecA superfamily)	0.72	9.04E-02	37.40	29.97	1.39E+04	9.73E+03
Ta0954m	putative peroxiredoxin	1.90	1.69E-01	20.08	15.60	3.24E+05	4.03E+05
Ta0967m	NADH dehydrogenase delta subunit	1.37	8.09E-01	42.83	35.72	3.03E+04	3.23E+04
Ta0971	glutamyl-tRNA(Gln) amidotransferase subunit E	1.74	4.18E-01	24.96	17.83	4.48E+04	7.60E+04
Ta0972	glutamyl-tRNA(Gln) amidotransferase subunit D	0.75	6.23E-03	5.75	30.15	4.13E+04	2.19E+04

Ta0977	methionine synthase	0.81	6.49E-02	29.07	22.58	8.19E+04	6.59E+04
Ta0978m	hypothetical protein	0.66	2.47E-03	9.77	23.75	5.07E+04	5.57E+04
Ta0987	arginine deiminase related protein	1.09	1.80E-01	11.62	36.62	1.30E+05	1.28E+05
Ta0994	gamma-glutamyltransferase related protein	1.11	1.48E-01	14.30	27.59	1.45E+04	9.40E+03
Ta0995	PcrB-like protein	1.62	4.79E-01	15.15	11.57	7.92E+03	1.71E+04
Ta0998	hypothetical protein	1.53	8.27E-01	20.70	5.45	3.66E+04	4.41E+04
Ta1000	DNA helicase II (UvrD) related protein	0.68	4.71E-02	26.82	43.93	1.56E+04	1.24E+04
Ta1001	succinate dehydrogenase	1.06	5.85E-02	16.82	12.84	8.33E+04	9.02E+04
Ta1011	peptidyl-prolyl cis-trans isomerase related protein	1.24	5.54E-01	36.10	34.42	1.03E+05	1.49E+05
Ta1014m	GTP cyclohydrolase II	0.61	1.10E-03	11.79	8.06	3.51E+04	1.38E+04
Ta1015	hypothetical protein	0.78	1.72E-01	47.89	24.40	5.83E+03	3.45E+03
Ta1016	DNA repair protein RAD25 related protein	0.62	2.19E-02	25.89	24.95	2.30E+03	1.43E+03
Ta1021	hypothetical protein	1.96	1.54E-01	20.38	17.74	8.98E+02	1.66E+03
Ta1022	molybdopterin biosynthesis protein (moeA-1) related protein	1.95	2.87E-01	17.77	30.50	3.34E+04	9.75E+04
Ta1030	DNA-directed RNA polymerase subunit D	1.82	4.40E-01	14.01	34.69	9.14E+04	1.30E+05
Ta1032	30S ribosomal protein S4	1.31	5.94E-01	27.60	25.36	2.31E+05	1.64E+05
Ta1035	flap endonuclease-1	0.56	8.50E-02	46.06	31.75	1.94E+04	1.07E+04
Ta1036	DNA repair and recombination protein RadB	1.31	7.51E-01	48.76	42.19	4.05E+03	4.42E+03
Ta1039	Signal recognition particle 19 kDa protein	0.80	1.68E-01	46.74	17.37	2.24E+04	4.08E+04
Ta1042	hypothetical protein	0.81	5.73E-02	20.56	39.87	1.09E+03	3.07E+03
Ta1055	DNA topoisomerase IV subunit B	0.90	1.67E-02	13.81	17.70	1.60E+04	1.84E+04
Ta1058	agmatinase related protein	1.50	9.91E-01	45.19	18.07	1.42E+04	2.63E+04
Ta1059	translation initiation factor IF-5A	0.80	3.97E-02	23.92	21.40	1.59E+05	1.19E+05
Ta1060m	metal-dependent hydrolase	1.43	9.00E-01	14.64	49.09	3.23E+04	2.53E+04
Ta1064	hypothetical protein	0.68	5.91E-02	34.01	27.48	2.74E+04	2.51E+04
Ta1066	phosphoribosylformylglycinamidine synthase subunit II	1.75	4.75E-01	14.27	30.57	2.80E+04	2.41E+04
Ta1073	probable sulfolipid biosynthesis protein	0.75	1.86E-02	17.40	28.14	8.29E+04	2.92E+04
Ta1075	phosphoglycerate kinase	1.33	6.43E-01	29.28	28.00	4.28E+04	4.52E+04
Ta1077	probable signal recognition particle protein	1.62	7.87E-01	48.87	16.93	2.41E+04	3.84E+04
Ta1078	aspartate aminotransferase related protein	1.68	5.42E-01	28.88	13.69	2.26E+04	3.88E+04
Ta1089m	DNA-directed RNA polymerase subunit E'	1.66	6.93E-01	15.36	39.72	6.56E+04	5.08E+04
Ta1092	30S ribosomal protein S24e	0.91	1.08E-01	19.09	43.16	7.78E+04	6.21E+04
Ta1103	glyceraldehyde-3-phosphate dehydrogenase	0.83	9.55E-03	9.59	24.24	5.26E+04	4.13E+04
Ta1104	DNA repair and recombination protein RadA	0.51	2.25E-02	29.10	34.77	8.10E+04	2.70E+04
Ta1105	hypothetical protein	1.77	3.96E-01	12.12	27.17	7.06E+03	9.05E+03
Ta1106	hypothetical protein	1.19	1.88E-01	10.99	23.75	1.47E+04	1.43E+04
Ta1107	phosphoribosylaminoimidazole carboxylase	1.75	3.83E-01	18.54	21.35	3.85E+04	6.80E+04
Ta1109	hypothetical protein	0.58	9.56E-02	45.70	42.44	5.25E+03	2.25E+03
Ta1112	translation initiation factor IF-2	1.22	1.91E-01	7.01	23.55	3.05E+04	4.31E+04
Ta1113	probable nucleoside diphosphate kinase	0.85	2.91E-02	17.24	25.44	3.66E+04	2.18E+04
Ta1128m	hypothetical protein	0.92	2.26E-01	42.07	31.45	9.96E+03	2.29E+04
Ta1131	30S ribosomal protein S15	1.05	1.45E-02	10.07	10.55	1.73E+05	2.26E+05
Ta1132m	Single-stranded DNA-specific exonuclease	1.07	1.99E-01	24.81	27.73	1.55E+04	1.73E+04
Ta1137	prefoldin subunit beta	1.08	4.58E-01	46.12	49.83	9.05E+04	1.55E+05
Ta1145	nicotinate phosphoribosyltransferase	0.50	2.32E-02	28.43	43.97	1.01E+05	3.55E+04
Ta1147	cysteinyl-tRNA synthetase	0.92	2.47E-01	42.03	39.68	5.85E+04	6.61E+04
Ta1149	hypothetical protein	1.29	4.79E-01	13.95	31.65	3.95E+04	8.62E+04
Ta1154m	Predicted 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase	2.00	1.55E-01	9.74	24.03	2.76E+04	5.80E+04
Ta1162	methionyl-tRNA synthetase	0.51	2.77E-02	32.56	26.91	1.87E+05	1.51E+04
Ta1163	lysyl-tRNA synthetase	0.78	1.51E-02	9.07	34.31	1.11E+05	3.73E+04
Ta1164	orotate phosphoribosyltransferase	1.82	3.42E-01	25.32	20.24	3.19E+05	2.40E+05
Ta1166	transcription regulator SlyA related protein	1.32	6.25E-01	22.84	35.07	2.08E+04	4.19E+04
Ta1167	30S ribosomal protein S3a	1.01	7.72E-02	19.19	20.54	2.87E+05	2.96E+05
Ta1179	homoserine dehydrogenase	0.97	1.77E-01	21.92	46.15	7.40E+04	4.79E+04
Ta1182	glucose-fructose oxidoreductase related protein	0.66	2.66E-02	27.76	12.83	2.36E+04	1.43E+04

Ta1190	30S ribosomal protein S2	0.70	4.26E-03	7.77	29.48	4.76E+05	3.04E+05
Ta1194	phosphopantetheine adenylyltransferase	0.51	3.59E-03	10.94	43.99	6.36E+04	3.33E+04
Ta1197	hypothetical protein	1.16	4.37E-01	41.34	22.36	9.42E+03	1.39E+04
Ta1201	hypothetical protein	1.26	4.90E-01	16.93	36.05	5.24E+04	4.49E+04
Ta1203	translation initiation factor IF-2 alpha subunit	0.82	6.29E-02	29.64	11.79	2.28E+04	4.27E+04
Ta1207	hypothetical protein	1.43	9.05E-01	39.83	45.92	1.10E+05	2.09E+05
Ta1210	alpha, alpha-trehalose-phosphate synthase (UDP-forming) related protein	1.15	1.14E-01	9.24	22.78	1.57E+04	1.20E+04
Ta1211	tryptophanyl-tRNA synthetase	0.62	1.15E-02	16.92	37.58	9.01E+04	4.52E+04
Ta1212	translation initiation factor IF-2B subunit alpha	0.97	1.16E-02	6.26	18.89	2.14E+05	1.74E+05
Ta1226	hypothetical protein	0.99	3.70E-01	48.96	45.15	4.11E+04	4.23E+04
Ta1227	probable carboxyphosphonoenolpyruvate phosphonomutase	0.84	4.79E-02	22.72	25.83	1.27E+04	1.48E+04
Ta1232	enoyl-CoA hydratase related protein	1.25	5.67E-01	42.77	16.41	8.62E+03	4.52E+03
Ta1237	conserved hypothetical GTP-binding protein	1.16	2.70E-01	18.60	29.88	7.39E+04	1.81E+05
Ta1247	adenylate kinase	1.64	6.38E-01	27.38	17.03	5.34E+04	4.52E+04
Ta1249	50S ribosomal protein L15	0.59	8.14E-03	20.56	15.24	2.57E+04	2.49E+04
Ta1251	30S ribosomal protein S5P	0.86	1.89E-02	15.35	19.70	9.38E+04	7.81E+04
Ta1252m	50S ribosomal protein L18	1.29	5.20E-01	26.12	24.65	1.83E+05	1.84E+05
Ta1262	30S ribosomal protein S17	1.04	4.25E-02	15.83	11.15	1.17E+05	9.70E+04
Ta1265	30S ribosomal protein S3	0.56	1.68E-03	12.65	17.92	1.70E+05	6.21E+04
Ta1266	50S ribosomal protein L22	0.69	1.68E-02	22.01	18.65	2.97E+05	2.74E+05
Ta1268	50S ribosomal protein L2	1.10	2.52E-01	28.38	24.97	7.68E+04	7.93E+04
Ta1269	50S ribosomal protein L23	1.46	9.09E-01	23.38	25.98	3.29E+05	3.00E+05
Ta1270	50S ribosomal protein L4	1.02	1.87E-02	12.99	8.77	1.29E+05	1.44E+05
Ta1272	hypothetical protein	0.88	1.98E-01	44.66	19.86	8.53E+03	5.62E+03
Ta1278	hydrogenase expression/formation protein (Anabaena hupE) related protein	0.63	8.16E-03	15.03	32.61	5.36E+04	4.28E+04
Ta1280	oligopeptide ABC transporter, ABC-binding protein related protein	1.21	3.58E-01	17.42	31.55	2.24E+04	5.02E+02
Ta1282m	Predicted transcription regulator	0.67	1.08E-01	46.03	10.83	7.99E+03	1.01E+04
Ta1291m	hypothetical protein	1.22	6.08E-01	39.74	47.59	1.48E+04	1.99E+04
Ta1292	exosome complex RNA-binding protein 1	0.80	6.29E-02	19.50	46.13	1.30E+05	6.19E+04
Ta1293	exosome complex exonuclease 1	0.80	1.69E-02	10.26	32.97	6.35E+04	5.61E+04
Ta1294m	exosome complex exonuclease 2	0.59	2.15E-03	13.14	17.13	2.34E+05	1.73E+05
Ta1296	hypothetical protein	0.60	7.96E-03	12.90	42.08	1.63E+04	5.62E+03
Ta1302	hypothetical protein	0.77	1.51E-01	40.40	47.41	1.57E+04	5.62E+04
Ta1304	hypothetical protein	1.38	6.76E-01	19.89	23.72	1.37E+04	8.03E+03
Ta1305	hypothetical protein	0.62	1.15E-02	17.45	35.49	6.52E+04	6.68E+04
Ta1311	probable Histone and other Protein Acetyltransferase (yeast HPA1)	0.96	4.92E-02	9.39	30.64	9.72E+03	2.33E+03
Ta1314	site-specific integrase/recombinase XerD related protein	0.86	1.04E-01	27.19	37.30	3.67E+02	3.58E+02
Ta1315	ribulose-phosphate 3-epimerase	1.10	4.18E-01	42.19	37.17	2.36E+04	2.22E+04
Ta1316	alcohol dehydrogenase related protein	1.16	3.25E-01	25.81	29.97	6.58E+04	4.20E+04
Ta1319	dihydroorotase	0.66	3.64E-02	29.00	25.27	3.56E+04	1.51E+04
Ta1323	probable beta-galactosidase	1.26	5.05E-01	35.51	9.69	6.57E+03	3.39E+03
Ta1324	hypothetical protein	1.00	2.28E-01	36.08	26.85	6.07E+03	5.81E+03
Ta1330	ornithine carbamoyltransferase	1.18	5.58E-01	41.52	49.31	3.59E+05	3.77E+05
Ta1341	mercuric reductase	1.11	3.76E-01	39.20	26.59	3.33E+02	4.90E+02
Ta1347	hypothetical protein	1.15	1.20E-01	12.68	20.57	6.32E+04	6.25E+04
Ta1349	hypothetical protein	0.82	8.40E-02	30.02	29.14	1.34E+04	7.33E+03
Ta1354	hypothetical protein	0.86	4.56E-03	11.58	9.71	2.71E+03	2.18E+03
Ta1355	hypothetical protein	1.36	4.92E-01	18.06	9.27	1.74E+03	1.77E+03
Ta1359	ATP binding cassette transporter, ExsD protein related	1.07	5.08E-02	16.76	8.78	1.94E+03	3.48E+03
Ta1361	GMP synthase subunit B	0.79	6.51E-02	29.90	21.88	1.81E+04	2.92E+04
Ta1382	hypothetical protein	1.03	2.03E-01	14.98	46.31	1.59E+03	1.36E+03
Ta1388	hypothetical protein	1.37	7.06E-01	21.39	31.45	5.39E+03	7.35E+03

Ta1394	Nta operon transcriptional regulator related protein	1.66	4.48E-01	16.41	14.34	1.39E+04	2.94E+04
Ta1396	hypothetical protein	1.12	1.55E-01	19.52	19.31	5.30E+02	1.46E+03
Ta1402	hypothetical protein	1.18	2.91E-01	19.30	28.19	3.80E+03	4.39E+03
Ta1411m	hypothetical protein	0.76	1.00E-02	8.00	32.48	1.18E+05	5.82E+04
Ta1418m	Predicted RNA-binding protein homologous to eukaryotic snRNP	0.64	1.31E-02	10.72	48.79	1.46E+04	1.21E+04
Ta1420	hypothetical protein	0.53	2.60E-03	13.47	28.25	3.58E+04	1.46E+04
Ta1423	hypothetical protein	0.84	4.39E-02	13.32	40.09	2.42E+03	2.09E+03
Ta1425	transcription antitermination protein NusG	1.49	9.90E-01	39.38	16.05	2.05E+04	4.98E+04
Ta1428	hypothetical protein	1.26	5.87E-01	37.73	29.45	2.11E+05	1.86E+05
Ta1429m	Predicted alternative thymidylate synthase	0.51	6.76E-06	3.57	2.93	4.77E+04	3.91E+03
Ta1432	2-phosphosulfolactate phosphatas	0.76	1.22E-02	19.39	4.70	1.79E+04	1.18E+04
Ta1435	dihydrolipoamide dehydrogenase	1.93	2.88E-01	27.75	23.61	2.79E+04	4.98E+04
Ta1436	dihydrolipoamide acetyltransferase	0.72	6.10E-03	5.09	33.25	4.66E+04	2.23E+04
Ta1438	probable 3-methyl-2-oxobutanoate dehydrogenase alpha chain precursor	0.86	1.16E-01	25.07	46.03	3.54E+04	3.00E+04
Ta1439	methionine aminopeptidase	0.78	3.41E-02	21.39	28.63	6.54E+04	4.49E+04
Ta1447	transcription factor	0.70	4.28E-02	27.30	32.52	1.96E+04	2.55E+04
Ta1451	hypothetical protein	1.21	4.50E-01	34.05	24.00	4.77E+04	5.63E+04
Ta1452	NDP-hexose 2, 3-enoyl reductase TylCII related protein	0.80	7.97E-02	27.86	37.34	3.72E+04	2.14E+04
Ta1453	hypothetical protein	0.81	6.90E-02	28.65	26.53	8.72E+02	1.19E+03
Ta1454	hypothetical protein	0.93	2.87E-01	48.08	36.42	7.17E+03	7.98E+03
Ta1476	probable glutamate formiminotransferase	1.25	4.11E-01	20.79	25.98	5.96E+04	2.11E+05
Ta1486	mannose-1-phosphate guanyltransferase related protein	1.01	1.79E-01	20.41	41.50	5.27E+04	5.21E+04
Ta1488	hypothetical protein	0.76	4.81E-02	25.34	31.89	1.84E+05	1.42E+05
Ta1492	probable RNase L inhibitor	0.74	2.48E-02	5.56	48.94	1.05E+05	4.11E+04
Ta1496	hypothetical protein	1.71	3.50E-01	11.27	18.04	1.38E+04	2.88E+04
Ta1498	probable glutamine synthetase	0.57	7.98E-03	17.94	32.81	3.47E+05	3.02E+05
Ta1506m	Exonuclease III	0.58	1.03E-02	19.24	32.85	6.84E+04	3.59E+04
Down-regulated anaerobically							
Ta0002	V-type ATP synthase subunit C	0.22	2.63E-03	20.29	61.35	2.77E+04	5.11E+03
Ta0003	V-type ATP synthase subunit F	0.20	2.42E-02	42.05	50.70	2.86E+04	6.90E+03
Ta0005	V-type ATP synthase subunit B	0.42	2.35E-03	16.75	24.24	8.20E+04	3.19E+04
Ta0006	V-type ATP synthase subunit D	0.17	4.98E-04	14.34	36.39	1.81E+04	3.10E+03
Ta0007a	ATP synthase subunit H	0.15	4.73E-03	26.76	64.42	3.57E+04	6.55E+03
Ta0018	hypothetical protein	0.20	4.93E-04	14.25	17.85	1.28E+04	1.44E+03
Ta0023m	hypothetical protein	0.49	5.42E-04	10.03	17.31	2.90E+03	1.19E+03
Ta0024	hypothetical protein	0.10	2.18E-04	10.81	94.55	1.43E+04	1.17E+03
Ta0025	hypothetical protein	0.39	3.84E-03	19.89	29.70	4.56E+03	1.87E+03
Ta0038	72K mitochondrial outer membrane protein related protein	0.17	3.15E-03	23.38	57.60	6.82E+04	9.20E+03
Ta0054	50S ribosomal protein L31	0.11	5.11E-04	9.70	173.21	1.55E+04	5.19E+02
Ta0059	S-adenosylmethionine synthetase	0.40	1.07E-03	14.49	13.21	3.64E+05	1.29E+05
Ta0069	hypothetical protein	0.36	3.99E-04	4.61	45.60	2.95E+04	1.17E+04
Ta0075	cobyric acid synthase	0.25	1.23E-02	32.63	35.56	1.10E+04	1.92E+03
Ta0083	hypothetical protein	0.14	4.61E-02	54.61	59.35	3.83E+05	1.88E+04
Ta0084	hypothetical protein	0.09	5.89E-03	28.25	173.21	1.73E+03	2.87E+02
Ta0086	hypothetical protein	0.31	1.23E-04	7.47	27.16	1.04E+04	1.61E+03
Ta0088m	ABC-type multidrug transport system, ATPase component	0.10	1.86E-02	40.41	173.21	1.10E+03	2.10E+02
Ta0097m	DNA primase	0.27	1.04E-02	30.24	42.37	3.08E+05	5.47E+04
Ta0098	hypothetical protein	0.45	3.07E-02	34.84	39.88	6.16E+04	3.94E+03
Ta0099	histidyl-tRNA synthetase	0.28	4.86E-03	24.61	19.68	9.64E+04	1.91E+04
Ta0106	hypothetical protein	0.43	1.83E-02	26.40	62.59	4.64E+03	1.77E+03
Ta0115	hypothetical protein	0.40	6.69E-03	23.01	32.08	5.02E+03	3.07E+03
Ta0118m	hypothetical protein	0.33	1.55E-03	16.29	29.66	2.50E+04	5.03E+04

