Genes concerned with synthesis of poly(glycerol phosphate), the essential teichoic acid in *Bacillus subtilis* strain 168, are organized in two divergent transcription units

CATHERINE MAUËL,¹ MICHAEL YOUNG^{1,2} and DIMITRI KARAMATA^{1*}

¹ Institut de Génétique et Biologie Microbiennes, Rue César-Roux 19, CH-1005 Lausanne, Switzerland

² Department of Biological Sciences, University College of Wales, Aberystwyth, Dyfed SY23 3DA, UK

(Received 25 September 1990; revised 19 November 1990; accepted 11 December 1990)

Insertional mutagenesis has revealed that a 22 kbp segment from the *hisA* region of the *Bacillus subtilis* 168 chromosome (310° on the genetic map) contains at least six independent transcription units, all apparently devoted to production of cell envelope components. Genes concerned with synthesis of poly(glycerol phosphate), poly(groP), an essential cell wall polymer in *B. subtilis* 168, are organized in two divergently transcribed operons denoted *tagABC* and *tagDEF*. Nucleotide sequence analysis indicates that three of these six genes encode extremely basic polypeptides. The deduced products of the *tagABC* operon may be involved in poly(groP) assembly and export, whereas those of the *tagDEF* operon, which are very hydrophilic, are more likely to be implicated in poly(groP) precursor biosynthesis. The first gene of the *tagDEF* operon encodes glycerol-3-phosphate cytidyltransferase (Pooley *et al.*, 1991, *Journal of General Microbiology* 137, 921–928) and its deduced product has significant homology with cholinephosphate cytidyltransferase from yeast. There is also substantial homology between the deduced products of *tagB* in the *tagABC* operon and *tagF* in the *tagDEF* operon.

Introduction

Synthesis of poly(glycerol phosphate), poly(groP), the major cell wall teichoic acid of *Bacillus subtilis* strain 168 (Baddiley, 1970) is normally essential for growth (Mauël *et al.*, 1989). All known mutations affecting genes specifically involved in teichoic acid biosynthesis (*tag* genes) result in a thermosensitive growth phenotype (Boylan *et al.*, 1972; Karamata *et al.*, 1972; Pooley *et al.*, 1987; Mauël *et al.*, 1989; Briehl *et al.*, 1989). They are clustered, together with the *gtaA*, *gtaB* and *gtaD* genes concerned with poly(groP) glucosylation (Young, 1967; Pooley *et al.*, 1987), at about 310° on the genetic map of *B. subtilis* (Piggot, 1989). Located nearby are other genes involved in cell envelope functions including structural genes encoding autolysins (Ph. Margot & D. Karamata, unpublished results) and enzymes concerned with bio-

synthesis of poly(glucose *N*-acetylgalactosamine 1-phosphate). Mutants impaired in biosynthesis of this secondary anionic polymer found in cell walls of strain 168 (Shibaev *et al.*, 1973) have been isolated on the basis of their resistance to phage ϕ 3T (Estrela *et al.*, 1991). A 60 kbp DNA segment encompassing all these genes has recently been cloned (Mauël *et al.*, 1989; Young *et al.*, 1989).

Physiological and biochemical investigations have indicated that synthesis of poly(groP) and its incorporation into the bacterial cell wall probably requires eight or nine enzymes (Ward, 1981). It is also apparent from these studies that expression of the genes concerned with anionic polymer biosynthesis is likely to be subject to a complex regulation. There is evidence of an interdependence between teichoic acid and peptidoglycan synthesis (Ward, 1981); moreover, under conditions of phosphate exhaustion, poly(groP) synthesis ceases and that of teichuronic acid, a phosphate-free acidic cell wall polymer, is initiated (Ellwood & Tempest, 1972).

To facilitate future characterization of this regulation, which could in turn shed light on the role of anionic polymers in cell envelope metabolism, we have embarked upon a detailed analysis of this region of the

Abbreviations: orf, open reading frame; poly(groP), poly(glycerol phosphate).

The nucleotide sequence data reported in this paper have been submitted to GenBank and have been assigned the accession number M57497.

chromosome. Initially, it was important to determine the number and organization of the *tag* genes cloned. We report here the identification and analysis of transcription units in a 22 kbp region and the nucleotide sequence of a 4263 bp DNA segment encompassing the *tagABC* operon, an intergenic region in which divergent transcription is initiated and a gene which, together with two recently sequenced genes, *tagE* and *tagF*, previously denoted *rodD* and *rodC*, respectively (Honeyman & Stewart, 1989), forms the *tagDEF* operon.

Methods

Bacterial strains, plasmids and phages. Bacterial strains are listed in Table 1. Fragments from the chromosomal DNA inserts in phages λ 53 and λ E51 (Mauël *et al.*, 1989; Young *et al.*, 1989) were subcloned in either pJH101 (Ap^R Tc^R Cm^R) (Ferrari *et al.*, 1983) or pMTL20EC (Ap^R Em^R Cm^R) (Chambers *et al.*, 1988) and maintained in *E. coli* strain DH5, except for the insert in plasmid 5307 (Mauël *et al.*, 1989), which was cloned in pMTL500C (Ap^R Cm^R) (Oultram *et al.*, 1988) and maintained in *B. subtilis*. The bacteriophages employed for genotype identification were laboratory stocks of ϕ 25, ϕ 29, ϕ 3T and defective bacteriophages PBSY and PBSZ harboured by *B. subtilis* strains S31 and W23, respectively.

Media and maintenance of strains. *E. coli* strain DH5 was grown and maintained on LB medium. When necessary, ampicillin and tetracycline were added at 50 μ g ml⁻¹ and 15 μ g ml⁻¹ final concentration, respectively. Strains of *B. subtilis* were grown and maintained on LB medium and tryptose blood agar base (TBAB). SA medium and media used for transformation were as previously described (Karamata & Gross, 1970). Chloramphenicol was added to media, as appropriate, at a final concentration of 3 μ g ml⁻¹.

DNA extraction and manipulation. DNA was extracted from λ phages according to Grossberger (1987). DNA fragments for subcloning were recovered from agarose gels using GeneClean supplied by Lucerna-Chem (Switzerland). Mini-preparations of plasmids from *E. coli* and *B. subtilis* were obtained by the alkaline lysis method (Birnboim & Doly, 1979) with an additional LiCl precipitation step to remove RNA (Birnboim, 1983). Restriction endonucleases, T4 DNA ligase and calf

intestinal phosphatase were obtained from Biofinex (Switzerland) and used according to the supplier's instructions.

Transformation and selection of transformants. Plasmids were introduced into *E. coli* DH5 by the transformation method of Chung & Miller (1988). Competent cells of *B. subtilis* were prepared as described by Karamata & Gross (1970) and stored frozen (Mauël *et al.*, 1989). Saturating concentrations of plasmid DNA (1–10 μ g DNA ml⁻¹) were used for transformations. Chloramphenicol-resistant transformants were selected on LB plates containing 3 μ g chloramphenicol ml⁻¹ and Tag⁺ recombinants were selected at 47 °C on TBAB plates (Briehl *et al.*, 1989).

Determination of phage susceptibility of chloramphenicol-resistant transformants of *B. subtilis*. Phage stocks (ϕ 25, ϕ 29, PBSY and PBSZ) were spotted onto fresh streaks of purified chloramphenicol-resistant transformants. Plates were incubated for 2–3 h at 37 °C and then at room temperature. ϕ 3T susceptibility was tested by spotting the phage stock on a soft agar overlay containing the strain under investigation together with appropriate divalent cations (10 mM-Ca²⁺, 10 mM-Mg²⁺, 1 mM-Mn²⁺). Growth inhibition and lysis, visible after 2–8 h, revealed sensitivity to the phage, whereas resistance was indicated by the absence of a zone of clearing.

