New Ways to Study Developmental Genes in Spore-Forming Bacteria

Philip Youngman, Peter Zuber, John B. Perkins
Kathleen Sandman, Michele Igo, Richard Losick

Summary. The regulated activation of numerous sets of genes in multiple chromosomal locations is a hallmark of cellular differentiation in both eukaryotes and prokaryotes. Certain species of bacteria that experience complex developmental cycles are especially attractive as systems in which to study the mechanisms of this kind of gene regulation because they are highly amenable to both biochemical and genetic approaches. *Bacillus subtilis*, which undergoes extensive cellular differentiation when it sporulates, is one such system. Many new methods are now available in this Gram-positive species for identifying, manipulating, and studying the regulation of genes involved in spore formation, including the use of transposable genetic elements that create gene fusions in vivo as an automatic consequence of insertions into genes.

The differentiation of cells into specialized types raises the following kinds of questions. In what ways are cells induced to differentiate? How are developmental genes regulated as temporal classes? How is gene expression restricted to differentiated cell types? In what ways are the products of developmental genes localized within cells and assembled into macromolecular structures? Among the kinds of bacteria that have proved to be especially useful for addressing such questions are the dimorphic bacteria of the genus *Caulobacter*, filamentous bacteria of the genus *Streptomyces*, social bacteria of the genus *Myxococcus*, symbiotic nitrogen-fixing bacteria of the genus *Rhizobium*, and spore-forming bacteria of the genus *Bacillus*. An essential element in the successful study of these model systems has been the creation of facile methods for applying the tools of bacterial genetics.

Differentiation in *Caulobacter crescentus* proceeds as a cell-cycle-dependent series of morphological changes, culminating in the formation of two cell types, a sessile stalked cell and a motile swarmer cell (*1*). One focus of interest in *C. crescentus* is the polar flagellum of its swarmer daughter, a cell-cycle-specific structure, whose synthesis and assembly represent a relatively simple system in which to study timing mechanisms in developmental gene expression and the role of positional information in morphogenesis. Differentiation in *Streptomyces coelicolor* is not governed by internal events, such as the cell-division cycle, but is rather an adaptive response to environmental conditions (nutrient limitation). When grown on solid media, *S. coelicolor* is fungus-like in morphology, propagating as a branched mat of multinucleate hyphae known as the substrate mycelium (*2*). The *S. coelicolor* developmental cycle involves the formation of upwardly projecting hyphae (the "aerial mycelium"), which form septa and then differentiate into chains of spore cells. This system lends itself to an analysis of how environmental cues induce cell differentiation and to investigations of the morphogenesis of complex multicellular structures. *Myxococcus xanthus* is a particularly useful system for studying the role of cell-cell communication in differentiation (*3*). During the *M. xanthus* developmental cycle, large numbers of cells coalesce into an aggregation center which, after a period of massive cell lysis, develops into a raised mound known as the "fruiting body." The differentiation of spore cells takes place in this fruiting body aggregate. A combination of genetic and biochemical approaches has shown that spore formation is dependent on the transmission of specific chemical signals between cells in the developing fruiting body. The developmental cycle of *Rhizobium meliloti* actually occurs in cooperation with the living root tissue of leguminous plants, which the bacteria invade and colonize (*4*). Once within the root tissue, the bacterial cells differentiate into "bacteroids," a cell type that is able to fix nitrogen, to the considerable advantage of the host plant. The distinctive features of the life cycles of these prokaryotic microorganisms make it possible to study different aspects of the developmental process.

The subject of this article, the Gram-positive bacterium *Bacillus subtilis*, is perhaps the most extensively studied of the differentiating prokaryotes. Its developmental cycle is an elaborate adaptive response to conditions of nutrient depletion, and involves the transformation of a vegetative cell into a complex, two-compartment sporangium within which a dormant cell type known as the endospore is produced (Fig. 1) (*5*, *6*). Development culminates with the lysis of the sporangium and the release of a mature endospore. Once triggered by nutrient limitation, the sporulation process follows a well-defined series of morphological events that is correlated with the activation of many gene sets in a highly ordered temporal progression that lasts for 7 to 10 hours. Indeed, it is these two aspects of sporulation—initiation by nutrient limitation and the subsequent playing out of a temporal program of gene expression—that have been and still are the principal topics of experimental interest. Several other features of the developmental cycle of this organism are also becoming accessible for study, such as differential gene expression in the two sporangium compartments and morphogenesis of the multilayered coat that encases the mature spore.

Although the various bacterial model systems are diverse in their biological features and in the aspects of cellular differentiation that are being studied, strategies for investigating the underlying mechanisms of gene regulation in these different kinds of bacteria have much in common. Of universal importance in these systems is the need to

Philip Youngman is a research associate, Peter Zuber and John Perkins are postdoctoral fellows, Kathleen Sandman and Michele Igo are graduate students, and Richard Losick is professor and chairman in the Department of Cellular and Developmental Biology, The Biological Laboratories, Harvard University, Cambridge, Massachusetts, 02138.

identify the regulated genes involved in the differentiation events and the need to analyze both in vitro and in vivo the factors that govern expression of such genes. Using the study of sporulation in *B. subtilis* as an example, we discuss in this article the kinds of genetic and biochemical approaches available in many bacteria for locating and analyzing genes subject to developmental regulation. We also consider how the knowledge of specific regulated genes can lead to an understanding of the regulatory mechanisms themselves.

