5
Genetic Engineering in *Bacillus subtilis*

ROY H. DOI

Department of Biochemistry and Biophysics, University of California, Davis, California 95616, USA

Introduction

Bacillus subtilis has the potential for becoming an extremely useful system for the cloning and expression of heterologous genes. Some of the practical advantages of this organism have been noted previously (Dean and Kaelbling, 1981; Debabov, 1982) and they include its ability to secrete proteins into the medium, its non-pathogenic nature to humans, its ability to grow at relatively high temperatures, its relatively heat-stable proteins, its ability to grow on either simple or complex media, and its historical use for the production of many industrial and food products normally synthesized by *B. subtilis*. With the advent of recombinant DNA technology it appears feasible to genetically engineer *B. subtilis* to be as versatile as the *Escherichia coli* and yeast systems for expressing various heterologous prokaryotic and eukaryotic genes.

The use of *B. subtilis* for the efficient production of biochemicals and proteins will depend on several major factors: (1) a better understanding of the genetic transcription and translation mechanisms including the regulatory elements which control these important functions; (2) the utilization of stable recombinant plasmids; (3) the development of *B. subtilis* strains which are reduced in their protease and nuclease contents, and (4) a systematic study of the protein-secretion process to determine the factors which would have to be manipulated to allow secretion of normally intracellular proteins. The analysis and improvement of these characteristics of *B. subtilis* are necessary for overcoming some of the inherent disadvantages and difficulties of the wild-type strains for their use as production organisms.

Once a better understanding of the features mentioned above has been obtained, the availability of current genetic engineering procedures should permit the creation of an 'ideal' cloning and gene-expression system in *B. subtilis*. The following scenario appears feasible (*Figure 1*). A gene, the

Abbreviations: bp, base pairs; CAT, chloramphenicol acetyltransferase; $CatO_2ase$, catechol:oxygen 2,3-oxidoreductase (catechol 2,3-dioxygenase); PSDR, portable Shine–Dalgarno regions; RBS, ribosome-binding site; SD, Shine–Dalgarno sequence; SES, stability-enhancing sequence.

Biotechnology and Genetic Engineering Reviews – Vol. 2, October 1984
0264-8725/84/02/121-35$10.00 + $0.00

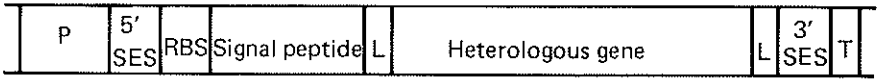

Figure 1. A schematic representation of a recombinant gene. P, promoter; 5′ SES, 5′ stability-enhancing sequence; RBS, ribosomal binding site; L, linker with restriction sites; 3′ SES, 3′ stability-enhancing sequence; T, transcription-termination site.

product of which is generally difficult to obtain, will be linked in phase to a fragment of DNA coding for a *B. subtilis* signal peptide. The signal sequence would be preceded by a very efficient translation initiation sequence or ribosome-binding site (RBS) which in turn would be preceded by mRNA stability-enhancing sequences (SES) at the 5′ end of the mRNA. The 'promoter' for this gene would actually consist of a cluster of efficient promoters which would be regulated either temporally and/or by conditions of the growth medium. Thus the gene would have the properties suitable not only for efficient transcription and translation, but would be under temporal control eliminating the need for any exogenous induction. The mRNA produced from this gene would also be stabilized by a 3′ SES, which would protect the mRNA from degradation by 3′ exonucleases. Thus the mRNA would have a longer half-life than normal *B. subtilis* mRNA and be translated in a polysomal mode. The protein product would have a typical *B. subtilis* signal peptide attached to its *N*-terminus and thus would be processed for secretion.

This engineered gene would then be inserted into a stable moderately high copy number *B. subtilis* plasmid. The plasmid would be transformed into a modified *B. subtilis* strain which does not cause gene rearrangements, contains fewer extracellular and intracellular proteases, has normal growth requirements, is stable and active during the stationary phase, and allows secretion of normally non-secreted proteins. Since this strain would produce reduced levels of extracellular and intracellular proteases, the extracellular protein product will not be degraded rapidly and possibly not at all if it is secreted efficiently. The heterologous gene would be expressed only after the culture had reached the stationary phase of growth or maximal cell density, since the promoter cluster would be designed to have a temporally controlled promoter in tandem with other temporal and stationary-phase-expressed promoters; this design would maximize the stability of the resident recombinant plasmid and allow long fermentation periods during the stationary phase with concomitant production of high levels of desired proteins.

What is the current understanding of gene expression and the genetic apparatus of *B. subtilis*? What is the relationship of our present knowledge to the scenario just presented for an 'ideal' gene? The purpose of this review is to provide a summary of our current knowledge of gene expression in *B. subtilis* and to point out areas which have to be investigated further before *B. subtilis* can be developed and utilized as an efficient system for heterologous gene expression. The recent rapid progress with *B. subtilis* indicates that the development of this system should occur within a few years.

Transcription

Transcription is one of the major sites for controlling gene expression in prokaryotes. Gene expression at this level can be regulated either temporally, nutritionally, or by the inherent characteristic of the promoter of the gene. Therefore a firm knowledge of the structure, function and regulation of the transcription machinery is necessary for construction of efficient genes. One must take into account the properties of the RNA polymerase holoenzyme (EC 2.7.7.6), their promoter specificities, regulatory sites near the promoter, and the various factors which regulate RNA polymerase activity.

TRANSCRIPTION APPARATUS OF *B. SUBTILIS*

Recent data have revealed the complex nature of the transcription apparatus present in *B. subtilis* in contrast to that found in *Escherichia coli*. A major difference between the two organisms is the presence of several different RNA polymerase holoenzymes in *B. subtilis* compared with the single RNA polymerase species which exists in *E. coli* (Losick and Pero, 1981; Doi, 1982a,b).

Four species of RNA polymerase holoenzyme have been observed in vegetative cells of *B. subtilis* (*Table 1*). They all contain the same RNA polymerase core enzyme consisting of $\alpha_2\beta\beta'$ subunits and only differ from each other by the sigma factor (σ) associated with the core enzyme. The sigma factors determine the promoter recognition specificity of the enzyme. The sigma factors associated with the vegetative cell core enzyme have molecular masses of 55 000 daltons (σ^{55}) (Shorenstein and Losick, 1973), 37 000 daltons (σ^{37}) (Haldenwang and Losick, 1980), 32 000 daltons (σ^{32}) (Johnson, Moran and Losick, 1983), and 28 000 daltons (σ^{28}) (Wiggs, Gilman and Chamberlin, 1981). One sigma factor, σ^{29}, has been found only in stationary phase cells (Haldenwang, Lang and Losick, 1981).

The σ^{55} enzyme comprises about 90–95% of the total RNA polymerase of vegetative cells and therefore the other forms are considered as minor forms of the enzyme (Doi, 1982a,b). The σ^{55} enzyme carries out most of the transcription during the growth phase: although the minor forms are present during the growth phase, albeit in small concentrations, their primary activities

Table 1. Multiple RNA polymerase holoenzymes of *B. subtilis*

Growth stage	RNA polymerase forms present*	References†
Vegetative	$E\sigma^{55}$, $E\sigma^{37}$, $E\sigma^{32}$, $E\sigma^{28}$	1,2,3,4
Sporulation		
Stage 0	$E\sigma^{55}$, $E\sigma^{37}$, $E\sigma^{32}$, $E\sigma^{28}$	1,2,3,4
Stage II	$E\sigma^{55}$, $E\sigma^{37}$, $E\sigma^{32}$, $E\sigma^{29}$	2,3,5,6,7
Stage III	$E\sigma^{55}$, $E\sigma^{29}$	5,6,7

* E stands for the core enzyme with the subunit composition $\alpha_2\beta\beta'$. The σ subunit superscript indicates the MW of the $\sigma \times 10^{-3}$.

† The data are from (1) Shorenstein and Losick, 1973; (2) Haldenwang and Losick, 1980; (3) Johnson, Moran and Losick, 1983; (4) Wiggs, Gilman and Chamberlin, 1981; (5) Fukuda and Doi, 1977; (6) Haldenwang, Lang and Losick, 1981; (7) Nakayama *et al.*, 1978.

may be associated with events which occur near the end of the log phase of growth and during early stationary phase when the cells are making a transition from rapid growth to a maintenance state and/or to sporulation. The activity of the σ^{55} enzyme is also modulated after the end of the growth phase, since there is reduced activity of σ^{55} enzyme after growth has ceased (Losick and Sonenshein, 1969; Gilman and Chamberlin, 1983). The possibility that σ^{55} enzyme activity is regulated by an anti-σ^{55} factor (Segall *et al.*, 1974) has been postulated, but not proved biochemically as yet (Tjian, Stinchcomb and Losick, 1974).

The σ^{37} enzyme is present in both log phase cells and early sporulating cells, but cannot be detected after the second hour of sporulation (Losick, 1982). Therefore the σ^{37} gene itself may be regulated during this stage. The σ^{28} enzyme is also present during the log phase of growth as determined by S1 nuclease mapping with RNA obtained from vegetative cells (Gilman and Chamberlin, 1983) and during the first two hours of sporulation, after which its transcripts decrease precipitously. Thus the σ^{28} enzyme-controlled genes appear to be repressed at this time.

The level of the σ^{37} enzyme is regulated by nutritional conditions, since cells grown on glucose medium contain only small amounts of this enzyme, while cells grown on richer, complex sporulation medium have increased amounts of the σ^{37} enzyme (Losick, 1982). An analysis of σ^{28} transcripts revealed that transcription from σ^{28} promoter-controlled genes is five times more active when cells are growing on glucose than on amino acids as the carbon source; furthermore, on nutrient sporulation medium there was a further twofold increase of transcription over the level with glucose (Gilman and Chamberlin, 1983). However, in very rich medium containing high concentrations of glucose — conditions which repress sporulation — transcription from σ^{28} promoters declined twofold from the level in sporulation medium. Moreover, a rapid loss in transcription from σ^{28} promoters occurred during early stationary phase, and this decline even preceded the decline in transcription from σ^{55} promoters. Therefore transcription from σ^{28} promoters was affected significantly by conditions of the medium.

The σ^{29} enzyme occurs only during the sporulation phase (Fukuda and Doi, 1977; Nakayama *et al.*, 1978; Haldenwang, Lang and Losick, 1981). Two other proteins with molecular weights of 34 000 and 23 000 and in association with core enzyme from sporulating cells have been reported, but their functions are still uncertain (Doi *et al.*, 1980; Doi, 1982a,b).

DEVELOPMENTAL CONTROL ON PROMOTER UTILIZATION

In addition to nutritional effects on σ^{37} and σ^{28} enzyme activity, there is a developmental control on expression of certain genes transcribed by these two forms of RNA polymerase. Several *spoO* genes, which are expressed at the end of the exponential phase of growth and are involved in the initiation of sporulation, appear to control transcription of genes by the minor forms of RNA polymerase. Cells bearing mutations on *spoOA*, *spoOB*, *spoOC*, *spoOE*, *spoOF*, *spoOH* and *spoOK* cannot transcribe the *spoVG* gene

(Ollington *et al.*, 1981) which is transcribed by σ^{37} enzyme (Moran *et al.*, 1981) and σ^{32} enzyme (Johnson, Moran and Losick, 1983). The expression of σ^{28}-controlled genes is also under control of *spoO* genes, since σ^{28} transcripts were substantially reduced in strains bearing mutations in *spoOA*, *spoOB*, *spoOE*, and *spoOF* (Gilman and Chamberlin, 1983); however, in *spoOH*, *spoOJ* and *spoOK* strains, σ^{28} transcripts were present at wild-type levels. Thus the ability to transcribe genes controlled by σ^{37} and σ^{28} promoters appears to be under the control of at least some of the *spoO* functions. These *spoO* functions which are also present during the growth phase have a major role in determining whether the sporulation cycle is to be initiated in *B. subtilis* (*see* Losick, 1982, for review). As *spoO* mutations tend to be pleiotropic and prevent development beyond the *spoO* stage, they could have significant deleterious consequences on gene expression during the stationary phase. Thus for practical purposes it would be best to avoid the use of *spoO* mutants for stationary phase expression of genes.

TRANSCRIPTION SPECIFICITY AND PROMOTERS OF *B. SUBTILIS*

The sigma factors play a critical part in promoter selection by the different RNA polymerase holoenzymes. This has been demonstrated by dissociation–reconstitution studies in which the sigma factor is dissociated from the core enzyme, which by itself shows no specificity for promoters; when the sigma factor is reassociated with the core, it restores promoter recognition specificity to the enzyme (Haldenwang and Losick, 1980; Haldenwang, Lang and Losick, 1981; Johnson, Moran and Losick, 1983). By these experiments, it was shown that each holoenzyme form recognized its own set of promoter sequences. Significant differences between these sets of promoters were noted particularly in the conserved regions initially observed in *E. coli* promoters (*see* Rosenberg and Court, 1979, for review).

