

A *Bacillus subtilis* Gene Encoding a Protein Similar to Nucleotide Sugar Transferases Influences Cell Shape and Viability

KIRSTEN D. PRICE,¹ STEVEN ROELS,² AND RICHARD LOSICK^{1*}

Department of Molecular and Cellular Biology, The Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138,¹ and Millennium Pharmaceuticals, Cambridge, Massachusetts 02139²

Received 24 April 1997/Accepted 2 June 1997

***Bacillus subtilis* gene *yfpP*, which is located at 196° on the genetic map, shows similarity to both the monogalactosyldiacylglycerol synthase gene of *Cucumis sativus*, which encodes a galactosyltransferase, and the *murG* genes of *B. subtilis*, *Escherichia coli*, *Haemophilus influenzae*, and *Synechocystis* sp. strain PCC6803, which encode *N*-acetylglucosaminyltransferases involved in peptidoglycan biosynthesis. Cells containing a null mutation of *yfpP* are shorter and rounder than wild-type cells during growth in Luria-Bertani medium and glucose minimal medium. In addition, the mutant cells preferentially undergo lysis when grown on solid Luria-Bertani medium.**

The chloroplast membranes of higher plants and eukaryotic algae contain two major galactolipids, monogalactodiacylglycerol (MGDG) and digalactosyldiacylglycerol. The final step in the synthesis of MGDG is catalyzed by MGDG synthase, a UDPgalactose:1,2-diacylglycerol 3-β-D-galactosyltransferase, which transfers galactose from UDPgalactose to 1,2-diacylglycerol (9). A sequence of an MGDG synthase from cucumber (*Cucumis sativus*) is known and was shown to be similar to those of the MurG peptidoglycan biosynthesis enzymes of *Escherichia coli* and *Bacillus subtilis* (17). The rigid peptidoglycan layer of the bacterial cell wall is composed of long glycan chains that are cross-linked through peptide side chains. The glycan chains are alternating copolymers of *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) that are held together by β1→4 glycosidic linkages. The chains are formed by polymerization of disaccharide units that are assembled on a lipid carrier molecule at the inner surface of the cytoplasmic membrane (11, 12, 18). Synthesis of the disaccharide is catalyzed by MurG, a UDP-*N*-acetylglucosamine:*N*-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol *N*-acetylglucosamine transferase, which transfers GlcNAc

from the nucleotide sugar UDP-GlcNAc to undecaprenyl-pyrophosphoryl-MurNAc-pentapeptide to form undecaprenyl-pyrophosphoryl-MurNAc-(pentapeptide)GlcNAc (2, 14). Here we report that *B. subtilis* contains a gene, *yfpP*, that influences cell shape and viability, and whose inferred product resembles MGDG synthase of *C. sativus* and shows similarity to some conserved regions of MurG proteins.

yfpP, which is 382 codons in length, was previously identified by the *B. subtilis* sequencing project and lies between *metB* and *cspD* at 196° on the chromosome (3). Our searches against available sequence databases have revealed that the inferred amino acid sequence of YpfP is most similar to that of MGDG synthase of *C. sativus*. An alignment of these proteins shows they exhibit 24% identity overall and 37% identity in four regions involving 183 YpfP residues (Fig. 1). In confirmation and extension of the results obtained by Shimojima et al. (17), we found that YpfP, like MGDG synthase, contains regions of great similarity to some conserved regions of the MurG proteins of *B. subtilis* (1, 7, 15), *Haemophilus influenzae* (5), *E. coli* (8, 13, 19), and *Synechocystis* sp. strain PCC6803 (10) (Fig. 2). The similarity is most extensive in an approximately 50-amino-

