Sequence of the 305°–307° region of the *Bacillus subtilis* chromosome

Blazenka Soldo, Vladimir Lazarevic, Catherine Mauël
and Dimitri Karamata

Author for correspondence: Dimitri Karamata. Tel: +41 21 320 60 75. Fax: +41 21 320 60 78.
e-mail: dkaramat@ulys.unil.ch

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

The nucleotide sequence of the *Bacillus subtilis* 168 chromosomal segment located between *yvhJ* (307°) and *secA* (305°) was determined. This 20.3 kb region encompasses 23 ORFs, 17 of which have been sequenced previously. Comparison of sequences obtained here with the previously obtained ones revealed seven discrepancies. The products of the sequenced genes are involved in the regulation of degradative enzymes, competence, flagellar motility and protein secretion. Putative functions of newly identified genes are based on sequence homologies.

Keywords: *Bacillus subtilis*, genome sequencing, *yvhJ*–*secA* region

INTRODUCTION

The nucleotide sequence of the *Bacillus subtilis* 168 chromosomal region located between 305° and 307°, obtained by piece-meal sequencing, consists essentially of two contigs of 7.9 and 9.0 kb, respectively. Among others, the former contains the *degSU* operon involved in degradative enzyme production (Kunst *et al.*, 1988; Henner *et al.*, 1988; Tanaka & Kawata, 1988), the competence operon *comFABC* (Londoño-Vallejo & Dubnau, 1993) and a σ^D -controlled incomplete flagellar operon containing *flgM* and the 5'-terminal part of *flgK* (Mirel *et al.*, 1994). The 9 kb contig encompasses the flagellin structural gene *hag* (Mirel & Chamberlin, 1989), a flagellar operon encoding the *FliD*, *FliS* and *FliT* proteins (Chen & Helmann, 1994), and the *secA* gene whose product is involved in protein secretion (Sadaie *et al.*, 1991).

In this contribution we report the sequence of the entire 20.3 kb region between 305° and 307°, i.e. 3.4 kb of new sequence. Discrepancies between sequences presented in this report and previously reported sequences are discussed.

METHODS

Bacterial strains, plasmids and phages. Segments of the *B. subtilis* chromosomal inserts in phages λ 63, λ 55, λ 38 (Mauël *et al.*, 1989) and in the plasmid p3900 were subcloned in pMTL20EC

(Chambers *et al.*, 1988) and maintained in *Escherichia coli* DH5 α (Hanahan, 1983) and JM83 (Yanisch-Perron *et al.*, 1985). Plasmid p3818, carrying a 1.95 kb *SphI*–*HindIII* chromosomal segment in pDIA5304 (Fig. 1; Glaser *et al.*, 1993), was inserted into the *B. subtilis* 168 chromosome. *PstI* chromosomal fragments of a chosen chloramphenicol-resistant transformant were recircularized and used to transform *E. coli* TP610 (Glaser *et al.*, 1993) yielding plasmid p3900 (Fig. 1). λ phages were propagated on *E. coli* P2392.

DNA preparation. Plasmids were prepared by the boiling method (Del Sal *et al.*, 1988). λ DNA was obtained as described by Grossberger (1987).

DNA sequencing and analysis. DNA sequencing (Sanger *et al.*, 1977) was performed on both strands by primer walking, using the subclones of the λ library and of the plasmids obtained by plasmid rescue. Labelling reactions were carried out with a Sequenase Version 2.0 Kit (USB) and [α -³⁵S]dATP (Amersham). Site-specific primers (18-mers) were synthesized with a SM oligonucleotide synthesizer (Beckman). The junctions between subclones were verified by sequencing of plasmids containing longer inserts or λ DNA. The sequence was compiled and analysed by the University of Wisconsin Computer Group software (Devereux *et al.*, 1984). Multiple protein sequence alignments were carried out with the PILEUP program (Feng & Doolittle, 1987) and used to construct a dendrogram.

Transformation. *E. coli* DH5 α and JM83 competent cells were prepared and transformed by the procedure of Chung & Miller (1988). Competent cells of the *E. coli* strain TP610 were prepared and transformed by the calcium chloride method (Sambrook *et al.*, 1989). Transformation of *B. subtilis* competent cells was carried out as described by Karamata & Gross (1970).

The GenBank accession number for the sequence reported in this paper is U56901.

Fig. 1. Organization of the chromosomal region between *yvhJ* and *secA*. Plasmid subclones used for sequencing are represented by bold lines. Thick arrows correspond to ORFs. Thin dashed arrows indicate proposed transcriptional units. ♀ represents putative stem-loops. Restriction sites are denoted: B, *Bam*HI; E, *Eco*RI; G, *Bgl*II; P, *Pst*I; S, *Sph*I; Sa, *Sal*I; Sc, *Sac*I; Sm, *Sma*I. Hatched bars indicate data bank entries. Promoters are denoted by filled triangles: A, σ^A -controlled promoter; D, σ^D -controlled promoter; ?, possible promoter.

RESULTS AND DISCUSSION

The analysis of the sequenced region revealed 23 complete ORFs, organized in nine putative transcription units (Fig. 1). At an appropriate distance, the putative start codons are preceded by sequences complementary to 16S rRNA (Fig. 2). Nineteen ORFs start with the AUG codon, two with GUG, one with UUG, and one with CUG. However, the usage of termination codons is not in agreement with *B. subtilis* preferences (Sharp *et al.*, 1990). The rare UGA and UAG stop codons are used five and eight times, respectively, while the generally frequent one, UAA, is used 10 times (Table 1).

The G+C content of this chromosomal segment is 41.1%, which is somewhat lower than the mean (43%)

for the *B. subtilis* genome (Fahmy *et al.*, 1985). For protein coding sequences, this parameter varies from 34.0% for *flgM* to 45.5% for *secA* (Table 1). The overall G+C content of non-coding regions is 35.6%. For each of the specific intergenic regions longer than 100 bp, the G+C content is lower than that of the flanking genes.