Developmental Genes and Their Regulation

Genes whose products are specifically required for endospore formation in *B. subtilis* are called *spo* genes (*7*). Such genes have been identified and defined by mutations that impair development but that have little or no effect on vegetative growth. Since the pioneering work of Schaeffer and Ionesco and Spizizen (*8*), hundreds of *spo* mutations have been isolated, and they have been grouped into more than 50 genetic loci on the basis of their positions on the chromosome and the stages at which they arrest development (Fig. 2); this is undoubtedly an underestimate of the number of genes involved in sporulation (*6*). Moreover, because pairwise complementation tests have been performed in very few cases, it is possible that many *spo* "loci" actually consist of several cistrons (or even multiple operons) that govern the same morphological event. Mutations at particular *spo* loci usually arrest development at a definite stage, and the loci are named according to the stage of blockage. For example, mutations at eight to ten loci called *spo0* prevent the formation of the polar septum that partitions the sporangium into mother cell and forespore compartments, the first definite morphological event of the sporulation sequence (Fig. 1). Mutations at eight known *spoII* loci allow septum formation, but they block the subsequent engulfment step by which an embryonic endospore (the forespore protoplast) is produced. Mutations in at least six *spoIII* loci allow development to proceed up to the stage at which a forespore protoplast is formed within the mother cell, but interfere with the subsequent synthesis or assembly of cortex or coat material (or both). Mutations that block at later stages (4 and 5) have also been isolated, as have mutations that impair spore germination (*ger*) and outgrowth (*out*) (Figs. 1 and 2).

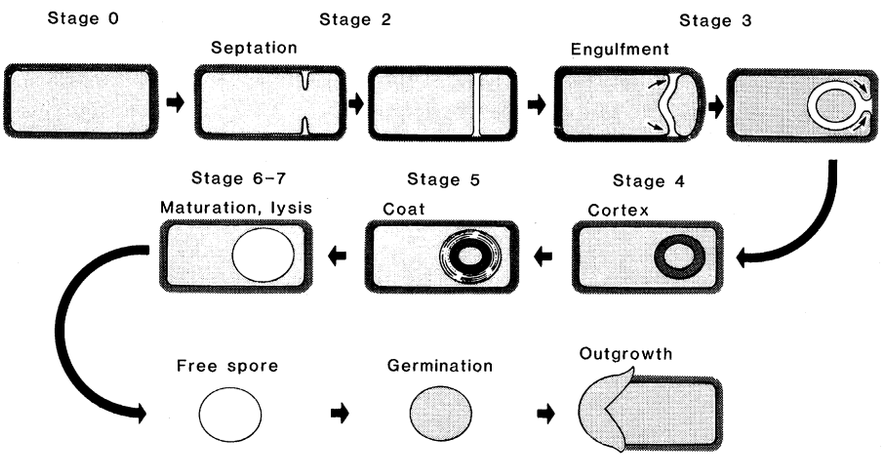

Fig. 1. The stages of sporulation and germination (*6*).

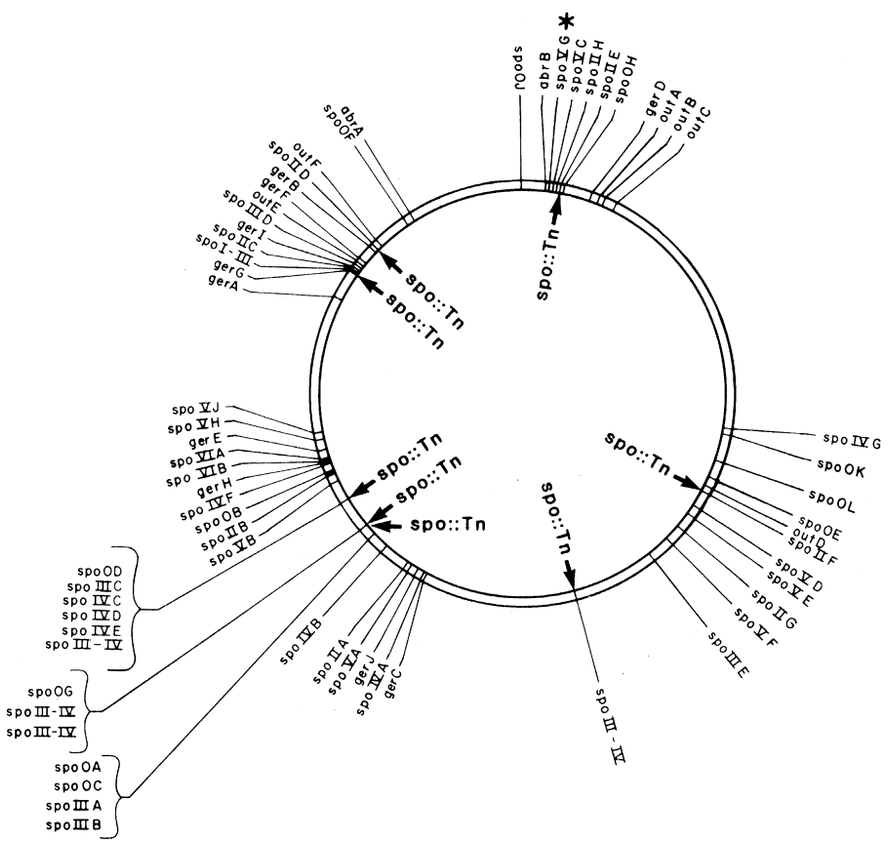

Fig. 2. A genetic map of the *B. subtilis* chromosome showing the positions of mutations that impair sporulation or germination, including eight Tn917-generated *spo* mutations.

It has become apparent through the genetic analysis of *spo* mutations that each stage of endospore formation is determined by the expression of many genes located at many chromosomal positions, and it has been widely assumed that the stage at which a particular mutation causes the arrest of sporulation reflects the time during spore development at which the mutated gene is active. While this assumption may often be valid, the information now available in a few well-studied cases suggests that Spo phenotypes are not always a reliable guide to the time of gene expression. It is clearly desirable, in any event, to determine by direct measurements of transcription the precise temporal pattern of activity for particular *spo* genes; indeed it is only after *spo* genes (or a large proportion of them) are sorted into distinct classes of temporal regulation that we can approach the question of how each class is regulated.

