Critical Reviews in Biochemistry and Molecular Biology, 38(3):173–198 (2003)

Bacillus Species Proteins Involved in Spore Formation and Degradation: From Identification in the Genome, to Sequence Analysis, and Determination of Function and Structure

Mark J. Jedrzejas and Wendy J.M. Huang*

Children's Hospital Oakland Research Institute, Oakland, CA 94609, USA

* **To whom correspondence should be addressed:** Children's Hospital Oakland Research Institute, 5700 Martin Luther King Jr. Way, Oakland, CA 94609. Phone 510-450-7932. Fax 510-450-7910. Email mjedrzejas@chori.org

TABLE OF CONTENTS

1040-9238/03/$.50

ABSTRACT: The members of *Bacillus* species are Gram-positive, ubiquitous spore-forming bacilli. Several genomic sequences have been made available during recent years, including *Bacillus subtilis*, a model organism among this genus, *Bacillus anthracis*, and their analyses provided a wealth of information about spore-forming bacteria. Some members of this species can cause serious diseases in livestock and humans. An important pathogen in this group of organisms is *B. anthracis*, which is the causative agent of anthrax. A summary of the *B. subtilis* genome information, based on the publicly released sequence, that allowed for the identification and characterization of new and novel proteins of this organism as well as similar proteins from other members of *Bacillus* species is provided. The primary goal for this work is to present a review of the genome sequence-identified genes that encode proteins involved in the sporulation, germination, and outgrowth processes. These three processes are essential for spore development and later its transformation into a vegetative cell. Additionally, for a few selected examples of the protein products of the identified genes, the application of bioinformatics and modeling tools is illustrated in order to determine their likely structure and function. Two three-dimensional models of the structures of such proteins, PrfA endonuclease and phosphatase PhoE, are presented together with the structure-based functional conclusions. The review of such studies provides an example of how the genomic sequence can be utilized in order to elucidate the structure and function of proteins, in particular proteins of the *Bacillus* species. Because only a limited number of proteins of *Bacillus* species organisms are involved in the synthesis and degradation of spores and have been characterized to date, this genome-based analysis may provide new insights into the developmental processes of bacterial organism.

I. INTRODUCTION

The objective of this review is to summarize the information about the genome-based identification of proteins of *Bacillus subtilis* and, to a smaller extent, *Bacillus anthracis*, which are involved in the spore-forming and spore-degradation (germination and outgrowth) processes. For a selected few of such identified proteins, example analyses using the tools of bioinformatics, structural genomics, and molecular modeling are illustrated in order to determine their function and structure. Another group among the genome identified proteins are proteins with known three-dimensional structures, determined experimentally either by means of X-ray crystallography or by nuclear magnetic resonance (NMR). The information about these proteins, their function, and structure is summarized. The availability of structural information, modeled or experimental, about a group of the identified proteins allows for their further analysis and for the relatively reliable assignment of their function. It is clearly evident that in the process of identification and analyses of these proteins, a better overall understanding of spore-forming pathogens and their interactions with human hosts can be developed. This insight may shed light on the elucidation of basic scientific processes and specific function(s) these proteins are involved in as well as discover the role of these proteins in the pathogenesis of spore-forming bacteria and possible other microbial organisms. The availability of such information in addition to advancing basic science may lead to the development of new prophylactic and/or therapeutic agents, such as vaccines and/or antibiotics, against *B. anthracis* and other spore-forming bacteria.

II. SPORE-FORMING BACTERIA

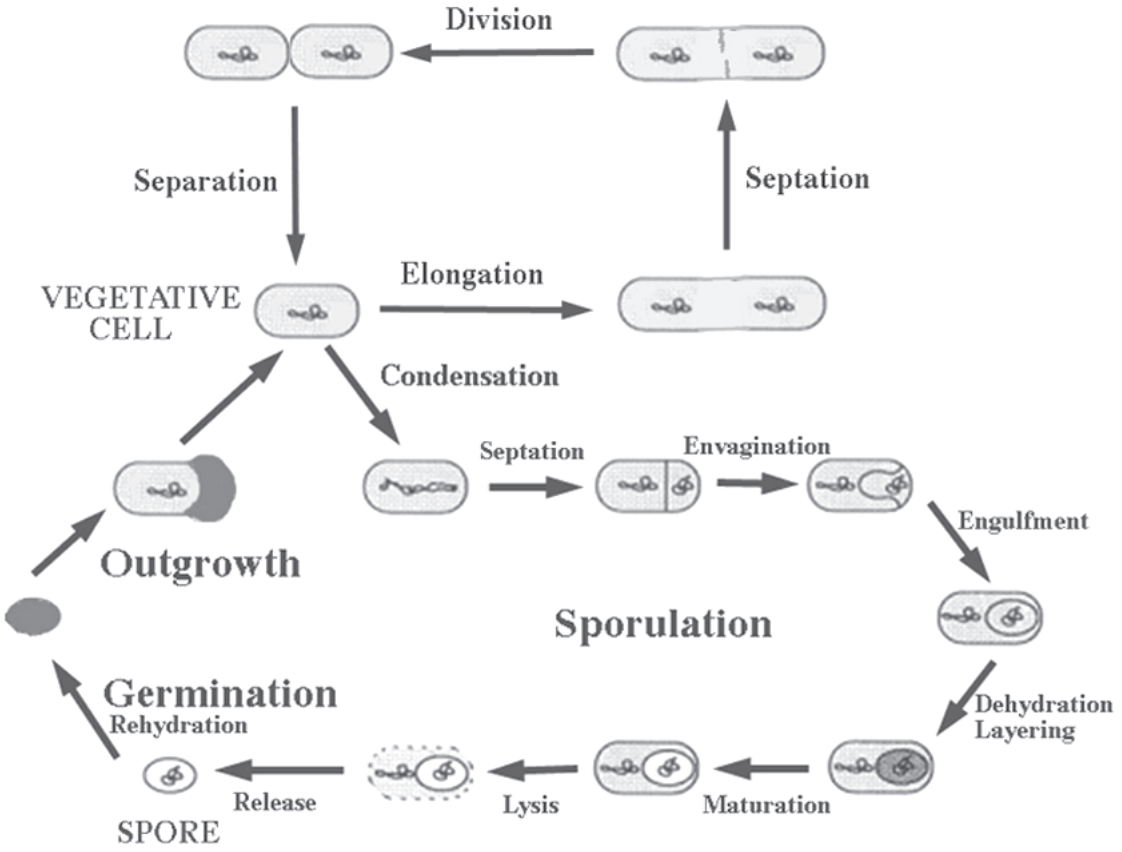
Members of *Bacillus* species, including *B. subtilis*, are Gram-positive, ubiquitous, spore-forming bacilli that can survive for years, possibly millions of years, even in harsh environments (Figure 1A and B) (Cano and Borucki, 1995; Kennedy *et al*., 1994; Slepecky, 1992). Under adverse conditions of nutrient limitations vegetative cells have the ability to undergo transformation into a spore in the sporulation process (Figure 1A). Spores have a totally different structure than a vegetative cell (Figure 1B, C) and are designed to survive under adverse conditions. Later when nutrients become available, the spore is triggered into a germination process followed by outgrowth, which results in the synthesis of a vegetative cell. The central part of the spore contains a core in which the genome, minimal set of proteins and other molecules, large amounts of Ca^{2+} and dipicolinic acid are stored. The spore and its core are protected on the outside by: (1) cortex which resembles peptidoglycan structures of Gram-positive bacteria but has a different structure than in vegetative cells, (2) the coat that consists largely of highly crosslinked proteins, and is sometimes surrounded by (3) exosporium that consists of polysaccharides, lipid structures, and proteins (Driks, 1999). These various structural parts of spores are often separated from one another by membrane layers (Figure 1B, C).

Some of these organisms cause serious diseases in livestock and in humans (Hanna and Ireland, 1999; Meselson *et al*., 1994; Thorne, 1993). The main pathogenic organism among this genus is *B. anthracis*, which is the causative agent of anthrax (Hanna and Ireland, 1999). *B. anthracis* is considered to be a major threat as a biological warfare and terrorist agent (Lalitha and Thomas, 1997; LaForce, 1994; Ivins *et al*., 1992; Irvins and Welkos, 1988). In part due to the recent terrorism events in the U.S., prevention and treatment of anthrax has become a priority from the public health perspective as well as for the medical-scientific community working in this field. Thus, there is a significant need to identify new strategies to prevent and to treat disease due to spore-forming bacterial pathogens, including *B. anthracis*.

Recent studies have suggested that certain *Bacillus* species proteins could be essential for normal spore development and degradation (Figure 1A) (Jedrzejas, 2002a,b). Such candidate proteins include, for example, cofactor independent phosphoglycerate mutase (iPGM), germination protease (GPR) (Ponnuraj *et al*., 2000a,b; Ponnuraj *et al*., 1999), penicillin-binding protein-related factor A endonuclease (PrfA-endonuclease) (Rigden *et al*., 2002b; Kelly *et al*., 2000), and NAD^+ synthetase (NADS) (Devedjiev *et al*., 2001). The availability of genomic sequence of *B. subtilis* and other spore-forming bacterial allowed for the additional identification of numerous other proteins that are uniquely essential for this group of bacteria.

III. GENOMIC SEQUENCES OF *Bacillus* SPECIES

The determination of a complete sequence of the *B. subtilis* genome (strain 168) (Kunst *et al*., 1997) has already been elucidated, and the information is publicly available at www.ncbi.nlm.nig.gov under Microbial Genomes, at www.tigr.org/tigr-scripts/CMR2/GenomePage3.spl?database=ntbs01, or at genolist.pasteur.fr/SubtiList. Completed genomic sequence for another member of bacilli is also available, alkaliphilic *Bacillus halodurans* strain C-125 (available at www.ncbi.nlm.nig.gov under Microbial Genomes) (Takami *et al*., 2000). The genome sequences of other members of *Bacillus* species, *B. anthracis* strain A2012 and Ames (Henderson *et al*., 1994) and *Bacillus stearothermophilus* (Nazina *et al*., 2001), are in the final stages of their determination and annotation, and the partially available sequence databases of these two bacterial organisms are available to BLAST searches (www.ncbi.nlm.nig.gov under Microbial Genomes) (Altschul *et al*., 1990). The sequence of the *Bacillus cereus* genome (Daffonchio *et al*., 1998) is also in progress. The majority of the identified proteins among the *Bacillus* genus are also present in non-spore-forming bacterial organisms. Some of the genome-encoded proteins, however, are specific only to spore-forming bacteria due to their distinct properties when compared with other bacteria. A number of other proteins are present in both spore-forming and non-spore-forming bacteria but those of spore formers evolved to specifically facilitate the needs and the requirement of making and degrading spores. For example, cofactor-independent phosphoglycerate mutase (pgm) of *Bacillus* species (the only pgm in *B. subtilis*) has evolved in spore-forming bacteria to possess a distinct pH-dependent activity that facilitates regulation of its activity during spore formation (lower pH) and degradation (higher pH, comparable to pH in non-spore-forming



A

FIGURE 1. Life-cycle of spore-forming bacteria such as *Bacillus subtilis.* **(A)** Life cycle of spore-forming bacteria. Under adverse conditions of nutrient limitations vegetative cells of *Bacillus* species have the ability to undergo transformation into a spore (sporulation process). Spores have totally different structure than a vegetative cell of this organism (see panel B of this figure) and are designed to survive for a long time under adverse conditions. When nutrients become available the spore is triggered into a germination process followed by outgrowth, which result in the synthesis of a vegetative cell. **(B)** Schematic structure of a spore. The spore core contains the genome, minimal set of proteins and other molecules, large amounts of Ca^{2+} and dipicolinic acid. The cortex resembles peptidoglycan structures of Gram-positive bacteria but has a different structure than in vegetative cells. The coat consists largely of highly crosslinked proteins, whereas the exosporium, that is present only in some species (e.g., *B. anthracis*), consists of polysaccharides, lipid structures, and proteins. These various structural parts of spores are often separated from one another by membrane layers (Driks, 1999). **(C)** Electron micrograph of *B. subtilis* spore. Inner, outer coats, cortex, and the core are labeled. The figure was generously provided by Dr. Adam Driks.