DNA sequencing. Plasmid DNA was obtained from minipreparations further purified by phenol extraction. Sets of deleted derivatives, prepared from several subclones in the pUC-related vector pMTL20EC (Chambers *et al.*, 1988), were obtained using a double-stranded nested deletion kit from Pharmacia. DNA sequencing was performed on both strands by the dideoxy-chain-termination method (Sanger *et al.*, 1977) with Sequenase Version 2.0 (USB Corporation) according to the supplier's instructions. 5'-[α -³⁵S]Deoxyadenosine thiotriphosphate (> 1000 Ci mmol⁻¹; > 37 TBq mmol⁻¹) was obtained from Amersham (UK). Primer and reverse primer were supplied by Biofinex (Switzerland). Two 17-mer oligonucleotides were kindly synthesized by Beckman Instrument International (Switzerland).

Nucleotide sequence assembly, verification and analysis was accomplished using the Sequence Analysis Software Package of the Genetic Computer Group, University of Wisconsin (Devereux *et al.*, 1984). The deduced products of open reading frames (orfs) were compared with sequences in the OWL database using the SWEEP and DAP database scanning and optimal alignment programs. The latter (Coulson *et al.*, 1987) provides an estimate of the statistical significance of each alignment (Collins *et al.*, 1988). The alignments shown in Fig. 4 were generated using the 100 PAM similarity scoring table (Dayhoff *et al.*, 1978).

Table 1. *Bacterial strains*

Name	Genotype*	Source
<i>Bacillus subtilis</i>		
L5047	<i>metB5 hisA35 pheA1 purA16 trpC2</i>	Mauël <i>et al.</i> (1989)
L5058	<i>tagB1 [tagA1] (Ts) hisA35 pheA1 purA16 trpC1</i>	Mauël <i>et al.</i> (1989)
L6601	<i>tagD11(Ts) hisA1 argC4 metC3</i>	Briehl <i>et al.</i> (1989)
L6602	<i>tagD12(Ts) hisA1 argC4 metC3</i>	Briehl <i>et al.</i> (1989)
<i>Escherichia coli</i>		
DH5	F ⁻ <i>endA1 hsdR17 (r⁻ m⁺) supE44 thi-1 λ⁻ recA1 gyrA96 relA1 Δ(argF-lacZYA) U169 ϕ80lacZΔM15</i>	Ph. Shaw, University of Lausanne

* The nomenclature of *tag* mutations has been modified according to information obtained from the nucleotide sequence so that the capital letter now refers to an identified orf. The previous designation(s) is indicated in square brackets. Other mutations or loci to which reference is made in the text are: *tagF3* [*tagB3*, *rodC1*] and *tagE* [*rodD*]. Designations of the *gtaA12* and *gtaD1* markers will be maintained until they have been unambiguously assigned to orfs.

Fig. 1. Transcription units in the *tag* region of the *B. subtilis* chromosome. DNA fragments were subcloned in plasmids pMTL20EC (Chambers *et al.*, 1988) or pJH101 (Ferrari *et al.*, 1983) and the resulting integrational plasmids used to transform *B. subtilis* L5047, with selection for chloramphenicol resistance. Insertions resulting in mutant (□) or wild type (▨) phenotypes served to identify transcription units I to V and to delimit the approximate extents of each of them. The transcription unit denoted (VI), which includes the gene encoding UDP-glucose pyrophosphorylase, is positioned with reference to the insert in plasmid 6312, which corrects mutations *gtaB100*, *gtaB101*, *gtaB114* and *gtaB515* previously identified by Pooley *et al.* (1987). DNA segments with transforming activity against *tagB1*, *tagF3*, *gtaA12* and *gtaD1* markers are also indicated. Restriction sites that delimit plasmid inserts are shown. They are abbreviated: B, *Bam*HI; E, *Eco*RI; G, *Bgl*II; H, *Hind*III; P, *Pst*I; Sa, *Sal*I; Sc, *Sac*I; Sp, *Sph*I. The extremities of inserts in EMBL3 or EMBL4 phages are delimited by *Sal*I sites deriving from the vector and denoted (Sa). The positions of the orfs comprising the *tagABC* and *tagDEF* operons are indicated with arrows.

Results

Transcription units in the tag region

Initial global characterization of the transcription units in a 22 kbp region of the *B. subtilis* 168 chromosome, containing all known markers strictly devoted to teichoic acid biosynthesis, was obtained by insertional mutagenesis with integrational plasmids (Piggot *et al.*, 1984; Mauël *et al.*, 1989).

A collection of integrational plasmids carrying different chromosomal fragments covering this region was used to transform strain L5047, a derivative of *B. subtilis* 168, with selection for chloramphenicol resistance. Absence of transformants was attributed to the essentiality of the disrupted transcription units; when present, transformants were tested for susceptibility to phages ϕ 25, ϕ 29, ϕ 3T, PBSY and PBSZ, whose receptors are formed, in part at least, by anionic cell wall polymers (Young, 1967; Karamata *et al.*, 1987; Estrela *et al.*, 1991; Fig. 1).

These experiments showed that the 22 kbp region contains at least five transcription units, denoted I to V in Fig. 1. Previously, it had been established that integrational plasmids containing small DNA segments

lying within either unit I or II are unable to transform *B. subtilis* (Honeyman & Stewart, 1988; Mauël *et al.*, 1989), from which it was concluded that expression of either of these transcription units was essential for growth under the culture conditions employed. Analysis with plasmids containing larger DNA segments indicated that units I and II have minimum sizes of 2.6 and 4.8 kbp, respectively (Fig. 1). All *tag* mutations so far identified are located within these transcription units (Boylan *et al.*, 1972; Karamata *et al.*, 1972; Pooley *et al.*, 1987; Mauël *et al.*, 1989; Briehl *et al.*, 1989). Insertion of plasmid 8106, which defines unit III, seriously impairs, but does not completely abolish growth, for an as yet unknown reason. Disruption of transcription units IV and V did not affect growth of the strains but it did influence their phage susceptibility. Recombinants arising from plasmid insertion in unit IV were characterized by limited clearing in the phage ϕ 29 spot test. Since glucosylated poly(groP) is the putative ϕ 29 receptor, this may suggest the existence of a locus involved in glucosylation of this polymer. However, previous studies on poly(groP) glucosylation have not identified mutations in this region (Young, 1967; Pooley *et al.*, 1987). Plasmid insertion in unit V generated ϕ 3T resistance. This is almost certainly due to disruption of the genes concerned with biosynthe-

Fig. 2. Sequencing strategy. The sequence of the 4263 bp *Bam*HI–*Cla*I segment, encoding the *tagABC* operon and part of the *tagDEF* operon, was determined on both DNA strands using sets of nested deletion derivatives prepared from plasmid subclones (□). Horizontal arrows indicate the extent of DNA sequences determined from individual clones; a filled circle indicates sequences obtained from a synthetic primer, whereas an open circle indicates those which encompass the *Bam*HI site in p5316. The sequence was determined for about 50 bp to the left of this site and was identical to that previously published (Honeyman & Stewart, 1989). Landmark restriction sites, as well as the positions of the deduced orfs, are shown. Restriction sites are denoted: C, *Cla*I; Bc, *Bcl*I; others are as in the legend to Fig. 1.

sis of poly(glucose *N*-acetylgalactosamine 1-phosphate) (Estrela *et al.*, 1991).

