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-B-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 Stransferase. 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-galactosyl-transferase; 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

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.

Copyright @ 1997 by The National Academy of Sciences of the USA 0027-8424/97/94333-52.00/0

PNAS is available online at http://www.pnas.org.

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-[(3cholamidopropyl) 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-[³H]galactose (1.85 TBq/mmol) and UDP-[¹⁴C]galactose (11.1 GBq/mmol) were purchased from Du-

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.

CCA																												ACA	2
	TTTCO	STTG	ATC	TCC	ATG	TAT	CTT	ACT	TAG	CTG	CAA	AAG	AAA																3
														м	R	N	P	S	т	V	V	Q	Е	N	G	Ş	v	s	1
GAT	TTCAT	TTC	TCA	ACT	GGG	TTA	TTT	TGC	GTT	TAG	TTC	AAG	GTI	TCT	TAA	CTT	GAA	TTC	AGA	AGG	GTG	TTC	TGG	TTC	GTC	GTC		ACTC	4
D	FI	s	Q	L	G	Y	F	A	F	s	s	R	F	L	N	L	N	s	Е	G	С	s	G	s	s	s	Ħ	s	4
vጉምም	TATT	אמביי	ccc	አመመ	መርገል	אתת	መመእ	TCC	መመረግ		תתת		~~~~	ACC	<u> </u>	CILC		mee	000	~~~~~	~~~	COR			~~~	200			5
	V K																												5
7																											Ŧ		
AGC	TTAAC	GAG	ATT	TGT	GAA	CGA	GTT	CAA	TAA	TGT	TAT	ААА	GTT	TCA	TTG	тса	CAA	ACC	ccc	TCT	TGG	GTT	TGC	TTC	GTT	AGG	TGC	TGT	6
S	L R	R	F	v	N	Е	F	N	N	v	I	ĸ	F	Ħ	С	Н	K	₽	₽	L	G	F	A	s	\mathbf{r}	G	G	v	1
	?rime:	: 1																									A	2-10	
TCA	JATG/	AAC	CAA	ŦGG	GAT	TAG	AGA	CGA	TGG	ATT	TGG	TGT	TTC	GCA	AGA	TGG	GGC.	ACT	GCC	ATT	GAA	TAA	AAT	CGA	.GGC	TGA	GAA	ACCC	7
<u>s</u>	DE	T	N	G	I	R	D	D	G	F	G	v	s	Q	D	G	Α	L	Ρ	\mathbf{r}	N	ĸ	I	Е	Α	Е	N	Ρ	1
מממ	CGGGI	ነመሮጥ	ተካልጣ	ատա	አልጥ	CAC	ጦር እ	CDC	maa	TCC	NGG	መር እ	The second	ccc	መምርሳ	maa	ሞር እ	ccc	አአጥ	<u>ر م م</u>	ccc	200	ر مىسىت	ממחיי	ሞርን	N C N	አመጣ	maa	8
	<u>r v</u>																												1
					P-9																								
	AATT																												9
N	N Y	Q	v	F	I	т	D	L	W	т	D	H	т	Ρ	W	Ρ	F	N	õ	L	Ρ	R	s	Y	N	F	L	v	1
ААА	CATGO	CAC	АТТ	GTG	GAA	GAT	GAC	ጥጥል	ста	TGT	GAC	TGC	TCC	ممم	AGT	GAT	TCA	TCA	GTC	ΔΔΑ	անան	TGC	TGC	AAC	ͲͲϹ	ААС	ΔΨΤ	ርልሞ	9
	H G																												2
	AP-4							AP-												AP-1									
	CGAGA																												1
<i>A</i>	R E	. V	Α	K	G	Ļ	М	K	Y	R	₽	D	I	I	I	s	v	H	P	L	М	Q	H	v	₽	I	R	I	2
гтG	AGGTC	GAA	GGG	сст	CTT	GAA'	таа	GAT	ͲႺͲ	արդար,	CAC	CAC	АСТ	AGT	CAC	AGA	ጥጥጥ	GAG	CAC	стб	CCA	ccc	AAC	ATG	GͲͲ	тса	саа	GCT	-
5	r s	ĸ		\mathbf{L}	L	N	К	I	v	F	Т	Т	v	v	T	D	L	S	т	С	Η	Ρ	т	W	F	Η	к		
L :	R S	K		L	L	N	ĸ	I	v	F		T AP-1		<u>v</u>	<u> </u>	D	L	s	т	С	Η	Ρ	т	W	F	H	к		
GTT.	ACAAG	ATG	G CTA	CTG	ccc.	ATC	TAC	GGA	GGT	AGC.	r AAA	AP-1 GAG	.3 GGC	TTT	GAA	AGC	TGG	ACT	CCA	GCC	TTC	CAA	ACT	AAA	ggt	TTT	TGG	<u> </u>	2 1
GTT.		ATG	G CTA	CTG C	CCC. P	ATC	TAC	GGA	GGT	AGC.	r AAA	AP-1 GAG	.3 GGC	TTT		AGC	TGG	ACT	CCA	GCC	TTC	CAA	ACT	AAA	ggt	TTT	TGG	<u> </u>	2 1
