

**Two-dimensional
reversed-phase liquid chromatography
coupled to MALDI TOF/TOF mass spectrometry:
an approach for shotgun proteome analysis**

Dissertation
zur Erlangung des Grades
des Doktors der Naturwissenschaften
der Naturwissenschaftlich- Technischen Fakultät III
Chemie, Pharmazie, Bio- und Werkstoffwissenschaften
Der Universität des Saarlandes

von

Maria Lasaosa

Saarbrücken

2008

Tag des Kolloquiums: 13 November 2008

Dekan: Prof. Dr. Uli Müller

Berichterstatter: Prof. Dr. Elmar Heinzle
P.D. Dr. Andreas Tholey
Dr. Markus Martin

Acknowledgements

First of all I would like to thank my supervisor Prof. Elmar Heinzle for his constant support, fruitful discussions and contribution to this work. I would also like to thank P.D. Andreas Tholey for the initial input in the work and the interesting debates during the regular meetings of the Functional Proteomics Group. I kindly thank Prof. Christian Huber for making possible a very successful collaboration with his group.

Special thanks to my colleagues of the Functional Proteomics Group (Masoud Zabet Moghaddam and Nathalie Selevsek) for the very helpful discussions, valuable advice and fantastic working atmosphere. I also thank my colleagues of the Biochemical Engineering Group for their kind support and friendly talks at the “Kaffeemaschine”. Many thanks to Rahul Deshpande for his stimulating remarks during the work, Dipl.Biol. Veronika Witte for her help with the in-gel digestions and Dr. Klaus Hollemeyer for his introduction to the gel electrophoresis.

Many thanks to my cooperation partner and “old” friend Nathanaël Delmotte of the Instrumental Analysis and Bioanalysis, as well as to Katja Melchior for a successful collaboration.

Thanks to Dr. Platsch (BASF AG) and his team colleagues for their advice with the 2D-PAGE technology. I would also like to thank Martina Schnoelzer and colleagues (DKFZ) for the information about in-gel digestion and MALDI TOF MS.

Special thanks to Mathias, my family in Spain and friends for their constant support and encourage during all this time.

Table of contents

ACKNOWLEDGEMENTS	3
TABLE OF CONTENTS	5
ABSTRACT	9
English version	9
German version	10
1. INTRODUCTION	11
1.1. Proteomics	11
1.2. Proteome analysis	12
1.3. Separation techniques for the proteome analysis	15
1.3.1. Two-dimensional gel electrophoresis.....	15
1.3.2. Multidimensional liquid chromatography	17
1.4. Mass spectrometry	18
1.4.1. MALDI TOF mass spectrometry.....	18
1.4.2. Electrospray mass spectrometry	21
1.4.3. Liquid chromatography-Mass spectrometry.....	22
1.4.4. Peptide fragmentation and tandem mass spectrometry	24
1.5. Protein identification by database search	26
1.6. <i>Corynebacterium glutamicum</i>	27
2. GOAL OF THE THESIS	29
3. MATERIALS AND METHODS	31
3.1. Materials	31
3.2. Cultivation of <i>Corynebacterium glutamicum</i>	32
3.3. Protein extraction from <i>Corynebacterium glutamicum</i> cells	32
3.4. Two-dimensional liquid chromatography	33

3.4.1. Digestion of cytosolic proteins	33
3.4.2. First dimension: reversed-phase HPLC at pH 10	33
3.4.3. Fraction screening by MALDI TOF MS	34
3.4.4. Second dimension: reversed-phase HPLC at pH 2.1	34
3.4.5. LC-MALDI spotting	35
3.4.6. LC-MALDI TOF/TOF analysis	35
3.4.7. Database search and protein identification	35
3.5. Two-dimensional gel electrophoresis	36
3.5.1. First dimension: isoelectric focusing	36
3.5.2. Second dimension: SDS-PAGE	37
3.5.3. Gel staining	37
3.5.4. Gel image analysis	38
3.5.5. In-gel digestion	38
3.5.6. Zip Tip™	39
3.5.7. MALDI TOF/TOF analysis	39
3.5.8. Database search	40
3.6. One-dimensional gel electrophoresis	40
4. RESULTS	41
4.1. Optimization of the protein extraction procedure	41
4.2. Evaluation and optimization of the LC-MALDI system	45
4.2.1. LC-MALDI spotting set-up	45
4.2.2. Optimization of LC-MALDI	47
Optimization of the spot deposition	47
Optimization of the mass accuracy	48
Collision induced dissociation (CID) versus post-source decay (PSD)	51
Performance of the Timed Ion Selector (TIS)	54
4.3. Proteome analysis of <i>Corynebacterium glutamicum</i> by two-dimensional liquid chromatography & MALDI TOF/TOF MS	55
4.3.1. First dimension: reversed-phase chromatography at high pH	56
4.3.2. Fraction screening by MALDI TOF MS	57
4.3.3. Second dimension: reversed-phase chromatography at low pH	59
4.3.4. MALDI TOF/TOF analysis	61
4.3.5. Database search and identified proteins	65

4.4. Proteome analysis of <i>Corynebacterium glutamicum</i> by two-dimensional gel electrophoresis and MALDI TOF/TOF MS	69
5. DISCUSSION	73
5.1. Evaluation of 2D-RP-HPLC combined with MALDI TOF/TOF for proteome analysis	73
5.1.1. Evaluation of the first dimension	73
5.1.2. Evaluation of the second dimension.....	76
5.1.3. Orthogonality of the separation.....	78
5.1.4. Evaluation based on peptide identification.....	80
5.1.5. Evaluation based on protein identification.....	82
5.2. Comparison of the novel 2D-RP-HPLC MALDI TOF/TOF approach with two alternative approaches for proteome analysis	85
5.2.1. Comparison with a top-down approach: 2D-PAGE.....	85
5.2.2. Comparison MALDI TOF/TOF vs. ESI-IT.....	89
5.2.3. Comparison of methods based on protein identification.....	96
TCA cycle	100
Glycolysis.....	101
Pentose phosphate pathway.....	101
5.2.4. Comparison of methods based on the codon adaptation index	103
5.3. Comparison with results from literature.....	105
5.3.1. Comparison with the MudPIT approach	105
5.3.2. Comparison with a 2D-PAGE approach	106
6. SUMMARY AND CONCLUSIONS.....	108
7. OUTLOOK	110
8. LIST OF ABBREVIATIONS, ACRONYMS AND SYMBOLS.....	111
9. LITERATURE	113
APPENDIX	121

Abstract

English version

Large-scale proteome analysis of complex biological mixtures requires high resolving power separations, high-throughput mass spectrometry and accurate detection, and bioinformatics tools. In the present work LC-MALDI TOF/TOF mass spectrometry was combined for the first time with a two-dimensional separation based on reversed-phase chromatography at high and low pH for the analysis of the cytosolic proteome of *Corynebacterium glutamicum*. The proteome coverage achieved by this approach, 55%, is the highest reported up to day for this bacterium. A total of 1644 proteins including single-peptide based identifications were identified.

The sample was also analysed by other conventional methods in the proteome analysis. The classical 2D-PAGE approach presents more limitations and delivered 166 different proteins including enzymes of the main metabolic pathways. Complementary results were found at peptide level for the LC-MALDI and 2D-PAGE approaches. Fractions collected during the first dimension at high pH were analysed by LC-MALDI MS and compared with the analysis carried out by LC-ESI-IT MS, which identified 745 proteins [1]. Further comparison with results found in literature confirmed that the two-dimensional reversed phase combination with LC-MALDI MS/MS is a promising tool in the proteome analysis. The MALDI based approach showed higher sensitivity than other approaches and was able to identify proteins over a larger dynamic range, based on the codon adaptation indexes of the proteins.

German version

An die Large-Scale Proteomanalyse von komplexen biologischen Mischungen sind eine Reihe von Anforderungen gestellt. Unter Anderem sind Trennungen mit hohem Auflösungsvermögen, High-Throughput Massenspektrometrie, präzise Massenbestimmung und Bioinformatik-Tools erforderlich. Gegenstand der vorliegenden Arbeit ist die Evaluierung der Kombination aus LC-MALDI-TOF/TOF Massenspektrometrie mit einer zwei-dimensionalen Trennung basierend auf Reverse-Phase-Chromatographie bei hohen und niedrigen pH-Werten. Des cytosolischen Proteoms von *Corynebacterium glutamicum* wurde mit der Methode analysiert. Dies ermöglichte die Identifizierung von 55% des Proteoms von *Corynebacterium glutamicum* und ist damit der nach momentanem Kenntnisstand effizienteste Ansatz. Insgesamt konnten 1644 verschiedene Proteine identifiziert werden.

Konventionelle Methoden der Proteomanalyse erreichten deutlich schlechtere Ergebnisse. So erzielte der klassische 2D-PAGE-Ansatz bei deutlich mehr Einschränkungen nur 166 Proteinidentifikationen inklusive wichtiger Enzyme der Stoffwechselwege. Komplementäre Ergebnisse konnten auf der Peptid-Ebene für LC-MALDI und 2D-PAGE Ansätze gefunden werden. Fraktionen der ersten Dimension bei hohem pH-Wert mittels LC-MALDI MS wurden analysiert und mit Ergebnissen aus LC-ESI-IT MS [1] Untersuchungen verglichen. Es hat sich gezeigt, dass der MALDI-basierte Ansatz höhere Sensitivität aufweist und in der Lage war Proteine über eine weitaus größere dynamische Reichweite bezogen auf den Codon Adaptation Index zu identifizieren.

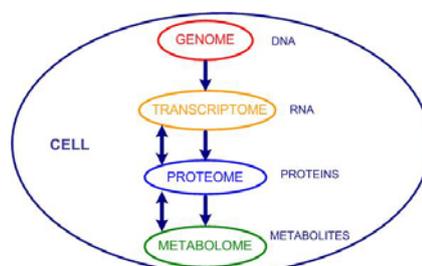
1. Introduction

1.1. Proteomics

In the past years new concepts in biochemistry have emerged as a requisite to define new fields of investigation. In the 1980s the term “genomics” was coined to describe the study of an organism’s entire genome, which is the hereditary information encoded in the DNA of the organism. The term “proteomics” was introduced for the first time in 1994 during the Siena meeting for protein separation based on two-dimensional gel electrophoresis [2]. This term involves the identification and quantification of all proteins present in a sample at a certain time and under certain conditions. Although the genome of an organism is steady the product of its expression, the proteome, is dynamic.

In cells there are different levels of organisation according to composition and function: DNA, RNA, proteins, metabolites (see Figure 1). The comprehensive analysis of each level is called respectively genomics, transcriptomics, proteomics, metabolomics. All these levels are interdependent and also influence the proteome composition of the cell. The proteins expressed in an organism, cell or tissue depend on the life cycle and external factors (temperature, nutrients and other environmental factors). Different cellular states lead to different gene expressions, which is revealed at RNA and protein level [3]. The global study of all “-omics” including metabolic fluxes is called systems biology and it aims the understanding of complex interactions across all levels in biological systems and the development of predictive models [4, 5].

Figure 1: Different organization levels in a cell. The genes are transcribed to RNAs, and these are translated to proteins, which catalyze the chemical reactions in the cell involving metabolites.



Protein expression is regulated by transcriptional and translational factors, which are proteins subjected also to regulatory mechanisms of synthesis and degradation [6]. Protein activity is further influenced by posttranslational modifications (e.g.

phosphorylation, glycosylation). The analysis of posttranslational modifications regulated by cellular environment, functional characterization of protein complexes and the functional characterization of protein pathways and their quantitative changes are currently some of the biological applications of proteomics [7, 8].

Since proteins represent the major part of the machinery of the cell, the study of the proteome of sequenced organisms yields information about occurrence and abundance of proteins and is essential for the understanding of protein activities and metabolic regulation. However, the proteome composition of an organism is very heterogeneous and its complexity depends on the type of organism. The human genome contains 20,000-25,000 protein-coding genes [9], whereas the genome of the bacterium *E. coli* encodes for 4,290 proteins [10]. Moreover, protein abundance can cover many orders of magnitude e.g. *Saccharomyces cerevisiae* has been reported to have a protein abundance in the cell spanning over six orders of magnitude [11]. As a consequence, it happens to be that high abundant proteins are identified multiple times while less abundant proteins remain under sampled [12]. In order to face the analysis of such complex samples it is essential to combine high-throughput techniques with sensitive and accurate mass spectrometers. The large amount of proteomic datasets usually generated requires software tools and computational approaches for data processing [13].

1.2. Proteome analysis

The proteome represents all the proteins expressed in the organism, cell or tissue that is object of study. Proteins are macromolecules of molecular mass between 10^4 - 10^6 Da responsible for a large variety of biological functions in the cell. Enzymes catalyze biochemical reactions; transport proteins carry molecules or ions to a specific site; regulatory proteins control the reactions of the metabolism. Other proteins form part of biological structures, have mechanical, transcriptional or signaling functions. Every protein has a unique sequence of amino acids, which is encoded in the corresponding gene. All the 20 amino acids contain a tetrahedral carbon linked to an amino group, a carboxylic group and a hydrogen atom. They differ in their side-chain, which determines the chemical properties of the amino acid. Proteins are linear polymers formed by a covalent link between the amino group of one amino acid and the carboxylic group of the next amino acid. The sequence of a protein determines the

protein folding, its three-dimensional structure and finally, its function. The side chain of some amino acid is ionizable and therefore carries a positive or negative charge depending on the pH (see pK_a in Table 1).

Table 1: Amino acids naturally occurring in proteins.

Name	Three (one) letter code	Elemental composition	Monoisotopic residue mass	pK_a	Hydrophobic Index*
Alanine	Ala (A)	C_3H_5NO	71.037	—	0.62
Arginine	Arg (R)	$C_6H_{12}N_4O$	156.101	12.5	-2.53
Asparagine	Asn (N)	$C_4H_6N_2O_2$	114.043	—	-0.78
Aspartic acid	Asp (D)	$C_4H_5NO_3$	115.027	4.1	-0.90
Cysteine	Cys (C)	C_3H_5NOS	103.009	8.3	0.29
Glutamic acid	Glu (E)	$C_5H_7NO_3$	129.042	4.1	-0.74
Glutamine	Gln (Q)	$C_5H_8N_2O_2$	128.058	—	-0.85
Glycine	Gly (G)	C_2H_3NO	57.021	—	0.48
Histidine	His (H)	$C_6H_7N_3O$	137.059	6.0	-0.40
Isoleucine	Ile (I)	$C_6H_{11}NO$	113.084	—	1.38
Leucine	Leu (L)	$C_6H_{11}NO$	113.084	—	1.06
Lysine	Lys (K)	$C_6H_{12}N_2O$	128.095	10.8	-1.50
Methionine	Met (M)	C_3H_9NOS	131.040	—	0.64
Phenylalanine	Phe (F)	C_9H_9NO	147.068	—	1.19
Proline	Pro (P)	C_5H_7NO	97.053	—	0.12
Serine	Ser (S)	$C_3H_5NO_2$	87.032	—	-0.18
Threonine	Thr (T)	$C_4H_7NO_2$	101.048	—	-0.05
Tryptophan (W)	Trp (W)	$C_{11}H_{10}N_2O$	186.079	—	0.81
Tyrosine (Y)	Tyr (Y)	$C_9H_9NO_2$	163.063	10.9	0.26
Valine (V)	Val (V)	C_5H_9NO	99.068	—	1.08
Terminal amino group	—	—	—	8.0	—
Terminal carboxyl group	—	—	—	3.1	—

* Reference: Eisenberg et al. [14]

Amino acids also show a different degree of hydrophobicity, which influences their location in a protein. As an example, trans-membrane domains are rich in very hydrophobic amino acids while hydrophilic amino acids are situated on the protein surface.

Proteome analysis is usually preceded by protein denaturation to unfold the three-dimensional structure of the proteins leaving only the linear chain of amino acids or primary structure. Protein unfolding is carried out by addition of detergents like sodium dodecyl sulfate (SDS), chaotropic agents like urea, organic solvents or heat. The covalent disulfide bridges between the side-chains of two cysteines can be reduced by addition of dithiotreitol or β -mercaptoethanol. The subsequent alkylation of the cysteines avoids refolding of the protein. A protease, an enzyme which cleaves the amino acid chain at specific residues, is added to digest the proteins and generate peptides with different number of amino acids. As a result, peptides of different lengths and with diverse physico-chemical properties are obtained [15, 16].

The first development concerning comprehensive analysis of all the proteins present in a sample was two-dimensional gel electrophoresis (2D-PAGE) [17, 18], which separates proteins according to two physico-chemical properties: isoelectric point and molecular mass. Alternatively, peptides resulting from enzymatic digestion of proteins with known proteases can be separated by liquid chromatography. Often the complexity of a proteome sample requires multidimensional chromatographic steps. Nevertheless, the key to high-throughput protein identification was the development of the so-called “soft-ionization” techniques in the mass spectrometry field. ESI and MALDI are ionization techniques developed in the 1980s which permit ionization of peptides and proteins followed by accurate mass measurements. In addition, fragmentation of peptides in the mass spectrometer can lead to complete or partial sequencing and thus unambiguous identification. In order to accelerate protein identification, specific algorithms were developed for systematic comparison of the masses in a spectra with the calculated masses obtained from the theoretical digestion of sequenced organisms. Currently proteome analysis can be classified in “bottom-up” and “top-down” approaches. When the separation takes place at protein level, it is called “top-down” whereas if the proteins are digested and separated at the peptide level the approach is called “bottom-up”.

At present most challenging questions in comprehensive proteome analysis are sensitivity, dynamic range and instrument capacity to acquire and process data fast and

effective. While the sensitivity has improved considerably in the past years, it is believed that progress in sample fractionation can improve the dynamic range. In addition, intelligent acquisition software can aid mass spectrometers to avoid redundant measurements on high abundant peptides and improve selectivity [19].

1.3. Separation techniques for the proteome analysis

1.3.1. Two-dimensional gel electrophoresis

Two-dimensional gel electrophoresis is a top-down approach based on protein mobility in an electric field. In the first dimension, the isoelectric focusing (IEF), proteins are placed on a strip containing an immobilized pH gradient (IPG) and an electric field is applied between the extremes of the strip. The proteins are charged due to amino-, carboxylic- groups and move towards the electrode of opposite sign until they reach the pH region corresponding to their isoelectric point (pI). When they reach their pI, proteins carry net charge zero and do not move in the applied electric field. Thus proteins are focused on the strip according to their pI.

After the first dimension proteins focused in the strip are dissolved in a SDS containing solution. This anionic detergent binds the protein, denatures it and provides a negative charge. SDS binds the protein in a ratio 1.4 g SDS/g protein, proportional to its molecular mass. In the second dimension, the strip containing the focused proteins saturated with SDS is placed on top of a polyacrylamide gel which has a constant pore size and acts as sieve. Proteins are separated in an electric field according to their hydrodynamic volume, ultimately according to their molecular mass (see Figure 2).

In order to visualise the proteins in the gel, a staining procedure is required. Staining can be performed with organic dyes like Coomassie blue, fluorescent dyes [20] or with silver nitrate. The interesting protein spots are excised from the gel and digested enzymatically. Finally, protein identification is performed by mass spectrometric measurement of the peptides released after in-gel digestion [21, 22].

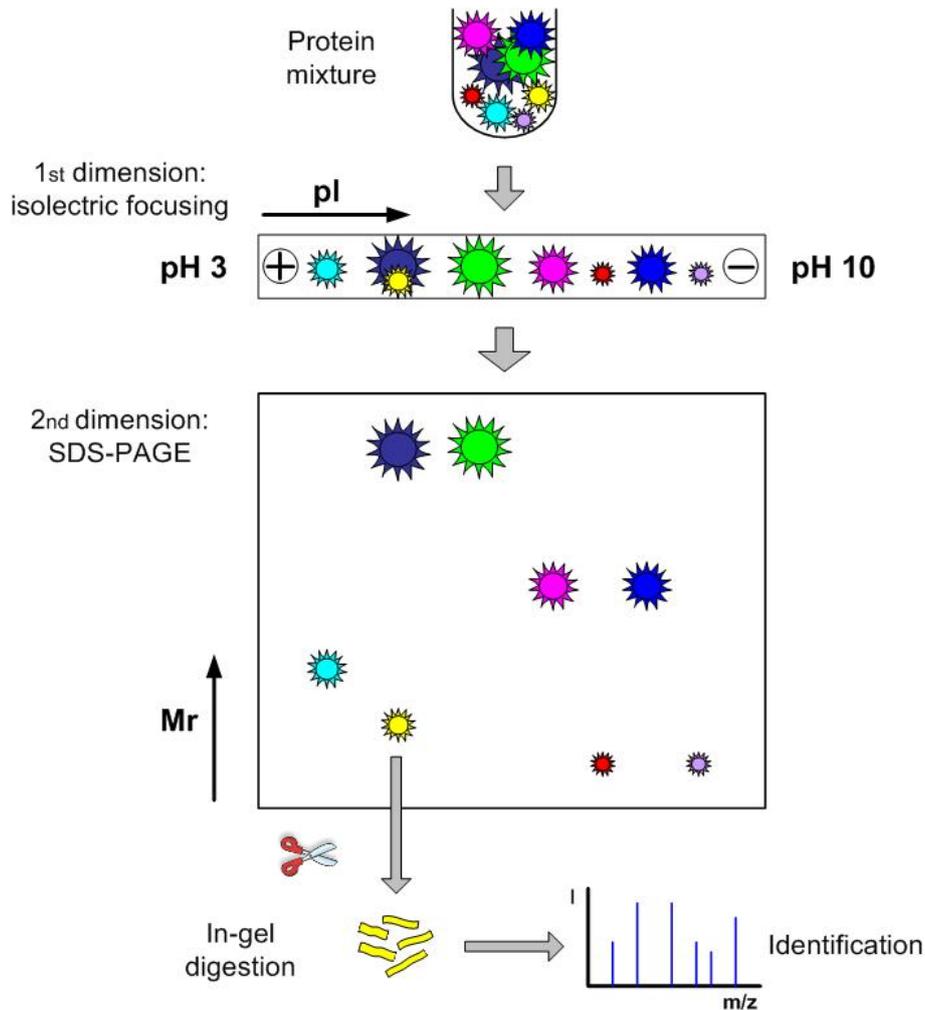


Figure 2: Scheme of 2D-PAGE. In the first dimension proteins are separated according to their isoelectric point. In the second dimension proteins are denatured in an SDS solution and separated according to molecular mass. Finally a protein is digested and identified by MS.

Although 2D-PAGE permits visualisation of up to 1000-2000 protein spots in a gel there are still a number of limitations not yet resolved [23, 24]:

- Sensitivity limits are in the ng region for organic dyes and better for silver and fluorescence staining [25]. The fluorescence dyes are, however, more expensive.
- Reproducibility of the gels is not always achieved. Experimental performance is laborious and no automation is available.
- Low dynamic range for quantification due to saturation effects of the dyes.
- Not all proteins can be resolved in the gels: very acidic, very basic, very hydrophobic and membrane proteins are underrepresented, as well as very small or very big proteins.

1.3.2. Multidimensional liquid chromatography

Multidimensional high-performance liquid chromatography (MD-HPLC) is the alternative approach to 2D-PAGE for the proteome analysis. Generally the proteome is digested generating a complex mixture of thousands of peptides which is subsequently separated by HPLC. Due to the complexity of the sample, separation of the peptides cannot be achieved by only one chromatographic step and therefore two-dimensional chromatography (2D-HPLC) is employed. The prerequisite to achieve good separation efficiency is the orthogonality between both chromatographic dimensions [26]. Traditionally, strong-cation exchange chromatography (SCX) is employed in the first dimension and reversed-phase chromatography (RP) in the second dimension [27-32]. SCX separates peptides according to their charge at the working pH. In reversed-phase chromatography (RPC) separation is based on the interaction of the peptides with the non-polar stationary phase. RP is the method of choice in the second dimension due to the compatibility of solvents with mass spectrometry, whereas the presence of salts in SCX has a signal suppression effect in the ionization. Frequently RPC of peptides is performed by addition of an ion-pairing reagent to the mobile phase generating reversed-phase ion-pair chromatography (RP-IPC). In RP-IPC elution of peptides is accomplished by a combination of hydrophobic and electrostatic interactions [33]. The SCX-RP set-up has been employed successfully for the proteome analysis with either on-line [30, 31] or off-line configurations [28, 34].

SCX and RP are orthogonal techniques and thus adequate for peptide separation by 2D-HPLC. However, it has been reported that the efficiency of SCX separations in 2D-SCX-RP-HPLC setups is lower than expected [35]. Due to the restricted number of charge states for the peptides at working conditions, mostly +2, +3 and +4, the separation by SCX is clustered. Recently, the orthogonality achieved by combination of different chromatographic systems has been studied. It has been reported that two-dimensional separations based on RP-RP-HPLC [36, 37], RP-IPRP-HPLC [1] and zwitterionic hydrophilic interaction liquid chromatography (ZIC-HILIC)-RP [38] represent a feasible alternative to the standard SCX-RP-HPLC combination. Moreover, compared to the 2D-SCX-RP scheme a 2-fold increase in identified proteins/peptides has been reported by switching to 2D-RP-RP [39] and more accurate retention prediction models [40].

The columns employed for peptide separation by reversed-phase chromatography in the second dimension has been traditionally capillary columns packed with

microparticulate sorbents. However, it has been revealed that the organic-polymer based monolith columns are suitable for proteomic applications due to their excellent chromatographic separation efficiency and high peak capacity in a short period of time [41-43]. A monolith column is a continuous rigid rod, highly porous resistant to high flow rates, which shows high mass transfer and consequently high separation efficiency [44, 45].

1.4. Mass spectrometry

As proteins and peptides are non-volatile and labile molecules under high temperature conditions, the development of “soft” ionization techniques was crucial to make possible their analysis by mass spectrometry. Currently there are two ionization methods introduced in 1988 most suitable for analysis of proteins and peptides: MALDI and ESI. Matrix Assisted Laser Desorption Mass Spectrometry (MALDI) was introduced by M. Karas and F. Hillenkamp et al. [46], and K. Tanaka et al. [47]. Electrospray ionization (ESI) of large biomolecules was introduced by Fenn et al. [48]. The volatilization of peptides and proteins permitted the application of mass spectrometry for accurate mass determination. Furthermore, it was observed that the fragmentation of a peptide delivers sufficient information to sequence it at least partially. Consequently, the protein can be identified by correlation of the acquired masses with the predicted masses obtained by a theoretical or “in-silico” digest of the protein. The improvements carried out in MS instrumentation during the past 25 years have increased the sensitivity in about 5 times every three years [49]. In the following lines the mass spectrometric techniques most commonly used in the proteome analysis will be described.

1.4.1. MALDI TOF mass spectrometry

In MALDI MS the sample is mixed with a matrix and deposited on a plate for analysis. The matrix is a small aromatic organic compound dissolved in an organic solvent like ethanol or acetonitrile at acidic pH. On the MALDI-plate the solvent evaporates and the sample co-crystallises with matrix crystals. The sample is brought into the mass spectrometer under high vacuum and irradiated with a pulsed laser (Figure 3). The matrix is characterized by strong absorption at the laser wavelength,

usually in the near-ultraviolet. Typically, N₂ lasers emitting at $\lambda = 337$ nm or Nd:YAG lasers emitting at $\lambda = 355$ nm, are employed in MALDI mass spectrometry of peptides and proteins.

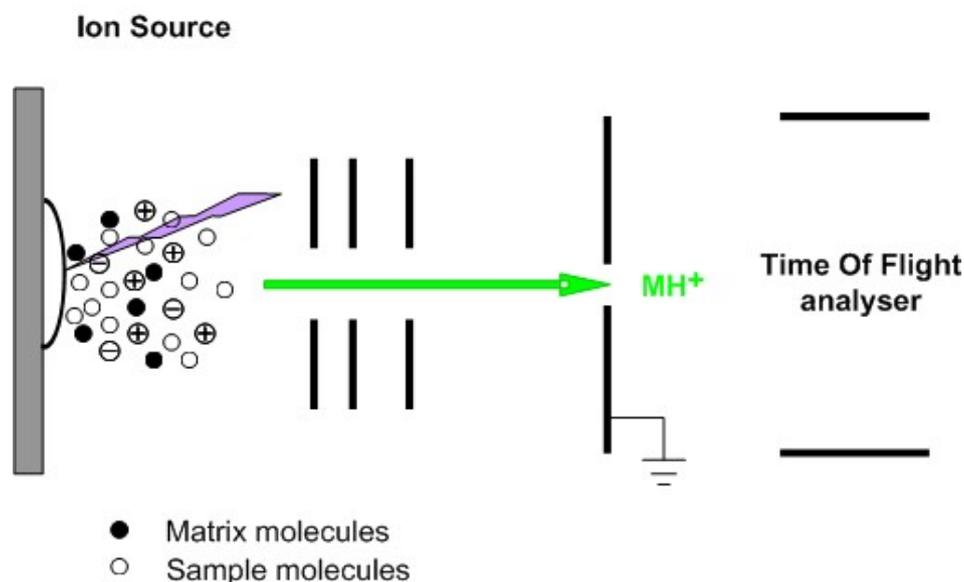


Figure 3: Ion source of a MALDI TOF instrument. A laser of typically 355 nm (Nd:YAG) or 337 nm (N₂ laser) irradiates the sample. Peptides are ionized by proton transfer from matrix ions. After that they are accelerated (+ 20 kV) by application of a high voltage to the time-of-flight (TOF) analyser.

The mechanism of the MALDI ionization is not completely understood but it is believed that the matrix absorbs most of the energy and assists desorption and ion formation. Basically MALDI ionization of peptides yields singly charged ions [50]. Following ionization, the ions are accelerated towards the analyser by a difference of potential (V_s) between an electrode and the extraction grid. The ions acquire a defined kinetic energy (E_k) given by the electric potential energy (E_{el}):

$$E_k = \frac{1}{2} \cdot (m \cdot v^2) = q \cdot V_s = E_{el} \quad (1)$$

Where m is the mass of the analyte and q is the total charge of the ion defined as:

$$q = z \cdot e \quad (2)$$

The time of flight (t) of the ion in the mass analyzer depends on its velocity (v) and the pathway length of the TOF (L) which is constant:

$$v = L / t \quad (3)$$

When the ions enter the field-free region they are separated according to their velocities. Because lighter ions travel faster, they reach the detector earlier than

heavier ions. In the time-of-flight (TOF) analyzer mass-to-charge ratios (m/z) are determined by measuring the time that ions take to reach the detector. Substituting (2) and (3) in (1) yields:

$$m/z = t^2 \cdot (2 \cdot V_s \cdot e / L^2) \quad (4)$$

Because the terms in parentheses are constant, m/z can be calculated from a measurement of the time of flight.

In order to improve mass resolution TOF analyzers are equipped with an electrostatic reflector also called reflectron (Figure 4). The reflectron is situated opposite to the ion source and creates an electric field that slows down the ions and sends them back to the TOF. Ions of same m/z but leaving the source with different E_k will be dispersed and enter the reflectron at different times. However the ions with higher E_k penetrate deeper in the reflector and leave the reflectron focused with ions of the same m/z thus reaching the detector at the same time. The reflectron corrects the kinetic energy dispersion, thus improving mass resolution [51].

MALDI TOF/TOF

Recently, a new generation of MALDI mass spectrometers have been developed to perform high-throughput MS/MS of peptides by collision induced dissociation [52-54]. Here, the performance of the 4800 MALDI TOF/TOF instrument commercialised by Applied Biosystems is described. The spectrum acquired in MS mode supplies a measurement of the ions present in a spot and makes possible the selection of the precursors for the subsequent MS/MS analysis. In MS/MS mode, the precursor is isolated before fragmentation. The desired ion travels through a first short linear TOF and is filtered by the Timed Ion Selector (TIS) that consists of two ion gates in series. The first gate deflects low mass ions by an applied voltage. When the selected precursor arrives, the gate opens (the deflection electrode is set to ground potential) to let the ion pass and the second gate closes after it to deflect the high mass ions. The precursor entering the collision induced dissociation (CID) cell is decelerated to 1 kV and fragmentations are induced by high-energy collisions (1-2 kV) with air molecules or a neutral gas like Ar, He or N₂. Ions leaving the CID region are accelerated in the second source to the second TOF analyser, which is a reflectron TOF, and are finally detected [55].

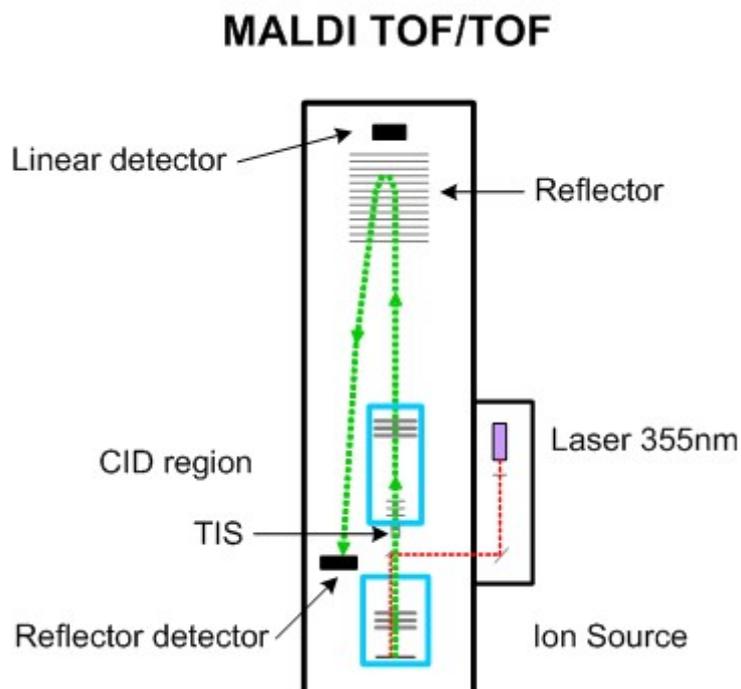


Figure 4: Schematic representation of a MALDI TOF/TOF mass spectrometer. The main difference with a standard MALDI TOF mass spectrometer is the presence of the TIS and the CID region.

It is also possible to induce peptide fragmentation without employing a gas in the CID region. Peptides that have acquired enough internal energy during the ionization process can spontaneously experience metastable decay in the field-free region. Because this fragmentation, also named post-source decay (PSD), occurs after acceleration, the fragment ions have the same velocity as its precursor and are selected together by the TIS. These fragments are then accelerated in the second source and due to their different masses are separated in the second TOF before arriving at the detector [51].

1.4.2. Electrospray mass spectrometry

ESI is, together with MALDI, the method of choice for soft ionization of proteins and peptides because it does not cause decomposition of biomolecules. In ESI, the analyte is dissolved in a solvent liquid and flows through a capillary tube to a needle where a high voltage is applied. Due to the electric field on the needle tip, the eluting liquid comes out as a fine mist of electrically charged droplets (electrospray) at atmospheric pressure. A potential drop (3-6 kV) between the needle and the inlet of the mass

spectrometer separated by 0.2-0.3 cm, guide the droplets towards the entrance of the mass spectrometer. The droplets pass through a curtain of heated inert gas or through a heated capillary to remove solvent molecules [51]. Solvent evaporation from the droplets occurs until the Coulomb repulsion between ions of the same sign becomes of the same order as the surface tension. At this point the Rayleigh limit is achieved and the droplet explodes generating smaller droplets; this process occurs several times until ions are in the gas-phase. To explain the mechanism of gas-ion formation two theories have been proposed. The charged residue model (CRM) proposes that evaporation of the solvent continues until one solute molecule remains and retains the charge to become an ion. The ion evaporation model (IEM) suggests that before the one solute molecule is achieved the charge density of the droplet is so high that some of the surface ions are ejected into the ambient gas [48, 56]. ESI differs from MALDI not only in the mechanism of ionization but also the type of ions generated as it is possible to observe multiple charged ions. Ions obtained from large molecules carry as many charges as the number of ionisable sites available. Due to the relation mass-to-charge (m/z), multiple charged ions allow analysis of high molecular mass compounds using a conventional mass analyzer.

In the proteome analysis the ESI ion source is often used with an Ion-Trap (IT) analyzer. Ions produced in the source are focused through a skimmer and two RF-only octapoles (see Figure 5). The vacuum in the trap is ensured by differential pumping between the ion source at atmospheric pressure and the mass analyzer. The IT consists of a cylindrical ring electrode and two end-cap electrodes to create a three-dimensional electric field. The ions' trajectories are governed by a radio frequency (RF) voltage and a helium bath gas. The ions are stored in the trap and ions of different m/z are consecutively ejected by ramping the RF to make ions trajectories sequentially become instable. For MS/MS analysis all ions except the precursor ions are ejected. The isolated m/z is excited by application of an excitation waveform to the end-cap electrodes, increasing the collisions with the He molecules. Finally the fragments are ejected according to their m/z and detected.

1.4.3. Liquid chromatography-Mass spectrometry

Interfacing the chromatographic and spectrometric components is fundamental for high-throughput proteome analysis and automatic acquisition of data. Because the

electrospray is formed directly from a solvent flow, ESI is preferably hyphenated with RP-HPLC for analysis of complex samples. The character of the organic solvents used in reversed phase chromatography and the absence of salts makes it fully compatible with ESI. The electrospray is formed at atmospheric pressure and the ions enter the mass spectrometer in the gas phase. This allows direct hyphenation of LC with ESI creating a robust on-line system (see Figure 5). One limitation of the on-line coupling LC-MS/MS in the analysis of complex samples is that peptide elution is faster than most mass spectrometers can collect MS/MS spectra [57].

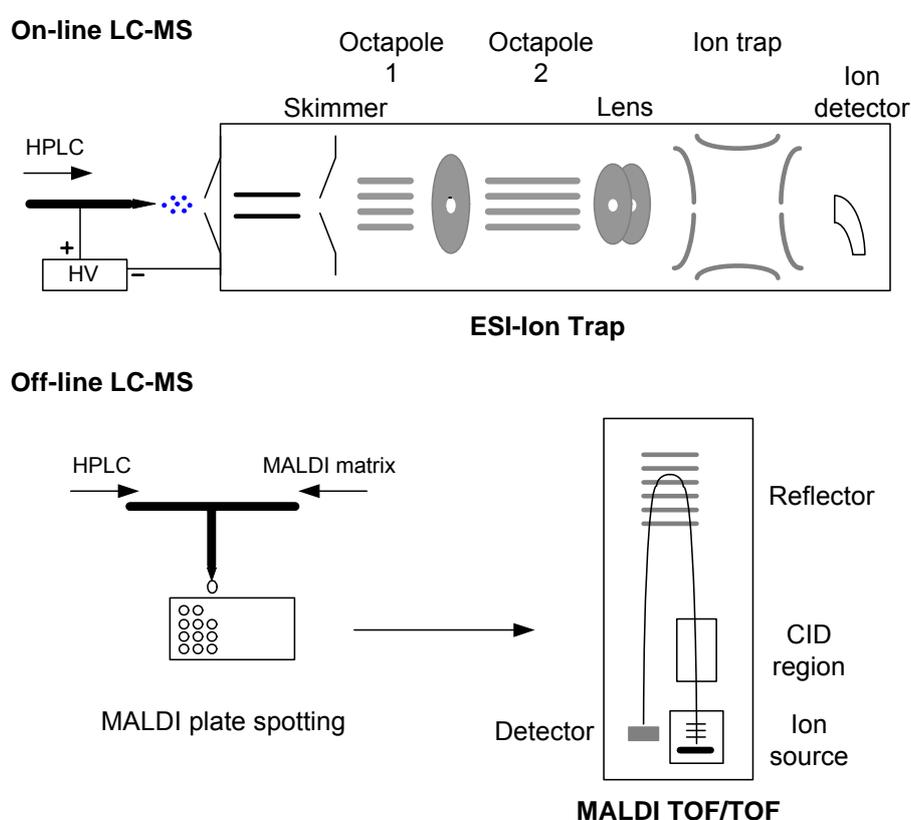


Figure 5: Hyphenation of liquid chromatography and mass spectrometry. On-line coupling of liquid chromatography and ESI-IT mass spectrometer (top). In the bottom figure is illustrated the off-line process that fixes the chromatographic separation on a MALDI plate followed by posterior analysis by MALDI TOF/TOF MS.

Dynamic exclusion is generally applied to avoid missing low-intensity peptides. If the number of co-eluting peptides is higher than the number of MS/MS acquirable in one-data analysis cycle. When low abundant peptides co-elute with high abundance peptides their intensity in the spectra exceed the dynamic range of the mass spectrometer and the weak ions are not observed [58]. Besides the MS/MS spectra

might not be acquired at the maximum of the chromatographic peak, producing a loss of sensitivity [59].

As mentioned before, recent developments in MALDI TOF mass spectrometry have permitted high-throughput MS/MS analysis of peptides. Coupling liquid chromatography with MALDI is therefore an attractive possibility for the proteome analysis and an alternative to the common LC-ESI combination. Nevertheless, MALDI ionization takes place in the mass spectrometer under vacuum and requires an off-line coupling (see Figure 5). In a LC-MALDI analysis, the chromatographically resolved peptides eluting from the column are mixed on-line with the matrix and deposited on the MALDI-plate. Separated peptides are thus immobilized on the MALDI-plate and the off-line mass spectrometric analysis is not restricted by the chromatographic process. Furthermore, the precursor list can be elaborated after complete MS acquisition in order to generate *in silico* ion chromatograms for each detected mass. MS/MS acquisition is thus optimized to perform MS/MS analysis at the maximum of the chromatographic peak, increasing the sensitivity. Besides the MALDI mass spectrometry shows a high tolerance to the presence of salts and detergents in the sample. For these reasons the off-line coupling of the LC separation and MALDI mass spectrometry provides fascinating possibilities for the proteome analysis.

1.4.4. Peptide fragmentation and tandem mass spectrometry

MS/MS or tandem mass spectrometry combines the information of the precursor mass and the mass of fragments derived from it for protein identification. The collision induced fragmentation of a protonated peptide in the mass spectrometer occurs by collision with air or neutral gas molecules (N₂, He, Ar). Depending on the collision energy of the precursor ion the CID is classified in high-energy CID, range of 1-2 keV as in MALDI MS/MS, or low-energy collisions in the range of 100 eV as in an ESI-IT or in a triple quadrupole. A peptide sequence is represented from the amino terminus (N-terminus) to the carboxyl terminus (C-terminus) as shown in Figure 6 A. Fragmentation occurs preferably at the peptide bond between two amino acids. The cleavage of one peptide bond generates two fragments containing either the N-terminus or the C-terminus of the peptide. The nomenclature of peptide fragmentation [60, 61] identifies the fragment containing the N-terminus from the original peptide as b-ion, whereas the fragment containing the C-terminus is called y-ion (see Figure 6 A). The b-type ions are all fragments containing the N-terminus derived from all

possible fragmentations at the peptide bonds. Similarly are defined the y-type ions, which comprise all possible C-terminus peptides. Internal fragments or fragments from the series x-, z-, a-, c- are more common in high-energy CID [62].

In the mass spectrometer, collision induced dissociation yields several fragments from all possible peptide bond fragmentation but only fragments retaining a charge can be detected. In a MS/MS mass spectrum the mass difference between two consecutive ions from the same series correspond exactly to the mass of the amino acid that differentiates them (Figure 6 B).

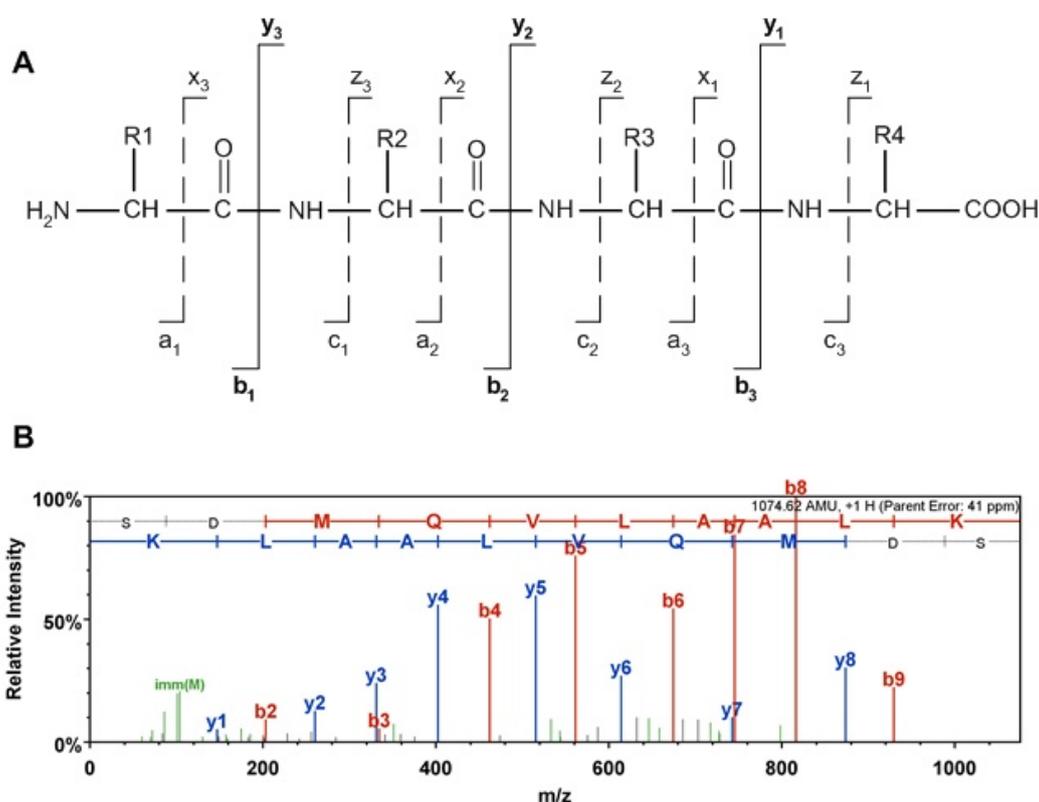


Figure 6: (A) Nomenclature for peptide fragmentation. (B) Spectrum generated by fragmentation of the peptide SDMQVLAALK by MALDI TOF/TOF with CID. Fragments corresponding to b-series and y-series are labelled. The difference between two ions of the same series permits calculation of the amino acid between them and thus peptide sequencing.

High-energy CID generally yields immonium ions $[H_2N=CH-R]^+$ from some individual amino acids in the peptide. These ions can be observed in the low mass region of the spectrum and can assist in the sequence assignment by confirming the presence of an amino acid. In the MS/MS spectrum in Figure 6 the immonium ion of methionine is observed at m/z 104.

1.5. Protein identification by database search

The identification of proteins is carried out by matching the acquired spectrum with data of a database containing the sequenced genome of the studied organism. There are two possible database searches that can be used for protein identification: peptide mass fingerprint (PMF) or MS/MS. Protein identification by PMF uses the masses of the peptides generated by digestion of a single protein. Because each protein sequence is unique, the digestion yields a distinctive “fingerprint” of peptides, which is compared with the theoretical digest of all proteins present in the database. Protein identification by tandem mass spectrometry or MS/MS uses the information generated by fragmentation of a peptide by collision induced dissociation (CID) or post source decay (PSD). The database search is performed by an algorithm that compares the experimental data with data calculated from a protein or nucleotide database until it matches the spectrum to an amino acid sequence. Moreover, the algorithm applies the cleavage rules of the enzyme used in the proteolysis of the proteins. The most widely used programs for correlation of spectra with theoretical data are Mascot [63] and Sequest [64]. In the present work Mascot, a search engine which incorporates a probability-based scoring, has been used. To perform a search the user has to define some parameters like the enzyme employed for digestion, the error window on the measured mass values, modifications expected or number of missed cleavages. The search engine returns a list of best matches to a spectrum and a score that permits determining if a result is significant or not. In proteomics large datasets of MS/MS data are submitted for database search and the rate of false-positives associated with the search can be calculated. Estimation of the false positive rate (FPR) of the identifications can be performed by searching in a composite “target-decoy” database [65]. While the “target database” is the one used for the routine identifications, the decoy database can be either the reversed or randomised of the “target” [28, 66]. The false positive rate can be estimated by doubling the number of decoy hits and dividing this by the total number of hits (decoy and target hits).

1.6. *Corynebacterium glutamicum*

Corynebacterium glutamicum is a gram-positive bacterium (Figure 7) of economic interest in biotechnology, it grows fast without forming clumps and it is a GRAS (generally recognized as safe) organism [67]. It belongs to the suborder Corynebacterineae, which includes also the pathogenic bacteria *Corynebacterium diphtheriae*, and the family related *Mycobacterium leprae* and *Mycobacterium tuberculosis*. Together with other well-known organisms like *Escherichia coli* or *Saccharomyces cerevisiae*, *C. glutamicum* is of industrial relevance and is therefore intensively studied.



Figure 7: *Corynebacterium glutamicum* wildtype cells. The cells are typically rod form cells (Source: J. Krömer).

The feature which has made *C. glutamicum* especially interesting in the last 50 years is its ability to produce amino acids. In 1957, Kinoshita found that *C. glutamicum* was able to secrete L-glutamic acid, which is used in its sodium glutamate form as a powerful food taste enhancer. The possibility of biotechnological production of L-glutamate was exploited by the industry and the interest in this bacterium increased. As a consequence, new strains were developed by random mutagenesis (chemical mutagenesis or UV irradiation), making possible large scale production of L-glutamate and L-lysine [68]. L-lysine is an essential amino acid usually present in the diet. However, manufactured cattle foodstuff lacks of this amino acid and therefore the addition of L-lysine is necessary to promote animal growth. Other amino acids produced industrially by *C. glutamicum* but to a less extent include L-threonine and L-glutamine [69]. One advantage of the biotechnological amino acid production is that it is a highly selective process and that yields exclusively the physiologically active L-isomer, whereas the production based on chemical methods yields a racemic mixture. Additionally, biotechnological processes promise easy waste disposals and consumption of renewable resources.

The complete genome of *C. glutamicum* was published in 2003 [70, 71] by two different institutions: Kitasato University (Japan) and the company Kyowa H.K., and the University of Bielefeld (Germany) in collaboration with the company Degussa. The genome size of *C. glutamicum* has 3.3 Mbp and the predicted open reading frames (ORFs) reported by Kitasato and Bielefeld Universities were 2993 and 3057 respectively. The information about the metabolic network and genome of *C. glutamicum* has permitted rational design of new strains by metabolic engineering (Ohnishi et al., 2002).

The proteome of *C. glutamicum* has been studied extensively by 2D-PAGE followed by identification by mass spectrometry [72-74]. The number of spots observable on a Coomassie Blue stained 2-gel of the cytosolic proteome was reported to be between 800 and 900. Based on this methodology Schaffer et al. reported the identification of 147 different soluble proteins and 13 membrane-associated proteins [74]. Moreover, one study on the closely related *Corynebacterium efficiens* based on 2D-PAGE and mass spectrometry identified up to 177 different proteins [75]. Recently, 2D-PAGE and identification by MS has been the method of choice to confirm results from transcriptome analysis of *C. glutamicum* cells grown on citrate [76]. While cytosolic proteins with pI 3-7 appear well resolved in 2D-gels, basic proteins with pI > 7 are underrepresented [74]. Membrane proteins are difficult to analyse by 2D-PAGE because they show low solubility and aggregate during IEF focusing. Therefore diverse strategies have been developed in order to identify the membrane proteome of *C. glutamicum*. The combination of anion-exchange chromatography followed by SDS-PAGE permitted the identification of 50 membrane-integral proteins [77]. Alternatively, a new protocol based on a predigest of isolated bacterial membranes of a lysine producing strain with trypsin, followed by digestion with trypsin /chymotrypsin in 60% methanol delivered the identification of 326 integral membrane proteins [78] in addition to 1181 cytosolic proteins. An approach using SDS-PAGE separation of *C. glutamicum* proteome followed by enzymatic digestion of individual gel bands prior to LC-MALDI analysis, reported the identification of 350 non-redundant proteins [79]. Recently, the combination of two-dimensional reversed phase chromatography and detection with an ESI-IT delivered the identification of 745 non-redundant cytosolic proteins [1]. Nevertheless, these numbers of identified proteins are still far from the roughly 3000 gene products predicted and improvements in the proteome analysis are necessary in order to achieve higher proteome coverage.

2. Goal of the thesis

Proteomics or proteome analysis is still a young research area of the bio-sciences. The study of the whole proteome of an organism is a huge task which requires the combination of different analytical techniques. Classically the separation of proteins has been achieved by 2D-PAGE. Alternatively the proteome can be digested yielding a complex mixture of peptides, which requires the combination of several chromatography stages for peptide separation. The standard set-up is a two-dimensional combination of SCX and RPC. The identification of the proteins is achieved by accurate mass spectrometric analysis of the peptides and/or peptide fragments.

The aim of this work was to develop an analytical method for the proteome analysis and more concrete to improve the methodology for the study of *C. glutamicum*. Because it is used for large-scale production of amino acids, *C. glutamicum* is a relevant organism in the industry. The importance of the achievements of the present work relies not only on the application of the method for general proteome analysis but also, from the biotechnological point of view, on the knowledge gained about this amino acid producing organism. The novel combination of two-dimensional RPC at acidic and basic pH with high-throughput analysis by LC-MALDI TOF/TOF is a very promising tool for the proteome analysis. In the first part of the thesis the establishment and optimization of the LC-MALDI technique for the proteome analysis is described. Further on, the results from the investigation of the proteome of *C. glutamicum* by means of 2D-RP-HPLC combined with MALDI MS/MS are presented. In the next part of the thesis the method is evaluated by comparison with two classical approaches for the proteome analysis as illustrated in Figure 8. In the first case the separation takes place at the protein level by 2D-PAGE. The LC-MALDI approach is compared with the standard 2D-PAGE followed by mass spectrometric analysis with the MALDI TOF/TOF instrument. In the second case the same separation is followed by a different mass spectrometric detection. Finally, the results are compared with results of previous proteome studies on *C. glutamicum* in the literature. The success obtained on our model organism *Corynebacterium glutamicum* suggests that the established method is also suitable for other organisms.

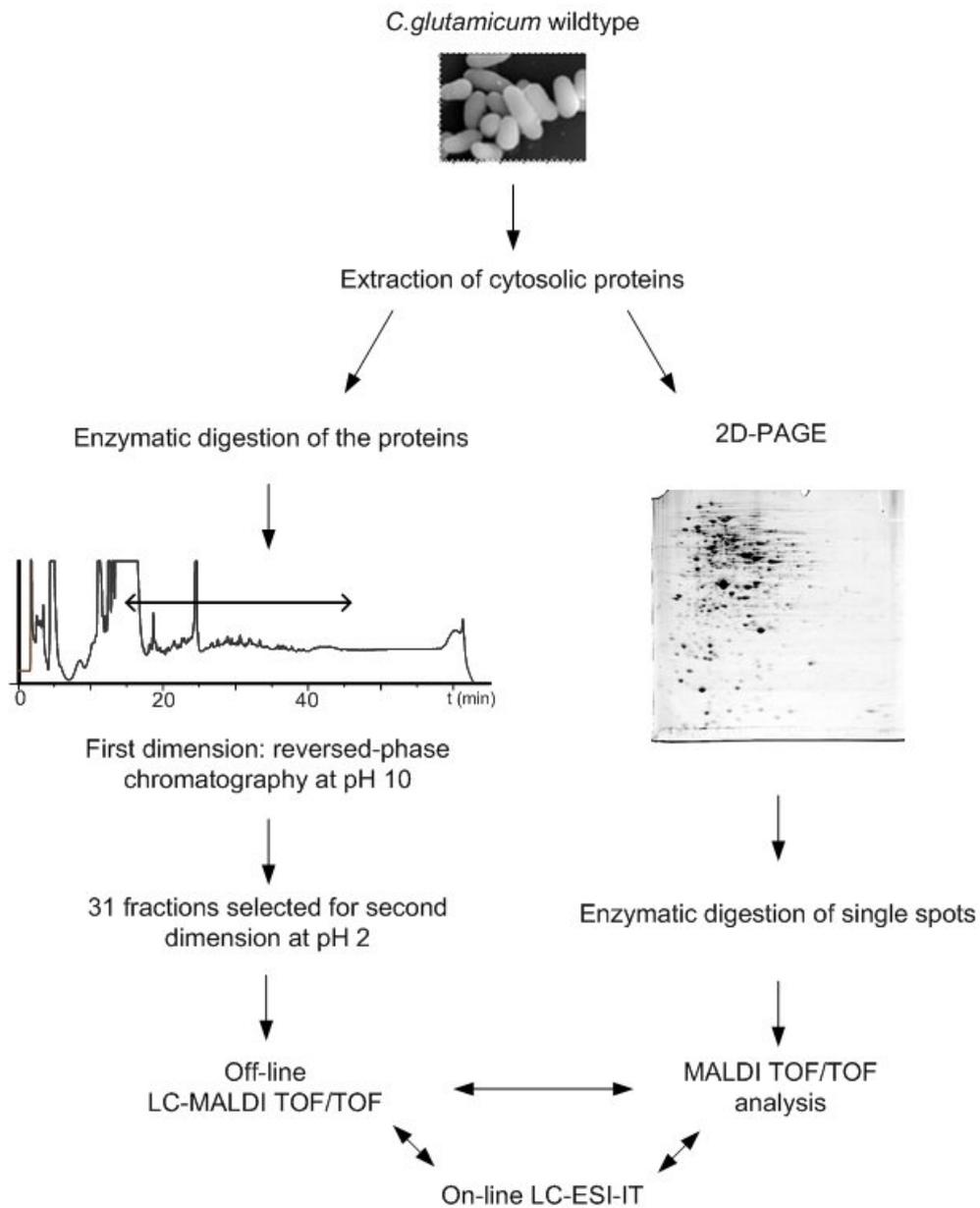


Figure 8: Strategy employed in this work for the proteome analysis of *C. glutamicum*. The digestion of the proteins extracted from the organism followed by the novel combination of 2D-RP-HPLC pH based and MALDI TOF/TOF MS is shown on the left side. This approach is compared with the 2D-PAGE approach shown on the right side and with the result obtained by using a ESI-IT mass spectrometer.

