Available online at www.sciencedirect.com
ScienceDirect
Journal of Chromatography B, 849 (2007) 129–140
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Review

Towards the entire proteome of the model bacterium *Bacillus subtilis* by gel-based and gel-free approaches☆

Susanne Wolff[a], Haike Antelmann[a], Dirk Albrecht[a], Dörte Becher[a], Jörg Bernhardt[a], Sierd Bron[b], Knut Büttner[a], Jan Maarten van Dijl[c], Christine Eymann[a], Andreas Otto[a], Le Thi Tam[a], Michael Hecker[a,*]

[a] *Institut für Mikrobiologie, Ernst-Moritz-Arndt-Universität Greifswald, Friedrich-Ludwig-Jahn-Str. 15, D-17487 Greifswald, Germany*
[b] *Department of Genetics, Groningen Biomolecular Sciences and Biotechnology Institute, Kerklaan 30, 9751 NN Haren, the Netherlands*
[c] *Department of Medical Microbiology, University Medical Centre Groningen and University of Groningen, Hanzeplein 1, P.O. Box 30001, 9700 RB Groningen, the Netherlands*

Received 2 June 2006; accepted 8 September 2006
Available online 20 October 2006

Abstract

With the emergence of mass spectrometry in protein science and the availability of complete genome sequences, proteomics has gone through a rapid development. The soil bacterium *Bacillus subtilis*, as one of the first DNA sequenced species, represents a model for Gram-positive bacteria and its proteome was extensively studied throughout the years. Having the final goal to elucidate how life really functions, one basic requirement is to know the entirety of cellular proteins. This review presents how far we have got in unraveling the proteome of *B. subtilis*. The application of gel-based and gel-free technologies, the analyses of different subcellular proteome fractions, and the pursuance of various physiological strategies resulted in a coverage of more than one-third of *B. subtilis* theoretical proteome.

Keywords: Proteome coverage; Stress and starvation signatures; Subproteomic fractions; Gel-based and gel-free proteomics

Contents

☆ This paper is part of a special volume entitled "Analytical Tools for Proteomics", guest edited by Erich Heftmann.

* Corresponding author. Tel.: +49 3834864200; fax: +49 3834864202.
E-mail address: hecker@uni-greifswald.de (M. Hecker).

1570-0232/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2006.09.029

1. Introduction

In the year 1995 proteomics [1] was coined to study the diverse properties of proteins combined with the elucidation of their cellular function. With the achievement to decipher entire genome sequences [2–4] protein researchers were supplied with the blueprint of every possible gene product of an organism. One initial goal in proteomics was the identification of all proteins expressed by a cell as the equivalent to the genome. Meanwhile also the exact determination of amino acid sequences, post-translational modifications, interaction partners, protein quantity, synthesis, stability, activity, localization and structure have been focused on, whereby the protein identification still forms the groundwork for such kind of studies. There is no single technology that could be utilized to answer all these versatile problems, but certainly these issues will be addressed profoundly by the still ongoing development of complementary technology platforms.

Whereas the complete sequencing of a genome is a finite undertaking, it is virtually impossible to determine an organisms proteome as a whole due to the significantly higher complexity mainly caused by protein targeting, post-translational modifications as well as mRNA splicing and protein processing. The invention of two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) by O'Farrell [5] and Klose [6] made an extensive separation of hundreds or even thousands of protein species possible and therefore brought the goal of a preferably complete proteome dissection closer. Due to the development of mild ionization methods [7,8] mass spectrometry was entering protein science. Combined with the introduction of database search algorithms [9–13], proteins separated on 2D-gels could now be identified in a significantly higher throughput compared with the hitherto performed N-terminal sequencing via Edman degradation [14].

Although 2D-PAGE proved to be a powerful separation technique, it has its downsides in resolving proteins with extremes in isoelectric point (p*I*) or molecular mass, and most unfavorable is certainly the poor resolution of hydrophobic proteins such as those spanning the cellular membrane. To gain access also to this gel-critical fraction of the proteome alternative techniques arose based on a gel-free sample separation at peptide rather than at protein level [15–17]. This approach became famous as shotgun proteomics and allowed the unbiased identification of hundreds and even thousands of proteins from highly complex protein mixtures [18,19]. Thus, a remarkable fraction of the proteome for several "low complexity" organisms could already be covered [20–24].

One of these model organisms is the well-studied soil bacterium *Bacillus subtilis*. Already by the mid-20th century some species of the genus *Bacillus* were known more generally for their roles in human and animal infections and others were valued as producers of antibiotics and industrial enzymes. Having become the paradigm for Gram-positive bacteria, *B. subtilis* is an ideal experimental system for studying mechanisms of gene regulation, metabolism and cell differentiation also in view of investigations of highly related microorganisms, such as the human pathogen *Staphylococcus aureus*. Because of the far-reaching history of research on *B. subtilis*, it was among the first species being fully DNA sequenced [25]. With the whole genome sequence available not only mRNA profiling of gene expression could be performed, but also the gate towards high-throughput protein identification was pushed open. Though we can look back on 20 years of proteome research in *B. subtilis* [26,27], still one-third of *B. subtilis* 4100 genes are not assigned to a defined function yet [28], and only for about one-third of all genes the corresponding protein could be demonstrated so far [24]. This review gives an overview on the strategies we explored with the objective to increase proteome coverage of *B. subtilis*. The results clearly evince the necessity to employ different technology platforms, but also tell us that it is still a long way to the implementation of complete proteome coverage, which will take us closer to the intrinsic goal of understanding how a single cell actually functions.

2. Cytosolic proteome map for growing and non-growing cells of *B. subtilis*

Two-dimensional gel electrophoresis with its high-resolution power is an obvious method to detect as many proteins as possible and, therefore, provides the basis for comprehensive physiological proteome studies. The technique was used successfully for a far-reaching proteome coverage in several bacteria, e.g. *Haemophilus influenzae* [29]. The introduction of proteomic approaches based on liquid chromatography coupled to mass spectrometry gave an important stimulus in the further increase of proteome coverage. Lipton et al. published the identification of 61% of the theoretical proteome of *Deinococcus radiodurans* [21]. With 1480 proteins identified, 35% of the *Escherichia coli* proteome could be covered [23], and in the "minimal" organism *Mycoplasma mobile* even 88% of the 635 open reading frames (ORFs) have been detected on protein level [30]. Combining 2D-PAGE to study cytosolic proteins and a gel-free approach for an analysis of membrane proteins Oesterhelt and coworkers were able to identify 34% of the theoretical proteome of the halophilic archaeon *Halobacterium salinarum* [31,32].

However, the depiction of an entire proteome does not solely depend on the analytical tools and their continuous improvement. A considerable part of the genome remains more or less silent under standard growth conditions and will be expressed only in response to specific stimuli. Therefore, two major classes of proteomes have to be defined: the proteome of growing cells (vegetative proteome) and the proteome of non-growing cells suffering from stress or starvation.

2.1. Cytosolic proteome map of growing cells

Macroarray analyses revealed that approximately 2500 of *B. subtilis* 4107 genes are actively transcribed under exponential growth conditions [33]. A first comprehensive map of the vegetative *B. subtilis* proteome was published by Büttner et al. [34], who were able to identify more than 300 cytosolic proteins by standard 2D-PAGE located in an analytical window of p*I* 4–7. Three years later Eymann et al. [33] expanded the cytosolic

proteome map of growing cells to a total number of 693 proteins in the region p*I* 4–7. This substantial increase in protein identifications can be ascribed to refined experimental protocols, to further sophistication in mass spectrometric technology and to the utilization of several ultra zoom gels (e.g. p*I* 4.5–5.5) which improved the resolution of protein separation in the first dimension on a 2D-gel. Proteins having a p*I* beyond 7, which theoretically applies to one-third of the *B. subtilis* proteome, are not covered by conventional 2D-PAGE (p*I* 4–7). Consequently, protocols have been worked out to enable the electrophoretic separation of alkaline proteins [35,36]. Ohlmeier et al. [37] used immobilized pH gradients (IPGs) covering a pH range from 4 to 12 to establish an alkaline master gel for *B. subtilis*. A remarkable fraction of the alkaline proteome is made up of ribosomal and highly hydrophobic proteins such as membrane proteins. From the alkaline 2D-gels 41 proteins with p*I* > 7 could be identified, most of which were ribosomal proteins. The goal to embed membrane proteins on alkaline 2D-gels failed. Their extreme hydrophobicity causes precipitation in IPG strips, which prevents their separation in the second dimension. Therefore, more attention has to be directed to membrane proteins since they seem to necessitate specific preparation and separation techniques. Additional 19 alkaline proteins were identified by Eymann et al. [33], who continued the analysis of the alkaline protein fraction on 2D-gels covering the pH range from 6 to 11.