Over 70% of the sequenced region has been sequenced and already published by others (Kunst *et al.*, 1988; Henner *et al.*, 1988; Tanaka & Kawata, 1988; Londoño-Vallejo & Dubnau, 1993; Mirel *et al.*, 1994; Mirel & Chamberlin, 1989; Chen & Helmann, 1994; Sadaie *et al.*, 1991; LaValle & Stahl, 1989). Comparison of these reports with our sequence data revealed seven discrepancies (Table 2). They do not necessarily represent sequencing errors, since they could be due to strain

<i>yvhJ</i>	<u>AUAAAGGAGACUGAAUAAUG</u>
<i>yvhK</i>	<u>UAGCAGGAGGGUCAGCUAGCAUG</u>
<i>degS</i>	<u>CGGAGGGAAUUAUG</u>
<i>degU</i>	<u>UACCAAGGAGGCGUGGCUUGUG</u>
<i>degV</i>	<u>AUAAGGCAGGUAUCGAAGCUAUG</u>
<i>comFA</i>	<u>CGAAAGGAGGCGUGCUAUGUG</u>
<i>comFB</i>	<u>GAUGAGGAGCUGUAAAAAUG</u>
<i>comFC</i>	<u>CAGGAGGAGGCAUUCAUAAACUG</u>
<i>yviB</i>	<u>AAUAAGGAGAAAGAAAGAUUG</u>
<i>flgM</i>	<u>CGAGAGGAUCCUAUG</u>
<i>yviC</i>	<u>AAAAAGGAGAAAGCCCAUG</u>
<i>flgK</i>	<u>AGAAAGGAUUCAGAAAAUG</u>
<i>flgL</i>	<u>AGGAAGGUAGGUAGAGUCACAUG</u>
<i>yviE</i>	<u>GUUCAGGAGGUAGAGUGAUG</u>
<i>yviF</i>	<u>GAAAAGGAUCGUGAACAAUG</u>
<i>csrA</i>	<u>UUGGAGGAGAAUCAUG</u>
<i>hag</i>	<u>AGGGAGGAACAAACAUG</u>
<i>yviH</i>	<u>AUAGGGGAUGAUCAAUUG</u>
<i>fliD</i>	<u>AGAAUAGGAGUGGUUUGAGAUG</u>
<i>fliS</i>	<u>AAUUUGGAGGAUGACACAUG</u>
<i>fliT</i>	<u>AGGCGGGAUCGCAUG</u>
<i>yviI</i>	<u>AAAAGGAGGCGUUCUUUGAUG</u>
<i>secA</i>	<u>AUAGAGGAGCGUUAUAAUG</u>

Fig. 2. Putative translation initiation regions. Bases complementary to 3' ends of 16S rRNA are underlined. Start codons are shown in bold type.

differences. Moreover, cloning and data handling may also account for the discrepancies. For instance, the two differences at the level of sequence termini could be explained by an interpretation error. Indeed, it is most likely that the previously reported termini were part of the cloning vector. Of the remaining five discrepancies, two are in the intergenic regions, and two could represent frame-shift mutations which extend the putative gene products of *yvhK* and *yviF* by 30 and 29 residues, respectively. Finally, a base change in *fliD* leads to the replacement of the Glu-9 by Ala in the deduced polypeptide sequence.

We have shown that the neighbouring 308°–311° chromosomal region consists mainly of divergents separated by inverted repeats encoding putative terminators (Lazarevic *et al.*, 1995). The genetic organization of the region presented here is different. Only two out of 23 genes are transcribed towards the origin of replication, i.e. 6% as compared to 28% characteristic of the 308°–311° segment. However, the fact that the most stable stem-loop structure followed by a run of U residues is located downstream of *yvhJ* (Table 3) confirms our previous finding, i.e. the strongest ρ -independent terminators are generally located between two convergent genes.

In the sequenced region, *yvhKJ* is the only operon transcribed in the direction opposite to that of the replication fork movement. The product of its distal gene, *yvhJ*, is a candidate for a transcriptional regulator. A membrane-anchored highly hydrophobic domain, extending between residues Ile-24 and Tyr-41, suggests that it might be involved in signal(s) transductions from the extracytoplasmic space to the cytoplasm. YvhI shares

homology with the *B. subtilis* LytR protein, which attenuates its own expression as well as that of the divergent *lytABC* operon (Lazarevic *et al.*, 1992), *Streptococcus pneumoniae* CpsA (Guidolin *et al.*, 1994) and *Enterococcus hirae* PsrA, the penicillin-binding protein gene repressor (Ligozzi *et al.*, 1993). However, YvhJ differs from its homologues by the presence of a region rich in residues with hydroxyl groups (Fig. 3). Out of 58 C-terminal amino acids, 63% are either Ser or Thr. Within this domain, five putative N-glycosylation sites (N-X-S/T) are found in a 19 aa stretch.

The homology of a previously sequenced segment of YvhK with the *E. coli* YigZ protein (Nakahigashi & Inokuchi, 1990), *Thermus aquaticus* Ypol (Akhmetzjanov & Vakhitov, 1992) and *Saccharomyces cerevisiae* Ycr59c has already been noted (Koonin *et al.*, 1994). Other members of this protein family of unknown function are *Haemophilus influenzae* YigZ (Fleischmann *et al.*, 1995) and the product of a *Thermus thermophilus* incomplete ORF situated upstream of the *polA* gene, which encodes thermostable DNA polymerase (Asakura *et al.*, 1993). The only eukaryotic member of this family, Ycr59c, is longer (Fig. 4) and, according to the dendrogram in Fig. 5, clearly distinct from its prokaryotic counterparts.

DegS–DegU acts as a two-component sensor-regulation system of degradative enzyme production (Kunst *et al.*, 1988; Henner *et al.*, 1988; Tanaka & Kawata, 1988). In addition, DegU plays a role in competence development (Tanaka & Kawata, 1988; Roggiani *et al.*, 1990). The product of the following gene, which belongs to the putative monocistronic *degV* operon, is similar to *Staphylococcus aureus* Orf2 (Dale *et al.*, 1993), to the plasmid pSK1-encoded protein Yp15 of the same organism (Rouch *et al.*, 1989) and to the *Mycoplasma genitalium* y326 gene product (Fraser *et al.*, 1995) (Fig. 6).

