The proteome of spore surface layers in food spoiling bacteria
Abhyankar, W.R.
General Discussion & Outlook
Throughout this thesis, the sub-proteomes of the surface layers of the bacterial spores have been analysed and discussed extensively. Beginning with the development of a gel-free method (Chapter 2) that focussed on the insoluble fraction of the spore coat protein from B. subtilis, this focus was broadened from the laboratory strain and a heat resistant food-isolate to the sub-proteomes of the spore coat and exosporium layers of two important pathogenic species - Bacillus cereus, a Gram positive, facultatively anaerobic organism and Clostridium difficile, a Gram positive, anaerobic organism (Chapter 3). These studies revealed that spore coat protein isolation and identification has not yet reached the ultimate and many more proteins can still be identified with the aid of newly developed technology provided precise and meticulous experimentation is done. Also the study on B. cereus and C. difficile strain shed light on the absolute requirement of proper and accurate genome annotations. Much of the identified B. cereus and C. difficile proteins have not been studied for their functional roles thereby opening the door for, yet unknown, possible mechanisms of spore survival and integrity.

In order to examine the possible interdependence of spore coat and exosporium proteins, some mutants in spore coat proteins of B. subtilis were also analysed. This study (Chapter 4), in agreement with the previous literature, reveals that the deletion of a single coat protein gene, with the exception of morphogenetic protein coding genes, does not significantly change the spore coat architecture. As seen in our study, in case of the cotE mutant the outer coat proteins are considerably reduced in amounts when compared to the wild-type strain. Such effects are not apparent in any other mutants analysed.

Inter-protein cross-linking, spore maturation and spore resistance to thermal stress are proposed to be linked together and therefore we also have probed the progress in spore coat protein cross-linking using 15N-labeled mature spores as a reference (Chapter 5). This study, for the first time, reveals a set of proteins that are really affected during spore maturation with regards to their protease resistance. These proteins are all said to be involved in protein cross-linking thus providing a strong evidence for the role of protein cross-linking in spore’s resistance towards thermal stress.

During all the studies that shaped this thesis, there are certain molecular properties of the spore coat and exosporium proteins that become evident from our investigations. In Chapter 6 such molecular properties are discussed by taking spore structure, resistance properties and potential applications of spores into consideration. This bioinformatics-based study made some hypothesis regarding the protein pI, surface charge of spores, the inherent protein sequences and repeats. These hypotheses need to be proved by further studies.

Though the comprehensiveness of our research has been demonstrated through this thesis there are certain observations that can be further acted upon in the future. These points are discussed below.
1. Study of cross-linked proteins in the bacterial spore coat

(a) Protein digestion

As mentioned in Chapter 6, the coat proteins are very peculiarly structured and functionally oriented. Though we developed a method for characterizing the insoluble fraction, the nature and abundance of cross-links has not been studied in detail. Also the stoichiometry of coat proteins involved in cross-linking needs to be focussed upon further. For the complete characterization of the insoluble fraction it is desirable to first achieve complete or maximum digestion of the insoluble fraction. The method of acid hydrolysis has been used previously by researchers to digest *Bacillus* spore proteins [1-3]. Unfortunately, these studies were limited only up to the estimation of amino acid compositions of these proteins and did not venture the possibility of identifying the cross-links. The chemical cleavage with acid provides an effective alternative to enzymatic digestion but also leads to formation of unspecific cleavage products post-digestion [4, 5]. Heijnis et al. [6] have shown the use of trypsin and glu-C treatments for the identification of di-tyrosine crosslinks from α-lactalbumin. In the work presented in this thesis we have used trypsin to achieve the protein digestion but use of such multiple protein digestion steps will allow the study of cross-linked proteins.

(b) Identification of cross-linked peptides

The insoluble fraction of spore coat proteins is said to comprise of three types of cross-links: di-tyrosine cross-links, the glutamyl-lysine cross-links and the disulfide linkages (Figure 1). Our bioinformatic analysis of amino acid compositions of spore surface proteins has shown that these proteins are, in general, rich in amino acids tyrosine, glutamic acid, glutamine, aspartic acid, asparagine, lysine and cysteine. All these amino residues reflect the possibilities of the presence of the above mentioned cross-links amongst the spore coat proteins.

In a previous study [2] the authors have suggested the presence of di-tyrosine residues amongst the coat proteins by studying the emission spectra of di-tyrosine (synthesized *in vitro*) along with the spore coat samples that were subjected to performic acid oxidation prior to the analysis. A shoulder in the region 390 - 400 nm, observed in the emission spectrum, was attributed to di-tyrosine and ~3 di-tyrosine moieties were predicted per 60000 Da of amino residues. However, the authors also have suggested that the increase in the di-tyrosine content of oxidized coat material might be due to the earlier performic acid oxidation. These authors also made use of acid hydrolysis in their study. Therefore, using a similar approach along with the spectroscopic methods the di-tyrosine estimation is possible. Di-tyrosine moiety has an absorption maximum centered around 318 nm [7] therefore the use of fluorescence spectroscopy is recommended for more precise determinations.