At the time Ohlmeier et al. [37] started their advent of an alkaline mastergel there were basically no alternatives to 2D-PAGE for a separation of such complex protein samples. Only after the emergence of gel-free and direct measurement techniques it became possible to fractionate complex samples of proteins with extreme physico-chemical properties, because such techniques allow separation steps at peptide instead of protein level [18,19].

A gel-free approach, in which the cytosolic proteome fraction of growing *B. subtilis* cells was dissected via two-dimensional liquid chromatography, yielded a total of 814 identified proteins of which 140 possessed a p*I* beyond 7 [24]. Surprisingly 129 of the 814 proteins are characterized by a p*I* between 4 and 7 and have no transmembrane domains but still they were not detected via standard 2D-PAGE. The incapability to detect these proteins on the gel could be caused by their low abundance, as 10 of the 129 proteins are transcriptional regulators that are most likely difficult to visualize by current staining methods. There are also proteins smaller than 10 kDa or larger than 100 kDa which will probably bottom out the gel or will not migrate into the 2D-gel, respectively. Despite the high-resolution power of 2D-PAGE one cannot rule out that some proteins could not be identified from the gel due to protein comigration, which results in masking of small protein spots by higher abundant proteins.

Combining the results from 2D-PAGE analyses with those obtained by a peptide fractionation via 2D-LC followed by MS/MS (tandem mass spectrometry) acquisition, it was possible to identify 1014 cytosolic proteins from exponentially growing cells. Most of the proteins of the central carbohydrate metabolism (glycolysis, pentose phosphate shunt and citric acid cycle), of almost all amino acid synthesis pathways, of purine and pyrimidine metabolism, of fatty acid metabolism, and of the main cellular functions like replication, transcription, translation and cell wall synthesis have been detected in 2D-PAGE studies already (Figs. 1 and 3). However, gel-free approaches allowed the identification of additional proteins, as alkaline, hydrophobic and low-abundant ones (Fig. 2). Thus, the foundation of a comprehensive monitoring of *B. subtilis* cellular physiology has been layed (Fig. 3).

2.2. Cytosolic proteome map of non-growing cells

The continuous requirement for adaptation of bacteria to physical stress and starvation in natural ecosystems has forced the development of a very complex adaptational network of gene regulation. In contrast to growing cells, which are characterized by a stable "vegetative core proteome" the proteome of non-growing cells is additionally determined by a specific stress or starvation condition which is reflected as "proteome signature" [38–41]. Tam et al. [42] complemented the vegetative proteome map of *B. subtilis* by a proteome map in response to stress and starvation using the 2D-gel-based approach. For that purpose the synthesized proteins of cytoplasmic proteomes of [^{35}S]-methionine labeled *B. subtilis* cells were analyzed in response to heat, salt, hydrogen peroxide and paraquat stress as well as after ammonium, tryptophan (*B. subtilis* carries an auxotrophy for tryptophan), glucose and phosphate starvation. The 2D-gel images of all stress and starvation experiments were combined by an image fusion approach to a stress (Fig. 4a and b) and starvation (Fig. 5a and b) proteome map, respectively [43]. In those maps all marker proteins induced specifically by one stimulus or generally by multiple stimuli were labeled with a defined color code of 15 different colors indicating their stress/starvation respondent induction profile. Such 2D-proteome maps with their color codes according to protein subsets with similar regulation have been used to classify stress and starvation proteins into specific and general regulons [42]. For example, the heat shock signature is characterized by an induction of heat-specific HrcA-dependent chaperones and the general induction of the SigB, CtsR and Spx regulons [44]. The treatment with hydrogen peroxide and paraquat resulted in the induction of the oxidative stress-specific PerR regulon, the iron starvation Fur regulon and the Spx regulon as general indicator for protein damage by non-native disulfide bond formation [45–48]. In contrast, the SOS regulon induction is specific for DNA damage caused by peroxides, only [49]. The specific adaptative function of stress specific regulons (e.g. HrcA, PerR) is to accomplish resistance against the stressor by neutralizing it, by adaptation to the specific stressor or by repair of the damage it caused [41,50]. In contrast, the SigB-dependent general stress regulon encodes proteins with different functions conferring a multiple, non-specific and preventive stress resistance to non-growing *B. subtilis* cells in anticipation of future stress possibly encountered during long-term stationary growth stages [41,51].

The proteome map summarizing the protein patterns of ammonium, tryptophan, glucose and phosphate starvation provided the tool to define specifically induced starvation regulons and more generally induced transition phase regulons. Starvation-specific proteins are required for high-affinity uptake of the limiting substrate or for the utilization of alternative

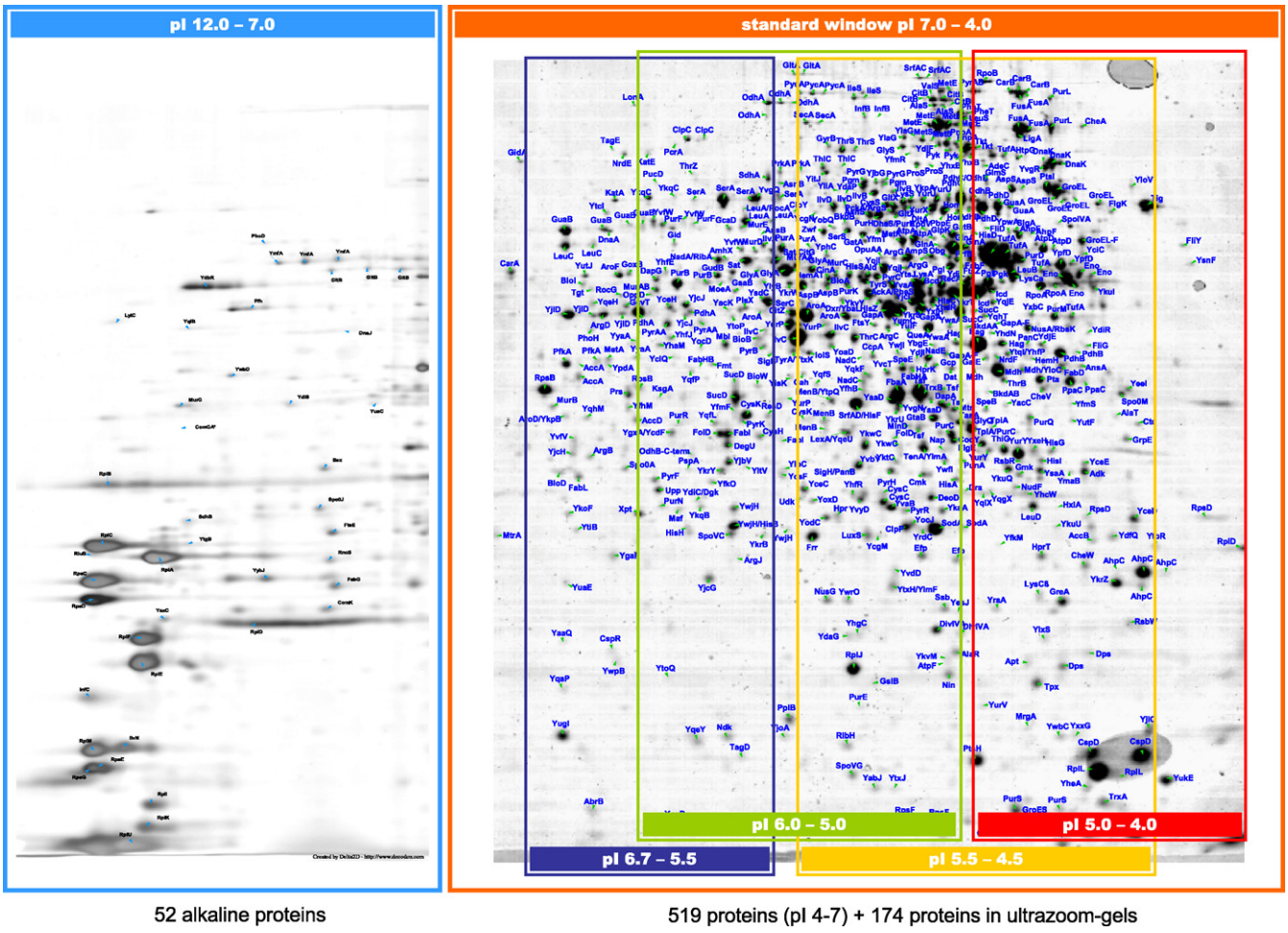

Fig. 1. The cytoplasmic vegetative proteome map of *B. subtilis* in the standard pH range (4–7), in the narrow pH ranges (4.0–5.0, 4.5–5.5, 5.0–6.0, 5.5–6.7) and in the alkaline pH range (7–12). Cytoplasmic protein extracts were separated by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and stained with Colloidal Coomassie Blue [33]. Spot identification was performed using matrix assisted laser desorption ionization tandem time of flight (MALDI-TOF-TOF) mass spectrometry.

