Determination of a 12 kb nucleotide sequence around the 76° region of the *Bacillus subtilis* chromosome

Hiroki Yamamoto,¹ Shigeki Uchiyama,¹ Aji N. Fajar,¹
Naotake Ogasawara² and Junichi Sekiguchi¹

Author for correspondence: Junichi Sekiguchi. Tel: +81 268 22 1215. Fax: +81 268 22 4079.
e-mail: jsekigu@gipsc.shinshu-u.ac.jp

¹ Department of Applied Biology, Faculty of Textile Science and Technology, Shinshu University, 3-15-1 Tokida, Ueda-shi, Nagano 386, Japan

² Graduate School of Biological Sciences, Nara Institute of Science and Technology, Ikoma, Nara 630-01, Japan

The nucleotide sequence of a 12361 bp DNA segment in the 76° region of the *Bacillus subtilis* chromosome has been determined. Ten putative ORFs were identified. The deduced amino acid sequences of the products of two of them (*glv-1* and *glv-2*) exhibited high similarity to those of *glvG* (6-phospho-β-glucosidase gene) and *glvC* [permease (the IIC domain) of the phosphotransferase system (PTS)], respectively, in the *glv* operon of *Escherichia coli*. The C-terminal region of Glv-2 exhibited similarity to the entire region of GlvB (the IIB domain of PTS) of *E. coli*, suggesting fusion of the *glvC* and *glvB* genes in *B. subtilis*. *glv-1*, *yfiA* and *glv-2* seem to form an operon of a phosphoenolpyruvate:sugar PTS, followed by a presumed four-membered operon of an ABC transport system. Moreover, a presumed sugar symporter and its regulatory genes were located in this region.

Keywords: *Bacillus subtilis*, genome sequencing, 76° region

INTRODUCTION

As participants in the international *Bacillus subtilis* genome cooperative project, we are responsible for the sequencing of an approximately 120 kb chromosomal region between *hsdC* and *sspE*. Since very few genes in this region have been cloned, we used as a probe the pNEXT23 plasmid (Itaya & Tanaka, 1991), which is a linking clone for the *NotI* site between *hsdC* and *sspE*. The physical map of the *B. subtilis* chromosome constructed by Itaya & Tanaka (1991) indicates that the map positions for *hsdC* and *sspE* are 69° and 81° instead of 58° and 63°, respectively, as reported by Anagnostopoulos *et al.* (1993).

METHODS

Bacterial strains, phages and plasmids. *B. subtilis* strain AC327 (*purB his-1 smo-1*) (Akamatsu & Sekiguchi, 1987) was used as the source of chromosomal DNA to construct a λ library. The fragment patterns of *SfiI*- and/or *NotI*-digested DNA of this strain were identical to those of strain 168 on PFGE (M. Itaya, personal communication). λ DASH II was used to construct a library after partial digestion of the DNA with *Sau3AI*.

Abbreviations: PTS, phosphotransferase system; SD, Shine–Dalgarno.

The GSDB, DDBJ, EMBL and NCBI accession number for the nucleotide sequence reported in this paper is D50543.

Escherichia coli XL1-Blue MRA (P2) (Stratagene) was used for propagation of the λ library. Clones were selected by plaque hybridization (Sambrook *et al.*, 1989) with ³²P-labelled pNEXT23 DNA as a probe. The insert in the selected clone was partially digested with DNase I, blunt-ended with T4 polymerase and the fragments were separated by agarose (1.2%, w/v) gel electrophoresis to obtain 700–1500 bp fragments. They were ligated to *SmaI*-digested and alkaline-phosphatase-treated pUC118 DNA. *E. coli* JM109 cells were transformed with the ligation mixture to obtain a randomly overlapping library for sequencing. Transformant cells were boiled in water and transferred to the PCR reaction mixture containing forward and reverse primers of pUC118, *Taq* polymerase (Applied Biosystems) and dNTP. After a small-scale PCR reaction, amplified DNAs, ranging from 700 to 1500 bp, were selected for sequencing. After large-scale amplification, the primers were removed with Microcon (Amicon).