Study of Cloned *spo* Genes

One way to assess the activity of a sporulation gene quantitatively (at the

level of transcription) is to use a cloned copy of the gene itself as a hybridization probe to measure the presence of the messenger RNA (mRNA) in the bacteria at different times during development. Such experiments can even distinguish between the transcripts that originate from different promoters (through the method of nuclease S1-protection mapping). Cloned copies of a gene can also serve as templates for in vitro studies of transcription, and thus create the possibility of analyzing biochemically the factors required for proper gene expression and regulation. Moreover, cloned genes can be mutated or otherwise altered in vitro (for the construction of gene fusions, for example) and then returned to the chromosome for analysis of their activity in vivo. The first developmentally regulated gene of *B. subtilis* to be isolated by molecular cloning and then subjected to this kind of analysis is a gene called *spoVG*, and the extensive and unexpected insights revealed through the study of this single gene illustrate the great utility of such experiments.

spoVG is a small sporulation gene (400 base pairs) that maps near the replication origin of the *B. subtilis* chromosome (asterisk in Fig. 2) (*9*). Transcription of *spoVG* is induced at the onset of sporulation, although its protein product appears to be required at later stages of development. Activation of *spoVG* must somehow be coupled to the initiation phase of sporulation, since its transcription depends on the products of a class of sporulation genes known as *spo0*, mutations of which prevent *B. subtilis* cells from entering the sporulation pathway (*10*).

An unusual feature of *spoVG* is the complexity of its transcription initiation region, which is actually composed of two overlapping promoters that are separately utilized by forms of RNA polymerase that contain different species of sigma factor, promoter recognition elements that are responsible for the selection of transcription initiation sites by RNA polymerase (*11*). Recognition of the upstream promoter P1 is determined by a 37,000-dalton sigma species known as σ^{37}, whereas a 32,000-dalton sigma species σ^{32} dictates utilization of the downstream promoter P2. As is the case for promoter selection by other forms of RNA polymerase holoenzyme, σ^{37}- and σ^{32}-RNA polymerases are believed to recognize promoter sites through contacts made by polymerase (and perhaps by the sigma subunits themselves) with conserved nucleotide sequences centered about 35 and 10 base pairs (−35 and −10, respectively) upstream from their respective transcription start points.

In the case of the overlapping *spoVG* promoters, whose start points are separated by ten base pairs, these −35 and −10 recognition sequences are thought to be arranged in alternating order, creating the mosaic structure shown in Fig. 3. A further complicating feature of *spoVG* is the existence of an upstream AT-rich "box" (A, adenine; T, thymine) (Fig. 3), which strongly enhances transcription from both the P1 and P2 start sites (*12*). The AT-rich box is composed largely of alternating stretches of A's and T's, a structural feature now known to be associated commonly with the upstream regions of strongly utilized promoters in *B. subtilis*.

The σ^{37}- and σ^{32}-RNA polymerases were discovered during studies in which cloned *spoVG* DNA was used as a transcription template in vitro (*11, 13*). These experiments were designed to detect biochemical components of the transcription apparatus required for the utilization of the *spoVG* promoter. Similar transcription experiments with other well-defined cloned DNA's as templates have led to the discovery of at least five holoenzyme forms in *B. subtilis* (*14, 15*), only one of which (σ^{55}-RNA polymerase) is homologous in specificity to the principal holoenzyme form of *Escherichia coli*, a Gram-negative bacterium in which evidence for RNA polymerase heterogeneity has only recently been reported (*16*). An important theme to emerge from the use of cloned DNA's as templates for in vitro RNA synthesis is that the activity of different sets of genes (including, in particular, developmental genes) may be determined, in part, by a spectrum of RNA polymerase holoenzyme forms whose promoter recognition specificities are intrinsic to their structures (*6, 14*).

Use of Gene Fusions to Study *spo* Gene Regulation

Given the existence in *B. subtilis* of multiple RNA polymerase forms that display distinct promoter specificities, we might have supposed that the regulated activation of different gene sets during sporulation might reflect simply the regulated appearance of particular sigma species specific for the promoters of those gene sets; this would be analogous on a grand scale to the temporal gene regulation program of *B. subtilis* phage SP01 (*17*). In fact, such a mechanism may account for the activation of some genes that are expressed during middle-to-late stages of sporulation, such as genes transcribed by σ^{29}-associated holoenzyme (*15, 18*). However, in the case of *spoVG*, as well as any other sporulation gene transcribed by σ^{32} or σ^{37}-associated holoenzyme, some other mechanism must be responsible for gene activation. The σ^{37}- and σ^{32}-RNA polymerase forms are present throughout the vegetative phase of development, a time when *spoVG* is relatively silent; and these polymerase forms do not increase in abundance at early times during spore development; that is, at the time when *spoVG* is induced to high levels of expression. Moreover, mutations in *spo0* loci that impair *spoVG* transcription do not detectably reduce the levels of σ^{37} and σ^{32} enzymes. Thus, superimposed on the complex transcriptional features of *spoVG*—namely, its overlapping promoters and the AT-rich box that enhances their utilization—are mechanisms for coupling its expression to events occurring at the onset of sporulation. These mechanisms probably involve one or more *spo0* gene products or a protein under their control, which must act in conjunction with σ^{37}- and σ^{32}-associated forms of RNA polymerase to stimulate *spoVG* RNA synthesis, for example, by binding to the promoter region.

