The proteome of spore surface layers in food spoiling bacteria
Abhyankar, W.R.

Citation for published version (APA):
Abhyankar, W. R. (2014). The proteome of spore surface layers in food spoiling bacteria
1

General Introduction
Stress response and Sporulation

The stress response in bacteria enables bacteria to survive extreme and fluctuating conditions in their immediate surroundings. Various bacterial mechanisms recognize different environmental changes and build an appropriate response. A bacterial cell can react simultaneously to a wide variety of stresses and the various stress response systems interact with each other by a complex of global regulatory networks[1]. These regulatory systems govern the expression of more effectors that maintain stability of the cellular equilibrium under the various conditions.

Most culturable bacteria grow and survive in rich media used to cultivate them in the laboratory. Nevertheless, in natural environments, the nutrient availability is a major hurdle for their growth. Nutrient supply is affected by diverse conditions like availability of water, competition with the other bacterial species in the environment etc. Also in certain environments, like oceans, the nutrients are extremely diluted. However, bacteria have evolved some characteristic mechanisms to adapt to such “starvation” conditions. The soil bacterium Bacillus subtilis, other Bacillus spp., anaerobic Clostridium spp. and related organisms can form endospores - small, metabolically dormant cellular structures that are remarkably resistant to heat, desiccation, radiation and chemical insult - in response to nutrient starvation. A variety of alternative responses can occur, including the activation of flagellar motility to search for new food sources by chemotaxis, the production of antibiotics to destroy competing soil microbes, the secretion of hydrolytic enzymes to scavenge extracellular proteins and polysaccharides, or the induction of ‘competence’ for uptake of exogenous DNA for consumption, with the occasional side-effect that new genetic information is stably integrated [2]. It is only after alternative responses have proven to be inadequate to relieve the stress, that sporulation is the fate chosen by a majority of the cells. Sporulation is an irreversible process where as other stress responses can be rapidly reversed by elimination of stress. Figure 1 (A) summarizes the possible fates for Bacillus subtilis cells upon nutrient starvation. The sporulation process holds to the fact that the state of dormancy achieves continuation of life and synchronization of developing stages of life forms with the environments or periods to which they are adapted. Sporulation is set-in as a response to a single stress (e.g. nutrient starvation) but the spores are resistant to many multiple stress conditions. Factors inducing sporulation and the germinants that break the dormancy both can be seen as the indicators of unfavourable and favourable periods, respectively. Thus, spore formation, albeit a stress response, may represent a timing device that perpetuates viability during unfavorable periods.

Sporulation cycle

The master regulator responsible for the decision to differentiate from a starving vegetative cell into a dormant spore is Spo0A[3]. An advanced sensing system orchestrates and directs the stress response into a phosphate flow through a network of kinases, which ends with the accumulation of Spo0A-P (phosphorylated Spo0A [4]). The level of Spo0A-P defines three stages in the decision-making process (see Figure 1 (B).
When the levels of Spo0A-P are low, the AbrB–Rok interaction cascade closes the path towards competence development. The Rap system together with two two-component sensing systems performs the “early assessment stage” [6, 7]. The assessment involves governing the probability to escape towards competence and governing the progression toward sporulation. After Spo0A-P reaches a threshold level $S_1$, the cell enters into the “decision stage”[8], during which the AbrB–Rok circuit opens a time window of opportunity (called a “competence window”) to escape into competence or make a final commitment to sporulation. The third and final “commitment stage,” is reached when Spo0A-P is accumulated above the threshold level $S_2$ [9]. The SinI-SinR signaling circuit is turned on by Spo0A-P leading to commitment to sporulation.

Figure 1. (A) Vegetative dormancy versus Sporulation. The $\sigma^B$-dependent general stress response and competence as components of a survival strategy alternative to sporulation. (Adapted from Hecker & Völker [10]) (B) Role of Spo0A in decision of sporulation. Depending on the levels of Spo0A, a cell goes through 3 stages of decision-making before it commits itself to sporulation. Spo0A* represents phosphorylated Spo0A or Spo0A-P. (Adapted from Schultz et al. [11])

The stages in the sporulation cycle of *Bacillus subtilis* are shown in Figure 2. The sequence of morphological changes resulting in the formation of a dormant endospore has been demonstrated by electron microscopy in the past and is similar for all species of *Bacillus* and *Clostridium* that have been examined. Although the process is continuous, it is convenient to divide it into different stages. Considering the vegetative cells at Stage 0, the sporulation division produces two distinct cells with very different fates, the smaller prespore (also known as the forespore), which develops into the spore, and the mother cell, which is necessary for spore formation but ultimately lyses (programmed cell death).

The condensation stage, where the two nuclei of the vegetative cell (post-asymmetric cell division) fuse to form a single axial filament of chromatin, was originally defined as Stage I. Completion of a septum formed by membrane invagination and growth at one pole of the cell takes places in Stage II. Soon after the division, distinct programs of gene expression are initiated in the two cell types. These are directed by sporulation-specific RNA polymerase $\sigma$ factors, $\sigma^F$ in the prespore and $\sigma^E$ in the mother cell. Freese [12] has suggested that at least some peptidoglycan synthesis is necessary at this time to give direction to the membrane synthesis and to ensure that a septum is formed. The
peptidoglycan that is synthesized during septation is then apparently digested away [13] so that the bacterium can proceed to Stage III. In this stage, a protoplast is formed within the mother cell. After division, the prespore is engulfed by the mother cell. On completion of the engulfment, there is another transition in transcription, with $\sigma^G$ becoming active in the prespore and $\sigma^K$ in the mother cell (Figure 3). Stage IV is the deposition of primordial germ cell wall and cortex between the membranes of the spore protoplast. Deposition of the spore coat around the cortex defines Stage V and Stage VI involves the "maturation" of the spore, at which time it develops its characteristic resistant properties. During Stage VII, the mother cell lyses and releases the completed spore. After the release, the spore undergoes further maturation [14]. These changes in gene regulation along with morphogenesis and the inter-compartmental signaling, finally lead to the development of the resistance characteristics of the mature spore.