The interdependence between the *yvhKJ* and *degSU* operons with face-to-face promoters is very likely (Fig. 7). The transcriptional regulation of these kinds of divergent operons could be complex due to the interference between RNA polymerase molecules moving in opposite directions, as well as partial duplex RNA formation (Beck & Warren, 1988) between 39 5'-terminal residues of the complementary transcripts (Fig. 7).

Products of *comFA* and *comFC* genes, belonging to the tricistronic σ^A -dependent operon, contribute to competence development (Londoño-Vallejo & Dubnau, 1993). Gene *comFC* is followed by three operons involved in flagellar motility and controlled by a σ^D -factor, known to be responsible for the transcription of cell wall hydrolases (Lazarevic *et al.*, 1992; Margot *et al.*, 1994), as well as that of chemotaxis genes (Márquez *et al.*, 1990). Mirel *et al.* (1994) found that the first of the former three σ^D -controlled operons contains homologues of *Salmonella typhimurium* late flagellar genes *flgM* (Gillen & Hughes, 1991) and *flgK* (Homma *et al.*, 1990). On the basis of our sequence data, this operon comprises at least one more gene, *flgL*. The following three genes, *yviE*, *yviF* and

Table 1. Features of the products deduced from the nucleotide sequences of putative ORFs

Gene	Other name(s)	Endpoints (nt)*	G + C (%)	Start/stop	Bases	Codons	Mass (kDa)	pI	Predicted or known function
<i>yvhJ</i>		45 < 1217	41.6	AUG/UAA	1173	391	43.2	9.62	Membrane-anchored; transcription regulator
<i>yvhK</i>	<i>yvyE</i>	1220 < 1870	40.8	AUG/UAA	651	217	24.8	6.20	Ycr59c/YigZ family (Koonin <i>et al.</i> , 1994)
<i>degS</i>		2087 > 3241	42.9	AUG/UGA	1155	385	45.0	6.33	Sensor; histidine kinase that phosphorylates DegU (Kunst <i>et al.</i> , 1988; Henner <i>et al.</i> , 1988)
<i>degU</i>		3327 > 4013	42.3	GUG/UAG	687	229	25.9	5.97	Transcription regulator of degradative enzymes (Kunst <i>et al.</i> , 1988; Henner <i>et al.</i> , 1988; Tanaka & Kawata, 1988)
<i>degV</i>	<i>degorf3</i>	4114 > 4956	42.9	AUG/UAA	843	281	31.6	5.35	—
<i>comFA</i>		5065 > 6453	42.7	GUG/UAG	1389	463	52.6	10.23	Involved in transformation (Londoño-Vallejo & Dubnau, 1993)
<i>comFB</i>		6516 > 6809	39.4	AUG/UGA	294	98	11.0	4.86	—
<i>comFC</i>		6809 > 7495	40.6	CUG/UAA	687	229	26.4	8.95	Involved in transformation; similar to ATP-dependent DNA/RNA helicases (Londoño-Vallejo & Dubnau, 1993)
<i>yviB</i>	<i>yvyF</i> , <i>orf139</i>	7572 > 7988	38.6	AUG/UGA	417	139	16.2	8.04	Flagellar protein (Mirel <i>et al.</i> , 1994)
<i>flgM</i>		8072 > 8335	34.0	AUG/UAA	264	88	10.0	10.35	Negative regulator of flagellin synthesis (Mirel <i>et al.</i> , 1994)
<i>yviC</i>	<i>yvyG</i> , <i>orf160</i>	8354 > 8833	40.6	AUG/UAG	480	160	18.2	9.12	Flagellar protein (Mirel <i>et al.</i> , 1994)
<i>flgK</i>		8855 > 10375	42.4	AUG/UAG	1521	507	54.4	4.40	Flagellar hook-associated protein 1 (HAP1); ATP/GTP binding
<i>flgL</i>		10389 > 11282	39.9	AUG/UAA	894	298	32.8	4.45	Flagellar hook-associated protein 3 (HAP3)
<i>yviE</i>		11332 > 11904	43.4	AUG/UAA	573	191	21.6	4.88	—
<i>yviF</i>		11928 > 12356	38.2	AUG/UAG	429	143	16.2	4.37	Transmembrane
<i>csrA</i>		12353 > 12574	40.0	AUG/UGA	222	74	8.1	6.53	Accelerates degradation of some transcripts
<i>bag</i>		12722 > 13633	41.3	AUG/UAA	912	304	32.6	4.82	Flagellin (Mirel & Chamberlin, 1989)
<i>yviH</i>	<i>orf99</i>	13869 > 14195	37.0	UUG/UAG	327	109	13.0	4.73	Flagellar protein (Chen & Helmann, 1994)
<i>fliD</i>		14216 > 15709	39.8	AUG/UAA	1494	498	54.5	4.83	Flagellar hook-associated protein 2 (Chen & Helmann, 1994)
<i>fliS</i>		15734 > 16132	42.1	AUG/UGA	399	133	15.1	5.29	Flagellar protein (Chen & Helmann, 1994)
<i>fliT</i>		16132 > 16470	38.9	AUG/UAG	339	113	13.5	5.85	Flagellar protein (Chen & Helmann, 1994)
<i>yviI</i>	<i>orf189</i>	17051 > 17617	43.0	AUG/UAA	567	189	22.0	5.27	σ^{54} modulator
<i>secA</i>		17788 > 20310	45.5	AUG/UAG	2523	841	95.5	5.46	Protein export (Sadaie <i>et al.</i> , 1991); ATP binding

* Stop codons are not included.