One way of identifying the regulatory factors required for induction of *spoVG* would be through the analysis of mutations and suppressors of mutations isolated on the basis of their effects on the expression of *spoVG* in vivo. In order to make such an analysis possible, the

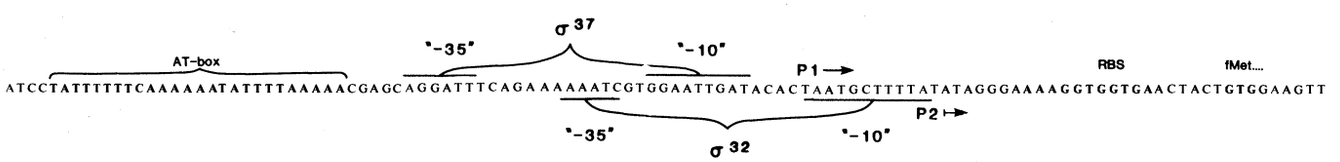

Fig. 3. The transcription initiation region of *spoVG*.

spoVG coding sequence was "fused" to the *E. coli* structural gene that encodes β-galactosidase, the *lacZ* gene (*19*). This in essence replaces the *spoVG* gene product with β-galactosidase, and permits the semi-quantitative assessment of *spoVG* expression (transcription and translation) in colonies of bacteria on agar plates that include the chromogenic substrate X-gal (5-bromo-4-fluoro-3-indolyl-β-D-galactoside). In one such construction, a 157–base pair segment from the transcription and translation initiation region of *spoVG* was fused in-frame to a *lacZ* coding sequence that has no promoter, and the resulting gene fusion was inserted into the temperate *B. subtilis* phage SPβ (*20*). This created a "portable" developmental promoter that could be conveniently moved into various genetic backgrounds (*spo0* mutants) simply by specialized transduction, and whose activity could be conveniently monitored with colorimetric substrates for β-galactosidase. Through the use of such an SPβ::*spoVG-lacZ* transducing phage, it was possible to demonstrate that a small promoter-containing segment of *spoVG* contained all the information necessary for replicating the normal pattern of *spo0*-dependent activation of *spoVG* (*20*) (Fig. 4, filled circles). Thus, in at least this particular case, developmental regulation is exerted at or near the region of transcription initiation.

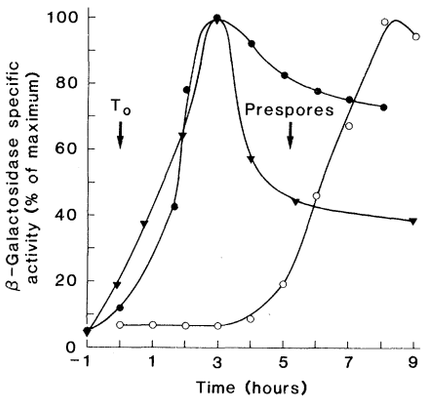

Fig. 4. Temporally regulated expression of in vitro constructed *spoVG-lacZ* (●) and *0.3 kb-lacZ* (○) gene fusions, and of a *spo* gene (▼) interrupted by an insertion of Tn917*lac*; all were assayed by measurement of β-galactosidase specific activity. T_0 marks the end of log-phase growth and the beginning of sporulation.

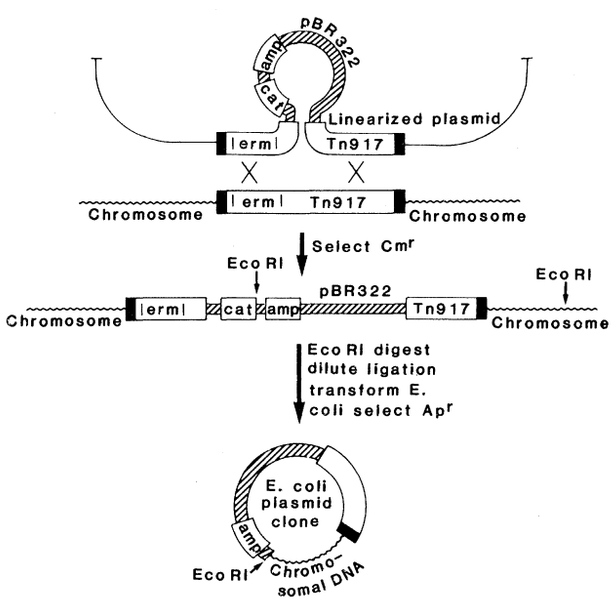

Fig. 5. The integration of pBR322 sequences by recombination into a chromosomal copy of Tn917 (*amp*, the Ap^r-conferring β-lactamase gene of pBR322; *cat*, a Gram-positive Cm^r-conferring gene). Competent bacteria containing a chromosomal insertion of Tn917 are transformed to Cm^r with linearized plasmid DNA (*32*).

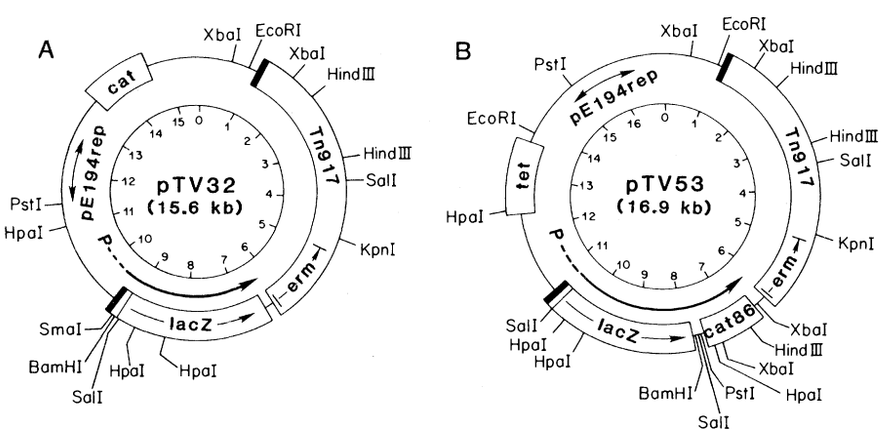

Fig. 6. Restriction maps of pTV32 (A) and pTV53 (B), two temperature-sensitive plasmids used to select for chromosomal insertions of Tn917*lac* and Tn917*cat-lac*.