3. Materials and methods

3.1. Materials

Ethylenediamine-tetraacetic acid disodium salt (EDTA, Approx. 99%), zwitterionic detergent 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS, for electrophoresis, min 98%), ammonium persulfate (min 98%), bromophenol blue sodium salt, sodium thiosulfate pentahydrate, the MALDI matrix α -cyano-4-hydroxycinnamic acid (CHCA, powder, $\geq 98\%$), the peptide for internal calibration [Glu¹]-fibrinopeptide B human ($\geq 90\%$) and all chemicals employed for chromatography (of HPLC grade) were obtained from Sigma-Aldrich (Steinheim, Germany). Thiourea (analysis grade), N,N,N',N' tetramethylethylenediamine (TEMED, min 99%), Coomassie brilliant blue for electrophoresis, ammonium hydrogencarbonate (extra pure) and ammonium sulfate were purchased from Merck (Darmstadt, Germany). Urea ($\geq 99.5\%$), Tris-(hydroxymethyl)-aminomethane ($\geq 99.9\%$), sodium dodecyl sulfate (SDS, $\geq 99.9\%$ for electrophoresis), acrylamide solution Rotiphorese[®] Gel 30 were purchased from Carl Roth GmbH (Karlsruhe, Germany). The protease inhibitor cocktail Complete Mini employed for protein extraction was purchased from Roche (Mannheim, Germany). Sodium carbonate (99.5%) was provided by Grüssing (Filsum, Germany). Chemicals employed for the 2D-PAGE like glycerin (99%, water free), iodoacetamide (for synthesis, $> 98\%$), ethanol (99%), acetic acid (99.5 - 99.8%), methanol (for synthesis), phosphoric acid (85%), silver nitrate (for analysis), formaldehyde for synthesis were obtained from Chemical Storage Facility of the University of Saarland. Dithiotreitol (DTT, molecular biology grade, $\geq 99\%$) was purchased from MBI Fermentas (St. Leon-Rot, Germany). The precast immobilized pH gradient Ready Strips[™] IPG and the carrier ampholytes Bio-Lyte 3-10 for the electrophoresis were purchased from Bio-Rad Laboratories (Hercules, CA, USA). Trifluoroacetic acid (TFA, $\geq 99.5\%$) was purchased from Fluka (Steinheim, Germany). Sequencing grade modified trypsin was obtained from Promega (Madison, WI, USA). All solvents were HPLC grade and water milliQ was prepared with a Millipore water purification system (Bedford, MA, USA).

3.2. Cultivation of *Corynebacterium glutamicum*

C. glutamicum wild-type strain ATCC 13032 (American Type and Culture Collection) cells were taken from -80 °C frozen stock culture and grown in agar plates at 30 °C. A single colony was taken for the first preculture and cells were grown for 8 h at 30°C in 50 mL complex medium (5 g L⁻¹ glucose, 5 g L⁻¹ yeast extract, 10 g L⁻¹ g tryptone, 5 g L⁻¹ NaCl) in 500 mL shake flask. Cells were harvested by centrifugation (8800 × g, 2 min, 30 °C) washed twice with sterile 0.9% NaCl and resuspended in 25 mL minimal medium for the second preculture in 250 mL shake flasks. Main cultivation was performed in minimal medium at 30°C and 230 rpm in an incubator shaker (Multitrom, Infors AG, Bottmingen, Switzerland) and cells were harvested at optical density (OD₆₆₀) 14 measured as described previously [80] when the growth rate reached $\mu = 0.37 \text{ h}^{-1}$ corresponding to 4.9 g L⁻¹ biomass dry weight. Minimal medium containing 15 g L⁻¹ glucose was prepared by combination of the following solutions after autoclaving: (A) 1 g NaCl, 55 mg CaCl₂ and MgSO₄ · H₂O in 559 mL deionized water; (B) 5 g (NH₄)₂SO₄ in 200 mL deionized water; 16 g K₂HPO₄ and 2 g KH₂PO₄ in 100 mL deionized water; (C) 20 mg FeSO₄ · 7 H₂O, pH 1 with HCl; (D) 15 g glucose in 100 mL deionized water; (E) 20 mL of vitamin solution (2.5 mg biotin, 5 mg thiamin HCl, 5 mg Ca-panthotenate in 100 mL deionized water); (F) 10 ml of 100X trace elements solution (200 mg L⁻¹ Fe₃Cl 6H₂O, 200 mg L⁻¹ MnSO₄ H₂O, 50 mg L⁻¹ ZnSO₄ H₂O, 20 mg L⁻¹ CuCl₂ 2H₂O, 20 mg L⁻¹ Na₂B₄O₇ 10H₂O, and 10 mg L⁻¹ (NH₄)₆Mo₇O₂₄ 4H₂O; and (G) 1 mL of a 10 mL DHB solution (30 mg 3,4-dihydroxybenzoic acid and 0.5 mL NaOH 6 M). Cells were harvested by centrifugation at 4°C for 5 min (8500 rpm, HFA 8.50 Highconic™ Rotor, Biofuge Stratos Sorvall), washed twice with 1 mL sterile 0.9% NaCl and stored at -75 °C.

3.3. Protein extraction from *Corynebacterium glutamicum* cells

Cell wall disruption was performed mechanically with a mixer mill (MM 301, Retsch GmbH, Haan, Germany). Approximately 200 mg (wet cell weight) of *C. glutamicum* cells were resuspended in lysis buffer 1:3 (20 mM TRIS, 5 mM EDTA, pH 7.5 and protease inhibitor cocktail). The vial and lysis buffer were kept in ice at 4 °C. About 0.8 g of glass beads of 0.25 - 0.5 mm external diameter was added. The vial was inverted 3 - 4 times to ensure homogeneity and placed in the mixer mill for cell

disruption. The mixer mill was set to maximal vibration intensity of 30 Hz (1800min⁻¹) for 15 min. After that, the vial was centrifuged (13000 × g, 5 min, 4°C) and the supernatant was collected in a fresh vial. Again fresh lysis buffer was added and the extraction procedure repeated for five times. The resulting supernatant containing the cytoplasmic proteins was centrifuged for 30 min at 50000 x g and 4°C to remove insoluble rests from cell debris. Protein concentration was determined by Bradford test (Bio-Rad Protein Assay, Bio-Rad Laboratories GmbH, Munich, Germany) [81] and the sample was stored in aliquots at -20 °C.

3.4. Two-dimensional liquid chromatography

3.4.1. Digestion of cytosolic proteins

About 300 µg of total protein were denatured in presence of 3M urea and 187 mM ammonium hydrogencarbonate for 1 hour at 37 °C. The disulfide bridges were reduced in 11 mM DTT for 2 h at 37°C and 600 rpm shaking in a Thermomixer comfort (Eppendorf). The solution was degassed and cysteines were alkylated in presence of 50 mM iodacetic acid for 30 min in a dark place. Excess of iodacetic acid was quenched by addition of β-mercaptoethanol to a 45 mM final concentration and the sample was incubated for 20 min at room temperature. The sample was dialyzed overnight in a dialysis cassette (Slide-A-Lyzer® 3500 MWCO, Pierce) against one liter distilled water. Proteins were digested with trypsin (Promega, Madison, WI, USA) in a ratio 1/50 (w/w) for 24 hours at 37 °C. Trypsin was dissolved in 50 mmol/L acetic acid and incubated for 30 min at 30 °C before digestion. Digestion was stopped by addition of TFA to 1% (v/v) end concentration.

3.4.2. First dimension: reversed-phase HPLC at pH 10

Peptide separation at high pH was performed on a Gemini C18 column, 150 mm x 2 mm, 3 µm, 110 Å (Phenomenex, Aschaffenburg, Germany). About 280 µg of digested peptides were manually injected with a 400 µL loop onto the column at 200 µL/min with eluent A (72 mM aqueous triethylamine, pH 10 adjusted with acetic acid). Peptide elution was accomplished with a gradient of 0 - 55 % B in 55 min using a low-pressure gradient pump (Rheos 2000, Flux Instruments, Basel, Switzerland) equipped

with a degasser (Knauer GmbH, Berlin, Germany) and an injection system (Model 7725, Rheodyne, Rohnert Park, CA, USA). Eluent B consisted of 72 mM triethylamine, 52 mM acetic acid in acetonitrile. Detection of peptide elution was performed at 280 nm. Fractions were collected every minute (approx. 200 μ L) and subsequently acetonitrile was removed by evaporation to 1/10 of the original volume in a vacuum concentrator (Model 5301, Eppendorf AG, Hamburg, Germany).

3.4.3. Fraction screening by MALDI TOF MS

In order to determine the fractions to be analysed by LC-MALDI, a rapid screening of all collected fractions was performed by MALDI TOF MS. A few microliters of each fraction were used to prepare sample dilutions. Dilutions 1:10 and 1:100 were prepared employing 0.1 % TFA. Each sample was spotted three times as follows: 0.5 μ L of the sample was deposited on a MALDI plate, followed by addition of 0.5 μ L of MALDI matrix (2.6 mg/mL CCA in 50% ACN, 0.1% TFA). Calibration was performed for the mass range 800 - 4000 Da and sample spectra were acquired under the following parameters: 50 shots per sub-spectrum and accumulation of 30 sub-spectra.

3.4.4. Second dimension: reversed-phase HPLC at pH 2.1

The selected fractions were reconstituted with 105 μ L of 0.10 % heptafluorobutyric acid previous to analysis and analysed in triplicate. A 10 μ L aliquot of each selected fraction was injected with a Famos/Ultimate HPLC instrument (LC Packings, Amsterdam, Netherlands) onto a PS-DVB trap column (10 mm x 0.2 mm i.d.) for peptide preconcentration and peptides were washed for 2.5 min with 10 % aqueous heptafluorobutyric acid at 10 μ L/min. Peptides were eluted onto the PS-DVB monolith column (60 mm x 0.1 mm i.d.) at flow rate 0.7 μ L/min. Mobile phase composition was (A) 0.05% TFA in H₂O, (B) 0.05% TFA in ACN. The gradient employed was a 60 min linear gradient from 0 to 20 % B followed by a 3 min isocratic elution at 100% B. The oven temperature was set to 25 °C and UV detection was performed at 214 nm.

3.4.5. LC-MALDI spotting

The effluent from the column was 0.7 $\mu\text{L}/\text{min}$ and the eluting peptides were mixed post-column with MALDI matrix in a ratio 1: 4 (v:v). The resulting flow rate for the spotting procedure was 3.5 $\mu\text{L}/\text{min}$. The matrix concentration used was 3 mg/ml in 70 % acetonitrile and 0.1% TFA aqueous solution. The spotting of the MALDI-plates was accomplished by a Probot microfraction collector (LC Packings, Dionex, Netherlands). Spots were deposited every 5 seconds on the MALDI-plate (Opti-TOF™ LC MALDI Insert, Applied Biosystems, USA) for 62 minutes. Spotting was started 12 minutes after injection corresponding to the delay time between the column and the spotter. The spots were distributed on the MALDI-plate in 32 rows and 49 columns which conferred a total capacity of 1568 spots per plate.

3.4.6. LC-MALDI TOF/TOF analysis

A MALDI TOF/TOF 4800 (Applied Biosystems, Darmstadt, Germany) mass spectrometer was used for acquisition and processing of the data. MS data from the spots was acquired in positive reflector ion mode in the mass range of 800-4000 m/z by accumulation of 1000 laser shots. In addition to the default calibration, MS spectra were internally calibrated with [Glu¹]-fibrinopeptide B (Sigma) added to the matrix to a concentration of 21 fmol/ μL . The list of precursors for MS/MS analysis was automatically generated by the instrument software according to the following selection criteria: minimum signal to noise ratio 35; precursor mass tolerance between spots \pm 200 ppm; minimum chromatogram peak of two spots and a maximal of six precursors per spot. An exclusion filter was used to eliminate the internal mass standard. MS/MS spectra were generated by one kilovolt collisions with air using a source voltage of 8 kV, a collision cell voltage of 7 kV and an accelerating voltage in the second source of 15 kV. Maximal 3000 laser shots were accumulated for a MS/MS spectrum. Stop conditions for MS/MS were defined as a minimal number of 10 peaks above 35 S/N with at least 12 accumulated sub-spectra.

3.4.7. Database search and protein identification

Raw files containing a peak list of the precursors and fragment ions were filtered with the Peak to Mascot tool (4000 Series Explorer Software, ABI) for generation of

MASCOT generic files. The settings for peak export were the following: maximal peak density of 20 peaks per 200 Da; mass range from 60 Da to the precursor mass minus 35 Da; minimal peak area 200; minimum signal/noise 10 and maximum 65 peaks per precursor. The filtered files were submitted for database search using MASCOT search engine [63] version 2.1 (Matrix Science, London, UK). The search was performed in *Corynebacterium glutamicum* ATCC 13032 Kitasato database (2993 protein coding genes), downloaded from the Comprehensive Microbial Research (CMR) public database of the Institute for Genomic Research (TIGR, www.tigr.org). The following search parameters were used: enzyme, trypsin; allowed missed cleavages, 1; fixed modifications, carboxymethylation of cysteine; variable modification, oxidation of methionine; mass tolerance for precursors was set to ± 50 ppm and for MS/MS fragment ions to ± 0.2 Da. The confidence interval for protein identification was set to $\geq 95\%$ ($p < 0.05$) and only peptides with an individual ion score above the identity threshold were considered correctly identified. Estimation of the false positive rate was calculated by MASCOT search in a composite database created by appending a random database to the normal database thus generating a database twice the size of the original [65, 66]. The number of hits in the random database was doubled and divided by the total number of hits (random and target) as described in [66].

3.5. Two-dimensional gel electrophoresis

3.5.1. First dimension: isoelectric focusing

The protein solution was diluted with rehydration buffer (8 M urea, 2 M thiourea, 4% CHAPS, 40 mM DTT and 1% Bio-lyte 3-10) up to 350 μ L. The amount of protein loaded in the strip depended on the staining method: 200 μ g of protein for Comassie blue staining and 75 μ g of protein for silver staining. The IPG strips used for isoelectric focusing (pH 4-7, 17cm) were unfrozen at room temperature and rehydrated overnight with 350 μ L of the rehydration solution in the rehydration tray covered with a lid. The rehydrated strips were transferred to a focusing tray and covered with mineral oil to prevent solvent evaporation. Paper wicks wet in distilled water were placed between the end of the strips and the electrodes. Isoelectric focusing was performed in the Protean IEF cell (Bio-Rad) at 20 °C and a maximum current of

50 μ A/strip was allowed. The focusing conditions were as follows: 500V gradient for 1500 Vh, 500 V constant for 2500 Vh, 3500 V gradient for 10 kV h and 3500 V constant until 70 kVh were completed. The strips which were not directly used after focusing were stored at -20°C .

3.5.2. Second dimension: SDS-PAGE

The gels employed for the second dimension consisted of two parts: stacking gel (T 5%, 0.125 Tris, pH 6.8, 0.1% SDS) and separating gel (T 12.5%, 0.375 Tris, pH 8.8, 0.1% SDS). Gels were 0.1 cm thick, 18.3 cm wide and 19.3 cm long, and were freshly polymerized prior to use. The 30% acrylamide/bis-acrylamide (37.5:1) solution was purchased from Roth. The polymerization was started by addition of 10% aqueous ammonium persulfate solution prepared fresh (1/200, v/v) followed by addition of TEMED (1/2000, v/v) under stirring. The separating gel was prepared firstly and was allowed to polymerize for 30 min. The stacking gel was allowed to polymerize for a minimum of two hours. The focused strips were equilibrated for 15 min in two different SDS containing buffers by gentle shaking previous to the second dimension. The first equilibration solution contained DTT for reduction of the sulfhydryl groups (6 M urea, 0.375 M Tris HCl, pH 8.8, 2% SDS, 20% glycerol, 2% (w/v) DTT and Bromophenol blue). After removal of the reducing buffer, the strips were equilibrated in an iodoacetamide containing buffer for alkylation of the reduced sulfhydryl groups (6 M urea, 0.375 M Tris HCl, pH 8.8, 2% SDS, 20% glycerol, 2.5% (w/v) iodoacetamide and Bromophenol blue). After removal of the alkylating buffer the strips were shortly equilibrated in running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3). Second dimension was run in a Protean II xi Cell (Bio-Rad) with the following settings: 20 mA constant for 2 h and 22 mA constant for 10 h.

3.5.3. Gel staining

In order to remove rests of running buffer which may interfere with the staining (especially SDS) and fix the proteins, gels were fixed in an aqueous solution of 30% ethanol and 10% acetic acid for at least two hours. Before staining, the gels were washed in distillate water for one hour. Coomassie Blue staining is based on ionic and hydrophobic interactions between the dye molecules and the proteins at acidic pH. The

staining protocol used is called “Blue silver” [25] and is based on the Neuhoff’s procedure [82]. Gels were incubated in equilibrating solution (10% methanol, 2% phosphoric acid) for one hour under gentle shaking. After that the equilibrating solution was replaced by the staining solution (aqueous solution of 10% phosphoric acid, 10% ammonium sulphate, 20% methanol and 0.12% of Coomassie Blue G-250 (w/v)). The equilibration solution was prepared as described by Candiano et al. (2004) and the gels were incubated under gentle shaking for 24 hours. Before scanning, the gels were incubated in distilled H₂O to remove the staining from the background of polyacrylamide gel. Gels were stored in a 20% ethanol, 2% glycerol aqueous solution.

3.5.4. Gel image analysis

Gels were digitalized by scanning the gel and image analysis was performed with the ImagemasterTM 2D Platinum Software version 5.0 (Amersham Biosciences, Uppsala, Sweden). Spots selected for analysis were excised from the gel and the 1 mm thick gel plugs were either digested or stored in Eppendorf tubes at -20 °C.

3.5.5. In-gel digestion

The vials containing the spots were placed in a Thermomixer comfort (Eppendorf). The gel plugs were washed with 100 µL milliQ water for 5 min at 37 °C and 600 rpm. Next, they were washed with 100 µL of a solution water/acetonitrile (50/50) for 5 min at 37 °C and 600 rpm. In the next step the gel plugs were incubated for 1 hour in 50 µL (when necessary more volume was used for complete coverage of the spot) of a 10 mM DTT solution at 56 °C and 600 rpm. After reducing of the proteins, the gel plugs were washed with 100 µL milliQ H₂O for 5 min at 37 °C and 600 rpm. Alkylation was performed at room temperature by addition of 50 µL of a 55 mM iodacetamide solution in 40 mM ammonium hydrogencarbonate. Gel plugs were incubated for 30 min in the dark. In order to remove rests of dye the gel plugs were washed for 5 min alternatively with water and water/acetonitrile (50:50) at 37 °C and 600 rpm. When the dye was completely removed, the gel plugs were incubated for 1 min in 100 µL of 100% acetonitrile. After removal of the acetonitrile the gel plugs were allowed to dry for 15 min with the cap of the vial open. Digestion was started by addition of 5 µL (or 10 µL for big gel plugs) of a trypsin solution (Promega Sequencing Grade Modified

Trypsin, Promega Corporation, Madison, USA). The trypsin solution was prepared by dilution 1:30 with 40 mM ammonium hydrogen carbonate. If necessary to keep the gel plug in solution, a few microliters of 40 mM ammonium hydrogen carbonate solution were added. Digestion was carried out overnight by incubation at 37 °C and 600 rpm shaking.

3.5.6. Zip Tip™

If peptide concentration was very low for detection, purifying and concentration of peptides was achieved with 10 µL ZipTip_{C18} pipette tips. The C18 chromatography media bed was prepared by aspiration and dispense of 10 µL of 100% acetonitrile solution twice. Then the media was equilibrated with 10 µL of 0.1% TFA solution three times. For peptide binding the sample was aspirated and dispensed up to 10 cycles. Tip washing was performed with a 0.1% TFA solution and elution was accomplished with 1-4 µL of 50% acetonitrile 0.1% TFA solution. The sample was directly deposited on the MALDI-plate and mixed with MALDI matrix for analysis.

3.5.7. MALDI TOF/TOF analysis

The identification of proteins from in-gel digests was performed with a MALDI-TOF/TOF 4800 mass spectrometer (Applied Biosystems). 0.5 µl of digest was deposited on the plate followed by addition of 0.5 µl of matrix. The matrix consisted of 5 mg/ml of CCA in 50% ACN and 0.1% TFA. The default calibration was updated with standard peptides before starting the measurements. Spectra were acquired automatically in positive mode and a total of 1500 shots were accumulated per spectrum. Mass range was selected between 800 and 4000 m/z. The three strongest precursors showing S/N better than 50 in the MS spectrum were selected for further MS/MS analysis. MS/MS was performed with the 1kV positive operating mode and CID was off. A maximum of 2000 laser shots were accumulated per MS/MS spectrum. The acquisition of MS/MS spectra stopped automatically when the following conditions were fulfilled: at least 10 of the peaks in the spectra were above 35 S/N after accumulation of a minimum of 12 a sub-spectra.

3.5.8. Database search

The spectra acquired by MALDI-TOF/TOF were searched in MASCOT[®] v 2.1 with the GPS Explorer[™] software (Applied Biosystems). A combined search with MS and MS/MS data was performed in the database of *C. glutamicum* (downloaded from TIGR website) and NCBI, the enzyme was trypsin and maximal missed cleavage was set to 1. Two variable modifications were considered: methionine oxidation and cysteine carboxymethylation. The mass tolerance for the precursor was set to ± 50 ppm and for MS/MS fragments was set to ± 0.2 Da.

3.6. One-dimensional gel electrophoresis

The gels, 7.4 cm wide x 6.8 cm length, were cast in the Mini-Protean System of Bio-Rad. 10 ml of solution (T 12.5%) were required to cast two separating gels for SDS-PAGE separation. The solution was prepared by adding 3.2 ml distillate water, 2.5 ml of 1.5 M Tris pH 8.8 solution, 100 μ l of 10% SDS solution, 4.1 ml of 30% acrylamide/bisacrylamide (37.5:1) solution, 5 μ l of TEMED and 50 μ l of 10% ammonium persulphate solution. The gel was allowed to polymerize for 30 min. Following that, the collecting gels were prepared. The solution consisted of 3.05 ml distillate water, 1.25 ml of 0.5 M Tris pH 6.8 solution, 50 μ l of 10% SDS solution, 0.65 ml of 30% acrylamide/bisacrylamide (37.5:1) solution, 5 μ l of TEMED and 50 μ l of 10% ammonium persulphate solution. The comb for the wells was placed on top of each gel and polymerization was allowed for 10 min. The samples for analysis were incubated with SDS reducing buffer for 4 min at 95 °C. The SDS reducing buffer had the following composition: 62.5 mM Tris-HCl, pH 6.8, 20% glycerol, 2% SDS, 5% β -mercaptoethanol. 10 μ l of sample in reducing buffer were placed on each well of the gel for separation of the proteins by SDS-PAGE. The running buffer composition was 25 mM Tris, 192 mM glycine, pH 8.3 and 0.1% SDS. Running conditions were set at 150 volts for one hour. Gel staining was achieved in the following steps: 10-20 min in fixing solution (30% ethanol (EtOH) 10% HAc), 20 min in staining solution (0.2% Coomassie R-250, 0.005% Coomassie G-250 in 10% EtOH, 40% MeOH, 2% glycerin). To destain the background, gels were placed in 40% MeOH, 10% HAc, 2% glycerin solution.

4. Results

4.1. Optimization of the protein extraction procedure

The cell wall of *C. glutamicum* has an unusual composition because although *C. glutamicum* is a gram-positive bacterium, its cell envelope has an additional outer membrane like in gram-negative bacteria [83]. This outer membrane, rich in mycolic acids, plays a crucial role in the resistance of the related organism *Mycobacterium* to various antibiotics [84].

Cell disruption efficiency is of major importance for proteomic studies. Optimization of the method for protein extraction is a requisite, especially for quantification studies where total protein extraction, including low abundant proteins, is required. Cell disruption methods are usually classified in mechanical processes (e.g. Mixer Mill, Sonicator, French Press) and non-mechanical processes (chemical, biological or physical disruption). Mechanical processes are the methods of choice in the laboratory because they disrupt completely the cell wall and liberate the cytosolic proteins. Chemical disruption by addition of detergents or antibiotics was discarded in order to avoid contamination of the sample. Enzymatic lysis of the cell wall with egg white lysozyme was not considered because *C. glutamicum* is resistant to it [85]. One of the limitations in the proteome analysis is that the volume of sample available is usually small and therefore cell disruption with ultrasonic disruption or French Press are not suitable methods. For proteome studies, especially for quantitative purposes, a fast and efficient method yielding high concentrations in the desired buffer is required. Therefore an improved protocol was developed for isolation of the cytosolic proteome when working with small samples from *C. glutamicum* by mechanical grinding.

Cell harvest of *C. glutamicum* was carried out at OD 12. Around 10 mL of cell culture were harvested by centrifugation, followed by two washing steps with saline solution and transferred to a 2 mL vial. Subsequently cell pellets were mixed with 1.5 times lysis buffer (w:w), and 2.5 times glass beads (w:w). The cell wall was disrupted in a mixer mill by the shear forces generated during vibration and enforced motion of glass beads [86]. All the cell pellets employed for the study had a similar wet cell weight (WCW) which was approximately 300 mg. After a defined shaking time in the mixer mill, the vial was centrifuged to separate the supernatant from cell debris, glass beads

and unbroken cells. The collected supernatant was stored for analysis and fresh lysis buffer was added to the cell debris. Then the vial was inverted a few times to homogenize the mixture before the subsequent extraction step in the mixer mill. In total, the extraction protocol was performed consecutively up to ten times. Different shaking times – 1, 8, 15, 30 min - were investigated in order to determine the optimal conditions for extraction of *C. glutamicum* proteins. The extraction yield was controlled after each grinding step by running a SDS gel and measuring the protein concentration of the recovered supernatant [81]. The same bands were observed in all the SDS gels from collected supernatants of different grinding times (see Figure 9). The bands differ only in the intensity which is due to a different protein concentration in the collected supernatants. Thus no effect was observed due to selective release of proteins with respect to protein molecular mass, in dependence with the extraction time.

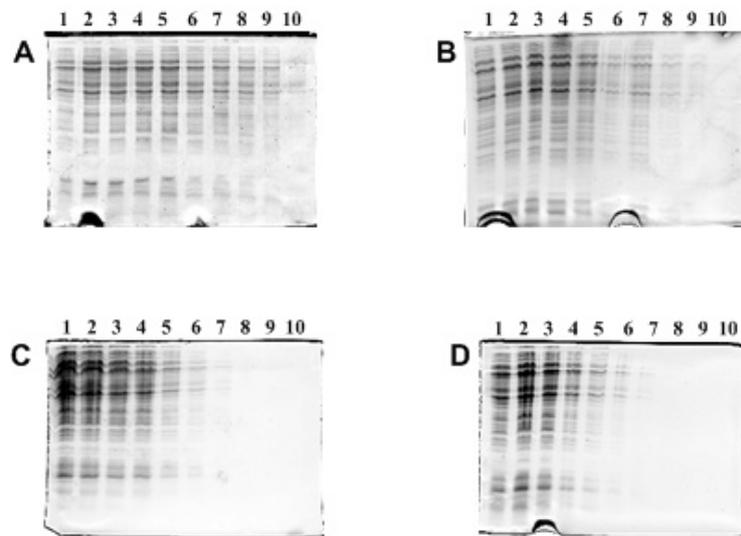


Figure 9: SDS-PAGE analysis of the supernatants collected after every disruption step for different grinding times: (A) 1 min, (B) 8 min, (C) 15 min, (D) 30 min. Cells were resuspended in lysis buffer (20 mM TRIS, 4mM EDTA, pH 7.5, proteases inhibitor) and mixed with glass beads. After a defined disruption time, the vial was centrifuged and the supernatant collected for SDS-PAGE analysis. On top of the gels the number of extraction in which the supernatant was collected is shown.

In Figure 10 the protein concentration of the supernatant collected after each consecutive extraction is plotted for the four different grinding times (1, 8, 15 and 30 min). It was found that short grinding times like 1 or 8 min, yield protein concentrations lower than 4.5 mg/ml per extraction step during the first 2-3 extraction steps. This is probably due to incomplete cell wall disruption. Increasing the grinding

time to 15 min improved the yield of the extraction up to 5-6 mg/ml per extraction step in the first four steps. Longer grinding times were also tested - 30 min - and the concentration of extracted protein in the first four extractions was slightly higher. It was observed that longer grinding times (15 min, 30 min) yielded higher concentrations with less extraction steps. Grinding times lower than 8 min were not sufficient for optimal cell wall disruption of *C. glutamicum*.

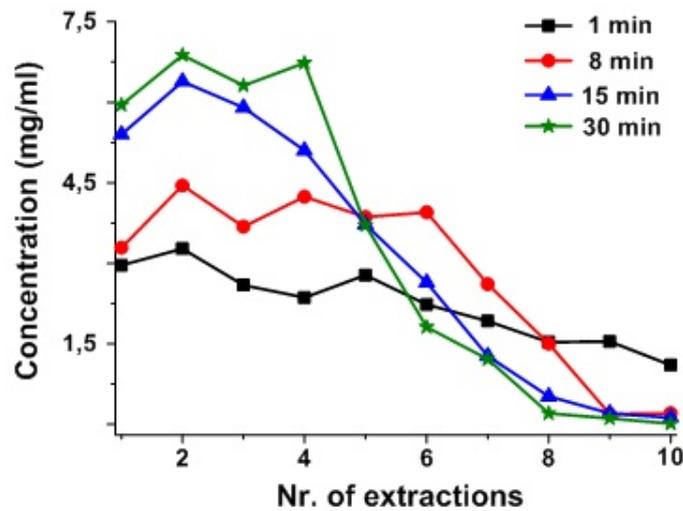


Figure 10: Protein concentration in the supernatant after the different extractions. Ten consecutive extractions were performed on *C. glutamicum* cell pellets for different grinding times in the mixer mill.

Extraction time periods of 15 min are basically better than 30 min because the improvement gained by a 30 min grinding over 15 min is very little, while the time needed for the extraction doubles. Besides that, longer grinding times generate additional sample heating. To avoid excessive heating of the sample during the grinding time, the plastic holders where the vials were placed, were stored at -20 °C before and between grinding steps. Sample temperature was checked after every grinding step. For grinding times of 15 min the sample temperature registered was in average 17 °C, while for 30 min grinding the average sample temperature raised to 25 °C. No significant improvement could be recognized when the experiment based on 15 min grinding was repeated in presence of more lysis buffer (1:3, cell pellet to buffer, w/w) data not shown.

Because complete filling of the vial negatively influences protein extraction [86], all experiments were carried out by filling the vial less than 3/4 of its capacity. Assuming

that all pellets had the same weight (300 ± 5 mg WCW) and that complete extraction was achieved after grinding ten times for 30 min, the percentage of extracted proteome accumulated after each consecutive extraction was calculated in dependence with the time and the results are shown in Figure 11.

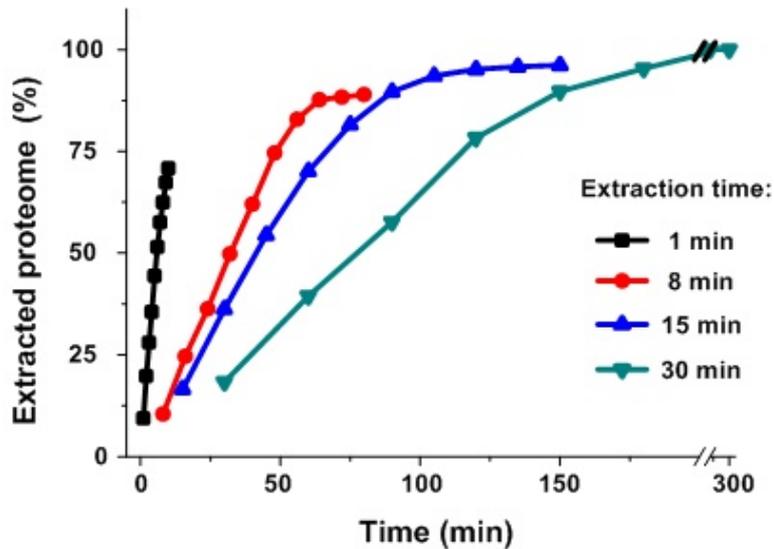


Figure 11: Percentage of total released proteome after each consecutive extraction in dependence with the total time employed for the extraction. Ten extractions were consecutively carried out for each of the four different grinding times.

It was assumed that the whole proteome was released after 10 consecutive extractions employing 30 min grinding times because the protein concentration determined for the last two extractions was close to zero. However, a 10 times 30 min extraction procedure represented a total extraction time of 300 min. Moreover, the 30 min extraction involves overheating of the sample. In contrast to that, the 1 min shaking time in the mixer mill provided less than 75% of the proteome after ten times. For an extraction time of fifteen minutes, it was possible to achieve a total extracted proteome of 81% after five consecutive extractions, in total 75 minutes shaking time. In order to achieve the same degree of extracted proteome with 8 min periods of shaking times, seven consecutive extractions and 56 min total time were necessary. The optimal conditions for extraction of the *C. glutamicum* cytosolic proteome were determined either 5 x 15 min or 7 x 8 min to gain 80% of the proteome. Higher extractions of the proteome were achieved for 10 x 8 min (89% and 80 min total time) or 7 x 15 min (93% and 105 min total time). More extractions or longer grinding times increase the total time unnecessarily and generate little improvement.

4.2. Evaluation and optimization of the LC-MALDI system

In a LC-MALDI experiment the chromatography is coupled off-line via deposition of the effluent of the column on a MALDI plate, the so called sample spotting. The subsequent mass spectrometric measurement is carried out by a MALDI TOF/TOF instrument. Consequently and in contrast to an LC-ESI-IT measurement, the time for MS acquisition in a LC-MALDI experiment is independent of the chromatographic process. However, a number of parameters are involved in a LC-MALDI experiment and their optimization is described in the next section.

4.2.1. LC-MALDI spotting set-up

The LC-MALDI spotting set-up employed in this work consisted of a HPLC system connected to an automatic fraction collector. The HPLC system was integrated by a Famos (LC Packings) autosampler equipped with a 10 μ L loop for sample injection. The trap-column for sample pre-concentration and the analytic column were monolith based. The outlet of the UV detector was connected to a microfraction collector (Probot, LC Packings) for deposition of the sample on the MALDI-plate (see Figure 12).

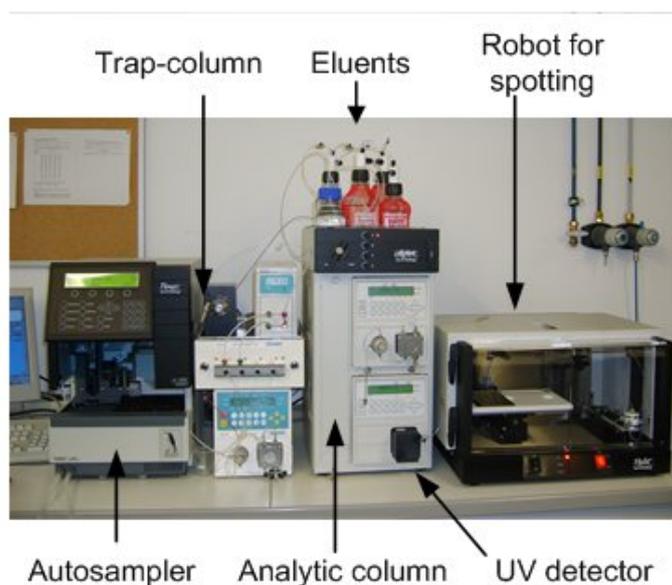


Figure 12: Set-up employed for LC-MALDI spotting. The sample was injected by the autosampler and trapped in the trap-column. Peptides were desorbed by an acetonitrile gradient and separated in the analytical column. The effluent was sent to the spotter and mixed on-line with the MALDI matrix.

The sample was injected by the autosampler on a PS-DVB trap-column (10 x 0.2 mm i.d.) which was mounted on a ten-port valve. The trap-column permitted pre-concentration of the peptides. The peptides retained in the trap-column were washed with a 0.1% heptafluorobutyric acid (HFBA) solution at 10 $\mu\text{L}/\text{min}$. HFBA was employed because it has been shown to enhance peptide retention during the washing step on the trap-column for peptide pre-concentration [34]. After 2.5 min the ten-port valve was switched (see Figure 13 A) and the gradient from the HPLC system flows in backflush mode into the trap-column at flow 0.7 $\mu\text{L}/\text{min}$. Desorbed peptides are, as a result, transferred to the PS-DVB monolith column (60 x 0.1 mm i.d.) where separation is accomplished by an acetonitrile gradient in 0.05% TFA at 25 $^{\circ}\text{C}$.

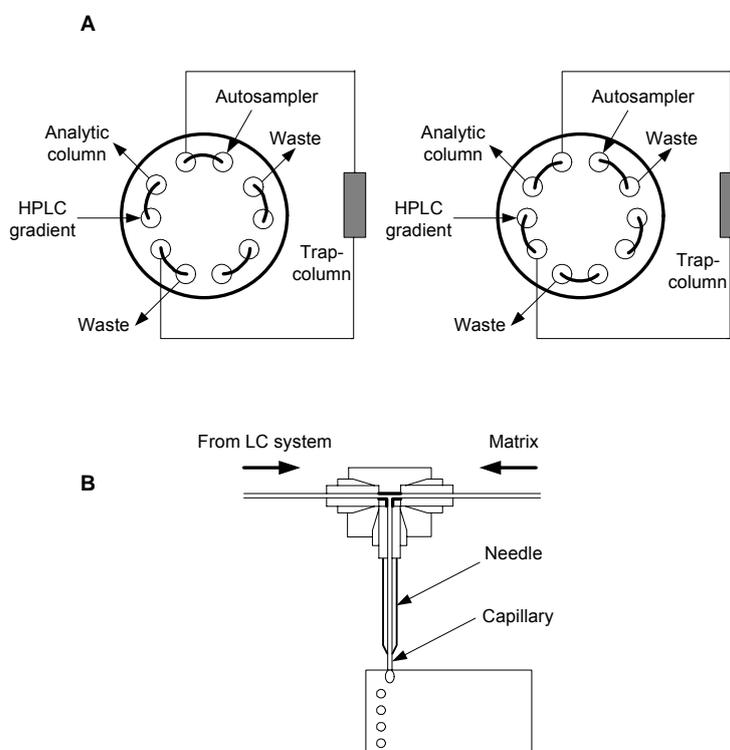


Figure 13: (A) Scheme of the ten-port valve employed for loading the sample on the trap-column after injection (left) and desorption of peptides trapped by the HPLC gradient followed by separation in the analytic column (right). (B) Automatic spot deposition on a MALDI-plate. Effluent from the column is mixed on-line with the matrix and deposited at regular intervals.

Peptides eluting from the monolith column pass the UV detector and reach a μ -Tee connection piece in which the column effluent is mixed on-line with MALDI matrix. The matrix employed in this work was α cyano-4-hydroxycinnamic acid (CHCA). The mixture effluent-matrix enters a needle in which a fused silica capillary is mounted

and directly deposited on the MALDI-plate at regular intervals (see Figure 13 B). The spot collection or plate spotting performed with a Probot microfraction collector is achieved by displacement of the table where the MALDI plate is placed.

4.2.2. Optimization of LC-MALDI

Optimization of the spot deposition

The optimal concentration of CHCA employed for manual spotting is about 5 mg/mL, mixed 1:1 (v/v) with the sample. As a result the matrix concentration on the plate is of 2.5 mg/mL and thus provides an optimal ionization. For the LC-MALDI application it was considered necessary to achieve the same matrix concentration on the plate. Despite a broad range of optimal matrix concentrations reported for LC-MALDI applications [87-89] in the range 3 - 7.5 mg/mL (70% acetonitrile, 0.1% TFA) it was decided to work with a low matrix concentration in order to avoid tube plugging due to possible matrix crystallization. For that reason the matrix concentration employed for automatic spotting was 3 mg/mL and was mixed on-line with the column flow rate effluent which was 0.7 $\mu\text{L}/\text{min}$. In Table 2 a list of possible ratios column flow rate to matrix flow rate is shown together with the final combined flow. In addition, the matrix concentration on the plate is shown for each ratio and the resulting spot volumes with two spotting times, 5 and 10 seconds, are considered.

Table 2: Flow rate after mixing the effluent from the column (0.7 $\mu\text{L}/\text{min}$) with matrix in different ratios. The matrix concentration on the plate and calculated volume per deposited spot (5 and 10 seconds spotting) are also shown.

<i>Ratio</i>	<i>Flow ($\mu\text{L}/\text{min}$) column+matrix</i>	<i>Matrix conc. (mg/mL)</i>	<i>Spot 5s V (μL)</i>	<i>Spot 10s V (μL)</i>
1:1	1.4	1.5	0.12	0.23
1:2	2.1	2	0.17	0.35
1:3	2.8	2.25	0.23	0.47
1:4	3.5	2.4	0.29	0.58
1:5	4.2	2.5	0.35	0.70
1:6	4.8	2.6	0.41	0.82

It was decided to use a ratio 1:4 (effluent/matrix) because it provides an appropriate matrix concentration on the plate and the spot volume is 0.3-0.6 μL . Bigger spot volumes are less homogeneous and generate large spot areas which need to accumulate more laser shots and thus longer measurement times. Small spot volumes below 0.2 μL are more homogeneous but generate small spot areas which are difficult to visualise. In addition, the low amount of sample hampers the performance of several MS/MS experiments on a single spot and results in loss of sensitivity [59].

A spot deposition of five seconds was chosen for the analysis of complex peptide mixtures by LC-MALDI. A five seconds spot generated a volume of about 300 nL that, after acetonitrile evaporation, showed a spot diameter of about 1200 μm . For optimal distribution of the space available on the plate the spots were arranged in 32 rows and 49 columns, which provided a capacity of 1568 spots.

Optimization of the mass accuracy

Mass accuracy is a key parameter for protein identification by database search because a high mass accuracy delivers more confident identifications and reduces the probability of false peptide matching. The mass accuracy indicates the deviation of the instrument response from a known calculated monoisotopic mass. The mass accuracy for a peptide is usually expressed in ppm and defined by the formula (5). To express the mass accuracy of a peptide in this work, the absolute error was employed.

$$\text{Mass accuracy (ppm)} = [(\text{observed mass} - \text{theoretical mass}) / \text{theoretical mass}] * 10^6 \quad (5)$$

A good mass accuracy is achieved with a MALDI MS instrument by either internal or external calibration. The internal calibration calibrates the spectrum based on the observed time of flight of known standards within the sample. The external calibration calibrates the spectrum based on the observed time of flight of known standards that have been acquired before in a different spot and calibrated internally in that spot. For routine measurement it is common to use the default calibration. The default calibration is based on the external calibration provided by eight calibration spots on the MALDI-plate. It uses a multi-parameter calibration equation with values based on the configuration of the system to provide calibration for the entire plate.

In order to test and improve the mass accuracy for the proteome analysis by LC-MALDI, a complex mixture of peptides of known composition was analyzed. The

peptide mixture was kindly provided by the group of Prof. Huber (Instrumental Analysis and Bioanalysis, University of Saarland, Germany). It was obtained by digestion of 10 standard proteins of different molecular masses and combining them to give the concentrations shown in Table 3. The mixture was diluted 1/10 and one microliter was employed for LC-MALDI analysis. Consequently the lowest amount of protein subjected to analysis was 0.250 pmol for transferrine and the highest amount was 1.61 pmol, which corresponded to cytochrome C.

Table 3: Digested protein mix used for mass accuracy improvement.

<i>Protein</i>	<i>MW (Da)</i>	<i>Concentration (pmol/μl)</i>
Cytochrom C	12360	16.18
BSA	66534	7.51
β -lactoglobulin A	18362	10.89
Carbonic Anhydrase	29022	6.89
Catalase	57585	3.47
Lysozyme	14306	13.98
Myoglobine	16951	11.79
Ribonuclease A	13682	14.61
Transferrine	79593	2.51
α -Lactalbumin	14180	14.11

In order to investigate if the mass accuracy was affected between LC-MALDI runs the ten protein digest mixture was spotted consecutively three times on a MALDI plate employing the same gradient and analysed by MALDI TOF/TOF MS. The default calibration was updated for MS and MS/MS mode, once before the consecutive measurement of the three spotted chromatograms (LC1, LC2, LC3) by MALDI TOF/TOF. The mass accuracy observed for each of the ten proteins in each analysed chromatographic run is shown Figure 14. After identification of the protein by database search, the calculated mass accuracy is the average of all the mass accuracy values assigned to the peptides correctly matched to the protein.

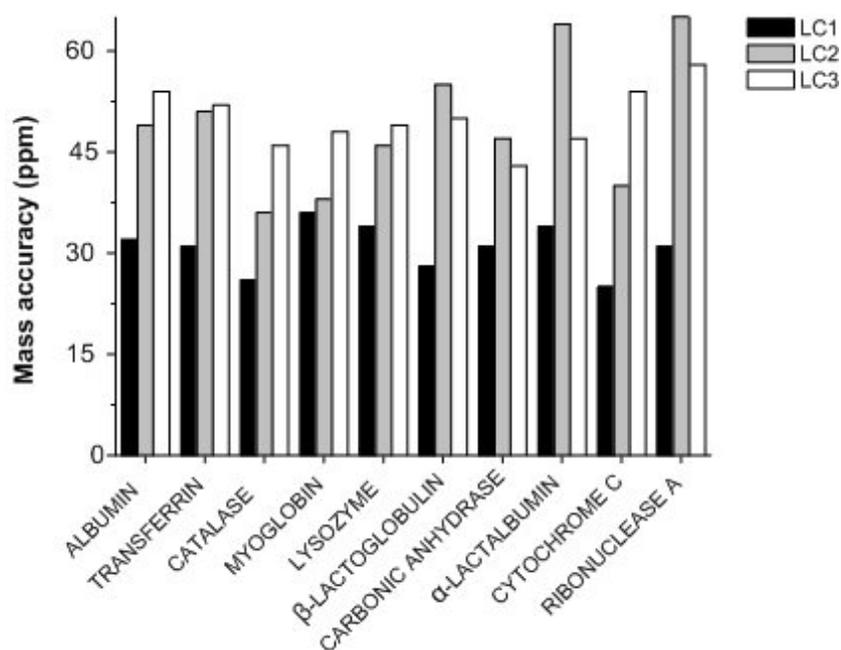


Figure 14: Mass accuracy observed for each protein on three consecutive LC-MALDI runs. Only LC1 was preceded by an update of the default calibration. Default calibration was not updated previous to LC2 and LC3 measurements.

In average, the mass accuracy observed for the identified proteins was 30 ppm during the first LC-MALDI run (LC1). However the mass accuracy was not constant between runs and got worse for the next two runs, LC2 and LC3. On the other hand the mass accuracy of the MS/MS mode was not affected between runs. To avoid a decrease in mass accuracy between LC-MALDI runs, the default calibration in MS mode was updated before each single run. Thus the mass accuracy was improved and the average mass accuracy of three consecutive analysis for the proteins was 34 ppm (see Figure 15A).

A further improvement in mass accuracy can be achieved by internal calibration, using a reference peptide present in the spot together with the sample. Internal calibration was performed on every single spot of the MALDI plate by addition of the peptide [Glu¹]-fibrinopeptide B (MW 1570.57 Da) to the matrix solution at a concentration of 21 fmol/ μ L. Because the matrix was mixed in the ratio 4:1 with the column effluent, the final concentration of the internal standard was 15.7 fmol/ μ L. The spot deposition was set to 5 seconds and consequently the spot volume was about 300 nL. The final amount of Fibrinopeptide on the plate per spot was of 4.7 fmol.

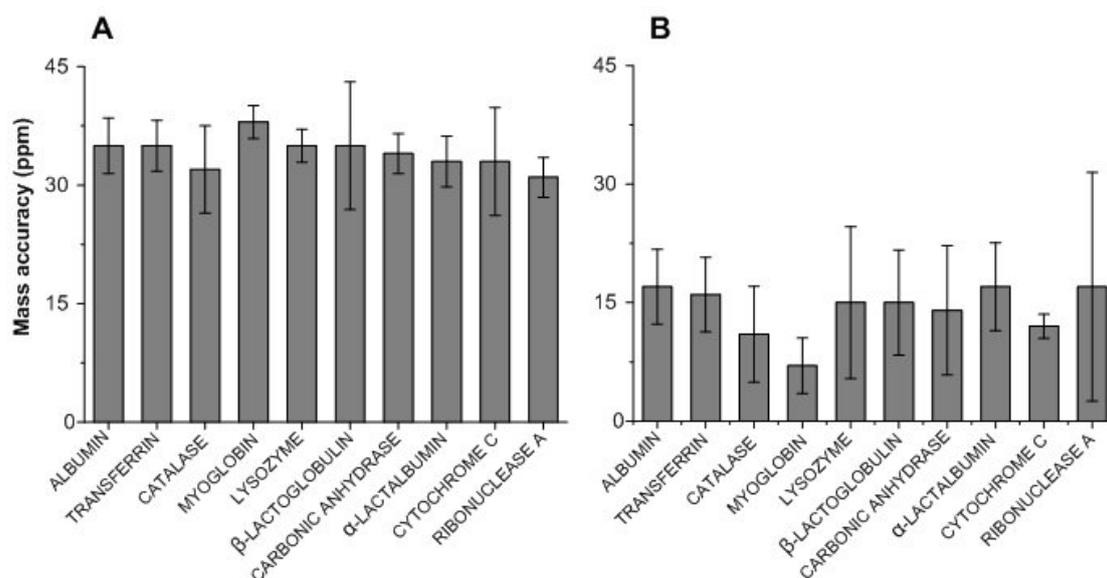


Figure 15: Mass accuracy in ppm calculated for the identified proteins by database search. In (A) only default calibration of the instrument was employed. In (B) the peptide Glu-Fibrinopeptide B was added to the matrix to perform internal calibration in addition to the default calibration.

The internal calibration improved the average mass accuracy of the proteins from 34 to 14 ppm (see Figure 15). Despite a quantitative improvement in the average mass accuracy by employing the internal calibration, the standard deviations observed were in some cases larger than the ones observed with external calibration (see Figure 15B). This is due to the fact that the amount of the Fibrinopeptide B on a spot is low and the presence of peptides at higher concentrations and different composition on the same spot can induce peak suppression of the Fibrinopeptide B [90]. For this reason in some spots the Fibrinopeptide B peak may not be present or not labelled because of a low S/N. In these cases internal calibration is not performed and only the default calibration is applied leading to higher deviations. In any case, the best mass accuracy was obtained when internal calibration was employed and a default calibration update was run before every single LC-MALDI analysis. Therefore this procedure was employed for the proteome analysis of *C. glutamicum*.

Collision induced dissociation (CID) versus post-source decay (PSD)

The MALDI TOF/TOF 4800 can be operated in the MS/MS mode with or without CID. Peptide fragmentation in the CID region is achieved by collisions between the precursor isolated prior to the CID region and gas molecules. When the CID is not employed, fragmentation happens by metastable decomposition occurring in the field-

free region after the ion source. This is the so called post source decay (PSD). In this case the precursor is isolated with its fragments according to their time of flight and re-accelerated in the second source. The spectra obtained by PSD deliver abundant y- and b- ions. Nevertheless CID spectra from peptide fragmentation of high-energy tandem mass spectrometry present some special features. Peptide fragmentation employing CID delivers immonium ions in the low-mass region (m/z 60-160). These signals are especially intensive for peptides containing histidine, isoleucine, leucine, phenylalanine, proline, tryptophan and tyrosine and are useful to confirm assignments about peptide composition. The partial loss of side-chain fragments during high-energy collisions leads to the formation of d- and w-ions which permit distinguishing the isomeric amino acids leucine and isoleucine. In addition, CID is known to deliver more internal fragments and a-ions aside from y- and b-series [52, 53].

In order to investigate the most advantageous conditions for the proteome analysis a digest of ten different proteins (Table 3) was analysed by LC-MALDI with CID and without CID. The sample was spotted three times and analysed by LC-MALDI MS/MS with the same acquisition parameters. The results were analysed by database search. Figure 16 shows that the total number of peptides counted per protein with CID and without it, is very similar except for the proteins presenting higher molecular weight. When employing peptide fragmentation without CID the three biggest proteins, BSA (66534 Da), transferrine (79593 Da) and catalase (57585), presented a higher number of peptides successfully sequenced.

Considering the sequence coverage achieved for the different proteins, the situation is different. Although the number of sequenced peptides was similar for six of the proteins (myoglobine, lysozyme, carbonic anhydrase, ribonuclease A, β -lactoglobuline and cytochrome) the sequence coverage achieved with CID was higher. Higher sequence coverage can be explained in terms of a higher number of different peptides successfully sequenced. The ion scores reported for both methods, with CID and without CID, were very similar.

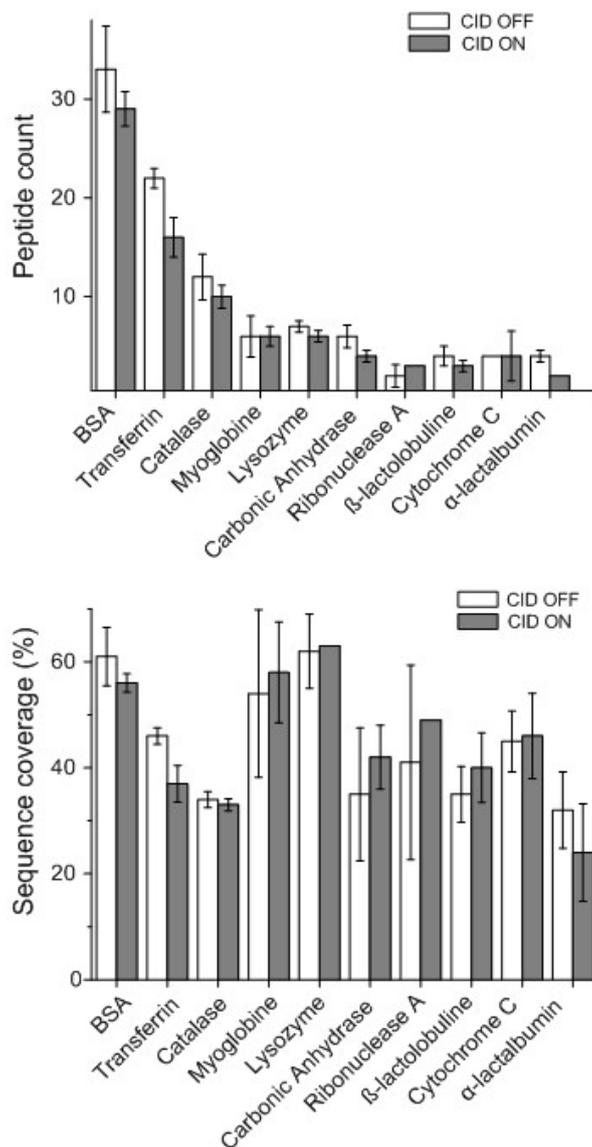


Figure 16: Comparison of peptide fragmentation in the MALDI TOF/TOF with CID and without CID. The peptide count (top figure) and sequence coverage (bottom figure) was evaluated for a digest of ten proteins.

CID was more successful in sequencing different peptides for some of the proteins. The capacity to deliver successful sequencing of different peptides of a single protein is of great importance in the proteome analysis. Proteins identified by one unique peptide, even if sequenced repetitively many times, are the so called one-hit wonders and usually not accepted as valid identifications. Besides, the calculated standard deviation was better with CID for the number of peptides count as well as for the sequence coverage. The absence of error bars in the plot indicates that the same results were achieved in the three subsequent measurements. Despite similar performance with and without CID, the results with CID on were slightly more reproducible and therefore more convenient for the proteome analysis of *C. glutamicum*.

Performance of the Timed Ion Selector (TIS)

The TIS is an ion gate situated before the CID region and deflects ions lighter than the selected precursors and therefore flying faster, as well as ions heavier than the precursor which are slower. The time during which the ion gate is open, is the so called “mass window for precursor selection”. The precursor mass window of the MALDI TOF/TOF ABI 4800, expressed in Da, can be set by the user as an absolute fixed value or relative to the resolution. The maximal TIS resolution given from the manufacturer is 400. The definition of resolution based on the full width half maximum (FWHM) is shown in Figure 17 A. High resolution values provide small windows for ion selection. As a consequence selectivity improves but too narrow precursor windows can compromise the sensitivity as the intensities of the fragments decrease [91].