Ta0124m	Predicted metal-dependent hydrolase related to alanyl-tRNA synthetase	0.37	3.36E-03	20.24	21.51	1.31E+05	5.09E+04
Ta0131	hypothetical protein	0.44	7.94E-03	2.56	82.38	2.89E+03	1.48E+03
Ta0136	nitrate reductase related protein	0.01	1.67E-02	43.42	173.21	6.96E+03	6.19E+01
Ta0139	pyrroline-5-carboxylate reductase	0.21	1.85E-03	19.34	44.11	4.90E+03	4.46E+02
Ta0143	probable ATP-binding protein	0.45	1.16E-02	27.39	4.72	6.39E+04	3.54E+05
Ta0158	hypothetical protein	0.18	1.92E-05	4.64	38.00	1.06E+04	7.23E+02
Ta0160	serotonin receptor related protein	0.48	5.90E-03	20.94	20.17	2.31E+04	7.72E+03
Ta0199	transcription factor IID	0.43	9.49E-03	20.96	54.92	6.37E+04	2.58E+04
Ta0214	tRNA pseudouridine synthase D	0.33	3.44E-05	5.87	13.48	2.56E+04	3.36E+03
Ta0229	acetyl-CoA synthetase related protein	0.24	3.86E-03	21.98	64.27	2.04E+02	2.54E+02
Ta0241	hypothetical protein	0.30	2.29E-03	19.00	29.91	7.54E+04	4.44E+03
Ta0242a	Predicted transcription regulator	0.02	1.04E-03	19.91	173.21	9.34E+03	2.21E+02
Ta0244	hypothetical protein	0.10	1.01E-03	18.35	56.15	6.96E+05	3.26E+04
Ta0247	3-deoxy-7-phosphoheptulonate synthase	0.11	7.71E-03	32.19	35.63	2.85E+04	1.89E+03
Ta0250	glucose-fructose oxidoreductase related protein	0.11	1.35E-04	7.99	103.46	2.98E+04	1.45E+03
Ta0265	probable sporulation protein (spo0KD)	0.25	3.40E-03	22.78	23.84	2.13E+04	6.08E+03
Ta0266	oligopeptide transport ATP-binding protein appF related protein	0.18	1.10E-02	33.85	6.54	1.50E+04	2.15E+03
Ta0270	sporulation protein (spo0KD) related protein	0.44	3.77E-02	15.04	124.52	1.13E+03	2.66E+02
Ta0276	molybdenum cofactor synthesis-step 1 (MOCS1) related protein	0.28	8.80E-03	18.24	121.80	1.72E+04	6.83E+02
Ta0305	hypothetical protein	0.19	3.27E-02	47.04	14.87	7.21E+03	6.92E+02
Ta0318m	Adenine-specific DNA methylase	0.49	8.15E-03	4.90	71.58	4.21E+02	1.41E+02
Ta0322	translation initiation factor IF-2 gamma subunit	0.48	5.98E-03	19.44	31.97	1.31E+05	3.97E+04
Ta0333	hypothetical protein	0.44	2.29E-03	16.87	17.89	2.33E+04	7.35E+03
Ta0336	hypothetical protein	0.32	2.72E-02	32.25	113.05	2.58E+04	6.38E+03
Ta0357	thiol-specific antioxidant related protein	0.22	9.12E-03	30.75	41.76	3.33E+04	1.08E+04
Ta0364	aspartate kinase related protein	0.00	4.63E-02	60.73	NA	2.28E+02	5.45E+02
Ta0411	cytolysin slyA related protein	0.40	5.93E-03	23.18	17.92	4.21E+03	2.23E+03
Ta0421	probable acetyl-coenzyme-A synthetase	0.30	1.61E-03	14.72	52.24	9.98E+02	1.17E+03
Ta0423	hypothetical protein	0.29	2.94E-03	8.74	99.73	4.37E+03	2.24E+02
Ta0424	hypothetical protein	0.12	1.73E-02	40.38	41.22	1.68E+05	8.75E+04
Ta0425	formate dehydrogenase related protein	0.06	4.90E-03	29.59	24.51	1.95E+05	4.04E+03
Ta0441	3-ketoacyl-(acyl-carrier-protein) reductase	0.42	2.32E-03	14.17	39.06	7.21E+04	2.84E+03
Ta0465	probable intracellular proteinase I	0.27	1.39E-02	27.01	113.10	1.20E+05	1.52E+04
Ta0489	hypothetical protein	0.46	1.40E-02	8.60	88.06	1.27E+03	5.22E+02
Ta0507	hypothetical protein	0.30	8.62E-03	27.17	49.40	2.24E+04	5.62E+03
Ta0512	tyrosyl-tRNA synthetase	0.45	1.14E-02	16.09	74.06	3.63E+04	1.46E+04
Ta0517	ferredoxin 2[4Fe-4S] related protin	0.27	2.27E-02	23.50	173.21	6.17E+02	6.63E+02
Ta0518	hypothetical protein	0.35	6.93E-04	13.32	18.08	1.73E+04	1.62E+04
Ta0523	adenylosuccinate lyase	0.42	2.91E-02	37.08	18.23	3.62E+04	1.86E+04
Ta0524	hypothetical protein	0.27	4.69E-05	6.54	20.49	2.53E+04	4.05E+03
Ta0534	peptide chain release factor 1	0.42	7.76E-03	23.23	32.64	9.26E+04	4.02E+04
Ta0536	glutamyl-tRNA reductase	0.19	3.90E-02	49.58	44.66	1.47E+04	2.10E+03
Ta0537m	Predicted ICC-like phosphoesterase	0.47	8.83E-03	18.35	53.22	1.77E+03	7.08E+02
Ta0549	hypothetical protein	0.20	9.90E-03	31.34	63.77	3.58E+03	3.67E+02
Ta0576	protoporphyrin IX magnesium chelatase related protein	0.49	8.05E-04	8.71	28.83	1.11E+04	3.35E+03
Ta0636	cell division control protein 6	0.29	1.38E-02	26.63	101.55	1.68E+03	4.95E+02
Ta0643	hypothetical protein	0.08	6.42E-03	30.63	113.48	3.80E+03	7.56E+01
Ta0644m	Site-specific DNA methylase	0.28	9.47E-03	23.47	98.03	9.97E+02	8.99E+01
Ta0660	hypothetical protein	0.39	7.69E-03	25.67	11.46	2.84E+05	2.42E+04
Ta0662m	Predicted metal-dependent hydrolase	0.44	7.28E-03	22.06	33.55	7.63E+03	2.17E+03
Ta0668	hypothetical protein	0.41	9.57E-04	14.34	4.61	1.58E+04	7.03E+03
Ta0670	hypothetical protein	0.37	2.11E-02	35.43	7.17	3.98E+04	1.23E+04
Ta0684	deoxyribose-phosphate aldolase	0.49	1.14E-02	25.21	22.15	1.48E+04	4.62E+03
Ta0687	hypothetical protein	0.37	6.32E-04	9.90	36.56	1.11E+03	6.08E+02

Ta0688	probable sugar ABC transporter, ATP-binding protein	0.24	1.19E-02	27.49	116.75	2.01E+03	2.08E+02
Ta0765	aspartate transaminase related protein	0.21	1.10E-03	16.77	39.05	5.28E+04	7.57E+03
Ta0792m	carbamoyl-phosphate synthase small subunit	0.39	2.59E-02	33.44	60.04	9.03E+03	2.80E+03
Ta0795	sugar fermentation stimulation protein	0.36	9.55E-03	26.57	40.10	2.93E+03	1.18E+03
Ta0816	pyruvate dehydrogenase	0.41	3.36E-02	37.79	40.49	5.78E+04	4.54E+03
Ta0819	citrate synthase	0.07	5.90E-03	29.95	173.21	1.01E+04	4.02E+01
Ta0842	hypothetical protein	0.29	3.00E-03	11.09	93.71	2.42E+03	9.74E+02
Ta0858	glycerate dehydrogenase	0.49	4.04E-03	18.23	21.05	3.22E+04	1.97E+04
Ta0868m	Predicted transcription regulator	0.08	2.10E-04	8.68	173.21	2.58E+03	3.04E+02
Ta0872	iron-dependent transcription repressor related protein	0.42	4.00E-02	40.12	38.07	1.36E+04	5.51E+03
Ta0890	hypothetical protein	0.38	4.16E-02	42.02	45.78	4.92E+04	2.25E+04
Ta0891	endonuclease IV related protein	0.10	4.02E-05	2.45	123.18	8.45E+03	7.10E+02
Ta0892	hypothetical protein	0.11	3.27E-02	49.57	88.82	1.57E+03	4.24E+01
Ta0913	hypothetical protein	0.12	3.05E-04	11.33	94.67	4.73E+03	5.96E+02
Ta0920	regulatory protein moxR related protein	0.32	1.17E-03	14.04	41.03	5.58E+03	2.56E+03
Ta0922	hypothetical protein	0.32	2.85E-02	38.26	59.73	8.66E+03	2.00E+03
Ta0941	GTP-binding protein related protein, GTP1/OBG-family	0.33	1.31E-03	12.81	50.27	2.09E+04	5.59E+03
Ta0942	glutamyl-tRNA synthetase	0.40	1.68E-02	31.92	3.69	3.06E+04	7.57E+03
Ta0943	hypothetical protein	0.12	3.60E-03	22.29	173.21	1.54E+03	8.33E+02
Ta0975m	DNA primase small subunit	0.35	1.05E-02	25.90	58.92	4.98E+03	1.04E+03
Ta0983	hypothetical protein	0.16	1.72E-02	37.99	99.88	1.97E+03	1.48E+02
Ta0988	hypothetical protein	0.31	1.84E-04	10.11	9.16	7.74E+03	1.49E+03
Ta1038	DNA primase large subunit	0.48	3.80E-02	37.29	31.58	1.85E+04	4.27E+03
Ta1047	hypothetical protein	0.47	1.88E-02	29.57	29.04	1.99E+04	9.99E+03
Ta1057m	50S ribosomal protein L10	0.10	1.27E-04	11.00	22.32	1.02E+05	6.08E+03
Ta1088	heat shock protein DnaJ related protein	0.22	1.01E-02	32.13	19.05	6.14E+03	7.35E+02
Ta1121	cob(I)alamin adenosyltransferase related protein	0.18	1.68E-02	37.94	49.08	1.02E+04	4.96E+02
Ta1127	hypothetical protein	0.30	1.57E-02	33.96	27.21	1.83E+05	1.18E+04
Ta1141	histidine triad (HIT) related protein	0.48	6.95E-03	17.00	47.25	3.12E+04	5.37E+04
Ta1155	ribonuclease Z	0.27	3.46E-06	3.01	13.74	2.56E+04	9.14E+03
Ta1158	glycerol-1-phosphate dehydrogenase	0.33	1.55E-02	12.39	141.61	1.85E+03	4.52E+02
Ta1159	hypothetical protein	0.38	9.97E-05	6.62	19.26	7.99E+04	2.29E+04
Ta1175	VAT-2 protein	0.31	6.43E-03	25.03	39.61	7.89E+04	1.78E+04
Ta1181	hypothetical protein	0.49	1.08E-03	3.62	40.07	8.16E+04	3.37E+04
Ta1186	conserved hypthetical protein	0.00	8.65E-04	19.37	NA	1.16E+03	2.06E+02
Ta1191	tRNA splicing endonuclease	0.36	2.12E-02	35.35	18.40	7.52E+03	2.32E+03
Ta1214	thymidine kinase related protein	0.38	4.09E-02	43.28	19.07	2.71E+04	3.69E+03
Ta1231	hypothetical protein	0.20	2.42E-02	40.26	100.92	3.37E+03	2.40E+02
Ta1238	hypothetical protein	0.35	1.61E-02	32.11	32.40	4.51E+04	1.91E+04
Ta1243m	Nop56p-related protein (ribosomal biogenesis)	0.32	7.15E-03	26.26	28.40	4.45E+03	3.23E+03
Ta1244	tRNA pseudouridine synthase B	0.34	3.10E-03	17.45	50.42	1.85E+04	5.90E+03
Ta1275	putative target YPL207w of the HAP2 transcriptional complex related protein	0.31	4.93E-03	24.31	11.27	1.41E+04	4.09E+03
Ta1298	hypothetical protein	0.46	9.52E-03	6.06	80.07	1.57E+04	1.06E+04
Ta1300m	rRNA dimethyladenosine transferase	0.39	2.72E-02	37.00	27.38	2.86E+03	1.22E+03
Ta1307	hypothetical protein	0.44	4.76E-04	10.76	15.80	1.14E+04	4.94E+03
Ta1350m	carbamate kinase	0.50	4.38E-02	37.62	38.75	6.25E+03	3.89E+03
Ta1356	heme biosynthesis protein (NirJ) related protein	AE	3.22E-03	27.44	NA	5.01E+02	0.00E+00
Ta1363	hypothetical protein	0.00	3.05E-04	14.80	NA	4.12E+02	9.38E+00
Ta1381	hypothetical protein	0.00	2.67E-04	14.31	NA	3.69E+02	2.67E+01
Ta1386	hypothetical protein	0.15	3.83E-03	23.73	96.84	6.64E+02	8.63E+01
Ta1387	hypothetical protein	0.12	2.95E-03	20.29	173.21	2.81E+02	3.00E+01
Ta1397	hypothetical protein	0.14	2.01E-02	40.04	136.96	9.04E+02	4.78E+01
Ta1398	hypothetical protein	0.44	3.44E-02	35.30	51.77	2.26E+04	9.76E+02
Ta1399	hypothetical protein	0.31	2.97E-03	19.80	35.63	8.35E+03	2.23E+03