Figure 2. Stages in the sporulation cycle. See text for the details.

Endospore structure

The resistance properties of spores are the result of its well assembled multilayered structure. The outer layers called exosporium and coat are mainly responsible for resistance and transmission of spores. Particularly, the proteinaceous coat surrounding the spore provides much of the chemical and enzymatic resistance. Beneath the coat resides a very thick layer of peptidoglycan called the cortex. Proper cortex formation is needed for dehydration of the spore core, which aids in resistance to high temperature. A germ cell wall resides under the cortex. This layer of peptidoglycan will become the cell wall of the bacterium after the endospore germinates. The inner membrane, under the germ cell wall, is a major permeability barrier against several potentially damaging chemicals. The center of the endospore, the core, exists in a very dehydrated state and houses the cell's DNA, ribosomes and large amounts of dipicolinic acid (DPA). DPA can comprise up to 10% of the spore’s dry weight and appears to play a role in maintaining spore dormancy. Small acid-soluble proteins (SASPs) are also only found in spores. They tightly bind and condense the DNA, and are in part responsible for resistance to UV light and DNA-damaging chemicals. Distinct spore layers are discussed below (see also Figure 4 (A) & (B)).
The core

The core is the innermost part of the spore and it contains the cytoplasm with cytoplasmic, ribosomal proteins and DNA. The physical state of the core cytoplasm is distinct with a water content of only 30-50%, as opposed to 70-88% in a vegetative cell cytoplasm [15]. This dehydrated state plays an important role in spore endurance, dormancy and resistance [16]. The pH in the spore core is ~6.5, which is lower than the pH of the vegetative cell cytoplasm [17]. The core contains large quantities of small acid soluble proteins i.e. SASP [18] which form a complex with the spore DNA and protect it against many types of damages [21] by keeping it in a compressed state [22]. Proteins in the core are largely immobile and the divalent cations in the core, mainly Ca$^{2+}$ form a complex with the spore-specific compound pyridine-2,6-dicarboxylic acid (dipicolinic acid or DPA [23]). DPA also plays a role in wet heat resistance [24] and UV resistance [25]. The amount of DPA in the core accounts for about 5-15% of the total spore weight [26]. DPA and calcium are excreted from the core during the first step of germination, and play important roles in further steps of the germination process [27].

Figure 3. Morphogenesis and gene regulation during spore formation. (Adapted from Piggot & Hilbert, 2004 [19])

Figure 4. Spore structure (A) B. subtilis, (B) B. anthracis. [Abbreviations: spore core (Cr), cortex peptidoglycan layer (Cx), undercoat region (Uc), inner (Ic) & the outer (Oc) coat layers, exosporium (Ex), basal layer (Bl) and a hair-like glycoprotein nap (Hn)] (Adapted from Henriques & Moran, 2007 [20])

...
Chapter 1

The inner membrane

The spore core is surrounded by the inner membrane. Upon germination it becomes the plasma membrane of the new vegetative cell and can expand to about 2-fold without any new lipid synthesis. The lipids in the inner membrane of dormant spores are also immobile, while in germinated spores and vegetative cells, the membrane is fluid and membrane lipids are highly mobile [28]. The inner membrane is proposed to be the main permeability barrier of spores. It accommodates the spore germination receptors [29]. Furthermore, the inner membrane is a target for several sporicidal chemicals [30].

The cortex and germ cell wall

The cortex is composed of a specifically modified peptidoglycan, layered around the inner membrane. The cortex is of central importance for maintaining the spore core dehydrated and thereby resistance and dormancy characteristics of spores [31]. Recent results indicated that the cortex merely serves as a static structure that maintains spore core dehydration [31]. The cortex peptidoglycan is loosely cross-linked [32] and variation in the degree of cross-linking within the cortex region may play a role in the pressure that the cortex exerts on the core [31]. This mechanistic pressure has been suggested to be important for core dehydration. The cortex peptidoglycan has two unique structural modifications [33] (Figure 5). Approximately, half of the muramic acid in the cortex is present in the form of muramic acid lactam (MAL). These lactam residues are found alternating with NAM (N-acetyl muramic acid) residues within the polyglycan strands. Also, about 25% of the NAM monomers have only an L-alanine side chain [34]. The peptides on the remaining NAM residues are cross-linked with each other [33, 34]. The function of the MAL residues remains unclear. The inner part of the cortex called the germ cell wall lacks the specific modifications that are characteristic of the cortex peptidoglycan. The specific structure of the cortex is conserved among species, and may play a role in spore heat resistance, but it has been shown that the volume ratio of the cortex peptidoglycan to the germ cell wall contributes to the formation of a more heat-resistant spore in C. perfringens [35]. During germination, the cortex peptidoglycan is rapidly degraded by the spore cortex lytic enzymes (SCLEs) already present in the dormant spore. The germ cell wall is not degraded upon germination and forms the initial cell wall of the freshly germinated spore [31].