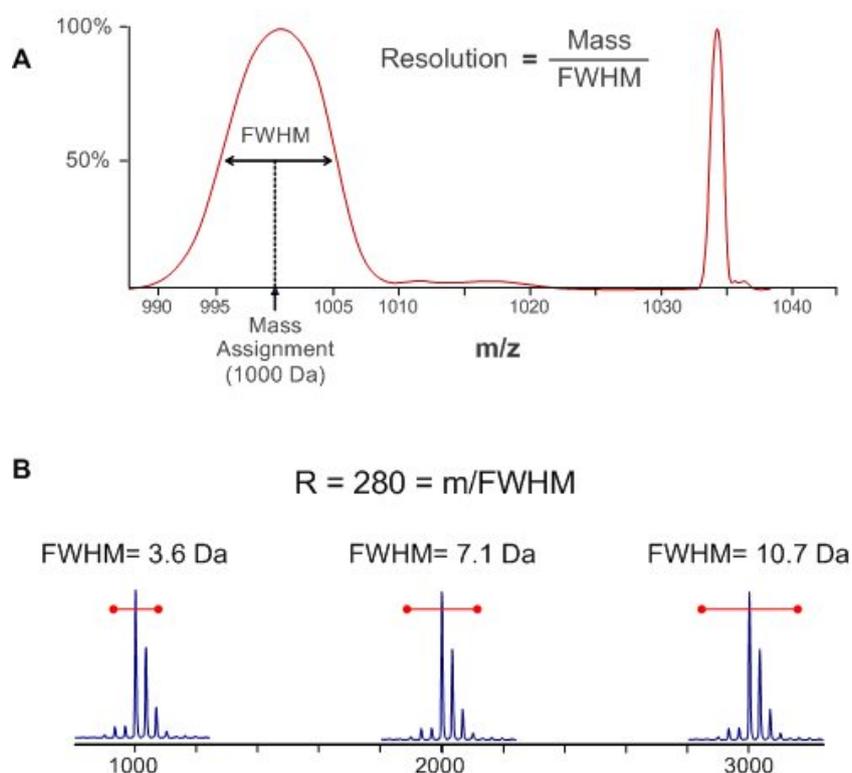


Figure 17: (A) Definition of resolution for the MALDI TOF/TOF mass spectrometer. (B) Relative precursor mass window applied by the TIS for three different masses when the resolution is set to 280.

The use of low resolution and therefore a broad mass window, compromises the selectivity of the precursor selection. In the present work the resolution for precursor selection was set to 280, relative to the mass of the precursor. The mass windows

calculated for three precursor masses (1000, 2000 and 3000 Da) with TIS resolution of 280 are shown in Figure 17 B. The resolution chosen for precursor selection offered a different mass window relative to the mass of the precursor. For instance, for a precursor mass of 1000 Da the full width at half maximum mass window is 3.6 Da, which allows selection of the monoisotopic peak and part of the isotopic cluster. For an ion of 2000 Da the mass window is 7.1 Da and for an ion of 3000, 10.7 Da. A 10.7 Da precursor mass window guarantees the selection of the monoisotopic peak together with the isotopes. However, in a complex sample like a proteome digest the presence of neighbour peaks within the mass window of the precursors can interfere in the identification. If more than one precursor is selected by the TIS, the MS/MS spectrum gets more complicated due to unassigned fragments. The presence of unassigned fragments in the spectrum complicates the assignment of a sequence by database search and yields a lower ion score for the identified peptide. This problem is more pronounced for peptides of high molecular mass above 3000 - 3500 Da.

4.3. Proteome analysis of *Corynebacterium glutamicum* by two-dimensional liquid chromatography & MALDI TOF/TOF MS

The cytosolic proteins of *C. glutamicum* wild-type extracted as described in section 3.3, were digested enzymatically with trypsin generating a mixture of peptides. The high complexity of the sample required fractionation before performing mass spectrometric analysis. Peptides were resolved by a novel bottom-up approach in which reversed phase (RP-HPLC) was employed in the first dimension and ion-pair reversed-phase (IP-RP-HPLC) in the second dimension [1, 36]. In the text we will refer to the two-dimensional separation as 2D-RP-HPLC. Because elution was accomplished in both dimensions by an increasing gradient of acetonitrile, the orthogonality of the separation was based on the use of a different pH in the mobile phase. The charge state shown by the different ionisable amino acids in a peptide sequence at pH 10 and pH 2.1 is shown in Table 4. A peptide carries different charges at different pH values depending on its amino acid composition. Also the polarity of the peptide is affected and thus the interaction with the stationary phase. As a result, the retention times of the peptides carrying ionisable amino acids are different in each dimension.

Table 4: Charge state of the ionizable amino acids at the working conditions for the first dimension (pH10) and the second dimension (pH 2.1). Values are based on the pKa values of the side chain of ionizable amino acids (see Table 1). The contribution of the amino-terminus and carboxyl-terminus is also shown.

<i>Amino Acid</i>	<i>pH 2.1</i>	<i>pH 10</i>
Asp	0	-1
Glu	0	-1
His	1	0
Cys	0	-1
Tyr	0	-0.5
Lys	1	0.5
Arg	1	1
N-terminus	1	0
C-terminus	0	-1

4.3.1. First dimension: reversed-phase chromatography at high pH

Peptides generated by digestion of the proteome of *C. glutamicum* were separated in the first dimension by reversed-phase chromatography carried out at pH 10. This dimension should show high orthogonality with the second dimension, with a working pH of 2.1.

The high pH was achieved in the mobile phase by addition of triethylamine (TEA) titrated to pH 10 with acetic acid. The amount of sample injected was 280 µg of peptides from the digest of *C. glutamicum* cytosolic proteins and the gradient employed was linear from 0 to 55% acetonitrile in 55 min. The chromatogram acquired at wavelength 280 nm is shown in Figure 18. In the chromatogram peptide signals are observed between minutes 17 and 45. However, not resolved peaks are observed due to the high complexity of the sample. The effluent from the column was collected in one minute intervals. Accordingly, a total of 55 fractions of 200 µL volume were collected for further analysis.

Since the first dimension was performed by reversed-phase chromatography instead of the common SCX, there were no salts present in the collected fractions but acetonitrile. As the second dimension was RP-HPLC and separation is based also on an acetonitrile gradient, it was necessary to remove the acetonitrile from the fractions.

The fractions were subjected to evaporation of acetonitrile in a vacuum concentrator thus increasing sample concentration 10 times.

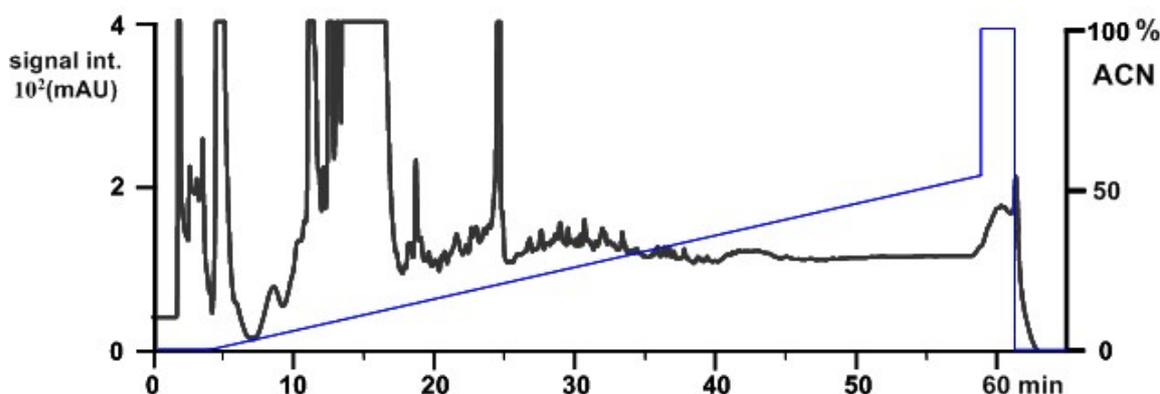


Figure 18: Separation of 280 μg of *C. glutamicum* digested proteome at high pH (pH 10). The column was a C18 Gemini from Phenomenex and mobile phase composition was: (A) 72 mmol/L TEA in H_2O , pH 10 with acetic acid; (B) 72 mmol/L TEA, 52 mmol/L acetic acid in acetonitrile. Elution at pH 10 was achieved by a 55 min linear gradient from 0 to 55% B followed by 2 min at 100% B. The chromatogram was acquired at wavelength 280 nm.

4.3.2. Fraction screening by MALDI TOF MS

The total number of fractions collected during the first dimension at pH 10 was 55 (one per minute). In the chromatogram peptide peaks are observed between fractions 17 and 45 but is not possible to discriminate clearly the peptide containing fractions. In order to elucidate the fractions to be analysed a fast screening method was developed employing the MALDI-TOF MS technology. All fractions were analyzed by preparation of two dilutions and each dilution was measured three times. Dilutions were prepared after evaporation of the acetonitrile from the elution gradient in a vacuum concentrator. The first dilution consisted of a 1:10 dilution of one microliter of the sample with a 0.1% TFA solution. The second dilution was prepared by dilution 1:100 of one microliter of the sample with a 0.1% TFA solution. Each dilution of a fraction was spotted three times on a MALDI-plate by deposition of 0.5 μL of the sample followed by 0.5 μL of matrix solution and it was allowed to co-crystallize for a few minutes. Following the MALDI-plate was measured by MALDI-TOF MS. After calibration, one spectrum was acquired from each spot in MS mode and the number of peaks labelled by the software was counted. A total of 1500 shots were accumulated per spectrum. Figure 19 depicts the average and standard deviation of the number of peaks counted in the MS spectra from the collected fractions.

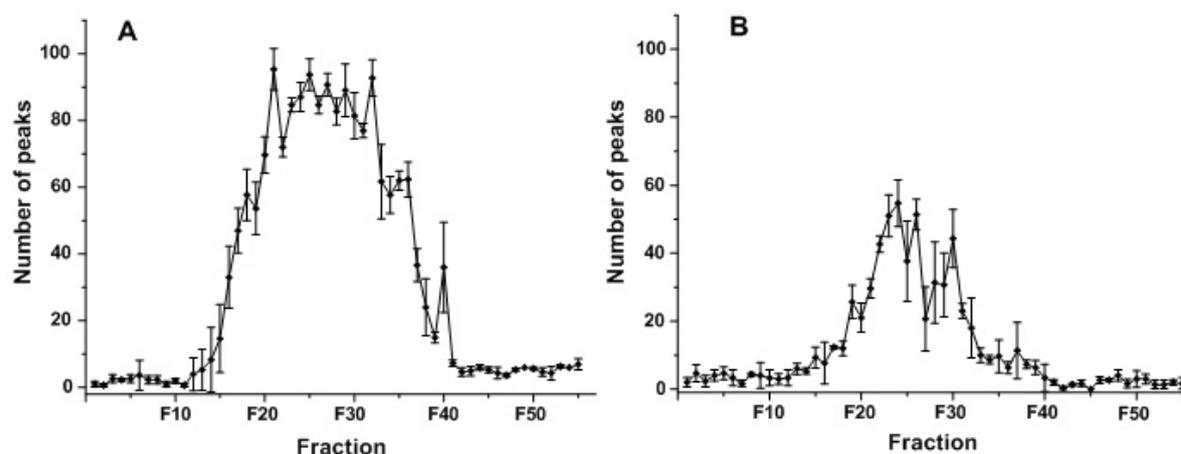


Figure 19: Number of peaks detected in each collected fraction by MALDI-TOF MS analysis. Two dilutions were prepared from each fraction, 1:10 (A) and 1:100 (B). The samples were spotted three times on the plate for measurement.

Almost no peaks are observed in the first and last ten fractions, where the few signals observed are derived from the matrix signals, CHCA clusters with sodium and potassium. The fractions where more peaks were counted correlate with the peptide peak signals in the chromatogram acquired during the first dimension (Figure 18). The dilution rate 1:10 (Figure 19 A) generated better results than dilution 1:100 and a higher number of peaks was observed per fraction as well as a smaller standard deviation. It can be speculated that the concentration of many peptides in the more diluted sample was below the limit of detection of the mass spectrometer for the acquisition parameters employed. The measurement of the fractions 1:100 could be improved by increasing the number of laser shots accumulated for each collected spectrum. However, because the purpose of the method is to screen the fractions, the preparation of two dilutions was sufficient to select the fractions for analysis. Fractions collected between minutes 15 and 45 were selected for further analysis by LC-MALDI. Because the number of peptides in a single fraction is very high and ion suppression during ionization [90] has a strong effect on the ionization efficiency of the peptides, the method employed does not provide quantitative information. However, the obtained information is sufficient to provide a criterion that permits rapid evaluation and selection of the fractions to analyze. The rapid screening method presented here minimizes the analysis time by rapid selection of the fractions to be analyzed; it facilitates the analysis planning and consumes only little sample volume.

4.3.3. Second dimension: reversed-phase chromatography at low pH

From a total of 55 collected fractions, thirty-one fractions corresponding to minutes 15 to 45 of the gradient were selected for further analysis by LC-MALDI. After acetonitrile removal the volume was reduced to 1/10. Each of the fractions was reconstituted with buffer, injected and spotted on a MALDI-plate three times generating a total of 93 chromatographic separations. The UV chromatograms of the three consecutive chromatographic separations of fraction 30 are illustrated in Figure 20. It can be observed the excellent reproducibility of the second dimension separation.

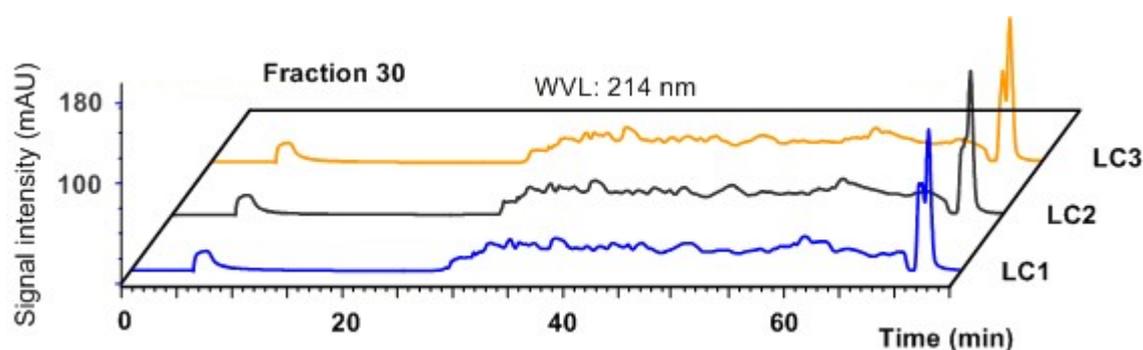


Figure 20: Three replicate injections of fraction 30 analysed by the LC-MALDI set-up described in section 4.2.1.

Spots were collected on the plate every 5 seconds for 62 minutes and one single chromatography run generated 747 spots on a MALDI-plate. Because it was possible to spot up to 1568 spots on each MALDI-plate, two chromatographic runs were spotted on each plate. The triplicate analysis of the 31 fractions generated a total of 47 MALDI-plates for analysis by MALDI TOF/TOF mass spectrometry. The total time required for the chromatographic separation and spotting was 116.25 hours.

Figure 21 shows the chromatograms of the 31 fractions later analysed by LC-MALDI TOF/TOF. Only one of the three replicate injections is shown per fraction. In the three first and last fractions there is almost no peptide signals observed, whereas fractions 20 to 37 show the presence of high amount of peptides and no resolved peaks can be observed.

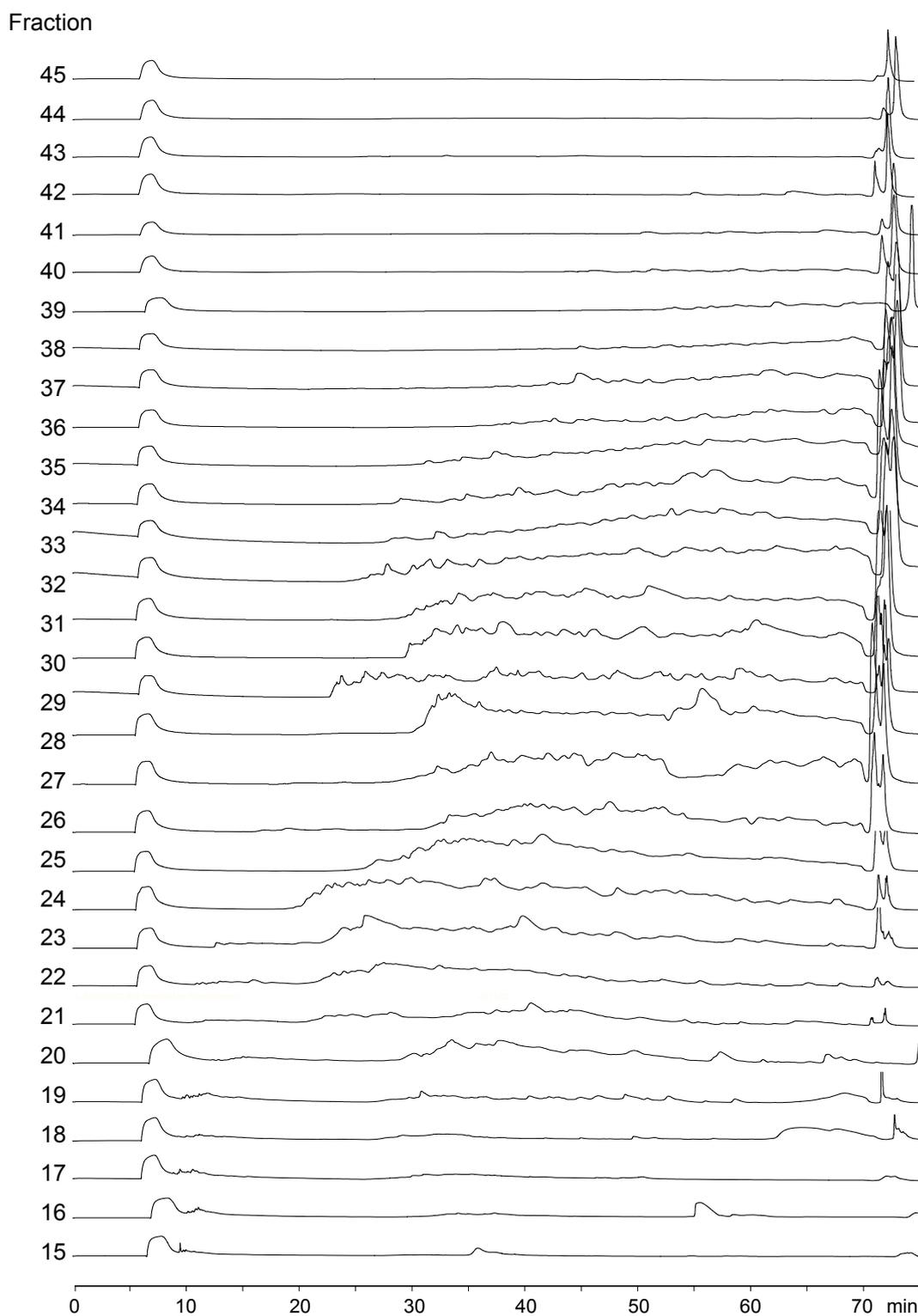


Figure 21: Analysis by LC-MALDI of 31 fractions collected during the first dimension. UV-chromatograms (214 nm) were acquired by the detector situated between the outlet of the monolithic column and the spotter. Mobile phase composition was: (A) 0.05% TFA in H₂O (B) 0.05% TFA in 100% ACN. Gradient: 60 min linear gradient from 0 to 20% B, followed by 3 min isocratic elution at 100% B. Flow rate 0.7 μ L/min. Column temperature was 25 °C.

4.3.4. MALDI TOF/TOF analysis

The 31 fractions selected for LC-MALDI analysis and spotted on MALDI-plate were subjected to high-throughput MALDI MS/MS analysis. The analysis of a MALDI-plate in which a chromatographic run is spotted, is carried out as follows: a MS acquisition on every spot of the plate for generation of a precursor list is followed by MS/MS of the precursors/peptides for protein identification. During the MS analysis one mass spectrum is acquired on each spot of the chromatographic run deposited on the plate. The mass range selected for the MS acquisition in this work was 800-4000 m/z. The instrument software evaluates all ions of the MS data acquired on each spot using the *fraction-to fraction tolerance* and the *minimum chromatogram width* in order to derive LC-MALDI peaks. The fraction-to-fraction precursor mass tolerance is expressed in ppm and prevents the redundant acquisition of a precursor in adjacent spots if the corresponding m/z values belong to the same elution profile. The first m/z in the list is assigned to the first LC-MALDI peak. The fraction-to-fraction mass tolerance (in the present work 200 ppm) is applied to the first m/z in the next spot. If the centroid mass is within the tolerance of the first LC-MALDI peak, the software merges the masses into the same LC-MALDI peak and derives a weighted average mass. If the centroid MS mass is outside the tolerance, the mass is selected as a new theoretical LC-MALDI peak. Every time a mass is merged into an existing LC-MALDI peak, the average mass is modified and the software examines all other masses to check if they are now within the tolerance. The minimum chromatogram peak width specifies the minimum number of spots to consider a precursor as LC-MALDI peak. This parameter facilitates to avoid selection of nonreproducible chemical noise. A gap of one spot is permitted in order to allow detection of low abundant peptides within a range of spots. The criteria specified for generation of the precursors list in the present work were as follows:

- A signal-to-noise (S/N) ratio equal to or above 35.
- A minimum chromatogram peak width of 2. This specifies that the peak must be at least in two spots.

- Fraction-to-fraction precursor mass tolerance of 200 ppm. m/z values in two adjacent spots within 200 ppm and consequently belonging to the same elution profile will be merged into a single LC-MALDI peak.
- Exclusion list: 850 ± 50 m/z in order to avoid selection of matrix peaks; peak 1571.57 m/z from Glu¹-Fibrinopeptide B used for internal calibration; masses above 3000 m/z .
- A maximum of 6 precursors per spot were selected for analysis, starting with the strongest precursor.

During the second stage of the analysis, MS/MS data was collected on the masses included in the precursor list according to the spot where the highest intensity of the LC-MALDI peak was registered. After ionization, the selection of the precursor was carried out by the timed ion selector (TIS). The typical resolution observed in the spectra acquired in MS mode for the selected mass range m/z 800-4000 was between 11000 and 20000. In MS/MS mode the resolution observed was up to 6500.

An example of protein identification by LC-MALDI MS/MS is illustrated in Figure 22. The measurement corresponds to the LC-MALDI analysis of the fraction 32 collected during the first dimension. In Figure 22 A is shown the extracted ion chromatogram (EIC) of m/z 2569.3 ± 0.5 Da. The EIC is obtained by plotting the sum of the signal intensities of all the MS peaks in a specific mass range as function of the spot number. In the plot the signal intensity was calculated as peak height for the ion m/z 2569.9 on each spot of the LC-MALDI analysis of fraction 32. The maximum of the LC-MALDI peak was found in the spot 419 and corresponds to an elution time of 46.9 min. The FWHM recorded for this peak was 15 seconds. The MS/MS fragmentation was performed on the precursor m/z 2569 selected on the spot number 419 and the subsequently mass spectrum acquired is shown in Figure 22 B. After database search, the MS/MS spectrum (see Figure 23) was matched to the peptide IATGFIADHPHLLQAPPADDEQGR, which belongs to the pyruvate carboxilase of *C. glutamicum* (locus ID NCgl0659). This protein was identified after analysis of all the fractions by 102 matches that corresponded to 38 different peptides across different fractions.

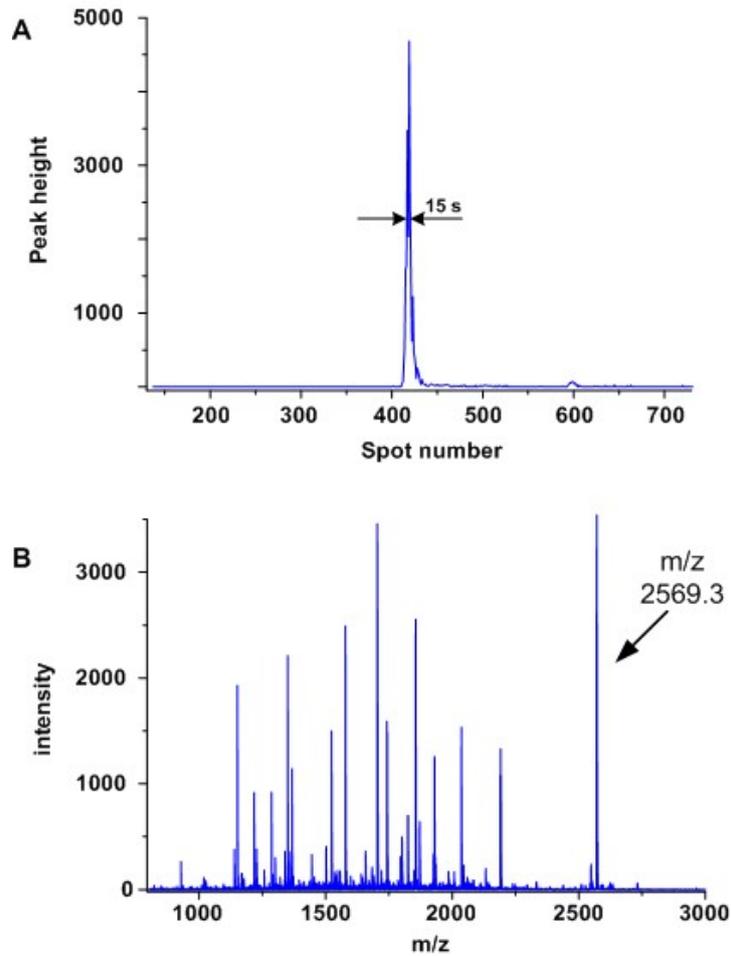


Figure 22: Peptide identification by LC MALDI MS/MS. The data shown here were obtained by analysis of fraction collected during minute 32 in the first dimension. (A) Extracted ion chromatogram of m/z 2569 from analysis by LC-MALDI of fraction 32. The maximum of the LC-MALDI peak was situated in the spot 419. The LC-MALDI peak width recorded at half maximum was 15 seconds. (B) MS spectrum obtained in the spot 419.

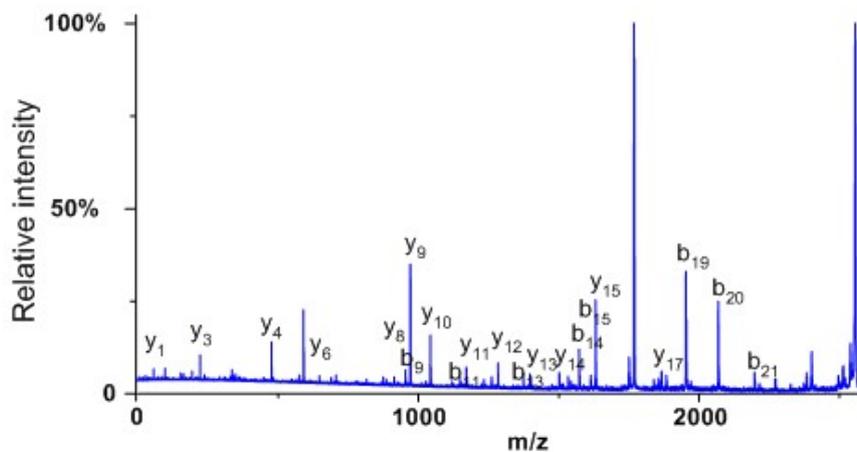


Figure 23: MS/MS spectrum of the precursor ion m/z 2569 performed on spot number 419 of the LC-MALDI analysis of fraction 18. The sequence assigned by database search with Mascot on *C. glutamicum* database was IATGFIADHPHLLQAPPADDEQGR.

The number of precursors selected for MS/MS analysis in each of the analysed fractions strongly depends on the complexity of the sample. The analysis of the 31 fractions from the first dimension generated 93 LC-MALDI runs. Figure 24 shows the number of masses included in the precursor list for subsequent MS/MS, resulting from the analysis of fractions from the first dimension. The average standard deviation of the three replicate measurements was 84 but in fraction 27 the standard deviation was higher than in the others. This was probably due to an experimental error in one of the analysis leading to a lower number of masses in the precursor list. However, the total number of peptides identified in this fraction was not affected as shown in Figure 25, where the number of different peptides successfully identified per fraction is shown.

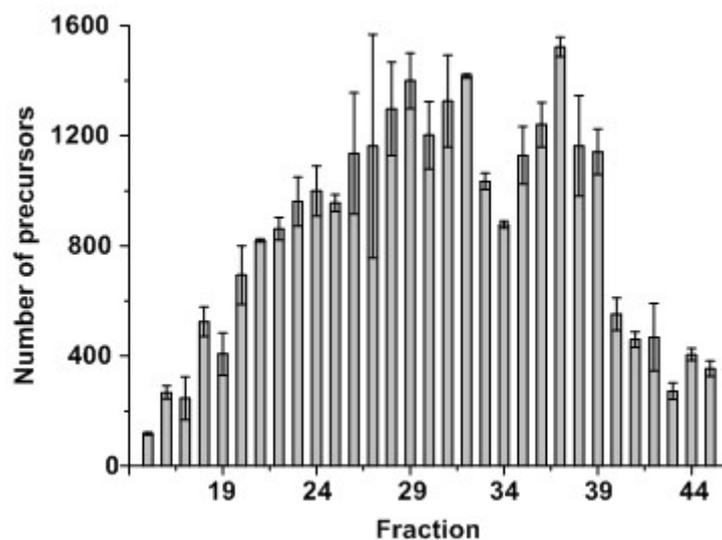


Figure 24: Number of masses selected for MS/MS in the list of precursors of each LC-MALDI analysis. The measurements correspond to the 31 fractions analyzed in triplicates.

In order to obtain the number of different identified peptides per fraction, the three replicate measurements of each fraction were merged to perform a single database search and the results are plotted in Figure 25. The highest number of different peptides sequenced per fraction was registered for fractions 30, 31 and 32. In these fractions a total of 612, 664 and 597 were, respectively, successfully sequenced. The lowest peptide content was found in the first and last analysed fractions. In these fractions 36 and 48 different peptides were identified.

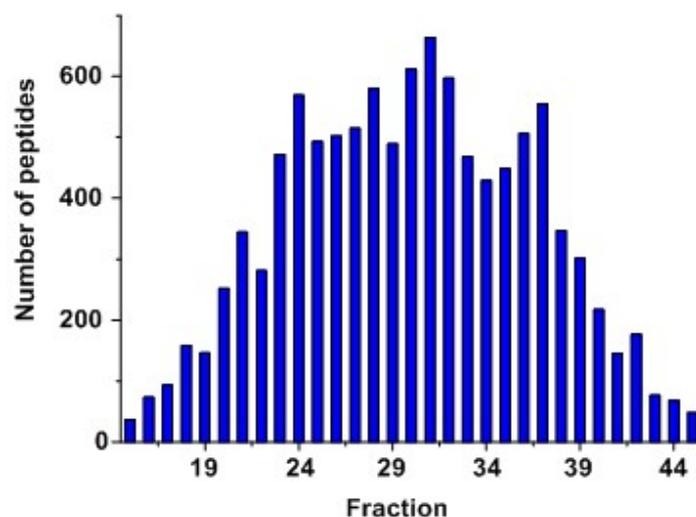


Figure 25: Number of different peptides identified by database search in each fraction. Results were obtained after pooling the three replicates of a fraction analyzed by LC-MALDI TOF/TOF MS.

The number of precursors registered per fraction (Figure 24) correlate well with the number of non-redundant peptides identified per fraction (Figure 25), as well as with the results obtained by MALDI-TOF fraction screening in section 4.3.2. The number of non-redundant peptides identified by database search per fraction was, on an average, 62% of the m/z values in the corresponding precursor list of the fraction.

4.3.5. Database search and identified proteins

All files generated by MALDI TOF/TOF analysis of the fractions were merged to perform one single search in the *C. glutamicum* database. This resulted in 75262 search queries. The Mascot search engine assigned an ion score to each experimental peptide correctly matched to a theoretical peptide sequence. The statistical significance threshold for proteins was $\geq 95\%$ and only peptides with score above the identity threshold were accepted. The number of identified proteins returned by Mascot was 1644 (see Appendix II) and the false discovery rate determined for this data set was 2.4%. The number of false positives (FP) was estimated by repeating the search in a composite database created by attaching a randomised *C. glutamicum* database to the original database thus generating a database twice the size of the original. The FPR

was calculated by doubling the number of hits in the randomised database and dividing the result by the total number of identified peptides from both databases [66]. The bar chart in Figure 26 shows the number of non-redundant proteins identified in each subsequent fraction with respect to the accumulated protein hits. The number of identified proteins increased till fraction eighteen and after that the number of newly identified proteins decreases in every fraction. The contribution to the overall number of identification in the last six fractions was very small.

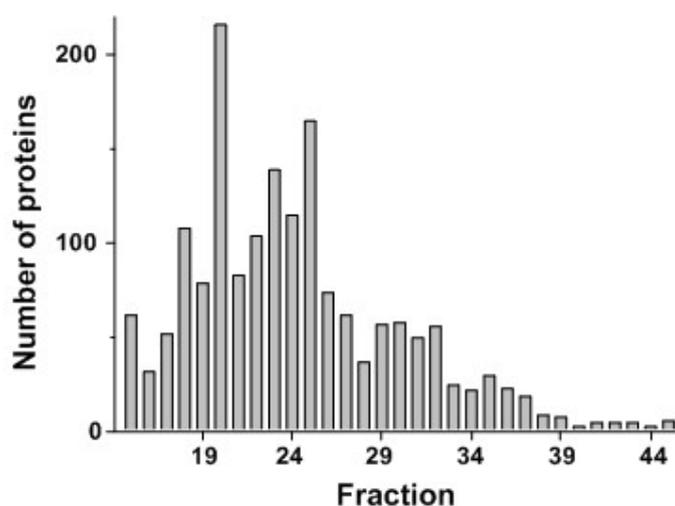


Figure 26: New protein hits identified in each fraction by database search with respect to the previously identified and starting with fraction 15. Proteins identified in a fraction by database search were not previously identified.

The identified proteins of *C. glutamicum* were classified based on the genome information of the Comprehensive Microbial Resources (CMR) of the TIGR database and is shown in Table 5. The percentage of proteins from each category identified in the present study is also shown. Almost all proteins involved in biosynthesis functions in the cell were identified in the present work: biosynthesis of amino acids (88%), biosynthesis of proteins (83%), nucleosides and nucleotides (75%), biosynthesis of cofactors, prosthetic groups and carriers (71%). Besides it is remarkable that over 50% of the predicted proteins involved in cellular processes (61%), energy metabolism (58%), DNA metabolism (55%), transcription (57%) and central intermediary metabolism (51%) were identified. In contrast, transport and binding proteins as well as proteins from the cell envelope are the underrepresented (34% and 32% identified respectively) as expected in a fraction of the proteome containing mainly cytosolic proteins. Especially interesting is the fact that almost 50% of the proteins involved in

regulatory functions in *C. glutamicum* were identified. The proteins involved in regulation of gene expression are usually present in the cell only at low concentration and thus difficult to identify in a complex mixture of proteins.

Table 5: Classification of identified proteins according to their biological function. Gene role categories were obtained from the Comprehensive Microbial Resources, Institute for Genomic Research (www.tigr.org). Note that some genes are assigned to more than one category.

<i>Gene role category</i>	<i># of genes</i>	<i>Identified by LC-MALDI TOF/TOF</i>	<i>%</i>
Amino acid biosynthesis	100	88	88
Biosynthesis of cofactors, prosthetic groups and carriers	100	71	71
Cell envelope	297	94	32
Cellular processes	95	58	61
Central intermediary metabolism	99	51	51
DNA metabolism	99	55	55
Energy metabolism	234	137	58
Fatty acid and phospholipids metabolism	33	10	30
Hypothetical proteins	139	19	14
Hypothetical proteins-Conserved	541	246	45
Mobile and extrachromosomal element functions	43	4	9
Protein fate	104	50	48
Protein synthesis	121	100	83
Purines, pyrimidines, nucleosides and nucleotides	71	53	75
Regulatory functions	174	86	49
Transcription	35	20	57
Transport and binding proteins	292	100	34
Unclassified	580	372	64
Unknown function	60	30	50

Most of the regulatory proteins in *C. glutamicum* are DNA-binding transcriptional regulators [92] containing a helix-turn-helix structural motif responsible of the protein-DNA interaction [93]. Recently, a list of 56 transcriptional regulators that have known function was published based on experimental characterization and prediction by bioinformatics tools [92]. Thirty-three of the fifty-six DNA-binding transcriptional regulators of known function were identified in the present study and are listed in the Appendix III.

About 22% of the encoded proteins in *C. glutamicum* are classified as membrane proteins [70]. Because the extraction procedure was focused on cytosolic proteins a low coverage of membrane proteins was expected. The results of the present work were compared with the membrane proteins reported by Fischer et al.[78]. As a result, 121 integral membrane proteins which include soluble components or subunits, were identified in this work and are listed in the Appendix IV. Within the identified proteins 48% of them were hypothetical proteins, 11 were predicted proteins and 12 proteins were permease components of ABC-type transporter. In the appendix, proteins are classified according to the number of transmembrane helices (TMH). In addition, the number of different peptides identified together with the number of hits that were matched to a protein is shown. 63 of these proteins were identified by one non-redundant peptide.

In the genome of *C. glutamicum*, 13 two-component signal transduction systems have been annotated [94]. The signal transduction systems are integrated by two domains: a kinase and the corresponding response regulator. As sensor kinases are integral membrane proteins with at least one trans-membrane helix, they were not extracted with the cytosolic proteins analyzed in this work and therefore not identified. However, it was possible to identify seven of the response regulators which form part of a two-component system (see Appendix V).

The identification of enzymes, which catalyse biochemical reactions of the central metabolism is essential in a proteomic study and relevant for further investigations on the organism. In the present study, all enzymes involved in the glycolysis and gluconeogenesis were identified. Further, all enzymes of the citrate cycle were identified except a hypothetical membrane protein (locus identification NCgl0359) and the succinyl-CoA synthetase β subunit (NCgl2477). Because only the cytosolic

fraction of the proteome was analyzed, the identification of membrane proteins was not probable. All enzymes of the pentose phosphate cycle were identified. All enzymes involved in the methionine biosynthesis and lysine biosynthesis pathway except the lysine exporter (NCgl1214) were identified (see Appendix X and XI).

During the first dimension peptide separation was performed at pH 10. It is known that peptides are likely to suffer deamidation under alkaline conditions, especially the sequence -Asn-Gly- [95]. To study how far peptides were affected by this modification in the present work, a new database search was performed and deamidation was chosen as variable modification in addition to oxidation of methionine. Deamidation of an Asn to Aps/IsoAsp does not modify peptide net charge but the mass is increased in 0.984 Da. The deamidation rate observed was 13% of the total identified peptides, whereas oxidation rate was 4%. One third of the deamidated peptides contained the sequence -Asn-Gly-. However, it is not possible to discern if the origin of these modifications was the basic pH employed in the first dimension or naturally occurring post-translational modifications.

4.4. Proteome analysis of *Corynebacterium glutamicum* by two-dimensional gel electrophoresis and MALDI TOF/TOF MS

Two-dimensional gel electrophoresis has been the most popular approach for proteome analysis during the past thirty years. The main advantage of this technique is the visualization of a high number of proteins after separation into two orthogonal dimensions. Visualization permits direct comparison of the spots from different samples and rapid detection of differences by overlapping of the gels with the appropriate software after gel scanning. Furthermore, the spot of interest can be excised from the gel for in-gel digestion and thus all resulting peptides that belong to the protein are analysed together in one single experiment. In contrast, digestion in a bottom-up approach leads to a mixture of peptides from all proteins present in the sample. Peptides are analysed after separation and peptides distributed all over the LC fractions but belonging to the same protein will be merged by the database search machine.

The 2D-PAGE top-down approach was applied to the analysis of the cytosolic proteins of *C. glutamicum* and employed as reference method to evaluate the results obtained

by the new shotgun approach exposed in section 4.3. Protein extraction was carried out as described before and 200 μg of proteins were dissolved in rehydration buffer. This solution was used to rehydrate the strip with the immobilized gradient used for isoelectric focusing. The pH range employed was 4 to 7 because most of the proteins of *C. glutamicum* show a pI in that range. According to the virtual 2D gel of the proteins of *C. glutamicum* (see Figure 27) there are also a number of basic proteins but preliminary studies testing different pH ranges (data not shown) did not allow the visualization of these proteins. This fact has been previously reported by Schaffer et al. [74].

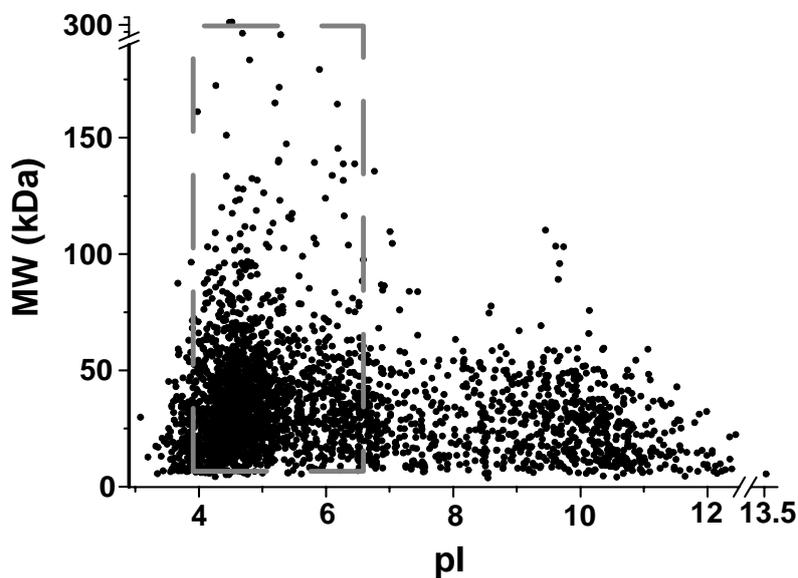


Figure 27: Theoretical 2D-gel of *C. glutamicum* based on the calculated isoelectric point and molecular weight of the 2993 proteins coded by the genome. The dashed line shows the area covered by the proteins identified by 2D-PAGE and MALDI TOF/TOF MS in this work. The borders of this area are represented by: $3.9 < \text{pI} < 6.6$ and $8.7 < M_r \text{ (Da)} < 315$.

Three technical replicate gels were performed and stained with Coomassie brilliant blue according to the protocol of Candiano et al. [25]. After gel digitalization and evaluation of the gels with the Imagemaster 2D Platinum software over 600 protein spots were reported per gel. One digitalized 2D-gel is shown in Figure 28. The spots selected for analysis were cut out of the gel and the resulting gel plugs were either subjected to digestion or stored in Eppendorf tubes at $-20\text{ }^{\circ}\text{C}$. In-gel digestion was carried out after de-staining of the spots by washing steps with water and water/acetonitrile (50:50) solutions. Disulfide bridges were reduced and cysteines alkylated with iodoacetamide to avoid side reactions with rests of unpolymerized

acrylamide. The gel plugs were incubated in acetonitrile for 1 min and allowed to dry for 15 min after removal of the acetonitrile. Then, trypsin solution was added and absorbed by the gel plug to ensure digestion of the protein. After 15 min of digestion, if necessary, the gel plugs were covered with buffer solution to facilitate peptide migration to the solution. Digestion was carried out overnight at 37 °C and was either stopped with 0.1% TFA for subsequent analysis or frozen at - 20 °C. Analysis of the spots digests was carried out by MALDI TOF/TOF MS. 0.5 µL of the digest was deposited on the plate followed by addition of 0.5 µL of MALDI matrix (5 mg/mL CHCA in a solution 50% ACN and 0.1% TFA).

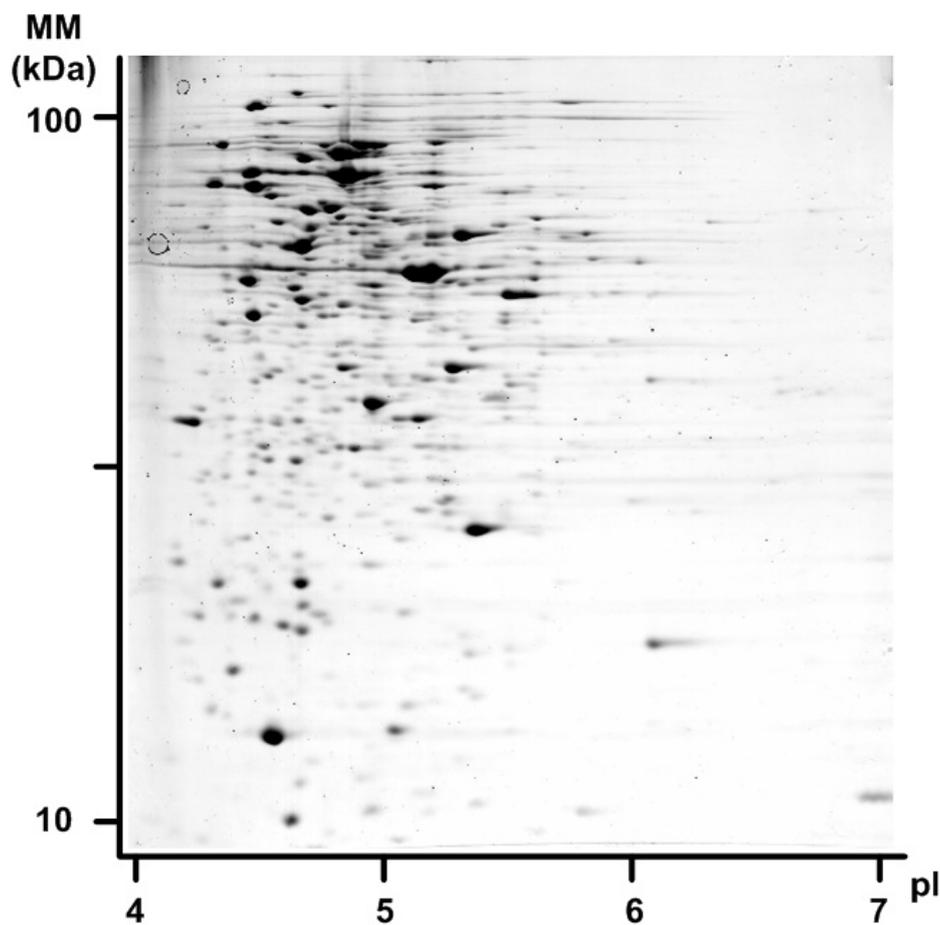


Figure 28: 2D-PAGE of *C. glutamicum* cytosolic proteins. The horizontal axis represents the pH range between 4 and 7. On the vertical axis proteins are separated according to their molecular mass; observe the scale is not linear. The gel was stained with Coomassie Brilliant Blue and digitalized with the Imagemaster software.

The analysis of a spot started with the acquisition of a PMF. Subsequently 3 of the most intensive peaks were selected for further MS/MS analysis. In the cases with very

low peptide concentration, the sample was concentrated by using ZipTip™ pipette tips with C18 chromatography beds. The MS and MS/MS information acquired on every spot was sent for database search performed by the Mascot search machine. In total 174 spots were successfully analyzed after enzymatic in-gel digestion with trypsin. The total number of different proteins identified by MALDI MS/MS was 166 (see Appendix I). In some cases, like in spot number 748 or 802, a mix of proteins was identified on a single spot. This problem is common due to overlapping of the proteins with similar molecular mass and pI [96]. The number of identified proteins of *C. glutamicum* in this work is similar to previous proteome studies by 2D-PAGE. Schaffer et al. [74] identified 152 different proteins of *C. glutamicum* in the cytosolic fraction, out of 970 protein spots detected by colloidal Coomassie staining. Similarly, 164 proteins of the 635 protein spots from the 2D-PAGE analysis of *Corynebacterium efficiens* were identified [75].

The pI of the identified proteins ranged from 3.9 to 6.63. The periplasmic component of an ABC-type transporter was identified in the spot number 105 with the lowest pI, whereas the ribosomal protein L5 with the pI 6.63 was identified in the spot number 73. Only one protein fell out of this range, the ribosomal protein L1 which has pI 8.84 (spot 74). The lowest molecular mass of the identified proteins was 8.7 kDa for the FGAM synthase (spot 834) and the highest was 315 kDa that corresponds to the fatty-acid synthase. According to these border proteins the window of proteins of *C. glutamicum* observable by 2D-PAGE method employed here is marked by a frame in Figure 27 with a dashed line. Whereas the 2D-PAGE approach is efficient to cover almost the whole range of masses (vertical axis) of the proteins in *C. glutamicum*, it is not capable of resolving the very acidic proteins and the proteins with pI above 6.63.

5. Discussion

5.1. Evaluation of 2D-RP-HPLC combined with MALDI TOF/TOF for proteome analysis

The novel combination of pH based two-dimensional reversed phase liquid chromatography with MALDI TOF/TOF mass spectrometry for the proteome analysis of *C. glutamicum* yielded the identification of 1644 proteins. This number of proteins is, to our knowledge, the highest number from *C. glutamicum* reported in a single study till now. Fischer et al. [78] reported the identification of 326 integral membrane proteins and 1181 cytosolic proteins of *C. glutamicum*. In the present work, the total number of peptides successfully sequenced by Mascot was 20553, which corresponded to 1644 proteins. 436 of these proteins were single-peptide based identifications.

5.1.1. Evaluation of the first dimension

In the first dimension, peptide separation was carried out by reversed phase liquid chromatography at pH 10 instead of the common SCX. Peptide net charge was predominantly negative because most ionizable amino acids carry negative charge (see Table 4) at high basic pH. Peptides carrying negative charges are rich in ionized carboxylic groups due to the presence of aspartic acid (D), glutamic acid (E), in addition to cysteines (C) and the C-terminus. In order to analyze peptide separation in the first dimension in terms of net charge the specific net charge of all identified peptides was calculated according to Table 4. Peptides were grouped according to their net charges and plotted against the elution time in the first dimension in Figure 29. The maximum of each elution profile is marked with an arrow in Figure 29. Peptides carrying the most negative net charges are more acidic (see Appendix VII and VIII) and eluted in the early fractions. These peptides are, consequently the most hydrophilic. As an example, the peptide DTSDDDDEISEEQQALIDR contains six D and three E, one R (charge +1) and the C-terminal (charge -1). This peptide has pI 4.9, carries net charge -9 at pH 10 and was identified in fraction 18 of the first dimension.

It is important to remark that the total number of identified peptides showing net charge -12, -11 and -10 was 19 including redundant peptides.

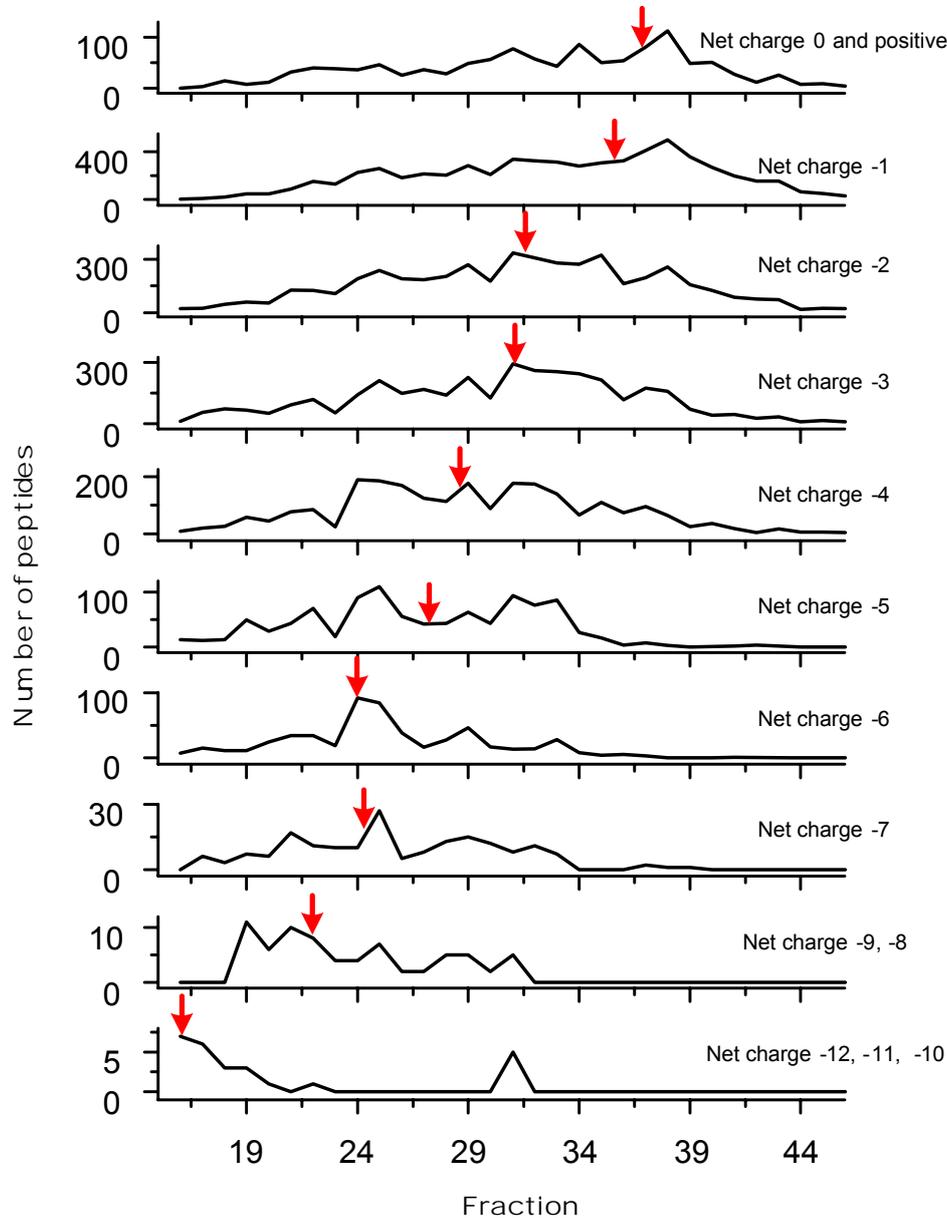


Figure 29: Separation of peptides during the first dimension at pH 10 in dependence with their net charge. Peptides carrying similar net charge were plotted against their elution time. The elution maximum for each category is pointed by the arrow. Acidic peptides carrying high negative net charge elute in the early fractions and the maximum displaces to later fractions according to increasing peptide basicity.

At high pH, basic peptides are more retained [37]. Peptides carrying net charge -1, zero and +1 have in general basic pI (see Appendix VII) show more hydrophobic character (see Appendix VIII) and tend to elute in the later fractions. The peptides presenting net charge +1 at pH 2.1 are generated by the presence of missed cleavage

sites in which at least two basic residues are present. Because only one missed cleavage was considered for database search, peptides with net charge higher than +1 were not found.

Figure 30 compares the net charge distribution for the identified peptides at basic and acidic pH. The number of different net charge states observed at pH 10 (B) is greater than at pH 2 (A). At acidic pH most peptides present net charge +2 (57%) or +3 (33%). At pH 10 the net charge range is broader and more equally distributed: net charge 0 and positive (12%), -1 (22%), -2 (24%), -3 (18%), -4 (12%), < -4 (10%).

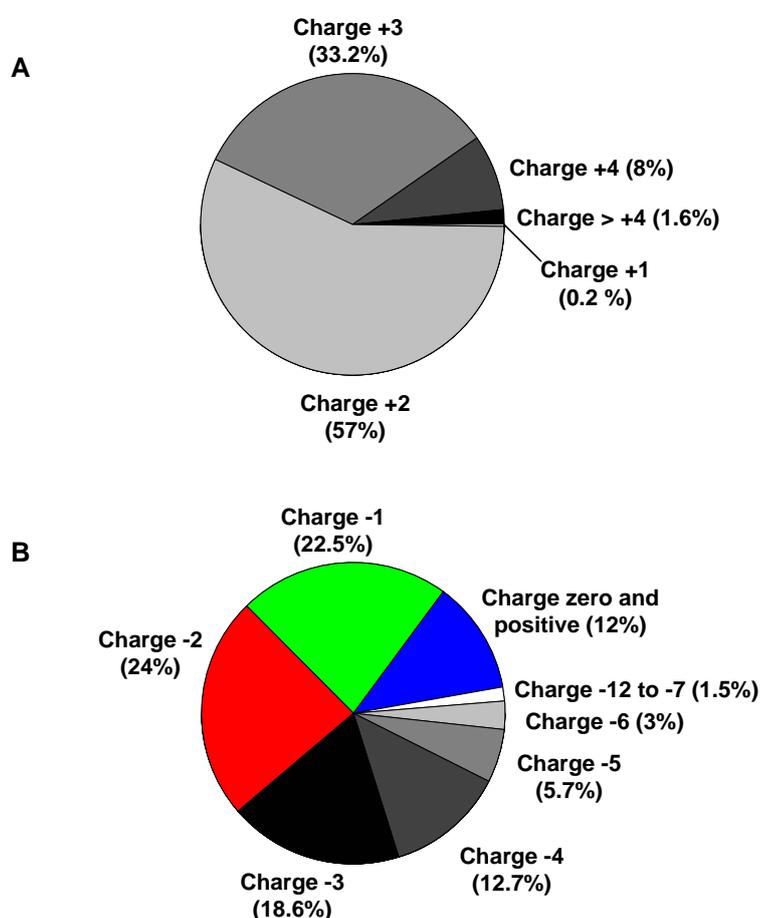


Figure 30: Distribution of peptide net charges at pH 2 (A) and pH 10 (B) for the peptides identified in this work.

