A 23·4 kb segment at the 69°–70° region of the *Bacillus subtilis* genome

Hiroki Yamamoto, Shigeki Uchiyama, Fajar Aji Nugroho and Junichi Sekiguchi

Author for correspondence: Junichi Sekiguchi. Tel: +81 268 21 5344. Fax: +81 268 21 5331.
e-mail: jsekigu@giptc.shinshu-u.ac.jp

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

Within the framework of the international project aimed at the sequencing of the *Bacillus subtilis* genome, a 23·4 kb chromosome segment has been cloned and sequenced. This region (23433 bp; 69°–70° of the genetic map) contains 17 complete ORFs and a partial one. A homology search for the products deduced from the 18 ORFs revealed that twelve of them had significant similarity to known proteins, including the quinolone-resistance protein, ABC transporter, aldehyde dehydrogenase, amino acid transporter, fosmidomycin-resistance protein, CDP-glucose 4,6-dehydratase, glucose-1-phosphate cytidyltransferase and cytochrome P450/NADPH-cytochrome P450 reductase.

Keywords: *Bacillus subtilis*, genome sequencing, 69°–70° region, cytochrome P450/NADPH-cytochrome P450 reductase

As participants in the international cooperative *Bacillus subtilis* genome sequencing project, we are responsible for the sequencing of an approximately 144 kb chromosomal region between pNEXT11 (69°) and *sspE* (81°) (Itaya & Tanaka, 1991). Recently, we determined a 12361 bp segment of the 76°–77° region (Yamamoto *et al.*, 1996a). We have sequenced the 73°–76° (Yamamoto *et al.*, 1996c), 77°–78° (Yamamoto *et al.*, 1996b) and 78°–81° (Yamamoto *et al.*, 1996d) regions. Here we report the nucleotide sequence and genome analysis of the 23·4 kb segment between 69° and 70°, and the finding of genes for aldehyde dehydrogenase, O-antigen biosynthesis and cytochrome P450/NADPH-cytochrome P450 reductase.

B. subtilis AC327 (*purB his-1 smo-1*) (Sekiguchi *et al.*, 1988) was used as the source of chromosomal DNA. A 37·0 kb region containing the *pel* locus (71°; Nasser *et al.*, 1993) has already been cloned by inverse PCR and sequenced. To amplify the next region, we performed long-range PCR with *B. subtilis* chromosomal DNA as a template, and a 19·6 kb product was obtained. Inverse PCR was used to clone the adjacent *PstI*–*inv* region (Fig. 1) and a 4·8 kb amplified fragment was obtained. We confirmed that the products of long-range and inverse PCR with DNA from both *B. subtilis* AC327 and 168 as templates showed the same restriction patterns on digestion with *Bam*HI, *Pst*I, *Eco*RI and *Hind*III, re-

spectively (data not shown). Two amplified fragments were used to obtain randomly overlapping libraries. Random sequencing as described previously (Yamamoto *et al.*, 1996a) followed by computer analysis indicated 17 complete ORFs and one incomplete ORF (the 5' region of *yfmO*) were identified (Table 1). Among the complete ORFs, 11 start with ATG, four with GTG, and two with TTG, and the direction of transcription of thirteen ORFs is the same as that of the movement of the replication fork (Fig. 1).

The aa sequence of YfmS exhibits partial similarity to the C-terminal regions of the methyl-accepting chemotaxis proteins (TlpB, TlpC, McpB, TlpA and McpA) of *B. subtilis* (Hanlon & Ordal, 1994) (Table 2). The bacterial chemotactic-signal transducers are composed of the two structural domains, in which an N-terminal periplasmic domain acts as the receptor and a C-terminal cytoplasmic domain contains the methylation sites (Hanlon & Ordal, 1994). However, there seems to be no transmembrane or periplasmic regions in YfmS. Therefore this may not be a chemotactic-signal transducer. The aa sequence of YfmT exhibits strong similarity to those of the aldehyde dehydrogenase and benzaldehyde dehydrogenase (XylC) of *Pseudomonas putida*, and salicylaldehyde dehydrogenase of *P. aeruginosa*. YfmT has glutamic acid and cysteine residues (corresponding to aas 253 and 287 with respect to the N-terminal aa, respectively) which have been implicated in the catalytic activity and are conserved in all the enzymes of aldehyde dehydrogenase family. The aa sequence of YfnG exhibits some similarity to those of the CDP-glucose-4,6-