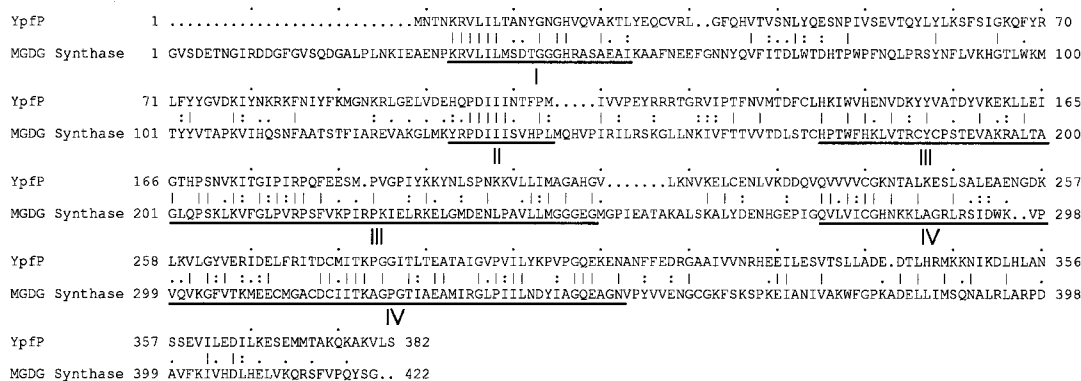


FIG. 1. Alignment, generated by the GAP program of the Wisconsin Genetics Computer Group sequence analysis package (version 9.0), between *B. subtilis* YpfP and the mature form of *C. sativus* MGDG synthase. Identical residues are indicated by lines. Highly and moderately similar residues are indicated by double dots and single dots, respectively. The database accession numbers corresponding to the *B. subtilis* YpfP and *C. sativus* MGDG synthase sequences are PS4166 and U62622, respectively. The roman numerals below the sequences indicate the regions of greatest identity.

* Corresponding author. Phone: (617) 495-1774. Fax: (617) 496-4642.

	****	
MurG (H.i.)	91	EEKPDAVLGM	100	155	VVGNPVEDLFE..MPNPD..IRFSDREKLRVLVVGGSQGARV	194
MurG (E.c.)	93	AYKPDVVLGM	102	157	VVGNPVRTDVLV..LPLPQ..QRLAGREGPVRVLVVGGSQGARI	196
MurG (B.s.)	90	EFKPDVAVIGT	99	159	FTGNFRASEVVS..IKTGRSLAEFGLSEDKKTVLIFGGSRGAAP	200
MurG (Syn.)	105	SKKINAVATT	114	174	WISTFVREQFRQPSLDLPI.....PPNRSGLIVVAGGSQGAVT	211
YpfP	102	EHQPDIINT	111	174	ITGIPIRPQFEESM.PVGPYKKNLSPNKKVLLIMAGAHGVLK	216
MGDG synthase	132	KYRPDIISV	141	209	VFGLPVRFSFVKPIRPKIELRKELGMDENLPAVLLMGGGEMGP	252
Consensus		KPD VI		G PVR	VLV GG G	
				*	*	*
	*****
MurG (H.i.)	250	WADVVICRSGALTVCEIAAVGAAAFVPPQH	300	KDRQOYLNAKYLSDVGA	AKI	300
MurG (E.c.)	253	WADVVCRRSGALTVSEIAAAGLPALFVPPQH	303	KDRQOYWNALPLEKAGA	AKI	303
MurG (B.s.)	259	AIDVIVARAGATTIAEITALGIPSVLIPSPYVTANHQEVNARSLGQHDAAIV	310			310
MurG (Syn.)	264	KADLAI SRAGAGTLELAVTQTPSILIPYFFAENHQMYNAQVFDAGAALM	315			315
YpfP	273	ITDCMITKPGGITLTEATAIGVPVILYKP...VPGQEKENANFFEDRGAATV	321			321
MGDG synthase	314	ACDCIITKAGPGTIAEAMIRGLPIILNDY...IAGQEAGNVFYVVENGCCKF	362			362
Consensus		D I R G T E	G P I	Q NA	D GAA	
		*	*	*	*	*

FIG. 2. Alignment of three regions of similarity between YpfP and *C. sativus* MGDG synthase and the MurG proteins of *H. influenzae*, *E. coli*, *B. subtilis*, and *Synechocystis* sp. strain PCC6803. This alignment was generated with the PILEUP and PRETTY programs of the Wisconsin Genetics Computer Group sequence analysis package (version 9.0). A consensus sequence obtained by using PRETTY with the Blosum62 amino acid substitution matrix (plurality, 5; threshold, 2) is shown below the sequences. An asterisk indicates residues that are identical in all six proteins. The dots above the sequences indicate residues that are identical in at least three of the four MurG proteins at that position. The database accession numbers for the MurG proteins of *B. subtilis*, *H. influenzae*, *E. coli*, and *Synechocystis* sp. strain PCC6803 are P37585, P45065, P17443, and D90917, respectively.

acid region near the carboxyl terminus. The alignment shown in Fig. 2 was used to generate a consensus sequence of 31 highly conserved residues. The YpfP sequence is identical to that of the consensus at 23 positions and contains a similar amino acid at 6 of the remaining 8 positions. In general, the residues appearing in the consensus sequence are those that are most highly conserved among the MurG proteins in those regions (Fig. 2), suggesting

that these regions in MGDG synthase and YpfP are functionally similar to those of the MurG proteins.