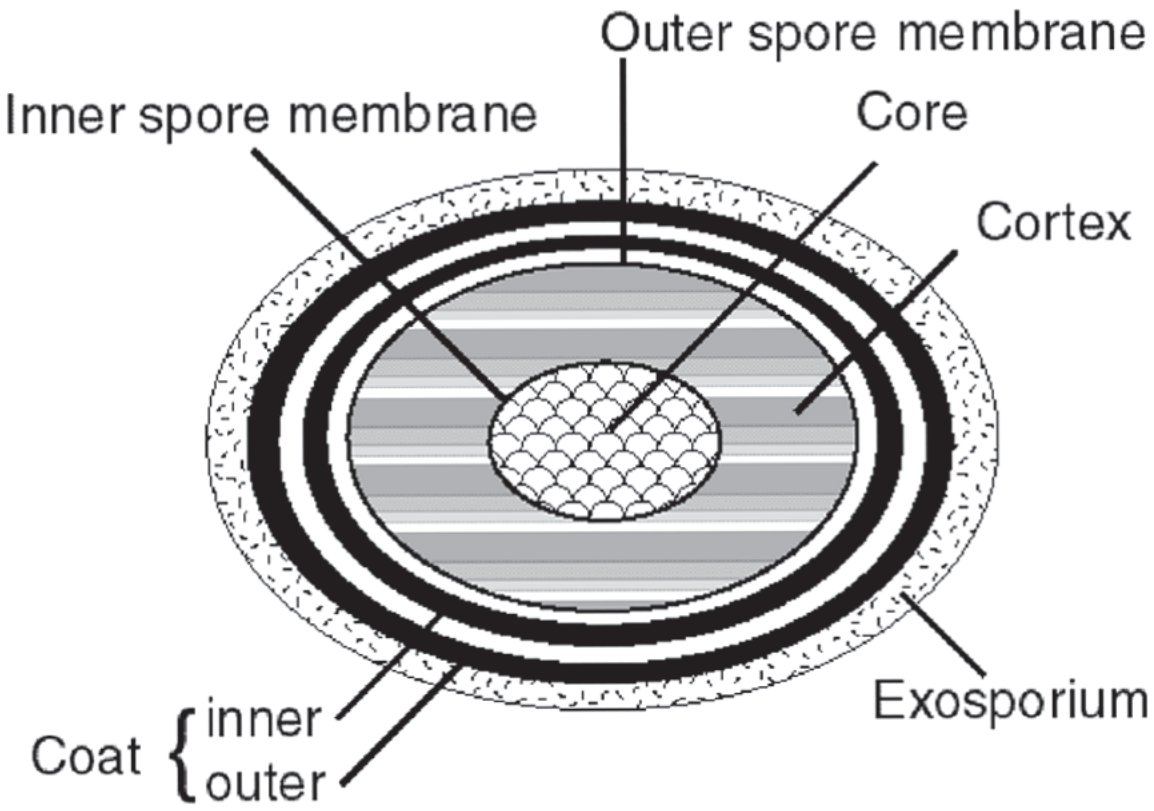


FIGURE 1B

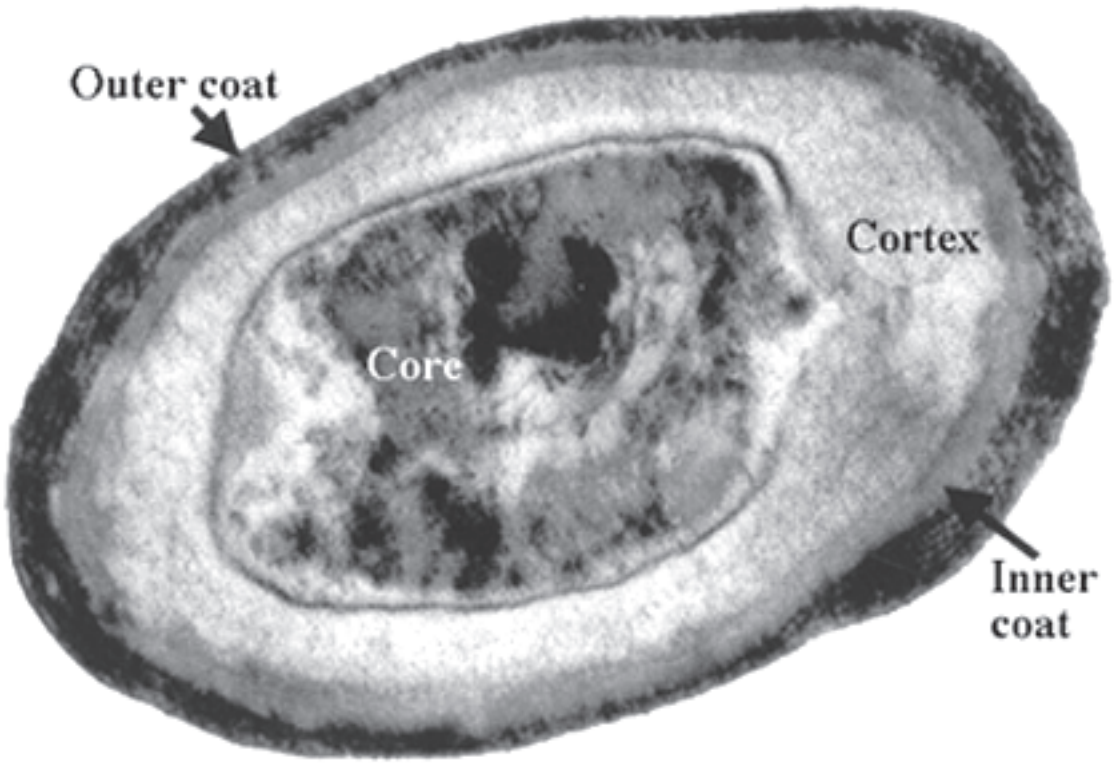


FIGURE 1C

bacteria). Therefore, this enzyme is essential for spore-forming organisms but is also present in other bacteria for which the pH sensitivity of this enzyme's activity is not present or is functionally not important (Jedrzejas and Setlow, 2001; Jedrzejas, 2000; Jedrzejas *et al*., 2000a,b).

IV. SEQUENCE OF *Bacillus subtilis* GENOME

The strain 168 of *B. subtilis* genome consists of a single circular chromosome, 4,214,810 base pairs in size, with a low average G+C content of 43.5%, but this content varies significantly throughout the chromosome. Based on the identification of ~4100 protein-coding genes (open reading frames), biological roles were assigned to only 58% of the predictions, 12% matched proteins of unknown function of other organisms, and 26% had no database match to any know protein (Kunst *et al*., 1997). In general, the analysis of sequence of this genome suggests that *B. subtilis* has a large number of transporters, including ATP-dependent transporters (ABC type), a large number of proteins secreted into extracellular space, and proteins interconverting low-molecular-weight compounds (e.g., protein members of metabolic pathways).

In addition to chromosomal DNA, *B. subtilis* or *B. anthracis* as well as other members of this species, harbor plasmids. The sequencing of four plasmids has been completed for *B. subtilis*, and these sequences are also publicly available at www.ncbi.nlm.nih.gov under Microbial Genomes. The number of extrachromosomal DNA in plasmids of *B. subtilis* varies depending on the strain of this organism, but sequence information is available for pTA1015 (5807 base pairs), pTA1040 (7837 base pairs), pTA1060 (natto strain; 8,737 base pairs) (Meijer *et al*., 1995), and p1414 (Russian soil isolate; 7949 base pairs) (Thorsted *et al.,* 1999) plasmids. For other members of *Bacillus* genus multiple plasmids have been sequenced and annotated. For example, two strains of *B. anthracis* (including strain A2012) contain two plasmids, pXO1 (181,677 and 181,654 base pairs) (Read *et al*., 2002; Okinaka *et al*., 1999a) and pXO2 (94,829 and 96,231 base pairs) (Read *et al*., 2002; Okinaka *et al*., 1999b), which were sequenced and annotated. In *B. anthracis* the pXO1 plasmid harbors the anthrax toxin genes, protective antigen, lethal and edema factors (Okinaka *et al*., 1999a). Additional plasmids for other members of *Bacillus* species have also been sequenced and annotated (www.ncbi.nlm.nih.gov:80/PMGifs/Genomes).

V. GENOME ENCODED PROTEINS THAT ARE DIRECTLY RELATED TO SPORES

The first thorough analysis of the genomic sequence of *B. subtilis* reported by Kunst *et al*. (1997) identified genes coding for proteins involved in various aspects of cellular functionality of this organism. A case-by-case analysis of these proteins allowed for the identification of those that are primarily involved in spore functionality such as spore formation, and degradation. The use of BLAST searches in order to confirm and to determine the functionality to these identified proteins was performed following standard methodology (Altschul *et al*., 1990). The identified spore-related proteins were grouped into categories as reported by Kunst *et al*. (1997). This selection of these macromolecules was also confirmed using the website listing of sporulation genes at www1.rhbnc.ac.uk/biological-sciences/cutting/spo/spo_frames.html. The results of such analyses are listed, recapitulated, and described in Table 1. Also, our own analysis of *B. anthracis* genome, strain A2012, revealed that majority of the spore-related proteins of *B. subtilis* are also present in the *B. anthracis* and they likely perform similar functions (Table 1). Therefore, the discussion of *B. subtilis* genes and their proteins products provided below is relevant to *B. anthracis* genes and the proteins they encode.

A. Cell Envelope and Cellular Processes

The analysis of identified protein and their groupings shows that the vast majority of spore-related proteins are involved in functions related to cell envelope and cellular processes. Out of 866 identified genes in this group, as many as 176 appear to be directly related to spores. Among this broad group the major subcategories are proteins related to cell wall (9 genes), sensors/signal transduction (3 genes), cell division (2 genes), sporulation (139 genes), and germination (23 genes).

The cell wall-related genes encode for proteins in two main subgroups. One is the cell wall degradation group (CwlC, CwlD, and CwlJ) and the other is a group of penicillin-binding proteins (DacB and DacF, PdbA, PdbB, PdbE, and PdbF). CwlC is

TABLE 1
Examples of Proteins and Protein Groups of *B. subtilis* and *B. anthracis* Related to Spores and Spore-Related Processes

Protein groups/ subgroups	Total number of genes/description of genes	Positions (*B. subtilis* genome)	*B. anthracis*[a]
	I. CELL ENVELOPE and CELLULAR PROCESSES, 866 genes		
Cell wall:	**93 genes**		
cwlC	N-acetylmuramoyl-L-alanine amidase, disruption of *cwlC* does not prevent mother cell lysis at the end of sporulation, CwlC together with LytC are involved in cell/cell-wall lysis in sporulation (255 residues)	1873	+
cwlD	N-acetylmuramoyl-L-alanine amidase, disruption of *cwlD* leads to a complete block of cortex degradation and germination, cortex-lytic enzyme specific for germination (237 residues)	157	+
cwlJ	cell wall hydrolase involved during the sporulation/germination process (142 residues)	282	+
dacB	penicillin-binding protein 5, essential for spore cortex peptidoglycan biosynthesis (3 proteins: 382, 196, and 179 residues)	2424	+
dacF	penicillin-binding protein, involved in peptidoglycan biosynthesis, not impair sporulation nor germination (389 residues)	2445	+
pbpA	penicillin-binding protein 2A, involved in spore outgrowth (716 residues)	2583	+
pbpB	penicillin-binding protein 2B, involved in cell-division septum formation (716 residues)	1581	+
pbpE	penicillin-binding protein 4, involved in spore cortex synthesis (451 residues)	3535	+
pbpF	penicillin-binding protein 1A, involved in germination (714 residues)	1083	+
Sensors	**38 genes**		
kinA	two-component sensor histidine kinase, involved in initiation of sporulation, mutation in *kinA* results in delay in sporulation-associated events and spore formation (606 residues)	2469	-