The GSDB, DDBJ, EMBL and NCBI accession numbers for the nucleotide sequence reported in this paper are D86418 and D87979.

Fig. 1. Physical map and gene organization of the 69°–70° region. A physical map of the sequenced 23.4 kb region from 69° (right end) to 70° (left end) of the *B. subtilis* chromosome is shown. The *NotI* site of pNEXT11 and the movement of the replication fork are indicated. The deduced 18 ORFs, and their transcriptional and translational directions are indicated by arrows and arrowheads. The deduced rho-independent terminators are denoted by asterisks before numbers: 1, $\Delta G = -21.4 \text{ kcal mol}^{-1}$ (89.5 kJ mol $^{-1}$); 2, $-15.7 \text{ kcal mol}^{-1}$ (65.7 kJ mol $^{-1}$); 3, $-30.7 \text{ kcal mol}^{-1}$ (128.4 kJ mol $^{-1}$); 4, $-32.4 \text{ kcal mol}^{-1}$ (135.6 kJ mol $^{-1}$); 5, $-31.1 \text{ kcal mol}^{-1}$ (130.1 kJ mol $^{-1}$); 6, $-28.6 \text{ kcal mol}^{-1}$ (119.7 kJ mol $^{-1}$); 7, $-27.7 \text{ kcal mol}^{-1}$ (115.9 kJ mol $^{-1}$); 8, $-37.1 \text{ kcal mol}^{-1}$ (155.2 kJ mol $^{-1}$); 9, $-23.7 \text{ kcal mol}^{-1}$ (99.2 kJ mol $^{-1}$); 10, $-20.4 \text{ kcal mol}^{-1}$ (85.4 kJ mol $^{-1}$); 11, $-23.4 \text{ kcal mol}^{-1}$ (97.9 kJ mol $^{-1}$). P, B and E indicate *PstI*, *BamHI* and *EcoRI* sites, respectively. The broken arrow indicates the partial sequenced ORFs (the 5' terminal region of *yfmO*).

Table 1. Putative ORFs in the 23433 bp sequence of the 69°–70° region

The transcriptional direction of the ORF is indicated by >. The thick arrow indicates a partial ORF. The sequence of the 3' terminal region of *yfmO* will be described elsewhere.

ORF	Endpoint (nucleotides)	Size of product (aa/kDa)	SD consensus sequence (upper case) and initiation codon (bold)
<i>yfmO</i>	975 > -213 →	396/41.5	gAtGGAGGTtAggtcaa atg
<i>yfmP</i>	1463 > 1044	140/16.5	gAAGGgGGattT ac cttg
<i>yfmQ</i>	2034 > 1591	148/16.9	AAcGGAGaTGATga ag tatg
<i>yfmR</i>	4046 > 2160	629/71.1	AcgtGAGGatgatagga acc atg
<i>yfmS</i>	5041 > 4184	286/30.3	AAAGGAGtg aa tg
<i>yfmT</i>	6512 > 5058	485/53.3	AAAGGAGaTGcTg acc atg
<i>yfnA</i>	6759 > 8144	462/49.0	ctAGGAGGaacT tt tg
<i>yfnB</i>	8239 > 8943	235/27.4	AAAGGgaGacAcacag atg
<i>yfnC</i>	9057 > 10283	409/43.5	tAAGGAGGaG Act caatg
<i>yfnD</i>	11252 > 10320	311/37.1	AttGGAtGTGgTata ag tg
<i>yfnE</i>	12431 > 11256	392/45.5	AcAGGAaGTGA aa atga ag tg
<i>yfnF</i>	13371 > 12463	303/35.0	AAAcGAGGTGAaCa ag actg tg
<i>yfnG</i>	14297 > 13398	300/34.0	ttAGGAaGctAT tt g tg
<i>yfnH</i>	15134 > 14373	254/28.7	AtgaGgGGT ag agatg
<i>yfnI</i>	17331 > 15373	653/75.0	AgAGcAGGTaA at atg
<i>yfnJ</i>	20921 > 17739	1061/119.5	AAAGGAatgGtTC g atg
<i>yfnK</i>	21071 > 22138	356/41.4	tAAGGtGagttTata at g
<i>yfnL</i>	23285 > 22179	369/41.0	AAAGGtGGTtAT tg atg

