Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 333–337, January 1997
Plant Biology

Cloning of the gene for monogalactosyldiacylglycerol synthase and its evolutionary origin

(galactosyltransferase/endosymbiosis/chloroplast membrane/peptidoglycan)

MIE SHIMOJIMA*, HIROYUKI OHTA*†, AKIHIRO IWAMATSU‡, TATSURU MASUDA*, YUZO SHIOI§, AND KEN-ICHIRO TAKAMIYA*

*Department of Biological Sciences, Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama, Kanagawa 226, Japan; ‡Central Laboratories for Key Technology, Kirin Brewery Company, 1-13-5 Fukuura, Kanazawa-ku, Yokohama, Kanagawa 236, Japan; and §Department of Biology and Geoscience, Faculty of Science, Shizuoka University, 836 Ohya, Shizuoka 422, Japan

Communicated by Takayoshi Higuchi, Kyoto University, Kyoto, Japan, November 6, 1996 (received for review August 26, 1996)

ABSTRACT **Monogalactosyldiacylglycerol (MGDG) synthase (UDPgalactose:1,2-diacylglycerol 3-β-D-galactosyltransferase; EC 2.4.1.46) catalyzes formation of MGDG, a major structural lipid of chloroplast. We cloned a cDNA for the synthase from cucumber cDNA library. The full-length cDNA clone was 2142 bp, and it contains a 1575-bp open reading frame encoding 525 aa. The open reading frame consists of the regions for a mature protein (422 aa; M_r of 46,552) and transit peptide to chloroplast (103 aa). Although the molecular weight of mature protein region matched that purified from cucumber cotyledons, it was quite different from those purified from spinach (≈20 kDa) reported by other groups. The mature region of the protein was expressed in *Escherichia coli* as a fusion protein with glutathione *S*-transferase. The expression in *E. coli* showed that the protein catalyzed MGDG synthesis very efficiently. Therefore, we concluded that the cDNA encodes MGDG synthase in cucumber. In addition, the deduced amino acid sequence of the MGDG synthase cDNA showed homology with MurG of *Bacillus subtilis* and *E. coli*, which encode a glycosyltransferase catalyzing the last step of peptidoglycan synthesis in bacteria. This sequence homology implies that the machinery of chloroplast membrane biosynthesis is evolutionarily derived from that of cell wall biosynthesis in bacteria. This is consistent with the endosymbiotic hypothesis of chloroplast formation.**

Chloroplast thylakoid membranes are the site of the plant's photochemical reactions and are composed of proteins, such as the photosynthetic reaction center, and lipids. In higher plants and eukaryotic algae, about 50% (wt/wt) of the membrane lipids of chloroplasts are two major galactolipids, monogalactosyldiacylglycerol (MGDG), and digalactosyldiacylglycerol (DGDG). MGDG and DGDG are not only major constituents of chloroplast membrane but have a major role in determining the physicochemical characteristics of thylakoid membranes (1, 2). A tightly bound MGDG molecule in the photosystem II reaction center complex has been identified as important in photosynthesis (3).

The final step of MGDG synthesis is catalyzed by MGDG synthase (UDPgalactose: 1,2-diacylglycerol 3-β-D-galactosyltransferase; EC 2.4.1.46). The enzyme transfers a galactose from UDPgalactose to 1,2-diacylglycerol in chloroplast envelope (4). Another major galactolipid, DGDG, is synthesized by dismutation of two molecules of MGDG (5). We have found that MGDG synthase activity in cucumber seedlings dramatically increased coincidentally with the accumulation of the two major galactolipids, MGDG and DGDG, during chloroplast development (6). This fact and other reports (4) strongly suggest that the reaction catalyzed by MGDG synthase is a key step in the formation of chloroplast membrane. Despite the importance of MGDG synthase for chloroplast membrane biogenesis, molecular characteristics of the enzyme remain ambiguous. Purification of MGDG synthase from spinach envelopes was reported by two groups (7, 8). According to their results, molecular mass of the purified proteins and the properties of the solubilized enzymes were different from each other. Teucher and Heinz (7) proposed that a 22-kDa polypeptide was associated with the activity. In contrast, Maréchal *et al.* (8) reported that they led to 90% enrichment of a 19-kDa polypeptide as the enzyme. In each case, there was no direct evidence to determine which polypeptide was indeed MGDG synthase, because the final amount of the purified enzyme was quite low when the purification was started from envelope membranes (7, 8). In addition to such discrepancy of the molecular weight, Covès *et al.* (9) reported lipid requirement for the activity of the MGDG synthase solubilized with 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonic acid (CHAPS), although cholate-solubilized enzyme did not show such requirement for the enzyme activity (7). The ambiguity of the molecular weight and the low abundance of the protein have prevented further investigations into the structure of the enzyme.