TABLE 1 (continued)

kinB	two-component sensor histidine kinase, involved in initiation of sporulation, when combined with kinA mutation, mutation in *kinB* severely reduces spore formation (428 residue)	3229	+
kinC	two-component sensor histidine kinase, involved in initiation of sporulation, phospho-relay-independent (428 residues)	1518	+
Cell division:	**21 genes**		
ftsA	cell-division protein, involved in septum formation, *Escherichia coli* homologs of *ftsA* and *ftsB* enzymes control initiation of cell-division, deletion leads to sick but viable cells (440 residues)	1596	+
ftsZ	cell-division protein, present in one operon with *ftsA*, involved in septum formation, essential for cell viability mutations in *ftsAZ* cause a filamentous phenotype during vegetative growth and prevent sporulation (382 residues)	1597	+
Sporulation:	**139 genes**		
bofA	inhibitor of the pro-σ^K processing machinery, mutation in *bofA* leads to defect in sporulation and in germination (87 residues)	30	-
bofC	forespore regulator of the σ^K checkpoint, might inhibit signaling action of SpoIVB (170 residues)	2837	+
cge (A, B, C, D, E)	maturation of the outermost layer of spores (*i.e.*, mutation in *cgeA* leads to spores with altered surface properties)	2148, -48, -48, -47, -46	-
cot (A, B, C, D, E, F, G, H, JA, JB, JC, K, L, M, N, S, T, V, W, X, Y, Z)	inner and outer spore coat or coat-associated proteins,	various	-,-,-,+,+,+ ,-,+,+,+,+ ,-,-,-,-,-,+ ,-,-,-,+,+
csgA	sporulation specific small, acid soluble protein (82 residues)	228	-
jag	SpoIIIJ-associate protein (208 residues)	4213	+
kapB	activator of KinB in the initiation of sporulation (128 residues)	3230	+
kapD	inhibitor of the KinA pathway to sporulation (205 residues)	3232	+
kbaA	activation of the KinB signaling pathway to sporulation, hydrophobic (198 residues), disruption of *kbaA* and *kinA* has a strong negative effect on sporulation	159	-
obg	GTP-binding protein involved in initiation of sporulation, (428 residues)	2853	+
phr (A, C, E, F, G, I, K)	group of phosphatase regulators	various	-
rapA	response regulator aspartate phosphatase, a phosphatase acting specifically on Spo0F~P (346 residues)	1315	-
rapB (= spo0P)	response regulator aspartate phosphatase, similar to RapA (377 residues)	3771	-
rap (C, D, E, F, G, H, I, J, K)	response regulator aspartate phosphatases	various	-
sinI and *R*	two-gene operon encoding SinI (57 residues) and its antagonist SinR (111 residues, repression or activation of gene expression by binding of SinR to DNA)	1461 2552	+ +

TABLE 1 (continued)

soj	centromere-like function involved in forspore chromosome partitioning / inhibition of SpoOA activation (253 residues)	4206	+
splB	spore photoproduct lyase (342 residues)	1461	+
spm (A, B)	spore maturation protein, involved in spore core dehydration (196 and 171 residues, respectively)	2423, -22	+
spo0B	sporulation initiation phosphoprotein, two-gene operon encoding Spo0B (192 residues) and Obg (428 residues)	2854	+
spo0E	negative sporulation regulation phosphatase (85 residues)	1430	+
spo0J	chromosome positioning near pole and transport via polar septum, two-gene operon encoding a 253-residue protein of unknown function and Spo0J (282 residues)	4206	+
spoIIA (A, B)	anti-anti-sigma-factor and anti-σ factor (and serine kinase, respectively, three-gene operon encoding SpoIIAA (117 residues), SpoIIAB (146 residues), and sigma F (255 residues)	2444	+
spoIIB	endospore development (332 residues)	2864	-
spoIID	required for the complete dissolution of the asymmetric septum (343-residue), N-terminal transmembrane domain	3777	+
spoIIE	serine phosphatase / asymmetric septum formation (826 residues) transmembrane segments	71	+
spoIIGA	protease processing pro-σ^E to active σ^E, two-gene operon encoding SpoIIGA (309 residues, transmembrane parts) and pro-σ^E (239 residues)	1603	-/+
spoIIM	required for dissolution of the septal cell wall, (215 residues) putative transmembrane segments	2450	+
spoIIP	required for dissolution of the septal cell wall, (401 residues)	2634	+
spoIIQ	required for completion of engulfment, membrane-bound (283 residues), initially targeted to the septum	3760	+
spoIIR	required for processing of pro-σ^E, secreted protein (224 residues)	3794	+
spoIIS (A, B)	SpoIISA is lethal when synthesized during vegetative growth without SpoIISB, SpoIISB disruption disrupts sporulation after septum formation, two-gene operon encoding SpoIISA (248 residues) and SpoIISB (56 residues)	1349, -48	-
spoIIIA (A, B, C, D, E, F, G, H)	mutants block sporulation after engulfment, eight-gene operon encoding SpoIIIAA (307 residues), SpoIIIAB (171 residues), SpoIIIAC (68 residues), SpoIIIAD (133 residues), SpoIIIAE (399 residues), SpoIIIAF (206 residues), SpoIIIAG (229 residues), and SpoIIIAH (218 residues)	2537, -36, -35, -35, -35, -34, -33, -32	+
spoIIIE	DNA translocate required for chromosome partitioning, (787 residues) possible transmembrane parts and nucleotide-binding site	1752	+
spoIIIJ	essential for σ^G activity, two-gene operon encoding SpoIIIJ (260 residues) and Jag (208 residues)	4214	+
spoIVA	required for spore cortex formation and coat assembly, acidic protein (492 residues), contains nucleoide-binding motif	2387	+
spoIVB	intercompartmental signalling of pro-σ^K processing in the mother cell, (425 residues), N-terminal transmembrane domain	2520	+
spoIVCA	site-specific DNA recombinase required to make *sigK* gene, (500 residues)	2654	-
spoIVF (A, B)	two-gene operon encoding SpoIVFA (inhibitor of SpoIVFB, 264 residues) and SpoIVFB (protease processing pro-σ^K to σ^K, 288 residues), both with putative transmembrane domains	2857, -56	+
spoVA (A, B, C, D, E,	mutants lead to formation of immature spores, six-gene operon encoding SpoVAA (200 residues), SpoVAB (141	2443, -42,	+

F)	residues), SpoVAC (150 residues), SpoVAD (338 residues), SpoVAE (323 residues), and SpoVAF (492 residues), likely all contain membrane-spanning domains	-41, -41, -40, -39	
spoVB	involved in spore cortex synthesis, (518 residues), many transmembrane domains	2829	+
spoVC	thermosensitive mutant blocs spore coat formation (188 residues)	60	+
spoVE	required for spore cortex synthesis (366 residues)	1590	+
spoVF (A, B)	dipicolinate synthase (Dpa) subunit A and B, respectively, two-gene operon encoding DpaA (297 residues) and DpaB (200 residues)	1744, -45	+
spoVG	required for spore cortex synthesis (97 residues)	56	+
spoVK	disruption leads to production of immature spores (322 residues) contains a putative nucleotide-binding motif)	1873	+
spoVM	required for normal spore cortex and coat synthesis (26 residues)	1655	+
spoVR	involved in spore cortex synthesis (468 residues)	1015	+
spoVS	required for dehydration of the spore core and assembly of the coat (86 residues)	1769	+
spoVID	required for the assembly of the spore coat, two-gene operon encoding SpoVID (575 residues), a very acidic protein, and Orf2 (341 residues).	2872	-
sps (A, B, C, D, E, F, G, I, J, K)	spore coat polysaccharide synthesis, operon encoding eleven proteins, SpsA-K (256, 472, 389, 289, 373, 239, 222, 117, 246, 315, and 432 residues respectively), some of which are similar to proteins involved in carbohydrate biosynthesis	3892, -91, -90, -89, -88, -87, -86, -85, -84, -83	-,-, -,-, -,-, -,+, +,+
ssp (A, B, C, D, E, F)	small acid-soluble proteins α, β, α/β, α/β, γ, α/β type, respectively; SspA-F (69, 67, 72, 64, 84, and 61 residues, respectively)	3025, 1050, 2156, 1413, 937, 53	+,+, +,+, -,-
usd	required for translation of *spoIIID* (36 residues)	3748	-
yknT	sporulation protein, σ^E controlled (321 residues)	1495	-
ykvU	spore cortex membrane protein (445 residues)	1449	-
ynzH	spore coat protein (86 residues)	1901	-
yobW	membrane protein σ^K controlled (188 residues)	2083	-
yqgT	γ-D-glutamyl-L-diamino acid endopeptidase I (376 residues)	2568	-
yqjG	lipoprotein SpoIIJ-like (275 residues)	2483	+
yra (D, E, F, G)	spore coat proteins (99, 65, 122, and 81 resides, repectively)	2754, -54, -52, -52	+,-, -,+
yrb (A, B, C)	spore coat proteins (240 resides)	2845, -44, -43	-,-, -
ytaA	spore coat protein (357 resides)	3161	-
ytgP	spore cortex protein (544 resides)	3074	+
ytpT	DNA translocase stage III sporulation protein (702 residues)	3051	+
yyaA	DNA-binding protein Spo0J-like (283 residues)	4208	+

TABLE 1 (continued)

Germination:	**23 genes**		
gerA (A, B, C)	germination response to L-alanine, three gene operon; GerAA (480 residues), GerAB (364 residues), and GerAC (373 residues); likely a membrane-located receptor for L-alanine.	3390, -91, 92	+,-, -
gerB (A, B, C)	germination response to glucose, fructose, asparagines, and KCl, three-gene operon encoding GerBA (482 residues), GerBB (367 residues), and GerBC (374 residues), likely a membrane receptor for germinants other than L-alanine	3689, -90, -91	+,-, -
gerC (A, B, C)	three-gene operon encoding GerCA (251 residues), GerCB (233 residues), and GerCC (348 residues)- A - heptaprenyl diphosphate synthase component I, B – meanaquinone biosynthesis methyltransferase, C - heptaprenyl diphosphate synthase component II.	2384, -83, -82	+,+, +
gerD	germination response to L-alanine, and the combination of glucose, fructose, asparagines, and KCl (185 residus)	159	+
gerK (A, B, C)	germination response to glucose, fructose, asparagines, and KCl, likely a membrane receptor for germinants other than L-alanine	20, 23, 21	-
gerM	germination (cortex hydrolysis) and sporulation (polar septa) (336 residues)	2902	+
gpr	spore germination protease (368 residues), initiates degradation of SASPs (the small, acid-soluble spore proteins) during the first minutes of germination	2635	+
sleB	spore-cortex lytic enzyme (305 residues)	2399	+
yfk (Q, R, T)	spore germination proteins (513, 383, and 358 residues respectively)	850, -48, -47	+,-, -
ykvt	spore-cortex lytic enzyme (208 residues)	1448	-
ynd (D, E, F)	spore germination protein (520, 363, and 404 residues respectively	1907, -08, -09	+,-, -
	II. INTERMEDIARY METABOLISM, 742 genes		
Glycolysis:	**28 genes**		
ipgm	cofactor independent phosphoglycerate mutase, (511 residues)	3478	+
Phosphate metabolism:	**9 genes**		
phoA	alkaline phosphatase A (416 residues)	1018	+
phoD	phosphodiesterase/alkaline phosphatase (556 residues)	284	-
phoE (yhfR)	broad specificity phosphatase (193 residues, missannotated as cofactor dependent *pgm*)	1109	+
phoH	phosphate starvation-induced protein (319 residues)	2615	+

III. INFORMATION PATHWAYS, 482 genes

Regulation:	**213 genes**		
abrB	transcription pleiotropic regulator of transcription state genes, (94 residues) hexamer *in vitro* and binds to specific DNA sequences	45	+
gerE	transcription regulator for expression of late spore coat genes (74 residues)	2904	+
hpr	transcription repressor of sporulation, and extracellular protease genes (203 residues)	1073	+
paiA	transcription repressor of sporulation, septation, and degradative enzyme genes (172 residues)	3304	-
paiB	transcription repressor of sporulation, and degradative enzymes enzyme genes (207 residues)	3304	+
sinR	transcription reglulator of post-exponential phase response genes, two-gene operon encoding SinI (57 residues) and SinR (previously Sin, 111 residues)	2552	+
slr	transcription activation of competence develop. and sporulation genes	3529	-
splA	transcription regulator of the spore photoproduct lyase operon (*splAB*), (342 residues), involved in repairing upon germination the damage caused by UV in spore DNA	1461	-
spo0A	two-component response regulator for the initiation of sporulation (267 residues)	2518	+
spo0F	two-component response regulator involved in initiation of sporulation (123 residues)	3809	+
spoIIID	transcription regulator of σ^E- and σ^K-dependent genes (93 residues), binds to DNA	3748	+
spoVT	transcription regulator of σ^G-dependent genes (178 residues)	64	+

IV. OTHER FUNCTIONS, 289 genes

Atypical conditions:			
	72 genes		
yveR	spore coat polysaccharide biosynthesis (344 residues)	3521	-
yvfE	spore coat polysaccharide biosynthesis (388 residues)	3515	-
Detoxification:	**68 genes**		
katA	vegetative catalase I, mutation in *katA* results in hydrogen peroxide sensitivity during sporulation, (483 residues) catalase II (686 residues)	960	+
katB	catalase, responsible for hydrogen peroxide resistance of germinating spores, *katX* mutation sensitive to hydrogen	4009	+
katX	peroxide during out-growth (547 residues)	3964	+

TABLE 1 (continued)

V. GENES CODING FOR PROTEINS SIMILAR TO UNKNOWN PROTEINS, 668 genes:

- from *B. subtilis*: 177 genes
- from other organisms: 491 genes

VI. GENES CODING FOR PROTEINS WITH NO SIMILARITY TO ANY KNOWN PROTEIN:

1,053 genes

---

[a] Genes also present in *B. anthracis* genome, strain A2012 **(+ : present, - : absent, ? : undetermined)** (Huang and Jedrzejas, 2003). The *B. anthracis* sequence data were obtained from the NCBI website www.ncbi.nlm.nih.gov under microbial genomes and were searched for similarity using the sequences of *B. subtilis* proteins as queries (described above). The majority of homologous proteins resulted in the sequence identity of 40% or higher.

