

# The Spo0A regulon of *Bacillus subtilis*

Virginie Molle,<sup>1†</sup> Masaya Fujita,<sup>1</sup> Shane T. Jensen,<sup>2</sup> Patrick Eichenberger,<sup>1</sup> José E. González-Pastor,<sup>1</sup> Jun S. Liu<sup>2</sup> and Richard Losick<sup>1\*</sup>

Departments of <sup>1</sup>Molecular and Cellular Biology and <sup>2</sup>Statistics, Harvard University, Cambridge, MA 02138, USA.

## Summary

The master regulator for entry into sporulation in *Bacillus subtilis* is the DNA-binding protein Spo0A, which has been found to influence, directly or indirectly, the expression of over 500 genes during the early stages of development. To search on a genome-wide basis for genes under the direct control of Spo0A, we used chromatin immunoprecipitation in combination with gene microarray analysis to identify regions of the chromosome at which an activated form of Spo0A binds *in vivo*. This information in combination with transcriptional profiling using gene microarrays, gel electrophoretic mobility shift assays, using the DNA-binding domain of Spo0A, and bioinformatics enabled us to assign 103 genes to the Spo0A regulon in addition to 18 previously known members. Thus, in total, 121 genes, which are organized as 30 single-gene units and 24 operons, are likely to be under the direct control of Spo0A. Forty of these genes are under the positive control of Spo0A, and 81 are under its negative control. Among newly identified members of the regulon with transcription that was stimulated by Spo0A are genes for metabolic enzymes and genes for efflux pumps. Among members with transcription that was inhibited by Spo0A are genes encoding components of the DNA replication machinery and genes that govern flagellum biosynthesis and chemotaxis. Also included in the regulon are many (25) genes with products that are direct or indirect regulators of gene transcription. Spo0A is a master regulator for sporulation, but many of its effects on the global pattern of gene transcription are likely to be mediated indirectly by regulatory genes under its control.

Accepted 9 September, 2003. \*For correspondence. E-mail losick@mc.b.harvard.edu; Tel. (+1) 617 495 4905; Fax (+1) 617 496 4642. †Present address: Institut de Biologie et Chimie des Protéines, Centre National de la Recherche Scientifique, Université de Lyon, Lyon, France.

## Introduction

Sporulation by the bacterium *Bacillus subtilis* is a multi-stage, developmental process that is responsible for the conversion of a growing cell into a dormant cell type known as the spore or endospore (Stragier and Losick, 1996; Piggot and Losick, 2002). The master regulator for entry into sporulation is the DNA-binding protein Spo0A (Hoch, 1993), which is a member of the response regulator family of transcription factors (Perego and Hoch, 2002). The activity of Spo0A is governed by a multicomponent phosphorelay, which consists of five histidine autokinases (KinA, KinB, KinC, KinD and KinE) and two phosphorelay proteins (Spo0F and Spo0B) (Jiang *et al.*, 2000). The kinases feed phosphoryl groups into the relay by phosphorylating Spo0F. Spo0F~P, in turn, transfers the phosphoryl groups to Spo0B, which, in the final step of the relay, phosphorylates Spo0A to create Spo0A~P (Burbulys *et al.*, 1991). The level of phosphorylation of Spo0A~P is also influenced by dedicated phosphatases that remove phosphoryl groups from Spo0F~P (e.g. RapA) and from Spo0A~P itself (Spo0E) (Perego *et al.*, 1994; Grossman, 1995; Perego and Hoch, 2002). Once activated by phosphorylation, the master regulator binds to a DNA sequence element known as the '0A-box', where it acts as both a repressor of certain vegetatively expressed genes (e.g. *abrB*; Perego *et al.*, 1988; Strauch *et al.*, 1989; Strauch and Hoch, 1993; Fujita and Sadaie, 1998) and an activator of genes involved in sporulation (for a review, see Piggot and Losick, 2002). Among the genes activated by Spo0A~P are those involved in remodelling the sister chromosomes of the sporulating cell into an 'axial filament' (Pogliano *et al.*, 2002; Ben-Yehuda *et al.*, 2003) and in the formation of an asymmetrically positioned (polar) septum that divides the developing cell into a small forespore compartment and a large mother cell compartment (Levin and Losick, 1996; Ben-Yehuda and Losick, 2002). Spo0A~P is also responsible for activating genes that lead to the appearance of the cell-specific regulatory proteins  $\sigma^F$  and  $\sigma^E$ , which act in the forespore and the mother cell respectively (Stragier and Losick, 1996; Piggot and Losick, 2002). Recent work indicates that Spo0A~P continues to function after the polar septum is formed, when it accumulates to high levels and directs transcription in the mother cell (Fujita and Losick, 2003).

Orthologues of *spo0A* have been identified in all endospore-forming bacteria (from the genera *Bacillus* and *Clostridium*) with genome sequences available. Interest-

ingly, however, pathways leading to the activation of Spo0A by phosphorylation differ between *Bacillus* and *Clostridium* in that orthologues of key components of the phosphorelay (i.e. Spo0B and Spo0F) are missing in the latter. It is therefore assumed that Spo0A is phosphorylated directly by a sensor kinase(s) in *Clostridium* rather than indirectly through a relay (Stephenson and Hoch, 2002; Stragier, 2002). Nevertheless, it is possible that many of the same genes are under Spo0A control in both genera and, hence, it would be of interest to examine the conservation of the Spo0A regulon members among endospore formers. It would also be interesting to search for orthologues of Spo0A-controlled genes in the two members of the *Listeria* genus for which genome sequences are available. *Listeria* is closely related to *Bacillus* (much more so than *Clostridium*), but is unable to sporulate and lacks an orthologue for *spo0A* and for many other sporulation genes as well.

The purpose of this investigation was to attempt to identify comprehensively genes that are likely to be under the direct control of Spo0A in *B. subtilis*. So far, 18 genes are known that are transcribed from Spo0A-controlled promoters, and these are organized in six single-gene units (*abrB*, *kinA*, *kinC*, *spo0A*, *spo0F* and *spolIE*) and four operons (*dLt*, *sin*, *spolIA* and *spolIG*) for a total of 10 transcription units. (Strauch *et al.*, 1990; 1992; 1993; Satola *et al.*, 1991; York *et al.*, 1992; Asayama *et al.*, 1995; Kobayashi *et al.*, 1995; Perego *et al.*, 1995; Fujita and Sadaie, 1998; Shafikhani *et al.*, 2002). However, a complete list of the members of the Spo0A regulon has not been available. Transcriptional profiling previously revealed over 500 genes with expression levels that were significantly influenced by Spo0A (but not by the later-acting sporulation regulatory protein  $\sigma^F$ ) at the start of sporulation, and over 200 genes were identified with expression that was stimulated or repressed in cells engineered to produce an activated form of Spo0A (Spo0A-Sad67; see below) during growth (Fawcett *et al.*, 2000). Clearly, Spo0A has a profound effect on the global pattern of gene expression. But many (indeed a high proportion) of the genes so identified could be indirect targets of the master regulator. For example, some of the previously observed transcriptional effects of Spo0A could result from the action of regulatory proteins (e.g. AbrB and SinI) produced under the control of Spo0A rather than Spo0A itself. Another important transcriptional regulator, the abundance of which is influenced by Spo0A, is the stationary phase sigma factor  $\sigma^H$ . Genome-wide analysis of the  $\sigma^H$  regulon (Britton *et al.*, 2002) has confirmed that the transcriptional profile of a  $\sigma^H$  mutant is similar to the transcriptional profile of a Spo0A mutant. However, computational searches for binding sites for each regulator in sequences upstream of  $\sigma^H$ - and Spo0A-regulated genes indicate that the lists of presumed direct targets of each

regulatory protein are largely different (Fawcett *et al.*, 2000; Britton *et al.*, 2002).

To catalogue genes that are likely to be under the direct control of Spo0A and to do so on a genome-wide basis, we took advantage of recently developed procedures for carrying out chromatin immunoprecipitation (ChIP) in conjunction with DNA microarray analysis (chip) in order to identify regions of the chromosome to which Spo0A binds *in vivo* (Ren *et al.*, 2000; Iyer *et al.*, 2001; see Laub *et al.*, 2002 and Molle *et al.*, 2003 for recent applications to DNA-binding proteins in bacteria). We have used this 'ChIP-on-chip' procedure in combination with transcriptional profiling and gel electrophoretic mobility shift assays to expand substantially the known members of the Spo0A regulon. Taking advantage of computational methods and our expanded list of Spo0A-controlled genes, we have updated the consensus recognition sequence for Spo0A (the 0A-box) and, by building null mutants, we have identified previously uncharacterized members of the regulon that influence sporulation.

## Results

### *Identification of DNA binding sites for Spo0A in vivo by ChIP-on-chip analysis*

Because Spo0A is normally subject to a complex pathway of regulation, we sought to simplify our analysis by bypassing the requirement for the phosphorelay in the activation of Spo0A. To do so, we took advantage of a previously described, constitutively active form of Spo0A called Spo0A-Sad67 that does not require phosphorylation in order to be active (Ireton *et al.*, 1993) (making the assumption that the spectrum of genes recognized by Spo0A-Sad67 is the same as that recognized by Spo0A-P). The ChIP-on-chip experiments were carried out using cells (strain RL1104) engineered to produce Spo0A-Sad67 during growth in response to IPTG. RL1104 cells that had been grown in the presence of inducer were treated with formaldehyde to cross-link proteins to DNA. DNA from the formaldehyde-treated cells was sheared, and complexes of Spo0A with DNA were then precipitated using anti-Spo0A antibodies. Next, after reversal of the cross-links, sheared DNA from the immunoprecipitation was amplified by polymerase chain reaction (PCR) using randomly annealed primers and fluorescently labelled with Cy5. As a reference, sheared DNA from the formaldehyde cross-linking that had not been subject to immunoprecipitation (representing bulk chromosomal DNA) was similarly amplified and labelled with Cy3. Finally, a mixture of the Cy5- and Cy3-labelled DNAs was annealed under hybridization conditions with DNA microarrays imprinted with PCR-amplified DNAs corresponding to the annotated open reading frames (ORFs)

in the *B. subtilis* genome. An enrichment factor was calculated that denotes the extent to which each gene was enriched by immunoprecipitation relative to the reference DNA.

Cells of strain RL1104 were induced for the synthesis of Spo0A-Sad67 in two independent experiments and, in each experiment, samples of cells were treated with cross-linker at 30 or 60 min after the addition of inducer.

This resulted in a total of four independent preparations of cross-linked DNAs, which were processed separately for chromatin immunoprecipitation and microarray analysis. The results in Table 1 are an average of the four ChIP-on-chip experiments. As summarized in Table 1, 132 PCR-amplified ORFs were identified as representing preferential sites of binding by Spo0A-Sad67 *in vivo*, corresponding to 64 regions of the chromosome (i.e. in most