Several other developmentally regulated promoters are becoming accessible to the kinds of analysis outlined above. For example, the *B. subtilis* serine exoprotease gene *sprE*, whose gene product is a classic marker for *spo0*-dependent events occurring at the onset of sporulation, is now known to be transcribed in vitro by σ^{37}-RNA polymerase, and its regulation in vivo is being studied through the use of an *sprE-cat* operon fusion (*21*). The *B. subtilis 0.3-kb* gene (*22*), *spoIIC* and *spoIIIC* genes (*18*), the *B. pumilus cat-86* gene (*23*), the *B. megaterium* genes encoding acid-soluble polypeptide components of the spore core (*24*), and the gene encoding the parasporal crystal toxin of *B. thuringiensis* (*25*) are all known to be associated with promoters whose utilization begins at intermediate stages of sporulation, and the regulation of these genes is a very active area of investigation. (The temporally delayed activation of the *0.3 kb* promoter as detected by means of a *lacZ* transcriptional fusion is shown in Fig. 4.) Even though they were cloned from diverse species, these genes are expressed in a developmentally appropriate fashion when introduced into *B. subtilis*, where their expression can be shown to exhibit a complex pattern of dependence on the products of other *spo* loci, and where their analysis has confirmed that *spo* genes comprise many distinct temporal expression classes.

Use of Transposons to Identify and Clone *spo* Genes

As explained above, the usual way of identifying genes whose products are involved in sporulation is to rely on mutations that block the sporulation process. These mutations can in turn be used to

clone *spo* genes, through the identification of *B. subtilis* phage or plasmid clones that complement a sporulation deficiency (*26*) or through the identification of phage lambda recombinants that can transform competent cells of a Spo^- mutant to a Spo^+ phenotype (*18, 27*). To facilitate the identification of *spo* genes and to simplify subsequent manipulations such as cloning, a system of transposon-mediated insertional mutagenesis has been developed in *B. subtilis* (*28*). This system makes use of a transposon called Tn917, which originates from the Gram-positive genus *Streptococcus* (*29*). The Tn917 transposon confers erythromycin resistance (Em^r), is 5.3 kilobases in length, and has terminal inverted repeats similar in sequence to those of transposons in the Tn3 family of Gram-negative insertion elements (*30*).

Because genes from Gram-negative bacteria are generally expressed very poorly in *B. subtilis*, it was not possible to use in this Gram-positive system the well-characterized transposons of *E. coli*, such as Tn5, Tn10, and Mu, which can generate a broad spectrum of insertional mutations. When Tn917 was introduced into *B. subtilis*, its ability to act as a "random" chromosomal mutagen was untested, and a study of its insertion-site specificity was necessary. It has since been established that the transposon displays a relatively high degree of target-site randomness (*31*). For example, of eight independently derived insertional *spo* mutations investigated thus far, all have proved to represent insertions in different genes (Fig. 2).

To make it possible to clone in *E. coli* the *B. subtilis spo* genes that can be identified by Tn917-mediated insertional mutagenesis, procedures have been developed (*32*) for integrating derivatives of the *E. coli* cloning vector pBR322 into the *B. subtilis* chromosome at the site of any existing Tn917 insertion by means of the "marker-replacement" strategy described in the legend to Fig. 5. This permits the cloning of *B. subtilis* chromosomal DNA adjacent to Tn917 in *E. coli* simply by transforming *E. coli* cells with "circularized" chromosomal DNA fragments (that is, restriction endonuclease digests of chromosomal DNA ligated at dilute concentration) and then selecting for pBR322-conferred resistance to ampicillin (Ap^r).

In addition to facilitating the cloning of DNA, a transposon insertion "tags" the mutated gene with a selectable marker, and thus greatly simplifies many routine genetic procedures in bacteria including mapping, the transfer of mutations with nonselectable phenotypes (like *spo* mutations) to different genetic backgrounds, the construction of partial diploids, and the isolation of second-site suppressor mutations (*33*). This has motivated the introduction of transposons into many different species, including differentiating members of the *Myxococcus* (*34*), *Caulobacter* (*35*), and *Rhizobium* (*36*) genera, where transposons have been successfully used for the identification and manipulation of developmental genes.

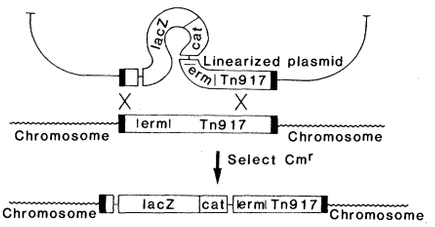

Fig. 7. The conversion of a simple insertional mutation into a *lacZ* fusion by transformation with linearized plasmid DNA containing a Tn917 derivative substituted with a promoterless *lacZ* gene and a *cat* gene that has its own promoter.

Transposon-Mediated Gene Fusions

Perhaps the single most powerful application of transposons for the analysis of gene expression and regulation in vivo is their use to generate transcriptional gene fusions as an automatic consequence of insertion into a gene, an application pioneered by Casadaban and Cohen (*37*) with their Mud(Ap^r,*lac*) derivatives of the transposon-phage Mu. The Mud(Ap^r,*lac*) elements are defective derivatives of Mu engineered to carry a promoterless copy of the *E. coli* lactose operon, near one end of the Mu genome, oriented such that insertions into a chromosomal transcription unit can place the *lac* operon under control of the promoter and other regulatory elements of the interrupted transcription unit.

Several groups of investigators, including some interested in developmental gene regulation in *C. crescentus* and *M. xanthus* (*38*), have recently taken advantage of the fact that any transposon can, in principle, be made to function like Mud(Ap^r,*lac*). The key requirement is to identify a site near one end of the transposon where foreign DNA (such as a promoterless *lacZ* gene) can be inserted without disrupting genes or sequences important for transposition activity. A thorough physical analysis of Tn917 resulted in the identification of such a site in this transposon, an Hpa I site located approximately 270 bp from its *erm*-proximal end (*30*) (*erm* is the erythromycin resistance determinant of Tn917). After this Hpa I site was modified to facilitate its use for cloning (*39*), a promoterless *lacZ* gene was inserted into Tn917 in an orientation that made its expression depend on transcription coming into the transposon from the outside, thus creating the fusion-generating transposon Tn917*lac* contained in plasmid pTV32 (Fig. 6A). In this construction, the *lacZ* gene was furnished with a ribosome-binding site appropriate for the efficient initiation of translation in *B. subtilis*. Such *lacZ*-carrying derivatives of Tn917 are unimpaired in their transposition and create transcriptional *lacZ* fusions when inserted into a chromosomal gene in the appropriate orientation (*31*).