DNA sequencing. All sequences were determined using a *Taq* Dye Primer Cycle Sequencing Kit and a 373A sequencer from Applied Biosystems. We determined the sequences of both DNA strands.

Computer analysis. The DNA sequences obtained were compiled using an Inherit Auto Assembler (Applied Biosystems) and further analysed for the location of possible ORFs using GENWORKS (IntelliGenetics). The amino acid sequences of the putative products of identified ORFs were examined for similarity to sequences reported previously in a non-redundant protein sequence data bank using the BLAST e-mail server at the Human Genome Center, Institute of Medical Science, Uni-

versity of Tokyo, Japan. According to the nomenclature proposed by A. Danchin (unpublished, Institut Pasteur, Paris, France), we used *yfi* for ORFs of unknown function.

RESULTS AND DISCUSSION

Cloning of the pNEXT23 region (76°) of the *B. subtilis* chromosome

Clones containing the pNEXT23 region were isolated by plaque hybridization. Among the 18 independent clones isolated, four were selected and their fragments were analysed with restriction enzymes. One clone (clone 4-1) containing a 14.5 kb insert with a common 9.4 kb *Pst*I fragment was selected and the inserted fragment was digested with DNaseI, followed by ligation to blunt-

ended pUC118, to construct a random library for sequencing. The inserted fragments were amplified by PCR from lysates of transformants and sequenced as described in Methods.

Putative ORFs were searched for all six possible translation frames. We considered valid the largest possible ORF, starting with ATG, GTG or TTG, which is preceded by a Shine–Dalgarno (SD) sequence complementary to the 3' end of the 16S rRNA of *B. subtilis* (3' UCUUUCUCCACUAG-). Based on this criterion, 10 putative ORFs were identified, as listed in Table 1. Among them, nine start with ATG and one with GTG. All 10 ORFs exhibit the same gene orientation (Fig. 1) and the *sspE* gene is located further downstream of *yfiH* (H. Yamamoto *et al.*, unpublished results). Therefore the

Table 1. Putative ORFs in the 12361 bp sequence near the 76° region

ORF	Endpoints (nucleotides)*	Size of product (aa/kDa)	SD consensus sequence (upper case) and initiation codon (bold)
<i>glv-1</i>	386 > 1732	449/50.5	AtcAAGGAGGTcgTCcatatg
<i>yfiA</i>	1800 > 2561	254/29.3	gaAAAtGgGGgGATCtgatatg
<i>glv-2</i>	2579 > 4159	527/58.1	tagcAGGgGGTtcTggtatg
<i>yfiB</i>	4268 > 5986	573/63.9	gGtttGGAGGcttTtcatatg
<i>yfiC</i>	5983 > 7794	604/67.4	gGAAGGGAGGgGAgCgagtcatg
<i>yfiD</i>	7952 > 8353	134/14.4	AtcAAGGAGGatAaaatg
<i>yfiE</i>	8374 > 9228	285/31.5	AGAAAaGAGGcatgCagcatg
<i>yfiF</i>	9325 > 10266	314/36.2	tGttAGGAGGcGtTtcatatg
<i>yfiG</i>	10444 > 11889	482/52.7	tGAAAGaAGGgGAagaaagccgtg
<i>yfiH</i>	11919 > 12361 →		taccAGGAGGaatctatcaatg

* > indicates the transcriptional direction of the ORF. The arrow indicates a partial ORF.

Fig. 1. Physical map and gene organization near the 76° region. A physical map of the sequenced 12.4 kb with respect to the *Not*I site (76°) is shown. E and P denote *Eco*RI and *Pst*I sites, respectively. In the lower part of the figure, the deduced 10 ORFs and their transcriptional and translational direction are shown. The direction corresponds to that of the movement of the replication fork. The deduced p-independent terminators and their free energy are also shown. In the case of *yfiH*, only the 5' region is presented, as a bold arrow.