dehydratases of *Yersinia enterocolitica* (RfbG), *Yersinia pseudotuberculosis* (AscB) and *Salmonella typhimurium* (RfbG). Moreover the product of *yfnH*, which locates

just upstream of *yfnG*, exhibits high similarity to those of the glucose-1-phosphate cytidyltransferases of *Y. pseudotuberculosis* (AscA), *Y. enterocolitica* (RfbF) and

Table 2. Similarity of the predicted ORF products to known proteins

Product	Similar protein in database*	Identity observed	FASTA score	Reference
YfmO	Quinolone-resistance protein (<i>norA</i>) of <i>Staphylococcus aureus</i> [sp:NORA_STAAU]	23.7% in 393 residues	348	Yoshida <i>et al.</i> (1990)
YfmR	ABC transporter homologue (<i>yjyK</i>) of <i>Escherichia coli</i> [sp:YJJK_ECOLI]	35.2% in the N-terminal region 545 residues	1173	Engel <i>et al.</i> (1991) Burland <i>et al.</i> (1995)
YfmS	Methyl-accepting chemotaxis protein (<i>tlpB</i>) of <i>Bacillus subtilis</i> [sp:TLPB_BACSU]	35.6% in the C-terminal region 177 residues	300	Hanlon & Ordal (1994)
YfmT	Aldehyde dehydrogenase of <i>Pseudomonas putida</i> [gp:PSEPADA_1, gi:995953]	44.9% in 485 residues	1390	
	Benzaldehyde dehydrogenase (<i>xylC</i>) of <i>Pseudomonas putida</i> [sp:XYLC_PSEPU]	41.5% in 480 residues	1196	Inoue <i>et al.</i> (1995)
YfnA	Amino acid transporter (<i>aat1</i>) of <i>Arabidopsis thaliana</i> [prf:2205352B]	34.7% in the N-terminal region 438 residues	875	
YfnB	Hypothetical protein (<i>yjyG</i>) of <i>Escherichia coli</i> [pir:S56598]	29.9% in 221 residues	372	Burland <i>et al.</i> (1995)
YfnC	Fosmidomycin-resistance protein (<i>fsr</i>) of <i>Escherichia coli</i> [gp:ECOFSR_1, gi:1019360]	44.6% in 386 residues	1147	
YfnG	TDP-glucose 4,6-dehydratase homologue of <i>Synechocystis</i> sp. [gp:SYCSLLE_73, gi:1001273]	30.3% in 287 residues	458	Kaneko <i>et al.</i> (1995)
	CDP-glucose 4,6-dehydratase (<i>rfbG</i>) of <i>Yersinia enterocolitica</i> [gp:YEU46859_4, gi:1197641]	26.0% in 311 residues	443	
	CDP-glucose dehydratase (<i>ascB</i>) of <i>Yersinia pseudotuberculosis</i> [prf:2017287C]	29.2% in 312 residues	425	Thorson <i>et al.</i> (1994)
	CDP-glucose 4,6-dehydratase (<i>rfbG</i>) of <i>Salmonella typhimurium</i> [sp:RFBG_SALTY]	27.5% in 313 residues	420	Jiang <i>et al.</i> (1991)
YfnH	Glucose cytidyltransferase (<i>ascA</i>) of <i>Yersinia pseudotuberculosis</i> [prf:2017287B]	45.9% in 257 residues	807	Thorson <i>et al.</i> (1994)
	Glucose-1-phosphate cytidyltransferase (<i>rfbF</i>) of <i>Yersinia enterocolitica</i> [gp:YEU46859_3, gi:1197640]	45.5% in 257 residues	803	
	Glucose-1-phosphate cytidyltransferase (<i>rfbF</i>) of <i>Salmonella typhimurium</i> [sp:RFBF_SALTY]	43.2% in 257 residues	759	Jiang <i>et al.</i> (1991)
YfnI	Hypothetical protein (<i>yqgS</i>) of <i>Bacillus subtilis</i> [gp:BACJH642_170, gi:1303866]	45.6% in 642 residues	1913	
YfnJ	Cytochrome P450 (BM-3)/NADPH-cytochrome P450 reductase (<i>cyp-102</i>) of <i>Bacillus megaterium</i> [sp:CPXB_BACME]	59.5% in 1055 residues	4272	Ruettinger <i>et al.</i> (1989)
YfnL	FAD-binding protein homologue of <i>Pseudomonas aeruginosa</i> [gp:PAU29897_1, gi:912582]	28.8% in the N-terminal region 243 residues	264	