The use of reversed-phase chromatography at pH 10 in the first dimension represents an advantage over the common SCX. A SCX separation is commonly carried out at acidic pH and peptides elute in narrow clusters according to their charge, mostly +2 and +3 [1, 37]. In consequence, the selectivity of the first dimension is restricted to

these two charge states. In contrast, the elution of peptides in RPC is based on peptide hydrophobicity. The employment of pH 10 increases the net charge states distribution of the peptides. As shown in the Appendix VIII peptide hydrophobicity calculated according to Eisenberg et al. [14] increases with peptide basicity. As a consequence peptide separation at pH 10 is dominated by a reversed-phase mechanism in which peptide net charge enhances the selectivity of the separation.

5.1.2. Evaluation of the second dimension

In order to achieve a high number of protein identifications, peptide selectivity must be dissimilar in both dimensions. Because the gradient in the two dimensions employed in this work was based on increasing the amount of acetonitrile, peptide elution was exclusively influenced by the mobile phase pH and the presence of ion-pairing reagents. The employment of different stationary phases to alter peptide selectivity is not as decisive as the mobile phase pH [37]. As shown in Table 4 the charge of the ionisable amino acids varies at pH 10 and 2.1.

In the following lines the influence of the peptide net charge on peptide identification in both dimensions will be studied. In order to see how peptide net charge was affected by the working pH values, the peptide net charge of the identified peptides was calculated at pH 10 and 2.1. In addition, the net charge of the peptides from the *in-silico* digest of *C. glutamicum* was calculated at the working pH and compared with the peptides identified in this study (see Figure 31). For the theoretical digest only peptides in the working mass range (800-4000 m/z) were considered for the calculations. As a result the number of peptides of the theoretical digest susceptible for detection was restricted to 95477 different peptides, including the possibility of one missed cleavage. In the theoretical digest most of the peptides are negatively charged at pH 10 (see Figure 31 A) and only 20% of the peptides carry positive or zero net charge. Although the distribution of net charges was similar for the identified and theoretical peptides at pH 10, the identification of peptides with zero or positive net charge was underrepresented compared to peptides with net charges -2, -3 and -4. Peptides carrying net charge -2, -3 and -4 at pH 10 are peptides containing aspartic acid and glutamic acid in their sequence, in addition to the negative charge provided

by the C-terminus. In section 5.2.2 it will be demonstrated that the relative occurrence of aspartic acid and glutamic acid is higher than predicted (see Appendix IX).

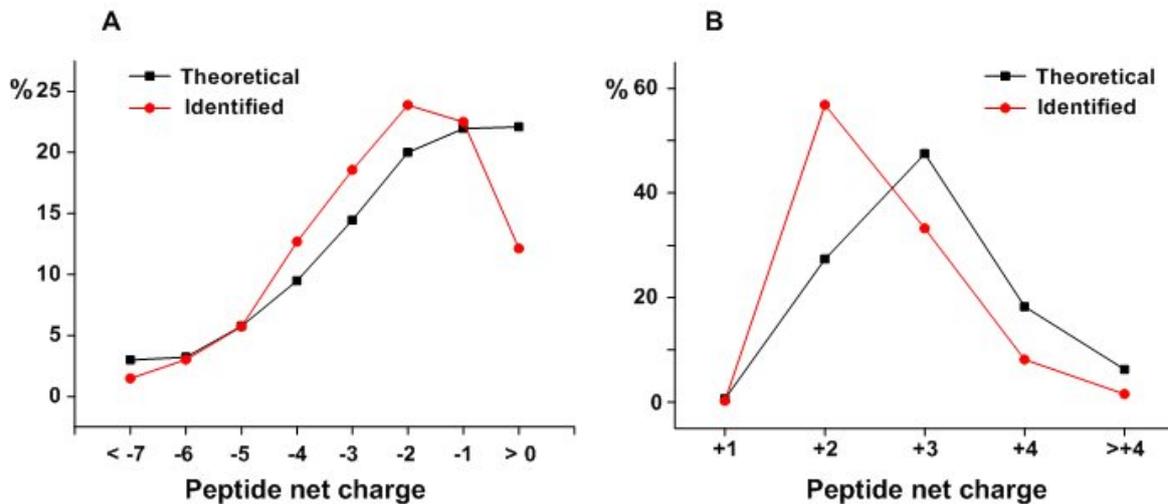


Figure 31: Percentage of peptides carrying the same net charge. Net charges were calculated for the identified peptides (●) and the peptides from the theoretical digest (■). Two situations were considered: the first dimension at pH 10 (A) and the second dimension at pH 2.1 (B).

Positive charges at pH 10 are due to the accumulation of arginine and lysine in the peptide; the presence of a missed cleavage yields a peptide with two basic residues and net charge +1. Because only one missed cleavage was considered during the database search, the finding of higher net charge states was excluded. All the underrepresented peptides showing zero or positive net charge at pH 10 are basic peptides ($pI > 7$). The distribution of charges for identified and theoretical peptides at pH 2.1 is shown in Figure 30 B. All identified peptides with net charge +1 were peptides derived from protein C-terminus without the presence of histidine (see Appendix VI), e.g. the C-terminus peptide WVTAQAMFG (m/z 1010.4) from 6-phosphofructokinase. Due to the absence of K or R, the positive charge was exclusively supplied by the N-terminus. Net charge +2 was supplied by the N-terminus and one K or R at the C-terminus, resulting from enzymatic cleavage with trypsin. Net charges +3, +4 and $> +4$ were due to the presence of one missed cleavage (one additional R or K) and the presence of histidine. At pH 2.1 there was a negative discrimination in the detection of peptides carrying net charge higher than +2. This is due to the fact that in the theoretical digest all peptides with one missed cleavage were considered for the calculations, whereas in

the real situation not all missed cleavages always occur. Only 22% of the identified peptides comprised the presence of one missed cleavage site. Peptides of net charge +2 at pH 2.1 were identified predominantly (57%). These peptides had mostly acidic or neutral isoelectric point: 0.02% had pI 4; 9% had pI 5; 53% had pI 6; 26% had pI 7 and only 12% had pI 8.

5.1.3. Orthogonality of the separation

The successful separation of peptides in a complex mixture by 2D-LC is achievable only if the selectivity in both dimensions is dissimilar, ideally orthogonal. In order to evaluate the orthogonality of the two dimensions the peptides were plotted on a two-dimensional spatial distribution according to the elution time in each dimension. The orthogonality of the 2D-LC employed in this work is shown in Figure 32.

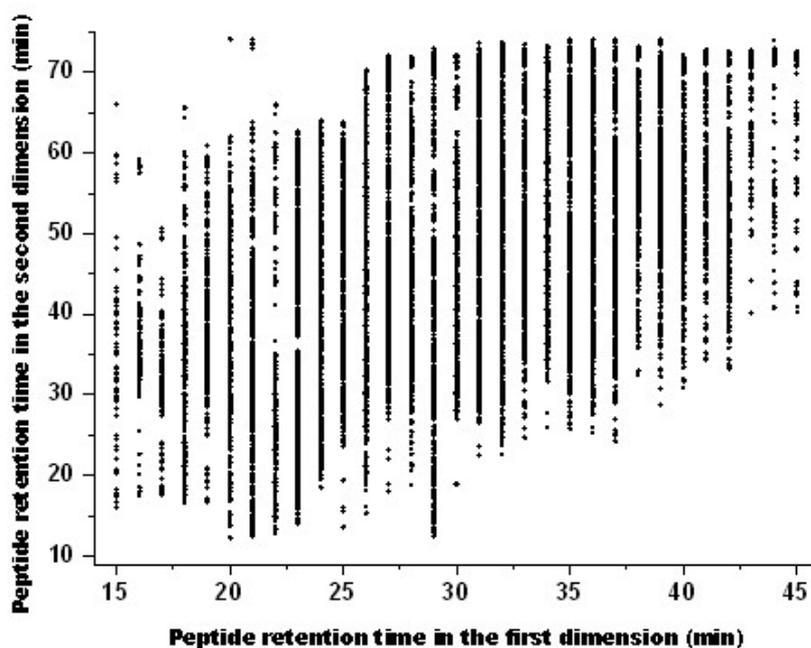


Figure 32: Orthogonality of the separation. Peptides are plotted according to their retention time in the first dimension (collected fraction) and in the second dimension (elution time assigned to the spot where a peptide was identified).

The elution time in the first dimension was given by the fraction in which a peptide was identified whereas the elution time in the second dimension was given by the spot number of the MALDI-plate where the peptide was successfully sequenced. A MATLAB-based program was employed to correlate the identified peptide with the spot where it was sequenced and assigned the corresponding retention time. Although peptides are distributed almost over the whole separation area there is an empty region in the bottom right corner and therefore the separation can be described as “nearly-orthogonal”. Because an acetonitrile gradient was employed in both dimensions for peptide separation, very hydrophobic peptides eluting in the latest fractions of the first dimension (35 - 45% of acetonitrile content) elute in the second dimension also at higher acetonitrile content. These peptides are displaced to the top right corner of the separation area leaving an unpopulated area in the bottom right corner. Also the area in the left top corner is partially unoccupied because in the first fractions of the first dimension elute very hydrophilic peptides.

In order to visualize peptide elution quantitatively peptide density was plotted in a 2D partitioned space (see Figure 33). The separation area was divided in 90 uniform-sized bins (30x30) and the number of identified peptides per bin was computed according to their retention times with MS Excel. For the calculations all identified peptides by MALDI TOF/TOF were considered including redundant peptides. Figure 33 illustrates that most of the peptides are concentrated along the diagonal of the separation space. Most of the peptides elute in the first dimension for an acetonitrile percentage of 23 to 38%. In the second dimension peptides are mostly well distributed except for the very hydrophobic ones present in the last fractions of the first dimension. These peptides start eluting on the minute 29 of the first dimension and in the second dimension elute during the 3 minutes washing step with 100% eluent B.

The orthogonality of the system could be improved by adapting the acetonitrile gradient to the peptide content. As shown above, the latest fractions from the first dimension are rich in hydrophobic peptides. In the second dimension, many peptides from fractions collected between 29 and 43 min in the first dimension elute during the washing step of the gradient. The original gradient for the second dimension was 0 to 20% of acetonitrile in 60 min. An improved gradient for fractions from the first dimension between minutes 29 and 45, could be faster from 0 to 10% of acetonitrile and extended from 10 to 25% acetonitrile.

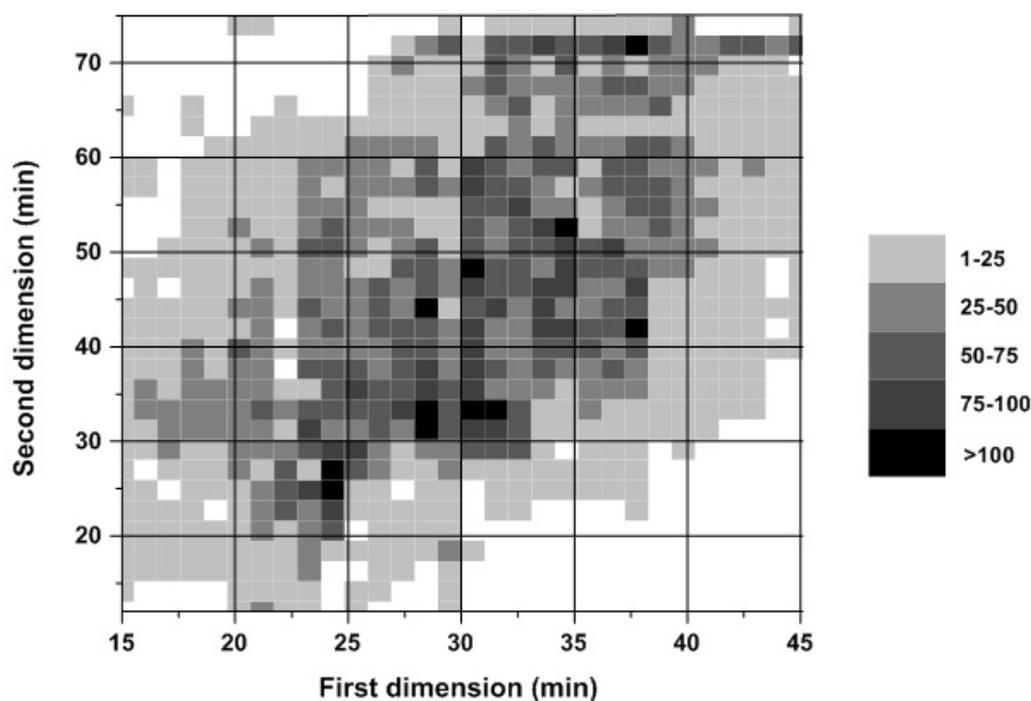


Figure 33: Peptide density in the two-dimensional separation space. The space was divided in 30x30 bins. In the legend (right side) the number of peptides identified per bin is shown in a gray scale.

Another improvement in order to reduce the analysis time could be to merge together the first and the latest 5-6 fractions of the first dimension. These fractions have the lowest peptide content. This would reduce the analysis from 31 to 20 fractions of the first dimension and save 33 LC-MALDI runs.

5.1.4. Evaluation based on peptide identification

The number of identified proteins strongly depends on the performance of the mass spectrometer and its capability to sequence peptides successfully. The 31 fractions analysed generated 93 LC-MALDI runs and 75262 precursors, but only the 27% of the masses selected for fragmentation yielded a successful identification by database search. Some of the reasons why a precursor was not correctly identified or did not match to a peptide are low concentration of the peptide, suppression effects and heterogeneous crystal morphology of the deposited samples. Peptides at low concentration are very much affected by sample depletion after several acquisitions on

the same spot. It has been reported that roughly 30% of the MS/MS spectra of a low intensity peptides was identified from the first five MS/MS acquisitions on the same spot [97]. Peak suppression can be reduced by employing longer gradients or smaller spots. However, the size of the spots on the plate is a compromise between the sample complexity and the analysis time. In the case of complex samples, smaller spot sizes (e.g. half the time) generate more spots and longer analysis time. A negative consequence is that the sample concentration per spot reduces and peptides present at low concentration may be under the limit of detection of the mass spectrometer. The number of times a peptide was successfully repeatedly sequenced by LC-MALDI MS/MS in this work is shown in Figure 34.

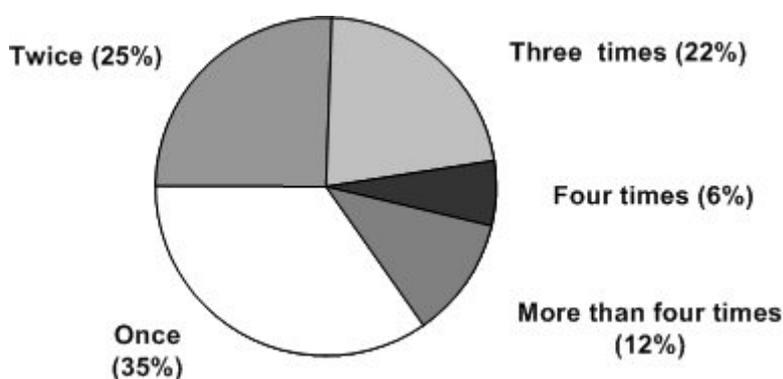


Figure 34: Number of times a unique peptide was successfully sequenced by MALDI TOF/TOF MS.

Each m/z recorded in the precursors list for MS/MS corresponds, in theory, to one peptide mass and it should be selected only once within a chromatographic run. The majority of the peptides was identified once, twice or three times due to the fact that three replicates were performed on each fraction. Peptides identified only once (35%) or twice (25%) are probably present in the sample at low concentration or have a low S/N. These peptides are available in not many spots and are more susceptible to be affected by peak suppression effects during the ionization process. Peptides identified three times were sequenced during each of the triplicate LC-MALDI measurements. These are abundant peptides are not much affected by peak suppression effects. The percentage of peptides measured more than four times was 17. Peptides sequenced more than three times are usually high abundant peptides, which were found in more than one fraction of the first dimension. Also variations in the mass accuracy during

the acquisition process can lead to the assignment of more than one precursor to one peptide during a LC-MALDI experiment.

A carry-over effect is frequently observed in SCX separations [27, 98] due to unspecific interactions with the stationary phase. Because in this work a RP separation was employed in the first dimension, a low carry-over of peptides was expected. Therefore the elution of peptides in the collected fractions was analyzed in the first dimension. These results are illustrated in Figure 35. 80% of the peptides identified by MALDI TOF/TOF MS eluted in one single fraction, 13% of the peptides were identified in two adjacent fractions and only 1% of the peptides were sequenced within three fractions.

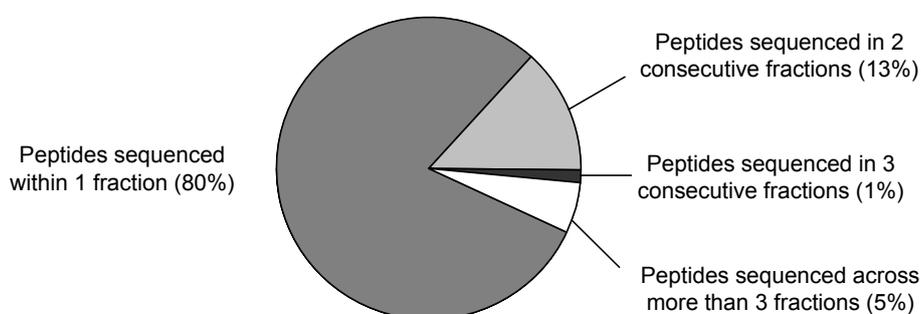


Figure 35: Peptide cross-over along different fractions. Most of the peptides (80%) eluted in one fraction whereas only 20% were identified in more than one fraction.

The residual 5% of the identified peptides were found in more than three fractions; the majority of them was identified more than four times and stem from highly abundant proteins. In conclusion, the observed cross-over effect of peptides during the first dimension was very small.

5.1.5. Evaluation based on protein identification

The total number of identified proteins returned by the Mascot search engine was 1644 proteins. 18% of these proteins were identified by matching two different peptides to their sequence, 12% by three different peptides and 44% by matching to them three or more peptides by database search. However, 26% of these proteins were single-peptide identifications (see Figure 36).

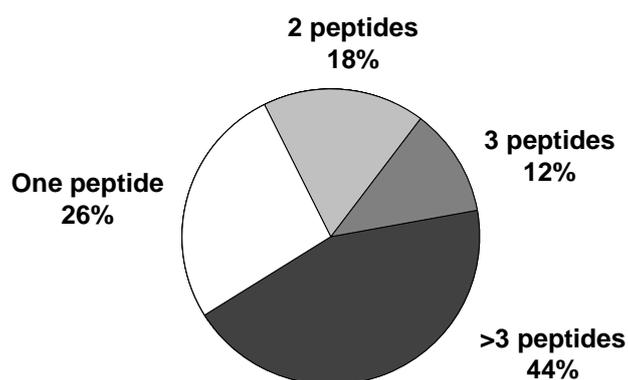


Figure 36: Protein distribution according to the number of peptides correctly assigned.

Assuming that proteins identified by at least two different peptides are confident identifications [99], 1208 proteins were definitely identified in this work. As a result, 1783 proteins of *C. glutamicum* were not identified. 52% of the not identified proteins were hypothetical proteins and 14% were membrane proteins.

It is generally assumed that high abundance proteins are more likely to be detected than proteins expressed at lower levels. An attempt to predict protein expression levels is the codon adaptation index (CAI), which is based on the degree of codon usage bias in genes. The degree of the codon usage bias has been proved to correlate with the level of gene expression in yeast [100] as well as in bacteria [101, 102]. The CAI values [103] are indicators of protein expression level and were employed in this work for estimation of protein expression. The CAI values for all predicted proteins in the *C. glutamicum* database as well as for the proteins identified by the 2D-LC-MALDI approach with more than one different peptide, were automatically generated (www.jcat.de) according to a procedure previously defined [104]. CAI close to unity indicates high expression, whereas CAI values close to zero point to low expression levels. Figure 37 illustrates the distribution of proteins of *C. glutamicum* according to their CAI value; additionally, the percentage of identified proteins from each CAI interval is shown. The graph shows that the percentage of identified proteins increases in agreement with their predicted expression level. 90% of the proteins encoded by genes that have a high CAI (CAI > 0.8) were successfully identified, as can be expected for highly expressed genes. In the range of CAI values between 0.8 and 0.4 the percentage of identified proteins is between 83 and 64%. Between 0.3 - 0.4 it was

50%, 30% between 0.2 - 0.3 and below 0.2 CAI it was only 19%. It is important to remark that the number of proteins identified within an interval of CAI values was not influenced by the number of total proteins in the interval.

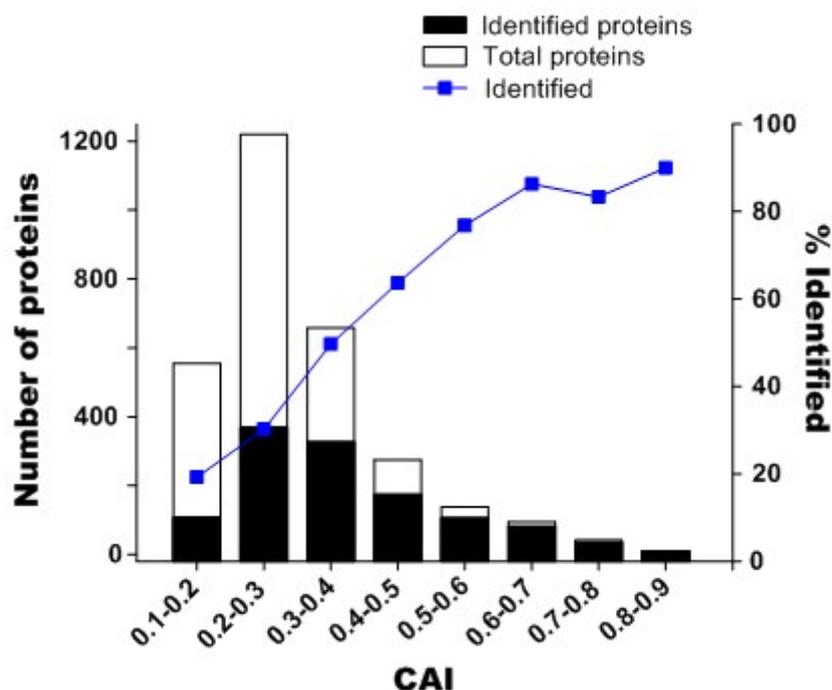


Figure 37: Distribution of CAI indexes for all predicted gene products of *C. glutamicum* (□) and the proteins identified by LC-MALDI (■). The line graph shows the percentage of identified proteins within the corresponding interval.

The CAI has been used as expression level indicators to evaluate the ability of the 2D-RP-HPLC MALDI TOF/TOF approach to identify proteins based on their abundance. A rough estimation shows that the approach presented here enables the identification of more than 80% of highly expressed proteins (CAI > 0.5), more than 50% of middle expressed proteins and more than 20% of low abundant proteins.

The proteome coverage achieved by the present method was 55% but reduces to 40% when considering only the proteins identified by more than one peptide. Taking in account that 22% of the proteins are classified as membrane proteins and the extraction protocol employed was specific for cytosolic proteins, 78% of the protein coding genes remain to be detected. In addition not all proteins are expressed under normal cultivation conditions [102]. The identification of half of them is a high-quality result as it is the highest number of proteins of *C. glutamicum* reported in a single study up to now.

5.2. Comparison of the novel 2D-RP-HPLC MALDI TOF/TOF approach with two alternative approaches for proteome analysis

In the following lines the novel 2D-RP-HPLC MALDI TOF/TOF approach for the proteome analysis of *C. glutamicum* is compared with two alternative platforms for proteome analysis. In order to estimate the efficiency of the method, the results obtained by the combination of 2D-RP-HPLC with MALDI TOF/TOF are weighed against the classical 2D-PAGE top-down approach described in section 4.4. and the approach described by Delmotte et. al. [1]. The later method employed the on-line LC-ESI-IT analysis of fractions collected by the first dimension at high pH. While the capabilities of the mass spectrometer employed the earlier study, an ion trap Esquire HCT (Bruker Daltonics Inc.), have been overcome by new instruments, the 4800 MALDI TOF/TOF represents the latest improvements in MALDI-TOF technology. As a consequence, the use of a more up to date mass spectrometer than the ESI-IT HCT for the analysis of fractions in the second dimension can lead to other improved results. Also the 2D-PAGE technique is a classical method for protein separation. However it has been little improved since it was established over twenty years ago [23] and its drawbacks limit its use for the proteome analysis [24].

5.2.1. Comparison with a top-down approach: 2D-PAGE

The cytosolic proteome of *C. glutamicum* was analysed by a classical top-down approach for proteome analysis as described in section 4.4. Separation was performed at the protein level by 2D-PAGE and identification was carried out by MALDI TOF/TOF analysis after digestion of each gel plug overnight. The results of this approach were compared with the novel 2D-RP-HPLC MALDI TOF/TOF system employed for bottom-up analysis of *C. glutamicum*.

Although the 2D-PAGE was capable of resolving around 600 protein spots, in practice around 200 gel plugs containing stained proteins were manually excised from the gel and 166 different proteins were successfully identified. As expected, the number of proteins identified by the bottom-up approach was superior to the 2D-PAGE approach. It has been reported that low abundant proteins, usually from genes with codon bias of

< 0.1, are usually not found by 2D-PAGE [96]. Thus, the potential of the 2D-PAGE technique is limited for the proteome analysis, in which a large range of expression levels are present in a sample. Therefore the comparison of the methods was performed at the peptide level. The MALDI TOF/TOF analysis of the digested spots was performed in two stages. In the first stage a mass spectrum was acquired from the spot on the plate, this is the so called peptide mass fingerprint (PMF). Subsequently, the three most intensive peaks of the mass spectrum were selected automatically for MS/MS analysis. Protein identification was carried out by combination of the information acquired by MS and MS/MS for the database search. In some cases fragmentation of the selected precursors was not successful and protein identification was subsequently based on PMF information. The comparison between the 2D-PAGE and LC-MALDI approach was based on the evaluation of the identified peptides of 46 proteins identified by both approaches. 560 peptides from the 2D-PAGE approach and 498 peptides from the LC-MALDI approach were studied. The degree of overlapping of the identified peptides is shown in Figure 38.

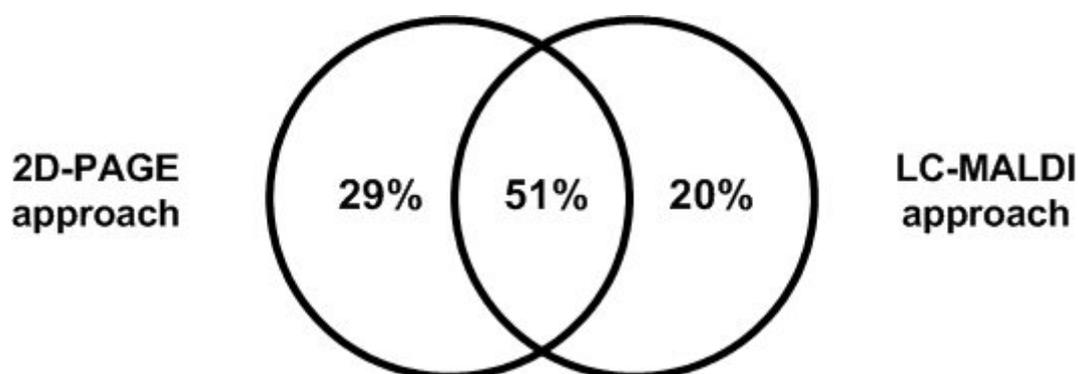


Figure 38: Venn diagram of the overlapping degree of the two sets of peptides from proteins identified by 2D-PAGE and LC-MALDI. Forty-six proteins identified by both approaches were employed for the data evaluation. The total number of non-redundant peptides was 703.

The total number of non-redundant peptides was 703 and 51% of them were successfully identified with both approaches. Each approach also supplied unique peptides, 29% of the peptides were only identified by the 2D-PAGE approach and 20% of the peptides were only identified by the LC-MALDI approach. Despite the use of the same mass spectrometer for peptide detection, the data in Figure 38 indicate that the bottom-up and the top-down approaches generated complementary information at

the peptide level. Identified peptides by each individual approach were analysed more closely in order to find out any discrimination that explains the complementarity of the methods. Figure 39 illustrates the distribution of peptides identified by each approach based on the peptide length or number of amino acids. Also peptides with one missed-cleavage were considered for analysis. The peptides solely identified by the 2D-PAGE approach were mostly small peptides of 5, 6 or 7 amino acids. The LC-MALDI approach was especially not successful to identify very small peptides composed of 5 or 6 amino acids. These peptides fall in the low mass range of MALDI MS mode which is below 900 Da. Because of the presence of matrix clusters in that m/z region these peptides were not chosen for fragmentation during LC-MALDI analysis. However, the LC-MALDI approach was capable of identifying more peptides of length 15 to 29 amino acids. This can be due to the fact that longer peptides originated by enzymatic digestion in the gel plugs can not diffuse as easy as smaller peptides from the polyacrylamide gel into the solution.

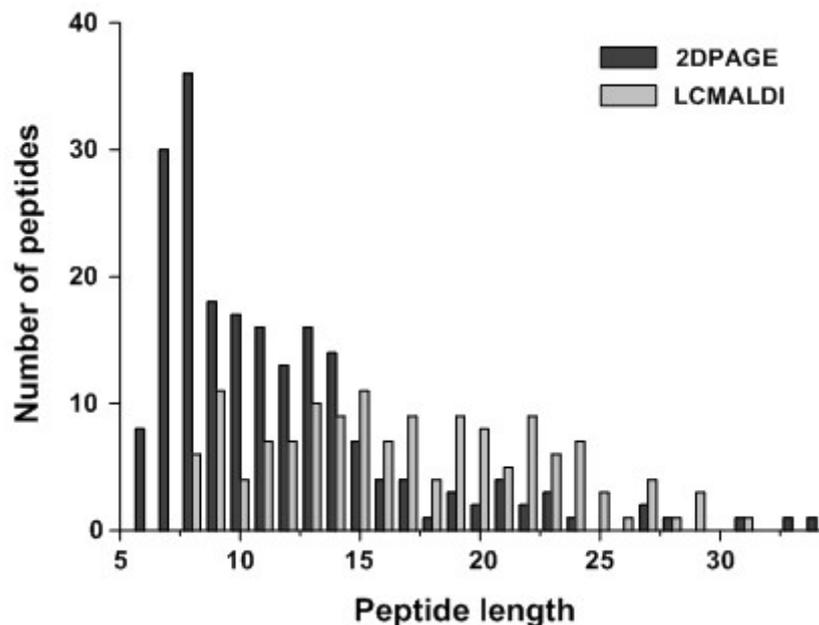


Figure 39: Distribution of peptides identified either by the 2D-PAGE approach or by the 2D-LC-MALDI approach. Peptides are represented based on the number of amino acids and belong to 46 proteins identified by both approaches.

Most of the peptides released after enzymatic digestion, which are susceptible of detection by MALDI MS are in the mass range between 800-2500 Da. Because all

peptides obtained by in-gel digest are ionized together for PMF analysis, peptide competition for the available protons can cause signal suppression [90]. Therefore peak suppression may play a role in detection of peptides originated from digestion of gel spots. These peak suppression effects are reduced if peptides are separated by liquid chromatography previous to mass spectrometric analysis. The combination of the bottom-up and top-down approaches delivered higher sequence coverage for most of the studied proteins. Figure 40 shows the sequence of a representative protein, Thioredoxin reductase of locus identification number NCgl2984 that is 317 amino acids long.



Figure 40: Sequence coverage of the protein Thioredoxin reductase achieved by the 2D-PAGE approach and by the 2D-LC-MALDI approach. Both approaches were able to identified unique peptides increasing the total sequence coverage to 83%.

The sequence coverage achieved by the 2D-PAGE approach was 58% whereas the LC-MALDI approach achieved 62%. Because each approach delivered unique peptides, the combination of both approaches raised the total sequence coverage to 83%. Consequently both methods supply complementary information, which improves the sequence coverage of the proteins by increasing the number of non-redundant peptides identified.

5.2.2. Comparison MALDI TOF/TOF vs. ESI-IT

The two mass spectrometric approaches employed for the analysis of the fractions collected during the first dimension present several differences. The different ionization methods, ESI and MALDI, are supported by different mass analyzer, ion-trap and TOF, respectively, which are based on different mass separation principles (see introduction). Furthermore the coupling of liquid chromatography and mass spectrometry requires an on-line set-up for the ESI-IT analysis and an off-line set-up for the MALDI approach. Despite the use of the same chromatography set-up and due to the intrinsic differences of the mass spectrometers mentioned above, some differences in the results could be expected. In Table 6 some of the features concerning the results delivered by MALDI MS and ESI MS are compared.

Table 6: Comparison of MALDI and ESI results based on the analysis of the digested proteome of *C. glutamicum* by 2D-RP-HPLC.

Parameters	MALDI TOF/TOF	ESI-IT
Spectra acquired	75262	25209
Total identified peptides	20553	10882
Unique identified peptides	7416	2709
Total identified proteins (% of the proteome)	1644 (55%)	745 (25%)
Identified proteins with two or more different peptides (% of the proteome)	1208 (40%)	468 (16%)
False positive rate	2.4%	2.3%
MS average mass accuracy	20 ppm	218 ppm

The MALDI approach delivered undoubtedly a higher number of identifications and consequently higher proteome coverage for all identified proteins as well as for proteins identified with more than one peptide. However the ESI approach yielded a better efficiency of MS/MS spectra successfully assigned (43%) compared to the MALDI approach (27%). The MS/MS efficiency was calculated by dividing the number of assigned MS/MS spectra by the total number of spectra acquired. One of the advantages of the MALDI approach relies on the off-line coupling because the

analysis time is not restricted to the chromatographic elution. Under these conditions only the amount of sample on the plate is the limiting factor for analysis of a spot. However, the time required for the complete experiment was longer than in the ESI approach: about 375 hours were required to analyse the thirty-one fractions in triplicates by MALDI TOF/TOF MS, whereas 124 hours were required for the analysis by ESI-IT. The improvement of the MS/MS efficiency by using a precursor ion exclusion list could yield more positive identifications [97] and thus reduce the analysis time. Such an exclusion list contains all the identified peptides in the preceding runs and avoids subsequent redundant acquisitions. As a consequence, sample depletion is reduced and identification of low abundant peptides is improved. As expected the mass accuracy in MS mode delivered by the TOF mass analyzer was ten times better than the mass accuracy provided by the ion trap. The calculated false positive rate was similar in both approaches.

The MALDI approach yielded 2.7 times more unique peptides than the ESI approach. As a consequence, a better sequence coverage of the identified proteins was achieved by the MALDI approach (see Figure 41). The sequence coverage of a protein is calculated relative to the number of amino acids of the protein and expressed in percentage. Both approaches delivered a similar number of proteins with sequence coverage between 15 and 25% but MALDI yielded more proteins than ESI with sequence coverage above 25%.

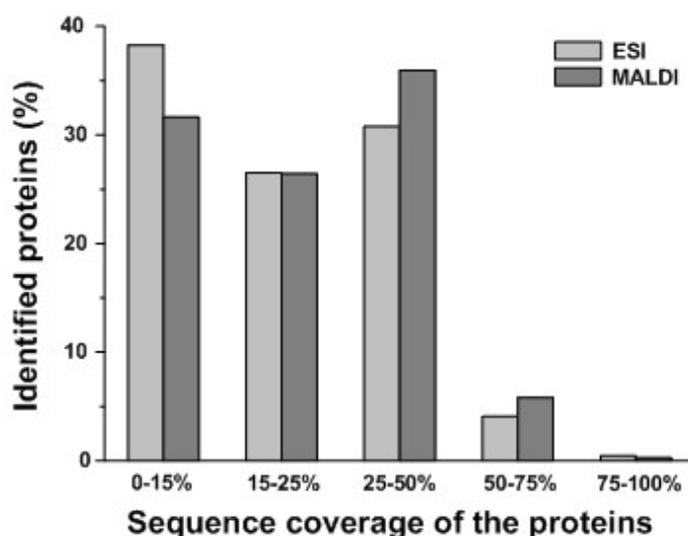


Figure 41: Comparison of the sequence coverage reported for the identified proteins by the ESI and MALDI approaches.

In the next lines, the differences between identified peptides by the ESI- and MALDI-approaches were investigated based on peptide mass, pI and hydrophobicity. The masses of the 95477 peptides from the theoretical digest were computed and compared with the masses of the peptides identified by both approaches (see Figure 42). Figure 42 illustrates the distribution of the peptides masses obtained by *in-silico* digest in the mass range of 800 to 4000 Da. For the MALDI approach the mass range of the fragmented precursors was m/z 800-3000, whereas in the ESI the range was 600-3900. Both approaches were especially not efficient in the detection of peptides with masses below 1000 Da in comparison with the theoretical data. Although 51 peptides with molecular masses below 800 Da were sequenced by the ESI approach, MALDI provided a superior rate of identified peptides for masses below 1000 Da.

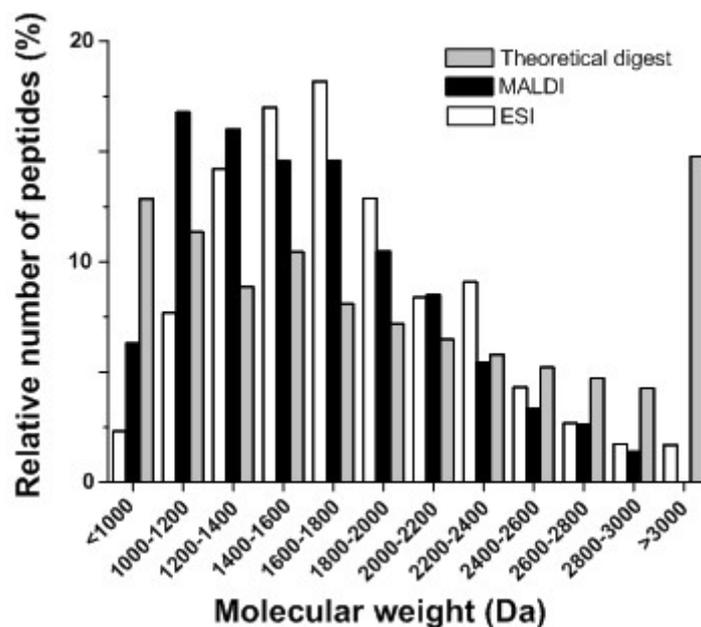


Figure 42: Distribution of peptides according to their masses expressed in percentage. The number of peptides obtained by theoretical digestion was 95477. The peptides used for the evaluation of data correspond to proteins identified by more than one peptide. The percentage was calculated with respect to all the peptides within a class.

In the mass range 1000 to 1400 Da MALDI showed better detection efficiency. The tendency inverts for higher masses and the ESI approach delivered a higher rate of successful identifications in the mass range 1400 - 2400 Da. Both approaches are equally comparable for mass ranges 2000 - 2200 and 2600 to 3000. With respect to peptide masses above 3000 Da the relative number of peptide identifications was lower than the predicted. This can be due to the fact that the theoretical approach

considers all possible one missed cleavage situations derived from the *in-silico* digest. The presence of one missed cleavage results always in a longer peptide. However in the experimental digest missed cleavages occur randomly. The ESI approach has an intrinsic advantage over MALDI for analysis of high mass peptides because during the ionization multiple charged ions (+2 and +3) are formed.

To compare the chemical properties of the identified peptides by MALDI and ESI, the pIs and hydrophobicities of the peptides were computed and are shown in Figure 43 and Figure 44. Figure 43 shows the distribution of identified peptides according to their calculated pI. The ESI approach identified more acidic peptides (pI < 6), whereas the MALDI approach tends to favor identification of peptides with pI above 6. Consequently it can be concluded that the MALDI approach yields preferentially more basic peptides. These observations are in accordance with data previously reported [105].

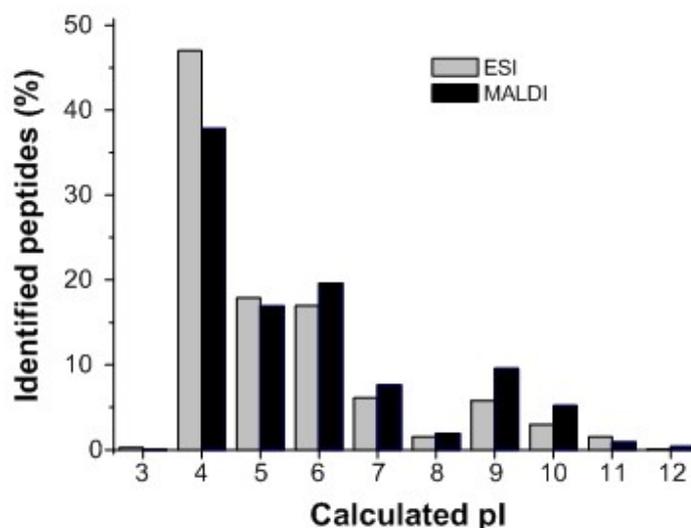


Figure 43: Calculated pI of the identified peptides by the ESI and MALDI approaches. Percentages are referred to the total number of peptides of the corresponding approach.

Hydrophobicities were computed according to the normalized scale of Eisenberg [14], where higher values indicate hydrophobic amino acids and negative numerical values indicate hydrophilic amino acids. In the present study it was found that the MALDI approach identified slightly more hydrophobic peptides (Figure 44). This fact could represent an advantage for the identification of peptides from proteins with transmembrane domains. However, there is no agreement about the preferential

ionization of hydrophobic or hydrophilic peptides by MALDI or ESI in previous studies. It has been reported that MALDI ionizes preferentially hydrophilic peptides [106], but it has also been reported that more hydrophobic peptides are preferentially ionized in CHCA matrix [90].

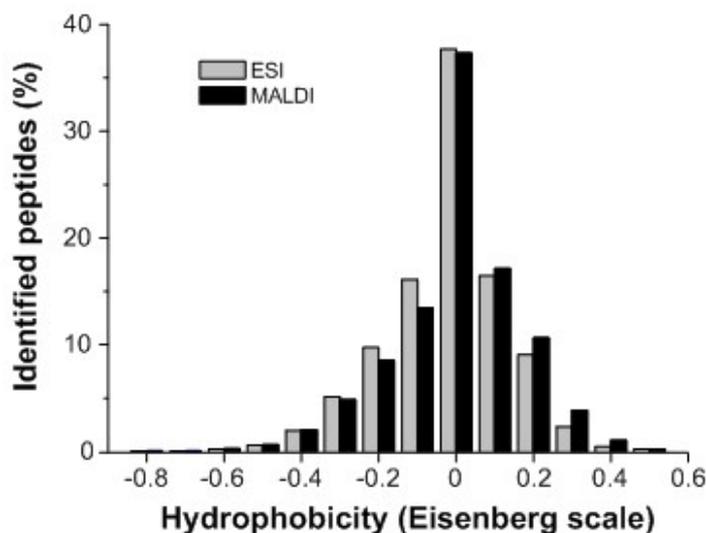


Figure 44: Hydrophobicity of the identified peptides by the MALDI and ESI approaches calculated according to the Eisenberg scale.

Previous studies have found that the MALDI technique is more prone to identified tryptic peptides ending with arginine, whereas ESI identifies rather peptides with a lysine at the C-terminus [105, 107]. The ratio of lysine to arginine (K/R) calculated for the theoretical digest was 0.68. The ratio for MALDI and ESI approaches was found to be 0.49 and 0.76, respectively. These results are in accordance to previous results that showed a tendency of MALDI to identified peptides with arginine, which is the most basic of all amino acids and has the highest proton affinity [108]. The frequencies of the different amino acids within the identified peptides were examined in Figure 45 for both MALDI and ESI techniques, and for the predicted peptides of the *in-silico* digest. Amino acids were grouped according to the properties of the side-chains for plotting. The acidic amino acids are not shown because the frequencies of both aspartic acid and glutamic acid were almost equal for the MALDI and ESI approaches (see Appendix IX). Besides that, the identification of these amino acids was higher than the predicted by the *in-silico* digest. The amino acids occurrences varied between 0.6% for cysteine identified by ESI and 7.8% for alanine identified by MALDI. The most pronounced differences are observed for the basic amino acids. As mentioned before,

ESI ionizes preferentially lysine containing peptides and MALDI arginine containing peptides. Besides, the identification of histidine is preferred by the two ionizations techniques compared to other amino acids. As a consequence, the frequency of occurrence of this amino acid for MALDI and ESI is higher than in the theoretical digest. The ESI approach showed a slight preference for aliphatic amino acids when compared with MALDI. MALDI showed a bias towards identification of aromatic amino acids, specially phenylalanine and tryptophan. In the case of amino acids with a hydroxyl group, they were preferentially identified by ESI, especially threonine. The identification of cysteine and methionine containing peptides was lower for both MALDI and ESI approaches than in the theoretical digest; however MALDI reported slightly more peptides than ESI.

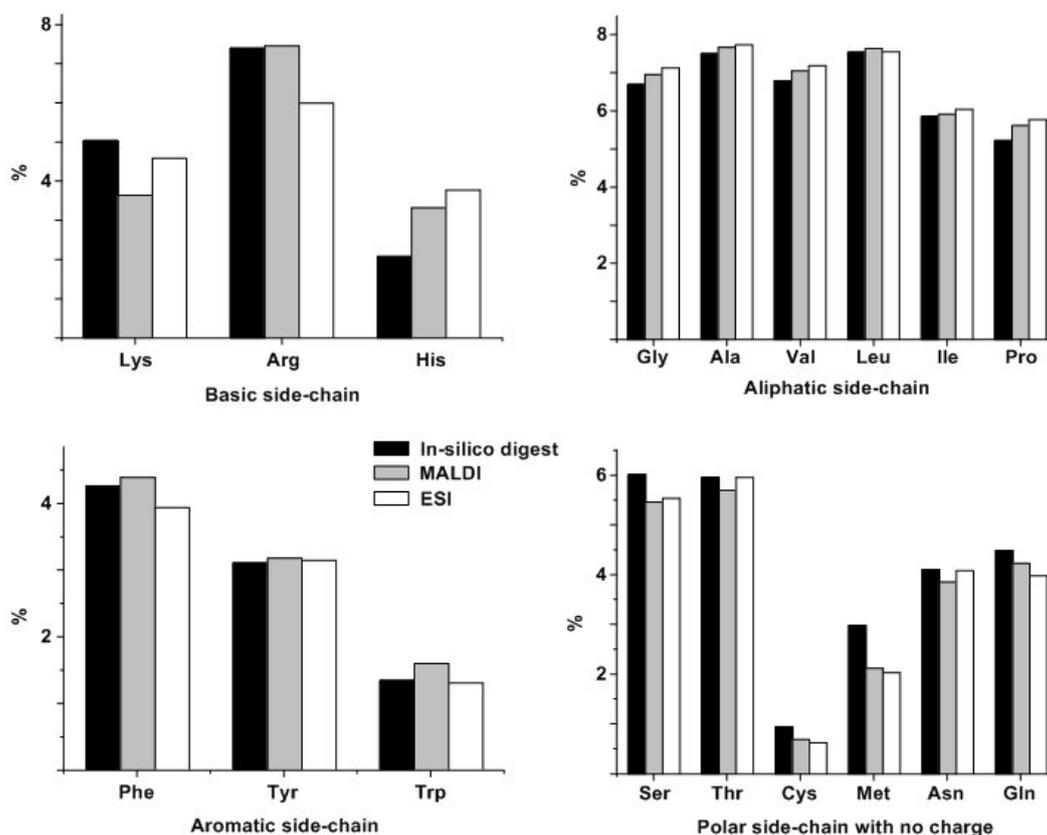


Figure 45: Relative occurrences of amino acids expressed as percentages for MALDI and ESI approaches. The percentage is referred to the number of amino acids from all identified peptides. The amino acids are grouped according to their side-chain properties.

In conclusion, the MALDI approach favors ionization of peptides containing arginine and aromatic amino acids. The ESI approach shows a preference to ionize peptides containing lysine, histidine and hydroxy amino acids. If the data are compared with

the predicted peptides, it can be concluded that both methods tend to preferentially ionize peptides containing histidine, aliphatic side-chain amino acids, aromatic amino acids and acidic amino acids.

The complementarity of the ESI and MALDI approaches at peptide level was investigated based on the unique peptides identified by each spectrometric approach (Figure 46).

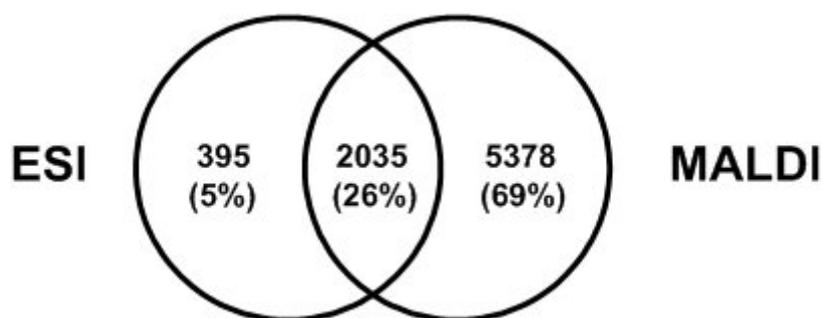


Figure 46: Venn diagram of the overlapping degree for the identified peptides by the MALDI approach and ESI approach. Only peptides of proteins identified by more than one unique peptide were considered.

The studied peptides belong to proteins identified by more than one peptide. The number of peptides considered was 2431 and 7413 for ESI and MALDI respectively. 26% of the peptides were common to both mass spectrometric approaches. However the MALDI approach delivered a higher number of unique peptides, 69%, compared to the 5% delivered by the ESI approach. Because the chromatographic separation employed was the same, the differences in peptide identification are due to the hyphenation to the mass spectrometer and differences in the ionization processes. Unique peptides identified exclusively by one of the methods were analyzed more in detail according to their peptide length. Peptide relative distribution according to the number of amino acids, expressed in percentage, is shown in Figure 47. The distribution of unique peptides in percentage is slightly different for both approaches. However the ESI approach shows a better detection for very short and very long peptides. Although the MALDI approach delivered higher number of unique peptides, a deficiency for the detection of short peptides (5-7 amino acids) and long ones (>26 amino acids) was observed. This is a direct consequence of the mass range selected for MS/MS analysis by LC-MALDI. Precursor's masses below 900 were not selected for fragmentation in order to avoid selection of matrix clusters.

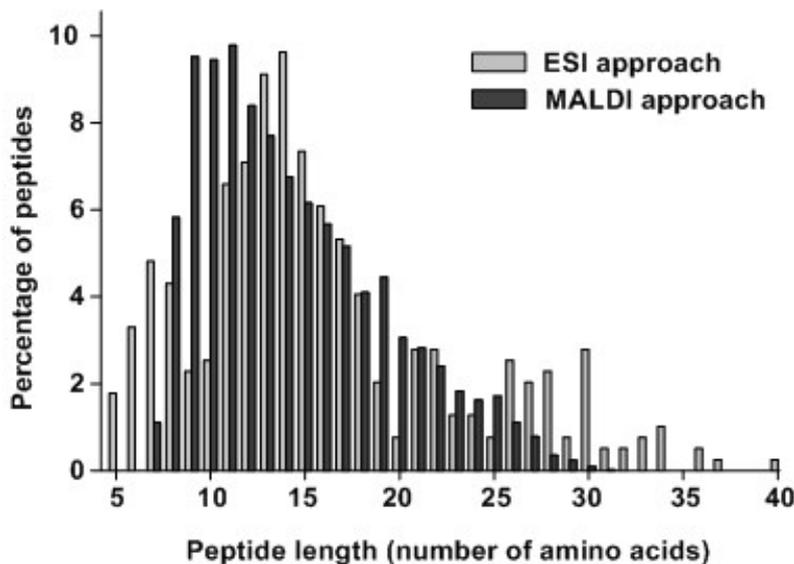


Figure 47: Distribution of peptides identified exclusively by one of the approaches, according to the number of amino acids in the peptide.

5.2.3. Comparison of methods based on protein identification

The evaluation of the proteins identified by the three proteomic approaches was based on proteins identified by more than one peptide. Therefore the number of proteins considered for comparison was: 166 proteins identified by 2D-PAGE, 468 proteins identified by 2D-RP-HPLC-ESI-IT and 1208 proteins identified by 2D-RP-HPLC MALDI TOF/TOF.

In previous studies, a complementary identification has been claimed between LC-MALDI MS/MS and LC-ESI MS/MS approaches [87, 89]. Although a small complementarity was observed at the peptide level between both approaches, at the protein level the LC-MALDI approach was clearly superior to the other approaches. Figure 48 illustrates the protein identifications shared by the different approaches. The LC-MALDI approach showed a higher total number of identifications including 726 unique identified proteins not identified by the other two approaches. 144 of the proteins overlapped for the three employed methods. These shared proteins are essentially enzymes present in the main metabolic pathways. These highly expressed proteins are more likely to be detected by all of the approaches.

Twenty-one proteins were identified by the MALDI and 2D-PAGE approach but not by the ESI approach. However, fifteen of them were found in the list of proteins identified by one unique peptide in the ESI approach. The other six proteins which could not be successfully identified by ESI have protein masses between 22 and 35

kDa. However, there is no apparent reason why they should not be identified by the ESI approach. The corresponding protein stained spots showed low intensity in the Coomassie stained polyacrylamide gel. The CAI values are between 0.3 and 0.6 and thus they are neither highly expressed nor low expressed proteins.

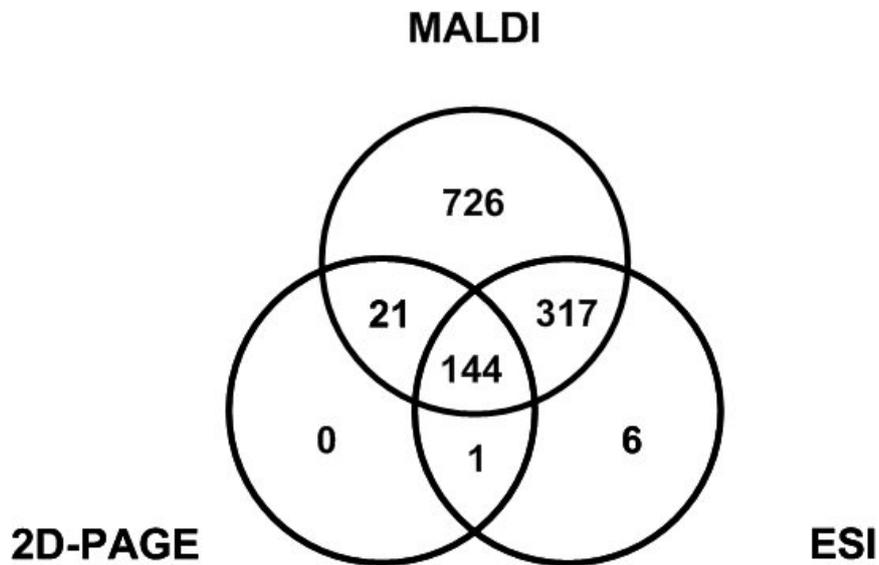


Figure 48: Comparison of methods based on identified proteins by the three approaches employed for the proteome analysis of *C. glutamicum*. The Venn diagram shows the degree of overlapping in the identifications.

These proteins identified by MALDI but not by the ESI approach were investigated more closely. Based on the retention time of the peptides from these proteins identified by the LC-MALDI approach, it was demonstrated that the identified peptides eluted from the column at times that show strong co-elution with other peptides. The retention times of these peptides are found in high peptide density areas from the Figure 33. Strong peptide co-elution appears to be the most probable cause of the failure to detect them by the ESI approach. The number of co-eluting peptides probably exceeds the number of MS/MS spectra that the mass spectrometer can acquire in one data-dependant analysis cycle. In the case of strong peptide co-elution the approach based on off-line coupling shows enhanced detection ability.

One single protein with locus ID NCgl0385 was exclusively identified by the 2D-PAGE and ESI approaches. Nevertheless one of the peptides was sequenced by the LC-MALDI approach and therefore belongs to the one-hit wonder proteins identified by MALDI. The six proteins identified exclusively by the ESI approach were

identified by the MALDI approach as one-hit proteins (see Table 7). Except NCgl1841 (MW 70kDa), the other proteins are low molecular mass proteins (below 16 kDa). Smaller proteins generate less possibly tryptic peptides within the m/z range of the mass spectrometer [30].