To investigate the function of *ypfP*, we replaced the entire *ypfP* open reading frame with a gene conferring resistance to kanamycin. To construct this mutant, *ypfP*, including flanking sequences extending 200 bp upstream and 400 bp downstream of the gene, was cloned following amplification by PCR. Next,

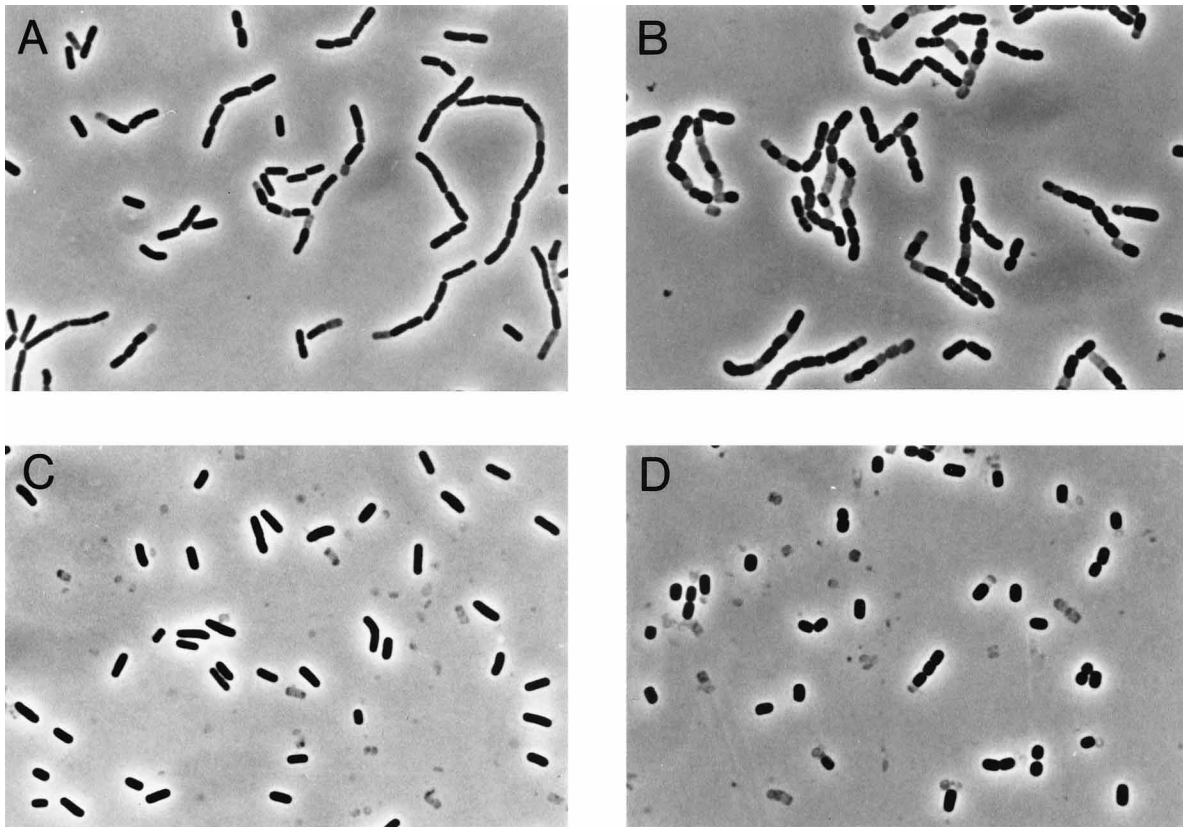


FIG. 3. Phase contrast microscopy of *B. subtilis* cells. Cultures were started at an optical density at 600 nm of 0.05 in LB liquid media, and photomicrographs were taken after 12 and 24 h. (A) Wild-type cells of strain PY79 after 12 h; (B) *ypfP* null mutant cells of strain KP261 after 12 h; (C) wild-type cells of strain PY79 after 24 h; (D) *ypfP* null mutant cells of strain KP261 after 24 h. Bar, 5 μ m.

an *NsiI-RsaI* fragment encompassing the entire *ypfP* open reading frame was replaced with a kanamycin resistance cassette (6). The resulting plasmid was then used to replace *ypfP* in the *B. subtilis* chromosome with the kanamycin resistance gene via marker replacement recombination to generate strain KP261 (*ypfPΔ::Km^r*). Loss of a spectinomycin resistance cassette present in the plasmid backbone during construction of this strain indicated that the plasmid had not integrated via single-crossover (Campbell-type) reciprocal recombination.