GTT. V	ACAAG FR	ATG C	G CTA Y	CTG C AP-	ссс. Р -5	ATC' S	TAC T	GGA E	GGT V	AGC. A	AAA K	AP-1 GAG _R	.3 GGC A	TTT L	GAA. K	AGC A	TGG. G	ACT L	CCA Q	GCC P	TTC S	CAA K	ACT L	AAA K	GGT V	TTT F	TGG G	L CCT L	2 1 3
STT. V	ACAAG	ATG C	G CTA Y TTC	CTG C AP- CTT	CCC. P -5 TGT	ATC S TAA	TAC T GCC	GGA E TAT	GGT V TCG	AGC. A	AAA K GAA	AP-1 GAG _R GAT	.3 GGC A TGA	TTT L GTT	gaa K Aagi	AGC A AAA	TGG. G AGA	ACT L ATT	CCA Q GGG	GCC P CAT	TTC S GGA	CAA K TGA	ACT L AAA	AAA K TCT	GGT V TCC	TTT F TGC	TGG G CGI	L CCT L	2 1 3 1
GTT. V CCT P	ACAAG <u>FR</u> GTGCG VR	ATG C GCC P	G CTA Y TTC S	CTG AP- CTT F	CCC. P -5 TGT V	ATC S TAA K	TAC T GCC P	GGA E TAT I	GGT V TCG R	AGC. A TCC P	AAA K GAA K	AP-1 GAG _R GAT I	.3 GGC A TGA E	TTT L GTT. L	GAA K AAG R	AGC A AAA K	TGG G AGA E	ACT L ATT L	CCA Q GGG G	GCC P CAT M	TTC S GGA D	CAA K TGA E	ACT L AAA N	AAA K TCT L	GGT V TCC P	TTT F TGC A	TGG G CGI V	L CCT L GTT L	2 1 3 1
GTT.	ACAAG <u>FR</u> STGCG VR ATGGG	ATG C GCC P	G CTA Y TTC S GGG	CTG AP- CTT F GGA	CCC. P -5 TGT V AGG	ATC S TAA K CATO	TAC T GCC P GGG	GGA E TAT I TCC	GGT V TCG R CAT	AGC. A TCC P TGA	AAA K GAA K GGC	AP-1 GAG R GAT I TAC	.3 GGC A TGA E TGC	TTT L GTT. L AAA	GAAI K AAGI R GGC0	AGC A AAAA K GCT.	TGG G AGA E AAG	ACT L ATT L	CCA Q GGG G GGC.	GCC P CAT M ATT	TTC S GGA D GTA	CAA K TGA E TGA	ACT L AAA N TGA	AAA K TCT L AAA	GGT V TCC P TCA	TTT F TGC A TGG	TGG G CGI V AGA	L CCT L CGTT L AGCC	2 1 3 1 3
GTT.	ACAAG <u>FR</u> GTGCG VR	ATG C GCC P	G CTA Y TTC S GGG	CTG AP- CTT F GGA	CCC. P -5 TGT V AGG	ATC S TAA K CATO	TAC T GCC P GGG	GGA E TAT I TCC	GGT V TCG R CAT	AGC. A TCC P TGA	AAA K GAA K GGC	AP-1 GAG R GAT I TAC	.3 GGC A TGA E TGC	TTT L GTT. L AAA	GAAI K AAGI R GGC0	AGC A AAAA K GCT.	TGG G AGA E AAG	ACT L ATT L	CCA Q GGG G GGC.	GCC P CAT M ATT L	TTC S GGA D GTA Y	CAA K TGA E TGA D	ACT L AAA N TGA	AAA K TCT L AAA	GGT V TCC P TCA	TTT F TGC A TGG	TGG G CGI V AGA	L CCT L CGTT L AGCC	2 1 3 1 3
GTT. V CCT P CTT.	ACAAG <u>FR</u> STGCG VR ATGGG	ATG C GCC P AGG G	G CTA Y TTC S GGG G	CTG AP- CTT F GGA E	CCC. P -5 TGT V AGG	ATC S TAA K CAT M	TAC T GCC P GGG G	GGA TAT I TCC P	GGT V TCG R CAT I	AGC A TCC P TGA E	AAA K GAA K GGC A	AP-1 GAG R GAT I TAC T	.3 GGC TGA E TGC A	TTT L GTT. L AAA K	GAA K AAG R GGCC A	AGC A AAAA K GCT. L	TGG G AGAI E AAG' S	ACT L ATT L TAA	CCA Q GGG G GGC A	GCC P CAT M ATT L A	TTC S GGA D GTA <u>Y</u> P-11	CAA K TGA E TGA D	ACT L AAA N TGA E	AAA K TCT L AAA N	GGT V TCC P TCA H	TTT F TGC A TGG G	TGG G CGI V AGA E	L CCT L CGTT L AGCC P	1 3 1 3 1 3
GTT. V CCT P CTT. L I	ACAAG <u>FR</u> GTGCG VR ATGGG 4 G	ATG C GCC P AGG G	G CTA Y TTC S GGG G TCT	CTG AP CTT F GGA E TGT	CCC P -5 TGT V AGG G	ATC S TAA K CATC M CTG	TAC T GCC P G G TGG	GGA E TAT I TCC P CCA	GGT V TCG R CAT I CAA	AGC A TCC P TGA E CAA	AAA K GAA K GGC A AAA	AP-1 GAG R GAT I TAC T	.3 GGC A TGA E TGC A AGC	TTT L GTT L AAA K TGG	GAA K AAG R GGCC A CAG	AGC A AAAA K GCT. L	TGG G AGA E AAG' S GCG	ACT L ATT L TAA K	CCA Q GGG G GGC. A AAT	GCC P CAT M ATT L A TGA	TTC S GGA D GTA <u>Y</u> P-11 TTG	CAA K TGA E TGA D L-1 GAA	ACT L AAA N TGA E GGT	AAA K TCT L AAA N	GGT V TCC P TCA H TGT	TTT F TGC A TGG G CCA	TGG G CGI V AGA E GGT	L CCT L CGTT L AGCC P	2 1 3 1 3 1 3 1 3
