Proteomic study of light responses in *Rhodospirillum centenum*

by

Fatih M. İPEK

A thesis submitted to

the Graduate Institute of Sciences and Engineering

of

Fatih University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Biology

September 2007
Istanbul, Turkey

APPROVAL PAGE

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Assist Prof. M.Serdal SAKÇALI

Head of Department

This is to certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Barık SALİH

Supervisor

Examining Committee Members

Assoc. Prof. Barık SALİH _____

Assist Prof. İffet İrem UZONUR _____

Assist Prof. M.Serdal SAKÇALI _____

It is approved that this thesis has been written in compliance with the formatting rules laid down by the Graduate Institute of Sciences and Engineering.

Assist. Prof. Dr. Nurullah ARSLAN

Director

September 2007
Istanbul, Turkey

PROTEOMIC STUDY OF LIGHT RESPONSES IN *RHODOSPIRILLUM CENTENUM*

Fatih Mehmet İPEK

M. S. Thesis – Biology
September 2007

Supervisor: Assoc. Prof. Barık SALİH
Co-Adviser: Prof. Bart DEVREESE

ABSTRACT

Most photosynthetic microorganisms have the capability of sensing light of different wave lengths and intensities. *Rhodospirillum centenum* is an anoxygenic photosynthetic bacterium that contains a hybrid photoreceptor protein called Ppr (PYP – phytochrome related). The objectives of this study were to determine the proteomic changes of *Rh. centenum* towards blue light intensities, and the physiological condition at which the Ppr photoreceptor best function. Method: A wild type (wt) *Rh. centenum* strain and a Ppr-deletion mutant (ΔPpr) strain were grown anaerobically for 48 hours at 37^{o}C. The cultures were exposed to continuous red and blue light exposure while the controls received no blue light exposure. For each strain, the protein constituents were analyzed by quantitative two-dimensional gel electrophoresis. A fluorescent staining protocol was used to detecte protein spots, which is sensitive and compatible with subsequent mass spectrometry. For comparison, two different normalization methods were applied based on (i) a fluorescently labeled internal protein standard (FLIS) and (ii) the default normalization algorithm implemented in the Proteomweaver analysis software (Pair Matched Based). Differentially expressed proteins were detected with statistical analysis. Student's t-test was applied for detecting statistical significance (P value was set at <0.05). Differentially regulated protein spots were manually excised. The proteins were guanidinated in-gel and desalted/destained. Subsequently, the guanidinated proteins were enzymatically cleaved with trypsin and, after extraction, one set of the peptides were sulfonated for MALDI MS/MS analysis. The other set of non-sulfonated peptides were used for LC-MS/MS analysis. Subsequent identification of differentially expressed proteins was based on mass spectrometric (MS) analysis of tryptic peptides and the use of MS-based database search algorithms (Mascot analysis software). Tandem MALDI results were also subjected to manual *de novo* sequnencing. Results: Following computational analysis, we detected approximately 500 protein spots per gel using a fluorescent staining protocol. For ΔPpr strain, 90 spots were significantly detected after PMB-normalization and 35 spots after FLIS-normalization. After setting a threshold of two times the SD (2σ) only 19 significantly regulated spots (13 up, 6 down) remained for the PMB-normalized data and 13 spots (6 up, 7 down) for the FLIS-normalized data. In the case of wt strain we detected 72

significant spots for the PMB-normalized data and 138 spots for the FLIS-normalization. After setting a 95% confidence interval which accounts for experimental variation, only 10 significantly regulated spots (4 up, 6 down) remained for the PMB-normalized data and 7 spots (0 up, 7 down) after the FLIS-normalized data. Each step has ruled out more false positive spots and brought us closer to a true candidate list of significantly regulated spots. 14 interpreted spots out of 32 differentially expressed significant spots were identified with effective combination of three identification methods. Mascot analysis software was used to match for PMF and PFF results of MALDI and LC-MS, and partial *de novo* sequence of tandem MALDI results were used to match within the primary protein sequence information of *Rh. centenum*. Conclusions: This study has generated a preliminary data on the proteomics of *Rh. centenum* that has an important impact on unraveling of the regulatory proteins involved in its photo responses.

Keywords: Proteomics, *Rhodospirillum centenum,* photosynthesis, Ppr photoreceptor, two-dimensional gel electrophoresis, MALDI MS/MS, LC-MS/MS, normalization, sulfonation, guanidination, *de novo* sequencing.

IŞIĞA DUYARLI *RHODOSPIRILLUM CENTENUM* İÇİN PROTEOMİK ÇALIŞMA

Fatih Mehmet İPEK

Yüksek Lisans Tezi – Biyoloji
Eylül 2007

Tez Yöneticisi: Doç. Dr. Barık SALİH
Tez Eş Yöneticisi: Prof. Dr. Bart DEVREESE

ÖZ

Fontosentetik mikroorganizmalar ışığın dalga boyunu ve yoğunluğunu algılayabilmektedir. *Rhodospirillum centenum* oksijensiz ortamda yetişebilen fotosentetik bir bakteridir ve Ppr adlı hibrit ışık almacına sahiptir. Bu projenin amacı mavi ışığın ve Ppr ışık almacının *Rh. centenum* üzerinde oynadığı rolü proteomic açıdan belirlemektir. Metod: Doğal ve Ppr reseptörü mutasyona uğratılmış *Rh. centenum* suşuları oksijensiz ortamda 48 saat 37^{o}C'de büyütüldü. Kültürlerden kontrol gurubu yanlızca kırmızı ışığa maruz bırakılırken örnek çalışma gurubu kırmızı ışığa ilave olarak mavi ışıklı ortamda yetiştirildi. Her bir suşudan ayrı ayrı proteinler izole edildi ve kantitatif iki boyutlu jel elektroforezinde yürütüldü. Proten spotları kütle spektrometrisine uyumlu, yüksek hassasiyete sahip floresan boyama metodu kullanılarak tespit edildi. Protein sentez farklılıkları iki farklı normalizasyon metodu temel alınarak belirlendi. Bunlar FLIS ve Proteomweaver programının içerdiği PMB normalizasyon metodlarıdır. Proteinlerin sentez farkları istatistiksel olarak analiz edildi. Analiz metodu olarak (olasılık değeri 0,05'ten küçük) Student' t-test kullanıldı. Sentez farkı belirlenen proteinler yürütülmüş olan iki boyutlu jelden kesilerek alındı, proteinler guanidine edilip boy eve iyonlardan ayrıştırıldı. Triptik enzim ile muamele edilerek parçalanan proteinler jel ortamından izole edildi. Elde edilen protein fragmanları ikiye bölündü ve aynı örneklerden oluşturulan ilk gurup MALDI MS/MS analizinde kullanılmak için sülfonik asit ile modifiye edildi. Modifiye edilmeyen gurup ise LC-MS/MS analizinde kullanınldı.Ardından sentez farkı belirlenen proteinlerin triptik enzim ile kesilmiş peptit fragmanlarına kütle spektrometri analizileri yapıldı ve sonuçlar uygulanan kütle spektrometrisini temel alan bilgi bankasının arama algoritmasına sahip Mascot analiz programı kullanılarak tesbit edildi. Ayrıca MALDI MS/MS sonuçları *de novo* sekanslama yoluylan da belirlenmeye çalışıldı. Sonuç: Analizler sonucunda kütle spektrometri ölçümüne olanak sağlayan hassaisyeti yüksek floresan boya ile boyanmış olan her bir jel için yaklaşık 500 protein spotu belirlendi. Protein spotlarındaki farklılıklar görüntü analiz programı Proteomweaver ve ardından istatistiksel analizler ile belirlendi. Ppr mutant suşusu için PMB normalizasyon metodu ile 90, FLIS normalizasyon metodu ile 35 protein spotunda sentez seviyesinin değiştiği gözlemlendi. İstatistiksel olarak eşik değeri standart sapma değerinin iki katı alındığında (2σ) PMB metodu için 19 spot (13 sentezde artma, 6 sentezde azalma) ve

FLIS metodu için 13 spot (6 sentezde artma, 7 sentezde azalma) geriye kalmıştır. Doğal suşu için ise PMB normalizasyon metodu ile 72, FLIS normalizasyon metodu ile 138 protein spotunda sentez seviyesinin değiştiği gözlemlendi. İstatistiksel olarak güven aralığı %95 alındığında PMB metodu için 10 spot (4 sentezde artma, 6 sentezde azalma) ve FLIS metodu için 13 spot (0 sentezde artma, 7 sentezde azalma) geriye kalmıştır. Her bir adım bizi yanlış pozitif sonuçtan uzaklaştırarak daha güvenilir sentezinde farklılık olan proteinlerin listesi yapmamıza olanak sağlamıştır. Sentez seviyesinde farklılık belirlenmiş olan 32 proteinden 14 tanesi üç farklı yol kullanılarak tespit edildi. Mascot programı MALDI ve LC-MS analizleri sonucunda elde edilen PMF ve PFF spectrogramlarını, *de novo* sekans yolu ise yanlızca MALDI MS/MS analizi sonuçlarını kullanarak oluşturulan peptit sekans dizilerini *Rh. centenum*'a ait birincil protein dizisi içinde tarayarak proteinlerin isimlendirildi. Nihai sonuç: Bu çalışma *Rh. centenum* proteomu hakkında ön bilgiler elde edilmesini sağlayarak, ışığın etkisi karşısında sentezinde değişim olan proteinlerin, belirlenmesinede katkıda bulunmuştur.

Anahtar Kelimeler: Proteomiks, *Rhodospirillum centenum,* fotosentez, Ppr ışık almacı, iki boyutlu jel elektroforezi, MALDI MS7MS, LC-MS/MS, normalizasyon, sülfonik asit modifikasyonu, guanidinasyon, *de novo* sekanslama.

DEDICATION

To my beloved ones

ACKNOWLEDGEMENT

I would like to thank my major professor Dr. Barık Salih for his scientific guidance and endless support. I am grateful to him for getting me involved in several of his ongoing research even before I was enrolled in the graduate program, for having me publish my first scientific paper in the SCI listed journal and for my first participation in an international scientific conference (FEBS). My appreciation also goes to him for encouraging me to participate in the Erasmus mobility program to work on my research project at Ghent University in Belgium.

I also express my thanks to Dr. Bart Devreese at Ghent University for giving me the chance to work in his proteomics laboratory.

For Samy Memmi my laboratory supervisor at Ghent University a special thanks for his great support and guidance during my work. I am grateful for him for being a very helpful researcher and above all a good friend.

Thanks to Prof. Dr. Carl Bauer for kindly willing to open genome sequence of *Rh. centenum* before publisih and giving *Rh. centenum* Ppr deletion mutant strain.

I would like to thank my parents; my father Hüseyin, my mother Afet for their patience and understanding during my research. In addition, my thanks go to İlyas Bora Büyüklü for financially support me during my thesis research.

My thanks go to my friends Bora Kazım Bölek, Mehmet Şener and Tuna Çakar for their support and helpful ideas about my study.

Special thanks go to the administrative people at Fatih University and Ghent University for their help during my work within the Erasmus program. I am grateful to the Scientific and Technological Research Council of Turkey (TUBİTAK) for providing me with a graduate scholarship.

TABLE OF CONTENTS

LIST OF TABLES

TABLE

LIST OF FIGURES

FIGURE

LIST OF SYSMBOLS AND ABBREVIATIONS

SYMBOL/ABBREVIATION

APS	Ammonium Persulfate
Arg	Arginine
BSA	Bovine Serum Albumin
CHAPS	3-[3-(Cholamidopropyl) Dimethylammonio]-1-Proanesulfonate
cm	Centimeter
CID	Collision-Induced Dissociation
CV	Coefficient Of Variation
CBB	Coomassie Brilliant Blue
DE	Delayed Extraction
DNA	Deoxyribonucleic Acid
DNAse	Deoxyribonucleic Acid Nucleases
DC	Direct Current
DTT	Dithiothreitol
ESI	Electrospray Ionization
EST’s	Expressed Sequence Tag's
FLIS	Fluorescently Labeled Internal Protein Standard
G	G Force
IPG	Immobilised pH Gradients
IEF	Isoelectric Focusing
kV	Kilovolt
kDa	Kilodalton
LC	Liquid Chromatography
Lys	Lysine
MS	Mass Spectrometry

MALDI	Matrix-Assisted Laser Desorption Ionization
mRNA	Messenger Ribonucleic Acid
mA	Milliamper
mg	Milligram
mL	Milliliter
mM	Millimolar
min	Minutes
M_r	Molecular Mass
µ	Micro
N	Normality
UV	Ultra Violet
PMB	Pair Matched Based
PFF	Peptide Fragmentation Fingerprinting
PMF	Peptide Mass Fingerprinting
PYP	Photoactive Yellow Protein
Ppr	PYP – Phytochrome Related
p*I*	Isoelectric Point
RF	Radio Frequency
RNAse	Ribonucleic Acid Nucleases
rRNA	Ribosomal Ribonucleic Acid
rpm	Rotations Per Minute
Q	Quadrupole
SDS	Sodium Dodecyl Sulfate
SVD	Singular Value Decomposition
SD	Standard Deviation
SPSS	Statistical Package for the Social Sciences
SPITC	4-Sulfophenyl Isothiocyanate
TEMED	Tetramethylethyleendiamine
TOF	Time Of Flight
TCA	Trichloroacetic Acid

TCEP	Tris (2-Carboxyethyl) Phosphine
v/v	Volume Per Volume
Vh	Volt–Hour
w/v	Weight Per Volume
Wt	Wild Type
2-DE	Two Dimensional Electrophoresis
1-DE	One Dimensional Electrophoresis
3-D	Tree Dimensional
PAGE	Polyacrylamide Gel Electrophoresis
ΔPpr	Ppr-Deletion Mutant
°C	Centigrade Celsius

CHAPTER 1

INTRODUCTION

Most photosynthetic microorganisms have the capability of photosensing light quality and intensity. Movement of motile photosynthetic microorganisms to locations that offer optimal physical and chemical conditions for light-dependent growth provides obvious selective advantages. Among phototrophs, many cyanobacteria and algae migrate towards or away from the direction of light, a process termed phototaxis. In contrast, anoxygenic photosynthetic bacteria are believed to respond to changes in light intensity in a manner that is not related to the direction of light, a process that is often described by the term "photophobic". However, swarm colonies of the purple photosynthetic bacterium *Rhodospirillum centenum* are capable of macroscopically visible phototactic behavior [**1**]. *Rh. centenum* is an anoxygenic photosynthetic bacterium that is able to differentiate into several cell types [**2**], and contains a unique hybrid photoreceptor protein called Ppr (PYP – phytochrome related). This Ppr has a photoactive yellow protein (PYP) N-terminal domain, a central domain with similarity to phytochromes and a C-terminal histidine kinase domain. It is presumed to regulate gene expression of a polyketide synthase, homologous to chalcone synthases in plants, in response to blue light, with autophosphorylation inhibited in vitro by blue light [**3**]. The role of chalcone synthase-like polyketide synthases in bacteria is unclear; however, in *Rh. centenum* it might be involved in cyst formation [**4**]. Since cyst formation is involved in a number of stress responses, e.g. UV light, a role for a light sensor like Ppr in the induction of cysts is possible. However, the direct signaling partner of Ppr or response regulator remains undetermined.

The aim of this project is to study the cellular response of *Rh. centenum* towards various light conditions, and in particular which physiological role is suited for the Ppr photoreceptor. Therefore, we conducted a proteomic experiment for both a wild type

(wt) *Rh. centenum* strain and a Ppr-deletion mutant (ΔPpr) strain, in which the organism was grown with or without blue light in addition of a background of photosynthetic red light. For each case, the protein complement was separated and analyzed by quantitative two-dimensional gel electrophoresis. Subsequent identification of differentially expressed proteins was based on mass spectrometric (MS) analysis of tryptic peptides and the use of MS-based database search algorithms.

The outcome of this research will form the basis for a more detailed study of specific signalling pathways in *Rh. centenum*. Apart for the general biological interest for this topic, this might lead to the discovery and the development of nanomaterials that convert light to chemical or electronic responses.

CHAPTER 2

REVIEW OF LITERATURE

2.1 MICROBIAL EVOLUTION AND DIVERSITY

It has been estimated that our planet is about 4.6 billion years old. Fossilized remains of prokaryotic cells around 3.5 to 3.8 billion years old have been discovered in stromatolites and sedimentary rocks.

Prokaryotic life arose very shortly after the earth cooled. The earliest prokaryotes were anaerobic. Cyanobacteria and oxygen-producing photosynthesis probably developed 2.5 to 3.0 billion years ago. Microbial diversity increased greatly as oxygen became more plentiful [**5**].

Carl Woese and his collaborators divided prokaryotic life into two distinct groups based on their studies on rRNA sequences, and constructed a universal phylogenetic tree [**Figure 1**]. The tree is divided into three major branches representing three primary groups: *Bacteria, Archeae,* and *Eucarya*. The *Archaea* and *Bacteria* first diverged, and then the *Eukaryotes* developed.

Volume 2 of the Bergey's Manual® of Systematic Bacteriology is devoted entirely to the *Proteobacteria*, sometimes called the purple bacteria because of the purple photosynthetic bacteria present through several of its subgroups. The *Proteobacteria* are the largest and most diverse group of bacteria; currently there are over 380 genera and 1300 species. Although the 16S rRNA studies show that they are phylogenetically related [**6**], *Proteobacteria* vary markedly in many respects on morphological and metabolic characteristics, and consist of five classes: *Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, Deltaproteobacteria,*

and *Epsilonproteobacteria.* Members of the purple photosynthetic bacteria are mainly found among the *α-, β-, γ-proteobacteria* [**6**].

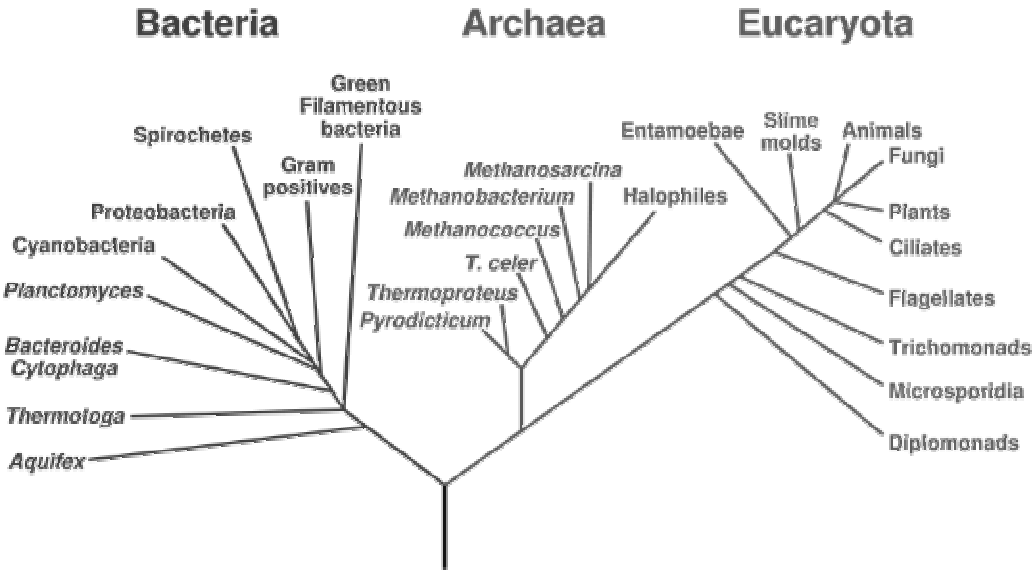


Figure 1. Universal phylogenetic tree [**7**].

2.2 *RHODOSPRILLUM CENTENUM*

Rh. centenum is a member of the *α-proteobacteria*, as an anoxygenic photosynthetic bacterium that is capable of differentiating into several cell types and one of several bacterial species capable of cyst cell formation [**8, 9**]. Cysts are bacterial resting cells that are extremely resistant to desiccation and mildly resistant to other environmental stresses, such as heat and UV light. Cyst formation is a complex process induced by a downshift in nutrient availability, which results in cells becoming spherical in shape, losing motility, and synthesizing a complex outer coat. The thick outer coat consists of two layers, an inner layer that consists of lipids and carbohydrates and outer layer that consists of lipopolysaccharides and lipoproteins, which function in protecting the cell from environmental stresses [**10**]. In *Rh. centenum*, cysts are most commonly observed in clusters of four or more cells surrounded by a common outer coat [**8, 11**]. The number of cells per cyst cluster varies from approximately four cells that are observed soon after the colony enters cyst formation to more than 10 cells per cyst cluster as the colony continues to age [**8**].

In liquid growth medium, this species forms swim cells with a single polar flagellum that exhibit a classic scotophobic response. However, when *Rh. centenum* is grown on an agar solidified medium, the cells differentiate into multi-flagellated swarm cells that are capable of exhibiting a visible phototactic response [**12, 13**] [**Figure 2**].

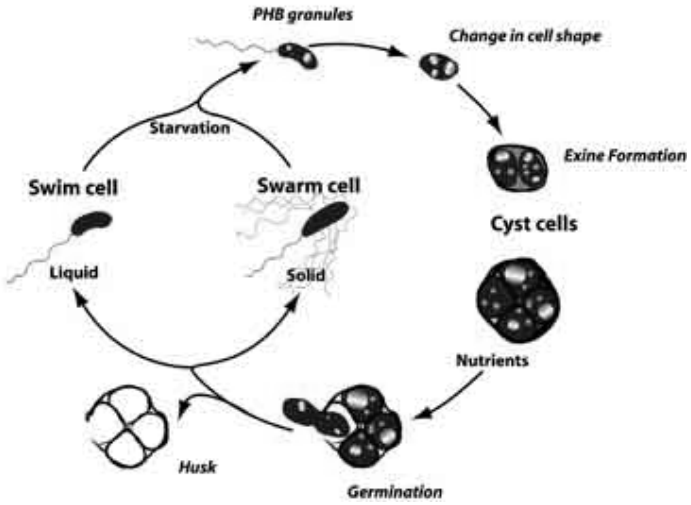


Figure 2. Diagram of the *Rh. centenum* life cycle, depicting swim cells, swarm cells and several stages of cyst formation. [**8**].

2.3 PHOTOBIOLOGY

Quantity and quality of light are important environmental factors on the response of all photosynthetic organisms. Plant cells use light as a signal for controlling many photomorphogenic processes (influence of light on plant development) such as chloroplast and seedling maturation and development. Motile unicellular microorganisms such as algae, cyanobacteria, and anoxygenic (non-oxygen-evolving) photosynthetic bacteria are capable of altering swimming behavior in response to light. These cells utilize photosensory behavior as a tool to migrate to locations that are optimal for photosynthetic growth [**14**].

Photosensory behavioral responses displayed by microorganisms generally fall into two categories [**15**]. One response is phototaxis, which is movement toward or away from a light source using a mechanism(s) that involves measuring the direction of propagation of a specific light beam. A second motility behavior is the scotophobic

response, which involves a random alteration in the direction of movement when motile cells experience a decrease in light intensity or cross a light/dark border [**14**].

2.3.1. Light Responses in *Rh. centenum*

The purple bacterium *Rh. centenum* was shown to exhibit a unique photosensory behavior. Swarm cells of *Rh. centenum* give response to both visible and infrared light aerobically grown colonies on solid CENS medium respectively move away and towards the light source [**14**]. However, there was no sign of swim cells of *Rh. centenum* moving towards or away from the sources of both visible and infrared light [**16**]. Swim cells exhibit a classic scotophobic response of random movement, rather than moving certain directions at a specific wave length of light.

In addition, the purple photosynthetic bacterium *Rh. centenum* has a unique hybrid photoreceptor protein called Ppr (PYP – phytochrome related). This Ppr has a photoactive yellow protein (PYP) N-terminal domain, a central domain with similarity to phytochromes and a C-terminal histidine kinase domain. Photoactive yellow proteins (PYPs) are 14-kDa water-soluble blue-light photoreceptors thought to be involved in mediating a photophobic response in *Halorhodospira halophila* against blue light to avoid UV radiation damage [**18**]. Phytochrome is a photoreceptor which has a function in photomorphogenic development. In plants, this photoreceptor plays a role in activation of seed germination, inhibition of stem elongation in dark-grown seedlings, induction of leaf expansion, and regulation of flowering in response to blue ligh [**18, 19**]. In prokaryotes, in the absence of protein serine / threonine / tyrosine kinases, protein histidine kinases play a major role in signal transduction involved in cellular adaptation to various environmental changes and stresses [**20**].

2.4 PROTEOMICS

2.4.1 General definition

Proteomics is often considered as the next step in the study of biological systems, after genomics. It is much more complicated than genomics, mostly because while an organism's genome is rather constant, a proteome differs from cell to cell and

constantly changes through its biochemical interactions with the genome and the environment. One organism has radically different protein expression in different parts of its body, at different stages of its life cycle and in different environmental conditions. Another major difficulty is the complexity of proteins relative to nucleic acids.

The proteome is defined as the complete protein complement of a genome, a photographic snapshot of the protein expression at a particular moment and under specific conditions [**21**]. One can say that there is one particular genome for every given organism or cell, but there is an infinite number of proteomes when referring to protein expression. A comprehensive description of the proteome of an organism not only provides a catalogue of all proteins encoded by the genome but also data on protein expression under defined conditions. Proteomics allows to obtain a quantitative description of protein expression and its changes under the influence of biological perturbations, the occurrence of post-translational modifications and the distribution of specific proteins within the cell [**22**].

Proteome analysis is a direct measurement of proteins in terms of their presence and relative abundance [**23**]. The overall aim of a proteomic study is characterization of the complex network of cell regulation. Neither the genomic DNA code of an organism nor the amount of mRNA that is expressed for each gene product (protein) yields an accurate picture of the state of a living cell [**24**], which can be altered by many conditions.

2.4.2 Classical approach

One of the greatest challenges in proteome analysis is the reproducible fractionation of these complex protein mixtures while retaining the qualitative and quantitative relationships. Therefore, proteomics has become formalized by combining techniques for large-scale protein separation with very precise, high-fidelity approaches that analyze, identify, and characterize the separated proteins.

Currently, two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) is the only method that can handle this task [**25, 26, 27**], and hence has gained special importance. Since 2D-PAGE is capable of resolving over 1,800 proteins in a single gel

[**28**], it is important as the primary tool of proteomics research where multiple proteins must be separated for parallel analysis. It allows hundreds to thousands of gene products to be analyzed simultaneously. In combination with computer assisted image evaluation systems for comprehensive qualitative and quantitative examination of proteomes, this electrophoresis technique allows cataloging and comparison of data among groups of researchers.

In the line of this project we performed a 2D-PAGE analysis on *Rh. centenum* wild-type and ΔPpr cell strains, grown under different light conditions. In this respect, we applied cell growth, protein extraction, first and second dimension electrophoresis, gel staining, visualization of a resulting, spot analysis, and mass spectrometric techniques [**Figure 3**].

2.5 ELECTROPHORESIS

The proteomes of cells are extremely complex, consisting of several thousand proteins. Because of this complexity, two dimensional polyacrylamide gel electrophoresis (2D-PAGE) has been widely used as the standard protein separation and display method. Two-dimensional gel electrophoresis separates proteins according to two independent parameters, which are the isoelectric point (p*I*) in the first dimension and the molecular mass (M_r) in the second dimension, by coupling isoelectric focusing (IEF) and sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) [**29**]. Proteins separated on 2-DE gels are visualized by staining with Coomassie blue dye, silver stains, fluorescent dyes, immunological detection or by radiolabelling, and subsequently quantified using densitometers, fluoro- and/or phosphor-imagers.