N-acetylmuramoyl-L-alanine amidase most likely associated with the mother cell wall and is responsible for lysis of the mother cell (Smith and Foster, 1995). CwlD is germination-specific lytic enzyme (Sekiguchi *et al*., 1995). It bears high homology to other cell lytic enzymes known, whereas CwlJ is cell wall hydrolase acting in sporulation and/or germination (Ishikawa *et al*., 1998). CwlJ seems to require dipicolinic acid for activity (Paidhungat *et al*., 2001). Another spore hydrolytic enzyme, SleB, is included among enzyme related to germination (Table 1) (Moriyama *et al*., 1996), and does not seem to have dipicolinic acid requirement for its activity (Paidhungat *et al*., 2001). The DacB and DacF gene products both are penicillin-binding proteins and both seem to have significant homology to D,D-peptidase (Buchanan and Ling, 1992; Wu *et al*., 1992). The remaining four penicillin-binding proteins, PdbA, PdbB, PdbE, and PdbF, are involved in spore outgrowth, septation, spore cortex synthesis, and germination processes during the spore life cycle.

The cellular sensor/signal transduction genes are members of two-component sensory cascade (KinA, KinB, and KinC), which are involved in the initiation of sporulation. KinA, KinB, as well as KinC have significant homology to the transmitter class histidine kinases (Trach and Hoch, 1993; Antoniewski *et al.,* 1990; Perego *et al*., 1989). KinA has the ability of *in vitro* autophosphorylation (Burbulys *et al*., 1991). The cell division genes code for cell division proteins FtsA and FtsZ that are involved in sporulation by initiation of cell division and septum formation. Homologous enzymes to FtsA and FtsZ can be identified in *Escherichia coli*, and there they control initiation of cell division (Bell and Lutkenhaus, 1992; 1991; Bell *et al*., 1988). Mutation in *ftsA* and/or *ftsZ* causes a filamentous phenotype during vegetative growth and prevents sporulation.

Among the sporulation genes constituting the largest subgroup in this category, the ones encoding proteins involved in the initiation and regulation of sporulation, septum formation, the synthesis of cortex or coat polysaccharides, and finally coat proteins or proteins protecting DNA from damage, small acid-soluble proteins, stand out from the large number of macromolecules involved in this process. For example, the *cge* operon consists of the *cgeA*, *cgeB*, *cgeC*, *cgeD*, and *cgeE* genes. For instance, *cgeA* contains two cistrons that encode CgeAA and CgeAB proteins. *CgeB*, on the other hand, contains one cistron encoding one protein. Disruption of these genes leads to spores with altered surface properties (www1.rhbnc.ac.uk/biological-sciences/cutting/spo/spo_frames.html). Spore coat proteins (inner and outer) are coded by 22 *cot* genes. These form a highly crosslinked external protein layer of spores that protects spores against various external damaging effects. Disruption of *cot* genes does not seem to interfere with sporulation or germination processes but often leads to spores with modified surface and, for example, sensitivity to lysozyme degradation. Other groups of genes coding for spore coat proteins are *yraD*, *yraE*, *yraF*, *yraG*, as well as *yrbA*, *yrbB*, *yrbC,* and *ytaA* genes (Table 1). Spore coat polysaccharides are synthesized by an 11-gene operon coding for SpsA through SpsK proteins. These proteins have some similarity to other carbohydrate synthesis enzymes (Glaser *et al*., 1993). The proteins primarily protecting spore DNA from damage are termed small acid-soluble proteins (SASP). There are several types of SASP: α-, β-, α/β-, γ-SASP proteins. The γ-SASP proteins, however, do not bind to DNA, as α- and β-SASP do, and their exact functionality is not yet fully understood. SASPs are coded by genes *sspA* through *sspF* (Ponnuraj *et al*., 2000b; Connors *et al*., 1986; Mason and Setlow, 1986).

The proteins coded by genes involved in germination constitute the second largest subgroup and can be divided into the ones coding for germination response to germinants (products of *ger* genes), spore lytic enzymes (SleB, YkvT), or spore germination proteins. The products of *ger* genes are highly hydrophobic proteins often with multiple transmembrane segments and are involved in initiation of germination of spores in response to germinants. Germination response to L-alanine is coded for by the three-gene *gerA* operon coding for GerAA, GerAB, and GerAC proteins (Moir and Smith, 1990; Zuberi *et al*., 1987; Feher *et al*., 1985). Germination response to glucose, fructose, asparagines, and KCl is coded by two three-gene operons, *gerB* and *gerK* (Moir and Smith, 1990). GerD codes for germination response to L-alanine and to glucose, fructose, asparagines, KCl (Kemp *et al*., 1991; Moir and Smith, 1990). Germination protease, Gpr, degrading the DNA protein protective coat deserves to be highlighted. The enzyme, likely a novel glutamic acid protease (Jedrzejas, 2002a,b, Ponnuraj *et al*., 2000a, b), initiates degradation of small, acid-soluble spore proteins protecting DNA from damage in spores.

B. Intermediary Metabolism

Among the intermediary metabolism category comprising 742 genes only a very limited number of them, if any, are directly related to spores. The proteins encoded by these genes are, understandably, involved in all aspects of vegetative cell metabolism and the majority are not specific to spore-forming bacteria. However, some of these metabolic enzymes have features specific to spores such as cofactor-independent phosphoglycerate mutase (iPGM) (main glycolytic pathway subgroup). In *B. subtilis*, *B. stearothermophilus*, and likely in all *Bacillus* species, this enzyme's activity is strictly Mn^{2+}-related pH-sensitive because its regulation is essential for preserving the deposits of spore's energy in form of 3-phosphoglycerate (3-PGA). Decreasing the activity of the enzyme as the pH drops in the sporulation process facilitates the buildup of 3-PGA deposits that are utilized immediately after initiation of germination when pH rises inside the spores. This rise activates iPGM enzyme and allows for 3-PGA conversion to 2-phosphoglycerate (2-PGA) that facilitates further metabolic chemistry performed by other enzymes of the glycolytic pathway and resulting in an energy-rich molecule ATP (Jedrzejas, 2002a,b; Jedrzejas and Setlow, 2001; Jedrzejas, 2000; Jedrzejas *et al*., 2000a,b). Selected proteins of 'phosphate metabolism' subgroup may also be related to the spore to some extent. For example, PhoE is a *Bacillus* species broad specificity phosphatase that originally was annotated as cofactor-dependent phosphoglycerate mutase (*yhfR*) (Rigden *et al*., 2002a; 2001). The known substrates of this enzyme are AMP, fructose-6-phjosphate, ribose-5-phosphate, CMP, 3-PGA, *p*-nitrophenlyphosphate, or α-napthylphosphate.

C. Information Pathways

Information pathways genes comprise another category of 482 genes, and the selected ones involved in the regulation subgroup (totaling 213 identified genes) comprise only 12 genes related to spores. No other subgroups were identified with spore-related genes among this category. The protein products of these genes are primarily involved in the regulation of transcription of spore coat proteins, for example, or are response regulators for the initiation of sporulation. The members of a two-component response regulator for the initiation of sporulation, Spo0A and Spo0F, are part of this group. SpoIIID and SpoVT are transcription regulators of σ^E-/σ^K-, and σ^G-dependent genes. SplA is another transcription regulator; it regulates transcription of *splAB* spore photoproduct lyase operon involved in UV-related DNA repair that is initiated after germination. GerE is among this group as the protein is involved in regulation of expression of late spore genes. The crystal structure of this molecule is also available (Ducros *et al*., 2001). GerE regulates transcription involved in σ^K-dependent promoters and represses transcription from other selected promoters (Zheng and Losick, 1990). A mutation in *gerE* gene abolishes *cotB* and *cotC* transcription and compromises the transcription of *cotD* gene (Zheng *et al*., 1992). As a consequence, the spores produced without *gerE* involvement are sensitive to lysozyme and have a modified coat structure (Feng and Aronson, 1986).

D. Proteins with Other Functions and Unknown Proteins

The 'other functions' category (289 genes) has a very limited number of spore-related genes. The only identified genes are those that protein products are implicated in atypical conditions and detoxification. For the first group only 2 out of 72 genes code for proteins related to spores, and they take part in synthesis of spore coat polysaccharides. The detoxification group genes (68 genes) have only three genes that are spore related: three catalases (*katA*, *katB*, and *katX*) that are related to hydrogen peroxide sensitivity and resistance during sporulation or germination processes. Mutations in *katA* gene result in hydrogen peroxide sensitivity during sporulation, but *katX* mutations are sensitive to hydrogen peroxide only during the outgrowth stage of germinating spores (Bagyan *et al*., 1998).

Finally, there are remaining genes in *B. subtilis* that either code for proteins similar to unknown proteins or completely novel proteins with no similarity to any known protein. For these reasons the functions of proteins in these groups are not known, and only future research may provide some additional information for these genes and their protein products. The two groups encompass a total of 1721 genes, a surprisingly large number, comprising ~42% of total number of ~4100 of *B. subtilis* genome identified open reading frames (Kunst *et al*., 1997). Some of the proteins encoded by these genes might be related to spores, and therefore the information included in this

review will likely be amended in the future. Similarly, Table 1 containing the list of known spore genes will need to be updated as more information is obtained about the unknown proteins. A large number of gene encoded protein groups in *B. subtilis* were not included in this review because they are involved in aspects of cellular functionality related specifically to vegetative growth of *B. subtilis* cells and are not directly related to spores.

VI. STRUCTURAL INFORMATION ABOUT SPORE-RELATED PROTEINS

Only relatively few proteins of *Bacillus* genus involved in either spore formation or degradation have known three-dimensional structures (Table 2). The majority of these known structures are related specifically to the sporulation process (as opposed to germination), and these are predominantly structures of proteins involved in regulation such as response regulators, or σ-factors' related proteins. These proteins include *Bacillus* species SinI/SinR, Spo0A, and Spo0F response regulators, SpoIIAB anti-σ factor, as well as GerE transcriptional regulator. Another group of structures is formed from proteins related to spore polysaccharides (modification or synthesis) such as penicillin-binding protein 5, DacB (structure of *E. coli* homolog), and spore coat polysaccharide synthesis protein, SpsA (from *B. subtilis*). In addition, metabolic proteins have some representation among those with the three-dimensional structures, and they include cofactor-independent phosphoglycerate mutase, iPGM, broad specificity phosphatase-PhoE, subtilisin E and serine protease, AprE, and AprX, respectively. Finally, the structure of *Bacillus megaterium* germination protease, GPR, is available as the single representative of the group of proteins involved in the germination process. Overall, the structural information about proteins related to spores, their formation, or dissipation, is very limited at present. The advancement in the bioinformatics and structural genomics tools available for the analysis of protein sequences to identify homologous molecules or even to reliably predict their three-dimensional structures will likely change this situation in the near future. These advances are coupled with the advances in techniques for protein production, crystallization, and in experimental determination of structures by X-ray or NMR. The utilization of these methods has the ability to generate a significant number, if not all, of new and needed three-dimensional structures of spore proteins of spore-forming bacteria. Two examples of the application of both the bioinformatics sequence analyses and the experimental or modeled structure determination for proteins of *Bacillus* species are presented below in order to illustrate the potential utilization for such studies. Once more structures become available, they will generate additional structural templates of proteins or protein domains. The availability of such templates will make future structure prediction and protein structure analyses, including functional annotation, significantly easier and more reliable.