substrates when the preferred one is limited or exhausted. The induction of the TnrA regulon involved in uptake and utilization of alternative nitrogen sources and the SigL/BkdR regulon for degradation of branched chain amino acids indicates an ammonium starvation-specific proteome signature [52–55]. In contrast, an induction of the TRAP-regulated tryptophan biosynthesis enzymes is the specific proteome signature for tryptophan starvation [56,57]. In response to glucose starvation several carbon catabolite-controlled marker proteins are specifically activated in the absence of glucose and repressed in the presence of glucose by CcpA, CcpN and AcoR [58–62]. Finally, the induction of the PhoPR regulon is the specific proteome signature for phosphate starvation [63–66]. Starvation regulons such as TnrA, TRAP, CcpA or PhoPR were shown to be specifically involved in the uptake and utilization of alternative nutrient sources in response to a particular nutrient limitation. The proteome signature for different kinds of nutrient limitation is mediated by the CodY, SigB and SigH transition phase regulon. These general starvation regulons that are induced in response to carbon, ammonium or phosphate starvation are required for the adaptation of the cell to post-exponential stationary phase processes such as survival under non-growing conditions, competence

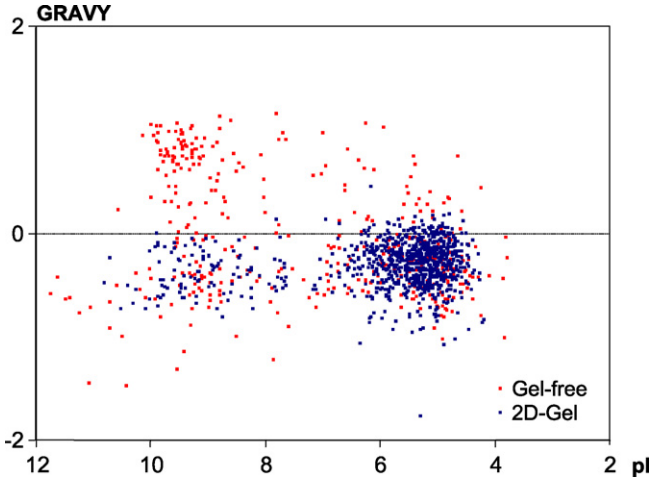

Fig. 2. Comparison of the subsets of proteins identified by a two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) approach and gel-free methods [24]. The plot GRAVY (grand average of hydropathy), which is a value for the hydrophobicity of proteins, versa isoelectric point clearly indicates the power of 2D-gels to separate proteins in an analytical window of p*I* 4–7. The gel-free approach could significantly increase the number of alkaline proteins and hydrophobic proteins which possess a more positive GRAVY value.

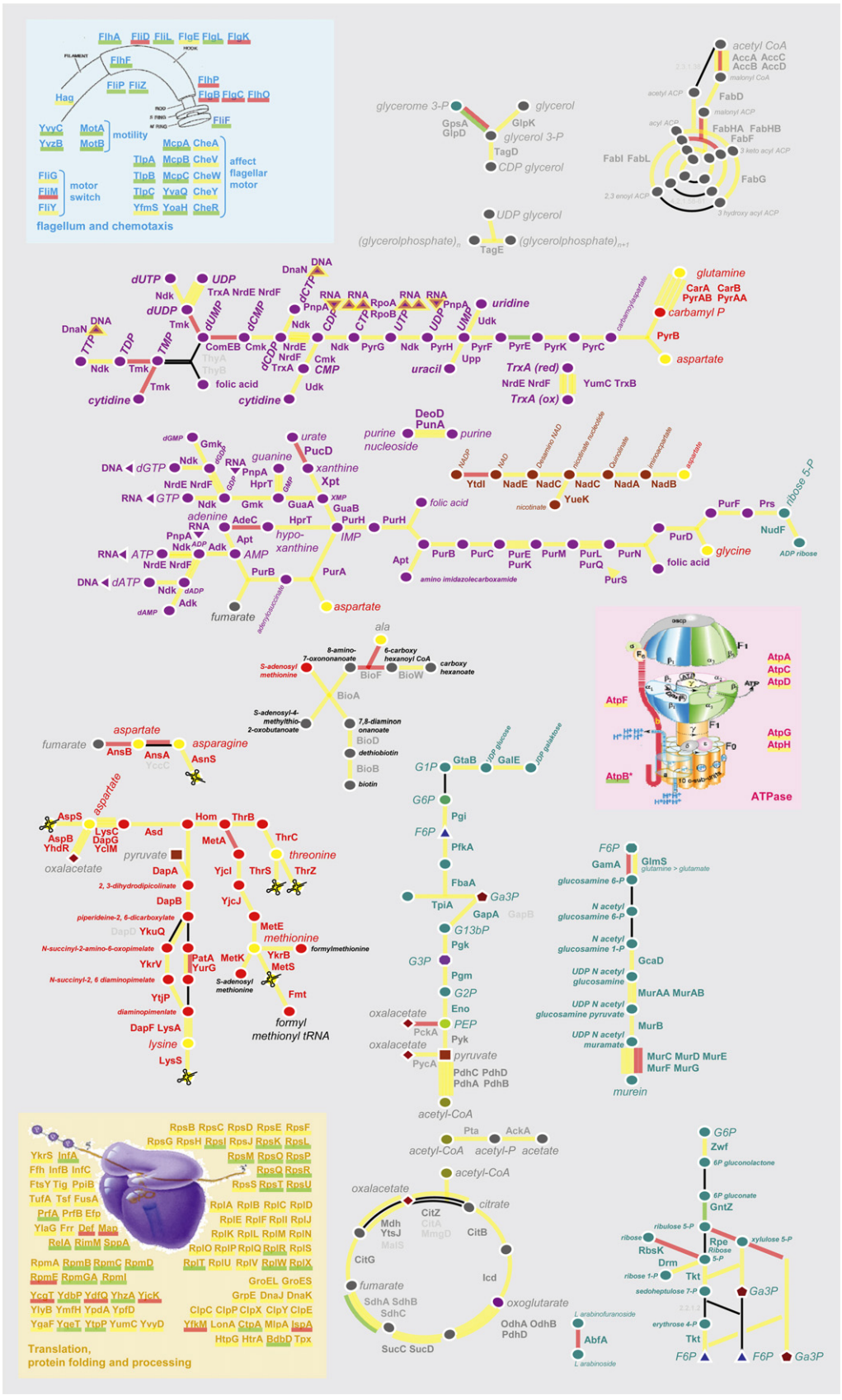

Fig. 3. Assignment of proteins identified to the different branches of cellular metabolism. Proteins that have not been identified in the 2D-gel images thus far are shaded light grey. Components of the carbohydrate metabolism such as glycolysis (center of the left page), pentose phosphate shunt (right side of the left page) and aminosugar synthesis for murein synthesis (right side of the left page) are indicated in dark cyan. Citric acid cycle (lower center of the left page), biotin (center of the left page) and fatty acid metabolism (upper right corner of the left page) are indicated by dark grey. Amino acid metabolism is colored red.