**Table 1.** The Spo0A regulon.

Region detected <sup>a</sup>	Gene	Enrichment factor <sup>b</sup>	Gel shift assay <sup>c</sup>	Transcriptional ratio <sup>d</sup>	Function <sup>e</sup>
Category I: genes previously known to be directly regulated by Spo0A					
<b><i>spollGA-sigE</i></b>	<i>spollGA</i>	1.8	+	49	Protease (processing of pro- $\sigma^E$ to $\sigma^E$ )
	<i>sigE</i>	3.1		20	Mother cell-specific early $\sigma$ factor
<b><i>spolIE</i></b>	<i>spolIE</i>	1.3	+	40	Ser phosphatase ( $\sigma^F$ activation)/ asymmetric septum formation
<b><i>spolIAA-spolIAB-sigF</i></b>	<i>spolIAA</i>	3.1	+	9.6	Anti-anti- $\sigma$ factor (antagonist of SpolIAB)
	<i>spolIAB</i>			8.1	Anti- $\sigma$ factor (antagonist of $\sigma^F$ )/Ser kinase
	<i>sigF</i>			17	Forespore-specific early $\sigma$ factor
<b><i>kinA</i></b>	<i>kinA</i>	1.3	+	2.0	Sensor histidine kinase (initiation of sporulation)
<b><i>ywjG-spo0F</i></b>	<i>ywjG</i>	2.8			
	<i>spo0F</i>	3.6	+	(1.6)	Response regulator (multicomponent P-relay)
<b><i>yabC-abrB-metS</i></b>	<i>yabC</i>	2.7			
	<i>abrB</i>	8.5	+	0.02	Transcriptional pleiotropic regulator of transition state genes
	<i>metS</i>	4.1	+	0.55	See category III
<b><i>abh-kinC</i></b>	<i>abh</i>	4.4			
	<i>kinC</i>	1.8	+	0.37	Sensor histidine kinase (initiation of sporulation)
<b><i>dLtA-dLtB-dLtC-dLtD-dLtE</i></b>	<i>dLtA</i>	1.6	+	0.42	Dcl (D-alanyl-D-Ala carrier protein ligase)
	<i>dLtB</i>	1.2		0.44	D-Ala transfer from Dcp to undecaprenol-P
	<i>dLtC</i>			0.46	Dcp (D-Ala carrier protein)
	<i>dLtD</i>			0.56	D-Ala esterification of lipoteichoic acid
	<i>dLtE</i>			0.46	Similar to ketoacyl reductase
<b><i>sinI-sinR</i></b>	<i>sinI</i>	2.2	+	NDE	Antagonist of SinR
	<i>sinR</i>			0.57	Transcriptional regulator of post-exponential phase responses genes
<b><i>spo0A-spoIVB</i></b>	<i>spo0A</i>	1.5	+	NA	Transcriptional regulator central for the initiation of sporulation
	<i>spoIVB</i>	2.5			
Category II: newly identified genes under the positive control of Spo0A					
<b><i>yxBB-yxBC</i></b>	<i>yxBB</i>	1.4			
	<i>yxBC</i>	(1.8)	+	52	Unknown/cupin metalloenzyme-like domain
<b><i>skfA (ybcO) → skfH (ybdE)</i></b>	<i>skfA</i>	2.0	+	8.5	Sporulation killing factor A
	<i>skfB</i>			15	Similar to subtilosin (antibiotic production)
	<i>skfC</i>			22	Unknown
	<i>skfD</i>			20	Unknown
	<i>skfE</i>			20	Similar to ABC transporter (binding protein)
	<i>skfF</i>			21	Similar to ABC transporter (permease)
	<i>skfG</i>			13	Unknown
	<i>skfH</i>			7.3	Unknown
<b><i>sdpA (yvaW)-sdpB (yvaX)-sdpC (yvaY)</i></b>	<i>sdpA</i>	3.1	+	5.1	Sporulation-delaying protein A
	<i>sdpB</i>	3.4		19	Sporulation-delaying protein B
	<i>sdpC</i>			(2.7)	Sporulation-delaying protein C
<b><i>yjCM-yjCN</i></b>	<i>yjCM</i>	1.9	-	17	Similar to alcohol dehydrogenase and to coat-associated protein YhbB
	<i>yjCN</i>	1.8			
<b><i>yppD-yppE-yppF-yppG</i></b> <sup>g</sup>	<i>yppD</i>	3.8	+	13	Unknown
	<i>yppE</i>	3.9		4.3	Unknown
	<i>yppF</i>	5.5	+	NDE	See category IV
	<i>yppG</i>	3.9		NDE	

Table 1. cont.

Region detected <sup>a</sup>	Gene	Enrichment factor <sup>b</sup>	Gel shift assay <sup>c</sup>	Transcriptional ratio <sup>d</sup>	Function <sup>e</sup>
<b>yttP-ytsP</b>	<i>yttP</i>	4.4	+	8.2	Similar to transcriptional regulator (TetR/AcrR family). Probably controlled by $\sigma^{\text{Hh}}$
<b>ywkD-racA (ywkC)</b>	<i>ytsP</i>	3.6			
	<i>ywkD</i>	2.2			
	<i>racA</i>	1.5	+	6.6	Remodelling and anchoring of the chromosome to the cell pole. Probably controlled by $\sigma^{\text{Hi}}$
<b>yocl-yoch</b>	<i>yocl</i>	1.6			
	<i>yoch</i>	(1.7)	+	5.4	Similar to cell wall-binding protein (LysM domain)
<b>yybN-yybM</b>	<i>yybN</i>	2.7	-	5.1	Unknown
<b>tkt-yneE-yneF</b>	<i>yybM</i>	1.6			
	<i>tkt</i>	2.4	+	0.26	See category III
	<i>yneE</i>	(4.1)	+	3.6	Unknown
<b>yfml</b>	<i>yneF</i>	3.4			
	<i>yfml</i>	2.4	+	3.6	Similar to macrolide efflux transporter
<b>yusE-yusD</b>	<i>yusE</i>	2.8	+	3.2	Unknown
	<i>yusD</i>	4.8			
	<i>ycgL</i>	4.3			
<b>ycgM</b>	<i>ycgM</i>	3.2	+	2.5	Similar to proline oxidase
	<i>yqzI</i>	1.6			
<b>yqzI-yqcG-yqcF</b>	<i>yqcG</i>	(2.1)	+	2.5	Similar to transposase (located in <i>skin</i> element)
	<i>yqcF</i>			2.6	Unknown (located in <i>skin</i> element)
	<i>cwIA</i>	3.3			
<b>cwIA-yqxl</b>	<i>yqxl</i>	(1.9)	+	2.4	Unknown (located in <i>skin</i> element)
	<i>yerB</i>	2.2	+	2.2	Unknown.
<b>ytsJ-accD (yttI)-accA</b>	<i>ytsJ</i>	1.9			
	<i>accD</i>	1.8	+	1.8	Acetyl-CoA carboxylase (beta subunit) <sup>k</sup>
	<i>accA</i>			1.6	Acetyl-CoA carboxylase (alpha subunit) <sup>k</sup>
<b>ykuK-ykzF-ykuL-ccpC (ykuM)<sup>l</sup></b>	<i>ykuK</i>	3.3			
	<i>ykzF</i>	2.5	+	NDE	Unknown
	<i>ykuL</i>	2.7	+	1.4	Unknown
	<i>ccpC</i>	2.0			
Category III: newly identified genes under the negative control of Spo0A					
<b>yueK-yuxH-yuzC</b>	<i>yueK</i>	3.2			
	<i>yuxH</i>	3.1	+	0.06	Unknown/EAL domain (putative diguanylate phosphodiesterase)
<b>yqgL-yqzD-yqzC</b>	<i>yuzC</i>	1.5			
	<i>yqgL</i>	3.1			
	<i>yqzD</i>	3.9	+	0.14	Unknown
	<i>yqzC</i>	1.3		0.10	Similar to aminodeoxychorismate lyase
<b>yjaV-med</b>	<i>yjaV</i>	6.1			
	<i>med</i>	2.8	+	0.12	Positive regulator of <i>comK</i>
<b>purT-mpr</b>	<i>purT</i>	2.6	-	0.12	Phosphoribosylglycinamide formyltransferase 2
	<i>mpr</i>	1.3			
<b>yrzB-yrrL</b>	<i>yrzB</i>	3.1			
	<i>yrrL</i>	2.0	+	0.17	Similar to folate metabolism (aminodeoxychorismate lyase)
	<i>cccB</i>	4.0			
<b>cccB-ftsE-ftsX</b>	<i>ftsE</i>	2.2	+	0.20	Cell division protein (ATP binding)
	<i>ftsX</i>			0.18	Cell-division protein (ABC transporter/permease)
	<i>codY</i>	2.4			
<b>codY-flgB-flgC → ylxL<sup>m</sup></b>	<i>flgB</i>	2.2	+	0.20	<i>FlaIche</i> operon (>26 kb cluster of 30 genes); motility and chemotaxis
	<i>flgC</i>	1.4		0.18	
<b>phoA-lytE</b>	<i>phoA</i>	1.5			
	<i>lytE</i>	2.5	+	0.19	Cell wall hydrolase (major autolysin)
<b>ykbA-ykaA</b>	<i>ykbA</i>	1.4			
	<i>ykaA</i>	2.5	+	0.22	Unknown/DUF47 domain
<b>tkt-yneE-yneF<sup>i</sup></b>	<i>tkt</i>	2.4	+	0.26	Transketolase <sup>k</sup>
	<i>yneE</i>	(4.1)	+	3.6	See category II
	<i>yneF</i>	3.4			

Table 1. cont.

Region detected <sup>a</sup>	Gene	Enrichment factor <sup>b</sup>	Gel shift assay <sup>c</sup>	Transcriptional ratio <sup>d</sup>	Function <sup>e</sup>
<i>yknZ-fruR</i>	<i>yknZ</i>	1.3			
	<i>fruR</i>	2.6	+	0.36	Transcriptional repressor of the fructose operon
<i>yprB-cotD-ypsA</i> <sup>h</sup>	<i>yprB</i>	4.0			
	<i>cotD</i>	4.8	+	0.39	Spore coat protein (inner)
	<i>ypsA</i>	1.8	-	0.36	Unknown
<i>divIVA</i> <i>rocR-rocD-rocE-rocF</i>	<i>divIVA</i>	1.9	+	0.38	Cell-division protein (septum placement)
	<i>rocR</i>	2.6			
	<i>rocD</i>	1.9	+	0.39	Ornithine aminotransferase
	<i>rocE</i>			0.41	Amino acid permease
<i>degS-yvyE-yvhJ</i>	<i>rocF</i>			0.47	Arginase
	<i>degS</i>	2.8			
	<i>yvyE</i>	3.5	+	0.47	Similar to Xaa-Pro dipeptidase ( <i>B.cereus</i> )
	<i>yvhJ</i>	1.8		0.40	Similar to LytR (cell envelope-related transcriptional attenuator)
<i>nfrA-ywchP</i>	<i>nfrA</i>	(2.8)	+	NDE	FMN-containing NADPH-linked nitro/flavin reductase
	<i>ywchP</i>	1.6	+	0.42	Similar to monooxygenase
<i>yyaB-soj-spo0J</i>	<i>yyaB</i>	2.4			
	<i>soj</i>	2.0	+	0.45	Negative regulation of sporulation initiation
	<i>spo0J</i>	1.9		0.44	Antagonist of Soj
<i>ykuV-rok</i> <i>rpmH-dnaA-dnaN</i>	<i>ykuV</i>	2.5			
	<i>rok</i>	3.0	+	0.47	Repressor of <i>comK</i>
	<i>rpmH</i>	3.4			
	<i>dnaA</i>	2.3	+	(0.63)	Initiation of chromosome replication <sup>k</sup>
<i>apt-relA-yrvI</i>	<i>dnaN</i>	1.8		0.50	DNA polymerase III (beta subunit) <sup>k</sup>
	<i>apt</i>	1.4			
	<i>relA</i>	2.4	-	0.50	GTP pyrophosphokinase (stringent response)
	<i>yrvI</i>			0.67	Unknown
<i>yaaR-holB-yaaT</i>	<i>yaaR</i>	1.9			
	<i>holB</i>	3.5	-	0.55	DNA polymerase III (delta' subunit) <sup>k</sup>
	<i>yaaT</i>	2.6		(0.68)	Unknown
<i>yabC-abrB-metS</i> <i>yqxD-dnaG</i>	<i>yabC</i>	2.7			
	<i>abrB</i>	8.5	+	0.02	See category I
	<i>metS</i>	4.1	+	0.55	Methionyl-tRNA synthetase <sup>k</sup>
	<i>yqxD</i>	3.7			
<i>dacA-yaaD-yaaE</i>	<i>dnaG</i>	2.7	+	0.65	DNA primase <sup>k</sup>
	<i>dacA</i>	1.3			
	<i>yaaD</i>	(1.4)	+	0.68	Similar to superoxide-inducible protein
	<i>yaaE</i>	1.2		0.66	Similar to aminotransferase
<i>ylmC-ylmD</i> → <i>ylmH</i>	<i>ylmC</i>	4.6			
	<i>ylmD</i>	2.8	+	(0.52)	Unknown/DUF152 domain
	<i>ylmE</i>			0.69	Similar to N-terminal domain of Ala racemase
	<i>ylmF</i>			0.71	Unknown/DUF552 domain
	<i>ylmG</i>			0.72	Similar to integral membrane protein (YGGT family)
	<i>ylmH</i>			0.82	Similar to RNA binding protein (S4 family)
	<i>yjoB-rapA-phrA</i>	<i>yjoB</i>	2.6		
<i>phrA</i>	<i>rapA</i>	2.8	+	(0.39)	Asp phosphatase (dephosphorylation of Spo0F~P)
	<i>phrA</i>			0.75	Inhibitor of RapA
Category IV: genes whose transcription was not detectably influenced by Spo0A					
<i>comK</i>	<i>comK</i>	2.8	+	NDE	Competence transcription factor (CTF)
<i>degQ-comQ-comX</i>	<i>degQ</i>	1.6		NDE	Degradative enzyme production
	<i>comQ</i>	2.8	-	NDE	Transcriptional regulator (competence)
	<i>comX</i>	2.5		NDE	Competence pheromone precursor
<i>fliT-yvyD</i>	<i>fliT</i>	1.6		NDE	Flagellar protein
	<i>yvyD</i>	4.6	+	NDE	Similar to sigma-54 modulating factor of Gram-negative bacteria (controlled by $\sigma^B$ and $\sigma^H$ )
<i>murAA-spolID</i>	<i>murAA</i>	1.6		NDE	UDP-NAG 1-carboxyvinyltransferase
<i>sigV</i>	<i>spolID</i>	1.9	+	NDE	Dissolution of the septal cell wall
	<i>sigV</i>	2.5	-	NDE	RNA polymerase ECF-type $\sigma$ factor

Table 1. cont.