Thus, in order to make fusions to genes whose regulation we would like to study, it is no longer necessary to go through the laborious process of cloning, making the fusion construction in vitro, and returning the fused gene to the chromosome without rearrangement. Instead, it is simply a matter of making appropriate transposon-mediated insertional mutations. This is such a reduction in labor that it is now realistic to consider recovering a sufficiently large collection of insertion-generated *spo* gene fusions to expect representation of all important temporal classes. Many such transposon-generated fusions have already been recovered, in preparation for analysis of the temporal classes they represent. The regulation of these temporal classes might then be analyzed by transferring the transposon-mediated fusions into various genetic backgrounds (by transformation) and then assessing the dependence of expression of the *spo*::Tn917*lac* fusion in question on the products of other known *spo* loci. In addition, such gene fusions should be very useful for the identification of new mutations that affect their regulation, such as second-site mutations that produce bluer or whiter colonies on X-gal plates. A potential simplifying aspect of this kind of analysis is that even studies involving a single representative of a given temporal class may provide important insights into the regulation of the whole class.

Because of the efficiency with which naturally "transformation-competent" *B. subtilis* bacteria can take up linear DNA strands and incorporate them by homologous recombination into host sequences, it is even possible in this system to turn already-existing *spo*::Tn917 insertions into *lacZ* transcriptional gene fusions (Fig. 7). The creation of *lacZ*

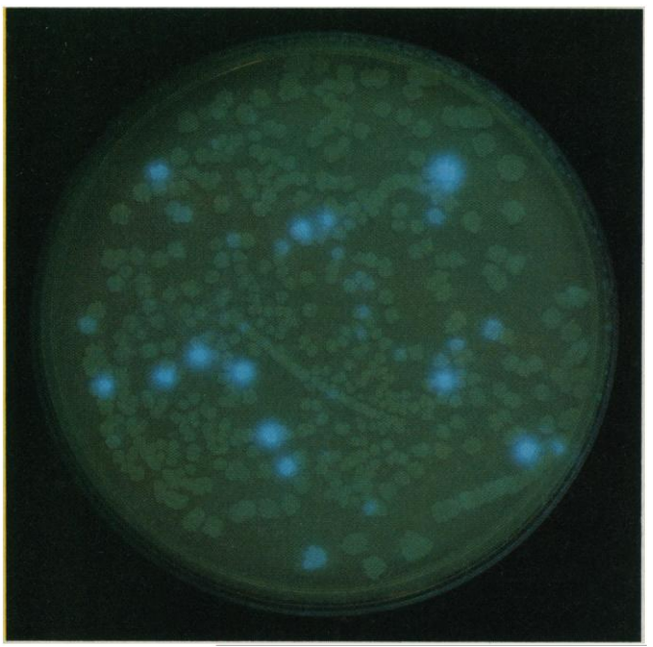

Fig. 8. Fluorescence in colonies of *B. subtilis* bacteria containing different chromosomal insertions of Tn917*lac*. The petri dish was sprayed (15 minutes before being photographed) with the β-galactosidase substrate 4-methylumbelliferyl-β-D-galactoside, which generates a highly fluorescent hydrolysis product. Bright colonies represent insertions into genes that are active during sporulation.

fusions in this fashion is possible whenever the existing transposon insertion is appropriately oriented relative to the promoter of the disrupted *spo* gene (theoretically 50 percent of all insertions), and thus should permit a more versatile analysis of many of our previously characterized *spo*::Tn917 mutations.

Identifying Developmentally Regulated Genes by Expression

Another important application of transposons that generate gene fusions is the identification of genes on the basis of expression or change of expression under specific conditions. One successful example was the use of Mud(Ap^r,*lac*) elements to identify genes of *E. coli* that become active in the "SOS" response to DNA damage (*40*). Two strategies have been developed in the *B. subtilis* system for identifying genes that become active during spore formation.

One approach involves a transposon that simultaneously generates an operon fusion to both *lacZ* and *cat* (Fig. 6B) (*cat* is a drug resistance gene that encodes chloramphenicol acetyltransferase). In this transposon, Tn917*cat-lac*, a promoterless *lacZ* gene and a promoterless *cat* gene (from *B. pumilus*) are arranged as a tandem pair, with no transcription terminator between them. Thus, insertions that activate the transposon-borne *cat-lac* "operon" (detected by blueness on X-gal plates) can be evaluated with a simple phenotypic test (sensitivity to chloramphenicol when cells are growing exponentially) to determine whether expression of the fusion is restricted to stationary phase (the period during which *spo* genes are active). Insertions that give rise to blue colonies that are sensitive to chloramphenicol identify genes whose activity is restricted to the post-vegetative phase of the life cycle.

Another approach is to use a β-galactosidase indicator in a way that provides a "real-time" assessment of gene activity. Promising results have been obtained with the indicator substrate 4-methylumbelliferyl-β-D-galactoside (MUG) (*41*). When hydrolyzed by β-galactosidase, MUG is converted to 4-methylumbelliferone, which is so highly fluorescent under long-wavelength ultraviolet light that it is easily detected even when present in very small amounts. For example, the amount of MUG substrate hydrolyzed within 10 minutes after application by a colony expressing a *lacZ* fusion at moderate levels may be readily visualized by its fluorescence, even when the observer is not "dark-adapted." (Hydrolysis of the commonly used chromogenic substrate X-gal cannot be detected with sufficient sensitivity for it to be used in this way.)