Examples of minor promoters recognized by the various minor *B. subtilis* RNA polymerases are illustrated in *Figures 2–4*. Several features of the promoters should be noted, since they would have a role in the construction of suitable promoters for efficient expression of engineered genes. The σ^{55} promoters of *B. subtilis* (Moran *et al.*, 1982) resemble the *E. coli* promoters with a consensus '−35' sequence of TTGACA and a '−10' sequence of TATAAT (Rosenberg and Court, 1979). Both conserved '−35' and '−10' regions of the minor promoters are quite distinct from those present in the σ^{55} promoters. Although it is premature to state that the sequence of the '−35' and '−10' regions of the minor sigma promoters are consensus sequences, they can be discerned quite readily from the σ^{55} promoters in both the '−10' and '−35' regions. Since both regions differ for the various promoters, Losick and Pero (1981) have proposed that the holoenzymes recognize both portions of the promoter, perhaps sequentially when the enzymes contact their cognate promoters.

Another recognizable region in the promoter is the interval between the '−35' and '−10' regions. For the σ^{55} promoters this interval is usually 17 base pairs (bp) long with some having 18 bp (Moran *et al.*, 1982). In four σ^{37}

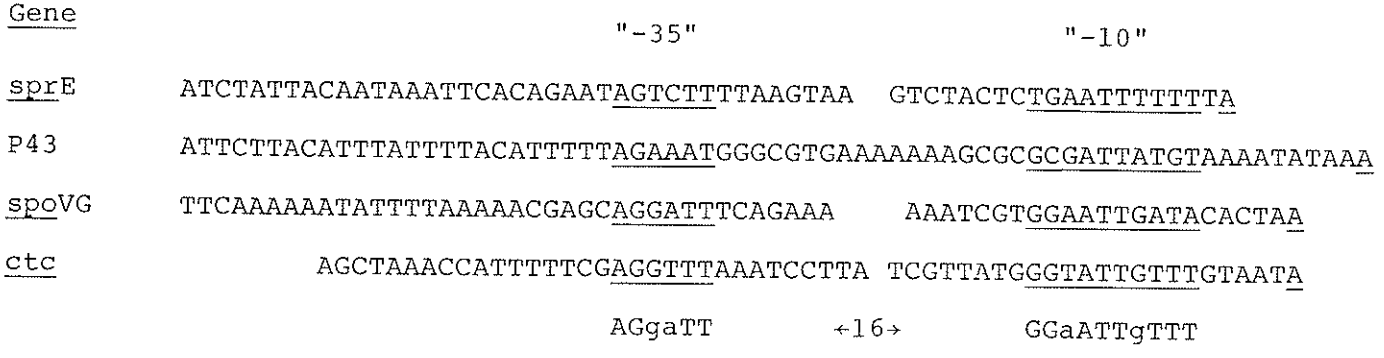

Figure 2. Sequences of *B. subtilis* sigma-37 promoters. The conserved '−10' and '−35' sequences are underlined. The bottom line indicates the consensus sequence with the large letters indicating either 4/4 or 3/4 agreement and the small letters indicating 2/4 agreement. The ←16→ indicates the average number of base pairs between the two conserved regions. The underlined letter to the far right is the site of transcription initiation. The references for the genes are as follows: *sprE* (Wong *et al.*, 1984); P43 (Wang and Doi, 1984); *spoVG* (Johnson, Moran and Losick, 1983); *ctc*, (Moran *et al.*, 1981).

```
Gene                "-35"                    "-10"
pMG102    TGATCAGAAGCTAAATGATTCTGTTTTTATGCCGATATAATCACTA
pMG201    CACAATGTCCCTAAAGTTCCGGGCACCAAAACCGATATTAACCATA

                    CTAAA        ←16→        CCGATAT
```

Figure 3. Sequences of *B. subtilis* sigma-28 promoters. The conserved '−10' and '−35' sequences are underlined. The bottom line shows the identity of the two promoters which have been sequenced to date. The ←16→ indicates the number of base pairs between the conserved regions. The underlined letter to the far right is the site of transcription initiation. The data are from Gilman, Wiggs and Chamberlin (1981).

```
                                                          ▽
sprE      TTATTCCATCTATTACAATAAATTCACAGAATAGTCTTTTAAGTAAGTCTA
                          "-35"                  "-10"

                                                          ▽
spoVG     CGAGCAGGATTTCAGAAAAAATCGTGGAATTGATACACTAATGCTTTTATA
                          "-35"                  "-10"

                          AAAT-        ←15→      T-A-G----T
```

Figure 4. Sequences of *B. subtilis* sigma-32 promoters. The conserved '−10' and '−35' sequences are underlined. The arrowhead points to the transcription-initiation point. In the bottom line a large letter indicates agreement between the two promoters. The ←15→ indicates the average number of base pairs between the two conserved regions. The data are from the following papers: *sprE* (Wong *et al.*, 1984); *spoVG* (Johnson, Moran and Losick, 1983).

promoters a 13–18 bp interval is present (*Figure 2*); for two σ^{28} promoters (*Figure 3*) the interval is 16 bp (Gilman and Chamberlin, 1983); for two σ^{32} promoters (*Figure 4*) the intervals are 14 and 15 bp (Johnson, Moran and Losick, 1983; Wong *et al.*, 1984). Although the data are sparse, it appears that the interval for the minor RNA polymerases is smaller (15–16 bp) than for the major σ^{55} enzyme (17–18 bp).

Many promoters are rich in AT base pairs upstream from the '−35' region (*Figure 5* and Moran *et al.*, 1982). Moran, Lang and Losick (1981) and Banner, Moran and Losick (1983) have shown by deletion analysis that this AT-rich region is essential for transcription from a σ^{37} promoter. Furthermore, a very strong σ^{55} promoter for the *veg* gene is preceded by a very AT-rich region (Moran *et al.*, 1982). They suggested that this region was important for efficient utilization of *B. subtilis* promoters and can be considered a transcription-enhancing region. The tandem promoters of *rrn* genes are also preceded by AT-rich regions. The P1 and P2 (first and second promoters) of the *rrnB* genes are preceded by 90% AT and 73% AT regions, respectively (Stewart and Bott,

1983). Two sets of *rrn* genes, *rrnA* and *rrnO*, are located near the replication origin of *B. subtilis* and they also have tandem promoters (Ogasawara, Moriya and Yoshikawa, 1983). The P1 and P2 of *rrnA* are preceded by AT-rich regions of 67% and 88%, respectively, and the P1 and P2 of *rrnO* are preceded by AT-rich regions of 67% and 88%, respectively. It is interesting to note that with the tandem *rrn* promoters the promoter with the richer upstream AT-rich region is preferred *in vivo*. These *rrn* promoters are very strong and can be subcloned in *E. coli* only with suitable termination sites distal to the promoters (Stewart and Bott, 1983). Thus these promoters are too strong for gene cloning unless suitably tailed by a strong transcription-termination signal (Gentz *et al.*, 1981).

A final segment of the promoter is the region which lies between the '−10' region and the +1 initiation site for transcription. Since many promoters have not been analysed by S1 nuclease mapping studies (Berk and Sharp, 1977), it is difficult to know the exact '−10' to +1 intervals for these promoters. However, based on initiation with a purine, most promoters have 4–7 base pairs between the '−10' and +1 site and this region is usually enriched in AT base pairs. The presence of AT base pairs in this interval may facilitate the localized melting of DNA which occurs during transcription initiation (Siebenlist, 1979).

From the available data, it is difficult to state what features of the *B. subtilis* promoters actually make them efficient or strong promoters. In the case of *E. coli*, there is evidence that promoters which have conserved −35 and −10 regions identical with the consensus sequence are stronger (de Boer, Comstock and Vasser, 1983). Only one *B. subtilis* σ^{55} promoter with both consensus −35 and −10 regions has been reported. This is the ϕ29 A1 or P_E1 promoter (Yoshikawa and Ito, 1982) but its exact relative strength is unknown. Generally, however, early phage promoters are considered to be strong promoters. Most of the σ^{55} promoters in fact vary considerably from the consensus sequence in either the '−10' or '−35' regions (*Figure 5*). Further data are required to determine the exact properties of highly efficient promoters. However by use of promoter-probe plasmids (*see* pages 132–134), one can obtain an idea of the relative strength of promoters *in vivo*.

PROMOTER ORGANIZATION IN *B. SUBTILIS*

Two types of complex promoter organization have been observed in *B. subtilis* which may influence the ultimate promoter designs for engineered genes. One basic design would have a promoter which is very efficient for gene expression during rapid growth of the cell. Another type of promoter would allow a controlled expression of genes for extended periods during the stationary phase of growth.

One kind of promoter organization is found in the *rrn* operons which code for the ribosomal RNAs of *B. subtilis*. The presence of *tandem promoters* in *rrn* operons has been reported by Stewart and Bott (1983) and Ogasawara, Moriya and Yoshikawa (1983). For *rrnB and rrnO*, 92 bp separate the '−10' regions of the two tandem promoters, whereas for *rrnA* a smaller spacing of 75 bp occurs. In *rrnB* the transcript initiated from P1 is 89 bp longer than the transcript from

Figure 5. AT-rich regions of *B. subtilis* sigma-55 promoters. The underlined letter at the far right is the site of transcription initiation. The '−10' and '−35' regions are underlined. The two arrowheads indicate the AT-rich region between nucleotides −40 and −60 of the promoters. See other sigma-55 promoter sequences in Murray and Rabinowitz (1982), Yoshikawa and Ito (1982), and Moran *et al.* (1982). The references for the promoters are as follows: P43, Wang and Doi, 1984; *rrnO* P2 and *rrnA* P1, Ogasawara, Moriya and Yoshikawa 1983; *rrnB* P1, Stewart and Bott, 1983; *rrnT* 2, Wawrousek and Hansen, 1983; SPO-26; Lee, Talkington and Pero, 1980; *veg*, Moran *et al.*, 1982; ϕ29; Yoshikawa and Ito, 1982.

P2. The P1 is used more extensively than P2 *in vivo* for both *rrnA* and *rrnB*, while P2 is used preferentially *in vivo* for *rrnO*. The sequence analysis of these tandem promoters indicated that both P1 and P2 were σ^{55} promoters.

The *rrn* promoters are extremely efficient *in vivo* and cause replication problems if they are inserted into *E. coli* plasmids (Stewart and Bott, 1983), unless the promoters are followed by an efficient *rho*-dependent transcription terminator such as the λtR1 (McKenney *et al.*, 1981). Since the *rrn* promoters are efficient σ^{55} promoters, the insertion of a properly engineered gene (e.g. with an efficient transcription terminator) distal to these promoters should allow maximum expression of the gene during the growth cycle. Since maximum expression occurs with only one of the two tandem promoters, the tandemness and double promoter *per se* may not be necessary. However, as we will discuss later, the idea of tandemness may be utilized for special cases.

The second type of promoter organization which has been observed recently is the presence of *overlapping promoters* in *B. subtilis* which control expression of *spoVG* (Moran *et al.*, 1981), *ctc* (Johnson, Moran and Losick, 1983), *sprE* (Wong *et al.*, 1984) and a 'cryptic gene' (Wang and Doi, 1984). The overlapping nature of these promoters is illustrated in *Figure 6*. Several interesting features of these promoters include (1) the overlapping of two different types of promoters, e.g. σ^{37} and σ^{32} promoters, σ^{55} and σ^{37} promoters; (2) staggered transcription initiation points or identical initiation points; (3) the possibility of temporally simultaneous or sequential use of the promoters (i.e. the use of both promoters during a particular growth phase or the use of one promoter during one growth phase and the other during another growth phase).

In the studies on the mapped genes *spoVG*, *ctc* and *sprE*, both promoters, which had staggered start points, were shown to be used during the stationary phase. Analyses were not done systematically at different growth phases in those studies.

The overlapping promoters of a 'cryptic gene' (Wang and Doi, 1984) were utilized during both growth and the stationary phase. Since these overlapping genes have an identical start point, it was not possible to discern whether one or both promoters were being used at the two growth phases (*Figure 6*). Quantitative S1 nuclease mapping (Gilman and Chamberlin, 1983) data indicated that the overlapping promoters were utilized much more highly during the early stationary phase than during the log phase of growth. The latter observation raises the intriguing possibility that the σ^{55} promoter is used only during growth and the σ^{37} promoter only during early stationary phase. This idea can be tested by obtaining temperature-sensitive (*ts*) mutants of the σ^{55} and σ^{37} genes. As the σ^{55} gene *(rpoD)* has been cloned (Price, Gitt and Doi, 1983), the isolation of a *ts* σ^{55} mutant should be feasible. With the proper *ts* mutant, a shift-up experiment at the early stationary phase would inactivate the σ^{55} enzyme and allow only the σ^{37} enzyme to transcribe from the overlapping cryptic gene promoter. Thus the contribution of the σ^{37} enzyme to the total 'cryptic gene' transcription can be determined.