We recently purified the MGDG synthase from cucumber cotyledons as a 47-kDa protein that was quite different from the spinach enzyme (10). The contradictory data on the size of the protein is suggestive of misassignment. Here, we elucidate the entire sequence of MGDG synthase and confirm its assignment by functional expression in *Escherichia coli*. The deduced amino acid sequence of the MGDG synthase cDNA shows homology with MurG, of *Bacillus subtilis* and *E. coli*, which encode a glycosyltransferase catalyzing the last step of peptidoglycan synthesis in bacteria (11–13). This sequence homology implies that the machinery of chloroplast membrane biosynthesis is evolutionarily derived from that of cell wall biosynthesis in bacteria. This is consistent with the endosymbiotic hypothesis of chloroplast formation.

MATERIALS AND METHODS

Materials. UDP-[^{3}H]galactose (1.85 TBq/mmol) and UDP-[^{14}C]galactose (11.1 GBq/mmol) were purchased from Du-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

0027-8424/97/94333-5$2.00/0
PNAS is available online at **http://www.pnas.org**.

Abbreviations: MGDG, monogalactosyldiacylglycerol; DGDG, digalactosyldiacylglycerol; GST, glutathione *S*-transferase; IPTG, isopropyl-β-D-thiogalactopyranoside.

Data deposition: The sequence reported in this paper has been deposited in the GenBank data base (accession no. U62622).

†To whom reprint requests should be addressed.