Region detected <sup>a</sup>	Gene	Enrichment factor <sup>b</sup>	Gel shift assay <sup>c</sup>	Transcriptional ratio <sup>d</sup>	Function <sup>e</sup>
<i>yabG-veg-sspF</i>	<i>yabG</i>	3.4		NDE	Protease (controlled by $\sigma^G$ )
	<i>veg</i>	2.8	+	NDE	Unknown (actively transcribed during growth)
	<i>sspF</i>	1.6		NDE	Small acid-soluble spore protein (controlled by $\sigma^G$ )
<i>yebG-purE-purK</i>	<i>yebG</i>	1.8		NDE	Unknown
	<i>purE</i>	3.7	-	NDE	Phosphoribosylaminoimidazole carboxylase I
	<i>purK</i>	1.9		NDE	Phosphoribosylaminoimidazole carboxylase II
<i>yhzB-ygaO</i>	<i>yhzB</i>	3.6		NDE	Unknown
	<i>ygaO</i>	3.6	+	NDE	Similar to RNA-dependent DNA polymerase ( <i>B. anthracis</i> )
<i>yjcO-yjcP-yjcQ</i>	<i>yjcO</i>	2.3		NDE	Unknown
	<i>yjcP</i>	2.1	+	NDE	Unknown
	<i>yjcQ</i>	2.8		NDE	Unknown
<i>yppD-yppE-yppF-yppG<sup>g</sup></i>	<i>yppD</i>	3.8	+	13	See category II
	<i>yppE</i>	3.9		4.3	
	<i>yppF</i>	5.5	+	NDE	Unknown
	<i>yppG</i>	3.9		NDE	Unknown
	<i>yrdA</i>	2.2		NDE	Unknown
<i>yrdA-yrdB-yrdC</i>	<i>yrdB</i>	4.8	-	NDE	Unknown
	<i>yrdC</i>	2.2		NDE	Similar to hydrolase (isochorismatase family)
	<i>yweA</i>	1.4		NDE	Unknown
<i>yweA-rocG (yweB)</i>	<i>rocG</i>	2.1	+	NDE	Glutamate dehydrogenase
	<i>ywqB</i>	3.0		NDE	Unknown
<i>ywqB-ywqC-ywqD</i>	<i>ywqC</i>	5.3	+	NDE	Similar to capsular polysaccharide biosynthesis
	<i>ywqD</i>	2.5		NDE	Similar to capsular polysaccharide biosynthesis

**a.** Gene regions that were preferentially precipitated by antibody to Spo0A, after *in vivo* cross-linking, as detected by hybridization to a genomic array. The likely target gene or operon is indicated in bold characters. The likely target gene was either the enriched gene or a neighbouring, promoter-proximal gene, transcription of which is strongly affected in an inducible Spo0A-Sad67 strain.

**b.** The enrichment factor indicates the extent to which a gene was preferentially precipitated compared with its abundance in total DNA. All enrichment factors have associated *P*-values < 0.01 as judged using the RESOLVER statistical package (Rosetta), except for enrichment factors indicated in parenthesis, where the corresponding gene was enriched in some but not all the experiments and thus has an associated *P*-value > 0.01.

**c.** A '+' indicates that a gel shift was observed. Results in category I represent data from the literature, whereas results for the remaining categories are from the present work.

**d.** All transcription ratios have associated *P*-values < 0.01 as judged using the RESOLVER statistical package (Rosetta), except for transcription ratios indicated in parenthesis, where the corresponding gene was enriched in some but not all the experiments and thus has an associated *P*-value > 0.01. Ratios are indicated for the likely target gene or operon.

**e.** Possible function according to SubtiList (<http://genolist.pasteur.fr/SubtiList/>) with additional annotations from SWISSPROT/TrEMBL (<http://us.expasy.org/>), InterPro (<http://www.ebi.ac.uk/interpro/>) and Pfam (<http://pfam.wustl.edu/>).

**f.** This region is represented twice in the table (category I and category III) as *abrB* and *metS* are divergently and hence separately transcribed.

**g.** This region is represented twice in the table (category II and category IV). The *yppD-yppE* operon and *yppF* are convergent transcription units. Transcriptional data were lacking for *yppF*, which is a very small gene (<200 bp in length), but both transcription units are likely to be direct targets of Spo0A as both exhibited a Spo0A binding site in gel retardation assays.

**h.**  $\sigma^H$  target according to Britton *et al.* (2002).

**i.**  $\sigma^H$  target according to Ben-Yehuda *et al.* (2003).

**j.** This region is represented twice in the table (category II and category III). A predicted terminator is present between *tkt* and *yneE*, and the genes exhibited opposite patterns of regulation by Spo0A, indicating that they belong to distinct transcription units.

**k.** Essential gene for growth in LB medium according to Kobayashi *et al.* (2003).

**l.** *yzkF* and *ykuL* are considered here as forming an operon as they are transcribed in the same direction and no terminator sequence has been found between the two genes. However, as both genes exhibited a shift in the gel retardation assays, it is equally possible that they constitute two independent transcription units regulated from distinct Spo0A binding sites.

**m.** For conciseness, only two of the 30 genes of the *flaIche* operon have been listed in the table, but expression of every gene in the operon was significantly downregulated (*P*-value < 0.01) in cells in which Spo0A-Sad67 synthesis was induced.

**n.** Although a predicted terminator is present between *cotD* and *ypsA*, the two genes are presumably co-regulated by Spo0A as only *cotD* was shifted in the gel retardation assays.

**o.** According to the results of Moch *et al.* (2000), *nfrA* and *ywch* are considered here as an operon. However, it is possible that they constitute two independent transcription units as a predicted terminator is present between the two genes, and both genes were shifted in the gel retardation assays.

NA, not applicable.

NDE, gene was not differentially expressed in induced cells compared with uninduced cells.

cases, a signal was obtained for adjacent genes, thus representing a common region of the chromosome and conceivably the same binding site). Even though the enrichment factors for all these genes were identified as significant ( $P$ -value  $< 0.01$ ) by the statistical program we used (Rosetta RESOLVER), 13 genes were only slightly enriched by chromatin immunoprecipitation (enrichment factor  $< 1.5$ ). Of these 13 genes, eight are adjacent to genes with an enrichment factor  $> 1.5$ , two (*spolIE* and *kinA*) are known targets of Spo0A, and the remaining three (*yxbB*, *dacA* and *yaaE*) are adjacent to two genes (*yxbC* and *yaaD*) that presented some variability in the four ChIP-on-chip experiments, i.e. they were detected in some but not all the experiments (perhaps corresponding to low-affinity Spo0A binding sites).

Binding sites for transcriptional regulatory proteins, such as Spo0A, generally lie upstream of and outside the ORFs for the genes they control. Yet the microarrays were spotted with PCR products corresponding to protein-coding sequences, not to intergenic regions. As a consequence, the signals obtained in the microarray analysis were expected to arise chiefly from the overlap of the fluorescently labelled probes from the sheared DNA fragments with a Spo0A binding site, on the one hand, and with a nearby ORF, on the other. This meant that, in some cases, and in particular in cases in which a Spo0A-binding site was located between two divergently transcribed genes, it was not immediately apparent which ORF was under the control of Spo0A. Also, it was possible that, in some cases, the sites at which Spo0A-Sad67 binding was observed were not physiologically significant as a control site at which Spo0A (or Spo0A-Sad67) stimulated or repressed transcription *in vivo*. With these considerations in mind, we also carried out transcriptional profiling in an effort to correlate the binding sites with genes whose transcription was influenced by Spo0A-Sad67.

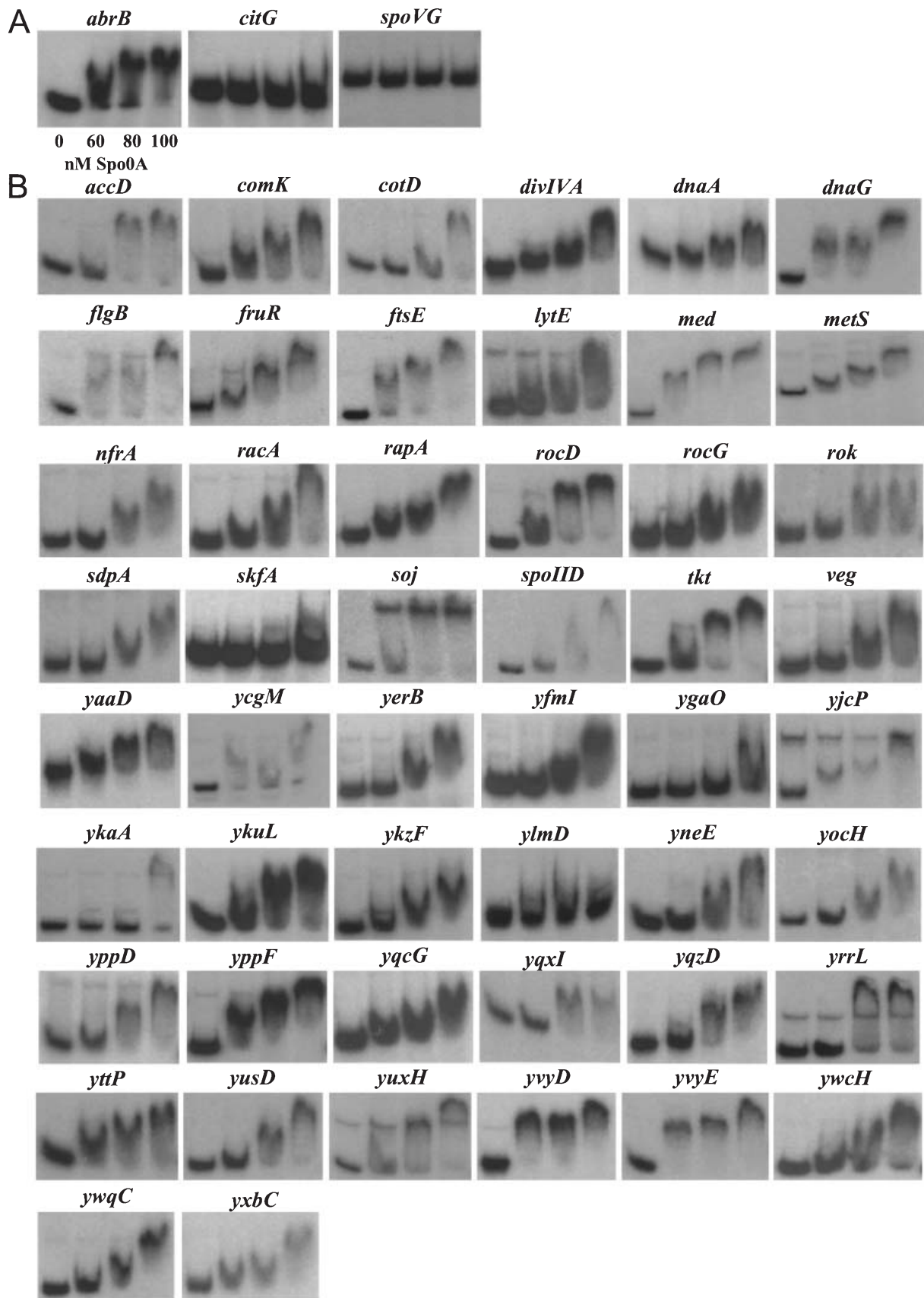
#### Transcriptional profiling

Transcriptional profiling was carried out using cells engineered to produce Spo0A-Sad67 in response to IPTG. Cells of strain RL1104 were grown to the mid-exponential phase of growth, at which time the cultures were split in two, and synthesis of Spo0A-Sad67 was induced by the addition of IPTG to one of the cultures. Samples of cells from the IPTG-treated culture were collected for RNA extraction at 30 min and 60 min after the addition of inducer. For reference, samples of cells were similarly collected in parallel at the corresponding times from the untreated culture (Fig. 1B). Next, cDNAs were generated from the IPTG-treated cells harvested at 30 min after the addition of inducer and labelled with Cy5. Likewise, cDNAs were generated from the untreated cells harvested at the equivalent time and labelled with Cy3. The two

cDNAs were mixed and annealed with DNA microarrays under hybridization conditions. Similarly, cDNAs were generated and differentially labelled with Cy5 and Cy3 from treated and untreated cells harvested at 60 min; As for the 30 min time point, the Cy3- and Cy5-labelled cDNAs from the 60 min time point were mixed and used to probe DNA microarrays. The entire procedure was carried out independently a total of four times for the 30 min time point and four times for the 60 min time point.