Although the limitations of the 2D-PAGE approach are well known, *i.e.* poor solubility of membrane proteins, limited dynamic range and difficulties in displaying and identifying low-abundance proteins, 2D-PAGE will remain a powerful and versatile tool and, at least in the immediate future, it is the most commonly used technique in proteome analysis [**29, 30, 31**].

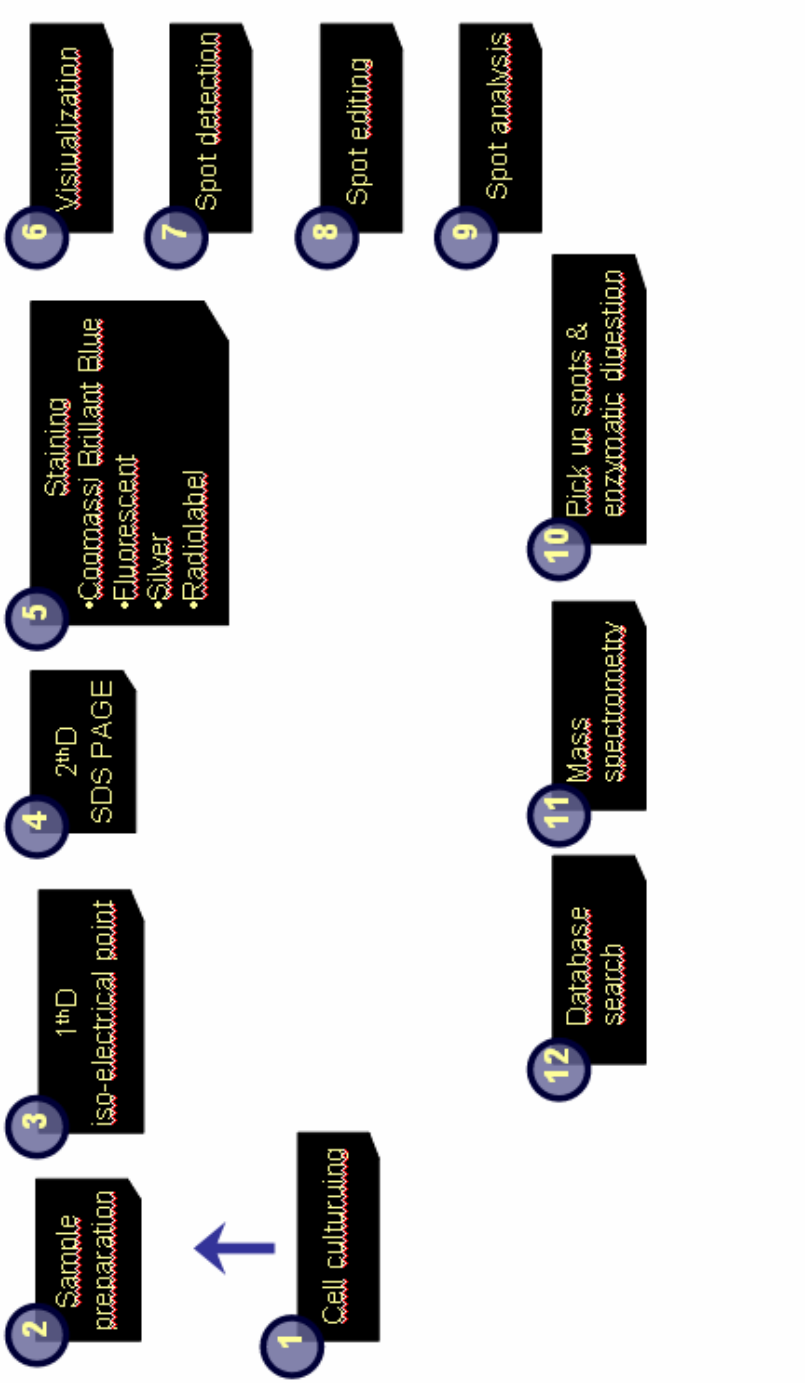


Figure 3. Necessary steps for proteome analysis.

2.5.1 Two Dimensional Polyacrylamide Gel Electrophoresis (2D-PAGE)

Any single-dimension method cannot resolve more than 80–100 different protein components. Two-dimensional gel electrophoresis exploits a combination of two different single dimension runs. Two-dimensional maps could be prepared by using virtually any combination of 1D method, but the one that has won universal recognition is that combining a charge (IEF) to a size (SDS-PAGE) fractionation, since these results in a more even distribution of components over the surface of the map. While this combination of separation methods was used at quite an early stage in the development of 2-DE macro-molecular mapping, it was the elegant work of O'Farrell [**32**] that really demonstrated the full capabilities of this approach. He was able to resolve and detect about 1100 different proteins from lysed *Escherichia coli* cells on a single 2-DE map.

2D-PAGE has become particularly significant in the past few years as a result of a number of developments. This new technique replaces the tube gels with gel strips supported by a plastic film backing [**33**]. Immobilised pH gradients are so precise that they allow excellent correlation between experimentally found and theoretically predicted p*I* values of both proteins and peptides. Methods for the rapid analysis of proteins have been improved. More powerful, less expensive computers and software are now available, allowing routine computerized evaluations of the highly complex 2-DE patterns. In addition, data on entire genomes for a number of organisms are now available, allowing rapid identification of the genes encoding a protein separated by 2-DE.

Today, the first dimension is preferably performed in individual IPG strips laid side by side on a cooling platform, with the sample often adsorbed into the entire strip during rehydration. At the end of the IEF step, the strips have to be interfaced with the second dimension, almost exclusively performed by mass discrimination via saturation with the anionic surfactant SDS. After the equilibration step, the strip is embedded on top of an SDS-PAGE slab, where the 2-DE run is carried out perpendicular to the 1DE migration. The 2-DE map displayed at the end of these steps is the stained SDS-PAGE slab, where polypeptides are seen, after staining, as spots, each characterized by an individual set of p*I* / M_r coordinates. Steps of proteome analysis are;

1. Sample preparation
2. First dimension: isoelectric focusing (IEF)
3. Second dimension: SDS-PAGE
4. Spot detection
5. Spot analysis

2.5.1.1 Sample preparation

Efficient and reproducible sample preparation methods are keys to successful 2-DE electrophoresis [**34, 35, 36**]. Sample preparation methods range from extraction with simple solubilization solutions to complex mixtures of chaotropic agents, detergents, and reducing agents. Sample preparation can include enrichment strategies for separating protein mixtures into reproducible fractions.

An effective sample preparation procedure will:

- Reproducibly solubilize proteins of all classes, including hydrophobic proteins.
- Prevent protein aggregation and loss of solubility during focusing.
- Prevent post-extraction chemical modification, including enzymatic or chemical degradation of the protein sample.
- Remove or thoroughly digest nucleic acids and other interfering molecules.
- Yield proteins of interest at detectable levels, which may require the removal of interfering abundant proteins or non-relevant classes of proteins.

Most protein mixtures will require some experimentation to determine optimum conditions for 2D-PAGE. Variations in the concentrations of chaotropic agents, detergents, ampholytes, and reducing agents can dramatically affect the 2-DE pattern.

The major problems concerning the visualization of extracted proteins of cells or tissues lie in the high dynamic range of expression, and the diversity of proteins with respect to molecular weight, isoelectric point and solubility. Although a one-step procedure for protein extraction would be highly desirable with regard to simplicity and

reproducibility, there is no single method of sample preparation that can be universally applied to all kinds of samples analyzed by 2D-PAGE.

The sample treatment is absolutely essential for a good 2-DE results. The protein composition of a sample must be reflected in the 2-DE gel pattern; hence, the sample must not be contaminated with foreign proteins. So, the method of sample collection is crucial. To avoid protein losses and modifications, sample treatment must be kept to a minimum.

The fundamental steps in sample preparation are (i) cell disruption, (ii) inactivation or removal of interfering substances and (iii) subsequent solubilization of the proteins.

The proteins are extracted with a so-called “lysis buffer”. The components of the lysis buffer have to convert all the proteins into a single conformation, prevent different oxidation steps and protein aggregates, get hydrophobic proteins into solution, deactivate proteases and cleave disulfide and hydrogen bonds. A standard lysis buffer contains a high urea concentration to convert proteins into a single conformation, to get and keep hydrophobic proteins in solution and to avoid protein–protein interactions [**37**]. Its action is through the disruption of hydrogen and hydrophobic bonds, the result of which has no effect on the intrinsic charge of the protein, making it ideal for IEF. With very hydrophobic proteins, such as membrane proteins, other, stronger denaturizing chaotrope like thiourea have to be added. Special care should be taken in preparing the samples to avoid chemical alterations in the proteins that could lead to changes in charges and doubling of spots. Temperatures above 37 °C for the lysis buffer are not recommended, as urea breaks down, and the resulting isocyanate would carbamylate the proteins, causing significant changes in their charges [**37, 38**], [**Figure 4**].

Zwitterionic or non-ionic detergent like 3-[3-(cholamidopropyl) dimethylammonio]-1-proanesulfonate (CHAPS), and Triton X-100 must be added to increase the solubility of hydrophobic proteins. Detergents help to disrupt membranes, solubilize lipids and delipidate proteins bound to vesicles or membranes. Hydrophobic

interactions can play an important role in the structural integrity of proteins and protein/protein interactions, and detergents are ideal for breaking these interactions.

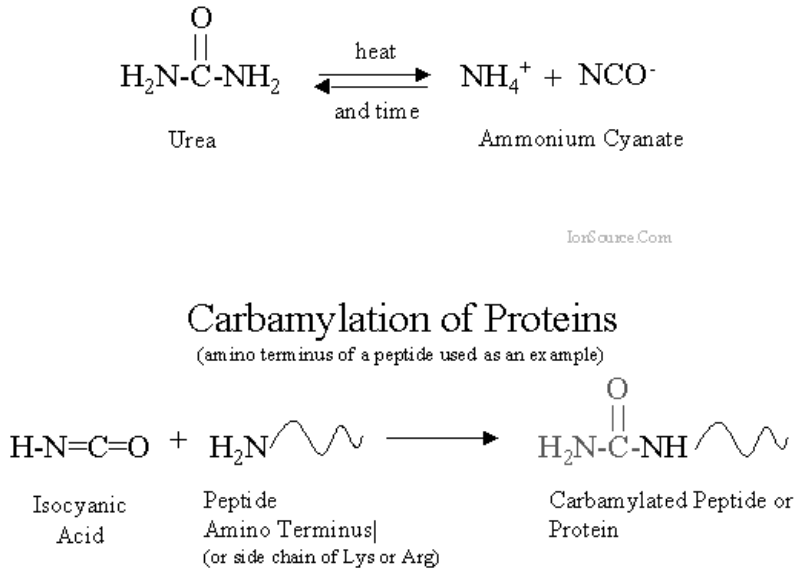


Figure 4. Decomposition of urea, and carbamylation of protein.

Reducing agents like dithiothreitol (DTT), beta-mercaptoethanol, and tris (2-carboxyethyl) phosphine (TCEP) prevent different oxidation steps. To fully denature proteins it is necessary to break any disulphide bridge that exists within a particular protein, or that links subunits of a complex together. This is undertaken by the use of a reductant, of which only relatively mild reducing agents are required for conversion of the disulphide bonds of proteins to sulfhydryl groups.

Careful sample handling is important when sensitive detection methods are employed. Silver-stained SDS-PAGE gels sometimes show artifacts in the 50 to 70 kD region and irregular but distinctive vertical streaking parallel to the direction of migration. This has been attributed to the reduction of skin keratin, a contaminant inadvertently introduced into the samples (keratin in the sample solution usually is focused near pH 5). Skin keratin is also a common contaminant seen in mass spectra. The best remedy for the keratin artifact is to avoid introducing it into the sample in the first place. Monomer solutions, stock sample buffers, gel buffers, and electrode buffers should be filtered through nitrocellulose and stored in well-cleaned containers. It also

helps to clean the gel apparatus thoroughly with detergent and to wear gloves while assembling the equipment.

2.5.1.1.a Sonication

Ultrasonic waves generated by a sonicator lyse cells through shear forces. Complete shearing is obtained when maximal agitation is achieved, but care must be taken to minimize heating and foaming. The sample must be cooled on ice between and during bursts.

2.5.1.1.b Bradford protein assay

Protein quantitation is often necessary before processing protein samples for isolation, separation and analysis by electrophoresis. The Bradford protein assay is the most commonly used method for the colorimetric detection and quantitation of total protein based on protein-dye binding (coomassie) chemistry. Use of Coomassie G-250 dye as a colorimetric reagent for the detection and quantitation of total protein was first described by Marion Bradford in 1976 [**39**].

In the acidic environment of the reagent, basic amino acids (arginine, lysine, and histidine) in the protein get positively charged. These charged amino acids tend to bind to the negatively charged sulphur trioxide group of the Coomassie dye. Protein concentrations are estimated by reference to absorbances obtained for a series of standard protein dilutions, which are assayed alongside the unknown samples. The best choice for a standard is a highly purified version of the predominant protein found in the samples. If a highly purified version of the protein of interest is not available or if it is too expensive to use as the standard, the alternative is to choose a protein that will produce a very similar color response curve. Bovine serum albumin (BSA) works well for a protein standard because it is widely available in high purity and is relatively inexpensive. For greatest accuracy in estimating total protein concentration in unknown samples, it is essential to include a standard curve each time the assay is performed.

Binding of the Coomassie dye to protein results in a spectral shift from the reddish / brown form of the dye (absorbance maximum at 465 nm) to the blue form of

the dye (absorbance maximum at 610 nm) [**Figure 5**]. The difference between the two forms of the dye is greatest at 595 nm, so that is the optimal wavelength to measure the blue color from the Coomassie dye-protein complex.

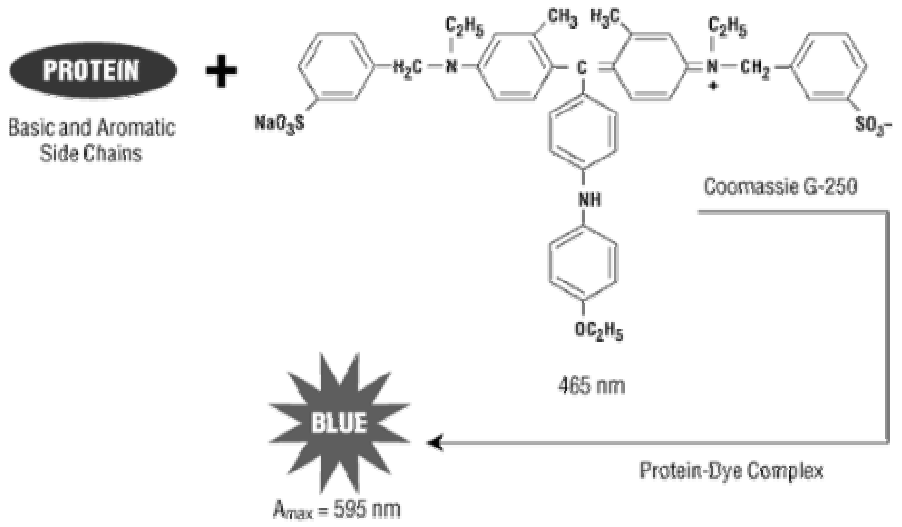


Figure 5. Reaction schematic for the Coomassie dye based protein assays [**40**].

In spectroscopy and spectrophotometry, the Beer-Lambert law, is an empirical relationship that relates the absorption of light to the properties of the material through which the light is travelling. There are several ways in which the law can be expressed,

$$A = \alpha\, l\, c$$

$$\frac{I_1}{I_0} = 10^{-A} = 10^{-\alpha l c}$$

where,

$$A = \log_{10}\left(\frac{I_0}{I_1}\right)$$

$$\alpha = \frac{4\pi k}{\lambda}$$

- A is absorbance

- I_0 is the intensity of the incident light
- I_1 is the intensity after passing through the material
- l is the distance that the light travels through the material (the path length)
- c is the concentration of absorbing species in the material
- α is the absorption coefficient or the molar absorptivity of the absorber
- λ is the wavelength of the light
- k is the extinction coefficient

In essence, the law states that there is a logarithmic dependence between the transmission of light through a substance and the concentration of the substance, and also between the transmission and the length of material that the light travels through. Thus if l and α are known, the concentration of a substance can be deduced from the amount of light transmitted by it [**Figure 6**].

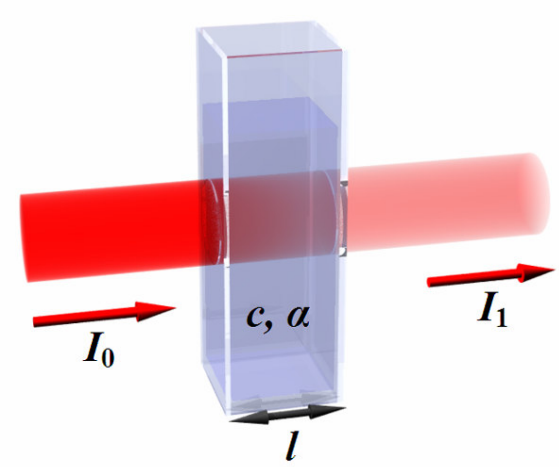


Figure 6. Representation of the Beer-Lambert absorption law, showing a beam of light being attenuated as it travels through a cuvette.

During or after cell lysis, interfering compounds have to be inactivated and/or removed from the sample solution due to the fact that they can disrupt 2D-PAGE application. Some of the interfering cellular chemicals are;

- Lipids
- Proteases
- Nucleic acids

- Polysaccharides
- Salts

(a) Lipids

Lipids can bind to proteins changing both their *pI* and M_r. For this reason it is important to add adequate detergent. Delipidation can be accomplished by extraction of the biological material with organic solvents (e.g. ethanol or acetone). However, severe losses in proteins may be experienced, either because certain proteins are soluble in organic solvent, or the precipitated proteins do not always resolubilize [**41**]. In several cases, high-speed centrifugation of lipid-rich material and subsequent removal of the lipid-layer has been recommended.

(b) Proteases

Protease activity can be a severe problem. The protein sample should be protected from proteolysis during cell disruption and subsequent preparation otherwise proteases degrade protein which can result in artificial spots. Proteases can be inhibited by disrupting the sample directly in strong denaturants of the lysis buffer such as urea, SDS, or use of the protein precipitant Trichloroacetic acid (TCA) [**42, 43, 44**]. Individual protease inhibitors are active only against specific classes of proteases, so it is usually advisable to use a combination of protease inhibitors. Broad-range protease inhibitor “cocktails” are available from a number of commercial sources.

(c) Nucleic acids

A crude extract can be contaminated with phospholipids and nucleic acids. Nucleic acids are visualized as horizontal streaks in the acidic part of the gel through interaction with carrier ampholytes and proteins. They can block gel pores and increase sample viscosity, they may also bind proteins, particularly nucleic acid binding proteins that will then be depleted and this way will not be fully represented on the gel. They can be removed from the protein sample by digestion with protease free nucleases (DNAse and RNAse).

(d) Polysaccharides

Polysaccharides cause severe smearing that is evident with both silver and Coomassie staining. Uncharged polysaccharides such as starch and glycogen can block gel pores, inhibiting migration of sample proteins resulting in poor focusing. These can simply be removed by ultracentrifugation. On the other hand, charged polysaccharides such as mucins and dextrans bind protein with their charge and therefore can deplete some proteins. They can also be very difficult to remove.

(e) Salts

High concentrations of salt can pose problems to the integrity of the IEF gel. Too much salt in the sample disturbs the isoelectric focusing and produces streaky patterns [**45, 46**]. In order to eliminate salt form the sample, use of concentrators or dialysis gives good patterns. However, the method involving protein precipitation with acetone and/or TCA and the subsequent resuspension in the same lysis buffer is also applied.

2.5.1.1.c Acetone precipitation

Acetone precipitation has been found very valuable for the inactivation of proteases to minimize protein degradation, removal of interfering compounds and, especially, for the enrichment of very alkaline proteins such as ribosomal proteins from total cell lysates [**47**]. However, attention has to be made to protein losses due to incomplete precipitation and/or resolubilization of proteins.

2.5.1.2 First dimension: isoelectric focusing

Differences in proteins net neutral p*I* are the basis of separation by IEF. The p*I* is defined as the pH at which a protein will not migrate in an electric field and is determined by the number and types of charged groups in a protein. Proteins are amphoteric molecules containing acidic and basic groups. In a basic environment, the acidic groups become negatively charged; in acidic environment the basic groups become positively charged. The net charge of a protein is the sum of all negative or positive charges of the amino acid side chains [**Figure 7**]. When an electric field is

applied, it will start to migrate towards the electrode of the opposite sign of its net charge. At the p*I*, the protein has no net charge and stops migrating [**Figure 8**].

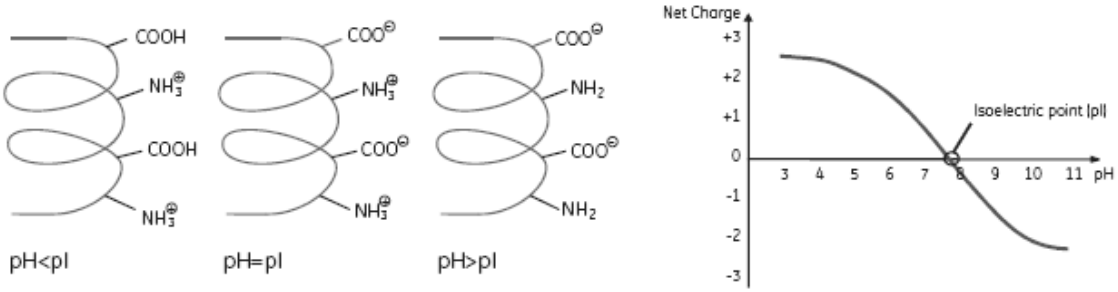


Figure 7. Amphoteric characteristics of proteins at different pH conditions [**48**].

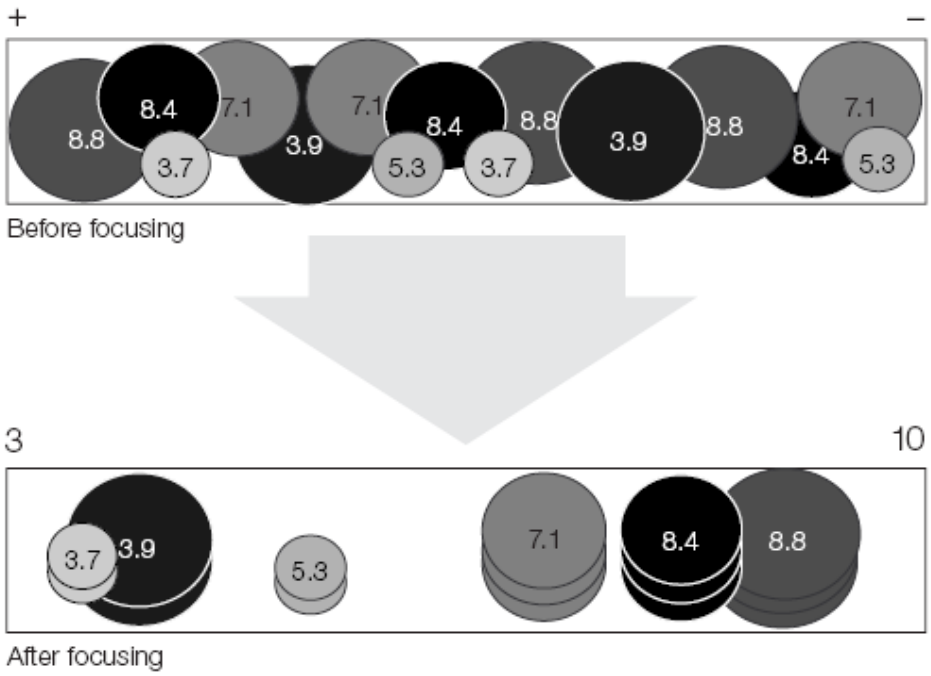


Figure 8. A mixture of proteins is resolved on pH 3-10 IPG strip according to each protein's p*I* and independently of its size [**49**].

Immobilised pH gradient (IPG) had been developed as an alternative to ampholytes by Bjellqvist [**33**]. The introduction of commercial IPG strips offer greater reproducibility and allow establishing comparisons among laboratories [**50, 51**]. Strips are plastic backed, come in various lengths, but are usually 3.0 mm wide and about 0.5

mm thick when rehydrated. They come, dehydrated, in a variety of pH ranges. The pH gradients in these gels are prepared by co-polymerising acrylamide monomers with acrylamide derivates containing carboxylic groups. Immobilised pH gradients are stable and capable of simultaneously focusing both acidic and basic proteins on a single gel prepared with a broad pH gradient. With IPG, several wide gradients have been used (3–10, 4–7). Important advantage of these strips is the possibility of loading greater amounts of sample, making the running of preparative gels possible for later characterization.

The complete voltage load during IEF is defined in volt–hour integrals (*Vh*). When the applied *Vh* are insufficient, not all spots are round and horizontal streaks are produced, and, in this way, higher *Vh* loads are needed, such as samples containing high molecular weight proteins, more hydrophobic proteins and preparative runs.

When the optimal conditions have been found, samples are run in at least five replicates, to check whether the pattern differences observed are caused by the noise of the system or by variations between different samples.

2.5.1.3 Second dimension: SDS PAGE

The second dimension of 2D-PAGE analysis separates proteins on the basis of their apparent molecular weight in polyacrylamide gels in the presence of SDS **[Figure 9]**. Large amounts of SDS are incorporated to form a SDS-protein complex in a ratio of approximately 1.4 g SDS/g protein, about 1 SDS for every 2 amino acid residues **[Figure 10]**. So, SDS masks the charge of the proteins and the resulting anionic complexes formed have a more or less constant net negative charge per unit mass. Thus, the electrophoretic mobility of proteins treated with SDS depends on the molecular weight of the protein. At a certain polyacrylamide percentage, there is an approximately linear relationship between the logarithm of the molecular weight and the relative migration distance of the SDS-polypeptide complexes of a certain molecular range. The molecular weights of the sample proteins can be estimated with the help of co-migrated standards of known molecular weights.

An appropriate acrylamide concentration is 12.5% (w/v). Lower percentages worsen the resolution of two-dimensional maps, and higher percentages (15%) make the extraction of proteins from the gel more difficult for later studies.