GTT. V CCT P CTT. L I I	ACAAG FR GTGCG / R ATGGG 4 G GGCC <i>P</i> <u>3 O</u>	ATG C GCC P AGG G AGT V	G CTA Y TTC S GGGG G TCT L	CTG AP- CTT F GGA E TGT V	CCC. P -5 TGT' V AGG G G AGG I	ATC: S TAAC K CATC M CTG: C	TAC T GCCC P G G G G G G	GGA TAT I TCC P CCA H	GGT V TCG R CAT I CAA N	AGC. A TCC P TGA E CAA K	AAA K GAA K GGCC A AAAA K	AP-1 GAG R GAT I TAC T ACT	3 GGC A TGA E TGC A AGC <u>A</u>	TTT L GTT. L AAAA K TGG G	GAA K AAG R GGCC A CAGC A R AP-8	AGC A AAAA K GCT. L GTT L	TGG G AGA E AAG' S GCG' R	ACT L ATT L TAA K S	CCA Q GGG G GGC. A AAT I	GCC P CAT M ATT L A TGA D	TTC S GGA D GTA <u>Y</u> P-11 TTG W	CAA K TGA E TGA L-1 GAA K	ACT L AAA N TGA E GGT V	AAA K TCT L AAA N TCC P	GGT V TCC P TCA H TGT V	TTT F TGC A TGG G CCA Q	TGG G CGI V AGA E GGT V	L CCT L CGTT L AGCC P CGAA K	2 1 3 1 3 1 3 1 4
GTT.	ACAAG <u>F</u> R GTGCC VR ATGGG 4 G 3 GCC 3 O FTTGT	ATG C GCCC P G G G AGG C CAC	G CTA Y TTCC S GGGG G TCT L AAAA	CTG AP- CTT F GGA E TGT V AAT	CCC. P -5 TGT' V AGGG G GATG -1 	ATC S TAAG K CATG M CTG GGA	TAC T GCC P G G G G G ATG	GGA TAT I TCCC P CCA	GGT V TCG R CAT I CAA N GGGG	AGC. A TCCC P TGA E CAA K AGC	AAAA K GAA K GGGC A AAAA K TTTG	AP-1 GAG R GAT I TAC T ACT. L TGA	3 GGC A TGA E TGC A AGC A TTG	TTT L GTT. L AAAA K TGG G ; CAT	GAAJ K R GGGCO A CAGO R AP-8 TATZ	AGC A AAAA K GCT. L GTT L	TGG G AGAJ E S GCG R AAAA	ACT L L TAA K S GGC	CCA Q GGG G GGC A AAT I GGG	GCC P CAT M ATT L ATT GA D CCC	TTC: S GGA D GTA Y P-11 TTG: W	CAA K TGA D L-1 GAA K AAC	ACT L AAA N TGA E GGT V GAT	AAA K TCT L AAA N TCC P TGC	GGT V TCC P TCA H TGT V	TTT F TGC A TGG G CCA Q AGC	TGG G CGI V AGA E GGT V CAT	L CCT L CGTT L AGCC P CGAA K	2 1 3 1 3 1 3 1 4 4 1
GTT. CCT CTT. L 1 ATA I GGG	ACAAG FR GTGCG / R ATGGG 4 G GGCC <i>P</i> <u>3 O</u>	ATG C GCCC P G G G AGG C CAC	G CTA Y TTCC S GGGG G TCT L AAAA	CTG AP- CTT F GGA E TGT V AAT	CCC. P -5 TGT' V AGGG G GATG -1 	ATC S TAAG K CATG M CTG GGA	TAC T GCC P G G G G G ATG	GGA TAT I TCCC P CCA	GGT V TCG R CAT I CAA N GGGG	AGC. A TCCC P TGA E CAA K AGC	AAAA K GAA K GGGC A AAAA K TTTG	AP-1 GAG R GAT I TAC T ACT. L TGA	3 GGC A TGA E TGC A AGC A TTG	TTT L GTT. L AAAA K TGG G ; CAT	GAAJ K R GGGCO A CAGO R AP-8 TATZ	AGC A AAAA K GCT. L GTT L	TGG G AGAJ E S GCG R AAAA	ACT L L TAA K S GGC	CCA Q GGG G GGC A AAT I GGG	GCC P CAT M ATT L A TGA D CCC	TTC: S GGA D GTA Y P-11 TTG: W TGG.	CAA K TGA D L-1 GAA K AAC	ACT L AAA N TGA E GGT V GAT	AAA K TCT L AAA N TCC P TGC	GGT V TCC P TCA H TGT V	TTT F TGC A TGG G CCA Q AGC	TGG G CGI V AGA E GGT V CAT	L CCT L CGTT L AGCC P CGAA K	2 1 3 1 3 1 3 1 4 4 1