Table 2. Similarity of predicted ORF products to known proteins

Product	Similar protein in database	Identity observed	BLAST score
Glv-1	Probable 6-phospho- β -glucosidase (<i>glvG</i>) of <i>E. coli</i>	79.3% in the N-terminal 198 residues	846
	6-phospho- β -glucosidase (<i>celF</i>) of <i>E. coli</i>	34.6% in the N-terminal 162 residues	291
YfiA	Hypothetical protein in <i>purL</i> 5' region (fragment) of <i>E. coli</i>	35.9% in the N-terminal 64 residues	108
Glv-2	PTS system, arbutin-like IIC component (<i>glvC</i>) of <i>E. coli</i>	66.3% in the N-terminal 368 residues	1165
	PTS system, arbutin-like IIB component (<i>glvB</i>) of <i>E. coli</i>	47.7% in the C-terminal 151 residues	207
YfiB	ABC transporter (<i>ywjA</i>) of <i>B. subtilis</i>	30.3% in 571 residues	416
	Probable transport ATP-binding protein (<i>msbA</i>) of <i>E. coli</i>	31.2% in 558 residues	416
	Multidrug resistance-like ATP-binding protein of <i>E. coli</i>	26.5% in 517 residues	473
	Haemolysin secretion ATP-binding protein (<i>blyB</i>) of <i>E. coli</i>	26.6% in 553 residues	442
YfiC	ABC transporter (<i>ywjA</i>) of <i>B. subtilis</i>	33.9% in 570 residues	905
	Probable transport ATP-binding protein (<i>msbA</i>) of <i>E. coli</i>	31.4% in 580 residues	520
YfiD	<i>E. coli</i> chromosomal region from 67.4 to 76.0 min	32.8% in the N-terminal 76 residues	129
YfiF	Multiple antibiotic resistance protein (<i>marA</i>) of <i>E. coli</i>	35.2% in the C-terminal 91 residues	120
	L-Rhamnose operon transcriptional activator (<i>rhaR</i>) of <i>E. coli</i>	21.1% in 247 residues	82
YfiG	Galactose-proton symport protein (<i>galP</i>) of <i>E. coli</i>	36.4% in 461 residues	237
	Arabinose-proton symport protein (<i>araE</i>) of <i>E. coli</i>	34.7% in 470 residues	221
	Xylose-proton symport protein (<i>xylE</i>) of <i>E. coli</i>	32.3% in 482 residues	190

orientation of the transcription and translation of the 10 ORFs must be identical to that of the movement of the replication fork.

Genes found in the 12 kb region and their functions (Table 2)

orf1. The deduced aa sequence of the product of *orf1* exhibits extremely high similarity to that of *E. coli glvG* (SP: glvg_ecoli, 212 aa), which is probably a 6-phospho- β -glucosidase (EC 3.2.1.86) gene (Reizer *et al.*, 1994). High similarity was also found with the *E. coli celF* gene, whose product is also a 6-phospho- β -glucosidase (SP: celF_ecoli, 372 aa) (Parker & Hall, 1990). *orf1* was named *glv-1*.

yfiA. The aa sequence of YfiA exhibited relatively high similarity to that of a hypothetical protein in the *purL* 5' region of the *E. coli* chromosome (SP: yfhh_ecoli, 91 aa) (Poulsen *et al.*, 1992).

orf2. The N-terminal aa sequence of Orf2 exhibited extremely high similarity (66.3% identity over 368 aa residues) to that of the product of the *E. coli glvC* gene, which is a permease (the IIC domain) of the phosphotransferase system (PTS) (SP: ptic_ecoli, 368 aa), and its C-terminal sequence to the product of the *glvB* gene, which is the IIB domain of the PTS function (SP: ptib_ecoli, 161 aa) (Reizer *et al.*, 1994). In *E. coli*, *glvC*, *glvB* and *glvG* seem to form an operon, in that order. Since the N-terminal and C-terminal regions of Orf2 exhibit sequence similarities to the entire *glvC* and *glvB* proteins, respectively (Fig. 2), it is very likely that these two proteins (PTS IIC and IIB domains) have been fused in *B. subtilis*,