Table 7 shows that the six proteins identified exclusively by the ESI approach were identified by the MALDI approach, however by one peptide hit only. In all cases the peptide was sequenced more than once. Two of the proteins shown in Table 7, NCgl0459 and NCgl2493, were identified by ESI based on two peptide hits. One of the peptides was also identified by MALDI and the sequence is given in Table 7. The second peptide had molecular mass over 3000 Da and charge +3. As already shown in Figure 42, the ESI approach was more successful than the MALDI approach for the identification of high molecular mass peptides due to the advantage of multiple charged peptides generated by the ESI ionization process.

Table 7: List of proteins identified by the ESI approach with more than one non-redundant peptide and by the MALDI approach with only one sequenced peptide.

Protein	Single-peptide sequenced by MALDI TOF/TOF	Times the peptide was sequenced
NCgl0459 Ribosomal protein L11	GNVVPVEITVYEDR	3
NCgl0519 Ribosomal protein L30/L7E	HTVIRPDTPEVR	6
NCgl0930 hypothetical protein	ATEFLDSDSGEQK	2
NCgl1448 pyrophosphohydrolase [3.6.1.31]	TFDSL YEELLNR	3
NCgl1861 phosphotransferase system [2.7.1.69]	FAGKPVIESGVK	3
NCgl2493 hypothetical protein	YSSPDMDLDSLQR	3

In the case of the protein NCgl1861 a complementarity of the methods was observed: the two peptides sequenced by the ESI approach were different from the peptide identified by the MALDI approach.

The comparison of the methods based on pI and molecular weight of the identified proteins is shown in Table 8. In addition, in the table are listed the proteins showing extreme pI and MW values from the predicted proteome of *C. glutamicum*. The MALDI approach identified the most acidic, most basic and smallest protein of all the approaches. The identified proteins with highest and lowest pI were hypothetical proteins which have pI 11.72 and 3.3 respectively. The protein with highest molecular weight was 3-oxoacyl-(acyl-carrier-protein) synthase, which has 316 kDa and the smallest protein identified with at least two different peptides was the ribosomal protein L33 (6.4 kDa).

Table 8: Comparison of the three methods for the proteome analysis based on pI and MW of the proteins identified from *C. glutamicum*.

Approach	Lowest pI	Highest pI	Lowest M_w (kDa)	Highest M_w (kDa)
2D-RP-HPLC	3.3	11.7	6.4	316
MALDI TOF/TOF	(NCgl1090)	(NCgl0838)	(NCgl0833)	(NCgl2409)
2D-RP-HPLC ESI-IT	3.8	11.6	6.8	316
	(NCgl1057)	(NCgl0834)	(NCgl0519)	(NCgl2409)
2D-PAGE &	3.9	6.63 (8.8)	8.7	315
MALDI TOF/TOF	(NCgl1915)	NCgl0501(0460)	(NCgl2501)	(NCgl0802)
Predicted proteins	3.08	13.62	3.7	316

The results generated by the ESI approach were very close to the MALDI approach. The 2D-PAGE was comparable to the other approaches for identification of high MW proteins and acidic proteins. As mentioned earlier it was not possible to separate and detect basic proteins by 2D-PAGE. In accordance with previous studies, 2D-PAGE is not suitable for identification of proteins of *C. glutamicum* which have basic pI. The present work also failed to identify small proteins (below 8 Da).

In order to investigate the ability of the different approaches to identify the enzymes involved in the central metabolic pathways, the enzymes of the TCA cycle, the glycolysis and the pentose phosphate pathway are represented in Figure 50, Figure 49

and Figure 51, respectively. Enzymes identified by the three methods, MALDI and ESI or just MALDI are marked with an ellipse, square or dashed square respectively. The information shown here is a compendium of the information provided by three different sources: KEGG (Kyoto Encyclopedia of Genes and Genomes; www.genome.jp/kegg/), ExPASyProteomics Server (www.expasy.org) and CoryneRegNet Release 4.0 of the Bielefeld University [109].

CENTRAL METABOLIC PATHWAYS

TCA cycle

None of the methods was able to identify the cytochrome subunit of the succinate dehydrogenase and the β -subunit of the succinyl-CoA synthetase. The cytochrome subunit has been described as a membrane protein and therefore is not achievable by the extracting protocol used here, which was optimized for cytosolic proteins.

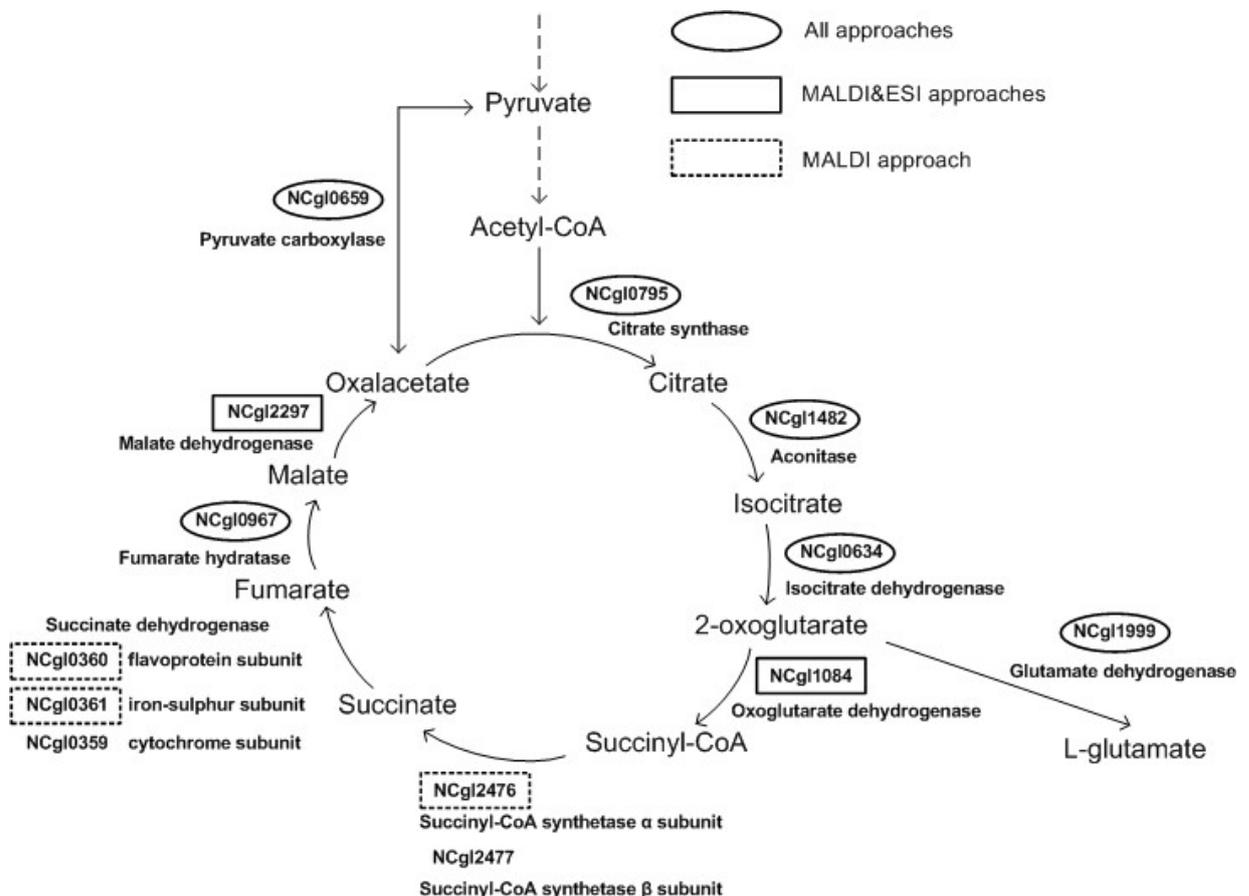


Figure 49: Enzymes of the TCA cycle which were identified by the three approaches, only by MALDI and ESI approaches or only by the MALDI experiment.

The β -subunit of the succinyl-CoA synthetase was, however, listed in the one-hit protein list of the MALDI approach and of the ESI approach. Interestingly the peptide reported by ESI was different (VIPVLIK) from the peptide sequenced by MALDI (MDLFEYQAR), which represents a case of complementary information provided by both techniques. The 2D-PAGE based approach failed in the identification of oxoglutarate dehydrogenase and malate dehydrogenase. MALDI exclusively identified the other two subunits of the succinate dehydrogenase as well as the α -subunit of the succinyl-CoA synthetase. The subunits of the succinyl-CoA synthetase have low CAI values and therefore it can be estimated that they are low expressed proteins. The biosynthesis of L-glutamate from oxoglutarate is also shown in Figure 49. The enzyme that catalyzes this reaction, the glutamate dehydrogenase, was identified by all three approaches.

Glycolysis

All the enzymes of the glycolysis pathway taking part in the transformation of the glucose to pyruvate were identified by the MALDI approach. Only one enzyme, locus ID NCgl1860, was not identified by the ESI approach. The 2D-PAGE approach was capable of identifying the enzymes which catalyze fructose-6-P to pyruvate. However the 2D-PAGE approach failed to identify the enzymes involved in phosphorylation and isomerization of glucose, as well as the subunits of the pyruvate dehydrogenase complex.

Pentose phosphate pathway

The pentose phosphate pathway is shown in Figure 51. All enzymes which catalyze the reactions of this metabolic pathway were identified by the LC-MALDI approach. Fewer enzymes were identified by the 2D-PAGE approach and one enzyme, NCgl 1536, was not identified by ESI. However it was found that the ESI approach reported one non-redundant peptide for this protein. Other representative pathways of amino acid biosynthesis in *C. glutamicum*, like lysine biosynthesis, and methionine biosynthesis are given in the Appendix X and XI.

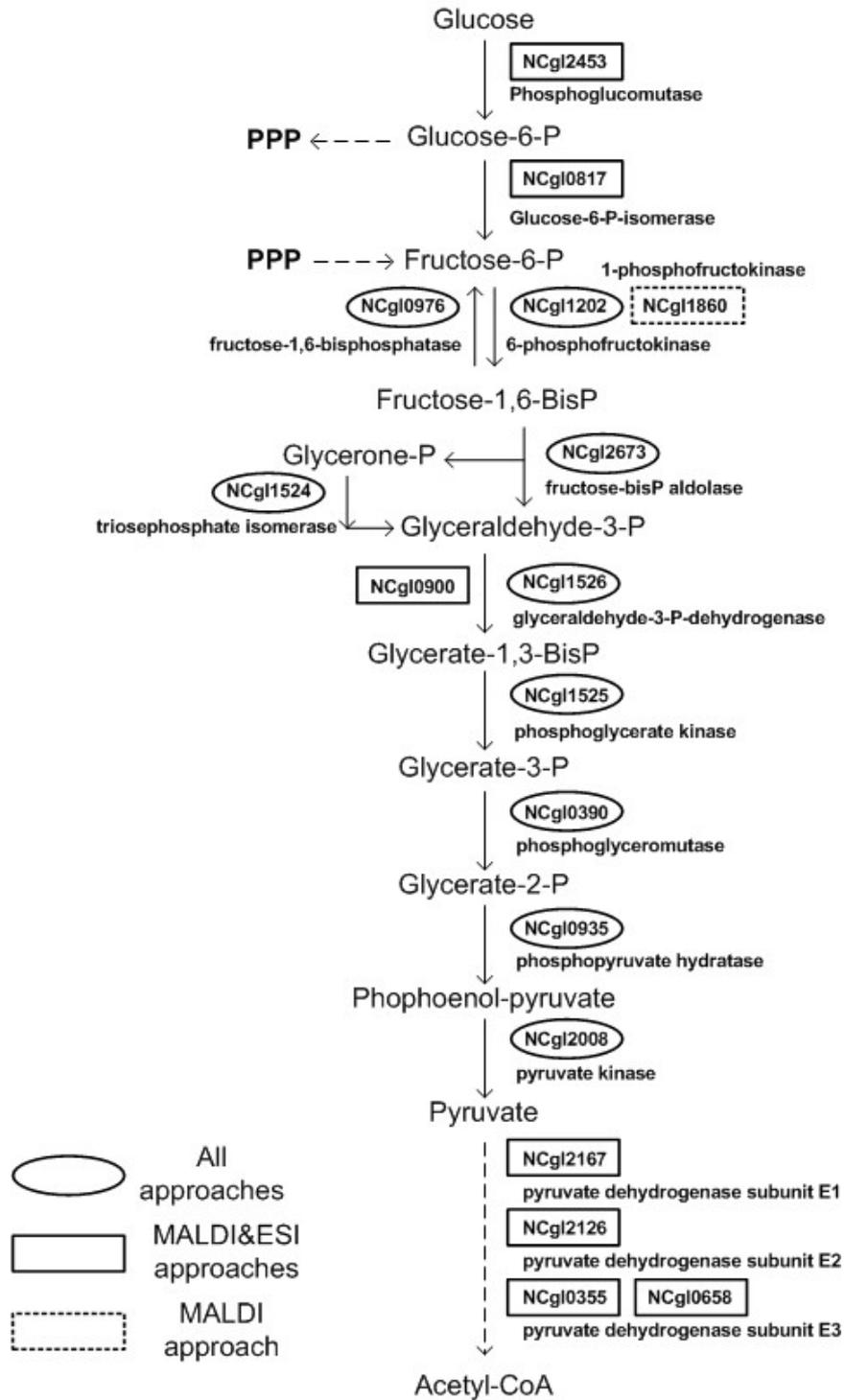


Figure 50: Enzymes of *C. glutamicum* participating in the glycolysis and identified by the different approaches.

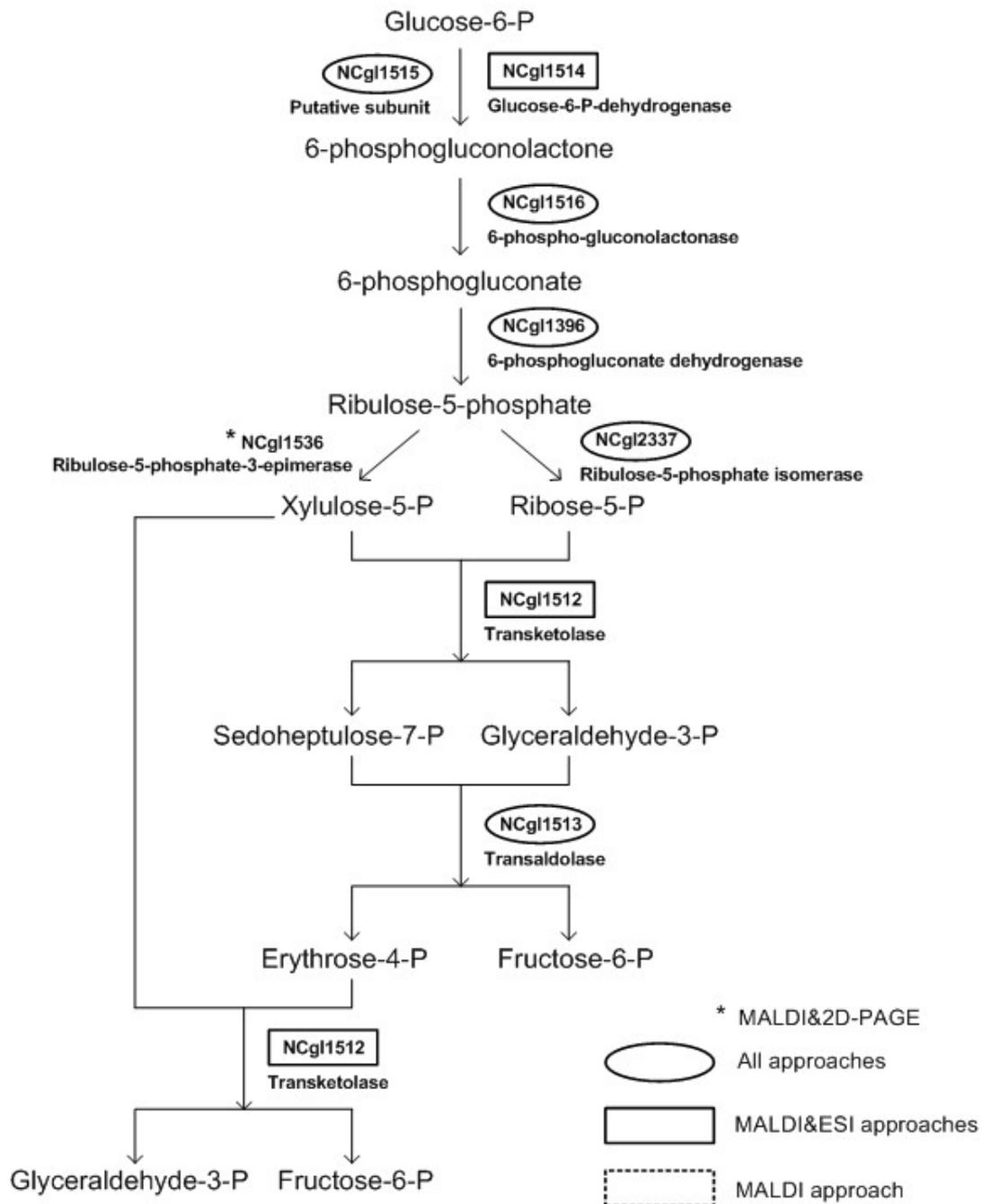


Figure 51: Biochemical reactions involved in the pentose phosphate pathway and enzymes that catalyze the reactions.

5.2.4. Comparison of methods based on the codon adaptation index

Protein expression levels have been shown to correlate well with codon adaptation index and mRNA expression for *Mycobacterium smegmatis* [102]. *C. glutamicum* is a family related organism and therefore a similar behaviour is expected. It was decided

to evaluate the three proteomic approaches employed on the proteome analysis of *C. glutamicum* according to protein expression levels predicted by CAI. For this purpose, the CAI values were computed for all proteins of *C. glutamicum* and for the proteins identified in each approach. Figure 52 shows the percentage of proteins identified by each analytical approach within different CAI intervals and referred to the total number of proteins in the considered interval. Considering that the CAI is an indicator of protein expression some conclusions can be drawn for the different sensitivities of the compared approaches.

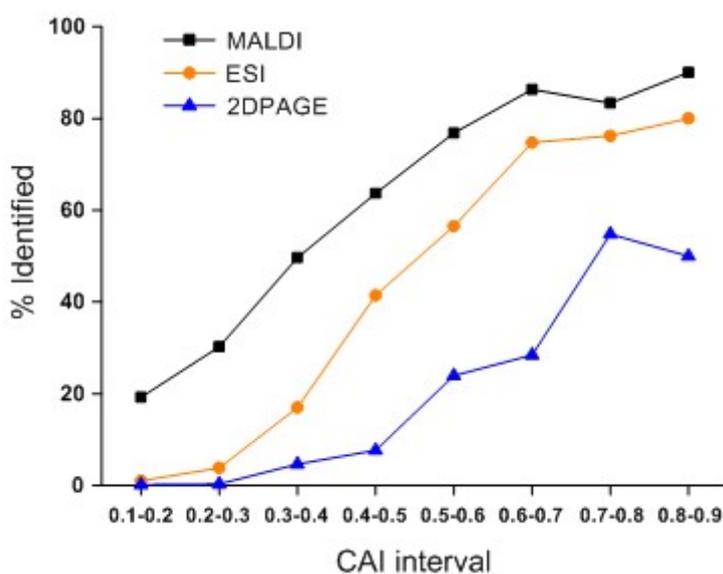


Figure 52: Sensitivity of the three approaches (2D-PAGE, 2D-RP-HPLC-ESI-IT, 2D-RP-HPLC MALDI TOF/TOF) based on the CAI. The relative abundance of identified proteins within a CAI interval is plotted for each analytical method employed in the analysis of *C. glutamicum* proteome. Percentages are calculated with respect to all the predicted proteins of *C. glutamicum* grouped into CAI intervals.

The most sensitive approach was the LC-MALDI TOF/TOF, followed by LC-ESI-IT and 2D-PAGE. In all the approaches protein identifications were in agreement with the protein expression levels indicate by the CAI. The MALDI-approach and ESI-approach delivered comparable results for high expressed proteins with CAI-values between 0.6 - 0.9. The percentage of identified proteins was 75 and 90% for ESI and MALDI, respectively. The electrophoresis based approach yielded significantly less identified proteins for this CAI values. At lower expression levels, the MALDI-based approach showed significantly better results. For example, in the CAI range between 0.2 - 0.3, MALDI yielded 30.2% (396 proteins) of the predicted proteins, but ESI only

3.9% (47 proteins). The lowest protein expression levels are indicated by CAI values between 0.1 and 0.2 and the MALDI approach yielded 107 proteins (19.3%) compared to 6 proteins (1.1%) by the ESI approach. No proteins having CAI below 0.3 were identified by the 2D-PAGE approach. Low abundant proteins with low CAI values are usually not identified by 2D-PAGE [96]. According to Figure 52 the limit of detection given by the CAI is situated in 0.3 for the 2D-PAGE and 0.2 for the ESI approach, whereas the MALDI approach was capable of identifying proteins of all CAI values.

The data presented here show that the three approaches are biased towards detection of highly expressed proteins. However, the MALDI approach, which was the most sensitive method, permitted the identification of 19% of the lower abundant proteins. Furthermore, the low abundant proteins were identified from a complex mixture of peptides obtained by digestion of the cytosolic proteins of *C. glutamicum*. These results suggest that the LC-MALDI approach is able to expand the dynamic range of protein identification further than the other two methods employed for the proteome analysis of *C. glutamicum*.

5.3. Comparison with results from literature

5.3.1. Comparison with the MudPIT approach

The highest number of identified proteins of *C. glutamicum* reported up to now in the literature was published by Fischer et. al. from a lysine producing strain [78]. Two optimized prefractionation protocols for extraction of membrane proteins were used for enrichment in membrane proteins. Both resulting samples were enzymatically digested by a trypsin/chymotrypsin mixture in an organic solvent. Subsequently, the peptides were separated using the multidimensional protein identification technology (MudPIT) [30, 31] coupled to a Finnigan LTQ ion trap mass spectrometer (Thermo Electron Cop., San Jose, CA) for detection. The combination of the results from both extraction protocols yielded the identification 326 integral membrane proteins and 1181 soluble proteins of *C. glutamicum*. The comparison of the proteins identified by the 2D-RP-HPLC MALDI TOF/TOF platform used in this work with the proteins identified by Fischer et al. [78] is illustrated in Figure 53. The identified proteins

include single-peptide based identifications in both approaches. Because nine of the 1181 soluble proteins identified by the MudPIT approach were not referred to a gene number, the comparison was carried out with the rest 1172 proteins.

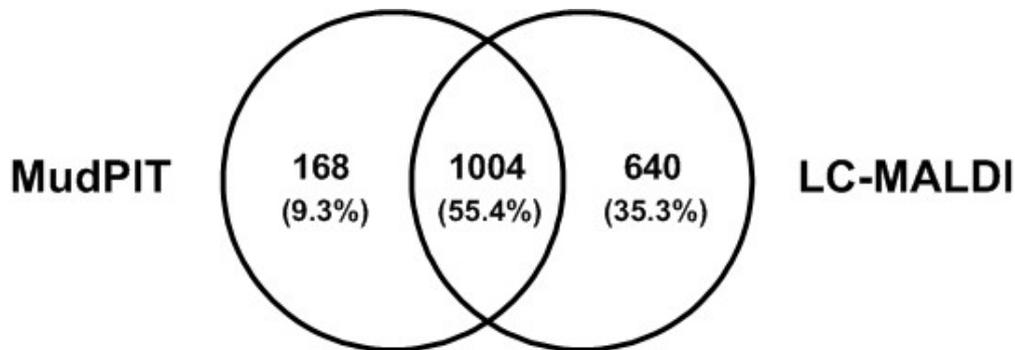


Figure 53: Overlapping degree of the identified proteins. A total of 1812 different proteins were identified by both proteomic platforms. 55% of them were common to both approaches.

In total, 1644 proteins identified by the LC-MALDI approach were compared with 1172 proteins identified by application of MudPIT technology. The total number of different proteins was 1812, which represents 60.5% proteome coverage of *C.glutamicum*. Although the extraction protocols employed in both studies were very different and focused on proteins with different properties, 55.4% of the proteins were common to both approaches. As expected, the approach employing an extraction protocol exclusive for cytosolic proteins identified more of these proteins: the LC-MALDI yielded 3.8 times more soluble proteins.

5.3.2. Comparison with a 2D-PAGE approach

The investigation of the cytoplasmic proteins of *C.glutamicum* by 2D-PAGE carried out by Schaffer et. al. [74] reported the detection of 970 Coomassie stained protein spots. The amount of protein loaded in the gel was 300 µg and the pI range used in the first dimension was 4 - 6. The successful identification of 169 spots by PMF MALDI TOF MS yielded 152 different proteins, including 35 proteins of no known function. The proteins identified by this approach will be compared with the proteins identified in the present work by 2D-PAGE analysis and described in 4.4. Because the proteins identified by Schaffer and colleagues did not have a gene number code, both

approaches were compared based on the protein name and the results are shown in Table 9.

Table 9: Cytosolic proteins of *C.glutamicum* identified by 2D-PAGE and MALDI MS by Schaffer et al. [74] (a) or in the present work (b) (see 4.4).

Identified proteins	Matched peptides (a)	Matched peptides (b)	Spot no. (Appendix I)
Glycolysis			
Phosphoglycerate kinase	16	20	8
Enolase	13	14	749
Phosphofructokinase	5	19	781
Triosephosphate isomerase	10	12	81
Phosphoglycerate mutase	14	12	793
Fructose-bisphosphate aldolase	12	x	x
Glyceraldehyde-3-phosphate dehydrogenase	10	15	768 &769
Fructose-1,6-bisphosphatase	x	12	745
Pyruvate kinase	x	16	724
Pentose phosphate pathway			
Glucose-6-phosphate-dehydrogenase	6	7	120
6-phospho-gluconolactonase	x	9	121
6-phosphogluconate dehydrogenase	x	22	753
Ribulose-5-phosphate isomerase	7	6	837
Ribulose-5-phosphate-3-epimerase	3	5	816
Transketolase	16	x	x
Transaldolase	x	16	761
TCA cycle			
Citrate syntase	x	18	742
Aconitase	11	33	50
Isocitrate dehydrogenase	16	32	715
Fumarase	19	25	735
Malate quinone oxidoreductase	17	15	722

Almost the same enzymes of the main metabolic pathways were identified by both approaches. The enzymes which were omitted in Table 9 were not identified by any of the methods. In 2D-PAGE approaches, protein detection is usually restricted to the abundant proteins of the sample [96].

6. Summary and conclusions

The common methods for the proteome analysis involve top-down and shotgun bottom-up approaches. In this work an alternative bottom-up approach based on 2D-RP-HPLC and MALDI TOF/TOF has been established as a method for the proteome analysis and tested with *C. glutamicum*. Besides that, an improved procedure for extraction of the cytosolic proteome of *C. glutamicum* was developed. The extraction method supplies a high recovery of proteins without contaminants and permits direct analysis of the sample without additional pre-concentration treatment.

The pH based 2D-RP-HPLC system represents some advantages over the classical 2D-SCX-RP application for the separation of peptides from a digested proteome sample. The absence of salts in the first dimension facilitates sample concentration by simple evaporation of the acetonitrile. A high selectivity of peptides in the first dimension was observed due to a broader net charge spectrum. Furthermore the use of RP-HPLC in the first dimension provides a low carry-over effect, in contrast to SCX. A method for rapid screening of the fractions from the first dimension was developed based on MALDI TOF MS. This method permits selection of the peptide containing fractions for further analysis.

The high-throughput LC-MALDI MS/MS technique based on an off-line coupling of liquid chromatography and mass spectrometry was compared with the common on-line set-up LC-ESI MS/MS. The MALDI approach yielded higher proteome coverage than the ESI approach: 1644 proteins (1208 with more than one peptide) vs. 745 (468). The number of proteins identified by MALDI represents more than twice the identification of the ESI approach but also the measurement time increased three times due to the off-line set-up of the LC-MALDI approach. With respect to the peptidome coverage, both methods shared 26% of all the identified peptides and a small complementarity was observed between the methods. The ESI approach yielded 5% unique peptides but the MALDI approach was capable of delivering 69% of unique peptides. Furthermore, the MALDI TOF/TOF mass spectrometry shows higher flexibility than other mass spectrometric approaches. Because the liquid chromatography is decoupled from the mass spectrometry, column flow rates can be varied and adjusted with the matrix flow rate for optimal spot composition. Besides that, the spotting frequency can be adapted to the sample complexity. MALDI

tolerates better the presence of salts and contaminants than ESI and is more compatible with the presence of ion-pairing reagents like TFA, which do not compromise the ionization process [110].

The novel 2D-RP-HPLC MALDI TOF/TOF approach was compared with a classical approach for separation of proteins: 2D-PAGE followed by in-gel digest analysis by MALDI TOF/TOF. The 2D-PAGE approach is not the most appropriate for a whole proteome study of *C. glutamicum* because it fails to detect low abundant proteins and it can not compete with the high number of identifications delivered by the LC-MS approaches. However, 2D-PAGE is a suitable choice when the object of the study involves visualization of expression changes in enzymes which show moderate to highly concentration and are consequently easily stained and detected [111, 112]. Although the bottom-up approach delivered a higher number of identified proteins a complementary peptide identification of both approaches was observed. Thus, the protein coverage of some of the identified proteins could be increased by the use of both methods.

The three approaches, MALDI-based, ESI-based and 2D-PAGE-based, were also compared at the protein level. Glycolysis, TCA cycle and pentose phosphate pathways were analyzed in detail. The three approaches were able to identify many of the enzymes participating in the main metabolic pathways, the MALDI-based approach was however superior for protein identification. Also the MALDI based approach showed the highest sensitivity and highest dynamic range based on the comparison of the identified proteins according to their CAI. The results of the proteome analysis by 2D-RP-HPLC MALDI TOF/TOF were compared with results of the literature like the MudPIT approach. In conclusion, we believe the approach proposed here is a robust method for the proteome analysis as demonstrated for the bacterium *C. glutamicum*.

7. Outlook

A total of 1208 proteins were identified by more than one peptide by the application of 2D-RP-HPLC MALDI TOF/TOF to the cytosolic proteome of *C. glutamicum* but, the analysis time necessary was 375 hours. Although this argument can be considered a drawback of the method, there are a number of improvements which could shorten the analysis time. With the aid of the fast screening method employing MALDI TOF mass spectrometry it is possible to determine the low peptide content fractions and pool them together thus reducing the total number of fractions to analyze. As an example, pooling the first and last five fractions would save up to 30 LC-MALDI runs thus reducing the analysis time by 90 hours. Also the use of an exclusion list across different runs can improve the measurements [97] because peptides showing intensive signals tend to be analyzed repetitively in different runs. In addition the chance of identifying low abundant peptides would be enhanced. The overall orthogonality can be improved by modifying the gradient of the second dimension for the latest fractions which contain the more hydrophobic peptides. The relative tolerance of the MALDI MS to the presence of surfactants and phosphate additives permits the application of alternative chromatography strategies not considered for the on-line LC-ESI-MS/MS. Alternatively, the use of ionic liquid matrices for proteome analysis can be considered [113].

Assuming that some proteins are not expressed under the cultivation conditions employed, we believe that the total number of identified proteins can be expanded using alternative cultivation conditions. Furthermore, the use of specific extraction protocols for enrichment of membrane proteins and cell wall associated proteins [77] could certainly increase the number of identifications of these proteins. Also other approaches like MudPIT could deliver more complementary results.

The data generated about *C. glutamicum* in this study can be used as reference platform for other proteome studies on mutant strains or modified culture conditions for *C. glutamicum*. Furthermore, the proposed method is fully compatible with quantification strategies like iTRAQ [114] or metabolic labeling [115] for comparative proteomics.

8. List of abbreviations, acronyms and symbols

μA	microampere
μL	microliter
2D	two dimensional
ACN	acetonitrile
CAI	codon adaption index
CHCA	α cyano-4-hydroxycinnamic acid
CID	collision induced dissociation
Da	Dalton
DTT	dithiotreitol
ESI	electrospray ionization
FPR	false positive rate
GRAS	generally regarded as safe
HFBA	heptafluorobutyric acid
HPLC	high performance liquid chromatography
i.d.	internal diameter
IEF	isoelectric focusing
IP	ion pair
iTRAQ	isobaric tag for relative and absolute quantification
L	liter
LC	liquid chromatography
M	molar
MALDI	matrix assisted laser desorption ionization
Mbp	megabase pair
MD	multidimensional
mg	milligram
min	minute
mL	milliliter
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MudPIT	multidimensional protein identification technology
MW	molecular weight
m/z	mass to charge ratio
OD	optical density
ORF	open reading frame

PAGE	polyacrylamide gel electrophoresis
pI	isoelectric point
PMF	peptide mass fingerprint
PS-DVB	polystyrene divinylbenzene
RPC	reversed phase chromatography
SCX	strong cation exchange chromatography
SDS	sodium dodecyl sulfate
TBS	tris buffered saline
TEMED	tetramethylethylenediamine
TFA	trifluoroacetic acid
TMH	trans-membrane helix
TOF	time of flight
UV	ultraviolet
V	volts
v/v	volume/volume
w/v	weight/volume
WCW	wet cell weight
ZIC-HILIC	zwitterionic hydrophilic interaction liquid chromatography

9. Literature

1. Delmotte, N., Lasasosa, M., Tholey, A., Heinzle, E., and Huber, C.G. (2007). Two-dimensional reversed-phase x ion-pair reversed-phase HPLC: an alternative approach to high-resolution peptide separation for shotgun proteome analysis. *J Proteome Res* 6, 4363-4373.
2. Wasinger, V.C., Cordwell, S.J., Cerpa-Poljak, A., Yan, J.X., Gooley, A.A., Wilkins, M.R., Duncan, M.W., Harris, R., Williams, K.L., and Humphery-Smith, I. (1995). Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis* 16, 1090-1094.
3. Cox, J., and Mann, M. (2007). Is proteomics the new genomics? *Cell* 130, 395-398.
4. Wendisch, V.F., Bott, M., and Eikmanns, B.J. (2006). Metabolic engineering of *Escherichia coli* and *Corynebacterium glutamicum* for biotechnological production of organic acids and amino acids. *Curr Opin Microbiol* 9, 268-274.
5. Wendisch, V.F., Bott, M., Kalinowski, J., Oldiges, M., and Wiechert, W. (2006). Emerging *Corynebacterium glutamicum* systems biology. *J Biotechnol* 124, 74-92.
6. Lottspeich, F. (1999). Proteome Analysis: A Pathway to the Functional Analysis of Proteins. *Angew Chem Int Ed Engl* 38, 2476-2492.
7. Cravatt, B.F., Simon, G.M., and Yates, J.R., 3rd (2007). The biological impact of mass-spectrometry-based proteomics. *Nature* 450, 991-1000.
8. Hopf, C., Bantscheff, M., and Drewes, G. (2007). Pathway proteomics and chemical proteomics team up in drug discovery. *Neurodegener Dis* 4, 270-280.
9. Int. Hum. Gen. Seq. Con. (2004). Finishing the euchromatic sequence of the human genome. *Nature* 431, 931-945.
10. Corbin, R.W., Paliy, O., Yang, F., Shabanowitz, J., Platt, M., Lyons, C.E., Jr., Root, K., McAuliffe, J., Jordan, M.I., Kustu, S., Soupene, E., and Hunt, D.F. (2003). Toward a protein profile of *Escherichia coli*: comparison to its transcription profile. *Proc Natl Acad Sci U S A* 100, 9232-9237.
11. Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. *Nature* 425, 737-741.
12. Malmstrom, J., Lee, H., and Aebersold, R. (2007). Advances in proteomic workflows for systems biology. *Curr Opin Biotechnol* 18, 378-384.
13. Nesvizhskii, A.I., Vitek, O., and Aebersold, R. (2007). Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Methods* 4, 787-797.
14. Eisenberg, D., Schwarz, M., Komaromy, M., and Wall, R. (1984). Analysis of Membrane and Surface Protein Sequences with the Hydrophobic Moment Plot. *J. Mol. Biol.* 179, 125-142.
15. Berg J.M., T.J.L., Stryer L. (2002). *Biochemistry*, Fifth Edition Edition (New York).
16. Lehninger A.L., N.D.L., Cox M.M. (1992). *Principios de Bioquímica*, Segunda Edición Edition (Omega).
17. Klose, J. (1975). Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* 26, 231-243.
18. O'Farrell, P.H. (1975). High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250, 4007-4021.
19. de Godoy, L.M., Olsen, J.V., de Souza, G.A., Li, G., Mortensen, P., and Mann, M. (2006). Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biol* 7, R50.

20. Alban, A., David, S.O., Bjorkesten, L., Andersson, C., Sloge, E., Lewis, S., and Currie, I. (2003). A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics* 3, 36-44.
21. Richert, S., Luche, S., Chevallet, M., Van Dorsselaer, A., Leize-Wagner, E., and Rabilloud, T. (2004). About the mechanism of interference of silver staining with peptide mass spectrometry. *Proteomics* 4, 909-916.
22. Winkler, C., Denker, K., Wortelkamp, S., and Sickmann, A. (2007). Silver- and Coomassie-staining protocols: detection limits and compatibility with ESI MS. *Electrophoresis* 28, 2095-2099.
23. Issaq, H., and Veenstra, T. (2008). Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE): advances and perspectives. *Biotechniques* 44, 697-698, 700.
24. Wang, H., and Hanash, S. (2003). Multi-dimensional liquid phase based separations in proteomics. *J Chromatogr B Analyt Technol Biomed Life Sci* 787, 11-18.
25. Candiano, G., Bruschi, M., Musante, L., Santucci, L., Ghiggeri, G.M., Carnemolla, B., Orecchia, P., Zardi, L., and Righetti, P.G. (2004). Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis* 25, 1327-1333.
26. Dugo, P., Cacciola, F., Kumm, T., Dugo, G., and Mondello, L. (2008). Comprehensive multidimensional liquid chromatography: theory and applications. *J Chromatogr A* 1184, 353-368.
27. Maynard, D.M., Masuda, J., Yang, X., Kowalak, J.A., and Markey, S.P. (2004). Characterizing complex peptide mixtures using a multi-dimensional liquid chromatography-mass spectrometry system: *Saccharomyces cerevisiae* as a model system. *J Chromatogr B Analyt Technol Biomed Life Sci* 810, 69-76.
28. Peng, J., Elias, J.E., Thoreen, C.C., Licklider, L.J., and Gygi, S.P. (2003). Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res* 2, 43-50.
29. Schley, C., Altmeyer, M.O., Swart, R., Muller, R., and Huber, C.G. (2006). Proteome analysis of *Myxococcus xanthus* by off-line two-dimensional chromatographic separation using monolithic poly-(styrene-divinylbenzene) columns combined with ion-trap tandem mass spectrometry. *J Proteome Res* 5, 2760-2768.
30. Washburn, M.P., Wolters, D., and Yates, J.R., 3rd (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 19, 242-247.
31. Wolters, D.A., Washburn, M.P., and Yates, J.R., 3rd (2001). An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* 73, 5683-5690.
32. Vollmer, M., Horth, P., and Nagele, E. (2004). Optimization of two-dimensional off-line LC/MS separations to improve resolution of complex proteomic samples. *Anal Chem* 76, 5180-5185.
33. Bartha, A., and Stahlberg, J. (1994). Electrostatic Retention Model of Reversed-Phase Ion-Pair Chromatography. *Journal of Chromatography A* 668, 255-284.
34. Schley, C., Swart, R., and Huber, C.G. (2006). Capillary scale monolithic trap column for desalting and preconcentration of peptides and proteins in one- and two-dimensional separations. *J Chromatogr A* 1136, 210-220.
35. Gilar, M., Daly, A.E., Kele, M., Neue, U.D., and Gebler, J.C. (2004). Implications of column peak capacity on the separation of complex peptide mixtures in single- and

- two-dimensional high-performance liquid chromatography. *J Chromatogr A* 1061, 183-192.
36. Gilar, M., Olivova, P., Daly, A.E., and Gebler, J.C. (2005). Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *J Sep Sci* 28, 1694-1703.
 37. Gilar, M., Olivova, P., Daly, A.E., and Gebler, J.C. (2005). Orthogonality of separation in two-dimensional liquid chromatography. *Anal Chem* 77, 6426-6434.
 38. Boersema, P.J., Divecha, N., Heck, A.J., and Mohammed, S. (2007). Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. *J Proteome Res* 6, 937-946.
 39. Dwivedi, R.C., Spicer, V., Harder, M., Antonovici, M., Ens, W., Standing, K.G., Wilkins, J.A., and Krokhin, O.V. (2008). Practical implementation of 2D HPLC scheme with accurate peptide retention prediction in both dimensions for high-throughput bottom-up proteomics. *Anal Chem* 80, 7036-7042.
 40. Spicer, V., Yamchuk, A., Cortens, J., Sousa, S., Ens, W., Standing, K.G., Wilkins, J.A., and Krokhin, O.V. (2007). Sequence-specific retention calculator. A family of peptide retention time prediction algorithms in reversed-phase HPLC: applicability to various chromatographic conditions and columns. *Anal Chem* 79, 8762-8768.
 41. Premstaller, A., Oberacher, H., Walcher, W., Timperio, A.M., Zolla, L., Chervet, J.P., Cavusoglu, N., van Dorselaer, A., and Huber, C.G. (2001). High-performance liquid chromatography-electrospray ionization mass spectrometry using monolithic capillary columns for proteomic studies. *Anal Chem* 73, 2390-2396.
 42. Toll, H., Wintringer, R., Schweiger-Hufnagel, U., and Huber, C.G. (2005). Comparing monolithic and microparticulate capillary columns for the separation and analysis of peptide mixtures by liquid chromatography-mass spectrometry. *J Sep Sci* 28, 1666-1674.
 43. Walcher, W., Oberacher, H., Troiani, S., Holzl, G., Oefner, P., Zolla, L., and Huber, C.G. (2002). Monolithic capillary columns for liquid chromatography-electrospray ionization mass spectrometry in proteomic and genomic research. *J Chromatogr B Analyt Technol Biomed Life Sci* 782, 111-125.
 44. Oberacher, H., Premstaller, A., and Huber, C.G. (2004). Characterization of some physical and chromatographic properties of monolithic poly(styrene-co-divinylbenzene) columns. *Journal of Chromatography A* 1030, 201-208.
 45. Svec, F., and Huber, C.G. (2006). Monolithic materials: Promises, challenges, achievements. *Anal Chem* 78, 2101-2107.
 46. Karas, M., Bachman, D., Bahr, U., and Hillekamp, F. (1987). Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *Int. J. Mass Spectrom. Ion Processes* 78, 53-68.
 47. Tanaka, T., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T., and Matsuo, T. (1988). Protein and polymer analysis up to m/z 100000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 2, 151-153.
 48. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246, 64-71.
 49. Zhou, M., and Veenstra, T. (2008). Mass spectrometry: m/z 1983-2008. *Biotechniques* 44, 667-668, 670.
 50. Karas, M., Glückmann, M., and Schäfer, J. (2000). Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. *J. Mass Spectrom.* 35, 1-12.
 51. E. de Hoffmann, V.S. (2007). *Mass Spectrometry Principles and Applications*, Third Edition.

52. Medzihradszky, K.F., Campbell, J.M., Baldwin, M.A., Falick, A.M., Juhasz, P., Vestal, M.L., and Burlingame, A.L. (2000). The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal Chem* 72, 552-558.
53. Suckau, D., Resemann, A., Schuerenberg, M., Hufnagel, P., Franzen, J., and Holle, A. (2003). A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. *Anal Bioanal Chem* 376, 952-965.
54. Bienvenut, W.V., Deon, C., Pasquarello, C., Campbell, J.M., Sanchez, J.C., Vestal, M.L., and Hochstrasser, D.F. (2002). Matrix-assisted laser desorption/ionization-tandem mass spectrometry with high resolution and sensitivity for identification and characterization of proteins. *Proteomics* 2, 868-876.
55. Vestal, M.L., and Campbell, J.M. (2005). Tandem time-of-flight mass spectrometry. *Methods Enzymol* 402, 79-108.
56. Nguyen, S., and Fenn, J.B. (2007). Gas-phase ions of solute species from charged droplets of solutions. *Proc Natl Acad Sci U S A* 104, 1111-1117.
57. Liu, H., Sadygov, R.G., and Yates, J.R., 3rd (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* 76, 4193-4201.
58. Ahn, N.G., Shabb, J.B., Old, W.M., and Resing, K.A. (2007). Achieving in-depth proteomics profiling by mass spectrometry. *ACS Chem Biol* 2, 39-52.
59. Chen, H.S., Rejtar, T., Andreev, V., Moskovets, E., and Karger, B.L. (2005). High-speed, high-resolution monolithic capillary LC-MALDI MS using an off-line continuous deposition interface for proteomic analysis. *Anal Chem* 77, 2323-2331.
60. Roepstorff, P., and Fohlman, J. (1984). Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* 11, 601.
61. Biemann, K. (1988). Contributions of mass spectrometry to peptide and protein structure. *Biomed Environ Mass Spectrom* 16, 99-111.
62. Papayannopoulos, I.A. (1995). The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mass Spectrometry Reviews* 14, 49-73.
63. Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551-3567.
64. Eng, J.K., McCormack, A.L., Yates, J.R. III. (1994). An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.*, 976-989.
65. Elias, J.E., Haas, W., Faherty, B.K., and Gygi, S.P. (2005). Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat Methods* 2, 667-675.
66. Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 4, 207-214.
67. Eggeling L., B.M. (2005). Handbook of *Corynebacterium glutamicum*, First Edition.
68. Koffas, M., and Stephanopoulos, G. (2005). Strain improvement by metabolic engineering: lysine production as a case study for systems biology. *Curr Opin Biotechnol* 16, 361-366.
69. Hermann, T. (2003). Industrial production of amino acids by coryneform bacteria. *J Biotechnol* 104, 155-172.
70. Ikeda, M., and Nakagawa, S. (2003). The *Corynebacterium glutamicum* genome: features and impacts on biotechnological processes. *Appl Microbiol Biotechnol* 62, 99-109.

71. Kalinowski, J., Bathe, B., Bartels, D., Bischoff, N., Bott, M., Burkovski, A., Dusch, N., Eggeling, L., Eikmanns, B.J., Gaigalat, L., Goesmann, A., Hartmann, M., Huthmacher, K., Kramer, R., Linke, B., McHardy, A.C., Meyer, F., Mockel, B., Pfefferle, W., Puhler, A., Rey, D.A., Ruckert, C., Rupp, O., Sahm, H., Wendisch, V.F., Wiegrabe, I., and Tauch, A. (2003). The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *J Biotechnol* *104*, 5-25.
72. Hermann, T., Finkemeier, M., Pfefferle, W., Wersch, G., Kramer, R., and Burkovski, A. (2000). Two-dimensional electrophoretic analysis of *Corynebacterium glutamicum* membrane fraction and surface proteins. *Electrophoresis* *21*, 654-659.
73. Hermann, T., Pfefferle, W., Baumann, C., Busker, E., Schaffer, S., Bott, M., Sahm, H., Dusch, N., Kalinowski, J., Puhler, A., Bendt, A.K., Kramer, R., and Burkovski, A. (2001). Proteome analysis of *Corynebacterium glutamicum*. *Electrophoresis* *22*, 1712-1723.
74. Schaffer, S., Weil, B., Nguyen, V.D., Dongmann, G., Gunther, K., Nickolaus, M., Hermann, T., and Bott, M. (2001). A high-resolution reference map for cytoplasmic and membrane-associated proteins of *Corynebacterium glutamicum*. *Electrophoresis* *22*, 4404-4422.
75. Hansmeier, N., Chao, T.C., Puhler, A., Tauch, A., and Kalinowski, J. (2006). The cytosolic, cell surface and extracellular proteomes of the biotechnologically important soil bacterium *Corynebacterium efficiens* YS-314 in comparison to those of *Corynebacterium glutamicum* ATCC 13032. *Proteomics* *6*, 233-250.
76. Polen, T., Schluesener, D., Poetsch, A., Bott, M., and Wendisch, V.F. (2007). Characterization of citrate utilization in *Corynebacterium glutamicum* by transcriptome and proteome analysis. *FEMS Microbiol Lett* *273*, 109-119.
77. Schluesener, D., Fischer, F., Kruij, J., Rogner, M., and Poetsch, A. (2005). Mapping the membrane proteome of *Corynebacterium glutamicum*. *Proteomics* *5*, 1317-1330.
78. Fischer, F., Wolters, D., Rogner, M., and Poetsch, A. (2006). Toward the complete membrane proteome: high coverage of integral membrane proteins through transmembrane peptide detection. *Mol Cell Proteomics* *5*, 444-453.
79. Brand, S., Hahner, S., and Ketterlinus, R. (2005). Protein profiling and identification in complex biological samples using LC-MALDI. In *DrugPlus International*. pp. 6-8.
80. Kiefer, P., Heinzle, E., Zelder, O., and Wittmann, C. (2004). Comparative metabolic flux analysis of lysine-producing *Corynebacterium glutamicum* cultured on glucose or fructose. *Appl Environ Microbiol* *70*, 229-239.
81. Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* *72*, 248-254.
82. Neuhoff, V., Arold, N., Taube, D., and Ehrhardt, W. (1988). 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.
83. Puech, V., Chami, M., Lemassu, A., Laneelle, M.A., Schiffler, B., Gounon, P., Bayan, N., Benz, R., and Daffe, M. (2001). Structure of the cell envelope of corynebacteria: importance of the non-covalently bound lipids in the formation of the cell wall permeability barrier and fracture plane. *Microbiology* *147*, 1365-1382.
84. Hashimoto, K., Kawasaki, H., Akazawa, K., Nakamura, J., Asakura, Y., Kudo, T., Sakuradani, E., Shimizu, S., and Nakamatsu, T. (2006). Changes in composition and content of mycolic acids in glutamate-overproducing *Corynebacterium glutamicum*. *Biosci Biotechnol Biochem* *70*, 22-30.

85. Hirasawa, T., Wachi, M., and Nagai, K. (2001). L-glutamate production by lysozyme-sensitive *Corynebacterium glutamicum* ltsA mutant strains. *BMC Biotechnol* *1*, 9.
86. Hummel, W., Kula M.-R. (1989). Simple method for small-scale disruption for bacteria and yeast. *J. Microbiological Methods* *9*, 201-209.
87. Bodnar, W.M., Blackburn, R.K., Krise, J.M., and Moseley, M.A. (2003). Exploiting the complementary nature of LC/MALDI/MS/MS and LC/ESI/MS/MS for increased proteome coverage. *J Am Soc Mass Spectrom* *14*, 971-979.
88. Hattan, S.J., Marchese, J., Khainovski, N., Martin, S., and Juhasz, P. (2005). Comparative study of [Three] LC-MALDI workflows for the analysis of complex proteomic samples. *J Proteome Res* *4*, 1931-1941.
89. Zhen, Y., Xu, N., Richardson, B., Becklin, R., Savage, J.R., Blake, K., and Peltier, J.M. (2004). Development of an LC-MALDI method for the analysis of protein complexes. *J Am Soc Mass Spectrom* *15*, 803-822.
90. Kratzer, R., Eckerskorn, C., Karas, M., and Lottspeich, F. (1998). Suppression effects in enzymatic peptide ladder sequencing using ultraviolet - matrix assisted laser desorption/ionization - mass spectrometry. *Electrophoresis* *19*, 1910-1919.
91. Wu, W.W., Wang, G., Baek, S.J., and Shen, R.F. (2006). Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF. *J Proteome Res* *5*, 651-658.
92. Brinkrolf, K., Brune, I., and Tauch, A. (2007). The transcriptional regulatory network of the amino acid producer *Corynebacterium glutamicum*. *J Biotechnol* *129*, 191-211.
93. Brune, I., Brinkrolf, K., Kalinowski, J., Puhler, A., and Tauch, A. (2005). The individual and common repertoire of DNA-binding transcriptional regulators of *Corynebacterium glutamicum*, *Corynebacterium efficiens*, *Corynebacterium diphtheriae* and *Corynebacterium jeikeium* deduced from the complete genome sequences. *BMC Genomics* *6*, 86.
94. Kocan, M., Schaffer, S., Ishige, T., Sorger-Herrmann, U., Wendisch, V.F., and Bott, M. (2006). Two-component systems of *Corynebacterium glutamicum*: deletion analysis and involvement of the PhoS-PhoR system in the phosphate starvation response. *J Bacteriol* *188*, 724-732.
95. Krokhin, O.V., Antonovici, M., Ens, W., Wilkins, J.A., and Standing, K.G. (2006). Deamidation of -Asn-Gly- sequences during sample preparation for proteomics: Consequences for MALDI and HPLC-MALDI analysis. *Anal Chem* *78*, 6645-6650.
96. Gygi, S.P., Corthals, G.L., Zhang, Y., Rochon, Y., and Aebersold, R. (2000). Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci U S A* *97*, 9390-9395.
97. Chen, H.S., Rejtar, T., Andreev, V., Moskovets, E., and Karger, B.L. (2005). Enhanced characterization of complex proteomic samples using LC-MALDI MS/MS: exclusion of redundant peptides from MS/MS analysis in replicate runs. *Anal Chem* *77*, 7816-7825.
98. Alpert, A.J., and Andrews, P.C. (1988). Cation-exchange chromatography of peptides on poly(2-sulfoethyl aspartamide)-silica. *J Chromatogr* *443*, 85-96.
99. Taylor, G.K., and Goodlett, D.R. (2005). Rules governing protein identification by mass spectrometry. *Rapid Commun Mass Spectrom* *19*, 3420.
100. Jansen, R., Bussemaker, H.J., and Gerstein, M. (2003). Revisiting the codon adaptation index from a whole-genome perspective: analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models. *Nucleic Acids Res* *31*, 2242-2251.
101. Gouy, M., and Gautier, C. (1982). Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res* *10*, 7055-7074.

102. Wang, R., Prince, J.T., and Marcotte, E.M. (2005). Mass spectrometry of the *M. smegmatis* proteome: protein expression levels correlate with function, operons, and codon bias. *Genome Res* 15, 1118-1126.
103. Sharp, P.M., and Li, W.H. (1987). The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* 15, 1281-1295.
104. Grote, A., Hiller, K., Scheer, M., Munch, R., Nortemann, B., Hempel, D.C., and Jahn, D. (2005). JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res* 33, W526-531.
105. Stapels, M.D., and Barofsky, D.F. (2004). Complementary use of MALDI and ESI for the HPLC-MS/MS analysis of DNA-binding proteins. *Anal Chem* 76, 5423-5430.
106. Yang, Y., Zhang, S., Howe, K., Wilson, D.B., Moser, F., Irwin, D., and Thannhauser, T.W. (2007). A comparison of nLC-ESI-MS/MS and nLC-MALDI-MS/MS for GeLC-based protein identification and iTRAQ-based shotgun quantitative proteomics. *J Biomol Tech* 18, 226-237.
107. Medzihradszky, K.F., Leffler, H., Baldwin, M.A., and Burlingame, A.L. (2001). Protein identification by in-gel digestion, high-performance liquid chromatography, and mass spectrometry: peptide analysis by complementary ionization techniques. *J Am Soc Mass Spectrom* 12, 215-221.
108. Tabb, D.L., Huang, Y., Wysocki, V.H., and Yates, J.R., 3rd (2004). Influence of basic residue content on fragment ion peak intensities in low-energy collision-induced dissociation spectra of peptides. *Anal Chem* 76, 1243-1248.
109. Baumbach, J., Wittkop, T., Rademacher, K., Rahmann, S., Brinkrolf, K., and Tauch, A. (2007). CoryneRegNet 3.0--an interactive systems biology platform for the analysis of gene regulatory networks in corynebacteria and *Escherichia coli*. *J Biotechnol* 129, 279-289.
110. Garcia, M.C., Hogenboom, A.C., Zappey, H., and Irth, H. (2002). Effect of the mobile phase composition on the separation and detection of intact proteins by reversed-phase liquid chromatography-electrospray mass spectrometry. *J Chromatogr A* 957, 187-199.
111. Barreiro, C., Gonzalez-Lavado, E., Brand, S., Tauch, A., and Martin, J.F. (2005). Heat shock proteome analysis of wild-type *Corynebacterium glutamicum* ATCC 13032 and a spontaneous mutant lacking GroEL1, a dispensable chaperone. *J Bacteriol* 187, 884-889.
112. Silberbach, M., Schafer, M., Huser, A.T., Kalinowski, J., Puhler, A., Kramer, R., and Burkovski, A. (2005). Adaptation of *Corynebacterium glutamicum* to ammonium limitation: a global analysis using transcriptome and proteome techniques. *Appl Environ Microbiol* 71, 2391-2402.
113. Zabet-Moghaddam, M., Heinzle, E., Lasaosa, M., and Tholey, A. (2006). Pyridinium-based ionic liquid matrices can improve the identification of proteins by peptide mass-fingerprint analysis with matrix-assisted laser desorption/ionization mass spectrometry. *Anal Bioanal Chem* 384, 215-224.
114. Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D.J. (2004). Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 3, 1154-1169.
115. Wang, Y.K., Ma, Z., Quinn, D.F., and Fu, E.W. (2002). Inverse ¹⁵N-metabolic labeling/mass spectrometry for comparative proteomics and rapid identification of protein markers/targets. *Rapid Commun Mass Spectrom* 16, 1389-1397.