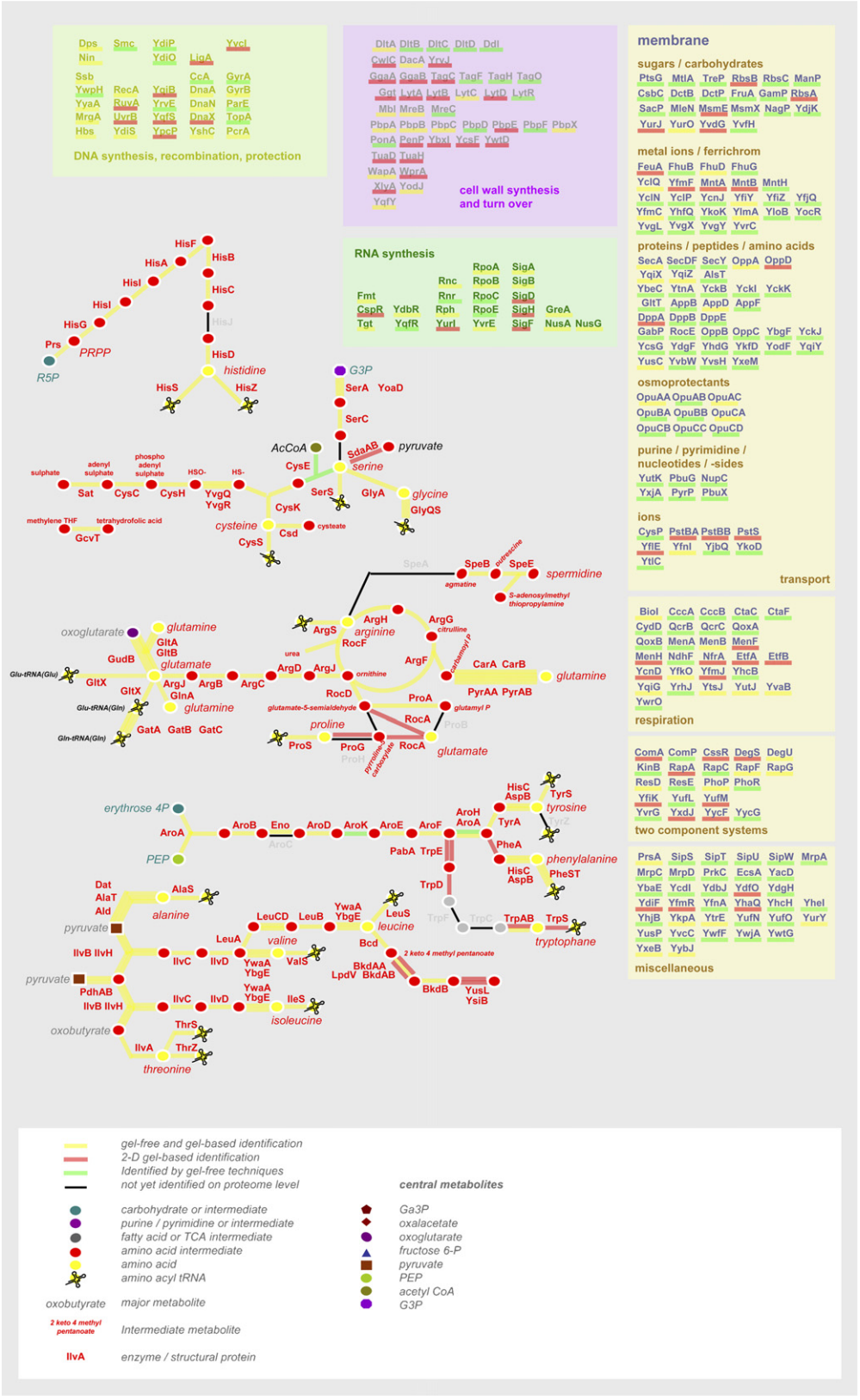

Fig. 3. (*Continued*) The essential amino acids are highlighted in yellow. Purin and pyrimidine (upper left page) metabolism are encoded in purple and the nicotinate metabolism (upper left page) in brown. Components not directly involved in metabolic pathways but in essential cell structures are presented in boxes. DNA-related functions in yellowish green (upper left corner right page), flagellum and chemotaxis-related components in azure (upper left corner left page) and ATPase components in pink (right side left page). Membrane proteins are highlighted cream-colored, components of the transcriptional machinery in green, and the ribosome and other components of the translational apparatus are encoded in ochre.

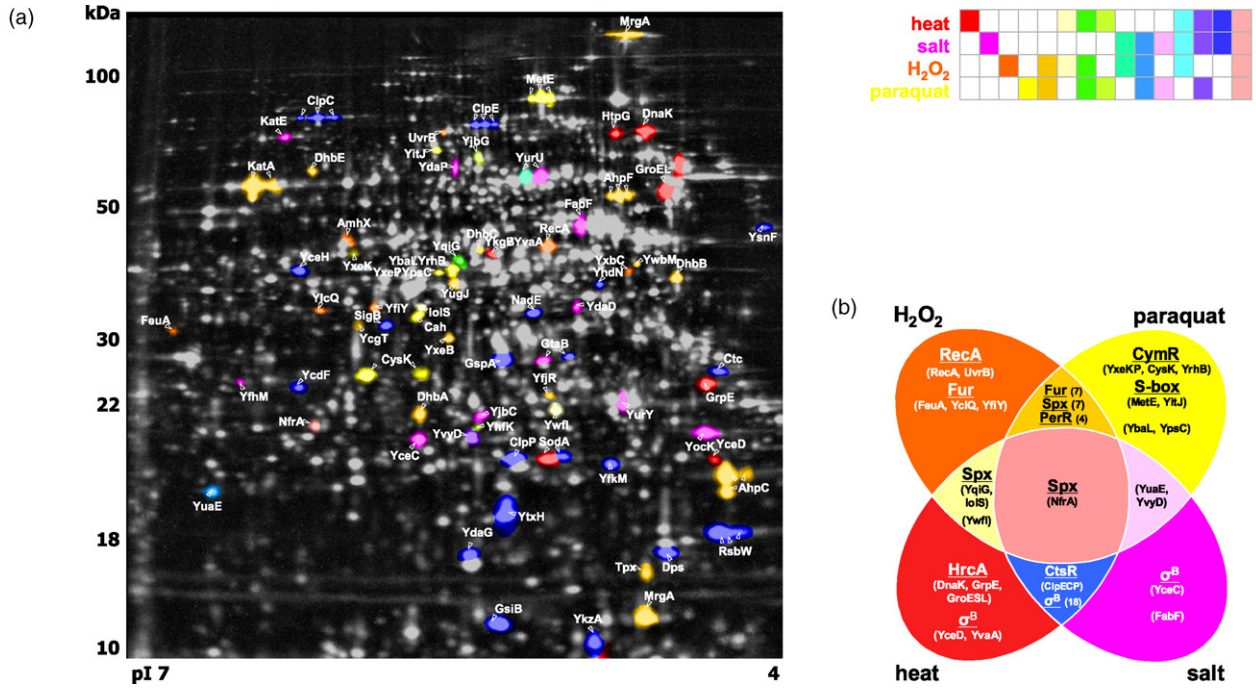

Fig. 4. (a) Cytoplasmic proteome map of *B. subtilis*. The realistically looking 2D-pattern is a composite image from five positionally corrected 2D-autoradiographs derived from exponentially grown as well as heat, salt, hydrogen peroxide and paraquat stressed *B. subtilis* cells. These autoradiographs showed the protein synthesis patterns from 5 min of cultivation incorporated [^{35}S]-methionine. After applying the union fusion algorithm of the Delta2D software ([42], DECODON) to this stack of autoradiograms the resulting composite image summarizes the spot information from the whole gel series including all stress proteins ever induced. This made it possible to detect a spot consensus containing all stress proteins and valid for the whole gel set. This spot consensus was transferred and adapted to any original autoradiograph. The quantitation was done according to the consensus' spot shapes. For color coding, each spot set on any autoradiograph was divided into a non-changed and a more than three-fold compared to the control induced fraction. The latter fraction was considered as a group of stress proteins. This binarization of any spot set made it possible to highlight all induced proteins according to the shown color legend. The labels indicate which protein had been identified behind the spots. (b) The specific and general stress regulons in *B. subtilis*. Commonly shared (generally induced) and unique (specifically induced) stress regulons and proteins in the proteome of *B. subtilis* exposed to heat, salt, hydrogen peroxide and paraquat stress according to the fused stress proteome map in (a) [42]. The specific and general stress regulons are underlined and the encoded specific or general stress proteins are listed in parentheses.

or sporulation. The CodY regulon encodes proteins that allow broader adaptation to nutrient depletion including extracellular degradative enzymes, transporter proteins, catabolic enzymes, factors involved in genetic competence, antibiotic synthesis pathways, chemotaxis proteins and sporulation proteins [67]. In addition, the SigB regulon is the general proteome signature for glucose and phosphate starvation and required for the stationary phase survival upon starvation [41,61]. SigH directs the transcription of several genes that function in the transition from exponential growth to stationary phase, including the initiation of spore formation, genetic competence and general adaptation to nutrient depletion [68]. The color coded fused proteome map for starvation categorized a subset of six SigH-dependent proteins (YvyD, YtxH, YisK, YpiB, Spo0A, YuxI and the CodY-dependent RapA) as general starvation proteins indicating the transition to stationary phase caused by nutrient limitation [42]. Thus, the fused stress/starvation proteome maps and the color code approach provide important leads for future research on protein function of the novel identified general stress/starvation proteins during stationary phase survival in *B. subtilis*.

