Construction and Application of Epitope- and Green Fluorescent Protein-Tagging Integration Vectors for *Bacillus subtilis*

Marcus Kaltwasser, Thomas Wiegert, and Wolfgang Schumann*

Institute of Genetics, University of Bayreuth, D-95445 Bayreuth, Germany

Here we describe the construction and application of six new tagging vectors allowing the fusion of two different types of tagging sequences, epitope and localization tags, to any *Bacillus subtilis* protein. These vectors are based on the backbone of pMUTIN2 and replace the *lacZ* gene with tagging sequences. Fusion of the tagging sequences occurs by PCR amplification of the 3′ terminal part of the gene of interest (about 300 bp), insertion into the tagging vector in such a way that a fusion protein will be synthesized upon integration of the whole vector via homologous recombination with the chromosomal gene. Three of these tagging sequences (FLAG, hemagglutinin, and c-Myc) allow the covalent addition of a short epitope tag and thereby detection of the fusion proteins in immunoblots, while three other tags (green fluorescent protein⁺, yellow fluorescent protein, and cyan fluorescent protein) are helpful in assigning proteins within one of the compartments of the cell. The versatility of these vectors was demonstrated by fusing these tags to the cytoplasmically located HtpG and the inner membrane protein FtsH.

Sequencing the *Bacillus subtilis* genome has revealed about 4,100 genes, the function of which approximately 60% has been identified either experimentally or by computer-based analysis (9). As a first step to elucidate the function of the remaining ~1,600 genes, a network of 18 European and 12 Japanese laboratories has systematically inactivated most of these genes of unknown function (17). To obtain this goal, the integration vector pMUTIN was constructed (18). This vector is unable to replicate in *B. subtilis*, but upon insertion of about 300 bp derived from the coding region of the gene to be inactivated by PCR, it can integrate into that gene by homologous recombination. Integration of the recombinant pMUTIN vector into the target gene transcriptionally fuses *lacZ* to that promoter of the gene, and downstream genes can be controlled by the isopropyl-β-D-thiogalactopyranoside (IPTG)-dependent Pspac promoter (18).

A further characterization of unknown gene products at the protein level requires their detection by antibodies. Production of antibodies requires purification of the protein followed by immunization of an animal, normally a rabbit. This procedure is time consuming and expensive, and the antibodies obtained often vary in their quality. To circumvent these problems, the method of choice is the use of epitope-tagging vectors and/or green fluorescent protein (GFP) fusions, both of which are important tools in eukaryotic systems (1, 12). While epitope-tagging vectors have never been described for *B. subtilis*, GFP fusion vectors are available, but they do not allow fusion of the localization tag to chromosomally located genes (4, 10, 11). We made use of the pMUTIN vector and modified it in such a way to allow a translational fusion of the three epitope tags FLAG, hemagglutinin (HA), and c-Myc and the localization tags GFP, cyan fluorescent protein (CFP), and yellow fluorescent protein

(YFP) to the 3′ end of any gene within the *B. subtilis* chromosome and of any other bacterial species not allowing replication of pMUTIN. While the FLAG tag is an artificial 8-amino-acid residue-long peptide (7), c-Myc (10 amino acid residues) and HA (9 amino acid residues) were derived from the human c-*myc* proto-oncogene and the HA of the influenza virus, respectively (3, 20). Antibodies specifically recognizing these tags are commercially available. GFP and its two variants are highly useful fluorescent tags for studying the localization and dynamics of proteins in living cells.

Construction of six tagging integration vectors. We started from the integration vector pDE01, a precursor of the pMUTIN2 derivative that carries *bgaB* instead of the *lacZ* reporter gene (E. Deuerling, unpublished work); this gene codes for a heat-stable β-galactosidase (5). First, the *Sal*I-*Sph*I fragment carrying the *bgaB* gene was replaced with a polylinker with several unique restriction enzyme sites (Table 1), resulting in the plasmid pMUTIN-Poly and thereby destroying the *Sal*I site. To ensure efficient termination of transcription immediately downstream of the hybrid gene, the *trpA* terminator of the *Escherichia coli* tryptophan operon (8), assembled from two complementary oligonucleotides (Table 1), was inserted into the *Spe*I-*Sph*I sites of pMUTIN-Poly in such a way that the *Sph*I site was destroyed, resulting in pMUTIN-Ter. This vector served as a backbone for the insertion of the six tagging sequences.