GTT. V CCT CTT. L I GGGG G	ACAAG <u>F</u> R GTGCG VR ATGGG 4 G 3 GCC 3 O FTTGT F V	ATG GCC P AGG G AGG G AGT V CAC	G CTA Y TTCC S GGGG G TCT L AAAA K	CTG AP- CTT F GGA E TGT V AAT M	CCC. P -5 TGT' V AGGG G GAT' I GGAT E	ATC S TAAC K CATC M CTG C C GGAI E	TAC T GCC P G G G G ATG C	GGA TAT I TCC P CCA H CAT	GGT V TCG R CAT I CAA N GGG G G	AGC. A TCCC P TGA E CAA K K AGC' A	J AAAA K GGAA K GGCC A AAAA K TTTG C	AP-1 GAG R GAT I TAC T ACT. L J	3 GGC A TGA E TGC A AGC A TTG C	TTT L GTT. L AAAA K TGG G I CAT I	GAAJ K R GGCC A CAGC R AP-8 TATZ I	AGC A AAAA K GCT. L GTT. L AAC. T	TGG. G AGAI E AAAG' S GCG' R AAAA K	ACT L ATT L TAA K S GGC A	CCA Q GGG G GGC. A AAT I GGG G	GCC P CAT M ATT IGA D CCC P	TTC: S GGA D GTA Y P-11 TTG: W TGG. G	CAA K TGA E I-1 GAA K AAC T	ACT L AAAA N TGA E GGT V GAT I	AAA K TCT L AAA N TCC P TCC A	GGT V TCC P TCA H TGT V TGA E	TTT F TGC A TGG G CCA Q AGC A	TGG G CGI V AGA E GGI V CAT M	L CCT L CGTT L AGCC P CGAA K CGAT I	2 1 3 1 3 1 3 1 4 4 1 4
GTT.	ACAAG FR GTGCG VR ATGGG ATGGG ATGGG ATGGG C ATGGG C ATGGG ATGGG ATGGG ATGGG ATGGG ATGGG ATGGG ATGGG ATGGG ATGGG ATGGG ATGGG ATGGG ATGGG ATGGGG ATGG ATGGG ATGG ATGGG ATGG ATGGG ATGGG ATGG ATGGG ATGGGGG	ATG GCC P AGG G AGG G AGT V CAC	G CTA Y TTC S GGG G G TCT L AAA K TAT	CTG AP- CTT F GGA E TGT V AAT M AAT	CCCC. P -5 TGT' V AGGG G G I GGA(E TCT(ATC: S TAAG K CATC M CTG: C C GGAA	TAC T 3GCC P 3GG G TGG G TGG C IGA	GGA TAT I TCC P CCA H CAT M CTA	GGT V TCG R CAT I CAA N GGG G G CAT	AGC. A TCCC P TGA E CAA K AGC A TGC	AAA K GAA K GGCC A AAAA K TTGG	AP-1 GAG R GAT I TAC T ACT. L D ACA	3 GGC A TGA E TGC A AGC A TTG C GGA	TTT L GTT. L AAAA K TGGG G ZAT I AGC	GAAA K AAAGI R GGCC A CAGC <u>R</u> I TGGJ	AGC A AAAA K GCT. L GTT L AAAC T	TGG. G AGAJ E AAAG' S GCG' <u>R</u> AAAA(K	ACT L ATT L TAA K S GGC A SGCC	CCA Q GGG G GGC. A AAT GGGG G ATA	GCC P CAT M ATT L A TGA D CCC P	TTC: S GGA D GTA Y P-11 TTG W TGG. G	CAA K TGA E TGA L-1 GAA K AAC T CGA	ACT L AAAA TGA GGT V GAT I AAAA	AAA K TCT L AAA N TCC P TGC A	GGT V TCC P TCA H TGT V TGA E ATG	TTT F TGC A TGG G CCA Q AGC A TGG	TGG G CGI V AGA GGT V CAT M GAA	L CCT L CGTT L AGCC P CGAA K CGAT I SGAT	2 11 3 11 3 11 3 11 4 4 11 4 1
GGG GGG GGG GGGA GGGA	ACAAC <u>F</u> R GTGCC7 R ATGGG 4 G GGCC2 <u>F</u> TTGT <u>F</u> V AP-1 GGTC1 <u>G</u> GTC1 <u>G</u> L	ATG GCC P AGG G AGG CAC. T TCC P	G CTA Y TTC S GGGG G TCT L AAAA K TAT. I	CTG C AP- CTT F GGA E TGT V AAT M AAT I	CCC. P -5 TGT' V AGG G G G G G G G G C C C C C C C C C C C C C	ATC: S TAAG K CATC M CTG' C C GGAA N	TAC T GCC P G G G G G ATG C IGA	GGA TAT I TCCC P CCA H CCA M CTA Y	GGT V TCG R CAT I CAA N GGG G G CAT I	AGC. A TCCC P TGA E CAA. K AGC' A TGCC A	AAAA K GAAA K GGCC A AAAA K TTGG C TGGG G	AP-1 GAG R GAT I TAC T ACT L J ACA Q	3 GGC A TGA E TGC A AGC A C GGA E	TTT L GTT. L AAAA K TGG G , CAT I AGC A	GAAI K AAGJ R GGCCC A CAGC <u>R</u> AP-8 TATJ I TGGJ G	AGC A AAAA K GCT. L GTT. L AAAC: T AAAA' N	TGG G AGAJ E AAAG' S GCG' R AAAAG K TGTG V	ACT L ATT L TAA K S GGC A SGCC, P	CCA Q GGG G G G G G G G G ATA Y	GCC P CAT M TGA D CCC P TGT V	TTC S GGA D GTA P-11 TTG W TGG. G CGT V	CAA K TGA D L-1 GAA K AAAC T CGA E	ACT L AAA N TGA GGT V GGT I AAA N	AAA K TCT L AAAA TCC P TGC A CGG G	GGT V TCC P TCA H TGT V TGA E ATG C	TTT F TGC A TGG G CCA Q AGC A G TGG G	TGG G CGI V AGA E GGT V CAT M GAA K	L CCT L GTT L GGCC P GGAA K GGAT I GTT F	22 11 33 11 33 11 33 11 44 11 44 11 44