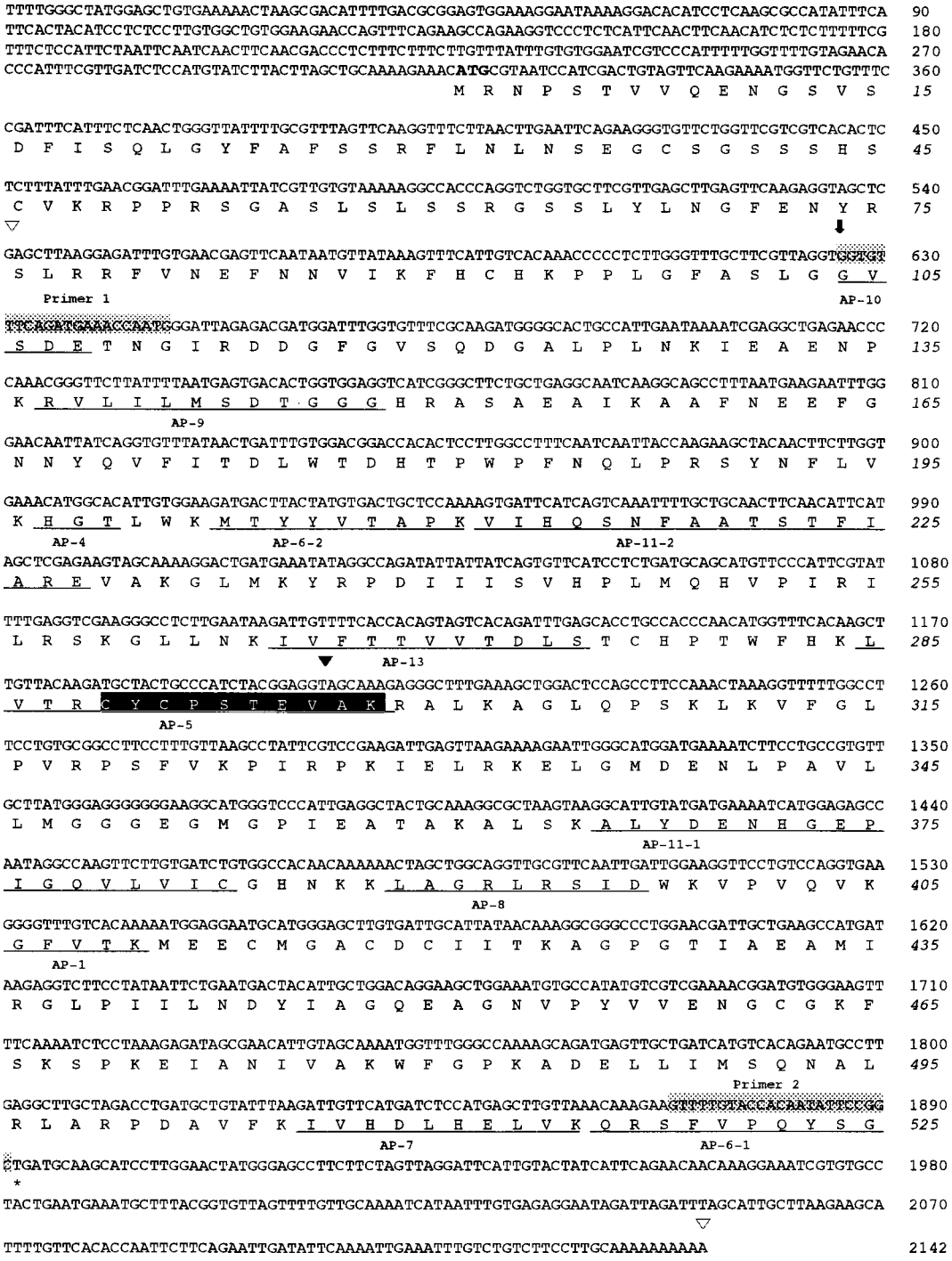

FIG. 1. Nucleotide and deduced amino acid sequences of the cucumber MGDG synthase. Amino acid sequences of peptides obtained from the purified protein are underlined. Numbering begins at the first nucleotide of the sequence. The putative N terminus of the mature protein region is indicated by arrow. Open triangles indicate the 5′ and 3′ ends of the first clone obtained. Closed triangles indicate those of PCR products by 5′ RACE.

Pont/NEN. Unlabeled UDP-galactose was obtained from Sigma. All other reagents were of analytical grade.

Peptide Sequencing. Purified enzyme, as described (10), was separated on 12% SDS/PAGE and then transferred onto polyvinylidene fluoride (PVDF) membrane (ProBlott, Applied Biosystems). The 47-kDa protein was excised and digested with *Achromobacter* protease I. Peptides were separated by reverse-phase HPLC, and subsequently the amino acid sequences were determined by the method of Iwamatsu (14).

cDNA Cloning of MGDG Synthase. A λgt11 cDNA library of cucumber (*Cucumis sativus* L. cv. Aonagajibai) was a gift from R. Tanaka. The library was prepared from poly(A)$^+$ RNA of 5-day-old dark seedlings illuminated for 6 h.

A mixture of ^{32}P-labeled degenerate oligonucleotide probes from a 10-aa sequence of the highlighted peptide (CYCPSTEVAK) in Fig. 1 was used to screen the cucumber λgt11 cDNA library. The sequence of the synthesized oligonucleotide probe was 5′-TGYTAYTGYCCIWSIACIGARGTIGCIAAR (IUB code). Hybridizations were carried out at 60°C in 5× standard saline citrate (SSC)/10 mM EDTA/10× Denhardt's solution containing 250 μg of salmon sperm DNA per ml. Filters were washed at 45°C in 5× SSC/0.1% SDS and autoradiographed.