Of the 64 regions that were identified in the ChIP-on-chip analysis, 52 contained at least one gene that revealed a significant increase or decrease in hybridization signal in the transcriptional profiling analysis. Genes that were located at or near a site of Spo0A-Sad67 binding, as judged by the ChIP-on-chip analysis, and with transcription that was influenced by Spo0A-Sad67 were considered to be likely targets of direct regulation by Spo0A and are represented in bold in the lefthand column of Table 1. In many cases, more than one gene in a region of the chromosome at which Spo0A-Sad67 binding was detected were altered in their level of transcription in response to Spo0A-Sad67. In all but two cases, the genes in question were contiguous, oriented in the same direction and responded in a co-ordinate manner to Spo0A-Sad67 and, hence, probably constituted a single Spo0A-controlled operon. However, in one case, two adjacent genes (*tkt* and *yneE*) responded oppositely to Spo0A-Sad67 (*tkt* being repressed and *yneE* being stimulated) and, in another case (*abrB* and *metS*), the two genes are arranged in a divergent orientation. We therefore interpret our results to indicate that *tkt*, *yneE*, *abrB* and *metS* are each separately under the control of Spo0A. With these considerations in mind, the total number of Spo0A-regulated transcription units comes to 54 (52 plus 2).

In some cases, the Spo0A-controlled operons detected in our analysis included genes, as judged by the transcriptional profiling analysis, that extended outside the region of the chromosome at which Spo0A-Sad67 binding was detected. For example, *ybcO*, which has been renamed *skfA*, is the first member of a Spo0A-controlled operon that contains eight genes (present findings; González-Pastor *et al.*, 2003). In this case, only the gene most proximal to the binding site for Spo0A was detected in the ChIP-on-chip analysis. Likewise, *yvaW*, which has been renamed *sdpA*, is the promoter-proximal gene of a three-cistron operon, of which only the first two members exhibited a signal in the ChIP-on-chip analysis (González-Pastor *et al.*, 2003). Therefore, based on the results of the transcriptional profiling experiments, we infer that the 54 transcription units correspond to 30 single-gene transcription units and 24 operons, which include 91 genes, for a grand total of 121 genes that are likely to be under the direct control of Spo0A (Table 1).



How successful was our analysis in detecting genes previously known to be under the direct control of Spo0A? Ten transcription units (six single-gene units and four operons, representing a total of 18 genes) were previously known to be under the direct control of Spo0A. Examples of such transcription units, which we refer to as category I, are *spolIA* and *spolIG*, which are induced during sporulation, and *abrB*, which is repressed during sporulation. For all 10 of these Spo0A-controlled transcription units, a statistically significant signal was detected in the ChIP-on-chip experiments at or near the gene (or the promoter-proximal member in the case of an operon), although in two cases (*spolIE* and *kinA*), the enrichment factor was low. The low signal was especially surprising in the case of *spolIE*, which is a strongly expressed target of Spo0A. A possible explanation is that genes adjacent to target genes were frequently observed to exhibit a stronger signal than the target gene itself. In the case of *spolIE*, the adjacent, upstream gene is a tRNA gene, which was not represented in our microarray. Furthermore, all the transcription units except *spo0F* exhibited a statistically significant alteration in their level of transcription in response to Spo0A. (Spo0A is both a positive and a negative regulator of *spo0F* and, hence, the effect of Spo0A on *spo0F* transcription would have been dependent on the precise concentration of the response regulator under the conditions of our experiments; Strauch *et al.*, 1993; Asayama *et al.*, 1995; Fujita and Sadaie, 1998.) Thus, all previously known, direct targets of Spo0A regulation were detected both in the ChIP-on-chip analysis and in the transcriptional profiling experiments, although in some cases the signals were at or near the limits of detection.

Category I, as we have seen, includes 10 transcription units and 18 genes. The remaining 44 transcription units (54 minus 10), representing 103 genes (121 minus 18), were candidates for previously unrecognized members of the Spo0A regulon. Of these, 18 transcription units (representing 31 genes) were genes and operons with transcription that was stimulated by Spo0A-Sad67 during growth and 26 transcription units (72 genes) with a level of transcription that was lower at one or both time points. Previously unrecognized members of the regulon with transcription that was stimulated by Spo0A-Sad67 are referred to as category II, and those with transcription that was inhibited by Spo0A-Sad67 are referred to as category III.

Of the 18 transcription units in category II, all but five (*yjcM*, *yppD–yppE*, *yfml*, *yusE*, *yerB*) had previously been

observed to be induced during sporulation in a Spo0A-dependent manner (Fawcett *et al.*, 2000). Thus, for 13 of the transcription units in category II, the results from ChIP-on-chip analysis and transcriptional profiling during growth with Spo0A-Sad67 were complemented by transcriptional profiling carried out during sporulation. Likewise, of the 26 transcription units in category III with transcription that was inhibited by Spo0A-Sad67, all but seven (*yaaD–yaaE*, *dnaG*, *dnaA–dnaN*, *holB–yaaT*, *rok*, *soj–spo0J* and *pureE*) had previously been observed to be repressed in a Spo0A-dependent manner upon entry into sporulation. As a word of caution, some of the repression effects were small and, in the case of *dnaA–dnaN*, experiments carried out by A. Goranov and A. Grossman (personal communication) under conditions of growth in minimal medium failed to detect Spo0A-Sad67-dependent inhibition of transcription.

The ChIP-on-chip analysis revealed 13 additional chromosomal regions (category IV) at which Spo0A-Sad67 binding was observed but for which transcriptional profiling revealed no genes with a level of transcription that was significantly influenced by Spo0A-Sad67. Possible explanations for genes in this category are: (i) the binding by Spo0A-Sad67 was not physiologically significant; (ii) the gene is under Spo0A control during sporulation but was transcribed at too low a level to have been detected during growth in response to Spo0A-Sad67; (iii) transcription of the gene requires an unknown positive regulatory protein in addition to Spo0A that was absent or inactive under the vegetative growth conditions of our experiment. Indeed, A. Goranov and A. Grossman (personal communication) have observed stimulation of the category IV gene pair *fliT yvyD*, repression of the *yweA rocG* gene pair and repression of the *ywqB ywqC ywqD* cluster by Spo0A-Sad67 under conditions of growth in minimal medium. Reinforcing the view that *ywqC* is a *bona fide* target of Spo0A, Fawcett *et al.* (2000) observed that it is repressed in a Spo0A-dependent manner during sporulation. At least two other genes in category IV are also likely to be *bona fide* targets of Spo0A, as their transcription was previously found to increase (in the case of *comQ*) or decrease (in the case of *ygaO*) during sporulation in a Spo0A-dependent manner (Fawcett *et al.*, 2000).

#### Gel electrophoretic mobility shift assays of Spo0A binding

Genes in categories II and III are strong candidates for previously unrecognized members of the Spo0A regulon

**Fig. 1.** Gel electrophoretic mobility shift analysis of the binding of Spo0A to putative regulatory regions for genes under its control. The regulatory regions of potential target genes were amplified by PCR using radioactive primers, incubated with the purified DNA-binding domain of Spo0A at the indicated concentrations and analysed by non-denaturing PAGE (see *Experimental procedures*). A. Gel shifts for controls. *abrB* is a known target of Spo0A and represents a positive control. *citG* and *spoVG* are negative controls. B. Gel shifts for genes in categories II, III and IV.

because they are at or near binding sites for Spo0A-Sad67 and because their transcription was significantly influenced by the regulatory protein. As a further test of whether these genes are direct targets of Spo0A, we investigated whether Spo0A would bind to the putative upstream regulatory regions for all the transcription units in categories II and III *in vitro* (in the case of putative multigene transcription units, we tested binding to the upstream region of the most upstream member of the unit). We also included all the genes in category IV in our analysis. Finally, we included *abrB*, which is known to be under the direct control of Spo0A, as a positive control, and *citG* and *spoVG* as negative controls. The binding experiments were carried out using radioactive DNAs that extended 200 bp upstream from the translational start site for the genes tested and a truncated form of Spo0A that corresponded to its DNA-binding domain, which is located in the C-terminal portion of the protein (Fig. 1). This truncated protein lacks the region in which phosphorylation takes place and is known to bind to DNA in a manner that is independent of phosphorylation (Rowe-Magnus and Spiegelman, 1998). DNA binding was assessed using a gel electrophoretic mobility shift assay (EMSA).

The results show that the truncated protein interacted measurably with the upstream regions for 17 out of 19 genes in category II and 24 out of 28 genes in category III (Fig. 1 and Table 1). Thus, in total, we were able to confirm biochemically the presence of a Spo0A binding site for 41 of our best candidates for newly identified members of the regulon. Of the remaining 13 genes (category IV) for which a binding site was detected *in vivo* but transcription of which was not found to be influenced by Spo0A-Sad67, 10 genes were identified with upstream regions that contained a binding site for the truncated Spo0A protein.

#### *Use of bioinformatics to identify an unbiased consensus binding sequence for Spo0A*

A consensus binding site for Spo0A, the seven nucleotide sequence 5'-TGTCGAA-3', was reported previously based on binding studies of the response regulator to the regulatory region of genes known to be under its direct control (Strauch *et al.*, 1990; Baldus *et al.*, 1994), and

mutational analysis has confirmed the importance of the G:C basepair at position five (5'-TGTCGAA-3') in Spo0A binding (Satola *et al.*, 1991; York *et al.*, 1992). Also, a co-crystal of the DNA-binding domain of Spo0A to a synthetic consensus binding sequence showed that the G at position 2 on the upper strand, the C at position 4 on the upper strand and the G at position 4 on the lower strand contact amino acid side-chains in the protein (Zhao *et al.*, 2002). We wished to take advantage of our identification of 64 regions of the chromosome for which Spo0A-Sad67 binding was observed in an independent computational approach to identifying the consensus binding site that was unbiased by the previously arrived at consensus.

The first step in our strategy was to use the Gibbs sampling algorithm BIOPROPECTOR (Liu *et al.*, 2001) to search for conserved motifs in the upstream region of the 64 Spo0A binding regions. Some of the Spo0A binding regions were inferred to contain more than one Spo0A-controlled transcription unit, bringing the total number of (non-overlapping) upstream sequences that were entered into the program to 69. We limited our search to 200 bp of upstream sequence, reasoning that binding sites for Spo0A should generally be located no further upstream than this distance from the gene under its control. In our analysis, we allowed for the possibility of multiple binding sites in an upstream region and in either orientation relative to the nearby gene. Because BIOPROPECTOR only accepts a fixed motif width, we ran the program separately for motif widths ranging from 7 bp to 11 bp in length. The best five motifs found by BIOPROPECTOR for each of the five motif widths were kept for the next step of the analysis, for a total of 25 possible motifs.