The previously identified *gtaB* locus (Young, 1967), denoted (VI) in Fig. 1, is associated with UDP-glucose pyrophosphorylase activity (Pooley *et al.*, 1987). Point mutations at this locus confer resistance to phages ϕ 25 and ϕ 29 and, in some instances, to defective bacteriophages PBSY and PBSZ (Pooley *et al.*, 1987). Since plasmid insertions in the region encompassing unit (VI) did not affect phage susceptibility, it would appear that the DNA fragment(s) containing this locus included either one or both extremities of the transcription unit.

Thus, experiments with integrational plasmids have revealed that the 22 kbp region investigated here contains at least five transcription units in addition to *gtaB*. With the possible exception of unit III, they all play a role in cell envelope biosynthesis.

Nucleotide sequence of transcription units I and II

The above analysis indicated that enzymes involved in poly(groP) synthesis are encoded by two transcription units, I and II, covering at least 2.6 and 4.8 kbp, respectively. To understand the biological role of this polymer, which is normally essential for cell surface extension, we first proceeded to complete the sequencing of this region. Two genes, denoted *rodC* and *rodD*, belonging to transcription unit II, have recently been

sequenced (Honeyman & Stewart, 1989). They lie on a DNA segment extending from just within the insert in plasmid 8105 to the *Bam*HI site of the insert in plasmid 5316 (Figs 1 and 2). The nucleotide sequence of the region contiguous with this *Bam*HI site, and extending through transcription unit I to a *Cla*I site in the plasmid 5105 insert, was determined on both strands. The sequencing strategy is summarized in Fig. 2.

Inspection of the sequence (Fig. 3), comprising 4263 bp, revealed the presence of four open reading frames (orfs). One of them, orfD, hereafter referred to as the *tagD* gene, is transcribed from the same DNA strand as the previously sequenced *rodD* and *rodC* genes (Honeyman & Stewart, 1989). These latter genes have been renamed *tagE* and *tagF*, respectively (see legend of Table 1). The *tagD* gene lies about 160 bp upstream from *tagE*. The three remaining orfs, A, B and C, hereafter referred to as the *tagA*, *tagB* and *tagC* genes, are transcribed in this order from the other DNA strand and appear to constitute a single operon. A 400 bp intergenic region, within which divergent transcription is presumably initiated, separates *tagA* from *tagD*.

Analysis of the open reading frames and their deduced products

The characteristics of the *tagA*, *tagB*, *tagC* and *tagD* genes, and their deduced products, are summarized in Table 2. All

Table 2. Features of the tagA, B, C and D genes and of their deduced products

Property	tagA	tagB	tagC	tagD
Start codon	AUG	AUG	AUG	AUG
Stop codon	UAA	UAA	UGA, UAA	UAA
A/T in wobble position (%)	66.9	68.6	69.8	70.8
CAI*	0.49	0.43	0.48	0.54
Length (amino acids)	256	381	442	129
Molecular mass (Da)	29499	44926	50710	15271
Isoelectric point	10.61	9.88	7.69	6.03
Charged residues† (%)	24	22	23	33
Lysine residues (%)	10.1	9.4	6.7	11.6
Adjacent basic residues‡ (%)	33	37	10	42

* CAI, codon adaptation index (Sharp & Li, 1986); values were calculated according to Shields & Sharp (1987).

† Arg, Lys, Glu, Asp.

‡ Ratio of adjacent basic residues to total basic residues (Arg, Lys).

four genes have AUG start codons and end with the UAA 'ochre' termination codon; tagC actually has two termination codons, UGA followed by UAA. There is a marked bias for A (adenine) or T (thymine) in the third position of codons in all four genes, suggesting that they are highly expressed (Shields & Sharp, 1987). A more sensitive indicator of the expression level of a gene is the codon adaptation index (CAI) of Sharp & Li (1986). CAI values were calculated for the four genes (Table 2), according to the data of Shields & Sharp (1987) for codon usage of highly expressed *B. subtilis* genes. The values obtained, between 0.43 and 0.54, are in the range expected for genes expressed at a moderate level.

The tagA, B and C genes form the tagABC operon. The predicted products of tagA and tagB, comprising 256 and 381 amino acids, respectively, are very basic polypeptides with a high lysine content (Table 2), in which a third or more of the basic residues form clusters of two or more amino acids. The former contains no cysteine residue and has a particularly hydrophilic domain at its COOH-terminus in which 15 of the 60 terminal residues are basic (Fig. 3). The predicted product of tagC is slightly basic and comprises 442 amino acids. Strong ribosome-binding sites are found just upstream from tagA and tagC; ΔG values, calculated according to the rules of Tinoco *et al.* (1973), are -23.2 and -17.4 kcal mol $^{-1}$ (-97.1 and -72.8 kJ mol $^{-1}$), respectively. In contrast, there is only a very weak potential ribosome-binding site with a ΔG value of -8 kcal mol $^{-1}$ (-33.5 kJ mol $^{-1}$) preceding tagB. The tagA and tagB genes are separated by a 32 bp non-coding region. Inspection of the 119 bp interval between tagB and tagC revealed the presence of a very short orf, denoted orfC' in Fig. 3, that could encode a 40 amino acid peptide commencing with two AUG codons immediately after the UAA stop codon

of tagB and ending with a UGA codon having a 2 bp overlap with the AUG start codon of tagC. There is no ribosome-binding site immediately preceding this orf.

The predicted product of tagD is an extremely hydrophilic 129 amino acid polypeptide. In spite of a high lysine content and a substantial proportion of basic residues lying adjacent to one to another (Table 2), it has an acidic character overall (predicted pI = 6.03). A strong ribosome-binding site ($\Delta G = -15.6$ kcal mol $^{-1}$; -65.3 kJ mol $^{-1}$) lies just upstream from the coding sequence.

Hydropathy plots (Kyte & Doolittle, 1982) of the predicted products of tagA, B, C and D (data not shown) did not reveal the presence of reiterated hydrophobic domains that are sometimes found in integral membrane proteins (Capaldi, 1982). However, the 25 amino acid residues at the amino-terminus of the tagB product resemble a signal sequence (Michaelis & Beckwith, 1982; Watson, 1984).

Comparison of the deduced products of the four tag genes with sequences in the OWL database using the Edinburgh DAP database search program revealed significant local homologies between the tagD product and several nucleotidyl transferases, including the cholinephosphate cytidyltransferase of *Saccharomyces cerevisiae* (Tsukagoshi *et al.*, 1987) [SD = 49.2], and also streptokinase from *Streptococcus equisimilis* (Malke *et al.*, 1985) [SD = 14.2]. These are illustrated in Fig. 4, together with the substantial homology that was detected between the predicted products of tagB and tagF (Honeyman & Stewart, 1989) [SD = 219]. The value given in square brackets is the number of standard deviations difference between the score for the alignment and that expected from the distribution of fortuitous alignments (see Collins *et al.*, 1988).

