doi:10.1016/j.jmb.2006.01.059 J. Mol. Biol. (2006) 358, 16–37

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

The Forespore Line of Gene Expression in *Bacillus subtilis*

Stephanie T. Wang[1], Barbara Setlow[2], Erin M. Conlon[3] Jessica L. Lyon[4], Daisuke Imamura[5], Tsutomu Sato[5] Peter Setlow[2], Richard Losick[1]* and Patrick Eichenberger[1,4]

[1]*Department of Molecular & Cellular Biology, Harvard University, Cambridge MA 02138, USA*

[2]*Department of Molecular Microbial and Structural Biology, University of Connecticut Health Center Farmington, CT 06032, USA*

[3]*Department of Mathematics and Statistics, University of Massachusetts, Amherst MA 01003, USA*

[4]*Department of Biology and Center for Comparative Functional Genomics New York University New York, NY 10003, USA*

[5]*International Agricultural Science, Tokyo University of Agriculture and Technology Fuchu, Tokyo 183-8509, Japan*

**Corresponding author*

Endospore formation by *Bacillus subtilis* involves three differentiating cell types, the predivisional cell, the mother cell, and the forespore. Here we report the program of gene expression in the forespore, which is governed by the RNA polymerase sigma factors σ^F and σ^G and the DNA-binding proteins RsfA and SpoVT. The σ^F factor turns on about 48 genes, including the gene for RsfA, which represses a gene in the σ^F regulon, and the gene for σ^G. The σ^G factor newly activates 81 genes, including the gene for SpoVT, which turns on (in nine cases) or stimulates (in 11 cases) the expression of 20 genes that had been turned on by σ^G and represses the expression of 27 others. The forespore line of gene expression consists of many genes that contribute to morphogenesis and to the resistance and germination properties of the spore but few that have metabolic functions. Comparative genomics reveals a core of genes in the σ^F and σ^G regulons that are widely conserved among endospore-forming species but are absent from closely related, but non-spore-forming *Listeria* spp. Two such partially conserved genes (*ykoU* and *ykoV*), which are members of the σ^G regulon, are shown to confer dry-heat resistance to dormant spores. The *ykoV* gene product, a homolog of the non-homologous end-joining protein Ku, is shown to associate with the nucleoid during germination. Extending earlier work on gene expression in the predivisional cell and the mother cell, we present an integrated overview of the entire program of sporulation gene expression.

Keywords: sporulation; forespore; sigma factor; transcriptional profiling; promoters

Introduction

The formation of a multicellular organism involves the generation of multiple types of specialized cells by cellular differentiation. The differentiation of each cell type is governed by its own program of gene expression and is coordinated with the differentiation of other cells by intercellular signaling pathways. A major challenge in the field of developmental biology is to comprehensively describe the programs of gene expression for all cell types in a developing organism and to elucidate the regulatory circuits to which they are subject. An attractive model system in which to undertake this challenge is endospore formation in *Bacillus subtilis* because of the relative simplicity of the organism and its amenability to the tools of traditional and molecular genetics.[1,2]

Endospore (henceforth simply spore) formation is a seven to eight hour process that is triggered by conditions of nutrient limitation.[1,2] Spore formation involves three cell types known as the predivisional cell, the forespore (or prespore) and the mother cell. Cells enter the pathway to sporulate in response to conditions of nutrient limitation, which results in the formation of the predivisional cell. Next, the predivisional cell undergoes a process of asymmetric division in which a septum is formed near

Abbreviations used: SASP, small, acid-soluble protein; FFL, feed-forward loop; GFP, green fluorescence protein.

E-mail address of the corresponding author: losick@mcb.harvard.edu

one pole, thereby creating dissimilar-sized progeny cells. These are the forespore (the smaller cell) and the mother cell. Initially, the forespore and the mother cell lie side-by-side but later in development the forespore is engulfed by the mother cell. This phagocytic-like process results in a cell-within-a-cell in which the forespore and its membrane are wholly surrounded by an outer layer of membrane derived from the engulfing mother cell. During subsequent morphogenesis, the forespore, which will become the core of the mature spore, undergoes dehydration driven in part by the replacement of water by calcium dipicolinate, and its chromosome becomes packaged by a family of small, acid-soluble proteins (SASPs) into a toroid-like structure in which it is protected against many types of DNA damage. Meanwhile, a thick layer of cell wall material known as the cortex is produced in the space between the membranes that separate the two cells, and an outer shell of protein termed the coat is deposited around the developing spore from within the mother cell. Thus, when morphogenesis is complete the spore core is encased by protective outer layers of cortex and coat material. Eventually, the spore is released by lysis of the mother cell. Upon release, the mature spore can remain dormant for long periods of time, but under favorable conditions can germinate and rapidly resume vegetative growth.