VII. EXAMPLES OF HOMOLOGY STUDIES, IMPROVED FUNCTIONAL ANNOTATION, AND THREE-DIMENSIONAL STRUCTURE DETERMINATION

A. Penicillin-Binding Protein-Related Factor A, PrfA, Is An Endonuclease Acting on DNA

The endonuclease penicillin-binding protein-related factor A, PrfA also known as RecU, of *B. subtilis* is encoded by a gene upstream of *ponA* in a two gene operon, with *ponA* encoding the major penicillin-binding protein of this bacterium, a class A penicillin-binding protein 1 (PBP1). This protein is transcribed predominantly during log-phase growth (Popham and Setlow, 1995). PBP1 was found to play an important role in cell division and this protein localizes to sites of cell division, in vegetative cells of *B. subtilis,* where it plays an important role in the formation of peptidoglycan in the division septum (Pedersen *et al.*, 1999). Due to *prfA* being in the same operon with *ponA*, the PrfA protein was assumed to be functionally linked with PBP1 and plays a role in cell wall synthesis as well. Mutation analysis of this two gene operon show that the mutation of *prfA* gene alone in *B. subtilis* causes a ~twofold decrease in cell growth rate, a severe defect in nucleoid segregation during cell division, and a *prfA*/*ponA* double mutant was found to severely limit cell growth (Pedersen and Setlow, 2000). The PBP1 protein was found to associate with the cell wall and membrane; however, the ~20 kDa recombinant PrfA was found to be a soluble protein when overexpressed in *Escherichia coli* (Kelly *et al.*, 2000). PBP1 was localized at division sites in

TABLE 2
Selected Identified Proteins of *Bacillus* species with Known Three-Dimensional Structures

Protein groups	Example of protein	Functional annotation	Availability of three-dimensional structure
I. CELL ENVELOPE AND CELLULAR PROCESSES			
Cell wall/ 93 genes	**DacB**, penicillin-binding protein 5	peptidoglycan synthesis of spore cortex	crystal structure of *E. coli* homolog (Davies *et al.*, 2001)
Cell division/ 21 genes	**FtsZ**, cell-division initiation protein	septum formation (cell-division)	crystal structure of *M. jannaschii* homolog (Lowe and Amos, 1998)
Sporulation/ 139 genes	**SinI**, antagonist of **SinR**	inhibition of sporulation	crystal structure of *B. subtilis* SinI- SinR complex (Lewis *et al.*, 1998)
	Spo0A, response regulator	regulator of sporulation	crystal structure of *B. stearothermophilus* of the revceiver/phosphoacceptor domain of Spo0A (Lewis *et al.*, 1999)
	Spo0F, response regulator	involved in many different signaling pathways, involved in initiation of sporulation	NMR structure of *B. subtilis* response regulator Spo0F (Feher *et al.*, 1997)
	SpoIIAA, cell fate determinant	functions as anti-σ factor antagonist	crystal structure of *B. sphaericus* SpoIIAA (Seavers *et al.*, 2001)
	SpoIIAB, anti-σ factor and serine kinase	anti-σ factor binding and negatively regulating σ^F also a serine kinase phosphorylating and inactivating anti-anti-σ SpoIIAA	crystal structure of *B. stearothermophilus* σ^F-SpoIIAB dimer (Compbell *et al.*, 2002)
	SpsA, spore coat polysaccharide synthesis	synthesis of glycosidic bond belonging to GT-2 family of inverting transferases	crystal structure of *B. subtilis* glycosyltransfer-ase SpsA (Tarbouriech *et al.*, 2001; Charnock and Davies, 1999)

TABLE 2 (continued)

Germination/ 44 genes	**Gpr**, spore protease	removal of DNA coat of spores	crystal structure of *B. megaterium* GPR zymogen and mode of active protease (Ponnuraj *et al.*, 2000b)
II. INTERMEDIARY METABOLISM			
Main glycolytic pathway/ 28 genes	**iPGM**, cofactor independent phosphoglycerate mutase	isomerisation of 3- and 2-phosphoglyceric acid	crystal structure of *B. stearothermophilus* iPGM (Jedrzejas *et al.*, 2000a, b)
	PhoE (YhfR), broad specificity phosphatase	removal of phosphate group from a variety of substrates	crystal structure and model of *B. stearothermophilus* PhoE (Rigden *et al.*, 2002a, 2001)
Metabolism of amino acids and related molec./ 205 genes	**AprE** and **AprX**, extracellular (subtilisin E) and intracellular alkaline serine protease	serine protease	crystal structure of *B. amyloliquefaciens* subtilisin (Bott *et al.*, 1988; Wright *et al.*, 1969)
III. INFORMATION PATHWAYS			
Regulation/ 213 genes	**GerE**, transcriptional regulator of spore formation	activator or repressor of gene expression together with σ^K	crystal structure of *B. subtilis* GerE (Ducros *et al.*, 2001)

vegetative cells of *B. subtilis* (Pedersen *et al.*, 1999). Overexpression in *E. coli* cells revealed that PrfA was localized in the nucleoid (Pedersen and Setlow, 2000). Furthermore, additional mutation data surprisingly showed the requirement of *prfA* for DNA repair and recombination in *B. subtilis* (Fernandez *et al.*, 1998). These effects of the *prfA* mutation on nucleoid morphology are similar to those of mutations inactivating proteins that are involved in chromosome segregation and/or chromosome condensation (Britton *et al.*, 1998, Moriya *et al.*, 1998) making PrfA involvement in DNA recombination a distinct possibility.

In order to distinguish between the two possibilities, PrfA involvement in cell wall synthesis or in recombination, bioinformatics and structural analysis were applied. Using a combination of sequence analysis, structural fold recognition, and three-dimensional model construction, a distant homology between *Bacillus* species PrfA protein and the *Proteus vulgaris* PvuII restriction enzyme (Cheung *et al.*, 1995, 1994) has been identified (Rigden *et al.*, 2002b).

Briefly, *Bacillus* species PrfA sequence showed no significant sequence similarity to any other protein (Rigden *et al.*, 2002b). Sequence identity of PrfA proteins from different organisms, orthologs, is at best ~36%. Therefore, sequence-based methods failed to identify significant proteins that would facilitate functional assignment. These methods also failed in identifying similarity to proteins of known structure. However, threading methods (Fischer *et al.*, 1999) and BLAST profile comparisons (Rychlewski *et al.*, 2000) identified PrfA's structural relationship with the PvuII enzyme. Based on this homology, a model for PrfA was constructed (Plate 1*) with the help and validation of statistical structural analysis with PROSA II (Sippl, 1993a,b). The detailed analysis of such a three-dimensional model, especially the analysis of its geometric and structural properties, shape, and electrostatic characteristics, suggested that PrfA should share with PvuII enzyme its DNA binding ability and nuclease activity (Plate 1B) (Rigden *et al.*, 2002b). Such analysis even allowed for the identification of possible amino acid residues likely involved in PrfA's activity. Based on these results, biochemical assays were carried out that confirmed both the ability of PrfA to bind DNA and its nuclease activity (Rigden *et al.*, 2002b). The mutation of selected residues suspected in PrfA activity rendered the enzyme inactive.

* Plate 1 appears following page 193.

These results might provide a paradigm for other proteins of *B. subtilis* or *Bacillus* species in general for determining reliable three-dimensional models and model-based function determination, examination of function, and revisitation of, often incorrect, functional annotation. In the case of spore-related proteins of *Bacillus* species that lack a significant number or a database of experimentally determined structures (Table 2), and that lack obvious sequence similarity to any known sequence or experimental structure does not necessarily prevent their detailed analysis. Such analysis may include reliable structural predictions and a reliable determination of function. Furthermore, the increase in the number of structures available or the development of a public database of experimental (X-ray or NMR) three-dimensional structures would make such a process significantly easier and more reliable. It is clearly evident that there is a necessity for a concerted effort to increase the number of spore protein structures available in order to provide significantly larger number of functional and structural templates to facilitate structure and function determination. Recent global efforts in genome-based structure determinations (structural genomics) will most likely facilitate such result.

B. Putative YhfR Phosphoglycerate Mutase Homolog Is Actually a Broad Specificity Phosphatase, PhoE

Another example of bioinformatics and modeling type analysis that facilitate functional and structural determinations involves protein with significant homology to more than one enzyme, a broad specificity phosphatase PhoE from *Bacillus* species. This enzyme was previously missannotated as a cofactor (2,3-phosphoglycerate) dependent phosphoglycerate mutase (PGM). The distribution of phosphoglycerate mutases in bacterial organisms is complex because numerous bacteria seem to have more than one PGM enzyme present within their cell, sometimes even as many as three PGM enzymes (i.e., in *E. coli*). In most cases, some bacteria have only a cofactor-dependent (dPGM), others only a cofactor-independent PGM (iPGM), and some organisms, usually those with larger genomes, seem to have both of these enzymes (Jedrzejas, 2002a,b, 2000). *Bacillus* species, for example, contain only a cofactor-independent PGM. The three-dimensional X-ray structures for both groups

of PGMs are available and are representative of these groups. The iPGM enzyme structure is from *B. stearothermophilus* (Jedrzejas *et al*., 2000a,b), and the dPGM structure is available from *E. coli* (Bond *et al*., 2002, 2001). In *B. subtilis* as well as other members of this genus, a protein homologous to cofactor-dependent PGMs has been identified by rudimentary sequence homology analysis. This dPGM protein homolog coded in *B. subtilis* by the *yhfR* gene does not have any PGM activity (Rigden *et al*., 2001), and this lack of activity raised a question of the true function of this protein. This example shows how sequence analysis coupled with careful molecular modeling of this homologous protein demonstrated that this enzyme is actually not a mutase but a phosphatase with broad substrate specificity.

In order to elucidate the real function of YhfR, sequence analysis and careful molecular modeling was performed as described in more detail below. Searching the databases for sequence homologies with YhfR showed significant similarity to dPGMs, fructose-2,6-bisphosphatases (F26Bpase), and α-ribazole-5-phosphate phosphatases (R5PPases) family of enzymes (Jedrzejas, 2000). For the first two of the homologous molecules, the experimental three-dimensional structures are available. This group of enzymes constitutes a dPGM superfamily because they all share similar α/β/α core fold motif so characteristic of even other than PGM glycolytic proteins (Jedrzejas, 2000). However, YhfR showed no preferred homology to any of the molecules identified by sequence searches. Different sequence-based methods of the identification of the homologous molecules yielded different top scoring, and therefore preferential, molecules. The phylogenetic analysis of YhfR showed, however, that this molecule is clearly different from any of the three identified main groups of enzymes: dPGMs, F26BPases, and R5PPases (Rigden *et al*., 2001).

Therefore, in addition to sequence homology, more extensive searches and sensitive analyses were engaged that utilize the identification of similarities in fold and in three-dimensional structure. As was the case for sequence homology analyses, the structural searches also identified similar molecules to those recognized by sequence analyses and did not identify a closest or a preferential structure or a structural template. However, the utilization of two structural templates identified, dPGM and F26Bpase, facilitated the determination of model structure (Bates and Sternberg, 1999). Using statistical structural analysis with PROSA II, the templates were analyzed to determine an appropriate single template at various positions in the molecule's sequence: dPGM or F26Bpase. Such careful model building led to the construction of a final three-dimensional model structure of the YhfR molecule (Plate 2*).

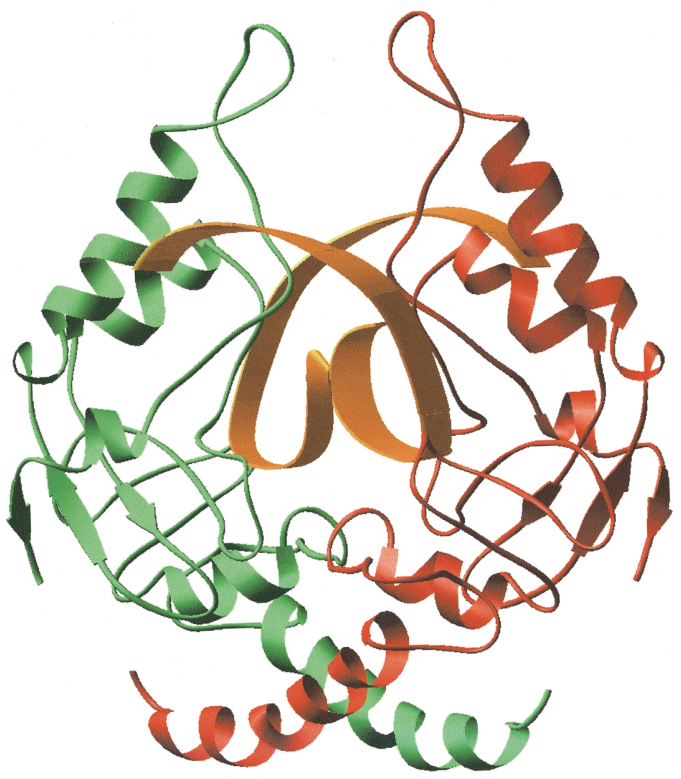
Comprehensive comparison of the YhfR structural model with the structures of *E. coli* dPGM (Bond *et al*., 2002, 2001) and rat testis F26Bpase (Yuen *et al*., 1999; Hasemann *et al*., 1996) showed that YhfR shares a slightly closer relationship with F26Bpase than with dPGM (Rigden *et al*., 2002a; 2001). As a consequence, conserved catalytic machinery was identified in YhfR that was similar to that of F26BPase but with a modified binding site for substrates (Rigden *et al*., 2003). When compared with F26BPase, the active site of the YhfR model was larger and more open to, for example, facilitate easier access of especially large substrates. An analysis of such an active site resulted in the suggestion that the enzyme would not have the mutase activity as the substrate reorientation necessarily associated with the mutase activity (Rigden *et al*., 1999) seemed impossible. The neutral and hydrophobic nature of the binding site of YhfR implied that a variety of even large hydrophobic substrates would be able to bind to the enzyme and be catalyzed by it. Therefore, the detailed structural analyses suggested that the YhfR might have a monophosphatase activity instead of mutase activity that involves phosphatase and substrate reorientation activities (Plate 2B).