Ta1401	hypothetical protein	0.00	6.44E-04	17.94	NA	8.13E+02	3.91E+01
Ta1417	hypothetical protein	0.36	5.30E-03	20.24	52.10	5.52E+03	2.75E+03
Ta1424	hypothetical protein	0.30	4.07E-02	45.27	50.02	4.46E+02	1.33E+02
Ta1443	transposon tn10 jemc related protein	0.00	1.85E-03	23.64	NA	9.43E+02	3.48E+01
Ta1445	hypothetical protein	0.00	3.47E-03	28.00	NA	1.02E+03	7.66E+01
Ta1455	hypothetical protein	0.17	7.45E-05	4.25	70.60	9.63E+04	2.40E+03
Ta1456	acetyl-CoA acetyltransferase	0.04	3.77E-03	27.77	65.43	1.05E+05	3.42E+03
Ta1457	hypothetical protein	0.09	1.31E-02	37.64	114.70	1.32E+05	5.59E+03
Ta1474	hypothetical protein	0.21	1.19E-02	22.94	173.21	9.44E+02	5.83E+02
Ta1475	ribonucleotide-diphosphate reductase alpha subunit	0.08	2.07E-05	6.34	57.83	3.10E+06	1.46E+05
Ta1489	GTP-binding protein	0.15	4.72E-02	54.78	7.80	6.32E+03	1.19E+03
Ta1491	hypothetical protein	0.41	4.63E-04	11.45	13.03	2.25E+04	1.10E+04
Ta1493	queuine tRNA-ribosyltransferase related protein	0.41	4.03E-04	8.36	28.85	3.64E+04	1.30E+04
Ta1500m	replication factor C small subunit	0.26	5.59E-03	25.80	32.15	1.15E+05	2.90E+04
Up-regulated anaerobically							
Ta0009	hypothetical protein	4.63	5.07E-04	13.14	10.61	1.71E+04	5.81E+04
Ta0011	UDP-glucose 4-epimerase related protein	4.22	1.93E-03	19.70	13.82	1.66E+04	9.24E+04
Ta0040	valyl-tRNA synthetase	3.16	4.62E-02	23.42	30.06	5.53E+04	5.35E+04
Ta0044m	CTP synthetase	7.49	1.25E-03	40.54	15.05	8.87E+04	2.02E+05
Ta0068	4-aminobutyrate aminotransferase	14.30	8.57E-05	33.39	8.95	6.05E+04	7.64E+05
Ta0085m	Galactonate dehydratase	2.95	1.49E-02	18.04	18.85	1.31E+05	3.23E+05
Ta0090	phosphoribosylaminoimidazole synthetase	3.80	1.22E-03	7.92	12.49	3.34E+04	1.55E+05
Ta0114	N-methyl-transferase related protein	7.28	2.84E-02	90.10	36.67	3.76E+02	8.19E+02
Ta0117	probable isocitrate dehydrogenase	6.43	2.70E-04	9.91	10.78	1.19E+05	3.94E+05
Ta0161	hypothetical protein	4.66	3.52E-04	31.69	2.43	1.99E+04	9.80E+04
Ta0165	hypothetical protein	Infinity	6.90E-03	NA	33.84	1.90E+02	3.14E+04
Ta0191	probable glucose 1-dehydrogenase	4.06	1.57E-02	33.36	24.24	2.56E+04	9.81E+04
Ta0194m	Predicted hydrolase (alpha/beta superfamily)	4.71	4.11E-04	31.47	4.45	4.82E+03	1.51E+04
Ta0196	death-associated protein kinase related protein	2.98	3.25E-02	21.61	24.72	1.73E+03	5.43E+03
Ta0208	dihydrolipoamide dehydrogenase	12.54	2.10E-02	40.11	41.05	9.56E+03	1.53E+05
Ta0227	hypothetical protein	13.65	1.68E-02	173.21	34.15	8.47E+02	4.11E+03
Ta0236m	hypothetical protein	2.67	1.87E-02	22.77	15.50	4.41E+04	9.48E+04
Ta0258	probable fumarase	3.98	2.51E-02	20.58	30.04	2.13E+05	1.18E+06
Ta0260	ferrodoxin oxidoreductase beta subunit	3.32	2.29E-04	10.71	5.85	3.93E+04	1.52E+05
Ta0294	acyl-CoA dehydrogenase (SHORT-CHAIN SPECIFIC) related protein	3.09	1.60E-02	32.84	15.82	4.86E+04	1.12E+05
Ta0296	acetyl-CoA acetyltransferase	2.89	3.00E-02	15.85	24.10	2.80E+04	6.96E+04
Ta0298	alpha-glucosidase related protein	9.82	2.21E-03	56.11	19.23	5.58E+02	5.64E+02
Ta0301	Tricorn protease interacting factor F2	3.37	1.83E-03	23.08	8.22	2.73E+05	3.53E+05
Ta0326	fixC protein related	8.19	3.70E-04	25.91	11.81	6.92E+04	4.19E+05
Ta0328	FIXA (related to carnitine metabolism) related protein	10.14	2.92E-03	44.45	21.84	8.31E+03	3.83E+05
Ta0329	probable electron transfer flavoprotein, alpha subunit	13.89	1.32E-03	9.64	19.29	3.97E+04	4.08E+05
Ta0343	nucleotide sugar epimerase related protein	3.82	1.97E-03	20.52	12.26	1.49E+05	5.30E+05
Ta0355	spermidine synthase	3.79	1.05E-03	24.12	7.89	5.25E+04	2.56E+05
Ta0361	50S ribosomal protein L11	2.21	1.56E-02	11.19	11.62	4.81E+04	7.03E+04
Ta0404	dihydroorotate dehydrogenase	2.74	4.39E-02	37.05	18.33	1.13E+04	2.61E+04
Ta0413	cofactor-independent phosphoglycerate mutase	4.48	2.46E-02	25.64	31.75	1.60E+03	1.54E+04
Ta0414	Glutamate synthase [NADPH] SMALL CHAIN related protein	9.54	2.88E-03	84.89	18.13	2.48E+04	9.39E+04
Ta0439	piperideine-6-carboxilic acid dehydrogenase related protein	4.00	7.36E-03	51.58	9.95	8.54E+04	6.81E+04
Ta0469	S-adenosyl-L-homocysteine hydrolase	2.37	1.43E-02	13.23	13.12	2.14E+05	3.85E+05
Ta0473	putative peroxiredoxin	6.57	4.71E-03	7.54	23.46	4.61E+04	1.20E+05
Ta0476	3-hydroxybutyryl-CoA dehydrogenase	4.01	3.85E-04	23.89	4.34	3.14E+04	2.05E+05
Ta0480	ornithine cyclodeaminase	3.03	1.11E-02	24.31	15.66	1.84E+04	5.16E+04

Ta0500	probable benzoylformate decarboxylase	4.70	6.86E-03	26.79	21.46	7.55E+03	2.28E+04
Ta0552	hypothetical protein	Infinity	1.22E-04	NA	11.71	6.90E+02	8.50E+03
Ta0573	uroporphyrinogen-III synthetase	4.23	2.46E-03	35.58	10.76	4.98E+03	2.02E+04
Ta0579	hypothetical protein	5.22	1.99E-04	18.70	7.82	4.49E+04	1.79E+05
Ta0582	acetyl-CoA acetyltransferase	4.62	6.76E-04	23.32	9.73	3.53E+04	1.46E+05
Ta0583	hypothetical protein	3.25	8.08E-03	25.45	15.21	3.10E+03	2.05E+04
Ta0584a	hypothetical protein	28.45	7.50E-03	66.00	32.66	2.97E+03	5.55E+04
Ta0603	aldehyde dehydrogenase related protein	6.49	1.06E-02	2.33	29.48	2.34E+05	2.96E+05
Ta0612	proteasome, beta chain	3.18	3.88E-02	61.69	9.43	2.05E+05	1.90E+05
Ta0626m	2-oxoacid-ferredoxin oxidoreductase, gamma subunit	307.57	1.81E-02	173.21	44.59	4.25E+01	1.38E+05
Ta0627	pyruvate ferredoxin oxidoreductase, delta subunit related protein	AN	3.36E-02	NA	54.51	0.00E+00	3.63E+04
Ta0628m	Pyruvate:ferredoxin oxidoreductase, alpha subunit	126.80	3.69E-04	152.83	15.26	7.49E+02	7.46E+04
Ta0629	ferrodoxin oxidoreductase beta subunit	53.66	1.36E-03	108.15	21.00	9.84E+01	3.02E+04
Ta0630m	SAM-dependent methyltransferase	3.59	8.88E-03	15.38	20.27	2.25E+03	5.57E+03
Ta0635	probable glutamate dehydrogenase	14.85	6.00E-03	52.83	28.80	1.05E+05	9.58E+05
Ta0648	hypothetical protein	5.02	3.16E-03	37.51	15.63	1.39E+04	1.27E+05
Ta0650	argininosuccinate synthase	76.15	3.19E-03	90.79	26.77	2.10E+02	2.59E+04
Ta0664	hypothetical protein	2.70	3.29E-02	30.64	17.33	3.46E+03	1.32E+04
Ta0693	hypothetical protein	4.17	4.36E-02	53.77	33.06	2.18E+04	7.63E+04
Ta0723	hypothetical protein	5.74	3.59E-02	86.60	34.52	1.42E+03	1.12E+04
Ta0751	hypothetical protein	3.44	2.26E-02	39.04	21.36	3.53E+03	2.05E+04
Ta0764m	metal-dependent hydrolase	3.68	1.02E-03	23.19	7.49	2.16E+04	6.80E+04
Ta0772m	ferrodoxin oxidoreductase beta subunit	9.13	4.13E-04	23.13	12.84	5.89E+04	5.27E+05
Ta0773	probable 2-oxoacid ferredoxin oxidoreductase, alpha chain	7.01	1.23E-03	34.20	15.05	1.21E+05	1.02E+06
Ta0777	leucyl-tRNA synthetase	3.39	3.94E-02	54.31	21.47	6.41E+04	6.64E+04
Ta0811	glycine hydroxymethyltransferase related protein	3.38	7.29E-03	33.99	12.05	3.74E+04	9.59E+04
Ta0815	tricorn protease interacting factor F3	4.44	2.02E-03	35.78	10.68	6.74E+04	1.37E+05
Ta0820	hypothetical protein	2.81	7.34E-03	28.55	5.69	1.11E+04	2.09E+03
Ta0823	hypothetical protein	3.35	4.22E-03	29.36	9.84	1.27E+03	9.77E+03
Ta0837	NADH peroxidase related protein	3.82	1.33E-02	20.94	23.53	8.99E+03	2.65E+04
Ta0840	VAT ATPase (VCP-like ATPase)	2.72	4.42E-02	26.62	22.80	1.68E+05	1.99E+05
Ta0845	hypothetical protein	4.68	5.15E-03	59.48	9.63	2.49E+02	2.92E+03
Ta0849	alanyl-tRNA synthetase	3.33	4.26E-03	33.68	6.36	7.95E+04	7.54E+04
Ta0878	ribose-5-phosphate isomerase A	6.77	3.95E-04	25.77	10.95	3.91E+03	6.11E+04
Ta0882	phosphopyruvate hydratase	3.62	2.27E-02	9.18	28.01	3.98E+04	7.20E+04
Ta0884	ferredoxin related protein	4.70	1.54E-02	32.67	27.22	5.05E+03	5.41E+04
Ta0886	pyruvate phosphate dikinase	4.38	2.46E-02	77.47	18.94	3.81E+03	4.61E+03
Ta0895	hypothetical protein	3.47	2.05E-02	14.41	25.87	1.35E+04	1.54E+04
Ta0896	pyruvate kinase	7.35	6.44E-03	87.30	19.65	1.99E+04	4.45E+04
Ta0897	glucose 1-dehydrogenase	10.95	9.62E-04	34.35	16.55	5.14E+04	5.42E+05
Ta0898	methylenetetrahydrofolate dehydrogenase (NADP+) related protein	5.04	9.49E-03	28.14	24.72	3.24E+04	1.97E+05
Ta0910	translation initiation factor IF-2B subunit delta	5.51	3.16E-02	27.66	38.23	1.08E+03	8.23E+03
Ta0932m	tRNA pseudouridine synthase A	3.49	1.12E-02	25.37	19.39	6.31E+02	2.10E+03
Ta0934m	succinyl-diaminopimelate desuccinylase	6.34	1.21E-02	48.11	28.26	2.55E+05	5.66E+05
Ta0952	malate dehydrogenase	9.50	1.57E-02	6.15	36.17	2.83E+05	2.27E+06
Ta0980	thermosome, alpha chain	2.40	3.57E-02	20.21	16.90	1.61E+06	1.36E+06
Ta0984	probable thioredoxin reductase	8.83	2.93E-04	47.61	9.14	3.50E+04	3.76E+05
Ta1013	indolepyruvate ferredoxin oxidoreductase	12.54	1.09E-02	87.00	32.42	7.33E+02	7.51E+03
Ta1053	N-carbamoylsarcosine amidase related protein	12.60	4.21E-02	42.21	51.57	1.49E+04	1.89E+05
Ta1111	UDP-glucose 4-epimerase related protein	12.38	9.99E-03	38.34	32.76	1.96E+04	2.58E+05
Ta1129	sulfide-quinone reductase related protein	6.67	1.48E-02	86.96	26.41	2.92E+02	3.05E+03
Ta1142	tungstealdehyde ferredoxaldehyde ferredoxin oxidoreductase related protein	4.65	2.81E-03	25.81	15.93	5.96E+03	2.20E+04

Ta1153	probable acetyl-CoA synthetase	50.30	5.78E-07	5.97	2.96	3.30E+04	6.28E+05
Ta1157	dihydrodipicoline synthase related protein	7.38	2.60E-03	39.40	19.07	4.17E+04	1.52E+05
Ta1193	aspartate aminotransferase related protein	4.02	7.84E-03	25.80	19.90	1.07E+05	2.51E+05
Ta1206	hypothetical protein	9.88	4.09E-04	39.61	12.16	2.32E+04	1.51E+05
Ta1208	nitrilase related protein	2.31	4.64E-02	20.57	16.95	2.15E+04	5.26E+04
Ta1239	hypothetical protein	2.27	4.76E-02	28.57	9.48	3.58E+04	2.47E+05
Ta1281	phosphomethylpyrimidine kinase	2.91	2.11E-02	38.83	11.22	1.58E+04	7.50E+04
Ta1331	succinyl-CoA synthetase alpha subunit	4.14	3.12E-03	40.96	9.18	6.45E+04	5.22E+05
Ta1332m	succinyl-CoA synthetase subunit beta	3.45	4.60E-02	7.86	34.25	2.63E+05	6.44E+05
Ta1351	purine phosphoribosyltransferase related protein	3.88	4.64E-02	16.20	36.88	1.77E+04	6.72E+04
Ta1357	glycine dehydrogenase subunit 2	4.87	3.21E-03	29.37	16.75	6.44E+03	5.74E+04
Ta1358	glycine dehydrogenase subunit 1	5.68	4.25E-03	33.38	19.99	7.69E+03	2.28E+04
Ta1362m	Fe2+ uptake regulation protein	3.75	9.07E-04	2.23	11.77	4.83E+03	2.18E+04
Ta1367m	adenylosuccinate synthetase	2.23	4.43E-02	29.12	3.41	4.17E+04	6.10E+04
Ta1368m	Peroxiredoxin	3.56	8.29E-03	27.26	17.26	8.66E+04	4.97E+05
Ta1374	hypothetical protein	8.58	1.08E-03	45.36	15.01	7.32E+04	8.79E+04
Ta1393	hypothetical protein	8.80	1.24E-02	20.36	33.11	2.22E+04	2.43E+05
Ta1395	mandelate racemase related protein	3.79	6.37E-03	13.81	19.34	5.59E+03	1.78E+04
Ta1412	hypothetical protein	20.48	4.18E-03	44.78	27.11	5.29E+04	1.15E+06
Ta1413	sulfide dehydrogenase related protein	5.83	4.76E-02	24.78	45.19	2.45E+05	1.12E+06
Ta1448m	Ferredoxin	6.76	3.57E-02	43.20	42.29	1.19E+04	5.64E+04
Ta1463	hypothetical protein	4.37	3.96E-02	49.64	33.94	1.59E+03	1.01E+04
Ta1466	hypothetical protein	14.41	1.87E-02	173.21	36.41	5.52E+01	1.40E+03
Ta1477	formiminotransferase cyclodeaminase related protein	4.26	4.15E-02	17.04	37.54	2.89E+03	2.57E+04
Ta1481	hypothetical protein	3.05	1.23E-02	12.51	19.50	8.24E+03	1.61E+04
Ta1485	hypothetical protein	3.19	8.46E-03	6.45	18.86	3.06E+04	5.21E+04
Ta1490	tricorn protease	2.78	1.44E-02	27.98	12.32	4.58E+04	3.75E+04
Ta1509	serine hydroxymethyltransferase	3.17	3.04E-02	8.25	27.67	5.93E+04	1.19E+05
Not quantified							
Ta0013	probable Fe-superoxide dismutase (sod)	0.85	2.45E-01	52.00	29.04	1.10E+06	4.27E+05
Ta0014m	N-glycosylase/DNA lyase	0.63	1.58E-01	33.66	112.34	4.01E+03	6.99E+02
Ta0020a	hypothetical protein	2.71	1.87E-01	14.81	48.39	7.99E+03	1.92E+04
Ta0022	hypothetical protein	0.66	1.65E-01	56.28	22.07	8.25E+04	3.91E+04
Ta0036m	DNA polymerase II large subunit	0.28	9.53E-02	62.62	79.96	2.61E+03	1.03E+03
Ta0037	hypothetical protein	1.80	5.64E-01	55.07	10.34	6.87E+04	7.70E+04
Ta0042	hypothetical protein	0.40	6.51E-02	48.36	49.89	8.41E+03	2.92E+03
Ta0042a	hypothetical protein	0.00	3.74E-01	173.21	NA	1.50E+03	1.84E+02
Ta0047	anaerobic sulfite reduction protein B related protein	2.07	3.40E-01	45.43	30.54	1.82E+03	3.20E+03
Ta0052	hypothetical protein	0.83	4.46E-01	91.00	11.37	1.76E+03	1.46E+03
Ta0053	50S ribosomal protein L39e	0.76	4.91E-01	87.33	137.85	1.57E+03	2.17E+03
Ta0058	N-terminal acetyltransferase complex ard1 subunit related protein	2.21	5.32E-01	107.20	39.34	8.17E+03	1.35E+04
Ta0060	phosphoribosylaminoimidazolecarboxamidef ormyltransferase	2.03	4.61E-01	28.67	52.44	1.05E+05	2.46E+05
Ta0061	probable 3-HYDROXYBUTYRYL-COA DEHYDRATASE	1.46	9.65E-01	25.62	73.49	1.01E+05	1.96E+05
Ta0064m	Acylphosphatase	6.16	1.33E-01	41.56	68.91	4.19E+03	1.06E+04
Ta0066	gliding motility protein (gldA) related protein	1.82	7.43E-01	27.75	86.67	3.63E+02	5.58E+02
Ta0073	orotidine 5'-phosphate decarboxylase	2.13	3.09E-01	9.32	44.31	1.09E+04	1.69E+04
Ta0091	30S ribosomal protein S12	1.27	6.20E-01	6.71	56.46	1.51E+04	9.81E+03
Ta0093	DNA-binding protein HTa	1.07	4.91E-01	54.43	46.55	1.54E+06	1.97E+06
Ta0095	hypothetical protein	0.73	2.29E-01	41.20	96.70	9.44E+03	4.98E+03
Ta0100	hypothetical protein	2.28	4.51E-01	104.94	20.23	9.49E+03	2.81E+04
Ta0108	peptidyl-tRNA hydrolase	0.43	6.93E-02	43.86	84.88	1.18E+04	3.57E+03
Ta0112	aconitate hydratase	2.20	1.46E-01	36.41	19.06	2.06E+05	1.39E+05
Ta0121	3-HYDROXYBUTYRYL-COA DEHYDRATASE related protein	1.00	6.16E-01	43.19	143.40	1.42E+04	9.37E+03