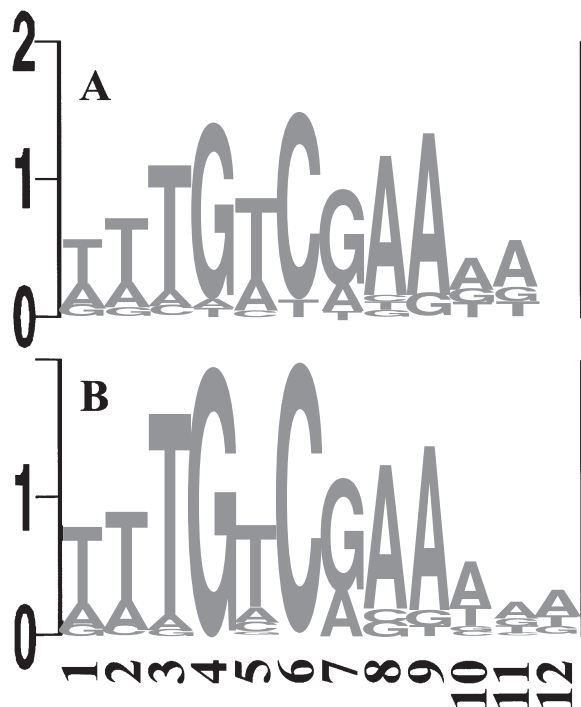
In the second step of the procedure, each of these BIOPROPECTOR searches was optimized using an algorithm called BIOOPTIMIZER that 'cleans up' (by adding or removing putative binding sites from the analysis) the BIOPROPECTOR motif as well as allowing the motif width to vary. Out of the 25 optimized motifs, the one with the highest score (refer to *Experimental procedures* for details of the score) was selected as the 'best' Spo0A DNA-binding motif. Interestingly, this 'best' motif resulted from the optimization of four of the 25 different BIOPROPECTOR motifs. A summary of the optimization results for this 'best' motif are given in Table 2. We see from the table that the

**Table 2.** Results of the optimization of Spo0A-binding motifs from BIOPROPECTOR by BIOOPTIMIZER.

BIOPROPECTOR starting point			Final optimization point		
Consensus	Motif width	Number of sites	Consensus	Motif width	Number of sites
TTGTCGAA	8	51	TTTGTGCGAAAA	12	51
ATTTGTGCGAAA	11	35	TTTGTGCGAAAA	12	51
TTTGTGCGAAAA	11	36	TTTGTGCGAAAA	12	51
TGTCGAAAAAT	11	34	TTTGTGCGAAAA	12	51

motifs resulting from the optimization procedure converged on exactly the same predicted sites, consensus sequence and motif width despite the fact that each of the four BIOPROPECTOR motifs differed in terms of the number of predicted sites, consensus sequence and motif width. The sequence logo for this 'best' motif is given in Fig. 2B, which conforms to the previously described consensus seven-nucleotide-long sequence but extends it by 5 bp (2 bp on the 5' side and 3 bp on the 3' side). The remaining 21 out of 25 optimized motifs were not exactly the same as this 'best' motif, but were quite similar. Figure 2A also shows a sequence logo derived from all previously known *Spo0A* binding sites, which extends the previous logo on the 5' and 3' sides, reinforcing the validity of the logo derived by BIOPROPECTOR/BIOOPTIMIZER.

The BIOPROPECTOR/BIOOPTIMIZER procedure was not restricted to finding a motif site in every single upstream sequence and, in fact, the 'best' motif found had 51 predicted sites within the 70 upstream regions. In a third stage of our analysis, we scanned all the upstream sequences as well as 100 nucleotides of downstream sequence to identify in each case the best matches to the 12-nucleotide-long consensus matrix of the 'best' motif as



**Fig. 2.** *Spo0A* consensus binding site. A. The consensus binding sequence for *Spo0A* as derived from all previously known *Spo0A* binding sites (category I). B. The consensus binding sequence for *Spo0A* as inferred by application of the BIOPROPECTOR AND BIOOPTIMIZER programs to *Spo0A*-binding regions identified by ChIP-on-chip. The results are presented as sequence logos in which the height of the letters is proportional to their frequency (Schneider and Stephens, 1990).

identified by BIOPROPECTOR/BIOOPTIMIZER as well as matches to a matrix based on the seven mostly highly conserved positions. Because this procedure finds a potential site for every upstream sequence regardless of the site's strength, we also calculated a *P*-value that represented the significance of that site and included only sites with *P*-values <0.1. The list of predicted sites from this scanning procedure is presented in *Supplementary material* (Table S1). The scanning procedure successfully identified at least one of the known *Spo0A* binding sites for the 10 genes in category I. It did not, however, detect all previously known sites and predicted some sites at which binding has not been observed. For comparison, in our previous study, BIOPROPECTOR successfully identified 18 out of 24 experimentally verified promoters for RNA polymerase containing the sporulation transcription factor  $\sigma^E$  (Eichenberger *et al.*, 2003). A high proportion of the sites predicted by BIOPROPECTOR/BIOOPTIMIZER were located upstream of the gene but, in a few cases, particularly for genes under the negative control of *Spo0A* (category III), predicted sites were located within the coding sequence (see the *Supplementary material*, Table S1).

#### *Creating null mutations in newly identified members of the Spo0A regulon*

In a final part of our analysis, we used a long flanking homology PCR strategy (Wach, 1996) to build null mutations in newly identified members of the *Spo0A* regulon and tested the resulting mutants for sporulation proficiency. Seven of the mutants proved to be of interest, and these fell into three groups: those that sporulated faster than the wild type, those that were mildly impaired in sporulation and one that was modestly impaired in sporulation.

Mutants of *yzkF* and *med* produced colonies that sporulated more rapidly than the wild type, the effect with the *yzkF* mutant being more pronounced than that for the *med* mutant. The *yzkF* gene, which has a binding site for *Spo0A in vivo* and is induced during sporulation in a *Spo0A*-dependent manner, is of unknown function. The gene immediately downstream of *yzkF* is *ykuL*, which was also detected in the ChIP-on-chip and gel shift experiments (using a fragment of the *ykuL* upstream region that only included the 3' end of the *yzkF* coding sequence). However, disruption of *ykuL* did not result in accelerated sporulation, indicating that *yzkF* alone is responsible for this phenotype. The *med* gene (Ogura *et al.*, 1997), which is a binding site for *Spo0A in vivo* and is repressed in a *Spo0A*-dependent manner, is a positive regulator of *comK*, which encodes the principal transcriptional activator for genetic competence. Thus, repression of *med* might contribute to inhibition of competence gene expression

during sporulation. Interestingly, *comK* is itself an *in vivo* binding site for Spo0A, although we detected no influence of Spo0A on *comK* expression. Conceivably, Spo0A contributes to the repression of *comK* during sporulation; if so, however, such an effect was not detected under the conditions of our analysis.

The second group consisted of four mutants that were mildly impaired in sporulation: *rocG*, *yerB*, *yqxI* and *ywkC*. When tested in liquid sporulation medium, the level of spore formation was reproducibly reduced by two- to threefold compared with the wild type. The case of *ywkC* is instructive as it shows the value of investigating the function of genes with mild sporulation phenotypes. A mutant of *ywkC* (renamed *racA*) had been built by Ben-Yehuda *et al.* (2003), who found that it encoded a protein with synthesis that is strongly induced in a Spo0A-dependent manner and which is responsible for remodelling sister chromosomes into an axial filament and for anchoring the chromosomes to the cell poles shortly after the onset of sporulation. A *racA* (*ywkC*) mutation causes a strong defect in chromosomal remodelling and anchoring but does not impair sporulation severely because spore formation is not strongly dependent upon axial filament formation. Like *racA*, *yerB* and *yqxI* are induced during sporulation in a Spo0A-dependent manner (category II genes). Also like *racA*, *yerB* and *yqxI* exhibit little or no similarity to genes of known function. We infer from the effects of mutations in these genes that they contribute to efficient sporulation, but the nature of their gene products and their possible function in sporulation is unknown. Interestingly, *yqxI* is located within the *skin* element, a genetic element that is excised at a late stage of sporulation to create an intact coding sequence for the sporulation transcription factor  $\sigma^K$ . It has been shown that strains devoid of *skin* sporulate normally, but it is possible that a mild sporulation defect could have been overlooked. Another direct target of Spo0A, *yqcG* (category II), is also found in the *skin* element and in the vicinity of *yqxI*, but it is transcribed in the opposite direction. The inferred *yqcG* gene product exhibits some similarity to a transposase, and disruption of *yqcG* does not seem to affect sporulation efficiency significantly (S. Ben-Yehuda and R. Losick, unpublished results). In any event, whatever the role of *yqxI* during sporulation, it is evidently specific to *B. subtilis* because *yqxI* and the *skin* element itself are not found in most other endospore formers. In contrast, orthologues of *racA* are found in most *Bacillus* spp. but not in *Clostridium* spp., whereas *yerB* has orthologues in *Bacillus halodurans* and *Ocaenobacillus iheyensis* but not in *Bacillus cereus* or *Bacillus anthracis*. Finally, we come to the case of *rocG*. The *rocG* gene encodes the enzyme glutamate dehydrogenase, which is needed for efficient utilization of arginine, ornithine and proline as nitrogen or carbon sources (Belitsky and Sonenshein, 1999; 1998). Although

*rocG*, which is a category IV gene, contains a binding site for Spo0A, no effect of Spo0A on *rocG* transcription was detected in our present study or in the investigation by Fawcett *et al.* (2000). Thus, the significance of the mild sporulation defect caused by a mutation in this gene is unclear, especially given the fact that *rocG* homologues are found in both endospore-forming and non-endospore-forming bacteria such as *Listeria*.

The last group is represented by the *ycgM* mutant, which was originally built and investigated by J. Silvaggi and P. Eichenberger (unpublished results). The *ycgM* mutant sporulated five- to 10-fold less efficiently than the wild type. The *ycgM* gene product is similar to proline oxidase. As pointed out to us by A. L. Sonenshein and B. Belitsky (personal communication), the effect of the *ycgM* mutation could result from the aberrant accumulation of a toxic compound, such as  $\delta$ -pyroline-5-carboxylate, which is generated during the breakdown and biosynthesis of proline and the breakdown of arginine. Interestingly, *ycgM* is conserved among *Bacillus* spp. and absent from *Clostridium* and *Listeria*.

## Discussion

Using a combination of three approaches, ChIP-on-chip analysis, transcriptional profiling and EMSAs, we have been able to assign 103 additional genes to the Spo0A regulon. With the 18 previously known members of the regulon, the tally is now 121 genes, which are organized as 30 single-gene units and 24 operons. Our analysis complements and extends that of Fawcett *et al.* (2000) who carried out transcriptional profiling during sporulation (as well as during growth using cells engineered to produce Spo0A-Sad-67) to identify genes with expression that is influenced by Spo0A-P. The analysis of Fawcett *et al.* (2000) was based on arrays that were less complete than those that were available for the present study and also relied on a less sensitive protocol for detecting alterations in gene expression than the fluorescence-based method used here. Nonetheless, among the 18 transcription units in category II (that is previously uncharacterized transcription units that were associated with a binding site for Spo0A-Sad67 and with transcription that was stimulated during growth by Spo0A-Sad67), 13 [*yxbC*, *skfA* (*ybcO*), *yttP*, *racA*, *yocH*, *yybN*, *sdpB* (*yvaX*), *yneE*, *ycgM*, *yqcG*, *yqxI*, *accA*, *ykzF*; Fawcett *et al.*, 2000] corresponded to transcription units that were induced during sporulation in a Spo0A-dependent manner. Meanwhile, other data have confirmed that *racA* (Ben-Yehuda *et al.*, 2003), *sdpA* (*yvaW*), *sdpB* (*yvaX*) and *sdpC* (*yvaY*), which appear to constitute an operon (González-Pastor *et al.*, 2003), and *skfA* (*ybcO*), *skfB* (*ybcP*), *skfC* (*ybcS*), *skfD* (*ybcT*), *skfE* (*ybdA*), *skfF* (*ybdB*), *skfG* (*ybdD*) and *skfH* (*ybdE*), which also appear to constitute an operon,

are induced during sporulation in a Spo0A-dependent manner (González-Pastor *et al.*, 2003). Thus, a high proportion of the genes assigned to the Spo0A regulon on the basis of harbouring a binding site for Spo0A-Sad67 and being induced by Spo0A-Sad67 during growth also meet the criterion of being expressed during sporulation in a Spo0A-dependent manner.

Also pertinent to our analysis are the transcriptional profiling experiments of Britton *et al.* (2002), who identified genes with transcription during sporulation that is dependent upon  $\sigma^H$ . Spo0A-P acts in conjunction with both the  $\sigma^A$ -containing and  $\sigma^H$ -containing forms of RNA polymerase and, therefore, some of the newly identified genes in the Spo0A regulon would be expected to be under the direct control of  $\sigma^H$ . Also,  $\sigma^H$ -RNA polymerase stimulates the synthesis of Spo0A during sporulation by transcription from a  $\sigma^H$ -controlled promoter just upstream of the *spo0A* gene. Thus, many Spo0A-controlled genes are expected to be directly or indirectly dependent on  $\sigma^H$  for their transcription. Indeed, among the 18 transcription units in category II genes, 13 [*yxbC*, *skfB* (*ybcP*), *yttP*, *racA*, *yocH*, *yvbN*, *sdpA* (*yvaW*), *yneE*, *yfml*, *yqcG*, *yqxl*, *accD* (*yttI*) and *ykuL*] had previously been found to be transcribed during sporulation in a  $\sigma^H$ -dependent manner and, in two cases (*yttP* and *racA*), a putative  $\sigma^H$ -controlled promoter was found upstream of the gene. Therefore, of the 18 transcription units herein assigned to the Spo0A regulon, all but four (*yjcM*, *yppD-yppE*, *yusE* and *yerB*) had been found to be induced during sporulation in a Spo0A- or  $\sigma^H$ -dependent manner in the studies of Fawcett *et al.* (2000) or Britton *et al.* (2002).