Table 2. Discrepancies between published sequences and sequence U56901

Accession no.	Reference	Position	Change		Position*	Remark*
			from	to*		
M23558	Henner <i>et al.</i> (1988)	22 < > 23	–	T	1377	YvhK extended for 30 aa residues
M21658	Tanaka & Kawata (1988)	1231†	T	C	4072	Non-coding region between <i>degU</i> and <i>degV</i>
M26948	Mirel & Chamberlin (1989)	52 < > 53	–	C	12299	YviF extended for 29 aa residues
M26947	LaVallie & Stahl (1989)	1†	C	T	12546	GCG (Ala-65) of <i>csrA</i> replaced by GTG (Val)
		65	C	–	12609 < > 12610	Non-coding region between <i>csrA</i> and <i>bag</i>
U07822	Chen & Helmann (1994)	701	A	C	14241	GAG (Glu-9) of <i>fliT</i> replaced by GCG (Ala)
		3451	T	A	16991	Non-coding region between <i>fliT</i> and <i>yviI</i>

* Refers to sequence U56901.

† Located at the relevant sequence extremity.

Table 3. Potential stem-loop structures and free energy of base pairing calculated according to the rules of Freier *et al.* (1986)

Stem-loop structure	ΔG [kcal mol ⁻¹ (kJ mol ⁻¹)]	Upstream ORF
<pre> UU AAAAAUGCCCCGUCCU UUUUUACGGGCUAGGA U GA </pre>	–22.0 (–92.1)	<i>yvhJ</i>
<pre> UU AGGAGACUUGCCU U UUUUCUGGACGGA A UUU UC </pre>	–15.1 (–63.2)	<i>degU</i>
<pre> UU AAUUCUCCGUU UUUAGAGGCGA U UUUU GA </pre>	–11.9 (–49.8)	<i>degV</i>
<pre> GCUA CAGAA GUCUU A UUUUU AGUA </pre>	–2.5 (–10.5)	<i>comFC</i>
<pre> – U AAGCGG CUC U UUCGCU GAG A UUUUU U G </pre>	–6.1 (–25.5)	<i>flgL</i>
<pre> AAA AUCCUCAC UAGGAGUG A UUUUUAUUUUUU AAA </pre>	–9.9 (–41.4)	<i>csrA</i>
<pre> G AAAAGACCUUGGC UUUUCUGGGACCG U UUUAA U </pre>	–14.4 (–60.3)	<i>bag</i>
<pre> A GGAUCGGAUCU C CCUAGUCUAGA C UUUAUUUUUUUCCGAA A </pre>	–12.7 (–53.2)	<i>fliT</i>
<pre> AU AGAAGCCUCCGUG UUUUUGGAAGGCGC G UUUUUG CU </pre>	–16.0 (–67.0)	<i>yviI</i>

YvhJ								70
LytR								
CpsA	MSRRFKKSRS	QKVKRSVNIV	LLTIYLLLV	FLLFLIFKYN	ILAFRYLNLV	ITALVLLVAL	LGLLLIYKK	
Psr								
YvhJ								140
LytR								
CpsA	AEKFTIFLLV	FSILVSSVSL	FAVQQFVGLT	NRLNATSNYS	EYSISVAVLA	DSDIENVTLQ	TSVTAPTGT	
Psr								
YvhJ		MAERVVRVR	KKKKSKRRKI	LKRIMLLFAL	ALLVVVGLGG	YKLYKTINAA	DESYDALSRG	210
LytR		MR	NERRKKKKT	LLTILTIIGL	LVLGTGGYAY	YLWHKAASV	ASIHESIDKS	
CpsA	NENIQKLLAD	IKSSQNTDLT	VNQSSSYLAV	YKSLIAGETK	AIVLNSVFEN	IIESEYPDYA	SKIKKIYTKG	
Psr				MKT	QKVILGLLTA	VVIVIGYFSF	EFVHGYSSTS	
YvhJ	NKSNLRNEVV	DMKKKPF	SILFMGIEDYATK	GQKGRSDSLI	VVTLDPKNKT	MKMLSIPRDT	RVQLAGDTTG	280
LytR	KKRD..KEVS	INKKDPF	SVLIMGVDE..RD	GDKGRADTLI	YMTVNPKNKT	TDMVSIPRDT	YTKIIGKGT.	
CpsA	FTKKV.EAPK	TSKNQSFNIY	VSGIDTYGPI	SSVSRSDVNI	LMTVNRDTKK	ILLTTTPRDA	YVPIADCGNN	
Psr	RADDVPKVEA	KSVPKIINVA	LIGSDA.RSK	EENGSRDSL	IAQYNQKTKK	AKLVSIMRDS	YVDIPGHG..	
YvhJ	SKTKINAAYS	KGGKDET	VETVENFLQIPID	KYVTVDFDGF	KDVINEVGGI	DVDVPFDFDE	KSDVDESKRI	350
LytR	.MDKINHSA	FGGTQMT	VDTVENFLDVPVD	YFVKVNMESE	RDVVDTLGGI	TVNSTFAFSY	DG.....Y	
CpsA	QKDKLTHA.G	IYGVDSIHT	LENLYGVDIN	YVRLNFTSF	LKMIDLLGGV	DVHNDQEF	FS....ALHGKF	
Psr	.QDKINAAYS	YGGIDLLNQT	LKENFKFETP	YYASITFQDF	IDCVNELF..			
YvhJ	YFKKGEMHLN	GEEALAYARM	RKQDKRGDFG	RNDRQKQILN	ALIDRMSSAS	NIAKIDKIAE	KASENVETNI	420
LytR	SFGKGEITLN	GKEALAYTRM	RKEDPRGDFG	RQDRQRQVIQ	GIINKGANIS	SITKFGDMFK	VVENNVKTNL	
CpsA	HFPVGNVHL	SEQALGFVRE	RYSLADGDRD	RGRNQKQVIV	AIQKLTSTE	VLKNYSSILQ	GLQDSLQTNM	
Psr								
YvhJ	RITEGLAL.Q	QIYSGFTSKK	IDTSLITGSD	LYLGP.....	.NNTYFEPD	ATNLEKVRKT	LQEHLDYTPD	490
LytR	TFDNMWDI.Q	SDYKG.ARKH	IKQHELKGTG	TKI.....	.NGIYYYQAD	ESALSDITKE	LKESLEK...	
CpsA	PIETMIDLVN	TQLESGGNYK	VNSQDLKGTG	RMVLPSYAMP	DSNLVMEID	DSSLAVVKAA	IQDVMGR...	
Psr								
YvhJ	TSTGTSGTED	GTDSSSSSGS	TGSTGTTTGD	TNGSSYSND	SSTSSNNSTT	NSTTDSSY	548	
LytR								
CpsA								
Psr								