Ta0123	phosphoenolpyruvate carboxykinase (GTP)	2.98	6.39E-02	31.43	30.21	1.41E+03	5.69E+03
Ta0125	alkyl hydroperoxide reductase subunit f related protein	1.21	6.28E-01	19.49	71.55	1.10E+05	1.02E+05
Ta0132	hypothetical protein	2.63	2.51E-01	50.57	47.92	1.57E+04	3.23E+04
Ta0135	hypothetical protein	0.15	5.13E-02	55.38	128.46	1.00E+04	1.20E+03
Ta0140	hypothetical protein	0.17	6.26E-02	59.77	28.56	4.21E+04	6.53E+03
Ta0147	hypothetical protein	0.65	2.42E-01	34.51	144.39	1.37E+03	2.59E+02
Ta0152	probable peroxiredoxin	1.30	7.21E-01	32.43	53.84	4.82E+05	3.12E+05
Ta0155m	hypothetical protein	0.91	2.53E-01	35.39	58.90	3.48E+03	3.76E+03
Ta0159	hypothetical protein	0.43	1.69E-01	72.88	34.17	1.50E+05	8.97E+03
Ta0162m	NADH:flavin oxidoreductase	0.52	1.06E-01	53.72	25.00	2.43E+05	2.95E+04
Ta0166	hypothetical protein	1.32	7.96E-01	70.33	3.49	1.27E+04	1.11E+04
Ta0169	citrate synthase	3.04	2.69E-01	28.44	67.24	2.22E+05	4.66E+05
Ta0171	hypothetical protein	AE	1.68E-01	102.99	NA	1.35E+03	0.00E+00
Ta0174	hypothetical protein	1.20	6.04E-01	54.86	33.45	5.50E+04	1.37E+05
Ta0176m	Predicted RNAse P protein subunit RPR2	1.17	7.97E-01	99.12	114.38	1.97E+03	8.14E+02
Ta0178	hypothetical protein	0.41	1.14E-01	59.01	70.12	1.46E+03	5.59E+02
Ta0182	hypothetical protein	0.60	3.05E-01	86.38	48.01	1.09E+05	2.32E+04
Ta0184	hypothetical protein	1.14	6.21E-01	75.90	16.30	1.83E+04	1.16E+04
Ta0185	hypothetical protein	0.89	3.11E-01	29.81	87.10	1.60E+03	1.37E+03
Ta0192m	SAM-dependent methyltransferase	0.60	3.22E-01	85.73	78.09	2.28E+03	2.59E+03
Ta0193m	Predicted phosphohydrolase (HD superfamily)	1.02	2.97E-01	15.25	62.50	1.21E+04	8.94E+03
Ta0195	hypothetical protein	3.31	6.65E-02	39.57	33.51	5.19E+03	8.78E+03
Ta0200	hypothetical protein	1.10	4.61E-01	31.60	63.08	4.59E+03	5.66E+03
Ta0202	FERRIC TRANSPORT ATP-BINDING PROTEIN AFUC related protein	4.77	6.03E-02	37.76	44.30	1.38E+05	3.88E+05
Ta0203	hypothetical protein	2.59	2.53E-01	76.67	33.18	8.02E+04	8.62E+04
Ta0204	hypothetical protein	2.96	1.42E-01	12.54	46.56	7.46E+04	1.40E+05
Ta0205	2, 4-dienoyl-CoA reductase (NADPH) precursor related protein	4.09	1.01E-01	86.90	41.08	7.38E+02	2.07E+03
Ta0211m	Electron transfer flavoprotein, beta-subunit	2.30	1.16E-01	40.15	16.14	3.89E+03	5.53E+03
Ta0212	electron transfer flavoprotein subunit alpha related protein	2.41	6.68E-02	37.22	12.94	5.46E+03	1.04E+04
Ta0216	hypothetical protein	0.57	2.17E-01	72.30	33.37	6.42E+03	1.10E+03
Ta0219	inositol-5-monophosphate dehydrogenase	2.65	6.33E-02	29.63	24.68	3.23E+05	1.73E+05
Ta0220	aldehyd dehydrogenase, mitochondrial 3 precursor related protein	1.97	4.64E-01	19.11	50.22	5.86E+03	7.94E+03
Ta0222	DNA polymerase II small subunit	0.97	6.28E-01	29.48	173.21	2.15E+03	7.13E+01
Ta0222a	hypothetical protein	0.55	1.18E-01	50.06	62.52	1.47E+05	4.47E+04
Ta0225	gliding motility protein related	0.57	1.84E-01	8.66	173.21	2.83E+02	1.09E+02
Ta0228	hypothetical protein	1.69	8.41E-01	95.43	43.13	1.56E+04	2.56E+04
Ta0235	hypothetical protein	0.14	6.33E-02	61.29	61.97	2.96E+04	1.69E+03
Ta0237	hypothetical protein	1.46	9.53E-01	22.17	54.36	6.87E+03	1.87E+04
Ta0238	imidazolonepropionase	0.39	1.19E-01	62.62	60.76	5.64E+02	2.83E+02
Ta0242	histidine ammonia-lyase	2.14	3.11E-01	26.09	41.28	1.10E+05	1.09E+05
Ta0245m	chorismate mutase	0.66	2.04E-01	35.84	120.31	3.89E+03	1.60E+03
Ta0246m	prephenate dehydrogenase	0.57	8.20E-02	26.66	99.21	1.26E+03	6.06E+02
Ta0255	hypothetical protein	0.49	5.39E-02	40.21	44.40	7.25E+03	1.90E+03
Ta0256	hypothetical protein	0.00	3.74E-01	173.21	NA	2.26E+02	1.94E+00
Ta0257	leucine-responsive-regulatory protein related protein	2.71	8.40E-02	12.74	33.26	4.35E+04	7.17E+04
Ta0259m	2-oxoacid--ferredoxin oxidoreductase, alpha subunit	7.09	1.59E-02	39.57	33.02	1.29E+05	7.96E+05
Ta0263	dipeptide transport system permease protein (DPPB) related protein	NA	1.00E+00	NA	NA	9.22E+01	3.25E+02
Ta0272	hypothetical membrane protein	0.00	4.06E-02	58.04	NA	8.55E+01	9.44E+01
Ta0280	hypothetical membrane protein	1.64	9.40E-01	86.68	173.21	1.13E+02	4.13E+01
Ta0282	3-phosphoshikimate 1-carboxyvinyltransferase	0.91	1.86E-01	27.05	53.87	4.48E+04	3.15E+04
Ta0283	shikimate kinase	0.65	2.08E-01	65.23	8.41	1.44E+03	1.24E+03
Ta0288m	Fe-S oxidoreductase	0.78	3.74E-01	77.64	50.16	1.19E+04	2.63E+03

Ta0290m	Predicted sterol carrier protein	4.90	3.05E-01	11.87	102.59	2.88E+03	1.28E+04
Ta0291	L-3-hydroxyacyl-CoA dehydrogenase precursor related protein	1.97	5.14E-01	26.76	55.45	5.29E+04	4.97E+04
Ta0292	acyl-CoA synthase	1.08	5.52E-01	73.93	4.74	1.18E+04	2.49E+03
Ta0295m	Acyl-CoA dehydrogenase	1.55	9.30E-01	36.34	62.46	1.09E+05	1.45E+05
Ta0297	probable DEAD box protein	AE	3.74E-01	173.21	NA	3.78E+01	0.00E+00
Ta0303m	Phosphoribosylcarboxyaminoimidazole carboxylase catalytic subunit	2.68	3.03E-01	56.14	56.81	2.27E+04	4.58E+04
Ta0309	phosphoribosylaminoimidazole-succinocarboxamidesynthase	1.19	5.25E-01	18.60	58.56	1.11E+05	1.24E+05
Ta0312	probable dCMP deaminase	0.84	1.98E-01	20.57	79.22	7.66E+03	5.05E+03
Ta0317	hypothetical protein	0.65	3.98E-01	95.78	89.86	1.20E+03	3.64E+02
Ta0324	O-sialoglycoprotein endopeptidase related protein	0.77	2.09E-01	53.13	36.22	6.60E+03	1.37E+03
Ta0327	hypothetical protein	9.50	2.03E-01	90.99	94.84	2.79E+03	8.83E+04
Ta0332	5'-methylthioadenosine phosphorylase	2.26	5.10E-01	78.15	63.48	7.79E+04	9.22E+04
Ta0335	mannose-1-phosphate guanyltransferase related protein	1.06	4.60E-01	41.99	63.26	1.33E+05	1.89E+05
Ta0337	hypothetical protein	2.76	8.20E-02	52.08	20.08	4.51E+03	1.79E+04
Ta0339m	Alpha-amylase	0.51	2.76E-02	24.25	68.30	6.81E+03	2.03E+03
Ta0340	hypothetical protein	1.38	9.42E-01	160.82	67.79	2.56E+02	4.55E+02
Ta0345	hypothetical protein	Infinity	1.20E-01	NA	87.83	1.62E+03	3.32E+03
Ta0346	hypothetical protein	1.00	1.98E-01	15.02	50.78	6.96E+03	7.65E+03
Ta0351	acetyl-CoA acetyltransferase	0.11	2.30E-01	112.76	173.21	5.29E+02	8.44E+01
Ta0363	aspartate-semialdehyde dehydrogenase	1.49	9.99E-01	88.28	56.23	4.21E+03	8.63E+03
Ta0365	hypothetical protein	5.24	8.04E-02	107.39	43.59	3.25E+02	1.19E+03
Ta0371	hypothetical protein	AE	3.74E-01	173.21	NA	1.69E+02	0.00E+00
Ta0372	hypothetical protein	0.00	1.19E-01	87.39	NA	3.63E+02	6.92E+02
Ta0375	protease synthase and sporulation negative regulatory protein (paiA) related	0.69	1.50E-01	22.21	102.98	1.13E+03	5.99E+02
Ta0375a	hypothetical protein	AN	2.46E-01	NA	127.49	0.00E+00	1.36E+03
Ta0387	hypothetical protein	0.74	5.61E-02	11.74	62.14	3.11E+04	2.59E+04
Ta0389	DNA-directed RNA polymerase subunit H	22.08	1.93E-01	94.31	103.14	1.01E+03	9.18E+03
Ta0390	DNA-directed RNA polymerase subunit beta	1.28	7.85E-01	81.17	15.41	1.10E+05	4.78E+04
Ta0394a	translation initiation factor IF-1A	1.04	5.21E-01	66.21	47.44	3.88E+04	4.26E+04
Ta0397	hypothetical protein	2.18	4.80E-01	89.27	34.13	4.25E+03	6.18E+03
Ta0400	hypothetical protein	4.10	8.79E-02	11.14	48.88	1.64E+03	5.20E+03
Ta0402	hypothetical protein	0.64	2.12E-01	66.34	21.09	2.28E+04	8.19E+03
Ta0408	hypothetical protein	0.83	1.88E-01	24.39	75.22	1.01E+04	3.72E+03
Ta0417	hypothetical protein	1.84	8.36E-01	118.66	114.45	1.88E+03	6.14E+02
Ta0429	electron transfer flavoprotein, alpha and beta subunits related protein	0.00	2.12E-01	116.78	NA	1.65E+02	7.47E+01
Ta0431	DNA-directed RNA polymerase subunit N	2.57	5.03E-01	105.37	77.55	1.33E+04	3.10E+04
Ta0433	50S ribosomal protein L13	1.33	8.23E-01	16.08	83.72	4.24E+05	3.29E+05
Ta0434	probable 50S ribosomal protein L18	0.92	4.36E-01	61.56	72.15	4.20E+05	3.82E+05
Ta0437	arsenite translocating ATPase (ASNA1) related protein	1.02	2.72E-01	26.15	50.89	8.30E+04	3.30E+04
Ta0438	hypothetical protein	1.78	7.47E-01	73.06	52.36	3.28E+04	5.20E+04
Ta0443	molybdenum cofactor biosynthesis protein A	0.66	7.37E-02	33.73	50.07	2.18E+03	1.47E+03
Ta0447	hypothetical protein	0.83	2.27E-01	40.92	64.13	5.49E+03	4.80E+03
Ta0448	hypothetical protein	AE	3.74E-01	173.21	NA	1.00E+03	0.00E+00
Ta0449	hypothetical protein	0.94	7.69E-01	173.21	173.21	2.03E+03	3.79E+03
Ta0453m	Predicted glycerate kinase	1.28	7.13E-01	42.79	52.73	4.69E+03	4.32E+03
Ta0454	N-carbamoylsarcosine amidase related protein	1.88	7.77E-01	98.16	89.69	9.48E+02	2.20E+03
Ta0455	hypothetical protein	1.73	6.78E-01	17.70	51.63	2.83E+04	6.33E+04
Ta0463	probable methylmalonyl-CoA mutase, alpha subunit, C-terminus	0.34	2.11E-01	87.05	100.36	1.31E+04	1.02E+04
Ta0471	small heat shock protein (hsp20) related protein	3.75	1.76E-01	37.78	61.73	7.64E+04	2.30E+05
Ta0481	short chain dehydrogenase	2.55	1.37E-01	24.44	36.03	1.23E+05	3.93E+05
Ta0490	hypothetical protein	1.29	6.75E-01	23.98	52.51	2.32E+04	2.07E+04

Ta0491	P-methyltransferase related protein	0.58	4.06E-01	113.41	44.02	4.83E+02	5.99E+01
Ta0492	molybdenum cofactor biosynthesis protein C	0.71	4.31E-01	91.02	107.16	4.26E+04	3.20E+03
Ta0495	threonine synthase	2.29	4.52E-01	40.62	67.41	1.55E+04	3.28E+04
Ta0501	hypothetical protein	1.70	8.25E-01	35.52	85.80	1.73E+04	2.47E+04
Ta0505	hypothetical protein	4.43	9.33E-02	24.45	51.70	8.33E+03	4.04E+04
Ta0513m	Lipoate-protein ligase A	3.57	1.69E-01	110.14	38.84	3.79E+03	6.74E+03
Ta0514	lipoate protein ligase related protein	0.56	8.38E-02	41.11	60.80	9.04E+04	4.13E+04
Ta0522	pyridoxine biosynthesis protein	2.82	2.12E-01	31.35	52.60	1.70E+05	4.03E+05
Ta0527	hypothetical protein	1.08	4.40E-01	33.96	60.02	4.98E+02	1.09E+03
Ta0529m	aspartate aminotransferase	0.39	1.02E-01	60.02	26.89	4.98E+04	1.60E+04
Ta0530	homoserine kinase	0.07	1.66E-01	97.32	86.60	1.31E+03	1.53E+02
Ta0531	probable cystathionine gamma-synthase	2.76	1.38E-01	47.22	34.48	3.09E+03	7.37E+03
Ta0532	hypothetical protein	38.76	7.68E-02	88.81	70.16	1.32E+02	1.35E+04
Ta0533	hypothetical protein	2.57	3.68E-01	49.06	65.73	4.01E+02	1.61E+03
Ta0538	conserved iron-sulfur protein	0.72	2.03E-01	57.04	27.33	3.78E+04	1.47E+04
Ta0544	hypothetical protein	0.09	1.44E-01	88.96	173.21	4.02E+03	3.89E+02
Ta0545	hypothetical protein	2.47	5.46E-01	92.66	87.45	5.49E+02	7.69E+02
Ta0546	hypothetical protein	1.13	4.82E-01	30.69	59.80	1.06E+04	8.89E+03
Ta0547	hypothetical protein	1.90	6.65E-01	26.29	77.17	5.39E+05	5.26E+05
Ta0548m	hypothetical protein	2.18	3.25E-01	49.98	34.63	1.00E+04	3.22E+04
Ta0550m	Predicted transcription regulator	0.27	6.64E-02	53.60	95.81	1.38E+03	1.83E+03
Ta0551	phenylalanyl-tRNA synthetase beta subunit	2.10	5.50E-01	59.87	64.25	3.00E+04	8.24E+03
Ta0554	hypothetical protein	0.58	3.40E-01	86.66	115.29	1.88E+03	6.22E+02
Ta0555	hypothetical protein	1.10	6.94E-01	65.24	114.49	1.50E+04	3.41E+03
Ta0558	hypothetical protein	1.80	7.40E-01	86.61	45.09	4.42E+02	2.41E+02
Ta0559	hypothetical protein	0.03	2.15E-01	115.62	173.21	9.85E+02	2.76E+02
Ta0561	hypothetical protein	1.21	7.01E-01	70.66	44.28	3.51E+04	9.46E+03
Ta0566	elongation factor 1-beta	0.68	1.24E-01	38.80	65.12	4.05E+04	9.26E+03
Ta0568	phosphomannomutase related protein	2.45	1.04E-01	23.94	28.89	7.35E+04	2.06E+05
Ta0570	hypothetical protein	0.56	1.50E-01	24.77	147.86	1.49E+03	2.76E+02
Ta0574	aspartate carbamoyltransferase regulatory subunit	1.29	7.29E-01	5.39	72.43	5.73E+04	3.07E+04
Ta0577	sarcosine oxidase related protein	2.22	3.57E-01	44.76	45.39	5.64E+03	1.29E+04
Ta0580	hypothetical protein	1.43	9.43E-01	86.93	13.61	3.30E+03	2.76E+03
Ta0588	hypothetical protein	1.16	4.44E-01	11.92	57.31	9.82E+03	9.65E+03
Ta0589	30S ribosomal protein S17	0.36	8.30E-02	52.83	87.68	8.11E+04	1.56E+04
Ta0590	hypothetical protein	Infinity	3.74E-01	NA	173.21	3.84E+02	4.66E+02
Ta0596	hypothetical protein	0.98	3.22E-01	51.44	17.85	2.36E+03	1.90E+03
Ta0600	hypothetical protein	3.33	6.40E-01	173.21	173.21	6.36E+02	2.83E+03
Ta0605	translation initiation factor IF-2 beta subunit	2.21	1.93E-01	9.06	35.67	1.61E+04	3.78E+04
Ta0606	hypothetical protein	AE	1.24E-01	89.00	NA	9.95E+01	0.00E+00
Ta0607m	Cell division GTPase	0.00	2.34E-01	123.60	NA	1.59E+04	8.26E+03
Ta0608	transcription regulator related protein	1.04	2.61E-01	8.53	56.85	1.37E+05	9.65E+04
Ta0611	inosine-5'-monophosphate dehydrogenase related protein	10.42	9.41E-02	44.63	67.60	2.83E+04	3.70E+05
Ta0615	hypothetical membrane protein	2.70	3.26E-01	21.69	68.27	2.92E+03	8.02E+03
Ta0616	transaldolase	3.41	9.60E-02	42.21	40.99	2.57E+05	3.93E+05
Ta0624	hypothetical protein	2.15	1.68E-01	31.08	22.99	1.69E+04	5.56E+04
Ta0625m	Phosphoglycerate mutase	1.54	9.49E-01	33.85	69.85	8.84E+03	1.20E+04
Ta0631m	Predicted coenzyme PQQ synthesis protein E	2.20	3.07E-01	53.03	31.76	8.46E+03	2.09E+04
Ta0632	hypothetical protein	1.61	8.63E-01	15.25	70.13	1.44E+04	2.11E+04
Ta0633	glycerol-3-phosphate dehydrogenase related protein	5.13	5.67E-02	39.03	44.87	3.10E+03	1.56E+04
Ta0634	hypothetical protein	2.46	2.56E-01	30.07	48.26	1.89E+03	3.14E+03
Ta0641m	ABC-type daunorubicin transport system, ATPase component	0.50	5.63E-01	173.21	173.21	6.06E+01	3.01E+01
Ta0649	argininosuccinate lyase	Infinity	7.63E-02	NA	72.89	4.48E+00	1.08E+04
Ta0653	S-adenosyl-L-methionine uroporphyrinogen methyltransferase related protein	0.77	2.37E-01	50.17	65.60	3.77E+04	2.62E+04