*gi, gp, pir, prf and sp denote the NCBI, GenBank Protein, PIR, PRF and SWISS-PROT databases, respectively.

S. typhimurium (*RfbF*). CDP-glucose-4,6-dehydratase and glucose-1-phosphate cytidyltransferase are enzymes required for the O-antigen biosynthetic pathway. In *S. typhimurium*, the *rfb* gene cluster has been cloned and sequenced (Jiang *et al.*, 1991). It has been reported that the *rfb* gene cluster is a large operon, which consists of approximately 16 ORFs, and that it is essential for O-antigen expression. *rfbF* and *rfbG* locate the sixth and seventh genes in the *rfb* operon, respectively. Judging from the genome structure, since there is no rho-independent terminator structure between *yfnH* and *yfnD*, *yfnH*–*G*–*F*–*E*–*D* genes may form

an operon (Fig. 1). However, the products of *yfnF*, *yfnE* and *yfnD* had no significant similarity to the other ORFs included in the *rfb* and *asc* gene clusters. The aa sequence of YfnJ exhibits extremely strong similarity to that of the cytochrome P450/NADPH-P450 reductase of *B. megaterium* (Cyp102). Cytochrome P450 reductase is required for electron transfer from NADP to cytochrome P450. Cyp102 is a large protein consisting of 1048 aa residues that contains an N-terminal P450 domain followed by a reductase domain (Ruettinger *et al.*, 1989; Ravichandran *et al.*, 1993). *yfnJ* is the second gene corresponding to the fusion of those for cytochrome

P450 and NADPH-cytochrome P450 reductase in a living organism.

Among the 18 ORFs in the 23.4 kb region, six did not exhibit any significant aa sequence similarity to those in the non-redundant protein sequence databases. Therefore, our research is now directed toward the isolation of gene-disrupted mutants in a systematic way under the international cooperative project.

ACKNOWLEDGEMENTS

We would like to thank M. Itaya for providing the *NotI* linking clone, pNEXT11. 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