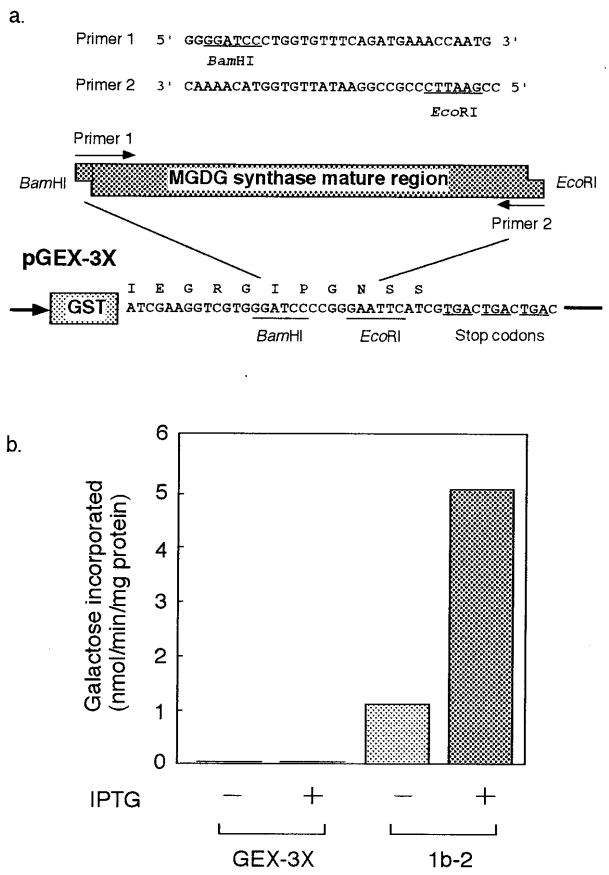

FIG. 2. MGDG synthase expressed in *E. coli*. (*a*) Construction of plasmid with GST fusion expression vector of pGEX-3X to generate the mature region of MGDG synthase as a fusion protein with GST. (*b*) MGDG synthase activity measured by the formation of [^{3}H]MGDG by cell-free extract of *E. coli* XL1-Blue transformed with pGEX-3X alone or pGEX-3X harboring the cDNA clone. We designated the expression vector pGEX-3X harboring the cDNA clone as pGEX-GT. GEX-3X and 1b-2 indicate *E. coli* XL1-Blue transformed with pGEX-3X and pGEX-GT, respectively. Plus (+) and minus (−) indicate the crude extracts from XL1-Blue grown in the presence or absence of IPTG, respectively.

λphage DNA from positive plaques was purified, and the cDNA inserts were subcloned into *Not*I site of pBluescript II SK$^+$ for sequencing.

The 5′ end of the MGDG synthase cDNA was cloned by the rapid amplification of cDNA ends (RACE) protocol using the Marathon cDNA amplification Kit (CLONTECH). Total RNA was isolated from cucumber cotyledons grown in darkness for 4 days and illuminated for 29 h. Poly(A)$^+$RNA was purified from total RNA with oligo(dT)-Latex (Takara Shuzo, Kyoto). Thirty cycles of PCR were carried out at 94°C for 1.5 min, 57°C for 2.5 min, and 72°C for 3 min with 10 pmol of 5′ end adaptor primer and 10 pmol of 5′ end amplification primer (nucleotides between 1179 and 1203; ACCTCCGTAGATGGGCAGTAGCATC). The amplified PCR products were cloned into pCR II vector using a TA Cloning Kit (Invitrogen) for sequencing.

Sequencing and Analysis of Cloned cDNA. Sequencing was performed by the cycle sequencing method, using kits from Li-Cor (Lincoln, NE), according to the supplier's instructions. A Li-Cor DNA sequencer was used. Sequence comparisons against the GenBank sequence data bases were performed using the BLAST programs (15).