Fig. 3. Comparison of the YvhJ protein to *Bacillus subtilis* LytR, *Streptococcus pneumoniae* CpsA and *Enterococcus hirae* Psr. Bold letters denote amino acids common to at least three of four proteins. Possible N-glycosylation sites are indicated by arrows.

csrA, could be part of the same transcription unit also. Indeed, the 12 bp spacer between the inverted repeat that follows *flgL* and the ribosome-binding site of *yviE* could hardly accommodate a promoter sequence. Thus, the stem-loop structure within this spacer would more likely correspond to a stabilizer of the mRNA molecule than function as a transcriptional terminator. FlgL shares similarity with several different flagellar proteins, notably with the flagellar hook-associated protein 3 in *E. coli* (Casaregola *et al.*, 1992). The highly hydrophilic YviE and hydrophobic YviF proteins do not display striking similarities with any of the translated nucleotide sequences deposited in GenBank (release 93.0). YviG is a 74 aa homologue of the *E. coli* regulatory factor known to affect glycogen biosynthesis, gluconeogenesis, cell size and surface properties (Romeo *et al.*, 1993), most likely by

accelerating the degradation of *glg* gene transcripts (Liu *et al.*, 1995).

The flagellin *bag* gene, transcribed by the E- σ^D holo-enzyme (Mirel & Chamberlin, 1989) is followed by the third σ^D -controlled tetracistronic operon encoding YviH and flagellar proteins FliD, FliS and FliT (Chen & Helmann, 1994).

A relatively long non-coding region of 580 bp separates *fliT* and *yviI*. The protein encoded by the latter gene is similar to the light-repressed protein LrtA of *Synechococcus* sp. (Tan *et al.*, 1994), the chloroplast-specific 30S ribosomal protein RPS30 of *Spinacia oleracea* (Johnson *et al.*, 1990) and the *Staphylococcus carnosus* YseA protein of unknown function (Klein *et al.*, 1995). The N-terminal

YvhK								70
OrfX								
Ypol								
YigZ.Ec								
YigZ.Hi								
Ycr59c	MDDDEHQLVE	ELEAVEAIYP	DLLSKKQEDG	SIIVVKVPQH	EYMTLQISFP	THYPSEEAPN	VIEVGVCTSL	
YvhK						MLHSY	FTVKEAGEHE	140
OrfX						MS	LTLADKVVEE	
Ypol								
YigZ.Ec						MMESW	LIPAAPVTVV	
YigZ.Hi						MAEY	LVPKSAVVE	
Ycr59c	AKRDLVDTKY	LQHLFQEVMD	SVFHRGSVCL	FDFLTLDGV	LYVEPEEETE	PVQQSDIPTD	PPEGWTASDP	
YvhK	IVIEKSRFIC	HLSRVSTEQE	AQEFIQKIKK	QH..WNATEN	CSAYVIGEND	HIQKAN..DD	GEPSTAGVVP	210
OrfX	EEIQKSRFIA	KAAPVASEEE	ALAFLAENRE	PEATEN	GHAYKIG...	..LLYRFSDD	GEPSTAGRP	
Ypol	.IIQRSRFLA	KAAPAASEEE	ALAFLAHRE	PQATEN	AYAYRIG...	..PLYRFSDD	GEPRA.PGRP	
YigZ.Ec	EEIKKSRFIT	MLAHTDGVFA	AKAFVESVRA	EH..PDAREH	CVAWVAGAPD	DSQQLGFSDD	GEPAGTAGKP	
YigZ.Hi	EEIKKSRFIT	YLQHTEGLED	ARAFWAKIKQ	EH..PNAREH	CWAAVAGKPT	DSLQLGFSDD	GEPAGTAGKP	
Ycr59c	ITDRGSTFMA	FAAHVTSEEQ	AFAMLDLLKT	DSKMRKANEV	MSAWRIKQDG	SAATYQDSDD	DGETA.AGSR	
YvhK	MLEVLKKRRL	KDTCVVTRY	FGGIKLQAGG	LIRAYGKSVS	EGLNHIGVVE	RKLMRIMHTS	ADYTWLGKIE	280
OrfX	ILHAIEAQGL	DRVAVLVVRY	FGGVKLQAGG	LVRAYGGVAA	EALRRAPKVP	LVERVGLAFL	VPFAEVGRVY	
Ypol	ILHAIEAAGL	DRVVVLVVRY	FGGVKLQAGG	LVRAYGGVAA	EALRRAPKAP	LLDWVEVAFR	VPFPEVGRVH	
YigZ.Ec	MLAQLMGSGV	GEITAVVVRY	YGGILLGTGG	LVKAYGGGVN	QALRQLTTQR	KTPLTEYTLQ	CEYHQLTGIE	
YigZ.Hi	MLSALQGSQ	GEISAVVVRY	YGGILLGTGG	LVRAYGNVQ	QALKLIESEI	KVERTLFKLD	CDYGQLRLVQ	
Ycr59c	MLHLITIMDV	WNVIVVVARW	FGGAHIGPDR	FKHINSTARE	AVVRAGFDS.			
YvhK	NELRESQFL	KEIHYSENVE	FETYVEEKET	NAFSEW.MTE	LTNGKSDIKE	GELTYLEKAV	NHIKETE	347
OrfX	ALLEARAPEG	RGDLHPGGVR	FALLLPKPER	EG.FLRALLD	ATRGQVALE.			
Ypol	GLLRARHLEA	EEAYLPEGVR	FALRLPPGGE	GGPCLKALSD	LTRGRLWVEG			
YigZ.Ec	ALLGQCDGKI	INSDYQAFVL	LRVALPAAKV	AEFSAK.LAD	FSRGS..QL	LAIEE.....		
YigZ.Hi	QLCEKYQVEI	LSQGFQANIH	LILGISEKTI	EAFSSE.LTE	KSSGRLVIQL	LEIGE.....		
Ycr59c								

Fig. 4. Sequence alignment of the *yvhK* gene product with the *Thermus thermophilus* incomplete OrfX, *Thermus aquaticus* Ypol, *E. coli* YigZ (YigZ.Ec), *Haemophilus influenzae* YigZ (YigZ.Hi) and *Saccharomyces cerevisiae* Ycr59c. Conserved residues are represented by bold type. Asterisks denote amino acids common to the prokaryotic sub-group only.