The outer membrane

Around the cortex lies the relatively poorly studied outer membrane. The outer membrane may have a function during spore formation but some studies in the past concluded that the outer membrane is quite permeable even to large molecules [36]. Nonetheless, the outer membrane is also identified as the main permeability barrier in bacterial spores [37]. It is reported to be intact in dormant B. megaterium spores [38] and a functional barrier to the diffusion of large ions in B. subtilis and B. cereus spores [27].
General Introduction

Figure 5. Structure of B. subtilis spore peptidoglycan. NAM carries side chains of L-alanine, the tetrapeptide L-ala-γ-D-glu-diaminopimelic acid-D-ala, or the tripeptide L-ala-γ-D-glu-diaminopimelic acid. Approximately 50% of the muramic acid residues have been converted to MAL, which is found with great regularity at every second muramic acid position. (Adapted from Popham et al. [34])

The coat

The coat is the most extensively studied layer of spores [20, 39]. It is built around the outer membrane, and is a dynamic, intricate protein structure generally consisting of three distinct layers - an amorphous undercoat layer, a lamellar inner coat, and an electron-dense striated outer coat [40]. The coat proteins constitute up to 10% of the spore dry weight and up to 25% of the entire spore proteome [41]. The coat protects the cortex peptidoglycan from enzymatic attack [42]. Furthermore, it is involved in resistance to environmental UV radiation [43, 44], to a variety of chemicals including oxidative agents [45, 46] but not significantly in resistance to wet heat [39, 47]. The coat contains enzymes, such as laccases, which may be active even when the spore core is devoid of metabolic activity [48, 49]. These enzymes may have a significant function in spore ecology [50]. The inner layer of the coat carries lytic enzymes, which help degrading the cortex during germination [51, 52]. Several other coat proteins are also involved in spore germination, by facilitating the passage of specific germinant molecules through the coat [53]. Recently, with the application of techniques new to spore-research, such as atomic force microscopy [54] and automated scanning microscopy [55], it has become clear that the coat is a very dynamic structure. However, 30% of the coat proteins constitute an insoluble protein fraction, characterized by extensive inter-protein cross-liking [20], that is resistant to proteolytic enzymes and thus remains to be studied in detail. Researchers have suggested the possible presence of di-tyrosine crosslinks [56, 57], ε-γ-glutamyl-lysine crosslinks [58] and the di-sulfide linkages among the proteins in this fraction. Nevertheless, the role of many coat proteins in cellular physiology remains to be elucidated.

The exosporium and the crust

In many species the spore coat is surrounded by a loose, membrane-like, glycoproteinaceous structure called the exosporium (Figure 3). The exosporium is important for spore hydrophobicity and adherence properties [59-61]. Since long, spores of B. cereus, B. anthracis and the Clostridium spp. are known to possess an exosporium [62]. B. subtilis spores are, in contrast, devoid of an exosporium. As it is difficult to
obtain sufficient quantities of the exosporium for study, the structure has received little attention over the years. Nevertheless, with renewed efforts in recent years the exosporium from *B. cereus* and *B. anthracis* has been analyzed in considerable detail. To this extent the BclA protein from the exosporium has been studied extensively for its structure as well as its role in spore-macrophage interactions [63, 64]. Similar to the important antigens that could serve for detection and identification [65, 66], the exosporium has been found to contain a number of enzymes, such as alanine racemase, nucleoside hydrolases, immune inhibitors etc. some of which are possibly involved in germination [67-69]. Another interesting finding is the presence of a manganese oxidizing enzyme in the exosporium from a marine *Bacillus* spp. which, in the natural environment, encases the spore in a metal shell, thereby increasing its resistance [70]. Recently, it was found that the ExsA protein is indispensable for anchoring the exosporium to the coat, while the ExsA equivalent i.e. SafA in *B. subtilis* 168 is essential for proper coat assembly [71]. Several other proteins of the *B. cereus* exosporium have homology to *B. subtilis* coat proteins, and therefore it is foreseen that the exosporium is a specialized and further decorated coat layer [71]. Lastly, a layer similar to the exosporium was assigned recently to the *Bacillus subtilis* spores named ‘crust’ [72]. This layer is claimed to be a glycoproteinaceous layer surrounding the coat. Interestingly, a layer similar to the *B. subtilis* crust spore surface was recently identified also from the *B. anthracis* spore in a freeze-etching study [73] indicating commonalities among the spore structure.

**Spores in Food Industry**

In the late 1700s, Nicolas Appert invented the process of appertization, now known as canning, to extend the food quality, shelf life and to prevent the spoilage of food over a longer time. He believed that the elimination of air was responsible for the stability of canned food. But later in 1875, Cohn & Koch for the first time discovered spores and at the same time Pasteur discovered that spore-forming bacteria caused food spoilage during investigations of butyric acid fermentation in wines. These contributions linked microbial activity with food quality and safety. In the late 1800s and early 1900s Prescott and Underwood (Massachusetts Institute of Technology, USA) along with Russell (University of Wisconsin, USA) found that spore-forming *Bacilli* caused the spoilage of thermally processed clams, lobsters, and corn [74]. Russell also showed that gaseous swelling with bad odors in canned peas was due to growth of heat-resistant bacterial spores [75]. Since then diseases and spoilage caused by spore formers are associated with thermally processed foods, as heat kills the vegetative cells but allows survival and growth of spore-forming organisms.