It is also possible to detect di-tyrosine by performing amino acid analysis with or without further mass spectrometric analysis. In such experiments though, a chromatographic separation of the peptide material is needed. Alternative approaches that can be used involve use of antibodies against di-tyrosine [8], thin-layer chromatography...
Figure 1. The types of cross-links predicted to comprise the spore coat insoluble fraction.

(TLC)-based as well as HPLC-based methods [9] and use of mass spectrometric quantification of di-tyrosine based on gas chromatography mass spectrometry (GC-MS), selected ion monitoring (SIM)-GC-MS as well as LC-ESI-MS/MS methods [10]. These approaches are rather costly, and require extensive sample preparation. Moreover, each of these methods has its own pitfalls which are comprehensively reviewed in the literature [10] and all the above methods are only applicable for preparative-scale experiments, to limited quantities of biological samples and can separate di-tyrosine from other hydrolyzed amino acids and are not applicable to a heterogeneous and complex protein sample such as the spore coat. The second type of linkages id est isopeptide linkages have been identified recently in the exosporium proteins [11]. With the use of mass spectrometry identification and quantification of isopeptide linkages has also been achieved [12, 13].

An MS-based method that can be a candidate to identify the cross-linked peptides from spore coat samples is the use of $^{18}$O-labelled water. This strategy facilitates cross-link analysis by providing peptides with a tag that enables distinction in a mass spectrum of signals belonging to linear and cross-linked proteins. Cross-linked peptides can be distinguished in mass spectra by a characteristic shift of 8 amu compared to the $^{16}$O-labelled ones. Cross-linked peptides contain two carboxy-termini and thus will show a shift of 8 amu when digested in H$_2^{18}$O due to the incorporation of two $^{18}$O atoms in each C-terminus post-digestion. Therefore, normal linear peptides are characterized in a mass spectrum by a 4 amu shift, while cross-linked peptides digested in H$_2^{18}$O will shift 8 amu [14]. Identification of the so-detected cross-linked peptide is then achieved through
interpretation of the MS/MS data of these cross-linked peptides. It needs to be investigated if this isotopic labeling method can be applied to insoluble cross-linked spore coat proteins.

(c) Additional oxidative coupling products of tyrosine residues

In several cases, di-tyrosine has been shown to be present in the structural proteins \cite{15, 16}. The presence of $\text{YXY}$ isodityrosine motifs \cite{17} in the coat proteins as mentioned in Chapter 6 and richness of certain coat proteins in amino residues proline, glycine and tyrosine raises a question if additional oxidative tyrosine coupling takes place. A range of additional oxidative coupling is shown in Figure 2. Isodityrosine, trityrosine, pulcherosine have all been identified to be present in structural protein components in various organisms. Isodityrosine has been observed in the cell wall glycoproteins of tomato \cite{18}, pulcherosine has been identified from plant cell wall proteins \cite{19}, trityrosine from the protein resilin in arthropods \cite{20} and in the fertilization envelops of sea urchin \cite{21}, isotrityrosine from the second-molt cuticle of \textit{Haemonchus contortus} infective larvae of sheep \cite{22}. Although the distribution of amino acids in the spore coat proteins suggests that these cross-links are likely to be present also in and between the spore coat proteins they have not been identified yet. Therefore there is a need to focus on identification of additional oxidative coupling. As mentioned previously, fluorescence

\includegraphics{figure2.png}

\textbf{Figure 2. Additional oxidative coupling products of tyrosine residues.} Figure is adapted from the work of Braddy et al. \cite{19} $\text{Tyr} =$ tyrosine; $\text{Idt} =$ isodityrosine
spectroscopy is a powerful tool to distinguish tyrosine moieties from di-tyrosine. However this method is not suitable to differentiate between the above mentioned oxidation products as all these products fall in the same range of excitation and emission maxima (Table 1). More efforts to develop new methodologies are required in order to distinguish and further characterize these crosslinks.

<table>
<thead>
<tr>
<th>Tyrosine derived oxidative coupling products</th>
<th>Fluorescence Excitation- Emission maxima (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pulcherosine (natural)</td>
<td>pH 2.0: 281 - 420</td>
</tr>
<tr>
<td></td>
<td>pH 12.0: 316 – 420</td>
</tr>
<tr>
<td>Pulcherosine (synthetic)</td>
<td>pH 2.0: 281 - 420</td>
</tr>
<tr>
<td></td>
<td>pH 12.0: 318 – 422</td>
</tr>
<tr>
<td>Di-tyrosine (synthetic)</td>
<td>pH 2.0: 283 - 410</td>
</tr>
<tr>
<td></td>
<td>pH 12.0: 3 15 – 422</td>
</tr>
<tr>
<td>Trityrosine (synthetic)</td>
<td>pH 2.0: 288 - 412</td>
</tr>
<tr>
<td></td>
<td>pH 12.0: 323 – 409</td>
</tr>
<tr>
<td>Isodityrosine</td>
<td>NA</td>
</tr>
<tr>
<td>Isotrityrosine</td>
<td>NA</td>
</tr>
</tbody>
</table>

Data adapted from Nomura et al. [21]. NA = not available.