The DNA sequences coding for the three different epitopes (Table 1) were ligated as complementary oligonucleotides into the *Xma*III and *Spe*I sites, resulting in pMUTIN-FLAG, -cMyc, and -HA (Fig. 1A); by this manipulation the *Spe*I site was destroyed in all three vectors. The codons for the different amino acids were designed on the basis of the codon usage table published for *B. subtilis* (http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?=Bacillus+subtilis+[gbbct]), and the open reading frames were terminated with two consecutive stop codons to ensure efficient termination of translation (Table 1). The correct DNA sequences of all three epitope tags

* Corresponding author. Mailing address: Institute of Genetics, University of Bayreuth, D-95445 Bayreuth, Germany. Phone: 49 (921) 552708. Fax: 49 (921) 552710. E-mail: wolfgang.schumann@uni-bayreuth.de.

TABLE 1. Complementary oligonucleotides used to create short insertions into the integration vector

Description	DNA sequence and restriction enzyme sites

Polylinker

```
    SalI*   KpnI    Eco47-3   ClaI    XmaIII              SpeI            SphI
    TCGAG GGTACC AGCGCT    ATCGAT  CGGCCG TTATCG ACTAGT TTATCG GCATG
        C CCATGG TCGCGA    TAGCTA  GCCGGC AATAGC TGATCA AATAGC C
```

Terminator

```
    SpeI                                          SphI*ᵃ
    CTAGT GCAGCCCGCCTAATGAGCGGGCTTTTTT CCATG
        A CGTCGGGCGGATTACTCGCCCGAAAAAA G
```

FLAG

```
    XmaIII                                                          SpeI*
    GGCCG  GAT TAT AAA GAT GAT GAT GAT AAA TAA  TAA T
        C  CTA ATA TTT CTA CTC ATA CTA TTT ATT  ATT AGATC
           Asp Tyr Lys Asp Asp Asp Asp Lys Stop Stop
```

cMyc

```
    XmaIII                                                          SpeI*
    GGCCG GAA CAA AAA CTT ATT AGC GAA GAA GAT CTT TAA  TAA T
        C CTT GTT TTT GAA TAA TCG CTT CTT CTA GAA ATT  ATT AGATC
          Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu Stop Stop
```

HA

```
    XmaIII                                                          SpeI*
    GGCCG TAT CCG TAT GAT GTT CCG GAT TAT GCA TAA  TAA T
        C ATA GGC ATA CTA CAA GGC CTA ATC CGT ATT  ATT AGATC
          Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Stop Stop
```

ᵃ These sites have been destroyed upon insertion of the DNA fragments.
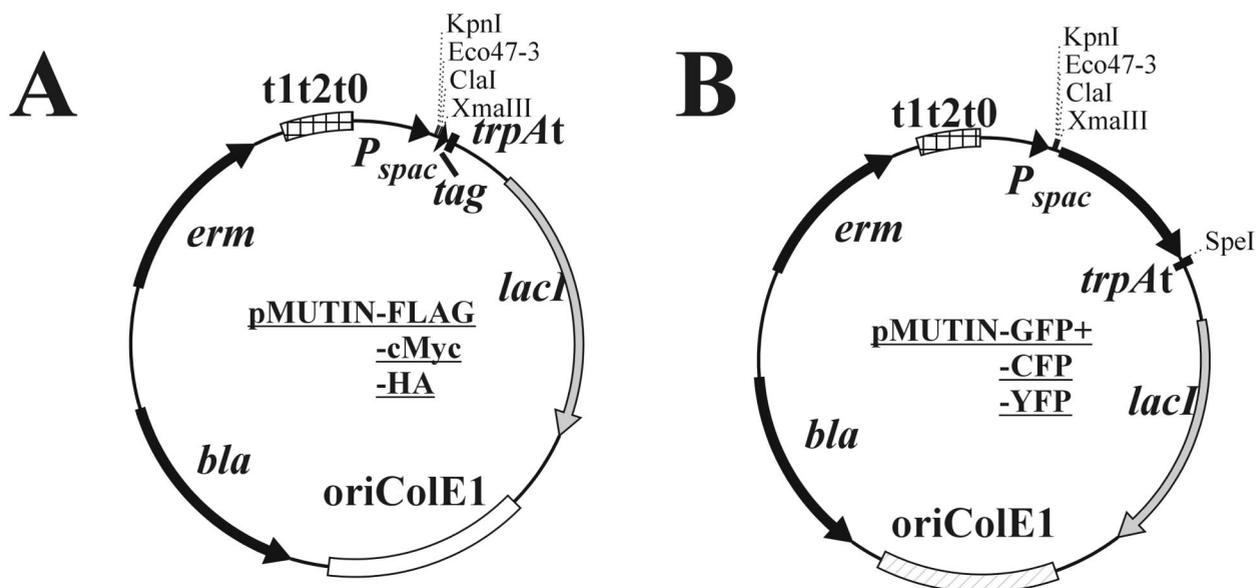