In total more than 200 stress or starvation induced proteins have been identified [42,45,58,64,69]. Eighty-three of them are absent in the vegetative proteome map, whereas the remaining stress and starvation induced proteins exhibit a basal expression level already in vegetative growing cells, indicating a basic physiological role in growing and non-growing cells.

3. The extracytoplasmic proteome of *B. subtilis* dissected by 2D-PAGE

As a Gram-positive bacterium *B. subtilis* lacks an outer membrane. It is therefore able to secrete large amounts of extracellular proteins directly into the environment. These proteins perform several very important functions, such as tapping of nutrients, environmental detoxification, cell-to-cell communication, combating competitors and, with regard to pathogenic bacteria, extracellular proteins play a critical role in virulence [70]. According to previous studies there are at least four distinct secretion pathways for approximately 300 proteins predicted to be located extracellularly [71,72]. Most proteins are secreted via the Sec pathway whereas the twin arginine translocation pathway, the pseudolipin export pathway and pathways using ATP-binding cassette transporters are employed for special purposes for only a few specific proteins [73]. The way a protein is translocated and its final destination are determined by the presence of particular signal peptides and retention signals. The first proteome analyses on secreted proteins in *B. subtilis* were performed in minimal medium with different carbon sources [74] as

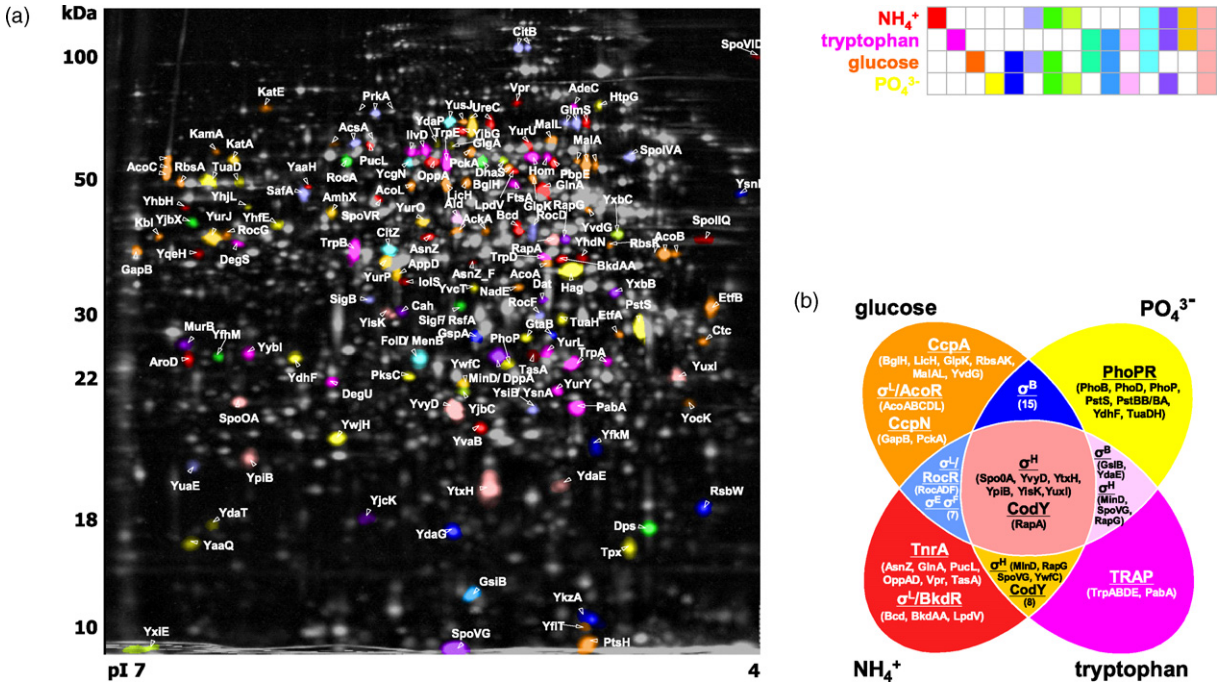

Fig. 5. (a) Cytoplasmic proteome map of *B. subtilis* after ammonium, tryptophan, glucose and phosphate starvation. The protein synthesis pattern (autoradiograms) of *B. subtilis* in response to starvation for ammonium, tryptophan, glucose and phosphate were combined to generate a fused starvation proteome map of *B. subtilis* using the union image fusion approach of the Delta2D software [42]. The induced marker proteins were color coded according to their expression profiles. All spots induced specifically after one starvation condition or generally also by other starvation conditions more than 3-fold are labeled in the fused proteome map by a defined color code that indicates the induction profile. (b) The specific and general starvation regulons in *B. subtilis.* Commonly shared (generally induced) and unique (specifically induced) starvation regulons and proteins in the proteome of *B. subtilis* after ammonium, tryptophan, phosphate and glucose starvation according to the fused starvation proteome map in (a) [42]. The specific and general starvation regulons are underlined and the encoded specific or general starvation proteins are listed in parentheses.

well as under phosphate starvation conditions [64]. However, the highest level of protein secretion in *B. subtilis* is observed when the cells have entered stationary growth phase in rich medium [75]. A 2D-PAGE approach was used to study the *B. subtilis* extracytoplasmic proteome which can be divided into three subproteomes: the true extracellular proteome, the lipoproteome and the cell wall associated proteome.

3.1. The extracellular proteome of B. subtilis

Most of the proteins that are completely transported across the cytoplasmic membrane are synthesized with an N-terminal signal peptide that is cleaved by a signal peptidase I (SPase I) after translocation which allows the protein to be released into the growth medium.

When *B. subtilis* is grown under conditions of phosphate starvation PhoPR-dependent proteins comprising extracellular phosphatases and phosphodiesterases (PhoA, PhoB, PhoD), the glycerophosphoryl phosphodiesterase GlpQ, the 5′nucleotidase YfkN, the ribonuclease YurI, the binding component of the high-affinity phosphate-specific transporter PstS and the lipoprotein YdhF are exceptionally strong induced [64,66,76,77]. These phosphate starvation-specific proteins remarkably account for 30% of the total phosphate starvation extracellular proteome (Fig. 6a). A master gel for the secreted proteome of *B. subtilis* was defined in Luria Broth (LB) medium during the stationary phase (Fig. 6b). In total, 113 different proteins of the *B. subtilis* secretome could be identified. Fifty-four of them were predicted to be secreted because of an N-terminal signal peptide with a SPase I cleavage site and the lack of retention signals [71,73]. The remaining proteins are unpredicted secretory proteins which either possess signal peptides and retention signals (18 lipoproteins, 6 cell wall-binding proteins and 6 membrane proteins) or lack signal peptides (17 cytoplasmic proteins, 6 phage-related proteins and 7 Flagella-related proteins). The identified extracellular proteins are involved in the uptake, transport or utilization of carbohydrates, proteins, nucleotides, lipids and phosphate, in the cell wall metabolism, in environmental detoxification or they show flagella- and phage-related functions [73,75].

3.2. The “lipoproteome” of B. subtilis

The lack of an outer membrane has, most likely, forced Gram-positive bacteria to modify several extracellular proteins with lipids to adhere them as “lipoproteins” at the membrane surface, as the high-affinity substrate-binding proteins of ABC-transporters. Lipoproteins in *B. subtilis* are Sec-dependently secreted, lipid modified by the diacyl glyceryl transferase Lgt and cleaved by the type II signal peptidase LspA [78]. To describe the “lipoproteome” fraction of *B. subtilis*, a washed cell membrane fraction was prepared that was insoluble in the non-detergent sulfobetaine (NDSB) and then extracted with