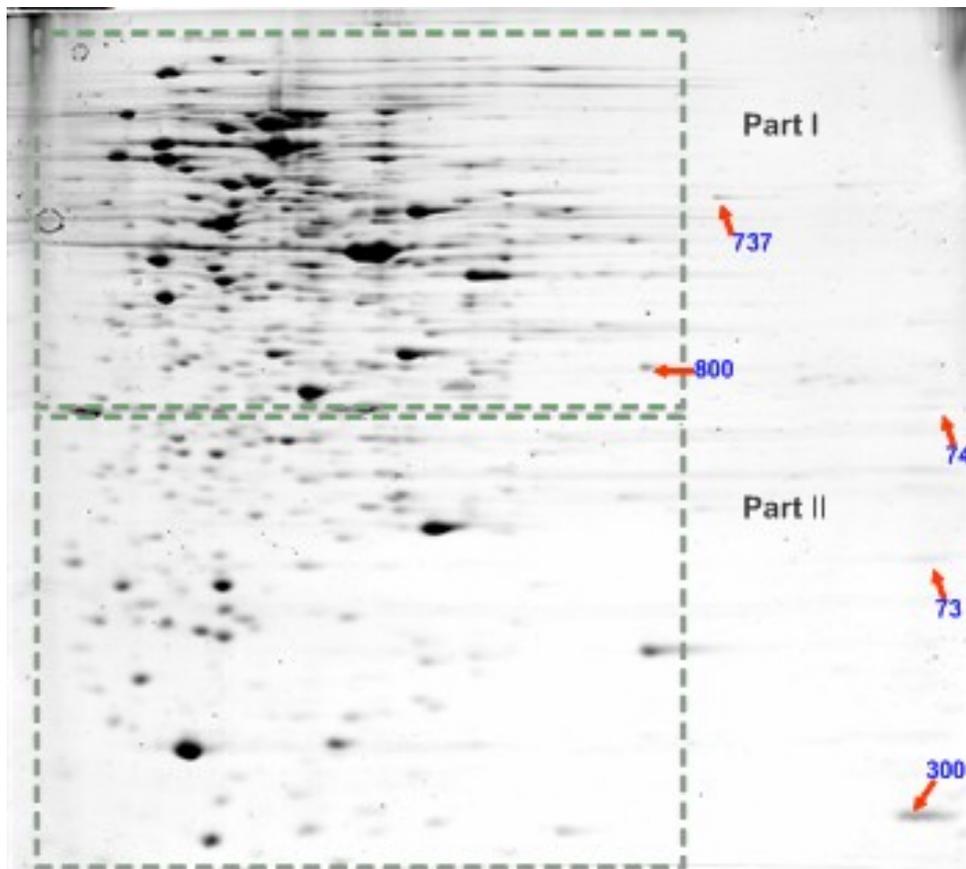
Appendix

Table of Contents

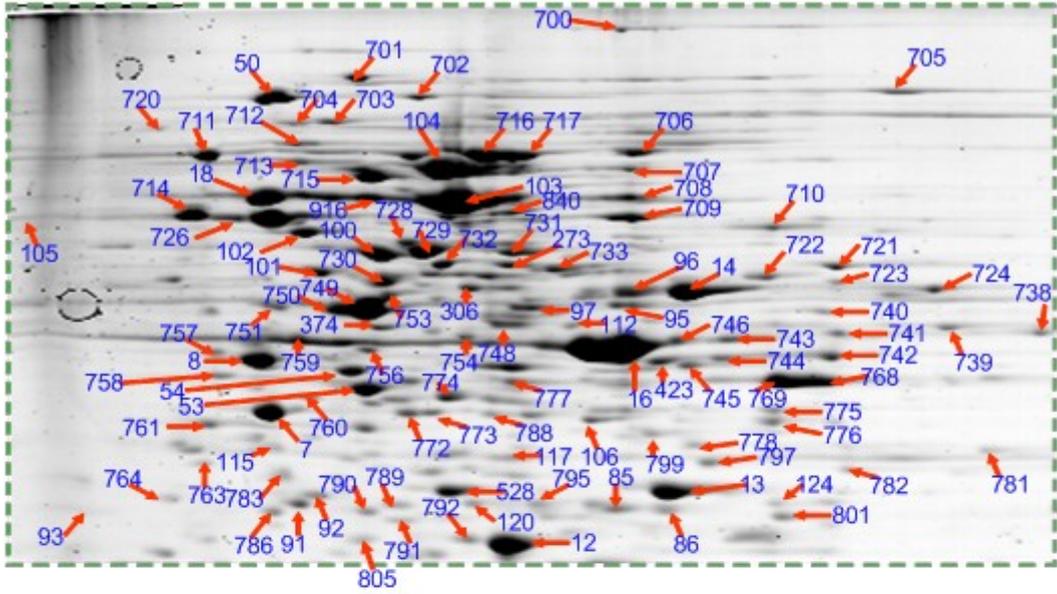
APPENDIX I.....	A 2
APPENDIX II.....	A 8
APPENDIX III.....	A 39
APPENDIX IV.....	A 41
APPENDIX V.....	A 44
APPENDIX VI.....	A 45
APPENDIX VII.....	A 46
APPENDIX VIII.....	A 46
APPENDIX IX.....	A 47
APPENDIX X.....	A 49
APPENDIX XI.....	A 50

Appendix I

Analysis of *C. glutamicum* cytosolic proteins by 2D-PAGE. About 200 μg of cytosolic proteins were loaded on a strip for separation between pH 4 and 7. Staining was carried out by incubation of the gel in a colloidal Coomassie Blue solution. The excised gel plugs which were successfully identified by MALDI TOF/TOF MS are indicated by an arrow and labelled with a number. Spot numbers and the corresponding identified proteins are listed in table 1.



Part I



Part II

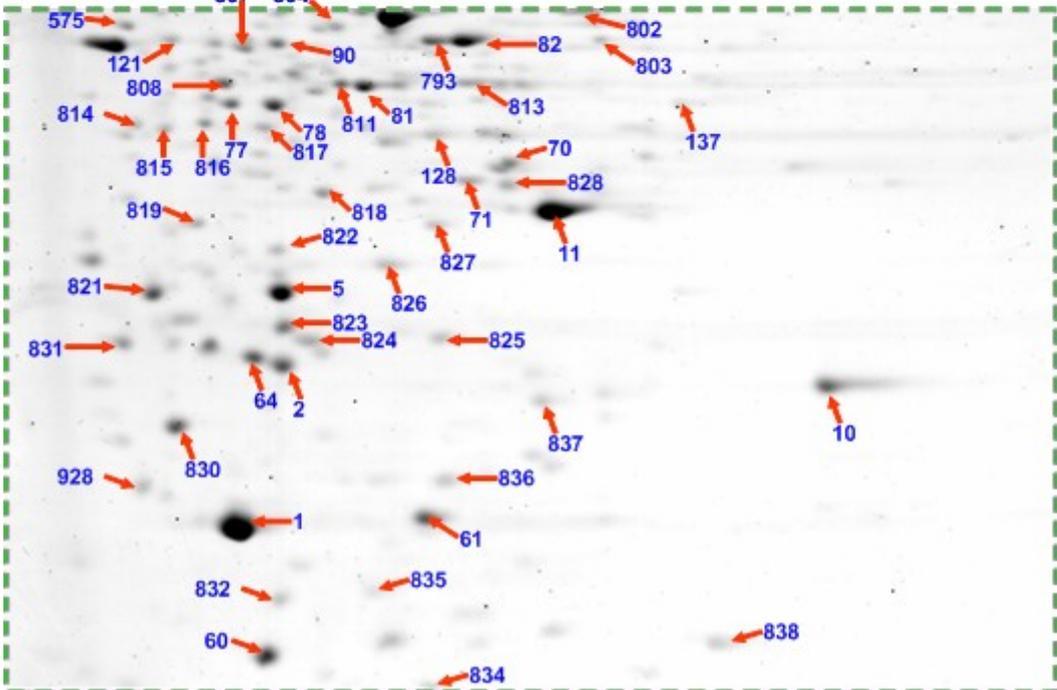


Table 1. List of proteins identified by 2D-PAGE separation followed by MALDI TOF/TOF analysis. Left column shows the number of the spot visualised in the gel.

Spot Nr.	Protein	NCgl no.
1	ribosomal protein L7/L12	NCgl0469
2	FHA-domain-containing protein	NCgl1385
5	peptidyl-prolyl cis-trans isomerase (rotamase) [5.2.1.8]	NCgl0033
7	ketol-acid reductoisomerase [1.1.1.86]	NCgl1224
8	DNA-directed RNA polymerase alpha subunit/40 kD subunit [2.7.7.6]	NCgl0540
8	3-phosphoglycerate kinase [2.7.2.3]	NCgl1525
10	ribosomal protein L10	NCgl0468
11	superoxide dismutase [1.15.1.1]	NCgl2826
12	translation elongation factor Ts	NCgl1949
13	fructose-bisphosphate aldolase [4.1.2.13]	NCgl2673
14	acetyl-CoA hydrolase [2.8.3.-]	NCgl2480
16	elongation factor Tu	NCgl0480
18	70 kDa heat shock chaperonin protein	NCgl2702
50	aconitase A [4.2.1.3]	NCgl1482
53	glutamyl- and glutaminyl-tRNA synthetases [6.1.1.17]	NCgl1244
54	phosphoserine aminotransferase [2.6.1.52]	NCgl0794
60	co-chaperonin GroES	NCgl0572
61	nucleoside diphosphate kinase [2.7.4.6]	NCgl2287
64	inorganic pyrophosphatase [3.6.1.1]	NCgl2607
70	FGAM synthase, glutamine amidotransferase domain [6.3.5.3]	NCgl2500
71	nitroreductase [1.6.99.3]	NCgl0328
73	ribosomal protein L5	NCgl0501
74	ribosomal protein L1	NCgl0460
77	hydroxymethylpyrimidine/phosphomethylpyrimidine kinase	NCgl2973
78	L-2,3-butanediol dehydrogenase	NCgl2582
81	triosephosphate isomerase [5.3.1.1]	NCgl1524
82	ABC-type transporter, ATPase component	NCgl1501
85	sulfate adenyltransferase subunit 2	NCgl2716
86	homoserine kinase [2.7.1.39]	NCgl1137
90	two-component system, response regulator	NCgl2518
91	thioredoxin reductase [1.6.4.5]	NCgl2984
92	phosphoribosylaminoimidazolesuccinocarboxamide (SAICAR) synthase [6.3.2.6]	NCgl2508
92	phosphoribosylaminoimidazolesuccinocarboxamide	NCgl2508
93	ABC-type transporter, periplasmic component	NCgl0610
95	3-isopropylmalate dehydratase large subunit [4.2.1.33]	NCgl1262
95	3-isopropylmalate dehydratase large subunit [4.2.1.33]	NCgl1262
96	fumarase [4.2.1.2]	NCgl0967
97	seryl-tRNA synthetase [6.1.1.11]	NCgl2793
100	phosphoglycerate dehydrogenase [1.1.1.95]	NCgl1235
101	cell division GTPase	NCgl2075
102	phosphoenolpyruvate-protein kinase [2.7.3.9]	NCgl1858
103	isopropylmalate/homocitrate/citramalate synthase [4.1.3.12]	NCgl0245
104	methionine synthase II [2.1.1.14]	NCgl1094
105	ABC-type transporter, periplasmic component	NCgl1915
106	putative inosine-5-monophosphate dehydrogenase	NCgl0579
112	phosphomannomutase	NCgl0558
115	exopolyphosphatase [3.6.1.11]	NCgl0938
117	NAD synthase [6.3.5.1]	NCgl2446

120	hypothetical protein	NCgl1515
121	6-phosphogluconolactonase [3.1.1.31]	NCgl1516
124	shikimate 5-dehydrogenase [1.1.1.25]	NCgl1087
128	uridylate kinase [2.7.4.-]	NCgl1948
273	putative ferredoxin/ferredoxin-NADP reductase [1.18.1.2]	NCgl2719
273	putative ferredoxin/ferredoxin-NADP reductase [1.18.1.2]	NCgl2658
300	ribosomal protein S6	NCgl2881
306	homoserine dehydrogenase [1.1.1.3]	NCgl1136
374	S-adenosylmethionine synthetase [2.5.1.6]	NCgl1541
423	O-acetylhomoserine sulfhydrylase [4.2.99.10]	NCgl0625
528	cysteine synthase [4.2.99.8]	NCgl2473
575	electron transfer flavoprotein alpha-subunit	NCgl1183
700	fatty-acid synthase [2.3.1.85]	NCgl0802
703	phenylalanyl-tRNA synthetase beta subunit [6.1.1.20]	NCgl1336
704	pullulanase	NCgl2026
705	pyruvate carboxylase [6.4.1.1]	NCgl0659
707	predicted membrane GTPase involved in stress	NCgl1053
708	detergent sensitivity rescuer dtsR2	NCgl0677
709	acyl-CoA carboxylase [6.3.4.14]	NCgl0670
710	catalase [1.11.1.6]	NCgl0251
711	phosphoribosylformylglycinamide (FGAM) synthase,synthetase domain [6.3.5.3]	NCgl2499
712	protease II	NCgl2507
714	FKBP-type peptidyl-prolyl cis-trans isomerase	NCgl2329
715	monomeric isocitrate dehydrogenase (NADP+)	NCgl0634
716	elongation factor G	NCgl0478
717	elongation factor G	NCgl0478
721	phosphoribosylaminoimidazolecarboxamide	NCgl0827
722	predicted dehydrogenase [1.1.99.16]	NCgl1926
723	adenylosuccinate synthase [6.3.4.4]	NCgl2669
724	pyruvate kinase [2.7.1.40]	NCgl2008
728	Asp-tRNAAsn/Glu-tRNA Gln amidotransferase B subunit	NCgl1211
729	chaperonin GroEL	NCgl2621
730	F0F1-type ATP synthase beta subunit [3.6.1.34]	NCgl1165
731	F0F1-type ATP synthase alpha subunit [3.6.1.34]	NCgl1163
732	chaperonin GroEL	NCgl0573
735	fumarase [4.2.1.2]	NCgl0967
737	FAD/FMN-containing dehydrogenase	NCgl0187
738	glutamate dehydrogenase/leucine dehydrogenase	NCgl1999
738	putative Asp-tRNAAsn/Glu-tRNA Gln amidotransferaseA subunit [6.3.5.-]	NCgl1199
739	glycine hydroxymethyltransferase [2.1.2.1]	NCgl0954
740	histidyl-tRNA synthetase [6.1.1.21]	NCgl1585
740	adenylosuccinate lyase [4.3.2.2]	NCgl2509
741	D-2-hydroxyisocaproate dehydrogenase	NCgl2528
742	citrate synthase [4.1.3.7]	NCgl0795
743	predicted iron-regulated ABC-type transporter SufB	NCgl1502
744	UDP-N-acetylglucosamine enolpyruvyl transferase [2.5.1.7]	NCgl0345
745	putative fructose-1,6-bisphosphatase/sedoheptulose1,7-bisphosphatase	NCgl0976
746	predicted iron-regulated ABC-type transporter SufB	NCgl1502
748 mix	cell division initiation protein	NCgl2070
748 mix	phosphotransacetylase [2.3.1.8]	NCgl2657
749	enolase [4.2.1.11]	NCgl0935
750	aspartokinase [2.7.2.4]	NCgl0247
751	phosphomannomutase [5.4.2.8]	NCgl0714
753	6-phosphogluconate dehydrogenase, family 1 [1.1.1.44]	NCgl1396

754	gamma-glutamyl phosphate reductase [1.2.1.41]	NCgl2272
756	diaminopimelate decarboxylase [4.1.1.20]	NCgl1133
757	Xaa-Pro aminopeptidase [3.4.13.9]	NCgl1430
758	predicted GTPase	NCgl0988
759	phosphoserine phosphatase [3.1.3.3]	NCgl2436
759	elongation factor Tu	NCgl0480
760	D-alanine-D-alanine ligase [6.3.2.4]	NCgl1267
761	transaldolase [2.2.1.2]	NCgl1513
763	predicted sugar phosphatase of the HAD superfamily	NCgl1355
764	archaeal fructose-1,6-bisphosphatase	NCgl0765
53	myo-inositol-1-phosphate synthase	NCgl2894
768 & 769	glyceraldehyde-3-phosphate dehydrogenase [1.2.1.12]	NCgl1526
772	phosphoribosylaminoimidazol (AIR) synthetase [6.3.3.1]	NCgl2494
773	PLP-dependent aminotransferase [2.6.1.-]	NCgl2227
774	class III Zn-dependent alcohol dehydrogenase [1.2.1.-]	NCgl0313
775	phenylalanyl-tRNA synthetase alpha subunit [6.1.1.20]	NCgl1335
776	Zn-dependent alcohol dehydrogenase [1.1.1.1]	NCgl2709
777	carbamoylphosphate synthase small subunit [6.3.5.5]	NCgl1548
778	cystathionine gamma-synthase [4.2.99.9]	NCgl2360
781	6-phosphofructokinase [2.7.1.11]	NCgl1202
782	putative Zn-NADPH:quinone dehydrogenase [1.1.1.1]	NCgl2449
783	hypothetical protein	NCgl1996
786	thiosulfate sulfurtransferase [2.8.1.1]	NCgl0671
788	histidinol-phosphate aminotransferase/tyrosine aminotransferase [2.6.1.9]	NCgl2020
789	dihydrodipicolinate reductase [1.3.1.26]	NCgl1898
790	2-keto-4-pentenoate hydratase/2-oxohepta-3-ene-1,7-dioic acid hydratase [5.3.3.-]	NCgl1241
791	predicted oxidoreductase [1.1.1.91]	NCgl0099
792	Zn-dependent hydrolase	NCgl0285
793	phosphoglycerate mutase 1 [5.4.2.1]	NCgl0390
795	pyridoxine biosynthesis enzyme	NCgl0754
797	L-lactate dehydrogenase [1.1.1.27]	NCgl2810
799 mix	UDP-glucose 4-epimerase [5.1.3.2]	NCgl1846
799 mix	acetylglutamate semialdehyde dehydrogenase [1.2.1.38]	NCgl1340
800	5,10-methylene-tetrahydrofolate dehydrogenase [1.5.1.5]	NCgl0620
801	phage shock protein A (IM30)	NCgl1886
802 mix	nucleoside-diphosphate-sugar epimerase	NCgl0317
802 mix	ATP phosphoribosyltransferase [2.4.2.17]	NCgl1447
803	trehalose-6-phosphatase [3.1.3.12]	NCgl2537
804	2-hydroxyhepta-2,4-diene-1,7-dioatesomerase	NCgl2919
805	oligoribonuclease (3->5 exoribonuclease) [3.1.-.-]	NCgl2386
807	hypothetical protein	NCgl2339
808	hypothetical protein	NCgl2246
811	predicted hydrolase of the HAD superfamily	NCgl2530
813	transcriptional regulator [2.7.1.63]	NCgl1835
814	uracil phosphoribosyltransferase [2.4.2.9]	NCgl0654
815	ProFAR isomerase	NCgl2015
816	pentose-5-phosphate-3-epimerase [5.1.3.1]	NCgl1536
817	Mn-dependent transcriptional regulator	NCgl1845
818	ATP-dependent Clp protease proteolytic subunit 2 [3.4.21.92]	NCgl2327
819	transcriptional regulator	NCgl2571
821	peroxiredoxin [1.11.1.-]	NCgl1041
822	predicted acyltransferase	NCgl2659
823	hypothetical protein	NCgl1123
824	adenylate kinase [2.7.4.3]	NCgl0533

825	predicted thioesterase	NCgl2365
826	ATP-dependent Clp protease proteolytic subunit 1 [3.4.21.92]	NCgl2328
827	3-isopropylmalate dehydratase small subunit [4.2.1.33]	NCgl1263
828	ribosome recycling factor	NCgl1947
830	universal stress protein UspA	NCgl1316
831	hypothetical protein	NCgl0385
832	hypothetical protein	NCgl1676
834	phosphoribosylformylglycinamide (FGAM) synthase,PurS component	NCgl2501
835	thioredoxin	NCgl2985
836	transcriptional regulator	NCgl2886
837	ribose 5-phosphate isomerase RpiB [5.3.1.26]	NCgl2337
838	ribosomal protein S6	NCgl2881
840	thiamine pyrophosphate-requiring enzyme [4.1.3.18]	NCgl1222
850	Zn-dependent oligopeptidase [3.4.15.5]	NCgl2219
915	4-alpha-glucanotransferase [2.4.1.25]	NCgl2217
916	putative acyl-CoA synthetase	NCgl2774
928	ferritin-like protein	NCgl2439

Appendix II

Table 2. List of proteins identified by LC-MALDI TOF/TOF mass spectrometry and database search, after separation by two-dimensional reversed phase chromatography at high and low pH. Accession numbers were obtained from the Comprehensive Microbial Resources, Institute for Genomic Research (www.tigr.org). Matched peptides include redundant hits.

No.	Accesión number	Matched peptides	Seq. Cov. (%)	Protein description
1	NCgl0001	11	21.9	ATPase involved in DNA replication initiation
2	NCgl0002	32	50.5	DNA polymerase III beta subunit [2.7.7.7]
3	NCgl0005	36	37.7	DNA gyrase (topoisomerase II) B subunit [5.99.1.3]
4	NCgl0006	1	3.7	predicted hydrolase or acyltransferase
5	NCgl0008	5	10.2	hypothetical protein
6	NCgl0012	48	30.8	DNA gyrase (topoisomerase II) A subunit [5.99.1.3]
7	NCgl0014	2	2.9	uncharacterized membrane protein
8	NCgl0024	3	3.6	predicted transcriptional regulator [3.6.1.-]
9	NCgl0029	2	11	ABC-type transporter: periplasmic component
10	NCgl0030	1	3.2	ABC-type transporter: permease component
11	NCgl0031	5	14.6	ABC-type transporter: ATPase component
12	NCgl0032	20	19.3	hypothetical protein
13	NCgl0033	33	64.2	peptidyl-prolyl cis-trans isomerase (rotamase) [5.2.1.8]
14	NCgl0040	7	12.1	serine/threonine protein kinase [2.7.1.-]
15	NCgl0042	4	6.5	cell division protein FtsI
16	NCgl0044	1	2.7	protein serine/threonine phosphatase [3.1.3.16]
17	NCgl0045	4	20.1	FHA-domain-containing protein
18	NCgl0046	4	11.8	hypothetical FHA-domain-containing protein
19	NCgl0049	12	17.1	NAD-dependent aldehyde dehydrogenase [1.2.1.16]
20	NCgl0055	1	12.9	hypothetical protein
21	NCgl0056	1	4.2	hypothetical protein
22	NCgl0058	1	3.8	hypothetical membrane protein
23	NCgl0060	1	3.8	Mg ²⁺ and Co ²⁺ transporter protein
24	NCgl0061	2	6.7	4-oxalocrotonate tautomerase-like protein
25	NCgl0064	1	3.1	ATPase related to phosphate starvation-inducible protein PhoH
26	NCgl0069	3	17.1	phosphoglycerate dehydrogenase or related dehydrogenase
27	NCgl0070	6	30.8	hypothetical protein
28	NCgl0077	7	17.2	hypothetical protein
29	NCgl0081	1	5.1	hypothetical protein
30	NCgl0083	3	9	urea amidohydrolase (urease) gamma subunit [3.5.1.5]
31	NCgl0084	3	19.1	urea amidohydrolase (urease) beta subunit [3.5.1.5]
32	NCgl0085	9	7.5	urea amidohydrolase (urease) alpha subunit [3.5.1.5]
33	NCgl0086	6	24.8	urease accessory protein UreE [3.5.1.5]
34	NCgl0092	13	13.1	hypothetical protein
35	NCgl0093	6	13	molecular chaperone: HSP90 family
36	NCgl0094	11	22	nucleoside phosphorylase [3.2.2.4]
37	NCgl0098	21	14.3	proline dehydrogenase [1.5.1.12]
38	NCgl0099	9	14	predicted oxidoreductase [1.1.1.91]
39	NCgl0101	13	19	metal-dependent enzyme

40	NCgl0104	4	8	acyl-CoA synthetase (AMP-forming)/AMP-acid ligase II
41	NCgl0105	7	28.1	deoR family transcriptional regulator of sugar metabolism
42	NCgl0106	4	16.9	lactoylglutathione lyase or related lyase
43	NCgl0110	4	16	deoR family transcriptional regulator of sugar metabolism
44	NCgl0111	1	3.5	sugar (pentulose and hexulose) kinase [2.7.1.17]
45	NCgl0112	9	19.4	panthothenate synthetase [6.3.2.1]
46	NCgl0113	6	18.6	hydroxymethyltransferase [2.1.2.11]
47	NCgl0118	10	29.4	predicted hydrolase of the HAD superfamily
48	NCgl0119	4	32.8	carbonic anhydrase/acetyltransferase [4.2.1.1]
49	NCgl0120	13	22.7	transcriptional regulator
50	NCgl0124	9	20.9	hypothetical membrane protein
51	NCgl0127	2	3.2	hypothetical membrane protein
52	NCgl0128	7	11.8	ankyrin repeat protein
53	NCgl0133	1	15.4	aspartate 1-decarboxylase [4.1.1.11]
54	NCgl0134	39	34.2	hypothetical protein
55	NCgl0144	1	14.4	hypothetical protein
56	NCgl0145	8	18.3	lactoylglutathione lyase-like protein
57	NCgl0147	1	3.4	hypothetical protein
58	NCgl0148	2	2.6	hypothetical protein
59	NCgl0151	30	29.1	predicted metalloendopeptidase
60	NCgl0154	5	20.6	transcriptional regulator
61	NCgl0155	5	24.2	sugar kinases: ribokinase family [2.7.1.4]
62	NCgl0156	4	10.3	hypothetical protein
63	NCgl0157	10	17.1	NAD-dependent aldehyde dehydrogenase [1.2.1.27]
64	NCgl0158	11	29.4	hypothetical protein
65	NCgl0159	6	4.1	thiamine pyrophosphate-requiring enzyme
66	NCgl0160	17	34.6	hypothetical protein
67	NCgl0162	4	21.7	sugar phosphate isomerase/epimerase
68	NCgl0163	1	3.3	permease of the major facilitator superfamily
69	NCgl0164	2	6.2	predicted dehydrogenase
70	NCgl0168	4	14.4	predicted dehydrogenase
71	NCgl0169	4	10.1	sugar phosphate isomerase/epimerase
72	NCgl0171	8	65.7	cold shock protein
73	NCgl0179	2	5.2	transposase
74	NCgl0180	13	39.3	PAS/PAC domain containing protein [2.7.3.-]
75	NCgl0181	6	4	glutamine 2-oxoglutarate aminotransferase large subunit [1.4.1.13]
76	NCgl0183	11	18.8	hypothetical protein
77	NCgl0184	5	5.1	putative arabinosyl transferase
78	NCgl0185	1	2.2	hypothetical membrane protein
79	NCgl0186	10	26.9	putative dehydrogenase
80	NCgl0187	26	50.4	FAD/FMN-containing dehydrogenase
81	NCgl0188	3	42.7	hypothetical protein
82	NCgl0189	1	13.5	hypothetical membrane protein
83	NCgl0190	4	16.6	hypothetical protein
84	NCgl0191	2	8.4	hypothetical protein
85	NCgl0194	4	19.4	hypothetical protein
86	NCgl0195	4	12.9	predicted glycosyltransferases
87	NCgl0196	2	14.1	hypothetical protein
88	NCgl0199	11	23.4	selenocysteine lyase
89	NCgl0200	3	11.3	NADPH:quinone reductase or related Zn-dependent oxidoreductases
90	NCgl0205	1	6.2	dinucleotide-utilizing enzyme
91	NCgl0206	1	9.9	molybdopterin converting factor: large subunit
92	NCgl0207	1	6.5	molybdopterin biosynthesis enzyme
93	NCgl0209	3	15.5	molybdopterin biosynthesis enzyme

94	NCgl0211	5	16	ABC-type transporter: periplasmic component
95	NCgl0212	1	11.8	hypothetical protein
96	NCgl0213	5	6.9	ABC-type transporter: ATPase component
97	NCgl0215	16	33.4	histidinol-phosphate aminotransferase/tyrosine aminotransferase [2.6.1.9]
98	NCgl0216	1	3.6	hypothetical protein
99	NCgl0221	2	4.6	Mg/Co/Ni transporter MgtE
100	NCgl0223	10	20.9	prephenate dehydrogenase
101	NCgl0224	1	7.9	hypothetical protein
102	NCgl0226	7	28.4	hypothetical protein
103	NCgl0229	7	19.9	queuine/archaeosine tRNA-ribosyltransferase [2.4.2.29]
104	NCgl0232	1	1.9	ABC-type multidrug/protein/lipid transport system: ATPase component
105	NCgl0233	2	4.8	glutamyl-/ glutaminyl-tRNA synthetase [6.1.1.17]
106	NCgl0234	2	5.2	hypothetical dioxygenase protein
107	NCgl0235	1	2.4	putative transposase
108	NCgl0236	1	7.9	hypothetical protein
109	NCgl0237	55	41.1	aspartate transaminase [2.6.1.1]
110	NCgl0238	3	15.8	hypothetical protein
111	NCgl0239	15	14.2	DNA polymerase III: gamma/tau subunits
112	NCgl0240	12	30.1	hypothetical protein
113	NCgl0241	5	24.3	recombinational DNA repair protein RecR
114	NCgl0242	6	22.8	predicted glutamine amidotransferase
115	NCgl0243	16	19.6	UDP-N-acetylmuramyl tripeptide synthase
116	NCgl0244	1	5.3	DNA polymerase III epsilon subunit
117	NCgl0245	366	63	isopropylmalate/homocitrate/citramalate synthase [4.1.3.12]
118	NCgl0247	24	27.8	aspartokinase [2.7.2.4]
119	NCgl0248	27	48.8	aspartate-semialdehyde dehydrogenase [1.2.1.11]
120	NCgl0249	1	3.1	hypothetical membrane protein
121	NCgl0251	61	52.9	catalase [1.11.1.6]
122	NCgl0253	2	6.5	transcriptional regulator
123	NCgl0259	2	10.1	protein-tyrosine-phosphatase
124	NCgl0262	1	12.6	hypothetical membrane protein
125	NCgl0272	6	17.5	predicted phosphohydrolase [3.1.-.-]
126	NCgl0273	4	16.6	hypothetical protein
127	NCgl0274	4	8.8	membrane carboxypeptidase
128	NCgl0276	4	25.9	hypothetical protein
129	NCgl0277	4	21.6	putative translation initiation inhibitor
130	NCgl0280	2	6.4	transcriptional regulator
131	NCgl0281	1	7.1	dehydrogenase [1.1.1.100]
132	NCgl0284	1	5.1	predicted acyl dehydratase
133	NCgl0285	10	28.1	Zn-dependent hydrolase
134	NCgl0286	35	64.8	cAMP-binding domain containing protein
135	NCgl0290	2	10.5	NTP pyrophosphohydrolase
136	NCgl0292	2	8.5	predicted hydrolase or acyltransferase [3.3.2.3]
137	NCgl0293	3	15.3	hypothetical membrane protein
138	NCgl0296	1	4.7	hypothetical membrane protein
139	NCgl0297	1	5.6	predicted ATPase involved in pili and flagella biosynthesis
140	NCgl0303	7	65.7	cold shock protein
141	NCgl0304	48	25.8	topoisomerase IA [5.99.1.2]
142	NCgl0305	2	5.1	hypothetical protein
143	NCgl0307	2	4.5	ATPase involved in DNA replication [2.7.7.7]
144	NCgl0309	2	8.6	putative adenylate kinase
145	NCgl0313	21	41.8	class III Zn-dependent alcohol dehydrogenase [1.2.1.-]
146	NCgl0314	3	25.6	Zn-dependent hydrolase or glyoxylase
147	NCgl0315	3	5.4	dehydrogenase [1.1.1.100]

148	NCgl0317	21	45.7	nucleoside-diphosphate-sugar epimerase
149	NCgl0319	4	16.1	glycosyltransferase involved in cell wall biogenesis
150	NCgl0321	2	7.5	predicted glycosyltransferase
151	NCgl0324	13	31.7	Zn-dependent alcohol dehydrogenase [1.1.1.2]
152	NCgl0325	1	5.6	dTDP-glucose pyrophosphorylase [2.7.7.24]
153	NCgl0326	11	13.7	dTDP-4-dehydrorhamnose 3:5-epimerase or related enzyme
154	NCgl0327	22	40.7	dTDP-D-glucose 4:6-dehydratase [4.2.1.46]
155	NCgl0328	13	38.9	nitroreductase [1.6.99.3]
156	NCgl0329	3	10.5	periplasmic component of ABC-type Fe ³⁺ -siderophore transport system
157	NCgl0332	11	16.3	aminopeptidase N
158	NCgl0333	17	17.4	serine protease [3.4.21.26]
159	NCgl0336	6	14	hypothetical esterase
160	NCgl0337	10	15.5	putative ATPase involved in chromosome partitioning
161	NCgl0338	10	39.3	protein-tyrosine-phosphatase [3.1.3.48]
162	NCgl0339	1	3.6	hypothetical protein
163	NCgl0340	7	14.8	predicted nucleoside-diphosphate sugar epimerase
164	NCgl0341	29	39	predicted pyridoxal phosphate-dependent enzyme
165	NCgl0342	3	8.3	sugar transferase involved in lipopolysaccharide synthesis
166	NCgl0343	5	6.1	predicted glycosyltransferase
167	NCgl0344	2	3.8	O-antigen and teichoic acid membrane export protein
168	NCgl0345	31	40	UDP-N-acetylglucosamine enolpyruvyl transferase [2.5.1.7]
169	NCgl0346	6	11	UDP-N-acetylmuramate dehydrogenase [1.1.1.158]
170	NCgl0347	13	24.9	cell wall biogenesis glycosyltransferase
171	NCgl0350	1	4.2	predicted acyltransferase
172	NCgl0351	22	37	predicted UDP-glucose 6-dehydrogenase [1.1.1.22]
173	NCgl0353	16	23.7	cell wall biogenesis glycosyltransferase
174	NCgl0354	9	27.4	acetyltransferase
175	NCgl0355	105	53.9	dihydrolipoamide dehydrogenase/glutathione oxidoreductase-like protein [1.8.1.1]
176	NCgl0356	11	24.3	UDP-glucose pyrophosphorylase [2.7.7.9]
177	NCgl0357	7	25.8	hypothetical protein
178	NCgl0358	22	39.2	predicted transcriptional regulator
179	NCgl0360	20	27	succinate dehydrogenase/fumarate reductase: flavoprotein subunit [1.3.99.1]
180	NCgl0361	10	14.1	succinate dehydrogenase/fumarate reductase Fe-S protein
181	NCgl0364	8	16	hypothetical protein
182	NCgl0365	1	2.6	uncharacterized membrane protein
183	NCgl0366	6	35	hypothetical protein
184	NCgl0368	1	8.9	transcriptional regulator
185	NCgl0371	24	49.3	formyltetrahydrofolate hydrolase [3.5.1.10]
186	NCgl0372	7	23.2	deoxyribose-phosphate aldolase [4.1.2.4]
187	NCgl0374	4	31	hypothetical protein
188	NCgl0375	4	2.3	cation transport ATPase [3.6.1.-]
189	NCgl0385	3	10.8	hypothetical protein
190	NCgl0386	5	10.3	UDP-N-acetylmuramate dehydrogenase [1.1.1.158]
191	NCgl0387	3	13.2	hypothetical protein
192	NCgl0388	27	26.6	acyl-CoA synthetase [6.2.1.3]
193	NCgl0389	7	14.4	predicted glycosyltransferase
194	NCgl0390	41	44.8	phosphoglycerate mutase 1 [5.4.2.1]
195	NCgl0392	8	36.2	two-component system: response regulators
196	NCgl0396	17	37.4	exopolyphosphatase
197	NCgl0398	7	14.3	pyrroline-5-carboxylate reductase [1.5.1.2]
198	NCgl0399	2	39.7	hypothetical protein
199	NCgl0400	3	4.5	phosphoserine phosphatase [3.1.3.3]
200	NCgl0402	4	8.6	glutamyl-tRNA reductase [1.2.1.-]
201	NCgl0403	9	12.9	porphobilinogen deaminase [4.3.1.8]

202	NCgl0406	1	3.2	permease of the major facilitator superfamily
203	NCgl0407	6	6.3	sugar phosphate isomerase/epimerase [4.2.1.-]
204	NCgl0413	2	4.2	ABC-type transporter periplasmic component
205	NCgl0414	8	10.5	uroporphyrinogen-III synthase/methylase [2.1.1.107]
206	NCgl0415	2	9.2	hypothetical protein
207	NCgl0416	14	38.1	delta-aminolevulinic acid dehydratase-like protein [4.2.1.24]
208	NCgl0420	16	18.2	uroporphyrinogen-III decarboxylase [4.1.1.37]
209	NCgl0421	8	13.2	protoporphyrinogen oxidase [1.3.3.4]
210	NCgl0422	11	22.4	glutamate-1-semialdehyde aminotransferase [5.4.3.8]
211	NCgl0423	3	16.3	phosphoglycerate mutase/fructose-2:6-bisphosphatase [5.4.2.1]
212	NCgl0428	3	6	hypothetical protein
213	NCgl0431	6	42	hypothetical protein
214	NCgl0433	1	3.7	1:4-dihydroxy-2-naphthoate octaprenyltransferase [2.5.1.-]
215	NCgl0434	3	2.8	hypothetical protein
216	NCgl0439	4	8.2	transcriptional regulator
217	NCgl0440	2	18.1	hypothetical protein
218	NCgl0443	1	8.7	hypothetical protein
219	NCgl0445	1	2	phosphate/sulphate permease
220	NCgl0446	8	15.7	dihydroxynaphthoic acid synthase [4.1.3.36]
221	NCgl0448	4	12.8	peptidase E
222	NCgl0449	8	16.6	O-succinylbenzoate synthase or related enzyme
223	NCgl0450	6	9	2-succinyl-6-hydroxy-2:4-cyclohexadiene-1-carboxylate synthase [4.1.1.71]
224	NCgl0452	1	3.4	predicted group 1 glycosyltransferase
225	NCgl0454	6	22.2	methylase [2.1.1.-]
226	NCgl0455	2	6.4	oxidoreductase
227	NCgl0456	4	17.4	geranylgeranyl pyrophosphate synthase [2.5.1.30]
228	NCgl0458	22	23.9	transcription antiterminator
229	NCgl0459	3	9.7	ribosomal protein L11
230	NCgl0460	64	49.2	ribosomal protein L1
231	NCgl0466	3	4	hypothetical protein
232	NCgl0467	4	12.9	hypothetical protein
233	NCgl0468	44	62.6	ribosomal protein L10
234	NCgl0469	35	44.5	ribosomal protein L7/L12
235	NCgl0471	87	37.4	DNA-directed RNA polymerase beta subunit/140 kD subunit [2.7.7.6]
236	NCgl0472	137	42.5	DNA-directed RNA polymerase beta subunit/160 kD subunit [2.7.7.6]
237	NCgl0475	1	5.6	hypothetical protein
238	NCgl0476	14	28.7	ribosomal protein S12
239	NCgl0477	50	54.8	ribosomal protein S7
240	NCgl0478	196	50.8	elongation factor G
241	NCgl0479	1	2.2	hypothetical protein
242	NCgl0480	300	63.6	elongation factor Tu
243	NCgl0481	1	6.9	hypothetical protein
244	NCgl0482	5	17.4	ABC-type transporter: ATPase component
245	NCgl0486	22	41.6	ribosomal protein S10
246	NCgl0487	92	61.5	ribosomal protein L3
247	NCgl0488	60	61	ribosomal protein L4
248	NCgl0489	15	49.5	ribosomal protein L23
249	NCgl0490	130	51.8	ribosomal protein L2
250	NCgl0491	37	48.9	ribosomal protein S19
251	NCgl0492	32	67.5	ribosomal protein L22
252	NCgl0493	32	35.9	ribosomal protein S3
253	NCgl0494	51	47.1	ribosomal protein L16/L10E
254	NCgl0495	19	76.3	ribosomal protein L29
255	NCgl0496	18	44.6	ribosomal protein S17

256	NCgl0498	1	9.5	hypothetical protein
257	NCgl0499	17	26.2	ribosomal protein L14
258	NCgl0500	9	20.2	ribosomal protein L24
259	NCgl0501	39	45.5	ribosomal protein L5
260	NCgl0503	4	11.2	aldo/keto reductase [1.1.1.-]
261	NCgl0505	2	12.1	formate dehydrogenase accessory protein [1.2.1.2]
262	NCgl0507	19	22.8	putative formate dehydrogenase [1.2.1.2]
263	NCgl0513	6	12.4	hypothetical protein
264	NCgl0514	2	5	predicted acetyltransferase
265	NCgl0515	12	29.5	ribosomal protein S8
266	NCgl0516	31	39.3	ribosomal protein L6
267	NCgl0517	9	27.6	ribosomal protein L18
268	NCgl0518	12	32.2	ribosomal protein S5
269	NCgl0519	6	19.7	ribosomal protein L30/L7E
270	NCgl0520	25	44.6	ribosomal protein L15
271	NCgl0523	1	2.4	NAD-dependent aldehyde dehydrogenase
272	NCgl0525	1	3.8	putative reductase
273	NCgl0527	1	6.5	dehydrogenase
274	NCgl0531	4	4.1	transcriptional regulator
275	NCgl0532	2	7.3	preprotein translocase subunit SecY
276	NCgl0533	26	41.4	adenylate kinase [2.7.4.3]
277	NCgl0534	7	23.1	methionine aminopeptidase [3.4.11.18]
278	NCgl0535	1	7.2	hypothetical protein
279	NCgl0536	14	44.4	translation initiation factor IF-1
280	NCgl0537	27	45.1	ribosomal protein S13
281	NCgl0538	5	11.2	ribosomal protein S11
282	NCgl0539	55	41.8	ribosomal protein S4
283	NCgl0540	69	51.2	DNA-directed RNA polymerase alpha subunit/40 kD subunit [2.7.7.6]
284	NCgl0541	46	66.3	ribosomal protein L17
285	NCgl0543	3	2	hypothetical membrane protein
286	NCgl0548	5	14.7	putative cyclopropane-fatty-acyl-phospholipid synthase [2.1.1.79]
287	NCgl0551	1	2.2	hypothetical membrane protein
288	NCgl0556	49	58.5	ribosomal protein L13
289	NCgl0557	27	23.1	ribosomal protein S9
290	NCgl0558	26	30.2	phosphomannomutase
291	NCgl0561	2	23.1	hypothetical protein
292	NCgl0562	24	34.4	hypothetical protein
293	NCgl0563	3	6.1	alanine racemase [5.1.1.1]
294	NCgl0564	10	29.1	hypothetical protein
295	NCgl0565	1	2.1	putative membrane protein
296	NCgl0567	1	5.3	conserved hypothetical protein
297	NCgl0568	2	13.6	acetyltransferase [2.3.1.128]
298	NCgl0569	12	31.7	metal-dependent protease [3.4.24.57]
299	NCgl0570	6	11.8	hypothetical protein
300	NCgl0572	33	84.8	co-chaperonin GroES
301	NCgl0573	40	51.3	chaperonin GroEL
302	NCgl0575	2	5.9	DNA-directed RNA polymerase specialized sigma subunit
303	NCgl0576	1	3.8	hypothetical protein
304	NCgl0578	50	45.7	inosine monophosphate dehydrogenase [1.1.1.205]
305	NCgl0579	32	44.1	putative inosine-5-monophosphate dehydrogenase
306	NCgl0582	39	44.7	GMP synthase [6.3.5.2]
307	NCgl0583	4	21.1	hypothetical protein
308	NCgl0585	1	3.4	two-component system sensor kinase
309	NCgl0586	3	13.5	two-component system response regulator

310	NCgl0591	2	3.4	hypothetical protein
311	NCgl0592	4	9.4	hypothetical protein
312	NCgl0593	1	5.9	hypothetical protein
313	NCgl0597	1	2.7	putative phytoene dehydrogenase [1.3.-.-]
314	NCgl0598	1	3.6	phytoene synthetase
315	NCgl0600	1	2.9	geranylgeranyl pyrophosphate synthase
316	NCgl0602	1	6	lipocalin
317	NCgl0603	11	13.6	predicted nucleoside-diphosphate-sugar epimerase
318	NCgl0604	15	24.1	deoxyribodipyrimidine photolyase
319	NCgl0605	1	8.4	cell wall biogenesis glycosyltransferase
320	NCgl0608	2	8	ABC-type transporter: permease component
321	NCgl0609	15	30.6	ABC-type transporter: ATPase component
322	NCgl0610	13	31.4	ABC-type transporter: periplasmic component
323	NCgl0611	2	1.9	DNA polymerase III alpha subunit
324	NCgl0614	4	14.3	Mn-dependent transcriptional regulator
325	NCgl0616	8	26.4	SIR2 family NAD-dependent protein deacetylase
326	NCgl0617	2	12.2	cytosine/adenosine deaminase
327	NCgl0618	2	7.8	ABC-type Fe ³⁺ -siderophores transport system: periplasmic component
328	NCgl0619	3	7	predicted SpoU class rRNA methylase
329	NCgl0620	19	40.5	5:10-methylene-tetrahydrofolate dehydrogenase [1.5.1.5]
330	NCgl0622	20	30	flotillin-like protein
331	NCgl0623	2	8.4	hypothetical protein
332	NCgl0624	9	22.5	homoserine O-acetyltransferase [2.3.1.11]
333	NCgl0625	37	50.3	O-acetylhomoserine sulfhydrylase [4.2.99.10]
334	NCgl0626	11	10.3	carbon starvation protein: predicted membrane protein
335	NCgl0627	5	20.6	uncharacterized protein
336	NCgl0628	4	12.9	uncharacterized protein
337	NCgl0629	3	18.1	PEP phosphonmutase or related enzyme
338	NCgl0630	5	9.7	citrate synthase [4.1.3.7]
339	NCgl0632	4	21.3	transcriptional regulator
340	NCgl0634	69	39.8	monomeric isocitrate dehydrogenase (NADP+) [1.1.1.42]
341	NCgl0641	15	28.9	exonuclease III
342	NCgl0644	1	7.1	ABC-type transporter: periplasmic component
343	NCgl0647	9	20.9	tryptophanyl-tRNA synthetase [6.1.1.2]
344	NCgl0648	2	6.4	multiple transmembrane domain containing protein
345	NCgl0649	1	5.3	hypothetical membrane protein
346	NCgl0651	10	12.5	hypothetical protein
347	NCgl0654	21	62.1	uracil phosphoribosyltransferase [2.4.2.9]
348	NCgl0655	4	24.2	predicted transcriptional regulator
349	NCgl0656	10	12.5	phosphomannomutase [5.4.2.8]
350	NCgl0657	5	6.3	metal-dependent amidase/aminoacylase/carboxypeptidase
351	NCgl0658	24	32.8	dihydrolipoamide dehydrogenase [1.8.1.4]
352	NCgl0659	102	47.9	pyruvate carboxylase [6.4.1.1]
353	NCgl0660	3	7.6	hypothetical protein
354	NCgl0664	4	6.8	hypothetical protein
355	NCgl0665	1	3.9	PEP phosphonmutase or related enzyme
356	NCgl0666	1	2.6	citrate synthase [4.1.3.7]
357	NCgl0668	4	7.5	predicted transcriptional regulator
358	NCgl0670	97	30.1	acyl-CoA carboxylase [6.3.4.14]
359	NCgl0671	41	69.4	thiosulfate sulfurtransferase [2.8.1.1]
360	NCgl0672	12	23.4	hypothetical protein
361	NCgl0673	7	29.7	hypothetical protein
362	NCgl0674	2	3.9	predicted acyltransferase
363	NCgl0677	51	38	detergent sensitivity rescuer dtsR2

364	NCgl0678	13	22.5	detergent sensitivity rescuer dtsR1
365	NCgl0679	9	25.7	biotin-(acetyl-CoA carboxylase) ligase [6.3.4.15]
366	NCgl0681	2	7.5	phosphoribosylaminoimidazole carboxylase (NCAIR synthetase) [4.1.1.21]
367	NCgl0682	5	1.4	K ⁺ transporter
368	NCgl0684	1	17.6	phosphoribosylaminoimidazole carboxylase [4.1.1.21]
369	NCgl0687	1	3.4	nitrilotriacetate monooxygenase [1.14.13.-]
370	NCgl0697	7	9.4	ABC-type transporter: periplasmic component
371	NCgl0698	3	8.1	ABC-type transporter: ATPase component
372	NCgl0700	5	3.6	helicase family member
373	NCgl0701	2	0.6	hypothetical protein
374	NCgl0702	6	18.4	hypothetical protein
375	NCgl0703	8	6.7	hypothetical protein
376	NCgl0704	4	6	putative helicase [3.6.1.-]
377	NCgl0705	9	3.3	putative helicase
378	NCgl0706	17	10.6	type II restriction enzyme: methylase subunits
379	NCgl0707	15	14.1	SNF2 family helicase
380	NCgl0709	12	23.8	predicted glycosyltransferase
381	NCgl0710	24	36.5	nucleoside-diphosphate-sugar pyrophosphorylase [2.7.7.13]
382	NCgl0711	2	9.6	hypothetical protein
383	NCgl0712	2	27.6	hypothetical protein
384	NCgl0713	2	8.5	hypothetical protein
385	NCgl0714	15	19.4	phosphomannomutase [5.4.2.8]
386	NCgl0715	4	16.7	hypothetical protein
387	NCgl0716	12	22.8	phosphomannose isomerase [5.3.1.8]
388	NCgl0717	1	6.9	hypothetical protein
389	NCgl0718	2	9.4	hypothetical protein
390	NCgl0719	35	38.7	S-adenosylhomocysteine hydrolase [3.3.1.1]
391	NCgl0720	3	20.7	thymidylate kinase [2.7.4.9]
392	NCgl0721	10	24.8	two-component system response regulator
393	NCgl0722	1	3.6	two-component system sensory transduction histidine kinase
394	NCgl0725	23	43.9	ribosome-associated protein Y
395	NCgl0726	52	32.8	preprotein translocase subunit SecA
396	NCgl0727	4	27.9	hypothetical protein
397	NCgl0728	6	14.3	hypothetical protein
398	NCgl0729	3	9.2	hypothetical protein
399	NCgl0730	21	32.8	5-enolpyruvylshikimate-3-phosphate synthase
400	NCgl0733	3	14.1	DNA-directed RNA polymerase specialized sigma subunits
401	NCgl0737	10	26.7	putative helicase
402	NCgl0738	5	40	hypothetical protein
403	NCgl0740	5	21.3	hypothetical protein
404	NCgl0741	10	8.1	putative helicase
405	NCgl0742	16	11.3	putative helicase
406	NCgl0743	4	14.2	predicted NAD-binding component of Kef-type K ⁺ transport system
407	NCgl0744	1	5	NTP pyrophosphohydrolase
408	NCgl0745	11	11.9	putative helicase [3.6.1.-]
409	NCgl0748	15	14.9	hypothetical protein
410	NCgl0749	1	5.7	predicted secreted protein
411	NCgl0750	5	7.2	hypothetical protein
412	NCgl0751	5	25.9	hypothetical protein
413	NCgl0752	2	3.1	hypothetical protein
414	NCgl0754	47	49.2	pyridoxine biosynthesis enzyme
415	NCgl0755	5	20.5	predicted glutamine amidotransferase
416	NCgl0764	1	13.6	hypothetical protein
417	NCgl0765	4	11.2	archaeal fructose-1:6-bisphosphatase

418	NCgl0766	6	12.7	archaeal fructose-1:6-bisphosphatase
419	NCgl0767	8	25	protein chain release factor B
420	NCgl0769	3	10.3	cell division protein
421	NCgl0771	2	10.3	hypothetical protein
422	NCgl0773	1	3.6	siderophore-interacting protein
423	NCgl0774	6	7.8	ABC-type cobalamin/Fe ³⁺ -siderophore transport system: periplasmic component
424	NCgl0776	20	27.5	ABC-type cobalamin/Fe ³⁺ -siderophore transport system: periplasmic component
425	NCgl0779	9	18.3	ABC-type cobalamin/Fe ³⁺ -siderophore transport system: ATPase component
426	NCgl0780	6	18.3	PLP-dependent aminotransferase [2.6.1.7]
427	NCgl0782	7	11.5	putative helicase
428	NCgl0783	2	3.5	hypothetical protein
429	NCgl0784	3	38.1	hypothetical protein
430	NCgl0786	2	7.1	cold shock protein
431	NCgl0788	1	9.3	glutamine cyclotransferase
432	NCgl0789	4	10.2	hypothetical protein
433	NCgl0791	2	10.1	rRNA methylase [2.1.1.66]
434	NCgl0792	4	8.3	hypothetical protein
435	NCgl0793	14	32.2	hypothetical protein
436	NCgl0794	29	58.5	phosphoserine aminotransferase [2.6.1.52]
437	NCgl0795	89	45.8	citrate synthase [4.1.3.7]
438	NCgl0796	6	18.6	FKBP-type peptidyl-prolyl cis-trans isomerase 1 [5.2.1.8]
439	NCgl0797	13	21.1	acetyl-CoA carboxylase beta subunit [6.4.1.2]
440	NCgl0799	7	1.8	Na ⁺ /proline: Na ⁺ /panthothenate symporter or related permease
441	NCgl0802	296	36	fatty-acid synthase [2.3.1.85]
442	NCgl0805	9	17.2	homoserine acetyltransferase [2.3.1.31]
443	NCgl0807	37	69.7	hypothetical protein
444	NCgl0809	5	6.6	dihydrofolate reductase [1.5.1.3]
445	NCgl0810	6	12	thymidylate synthase [2.1.1.45]
446	NCgl0811	11	31.3	inositol monophosphatase family protein
447	NCgl0812	3	1.8	Lhr-like helicase
448	NCgl0813	1	3.5	formamidopyrimidine-DNA glycosylase [3.2.2.23]
449	NCgl0814	3	15.4	alkaline phosphatase like protein
450	NCgl0815	1	5	hypothetical membrane-associated protein
451	NCgl0817	29	28.1	glucose-6-phosphate isomerase [5.3.1.9]
452	NCgl0819	1	8.9	chorismate mutase
453	NCgl0820	24	19.1	hypothetical helicase [3.6.1.-]
454	NCgl0821	1	2.2	ABC-type transporter: permease component
455	NCgl0823	1	7.1	predicted transcriptional regulator
456	NCgl0826	9	40.1	phosphoribosylglycinamide formyltransferase PurN [2.1.2.2]
457	NCgl0827	49	42.7	phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase
458	NCgl0828	7	23.4	citrate lyase beta subunit [4.1.3.6]
459	NCgl0829	10	27.9	transcriptional regulator
460	NCgl0830	4	15.1	hypothetical protein
461	NCgl0831	12	25.3	ribosomal protein S18
462	NCgl0832	10	24.8	ribosomal protein S14
463	NCgl0833	5	37	ribosomal protein L33
464	NCgl0834	32	52.6	ribosomal protein L28
465	NCgl0837	13	38.6	ribosomal protein L31
466	NCgl0838	6	38.6	hypothetical protein
467	NCgl0839	2	6	two-component system: response regulator
468	NCgl0841	5	8.2	trypsin-like serine protease [3.4.21.-]
469	NCgl0842	4	27.2	molybdopterin biosynthesis enzyme
470	NCgl0845	1	13.2	5-formyltetrahydrofolate cyclo-ligase [6.3.3.2]
471	NCgl0846	8	28.6	UDP-glucose pyrophosphorylase [2.7.7.9]