GlvC	1:	MLSQIQRFPGGAMFTFVLLFFPAGIVVGLAILLQNPMPUGESLTDPNLSLFAQIVHIEBGG
		*.*****.*****.*****.*****.*****.*****.*****.*****.*****.
Glv-2	1:	MMQKIQRFPGSAMFVPLLFAFAGIIVGISTLFRNKTLNG-PLADPDGFWYQWYIIEBGG
GlvC	61:	WTVFRNMLPIFAVGLPIGLAKQAQGRACLAVMSFLTWYFTINAMGTMGSGYFGVDFTQD
		****.***.***.***.***.***.***.***.***.***.***.***.***.***.***.***.
Glv-2	60:	WTVFNQMPILLFAIGIPVALAKKAQARACLEALTLYLTNYFVSAILTVWGGAFGVDMNQE
GlvC	121:	AVAGSLTMMAGIKTLDTSIIGAIISIGVITLHNRLDKKLPVFLGIFQGSTSYVVIAP
		*****.*****.*****.*****.*****.*****.*****.*****.*****.
Glv-2	120:	VGGTSLTMMIAGIKTLDTINIIGAIPISSIVVFLHNRYFDKKLPDFLGIQGSTYIVMISF
GlvC	181:	LVMIPCAWLTLLGMPKVMQGIESTLQAFIRSAGALGVWYTFLEIRILPTGLHPIYQGGI
		*****.*****.*****.*****.*****.*****.*****.*****.*****.
Glv-2	180:	FIMIPIALAVSYIWMVQSGIGSLQSPVAVSGAVGWYITFLERILPTGLHPIYITPFI
GlvC	241:	PGPAVEGGIQMYAQHLQEFSLSAEPLKSLFPEGGFALHNSKIPGAVGILAMVYFTAA
		*****.*****.*****.*****.*****.*****.*****.*****.*****.
Glv-2	240:	YGPAAVEGGIVTYWAQHLGEYSQSAKPLKELFPQGGFALHNSKIPGIPGIALAFVYTA
GlvC	301:	PENRVKVGALLIPATLTAMLVGITEPLEFTFLFISPLLFAVHAVLAASMTVMYLVGVVG
		*.*****.*****.*****.*****.*****.*****.*****.*****.*****.
Glv-2	300:	KEKKLVAGLLIPVTLTAIVAGITEPIETFLFISPLLFAVHAVLAATMSTVMYVAGVVG
GlvC	361:	NMGGLID
		*****.
Glv-2	360:	NMGGLIEAVTLAWIPLFGSHGTYVYQILIGLSFTAYFFVFRFLILKFNIAFPGREKD
		*****.*****.*****.*****.*****.*****.*****.*****.*****.
GlvB	1:	MFSNHADMILTQIAGLCFTLLYFVVFRTLLQFNCTPGR---
Glv-2	421:	EQQETKLYSKKEYRERKNDKDETAASAEATDAPLYIEALGGKDNITEVNCATRLRVSV
		*****.*****.*****.*****.*****.*****.*****.*****.*****.
GlvB	43:	EDAENVKLYSKAEYKASRGQTAAEPKKELDQAAGI-LQALGGVGNISSINNCATRLRIAL
Glv-2	481:	KDETKEPDSVFRALGAHGVRNGKAFQVIIGLSVPQMRERVEKILAQ
		*.*****.*****.*****.*****.*****.*****.*****.*****.*****.
GlvB	102:	HMSQTLDDDEVFKLGAHGVPFRSGDAIQVIIGLHVSQLRQQLDSLINSQSAENVATTEA

Fig. 2. Alignment of the amino acid sequences of Glv-2, *E. coli* GlvC (SP: ptic_ecoli) and *E. coli* GlvB (SP: ptib_ecoli). Numbers are the positions with respect to the N-terminal aa of each protein. Identical and similar aa are indicated by asterisks and dots, respectively.

probably without loss of either function. Fusion of the PTS IIC and IIB domains was also found in several other proteins (Saier & Reizer, 1994; Reizer *et al.*, 1994), e.g. *E.*