FIG. 1. Tagging integration vectors. (A) Epitope-tagging vectors carrying the coding sequences for the three epitopes FLAG, c-Myc, and HA. (B) Localization-tagging vectors containing the coding sequences for *gfp*[+], *yfp*, and *cfp*. All vectors are derivatives of pBR322 and do not replicate in *B. subtilis*.

TABLE 2. Oligonucleotides used in this study

Primer	Sequence (5′ to 3′)[a]	Description; position
ON1	**GGCCAT** <u>CGGCCG</u> GAAGAAGATATACATATGGCT	*Xma*III; 5′ end of *gfp*[+]
ON2	**GGCCAT** <u>ACTAGT</u> GCTCGAATTCATTATTTGTAG	*Spe*I; 3′ end of *gfp*[+]
ON3	**GGCCAT** <u>CGGCCG</u> ATGGTGAGCAAGGGCGAG	*Xma*III; 5′ end of *cfp/yfp*
ON4	**GGCCAT** <u>ACTAGT</u> TTACTTGTACAGCTCGTC	*Spe*I; 3′ end of *gfp/yfp*
ON5	**GGCCAT** <u>GGTACC</u> GTCAGAATCTCAAAACGGCTT	*Kpn*I; 5′ end of *htpG*
ON6	**GGCCAT** <u>CGGCCG</u> CACCATGACCTTGCAAATATT	*Xma*III; 3′ end of *htpG*
ON7	**GGCCAT** <u>GGTACC</u> GATTTCAACAACGAACAGAAC	*Kpn*I; 5′ end of *ftsH*
ON8	**GGCCAT** <u>CGGCCG</u> CTCTTTCGTATCGTCTTTCTT	*Xma*III; 3′ end of *ftsH*

[a] G/C clamps are shown in boldface; restriction sites are underlined.

were confirmed by DNA sequencing. The coding regions for the GFP and its two variants were generated by PCR and flanked with *Xma*III and *Spe*I sites (Table 2). As templates, we used the plasmids pMN402 (coding for GFP[+], a variant of the wild-type *gfp* exhibiting increased fluorescence [14] [using oligonucleotides ON1 and ON2] [Table 2]) and pSG1186 and pSG1187 (ON3 and ON4, coding for *cfp* and *yfp*, respectively) (4), resulting in the three plasmids pMUTIN-GFP[+], -CFP, and -YFP (Fig. 1B). All six tagging vectors retain the unique restriction sites *Kpn*I, *Eco*47-3, *Cla*I, and *Xma*III for insertion of PCR fragments.