Interestingly, a high proportion of category II genes either had little similarity to other genes in the databases or were similar to genes of unknown function. Recently, functions have been elucidated for several of the genes in category II, such as *racA*, which was discussed above (Ben-Yehuda *et al.*, 2003). Likewise, the eight-cistron *skfABCDEFGH* (*ybcOPST ybdABDE*) operon has been shown to be responsible for the production of an antibiotic-like killer factor, and the three-cistron *sdpABC* (*yvaWXY*) operon has been shown to be responsible for the production of an exported signalling protein (González-Pastor *et al.*, 2003). Mutations of both operons cause an accelerated sporulation phenotype, similar to that seen here for mutations of *yzkF* and *med*. The killing factor and signalling protein are part of a system that delays sporulation by killing cells that have not entered the pathway to sporulate. The dead cells release nutrients that allow cells with activated Spo0A to continue growing. It will be interesting to see whether *yzkF* and *med* participate in this unusual system of cannibalistic behaviour.

Other members of category II are known genes or resemble genes of known function. These are: the *accD-accA* operon, which encodes the beta and alpha subunits

of acetyl-coenzyme A (CoA) carboxylase (an enzyme involved in long-chain fatty acid biosynthesis); *yocH*, the product of which is similar to a cell wall-binding protein; *yfml*, the product of which is similar to a macrolide efflux transporter; *yttP*, the product of which is similar to a transcriptional regulator of the TetR/AcrR family; *yqcG*, the product of which is similar to a transposase; *yjcM*, the product of which is similar to an alcohol dehydrogenase, and *ycgM*, the product of which is similar to a proline oxidase. A mutant of *ycgM* was found to be impaired in sporulation but, as indicated in the *Results*, this phenotype is probably an indirect consequence of the accumulation of a toxic intermediate in amino acid catabolism. As noted above, many of the genes in category II are expressed during sporulation in a  $\sigma^H$ -dependent manner (Britton *et al.*, 2002). Furthermore, some of the genes in category II appear to be functionally related to  $\sigma^H$ -controlled genes that are evidently not members of the Spo0A regulon. For example, the  $\sigma^H$ -controlled gene *yojL*, which is inferred to encode a cell wall-binding protein, is homologous to the Spo0A-activated gene *yocH*. Similarly, the  $\sigma^H$ -controlled genes *yoeA* and *ywfF*, which are thought to be involved in the export of toxic compounds, could act in parallel with the category II gene *yfml*, the product of which is similar to a macrolide efflux transporter.

Most of the genes in category III, that is genes with transcription that was repressed by Spo0A-Sad67, are of known function and, in six cases, were described recently as being essential for growth of *B. subtilis* in LB medium (Kobayashi *et al.*, 2003). Several of these genes are of particular interest. The product of the *soj* gene represses several genes that are activated by Spo0A-P, including *spo0A* itself (Quisel and Grossman, 2000). Thus, by inhibiting transcription of *soj*, Spo0A-P could contribute to a self-reinforcing cycle in which Spo0A stimulates its own synthesis. The product of *divIVA* is responsible for sequestering the cell division inhibitor, MinCD, to the cell poles during vegetative growth (Marston *et al.*, 1998). Spo0A-P-mediated inhibition of *divIVA* transcription could lead to reduced levels of DivIVA after the start of sporulation, contributing to the release of MinCD from the poles and facilitating the switch to polar division. By inhibiting the transcription of *dnaA*, *dnaG*, *dnaN* and *holB*, which encode the replication initiation protein DnaA, DNA primase and the  $\beta$  and  $\delta'$  subunits of the DNA polymerase III holoenzyme, respectively, Spo0A-P could contribute to blocking new rounds of chromosome duplication after the start of sporulation. The 30 genes of the 26 kb *fla-che* operon encode most of the functions required for flagellum biosynthesis and chemotaxis (West *et al.*, 2000). It is known that cells stop swimming shortly after they enter the pathway to sporulate, perhaps because of inhibition of transcription of the *fla-che* operon by Spo0A-P.

Next, we comment on the anomalous case of the spore coat protein gene *cotD*, which was found to contain a Spo0A binding site and transcription of which in growing cells was found to be inhibited by Spo0A-Sad67 both in our present experiments and as reported previously (Fawcett *et al.*, 2000). The promoter for *cotD* is under the control of the late-acting sporulation regulatory protein  $\sigma^K$ , which is absent in vegetative cells. Hence, transcription of *cotD* during growth is unexpected. Evidently, *cotD* is subject to a second mode of transcription during growth. If so, then repression by Spo0A-P could serve to prevent inappropriate expression of the spore coat protein gene during early stages of sporulation. Also, repression of *cotD* by Spo0A-P could serve to downregulate readthrough transcription into *ypsA*, which is located immediately downstream of *cotD* and was also found to be negatively regulated by Spo0A-Sad67. Gel electrophoretic mobility experiments revealed a Spo0A binding site located upstream of *cotD* but not upstream of *ypsA*, findings that are consistent with the idea that the influence of Spo0A-Sad67 on the transcription of both genes originates from a single Spo0A binding site upstream of *cotD*.

One gene in category IV deserves special mention. The *spoIID* gene was identified as a Spo0A binding site in both ChIP-on-chip analysis and electrophoretic mobility shift experiments, but no signal was detected for *spoIID* in our transcriptional profiling experiments. The *spoIID* gene is a mother cell-expressed gene and is known to be under the direct control of  $\sigma^E$ . Thus, we would not have expected it to be transcribed during growth in cells engineered to produce Spo0A-Sad67. What then is the physiological significance of a Spo0A binding site near *spoIID*? Recent work has shown that Spo0A-P continues to accumulate after the initiation phase of sporulation, reaching high

levels in the mother cell (Fujita and Losick, 2003). Perhaps Spo0A-P helps to fine tune the  $\sigma^E$ -dependent expression of *spoIID* after polar septation.

We were also interested in determining how many of the Spo0A regulon members were conserved in other low-GC Gram-positive bacteria related to *B. subtilis*. Complete genome sequences are currently available for four other *Bacilli*, i.e. *B. cereus* (Ivanova *et al.*, 2003), *B. anthracis* (Read *et al.*, 2003), *B. halodurans* (Takami *et al.*, 2000) and *O. iheyensis* (Takami *et al.*, 2002), three *Clostridium* species, i.e. *C. acetobutylicum* (Nolling *et al.*, 2001), *C. perfringens* (Shimizu *et al.*, 2002) and *C. tetani* (Bruggemann *et al.*, 2003) and two *Listeria* species, i.e. *L. monocytogenes* and *L. innocua* (Glaser *et al.*, 2001). Spo0A itself is conserved among the endospore-forming bacteria of the genera *Bacillus* and *Clostridium* but is absent from the closely related but non-endospore-forming genus *Listeria*. Thus, given the taxonomic proximity between *Bacillus* and *Listeria*, many *Bacillus* genes are expected to be conserved in *Listeria* with the exception of sporulation genes (unless they have a function in addition to sporulation). Conversely, sporulation genes should be more frequently conserved between *Bacillus* and *Clostridium* compared with other gene categories. We present the results of the comparative genomics analysis in Table 3. Strikingly, the majority of genes repressed by Spo0A, which are consequently expressed during vegetative growth and turned off during entry into sporulation, are widely conserved among the three genera, with many genes being conserved between *Bacillus* and *Listeria*. This high degree of conservation is also consistent with the presence of a significant number of essential genes in category III. On the other hand, genes activated by Spo0A are less widely conserved and clearly under-

**Table 3.** Comparative genomics.

	<i>Bacilli</i>					<i>Listeria</i>		<i>Clostridia</i>		
	<i>B. subtilis</i>	<i>B. anthracis</i>	<i>B. cereus</i>	<i>B. halodurans</i>	<i>O. iheyensis</i>	<i>L. innocua</i>	<i>L. monocytogenes</i>	<i>C. acetobutylicum</i>	<i>C. perfringens</i>	<i>C. tetani</i>
Orthologues to transcription units <sup>a</sup> activated by Spo0A in <i>B. subtilis</i>	23 (100%)	14 (61%)	14 (61%)	14 (61%)	13 (57%)	5 (22%)	5 (22%)	6 (26%)	7 (30%)	6 (26%)
Orthologues to transcription units repressed by Spo0A in <i>B. subtilis</i>	30 (100%)	24 (80%)	24 (80%)	24 (80%)	23 (77%)	21 (70%)	21 (70%)	18 (60%)	15 (50%)	15 (50%)
Total orthologues <sup>b</sup>	53 (100%)	38 (72%)	38 (72%)	38 (72%)	36 (68%)	26 (49%)	26 (49%)	24 (45%)	22 (42%)	21 (40%)

**a.** Orthologues were identified for Spo0A-controlled transcription units instead of genes, in order to avoid bias introduced by large operons (e.g. the *fla/che* operon, which contains 30 genes). When some but not all the genes in a given operon were conserved in other species, the operon was nevertheless considered to be orthologous to the corresponding *B. subtilis* transcription unit.

**b.** *spo0A* was not included in the table, but orthologues of *spo0A* are found in all the species represented in the table, except *Listeria monocytogenes* and *Listeria innocua*.

represented in *Listeria*. These genes are more likely to be directly involved in sporulation-related processes and are expected to be absent in non-endospore-forming bacteria. It may appear surprising that only a small number of genes activated by Spo0A in *B. subtilis* are found in *Clostridium*; however, this is consistent with the fact that initiation of sporulation in *Clostridium* appears to occur through a different pathway (one that apparently lacks a multicomponent phosphorelay) and under very different environmental conditions (*Clostridia* are anaerobes).

Finally, we comment on the size of the Spo0A regulon. Judging from our success in detecting essentially all genes previously known to be under the direct control of Spo0A, we believe that our current estimate for the size of the Spo0A regulon (121 genes, representing <3% of the genes in the genome) is likely to represent a high proportion of the genes that are under the direct control of the master regulator for entry into sporulation. Meanwhile, previous work has shown that Spo0A influences the transcription of more than 500 genes during sporulation, representing over 10% of the genes in the genome. Evidently, and assuming that 121 genes is not a gross underestimate of the size of the regulon, Spo0A must exert its effect on many genes in an indirect manner. Indeed, many of the genes that are under the direct control of Spo0A are themselves involved in the regulation of gene expression, either directly, as in the cases of *abrB*, *dnaA*, *sigF* (*spolIAC*), *sigE* (*spolIGB*), *sinR*, *rok*, *soj*, *fruR* and *sigD*, or indirectly, as in the case of *rapA*, *phrA*, *kinA*, *kinC*, *med*, *sinI*, *relA*, *spolIAA*, *spolIAB*, *spolIE*, *spo0F*, *spo0J* and *sdpA* (*yvaW*), *sdpB* (*yvaX*) and *sdpC* (*yvaY*) (González-Pastor *et al.*, 2003). In addition, the inferred product of *yttP* resembles members of the TetR/AcrR family of transcription regulators. Genes involved in the regulation of gene expression are also well represented in category IV (i.e. *comQ*, *sigV*, *comK* and *yvyD*). Spo0A is indeed a master regulator for sporulation, but much of its influence on the global pattern of gene expression is likely to be an indirect consequence of its effects on the synthesis of other regulatory proteins.