The assays for phosphatase activity performed as a consequence of modeling studies clearly confirmed that *Bacillus* species YhfR is a phosphatase not a mutase. Such activity was identified for a variety of substrates like nucleoside monophosphates, 3-PGA, sugar phosphates, and two aromatic phosphomonoesters (Rigden *et al*., 2001). The experimental results support the model-based predictions of function/activity exceptionally well. As a result of this study, *Bacillus* species YhfR was confirmed to be a broad specificity phosphatase and was renamed PhoE. The experimental three-dimensional X-ray crystal structure of the enzyme was recently determined (Plate 3*). The structure shows remarkable similarity to the model (Plates 3 and 4) (Rigden *et al*., 2003).

The functional annotations of bacterial PGM homologs, especially dPGM homologs, must be carefully examined and confirmed experimentally because many of them are likely to possess phosphatase activ-

* Plates appears following page 193.

A



B

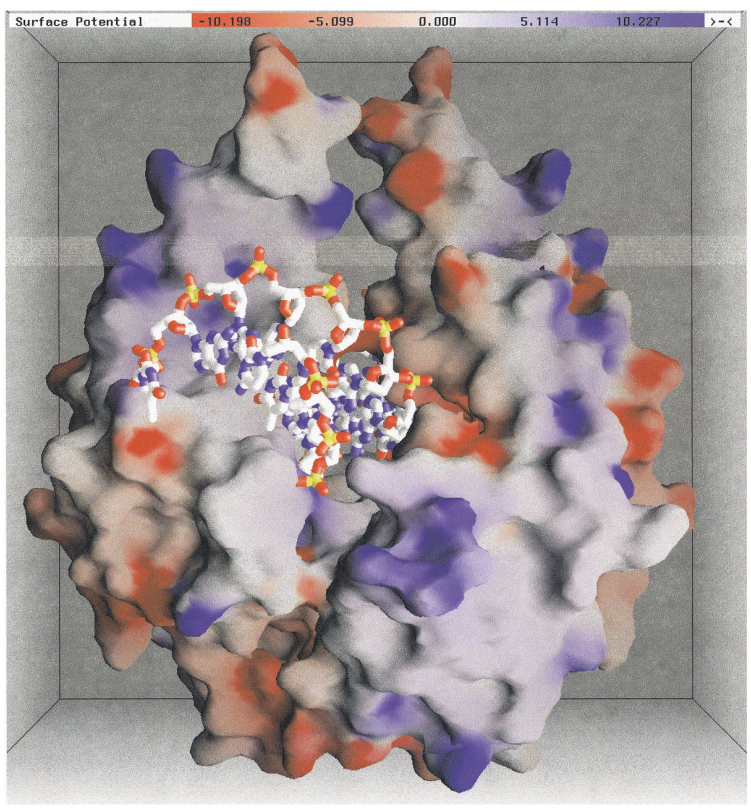
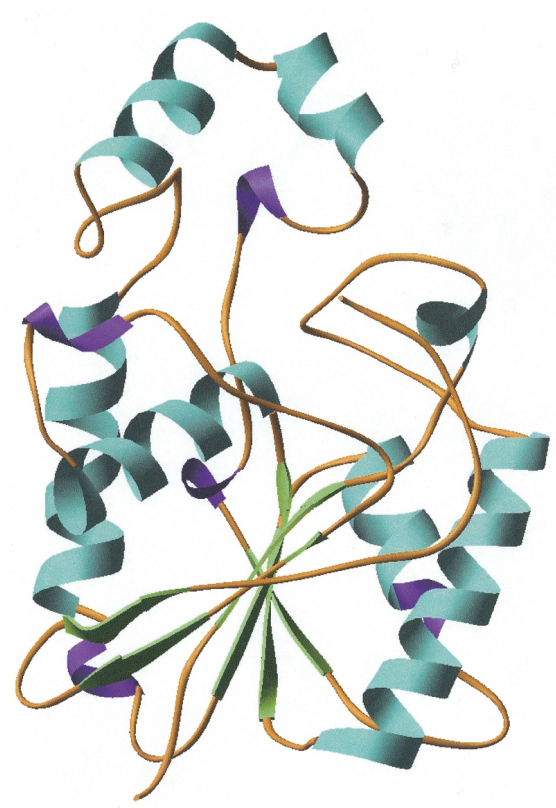


PLATE 1. Model three-dimensional structure of *B. stearothermophilus* PrfA. (A) A model of the three-dimensional structure of PrfA dimer with one subunit shown as a red ribbon structure and the other as a green with bound DNA (in gold) is shown (Rigden *et al.*, 2002b). The deep crevice in the middle of a dimer is reminiscent of a DNA binding crevice. Such a crevice is characteristic of other nucleases' structures and the suggested sliding mechanism of acting on DNA (Rigden *et al.*, 2002b; Breyer and Matthews, 2001). **(B)** Overall distribution electrostatic of potential. Positive potential is shown in blue and negative potential in red. The units of the scale are *kT* where *k* is the Boltzmann constant and *T* is temperature (scale at the top of the figure). The catalytic site and DNA binding cleft are shown in the center of the dimer as indicated by a bound DNA molecule shown in ball and stick fashion with bonds and atoms colored by the atom type (carbon atoms in green, oxygen atoms in red, nitrogen atoms in blue, and phosphate ions in yellow).

A



B

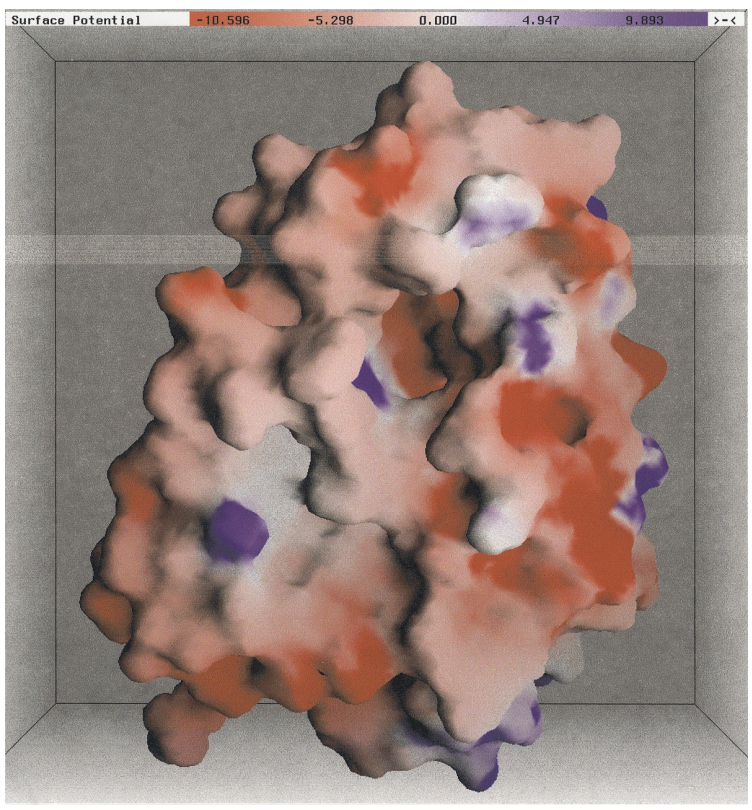
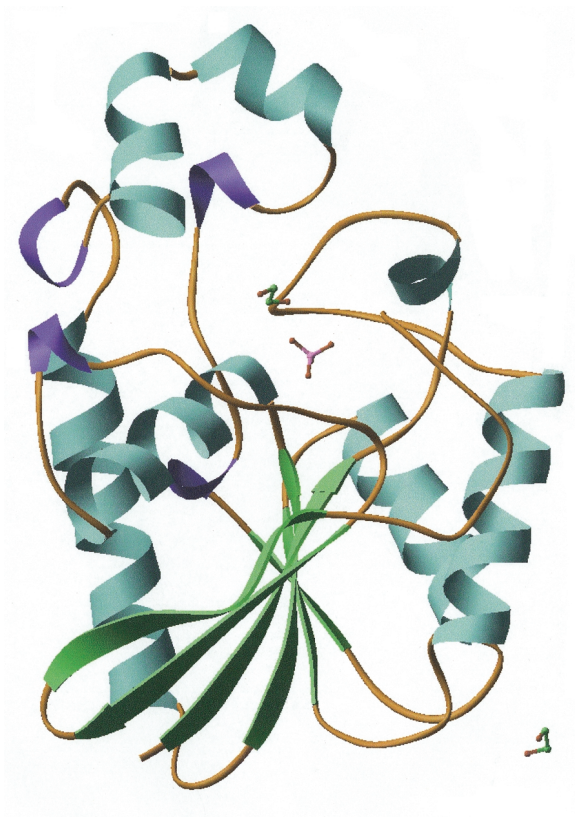


PLATE 2. Three-dimensional structure of *B. stearothermophilus* PhoE model. (A) A model of the three-dimensional structure of the phosphatase is shown as ribbon (Rigden *et al.*, 2001). The ribbon is color coded by the secondary structure elements (α-helices in blue, 3_{10} helices in purple, β-sheets in green). The structure consists from two domains: a major one with characteristic of glycolytic enzymes α/β/α core domain (bottom) and an additional small domain built from two α-helices and one 3_{10} helix (top). **(B)** Overall distribution electrostatic of potential is shown as in Plate 1b. The substrate binding and the catalytic site are located in the center of the structure and in the deep opening of the structure between the two domains. The orientation of the molecule is similar to that in part A of the figure.