Ta0654	precorrin-8X methylmutase related protein	0.75	2.78E-01	67.59	28.53	5.13E+04	3.83E+04
Ta0663	hypothetical protein	3.36	6.88E-02	44.73	33.51	8.75E+04	2.26E+05
Ta0666	hypothetical protein	6.74	3.87E-01	97.49	137.21	9.16E+02	5.32E+04
Ta0669	tryptophan synthase subunit beta	0.48	2.70E-01	86.88	93.12	1.22E+03	1.07E+02
Ta0673	hypothetical protein	0.64	1.78E-01	58.84	34.15	1.67E+03	8.54E+02
Ta0680	hypothetical protein	3.71	4.01E-01	23.05	109.84	3.46E+04	1.10E+05
Ta0681	hypothetical protein	0.96	4.46E-01	71.41	21.71	4.82E+03	2.42E+03
Ta0686	hypothetical protein	0.83	3.57E-01	70.06	38.95	3.17E+03	7.68E+02
Ta0695	hypothetical protein	0.00	1.65E-01	102.18	NA	3.74E+02	1.47E+01
Ta0696	hypothetical protein	0.33	1.42E-01	71.36	87.52	3.43E+02	8.70E+01
Ta0701	hypothetical protein	0.52	1.26E-01	52.70	71.58	2.45E+03	3.65E+04
Ta0719	hypothetical protein	0.30	2.50E-03	19.66	28.48	1.76E+04	2.91E+03
Ta0729	hypothetical protein	0.80	2.99E-01	65.49	25.83	5.49E+03	1.90E+03
Ta0740	hypothetical protein	0.93	3.77E-01	56.89	51.46	1.83E+04	1.44E+04
Ta0742	hypothetical protein	0.59	1.40E-01	46.82	81.37	2.78E+03	1.35E+03
Ta0745	dihydrodipicoline synthase related protein	0.90	5.71E-01	95.68	97.13	2.39E+02	1.64E+02
Ta0749	sorbitol dehydrogenase related protein	9.80	6.11E-02	11.37	56.83	7.80E+02	4.55E+03
Ta0752	starvation-sensing protein rspA related protein	3.44	1.43E-01	15.89	53.55	1.78E+04	5.61E+04
Ta0754	glucose-1-dehydrogenase	4.61	6.79E-02	33.98	45.87	1.77E+04	8.57E+04
Ta0756	muconate cycloisomerase related protein	3.11	1.23E-01	6.60	46.24	1.94E+04	5.05E+04
Ta0758	DNA-dependent DNA polymerase X related protein	2.21	3.16E-01	70.83	11.03	2.16E+04	9.83E+03
Ta0759	hypothetical protein	1.92	6.86E-01	86.33	59.42	1.01E+03	4.88E+03
Ta0761	hypothetical protein	0.37	1.01E-01	60.31	44.55	8.44E+03	2.52E+03
Ta0763	hypothetical protein	1.56	9.41E-01	47.07	84.37	3.70E+03	4.24E+03
Ta0767	hypothetical protein	0.63	8.77E-02	24.25	87.43	2.81E+03	2.97E+03
Ta0774	nicotinamide-nucleotide adenylyltransferase	3.88	5.94E-02	97.19	16.52	9.40E+03	1.38E+04
Ta0776	probable glutamate dehydrogenase	0.75	1.84E-01	52.32	25.02	2.46E+04	6.60E+03
Ta0779	2-hydroxyacid dehydrogenase related protein	2.15	3.70E-01	34.98	46.72	2.43E+03	9.66E+03
Ta0780	benzoylformate decarboxylase related proten	2.35	2.46E-01	61.69	24.95	4.01E+02	3.93E+02
Ta0790m	Endonuclease III	0.36	8.21E-02	55.02	60.86	5.00E+03	2.75E+03
Ta0791	carbamoyl-phosphate synthase large subunit	1.12	4.84E-01	54.01	20.66	1.54E+04	3.82E+03
Ta0797m	Predicted thioesterase	2.68	1.85E-01	31.83	44.58	1.39E+04	4.37E+04
Ta0798	hypothetical protein	2.65	1.64E-01	59.94	28.93	3.66E+03	1.26E+04
Ta0800	hypothetical protein	3.19	5.32E-01	173.21	108.29	7.32E+02	1.07E+03
Ta0808	indole-3-glycerol-phosphate synthase	0.79	6.04E-01	113.20	173.21	6.22E+02	7.94E+02
Ta0810	probable aldehyde ferredoxin oxidoreductase	13.57	6.37E-02	86.80	59.83	3.30E+02	7.93E+03
Ta0817	hypothetical protein	0.08	1.82E-01	100.72	173.21	1.15E+03	1.05E+02
Ta0818	hypothetical protein	0.00	7.50E-02	72.41	NA	1.11E+03	7.56E+01
Ta0822	3-oxoacyl-[acyl-carrier-protein] reductase related protein	0.43	5.35E-02	44.53	31.73	1.04E+04	8.65E+03
Ta0826	hypothetical protein	2.22	4.34E-01	53.00	55.25	5.11E+03	9.77E+03
Ta0829	hypothetical protein	0.58	2.62E-01	72.27	95.42	2.91E+03	9.97E+02
Ta0830	proline iminopeptidase	1.37	8.21E-01	50.49	35.22	1.38E+04	1.49E+04
Ta0841	alcohol dehydrogenase related protein	1.80	6.45E-01	25.95	55.72	1.89E+04	3.59E+04
Ta0843	hypothetical protein	2.93	5.52E-01	173.21	97.42	1.98E+03	4.19E+03
Ta0844	molybdopterin biosynthesis protein moeB related protein	2.35	1.25E-01	35.81	23.61	1.68E+04	2.38E+04
Ta0847m	hypothetical protein	3.01	2.83E-01	13.63	70.21	9.90E+04	2.66E+05
Ta0848	hypothetical protein	1.10	4.23E-01	11.63	67.39	9.26E+03	7.80E+03
Ta0854	probable phosphoheptose isomerase	3.84	9.36E-02	25.54	47.32	4.33E+03	1.65E+04
Ta0864	small heat shock protein (hsp20) related protein	12.75	1.20E-01	33.48	77.50	8.31E+04	4.77E+05
Ta0866	thioredoxin related protein	2.08	4.06E-01	49.54	38.73	2.83E+03	8.05E+03
Ta0876	hypothetical protein	2.12	5.89E-01	26.04	85.13	5.69E+04	6.97E+04
Ta0879	isoleucyl-tRNA synthetase	2.03	1.54E-01	33.42	8.65	3.96E+04	2.17E+04
Ta0883	diphthine synthase	3.25	1.18E-01	47.30	41.81	3.93E+03	7.18E+03
Ta0885m	Predicted DNA-binding protein	4.88	1.46E-01	59.27	64.22	8.06E+04	4.50E+05
Ta0887	triacylglycerol lipase related protein	1.24	6.60E-01	52.33	37.45	7.31E+03	7.33E+03

Ta0888m	Predicted amidohydrolase	4.16	1.89E-01	30.98	69.38	5.09E+03	4.97E+04
Ta0889	hypothetical protein	4.06	9.78E-02	13.89	50.69	8.50E+04	3.03E+05
Ta0894	hypothetical protein	1.32	7.67E-01	51.19	44.00	1.73E+03	2.84E+03
Ta0904	export type III nucleotide binding protein (Vir B11) related protein	0.00	3.74E-01	173.21	NA	5.93E+01	4.55E+00
Ta0906	hypothetical protein	2.39	1.06E-01	44.23	14.73	1.81E+03	8.10E+02
Ta0911	hypothetical protein	2.07	2.33E-01	20.54	30.98	1.17E+03	8.08E+03
Ta0915	chorismate mutase/prephenate dehydratase related protein	0.29	3.00E-01	111.77	173.21	7.11E+02	4.32E+02
Ta0916	hypothetical protein	2.18	2.60E-01	26.34	37.66	4.88E+03	3.41E+04
Ta0921	hypothetical protein	NA	1.00E+00	NA	NA	3.54E+01	6.90E+02
Ta0924	hypothetical protein	1.33	7.80E-01	57.70	23.16	2.99E+04	1.09E+04
Ta0926	hypothetical protein	0.64	4.53E-02	20.96	64.52	5.65E+03	1.18E+03
Ta0927	hypothetical protein	6.65	2.92E-01	34.95	110.59	1.35E+03	1.65E+04
Ta0931	hypothetical protein	0.78	4.41E-02	4.50	53.50	1.16E+04	6.32E+03
Ta0936	hypothetical protein	1.03	6.36E-01	86.80	88.04	7.51E+02	6.84E+02
Ta0938	hypothetical protein	AE	3.74E-01	173.21	NA	2.76E+02	0.00E+00
Ta0939	hypothetical protein	0.00	3.74E-01	173.21	NA	2.80E+02	1.86E+01
Ta0940	transcription initiation factor IIB	1.57	8.96E-01	20.48	55.12	6.30E+03	4.56E+03
Ta0940a	RNA-binding protein involved in rRNA processing	1.21	7.87E-01	91.97	76.20	2.64E+03	1.83E+03
Ta0944m	GMP synthase subunit A	1.55	9.62E-01	44.64	131.95	4.17E+03	4.98E+03
Ta0945	transcription initiation factor IIB	0.85	2.48E-01	38.16	68.61	1.56E+04	5.46E+03
Ta0947	beta-hydroxybutyryl-CoA dehydrogenase related protein	2.46	1.13E-01	29.50	28.63	3.08E+03	2.34E+04
Ta0947a	hypothetical protein	0.46	8.35E-02	48.02	68.70	2.62E+04	6.61E+03
Ta0948	transcription factor	1.66	8.14E-01	28.27	63.38	7.84E+03	1.01E+04
Ta0950	hypothetical protein	1.15	5.63E-01	21.63	77.76	3.90E+04	2.46E+04
Ta0951	transcription regulator (exsB) related protein	2.12	9.19E-02	25.00	15.34	3.27E+04	5.58E+04
Ta0955	delta-aminolevulinic acid dehydratase	1.38	8.28E-01	15.17	58.94	1.99E+04	2.62E+04
Ta0956	hypothetical protein	0.78	7.61E-02	4.96	64.99	5.29E+04	4.46E+04
Ta0957	hypothetical protein	0.73	3.83E-01	89.12	28.80	2.89E+03	1.51E+03
Ta0958	hypothetical protein	2.10	5.01E-01	85.83	29.15	9.75E+03	9.45E+03
Ta0965	NADH dehydrogenase subunit I	0.60	3.52E-01	86.61	116.33	2.17E+03	8.84E+02
Ta0968	NADH dehydrogenase subunit C	2.58	3.22E-01	96.04	33.24	1.44E+03	1.15E+04
Ta0969m	NADH dehydrogenase beta subunit	1.29	7.07E-01	50.80	33.75	3.57E+03	5.33E+03
Ta0973	hypothetical protein	0.59	5.14E-02	27.56	65.68	2.21E+03	1.03E+03
Ta0974	probable acyl-CoA dehydrogenase	0.63	1.47E-01	55.16	19.92	6.58E+04	3.28E+04
Ta0979	5, 10-methylenetetrahydrofolate reductase related protein	0.41	1.17E-01	62.80	26.37	1.22E+04	4.31E+03
Ta0981	transcription regulator LrpA related protein	1.09	6.71E-01	87.14	73.79	3.11E+04	3.98E+04
Ta0989	hypothetical protein	0.56	1.21E-01	48.48	67.75	3.32E+03	2.04E+03
Ta0990	hypothetical protein	0.81	2.28E-01	48.48	52.19	2.41E+03	1.04E+03
Ta0991	hypothetical protein	0.55	2.91E-02	24.97	57.72	2.93E+04	1.87E+04
Ta0997	N(2),N(2)-dimethylguanosine tRNA methyltransferase	0.90	1.33E-01	19.22	50.86	4.63E+03	3.29E+03
Ta0999m	hypothetical protein	1.13	6.53E-01	41.69	98.03	1.13E+05	8.25E+04
Ta1002	succinate dehydrogenase	0.85	3.30E-01	64.63	35.75	3.18E+04	1.56E+04
Ta1005	hypothetical protein	0.19	5.35E-02	55.24	62.69	2.08E+04	2.59E+03
Ta1006	DNA-directed RNA polymerase subunit M	2.06	4.34E-01	45.02	43.82	3.44E+03	7.69E+03
Ta1007	hypothetical membrane protein	0.85	2.91E-01	57.57	38.19	8.81E+03	2.37E+03
Ta1009	hypothetical protein	0.77	5.05E-01	96.81	120.78	8.82E+02	3.80E+02
Ta1010m	Predicted rRNA methylase	1.19	7.14E-01	51.56	90.22	6.34E+03	9.14E+03
Ta1012	indolepyruvate ferredoxin oxidoreductase, alpha subunit related protein	0.58	8.41E-02	34.33	80.58	1.09E+04	9.42E+02
Ta1017	hypothetical protein	0.50	1.10E-01	55.80	27.51	5.12E+03	1.99E+03
Ta1018	serine-glyoxylate aminotransferase related protein	2.36	2.93E-01	12.63	52.07	3.17E+04	6.49E+04
Ta1019	MoaD (involved in molybdopterin synthesis) related protein	1.54	9.44E-01	31.29	61.04	3.13E+03	1.03E+04
Ta1023	molybdopterin biosynthesis protein (moeA-	1.72	7.52E-01	19.75	65.03	9.23E+03	7.47E+03

	2) related protein						
Ta1027	ABC transporter, ATP-binding protein related protein	1.32	8.21E-01	79.98	23.03	3.95E+03	1.00E+04
Ta1031	30S ribosomal protein S11	3.20	4.33E-01	11.51	106.13	7.23E+04	2.75E+05
Ta1033	30S ribosomal protein S13	1.43	9.17E-01	52.91	37.68	1.25E+05	9.07E+04
Ta1037	proline dipeptidase related protein	2.01	1.49E-01	13.78	22.93	1.00E+05	2.04E+05
Ta1040	30S ribosomal protein S8	1.37	8.90E-01	91.66	29.85	5.51E+04	9.43E+04
Ta1041m	hypothetical protein	2.67	3.65E-01	16.67	74.27	4.27E+03	4.47E+04
Ta1044m	phosphoribosylformylglycinamidine synthase	0.99	4.47E-01	52.68	68.56	6.78E+04	5.03E+04
Ta1046	probable chaperone (csaA)	2.48	3.86E-01	43.50	65.81	5.44E+03	8.31E+03
Ta1047a	hypothetical protein	40.10	1.37E-01	173.21	89.62	1.62E+02	7.48E+03
Ta1050	transport-ATP binding protein related protein	0.65	3.96E-01	86.84	127.73	2.59E+02	6.78E+02
Ta1052	hypothetical protein	3.66	3.71E-01	116.95	90.22	4.28E+02	1.49E+03
Ta1054	DNA gyrase subunit A	2.22	5.49E-02	4.98	20.91	2.41E+04	1.93E+04
Ta1056	hypothetical protein	0.55	9.84E-02	45.48	61.95	9.46E+04	6.92E+03
Ta1061m	Predicted transcription regulator	1.35	8.51E-01	27.33	86.62	4.10E+02	6.59E+02
Ta1065	hypothetical protein	1.03	4.67E-01	53.93	58.11	7.59E+03	7.50E+03
Ta1068	hypothetical protein	1.60	8.81E-01	15.09	70.25	2.56E+03	4.81E+03
Ta1072	ABC transporter ATP-binding protein related protein	0.00	1.36E-01	92.94	NA	2.88E+02	3.98E+01
Ta1074	glucose-1-phosphate thymidylyltransferase related protein	1.06	4.87E-01	41.32	70.67	1.06E+04	1.28E+04
Ta1076m	prefoldin subunit alpha	0.54	6.31E-02	36.34	65.47	7.61E+04	4.30E+04
Ta1081	ATP-dependent proteinase La (Lon) related protein	0.52	1.45E-02	13.31	68.46	3.68E+03	5.82E+02
Ta1082m	hypothetical protein	0.54	3.50E-01	99.09	87.18	6.48E+02	6.20E+02
Ta1084	hypothetical protein	4.04	1.13E-01	29.65	52.95	1.28E+04	3.17E+04
Ta1086	heat shock protein GrpE related protein	1.75	6.87E-01	50.44	39.35	4.78E+03	3.33E+03
Ta1087	molecular chaperone DnaK	3.29	7.00E-02	21.47	37.29	5.08E+04	1.56E+05
Ta1090	DNA-directed RNA polymerase subunit E'	6.25	1.19E-01	86.70	63.47	3.01E+03	2.18E+04
Ta1091	hypothetical protein	2.26	2.25E-01	15.31	39.90	1.70E+04	2.61E+04
Ta1093	30S ribosomal protein S27ae	Infinity	1.20E-01	NA	87.75	5.16E+03	9.17E+03
Ta1093a	hypothetical protein	AE	6.01E-02	66.63	NA	2.31E+04	0.00E+00
Ta1095	hypothetical protein	NA	1.00E+00	NA	NA	3.11E+01	4.41E+01
Ta1096	probable glycerol kinase	0.80	3.36E-01	72.79	25.45	2.94E+03	2.24E+02
Ta1098	cell division protein pelota related protein	0.49	5.83E-02	39.61	58.45	9.87E+03	4.43E+03
Ta1099	hypothetical protein	0.79	2.04E-01	5.76	101.63	5.55E+02	1.44E+03
Ta1100	3-octaprenyl-4-hydroxybenzoate carboxy-lyase	2.54	2.89E-01	84.20	31.52	1.02E+04	2.83E+04
Ta1110	transcription regulator ArsR related protein	2.38	4.63E-01	86.60	58.24	8.88E+03	1.46E+04
Ta1114	50S ribosomal protein L24	1.43	9.08E-01	56.12	26.97	1.95E+05	1.09E+05
Ta1115	30S ribosomal protein S28e	0.81	1.84E-01	32.70	67.78	3.07E+04	1.94E+04
Ta1116	50S ribosomal protein L7Ae	1.46	9.57E-01	18.93	55.15	3.01E+05	3.52E+05
Ta1119	hypothetical protein	0.46	5.97E-02	28.17	119.20	7.91E+02	1.61E+02
Ta1123	hypothetical protein	1.26	6.63E-01	16.18	65.98	4.22E+03	4.06E+03
Ta1124m	ribosome biogenesis protein	4.28	8.46E-02	48.04	46.51	1.67E+03	1.89E+04
Ta1125	hypothetical protein	0.00	1.21E-01	88.25	NA	5.35E+02	1.12E+02
Ta1134	hypothetical protein	NA	1.00E+00	NA	NA	1.75E+02	9.27E+01
Ta1136	hypothetical protein	Infinity	1.16E-01	NA	86.63	4.94E+00	2.52E+01
Ta1138	hypothetical protein	1.27	6.72E-01	9.53	66.03	6.94E+03	6.38E+03
Ta1139	Aut related protein	2.67	5.37E-01	20.80	112.80	9.34E+03	1.86E+04
Ta1140	N-terminal acetyltransferase related protein	1.98	7.32E-01	90.29	94.44	2.41E+03	3.60E+03
Ta1144	hypothetical protein	NA	1.00E+00	NA	NA	1.86E+02	1.03E+03
Ta1148	DNA ligase	0.79	6.50E-02	10.56	58.23	3.57E+04	3.93E+04
Ta1150a	Small primase-like proteins (Toprim domain)	0.12	2.63E-01	121.22	173.21	7.56E+02	1.18E+02
Ta1168	probable adenine specific DNA	0.00	1.19E-01	87.66	NA	1.76E+02	2.96E+01