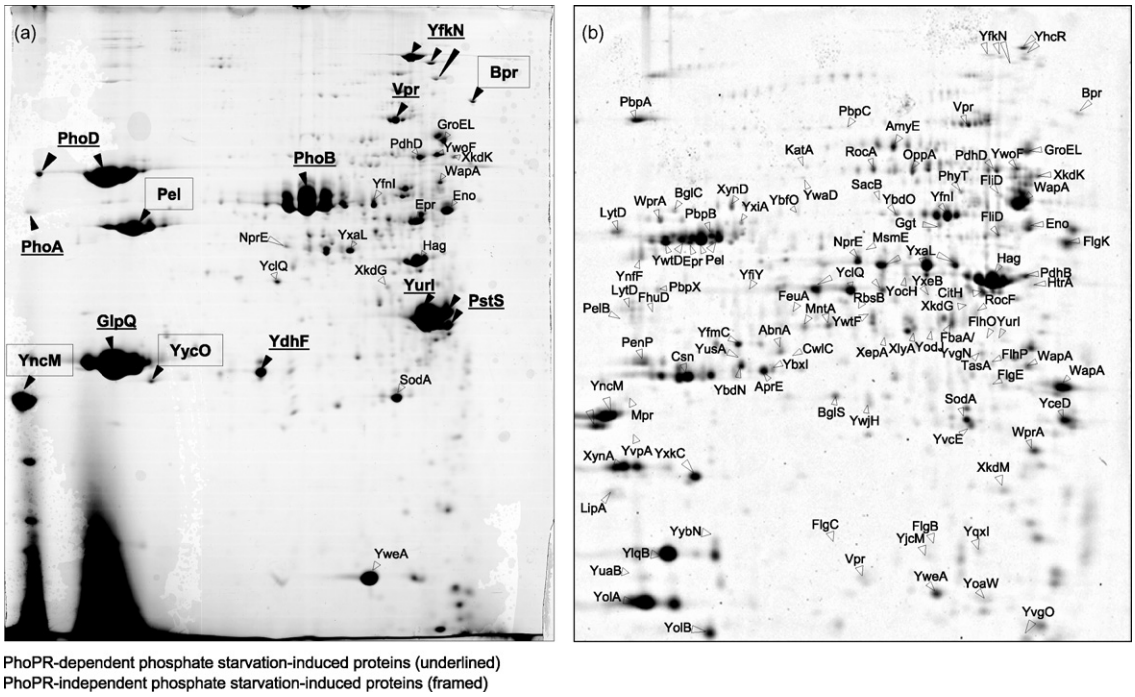

Fig. 6. The extracellular proteome of *B. subtilis* 168 under conditions of phosphate starvation (a) and in complete medium (b). Cells of *B. subtilis* 168 were grown in minimal medium under the conditions of phosphate starvation (a) and in Luria Broth (b). Proteins in the growth medium were harvested 1 h after entry into the stationary phase. After precipitation with trichloric acid (TCA), the extracellular proteins were separated by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and stained with Sypro Ruby [63,74].

the detergent *n*-dodecyl-β-D-maltoside as described previously [79]. A separation of this lipoprotein containing fraction by 2D-PAGE yielded 10 lipoproteins which have not been identified in the cytoplasmic proteome [76]. Interestingly, nine of these lipoproteins are also components of the extracellular proteome. They are most likely released into the growth medium by proteolytic shaving [80].

3.3. The cell wall proteome of *B. subtilis*

Several enzymes involved in cell wall turnover contain a variable number of repeat domains [71] which show an affinity for components of the cell wall [81]. These repeats are thought to facilitate the direction of enzymes to distinct sites in the cell wall. In Gram-positive organisms also a covalent anchoring of surface proteins to the cell wall is known [82]. It requires a C-terminal cell wall sorting signal that will be recognized by a specific transpeptidase, the sortase A, which covalently binds the protein to the cell wall [83]. The LiCl-extracted cell wall proteome of *B. subtilis* consists of 12 abundant non-covalently linked cell wall proteins including the WapA processing products (CWBP105 and CWBP62), the processing products of the major cell wall protease WprA (CWBP52 and CWBP23) and the autolysins LytC and LytB [84]. Interestingly, some cell wall proteins have been identified exclusively in the extracellular proteome (e.g. LytD, an autolysin), but not in the cell wall fraction. This indicates that the presence of cell wall binding repeats is not a guarantee for retention in the cell wall. Cell wall proteins might be released into the medium due to proteolytic shedding and cell wall turnover [73].

In total the comprehensive analyses of the *B. subtilis* extracytoplasmic proteome by a 2D-gel-based approach lead to the identification of 119 additional proteins which are either secreted, lipid anchored or associated to the cell wall.

4. The membrane proteome of vegetative cells

Membrane proteins constitute an important facet in physiological proteomics due to their importance in maintaining cell integrity, signal sensing, transport processes, energy conservation and their role in virulence regarding pathogenic bacteria. Despite their functional relevance the study of membrane proteins is notoriously difficult [85] and thus lagging behind analyses on other subproteomic fractions as cytosolic [86], cell wall-associated [76,84] and secreted [73,75,76] proteins. Similar to the extracellular proteins membrane proteins possess signal peptides marking them for cellular export. However, in most cases these signal peptides are not removed by a SPase and, consequently, they serve in the anchoring of proteins in the membrane. Furthermore, membrane proteins may contain additional membrane-spanning domains, which determine the topology of these proteins in the cytoplasmic membrane. According to the TMHMM algorithm to predict transmembrane domains [87] more than one quarter of all *B. subtilis* proteins is characterized by one or more membrane-spanning domains, emphasizing the need to focus on this barely characterized proteome fraction.

With a profiling of ABC-transporter solute-binding proteins, Bunai et al. initiated the work on the *B. subtilis* membrane protein fraction [79]. Dreisbach and coworkers established a technique

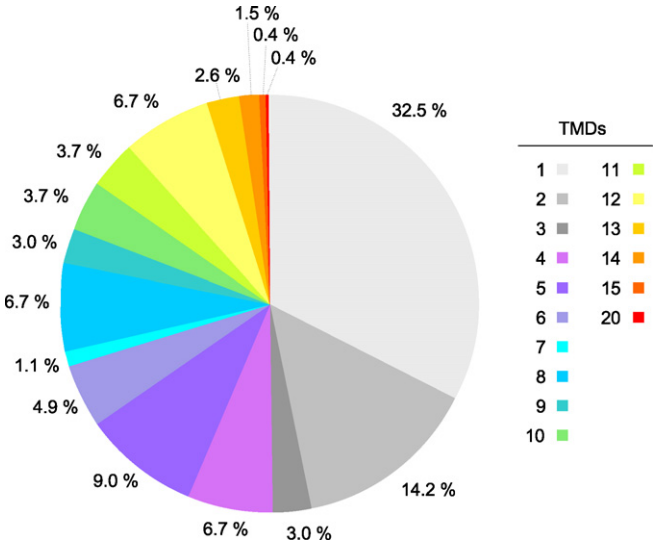

Fig. 7. Depiction of the number of transmembrane domains in the 268 membrane proteins identified by semi-gel-based proteomics.

in which the cytoplasmic membrane is washed in high-salt and alkaline buffers before the membrane proteins are solubilized with the detergent *n*-dodecyl-β-D-maltoside [33]. The membrane proteins are subsequently separated by one-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis (1D-SDS-PAGE), enzymatically digested and their peptides resolved via reverse phase chromatography followed by MS/MS analysis. With a result of 268 identified proteins which contain membrane-spanning domains this technique was successfully applied to start the dissection of the *B. subtilis* membrane proteome of exponentially growing cells [24,33]. Remarkably, 134 proteins possess four or more transmembrane domains (TMDs), which is consistent with their strong hydrophobicity and explains why this semi-gel-based approach is needed for their detection. Fig. 7 depicts the numbers of membrane-spanning domains in the 268 identified membrane proteins. Most of the detected membrane proteins are still of unknown function. The functionally defined proteins can be assigned to several essential categories, such as permeases and transporters, components of respiratory chains and ATP-synthase, redoxactive enzymes (e.g. oxidoreductases and dehydrogenases), two component systems, proteins involved in cell motility, cell division, autolysis, chemotaxis, osmoregulation, penicillin-binding proteins, parts of translocation systems and signal peptidases, as well as extracellularly acting proteases. Of the 268 membrane proteins, 16 are predicted lipoproteins according to Tjalsma et al. [71], 11 of those have not been detected in 2D-gel-based studies.

5. Conclusion

The genome of *B. subtilis* contains more than 4100 genes. In our proteome study 1395 proteins were identified so far. Most of them are vegetative proteins which are synthesized in growing cells to put on house-keeping functions (950 cytosolic proteins, 268 membrane proteins, 12 cell wall-bound proteins). Some vegetative proteins are characterized by a basal expression level indicating basic cell functions even in growing cells, but they are strongly induced in response to stress stimuli, like chaperones as prominent members of this class. In response to nutrient starvation 155 proteins were determined to be induced and can be used as proteomic signatures for carbon, nitrogen or phosphate starvation. Finally, our work resulted in the detection of 113 extracellular and 21 lipid-anchored proteins, mostly expressed in non-growing cells. The total proteome information we obtained covers more than one-third of *B. subtilis* theoretical proteome, now ready for physiological application. Most proteins of the main metabolic pathways, predominantly located in the proteomic window of p*I* 4–7, were covered by our proteomic approaches. Many enzymes, already known for a long time, have been visualized for the first time. What can we do with this new kind of information? First of all we have the chance now to follow the regulation of entire metabolic pathways on a proteome-wide scale, as it was already done for the carbon core metabolism [42,58,60,61].