GTT.	ACAAC F <u>R</u> STGCC 7 R ATGGG 4 G GGCC 4 G GGCC 4 G GGCC 7 V AP-1 GTCT G L	ATG GCC P AGG G AGG G AGT V CAC. T TCC P	G CTA Y TTC S GGG G G TCT L AAAA X TAT. I TAA	CTG C AP- CTT F GGA E TGT V AAAT M I AAAT	CCC. P -5 TGT' V AGGG G GAT' GAT'	ATC: S TAAG K CATO M CTG? C GGAA S GGAA? N AGCCO	TAC T GCC P G G G G G G ATG C IGA D GAA	GGA E TAT I TCCO P CCA H CCAT M CTA Y CAT	GGT V TCG R CAT I CAA GGG G G CAT I TGT.	AGC. A TCC: P TGA E CAA K CAA K AGC: A AGC.	AAAA K GAA K GGCC A AAAA K TTGG C TGGG G AAAA	AP-1 GAG R GAT I TAC T ACT L D ACA Q ATG	3 GGC A TGA E TGC A AGC A C C GGA E GTT	TTT L GTT. L AAAA K TGGG G J CAT I AGC A TGGG	GAAJ K AAGJ R GGCC A CAGC <u>R</u> AP-8 TATJ I G GGCCJ	AGC A AAAA K GCT. L GTT AAAC T AAAA N AAAA	TGG G AGA E S GCG' R AAAA K TGT V	ACT L ATT L TAA S GGC A GGC A AGA	CCA Q GGG G G G G G G G G G G G ATA Y TGA	GCC P CAT M TGA D CCCC P TGT V GTT	TTC S GGA D GTA P-11 TTG W TGG. G CGT V GCT	CAA K TGA E I-1 GAA K AAC T CGA E GAT	ACT L AAA N TGA E GGT V GAT I AAA N CAT	AAA K TCT L AAAA N TCC P TGC A CGG G CGG	GGT V TCC P TCA H TGT V TGA E C ACA	TTT F TGC A TGG G A G G A G A G A A G G A A	TGG G CGT V AGA E GGT V CAT M GAA K TGC	L CCT L GTT L GGCC P GGAA K GGAT I GTT F CCTT	2 1 3 1 3 1 3 1 3 1 3 1 1 4 4 1 4 1 1
GTT. V CCT P CTT. L I GGG G G G G G G G G G G G G G G C C T C A C A G A G A G A G A C T T.	ACAAC <u>F</u> R GTGCC7 R ATGGG 4 G GGCC2 <u>F</u> TTGT <u>F</u> V AP-1 GGTC1 <u>G</u> GTC1 <u>G</u> L	ATG GCC P AGG G AGG G AGT V CAC. T TCC P	G CTA Y TTC S GGG G G TCT L AAAA X TAT. I TAA	CTG C AP- CTT F GGA E TGT V AAAT M I AAAT	CCC. P -5 TGT' V AGGG G GAT' GAT'	ATC: S TAAG K CATO M CTG? C GGAA S GGAA? N AGCCO	TAC T GCC P G G G G G G ATG C IGA D GAA	GGA E TAT I TCCO P CCA H CCAT M CTA Y CAT	GGT V TCG R CAT I CAA GGG G G CAT I TGT.	AGC. A TCC: P TGA E CAA K CAA K AGC: A AGC.	AAAA K GAA K GGCC A AAAA K TTGG C TGGG G AAAA	AP-1 GAG R GAT I TAC T ACT L D ACA Q ATG	3 GGC A TGA E TGC A AGC A C C GGA E GTT	TTT L GTT. L AAAA K TGGG G J CAT I AGC A TGGG	GAAJ K AAGJ R GGCC A CAGC <u>R</u> AP-8 TATJ I G GGCCJ	AGC A AAAA K GCT. L GTT AAAC T AAAA N AAAA	TGG G AGA E S GCG' R AAAA K TGT V	ACT L ATT L TAA S GGC A GGC A AGA	CCA Q GGG G G G G G G G G G G G ATA Y TGA	GCC P CAT M TGA D CCCC P TGT V GTT	TTC S GGA D GTA P-11 TTG W TGG. G CGT V GCT	CAA K TGA E I-1 GAA K AAC T CGA E GAT	ACT L N TGA GGT V GAT I AAAA N CAT M	AAA K TCT L AAAA N TCC P TGC A CGG G CGG	GGT V TCC P TCA H TGA TGA C ATGA C ACA	TTT F TGC A TGG G AGC A G AGC G A G A A G C A A G C A A G C A A G C A A C A A A C A A A C A A A A	TGG G CGT V AGA E GGT V CAT M GAA K TGC	L CCT L GTT L GGCC P GGAA K GGAT I GTT F CCTT	1 2 1 3 1 3 1 3 1 4 4 1 4 1 4 4 1 4 4