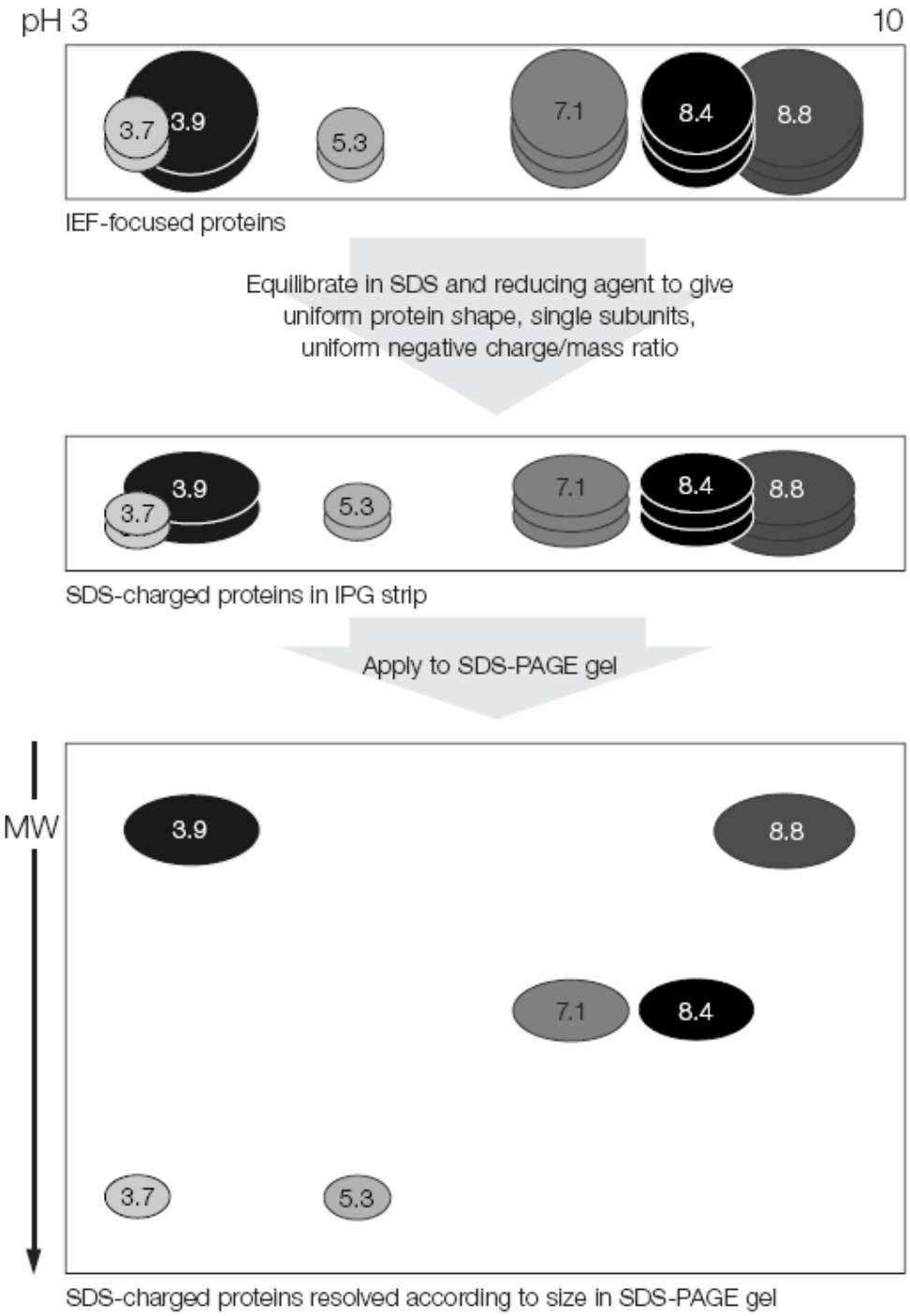


Figure 9. Separation of proteins by SDS-PAGE after separation by IEF [**52**].

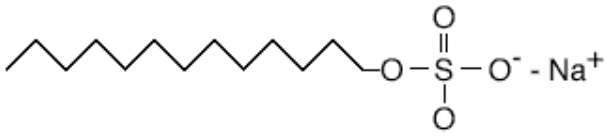


Figure 10. Sodium dodecyl sulfate, SDS.

The two-dimensional protein maps that allow only the visualisation of spots are called “analytical gels”. These gels present the best patterns for a comparative analysis. To increase the concentration of several specific proteins for posterior characterization, either by mass spectrometry or by any other analytical method, a greater sample load is applied in the first dimension, thereby obtaining protein maps containing greater amounts of protein. These gels are referred to as “preparative gels”.

The absence of spots overlapping and the clear spots separation make the analysis and comparison of gels easier, also facilitating the isolation of proteins from the 2-DE gels for later step of analysis. Additionally, the free spaces between spots can be covered by new spots, corresponding to small amounts of sample proteins. These spots could be detected by increasing the load or the sensitivity of the staining method [**53**].

2.5.1.4 Spot detection

Gels are run for either analytical or preparative purposes. The intended use of the gel determines the amount of protein to load and the means of detection. It is most common to make proteins in gels visible by staining them with dyes.

Once the electrophoresis is finished, the proteins are detected using a variety of staining methods of different sensitivities. There are several properties required for the ideal spot detection technique in 2-DE gels in proteomics. It should be sensitive enough for low copy number proteins, and allow for quantitative analysis. It should have wide linearity and dynamic range, and be compatible with mass spectrometry. Moreover, the technique should be environmentally friendly.

The sensitivity that is achievable in staining is determined by: 1) the amount of stain that binds to the proteins; 2) the intensity of the coloration; 3) the difference in coloration between stained proteins and the residual background in the body of the gel

(the signal-to-noise ratio). Unbound stain molecules should be able to be washed out of the gels without removing much stain from the proteins.

All stains interact differently with different proteins [**54**]. No stain will universally stain all proteins in a gel in proportion to their abundance. The only observation that seems to hold for most stains is that they interact best with basic amino acids. For critical analysis, replicate gels should be stained with two or more different stains. In practice, the techniques most applied in proteomics laboratories are Coomassie Brilliant Blue, and fluorescence staining.

2.5.1.4.a Coomassie Brilliant Blue

It is important to know that different staining techniques stain proteins differently. There are proteins that do not stain with CBB, but that do with fluorescent staining, and vice versa. The "classical" CBB recipe has a problem because during destaining, the spots are partly destained as well and because no steady state is reached, quantification is less reliable and less reproducible.

Colloidal Coomassie Blue staining contains ammonium sulfate, which increases the strength of hydrophobic interactions between the proteins and the dye. This staining method takes a very long time and requires several steps.

Coomassie Blue stained gels are usually compatible with mass spectrometry analysis because the dye can be completely removed from the proteins. Moreover, CBB stained gels contain enough protein for the identification and characterization using mass spectrometry.

2.5.1.4.b Fluorescent stains

An important recent advance was the development of high-sensitivity non-covalent fluorescent stains for general proteins, glycoproteins and phosphoproteins [**55**]. A major advantage of fluorescent stains with conventional chromogenic stains is a wider linear detection range, and they are therefore very well suited for quantification of proteins. Most of them show sensitivity similar or higher than that of the Coomassie

Blue dyes. Fluorescence staining is compatible with mass spectrometry. A fluorescence scanner is required for visualisation and detection.

In our experiment we used Flamingo (Flamingo™ Fluorescent Gel Stain, Bio-Rad) and Krypton (Krypton™ Protein Stain, Pierce Biotechnology) stains.

2.6 USE OF AN INTERNAL STANDARD

As already thoroughly explained, 2-DE is a powerful separation method for complex protein mixtures. However, large intergel variations in spot intensity limit its use for quantitative proteomics studies. Intergel variations are mainly caused by experimental dissimilarities in protein migration patterns and inefficiencies during protein loading through IPG strip reswelling and/or protein transfer between the first and second dimensions [**56, 57**].

Under the assumption that all proteins are absorbed into the IPG strip with equal efficiency, the incorporation of an internal protein standard covalently labeled with a fluorescent dye (*i.e.* DyLight™ $_{649}$) provides a method to quantify the experimental variability. The sample of interest is spiked with the fluorescently labeled internal protein standard (FLIS) prior to separation with 2-DE. Post-2-DE, total proteins are stained with a fluorescent protein stain that has sufficient spectral separation from the FLIS. Due to the high extinction coefficient of the DyLight™ $_{649}$ fluorescent molecule, incorporation of 0.1% of total protein is sufficient to allow visualization of the FLIS.

Thus, the FLIS and the total protein can be visualized and quantified independently at different wavelengths. Following quantification, the spot volumes of the total protein stain are normalized (divided) with the FLIS output from the same gel. Normalization allows a means to correct for variations in reswelling efficiency of each individual IPG strip and other reproducibility problems including incomplete protein transfer from the first dimension to the second dimension [**56**]. The procedure is outlined in **Figure 11**. Due to covalent labeling of the internal standard proteins with the DyLight™ $_{649}$ Maleimide, the spot pattern will be shifted in comparison to the unmodified sample proteins. However, a shift in p*I* and molecular weight does not influence the outcome of normalization by the FLIS method. The average (or median)

spot volume from a selection of representative protein spots in the FLIS standard is utilized for global ratiometric normalization, *i.e.*, the calculated average FLIS value from one 2-DE gel is used for normalization of all sample protein volumes in that same 2-DE gel.

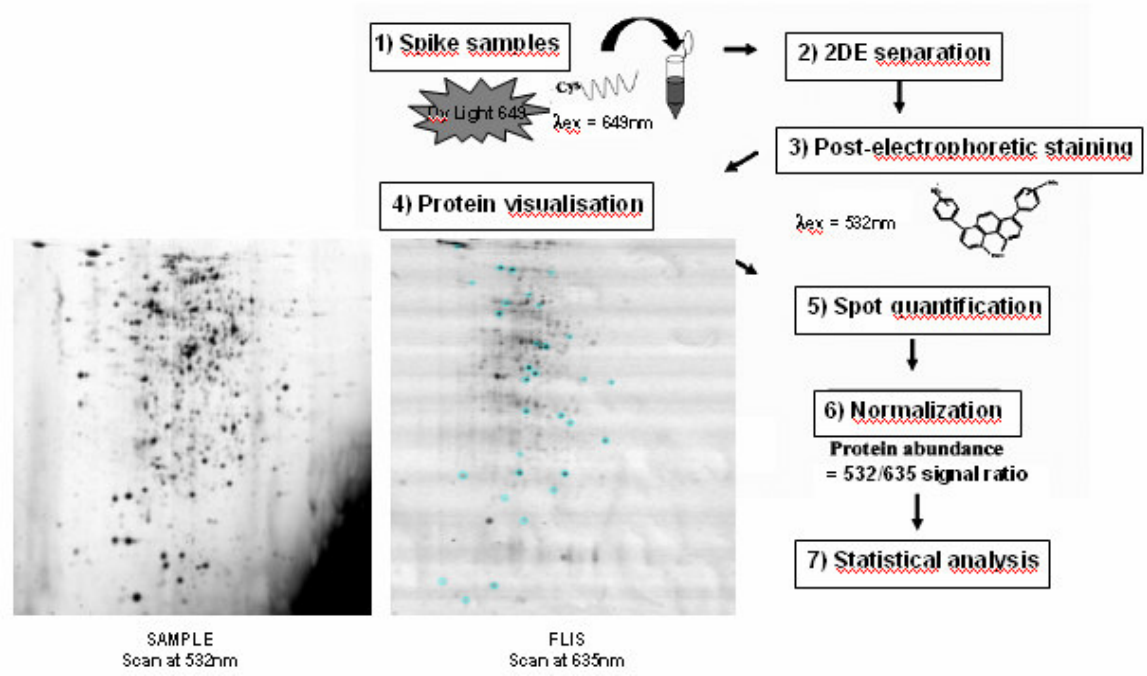


Figure 11. Flowchart diagramming the use of FLIS for normalization in quantitative proteomics using 2-DE separation. (1) Protein samples to be analyzed are spiked with 0.1% of the total protein load of FLIS DyLight$^{TM}_{649}$ prior to separation with 2-DE. (2) Protein sample is separated with 2-DE in the dark.(3) Following 2-DE separation, the proteins are stained with a fluorescent protein stain spectrally separated from FLIS. (4) Total protein and FLIS are visualized with a laser scanner using different channels. (5) Protein spots visualized using total protein stain (λex = 532nm) and FLIS (λex = 635nm) are quantified separately using 2-DE software, and quantities are exported to Excel. (6) Spot volumes of the total protein stain are normalized with the output of FLIS from the same 2-DE gel. (7) Statistical analysis of alterations in protein abundance is performed followed by identification of proteins of interest.

2.7 QUANTIFICATION OF 2-DE PATTERNS

The use of 2-DE in classical differential or comparative proteomics relies on using 2-DE images of related samples acquired from different gels. A number of

commercially available software packages exist that can analyse 2-DE images. Examples are PDQuest™ (Bio-Rad), Proteomweaver™ (Defines), Progenesis™ (Nonlinear Dynamics), Imagemaster 2-DE Platinum™ (GE-Healthcare), and DeCyder™ (GE-Healthcare). Although specific algorithms may differ, all software packages principally handle pretty much the same workflow: in brief, a 2-DE analysis subsequently consist of spot detection, spot matching across different gels, spot volume normalization to encompass for intergel verification in spot intensity and finally, the use of statistical test to detect significantly up or down regulated protein spots as well newly expressed or suppressed ones.

Computer-assisted image analysis software is an indispensable tool for the evaluation of complex 2-DE gels. It allows:

• Storage and structuring of large amounts of collected experimental image data.
• Rapid and sophisticated analysis of experimental information.
• Supplementation and distribution of data among laboratories.
• Establishment of 2-DE protein data banks.

Image analysis systems deliver comprehensive qualitative and quantitative data from a large number of 2-DE gels. The first steps in the differential analysis of 2-DE images are spot detection and matching. Before the software automatically detects the protein spots of a 2-DE gel, the raw image data are corrected and the gel background is subtracted. The process is executed with menus and “wizard tools”.

Software takes protein spots mathematically as 3-D Gaussian distributions and uses the models to determine absorption maxima. This enables automatic detection and resolution of merged spots. 2-DE analysis software generates a set of spots detected in each of a number of gel replicates. Following this procedure, spot intensities are obtained by integration of the Gaussian function.

The next step in 2-DE gel evaluation is the identification of proteins that are present in all gels of a series. This task is made difficult primarily because of inherent irreproducibility in gels, which affects the positions of spots within a gel series.

Gel analysis software must detect minor shifts in individual spot position within the gel series. Automatic gel comparison is created with the assumption that the relative positions of spots are altered only slightly relative to each other, and the spots are allocated on this basis, prior to automatic gel comparison. Before the software can detect and document matching of different spots, a number of landmarks, or identical spots in the gel series, must be identified. The matching tool speeds the process by making "best guess" assignments of landmark spots to images in the gel series.

Using the landmarks, the image analysis software first attempts to compare all spots lying very near these fixed points and then uses the matched spots as starting points for further comparisons. Thus, the entire gel surface is systematically investigated for the presence or absence of matching spots in a gel series. The results of the automatic gel comparison require verification, as does automatic spot detection. To compare gels from different experiments, the reference images are compared.

Subsequently, spot volumes are normalized to encompass for intergel variation inherent to 2-DE. Normalization adjusts the spot intensities or protein expression levels in such a way that they become comparable between different gels. Variability is introduced at several stages including inefficiencies in the initial protein loading techniques in the first dimension during rehydration and incomplete transfer of proteins from the first to the second dimension of separation. This causes an unknown variation in the total amount of protein present in any given 2-DE gel and drastically impairs the ability to quantitatively determine alterations in protein abundances. Typically, the normalization method used is based on the assumption that the majority of protein spots have not changed in expression level.

Several normalization methods have been described for 2-DE analysis. Most commonly used is Total Valid Spot Volume Normalization in which each individual spot volume is transformed into a percentage of the sum of spot volumes from the entire 2-DE gel.

Another possible method is used by Proteomweaver^{TM} analysis software, which consists of a pair matched based (PMB) normalization algorithm. It is performed after calculating pair matches between spot volumes from two gels. In brief, this algorithm

first computes a normalization factor between all pairs of gels for which a pair match exists: for every primary or hand matched in the pair match, the ratio between the pair matched spots is calculated. The normalization factor is the median of these ratios. Then it tries to compute for every gel an intensity factor that makes all the normalization factors as close to one as possible. The intensity factor is calculated by linear equation system using Singular Value Decomposition (SVD).

Using a fluorescently labeled internal protein standard (FLIS) for normalization purposes, each individual spot volume obtained at the excitation wavelength of the total protein stain is divided by the average value of a number of representative spots of the FLIS output of that same gel.

Finally, in an ideal situation, a representative of an individual spot is found in each replicate gel and their spot volumes show a normal distribution in the case of reproducible 2-DE. Consequently, the mean (μ) value of spot volumes of a matched protein spot can be considered as a representative expression level of that protein in the stated proteome. When population sample (N) is large, the standard deviation of population mean is decreasing according to;

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}.$$

So, including multiple biological replicates in the experiment improves the detection of a real protein expression level in the stated proteome.

Finally, I will describe two variation estimation techniques available by which the reproducibility (experimental variation) of the proteomic technique used (in this case 2-DE) can be verified: regression analysis and the coefficient of variation.

In the first method, the relatedness between two groups can be analyzed by plotting the average volume of each spot in the first group on a log scale against its average volume in the second group. If the spot's volume is the same in both, its point will fall on a straight line (slope = 1). If not, the point will fall above or below this

central line. The correlation coefficient for the regression line indicates the strength of the relationship between the average spot volumes of the two groups. A coefficient of 1.00 indicates that the two gel images are perfectly similar, while a low coefficient value indicates that the two are not very similar. Assuming that the majority of the proteins are not regulated between the two conditions, a highly reproducible 2-DE experiment should result in a correlation close to 1.00, as differential proteomics is mainly interested in detection of the small fraction of regulated or newly synthesized or suppressed spots.

The coefficient of variation (CV) or relative standard deviation is defined as the ratio of the standard deviation (σ) to the mean (μ). It is a dimensionless number that allows comparison of the variation of populations that have significantly different mean values, such as a set of e.g. 1,000 detected protein spots. It is often reported as a percentage (%) by multiplying the above calculation by 100.

$$\mathbf{CV\ (\%) = (\sigma / \mu) \times 100}$$

For 2-DE analysis, a low CV range is favored: the lower the resulting CVs, the lower the cutoff for detection of significant alterations will be. More importantly, a normal distribution of the relative standard deviation of the data set is required. If not, it indicates that certain spots (e.g. low abundance proteins) show more variation in their spot intensities than others.

2.8 MASS SPECTROMETRY

Determination of the protein composition in a sample is a complex analytical problem, as proteins show high physico-chemical variability. Different strategies have been developed to cope with this technically challenging task. These strategies all involve the separation of the complex protein mixture, followed by identification. The end result will be the protein composition of the investigated sample. In addition, samples can be compared with each other, resulting in a differentially display of the present protein. However, this goal can only be accomplished if a quantification technology is implemented in the profiling proteomics strategy. Nonetheless, mass

spectrometry (MS) plays a key role in each strategy, regardless of the method used for separation of the complex protein mixture.

MS has been described as the smallest scale in the world, not because of the mass spectrometer's size but because of the size of what it weighs. Over the past decade MS has undergone tremendous technological improvements allowing for its application to proteins, peptides, carbohydrates, DNA, drugs, and many other biological relevant molecules. First MS studies were performed in the beginning of the 20th century [**58, 59**]. However, it was only from the beginning of the 1980s that MS started to play a significant role in biological sciences. It took that time because MS requires charged, gaseous molecules for analysis and biomolecules, being large and polar, are not easily transfered into the gas phase, and ionized.

MS determines the mass of a molecule by measuring its mass to charge ratio (m/z). Ions are generated by inducing either the loss or gain of a charge from a natural species. Once formed, ions are electrostatically directed into a mass analyzer where they are separated according to their m/z ratio and finally detected. The result of molecular ionization, ion separation, and ion detection is a spectrum that can provide molecular mass or even structural information. The development of two soft ionization techniques, electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI), has overcome the harms and these methods are now the basis for diverse, powerful instrumentation. As ESI and MALDI were becoming routine in biological MS, the concept of proteomics was emerging. Further development of MS, *i.e.* more sensitive, accurate and high-throughput instruments, goes together with the expansion of proteomics, a trend that shows no signs of slowing down.

Basically, a mass spectrometer consists of three components: an ion source, to produce ions from the sample; one (or more) mass analyzer(s), to separate the ions according to their m/z ratio, and an ion detector, to register the number of ions arriving. The mass spectrometer is controlled by specialized software which allows processing the data and producing the mass spectrum in a suitable form. MS can be divided into two categories according to their ion source.

2.8.1 Ion Source

2.8.1.1 Matrix-Assisted Laser Desorption Ionization (MALDI)

Matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) was first introduced in 1988 by Tanaka and colleagues [**60, 61**]. It has since become a widespread analytical tool for peptides, proteins, and most other biomolecules. The efficient and direct energy transfer during a matrix-assisted laser-induced desorption event provides high ion yields of the intact analyte, and allows for the measurement of compounds with sub-picomole sensitivity. In addition, the utility of MALDI for the analysis of heterogeneous sample makes it very attractive for the mass analysis of complex biological samples.

The fundamental process of ion generation and desorption are still poorly understood and the subject of an active debate [**62**]. The most widely accepted mechanism states that MALDI causes the ionization and transfer of a sample from the condensed phase to the gas phase via excitation and vaporization of the sample matrix [**Figure 12**].

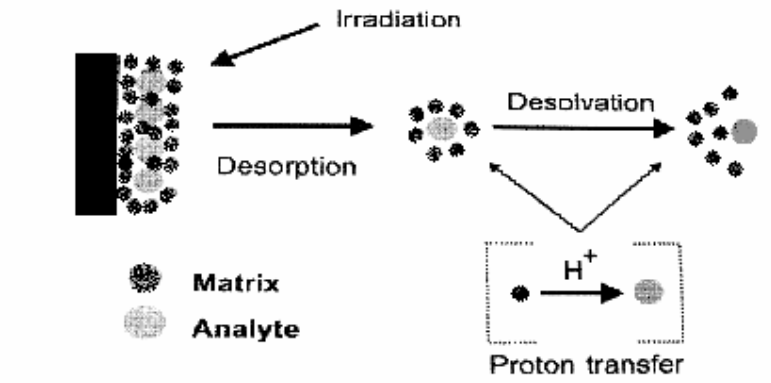


Figure 12. The efficient and direct energy transfer of the UV-laser pulse during a MALDI event allows for relatively small quantities of sample (femtomole to picomole) to be analyzed. In addition, the utility of MALDI-MS for the analysis of heterogeneous sample makes it very attractive for the mass analysis of biological sample.

In MALDI analysis, the analyte is first co-crystallized with a large molar excess of a matrix compound [**63, 64**], usually a UV-absorbing weak organic acid. The target is

then irradiated by a pulsed laser, mostly a nitrogen laser at 337 nm, which is used to deposit energy into the matrix. The energy causes rapid thermal heating and localized sublimation of the matrix crystals. The matrix contains a chromophore which can absorb nearly all of the laser radiation. The matrix also serves to minimize sample damage from laser radiation by absorbing most of the energy. The matrix expends into the gas phase and takes with it intact analyte molecules, which are desorbed into a gas phase. The most commonly used matrix molecules are α-cyano-4-hydroxycinnamic acid for peptides and polypeptides less than 5 kDa, and 3,5-dimethoxy-4-hydroxy-cinnamic acid (sinapinic acid) for proteins. Usually this procedure gives singly charged ions that are subsequently analyzed in the MS.

2.8.1.2 Electrospray Ionization (ESI)

Electrospray is an atmospheric-pressure ionization method that produces small charged droplets from a liquid medium under the influence of an electrical field. First ESI were reported in the late 1980s by the group of Fenn [**65, 66**], at Yale University. In conventional ESI, a liquid is passed through a thin conducting needle at high voltage (3-4 kV) and a potential difference is applied between the needle and a counter electrode. Depending on the pH of the solvent, the analytes will exist in an ionized form and like charges will accumulate at the tip of the needle. The highest density of charges at the tip leads to the formation of a Taylor cone [**67**]. When the repulsive forces become stronger than the surface tension at the tip of the cone, the liquid is ejected as a mist, which breaks up into smaller charged droplets when moving towards the counter electrode. This movement is directed by a potential and pressure gradient and a counter-current flow of gas facilitates solvent evaporation. Solvent evaporation in the charged droplets results into droplets breaking up into smaller droplets [**68**]. This is caused by the surface Coulomb forces exceeding the surface tension, the Rayleigh limit [**69**]. The process continues until droplets with diameters in the nanometer range are generated. This process is schematically present in **Figure 13**.

Two mechanisms, the charged-residue and the ion evaporation model [**70**], have been proposed for the formation of gas phase ions out of these small charged droplets. In the charged-residue model, it is proposed that solvent evaporates and droplets break up until those with only a single analyte ion are created and the evaporation continues

until a gas phase ion is formed. In the ion evaporation model, it is started that droplets with a radius less than 10 nm allow direct emission of a gaseous ion (field desorption). The charge state of the ion will depend on the number of charges transferred from the droplet to the ion during desorption. As in MALDI, the mechanism of ionization is still under debate [**71, 72, 73, 74, 75**]. If several ionizable sites are present, multiple charged ions ($[M + nH]^{n+}$) will be formed [**Figure 14**].

2.8.2 Mass Analyzers

When gaseous ions are formed, they are transported to the analyzer and separated according to their (m/z) ratio. A mass analyzer mainly determines the quality of a mass spectrum and the characteristics for these devices are: (i) mass range, (ii) resolution, (iii) transmission. The mass range limits the m/z value over which ions can be detected, while the resolution is defined as the smallest difference in mass still able to display two ions as separate peaks in the mass spectrum. The transmission is the ratio of ions reaching the detector and ions produced in the source.

The mass analyzers are diverse in terms of design and performance. Hence, they differ in sensitivity, speed, mass accuracy, mass resolution, and mass range. The three basic types of mass analyzers currently in use for proteomics research are: the quadrupole, the ion trap, and the time-of-flight.

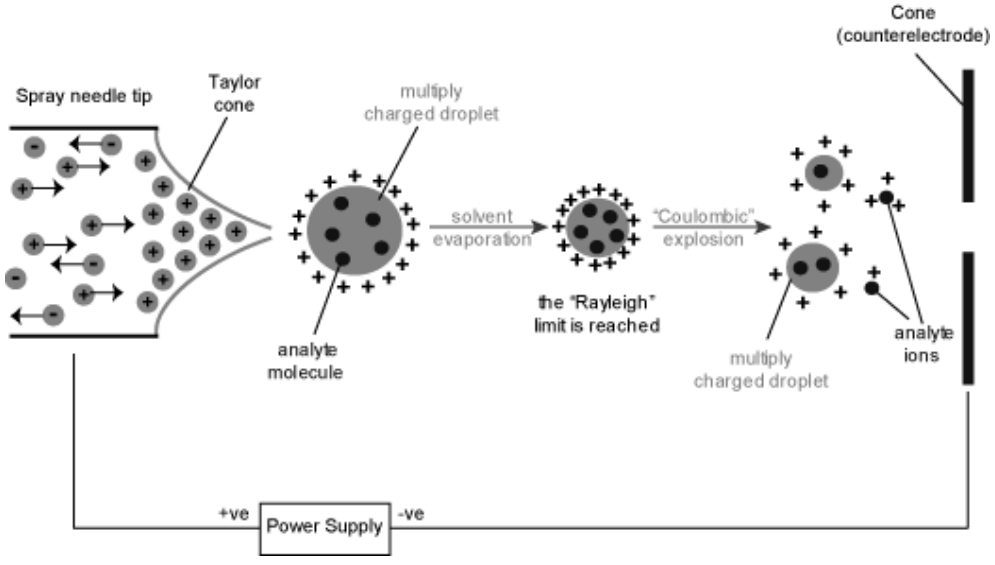


Figure 13. A schematic of the mechanism of ion formation in ESI.

