Use Of Proteomics For The Analysis Of Hard-To-Dissect Biological Samples

by

Nikhil D Phadke

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular, Cellular, and Developmental Biology)
in The University of Michigan
2011

Doctoral Committee:

Professor Janine R. Maddock, Chair
Professor Philip C. Andrews
Associate Professor Laura J. Olsen
Assistant Professor Lyle A. Simmons

ACKNOWLEDGMENTS

I would like to thank my advisors Dr. Janine Maddock and Dr. Philip Andrews for their guidance and patience over the years and my committee members and the faculty and staff of the MCDB department for making this possible. Several individuals have directly or indirectly helped or supported me over the years, including family, friends, and colleagues. You know who you are, my deepest gratitude to you.

Thank you,

Nikhil Phadke

November 2010

Ann Arbor.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter I

Chapter II

Chapter III

Chapter IV

LIST OF TABLES

Chapter I

Chapter II

Chapter III

Chapter IV

CHAPTER I

QUANTITATIVE PROTEOMICS OF MICROBIAL MEMBRANE PROTEINS BY ITRAQ

The bulk of my research deals with the development and standardization of improved methods to carry out meaningful proteomic analyses of proteins from hard-to-dissect fractions biological fractions. I also describe some of the earliest efforts where proteomic analysis was carried out concomitantly with the genome sequencing project. These are illustrated with proteomic analyses of membrane fractions from the previously unannotated Gram negative bacterium *C. crescentus*, from the refractile fractions of spore protective structures of the Gram positive bacterium *B. subtilis* and from our preliminary investigation of the spore protective structures of its close relative, the biowarfare agent *B. anthracis*. This first chapter provides an introduction to the field of proteomics, and it's methodologies with a special focus on membrane proteins from microbial systems. It reviews the newest techniques for quantitative proteomics in these systems, with a special focus on the iTRAQ approach, currently one of the methods of choice for quantitative proteomics, and illuminates options available to investigators wishing to carry out iTRAQ based quantitative proteomic studies on microbes.

CHAPTER I

QUANTITATIVE PROTEOMICS OF MICROBIAL MEMBRANE PROTEINS BY ITRAQ

The bulk of my research deals with the development and standardization of improved methods to carry out meaningful proteomic analyses of proteins from hard-to-dissect fractions biological fractions. I also describe some of the earliest efforts where proteomic analysis was carried out concomitantly with the genome sequencing project. These are illustrated with proteomic analyses of membrane fractions from the previously unannotated Gram negative bacterium *C. crescentus*, from the refractile fractions of spore protective structures of the Gram positive bacterium *B. subtilis* and from our preliminary investigation of the spore protective structures of its close relative, the biowarfare agent *B. anthracis*. This first chapter provides an introduction to the field of proteomics, and it's methodologies with a special focus on membrane proteins from microbial systems. It reviews the newest techniques for quantitative proteomics in these systems, with a special focus on the iTRAQ approach, currently one of the methods of choice for quantitative proteomics, and illuminates options available to investigators wishing to carry out iTRAQ based quantitative proteomic studies on microbes.

Abstract

Classical proteomic techniques, based on two-dimensional electrophoresis combined with mass spectrometry, have been invaluable for identifying proteins and cataloguing the proteomes of several microbes. They have, however, for the most part, been limited to qualitative rather than quantitative observations. Furthermore, technical limitations of these approaches have resulted in the under-representation of proteins from several fractions of the proteome such as hydrophobic membrane proteins, spore proteins, alkaline proteins, highly acidic proteins, and very small and very large proteins. Several new methods have recently been developed to address the limitation of quantitative proteomic profiling. Isobaric tagging for relative and absolute quantification of proteins (iTRAQ) is one of the most popular of these methods. In addition to enabling quantitative proteomics, it has also enhanced the analysis of protein fractions that were not previously amenable to proteomic profiling. I review the use of iTRAQ for quantitative analysis of the membrane proteomes of microbial systems. I visit some of the landmark microbial membrane iTRAQ studies and discuss the advances made. I also describe a previously unpublished iTRAQ based quantitative analysis of the membrane fractions of the predatory bacterium *Bdellovibrio bacteriovorus* comparing host-dependent and host-independent lifestyles. 718 unique proteins were identified, 322 of these with high confidence. Between 117-150 of the identified proteins were predicted to be membrane proteins by different algorithms. 66 of the identified proteins, many of which are membrane proteins, are differentially expressed with high confidence under the

conditions compared. Finally I discuss the options available to investigators wishing to carry out the analysis of microbial membrane fractions using iTRAQ.

Introduction

Even though protein sequencing, in the form of the Edman degradation reaction [1], pre-dates DNA sequencing by nearly two decades [2, 3], it was the latter that advanced more rapidly. This was for a number of reasons, including the relative simplicity of DNA as compared to proteins, the lower cost of DNA sequencing and the relative ease of automating DNA sequencing techniques - particularly the Sanger method and its derivatives [4]. The increases in DNA sequencing capacity were accompanied by a drive to gain a more holistic view of biological systems. The initiation of several massive genome projects resulted in the completion of over a thousand genomes, giving rise to an explosion of DNA sequence data [4, 5]. However, it quickly became evident that sequence information by itself only provides a one-dimensional view of the situation. Mapping of temporal and spatial expression patterns and quantitation of gene products would be necessary for a more complete and meaningful understanding of biological systems and their regulation.

The development of microarrays capable of measuring RNA levels significantly enhanced our ability to quantitate and monitor changes in gene expression. It was suddenly possible to assay the expression levels of thousands of genes in parallel in a single experiment [6, 7]. Nevertheless, while gene expression levels can provide an indication of protein expression levels, mRNA expression profiling does take into account post-translational control and modification, and co-operative interactions between proteins [7, 8]. Furthermore, it is becoming more and more clear that mRNA and protein expression patterns can, at times, be divergent [6, 9]. Because, for the most part, proteins are the final effector molecules in biological systems, ultimately it is important

to assay their expression levels to get a deeper understanding of these systems. Unfortunately the ability to carry out large-scale protein analyses, comparable in scale to genome sequencing projects and microarrays analyses, has been hampered by the complexity and enormous biological variations in proteins. The field of proteomics was born out of a requirement to address this challenge.

Classical proteomics

The core of modern proteomics lies in the development of two-dimensional electrophoresis (2-DE), as a method for arraying proteins on gels, by O'Farrell in the mid 1970s [10, 11]. Two properties of protein molecules are exploited to resolve them electrophoretically – the isoelectric point (pI) of the protein i.e. the pH at which it has no net charge, and the molecular weight of the protein. Proteins are initially separated on the basis of their pI using isoelectric focusing (IEF) followed by separation on the basis of molecular weight using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The arrayed proteins are typically visualized by Coomassie Brilliant Blue (CBB) staining or silver staining. The latter can be about 20-100 times more sensitive than CBB staining, but is less linear across its dynamic range and can interfere with subsequent downstream analysis [12]. Alternative methods for visualization include autoradiography, which is highly sensitive but logistically complex [13], and more recently fluorescent staining (e.g. with the SYPRO family of dyes), which is nearly as sensitive as silver staining and significantly more linear [12]. Originally protein identification was carried out either by Edman sequencing of the resolved proteins (spots) on the gels or by immuno-blotting, but these methods are laborious and not easily amenable to high throughput analyses.

Advances in mass spectrometry (MS) coupled with the availability of sequence databases, as a result of the genome sequencing projects, saw the birth of the modern proteomics era. The development of 'soft ionization' techniques in the early 1980s allowed investigation of biomolecules like proteins and peptides which were previously not easily amenable to mass spectrometric analysis. The two most frequently used soft ionization methods are matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) which is often carried out in-line with high performance liquid chromatography (HPLC) [14]. The most commonly used approach for identifying proteins by MS is peptide mass fingerprinting (PMF). Proteins are digested, after 2-DE separation, with proteolytic enzymes that cleave at fixed residues creating peptides whose masses can be measured by MS. The pattern of measured masses forms a signature for each of the proteins and can be used to identify the protein when matched against a theoretical digest of all the proteins from the organism [14]. Trypsin, which cleaves proteins at lysine and arginine residues, is one of the most frequently used enzymes for this purpose. Separated proteins are typically digested within the gel matrix itself, and digested peptides are extracted with organic solvents prior to MS analysis. Detergents and chaotropes, which can interfere with MS analysis of the peptides, are removed during the course of two electrophoretic steps and the peptide gel-extraction procedure. Recently, with the development of more advances mass spectrometers, tandem MS/MS for *de novo* amino acid sequencing has become more common [14].

This "classical" approach of using 2-DE coupled with MS has been the gold standard for proteomic analyses of complex protein mixtures for several years. Unfortunately, there are some serious limitations to the approach. These include the

inability to efficiently resolve very small (<10 kDa) or very large (>150 kDa) proteins, a large subset of basic proteins (pI > 9), highly acidic proteins, and hydrophobic proteins with grand average of hydropathy values (GRAVY) > 0.3 – 0.4 [15]. Furthermore, the technique has a relatively low dynamic range of resolution, being able to resolve across a range of 3-4 orders of magnitude at best. Biological variations can be up to 5 - 9 orders of magnitude and can even extend up to 12 orders in plasma [16-18]. This can result in missing or masking of low-abundance but biologically significant proteins [19]. Although robotics-based tools have been introduced into the workflow to assist with some of the repetitive precision steps, they remain expensive and hard to integrate into standard laboratory processes [20].

Membrane proteomics

One of the most serious consequences of the reduced capacity to separate hydrophobic proteins by 2-DE is the inability of the technique to resolve membrane proteins efficiently. Membrane proteins constitute between 20 – 30 % of the total proteins in all genomes [21]. They play critical roles in cell structure, cell specialization, communication, signaling, cell-cell interactions, transport and energy metabolism. From a biomedical standpoint, they are also critical for our understanding of pathogenesis and drug resistance mechanisms. Nearly 70% of all pharmaceutical drug targets are membrane proteins [22]. Furthermore, a detailed knowledge of membrane and surface proteins is necessary for the development of vaccines and identification of targets for diagnostics.

In order to traverse the plasma membrane, which is hydrophobic in the middle and hydrophilic on the surfaces, membrane proteins are amphipathic i.e. they have

hydrophobic sections and hydrophilic sections. To enable integration with the plasma membrane, two protein structures have evolved - a β-barrel structure, which is common in the outer membrane proteins (OMPs) of Gram negative bacteria and organelles like mitochondria, and an α-helical structure, which is common feature of inner membrane proteins (IMPs) in Gram-negative bacteria and in most other membrane systems. β-barrel proteins have alternating hydrophilic and hydrophobic amino acid residues in regions that are in contact with hydrophobic parts of the plasma membrane, resulting in relatively low overall hydrophobicity [23]. As a result, this class of membrane proteins is more amenable to analysis by 2-DE, and several successful studies have been carried out on microbial OMPs [24-26]. α-helical membrane proteins, however, are harder to solubilize and also have a propensity to precipitate during IEF owing mostly to stretches of hydrophobic amino acid-rich regions in their plasma membrane spanning transmembrane domains (TMDs). Typically, only IMPs with low GRAVY scores, and with two or fewer TMDs, are efficiently resolved by IEF [7, 27, 28]. Furthermore, the lower number of tryptic digestion sites in membrane spanning regions, and the decreased access to proteases due to low aqueous solubility, result in insufficient number of peptides for MS analysis [28]. Membrane proteins with large hydrophilic exposed domains can be analyzed by some proteomic techniques [29], but the size of these exposed domains varies considerably among membrane proteins, so this property cannot be relied upon for complete proteome coverage [30, 31]. Finally, the low abundance of these membrane proteins, relative to many cytoplasmic proteins, can conspire with the limited dynamic range of the method resulting in the masking of these proteins on gels.

Improved analysis of membrane proteins by 2-DE

In spite of these limitations, there are still some clear advantages to using 2-DE based proteomics. It remains the only method that allows direct visualization of proteins and the ability to see isoforms [32]. Therefore several efforts have been made to address the limitations of 2-DE based proteomics, especially to address the membrane protein resolving deficiencies. These are reviewed in detail elsewhere [33]. Not surprisingly, since some of the biggest challenges of proteomic analyses occur during the preparation and handling of membrane fractions, many of these techniques are also relevant for proteomic analyses by non 2-DE methods including iTRAQ [22, 29], so they are discussed in some detail here.

The issue of dynamic range is largely addressed by two approaches - enrichment of membrane fractions and removal of contaminating high abundance soluble proteins. The enrichment approach is sometimes also referred to as fractionation or sub-proteomics [19, 34]. For most single celled microbes, crude membrane preparations can be obtained simply by high-speed centrifugation of cell lysates, which separates soluble proteins from the insoluble membrane fractions. For microbes with more complex membrane systems e.g. cyanobacteria and protozoa, enrichment of specific membranes can be achieved by density gradient centrifugation [23]. Commonly used gradients include sucrose, sorbitol, and Percoll™ (GE Healthcare). Affinity based techniques are also a popular tool for enrichment of specific, or subsets of, membrane proteins. These include immunoisolation and affinity purification [35]. The enrichment of hydrophobic membrane glycoproteins by lectin affinity using wheat germ agglutinin (WGA), a technique more commonly used in eukaryotes, is an example of the latter [7, 35]. Biotinylation and glycosylation affinity

purifications are two more affinity purification strategies that have seen use in proteomic studies [35]. The use of colloidal silica is a novel approach to affinity purification of membranes [36]. The technique exploits the affinity of positively charged silica for the anionic phospholipid head groups of the membrane and the carbohydrate groups of glycoproteins. Aluminum chlorohydroxide coated silica particles are applied to intact cells, and the particles are cross-linked with anionic poly acrylic acid, which forms a supporting matrix. On disruption of the target cells, the membranes remain bound the matrix in the form of open sheets, and can be easily separated by centrifugation [23, 35].

Phase-partitioning using mixtures of hydrophobic and hydrophilic solutions offers yet another powerful approach for enrichment of membrane proteins. Mixtures of dextran and polyethylene glycol (PEG) are frequently used for this purpose. Hydrophobic membrane proteins partition into the relatively hydrophobic PEG layer, while soluble proteins partition into the hydrophilic dextran layer. This approach has been used successfully with cyanobacteria, which possess intricate subcellular membrane organizations [37]. Another partitioning method exploits the temperature dependent phase properties of the detergent Triton X-114. Bordier *et al* developed a method where initial extraction is carried out in Triton X-114 at temperatures between 0°C – 4°C [38]. Heating the solution to above 20°C induces phase partition; hydrophobic membrane proteins partition into the detergent while the less hydrophobic proteins partition into the aqueous phase [19]. The Triton X-114 phase partitioning approach been used successfully in several mycobacterial studies [39-42]. Other phase partitioning systems that have been used include a system comprised of the protein stabilizing co-solvent trifluoroethanol (TFE) and chloroform [43]. This system has been successfully used with

Escherichia coli [43] and *Staphylococcus aureus* [44] membranes. Another popular binary phase partitioning system, that utilizes a mixture of chloroform and methanol, is useful for delipidation [35, 45]. Proteins aggregate at the chloroform-methanol interface, while lipids are solubilized into the chloroform fraction. Precipitation is another approach used to enrich membrane fractions and remove contaminants, including lipids that might interfere with downstream analysis [23].

Precipitation is frequently carried our using cold trichloroacetic acid (TCA) and acetone. These two reagents can be used independently, as a mixture, or sequentially. Precipitation is often used in conjunction with other enrichment techniques.

Removal of contaminating high abundance soluble proteins can be carried out by washing membrane fractions with high salt or high pH solutions or by using a combination of both. A high pH, cold sodium carbonate wash [24] has been used extensively to enrich membrane proteins by removing peripheral and soluble proteins [26, 46-48]. The ions disrupt the electrostatic interactions between membrane-associated proteins, and the high pH opens the membrane vesicles preventing entrapment of the associated proteins inside the vesicles. Carbonate washes can be used along with other enrichment techniques. Srivastava *et al* combined sucrose density gradient centrifugation, with dextran/PEG phase partitioning, and the carbonate wash procedure, to identify 76 proteins from the thylakoid membrane fractions of the cyanobacterium *Synechocystis* sp. PCC6803 [37]. While the carbonate wash procedure is one of the most popular approaches for removal of non-membrane protein contaminants, there are some reports that the procedure either has no effect [49] or results in the loss of some membrane proteins [28, 50]. In the latter cases, EDTA and sodium bromide (NaBr) have been used

successfully in place of carbonate [51]. Washes with sodium chloride (NaCl) have also been used either on their own or in addition to the carbonate wash [23].

A lot of attention has been paid to the issue of solubilization of proteins prior to 2-DE analysis. Most efforts have relied on the use of detergents or organic solvents. Strong ionic detergents like sodium dodecyl sulfate (SDS) and lithium dodecyl sulfate (LDS) are highly effective in denaturing membranes and extracting proteins from them [46], however they are incompatible with IEF at all but the lowest concentrations, and also interfere with downstream MS analysis [23]. Mild ionic detergents like Triton X-100 (TX100) have been used for several years. While they are compatible with IEF and downstream MS analysis at higher concentrations than ionic detergents, they have relatively low efficacy of solubilization. The introduction of sulfobetaines like CHAPS provide significantly improved solubilization abilities compared with non-ionic detergents. The development of aggressive zwitterionic amidosulfobetaines like ASB-14 and C8Φ by the Rabilloud group in the 1990s led to even more spectacular advances in the solubilization of membrane proteins [52]. Combined usage of these zwitterionic amidosulfobetaines with traditional non-ionic detergents like TX100 or n-dodecyl-b-D-maltoside (DM) can provide excellent separation of OMPs by 2-DE [26].

More recently there has been development of a new class of detergents called acid cleavable or acid labile surfactants such as ALS [53-61]. These detergents provide similar solubilizing power to the ionic detergents, but degrade on exposure to low pH, thus creating minimal interference with MS analysis [23, 35]. Many of these are commercially available under brand names such as the ionic RapiGest (Waters), the zwitterionic PPS (Protein Discovery) and ProteaseMax (Promega) [23]. Several successful membrane

proteomic studies have been carried out using this class of reagents, however, there is a caveat; they can co-precipitate with hydrophobic peptides decreasing yields [35, 62]. Another new reagent with promising potential is the proprietary MS-compatible detergent Invitrosol (Invitrogen). Invitrosol is not acid cleavable, but elutes in distinct peaks from most digested peptides, providing minimal interference during MS analysis [61].

Like detergents, organic solvents such as methanol, acetone and acetonitrile, trifluoroethanol (TFE), and organic acids like formic acid, and trifluoroacetic acid (TFA) are also powerful reagents for the solubilization of membrane proteins [23]. These have seen increased usage in some of the quantitative non-gel based methods and are discussed in greater detail in later sections.

Chaotropes are another class of chemicals that aid in solubilization and reduction of precipitation during IEF. Chaotropes enable the unfolding of proteins and exposure of central hydrophobic regions. When used in conjunction with detergents, the resultant mixtures have greatly augmented solubilizing power. The chaotropes urea and guanidium chloride have been used in protein studies for several years. Rabilloud *et al* pioneered the use of thiourea, which significantly improves solubilization of membrane proteins when used in combination with urea and zwitterionic detergents [63].

Another important component of protein solubilizing mixtures are reducing agents, which are added to completely unfold proteins containing disulfide bonds. β-mercaptoethanol (BME) or dithiothreitol (DTT) were initially used for reduction but these have low stability. Tributyl phosphine (TBP) functions stoichiometrically and hence can be used at lower concentrations, and because it has no charge, reducing power is consistent across the pH gradient [64].

Traditional tube gels made with carrier ampholytes were amongst the most problematic parts of 2-DE. The current preferred approach for IEF is through the use of immobilized pH gradients (IPGs), which are formed by copolymerization of buffering and titrant groups of acrylamido derivates into a gel matrix and stabilization onto a firm immobilized support. IPGs address many of the issues of the older-style tube gels by offering easier handling of the gels, increased load capacities, greater physical stability, and lower variations from run to run [63, 65, 66]. The use of narrow range IPGs allows even higher loads and the detection of more protein spots in a given pH range [67], while the development of alkaline IPGs has allowed the resolution of some basic proteins that were not previously amenable to separation by IEF [47]. Some more radical approaches to address the issue of precipitation during IEF involve the elimination of the IEF step altogether.

With the increased resolving power afforded by MS/MS over conventional primary MS PMF, it is possible to simply couple SDS-PAGE 1-DE to mass spectrometric analysis. Such an approach has been used in several studies, including a study by our group on spore preparations from *Bacillus* [46]. Nevertheless, it is still desirable to use orthogonal or semi-orthogonal separation techniques, prior to MS analysis, for more complex samples. Some common approaches include the use of 1-DE gels paired with some form of chromatographic separation e.g. ion exchange chromatography (IEC/SDS-PAGE) or replacement of the IEF step from 2-DE with other electrophoretic methods [7]. Two such methods are 16-benzyldimethyl-n-hexadecylammonium chloride (BAC)-PAGE and the closely related cetyl trimethyl ammonium bromide (CTAB)-PAGE [68-70]. These techniques make use of the cationic detergents BAC and CTAB respectively,

and rely on the same separation principle as SDS-PAGE i.e. on the basis of molecular weight. However, because BAC and CTAB bind to proteins in different ratios as compared to SDS, the use of combinations of BAC-PAGE with SDS-PAGE or CTAB-PAGE with SDS-PAGE can yield decent orthogonal separations. Using a BAC/SDS-PAGE based separation approach, Bisle *et al* were able to identify several hydrophobic proteins from *Halobacterium salinarum* (including the model protein for multi-TMD containing α-helical proteins, bacteriorhodopsin) while Schluesener *et al* were able to identify several membrane proteins from *Corynebacterium glutamicum* [7, 45, 71]. Likewise, it is possible to use 2 successive rounds of SDS-PAGE, with varying parameters of acrylamide concentrations, buffer components, etc., which significantly alter electrophoretic migration patterns, to resolve mixtures of hydrophobic proteins [68]. Williams *et al* used a bicine buffer based SDS/SDS-PAGE approach and demonstrated significant improvements in the separation of membrane proteins [72].

In all the methods described above, electrophoresis in the first dimension is carried out under protein denaturing conditions. It is also possible to carry out separations under non-denaturing conditions. Blue native (BN)-PAGE utilizes the binding affinity of the negatively charged dye molecule Coomassie Brilliant Blue (CBB) for proteins. The use of relatively mild non-ionic detergents like Triton X-100 and DM for BN-PAGE can keep proteins active and protein complexes together during the first electrophoretic separation [50]. Lasserre *et al* were able to resolve 160 *E. coli* proteins, which included 124 membrane proteins, by the BN/SDS-PAGE method [73]. Clear native (CN)-PAGE, which substitutes CBB with negatively charged detergents like sodium deoxycholate, is another effective non-denaturing separation technique for complex proteomes [7].

However, this technique is not commonly employed in bacterial studies, as it does not provide sufficient resolution for their relatively unmodified proteins [7].

The difficulty in obtaining high sequence coverage of membrane proteins due to lack of, or decreased, access to tryptic digestion sites can be addressed through the use of additional alternative proteases like Lys-C [15] and chymotrypsin [28] or by chemical cleavage with agents like cyanogen bromide (CNBr) [74, 75]. The chemical agent CNBr, which generates large peptides, or the enzymes Lys-C and chymotrypsin are often used in conjunction with trypsin to yield peptides that can be easily analyzed by MS. Use of compatible solvents such as formic acid for CNBr or methanol for trypsin provide the proteolytic agents improved access to hydrophobic proteolytic sites while retaining their activity [22, 29, 76]. These strategies are commonly used in many of the gel-free proteomic methods, and are discussed in later sections.

Quantitative 2-DE based proteomics

The modified electrophoretic techniques described above have been used successfully to carry out several proteomic studies in microbes and have resulted in vastly improved coverage of membrane proteins [33]. Nevertheless, while extremely informative, the bulk of these studies have mostly yielded non-quantitative results. These have been either in the form of reference mappings with localization of proteins on gels and subsequent identification, or protein expression profiling studies where the presence or absence of protein spots between compared conditions is revealed [5]. As stated earlier, simply identifying and cataloguing the protein complement of an organism or even revealing binary differences between conditions is not sufficient to get a true picture of the dynamic interactions occurring within. There is a need to quantify changes in

protein expression between compared cell types, during time courses, between diseased and non-diseased states, etc. The initial approaches to quantitative proteomics were extensions of gel-based techniques; many of which were semi-quantitative in nature. We have previously reported the response of change in growth medium on the expression of several receptor proteins in outer membrane of the Gram-negative bacterium *Caulobacter crescentus* [26]. However, these changes were discerned only by visualization and no efforts were made to quantify them. A more sophisticated approach makes use of densitometric analysis of gel images of the compared conditions using software packages [77]. This can yield quantitative information on protein expression; however, gel-to-gel variation remains a serious problem.

One of the first truly quantitative proteomics technologies to be developed was two dimensional fluorescence difference gel electrophoresis (2D-DIGE) [78-80]. It is essentially an extension of standard 2-DE that incorporates fluorescent tagging of lysine or cysteine residues. Most often tagging is carried out using the cyanine dyes Cy3™, Cy5™ and Cy2™(GE Healthcare). These dyes have an N-hydroxysuccinimidyl ester reactive group that enables protein binding via the ε-amino group of lysine residues. Presence of bound dye molecules does not significantly alter the migration of proteins in either dimension of electrophoresis. Up to three different samples to be compared can be labeled with the different dyes and then mixed and resolved together on a single gel. In practice often only two samples or conditions are compared while the third dye is often used for controls or standards. Because samples from the conditions being compared are run together on the same gel, the method reduces gel-to-gel variability as compared with prior 2-DE densitometric quantitative methods. However, the technique suffers from

lower sensitivity as compared with silver staining and a relatively narrow dynamic range of resolution, in addition to being hampered by all the other limitations of standard 2-DE. There are additional complications in quantification due to difficulty in resolving co-migrating spots from different proteins. In spite of these limitations, this approach has been used quite extensively in microbial proteomics. In their 2D-DIGE analysis of the effect of benzoic acid treatment on *Escherichia coli*, Yan *et al* identified 179 differentially expressed proteins, several of which were OMPs [80]. Nevertheless, most 2D-DIGE studies have focused on cytosolic proteins due to the limitations of IEF/SDS-PAGE in resolving membrane proteins. To overcome these limitations, Bisle *et al* used 2D-DIGE based on BAC/SDS-PAGE rather than IEF/SDS-PAGE in their analysis of the membrane proteome of the halophilic archaeon *Halobacterium salinarum* [45]. To further enhance the recovery of membrane proteins, membrane vesicles were enriched using sucrose density gradient centrifugations and membranes were delipidated using a methanol-chloroform extraction protocol. Using a combination of this modified DIGE approach with another non-gel based proteomic technique (discussed ahead) Bisle *et al* were able to identify an impressive 155 membrane or membrane associated proteins, 101 of these containing TMDs. Equally importantly, a good quantitation correlation was seen between the two distinct proteomic approaches thereby validating both the methods.

Gel free and shotgun proteomics

The modified 2-DE methods described above have greatly improved our ability to cope with membrane proteins, however the overall resolving power of these approaches has not been significantly greater than just 1-DE on its own [23]. By the late 1990s and early 2000s it had become clear that a parallel alternative approach to classical 2-DE +

MS based proteomics needed to be pursued [20]. Many of these alternative approaches exploited the recent advances in mass spectrometry. The gel-based methods described above used the gels for quantification of proteins, while the MS step was only used for identification.

In contrast, most of the newer so called "shotgun methods" utilize the power of MS for quantification of proteins. Most of the shotgun approaches rely on the generation and analysis, by MALDI-MS or ESI-MS, of proteolytic peptide mixes rather than intact proteins, following some form of peptide separation. This is often referred to as a 'bottom-up' approach [17]. In a tactic analogous to 2-DE, which uses two different electrophoretic techniques to achieve intact protein separation, separations of peptides in these shotgun methods are usually achieved by coupled orthogonal chromatographic techniques. Strong cation exchange (SCX) chromatography is often used as the first method of separation. Peptide mixtures are loaded onto a strong cation exchange immobilized phase and the samples gradually eluted in a gradient of increasing salt concentration. Most commonly, each fraction is subsequently separated by reverse phase (RP) chromatography, which separates peptides on the basis of their hydrophobicity. RP separation is often carried out using a high performance liquid chromatography (HPLC) system directly in-line with a mass spectrometer (RP-HPLC-MS). The primary MS and MS/MS data acquired are used to search databases and identify proteins. Greater the number of peptide identified corresponding to a given protein, better is the confidence in the identification of that protein.

In one of the earliest demonstration of the power of such a shotgun approach, researchers from the Yates group at The Scripps Institute, La Jolla carried out an

extensive analysis of *Saccharomyces cerevisiae* [30]. They modified an online orthogonal separation and identification protocol, consisting of strong cation exchange (SCX) directly linked to reverse phase (RP) chromatography and tandem MS/MS, originally described by Link *et al* [20]. In a procedure they termed multidimensional protein identification technology (MudPIT), Washburn *et al* married this modified separation scheme to an enhanced sample preparation methodology. Soluble fractions were proteolytically digested initially with the endoprotease LysC followed by trypsin. Insoluble fractions were dissolved in 90% formic acid, initially digested with cyanogen bromide (CNBr), and thereafter with the endoprotease LysC and trypsin. Formic acid is an efficient solubilizing agent for membrane proteins, and CNBr, which remains active in the presence of formic acid, is capable of cleaving hydrophobic proteins that often lack tryptic digestion sites in their TMDs. Subsequent digestion with LysC and trypsin results in the generation of smaller peptides, which are better suited for MS analysis. The peptides were analyzed in a Finnigan LCQ ion trap mass spectrometer paired with a nano-LC electrospray ionization source after online separation via a directly coupled SCX-RP chromatographic system. Protein identifications were made using the SEQUEST algorithm (Thermo Finnigan, San Jose, CA) [81]. In the study, 1484 unique proteins were identified from 5540 spectra from MS/MS analyses of soluble, and lightly and heavily washed insoluble fractions. Of these, 131 proteins were predicted to be membrane proteins with three or more TMDs. 26 of these had 10 or more TMDs and one even had 20 predicted TMDs. The resolving power and throughput of this approach was demonstrated in a comparison with the previous record for highest number of proteins identified in a single experiment, in which 279 identifications were made by 2-DE in a

single experiment [30]. Equally impressive was the excellent recovery of membrane proteins by this method. Analysis of the results also showed that the method was largely unbiased against identification of hydrophobic proteins, proteins at extremes of molecular weight, and at extremes of pIs, all of which have proved problematic for 2-DE analyses.

Even though classical 2-DE based separations of intact proteins are not employed during most shotgun proteomic approaches, solubilizing membrane proteins and obtaining peptides from those proteins, nevertheless, remains a challenge [22, 29]. Some of the advances made to enhance recovery of membrane proteins by 2-DE have been adapted to the shotgun methods and several novel approaches have also been made to improve recovery of membrane proteins by these techniques. Wu *et al* introduced a method known called high pH and proteinase K (hpPK) or 'membrane shaving', which exploits the ability of high pH solutions to disrupt membranes without denaturing proteins, and the slightly reduced activity of proteinase K under those conditions, which produces peptides of optimal size for MS analysis [29]. They carried out a MudPIT analysis of brain homogenate using their hpPK protocol. Analysis of the peptides by ESI-MS/MS, on a Thermo Finnigan LCQ-Deca mass spectrometer, resulted in the identification of 1610 proteins with 2 or more peptides at >95% confidence. 454 of these proteins (~29%) were predicted to have between 1 – 23 TMDs reflecting no bias against TMD containing proteins by the method. In addition to the extensive unbiased proteomic coverage offered by this approach, sometimes referred to a 'shaving', it also allows the detection of post-translational covalent modifications and the uncovering of membrane topologies. In a variation of this method Rodriguez-Ortega *et al* carried out tryptic digestion of exposed proteins of intact group A *Streptococcus* (GAS) cells to identify

surface proteins as vaccine candidates [82]. SCX-RP-MS/MS analysis of the peptides identified 68 proteins that were predicted to be surface-associated, including most of the protective antigens described in the literature. A similar analysis of *Bacillus subtilis* by Tjalsma *et al* identified 41 proteins including several containing TMDs [83].

Blonder *et al* introduced a novel sample preparation approach during their analysis of the membrane subproteome of the Gram-positive bacterium *Deinococcus radiodurans* [76]. Even though it is classified as a Gram-positive organism, it has a particularly robust envelope with many components typical of Gram-negative bacteria, and as such poses a challenge for membrane proteomic analysis. *D. radiodurans* cells were disrupted in a French Pressure Cell and membranes were harvested by centrifugation. Contaminating proteins were removed using the carbonate wash procedure, and membranes were solubilized in a 60% methanol solution. Proteolytic digestion was carried out directly in this organic-aqueous solution. Peptides were separated using RP chromatography and analyzed using a Thermo Finnigan LCQ ion trap mass spectrometer. 503 proteins were identified. PSORTb [84, 85] predicted 215 to be IMPs and 53 to be OMPs. Independently, on the basis of their GRAVY values, 135 proteins were predicted to be hydrophobic. The identified membrane proteins had between 1-16 TMDs. Impressively, several of the putative membrane proteins were identified with at least one hydrophobic peptide in the MS analysis. This fact and the extensive coverage of the membrane proteome amply demonstrate the power of the method to recover hydrophobic peptides for MS analysis.

In a separate study Blonder *et al* were also able to demonstrate that trypsin retained sufficient activity for thorough digestion of solubilized membranes in the

methanol solvent system [86]. Using methanol-solubilized membranes from *Halobacterium halobium*, which were enriched for the model membrane protein bacteriorhodopsin, they were able to demonstrate that methanol solubilization makes the entire protein, including the TMD regions, accessible for tryptic digestion [86]. This was achieved by comparing theoretical and observed digests of the protein. They were also able to identify 40 additional *H. halobium* membrane proteins during these experiments. In a parallel series of experiments in the same study they also analyzed human epidermal plasma membranes. Sucrose gradient fraction were enriched for α6-integrin, a common marker for this fraction, and 117 unique proteins, 64 of which were known to be plasma membrane associated, were identified. Zhang *et al* used methanol-assisted solubilization on *E. coli* inner membranes to identify 358 proteins, 159 of which were predicted to be integral membrane proteins [87]. Goshe *et al* coupled this technique with a cysteine-specific biotinylation affinity strategy to improve the recovery of low abundance proteins [31]. Using such a combined approach, they were able to identify 89 proteins from *D. radiodurans*, 40 of which were predicted to be membrane proteins with 1-9 TMDs. Application of the same strategy to the Gram-negative bacterium *Pseudomonas aeruginosa* yielded 768 protein identifications [88]. 333 of these of which were classified as membrane proteins, while 195 were classified as hydrophobic on the basis of GRAVY scores.

In their comprehensive analysis of the quantitative differences between the membrane proteomes of a *Corynebacterium glutamicum* L-lysine producing strain and a standard lab strain, Fischer *et al* combined elements of both the hpPK and methanol-solubilization strategies [28]. Trypsin and CNBr were used to digest intact membranes to

identify exposed domains of membrane proteins. In parallel, aqueous trypsin digestion was used after high salt wash to remove membrane-associated proteins. In the latter case, membranes were solubilized with methanol and hydrophobic domains were digested by trypsin and chymotrypsin in the methanol-aqueous solution. Using this approach, they were able to identify 326 integral membrane proteins, representing approximately 50% of predicted membrane proteome. Similarly, Blackler *et al* combined elements of hpPK with organic solvent-based solubilization [89]. The original hpPK or 'shaving' method only yields protease accessible peptides (PAPs). The modified method of Blackler *et al* augments those peptides with peptides obtained from methanol solubilized, CNBr digests of the 'shaved' membranes i.e. membrane embedded peptides (MEPs).

Barrios-Llerena *et al* utilized a solubilization strategy more commonly used in 2-DE separations for their shotgun analysis of the cyanobacterium *Anabaena variabilis* ATCC 29413 [90]. Cell pellets were solubilized in a buffer containing the 7M Urea and 2M Thiourea and a mixture of the detergents ASB-14 and dodecyl-b-d-maltoside (DM) at 2% and 1% concentrations respectively, and subsequently disrupted by grinding under liquid nitrogen. Protein fractions were then digested with the endoproteases Lys-C followed by trypsin. The resulting peptides were separated by SCX and followed by RP in line with a nano-ESI-MS/MS analysis on a QSTAR XL instrument. A total of 646 proteins corresponding to ~13% of theoretical proteome were identified. 44 of these proteins (~5% of total predicted IMPs) were localized to the inner membrane, while 6 (~7% of predicted OMPs) were localized to the outer membrane. While membrane proteins were underrepresented in the results, it is not particularly surprising given that no special membrane enrichment techniques were used. The results were all the more

encouraging when they were compared with a previous 2-DE based analysis of the same organism, using a similar sample preparation protocol, which resulted in the identification of only 9 cytoplasmic proteins (~1% of total predicted cytoplasmic proteins).

Researchers from the Yates group, showed that the addition of the new MS-compatible detergents like Invitrosol and the acid labile surfactants PPS and RapiGest, could dramatically improve protein solubilization and proteolytic efficiency [61], confirming similar observations by Blackler *et al* and Ruth *et al* [23, 60]. By pooling results from multiple different detergent-assisted trypsin digestions in organic-aqueous or aqueous systems, they were able to identify over 700 proteins from mouse pancreatic cell line samples. They showed that the choice of detergent and organic solvent introduced considerable variability in the number and types of peptides analyzed by MS, and hence in the proteins identified. In their hands the acid labile detergent PPS, in aqueous buffer or 80% acetonitrile, yielded the highest number of hydrophobic peptides, and thus had potential in the analysis of membrane fractions.

Masuda *et al* devised a protocol that makes use of the bile salt sodium deoxycholate (SDC) to aid tryptic digestion of membrane proteins [62]. They demonstrated that the presence of the 'phase transfer surfactant' SDC improves the solubility of hydrophobic proteins and at the same time enhances the activity of trypsin while improving its accessibility to the membrane proteins. SDC can be removed post-digestion by addition of a water immiscible solvent into which it partitions, leaving the extracted peptides in the aqueous phase. Using this protocol, they were able to identify 1450 proteins from human cervical cancer HeLa cells by SCX-RP-MS/MS. 764 (53%) of

these were predicted to be membrane proteins. In a modification of this protocol they added the surfactant sodium laurylsarcosinate (SLS/sarkosyl) to SDC and used immobilized trypsin to digest peptides [91]. Using this approach they were able to identify 1453 *E. coli* proteins, 545 of which were membrane proteins.

Modifications are continuously being carried out to improve solubilization and digestion of proteins as well as recovery and separation of peptides for the shotgun methods. Zhong *et al* developed a microwave-assisted acid hydrolysis (MAAH) protocol, which builds on the original Edman sequencing technique, to digest proteins prior to LC-MS analysis [92]. Membrane enriched human breast cancer cell line lysates were solubilized in 25% trifluoroacetic acid (TFA) and subjected to 10 minutes of microwave irradiation. 119 proteins were identified, 41 of which were integral membrane or membrane proteins containing up to 12 TMDs. Speers *et al* demonstrated that increasing temperature up to 60°C during RP-LC separations of peptides could significantly improve the recovery of hydrophobic peptides, with a 500% increase in peptide identification and a 400% increase in protein identification as compared to standard room-temperature separations [93]. Lu *et al* developed a protocol known as 'tube-gel' digestion in which detergent solubilized cell lysates or membrane preparations were directly incorporated into polyacrylamide gels without electrophoresis [94]. This was achieved by mixing the protein sample with monomeric acrylamide, and carrying out polymerization of the gels in the presence of the sample. Detergents were removed by washing after polymerization and peptides were obtained by tryptically digesting the samples in the gel matrix using standard in-gel digestion protocols followed by standard peptide gel-extractions protocols. With such an approach they were able to identify 178

membrane proteins from prostrate cancer cells using a nano-LC-ESI-MS/MS system. Han *et al* modified this system to include a reduction and alkylation step prior to gel incorporation [95]. Using their modification, they were able to identify 330 integral membrane proteins containing up to 19 TMDs from kidney cells. They obtained good recovery of hydrophobic peptides, and high sequence coverage of 14.1 peptides per protein. Zhou *et al* showed that increasing the temperature during polymerization and gel-embedment, and carrying out digestion in the presence of sodium deoxycholate significantly enhances the recovery of hydrophobic peptides [96].

In a separate study, they also demonstrated that it is possible to incorporate proteins into gels by absorption into vacuum-dried polyacrylamide gels rather than incorporation during polymerization [97]. They were able to obtain 190 membrane protein identifications from rat liver cell membrane preparations as compared to 152 identifications with the tube-gel copolymerization approach. With these modifications, they were able to identify 326 integral membrane proteins from rat liver cell plasma membrane fractions. These gel-enhanced digestion techniques enable the use of higher concentrations of tryptic enzyme incompatible and MS incompatible detergents and chaotropes for solubilization since they are removed prior to digestion.

MudPIT style analysis, comprising of SCX coupled with RP, is the most common separation method for shotgun proteomics [23]. One of the main reasons for this is that it is relatively easy to integrate SCX and RP in an online fashion and directly couple them to MS instrumentation, thereby enabling automation. However alternative peptide separation methods exist as well. IEF based separation of peptides is a powerful alternative. Gan *et al* and Chong *et al* demonstrated, on membrane preparations from the

cyanobacterium *Synechocystis* sp. PCC6803 and the archaeon *Sulfolobus solfataricus* respectively, that IPG-IEF based fractionation of peptides can result in excellent separation of peptides and MS identification of proteins [98, 99]. The approach was also used successfully by Scherl *et al* in their analysis of antibiotic resistant *Staphylococcus aureus*, which is discussed in a later section [100]. However, the IPG-IEF peptide fractioning method requires the gel matrix to be cut into small pieces after IEF, and is quite tedious and not particularly adapted to automation or high throughput analyses. The OFFGEL™ system (Agilent), a free flow isoelectric focusing method, in which IEF and subsequent peptide recovery can both be carried out entirely in the liquid phase, offers an alternative [101]. OFFGEL-IEF fractionation has been successfully coupled with RP chromatography [102] as well as capillary electrophoresis [103] to provide robust post-digestion separation of peptide mixtures prior to MS. Another method, GeLC offers a different approach to orthogonal separation [7]. It combines the use of standard 1-DE based separation of proteins prior to digestion, followed by liquid chromatographic separation of peptides, usually by RP, after digestion.

Quantitative shotgun proteomics

Powerful as they are, shotgun methods such as MudPIT are not inherently quantitative [104]. However, they can be relatively easily coupled with other techniques to provide quantitative results. Stable isotopic labeling methods have provided a powerful platform for quantitative proteomic studies [23]. Isotopic labeling is based on the premise that the physical and chemical properties of isotopically labeled biomolecules are identical, but they can be distinguished by mass spectrometry. Typically, different isotopic variants of the stable isotope labels are incorporated into the samples to be

compared after which the samples are combined and subjected to orthogonal separations and analysis by mass spectrometry. Relative quantification of the peak areas of the peptide ion MS spectra is used to calculate relative protein abundance [23]. This approach may also be referred to as stable isotope ratio mass spectrometry (SIRMS) [17]. From the point they are combined, up to the generation of mass spectra, the samples to be compared are always together and are exposed to identical conditions, hence sample to sample variations are minimized. Both gel based and non-gel based proteomic techniques are compatible with stable isotope labeling. Depending on the nature of the isotopic labels used, and the proteomic strategy used, the labels may be incorporated either pre-digestion, during digestion, or post-digestion. There are two broad approaches for using stable isotopes in proteomics - metabolic i.e. *in vivo* labeling, and chemical or enzymatic i.e. *in vitro* labeling.

Oda *et al* provided one of the first demonstrations of *in vivo* metabolic labeling through the use of ^{15}N-enriched cell culture media in their comparisons of protein abundance in wild type and mutant *Saccharomyces cerevisiae* cells [105]. The method, known as $^{15}N/^{14}N$ metabolic labeling, allows the introduction of the stable isotope at one of the earliest points in the experimental chain, often through the use of compounds such as ^{15}N-labelled ammonium sulfate (NH_4SO_4) as sole nitrogen source [7]. Cells to be compared, are grown in media containing the light or heavy isotopes, and then mixed and processed together. This reduces differential protein loss during sample preparation and separation. Comparison of the mass peaks from the heavy and light isotopes is used to calculate the relative abundance of the proteins in the sample. Other stable isotopes such as ^{13}C, ^{18}O, and ^{2}H (^{2}D) can also be used for in a similar manner, though the use of ^{2}H is

less common due to its effect on enzymatic activity [105]. Even though metabolic labeling-based quantitative proteomics techniques are increasingly being applied to higher organisms, incorporation of the labels is quite complex [106]. The methods are, however, particularly well suited for use in microbial systems that can be grown in culture. Using a ^{15}N-based metabolic labeling strategy, Becher *et al* were able to identify over 1700 proteins and quantitate 1450 proteins, without any bias against membrane proteins, in their large-scale analysis of the *Staphylococcus aureus* proteome [107].

Mann and colleagues demonstrated another means of introducing stable isotopic labels, during protein synthesis, through labeled amino acids [108]. The stable isotope labeling by amino acids in cell culture (SILAC) method involves incorporation of heavy amino acids, typically ^{13}C-arginine and ^{13}C-lysine, but also heavy leucine, isoleucine, or tyrosine, during growth. Like the ^{15}N metabolic labeling method, SILAC is also more suited to microbes that can be cultured and grown, as compared with higher organisms e.g. tissue. It also requires an auxotrophy for the labeled amino acid, and necessitates growth on minimal media, thereby potentially introducing a bias towards that growth condition. In spite of these possible drawbacks, SILAC is extremely popular and has seen widespread usage in microbial proteomics [109].

In vitro labeling techniques such as enzymatic and chemical labeling are also very popular. Yao *et al* introduced the $^{18}O/^{16}O$ enzymatic or proteolytic labeling technique, in which the ^{18}O label is incorporated into peptides during proteolytic digestion in the presence of $H_2{}^{18}O$ [110]. Tryptic digestion in the presence of $H_2{}^{18}O$ introduces 2 ^{18}O atoms into the carboxy termini of digested peptides thereby creating a 4Da mass shift in comparison to proteins digested in the presence of normal water ($H_2{}^{16}O$). Other

proteolytic enzymes such as Glu-C, Lys-C, and chymotrypsin can also be used for incorporation of the heavy oxygen atom. The method is extremely powerful, however it has not seen widespread acceptance in proteomic analyses due to variability in the ^{18}O incorporation step, and post-labeling back-exchange between ^{18}O and ^{16}O atoms both of which can complicate quantitation [17]. Furthermore, due to the dependence on the proteolytic enzyme for incorporation of the stable isotope, the method could bias itself against highly hydrophobic proteins that have reduced access to proteases.

Isotope coded affinity tagging (ICAT® - Applied Biosystems) is a prototypic chemical stable isotopic labeling strategy, and remains one of the most commonly used methods for quantitative proteomics [111]. Cysteine-containing residues from proteins are labeled with either light (^{12}C) or heavy tags (^{13}C) that also have a biotin group. A modified form of the reagent has a cleavable tag and is sometimes referred to as cleavable ICAT (cICAT). After tagging, proteins are mixed in equal amounts, and digested tryptically, after which they are desalted by cation exchange chromatography and purified using avidin. Mass spectrometry results in dual peaks, the ratios of which correspond to the relative abundance of the starting analytes. ICAT is an extremely robust method, and is often the first choice method for quantitative proteomic studies. However, ICAT is not without limitations. Many proteins, particularly membrane proteins, either completely lack or have limited number of cysteine residues making it difficult to analyze them by ICAT. Additionally, the reagents used for membrane solubilization such as the detergent SDS can interfere with ICAT alkylation. While the use of cysteine based tagging may be a limitation due to the low frequency of cysteine residues in proteins, cysteine specific tagging can be exploited for novel applications. In a unique variation of

ICAT, termed as OxiCAT, researchers in the Jakob group have modified the ICAT chemistry to measure the susceptibility of thiol containing proteins to oxidative stress [112].

Two other tagging methods similar to ICAT are HysTAG [113] and ICPL [114]. Like ICAT, HysTAG is a cysteine-specific label that makes use of a decapeptide, which binds cysteine residues through a disulfide bond. The tag contains a string of 6 histidine resides that allow recovery by immobilized metal affinity chromatography (IMAC) e.g. by Ni-NTA columns, and an internal tryptic cleavage site that allows the removal of the bulk of the tag after purification. The HysTAGs themselves are available in heavy and light versions with a 4Da difference in masses. Isotope coded protein labeling (ICPL), in contrast, makes use of an isotopic tag that binds to lysine residues thereby side-stepping some of the issues introduced by cysteine-labeling. Bisle *et al* have successfully used ICPL in conjunction with BAC/SDS-PAGE 2D-DIGE to identified and quantify 175 proteins, including 101 integral membrane proteins in *Halobacterium salinarum* [45]. As noted earlier, there was excellent correlation between quantification by both methods.

Most of the above-described methods offer relative quantification of proteins between compared samples. In contrast, in the absolute quantification of proteins (AQUA) approach, predetermined amounts of stable-isotope labeled synthetic peptides are spiked into protein digests [115]. A comparison of the mass spectrometric signal of these synthetic peptides with the endogenous peptides from the samples can be used to determine absolute quantification of the samples.

While isotope based quantitative proteomics remain extremely popular, there are nevertheless, some complication associated with isotopic labeling. Even though the

biological complexity of the samples being compared remains the same, the mass spectrometric complexity is doubled with a simultaneous reduction in individual peak intensities [116]. Two alternatives to isotopic labeling are label-free quantitative proteomics or isobaric labeling techniques. Label free strategies usually involve one of two approaches to provide relative protein expression levels – (i) measurement and comparison of mass spectrometric signal intensities of spectra belonging to the protein of interest, or (ii) comparative counting of the number of spectra identifying the protein of interest [115, 117]. The latter can be extended to provide an indication of absolute protein expression levels through the protein abundance index (PAI), which is the ratio of number of observed peptides to the theoretical maximum number of peptides that could have been obtained from the protein, or the logarithmically related exponentially modified PAI (emPAI) [115, 117]. Absolute protein expression (APEX) is a modification of PAI that takes into account a correction factor based on machine learning to provide more accurate quantification [118]. Isobaric labeling techniques like TMT [119] and iTRAQ [116] offer another alternative to isotopic labeling, and are discussed in depth in the next sections.

Isobaric tagging for relative and absolute quantitation (iTRAQ)

Ross *et al* developed an elegant method for multiplexed quantitative proteomics that does not depend on gel-based separation or on the use of isotopic tags [116]. The method is known as isobaric tagging for relative and absolute quantitation (iTRAQ® - Applied Biosystems). The initially described iTRAQ procedure combines the separation power of MudPIT with the ability to multiplex samples and quantify the results.

iTRAQ makes use of a set of amine reactive reagents of equal mass that can be used to generate derivatized peptides which are indistinguishable from each other during chromatographic and electrophoretic separation and in single MS mode, but which yield signature ions for each variant of the reagent, when fragmented in MS/MS mode. In their original form the iTRAQ reagents consist of a set of four reagents, each of which is a molecule made up of three functional parts. The amine specific reactive part (peptide-reactive group) - an NHS ester, enables the derivatization of peptides at free amine groups. The balance part (carbonyl group) and the reporter part (N-methylpiperazine derivatives) of the reagent have a combined mass of 145 Da in all the four forms of the mix, but fragment differently to result in reporter ions with masses of 114, 115, 116 and 117 Da and balance parts with masses of 31, 30, 29 and 28 Da respectively. The differential masses of the balance and reporter parts are achieved through the use of different combinations of atoms of ^{14}C, ^{15}N and ^{18}O isotopes in the two parts. Biological samples that are to be compared are processed in parallel. In general, proteins or whole lysates are reduced, alkylated, tryptically digested and labeled with the iTRAQ reagents. In a similar manner to the ICPL method described earlier, the iTRAQ reagents attach at free amine residues at the N-terminus and at lysine residues. Free amine groups are present in all peptides in contrast with cysteines, which are only present in 95% of peptides. This affords the technique an advantage over ICAT and other methods that employ cysteine-modifying tags.

The iTRAQ labeled peptides are then combined in equal amounts on the basis of carefully quantitated protein content, and the resultant mixtures are separated by orthogonal methods. These usually comprise of strong cation exchange (SCX)

chromatography followed by reverse phase HPLC to reduce sample complexity i.e. MudPIT. The separated fractions are subsequently analyzed by tandem mass spectrometry (MS/MS). The primary MS peaks are the sum total of the ionization of peptides from all the combined samples giving strong peaks. In contrast to the complication observed with isotopic methods, there is no increase in mass spectrometric complexity in primary mode. During subsequent MS/MS analysis on the selected peaks, CID causes the isobaric tags to break allowing the ions coming from the different combined samples to be resolved and quantified relative to each other.