GTT. V CCTT. L I GGG G AGA R TCA S TCA	ACAAC F <u>R</u> STGCC 7 R ATGGG 4 G GGCC 4 G GGCC 4 G GGCC 7 V AP-1 GTCT G L	ATG GCC P AGG G AGG G AGT V CAC T T CCAC P	G CTA Y TTCC S GGGG G TCT L AAAA K I TAA K	CTG AP- CTT F GGA E TGT V AAAT I AAAT I AGA E	CCC. P -5 TGT' V GGAT(L GGAT) I	ATC: S TAAC K CATC M CTG' CC CTG' C GGAA: N AGCC A	TAC T 3GCC P 3GGG G TGGG G ATG C IGA D SAA N	GGA TAT I TCCC P CCA H CCA M CCA Y I	GGT V TCG R CAT I CAT GGG G CAT I TGT V	AGC. A TCCC P TGA E CAA K AGCC A TGC' A AGC. A	J AAAA K GGAA K GGGC A AAAA K TTGG G TGGG G AAAA K	AP-1 GAG R GAT I TAC T ACT L D ACA Q W	GGCC A TGA E TGC A AGC A C GGA E GGTT F	TTT L GTT. L AAAA K TGGG G CAT I AGC' A G G	GAAJ K R GGCCA A CAGCA R A A P-8 TATJ I G G G CCCA P	AGC A AAAA K GCT. L GTT L AAAC T AAAA N K	TGG G AGAJ E AAAG' S GCG' R AAAA K TGTG V AGC/ A	ACT L ATT L TAA K SGCC A GGCC A GCCC P AGA	CCA Q GGGG G GGC. A AAAT I GGGG G AATA Y TGAA E	GCCC P CAT M TGA D CCCC P TGT V GTT L	TTC: S GGA D GTA Y P-11 TTG W TGG. G CGT V SGCT L	CAA K TGA D L-1 GAA K CGA T CGA E GAT I	ACT L AAAA N TGA E GGT V GAT I AAAA N CAT M	AAA K TCT L AAAA TCC P TGC A CGGG G CGGG GTC. S	GGT V TCC P TCA H TGA TGA E C ATGA C ACA Q er 2	TTT F TGC A TGG G CCA Q AGC A G G AGC A C A G C A C C A C A C C A C C A C A	TGG G CGI V AGA GGT V CAT M GAA K TGC A	L CCT L GGTT L GGCC P GGAA K GGAT I GGTT F CCTT L	2 1 3 1 3 1 3 1 3 1 4 4 1 4 1 4 1
GTT.	ACAAG FR GTGCCC 7 R ATGGCC 4 G GGCCA 5 O TTTGT 7 V AP-1 GGTC1 3 L AAAATCC (S	ATG C C C C C C C C C C C C C C C C C C C	G CTA Y TTC: S GGGG G TCT L AAAA K TAA. K AACC	CTG AP- CTT F GGA E TGT M AAAT I AAAT I AGA E TGA	CCC. P -5 TGT V AGG G G G G C C C C C C C C C C C C C	ATC' S TAAG K CATC M CTG' GGAA S GGAA N S GGAA N AGCC A TGT/	TAC T 3GCC P 3GGG G TGGG G ATG C IGAA D SAAA N ATT	GGA TAT I TCC P CCA H CAT Y CAT I I TAA	GGT V TCG R CAT I CAA N GGGG G CAT I TGT. V GAT	AGC. A TCCC P TGA E CAA K AGCC A TGC' A TGC' A	J AAAA K GGGC A AAAA K TTGG G AAAA K TCAA H	AP-1 GAG R GAT I TAC T ACT. L TGA D ACA Q TGA W TGA	3 GGC A TGA E TGC A A GC A C GGA E GTT F TCT L	TTT L GTT. L AAAA K TGGG G CCA'	GAAJ K AAGJ R GGCCA A CAGC R AP-8 TAT/J I TGGJ G GCC/J P TGAC	AGC A AAAA K GCT L GTT AAAC T AAAA N AAAA K GCT	TGG G AGAJ E AAAG' S GCG' R AAAA K TGTC A AGC/ A	ACT L ATT L TAA K S SGCC A SGCC A SGCC A D D TAA	CCA Q GGG G GGC. A AAAT I GGGG G AATA Y TGA E ACA	GCC P CAT M ATT A ATT A A CCC P CCC P TGT V GTT L AAAG	TTC: S GGA D GTA Y P-111 TTG G TTGG. G CGT V SGCT L AAS	CAA K TGA D L-1 GAA K AACC T CGA E GAT I	ACT L AAA N TGA GGT I GAT I AAA N CAT M I TST	AAA K TCT L AAAA P TCC P TGC A CGG G CGG G CGG G TC. S ?rim ACC P	GGT V TCC P TCA H TGT V TGA E ATG C ACA Q Q Z ACA	TTT F TGC A TGG CCA Q AGC A CCA Q AGC A CCA A C C C A C C C C A C	TGG G CGI V AGA E GGT V CAT M GAA K TGC A	L CCT L GTT L GCC P GAA K GAA I I GTT F CTT L	$ \begin{array}{c} 2 \\ 1 \\ 3 \\ 1 \\ 3 \\ 1 \\ 4 \\ 1 \\ 4 \\ 1 \\ 4 \\ 1 \\ $