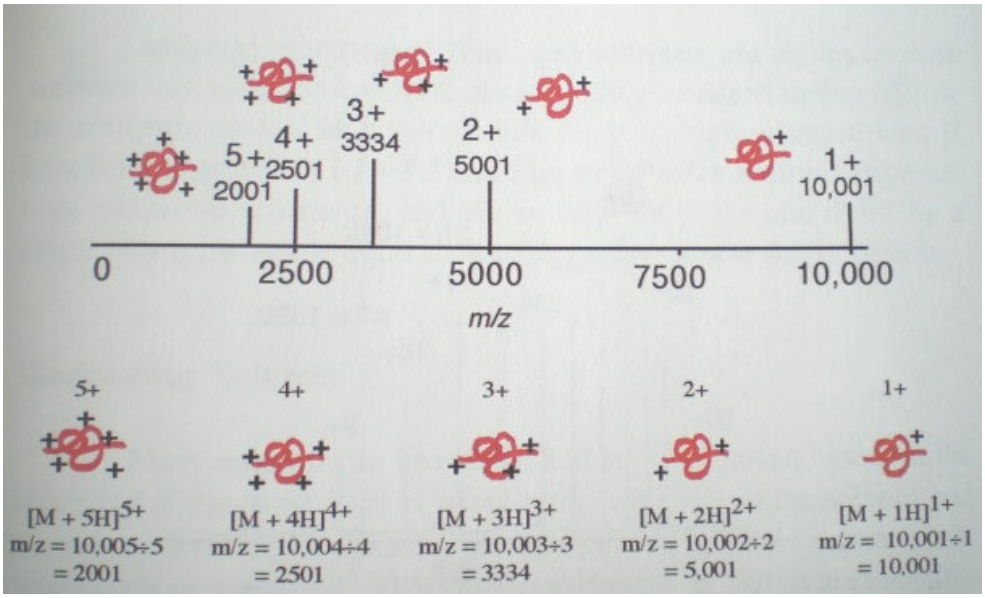


Figure 14. A protein with a molecular weight of 10,000 Da generates five different peaks with the ion containing 5, 4, 3, 2, and 1 charge(s), respectively. The mass spectrometer detects each of the protein ions at m/z 2001, 2005, 3334, 5001, and 10001, respectively [**76**].

2.8.2.1 Quadrupole Mass Filter

The quadrupole mass filter, introduced in 1953 [**77**], is constructed of four electrically conducting cylindrical rods and mass selection is achieved by properly choosing a combination of radio frequency (RF) and direct current (DC) voltages. A two dimensional quadrupole field is established between the four cylindrical, equidistantly spaced electrodes and with the two opposite rods connected electrically [**Figure 15**]. The DC and RF voltage are set such that only the ion of interest has a stable trajectory resonance through the quadrupole, and subsequently reaches the detector. All other out of range ions do not have a stable trajectory through the quadrupole mass analyzer and collide with the rods of the quadrupole [**Figure 16**]. The mass filter is a continuous analyzer and can be set to transmit a single isotope or to scan over a wide m/z range.

Quadrupoles offer three main advantages. First, they tolerate relatively high pressure (~5 x 10^{-5} torr), making them well suited for ESI sources where the ions are produced at atmospheric pressure. Secondly, quadrupoles have a significant mass range

with the capability of analyzing a m/z up to 4000, which is useful because ESI of proteins and other biomolecules commonly produce charge distributions from m/z 1000 to 3500. Finally, quadrupole mass spectrometers are relatively low cost instruments.

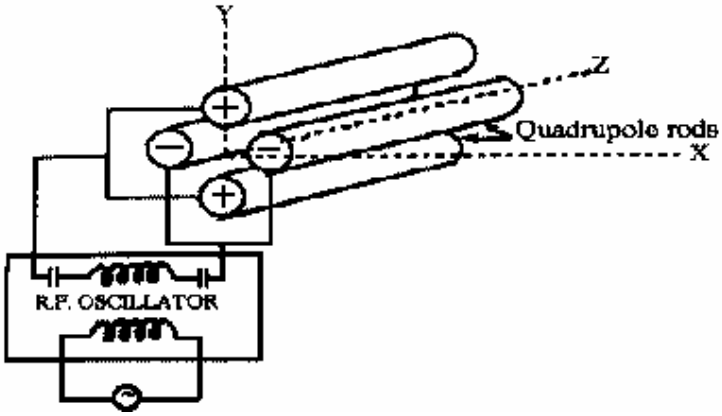


Figure 15. Arrangement of quadrupole rods and electrical connection to RF generator.

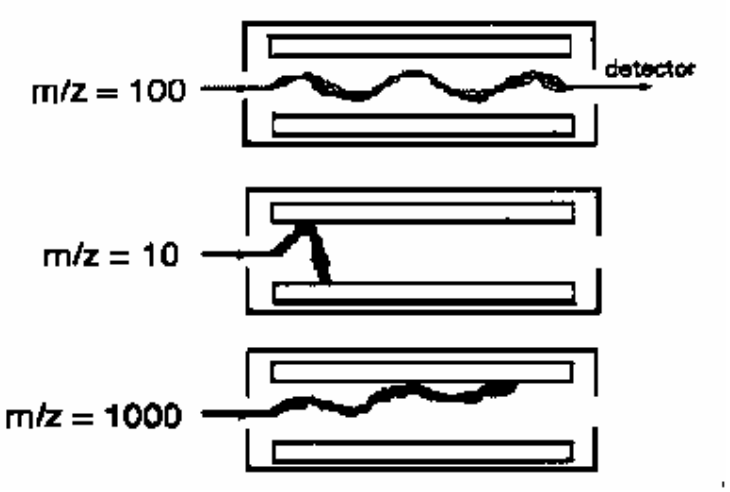


Figure 16. A cross section of a quadrupole mass analyzer taken as it analyzes for m/z 100, 10, and 1000, respectively. It is important to note that both the DC and RF fields are the same in all three cases and only ion with m/z = 100 travel the total length of the quadrupole and reach the detector, the other ions are filtered out.

In order to apply tandem mass analysis with a quadrupole instrument it is necessary to place two or more quadrupoles in series. Each quadrupole has a separate function: the first quadrupole (Q1) is used to scan across a present m/z range or to select an ion of interest. The second quadrupole (Q2), also known as the collision cell, transmits the ions while introducing a collision gas (argon or helium) into the flight path

of the selected ion, and the third quadrupole (Q3) serves to analyze the fragment ions generated in the collision cell (Q2) [**Figure 17**]

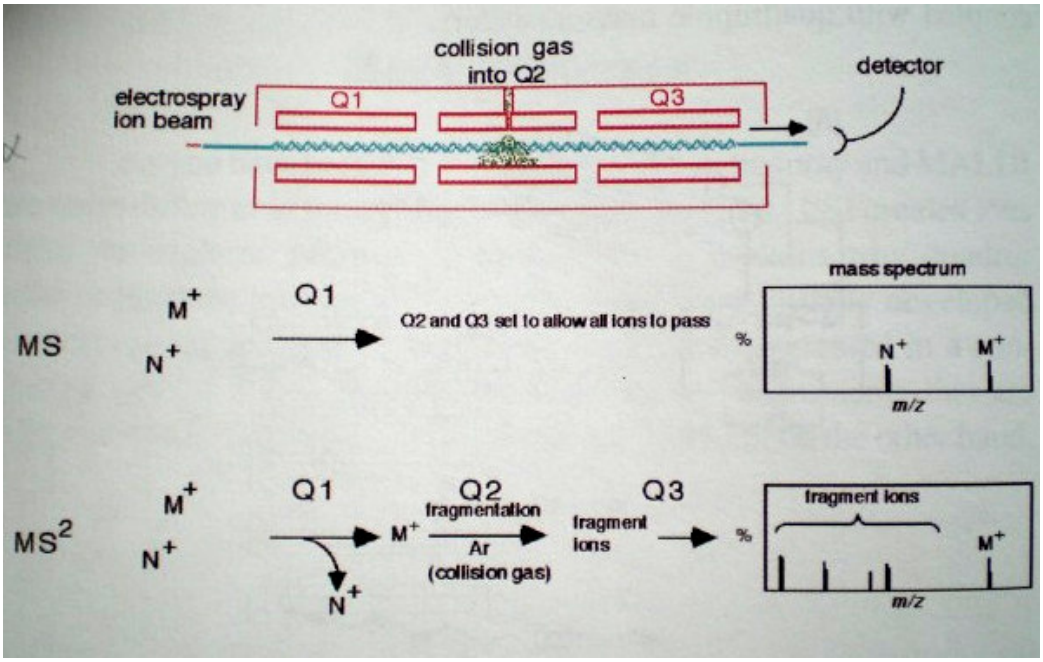


Figure 17. A triple quadrupole ESI mass spectrometer possesses ion selection and fragmentation capabilities allowing for tandem mass spectra [**78**].

2.8.2.2 Time of Flight

The principle of the time of flight (TOF) mass analyzer is to measure the flight time of an ion accelerated out of an ion source into a field free drift tube to a detector. TOF is related to the m/z values of the ions according to the following formula;

$$t^2 = \frac{m}{z}\left(\frac{d^2}{2Ve}\right)$$

t = time

d = distance

m = mass

z = charge

V = voltage

e = kinetic energy

TOF analysis is based on accelerating a group of ions to a detector where all of the ions are given the same amount of energy through an accelerating potential. Because the ions have the same energy, but different mass, the lighter and heavier ions reach the detector at different times. The lighter ions reach the detector first because of their greater velocity. Hence, the analyzer is called TOF because the mass is determined from the ions' time of arrival. Mass, charge and kinetic energy of the ion all play part in the arrival time at the detector.

The resolving power of TOF analyzer is limited by the initial spatial spread and initial velocity of the ions. For example, in a MALDI source, the ionization creates a burst of ions that will have different kinetic energies and the ions will be at different distances from the detector. Therefore ions with the same m/z value but with different kinetic energy or at different distances will be detected at different times and will cause decreasing resolution. Two powerful techniques have been developed to compensate for time of ion formation, location of ion formation, and kinetic energy of ion formation distribution. The first technique is the introduction of ion mirrors, or reflectrons [**79**], which compensates for the initial energy spread of the ions. The reflectron creates a retarding field that deflects the ions and sends them back through the flight tube [**Figure 18**]. The more energy the ion has, the further it penetrates the retarding field of the reflectron before it is being reflected. Therefore, a more energetic ion will travel a longer flight path and arrive at the detector at the same time as less energetic ions with the same mass. The kinetic energy distribution of ions is also the result from their acceleration through the gas phase plume created during desorption [**80**]. Brown and Lennon [**81**] have observed that both mass accuracy and mass resolution are significantly improved by introducing a delay between ion formation and extraction of the ions from the source. In theory, delayed extraction is a relatively simple way of cooling and focusing the ions immediately after the MALDI ionization event, yet in practice it was initially a challenge to pulse 10 kV on and off within a nanosecond time scale. In traditional MALDI instruments, the ions were accelerated out of the ionization source immediately as they were formed. However, with delayed extraction the ions are allowed to cool for ~150 nanoseconds before being accelerated to the analyzer [**Figure 19**]. This cooling period generates a set of ions with a much smaller kinetic energy distribution. Overall, this result increased resolution and accuracy.

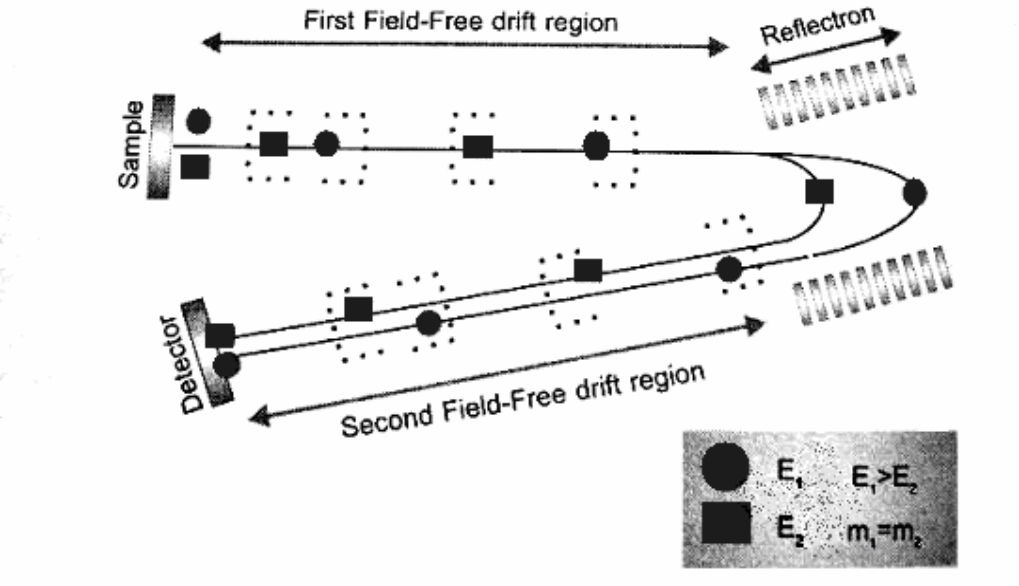


Figure 18. Two ions of the same mass but different initial kinetic energies, E1 and E2, are produced and extracted into the first field free region. The ions are then reflected by the ion mirror, the reflectron, and travel through the second field free drift region to be detected at the same time by the detector.

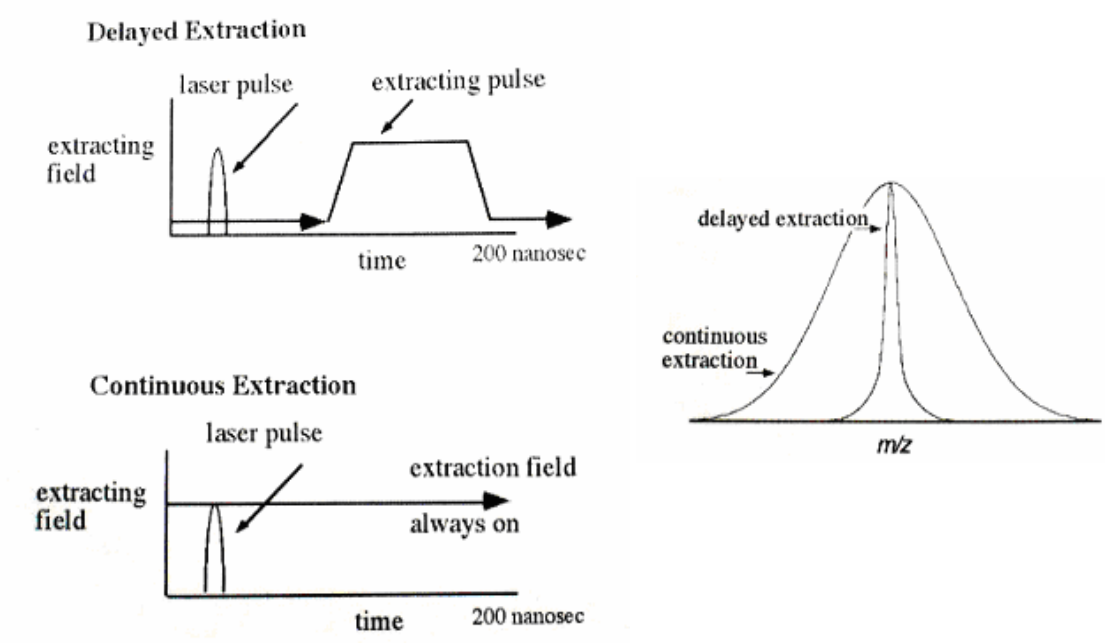


Figure 19. Delayed extraction (DE) is a technique applied in MALDI which allows ions to be extracted form the ionization source after a cooling period. This cooling period effectively narrows the kinetic energy distribution of the ions, thus providing higher resolution than in continuous extraction techniques [**82**].

2.8.2.3 Ion Trap

The ion trap consist of a chamber surrounded by a ring electrode and two end cap electrodes and can be regarded as a quadrupole mass filter bent around on itself to form a closed loop [**83, 84**]. Hence, the ion trap is also referred to as the quadrupole ion trap. Unlike the quadrupole and TOF analyzers, which separate ions in space, the ion trap separates ions in time. Ions of all m/z values enter the trap at the same time and are subjected to oscillating electric fields generated by an RF voltage applied to the ring electrode only. As ions will repel each other, ion losses are prevented by introducing a gas, typically helium, inside the trap, which will remove excess energy from the ions by collision. This has the effect of collisionally damping the ion motion and causes the ion to migrate to the center of ion trap. Ions above the threshold m/z ratio remain in the trap while others are ejected through small holes in the distal end-cap electrode. By gradually increasing the voltage in the ring electrode, ions with increasing m/z ratios are ejected over time. This process is called mass selective ejection.

2.9 TANDEM MASS SPECTROMETRY

Tandem mass spectrometers [**85**] mostly use collision-induced dissociation (CID) [**86, 87**] for the acquisition of MS-MS spectra. While the molecular weight information obtained from ESI, and MALDI are useful in the preliminary stages of characterization, it can also be very important to gain more detailed structural information through fragmentation. Tandem mass spectrometry, the ability to induce fragmentation and perform successive mass spectrometry experiments on these ions, is generally used to obtain this structural information [**Figure 20**].

In tandem mass spectrometry, a precursor ion is selected in the first analyzer and focused into a collision cell. In the cell, filled with an inert gas, collision occurs between the gas and the precursor ion and translational energy of the precursor ion will be partly converted into internal energy, exciting the ions to an unstable state. The precursor ions dissociate and fragment to product ions. The masses of the product ions are then determined in a second mass analyzer. This process is called collision induced dissociation.

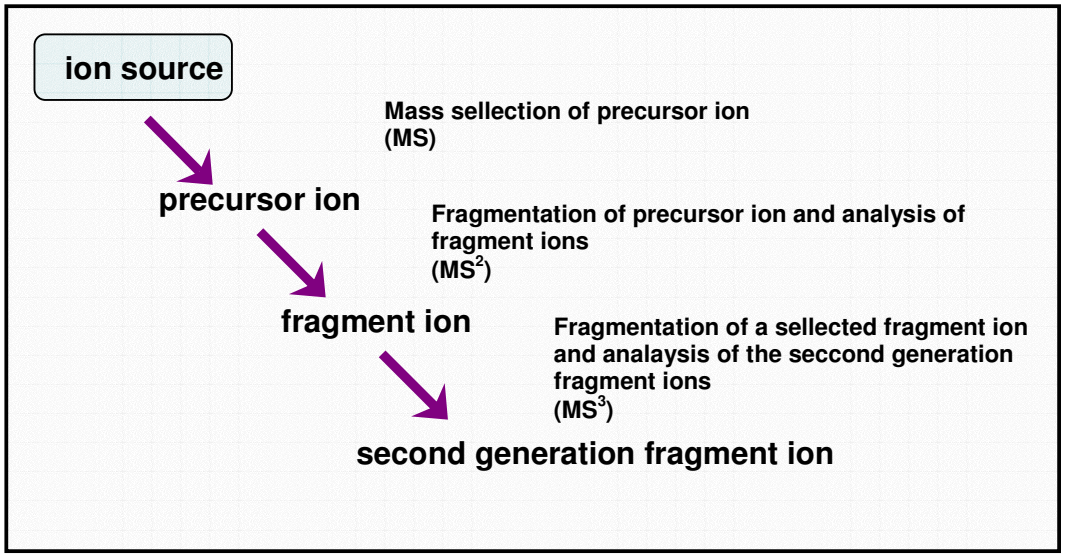


Figure 20. Description of tandem mass spectrometry experiment. (MS^n – n refers to the number of generations of fragment ions being analyzed).

In order to obtain peptide sequence information by mass spectrometry, fragments of an ion must be produced that reflect structural features of the original compound. Fortunately, most peptides are linear molecules, which allow for relatively straightforward interpretation of the fragmentation data. This process is initiated by converting some of the kinetic energy from the peptide ion into vibrational energy. This is achieved by introducing the selected ion, usually an $(M + H)^+$ or $(M + nH)^{+n}$ ion, into a collision cell where it collides with natural argon, xenon, or helium atoms, resulting in fragmentation. The fragments are then monitored via mass analysis. Tandem mass spectrometry allows for a heterogeneous solution of peptides to be analyzed and then by filtering the ion of interest into the collision cell, structural information can be derived on each peptide from the complex mixture. The fragment ions produced in this process can be separated into two classes. One class retains the charge on the N-terminal and occurs at three different positions, designated as types a_n, b_n, and c_n. The second class of fragment ions retains the charge on the C-terminus and fragmentation occurs at three different positions, type x_n, y_n, and z_n. Most fragment ions are obtained from cleavage between the carbonyl carbon and amine nitrogen of the amide peptidebond, resulting in b- and y-type ions when the charge is retained on the N-terminal or C-terminal end of the molecule. [**Figure 21**].

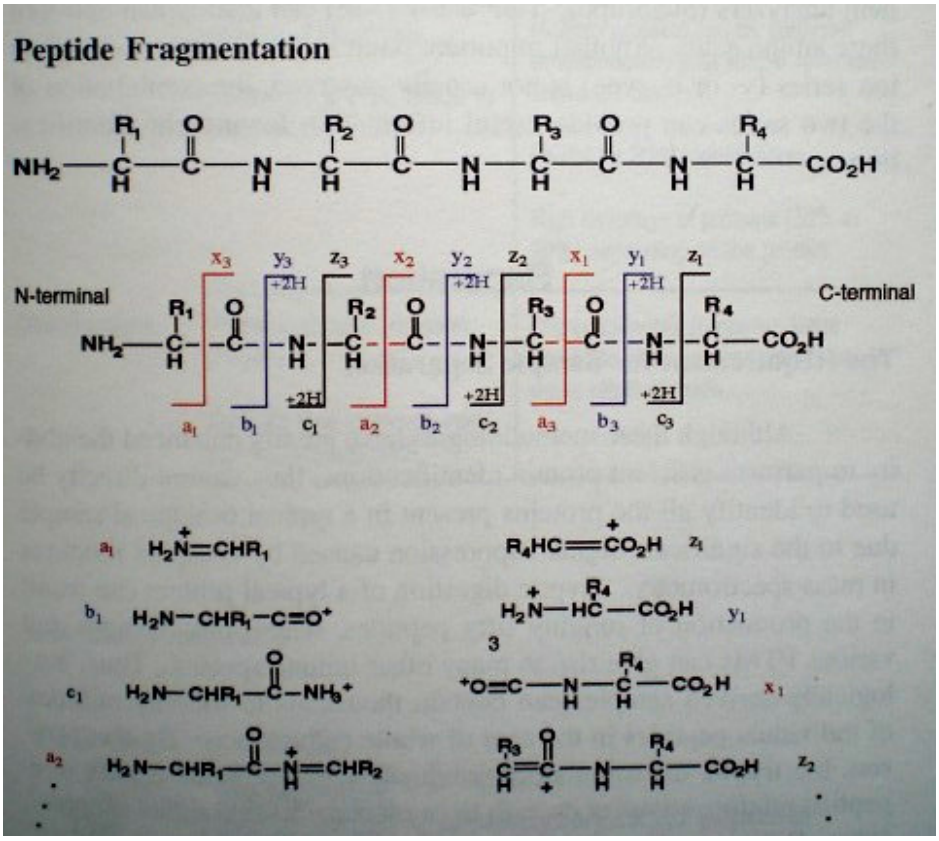


Figure 21. Peptide fragmentation results in the ions shown above. Collision-induced dissociation (CID) spectra often result in the dominant fragmentation at the amide bonds in the polyamide backbone, producing ions of the type b or y [**88**].

2.10 PROTEIN PROFILING

Mass spectrometry has rapidly become one of the most important tools for identification of proteins. Using ESI or MALDI-MS, proteolytic peptides can be ionized intact into the gas phase and their masses accurately measured. Based on this information, proteins can readily be identified using a methodology called protein mass mapping or peptide mass mapping, in which these measured masses are compared to predicted values derived from a protein database. Further sequence information can also be obtained by fragmenting individual peptides in tandem MS experiments.

2.10.1 Peptide Mass Fingerprint

The first approach for characterization of complex protein mixtures was separation by 2D-PAGE followed by in-gel tryptic digestion of visible protein spots and subsequent off-line analysis of each digestion mixture by MALDI-MS.

In this technique each protein can be uniquely identified by the masses of its constituent peptides, this unique signature being known as the fingerprint. In peptide mapping, also referred to as peptide mass mapping or peptide mass fingerprinting (PMF), proteins are identified by matching the list of observed peptide masses with a calculated list of all the expected peptide masses for each entry in a database. The enzyme trypsin is a commonly used protease that cleaves peptides on the C-terminal side of the relatively abundant amino acid arginine (Arg) and lysine (Lys). Thus, trypsin cleavage results in a large number of reasonably sized fragments from 500 to 3,000 Da, offering a significant probability for unambiguously identifying the target protein. The observed masses of the proteolytic fragments are compared with theoretical "in silico" digests of all the proteins listed in a sequence database [**Figure 22**]. The matches or hits are then statistically evaluated and ranked according to the highest probability.

Five laboratories independent from each other, simultaneously developed algorithms allowing database searching on the basis of peptide mass data [**89, 90, 91, 92, 93**]. Things that need to be considered for reliable protein identification are: the quality and relative intensity of the peaks, the mass accuracy and the coverage of the protein.

Clearly, the success of this strategy is predicted on the existence of the correct protein sequence within the database searched. However, the quality and content of such databases are continually improving as a result of genomic sequencing of entire organisms, and the likelihood for obtaining matches is now reasonably high. While exact matches are readily identified, proteins that exhibit significant homology to the sample are also often identified with lower statistical significance. This ability to identify proteins that share homology with poorly characterized sample makes protein mass mapping a valuable tool in the study of protein structure and function. This method of identification relies on the ability of mass spectrometry to measure the

masses of the peptides with reasonably accuracy, with typical values ranging from roughly 5 to 50 ppm (5 ppm = ±0.005 Da for a 1,000 Da peptide).

One obvious limitation of this methodology is that two peptides having different amino acid sequences can still have the same exact mass. In practice, matching 5-8 different tryptic peptides is usually sufficient to unambiguously identify a protein with an average molecular weight of 50 kDa. The other limitation of PMF is generally restricted to identifying simple protein mixtures from organisms with fully sequenced genomes.