Fig. 5. Relatedness of YvhK and homologous proteins. See legend to Fig. 4 for details.

region of YviI is homologous to a family of small regulators of σ^{54} activity in Gram-negative bacteria (Merrick & Coppard, 1989). YviI and homologous proteins have different lengths and their common 27 aa domain occupies different positions with respect to the C and N termini (Fig. 8). The domain corresponding to the N-terminal portion of RPS30, responsible for targeting to chloroplasts, is not present in the other four homologues. This is not surprising, since the latter are of prokaryotic origin. Taking into account the above-mentioned differences between YviI and its homologues, the function(s) of the uncharacterized members of this group might be unrelated to those presently attributed to some of its members. An interesting similarity between two Gram-positive organisms, *B. subtilis* and *Staph. carnosus*, was found at the level of genetic organization, i.e. both *yviI* and its *Staph. carnosus* homologue are followed by *secA* (*div*⁺), a gene responsible for protein secretion and known to affect processes such as cell division, autolysis, competence, sporulation and germination (Sadaie *et al.*, 1991).

DegV	GTELGEALG	ICWCFK..	298
Y326	IVETGIDAFS	ISLLIDNK	
Orf2			
Yp15			

Fig. 6. Alignment of the deduced amino acid sequence of DegV with that of *Mycoplasma genitalium* Y326, *Staphylococcus aureus* chromosomal Orf2 and plasmid pSK1-encoded Yp15. Residues common to DegV and at least one other protein are represented by bold type.

1861 TGTGCAGCATGCTAGCTGACCC⁻³⁵TCTGCTAAGCATAAAAGACTGCCTATACAAATTCGTACAGTCTTTTAGAATTTTGTG
ACACGTCGTACGATCGACTGGGAGGACGATTTCGTATTTTCTGACGGATATGTTTAAAGCATGTCAGAAATCTTAAAAACAC
S H L M yvhK

1941 CGTATTTTGGTATCATAAAGAGTAGATAGTATATAAAAATGTTTTTTTCTAGAATATACGCATTCTTTCATTATAATTCTG
GCATAAAACCATAGTATTTCTCATCTATCATATATTTTTTACAAAAAAGATCTTATATGCGTAAGAAAGTAAATATTAAGC
←⁻¹⁰

2021 ACATAATTTGCAGATCAATTACATTTATAATAAAAAATATATGACAACGCCGTGACGAGGGGAAATTATGAATAAAACAAA
TGTATTAAACGTCTAGTTAATGTAAATATATTTTTATATACTGTTGCGGCAC⁻³⁵TGCCTCCCTTTAATACTTTATTTTGTTT
degS M N K T K

Fig. 7. Nucleotide sequence of the *yvhK-degS* intergenic region. – 10 and –35 promoter regions are underlined. Directions of transcription from the 5' end of *degSU* (Msadek *et al.*, 1990) and *yvhKJ* (V. Lazarevic, unpublished) are denoted by arrows. Shine–Dalgarno sequences are double-underlined.

REFERENCES

- Akhmetzjanov, A. A. & Vakhitov, V. A. (1992).** Molecular cloning and nucleotide sequence of the DNA polymerase gene from *Thermus flavus*. *Nucleic Acids Res* **20**, 5839.
- Asakura, K., Komatsubara, H., Soga, S., Yomo, T., Oka, M., Emi, S. & Urabe, I. (1993).** Cloning, nucleotide sequence, and expression in *Escherichia coli* of DNA polymerase gene (*polA*) from *Thermus thermophilus* hb8. *J Ferment Bioenerg* **76**, 265–269.
- Beck, C. F. & Warren, R. A. J. (1988).** Divergent promoters, a common form of gene organization. *Microbiol Rev* **52**, 318–326.
- Casaregola, S., Jacq, A., Laoudj, D., McGurk, G., Margaron, S., Tempete, M., Norris, V. & Holland, I. B. (1992).** Cloning and

analysis of the entire *Escherichia coli* *ams* gene. *ams* is identical to *bmp1* and encodes a 114 kDa protein that migrates as a 180 kDa protein. *J Mol Biol* **228**, 30–40.

- Chambers, S. P., Prior, S. E., Barstow, D. A. & Minton, N. P. (1988).** The pMTL *nic*⁻ cloning vectors. I. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. *Gene* **68**, 139–149.
- Chen, L. & Helmann, J. D. (1994).** The *Bacillus subtilis* σ^D -dependent operon encoding the flagellar proteins FliD, FliS and FliT. *J Bacteriol* **176**, 3093–3101.
- Chung, C. T. & Miller, R. H. (1988).** A rapid and convenient method for the preparation and storage of competent bacterial cells. *Nucleic Acids Res* **16**, 3580.