The presence of *tandem* and *overlapping* promoters suggests the following design for clustered promoters (*see Figure 7*):

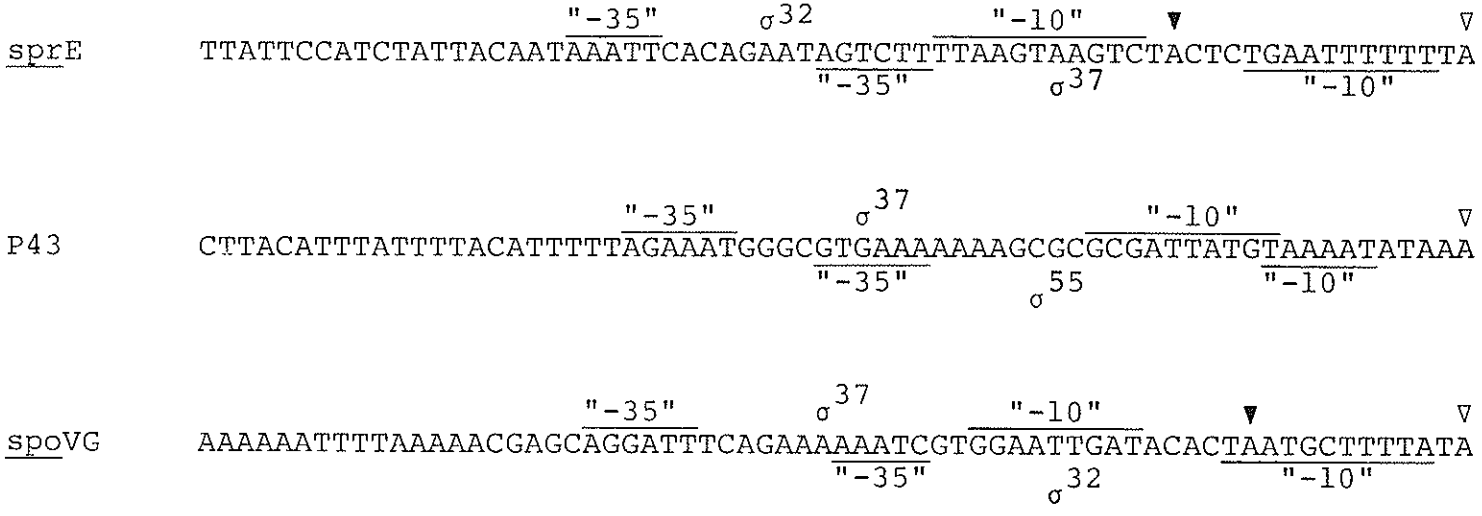

Figure 6. Overlapping promoters in *B. subtilis*. The solid and open arrowheads indicate the transcription initiation points for P1 and P2, respectively. For P43, the same transcription initiation point is used by both P1 and P2. The '−10' and '−35' regions for the P1 promoters are overlined and for the P2 promoters underlined. Each type of promoter is indicated by a sigma with a superscript. The data are from the following references: *sprE*, Wong *et al.*, 1984; P43, Wang and Doi, 1984; *spoVG*, Johnson, Moran and Losick, 1983.

1. Tandem promoters containing a strong σ^{55} promoter followed by a σ^{37} promoter. The purpose of this arrangement is to have a gene expressed efficiently during both growth and the stationary phases.
2. Tandem promoters containing two overlapping promoters (σ^{37} and σ^{32} promoters) followed by a σ^{29} promoter. In this arrangement a gene would be turned on after growth had ceased (most σ^{37}-controlled genes appear to be expressed strongly during the early stationary phase and are controlled by *spoO* genes (Losick, 1982)) and the expression initiated at the early stationary phase will be continued by σ^{29} enzyme which is produced about 2–3 hours after initiation of the stationary phase (Haldenwang, Lang and Losick, 1981). This should favour a long period of gene expression during the stationary phase. A relatively high expression of the chloramphenicol acetyltransferase gene was observed during the stationary phase when it was fused to the *sprE* σ^{37}–σ^{32} promoter (Goldfarb *et al.*, 1983; Wong *et al.*, 1984).

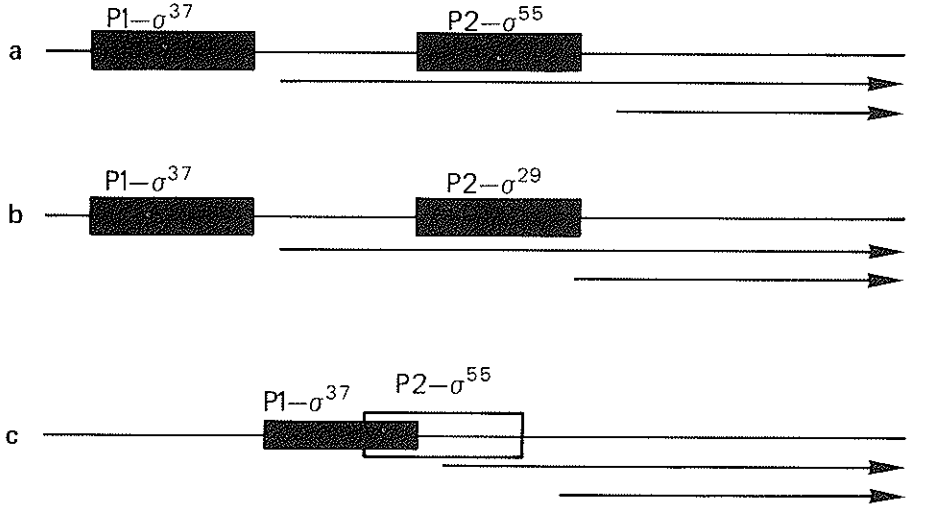

Figure 7. Schematic models for constructing promoters. (a) Two tandem promoters with different sigma specificities are linked to a gene to be expressed during growth (sigma-55) and during stationary phase (sigma-37). (b) Two tandem promoters are used to obtain expression of a gene during early stationary phase (sigma-37) and late stationary phase (sigma-29). (c) Two overlapping promoters (sigma-37 and sigma-55) are used to obtain expression during both growth and sporulation. The arrows indicate the site and direction of transcription.

Selection of promoters useful for genetic engineering

EXPRESSION-PROBE PLASMIDS

Several types of promoter-probe plasmids are currently available for the isolation of *B. subtilis* promoters (*Table 2*). The first type facilitates the selection of *B. subtilis* promoters contained on *Hin*d III DNA fragments which

Table 2. Promoter-cloning plasmids for *B. subtilis*

Plasmid	Cloning sites	Selection marker	Hosts	Ref†
pGR71	*Hind* III	CAT*	*B. subtilis* and *E. coli*	1
pPL703	*Eco*RI, *Pst* I, *Sal* I, *Bam* HI	CAT	*B. subtilis*	2
pCPP–3	*Eco* RI	CAT	*B. subtilis* and *E. coli*	3
pCPP–4	*Bam* HI	CAT	*B. subtilis* and *E. coli*	3
pTG402	*Bam* HI, *Hpa* I, *Kpn* I	$CatO_2$ase	*B. subtilis* and *E. coli*	4
pCED6	*Hind* III	β-galactosidase	*B. subtilis* and *E. coli*	5

* CAT, chloramphenicol acetyltransferase; $CatO_2$ase, catechol 2,3-dioxygenase.
† (1) Goldfarb, Doi and Rodriguez, 1981; (2) Mongkolsuk *et al.*, 1983; (3) Band, Yansura and Henner, 1983; (4) Zukowski *et al.*, 1983; (5) Donnelly and Sonenshein, 1984.

cause expression of a chloramphenicol acetyltransferase (CAT) gene (*cat*) (Goldfarb, Doi and Rodriguez, 1981). This plasmid, pGR71, was constructed from an *E. coli* plasmid pBR350 and a *Staphylococcus aureus* plasmid pUB110 and is able to replicate in both *E. coli* and *B. subtilis* as a shuttle vector. Inserted into this cointegrate plasmid is a promoterless *cat* gene from *E. coli* transposon Tn9; directly in front of this *cat* gene is located a unique *Hind* III site into which promoter-containing *Hind* III fragments from *B. subtilis* can be inserted. In the absence of a promoter-containing DNA fragment the host cell is chloramphenicol sensitive; if the proper promoter-containing fragment is inserted into the *Hind* III site, the host cell becomes chloramphenicol resistant. One of the limiting features of this plasmid is that transcription and translation in *B. subtilis* initiate from the promoter and the ribosome-binding site (RBS) furnished by the inserted *Hind* III fragment. Thus the insert must contain a translation reading frame in phase with the open reading frame of the leader sequence of the *cat* gene. Translation from the RBS on the insert results in the formation of an enzymatically active fusion product which has an *N*-terminal *B. subtilis* polypeptide fused to the CAT (Goldfarb, Rodriguez and Doi, 1982). Translation cannot be initiated from the *E. coli* RBS present on the Tn9 *cat* gene, because the *B. subtilis* translation mechanism uses *E. coli* RBS very poorly if at all (Goldfarb, Rodriguez and Doi, 1982).

Theoretically, it is possible to isolate 1/3 of the *B. subtilis* promoters with this plasmid, assuming that 1/3 of the *Hind* III fragments will be inserted in the correct reading frame with the *cat* gene. Another disadvantage of this plasmid is that promoters which are expressed after the growth phase cannot be selected readily, since cells containing such recombinant plasmids will be chloramphenicol sensitive during growth. This difficulty has been partially overcome by adding anti-CAT antibody to the growth plate instead of chloramphenicol. A precipitin line forms around colonies which produce CAT during the stationary phase, because cells lyse during sporulation allowing the released CAT to interact with the anti-CAT in the agar (Goldfarb *et al.*, 1983). Despite some of these difficulties this plasmid has been used to isolate promoters which are expressed during growth (Goldfarb, Doi and Rodriguez, 1981) and the stationary phase (Goldfarb *et al.*, 1983; Wong *et al.*, 1984).

Another useful plasmid for selecting *B. subtilis* promoters is designated pPL703 (Mongkolsuk *et al.*, 1983). It was derived from pPL603 (Williams,

Duvall and Lovett, 1981) and contains a promoter-less CAT gene (*cat*) originating from *B. pumilus* NCIB 8600. In front of this gene are four unique restriction sites *Eco* RI, *Pst* I, *Sal* I and *Bam* HI. The insertion of promoter-containing fragments into these sites activates the expression of the *cat* gene. The advantage of pPL703 over pGR71 mentioned previously is that the *cat* gene in pPL703 contains a *B. subtilis*-like RBS (Duvall *et al.*, 1983; Harwood, Williams and Lovett, 1983) and results in the production of a native CAT product in contrast to the usual fusion product made from pGR71. This structure eliminates the requirement that the inserted DNA fragment be linked to the *cat* gene in an open reading frame from the translation start point in the insert. Thus in theory any promoter-containing DNA fragment should activate the *cat* gene. Although this plasmid has an origin of replication for only *B. subtilis* and therefore is not a shuttle vector, another advantage is its extreme stability in *B. subtilis* and the relative absence of gene rearrangements during growth.

A type of plasmid which features a combination of pGR71 and pPL703 is represented by plasmids pCPP-3 and pCPP-4 which contain the origins of replication from pBR322 and pBU110 and a promoter-less *Cm*^{r} gene from pC194 (Band, Yansura and Henner, 1983). The RBS is still intact for the *Cm*^{r} gene and contains either an *Eco* RI site (pCPP-3) or a *Bam* HI site (pCPP-4) in front of the gene for insertion of promoter-containing DNA fragments. In addition a terminator sequence λt_0 was inserted after the *Cm*^{r} gene in order to enhance the cloning of very strong promoters. The advantages of these plasmids are: (1) they are shuttle vectors which can replicate in *B. subtilis* and *E. coli*; (2) they contain an RBS which can be utilized by the *B. subtilis* translation system; (3) the *Bam* HI and *Eco* RI sites are particularly useful for cloning; (4) the selection process is simple and sensitive; (5) the sequence of the *Cm*^{r} gene is known. These plasmids have been utilized to select *B. subtilis* promoters which confer resistance to chloramphenicol from 5 μg to 250 μg/ml.

A chromogenic promoter-probe plasmid has been developed by inserting the *Pseudomonas putida* TOL plasmid gene *xylE*, which codes for catechol 2,3-dioxygenase ($CatO_2$ase; catechol:oxygen 2,3-oxidoreductase (decyclizing), (EC1.13.11.2)), into pHV33, a plasmid vector that functions in *E. coli* and *B. subtilis* (Primrose and Ehrlich, 1981) to yield plasmid pTG402 (Zukowski *et al.*, 1983). Since the *xylE* gene is the second gene of a regulated operon on the *P. putida* TOL plasmid (Franklin *et al.*, 1981; Inouye, Nakazawa and Nakazawa, 1981) it does not have its own promoter and thus insertion of promoter-containing DNA fragments into unique *Bam* HI, *Hpa* I and *Kpn* I sites on the cloning vector activates expression of the $CatO_2$ase gene. The special feature of this system is the method for detecting the activated gene. The colonies which express *xylE* become yellow within seconds after selection plates are sprayed with catechol, a colourless compound that is converted by the enzyme to the yellow product, 2-hydroxymuconic semialdehyde. Although this is a *Pseudomonas* gene, the *B. subtilis* translation system is able to initiate translation from the *xylE* ribosome-binding site and synthesize a native-sized $CatO_2$ase. Thus this is also a pure promoter-probe plasmid and the specific activity of the $CatO_2$ase in different clones probably reflects the promoter

efficiency, although the published data (Zukowski *et al.*, 1983) were from stationary phase cells and therefore subject to some interpretation; e.g. enzyme produced by promoters which turned on the gene prior to sporulation may be subjected to more turnover during sporulation than enzyme produced from promoters turned on in late sporulation. Nevertheless, this system has the advantage that special indicator plates are unnecessary, the substrate catechol is very inexpensive, and the detection method is simple and quite sensitive.