Nevertheless, there is a long way to go to the entire proteome of *B. subtilis* with more than 4100 predicted gene products. To reach this final goal we have two major approaches in mind, first of all an analytical approach: relying on gel-free techniques we were able to add many more proteins to the "gel-based proteome map", particularly to the membrane proteome fraction which still escapes detection by gel-based proteomics. This is, however, only a beginning. By a consequent extension of the application of gel-free techniques we are convinced to increase the proteome coverage in a substantial way already in the near future.

The second approach follows a physiological consideration: a substantial part of the genome is expressed only under definite physiological circumstances. Concurrent transcriptome and proteome studies revealed that even in cells that were exposed to various stress or starvation stimuli only a subfraction of the induced genes, in most cases between 30 and 60%, were covered by proteomics [33]. The still missing genes code for membrane proteins, low abundant proteins and many others.

Furthermore, we almost ignored the proteome of sporulating cells which may consist of more than 300 or even 400 proteins. And finally, one has to bear in mind that we exposed the cells to laboratory stress and starvation stimuli which can only simulate growth-restricting conditions in nature. The list of these "natural stimuli" is far off from being complete. We have only limited information on the proteome of cells grown under nitrate respiration or fermentation conditions, and we do not know much about the proteome of cells under cell wall stress, alkaline or acidic stress, to mention some of them. Furthermore, many stimuli cells typically have to face in their natural habitat were ignored. Cell growth in nature is characterized by a characteristic "multicellular behaviour" including the formation of microcolonies and biofilms, for example. The growth of *B. subtilis* adjacent to plant roots in the soil is probably very important regarding its natural life style. It is an attractive and challenging task for future studies to analyze the proteome of cells grown under environmentally related conditions in order to increase the coverage of the proteome on the one side and to gain new and essential information on its ecophysiology on the other.

Acknowledgements

We are grateful to Annette Dreisbach and Prof. Uwe Völker for the good collaboration in the purification of membrane proteins of *B. subtilis*. We thank DECODON GmbH for supporting us with the Delta2D 3.4 gel analysis software. This work was supported by grants from the Deutsche Forschungsgemeinschaft (HE 1887/7-4), the European Union (QLK3-CT-1999-00413 and LSHG-CT-2004-0005257), the Bundesministerium für Bildung und Forschung (03ZIK011 and 0313313A) and the Bildungsministerium of the country Mecklenburg-Vorpommern (EMAU 0202120).

References

[1] M.R. Wilkins, J.C. Sanchez, A.A. Gooley, R.D. Appel, I. Humphery-Smith, D.F. Hochstrasser, K.L. Williams, Biotechnol. Genet. Eng. Rev. 13 (1996) 19.
[2] R.D. Fleischmann, M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage, C.J. Bult, J.F. Tomb, B.A. Dougherty, J.M. Merrick, et al., Science 269 (1995) 496.
[3] A. Goffeau, R. Aert, M.L. Agostini-Carbone, A. Ahmed, M. Aigle, L. Alberghina, K. Albermann, M. Albers, M. Aldea, D. Alexandraki, et al., Nature 387 (1997) 5.
[4] J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans, R.A. Holt, et al., Science 291 (2001) 1304.
[5] P.H. O'Farrell, J. Biol. Chem. 250 (1975) 4007.
[6] J. Klose, Humangenetik 26 (1975) 231.
[7] M. Karas, F. Hillenkamp, Anal. Chem. 60 (1988) 2299.
[8] J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, C.M. Whitehouse, Science 246 (1989) 64.
[9] W.J. Henzel, T.M. Billeci, J.T. Stults, S.C. Wong, C. Grimley, C. Watanabe, Proc. Natl. Acad. Sci. U.S.A. 90 (1993) 5011.
[10] M. Mann, P. Hojrup, P. Roepstorff, Biol. Mass. Spectrom. 22 (1993) 338.
[11] D.J. Pappin, P. Hojrup, A.J. Bleasby, Curr. Biol. 3 (1993) 327.
[12] P. James, M. Quadroni, E. Carafoli, G. Gonnet, Biochem. Biophys. Res. Commun. 195 (1993) 58.
[13] J.R. Yates 3rd, S. Speicher, P.R. Griffin, T. Hunkapiller, Anal. Biochem. 214 (1993) 397.
[14] K. Heirwegh, P. Edman, Biochim. Biophys. Acta 24 (1957) 219.
[15] D.F. Hunt, R.A. Henderson, J. Shabanowitz, K. Sakaguchi, H. Michel, N. Sevilir, A.L. Cox, E. Appella, V.H. Engelhard, Science 255 (1992) 1261.
[16] R.A. Henderson, H. Michel, K. Sakaguchi, J. Shabanowitz, E. Appella, D.F. Hunt, V.H. Engelhard, Science 255 (1992) 1264.
[17] D.F. Hunt, H. Michel, T.A. Dickinson, J. Shabanowitz, A.L. Cox, K. Sakaguchi, E. Appella, H.M. Grey, A. Sette, Science 256 (1992) 1817.
[18] A.J. Link, J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik, J.R. Yates 3rd, Nat. Biotechnol. 17 (1999) 676.
[19] D.A. Wolters, M.P. Washburn, J.R. Yates 3rd, Anal. Chem. 73 (2001) 5683.
[20] M.P. Washburn, D. Wolters, J.R. Yates 3rd, Nat. Biotechnol. 19 (2001) 242.
[21] M.S. Lipton, L. Pasa-Tolic, G.A. Anderson, D.J. Anderson, D.L. Auberry, J.R. Battista, M.J. Daly, J. Fredrickson, K.K. Hixson, H. Kostandarithes, C. Masselon, L.M. Markillie, R.J. Moore, M.F. Romine, Y. Shen, E. Stritmatter, N. Tolic, H.R. Udseth, A. Venkateswaran, K.K. Wong, R. Zhao, R.D. Smith, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 11049.
[22] J. Peng, J.E. Elias, C.C. Thoreen, L.J. Licklider, S.P. Gygi, J. Proteome Res. 2 (2003) 43.
[23] M. Taoka, Y. Yamauchi, T. Shinkawa, H. Kaji, W. Motohashi, H. Nakayama, N. Takahashi, T. Isobe, Mol. Cell Proteomics 3 (2004) 780.
[24] S. Wolff, A. Otto, D. Albrecht, J. Stahl Zeng, K. Büttner, M. Glückmann, M. Hecker, D. Becher, Mol. Cell Proteomics 21 (2006) 21.
[25] F. Kunst, N. Ogasawara, I. Moszer, A.M. Albertini, G. Alloni, V. Azevedo, M.G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, et al., Nature 390 (1997) 249.
[26] U.N. Streips, F.W. Polio, J. Bacteriol. 162 (1985) 434.
[27] A. Richter, M. Hecker, FEMS Microbiol. Lett. 36 (1986) 69.
[28] A.L. Sonenshein, J.A. Hoch, R. Losick (Eds.), *Bacillus subtilis* and its Closest Relatives, ASM Press, Washington, DC, USA, 2001, p. 4.
[29] H. Langen, B. Takacs, S. Evers, P. Berndt, H.W. Lahm, B. Wipf, C. Gray, M. Fountoulakis, Electrophoresis 21 (2000) 411.
[30] J.D. Jaffe, N. Stange-Thomann, C. Smith, D. DeCaprio, S. Fisher, J. Butler, S. Calvo, T. Elkins, M.G. FitzGerald, N. Hafez, C.D. Kodira, J. Major, S. Wang, J. Wilkinson, R. Nicol, C. Nusbaum, B. Birren, H.C. Berg, G.M. Church, Genome Res. 14 (2004) 1447.
[31] A. Tebbe, C. Klein, B. Bisle, F. Siedler, B. Scheffer, C. Garcia-Rizo, J. Wolfertz, V. Hickmann, F. Pfeiffer, D. Oesterhelt, Proteomics 5 (2005) 168.
[32] C. Klein, C. Garcia-Rizo, B. Bisle, B. Scheffer, H. Zischka, F. Pfeiffer, F. Siedler, D. Oesterhelt, Proteomics 5 (2005) 180.
[33] C. Eymann, A. Dreisbach, D. Albrecht, J. Bernhardt, D. Becher, S. Gentner, T. Tam le, K. Büttner, G. Buurman, C. Scharf, S. Venz, U. Völker, M. Hecker, Proteomics 4 (2004) 2849.
[34] K. Büttner, J. Bernhardt, C. Scharf, R. Schmid, U. Mäder, C. Eymann, H. Antelmann, A. Völker, U. Völker, M. Hecker, Electrophoresis 22 (2001) 2908.
[35] A. Görg, C. Obermaier, G. Boguth, A. Csordas, J.J. Diaz, J.J. Madjar, Electrophoresis 18 (1997) 328.
[36] A. Görg, G. Boguth, C. Obermaier, W. Weiss, Electrophoresis 19 (1998) 1516.
[37] S. Ohlmeier, C. Scharf, M. Hecker, Electrophoresis 21 (2000) 3701.
[38] J. Bernhardt, U. Völker, A. Völker, H. Antelmann, R. Schmid, H. Mach, M. Hecker, Microbiology 143 (1997) 999.
[39] H. Antelmann, J. Bernhardt, R. Schmid, H. Mach, U. Völker, M. Hecker, Electrophoresis 18 (1997) 1451.
[40] R.A. VanBogelen, E.E. Schiller, J.D. Thomas, F.C. Neidhardt, Electrophoresis 20 (1999) 2149.
[41] M. Hecker, U. Völker, Proteomics 4 (2004) 3727.
[42] L.T. Tam, H. Antelmann, C. Eymann, D. Albrecht, J. Bernhardt, M. Hecker, Proteomics 6 (2006) 4565.
[43] S. Luhn, M. Berth, M. Hecker, J. Bernhardt, Proteomics 3 (2003) 1117.
[44] M. Hecker, U. Völker, Adv. Microb. Physiol. 44 (2001) 35.
[45] J. Mostertz, C. Scharf, M. Hecker, G. Homuth, Microbiology 150 (2004) 497.
[46] J.D. Helmann, M.F. Wu, A. Gaballa, P.A. Kobel, M.M. Morshedi, P. Fawcett, C. Paddon, J. Bacteriol. 185 (2003) 243.
[47] J. Ollinger, K.B. Song, H. Antelmann, M. Hecker, J.D. Helmann, J. Bacteriol. 188 (2006) 3664.
[48] S. Nakano, E. Kuster-Schock, A.D. Grossman, P. Zuber, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 13603.
[49] S. Fernandez, S. Ayora, J.C. Alonso, Res. Microbiol. 151 (2000) 481.
[50] M. Hecker, U. Völker, Mol. Microbiol. 29 (1998) 1129.
[51] T.A. Gaidenko, C.W. Price, J. Bacteriol. 180 (1998) 3730.
[52] S.H. Fisher, Mol. Microbiol. 32 (1999) 223.
[53] L.V. Wray Jr., A.E. Ferson, K. Rohrer, S.H. Fisher, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 8841.
[54] K. Yoshida, H. Yamaguchi, M. Kinehara, Y.H. Ohki, Y. Nakaura, Y. Fujita, Mol. Microbiol. 49 (2003) 157.
[55] M. Debarbouille, R. Gardan, M. Arnaud, G. Rapoport, J. Bacteriol. 181 (1999) 2059.
[56] P. Babitzke, P. Gollnick, J. Bacteriol. 183 (2001) 5795.
[57] P. Gollnick, P. Babitzke, A. Antson, C. Yanofsky, Annu. Rev. Genet. 39 (2005) 47.
[58] J. Bernhardt, J. Weibezahn, C. Scharf, M. Hecker, Genome Res. 13 (2003) 224.
[59] H.M. Blencke, G. Homuth, H. Ludwig, U. Mäder, M. Hecker, J. Stülke, Metab. Eng. 5 (2003) 133.
[60] K. Yoshida, K. Kobayashi, Y. Miwa, C.M. Kang, M. Matsunaga, H. Yamaguchi, S. Tojo, M. Yamamoto, R. Nishi, N. Ogasawara, T. Nakayama, Y. Fujita, Nucleic Acids Res. 29 (2001) 683.
[61] T. Koburger, J. Weibezahn, J. Bernhardt, G. Homuth, M. Hecker, Mol. Genet. Genomics 274 (2005) 1.
[62] P. Servant, D. Le Coq, S. Aymerich, Mol. Microbiol. 55 (2005) 1435.