A microbial system, yeast, was used for the first demonstration of the iTRAQ system [116]. *Saccharomyces cerevisiae* strains deficient in the nonsense-mediated mRNA decay pathway (upf1Δ) and the general 5' – 3' decay pathway (xrn1Δ) were compared for quantitative global proteome expression changes with an isogenic wild type (WT) yeast strain using iTRAQ reagents. Cultures of each strain were mechanically disrupted in buffers containing Triton X-100 and guanidine. Whole cell lysates were reduced, alkylated, acetone precipitated, trypsin digested, lyophilized, quantitated, labeled in parallel with the differential 4-plex iTRAQ reagents, and mixed in equal protein content. The WT strain was labeled with the $iTRAQ_{116}$ label, while the xrn1Δ and upf1Δ strains were labeled with the $iTRAQ_{114}$ and $iTRAQ_{115}$, labels respectively. The $iTRAQ_{117}$ reagent was used to label a known amount of a synthetic control peptide for absolute quantification. The mixtures were resolved first using strong cation exchange (SCX) chromatographic separation with a KCl gradient on a PolySulfoethyl-A column followed by reverse phase (RP) separation using an ACN gradient on a C_{18} RP column. MALDI MS/MS was employed to analyze these fractions. The ~4500 peptides generated

were used to identify 1217 unique proteins (with > 95% confidence on peptide ion scores with p-values of 0.05 or lower) using the Mascot search engine [120]. Out of these, 685 unique proteins were identified on the basis of 2 or more unique peptides each. Since no specific effort was made to enrich for membrane proteins, unsurprisingly not many membrane proteins were identified. This latter set of 685 proteins was used for comparative expression analysis between the three strains.

The expression levels of peptides were calculated using the signature ion peak areas using the formula: $\text{area}_{(\text{mutant})}/\text{area}_{(\text{mutant})} + \text{area}_{(\text{wild type})}$. In the comparison between strains xrn1Δ and WT, 48 proteins were found to be up-regulated and 39 down-regulated. In the comparison between strain upf1Δ and WT, 62 proteins were classified as up-regulated, and 23 as down-regulated. These 4-plex iTRAQ results correlated well with pilot 2-plex iTRAQ experiments and an independent ICAT experiment.

The varying workflows between the ICAT and iTRAQ experiments resulted in non-perfect overlap between the two datasets. Nevertheless, there was sufficient correlation between the data to validate the iTRAQ approach. For the proteins identified by both approaches, better peptide coverage was seen with iTRAQ (4.5 peptides/protein) as compared with ICAT (~2 peptides/protein), further underscoring the utility of the approach. Analysis of the identified proteins revealed a significant overlap in the proteins up-regulated in the xrn1Δ and upf1Δ strains, with several proteins involved in general nitrogen metabolism and amino acid biosynthesis are up-regulated in both. In contrast, there was no significant overlap between proteins down-regulated in these strains. Several proteins involved in DNA replication and RNA transcription were down-regulated in the upf1Δ strain, while in the xrn1Δ strain, several proteins involved with

translation were down-regulated. Interestingly, when the data were compared to parallel microarray experiments, limited correlation was found between mRNA expression and protein expression patterns. This further validated previous studies and provided additional justification for the importance of quantitative proteomic analyses alongside transcriptomic studies [6]. Another innovation from this study was the introduction of an internal standard control for absolute quantification.

While the rest of this manuscript is focused on iTRAQ, it is important to note that another isobaric tagging method, Tandem Mass Tagging (TMT) [119], was described slightly earlier than iTRAQ. However, the rapid commercialization of iTRAQ technology by Applied Biosystems resulted in quicker and wider adoption of that method. Furthermore, because the initial TMT method was only 2-plex as compared with the availability of 4-plex iTRAQ from the outset, as well as the preference of a 1Da mass difference between the iTRAQ labels as compared with a 3Da difference between the TMT reagents [121], the iTRAQ reagents were preferred. More recently a 6-plex TMT set has become available commercially from Proteome Sciences / Thermo Fisher Scientific and has seen some use [122, 123]. Recently some additional isobaric reagents like ExacTAG (Perkin Elmer), DiART [124], DiLeu [121] have also been introduced.

iTRAQ remains the most popular isobaric proteomic methods, and has been successfully used to investigate several prokaryotic and eukaryotic systems. Aggarwal *et al* used iTRAQ reagents to study the global protein expression patterns of *Escherichia coli* overexpressing the rhsA element [125]. 780 unique proteins were identified in that study, a majority on the basis of two or more peptides. No specific effort, however, was made to identify membrane proteins. In contrast, Chen *et al* specifically applied the

technique to the analysis membranes of zymogen granules from rat pancreas [126]. Using a combinations of 2-DE and iTRAQ, they identified 101 proteins from the membranes of zymogen granules, including several previously validated or predicted membrane proteins, elegantly demonstrating the power of the approach to analyze membrane proteins.

In one of the early bacterial iTRAQ studies, Redding et al investigated the proteomic response of *Desulfovibrio vulgaris* Hildenborough (DvH) to nitrate stress [127]. The organism is of interest because it plays an important role in global sulfur cycling and has utility in bioremediation and georemediation since it can metabolize uranium and chromate [127]. Contaminated sites often have high concentrations of nitrates, so understanding how the organisms react to their environment is important. Stressed DvH cells grown in 105 mM sodium nitrate ($NaNO_3$) causing 50% growth inhibition were compared with unstressed cells growing under standard conditions. In order to minimize biological variation, triplicate samples were pooled to improve confidence in the results [128]. The $iTRAQ_{114}$ reagent was used to label the pre-nitrate stress control (TC_0) sample, while the iTRAQ115 label was used to label the 480-minute control sample (TC_{480}). The $iTRAQ_{116}$ and $iTRAQ_{117}$ reagents were used to label technical replicates of the 480-minute nitrate stressed samples (TN_{480}). Pooled labeled samples from the different conditions were separated by SCX and analyzed by RP-HPLC-MS using an ABI QSTAR Quadrupole TOF instrument. 1166 proteins were identified by ProQuant (ABI), and 1221 by Mascot [120]. 1047 identifications were common to both software packages. Out of these, 737 proteins were identified with 2 or more unique peptides with >95% CI, providing coverage of approximately 22% of the

total predicted proteome at a high level of confidence. Out of the 3396 predicted ORFs in the genome ~110 either do not have tryptic sites or do not generate peptides between 800 and 3000 Da for MS, so these would not be expected to be detected. 185 proteins were detected with changed expression above internal error levels. After taking into account the error between replicates, 65 proteins appeared to be differentially regulated between experimental conditions in a statistically significant manner. Proteins belonging to the central metabolic pathways and cellular machinery functional groups did not appear to be majorly perturbed. Several hypothetical proteins, however, had altered expression profiles. In light of these findings, the authors concluded that the response to nitrate stress and ionic (osmotic) shock could possibly be controlled by novel mechanisms in this organism. Since the study was not geared towards membrane proteins, there was no attempt to enrich for, or identify, membrane proteins specifically. Nevertheless, out of the 737 final identified proteins at least 16, and possibly several more, were potentially membrane proteins (Ref: text and supplementary tables from the paper). This is promising given the relative low abundance of membrane proteins as compared to cytoplasmic proteins in non-enriched preparations. The potential membrane proteins included an ATP synthase (F1 b subunits, which are membrane-bound), ABC transporters (perhaps important in hyperionic stress response), a sensory box histidine kinase, permeases and a quinone-interacting membrane-bound oxidoreductase.

Recently Choe *et al* demonstrated the use of 8-plex iTRAQ with a time course study of protein changes in the cerebrospinal fluid (CSF) of Alzheimer's disease patients undergoing intravenous immunoglobulin treatment [129]. The 8-plex iTRAQ reagents are based on the same chemistry as the 4-plex reagents, but utilize reporter ions with masses

of 113, 114, 115, 116, 117, 118, 119 and 121 Da. Ow *et al* utilized the 8-plex reagents to compare protein expression in heterocysts and vegetative cells of *Nostoc* PCC 7120 and simultaneously compared growth under N_2-fixing and non-fixing conditions [130]. The use of the 8-plex reagents allowed for the inclusion of biological replicates in the same experiment. They identified 506 proteins, of which 402 were quantified.

The power of iTRAQ for multiplexing studies was also particularly evident in another time-course experiment. Jagtap *et al* used iTRAQ to study the early events in *Bacillus anthracis* spore germination [131]. 4-plex iTRAQ was used to analyze spores immediately prior to induction of germination (T_0) and after induction of germination at 2 (T_2), 7 (T_7) and 17 (T_{17}) minutes. Spores were disrupted mechanically using zirconium beads and proteins were solubilized in 0.1% SDS after precipitation with cold acetone. Samples were reduced, alkylated, trypsin digested and tagged with iTRAQ reagents using standard protocols. The pooled labeled peptides were separated by SCX- and RP-HPLC and subsequently analyzed by MALDI-MS/MS using an ABI 4700 Proteomics Analyzer. Peptide identifications were made using the Mascot search engine [120]. From two biological replicates, 295 and 273 unique protein identifications were obtained from 3096 and 1879 peptides respectively with >90% confidence. When a more stringent, but realistic criteria of 2 unique peptides per identification was applied, the number of identified proteins from the two replicates was reduced to 261 and 203 respectively. 167 proteins were commonly identified using this criterion in both replicates. Of these, 39 proteins changed expression levels during the germination time course. 19 proteins showed decreased expression, 20 showed increased expression, and 115 proteins did not show a statistically significant change. Additionally 13 proteins had inconsistent

expression level changes across the data sets. Since metabolism is regarded to be stationary during early stages of germination and degradation of proteins is expected during this phase, the decreases in the levels of several proteins can be explained. However the increase in the levels of certain proteins is a little harder to account for. The authors have hypothesized that the increase in the levels of certain proteins could be due to improved access to already present proteins as the spore coat degrades, or due to the onset of fresh metabolism as the cell transitions to an active vegetative state or perhaps a combination of both. In addition to identifying several proteins and thereby increasing understanding of the germination process, the study also resulted in the identification of previously un-annotated gene products, which were also present in other *B. anthracis* strains or in other closely related *Bacillus* species. The study is important not only because it illustrates the applicability of iTRAQ for a multi-point time course experiment, but it also demonstrates the method's utility in investigating hard to analyze proteomic samples such as spore proteins.

These initial studies demonstrated that iTRAQ can provide robust quantitation. Wu *et al* compared the quantitative ability of iTRAQ to ICAT and 2D-DIGE using simple defined protein mixtures, as well as complex biological samples [132]. They demonstrated reasonable correlation between all three methods for the simple protein mixtures, while there were some discrepancies in the analysis of the complex biological samples. Nevertheless, of the three methods, iTRAQ showed the most sensitivity followed by ICAT and 2D-DIGE, which had similar performance to each other [132]. The lower variability seen with the iTRAQ data agree with the observations of Choe *et al* who compared iTRAQ to DIGE [133]. The other advantages demonstrated by iTRAQ

were a greater number of identified peptides, which leads to greated confidence in identifications, and no cysteine-labeling bias as seen with ICAT [132].

Microbial membrane iTRAQ studies

In their extensive review of bacterial proteomics from 2008, Poetsch and Wolters [7] reported that quantitative proteomics of bacterial systems was a nascent albeit rapidly growing field. Since that review there have been several quantitative shotgun proteomic analyses of microbial systems, some of which have focused on membrane proteins. Several of these quantitative bacterial membrane proteomic studies have utilized iTRAQ as their method of quantitation. In this section we review some of the recent studies that have demonstrate the utility of iTRAQ for analysis of microbial membrane proteins. An overview of these studies is provided in Table 1.2.

In their comprehensive gel-free and gel-based proteomics analysis of the model bacterium *Bacillus subtilis*, Wolff *et al* carried out an iTRAQ analysis on proteins differentially expressed in cytosolic fractions as a result of heat shock [18]. While iTRAQ was not carried out on proteins from the enriched membrane fraction, they were analyzed using a semi-gel approach. Membrane preparations were washed with salt and sodium carbonate and solubilized with 15% dodecyl maltoside (DM). The purity of membrane fractions was verified by immunoblots, which were used to check for the presence of marker proteins from the membranes and common cytoplasmic contaminants [134]. The authors had previously determined that 2-DE separations of membrane preparations resulted in non-satisfactory recovery of proteins with membrane spanning domains, possibly due to loss during IEF [134]. Therefore membrane fractions were analyzed by 1D SDS-PAGE followed by ESI-MS/MS. This semi-gel analysis of membrane fractions

resulted in 453 protein identifications, 265 of which were made with 2 or more peptide matches. 232 of the 453 identified proteins (~51%) had one or more predicted TMD. In a subsequent publication, the authors reported 268 membrane proteins [135]. Perhaps additional membrane proteins were discovered upon re-analysis of the data, or by a combination of data from additional experiments. Out of these, 134 proteins were predicted to have four or more transmembrane domains [135], 204 of the proteins identified were not seen in standard 2-DE-MS/MS or 2-D LC-MS/MS highlighting the importance of enrichment and special preparation of membrane fractions and also the use of distinct parallel proteomic analysis approaches.

The understanding of membrane protein dynamics is particularly important in pathogenic microbes, because proteins present on the surfaces directly interact with hosts and membrane proteins are primary targets for therapeutic intervention, vaccine development and as diagnostic targets [46, 82]. One of the earliest membrane focused iTRAQ studies on a microbe was carried out on antibiotic resistant strains of the common Gram-positive pathogen *Staphylococcus aureus* [100]. Multidrug-resistant strains of methicillin-resistant *S. aureus* (MRSA), with increasing resistance to glycopeptides, have been emerging around the world [136-138]. Some strains with high-level glycopeptide resistance to the drug vancomycin (VRSA) acquired through the *vanA* gene from *Enterococcus faecalis* (VRE) have been well studied [139]. However, less well understood are the distinct intermediate glycopeptide-resistant strains (GISA); the molecular basis of whose resistance remains unclear [140]. GISA strains are characterized by increased cell wall thickness, with 30 to 40 cross-linked layers of peptidoglycan present as against 20 layers in wild type strains [100]. It is highly likely

that no single genetic or biochemical change leads to the GISA phenotype. Increasing evidence points to multiple factors like cell wall synthesis and processing, autolysis, and regulatory events being involved.

In order to understand these strains better, Scherl *et al* carried out a comprehensive comparative proteomic and transcriptomic analysis of isogenic clinical GISA strains of *S. aureus* with variable susceptibility to vancomycin and teicoplanin [100].. The strains used for the study were a clinical isolate glycopeptide-sensitive strain (MRGR3), a GISA strain (14-4), and a sensitive revertant strain (14-4Rev). mRNA expression levels were determined using microarrays and quantitative polymerase chain reaction (qPCR) while standard 2-DE and IPG-IEF coupled with LC-MS/MS and iTRAQ were used for proteomic analyses. Special attention was paid to the analysis of membrane-enriched fractions because membrane proteins are often implicated in the development of antibiotic resistance.

For proteomic sample preparation, cells were washed in 1.1M sucrose buffer, and the lytic enzyme lysostaphin was used to digest the protective peptidoglycan layer yielding protoplasts that were disrupted by hypo-osmotic shock. Initial 2-DE experiments were carried out on insoluble membrane fractions to determine if differences could be discerned between the strains. IEF for the 2-DE analysis was carried out in a buffer containing a 50% concentration of the co-solvent TFE in place of the detergent [43]. The antibiotic resistant strains MRGR3 and 14-4Rev showed similar 2-DE patterns to each other while the sensitive strain 14-4 had a distinct pattern.

For iTRAQ analysis, membrane fractions were solubilized, reduced, alkylated, digested and labeled with iTRAQ reagents as per the standard manufacturer

recommended protocol. The iTRAQ labels were varied in the two replicates. The labeled peptides were concentrated, desalted, and subjected to IEF in a buffer containing 4M Urea and 50% TFE using IPGs. After IEF, slices of the IPGs were manually excised, and the peptides were extracted using trifluoroacetic acid (TFA) and acetonitrile (AcN). Additional separation of the fractions was carried out using a C_{18} RP column. Mass spectrometric analysis was carried out with ABI 4700 MALDI-TOF/TOF instrument in MS and MS/MS mode. Protein identifications were made using the Phenyx software platform [102, 141].

In the two separate iTRAQ experiments on membrane-enriched fractions, 3724 and 3719 unique peptides identifying 632 proteins and 754 proteins respectively with high confidence with at least 2 unique peptides per protein were detected. A total of 835 unique proteins providing approximately 32% genome coverage were identified, which was a significant improvement from the 23% proteome coverage obtained from all the previous attempts. 551 of the uniquely identified proteins were common to both experimental replicates. The average peptide coverage of the identified proteins was 4.9 peptides per protein. Approximately 20% of the identified proteins were predicted to be membrane proteins. This compared favorably with the theoretical membrane protein predictions from the genome where 637 of the 2575 ORFs (~24%) are predicted to be membrane proteins with at least one TMD. Relative quantifications were obtained on 835 proteins in comparisons between the sensitive MRGR3 strain and the resistant 14-4 strain. Similarly 826 quantifications were obtained in comparisons of the resistant 14-4 strain and the sensitive 14-4Rev strain, while 826 quantifications were obtained in the comparison between sensitive strains MRGR3 and 14-4Rev. Because the coefficient of

variation (CV) of quantification between individual peptides of the same protein was relatively high, stringent criteria were used for the cutoff. 178 unique proteins, corresponding to approximately 4% of the genome, were considered to be differentially expressed in the GISA strain. 155 proteins were differentially expressed between strains 14-4 and MRGR3, and 110 between strains 14-4 and 14-4Rev. 65% of the 178 unique proteins were common to the two comparisons.

The combined proteomic and transcriptomic study identified several targets potentially involved in glycopeptide resistance mechanisms, including proteins associated with cell wall synthesis, imipenem resistance, signal transduction, purine metabolism and ABC transporters. Functional classification of the differentially expressed proteins using the cluster of orthologous groups (COG) system of classification [142]placed them into energy metabolism, amino-acids transport, cell envelope biosynthesis, protein turnover and inorganic ion transport categories. Many of the proteins that were differentially expressed in a similar manner between the resistant strain and the two sensitive strains showed correlation between mRNA expression and protein expression patterns. Interestingly, however, several proteins belonging to the last two functional categories, protein turnover and inorganic transport, showed divergent expression patterns at the mRNA and protein level, indicating the possibility of extensive post-translational regulation in those categories. Overall, even though there was a good correlation between proteomic and transcriptomic data, the divergence seen in expression patterns of certain protein categories is in agreement with the observations of Gygi et al [6] and validates the need for simultaneous proteomic and transcriptomic expression studies.

Radosevich *et al* used an iTRAQ based proteomic approach to investigate pathogenicity in *Mycobacterium avium* subspecies paratuberculosis [143]. The organism is the causative agent of Johne's disease, a debilitating and fatal condition in cattle and sheep [144]. Extensive microarray data was available for this organism [145] and 25% of the proteome had been previously identified [146]. Proteomic analysis of membrane fractions of these Gram-positive organisms are further complicated by the presence of particularly resilient cell walls containing mycolic acids. Radosevich *et al* compared two strains of the bacterium – 'K-10' a laboratory-adapted strain that has undergone several serial passages, and '187' a fresh clinical isolate from an infected animal. Preliminary studies on these strains showed differences in growth rates as well as differential banding patterns of both membrane and soluble fractions on silver stained 1D SDS-PAGE gels indicating phenotypic and biochemical variations between the strains. Membrane fractions were prepared by retaining the insoluble pellets after ultracentrifugation of sonicated cultures, while the supernatant was processed as the soluble fraction; no additional preparation was carried out. iTRAQ labeling was carried out by the standard manufacturer's recommended protocol using the $iTRAQ_{114}$ label for strain '187' and the $iTRAQ_{117}$ label for strain 'K-10'. MS/MS analysis was carried out with a Q-TOF Ultima API mass spectrometer following separation by SCX and RP-HPLC. Protein identifications were made using the Mascot software [120]. Protein expression data were obtained for 550 proteins in the membrane fractions, 385 of which were uniquely identified in the membrane fractions and not in the cytosolic fractions. Of these proteins, 266 were deemed hypothetical proteins by Swiss Prot. Out of the predicted membrane proteins in this fraction, 37 were up-regulated in the 'K-10' strain while 35 were up-

regulated in strain '187'. From the cytosolic fractions, 487 proteins were identified, 324 of which were unique to this fraction. 185 of these were predicted to be hypothetical proteins by Swiss Prot. 22 were up-regulated in strain 'K-10' while 18 were up-regulated in strain '187'. In all, 874 proteins were identified and quantified from the membrane and cytoplasmic fractions, 165 of which were common to both fractions. 111 proteins showed significant changes in expression levels between the two strains while 763 showed no significant changes. All identifications were within 95% CI and with at least 2 peptides identities per protein in addition to quantitation data for both the iTRAQ labels. Many of the results were confirmed by immunoblot. These differentially expressed proteins could be candidates for understanding pathogenesis, as well as for diagnostic targets and vaccine targets.

Another mycobacterial pathogen subjected to an iTRAQ based membrane proteomic analysis is *Mycobacterium ulcerans* [49]. ***M. ulcerans***, a Gram-positive bacterium, is the causative agent of Buruli ulcer, a devastating human necrotic skin disease [147]. Skin damage is caused due to the release of mycolactone, a cytotoxic and immunosuppressive macrocyclic polyketide [147]. Strains deficient in mycolactone production are also unable to colonize the salivary glands of *Naucoris cimicoides*, a carnivorous water bug that is believed to be a host and reservoir for ***M. ulcerans*** [148]. Hence there is a potential correlation between mycolactone production and pathogenicity. Tafelmeyer *et al* followed a similar approach to Wolff et al in their analysis *Bacillus subtilis*, to investigate these mycolactone deficient strains [18]. They utilized a combination of gel based and gel free proteomic techniques to identify 1074 unique proteins, which corresponds to ~25% of predicted ORFs in the genome. Membrane

proteins were identified either by a combination of 1-D-SDS-PAGE-LC or 2D-LC followed by ESI-MS/MS. 481 proteins were identified from the membrane fractions, 240 which were unique to these fractions. The utility of this approach for investigation of highly hydrophobic membrane proteins was demonstrated by the fact that 20 of the identified proteins had 10 or more predicted TMDs with one protein having 15 predicted TMDs. Differentially expressed proteins belonged to information pathways, lipid metabolism and stress response. An interesting observation from this study was that the carbonate wash step, which has been employed to such good effect in several proteomic studies, appeared to have no effect on the membranes of *M. ulcerans* in the hands of these investigators. The overall experimental results also demonstrate once again that no single method can provide complete proteomic coverage, they underscore the importance of combining distinct proteomic separations for comprehensive coverage of the proteome.

Mammary pathogenic *E. coli* (MPEC) strains are capable of infecting and growing in mammary glands [149]. Lippolis *et al* used an iTRAQ approach to elucidate the mechanisms that enable the growth of this organism in milk, a medium that does not favor bacterial growth, with a hope that it could shed some light on the mechanism of their pathogenesis [150]. Proteomic comparisons of cells grown in standard Luria Bertani (LB) growth medium or fresh whole bovine milk were carried out. MPEC cells grown in milk were purified using sucrose gradients to remove contaminants from the growth medium. Cells were disrupted by sonication and separated to give soluble and membrane fractions. Soluble fractions and membrane fractions were tryptically digested, reduced, alkylated, labeled with iTRAQ reagents using the manufacturer's recommended protocol, and samples from the fractions were combined. The pooled labeled peptides were

separated by SCX and analyzed by RP-MS/MS on a Waters Q-TOF Ultima API mass spectrometer. A total of 1000 proteins were identified using the Mascot algorithm [120], 633 of these with high confidence. 336 of these (256 with high confidence) were from the membrane fraction, 356 (244 with high confidence) from the soluble fraction and 308 (133 with high confidence) were identified in both fractions. Approximately 20% of identified proteins were up-regulated in cells grown in milk, while 10% down-regulated. Not surprisingly, amongst the up-regulated proteins were several proteins involved in galactose metabolism, including beta galactosidase (4 fold overexpression) and UDP-glucose 4-epimerase (5 fold overexpression). The authors hypothesized that these could be associated with altered cell wall lipopolysaccharides and could play a role in pathogenesis. Many other proteins of interest were also up-regulated including several outer membrane siderophore receptors, which were likely up-regulated to counter effect of the iron sequestering protein lactoferrin present in milk, and the protein LuxS, which is involved in bacterial quorum sensing and also associated with several virulence genes [150]. In contrast, several structural flagellar genes were down-regulated. The authors suggested that this could be to evade host immune defenses [150].

Leptospira interrogans is the causative agent of the zoonosis leptospirosis, which can be fatal [151]. Lo *et al* were interested in global protein expression changes on a temperature upshift from 30 °C to 37 °C, which reflects conditions associated with pathogenesis [152]. The investigation focused on leptospiral outer membrane proteins, which are known to play a role in pathogenesis. Earlier microarray experiments from the group that had revealed divergent expression patterns between OMP mRNA and protein expression levels [153], provided an additional motivation to analyze the OM proteome.

L. interrogans OMPs were extracted using a standard leptospiral OMP extraction protocol [154, 155] with the phase partitioning detergent Triton X-114. A methanol/chloroform extraction was utilized for detergent removal and OMP enrichment. The enrichment of OMPs and presence of minimal contamination by IMPs and cytoplasmic proteins was confirmed by immunoblots against representative proteins from the 3 cellular compartments. Dissolved protein samples were reduced, alkylated, tryptically digested and labeled with iTRAQ reagents according to the manufacturer's suggested protocol. Peptides from the 30°C samples were labeled with $iTRAQ_{114}$ and $iTRAQ_{115}$ labels, while peptides from the 37°C upshift samples were labeled with $iTRAQ_{116}$ and $iTRAQ_{117}$ labels. The labeled samples were pooled, and peptides were separated and analyzed by SCX followed by RP-MS/MS on an ABI QSTAR XL mass spectrometer. 1026 proteins, representing 28.4% of the predicted proteome, were identified at 99% CI using the Paragon software platform (ABI) [156]. Of the identified proteins, 22 were predicted to be OMPs, 58 lipoproteins, 38 IMPs, and 754 were predicted to be cytoplasmic proteins. The OMPs and lipoproteins were underrepresented, with only 26% of predicted OMPs and 34% of predicted lipoproteins from the genome identified in the study. A possible explanation for the low yield of OMPs was that not all OMPs were expressed under the conditions tested. It has been suggested that the Triton X-114 method is not an efficient method for the recovery of OMPs, however, the authors noted that they had previously obtained nearly 87% OMP recovery in *Campylobacter jejuni* using a similar procedure [157]**.** The authors posited that either insufficient tryptic peptides were generated during digestion or the size of the digested peptides was too large for MS analysis. On the other hand, the large number of IMPs and cytoplasmic

proteins identified, in spite of the confirmed purity of the OMP preparations by immunoblot, demonstrates the high sensitivity of the iTRAQ approach. Proteins with 1.5 fold difference in quantification between the two conditions with at least 95% CI were considered differentially expressed. Using this criterion, 27 proteins were deemed as up-regulated after the 37 °C upshift, while 66 were deemed down-regulated. Comparison with previous transcriptomic studies showed that there was some correlation between protein and gene expression levels [153]. However, a subset of proteins showed a change in protein expression level, even though no change was seen in their mRNA expression levels. These proteins constitute strong candidates for post-translational regulation. As an explanation for the proteins down-regulated in response to temperature shift, the authors suggested that it could be a pathogenic response to evade host defense systems.

Leptospira interrogans has also been the subject of another global proteomic study. Eshghi *et al* used 2-DE and iTRAQ to monitor the changes in protein expression in response to iron limitation and presence of serum, which mimics *in vivo* conditions [158]. 563 proteins were identified in the study, 65 of which showed altered expression patterns under the compared conditions. Many of these proteins are potentially involved in the infection process. Even though no special effort was made to analyze membrane proteins, several membrane proteins were identified, like the OMP TolC were identified in the study.

Acinetobacter is one of the leading causes of hospital infections due to its ability to rapidly acquire resistance to commonly used antibiotics [159]. Prior proteomics studies have shows altered membrane protein expression patterns, and quantitative protein expression changes, in multidrug resistant and colistin resistant strains [160-163] . Soares

et al monitored the growth of *Acinetobacter baumannii* cultures at various stages of their growth cycle to investigate the effect of oxidative stress and nitrosative stress [164]. The authors demonstrated that reactive oxygen species (ROS) and reactive nitrogen intermediates (RNI) accumulated during growth. *Acinetobacter* cells were harvested at exponential, early stationary and late stationary stages of their growth cycles. Membrane fractions were obtained by ultracentrifugation and enriched with a carbonate wash. Samples for iTRAQ were processed by the manufacturer's suggested protocol. Pooled labeled peptides were separated by SCX and RP and analyzed on a QSTAR Elite mass spectrometer. Identifications were made with the Mascot algorithm [120]. A combination of 2-DE and iTRAQ resulted in the identification of 107 differentially expressed proteins during the growth cycle. 76 proteins of these were identified by 2-DE, comprising of 13 membrane proteins and 63 cytosolic proteins, while 31 differentially expressed proteins were identified by iTRAQ. There was some overlap between the proteins identified by both approaches and for those, a similar pattern of protein expression change was noted. This validates both the techniques, but also re-enforces the importance of using complementary proteomic techniques for complete proteome coverage. Amongst the identified proteins were proteins involved in signaling, potential virulence factors, and general stress response. Once again, the authors reported poor correlation between protein expression patterns and mRNA expression patterns as measured by qPCR.

The parasitic protozoan *Plasmodium falciparum* is the causative agent of some of the most serious forms of malaria in humans, accounting for nearly one million deaths per year and nearly a quarter of a billion cases of malaria [165]. The emergence of strains resistant to traditional anti-malarial compounds has led to a renewed search for new anti-

malarial drugs as well as evaluation of anti-malarial activity of existing drugs used for other conditions. One of the current recommended treatments is a daily dose of the antibiotic doxycycline (DOX), a tetracycline derivative, often used in combination with quinine or artemesinin derivatives [166, 167]. The mode of action of doxycycline against bacteria is well documented - it binds to several proteins in the 30S ribosomal subunit, and RNA molecules in the 16S ribosomal subunit preventing binding of the charged tRNA molecules to the acceptor site on the ribosome [168]. However, the mode of action of DOX against *Plasmodium falciparum* is still not clear at the molecular level. Recent research points to the fact that this family of antibiotics probably targets the plasmodial organelles - mitochondria and plastids [169-171].

In order to further understand the mode of action of DOX on *P. falciparum* Briolant *et al* carried out comparative quantitative proteomic studies of changes at the schizont stage of the parasite's lifecycle in samples that were exposed or not-exposed to DOX using two complementary proteomic techniques, 2D-DIGE and iTRAQ [172]. Chloroquinine resistant clones of *P. falciparum* were maintained in continuous cultures. At the ring stage of the parasitic cycle, the cultures were either exposed or not exposed to 10μM doxycycline, which was previously determined to be the IC_{50}, for 24 hours. This was followed by a chase because continuous exposure to DOX would have be 100% lethal. RBCs were lysed with 0.1% saponin to release free parasites, which were washed with PBS and disrupted by ultrasonication. Soluble proteins were obtained from the supernatant, and the membrane protein pellet was solubilized in 4% CHAPS. All samples were acetone precipitated to remove lipids. Four biological replicates were processed by 2D-DIGE, while three biological replicates were processed by iTRAQ. For iTRAQ,

samples were treated as per the manufacturer's instructions. From the soluble fractions, the control sample was labeled with the $iTRAQ_{114}$ label while the DOX-treated sample was labeled with the $iTRAQ_{117}$ label, while from the membrane fractions, the control and DOX-treated samples were labeled with the $iTRAQ_{115}$ and $iTRAQ_{116}$ labels respectively. Pooled peptide mixtures were separated by SCX and RP and analyzed on a Waters Q-TOF Ultima system. Protein identifications were made with the Mascot algorithm [120].

Analysis of changing spots on the 2D-DIGE platform resulted in the identification of 32 distinct proteins using a combination of soluble fractions run on 18cm pI 3-10 broad-range IPG strips, and membrane fractions run on pI 4-7 and pI 6-11 narrow-range IPG strips. Spots with intensity ratios <0.74 or >1.35 between the conditions were considered as differentially regulated. Based on this, 22 proteins were classified as up-regulated and 10 as down-regulated The PlasmoDB database localized 2 of these to the membrane - the plasmepsin 1 precursor protein involved in hemoglobin catabolism was up-regulated, while 1 unknown plasmodial protein was down-regulated. iTRAQ analysis of the soluble fractions resulted in the identification 422 unique proteins. 246 of these were plasmodial proteins and 176 were human proteins. 22 proteins showed significant changes in expression levels; 18 were up-regulated, and 4 were down-regulated. iTRAQ analysis of the membrane fractions yielded 308 unique protein identifications; 204 were plasmodial proteins and 104 were human proteins. 18 proteins showed significant changes in expression; 14 were up-regulated and 4 were down-regulated. 6 of these were confirmed as membrane proteins, 5 of which were up-regulated, 1 was down-regulated. Of the 32 identified with altered regulation by DIGE, 6 (19%) were also identified in the iTRAQ analysis. Most proteins appeared to have similarly altered expression profiles by

both methods. A combination of results from both methods revealed 64 plasmodial proteins to be differentially regulated. Of these 14 were cytoplasmic (13 by iTRAQ and 1 by DIGE), 12 were localized to the apicoplasts (10 by iTRAQ and 2 by DIGE), 8 to the plasma membrane (6 by iTRAQ and 2 by DIGE), 8 to the nucleus (2 by iTRAQ and 6 by DIGE), 2 mitochondrion (both by DIGE), while 20 had unknown localization (16 by iTRAQ and 4 by DIGE). The down-regulated proteins largely belonged to the protein synthesis and transport functional category, while the up-regulated proteins were involved in protein metabolism and anti-oxidant response metabolism. Analysis of the differentially expressed proteins provided evidence that apicoplasts and mitochondria are targets for DOX action in *Plasmodium*. The expression patterns for three apicoplast genes were tested by qPCR and the results showed correlation with the proteomic results. The greater number of proteins identified by iTRAQ in both fractions demonstrates the power of the method. Nevertheless, only 19% protein identifications were common to both approaches, showing once again that multiple methods are necessary for complete proteome coverage.

Tannerella forsythia is a Gram negative bacterium implicated in periodontitis [173]. During pathogenesis, these bacteria exist as part of a biofilm attached to tooth surfaces i.e. as a subgingival plaque [173]. In many organisms, cells that are adapted to a biofilm lifestyle show altered metabolism, including enhanced drug resistance, when compared with their free-living equivalents [174]. Pham et al used iTRAQ to compare protein expression in biofilm and planktonic *T. forsythia* cultures [175]. Cells were lysed and solubilized in a 2-DE like solubilization buffer containing urea, thiourea, and CHAPS by freeze thawing in liquid nitrogen (LN_2). Proteins were precipitated using acetone,

TCA and DTT. Reconstituted proteins were reduced, alkylated, digested and labeled with the iTRAQ reagents according to the manufacturers protocol. The labeled peptides were combined, vacuum concentrated, and separated and analyzed by SCX and RP-HPLC-MS using an ABI QStar XL Hybrid ESI Quadrupole Time-Of-Flight tandem mass spectrometer. Protein identifications were made using the Phenyx algorithm [102, 141]. 348 proteins were identified and quantified, 44 of which were found to be differentially expressed. Several outer membrane proteins were up-regulated in the biofilm cells including transport system proteins, S-layer proteins, and Ton-B receptors, several of which are potentially involved with iron transport and in the transport of complex carbohydrates like starch [176, 177]. Additionally the authors showed that biofilm cells were 10-20 times more resistant to oxidative, and identified several up-regulated proteins putatively involved in oxidative stress response. They suggested this as a possible mechanism for the organism's survival in the oral cavity.

Cyanobacteria, which are important models for photosynthesis in higher organisms [178, 179], have been the subjects of several high quality quantitative proteomic studies. Many of these have been carried out by investigators from the prolific group of Philip Wright, at the University of Sheffield, UK. Stensjö *et al* used iTRAQ to measure differential protein expression in the oxygen-evolving phototrophic nitrogen-fixing filamentous cyanobacterium, *Nostoc* species PCC 7120, under nitrogen fixing and non-fixing conditions to shed light on how nitrogen-fixing filamentous cyanobacteria alter their metabolic pathways to enable hydrogen production [180]. Protein expression patterns of cells grown either in the presence (nitrogen fixing), or in the absence (non-fixing condition), of ammonium chloride (NH_4Cl) in the medium were compared.

Additionally, to study the effect of hydrogen (H_2) supplementation, growth in 9% H_2 bubbled media under both the above conditions was also compared. As is common with iTRAQ studies, the experiment was independently carried out two times to evaluate the impact of experimental and biological variations on the study. Cells were disrupted using glass beads, and proteins were obtained by precipitating the cell lysates with cold acetone. Proteins were reduced, alkylated, tryptically digested and labeled with the iTRAQ reagents using standard protocols. The pooled labeled peptides were separated and analyzed by SCX followed by RP in line with a QSTAR XL tandem ESI-MS instrument.

The authors identified 486 unique proteins in the study. Of these, 313 in were identified in experiment set 1, and 330 in experiment set 2 with > 95% CI. Using a criteria for proteins with >1.8 fold change being considered as up-regulated and proteins <0.6 fold change being considered as down-regulated, approximately 30% of the identified proteins were classified as having altered expression patterns under nitrogen-depletion conditions. Of the 122 differentially regulated proteins, 94 were identified from experiment 1, and 95 from experiment 2. Of the 94 proteins with altered expression in experiment 1, 80 were up-regulated, while 56 were up-regulated in experiment 2. Addition of hydrogen to the actively nitrogen-fixing cells resulted in only 5% of identified proteins showing significant expression change.

While the authors were able to identify a significant number of membrane proteins from the photosystem I and II (PSI, PSII) membrane complexes and other pathways, they considered the total number of hydrophobic proteins identified as underrepresented; PSORTb [84, 85] predicted 8.9% of the identified proteins to be

membrane proteins, while for the genome complement of 5366 chromosomal ORFs 20% were predicted to be membrane localized. Supplementary analysis by the SOSUI membrane prediction algorithm [181, 182] and addition of membrane associated proteins brought the percentage of total predicted membrane and membrane-associated proteins up to 18.9%. However applying the same logic to the predicted ORFs in the genome, one would also expect to see a proportional increase in predicted membrane and membrane-associated proteins, so it is unclear if this additional analysis provides a meaningful increase in the membrane coverage of the study. It should be noted that the underrepresentation of membrane proteins in this study is not surprising because no special attempt were made to enrich the membrane fractions. In light of this, the recovery of membrane proteins is very promising.

Prochlorococcus marinus MED4, which was first cultured in 1988, is an extremely small oxygenic phototrophic cyanobacterium that is capable of thriving under extremely adverse conditions such as desiccation, hypersalinity, high temperatures and extremes of pH [183]. *Prochlorococcus* was first observed in nutrient-poor regions of the ocean that were, at the time, thought to be free of microorganisms [183]. The organism is now believed to be one of the most abundant photosynthetic organisms on the planet, significantly contributing to biogeochemical cycling and climate control [184].

Pandhal *et al* carried out an iTRAQ based study to investigate the response to varying light levels mimicking natural oceanic conditions [185]. They also compared the protein expression patterns with those of *Synechocystis* and *Anabaena*, which had previously been investigated in their research group [90, 98]. In each of two series of experiments, *Prochlorococcus* cultures were grown under varying intensities of light –

low intensity (20 μEinstein m^{-2} s^{-1}), medium intensity (60 μEinstein m^{-2} s^{-1}) and high intensity (100 μEinstein m^{-2} s^{-1}) and compared using standard iTRAQ methodology. Cells were washed in a sucrose buffer for removal of salts and polysaccharides, and disrupted, in a buffer containing 9M urea and 1% of the detergent CHAPS, through a combination of mechanical cracking and liquid nitrogen. Proteins were extracted by TCA/acetone precipitation, and then reduced, alkylated, digested, and labeled with iTRAQ reagents according to the manufacturer's protocol. Pooled peptides from both conditions were mixed and then separated and analyzed by SCX and RP coupled with a QSTAR XL tandem ESI mass spectrometer. The study identified a total of 184 unique proteins, which is 11% of the total predicted coding regions for the organism. 53 of these uniquely identified proteins were common to both experiments, while 94 proteins were unique to experiment 1 and 37 unique to experiment 2. The distribution of the identified proteins was 73.7% acidic and 26.3% basic in contrast to the predicted distribution, which was 60% basic. The authors, however, were unable to explain why their results were biased to the acidic side. Using PSORTb [84, 85], 24 proteins (12.8%) were identified as localizing to the membranes. 21 of these were predicted to be IMPs and 3 were predicted to be OMPs. This is higher, both in terms of actual number of proteins identified as well as a percentage of total proteins identified, when compared with the studies on *Synechocystis* (2.8%) [98] and *Anabaena* (7.7%) [90] using non-iTRAQ methods. Additional analysis using the LipoP algorithm [186] to predict specific signal sequences in lipoproteins, further increased the total number of identified proteins predicted to localize to the membranes to 35. The exact cause of improved membrane protein recovery *vis a vis* the earlier cyanobacterial studies was not clear, though the

authors surmised that it could be ascribed to the use of a high concentration of urea in the protein solubilization step.

We have previously demonstrated that cross-species PMF between closely related species is possible [48]. Pandhal *et al* attempted a cross species proteomic analysis between closely related cyanobacteria by comparing *Euhalothece* with the genetically and morphologically similar, *Synechocystis* spp PCC6803 [179]. Proteomic analyses were carried out using a combination on ^{15}N metabolic labeling and iTRAQ. Since the *Euhalothece* genome sequence was not available, the *Synechocystis* genome database was used for protein identification. *Euhalothece* spp BAA001 is an extremely halotolerant cyanobacterium isolated from a lake in the heart of the Sahara which grows optimally at 3% salt concentration, but can tolerate concentrations from up to 12% - 15% [179]. *Synechocystis*, whose genome has been sequenced, grows optimally at 0% salt concentration, though it can tolerate up to 7% salt [187]. The approach used in the study is based on a premise, previously demonstrated in the plant *Arabidopsis*, where increased salt tolerance is generated by differential expression and regulation of shared components rather than through novel mechanisms [188].

For the iTRAQ analysis *Euhalothece* and *Synechocystis* cells grown at 6% salt were disrupted by mechanical cracking in liquid nitrogen and proteins were precipitated using TCA and acetone. Samples were reduced, alkylated, tryptically digested and labeled with iTRAQ reagents as recommended by the manufacturer. Biological duplicates of the 6% salt *Synechocystis* culture were labeled with the $iTRAQ_{115}$ and $iTRAQ_{116}$ labels while the 6% salt *Euhalothece* sample was labeled with the $iTRAQ_{117}$ label. Pooled samples were separated and analyzed by SCX-RP-MS on a QStar XL

Hybrid ESI Q-TOF-MS/MS instrument. Protein identifications were made using the Paragon algorithm [156]. 207 unique proteins were identified by iTRAQ at 95% CI, 39 of these were predicted to be membrane proteins by PSORTb [84, 85] and LipoP [186], and 21 proteins were deemed to be differentially regulated using a 2 fold change in expression as a cutoff. As a result of this study, the authors were able to demonstrate the applicability of iTRAQ to cross-species proteomics. Furthermore, they demonstrated differences in cell behavior under changing salt conditions, and showed differences in protein abundance levels across species for proteins orthologous to both.

Recently Rowland *et al* carried out an iTRAQ based proteomic analysis on the thylakoid membranes of *Synechocystis* sp PCC6803 comparing cells grown at 25°C and 38°C [189]. Growth at 38 °C confers increased thermotolerance to the highly heat sensitive components of the *Synechocystis* photosystem II (PSII) photosynthetic apparatus [189]. 385 distinct proteins were identified in the study, 203 of which were common to 3 biological replicates. 168 of these were identified with 2 or more peptides. 48 proteins were found to be differentially regulated under the compared conditions. Of these 15 were up-regulated and 33 were down-regulated. The up-regulated proteins included several proteins associated with electron transport, while many of the down-regulated ones are hypothetical proteins whose functions remain to be deduced. The authors concluded that the enhanced thermotolerance at 38°C was due to the additional activity of several proteins as 38°C.

Archaeal systems have also been the subjects of iTRAQ based proteomic studies. In an attempt to create a comprehensive proteomic database called Peptide Atlas for the extremely halophilic archaeon *Halobacterium salinarum* NRC-1, Van *et al* used a

multitude of proteomic technologies including ICAT, iTRAQ, immunoprecipitation (IP) and mass spectrometry to analyze soluble and membrane fractions from the organism [190]. Over 630,000 MS spectra were generated in 497 individual runs from 88 sets of proteomic experiments. The authors identified 1646 proteins representing 63% of the predicted genome. Aided by fractionation and subsequent detergent solubilization, they were able to identify 188 out of the 550 proteins predicted to have transmembrane domains. Despite enrichment of membrane fractions only 34% of the predicted membrane proteins were identified as against 70% of soluble proteins (1458 out of 2077 predicted soluble proteins). Furthermore, analysis of the data showed a bias against the detection of basic proteins, high molecular weight proteins and peptides at both extremes of the hydrophobicity scale. However, considering that the data were generated using a multitude of techniques, it is not immediately clear as to how each analytical method contributed to these biases. Nevertheless, the authors noted that iTRAQ could introduce a significant bias in peptide detection by MS. The exact basis of this bias remains unclear, however, in light of this observation, they suggested that an empirical strategy was best suited to selecting proteotypic peptides for quantitation. Interestingly, they also demonstrated that for this system there was good correlation between mRNA expression patterns and protein expression patterns.

Williams *et al* carried out a couple of extensive iTRAQ based proteomic analysis of the secreted, soluble and insoluble proteins fractions of *Methanococcoides burtonii*, a psychrophilic methanogenic archaeon adapted to grow at temperatures of 1°C – 2°C at the bottom of Ace Lake in Antarctica [191, 192]. Methylamines and methanol are the only known carbon and energy substrates for *M. burtonii*. In two parallel studies they

compared the protein expression patterns of cultures growing at 4°C and 23°C [192] as well as cultures growing in media containing trimethylamine or methanol as carbon sources [191]. Because genome studies on the organism had revealed the heightened role of genes involved in membrane and cell envelope biogenesis and genes encoding integral membrane proteins in its ability to survive extreme conditions, special effort was made to proteomically analyze this group of proteins. Secreted proteins were enriched in the supernatant fractions by collecting the filtrate of the culture. Soluble protein fractions were collected after ultrasonic disruption, and filtration of the supernatant. Insoluble membrane fractions were prepared by washing the insoluble pellet with carbonate to remove loosely associated soluble proteins, heat denaturing at 90°C for 2 minutes, and then reconstituting the pellet in 60% methanol, based on the method of Blonder *et al* [76]. Samples were reduced, alkylated, digested, and labeled with the iTRAQ reagents as per the manufacturer's protocol. Combined peptides were separated and analyzed by SCX and RP coupled to an ABI API QStar Pulsar i hybrid tandem mass spectrometer. Protein identifications were made with the Mascot algorithm [120]. A total of 698 unique proteins were identified from all the fractions out of the 2431 predicted ORFs in the genome (24%). Of the 698 identified proteins, 362 were unique to single fractions. 193 were identified in the insoluble fraction, 153 in the soluble fraction and 16 in the supernatant. From the first set of experiments, 166 of the identified proteins were found to have significantly altered expression patterns of at least 1.5 fold change, after a change in growth temperature [192]. 79 proteins were up-regulated in one or more fractions at 4°C, while 68 proteins were up-regulated in one or more fractions at 23°C. 19 proteins were found to be variably regulated in the different sample fractions. 62 of differentially

regulated proteins had been seen in previous non-iTRAQ studies [193-196]. The increased abundance of several surface layer protein, a likely adaptation for growth at 4°C, suggested an extensive remodeling of the cell envelope. Other proteins overexpressed at 4°C included several proteins potentially linked to countering the cold conditions including potential RNA chaperones, other proteins involved with protein folding and translation. On the other hand, proteins overexpressed at 23°C included several oxidative stress proteins and membrane proteins of unknown functions. In the second set of experiments, the authors showed that growth in methanol appeared to result in oxidative stress in the cells with the overexpression of universal stress proteins and nucleic acid binding proteins, and caused a change in cell envelope proteins to counter the solubilizing effect of methanol [191].

Sulfolobus solfataricus P2, a thermophilic archaeon isolated from sulfur-rich hot springs in Naples, Italy, grows optimally at 80°C and at pH 3-4 [197, 198]. Researchers in the Wright group compared protein expression in cells grown at optimal temperature with cells grown at reduced temperatures of 70°C and 65°C using iTRAQ [199, 200]. Late exponential phase cell cultures were used to generate soluble and membrane fractions, which were analyzed separately. The investigators had a special interest in membrane proteins, and therefore, three methods were used for preparation of membrane fractions. These comprised of (i) the standard ABI recommended sample preparation methodology, (ii) a sample preparation scheme augmented with additional digestion with trypsin and chymotrypsin, and (iii) a sample preparation scheme even further augmented to include delipidation with sodium deoxycholate (SDC), based on the method of Masuda *et al* [62]. Samples grown at different temperatures were processed by each of these

methods, and then reduced, alkylated and iTRAQ labeled using the manufacturer's recommended method. Pooled labeled peptides were separated and analyzed by SCX and RP coupled to an ABI QStar XL Hybrid ESI Quadrupole time-of-flight tandem mass spectrometer. Protein identifications were made with the Phenyx algorithm [102, 141]. By combining data from 3 iTRAQ experiments representing the 3 sample preparation conditions, 395 unique proteins were identified with 2 or more peptides, 373 of these were predicted to be membrane proteins. This represents nearly 45% of the 833 predicted membrane proteins from the genome. Approximately 20% of the identified proteins were shown to have >1.5 fold altered expression patterns after lowering of the growth temperature. From a methodology standpoint, SDC delipidation-based sample preparation provided the largest number of identifications from a single experiment with 284 proteins detected, 246 of which were membrane proteins. This improvements provided by the modified protocol are highlighted by the results from the standard suggested iTRAQ protocol in which only 147 proteins, 133 of which were predicted to be membrane proteins, were detected. The authors have suggested the use of the technique for analysis of other archaeal membrane proteomes.

iTRAQ-based proteomic analyses have also been carried out on industrially and biotechnologically important organisms. *Streptomyces* are Gram-positive bacteria, which are medically and industrially important for antibiotic and antitumor agent production [9, 201, 202]. Due to their complex life cycles, they also serve as valuable model for development in multicellular organisms. In order to better understand growth and development, Manteca *et al* carried out 2 iTRAQ-based proteomic studies in *Streptomyces coelicolor* [201, 202]. In the first study, iTRAQ, on cytoplasmic, membrane

and membrane extrinsic fractions, was used to follow a *Streptomyces coelicolor* culture growth time course [202]. All experiments were carried out in duplicate. Mycelia were scraped of solid cultures at 12, 24 and 72 hrs. Mycelia were mechanically disrupted, and cells were lysed by sonication. Ultracentrifugation was used to separate soluble and membrane fractions, which were subjected to a carbonate wash. Membranes were boiled with SDS-gel-loading buffer, and resolved by SDS-PAGE. The gels were manually cut into slices, and reduction, alkylation and tryptic digestion was carried out in-gel. Peptides were extracted using formic acid. Reconstituted peptides were labeled with $iTRAQ_{113}$, $iTRAQ_{114}$ and $iTRAQ_{115}$ labels (12, 24 and 72 hrs, respectively). Labeled peptides were analyzed using RP-MS/MS on a Waters Q-TOF tandem mass spectrometer, and protein identifications were made with the Mascot search engine [120]. A total of 626 proteins were identified, representing 8% of predicted proteome. 361 identifications were common between both biological replicates, of which 345 were quantitated.

Out of the proteins common to both biological replicates, 107 proteins were identified in the membrane fractions. Several of these were also seen in the other two fractions. Of these, approximately 80 proteins had 1 or more predicted TMDs and nearly 30 proteins had predicted signal peptides. Amongst the proteins from the soluble fractions, approximately 50 proteins had predicted TMDs and 10 had predicted signal peptides.

In the second study, the authors compared protein expression patterns between non-sporulating liquid cultures and solid cultures [201]. It was previously assumed that no differentiation took place in liquid cultures, however, it has recently been demonstrated that compartmentalization occurs in liquid cultures as well. 8-plex iTRAQ

was used to compare protein expression in liquid cultures at 14 hrs and 90 hrs and solid cultures at 12 hrs and 72 hrs [129, 201, 202]. 642 proteins were identified in this study with a very similar pattern to the first study [201, 202]. Analysis of the data from both studies revealed that as cells progressed from the initial compartmentalized mycelial stage to the multinucleated hyphal stage, there was a change protein expression from primary metabolism to secondary metabolism. In the second study, the bulk of the identified proteins showed similar expression patterns between solid and liquid cultures [201]. Only ~17% proteins showed a significant change in abundance, some of which were involved with hyphal compartmentalization and spore formation.