BamHI

CTTAGGATTTAATATTTGGTCGTTAAAGGAGAACTAGAAACAACATCTCTACGGAAGACAAGAACCTCCATTGTTGGAGTGTGACTAGAACTCTTTCAGTTTAAATGGAAGGGTCA 120
 3' * L G A I E E K I K T T S I G E T R P L Y V V E C Q D K L F D F K G E W
 End OrfD

BclI

GTAGGGGATTTGTTTTGTAGTTATAATCTAGTTTATTATAGAACGAAAAACGAGGGTTAAGAAAAGTCCTTAATGAAGCAGGTGTATAGCTTACCAAGATTTTAATTAATGCCA 240
 D D G M V F V D I N H D I I D Q K K Q E W N K E P I V E D V Y R I T E L I L K R

HindIII

EcoRI

CGAGTATCGACACTATTCGAAACAAAAACGAAACATTAACTTAAGTAGGCAACTTTATCGTTGTTGATTATTAGTGGTTCGACGAAACGTCGAGTTCGTTAAATTACACTGGGGTTA 360
 H E Y S H Y A K K Q K Q L N F E D T S I A V V L Y D G L Q K A R E L L K I H G W

5'TTAAAGAAACGCTCCTTCCTAATGTTAAACACTATTAACAACTACTGACGTTAGCGATTAAATTCATAAAACCTGTTT

CGTCATTAGTTTTCAAGGTATACACTATTGAAAAAGTAAATTTCTTTGCGAGGAAGGATTACAATTTGTGATAATTTTGTGATGACTGCAATCGCTAATTTAAGATTTTGGACAAA 480
 H L L D F T G Y T I V K K M
 Start OrfD

BclI

GTATTCCTTATATCGACATATTTCTCGTATGTCTTCACTGATCATTAATTATAGCATACCGCCTAAAAAGGTTAAGTACTCATGAGGAGAAAAACATTCCAAATATGTAATTACACTC 600
 CATAAGGAATATAGCTGTATAAAGAGCATACAGAAGTACTAGTAATTAATATCGTATGGCGGATTTTTCCAATTCATGAGTACTCTCTTTTGTAAAGTTTATACATTAAATGTGAG

AGACTCCATTGGCCACCAAGGTCTGTTGTTTCATTCAATGTAAAATTGGAACATATTTAAATTTAATGTAAAGATAATGTTAATCCTTTGTTGAAGATTTATGTTAAATAT 720
 TCTGAGGTAACCGTGGTCCAGACAACAAGTAAGCAAGTAAGTTAACATTTTAACCTTTGTATAAATTTAAATTACATTTCTATTACAATTAGGAAACAACCTCTAATAACAATTTTATA

Start OrfA

M Q T E T I H N I P Y V N S

AAGGTAGCTTGATTTTGCTTTTAGAACTCTCGGTTTTAACTATATAAATAGAATCTTAAAGGAAGGAGGAGCCAGACATGCAACAGAGACTATTCAATATTCCTTATGTTAACAG 840
 TTCCATCGAACTAAAACGAAATCTTTGAGAGCCAAAATTGATATATTTATCTTAGAATTTCTCTCTCTCGGTCTG 5'

N L T S F I D Y L E K H Y I D Q K I G A V I S T V N P E I A F A A I K D R D Y F
 TAACCTAACATCATTTCGATTAACCTAGAGAAACATTATATGATCAAAAGATTGGTGCTGTCATTCCACAGTCAACCTGAAATTCGATTGCGAGCTATTAAGATCGGGACTACTT 960
 ClaI BclI

D V L S S S N F I L P D G I G V V M M S R L T N N R L Q S R I A G Y D V F K E L
 TGATGACTATCTTCTCAAACTTCATTTGCTGATGGGATCGGTGTAGTCATGATGTCACGTTTGACTAACACAGGCTTCAGTCCGGAATCGCAGGTACGATGTTTTTAAAGAACT 1080

L G V A N K K K K R I F L Y G A K K D V I K G V V S K I S S E Y P N I K I A G Y
 TCTGGCGTTGCTAACAAAAAGAAAAACGCATATTTCTGTACGGTGCGAAGAAAGATGTTATAAAGGTGTTGATCCAAAATTTCTTCTGAATACCAACATTAAATCGCAGGTTA 1200

S D G Y V Q D R T L V A K Q I A R A N P D M V F V A L G Y P H Q E K F I H N Y R
 TTCAGACGGTTATGTTCAAGATCGCACCTCTAGTGGCAAAACAAATTGCAAGAGCAAATCCTGACATGGTATTCTGATGCTTTAGGGTATCCTCATCAAGAGAAATTCATTACATACTACAG 1320

H L F P K A V S I G L G G S F D V F S G N V K R A P S W M I R L N L E W F Y R L
 ACACTTGTTCCCTAAAGCGGTATCTATAGTCTAGGAGGAGCTTTGATGATTATTAGCGGTAATGTCAAACGGGCCCTTCCTGGATGATACGTTTAAATTTAGAAATGGTTCTACCGGTT 1440
 HindIII

I L N P W R W K R M L S I P K Y A L T V L K E E K N K K T F Y P K P E K D H T K
 AATCTTGAATCCATGGCGATGGAAAGAAATGCTTAGCATCCAAAATATGCTTTAACGGTTTTAAAGAAAGAAAAAACAAAAAACTTTCTATCCAAAGCCTGAAAAAGATCATACAAA 1560

End OrfA

Q I *

Start OrfB

M K I R S L L A N C Y L Y F L S A I A F F L O W V K

ACAGATTTAAAAACAAAACCTTCAAATAAATGGATGATAAAAAATGAAAAAAGATCACTACTGGCGAATTGCTATTTGATTTTTTATCAGCAATCGCCTTTTCTGCAATGGGTGAAA 1680

P E S K V T L L I S F E A N A K A I L E E Y E Q G Q Y S Y K L N I L Y T Q Q A S
 CCTGAAAGCAAAGTGACCTCTCTGATCTCATTCGAAGCAAACGCAAAGCGATTCTTGAGGAATATGAACAAGGACAATACTCTTATAAATTAACATCCTGTATACCCAGCAGGCCTCT 1800

A I A E S F P N V D A Y L L Q E K N P I H L I K A V Y L M F N S K V I I T D N Y
 GCTATTGCCGAGAGCTTCCGAATGTAGATGCATATTTATTACAAGAAAAACCCATCCATCTCATTAAAGCTGTATACCTAATGTTTAAACAGTAAAGTGATCATAACAGATAATTAC 1920
 BclI

F L L T S V L N K R K Q T K C I Q V W H A N G S L K K F G L E D I T N M Q R T K
 TTTCTATTAACAAGTGTGTTAAATAAGCGGAAGCAAACGAAATGTATCCAAGTATGGCAGCAAATGGCTCACTCAAAAAATTTGGTTTGGGAAGACATTACAAACATGCAACGGGACGAAA 2040

T D I K R F Q K V Y S S Y D Y L T V G S E E M A N I F K K S F G I K D N Q L L K
 ACTGATATTAAGGTTCCAAAAAGTATACAGTTCTGATGATTATTTAACTGTGGCTCAGAAGAGATGGCAAATATCTTCAAGAAATCATTTGGAATAAAGGATAACCAATATTAAAAA 2160

[illegible]

Fig. 3. Nucleotide sequence of the *B. subtilis* *tagABC* operon and part of the *tagDEF* operon. The complete DNA sequence of 4263 nucleotides derived from the clones shown in Fig. 2 is presented together with landmark restriction sites. Except for the intergenic region between nucleotides 401 and 799, where both strands are shown, only the non-coding strand is given. Potential ribosome-binding sites are indicated (= =, paired bases; --, unpaired bases). The inferred amino acid sequences of the orfs are given above (orfs A, B, C' and C) or below (orfD) the nucleotide sequence, in one-letter code. The hydrophobic domain comprising 25 amino acids at the amino terminus of the predicted product of orfB is underlined. The hyphenated region of dyad symmetry immediately following orfC is also shown (> > > > < < < <)