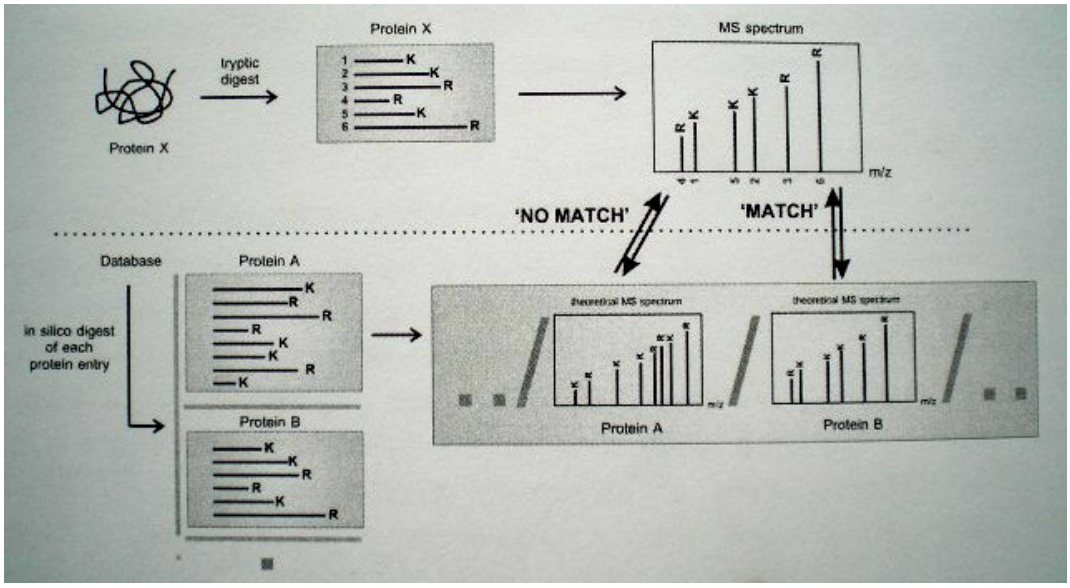


Figure 22. Protein identification through the comparison of tryptic peptides of an unknown protein to the theoretical digest of known proteins. This method does not apply to complex protein mixtures as only peptide masses are measured. Therefore, this method is used for protein mixtures resolved by electrophoresis or for the identification of single proteins. Because MALDI generates singly charged ions, it is generally used in this type of analysis [**94**].

2.10.2 Peptide Fragmentation Fingerprint

Important information is obtained by the acquisition of an MS-MS spectrum of one or more peptides. The collision-induced disassociation (CID) spectrum provides primary structural information which is not present in PMF and therefore, protein

identification is more reliable. The partial sequence information contained in tandem MS experiment is much more specific than simply using the mass of a peptide, since two peptides with identical amino acid contents but different sequences will exhibit different fragmentation patterns. As peptide masses are a fingerprint for a protein, the set of fragment ions generated by MS-MS act as a fingerprint for an individual peptide. In peptide fragmentation fingerprinting (PFF) [**95**], the un-interpreted fragment ion masses are used in a correlative database search to identify the peptide which yields a similar fragmentation pattern under similar conditions [**Figure 23**].

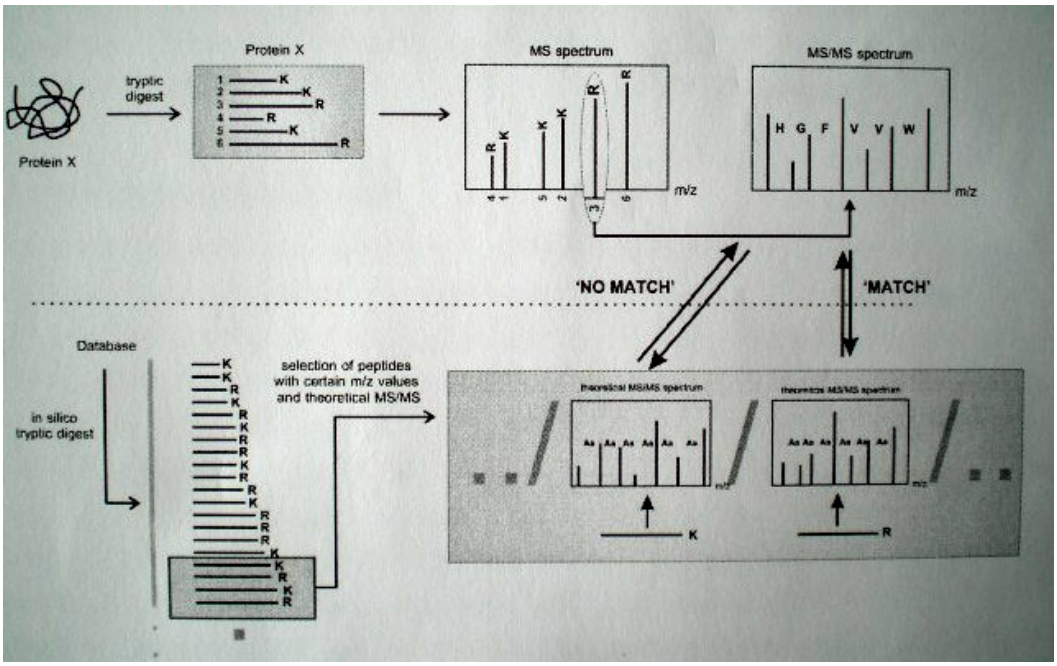


Figure 23. A peptide is selected and subjected to MS-MS fragmentation in the mass spectrometer. The obtained un-interpreted MS-MS spectrum is compared with *in silico* generated MS-MS spectra. The identified peptide is assigned to the protein [**96**].

Two or more peptides identified in this way are usually sufficient to unambiguously identify a protein. Several algorithms have been developed and the Mascot [**97**] search engine being the most popular. The Mascot method is a probability-based matching algorithm. The experimentally observed peaks are compared with the fragment ion masses calculated from peptide sequences in the database, and a score is calculated that reflects the statistical significance of the match. These programs have one main prerequisite: the sequence, either protein or nucleotide must be known.

Therefore, these database searching tools are excellent to study organisms of which the complete genome information is accessible.

Advantageous to PMF, beside more reliable protein identification, un-interpreted fragment ion masses may be used to search in databases that contain incomplete gene information (EST's) and the identification of proteins from complex mixture becomes possible.

2.10.3 *De novo* Sequence Analysis

The obtained fragment ion spectrum can also be interpreted, either manually or automatically, to derive partial *de novo* peptide sequences. The extracted short, unambiguous section of the amino acid sequence of the peptide, together with the measured mass of the precursor ion, is then used in standard database search queries [9**8**]. This correlative search method is more accurate than the above mentioned, because additional information is provided. Moreover, if longer stretches of amino acid sequences are determined from the MS-MS spectrum, non-correlative database searching programs, such as the BLAST [**99**] and FASTA [**100**] sequence alignment tools can be brought into play to identify the peptide and concomitantly the protein. However, the MS-MS spectra are mainly manually interpreted, because *de novo* sequencing algorithms are computationally intensive and require high quality data [**101**].

2.10.4 Sulfonic Acid Derivatization

Many derivatization strategies have been developed over the past 20 years to increase sensitivity and the amount of sequence information that can be obtained. Most of those derivatives are cationic or contain a very basic group because it improves peptide detection sensitivity. Recently, an entirely new derivatization strategy was developed specifically to enhance *de novo* sequencing of peptides under low internal energy conditions. This approach uses anionic sulfonic acid groups instead of basic or cationic groups, and that change dramatically improves the quality of sequence data obtained with MALDI MS. Sulfonic acid derivatives fragment well under both

electrospray ionization (ESI) and MALDI conditions, and easily interpreted fragmentation patterns are obtained with a variety of mass analyzers [**102**].

Tandem MS is much more versatile than single-stage MS for peptide sequencing because it is directly applicable to complex peptide mixtures [**103, 104**]. Unfortunately, peptide fragmentation patterns produced by this method are often difficult or impossible to interpret *de novo*.

Peptide fragmentation patterns produced at high-collision energies were strongly dependent on the presence and location of the basic amino acids Lys and Arg [**105**]. Peptides containing a single basic residue at the N or C-terminus are preferable for sequencing purposes because the resulting fragmentation patterns are relatively simple compared with the spectra obtained from peptides containing an internal basic residue.

Formation of cationic derivatives of peptides actually hinders peptide sequencing under low-energy collision conditions [**106**]. Structure-specific fragmentation of the backbone could be recovered simply by adding a proton to the precharged derivative to form a doubly charged ion. The added proton is free to protonate anywhere along the peptide backbone, and it induces fragmentation at the backbone amide bonds. Protonation of amide nitrogen atoms significantly lowers backbone amide bond strengths, which should favor cleavage of amide bonds under low energy conditions [**107**].

Single charged tryptic peptides formed either by ESI or MALDI do not fragment readily because the ionizing proton is localized on the basic side chain of the C-terminal Lys or Arg. There is not enough internal energy available to move the ionizing proton from the basic side chain to the peptide backbone.

To create a tryptic peptide with an extra proton, but not a double charge, we intentionally added a sulfonic acid group to the N-terminus. This strategy concedes that the side chain of the basic C-terminal residue will be protonated under MALDI conditions. The strong acid group was chosen so it would remain deprotonated under MALDI conditions, counter balancing the C-terminal positive charge. One additional proton, which is required to ionize the peptide for MS analysis, is more or less free to

randomly protonate backbone amide bonds, because the most basic site in the molecule is already occupied.

Sulfonation reagents react with amino groups. Arg-terminated tryptic peptides add a single sulfonate group to the N-terminus producing the desired monosulfonate derivative. Unprotected Lys-terminated peptides often form disulfonate derivatives, with one sulfonate group attached to the N-terminus and the second group attached to the amino group of the Lys side chain. Disulfonate derivatives are undesirable because they exhibit poor sensitivity in the positive-ion mode and relatively poor fragmentation behavior under negative- ion conditions. Negative-ion sequencing of disulfonate derivatives was not as reliable as positive-ion sequencing of Arg-terminated monosulfonate derivatives.

Guanidination of Lys side chains is a simple, one-step method to protect Lys while leaving the N-terminus of peptides free for subsequent sulfonation [**108, 109**]. Guanidination converts Lys to homoarginine, which differs from Arg by a single methylene in the side chain. Importantly, the gas-phase basicity of homoarginine is significantly greater than that of Lys, which improves the positive-ion response for Lys-terminated peptides [**110**]. Guanidination often improves protein sequence coverage in mapping experiments, and it does not complicate database searching because the added mass increment can be programmed into database search algorithms.

.

CHAPTER 3

MATERIALS AND METHODS

3.1 CELL CULTURING

Wild type *Rh. centenum* (DSM 9894) was obtained from the DSMZ microbial collection. A *Rh. centenum* Ppr deletion mutant was kindly provided by Prof. Carl Bauer. Liquid cultures of both *Rh. centenum* strains were grown anaerobically under tungsten illumination at 30°C in medium 27 described by DSMZ. Synchronous cultures, containing only vegetative cells, were obtained by passage down a volume of sterilized 150-212 µm sized glass beads (Sigma), and from this filtered fractionation 5 mL was used to inoculate 100 mL anaerobically prepared medium. Cultures were grown anaerobically at 37°C in the presence or absence of blue light (>390 nm and <510 nm) in addition of red light (>600 nm) for photosynthetic growth. Five independent cultures were conducted for each light condition. Cultures were harvested 48 hours after inoculation when the cells reached the end of the exponential growth phase.

3.1.1 Preparation of Medium 27

For 1 liter:	
Yeast extracts	0.3 g
Pyruvate Na	2 g
Ethanol	0.5 mL
Na_2 succinate	1 g
Fe (III) citrate (0.1 % in dH_2O)	5 mL
KH_2PO_4	0.5 g
$MgSO_4$ x $7H_2O$	0.4 g
NaCl	0.4 g

NH_4Cl	0.4 g
$CaCl_2$ x $2H_2O$	0.05 g
Vitamin B12	0.4 g
Trace element solution SL-6	1 mL
dH_2O	1050 mL

Vitamin B12

Biotin	10.0 mg
Nicotinamide	35.0 mg
Thiamine-HCl x 2 H_2O	30.0 mg
p-Aminobenzoic acid	20.0 mg
Pyridoxal hydrochloride	10.0 mg
Ca-pantothenate	10.0 mg
Vitamin B12	5.0 mg
Distilled water	100.0 mL

Trace element SL-6

$ZnSO_4$ x $7H_2O$	0.10 g
$MnCl_2$ x $2H_2O$	0.03 g
H_3BO_3	0.30 g
$CoCl_2$ x $6H_2O$	0.20 g
$CuCl_2$ x $2H_2O$	0.01 g
$NiCl_2$ x $6H_2O$	0.02 g
Na_2MoO_4 x $2H_2O$	0.03 g
Distilled water	1000 mL

Adjust pH to 6.8. Autoclave at 121°C for 15 min. Distribute 95 mL medium into 100 mL screw-capped bottles. Incubate in the anaerobic chamber over night in order to remove all oxygen in the medium. Close bottles with a rubber septum and screw tight. Inoculate organism from stock culture to fresh medium with sterile syringes under laminar flow.

3.1.2 Equipments

Microcentrifuge (Eppendorf), Ultra high speed centrifuge (Beckman Coulter), Sonicator (Branson), pH meter, 0,22 µm filter paper (Milipore), anaerobic chamber (COY Laboratory Products Inc.).

3.1.3 Chemicals

Urea (USB), thiourea (Fluka), CHAPS (Fluka), TritronX 100 (Bio-Rad), TCEP (Pierce), EDTA-free protease inhibitor cocktail (Roche), RNAseA (Sigma), DNAseI (Sigma), $MgCl_2$, Bovine Serum Albumin (Pierce), Coomassie Plus Reagent (Pierce), acetone (VWR), Dy Light^{TM} $_{649}$-Maleimide (Pierce).

3.1.4 Method of Sample Preparation

The sample cells were harvested by centrifugation for 20 minutes at 11,000 x g at 4°C. The cell pellet was resuspended and washed with 20 mL Ex H_2O and centrifuged for another 20 minutes at 4,000 x g at 4°C. The pellet was then resuspended in freshly prepared lysis buffer containing 9 M urea, 2 M thiourea, 2% (w/v) CHAPS, 1% (w/v) DTT, 10 mL 20 mM Tris-HCl pH 7. Subsequently 400 µL Complete EDTA-free protease inhibitor cocktail (Roche) was added to the suspension. Successively, the suspension was sonicated four times for 30 seconds with a 30% amplitude (1 second pulse on; 0.5 second pulse off) to facilitate cell lysis and protein solubilization. In the next step, the sample was transferred in 2 mL tubes and centrifuged for 10 minutes at 20,000 x g at 4°C. The supernatants were collected and 0.1 volume of 0.5 mg/mL DNAseI, 0.25 mg/mL RNAse A, 50 mM $MgCl_2$ was added. The sample was mixed gently and incubated for 1 hour on ice. After incubation, the sample was centrifuged for 1 hour at 75,000 x g at 4°C. The supernatant was then collected and aliquotted in 1,5 mL tubes and stored at -20°C.

Short-time storage (several hours to overnight) of extracts is often possible in the refrigerator (4°C). For a longer time, storage in a freezer at -20°C is preferred. However, repeated freezing and thawing of the sample must be avoided. Better make aliquots and thaw only once.

Sample must be holding on ice during the experiment to avoid carbamylation of proteins. Lysis buffer should always be prepared freshly.

Protein extracts should not be too diluted to avoid loss of protein due to adsorption to the wall of the vessel (glass or plastic). The minimum protein concentration should not be less than 0.1 mg/mL, and optimum concentration is 1-5 mg/mL.

3.1.5 Protein Assay with the method according to Bradford

A serial dilution of Bovine Serum Albumin (BSA) protein standards was prepared using **Table 1** as a guide to prepare a set of protein standards. The protein standards were preferably prepared in the same diluents as the sample(s). 0.05 mL of each standard or unknown sample was pipeted into appropriately labeled test tubes. 1.5 mL of the Coomassie Plus Reagent added into each tube and mix well. For the most consistent results, samples were incubated for 10 minutes at room temperature. With the spectrophotometer set to 595 nm, zero the instrument on a cuvette filled with 1.5 mL of the Coomassie Plus Reagent mixed with 0.05 mL of 20 mM Tris-HCl pH 7. Subsequently, measure the absorbance of all the samples. A standard curve was prepared in order to by plotting the measurement for each BSA standard and its concentration in µg/mL. The standard curve used to determine the protein concentration of each unknown sample.

3.1.6 Synthesis of Fluorescently Labeled Internal Standard (FLIS)

Protein extraction from wild type *Rh. centenum* cells, grown anaerobically under full visible light condition, was performed according to the method described above except 1,5 mg/mL of the reducing agent TCEP replaced DTT in the lysis buffer. Protein concentration was determined using the protein assay method of Bradford. The DyLight^{TM} $_{649}$ maleimide labeling reaction was performed in the dark. In brief, 1 mg of DyLight^{TM} $_{649}$ maleimide was diluted in 100 µL of DMSO. An aliquot of crude protein extract containing 2,46 mg protein was diluted with 1 mL lysis buffer (7 M urea, 2 M thiourea, 4% (w/v) CHAPS, 1 x Protease inhibitor cocktail, 1.5 mg TCEP/mL), and the pH was adjusted to 7.3 by the addition of 45 µL of 1.6 M Tris-HCl buffer, pH 7.14. The

optimal pH is 7.0–7.5 according to the manufacturer. The protein solution was added to the 20 µL DyLight^{TM} $_{649}$ dye solution, and the mixture was stirred at room temperature for 2 hours in the dark. The reaction was quenched by the addition of glutathione in two-fold molar excess of the DyLight^{TM} $_{649}$-maleimide dye (21µL of 20 mM glutathione in dH_2O), and the reaction mixture was allowed to stir at room temperature for an additional hour. The sample was then diluted ten-fold in lysis buffer, aliquotted, and stored at -20°C until use.

Table 1. Preparation of diluted albumin (BSA) standards.

Tube	Volume of Diluent	Volume and Source of BSA	Final BSA Concentration
A	0	300 µL of stock	2000 µg/mL
B	125 µL	375 µL of stock	1500 µg/mL
C	325 µL	325 µL of stock	1000 µg/mL
D	175 µL	175 µL of tube B	750 µg/mL
E	325 µL	325 µL of tube C	500 µg/mL
F	325 µL	325 µL of tube E	250 µg/mL
G	325 µL	325 µL of tube F	125 µg/mL
H	400 µL	100 µL of tube G	25 µg/mL
I	400 µL	0	0 µg/mL = Blank

Dilution Scheme for Standard Test Tube Protocols (Working range = 100-1500 µg/mL)

3.2 FIRST DIMENSION: ISOELECTRIC FOCUSING

3.2.1 Apparatus

Ready Strip^{TM} IPG strip, (Bio-Rad), pH 3-10 and pH 4-7, 0.5 × 3 × 170 mm, IEF cell Protean (Bio-Rad), wicks (Bio-Rad).

3.2.2 Chemicals

Urea (USB), CHAPS (Fluka), DTT (Fluka), glycerol, Bio-Lyte® Ampholyte buffer (Bio-Rad), mineral oil (Pharmacia), SDS (Sigma), iodoacetamide (Riedel-de Haen).

3.2.3 Reswelling of IPG Strip

Isoelectric focusing (IEF) was carried out by using IPG strips. Dehydrated IPG strips were directly reswollen with 100 µg protein extracts which were first mixed with IPG rehydration buffer (9 M urea, 2 M thiourea, 4% (w/v) CHAPS, 1% (w/v) DTT, 2% ampholytes and 0,5% (v/v) glycerol. Glycerol prevents the IPG strip from drying during the first dimension which is run at high voltages. Ampholytes are added in order to stabilize the pH on the strip and to help protein migration.

A final sample volume of 360 µL was pipetted into the grooves of the reswelling tray. The protective cover sheets were peeled off from the IPG strips and subsequently the strips were placed into the groove with the gel side down. Care is taken to avoid trapping air bubbles.

Finally, the strips were rehydrated for minimum 8 hours either actively under low voltage (50V) to improve entry of high *Mr* proteins or passively without any voltage application.

3.2.4 Isoelectric Focusing

After suitable rehydration time, the IPG strip is ready for the first dimension, with the proteins already uniformLy distributed within the gel matrix. The clear advantage of in-gel rehydration is the large volume of sample that can be applied. Temperatures of around 20^{o} C should be used and this should be kept constant as it is a variable and may affect the *pI* of sample proteins. Increasing the temperature too much above this may result in carbamylation of the proteins. Much lower temperatures may cause precipitation of components such as urea, and urea crystal can be formed.

IEF was performed using the Protean IEF Cell Unit. Prior to IEF, strips were covered with mineral oil to prevent them from drying. Humid wicks were inserted on the each pole of electrode strip plate before running 1D. Proteins were focused at a maximum of 10,000 Volt for 35,000 volt-hours based on the **Table 2**.

Table 2. Voltage changes versus time during isoelectric focusing of proteins.

Phase	**Voltage (V)**	**Current (mA)**	**Duration (min)**
1	↑ 150	2	30
2	↔ 150	2	60
3	↑ 300	2	30
4	↔ 300	2	45
5	↑ 3500	2	90
6	↔ 3500	2	540
7	↓ 500	2	10
8	↔ 500	2	540

3.2.5 Equilibration of IPG Strip

Equilibration is performed twice on a shaker. SDS at 2% (w/v) is sufficient for preparative protein loads. Increasing the SDS concentration improves the solubilisation of hydrophobic proteins [**111**]. DTT is necessary in a first step because after the proteins have been focused they have to be treated with the reductant again. In a second step, the iodoacetamide functions to react with any unreduced DTT, increase spot sharpness and improve protein identification using mass spectrometry. The first step in the equilibration is to reduce all of the cysteines present in the protein, this step is undertaken to reduce any disulphide bridges that may have formed between adjacent cysteine residues. This is a reversible process and oxidation may occur if reducing conditions are lost. For this reason, once we have carried out the reduction step we then treat the strips with iodoacetamide. A non-reversible reaction then occurs that places a functional group on the sulphide group, blocking it and ensuring that no further reactions can occur. 10 minutes incubation of each equilibration buffer was needed. Content of buffers were 6 M urea, 10% (v/v) glycerol, 2% (w/v) SDS in 50 mM Tris-HCl pH 8,8 , 1% (w/v) DTT for the first, and 5% (w/v) iodoacetamide for second one.

3.3 SECOND DIMENSION: SDS-PAGE

Equilibrated strips were placed on the home-casted vertical SDS-PAGE gels (12,5% acrylamide concentration) and sealed with 0.5% (w/v) agarose. The second dimension was performed on a Protean Plus Dodeca Cell (Bio-Rad) at 10 mA/gel for 15 minutes, followed by 20 mA/gel until the bromephenol blue front reached the bottom of the gel.

3.3.1 Equipments

2D electrophoresis tank for 2 to 4 gels (Bio-Rad Protean® II xi Cell), 2D electrophoresis tank for 12 gels (Bio-Rad Protean® Plus Dodeca™ Cell), power supplier (Bio-Rad Power Pac 200).

3.3.2 Chemicals

Acrylamide / Methylenebisacrylamide (National Diagnostic), Tetramethylethyleendiamine (Fluka), ammonium persulfate, isobutanol (Riedel-de Haen), agarose (Amersham Bioscience), bromephenol blue (Aldrich).

3.3.3 Preparation of 12% SDS-PAGE

a. For two gels;

- 26.5 mL MQ H_2O
- 20 mL 1.5 M Tris-HCl pH 8.8 + 0.4% (w/v) SDS
- 30.8 mL 30% (w/v) acrylamide + 0.8% (w/v) methylenebisacrylamide
- 90 µL Temed (tetramethylethyleendiamine)
- 250 µL 10% (w/v) APS (ammonium persulfate). APS solution should be prepared freshly just before use.

b. For twelve (dodeca) gels;

- 380 mL MQ H_2O
- 285 mL 1.5 M Tris-HCl pH 8,8 + 0.1% (w/v) SDS
- 475 mL 30% (w/v) acrylamide + 0.8% (w/v) methylenebisacrylamide
- 720 µL Temed (tetramethylethyleendiamine)
- 2 mL 10% (w/v) APS (ammonium persulfate). APS solution should be prepared freshly just before use.

Acrylamide and methylenebisacrylamide are able to polimerize with the contribution of Temed and APS. Make sure solution gets homogeneously mixed [**Figure 24**]. It prevents unequal gel solidification. Then load the solution rapidly on the vertical SDS-gel cassettes. Remove air bubbles from the gel with clean thin capillary wire before solidification.

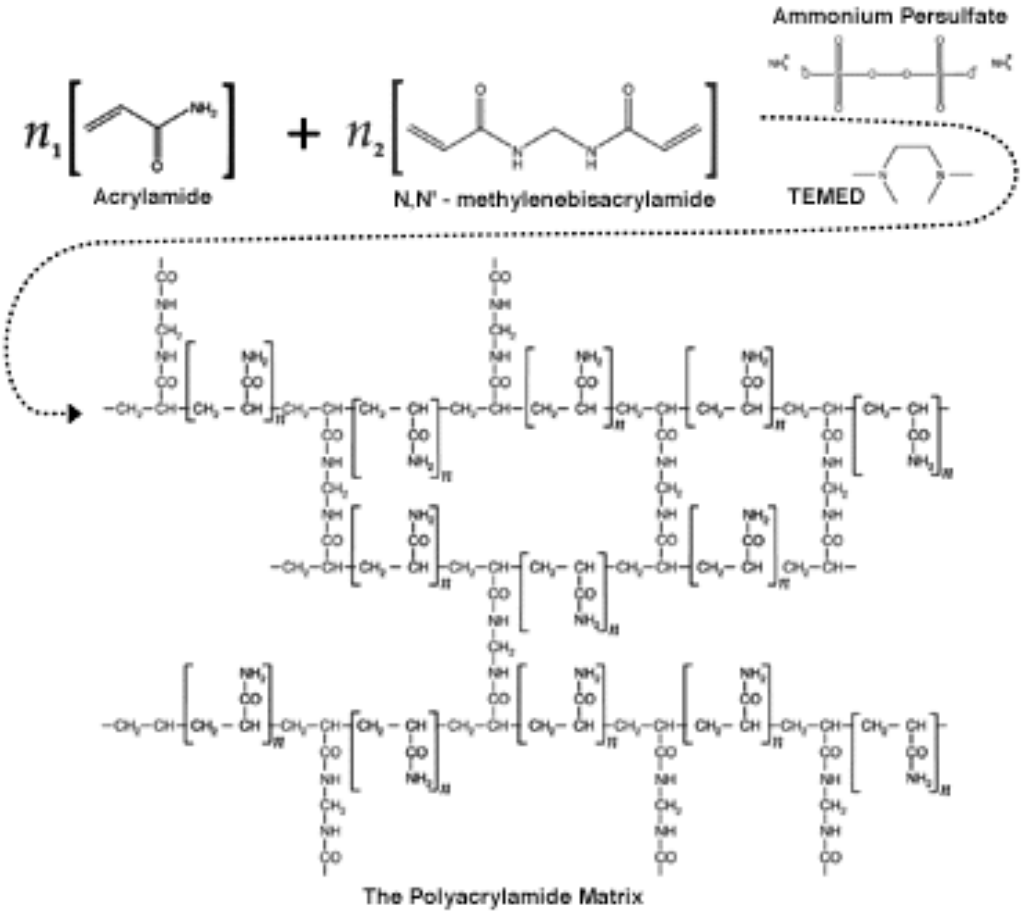


Figure 24. The polymerization of acrylamide and N,N'-methylenebisacrylamide to form a cross-linked polyacrylamide gel [**112**].

Finally, add isobutanol + water solution on top of the gel. Isobutanol ensures a linear gel surface and prevents air to make contact with gel. Air can distrupt polymerization of polyacrylamide, and dry the gel surface. Keep the gel 1 to 2 hours at room temprature, and then keep it in 4°C cold room overnight in order to complete solidification of gels.

After the IPG strip transfer on the 2-DE gel, add 0.5 % (w/v) agarose + bromephenol blue on top of it. It enhances the transfer of proteins between the first and second dimension. Use spatula and make sure that all air bubbles are removed between SDS-PAGE gel and gel strip.