Stable Expression of MGDG Synthase in *E. coli*. The PCR primers for the amplification of the mature form of MGDG synthase were primer 1 (5′-GGGGATCCCTGGTGTTTCA-

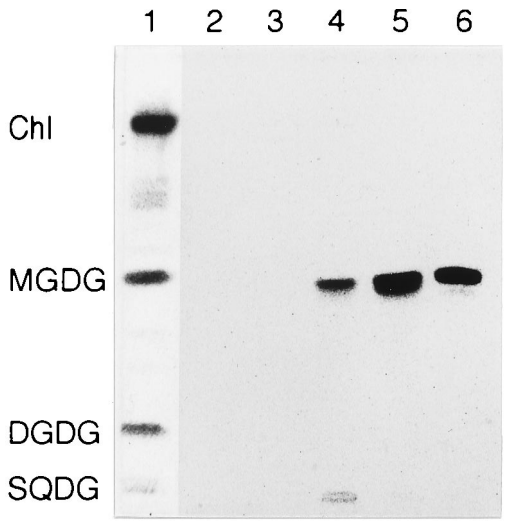

FIG. 3. TLC analysis of galactosylation products produced by the expressed fusion protein of MGDG synthase and GST in *E. coli* extracts. Lane 1, total lipids extracted from spinach, which were stained with anthrone–sulfuric acid; lanes 2 and 3, the reaction products of the extract from GEX-3X, which are untreated and treated with IPTG (1 mM final concentration), respectively; lanes 4 and 5, the reaction products of the extract from 1b-2, which are untreated and treated with IPTG, respectively; and lane 6, the reaction products of MGDG synthase purified from cucumber cotyledons. SQDG, sulfoquinovosyldiacylglycerol.

GATGAAACCAATG-3′) and primer 2 (5′-CAAAACATGGTGTTATAAGGCCGCCCTTAAGCC-3′) (see Figs. 1 and 2). The cDNA was subjected to 30 cycles of PCR amplification (1.5 min at 94°C, 2 min at 60°C, and 3 min at 72°C) according to the manufacturer's instructions. The amplified cDNA was digested with *Bam*HI and *Eco*RI and inserted into the expression vector pGEX-3X. XL1-Blue cells were transformed with the construct and used for expression.

The transformed XL1-Blue (Stratagene) was inoculated to 5 ml of Luria–Bertani (LB) medium and grown at 37°C for 16 h. Approximately 0.5 ml of the culture was added to 50 ml of LB medium and grown at 37°C for 3 h. The fusion protein was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG; 1 mM final concentration), and cells were harvested after 2 h by centrifugation (5 min at 3000 rpm). The cell pellet was resuspended in 5 ml of 50 mM Mops-NaOH buffer (pH 7.9) containing 10 mM dithiothreitol (MOD buffer) and sonicated. The sonicated suspension was centrifuged (5 min at 12,000 rpm) and the supernatant (cell-free extract) was used for the assay of MGDG synthase.

Assay for MGDG Synthase Activity. MGDG synthase activity was assayed by the method of Teucher and Heinz (7) with 400 μM UDP-galactose as substrate. The solution of enzyme (30 μl) was mixed with 50 μl of 1,2-dioleoyl-*sn*-glycerol [200 μg; dispersed in 0.01% (wt/vol) Tween 20], and brought to 190 μl with MOD buffer. The mixture was preincubated at 30°C for 5 min, and then 10 μl of 8 mM UDP-4,5-[^{3}H]galactose (92.5 Bq·nmol^{-1}) was added. The reaction mixture was incubated at 30°C for 15 min, and the reaction was stopped by the addition of 1 ml of ethyl acetate. After the reaction mixture had been stirred vigorously with vortex mixer, 0.5 ml of 0.45% (wt/vol) NaCl was added, and the mixture was centrifuged at 500 × *g* for 5 min. The lower layer was removed, and another 0.5 ml of 0.45% (wt/vol) NaCl was added. The mixture was centrifuged, and the upper layer was transferred to a scintillation vial that contained scintillation fluid (4 ml; ACS-II, Amersham) and methanol (0.5 ml). Radioactivity was determined in a liquid scintillation counter (model LS6500CE, Beckman–RIIC, Scotland).

Table 1. Change in lipid composition of *E. coli*

Transformant of *E. coli*	Lipid, mol%			
	PG	CL	PE	MGDG
GEX-3X	13.8	5.9	80.3	—
1b-2	7.6	5.0	70.0	17.4

PG, phosphatidylglycerol; CL, cardiolipin; PE, phosphatidylethanolamine.