In comparison to the other food-borne pathogenic bacteria, spores survive better under conditions prevailing in the food and during food processing and thus causing big problems for the food industries. Three species of spore-formers - *Clostridium botulinum*, *Clostridium perfringens*, and *Bacillus cereus*, are especially notorious toxin producers. Other spore forming species cause spoilage. Sporulating bacteria causing food-borne illness and spoilage are particularly important in low-acid foods (pH $\geq 4.6$) packaged in cans *id est* “canned” foods, which are processed by heat.
Other sporeformers cause spoilage of high-acid foods (pH ≤ 4.6). Psychrotrophic sporeformers such as *Bacillus weihenstephanensis* cause spoilage of refrigerated foods (See Table 1). Meat products are also often contaminated with spores of anaerobic *Clostridia* [76]. Modeling inactivation of food-borne pathogens has been one of the first achievements in predictive microbiology. Inactivation models were initially focused on the destruction of *C. botulinum* spores in low acid canned foods [77, 78]. According to the recent report of European Food Safety Authority (EFSA) [79], mild heat treatment (e.g. very lightly cooked food or pasteurization) of a few seconds at 70°C permits several log<sub>10</sub> units of inactivation of vegetative bacterial pathogens and parasites, but may not be sufficient for food-borne viruses and will not inactivate bacterial spores nor bacterial toxins formed in foods. Also cooking of foods for several minutes at 90°-100°C practically eliminates vegetative bacterial pathogens, food-borne viruses, and parasites allowing several log<sub>10</sub> units of inactivation of spores of non-proteolytic *C. botulinum* (psychrotolerant *C. botulinum*), but not of pathogenic *Bacillus* and other pathogenic *Clostridium spp*. Additionally the toxins from *C. botulinum* would be inactivated but not the emetic toxin from *B. cereus*. In general, heating to 100°C may allow survival of thermoduric spores, while lower temperatures may lead to spore activation. Spores cannot grow in dry fat, but they may grow in formulated products or in formulations with moist materials in processed foods [80]. However, according to the EFSA report a reliable inactivation of spores of pathogenic bacteria in high a<sub>w</sub> (water activity) foods can be achieved by sterilization treatments (e.g. 3 min at 120°C). Killing of spores by inducing their germination and subsequently inactivating the vegetative cells by pasteurization level thermal treatments are thought to be good preservation options for the food industries. In addition, preservatives like weak organic acids (e.g. sorbic acid), are being studied for their effects on spore germination and outgrowth [81, 82]. In order to minimize or eliminate the problems caused by the spores, it is very important to detect and estimate the amount of spores in the food and from the patient’s sample. Efficient removal of spores is possible when simple and quick spore detection and removal systems are available. To build such systems it becomes necessary to understand the properties of spores as well as to gain the knowledge about the entire sporulation and germination pathways. Efforts in the past have led to the understanding of the mechanism of sporulation and the signaling cascade that initiates the whole cycle yet the complete sporulation models have not been mapped for most of the spore formers. Also the signaling mechanisms that operate when a dormant spore germinates yet remain to be uncovered although the gene-expression processes operative during germination have been described [83, 84].

**Medical significance of spores**

Bacterial spores came into focus in 2001 after the Anthrax bio-terror attacks in the USA. The attacks used spore powder from *B. anthracis* and the spores after germination spread the Anthrax disease to people killing 22 of them. According to the Federal Bureau of Investigations (FBI), the ensuing investigation became “one of the largest and most complex in the history of law enforcement”. In hospitals, spores from *C. difficile* are often
### Table 1. Spoilage of canned foods by sporeformers.

<table>
<thead>
<tr>
<th>Type of spoilage</th>
<th>pH</th>
<th>Major sporeformers responsible</th>
<th>Spoilage defects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flat sour</td>
<td>≥ 5.3</td>
<td><em>B. coagulans</em></td>
<td>No gas, pH lowered. May have abnormal odor and cloudy liquor.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>B. stearothermophilus</em></td>
<td></td>
</tr>
<tr>
<td>Thermophilic anaerobe</td>
<td>≥ 4.8</td>
<td><em>C. thermosaccharolyticum</em></td>
<td>The can swells, may burst. Anaerobic end products give sour, fermented, or butyric odor. Typical foods are spinach, corn.</td>
</tr>
<tr>
<td>Sulfide spoilage</td>
<td>≥ 5.3</td>
<td><em>D. nigrificans</em></td>
<td>Hydrogen sulfide produced, giving rotten egg odor. Iron sulfide precipitate gives blackened appearance. Typical foods are corn, peas.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>C. bifermentans</em></td>
<td></td>
</tr>
<tr>
<td>Putrefactive anaerobe</td>
<td>≥ 4.8</td>
<td><em>C. sporogens</em></td>
<td>Plentiful gas. Disgusting putrid odor. pH often increased. Typical foods are corn, asparagus.</td>
</tr>
<tr>
<td>Psychrotrophic <em>Clostridia</em></td>
<td>&gt; 4.6</td>
<td><em>C. estertheticum</em></td>
<td>Spoilage of vacuum-packaged chilled meats. Production of gas, off flavors and odors, discoloration.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>C. algidicarnis</em></td>
<td></td>
</tr>
<tr>
<td>Aerobic sporeformers</td>
<td>≥ 4.8</td>
<td><em>Bacillus spp.</em></td>
<td>Gas usually absent except for cured meats; milk is coagulated. Typical foods are milk, meat, beets.</td>
</tr>
<tr>
<td>Butyric spoilage</td>
<td>≥ 4.0</td>
<td><em>C. butyricum</em></td>
<td>Gas, acetic and butyric odor. Typical foods are tomatoes, peas, olives, cucumbers.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>C. tertium</em></td>
<td></td>
</tr>
<tr>
<td>Acid spoilage</td>
<td>≥ 4.2</td>
<td><em>B. thermoacidurans</em></td>
<td>Flat (<em>Bacillus</em>) or gas (butyric anaerobes). Off odors depend on organism. Common foods are tomatoes, tomato products etc.</td>
</tr>
<tr>
<td></td>
<td>&lt; 4</td>
<td><em>Alicyclobacillus acidoterrestris</em></td>
<td>Flat spoilage with off flavors. Most common in fruit juices, acid vegetables, and also reported to spoil iced tea.</td>
</tr>
</tbody>
</table>