Construction of recombinant tagging vectors. To prove that these vectors function appropriately, we decided to fuse the tagging sequences to two different genes, one coding for the cytoplasmic heat shock protein HtpG and the other for the ATP- and Zn[2+]-dependent protease FtsH, which is anchored to the cytoplasmic membrane (2, 16). To accomplish this goal, about 300 bp from the 3′ end of both genes excluding their stop codon was amplified by PCR using the oligonucleotides ON5 and ON6 in the case of *htpG* and ON7 and ON8 for *ftsH* (Table 2). Both PCR products were cleaved with *Kpn*I and *Xma*III and ligated into the six tagging vectors treated with the same enzymes. In the last step, the 12 recombinant plasmids were used to transform competent *B. subtilis* strain 1012 cells (13). Transformants were selected on Luria-Bertani plates containing erythromycin and were further analyzed by PCR for inte-
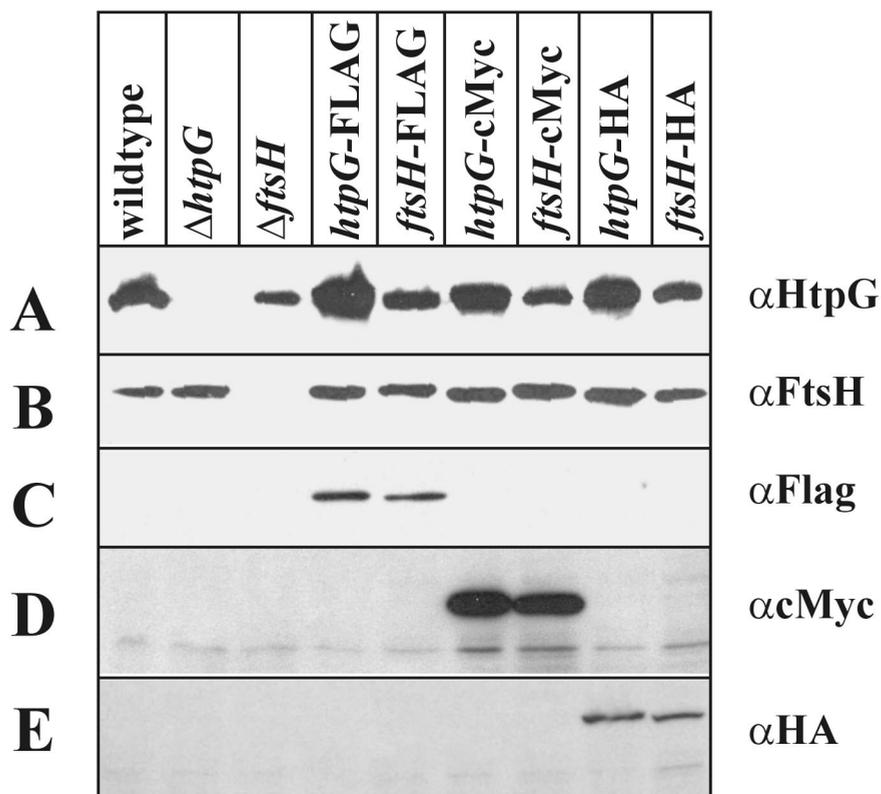