3.3.4 Preparation of 0.4% (w/v) Agarose Gel

- 30 mL dH_2O
- 10 mL 50 mM Tris-HCl pH: 8,8
- 0,16 g agarose
- Trace amount of bromophenol blue reagent

Solution was boiled to solubilize the agarose. Then keep it as a solid form under room temperature. Solution is heated until boiling just before usage.

3.3.5 Running of 2-DE Gel

Run the gel at 400 V (~10V/cm^2) and 15mA/gel for first 15 minutes and increase current to 30 mA / gel till the end. Proteins were run until the bromophenol blue front eluted from the bottom of the gel. Protein transfer from the first dimension (IPG-strip) to the second (SDS-PAGE gel) was performed slowly in order to avoid streaking and to minimize loss of high M_r proteins.

3.4 SPOT DETECTION

3.4.1 Equipments

Shaker, densitometers (Bio-Rad, GS 800 calibrated densitometer), fluorescence scanners (Bio-Rad, Pharos FX Plus™ molecular imager), Proteomweaver™ (Definiens).

3.4.2 Chemicals

MeOH (BioSolve), H_3PO_4 (Acros Organics), $(NH_4)_2SO_4$, CCB G-250 (Fluka), MQ water, ethanol (), Flamingo™ fluorescent stain (Bio-Rad), Krypton™ fluorescent stain (Pierce).

3.4.3 Staining with CBB

The colloidal CBB solution was prepared as followed:

- MeOH 340 mL
- H_3PO_4 35 mL
- $(NH_4)_2SO_4$ 170 g
- CCB G-250 2 g
- MQ water fill up until 1 L

Following electrophoresis, the proteins are fixated on the gels by using a fixative solution for 2 hours on orbital shaker at room temperature. Then it is replaced with CBB stain. Subsequently the gels are incubated overnight. Finally, the gels are washed with 30% methanol for an hour to remove excess CBB dye and reduce background staining.

Fixative solution:

- Ethanol 500 mL
- H_3PO_4 23.5 mL
- MQ water fill up until 1 L

3.4.4 Staining with Flamingo Fluorescent Dye

Following electrophoresis, the proteins on the gels are fixed using 40% (v/v) ethanol, 10% (v/v) acetic acid for 2 hours on an orbital shaker at room temperature. Then the solution is replaced with 1x Flamingo™ stain. The gels were incubated for at least 3 hours under the dark condition.

3.4.5 Staining with Krypton Fluorescent Dye

After electrophoresis, the protein gels were fixed in 40% (v/v) ethanol, 10% (v/v) acetic acid for at least 30 minutes. The fixing solution is refreshed once, and gently agitated for another 30 minutes. Afterwards, the fixing solution was discarded, and gels were agitated with distilled water for 5 minutes. After that, water was replaced with 1x Krypton™ protein stain and gels were incubated on orbital shaker at room temperature under dark conditions over night. Fluorescent stain was replaced with destaining solution (5% (v/v) acetic acid) and agitated for 5 minutes. Finally, the solution was substituted with distilled water to agitate for 15 minutes.

3.4.6 Scanning Gels After Staining

Before 2-DE gels can be analyzed with an image evaluation system, they must be digitized. The most commonly used devices are camera systems, densitometers, phosphor imagers, and fluorescence scanners.

Both Flamingo and Krypton stained gels were scanned by a Pharos FX Plus™ molecular imager using a 532 nm excitation laser. The same gels were also scanned with a 635 nm excitation laser in order to visualize the DyLight™ $_{649}$ internal protein standard. Coomassie Brilliant Blue stained gels were scanned via GS-800 calibrated densitometer.

3.5 SPOT ANALYSIS

3.5.1 Spot Quantification

Proteomweaver (Definiens, Munich, Germany) was used to analyze 2-DE gel images. Following cropping and rotation to align the gel images, automatic image warping, spot detection, and matching were performed for both images of each gel (total protein (λ_{ex} = 532 nm) and $FLIS_{649}$ (λ_{ex} = 635 nm). Matching and spot detection were carefully reviewed and edited manually. For images of total protein, a set of spots that could be found in at least 4 gels per replicate group were selected. For images of the $FLIS_{649}$, sets containing approximately 30 spots representative for the entire *pI* and molecular weight range of the gel were selected.

Spot intensities given as spot volumes were normalized according to the following methods. First, a pre-normalization step is done individually for every gel, intended to bring the absolute numeric values of spot intensities into a reasonable range. The pre-normalization takes the intensity of a "normalization reference spot" and sets this intensity to 1. All other spots in a gel are adjusted accordingly. The "normalization reference spot" has an intensity higher than 95% of all spot intensities. This way, a large spot always has an intensity factor of about 1, huge spots have intensities larger than 1, and small spots less than 1. The introduction of an absolute scaling is intended to diminish strong variation between experiments. For comparison, two different normalization algorithms were subsequently applied using (i) the fluorescently labeled internal standard and (ii) the default normalization algorithm implemented in the Proteomweaver software. The latter applies a pair matched based (PMB) normalization step to the pre-normalized values. It is performed after calculating pair matches between spot volumes from two gels. In brief, this algorithm first computes a normalization factor between all pairs of gels for which a pair match exists: for every primary or hand-match in the pair match, the quotient between the pair matched spots is calculated. The normalization factor is the median of these quotients. Then it tries to compute for every gel an intensity factor that makes all the normalization factors as close to one as possible. This intensity factor is calculated by a linear equation system using 'Singular Value Decomposition' (SVD). As for most normalization methods, PMB-normalization only works as long as the percentage of up- or down-regulated spots is less than 50%.

For the $FLIS_{649}$ normalization method, pre-normalized values were exported to a spreadsheet program (Microsoft Excel 2003, Microsoft, Redmond), after which each individual spot volume obtained at the 532 nm excitation wavelength was divided by the average (pre-normalized) value of 30 representative spots of the $FLIS_{649}$ output of the same gel.

No background subtraction algorithm was performed for neither normalization method. Finally, all expression data were further analyzed in Excel as described below.

3.5.2 Statistical Analysis

For each *Rh. centenum* strain (wild-type and Ppr deletion mutant), blue light (experimental) conditions were compared to red light (control) conditions. The mean, SD, CV and log volume ratio were calculated for each protein spot across the different replicate gels in the control and experimental group. Regulated protein spots were detected as followed: the mean spot volumes from both light conditions were compared through a two-sample T-test and a 95% confidence interval for experimental variation was set out of the distribution of the log volume ratios. Protein spot with a t-statistic $\leq$ 0.05 and a log volume ratio which resides outside the confidence interval, were assigned to be significantly up or down regulated. The normality of the distribution of log volume ratios were tested with the Kolmogorov-Smirnov test and Q-Q plots using SPSS (Chicago, IL). The Kolmogorov-Smirnov statistic tests the hypothesis that the data are normally distributed. A low significance value (generally less than 0.05) indicates that the distribution of the data differs significantly from a normal distribution. A Q-Q normal plot refers to the sample quantiles plotted against the theoretical quantiles in a normal distribution. A resulting straight line implies that the sample has a normal distribution.

3.6 MASS SPECTROMETRY

3.6.1 Equipments

Water bath (Memmert), incubator (Sellecta), Speed Vac (Thermo Savant), microcentrifuge (Eppendorf), 4700 Proteomics Analyser (Applied Biosystems), ESI-Q-Trap mass spectrometer (Applied Biosystems/ MDS Sciex), Ultimate^{TM} nano-HPLC (LC Packings).

3.6.2 Chemicals

Milli-Q water, ammonium hydroxide (Merck), *O*-methylisourea hemisulfate (Acros), ammonium bicarbonate (Valkenswaard), acetonitril (Valkenswaard), trypsin (Promega), 4-sulfophenyl isothiocyanate (Aldrich),α-cyano-4-hydroxycinnamic acid (Sigma).

3.6.3 'In-Gel' Digestion

Differential protein spots were excised from the gel and washed twice with 150 µL of 200 mM ammonium bicarbonate in 50% ACN (30 min at 30°C). After drying at room temperature, for 10 min, a volume of 10 µL digestion buffer (50 mM ammonium bicarbonate, pH 7.8) containing 200 ng modified trypsin per microliter was added to the dried gel spots and the tubes were kept on ice for 45 min to allow the gel pieces to be completely soaked with the protease solution. Then 20 µL of digestion buffer was added. Digestion was performed overnight at 37°C, after which the supernatants were recovered and the resulting peptides were extracted twice with 30 µL of 60% ACN/0.1% TFA. The extracts were pooled and dried in the SpeedVac. The peptides were redissolved in 12 µL 0.1% formic acid.

3.6.4 Mass Spectrometric Analysis

A 4700 Proteomics Analyzer (Applied Biosystems) with TOF/TOF optics was used for all MALDI-MS and MS/MS applications. Samples were prepared by mixing 0.5 µL of the sample with 0.5 µL matrix solution (7 mg/mL alfa-cyano-4-hydroxy-

cinnamic acid in 50% ACN/0,1 % TFA), and spotted on a stainless steel 192-well target plate. They were allowed to air-dry at room temperature, and were then inserted in the mass spectrometer and subjected to mass analysis.

If the identification, based on MALDI TOF/TOF mass spectral data, was unconclusive, the peptide mixture was separated by nano-HPLC and detected on-line by an ESI-Q-TRAP mass spectrometer.

3.6.5 *De novo* Sequencing

3.6.5.1 'In-Gel' Guanidination

Guanidination was performed according to [**113**]. 5 µL of Milli-Q (MQ) water, 11 µL of 7 N ammonium hydroxide, and 3 µL of a freshly prepared 7.5 M *O*-methylisourea hemisulfate solution was added to the excised spots. The samples were incubated at 65°C for 2 h, after which the guanidinated samples were taken from the oven and the remainder of the solution was discarded. The gel pieces containing the guanidinated samples were desalted and destained in one step. Two washes using 150 mL of 200 mM ammonium bicarbonate in 50% ACN (30 min at 30°C) were performed and subsequently the gel pieces were dried in a SpeedVac.

3.6.5.2 Trypsin Digestion and Sulfonylation

A volume of 8 µL digestion buffer (50 mM ammonium bicarbonate,pH 7.8) containing150 ng modified trypsin per microliter was added to the dried gel spots and the tubes were kept on ice for 45 min to allow the gel pieces to be completely soaked with the protease solution. Digestion was performed overnight at 37°C, the supernatants were recovered and the resulting peptides were extracted twice with 35 µL of 60% ACN/0.1% TFA. The extracts were pooled and dried in the SpeedVac. The peptides were redissolved in 4 mL of 12.5 mM ammonium bicarbonate and 50% ACN/MQ, and 2 µL was mixed with 2 µL of the sulfonation solution. The sulfonation reagent was prepared by dissolving 10 mg 4-sulfophenyl isothiocyanate (SPITC) in 1 mL 12.5 mM ammonium bicarbonate. The tubes were briefly vortexed and reacted for 30 min at 37°C. Upon sulfonation, the samples were not desalted, a fraction of the samples was

mixed with matrix solution (7 mg/mL α-cyano-4-hydroxycinnamic acid in 50% ACN/0,1 % TFA) and spotted on the MALDI plate.

3.6.5.3 MS and MS/MS

Samples for *de novo* sequencing were ananlyzed using the Proteomics Analyzer as described in 3.6.4.

3.7 PROTEIN IDENTIFICATION

The genome sequencing project of *Rh. centenum* is currently still in progress **[114]** and is expected to be released to the public in the beginning of 2008. However, we were kindly given access to the annotated genome by Prof. Carl Bauer. The database was formatted to make it accessible for the database searching program MASCOT **[115]**. MS data are submitted to Mascot in the form of a peak list, which is then compared with a calculated list of all the expected peptide and/or fragment masses for each sequence entry in the database. Mascot uses a probability based scoring algorithm to report statistically significant matches. If an exact match is not present in the database, the highest scoring matches will be those entries which exhibit the greatest homology.

Protein identification using manually derived *de novo* sequences was based on finding exact matches within the *Rh. centenum* protein sequence database.

CHAPTER 4

RESULTS

4.1 BACTERIAL GROWTH

Liquid cell cultures of wild-type (wt) *Rh. centenum* and its Ppr deletion mutant (ΔPpr) showed similar anaerobic photosynthetic growth at all light conditions, as shown in **Figure 25**.

4.2 2-DE

Protein extracts from wild-type (wt) *Rh. centenum* and its Ppr deletion mutant (ΔPpr) were subjected to 2-DE to study blue-light photoresponses of a phototrophic bacterium in which a particular photoreceptor has been deleted. 2-DE was performed using broad range pH strips (pH 3-10) for the first dimensional separation and 12,5% polyacrylamide gels were used in the second dimension. For comparison, two different normalization methods were applied based on (i) a fluorescently labeled internal protein standard (FLIS) and (ii) the default normalization algorithm implemented in the Proteomweaver analysis software (Pair Matched Based), as described in chapter 3 (section 3.5.1). Following computational analysis, we detected approximately 500 protein spots per gel using a fluorescent staining protocol which is sensitive and compatible with subsequent mass spectrometry. Protein spot patterns on 2-DE gels appeared very similar upon visual inspection. However, differences could be detected by using the image analysis software Proteomweaver and subsequent statistical analysis of spot volumes in Excel. 2-DE gels of both strains grown under red and/or blue light conditions are presented in **Figures 26 to 29**.

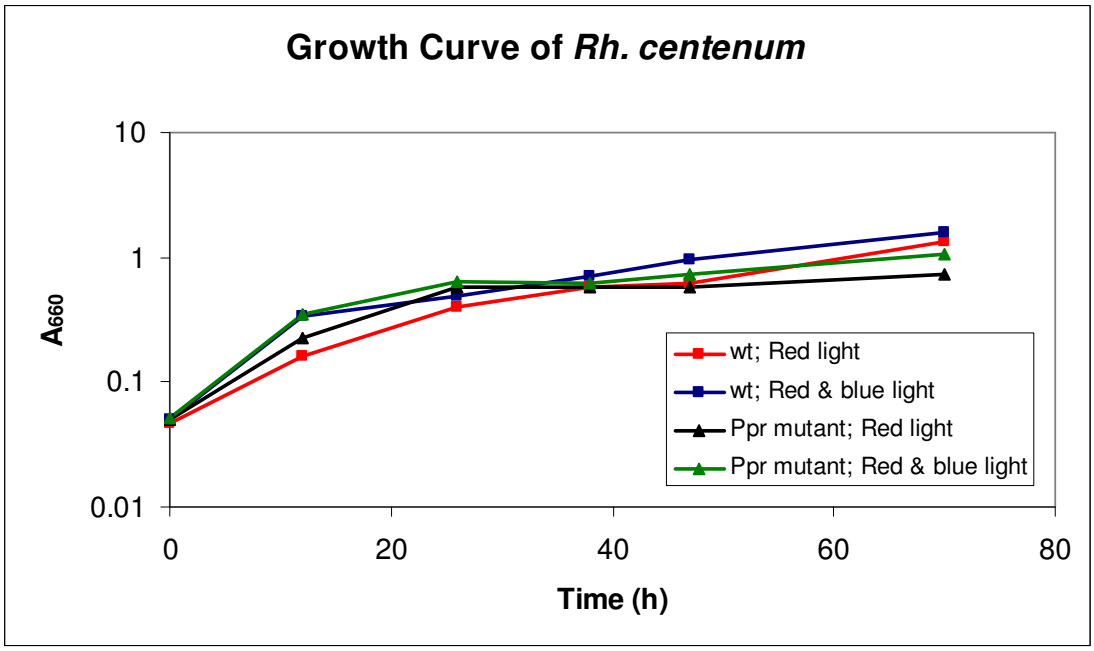


Figure 25. Typical growth curve of *R. centenum* cells at different growth conditions, obtained by turbidity measurements at 660 nm. Liquid cell cultures were grown anaerobically at 37°C in the presence or absence of blue light (>390 nm and <510 nm) in addition of red light (>600 nm) for photosynthetic growth.

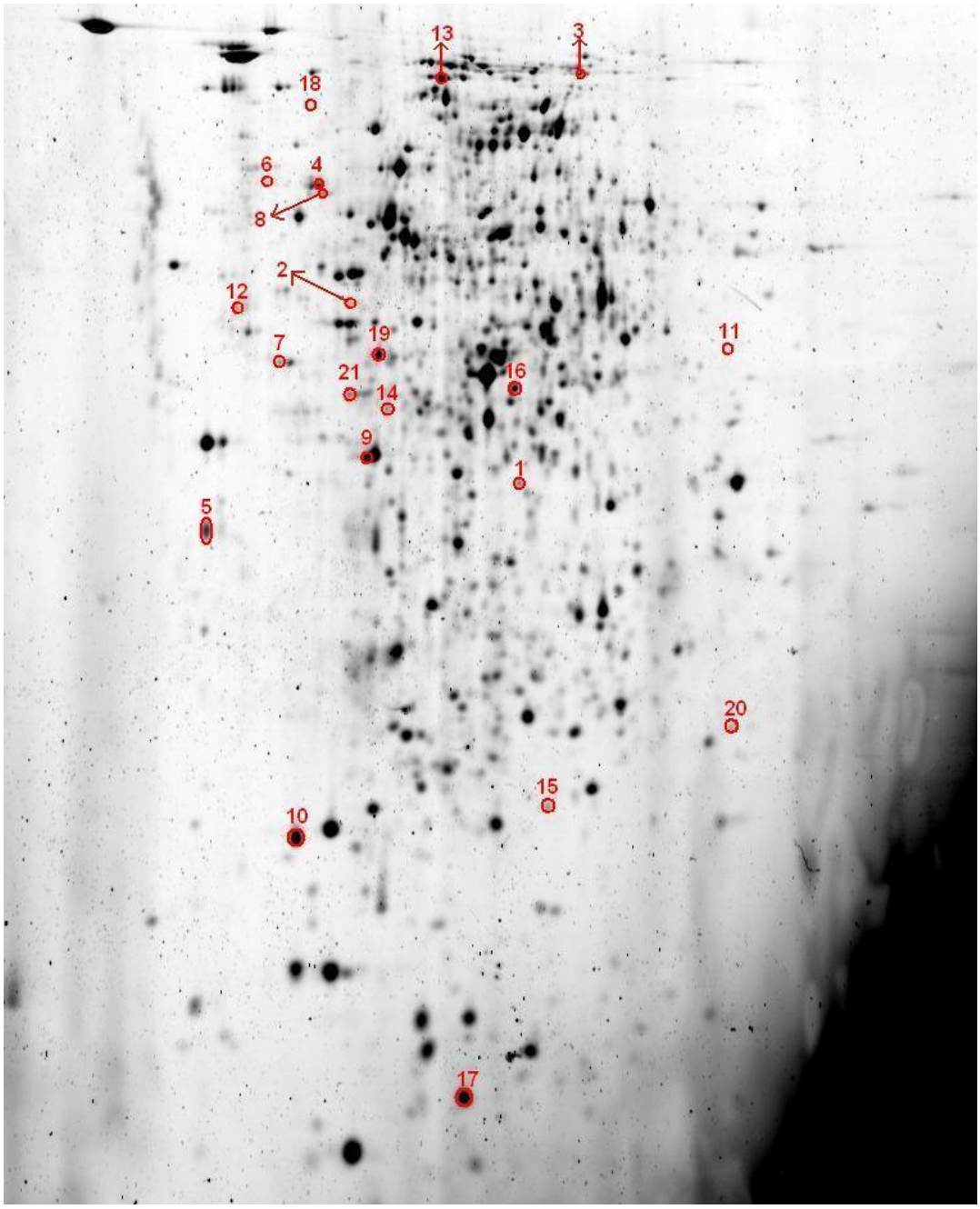


Figure 26. 2D-PAGE map of *Rh. centenum* ΔPpr protein extracts, after growth under red light conditions. 100 µg protein was loaded on the IPG-strip with a pH gradient of 3-10. Protein spots displaying significant up or down regulation are marked and numbered.

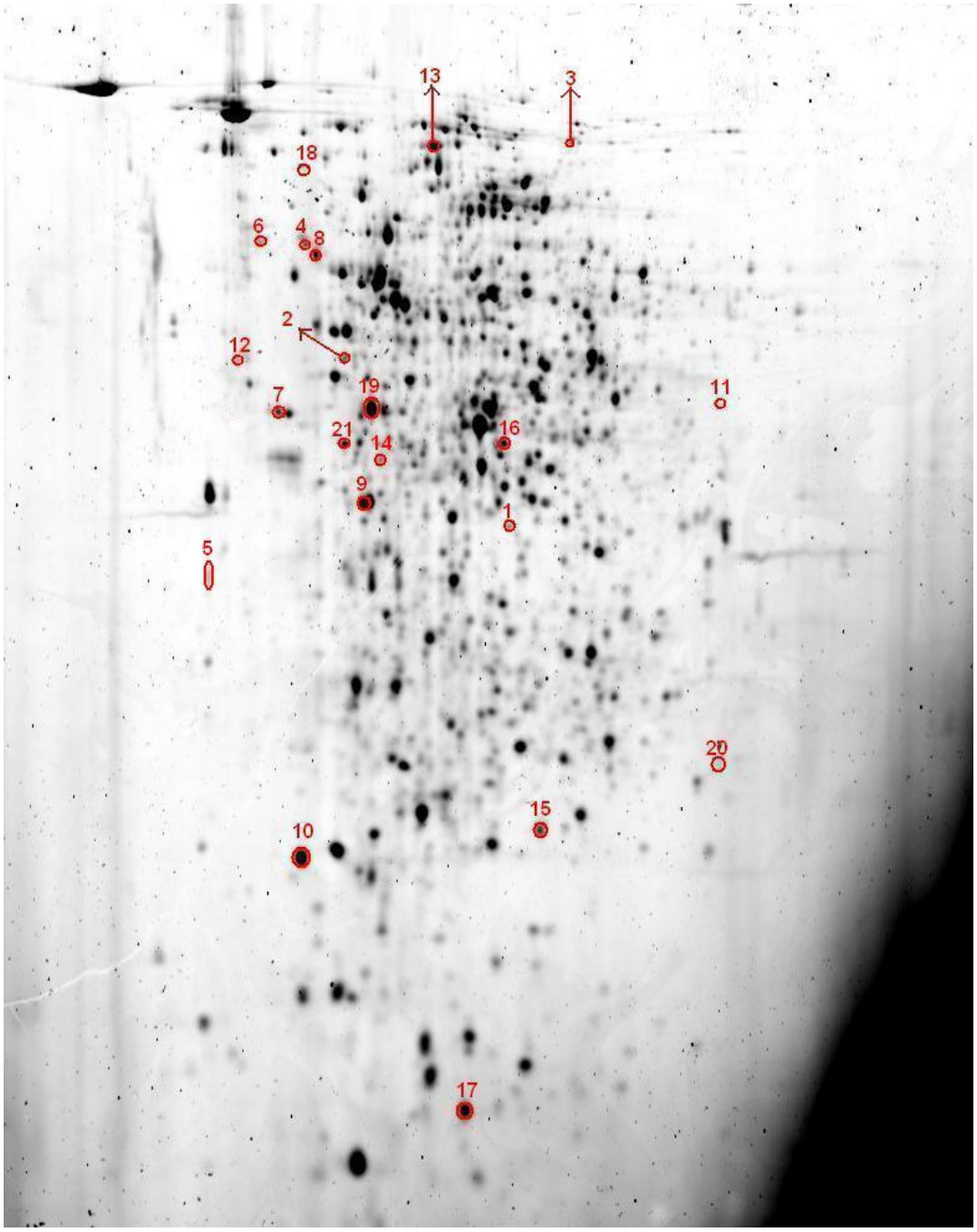


Figure 27. 2D-PAGE map of *Rh. centenum* ΔPpr protein extracts, after growth under red and blue light conditions. 100 µg protein was loaded on the IPG-strip with a pH gradient of 3-10. Protein spots displaying significant up or down regulation are marked and numbered.

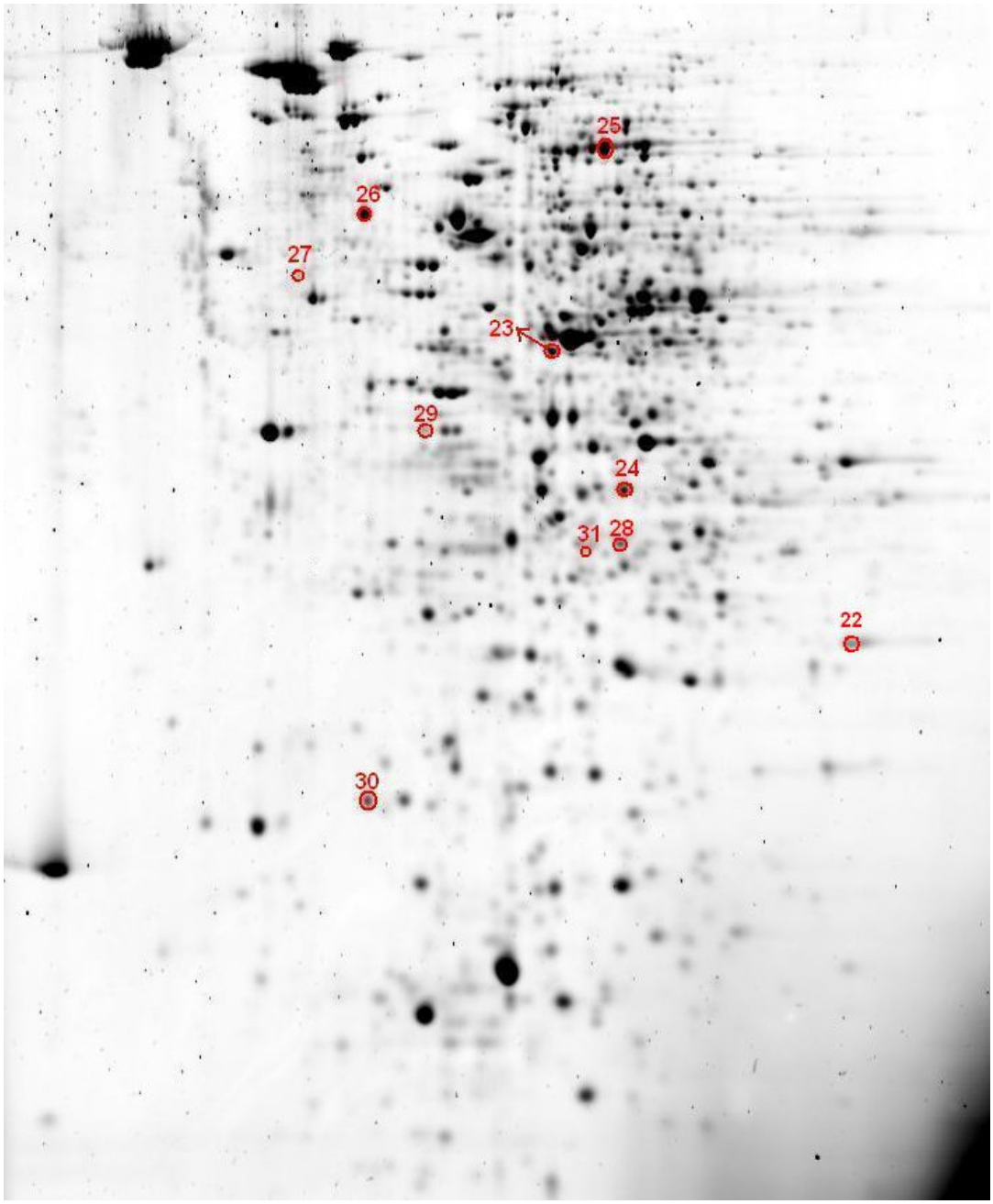


Figure 28. 2D-PAGE map of *Rh. centenum* wt protein extracts after growth under red light conditions. 100 µg protein was loaded on the IPG-strip with a pH gradient of 3-10. Protein spots displaying significant up or down regulation are marked and numbered.