---

Adapted from Food Microbiology: Fundamentals and Frontiers, 3rd Ed. [85]
attack. In a recent study, using simulations of endospore-laden surfaces under laboratory conditions, the effects of exposure time, disinfectant concentration and possible synergies with endospore germination stimulant were evaluated. None of the disinfectants tested was found to be sporicidal [87]. Also, killing the non-spore-laden vegetative cells causing the infection does not ensure the death of the spores at the same time. The spores can germinate & further re-initiate the infection later. Hence, for spore forming bacteria, the medical procedure has to be designed to get rid of the bacteria as well as the spores.

**Mass spectrometry (MS)-driven Proteomics**

The word “proteome” is derived from the proteins expressed by a genome, and it refers to all the proteins produced by an organism. Proteins are important players in a living organism, involved in a plethora of activities which eventually lead to a particular cellular phenotype. As a consequence, protein analysis is of major interest in cellular and molecular biology and involves protein identification, quantification, localization as well as the study of post-translational modifications. To achieve this, many methods starting from Edman sequencing methodology [88] to mass spectrometry (MS) - based proteomics, have been invented and developed over the past few decades. Prior to mass spectrometry-based proteomics, proteome analysis required separation of proteins followed by protein sequencing through Edman sequencing. Thus several different gel-based techniques have been developed that make use of physicochemical properties of proteins to separate them. These methodologies involved separation based on - the isoelectric potentials of proteins i.e. isoelectric focusing (IEF) [89], the molecular weight of proteins i.e. SDS-poly acrylamide gel electrophoresis (SDS-PAGE) [90] and both isoelectric point as well as molecular weight in a two-dimensional gel electrophoresis (2-DE) approach [91]. After separation, proteins could be visualized on the gel by different means such as Coomassie Brilliant blue staining, silver staining or immunoblotting methods followed by Edman sequencing of the proteins from the spots evident on the gels [92]. This approach required a large amount of protein material allowing in most cases only the abundant proteins to fall in the scope of the method.

Through the advent of new ionization methods such as electrospray ionization (ESI) [93], matrix-assisted laser desorption/ionization (MALDI) [94, 95], much more sensitive techniques of mass spectrometry became available for protein detection and analysis, requiring much less protein material. Inside the mass spectrometer, the peptides are fragmented by cleavage of, most commonly, peptide bonds thereby generating fragments ions of different charge states. Cleavage of peptide bond leads to formation of two fragments. With regards to these fragments if one peptide fragment retains the positive charge at the carboxyl (C) - terminus of the peptide ion, the ion is called a \( y \)-ion. If the fragment retains the positive charge at the amino (N) - terminus, the ion is known as a \( b \)-ion. The amount of the \( y \) and \( b \)-ions generated from the peptides is generally an indicative of the quality of the spectrum generated for that peptide. Thus protein identification is achieved by matching all the spectra generated from the peptides identified from the protein to the theoretical spectra generated for the peptides with the help of sequence databases. Accurate mass measurements of peptides by mass
spectrometry essentially developed further as more and more genome and/or protein sequences were made available in databases wherein recorded masses could be matched. In a routine proteomic analysis, after gel-based separation, a protein spot is subjected to in-gel digestion using proteases such as trypsin, after which the resulting peptides are extracted and their masses are determined by mass spectrometry. The list of peptide masses, resulting from a single gel-spot, is then used as a fingerprint to identify the protein using a protein sequence database and a specialized search engine in a method denoted as peptide mass fingerprinting (PMF). The major drawbacks of PMF are that the amino acid sequence of protein cannot be directly determined in addition to the fact that the proteins must be purified prior to digestion. As an alternative, peptide-centric proteomics developed as a new standard [96] for proteome analyses. This approach focuses on the separation of peptides by liquid chromatography rather than the separation of proteins. Chromatography can be performed in-line coupled to tandem mass spectrometry (MS/MS), where resulting spectra allow determination of peptide sequences, which can then be used to infer the parent proteins [97]. In this approach multiple peptides can be utilized to identify proteins and since peptides are less extreme in their physicochemical parameters such as the molecular mass, the hydrophobic or hydrophilic nature, the isoelectric points etc., peptide-centric proteomics becomes highly sensitive. In addition, peptide-centric proteomics also allows qualitative proteomics (comprehensively mapping the presence of all the proteins in the sample) and quantitative proteomics (quantification of changes in protein abundance between samples). Both areas of research are covered in some detail below.