Another promoter-selection method based on chromogenic detection has been developed by Donnelly and Sonenshein (1984). A promoter-less *trpB'A'lacZ* fusion derived from *E. coli* has been inserted into a shuttle vector derived from pBR322 and pUB110 to form pCED6. The plasmid has a unique *Hin*dIII site in front of the *trpB'A'lacZ* fusion into which promoter-containing *Hin*dIII fragments can be inserted. A mRNA is made containing the sequence of the *N*-terminus of the inserted gene fused to the *trpB'A'lacZ* genes. Translation of the mRNA occurs from an *E. coli* ribosome-binding site present on *trpA'* and results in a *trpA'–lacZ* fusion product which has β-galactosidase activity. When cells containing pCED6 with DNA inserts are grown on X-gal (5-bromo-4-chloro-3-indoyl-β-D-galactoside) plates, the promoter-containing $LacZ^+$ colonies are blue and the $LacZ^-$ colonies are white. Thus the detection of activation of the *lacZ* gene by inserted promoters is simple and this method facilitates the screening of large numbers of clones. As the amount of β-galactosidase produced is a reflection of the amount of mRNA synthesized, pCED6 is also a pure promoter-probe plasmid.

The advantages of this system are the ease of detection of promoter-containing clones and the ability to detect all types of promoters by use of the special selection plate. Furthermore, the *in vitro* enzyme assays are relatively easy to conduct.

The practical use for these promoter-probe plasmids is that they can be used to select strong or efficient promoters, and nutritionally (Goldfarb *et al.*, 1983), developmentally and temporally (Goldfarb *et al.*, 1983; Mongkolsuk *et al.*, 1983; Wong *et al.*, 1984) regulated promoters. They can also be used to test promoters under heterologous host conditions, e.g. *B. subtilis* promoters in *E. coli* and vice versa.

TEMPORALLY EXPRESSED PROMOTERS

Several temporally regulated promoters which are expressed during early stationary phase have been characterized (Haldenwang, Lang and Losick, 1981; Moran *et al.*, 1981: Goldfarb *et al.*, 1983; Zuber and Losick, 1983; Wong *et al.*, 1984). One controls expression of the extracellular serine protease subtilisin E gene *sprE* (Wong *et al.*, 1984) and actually consists of two overlapping promoters recognized by σ^{37} and σ^{32}-like enzyme. The *sprE* gene expression is reduced strongly by *spoOA*, to a lesser extent by *spoOB*, *spoOE* and *spoOF*, and hardly at all by *spoOH* (Brehm, Staal and Hoch, 1983). The *spoVG* and *ctc* genes are also temporally regulated and transcribed from two overlapping promoters, a σ^{37} and a σ^{32} promoter (Johnson, Moran and Losick,

1983). The expression from the *spoVG* gene is regulated by at least seven *spoO* genes including *spoOA*, *spoOB*, *spoC*, *spoOE*, *spoOF*, *spoOH* and *spoOK* (Zuber and Losick, 1983).

Two temporally regulated promoters have also been reported which are expressed in *spoOA* cells. Their expression was reduced when grown in the presence of high concentrations of glucose suggesting that some type of glucose effect was being exerted on them (Mongkolsuk *et al.*, 1983).

The use of these promoters for gene expression in $spoO^+$ cells may be particularly useful, since external inducers or shifts in temperature may not be necessary and they may allow extended expression of genes during the stationary phase.

EXPRESSION OF HETEROLOGOUS PROMOTERS IN *B. SUBTILIS*

Very few data are available to show that promoters from foreign organisms are used in *B. subtilis*. Generally there appears to be very poor usage of *E. coli* promoters by *B. subtilis in vivo* since *E. coli* DNA fragments shotgunned into the promoter-probe plasmid pTG402 yielded very few clones which turned on the probe enzyme (Zukowski *et al.*, 1983). In related studies DNA fragments from *B. licheniformis* and *B. pumilus* were able to express the probe enzyme at about the same frequency as *B. subtilis* DNA. Thus the transcriptional machinery apparently recognized *Bacillus* sp. promoters reasonably well, but *E. coli* promoters only poorly at best. The most precise study concerning the expression of *E. coli* promoters in *B. subtilis* was reported by Kreft *et al.* (1983) who showed by S1 nuclease mapping (Berk and Sharp, 1977) and promoter-deletion studies that the tetracycline resistance gene (*tet*) from pBR322 could be expressed in *B. subtilis* from the same promoter as that used in *E. coli* and that the *tet* gene was not being expressed from some *S. aureus* promoter on the plasmid. These investigators also demonstrated that the β-lactamase (*bla*) and chloramphenicol (*cat*) resistance genes from *E. coli* plasmids pBR322 and pACYC184, respectively, could not be transcribed in *B. subtilis*.

Promoters from other organisms which apparently are used in *B. subtilis* include the promoters from the *B. amyloliquefaciens* subtilisin (Wells *et al.*, 1983) and α-amylase (Palva, 1982; Palva *et al.*, 1982) genes, the promoters for *B. licheniformis* penicillinase (Neugebauer, Sprengel and Schaller, 1981; Gray and Chang, 1981) and α-amylase (Ortlepp, Ollington and McConnell, 1983) genes, the promoter for *B. cereus* cereolysin gene (Kreft *et al.*, 1983) and the promoter for the *B. thuringiensis* strain *berliner* crystal-protein gene (Klier *et al.*, 1982; *see also* Chapter 12 of this volume). However, until S1 nuclease mapping or promoter-deletion studies have been carried out for these genes, it is not possible to state unequivocally that the native promoters for these heterologous genes were transcribed in *B. subtilis*.

For the purpose of expressing heterologous genes in *B. subtilis*, it would be most advantageous to use a well-characterized *B. subtilis* promoter, the sequence, efficiency and regulation of which were understood.

THE IDEAL PROMOTER FOR GENE EXPRESSION IN *B. SUBTILIS*

In conclusion, the tandem or overlapping promoters for an engineered gene should have the following characteristics:

1. A favourable transcription-enhancing sequence (TES) in the −40 to −80 region containing A and T stretches;
2. An ideal consensus sequence in the '−35' and '−10' regions with a consensus spacing between these regions;
3. A favourable spacing and AT-rich composition between the '−10' and +1 regions;
4. Tandem or overlapping promoters selected to permit regulated transcription under desired temporal, nutritional or developmental conditions. This promoter or promoter cluster would be efficiently utilized under the desired conditions and would allow maximum transcription by the RNA polymerase holoenzymes which are present in the cell at that time.

Translation

Translation efficiency of a gene can be influenced by several factors, including the rate of translation initiation as affected by the ribosomal binding site (RBS), the stability of the mRNA, the codon usage by the translation machinery, availability of substrates, presence of ribonucleases, energy supply and toxicity of product. The parameters which could be controlled by genetic engineering techniques include the use of favourable RBS on the mRNA, the presence of stabilization-enhancing sequences (SES) on the 5′ and 3′ ends of the mRNA, and the presence of tRNAs which can recognize codons which are used infrequently in *B. subtilis*.

TRANSLATION BARRIER IN *B. SUBTILIS*

A number of studies have shown that the translational machinery from Gram-positive cells is not capable of efficiently translating mRNAs from Gram-negative organisms (Stallcup and Rabinowitz, 1973; Legault-Demare and Chambliss, 1975; McLaughlin, Murray and Rabinowitz, 1981). The site of the translation barrier appears to be at the level of 30S ribosome–mRNA interaction (Legault-Demare and Chambliss, 1975) and not at the level of initiation factors. When the *B. subtilis* translation system was offered a mRNA containing both a *B. subtilis* and an *E. coli* RBS, translation of the mRNA occurred *in vivo* only from the *B. subtilis* RBS (Goldfarb, Rodriguez and Doi, 1982), again indicating the preference of the *B. subtilis* translation system for its own RBS. Thus it is important when expressing a heterologous gene in *B. subtilis* that an efficient *B. subtilis*-like RBS is provided for the gene.

RIBOSOMAL BINDING SITES (RBS) ON *B. SUBTILIS* mRNAS

The RBS of different genes show a wide variation in rate of use and could vary by nearly a thousand-fold in relative efficiency (*see* Gold *et al.*, 1981, for

review). The examination of a number of RBS has revealed three common features: (1) a Shine–Dalgarno (SD) sequence (Shine and Dalgarno, 1974); (2) a spacer region between the SD sequence and the initiation codon; and (3) the initiation codon, AUG or GUG. Besides these features, Gold *et al.* (1981) proposed that the strength of the initiation signal also depended on the *spatial organization* of the initiation domain of the mRNA. The optimization of these features by genetic engineering should result in the formation of a strong initiation signal on the mRNA.

The analysis of RBS of *B. subtilis* and other Gram-positive organisms indicated that a high complementarity was required between the SD sequence at the 3′ end of *B. subtilis* 16S rRNA and the 5′ end of the mRNA for efficient translation initiation (McLaughlin, Murray and Rabinowitz, 1981). The data indicated that the calculated free energies of interaction of the SD regions in rRNA and mRNA was on average −17·6 kcal for Gram-positive organisms (McLaughlin, Murray and Rabinowitz, 1981) and −11 kcal for Gram-negative organisms (Gold *et al.*, 1981). The sequences at the 3′ end of 16S rRNA are $3'_{HO}$ UCUUUCCUCCACUAG-5′ and $3'_{HO}$ AUUCCUCCAUAG-5′ for *B. subtilis* (Murray and Rabinowitz, 1982) and *E. coli* (Gold *et al.*, 1981), respectively. The complementary sequence in the mRNA of *B. subtilis* very frequently is the five-base sequence GGAGG (McLaughlin, Murray and Rabinowitz, 1981; Moran *et al.*, 1982; Murray and Rabinowitz, 1982) whereas in *E. coli* there is often a four-base sequence AGGA (Gold *et al.*, 1981). The data in *Table 3* support the previous observations that the *B. subtilis* RBS has a relatively high calculated free energy of interaction of the SD regions. However, it is interesting to note that not all *B. subtilis* SD have the GGAGG sequence; in addition, some free-energy values are considerably lower than the average of −17·6 kcal. Therefore there are probably other parameters involved in the interaction between the mRNA and ribosome (Gold *et al.*, 1981).

If an increased complementarity between the SD regions is an important factor in increasing translational efficiency, then an optimization of the SD

Table 3. Putative ribosome-binding sites of *Bacilli*[*]

mRNA	Nucleotide sequence	ΔG′	Ref[†]
Subtilisin E (*spr*E)	UUUUUAAAAGGAGAGGUAAAGAGUG	14	1
Subtilisin	UGAAAAAAAGGAGAGGAUAAAGAGUG	15	2
α-amylase	UCAAAUAAGGAGUGUCAAGAAUG	13	3
spoOF	CGUUAAAAGGAGGAUUCACUUAUG	19	4
purF	CGUAUGGAAAGGAGCUAUCCCAUG	16	5
trpE	UAAGGAGUGAGAGCAAUG	13	6
trpD	CAUAGCAAGGAGGAUAAAGCUAUG	18	6
trpC	AAAGCAGAAAGAGGAAGAGAUCUAUG	14	6
crystal protein	AAAGAGAUGGAGGUAACUUAUG	17	7

[*] See Moran *et al.* (1982) and McLaughlin, Murray and Rabinowitz (1981) for additional data. ΔG′ was calculated by the method of Tinoco *et al.* (1973). The large letters represent complete homology with the Shine–Dalgarno sequence of *B. subtilis* (McLaughlin, Murray and Rabinowitz, 1981).

[†] (1) Wong *et al.*, 1984; (2) Wells *et al.*, 1983; (3) Yamazaki *et al.*, 1983; (4) Shimotsu *et al.*, 1983; (5) Makaroff *et al.*, 1983; (6) Band, Shimotsu and Henner, 1984; (7) Wong, Schnepf and Whiteley, 1983.

region in the mRNA should enhance the rate of protein synthesis. Band and Henner (1984) investigated this possibility by inserting portable RBS (de Boer *et al.*, 1983) in front of the leukocyte interferon A gene in a series of plasmids derived from plasmid pBS42 (Band and Henner, 1984). The only difference between these derived plasmids was the portable Shine–Dalgarno regions (PSDR). They found that the interferon activity in crude extracts of *B. subtilis* increased as the number of complementary bases in the PSDR increased from four to 12. When only four bases in the PSDR were complementary, little or no interferon activity was seen in *B. subtilis*, whereas good expression was seen in *E. coli* (the derived plasmids of pBS42 are capable of replicating in both host organisms). However, with seven and 12 complementary SD bases, increasing expression of interferon activity was observed in *B. subtilis*. Conversely, as the degree of complementarity increased from four to 12 bases, the activity in *E. coli* was actually reduced two- to sixfold depending on the specific insert.