In a report by Steen *et al*, an iTRAQ analysis was carried out on a *Lactococcus lactis* system used as an expression system for the production of human cystic fibrosis transmembrane regulator (CFTR), mutations in which lead to the disease cystic fibrosis [203]. The *L. lactis* expression system is a preferred eukaryotic protein expression system over *E. coli* for multiple reasons, including a slower doubling rate, presence of chaperones that potentially aid the folding of expressed proteins, and favorable cytoplasmic and membrane environment for foreign proteins [204]. In their analysis, however, Steen *et al* noted problems like growth arrest and low yields of CFTR.

To investigate potential causes for this, they used 4-plex and 8-plex iTRAQ reagents to analyze membrane and soluble fractions. CFTR expressing cells and control cells were compared as a function of pre- and post- induction time at 0, 1 and 4 hours. The use of biological replicates lead to a total of 24 samples for analysis. Cells were disrupted mechanically and membranes and soluble fractions were separated by centrifugation. The fractions were reduced, alkylated, tryptically digested, and labeled

with iTRAQ reagents using standard protocols. Pooled peptides were separated by SCX-RP and spotted onto a MALDI plate. MALDI-MS analysis was carried out using an ABI 4800 MALDI-TOF/TOF proteomics analyzer, and protein identifications were carried out using the Mascot search engine [120]. 744 proteins were quantified from membrane fractions and 688 proteins from the soluble fractions. A total of 846 unique proteins were identified representing 35% of predicted proteins in the genome. 163 of these were predicted to be integral membrane proteins. Approximately 20% of identified proteins were up-regulated in CFTR overexpressing strains while 10% were down-regulated. Amongst the up-regulated proteins were stress related proteins, including cell envelope stress proteins and heat shock proteins, which is indicative of misfolded proteins in the membrane. Analysis of the proteins revealed that *L. lactis* responded differently to *E. coli* during overexpression of CFTR and hence different approaches were needed to improve expression yields.

Sphingopyxis alaskensis is an abundant marine bacterium that is used as a model system for physiological studies [205]. The organism is part of a group of marine bacteria that might serve as sensitive indicators of changes in ultraviolet radiation on the Earth's surface [205]. Matallana-Surget *et al* utilized iTRAQ to investigate the effect of solar radiation and varying wavelengths of UV light on *S. alaskensis* [206]. 12 different growth conditions, under varying levels UV intensity and wavelength, simulating various depths under the ocean, in addition to full sunlight and complete darkness controls, were compared by iTRAQ. No effort was made to enrich membrane fractions. Samples were processed for iTRAQ labeling using standard sample preparation and peptide labeling protocols. Pooled labeled peptides were separated and analyzed by SCX and RP coupled

to an ABI QStar Pulsar *i* hybrid LC-MS/MS system, and identifications were made using the Mascot engine [120]. A total of 811 proteins were identified corresponding to 27% of the predicted ORFs. Comparison of pI-molecular weigh distributions between identified and predicted proteins showed a standard bimodal distribution in both sets, with a slight bias towards acidic side amongst the identified proteins but no visible bias based on molecular weight, demonstrating no significant overall bias in the methodology. Functional distribution by COGs [142] also revealed a similar distribution between observed and predicted proteins, except for proteins belonging to the COG V category, which encompasses proteins involved in defense mechanisms. This could be because many COG V proteins are membrane proteins with multiple membrane spanning domains like permeases and ABC transporters. Cellular localizations were predicted for approximately 57% of the identified proteins. 12.6% were predicted to localize either to the membranes or the periplasmic space. 62 were classified as IMPs (7.7%), 23 as OMPs (2.9%), 17 as periplasmic (2.1%), 2 were classified as extracellular (0.2%). 350 had unknown localization, though the authors suggested that some of these could be membrane localized. Interestingly 33% of the proteins identified in the pI 4-7 range were classified as hydrophobic based on GRAVY scores. 119 proteins were identified as differentially expressed between the compared conditions. These included proteins involved in protecting DNA from damage, detoxification, minimization of oxidative stress, chaperones as well as protein involved in nitrogen metabolism and in information processing pathways.

iTRAQ investigation of *Bdellovibrio bacteriovorus*

Jagtap *et al* (unpublished to date) carried out an elegant quantitative membrane

proteomic study of the predatory Gram-negative bacterium *Bdellovibrio bacteriovorus* using iTRAQ. The bacterium has an unusual lifestyle in which it spends a significant part of its lifecycle as an intracellular parasite of other Gram-negative bacteria [207-210]. The free form of the bacterium uses its flagellum to swim towards its host. It collides with and attaches to the host outer membrane. After which it enters through a pore it creates, and then reseals the pore to prevent entry of additional *Bdellovibrio* cells. Once inside the host cell, the parasitic bacterium attaches itself to the host inner membrane in the periplasmic space and makes use of the host-derived macromolecules for its metabolism and reproduction. Once the host resources are exhausted, the parasite multiplies into several motile *Bdellovibrio* daughter cells within the host cell, which subsequently lyse the host membrane and escape to initiate another round of their life cycle. Several membrane proteins have been implicated in the predatory lifecycle of *Bdellovibrio* [208, 210]. A global quantitative membrane proteomic analysis offers a powerful method to get a glimpse of some of the proteins involved in this.

Jagtap *et al* utilized a mutant host-independent strain HID2 that had evolved the ability to grow in host-free nutrient media and had lost its predatory ability for the study [211]. They carried out an iTRAQ comparison of this host-independent *Bdellovibrio* strain (HID2) with the wild type strain (HD100) grown on three different hosts - *Escherichia coli* (HDE), *Pseudomonas putida* (HDP) and *Salmonella typhimureum* (HDS). HID2 cells were grown free and membrane fractions were enriched. Wild type cells were grown on the three hosts. *Bdellovibrio* cells were separated from hosts by Percoll™ gradient ultracentrifugation and then membrane fractions were washed and enriched using sodium chloride and sodium carbonate. The quality of the membrane

preparations was tested using standard 2-DE. Once the preparations were deemed satisfactory, they were processed for iTRAQ analysis. For the iTRAQ experiments, membrane fractions were reduced, alkylated, tryptically digested and labeled with iTRAQ reagents as follows: HID2 ($iTRAQ_{114}$), HDE ($iTRAQ_{115}$), HDP ($iTRAQ_{116}$), and HDS ($iTRAQ_{117}$). The labeled membrane samples were quantitated and mixed in equal protein content. The mixtures were separated by 2D-LC comprising of SCX and RP. Mass spectrometric analysis on the fractions was carried out using an Applied Biosystems 4800 TOF/TOF instrument. Two biological replicates were used for the experiment to improve confidence in the results.

Multiple algorithms – Sequest [81], X! Tandem [212], Mascot [120], Phenyx (GeneBio, Geneva, Switzerland)[102, 141], and Paragon (ABI) [156] - with similar search parameters were used to identify proteins. There was extensive overlap between all algorithms, however there were some spectra identified exclusively by each software package (Figure 1.1). The results from all algorithms were aggregated to generate a combined raw data set (CRDS). Conflicts were handled by selecting results common to the majority of algorithms or discarding spectra in case of a tie. A false positive rate (1% FPR) was calculated using a reverse database of host sequences [213]. Results from biological replicate 1 resulted in a CDRS of 3363 peptides (at 1% FPR) yielding 486 unique *Bdellovibrio* proteins. 333 of these were identified with at least 2 peptides and 247 with at least 2 unique peptide matches. Replicate 2 resulted in 3995 peptides yielding 577 unique protein identifications. 390 of these proteins were identified with at least 2 peptides and 251 with at least 2 unique peptide matches. A combination of the data sets from both biological replicates resulted in 718 unique proteins. Of these 718 proteins

from the combined dataset of both replicates, 322 were identified with 2 or more unique peptides, while 263 were present in both biological replicates with at least 2 unique peptides in one set. Using a more stringent criterion of at least 2 unique peptides required in each replicate, 176 unique proteins were identified.

A pI versus molecular weight scatter plot of total predicted proteins in the genome (3583) overlaid with that of the 718 unique proteins, and the 322 proteins identified with 2 or more peptides, showed a similar bimodal pattern in all three sets (Figure 1.2). The acidic bias normally seen in 2-DE based proteomics does not appear in the identified proteins, and if anything, there may be a better representation of basic proteins based on a visual inspection of the scatter plot (discussed ahead). Interestingly the higher molecular weight proteins appear slightly better represented in the identified proteins. This could be due to a larger number of peptides generated from, as well as a potentially larger number of total and accessible tryptic sites present in the higher molecular weight proteins.

Of the 322 proteins identified from the combined dataset on the basis of 2 or more unique peptides, 94 (~29%) were classified as inner membrane proteins (IMPs) by PSORTb v3.0 [84, 85] and 23 (~7%) were classified as outer membrane proteins (OMPs). 107 (~33%) proteins were localized to the cytoplasm, 18 (5.6%) to the periplasmic space and 3 were classified as extracellular. 77 (~24%) proteins were designated as unknown or had no prediction data available (Figure 3.3). In addition to the assignments made by PSORTb [84, 85], membrane localization or association information was also computed by more methods such as the automated pre-annotation from the United States Department of Energy Joint Genome Institute (DOE JGI) Integrated Microbial Genomes (IMG) database [214], and the SOSUI [181, 182],

TMHMM [215, 216] and LipoP [186] algorithms. These resulted in 69 additional proteins (21%) being predicted as integral membrane or membrane associated. Taking the predictions from all algorithms additively, potentially up to 58% of proteins from these 322 proteins could be membrane proteins. A more conservative and realistic estimate can be reached by using a consensus between all the algorithms where a protein is considered to a membrane protein if it classified as a membrane protein by at least 3 of the 5 algorithms, This approach leads to an prediction of approximately 36% of identified proteins as membrane proteins. Regardless of where the actual percentage of proteins lies within this range, the results demonstrate a good representation of this class of proteins in the experiments, and validate the enrichment techniques as well as the utility of iTRAQ for the analysis of membrane proteins.

Of the 176 proteins identified in both biological replicates with 2 or more unique peptides in each replicate, 45 were predicted to be IMPs and 17 OMPs by PSORTb v3.0 [84, 85]. 58 proteins were predicted to be cytoplasmic, 10 periplasmic, and 1 was predicted to be extracellular. No localization predictions were made for 43 of the proteins, while 2 of the proteins were identified as *E. coli* host proteins. Of the proteins not classified as IMPs or OMPs, a further 35 are predicted to have transmembrane domains or are lipoproteins based on the additional algorithms described above. Taking the data together, between 35 to 55 % of these proteins could be membrane proteins (data not shown; available upon request). All proteins from this pool with a fold change of 1.7 (up or down) and a p-value of 0.5 or lower in at least one replicates of the three host conditions were deemed as differentially expressed. 66 proteins were classified as significantly differentially expressed in one or more condition. 30 of these were up-

regulated on average while 36 were down-regulated. Many of the down-regulated proteins were predicted to be membrane proteins. Amongst the down-regulated proteins, PSORTb [84, 85] predicted 10 IMPs, and 7 OMPs (Figure 1.3). Up to 11 more proteins could potentially be membrane, membrane-associated or lipoproteins based on analysis by the additional algorithms described earlier as well as two more algorithms for the prediction of outer membrane proteins BOMP [217] and TMB-Hunt (BBTM) [218, 219]. Thus potentially 78% of down-regulated proteins could be membrane proteins. However, using the more realistic consensus approach described in the previous paragraph, 20 of these proteins are highly likely to be membrane proteins. The down-regulated proteins belong to COG groups C, M, N, O, P, T, and U (Figure 1.4) e.g. TolC. Amongst the up-regulated proteins PSORTb [84, 85] predicted 2 IMPs and 3 OMPS. Integrating the results from the other algorithms, at the most 3 more protein are likely to be membrane or membrane-associated proteins, taking the total up to 8. A more realistic estimate of membrane proteins using a consensus approach between all the algorithms is between 3-5. Ribosomal proteins, which belong to 'COG J' (Figure 1.4, Figure 1.5b), constitute the largest group of up-regulated proteins in the host dependent lifestyle. These include RplD, RplM, RpsD, RplF, RpsU, Rpl32, RplQ RpsG, RpsI, RpsB and RplE. The large number of ribosomal proteins probably account for the apparent alkaline bias of the low molecular weight proteins (Figure 1.2). The investigators hypothesized that ribosomal proteins could be overexpressed to meet an increased demand during utilization of host biomolecules. However, it should be noted that ribosomal proteins are frequently seen as contaminants in membrane preparations [47]. The remaining up-regulated proteins belong to the COG groups M, N and P. These include MotA (which is expressed at

higher levels in HDE as compared to HDP and HDS) MurD. In addition to the differentially regulated *Bdellovibrio* proteins, host proteins Omp3A from *E. coli* and LRPP from *S. typhimureum* were also shown to be up-regulated, the exact reason for which was unclear. Utilizing a less stringent criteria for up and down regulation of proteins i.e. selecting from the dataset of 322 unique proteins with 2 or more peptides (rather than the dataset of 176 proteins that appeared in both replicates with 2 or more unique peptides in each replicate), several additional proteins were reported as up regulated and down-regulated (126 proteins totally, 54 were up-regulated and 72 were down-regulated on average). Amongst the up-regulated proteins are FolE and Pur, while DsbA, and PhoR are amongst the down-regulated proteins. These differentially expressed *Bcellovibrio* proteins (Figures 1.5a-d) are prime candidates for additional analysis to better understand how this interesting organism adapts to its unique lifestyle.

Conclusions

A review of the quantitative proteomics literature brings several important points to light. The clearest message that emerges from all the reported studies is the critical importance of carrying out quantitative proteomics analyses. The majority of studies reviewed here report discrepancies between transcriptomic and proteomic data, confirming the results of Gygi *et al* [6]. Thus it is not sufficient to carry out only mRNA based global expression analyses; incorporation of global protein expression changes is a must for a proper understanding of biological networks.

Another fact that becomes clear is the tremendous progress that has been made in proteomic technologies. Significant strides have been made to address two of the biggest limitations of traditional proteomics i.e. the inefficiency in analyzing hydrophobic membrane proteins, and the limited ability to quantitate changes in protein expression.

Shotgun methods like MudPIT combined with quantitative proteomic approaches like iTRAQ have driven these advances. In many of the studies described, traditional 2-DE type proteomic analyses were carried out in parallel with shotguns methods. For the proteins that have been quantitated in both methods, in general there has been agreement between the quantitative results, which validates the quantitative ability of both methods. However, there are cases when the two approaches yield divergent data [132]. More importantly, the limited overlap between proteins identified by both approaches, highlights the need of using multiple proteomic technologies to obtain reliable and complete proteome coverage. This is because each method exploits slightly different physico-chemical properties of the assayed proteins to provide separation. The results also underscore the need for continued development of non-shotgun approaches, such as

2-DE, side by side with shotgun methods, which leave plenty of room for further improvement in spite of their advances.

We have described some of the advances made in quantitative proteomics for the analysis of microbial membrane proteins with a focus on the iTRAQ methodology. All areas of the proteomic analytical chain have seen improvement. These include sample preparation, protein and peptide separations, proteolytic digestion, and identifications due to improved instrumentation, improved databases, advances in algorithms and statistical methods. The studies demonstrate that iTRAQ is one of the most powerful methods of choice for quantitative proteomics studies, and is particular suited for microbial membrane analyses. Even though it may not be possible to obtain complete proteome coverage using a single proteomic methods, iTRAQ based methods can offer among the best proteomic coverage when compared with other single methods [132, 133]. iTRAQ has some documented limitations such as the need for biological replicates and the need for multiple analyses of samples for robust quantification [128, 220], interference of peptides during precursor ion selection for MS/MS and potential contamination of the 121 m/z signal during 8-plex iTRAQ [221, 222] and decreased identification rates when higher '–plex' multiplexing is carried out [223]. Nevertheless, when these limitations are understood and accounted for, the technique provides unparalleled proteomic coverage.

Investigators wishing to pursue an iTRAQ based microbial membrane proteomic analysis are confronted with a bewildering array of choices with respect to sample handling methodologies. Given the conflicting data from the literature, these need to be determined empirically on a case-by-case basis with careful attention being paid to the biological questions being posed by the study. These include the choice of strategies for

enrichment, wash, solubilization, digestion, separation, identification, and data handling. I highlight some of the options available.

For microbial membrane studies, the first stage of the proteomic analysis chain is the enrichment of membranes and removal of contaminants. Ultracentrifugation to sediment insoluble membranes is the method of choice, though density gradient centrifugation may be used for more complex membrane systems such as cyanobacteria, or if binary membrane systems are present as is the case with *Bdellovibrio* cells in their hosts, or when additional purity is desired [23]. Other options for enrichment include precipitation, phase partitioning, and affinity purification. Delipidation using the chloroform/methanol system can be quite useful to remove contaminating lipids [23]. The use of a wash step is equally important. The carbonate wash protocol is very popular, however it needs to be determined on a case-by-case basis whether this step can adversely affect the sample being analyzed [24, 28, 50]. It is possible, and perhaps desirable, to use multiple enrichment techniques along with thorough washing because high purity is desirable for iTRAQ. This is due to the sensitivity of the method, which is evidenced in many of the studies described above, where proteins from contaminating fractions were detected in spite of robust enrichment.

The choices of solubilization and proteolytic digestion strategies are tightly tied together. Solubilization is often carried out using variations of standard 2-DE buffers containing detergent and chaotropes. Marked improvements have been made with the use of alternative solubilization strategies such as through the use of organic solvents like methanol and acetonitrile, organic acids like formic acid, and modern MS-compatible detergents like RapiGest, PPS, and Invitrosol [61, 76]. Another approach that has resulted

in significant improvements in solubilization and digestion is through the incorporation of the bile salt sodium deoxycholate (SDC), which can be further augmented by sarkosyl [62, 91]. All these strategies improve access of proteolytic enzymes like trypsin to hydrophobic domains of proteins in the membranes. Depending on the scenario, trypsin may be replaced or augmented with other proteolytic agents including the enzymes chymotrypsin, Lys-C, Glu-C, the non-specific protease Proteinase K, or the chemical cleavage agent cyanogen bromide (CNBr) [28-30]. More recently Rietschel *et a*l demonstrated the power elastase and pepsin for proteomic analysis of membranes for *Halobacterium salinarum* and *Corynebacterium glutamicum* [224-226]. The use of immobilized proteins, through embedment in gels, or immobilized proteolytic enzymes, also appears to enhance digestion [94, 227]. Evaluation of the synergistic potential of combinations of some or all of these approaches is a promising avenue of future research.

Separation of digested iTRAQ-labeled peptides is usually carried out using variations of the MudPIT approach i.e. through the orthogonal chromatographic techniques strong cation exchange (SCX) chromatography and reverse phase (RP) chromatography [30]. However IEF based separations, which may be augmented by the addition of trifluoroethanol (TFE), used in place of SCX also provide good result [98-100]. It is also possible to carry out 1-DE SDS-PAGE of intact proteins prior to digestion and labeling followed by post-digestion RP separation [201, 202]. The use of elevated temperatures during chromatographic steps also significantly improves peptide yields [93].

The final stage in the proteomic pipeline is the post MS analysis of the data For most studies, investigators choose a single algorithm to provide protein identifications. In

contrast Redding *et al* used two algorithms in their analysis of *Desulfovibrio*, while Jagtap *et al* combined 5 algorithms in their analysis of *Bdellovibrio.* The use of multiple algorithms can result in additional identifications, and more importantly can increase confidence in identifications.

References

[1] Edman, P., *Arch Biochem* 1949, *22*, 475.

[2] Maxam, A. M., Gilbert, W., *Proc Natl Acad Sci U S A* 1977, *74*, 560-564.

[3] Sanger, F., Nicklen, S., Coulson, A. R., *Proc Natl Acad Sci USA* 1977, *74*, 5463-5467.

[4] Hutchison, C. A., *Nucleic Acids Res* 2007, *35*, 6227-6237.

[5] Phillips, C. I., Bogyo, M., *Cell Microbiol* 2005, *7*, 1061-1076.

[6] Gygi, S. P., Rochon, Y., Franza, B. R., Aebersold, R., *Mol Cell Biol* 1999, *19*, 1720-1730.

[7] Poetsch, A., Wolters, D., *Proteomics* 2008, *8*, 4100-4122.

[8] Flory, M. R., Griffin, T. J., Martin, D., Aebersold, R., *Trends Biotechnol* 2002, *20*, S23-29.

[9] Jayapal, K. P., Philp, R. J., Kok, Y.-J., Yap, M. G. S., Sherman, D. H., Griffin, T. J., Hu, W.-S., *PLoS ONE* 2008, *3*, e2097.

[10] O'Farrell, P. H., *J Biol Chem* 1975, *250*, 4007-4021.

[11] O'Farrell, P. H., *Proteomics* 2008, *8*, 4842-4852.

[12] Kolkman, A., Dirksen, E. H. C., Slijper, M., Heck, A. J. R., *Mol Cell Proteomics* 2005, *4*, 255-266.

[13] Patton, W. F., *J Chromatogr B Analyt Technol Biomed Life Sci* 2002, *771*, 3-31.

[14] Graham, R. L., Graham, C., McMullan, G., *Microb Cell Fact* 2007, *6*, 26.

[15] Molloy, M. P., Donohoe, S., Brzezinski, E. E., Kilby, G. W., Stevenson, T. I., Baker, J. D., Goodlett, D. R., Gage, D. A., *Proteomics* 2005, *5*, 1204-1208.

[16] Linscheid, M. W., *Anal Bioanal Chem* 2005, *381*, 64-66.

[17] Schneider, L. V., Hall, M. P., *Drug Discov Today* 2005, *10*, 353-363.

[18] Wolff, S., Otto, A., Albrecht, D., Zeng, J. S., Büttner, K., Glückmann, M., Hecker, M., Becher, D., *Mol Cell Proteomics* 2006, *5*, 1183-1192.

[19] Josic, D., Brown, M. K., Huang, F., Callanan, H., Rucević, M., Nicoletti, A., Clifton, J., Hixson, D. C., *Electrophoresis* 2005, *26*, 2809-2822.

[20] Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., Yates, J. R., *Nat Biotechnol* 1999, *17*, 676-682.

[21] Wallin, E., von Heijne, G., *Protein Sci* 1998, *7*, 1029-1038.

[22] Wu, C. C., Yates, J. R., *Nat Biotechnol* 2003, *21*, 262-267.

[23] Speers, A. E., Wu, C. C., *Chem Rev* 2007, *107*, 3687-3714.

[24] Molloy, M. P., Herbert, B. R., Slade, M. B., Rabilloud, T., Nouwens, A. S., Williams, K. L., Gooley, A. A., *Eur J Biochem* 2000, *267*, 2871-2881.

[25] Nouwens, A. S., Cordwell, S. J., Larsen, M. R., Molloy, M. P., Gillings, M., Willcox, M. D., Walsh, B. J., *Electrophoresis* 2000, *21*, 3797-3809.

[26] Phadke, N. D., Molloy, M. P., Steinhoff, S. A., Ulintz, P. J., Andrews, P. C., Maddock, J. R., *Proteomics* 2001, *1*, 705-720.

[27] Bunai, K., Yamane, K., *J Chromatogr B Analyt Technol Biomed Life Sci* 2005, *815*, 227-236.

[28] Fischer, F., Wolters, D., Rögner, M., Poetsch, A., *Mol Cell Proteomics* 2006, *5*, 444-453.

[29] Wu, C. C., MacCoss, M. J., Howell, K. E., Yates, J. R., *Nat Biotechnol* 2003, *21*, 532-538.

[30] Washburn, M. P., Wolters, D., Yates, J. R., *Nat Biotechnol* 2001, *19*, 242-247.

[31] Goshe, M. B., Blonder, J., Smith, R. D., *J Proteome Res* 2003, *2*, 153-161.

[32] Gorg, A., Weiss, W., Dunn, M. J., *Proteomics* 2004, *4*, 3665-3685.

[33] Rabilloud, T., Chevallet, M., Luche, S., Lelong, C., *Journal of Proteomics* 2010, *73*, 2064-2077.

[34] Cordwell, S. J., Thingholm, T. E., *Proteomics* 2010, *10*, 611-627.

[35] Lu, B., McClatchy, D. B., Kim, J. Y., Yates, J. R., *Proteomics* 2008, *8*, 3947-3955.

[36] Chaney, L. K., Jacobson, B. S., *J Biol Chem* 1983, *258*, 10062-10072.

[37] Srivastava, R., Pisareva, T., Norling, B., *Proteomics* 2005, *5*, 4905-4916.

[38] Bordier, C., *J Biol Chem* 1981, *256*, 1604-1607.

[39] Sinha, S., Kosalai, K., Arora, S., Namane, A., Sharma, P., Gaikwad, A. N., Brodin, P., Cole, S. T., *Microbiology* 2005, *151*, 2411-2419.

[40] Mattow, J., Siejak, F., Hagens, K., Schmidt, F., Koehler, C., Treumann, A., Schaible, U. E., Kaufmann, S. H. E., *Proteomics* 2007, *7*, 1687-1701.

[41] Zheng, J., Wei, C., Leng, W., Dong, J., Li, R., Li, W., Wang, J., Zhang, Z., Jin, Q., *Proteomics* 2007, *7*, 3919-3931.

[42] Målen, H., Berven, F. S., Søfteland, T., Arntzen, M. Ø., D'Santos, C. S., De Souza, G. A., Wiker, H. G., *Proteomics* 2008, *8*, 1859-1870.

[43] Deshusses, J., Burgess, J., Scherl, A., Wenger, Y., Walter, N., Converset, V., Paesano, S., Corthals, G., Hochstrasser, D., Sanchez, J., *Proteomics* 2003, *3*, 1418-1424.

[44] Scherl, A., François, P., Bento, M., Deshusses, J. M., Charbonnier, Y., Converset, V., Huyghe, A., Walter, N., Hoogland, C., Appel, R. D., Sanchez, J.-C., Zimmermann-Ivol, C. G., Corthals, G. L., Hochstrasser, D. F., Schrenzel, J., *J Microbiol Methods* 2005, *60*, 247-257.

[45] Bisle, B., Schmidt, A., Scheibe, B., Klein, C., Tebbe, A., Kellermann, J., Siedler, F., Pfeiffer, F., Lottspeich, F., Oesterhelt, D., *Mol Cell Proteomics* 2006, *5*, 1543-1558.

[46] Lai, E.-M., Phadke, N. D., Kachman, M. T., Giorno, R., Vazquez, S., Vazquez, J. A., Maddock, J. R., Driks, A., *J Bacteriol* 2003, *185*, 1443-1454.

[47] Molloy, M. P., Phadke, N. D., Chen, H., Tyldesley, R., Garfin, D. E., Maddock, J. R., Andrews, P. C., *Proteomics* 2002, *2*, 899-910.

[48] Molloy, M. P., Phadke, N. D., Maddock, J. R., Andrews, P. C., *Electrophoresis* 2001, *22*, 1686-1696.

[49] Tafelmeyer, P., Laurent, C., Lenormand, P., Rousselle, J.-C., Marsollier, L., Reysset, G., Zhang, R., Sickmann, A., Stinear, T. P., Namane, A., Cole, S. T., *Proteomics* 2008, *8*, 3124-3138.

[50] Aivaliotis, M., Karas, M., Tsiotis, G., *J. Proteome Res.* 2007, *6*, 1048-1058.

[51] Kouyianou, K., Aivaliotis, M., Gevaert, K., Karas, M., Tsiotis, G., *Photosynth Res* 2010, *104*, 153-162.

[52] Chevallet, M., Santoni, V., Poinas, A., Rouquié, D., Fuchs, A., Kieffer, S., Rossignol, M., Lunardi, J., Garin, J., Rabilloud, T., *Electrophoresis* 1998, *19*, 1901-1909.

[53] Zeller, M., Brown, E., Bouvier, E., Konig, S., *J Biomol Tech* 2002, *13*, 1-4.

[54] Ross, A., Lee, P., Smith, D., Langridge, J., Whetton, A., Gaskell, S., *Proteomics* 2002, *2*, 928-936.

[55] Meng, F., Cargile, B. J., Patrie, S. M., Johnson, J. R., McLoughlin, S. M., Kelleher, N. L., *Anal Chem* 2002, *74*, 2923-2929.

[56] König, S., Schmidt, O., Rose, K., Thanos, S., Besselmann, M., Zeller, M., *Electrophoresis* 2003, *24*, 751-756.

[57] Yu, Y.-Q., Gilar, M., Lee, P. J., Bouvier, E. S. P., Gebler, J. C., *Anal Chem* 2003, *75*, 6023-6028.

[58] Norris, J. L., Porter, N. A., Caprioli, R. M., *Anal Chem* 2003, *75*, 6642-6647.

[59] Nomura, E., Katsuta, K., Ueda, T., Toriyama, M., Mori, T., Inagaki, N., *J Mass Spectrom* 2004, *39*, 202-207.

[60] Ruth, M. C., Old, W. M., Emrick, M. A., Meyer-Arendt, K., Aveline-Wolf, L. D., Pierce, K. G., Mendoza, A. M., Sevinsky, J. R., Hamady, M., Knight, R. D., Resing, K. A., Ahn, N. G., *J. Proteome Res.* 2006, *5*, 709-719.

[61] Chen, E. I., Cociorva, D., Norris, J. L., Yates, J. R., *J. Proteome Res.* 2007, *6*, 2529-2538.

[62] Masuda, T., Tomita, M., Ishihama, Y., *Journal of proteome research* 2008, *7*, 731-740.

[63] Rabilloud, T., Adessi, C., Giraudel, A., Lunardi, J., *Electrophoresis* 1997, *18*, 307-316.

[64] Herbert, B. R., Molloy, M. P., Gooley, A. A., Walsh, B. J., Bryson, W. G., Williams, K. L., *Electrophoresis* 1998, *19*, 845-851.

[65] Bjellqvist, B., Ek, K., Righetti, P. G., Gianazza, E., Gorg, A., Westermeier, R., Postel, W., *J Biochem Biophys Methods* 1982, *6*, 317-339.

[66] Gorg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R., Weiss, W., *Electrophoresis* 2000, *21*, 1037-1053.

[67] Westbrook, J. A., Yan, J. X., Wait, R., Welson, S. Y., Dunn, M. J., *Electrophoresis* 2001, *22*, 2865-2871.

[68] Rais, I., Karas, M., Schagger, H., *Proteomics* 2004, *4*, 2567-2571.

[69] Braun, R. J., Kinkl, N., Beer, M., Ueffing, M., *Analytical and bioanalytical chemistry* 2007, *389*, 1033-1045.

[70] Rabilloud, T., *Electrophoresis* 2009, *30 Suppl 1*, S174-180.

[71] Schluesener, D., Rogner, M., Poetsch, A., *Anal Bioanal Chem* 2007, *389*, 1055-1064.

[72] Williams, T. I., Combs, J. C., Thakur, A. P., Strobel, H. J., Lynn, B. C., *Electrophoresis* 2006, *27*, 2984-2995.

[73] Lasserre, J.-P., Beyne, E., Pyndiah, S., Lapaillerie, D., Claverol, S., Bonneu, M., *Electrophoresis* 2006, *27*, 3306-3321.

[74] van Montfort, B. A., Canas, B., Duurkens, R., Godovac-Zimmermann, J., Robillard, G. T., *J Mass Spectrom* 2002, *37*, 322-330.

[75] van Montfort, B. A., Doeven, M. K., Canas, B., Veenhoff, L. M., Poolman, B., Robillard, G. T., *Biochim Biophys Acta* 2002, *1555*, 111-115.

[76] Blonder, J., Goshe, M. B., Moore, R. J., Pasa-Tolic, L., Masselon, C. D., Lipton, M. S., Smith, R. D., *J. Proteome Res.* 2002, *1*, 351-360.

[77] Lai, E.-M., Nair, U., Phadke, N. D., Maddock, J. R., *Mol Microbiol* 2004, *52*, 1029-1044.

[78] Unlu, M., Morgan, M. E., Minden, J. S., *Electrophoresis* 1997, *18*, 2071-2077.

[79] Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins, E., Currie, I., Davison, M., *Proteomics* 2001, *1*, 377-396.

[80] Yan, J. X., Devenish, A. T., Wait, R., Stone, T., Lewis, S., Fowler, S., *Proteomics* 2002, *2*, 1682-1698.

[81] Eng, J. K., McCormack, A. L., Yates Iii, J. R., *Journal of the American Society for Mass Spectrometry* 1994, *5*, 976-989.

[82] Rodríguez-Ortega, M. J., Norais, N., Bensi, G., Liberatori, S., Capo, S., Mora, M., Scarselli, M., Doro, F., Ferrari, G., Garaguso, I., Maggi, T., Neumann, A., Covre, A., Telford, J. L., Grandi, G., *Nat Biotechnol* 2006, *24*, 191-197.

[83] Tjalsma, H., Lambooy, L., Hermans, P. W., Swinkels, D. W., *Proteomics* 2008, *8*, 1415-1428.

[84] Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S. C., Ester, M., Foster, L. J., Brinkman, F. S., *Bioinformatics* 2010, *26*, 1608-1615.

[85] Gardy, J. L., Laird, M. R., Chen, F., Rey, S., Walsh, C. J., Ester, M., Brinkman, F. S., *Bioinformatics* 2005, *21*, 617-623.

[86] Blonder, J., Conrads, T. P., Yu, L.-R., Terunuma, A., Janini, G. M., Issaq, H. J., Vogel, J. C., Veenstra, T. D., *Proteomics* 2004, *4*, 31-45.

[87] Zhang, N., Chen, R., Young, N., Wishart, D., Winter, P., Weiner, J. H., Li, L., *Proteomics* 2007, *7*, 484-493.

[88] Blonder, J., Goshe, M. B., Xiao, W., Camp, D. G., Wingerd, M., Davis, R. W., Smith, R. D., *J. Proteome Res.* 2004, *3*, 434-444.

[89] Blackler, A. R., Speers, A. E., Ladinsky, M. S., Wu, C. C., *Journal of proteome research* 2008, *7*, 3028-3034.

[90] Barrios-Llerena, M. E., Chong, P. K., Gan, C. S., Snijders, A. P. L., Reardon, K. F., Wright, P. C., *Briefings in functional genomics & proteomics* 2006, *5*, 121-132.

[91] Masuda, T., Saito, N., Tomita, M., Ishihama, Y., *Mol Cell Proteomics* 2009, *8*, 2770-2777.

[92] Zhong, H., Marcus, S. L., Li, L., *J Am Soc Mass Spectrom* 2005, *16*, 471-481.

[93] Speers, A. E., Blackler, A. R., Wu, C. C., *Anal Chem* 2007, *79*, 4613-4620.

[94] Lu, X., Zhu, H., *Mol Cell Proteomics* 2005, *4*, 1948-1958.

[95] Han, C.-L., Chien, C.-W., Chen, W.-C., Chen, Y.-R., Wu, C.-P., Li, H., Chen, Y.-J., *Molecular & Cellular Proteomics* 2008, *7*, 1983-1997.

[96] Zhou, J., Huang, S., Bi, D., Zhang, H., Li, J., Lin, Y., Chen, P., Wang, X., Liang, S., *Electrophoresis* 2009, *30*, 4109-4117.

[97] Zhou, J., Xiong, J., Li, J., Huang, S., Zhang, H., He, Q., Lin, Y., Chen, P., Wang, X., Liang, S., *Analytical Biochemistry* 2010, *404*, 204-210.

[98] Gan, C. S., Reardon, K. F., Wright, P. C., *Proteomics* 2005, *5*, 2468-2478.

[99] Chong, P. K., Wright, P. C., *J Proteome Res* 2005, *4*, 1789-1798.

[100] Scherl, A., François, P., Charbonnier, Y., Deshusses, J. M., Koessler, T., Huyghe, A., Bento, M., Stahl-Zeng, J., Fischer, A., Masselot, A., Vaezzadeh, A., Gallé, F., Renzoni, A., Vaudaux, P., Lew, D., Zimmermann-Ivol, C. G., Binz, P.-A., Sanchez, J.-C., Hochstrasser, D. F., Schrenzel, J., *BMC Genomics* 2006, *7*, 296.

[101] Ernoult, E., Gamelin, E., Guette, C., *Proteome Sci* 2008, *6*, 27.

[102] Heller, M., Ye, M., Michel, P. E., Morier, P., Stalder, D., Junger, M. A., Aebersold, R., Reymond, F., Rossier, J. S., *J Proteome Res* 2005, *4*, 2273-2282.

[103] Busnel, J. M., Lion, N., Girault, H. H., *Anal Chem* 2007, *79*, 5949-5955.

[104] Cordwell, S. J., *Curr Opin Microbiol* 2006, *9*, 320-329.

[105] Oda, Y., Huang, K., Cross, F. R., Cowburn, D., Chait, B. T., *Proc Natl Acad Sci U S A* 1999, *96*, 6591-6596.

[106] Fenselau, C., *J Chromatogr B Analyt Technol Biomed Life Sci* 2007, *855*, 14-20.

[107] Becher, D., Hempel, K., Sievers, S., Zühlke, D., Pané-Farré, J., Otto, A., Fuchs, S., Albrecht, D., Bernhardt, J., Engelmann, S., Völker, U., van Dijl, J. M., Hecker, M., *PLoS ONE* 2009, *4*, e8176.

[108] Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., Mann, M., *Mol Cell Proteomics* 2002, *1*, 376-386.

[109] Han, X., Aslanian, A., Yates, J. R., *Curr Opin Chem Biol* 2008, *12*, 483-490.

[110] Yao, X., Freas, A., Ramirez, J., Demirev, P. A., Fenselau, C., *Anal Chem* 2001, *73*, 2836-2842.

[111] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., Aebersold, R., *Nat Biotechnol* 1999, *17*, 994-999.

[112] Leichert, L. I., Gehrke, F., Gudiseva, H. V., Blackwell, T., Ilbert, M., Walker, A. K., Strahler, J. R., Andrews, P. C., Jakob, U., *Proc Natl Acad Sci USA* 2008, *105*, 8197-8202.

[113] Olsen, J. V., Andersen, J. R., Nielsen, P. A., Nielsen, M. L., Figeys, D., Mann, M., Wisniewski, J. R., *Mol Cell Proteomics* 2004, *3*, 82-92.

[114] Schmidt, A., Kellermann, J., Lottspeich, F., *Proteomics* 2005, *5*, 4-15.

[115] Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., Kuster, B., *Anal Bioanal Chem* 2007, *389*, 1017-1031.

[116] Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., Pappin, D. J., *Mol Cell Proteomics* 2004, *3*, 1154-1169.

[117] Zhu, W., Smith, J. W., Huang, C.-M., *J. Biomed. Biotech.* 2010, *2010*, 840518.

[118] Kuntumalla, S., Braisted, J. C., Huang, S.-T., Parmar, P. P., Clark, D. J., Alami, H., Zhang, Q., Donohue-Rolfe, A., Tzipori, S., Fleischmann, R. D., Peterson, S. N., Pieper, R., *Proteome Sci* 2009, *7*, 22.

[119] Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A. K. A., Hamon, C., *Anal Chem* 2003, *75*, 1895-1904.

[120] Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S., *Electrophoresis* 1999, *20*, 3551-3567.

[121] Xiang, F., Ye, H., Chen, R., Fu, Q., Li, L., *Anal Chem* 2010, *82*, 2817-2825.

[122] Dayon, L., Hainard, A., Licker, V., Turck, N., Kuhn, K., Hochstrasser, D. F., Burkhard, P. R., Sanchez, J.-C., *Anal Chem* 2008, *80*, 2921-2931.

[123] van Ulsen, P., Kuhn, K., Prinz, T., Legner, H., Schmid, P., Baumann, C., Tommassen, J., *Proteomics* 2009, *9*, 1771-1781.

[124] Zhang, J., Wang, Y., Li, S., *Anal Chem* 2010, *82*, 7588-7595.

[125] Aggarwal, K., Choe, L. H., Lee, K. H., *Proteomics* 2005, *5*, 2297-2308.

[126] Chen, X., Walker, A. K., Strahler, J. R., Simon, E. S., Tomanicek-Volk, S. L., Nelson, B. B., Hurley, M. C., Ernst, S. A., Williams, J. A., Andrews, P. C., *Mol Cell Proteomics* 2006, *5*, 306-312.

[127] Redding, A. M., Mukhopadhyay, A., Joyner, D. C., Hazen, T. C., Keasling, J. D., *Brief Funct Genomic Proteomic* 2006, *5*, 133-143.

[128] Gan, C. S., Chong, P. K., Pham, T. K., Wright, P. C., *J. Proteome Res.* 2007, *6*, 821-827.

[129] Choe, L., D'Ascenzo, M., Relkin, N. R., Pappin, D., Ross, P., Williamson, B., Guertin, S., Pribil, P., Lee, K. H., *Proteomics* 2007, *7*, 3651-3660.

[130] Ow, S. Y., Cardona, T., Taton, A., Magnuson, A., Lindblad, P., Stensjö, K., Wright, P. C., *J Proteome Res* 2008, *7*, 1615-1628.

[131] Jagtap, P., Michailidis, G., Zielke, R., Walker, A. K., Patel, N., Strahler, J. R., Driks, A., Andrews, P. C., Maddock, J. R., *Proteomics* 2006, *6*, 5199-5211.

[132] Wu, W. W., Wang, G., Baek, S. J., Shen, R.-F., *J. Proteome Res.* 2006, *5*, 651-658.

[133] Choe, L. H., Aggarwal, K., Franck, Z., Lee, K. H., *Electrophoresis* 2005, *26*, 2437-2449.

[134] Eymann, C., Dreisbach, A., Albrecht, D., Bernhardt, J., Becher, D., Gentner, S., Tam, L. T., Büttner, K., Buurman, G., Scharf, C., Venz, S., Völker, U., Hecker, M., *Proteomics* 2004, *4*, 2849-2876.

[135] Wolff, S., Antelmann, H., Albrecht, D., Becher, D., Bernhardt, J., Bron, S., Büttner, K., van Dijl, J. M., Eymann, C., Otto, A., Tam, L. T., Hecker, M., *J Chromatogr B Analyt Technol Biomed Life Sci* 2007, *849*, 129-140.

[136] Hiramatsu, K., Aritaka, N., Hanaki, H., Kawasaki, S., Hosoda, Y., Hori, S., Fukuchi, Y., Kobayashi, I., *Lancet* 1997, *350*, 1670-1673.

[137] Kim, M. N., Pai, C. H., Woo, J. H., Ryu, J. S., Hiramatsu, K., *J Clin Microbiol* 2000, *38*, 3879-3881.

[138] Tenover, F. C., Lancaster, M. V., Hill, B. C., Steward, C. D., Stocker, S. A., Hancock, G. A., O'Hara, C. M., McAllister, S. K., Clark, N. C., Hiramatsu, K., *J Clin Microbiol* 1998, *36*, 1020-1027.

[139] Weigel, L. M., Clewell, D. B., Gill, S. R., Clark, N. C., McDougal, L. K., Flannagan, S. E., Kolonay, J. F., Shetty, J., Killgore, G. E., Tenover, F. C., *Science* 2003, *302*, 1569-1571.

[140] Cui, L., Iwamoto, A., Lian, J. Q., Neoh, H. M., Maruyama, T., Horikawa, Y., Hiramatsu, K., *Antimicrob Agents Chemother* 2006, *50*, 428-438.

[141] Colinge, J., Masselot, A., Giron, M., Dessingy, T., Magnin, J., *Proteomics* 2003, *3*, 1454-1463.

[142] Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Smirnov, S., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J., Natale, D. A., *BMC Bioinformatics* 2003, *4*, 41.

[143] Radosevich, T. J., Reinhardt, T. A., Lippolis, J. D., Bannantine, J. P., Stabel, J. R., *Journal of Bacteriology* 2007, *189*, 1109-1117.

[144] Stabel, J. R., *J Dairy Sci* 1998, *81*, 283-288.

[145] Gao, Q., Kripke, K., Arinc, Z., Voskuil, M., Small, P., *Tuberculosis (Edinb)* 2004, *84*, 188-196.

[146] Mawuenyega, K. G., Forst, C. V., Dobos, K. M., Belisle, J. T., Chen, J., Bradbury, E. M., Bradbury, A. R., Chen, X., *Mol Biol Cell* 2005, *16*, 396-404.

[147] van der Werf, T. S., Stinear, T., Stienstra, Y., van der Graaf, W. T., Small, P. L., *Lancet* 2003, *362*, 1062-1064.

[148] Marsollier, L., Aubry, J., Coutanceau, E., Andre, J. P., Small, P. L., Milon, G., Legras, P., Guadagnini, S., Carbonnelle, B., Cole, S. T., *Cell Microbiol* 2005, *7*, 935-943.

[149] Shpigel, N. Y., Elazar, S., Rosenshine, I., *Curr Opin Microbiol* 2008, *11*, 60-65.

[150] Lippolis, J. D., Bayles, D. O., Reinhardt, T. A., *J. Proteome Res.* 2009, *8*, 149-158.

[151] Levett, P. N., *Clin Microbiol Rev* 2001, *14*, 296-326.

[152] Lo, M., Cordwell, S. J., Bulach, D. M., Adler, B., *PLoS Negl Trop Dis* 2009, *3*, e560.

[153] Lo, M., Bulach, D. M., Powell, D. R., Haake, D. A., Matsunaga, J., Paustian, M. L., Zuerner, R. L., Adler, B., *Infect Immun* 2006, *74*, 5848-5859.

[154] Haake, D. A., Walker, E. M., Blanco, D. R., Bolin, C. A., Miller, M. N., Lovett, M. A., *Infect Immun* 1991, *59*, 1131-1140.

[155] Haake, D. A., Chao, G., Zuerner, R. L., Barnett, J. K., Barnett, D., Mazel, M., Matsunaga, J., Levett, P. N., Bolin, C. A., *Infect Immun* 2000, *68*, 2276-2285.

[156] Shilov, I. V., Seymour, S. L., Patel, A. A., Loboda, A., Tang, W. H., Keating, S. P., Hunter, C. L., Nuwaysir, L. M., Schaeffer, D. A., *Mol Cell Proteomics* 2007, *6*, 1638-1655.

[157] Cordwell, S. J., Len, A. C., Touma, R. G., Scott, N. E., Falconer, L., Jones, D., Connolly, A., Crossett, B., Djordjevic, S. P., *Proteomics* 2008, *8*, 122-139.

[158] Eshghi, A., Cullen, P. A., Cowen, L., Zuerner, R. L., Cameron, C. E., *J. Proteome Res.* 2009, *8*, 4564-4578.

[159] Dijkshoorn, L., Nemec, A., Seifert, H., *Nat Rev Microbiol* 2007, *5*, 939-951.

[160] Marti, S., Sanchez-Cespedes, J., Oliveira, E., Bellido, D., Giralt, E., Vila, J., *Proteomics* 2006, *6 Suppl 1*, S82-87.

[161] Siroy, A., Cosette, P., Seyer, D., Lemaitre-Guillier, C., Vallenet, D., Van Dorsselaer, A., Boyer-Mariotte, S., Jouenne, T., De, E., *J Proteome Res* 2006, *5*, 3385-3398.

[162] Soares, N. C., Cabral, M. P., Parreira, J. R., Gayoso, C., Barba, M. J., Bou, G., *Proteome Sci* 2009, *7*, 37.

[163] Fernandez-Reyes, M., Rodriguez-Falcon, M., Chiva, C., Pachon, J., Andreu, D., Rivas, L., *Proteomics* 2009, *9*, 1632-1645.

[164] Soares, N. C., Cabral, M. P., Gayoso, C., Mallo, S., Rodriguez-Velo, P., Fernández-Moreira, E., Bou, G., *J. Proteome Res.* 2010, *9*, 1951-1964.

[165] Mayor, S., *BMJ* 2008, *337*, a1678.

[166] Briolant, S., Almeras, L., Fusai, T., Rogier, C., Pradines, B., *Med Trop (Mars)* 2007, *67*, 86-96.

[167] *Med Mal Infect* 2008, *38*, 68-117.

[168] Chopra, I., Roberts, M., *Microbiol Mol Biol Rev* 2001, *65*, 232-260 ; second page, table of contents.

[169] Lin, Q., Katakura, K., Suzuki, M., *FEBS Lett* 2002, *515*, 71-74.

[170] Dahl, E. L., Shock, J. L., Shenai, B. R., Gut, J., DeRisi, J. L., Rosenthal, P. J., *Antimicrob Agents Chemother* 2006, *50*, 3124-3131.

[171] Goodman, C. D., Su, V., McFadden, G. I., *Mol Biochem Parasitol* 2007, *152*, 181-191.

[172] Briolant, S., Almeras, L., Belghazi, M., Boucomont-Chapeaublanc, E., Wurtz, N., Fontaine, A., Granjeaud, S., Fusaï, T., Rogier, C., Pradines, B., *Malar J* 2010, *9*, 141.

[173] Socransky, S. S., Haffajee, A. D., Cugini, M. A., Smith, C., Kent, R. L., Jr., *J Clin Periodontol* 1998, *25*, 134-144.

[174] Mah, T. F., O'Toole, G. A., *Trends Microbiol* 2001, *9*, 34-39.

[175] Pham, T. K., Roy, S., Noirel, J., Douglas, I., Wright, P. C., Stafford, G. P., *Proteomics* 2010, *10*, 3130-3141.

[176] Koebnik, R., *Trends Microbiol* 2005, *13*, 343-347.

[177] Koropatkin, N., Martens, E. C., Gordon, J. I., Smith, T. J., *Biochemistry* 2009, *48*, 1532-1542.

[178] Wang, Y., Xu, W., Chitnis, P. R., *Proteome Sci* 2009, *7*, 11.

[179] Pandhal, J., Ow, S. Y., Wright, P. C., Biggs, C. A., *J. Proteome Res.* 2009, *8*, 818-828.

[180] Stensjö, K., Ow, S. Y., Barrios-Llerena, M. E., Lindblad, P., Wright, P. C., *J. Proteome Res.* 2007, *6*, 621-635.

[181] Mitaku, S., Ono, M., Hirokawa, T., Boon-Chieng, S., Sonoyama, M., *Biophys Chem* 1999, *82*, 165-171.

[182] Hirokawa, T., Boon-Chieng, S., Mitaku, S., *Bioinformatics* 1998, *14*, 378-379.

[183] Chisholm, S. W., Olson, R. J., Zettler, E. R., Goericke, R., Waterbury, J. B., Welschmeyer, N. A., *Nature* 1988, *334*, 340-343.

[184] Rocap, G., Larimer, F. W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. A., Arellano, A., Coleman, M., Hauser, L., Hess, W. R., Johnson, Z. I., Land, M., Lindell, D., Post, A. F., Regala, W., Shah, M., Shaw, S. L., Steglich, C., Sullivan, M. B., Ting, C. S., Tolonen, A., Webb, E. A., Zinser, E. R., Chisholm, S. W., *Nature* 2003, *424*, 1042-1047.

[185] Pandhal, J., Wright, P. C., Biggs, C. A., *J. Proteome Res.* 2007, *6*, 996-1005.

[186] Juncker, A. S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H., Krogh, A., *Protein Sci* 2003, *12*, 1652-1662.

[187] Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M., Tabata, S., *DNA Res* 1996, *3*, 185-209.

[188] Taji, T., Seki, M., Satou, M., Sakurai, T., Kobayashi, M., Ishiyama, K., Narusaka, Y., Narusaka, M., Zhu, J. K., Shinozaki, K., *Plant Physiol* 2004, *135*, 1697-1709.

[189] Rowland, J. G., Simon, W. J., Nishiyama, Y., Slabas, A. R., *Proteomics* 2010, *10*, 1917-1929.

[190] Van, P. T., Schmid, A. K., King, N. L., Kaur, A., Pan, M., Whitehead, K., Koide, T., Facciotti, M. T., Goo, Y. A., Deutsch, E. W., Reiss, D. J., Mallick, P., Baliga, N. S., *J Proteome Res* 2008, *7*, 3755-3764.

[191] Williams, T. J., Burg, D. W., Ertan, H., Raftery, M. J., Poljak, A., Guilhaus, M., Cavicchioli, R., *J Proteome Res* 2010, *9*, 653-663.

[192] Williams, T. J., Burg, D. W., Raftery, M. J., Poljak, A., Guilhaus, M., Pilak, O., Cavicchioli, R., *Journal of proteome research* 2010, *9*, 640-652.

[193] Saunders, N. F., Ng, C., Raftery, M., Guilhaus, M., Goodchild, A., Cavicchioli, R., *J Proteome Res* 2006, *5*, 2457-2464.

[194] Goodchild, A., Raftery, M., Saunders, N. F., Guilhaus, M., Cavicchioli, R., *J Proteome Res* 2005, *4*, 473-480.

[195] Goodchild, A., Raftery, M., Saunders, N. F., Guilhaus, M., Cavicchioli, R., *J Proteome Res* 2004, *3*, 1164-1176.

[196] Goodchild, A., Saunders, N. F., Ertan, H., Raftery, M., Guilhaus, M., Curmi, P. M., Cavicchioli, R., *Mol Microbiol* 2004, *53*, 309-321.

[197] Snijders, A. P., Walther, J., Peter, S., Kinnman, I., de Vos, M. G., van de Werken, H. J., Brouns, S. J., van der Oost, J., Wright, P. C., *Proteomics* 2006, *6*, 1518-1529.

[198] Chong, P. K., Burja, A. M., Radianingtyas, H., Fazeli, A., Wright, P. C., *J. Proteome Res.* 2007, *6*, 1430-1439.