FIG. 2. Detection of the two proteins HtpG and FtsH by immunoblot analysis. *B. subtilis* strain 1012 and its isogenic derivatives were grown at 37°C to mid-exponential growth phase, and sample preparation for sodium dodecyl sulfate-polyacrylamide gel electrophoresis and for immunoblot analysis was performed as described previously (6). An equal amount of protein was applied per lane. The blots were probed with αHtpG (A), αFtsH (B), αFLAG (C), αc-Myc (D), and αHA (E). The bands representing the tagged proteins are marked by an arrow. Antibodies against HtpG and FtsH have been raised against purified His-tagged proteins (2, 16). Antibodies raised against the epitope tags were purchased from Sigma (αFLAG and αc-Myc, clone 9E10) and Roche Diagnostics (αHA, clone 12CA5).
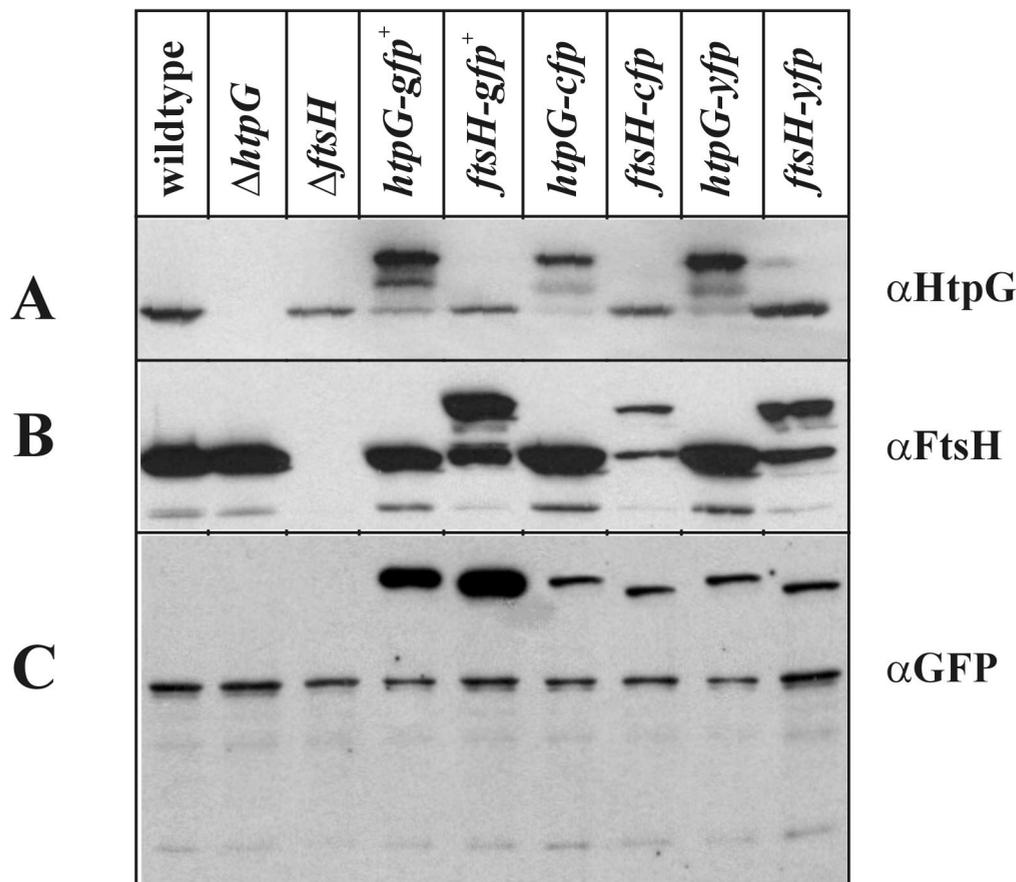
FIG. 3. Detection of HtpG and FtsH fusion proteins by immunoblot analysis. Preparation of the immunoblots followed the procedure described in the Fig. 2 legend. The blots were probed with αHtpG, αFtsH, and αGFP. The bands representing the tagged proteins are marked by an arrow. αGFP was obtained from Clontech.

gration of one copy each of the plasmid at the correct locus (data not shown); one strain each was used in the following experiments.

Fusion proteins carrying the three epitope tags can be detected using specific antibodies. In the next step, we tested for the production of fusion proteins carrying the different epitope tags. Cells carrying *htpG* or *ftsH* fused to either the FLAG, c-Myc, or HA epitope were analyzed for the presence of the appropriate fusions. As can be seen in Fig. 2A, αHtpG cross-reacted with a protein of about 72 kDa present in all strains tested with the exception of the *htpG* knockout. The same result was obtained for FtsH, which could be detected in all strains but the *ftsH* knockout (Fig. 2B). When antibodies raised against the three epitopes were used to probe the extracts, they specifically detected the two fusion proteins containing FLAG (Fig. 2C), c-Myc (Fig. 2D), and HA (Fig. 2E). We conclude from these results that all three integration vectors function as expected and can be used to add one of the three epitopes to either a cytoplasmic or an integral membrane protein.