YviI								70
YseA								
LrtA								
RPS30	MATLCTSAIN	MNPNLTNSLS	NSINLSSTPT	NLSSLRSTFT	NSCSLGLNVA	VKSVQISRNK	PNVVCMSWDG	
Rp5M								
YviI	MNYNIR	GENIEVTPAL	KDHVERKIGK	LERYFDHSVD	ADVNVNLKFY	NDKESKVE..	VTIPM	140
YseA								
LrtA	MKLLIQ	GNNIAVTESI	HDYVESKLEK	ATKHFQT.FA	TKVDVHLS..	VANNARITDK	HKA EVT VYAN	
RPS30	PLSSVKLILQ	GRNLEVSDNV	RSHVEDKVVK	SVAKHSH.LV	REVDVRLSAR	GGDL SKGP KL	RRC E V T L F T K	
Rp5M	MQLNIT	GHNVEITPAM	RDFVTAKFSK	LEQFFDRINQ	VYIVL.....	KVEKV	TQIADANLHV	
YviI	TDLALRSEVH	NEDMYNAIDL	ATNKLERQIR	KHKTQVNRKF	REQGSPKYLL	ANGLGSDTDI		210
YseA		..MIYAGIDK	ITNKLECQVR	KYKTRVNRKK	RKESEHEPFP	ATP.ETPPET		
LrtA	.GTVIRAQEG	SENLYASIDL	VSDKIAQLR	KYKEKNFGKK	THVQE.KTSE	VLPEDPVPD.	NSLRS	
RPS30	RHGVIKAEED	AESLYSSIDL	VSSIIQRKLR	KIKDKVSDHG	RHMKGFNRSK	VRDPEPVRIT	REEVLEEVE	
Rp5M	NGGEIHASAE	GQDMYAAIDG	LIDKLARQLT	KHKDKLKQH.				
YviI	..AVQDDIEE	EESLDIVRQK	RFNLKPMDS	EAILQMNMLG	HNFFVFTNAE	TNLTNVVYRR	NDGKYGLIEP	280
YseA	..AVDHDKDD	E..IEIIRSK	QFSLKPMDS	EAVLQMDLLG	TDFFI FNDRE	TDGTSIVYRR	KDGKYGLIET	
LrtA	RPRASS....	...EVVRMK	YFAMPPTID	QALEQLQLVD	HDFYMFNLKD	TNAINVIYIR	NHGGYGVIIQ	
RPS30	APAPVSVEDD	DFIEEVVTRK	YFDMPTLTIT	EAVEQLENVD	HDFYAFRNEE	TGDINILYKR	KEGGYGLIIP	
Rp5M								
YviI	TE.....							303
YseA	VEKLICDI..							
LrtA	RLGKE.....							
RPS30	KDGKTEKLES	LPVQTDKQPS	FAE					
Rp5M								

Fig. 8. Alignment of *B. subtilis* YviI with *Staphylococcus carnosus* YseA, *Synechococcus* sp. LrtA, *Spinacia oleracea* RPS30 and *Klebsiella pneumoniae* Rp5M as a representative of the σ^{54} modulation protein family.

Dale, G. E., Then, R. L. & Stuber, D. (1993). Characterization of the gene for chromosomal trimethoprim-sensitive dihydrofolate reductase of *Staphylococcus aureus* ATCC 25923. *Antimicrob Agents Chemother* **37**, 1400–1405.

Del Sal, G., Manfioletti, G. & Schneider, C. (1988). A one-tube plasmid DNA mini-preparation suitable for sequencing. *Nucleic Acids Res* **16**, 9878.

Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* **12**, 387–395.

Fahmy, F., Flossdorf, J. & Claus, D. (1985). The DNA base composition of the type strains of the genus *Bacillus*. *Syst Appl Microbiol* **6**, 60–65.

Feng, D. & Doolittle, R. (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J Mol Evol* **25**, 351–360.

Fleischmann, R. D. and others (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512.

Fraser, C. M. and others (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397–403.

Gillen, K. L. & Hughes, K. T. (1991). Molecular characterization of *flgM*, a gene encoding a negative regulator of flagellin synthesis in *Salmonella typhimurium*. *J Bacteriol* **173**, 6453–6459.

Glaser, P., Kunst, F., Arnaud, M., Coudart, M.-P., Gonzales, W., Hullo, M.-F., Ionescu, M., Lubochinsky, B., Marcelino, L., Moszer, I., Presecan, E., Santana, M., Schneider, E., Schweizer, J., Vertes, A., Rapoport, G. & Danchin, A. (1993). *Bacillus subtilis* genome

project: cloning and sequencing of the 97 kb region from 325° to 333°. *Mol Microbiol* **10**, 371–384.

Grossberger, D. (1987). Minipreps of DNA from bacteriophage lambda. *Nucleic Acids Res* **15**, 6737.

Guidolin, A., Morona, J. K., Morona, R., Hansman, D. & Paton, J. C. (1994). Nucleotide sequence analysis of genes essential for capsular polysaccharide biosynthesis in *Streptococcus pneumoniae* type 19F. *Infect Immun* **62**, 5384–5396.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**, 557–580.

Henner, D. J., Yang, M. & Ferrari, E. (1988). Localization of *Bacillus subtilis* *sacU*(Hy) mutations of two linked genes with similarities to the conserved procaryotic family of two-component signalling systems. *J Bacteriol* **170**, 5102–5109.

Homma, M., DeRosier, D. J. & Macnab, R. M. (1990). Flagellar hook and hook-associated proteins of *Salmonella typhimurium* and their relationship to other axial components of the flagellum. *J Mol Biol* **213**, 819–832.

Johnson, C. H., Kruft, V. & Subramanian, A. R. (1990). Identification of a plastid-specific ribosomal protein in the 30S subunit of chloroplast ribosomes and isolation of the cDNA clone encoding its cytoplasmic precursor. *J Biol Chem* **265**, 12790–12795.

Karamata, D. & Gross, J. D. (1970). Isolation and genetic analysis of temperature sensitive mutants of *B. subtilis* defective in DNA synthesis. *Mol Gen Genet* **108**, 277–287.