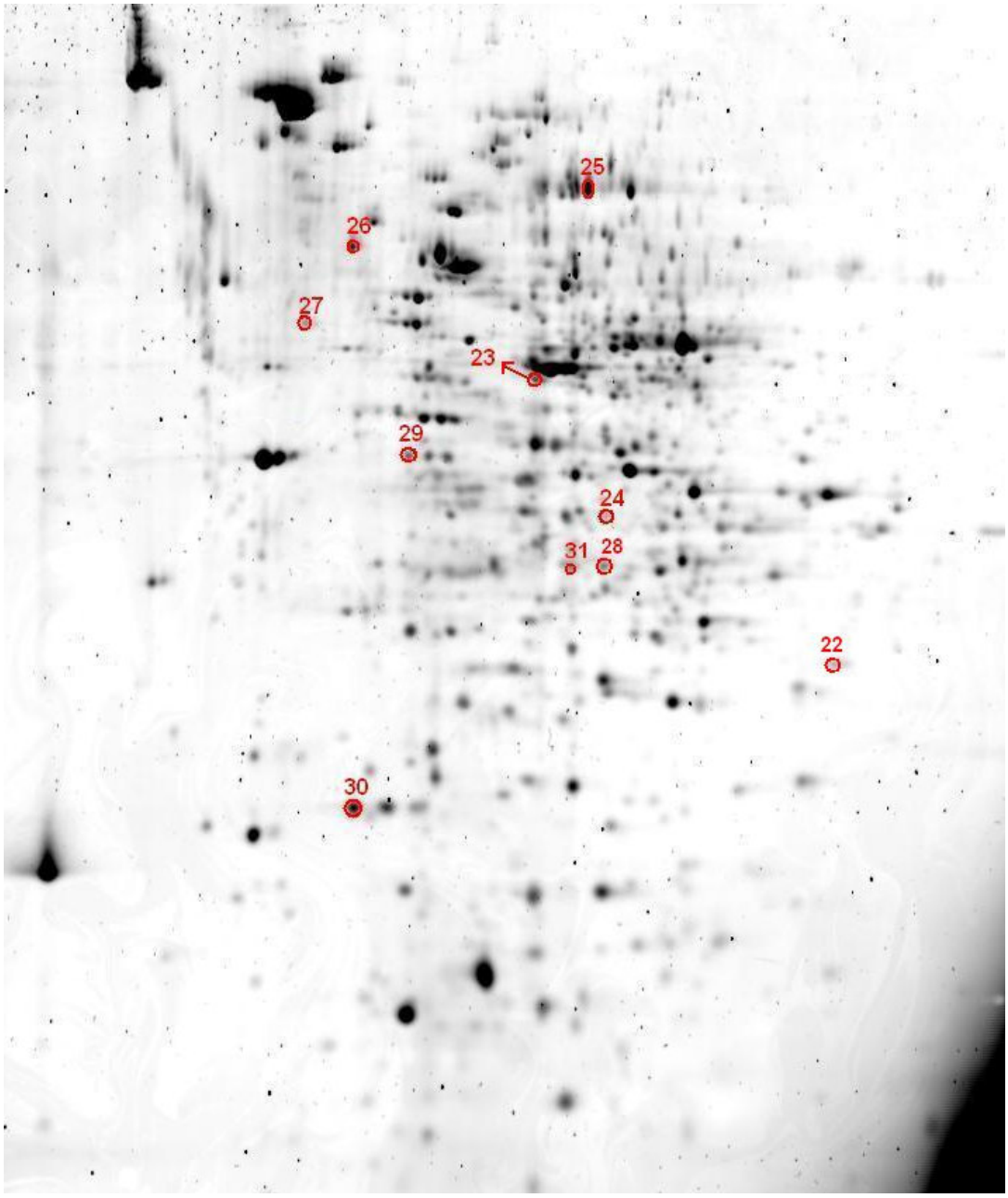


Figure 29. 2D-PAGE map of *Rh. centenum* wt protein extracts after growth under red and blue light conditions. 100 µg protein was loaded on the IPG-strip with a pH gradient of 3-10. Protein spots displaying significant up or down regulation are marked and numbered.

4.2.1 Quantitative Image Analysis

4.2.1.1 *Rh. centenum* ΔPpr

After careful evaluation of the gel images, only 4 control gels and 3 treated gels and their respective FLIS images were retained for further analysis. The reproducibility of the 2-DE protein profiles is high. Analysis of variability in protein abundances revealed a correlation coefficient of 0.97, and was not influenced by any normalization method used. Subsequently, we found that the coefficient of variation (CV) of all spots over the replicate gels was in the same range (5-74%) and 95% of the matched spots had a CV <50%.

In order to detect significantly different spots, we applied a Student's t-test to each data set which resulted in different amounts of significant ($p < 0.05$) spots depending on the normalization method used: 90 spots were significantly detected after PMB-normalization and 35 spots after FLIS-normalization. Next, we tried to determine thresholds above which a change in expression levels can be considered significant at a defined confidence level. Therefore we examined the frequency distribution of the log volume ratios (between treated and control gels, *i.e.* blue vs. red light conditions) of average spot volumes. After setting a threshold of two times the SD (2σ) only 19 significantly regulated spots (13 up, 6 down) remained for the PMB-normalized data and 13 spots (6 up, 7 down) for the FLIS-normalized data. However, for none of the normalization methods applied, the distribution of the log volume ratios was significantly normal ($p > 0.05$), which would make the use of "SD based" threshold statistically unjustified. On the other hand, the deviation from normality for all data sets was not visible by merely examining the distribution diagram and the normal Q-Q plot [**Figure 30A** and **30B**], which lead us to assume that the data were normally distributed. In addition, when those 'SD based' significant spots (19 out of 481 spots) were excluded from analysis, the PMB-normalized log volume ratios became significantly normal ($p = 0.2$), not like in the case of the FLIS-normalized data ($p = 0.034$). For the latter to become significantly normal, significant spots detected by the Students' t-test (35 out of 481 spots) had to be omitted from analysis ($p = 0.057$). Of the spots assigned as 'differentially expressed with statistical significance' there is an overlap of 11 protein spots between the two normalization approaches as depicted in [**Figure 32A**].

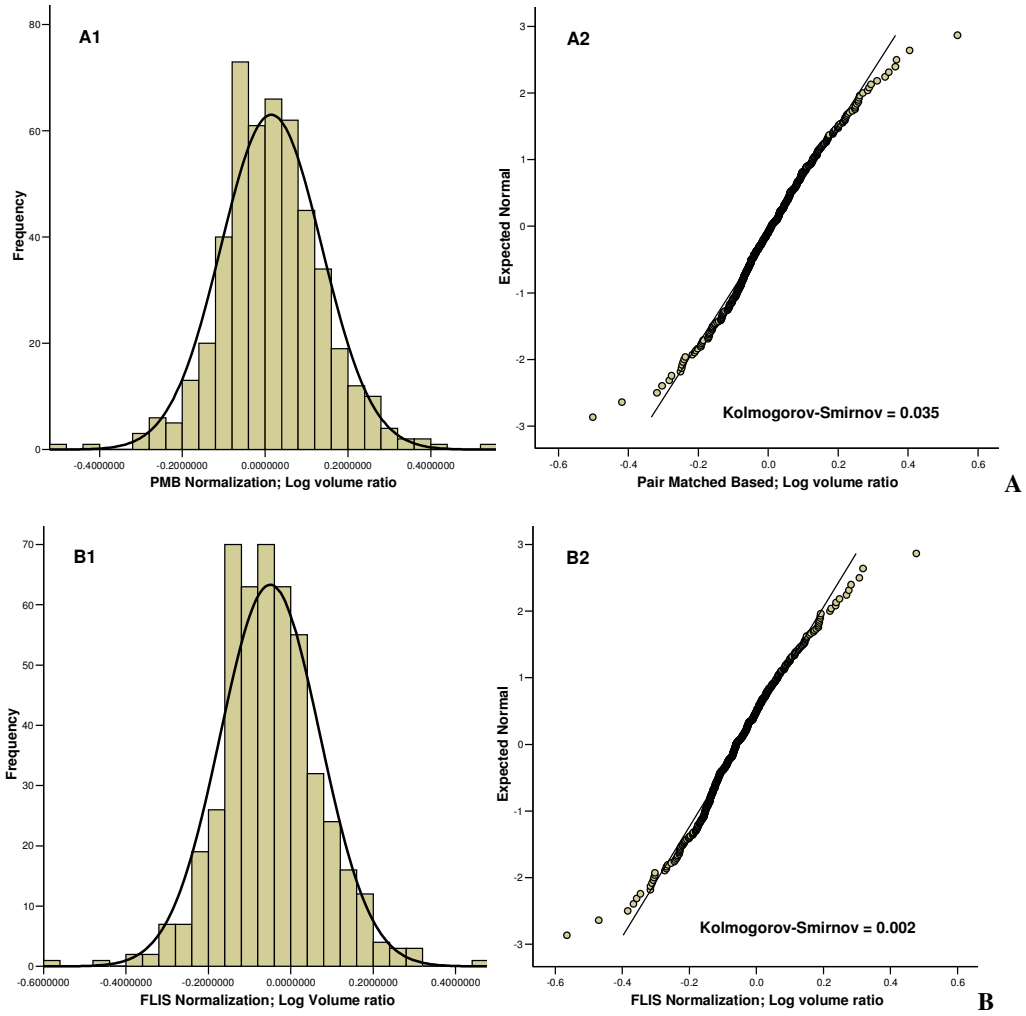


Figure 30. Histograms (left) and Q-Q normal plots (right) of the distribution of the log volume ratios from the quantification of 481 spots from the set of 2-DE gels (control and treated; n = 7) of *Rh. centenum* ΔPpr extracts, of which the values were normalized using two different methods: (**A**) Pair Matched Based normalization and (**B**) normalization with the average of 31 representative FLIS-spots. *p*-Values of the Kolmogorov-Smirnov normality test are displayed in each of the Q-Q normal plots (right).

4.2.1.2 *Rh. centenum* wt

Also for wild-type *Rh. centenum* extracts the reproducibility of the 2-DE protein profiles was high, with a calculated correlation coefficient of 0.96, for both normalization methods used. However, the range of CV of all detected spots after FLIS-normalization (17-105%) was somewhat shifted to higher values compared to the PMB-normalized data (4-83%), and 95% of the matched spots had a CV <56% and <46% respectively.

Following the same procedures as for the 2-DE gels from the *Rh. centenum* ΔPpr deletion mutant strain, *i. e.* using t-statistics, we detected 72 significant spots for the PMB-normalized data and 138 spots for the FLIS-normalization. After setting a 95% confidence interval which accounts for experimental variation, only 10 significantly regulated spots (4 up, 6 down) remained for the PMB-normalized data and 7 spots (0 up, 7 down) after FLIS normalization. Also in this case, the log volume ratios of both 'Pair matched based' and FLIS-normalized data did not show a significant normal distribution according to the Kolmogorov-Smirnov normality test as shown in [**Figure 31A** and **31B**]. We obtained a normal distribution ($p > 0.2$) of the log volume ratios of the PMB-normalized data by omitting 'SD based' significant spots (10 out of 329). Exclusion of 'SD based' (7 out of 329) or Students' t-tested (138 out of 329) significant spots on the FLIS-data did not transform the log volume ratios into a normal distribution according to the Kolmogorov-Smirnov normality test ($p = 0.04$ and $p = 0.01$ respectively).

Of the spots assigned as 'differentially expressed with statistical significance' there is an overlap of 6 protein spots between the two test approaches [**Figure 32B**].

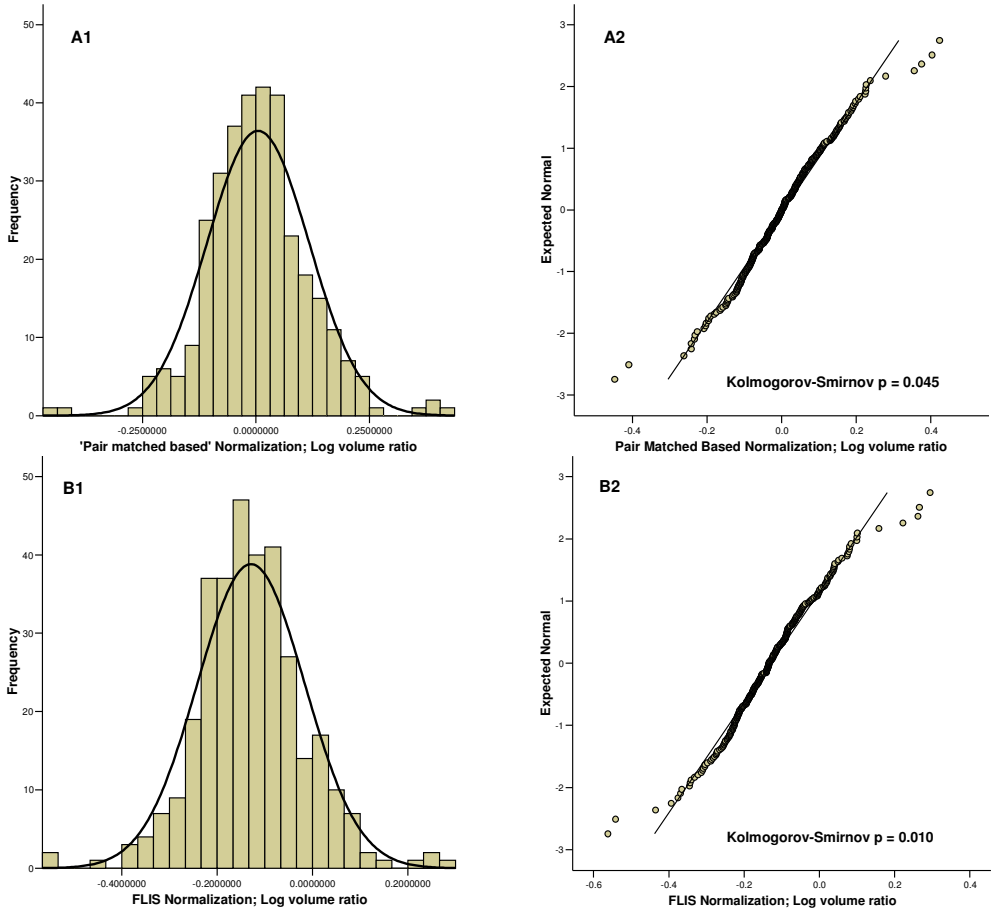
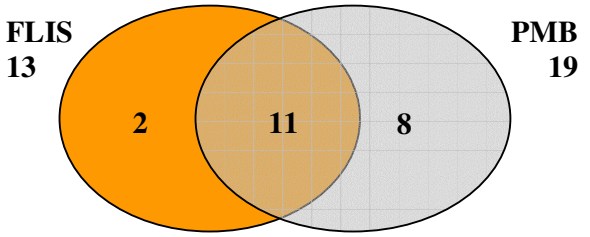
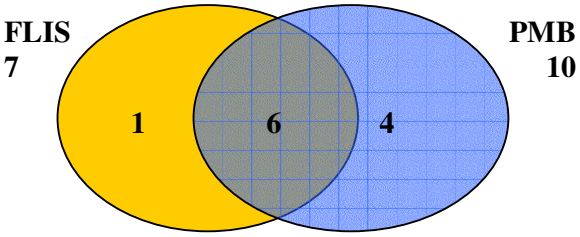


Figure 31. Histograms (left) and Q-Q normal plots (right) of the distribution of the log volume ratios from the quantification of 481 spots from the set of 2-DE gels (control and treated; n = 10) of *Rh. centenum* wild-type extracts, of which the values were normalized using two different methods: (**A**) 'Pair matched based' normalization and (**B**) normalization with the average of 33 representative FLIS spots. *p*-Values of the Kolmogorov-Smirnov normality test are displayed in each of the Q-Q normal plots (right).



A



B

Figure 32. Venn diagram showing the numbers of differentially expressed protein spots using FLIS and PMB normalization during 2-DE analysis of protein extracts from ΔPpr-deletion (A) and wild-type (B) *Rh. centenum* strains, grown under blue and/or red light conditions.

4.3. MASS SPECTROMETRY AND PROTEIN IDENTIFICATION

The 32 differentially regulated protein spots were manually excised, enzymatically cleaved and identified through a combination of mass spectrometric techniques as described in 3.6 and 3.7.

Using the *Rh. centenum* protein database we identified 3 protein spots via their 'Peptide Mass Fingerprint' in Mascot. Another 6 spots were identified based on MS/MS ions searches with Mascot as shown in **Table 3**.

In addition of these two methods, we also implemented manual *de novo* sequencing to identify our proteins. For 8 spots we observed a good 'Peptide Mass Fingerprint' (PMF) of the sulfonated peptides suitable for *de novo* MALDI MS/MS analysis. For the other spots, we observed a very weak PMF, with signal intensities that

were to low for MALDI MS/MS fragmentation. All fragmentation experiments were performed with the collision energy set at 1 keV and no gas in the collision chamber (low energy CID). Typically, the most intense peaks in the PMF were selected for MS/MS analysis. An example of a MALDI MS/MS fragment spectrum is given in **Figure 33**, in which the initial loss of the sulfonic acid derivative was observed as Δm = 215. By simple calculation of the differences between the adjacent y-ion fragments, the amino acid sequence could readily be interpreted. However, low energy CID is not able to distinguish between the isobaric amino acids Leu and Ile, indicated as (I/L).

In conclusion, 14 out of 32 spots assigned as differentially expressed were successfully identified using the combination of 3 MS-based protein identification tools and knowledge of the *Rh. centenum* protein sequence database.

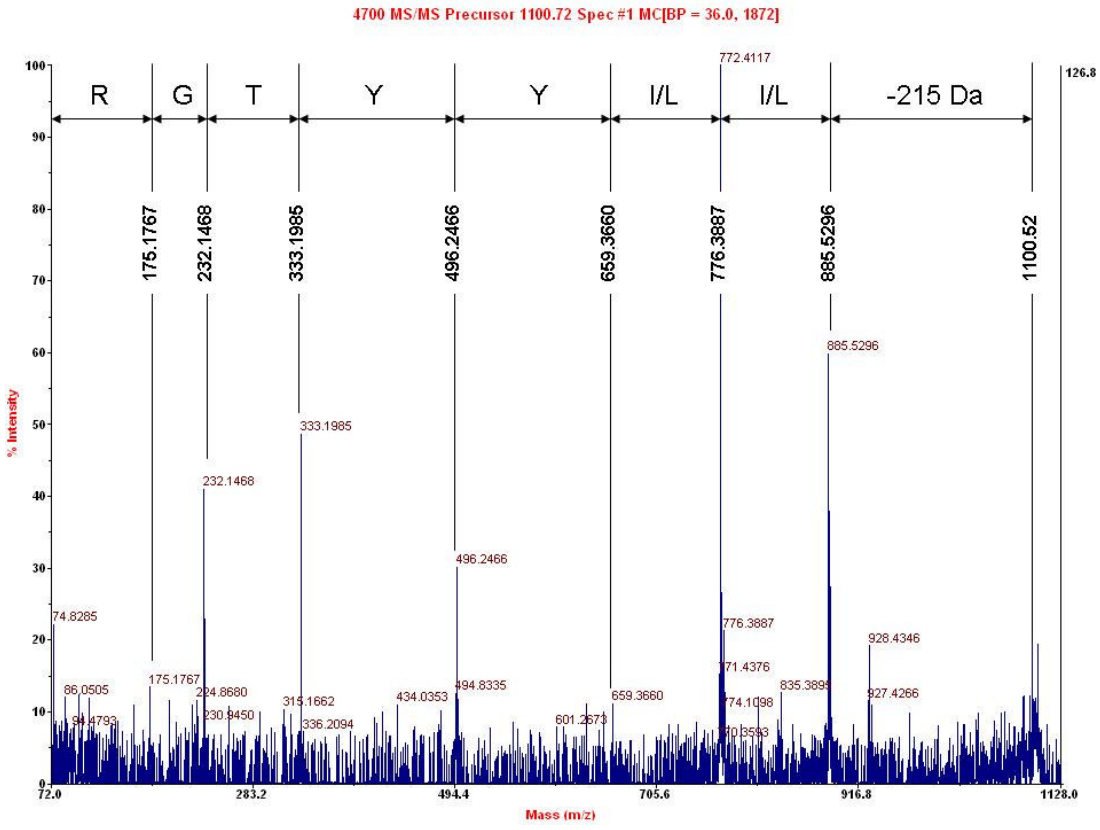


Figure 33. MALDI MS/MS fragment spectrum (positive ion mode) of the sulfonated tryptic peptide ILYYTGR of spot 13 (**Figure 26**). All labeled fragment ions are y-ions. *De novo* derived sequence information is indicated in the one-letter code. Loss of the sulfonation label is indicated as -215 Da.

Table 3. Identified proteins from *Rhodospirillum centenum* using 3 MS-based identification tools.

Rhodospirillum centenum ΔPpr-mutant					
No.[a]	Proteom ID[b]		Sequence[c]	Gene ID[d]	MS-tool[e]	Identification[f]
1	75786		-	-	-	-
2	75522		-	ORF02208	MS/MS ion search	ATP synthase F1, beta subunit
3	75471		-	-	-	
4	75788		GPQIFLSR	ORF02863	De novo	Transcription elongation factor NusA
			AFLEEQNAR			
5	77940		-	-	-	-
6	75720		-	-	-	-
7	75630		-	-	-	-
8	75316		-	-	-	-
9	75299		-	-	-	-
10	75056		AYFR	ORF02876	De novo	Outer membrane protein
			AGLLATPDTLVYGR		MS/MS ion search	
			IGGGVEHALTENVLAR			
11	75311		-	-	-	-
12	75584		YEVNDNLR	ORF03674	De novo	Vitamin B12 transporter btuB precursor, putative
13	75044		ILYYTGR	ORF03485	De novo	Translation elongation factor G FusA
					MS/MS ion search	
14	75050		-	-	-	-
15	75624		-	ORF01759	MS/MS ion search	Phosphoserine aminotransferase serC
16	75660		-	ORF01638	MS/MS ion search	HAMP domain protein
17	75259		-	ORF00292	PMF	TonB-dependent receptor, putative
18	78422		YILTVAPQQASLVR	ORF00977	De novo	Phosphoribosylformylglycinamidine synthase II purL
19	75349		HVEYQTVRR	-	De novo	-
			Exxx(I/L)(I/L)ASR			
20	75406		-	ORF00844	MS/MS ion search	Conserved hypothetical protein
21	78456		-	ORF01638	PMF	

				Rhodospirillum centenum wild-type			
Nr[a]	Proteom ID[b]			Sequence[c]	Gene ID[d]	MS-tool[e]	Identification[f]
22	149			-	-	-	-
23	120			QVVLLENTR	ORF03330	De novo	Phosphoglycerate kinase pgk
24	170			-	-	-	-
25	241			-	ORF02290	PMF	Conserved hypothetical protein
					ORF04004	PMF	Transketolase cbbT
					ORF00454	PMF	RNA pseudouridine synthase, putative
26	270			QVFDFFR	ORF01230	De novo	Trigger factor TF, putative
27	288			-	-	-	-
28	227			-	-	-	-
29	72			-	-	-	-
30	137			-	-	-	-
31	163			-	-	-	-
32	144			-	-	-	-
[a]	Spot number according to the position on the 2-D PAGE						
[b]	ID number accorded by Proteomweaver during analysis						
[c]	Manually derived amino acid sequence from MS/MS fragmentation spectra						
[d]	*Rh. centenum* gene ID						
[e]	MS-method(s) giving significant identification						
[f]	Protein identification according to the annotated *Rh. centenum* database						
				Downregulated			
				Upregulated			
				FLIS Normalization			
				FLIS + PMB Normalization			
				PMB Normalization			

78

CHAPTER 5

DISCUSSION

Rh. centenum is able to form cyst cell under limited nutrients which is a combination of at least 4 individual cells surrounded by thick coat [**116**]. This physiological change could be the reason of increasing absorption rate at the early stage of the stationary phase. Different light conditions did not change the growth rate of wt *Rh. centenum* and the Ppr deletion mutant (ΔPpr).

Through comparative proteomics we were interested in those protein spots that show a different expression level in *Rh. centenum* under the blue-light growth conditions. In order to detect such spots, a number of statistical approaches were applied. First, the means of the spot volumes from our proteomic state of interest (*i.e.* treated with blue light) was compared with the means of the spot volumes from their matched spots of a control group (*i.e.* red light cultivated cells). Therefore, we transformed our problem in the following statistical question: are the two population means significantly different and at which confidence interval? Conservatively, we used a two-sample Student's t-test statistic, in which the null hypothesis

$$\mathbf{H_0 :} \quad \mu_{\mathbf{Control}} = \mu_{\mathbf{Treated}}$$

is tested in favor of the alternative hypothesis

$$\mathbf{H_a :} \quad \mu_{\mathbf{Control}} \neq \mu_{\mathbf{Treated}}$$

and probability values (*p* values) of less than 0.05 are applied to define statistical significance. However, there exist some drawbacks using solely this method. For

example, Student's t-statistics only apply on normally distributed data (*i.e*. following a normal or Gaussian distribution) and therefore should the normal distribution of the spot volumes from each group be validated prior to analysis. On the other hand, the "Central Limit Theorem" of probability theory states that when n is large, the mean (μ) of a random sample of size n and with a finite variance, will be approximately normally distributed. This underlines the importance of including multiple biological replicates in the experiment. Unfortunately, in most 2-DE experiments, groups do not comprehend more than 10-20 data points (e.g. individuals, gels) and therefore, a normal distribution of the measured values cannot be concluded. Thus, most tests for normal distribution (e.g. Kolmogorov-Smirnov, Shapiro-Wilks) find that the normal distribution of the experimental values cannot be rejected, which does not imply that a normal distribution is the true distribution of the data. Despite these problems, most proteomic studies assume a normal distribution of values, and for this reason we also considered our data distribution to be normal. An alternative could be using non-parametric tests such as the Mann-Whitney-Wilcoxon test.

Next, we were not yet convinced that the use of a parametric (such as Student's t) or non-parametric test with a 95% confidence interval ($p < 0.05$) was sufficient enough to detect significant up - or down regulation of protein spots, although gel-to-gel variation of each spot (σ) is taken into account during calculation of the t-statistic,

$$t = (\mu_{Control} - \mu_{Treated}) / \sqrt{[(\sigma^2_{Control}/n_{Control}) + (\sigma^2_{Treated}/n_{Treated})]}.$$

Most importantly, using probability values of 0.05 or less also means that an analysis of 1,000 resolved proteins will yield at least 50 potential false positives. Thus, we needed an additional method to reduce this rate of false positives.

We attempted filtering out the false positives by setting up a confidence interval for significant regulation based on the evaluation of experimental variation. Therefore, we reported the difference between spot volumes as a volume ratio. In this approach we assume that the majority of the spots in the gel are not regulated, and thus the distribution of the log volume ratios should follow a normal distribution with a mean close to 0. If so, a threshold of two times the standard deviation is suggested that will include 95% of the spot changes that arise from experimental variation. Consequently,

protein spots with a fold change greater than this threshold are considered significant changes with a 95% confidence that the change is a difference between the two samples not due to random chance.