The base composition of the region immediately following the region of complementarity also had a significant effect on the translation efficiency of the interferon gene. Translation was much more efficient with As and Ts following the SD region than Cs and Gs, as the presence of the latter bases reduced the expression level by a factor of 15 and 50, respectively. There was less effect of these bases in *E. coli*. Band and Henner (1984) concluded from these experiments that the increased complementarity of the SD region resulted in the higher expression of the interferon gene. The spacing between the SD region and ATG (initiation codon) was less important, although the composition of the spacer bases had a significant impact on expression with As and Ts favouring greater expression. On the other hand Chang *et al.*, (1983) did find that spacing distance between the SD sequence and the initiation codon ATG had a significant effect when the Pen P (β-lactamase gene) translation initiation sequence (i.e. SD sequence) was used to express interferon activity.

Most of the genes have an initiation codon of AUG; however, GUG has been found in three cases. They include the *sprE* gene (Wong *et al.*, 1984) and the *spoVG* and *0·3 kb* genes (Moran *et al.*, 1982). No evidence is available as to whether increased efficiency of initiation can be attributed to GUG.

Thus to optimize the expression of a gene at the translation level in *B. subtilis*, the translation initiation signal should be modified in the following ways:

1. Insertion of a PSDR containing 12 bases complementary to the 3′ end of the *B. subtilis* 16S rRNA;
2. Attachment of an eight-base spacer rich in As and Ts between the PSDR and the ATG (initiation codon).

STABILIZATION-ENHANCING SEQUENCES (SES)

Another factor which has to be considered for translation efficiency is the stability of the mRNA itself. Most mRNAs in *B. subtilis* have a short half-life of about 1 minute (Levinthal, Keynan and Higa, 1962; Leighton and Doi, 1971). In this regard it is possible to imagine structural modifications of the mRNA

which could result in greater resistance to exonucleases. Double-stranded regions of RNA are more resistant to attack from nucleases (Gillespie, 1968). Therefore, by inserting inverted repeat regions in the 5′ end of the mRNA preceding the SD region, one could conceivably form nuclease-resistant hairpin regions. This type of structural modification must not interfere with ribosome interaction with the SD sequence, must not create extensive hydrogen bonds with other regions of the 5′ of the mRNA, and must not be a target site for nucleases which recognize hairpin regions of RNA (e.g. nucleases which cleave precursor RNA (Abelson, 1979)).

Interestingly, the transcription-termination signal for many prokaryotic genes consist of CG-rich inverted repeats followed by several Us (Rosenberg and Court, 1979; Brennan and Geiduschek, 1983). The presence of this inverted region has been postulated to serve as a pause region for the RNA polymerase in order to facilitate the reactions necessary for transcription termination (Gilbert, 1976). Another possible function of this stem and loop structure at the 3′ end of the mRNA is to protect the mRNA from exonucleolytic attack during translation. The engineered addition of another inverted repeat proximal to the normal termination signals could serve the dual function of signalling a termination site as well as acting as a 3′ SES.

The efficacy of this idea will have to be tested somewhat empirically, because one cannot predict the consequence of inserted SES in mRNA translation. It is possible that an improper 5′ SES would hydrogen bond with the RBS or coding regions of the mRNA and actually depress the translation rate. Prior consideration of the 5′ sequence of the gene therefore is necessary to avoid this complication. The 3′ SES has to be designed to allow hairpin formation without negatively affecting the distal termination signal. The insertion of an excessively CG-rich inverted repeat may hinder passage of the RNA polymerase through the inserted hairpin sequence and the normal transcription termination signal and result in an excessive pause at this site, again resulting in delayed transcription and/or translation. In any case, the presence of SES could have an important role, especially during the stationary phase when the overall rate of RNA synthesis is reduced and the frequent utilization of mRNA by ribosomes would be advantageous for high protein production.

STRAIN IMPROVEMENT FOR TRANSLATION EFFICIENCY

If a heterologous cloned gene contained codons which were used infrequently by *B. subtilis*, a reduced rate of translation could result. Sequence data for *B. subtilis* genes are accumulating rapidly and it should be possible to determine which codons are not used frequently. On the assumption that the concentration of cognate tRNAs for these codons is low in *B. subtilis* and could limit the rate of translation (Ikemura, 1981), it should be possible to increase the concentration of those tRNAs by inserting the tRNA genes in a suitable plasmid. The tRNA genes would be linked to promoters which are expressed constitutively to produce increased amounts of these tRNAs. A number of tRNA genes in two operons have been cloned in *B. subtilis* (Green and Vold, 1983; Wawrousek and Hansen, 1983) and thus it should be possible to isolate

and construct suitable tRNA enrichment vectors for this purpose. The tRNA genes could be placed either with the gene of interest on the cloning plasmid or on a separate plasmid compatible with the cloning vector. In this manner the translation of codons used infrequently in *B. subtilis* would be facilitated.

Cloning vectors

PLASMID CLONING VECTORS

Two types of cloning vectors are used in *B. subtilis*: plasmids and phages. Many of the useful plasmids for cloning in *B. subtilis* have actually been derived from *S. aureus* and their properties have been summarized by Gryczan (1982). The properties of plasmids which make them suitable for cloning include a small molecular weight, unique restriction sites, relatively high copy number and favourable antibiotic markers. Many of the natural plasmids have also been recombined *in vitro* to form chimeric plasmids with multiple antibiotic-resistance markers. These plasmids with unique restriction sites in the antibiotic markers have been useful for cloning, because insertional inactivation of a marker can be used to identify recombinants readily.

The availability of *B. subtilis–E. coli* co-integrate shuttle plasmids has been extremely helpful. In most cases these shuttle vectors are recombinants of an *S. aureus* plasmid with *E. coli* plasmid pBR322 or its derivatives (Gryczan, 1982). These co-integrate plasmids are able to replicate DNA fragments with a relatively high degree of stability in *E. coli*. In addition, *E. coli* is able to express many heterologous genes including those regulated by *B. subtilis* transcription–translation signals and make it a suitable organism for detecting cloned genes. The gene of interest can be transferred by the co-integrate plasmid to *B. subtilis* where expression and secretion can be tested.

STABILIZATION OF PLASMIDS

For small-scale studies on gene cloning and expression, many of the currently used plasmids are suitable. However, for large-scale production a major problem facing the producer is the instability of plasmids containing recombinant DNA (Imanaka and Aiba, 1981). This instability can be manifested by either a complete or a partial loss of plasmids from the host cell or by rearrangements in the plasmid which result in partial or complete loss of the gene being cloned or expressed. In biotechnology processes there could be decreased product yield concomitant with the reduction in plasmid copy number. In addition, any plasmid-free segregant which arises could replace the resident population over a relatively few generation times, as bacteria carrying plasmids generally have a growth disadvantage in competition with plasmid-free cells under conditions which do not select for plasmid-coded functions (Jones *et al.*, 1980). This could have serious consequences if starter cultures for fermentations are prepared by a stepwise sequence involving increasing volumes of media or if a long fermentation period is used.

The maintenance of plasmids from generation to generation involves three

general types of functions: (1) replication; (2) cell division, and (3) partition (Ogura and Hiraga, 1983a,b). The replication functions involve the *ori* (origin of replication) locus and the machinery for replication of DNA; the cell division functions involve the *ccd* locus which is involved in coupling cell division to plasmid proliferation (Ogura and Hiraga, 1983b); and the partition functions involve the *par* and *sop* loci which are essential for normal partition of the plasmid into daughter cells during cell division (Meacock and Cohen, 1980; Ogura and Hiraga, 1983a). Most of these studies have been carried out with the *E. coli* plasmids such as the mini-F plasmid pSC138 (Timmis, Cabello and Cohen, 1975) and its recombinant derivatives (Ogura and Hiraga, 1983a,b), and pSC101 (Meacock and Cohen, 1980). Even among *E. coli* plasmids, not all have the same functions and requirements for plasmid stability. Very little is known about such functions in *B. subtilis* plasmids, but it can be assumed that similar functions are probably involved in the stabilization of plasmids in *B. subtilis*.

A start in understanding plasmid stability in *B. subtilis* was reported by Chang *et al.* (1983) who inserted an *Mbo*I DNA fragment obtained from plasmid pLS11 (Tanaka, Kuroda and Sakaguchi, 1977), a naturally occurring *B. subtilis* plasmid, into pOG2326, a bifunctional plasmid which was able to replicate in both *E. coli* and *B. subtilis*. pOG2326 was constructed from plasmids pOG1196 and pBR322. The pOG2326–*Mbo*I plasmid was transformed into *B. subtilis* BD224 and grown in the presence of chloramphenicol (one of the genes on pOG2326 was for chloramphenicol resistance) and then in the absence of chloramphenicol. After more than 100 generations in the absence of chloramphenicol, virtually 100% of the *B. subtilis* BD224 still carried the plasmid pOG2381 with an *Mbo*I fragment and was chloramphenicol resistant. This plasmid was extremely stable. The authors interpreted these results as stabilization of the plasmid by insertion of a normal *B. subtilis* plasmid *par* (partition function) function into the co-integrate plasmid lacking such a function. These results suggest that all the regulatory signals for plasmid maintenance and partition should be of *B. subtilis* origin and that they should be present in the recombinant plasmid.

PRACTICAL CONSIDERATIONS FOR STABILIZING RECOMBINANT PLASMIDS

Although the genetic factors for maintaining the stability of plasmids in *B. subtilis* are not known, many factors have been analysed which affect the maintenance of recombinant plasmids in bacteria. These factors include (1) the growth rate of the host cell; (2) the environmental stress imposed on the host cell; (3) characteristics of the host cell; (4) characteristics of the plasmid; (5) presence of a recombinant gene on the plasmid (for review *see* Imanaka and Aiba (1981)).

Growth-limiting conditions may lead to plasmid segregation and loss (Noack *et al.*, 1981). Substrate limitation which inhibits the efficiency of plasmid replication results in the reduction of copy number of the plasmid and ultimately in a plasmid-free cell. As plasmid-carrying cells generally have a growth disadvantage in competition with plasmid-free cells, plasmid-free cells

will overgrow the plasmid carriers. Therefore cultivation conditions should be controlled to prevent reduction in plasmid copy number.

Environmental stress, e.g. the presence of an antibiotic, will in many cases force the host cell to maintain a plasmid containing the antibiotic-resistance gene. Although this does result in maintenance of the plasmid and the selective gene, it still could lead to rearrangements or even to complete loss of the recombinant gene of interest. At times the selective gene will become integrated into the host chromosome and leave a partially deleted plasmid or a plasmid-free cell. For large-scale maintenance of plasmids the use of environmental stress may not be practical, although one could imagine a situation in which a chromosomal gene is making an essential, but temperature-sensitive enzyme and the plasmid-borne gene is making a temperature-resistant form of the enzyme. In this case the use of *rec*$^-$ (recombinant-deficient) host and high growth temperature may favour the maintenance of the plasmid and the required gene.

An important consideration for the maintenance of plasmids would be the host and plasmid genotypes. Although this could vary for the particular plasmid being used, certain basic properties of the host should be considered. Of utmost importance is the compatibility of the host with the proliferation properties of the plasmid. As mentioned earlier in this discussion, the replication of DNA, cell division coupled to plasmid replication, and partitioning of plasmids into daughter cells, would have to be carried out efficiently. The recognition of the specificity of protein–protein and protein–DNA interactions forces one to conclude that the use of native *B. subtilis* plasmids would be the best starting point for the development of stable *B. subtilis* plasmids. Investigators have tended not to use *B. subtilis* plasmids, because they do not have selectable markers. However, it would seem logical that the *B. subtilis* replication and cell-division machinery must work efficiently with the *ori*, *ccd*, and *par* genes (Ogura and Hiraga, 1983a,b) and the recognition sites for separation enzymes of naturally resident plasmids. In addition, it is likely that sequences in the plasmid which might lead to illegitimate recombination events have been removed during the long-term residence of the plasmid in the host. Thus a homologous *B. subtilis*-plasmid system should be most stable and suitable for cloning and gene expression in *B. subtilis*.

The use of *rec*$^-$ host cells has not been the complete answer for increasing plasmid stability in *B. subtilis*, as illegitimate recombination occurs between heterologous DNA and *B. subtilis* DNA. Illegitimate recombination appears to be mediated by DNA gyrase in *E. coli* (Ikeda, Aoki and Naito, 1982). A highly active DNA gyrase system in *B. subtilis* may present a formidable barrier to improvement of plasmid stability. Although there is some relationship between high copy number of a recombinant plasmid and the yield of desired products (i.e. by dose response), a moderate copy number may favour product yield and plasmid stability more than extremely high copy number plasmids (Aiba, Tsunekawa and Imanaka, 1982).