## Experimental procedures

### Chromatin immunoprecipitation microarray (ChIP-on-chip) experiments

An overnight culture of *B. subtilis* strain RL1104 (*amyE::P<sub>spac</sub>-spo0A-sad67 cat, spo0AΔerm*) (Fawcett *et al.*, 2000) on LB agar was used to inoculate a 50 ml culture in DS medium to give an initial OD<sub>600</sub> of 0.05. Cross-linking with formaldehyde, extraction and shearing of DNA and immunoprecipitation generally followed the protocol of Quisel *et al.* (1999) with differences in details noted. When the cells grown at 37°C reached an OD<sub>600</sub> of 0.3–0.4, the induction of P<sub>spac</sub>-*spo0A-sad67* was accomplished by adding IPTG at a final concen-

tration of 1 mM. At times T<sub>0.5</sub> (30 min) and T<sub>1</sub> (1 h) after induction (the end of exponential growth), the cultures were treated with formaldehyde (1% final concentration) and NaPO<sub>4</sub> (10 mM final concentration) for 30 min. Glycine was added to 125 mM final concentration, and the cultures were incubated for an additional 5 min. Cells were washed twice with 40 ml of phosphate-buffered saline (PBS; pH 7.3) (Ausubel *et al.*, 1990), resuspended in 1 ml of IP buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5% Triton X-100) supplemented with 50 µl of 1× protease inhibitor cocktail (Roche) and 10 mg of lysozyme and incubated at 37°C for 20 min. DNA in the lysate was sheared by sonication (Branson 250 microtip sonicator) to give an average fragment size of 300–1000 bp. Ten microlitres of the supernatant fluid of subsequent centrifugation was removed and saved for later analysis ('total DNA'). The remainder of the supernatant fluid was precleared by incubation with 1/10th volume of 50% protein A–Sepharose slurry (Sigma) for 1 h at 4°C. After centrifugation, Spo0A and Spo0A–DNA complexes in the supernatant fluid were immunoprecipitated overnight at 4°C using rabbit antibody to Spo0A, followed by incubation with 50 µl of a 50% Protein A–Sepharose slurry (1 h, 4°C). Complexes were washed four times (15 min each) with 1 ml of IP buffer. The slurry was resuspended in 150 µl of elution buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1% SDS). The 10 µl total DNA sample was mixed with 150 µl of elution buffer. To reverse formaldehyde-induced cross-links, the immunoprecipitated and total DNA samples were incubated at 65°C overnight. Supernatants were collected and treated at 37°C for 2 h with 150 µl of TE buffer containing glycogen (0.27 mg ml<sup>-1</sup>) and proteinase K (100 µg ml<sup>-1</sup>). The DNA was purified by phenol–chloroform extraction, precipitated with isopropanol and washed with 70% ethanol. Immunoprecipitated DNA was resuspended in 25 µl of TE, and the total DNA sample was resuspended in 100 µl of TE.

PCR amplification of DNA, differential fluorescent labelling, hybridization to microarrays and array scanning were done according to the protocols at <http://microarrays.org/protocols.html>. The arrays were spotted with 4074 PCR products corresponding to *B. subtilis* ORFs as well as with four *Escherichia coli* genes as negative controls (Britton *et al.*, 2002). The hybridization results were scanned using a GenePix 4000B scanner (Axon Instruments) and analysed with GENEPIX 3.0 software (Axon Instruments) (Britton *et al.*, 2002). The entire procedure was carried out four independent times at T<sub>0.5</sub> and T<sub>1</sub>, and the results were averaged. The enrichment factor for a given gene was calculated as the ratio of hybridization of immunoprecipitated DNA to total DNA, normalized using RESOLVER software (Rosetta).

### Growth of cells and extraction of RNA for microarray analysis

Cultures of strain RL1104 harbouring P<sub>spac</sub>-*spo0A-sad67* were grown in LB medium at 37°C to an OD<sub>600</sub> of 0.3–0.4, at which time the cultures were split in two, and *spo0A* was induced by the addition of 1 mM IPTG to one of the cultures. Samples of cells for RNA extraction were taken 30 and 60 min after induction with IPTG, pelleted and frozen immediately in liquid N<sub>2</sub>, and were compared with the same time points in the parallel cultures without IPTG. RNA was har-

vested from each culture and prepared for hybridization as described by Britton *et al.* (2002). The entire procedure was carried out four independent times for  $T_{0.5}$  and  $T_1$ , and the results were averaged. The microarrays used came from the same set as those used in ChIP-on-chip experiments.

### Gel mobility shift assays

DNA fragments used in this assay correspond to the region 250 bp upstream of the start codon and were amplified by PCR. The 5' ends of DNA were labelled using [ $\gamma$ - $^{32}$ P]-ATP and T4 polynucleotide kinase. A typical assay mixture contained in 20  $\mu$ l: 10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol (DTT), 5% (v/v) glycerol, 0.5  $\mu$ g of BSA, radioactive DNA probe (2000 c.p.m. ml $^{-1}$ ) and various amounts of the purified DNA-binding domain of Spo0A protein (Rowe-Magnus and Spiegelman, 1998). After 30 min of incubation at room temperature, 20  $\mu$ l of this mixture was loaded onto a native 4% (w/v) polyacrylamide TBE Ready Gel (Bio-Rad) and electrophoresed in 1 $\times$  TBE buffer for 1 h at 100 V cm $^{-1}$ . Radioactive species were detected by autoradiography after exposure to Kodak X-OMAT film at  $-70^\circ\text{C}$ .

### Purification of the DNA-binding domain of Spo0A protein

A DNA fragment encoding the C-terminal, DNA-binding domain of Spo0A (codons 143–267; Rowe-Magnus *et al.*, 1998) was amplified by PCR from chromosomal DNA of PY79 using oligonucleotides (5'-TACATATGAGCAGCCAGCCTGAACC-3' and 5'-TGCTCGAGTTATTAAGGTTTTGAA-3') that introduced a *Nde*I site at the 5' end and a *Xho*I site at the 3' end. The fragment was cloned into pET21b, thereby generating an extension of six histidine codons at the 3' end of the gene (pMF14). Expression and purification of protein was performed as described previously (Fujita and Losick, 2003).

### Bioinformatics

Our BIOPROPECTOR/BIOOPTIMIZER procedure was designed to find the set of sites in the upstream sequences of the 69 genes identified in the ChIP-on-chip analysis that best fitted the Bayesian model for binding motifs (Lawrence *et al.*, 1993; Liu, 1994; Liu *et al.*, 1995; reviewed by Jensen *et al.*, 2004). If we represent the sequence data as  $S$ , then the best fit of the model is the set of motif start sites  $A$  and motif width  $w$  that gives the maximum posterior density  $p(A, w|S)$ . As long as prior information is vague, maximizing the posterior density is equivalent to maximum likelihood estimation. We took the logarithm of this posterior distribution and called this our scoring function  $\psi(A, w) = \log p(A, w|S)$ , so that our goal was then to maximize the scoring function, or score,  $\psi$ . This maximization was accomplished by a simple optimization algorithm. Given an initial configuration of sites,  $A$ , we systematically scanned through every position of every sequence in the data set and decided whether to add this position to  $A$  as the start of a motif site or to remove it from  $A$  if it was already in  $A$ . If we denote  $A'$  as  $A$  with this change made, then we accept the change only if  $\psi(A', w) > \psi(A, w)$ .

Similarly, after each scan through every position, we considered all possible changes in  $w$  to either  $w' = w + 1$  or  $w' = w - 1$  and in either ends of the motif and selected the one that increased the score, i.e.  $\psi(A, w') > \psi(A, w)$ . As this is a deterministic strategy, it was important that we started the algorithm in a configuration  $A$  that is already close to the maximum, for which we used the program BIOPROPECTOR (Liu *et al.*, 2001). BIOPROPECTOR is a stochastic implementation of the motif model, which gives it more flexibility to find good motif signals, although it is not designed to find an optimal motif signal, nor will exactly the same motif signals be found in independent runs of the program. The program is also restricted to a fixed motif width  $w$ . Therefore, we used a combined procedure for finding these unknown motif sites, first using BIOPROPECTOR to find putative Spo0A binding sites, and then the optimization algorithm described previously (which we have named BIOOPTIMIZER) to 'clean up' the motif signal via the possible addition of new sites or the removal of current sites, as well as allowing changes in motif width  $w$  in order to find the optimal motif width.

Having found an optimal motif with this combined BIOPROPECTOR/BIOOPTIMIZER procedure, we implemented an additional scanning procedure to find more potential Spo0A sites. Using the optimal motif estimates  $\hat{\theta}_{j,k}$  (the estimated proportion of nucleotide  $k$  in position  $j$  of the motif) and background estimates  $\hat{\theta}_{0,k}$  (the estimated proportion of nucleotide  $k$  in the background) provided by the BIOPROPECTOR/BIOOPTIMIZER procedure, we scanned all upstream (and 100 bp downstream) sequences to see whether there were additional sites that matched the motif model closely but were not strong enough to be detected by the BIOPROPECTOR/BIOOPTIMIZER procedure. In each sequence, for each potential starting position  $i$ , we had a potential site,  $S_i = (r_i, r_{i+1}, \dots, r_{i+w-1})$ , for which we computed the following score:

$$\text{Strength}_i = \log \frac{P(S_i | \text{motif})}{P(S_i | \text{background})} = \log \left( \frac{\prod_{j=1}^w \hat{\theta}_{j, r_{i+j-1}}}{\prod_{j=1}^w \hat{\theta}_{0, r_{i+j-1}}} \right)$$

Only positions in each sequence with large *Strength* values are considered as potential additional sites. However, for any additional sites found by the scanning procedure, one must be cautious about the strength of these sites, as each position will have a *Strength* value regardless of how well a site starting at that position fitted the motif model. Therefore, we also calculated a *P*-value for each position by comparing the *Strength* calculated for that position with the best *Strength* value found in each of 50 000 random sequences. The *P*-value of each potential site was then calculated as the proportion of these 50 000 random sequences that had the best *Strength* values that were larger than the *Strength* value calculated for the potential site. Only potential sites with low *P*-values (<0.10) were considered as additional sites. If a sequence already contained a site found by our BIOPROPECTOR/BIOOPTIMIZER procedure, we would expect that this same site would be found by the upstream scanning procedure. We performed this scanning procedure twice, first using the optimal 12 bp motif found by our BIOPROPECTOR/BIOOPTIMIZER procedure, and then again using our optimal 12 bp motif with the first two and last three columns removed, giving us a reduced 7 bp motif that represented the most highly conserved positions (TGTCGAA).

## Acknowledgements

We thank J. Silvaggi, S. Ben-Yehuda, A. Goranov and A. Grossman for sharing unpublished results, A. L. Sonenshein and B. Belitsky for advice on the interpretation of the *ycgM* mutant, and S. Ben-Yehuda, P. Fawcett and A. L. Sonenshein for advice on the manuscript. V.M. was a fellow of the European Molecular Biology Organization. P.E. was supported by the Swiss National Science Foundation and a Merck Core Educational Support Program. This work was supported by a NIH grant GM18568 to R.L. and NSF grant DMS-0204674 to J.L.

## Supplementary material

The following material is available from <http://www.blackwellpublishing.com/products/journals/suppmat/mmi/mmi3818/mmi3818sm.htm>