A



B

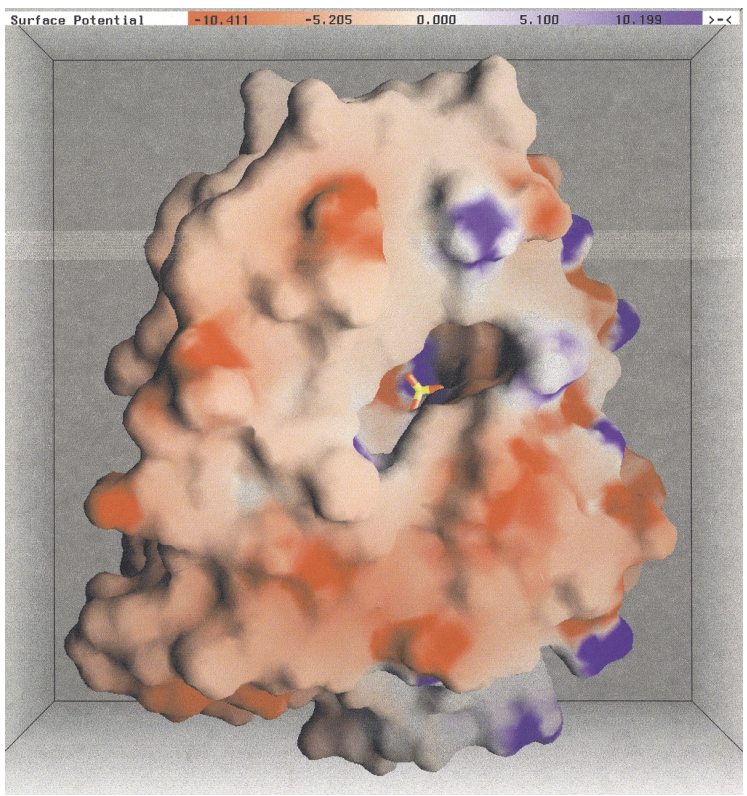


PLATE 3. Experimental, X-ray crystal structure of *B. stearothermophilus* PhoE. (A) An experimental structure of PhoE is shown in similar orientation as the model structure in Plate 2A,B (Rigden *et al.*, 2002a). The general arrangement of both domains, the arrangement of individual structural elements such as α-helices and β-sheets, are very similar to those predicted in the model. The X-ray structure confirmed the high accuracy of the earlier model of this broad specificity phosphatase and showed that careful modeling studies can be extremely useful in functional prediction(s) for proteins. Such functional predictions have to be, however, tested experimentally to avoid miss-assignment of functional annotation (Rigden *et al.*, 2001). The active site contains a bound phosphate ion. Two molecules of 1,2-ethanediol are also bound to the enzyme and shown. The bound molecules atoms of the substrate are depicted in a ball and stick fashion with bonds as in Plate 1B). **(B)** Overall distribution electrostatic of potential is shown as in Plate 1B and 2B. The substrate binding and the catalytic site are located in the center of the structure where a bound phosphate molecule is visible. The orientation of the molecule is similar to that in Plate 2B as well as part A of this plate.

ity instead of mutase. The majority of bacteria with more than one PGM enzyme identified should have the annotation of the putative PGM enzymes revisited, analyzed using bioinformatics and modeling tools available, and the results confirmed experimentally. The computational methods of sequence homology and structure-based similarity searches are developed enough to attempt making reliable predictions about protein structure and then function. Such predictions, however, always need to be subsequently confirmed by experiments. Also, three-dimensional modeling can facilitate functional determinations of proteins even if multiple possible structural templates are available and no obvious preferred similarities, either in sequence and/or in structure, are evident. Careful statistical analysis in these cases to choose a proper local structural template is necessary and has been proven successful.

VIII. OTHER MEMBERS OF *Bacillus* Species

It is unlikely that the proteins encoded by the *B. subtilis* genome would be specific only to *Bacillus* family and would not be expressed by other members of the *Bacillus* genus or other bacterial organisms. An analysis of genomic sequences from other finished and unfinished microbial genomes available at www.ncbi.nlm.nih.gov under Microbial Genomes allowed for the identification of multiple proteins that are similar to those of *B. subtilis* in a variety of other bacterial organisms. Such analysis highlighted molecules with significant sequence similarity even from vastly different and evolutionarily distant bacteria, and therefore these molecules are likely related in structure and function. The presence of molecules homologous to those of *B. subtilis* was especially evident in other spore-forming bacteria, primarily other members of *Bacillus* and *Clostridium* genera. Specific published examples of proteins for which homologs were identified in other *Bacillus* and *Clostridium* genera organisms include cofactor-independent phosphoglycerate mutase (Jedrzejas, 2002a,b), germination protease, and small acid-soluble proteins (Ponnuraj *et al*., 2000a, b, 1999), NAD^+ synthetase (Devedjiev *et al*., 2001), broad specificity phosphatase PhoE (Rigden *et al*., 2000a; 2001), and endonuclease penicillin-binding protein-related factor A — PrfA (Rigden *et al*., 2002b). Selected examples of these molecules have been discussed and analyzed in this study.

CONCLUSIONS

The identified spore-related proteins in the genome of *B. subtilis* were recognized based on the earlier biochemical, molecular biology, or microbiology studies performed and were also based on the homology searches of available databases (Kunst *et al*., 1997; Altschul, 1990). Additional proteins in *B. subtilis* when compared with non-spore-forming bacteria might have specialized functions specific to spores but may also be present in other bacterial organisms that do not have the ability to form spores. More spore proteins might be identified in the future; some of them will likely be recognized among the vast number of proteins coded for in the genome that still have unknown function(s).

The examples of sequence, functional, and structural analyses presented above show an excellent illustration of the exploration of genomic sequences through the sequence analysis and three-dimensional modeling. Such carefully performed studies can produce reliable structural and functional predictions for genome-encoded macromolecules. However, it seems unlikely that automatically generated model(s) and functional annotations using the distantly related sequence and/or structure-based templates could produce reliable results. A rigorous homology and modeling procedure must be applied in such instances. In both cases presented here, YhfR/PhoE and PrfA-endonuclease automatic sequence and/or structure analyses led to errors in functional annotation and would likely lead to errors in automated model structure determination. An improvement of methods is necessary for an easier and more reliable generation of structures or functional annotations for cases where no close homologous macromolecules are available or where several homologs are identified with different, sometimes contradictory, characteristics. Currently, close and highly proficient human intervention is necessary to accomplish three-dimensional homology structure modeling or functional annotations, especially in more difficult cases.

The increased number of structures and structural domains available for proteins, in this case for spore-related proteins, will greatly facilitate model determination and functional predictions. Therefore, the creation of larger databases of three-dimensional structures or domains is necessary. The number of experimental structures for proteins related to spores is very small, and significantly increased effort in the generation of such structures is clearly evident. The

availability of such database(s) in the future might even allow for the automation of a reliable model and function determination for molecules with unknown structures and either unknown or improperly annotated function(s). In *B. subtilis* ~42% of all proteins identified in the genome have unknown function, and therefore unknown structure (Table 1). Such a large number of uncharacterized proteins likely contains significant scientific information that warrants further investigation. The knowledge gained from such studies will increase our understanding of *B. subtilis*, all bacteria, and ultimately might contribute to better cures for microbial diseases and their prevention.

ACKNOWLEDGMENTS

The author would like to thank Dr. Adam Driks, Loyola University Medical Center, Maywood, IL, for providing the micrograph of *B. subtilis* spore. Preliminary sequence data for *B. anthracis*, strain A2012, were obtained from The Institute of Genomic Research available on website www.ncbi.nlm.nih.gov under microbial genomes. The *Bacillus* species-related studies in the author's laboratory are supported by the Defense Advanced Research Projects Agency contract N66001-01-C-8013 and National Institutes of Health grant AI 44079.

REFERENCES

Altschul, S.F. *et al.,* Basic local alignment search tool, *J. Mol. Biol.,* **215**, 403, 1990.

Antoniewski, C., Savelli, B., and Stragier, P., The *spoIIJ* gene, which regulates early developmental steps in *Bacillus subtilis*, belongs to a class of environmentally responsive genes, *J. Bacteriol.,* **172**, 86, 1990.

Bagyan, I., Casillas-Martinez, L., and Setlow, P., The *katX* gene, which codes for the catalase in spores of *Bacillus subtilis*, is a forespore-specific gene controlled by sigmaF, and KatX is essential for hydrogen peroxide resistance of the germinating spore, *J. Bacteriol.,* **180**, 2057, 1998.

Bates, P.A. and Sternberg, M.J., Model building by comparison at CASP3: Using expert knowledge and computer automation, *Proteins* **Suppl 3**, 47, 1999.

Beall, B., Lowe, M., and Lutkenhaus, J., Cloning and characterization of *Bacillus subtilis* homologs of *Escherichia coli* cell division genes *ftsZ* and *ftsA*, *J. Bacteriol.,* **170**, 4855, 1988.

Beall, B. and Lutkenhaus, J., FtsZ in *Bacillus subtilis* is required for vegetative septation and for asymmetric septation during sporulation, *Genes Dev.*, **5**, 447, 1991.

Beall, B. and Lutkenhaus, J., Impaired cell division and sporulation of a *Bacillus subtilis* strain with the *ftsA* gene deleted, *J. Bacteriol.,* **174**, 2398, 1992.

Bond, C.S., White, M.F., and Hunter, W.N., High resolution structure of the phosphohistidine-activated form of *Escherichia coli* cofactor-dependent phosphoglycerate mutase, *J. Biol. Chem.,* **276**, 3247, 2001.

Bond, C.S., White, M.F., and Hunter, W.N., Mechanistic implications for *Escherichia coli* cofactor-dependent phosphoglycerate mutase based on the high-resolution crystal structure of a vanadate complex, *J. Mol. Biol.,* **316**, 1071, 2002.

Bott, R., *et al.*, The three-dimensional structure of *Bacillus amyloliquefaciens* subtilisin at 1.8 Å and an analysis of the structural consequences of peroxide inactivation, *J. Biol. Chem.,* 263, 7895, 1988.

Breyer, W.A. and Matthews, B.W., A structural basis for processivity, *Protein Sci.,* **10**, 1699, 2001.

Britton, R.A., Lin, D.C., and Grossman, A.D., Characterization of a prokaryotic SMC protein involved in chromosome partitioning, *Genes Dev.,* **12**, 1254, 1998.

Buchanan, C.E. and Ling, M.L., Isolation and sequence analysis of *dacB*, which encodes a sporulation-specific penicillin-binding protein in *Bacillus subtilis*, *J. Bacteriol.,* **174**, 1717, 1992.

Burbulys, D., Trach, K.A., and Hoch, J.A., Initiation of sporulation in *B. subtilis* is controlled by a multicomponent phosphorelay, *Cell*, **64**, 545, 1991.

Campbell, E.A., *et al.*, Crystal structure of the *Bacillus stearothermophilus* anti-sigma factor SpoIIAB with the sporulation sigma factor *sigmaF*, *Cell,* **108**, 795, 2002.

Cano, R.J. and Borucki, M.K., Revival and identification of bacterial spores in 25- to 40-million-year-old Dominican amber, *Science*, **268**, 1060, 1995.

Charnock, S.J. and Davies, G.J., Structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms, *Biochemistry,* **38**, 6380, 1999.

Cheng, X., *et al.*, Structure of *PvuII* endonuclease with cognate DNA, *EMBO J.*, **13**, 3927, 1994.

Cheng, X., *et al.,* Crystal structure of the *Pvu*II restriction endonuclease, *Gene*, **157**, 139, 1995.

Connors, M. J., Mason, J. M., and Setlow, P., Cloning and nucleotide sequencing of genes for three small, acid-soluble proteins from *Bacillus subtilis* spores, *J. Bacteriol.,* **166**, 417, 1986.

Daffonchio, D., *et al.*, PCR fingerprinting of whole genomes: The spacers between the 16S and 23S

rRNA genes and of intergenic tRNA gene regions reveal a different intraspecific genomic variability of *Bacillus cereus* and *Bacillus licheniformis*, *Int. J. Syst. Bacteriol.,* **48**, 107, 1998.

Davies, C., White, S.W., and Nicholas, R.A., Crystal structure of a deacylation-defective mutant of penicillin-binding protein 5 at 2.3-Å resolution, *J. Biol. Chem.*, **276**, 616, 2001.

Devedjiev, Y., *et al*., Stabilization of active-site loops in NH3-dependent NAD^+ synthetase from *Bacillus subtilis*, *Acta. Crystallogr. D*., **57**, 806, 2001.

Driks A. *Bacillus subtilis* spore coat. *Microbiol. Mol. Biol. Rev*., **63**, 1, 1999.

Ducros, V.M., *et al*., Crystal structure of GerE, the ultimate transcriptional regulator of spore formation in *Bacillus subtilis, J. Mol. Biol.,* **306**, 759, 2001.

Feavers, I.M., Miles, J.S., and Moir, A., The nucleotide sequence of a spore germination gene (*gerA*) of *Bacillus subtilis* 168, *Gene*, **38**, 95, 1985.

Feher, V.A., *et al*., High-resolution NMR structure and backbone dynamics of the *Bacillus subtilis* response regulator, Spo0F: implications for phosphorylation and molecular recognition, *Biochemistry*, **36**, 10015, 1997.

Feng, P. and Aronson, A.I., Characterization of a *Bacillus subtilis* germination mutant with pleiotropic alterations in spore coat structure, *Curr. Microbiol*., **13**, 221, 1986.

Fernandez, S., Sorokin, A., and Alonso, J.C., Genetic recombination in *Bacillus subtilis* 168: effects of *recU* and *recS* mutations on DNA repair and homologous recombination, *J. Bacteriol.,* 1803405, 1998.

Fischer, D., *et al*., CAFASP-1: critical assessment of fully automated structure prediction methods, *Proteins* **Suppl. 3**, 209, 1999.

Glaser, P., *et al*., *Bacillus subtilis* genome project: cloning and sequencing of the 97 kb region from 325 degrees to 333 degrees, *Mol. Microbiol.,* **10**, 371, 1993.

Hanna, P.C. and Ireland, J.A., Understanding *Bacillus anthracis* pathogenesis, *Trends Microbiol*., **7**, 180, 1999.