[63] C. Eymann, H. Mach, C.R. Harwood, M. Hecker, Microbiology 142 (1996) 3163.
[64] H. Antelmann, C. Scharf, M. Hecker, J. Bacteriol. 182 (2000) 4478.
[65] F.M. Hulett, Mol. Microbiol. 19 (1996) 933.
[66] N.E. Allenby, N. O'Connor, Z. Pragai, A.C. Ward, A. Wipat, C.R. Harwood, J. Bacteriol. 187 (2005) 8063.
[67] V. Molle, Y. Nakaura, R.P. Shivers, H. Yamaguchi, R. Losick, Y. Fujita, A.L. Sonenshein, J. Bacteriol. 185 (2003) 1911.
[68] R.A. Britton, P. Eichenberger, J.E. Gonzalez-Pastor, P. Fawcett, R. Monson, R. Losick, A.D. Grossman, J. Bacteriol. 184 (2002) 4881.
[69] D. Höper, J. Bernhardt, M. Hecker, Proteomics 6 (2006) 1550.
[70] P.R. Jungblut, U.E. Schaible, H.J. Mollenkopf, U. Zimny-Arndt, B. Raupach, J. Mattow, P. Halada, S. Lamer, K. Hagens, S.H. Kaufmann, Mol. Microbiol. 33 (1999) 1103.
[71] H. Tjalsma, A. Bolhuis, J.D. Jongbloed, S. Bron, J.M. van Dijl, Microbiol. Mol. Biol. Rev. 64 (2000) 515.
[72] J.M. van Dijl, P.G. Braun, C. Robinson, W.J. Quax, H. Antelmann, M. Hecker, J. Müller, H. Tjalsma, S. Bron, J.D. Jongbloed, J. Biotechnol. 98 (2002) 243.
[73] H. Tjalsma, H. Antelmann, J.D. Jongbloed, P.G. Braun, E. Darmon, R. Dorenbos, J.Y. Dubois, H. Westers, G. Zanen, W.J. Quax, O.P. Kuipers, S. Bron, M. Hecker, J.M. van Dijl, Microbiol. Mol. Biol. Rev. 68 (2004) 207.
[74] I. Hirose, K. Sano, I. Shioda, M. Kumano, K. Nakamura, K. Yamane, Microbiology 146 (2000) 65.
[75] H. Antelmann, H. Tjalsma, B. Voigt, S. Ohlmeier, S. Bron, J.M. van Dijl, M. Hecker, Genome Res. 11 (2001) 1484.
[76] I. Humphery-Smith, M. Hecker (Eds.), Microbial Proteomics, Wiley-VCH, Weinheim, 2006, pp. 179–208.
[77] A.L. Sonenshein, J.A. Hoch, R. Losick (Eds.), *Bacillus subtilis* and its Closest Relatives, ASM Press, Washington, DC, USA, 2001, pp. 193–201.
[78] S. Leskela, E. Wahlstrom, V.P. Kontinen, M. Sarvas, Mol. Microbiol. 31 (1999) 1075.
[79] K. Bunai, M. Ariga, T. Inoue, M. Nozaki, S. Ogane, H. Kakeshita, T. Nemoto, H. Nakanishi, K. Yamane, Electrophoresis 25 (2004) 141.
[80] H. Tjalsma, J.M. van Dijl, Proteomics 5 (2005) 4472.
[81] J.M. Ghuysen, J. Lamotte-Brasseur, B. Joris, G.D. Shockman, FEBS Lett. 342 (1994) 23.
[82] O. Schneewind, D. Mihaylova-Petkov, P. Model, Embo J. 12 (1993) 4803.
[83] W.W. Navarre, O. Schneewind, Mol. Microbiol. 14 (1994) 115.
[84] H. Antelmann, H. Yamamoto, J. Sekiguchi, M. Hecker, Proteomics 2 (2002) 591.
[85] V. Santoni, M. Molloy, T. Rabilloud, Electrophoresis 21 (2000) 1054.
[86] R. Schmid, J. Bernhardt, H. Antelmann, A. Völker, H. Mach, U. Völker, M. Hecker, Microbiology 143 (1997) 991.
[87] A. Krogh, B. Larsson, G. von Heijne, E.L. Sonnhammer, J. Mol. Biol. 305 (2001) 567.