YfiB	1:MQKAL--SF-L-K-PYSLLAGIALILML--T--EL-AV-E-LMQP--LLI-A--K-IIDQ
YfiC	1:L-DIRKP.QYP.L-IDKKE.AKKAKAKD.KGT.RRIMSY.AERKG...LVMLMV.SA
YwjA	1:-----M-RQP-PSYTKPKTLF-F-D-PFSA--G-GL-ME-
MsbA	1:-----MHNDKDL.S.WQTFRLNPTIAPFKAG.V.GVAL.LNA
YfiB	43:--GILKQ-DL-RHV--WIMQ-TV--MI-G-LTVLS-F-AAGM-L--NSFYAAHVQSQSPSY
YfiC	61:IF.L.GPFIKKAIDHF.V.K.SGL.PVL.LL.AIVTIQSL.S.WFQNYWMITI..GTVF
YwjA	31:LS--FPL-IVNYFIDTLLP.RDWGLI.ATSIGLPAVLAASSALQYIVTVWGHMLGINIET
MsbA	39:ASDTFMLS.L.KPLDDGFGKTRDSVLVWMP.V.IGLMILR.ITSYVS.YCISN..GKVV
YfiB	87:DTRKGLFQKIQSFYSTFGQFSSSSYITRLINDVTVQNMIFMSLRFLMRAPLMIAGGIV
YfiC	121:RM.SE..THLHELPIFF.DKQRHGLMS.V...IEN.SSTLNT.VIQI.SSVITFV.T.A
YwjA	88:M..S..DHL.KL.FKPYDNKKTOTLMSK...LMYIGEVAHNGPREDLFI.VMT.L.APG
MsbA	99:TM.RR..GHRMGMPV.F.DKQ.TGTLLS.L.Y.SE..ASSSSGA.ITVV.EGAS.I.LFI
YfiB	147:LSLAVNVKLGFFLLVITPILIFLLVLRKGGALFRSVQKRLDQVNTIMQE-NLIATKLI
YfiC	181:VM.YMSP.LTIT.TI..VMAAS.K.ITNRT.K..KEQ..N.GDL.GYIE.-SVSGA.V.
YwjA	148:VM.FI.WQ.ALLTPIIM..V.WLA.-YFN.KMTKAPTILAKDIGDFSARV.N.IGG.R.I
MsbA	159:MMPTYSWQ.SII..IVLAPVTSIAIR.VS.RFRNISQNMQNTMGQV.TSA.QM.KGH.EV
YfiB	206:KALLRGSHVEIKRIFKANTRIMKETSAPQVLEFTMPVLMMLNLCILLIMWAGS-YSITS
YfiC	240:..YS.EKQITAE.LEK.AA.KTSGFW.QTISG.IPK.MNS.N..SFTM.AAI.GLVAL-K
YwjA	207:Q.FQNEAF.KE..AVN.Q.FRVIKL.SYKIMAKNGISISM.TRFVT.FV.LC.TMFV.RG
MsbA	218:LIFGGQEV.T...D.VSN.MRLQGMVSSASSISD.IIQ.IAS.ALAPV.Y.A.FP.VMD
YfiB	265:GSTQGVDDVVAIINYATRITGALSMFFLLIMIFTRAKASGDRIGEV-ETEGDEREBOGIS
YfiC	299:WIST.SI.VPAE.SRQF.RP.NDLANQFNTILS.I.GAE.VFD..D.K.ER.D.KNAVH
YwjA	267:SLSYGEF...FVLLTNVLFRRPIDKINAI.EMYP.GI.GPKSYM.LMETEFDIQSDPSKQD
MsbA	278:SL.ATGTT.VFSSMI-ALMRP.KSLTNVNAQ.Q.QM.ACQTLPTI.DSEQE-KD.-.KRV
YfiB	324:DRL-SGRIFQHVSPRYEMDREALRNVSFSAPKPRETIAILGATGSGKSTLFLQIPRLQ
YfiC	359:QPIQT.S...RD...G.-DKSQQT.KHLQ.TVPAGQS..FV.P..A..T.VIN.LA.F.E
YwjA	326:VSGLK.N.RYK...G.DDHH-NV.NDINL.IQAG..V.FV.PS.A...CS.L..F.E
MsbA	335:IERAT.DV..RN.T.T..GR.VP...INLKIPAGK.V.LV.RS...IAS..T.F.D
YfiB	383:PDSGRYIYDEKPVQDIPABGLRRQIGYVQPEVLLFSGTIKENIANGKENA-SLDEIMDAA
YfiC	418:ND.K.L..GTDIKTLTRAS..KNM.F.L.DSF..Q...R..RY.RLD..DQ.VEA..
YwjA	385:ASE.D.T.GISIK.MTLSS..G...V.Q.D.F...LR...Y.RLG.-EED.WQ.V
MsbA	395:I.E.E.LM.GHDLREYTLAS..N.VAL.S.N.H..ND.VAN..YARTEQY.REQ.EE..
YfiB	442:KLAQIHETILKLPNGYDVTVLQQRGNVLSGGQKQRISTARALIRKPAILLDDSTSLDLQ
YfiC	477:T.NA.SP.ER..K...T.N.SGI.Q...L...VLAD.VL.I..EA..NI.TV
YwjA	444:Q.HLE.LVHNM.D.L..MI.E..K.....L...MFLKN.S..I..EA...TE
MsbA	455:RM.YAMDF.N.MD...I.EN..L...R...A...L.DSP..I..EA...TE
YfiB	502:TEAKLLEAISTY-HC-TTLITQKITTAMKADQILLLEDGELIEKGTSHSELLSESQLYKR
YfiC	537:..VNIQ..LARLMGR.SV..AHLN.IQR...VV.KN..M...S.D..IRQGF.SD
YwjA	504:..AIQK.LQELSEGR...V.AHLA.IKD..R.VVVTNNGIE.Q.R.QD.IEAGG..S.
MsbA	515:S.RAIQA.LDELQNR.S.V.AHRLS.IE...E.VVV...VIV.R...ND..EHRGV.AQ
YfiB	560:IYBSQFGRGSESC
YfiC	597:L....EK-----
YwjA	564:LHQA..-GQMVHR-
MsbA	575:LHKM..-GQ-----