DNA Manipulations. *E. coli* XLI-Blue was used in the cloning experiments. Standard molecular cloning techniques were used (16).

Protein Assay. Proteins were quantitated by the method of Bensadoun and Weinstein (17) with bovine serum albumin as the standard.

Lipid Analysis. The ethyl acetate fraction obtained as above (assay for MGDG synthase activity) was evaporated *in vacuo* and the residual lipids were analyzed by TLC in chloroform:methanol:water (65:15:2, vol/vol/vol). UDP-[^{3}H]galactose, the substrate used in the usual assay mentioned above, was replaced with UDP-[^{14}C]galactose to detect the reaction products by Image Analyzer (Fujix BAS2000, Fuji Photo Film). As a control, total lipids from spinach leaves were extracted by the method of Bligh and Dyer (18) and separated by TLC, simultaneously. The glycolipids were stained by spraying anthrone–sulfuric acid.

RESULTS AND DISCUSSION

Molecular Cloning of a cDNA Encoding MGDG Synthase. Purified MGDG synthase from cucumber (10) was *S*-carboxymethylated and digested by lysylendopeptidase (*Achromobacter* protease I) after blotting on PVDF membrane. The amino acid sequence of 12 peptides (Fig. 1, AP-1–AP-13) was determined (14). A cucumber λgt 11 cDNA library was screened with the 30-mer oligonucleotide based on the peptide sequence of AP-5 (Fig. 1, highlighted), and a 1647-bp cDNA clone (base pairs 496-2142, indicated in Fig. 1) was isolated. Although the cDNA clone encoded all of the peptide sequences identified in the native protein (Fig. 1), it seemed to be truncated at the 5′ end, since it lacked a readily identifiable initiation codon. We subsequently amplified the region of the 5′ end of the full-length cDNA using 5′ RACE (CLONTECH). Fig. 1 illustrated the combined nucleotide and amino acid sequences of MGDG synthase from 5′ end to the poly(A) tail. The 2142-bp cDNA clone contains a 1575-bp open reading frame between nucleotides 317 and 1891, encoding 525 aa. The ATG codon at position 317 is identified as the initiation site for translation and Gly-104 is probably the N terminus of the mature protein, because lysine residue was not connected to the peptide (AP-10) at N terminus although the protein was digested by lysylendopeptidase. The molecular weight predicted from the mature sequence (46,552) matches that estimated by SDS/PAGE (46,000 ± 1000; ref. 10). Since MGDG synthase is present in envelope membranes from all chloroplasts and nongreen plastids (4), the leader peptide of 103 amino acid residues upstream of Gly-104 is likely to be a transit peptide required for import into the chloroplast envelope.

Expression of Enzymatically Active MGDG Synthase in *E. coli*. To confirm that the cDNA encodes MGDG synthase, a fusion protein with glutathione *S*-transferase (GST) was expressed in *E. coli*, and subsequently the MGDG synthase activity was analyzed. The mature region of the clone (1266 bp between nucleotides 626 and 1891) was ligated with pGEX-3X vector to produce MGDG synthase (without the first 103 aa) fused downstream of GST in *E. coli* XL1-Blue. The plasmid construct is shown in Fig. 2*a*. We term this transformant 1b-2. The expression of the fusion protein as a 72-kDa polypeptide was confirmed by SDS/PAGE. Cell-free extracts of 1b-2 exhibited galactose incorporation activity even without IPTG induction (Fig. 2*b*). Additionally, an ≈8-fold increase in MGDG synthase activity was observed after addition of IPTG. There was no activity detected in cell-free extracts from GEX-3X, transformant expressing only GST (Fig. 2*b*). The reaction products were further analyzed by TLC. Fig. 3 clearly shows that the cell-free extracts of the 1b-2 specifically produce MGDG. These results provide direct evidence that the cDNA clone obtained here is MGDG synthase.