Through this method, we observed a serious reduction of significantly regulated spots compared to using Student's t-test alone. However, in some cases the normal distribution of log volume ratios were not apparent, which would make the use of "SD based" threshold statistically unjustified. The normal distributions were easily restored by omitting these 'significant' spots. It is not unlogical for outlier data to distort the normality of distributions, as observed above. Unfortunately, when these distributions concern proteomic data, outlier data are our main points of interests as they represent significantly altered protein expression levels in response to changed biological conditions. But, when seeking for expression changes with statistical confidence through studying log volume ratios, a normal distribution is required. Maybe it is therefore necessary to exclude a certain fraction of outlier data prior to setting a confidence interval? In some 2-DE software packages this step is executed by default (e.g. DeCyder^{TM}). But it is unclear how to set a cutoff for the fraction of data to be omitted?

Apart from the normalization algorithm used (which will be discussed further in this chapter), we believe that through combination of both methods previously discussed, we obtained a reliable detection of 32 significantly regulated protein spots with a statistical confidence.

Several normalization algorithms have been developed to account for experimental variabilities inherent to 2-DE. Normalization adjusts the spot intensities or protein expression levels in such a way that they become comparable between different gels. Although a good normalization method is indispensable for a reliable 2-DE analysis, it should be noted that no normalization protocol exist that is able to correct poor 2-DE data.

We evaluated two normalization methods, described as FLIS and PMB-normalization. Our results show that each method creates a divergent set of significant proteins and of which a fraction are commonly shared between the two methods.

Moreover, our all cross-sectional 17 spots (11 for ΔPpr mutant and 6 for wild-type) most likely represent true regulated spots. Specific spots for each normalization method can be interpreted as false positives inherent to the normalization algorithm used. Although this is not necessarily true. It is also possible that some regulated spots are missed (false negatives) due to the normalization algorithm. Especially lightly regulated protein spots will be sensitive to such spot volume changing algorithms and will reside on the border of significant detection.

In conclusion, comparing different normalization method aids the reliable analysis of 2-DE data but it is not feasible to be applied in high-throughput proteomics due to time-costs and intensive effort.

We used three different MS-based tools to identify our significant proteins. Two of these methods are using the PMF and PFF results of MALDI-TOF/TOF and LC-MS/MS respectively. Experimentally measured m/z values of peptide fragments were searched against a theoretically cleaved *Rh. centenum* protein sequence database. Alternatively, *de novo* sequencing was used as a third method for identification.

LC-MS/MS is a useful method for identification of low abundance proteins. In our experiments, some spots have a low intensity on the 2-DE PAGE image, so LC-MS/MS measurement gives us the opportunity to detect these low abundance spots. LC will separate and concentrate the peptide mixture before mass spectrometry. It leads to an increasing quality of peaks on the spectrogram.

MALDI TOF/TOF is the most rapid and easiest method to obtain mass spectra of protein peptides. Despite being a high-throughput method, it is only able to detect well abundant proteins.

De novo sequencing uses manually derived (partial) amino acid sequences, based on PFF results, to search against the protein sequences of an organism.

The first two methods rely on Mascot analysis software, which has been developed to aid in interpreting MS-data for protein identification. In the last method experimentally obtained partial sequence information is directly used for searching the

primary protein sequences of our organism without using any sophisticated analysis software.

The success of all three methods is improved if full genome sequence information is available, which in our case of *Rh. centenum* was kindly provided by Prof. Carl Bauer. The combinations of all three methods is proven to be complementary and allowed us to interpret 14 out of 32 regulated proteins, giving us a success rate of our identification of 44%. Using more than one interpretation methods is boosting the identification of proteins.

CHAPTER 6

CONCLUSION

The combination of two dimensional gel electrophoresis and mass spectrometry in protein biochemistry, is an established powerful strategy for the analysis of quantitative and qualitative differences between two or more related biological states. For this thesis work we applied those methods to investigate certain light responses of a phototrophic bacterium, of which we hope they would correlate with the function of a unique photoreceptor the organism expresses.

During analysis we evaluated two normalization methods, described as FLIS and PMB-normalization, to encounter for experimental variation inherent to 2-DE. Subsequently, the expression data were investigated through several statistical approaches to find reliable regulated spots, each step ruling out more false positive spots and bringing us closer to a true candidate list of significantly regulated spots.

Finally, relevant protein spots were subjected to mass spectrometric analysis for the purpose of identification, after an 'in-gel' digestion protocol with trypsin. In proteomics, high-throughput identification strategies rely on comparisons with predicted values derived from a protein database. Therefore, for purpose of identification, we applied MALDI MS/MS and LC-MS/MS coupled to the Mascot search engine. *De novo* sequence information relies on measuring the mass differences between adjacent fragment ion peaks of one of the major ion series. This sequence information was matched with entries of the primary protein sequence database of *Rh. centenum*. We perfomed a sulfonic acid derivatization protocol to enhance our MALDI MS/MS fragmentation spectra.

Our 2-DE analysis revealed a set of 32 spots involved in a cellular response of the phototrophic organism towards blue light conditions. So far, 14 spots were successfully identified and their significance in the photoresponse will have to be further investigated. We also learned that no single method exists that unambiguously detects true regulation of protein expression. Multiple strategies to encounter for experimental variation as well as documented statistical approaches are useful in order reach true significance.

REFERENCES

[1] Ragatz, L.; Jiang, Z.Y.; Bauer, C.E.; Gest, H. (1995) Macroscopic phototactic behavior of the purple photosynthetic bacterium *Rhodospirillum centenum*. *Arch Microbiol.* 163: 1-6.

[2] Berleman, J.E.; Bauer, C.E. (2004) Characterization of cyst cell formation in the purple photosynthetic bacterium *Rhodospirillum centenum. Microbiology*, 150: 383-90.

[3] Jiang, Z. (1999) Bacterial photoreceptor with similarity to photoactive yellow protein and plant phytochromes. *Science*, 285: 406-9.

[4] Berleman, J.E.; Hasselbring, B.M.; Bauer, C.E. (2004) Hypercyst mutants in *Rhodospirillum centenum* identify regulatory loci involved in cyst cell differentiation. *J Bacteriol,* 186: 5834-41.

[5] Prescott, L.M.; Harley, J.P.; Klein, D.A. (2002) Microbiology 5th edition, 423-424.

[6] Prescott, L.M.; Harley, J.P.; Klein, D.A. (2002) Microbiology 5th edition, 487.

[7] Margulis, L. Origin of Eukaryotic cells. (1970) Yale University Press, New Haven, CT.

[8] Berleman, J.E.; Bauer, C.E. (2004) Characterization of cyst cell formation in the purple photosynthetic bacterium *Rhodospirillum centenum. Microbiology*, 150: 383-390.

[9] Favinger, J.; Stadtwald, R.; Gest, H. (1989) *Rhodospirillum centenum*, sp. nov., a thermotolerant cyst-forming anoxygenic photosynthetic bacterium. *Antonie Leeuwenhoek*. 55: 291–296.

[10] Berleman, J.E.; Hasselbring, B.M.; Bauer, C.E. (2004) Hypercyst mutants in *Rhodospirillum centenum* identify regulatory loci involved in cyst cell differentiation*. J. Bacteriol.* 186: 5834-41.

[11] Stadtwald-Demchick, R.; Turner, F. R.; Gest, H. (1990) Physiological properties of the thermotolerant photosynthetic bacterium, *Rhodospirillum centenum*. *FEMS Microbiol. Lett.* 67: 139–144.

[12] Ragatz, L.; Jiang, Z.Y.; Bauer, C. E.; Gest, H. (1994) Phototactic purple bacteria. *Nature.* 370: 104.

[13] Ragatz, L.; Jiang, Z.Y.; Bauer, C.E.; Gest, H. (1995) Macroscopic phototactic behavior of the purple photosynthetic bacterium *Rhodospirillum centenum. Arch. Microbiol.* 163: 1–6.

[14] Jiang, Z.Y.; Bauer, C.E. (2001) Component of the *Rhodospirillum centenum* photosensory apparatus with structural and functional similarity to methyl-accepting chemotaxis protein chemoreceptors. *J Bacteriol.* 183: 171-7.

[15] Rushing, B.G.; Jiang, Z.Y.; Gest, H.; Bauer C.E. (2000) Phototactic behavior In J. Lederberg, Encyclopedia of Microbiology, 2nd ed., vol. 3. Academic Press. 618–624,

[16] Marcella, J.S.; Judith, P.A; Elizabeth, E.S.; Thomas, P.P. (1997) Photoresponses of the Purple Nonsulfur Bacteria *Rhodospirillum centenum* and *Rhodobacter sphaeroides*. *J. Bacteriol*. 179: 6764–6768.

[17] Sprenger, W.W.; Hoff, W.D.; Armitage, J.P.; Hellingwerf, K.J. (1993) The eubacterium Ectothiorhodospira halophila is negatively phototactic, with a wavelength dependence that fits the absorption spectrum of the photoactive yellow protein. *J. Bacteriol*. 175: 3096-104.

[18] Winslow, R.B.; Margaret, A.O. (2001) Photoreceptors in Plant Photomorphogenesis to Date. Five Phytochromes, Two Cryptochromes, One Phototropin, and One Superchrome. *Plant Physiol.* 125: 85-88.

[19] Jiang, Z.; Swem, L.R.; Rushing, B.G.; Devanathan, S.; Tollin, G.; Bauer, C.E. (1999) Bacterial photoreceptor with similarity to photoactive yellow protein and plant phytochromes. *Science*. 285: 406-409.

[20] Park, H.; Saha, S.; Inouye, M. (1998) Two-domain reconstitution of a functional protein histidine kinase. *Biochemistry*. 95: 6728-6732.

[21] Wilkins, M.R.; Pasquali, C.; Appel, R.D.; Ou. K; Golaz, O.; Sanchez, J.C.; Yan, J.X.; Gooley, A.A.; Hughes, G.; Humphery-Smith, I.; Williams. K.L.; Hochstrasser, D.F. (1996) From proteins to proteome: Large-scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology*. 14: 61-65.

[22] Anderson, N.G.; Anderson, N.L; (1998) Proteome and proteomics: New technologies, new concepts, and new words. *Electrophoresis*. 19: 1853-1861.

[23] Wilkins, M.R.; Sanchez, J.C.; Gooley, A.A.; Appel, R.D.; Humphery-Smith, I.; Hochstrasser, D.F.; Williams, K.L. (1996) Progress with proteome projects: Why all proteins expressed by a genome should be identified and how to do it ? *Biotechnol. Genet. Eng. Rev.* 13: 19-50.

[24] Lubec, G.; Nonaka, M.; Krapfenbauer, K.; Gratzer, M.; Cairns, N.; Fountoulakis, M. (1999) Expression of the dihydropyrimidinase related protein 2 (DRP-2) in Down syndrome and Alzheimer's disease brain is downregulated at the mRNA and dysregulated at the protein level. *J. Neural Transm. Suppl*. 57: 161–177.

[25] Cutler, P.; Bell, D.J.; Birrell, H.C.; Connelly, J.C.; Connor, S.C.; Holmes, E.; Mitchell, B.C.; Monte, S.Y.; Neville, B.A.; Pickford, R.; Polley, S.; Schneider, K.; Skehel, J.M. (1999) An integrated proteomic approach to studying glomerular nephrotoxicity. *Electrophoresis*. 20: 3647–3658.

[26] Fegatella, F.; Ostrowski, M.; Cavicchioli, R. (1999) An assessment of protein profiles from the marine oligotrophic ultramicrobacterium *Sphingomonas* sp. Strain RB2256. *Electrophoresis*. 20: 2094–2098.

[27] Görg, A.; Obermaier, C.; Boguth, G.; Harder, A.; Scheibe, B.; Wildgruber, R.; Weiss, W. (2000) The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis*. 21: 1037–1053.

[28] Choe, L.H.; Lee, K.H. (2000) A comparison of three commercially available isoelectric focusing units for proteome analysis: The multiphor, the IPGphor and the protean IEF cell. *Electrophoresis*. 21: 993–1000

[29] Celis, J.E.; Gromov, P. (1999) 2D protein electrophoresis: can it be perfected? *Curr. Opin. Biotechno*.10: 16-21.

[30] Rabilloud, T. (2002) Two-dimensional gel electrophoresis in proteomics: old fashioned, but it still climbs up the mountains. *Proteomics* 2: 3-10.

[31] Fey, S.J.; Larsen, P.M. (2001) 2D or not 2D. Two-dimensional gel electrophoresis. *Curr. Opin. Chem. Biol*. 5: 26-33.

[32] O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 250: 4007-4021.

[33] Bjellqvist, B.; Ek, K.; Righetti, P.G.; Gianazza, E.; Görg, A.; Westermeier, R.; Postel, W. (1982) Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *J. Biochem. Biophys. Methods.* 6: 317-339.

[34] Link, A.J.; (1999) 2-DE Proteome Analysis Protocols, *Methods Mol. Biol*. 112.

[35] Rabilloud, T. (1999) Solubilization of proteins in 2-DE electrophoresis *Methods Mol. Biol.* 112: 9–19.

[36] Macri, J.; McGee, B.; Thomas, J.N.; Du, P.; Stevenson, T.I.; Kilby, G.W.; Rapundalo, S.T. (2000) Cardiac sarcoplasmic reticulum and sarcolemmal proteins separated by two dimensional electrophoresis: surfactant effects on membrane solubilization. *Electrophoresis*. 21: 1685–1693.

[37] Lopez, J.L. (2006) Two-dimensional electrophoresis in proteome expression analysis. *J. Chromatogr B. Analyt. Technol*. Biomed. Life. Sci.

[38] Santaren, J.F. (1990) Towards establishing a protein database of *Drosophila*. *Electrophoresis*, 11: 254-267.

[39] 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.

[40] Protein Assay, Technical Handbook, Pierce, p:19.

[41] Mastro, R.; Hall, M. (1999) Protein delipidation and precipitation by tri-n-butylphosphate, acetone, and methanol treatment for isoelectric focusing and two-dimensional gel electrophoresis. *Analytical Biochemistry*. 273: 313-315.

[42] Damerval, C.; De Vienne, D.; Zivy, M.; Thiellement, H. (1986) Technical improvements in two-dimensional electrophoresis increase the level of genetic variation detected in wheat-seedling proteins. *Electrophoresis*. 7: 52-54.

[43] Granzier, H.L.M.; Wang, K. (1993) Gel electrophoresis of giant proteins: solubilization and silver-staining of titin and nebulin from single muscle fiber segments. *Electrophoresis*. 14: 56-64.

[44] Colas des Francs, C.; Thiellement, H.; De Vienne, D. (1985) Analysis of Leaf Proteins by Two-Dimensional Gel Electrophoresis: Protease Action as Exemplified by Ribulose Bisphosphate Carboxylase/ Oxygenase Degradation and Procedure to Avoid Proteolysis during Extraction. *Plant. Physiol*. 78: 178-182.

[45] Wilkins, M.R.; Sanchez, J.C.; Gooley, A.A.; Appel, R.D.; Humphery-Smith, I.; Hochstrasser, D.F.; Williams, K.L. (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol. Genet. Eng. Rev*. 13: 19-50.

[46] Görg, A.; Weiss, W.; Dunn, M. (2004) Current two-dimensional electrophoresis technology for proteomics. *Proteomics*. 4: 3665-3685.

[47] Görg, A.; Weiss, W. (1998) High-resolution two-dimensional electrophoresis of proteins using immobilized pH gradients. Cell Biology, A Laboratory Handbook, Academic Press. 386-397.

[48] Biochemistry and Biotechnology Laboratory Manual (2006), Gent University Press, p:11

[49] 2-D Electrophoresis for Proteomics, Method and Product Manual, Bio-Rad, p:8.

[50] Corbett, J.M.; Wheeler, C.H.; Baker, C.S.; Yacoub, M.H.; Dunn, M.J. (1994) The human myocardial two-dimensional gel protein database: update 1994. *Electrophoresis*. 15: 1459-1465.

[51] Blomberg, A.; Blomberg, L.; Norbeck, J.; Fey, S.J.; Larsen, P.M.; Roepstorff, P.; Degand, H.; Boutry, M.; Posh, A. (1995) Interlaboratory reproducibility of yeast protein patterns analyzed by immobilized pH gradient two-dimensional gel electrophoresis. *Electrophoresis*, 16: 1935-1945.

[52] 2-D Electrophoresis for Proteomics, Method and Product Manual, Bio-Rad, p:13.

[53] Klose, J.; Kobalz, U. (1995) Two-dimensional electrophoresis of proteins: An updated protocol and implications for a functional analysis of the genome. *Electrophoresis*. 16: 1034-1059.

[54] Carroll, K.; Ray, K.; Helm, B.; Carey, E. (2000) Two-dimensional electrophoresis reveals differential protein expression in high- and low-secreting variants of the rat basophilic leukaemia cell line. *Electrophoresis*. 21: 2476–2486.

[55] Patton, W.F. (2002) Detection technologies in proteome analysis. *J. Chromatogr. B*. 771: 3-31.

[56] Zhou, S.; Bailey, M.J.; Dunn, M.J.; Preedy, V.R. (2005) Measurement of specific radioactivity in proteins separated by two-dimensional gel electrophoresis. *Proteomics*. 5: 2739–2747.

[57] Molloy, M.P. (2000) Two-dimensional electrophoresis of membrane proteins using immobilized pH gradients. *Anal. Biochem*. 280: 1–10.

[58] Thomson, J.J. (1913) Rays of positive electricity and their applications to chemical analysis. Longmans Green, London.

[59] Aston, F.W. (1933) Mass spectra and isotopes. Arnold. London.

[60] Tanaka, K.; Waki, H.; Ido, Y. (1988) Protein and polymer analysis up to m/z 100.000 by laser ionization time of flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 2: 151-153.

[61] Tanaka, K. (2003) The origin of macromolecule ionization by laser irradiation (Nobel lecture) *Angew. Chem. Int.* Ed. 42: 3860-3870.

[62] Cotter, R.J. (1997) Time of flight mass spectrometry: instrumentation and applications in biological research. *Am. Chemical Society*. 131-134.

[63] Karas, M.; Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10.000 Daltons. *Anal. Chem.*, 60: 2299-2301.

[64] Hillenkamp, F.; Karas, M.; Beavis, R.C. (1991) Matrix assisted laser desorption ionization mass spectrometry of biopolymers. *Anal. Chem.* 63: 1193-1203.

[65] Meng, C.K.; Mann, M.; Fenn, J.B. (1988) Of protons or proteins", Z. *Phys. D*. 10: 361-368.

[66] Fenn, J.B.; Mann, M.; Meng, C.K. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science*. 246: 64-71.

[67] Taylor, G. (1964) Desintegration of water drops in an electrical field. *Proc. R. Soc. Lond.* 280: 383-397.

[68] Kebarle, P.; Ho, Y. (1997) On the mechanism of electrospray mass spectrometry. In Electrospray ionization mass spectrometry: fundamentals, instrumentation and application. *Wiley*. 3-63.

[69] Rayleight, L. (1882) On the equilibrium of liquid conducting masses charged with electricity. *Philos. Mag*. 14: 184-186.

[70] Iribarne, J.V.; Thomson, B.A. (1976) On the evaporation of small ions from charged droplets. *J. Chem. Phys.* 64: 2287-2294.

[71] Fenn, B.J. (1993) Ion formation from charged droplets: roles of geometry, energy, and time. *J. Am. Soc. Mass Spectrom*. 4: 424-535.

[72] Kebarle, P.; Tang, L. (1993) From ions in solution to ions in the gas phase. *Anal. Chem*. 65: 972-986.

[73] Cole, R.B. (1997) Electrospray ionization mass spectrometry: fundamentals, instrumentations and applications. *Wiley*.

[74] Kabarle, J.A. (2000) A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J. Mass Spectrom*. 35: 804-817.

[75] Cole, R.B. (2000) Some tenets pertening to electrospray ionization mass spectrometry. *J. Mass Spectrom*. 35: 763-772.

[76] Siuzdak, G. (2003) The expending role of mass spectrometry in biotechnology. MMC Press, p: 17.

[77] Paul, V.W.; Steinwedel, H. (1953) "Ein neues massenspektrometer ohne magnetfeld. *Z. Naturforsch*. 8a: 448-450.

[78] Siuzdak, G. (2003) The expending role of mass spectrometry in biotechnology. MMC Press, p: 48.

[79] Mamyrin, B.A.; Karataev, V.I.; Shmikk, D.V. (1973) The mass reflectron, a new non-magnetic time of flight spectrometer with high resolution. *Sov. Phys. JEPT*. 37: 45-48.

[80] Beavis, R.C.; Chait, B.T. (1991) Velocity distribution of intact high mass polypeptide molecule ions produced by matrix assisted laser desorption. *Chem. Phys. Lett.* 181-497.

[81] Brown, R.S.; Lennon, J.J. (1995) Mass resolution improvement by incorporation of pulsed ion extraction in a matrix assisted laser desorption ionization linear time of flight mass spectrometer. *Anal. Chem*. 67: 1998-2003.

[82] Siuzdak, G. (2003) The expending role of mass spectrometry in biotechnology. MMC Press, p: 54.

[83] March, R.E. (1996) An introduction to quadrupole ion trap mass spectrometry. *J. Mass Spectrom.* 32: 351-369.

[84] Jonscher, K.R.; Yates, I.; John, R. (1997) The quadrupole ion trap mass spectrometer: a small solution to a big challenge. *Anal. Biochem*. 244: 1-15.

[85] Hunt, D.F.; Yates, I.; John, R.; Shabanowitz, J. (1986) Protein sequencing by tandem mass spectrometry. *Proc. Natl. Acad. Sci*. 83: 6233-6237.

[86] Jennings, K.R. (1968) Collision-induced decomposition of aromatic molecular ions. *Int. J. Mass Spectrom. Ion Physics*. 1: 227-235.

[87] Mc Laffery, F.W.; Bente, P.F.; Kornfield, R. (1973) Collision activation spectra of organic ions. *J. Am. Chem. Soc*. 95: 2120-2129.

[88] Siuzdak, G. (2003) The expending role of mass spectrometry in biotechnology. MMC Press, p: 129.

[89] Hanzel, W.J.; Billeci, T.M.; Stults, J.T. (1993) Identifying proteins from 2 dimensional gels by molecular mass searching of peptide-fragments in protein sequence databases. *Proc. Natl. Acad. Sci*. 90: 5011-5015.

[90] Pappin, D.J.C.; Hojrup, P.; Bleasby, A.J. (1993) Rapid identification of proteins by peptide mass fingerprinting. *Curr. Biol*. 6: 327-332.

[91] Yates, J.R.; Specier, S.; Griffin, P.R. (1993) Peptide mass map: a highly informative approach to protein identification. *Anal. Biochem*. 214: 397-408.

[92] James, P.; Quadroni, M.; Carafoli, E.; Gonnet, G. (1993) Protein identification by mass profile fingerprinting. *Biochem. Biophys. Res. Commun*. 195: 58-64.

[93] Mann, M.; Hojrup, P.; Roepstorff, P. (1993) Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol. Mass Spectrom*. 22: 338-345.

[94] Vanrobaeys, F. (2005) Implementation of nano-liquid chromatography hypernated to tandem mass spectrometry for protein identification in gel and non-gel based proteomics. *Gent University Press*. p: 51.

[95] Chamrad, D.C.; Korting, G.; Stuhler, K. (2004) Evaluation of algorithms for protein identification from sequence databases using mass spectrometry data. *Proteomics*. 4: 619-628.

[96] Vanrobaeys, F. (2005) Implementation of nano-liquid chromatography hypernated to tandem mass spectrometry for protein identification in gel and non-gel based proteomics. *Gent University Press*. p: 52.

[97] Perkins, D.N.; Pappin, D.J.; Creasy, D.M. (1999) Probability based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*. 20: 3551-3567.

[98] Mann, M.; Wilm, M.; (1994) Error tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem*. 66: 4390-4399.

[99] Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; (1990) Basic local alignment search tool. *J. Mol. Biol*. 215(3): 403-410.

http://www.ncbi.nlm.nih.gov/BLAST

[100] Lipman, D.J.; Pearson, W.R.; (1985) Rapid and sensitive protein similarity searches. *Science*. 22; 227(4693): 1435-1441.
http://www.ebi.ac.uk/fasta33/index.htmL

[101] Nesvizhskii, A.I.; Aebersold, R. (2004) Analysis, statistical validation and dissemination of large scale proteomics datasets generated by tandem MS. *Drug. Discov. Today.* 9: 173-181.

[102] Keough T.; Youngquist R.S.; Lacey M.P. (2003) Sulfonic acid derivatives for peptide sequencing by MALDI MS. *Anal. Chem.* 75:156A-165A.

[103] Hunt, D. F.; Yates, J.R.; Shabonowitz, J.; Winston, S.; Hauer, C. R. (1986) *Proc. Natl. Acad. Sci.* 83: 6233.

[104] Biemann, K.; Martin, S. A. (1990) *Mass Spectrom. Rev.* 6, 1.

[105] Biemann, K.; McCloskey, J.A. (1990) In Methods in Enzymology. Academic Press: San Diego, CA.p: 455.

[106] Burlet, O.; Orkiszewski, R.S.; Ballard, K.D.; Gaskell, S.J. (1992) *Rapid Commun. Mass Spectrom.* 6: 658.

[107] McCormack, A.L.; Somogyi, A.; Dongre, A.R.; Wysocki, V.H. (1993) *Anal. Chem.* 65: 2859.

[108] Brancia, F.L.; Oliver, S.G.; Gaskell, S.J. (2000) *Rapid Commun. Mass Spectrom.* 14: 2070.

[109] Hale, J.E.; Butler, J.P.; Knierman, M.D.; Becker, G.W. (2000) *Anal. Biochem.* 287: 110.

[110] Krause, E.; Wenschuh, H.; Jungblut, P.R. (1999) *Anal. Chem*. 7: 4160.

[111] Mc Donough, J.; Marban E. (2005) Optimization of IPG strip equilibration for the basic membrane protein mABC1. *Proteomics*. 5: 2892-2895.

[112] Biochemistry and Biotechnology Laboratory Manual (2006), Gent University Press, p:13

[113] Samyn, B.; Sergeant, K.; Memmi, S.; Debyser, G.; Devreese, B.; Van Beeumen, J. (2006) MALDI-TOF/TOF de novo sequence analysis of 2-DE PAGE-separated proteins from *Halorhodospira halophila,* a bacterium with unsequenced genome. *Electrophoresis*. 27(13): 2702-2711.

[114] Yildiz, F.H.; Gest, H.; Bauer, C.E. (1991) Genetic analysis of photosynthesis in *Rhodospirillum centenum*. *J. Bacteriol. 173*: 4163-4170.
http://genomes.tgen.org

[115] Perkins, D.N.; Pappin, D.J.C.; Creasy, D.M.; Cottrell, J.S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis.* 20(18): 3551-3567.
http://www.matrixscience.com

[116] Yildiz, F.H.; Gest, H.; Bauer, C.E. (1991) Attenuated effect of oxygen on photopigment synthesis in *Rhodospirillum centenum*. *J Bacteriol.* 173(17): 5502–5506.