	methyltransferase						
Ta1170;Ta1414	hypothetical protein	19.51	5.50E-02	41.68	59.52	4.50E+04	7.25E+05
Ta1174	hypothetical protein	1.64	8.61E-01	27.01	82.69	1.55E+04	2.84E+04
Ta1178m	ABC-type nitrate transporter, ATPase component	2.53	5.50E-01	173.21	37.39	1.14E+03	1.64E+03
Ta1183	hypothetical protein	4.08	3.34E-01	25.17	99.63	1.13E+03	8.73E+03
Ta1189	8-amino-7-oxononanoate synthase	0.96	4.11E-01	53.09	62.48	6.51E+04	5.37E+04
Ta1192	hypothetical GTP-binding protein	1.05	2.36E-01	4.54	51.25	6.50E+04	8.16E+04
Ta1199	4-hydroxybenzoate decarboxylase related protein	0.67	2.04E-01	62.52	13.35	3.10E+04	2.35E+04
Ta1200	probable cobyrinic acid a, c-diamide synthase (cbiA)	0.33	8.23E-02	57.03	58.15	6.03E+03	2.27E+03
Ta1204	30S ribosomal protein S27	1.31	8.10E-01	42.95	77.01	2.23E+04	2.67E+04
Ta1205	50S ribosomal protein L44e	1.95	6.42E-01	45.65	73.59	4.06E+03	6.08E+03
Ta1209m	Trehalose-6-phosphatase	1.31	7.67E-01	39.32	59.42	3.39E+03	3.95E+03
Ta1217	hypothetical protein	2.27	5.40E-01	60.54	79.12	7.27E+04	8.07E+04
Ta1218	transcription regulator LrpA related protein	1.05	3.12E-01	20.02	57.00	6.44E+03	7.79E+03
Ta1223a	Sec-independent protein secretion pathway component	0.07	1.14E-02	37.26	12.76	1.59E+05	4.14E+03
Ta1229m	Rieske Fe-S protein	0.00	1.45E-01	95.83	NA	2.67E+02	1.97E+01
Ta1230	hypothetical protein	2.65	4.96E-01	91.64	87.09	4.13E+02	7.42E+02
Ta1240	small nuclear ribonucleoprotein	3.15	1.68E-01	8.10	54.07	1.91E+04	3.78E+04
Ta1242	amidophosphoribosyltransferase	1.70	7.90E-01	69.75	40.32	1.05E+04	1.43E+04
Ta1245	cytidylate kinase	1.88	6.33E-01	46.39	59.27	7.19E+03	8.10E+03
Ta1250m	50S ribosomal protein L30	1.17	5.22E-01	9.94	65.39	3.66E+05	2.64E+05
Ta1253	50S ribosomal protein L19e	6.21	1.75E-01	43.10	79.15	6.27E+03	1.13E+05
Ta1254	50S ribosomal protein L32e	0.60	1.78E-02	15.65	52.97	9.26E+03	9.33E+03
Ta1255	50S ribosomal protein L6	1.26	6.85E-01	54.38	32.14	2.03E+05	1.32E+05
Ta1256	30S ribosomal protein S8	1.25	5.83E-01	13.89	52.82	1.91E+05	3.44E+05
Ta1258	50S ribosomal protein L5	1.60	8.50E-01	1.09	58.08	5.98E+04	2.03E+05
Ta1259	30S ribosomal protein S4	1.26	6.45E-01	33.20	51.67	2.37E+05	2.50E+05
Ta1259a	50S ribosomal protein L24	1.10	6.71E-01	84.52	68.34	1.17E+05	7.58E+04
Ta1261	50S ribosomal protein L14	0.80	8.14E-02	19.08	53.97	2.50E+05	1.16E+05
Ta1263b	translation initiation factor IF-1	2.85	4.25E-01	7.64	92.75	6.16E+04	6.62E+04
Ta1264a	50S ribosomal protein L29	11.35	7.10E-02	46.52	61.30	3.74E+03	8.25E+04
Ta1267	30S ribosomal protein S19P	1.33	8.30E-01	25.77	87.09	1.71E+05	1.22E+05
Ta1271	50S ribosomal protein L3	1.60	8.50E-01	26.84	50.40	1.70E+05	2.70E+05
Ta1276	thermosome beta chain	2.01	2.00E-01	27.40	20.80	1.75E+06	1.37E+06
Ta1277	thiamine-phosphate pyrophosphorylase	1.27	7.66E-01	38.26	81.27	7.82E+03	7.97E+03
Ta1284	fragment of probable methionine--tRNA ligase	1.12	4.71E-01	16.08	69.07	2.60E+05	1.94E+05
Ta1285	replication factor C large subunit	0.49	9.10E-02	49.48	54.82	3.16E+04	9.11E+03
Ta1286m	50S ribosomal protein L15	2.59	2.42E-01	11.21	53.24	3.47E+04	8.73E+04
Ta1288	proteasome subunit alpha	1.86	6.48E-01	31.91	65.57	6.33E+05	9.29E+05
Ta1295	50S ribosomal protein L37	4.63	1.89E-01	58.87	71.75	1.35E+05	1.34E+05
Ta1297	50S ribosomal protein L21	0.87	4.39E-01	70.38	79.42	2.06E+05	1.53E+05
Ta1299	hypothetical protein	1.00	4.11E-01	59.47	26.46	3.59E+04	2.54E+04
Ta1306	hypothetical protein	1.58	9.44E-01	44.33	117.29	5.55E+02	1.05E+03
Ta1313	geranylgeranyl pyrophosphate synthase related protein	2.02	2.16E-01	18.78	27.79	1.62E+04	1.70E+04
Ta1318m	phosphoribosylformylglycinamidine synthase subunit I	1.17	6.33E-01	45.92	69.77	3.18E+04	2.59E+04
Ta1320	hypothetical protein	0.84	7.25E-01	173.21	173.21	4.74E+02	2.58E+02
Ta1325	probable oligopeptide ABC transport system, ATP-binding protein appD	0.00	1.16E-01	86.61	NA	7.16E+02	5.40E+00
Ta1339	cell-cycle regulation histidine triad (HIT) related protein	1.04	4.34E-01	55.72	31.07	6.32E+03	9.17E+03
Ta1348	hypothetical protein	3.11	1.94E-01	36.05	55.12	1.64E+03	2.88E+03
Ta1352	hypothetical protein	3.34	3.30E-01	47.27	84.02	5.05E+03	1.91E+04
Ta1353m	hypothetical protein	0.97	4.38E-01	63.64	46.75	1.10E+05	8.07E+04
Ta1364	hypothetical protein	0.93	3.91E-01	60.81	46.17	1.59E+03	1.82E+03

Ta1365	hypothetical protein	2.69	3.21E-01	47.77	63.04	8.53E+03	5.42E+03
Ta1366m	glycine cleavage system protein H	1.79	7.02E-01	14.55	70.06	3.92E+04	1.37E+05
Ta1370	hypothetical protein	1.51	9.84E-01	47.39	86.02	2.89E+04	4.10E+04
Ta1376	hypothetical protein	2.23	5.61E-01	94.33	65.18	5.90E+02	5.68E+02
Ta1380	hypothetical protein	0.00	3.74E-01	173.21	NA	1.54E+01	7.91E+01
Ta1392	hypothetical protein	7.21	1.63E-01	32.12	80.21	1.13E+04	9.83E+04
Ta1403	thermopsin precursor related protein	AE	4.39E-02	59.61	NA	1.15E+02	0.00E+00
Ta1410	ferrichrome ABC transporter (ATP-binding p) related protein	0.71	6.59E-01	173.21	173.21	5.55E+02	6.03E+02
Ta1415	hypothetical protein	6.20	1.13E-01	91.87	61.15	1.72E+05	1.26E+06
Ta1416	DNA-directed RNA polymerase subunit L	0.99	3.38E-01	32.92	63.47	1.80E+05	1.22E+05
Ta1419	hypothetical protein	1.94	5.20E-01	31.16	51.37	1.79E+05	2.76E+05
Ta1422	hypothetical protein	0.94	7.70E-01	173.21	173.21	3.85E+03	1.60E+03
Ta1429a	hypothetical protein	7.13	2.05E-01	16.59	90.59	4.06E+04	2.34E+05
Ta1431	hypothetical protein	0.00	1.77E-01	105.85	NA	2.30E+02	5.34E+01
Ta1434	hypothetical protein	1.12	3.94E-01	25.44	50.37	4.57E+04	3.95E+04
Ta1437	probable 3-methyl-2-oxobutanoate dehydrogenase chain E1-beta	0.53	2.14E-02	22.15	57.94	4.99E+04	2.39E+04
Ta1440	hypothetical protein	2.26	2.62E-01	45.27	33.77	1.33E+04	2.54E+04
Ta1441	hypothetical protein	3.19	3.14E-01	39.30	77.95	6.74E+03	2.97E+04
Ta1442	hypothetical protein	0.00	3.74E-01	173.21	NA	3.92E+02	8.52E+02
Ta1446	trna nucleotidyltransferase related protein	0.56	1.02E-01	45.09	64.57	1.79E+04	9.33E+03
Ta1450	hypothetical protein	2.57	6.64E-01	173.21	117.95	2.47E+02	9.03E+02
Ta1458	ribonuclease HII	2.81	7.38E-02	23.38	31.34	6.56E+03	3.07E+03
Ta1465	acyl-CoA dehydrogenase related protein	0.91	7.56E-01	173.21	173.21	9.91E+01	2.28E+02
Ta1467	hypothetical protein	1.28	7.86E-01	80.97	31.77	1.09E+03	2.93E+03
Ta1468	hypothetical protein	5.49	5.08E-02	57.24	42.93	1.04E+04	8.23E+03
Ta1478	probable formate-tetrahydrofolate ligase	4.33	5.57E-02	56.95	37.72	1.39E+05	1.76E+05
Ta1479m	ATPase involved in chromosome partitioning	2.66	2.63E-01	45.04	52.73	4.73E+04	1.08E+05
Ta1480	molybdopterin-synthase large subunit related protein	2.38	2.24E-01	21.55	42.95	1.07E+04	3.52E+04
Ta1495	geranylgeranyl reductase related protein	0.24	1.96E-01	93.84	34.00	2.03E+03	2.23E+04
Ta1501	RNA helicase (RIG-I) related protein	0.76	3.23E-01	75.83	5.92	2.92E+04	2.98E+03
Ta1507	hypothetical protein	0.00	2.25E-01	120.77	NA	3.01E+04	1.63E+01
Ta1508m	Acyl-coenzyme A synthetase	0.58	3.24E-01	90.44	68.63	3.93E+03	1.42E+02

Table S3: Quantitative transcriptomics data of 445 regulated mRNAs. The gene IDs, annotations, changes in mRNA expression level (anaerobic/aerobic), p-values of the Student's t-test, mRNA levels under aerobic conditions and mRNA levels under anaerobic conditions were listed.

Gene ID	***Annotation***	***Ratio$_{anae/ae}$***	***p-value***	***mRNA level (aerob)***	***mRNA level (anaerob)***
Down-regulated cytosolic protein-encoding mRNAs (anaerobically)					
Ta0001	V-type ATP synthase subunit E	0.47	7.95E-04	13.75	12.65
Ta0002	V-type ATP synthase subunit C	0.40	1.03E-04	13.73	12.42
Ta0003	V-type ATP synthase subunit F	0.41	1.35E-04	14.12	12.82
Ta0004	V-type ATP synthase subunit A	0.50	4.84E-05	14.03	13.02
Ta0006	V-type ATP synthase subunit D	0.50	1.17E-04	14.07	13.06
Ta0014m	N-glycosylase/DNA lyase	0.19	2.20E-05	13.72	11.29
Ta0016	hypothetical protein	0.50	1.55E-03	9.88	8.87
Ta0018	hypothetical protein	0.25	6.45E-05	13.17	11.16
Ta0023m	hypothetical protein	0.47	4.24E-04	12.49	11.39
Ta0024	hypothetical protein	0.29	2.17E-05	14.46	12.69
Ta0029	hypothetical protein	0.37	5.68E-05	12.08	10.65
Ta0038	72K mitochondrial outer membrane protein related protein	0.24	6.70E-04	14.41	12.34
Ta0042	hypothetical protein	0.49	1.50E-04	12.91	11.89
Ta0055m	translation initiation factor IF-6	0.38	1.57E-04	13.38	11.98
Ta0058	N-terminal acetyltransferase complex ard1 subunit related protein	0.43	1.26E-04	11.09	9.87
Ta0078	hypothetical protein	0.50	3.56E-04	15.10	14.08
Ta0081	thymidylate kinase	0.44	1.26E-04	12.57	11.37
Ta0083	hypothetical protein	0.29	1.09E-03	12.70	10.93
Ta0084	hypothetical protein	0.37	5.78E-05	12.83	11.40
Ta0088m	ABC-type multidrug transport system, ATPase component	0.19	5.25E-05	13.37	10.94
Ta0095	hypothetical protein	0.48	5.14E-04	13.62	12.56
Ta0098	hypothetical protein	0.45	4.87E-05	15.25	14.10
Ta0099	histidyl-tRNA synthetase	0.30	5.76E-05	14.17	12.43
Ta0100	hypothetical protein	0.34	3.30E-04	12.24	10.68
Ta0101m	C4-type Zn finger protein	0.42	1.64E-04	14.03	12.76
Ta0102	isopentenyl pyrophosphate isomerase	0.47	1.52E-04	13.38	12.28
Ta0112	aconitate hydratase	0.47	8.83E-04	14.60	13.51
Ta0118m	hypothetical protein	0.40	2.99E-04	12.21	10.89
Ta0122	xylulose kinase related protein	0.47	1.51E-04	13.98	12.89
Ta0132	hypothetical protein	0.14	4.90E-05	13.93	11.04
Ta0133m	Pirin-related protein	0.14	9.84E-06	13.05	10.18
Ta0135	hypothetical protein	0.14	1.85E-05	11.87	8.99
Ta0136	nitrate reductase related protein	0.23	4.19E-05	10.77	8.67
Ta0139	pyrroline-5-carboxylate reductase	0.39	2.46E-04	10.66	9.28
Ta0140	hypothetical protein	0.20	1.83E-04	10.57	8.25
Ta0143	probable ATP-binding protein	0.41	4.04E-04	12.19	10.91
Ta0147	hypothetical protein	0.26	7.51E-05	10.57	8.63
Ta0152	probable peroxiredoxin	0.08	7.03E-05	13.19	9.53
Ta0157	chromosome segregation protein	0.50	6.29E-05	14.32	13.32
Ta0162m	NADH:flavin oxidoreductase	0.44	1.02E-04	14.56	13.38
Ta0168	hypothetical protein	0.25	9.41E-05	11.50	9.49
Ta0193m	Predicted phosphohydrolase (HD superfamily)	0.31	4.80E-05	13.90	12.23
Ta0194m	Predicted hydrolase (alpha/beta superfamily)	0.31	8.84E-06	12.41	10.71
Ta0205	2, 4-dienoyl-CoA reductase (NADPH) precursor related protein	0.36	1.60E-04	12.13	10.67
Ta0214	tRNA pseudouridine synthase D	0.45	1.79E-04	14.47	13.32
Ta0222	DNA polymerase II small subunit	0.40	5.65E-05	14.91	13.57
Ta0222a	hypothetical protein	0.35	7.43E-04	12.89	11.37
Ta0228	hypothetical protein	0.21	4.17E-04	12.56	10.32