[199] Pham, T. K., Sierocinski, P., van der Oost, J., Wright, P. C., *J Proteome Res* 2010, *9*, 1165-1172.

[200] Zaparty, M., Esser, D., Gertig, S., Haferkamp, P., Kouril, T., Manica, A., Pham, T. K., Reimann, J., Schreiber, K., Sierocinski, P., Teichmann, D., van Wolferen, M., von Jan, M., Wieloch, P., Albers, S. V., Driessen, A. J. M., Klenk, H.-P., Schleper, C., Schomburg, D., van der Oost, J., Wright, P. C., Siebers, B., *Extremophiles* 2010, *14*, 119-142.

[201] Manteca, A., Jung, H. R., Schwämmle, V., Jensen, O. N., Sanchez, J., *Journal of proteome research* 2010.

[202] Manteca, A., Sanchez, J., Jung, H. R., Schwämmle, V., Jensen, O. N., *Molecular & Cellular Proteomics* 2010, *9*, 1423-1436.

[203] Steen, A., Wiederhold, E., Gandhi, T., Breitling, R., Slotboom, D. J., *Molecular & Cellular Proteomics* 2010.

[204] Kunji, E. R., Chan, K. W., Slotboom, D. J., Floyd, S., O'Connor, R., Monne, M., *Curr Opin Biotechnol* 2005, *16*, 546-551.

[205] Eguchi, M., Ostrowski, M., Fegatella, F., Bowman, J., Nichols, D., Nishino, T., Cavicchioli, R., *Appl Environ Microbiol* 2001, *67*, 4945-4954.

[206] Matallana-Surget, S., Joux, F., Raftery, M. J., Cavicchioli, R., *Environ Microbiol* 2009, *11*, 2660-2675.

[207] Dori-Bachash, M., Dassa, B., Pietrokovski, S., Jurkevitch, E., *Applied and Environmental Microbiology* 2008, *74*, 7152-7162.

[208] Lambert, C., Chang, C.-Y., Capeness, M. J., Sockett, R. E., *PLoS ONE* 2010, *5*, e8599.

[209] Medina, A. A., Kadouri, D. E., *Research in Microbiologoy* 2010, *160*, 224-231.

[210] Wolfe, A. J., *J Bacteriol* 2010.

[211] Cotter, T. W., Thomashow, M. F., *J Bacteriol* 1992, *174*, 6018-6024.

[212] Craig, R., Beavis, R. C., *Bioinformatics* 2004, *20*, 1466-1467.

[213] Elias, J. E., Gygi, S. P., *Nat Methods* 2007, *4*, 207-214.

[214] Markowitz, V. M., Chen, I.-M. A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A., Anderson, I., Lykidis, A., Mavromatis, K., Ivanova, N. N., Kyrpides, N. C., *Nucleic Acids Research*.

[215] Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E. L., *J Mol Biol* 2001, *305*, 567-580.

[216] Sonnhammer, E. L., von Heijne, G., Krogh, A., *Proc Int Conf Intell Syst Mol Biol* 1998, *6*, 175-182.

[217] Berven, F. S., Flikka, K., Jensen, H. B., Eidhammer, I., *Nucleic Acids Res* 2004, *32*, W394-399.

[218] Garrow, A. G., Agnew, A., Westhead, D. R., *Nucleic Acids Res* 2005, *33*, W188-192.

[219] Garrow, A. G., Agnew, A., Westhead, D. R., *BMC Bioinformatics* 2005, *6*, 56.

[220] Chong, P. K., Gan, C. S., Pham, T. K., Wright, P. C., *J. Proteome Res.* 2006, *5*, 1232-1240.

[221] Ow, S. Y., Salim, M., Noirel, J., Evans, C., Rehman, I., Wright, P. C., *J Proteome Res* 2009, *8*, 5347-5355.

[222] Perkel, J. M., *J. Proteome Res.* 2009, *8*, 4885.

[223] Pichler, P., Köcher, T., Holzmann, J., Mazanek, M., Taus, T., Ammerer, G., Mechtler, K., *Anal Chem* 2010, *82*, 6549-6558.

[224] Rietschel, B., Bornemann, S., Arrey, T. N., Baeumlisberger, D., Karas, M., Meyer, B., *Proteomics* 2009, *9*, 5553-5557.

[225] Rietschel, B., Arrey, T. N., Meyer, B., Bornemann, S., Schuerken, M., Karas, M., Poetsch, A., *Mol Cell Proteomics* 2009, *8*, 1029-1043.

[226] Gilmore, J. M., Washburn, M. P., *Journal of Proteomics* 2010.

[227] Pereira-Medrano, A. G., Forster, S., Fowler, G. J. S., Mcarthur, S. L., Wright, P. C., *Lab on a chip* 2010.

Figure 1.1: Five independent algorithms identified non-perfectly overlapping, but similar number of spectra (first 5 bars of chart). The total spectra identified (6^{th} bar of chart) were a consensus of the five algorithms, and ranged from being identified by all algorithms to only 1 with all the different combinations in between. The distribution between the algorithms is shown in the 6^{th} bar of the chart.

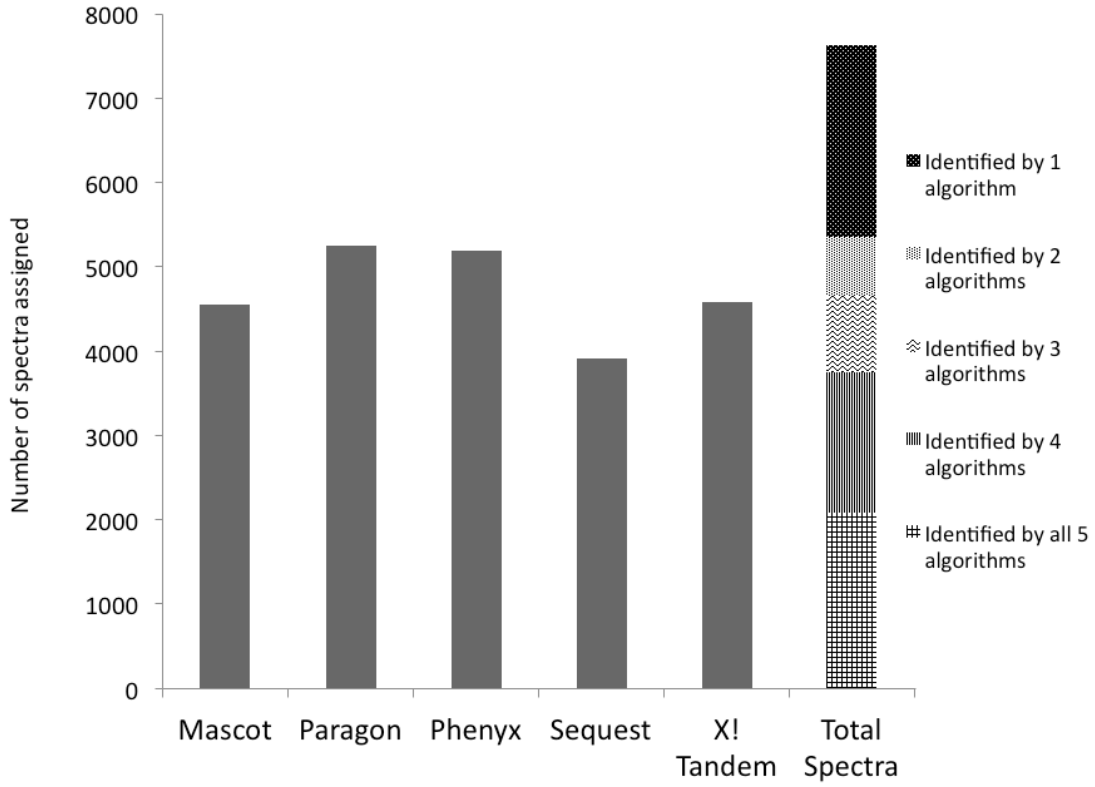

8000
7000
6000
5000
4000
3000
2000
1000
0
Number of spectra assigned
Mascot
Paragon
Phenyx
Sequest
X! Tandem
Total Spectra
Identified by 1 algorithm
Identified by 2 algorithms
Identified by 3 algorithms
Identified by 4 algorithms
Identified by all 5 algorithms

Figure 1.2: pI versus molecular weight plot of *Bdellovibrio bacteriovorus* predicted proteins overlaid with pI versus molecular weight plot of unique proteins identified by iTRAQ and unique proteins identified with high confidence of 2 or more unique peptides.

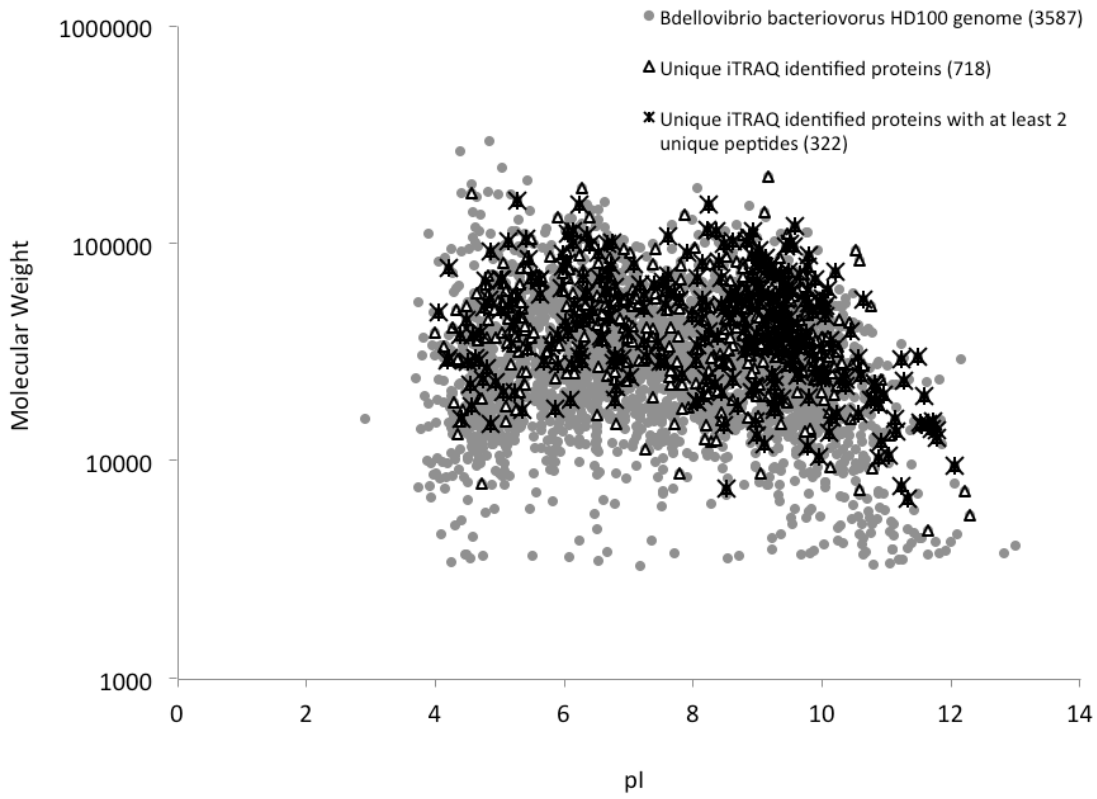

Bdellovibrio bacteriovorus HD100 genome (3587)
Unique iTRAQ identified proteins (718)
Unique iTRAQ identified proteins with at least 2 unique peptides (322)
Molecular Weight
1000000
100000
10000
1000
0
2
4
6
8
10
12
14
pI

Figure 1.3: Localization of predicted, identified, and differentially regulated proteins by PSORTb v3.0.

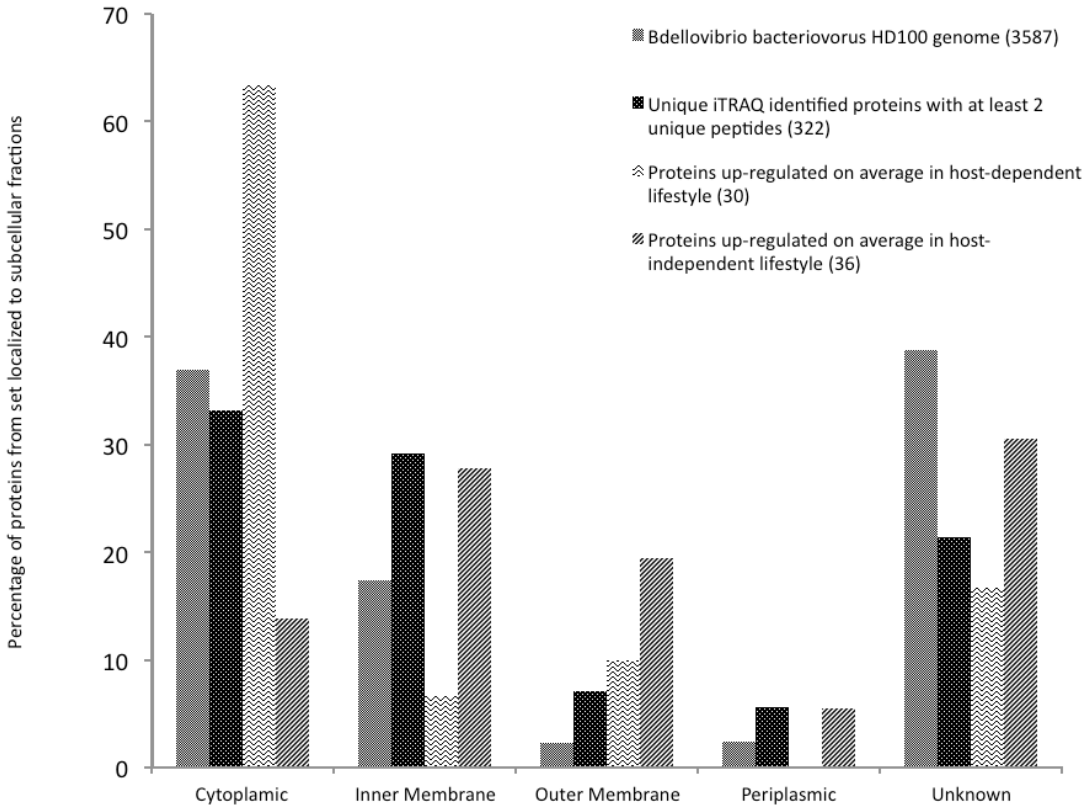
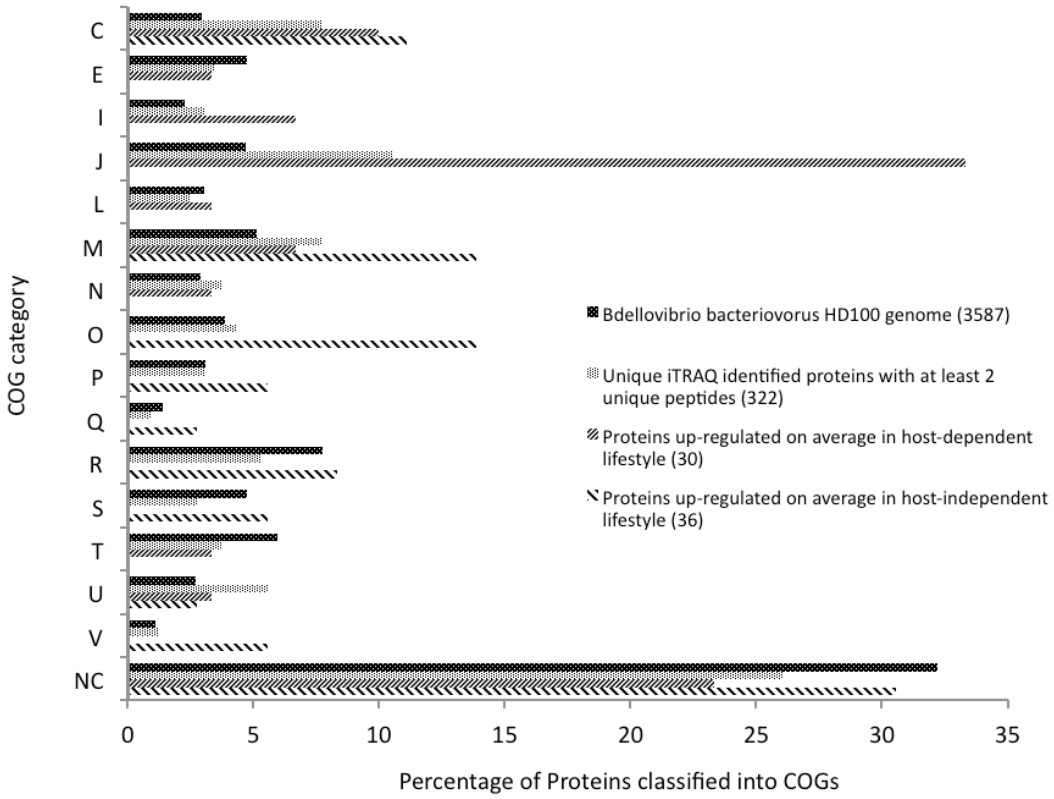

Percentage of proteins from set localized to subcellular fractions
70
60
50
40
30
20
10
0
Cytoplamic
Inner Membrane
Outer Membrane
Periplasmic
Unknown
Bdellovibrio bacteriovorus HD100 genome (3587)
Unique iTRAQ identified proteins with at least 2 unique peptides (322)
Proteins up-regulated on average in host-dependent lifestyle (30)
Proteins up-regulated on average in host-independent lifestyle (36)

Figure 1.4: Functional categorization of predicted, identified and differentially regulated proteins based on COG classification



[C] Energy production and conversion
[E] Amino acid transport and metabolism
[I] Lipid transport and metabolism
[J] Translation, ribosomal structure and biogenesis
[L] Replication, recombination and repair
[M] Cell wall/membrane/envelope biogenesis
[N] Cell motility
[O] Posttranslational modification, protein turnover, chaperones
[P] Inorganic ion transport and metabolism
[Q] Secondary metabolites biosynthesis, transport and catabolism
[R] General function prediction only
[S] Function unknown
[T] Signal transduction mechanisms
[U] Intracellular trafficking, secretion, and vesicular transport
[V] Defense mechanisms
[NC] No COG

Figure 1.5a: Relative expression levels of proteins belonging to the 'Cellular Processes and Signaling' functional category of COGs, in host-dependent versus host-independent lifestyles. These proteins occur in both biological replicates and have at least 2 unique peptides in each replicate with p-values of 0.05 or lower.

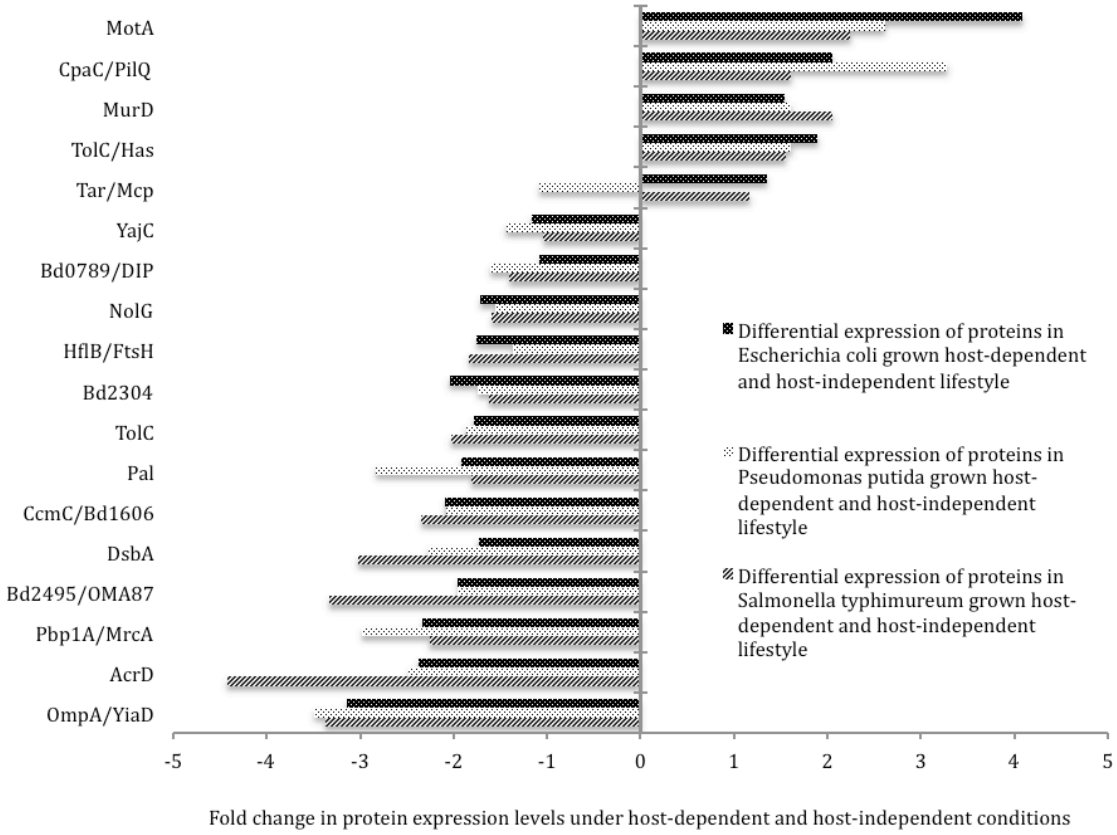

MotA
CpaC/PilQ
MurD
TolC/Has
Tar/Mcp
YajC
Bd0789/DIP
NolG
HflB/FtsH
Bd2304
TolC
Pal
CcmC/Bd1606
DsbA
Bd2495/OMA87
Pbp1A/MrcA
AcrD
OmpA/YiaD
-5
-4
-3
-2
-1
0
1
2
3
4
5
Differential expression of proteins in Escherichia coli grown host-dependent and host-independent lifestyle
Differential expression of proteins in Pseudomonas putida grown host-dependent and host-independent lifestyle
Differential expression of proteins in Salmonella typhimureum grown host-dependent and host-independent lifestyle
Fold change in protein expression levels under host-dependent and host-independent conditions

Figure 1.5b: Relative expression levels of proteins belonging to the 'Information Storage and Processing' functional category of COGs, in host-dependent versus host-independent lifestyles. These proteins occur in both biological replicates and have at least 2 unique peptides in each replicate with p-values of 0.05 or lower.

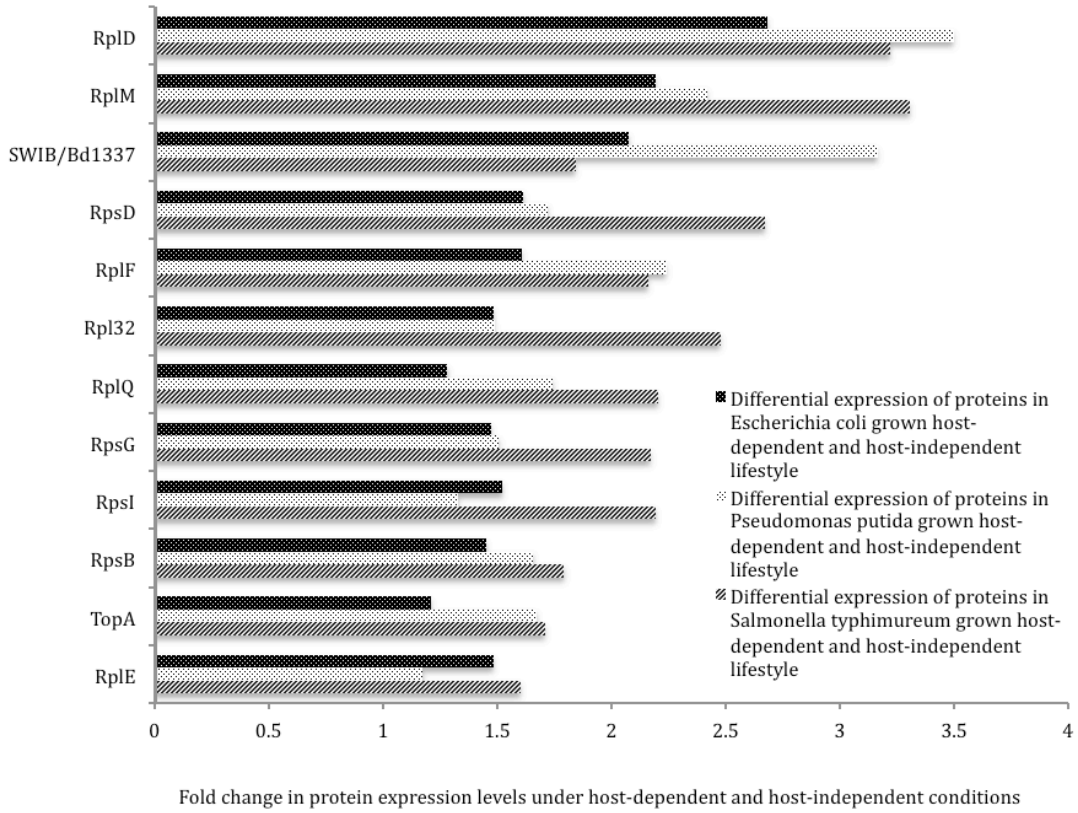

RplD
RplM
SWIB/Bd1337
RpsD
RplF
Rpl32
RplQ
RpsG
RpsI
RpsB
TopA
RplE
0
0.5
1
1.5
2
2.5
3
3.5
4
Differential expression of proteins in Escherichia coli grown host-dependent and host-independent lifestyle
Differential expression of proteins in Pseudomonas putida grown host-dependent and host-independent lifestyle
Differential expression of proteins in Salmonella typhimureum grown host-dependent and host-independent lifestyle
Fold change in protein expression levels under host-dependent and host-independent conditions

Figure 1.5c: Relative expression levels of proteins belonging to the 'Metabolism' functional category of COGs, in host-dependent versus host-independent lifestyles. These proteins occur in both biological replicates and have at least 2 unique peptides in each replicate with p-values of 0.05 or lower.

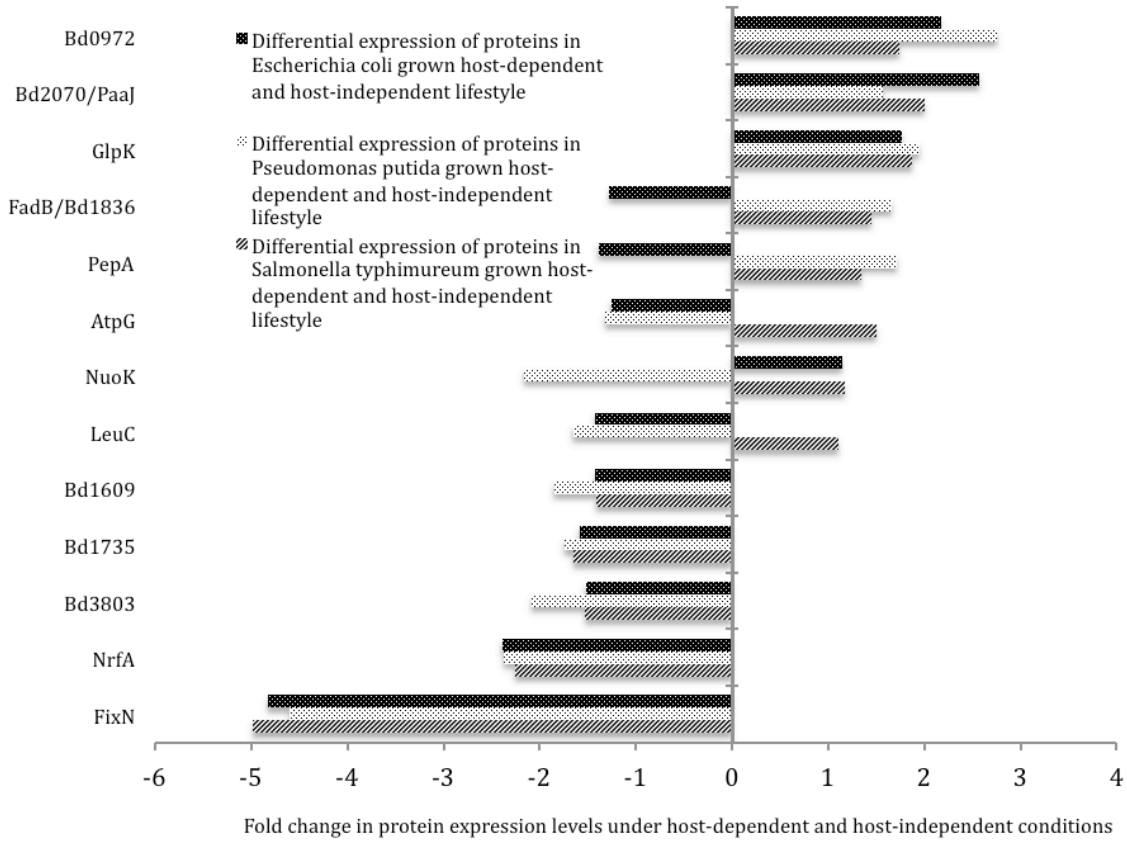

Bd0972
Bd2070/PaaJ
GlpK
FadB/Bd1836
PepA
AtpG
NuoK
LeuC
Bd1609
Bd1735
Bd3803
NrfA
FixN
Differential expression of proteins in Escherichia coli grown host-dependent and host-independent lifestyle
Differential expression of proteins in Pseudomonas putida grown host-dependent and host-independent lifestyle
Differential expression of proteins in Salmonella typhimureum grown host-dependent and host-independent lifestyle
-6
-5
-4
-3
-2
-1
0
1
2
3
4
Fold change in protein expression levels under host-dependent and host-independent conditions

Figure 1.5d: Relative expression levels of proteins belonging to the 'Poorly Characterized' functional category of COGs or which are not classified into COGs, in host-dependent versus host-independent lifestyles. These proteins occur in both biological replicates and have at least 2 unique peptides in each replicate with p-values of 0.05 or lower. Note the fold change for hypothetical protein Bd3054 is outside the range represented in this chart.

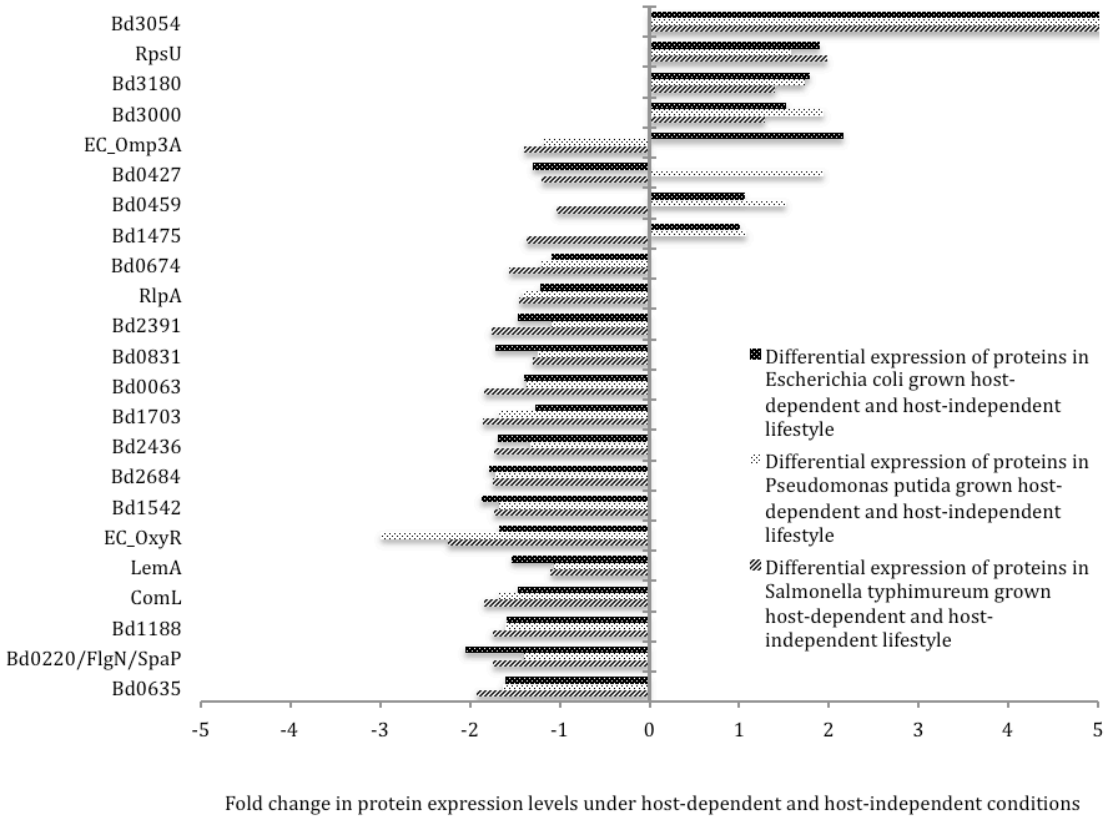

Bd3054
RpsU
Bd3180
Bd3000
EC_Omp3A
Bd0427
Bd0459
Bd1475
Bd0674
RlpA
Bd2391
Bd0831
Bd0063
Bd1703
Bd2436
Bd2684
Bd1542
EC_OxyR
LemA
ComL
Bd1188
Bd0220/FlgN/SpaP
Bd0635
-5
-4
-3
-2
-1
0
1
2
3
4
5
Differential expression of proteins in Escherichia coli grown host-dependent and host-independent lifestyle
Differential expression of proteins in Pseudomonas putida grown host-dependent and host-independent lifestyle
Differential expression of proteins in Salmonella typhimureum grown host-dependent and host-independent lifestyle
Fold change in protein expression levels under host-dependent and host-independent conditions

Table 1.1: List of abbreviations used in this chapter

1-DE	one dimensional electrophoresis
2-DE	two dimensional electrophoresis
2D-DIGE	two dimensional fluorescence difference gel electrophoresis
2D-LC	two dimensional liquid chromatography
ALS	acid labile surfactant
APEX	absolute protein expression
AQUA	absolute quantification of proteins
BAC	16-benzyldimethyl-n-hexadecylammonium chloride
BME	beta mercaptoethanol
BN	blue native
CBB	Coomassie Brilliant Blue
CFTR	cystic fibrosis transmembrane regulator
CI	confidence interval
cICAT	cleavable isotope coded affinity tagging
CN	clear native
CNBr	cyanogen bromide
COG	cluster of orthologous groups
CRDS	combined raw data set
CSF	cerebrospinal fluid
CTAB	cetyl trimethyl ammonium bromide
DIGE	difference gel electrophoresis
DM	n-dodecyl-b-D-maltoside
DNA	deoxyribonucleic acid
DOE-JGI	Department of Energy - Joint Genome Institute
DOX	doxycycline
DTT	dithiothreitol
DvH	*Desulfobibrio vulgaris* Hildenborough
EDTA	ethylenediaminetetraacetic acid
emPAI	exponentially modified protein abundance index
ESI	electrospray ionization
FPR	false positive rate
GAS	Group A *Streptococcus*
GISA	glycopeptide intermediate resistant *Staphylococcus aureus*
GRAVY	grand average of hydropathy values
HPLC	high performance liquid chromatography
hpPK	high pH and proteinase K
ICAT	isotope coded affinity tagging
ICPL	isotope coded protein labeling
IEC	ion exchange chromatography
IEF	isoelectric focusing
IMAC	immobilized metal affinity chromatography

IMG	Integrated Microbial Genomes
IMP	inner membrane protein
IP	immunoprecipitation
IPG	immobilized pH gradient
iTRAQ	isobaric tagging for relative and absolute quantification of proteins
LB	Luria Bertani
LDS	lithium dodecyl sulfate
LN2	liquid nitrogen
MAAH	microwave-assisted acid hydrolysis
MALDI	matrix-assisted laser desorption ionization
MEP	membrane embedded peptide
MPEC	mammary pathogenic *Escherichia coli*
mRNA	messenger ribonucleic acid
MRSA	methicillin resistant *Staphylococcus aureus*
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MudPIT	multidimensional protein identification technology
OMP	outer membrane protein
ORFs	open reading frames
PAGE	polyacrylamide gel electrophoresis
PAI	protein abundance index
PAP	protease accessible peptide
PBS	phosphate buffered saline
PEG	polyethylene glycol
pI	isoelectric point
PMF	peptide mass fingerprinting
PSI	photosystem I
PSII	photosystem II
qPCR	quantitative polymerase chain reaction
RBC	red blood cells
RNA	ribonucleic acid
RNI	reactive nitrogen intermediates
ROS	reactive oxygen species
RP	reverse phase
SCX	strong cation exchange
SDC	sodium deoxycholate
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SILAC	stable isotope labeling by amino acids in cell culture
SIRMS	stable isotope ratio mass spectrometry
SLS	sodium lauryl sarcosinate (Sarkosyl)
TBP	tributyl phosphine
TCA	trichloroacetic acid

TFA	trifluoroacetic acid
TFE	trifluorethanol
TMD	transmembrane domain
TMT	tandem mass tagging
tRNA	transfer ribonucleic acid
TX-100	Triton x-100
UV	ultra violet
VRE	vancomycin resistant *Enterococcus*
WGA	wheat germ agglutinin
WT	wild type

Table 1.2: Summary of recent iTRAQ-based quantitative studies focused on microbial membranes

Organism	Parameter investigated	Salient feature /approach	Results	Reference
Saccharomyces cerevisiae	Strains deficient in decay pathways xrn1Δ and upf1Δ compared with isogenic wild type strains	Introduced iTRAQ. Tandem LC separation using SCX followed by RP-HPLC and MS/MS analysis. Introduced the use of isobarically tagged synthetic internal standard for absolute quantification. No effort to enrich or study membrane proteins	1217 unique proteins, 172 differentially expressed	Ross *et al* 2004
Desulfovibrio vulgaris Hildenborough	Response to nitrate stress	Used standard iTRAQ methodology. No special effort to enrich for membrane proteins	737 proteins, 67 differentially regulated. At least 16 membrane proteins, possibly more.	Redding *et al* 2006
Staphylococcus aureus (GISA)	Compared intermediate glycopeptide resistant strains with isogenic susceptible strains at proteomic and transcriptomic levels	Fractionation of insoluble membranes. Separation of post-digestion peptides by IEF followed by RP-HPLC-MS.	Increased proteome coverage to 32% compared to previous attempts. 835 unique proteins (approx 20% membrane). 178 differentially expressed proteins. Showed divergence between mRNA and protein expression levels for several genes.	Scherl *et al* 2006

Bacillus subtilis	Effect of heat shock	Combined standard iTRAQ on soluble fractions with semi-gel (1-DE SDS + MS/MS) analysis of membrane fractions. Used 15% dodecyl maltoside for solubilization. Used carbonate wash.	Identified 453 proteins; 232 -268 of these predicted to be membrane proteins.	Wolff *et al* 2006, Wolff *et al* 2007
Mycobacterium avium subsp. paratuberculosis	Compared serially passed attenuated laboratory strain with fresh clinical isolate	Separation of soluble and membrane fractions. Enrichments of membranes. Standard iTRAQ methodology on each fraction.	874 proteins from cytoplasmic and membrane fractions. 550 proteins in membrane fractions, 385 unique to membrane fractions. 111 proteins differentially expressed, 72 of these predicted to be membrane proteins.	Radosevic *et al* 2007
Nostoc sp. PCC7120	Growth under nitrogen fixing and non-fixing conditions	No special effort made to enrich for membrane proteins	486 unique proteins, 122 differentially expressed. Membrane proteins underrepresented, but several photosynthetic aparatus membrane proteins detected 8.9% of identified proteins predicted to be membrane. Additional algorthms bring up membrane proteins to nearly 19%.	Stensjö *et al* 2007
Prochlorococcus marinus MED4	Investigated organism's response to varying light conditions	High (9M) urea in solubilization	184 unique proteins, out of which 24 - 35 membrane proteins by different algorithms. Membrane proteins over-represented.	Pandhal *et al* 2007

Bdellovibrio bacteriovorus	Compared Bdellovibrio protein expression in 3 different hosts and host independent life style	Density gradient centrifugation to purify and enrich membranes. Carbonate wash.	718 unique proteins. Between 36% to 47% membrane. 66 proteins differentially regulated. Between 17 - 24 membrane proteins downregulated in host dependent lifestyle and 5 upregulated.	Jagtap *et al* unpublished , this work
Mycobacterium ulcerans	Compared mycolactone deficient strain with wild type strain	Fractionation into supernatant, soluble and membrane fractions.	1074 unique proteins.481 proteins from membrane fractions. About 20 with 10 or more predicted TMs	Tafelmayer *et al* 2008
Halobacterium salinarum	Creation of comprehensive proteome atlas	Combination of multiple approaches - iTRAQ, ICAT and IP. Used fractionation and detergent solubilization.	1646 proteins, 188 membrane proteins.	Van *et al* 2008
Escherichia coli (mammary pathogenic)	Effect of growth in milk compared with growth in standard bacterial growth medium	Fractionation into soluble and membrane fractions.	1000 proteins, 336 in membrane fractions. Appoximately 30% of identified proteins differentially expressed.	Lippolis *et al* 2009
Leptospira interrogans	Effect of temperature upshift from 30°C to 37°C	Membranes extracted using Triton X-114 phase partitioning followed by CaCl2 addition, and then metanol/chloroform for detergent removal.	1026 proteins identified, 93 differentially expressed. 38 IMPs, 22 OMPs, 58 lipoproteins.	Lo *et al* 2009

Euhalothece sp. BAA001 & *Synechocystis* sp. PCC6803	Cross species protein identification using database from closely related Synechocystis	Proteomic analysis of an unsequenced organism using genome of closely related organism	207 unique proteins by iTRAQ, 39 of these predicted to be membrane. 21 proteins with changed expression levels	Pandhal *et al* 2009
Sphingopyxis alaskensis	Investigated the efects of twelve different conditions of varying light intensity and wavelength	Standard iTRAQ methodology. No special effort was made to enrich membrane proteins.	811 proteins identified. 62 IM and 23 OM. Potentially more of the unknown identified proteins could be membrane proteins.	Matallana-Surget *et al* 2009
Plasmodium falciparum	Effect of the antibiotic doxycycline	Used a combination of DIGE and iTRAQ. Soluble and membrane fractions analyzed in parallel. Membranes solubilized in CHAPS	422 proteins including 246 plasmodial proteins identified from soluble fractions; 308 proteins including 204 plasmodial proteins from membrane fractions. 64 proteins differentially expressed in both fractions by iTRAQ and DIGE, 8 of them predicted to be membrane.	Briolant *et al* 2010
Methanococcoides burtonii	Compared growth of psychrophile at increased temperature and also compared growth on methanol and trimethylamine as carbon sources.	Separate fractions for soluble, insoluble and secreted proteins. Insoluble fractions carbonate washed and solubilized with methanol.	698 unique proteins, 198 unique to membrane fractions. 166 of identified proteins differentially regulated.	Williams *et al* 2010 (2 studies)
Synechocystic sp PCC6803 thylakoid membranes	Studies effect of growth temperature at 38°C versus 25°C on thermotolerance of photosystem II	Used Triton X-100 enriched membrane fractions.	385 distinct proteins, 168 of these with 2 or more unique peptides. 48 proteins differentially expressed	Rowland *et al* 2010

Sulfolobus solfataricus P2	Effect of lowered growth temperature on thermophilic archaeon	Enrichment of membrane fractions by sucrose density gradient centrifugation. Delipidation with chloroform/methanol. Addition of sodium deoxycholate to aid proteolytic digestion. Additional digestion with trypsin + chymotrypsin. 8-plex iTRAQ.	395 proteins identified from membrane fractions, out of which 373 were predicted as membrane proteins. Approximately 20% of identified proteins differentially expressed.	Pham *et al* 2010, Zaparty *et al* 2010
Acinetobacter baumannii	Analysis of growth phases and understanding adaptation to oxidative and nitrosative stress	2-DE and iTRAQ on soluble and membrane enriched fractions.	31 differentially expressed proteins by iTRAQ and 76 by 2-DE, 13 of which were membrane.	Soares *et al* 2010
Lactococcus lactis	Compared engineered stains artificially epressing human CFTR with wild type.	4-plex and 8-plex iTRAQ on soluble and membrane fractions.	846 unique proteins, 163 predicted integral membrane proteins. Approximately 30% of identified proteins differentially expressed.	Steen *et al* 2010
Streptomyces coelicolor	Studied solid culture growth time course, and comparison of liquid cultures with solid cultures	Fractionation into soluble, membrane associated and membrane fractions. Used carbonate wash.	626 and 642 proteins identified in different studies from cytoplasmic and membrane fractions. ~80 proteins from membrane fractions were predicted to have 1 or more TM.	Manteca *et al* 2010 (2 studies)

Tannerella forsythia	Compared planktonic and biofilm grown cultures of periodontal pathogen	No special effort to enrich insoluble membranes. Standard iTRAQ protocol used.	348 proteins identified, 44 with differential expression. Many upregulated proteins are membrane proteins, ~18% are OMPs.	Pham *et al* 2010

CHAPTER II

ANALYSIS OF THE OUTER MEMBRANE PROTEOME OF *CAULOBACTER CRESCENTUS* BY TWO-DIMENSIONAL ELECTROPHORESIS AND MASS SPECTROMETRY

This chapter describes the work I carried out towards the creation of the standard *C. crescentus* outer membrane proteome. The work was published in its entirety in the journal *Proteomics* (2001, *1*, 705–720) with the title, "Analysis of the Outer Membrane Proteome of *Caulobacter crescentus* by Two-dimensional electrophoresis and Mass Spectrometry" by Nikhil D. Phadke, Mark P. Molloy, Stephanie A. Steinhoff, Peter J. Ulintz, Philip C. Andrews and Janine R. Maddock. During this study, I collected additional data and carried out additional analysis describing the conservation of cross species outer membrane protein expression patterns and the feasibility of cross species peptide mass fingerprinting in bacteria . These were published as part of an article in the journal *Electrophoresis* (2001, *22*, 1686–1696) titled "Two-dimensional electrophoresis and peptide mass-fingerprinting of bacterial outer membrane proteins" by Mark P. Molloy, Nikhil D. Phadke, Janine R. Maddock, and Philip C. Andrews. A large section of the work described in this chapter was carried out concurrently with the sequencing of the *Caulobacter crescentus* genome at the Institute for Genomic Research, Rockville, MD. Our contributions to the annotation effort were published as sections of the *Caulobacter crescentus* genome paper in the *Proceedings of the National Academy of Science* (2001,

98 (7), 4136–4141) under the title ‘Complete genome sequence of *Caulobacter crescentus*’ by W.C. Nierman *et al*.

Abstract

Caulobacter crescentus, a Gram negative α-purple bacterium that displays an invariant asymmetric cell division pattern, has become a pivotal model system for the study of bacterial development. Membrane proteins play key roles in cell cycle events, both as components of landmark morphological structures and as critical elements in regulation of the cell cycle. Recent advances made in our laboratories for the isolation and solubilization of bacterial membrane proteins prior to isoelectric focusing have allowed us to significantly improve the separation of outer membrane proteins by 2-D electrophoresis. In this work we describe the analysis of the outer membrane proteome of *Caulobacter crescentus*. Proteins were identified using 2-D gel electrophoresis and peptide mass fingerprinting by MALDI-TOF mass spectrometry. We identified 54 unique proteins out of which 41 were outer membrane proteins. Of the outer membrane proteins, 16 were identified as Ton-B dependent receptor proteins. These studies were executed simultaneously with the *Caulobacter* genome sequencing project and advantages and limitations of proteomic analysis of a non-annotated genome are discussed. Finally, differences in protein expression patterns between different cells grown in rich and minimal media are presented which demonstrate that many of the differences are due to differential expression of TonB-dependent receptor proteins.

Introduction

Caulobacter crescentus is an aquatic Gram negative α-Proteobacterium (α-purple bacterium) that is characterized by an asymmetric cell division. Every pre-divisional *C. crescentus* cell divides unequally to yield a motile swarmer cell and a sessile stalked cell. As part of its cell cycle, the swarmer cell undergoes polar morphogenesis resulting in the replacement of its flagellum by a stalk, a cylindrical extension of the cell envelope. Stalked cells mature to form pre-divisional cells which then undergo cell division. The cell cycle continues invariantly in this manner [1] until the cell ages and dies [2]. This dimorphic life cycle coupled with the relative simplicity of a prokaryotic system makes *C. crescentus* an excellent model for the study of development in biological systems.

An intriguing feature of the *C. crescentus* cell cycle is the morphogenesis of the swarmer cell into a stalked cell through loss of the polar flagellum and growth of the polar stalk. At a molecular level, this developmental change involves spatial and temporal regulation of several proteins including inner membrane proteins such as McpA [3, 4] and CckA [5]. Additional membrane proteins, primarily histidine protein kinases, have been implicated in the control of cell cycle events in *C. crescentus* [5-12]. It has long been hypothesized that a polar 'complex' of proteins termed an 'organizational center' may be involved in regulating cell cycle specific events in the cell [13]. To date, the identity of the components of the putative organizational center is unknown.

The presence and expression of polar complex components would be highly amenable to study through the use of proteomics. In fact, 2-D electrophoresis has been successfully used to demonstrate spatial protein expression patterns in *C. crescentus* cells

[14]. However, due to technical limitations, these older studies were confined to the visualization of protein expression patterns without identification of individual proteins. The combination of recent advances in protein identification using peptide mass fingerprinting through mass spectrometry, the increased sensitivity of Delayed Extraction Matrix Assisted Laser Desorption/Ionization Time of Flight Mass Spectrometry (DE-MALDI-TOF MS), and improved techniques for separation of membrane proteins in our laboratories, as well as the availability of whole genome sequence information, however, has made identification of these protein spots on a 2-D gel a reality.

In this report we describe the separation and identification of outer membrane proteins (OMPs) from *C. crescentus* by 2-DE and MALDI-TOF MS. Central to this work was access to the genome sequence of *C. crescentus* which was made available by the Institute for Genomic Research (Rockville, MD), and we describe the process we used to identify proteins from an unannotated genome. Our results demonstrate the utility of concurrent genome sequencing and proteome analysis. Finally, we demonstrate that the separation of outer membrane proteins by 2D electrophoresis is useful for comparing changes in membrane protein expression that accompany changes in cell physiology.

Materials and Methods

Growth of Caulobacter crescentus cultures

Late exponential phase cultures of *Caulobacter crescentus* CB15N were used to inoculate 1 L cultures of Peptone Yeast Extract (PYE) [15] or M2G minimal medium [16] at a 1:500 dilution. Cells were grown at 30°C with constant shaking at 250 rpm in 4L baffled flasks to an OD_{600} of 0.8. Cells were harvested by centrifugation at 7000 g, washed with 50 mM Tris-HCl pH 8.0, and either used immediately or stored at –80°C.

Sample preparation

The harvested cells were resuspended in 50 mM Tris-HCl (pH 8.0) with freshly added 1X protease inhibitor cocktail (Boehringer Mannheim, Germany), 1 mM PMSF and 0.5 mM EDTA. Cells were disrupted by two passages through a French pressure cell (AMINCO, Silver Spring, MD) at 16,000 psi. Cellular debris was removed by centrifugation (2X) at 7000 g for 10 min at 4°C followed by centrifugation at 17000 g for 25 min at 4°C. Total protein (30 mg) was incubated with ice-cold 100 mM Na_2CO_3 (sodium carbonate) in a total volume of 60 ml for 1 h with constant stirring [17]. Membrane vesicles were obtained by ultracentrifugation at 170,000 g for 1 h in a Beckman 50.2 Ti rotor at 4°C. Membrane vesicles were washed twice (50,000 rpm for 15 min at 4°C in a Sorvall RP80-AT rotor) (Kendro Laboratory Products, Newtown, CT) with 50 mM Tris-HCl pH 8.0.

2-Dimensional electrophoresis

Immobilized pH Gradients (IPGs) (18 cm; Amersham Pharmacia, Sweden) were used for the first dimension. Solubilization of membrane vesicles was carried out by sonicating 1 mg of total protein (pH 3-10 IPG) or 1.5 mg protein (pH 4-7 IPG) in 450 µl of rehydration buffer [7 M urea, 2 M thiourea, 2 mM tributyl phosphine (TBP), 0.5% (v/v) Biolytes 3-10 (Bio-Rad, CA, USA), and 40 mM Tris-base]. Detergents were added as indicated in Table 2.1 [18]. IPGs were rehydrated overnight in 450 µl of rehydration buffer. Isoelectric focusing was carried out for 60,000 Vh at a maximum of 6000 V using the Multiphor II system (Amersham-Pharmacia, Sweden). Equilibration buffer [6 M urea, 2% (w/v) SDS, 20% (v/v) glycerol, 0.15 M BisTris/0.1 M HCl] was freshly prepared. IPGs were incubated for 10 min with shaking at room temperature in equilibration buffer containing 0.5% (w/v) Dithiothreitol (DTT) followed by 10 min in equilibration buffer containing 1.5% (w/v) iodoacetamide. The IPGs were then embedded onto SDS-PAGE gels using 1% low melting agarose in 0.15 M BisTris/0.1 M HCl. Acrylamide (10%) with 0.675% (w/v) piperazine diacrylamide in 0.165 M Tris-Base/0.15 M HCl was used for the SDS-PAGE. The anode buffer was 0.384 M glycine/50 mM Tris-base, 0.1% (w/v) SDS and the cathode buffer contained 0.2 M taurine/25 mM Tris-base, 0.1% (w/v) SDS [17]. Second dimensional electrophoresis was performed using the PROTEAN II system (Bio-Rad, CA, USA). After an initial ramp up period of 2 h at 50 V, the gels were run at 100 V for 16 h at 4°C.