Fusion proteins carrying the three GFP variants can be demonstrated using anti-GFP antibodies. Next, we wanted to detect the six fusion proteins carrying the localization tags at their C termini by immunoblotting. Again, the proteins of the different cell extracts were separated by electrophoresis, blotted, and probed with three different antibodies. When the blot

was treated with anti-HtpG (αHtpG), cross-reacting material was detected in all cell lysates, with the exception of the Δ*htpG* strain (Fig. 3A). Cells from strains expressing a tagged HtpG protein exhibited a band which by its molecular mass represented the fusion protein. Two additional bands can be visualized, one with the molecular mass of HtpG and the second migrating between the two others (Fig. 3A), both most probably representing degradation products. A similar observation was made when the blot was probed with αFtsH. Cross-reactive material was present in all extracts but those from the Δ*ftsH* strain (Fig. 3B). Again, strains producing the tagged FtsH exhibited two additional bands, one major with the molecular mass of FtsH and one minor migrating between the two others (Fig. 3B). These data suggest that the fusion proteins are a target for one or more proteases within the *B. subtilis* cells. Furthermore, the antibodies reacted with a protein of about 50 kDa, not present in the *ftsH* knockout and therefore representing a degradation product of FtsH. When αGFP antibodies were used, they detected the fusion protein only and another protein present in all extracts (Fig. 3C).

Since the tagged proteins could be directly visualized using fluorescence microscopy only in the case of the GFP⁺ versions (only filter sets for GFP were available; data not shown) and since descriptions of cells expressing FtsH-GFP have already been published (19), we decided to measure the fluorescence

TABLE 3. Fluorescence measurements in cells synthesizing HtpG labeled with three different GFP tags

Fusion protein	Excitation at (nm):	Emission at (nm):	Relative fluorescence after heat shock	
			0 min	30 min
HtpG-CFP	433	475	0	5
HtpG-YFP	513	527	0	8
HtpG-GFP$^+$	488	510	10	25

[a] The strains expressing the three different fusion proteins were grown in Luria-Bertani medium at 37°C to mid-exponential growth phase and were then shifted to 48°C to induce a heat shock. One-milliliter samples were withdrawn immediately before and 30 min after challenge of the cells with heat stress. The cells were pelleted by centrifugation, resuspended in phosphate-buffered saline (0.1 M sodium phosphate, 1 mM MgCl$_2$, pH 7.0), pelleted again, resuspended in the same buffer, and kept on ice. Then 0.1 V each of optical density at 578 nm was withdrawn and pipetted into cuvettes prewarmed at 25°C, and fluorescence measurements were carried out on an FS-4010 spectrophotometer (Hitachi). The wavelengths of fluorescence excitation and emission for the three are given.

of all three versions using a fluorometer. We chose the three strains producing HtpG fusion proteins, allowing the induction of its gene by heat shock (15). It turned out that only HtpG-GFP$^+$ could be detected in the absence of heat stress (Table 3), while all three fusion proteins were clearly present after a heat shock. This experiment independently confirms that all three proteins are indeed produced.

Conclusions. (i) Six tagging integration vectors were constructed allowing translational fusions of two types of tagging sequences, epitope, and localization, to the 3′ end of any chromosomal gene of interest. If this gene is part of a polycistronic operon and if one of the downstream genes is essential, its transcription can be ensured by an IPTG-regulable promoter. These vector plasmids were primarily constructed for use in *B. subtilis* but can be applied to any bacterial species not allowing replication of the pBR322-based plasmids. But expression of downstream genes depends on synthesis of the LacI repressor protein and proper functioning of the Pspac promoter. (ii) The epitope sequences (FLAG, c-Myc, and HA) allow detection of tagged proteins within the cell using commercially available antibodies. Furthermore, the FLAG and HA tags can be used to purify the fusion proteins by affinity chromatography. (iii) The localization tags GFP$^+$, which produces enhanced fluorescence; YFP; and CFP can be used to localize a protein to a specific compartment within the cell. (iv) To prove that all six tagging vectors are functional, they were fused to two different genes, one coding for the cytoplasmic protein HtpG and the other for the integral membrane protein FtsH. All 12 fusion proteins could be detected using specific antibodies. (v) The six plasmids can be ordered from the *Bacillus* Genetic Stock Center (http://bacillus.biosci.ohio-state.edu). The DNA sequences of the plasmids and their plasmid maps can be found at http://btbgn1.bio.uni-bayreuth.de/lsgenetik1/frames.htm.