Ta0229	acetyl-CoA synthetase related protein	0.16	4.19E-05	13.06	10.38
Ta0230	probable isovaleryl-CoA dehydrogenase	0.31	2.65E-05	10.97	9.26
Ta0241	hypothetical protein	0.07	3.55E-06	12.62	8.68
Ta0245m	chorismate mutase	0.43	1.60E-02	11.98	10.77
Ta0246m	prephenate dehydrogenase	0.37	1.56E-03	14.22	12.77
Ta0247	3-deoxy-7-phosphoheptulonate synthase	0.36	1.29E-04	14.27	12.81
Ta0250	glucose-fructose oxidoreductase related protein	0.41	4.88E-05	12.89	11.60
Ta0255	hypothetical protein	0.29	3.19E-04	13.36	11.58
Ta0256	hypothetical protein	0.26	1.44E-04	13.86	11.94
Ta0265	probable sporulation protein (spo0KD)	0.31	2.54E-04	14.84	13.14
Ta0266	oligopeptide transport ATP-binding protein appF related protein	0.28	2.30E-04	14.13	12.27
Ta0267	hypothetical protein	0.22	4.00E-04	11.72	9.53
Ta0276	molybdenum cofactor synthesis-step 1 (MOCS1) related protein	0.42	1.87E-04	13.24	12.00
Ta0286m	hypothetical protein	0.29	4.83E-05	14.36	12.57
Ta0297	probable DEAD box protein	0.16	5.94E-05	12.17	9.54
Ta0298	alpha-glucosidase related protein	0.44	1.79E-04	10.37	9.19
Ta0302	sua5 related protein	0.44	1.24E-04	14.37	13.19
Ta0305	hypothetical protein	0.37	6.35E-05	13.14	11.70
Ta0317	hypothetical protein	0.48	2.71E-04	13.46	12.41
Ta0318m	Adenine-specific DNA methylase	0.50	1.64E-04	12.30	11.30
Ta0347	hypothetical protein	0.35	1.31E-04	12.43	10.90
Ta0351	acetyl-CoA acetyltransferase	0.43	7.57E-04	11.04	9.82
Ta0357	thiol-specific antioxidant related protein	0.32	3.19E-04	12.99	11.35
Ta0371	hypothetical protein	0.08	2.70E-05	11.97	8.38
Ta0372	hypothetical protein	0.25	8.38E-05	10.34	8.31
Ta0386	hypothetical protein	0.50	5.17E-04	9.94	8.94
Ta0394a	translation initiation factor IF-1A	0.46	2.44E-04	14.06	12.93
Ta0396m	Predicted serine/threonine protein kinase	0.36	4.09E-05	12.90	11.42
Ta0402	hypothetical protein	0.44	8.70E-05	13.69	12.49
Ta0408	hypothetical protein	0.24	2.67E-04	13.04	10.96
Ta0419	hypothetical protein	0.46	6.73E-04	11.08	9.97
Ta0421	probable acetyl-coenzyme-A synthetase	0.19	1.25E-04	11.69	9.31
Ta0423	hypothetical protein	0.23	1.03E-04	15.03	12.89
Ta0424	hypothetical protein	0.30	1.66E-04	15.49	13.77
Ta0425	formate dehydrogenase related protein	0.41	1.59E-04	15.68	14.41
Ta0428	biphenyl dioxygenase, Rieske iron-sulfur component related protein	0.49	3.80E-04	9.73	8.69
Ta0429	electron transfer flavoprotein, alpha and beta subunits related protein	0.35	4.87E-05	10.12	8.60
Ta0443	molybdenum cofactor biosynthesis protein A	0.40	3.73E-04	13.25	11.94
Ta0448	hypothetical protein	0.15	4.89E-05	11.58	8.87
Ta0449	hypothetical protein	0.10	2.65E-04	12.96	9.67
Ta0450	DNA polymerase II	0.20	6.92E-04	12.85	10.51
Ta0461	hypothetical protein	0.41	1.62E-04	11.12	9.85
Ta0462	probable methylmalonyl-CoA mutase, alpha subunit, N-terminus	0.40	1.09E-03	13.37	12.05
Ta0463	probable methylmalonyl-CoA mutase, alpha subunit, C-terminus	0.34	5.29E-03	11.89	10.35
Ta0464	hypothetical protein	0.30	4.15E-04	13.61	11.85
Ta0479	glucosamine-fructose-6-phosphate aminotransferase related protein	0.40	1.17E-04	11.86	10.55
Ta0489	hypothetical protein	0.31	1.71E-04	12.09	10.40
Ta0491	P-methyltransferase related protein	0.25	1.45E-03	14.11	12.09
Ta0524	hypothetical protein	0.49	4.92E-04	13.00	11.96
Ta0536	glutamyl-tRNA reductase	0.37	1.87E-04	13.92	12.51
Ta0537m	Predicted ICC-like phosphoesterase	0.30	3.01E-04	11.84	10.09
Ta0576	protoporphyrin IX magnesium chelatase related protein	0.47	4.43E-04	13.15	12.05
Ta0643	hypothetical protein	0.35	6.76E-03	12.44	10.93
Ta0652	hypothetical protein	0.39	1.31E-02	13.59	12.22
Ta0653	S-adenosyl-L-methionine uroporphyrinogen methyltransferase related protein	0.37	9.04E-03	14.09	12.65

Ta0654	precorrin-8X methylmutase related protein	0.35	4.86E-03	14.78	13.25
Ta0655	probable cobalamin biosynthesis precorrin-3 methylase	0.45	2.69E-03	15.42	14.27
Ta0656	cobalt-precorrin-6A synthase	0.33	3.51E-03	14.47	12.88
Ta0657	precorrin-8W decarboxylase related protein	0.34	4.98E-03	14.40	12.84
Ta0658	precorrin-2 methyltransferase related protein	0.37	8.28E-03	14.94	13.49
Ta0659	probable precorrin-4 methylase	0.42	9.57E-03	14.92	13.67
Ta0660	hypothetical protein	0.36	4.20E-03	14.97	13.49
Ta0669	tryptophan synthase subunit beta	0.08	5.74E-05	14.36	10.71
Ta0747	glucose 1-dehydrogenase related protein	0.48	1.79E-04	11.46	10.39
Ta0790m	Endonuclease III	0.23	3.34E-04	12.14	10.01
Ta0791	carbamoyl-phosphate synthase large subunit	0.49	3.97E-03	12.24	11.22
Ta0795	sugar fermentation stimulation protein	0.44	6.45E-04	12.15	10.96
Ta0803m	Tryptophan synthase alpha chain	0.02	4.16E-05	14.95	9.01
Ta0804	anthranilate phosphoribosyltransferase	0.02	5.33E-05	14.59	8.59
Ta0805m	N-(5'-phosphoribosyl)anthranilate isomerase	0.02	4.18E-05	15.07	9.44
Ta0806m	anthranilate synthase component I	0.02	2.83E-05	15.17	9.28
Ta0807	anthranilate synthase component II	0.02	2.68E-05	15.30	9.91
Ta0808	indole-3-glycerol-phosphate synthase	0.02	3.99E-05	14.67	9.08
Ta0816	pyruvate dehydrogenase	0.31	2.03E-04	13.15	11.47
Ta0817	hypothetical protein	0.17	1.89E-04	11.73	9.17
Ta0818	hypothetical protein	0.19	7.64E-04	11.43	9.02
Ta0819	citrate synthase	0.12	2.25E-04	12.18	9.17
Ta0822	3-oxoacyl-[acyl-carrier-protein] reductase related protein	0.40	2.06E-04	10.83	9.52
Ta0857	ABC transporter cdd4-like protein ScnF related protein	0.29	1.33E-04	11.53	9.74
Ta0891	endonuclease IV related protein	0.46	5.26E-04	13.81	12.68
Ta0892	hypothetical protein	0.28	4.18E-05	13.65	11.82
Ta0904	export type III nucleotide binding protein (Vir B11) related protein	0.49	2.36E-04	10.87	9.84
Ta0913	hypothetical protein	0.21	1.79E-04	11.84	9.60
Ta0926	hypothetical protein	0.35	4.00E-05	11.60	10.10
Ta0947a	hypothetical protein	0.33	1.57E-03	12.93	11.32
Ta0975m	DNA primase small subunit	0.31	2.55E-05	12.83	11.16
Ta0994	gamma-glutamyltransferase related protein	0.18	9.39E-05	13.74	11.27
Ta1014m	GTP cyclohydrolase II	0.31	2.12E-04	12.36	10.69
Ta1020	transcription initiation factor IIB related protein	0.33	2.30E-04	10.98	9.39
Ta1021	hypothetical protein	0.46	1.29E-04	13.99	12.88
Ta1065	hypothetical protein	0.47	9.67E-04	14.01	12.94
Ta1089m	DNA-directed RNA polymerase subunit E'	0.43	1.92E-04	13.75	12.54
Ta1093a	hypothetical protein	0.17	4.92E-05	12.42	9.89
Ta1098	cell division protein pelota related protein	0.41	1.93E-04	13.32	12.04
Ta1100	3-octaprenyl-4-hydroxybenzoate carboxy-lyase	0.42	3.41E-04	12.45	11.21
Ta1109	hypothetical protein	0.22	2.67E-04	12.09	9.93
Ta1110	transcription regulator ArsR related protein	0.47	1.28E-04	11.36	10.25
Ta1127	hypothetical protein	0.42	2.75E-05	14.00	12.74
Ta1128m	hypothetical protein	0.48	2.44E-04	12.56	11.51
Ta1129	sulfide-quinone reductase related protein	0.15	1.07E-04	13.72	10.95
Ta1144a	hypothetical protein	0.40	1.24E-03	12.25	10.91
Ta1152	hypothetical protein	0.20	3.26E-05	14.07	11.72
Ta1162	methionyl-tRNA synthetase	0.41	4.52E-05	14.13	12.84
Ta1178m	ABC-type nitrate transporter, ATPase component	0.49	6.91E-05	13.42	12.41
Ta1181	hypothetical protein	0.46	2.54E-04	13.11	11.98
Ta1191	tRNA splicing endonuclease	0.27	1.03E-04	13.50	11.60
Ta1200	probable cobyrinic acid a, c-diamide synthase (cbiA)	0.29	6.88E-05	13.55	11.77
Ta1229m	Rieske Fe-S protein	0.42	4.75E-03	12.03	10.77
Ta1232	enoyl-CoA hydratase related protein	0.32	3.73E-04	11.40	9.74
Ta1238	hypothetical protein	0.45	2.20E-03	11.37	10.21
Ta1243m	Nop56p-related protein (ribosomal biogenesis)	0.33	1.16E-04	12.19	10.59

Ta1244	tRNA pseudouridine synthase B	0.31	5.23E-05	13.72	12.04
Ta1245	cytidylate kinase	0.45	5.82E-04	14.39	13.24
Ta1275	putative target YPL207w of the HAP2 transcriptional complex related protein	0.26	1.29E-04	14.48	12.51
Ta1296	hypothetical protein	0.49	1.07E-04	14.16	13.14
Ta1299	hypothetical protein	0.45	1.42E-04	13.01	11.86
Ta1300m	rRNA dimethyladenosine transferase	0.26	1.02E-04	13.39	11.42
Ta1307	hypothetical protein	0.29	5.84E-05	14.09	12.30
Ta1312	hypothetical protein	0.39	9.82E-04	12.83	11.46
Ta1313	geranylgeranyl pyrophosphate synthase related protein	0.49	5.57E-04	12.95	11.93
Ta1316	alcohol dehydrogenase related protein	0.45	4.76E-04	14.35	13.18
Ta1319	dihydroorotase	0.38	5.82E-05	13.98	12.59
Ta1325	probable oligopeptide ABC transport system, ATP-binding protein appD	0.35	1.27E-04	10.41	8.91
Ta1326	probable dipeptide transport system, ATP-binding protein dppF	0.40	1.60E-04	10.85	9.52
Ta1341	mercuric reductase	0.50	3.48E-04	13.04	12.04
Ta1354	hypothetical protein	0.32	1.89E-04	11.94	10.29
Ta1355	hypothetical protein	0.12	5.64E-05	12.36	9.35
Ta1356	heme biosynthesis protein (NirJ) related protein	0.30	3.99E-04	12.49	10.74
Ta1361	GMP synthase subunit B	0.40	1.80E-04	13.78	12.44
Ta1364	hypothetical protein	0.39	2.06E-04	12.96	11.61
Ta1372	hypothetical protein	0.47	6.58E-04	11.96	10.89
Ta1381	hypothetical protein	0.48	8.57E-04	11.79	10.72
Ta1383	hypothetical protein	0.46	1.74E-03	9.47	8.34
Ta1397	hypothetical protein	0.47	1.08E-03	9.98	8.89
Ta1398	hypothetical protein	0.33	1.64E-04	12.62	11.01
Ta1399	hypothetical protein	0.37	4.91E-04	13.06	11.63
Ta1408	hypothetical protein	0.41	1.89E-04	12.62	11.35
Ta1409m	Predicted GTP:adenosylcobinamide-phosphate guanylyltransferase	0.33	1.30E-04	12.40	10.82
Ta1410	ferrichrome ABC transporter (ATP-binding p) related protein	0.45	2.67E-04	13.04	11.89
Ta1417	hypothetical protein	0.42	6.76E-05	13.21	11.96
Ta1423	hypothetical protein	0.49	7.08E-04	13.59	12.57
Ta1424	hypothetical protein	0.48	1.01E-04	14.14	13.08
Ta1431	hypothetical protein	0.26	1.89E-04	13.31	11.36
Ta1432	2-phosphosulfolactate phosphatas	0.36	3.22E-04	11.60	10.11
Ta1436	dihydrolipoamide acetyltransferase	0.47	8.63E-04	13.56	12.47
Ta1437	probable 3-methyl-2-oxobutanoate dehydrogenase chain E1-beta	0.45	1.06E-03	13.13	11.99
Ta1445	hypothetical protein	0.37	6.52E-04	11.69	10.24
Ta1446	trna nucleotidyltransferase related protein	0.46	1.24E-04	13.26	12.15
Ta1453	hypothetical protein	0.36	2.67E-04	10.80	9.34
Ta1455	hypothetical protein	0.43	1.20E-02	13.94	12.71
Ta1456	acetyl-CoA acetyltransferase	0.44	1.83E-02	13.84	12.67
Ta1457	hypothetical protein	0.38	1.23E-02	10.38	9.00
Ta1460	hypothetical protein	0.28	1.89E-04	10.99	9.15
Ta1465	acyl-CoA dehydrogenase related protein	0.28	1.24E-04	10.68	8.83
Ta1473	hypothetical protein	0.46	4.03E-04	11.60	10.49
Ta1474	hypothetical protein	0.41	2.14E-04	14.96	13.68
Ta1475	ribonucleotide-diphosphate reductase alpha subunit	0.35	5.95E-05	15.77	14.25
Ta1489	GTP-binding protein	0.37	4.39E-04	13.11	11.68
Ta1493	queuine tRNA-ribosyltransferase related protein	0.37	3.45E-05	15.31	13.89
Ta1500m	replication factor C small subunit	0.46	9.81E-04	13.94	12.82
Ta1507	hypothetical protein	0.20	3.71E-04	10.57	8.24
Ta1508m	Acyl-coenzyme A synthetase	0.37	5.94E-05	10.13	8.69
Up-regulated cytosolic protein-encoding mRNAs (anaerobically)					
Ta0012	hypothetical protein	3.32	5.47E-04	10.86	12.60
Ta0068	4-aminobutyrate aminotransferase	5.24	4.25E-05	12.36	14.75
Ta0117	probable isocitrate dehydrogenase	2.45	3.75E-04	11.84	13.13

Ta0165	hypothetical protein	2.85	1.29E-04	8.34	9.85
Ta0326	fixC protein related	2.30	2.33E-04	14.08	15.29
Ta0327	hypothetical protein	3.00	2.73E-04	13.32	14.91
Ta0329	probable electron transfer flavoprotein, alpha subunit	2.08	5.27E-05	14.43	15.49
Ta0413	cofactor-independent phosphoglycerate mutase	3.71	1.23E-04	10.21	12.10
Ta0456m	Malate oxidoreductase (malic enzyme)	2.59	1.32E-02	11.58	12.95
Ta0469	S-adenosyl-L-homocysteine hydrolase	2.21	6.52E-04	12.02	13.17
Ta0475	transposase related protein	2.05	5.10E-04	7.92	8.96
Ta0500	probable benzoylformate decarboxylase	2.71	1.07E-03	11.50	12.94
Ta0532	hypothetical protein	3.37	5.77E-05	9.54	11.29
Ta0577	sarcosine oxidase related protein	2.29	4.08E-03	9.44	10.63
Ta0582	acetyl-CoA acetyltransferase	3.01	6.75E-05	11.64	13.23
Ta0596	hypothetical protein	2.05	5.26E-04	9.17	10.21
Ta0601	hypothetical protein	2.60	1.56E-03	9.73	11.11
Ta0603	aldehyde dehydrogenase related protein	2.44	4.86E-05	12.93	14.22
Ta0605	translation initiation factor IF-2 beta subunit	2.56	3.20E-04	11.92	13.27
Ta0608	transcription regulator related protein	3.28	1.79E-04	7.82	9.54
Ta0611	inosine-5'-monophosphate dehydrogenase related protein	2.77	6.70E-04	12.62	14.08
Ta0619	probable 2-keto-3-deoxy gluconate aldolase	2.47	9.34E-05	9.52	10.83
Ta0624	hypothetical protein	2.55	2.53E-04	9.87	11.22
Ta0625m	Phosphoglycerate mutase	2.33	1.07E-04	7.64	8.86
Ta0626m	2-oxoacid-ferredoxin oxidoreductase, gamma subunit	38.34	4.32E-06	7.48	12.74
Ta0627	pyruvate ferredoxin oxidoreductase, delta subunit related protein	25.22	2.45E-05	9.69	14.34
Ta0628m	Pyruvate:ferredoxin oxidoreductase, alpha subunit	12.71	6.92E-05	10.64	14.31
Ta0629	ferrodoxin oxidoreductase beta subunit	9.73	4.81E-05	9.94	13.22
Ta0632	hypothetical protein	2.08	4.88E-04	10.71	11.77
Ta0634	hypothetical protein	9.03	6.70E-05	8.64	11.82
Ta0635	probable glutamate dehydrogenase	3.41	4.94E-05	11.74	13.51
Ta0650	argininosuccinate synthase	2.60	1.66E-03	10.71	12.09
Ta0680	hypothetical protein	5.41	2.25E-05	8.01	10.45
Ta0684	deoxyribose-phosphate aldolase	2.83	9.35E-05	7.79	9.29
Ta0693	hypothetical protein	6.05	9.74E-05	8.53	11.13
Ta0694	hypothetical protein	2.40	9.75E-04	9.07	10.34
Ta0695	hypothetical protein	2.83	6.74E-05	8.53	10.03
Ta0696	hypothetical protein	2.41	1.24E-04	8.55	9.82
Ta0701	hypothetical protein	4.39	4.84E-05	8.17	10.31
Ta0702	hypothetical protein	4.97	2.41E-05	7.59	9.91
Ta0707	hypothetical protein	4.42	2.29E-05	7.66	9.81
Ta0736	hypothetical protein	2.35	2.73E-04	7.85	9.09
Ta0744	hypothetical protein	2.50	2.54E-04	8.25	9.57
Ta0749	sorbitol dehydrogenase related protein	2.51	7.45E-04	10.40	11.73
Ta0754	glucose-1-dehydrogenase	2.20	1.97E-04	11.42	12.56
Ta0756	muconate cycloisomerase related protein	3.92	1.43E-04	10.49	12.47
Ta0759	hypothetical protein	2.32	2.33E-04	9.03	10.25
Ta0767	hypothetical protein	4.04	1.79E-04	9.55	11.56
Ta0774	nicotinamide-nucleotide adenylyltransferase	3.48	2.72E-04	8.59	10.39
Ta0778	pyruvoyl-dependent arginine decarboxylase	4.28	1.46E-04	7.79	9.88
Ta0788m	Diaminopimelate decarboxylase	3.81	3.58E-04	9.22	11.15
Ta0793a	hypothetical protein	2.17	5.55E-03	9.39	10.51
Ta0798	hypothetical protein	2.65	1.38E-03	11.53	12.93
Ta0810	probable aldehyde ferredoxin oxidoreductase	8.28	5.97E-05	11.61	14.66
Ta0811	glycine hydroxymethyltransferase related protein	2.62	1.15E-03	11.68	13.06
Ta0813	hypothetical protein	2.33	5.29E-03	11.97	13.19
Ta0828	hypothetical protein	3.09	6.02E-05	7.91	9.54
Ta0841	alcohol dehydrogenase related protein	2.61	1.01E-04	12.13	13.51
Ta0842	hypothetical protein	2.84	2.76E-04	12.58	14.09