Thus, when a population of *B. subtilis* bacteria containing Tn917*lac*-generated inserts is plated under conditions that favor sporulation, and then sprayed with a solution of MUG *after* colonies have developed to a stage where growth has virtually ceased and spores are forming (approximately 24 hours), colonies of bacteria that contain fusions of *lacZ* to genes active during sporulation are immediately apparent from their fluorescence under a hand-held ultraviolet lamp (Fig. 8). For example, Fig. 4 (triangles) shows that one brightly fluorescent colony identified in this way contains an insertion in a gene that is relatively silent during growth, but is induced to much higher levels of expression in the stationary phase. Interestingly, a high proportion of insertions that generate bright colonies actually turn out to cause developmental abnormalities—in fact, the example in Fig. 4 was found to cause an insertional *spo* mutation—although an important advantage of this approach is that it enables us to identify genes that are active during sporulation regardless of whether their products are required for spore formation.

Conclusion

One challenge to understanding the way in which differentiating cells, such as sporulating bacteria, regulate the expression of their genes is the extreme complexity of such differentiation events. Many different kinds of gene products are involved, which function in many different structural and regulatory roles. Once the regulated genes can be identified, however, powerful genetic and biochemical methods can be made available for their analysis. One advantage of many bacterial cell differentiation model systems, including the *B. subtilis* model, is the existence of rapid and convenient procedures for identifying, isolating, and manipulating the regulated genes of interest. The ease with which genes may be identified and studied in these systems makes the understanding of even complex regulatory networks a realistic objective.

References and Notes

1. B. Ely and L. Shapiro, in *Microbial Development*, R. Losick and L. Shapiro, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1984), p. 1.
2. K. Chater, *ibid.*, p. 89.
3. D. Kaiser, *ibid.*, p. 197.
4. F. Ausubel, *ibid.*, p. 275.
5. J. Hoch, in *Advances in Genetics*, E. W. Caspari, Ed. (Academic Press, New York, 1976), p. 67; R. H. Doi, *Annu. Rev. Genet.* **11**, 29 (1977).
6. R. Losick and P. Youngman, in *Microbial Development*, R. Losick and L. Shapiro, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1984), p. 63; R. H. Doi, *Arch. Biochem. Biophys.* **214**, 772 (1982).
7. P. J. Piggot and J. G. Coote, *Bacteriol. Rev.* **40**, 908 (1976); P. J. Piggot, A. Moir, D. A. Smith, in *Sporulation and Germination*, H. Levinson, A. L. Sonenshein, D. J. Tipper, Eds. (American Society for Microbiology, Washington, D.C., 1981), p. 29.
8. P. Schaeffer and H. Ionesco, *C. R. Acad. Sci.* **251**, 3125 (1960); J. Spizizen, in *Spores III*, H. O. Halvorson, Ed. (Burgess, Minneapolis, Minn., 1961), p. 142.
9. J. F. Ollington, W. G. Haldenwang, T. V. Huynh, R. Losick, *J. Bacteriol.* **147**, 432 (1981); A. Rosenbluh, C. D. B. Banner, R. Losick, P. C. Fitz-James, *ibid.* **148**, 341 (1981).
10. J. A. Hoch, M. A. Shiflett, J. Trowsdale, S. M. H. Chen, in *Spores VII*, G. Chambliss and J. C.

Vary, Eds. (American Society for Microbiology, Washington, D.C. 1978), p. 131.
11. W. C. Johnson, C. P. Moran, Jr., R. Losick, *Nature (London)* **302**, 800 (1983).
12. C. D. B. Banner, C. P. Moran, Jr., R. Losick, *J. Mol. Biol.* **168**, 351 (1983).
13. W. G. Haldenwang and R. Losick, *Nature (London)* **282**, 256 (1979).
14. J. L. Wiggs, M. Z. Gilman, M. Chamberlin, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2762 (1981); M. Z. Gilman and M. Chamberlin, *Cell* **35**, 285 (1983).
15. W. G. Haldenwang, N. Lang, R. Losick, *Cell* **23**, 615 (1981).
16. R. Lardick, V. Vaughn, E. T. Lau, R. A. Van Bogeln, J. W. Erickson, F. C. Neidhardt, *ibid.* **38**, 175 (1984); A. Grossman, J. W. Erickson, C. Gross, *ibid.*, p. 383.
17. J. Pero, J. Nelson, T. D. Fox, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 1589 (1975); J. Duffy and E. P. Geiduschek, *Nature (London)* **270**, 28 (1977); R. Tjian and J. Pero, *ibid.* **262**, 753 (1976); G. Spiegelman and H. R. Whiteley, *J. Biol. Chem.* **249**, 1476 (1974); C. Talkington and J. Pero, *Virology* **83**, 365 (1978).
18. M. S. Rosenkrantz, S. Rong, A. L. Sonenshein, in preparation; J. Trempy, D. Plummer, W. G. Haldenwang, in preparation.
19. J. H. Miller, W. S. Reznikoff, A. E. Silverstone, K. Ippen, E. R. Signer, J. R. Beckwith, *J. Bacteriol.* **104**, 1273 (1970); M. J. Casadaban and S. N. Cohen, *J. Mol. Biol.* **138**, 198 (1980); C. E. Donnelly and A. L. Sonenshein, in *Molecular Cloning and Gene Regulation in Bacilli*, A. T. Ganesan, S. Chang, J. A. Hoch, Eds. (Academic Press, New York, 1984), p. 63.
20. P. Zuber and R. Losick, *Cell* **35**, 275 (1983).
21. D. S. Goldfarb, S.-L. Wong, T. Kudo, R. Doi, *Mol. Gen. Genet.* **191**, 319 (1983); S.-L. Wong, C. Price, D. S. Goldfarb, R. Doi, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1184 (1984).
22. M. Stephens, N. Lang, K. Sandman, R. Losick, *J. Mol. Biol.* **176**, 333 (1984).
23. S. Mongkolsuk and P. Lovett, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3457 (1984).
24. S. Goldrick and P. Setlow, *J. Bacteriol.* **155**, 1459 (1983); J. Mason and P. Setlow, *ibid.* **157**, 931 (1984).
25. H. C. Wong, H. E. Schnepf, H. R. Whiteley, *J. Biol. Chem.* **258**, 1960 (1983); A. Klier, F. Fargotte, G. Rapoport, *EMBO J.* **1**, 791 (1982).
26. H. Hirochika, Y. Kobayashi, F. Kawamura, H. Saito, *J. Bacteriol.* **146**, 494 (1981); J. Weir, E. Dubnau, N. Ramakrishna, I. Smith, *ibid.* **157**, 405 (1984).
27. E. Ferrari, D. J. Henner, J. A. Hoch, *ibid.* **146**, 430 (1981).
28. P. J. Youngman, J. B. Perkins, R. Losick, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2305 (1983).
29. P. Tomich, F. Y. An, D. B. Clewell, *J. Bacteriol.* **141**, 1366 (1980).
30. J. B. Perkins and P. Youngman, *Plasmid* **12**, 119 (1984).
31. P. Youngman, J. B. Perkins, K. Sandman, in *Genetics and Biotechnology of Bacilli*, A. T. Ganesan and J. A. Hoch, Eds. (Academic Press, New York, 1984).
32. P. Youngman, J. B. Perkins, R. Losick, *Mol. Gen. Genet.* **195**, 424 (1984).
33. N. Kleckner, J. Roth, D. Botstein, *J. Mol. Biol.* **116**, 125 (1977).
34. J. Kuner and D. Kaiser, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 425 (1981).
35. B. Ely and R. Croft, *J. Bacteriol.* **149**, 620 (1982).
36. J. E. Beringer, J. L. Beyon, A. V. Buchanan-Wallaston, A. W. B. Johnston, *Nature (London)* **276**, 633 (1978); R. Simon, U. Priefer, A. Puhler, *Biotechnology* **1**, 784 (1983).
37. M. Casadaban and S. Cohen, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4530 (1979).
38. V. Bellofatto, L. Shapiro, D. Hodgson, *ibid.* **81**, 1035 (1984); L. Kroos and D. Kaiser, personal communication.
39. P. Youngman, J. B. Perkins, R. Losick, *Plasmid* **12**, 1 (1984).
40. C. Kenyon and G. C. Walker, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2816 (1980).
41. The use of MUG on agar plates as an indicator substrate for β-galactosidase was suggested to us by David Gelfand of the Cetus Corporation.