Staining of 2D gels and imaging

Coomassie staining was carried out as described previously [19]. Briefly, gels were stained with Colloidal Coomassie Brilliant Blue G-250 for 24 h, and destained with 10% (v/v) acetic acid. MS compatable silver staining was carried out according to the Vorum protocol (http://www.protana.com/PDF/ASMS/ExAbSilverstain.pdf). The gel images were digitized using the UMAX Power Look II scanner with the Magic Scan software (UMAX Technologies Inc., Fremont, CA). Image manipulation and analysis was performed using Adobe Photoshop 5.5.

Preparation of samples for MALDI and peptide mass fingerprinting

Spots were excised from the Coomassie blue or Silver stained gels and washed with a solution of 50% (v/v) acetonitrile and 100 mM ammonium bicarbonate. In-gel digestions were performed using 150 – 500 ng modified porcine trypsin (Promega, Madison, WI) in 10 µl 100 mM ammonium bicarbonate at 37°C overnight. The peptides were extracted using 60% (v/v) acetonitrile – 1% (v/v) trifluoroacetic acid (TFA), concentrated to dryness in a speed vac, and resuspended in 8 µl 3% (v/v) TFA. 0.8 µl of each sample was loaded onto a gold-plated MALDI plate (PerSeptive Biosystems, Framingham, MA) with an equal volume of 10 mg/ml (w/v) α-cyano-4-hydroxy cinnamic acid solution in 50% (v/v) acetonitrile and 1% (v/v) TFA. MALDI-TOF MS was performed on a Voyager-DE STR instrument (PerSeptive Biosystems, Framingham, MA) run in delayed extraction reflector mode with the following parameters: 1982 laser intensity, 25 kV accelerating voltage, 72% grid voltage, 0% guide wire voltage, 100 ns delay and a low mass gate of 500 Da. Spectra were calibrated with 842.5 Da and 2211.1 Da trypsin peaks using Data Explorer (PE Biosystems, Foster City CA). Monoisotopic

peptide masses were obtained using an in-house virtual instrument created in the Lab View graphical programming language (G. Rymar and P. Andrews, unpublished). The resulting peptide mass fingerprints were searched using a local copy of the program MS-Fit (part of the Protein Prospector package by P. Baker and K. Clauser http://prospector.ucsf.edu/) against a local *C. crescentus* database. A mass accuracy of 150 ppm was used. A maximum of 1 missed enzymatic cleavage, and modification of cysteines by carbamidomethylation plus possible modification by acrylamide were considered during the searches.

Database construction

C. crescentus genomic DNA sequences were made available prior to publication by the Institute for Genomic Research (TIGR, Rockville, MD). The putative open reading frames (ORFs) were identified using GLIMMER [20]. The initial functional assignments were made by TIGR using the Blast Extend Repraze (BER) pairwise alignment method. Protein translations of genes identified by GLIMMER were searched against a non-redundant amino acid database using BLASTP [21]. The open reading frames were further checked using an in-house modification of the third position GC profiling program FramePlot [22]. The presence of predicted signal peptides was determined using the program SignalP V2 [23] (**http://www.cbs.dtu.dk/services/SignalP-2.0/**). The theoretical molecular weights and pIs after removal of the signal peptides were calculated using the ProtParam tool at EXPASY (**http://expasy.cbr.nrc.ca/tools/protparam.html**). Additional functional annotation was performed with NCBI's COGNITOR program [24] (**http://www.ncbi.nlm.nih.gov/COG/cog99nitor.html**). The FA-Index program (part of

the Protein Prospector package) was used to create an MS-Fit searchable indexed database from a FASTA formatted translation of the putative ORFs.

Results

Identification of putative open reading frames (ORFs), preliminary analysis of the *C. crescentus* genome, and identification of putative membrane proteins

The *C. crescentus* genome was sequenced by TIGR (Rockville, MD) using a novel variation of shotgun sequencing [25, 26]. The size of the *Caulobacter* genome is approximately 4 Mb. Based on the number of ORFs from other genomes of similar size and predictions by GLIMMER, the *C. crescentus* genome is estimated to contain 3767 ORFs.

Because these studies preceded the annotation of the genome, we analyzed each of the putative ORFs using the Cluster of Orthologous Group (COG) system of protein classification [24]. The COG system is based on position specific score matrices of sequence similarity and offers a very powerful method for the functional classification of proteins on a genomic scale. We analyzed the *C. crescentus* genome using the COGNITOR tool [24] and placed each ORF into the highest scoring COG group. To simplify the analysis, proteins with multiple domains were only placed into the COG group with the highest sequence similarity matches. This system of classification is useful because it allows for genomic scale comparisons with other organisms that have been analyzed by this method, as well as providing relatively accurate functional classification of the ORFs.

In order to predict integral membrane proteins, we analyzed the predicted ORFs using a combination of the SOSUI [27, 28] and PSORT [29] algorithms. We predict 731 (19.4%) of the putative *C. crescentus* ORFs to be integral membrane proteins. These results are in agreement with statistical estimates which predict that approximately 20 %

of all the identified ORFs in bacteria, archaea and eukarya encode putative integral membrane proteins [30; http://cubic.bioc.columbia.edu/genomes/].

Although these programs accurately predict inner membrane proteins (IMPs), they are not efficient at predicting OMPs. Much of the difficulty in predicting OMPs is due to inherent structural differences between inner and outer membrane proteins. For example, while IMPs often have one or more hydrophobic membrane spanning domains and consensus domains like the core transmembrane amino acids [30], OMPs often span the membrane as a β-barrel structure with alternate hydrophobic and hydrophilic amino acids and have hydrophobicity values similar to those of cytoplasmic proteins. Nonetheless, OMPs can be predicted with reasonable confidence using a combination of criteria such as sequence similarities to bona fide OMPs, presence of signal peptides for translocation to the outer membrane, and presence of aromatic anchor residues at the C-terminus [17, 31]. To predict OMPs in *C. crescentus*, we used a combination of the PSORT algorithm [29], SignalP [23], COGNITOR [24], and PSI-BLAST [21]. From these analyses, we predict approximately 140 ORFs (4%) encoding outer membrane proteins (OMPs). While this number cannot be directly compared with the 86 OMPs (2%) predicted for *E. coli* by manual annotation (*E. coli* cell envelope protein data collection (http://www.cf.ac.uk/biosi/staff/ehrmann/head.html), it seems likely that the *C. crescentus* genome encodes a higher percentage of proteins that are involved in cell envelope biogenesis or that are components of the outer membrane as compared with *E. coli*. This fact is supported by COG analysis which places a total of 145 (5.2 %) *C. crescentus* ORFs into the COG group M category, that includes the cell envelope biogenesis and outer membrane proteins, as compared with 189 (4.4 %) in *E.coli*.

Separation of membrane proteins using 2-D Electrophoresis

Our initial attempts at solubilization of *C. crescentus* OMPs using conditions described for *E. coli* membranes [32] resulted in very poor resolution on 2-D gels (Fig 2.1A, 2.1B). The use of combined surfactants CHAPS and SB (3-10), which were highly effective in solubilizing some abundant *E. coli* OMPs, provided disappointing results with *C. crescentus* membrane preparations (data not shown). Therefore, we conducted a systematic evaluation of various surfactant, chaotrope, and reducing agent combinations for effective solubilization of these membranes (Table 1, Fig 1). The use of the chaotrope thiourea [17] in conjunction with urea dramatically increased the extracting power of the solubilizing solution (data not shown). Similarly, the use of the phosphine reducing agent TBP [33] instead of DTT resulted in superior resolution and decreased horizontal streaking (data not shown). Optimal results were obtained when the base rehydration solutions was 7 M urea, 2 M thiourea, 2 mM TBP, 40 mM Tris-base and 0.5% 3-10 Biolytes [except in the case of SB (3-10) which has a low tolerance to urea; 5 M urea was used]. Excellent resolution was seen when the amidosulfobetaine detergents ASB-14 and C8ϕ [34] or Triton X-100 were supplemented to this solution, although the resolution of spots and the recovery of proteins show slight variations between different conditions (Fig 2.1D, 2.1E, 2.1F). We observed highest resolution and recovery using a combination of 2 % C8ϕ + 1% Triton X-100 (Fig 2.2) and 1% ASB-14 + 1% Triton X-100 (data not shown). Using these conditions, we obtained highly reproducible patterns of membrane protein resolution by 2D-electrophoresis using broad range pH 3-10 IPGs (Fig 2.2, 2.3).

We routinely visualize approximately 200 distinct spots on gels stained with either Coomassie Blue (1 mg protein loaded) or silver nitrate (0.1 mg protein loaded). As expected, the silver stained gels revealed spots that cannot be seen with Coomassie Blue stained gels run in parallel. However, several of the protein spots that are reproducibly seen on Coomassie Blue stained gels fail to appear on the silver stained gels, and it may be that those proteins have markedly different affinities for Coomassie and silver stains and hence do not stain stoichiometrically. It might also be possible that some of these proteins are on the threshold of detection on the Coomassie stained gels and when reduced to 10% of their original concentration, cannot be detected by silver staining.

A large number of highly expressed proteins were observed between molecular weight of 65 kDa and 120 kDa, in agreement with earlier reports that the outer membrane of *C. crescentus* contains many proteins with a molecular weight greater than ~70 kDa [35]. Most of the spots on our gels were resolved between a pH of 4 and 8.2. Using theoretical pI versus molecular weight distributions of *Caulobacter* ORFs, we predict between 65-95 putative OMPs in our resolving range (pI range between 4 - 8.2, molecular weight > 10 kDa).

Identification of spots by MALDI-TOF MS

In this study, we initially examined the outer membrane proteome of mid-log *C. crescentus* cells grown in minimal medium in order to compare our data with existing *C. crescentus* physiological and cell-cycle studies which are typically performed with cells grown under similar conditions [1]. Figure 2.2 shows pH 3-10 and pH 4-7 gels run under identical conditions. The higher molecular weight spots between 68-120 kDa, which resolved in the pH 4.7 - 6.7 range, were identified from pH 4-7 IPGs which provide better

separation of the spots in that region (Fig. 2.2B). From these gels, we excised all the distinct visible spots (or spots from a duplicate gel run in parallel) and successfully identified 120 out of these spots by MALDI-TOF (Tables 2.2 and 2.3). In many cases, several spots represented a single putative ORF, reducing the number of total unique proteins identified to 54. Thus, we have demonstrated that these 54 proteins are bona fide gene products and are expressed in minimal media. Many weaker spots are also expressed (Fig. 2.2, spots not circled), but confirmed identities were not obtained in this study.

Of the 54 unique proteins identified, 41 were putative OMPs, 7 were inner membrane proteins, 3 were cytoplasmic, 2 were periplasmic, and 1 was secreted (see Table 2.2). Of the proteins identified, 89% were predicted to be membrane proteins, demonstrating the purity of our membrane protein enrichment. As expected, the vast majority (86%) of these were predicted to be OMPs. The introduction of the alkaline Na_2CO_3 wash, which effectively solubilizes cytoplasmic and periplasmic proteins and reduces these protein contaminants in the membrane fractions, likely contributed to the efficiency of our membrane fractionation [17].

Verification of annotation / gene-products

During the analysis phase of this study we noted that most of the identified proteins migrated according to their predicted pIs and molecular weights, confirming that these genes are correctly annotated and are expressed. However, we also observed a few proteins that migrated differently on the 2D gels than predicted, both before and after removal of putative signal peptides. While some of these discrepancies could be

attributed to potential post-translational modifications of the proteins, it became apparent in other cases that the putative ORFs automatically identified by GLIMMER might have incorrect start or end sites. An interesting example is seen with spots 113-120 (Fig 2.2B). These spots migrate in the molecular weight range of 100 kDa and a pI range of 4.8 – 5.2, and provide strong MALDI signals. A database query matches each of these spots to two putative ORFs with molecular weights of 60 kDa and 32 kDa respectively. Peptide coverage over both ORFs is uniformly high as are the MOWSE scores. Significantly, the two ORFs are adjacent to each other and their combined masses plus the mass of the intergenic region would result in a protein the size of the identified spots. Furthermore, ORF00440 and ORF00441 are homologous to the C- and N-terminal regions of the outer membrane iron receptor family of proteins, respectively. Thus, it appears that ORFs 0440 and 0441 are mis-annotated and represent a single large ORF that when translated, produces a protein of 100 kDa.

Comparison of membrane protein gels from exponentially growing cells in minimal and rich medium.

Because *C. crescentus* is a free-living oligotroph, its nutrition needs are relatively less stringent than those of the enteric *E. coli*. For routine studies, *C. crescentus* cells are grown in the relatively rich medium PYE (less enriched for nutrients than the Luria-Bertani medium, which is commonly used for *E. coli*) whereas for cell cycle studies, the cells are typically grown in minimal medium, M2G [1]. To address the consequences of growing *C. crescentus* cells in rich vs. minimal media, we compared the OMP profiles of cells grown under these two conditions (Fig 2.3). From multiple independent sample preparations we observed that the basic pattern of OMP expression is similar under these

two growth conditions. However there are significant highly reproducible increases in spot abundance in the high molecular weight range (65 kDa and 90 kDa) under the minimal media growth conditions. The identification of the protein spots that appear to be overexpressed, by visual examination, in minimal medium is shown in Table 2.5. Interestingly, COG analysis places most of these proteins into a single COG group, outer membrane receptor proteins that are often involved in iron transport. All these proteins belong to the family of TonB-dependent outer membrane receptors, which utilize energy from the periplasm spanning inner membrane protein TonB to translocate nutrients across the outer membrane [36].

Discussion

The 4 Mb genome of *C. crescentus* tentatively contains 3767 putative ORFs out of which 731 (19.4%) are predicted to be integral membrane proteins (Table 3). In *E. coli*, there are 86 predicted OMPs by manual annotation whereas in *C. crescentus*, we predict 140 (3.7%) OMPs by automatic annotation. Comparisons between *E. coli* and *C. crescentus* OMPs reveal that the bulk of *E. coli* OMP groups, such as the hydrolases (e.g. MltABC), the transport and receptor proteins (e.g. OmpA, BtuB, FadL, FecA, FepA), lipoproteins (e.g. Blc, Pal, VacJ), flagellar proteins (e.g. FlgH) and other hypothetical proteins (e.g. YccZ, NlpA, YjcP), have representative homologs in *C. crescentus*. *C. crescentus* does not possess the proteins involved in the export and assembly of fimbrae (i.e. CssD, PapC, FasD, and FimD). Also absent are homologs of the porins LamB (maltoporin), OmpX, OmpG, OmpF , OmpC, PhoE and Tsx although this is not surprising since porins with widely divergent primary sequences can form similar tertiary 3-D structures [37].

In addition to these conserved proteins, *C. crescentus* has several other predicted OMPs. Nearly 70 of the predicted OMPs belong to either the TonB-dependent receptor group of proteins or contained sequence signatures for OmpA family proteins. Of these, at least 60 contain the TonB box, three times as many as are present in *E. coli.* TonB-dependent receptors and related proteins are involved in the uptake of iron and heme complexes into Gram negative cells [36, 38]. A few are known to be involved in the uptake of Vitamin B12 (i.e. BtuB*)* and colicins, and others function as phage receptors. It is highly improbable that *C. crescentus* would require such a large number of outer membrane receptors for transport of iron and heme. In fact, initial inspection of the

genome reveals that only 19 of these TonB receptor proteins are organized in the classical operon structure (or variations thereof) characteristic of outer membrane receptors involved in iron metabolism. More likely, *C. crescentus*, as a free-living organism, might have an elaborate cellular machinery to import nutrients from its highly variable external environment. Many of the TonB-dependent receptors in *C. crescentus* could play a role in the transport of other nutrients and macromolecules into the cell, possibly by mechanisms that have not been elucidated in other systems. Interestingly, despite the presence of a large number of TonB-dependent receptors, the *C. crescentus* genome appears to contain only one copy each of the periplasm-spanning TonB protein, and the associated energy transducing proteins ExbBD and the functionally overlapping TolQR proteins.

Recent improvements in the solubilization and separation of membrane fractions from Gram negative bacteria allow for excellent resolution of the OMPs [17] (M. Molloy and N. Phadke, unpublished), although the resolving of integral inner membrane proteins on IPGs remains problematic perhaps due to poor solubilization and precipitation of the hydrophobic inner membrane proteins during IEF [18]. The pattern of spots on our gels is similar to the pattern obtained on theoretical 2D gels, created by plotting the pIs of the predicted *C. crescentus* OMPs vs. molecular weights on a logarithmic scale (data not shown). The many spots in the 70 – 120 kDa range confirms earlier observations of a large number of high molecular weight proteins in the outer membrane of *C. crescentus* [35], several of which are represented by large 'trains' of spots that vary by pI. This pattern is indicative of multiple forms of the same protein and may be caused by post-translational processing or experimental artifacts such as deamidation [39, 40]. Other

biologically significant post-translational modifications such as phosphorylations, and glycosylations are also possible, but unlikely in prokaryotes. Further investigation using MS/MS will be required to elucidate the nature of these modifications.

Although we separated our membrane proteins over a pH range of 3-10, few protein spots were detected above pH 8 even though approximately 40 OMPs are predicted to lie within pH 8-10 in the molecular weight that we examined. We envision two possible reasons for this low recovery, (i) problematic recovery or resolution of alkaline proteins by IEF and (ii) low expression of these proteins under our current growth conditions. Based on our previous experiences with IPGs, the former appears to be more likely. The resolution of these alkaline membrane proteins may improve with the development of improved alkaline IEF gels.

We identified 54 unique proteins in this study and as expected, the majority of these were OMPs (76%). The presence of a larger number of OMPs in *C. crescentus* as compared with *E.coli* might signify a greater specificity of function in this organism. *C. crescentus* might have a larger number of outer membrane proteins which are expressed only under certain conditions as opposed to *E. coli* which has several OMPs with generalized functions expressed under a multitude of conditions. Based on a functional distribution of the identified proteins (Table 2.4), the largest group of identified OMPs belongs to the class of TonB-dependent receptors. Not surprisingly, many of the TonB receptors were upregulated in cells grown under nutrient limiting conditions (minimal media) suggesting a role for these receptors in nutrient uptake. However, only 16 of the 60 predicted TonB receptors were identified in this study. Although some would not be resolved on our gels, others may not be expressed at detectable levels under the

conditions examined. Clearly a major challenge ahead is to elucidate the functional role of all of the TonB receptors.

This study was initiated well before the completion of the *C. crescentus* genome although identification of protein spots by PMF requires comparing peptide masses with masses of predicted peptides derived from predicted ORFs. Surprisingly, we were able to obtain significant identities even during the early stages of the sequencing project. To do this, the entire set of contigs were randomly concatenated and the sequence analyzed for ORFs using GLIMMER. With each genome release, new ORF assignments were made and additional protein identifications were realized.

One of the challenges in our study was to make predictions for both likelihood of membrane association and putative gene or functional assignment prior to annotation of the genome. We solved these problems by making predictions of cellular location based on various membrane prediction programs and database homology searches. Membrane predictions were initially done automatically and then confirmed for each relevant ORF. Functional assignments of all of the genes were made by placing each ORF into the top-hitting COG category. Although this approach results in classifying bimodal proteins incorrectly, we have found that our functional assignment based on COG category is very similar to the functional assignments made by TIGR.

Identification of a protein through proteomics is concrete evidence that a predicted gene encodes a bona fide protein product. In this study we have demonstrated that 54 of the predicted *C. crescentus* genes are indeed expressed. In addition, we have used our proteome data to confirm or correct ORF assignments as described in the case of ORF00440/0441. Several additional proteins migrate aberrantly and are under

investigation. Finally, with the outer membrane proteome in hand, we are well poised to examine the changes in proteome profiles of *C. crescentus* cells throughout the cell cycle and under various physiological conditions. These studies will aid in our understanding of *C. crescentus* lifestyle as well as in uncovering the role of these proteins.

References

[1] Poindexter, J. S., *Microbiol. Rev.* 1981, 45**,** 123-179.

[2] Pennisi, E., *Science* 2000, 289**,** 1131-1133.

[3] Alley, M. R., Maddock, J. R. & Shapiro, L., *Genes Dev.* 1992, 6**,** 825-836.

[4] Maddock, J. R., Alley, M. R. & Shapiro, L., *J. Bacteriol.*1993, 175**,** 7125-7129.

[5] Jacobs, C., Domian, I. J., Maddock, J. R. & Shapiro, L., *Cell* 1999, 97**,** 111-120.

[6] Ohta, N., Lane, T., Ninfa, E. G., Sommer, J. M. & Newton, A., *Proc. Natl. Acad. Sci. U S A* 1992, 89**,** 10297-10301.

[7] Wang, S. P., Sharma, P. L., Schoenlein, P. V. & Ely, B., *Proc. Natl. Acad. Sci. U S A* 1993, 90**,** 630-634.

[8] Hecht, G. B., Lane, T., Ohta, N., Sommer, J. M. & Newton, A., *EMBO J.* 1995, 14**,** 3915-3924.

[9] Domian, I. J., Quon, K. C. & Shapiro, L., *Cell* 1997, 90**,** 415-424.

[10] Domian, I. J., Quon, K. C. & Shapiro, L., *Curr. Opin. Genet. Dev.* 1996, 6**,** 538-544.

[11] Wu, J., Ohta, N. & Newton, A., *Proc. Natl. Acad. Sci. U S A* 1998, 95**,** 1443-1448.

[12] Wu, J., Ohta, N., Zhao, J. L. & Newton, A., *Proc. Natl. Acad. Sci. U S A* 1999, 96**,** 13068-13073.

[13] Huguenel, E. D. & Newton, A., *Differentiation* 1982, 21**,** 71-78.

[14] Huguenel, E. & Newton, A., *Proc. Natl. Acad. Sci. U S A* 1984, 81**,** 3409-3413.

[15] Poindexter, J. S., *Bacteriol. Rev.* 1964, 28**,** 231-295.

[16] Ely, B., *Methods Enzymol.* 1991, 204**,** 372-384.

[17] Molloy, M. P., Herbert, B. R., Slade, M. B., Rabilloud, T., Nouwens, A. S., Williams, K. L. & Gooley, A. A., *Eur. J. Biochem.* 2000, 267**,** 2871-2881.

[18] Molloy, M. P., *Anal. Biochem.* 2000, 280**,** 1-10.

[19] Molloy, M. P., Herbert, B. R., Williams, K. L. & Gooley, A. A., *Electrophoresis* 1999, 20**,** 701-704.

[20] Salzberg, S. L., Delcher, A. L., Kasif, S. & White, O., *Nucleic Acids Res.* 1998, 26**,** 544-548.

[21] Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J., *Nucleic Acids Res.* 1997, 25**,** 3389-3402.

[22] Ishikawa, J. & Hotta, K., *FEMS Microbiol. Lett.* 1999, 174**,** 251-253.

[23] Nielsen, H., Brunak, S. & von Heijne, G., *Protein Eng.* 1999, 12**,** 3-9.

[24] Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V., *Nucleic Acids Res.* 2000, 28**,** 33-36.

[25] Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M. & et al., *Science* 1995, 269**,** 496-512.

[26] Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., Chen, L. X., Brandon, R. C., Rogers, Y. H., Blazej, R. G., Champe, M., Pfeiffer, B. D., Wan, K. H., Doyle, C., Baxter, E. G., Helt, G., Nelson, C. R., Gabor Miklos, G. L., Abril, J. F., Agbayani, A., An, H. J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R. M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E. M., Beeson, K. Y., Benos, P. V., Berman, B. P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M. R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K. C., Busam, D. A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J. M., Cawley, S., Dahlke, C., Davenport, L. B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A. D., Dew, I., Dietz, S. M., Dodson, K., Doup, L. E., Downes, M., Dugan-Rocha, S., Dunkov, B. C., Dunn, P., Durbin, K. J., Evangelista, C. C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A. E., Garg, N. S., Gelbart, W. M., Glasser, K., Glodek, A., Gong, F., Gorrell, J. H., Gu, Z., Guan, P., Harris, M., Harris, N. L., Harvey, D., Heiman, T. J., Hernandez, J. R., Houck, J., Hostin, D., Houston, K. A., Howland, T. J., Wei, M. H., Ibegwam, C., et al., *Science* 2000, 287**,** 2185-2195.

[27] Hirokawa, T., Boon-Chieng, S. & Mitaku, S., *Bioinformatics* 1998, 14**,** 378-379.

[28] Mitaku, S., Ono, M., Hirokawa, T., Boon-Chieng, S. & Sonoyama, M., *Biophys. Chem.* 1999, 82**,** 165-171.

[29] Nakai, K. & Horton, P., *Trends Biochem. Sci.* 1999, 24**,** 34-36.

[30] Santoni, V., Molloy, M. & Rabilloud, T., *Electrophoresis* 2000, 21**,** 1054-1070.

[31] Nakai, K., *Adv. Protein Chem.* 2000, 54**,** 277-344.

[32] Molloy, M. P., Herbert, B. R., Walsh, B. J., Tyler, M. I., Traini, M., Sanchez, J. C., Hochstrasser, D. F., Williams, K. L. & Gooley, A. A., *Electrophoresis* 1998, 19**,** 837-844.

[33] Herbert, B. R., Molloy, M. P., Gooley, A. A., Walsh, B. J., Bryson, W. G. & Williams, K. L., *Electrophoresis* 1998, 19**,** 845-851.

[34] Chevallet, M., Santoni, V., Poinas, A., Rouquie, D., Fuchs, A., Kieffer, S., Rossignol, M., Lunardi, J., Garin, J. & Rabilloud, T., *Electrophoresis* 1998, 19**,** 1901-1909.

[35] Agabian, N. & Unger, B., *J. Bacteriol.* 1978, 133**,** 987-994.

[36] Sigel, A., Sigel, H. (Eds), *Metal Ions in Biological Systems. Vol 35 Iron Transport and Storage in Microorganisms, Plants, and Animals,* Dekker, New York 1978, pp. 67-145.

[37] Paulsen, I. T., Nguyen, L., Sliwinski, M. K., Rabus, R. & Saier, M. H., Jr., *J. Mol. Biol.* 2000, 301**,** 75-100.

[38] Sigel, A., Sigel, H. (Eds), *Metal Ions in Biological Systems. Vol 35 Iron Transport and Storage in Microorganisms, Plants, and Animals,* Dekker, New York 1978, pp. 355-401.

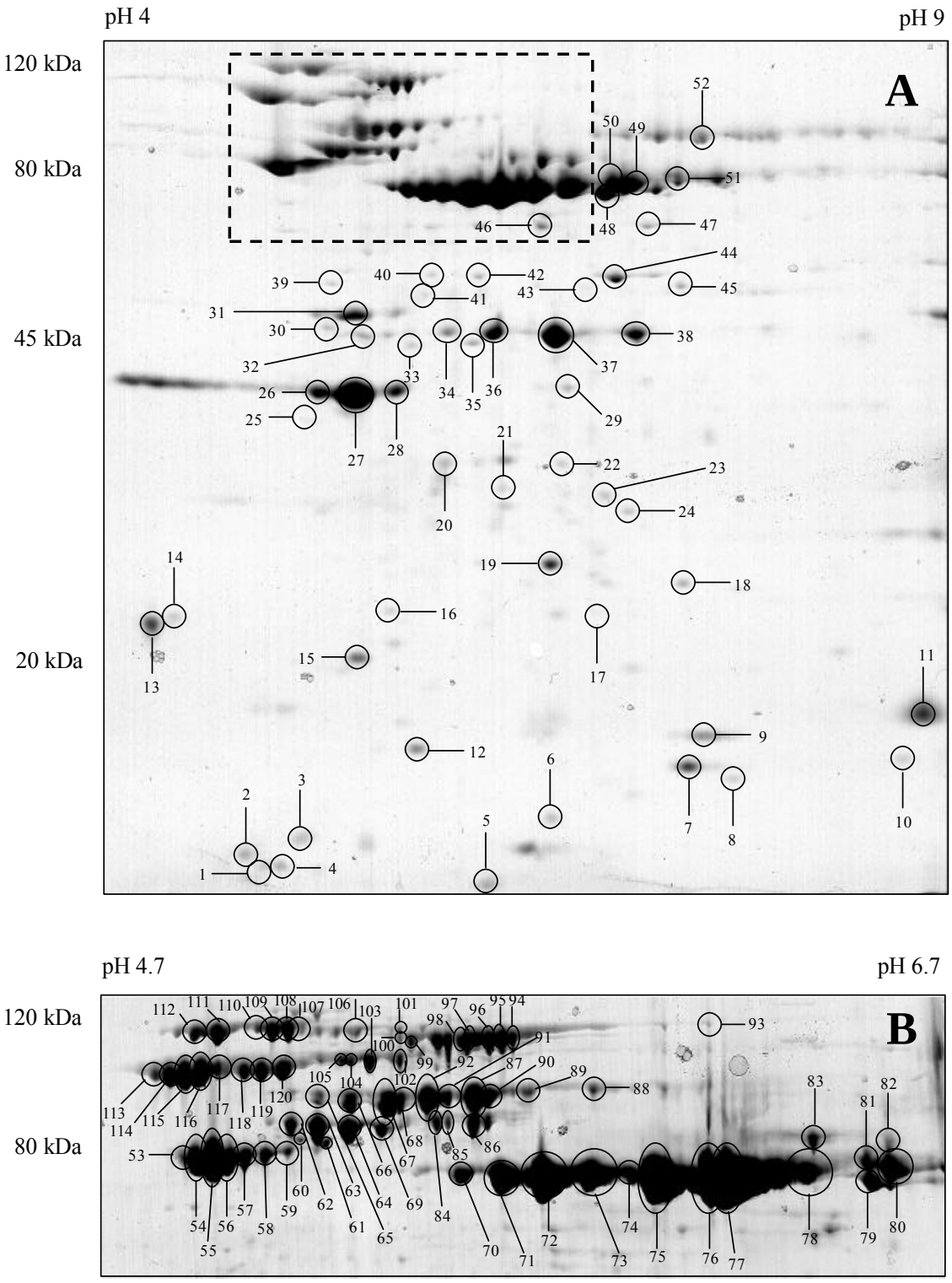
[39] Gianazza, E., *J. Chromatogr. A* 1995, 705**,** 67-87.

[40] Sarioglu, H., Lottspeich, F., Walk, T., Jung, G. & Eckerskorn, C., *Electrophoresis* 2000, 21**,** 2209-2218.

Figure 2.1: Membrane proteins obtained from late exponential growth phase *C. crescentus* cells in PYE separated by 2-D electrophoresis using pH 3–10 IPGs and 10% SDS-PAGE. Samples were solubilized with (A) 5 M urea, 2 M thiourea and 2 mM TBP with 2% SB (3–10) or 7 M urea, 2 M thiourea and 2 mM TBP with detergents, (B) 3% CHAPS, (C) 2% n-dodecylmaltoside, (D) 1% ASB-14, (E) 2% C8f, and (F) 1% Triton X-100 respectively. Gels were stained with Coomassie Blue G-250. Abbreviations are as defined in Table 2.1.

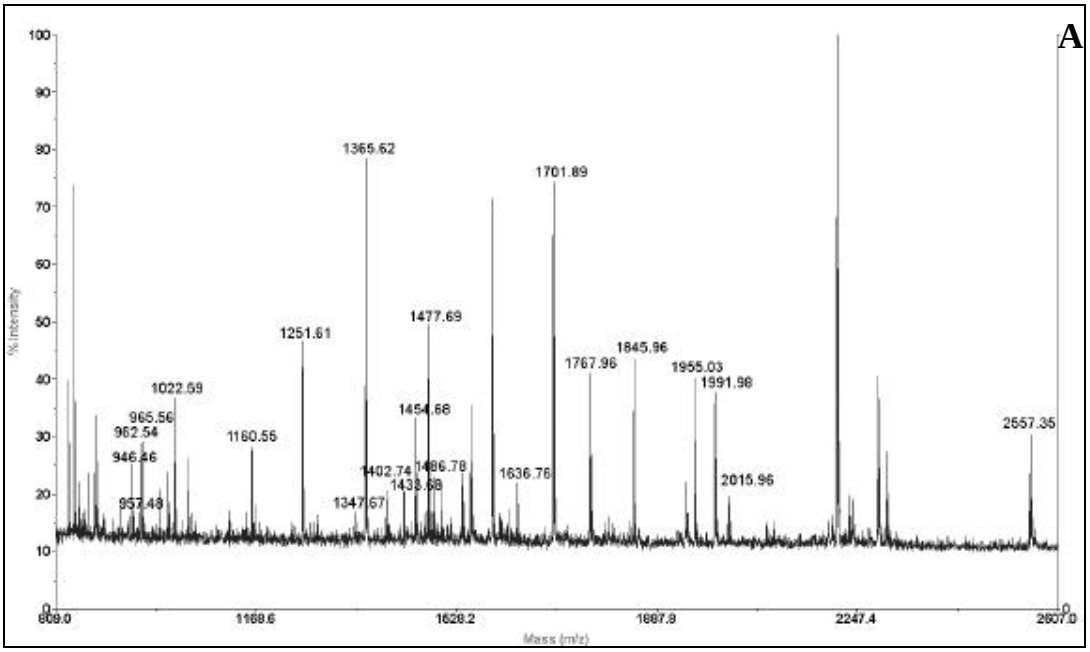
A
B
C
D
E
F
pH 3
pH 10
pH 3
pH 10
pH 3
pH 10

Figure 2.2: Identification of protein spots by peptide mass fingerprinting. Coomassie stained (A) pH 3– 10, and (B) pH 4–7 gel of exponentially growing C. crescentus cells in M2G. An expanded separation of the clustered high molecular range proteins in the boxed region shown in panel A, was obtained by separation on a pH 4–7 gel (B). The protein identifications of the circled spots are in Table 2.3.


pH 4
pH 9
120 kDa
80 kDa
45 kDa
20 kDa
A
pH 4.7
pH 6.7
120 kDa
80 kDa
B

..... Figure 2.3: Peptide mass fingerprint of a single spot, spot No. 70, from Fig. 2.2B (A). Monoisotopic pep- tide masses (matching masses are indicated) were obtained and a search was carried out against the *C. crescentus* database. Identification of the spot as a putative ferri- chrome iron receptor along with the score and coverage is detailed in Table 2.2. The peptide fragments matching the protein sequence are highlighted in bold (B).

A



B

```
1          11         21         31         41         51         61         71
MTSKALLFAL CAGVSQFAFA DLAVAQDKTD EVEGVVVTGS RAASASINGV VIDPMRLPQS VRVLDEALIV DTGVTNLSDL
81         91         101        111        121        131        141        151
FDFAGGMARQ NSFGGAWDAY AIRGFSGDIN QGPDLLVNRF TANRGFNARR DVATVERFQV LKGPASALSG KGEPGGSINI
161        171        181        191        201        211        221        231
VTKAPTETAQ GSGELSYGSF DAKRIMGDLS GPLGGGVSAR MIAVYQDTDG WRDHVGSDRL LLAPSLAWTP SDDLRLLYQL
241        251        261        271        281        291        301        311
EANTVHFVHD RGLVAVAGNG KALPRERFLG EPNDGDITQK TLQHQLTTTY NFSPSVAVEA GVQYRDGSFR GQSTHNGALI
321        331        341        351        361        371        381        391
GTQLRRQLRI HDYTWDDLSG RIEVSFDGKL GGLEHQLRAG ADAFTYEQHR IFYRFNPTAA TPYAIDILNP VYGQAKPVAP
401        411        421        431        441        451        461        471
LNQNVLEELR GESLYLQDLV TLNSQFTLLV GVRQDWIRQT NTNYRNNTVT RQSPSQASPR AALTWAPNEA FSAYVSWGRS
481        491        501        511        521        531        541        551
FRYNQGSDAV GGAFPPEKGE AWEAGVKWDL AGRLTGTASL FRIDKENILV NDPANSGFFI PVGAARSQGF EAETNLRLPK
561        571        581        591        601        611        621        631
GITATAVYAY TDTEITRDTR TNMIGSSLSN VPKHSGAVYA NWRSDGDAPG SVTLGGGVVY VGERAGDDVN TGFKLPDYVT
641        651        661        671        681        691
VRANLAYNVS KAVSLHLDVE NLFDTYYLES SYSNVWITPG APRTITGRLR VSF
```

The matched peptides cover **45%** (316/693 AA's) of the protein.

Figure 2.4: 2-D electrophoresis of membrane vesicles from cells grown in (A) M2G and (B) PYE. Gels were stained by MS compatible silver stain. Identified protein spots showing visually different expression patterns are boxed (new) or circled (shown in Fig. 2.2). A summary of the identified proteins putatively enriched in minimal media is in Table 2.4 while more detailed information of the spot identities is in Table 2.3.

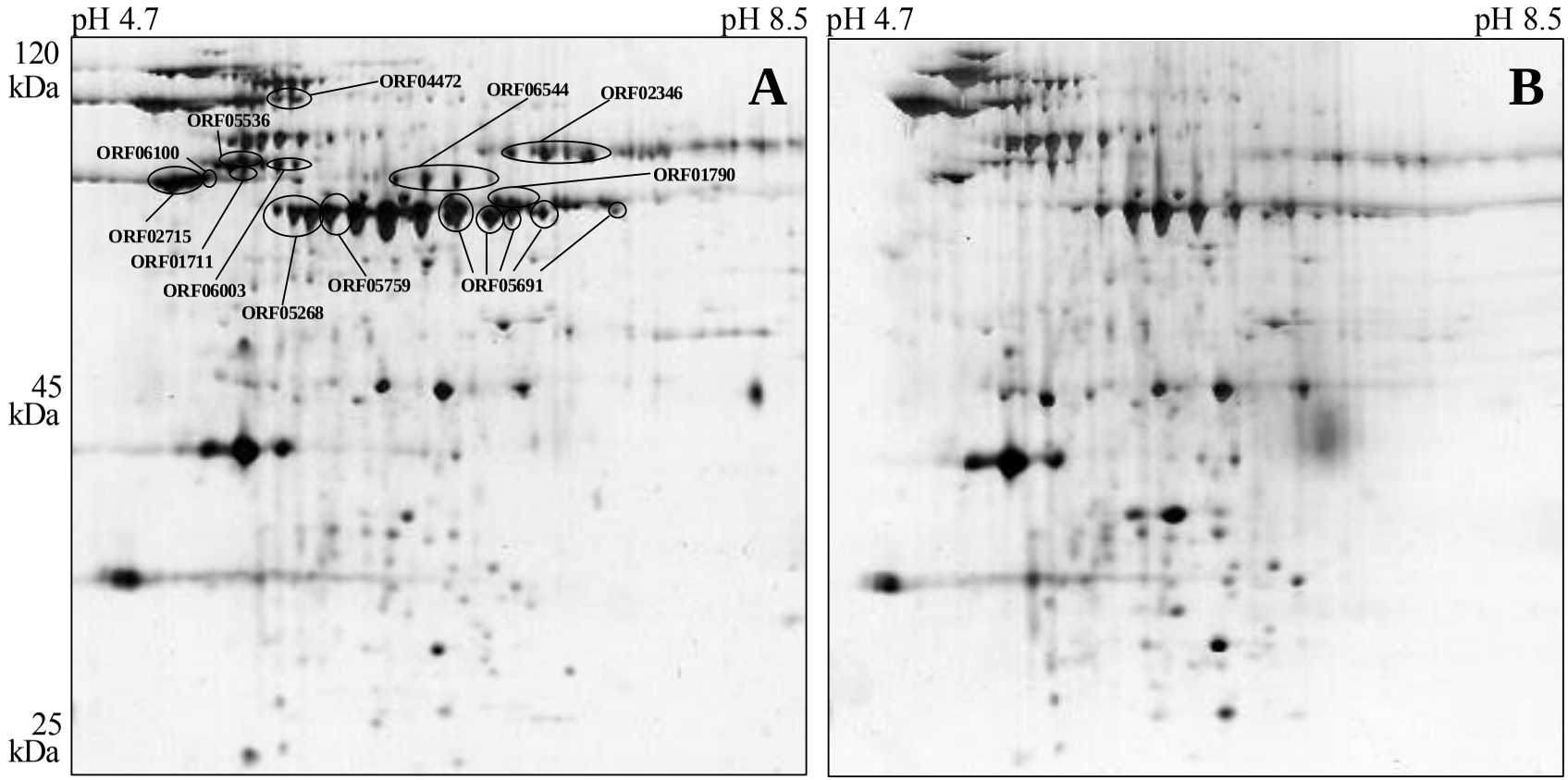

pH 4.7
pH 8.5
120 kDa
45 kDa
25 kDa
A
ORF04472
ORF06544
ORF02346
ORF05536
ORF06100
ORF01790
ORF02715
ORF01711
ORF06003
ORF05268
ORF05759
ORF05691
pH 4.7
pH 8.5
B

Table 2.1: Conditions used for *C. crescentus* membrane protein solubilization

Detergents used [a]	Gel images [b]
2% SB (3-10) [c]	Very poor
2% CHAPS + 2% SB (3-10) [c]	Poor
3% CHAPS [d]	Poor
3% CHAPS + 1% HFIP [d]	Fair
2% DM [d]	Fair
1% Triton X-100 [d]	Good
1% ASB-14 [d]	Good
2% C8ϕ [d]	Good
1% ASB-14 + 1% Triton X100 [d]	Very good
2% C8ϕ + 1% Triton X-100 [d]	Very good

[a] Detergents and solvents used decyldimethlammoniopropanesulfonate [SB (3-10)], 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfate (CHAPS), 3,3,3,3',3',3'-hexafluoro-2-propanol (HFIP), n-Dodecylmaltoside (DM), Triton X-100 (TX100), tetradecanoylamidopropyldimethylammoniopropanesulfonate (ASB-14) and 4-octyl benzoylamidopropyldimethylammonio-propanesulfonate (C8ϕ)

[b] Relative qualitative assessment of spot separation and resolution

[c] Rehydration solutions were in 5 M urea, 2 M thiourea, 2 mM TBP, 0.5 % 3-10 Biolytes and 40 mM Tris-base due to incompatibility of the SB 3-10 detergent with the chaotrope

[d] Rehydration solutions were in 7 M urea, 2 M thiourea, 2 mM TBP, 0.5 % 3-10 Biolytes and 40 mM Tris-base

Table 2.2: Identification of *C. crescentus* membrane proteins by peptide mass fingerprinting.

Spot no.	ORF no.	MOWSE	Peptides matched	Coverage %	Protein description	Predicted MW/pI	Predicted localization
1	ORF02943	8.31E+03	6	44	biopolymer transport exbd protein [a]	11115.8/4.92	Inner membrane
2	ORF03027	4.20E+02	4	32	FoF1-type ATP synthase B subunit [b]	13103.8/5.05 [c]	Inner membrane
3	ORF04095	1.57E+03	4	41	MarR family, putative [a]	16453.1/6.31	cytoplasm
4	ORF02943	1.67E+03	5	34	biopolymer transport exbd protein [a]	11403.1/4.92 [c]	Inner membrane
5	ORF08037	3.94E+03	5	46	hypothetical protein [a]	11732.2/6.79 [c]	Outer membrane
6	ORF02264	3.72E+03	4	37	hypothetical protein [a]	15777.6/6.95 [c]	Outer membrane
7	ORF02535	3.69E+04	7	40	hypothetical protein [a]	20040.2/7.85 [c]	Outer membrane
8	ORF00985	1.73E+05	7	48	hypothetical protein [a]	19713.1/7.16 [c]	Outer membrane
9	ORF03662	2.57E+02	4	39	conserved hypothetical protein [a]	23177.6/9.12	Outer membrane
10	ORF02535	3.29E+04	6	38	hypothetical protein [a]	19740.8/8.44 [c]	Outer membrane
11	ORF01250	4.72E+04	8	43	Flagellar motor protein MotB and related proteins, PAL/OmpA family [b]	16879.8/8.27 [c]	Outer membrane
12	ORF03029	2.70E+03	6	33	ATP synthase b chain precursor [a]	18464.9/5.65	Inner membrane
13	ORF00130	5.60E+03	5	31	hypothetical protein [a]	21964.3/4.57 [c]	Outer membrane
14	ORF00130	3.08E+04	6	37	hypothetical protein [a]	21964.3/4.57 [c]	Outer membrane
15	ORF02690	3.23E+03	7	39	Flagellar motor protein MotB and related proteins, PAL/OmpA family [b]	21466/5.85 [c]	Outer membrane
16	ORF00985	8.41E+03	5	35	hypothetical protein [a]	19713.1/7.16 [c]	Outer membrane
17	ORF05269	3.20E+05	10	13	ABC-type multidrug transport system, ATPase component [b]	106977.9/6.09 [c]	Inner membrane
18	ORF05269	2.53E+08	15	19	ABC-type multidrug transport system, ATPase component [b]	106977.9/6.09 [c]	Inner membrane
19	ORF06872	1.70E+04	7	39	Chromosome segregation ATPases [b]	24969.1/5.75	cytoplasm
20	ORF05691	1.37E+03	5	12	vitamin b12 receptor precursor, putative [a]	67230.5/5.89 [c]	Outer membrane
21	ORF05691	2.98E+05	9	19	vitamin b12 receptor precursor, putative [a]	67230.5/5.89 [c]	Outer membrane
22	ORF03019	504	3	14	phnD protein [a]	34729.2/6.74 [c]	periplasmic

							space
23	ORF01911	3.35E+05	9	25	ABC-type multidrug transport system, ATPase component [b]	44324.1/10.24	Inner membrane
24	ORF06104	3.66E+05	10	26	Urea amidohydrolase (urease) alpha subunit [b]	44156.8/5.42 [c]	periplasmic space
25	ORF00907	1.19E+03	4	18	protein export protein prsa precursor, putative [a]	31562.9/5.16	Outer membrane
26	ORF01781	3.94E+03	9	28	OprF, putative [a]	46436.1/7.88 [c]	Outer membrane
27	ORF01781	6.88E+04	11	35	OprF, putative [a]	46436.1/7.88 [c]	Outer membrane
28	ORF01781	1.72E+04	7	23	OprF, putative [a]	46436.1/7.88 [c]	Outer membrane
29	ORF05691	9.73E+04	8	18	vitamin b12 receptor precursor, putative [a]	67230.5/5.89 [c]	Outer membrane
30	ORF01430	1.04E+06	10	30	putative transglycosylase [a]	44091.9/5.47 [c]	Outer membrane
31	ORF01781	1.50E+03	8	26	OprF, putative [a]	46436.1/7.88 [c]	Outer membrane
32	ORF01430	7.62E+06	11	35	putative transglycosylase [a]	44091.9/5.47 [c]	Outer membrane
33	ORF06344	2.41E+08	13	42	OMP P1, putative [a] Long-chain fatty acid transport protein FadL [b]	43037/5.50 [c]	Outer membrane
34	ORF02621	6.96E+04	9	22	hypothetical protein [a]	43538/6.44 [c]	Outer membrane
35	ORF01252	7.99E+05	10	26	tolB protein [a] component of the Tol biopolymer transport system [b]	44067.5/6.96 [c]	Outer membrane
36	ORF02621	2.68E+08	15	40	hypothetical protein [a]	43538/6.44 [c]	Outer membrane
37	ORF02621	8.49E+08	15	47	hypothetical protein [a]	43538/6.44 [c]	Outer membrane
38	ORF02621	1.71E+08	13	39	hypothetical protein [a]	43538/6.44 [c]	Outer membrane
39	ORF03881	1.94E+08	13	37	OprM	47648.6/5.39 [c]	Outer membrane
40	ORF04087	2.16E+05	9	23	hypothetical protein [a]	49070.7/5.71 [c]	periplasmic space
41	ORF05691	1.05E+08	12	32	vitamin b12 receptor precursor, putative [a]	67230.5/5.89 [c]	Outer membrane
42	ORF04889	1.52E+10	14	45	Putative outer membrane protein [a]	47179.5/6.85 [c]	Outer membrane
43	ORF02873	1.35E+03	5	17	phosphate regulon sensor protein PhoR [a]	42799.9/6.05 [c]	Inner membrane
44	ORF04889	1.93E+11	18	50	Putative outer membrane protein [a]	47179.5/6.85 [c]	Outer membrane
45	ORF04352	1.13E+07	12	40	hypothetical protein [a]	54726.4/8.98 [c]	Outer membrane
46	ORF04313	3.32E+04	7	17	type I secretion system outer membrane protein RsaF (rsaF) [a]	53799.3/7.98 [c]	Outer membrane
47	ORF04313	4.73E+06	10	28	type I secretion system outer membrane protein RsaF (rsaF) [a]	53799.3/7.98 [c]	Outer membrane
48	ORF05759	3.30E+13	24	46	Outer membrane receptor proteins, mostly Fe transport [b]	71162.3/6.41 [c]	Outer membrane
49	ORF01790	1.73E+13	23	46	TonB-dependent receptor [a]	72484.2/6.65 [c]	Outer membrane

50	ORF01790	3.93E+13	25	48	TonB-dependent receptor [a]	72484.2/6.65 [c]	Outer membrane
51	ORF05759	5.14E+10	18	35	Outer membrane receptor proteins, mostly Fe transport [b]	71162.3/6.41 [c]	Outer membrane
52	ORF02346	1.14E+08	15	24	Outer membrane receptor proteins, mostly Fe transport [b]	90124.3/9.05	Outer membrane
53	ORF02715	5.74E+10	18	37	TonB-dependent receptor, putative [a]	69941.5/4.95 [c]	Outer membrane
54	ORF02715	1.87E+14	25	48	TonB-dependent receptor, putative [a]	69941.5/4.95 [c]	Outer membrane
55	ORF02715	5.88E+16	30	58	TonB-dependent receptor, putative [a]	69941.5/4.95 [c]	Outer membrane
56	ORF02715	8.44E+13	26	49	TonB-dependent receptor, putative [a]	69941.5/4.95 [c]	Outer membrane
57	ORF02715	1.34E+10	18	33	TonB-dependent receptor, putative [a]	69941.5/4.95 [c]	Outer membrane
58	ORF06100	4.23E+08	17	28	Outer membrane receptor proteins, mostly Fe transport [b]	84322.1/5.96	Outer membrane
59	ORF06100	2.91E+08	16	25	Outer membrane receptor proteins, mostly Fe transport [b]	84322.1/5.96	Outer membrane
60	ORF01711	3.51E+08	15	22	TonB-dependent receptor protein, putative [a]	89242.5/5.42 [c]	Outer membrane
61	ORF01711	1.93E+09	15	25	TonB-dependent receptor protein, putative [a]	89242.5/5.42 [c]	Outer membrane
62	ORF05536	1.07E+12	22	40	TonB-dependent receptor, putative [a]	80843.9/5.48	Outer membrane
63	ORF05536	4.80E+14	28	52	TonB-dependent receptor, putative [a]	80843.9/5.48	Outer membrane
64	ORF05536	3.75E+14	26	47	TonB-dependent receptor, putative [a]	80843.9/5.48	Outer membrane
65	ORF01459	2.68E+07	11	22	conserved hypothetical protein [a]	78160.3/5.91	Outer membrane
66	ORF01459	1.83E+06	10	21	conserved hypothetical protein [a]	78160.3/5.91	Outer membrane
67	ORF01459	1.29E+10	15	29	conserved hypothetical protein [a]	78160.3/5.91	Outer membrane
68	ORF01459	5.14E+06	10	22	conserved hypothetical protein [a]	78160.3/5.91	Outer membrane
69	ORF02965	2.98E+11	17	30	TonB-dependent receptor, putative [a]	89918.4/8.71	Outer membrane
70	ORF05268	1.41E+13	22	45	ferrichrome iron receptor, putative [a]	72997.2/5.61 [c]	Outer membrane
71	ORF05268	2.88E+14	22	46	ferrichrome iron receptor, putative [a]	72997.2/5.61 [c]	Outer membrane
72	ORF05268	2.32E+11	17	34	ferrichrome iron receptor, putative [a]	72997.2/5.61 [c]	Outer membrane
73	ORF05691	3.05E+13	22	44	vitamin b12 receptor precursor, putative [a]	67230.5/5.89 [c]	Outer membrane
74	ORF05691	1.38E+07	13	35	vitamin b12 receptor precursor, putative [a]	67230.5/5.89 [c]	Outer membrane
75	ORF05691	4.46E+11	20	49	vitamin b12 receptor precursor, putative [a]	67230.5/5.89 [c]	Outer membrane
76	ORF05691	4.77E+12	24	62	vitamin b12 receptor precursor, putative [a]	67230.5/5.89 [c]	Outer membrane
77	ORF05691	6.43E+11	21	48	vitamin b12 receptor precursor, putative [a]	67230.5/5.89 [c]	Outer membrane
78	ORF05691	1.24E+11	19	42	vitamin b12 receptor precursor, putative [a]	67230.5/5.89 [c]	Outer membrane
79	ORF05759	1.74E+12	21	39	Outer membrane receptor proteins, mostly Fe transport [b]	71162.3/6.41 [c]	Outer membrane
80	ORF01790	3.42E+10	21	41	TonB-dependent receptor [a]	72484.2/6.65 [c]	Outer membrane
81	ORF01790	1.10E+10	17	33	TonB-dependent receptor [a]	72484.2/6.65 [c]	Outer membrane
82	ORF06544	2.32E+03	7	11	outer membrane hemin receptor, putative [a]	75226.5/6.02 [c]	Outer membrane
83	ORF06544	6.96E+04	10	15	outer membrane hemin receptor, putative [a]	75226.5/6.02 [c]	Outer membrane
84	ORF06003	5.31E+08	14	22	outer membrane protein [a]	84154.8/5.63 [c]	Outer membrane

85	ORF06003	8.72E+08	17	24	outer membrane protein [a]	84154.8/5.63 [c]	Outer membrane
86	ORF06003	1.58E+16	31	42	outer membrane protein [a]	84154.8/5.63 [c]	Outer membrane
87	ORF04258	7.17E+08	13	23	outer membrane receptor, putative [a]	87590.4/5.38 [c]	Outer membrane
88	ORF02709	652	5	9	Outer membrane receptor proteins, mostly Fe transport [b]	85580.8/5.61 [c]	Outer membrane
89	ORF02709	3.85E+05	8	15	Outer membrane receptor proteins, mostly Fe transport [b]	85580.8/5.61 [c]	Outer membrane
90	ORF02709	1.05E+10	17	26	Outer membrane receptor proteins, mostly Fe transport [b]	85580.8/5.61 [c]	outer membrane
91	ORF02709	1.07E+10	18	25	Outer membrane receptor proteins, mostly Fe transport [b]	85580.8/5.61 [c]	outer membrane
92	ORF02709	7.09E+13	22	37	Outer membrane receptor proteins, mostly Fe transport [b]	85580.8/5.61 [c]	outer membrane
93	ORF07045	2.15E+03	6	9	virulence-associated protein [a] Exoribonuclease [b]	115160.3/9.73	secreted
94	ORF04119	8.41E+12	20	25	ompa-related protein precursor, putative [a]	108754.8/5.63 [c]	outer membrane
95	ORF04119	8.27E+08	14	17	ompa-related protein precursor, putative [a]	108754.8/5.63 [c]	outer membrane
96	ORF04119	7.08E+16	27	34	ompa-related protein precursor, putative [a]	108754.8/5.63 [c]	outer membrane
97	ORF04119	7.00E+12	22	25	ompa-related protein precursor, putative [a]	108754.8/5.63 [c]	outer membrane
98	ORF04119	2.49E+13	22	27	ompa-related protein precursor, putative [a]	108754.8/5.63 [c]	outer membrane
99	ORF04119	5.97E+10	17	21	ompa-related protein precursor, putative [a]	108754.8/5.63 [c]	outer membrane
100	ORF04119	6.07E+06	11	14	ompa-related protein precursor, putative [a]	108754.8/5.63 [c]	outer membrane
101	ORF04119	9.21E+04	10	11	ompa-related protein precursor, putative [a]	108754.8/5.63 [c]	outer membrane
102	ORF04472	231	4	5	conserved hypothetical protein [a]	103679.7/5.51	outer membrane
103	ORF04472	2.83E+04	7	11	conserved hypothetical protein [a]	103679.7/5.51	outer membrane
104	ORF04472	1.73E+03	6	9	conserved hypothetical protein [a]	103679.7/5.51	outer membrane
105	ORF04472	2.26E+03	6	9	conserved hypothetical protein [a]	103679.7/5.51	outer membrane
106	ORF00797	3.98E+03	6	7	conserved hypothetical protein [a]	116453.7/5.15 [c]	outer membrane
107	ORF00797	1.67E+03	6	6	conserved hypothetical protein [a]	116453.7/5.15 [c]	outer membrane
108	ORF00797	1.89E+09	15	19	conserved hypothetical protein [a]	116453.7/5.15 [c]	outer membrane
109	ORF00797	1.14E+10	16	23	conserved hypothetical protein [a]	116453.7/5.15 [c]	outer membrane
110	ORF00797	925	5	6	conserved hypothetical protein [a]	116453.7/5.15 [c]	outer membrane
111	ORF02871	1.17E+04	7	9	hypothetical protein [a]	120690.1/6.06	outer membrane
112	ORF02871	8.28E+06	11	15	hypothetical protein [a]	120690.1/6.06	outer membrane
113	ORF00440/00441	See text [d]	na	na	Outer membrane receptor proteins, mostly Fe transport [b]	see text [d]	outer membrane
114	ORF00440/00441	See text [d]	na	na	Outer membrane receptor proteins, mostly Fe transport [b]	see text [d]	outer membrane
115	ORF00440/00441	See text [d]	na	na	Outer membrane receptor proteins, mostly Fe transport [b]	see text [d]	outer membrane
116	ORF00440	See text [d]	na	na	Outer membrane receptor proteins, mostly Fe transport [b]	see text [d]	outer membrane

	/00441						
117	ORF00440 /00441	See text [d]	na	na	Outer membrane receptor proteins, mostly Fe transport [b]	see text [d]	outer membrane
118	ORF00440 /00441	See text [d]	na	na	Outer membrane receptor proteins, mostly Fe transport [b]	see text [d]	outer membrane
119	ORF00440 /00441	See text [d]	na	na	Outer membrane receptor proteins, mostly Fe transport [b]	see text [d]	outer membrane
120	ORF00440 /00441	See text [d]	na	na	Outer membrane receptor proteins, mostly Fe transport [b]	see text [d]	outer membrane
121	ORF00675 [e]	3.64E+03	6	28	hypothetical protein [a]	21536.2/5.95 [c]	outer membrane
122	ORF05415 [e]	773	6	7	sensor protein kdpd [a]	96844.5 / 6.47	inner membrane

[a] Protein description from TIGR automatic annotation.
[b] Protein description from COG function.
[c] Indicates presence of predicted Signal Peptide. Theoretical MW/pI after removal of signal peptide.
[d] See text, results section 3.4.
[e] Spots identified from gel in Fig 3B

Table 2.3: Estimated and observed proteins from *C. crescentus* based on genome sequence analysis and 2-D electrophoresis of membrane proteins

Category	Number
ORFs in genome	3767
Integral membrane proteins	731
Outer membrane proteins	140
Putative outer membrane proteins within 2-D gel resolution range (MW > 10 kDa, pI 4 – 8.2)	65 –95
Protein spots identified on gel	120
Unique proteins identified	54
Predicted membrane proteins	48
Predicted outer membrane proteins	41

[a] Predicted numbers

[b] Observed numbers

Table 2.4: Functional characterization of identified spots based on COGs.