Part of the instability of plasmids might, however, be linked directly to the presence of the recombinant gene on the plasmid, because its presence can affect stability in several ways. The high expression of the gene could cause

considerable stress on cell metabolism and decrease growth rate to such a point that plasmid replication itself is affected (Aiba, Tsunekawa and Imanaka, 1982). Expression from particular efficient promoters on an inserted gene can interfere with plasmid replication by transcriptional read-through from the inserted gene into the promoter priming DNA replication (Steuber and Bujard, 1982). Furthermore, transcriptional read-through could lead to the production of *rop* (repressor of primer) gene product (Twigg and Sherratt, 1980; Cesareni, Muesing and Polisky, 1982) which negatively affected copy number of plasmid pBR322 (Steuber and Bujard, 1982).

In summary, plasmid instability in *B. subtilis* is one of the major barriers for the use of plasmid-carried recombinant genes in large-scale production. However, the following points might be considered to overcome this hurdle:

1. The use of media which do not limit cell growth and plasmid replication;
2. The use of a *native B. subtilis* plasmid with suitable replication and partitioning properties which are compatible with the host and selected naturally for stability;
3. The use of moderate copy number plasmid to prevent excessive pressures on the cell's metabolism. Thus, normal cell growth will be maintained, as well as plasmid replication;
4. The use of strong transcription terminators on inserted genes to prevent disruption of normal plasmid replication functions by read-through transcription;
5. The use of promoters on the inserted gene which can be turned on after maximum cell and plasmid growth has been attained. This may prevent gene expression from causing plasmid instability during cell growth, especially during build-up of inoculum for large-scale fermentation.

PHAGES USED IN CLONING STUDIES

Two systems have been developed for cloning genes by *B. subtilis* phages. Kawamura *et al.* (1981) have constructed specialized transducing phages from temperate phages ρ11 and ϕ105 by using a novel prophage transformation method. They cloned the *spoOF*$^+$ gene of *B. subtilis* in the following way. *B. subtilis spo*$^+$ DNA and phage ρ11 DNA were restricted with *Eco* RI, mixed and then ligated. This ligated DNA was then added to competent cells of *B. subtilis* which were *spoOF* and lysogenic for ρ11. When the *Eco* RI fragment containing the *spoOF*$^+$ gene was ligated between two naturally flanking ρ11 *Eco* RI fragments and taken up by the competent cell, two different events could occur to produce *spo*$^+$ transformants: (1) the *spoOF*$^+$ fragment could transform the *spoOF* recipient to *spo*$^+$ by transformation at the *spoOF* locus and leave the ρ11-lysogen intact, or (2) the ρ11 DNA fragments flanking the *spoOF*$^+$ DNA fragment could be integrated into the ρ11-lysogen and insert the *spoOF*$^+$ fragment simultaneously in the prophage. Induction of the two types of lysogen will result in two different types of phage. In case (1), induction will produce a completely normal ρ11 phage; in case (2), induction will produce a ρ11 phage carrying the *spoOF*$^+$ gene. These two types of phage could be

separated by infecting *spoOF*$^-$ cells which were non-lysogenic. Only the phage carrying the *spoOF*$^+$ fragment would transduce these recipients to *spo*$^+$ by integrating into the *B. subtilis* chromosome at the proper ρ11 attachment site. Induction of these cells would then result in a population of specialized transducing ρ11 phage containing the cloned *spoOF*$^+$ gene. Similar types of construction were also carried out with ϕ105, and *spoOF*$^+$ and *spoOB*$^+$ genes were cloned. Each phage has its advantages: ρ11 was found to be the most efficient vector for shotgun experiments with various *B. subtilis* genes, perhaps because of its larger size; ϕ105 is a smaller phage and is convenient for isolating individual DNA fragments from the restricted recombinant phage. These temperate phages do have the advantage of inserting single copies of a desired gene into the chromosome of *B. subtilis*. They would, therefore, be convenient for expression of a recombinant gene which might be deleterious to the cell when expressed in excessive amounts from high copy number plasmids. Furthermore, ϕ105 phage carrying recombinant genes can be used for gene conversion (Iglesias *et al.*, 1981) of certain markers (F. Kawamura and R.H. Doi, unpublished data).

Heilmann and Reeve (1982) have constructed a viral cloning vector by use of the virulent bacteriophage SPP1 (Riva, Polsinelli and Falaschi, 1968). This well-characterized phage contains non-essential regions which comprise up to 10% of the genome (Behrens *et al.*, 1979). A unique *Bam* HI restriction site was inserted into the non-essential region of the genome of a deletion mutant (SPP1v) of SPP1. SPP1v has been used to clone *Bam* HI, *Bgl II and Bcl* I fragments by restricting the phage with *Bam* HI, insertion of the fragments, ligation and transfection of a suitable *B. subtilis* host. An enrichment method for recombinant molecules involves the insertion of *Bgl* II or *Bcl* I fragments into the *Bam* HI site which destroys the *Bam* HI site. Treatment of the recombinant molecules with *Bam* HI destroys *Bam* HI-treated phage molecules which have reclosed without an insert, and only intact phage containing either a *Bgl* II or *Bcl* I fragment can transfect. The presence of two *Xma* III sites directly adjacent to, and on both sides of, the inserted *Bam* HI site allows the precise excision of cloned DNA even when the *Bam* HI site has been destroyed by insertion of *Bgl* II or *Bcl* I fragments. The size of the DNA which can be cloned is about 4 mD (6 kbp) which may be a limiting factor in the use of this system.

The practical use of a plasmid or phage vector will depend on the goals of the investigator and/or producer. For small-scale studies both systems can be used relatively well. However, for large-scale production of proteins the utilization of a stable plasmid system is more appealing.

Secretion and signal peptides

One of the major advantages for using *B. subtilis* as a host for expression of cloned genes is its ability to secrete proteins to the extracellular medium. This capability would increase the stability and facilitate the purification of cloned gene products. Studies on both prokaryotic and eukaryotic extracellular proteins have revealed that these proteins are made in a precursor form.

Several genes for secreted proteins have been cloned and sequenced from *Bacillus* species (*Figure 8*). A consistent feature of these proteins is the presence of a signal peptide (Milstein *et al.*, 1972) at the *N*-terminus of the precursor and it is believed that the hydrophobic properties of this signal peptide play a critical part in the transport of the molecule across the membrane (Blobel and Dobberstein, 1975). There is a signal peptidase site between the signal peptide and the mature protein which serves as the cleavage site to remove the signal peptide from the mature protein after it has crossed the membrane (Kreil, 1981). As the signal peptide has a major role in transmembrane transport of extracellular proteins, it is logical that much attention has been focused on the possibility of using a *B. subtilis* signal peptide linked to a protein of interest, for secretion of that protein.

The complex nature of the secretion process has precluded the successful secretion of a protein which ordinarily is not secreted. The simple process of constructing a recombinant gene containing the promoter and signal peptide sequence linked to a gene in the correct reading frame can lead to the synthesis of a fusion product, but this product may not be secreted and could be lethal to the cell. In one example the promoter and signal peptide from *B. licheniformis* β-lactamase was attached to the chloramphenicol acetyltransferase gene from Tn9 in the proper reading frame in the shuttle plasmid pGR71 (Goldfarb, Doi and Rodriguez, 1981). Although some *E. coli* clones of this recombinant plasmid were chloramphenicol resistant, they were very slow growers and the cells grew as long snakes. When the same recombinant plasmid was inserted into *B. subtilis*, it apparently was lethal, for no chloramphenicol-resistant *B. subtilis* clones were ever obtained (D.S. Goldfarb and R.H. Doi, unpublished data).

A successful recombinant gene has been constructed for secretion when the promoter and signal peptide sequence from one extracellular protein gene was attached to the sequence of a mature protein coded by another extracellular gene (Palva *et al.*, 1982). In this case the promoter and signal peptide from *B. amyloliquefaciens* α-amylase gene was fused through a *Hin*d III linker to the structural part of the *E. coli* β-lactamase gene which lacked its own signal sequence. When this recombinant gene was inserted into pUB110 and then transformed into *B. subtilis*, more than 95% of the β-lactamase was secreted into the growth medium (Palva *et al.*, 1982). If the six carboxy-terminal amino acids (−6) of the signal peptide were absent, 90% of the β-lactamase remained cell-bound and the total activity was only 10% of that obtained with the complete signal peptide. If the β-lactamase was fused to the *N*-terminal 14 amino acids (+14) of the α-amylase, the total activity was 20% less than the fusion at −1 or +4 amino acids. The size of the extracellular product indicated that the signal sequence was processed at or near the correct position in these constructs. Palva *et al.* (1983) have also used the α-amylase signal peptide to secrete mature human interferon α2. Again, cleavage occurred at the end of the signal peptide, precisely as with α-amylase. These results demonstrated that the use of a *Bacillus* signal peptide could be used to secrete a protein which is normally secreted if fusion is at the correct position, that the amino-acid sequence after the signal peptide may not be critical for the specificity of the *B.*

Extracellular Protein	Ref.*
B. subtilis subtilisin E	
MetArgSerLysLys|LeuTrpIleSerLeuLeuPheAlaLeuThrLeuIlePheThrMetAlaPheSerAsnMetSerAlaGlnAlaAla	1
B. amyloliquefaciens subtilisin	
MetArgGlyLysLys|ValTrpIleSerLeuLeuPheAlaLeuAlaLeuIlePheThrMetAlaPheGlySerThrSerSerAlaGlnAlaAla	2
B. subtilis α-amylase	
MetPheAlaLysArgPheLys|ThrSerLeuLeuProLeuPheAlaGlyPheLeuLeuLeuPheTyrLeuValLeuAlaGlyProAlaAlaAlaSerAla	3
B. amyloliquefaciens α-amylase	
MetIleGluLysArgLysArg|ThrValSerPheArgLeuValLeuMetCysThrLeuLeuPheValSerLeuProIleThrIleThrSerAla↓Val	4
B. licheniformis β-lactamase	
MetLysLeuTrpPheSerThrLeuLysLeuLysLys|AlaAlaAlaValLeuLeuPheSerCysValAlaLeuAlaGlyCysAlaAsnAsnGlnThrAsnAla↓Ser	5
B. cereus β-lactamase	
MetMetIleLeuLysAsnLysArgMetLeuLys|IleGlyIleCysValGlyIleLeuGlyLeuSerIleThrSerLeuGluAlaPheThrGlyGluSer	6

Figure 8. Signal peptide sequences in *Bacilli*. The vertical line demarcates the *N*-terminal charged residues from the hydrophobic region of the signal peptide. The underlined residues indicate putative signal peptidase sites. The arrows indicate the site of signal peptidase activity on these propeptides. References: (1) Wong *et al.*, 1984; (2) Wells *et al.*, 1983; (3) Yamazaki *et al.*, 1983 and Yang, Galizzi and Henner, 1983; (4) Takkinen *et al.*, 1983; (5) Neugebauer, Sprengel and Schaller, 1981; Kroyer and Chang, 1981; (6) Mezes *et al.*, 1983.

subtilis signal peptidase, and that synthesis and secretion time is controlled by the promoter. Thus, with the isolation of an increasing number of genes for extracellular enzymes for *B. subtilis*, this strategy with signal peptides will be useful for studying the secretion of proteins which normally are not secreted by the cell.

This latter point is critical, since normally intracellular proteins may never be secreted even when fused to signal peptides and may require the development and use of leaky cells (Kudo, Kato and Horikoshi, 1983) or other strategies.

Conclusions

Although significant progress has been made concerning our knowledge of genetic expression in *B. subtilis*, several areas require further study for obtaining optimum expression of recombinant genes on a large scale:

1. The properties of efficient and regulated promoters will have to be elucidated. At the moment it is not clear what sequence properties make a promoter efficient. At present there are techniques to obtain naturally efficient promoters, but they may not serve the proper purpose and may require modifications, e.g. for temporal regulation of their expression.
2. The properties of the ribosome-binding site should be optimized. Although the Shine–Dalgarno sequence plays a part, there may be other factors, such as secondary structures at the 5′ end of mRNAs, that have a critical role for efficient initiation of translation.
3. The stabilization of recombinant plasmids to allow maximum gene-dosage effects for the production of proteins is a major problem facing the *Bacillus* investigators. The presence of an efficient illegitimate recombination system may prove to be a big hurdle. The utilization of naturally resident *B. subtilis* plasmids may be helpful.
4. The effective use of *B. subtilis* signal peptides for the secretion of heterologous proteins has been demonstrated for normally secreted proteins which have evolved to pass through the membrane. A more difficult problem is to find the proper conditions for secretion of proteins which normally are *not* secreted, because their chemical properties may prevent their passage through the hydrophobic environment of the membrane. This may require a modification of this strategy.
5. *B. subtilis* strains that are low in extracellular proteases, that are compatible with plasmid vectors, and that possess low illegitimate recombination activity should be isolated.