**Qualitative proteomics**

Proteome coverage is mainly influenced by three factors - (a) the sensitivity of the mass spectrometer *id est*, in this context, the lowest amount of the sample that can be detected; (b) the dynamic range of the instrument or the signal intensity range of the instrument, (c) the duty cycle of the mass spectrometer *id est* the number of fragmentation spectra (with a fair amount of complementary \( b \) and \( y \)-ions) that the mass spectrometer can produce within a given time frame. In tandem mass spectrometry (MS/MS), data acquisition with the aid of computers is a routine practice. The data acquisition depends on the ion abundance of the peptides in the sample. Also in a complex peptide mixture, co-elution of certain peptides is routinely observed. Therefore the number of ions co-eluting can significantly affect the number of ions for which tandem mass spectra can be acquired. As a result, the data acquisition can be biased against the low abundant ion signals representing the peptides present at low levels. This also leads to random sampling of ions for fragmentation in the mass spectrometer. Requirement of sample fractionation to reduce complexity of samples prior to the mass spectrometry is also essential for complex proteome sample. Moreover, since a proteomics method does not inherently involve amplification of sample peptides, when a low abundant or any protein is not identified, no inference can be drawn on whether the protein is absent in the sample or it falls outside the detection limits of the instrument used. This has not constrained the qualitative proteomics to produce continuously increasing lists of identified peptides.
General Introduction

thereby increasing the proteome coverage. Over the years, researchers have invented strategies to avoid problems arising due to random sampling and to reduce complexity of the peptide mixture resulting from proteome digestion. The digestion of proteins followed by separation of obtained peptides by liquid chromatography (LC) - MS/MS analysis *id est* a shotgun proteomics strategy or selection and analysis of only a targeted set of peptides related to proteomic experiment *id est* a targeted proteomics strategy are the main approaches in proteomics studies. An overview of the proteome composition is readily generated by shotgun proteomics strategies (*e.g.* Multidimensional Protein Identification Technology (MUDPIT) [98] and Gel LC-MS/MS [99]) in which many proteins are identified by multiple peptide sequences per protein, which increases the reliability of such identifications. Whereas in targeted proteomics strategies (*e.g.* Combined Fractional Diagonal Chromatography (COFRADIC) and Multiple Reaction Monitoring (MRM) [100-103]) the selection of set of peptides is such that it represents the analyzed proteome or target proteins, respectively. Since a selection of a subset of peptides yields a less dense peptide mixture, random sampling tends to be reduced. In addition, targeted proteomics can also be applied in identification of post-translationally modified peptides and yet many possible protein modifications remain challenge in current proteomics framework. With the development of software tools such as MASCOT [104] and OMSAA [105], the identification of proteins by database searching has becomes more accurate as in such cases probabilistic scoring matrices are used to score for the peptide identity. In routine practice protein or peptide MASCOT scores above 20 are accepted as reliable identifications. To consider the role of post-translational modifications in protein chemistry and to study them in the perspective of “omics” strategies and systems biology, measurement of their abundance in a variety of conditions is mandatory.