COG functional category (Symbol)	No. of identified proteins
Transcription (K)	2
Cell division and chromosome partitioning (D)	1
Cell envelope biogenesis, outer membrane (M)	2
Cell motility and secretion (N)	4
Inorganic ion transport and metabolism (P)	16
Signal transduction mechanisms (T)	2
Energy production and conversion (C)	2
Amino acid transport and metabolism (E)	1
Lipid metabolism (I)	1
General function prediction only (R)	2
Function unknown (S)	3
Not in any COG	15

Table 2.5: Identified spots apparently overexpressed on minimal medium.

ORF no.	Protein description	Predicted MW/pI
ORF06100	Outer membrane receptor proteins, mostly Fe transport [a]	84322.1/5.96
ORF05536	TonB-dependent receptor, putative [b]	80843.9/5.48
ORF04472	conserved hypothetical protein [b]	103679.7/5.51
ORF06544	outer membrane hemin receptor, putative [b]	75226.5/6.02 [c]
ORF02346	Outer membrane receptor proteins, mostly Fe transport [a]	90124.3/9.05
ORF01790	TonB-dependent receptor [b]	72484.2/6.65 [c]
ORF05759	Outer membrane receptor proteins, mostly Fe transport [a]	71162.3/6.41 [c]
ORF05691	vitamin b12 receptor precursor, putative [b]	67230.5/5.89 [c]
ORF05268	ferrichrome iron receptor, putative [b]	72997.2/5.61 [c]
ORF06003	outer membrane protein [b]	84154.8/5.63 [c]
ORF01711	TonB-dependent receptor protein, putative [b]	89242.5/5.42 [c]
ORF02715	TonB-dependent receptor, putative [b]	69941.5/4.95 [c]

[a] Protein description from COG function.

[b] Protein description from TIGR automatic annotation.

[c] Indicates presence of predicted Signal Peptide. Theoretical MW/pI after removal of signal peptide.

CHAPTER III

PROFILING THE ALKALINE MEMBRANE PROTEOME OF *CAULOBACTER CRESCENTUS* WITH TWO-DIMENSIONAL ELECTROPHORESIS AND MASS SPECTROMETRY

The work described in this chapter discusses the development of techniques to increase the proteomic coverage of *Caulobacter cresentus* membrane proteins. I carried out the *Caulobacter crescentus* 2-DE gel optimizations, mass spectrometry, and data analysis. This work was presented at the joint AES/AIChE Meeting (Annual Meeting of American Electrophoresis Society, Nov 2001/18th Annual Meeting of the American Institute of Chemical Engineers), in Reno, NV as an equal contribution by Mark P. Molloy and Nikhil D. Phadke, and is published in it's entirety in the journal *Proteomics* (2002, *2,* 899–910) under the title "Profiling the alkaline membrane proteome of *Caulobacter crescentus* with two-dimensional electrophoresis and mass spectrometry" by Mark P. Molloy, Nikhil D. Phadke, Hong Chen, Richard Tyldesley, David E. Garfin, Janine R. Maddock, and Philip C. Andrews.

Abstract

Attempts at protein profiling in the alkaline pH region using isoelectric focusing have often proved difficult, greatly limiting the scope of proteome analysis. We investigated several parameters using custom pH 8–11 immobilized pH gradients to separate a *Caulobacter crescentus* membrane preparation. These included sample application, quenching endoosomotic flow and gel matrix composition. Among these factors, the sample application position was the predominant parameter to affect two-dimensional gel quality. Separated proteins were silver stained and profiled using matrix-assisted laser desorption/ionization (MALDI) mass spectrometry. The use of a prototype MALDI-Q-Tof mass spectrometer assisted identification of several proteins by providing highly informative peptide fragmentation data from the sample digests. Thirty-two unique alkaline proteins were identified in this study, which complements our previously described *C. crescentus* membrane proteome. Our experiments point towards new options for proteomic researchers aiming to both extend the scope of analysis, and simplify methods of identifying proteins with high confidence.

Introduction

The recent completion of the *Caulobacter crescentus* genome sequencing [1], and proteomic research initiated in our laboratories [2] indicates that this Gram negative heterotroph possesses a large number of alkaline outer membrane proteins (OMPs). This is in contrast to *Escherichia coli* and other enterobacteria where the majority of OMPs lie between p*I* 4–7. We have previously demonstrated that OMPs within the p*I* 4–7 region are amenable to routine 2-DE [3, 4]. While approaches for the enrichment and solubilization of hydrophobic proteins and membrane proteins have advanced significantly over recent years [5, 6], the development of methods for IEF in the alkaline pH region have seen slow progress. This is despite our knowledge that within a given prokaryotic proteome there is a bimodal distribution of proteins based on p*I*, with clusters centered around pH 5 and pH 9 [7]. Furthermore, the number of proteins estimated with alkaline p*I*'s (*i.e.* F pH 7.5) forms a significantly large portion of the proteome: 38% in *E. coli*, 49% in *Methanococcus jannaschii* and 62% in *Helicobacter pylori* [8]. In eukaryotes the situation appears more complex with a trimodal distribution clustered around pH 5, pH 7 and pH 9 [8]. The poor coverage of proteins within the alkaline cluster using 2-DE is a serious problem for proteomic efforts that strive for large- scale, global analyses.

The problem of cathodic drift seen with carrier ampholyte formed pH gradients imparts its greatest effect on IEF of alkaline proteins, resulting in a dramatic loss of these proteins with standard focusing protocols [9]. However, by decreasing the focusing time

to achieve a nonequilibrium, transient state (*i.e.* NEGPHE), alkaline proteins have been resolved by classical tube gel systems, albeit with decreased resolution compared to steady-state focusing experiments. The introduction of IPGs for IEF alleviated the problem of cathodic drift, allowing for highly reproducible, steady-state focusing of milligram quantities of protein [10, 11]. Nonetheless, even with the issue of cathodic drift solved by IPGs, isoelectric focusing in the alkaline region has remained problematic, with two key factors identified.

When producing shallow alkaline IPGs the buffering power of water becomes a significant factor requiring the use of very basic (pK 10.3, pK F 13) acrylamido buffers in formulating the pH gradient [12]. An IPG matrix formed with very basic acrylamido buffers exhibits a strong positive charge, establishing a "reverse-endosmotic flow" (REOF) of ionized water towards the anode. The transport of water under these conditions reportedly impinges upon steady-state focusing of proteins [13]. To counteract this effect, reagents such as sorbitol [14], methylcellulose and isopropanol [13] have been added to the gel matrix in efforts to quench this process. A second problem stems from the instability of acrylamide under extremely acidic or basic conditions where it undergoes hydrolysis, forming acrylic acid [15, 16]. The presence of these charged groups within the matrix establishes an EOF and compromises protein resolution. Solutions have been sought by developing alternative, *N*-substituted acrylamide derivatives that sterically hinder hydrolysis of the amido group. One such IPG matrix formed with dimethylacrylamide (DMA) has shown improved alkaline separations of histones and ribosomal proteins [13]. Several other *N*-substituted acrylamides have been

described [17], but to date have not been put to use in IPGs for narrow range alkaline IEF.

In our efforts to further catalogue *C. crescentus* OMPs it became necessary to investigate the parameters for high resolution alkaline IEF using IPGs. We constructed a series of three pH 8–11 IPGs composed of either acrylamide, or the *N*-substituted derivatives, DMA, and acryloylaminoethoxyethanol (AAEE) [15]. These matrices were tested under several conditions to separate *C. crescentus* OMPs. Peptide mass mapping of silver stained spots was conducted for protein identification. In some cases protein identification was facilitated by conducting the analysis using a hybrid, MALDI-Q-Tof mass spectrometer.

Materials and methods

Growth of *C. crescentus* cultures

Caulobacter crescentus cells were grown in rich medium and harvested as described previously [2]. Briefly, late exponential phase cultures of *C. crescentus* CB15N were grown at 30UC with aeration to an OD600 of 0.8, and harvested by centrifugation at 7000 *g*. Cells were washed with 50 mM Tris-HCl pH 8.0, and either used immediately or stored at –80°C.

Sample preparation

Samples for IEF were prepared as described earlier [2], with the exception of the sodium carbonate enrichment step. The harvested cells were resuspended in 50 mM Tris-HCl (pH 8.0) with freshly prepared 1X protease inhibitor cocktail (Boehringer Mannheim, Mannheim, Germany), 1 mM PMSF and 0.5 mM EDTA. Cells were lysed by two passages through a French pressure cell (AMINCO, Silver Spring, MD, USA) at 16000 psi. Cell debris was separated by centrifugation (2X) at 7000 *g* for 10 min at 4UC and centrifugation at 17000 *g* for 25 min at 4UC. Membrane vesicles were pelleted from the supernatant by ultracentrifugation at 170 000 *g* for 1 h in a Beckman 50.2 Ti rotor (Fullerton, CA, USA) at 4°C. Membrane vesicles were resuspended using a Fisher Scientific ultrasonic dismembrator 60 at 6 W RMS for 2 bursts of 20 s and washed twice (50 000 rpm for 45 min at 4UC in a Sorvall RP80-AT rotor; Kendro, Newtown, CT, USA) with 50 mM Tris-HCl pH 8.0. Total protein concentration was calculated using a Bradford assay and membranes were stored in 0.4 mg aliquots of 400 ml 20% w/v glycerol, 50 mM Tris-HCl pH 8.0 at –80°C.

Preparation of alkaline IPGs

DMA was purchased from Fluka (St. Louis, MO, USA) and AAEE was synthesized essentially as described [15]. All IPG monomers were synthesized as described [9], except IPG monomer pK F 12 (Aldrich, St. Louis, MO, USA) and prepared as 0.2 M stock solutions in n-propanol. 17 cm pH 8–11 IPGs were designed by modifying published recipes [9, 18] as described in Table 3.1. Acrylamide gels were 4%T, 3%C (bis-acrylamide), while DMA and AAEE IPGs were 5%T, 3%C (bis-acrylamide), each with a mean buffering capacity of 2.9 meq/v/pH/L. All gels were washed four times with water and once with 0.5–2% glycerol solution. The gels were then dried, covered with polyester films, and cut into 3 mm strips.

Two-dimensional electrophoresis

Acrylamide, DMA, and AAEE IPGs were used for the first dimension. IPGs were rehydrated overnight with 380 mL of rehydration buffer (7 M urea, 2 M thiourea, 2 mM tributyl phosphine (TBP), 0.5% v/v Bio-Lyte 3/10 (Bio-Rad, Hercules, CA, USA), 40 mM Tris base, 1% w/v ASB-14 (Calbiochem, La Iolla, CA, USA), 0.5% w/v Triton X-100). In some cases 15% v/v isopropanol was also included in the rehydration buffer. Membrane vesicles were solubilized by sonicating 0.5 mg of total protein in 80 ml of rehydration buffer. Samples were applied to the IPGs using loading cups (Amersham Pharmacia Biotech, Uppsala, Sweden) or included in the rehydration solution. IEF was carried out for 80 000 Vh at a maximum of 6000 V (150 V/3 h, 300 V/1 h, 600 V/1 h, 1000 V/1 h, 3000 V/1 h, 6000 V until 80 kVh) using the Multiphor II system (Amersham

Pharmacia Biotech). IPG equilibration and SDS-PAGE were as previously described [3].

Staining of 2D gels and imaging

MALDI mass spectrometry compatible silver staining was carried out according to the protocol of Vorum as described earlier [2]. The gel images were scanned with a UMAX Power Look II scanner (UMAX Technologies, Fremont, CA, USA). The gel images were manipulated and analyzed using Adobe Photoshop 6.0.

Mass spectrometric analysis

Manual tryptic digestion and MALDI-MS by PerSeptive BioSystems Voyager DE-STR

Peptide mass fingerprinting (PMF) was carried out as described previously [2] with minor modifications. Spots were excised from the silver stained gels and washed with a solution of 50% v/v acetonitrile and 50 mM ammonium bicarbonate. In-gel digestions were performed using 150–300 ng modified porcine trypsin (Promega, Madison, WI, USA) in 12 mL of freshly prepared 100 mM ammonium bicarbonate at 37°C overnight. The peptides were extracted using 60% v/v acetonitrile, 1% v/v TFA, concentrated to near dryness, then resuspended in 6 mL 3% v/v TFA. 0.8 mL of each sample was loaded onto a gold plated MALDI plate (PerSeptive Biosystems, Framingham, MA, USA) with an equal volume of 10 mg/mL w/ v a-cyano-4-hydroxy cinnamic acid solution in 50% v/v acetonitrile and 1% v/v TFA. MALDI-MS was performed on a Voyager-DE STR instrument (PerSeptive Biosystems) run in delayed extraction reflector mode with the following parameters: 1982 laser intensity, 25 kV accelerating voltage, 72% grid voltage, 0% guide wire voltage, 100 ns delay and a low mass gate of 500 Da. Spectra were

calibrated with 842.5 Da and 2211.1 Da trypsin peaks using Data Explorer (PE Biosystems, Foster City CA, USA). The resulting peptide mass fingerprints were searched using a local copy of the program MS-Fit [19] against a local *C. crescentus* database.

Automated tryptic digestion

Fifteen protein spots representing both lightly and heavily silver stained proteins were manually excised from the gel and washed as described above. The spots were dissected into 1 mm cubes and transferred into a 96 well plate, then automatically digested using a Micro- mass MassPREP Station (Micromass, Wythenshawe, UK). Briefly, the silver stained gel pieces were destained with alternate potassium ferricyanide and sodium thiosulphate washes. The proteins were then reduced and alkylated with the addition of DTT and iodoacetamide respectively followed by extensive washes with ammonium bicarbonate and acetonitrile. Protein digestion was performed with the addition of 25 mL trypsin solution at 6 ng/mL (Promega) at 37°C for 5 h. The resulting peptides were extracted in an aqueous solution of 1% v/v formic acid, 2% v/v acetonitrile.

For the M@LDI instrument (Micromass), samples were spotted on to a standard 96 well M@LDI target plate and a-cyano-4-hydroxycinnamic acid (10 mg/mL) added. For MALDI-Q-Tof analysis, samples were manually spotted onto a 10 well target plate and a saturated solution of dihydroxybenzoic acid in 80% acetonitrile was added.

MS and MS/MS by Micromass MALDI-Q-Tof

Peptides were analyzed using a Q-Tof instrument (Micromass) fitted with an experimental MALDI source [20]. The prototype utilized a strip sample plate and holder in place of the standard electrospray ion source. The holder carried a target plate that contained 10 sample wells which when introduced *via* the vacuum lock moved along a horizontal track under stepper motor control. MALDI conditions were achieved by using a 337 nm laser that illuminated the target with a beam diameter of approximately 300 mm. After the sample was consumed, the target plate was stepped and data recorded from this 'track' across the target sample. The rate of movement was set so that without user intervention approximately 6 min of data could be acquired. The actual acquisition used the previously acquired data to determine whether or not the data quality would be improved by adding more data to the same experiment. This then allowed multiple MS/MS experiments to be per- formed from the same spot. To ensure efficient transfer of the ions into the analyzer a capillary line and needle valve were used to introduce gas into the source housing to provide collisional cooling of the ion beam [21] and to provide an operating pressure of approximately 0.25 mbar. The instrument was operated at a resolution of D4500 (full width at half maximum) and spectra were recorded and summed, with a 5 s integration. In the MS/MS mode, argon collision gas was used with the collision energy being set to approximately (0.05 V/Da precursor mass). Whilst observing the ion distribution in real time, the collision energy was further varied to obtain a fragmentation pattern across the (MS/MS) mass range.

MALDI-MS by Micromass M@LDI-R

An aliquot of digest prepared as described above was analyzed using a Micromass M@LDI-R instrument. The instrument was operated in positive ion mode with the source voltage set to 15 kV. The pulse voltage was optimized at 3125 V, the detector voltage was 2000 V and the reflectron voltage was set to 500 V. Laser position and energy were automatically optimized using the MAXSpec software algorithm.

Protein identification criteria

The *C. crescentus* peptide mass fingerprint searchable database was created as described previously [2]. Briefly, ORFs from the genome were annotated and formatted to enable searching with the peptide mass data using search engines. For PMF searches, a mass accuracy of 50 ppm was used. A maximum of one missed enzymatic cleavage, and modification of cysteines by carboxyamidomethylation or modification by acrylamide were considered during the searches. For protein identification using Protein Prospector (Voyager DE-STR MALDI) the requirement for successful matches required a minimum of four peptides and a MOWSE score F 1000. For MALDI- Q-Tof MS/MS, MALDI-Q-Tof PMF and M@LDI PMF searching, the top ranked candidate protein returned by the ProteinLynx Global server engine was recorded. At this point highly confident, unambiguous matches returned a % probability score of approximately 100% for PMF data, or an MS/MS score F 30 for fragmentation data. However, to improve confidence, results were combined so that identifications were only considered successful if MS/MS data obtained for two or more peptides matched the same protein, or if MS/MS from one peptide and PMF data from either MALDI-Q-Tof or M@LDI were in agreement.

Results

Investigation of the *C. crescentus* genome sequence indicates that 42% of predicted proteins possess a p*I* F 7.5 (Fig. 3.1). Furthermore, our initial proteomic investigation of *C. crescentus* using standard IEF conditions revealed numerous alkaline OMPs [2], contrasting the situation observed for other enterobacteriaceae [4]. To facilitate a detailed investigation of the *C. crescentus* alkaline membrane proteome, under the previously described rich medium growth conditions [2], we prepared three sets of narrow range alkaline IPGs (pH 8–11), with a matrix backbone constructed of acrylamide, DMA or AAEE. The two additional *N*-modified polymers (DMA and AAEE) were chosen as previous reports have suggested they may be more suitable to protein separation under basic pH conditions [13, 15]. IPGs constructed of DMA are stable at alkaline pH and have previously been described for separation of some basic proteins such as ribosomes and histones [13]. However, stabilizing the amido group through the incorporation of additional methyl groups markedly increases the hydrophobic property of DMA [15], which may impart an unwanted, profound effect towards protein losses using 2-D gels. On the other hand, AAEE is base stable and has superior hydrophilic properties over both DMA and acrylamide [15], and may be a more suitable choice for alkaline IEF.

Sample application for alkaline IPGs

In a series of preliminary investigations we used a *C. crescentus* membrane preparation and a yeast cell lysate to examine the optimal sample application point for narrow range alkaline IPGs. Three separate loading conditions were tested, in-gel rehydration loading [22], anode cup loading and cathode cup loading. Anode cup loading

produced 2-D gels of high quality (Fig. 3.2B), while loading at the alkaline cathode position poorly resolved gels with prominent horizontal streaking (not shown). The rehydration loading technique that is highly efficient and suitable for high protein loads when used with standard IPGs (pH 4–7, pH 3–10), consistently produced poorly focused, often streaky gels with obvious protein loss (Fig. 3.2A). We were initially surprised by these results, because for routine 2-DE experiments we find that with moderately high protein loads (500 mg), rehydration loading is the preferred loading method. As an additional point, we observed consistently poor results for the rehydration loading method independent of the gel matrices tested, however, by reverting to anode cup loading good quality gels were obtained. This observation indicates that the loading technique and not the gel matrix were the basis for the varied resolution.

IPG matrix effects on protein separation

After determining that anode cup loading was the preferred sample application point for narrow range alkaline IPGs we investigated the effects imparted by the gel matrix for separating *C. crescentus* OMPs. Initial experiments involved separating OMPs with the pH 8–11 IPGs under standard running conditions. For these experiments no additives to quench REOF were included (Fig. 3.3A-C). Surprisingly, for each gel matrix tested the OMP preparation was satisfactorily separated with well defined spot shape. Overall, only minor differences in resolution quality were noted between each gel matrix, with the acrylamide and AAEE gels showing arguably less streaking than DMA. However, we observed that towards the extreme basic portion of the DMA and AAEE gels, additional proteins were resolved that were missing from the acrylamide gel. This is a significant point, as at least 20 very alkaline proteins including several abundant

polypeptides were detected with both the *N*-substituted matrices, but not with the standard acrylamide gel. One explanation consistent with the literature is hydrolysis of the acrylamide gel at the basic pH extremity compromising protein resolution [15].

Figure 3.3D-F shows a second set of experiments examining the effect of adding isopropanol to the sample solution as a quenching reagent against REOF [13]. As was the case in Fig. 3.3A-C that contained no reagent to quench REO, the inclusion of 15% isopropanol to the gels shown in Fig. 3.3D-F caused only subtle differences in the protein patterns for each gel matrix. As reported for Fig. 3.3A-C, DMA and AAEE gels contained additional protein spots at the extreme basic section of the gel that were absent from the acrylamide gel. In comparing Fig. 3.3A-C (no isopropanol) with Fig. 3.3D-F (15% isopropanol) only minor differences in protein resolution were noted, although we consider that spot shape was slightly superior (less streaked) with the addition of 15% isopropanol. These subtle differences in protein patterns leads us to conclude that REO had a superficial effect on protein resolution, or the addition of isopropanol played only a minor role in dampening REO. Additionally, we did not observe wholesale losses or gains of proteins based upon matrix composition suggesting that polymer hydrophobicity did not influence recovery of proteins in the second dimension as was anticipated for DMA gels.

Profiling *C. crescentus* alkaline OMPs by MALDI-MS

We had previously noted that *C. crescentus* contained numerous alkaline OMPs that did not resolve to high reso- lution using standard pH 3–10 IPGs, and that improved separation of these proteins was needed to more fully profile *C. crescentus*. Each silver

stained protein spot that we considered by visual examination to contain sufficient protein for MALDI-MS was excised and manually digested with trypsin *in situ*. Table 3.2 lists MS results for the identification of silver stained spots annotated in Fig. 3.4. For manual digestion procedures and PMF using the PE Biosystems DE-STR instrument, we trialed various peptide preparation methods that had been reported for enhanced MS analysis of silver spots including destaining prior to digestion [23] and reverse-phase peptide concentration after digestion using C18 ZipTips (Millipore, Bedford, MA, USA). In our experience, however, using a manual approach for digestion, the highest quality data were obtained when we adopted our protocol designed for the analysis of Coomassie stained spots, with minor variations to ensure maximum concentration of the analytes at all stages. In most cases an apparent successful spectra was obtained, and with filtering to remove the many spurious background ions, at least six peptides were recovered *per* protein spot. We noted that unlike the case for Coomassie stained protein spots, the interpretation of MALDI data from silver stained spots was significantly more labor intensive, as careful interrogation was required to select monoisotopic ions derived from the peptide analyte and exclude background "junk" peaks. We also experienced several cases where very few peptides were recovered from heavily stained silver spots (*e.g.* spots 23, 30, 32, 33, 36, 39, 49, 52, 56, 57, 60, 61). Furthermore, several small, weakly stained spots gave surprisingly good spectra with many peptides (*e.g.* spots 16, 19, 46). We conclude here that due to non uniform staining intensity when using silver, it may be wise for future experimentation to select all visible protein spots. In our previous experiences we had found a better correlation between stain intensity and successful PMF when using Coomassie for detection, such that very faintly stained Coomassie spots

usually generated poor spectra.

Profiling OMPs using MALDI-Q-Tof

During the course of this work, despite repeated attempts, certain heavily stained silver spots (*e.g.* spots 31, 35, 40) could not be identified using our Voyager DE-STR MALDI instrument. For further investigation we selected 15 silver stained spots of varying intensities from a replicate gel and analyzed tryptic digests of these samples using a recently described MALDI-Q-Tof mass spectrometer [20]. MALDI-Q-Tof provides highly informative, peptide fragmentation data from a sample digest applied to a MALDI plate, eliminating the need for often cumbersome, labor intensive liquid MS interfaces. Furthermore, the single charge state of precursor ions simplifies subsequent data analysis. Table 3.3 shows data obtained from the MALDI-Q-Tof experiments. This approach permitted protein identification using the standard, PMF method, and importantly, higher confidence assignments using peptide fragmentation data. MALDI-Q-Tof data allowed assignment of candi- date proteins with high probability in 9/15 spots examined using solely the PMF approach. As an example the spectra obtained for spot 18 is shown in Fig. 3.5A (MS) and Fig. 3.5B (MS/MS). MS/MS data showed no discrepancy in protein assignments when compared to assignments established solely in MS mode. Even with this small data set the advantage of conducting MALDI-Q-Tof was quickly apparent, as highly informative fragmentation data was obtained for an additional two samples (spots 25, and 43) that would have otherwise remained inconclusively identified. In 3/15 cases no useful MS/MS data was obtained (spots 26, 42, 45).

As a further check of the validity of the MALDI-Q-Tof MS/ MS data we subjected

the same digests to PMF using a conventional ionization M@LDI-R instrument. Fig. 3.5C shows the spectrum of spot 18 obtained with the M@LDI-R. Unfortunately, for the additional four candidates indicated by MS/MS, in only one case using the M@LDI-R instrument could we obtain data of sufficient quality to indicate a confident assignment, nonetheless this assignment was in agreement to that established using MALDI-Q-Tof. Two out of four of these samples were not considered positive matches as they failed our criteria of two or more MS/MS peptides, or one MS/MS peptide plus consistency with PMF data. It is noteworthy that one of these four samples (spot 25) was positively assigned based upon solely fragmentation spectra from two peptides. In this case, without the benefit of MS/MS, this protein would remain unassigned as both the conventional and modified MALDI instruments failed to produce a confident result using a PMF search strategy. As a final point it is important to note that no discrepancies were observed between the candidates returned for the entire M@LDI and MALDI-Q-Tof matching data set (10/15).

Profiling the *C. crescentus* alkaline membrane proteome

Overall, confident identifications were made to 36 spots, four of which contained multiple proteins (spots 12, 31, 42, 43). As anticipated, the majority of these proteins were categorized as membrane proteins, although a significant portion of the smaller, alkaline proteins were ribosomal components. The presence of cytoplasmic ribosomal proteins was not unexpected, as these samples did not receive an alkaline carbonate wash, to strip transient and loosely associated proteins from the membrane [3, 4]. Nine of the identified membrane proteins were putative OMPs, while an additional nine were putative inner membrane proteins. Of the 32 unique proteins identified in this study, only

two overlap with the 54 unique proteins identified in our earlier study, carried out using standard pH 3–10 IEF strips [2]. This is noteworthy because the bulk of the proteins identified here have p*I*s between 8–10, and should have theoretically been resolved on a pH 3–10 gel.

A functional distribution of the identified proteins based on the cluster of orthologous groups (COG) system of classification is shown in Table 3.6 [24]. With the exception of ribosomal proteins, the functional distribution of the proteins is similar to what we have reported previously [2]. Transporters of inorganic ions comprise the largest group of identified proteins, the most abundant of which are the TonB dependent outer membrane receptors. We had previously identified 16 members of this family, and in the present study have shown the presence of an additional 5 TonB dependent receptors. Overall, we have confirmed the expression of 21 of the approximately 70 predicted TonB dependent receptors in the *C. crescentus* genome.

Discussion

A significant number of *C. crescentus* OMPs possess alkaline p*I*s and are poorly separated using conventional IPGs. In this report we have used narrow range alkaline IPGs to greatly improve separation of these OMPs. The improved resolution allowed MS identification of protein spots that now contribute towards an alkaline master gel of *C. crescentus* membranes. This alkaline master gel complements our previous work in profiling *C. crescentus* OMPs over the pH 3–10 range [2]. Using these conditions, we have been able to identify 30 additional proteins from the *C. crescentus* proteome, many of which are OMPs. Of the newly identified proteins we have confirmed expression of five previously unseen TonB dependent receptors, which are the largest subgroup of OMPs in this genome. With this information complementing our previous results, we are poised to make a more comprehensive functional analysis of the expression patterns of the OMPs in the *C. crescentus* genome. Additionally, nine of the proteins identified in this study, are putative inner membrane proteins. It is widely believed that due to their hydrophobic nature, inner membrane proteins are difficult to resolve by 2-DE. It is plausible that we were able to resolve these proteins in this study, due to the fact that not one of these proteins is predicted to have more than a single transmembrane domain and their average hydropathicity values range between –0.3 and –0.09, *i.e.* they are not highly hydrophobic.

Of the IEF parameters tested during this study, the position of sample application appeared most critical in producing high resolution gels with limited protein losses and minimal streaking. Our observation that anode cup loading provided the optimal loading

position in the pH 8–11 gel is consistent with the work of Ohlmeier *et al.* [25] who presented *Bacillus subtilis* gels of high quality using a pH 4–12 IPG, and also reported poor resolution when the in-gel rehydration method was used. One possible explanation for our observations is an ionic repulsion effect between basic proteins and the strongly charged gel matrix resulting in severe protein loss during the passive diffusion of sample entry. This effect was not observed when the sample was applied under an electric field at the anode (least charged position for this sample).

The literature recounts only a few applications of IEF in the alkaline region using IPGs. Previous research has demon- strated that IEF in the alkaline region can be problematic, with issues stemming from REOF and instability of the gel matrix [13, 15]. Our experience with *C. crescentus* OMPs indicated that these are only minor issues, and attempts to treat them by substituting gel polymer compositions and quenchers of REOF resulted in only superficial changes to protein patterns. Furthermore, our conclusions are consistent with the quality gels displayed by Ohlmeier *et al*. [25] that were composed of polyacrylamide and contained no additive to quench REOF. It should be pointed out that in both Ohlmeier *et al.* and in our study, strong solubilizing conditions were used (*i.e.* thiourea was included), and this may have contributed to the quality of our separations. Thiourea was not included in earlier studies with alkaline IPGs that required additives to improve resolution [*e.g.* 13].

The use of the alternative gel polymers DMA and AAEE showed a greater alkaline separation distance than the reference acrylamide gel. Based upon protein identities we

consistently detected proteins towards the cathodicend of the IPG with p*I*s F 10. The most alkaline protein identified, ORF 4783 (spot 64), possessed a p*I* of 10.41. Ohlmeier *et al*. reported a similar level of coverage with alkaline IPGs. The two most basic ribosomal proteins iden- tified in their study, RplB and RplD from *B. subtilis*, possessed p*I*s of 11.01 and 10.49 respectively [25]. It should be noted that although RplB was identified in that study, its spot shape was very elongated and its position on the gel was no more basic than that of RplD and several other proteins with lower p*I*s, suggesting the analyzed protein did not reach its theoretical p*I*. Their observations, combined with our data as presented here, imply a fundamental technical limitation to our present 2-DE techniques for separation of extremely basic polypeptides. Most likely this problem is centered on difficulties encountered in generating extremely basic IPGs that would be capable of resolving such polypeptides. The most alkaline acrylamido derivatives available to provide buffering capacity for IPGs has a pK of 10.3. A quaternary amine of pK F 13 is also used as a titrant, but most effective buffering can only occur within a narrow window around the pK 10.3 acrylamido buffer. The lack of sufficient buffering capacity greater than approximately pH 10.5 will continue to hamper efforts to extend the utility of any alkaline IPG unless newly formulated acrylamido buffers with pK F 10.3 can be developed.

The use of MALDI-MS to generate peptide mass maps generally provides a good degree of certainty in the identification of proteins from small genome organisms. However, as the genome size increases, there is greater ambiguity towards protein identification when PMF is adopted as the sole approach [26, 27]. Several supplementary

methods have been used to improve the certainty of protein identifications from PMF of large genome organisms including protein digestion using multiple proteases [28] and increasing the mass accuracy of analysis [29, 30]. A clear advantage of instruments like the MALDI- Q-Tof is the ability to obtain peptide mass fingerprints and MS/MS spectra from the same sample, allowing proteins to be identified with a high degree of certainty even from large genome organisms. This ability will be greatly enhanced once control software is developed that allows data-dependent acquisition. Data-dependent acquisition is particularly useful for deciphering protein mixtures as well as for proteins that are highly modified, both of which represent difficulties for classical PMF approaches.

References

[1] Nierman, W. C., Feldblyum, T. V., Laub, M. T., Paulsen, I. T., *et al.*, *Proc. Natl. Acad. Sci. USA* 2001, *98*, 4136–4141.

[2] Phadke, N. D., Molloy, M. P., Steinhoff, S. A., Ulintz, P. J., *et al.*, *Proteomics* 2001, *1*, 705–720

[3] Molloy,M.P.,Herbert,B.R.,Slade,M.B.,Rabilloud,T.,*etal.*, *Eur. J. Biochem.* 2000, *267*, 2871–2881.

[4] Molloy,M.P.,Phadke,N.D.,Maddock,J.R.,Andrews,P.C., *Electrophoresis* 2001, *22*, 1686–1696.

[5] Molloy, M. P., *Anal. Biochem*. 2000, *280*, 1–10.

[6] Santoni, V., Molloy, M., Rabilloud, T., *Electrophoresis* 2000, *21*, 1054–1070.

[7] VanBogelen, R. A., Schiller, E. E., Thomas, J. D., Neidhardt, F. C., *Electrophoresis* 1999, *20*, 2149–2159.

[8] Schwartz, R., Ting, C. S., King, J., *Genome Res*. 2001, *11*, 703–709.

[9] Righetti, P. G., *Isoelectric Focusing: Theory, Methodology and Applications,* Elsevier, Amsterdam 1983.

[10] Sanchez, J.-C., Rouge, V., Pisteur, M., Ravier, F., *et al.*, *Electrophoresis* 1997, *18*, 324–327. [11] Görg, A., Obermaier, C., Boguth, G., Harder, A., *et al.*, *Electrophoresis* 2000, *21*, 1037–1053.

[12] Righetti,P.G.,*ImmobilizedpHGradients:TheoryandMethodology*, Elsevier, Amsterdam 1990.

[13] Görg, A., Obermaier, C., Boguth, G., Csordas, A., *et al.*, *Electrophoresis* 1997, *18*, 328–337.

[14] Bossi, A., Gelfi, C., Orsi, A., Righetti, P. G., *J. Chromatogr. A* 1994, *686*, 121–128.

[15] Chiari, M., Micheletti, C., Nesi, M., Fazio, M., Righetti, P. G., *Electrophoresis* 1994, *15*, 177–186.

[16] Miertus, S., Righetti, P. G., Chiari, M., *Electrophoresis* 1994, *15*, 1104–1111.

[17] Chiari, M., Righetti, P. G., *Electrophoresis* 1995, *16*, 1815–1829.

[18] Görg, A., Boguth, G., Obermaier, C., Weiss, W., *Electrophoresis* 1998, *19*, 1516–

1519.

[19] Clauser, K. R., Baker, P., Burlingame, A. L., *Anal Chem*. 1999, *71*, 2871–2882.

[20] Harvey, D. J., Bateman, R. H., Bordoli, R. S., Tyldesley, R., *Rapid Commun. Mass Spectrom*. 2000, *14*, 2135–2142.

[21] Krutchinsky, A. N., Loboda, A. V., Spicer, V. L., Dworschak, R., *et al.*, *Rapid Commun. Mass Spectrom*. 1998, *12*, 508– 518.

[22] Rabilloud, T., Valette, C., Lawrence, J. J., *Electrophoresis* 1994, *15*, 1552–1558.

[23] Gharahdaghi,F.,Weinberg,C.R.,Meagher,D.A.,Imai,B.S., Mische, S. M., *Electrophoresis* 1999, *20*, 601–605.

[24] Tatusov, R. L., Galperin, M. Y., Natale, D. A., Koonin, E. V., *Nucleic Acids Res.* 2000, *28*, 33–36.

[25] Ohlmeier, S., Scharf, C., Hecker, M., *Electrophoresis* 2000, *21*, 3701–3709.

[26] Kuster, B., Mann, M., *Curr. Opin. Struct. Biol*. 1998, 8, 393– 400.

[27] Yates J. R., *J. Mass Spectrom*. 1998, 33, 1–19.

[28] James, P., Quadroni, M., Carafoli, E., Gonnet, G., *Protein Sci.* 1994, *3*, 1347–1350.

[29] Jensen, O. N., Podtelejnikov, A., Mann, M., *Rapid Commun. Mass Spectrom.* 1996, *10*, 1371–1378.

[30] Takach, E. J., Hines, W. M., Patterson, D. H., Juhasz, P., *et al.*, *J. Protein Chem.* 1997, *16*, 363–369.

Figure 3.1: Bimodal distribution of *C. crescentus* proteins illustrated through a p*I vs M*r plot of ORFs before removal of predicted signal peptides. The X-axis shows the predicted p*I*s of the proteins on a linear scale and the Y-axis shows the predicted *M*r of the proteins on a logarithmic scale.

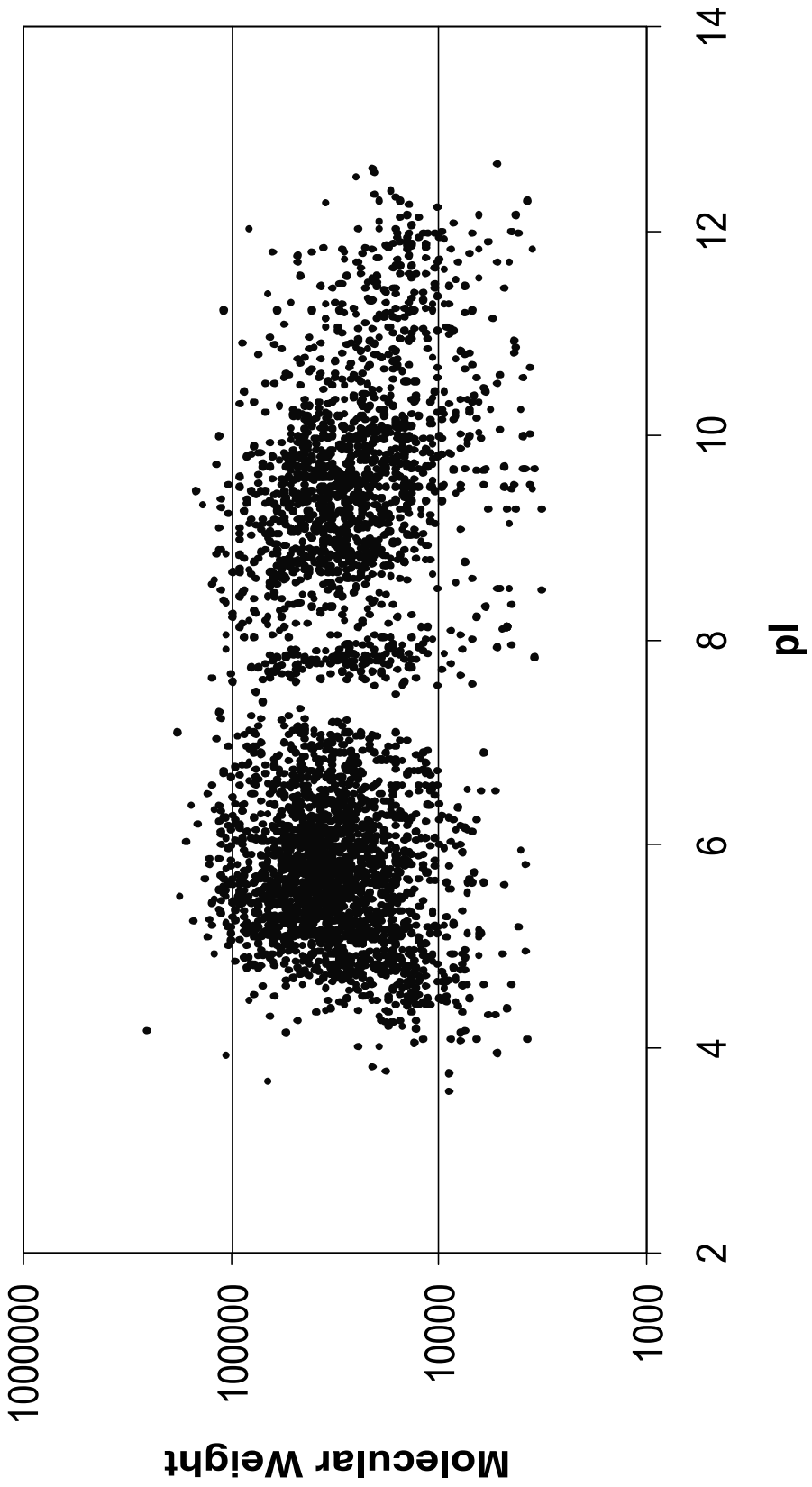

Molecular Weight
1000000
100000
10000
1000
pI
2
4
6
8
10
12
14

Figure 3.2: Effect of sample loading method on resolution. pH 8– 11 acrylamide gels loaded by (A) standard rehydration methods and (B) anodic cup loading.

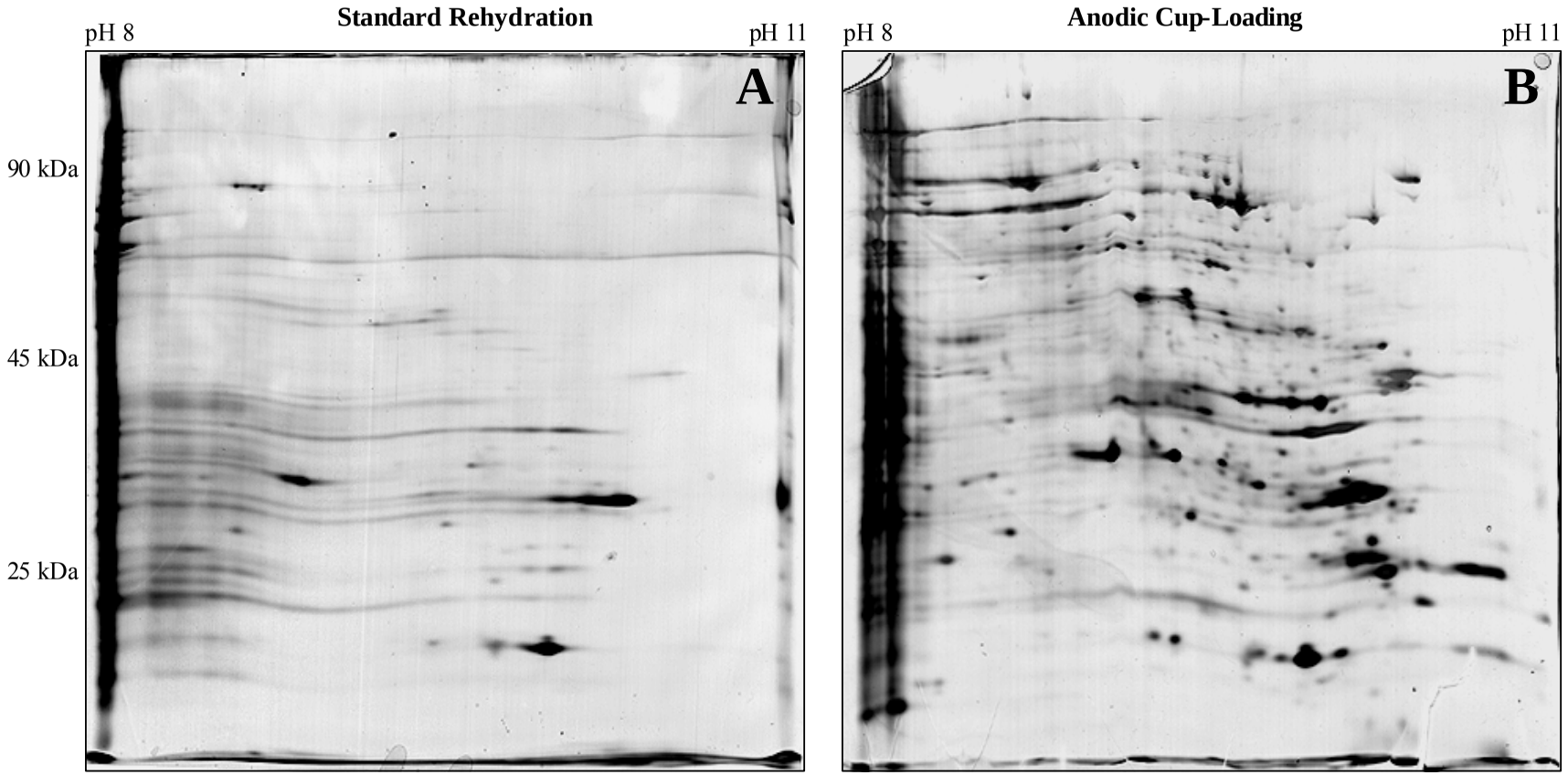

Standard Rehydration
pH 8
pH 11
A
90 kDa
45 kDa
25 kDa
Anodic Cup-Loading
pH 8
pH 11
B

Figure 3.3: Effect of isopropanol on separation in different IEF gel matrices. Alkaline gels run without isopropanol (A, B, C) and with isopropanol added (D, E, F). Matrices were made of acrylamide (A, D), AAEE (B, E) and DMA (C, F).

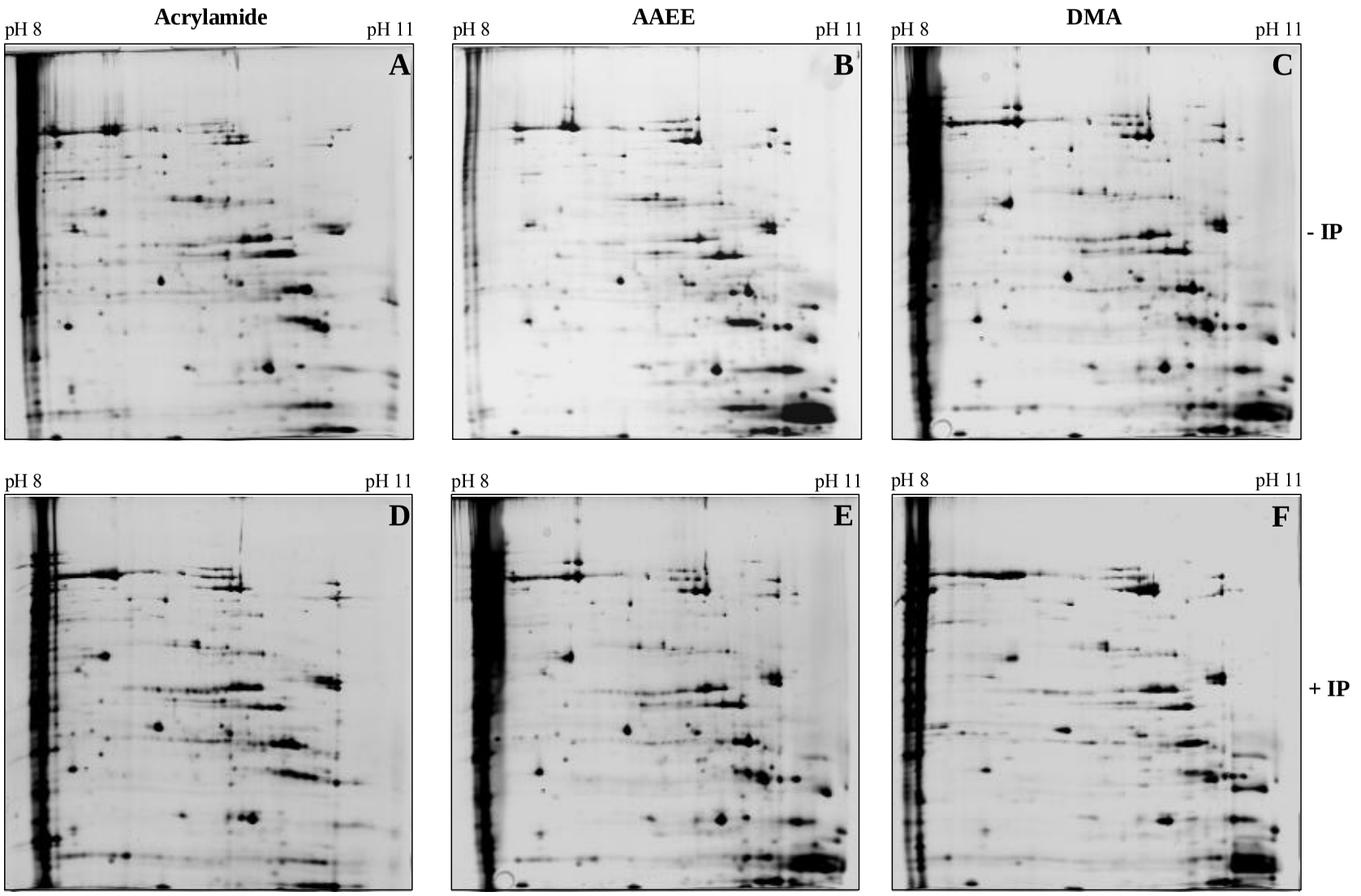

Acrylamide
AAEE
DMA
pH 8
pH 11
A
B
C
D
E
F
- IP
+ IP

Figure 3.4: Alkaline pH 8–11 AAEE master gel of *C. crescentus* membrane preparations. Analyzed spots are indicated with numbers. Spots identified by PMF and/or MS/MS are in **bold** and underlined. Identifications are described in Tables 3.2, 3.3, 3.4 and 3.5.