<i>tagF</i> 463	KPSHTTYLQTHWGT-PLKRLAMDMEEVHMPGTNTKKYKRNFIKEASNWDYLISPNG	518
	* . * . * * . . * * * . * * * * . * . * .	
<i>tagB</i> 115	KRKQTKCIQVWHANGSLKKFGLD-DITNMQRKTQDI-KR-FQKVYSSYDYLTVGSE	168
<i>tagF</i> 519	YSTEIFTRAFQF-NKTMIESGYPRNDFLHNDNNEETISLIKSRLNIPRDKKVILYA	572
	. . * * . * * * * . * . * * . * * * * * . * . * .	
<i>tagB</i> 169	EMANIFKKSFGIKDNQLLKIGVPLTDPYRENKK-KIS--DT-LNIQR-KKIILYA	215
<i>tagF</i> 573	PTWRDDQFYAKGRYKFDLDDLHQLRQELGNEYIVILRMHYLVAENFDLGPFEFGFA	628
	** ** * . . . * . * . * . * . * . * . * . * . * .	
<i>tagB</i> 216	PTRFD---YNM-Q-SIQLPFTEEQLIHLKKEEYVLFVKLHPAIQNNIDIKYSSDYI	768
<i>tagF</i> 629	YDFSAYEDIRELYMVSDLLITDYSSVFFDFANLKRPMLEFFVPDIETYRDKLRG	681
	* * * . . * * . * . * . * . * . * . * . * . * . * .	
<i>tagB</i> 269	KDVSNYA-LFDLLMAADILITDYSSVPFEFSILNKPILFYTYDLKLYQQK-RG	320
<i>cct</i> 110	GVFDLFHLGHMKQLEQCKKAFFNPVTLIVGVPSDKIT-HKLKGLTV	153
	* * * * * * . * * . * . * . * . * . * . * .	
<i>tagD</i> 8	GTFDLLHWGHIKLLERAKQ-LGDY-LVVAISTDEFNLQKQKK-AY	49
	 * * * * . * .	
<i>str</i> 296	ISEKYYVLKKGEK-PY	310
<i>cct</i> 154	LTDKQRCETLTHCRWVDEVVPNAPWCVTPEFLLEHKIDH-VAHDD	197
	. . * * * * * . * . * . . . * . * . * . * .	
<i>tagD</i> 50	HSYEHRKLILETIRYVDEVIPEKNWEQKKQDIIDHNIDVFVMGDD	94
	 * * * . * .	
<i>str</i> 311	DPFDRSHLKLFTIKYVD	327

Fig. 4. Alignment of the predicted amino acid sequence of the *tagB* product with that of the *B. subtilis tagF* product (upper part) and alignment of *tagD* product with the sequences of the *Saccharomyces cerevisiae* cholinephosphate cytidyltransferase (*cct*) and the *Streptococcus equisimilis* streptokinase precursor (*str*) (lower part). Identical residues are indicated by (*) and conserved residues by (·).

Potential transcription signals

The 400 bp intergenic region between the two transcription units was inspected for sequences having appreciable homology with consensus promoters recognized by the various forms of RNA polymerase found in *B. subtilis*. Several potential candidates were found. Among them were two possible SigA promoters (Moran *et al.*, 1982), one 30 bp upstream from the AUG start codon of *tagA* and the other about 260 bp upstream from the AUG start codon of *tagD*. This intergenic region also contains sequences on both strands showing homology (11 out of 13 positions matching) with the consensus -10 and -35 regions recognized by a minor form of RNA polymerase associated with SigC (Johnson *et al.*, 1983). There is a 4 bp overlap (nucleotides 463-466) between the -35 regions of these putative, divergent

SigC promoters. Which, if any, of these sequences represents a functional promoter remains to be determined. Inspection also revealed sequences on both DNA strands - CgTTCATtcaAAtGTaAa (N₁₀) TATtta and tTaaCATAAATCTtcaAC (N₁₀) cATtAT, 153 and 281 nucleotides upstream from the AUG start codon of *tagA* and *tagD*, respectively - showing appreciable homology with the 'pho box' motif associated with *E. coli* genes subjected to phosphate regulation - CT^G/_TTCATA^T/_A A^T/_ACTGTCA^C/_T (N₁₀) TATAAT (Makino *et al.*, 1986, 1988).

Separated by one base from the two *tagC* termination codons (UGA, UAA) there is a 39 bp sequence capable of forming a potential hairpin-loop structure with a ΔG value of -29.4 kcal mol⁻¹ (123.0 kJ mol⁻¹). This structure, neither particularly G+C-rich nor immediately followed by a run of T residues, could also serve as the

Fig. 5. Assignment of genetic markers to orfs. Subclones and their deleted derivatives are indicated together with their transforming activities against *tag-1*, *tag-11* and *tag-12* recipients. Symbols + and - in the box indicate the presence ($> 10^3$ transformants per μ g DNA) and absence of activity, respectively. Those indicated (+) showed weak activity (see text). ND, Not determined.

terminator for an adjacent convergently transcribed gene (Ph. Margot & D. Karamata, unpublished results).

Assignment of mutations to genes

A number of conditional lethal mutations affecting poly(groP) biosynthesis have been isolated and their approximate physical locations determined (Briehl *et al.*, 1989; Mauël *et al.*, 1989). Expression of three closely linked mutations confers, at a non-permissive temperature, one of two different phenotypes: mutants bearing *tag-11* or *tag-12* mutations are devoid of a CDP-glycerol pool, whereas mutants bearing the *tag-1* mutation accumulate this precursor (Pooley *et al.*, 1991). All three mutations can be corrected by transformation with plasmid 5307 (Mauël *et al.*, 1989), which contains *tagA*, *B* and *D* as well as parts of *tagC* and *tagE* (Fig. 5). To assign the group I mutations to specific genes, appropriate recipient strains were transformed with smaller subclones. Both *tag-11* and *tag-12* mutations were corrected by p5316, which covers parts of *tagD* and *tagE* (Fig. 5). Assignment of these mutations to *tagD* was achieved by transforming appropriate strains with various deletion derivatives of p5316. While correction of *tag-11* and *tag-12* mutations was very efficient ($> 10^3$ transformants per μ g DNA) with p5316 and derivatives p5316-1 to p5316-4, it fell to only 50 per μ g DNA with p5316-5. No transformants were obtained with p5314. Therefore, the

tag-11 and *tag-12* mutations affect *tagD*; they presumably lie between nucleotides 115 and 255 (Figs 3 and 5). The same approach was used to localize the *tag-1* mutation, previously shown to be corrected by p5312 (Mauël *et al.*, 1989), which contains DNA sequences overlapping *tagA* and *tagB*. Since *tag-1* can be corrected by deletion derivatives p5310-1, p5310-2 and p5310-3 but not by derivative p5310-4, it was assigned to *tagB*, and tentatively placed between nucleotides 1966 and 2163 (Figs 3 and 5).