Fig. 3. Alignment of the deduced amino acid sequences of YfiB, YfiC, *B. subtilis* YwjA (gp: bsdna320d_2) and *E. coli* MsbA (SP: msba_ecoli). Numbers are the positions with respect to the N-terminal aa of each protein. Dots and dashes indicate identical aa to those of YfiB and the introduction of gaps in the alignment, respectively. Asterisks indicate aa common to the four proteins.

coli IIBCfrv (Plunkett *et al.*, 1993), *Xanthomonas campestris* FruII (De Crécy-Lagard *et al.*, 1991), *Rhodobacter capsulatus* FruII (Wu & Saier, 1990) and *E. coli* FruII (Prior & Kornberg, 1988). The C-terminal aa sequence of Orf2 also exhibited similarity (30.9% identity over 68 aa residues) to the N-terminal aa sequence of a member of the PTS of *B. subtilis*, BglP (SP: ptba_bacsu, 609 aa) (Yoshida *et al.*, 1995; Le Coq *et al.*, 1995). *orf2* was named *glv-2*.

Regarding gene organization, a gene encoding an enzyme IIA is lacking in the *glv* operon of *E. coli*. In *B. subtilis*, it

is unlikely that the *yfiA* product corresponds to the enzyme IIA domain. Therefore, an enzyme IIA-encoding gene may also be lacking in the deduced *glv* operon of *B. subtilis*.

yfiB. The aa sequence of YfiB exhibited high similarity to that of the product of *E. coli* *msbA*, which is probably a transport ATP-binding protein (ABC transporter; SP: mdLecoli, 1143 aa) (Allikmets *et al.*, 1993). The sequence of YfiB also exhibited high similarity to many ABC transporters, e.g. the haemolysin secretion ATP-binding protein (SP: hlyb_ecoli, 707 aa) (Felmlee *et al.*, 1985), multidrug resistance protein 1 (SP: mdr1_leien, 1280 aa) (Chow *et al.*, 1993) and the colicin V secretion ATP-binding protein (SP: cvab_ecoli, 698 aa) (Gilson *et al.*, 1990). *B. subtilis* YwjA (GP: bsdna320d_2), which is assumed to be an ABC transporter, was recently deposited in a database. Its aa sequence exhibited high similarity to that of the YfiB.