RESEARCH ARTICLE

Redesigning Trypsin: Alteration of Substrate Specificity

Charles S. Craik, Corey Largman, Thomas Fletcher
Steven Roczniak, Philip J. Barr
Robert Fletterick, William J. Rutter

Studies relating the structure of the serine proteases to their functions as catalysts of amido and ester bond cleavage have influenced current ideas concerning the mechanism of enzyme action. X-ray crystallographic analyses of eukaryotic and prokaryotic members of this family of enzymes (*1, 2*) have revealed a common three-dimensional structure. The results of structural analysis and ancillary data on selective chemical modification have revealed the location and function of essential residues for zymogen activation (*3*) and catalysis (*4*). The catalytic mechanism appears to be conserved among the members of this extensive family. The diverse activities of the serine proteases are the result of the different sets of amino acids that are utilized by each enzyme for substrate binding.

The genes and complementary DNA's (cDNA's) for most of the pancreatic serine proteases, including trypsin (*5, 6*), chymotrypsin (*7*), elastase (*8*), and kallikrein (*9*), have been cloned and sequenced. We elected to substitute, delete, or insert the codons for specific amino acids in the coding regions of these genes. Subsequent expression of the modified genes and characterization of the amino acid–altered gene products provide means to evaluate critically the postulates about specific amino acid side chains as they relate to catalysis, substrate specificity, and zymogen activation of the serine proteases. Using this system, we have designed and partially characterized three trypsin mutants that influence substrate specificity. Two of the mutants may assume a zymogen-like conformation in the absence of substrate.

Modeling wild-type and mutant rat trypsins. Although the three-dimensional structure for rat pancreatic trypsin II is not known, the primary structure has 74 percent identity with bovine cationic trypsin (*5, 6*), whose three-dimensional crystal structure is known (*1*). Of the 57 differences in amino acids, 48 are at the solvent-accessible surface, 5 are conservative changes within the internal structure, and the other 4 (88 Ser→Ile; 112 Ala→Val; 181 Phe→Val; 183 Ala →Val) are nonconservative (*10*). Active site residues His^{57}, Asp^{102}, and Ser^{195} of rat and bovine trypsin are conserved and located in regions with sequence identity. Similarly, Asp^{189} at the base of the substrate-binding pocket, which is thought to confer the substrate specificity for arginyl and lysyl side chains upon trypsin, is present in the rat sequence. The glycine residues at positions 216 and 226, which seem to permit entry of large amino acid side chains into the hydrophobic pocket, are also conserved. Indeed, when the structure of rat trypsin is compared with that of bovine trypsin complexed with pancreatic trypsin inhibitor (*11*) (with Lys^{15} of the inhibitor in the substrate-binding pocket), there are no substitutions within 7.6 Å of the substrate side chain. We therefore used the bovine trypsin structures complexed with either the pancreatic trypsin inhibitor or benzamidine [an arginine analog that is an accurate structural model for arginine (*12, 13*)] as a model for the rat trypsin complexed with lysine and arginine substrates, respectively.

C. S. Craik, S. Roczniak, and W. J. Rutter are members of the Hormone Research Institute and the Department of Biochemistry and Biophysics, and R. Fletterick is a member of the Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143. P. J. Barr is a member of Chiron Corporation, 4560 Horton Street, Emeryville, California 94608. C. Largman and T. Fletcher are members of the Biochemistry Research Laboratories, Martinez Veterans Administration Medical Center, Martinez, California 94553. Correspondence should be sent to C. S. Craik.