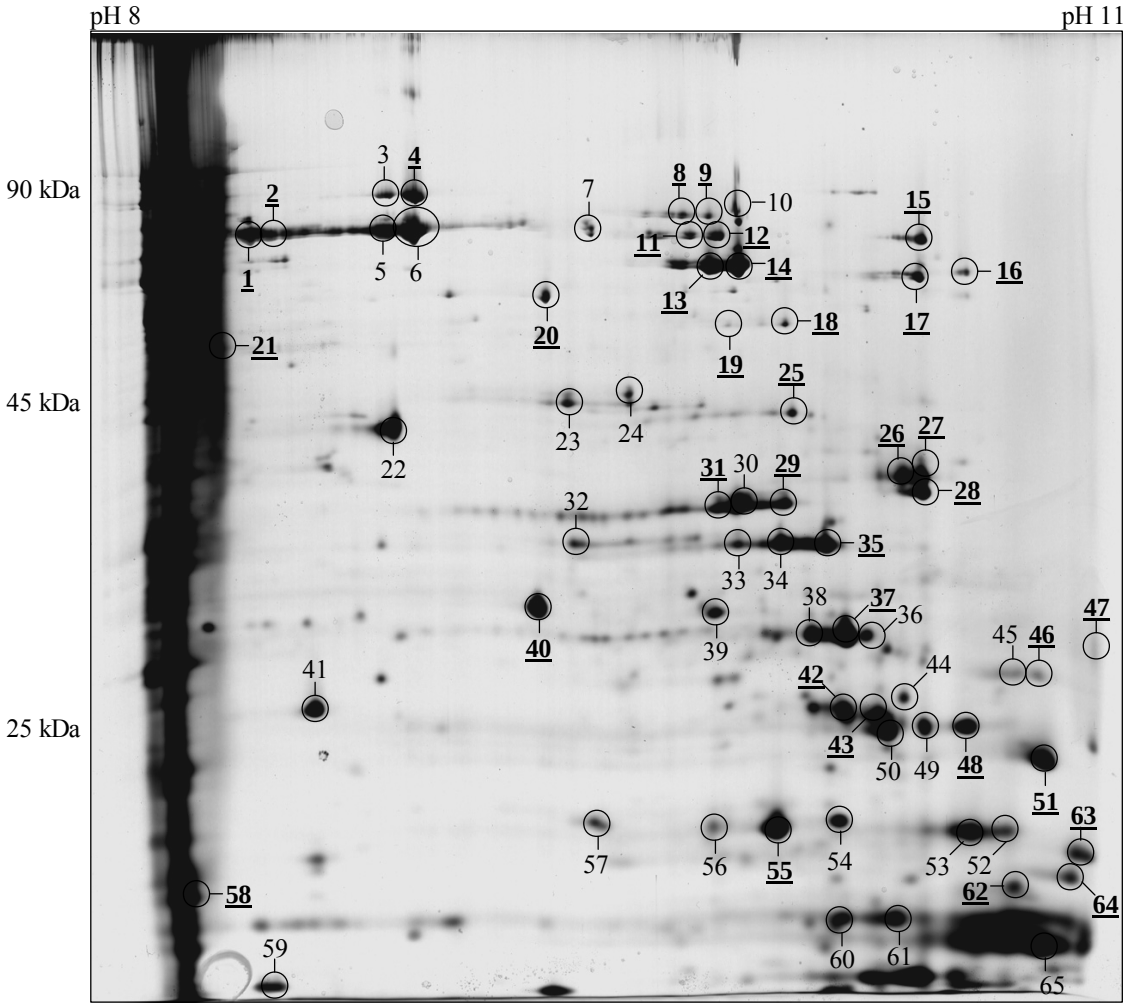
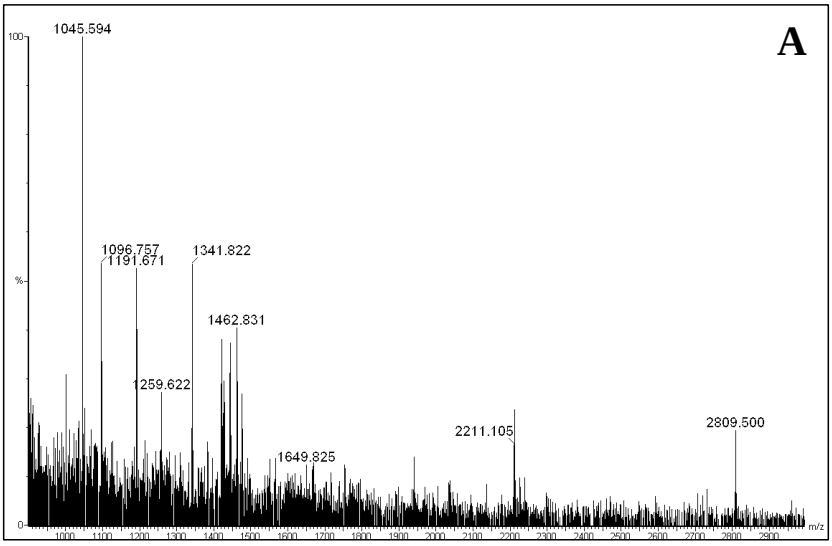
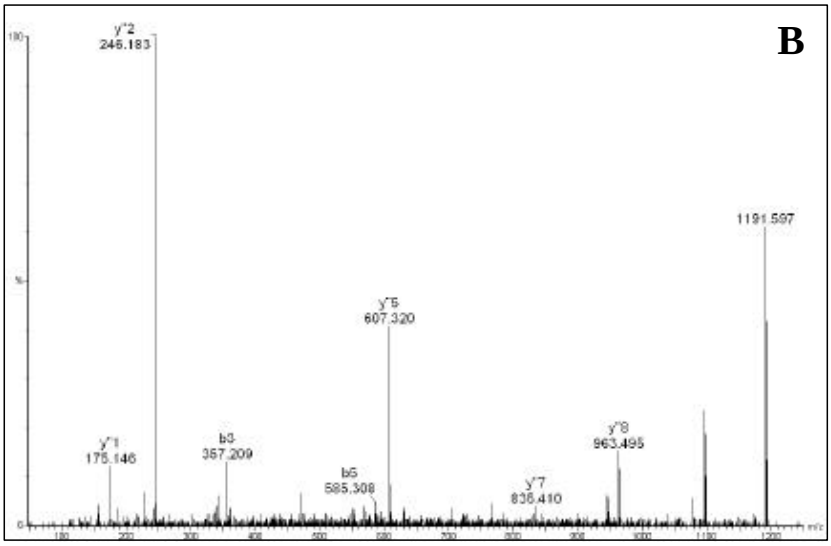

pH 8
pH 11
90 kDa
45 kDa
25 kDa
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Figure 3.5: Spectra obtained from spot 18. MS spectrum (A) and MS/MS spectrum (B) of the *m/z* 1191.6 peptide fragment from spot 18 (B) both using the Micromass MALDI-Q-Tof instrument. MS spectrum (C) for spot 18 obtained with the M@LDI-R.


A
1045.594
1096.757
1191.671
1341.822
1462.831
1259.622
1649.825
2211.105
2809.500


B
y"2
246.183
1191.597
y"5
607.320
y"1
175.146
b3
357.209
b5
585.308
y"7
836.410
y"8
963.495

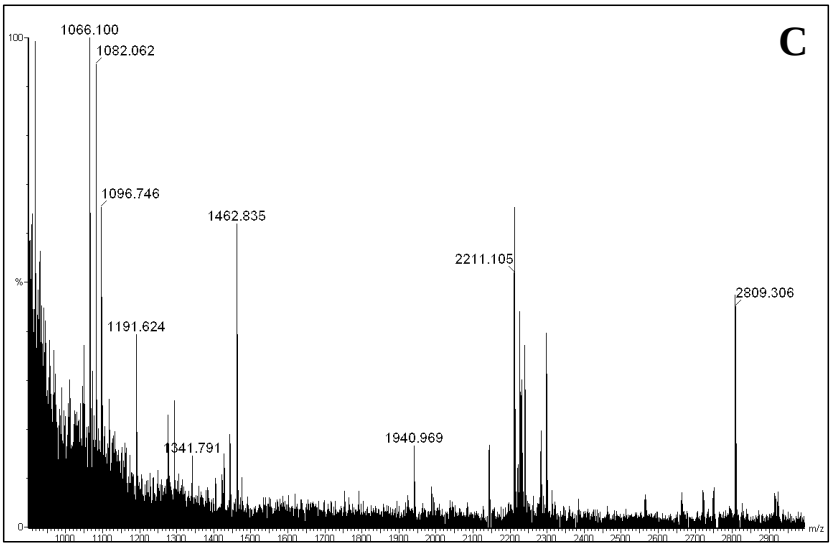

C
1066.100
1082.062
1096.746
1462.835
2211.105
2809.306
1191.624
1341.791
1940.969

Table 3.1. Preparation of pH 8-11 IPGs

Chemicals	Acrylamide gels pH8	Acrylamide gels pH 11[a]	AAEE gels pH8	AAEE gels pH 11[a]	DMA gels pH8	DMA gels pH 11[a]
IPG monomer pK 3.6	842 µl	216 µl	842 µl	216 µl	842 µl	216 µl
IPG monomer pk 8.5	420 µl	300 µl	420 µl	300 µl	420 µl	300 µl
IPG monomer pK 9.3	107 µl	108 µl	109 µl	110 µl	111 µl	112 µl
IPG monomer pK 10.3	319 µl	339 µl	319 µl	339 µl	319 µl	339 µl
IPG monomer pK F 12	100 µl	219 µl	100 µl	219 µl	100 µl	219 µl
Acrylamide / Bis (38.8/1.2)	1500 µl	1500 µl	0 µl	0 µl	0 µl	0 µl
AAEE/Bis (38.8/1.2)	0 µl	0 µl	1875 µl	1875 µl	0 µl	0 µl
DMA/Bis (38.8/1.2)	0 µl	0 µl	0 µl	0 µl	1875 µl	1875 µl
Deionized water	9.26 ml	12.25 ml	8.88 ml	11.87 ml	8.88 ml	11.87 ml
Glycerol (100%)	3.0 g	0 g	3.0 g	0 g	3.0 g	0 g
TEMED (100%)	6.1 µl	6.1 µl	6.7 µl	6.7 µl	6.7 µl	6.7 µl
Ammonium persulfate (10%)	66 µl	66 µl	66 µl	66 µl	66 µl	66 µl
Final volume	15.0 ml	15.0 ml	15.0 ml	15.0 ml	15.0 ml	15.0 ml

[a] pH 11 solutions are adjusted to pH 7 with 50% acetic acid before polymerization

Table 3.2: Proteins identified by PMF using PE Biosystems DE-STR MALDI-TOF

Spot No.	ORF No.	Gene name	MOWSE	Peptides Matched	Cover-age %	ORF description	Putative final mass/pI	Predicted localization
1	ORF 01077	CC3146	9.06E+04	11	14	Outer membrane receptor proteins, mostly Fe transport (TonB-dependent receptor)	86646.7/7.82b	OM
2	ORF 01077	CC3146	1.29E+09	15	26	Outer membrane receptor proteins, mostly Fe transport (TonB-dependent receptor)	86646.7/7.82b	OM
4	ORF 01077	CC3146	1.36E+08	17	30	Outer membrane receptor proteins, mostly Fe transport (TonB-dependent receptor)	86646.7/7.82b	OM
8	ORF 05578	CC1689	5.18E+08	16	24	Organic solvent tolerance protein OstA a)	85086.5/8.90b	OM
11	ORF 03898	CC0815	9.58E+08	17	28	Outer membrane receptor proteins, mostly Fe transport	76503.6/9.03b	OM
12	ORF 03898	CC0815	3.97E+09	19	30	Outer membrane receptor proteins, mostly Fe transport	76503.6/9.03b	OM
12	ORF 00777	CC3001	7.70E+05	11	16	Outer membrane receptor proteins, mostly Fe transport (TonB-dependent receptor)	08679.5/8.52b	OM
13	ORF 05752	CC1778	3.62E+09	16	31	Outer membrane receptor proteins, mostly Fe transport (TonB-dependent receptor)	78252.1/9.48	OM
14	ORF 05752	CC1778	1.95E+08	15	26	Outer membrane receptor proteins, mostly Fe transport (TonB-dependent receptor)	78252-1/9.48	OM
15	ORF 06147	CC1996	1.60E+07	13	29	Membrane proteins related to metallo-endopeptidases (peptidase, M23/M37 family)	60641/9.76b)	Unclear
16	ORF 03596	CC0663	5.41E+13	26	52	Outer membrane receptor proteins, mostly Fe transport (TonB-dependent receptor)	66094.7/9.55b	OM

17	ORF 03596	CC0663	3.83E+16	30	61	Outer membrane receptor proteins, mostly Fe transport (TonB-dependent receptor)	66094-7/9.55b	OM
19	ORF 05429	CC1603	6.36E+09	21	39	Predicted outer membrane protein a)	68203.3/9.36	OM
20	ORF 00474	CC1603	5.62E+06	12	28	Predicted aminopeptidases a)	57622.6/8.66b	IM/Periplasm
21	ORF 01790	CC3500	2.52E+11	20	37	Outer membrane receptor proteins, mostly Fe transport (TonB-dependent receptor)	72484.2/6.65b	OM
27	ORF 05264	CC1515	1.89E+06	10	41	Phosphate ABC transporter, periplasmic phosphate-binding protein PstS, putative	34496.6/9.65b	Periplasm
28	ORF 92868	CC0286	4.99E+07	10	34	Periplasmic sulfate/thiosulfate binding proteins (sulfate-binding protein (sbp))	34759.6/9.69b	Periplasm
29	ORF 01683	CC3448	1.59E+05	9	38	FoF1-type ATP synthase gamma subunit (atpG)	31291.1/9.23	IM
37	ORF 03553	CC0640	6.08E+07	16	77	Ribosomal protein L1 (rplA)	23735.5/9.52	Cytoplasm
40	ORF 05510	CC1652	7.10E+04	9	55	TPR-repeat-containing proteins a)	26485.4/8.61	IM
46	ORF 04755	CC1249	5.86E+04	8	48	Ribosomal protein L4 (rplD)	22792.5/10.30	Cytoplasm
51	ORF 04777	CC1263	4.19E+03	6	38	Ribosomal protein L6 (rplF)	18886.2/10.07	Cytoplasm
55	ORF 03279	CC0496	1.57E+04	6	47	Ribosomal protein L10 (rplJ)	18039.1/9.27	Cytoplasm
58	ORF 01250	CC3229	8.70E+03	6	38	Flagellar motor protein MotB and related proteins, PAL/OmpA family a)	17220.2/9.15b	OM
62	ORF 04794	CC1273	2.94E+03	6	44	Ribosomal protein L17 (rplQ)	15325.8/10.09	Cytoplasm
63	ORF 01188	CC3201	4.90E+05	9	67	Ribosomal protein S7 (rpsG)	18001.7/10.23	Cytoplasm
64	ORF 04783	CC1267	2.63E+03	7	39	Ribosomal protein L15 (rplO)	16924.8/10.41	Cytoplasm

[a] Protein description from COG function.
[b] Indicates presence of predicted Signal Peptide. Theoretical MW/pI after removal of signal peptide.

Table 3.3: Proteins identified by PMF using Micromass MALDI-Q-Tof.

Spot No.	ORF No.	Gene name	MQT % probability	Peptides Matched	Cover-age %	ORF description	Putative final mass/pI	Predicted localization
9	ORF 05578	CC1689	100	27 (1 ox)	35	Organic solvent telerance protein OstAa)	85086.5/8.90[b]	OM
18	ORF 02084	CC3653	1.3	16 (4 ox)	31	Signal recognition particle protein (ffh)/ Signal recognition particle GTPase	53432.9/9.39	Unclear
25	ORF 01338	CC3273	3.2	16 (2 ox)	35	Hypothetical protein	45761.5/9.19[b]	IM
31	ORF 01683	CC3448	99.5	10 (1 ox)	30	FoF1-type ATP synthase gamma subunit (atpG)	31291.1/9.23	IM
35	ORF 04754	CC1248	73	5	15	Ribosomal protein L3 (rplC)	28079.9/9.49	Cytoplasm
40	ORF 05510	CC1652	99.9	7 (3 ox)	29	RPR-repeat-containing proteinsa)	26485.4/8.61	IM
42	ORF 07117	CC2493	58.7	6 (1 ox)	19	Transcriptional regulator, TetR family	24493.1/5.61	IM
43	ORF 07173	CC2523	25.4	6	27	Hypothetical protein	26257.4/10.08	IM
43	ORF 04773	CC1260	12.1	5 (2 ox)	25	Ribosomal protein L5 (rplE)	20965.4/9.52	Cytoplasm
48	ORF 01688	CC3450	93.1	6 (1 ox)	29	ATP synthase F1, delta subunit (atpH)/FoF1-type ATP synthase delta subunit	19181.1/9.84	IM

[a] Protein description from COG function.
[b] Indicates presence of predicted Signal Peptide. Theoretical MW/pI after removal of signal peptide.

Table 3.4: Proteins identified by PMF using Micromass MALDI-Q-Tof.

Spot No.	ORF No.	Gene name	MS/MS #1	MS/MS #2	MS/MS #3	MS/MS Score	Peptides	ORF description	Putative final mass/pI	Predicted localization
9	ORF 05578	CC1689	1012.5	1767.8	–	56.8	2	Organic solvent tolerance protein OstAa)	85086.5/8.90b)	OM
18	ORF 02084	CC3653	1096.7	1191.625	1341.77	55.6	3	Signal recognition particle protein (ffh)/Signal recognition particle GTPase	53432.9/9.39	Unclear
25	ORF 01338	CC3273	1291.7	1383.7	–	52.9	2	Hypothetical protein	45761.5/9.19b)	IM
31	ORF 01683	CC3448	1329.1	–	–	23.1	1	FoF1-type ATP synthase gamma subunit (atpG)	31291.1/9.23	IM
35	ORF 04754	CC1248	1200.6	1264.64	–	72.5	2	Ribosomal protein L3 (rplC)	28079.9/9.49	Cytoplasm
40	ORF 05510	CC1652	971.5	1428.722	–	42.7	2	RPR-repeat-containing proteins)	26485.4/8.61	IM
43	ORF 04773	CC1260	1439.77	–	–	21.6	1	Ribosomal protein L5 (rplE)	20965.4/9.52	Cytoplasm
47	ORF 04762	CC1254	1185.5	1183.5	–	69.3	1	Ribosomal protein S3 (rpsC)	28139.4/10.36	Cytoplasm
48	ORF 01688	CC3450	1156.6	1248.6	–	54.6	1	ATP synthase F1, delta subunit (atpH)/FoF1-type ATP synthase delta subunit	19181.1/9.84	IM

[a] Protein description from COG function.
[b] Indicates presence of predicted Signal Peptide. Theoretical MW/pI after removal of signal peptide.

Table 3.5: Proteins identified by PMF using Micromass M@LDI-R

Spot No.	ORF No.	Gene name	M@LDI % probaliity	Peptides Matched	Cover-age %	ORF description	Putative final mass/pI	Predicted localizatio n
9	ORF 05578	CC1689	100	33 (2 ox)	49	Organic solvent tolerance protein OstAa)	85086.5/8.90b)	OM
18	ORF 02084	CC3653	82.4	15 (2 ox)	28	Signal recognition particle protein (ffh)/Signal recognition particle GTPase	53432.9/9.39	Unclear
25	ORF 01338	CC3273	0.6	7(1 ox)	17	Hypothetical protein	45761.5/9.19b)	IM
26	ORF 05264	CC1515	79.6	7(1 ox)	23	Phosphate ABC transporter, periplasmic phosphate-binding protein PstS, putative	34496.6/9.66b)	Periplasm
31	ORF 01683	CC3448	100	14	41	FoF1-type ATP synthase gamma subunit (atpG)	31291.1/9.23	IM
31	ORF 01177	CC3196	0	7(2 ox)	68	Membrane permeases, predicted cation efflux pumps (HlyD family secretion protein)	34145.5/9.55	IM
35	ORF 04754	CC1248	100	6(1ox)	32	Ribosomal protein L3 (rplC)	28079.9/9.49	Cytoplasm
40	ORF 05510	CC1652	100	14 (5 ox)	46	TPR-repeat-containing proteinsa)	26485.4/8.61	IM
42	ORF 02425	CC0072	73.4	7(1 ox)	26	Porphobilinogen deaminase (hemC)	30835.3/5.02b)	Cytoplasm
47	ORF 04762	CC1254	100	13 (2 ox)	47	Ribosomal protein S3 (rpsC)	28139.4/10.36	Cytoplasm
48	ORF 01688	CC3450	38.4	8	49	ATP synthase F1, delta subunit (atpH)/FoF1-type ATP synthase delta subunit	19181.1/9.84	IM

[a] Protein description from COG function.
[b] Indicates presence of predicted Signal Peptide. Theoretical MW/pI after removal of signal peptide.

Table 3.6: Functional characterization of identified proteins based on COGs

COG functional category	No of identified proteins
Translation (J)	10
Transcription (K)	1
Cell envelope biogenesis, outer membrande (M)	3
Cell motility and secretion (N)	2
Inorganic ion transport and metabolism (O	8
Energy production and conversion (C)	2
Coenzyme metabolism (H)	1
General function prediction only (R)	3
Not in any COG	2

CHAPTER IV

PROTEOMIC ANALYSIS OF THE SPORE COATS OF *BACILLUS SUBTILIS* AND *BACILLUS ANTHRACIS*

This chapter describes the progress made towards the analysis of the *Bacillus subtilis* spore coat proteome, as well as the preliminary analysis of the *Bacillus anthracis* genome and spore coat proteome. The work was carried out as a collaborative project with Erh-Min Lai. My contributions were the design of the sample preparation methodology, and the 2-DE gel electrophoresis. Mass spectrometry and data analysis were carried out jointly. The work was published in its entirety in the Journal of Bacteriology (2003, *185* (4),1443–1454) titled "Proteomic Analysis of the Spore Coats of *Bacillus subtilis* and *Bacillus anthracis*" by Erh-Min Lai, Nikhil D. Phadke, Maureen T. Kachman, Rebecca Giorno, Santiago Vazquez, Jenny A. Vazquez, Janine R. Maddock, and Adam Driks, as an equal contribution between the first two authors.

Abstract

The outermost proteinacious layer of bacterial spores, called the coat, is critical for spore survival, germination and, for pathogenic spores, disease. To identify novel spore coat proteins, we have carried out a preliminary proteomic analysis of *Bacillus subtilis* and *B. anthracis* spores, using a combination of standard SDS-PAGE separation and improved 2-dimensional electrophoretic separations, followed by MALDI-TOF and/or MS/MS mass spectrometry. We identified 38 *B. subtilis* spore proteins, 12 of which are known coat proteins. We propose that of the novel proteins, YtaA, YvdP and YnzH are *bona fide* coat proteins and have renamed them CotI, CotQ and CotU, respectively. In addition, we initiated a study of coat proteins in *B. anthracis* and identified 11 spore proteins, 6 of which are candidate coat or exosporium proteins. We also queried the unfinished *B. anthracis* genome for potential coat proteins. Our analysis suggests that the *B. subtilis* and *B. anthracis* coats have roughly similar numbers of proteins and that a core group of coat protein species is shared between these organisms, including the major morphogenetic proteins. Nonetheless, a significant number of coat proteins are probably unique to each species. These results should accelerate efforts to develop *B. anthracis* detection methods and understand the ecological role of the coat.

Introduction

Bacteria build a variety of precisely positioned macromolecular structures. These include the divisome, which mediates cell division (43), the flagellum (2) and the type III secretory apparatus, critical for pathogenesis in many organisms (40, 42). To understand how such structures are built and how they function, it is critical to identify their protein components. In this report, we seek to identify the proteins that make up the coat that encases spores of *B. subtilis* and *B. anthracis*.

Spores are produced by many species of Bacilli and Clostridia in response to severe external stress (72, 73). These highly resilient dormant cell types are able to withstand extremes of temperature, radiation, chemical assault, time and even the vacuum of outer space (58). Upon the return of favorable environmental conditions, spores can readily convert to actively growing vegetative cells through a process known as germination (48, 65). These abilities enable spores not only to survive in extreme conditions but, in some species, to cause significant disease. In the case of *B. anthracis*, contact between the host and the spore is essential for infection (17). The most prominent structural feature common to all bacterial spores is a multilayered proteinaceous shell called the coat (19, 21, 31). The coat is critical for resistance properties as well as germination. It provides mechanical integrity and excludes large toxic molecules while, at the same time, allowing small nutrient molecules to penetrate and interact with the germination receptors located toward the spore interior (49, 65).

Thin-section electron microscopy of the *B. subtilis* spore reveals two major layers in the coat: a lightly staining lamellar inner layer and a darkly staining outer layer (4, 81). The morphogenetic coat protein CotE directs the assembly of most, if not all, outer coat proteins and some of the inner coat proteins (23). Additional morphogenetic proteins, such as SafA (YrbA) and SpoVID, further guide coat protein deposition (8, 63, 64, 78). The *B. anthracis* coat is much thinner than that of *B. subtilis*, and its layered architecture is less striking, although inner and outer coat layers can still be discerned (20, 28, 30, 47, 67). In many Bacilli, including *B. anthracis* but not *B. subtilis*, the spore is encased by an additional structure known as the exosporium. The exosporium surrounds the entire spore, including the coat, from which it is structurally and biochemically distinct (45). Notably, the exosporium is heavily glycosylated (26), appears to be composed of relatively few protein species, and harbors at least one major glycoprotein (74). Its function is unknown although it appears not to have a major role in pathogenesis (74).

Approximately 30 confirmed or putative *B. subtilis* coat proteins have been identified, but how they participate in spore survival and germination remains, for the most part, obscure (21). A subset of coat proteins has significant, although poorly understood, roles in assembly. Importantly, deletion of any one of the large number of remaining coat proteins has little or no detectible phenotype. As a result, the functions of most of the coat proteins are unknown and, furthermore, most roles of the coat cannot be ascribed to specific proteins. It is very likely that many coat functions are emergent properties of the interactions of multiple coat proteins (80). Therefore, a detailed understanding of the coat's role may require knowledge of most or all of its components.

Even though a large number of coat proteins has been found, it is clear that additional coat proteins remain undetected (19). Identification of these proteins will require a comprehensive characterization of the coat at the molecular level, which has not yet been achieved.

Proteomics offers a powerful platform for the analysis of components of many macromolecular assemblies through separation of complex protein mixtures by 2-dimensional (2D) electrophoresis and identification by mass spectrometry (MS) (see, for example, (27, 33)). Several excellent proteomic studies of *B. subtilis* have been carried out. Notable amongst these are a recent alkaline two-dimensional map and a comprehensive two-dimensional map (12, 61). These studies focused on the analysis of soluble proteins from vegetative cells. In contrast, analysis of spore proteins has been hindered by the requirement for solublization of tightly associated coat proteins (4). Recent advances in protein solubilization and separation made in our laboratories (51, 52, 66), coupled with significant increases in mass spectrometric sensitivity and the availability of genomic sequence data from several organisms, have made the identification of proteins from hard-to-dissect biological samples more realistic. Here, we take advantage of these technological improvements to identify spore proteins in *B. subtilis* and *B. anthracis*.

Materials and Methods

General methods

Bacillus subtilis (PY79 (82)) or *Bacillus anthracis* (Sterne strain, from Paul Jackson, Los Alamos National Laboratory) spores were prepared essentially identically by exhaustion in Difco Sporulation Medium (DSM) (16). The only difference was that for the case of *B. anthracis*, initial growth was on LB agar plates supplemented with 5 g of nutrient broth (Difco) per liter. Spores were washed three times in double distilled H_2O (ddH_2O) and then resuspended in ddH_2O. Resuspended spores were either used immediately or stored at 4°C for no longer than three weeks. For 1D gel electrophoresis of *B. subtilis* spore extracts, we prepared protein samples as described (41).

2D electrophoresis

2D gel electrophoresis was performed as described (66) with the following modifications during the isoelectric focusing step. 45 μl of the boiled 2% SDS (sodium dodecyl sulfate) or 2% LDS (lithium dodecyl sulfate) solubilized suspension was mixed by sonication with 405 μl of 1.1X rehydration buffer [7.7 M urea, 2.2 M thiourea, 2.2 mM tributyl phosphine (TBP), 0.55% (v/v) Biolytes 3-10 (Bio-Rad, CA, USA), and 44 mM Tris-base] containing 2% benzoylamidopropyldimethylammoniopropanesulfonate (C8ϕ) and 1% Triton X-100 (TX-100) and allowed to incubate with constant mixing at 30°C. Samples were spun at 13,000g to remove insoluble material. Immobilized pH Gradients (IPGs) (18 cm; Amersham Pharmacia, Sweden) were rehydrated overnight in 450 μl of solubilized spore suspension in rehydration buffer. Isoelectric focusing was

carried out as described (66) with the addition of frequent filter paper wick changes during isoelectric focusing.

MALDI-MS and MS/MS analysis

Manually excised Coomassie protein spots from 1D gels were digested with 150 – 300 ng modified porcine trypsin (Promega, Madison, WI) in 30 μl of freshly prepared 100 mM ammonium bicarbonate at 37°C overnight as described (66). Protein spots from 2D gels were manually excised from the gel using a 2 mm dermal punch and transferred to a 96 well plate. The spots were automatically digested using a Micromass® MassPREP™ Station (Micromass Ltd., Manchester, UK). Briefly, the silver stained gel pieces were destained with two washes of a 1:1 mixture of 30 mM potassium ferricyanide and 100 mM sodium thiosulfate followed by dehydration of the gel pieces with acetonitrile. Protein spots were reduced and alkylated with 10 mM DTT and 55 mM iodoacetamide, both buffered with 100 mM ammonium bicarbonate, followed by a wash with 100 mM ammonium bicarbonate and two dehydration cycles with acetonitrile. Protein digestion was performed with the addition of 25 μl of 6 ng/μl sequencing grade modified trypsin solution in 100 mM ammonium bicarbonate (Promega, Madison, WI) and incubation of the solution at 37˚C for 4.5 hours. The resulting peptides were extracted in an aqueous solution of 1% (v/v) formic acid, 2% (v/v) acetonitrile. 2.0 μl of extracted peptides and 1.6 μl of α-cyano-4-hydroxycinnamic acid (10 mg/ml in 50% acetonitrile/H_2O and 0.1% TFA) were mixed in the dispensing tip and spotted onto a Micromass® 96 well MALDI target plate.

Analysis using the PerSeptive Biosystems Voyager-DE-STR was as described (66). The resulting peptide mass fingerprints were searched using a local copy of the program MS-Fit (a component of the Protein Prospector package) against an internal database as described below. For peptide mass fingerprinting searches, MALDI spectra were analyzed using the Micromass® MassLynx™ 3.5 software package (Micromass Ltd., Manchester, UK). Spectra were manually recalibrated using the 2211.104 Da and 842.509 Da trypsin autodigestion peptides, and the data were exported to text files. Monoisotopic peptide masses were obtained manually or by using an in-house virtual instrument created in the LabView graphical programming language (G. Rymar and P. Andrews, unpublished). The resulting peptide mass fingerprints were searched using a local copy of the program MS-Fit against an internal database. A mass accuracy of 75 ppm was used. A maximum of 1 missed enzymatic cleavage, and modification of cysteines by carbamidomethylation plus possible modification by acrylamide were considered during the searches.

For MS/MS, 0.5 µl of gel digest was spotted onto the target plate followed by 0.5 uL of a-cyano-4-hydroxycinnamic acid (a-CHCA) (5 mg/ml in 50% acetonitrile, 1% trifluoroacetic acid) and the spot was allowed to air dry at room temperature. MS/MS was performed on a Micromass MALDI-QTOF Ultima mass spectrometer. This instrument features an unattenuated nitrogen laser, operating at 337 nm and firing at 10 Hz, which is rastered over the sample spot (2.5 mm diameter) at 1 Hz. Ions are introduced into the instrument by a nitrogen flow (5 psi) that also serves to cool the ions. Ions are selected for MS/MS in the quadrupole, with the mass tolerance set to ± 5 Da.

This window was closed to 2 Da if another peptide occurred within the 5 Da window of the peptide of interest. The selected ion was fragmented in the hexapole collision cell, with the argon gas pressure set to 25 psi and the collision energy varied from 50-150 V according to the mass of the parent ion. The parent and product ions are resolved in the Time of Flight (TOF) region, equipped with a reflectron which produces monoisotopic resolution for ions from 100-4000 Da, and detected by a microchannel plate detector set to 2250 V. Parent ion masses were taken from the peptide mass fingerprint spectrum after internal calibration using trypsin autolysis peaks. Product ion masses were calibrated using instrument calibration only. Parent and product ion masses were submitted to Mascot Ions Search (www.matrixscience.com), searching the NCBInr database for ALL species, for protein identification. The parent ion tolerance was set to 100 ppm and the product ion tolerance to 0.5 Da. Alternative modes of searching (no enzyme, specifying PTMs, searching additional databases) were performed as needed. An identification based on MS/MS on one peptide was considered adequate if the Mascot score was above the significance level, although in many cases, we obtained additional confirmation by MS/MS on a second peptide or from a peptide mass fingerprint database search (Table 1).

Creating a peptide mass fingerprint searchable *B. anthracis* database

The latest releases of the *B. anthracis* genome sequence were made available prior to publication by the Institute for Genomic Research (TIGR, Rockville, MD). Sequence data was in the form of several large contigs in FASTA format. Contigs were assembled into a single contiguous DNA sequence in FASTA format using a script written in the Perl programming language. Putative open reading frame (ORF)

predictions were carried out using the Interpolated Markov Model gene prediction package, GLIMMER (69). Briefly, a trained model was created, using the program 'Build-icm', on a data set of sequences from closely related species available in the NCBI non-redundant database. Complete nucleotide sequences of known and predicted chromosomal ORFs from *B. anthracis*, *B. cereus*, *B. thuringiensis*, *B. halodurans*, and *B. subtilis* were used to train the model. The program 'Glimmer2' was used to predict putative ORFs from the concatenated DNA sequence file based on the trained model. FASTA formatted nucleotide sequences of predicted ORFs were obtained from the output of Glimmer2 using an in-house Perl program. The resulting nucleotide sequences were used to retrain the model using 'Build-icm', and the gene prediction process was repeated with the newly trained model. The predicted ORF nucleotide sequences were translated to amino acid sequences in reading frame 1 using the 'Transeq' program from the EMBOSS suite of molecular biology applications (http://www.emboss.org/). Rapid preliminary annotation was carried out using scripts written in the Python programming language to submit each predicted sequence to BLAST (3), COGNITOR (79), FramePlot (36), SignalP2 (59), ProtParam (http://us.expasy.org/tools/protparam.html), PSORT (57), and SOSUI (32). The translated and annotated sequences were concatenated into a file containing known and predicted protein sequences from from *B. anthracis*, *B. cereus*, *B. thuringiensis*, *B. halodurans*, and *B. subtilis*. The resulting file was indexed for searching with MS-Fit using the 'Faindex' program from the Protein Prospector package.

Results and Discussion

Analysis of *B. subtilis* spore proteins using 1D and 2D electrophoresis

While 1D analysis may not possess the separation power of 2D electrophoresis, it has several advantages over 2D analysis, including the ability to separate hydrophobic proteins such as those found in membranes, the ability to tolerate significantly higher protein loads, and the ability to resolve proteins linearly across the pH scale. Thus, in order to identify proteins difficult to resolve on 2D gels, we first identified proteins in the major bands visualized on Coomassie-stained 1D gels (Fig 4.1). *B. subtilis* spore samples were prepared for SDS-PAGE analysis as described in the materials and methods section. From these studies, we identified 19 bands comprising 27 unique proteins (Table 4.1). Not surprisingly, we identified a number of *bona fide* coat proteins such as CotA (18), CotB (18), CotE (84), CotF (15), CotG (68), CotJC (71), CotR (41), CotS (1), CwlJ (7) and YaaH (38). We also identified YxeE, identified in spores previously and known to be a good coat protein candidate (77). We identified eight spore proteins that are very unlikely to be in the coat [CoxA (YrbB) (75), CspD (29, 70), Hbs (46), PhoA (34), SleB (10, 53-55), SspA (21), SspE (21) and YhcN (6)]. Most importantly, we identified the unknown hypothetical proteins YckK, YdhD, YjdH, YodI, YpeP, YpzA, YtaA, and YusA which we regard as candidate coat proteins.

As anticipated, several Coomassie-staining bands on our 1D gels contained more than one protein (bands 3, 12, 13, 16, 18 and 20; Table 1). Since one long-term goal is to examine how the proteomic profile changes in various coat gene mutant backgrounds and during the course of sporulation, it was critical to develop 2D separation of coat proteins

such that the level of individual proteins could be monitored. We found that conventional solubilization techniques (60) and even aggressive solubilization techniques designed for separation of membrane proteins by 2D electrophoresis, involving strong chaotropes, amidosulfobetaine detergents and strong reducing agents (14, 50, 66) proved ineffective for solubilization of *B. subtilis* coat-enriched fractions. Large insoluble pellets remained after centrifugation of coat samples solubilized under these conditions (data not shown). Use of the ionic detergents SDS and LDS or chaotropes like guanidium-HCl resulted in much better solubilization. However, these reagents are incompatible with IEF at the concentrations used for solubilization. Several groups have suggested solubilization by these strong agents followed by dilution or detergent exchange with IEF compatible reagents (25, 50). We investigated the use of strong ionic detergent solubilization followed by dilution with large volumes of IEF compatible reagents as a method for preparing 1st dimension samples. Coat-enriched *B. subtilis* fractions were solubilized in a modified 1D loading buffer containing SDS or LDS. These samples were subsequently resolubilized in an excess of IEF compatible reagents and subjected to 2D electrophoresis. Excellent separation was obtained when IEF was carried out for extended periods up to 120 kVh, with frequent wick changes to remove unfocussed material. The best results were seen when initial solubilization was carried out on very fresh samples using small volumes of SDS buffer, followed by dilution with standard rehydration buffer containing 2% C8ϕ + 1% Triton X-100 (Figs. 2 and 3). We identified 19 spots consisting of 14 unique proteins from pH 3-10 gels (Fig. 4.2, Table 4.2). Not surprisingly, we found a number of proteins also found in the 1D gels such as CotA, CotE, CotJC, and YaaH as well as coat proteins not found on the 1D gel (CotC and CotY). A protein unlikely to be a

coat protein, CggR (24), was also identified. Finally, we identified YhbA, YirY, YisY, YnzH, YopQ, YvdP, and YwqH, which we regard as candidate coat proteins.

B. subtilis coat proteins

About 30 coat proteins or coat protein candidates have been described (21). We found 12 of these known proteins and 17 additional coat proteins or coat protein candidates. Of these, we regard YtaA, YvdP and YnzH as highly likely to be coat proteins and we have renamed them CotI, CotQ and CotU, respectively. We regard YtaA as a coat protein based on 1) significant similarity to CotS (BLAST score of 3e-25), as already noted in the *B. subtilis* genome sequence annotation, and 2) its gene is adjacent to, and oriented away from, the operon harboring *cotS*. We had already identified YvdP in a separate study and showed that it is synthesized at a late time in sporulation and associated with the spore in a manner dependent on the coat morphogenetic protein CotE (41, 62, 84). Furthermore, its gene is adjacent to *cotR* and transcribed divergently from it. *cotR* (41) and *cotQ* could share upstream regulatory sequences. We regard YnzH as a coat protein because we have also shown that it is synthesized late in sporulation and is assembled in a CotE-dependent manner (62). As was previously noted in the *B. subtilis* genome sequence annotation (http://genolist.pasteur.fr/SubtiList/), the YnzH sequence is strikingly similar to that of CotC.

Of the 14 additional coat protein candidates, we predict that many will also be *bona fide* coat proteins. Of particular interest is YisY, very likely to be a chloride peroxidase, CotQ, highly similar to reticuline oxidase from plants, and YdhD, predicted to be a cortex lytic enzyme (39). Precedence for a cortex lytic enzyme in the coat is

provided by CwlJ (7). Several other uncharacterized proteins identified in our study are also good coat protein candidates. For example, YhbA is an iron-sulfur cluster-binding protein and likely the first gene in a 6-gene operon. Most of the other genes in the operon (*yhbB*, *yhbD*, *yhbE* and *yhbF*) are found only among the Bacilli and Clostridria, consistent with a role in sporulation. YpeP is a novel protein that appears to be cotranscribed with *ypeQ* and transcribed divergently from the operon containing *ypdP* and *ypdQ*. YpdQ is predicted to be a cell wall synthesis protein. Interestingly, the gene downstream of the *ypeP* operon encodes YpzA, another protein identified in our study.

The possibility that CotQ and YisY are oxidases is notable in light of the recent demonstration that CotA is a multicopper oxidase of the laccase class (35, 44) and the similarity of CotJC to peroxidase (71). Oxidases can participate in a wide variety of biosynthetic activities. Potentially, they could play a role in coat protein crosslinking, as has been suggested previously (4, 19, 31), in detoxification of environmental contaminants or in symbiosis with other soil organisms (22). The identification of these enzymes in the coat is consistent with the view that the coat plays active roles in spore protection and germination, rather than acting solely as a passive barrier (21, 23). Whatever their roles, CotA (18), CotJC (71) and CotQ (62) are dispensable for coat resistance and germination, as usually measured in the laboratory.

We found 12 previously identified *B. subtilis* coat proteins (Tables 4.1 and 4.2). Several coat proteins are known to be present in a form smaller than the presumed translation product, including CotF, CotT and SafA (5, 11, 15, 76). Consistent with this

and extending the list of processed coat proteins, we found that CotA, CotC, CotE and CotF were present in multiple forms. The two forms of CotF are the known mature forms (15). CotC was present in several spots of different mobility and with a charge very different than its predicted pI of 8.6 (Fig. 4.2), possibly the result of proteolysis or modification during coat assembly. Likewise, we found multiple forms of CotE. YjdH, YodI and YirY are three additional proteins that may also be processed based on their migration (Figs. 4.1 and 4.2). YirY is a particularly striking example. YirY is predicted to be a 129 kDa protein that migrates at ~50 kDa (Fig. 4.2). We predict that either YirY is processed or that the original annotation that separated YirY into two distinct proteins (YirY and YirZ) is, in fact, correct. Finally, some proteins migrate slower than predicted, including CotA, CotB, CotG and YpzA (Figs. 4.1 and 4.2). These proteins may be modified or crosslinked to additional protein.

Deletions of most of the known coat proteins have minimal or undetectable phenotypes in standard laboratory assays, as already pointed out (21). Given the complexity of coat composition and the likely enzymatic roles of some of the components, it seems reasonable to suppose that the coat has functions beyond those measured in the laboratory. These may include roles for coat-associated enzymes in coat protein crosslinking and degradation of environmental toxins, symbiosis with plants and protection against competitor organisms. An additional important function of the coat is in germination. While this role has been appreciated for some time (4), it is only recently that its molecular basis has begun to become understood. For example, in addition to the discovery that the cortex lytic enzyme CwlJ is in the coat (7), the Moir laboratory has

shown that the *gerP* operon, likely encoding coat protein(s), participates in germination (9). Specifically, that study indicates that the coat does more than merely act as a passive sieve through which germinants flow.

We note that less than half of the known coat proteins were found in our experiments. This is likely to be for several reasons. First, we have not exhausted our analysis of faint bands present on 1D gels (Fig. 4.1) or of focused spots on the 2D gels (Fig. 4.2). As MS/MS technologies become more sensitive, we are certain that we will also identify these proteins. Second, the identification of any protein by mass spectrometry relies on the ionization efficiency of the trypsin fragments and the sensitivity of the mass spectrometers. MALDI-MS mapping is sensitive but requires multiple peptides for identification. MS/MS identification using the QTOF requires fewer peptides but often a stronger ionization signal. Thus, low abundance proteins or those with only a few detected peptides may be missed. Third, many of the coat proteins are basic and would not be resolved on the 2D gels used in this study. Although the theoretical separation on 3-10 IPGs should allow for separation of proteins up to pI 10, we rarely focus on proteins with pIs greater than 8 using these IPGs (51, 66). The development of separation procedures for alkaline coat proteins, similar to what we have achieved for membrane proteins (51), will be required to visualize these proteins. Fourth, small proteins (such as the 29 amino acid SpoVM) will be absent from the 2D gels and are not amenable to MALDI-MS mapping, as the required number of peptides to identify the protein may not be available. Finally, some of the covalently crosslinked coat proteins may not have been solubilized in this study and would, therefore, be missed altogether.

Despite these limitations, we have identified many of the known coat proteins and have uncovered 14 new coat candidates (Table 4.3). Our survey is clearly less than saturated, and it is likely that a significant number of novel coat protein species remain to be discovered. Future analysis by MS/MS coupled with the use of alkaline 2D gels (51) should readily lead to a proteome map of most, if not all, spore proteins. Such a proteomic map will be instrumental for biochemical characterization of coat assembly.

Analysis of *B. anthracis* spore proteins using 1D electrophoresis

In spite of its importance to spore survival, germination and pathogenesis, the protein composition of the *B. anthracis* coat has received very little study. None of the coat proteins in this organism have been identified, nor has coat assembly been characterized, although one exosporium protein has been characterized (74) and some spore-associated proteins have been identified in the close relative *B. cereus* (13). Comparison of the *B. subtilis* and *B. anthracis* genomes shows that coat proteins with key roles in morphogenesis are present in both organisms (Table 4.3) and, therefore, it is plausible that coat assembly follows largely the same program in the two species (20). Interestingly, this same analysis suggested there are important differences in the protein compositions of the coats of these organisms, possibly among the outer layers. Identification of novel coat proteins in *B. anthracis* will not only help in understanding coat assembly and function but could also be of significant value in development of vaccines and methods for decontamination, detection and inhibition of germination.

Using standard 1D SDS-PAGE with 15% gels, we identified 11 *B. anthracis* spore proteins. We consider ORF1326 (similar to CotJC), NP_654287 (more distantly

related to CotB and YwrJ) and NP_655129 (related to CotY and CotZ) to be coat proteins. NP_657011 encodes a protein that is 49% similar to amino acids 23-173 of the predicted 404 amino acid *B. subtilis* protein YndF. NP_657011 also resembles GerBC and GerBA (27% and 31% identical, respectively). NP_654944 encodes an unknown protein closely related to a Clostridium protein and therefore is a good candidate for a coat or exosporium protein. Finally, we consider NP_655132 to be a coat protein as the gene is flanked by genes whose products we predict will be coat proteins (NP_655129 and ORF1889, Table 4.3). In this preliminary study, we did not detect several expected *B. anthracis* coat proteins (based on analysis of the *B. subtilis* genome, see Table 4.3) and, therefore, we anticipate that a significant number of coat proteins have yet to be identified. Taken together, the SDS-PAGE analysis (Fig. 4.4) and the genomic comparisons (Table 4.3) suggest that the number of coat proteins in *B. subtilis* and *B. anthracis* are similar, although we predict that spores from each organism will also contain species-specific proteins.

From our initial MALDI mapping of a 1D gel, we identified a homolog of CotJC and proteins related to CotB and CotZ. The CotZ-like protein has two electrophoretic variants (Fig 4.2), which may indicate processing or protein modification. In *B. subtilis*, CotB is on the spore surface (37), CotJC is likely to be in the inner coat (71), and CotZ is part of a relatively insoluble portion of the coat (83). Therefore, if these features hold true in *B. anthracis*, our approach succeeded in extracting proteins at different positions in the coat and included a protein that we predict will be relatively highly crosslinked. Interestingly, the *B. anthracis* CotB-like protein is predicted to be much smaller than the

B. subtilis homolog (19 kDa and 43 kDa, respectively) and, therefore, these proteins may function differently in the two species. The smaller size is not likely due to misannotation, as the *B. anthracis* protein migrates at 21 kDa (Fig. 4.4) whereas in *B. subtilis*, CotB migrates at 62 kDa (Fig. 4.1). Among the novel proteins identified, we predict that some will be *B. anthracis*-specific coat proteins. Although we cannot yet confirm that any of these candidate proteins are in the coat, our general success at extracting *bona fide* coat proteins, coupled with the likelihood that some *B. anthracis* coat proteins are not present in *B. subtilis* (20), suggests that at least some of these candidates are coat components. 9 of the novel proteins we identified in *B. subtilis* have homologs in *B. anthracis* (Table 4.3) and, therefore, we regard them as candidate *B. anthracis* coat proteins as well. As for *B. subtilis* coat proteins (see above), we anticipate that additional coat proteins remain to be found.

The complexity of the coat suggests it has important roles in adaptations to the diverse niches in which Bacilli flourish and argues that much more sophisticated tests of coat function are needed. Likewise, much deeper investigations of the ecology of spore-formers will be required if we are to understand the evolutionary pressures that resulted in a protective organelle with such diverse and powerful protective features. Given the likelihood that many properties of the coat are emergent phenomena (see introduction), ascribing coat functions to specific proteins will likely require deleting specific, as yet unknown, combinations of coat protein genes.

References

1. Abe, A., H. Koide, T. Kohno, and K. Watabe. 1995. A *Bacillus subtilis* spore coat polypeptide gene, *cotS*. Microbiology. 141:1433-1442.

2. Aizawa, S.-I. 2001. Bacterial flagella and type III secretion systems. FEMS Microbiol. Lett. 202:157-164.

3. Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402.

4. Aronson, A. I., and P. Fitz-James. 1976. Structure and morphogenesis of the bacterial spore coat. Bacteriol. Rev. 40:360-402.

5. Aronson, A. I., H. Y. Song, and N. Bourne. 1989. Gene structure and precursor processing of a novel *Bacillus subtilis* spore coat protein. Mol. Microbiol. 3:437-444.

6. Bagyan, I., M. Noback, S. Bron, M. Paidhungat, and P. Setlow. 1998. Characterization of yhcN, a new forespore-specific gene of Bacillus subtilis. Gene. 212(2):179-88.

7. Bagyan, I., and P. Setlow. 2002. Localization of the cortex lytic enzyme CwlJ in spores of *Bacillus subtilis*. J. Bacteriol. 184:1219-1224.

8. Beall, B., A. Driks, R. Losick, and C. P. Moran Jr. 1993. Cloning and characterization of a gene required for assembly of the *Bacillus subtilis* spore coat. J. Bacteriol. 175:1705-1716.

9. Behravan, J., H. Chirakkal, A. Masson, and A. Moir. 2000. Mutations in the *gerP* locus of *Bacillus subtilis* and *Bacillus cereus* affect access of germinants to their targets in spores. J. Bacteriol. 182:1987-1994.

10. Boland, F. M., A. Atrih, H. Chirakkal, S. J. Foster, and A. Moir. 2000. Complete spore-cortex hydrolysis during germination of Bacillus subtilis 168 requires SleB and YpeB. Microbiology. 146(Pt 1):57-64.

11. Bourne, N., P. C. Fitz-James, and A. I. Aronson. 1991. Structural and germination defects of *Bacillus subtilis* spores with altered contents of a spore coat protein. J. Bacteriol. 173:6618-6625.