None of these problems are insurmountable. It will be interesting to see whether empirical methods of selecting desired properties or the new recombinant DNA techniques will provide the answers. Probably a combination of the two approaches will be most suitable for the near future.

Acknowledgement

The author's research has been supported by grants from the National Science Foundation and the National Institute of General Medical Sciences. He thanks Chester W. Price, Fujio Kawamura and Pei-Zhi Wang for helpful discussions.

References

ABELSON, J. (1979). RNA processing and the intervening sequence problem. *Annual Review of Biochemistry* **48**, 1035–1069.

AIBA, S., TSUNEKAWA, H. AND IMANAKA, T. (1982). New approach to tryptophan production by *Escherichia coli*: genetic manipulation of composite plasmids *in vitro*. *Applied and Environmental Microbiology* **43**, 289–297.

BAND, L. AND HENNER, D.J. (1984). *Bacillus subtilis* requires a "stringent" Shine–Dalgarno region for gene expression. *DNA* **3**, 17.

BAND, L., SHIMOTSU, H. AND HENNER, D.J. (1984). Nucleotide sequence of the *B. subtilis trp*E and *trp*D genes. *Gene* **27**, 55–65.

BAND, L., YANSURA, D.G. AND HENNER, D.J. (1983). Construction of a vector for cloning promoters in *Bacillus subtilis*. *Gene* **26**, 313–315.

BANNER, C.D.B., MORAN, C.P., JR AND LOSICK, R. (1983). Deletion analysis of a complex promoter for a developmentally regulated gene from *Bacillus subtilis*. *Journal of Molecular Biology* **168**, 351–365.

BEHRENS, B., LUDER, G., BEHNCKE, M., TRAUTNER, T.A. AND GANESAN, A.T. (1979). The genome of *B. subtilis* phage SPP1. Physical arrangement of phage genes. *Molecular and General Genetics* **175**, 351–357.

BERK, A.J. AND SHARP, P.A. (1977). Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. *Cell* **12**, 721–732.

BLOBEL, G. AND DOBBERSTEIN, B. (1975). Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *Journal of Cell Biology* **67**, 835–851.

BREHM, S.P., STAAL, S.P. AND HOCH, J.A. (1973). Phenotypes of pleiotropic-negative sporulation mutants of *Bacillus subtilis*. *Journal of Bacteriology* **115**, 1063–1070.

BRENNAN, S.M. AND GEIDUSCHEK, E.P. (1983). Regions specifying transcriptional termination and pausing in the bacteriophage SPO1 terminal repeat. *Nucleic Acids Research* **11**, 4157–4175.

CESARENI, G., MUESING, M.A. AND POLISKY, B. (1982). Control of ColE1 DNA replication: The *rop* gene product negatively affects transcription from the replication primer promoter. *Proceedings of the National Academy of Sciences of the United States of America*. **79**, 6613–6317.

CHANG, S., HO, D., GRAY, O., CHANG, S.-Y. AND MCLAUGHLIN, J. (1983). Functional expression of human interferon genes and construction of a partition-proficient plasmid vector in *B. subtilis*. In *Genetics of Industrial Microorganisms 1982* (Y. Ikeda and T. Beppu, Eds), pp. 227–231. Kodansha Ltd., Tokyo.

DEAN, D.H. AND KAELBLING, M.J. (1981). A genetic engineering manifesto for the genus *Bacillus*. *Annals of the New York Academy of Sciences* **369**, 23–32.

DEBABOV, V.G. (1982). The industrial use of *Bacilli*. In *The Molecular Biology of the Bacilli, Volume 1:* Bacillus subtilis (D.A. Dubnau, Ed.), pp. 331–370. Academic Press, New York.

DE BOER, H.A., COMSTOCK, L.J. AND VASSER, M. (1983). The TAC promoter — a functional hybrid derived from the Trp and Lac promoters. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 21–25.

DE BOER, H.A., COMSTOCK, L.J., HUI, A., WONG, E. AND VASSER, M. (1983). Portable Shine–Dalgarno regions: nucleotides between the Shine–Dalgarno sequence and the start codon affect the translation efficiency. In *Gene Amplification and Analysis*

(T.S. Papas, M. Rosenberg, and J.G. Chirikjian, Eds), volume 3. Elsevier Science Publishing Co., Inc., New York.

Doi, R.H. (1982a). Multiple RNA polymerase holoenzymes exert transcriptional specificity in *Bacillus subtilis*. *Archives of Biochemistry and Biophysics* **214**, 772–781.

Doi, R.H. (1982b). RNA polymerase of *Bacillus subtilis*. In *The Molecular Biology of the Bacilli, Volume I:* Bacillus subtilis (D.A. Dubnau, Ed.), pp. 72–110. Academic Press, New York.

Doi, R.H., Halling, S.M., Williamson, V.M. and Burtis, K.C. (1980). The transcriptional apparatus of *Bacillus subtilis:* a model for Gram-positive prokaryotes. In *Genetics and Evolution of RNA Polymerase, tRNA and Ribosomes* (S. Osawa, H. Ozeki, H. Uchida and T. Yura, Eds), pp. 117–136. University of Tokyo Press, Tokyo.

Donnelly, C.E. and Sonenshein, A.L. (1984). Promoter-probe plasmid for *Bacillus subtilis*. *Journal of Bacteriology*, **157**, 965–967.

Duvall, E.J., Williams, D.M., Lovett, P.S., Rudolph, C., Vasantha, N. and Guyer, M. (1983). Chloramphenicol-inducible gene expression in *Bacillus subtilis*. *Gene* **24**, 171–177.

Franklin, F.C.H., Bagdasarian, M., Bagdasarian, M.M. and Timmis, K.N. (1981). Molecular and functional analysis of the TOL plasmid pWWO from *Pseudomonas putida* and cloning of genes for the entire regulated aromatic ring *meta* cleavage pathway. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 7458–7462.

Fukuda, R. and Doi, R.H. (1977). Two polypeptides associated with ribonucleic acid polymerase core of *Bacillus subtilis* during sporulation. *Journal of Bacteriology* **129**, 422–432.

Gentz, R., Langner, A., Chang, A.C.Y., Cohen, S.N. and Bujard, H. (1981). Cloning and analysis of strong promoters is made possible by the downstream placement of a RNA termination signal. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 4936–4940.

Gilbert, W. (1976). Starting and stopping sequences for the RNA polymerase. In *RNA Polymerase* (R. Losick and M. Chamberlin, Eds), pp. 193–205. Cold Spring Harbor Laboratory, New York.

Gillespie, D. (1968). The formation and detection of DNA–RNA hybrids. *Methods in Enzymology* **12**, Part B, 641–668.

Gilman, M.Z. and Chamberlin, M.J. (1983). Developmental and genetic regulation of the *Bacillus subtilis* genes transcribed by σ^{28}–RNA polymerase. *Cell* **35**, 285–293.

Gilman, M.Z., Wiggs, J.L. and Chamberlin, M.J. (1981). Nucleotide sequences of two *Bacillus subtilis* promoters used by *Bacillus subtilis* sigma-28 RNA polymerase. *Nucleic Acids Research* **9**, 5991–6000.

Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B.S. and Stormo, G. (1981). Translation initiation in prokaryotes. *Annual Review of Microbiology* **35**, 365–403.

Goldfarb, D.S., Doi, R.H. and Rodriguez, R.L. (1981). Expression of Tn9-derived chloramphenicol resistance in *Bacillus subtilis*. *Nature* **293**, 309–311.

Goldfarb, D.S., Rodriguez, R.L. and Doi, R.H. (1982). Translation block to the expression of the *Escherichia coli* Tn9-derived chloramphenicol resistance gene in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 5886–5890.

Goldfarb, D.S., Wong, S.-L., Kudo, T. and Doi, R.H. (1983). A temporally regulated promoter from *Bacillus subtilis* is transcribed only by an RNA polymerase with a 37,000 dalton sigma factor. *Molecular and General Genetics* **191**, 319–325.

Gray, O. and Chang, S. (1981). Molecular cloning and expression of *Bacillus licheniformis* β-lactamase gene in *Escherichia coli* and *Bacillus subtilis*. *Journal of Bacteriology* **145**, 422–428.

GREEN, C.J. AND VOLD, B.S. (1983). Sequence analysis of a cluster of twenty-one tRNA genes in *Bacillus subtilis*. *Nucleic Acids Research* **11**, 5763–5774.

GRYCZAN, T.J. (1982). Molecular cloning in *Bacillus subtilis*. In *The Molecular Biology of the Bacilli, Volume 1*: Bacillus subtilis (D. Dubnau, Ed.), pp. 307–329. Academic Press, New York.

HALDENWANG, W.G. AND LOSICK, R. (1980). A novel RNA polymerase σ factor from *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America* **77**, 7000–7004.

HALDENWANG, W.G., LANG, N. AND LOSICK, R. (1981). A sporulation-induced sigma-like regulatory protein from *B. subtilis*. *Cell* **23**, 615–624.

HARWOOD, C.R., WILLIAMS, D.M. AND LOVETT, P.S. (1983). Nucleotide sequence of a *Bacillus pumilus* gene specifying chloramphenicol acetyltransferase. *Gene* **24**, 163–169.

HEILMANN, H. AND REEVE, J.N. (1982). Construction and use of SPP1v, a viral cloning vector for *Bacillus subtilis*. *Gene* **17**, 91–100.

IGLESIAS, A., BENSI, G., CANOSI, U. AND TRAUTNER, T.A. (1981). Plasmid transformation in *Bacillus subtilis*. Alterations introduced into the recipient homologous DNA of hybrid plasmids can be corrected in transformation. *Molecular and General Genetics* **184**, 405–409.

IKEDA, H., AOKI, K. AND NAITO, A. (1982). Illegitimate recombination mediated *in vitro* by DNA gyrase of *Escherichia coli*: structure of recombinant DNA molecules. *Proceedings of the National Academy of Sciences of the United States of America*. **79**, 3724–3728.

IKEMURA, T. (1981). Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein gene: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *Journal of Molecular Biology* **151**, 389–409.

IMANAKA, T. AND AIBA, S. (1981). A perspective on the application of genetic engineering: stability of recombinant plasmid. *Annals of the New York Academy of Sciences* **369**, 1–14.

INOUYE, S., NAKAZAWA, A. AND NAKAZAWA, T. (1981). Molecular cloning of TOL genes *xylB* and xylE in *Escherichia coli*. *Journal of Bacteriology* **145**, 1137–1143.

JOHNSON, W.C., MORAN, C.P. JR AND LOSICK, R. (1983). Two RNA polymerase sigma factors from *Bacillus subtilis* discriminate between overlapping promoters for a developmentally regulated gene. *Nature* **302**, 800–804.

JONES, I.M., PRIMROSE, S.B., ROBINSON, A. AND ELLWOOD, D.C. (1980). Maintenance of some ColE1-type plasmids in chemostat culture. *Molecular and General Genetics* **180**, 579–584.

KAWAMURA, F., SHIMOTSU, H., SAITO, H., HIROCHIKA, H. AND KOBAYASHI, Y.(1981). Cloning of *spoO* genes with bacteriophage and plasmid vectors in *Bacillus subtilis*. In *Sporulation and Germination* (H.S. Levinson, A.L. Sonenshein and D.J. Tipper, Eds), pp. 109–113. American Society for Microbiology, Washington, DC.

KLIER, A., FARGETTE, F., RIBIER, J. AND RAPOPORT, G. (1982). Cloning and expression of the crystal protein genes from *Bacillus thuringiensis* strain *berliner* 1715. *EMBO Journal* **1**, 791–799.

KREFT, H., BERGER, H., HARTLEIN, M., MULLER, B., WEIDINGER, G. AND GOEBEL, W. (1983). Cloning and expression in *Escherichia coli* and *Bacillus subtilis* of the hemolysin (Cereolysin) determinant from *Bacillus cereus*. *Journal of Bacteriology* **155**, 681–689.

KREIL, G. (1981). Transfer of proteins across membranes. *Annual Review of Biochemistry* **50** 317–348.

KROYER, J. AND CHANG, S. (1981). The promoter-proximal region of the *Bacillus licheniformis* penicillinase gene: nucleotide sequence and predicted leader peptide sequence. *Gene* **15**, 343–347.

KUDO, T., KATO, C. AND HORIKOSHI, K. (1983). Excretion of the penicillinase of an alkalophilic *Bacillus* sp. through the *Escherichia coli* outer membrane. *Journal of*

Bacteriology **156**, 949–951.

Lee, G., Talkington, C. and Pero, J. (1980). Nucleotide sequence of a promoter recognized by *Bacillus subtilis* RNA polymerase. *Molecular and General Genetics* **180**, 57–65.

Legault-Demare, L. and Chambliss, G.H. (1975). Selective messenger translation by *Bacillus subtilis* ribosomes. *Molecular and General Genetics* **142**, 277–287.

Leighton, T.J. and Doi, R.H. (1971). The stability of messenger ribonucleic acid during sporulation in *Bacillus subtilis*. *Journal of Biological Chemistry* **246**, 3189–3195.