GTT. CCT CTT. L GGG GGG GGG CA S L CCA S L CCA S L CCA S L CCA S L CCA S L CCA CCA CCA CCA CCA CCA CCA CCA CCA C	ACAAC F R STGCC 7 R ATGGC 4 G GGCC 4 G GGCC 4 G TTTGT 5 V AP-1 GGTC 5 L AAATC C S CTTGC C A	AGG G AGG G AGG G AGT V CAC. T TCC P TCC P TAG. R	G CTA Y TTC S GGG G TCT L AAAA K TAA. K TAAA K ACCC P	CTG AP- CTT F GGA E TGT V AAT N AAAT I AGA E TGA D	CCC. P -5 TGT' V AGGG G G A G G TCT' L I TGC' A	ATC: S TAAG K CTG' C C GGAA S GGAA N AGCC A TGT/ V	TAC T GCCC P GGGG G G G ATG C IGA D SAAA N F	GGA TAT I TCC P CCA H CAT M CAT I I TAA K	GGT V TCG R CAT I CAA N GGGG G G TGT. V GAT	AGC. P TGA E CAA K AGC A TGC A TGC V	I AAAA K GGAC A AAAA K TTGG G TGGG G AAAA K TCAA K	AP-1 GAG R GAT I TAC T ACT L D ACA Q ATG W TGA	3 GGC A TGA E TGC A AGC C GGA E GTT F TCT L	TTT L GTT. L AAAA K TGGG G CCAT I AGC G G CCAA	GAAA K R GGGCC A CAGC R AP-8 TAT/ I TGG2 GCC/ P TGAC E	AGC A AAAA K GCT L GTT AAAC N AAAA K GCT L	TGG. G AGAJ E AAAG' S GCG' R AAAA K TGTC A GCG' J TGTC A	ACT L ATT L TAA K S SGCO A SGCO A SGCO A SGCO A TAA L L	CCA Q GGG G GGC. A AAAT I GGGG G ATA Y TGA Y ACA Q	GCCC P CAT M ATT L AAT D CCCC P TGT V GTT L AAG. R	TTC: S GGA D GTA TTG TTG G CGT V CGT L S	CAA K TGA D L-1 GAA K AAAC T CGA E GAT I RTT A	ACT L AAA N TGA E GGT I AAA N CAT N L U PP-6-	AAA K TCT L AAA P TCC P TGC C G G TC S CGG G G TC S Prim ACC P -1	GGT V TCC P TCA E TGT V TGA E ATG C ACA Q Q Q	TTT F TGC A CCA Q AGC A G GAA N 2 ATA Y	TGC G CGT V AGA GGT V CAT M GAA K TGC A	L CCT L GTT L GGAA K GAA I GAA I GTT L CTT L GG G G G G G	22 11 33 11 33 11 33 11 33 11 44 11 44 11 55
GTT. V CCT P CTT. L I G G G G G G G G G G G G G G G G G G	ACAAG FR GTGCC 7 R ATGGC 4 G GGCC 4 G GGCC 7 V AP-1 GGTC 7 L AAATCC (S CTTGC	AGG G AGG G AGG G AGT V CAC. T TCC P TCC P TAG. R	G CTA Y TTC S GGG G TCT L AAAA K TAA. K TAAA K ACCC P	CTG AP- CTT F GGA E TGT V AAT N AAAT I AGA E TGA D	CCC. P -5 TGT' V AGGG G G A G G TCT' L I TGC' A	ATC: S TAAG K CTG' C C GGAA S GGAA N AGCC A TGT/ V	TAC T GCCC P GGGG G G G ATG C IGA D SAAA N F	GGA TAT I TCC P CCA H CAT M CAT I I TAA K	GGT V TCG R CAT I CAA N GGGG G G TGT. V GAT	AGC. P TGA E CAA K AGC A TGC A TGC V	I AAAA K GGAC A AAAA K TTGG G TGGG G AAAA K TCAA K	AP-1 GAG R GAT I TAC T ACT L D ACA Q ATG W TGA	3 GGC A TGA E TGC A AGC C GGA E GTT F TCT L	TTT L GTT. L AAAA K TGGG G CCAT I AGC G G CCAA	GAAA K R GGGCC A CAGC R AP-8 TAT/ I TGG2 GCC/ P TGAC E	AGC A AAAA K GCT L GTT AAAC N AAAA K GCT L	TGG. G AGAJ E AAAG' S GCG' R AAAA K TGTC A GCG' J TGTC A	ACT L ATT L TAA K S SGCO A SGCO A SGCO A SGCO A TAA L L	CCA Q GGG G GGC. A AAAT I GGGG G ATA Y TGA Y ACA Q	GCCC P CAT M TGA D CCCC P TGT V GTT L AAG. R	TTC: S GGA D GTA TTG TTG G CGT V CGT L S	CAA K TGA D L-1 GAA K AAAC T CGA E GAT I RTT A	ACT L AAA N TGA E GGT I AAA N CAT N L U PP-6-	AAA K TCT L AAA P TCC P TGC C G G TC S CGG G G TC S Prim ACC P -1	GGT V TCC P TCA E TGT V TGA E ATG C ACA Q Q Q	TTT F TGC A CCA Q AGC A G GAA N 2 ATA Y	TGC G CGT V AGA GGT V CAT M GAA K TGC A	L CCT L GTT L GGAA K GAA I GAA I GAT I CTT L CTT L GG	2 1 3 1 3 1 3 1 4 4 1 4 1 4 1 4 1 4