- Burland, V., Plunkett, G., III, Sofia, H. J., Daniels, D. L. & Blattner, F. R. (1995). Analysis of the *Escherichia coli* genome VI: DNA sequence of the region from 92.8 through 100 minutes. *Nucleic Acids Res* **23**, 2105–2119.
- Engel, H., Kazemier, B. & Keck, W. (1991). Murein-metabolizing enzymes from *Escherichia coli*: sequence analysis and controlled overexpression of the *slt* gene, which encodes the soluble lytic transglycosylase. *J Bacteriol* **173**, 6773–6782.
- Hanlon, D. W. & Ordal, G. W. (1994). Cloning and characterization of genes encoding methyl-accepting chemotaxis proteins in *Bacillus subtilis*. *J Biol Chem* **269**, 14038–14046.
- Inoue, J., Shaw, J. P., Rekik, M. & Harayama, S. (1995). Overlapping substrate specificities of benzaldehyde dehydrogenase (the *xylC* gene product) and 2-hydroxymuconic semialdehyde dehydrogenase (the *xylG* gene product) encoded by TOL plasmid pWWO of *Pseudomonas putida*. *J Bacteriol* **177**, 1196–1201.
- 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.
- Jiang, X. M., Neal, B., Santiago, F., Lee, S. J., Romana, L. K. & Reeves, P. R. (1991). Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar *typhimurium* (strain LT2). *Mol Microbiol* **5**, 695–713.
- Kaneko, T., Tanaka, A., Sato, S., Kotani, H., Sazuka, T., Miyajima, N., Sugiura, M. & Tabata, S. (1995). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome. *DNA Res* **2**, 153–166.
- Nasser, W., Awade, A. C., Reverchon, S. & Robert-Baudouy, J. (1993). Pectate lyase from *Bacillus subtilis*: molecular characterization of the gene, and properties of the cloned enzyme. *FEBS Lett* **335**, 319–326.
- Ravichandran, K. G., Boddupalli, S. S., Hasermann, C. A., Peterson, J. A. & Deisenhofer, J. (1993). Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450s. *Science* **261**, 731–736.
- Ruettinger, R. T., Wen, L. P. & Fulco, A. J. (1989). Coding nucleotide, 5' regulatory, and deduced amino acid sequences of P-450BM-3, a single peptide cytochrome P-450: NADPH-P-450 reductase from *Bacillus megaterium*. *J Biol Chem* **264**, 10987–10995.
- Sekiguchi, J., Ezaki, B., Kodama, K. & Akamatsu, T. (1988). Molecular cloning of a gene affecting the autolysin level and flagellation in *Bacillus subtilis*. *J Gen Microbiol* **134**, 1611–1621.
- Thorson, J. S., Lo, S. F., Ploux, O., He, X. & Liu, H. (1994). Studies of the biosynthesis of 3,6-dideoxyhexoses: molecular cloning and characterization of the *asc* (ascarylose) region from *Yersinia pseudotuberculosis* serogroup VA. *J Bacteriol* **176**, 5483–5493.
- Yamamoto, H., Uchiyama, S., Fajar, A. N., Ogasawara, N. & Sekiguchi, J. (1996a). Determination of a 12 kb nucleotide sequence around the 76° region of the *Bacillus subtilis* chromosome. *Microbiology* **142**, 1417–1421.
- Yamamoto, H., Uchiyama, S. & Sekiguchi, J. (1996b). The *Bacillus subtilis* chromosome region near 78° contains the genes encoding a new two-component system, three ABC transporters and a lipase. *Gene* **181**, 147–151.
- Yamamoto, H., Uchiyama, S. & Sekiguchi, J. (1996c). Cloning and sequencing of a 40.6 kb segment in the 73°–76° region of the *Bacillus subtilis* chromosome containing genes for trehalose metabolism and acetoin utilization. *Microbiology* **142**, 3057–3065.
- Yamamoto, H., Uchiyama, S. & Sekiguchi, J. (1996d). Cloning and sequencing of a 27.8 kb nucleotide sequence of the 79°–81° region of the *Bacillus subtilis* genome containing the *sspE* locus. *DNA Res* **3**, 257–262.
- Yoshida, H., Bogaki, M., Nakamura, S., Ubukata, K. & Konno, M. (1990). Nucleotide sequence and characterization of the *Staphylococcus aureus* *norA* gene, which confers resistance to quinolones. *J Bacteriol* **172**, 6942–6949.

Received 10 September 1996; revised 2 December 1996; accepted 13 December 1996.