Quantitative proteomics

Qualitative analysis provides a compositional map of proteins which can be extended with relative or absolute abundance information by quantitative proteomics. Quantitative proteomics deals with samples that may vary in cellular phenotypes, provided stimuli, time durations of the stimulus or many other cellular states in which difference in protein composition can be expected [106]. Quantitative studies are mainly done by two approaches. In one approach, mass tags, such as isotopic labels, are added which allow differentiation between peptides from distinct samples. Such mass tags can be administered into proteins or peptides metabolically, by chemical means and enzymatically [107, 108]. Figure 6 illustrates different quantitative proteomics strategies based on the use of mass tags. As seen, the *in vivo* techniques involve incorporation of a stable isotope into the proteins which can be achieved via the addition of an isotope of an element (*e.g.* $^{13}$C, $^{15}$N, or $^{18}$O as salts or amino acids) to the growth media in a form that makes it suitable for incorporation into the cell, tissue or even the entire organism [109]. In these cases, protein identification is carried out by analyzing the fragmentation spectra of at least one of the co-eluting ‘heavy’ and ‘light’ peptides and subsequently the relative quantitation is achieved by comparing the intensities of isotope clusters of the intact
peptides in the combined LC-MS spectrum. The heavy isotope of nitrogen ($^{15}$N) is more commonly used as the stable isotope due to the lower number of nitrogen atoms in a peptide compared to the number of carbon atoms. For the same reason, with $^{13}$C as a label, the mass shift in the heavy peak compared to its light equivalent can be very large, making quantitative analysis challenging. Higher number of carbon atoms also makes it difficult to get a completely labeled proteome in short time, as a considerable exchange of $^{12}$C and $^{13}$C takes place within the cell during various cellular activities. Deuterium ($^2$H) labeling, though efficient, is rarely used due to its potential toxic effects on the organism under study [110]. The technique developed by Ong and Mann [111], called stable isotope labeling with amino acids in cell culture (SILAC; Figure 6 (A)), has been widely applied and has proved suitable for a variety of organisms [112-115]. In a routine SILAC experiment, $^{13}$C/$^{15}$N-labeled lysine and/or arginine is used, that gives a fixed mass difference for the peptides when trypsin or Lys-C is used as a protease. However, one of the limitations faced by this technique is that the cells or the organism to be used in these experiments needs to be an auxotroph for the labeled amino acid(s) to confirm that the labeled amino acid is the only source for protein synthesis. Also especially in eukaryotes, arginine can be metabolically converted to proline [116] thereby interfering with the protein quantification if arginine is used for SILAC-labeling [109]. For in vitro labeling (Figure 6 (B)), typically, isotope coded affinity tags (ICAT) that label the cysteines in the proteins can be used but the success of the method depends on the amount of cysteines available in the proteins as well as on good purification methods. The relative abundances of the peptides can then be determined comparing the ratio of the light and heavy forms of each peptide. Likewise, using isobaric tags for relative and absolute quantification (iTRAQ) or tandem mass tags (TMT) that label the N-termini and the Lysine-side chains in the digested peptides, quantification can be achieved. These tags contain four regions - a mass reporter (M), a cleavable linker (F), a mass normalizer (N) and a protein reactive group (R). Structurally all the tags are identical but they differ in the positions of isotope substitutions, such that the mass reporter and mass normalization regions differ in their molecular masses in each tag. These combined M-F-N-R regions of the tags have the same total molecular weights and thus during chromatographic or electrophoretic separation and in single MS mode, molecules labeled with different tags are not distinguished. Further, during peptide fragmentation, the isobaric amine groups, introduced by the iTRAQ and TMT labels, are also fragmented giving reporter ions with distinct m/z ratios. Relative peptide abundances are then estimated by comparing the intensities of these reporter ions [117, 118]. Therefore, for such studies sufficient fragmentation of the labelled peptides is of prime importance. Since, the in vitro labeling is performed after protein extraction it remains independent of the source and preparation of the sample allowing practically any type of biological sample to be labelled. Also, the time needed for the chemical or in vitro tagging is generally much shorter than when a label is incorporated metabolically where it may take weeks to in vivo label the organisms or cells depending on the growth rate. The in vivo labeling techniques are still preferred over the in vitro labeling methods due to their ingrained advantages that (i) the in vivo labeling does not suffer from side reactions or incomplete labeling that might occur in
General Introduction

chemical derivatization and (ii) the metabolic labeling occurs at the earliest possible moment in the sample preparation process, thus minimizing the errors in quantification.

Isotope labeling may involve a number of technical difficulties such as the specific requirements for metabolic labeling or the problems of reproducible chemical labeling. Thus, in the second quantitative proteomics approach, *id est* label-free quantification, aligned peak intensity profiles from LC-MS or LC-MS/MS analyses are integrated to find the differences in protein abundances. Quantitative protein comparisons are based on the relative intensities of extracted ion chromatograms from their tryptic peptides. Resulting datasets, usually in triplicate, are aligned using peptide mass and LC retention times. The spectral count (SC) based label-free method is easy, fast and correlates well with isotope-labeling quantification \[119\]. It involves counting the number of peptide MS/MS spectra assigned to a protein in an LC-MS/MS experiment and allows both relative and absolute quantification of protein abundance. The second method takes into account the chromatographic peak area under the curve (AUC) or the signal intensity measurement of the peptide precursor ion in the MS spectra \[120\]. In an alternate approach proposed by Rappsilber et al. \[121\], the protein abundance index (PAI) can be calculated by calculating the ratio of the observed unique peptides and theoretical number of tryptic peptides within a given m/z range for a certain protein in the sequence database. This method was further improved by Ishihama and co-workers \[122\], where the

Figure 6. (A) *In vivo* and (B) *In vitro* approaches for labeling of proteins and/or peptides for mass spectrometry based quantitative proteomics. See the text for details.
exponentially modified PAI (emPAI) was proposed as a direct estimation of the protein amount in the sample. However, the emPAI method fails in case that the sample contains highly abundant proteins or if mass spectrometers with low resolution are used for the analysis [107]. A recently proposed method, called absolute protein expression (APEX) [123], based on the probability of detection of the peptides by MS makes use of a machine learning classification algorithm for peptide length and residue composition. However for this purpose the initial training set of peptide queries for the algorithm is crucial especially if the sample contains unknown proteins. To circumvent these problems the use of the average of the total ion count for the identified tryptic peptides of a the protein id est the total ion chromatogram (TIC) has also been proposed [124]. Albeit all these efforts, the label-free experiments need to be more carefully controlled, due to possible error caused by run-to-run variations in performance of LC and MS. However, the development of highly reproducible nano-HPLC separation, high resolution mass spectrometers and delicate computational tools has been aimed at to improve the reliability and accuracy of label-free, quantification. Yet, the label-free approaches are less accurate among the mass spectrometry based quantification techniques when considering the overall experimental process because all the variations between experiments are reflected in the obtained data. Consequently, the number of experimental steps should be minimum and reproducibility at each step needs to be controlled. Nevertheless, label-free quantification is worth considering for a number of reasons including costs, simpler experimental steps, the proteome coverage of the sample and comparison of multiple samples or experimental conditions.