472	NCgl0847	5	10.5	molybdopterin biosynthesis enzyme
473	NCgl0853	10	15	glycosidase [3.2.1.54]
474	NCgl0855	10	15.6	predicted methyltransferase
475	NCgl0857	19	24.8	methionyl-tRNA synthetase [6.1.1.10]
476	NCgl0858	3	6.3	putative helicase
477	NCgl0860	1	2.8	hypothetical protein
478	NCgl0861	1	6.5	hypothetical protein
479	NCgl0865	19	32.7	FAD/FMN-containing dehydrogenase [1.1.1.28]
480	NCgl0871	6	18.4	Mg-dependent DNase
481	NCgl0872	1	5.3	hypothetical protein
482	NCgl0873	6	17.1	dimethyladenosine transferase [2.1.1.-]
483	NCgl0874	1	8.4	putative isopentenyl monophosphate kinase [2.7.1.-]
484	NCgl0875	9	10.8	ABC-type transporter
485	NCgl0877	4	11.3	hypothetical protein
486	NCgl0878	7	10.9	hypothetical protein
487	NCgl0879	4	39.3	hypothetical protein
488	NCgl0880	35	49.7	hypothetical protein
489	NCgl0881	11	19.7	hypothetical protein
490	NCgl0882	5	8.8	enoyl-CoA hydratase/carnithine racemase [4.2.1.17]
491	NCgl0884	1	3.5	hypothetical protein
492	NCgl0886	3	13.5	transcriptional regulator
493	NCgl0892	24	30.2	peptide chain release factor 3
494	NCgl0895	3	2.8	ABC-type transporter: permease component
495	NCgl0898	3	13.2	peptidyl-tRNA hydrolase [3.1.1.29]
496	NCgl0899	5	14.7	dioxygenase [1.13.11.32]
497	NCgl0900	16	21	glyceraldehyde-3-phosphate dehydrogenase/erythrose-4-phosphate dehydrogenase
498	NCgl0902	25	48.5	ribosomal protein L25
499	NCgl0903	8	32.2	predicted lactoylglutathione lyase [4.4.1.5]
500	NCgl0905	31	33.2	phosphoribosylpyrophosphate synthetase [2.7.6.1]
501	NCgl0906	35	42.5	N-acetylglucosamine-1-phosphate uridylyltransferase
502	NCgl0908	1	2.9	putative multicopper oxidase
503	NCgl0909	1	2.6	ABC-type transporter: ATPase component
504	NCgl0914	1	4.1	putative ABC transporter ATPase component
505	NCgl0916	23	34.9	gamma-glutamyltranspeptidase [2.3.2.2]
506	NCgl0919	1	6.4	transposase
507	NCgl0922	1	4.5	hypothetical protein
508	NCgl0924	17	11.2	transcription-repair coupling factor - superfamily II helicase
509	NCgl0925	1	2.3	ABC-type transporter: ATP-binding component
510	NCgl0927	1	1.2	hypothetical protein
511	NCgl0930	2	18.6	hypothetical protein
512	NCgl0932	5	17.9	hypothetical protein
513	NCgl0933	7	52.4	hypothetical protein
514	NCgl0935	132	47.8	enolase [4.2.1.11]
515	NCgl0937	15	49.5	hypothetical protein
516	NCgl0938	6	13.7	exopolyphosphatase [3.6.1.11]
517	NCgl0946	2	17.2	transcription elongation factor
518	NCgl0948	13	42.3	uncharacterized protein
519	NCgl0950	14	27	3-Deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase [4.1.2.15]
520	NCgl0951	1	5.3	undecaprenyl pyrophosphate synthase [2.5.1.31]
521	NCgl0952	5	12.7	hypothetical protein
522	NCgl0953	6	11.8	panthothenate kinase [2.7.1.33]
523	NCgl0954	64	47.9	glycine hydroxymethyltransferase [2.1.2.1]
524	NCgl0957	12	37.4	hypothetical protein
525	NCgl0959	2	6.1	sortase or related acyltransferase

526	NCgl0960	1	3.8	allophanate hydrolase subunit 2
527	NCgl0962	3	12.3	uncharacterized protein
528	NCgl0966	8	18.1	hypothetical protein
529	NCgl0967	49	39.9	fumarase [4.2.1.2]
530	NCgl0972	3	3.1	putative reductase
531	NCgl0974	1	2.8	acyl-CoA dehydrogenase
532	NCgl0976	30	33.4	putative fructose-1:6-bisphosphatase/sedoheptulose 1:7-bisphosphatase
533	NCgl0977	2	13.6	hypothetical protein
534	NCgl0981	8	24	exonuclease VII: large subunit [3.1.11.6]
535	NCgl0982	9	15.1	penicillin tolerance protein
536	NCgl0986	2	6.6	Na ⁺ -dependent transporter of the SNF family
537	NCgl0987	3	9.5	hypothetical protein
538	NCgl0988	18	34.3	predicted GTPase
539	NCgl0989	1	9.1	plasmid maintenance system antidote protein
540	NCgl0990	2	9.1	ornithine carbamoyltransferase [2.1.3.3]
541	NCgl0992	1	5.4	dehydrogenase [1.1.1.105]
542	NCgl0994	16	2.1	GGDEF family protein
543	NCgl0999	1	11.2	hypothetical protein
544	NCgl1011	3	4.1	hypothetical protein
545	NCgl1012	17	35.9	Mg-chelatase subunit ChII
546	NCgl1013	4	12.4	phosphoglycerate mutase [5.4.2.1]
547	NCgl1015	18	23.2	putative PEP phosphonmutase [2.7.8.23]
548	NCgl1016	2	4	ABC-type transporter: duplicated ATPase component
549	NCgl1022	4	16.2	cysteine sulfinatase desulfinate
550	NCgl1023	15	25.4	nicotinate-nucleotide pyrophosphorylase [2.4.2.19]
551	NCgl1024	37	27.3	quinolinate synthase [1.4.3.-]
552	NCgl1025	4	20.6	ADP-ribose pyrophosphatase
553	NCgl1027	6	17.7	hypothetical protein
554	NCgl1029	3	6	lipoate-protein ligase A
555	NCgl1030	1	12.4	hypothetical protein
556	NCgl1032	8	11.9	2-polyprenyl-6-methoxyphenol hydroxylase [1.14.13.2]
557	NCgl1035	1	6.4	ABC-type transporter: permease component
558	NCgl1040	7	9.9	excinuclease ATPase subunit
559	NCgl1041	16	23.2	peroxiredoxin [1.11.1.-]
560	NCgl1042	2	1.4	hypothetical membrane protein
561	NCgl1043	3	5.3	hypothetical protein
562	NCgl1045	1	17.7	hypothetical protein
563	NCgl1048	1	3.2	hypothetical protein
564	NCgl1049	7	53	arsenate reductase
565	NCgl1051	10	24	hypothetical protein
566	NCgl1053	49	43.6	predicted membrane GTPase involved in stress response
567	NCgl1055	18	40.7	hypothetical protein
568	NCgl1057	17	45.7	ferredoxin 3
569	NCgl1058	10	14.4	PLP-dependent aminotransferase [2.6.1.1]
570	NCgl1061	4	10.1	tetrahydrodipicolinate N-succinyltransferase [2.3.1.117]
571	NCgl1062	1	2.6	gamma-aminobutyrate permease
572	NCgl1063	13	29.4	tetrahydrodipicolinate N-succinyltransferase
573	NCgl1064	5	13.8	succinyl-diaminopimelate desuccinylase [3.5.1.18]
574	NCgl1065	4	12.1	predicted Rossmann fold nucleotide-binding protein
575	NCgl1066	6	8.7	putative dihydropteroate synthase [2.5.1.15]
576	NCgl1068	1	7.8	hypothetical protein
577	NCgl1069	3	23.6	hypothetical protein
578	NCgl1070	20	37.8	SAM-dependent methyltransferase
579	NCgl1071	17	21.5	beta-fructosidase [3.2.1.26]

580	NCgl1072	13	23	predicted glycosyltransferase
581	NCgl1073	12	23.7	ADP-glucose pyrophosphorylase [2.7.7.27]
582	NCgl1074	15	39.9	hypothetical protein
583	NCgl1077	7	34.6	hypothetical protein
584	NCgl1078	9	14.7	ATPase
585	NCgl1079	4	13.5	hypothetical membrane protein
586	NCgl1080	5	10	Mg/Co/Ni transporter MgtE
587	NCgl1081	1	4.8	hypothetical membrane protein
588	NCgl1083	1	9.6	hypothetical protein
589	NCgl1084	142	41.1	2-oxoglutarate dehydrogenases: E1 component [1.2.4.2]
590	NCgl1085	2	2.2	ABC-type transporter: duplicated ATPase component
591	NCgl1086	3	11.3	hypothetical membrane protein
592	NCgl1087	8	20.9	shikimate 5-dehydrogenase [1.1.1.25]
593	NCgl1088	3	7.6	carboxylesterase type B [3.1.1.-]
594	NCgl1089	2	14.3	hypothetical protein
595	NCgl1090	9	17.6	hypothetical protein
596	NCgl1091	1	6.1	hypothetical protein
597	NCgl1094	249	56.5	methionine synthase II [2.1.1.14]
598	NCgl1096	2	4.5	predicted flavoprotein involved in K ⁺ transport
599	NCgl1099	2	3	predicted hydrolase
600	NCgl1100	1	1.5	non-ribosomal peptide synthetase modules and related proteins
601	NCgl1102	1	2	ABC-type transporter: ATPase component and permease component
602	NCgl1106	8	22.1	putative mutT-like protein
603	NCgl1108	3	4.5	putative gamma-aminobutyrate permease
604	NCgl1109	29	22.7	putative helicase
605	NCgl1112	2	8.7	alcohol dehydrogenase IV [1.3.1.32]
606	NCgl1113	1	5.4	protocatechuate 3:4-dioxygenase beta subunit [1.13.11.1]
607	NCgl1114	1	10.2	hypothetical membrane protein
608	NCgl1116	1	3.8	putative Na ⁺ /proline: Na ⁺ /panthothenate symporter
609	NCgl1117	8	5.9	putative helicase
610	NCgl1118	4	9.5	hypothetical protein
611	NCgl1120	1	1.8	ATPase involved in DNA repair
612	NCgl1123	19	71.5	hypothetical protein
613	NCgl1124	3	14.2	transcriptional regulator
614	NCgl1131	2	5	predicted hydrolase or acyltransferase
615	NCgl1132	33	30.4	arginyl-tRNA synthetase [6.1.1.19]
616	NCgl1133	10	15.5	diaminopimelate decarboxylase [4.1.1.20]
617	NCgl1134	3	8	hypothetical protein
618	NCgl1136	49	46.7	homoserine dehydrogenase [1.1.1.3]
619	NCgl1137	11	22.3	homoserine kinase [2.7.1.39]
620	NCgl1138	3	6.1	hypothetical protein
621	NCgl1140	6	19.7	nitrate reductase delta subunit [1.7.99.4]
622	NCgl1141	28	42.7	nitrate reductase beta chain [1.7.99.4]
623	NCgl1142	34	22.6	anaerobic dehydrogenase [1.7.99.4]
624	NCgl1144	3	19.1	molybdopterin biosynthesis enzyme
625	NCgl1149	8	10.2	molybdopterin biosynthesis enzyme
626	NCgl1150	2	5	molybdenum cofactor biosynthesis enzyme
627	NCgl1151	2	5.6	acyl-CoA synthetase [6.2.1.-]
628	NCgl1152	31	17.7	transcription termination factor
629	NCgl1153	25	41.6	protein chain release factor A
630	NCgl1155	6	21.8	putative translation factor (SUA5)
631	NCgl1156	1	2.9	UDP-N-acetylmuramyl pentapeptide phosphotransferase
632	NCgl1159	3	10.4	F0F1-type ATP synthase a subunit [3.6.1.34]
633	NCgl1160	13	20	F0F1-type ATP synthase c subunit [3.6.1.34]

634	NCgl1161	6	21.3	F0F1-type ATP synthase b subunit [3.6.1.34]
635	NCgl1162	17	24.7	F0F1-type ATP synthase delta subunit [3.6.1.34]
636	NCgl1163	45	41.4	F0F1-type ATP synthase alpha subunit [3.6.1.34]
637	NCgl1164	22	32.6	F0F1-type ATP synthase gamma subunit [3.6.1.34]
638	NCgl1165	100	58.6	F0F1-type ATP synthase beta subunit [3.6.1.34]
639	NCgl1166	2	11.3	F0F1-type ATP synthase epsilon subunit [3.6.1.34]
640	NCgl1167	2	12.1	hypothetical protein
641	NCgl1168	6	23.9	predicted nuclease of the RecB family
642	NCgl1170	8	38.4	lactoylglutathione lyase
643	NCgl1171	2	16.2	hypothetical protein
644	NCgl1172	8	7.8	thioredoxin domain-containing protein
645	NCgl1173	1	3.1	coenzyme F420-dependent N5:N10-methylene tetrahydromethanopterin reductase
646	NCgl1175	2	9.5	ABC-type transporter: ATPase component
647	NCgl1177	39	21.9	1:4-alpha-glucan branching enzyme [2.4.1.18]
648	NCgl1178	28	24.9	glycosidase [3.2.1.1]
649	NCgl1179	4	8.5	ABC-type transporter: ATPase component
650	NCgl1180	2	6.8	hypothetical protein
651	NCgl1181	5	10.9	hypothetical protein
652	NCgl1182	12	23.7	electron transfer flavoprotein beta-subunit
653	NCgl1183	10	20.5	electron transfer flavoprotein alpha-subunit
654	NCgl1184	5	20.2	cysteine sulfinase
655	NCgl1189	4	12.1	spermidine synthase
656	NCgl1190	5	12.6	hypothetical protein
657	NCgl1192	8	26	predicted tRNA (5-methylaminomethyl-2-thiouridylate) methyltransferase [2.1.1.1]
658	NCgl1193	2	12.1	hypothetical protein
659	NCgl1195	1	8	DNA polymerase III alpha subunit
660	NCgl1196	16	19.1	NAD-dependent DNA ligase [6.5.1.2]
661	NCgl1197	8	22.6	hypothetical protein
662	NCgl1198	3	18.2	Asp-tRNAAsn/Glu-tRNA ^{Gln} amidotransferase C subunit
663	NCgl1199	25	32.6	putative Asp-tRNAAsn/Glu-tRNA ^{Gln} amidotransferase A subunit [6.3.5.-]
664	NCgl1200	17	39.2	siderophore-interacting protein
665	NCgl1201	1	19.8	hypothetical membrane protein
666	NCgl1202	43	55.2	6-phosphofructokinase [2.7.1.11]
667	NCgl1203	4	2.4	transcriptional regulator
668	NCgl1204	2	5.2	ABC-type transporter: duplicated ATPase component
669	NCgl1205	1	4.9	ABC-type transporter: permease component
670	NCgl1208	1	4.2	acetyltransferase
671	NCgl1209	4	14.8	ABC-type transport system: periplasmic component
672	NCgl1211	35	35.7	Asp-tRNAAsn/Glu-tRNA ^{Gln} amidotransferase B subunit
673	NCgl1212	1	4.9	predicted dinucleotide-binding enzyme [1.-.-.-]
674	NCgl1213	13	30.5	predicted oxidoreductases
675	NCgl1215	4	10	transcriptional regulator
676	NCgl1216	15	42.6	predicted glutathione S-transferase
677	NCgl1218	3	9.7	hypothetical membrane protein
678	NCgl1219	39	33.1	dihydroxyacid dehydratase/phosphogluconate dehydratase
679	NCgl1220	2	6.2	hypothetical protein
680	NCgl1222	36	32.9	thiamine pyrophosphate-requiring enzyme [4.1.3.18]
681	NCgl1223	25	70.3	acetolactate synthase: small subunit [4.1.3.18]
682	NCgl1224	39	30.2	ketol-acid reductoisomerase [1.1.1.86]
683	NCgl1230	1	7.5	hypothetical protein
684	NCgl1231	2	15.4	hypothetical protein
685	NCgl1233	30	26.6	hypothetical protein
686	NCgl1235	121	66.8	phosphoglycerate dehydrogenase [1.1.1.95]
687	NCgl1236	5	23.5	hypothetical protein

688	NCgl1237	17	37.1	isocitrate/isopropylmalate dehydrogenase [1.1.1.85]
689	NCgl1239	16	24.4	predicted signal-transduction protein
690	NCgl1240	2	11	DNA polymerase III epsilon subunit
691	NCgl1241	11	31	2-keto-4-pentenoate hydratase/2-oxohepta-3-ene-1:7-dioic acid hydratase [5.3.1.11]
692	NCgl1243	7	18.6	isochorismate synthase [5.4.99.6]
693	NCgl1244	42	42.6	glutamyl- and glutaminyl-tRNA synthetases [6.1.1.17]
694	NCgl1247	2	12.6	hypothetical protein
695	NCgl1251	3	4.9	hypothetical protein
696	NCgl1253	1	1.4	thiamine biosynthesis protein ThiC
697	NCgl1255	157	55.2	glucan phosphorylase [2.4.1.1]
698	NCgl1257	7	22.4	Zn-dependent hydrolase
699	NCgl1261	8	29.4	transcriptional regulator
700	NCgl1262	33	31.6	3-isopropylmalate dehydratase large subunit [4.2.1.33]
701	NCgl1263	14	47.7	3-isopropylmalate dehydratase small subunit [4.2.1.33]
702	NCgl1264	3	12.8	NTP pyrophosphohydrolase [3.6.1.-]
703	NCgl1266	12	24.1	glycerol 3-phosphate dehydrogenase [1.1.1.94]
704	NCgl1267	11	22.5	D-alanine-D-alanine ligase [6.3.2.4]
705	NCgl1269	4	20.7	thiamine monophosphate kinase [2.7.4.16]
706	NCgl1270	3	9.6	uracil DNA glycosylase [3.2.2.-]
707	NCgl1271	1	2.1	predicted kinase
708	NCgl1272	1	1.3	RecG-like helicase [3.6.1.-]
709	NCgl1273	1	29.6	pyruvate carboxylase: C-terminal domain/subunit [2.1.3.1]
710	NCgl1274	3	9.3	N6-adenine-specific methylase
711	NCgl1275	2	23.8	phosphopantetheine adenylyltransferase
712	NCgl1276	4	15.4	ABC-type transporter: ATPase component
713	NCgl1278	2	5.4	ABC-type transporter: periplasmic component
714	NCgl1279	1	5.2	hypothetical protein
715	NCgl1284	1	13.9	hypothetical protein
716	NCgl1285	1	4.5	hypothetical protein
717	NCgl1288	2	8.3	hypothetical protein
718	NCgl1289	1	6.7	hypothetical protein
719	NCgl1291	1	7.6	hypothetical protein
720	NCgl1292	1	10.4	hypothetical protein
721	NCgl1299	33	26.6	DNA polymerase I [2.7.7.7]
722	NCgl1302	2	5.2	aldo/keto reductase [1.1.1.218]
723	NCgl1303	3	11.2	SAM-dependent methyltransferase
724	NCgl1304	145	47.7	ribosomal protein S1
725	NCgl1305	19	17.1	phosphotransferase system IIC component: glucose/maltose/N-acetylglucosamine
726	NCgl1306	5	21	dephospho-CoA kinase
727	NCgl1307	1	2.2	hypothetical protein
728	NCgl1309	7	13.8	inosine-uridine nucleoside N-ribohydrolase [3.2.2.1]
729	NCgl1311	4	5.5	ribokinase family sugar kinase [2.7.1.15]
730	NCgl1315	12	17.9	helicase subunit of the DNA excision repair complex
731	NCgl1316	14	38.1	universal stress protein UspA
732	NCgl1318	12	24.5	predicted nucleoside-diphosphate-sugar epimerase
733	NCgl1319	1	1.7	hypothetical protein
734	NCgl1320	25	37.2	hypothetical protein
735	NCgl1321	6	24.5	hypothetical protein
736	NCgl1322	21	20.9	excinuclease ATPase subunit
737	NCgl1324	5	31.2	translation initiation factor IF3
738	NCgl1326	7	33.1	ribosomal protein L20
739	NCgl1327	2	3.3	hypothetical protein
740	NCgl1328	2	14.9	hypothetical protein
741	NCgl1333	4	12.8	glycerophosphoryl diester phosphodiesterase [3.1.4.46]

742	NCgl1334	5	27.8	rRNA methylase [2.1.1.34]
743	NCgl1335	19	33.3	phenylalanyl-tRNA synthetase alpha subunit [6.1.1.20]
744	NCgl1336	43	36.3	phenylalanyl-tRNA synthetase beta subunit [6.1.1.20]
745	NCgl1337	4	9.9	hypothetical protein
746	NCgl1340	6	13.7	acetylglutamate semialdehyde dehydrogenase [1.2.1.38]
747	NCgl1341	9	30.2	Ornithine acetyltransferase [2.3.1.35]
748	NCgl1342	17	36	N-acetylglutamate kinase [2.7.2.8]
749	NCgl1343	30	44.8	PLP-dependent aminotransferase [2.6.1.11]
750	NCgl1344	18	33.9	ornithine carbamoyltransferase [2.1.3.3]
751	NCgl1345	4	17.9	arginine repressor
752	NCgl1346	46	48.6	argininosuccinate synthase [6.3.4.5]
753	NCgl1347	17	24.5	argininosuccinate lyase [4.3.2.1]
754	NCgl1349	1	1.9	hypothetical protein
755	NCgl1351	2	44.1	hypothetical protein
756	NCgl1352	16	30	tyrosyl-tRNA synthetase [6.1.1.1]
757	NCgl1353	10	43.9	universal stress protein UspA
758	NCgl1354	11	22.4	TPR-repeat-containing protein
759	NCgl1355	2	11.9	predicted sugar phosphatase of the HAD superfamily
760	NCgl1356	1	31.5	hypothetical protein
761	NCgl1357	4	9.5	predicted rRNA methylase
762	NCgl1358	3	12.8	predicted kinase
763	NCgl1360	8	15.1	hypothetical protein
764	NCgl1361	5	13.4	hypothetical protein
765	NCgl1362	18	23.8	CTP synthase [6.3.4.2]
766	NCgl1364	5	11.5	integrase
767	NCgl1366	11	10.7	ATPase involved in chromosome partitioning
768	NCgl1367	2	11	hypothetical protein
769	NCgl1368	4	17.6	acetyltransferase
770	NCgl1369	4	16.6	rhodanese-related sulfurtransferase
771	NCgl1370	2	8.3	predicted transcriptional regulator
772	NCgl1371	30	41.6	16S rRNA uridine-516 pseudouridylate synthase [4.2.1.70]
773	NCgl1372	9	13.6	cytidylate kinase [2.7.4.14]
774	NCgl1373	27	35.4	predicted GTPase
775	NCgl1378	1	2.8	ABC-type transport system ATPase component
776	NCgl1380	6	6.4	NhaP-type Na ⁺ /H ⁺ and K ⁺ /H ⁺ antiporter
777	NCgl1382	2	24	hypothetical protein
778	NCgl1383	2	4.2	hydrolase of the alpha/beta superfamily
779	NCgl1384	27	24.9	preprotein translocase subunit SecA
780	NCgl1385	27	72	FHA-domain-containing protein
781	NCgl1387	8	11.8	hypothetical protein
782	NCgl1388	1	6.8	predicted transcriptional regulator
783	NCgl1389	1	4.7	hypothetical membrane protein
784	NCgl1390	3	4.3	hypothetical protein
785	NCgl1391	3	3.4	hypothetical protein
786	NCgl1394	14	28.1	putative helicase
787	NCgl1396	45	32.2	6-phosphogluconate dehydrogenase: family 1 [1.1.1.44]
788	NCgl1397	5	30.5	uncharacterized protein
789	NCgl1398	4	17	SAM-dependent methyltransferase
790	NCgl1401	1	3.4	transcriptional regulator
791	NCgl1402	1	3.5	ABC-type transporter: permease component
792	NCgl1407	4	7.8	hydroxymethylpyrimidine/phosphomethylpyrimidine kinase [2.7.4.7]
793	NCgl1408	2	11.2	hydroxyethylthiazole kinase [2.7.1.50]
794	NCgl1409	70	40.7	NADH dehydrogenase: FAD-containing subunit [1.6.99.3]
795	NCgl1410	3	4.8	cyclopropane fatty acid synthase [2.1.1.79]

796	NCgl1412	6	25.3	predicted phosphoribosyltransferase
797	NCgl1415	2	10.6	hypothetical protein
798	NCgl1416	5	14.7	hypothetical protein
799	NCgl1420	1	4	hypothetical protein
800	NCgl1422	15	22.7	hypothetical protein
801	NCgl1423	9	20.4	glycosyltransferase
802	NCgl1426	2	5.3	hypothetical protein [3.1.1.3]
803	NCgl1428	6	10.8	precorrin-6B methylase 1 [2.1.1.132]
804	NCgl1430	10	18.2	Xaa-Pro aminopeptidase [3.4.13.9]
805	NCgl1431	1	6.1	hypothetical protein
806	NCgl1432	15	16.9	putative helicase
807	NCgl1435	3	7.3	predicted transcriptional regulator
808	NCgl1437	8	14.8	hypothetical protein
809	NCgl1438	1	15.6	hypothetical protein
810	NCgl1439	13	17.1	hypothetical protein
811	NCgl1440	8	12.9	ATPase of the AAA+ class
812	NCgl1441	16	27.7	predicted SAM-dependent methyltransferase involved in tRNA-Met maturation
813	NCgl1442	18	36	aspartyl aminopeptidase
814	NCgl1443	3	8.6	RecB family exonuclease
815	NCgl1446	9	13.5	aspartate ammonia-lyase [4.3.1.1]
816	NCgl1447	14	45.2	ATP phosphoribosyltransferase [2.4.2.17]
817	NCgl1448	3	13.8	phosphoribosyl-ATP pyrophosphohydrolase [3.6.1.31]
818	NCgl1449	3	15.6	predicted phosphatase/phosphohexomutase
819	NCgl1450	66	33.8	Methionine synthase I: cobalamin-binding domain [2.1.1.13]
820	NCgl1452	4	6.4	predicted flavoprotein involved in K ⁺ transport [1.6.4.-]
821	NCgl1454	1	15.7	protein-tyrosine-phosphatase
822	NCgl1455	3	4.7	protein-tyrosine-phosphatase
823	NCgl1456	13	64.3	hypothetical protein
824	NCgl1457	9	17.9	cysteinyl-tRNA synthetase [6.1.1.16]
825	NCgl1459	1	3.2	predicted oxidoreductases
826	NCgl1460	3	8.4	hypothetical protein
827	NCgl1461	13	36.9	dihydroorotate dehydrogenase [1.3.3.1]
828	NCgl1466	10	34.5	phospholipid-binding protein
829	NCgl1470	7	16.3	putative periplasmic protein kinase ArgK [2.7.-.-]
830	NCgl1471	33	33.6	methylmalonyl-CoA mutase: N-terminal domain/subunit [5.4.99.2]
831	NCgl1472	29	22.2	methylmalonyl-CoA mutase: N-terminal domain/subunit [5.4.99.2]
832	NCgl1475	5	4.6	membrane protease subunit
833	NCgl1477	1	3.2	hypothetical protein
834	NCgl1478	1	5	hypothetical protein
835	NCgl1479	22	31.1	protoheme ferro-lyase [4.99.1.1]
836	NCgl1480	3	4.6	cell wall-associated hydrolase
837	NCgl1482	124	37.2	aconitase A [4.2.1.3]
838	NCgl1483	5	12.2	transcriptional regulator
839	NCgl1484	7	22.6	GMP synthase [6.3.5.2]
840	NCgl1485	1	5.8	predicted nucleoside-diphosphate-sugar epimerase
841	NCgl1488	3	0.9	cation transport ATPase [3.6.1.-]
842	NCgl1489	2	5.7	hypothetical protein
843	NCgl1493	13	18.8	ABC-type transporter: duplicated ATPase component
844	NCgl1498	1	6.6	putative aromatic ring hydroxylating enzyme
845	NCgl1499	4	14.1	NifU-like protein
846	NCgl1500	11	23.8	selenocysteine lyase
847	NCgl1501	24	43.3	ABC-type transporter: ATPase component
848	NCgl1502	26	46.4	predicted iron-regulated ABC-type transporter SufB
849	NCgl1503	32	52.8	predicted iron-regulated ABC-type transporter SufB

850	NCgl1504	1	8.8	predicted transcriptional regulator
851	NCgl1509	4	2.7	putative helicase
852	NCgl1510	11	29.6	NADPH:quinone reductase [1.6.5.5]
853	NCgl1512	59	42	transketolase [2.2.1.1]
854	NCgl1513	26	44.7	transaldolase [2.2.1.2]
855	NCgl1514	31	29.8	glucose-6-phosphate 1-dehydrogenase [1.1.1.49]
856	NCgl1515	21	35.1	hypothetical protein
857	NCgl1516	18	35.3	6-phosphogluconolactonase [3.1.1.31]
858	NCgl1520	1	2.9	predicted ornithine cyclodeaminase
859	NCgl1523	37	22.4	phosphoenolpyruvate carboxylase [4.1.1.31]
860	NCgl1524	53	55.2	triosephosphate isomerase [5.3.1.1]
861	NCgl1525	42	54.1	3-phosphoglycerate kinase [2.7.2.3]
862	NCgl1526	8	24.6	glyceraldehyde-3-phosphate dehydrogenase [1.2.1.12]
863	NCgl1527	5	14.7	hypothetical protein
864	NCgl1528	11	24.6	hypothetical protein
865	NCgl1529	11	30.1	predicted P-loop-containing kinase
866	NCgl1530	2	3.7	nuclease subunit of the excinuclease complex
867	NCgl1532	5	34	riboflavin synthase beta-chain [2.5.1.9]
868	NCgl1533	10	19	GTP cyclohydrolase II [3.5.4.25]
869	NCgl1534	5	19.9	riboflavin synthase alpha chain [2.5.1.9]
870	NCgl1536	5	25.6	pentose-5-phosphate-3-epimerase [5.1.3.1]
871	NCgl1537	6	14.2	tRNA and rRNA cytosine-C5-methylase
872	NCgl1538	9	32.7	methionyl-tRNA formyltransferase [2.1.2.9]
873	NCgl1539	8	23.7	N-formylmethionyl-tRNA deformylase [3.5.1.31]
874	NCgl1540	5	3.6	primosomal protein N
875	NCgl1541	33	44	S-adenosylmethionine synthetase [2.5.1.6]
876	NCgl1542	7	18.8	phosphopantothencysteine synthetase/decarboxylase
877	NCgl1543	6	30.5	DNA-directed RNA polymerase subunit K/omega
878	NCgl1544	9	26.3	guanylate kinase [2.7.4.8]
879	NCgl1545	16	38.7	hypothetical protein
880	NCgl1546	3	8.6	orotidine-5-phosphate decarboxylase [4.1.1.23]
881	NCgl1547	71	37.6	carbamoylphosphate synthase large subunit [6.3.5.5]
882	NCgl1548	22	44	carbamoylphosphate synthase small subunit [6.3.5.5]
883	NCgl1549	19	27.7	dihydroorotase [3.5.2.3]
884	NCgl1550	19	25.3	aspartate carbamoyltransferase: catalytic chain [2.1.3.2]
885	NCgl1551	8	33.3	pyrimidine operon attenuation protein [2.4.2.9]
886	NCgl1552	5	9	predicted Sula family nucleoside-diphosphate sugar epimerase
887	NCgl1553	4	27	hypothetical protein
888	NCgl1556	2	4.4	transcription termination factor
889	NCgl1557	34	48.1	translation elongation factor P
890	NCgl1558	23	38.8	Xaa-Pro aminopeptidase
891	NCgl1559	10	24.9	3-dehydroquinate synthetase [4.6.1.3]
892	NCgl1560	8	30	shikimate kinase [2.7.1.71]
893	NCgl1561	22	32.9	chorismate synthase [4.6.1.4]
894	NCgl1568	3	8.5	predicted periplasmic solute-binding protein
895	NCgl1569	10	38.5	predicted endonuclease involved in recombination
896	NCgl1570	44	24.9	alanyl-tRNA synthetase [6.1.1.7]
897	NCgl1571	4	10.5	uncharacterized ATPase related to the helicase subunit of the Holliday junction
898	NCgl1572	13	32.8	hypothetical protein
899	NCgl1573	54	36.2	aspartyl-tRNA synthetase [6.1.1.12]
900	NCgl1574	3	6.6	predicted metalloprotease
901	NCgl1575	3	4.3	hypothetical helicase [3.2.1.3]
902	NCgl1576	2	6	predicted membrane protein
903	NCgl1578	1	6.6	transcriptional regulator

904	NCgl1579	2	12.6	hypothetical protein
905	NCgl1580	2	3	coenzyme F420-dependent N5:N10-methylene tetrahydromethanopterin reductase
906	NCgl1584	20	20.2	glycerol-3-phosphate dehydrogenase
907	NCgl1585	31	47.6	histidyl-tRNA synthetase [6.1.1.21]
908	NCgl1586	3	5.6	putative Zn-dependent hydrolase
909	NCgl1587	3	7.1	peptidyl-prolyl cis-trans isomerase (rotamase) [5.2.1.8]
910	NCgl1588	6	16.2	hypothetical protein
911	NCgl1590	21	18.6	guanosine polyphosphate pyrophosphohydrolase/synthetase [2.7.6.5]
912	NCgl1591	11	40.5	adenine/guanine phosphoribosyltransferase [2.4.2.7]
913	NCgl1592	4	6.6	dipeptide-binding protein DciAE
914	NCgl1594	2	3	preprotein translocase subunit SecD
915	NCgl1596	1	6.6	Holliday junction resolvosome helicase subunit
916	NCgl1597	3	7.3	Holliday junction resolvosome DNA-binding subunit
917	NCgl1599	13	23.5	hypothetical protein
918	NCgl1600	4	11.7	acyl-CoA thioesterase [3.1.2.-]
919	NCgl1603	1	7.5	predicted glycosyltransferase [2.4.1.-]
920	NCgl1604	2	9	lauroyl/myristoyl acyltransferase
921	NCgl1606	2	9.4	diadenosine tetraphosphate (Ap4A) hydrolase
922	NCgl1607	54	38.2	threonyl-tRNA synthetase [6.1.1.3]
923	NCgl1608	3	5.5	predicted iron-dependent peroxidase
924	NCgl1609	3	11	hypothetical protein
925	NCgl1611	2	2.1	hypothetical protein
926	NCgl1614	5	24.9	histone acetyltransferase HPA2 [2.3.-.-]
927	NCgl1615	2	11.4	hypothetical protein
928	NCgl1623	1	5.5	ABC-type transporter: ATPase component
929	NCgl1626	1	4.1	phosphopantothenoylecysteine synthetase/decarboxylase
930	NCgl1627	1	3.9	hypothetical protein
931	NCgl1629	3	14.3	hypothetical protein
932	NCgl1631	4	4.1	hypothetical protein
933	NCgl1633	5	6.5	hypothetical protein
934	NCgl1640	1	4.6	hypothetical protein
935	NCgl1647	1	12	hypothetical protein
936	NCgl1650	6	15	transcriptional regulator
937	NCgl1654	1	3.6	hypothetical protein
938	NCgl1658	1	1.5	hypothetical protein
939	NCgl1659	4	1.6	hypothetical protein
940	NCgl1663	1	6.7	hypothetical protein
941	NCgl1669	2	3.7	predicted ATPase
942	NCgl1670	5	1.7	hypothetical protein
943	NCgl1674	1	9.6	hypothetical protein
944	NCgl1676	22	56	hypothetical protein
945	NCgl1685	1	5.2	hypothetical membrane protein
946	NCgl1699	1	26.7	hypothetical protein
947	NCgl1700	3	1.3	hypothetical protein
948	NCgl1702	7	2.8	hypothetical protein
949	NCgl1703	14	18.5	site-specific DNA methylase or
950	NCgl1704	4	6.4	stress-sensitive restriction system protein 1
951	NCgl1705	4	6.2	stress-sensitive restriction system protein 2
952	NCgl1708	2	3.6	hypothetical protein
953	NCgl1715	2	4.6	hypothetical protein
954	NCgl1716	1	1.7	ATPase with chaperone activity: ATP-binding subunit
955	NCgl1720	2	15	hypothetical protein
956	NCgl1721	1	1.4	hypothetical protein
957	NCgl1728	1	4.1	hypothetical protein

958	NCgl1731	5	34.3	hypothetical protein
959	NCgl1732	3	18.9	hypothetical protein
960	NCgl1734	3	11.2	hypothetical protein
961	NCgl1738	3	11.9	hypothetical protein
962	NCgl1739	2	3.5	hypothetical protein
963	NCgl1744	2	7.1	hypothetical protein
964	NCgl1745	2	18.9	predicted transcriptional regulator
965	NCgl1747	1	8.2	hypothetical protein
966	NCgl1748	4	8	periplasmic serine protease
967	NCgl1755	1	5.4	hypothetical protein
968	NCgl1756	4	12	hypothetical protein
969	NCgl1761	1	6.2	hypothetical protein
970	NCgl1766	4	5.7	hypothetical protein
971	NCgl1774	2	1.9	hypothetical protein
972	NCgl1806	2	7	integrase
973	NCgl1816	1	4.8	hypothetical protein
974	NCgl1821	2	8.2	pyrimidine reductase
975	NCgl1823	10	50	hypothetical protein
976	NCgl1824	21	44.8	hypothetical protein
977	NCgl1825	6	26.9	hypothetical protein
978	NCgl1826	1	2.2	ribonuclease D [3.1.26.3]
979	NCgl1827	21	26.6	deoxyxylulose-5-phosphate synthase
980	NCgl1828	7	20.1	SAM-dependent methyltransferase [2.1.1.-]
981	NCgl1829	7	17.8	hypothetical protein
982	NCgl1830	9	61.7	dUTPase [3.6.1.23]
983	NCgl1831	3	16.9	hypothetical membrane protein
984	NCgl1833	14	46.4	hypothetical protein
985	NCgl1834	3	16.4	archaeal fructose-1:6-bisphosphatase
986	NCgl1835	20	43.2	transcriptional regulator [2.7.1.63]
987	NCgl1836	34	28.3	DNA-directed RNA polymerase sigma subunit SigA
988	NCgl1839	5	9.1	hypothetical protein
989	NCgl1842	8	12.8	predicted rRNA or tRNA methylase
990	NCgl1844	13	20.5	DNA-directed RNA polymerase sigma subunit SigB
991	NCgl1845	24	53.1	Mn-dependent transcriptional regulator
992	NCgl1846	8	23.4	UDP-glucose 4-epimerase [5.1.3.2]
993	NCgl1847	1	2.7	hypothetical protein
994	NCgl1848	7	14.4	hypothetical protein
995	NCgl1849	12	14.5	putative helicase
996	NCgl1850	10	23.5	transcriptional regulator
997	NCgl1852	18	12.1	HrpA-like helicase
998	NCgl1853	1	6.7	predicted transcriptional regulator
999	NCgl1855	6	26.1	SOS-response transcriptional repressor [3.4.21.88]
1000	NCgl1856	4	13.1	transcriptional regulator of sugar metabolism
1001	NCgl1857	4	13.8	fructose-1-phosphate kinase [2.7.1.56]
1002	NCgl1858	52	40.1	phosphoenolpyruvate-protein kinase [2.7.3.9]
1003	NCgl1859	2	3.8	transcriptional regulator of sugar metabolism
1004	NCgl1860	6	16.4	putative fructose-1-phosphate kinase [2.7.1.11]
1005	NCgl1861	3	1.7	phosphotransferase system: fructose-specific IIC component [2.7.1.69]
1006	NCgl1862	1	13.5	hypothetical protein
1007	NCgl1863	6	22.3	hypothetical protein
1008	NCgl1865	4	9.5	GTPase
1009	NCgl1866	4	12.7	hypothetical protein
1010	NCgl1867	3	12.3	hypothetical protein
1011	NCgl1868	9	30.3	diaminopimelate epimerase [5.1.1.7]

1012	NCgl1869	5	18.3	tRNA delta(2)-isopentenylpyrophosphate transferase [2.5.1.8]
1013	NCgl1871	52	51.4	hypothetical protein
1014	NCgl1872	1	6.2	hypothetical protein
1015	NCgl1874	16	18.8	2-methylthioadenine synthetase
1016	NCgl1875	8	20.8	glutamate ABC-type transporter: ATPase component
1017	NCgl1876	18	34.2	glutamate ABC-type transporter: periplasmic component
1018	NCgl1877	5	13.6	glutamate ABC-type transporter: permease component
1019	NCgl1879	1	4.7	hypothetical protein
1020	NCgl1880	20	26.1	RecA/RadA recombinase
1021	NCgl1885	3	10.3	ABC-type transporter: permease component
1022	NCgl1886	37	45.3	phage shock protein A (IM30)
1023	NCgl1890	4	27.4	hypothetical protein
1024	NCgl1892	1	7	membrane protein TerC
1025	NCgl1893	5	4	DNA translocase spoIIIE-like protein
1026	NCgl1894	7	28.9	hypothetical protein
1027	NCgl1895	37	28.3	predicted hydrolase of the metallo-beta-lactamase superfamily
1028	NCgl1896	24	44.2	dihydrodipicolinate synthase/N-acetylneuraminic lyase [4.2.1.52]
1029	NCgl1897	5	17.6	predicted alternative thymidylate synthase
1030	NCgl1898	20	40.7	dihydrodipicolinate reductase [1.3.1.26]
1031	NCgl1899	1	6.4	hypothetical protein
1032	NCgl1900	52	40.5	polyribonucleotide nucleotidyltransferase [2.7.7.8]
1033	NCgl1901	25	69.7	ribosomal protein S15P/S13E
1034	NCgl1902	10	4.7	inosine-uridine nucleoside N-ribohydrolase
1035	NCgl1903	8	12.9	FAD synthase [2.7.1.26]
1036	NCgl1904	2	3.4	pseudouridine synthase [4.2.1.70]
1037	NCgl1905	7	14.3	phosphopantetheinyl transferase component of siderophore synthetase
1038	NCgl1906	6	19.8	hypothetical protein
1039	NCgl1908	4	16	exopolyphosphatase-related protein
1040	NCgl1909	14	52.3	ribosome-binding factor A
1041	NCgl1910	44	22.1	translation initiation factor 2
1042	NCgl1911	1	15.2	predicted nucleic-acid-binding protein
1043	NCgl1912	9	18.1	transcription terminator
1044	NCgl1913	7	16.9	hypothetical protein
1045	NCgl1915	12	11	ABC-type transporter: periplasmic component
1046	NCgl1918	5	5.7	ABC-type transporter: duplicated ATPase component
1047	NCgl1919	43	48.6	prolyl-tRNA synthetase [6.1.1.15]
1048	NCgl1920	1	10.6	hypothetical protein
1049	NCgl1922	1	3	Mg-chelatase subunit ChII
1050	NCgl1923	2	13.2	uroporphyrinogen-III methylase
1051	NCgl1924	9	18.6	hypothetical protein
1052	NCgl1925	1	5.2	hypothetical protein
1053	NCgl1926	36	46.8	predicted dehydrogenase [1.1.99.16]
1054	NCgl1927	1	4.7	hypothetical protein
1055	NCgl1928	11	24.5	dihydrodipicolinate dehydrogenase [1.6.4.2]
1056	NCgl1930	4	5.9	hypothetical membrane protein
1057	NCgl1932	16	31.6	methionine aminopeptidase [3.4.11.18]
1058	NCgl1933	1	1.6	cell division protein FtsI
1059	NCgl1934	3	16.7	two-component system: response regulator
1060	NCgl1938	21	41.2	gcpE-like protein
1061	NCgl1940	9	15.8	1-deoxy-D-xylulose 5-phosphate reductoisomerase [1.1.1.-]
1062	NCgl1941	4	21.1	hypothetical membrane protein
1063	NCgl1942	1	1.5	ABC-type transporter: permease component
1064	NCgl1943	4	15.1	ABC-type transporter: ATPase component
1065	NCgl1944	13	31.1	predicted Fe-S-cluster redox enzyme

1066	NCgl1947	9	24.9	ribosome recycling factor
1067	NCgl1948	30	46.1	uridylate kinase [2.7.4.-]
1068	NCgl1949	45	62.5	translation elongation factor Ts
1069	NCgl1950	80	47.8	ribosomal protein S2
1070	NCgl1952	1	2.9	integrase
1071	NCgl1953	2	7.6	predicted Rossmann-fold nucleotide-binding protein
1072	NCgl1956	6	44.6	hypothetical protein
1073	NCgl1958	1	6.5	signal peptidase I
1074	NCgl1959	1	5.1	ABC-type transport systems: periplasmic component
1075	NCgl1960	17	18.6	ribosomal protein L19
1076	NCgl1962	3	5.3	glycine/D-amino acid oxidases
1077	NCgl1966	23	21.1	predicted RNA binding protein
1078	NCgl1969	1	2.6	adenylosuccinate lyase [5.5.1.2]
1079	NCgl1972	9	15.4	tRNA-(guanine-N1)-methyltransferase [2.1.1.31]
1080	NCgl1973	1	8.3	hypothetical protein
1081	NCgl1974	8	38	RimM protein
1082	NCgl1975	2	13.8	hypothetical protein
1083	NCgl1976	19	38.8	ribosomal protein S16
1084	NCgl1977	1	7	ankyrin repeat containing protein
1085	NCgl1980	21	24.1	signal recognition particle GTPase
1086	NCgl1982	7	44.6	nitrogen regulatory protein PII
1087	NCgl1984	4	10	signal recognition particle GTPase
1088	NCgl1985	13	7.6	hypothetical protein
1089	NCgl1986	10	8.1	chromosome segregation ATPase
1090	NCgl1987	9	59.6	acylphosphatases [3.6.1.7]
1091	NCgl1988	2	2.2	hypothetical protein
1092	NCgl1993	1	7	formamidopyrimidine-DNA glycosylase [3.2.2.23]
1093	NCgl1994	7	22.7	dsRNA-specific ribonuclease [3.1.26.3]
1094	NCgl1995	5	17.4	hypothetical protein
1095	NCgl1996	20	30	hypothetical protein
1096	NCgl1997	2	6.8	ABC-type transporter: ATPase component and permease component
1097	NCgl1998	2	5.1	ABC-type transporter: ATPase component and permease component
1098	NCgl1999	117	62.4	glutamate dehydrogenase/leucine dehydrogenase [1.4.1.4]
1099	NCgl2000	3	2.7	glycerate kinase
1100	NCgl2001	17	59.8	hypothetical protein
1101	NCgl2002	1	2.5	hypothetical protein
1102	NCgl2003	3	6.2	metal-dependent amidase/aminoacylase/carboxypeptidase [3.5.1.32]
1103	NCgl2008	50	45.7	pyruvate kinase [2.7.1.40]
1104	NCgl2012	2	13.6	phosphoribosyl-AMP cyclohydrolase [3.5.4.19]
1105	NCgl2013	7	34.5	imidazoleglycerol-phosphate synthase and cyclase hisF
1106	NCgl2014	1	6.5	archaeal fructose-1:6-bisphosphatase
1107	NCgl2015	10	26	phosphoribosylformimino-5-aminoimidazole carboxamide ribonucleotide (ProF)
1108	NCgl2016	7	19.9	glutamine amidotransferase [2.4.2.-]
1109	NCgl2020	22	40.4	histidinol-phosphate aminotransferase/tyrosine aminotransferase [2.6.1.9]
1110	NCgl2021	22	23.8	histidinol dehydrogenase [1.1.1.23]
1111	NCgl2022	3	2.3	hypothetical protein
1112	NCgl2024	8	4.2	hypothetical protein
1113	NCgl2025	3	4.7	transcriptional regulator
1114	NCgl2026	26	22.2	pullulanase
1115	NCgl2027	4	7.1	SAM-dependent methyltransferase
1116	NCgl2030	2	6	transcriptional regulator
1117	NCgl2031	1	3.8	ABC-type transporter: ATPase component
1118	NCgl2035	6	15.7	DNA polymerase III epsilon subunit [2.7.7.7]
1119	NCgl2037	27	24.9	maltooligosyl trehalose synthase

1120	NCgl2038	8	19.4	hypothetical protein
1121	NCgl2041	2	11.2	coenzyme F420-dependent N5:N10-methylene tetrahydromethanopterin reductase
1122	NCgl2044	1	3.2	hypothetical protein
1123	NCgl2045	30	25.2	1:4-alpha-glucan branching enzyme
1124	NCgl2046	13	19.5	threonine dehydratase [4.2.1.16]
1125	NCgl2047	3	11.7	hypothetical membrane protein
1126	NCgl2048	33	29.7	methionine synthase II [2.1.1.14]
1127	NCgl2049	15	10.9	DNA polymerase III alpha subunit
1128	NCgl2053	16	41.7	dehydrogenase [1.-.-.]
1129	NCgl2054	3	10.7	diaminopimelate decarboxylase [4.1.1.20]
1130	NCgl2055	1	7.3	cysteine synthase [4.2.99.8]
1131	NCgl2057	5	19.7	23S RNA-specific pseudouridylate synthase [4.2.1.70]
1132	NCgl2059	2	12.3	hypothetical protein
1133	NCgl2061	2	9.5	hypothetical protein
1134	NCgl2062	1	8.3	L-asparaginase [3.5.1.1]
1135	NCgl2064	3	1.9	nucleotidyltransferase
1136	NCgl2066	1	3	predicted transcriptional regulator
1137	NCgl2067	3	10.8	hypothetical protein
1138	NCgl2068	56	34.1	isoleucyl-tRNA synthetase [6.1.1.5]
1139	NCgl2070	25	36.4	cell division initiation protein
1140	NCgl2071	2	11.6	hypothetical membrane protein
1141	NCgl2072	15	76.3	hypothetical protein
1142	NCgl2073	3	16.2	predicted enzyme with a TIM-barrel fold
1143	NCgl2074	6	25.6	hypothetical protein
1144	NCgl2075	31	41.2	cell division GTPase
1145	NCgl2076	1	7.2	cell division septal protein
1146	NCgl2077	31	38.9	UDP-N-acetylmuramate-alanine ligase [6.3.2.8]
1147	NCgl2078	7	16.3	UDP-N-acetylglucosamine:LPS N-acetylglucosamine transferase [2.4.1.-]
1148	NCgl2079	1	1.5	bacterial cell division membrane protein
1149	NCgl2080	9	14.4	UDP-N-acetylmuramoylalanine-D-glutamate ligase [6.3.2.9]
1150	NCgl2082	19	24.9	UDP-N-acetylmuramyl pentapeptide synthase [6.3.2.15]
1151	NCgl2083	14	17.3	UDP-N-acetylmuramyl tripeptide synthase [6.3.2.13]
1152	NCgl2084	1	1.8	cell division protein FtsI
1153	NCgl2085	2	8.3	hypothetical protein
1154	NCgl2086	17	33.5	predicted S-adenosylmethionine-dependent methyltransferase
1155	NCgl2091	12	32.2	5:10-methylenetetrahydrofolate reductase [1.7.99.5]
1156	NCgl2092	7	17.3	geranylgeranyl pyrophosphate synthase [2.5.1.1]
1157	NCgl2094	4	15.4	hypothetical protein
1158	NCgl2095	6	8.9	serine/threonine protein kinase
1159	NCgl2098	26	49.4	3-Deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase
1160	NCgl2099	2	8.3	hypothetical protein
1161	NCgl2103	1	4.1	hypothetical membrane protein
1162	NCgl2104	10	23.3	1-acyl-sn-glycerol-3-phosphate acyltransferase [2.3.1.51]
1163	NCgl2105	7	16.7	glucose kinase [2.7.1.2]
1164	NCgl2106	1	4.7	predicted glycosyltransferase
1165	NCgl2107	1	2.6	cell wall-associated hydrolase
1166	NCgl2109	9	9.8	cytochrome b subunit of the bc complex
1167	NCgl2110	10	23.8	Rieske Fe-S protein
1168	NCgl2111	2	5.1	cytochrome C
1169	NCgl2114	1	11.2	hypothetical membrane protein
1170	NCgl2115	9	15	cytochrome C oxidase subunit II [1.9.3.1]
1171	NCgl2116	41	38	asparagine synthase
1172	NCgl2117	7	36.8	hypothetical protein
1173	NCgl2120	11	19.6	NaMN:DMB phosphoribosyltransferase [2.4.2.21]

1174	NCgl2122	9	36.8	putative short-chain dehydrogenase
1175	NCgl2123	30	43.1	branched-chain amino acid aminotransferase [2.6.1.42]
1176	NCgl2124	6	11.4	leucyl aminopeptidase [3.4.11.1]
1177	NCgl2126	49	31	dihydrolipoamide acyltransferase [2.3.1.61]
1178	NCgl2127	6	25.1	lipoate-protein ligase B
1179	NCgl2128	15	33.6	lipoate synthase
1180	NCgl2129	2	11.5	hypothetical membrane protein
1181	NCgl2130	1	2.6	predicted permease
1182	NCgl2132	2	12.3	hypothetical membrane protein
1183	NCgl2133	63	33.1	glutamine synthase [6.3.1.2]
1184	NCgl2135	2	25.6	hypothetical protein
1185	NCgl2136	10	30.3	NTP pyrophosphohydrolase
1186	NCgl2139	42	43.5	threonine synthase [4.2.99.2]
1187	NCgl2143	1	7.4	hypothetical protein
1188	NCgl2145	7	20	hypothetical protein
1189	NCgl2146	9	27.4	hypothetical protein
1190	NCgl2147	11	9.5	glutamine synthetase adenylyltransferase [2.7.7.42]
1191	NCgl2148	12	22	glutamine synthase [6.3.1.2]
1192	NCgl2150	29	28.1	hypothetical protein
1193	NCgl2151	2	48.3	hypothetical protein
1194	NCgl2152	4	11.4	galactokinase [2.7.1.6]
1195	NCgl2153	10	19.8	exoribonucleases
1196	NCgl2155	9	18.3	phosphoglycerate mutase
1197	NCgl2156	10	34.3	Zn-ribbon protein
1198	NCgl2157	6	14.5	hypothetical protein
1199	NCgl2158	6	13.8	predicted phosphatase [3.1.3.18]
1200	NCgl2159	8	31.2	protein-tyrosine-phosphatase [3.1.3.48]
1201	NCgl2160	1	5	hypothetical protein
1202	NCgl2165	16	41.3	hypothetical protein
1203	NCgl2167	149	52.3	pyruvate dehydrogenase: decarboxylase component [1.2.4.1]
1204	NCgl2168	1	3.8	ABC-type transporter: ATPase component
1205	NCgl2170	6	6.1	hypothetical protein
1206	NCgl2172	4	30.6	hypothetical protein
1207	NCgl2174	3	35.1	acyl carrier protein
1208	NCgl2175	10	20	predicted sugar phosphatase of the HAD superfamily
1209	NCgl2176	3	7	hypothetical protein
1210	NCgl2178	7	18.2	beta-lactamase class C
1211	NCgl2186	21	26.1	hypothetical protein
1212	NCgl2188	4	5.4	bacterial type DNA primase [2.7.7.-]
1213	NCgl2190	2	16	hypothetical protein
1214	NCgl2191	36	32.9	glucosamine 6-phosphate synthetase [2.6.1.16]
1215	NCgl2193	3	11.3	dGTP triphosphohydrolase
1216	NCgl2194	6	16.9	hypothetical protein
1217	NCgl2195	7	7.3	chromosome segregation ATPase
1218	NCgl2198	47	46.6	class II glycyl-tRNA synthetase [6.1.1.14]
1219	NCgl2199	2	8.9	predicted transcriptional regulator
1220	NCgl2200	6	30.6	Fe ²⁺ /Zn ²⁺ uptake regulation protein
1221	NCgl2203	7	19.8	undecaprenyl pyrophosphate synthase [2.5.1.31]
1222	NCgl2205	11	23.9	GTPase
1223	NCgl2208	6	12	phosphate starvation-inducible protein PhoH
1224	NCgl2209	4	13.7	hypothetical protein
1225	NCgl2210	17	29.6	molecular chaperone
1226	NCgl2211	4	14.1	transcriptional regulator of heat shock gene
1227	NCgl2215	3	12.7	hypothetical protein

1228	NCgl2216	19	16.1	long-chain acyl-CoA synthetase (AMP-forming) [6.2.1.3]
1229	NCgl2217	57	31.9	4-alpha-glucanotransferase [2.4.1.25]
1230	NCgl2218	1	2.1	ABC-type transporter: ATPase component and permease component
1231	NCgl2219	23	25.1	Zn-dependent oligopeptidase [3.4.15.5]
1232	NCgl2221	9	15.7	putative trehalose synthase
1233	NCgl2222	4	10.2	hypothetical protein
1234	NCgl2223	2	9.7	isopentenylidiphosphate isomerase [5.3.3.2]
1235	NCgl2227	13	22.6	PLP-dependent aminotransferase [2.6.1.-]
1236	NCgl2229	2	12.3	coenzyme F420-dependent N5:N10-methylene tetrahydromethanopterin reductase
1237	NCgl2241	1	3.5	ABC-type transporter: duplicated ATPase component
1238	NCgl2243	1	4	sugar kinase [2.7.1.15]
1239	NCgl2246	26	46.8	hypothetical protein
1240	NCgl2247	16	12.2	malate synthase [4.1.3.2]
1241	NCgl2248	8	13.4	isocitrate lyase [4.1.3.1]
1242	NCgl2249	3	5.3	thiamine biosynthesis protein x
1243	NCgl2250	1	6.8	hypothetical protein
1244	NCgl2251	6	3.7	choline-glycine betaine transporter
1245	NCgl2253	2	7.5	hypothetical protein
1246	NCgl2259	28	38.4	membrane GTPase LepA
1247	NCgl2260	2	3.9	hypothetical protein
1248	NCgl2261	41	46	ribosomal protein S20
1249	NCgl2263	4	28.9	ankyrin repeat containing protein
1250	NCgl2267	5	12	hypothetical protein
1251	NCgl2268	7	17.8	phosphoglycerate mutase
1252	NCgl2269	3	15.9	hypothetical protein
1253	NCgl2270	3	6	nicotinic acid mononucleotide adenyltransferase
1254	NCgl2272	22	24.5	gamma-glutamyl phosphate reductase [1.2.1.41]
1255	NCgl2273	26	30.9	putative phosphoglycerate dehydrogenase
1256	NCgl2274	9	11.4	glutamate 5-kinase [2.7.2.11]
1257	NCgl2275	9	12.6	predicted GTPase
1258	NCgl2276	1	2	xanthine/uracil permease
1259	NCgl2277	28	34.9	aldo/keto reductase [1.1.1.-]
1260	NCgl2278	5	19.8	hypothetical protein
1261	NCgl2279	18	62.5	ribosomal protein L27
1262	NCgl2280	8	42.6	ribosomal protein L21
1263	NCgl2281	38	27.2	ribonuclease E [3.1.4.-]
1264	NCgl2282	1	4.1	hypothetical membrane protein
1265	NCgl2287	6	32.4	nucleoside diphosphate kinase [2.7.4.6]
1266	NCgl2289	9	55.3	predicted acetyltransferase
1267	NCgl2290	3	10.3	hypothetical protein
1268	NCgl2292	12	17.9	folylpolyglutamate synthase
1269	NCgl2293	50	32.1	valyl-tRNA synthetase [6.1.1.9]
1270	NCgl2294	4	6.9	ABC-type transport systems: periplasmic components
1271	NCgl2295	15	26	molecular chaperone
1272	NCgl2296	5	30.3	predicted Rossmann fold nucleotide-binding protein
1273	NCgl2297	12	24.7	malate/lactate dehydrogenase [1.1.1.37]
1274	NCgl2298	5	7.7	transcriptional regulator
1275	NCgl2299	2	6.8	predicted transcriptional regulator
1276	NCgl2304	19	19.7	ATP-dependent protease Clp: ATPase subunit
1277	NCgl2305	10	21.3	hypothetical protein
1278	NCgl2306	3	10	acyl-CoA:acetate CoA transferase beta subunit [2.8.3.6]
1279	NCgl2307	5	20.4	acyl-CoA:acetate CoA transferase alpha subunit [2.8.3.6]
1280	NCgl2308	2	5.1	transcriptional regulator
1281	NCgl2309	8	16.2	acetyl-CoA acetyltransferase