Discussion

Analysis by insertional mutagenesis of a 22 kbp region of the *B. subtilis* chromosome, located at about 310° on the genetic map (Piggot, 1989), has revealed that it contains at least six transcription units. Two are concerned with poly(groP) biosynthesis, one with poly(groP) glucosylation and one with poly(glucose *N*-acetylgalactosamine 1-phosphate) synthesis. Another, insertions in which cause decreased susceptibility to phage $\phi 29$, is probably involved in the biosynthesis of a surface component required for phage infection. The function of the sixth transcription unit, identified by plasmid insertions that seriously impair bacterial growth and viability, is not known. These findings suggest that the entire 22 kbp region is probably devoted to production of cell envelope

components. On either side of this region lie genes encoding autolysins (R. Studer, Ph. Margot & D. Karamata, unpublished results). Additional cell surface determinants located on the *hisA*-proximal side of *gtaB* are concerned with motility and competence. The *degU* locus, which encodes a sensor/regulator couple, the sensor being located in the cell envelope (Kunst *et al.*, 1988; Henner *et al.*, 1988), also lies in this region. Thus, all mutations that have so far been mapped in a substantially larger DNA segment (> 60 kbp), of which the 22 kbp region is a part (Young *et al.*, 1989), affect cell surface components.

The *tag* genes lie to one side of the 22 kbp region and insertional mutagenesis has shown that they are organized in two distinct transcription units. The nucleotide sequence data are entirely in accord with this conclusion (Fig. 3) and provide evidence for the existence of six (or possibly seven) genes organized in two divergently transcribed operons denoted *tagABC* and *tagDEF*. Analysis by transformation has shown that mutations conferring a Tag⁻ phenotype are all located within these two operons. Since some eight or nine steps are apparently involved in the biosynthesis and incorporation of poly(groP) into the bacterial cell wall (Ward, 1981), additional *tag* genes may well await discovery elsewhere. However, an alternative possibility, which cannot be ruled out at this stage, is that some of the genes identified here may encode multifunctional enzymes.

According to the nucleotide sequence, the *tagABC* operon comprises three (or possibly four) genes. The extremities of the operon, as indicated by the nucleotide sequence, are entirely consistent with the results of insertional mutagenesis (Fig. 1).

Analysis of the sequence data, in conjunction with previously published results, reveals several features of interest concerning the *tagB* product. Growth of strains bearing the *tagB1* mutation is thermosensitive. At a non-permissive temperature, bacteria become irregularly rounded and display abnormal cell wall thickening (Boylan *et al.*, 1972; Karamata *et al.*, 1972). Recent evidence has revealed an absence of phosphate incorporation into their cell walls as well as an accumulation of CDP-glycerol, a poly(groP) precursor (Pooley *et al.*, 1991).

Since poly(groP) is a cell wall component, it might be anticipated that the enzymes concerned with its biosynthesis would be membrane-associated. Indeed, biochemical evidence indicates a membrane location for at least two of them, poly(groP) polymerase and poly(groP) translocase (Ward, 1981). The predicted product of *tagB* has a 25 amino acid amino-terminal domain that might function as a membrane anchor (Watson, 1984; Singer *et al.*, 1987). This domain also contains the sequence L-A-N-C-Y-L, which is similar to the consensus of the

cleavage site for precursors of lipoproteins (Gilson *et al.*, 1988), suggesting perhaps that the *tagB* product may be anchored to the cell membrane via an amino-terminal glyceride-cysteine linkage. None of the predicted products of the *tagABC* operon have reiterated hydrophobic alpha-helical domains of sufficient length to serve as membrane-spanning segments. However, this observation does not rule out the possibility of a membrane location for these gene products. Like porins (Rosenbusch, 1990) and the type IV transmembrane integral proteins (Singer *et al.*, 1987), the membrane-associated poly(groP) biosynthetic machinery might adopt a conformation whereby the external surface in contact with the apolar tails of membrane phospholipids is hydrophobic, whereas the environment in which the highly negatively charged poly(groP) precursors are bound, sequentially assembled and presumably extruded, may be hydrophilic. The predicted products of *tagA* and *tagB* are extremely basic polypeptides (Table 2). The substantial net positive charge that they would bear at physiological pH could play a significant role in electrostatic interactions with the negatively charged surface(s) of the bacterial cell membrane or, perhaps, with teichoic acids themselves, as previously suggested by Honeyman & Stewart (1989) for the *tagF* product.

The *tagB* and *tagF* products are likely to be functionally related because the phenotypes associated with the *tagB1* and *tagF3* mutations are indistinguishable (Boylan *et al.*, 1972; Karamata *et al.*, 1972; Pooley *et al.*, 1987). It is of interest, in this context, that a striking similarity was found between the predicted products of *tagB* and *tagF* (Fig. 4), although the biological significance of this similarity is not yet understood. The *tagB* gene has a CAI value of 0.43 (Table 2) and a weak ribosome-binding site, suggesting perhaps that it may be relatively weakly expressed (Sharp *et al.*, 1990).

Very little can be deduced concerning the precise functions of the predicted products of *tagA* and *tagC*. They show no statistically significant homologies with other sequenced proteins and the phenotypes associated with mutations in these genes remain to be determined. Whether or not the predicted product of *orfC'* has any biological significance remains to be elucidated.

The *tagDEF* operon comprises three genes, two of which, *tagE* and *tagF*, had formerly been sequenced by Honeyman & Stewart (1989). Within the intergenic region preceding *tagE* (160 bp), these authors identified a putative SigA promoter, preceded by an inverted repeat sequence to which a regulatory element might bind, just upstream from the coding sequence of *tagE*. The integrational plasmid, 5316, which contains this putative SigA promoter, together with substantial parts of the coding sequences of *tagD* (upstream) and *tagE* (downstream), could not be established by homologous

recombination in the bacterial chromosome (Fig. 1). This finding is inconsistent with the proposal (Honeyman & Stewart, 1989) that transcription of *tagE* and *tagF* initiated at the SigA promoter suffices to support growth, at least under the conditions employed here. It suggests instead that *tagD*, *tagE* and *tagF* form an operon. Analysis of RNA will enable the start point(s) of transcription to be determined precisely.

Two mutations, *tag-11* and *tag-12*, have been localized to *tagD*. Their associated phenotypes are thermosensitivity of growth (Briehl *et al.*, 1989) and an absence of CDP-glycerol, a poly(groP) precursor, at the non-permissive temperature (Pooley *et al.*, 1991). The extremely hydrophilic nature of the predicted product of *tagD* and the local homology that it shares with cholinephosphate cytidyltransferase of *Saccharomyces cerevisiae* (Fig. 4), are consistent with the conclusion that *tagD* encodes the cytoplasmic enzyme, glycerol-3-phosphate cytidyltransferase (Pooley *et al.*, 1991). This enzyme produces the principal precursor required for biosynthesis of poly(groP), the major teichoic acid found in the bacterial cell wall. The CAI value of 0.54 obtained for the predicted product of *tagD* is not inconsistent with a high level of expression (Sharp & Li, 1986; Shields & Sharp, 1987; Sharp *et al.*, 1990), as would be expected for the gene encoding this enzyme.

The predicted products of all three genes in the *tagDEF* operon are extremely hydrophilic (Honeyman & Stewart, 1989; Table 2), suggesting that they are cytosolic (soluble) proteins. The earlier conclusion that mutation *gtaA12*, located in *tagE*, affects the membrane-associated enzyme UDP-glucose poly(groP) glucosyltransferase (Young, 1967) would merit re-evaluation. In the light of the nucleotide sequence information, and the evidence that *tagD* encodes glycerol-3-phosphate cytidyltransferase, it seems possible that the *tagDEF* operon is dedicated to production of soluble enzymes concerned with biosynthesis of poly(groP) precursors.