yfiC. The aa sequence of YfiC also exhibited high similarity to those of ABC transporters and YfiB. The alignment of the YfiA, YfiB, YwjA and MsbA sequences is shown in Fig. 3. ABC transporters comprise a membrane-spanning domain and an ATP-binding domain. As judged from structural models of various ABC transporters, bacterial export complexes may form dimers (Fath & Kolter, 1993). In eukaryotes, the multidrug resistance protein 1, MDR, is an ABC transporter which seems to be a fused protein comprising two similar ABC transporters in the same orientation. Therefore, YfiB and YfiC may form a complex. But another hypothesis (YfiB or YfiC forms dimers) is equally possible.

yfiD. The aa sequence of YfiD exhibited high similarity to that of a hypothetical protein of *E. coli* (GP: ecouw67_32, 160 aa).

yfiF, yfiG and yfiH may form a sugar symport operon. The aa sequence of YfiF exhibited similarity to a multiple antibiotic resistance protein, which plays the role of a transcriptional activator (SP: mara_ecoli, 129 aa) (Cohen *et al.*, 1993). YfiF also exhibited similarity to several regulatory proteins, e.g. the L-rhamnose operon transcriptional activator (SP: rhar_ecoli, 312 aa) (Plunkett *et al.*, 1993) and the arabinose operon regulatory protein (SP: arac_erwch, 310 aa) (Lei *et al.*, 1985).

yfiG. The aa sequence of YfiG exhibited high similarity to the galactose-proton symport protein of *E. coli* (SP: galp_ecoli, 464 aa), the arabinose-proton symport protein of *E. coli* (SP: arae_ecoli, 472 aa) (Maiden *et al.*, 1987) and the xylose-proton symport protein of *E. coli* (xylose transporter; SP: xyle_ecoli, 491 aa) (Davis & Henderson, 1987).

Concluding remarks

Fig. 1 also shows that there are two ρ -independent terminators downstream of the *glv-2* and *yfiE* genes. Judging from the similar genetic organization of the *E. coli* *glv* operon (Reizer *et al.*, 1994) and the similar functions deduced for YfiB and YfiC (ABC transporters) and for

YfiF and YfiG (sugar symporter and its regulator), it is likely that the 12361 bp region contains three operons, all of which are related to membrane transport.

ACKNOWLEDGEMENTS

The authors wish to thank M. Itaya for providing the *NotI* linking clones, pNEXT23 and pNEXT47, and PFGE of *NotI* and/or *SfiI* fragments of the *B. subtilis* AC327 chromosome. This research was supported by a Grant-in-Aid for Creative Research (Human Genome Project) from the Ministry of Education, Science and Culture of Japan.