Hasemann, C.A., *et al*., The crystal structure of the bifunctional enzyme 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase reveals distinct domain homologies, *Structure*, **4**, 1017, 1996.

Henderson, I., Duggleby, C.J., and Turnbull, P.C., Differentiation of *Bacillus anthracis* from other *Bacillus cereus* group bacteria with the PCR, *Int. J. Syst. Bacteriol.,* **44**, 99, 1994.

Huang, W.J.M. and Jedrzejas, M.J., *Bacillus anthracis* germination genes. Submitted.

Ishikawa, S., Yamane, K., and Sekiguchi, J., Regulation and characterization of a newly deduced cell wall hydrolase gene (*cwlJ*) which affects germination of *Bacillus subtilis* spores, *J. Bacteriol.,* **180**, 1375, 1998.

Ivins, B.E. and Welkos, S.L., Recent advances in the development of an improved, human anthrax vaccine, *Eur. J. Epidemiol*., **4**, 12, 1988.

Ivins, B.E., *et al*., Immunization against anthrax with *Bacillus anthracis* protective antigen combined with adjuvants, *Infect. Immun*., **60**, 662, 1992.

Jedrzejas, M.J., *et al.,* Structure and mechanism of action of a novel phosphoglycerate mutase from *Bacillus stearothermophilus*, *EMBO J.,* **19**, 1419, 2000a.

Jedrzejas, M.J., *et al*., Mechanism of catalysis of the cofactor-independent phosphoglycerate mutase from *Bacillus stearothermophilus*. Crystal structure of the complex with 2-phosphoglycerate, *J. Biol. Chem*., **275**, 23146, 2000b.

Jedrzejas, M.J., Structure, function, and evolution of phosphoglycerate mutases: comparison with fructose-2,6-bisphosphatase, acid phosphatase, and alkaline phosphatase, *Prog. Biophys. Mol. Biol*., **73**, 263, 2000.

Jedrzejas, M.J. and Setlow, P., Comparison of the binuclear metalloenzymes diphosphoglycerate-independent phosphoglycerate mutase and alkaline phosphatase: their mechanism of catalysis via a phosphoserine intermediate, *Chem. Rev*., **101**, 607, 2001.

Jedrzejas, M.J., Three-dimensional structure and molecular mechanism of novel enzymes of spore-forming bacteria, *Med. Sci. Monit*., **8**, RA183, 2002a.

Jedrzejas, M., The structure and function of novel enzymes of *Bacillus anthracis* and other spore-forming bacteria: development of novel prophylactic and therapeutic agents, *Crit. Rev. Biochem. Mol. Biol.,* **37**, 339–373, 2002b.

Kelly, S.J., et al., Structural characterization of penicillin-binding protein-related factor A (PrfA) from *Bacillus* species, *J. Struct. Biol.,* **131**, 90, 2000.

Kemp, E.H., *et al*., Analysis of transcriptional control of the *gerD* spore germination gene of *Bacillus subtilis* 168, *J. Bacteriol.,* **173**, 4646, 1991.

Kennedy, M.J., Reader, S.L., and Swierczynski, L.M., Preservation records of micro-organisms: evidence of the tenacity of life, *Microbiology*, **140**, 2513, 1994.

Kunst, F., *et al*., The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*, *Nature*, **390**, 249, 1997.

LaForce, F.M., Anthrax, *Clin. Infect. Dis.,* **19**, 1009, 1994.

Lalitha, M.K. and Thomas, M.K., Penicillin resistance in *Bacillus anthracis*, *Lancet,* **349**, 1522, 1997.

Lewis, R.J., *et al.,* An evolutionary link between sporulation and prophage induction in the structure of a repressor:anti-repressor complex, *J. Mol. Biol.,* **283**, 907, 1998.

Lewis, R.J., *et al*., Phosphorylated aspartate in the structure of a response regulator protein, *J. Mol. Biol.,* **294**, 9, 1999.

Lowe, J. and Amos, L.A., Crystal structure of the bacterial cell-division protein FtsZ, *Nature*, **391**, 203, 1998.

Mason, J.M. and Setlow, P., Essential role of small, acid-soluble spore proteins in resistance of *Bacillus subtilis* spores to UV light, *J. Bacteriol.,* **167**, 174, 1986.

Meijer, W.J., *et al*., The endogenous *Bacillus subtilis* (natto) plasmids pTA1015 and pTA1040 contain signal peptidase-encoding genes: identification of a new structural module on cryptic plasmids, *Mol. Microbiol.,* **17**, 621, 1995.

Meselson, M., *et al*., The Sverdlovsk anthrax outbreak of 1979, *Science*, **266**, 1202, 1994.

Moir, A. and Smith, D.A., The genetics of bacterial spore germination, *Annu. Rev. Microbiol.,* **44**, 531, 1990.

Moriya, S., *et al*., A *Bacillus subtilis* gene-encoding protein homologous to eukaryotic SMC motor protein is necessary for chromosome partition, *Mol. Microbiol*., **29**, 179, 1998.

Moriyama, R., *et al*., A gene (*sleB*) encoding a spore cortex-lytic enzyme from *Bacillus subtilis* and response of the enzyme to L-alanine-mediated germination, *J. Bacteriol*., **178**, 6059, 1996.

Nazina, T.N., *et al*., Taxonomic study of aerobic thermophilic bacilli: Descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and Geobacillus uzenensis sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus, Bacillus thermocatenulatus, Bacillus thermoleovorans, Bacillus kaustophilus, Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *Int. J. Syst. Evol. Microbiol*., **51**, 433, 2001.

Okinaka, R.T., *et al*., Sequence and organization of pXO1, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes, *J. Bacteriol*., **181**, 6509, 1999a.

Okinaka, R., *et al*., Sequence, assembly and analysis of pX01 and pX02, *J. Appl. Microbiol.,* **87**, 261, 1999b.

Paidhungat, M., Ragkousi, K., and Setlow, P., Genetic requirements for induction of germination of spores of *Bacillus subtilis* by Ca(2+)-dipicolinate, *J. Bacteriol*., **183**, 4886, 2001.

Pedersen, L.B., Angert, E.R., and Setlow, P., Septal localization of penicillin-binding protein 1 in *Bacillus subtilis*, *J. Bacteriol.,* **181**, 3201, 1999.

Pedersen, L.B. and Setlow, P., Penicillin-binding protein-related factor A is required for proper chromosome segregation in *Bacillus subtilis, J. Bacteriol.,* **182**, 1650, 2000.

Perego, M., *et al*., Characterization of the gene for a protein kinase which phosphorylates the sporulation-regulatory proteins Spo0A and Spo0F of *Bacillus subtilis*, *J. Bacteriol*., **171**, 6187, 1989.

Ponnuraj, K., Nessi, C., Setlow, P., and Jedrzejas, M. J., Structural studies of a novel germination protease from spores of *Bacillus megaterium*, *J. Struct. Biol.,* **125**, 19, 1999.

Ponnuraj, K., *et al*., Crystallization and preliminary diffraction studies of a truncated form of a novel protease from spores of *Bacillus megaterium*, *Acta Crystallogr. D*., **56**, 70, 2000a.

Ponnuraj, K., *et al*., Crystal structure of a novel germination protease from spores of *Bacillus megaterium*: structural arrangement and zymogen activation, *J. Mol. Biol.,* **300**, 1, 2000b.

Popham, D.L. and Setlow, P., Cloning, nucleotide sequence, and mutagenesis of the *Bacillus subtilis ponA* operon, which codes for penicillin-binding protein (PBP) 1 and a PBP-related factor, *J. Bacteriol.,* **177**, 326, 1995.

Read, T.D., *et al*., Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*, *Science*, **296**, 2028, 2002.

Rigden, D.J., *et al.,* Sulfate ions observed in the 2.12 Å structure of a new crystal form of *S. cerevisiae* phosphoglycerate mutase provide insights into understanding the catalytic mechanism, *J. Mol. Biol.,* **286**, 1507, 1999.

Rigden, D.J., *et al.,* A cofactor-dependent phosphoglycerate mutase homolog from *Bacillus stearothermophilus* is actually a broad specificity phosphatase, *Protein Sci*., **10**, 1835, 2001.

Rigden, D.J., *et al*., Structure and mechanism of action of a cofactor-dependent phosphoglycerate mutase homolog from *Bacillus stearothermophilus* with broad specificity phosphatase activity, *J. Mol. Biol.,* **315**, 1129, 2002a.

Rigden, D.J., *et al*., PrfA protein of *Bacillus* species: prediction and demonstration of endonuclease activity on DNA, *Protein Sci.,* **11**, 2370, 2002b.

Rigden *et al*., Structures of phosphate and trivanadate complexes of *Bacillus stearothermophilus* phosphatase PhoE: structural and functional analysis in

the cofactor-dependent phosphoglycerate mutase superfamily, *J. Mol. Biol.*, **325**, 411, 2003.

Rychlewski, L., *et al.*, Comparison of sequence profiles. Strategies for structural predictions using sequence information, *Protein Sci.,* **9**, 232, 2000.

Seavers, P.R., *et al.*, Structure of the *Bacillus* cell fate determinant SpoIIAA in phosphorylated and unphosphorylated forms, *Structure,* **9**, 605, 2001.

Sekiguchi, J., *et al.*, Nucleotide sequence and regulation of a new putative cell wall hydrolase gene, *cwlD*, which affects germination in *Bacillus subtilis*, *J. Bacteriol.,* **177**, 5582, 1995.

Sippl, M.J., Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures, *J. Comput. Aided Mol. Des.,* **7**, 473, 1993a.

Sippl, M.J., Recognition of errors in three-dimensional structures of proteins, *Proteins*, **17**, 355, 1993b.

Slepecky, R.A., What is *Bacillus*?, in *Biology of Bacilli*, Doi, R.H. and McGloughlin, M., Eds., Butterworth-Heinemann Press, 1992, pp. 1–21.

Smith, T.J. and Foster, S.J., Characterization of the involvement of two compensatory autolysins in mother cell lysis during sporulation of *Bacillus subtilis* 168, *J. Bacteriol.,* **177**, 3855, 1995.

Takami, H., *et al.*, Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*, *Nucleic Acids Res.*, **28**, 4317, 2000.

Tarbouriech, N., Charnock, S.J., and Davies, G.J., Three-dimensional structures of the Mn and Mg dTDP complexes of the family GT-2 glycosyltransferase SpsA: a comparison with related NDP-sugar glycosyltransferases, *J. Mol. Biol.,* **314**, 655, 2001.

Thorne, C.B., *Bacillus anthracis*, in *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology, and Molecular Genetics*, Hoch, J.A., Losick, R., and Sonenshein, A.L., Eds., American Society of Microbiology, Washington, D.C., 1993, pp. 113–124.

Thorsted, P.B., *et al.*, Complete sequence of *Bacillus subtilis* plasmid p1414 and comparison with seven other plasmid types found in Russian soil isolates of *Bacillus subtilis*, *Plasmid,* **41**, 274, 1999.

Trach, K.A. and Hoch, J.A., Multisensory activation of the phosphorelay initiating sporulation in *Bacillus subtilis*: identification and sequence of the protein kinase of the alternate pathway, *Mol. Microbiol.*, **8**, 69, 1993.

Wright, C.S., Alden, R.A., and Kraut, J., Structure of subtilisin BPN' at 2.5 angstrom resolution, *Nature*, **221**, 235, 1969.

Wu, J.J., Schuch, R., and Piggot, P.J., Characterization of a *Bacillus subtilis* sporulation operon that includes genes for an RNA polymerase sigma factor and for a putative DD-carboxypeptidase, *J. Bacteriol.*, **174**, 4885, 1992.

Yuen, M.H., *et al.*, Crystal structure of the H256A mutant of rat testis fructose-6-phosphate,2-kinase/fructose-2,6-bisphosphatase. Fructose 6-phosphate in the active site leads to mechanisms for both mutant and wild type bisphosphatase activities, *J. Biol. Chem.*, **274**, 2176, 1999.

Zheng, L.B. and Losick, R., Cascade regulation of spore coat gene expression in *Bacillus subtilis*, *J. Mol. Biol.,* **212**, 645, 1990.

Zheng, L., *et al.*, Sporulation regulatory protein GerE from *Bacillus subtilis* binds to and can activate or repress transcription from promoters for mother-cell-specific genes, *J. Mol. Biol.*, **226**, 1037, 1992.

Zuberi, A.R., Moir, A., and Feavers, I.M., The nucleotide sequence and gene organization of the *gerA* spore germination operon of *Bacillus subtilis* 168, *Gene,* **51**, 1, 1987.