Klein, M., Meens, J. & Freudl, R. (1995). Functional characterization of the *Staphylococcus carnosus* SecA protein in *Escherichia coli* and

- Bacillus subtilis* *secA* mutant strains. *FEMS Microbiol Lett* **131**, 271–277.
- Koonin, E. V., Bork, P. & Sander, C. (1994). Yeast chromosome III: new gene functions. *EMBO J* **13**, 493–503.
- Kunst, F., Débarbouillé, M., Msadek, T., Young, M., Mauël, C., Karamata, D., Klier, A., Rapoport, G. & Dedonder, R. (1988). Deduced polypeptides encoded by the *Bacillus subtilis* *sacU* locus share homology with two-component sensor-regulator systems. *J Bacteriol* **170**, 5093–5101.
- LaVallie, E. R. & Stahl, M. L. (1989). Cloning of the flagellin gene from *Bacillus subtilis* and complementation studies of an *in vitro*-derived deletion mutation. *J Bacteriol* **171**, 3085–3094.
- Lazarevic, V., Margot, P., Suldo, B. & Karamata, D. (1992). Sequencing and analysis of the *Bacillus subtilis* *lytRABC* divergon: a regulatory unit encompassing the structural genes of the *N*-acetylmuramoyl-L-alanine amidase and its modifier. *J Gen Microbiol* **138**, 1949–1961.
- Lazarevic, V., Mauël, C., Soldo, B., Freymond, P.-P., Margot, P. & Karamata, D. (1995). Sequence analysis of the 308° to 311° segment of the *Bacillus subtilis* 168 chromosome, a region devoted to cell wall metabolism, containing non-coding grey holes which reveal chromosomal rearrangements. *Microbiology* **141**, 329–335.
- Ligozzi, M., Pittaluga, F. & Fontana, R. (1993). Identification of a genetic element (*psr*) which negatively controls expression of *Enterococcus hirae* penicillin-binding protein 5. *J Bacteriol* **175**, 2046–2051.
- Liu, M. Y., Yang, H. & Romeo, T. (1995). The product of the pleiotropic *Escherichia coli* gene *csrA* modulates glycogen biosynthesis via effects on mRNA stability. *J Bacteriol* **177**, 2663–2672.
- Londoño-Vallejo, J. A. & Dubnau, D. (1993). *comF*, a *Bacillus subtilis* late competence locus, encodes a protein similar to ATP-dependent RNA/DNA helicases. *Mol Microbiol* **9**, 119–131.
- Margot, P., Mauël, C. & Karamata, D. (1994). The gene of the *N*-acetylglucosaminidase, a *Bacillus subtilis* 168 cell wall hydrolase not involved in vegetative cell autolysis. *Mol Microbiol* **12**, 535–545.
- Márquez, L. M., Helmann, J. D., Ferrari, E., Parker, H. M., Ordal, G. W. & Chamberlin, M. J. (1990). Studies of σ^D -dependent functions in *Bacillus subtilis*. *J Bacteriol* **172**, 3435–3443.
- Mauël, C., Young, M., Margot, P. & Karamata, D. (1989). The essential nature of teichoic acids in *Bacillus subtilis* as revealed by insertional mutagenesis. *Mol Gen Genet* **215**, 388–394.
- Merrick, M. J. & Coppard, J. R. (1989). Mutations in genes downstream of the *rpoN* gene (encoding σ^{54}) of *Klebsiella pneumoniae* affect expression from σ^{54} -dependent promoters. *Mol Microbiol* **3**, 1765–1775.
- Mirel, D. B. & Chamberlin, M. J. (1989). The *Bacillus subtilis* flagellin gene (*bag*) is transcribed by the σ^{28} form of RNA polymerase. *J Bacteriol* **171**, 3095–3101.
- Mirel, D. B., Lauer, P. & Chamberlin, M. J. (1994). Identification of flagellar synthesis regulatory and structural genes in a σ^D -dependent operon of *Bacillus subtilis*. *J Bacteriol* **176**, 4492–4500.
- Msadek, T., Kunst, F., Henner, D., Klier, A., Rapoport, G. & Dedonder, R. (1990). Signal transduction pathway controlling synthesis of a class of degradative enzymes in *Bacillus subtilis*: expression of the regulatory genes and analysis of mutations in *degS* and *degU*. *J Bacteriol* **172**, 824–834.
- Nakahigashi, K. & Inokuchi, H. (1990). Nucleotide sequence between the *fadB* gene and the *rrnA* operon from *Escherichia coli*. *Nucleic Acids Res* **18**, 6439.
- Roggiani, M., Hahn, J. & Dubnau, D. (1990). Suppression of early competence mutations in *Bacillus subtilis* by *mec* mutations. *J Bacteriol* **172**, 4056–4063.
- Romeo, T., Gong, M., Liu, M. Y. & Brun-Zinkernagel, A.-M. (1993). Identification and molecular characterization of *csrA*, a pleiotropic gene from *Escherichia coli* that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. *J Bacteriol* **175**, 4744–4755.
- Rouch, D. A., Messerotti, L. J., Loo, L. S., Jackson, C. A. & Skurray, R. A. (1989). Trimethoprim resistance transposon Tn4003 from *Staphylococcus aureus* encodes genes for a dihydrofolate reductase and thymidylate synthetase flanked by three copies of IS257. *Mol Microbiol* **3**, 161–175.
- Sadaie, Y., Takamatsu, H., Nakamura, K. & Yamane, K. (1991). Sequencing reveals similarity of the wild-type *div⁺* gene of *Bacillus subtilis* to the *Escherichia coli* *secA* gene. *Gene* **98**, 101–105.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**, 5463–5467.
- 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. M. Zukowski, A. T. Ganesan & J. A. Hoch. San Diego: Academic Press.
- Tan, X., Varughese, M. & Widger, W. R. (1994). A light-repressed transcript found in *Synechococcus* PC 7002 is similar to a chloroplast-specific small subunit ribosomal protein and to a transcription modulator protein associated with sigma 54. *J Biol Chem* **269**, 20905–20912.
- Tanaka, T. & Kawata, M. (1988). Cloning and characterization of *Bacillus subtilis* *iep*, which has positive and negative effects on production of extracellular proteases. *J Bacteriol* **170**, 3593–3600.
- Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119.

Received 8 May 1996; revised 28 June 1996; accepted 10 July 1996.