Ta0843	hypothetical protein	3.57	3.69E-03	7.66	9.50
Ta0854	probable phosphoheptose isomerase	2.68	1.46E-04	8.56	9.98
Ta0866	thioredoxin related protein	3.16	7.78E-05	8.10	9.76
Ta0880	ribokinase related protein	2.06	3.73E-04	10.61	11.65
Ta0884	ferredoxin related protein	2.11	1.79E-04	11.33	12.41
Ta0885m	Predicted DNA-binding protein	2.22	5.49E-04	12.74	13.89
Ta0886	pyruvate phosphate dikinase	2.76	3.15E-04	9.63	11.10
Ta0895	hypothetical protein	2.13	7.13E-04	10.19	11.28
Ta0897	glucose 1-dehydrogenase	2.39	2.96E-04	13.70	14.96
Ta0916	hypothetical protein	3.01	1.51E-04	8.85	10.44
Ta0925	hypothetical protein	3.40	9.03E-04	8.86	10.63
Ta0934m	succinyl-diaminopimelate desuccinylase	2.68	4.93E-05	13.30	14.73
Ta0944m	GMP synthase subunit A	2.65	2.06E-04	12.46	13.87
Ta0945	transcription initiation factor IIB	2.10	1.64E-04	14.49	15.56
Ta0952	malate dehydrogenase	2.93	1.30E-04	13.50	15.06
Ta0969m	NADH dehydrogenase beta subunit	2.50	6.06E-04	13.08	14.40
Ta1009	hypothetical protein	2.28	1.79E-04	8.07	9.25
Ta1044m	phosphoribosylformylglycinamidine synthase	3.09	6.62E-04	8.70	10.33
Ta1078	aspartate aminotransferase related protein	11.57	4.84E-05	11.56	15.09
Ta1121	cob(I)alamin adenosyltransferase related protein	2.16	6.56E-04	8.61	9.72
Ta1124m	ribosome biogenesis protein	2.29	5.57E-04	9.95	11.14
Ta1217	hypothetical protein	2.11	1.59E-04	9.30	10.38
Ta1302	hypothetical protein	2.00	6.92E-04	10.16	11.16
Ta1336	type IV site-specific deoxyribonuclease Eco57I related protein	3.14	2.58E-04	9.48	11.13
Ta1339	cell-cycle regulation histidine triad (HIT) related protein	2.08	4.98E-04	8.21	9.27
Ta1374	hypothetical protein	4.06	2.94E-04	8.83	10.85
Ta1392	hypothetical protein	6.99	2.18E-04	7.19	9.99
Ta1393	hypothetical protein	2.26	1.91E-04	12.49	13.67
Ta1394	Nta operon transcriptional regulator related protein	2.98	2.32E-04	8.49	10.06
Ta1477	formiminotransferase cyclodeaminase related protein	2.00	2.93E-04	12.70	13.70
Ta1496	hypothetical protein	5.48	1.78E-04	9.34	11.80
Down-regulated membrane protein-encoding mRNAs (anaerobically)					
Ta0008	ATP synthase subunit I	0.33	2.33E-06	14.51	12.93
Ta0015	hypothetical membrane protein	0.12	4.83E-05	12.25	9.17
Ta0017	hypothetical membrane protein	0.19	2.99E-05	12.04	9.68
Ta0032	hypothetical protein	0.41	1.43E-04	12.19	10.90
Ta0041	hypothetical protein	0.45	2.60E-05	11.57	10.41
Ta0048	phosphate permease related protein	0.24	1.81E-03	13.08	11.02
Ta0062	hypothetical protein	0.19	5.71E-05	13.15	10.77
Ta0065m	Predicted permease	0.49	3.16E-04	12.61	11.59
Ta0076	cobalamin biosynthesis protein	0.35	4.37E-04	11.95	10.45
Ta0087	hypothetical protein	0.16	1.01E-04	13.10	10.42
Ta0094	hypothetical protein	0.40	5.32E-04	12.60	11.28
Ta0110a	hypothetical protein	0.42	2.45E-04	11.74	10.50
Ta0129	amylopullulanase related protein	0.33	1.79E-04	11.94	10.36
Ta0137	hypothetical protein	0.26	4.15E-05	10.87	8.94
Ta0141	proline/betaine transport protein related	0.26	4.97E-04	11.40	9.44
Ta0142	amino acid transporter related protein	0.19	1.24E-04	11.58	9.17
Ta0144	hypothetical protein	0.22	2.48E-04	12.59	10.42
Ta0145	trehalose/maltose transporter, inner membrane protein MalF related protein	0.17	3.21E-04	14.05	11.51
Ta0146	trehalose/maltose transporter, inner membrane protein (MalG) related protein	0.11	1.23E-04	13.30	10.16
Ta0148	cationic amino acid transporter 3 (HCAT3) related protein	0.35	5.42E-04	11.50	9.99
Ta0149	hypothetical protein	0.27	5.96E-05	11.65	9.78
Ta0151m	Signal peptidase I	0.27	1.84E-04	11.35	9.43

Ta0154	hypothetical protein	0.31	1.29E-04	11.52	9.81
Ta0167	thermopsin precursor related protein	0.47	1.89E-04	13.63	12.52
Ta0170m	Arsenite transport permease	0.21	4.04E-05	14.33	12.04
Ta0171	hypothetical protein	0.43	8.37E-04	15.51	14.30
Ta0173	hypothetical membrane protein	0.38	1.87E-04	13.91	12.50
Ta0181m	Ammonia permease	0.45	2.16E-04	10.29	9.13
Ta0183	hypothetical protein	0.48	2.95E-03	11.96	10.90
Ta0210	hypothetical membrane protein	0.43	4.82E-05	13.04	11.84
Ta0221	PAGO protein related protein	0.32	3.17E-06	11.71	10.07
Ta0224	self-defense gene tcr3 related protein	0.40	7.43E-05	14.50	13.17
Ta0226	hypothetical protein	0.46	1.56E-04	13.98	12.86
Ta0232	hypothetical membrane protein	0.20	6.09E-05	11.88	9.56
Ta0240	hypothetical protein	0.49	3.57E-03	9.38	8.34
Ta0254	hypothetical membrane protein	0.25	1.97E-04	12.13	10.12
Ta0264	probable dipeptide transport system permease protein dppC	0.39	4.77E-04	14.89	13.54
Ta0275	hypothetical protein	0.17	7.13E-04	12.05	9.52
Ta0300	hypothetical protein	0.18	1.81E-04	12.33	9.83
Ta0306	hypothetical membrane protein	0.41	6.09E-05	13.09	11.82
Ta0348	hypothetical protein	0.23	5.28E-05	13.34	11.19
Ta0354	hypothetical membrane protein	0.50	1.53E-03	11.98	10.98
Ta0376	lhypothetical protein	0.39	3.00E-04	12.47	11.12
Ta0407	hypothetical protein	0.26	4.01E-05	11.97	10.04
Ta0409	hypothetical membrane protein	0.25	8.43E-05	12.09	10.11
Ta0418	hypothetical membrane protein	0.29	1.94E-04	10.60	8.82
Ta0420	hypothetical protein	0.28	1.38E-04	11.41	9.55
Ta0426	hypothetical membrane protein	0.26	4.97E-04	15.12	13.19
Ta0427	L-ASPARAGINE PERMEASE related protein	0.27	1.89E-04	10.74	8.84
Ta0442	transport membrane protein related protein	0.46	1.30E-04	14.00	12.88
Ta0488	hypothetical protein	0.45	1.24E-04	12.11	10.96
Ta0503	hypothetical protein	0.35	2.66E-04	12.71	11.21
Ta0503a	hypothetical protein	0.38	1.49E-04	12.18	10.79
Ta0560	flagellar accessory protein FabJ	0.41	1.28E-04	12.86	11.58
Ta0590a	hypothetical protein	0.37	3.85E-04	11.84	10.39
Ta0591	hypothetical protein	0.17	5.69E-04	13.35	10.76
Ta0592	hypothetical membrane protein	0.20	8.34E-04	13.11	10.82
Ta0620	amino acid transporter protein related protein	0.32	6.24E-05	11.89	10.23
Ta0645	amino acid permease related protein	0.19	6.29E-04	11.06	8.68
Ta0661	hypothetical protein	0.19	6.21E-05	13.37	10.96
Ta0722	hypothetical protein	0.46	7.58E-04	11.69	10.58
Ta0726	hypothetical membrane protein	0.39	4.00E-04	11.54	10.18
Ta0785	hypothetical protein	0.41	2.30E-04	11.37	10.06
Ta0789	amino acid transport protein related protein	0.08	4.87E-05	12.47	8.85
Ta0839	hypothetical protein	0.43	1.85E-03	12.40	11.17
Ta0851	hypothetical protein	0.48	2.78E-03	11.84	10.78
Ta0856	hypothetical membrane protein	0.40	1.78E-04	13.59	12.28
Ta0870	hypothetical protein	0.49	7.64E-04	13.25	12.23
Ta0877	amino acid permease related protein	0.42	3.67E-03	13.25	12.01
Ta0918	hypothetical protein	0.46	1.04E-04	12.60	11.48
Ta0933	multidrug resistance protein NorA related protein	0.41	5.63E-05	11.99	10.72
Ta0976	xanthomonapepsin related protein	0.47	2.73E-04	14.37	13.27
Ta0992	cytochrome bd ubiquinol oxidase related protein	0.05	4.86E-05	15.03	10.81
Ta0993	hypothetical membrane protein	0.05	2.62E-05	14.56	10.24
Ta1069	multidrug resistance protein related protein	0.37	1.83E-04	13.19	11.76
Ta1070	hypothetical protein	0.38	2.28E-05	11.65	10.25
Ta1071	hypothetical protein	0.31	3.99E-05	12.49	10.82
Ta1081	ATP-dependent proteinase La (Lon) related protein	0.42	5.57E-05	15.24	14.01

Ta1085	multidrug-efflux transporter related protein	0.35	4.78E-05	11.81	10.31
Ta1102	hypothetical membrane protein	0.45	7.02E-03	11.83	10.69
Ta1108	hypothetical protein	0.19	1.69E-04	12.71	10.32
Ta1118	hypothetical protein	0.38	5.64E-05	10.87	9.47
Ta1143	heavy-metal transporting P-type ATPase related protein	0.31	1.55E-03	14.44	12.76
Ta1161m	Predicted cationic amino acid transporter	0.16	6.38E-05	13.64	11.02
Ta1165	sugar transport related membrane protein	0.36	7.29E-04	13.11	11.63
Ta1169	probable high-affinity nickel-transport protein	0.06	5.18E-04	14.37	10.31
Ta1184	hypothetical membrane protein	0.29	7.41E-05	12.93	11.12
Ta1185m	Kef-type K+ transport system, membrane component	0.36	2.67E-04	11.33	9.86
Ta1195	hypothetical protein	0.39	8.79E-06	14.00	12.66
Ta1198	4-hydroxybenzoate octaprenyltransferase	0.38	5.30E-05	13.15	11.74
Ta1213	purine-cytosine permease related protein	0.34	1.73E-04	12.73	11.17
Ta1218a	hypothetical protein	0.43	1.84E-04	12.40	11.18
Ta1220	hypothetical membrane protein	0.35	4.01E-05	13.01	11.50
Ta1221	hypothetical membrane protein	0.35	2.35E-04	13.19	11.67
Ta1222	cytochrome b related protein	0.43	2.67E-04	14.59	13.38
Ta1225	hypothetical protein	0.35	1.19E-04	11.82	10.30
Ta1228	cytochrome b related protein	0.34	2.30E-03	12.43	10.87
Ta1234	hypothetical membrane protein	0.27	1.81E-04	11.45	9.55
Ta1235	hypothetical membrane protein	0.23	1.02E-04	11.75	9.60
Ta1273	hypothetical membrane protein	0.35	2.51E-04	10.63	9.14
Ta1274	sterol-regulatory element-binding proteins intramembrane protease related protein	0.20	5.18E-05	13.24	10.91
Ta1301	membrane protein 7, erythrocyte (human) related protein	0.24	4.89E-05	12.22	10.16
Ta1302a	preprotein translocase	0.37	2.19E-03	13.82	12.37
Ta1308m	potassium-transporting ATPase subunit C	0.22	6.77E-05	15.10	12.89
Ta1309	potassium-transporting ATPase subunit B	0.30	1.30E-04	14.91	13.17
Ta1310	potassium-transporting ATPase subunit A	0.28	1.92E-04	14.32	12.47
Ta1317	hypothetical membrane protein	0.41	2.32E-04	12.00	10.71
Ta1322	surface antigen genes (Methanosarcina mazei) related protein	0.47	4.17E-04	11.94	10.86
Ta1327	oligopeptide transport system, permease protein appC related protein	0.35	1.29E-04	12.28	10.76
Ta1328	dipeptide transport system, permease protein (dppB) related protein	0.42	1.43E-04	12.25	11.01
Ta1329	dipeptide ABC transport system, periplasmic dipeptide-binding protein dppA	0.30	1.23E-04	11.70	9.95
Ta1333	hypothetical membrane protein	0.49	2.11E-04	10.76	9.72
Ta1340	transport membrane protein (permease) related protein	0.43	2.73E-04	10.99	9.78
Ta1344m	Predicted membrane-associated Zn-dependent protease	0.42	4.28E-05	14.82	13.56
Ta1345	hypothetical membrane protein	0.32	2.73E-04	13.44	11.77
Ta1346	multidrug-efflux transporter related protein	0.46	8.33E-04	13.03	11.91
Ta1359a	hypothetical protein	0.42	4.23E-04	14.07	12.82
Ta1371	hypothetical protein	0.37	3.33E-04	12.59	11.16
Ta1378	hypothetical protein	0.01	3.85E-06	15.02	8.64
Ta1379	phosphate transporter related protein	0.09	4.20E-05	12.71	9.18
Ta1385	hypothetical protein	0.24	2.58E-05	14.45	12.40
Ta1391	phosphate transporter related protein	0.49	3.75E-03	9.91	8.88
Ta1403	thermopsin precursor related protein	0.12	9.69E-04	15.41	12.38
Ta1404	multidrug resistance (emrA) protein related	0.06	8.78E-06	14.28	10.32
Ta1421	hypothetical protein	0.40	1.61E-04	13.76	12.43
Ta1430	hypothetical protein	0.45	8.26E-04	13.92	12.78
Ta1462	penicillin amidase precursor related protein	0.39	1.25E-04	13.32	11.96
Ta1469	hypothetical protein	0.43	9.41E-05	11.43	10.21
Ta1470	hypothetical protein	0.45	6.68E-05	10.75	9.59
Ta1494	hypothetical protein	0.23	4.80E-05	14.02	11.89
Up-regulated membrane protein-encoding mRNAs (anaerobically)					
Ta0067	hypothetical protein	2.43	1.89E-04	10.48	11.76

Ta0467	hypothetical protein	2.41	1.75E-02	9.71	10.98
Ta0633	glycerol-3-phosphate dehydrogenase related protein	2.76	5.85E-04	11.26	12.73
Ta0640	hypothetical membrane protein	18.93	7.46E-04	10.27	14.51
Ta0710	hypothetical protein	2.06	2.80E-05	7.29	8.33
Ta0719	hypothetical protein	4.74	5.80E-05	7.77	10.01
Ta0793	multidrug resistance protein related protein	2.43	8.37E-04	10.31	11.59
Ta0826	hypothetical protein	3.51	1.77E-04	8.27	10.08
Ta0853	ethanolamine permease related protein	3.85	1.97E-04	11.13	13.08
Ta0876	hypothetical protein	3.72	1.25E-04	7.55	9.44
Ta0970m	NADH dehydrogenase alpha subunit	3.34	5.23E-04	11.64	13.38
Ta1079	cobalamin synthase	2.39	5.06E-04	13.82	15.08
Ta1120	probable cationic amino acid transporter	3.29	5.65E-04	12.17	13.89
Ta1172	hypothetical protein	13.42	9.84E-04	9.53	13.28
Ta1173	hypothetical protein	2.51	5.99E-03	7.49	8.82