The *tagABC* and *tagDEF* operons are separated by an intergenic region, within which divergent transcription is presumably initiated. Several motifs that might play a role in transcriptional regulation, including sequences resembling SigA and SigC promoters, are present on both DNA strands, as are sequences resembling (12/18 positions matching) the 'pho box' motif characteristic of *E. coli* genes subjected to phosphate regulation (Makino *et al.*, 1986). A sequence resembling the 'pho box' motif (11/18 positions matching) has also been found upstream from the *phoP* gene in *B. subtilis* (Seki *et al.*, 1987). In *E. coli* the 'pho box' motif is the sequence to which the PhoB protein binds to promote transcription (Makino *et al.*, 1988). It would not be surprising were the *tag* operons under phosphate regulation, since it has long been known that during phosphate limitation the main cell

wall teichoic acid is replaced by an alternative phosphate-free anionic polymer, teichuronic acid (Ellwood & Tempest, 1972). However, some form of negative control would be expected, to reduce transcription of the *tagABC* and *tagDEF* operons during phosphate limitation. Experiments are in progress to determine the roles, if any, of the various possible transcriptional elements identified by sequence inspection in the 400 bp intergenic region between the *tagABC* and *tagDEF* operons.

The organization of functionally related genes in two transcription units that are controlled by divergent promoters is a common feature of gene organization in prokaryotes (Beck & Warren, 1988). A previously reported example in *B. subtilis* is that of the *gltA* and *gltC* loci involved in glutamate biosynthesis (Bohannon & Sonenshein, 1989). This form of gene organization provides several advantages, among which are the potential for tight co-ordinate control and for translocation of these related functions as a self-contained module to new genetic locations without loss of autonomy (Beck & Warren, 1988). The latter notion finds here convincing support. In 168/W23 interstrain hybrids of *B. subtilis* there is integral replacement of the 168 *tag* genes, concerned with poly(groP) biosynthesis, by the functionally equivalent, but non-homologous, W23 *tar* genes, concerned with poly(ribitol phosphate) biosynthesis. In genetic crosses, the *tag* and *tar* genes behave as a mutually exclusive pseudo-allelic pair of cassettes or modules (Karamata *et al.*, 1987; Young *et al.*, 1989).

Finally, the availability of the cloned genes will facilitate the study of both the enzymology and the regulation of anionic polymer biosynthesis, providing possible new insights into factors controlling cell surface extension.

Sequence analysis was performed using the UK SEQNET and Swiss Friederich Miescher Institute facilities. We thank A. F. W. Coulson for help with the Edinburgh DAP database scanning and optimal alignment programs and P. Sharp for sending the CODONS sequence analysis software for determination of CAI values. The excellent technical assistance of Martine Pagnier is gratefully acknowledged. Part of this work was supported by the Foundation of the 450th Anniversary of the University of Lausanne.

References

- BADDILEY, J. (1970). Structure, biosynthesis and function of teichoic acids. *Accounts of Chemical Research* **3**, 98–105.
- BECK, C. F. & WARREN, R. A. J. (1988). Divergent promoters, a common form of gene organization. *Microbiological Reviews* **52**, 318–326.
- BIRNBOIM, H. C. (1983). A rapid alkaline method for the isolation of plasmid DNA. *Methods in Enzymology* **100**, 243–255.
- BIRNBOIM, H. C. & DOLY, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research* **7**, 1513–1523.