Spore Proteomics

Due to their structure and composition, spores are particularly difficult to lyse. Various enzymatic, chemical and physical methods have been mentioned in the literature to solubilize the spore coat proteins from pure spore preparations. The thick layers of extensively cross-linked coat proteins and the complexity of the protein sample that can be solubilized from wild-type coats pose an analytical challenge. Approximately one-third of total coat protein in wild-type spores is resistant to extraction procedures normally used to solubilize the majority of proteins [125]. Differences have also been found in the solubility of different coat fractions [126] which could complicate the spore protein extraction further. In 1978, Goldman and Tipper thoroughly analyzed the spore coat protein fraction from *B. subtilis* strain 168. They could solubilize 65% of the protein from the spore coat whereby they identified a wide mass range (9000 – 16000 Da) of low-molecular weight peptides [127]. In their study, Goldman & Tipper used a two-step protein extraction procedure. In the first step, the coat fraction was subjected to 1% SDS & 50 mM dithiothreitol in sodium carbonate buffer (pH 10) followed by incubation at 37°C for 20 min. After centrifugation, the coat fractions were then subjected to 3% SDS & 2% β-mercaptoethanol in tris (hydroxymethyl)-aminomethane (Tris) - hydrochloride buffer (pH 6.8) followed by boiling at 100°C for 3 min. The protein purification, fractionation by classical SDS-gel electrophoresis and Sephadex chromatography was followed by protein characterization by either molecular approaches or more recently by
bioinformatics analyses. This effort paved the way for future studies as many groups used strong alkalis, detergents and reducing agents alone or in combination to solubilize, purify, identify and characterize the spore coat proteins [126, 128, 129].

With the advent of GFP-fusion studies, where a copy of GFP is tagged to the protein of interest, as well as with the progress in the transcriptomic i.e. microarray field, studying the expression and localization of spore coat proteins as well as the exosporium proteins in species forming spores with this additional outer layer, was possible. But it was by virtue of mass spectrometry that the study of proteome-wide data became easier. Thus combining all the methods discussed above, proteome characterization of B. megaterium [130, 131], B. cereus [132], B. anthracis [133], B. thuringiensis [134] as well as the members of the Clostridium family like C. perfringens [135], C. novyi [136], C. difficile [137-139] spore coats has been approached successfully. With all these efforts coalesced, till now, at least 70 proteins have been assigned to the coat. However, the spore coat structures from Clostridia (compared to the Bacilli) as well as the insoluble protein fraction from the coat have not yet attracted the focus. This fraction is the focus of our study presented in this thesis.

These extensive efforts to study the spore coats have up till now allowed researchers to employ endospores in various applications. Cutting and co-workers have studied the potential of spores as probiotics in food [140, 141] as well as drug vehicles [142]. Use of spores as surface display systems has been suggested [143] and has been shown to be effective [144] in recent years. Further on, applications of spores as platforms for bio-analytical and biomedical applications are very well reviewed by Knecht and co-workers [145]. Food industries are especially in search of efficient, simple and quick methods to detect and control spores from processed food samples. Spore coat and exosporium proteins can prove apt for such purposes as seen from the studies mentioned above and thus it is necessary to pay focus on proteome analysis of the spore surface layers from many different spore formers.

Outline of the thesis

Bacterial spores are a major problem in the food industry as well as the hospitals. Their high resistance allows them to escape the food processing treatments designed to inactivate bacteria. In the final food products, the spores can germinate, outgrow and multiply, leading to food-spoilage and intoxication. To reduce spore-related problems, the industries invest a substantial amount in food quality control and special cares are assigned in hospitals. Consumers nowadays prefer less processed food thereby allowing easy survival of spores. Also on the contrary, to kill the spores if food is treated with extreme measures then the nutritious properties as well as the organoleptic characters of food items are lost. Thus more thorough understanding of the resistant mechanisms of spores as well as the possible structural differences in spores of different species might enable the industries to design processes that will more efficiently eliminate spores. This would lead to an extended shelf-life of food products, a reduction of spoilage and poisoning events, cheaper processes and better product quality as well as restricted spore-mediated outbreaks.
In this chapter the basic knowledge about sporulation, spore structures and the importance of spores are described. Mass spectrometry based proteomics strategies for qualitative and quantitative proteomics are also briefly discussed in this introductory chapter. With regards to the problem of gel-based identification of spore coat proteins, **Chapter 2** describes the establishment of a comprehensive proteomics method that allowed us to focus on the insoluble protein fraction of spore coats in a lab-isolate as well as a food isolate. In **Chapter 3**, the extension of the newly developed method to genetically distinct spore forms that even belong to different bacterial domains (aerobe and anaerobe) is discussed and along with identification of few potential candidate marker proteins the relevance of the identified proteins to the spores is also discussed. **Chapter 4** focusses on mass spectrometry based identification of possible interdependences amongst spore coat proteins in *B. subtilis*. **Chapter 5** shows how mass spectrometry can be tailored to monitor the progress in spore coat maturation and thereby stresses the role of spore coat protein cross-linking. Finally, in **Chapter 6**, molecular properties of the proteins identified in our studies are summarized and placed in the perspective with future directions and a discussion on their practical applications.

**References**

23. Powell JF. Isolation of dipicolinic acid (pyridine-2:6-dicarboxylic acid) from spores of Bacillus megatherium. The Biochemical journal. 1953;54(2):210-1. Epub 1953/05/01.
General Introduction

Chapter 1


116. Van Hoof D, Pinkse MWH, Oostwaard DW-V, Mummery CL, Heck AJR, Krijgsveeld J. An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics. Nat Meth. 2007;4(9):677-8.