II. Salt-bridge linkages between the spore coat proteins

The MS analyses of the tryptic digests of the spore coat described in this thesis have shown very frequent occurrence of peptides with a glutamic acid (E) at the C-terminus. These peptides cannot be formed by proteolytic cleavage using trypsin or chymotrypsin. Cleavage of the glutamine acid peptide bond can be realized using Glu-C protease or in strongly acidic media at elevated temperatures and will cleave the protein both before and after glutamic acid. This does not explain the observed specific protein cleavage after glutamic acid.

Glutamic (E) and aspartic (D) acid as well as the basis amino acid lysine (K) appear to be overpopulated in many coat proteins as shown in Chapter 6. This suggests that these residues are involved in the assembly of the spore coat. Maturation of the spore coat is associated with the formation of cross-links in the outer layer and crust as described in Chapter 5. The hypothesis is that during initial maturation the dityrosine and ε-(γ)-glutamyl-lysine link are formed. As the number of cross-links grows, the water is squeezed out between the proteins. In water free environment the acidic residues D and E will transfer a protein to the basic lysine (K) residue resulting in an ionic bond shown in Figure 3. The large number of E, D and K residues in the outer layer proteins may result in many ionic bonds between the proteins which will contribute significantly to the assembly and stability of the spore coat. However, during the processing of the coat for MS analysis the bead beating event induces heat and mechanical stress. This stress may drive a reaction which cleaves the peptide bond after glutamic acid as depicted in Figure 3 (A). The cleavage is activated by the ionic bond which facilitates the negatively charged carboxyl oxygen to undergo a BAc2 reaction with the peptide carbonyl via an energetically favorable 6-member ring transition state. The resulting cleavage of the peptide bond after the glutamic acid residue yield a cyclic anhydride, which under the
bead beating aqueous conditions will be hydrolyzed to C-terminal glutamic acid. This explains the occurrence of many glutamic acid C-terminal peptides after tryptic digestion of the coat proteins. Likewise, aspartic acid (D) can form an ionic bond with the basic lysine residue. Yet, no aspartic acid C-terminal peptides have been identified after tryptic digestion of the coat proteins. This is consistent with the expected higher activation energy for the BAC2 reaction with the peptide carbonyl which for the aspartic acid residue proceeds via an energetically unfavorable 5-membered ring transition state as depicted in Figure 3 (B). Amino acid sequences of coat proteins such as CotC, CotU, CotT suggest a lack of ordered secondary structure. Thus these proteins may function as molecular glue by sealing the gaps in the coat via cross-links and ionic bond formation (Figure 4). These molecular glues may mimic a zipper structure and this zip can be opened once the water enters the spore upon initiation of germination which breaks the ionic bond, leading further to activation of cortex degrading enzymes. The supramolecular-structure based on above mentioned hypothesis is shown in Figure 4.

Figure 3. Cleavage of the peptide bond after (A) glutamic acid and (B) aspartic acid activated by the ionic bond between carboxylic group and amino group of lysine.
Figure 4. Model of the supra-molecular structure of the spore coat. The neighbouring coat proteins (dark blue, green, red) are stabilized by di-tyrosine, glutamyl-lysine and ionic bonds. Small proteins, such as CotC, can act as molecules glues (faint blue) in the coat.

Outlook

The studies presented in this thesis have led to the identification of many putative proteins from the insoluble coat fraction. For the most interesting protein candidates based on criteria such as amino acid sequence, secondary structure, size as well as dependency on the known morphogenetic proteins, their localization and function can be analysed. To achieve this, the use of GFP-tags and point mutations in the protein sequences, respectively, has been proven successful [11, 23].

In Chapter 5 the role has been discussed of a subset of coat proteins that are found to participate in spore maturation by assembly of cross-links. The progress of this cross-linking is shown to be coupled with the increase in wet-heat resistance. It will be interesting to monitor this set of cross-linking coat proteins during spore germination. A recent ICAT-based quantitative proteomics study [24] of B. anthracis has already shown that the protein levels of CotJA, CotJB, CotJC, CoxA etc. are decreased during spore germination. The 15N-metabolic labeling method discussed in Chapter 5, can be used to monitor the digestion efficiency and hence the density of cross-links during germination. For this, a short pulse of germinants can be given to the spores followed by quenching the germination by addition germination inhibitor such as phenyl methyl sulfonyl fluoride (PMSF [24]).

In conclusion, the bio-information on the proteome of spore coats of food spoiling bacteria obtained with developed proteomic methods described in this thesis has provided novel insights in the molecular structure and assembly of spore coats and offers new opportunities for further studies to unravel the mysteries of bacterial sporulation and germination.
References