**Table S1.** Binding sequences for Spo0A as inferred by scanning are listed in columns D, G and J.

## References

- Asayama, M., Yamamoto, A., and Kobayashi, Y. (1995) Dimer form of phosphorylated Spo0A, a transcriptional regulator, stimulates the *spo0F* transcription at the initiation of sporulation in *Bacillus subtilis*. *J Mol Biol* **250**: 11–23.
- Ausubel, F., Brent, R., Kingston, R., Moore, D., Seidman, J., Smith, J., and Struhl, K. (1990) *Current Protocols in Molecular Biology*. New York: John Wiley and Sons.
- Baldus, J.M., Green, B.D., Youngman, P., and Moran, C.P., Jr (1994) Phosphorylation of *Bacillus subtilis* transcription factor Spo0A stimulates transcription from the *spolIG* promoter by enhancing binding to weak 0A boxes. *J Bacteriol* **176**: 296–306.
- Belitsky, B.R., and Sonenshein, A.L. (1998) Role and regulation of *Bacillus subtilis* glutamate dehydrogenase. *J Bacteriol* **180**: 6298–6305.
- Belitsky, B.R., and Sonenshein, A.L. (1999) An enhancer element located downstream of the major glutamate dehydrogenase gene of *Bacillus subtilis*. *Proc Natl Acad Sci USA* **96**: 10290–10295.
- Ben-Yehuda, S., and Losick, R. (2002) Asymmetric cell division in *B. subtilis* involves a spiral-like intermediate of the cytokinetic protein FtsZ. *Cell* **109**: 257–266.
- Ben-Yehuda, S., Rudner, D.Z., and Losick, R. (2003) RacA, a bacterial protein that anchors chromosomes to the cell poles. *Science* **299**: 532–536.
- Britton, R.A., Eichenberger, P., Gonzalez-Pastor, J.E., Fawcett, P., Monson, R., Losick, R., and Grossman, A.D. (2002) Genome-wide analysis of the stationary-phase sigma factor ( $\sigma^H$ ) regulon of *Bacillus subtilis*. *J Bacteriol* **184**: 4881–4890.
- Bruggemann, H., Baumer, S., Fricke, W.F., Wiezer, A., Liesegang, H., Decker, I., et al. (2003) The genome sequence of *Clostridium tetani*, the causative agent of tetanus disease. *Proc Natl Acad Sci USA* **100**: 1316–1321.
- Burbulys, D., Trach, K.A., and Hoch, J.A. (1991) Initiation of sporulation in *Bacillus subtilis* is controlled by a multicomponent phosphorelay. *Cell* **64**: 545–552.
- Eichenberger, P., Jensen, S.T., Conlon, E.M., van Ooij, C., Silvaggi, J., Gonzalez-Pastor, J.E., et al. (2003) The  $\sigma^E$  regulon and the identification of additional sporulation genes in *Bacillus subtilis*. *J Mol Biol* **327**: 945–972.
- Fawcett, P., Eichenberger, P., Losick, R., and Youngman, P. (2000) The transcriptional profile of early to middle sporulation in *Bacillus subtilis*. *Proc Natl Acad Sci USA* **97**: 8063–8068.
- Fujita, M., and Losick, R. (2003) The master regulator for entry into sporulation in *Bacillus subtilis* becomes a cell-specific transcription factor after asymmetric division. *Genes Dev* **17**: 1166–1174.
- Fujita, M., and Sadaie, Y. (1998) Feedback loops involving Spo0A and AbrB in *in vitro* transcription of the genes involved in the initiation of sporulation in *Bacillus subtilis*. *J Biochem* **124**: 98–104.
- Glaser, P., Frangeul, L., Buchrieser, C., Amend, A., Baquero, F., Berche, P., et al. (2001) Comparative genomics of *Listeria* species. *Science* **294**: 849–852.
- González-Pastor, J.E., Hobbs, E.C., and Losick, R. (2003) Cannibalism by sporulating bacteria. *Science* **301**: 510–513. Published online 19 June 2003; 10.1126/Science.1086462.
- Grossman, A.D. (1995) Genetic networks controlling the initiation of sporulation and the development of genetic competence in *Bacillus subtilis*. *Annu Rev Genet* **29**: 477–508.
- Hoch, J.A. (1993) Regulation of the phosphorelay and the initiation of sporulation in *Bacillus subtilis*. *Annu Rev Microbiol* **47**: 441–465.
- Ireton, K., Rudner, D.Z., Siranosian, K.J., and Grossman, A.D. (1993) Integration of multiple developmental signals in *Bacillus subtilis* through the Spo0A transcription factor. *Genes Dev* **7**: 283–294.
- Ivanova, N., Sorokin, A., Anderson, I., Galleron, N., Candelon, B., Kapatral, V., et al. (2003) Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. *Nature* **423**: 87–91.
- Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M., and Brown, P.O. (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409**: 533–538.
- Jensen, S.T., Liu, X.S., Zhou, Q., and Liu, J.S. (2004) Computational discovery of gene regulatory binding motifs: a Bayesian perspective. *Stat Sci* (in press).
- Jiang, M., Shao, W., Perego, M., and Hoch, J.A. (2000) Multiple histidine kinases regulate entry into stationary phase and sporulation in *Bacillus subtilis*. *Mol Microbiol* **38**: 535–542.
- Kobayashi, K., Shoji, K., Shimizu, T., Nakano, K., Sato, T., and Kobayashi, Y. (1995) Analysis of a suppressor mutation *ssb* (*kinC*) of *surOB20* (*spo0A*) mutation in *Bacillus subtilis* reveals that *kinC* encodes a histidine protein kinase. *J Bacteriol* **177**: 176–182.
- Kobayashi, K., Ehrlich, S.D., Albertini, A., Amati, G., Andersen, K.K., Arnaud, M., et al. (2003) Essential *Bacillus subtilis* genes. *Proc Natl Acad Sci USA* **100**: 4678–4683.
- Laub, M.T., Chen, S.L., Shapiro, L., and McAdams, H.H. (2002) Genes directly controlled by CtrA, a master regulator of the *Caulobacter* cell cycle. *Proc Natl Acad Sci USA* **99**: 4632–4637.

- Lawrence, C.E., Altschul, S.F., Boguski, M.S., Liu, J.S., Neuwald, A.N., and Wootton, J. (1993) Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science* **14**: 208.
- Levin, P.A., and Losick, R. (1996) Transcription factor Spo0A switches the localization of the cell division protein FtsZ from a medial to bipolar pattern in *Bacillus subtilis*. *Genes Dev* **10**: 478–488.
- Liu, J.S. (1994) The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *J Am Stat Assoc* **94**: 958–966.
- Liu, J.S., Neuwald, A.N., and Lawrence, C.E. (1995) Bayesian models for multiple local sequence alignment and Gibbs sampling strategies. *J Am Stat Assoc* **90**: 1156–1170.
- Liu, X., Brutlag, D.L., and Liu, J.S. (2001) BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac Symp Biocomput* **6**: 127–138.
- Marston, A.L., Thomaidis, H.B., Edwards, D.H., Sharpe, M.E., and Errington, J. (1998) Polar localization of the MinD protein of *Bacillus subtilis* and its role in selection of the mid-cell division site. *Genes Dev* **12**: 3419.
- Moch, C., Schrogel, O., and Allmansberger, R. (2000) Transcription of the *nfrA-ywch* operon from *Bacillus subtilis* is specifically induced in response to heat. *J Bacteriol* **182**: 4384–4393.
- Molle, V., Nakaura, Y., Shivers, R.P., Yamaguchi, H., Losick, R., Fujita, Y., and Sonenshein, A.L. (2003) Additional targets of the *Bacillus subtilis* global regulator CodY identified by chromatin immunoprecipitation and genome-wide transcript analysis. *J Bacteriol* **185**: 1911–1922.
- Nolling, J., Breton, G., Omelchenko, M.V., Makarova, K.S., Zeng, Q., Gibson, R., et al. (2001) Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J Bacteriol* **183**: 4823–4838.
- Ogura, M., Ohshiro, Y., Hirao, S., and Tanaka, T. (1997) A new *Bacillus subtilis* gene, *med*, encodes a positive regulator of *comK*. *J Bacteriol* **179**: 6244–6253.
- Perego, M., and Hoch, J.A. (2002) Two-component systems, phosphorelays, and regulation of their activities by phosphatases. In *Bacillus subtilis and its Closest Relatives: from Genes to Cells*. Sonenshein, A.L., Hoch, J.A., and Losick, R. (eds). Washington, DC: American Society for Microbiology Press, pp. 473–481.
- Perego, M., Spiegelman, G.B., and Hoch, J.A. (1988) Structure of the gene for the transition state regulator, *abrB*: regulator synthesis is controlled by the *spo0A* sporulation gene in *Bacillus subtilis*. *Mol Microbiol* **2**: 689–699.
- Perego, M., Hanstein, C.G., Welsh, K.M., Djavakhishvili, T., Glaser, P., and Hoch, J.A. (1994) Multiple protein aspartate phosphatases provide a mechanism for the integration of diverse signals in the control of development in *Bacillus subtilis*. *Cell* **79**: 1047–1055.
- Perego, M., Glaser, P., Minutello, A., Strauch, M.A., Leopold, K., and Fischer, W. (1995) Incorporation of D-alanine into lipoteichoic acid and wall teichoic acid in *Bacillus subtilis*. Identification of genes and regulation. *J Biol Chem* **270**: 15598–15606.
- Piggot, P.J., and Losick, R. (2002) Sporulation genes and intercompartmental regulation. In *Bacillus subtilis and its Closest Relatives: from Genes to Cells*. Sonenshein, A.L., Hoch, J.A., and Losick, R. (eds). Washington, DC: American Society for Microbiology Press, pp. 483–517.
- Pogliano, J., Sharp, M.D., and Pogliano, K. (2002) Partitioning of chromosomal DNA during establishment of cellular asymmetry in *Bacillus subtilis*. *J Bacteriol* **184**: 1743–1749.
- Quisel, J.D., and Grossman, A.D. (2000) Control of sporulation gene expression in *Bacillus subtilis* by the chromosome partitioning proteins Soj (ParA) and Spo0J (ParB). *J Bacteriol* **182**: 3446–3451.
- Quisel, J., Lin, D., and Grossman, A.D. (1999) Control of development by altered localization of a transcription factor in *B. subtilis*. *Mol Cell* **4**: 665–672.
- Read, T.D., Peterson, S.N., Tourasse, N., Baillie, L.W., Paulsen, I.T., Nelson, K.E., et al. (2003) The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* **423**: 81–86.
- Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., et al. (2000) Genome-wide location and function of DNA binding proteins. *Science* **290**: 2306–2309.
- Rowe-Magnus, D.A., and Spiegelman, G.B. (1998) DNA strand separation during activation of a developmental promoter by the *Bacillus subtilis* response regulator Spo0A. *Proc Natl Acad Sci USA* **95**: 5305–5310.
- Satola, S., Kirchman, P.A., and Moran, C.P., Jr (1991) Spo0A binds to a promoter used by sigma A RNA polymerase during sporulation in *Bacillus subtilis*. *Proc Natl Acad Sci USA* **88**: 4533–4537.
- Schneider, T.D., and Stephens, R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* **18**: 6097–6100.
- Shafikhani, S.H., Mandic-Mulec, I., Strauch, M.A., Smith, I., and Leighton, T. (2002) Postexponential regulation of *sin* operon expression in *Bacillus subtilis*. *J Bacteriol* **184**: 564–571.
- Shimizu, T., Ohtani, K., Hirakawa, H., Ohshima, K., Yamashita, A., Shiba, T., et al. (2002) Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. *Proc Natl Acad Sci USA* **99**: 996–1001.
- Stephenson, K., and Hoch, J.A. (2002) Evolution of signalling in the sporulation phosphorelay. *Mol Microbiol* **46**: 297–304.
- Stragier, P. (2002) A gene odyssey: exploring the genomes of endospore-forming bacteria. In *Bacillus subtilis and its Closest Relatives: from Genes to Cells*. Sonenshein, A.L., Hoch, J.A., and Losick, R. (eds). Washington, DC: American Society for Microbiology Press, pp. 483–517.
- Stragier, P., and Losick, R. (1996) Molecular genetics of sporulation in *Bacillus subtilis*. *Annu Rev Genet* **30**: 297–341.
- Strauch, M.A., and Hoch, J.A. (1993) Signal transduction in *Bacillus subtilis* sporulation. *Curr Opin Genet Dev* **3**: 203–212.
- Strauch, M.A., Spiegelman, G.B., Perego, M., Johnson, W.C., Burbulys, D., and Hoch, J.A. (1989) The transition state transcription regulator AbrB of *Bacillus subtilis* is a DNA binding protein. *EMBO J* **8**: 1615–1621.
- Strauch, M., Webb, V., Spiegelman, G., and Hoch, J.A. (1990) The Spo0A protein of *Bacillus subtilis* is a repressor of the *abrB* gene. *Proc Natl Acad Sci USA* **87**: 1801–1805.

- Strauch, M.A., Trach, K.A., Day, J., and Hoch, J.A. (1992) Spo0A activates and represses its own synthesis by binding at its dual promoters. *Biochimie* **74**: 619–626.
- Strauch, M.A., Wu, J.J., Jonas, R.H., and Hoch, J.A. (1993) A positive feedback loop controls transcription of the *spoOF* gene, a component of the sporulation phosphorelay in *Bacillus subtilis*. *Mol Microbiol* **7**: 967–974.
- Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N., *et al.* (2000) Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. *Nucleic Acids Res* **28**: 4317–4331.
- Takami, H., Takaki, Y., and Uchiyama, I. (2002) Genome sequence of *Oceanobacillus iheyensis* isolated from the Iheya Ridge and its unexpected adaptive capabilities to extreme environments. *Nucleic Acids Res* **30**: 3927–3935.
- Wach, A. (1996) PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in *S. cerevisiae*. *Yeast* **12**: 259–265.
- West, J.T., Estacio, W., and Marquez-Magana, L. (2000) Relative roles of the *fla/che* P<sub>A</sub>, P<sub>D-3</sub>, and P<sub>sigD</sub> promoters in regulating motility and *sigD* expression in *Bacillus subtilis*. *J Bacteriol* **182**: 4841–4848.
- York, K., Kenney, T.J., Satola, S., Moran, C.P., Jr, Poth, H., and Youngman, P.J. (1992) Spo0A controls the sigma A-dependent activation of *Bacillus subtilis* sporulation-specific transcription unit *spoIIIE*. *J Bacteriol* **174**: 2648–2658.
- Zhao, H., Msadek, T., Zapf, J., Madhusudan, Hoch, J.A., and Varughese, K.I. (2002) DNA complexed structure of the key transcription factor initiating development in sporulating bacteria. *Structure* **10**: 1041–1050.