Changes in the Lipid Composition of the *E. coli* Transformant, 1b-2. These results also suggest that the *E. coli* transformant could be producing MGDG *in vivo* using the fusion protein of MGDG synthase and GST. Therefore, we analyzed the lipid composition of the membranes of GEX-3X and 1b-2 (Table 1). Surprisingly, ≈17 mol% of the total lipids of the membrane contained MGDG in 1b-2 cells, while MGDG was not detected in the membrane lipids of

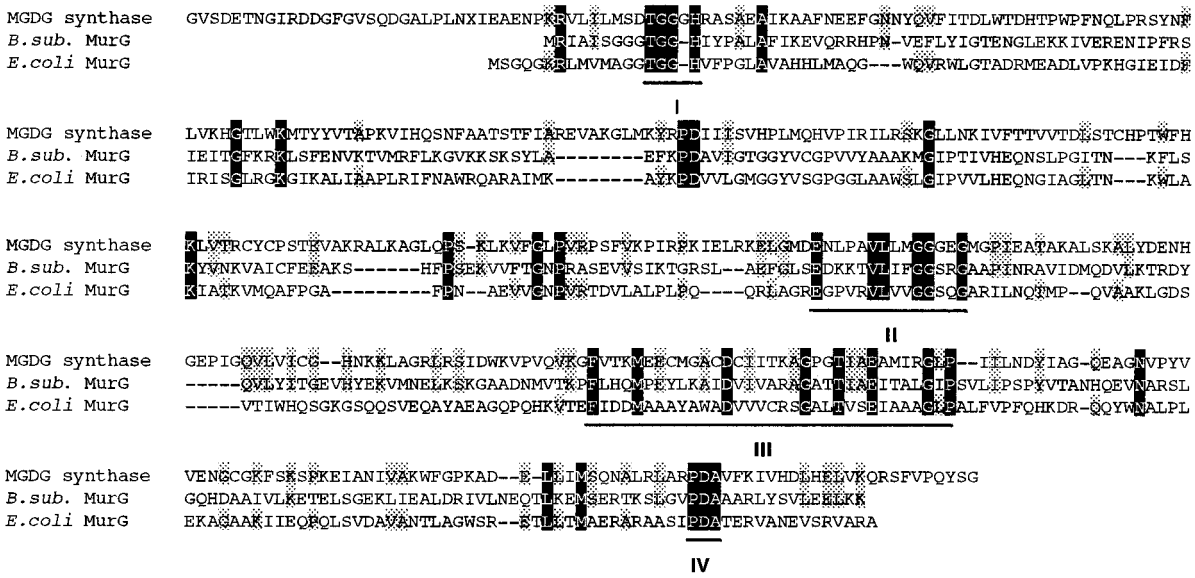

FIG. 4. Alignment of the deduced amino acid sequence of MGDG synthase (mature region) with those of the MurG protein from *B. subtilis* (12) and *E. coli* (19, 20). Amino acids that are identical between MGDG synthase and the other proteins are highlighted in black. Four typical homologous regions (I, II, III, and IV) are indicated by underlining.

GEX-3X. The growth of 1b-2 cells was much slower than that of GEX-3X cells.

A Homology of the Deduced Amino Acid Sequence of MGDG Synthase cDNA with That of MurG in Bacteria. A homology search of all available databases revealed sequence similarity of MGDG synthase to the *murG* genes of *B. subtilis* and *E. coli*. These encode a glycosyltransferase involved in the biosynthesis of bacterial peptidoglycan. The deduced amino acid sequence of MGDG synthase shares 23.3% and 19.0% identity over 258 and 116 aa for the MurG protein in *B. subtilis* (12) and *E. coli* (19, 20), respectively (Fig. 4). *MurG* gene encodes UDP-*N*-acetylglucosamine:*N*-acetylmuramyl (pentapeptide) pyrophosphoryl-undecaprenol *N*-acetylglucosamine transferase (11). The *N*-acetylglucosaminyltransferase is responsible for the final step in the formation of the lipid-linked disaccharide-pentapeptide subunit of peptidoglycan, a universal bacterial cell wall constituent (21). The deduced amino acid sequence of MGDG synthase cDNA showed negligible homology with any other prokaryotic and eukaryotic glycosyltransferases. Comparison of the hydropathy plot with MurG protein of *B. subtilis* also demonstrated that the profiles are not identical but substantially similar between these proteins. MurG proteins are membrane proteins that are weakly associated with the inner face of the bacterial cytoplastic membranes (13). In addition, MurG proteins and MGDG synthase are rich in lysine residues and have very high theoretical pI values ≈9.5, suggesting that these proteins are weakly associated with the membranes by electrostatic interaction. In fact, the purified MGDG synthase is a basic protein (10).