REFERENCES

1. **Chubet, R. G., and B. L. Brizzard.** 1996. Vectors for expression and secretion of FLAG epitope-tagged proteins in mammalian cells. BioTechniques **20:**136–141.
2. **Deuerling, E., B. Paeslack, and W. Schumann.** 1995. The *ftsH* gene of *Bacillus subtilis* is transiently induced after osmotic and temperature up-shock. J. Bacteriol. **177:**4105–4112.
3. **Evan, G. I., G. K. Lewis, G. Ramsay, and J. M. Bishop.** 1985. Isolation of monoclonal antibodies specific for human c-*myc* proto-oncogene product. Mol. Cell. Biol. **5:**3610–3616.
4. **Feucht, A., and P. J. Lewis.** 2001. Improved plasmid vectors for the production of multiple fluorescent protein fusions in *Bacillus subtilis*. Gene **264:**289–297.
5. **Hirata, H., S. Negoro, and H. Okada.** 1985. High production of thermostable β-galactosidase of *Bacillus stearothermophilus* in *Bacillus subtilis*. Appl. Environ. Microbiol. **49:**1547–1549.
6. **Homuth, G., M. Heinemann, U. Zuber, and W. Schumann.** 1996. The genes *lepA* and *hemN* form a bicistronic operon in *Bacillus subtilis*. Microbiology **142:**1641–1649.
7. **Hopp, T. P., K. S. Prickett, V. L. Price, R. T. Libby, C. J. March, P. Cerretti, D. L. Urdal, and P. J. Conlon.** 1988. A short polypeptide marker sequence useful for recombinant protein identification and purification. Bio/Technology **6:**1204–1210.
8. **Keiler, K. C., P. R. H. Waller, and R. T. Sauer.** 1996. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. Science **271:**990–993.
9. **Kunst, F., N. Ogasawara, I. Moszer, et al.** 1997. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. Nature **390:**249–256.
10. **Lewis, P. J., and J. Errington.** 1996. Use of green fluorescent protein for detection of cell-specific gene expression and subcellular protein localization during sporulation in *Bacillus subtilis*. Microbiology **142:**733–740.
11. **Lewis, P. J., and A. L. Marston.** 1999. GFP vectors for controlled expression and dual labelling of protein fusions in *Bacillus subtilis*. Gene **227:**101–109.
12. **Margolin, W.** 2000. Green fluorescent protein as a reporter for macromolecular localization in bacterial cells. Methods **20:**62–72.
13. **Saito, H., T. Shibata, and T. Ando.** 1979. Mapping of genes determining nonpermissiveness and host-specific restriction to bacteriophages in *Bacillus subtilis* Marburg. Mol. Gen. Genet. **170:**117–122.
14. **Scholz, Q., A. Thiel, W. Hillen, and M. Niederweis.** 2000. Quantitative analysis of gene expression with an improved green fluorescent protein. Eur. J. Biochem. **267:**1565–1570.
15. **Schulz, A., S. Schwab, S. Versteeg, and W. Schumann.** 1997. The *htpG* gene of *Bacillus subtilis* belongs to class III heat shock genes and is under negative control. J. Bacteriol. **10:**3103–3109.
16. **Schulz, A., B. Tzschaschel, and W. Schumann.** 1995. Isolation and analysis of mutants of the *dnaK* operon of *Bacillus subtilis*. Mol. Microbiol. **15:**421–429.
17. **Schumann, W., S. D. Ehrlich, and N. Ogasawara.** 2001. Functional analysis of bacterial genes. John Wiley & Sons, Ltd., Chichester, United Kingdom.
18. **Vagner, V., E. Dervyn, and S. D. Ehrlich.** 1998. A vector for systematic gene inactivation in *Bacillus subtilis*. Microbiology **144:**3097–3104.
19. **Wehrl, W., M. Niederweis, and W. Schumann.** 2000. The FtsH protein accumulates at the septum of *Bacillus subtilis* during cell division and sporulation. J. Bacteriol. **182:**3870–3873.
20. **Wilson, I. A., H. L. Niman, R. A. Houghten, A. R. Cherenson, M. L. Connolly, and R. A. Lerner.** 1984. The structure of an antigenic determinant in a protein. Cell **37:**767–778.