The master regulator for entry into sporulation is Spo0A, a member of the response regulator family of DNA-binding proteins.[3,4] Spo0A, which is activated by phosphorylation, orchestrates gene expression in the predivisional cell, acting both as an activator and a repressor. Some gene expression in the predivisional cell is activated indirectly by Spo0A in a pathway involving Spo0A-mediated repression of a repressor gene called *abrB*.[5–8] The product of this gene, AbrB, prevents the expression during growth of certain genes that are turned on at the start of sporulation when AbrB is depleted from cells in which Spo0A has been activated.

Approximately 121 genes, which are organized as 30 single-gene units and 24 operons, are under the direct control of Spo0A.[9] Forty of these genes are up-regulated by Spo0A and 81 are down-regulated. Interestingly, the levels of Spo0A rise gradually in the early stages of sporulation with different genes being turned on or off at different levels of phosphorylated Spo0A.[10] Thus, some genes in the regulon are low-threshold genes (requiring a low level of Spo0A to be turned on or off) and others high-threshold genes. The *abrB* gene, for example, is a low-threshold gene that is repressed at low levels of the master regulator.

Asymmetric division sets in motion two parallel but interconnected programs of gene expression.[11] The earliest-acting, cell-specific regulatory protein in the mother cell line of gene expression is the RNA polymerase sigma factor σ^E. The σ^E factor turns on the expression of about 262 genes, including the genes for GerR and SpoIIID.[12,13] These DNA-binding proteins repress the expression of many of the genes in the σ^E regulon. In addition, SpoIIID, acting in conjunction with σ^E, turns on about ten genes including genes involved in the appearance of the next regulatory protein in the mother cell line of gene expression, σ^K. The σ^K factor, in turn, switches on about 75 genes, including the gene for GerE. Finally, GerE, acting both as a repressor and an activator, switches off many of the genes in the σ^K regulon while turning on the expression of a terminal gene set of about 36 genes. Thus, the mother cell program of gene expression is governed by a hierarchical regulatory cascade consisting of a linked series of feed-forward loops (FFLs).[14–16] FFLs are wide spread regulatory motifs in which a primary regulator (such as σ^E) directs the synthesis of a secondary regulator (such as SpoIIID) and both regulatory proteins control the expression of a set of target genes.

Meanwhile, a parallel program of gene expression is played out in the forespore compartment of the sporangium. The earliest-acting regulatory protein in the forespore is σ^F, which turns on the synthesis of RsfA[17] and σ^G. RsfA is a DNA-binding protein that represses one of the genes in the σ^F regulon, *spoIIR*, a signaling gene that is involved in triggering the appearance of σ^E in the mother cell.[18,19] The σ^G factor is an activator that turns on the next set of genes in the cascade, including genes for the SASP family of proteins (see above), and the gene for the DNA-binding protein SpoVT.[20] SpoVT (as we shall see) is both a repressor of genes in the σ^G regulon as well as an activator of the terminal gene set in the forespore line of gene expression.

Finally, the forespore and mother cell lines of gene expression are linked to each other by a series of intercellular signaling pathways in which σ^F, by turning on the synthesis of the secreted signaling protein SpoIIR (above), triggers the appearance of σ^E in the mother cell and σ^G, by turning on the synthesis of the secreted signaling protein SpoIVB, triggers the appearance of σ^K in the mother cell.[11] Meanwhile, σ^E sets in motion a still poorly understood chain of events that triggers the activation of σ^G in the forespore. Thus, gene expression in the forespore and the mother cell is linked in criss-cross fashion by a successive series of signals that go back and forth from one cell to the other.

As enumerated above, gene expression in the predivisional cell and in the mother cell were elucidated previously in a comprehensive manner involving multiple complementary approaches, including transcriptional profiling during sporulation, transcriptional profiling in cells engineered to produce regulatory proteins during growth, biochemical analysis with purified regulatory proteins, chromatin immunoprecipitation, promoter identification by transcriptional start site mapping, bioinformatics and the use of gene reporters, and systematic gene inactivation. Recently, and while this work was in progress, an analysis of genes under the control of σ^F and σ^G was reported largely based on transcriptional profiling during