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 λ gt11 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 ³²P-labeled degenerate oligonucleotide probes from a 10-aa sequence of the highlighted peptide (CYCPSTE-VAK) in Fig. 1 was used to screen the cucumber λ gt11 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.

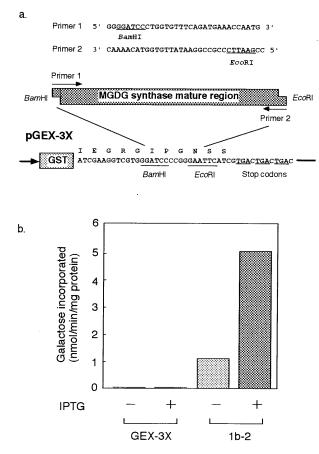


FIG. 2. MGDG synthase expressed in *E. coli*. (*a*) Construction of plasmid with GST fusion expression vector of pGEX-3X to generate the mature region of MGDG synthase as a fusion protein with GST. (*b*) MGDG synthase activity measured by the formation of [³H]MGDG by cell-free extract of *E. coli* XL1-Blue 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.

 λ phage 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; ACCTCCGTAGAT-GGGCAGTAGCATC). 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-

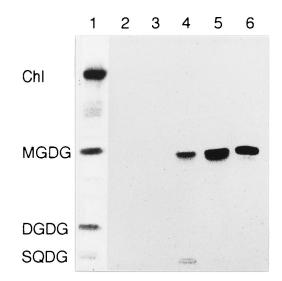


FIG. 3. TLC analysis of galactosylation products produced by the expressed fusion protein of MGDG synthase and GST in *E. coli* extracts. Lane 1, total lipids extracted from spinach, which were stained with anthrone–sulfuric acid; lanes 2 and 3, the reaction products of the extract from GEX-3X, which 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, sulfo-quinovosyldiacylglycerol.

GATGAAACCAATG-3') and primer 2 (5'-CAAAACATG-GTGTTATAAGGCCGCCCTTAAGCC-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-thiogalactopyr-anoside (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 μ M UDP-galactose as substrate. The solution of enzyme (30 μ l) was mixed with 50 μ l of 1,2-dioleoyl-snglycerol [200 μ g; 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 μ l of 8 mM UDP-4,5-[³H]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 $500 \times g$ 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 <i>E</i> .	coli

Transformant	Lipid, mol%							
of E. coli	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 chloro-form:methanol:water (65:15:2, vol/vol/vol). UDP-[³H]galactose, the substrate used in the usual assay mentioned above, was replaced with UDP-[¹⁴C]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 Scarboxymethylated 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 λ gt 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 \pm 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. 2a. 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. 2b). Additionally, an \approx 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. 2b). 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, $\approx 17 \text{ mol}\%$ of the total lipids of the membrane contained MGDG in 1b-2 cells, while MGDG was not detected in the membrane lipids of

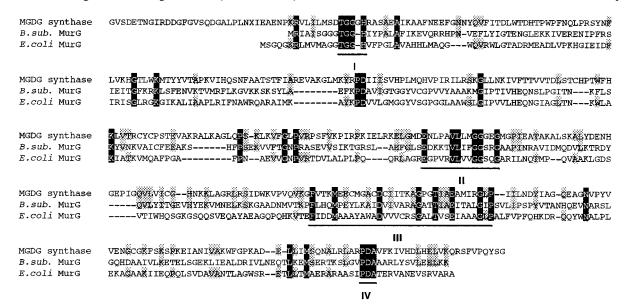


FIG. 4. Alignment of the deduced amino acid sequence of MGDG synthase (mature region) with those of the MurG protein from *B. subtilis* (12) and *E. coli* (19, 20). Amino acids that are identical between MGDG synthase and the other proteins are highlighted in black. Four typical homologous regions (I,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

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-Nacetylglucosamine: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 disaccharidepentapeptide 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.
- Joyard, J. & Douce, R. (1987) in *Biochemistry of Plants*, eds. Stumpf, P. K. (Academic, New York), pp. 215–274.
- 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.
- Maréchal, É., Block, M. A., Joyard, J. & Douce, R. (1991) C. R. Acad. Sci. 313, 521–528.
- Covès, J., Joyard, J. & Douce, R. (1988) Proc. Natl. Acad. Sci. USA 85, 4966–4970.
- 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.
- Mengin-Lecreulx, D., Laurent, T., Rousseau, M. & van Heijenoort, J. (1991) J. Bacteriol. 173, 4625–4636.
- 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.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) J. Mol. Biol. 215, 403–410.
- 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.
- Mengin-Lecreulx, D., Texier, L. & van Heijenoort, J. (1990) Nucleic Acids Res. 18, 2810.
- 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.