12. Buttner, K., J. Bernhardt, C. Scharf, R. Schmid, U. Mader, C. Eymann, H. Antelmann, A. Volker, U. Volker, and M. Hecker. 2001. A comprehensive two-dimensional map of cytosolic proteins of Bacillus subtilis. Electrophoresis. 22(14):2908-35.

13. Charlton, S., A. J. Moir, L. Baillie, and A. Moir. 1999. Characterization of the exosporium of *Bacillus cereus*. J. Appl. Microbiol. 87:241-245.

14. Chevallet, M., V. Santoni, A. Poinas, D. Rouquie, A. Fuchs, S. Kieffer, M. Rossignol, J. Lunardi, J. Garin, and T. Rabilloud. 1998. New zwitterionic detergents improve the analysis of membrane proteins by two-dimensional electrophoresis. Electrophoresis. 19:1901-1909.

15. Cutting, S., L. Zheng, and R. Losick. 1991. Gene encoding two alkali-soluble components of the spore coat from *Bacillus subtilis*. J. Bacteriol. 173:2915-2919.

16. Cutting, S. M., and P. B. Vander Horn. 1990. Molecular Biological Methods for *Bacillus*. John Wiley & Sons Ltd., Chichester, United Kingdom.

17. Dixon, T. C., M. Meselson, J. Guillemin, and P. C. Hanna. 1999. Anthrax. N. Engl. J. Med. 341:815-826.

18. Donovan, W., L. Zheng, K. Sandman, and R. Losick. 1987. Genes encoding spore coat polypeptides from *Bacillus subtilis*. J. Mol. Biol. 196:1-10.

19. Driks, A. 1999. The *Bacillus subtilis* spore coat. Microbiol. Mol. Biol. Rev. 63:1-20.

20. Driks, A. 2002. Maximum shields: the armor plating of the bacterial spore. Trends Microbiol. 10:251-254.

21. Driks, A. 2002. Proteins of the spore core and coat, p. 527-536. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and its closest relatives. American Society for Microbiology, Washington.

22. Driks, A. 1999. Spatial and temporal control of gene expression in prokaryotes. *In* V. E. A. Russo, D. J. Cove, L. G. Edgar, R. Jaenisch, and F. Salamini (ed.), Development: genetics, epigenetics and environmental regulation. Springer, Berlin.

23. Driks, A., and P. Setlow. 2000. Morphogenesis and properties of the bacterial spore. *In* Y. V. Brun and L. J. Shimkets (ed.), Prokaryotic Development. American Society for Microbiology, Washington, D.C.

24. Fillinger, S., S. Boschi-Muller, S. Azza, E. Dervyn, G. Branlant, and S. Aymerich. 2000. Two glyceraldehyde-3-phosphate dehydrogenases with opposite physiological roles in a nonphotosynthetic bacterium. J. Biol. Chem. 275:14031-14037.

25. Fountoulakis, M., and T. B. 2001. Effect of strong detergents and chaotropes on the detection of proteins in two-dimensional gels. Electrophoresis. 22:1593-1603.

26. Garcia-Patrone, M., and J. S. Tandecarz. 1995. A glycoprotein multimer from *Bacillus thuringiensis* sporangia: dissociation into subunits and sugar composition. Mol. Cell. Biochem. 145:29-37.

27. Gavin, A. C., M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A. M. Michon, C. M. Cruciat, M. Remor, C. Hofert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M. A. Heurtier, R. R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer, and G. Superti-Furga. 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature. 415:141-147.

28. Gerhardt, P., and E. Bibi. 1964. Ultrastructure of the exosporium enveloping spores of *Bacillus cereus*. J. Bacteriol. 88:1774-1789.

29. Graumann, P. L., and M. A. Marahiel. 1999. Cold shock proteins CspB and CspC are major stationary-phase-induced proteins in *Bacillus subtilis*. Arch. Microbiol. 171:135-138.

30. Hachisuka, Y., K. Kojima, and T. Sato. 1966. Fine filaments on the outside of the exosporium of *Bacillus anthracis* spores. J. Bacteriol. 91:2382-2384.

31. Henriques, A. O., and C. P. Moran, Jr. 2000. Structure and assembly of the bacterial endospore coat. Methods. 20:95-110.

32. Hirokawa, T., S. Boon-Chieng, and S. Mitaku. 1998. SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics. 14:378-379.

33. Ho, Y., A. Gruhler, A. Heilbut, G. D. Bader, L. Moore, S. L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff, J. Shewnarane, M. Vo, J. Taggart, M. Goudreault, B. Muskat, C. Alfarano, D. Dewar, Z. Lin, K. Michalickova, A. R. Willems, H. Sassi, P. A. Nielsen, K. J. Rasmussen, J. R. Andersen, L. E. Johansen, L. H. Hansen, H. Jespersen, A. Podtelejnikov, E. Nielsen, J. Crawford, V. Poulsen, B. D. Sorensen, J. Matthiesen, R. C. Hendrickson, F. Gleeson, T. Pawson, M. F. Moran, D. Durocher, M. Mann, C. W. Hogue, D. Figeys, and M. Tyers. 2002. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature. 415:180-183.

34. Hulett, F. M. 2002. The Pho Regulon, p. 193-202. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and its closest relatives. American Society for Microbiology, Washington.

35. Hullo, M. F., I. Moszer, A. Danchin, and I. Martin-Verstraete. 2001. CotA of *Bacillus subtilis* is a copper-dependent laccase. J. Bacteriol. 183(1):5426-5430.

36. Ishikawa, J., and K. Hotta. 1999. FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G + C content. FEMS Microbiol. Lett. 174:251-253.

37. Isticato, R., G. Cangiano, H. T. Tran, A. Ciabattini, D. Medaglini, M. R. Oggioni, M. De Felice, G. Pozzi, and E. Ricca. 2001. Surface Display of Recombinant Proteins on *Bacillus subtilis* Spores. J. Bacteriol. 183:6294-6301.

38. Kodama, T., H. Takamatsu, K. Asai, K. Kobayashi, N. Ogasawara, and K. Watabe. 1999. The *Bacillus subtilis yaaH* gene is transcribed by SigE RNA polymerase during sporulation, and its product is involved in germination of spores. J. Bacteriol. 181:4584-4591.

39. Kodama, T., H. Takamatsu, K. Asai, N. Ogasawara, Y. Sadaie, and K. Watabe. 2000. Synthesis and characterization of the spore proteins of *bacillus subtilis* YdhD, YkuD, and YkvP, which carry a motif conserved among cell wall binding proteins. J. Biochem. (Tokyo). 128:655-663.

40. Lee, V. T., and O. Schneewind. 2001. Protein secretion and the pathogenesis of bacterial infections. Genes Dev. 15:1725-1752.

41. Little, S., and A. Driks. 2001. Functional analysis of the *Bacillus subtilis* morphogenetic spore coat protein CotE. Mol. Microbiol. 42:1107-1120.

42. Macnab, R. M. 1999. The bacterial flagellum: reversible rotary propellor and type III export apparatus. J. Bacteriol. 181:7149-7153.

43. Margolin, W. 2001. Spatial regulation of cytokinesis in bacteria. Curr. Opin. Microbiol. 4:647-652.

44. Martins, L. O., C. M. Soares, M. M. Pereira, M. Teixeira, T. Costa, G. H. Jones, and A. O. Henriques. 2002. Molecular and biochemical characterization of a highly stable bacterial laccase that occurs as a structural component of the *Bacillus subtilis* endospore coat. J. Biol. Chem. 277(2):18849-18859.

45. Matz, L. L., T. C. Beaman, and P. Gerhardt. 1970. Chemical composition of exosporium from spores of *Bacillus cereus*. J Bacteriol. 101:196-201.

46. Micka, B., N. Groch, U. Heinemann, and M. A. Marahiel. 1991. Molecular cloning, nucleotide sequence, and characterization of the *Bacillus subtilis* gene encoding the DNA-binding protein HBsu. J. Bacteriol. 173:3191-3198.

47. Moberly, B. J., F. Shafa, and P. Gerhardt. 1966. Structural details of anthrax spores during stages of transformation into vegetative cells. J. Bacteriol. 92:220-228.

48. Moir, A. 1981. Germination properties of a spore coat-defective mutant of *Bacillus subtilis*. J. Bacteriol. 146:1106-1116.

49. Moir, A., B. M. Corfe, and J. Behravan. 2002. Spore germination. Cell. Mol. Life Sci. 59:403-409.

50. Molloy, M. P. 2000. Two-dimensional electrophoresis of membrane proteins using immobilized pH gradients. Anal. Biochem. 280:1-10.

51. Molloy, M. P., N. D. Phadke, H. Chen, R. Tyldesley, D. E. Garfin, J. M. Maddock, and P. C. Andrews. 2002. Profiling the alkaline membrane proteome of *Caulobacter crescentus* with two-dimensional electrophoresis and mass spectrometry. Proteomics. 2:899-910.

52. Molloy, M. P., N. D. Phadke, J. R. Maddock, and P. C. Andrews. 2001. Two-dimensional electrophoresis and peptide mass-fingerprinting of bacterial outer membrane proteins. Elecrophoresis. 22:1686-1696.

53. Moriyama, R., H. Fukuoka, S. Miyata, S. Kudoh, A. Hattori, S. Kozuka, Y. Yasuda, K. Tochikubo, and S. Makino. 1999. Expression of a germination-specific amidase, SleB, of Bacilli in the forespore compartment of sporulating cells and its localization on the exterior side of the cortex in dormant spores. J. Bacteriol. 181:2373-2378.

54. Moriyama, R., A. Hattori, S. Miyata, S. Kudoh, and S. Makino. 1996. A gene (*sleB*) encoding a spore cortex-lytic enzyme from *Bacillus subtilis* and response of the enzyme to L-alanine-mediated germination. J. Bacteriol. 178:6059-6063.

55. Moriyama, R., S. Kudoh, S. Miyata, S. Nonobe, A. Hattori, and S. Makino. 1996. A germination-specific spore cortex-lytic enzyme from *Bacillus cereus* spores: cloning and sequencing of the gene and molecular characterization of the enzyme. J. Bacteriol. 178:5330-5332.

56. Mortz, E., T. N. Krogh, H. Vorum, and A. Gorg. 2001. Improved silver staining protocols for high sensitivity protein identification using matrix-assisted laser desorption/ionization-time of flight analysis. Proteomics. 1:1359-1363.

57. Nakai, K. 2000. Protein sorting signals and prediction of subcellular localization. Adv. Protein Chem. 54:277-344.

58. Nicholson, W. L., N. Munakata, G. Horneck, H. J. Melosh, and P. Setlow. 2000. Resistance of Bacillus endospores to extreme terrestrial and extraterrestrial environments. Microbiol. Mol. Biol. Rev. 64:548-572.

59. Nielsen, H., S. Brunak, and G. von Heijne. 1999. Machine learning approaches for the prediction of signal peptides and other protein sorting signals. Protein Eng. 12:3-9.

60. O'Farrell, P. H. 1975. High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. 250:4007-4002.

61. Ohlmeier, S., C. Scharf, and M. Hecker. 2000. Alkaline proteins of Bacillus subtilis: first steps towards a two- dimensional alkaline master gel. Electrophoresis. 21(17):3701-9.

62. Otte, M., and A. Driks. Unpublished observations.

63. Ozin, A. J., A. O. Henriques, H. Yi, and C. P. Moran, Jr. 2000. Morphogenetic proteins SpoVID and SafA form a complex during assembly of the *Bacillus subtilis* spore coat. J. Bacteriol. 182:1828-1833.

64. Ozin, A. J., C. S. Samford, A. O. Henriques, and C. P. Moran, Jr. 2001. SpoVID guides SafA to the spore coat in *Bacillus subtilis*. J. Bacteriol. 183:3041-3049.

65. Paidhungat, M., and P. Setlow. 2002. Spore germination and outgrowth, p. 527-536. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and its closest relatives. American Society for Microbiology, Washington.

66. Phadke, N. D., M. P. Molloy, S. A. Steinhoff, P. J. Ulintz, P. C. Andrews, and J. R. Maddock. Analysis of the outer membrane proteome of *Caulobacter crescentus* by two-dimensional electrophoresis and mass spectrometry. Proteomics. In press.

67. Roth, I. L., and R. P. Williams. 1963. Comparison of the fine structure of virulent and avirulent spores of *Bacillus anthracis*. Texas Rept. Biol. Med. 21:394-399.

68. Sacco, M., E. Ricca, R. Losick, and S. Cutting. 1995. An additional GerE-controlled gene encoding an abundant spore coat protein from *Bacillus subtilis*. J. Bacteriol. 177:372-377.

69. Salzberg, S. L., A. L. Delcher, S. Kasif, and O. White. 1998. Microbial gene identification using interpolated Markov models. Nucleic Acids Res. 26:544-548.

70. Schindler, T., P. L. Graumann, D. Perl, S. Ma, F. X. Schmid, and M. A. Marahiel. 1999. The family of cold shock proteins of *Bacillus subtilis*. Stability and dynamics *in vitro* and *in vivo*. J. Biol. Chem. 274:3407-3413.

71. Seyler, R. W. J., A. O. Henriques, A. J. Ozin, and C. P. J. Moran. 1997. Assembly and interactions of *cotJ*-encoded proteins, constituents of the inner layers of the *Bacillus subtilis* spore coat. Mol. Microbiol. 25:955-966.

72. Sonenshein, A. L. 2000. Endospore-forming bacteria: an overview. *In* Y. V. Brun and L. J. Shimkets (ed.), Prokaryotic Development. American Society for Microbiology, Washington, D.C.

73. Sonenshein, A. L., J. A. Hoch, and R. Losick. 2002. *Bacillus subtilis* and its closest relatives. American Society for Microbiology, Washington.

74. Sylvestre, P., E. Couture-Tosi, and M. Mock. 2002. A collagen-like surface glycoprotein is a structural component of the *Bacillus anthracis* exosporium. Mol. Microbiol. 45:169-178.

75. Takamatsu, H., T. Hiraoka, T. Kodama, H. Koide, S. Kozuka, K. Tochikubo, and K. Watabe. 1998. Cloning of a novel gene *yrbB*, encoding a protein located in the spore integument of *Bacillus subtilis*. FEMS Microbiol. Lett. 166:361-367.

76. Takamatsu, H., A. Imamura, T. Kodama, K. Asai, N. Ogasawara, and K. Watabe. 2000. The *yabG* gene of *bacillus subtilis* encodes a sporulation specific protease which is involved in the processing of several spore coat proteins. FEMS Microbiol. Lett. 192:33-38.

77. Takamatsu, H., T. Kodama, A. Imamura, K. Asai, K. Kobayashi, T. Nakayama, N. Ogasawara, and K. Watabe. 2000. The *Bacillus subtilis yabG* gene is transcribed by SigK RNA polymerase during sporulation, and *yabG* mutant spores have altered coat protein composition. J. Bacteriol. 182:1883-1888.

78. Takamatsu, H., T. Kodama, T. Nakayama, and K. Watabe. 1999. Characterization of the *yrbA* gene of *Bacillus subtilis*, involved in resistance and germination of spores. J. Bacteriol. 181:4986-4994.

79. Tatusov, R. L., M. Y. Galperin, D. A. Natale, and E. V. Koonin. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 28(1):33-6.

80. Thomas, J. H. 1993. Thinking about genetic redundancy. Trends Genet. 9:395-399.

81. Warth, A. D., D. F. Ohye, and W. G. Murrell. 1963. The composition and structure of bacterial spores. J. Cell Biol. 16:579-592.

82. Youngman, P., J. B. Perkins, and R. Losick. 1984. Construction of a cloning site near one end of Tn*917* into which foreign DNA may be inserted without affecting transposition in *Bacillus subtilis* or expression of the transposon-borne *erm* gene. Plasmid. 12:1-9.

83. Zhang, J., P. C. Fitz-James, and A. I. Aronson. 1993. Cloning and characterization of a cluster of genes encoding polypeptides present in the insoluble fraction of the spore coat of *Bacillus subtilis*. J. Bacteriol. 175:3757-3766.

84. Zheng, L., W. P. Donovan, P. C. Fitz-James, and R. Losick. 1988. Gene encoding a morphogenic protein required in the assembly of the outer coat of the *Bacillus subtilis* endospore. Genes Dev. 2:1047-1054.

Figure 4.1: *B. subtilis* coat proteins resolved by 1D standard SDS–15% PAGE. The gel was stained with Coomassie blue G-250. The indicated bands were prepared for MALDI MS and MS/MS mapping as described in Materials and Methods. Protein identifications correspond to those described in Table 4.1.

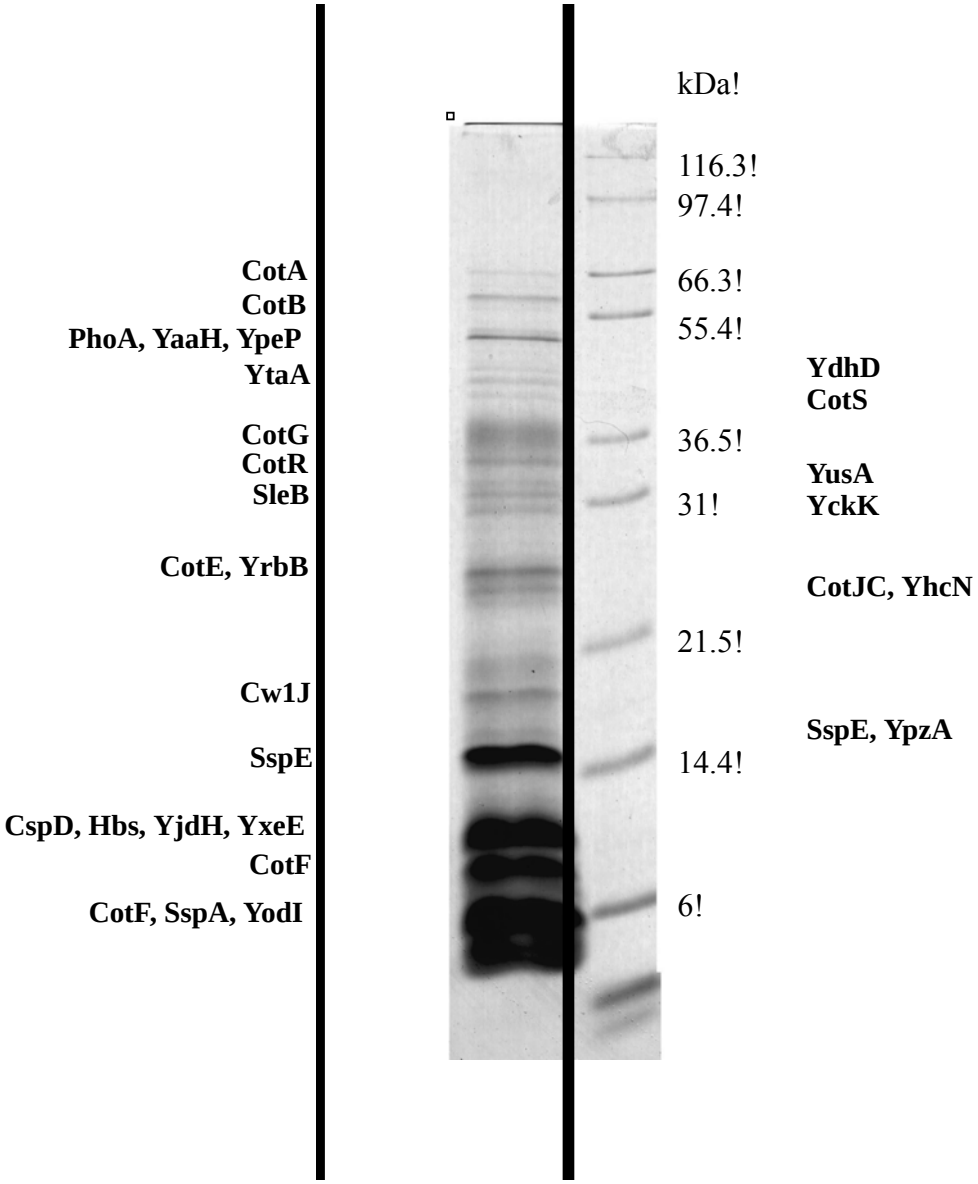
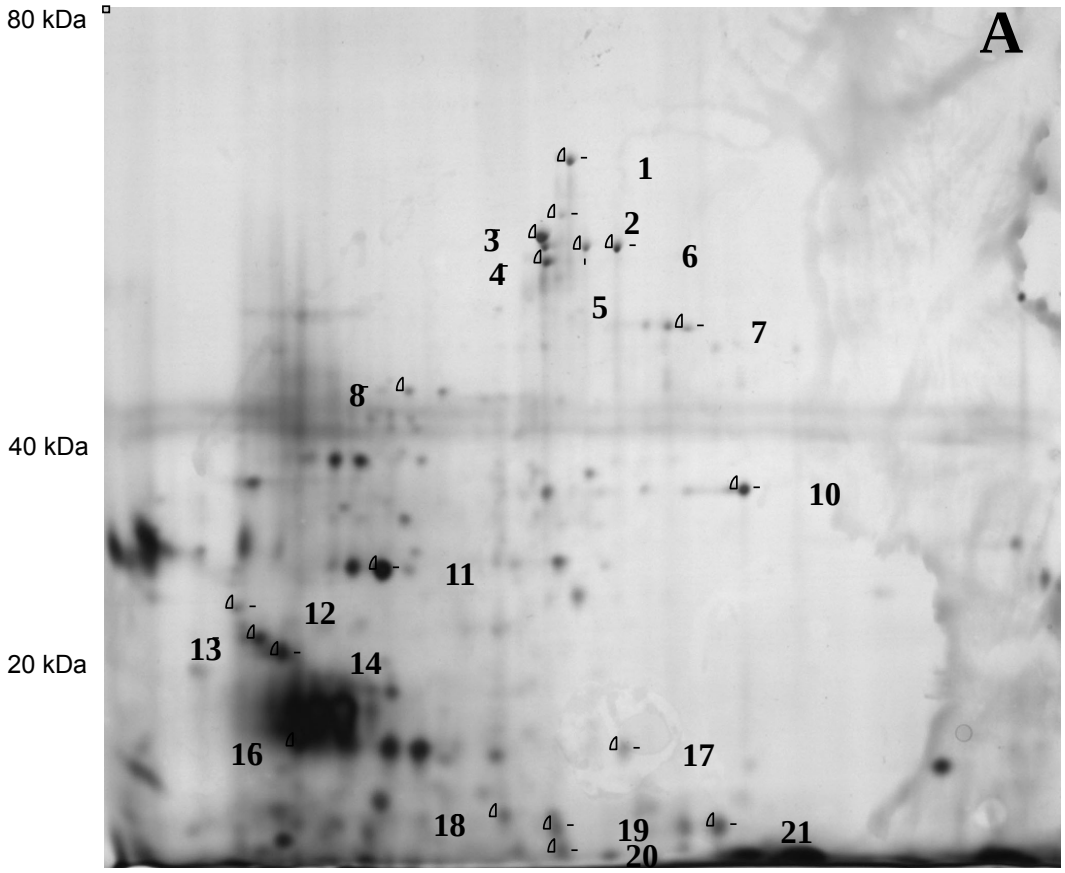

kDa!
116.3!
97.4!
66.3!
55.4!
36.5!
31!
21.5!
14.4!
6!
CotA
CotB
PhoA, YaaH, YpeP
YtaA
CotG
CotR
SleB
CotE, YrbB
Cw1J
SspE
CspD, Hbs, YjdH, YxeE
CotF
CotF, SspA, YodI
YdhD
CotS
YusA
YckK
CotJC, YhcN
SspE, YpzA

Figure 4.2: Preliminary master gel of *B. subtilis* coat preparations and MALDI analysis. (A) pH 3 to 10 gel of *B. subtilis* coat preparations. Sample was solubilized in 10% LDS buffer. The solution was diluted in an excess of standard solubilization buffer containing 2% C8ϕ and 1% TX-100 and resolved by IEF. The second dimension was carried out on an SDS–11% polyacrylamide gel and silver stained (56). Spots identified by peptide mass fingerprinting are circled and numbered, and the identifications are presented in Table 4.2. (B) MALDI-MS spectrum for spot 1.


A
80 kDa
40 kDa
20 kDa
1
2
3
4
5
6
7
8
10
11
12
13
14
16
17
18
19
20
21
pH 3
pH 10

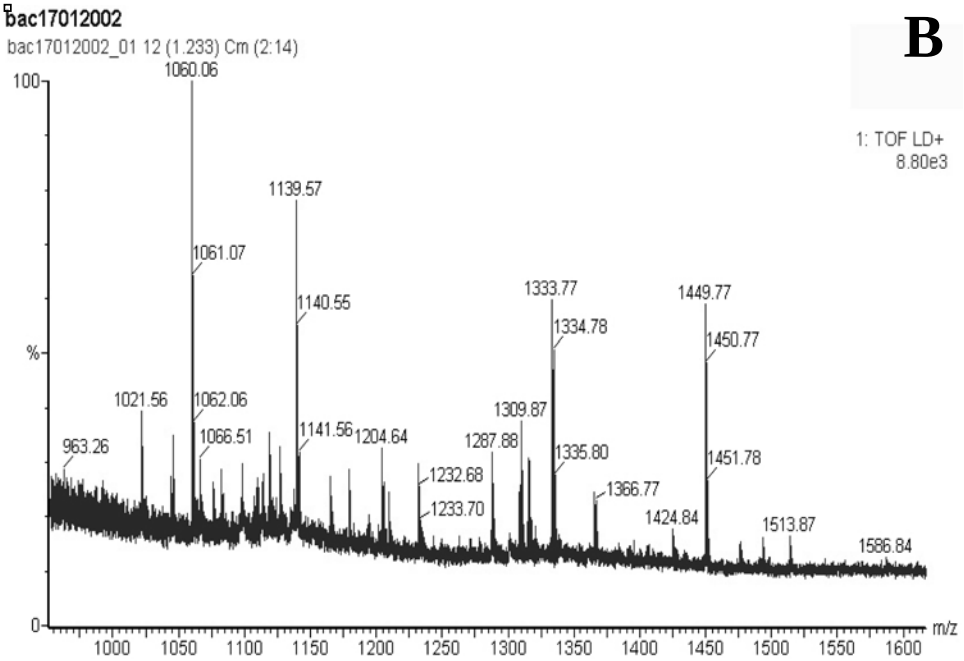

bac17012002
bac17012002_01 12 (1.233) Cm (2:14)
B
1: TOF LD+
8.80e3
100
%
0
1060.06
1139.57
1061.07
1140.55
1333.77
1334.78
1449.77
1450.77
1021.56
1062.06
1066.51
1141.56
1204.64
1309.87
1287.88
1335.80
963.26
1232.68
1233.70
1366.77
1451.78
1424.84
1513.87
1586.84
1000
1050
1100
1150
1200
1250
1300
1350
1400
1450
1500
1550
1600
m/z

Figure 4.3: Improved 2D separation. One hundred eighty microliters of a spore pellet was solubilized in 100 μl of 1x SDS buffer after being boiled for 10 min. Ninety microliters of this solution was added to 810 μl of 1.1x 2% C8Φ–1% TX-100 in standard rehydration buffer and incubated at 30°C for 1 h before rehydration. Four hundred fifty microliters each of this solution was run on a pH 3 to 10 (A) and a pH 4 to 7 (B) IEF strip. Frequent changes of filter paper wicks during IEF and focusing up to 120 kVh until the dye ran out completely at the acidic end greatly improved separation. The second dimension is as in Fig. 4.2.

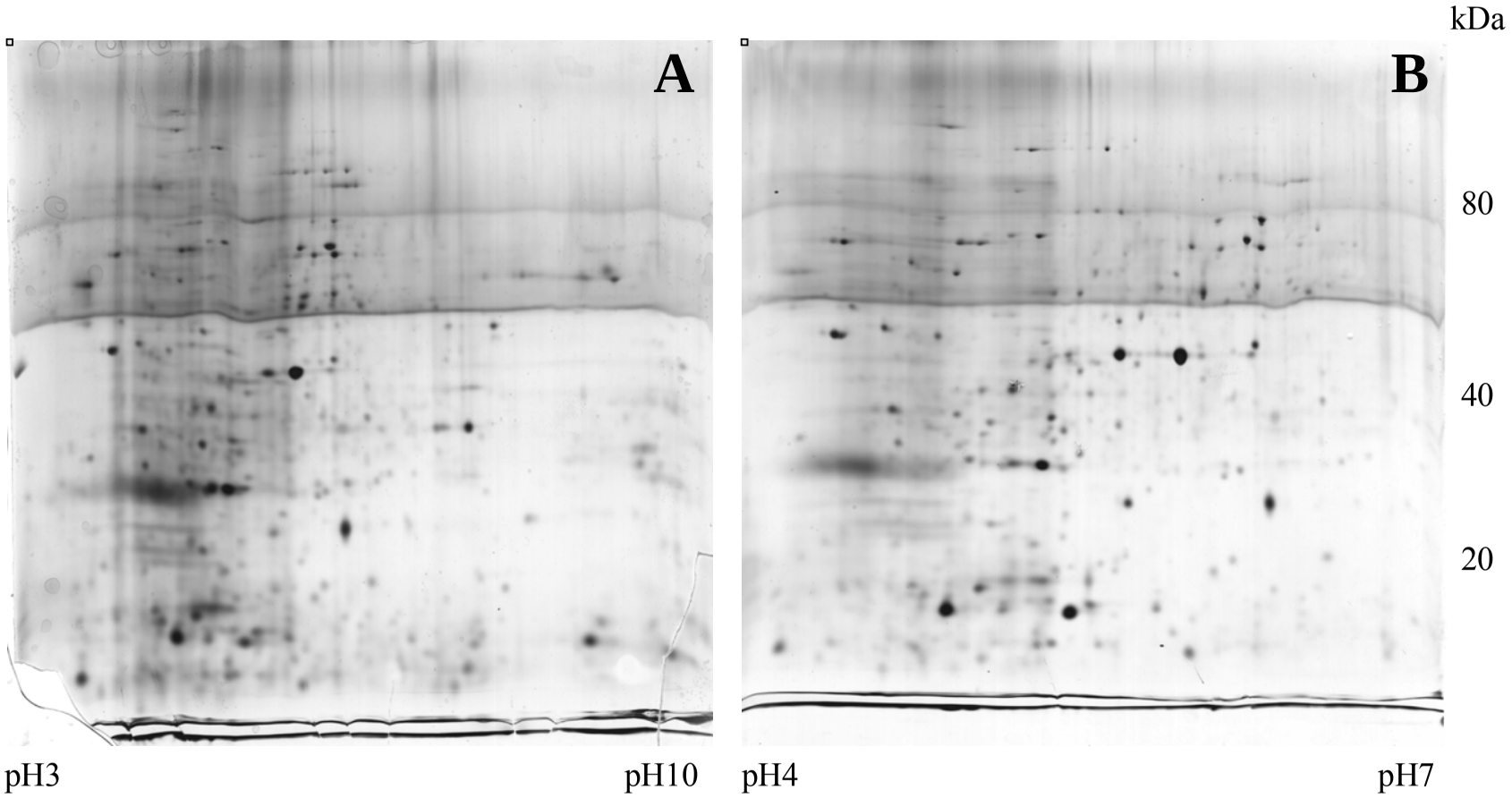

A
B
kDa
80
40
20
pH3
pH10
pH4
pH7

Figure 4.4: *B. anthracis* coat proteins resolved by 1D standard SDS– 15% PAGE. The samples were solubilized with SDS, and the resulting gel was stained with Coomassie blue G-250. Proteins were identified by MALDI-TOF MS, and the identifications are presented in Table 4.4.

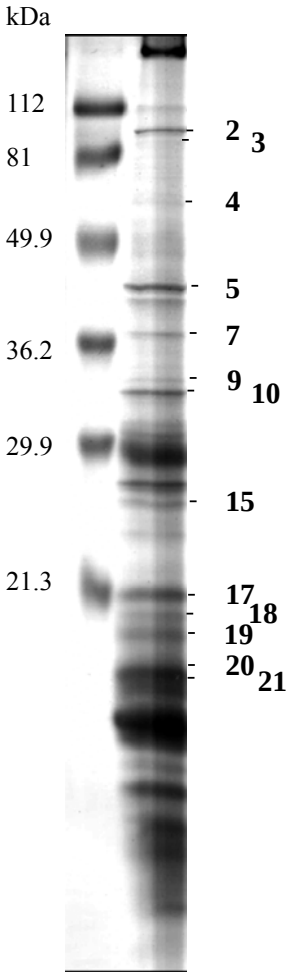

kDa
112
81
49.9
36.2
29.9
21.3
2
3
4
5
7
9
10
15
17
18
19
20
21

Table 4.1: *B. subtilis* proteins identified from a 1D gel.

Band	Protein	MALDI-MS[a]	MS/MS[b]	Predicted MW (kDa)/ pI	Observed migration (kDa)
1	CotA	yes	1 (1514)	58.5 / 5.91	71
2	CotB		2 (1305, 1530)	43.0 / 9.47	62
3	YaaH	yes	2 (1414, 1366)	48.6 / 5.72	53
3	YpeP		2 (1542, 1635)	51.2 / 6.31	53
3	PhoA		1 (2555)	50.2 / 9.50	53
4	YdhD		2 (1317, 1784)	49.0 / 9.01	48
5	YtaA		1 (1401)	41.2 / 5.16	46
6	CotS		1 (1430)	41.1 / 6.61	44
7	CotG	yes	2 (1249, 1676)	24.0 / 10.26	40
8	CotR		2 (1562, 2123)	35.4 / 9.87	37
9	YusA		1 (1998)	30.4 / 8.26	35
10	SleB	yes	1 (1800)	34.0 / 9.27	33
11	YckK		1 (1893)	29.4 / 7.64	31
12	CotE	yes	2 (1994, 1549)	21.0 / 4.36	26
12	YrbB		2 (2498, 1650)	19.5 / 8.89	26
13	CotJC	yes	1 (1909)	21.7 / 5.08	25
13	YhcN		2 (1851, 2249)	21.0 / 5.56	25
15	CwlJ		1 (1589)	16.4 / 9.38	17
16	YpzA		1 (1551)	10.1 / 4.3	18
16	SspE	yes		9.3 / 8.19	18
17	SspE	yes	4 (1481, 1812, 1871, 2968)	9.3 / 8.19	15
18	YxeE		2 (1585, 1821)	14.7 / 8.03	12
18	YjdH		1 (1303)	15.3 / 5.25	12
18	CspD		1 (1922)	7.3 / 4.51	12
18	Hbs		1 (1393)	9.9 / 8.96	12
19	CotF	yes	1 (1634)	18.7 / 6.98	8
20	CotF		2 (1452, 1744)	18.7 / 6.98	5
20	YodI		1 (1070)	9.2 /10.28	5
20	SspA		2 (1656, 1879)	7.1 / 4.94	5

[a] Top MALDI-MS hit
[b] Number of peptides providing fragment ions and sizes (parentheses)

Table 4.2: Proteins identified by MALDI-TOF-MS from 2D separation of *B. subtilis* spore coat fractions

Spot #	MOWSE score	Peptides matched	MW/pI	Name	Function
1	2.15E+06	11	58499.3/5.91	CotA	Spore coat protein A
2	5.50E+05	13	48637.0/5.72	YaaH	Unknown protein
3	NA[a]	9	48534.4/7.5	YhbA	Unknown protein
4	1.14E+06	10	58499.3/5.91	CotA	Spore coat protein A
5	8.49E+03	6	50085.0/6.12	YvdP	Similar to spore coat protein, CotQ
6	8.16E+02	5	53503.9/5.41	YopQ	Unknown protein
7	NA[a]	8	129334/5.44	YirY	Unknown protein
8	NA[a]	8	37382.4/5.8	CggR	Transcriptional regulator
10	3.44E+04	6	30559/6.8	YisY	Unknown protein
11	NA[a]	7	21695.8/5.1	CotJC	Spore coat protein
12	1453	4	20977/4.4	CotE	Spore coat morphogenetic protein E
13	7589	5	20977/4.4	CotE	Spore coat morphogenetic protein E
14	6.26E+05	8	20977/4.4	CotE	Spore coat morphogenetic protein E
16	NA[a]	5	18728/5.02	CotY	Spore coat protein Y
17	NA[a]	5	15857/5.01	YwqH	Unknown protein
18	NA[a]	5	14785/9.4	CotC	Spore coat protein C
19	NA[a]	5	14785/9.4	CotC	Spore coat protein C
20	NA	5	14785/9.4	CotC	Spore coat protein C
21	5.62E+03	6	11562.5/6.28	YnzH	Similar to spore coat protein, CotU

[a] Identifications were made with the MASCOT algorithm or with the Micromass Protein Lynx Software Package

Table 4.3: Known and predicted spore coat proteins in *B. subtilis* and *B. anthracis*[a]

B. subtilis **previously known spore coat protein**	**Comments on *B. subtilis* proteins**	***B. anthracis*** **Protein ID**[b]	**Identities / length (%)**	**Comments on *B. anthracis* proteins**
CotA	Multi-copper oxidase	655766	39/133 (29%) 51/213 (24%)	2 regions of similarity to CotA.
CotB	Encoded downstream of *cotH*	654287 654288	47/156 (30%) 55/153 (33%)	Adjacent genes.
CotC		none		
CotD		655458	32/70 (45%)	Likely homolog.
CotE	Controls assembly of most outer coat proteins and some inner coat proteins	657736	106/181 (58%)	Homolog
CotF	Proteolytically processed from a 23kD precursor	656983	97/154 (62%)	Homolog
CotG	Gene is divergent from *cotH*	none		Lack of homology could be due to lack of complexity.
CotH	Gene is divergent from *cotG*, upstream of *cotB*	655906	198/356 (55%)	Homolog
CotJA	Encoded in *cotJABC* operon	654751	46/66 (69%)	Homolog
CotJC	Encoded in *cotJABC* operon	1326[c] 654749 656966	170/189 (89%) 63/68 (92%) 48/201 (23%)	654749 is a gene fragment. 1326 is the homolog.
CotM	Endoded in *sspO sspP cotM* operon	657526 656113	26/111 (23%) 20/83 (24%)	657526 is CotM-like.
CotR (YvdO)	Gene is adjacent to and divergent from *cotQ* (*yvdP*)	none		
CotSA	Encoded in *cotSA cotS* operon. Gene is divergent from *cotI* (*ytaA*)	658792 655436 653730	57/215 (26%) 59/221 (26%) 52/212 (24%)	
CotS	Encoded in *cotSA cotS* operon	653420	48/183 (26%)	C-terminus only.
CotT	Proteolytically processed from a 8 kD precursor	none		

CotV	Encoded in *cotVWXYZ* operon	none		Slightly similar to several ORFs.
CotW	Encoded in *cotVWXYZ* operon	none		Slightly similar to several ORFs.
CotY	Encoded in *cotVWXYZ* operon	1889[c] 655129	53/154 (34%) 54/153 (35%)	Both CotY/Z-like.
CotZ	Encoded in *cotVWXYZ* operon	655129 1889[c]	50/141 (35%) 48/139 (34%)	Both CotY/Z-like.
CwlJ (YcbQ)	Cortex-lytic enzyme	653856 656463 656634 657724	87/140 (62%) 82/141 (58%) 34/119 (28%) 40/126 (31%)	Both 653856 and 656463 are homologs. 656634 and 657724 have a CwlJ-like C-terminal motif.
SpoIVA	Appears to connect the coat to the forespore and is required for cortex formation	655410	434/492 (88%)	Homolog
SpoVM	Partially required for SpoIVA localization	657826	21/26 (84%)	Homolog
SpoVID	Partially required for coat localization	658491	100/300 (33%)	Gene fragment?
GerP[d]	Some product(s) of the 6 gene *gerP* operon likely controls entry of germinant into the spore interior			
YaaH	Peptidoglycan hydrolase. Affects germination. Similar to YvbX and YkvQ	657516 657340 654321	207/424 (48%) 64/247 (25%) 37/163 (22%)	657516 is a homolog.
YabG	Affects coat protein composition and germination. Has protease activity	653990	170/288 (59%)	Homolog

YrbA	Affects germination and coat assembly. 31kD species likely to be a product of proteolysis of a 43kD precursor	658457	56/112 (50%)	Gene fragment?
***B. subtilis* candidate spore coat proteins**				
CotQ (YvdP)	Gene is adjacent to and divergent from *cotR*	none		
CotU (YnzH)	Similar to CotC	none		Small ORF
CotI (YtaA)	Similar (47% over most of its length) to CotS. Some similarity to YutH. Gene is divergent from *cotSA*	653420	41/190 (21%)	C-terminus of 653420 similar to C-terminus of CotI.
YckK	Resembles an ABC transporter	654795 654301 654584	138/267 (51%) 81/267 (30%) 75/276 (31%)	654795-homolog. All are clearly related along entire length.
YdhD	Predicted to be a cortex lytic enzyme. Similar to YaaH and YvbX	657516 657340	125/430 (30%) 55/213 (25%)	
YjdH	Unknown protein	none		
YhbA	Resembles an iron-sulfur cluster-binding protein	654477	263/375 (70%)	Homolog
YhdE	Resembles a RRF2 family protein. Slightly similar to YwgB. Transcribed divergently from *spoVR*	654902 658427 657309	37/130 (28%) 38/137 (27%) 29/135 (21%)	All YhdE-like along entire length.
YirY	Potentially an exonuclease	657817 656214	192/1013 (18%) 92/349 (26%)	
YisY	Resembles a chloride peroxidase, 45% similar to YdjP	658807	115/268 (42%)	14 more ORFs with weak similarity over the entire sequence.

YodI		657487	20/42 (47%)	Likely homolog, gene fragment.
YopQ		657926	36/129 (27%)	
YpeP/YpeB	*ypeB*, *ypzA* and *cspD* are within 1.5 kb region	656633	257/448 (57%)	Homolog
YpzA	Homologue identified only in *B. anthracis.* *ypeB*, *ypzA* and *cspD* are within 1.5 kb region	655504	25/73 (34%)	Homolog
YusA	ABC transporter substrate-binding protein	653451 653452 654118 654245	165/275 (60%) 156/275 (56%) 103/277 (37%) 95/277 (34%)	653451 and 653452 are adjacent genes.
YwqH	Some similarity to YxiB	657122	34/135 (25%)	
YxeE		657441	50/109 (45%)	Homolog
Additional proteins identified in this study				
CspD	Cold shock protein. *ypeB*, *ypzA* and *cspD* are within 1.5 kb region	658893 655505 655043 653638 657446	57/65 (87%) 54/63 (85%) 52/63 (82%) 50/63 (79%) 45/56 (80%)	All appear to be homologs.
Hsb	Histone-like protein. Gene is downstream from *spoIVA*	655411 656232 5344[c]	80/89 (89%) 65/89 (73%) 58/89 (65%)	All appear to be homologs
PhoA	Alkaline phosphatase	658376	256/449 (57%)	Homolog
SleB	Cortex-lytic enzyme	656634 657724 656463 653856	156/277 (56%) 68/203 (33%) 36/126 (28%) 34/124 (27%)	656634-homolog. 657724, 656463 and 653856 similar to C-terminus of SleB.
SspA	Alpha-type small acid-soluble spore protein	656989 656992 654798 658678 655848 655208	52/67 (77%) 50/66 (75%) 44/66 (66%) 44/61 (72%) 47/67 (70%) 39/58 (67%)	All appear to be homologs.
SspE	Gamma-type small acid-soluble spore protein			
YhcN	Forespore protein, some similarity to YlaJ	656632 658456	47/182 (25%) 52/196 (26%)	
YrbB (CoxA)	Cortex protein.	658456	43/143 (30%)	Gene fragment?

	Gene is downstream of *safA* (*yrbA*)			
CggR	A transcriptional regulator	653588	206/338 (60%)	Homolog

[a]Proteins in bold were identified in this study.
[b]NP_ numbers accessible through NCBI
[c] Our working ORF identification, gene not in public database.
[d]It is unknown which proteins encoded in the *gerP* operon are coat proteins. Tallies of coat proteins in the text do not include those potentially encoded by the *gerP* operon.

Table 4.4: Proteins identified by MALDI-TOF-MS from 1D of *B. anthracis* spore coat fractions

Band	Protein ID[a]	MOWSE score	Peptides matched	Coverage (%)	MW/pI	Name
2	654830	1.49E+11	18	28	91362.5 / 5.70	Cell surface antigen, contains SLH domain
3	654030	663	6	7	90530.8 / 6.06	81% identical to *B. subtilis* ClpC, class III stress response ATPase
4	654198	1.56E+08	10	21	57432.2 / 4.79	78% identical to *B. subtilis* GroEL
5	657011	1.21E+03	4	19	24508.4 / 8.17	49% identical to *B. subtilis* YndF (amino acids 23-173), Putative spore germination protein.
7	656768	1.01E+07	12	36	36244.2 / 5.06	Nucleoside hydrolase
9	657458	8.93E+03	6	21	41951.3 / 8.82	Nucleoside hydrolase
10	654944	9.59E+08	12	39	38373.2 / 6.17	Unknown protein, also in *Clostridium perfringens*
15	1326[b]	3.17E+03	5	51	21650.7 / 5.15	89% identical to *B. subtilis* CotJC.
17	654287	177	3	15	19389.8 / 5.31	37% identical to *B. subtilis* CotB (amino acids 18-146) and 25% identical to YwrJ.
18	655132	3.29E+03	3	35	17672.4 / 4.59	Unknown protein, adjacent to genes encoding CotY/Z-like proteins
19	655132	168	2	19	17672.4 / 4.59	Unknown protein, adjacent to genes encoding CotY/Z-like proteins
20	655129	8.36E+03	4	20	16459.9 / 4.95	35% identical to *B. subtilis* CotZ and CotY
21	655129	5.09E+04	6	47	16459.9 / 4.95	35% identical to *B. subtilis* CotZ and CotY

[a]NP_ numbers accessible through NCBI
[b] Our working ORF identification, gene not in public database.

CHAPTER V

CONCLUSIONS AND OUTLOOK

The bulk of this work addreses the development of methodologies to carry out meaningful proteomic analyses of proteins from hard-to-dissect biological samples, such as membrane fractions. In chapter 1, I provide an introduction to the field of proteomics with a special focus on our primary biological systems of interest i.e. membrane proteins from microbes. I review in depth, some of the most recent microbial membrane quantitative proteomic studies that utilize the preferred state-of-the art technique at the time of writing – iTRAQ, and discuss some of the options available to investigators wishing to carry out such analyses.

The rest of the chapters describe the work carried out in my research group for the analysis of these hard-to-dissect proteomic samples. The approaches that I have used, and the methods that I have described, to-date remain relevant to the analysis of hard-to-dissect proteomic samples. They also provide a guideline for investigators wishing to carry out proteomic analyses concurrently with genome sequencing projects.

Proteomic analysis of membrane fractions from the model system, *C. crescentus,* was unusually complicated due to lack of a completed genome sequence project at the time of initiation of this work. Chapter 2 describes my efforts towards the analysis of membrane proteins from an incomplete and un-annotated genome. Concomitantly with

the annotation process, I established conditions for reproducible high quality separations of *C. crescentus* outer membrane proteins by 2-DE using pH 3-10 and pH 4-7 IPGs. Using peptide mass fingerprinting to identify proteins separated by these gels, I established a catalogue of the most abundant proteins from the *C. crescentus* outer membrane, which I refer to as the standard outer membrane proteome of *C. crescentus*. Since this work was carried out in parallel with the sequencing effort, at times results from the proteomic survey were used to correct discrepancies in the annotation process such as prediction of ORF size. This demonstrates the utility and feasibility of initiating proteomic research concomitantly with a sequencing project. The usefulness of a standard proteome is that it provides a stable reference point to compare variations in protein expression patterns in response to physiological or external conditions or between different organisms or fractions. I demonstrated this by comparing the protein expression profile of *C. crescentus* cells grown under nutrient limiting and nutrient non-limiting conditions. Several TonB-dependent receptors, which have been implicated in the uptake of nutrients in *C. crescentus*, were shown to be upregulated under nutrient limiting conditions..

In chapter 3, I described methods used to further expand the proteomic coverage of the samples. I tackled one of the most problematic areas in 2-DE mediated proteomics, the separation of alkaline proteins. Using modified IPG matrices and conditions for IEF, I successfully created high quality reproducible separations of *C. crescentus* membrane fractions. Identification of proteins from these gels was greatly assisted by the use of a prototype hybrid MALDI-Q-Tof instrument from Micromass that was capable of carrying out MS/MS analysis using a MALDI ion source. The ability of this instrument to make

confident identifications with limited peptide mass data greatly enhances the power of proteomics. Furthermore, the advances offered by such devices promise to offer even greater benefits when analyzing proteins from large genomes.

Finally in chapter 4, I utilized some of the lessons learned during my analysis of *C. crescentus* membrane fractions for the analysis of spore coat fractions from *B. subtilis*. I described a methodology, which I present as a general approach for analyzing hard-to-dissect biological samples by proteomics. Using a combination of 1-DE and 2-DE with MALDI-TOF-MS and MS/MS analysis, I detected the presence of a significant number of known spore coat proteins from *B. subtilis*. I also identified several more proteins from the spore coat fractions that are good candidates for further investigation, as putative coat components, by intensive genetic and biochemical techniques. Additionally, with the example of *B. anthracis*, I once again demonstrated the validity and utility of carrying out proteomic analysis concurrently with a genome sequencing effort. Using a bioinformatics approach, I detected putative homologues of several spore coat and surface layer proteins from this pathogen. These data were complemented with a preliminary proteomic survey of spore coat samples from the organism as well. These two data sets provided a large number of putative *B. anthracis* spore surface proteins that are potential targets for further investigation. These proteins could be a key towards development of strategies for disassembly or destruction of spores. Additionally, they promise to prove invaluable in the efforts to devise vaccines and rapid detection instruments against this dangerous organism.

APPENDIX

[Below is a sample of a functional Python script for one of our 'web-robots']

```
#!/usr/local/bin/python

#-----------------------------------------------------------
# SCRIPT   : cog_query.py
# AUTHOR   : NP, PJU
#-----------------------------------------------------------

from Requests import GETRequest
import string, re, sys

inputfile = sys.argv[1]
outputfile = sys.argv[2]

def get_params():
   datafile = open(inputfile, "r")
   text = datafile.read()
   lines = string.split(text, "\n")
   myseq = ""
   myorf = ""
   myfunc = ""
   seqlist = []
   dict = {}
   for line in lines:
      mseq = ""
      orf = re.match(r"^>(ORF.[0-9]*)(.*)$", line)
      seq = re.match(r"([A-Za-z]*)$", line)
      if orf:
         oldorf = myorf
         oldfunc = myfunc
         myorf = orf.group(1)
         myfunc = orf.group(2)
         mytitle = oldorf+'%OD%OA'+myseq
         if oldorf:
            dlist = [mytitle, oldfunc]
            dict[oldorf] = dlist
```

```
            seqlist.append(mytitle)
         myseq = ""
      if seq:
         mseq = seq.group(1)
         myseq = myseq + mseq
   #print seqlist
   #print dict
   return dict



def grab_results(dict):
   machine = "www.ncbi.nlm.nih.gov"
   uri = "/cgi-bin/COG/nph-cognitor99"
   port = 80

   for key in dict.keys():
      my_dic = {}
      orf_name = key
      seq = dict[key][0]
      func = dict[key][1]

      my_dic[key] = [func]
      param1 = 3
      param2 = 'seq='+seq

      #values = []

      params = {"hit=3":"hit=3", param2: param2}
      G = GETRequest(machine, uri, params, port)
      response = G.retrieve()

         print response

      res_list = parse_response(response)
      for item in res_list:
         my_dic[key].append(item)
         dict[key].append(item)
      output_results(my_dic)

def parse_response(resp):
   lines = string.split(resp, "\n")
   list = []
   for line in lines:
     cg = re.match(r"<th.*\scolor=#000000>([A-Z]*)<.*$", line)
```

```
            cn = re.match(r"<th bgcolor=#[a-z0-9]*><a href=palog?.*>(COG[0-
9]*)<.*$", line)
            fn = re.match(r"<th bgcolor=#[a-z0-9]*>([^<]*)</th>.*$", line)
            ft = re.match(r"<td bgcolor=#[a-z0-9]*>.*1>\s(.*)\s</font>", line)
            if cg:
               list.append(cg.group(1))
            if cn:
               list.append(cn.group(1))
            if fn:
               list.append(fn.group(1))
            if ft:
               list.append(ft.group(1))
         return list


      def output_results(dict):
         """ Takes the dict results and outputs it to a flatfile."""
         for key in dict.keys():
            format = "%s;%s;%s;%s;%s;$s\n"
            print dict
            if (len(dict[key]) > 2):
               print "HIT!!!!"
               a = key + ";" + dict[key][0] + ";" + dict[key][1] + ";" + dict[key][2] + ";" +
dict[key][3] + ";" + dict[key][4] + "\n"
               output.write(a)
               print "\n\n" + a + "\n\n"
            else:
               a = key + ";" + dict[key][0] + ";;No cog;;\n"
               output.write(a)



      if __name__ == '__main__':
         output = open(outputfile, "w")
         myseqlist = get_params()
         results = grab_results(myseqlist)
```