From evolutionary considerations of chloroplast biogenesis, the possibility exists that, following symbiosis between a bacterium (e.g. cyanobacterium) and a host cell, the *murG* gene was transferred to the nucleus of the host and its function diverged to catalyze the final step of MGDG synthesis instead of peptidoglycan synthesis. In the process, these bacterial genes could be integrated into eukaryotic cells and be transferred to the nucleus accompanying functional alterations. The results shown here provide additional evidence for the endosymbiotic origin of the chloroplast, and clearly resolves the differences in molecular weight reported in the literature for this key enzyme in thylakoid lipid synthesis.

We thank Mr. Ryoichi Tanaka (Kyoto University) for providing cDNA library and Professor Antoni R. Slabas (University of Durham) for critical reading of the manuscript. This work was supported by a Grant-in-Aid for Scientific Research on Priority Areas (04273101) from the Ministry of Education, Science and Culture, Japan.

1. Murphy, D. J. (1986) *Biochim. Biophys. Acta* **864,** 33–94.
2. Murphy, D. J. (1982) *FEBS Lett.* **150,** 19–26.
3. Murata, N., Higashi, S. & Fujimura, Y. (1990) *Biochim. Biophys. Acta* **1019,** 261–268.
4. Joyard, J. & Douce, R. (1987) in *Biochemistry of Plants*, eds. Stumpf, P. K. (Academic, New York), pp. 215–274.
5. Heemskerk, J. W. M., Storz, T., Schmidt, R. R. & Heinz, E. (1990) *Plant Physiol.* **93,** 1286–1294.
6. Ohta, H., Shimojima, M., Ookata, K., Masuda, T., Shioi, Y. & Takamiya, K. (1995) *Plant Cell Physiol.* **36,** 1115–1120.
7. Teucher, T. & Heinz, E. (1991) *Planta* **184,** 319–326.
8. Maréchal, É., Block, M. A., Joyard, J. & Douce, R. (1991) *C. R. Acad. Sci.* **313,** 521–528.
9. Covès, J., Joyard, J. & Douce, R. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 4966–4970.
10. Ohta, H., Shimojima, M., Arai, T., Masuda, T., Shioi, Y. & Takamiya, K. (1995) in *Plant Lipid Metabolism*, eds. Kader, J.-C. & Mazliak, P. (Kluwer Academic, Dordrecht, The Netherlands), pp. 152–155.
11. Mengin-Lecreulx, D., Laurent, T., Rousseau, M. & van Heijenoort, J. (1991) *J. Bacteriol.* **173,** 4625–4636.
12. Miyao, A., Yoshimura, A., Sato, T., Yamamoto, T., Theeragool, G. & Kobayashi, Y. (1992) *Gene* **118,** 147–148.
13. Bupp, K. & van Heijenoort, J. (1993) *J. Bacteriol.* **175,** 1841–1843.
14. Iwamatsu, A. (1992) *Electrophoresis* **13,** 142–147.
15. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
16. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
17. Bensadoun, A & Weinstein, D. (1976) *Anal. Biochem.* **70,** 241.
18. Bligh, E. G. & Dyer, W. J. (1959) *Can. J. Biochem. Physiol.* **37,** 911–917.
19. Mengin-Lecreulx, D., Texier, L. & van Heijenoort, J. (1990) *Nucleic Acids Res.* **18,** 2810.
20. Ikeda, M., Wachi, M., Jung, H. K., Ishino, F. & Matsuhashi, M. (1990) *Nucleic Acids Res.* **18,** 4014.
21. Roten, C.-A. H. & Karamata, D. (1992) *Experientia* **48,** 921–931.