KP261 was able to form colonies on Luria-Bertani (LB) solid medium (16), but in contrast to its cogenic parent, the mutant exhibited substantial lysis after about 36 h of growth at 37°C. Preferential lysis by the mutant was not observed on Tris Spizizen salts (TSS) solid medium (4), a glucose minimal medium, or in LB or TSS liquid media. In addition, in both types of liquid media the growth curves of the mutant and wild-type parent were similar (data not shown). On LB and TSS solid media and in LB and TSS liquid media, the mutant cells were shorter and rounder than those of the wild type. In both types of liquid media, this phenotype was visible as early as the exponential phase of growth (data not shown). In LB medium, however, the phenotype was less striking earlier during growth but became more pronounced the further the cells progressed into the stationary phase (Fig. 3). In another characterization, the mutant cells showed no measurable defect in their capacity to sporulate when tested on Difco sporulation medium (data not shown).

That the lysis and shape phenotypes were due to the absence of *ypfP* and not to a polar effect of the deletion/insertion mutation on the expression of a downstream gene was demonstrated by a complementation experiment. *ypfP*, including flanking sequences extending 375 bp upstream and 200 downstream of the open reading frame, was cloned into pDG364 (4), an *amyE* integration vector. We then introduced this plasmid into KP261 and isolated a transformant in which the plasmid-borne copy of *ypfP* had integrated at (and thereby inactivated) the *amyE* locus via double recombination. The genotype of the resulting strain, KP278 (*ypfPΔ::Km^r amyE::ypfP⁺*), was verified by backcross experiments with the wild-type strain. KP278 behaved indistinguishably from the wild type with respect to lysis and cell shape (data not shown).

Because the *ypfP* null mutation affects cell shape, which is determined by the cell envelope, and, in keeping with its similarity to MGDG synthase and the MurG peptidoglycan biosynthesis enzymes, we propose that YpfP plays a role in cell envelope biosynthesis or maintenance. That MGDG synthase is 24% identical to YpfP is significant but not sufficiently so to indicate that both proteins have the same function. More noteworthy is that MGDG synthase and YpfP share regions of significant similarity with known MurG proteins. What could be the function of these domains? Both MGDG synthase and MurG catalyze the formation of a glycosidic bond between a sugar residue and a lipid-linked substrate. MGDG synthase transfers galactose from UDPgalactose to 1,2-diacylglycerol, whereas MurG transfers GlcNAc from UDP-GlcNAc to lipid-linked MurNAc-pentapeptide. Perhaps the regions of similarity comprise binding sites for the nucleotide sugar substrates of these enzymes or the catalytic domain that is responsible for the glycosyltransferase activity. In any event, it seems possible that YpfP is a nucleotide sugar transferase and is involved in the synthesis or modification of a component of the cell envelope.

We thank A. Sorokin and D. Ehrlich for providing the DNA se-

quence prior to publication and C. Webb for critical reading of the manuscript.

This work was supported by a grant from the Eli Lilly Corporation and by NIH grant GM18568 to R.L.