Levinthal, C., Keynan, A. and Higa, K. (1962). Messenger RNA turnover and protein synthesis in *Bacillus subtilis* inhibited by actinomycin D. *Proceedings of the National Academy of Sciences of the United States of America* **48**, 1631–1638.

Losick, R. (1982). Sporulation genes and their transcription. In *The Molecular Biology of the Bacilli, Volume I*: Bacillus subtilis (D.A. Dubnau, Ed.), pp. 179–201. Academic Press, New York.

Losick, R. and Pero, J. (1981). Cascades of sigma factors. *Cell* **25**, 582–584.

Losick, R. and Sonenshein, A.L. (1969). Change in the template specificity of RNA polymerase during sporulation of *Bacillus subtilis*. *Nature* **224**, 35–37.

McKenney, K., Shimatake, H., Court, D., Schmeissner, U., Brady, C. and Rosenberg, M. (1981). A system to study promoter and terminator signals recognized by *Escherichia coli* RNA polymerase. In *Gene Amplification and Analyses, Volume 1. Structural Analysis of Nucleic Acids* (J.G. Chirikjian and T.S. Papes, Eds), pp. 383–415. Elsevier/North-Holland.

McLaughlin, J.R., Murray, C.L. and Rabinowitz, J.C. (1981). Unique features in the ribosome binding site sequence of the Gram-positive *Staphylococcus aureus* β-lactamase gene. *Journal of Biological Chemistry* **256**, 11283–11291.

Makaroff, C.A., Zalkin, H., Switzer, R.L. and Vollmer, S.J. (1983). Cloning of the *Bacillus subtilis* glutamine phosphoribosylpyrophosphate amidotransferase gene in *Escherichia coli*. *Journal of Biological Chemistry* **258**, 10586–10593.

Meacock, P.A. and Cohen, S.N. (1980). Partitioning of bacterial plasmids during cell division: a *cis*-acting locus that accomplishes stable plasmid inheritance. *Cell* **20**, 529–542.

Mezes, P.S.F., Yang, Y.O., Hussain, M. and Lampen, J.O. (1983). *Bacillus cereus* 569/H β-lactamase I: cloning in *Escherichia coli* and signal sequence determination. *FEBS Letters* **161**, 195–200.

Milstein, C., Brownlee, G.G., Harrison, T.M. and Matthews, M.B. (1972). A possible precursor of immunoglobulin light chains. *Nature New Biology* **239**, 117–120.

Mongkolsuk, S., Chiang, Y.-W., Reynolds, R.B. and Lovett, P.S. (1983). Restriction fragments that exert promoter activity during postexponential growth of *Bacillus subtilis*. *Journal of Bacteriology* **155**, 1399–1406.

Moran, C.P., Jr, Lang, N. and Losick, R. (1981). Nucleotide sequence of a *Bacillus subtilis* promoter recognized by *Bacillus subtilis* RNA polymerase containing σ^{37}. *Nucleic Acids Research* **9**, 5979–5990.

Moran, C.P., Jr, Lang, N., Banner, C.D.B., Haldenwang, W.G. and Losick, R. (1981). Promoter for a developmentally regulated gene in *Bacillus subtilis*. *Cell* **25**, 783–791.

Moran, C.P., Jr, Lang, N., Legrice, S.F.J., Lee, G., Stephens, M., Sonenshein, A.L., Pero, J. and Losick, R. (1982). Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. *Molecular and General Genetics* **186**, 339.

Murray, C.L. and Rabinowitz, J.C. (1982). Nucleotide sequences of transcription and translation initiation regions in *Bacillus* phage ϕ29 early genes. *Journal of Biological Chemistry* **257**, 1053–1062.

Nakayama, T., Williamson, V., Burtis, K. and Doi, R.H. (1978). Purification and properties of two RNA polymerases from sporulating cells of *Bacillus subtilis*. *European Journal of Biochemistry* **88**, 155–164.

NEUGEBAUER, K., SPRENGEL, R. AND SCHALLER, H. (1981). Penicillinase from *Bacillus licheniformis:* nucleotide sequence of the gene and implications for the biosynthesis of a secondary protein in a Gram-positive bacterium. *Nucleic Acids Research* **9**, 2577–2588.

NOACK, D., ROTH, M., GEUTHER, R., MULLER, G., UNDISZ, K., HOFFMEIER, C. AND GASPAR, S. (1981). Maintenance and genetic stability of vector plasmids pBR322 and pBR325 in *Escherichia coli* K12 strains grown in a chemostat. *Molecular and General Genetics* **184**, 121–124.

OGASAWARA, N., MORIYA, S. AND YOSHIKAWA, H. (1983). Structure and organization of rRNA operons in their region of the replication origin of the *Bacillus subtilis* chromosome. *Nucleic Acids Research* **11**, 6301–6318.

OGURA, T. AND HIRAGA, S. (1983a). Partition mechanism of F plasmid: two plasmid gene-encoded products and a *cis*-acting region are involved in partition. *Cell* **32**, 351–360.

OGURA, T. AND HIRAGA, S. (1983b). Mini-F plasmid genes that couple host cell division to plasmid proliferation. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 4784–4788.

OLLINGTON, J.F., HALDENWANG, W.G., HUYNH, T.V. AND LOSICK, R. (1981). Developmentally regulated transcription in a cloned segment of the *Bacillus subtilis* chromosome. *Journal of Bacteriology* **147**, 432–442.

ORTLEPP, S.A., OLLINGTON, J.F. AND MCCONNELL, D.J. (1983). Molecular cloning in *Bacillus subtilis* of a *Bacillus licheniformis* gene encoding a thermostable alpha amylase. *Gene* **23**, 267–276.

PALVA, I. (1982). Molecular cloning of α-amylase gene from *Bacillus amyloliquefaciens* and its expression in *Bacillus subtilis*. *Gene* **19**, 81–87.

PALVA, I., SARVAS, M., LEHTOVAARA, P., SIBAZKOV, M. AND KAARIAINEN, L. (1982). Secretion of *Escherichia coli* β-lactamase from *Bacillus subtilis* by the aid of α-amylase signal sequence. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 5582–5586.

PALVA, I., LEHTOVAARA, P., KAARIANINEN, L., SIBAKOV, M., CANTELL, K., SCHEN, C.H., KASHIWAGI, K. AND WEISSMANN, C. (1983). Secretion of interferon by *Bacillus subtilis*. *Gene* **22**, 229–235.

PRICE, E.W., GITT, M.A. AND DOI, R.H. (1983). Isolation and physical mapping of the gene encoding the major σ factor of *Bacillus subtilis* RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 4074–4078.

PRIMROSE, S.B. AND EHRLICH, S.D. (1981). Isolation of plasmid deletion and study of their instability. *Plasmid* **6**, 193–201.

RIVA, S., POLSINELLI, M. AND FALASCHI, A. (1968). A new phage of *B. subtilis* with infectious DNA having separable strands. *Journal of Molecular Biology* **35**, 347–356.

ROSENBERG, M. AND COURT, D. (1979). Regulatory sequences involved in the promotion and termination of RNA transcription. *Annual Review of Genetics* **13**, 319–353.

SEGALL, J., TJIAN, R., PERO, J. AND LOSICK, R. (1974). Chloramphenicol restores sigma factor activity to sporulating *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America* **71**, 4860–4863.

SHIMOTSU, H., KAWAMURA, F., KOBAYASHI, Y. AND SAITO, H. (1983). Early sporulation gene *spoOF* — nucleotide sequence and analysis of gene product. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 658–662.

SHINE, J. AND DALGARNO, L. (1974). The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proceedings of the National Academy of Sciences of the United States of America* **71**, 1342–1346.

SHORENSTEIN, R.G. AND LOSICK, R. (1973). Comparative size and properties of the sigma subunits of RNA polymerase from *B. subtilis* and *E. coli*. *Journal of Biological Chemistry* **248**, 6170–6173.

SIEBENLIST, U. (1979). RNA polymerase unwinds an 11-base pair segment of a phage T7

promoter. *Nature* **279**, 651–652.

Stallcup, M.R. and Rabinowitz, J.C. (1973). Initiation of protein synthesis *in vitro* by a Clostridial system. I. Specificity in the translation of natural messenger ribonucleic acids. *Journal of Biological Chemistry* **248**, 3209–3215.

Steuber, D. and Bujard, H. (1982). Transcription from efficient promoters can interfere with plasmid replication and diminish expression of plasmid specified genes. *EMBO Journal* **1**, 1399–1404.

Stewart, G.C. and Bott, K.F. (1983). DNA sequence of the tandem ribosomal RNA promoter for *B. subtilis* operon *rrnB*. *Nucleic Acids Research* **11**, 6289–6300.

Takkinen, K., Pettersson, R.F., Kalkkinen, N., Palva, K., Soderlund, H. and Kaariainen, L. (1983). Amino acid sequence of α-amylase from *Bacillus amyloliquefaciens* deduced from the nucleotide sequence of the cloned gene. *Journal of Biological Chemistry* **258**, 1007–1013.

Tanaka, T., Kuroda, M. and Sakaguchi, K. (1977). Isolation and characterization of four plasmids from *Bacillus subtilis*. *Journal of Bacteriology* **129**, 1487–1494.

Timmis, K., Cabello, and Cohen, S.N. (1975). Cloning, isolation, and characterization of replication region of complex plasmid genomes. *Proceedings of the National Academy of Sciences of the United States of America* **72**, 2242–2246.

Tinoco, E., Jr, Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973). Improved estimation of secondary structure in ribonucleic acids. *Nature New Biology* **246**, 40–41.

Tjian, R., Stinchcomb, D. and Losick, R. (1974). Antibody directed against *Bacillus subtilis* σ factor purified by sodium dodecyl-sulfate slab gel electrophoresis. Effect of transcription by RNA polymerase in crude extracts of vegetative and sporulating cells. *Journal of Biological Chemistry* **250**, 8824–8828.

Twigg, A.J. and Sherratt, D. (1980). Trans-complementable copy-number mutants of plasmid ColE1. *Nature* **283**, 216–218.

Wang, P.-Z. and Doi, R.H. (1984). Overlapping promoters transcribed by *Bacillus subtilis* sigma-55 and sigma-37 RNA polymerase holoenzymes during growth and stationary phases. *Journal of Biological Chemistry* **259**, 8619–8625.

Wawrousek, E.F. and Hansen, J.N. (1983). Structure and organization of a cluster of six tRNA genes in the space between tandem ribosomal RNA gene sets in *Bacillus subtilis*. *Journal of Biological Chemistry* **258**, 291–298.

Wells, J.A., Ferrari, E., Henner, D.J., Estell, D.A. and Chen, E.Y. (1983). Cloning, sequencing, and secretion of *Bacillus amyloliquefaciens* subtilisin in *Bacillus subtilis*. *Nucleic Acids Research* **11**, 7911–7925.

Wiggs, J.L., Gilman, M.Z. and Chamberlin, M.J. (1981). Heterogeneity of RNA polymerase in *Bacillus subtilis*: evidence for an additional σ factor in vegetative cells. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 2762–2766.

Williams, D.M., Duvall, E.J. and Lovett, P.S. (1981). Cloning restriction fragments that promote expression of a gene in *Bacillus subtilis*. *Journal of Bacteriology* **146**, 1162–1165.

Wong, H.C., Schnepf, H.E. and Whiteley, H.R. (1983). Transcriptional and translational start sites for the *Bacillus thuringiensis* crystal protein gene. *Journal of Biological Chemistry* **258**, 1960–1967.

Wong, S.-L., Price, C.W., Goldfarb, D.S. and Doi, R.H. (1984). The subtilisin E gene of *Bacillus subtilis* is transcribed from σ^{37}promoter *in vivo*. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 1184–1188.

Yamazaki, H., Ohmura, K., Nakayama, A., Takeichi, Y., Otozai, K., Yamasaki, M., Tamura, G. and Yamane, K. (1983). α-amylase genes (*amyR2* and *amyE*$^+$) from an α-amylase-hyperproducing *Bacillus subtilis* strain: molecular cloning and nucleotide sequences. *Journal of Bacteriology* **156**, 327–337.

Yang, M., Galizzi, A. and Henner, D. (1983). Nucleotide sequence of the amylase gene from *Bacillus subtilis*. *Nucleic Acids Research* **11**, 237–249.

YOSHIKAWA, H. AND ITO, J. (1982). Nucleotide sequence of the major early region of bacteriophage ϕ29. *Gene* **17**, 323–335.

ZUBER, P. AND LOSICK, R. (1983). Use of a *LacZ* fusion to study the role of the *spoO* genes of *Bacillus subtilis* in developmental regulation. *Cell* **35**, 275–283.

ZUKOWSKI, M.M., GAFFNEY, D.F., SPECK, D., KAUFFMANN, M., FINDELI, A., WISECUP, A. AND LECOCQ, J.P. (1983). Chromogenic identification of genetic regulatory signals in *Bacillus subtilis* based on expression of a cloned *Pseudomonas* gene. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 1101–1105.