- BOHANNON, D. E. & SONENSHEIN, A. L. (1989). Positive regulation of glutamate biosynthesis in *Bacillus subtilis*. *Journal of Bacteriology* **171**, 4718–4727.
- BOYLAN, R. J., MENDELSON, N. H., BROOKS, D. & YOUNG, F. E. (1972). Regulation of the bacterial cell wall: analysis of a mutant of *Bacillus subtilis* defective in biosynthesis of teichoic acids. *Journal of Bacteriology* **110**, 281–290.
- BRIEHL, M., POOLEY, H. M. & KARAMATA, D. (1989). Mutants of *Bacillus subtilis* thermosensitive for growth and wall teichoic acid synthesis. *Journal of General Microbiology* **135**, 1325–1334.
- CAPALDI, R. A. (1982). Structure of intrinsic membrane proteins. *Trends in Biochemical Sciences* **7**, 292–295.
- CHAMBERS, S. P., PRIOR, S. E., BARSTOW, D. A. & MINTON, N. P. (1988). The pMTL *nic*-cloning vectors. I. Improved polylinker region to facilitate the use of sonicated DNA for nucleotide sequencing. *Gene* **68**, 139–149.
- CHUNG, C. T. & MILLER, R. H. (1988). A rapid and convenient method for the preparation and storage of competent bacterial cells. *Nucleic Acids Research* **16**, 3580.
- COLLINS, J. F., COULSON, A. F. W. & LYALL, A. (1988). The significance of protein sequence similarities. *CABIOS* **4**, 67–71.
- COULSON, A. F. W., COLLINS, J. F. & LYALL, A. (1987). Protein and nucleic acid sequence database searching: a suitable case for parallel processing. *Computer Journal* **30**, 420–424.
- DAYHOFF, M. O., SCHWARTZ, R. M. & ORCUTT, B. C. (1978). A model of evolutionary change in proteins. In *Atlas of Protein Sequence and Structure*, vol. 5, supplement 3, pp. 345–353. Edited by M. O. Dayhoff. Washington, DC: NBRF.
- DEVEREUX, J., HAEBERLI, P. & SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387–395.
- ELLWOOD, D. C. & TEMPEST, D. W. (1972). Effect of environment on bacterial wall content and composition. *Advances in Microbial Physiology* **7**, 83–117.
- ESTRELA, A.-I., POOLEY, H. M., DE LENCASTRE, H. & KARAMATA, D. (1991). Genetic and biochemical characterization of *Bacillus subtilis* 168 mutants specifically blocked in the synthesis of the teichoic acid, poly(3-*O*- β -D-glucopyranosyl-*N*-acetylgalactosamine 1-phosphate); *gneA*, a new locus, is associated with UDP-*N*-acetylglucosamine 4-epimerase activity. *Journal of General Microbiology* **137**, 943–950.
- FERRARI, F. A., NGUYEN, A., LANG, D. & HOCH, J. A. (1983). Construction and properties of an integrable plasmid for *Bacillus subtilis*. *Journal of Bacteriology* **154**, 1513–1515.
- GILSON, E., ALLOING, G., SCHMIDT, T., CLAVERYS, J.-P., DUDLER, R. & HOFNUNG, M. (1988). Evidence for high affinity binding-protein dependent transport systems in Gram-positive bacteria and in *Mycoplasma*. *EMBO Journal* **7**, 3971–3974.
- GROSSBERGER, D. (1987). Minipreps of DNA from bacteriophage lambda. *Nucleic Acids Research* **15**, 6737.
- HENNER, D. J., YANG, M. & FERRARI, E. (1988). Localization of *Bacillus subtilis* *sacU*(Hy) mutations to two linked genes with similarities to the conserved procaryotic family of two-component signalling system. *Journal of Bacteriology* **170**, 5102–5109.
- HONEYMAN, A. L. & STEWART, G. C. (1988). Identification of the protein encoded by *rodC*, a cell division gene from *Bacillus subtilis*. *Molecular Microbiology* **2**, 735–741.
- HONEYMAN, A. L. & STEWART, G. C. (1989). The nucleotide sequence of the *rodC* operon in *Bacillus subtilis*. *Molecular Microbiology* **3**, 1257–1268.
- JOHNSON, W. C., MORAN, C. P., JR & LOSICK, R. (1983). Two RNA polymerase σ factors from *Bacillus subtilis* discriminate between overlapping promoters for a developmentally regulated gene. *Nature, London* **302**, 800–804.
- KARAMATA, D. & GROSS, J. (1970). Isolation and genetic analysis of temperature sensitive mutants of *Bacillus subtilis* defective in DNA synthesis. *Molecular and General Genetics* **108**, 277–287.
- KARAMATA, D., MCCONNELL, M. & ROGERS, H. J. (1972). Mapping of *rod* mutants of *Bacillus subtilis*. *Journal of Bacteriology* **111**, 73–79.
- KARAMATA, D., POOLEY, H. M. & MONOD, M. (1987). Expression of heterologous genes for wall teichoic acids in *Bacillus subtilis* 168. *Molecular and General Genetics* **207**, 73–81.
- KUNST, F., DEBARBOUILLE, M., MSADEK, T., YOUNG, M., MAUËL, C., KARAMATA, D., Klier, A., RAPOPORT, G. & DEDONDER, R. (1988). Deduced polypeptides encoded by the *Bacillus subtilis* *sacU* share homology with two-component sensor-regulator systems. *Journal of Bacteriology* **170**, 5093–5101.
- KYTE, J. & DOOLITTLE, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**, 105–132.
- MAKINO, K., SHINAGAWA, H., AMEMURA, M. & NAKATA, A. (1986). Nucleotide sequence of the *phoB* gene, the positive regulatory gene for the phosphate regulon of *Escherichia coli*. *Journal of Molecular Biology* **190**, 37–44.
- MAKINO, K., SHINAGAWA, H., AMEMURA, M., KIMURA, S., NAKATA, A. & ISHIHAMA, A. (1988). Regulation of the phosphate regulon of *Escherichia coli*. Activation of *pstS* transcription by PhoB protein *in vitro*. *Journal of Molecular Biology* **203**, 85–95.
- MALKE, H., ROE, B. & FERRETTI, J. J. (1985). Nucleotide sequence of the streptokinase gene from *Streptococcus equisimilis* H46A. *Gene* **34**, 357–362.
- MAUËL, C., YOUNG, M., MARGOT, PH. & KARAMATA, D. (1989). The essential nature of teichoic acids in *Bacillus subtilis* as revealed by insertional mutagenesis. *Molecular and General Genetics* **215**, 388–394.
- MICHAELIS, S. & BECKWITH, J. (1982). Mechanism of incorporation of cell envelope proteins in *Escherichia coli*. *Annual Review of Microbiology* **36**, 435–465.
- MORAN, C. P., JR, LANG, N., LEGRICE, S. F. J., LEE, G. & STEPHENS, M. (1982). Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. *Molecular and General Genetics* **186**, 339–346.
- OUTRAM, J. D., LOUGHLIN, M., SWINFELD, T.-J., BREHM, J. K., THOMPSON, D. E. & MINTON, N. P. (1988). Introduction of plasmids into whole cells of *Clostridium acetobutylicum* by electroporation. *FEMS Microbiology Letters* **56**, 83–88.
- PIGGOT, P. J. (1989). Revised genetic map of *Bacillus subtilis* 168. In *Regulation of Prokaryotic Development*, pp. 1–41. Edited by I. Smith, R. A. Slepecky & P. Setlow. Washington, DC: American Society for Microbiology.
- PIGGOT, P. J., CURTIS, A. M. & DE LENCASTRE, H. (1984). Use of integrational plasmid vectors to demonstrate the polycistronic nature of a transcriptional unit (*spoIIA*) of *Bacillus subtilis*. *Journal of General Microbiology* **130**, 2123–2136.
- POOLEY, H. M., PASCHOUD, D. & KARAMATA, D. (1987). The *gtaB* marker in *Bacillus subtilis* 168 is associated with a deficiency in UDP-glucose pyrophosphorylase. *Journal of General Microbiology* **133**, 3481–3493.
- POOLEY, H. M., ABELLAN, F.-X. & KARAMATA, D. (1991). A conditional-lethal mutant of *Bacillus subtilis* 168 with a thermosensitive glycerol-3-phosphate cytidyltransferase, an enzyme specific for the synthesis of the major wall teichoic acid. *Journal of General Microbiology* **137**, 921–928.
- ROSENBUSCH, J. P. (1990). Structural and functional properties of porin channels in *E. coli* outer membranes. *Experientia* **46**, 167–173.
- SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5463–5467.
- SEKI, T., YOSHIKAWA, H., TAKAHASHI, H. & SAITO, H. (1987). Cloning and nucleotide sequence of *phoP*, the regulatory gene for alkaline phosphatase and phosphodiesterase in *Bacillus subtilis*. *Journal of Bacteriology* **169**, 2913–2916.
- SHARP, P. M. & LI, W. H. (1986). The codon adaptation index – a measure of directional synonymous codon usage bias and its potential applications. *Nucleic Acids Research* **14**, 7737–7749.
- SHARP, P. M., HIGGINS, D. G., SHIELDS, D. C., DEVINE, K. M. & HOCH, J. A. (1990). *Bacillus subtilis* gene sequences. In *Genetics and Biotechnology of Bacilli*, vol. 3, pp. 89–98. Edited by M. Zukowski, A. T. Ganesan & J. A. Hoch. Orlando, Florida: Academic Press.
- SHIBAEV, V. N., DUCKWORTH, M., ARCHIBALD, A. R. & BADDILEY, J. (1973). The structure of a polymer containing galactosamine from walls of *Bacillus subtilis* 168. *Biochemical Journal* **135**, 383–384.

- SHIELDS, D. C. & SHARP, P. M. (1987). Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational biases. *Nucleic Acids Research* **15**, 8023–8040.
- SINGER, S. J., MAHER, P. A. & YAFFE, M. P. (1987). On the transfer of integral proteins into membranes. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 1960–1964.
- TINOCO, I., JR, BORER, P. N., DENGLER, B., LEVINE, M. D., UHLENBECK, O. C., CROTHERS, D. M. & GRALLA, J. (1973). Improved estimation of secondary structure in ribonucleic acids. *Nature New Biology* **246**, 40–41.
- TSUKAGOSHI, Y., HIKAWA, J. & YAMASHITA, S. (1987). Molecular cloning and characterization of the gene encoding cholinephosphate cytidyltransferase in *Saccharomyces cerevisiae*. *European Journal of Biochemistry* **169**, 477–486.
- WARD, J. B. (1981). Teichoic and teichuronic acids: biosynthesis, assembly and location. *Microbiological Reviews* **45**, 211–243.
- WATSON, M. E. E. (1984). Compilation of published signal sequences. *Nucleic Acids Research* **12**, 5145–5164.
- YOUNG, F. E. (1967). Requirement of glucosylated teichoic acid for adsorption of phage in *Bacillus subtilis* 168. *Proceedings of the National Academy of Sciences of the United States of America* **58**, 2377–2384.
- YOUNG, M., MAUËL, C., MARGOT, PH. & KARAMATA, D. (1989). Pseudoallelic relationship between non-homologous genes concerned with biosynthesis of polyglycerol phosphate and polyribitol phosphate teichoic acids in *Bacillus subtilis* strains 168 and W23. *Molecular Microbiology* **3**, 1805–1812.