REFERENCES

1. Beall, B., and J. Lutkenhaus. 1989. Nucleotide sequence and insertional inactivation of a *Bacillus subtilis* gene that affects cell division, sporulation, and temperature sensitivity. *J. Bacteriol.* **171**:6821–6834.
2. Bupp, K., and J. van Heijenoort. 1993. The final step of peptidoglycan subunit assembly in *Escherichia coli* occurs in the cytoplasm. *J. Bacteriol.* **175**:1841–1843.
3. Capuano, V., N. Galleron, P. Pujic, A. Sorokin, and S. D. Ehrlich. 1996. Organization of the *Bacillus subtilis* 168 chromosome between *kdg* and the attachment site of the SPβ prophage: use of Long Accurate PCR and yeast artificial chromosomes for sequencing. *Microbiology* **142**:3005–3015.
4. Cutting, S. M., and P. B. Vander Horn. 1990. Genetic analysis, p. 27–74. In C. R. Harwood and S. M. Cutting (ed.), *Molecular biological methods for Bacillus*. John Wiley & Sons Ltd., Chichester, United Kingdom.
5. Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J.-F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. Sutton, W. FitzHugh, C. Fields, J. D. Gocayne, J. Scott, R. Shirley, L.-I. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. R. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. M. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith, and J. C. Venter. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**:496–512.
6. Guérout-Fleury, A.-M., K. Shazand, N. Frandsen, and P. Stragier. 1995. Antibiotic-resistance cassettes for *Bacillus subtilis*. *Gene* **167**:335–336.
7. Henriques, A. O., H. de Lencastre, and P. J. Piggot. 1992. A *Bacillus subtilis* morphogene cluster that includes *spoVE* is homologous to the *mra* region of *Escherichia coli*. *Biochimie* **74**:735–748.
8. Ikeda, M., M. Wachi, H. K. Jung, F. Ishino, and M. Matsuhashi. 1990. Nucleotide sequence involving *murG* and *murC* in the *mra* gene cluster region of *Escherichia coli*. *Nucleic Acids Res.* **18**:4014.
9. Joyard, J., and R. Douce. 1987. Galactolipid synthesis, p. 215–274. In P. K. Stumpf (ed.), *The biochemistry of plants*. Academic Press, New York, N.Y.
10. Kaneko, T., S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirose, M. Sugiura, S. Sasamoto, T. Kimura, T. Hosouchi, A. Matsumoto, A. Muraki, N. Nakazaki, K. Naruo, S. Okumura, S. Shimpo, C. Takeuchi, T. Wada, A. Watanabe, M. Yamada, M. Yasuda, and S. Tabata. 1996. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**:109–136.
11. Labischinski, H., and H. Maidhof. 1994. Bacterial peptidoglycan: overview and evolving concepts, p. 23–38. In J.-M. Ghuysen and R. Hakenbeck (ed.), *Bacterial cell wall*. Elsevier, Amsterdam, The Netherlands.
12. Matsuhashi, M. 1994. Utilization of lipid-linked precursors and the formation of peptidoglycan in the process of cell growth and division: membrane enzymes involved in the final steps of peptidoglycan synthesis and the mechanism of their regulation, p. 55–71. In J.-M. Ghuysen and R. Hakenbeck (ed.), *Bacterial cell wall*. Elsevier, Amsterdam, The Netherlands.
13. Mengin-Lecreux, D., L. Texier, and J. van Heijenoort. 1990. Nucleotide sequence of the cell-envelope *murG* gene of *Escherichia coli*. *Nucleic Acids Res.* **18**:2810.
14. Mengin-Lecreux, D., L. Texier, M. Rousseau, and J. van Heijenoort. 1991. The *murG* gene of *Escherichia coli* codes for the UDP-*N*-acetylglucosamine-*N*-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol *N*-acetylglucosamine transferase involved in the membrane steps of peptidoglycan synthesis. *J. Bacteriol.* **173**:4625–4636.
15. Miyao, A., A. Yoshimura, T. Sato, G. Yamamoto, G. Theeragool, and Y. Kobayashi. 1992. Sequence of the *Bacillus subtilis* homolog of the *Escherichia coli* cell-division gene *murG*. *Gene* **118**:147–148.
16. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
17. Shimajima, M., H. Ohta, A. Iwamoto, T. Masuda, Y. Shioi, and K. Takamiya. 1997. Cloning of the gene for monogalactosyldiacylglycerol synthase and its evolutionary origin. *Proc. Natl. Acad. Sci. USA* **94**:333–337.
18. van Heijenoort, J. 1994. Biosynthesis of the bacterial peptidoglycan unit, p. 39–54. In J.-M. Ghuysen and R. Hakenbeck (ed.), *Bacterial cell wall*. Elsevier, Amsterdam, The Netherlands.
19. Yura, T., H. Mori, H. Nagai, T. Nagata, A. Ishihama, N. Fujita, K. Isono, K. Mizobuchi, and A. Nakata. 1992. Systematic sequencing of the *Escherichia coli* genome: analysis of the 0–2.4 min region. *Nucleic Acids Res.* **20**:3305–3308.