REFERENCES

- Akamatsu, T. & Sekiguchi, J. (1987). Genetic mapping and properties of filamentous mutations in *Bacillus subtilis*. *Agric Biol Chem* **51**, 2901–2909.
- Allikmets, R., Gerrard, B., Court, D. & Dean, M. (1993). Cloning and organization of the *abc* and *mdl* genes of *Escherichia coli*: relationship to eukaryotic multidrug resistance. *Gene* **136**, 231–236.
- Anagnostopoulos, C., Piggot, P. J. & Hoch, J. A. (1993). The genetic map of *Bacillus subtilis*. In *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology and Molecular Genetics*, pp. 425–461. Edited by A. L. Sonenshein, J. A. Hoch & R. Losick. Washington, DC: American Society for Microbiology.
- Chow, L. M., Wong, A. K., Ullman, B. & Wirth, D. F. (1993). Cloning and functional analysis of an extrachromosomally amplified multidrug resistance-like gene in *Leishmania enriettii*. *Mol Biochem Parasitol* **60**, 195–208.
- Cohen, S. P., Hächler, H. & Levy, S. B. (1993). Genetic and functional analysis of the multiple antibiotic resistance (*mar*) locus in *Escherichia coli*. *J Bacteriol* **175**, 1484–1492.
- Davis, E. O. & Henderson, P. J. (1987). The cloning and DNA sequence of the gene *xyfE* for xylose–proton symport in *Escherichia coli* K12. *J Biol Chem* **262**, 13928–13932.
- De Crécy-Lagard, V., Bouvet, O. M. M., Lejeune, P. & Danchin, A. (1991). Fructose catabolism in *Xanthomonas campestris* pv. *campestris*. Sequence of the *pts* operon, characterization of the fructose-specific enzymes. *J Biol Chem* **266**, 18154–18161.
- Fath, M. J. & Kolter, R. (1993). ABC transporters: bacterial exporters. *Microbiol Rev* **57**, 995–1017.
- Felmler, T., Pellett, S. & Welch, R. A. (1985). Nucleotide sequence of an *Escherichia coli* chromosomal hemolysin. *J Bacteriol* **163**, 94–105.
- Gilson, L., Mahanty, H. K. & Kolter, R. (1990). Genetic analysis of an MDR-like export system: the secretion of colicin V. *EMBO J* **9**, 3875–3894.
- Itaya, M. & Tanaka, T. (1991). Complete physical map of the *Bacillus subtilis* 168 chromosome constructed by a gene-directed mutagenesis method. *J Mol Biol* **220**, 631–648.
- Le Coq, D., Lindner, C., Krüger, S., Steinmetz, M. & Stülke, J. (1995). New β -glucoside (*bgl*) genes in *Bacillus subtilis*: the *bglP* gene product has both transport and regulatory functions similar to those of BglF, its *Escherichia coli* homolog. *J Bacteriol* **177**, 1527–1535.
- Lei, S.-P., Lin, H.-C., Heffernan, L. & Wilcox, G. (1985). *araB* gene and nucleotide sequence of the *araC* gene of *Erwinia carotovora*. *J Bacteriol* **164**, 717–722.
- Maiden, M. C. J., Davis, E. O., Baldwin, S. A., Moore, D. C. M. & Henderson, P. J. F. (1987). Mammalian and bacterial sugar transport proteins are homologous. *Nature* **325**, 641–643.
- Parker, L. L. & Hall, B. G. (1990). Characterization and nucleotide sequence of the cryptic *cel* operon of *Escherichia coli* K12. *Genetics* **124**, 455–471.
- Plunkett, G., Burland, V., Daniels, D. L. & Blattner, F. R. (1993). Analysis of the *Escherichia coli* genome. III. DNA sequence of the region from 87.2 to 89.2 minutes. *Nucleic Acids Res* **21**, 3391–3398.
- Poulsen, L. K., Larsen, N. W., Molin, S. & Andersson, P. (1992). Analysis of an *Escherichia coli* mutant strain resistant to the cell-killing function encoded by the *gef* gene family. *Mol Microbiol* **6**, 895–905.
- Prior, T. I. & Kornberg, H. L. (1988). Nucleotide sequence of *fruA*, the gene specifying enzyme II_{fru} of the phosphoenolpyruvate-dependent sugar phosphotransferase system in *Escherichia coli* K12. *J Gen Microbiol* **134**, 2757–2768.
- Reizer, J., Michotey, V., Reizer, A. & Saier, M. H., Jr (1994). Novel phosphotransferase system genes revealed by bacterial genome analysis: unique, putative fructose- and glucoside-specific systems. *Protein Sci* **3**, 440–450.
- Saier, M. H., Jr & Reizer, J. (1994). The bacterial phosphotransferase system: new frontiers 30 years later. *Mol Microbiol* **13**, 755–764.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Wu, L.-F. & Saier, M. H., Jr (1990). Nucleotide sequence of the *fruA* gene, encoding the fructose permease of the *Rhodobacter capsulatus* phosphotransferase system, and analyses of the deduced protein sequence. *J Bacteriol* **172**, 7167–7178.
- Yoshida, K., Sano, H., Seki, S., Oda, M., Fujimura, M. & Fujita, Y. (1995). Cloning and sequencing of a 29 kb region of the *Bacillus subtilis* genome containing the *hut* and *wapA* loci. *Microbiology* **141**, 337–343.

Received 6 October 1995; revised 5 January 1996; accepted 29 January 1996.