1282	NCgl2310	4	9.6	predicted hydrolase/acyltransferase [4.1.1.44]
1283	NCgl2311	5	5.2	DNA-binding HTH domain-containing protein
1284	NCgl2312	3	19.7	hypothetical protein
1285	NCgl2314	3	15.7	protocatechuate 3:4-dioxygenase beta subunit [1.13.11.3]
1286	NCgl2315	1	9.6	protocatechuate 3:4-dioxygenase beta subunit [1.13.11.3]
1287	NCgl2318	2	3.2	putative chloromuconate cycloisomerase
1288	NCgl2319	4	7.7	protocatechuate 3:4-dioxygenase beta subunit [1.13.11.1]
1289	NCgl2321	1	10.4	benzoate dioxygenase small subunit
1290	NCgl2324	1	1.3	DNA-binding HTH domain containing protein
1291	NCgl2325	1	2.7	putative benzoate transporter
1292	NCgl2327	7	22.6	ATP-dependent Clp protease proteolytic subunit 2 [3.4.21.92]
1293	NCgl2328	24	52.3	ATP-dependent Clp protease proteolytic subunit 1 [3.4.21.92]
1294	NCgl2329	26	31.6	FKBP-type peptidyl-prolyl cis-trans isomerase
1295	NCgl2333	4	30.1	hypothetical protein
1296	NCgl2336	6	16.6	hypothetical protein
1297	NCgl2337	10	29.9	ribose 5-phosphate isomerase RpiB [5.3.1.26]
1298	NCgl2339	9	29.6	hypothetical protein
1299	NCgl2340	61	37	aminopeptidase N [3.4.11.2]
1300	NCgl2342	3	7.8	hypothetical protein
1301	NCgl2343	2	4.5	hypothetical protein
1302	NCgl2347	1	3.6	phytoene/squalene synthetase [2.5.1.-]
1303	NCgl2349	22	30.9	hypothetical protein
1304	NCgl2350	2	6.9	ABC-type transporter: duplicated ATPase component
1305	NCgl2353	6	6	ABC-type transporter: periplasmic component
1306	NCgl2354	12	16.2	hypothetical protein
1307	NCgl2355	13	23.5	PLP-dependent aminotransferase [2.6.1.11]
1308	NCgl2358	4	14.8	oxidoreductase [1.1.1.36]
1309	NCgl2359	10	28	transcriptional regulator
1310	NCgl2360	30	34.5	cystathionine gamma-synthase [4.2.99.9]
1311	NCgl2361	7	14.4	ABC-type transporter: ATPase component
1312	NCgl2362	5	32.8	hemoglobin-like protein
1313	NCgl2364	2	13.4	hypothetical protein
1314	NCgl2365	13	63.9	predicted thioesterase
1315	NCgl2368	49	33.6	ABC-type transporter: duplicated ATPase component
1316	NCgl2373	2	3.9	ABC-type transporter: permease component
1317	NCgl2374	1	3.2	ABC-type transporter: permease component
1318	NCgl2375	64	57.6	ABC-type transporter: periplasmic component
1319	NCgl2376	10	22.7	hypothetical protein
1320	NCgl2377	43	51.9	ABC-type transporter: ATPase component
1321	NCgl2383	8	13.2	hypothetical protein
1322	NCgl2384	4	15.6	hydroxypyruvate isomerase
1323	NCgl2385	16	27.8	short-chain dehydrogenase
1324	NCgl2386	5		oligoribonuclease
1325	NCgl2387	3	3	hypothetical protein
1326	NCgl2388	5	5	hypothetical protein
1327	NCgl2395	4	4.6	glutaminase [3.5.1.2]
1328	NCgl2396	2	17.9	transcriptional regulator
1329	NCgl2401	5	12.4	amidase
1330	NCgl2402	2	22	hypothetical protein
1331	NCgl2403	10	46.2	bacterioferritin comigratory protein
1332	NCgl2404	4	10.4	transcriptional regulator
1333	NCgl2409	144	28.8	3-oxoacyl-(acyl-carrier-protein) synthase [2.3.1.85]
1334	NCgl2414	10	17.5	xanthosine triphosphate pyrophosphatase
1335	NCgl2415	12	38.4	RNase PH [2.7.7.56]

1336	NCgl2422	9	27.1	metal-dependent hydrolase [3.1.6.1]
1337	NCgl2423	7	8.5	glutamate racemase [5.1.1.3]
1338	NCgl2424	2	6.3	hypothetical protein
1339	NCgl2425	3	12.8	transcriptional regulator
1340	NCgl2428	13	31.8	hypothetical protein
1341	NCgl2429	3	22.6	hypothetical protein
1342	NCgl2430	4	25.1	hypothetical protein
1343	NCgl2431	30	45.3	nicotinic acid phosphoribosyltransferase
1344	NCgl2434	1	4.4	hypothetical membrane protein
1345	NCgl2435	1	6	hypothetical protein
1346	NCgl2436	29	44.1	phosphoserine phosphatase [3.1.3.3]
1347	NCgl2437	13	13.4	heme/copper-type cytochrome/quinol oxidase: subunit 1 [1.9.3.1]
1348	NCgl2438	12	18.6	ribonucleotide reductase beta subunit
1349	NCgl2439	9	34	ferritin-like protein
1350	NCgl2440	8	28.4	transcriptional regulator
1351	NCgl2443	39	35.7	ribonucleotide reductase alpha subunit
1352	NCgl2444	1	14.2	ribonucleotide reduction protein
1353	NCgl2446	17	26	NAD synthase [6.3.5.1]
1354	NCgl2448	9	45.3	hypothetical protein
1355	NCgl2449	12	25	putative Zn-NADPH:quinone dehydrogenase [1.1.1.1]
1356	NCgl2450	4	8.8	hypothetical protein
1357	NCgl2452	3	12.5	hypothetical protein
1358	NCgl2453	29	39.9	phosphoglucomutase [5.4.2.2]
1359	NCgl2456	2	11.7	hypothetical protein
1360	NCgl2464	1	1.3	ABC-type transporter: permease component
1361	NCgl2465	7	27.1	ABC-type transporter: ATPase component
1362	NCgl2466	2	14.2	hypothetical protein
1363	NCgl2470	29	27.3	UDP-N-acetylglucosamine enolpyruvyl transferase [2.5.1.7]
1364	NCgl2471	9	32.6	hypothetical protein
1365	NCgl2472	28	44.8	regulatory-like protein
1366	NCgl2473	49	72.3	cysteine synthase [4.2.99.8]
1367	NCgl2474	12	30.2	serine acetyltransferase [2.3.1.30]
1368	NCgl2475	14	61.5	predicted acetyltransferase
1369	NCgl2476	8	17.7	succinyl-CoA synthetase alpha subunit [6.2.1.5]
1370	NCgl2477	2	2.3	succinyl-CoA synthetase beta subunit [6.2.1.5]
1371	NCgl2478	6	13.9	predicted dithiol-disulfide isomerase
1372	NCgl2480	108	46.2	acetyl-CoA hydrolase [2.8.3.-]
1373	NCgl2481	11	25.5	predicted TIM-barrel containing enzyme
1374	NCgl2482	18	39.3	phosphate uptake regulator
1375	NCgl2486	1	5.6	ABC-type transporter: periplasmic component
1376	NCgl2487	11	40.8	histone acetyltransferase HPA2-like protein
1377	NCgl2489	3	4.3	hypothetical protein
1378	NCgl2490	12	24.5	hypothetical protein
1379	NCgl2491	3	15.9	branched-chain amino acid aminotransferase [2.6.1.42]
1380	NCgl2492	21	36.2	predicted aminomethyltransferase related to GcvT
1381	NCgl2493	3	18.3	hypothetical protein
1382	NCgl2494	28	38	phosphoribosylaminoimidazol (AIR) synthetase [6.3.3.1]
1383	NCgl2495	14	26.2	glutamine phosphoribosylpyrophosphate amidotransferase
1384	NCgl2496	3	20.8	hypothetical protein
1385	NCgl2497	16	24.2	acyl-CoA hydrolase
1386	NCgl2498	1	3.1	predicted membrane proteins
1387	NCgl2499	54	42.5	phosphoribosylformylglycinamide (FGAM) synthase: synthetase domain [6.3.
1388	NCgl2500	38	55.2	phosphoribosylformylglycinamide (FGAM) synthase: glutamine amidotransfer
1389	NCgl2501	8	48.1	phosphoribosylformylglycinamide (FGAM) synthase: PurS component

1390	NCgl2502	7	20.8	glutathione peroxidase [1.11.1.9]
1391	NCgl2507	24	24.6	protease II
1392	NCgl2508	20	33.7	phosphoribosylaminoimidazolesuccinocarboxamide (SAICAR) synthase [6.3.2.2]
1393	NCgl2509	45	41.6	adenylosuccinate lyase [4.3.2.2]
1394	NCgl2510	10	28.8	PLP-dependent aminotransferase [2.6.1.1]
1395	NCgl2511	21	40.9	phosphoribosylamine-glycine ligase [6.3.4.13]
1396	NCgl2512	5	22.8	diadenosine tetraphosphate (Ap4A) hydrolase
1397	NCgl2517	1	2.3	two-component system: sensory transduction histidine kinase
1398	NCgl2518	11	20	two-component system: response regulator
1399	NCgl2520	4	6.7	hypothetical protein
1400	NCgl2521	39	26.1	thiamine pyrophosphate-requiring enzyme [1.2.2.2]
1401	NCgl2522	1	1.4	permease of the major facilitator superfamily
1402	NCgl2525	4	4.9	hypothetical membrane protein
1403	NCgl2526	1	5.6	succinate dehydrogenase/fumarate reductase: flavoprotein subunit
1404	NCgl2528	16	20.6	D-2-hydroxyisocaproate dehydrogenase
1405	NCgl2529	1	4.1	hypothetical protein
1406	NCgl2530	24	37.6	predicted hydrolase of the HAD superfamily
1407	NCgl2531	4	25.5	hypothetical protein
1408	NCgl2533	1	2	hypothetical protein
1409	NCgl2534	3	21.1	hypothetical protein
1410	NCgl2537	7	30.1	trehalose-6-phosphatase [3.1.3.12]
1411	NCgl2538	4	9.7	transcriptional regulator
1412	NCgl2540	1	6.1	ABC-type transporter: ATPase component
1413	NCgl2544	1	2.8	3-ketosteroid 1-dehydrogenase
1414	NCgl2547	1	7.6	hypothetical protein
1415	NCgl2550	3	20.6	transcriptional regulator
1416	NCgl2551	10	19.8	rRNA methylase [2.1.1.-]
1417	NCgl2552	13	20	cysteinyl-tRNA synthetase [6.1.1.16]
1418	NCgl2553	4	6.8	phosphotransferase system IIC component [2.7.1.69]
1419	NCgl2554	23	28.6	beta-fructosidase [3.2.1.26]
1420	NCgl2555	9	24.9	glucosamine-6-phosphate isomerase [5.3.1.10]
1421	NCgl2558	3	10.2	transcriptional regulator [2.7.1.2]
1422	NCgl2559	1	7.3	putative N-acetylmannosamine-6-phosphate epimerase
1423	NCgl2561	4	17.5	transcriptional regulator
1424	NCgl2562	6	15.8	ABC-type transporter: periplasmic component
1425	NCgl2564	2	5.6	ABC-type transporter: ATPase component
1426	NCgl2569	1	6.3	2C-methyl-D-erythritol 2:4-cyclodiphosphate synthase
1427	NCgl2570	1	4.7	4-diphosphocytidyl-2-methyl-D-erythritol synthase
1428	NCgl2571	4	18.2	transcriptional regulator
1429	NCgl2572	5	17	two-component system: response regulator
1430	NCgl2574	1	7.7	hypothetical protein
1431	NCgl2575	4	14	DNA repair protein
1432	NCgl2576	5	14.2	predicted nucleic-acid-binding protein
1433	NCgl2578	24	38.8	NAD-dependent aldehyde dehydrogenase [1.2.1.-]
1434	NCgl2579	16	32.9	carbonic anhydrase [4.2.1.1]
1435	NCgl2580	1	3.1	A/G-specific DNA glycosylase
1436	NCgl2582	21	51.2	L-2,3-butanediol dehydrogenase
1437	NCgl2584	6	33	hypothetical protein
1438	NCgl2585	49	32	ATPase with chaperone activity: ATP-binding subunit
1439	NCgl2586	23	36.3	inositol-monophosphate dehydrogenase
1440	NCgl2587	2	3.3	AraC-type DNA-binding domain-containing protein
1441	NCgl2588	3	5.4	2-polyprenyl-6-methoxyphenol hydroxylase [1.14.13.7]
1442	NCgl2589	4	7.3	hypothetical protein
1443	NCgl2590	3	6.2	hypothetical protein

1444	NCgl2591	5	8.5	hypothetical protein
1445	NCgl2594	27	23.6	lysyl-tRNA synthetase class II [6.1.1.6]
1446	NCgl2596	2	7.8	hypothetical protein
1447	NCgl2597	1	3.4	hypothetical membrane protein
1448	NCgl2599	2	13.8	7:8-dihydro-6-hydroxymethylpterin-pyrophosphokinase [2.7.6.3]
1449	NCgl2601	2	11.1	dihydropteroate synthase [2.5.1.15]
1450	NCgl2602	8	26.9	GTP cyclohydrolase I [3.5.4.16]
1451	NCgl2604	6	33.5	hypoxanthine-guanine phosphoribosyltransferase
1452	NCgl2606	6	9.1	D-alanyl-D-alanine carboxypeptidase [3.4.16.4]
1453	NCgl2607	23	46.8	inorganic pyrophosphatase [3.6.1.1]
1454	NCgl2611	3	21.1	hypothetical protein
1455	NCgl2615	4	11.9	uncharacterized NAD(FAD)-dependent dehydrogenase
1456	NCgl2616	14	69.8	rhodanese-related sulfurtransferase
1457	NCgl2617	1	13.5	transcriptional regulator
1458	NCgl2620	26	33.3	hypothetical protein
1459	NCgl2621	57	33.9	chaperonin GroEL
1460	NCgl2623	5	7.2	hypothetical protein
1461	NCgl2624	5	9.7	hypothetical protein
1462	NCgl2625	5	5.7	hypothetical protein
1463	NCgl2626	14	13.7	hypothetical protein
1464	NCgl2627	5	2.1	hypothetical protein
1465	NCgl2628	52	28.5	hypothetical protein
1466	NCgl2629	3	5.8	hypothetical membrane protein
1467	NCgl2631	14	16.4	hypothetical protein
1468	NCgl2638	1	9.5	multisubunit Na ⁺ /H ⁺ antiporter
1469	NCgl2639	1	4	hypothetical protein
1470	NCgl2640	10	17	hypothetical protein
1471	NCgl2643	2	7.8	N-formylmethionyl-tRNA deformylase [3.5.1.31]
1472	NCgl2644	5	13.1	histone acetyltransferase HPA2-like protein
1473	NCgl2645	10	25.1	exonuclease III [3.1.11.2]
1474	NCgl2649	9	43.9	predicted epimerase
1475	NCgl2651	1	6.8	ABC-type transporter: ATPase component
1476	NCgl2652	1	10.2	NTP pyrophosphohydrolase
1477	NCgl2656	16	33	acetate kinase [2.7.2.1]
1478	NCgl2657	16	31.5	phosphotransacetylase [2.3.1.8]
1479	NCgl2658	38	46.6	putative ferredoxin/ferredoxin-NADP reductase [1.18.1.2]
1480	NCgl2659	4	22.5	predicted acyltransferase
1481	NCgl2663	13	30.4	phosphoribosylglycinamide formyltransferase 2
1482	NCgl2667	1	2.1	two-component system: sensory transduction histidine kinases
1483	NCgl2668	1	3.9	two-component system: response regulator
1484	NCgl2669	23	28.4	adenylosuccinate synthase [6.3.4.4]
1485	NCgl2670	3	16.6	hypothetical protein
1486	NCgl2673	87	59.6	fructose-bisphosphate aldolase [4.1.2.13]
1487	NCgl2674	1	5.9	hypothetical protein
1488	NCgl2675	2	8.7	rRNA methylase
1489	NCgl2676	12	38	orotate phosphoribosyltransferase [2.4.2.10]
1490	NCgl2677	6	12.5	hypothetical protein
1491	NCgl2678	1	3.3	thiosulfate sulfurtransferase
1492	NCgl2682	39	25.5	ATPase with chaperone activity: ATP-binding subunit
1493	NCgl2684	2	14.7	predicted transcriptional regulator
1494	NCgl2687	5	10.1	alkanal monooxygenase
1495	NCgl2692	4	8	hypothetical protein
1496	NCgl2694	3	8.8	putative 2-polyprenylphenol hydroxylase
1497	NCgl2695	9	27.1	putative amidohydrolase

1498	NCgl2698	19	34.6	NAD-dependent aldehyde dehydrogenase [1.2.1.3]
1499	NCgl2699	2	14.4	predicted transcriptional regulator
1500	NCgl2700	11	15.9	molecular chaperone
1501	NCgl2701	9	8.3	molecular chaperone GrpE
1502	NCgl2702	83	48.1	70 kDa heat shock chaperonin protein
1503	NCgl2703	1	4.4	predicted permease
1504	NCgl2706	5	6.5	hypothetical protein
1505	NCgl2709	11	19.7	Zn-dependent alcohol dehydrogenase [1.1.1.1]
1506	NCgl2715	26	48.5	sulfate adenylate transferase subunit 1
1507	NCgl2716	27	44.4	sulfate adenylyltransferase subunit 2
1508	NCgl2717	3	9.5	phosphoadenosine phosphosulfate reductase
1509	NCgl2718	45	34.4	putative nitrite reductase
1510	NCgl2719	59	57.3	putative ferredoxin/ferredoxin-NADP reductase
1511	NCgl2720	11	15.9	hypothetical protein
1512	NCgl2722	5	29	hypothetical protein
1513	NCgl2730	15	32.2	putative peptidase
1514	NCgl2737	25	36.9	putative membrane protease subunit
1515	NCgl2739	3	11.2	3-methyladenine DNA glycosylase [3.2.2.20]
1516	NCgl2741	4	14.9	hypothetical membrane protein
1517	NCgl2747	22	38.2	PLP-dependent aminotransferase [2.6.1.1]
1518	NCgl2750	7	11.8	predicted UDP-glucose 6-dehydrogenase [1.1.1.22]
1519	NCgl2751	9	38.6	deoxycytidine deaminase [3.5.4.13]
1520	NCgl2754	1	5.1	beta-N-acetylglucosaminidase-like protein
1521	NCgl2755	5	31.8	hypothetical protein
1522	NCgl2757	1	1.6	hypothetical membrane protein
1523	NCgl2759	1	3.4	predicted acyltransferase
1524	NCgl2763	4	22.1	hypothetical protein
1525	NCgl2765	36	35.2	phosphoenolpyruvate carboxykinase (GTP) [4.1.1.32]
1526	NCgl2767	10	19.6	predicted S-adenosylmethionine-dependent methyltransferase
1527	NCgl2768	5	19.3	hypothetical protein
1528	NCgl2772	17	25	acetyl-CoA carboxylase beta subunit [6.4.1.3]
1529	NCgl2773	57	24.7	putative polyketide synthase
1530	NCgl2774	27	21.1	putative acyl-CoA synthetase
1531	NCgl2775	3	7.8	hypothetical protein
1532	NCgl2776	3	21.1	hypothetical protein
1533	NCgl2777	19	19.8	putative esterase
1534	NCgl2779	6	7.9	putative esterase
1535	NCgl2780	1	1.6	hypothetical membrane protein
1536	NCgl2782	4	36.9	membrane-associated phospholipid phosphatase
1537	NCgl2783	34	30.2	hypothetical protein
1538	NCgl2784	5	14.2	hypothetical membrane protein
1539	NCgl2787	21	20.8	predicted flavoprotein involved in K ⁺ transport
1540	NCgl2788	21	28.4	UDP-galactopyranose mutase [5.4.99.9]
1541	NCgl2789	8	9.5	hypothetical protein
1542	NCgl2790	15	27.7	glycerol kinase [2.7.1.30]
1543	NCgl2792	3	10.3	1-acyl-sn-glycerol-3-phosphate acyltransferase
1544	NCgl2793	21	43.8	seryl-tRNA synthetase [6.1.1.11]
1545	NCgl2794	11	38.7	transcriptional regulator
1546	NCgl2795	7	20.8	hypothetical protein
1547	NCgl2796	2	15.8	hypothetical protein
1548	NCgl2798	3	17.5	putative phosphoglycerate mutase
1549	NCgl2799	9	18.7	prephenate dehydratase [4.2.1.51]
1550	NCgl2804	8	51.4	hypothetical protein
1551	NCgl2805	1	3.3	hypothetical protein

1552	NCgl2806	12	17.6	hypothetical protein
1553	NCgl2808	2	3.7	putative gluconate permease
1554	NCgl2809	11	14.1	pyruvate kinase-like protein
1555	NCgl2810	26	35	L-lactate dehydrogenase [1.1.1.27]
1556	NCgl2811	2	2.5	hypothetical protein
1557	NCgl2812	3	8	predicted hydrolase of the HAD superfamily
1558	NCgl2813	12	32	predicted flavoprotein
1559	NCgl2814	4	13.9	transcriptional regulator
1560	NCgl2816	1	2.8	putative integral membrane transport protein
1561	NCgl2817	19	25.5	L-lactate dehydrogenase
1562	NCgl2825	10	31.3	peptide methionine sulfoxide reductase [1.8.4.6]
1563	NCgl2826	36	40.5	superoxide dismutase [1.15.1.1]
1564	NCgl2830	3	3.8	hypothetical protein
1565	NCgl2831	5	14.7	hypothetical protein
1566	NCgl2833	1	9.2	transcriptional regulator
1567	NCgl2834	10	19.8	two-component system: response regulator
1568	NCgl2839	3	5.2	hypothetical membrane protein
1569	NCgl2840	16	39.4	transcriptional regulator
1570	NCgl2842	9	25.9	hypothetical protein
1571	NCgl2843	1	5.2	alkanal monooxygenase-like protein
1572	NCgl2844	2	6.8	23S RNA-specific pseudouridylate synthase
1573	NCgl2847	21	25.6	hypothetical protein
1574	NCgl2848	25	33.7	hypothetical protein
1575	NCgl2857	3	31.5	hypothetical protein
1576	NCgl2863	4	10.4	two-component system: response regulator
1577	NCgl2874	8	37.1	thioredoxin
1578	NCgl2875	1	33.8	copper chaperone
1579	NCgl2876	3	8.9	putative transmembrane transport protein
1580	NCgl2878	7	11.4	replicative DNA helicase [3.6.1.-]
1581	NCgl2879	14	29.3	ribosomal protein L9
1582	NCgl2880	3	16.9	single-stranded DNA-binding protein
1583	NCgl2881	11	50.5	ribosomal protein S6
1584	NCgl2883	1	2.7	hypothetical membrane protein
1585	NCgl2884	4	6.3	penicillin-binding protein
1586	NCgl2885	3	23.8	hypothetical protein
1587	NCgl2886	13	41.4	transcriptional regulator
1588	NCgl2887	5	15	hypothetical protein
1589	NCgl2888	1	10.9	hypothetical membrane protein
1590	NCgl2889	7	54.2	hypothetical protein
1591	NCgl2890	16	49.4	hypothetical protein
1592	NCgl2892	2	10.4	ABC-type transporter: ATPase component
1593	NCgl2893	3	1.5	efflux system protein
1594	NCgl2894	72	64.7	myo-inositol-1-phosphate synthase
1595	NCgl2896	23	30.1	hypothetical protein
1596	NCgl2897	13	47.3	starvation-inducible DNA-binding protein
1597	NCgl2898	6	12.2	formamidopyrimidine-DNA glycosylase [3.2.2.23]
1598	NCgl2899	1	3.9	hypothetical protein
1599	NCgl2901	3	20.9	methylated DNA-protein cysteine methyltransferase [2.1.1.63]
1600	NCgl2902	3	8	hypothetical protein
1601	NCgl2904	20	32.7	malic enzyme [1.1.1.40]
1602	NCgl2905	11	13.2	sugar kinase [2.7.1.12]
1603	NCgl2908	1	2.5	putative mercuric reductase [1.16.1.1]
1604	NCgl2909	10	12.4	D-amino acid dehydrogenase subunit
1605	NCgl2910	1	2.8	hypothetical protein

1606	NCgl2913	2	4.4	nitroreductase
1607	NCgl2915	56	28.5	leucyl-tRNA synthetase
1608	NCgl2918	1	3.7	hypothetical protein
1609	NCgl2919	6	11.8	2-hydroxyhepta-2:4-diene-1:7-dioatesomerase
1610	NCgl2923	1	2.3	putative hydroxylase/monooxygenase
1611	NCgl2927	18	26.1	anthranilate synthase component I [4.1.3.27]
1612	NCgl2929	20	36.8	anthranilate phosphoribosyltransferase [2.4.2.18]
1613	NCgl2930	25	30.6	indole-3-glycerol phosphate synthase [5.3.1.24]
1614	NCgl2931	8	18.5	tryptophan synthase beta chain [4.2.1.20]
1615	NCgl2932	12	31.8	tryptophan synthase alpha chain [4.2.1.20]
1616	NCgl2934	2	7.8	hypothetical protein
1617	NCgl2938	6	16.5	NADH:flavin oxidoreductase
1618	NCgl2940	2	6.2	hypothetical protein
1619	NCgl2941	1	9.8	predicted transcriptional regulator
1620	NCgl2942	6	4.8	NADH:flavin oxidoreductase
1621	NCgl2943	10	41.2	hypothetical protein
1622	NCgl2950	3	7.6	transcriptional regulator
1623	NCgl2951	2	10.3	putative hydroxyquinol/catechol 1:2-dioxygenase
1624	NCgl2952	1	4.4	maleylacetate reductase
1625	NCgl2954	2	9.1	transcriptional regulator
1626	NCgl2956	1	5.1	hypothetical protein
1627	NCgl2958	2	10.2	putative dehydrogenase
1628	NCgl2961	1	5	putative proline-betaine transporter
1629	NCgl2964	10	6.7	putative helicase
1630	NCgl2970	11	22.5	ABC-type transport systems: periplasmic component
1631	NCgl2971	3	6.2	putative oxidoreductase/dehydrogenase
1632	NCgl2973	20	47	hydroxymethylpyrimidine/phosphomethylpyrimidine kinase
1633	NCgl2975	2	25.4	putative copper chaperone
1634	NCgl2976	2	11.3	hypothetical membrane protein
1635	NCgl2978	2	7	hypothetical protein
1636	NCgl2979	24	26.1	putative polynucleotide polymerase
1637	NCgl2980	9	5	hypothetical protein
1638	NCgl2982	2	2.7	putative virulence factor
1639	NCgl2984	26	48.6	thioredoxin reductase [1.6.4.5]
1640	NCgl2985	7	35.5	thioredoxin
1641	NCgl2986	11	16.4	N-acetylmuramoyl-L-alanine amidase [3.5.1.28]
1642	NCgl2988	13	14.8	putative cell division protein ParB
1643	NCgl2989	11	29.6	putative cell division protein ParA
1644	NCgl2990	10	25.1	glucose-inhibited division protein B

Appendix III

Table 3. List of the DNA-binding transcriptional regulators identified by MALDI TOF/TOF MS and database search after separation with two-dimensional reversed phase chromatography at high and low pH. The list is based on the work of Brinkrolf et al. (2007).

Category	NCgl no.	Description of regulatory function
Carbohydrate metabolism	NCgl0110	Repressor of xylulose kinase gene
	NCgl0154	Repressor of iositol metabolism
	NCgl0286	Repressor of acetate/gluconate metabolism
	NCgl0358	Repressor of acetate metabolism
	NCgl1203*	Repressor of ribose uptake
	NCgl1483	Repressor of aconitate gene
	NCgl1856	Repressor in fructose PTS gene region
	NCgl1859*	Repressor of fructose PTS system
	NCgl2299	Repressor of vanillate metabolism
	NCgl2308*	Repressor of protocatechuate degradation
	NCgl2311	Activator of aromatic compound degradation
	NCgl2324*	Activator of aromatic compound degradation
	NCgl2472	Activator of acetate metabolism and <i>cspB</i> gene
	NCgl2587*	Activator of phenol degradation
	NCgl2954	Activator of hydroxyquinol pathway
Specific biosynthesis pathways and transport systems	NCgl0253*	Activator of amino acid export system BrnEF
	NCgl0368*	Repressor of MFS-type transporters
	NCgl0886	Repressor of MmpL-type transporter
	NCgl1215	Activator of amino acid transport system LysE
	NCgl1345	Repressor of arginine biosynthesis
	NCgl1551	Repressor of <i>pyrH</i> and <i>pyrBC</i> genes
	NCgl1578*	Repressor of genes cg1844-cg1847
NCgl1853*	Repressor of ribonucleotid	

Macroelement and metal homeostasis	NCgl0120	Activator of assimilatory sulphate reduction
	NCgl0829	Repressor exerting global nitrogen control
	NCgl1845	Dual regulator of iron metabolism
	NCgl2199*	Repressor of <i>furB</i> regulatory gene
	NCgl2684	Repressor of cg3083–cg3085 genes
SOS and stress response	NCgl2840	Repressor of sulphur metabolism
	NCgl1504*	Repressor of iron-sulfur cluster biogenesis
	NCgl1855	Repressor of SOS response
	NCgl2211	Repressor of heat shock response
	NCgl2699	Repressor of heat shock response

*Identified by one non-redundant peptide

Appendix IV

Table 4. Integral membrane proteins identified in the present by MALDI TOF/TOF mass spectrometry and database search after separation with two-dimensional reversed phase chromatography at high and low pH. The number of transmembrane helices (TMH) of the proteins was obtained from Fischer et al.

Nr. of TMH	Locus ID	Protein	No. of hits	No. of different peptides
1	NCgl0032	hypothetical protein	20	7
1	NCgl0044	protein serine/threonine phosphatase [3.1.3.16]	1	1
1	NCgl0070	hypothetical protein	6	3
1	NCgl0195	predicted glycosyltransferases	4	3
1	NCgl0321	predicted glycosyltransferase	2	2
1	NCgl0400	phosphoserine phosphatase [3.1.3.3]	3	1
1	NCgl0576	hypothetical protein	1	1
1	NCgl0603	predicted nucleoside-diphosphate-sugar epimerase	11	5
1	NCgl0677	detergent sensitivity rescuer dtsR2	51	15
1	NCgl0678	detergent sensitivity rescuer dtsR1	13	8
1	NCgl0841	trypsin-like serine protease [3.4.21.-]	5	2
1	NCgl1360	hypothetical protein	8	5
1	NCgl1409	NADH dehydrogenase: FAD-containing subunit [1.6.99.3]	70	13
1	NCgl2095	serine/threonine protein kinase	6	5
1	NCgl2402	hypothetical protein	2	2
2	NCgl0060	Mg ²⁺ and Co ²⁺ transporter protein	1	1
2	NCgl0124	hypothetical membrane protein	9	7
2	NCgl0293	hypothetical membrane protein	3	2
2	NCgl0337	putative ATPase involved in chromosome partitioning	10	5
2	NCgl0365	uncharacterized membrane protein	1	1
2	NCgl0722	two-component system sensory transduction histidine kinase	1	1
2	NCgl1079	hypothetical membrane protein	4	2
2	NCgl1081	hypothetical membrane protein	1	1
2	NCgl1160	F ₀ F ₁ -type ATP synthase c subunit [3.6.1.34]	13	1
2	NCgl1201	hypothetical membrane protein	1	1
2	NCgl1389	hypothetical membrane protein	1	1
2	NCgl1831	hypothetical membrane protein	3	2
2	NCgl1930	hypothetical membrane protein	4	1
2	NCgl1941	hypothetical membrane protein	4	2
2	NCgl1988	hypothetical protein	2	1
2	NCgl2071	hypothetical membrane protein	2	1
2	NCgl2111	cytochrome C	2	1
2	NCgl2129	hypothetical membrane protein	2	2
2	NCgl2517	two-component system: sensory transduction histidine kinase	1	1
2	NCgl2597	hypothetical membrane protein	1	1
2	NCgl2627	hypothetical protein	5	2

2	NCgl2629	hypothetical membrane protein	3	3
2	NCgl2737	putative membrane protease subunit	25	8
2	NCgl2741	hypothetical membrane protein	4	2
3	NCgl0058	hypothetical membrane protein	1	1
3	NCgl0649	hypothetical membrane protein	1	1
3	NCgl0743	predicted NAD-binding component of Kef-type K ⁺ transport system	4	3
3	NCgl1042	hypothetical membrane protein	2	1
3	NCgl1086	hypothetical membrane protein	3	2
3	NCgl2110	Rieske Fe-S protein	10	6
3	NCgl2114	hypothetical membrane protein	1	1
3	NCgl2132	hypothetical membrane protein	2	1
3	NCgl2282	hypothetical membrane protein	1	1
3	NCgl2638	multisubunit Na ⁺ /H ⁺ antiporter	1	1
3	NCgl2888	hypothetical membrane protein	1	1
4	NCgl0221	Mg/Co/Ni transporter MgtE	2	2
4	NCgl0249	hypothetical membrane protein	1	1
4	NCgl0769	cell division protein	3	2
4	NCgl0815	hypothetical membrane-associated protein	1	1
4	NCgl0914	putative ABC transporter ATPase component	1	1
4	NCgl1035	ABC-type transporter: permease component	1	1
4	NCgl1043	hypothetical protein	3	2
4	NCgl1114	hypothetical membrane protein	1	1
4	NCgl2434	hypothetical membrane protein	1	1
5	NCgl0340	predicted nucleoside-diphosphate sugar epimerase	7	5
5	NCgl0585	two-component system sensor kinase	1	1
5	NCgl0608	ABC-type transporter: permease component	2	1
5	NCgl0925	ABC-type transporter: ATP-binding component	1	1
5	NCgl1218	hypothetical membrane protein	3	2
5	NCgl1893	DNA translocase spoIIIE-like protein	5	2
5	NCgl1997	ABC-type transporter: ATPase component and permease component	2	2
5	NCgl1998	ABC-type transporter: ATPase component and permease component	2	2
5	NCgl2564	ABC-type transporter: ATPase component	2	1
5	NCgl2839	hypothetical membrane protein	3	1
6	NCgl0648	multiple transmembrane domain containing protein	2	2
6	NCgl1159	F ₀ F ₁ -type ATP synthase a subunit [3.6.1.34]	3	2
6	NCgl1576	predicted membrane protein	2	2
6	NCgl1594	preprotein translocase subunit SecD	2	1
6	NCgl1877	glutamate ABC-type transporter: permease component	5	3
6	NCgl2373	ABC-type transporter: permease component	2	1
6	NCgl2374	ABC-type transporter: permease component	1	1
7	NCgl0445	phosphate/sulphate permease	1	1
7	NCgl0752	hypothetical protein	2	2
7	NCgl2498	predicted membrane proteins	1	1
8	NCgl0375	cation transport ATPase [3.6.1.-]	4	2

8	NCgl1739	hypothetical protein	2	1
8	NCgl2079	bacterial cell division membrane protein	1	1
8	NCgl2703	predicted permease	1	1
9	NCgl0551	hypothetical membrane protein	1	1
9	NCgl0565	putative membrane protein	1	1
9	NCgl0927	hypothetical protein	1	1
9	NCgl0994	GGDEF family protein	16	2
9	NCgl1488	cation transport ATPase [3.6.1.-]	3	1
9	NCgl1861	phosphotransferase system: fructose-specific IIC component [2.7.1.69]	3	1
9	NCgl1892	membrane protein TerC	1	1
9	NCgl2109	cytochrome b subunit of the bc complex	9	4
9	NCgl2130	predicted permease	1	1
9	NCgl2533	hypothetical protein	1	1
10	NCgl0127	hypothetical membrane protein	2	1
10	NCgl0532	preprotein translocase subunit SecY	2	2
10	NCgl0543	hypothetical membrane protein	3	2
10	NCgl0674	predicted acyltransferase	2	2
10	NCgl0821	ABC-type transporter: permease component	1	1
10	NCgl1305	phosphotransferase system IIC component: glucose/maltose/N-acetylglucosamine-specific	19	6
10	NCgl2103	hypothetical membrane protein	1	1
10	NCgl2464	ABC-type transporter: permease component	1	1
10	NCgl2553	phosphotransferase system IIC component [2.7.1.69]	4	3
10	NCgl2780	hypothetical membrane protein	1	1
10	NCgl2883	hypothetical membrane protein	1	1
11	NCgl0406	permease of the major facilitator superfamily	1	1
11	NCgl1380	NhaP-type Na ⁺ /H ⁺ and K ⁺ /H ⁺ antiporter	6	3
11	NCgl2961	putative proline-betaine transporter	1	1
12	NCgl1085	ABC-type transporter: duplicated ATPase component	2	2
12	NCgl1108	putative gamma-aminobutyrate permease	3	2
12	NCgl2251	choline-glycine betaine transporter	6	2
12	NCgl2276	xanthine/uracil permease	1	2
12	NCgl2437	heme/copper-type cytochrome/quinol oxidase: subunit 1 [1.9.3.1]	13	5
12	NCgl2757	hypothetical membrane protein	1	1
12	NCgl2876	putative transmembrane transport protein	3	2
13	NCgl0185	hypothetical membrane protein	1	1
13	NCgl0799	Na ⁺ /proline: Na ⁺ /panthothenate symporter or related permease	7	1
13	NCgl0986	Na ⁺ -dependent transporter of the SNF family	2	2
13	NCgl1116	putative Na ⁺ /proline: Na ⁺ /panthothenate symporter	1	1
14	NCgl0184	putative arabinosyl transferase	5	4
15	NCgl2982	putative virulence factor	2	2
16	NCgl0626	carbon starvation protein: predicted membrane protein	11	5

Appendix V

Table 5. Two-component signal transduction systems of *C. glutamicum* identified by MALDI TOF/TOF mass spectrometry and database search after separation with two-dimensional reversed phase chromatography at high and low pH. Out of the two components, only the response regulators were identified. The sensor kinases are integral membrane proteins which were probably not extracted with the protocol employed in this work.

Gene	NCgl no.	Component	Regulatory role
<i>mtrB</i>	NCgl0722	Sensor kinase	Cell wall metabolism and osmoregulation (Möker et al..2004)
<i>mtrA</i>	NCgl0721	Response regulator	
<i>phoS</i>	NCgl2517	Sensor kinase	Adaptation to phosphate starvation (Kočan et al..2006)
<i>phoR</i>	NCgl2518	Response regulator	
<i>cgtS4</i>	NCgl0391	Sensor kinase	No data
<i>cgtR4</i>	NCgl0392	Response regulator	
<i>cgtS5</i>	NCgl2573	Sensor kinase	No data
<i>cgtR5</i>	NCgl2572	Response regulator	
<i>cgtS7</i>	NCgl0585	Sensor kinase	No data
<i>cgtR7</i>	NCgl0586	Response regulator	
<i>cgtS9</i>	NCgl2862	Sensor kinase	Copper metabolism (Kočan et al..2006)
<i>cgtR9</i>	NCgl2863	Response regulator	
<i>cgtS11</i>	NCgl2835	Sensor kinase	No data
<i>cgtR11</i>	NCgl2834	Response regulator	

Appendix VI

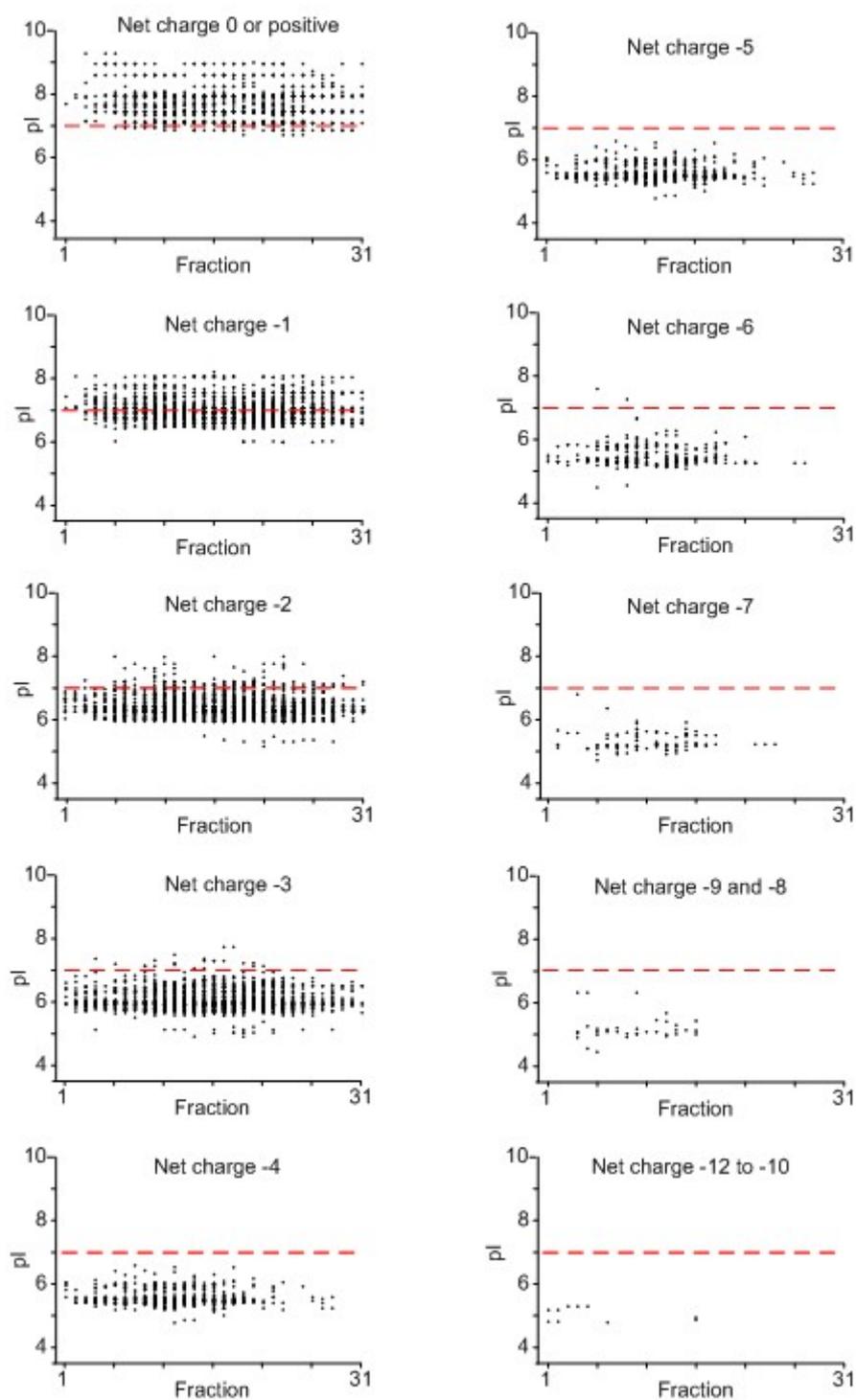
Table 6. List of peptides identified by LC-MALDI that carry net charge +1 at pH 2.1.

All these peptides corresponded to a C-terminal end of a protein.

Charge	Sequence	Protein ID	Name	
+1	MGPLSQGMGGLF	NCgl0240	hypothetical protein	Cterminus
+1	AVNEYLDSSFYEAG	NCgl0399	hypothetical protein	Cterminus
+1	GVGGEVLAYVW	NCgl0515	ribosomal protein S8	Cterminus
+1	APVIPEGLF	NCgl0628	uncharacterized protein	Cterminus
+1	AAQLGLVDVEQF	NCgl0838	hypothetical protein	Cterminus
+1	WVTAQAMFG	NCgl1202	6-phosphofructokinase [2.7.1.11]	Cterminus
+1	LFQWQED	NCgl1231	hypothetical protein	Cterminus
+1	FTLPFGIQA	NCgl1335	phe-tRNA synthetase alpha subunit	Cterminus
+1	LVFLAGPAE	NCgl1385	FHA-domain-containing protein	Cterminus
+1	QTGADDLSAEAFPEFE	NCgl1456	hypothetical protein	Cterminus
+1	SIVWGWYEN	NCgl1466	phospholipid-binding protein	Cterminus
+1	IAALIAQDLDAE	NCgl1862	hypothetical protein	Cterminus
+1	LETANFEGDLQL	NCgl2126	dihydrolipoamide acyltransferase	Cterminus
+1	LDDPTSVSVDPNAPPEE	NCgl2167	pyruvate dehydrogenase	Cterminus
+1	EYFGEEEAETSEEA	NCgl2329	FKBP-type cis-trans isomerase	Cterminus
+1	VIEASTLAEALAAVSL	NCgl2575	DNA repair protein	Cterminus
+1	NLLVSYDPNPTGLF	NCgl2698	NAD- aldehyde dehydrogenase	Cterminus

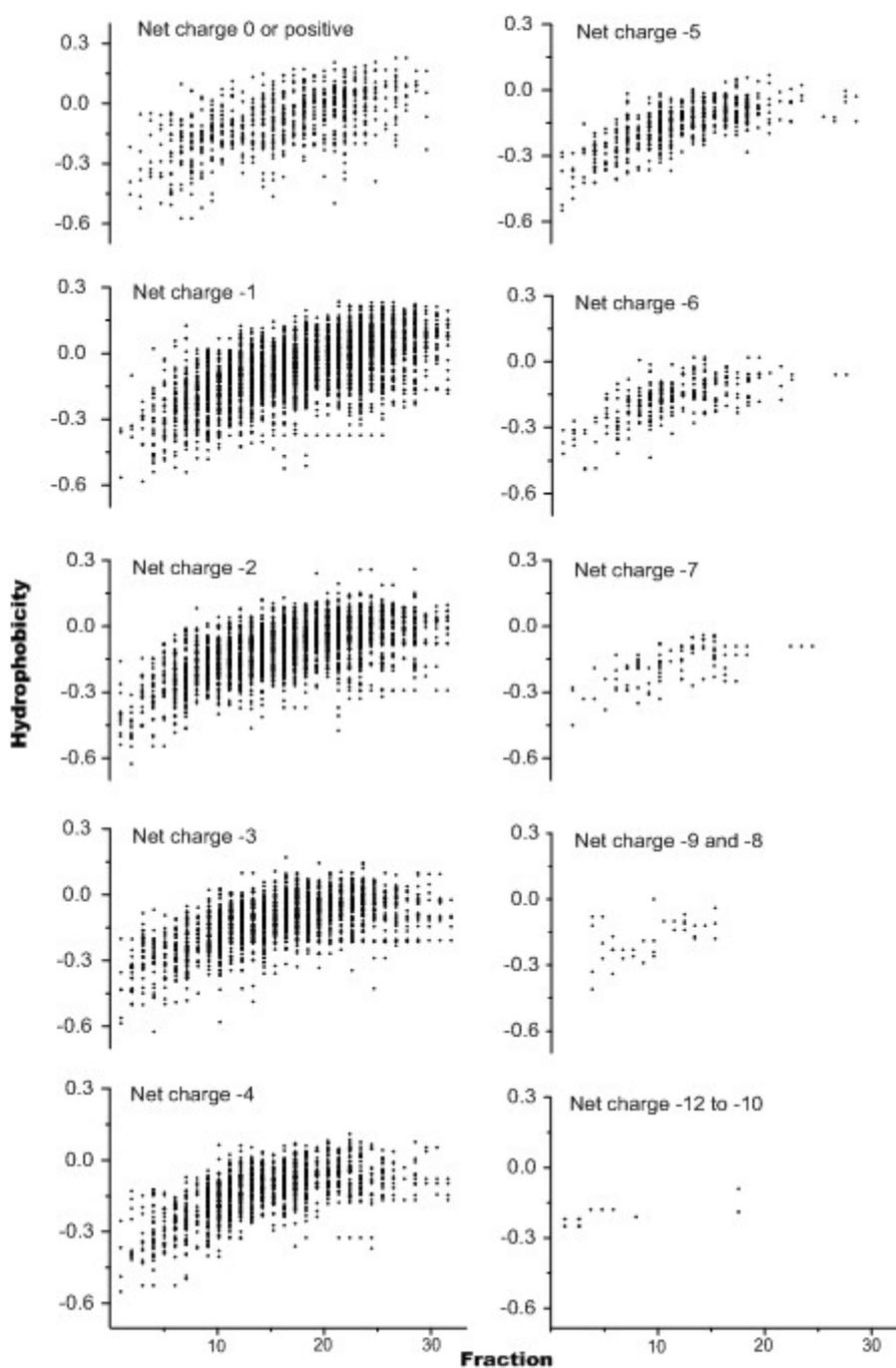
Appendix VII

Calculated isoelectric point for the identified peptides with respect to the fraction were they eluted during the first dimension of the 2D-RP-HPLC separation at high and low pH. Peptides are grouped according to their net charge at pH 10.



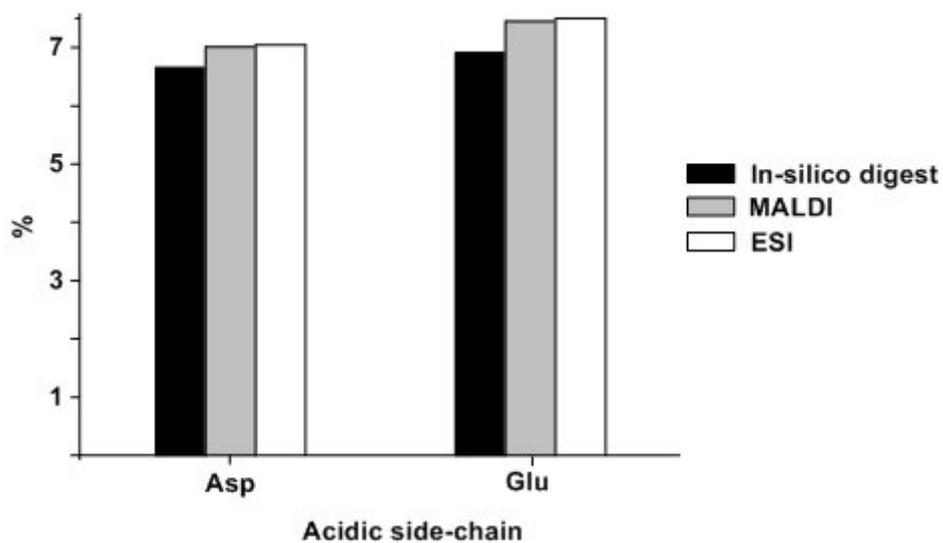
Appendix VIII

Hydrophobic character of the identified peptides calculated according to the Eisenberg scale. Peptides are plotted according to the fraction were they eluted during the first dimension of the 2D-RP-HPLC separation at high and low pH. Peptides are grouped according to their net charge at pH 10.



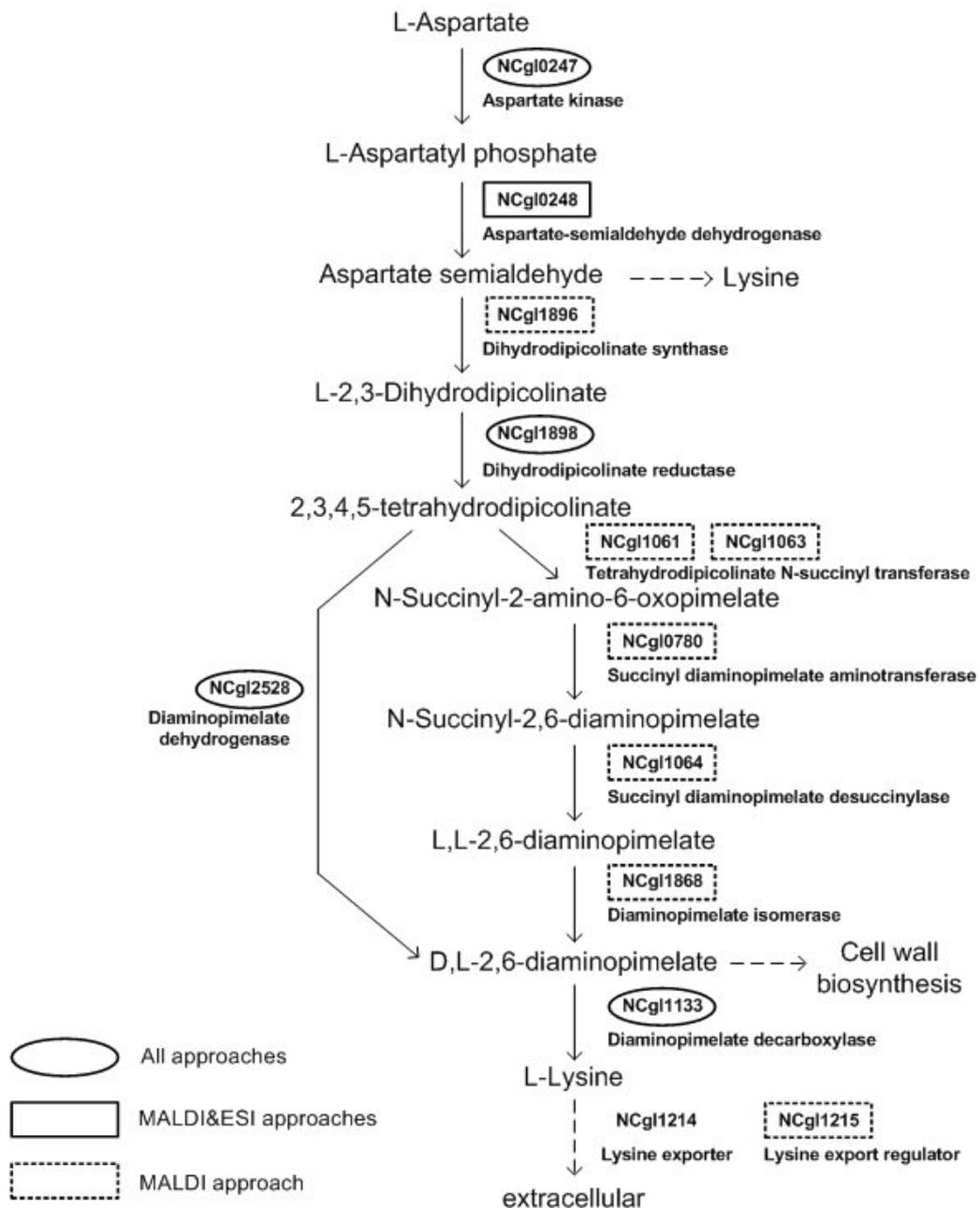
Appendix IX

The relative occurrence of aspartic acid and glutamic acid in the peptides identified by MALDI MS and ESI MS is higher than the relative occurrence predicted for the theoretical digest.



Appendix X

Enzymes catalyzing the reactions of the lysine biosynthesis pathway identified by the different proteomic approaches described in the present work: 2D-PAGE and two-dimensional RP x IPRP-HPLC followed by either MALDI TOF/TOF MS or ESI-IT MS.



Appendix XI

Enzymes of the methionine biosynthesis pathway identified. The proteomic approaches compared in the present work are: 2D-PAGE and 2D- RP x IPRP-HPLC followed by either MALDI TOF/TOF MS or ESI-IT MS.

