Microbiology (2001), **147**, 1937–1946 Printed in Great Britain

Analysis of the *wee* gene cluster responsible for the biosynthesis of the polymeric bioemulsifier from the oil-degrading strain *Acinetobacter lwoffii* RAG-1

David Nakar and David L. Gutnick

Author for correspondence: David L. Gutnick. Tel: +972 3 6409834. Fax: +972 3 6409407. e-mail: davidg@post.tau.ac.il.

Department of Molecular Microbiology and Biotechnology, Tel-Aviv University, Ramat Aviv 69978, Israel

A cluster (27 kbp) of genes responsible for the biosynthesis of the amphipathic, polysaccharide bioemulsifier emulsan from the oil-degrading *Acinetobacter lwoffii* RAG-1 was isolated and characterized. The complete sequence of this cluster, termed *wee*, consisted of 20 ORFs. One set of 17 ORFs was transcribed in one direction, while a second set of three ORFs, 607 bp upstream of the first, was transcribed in the opposite direction. Mutations in either of the two regions caused defects in emulsan production, yielding specific activities of 5–14% of parental emulsifying activity. Putative functions could be assigned to proteins involved in production of nucleotide amino sugar precursors, transglycosylation, transacetylation, polymerization and transport. However, no JUMPstart or *ops* sequences, normally found associated with some polysaccharide biosynthetic gene clusters, were identified. Evidence is presented suggesting that the bioemulsifier may be a member of the group 1 or group 4 polysaccharides.

Keywords: emulsan, biosurfactant, amphipathic biopolymer, polysaccharide biosynthesis

INTRODUCTION

The hydrocarbon-degrading *Acinetobacter lwoffii* RAG-1 produces a potent, amphipathic, polyanionic (molecular mass 10^6 Da), galactosamine-containing bioemulsifier, emulsan (Rosenberg *et al.*, 1979; Zuckerberg *et al.*, 1979), which stabilizes a wide variety of oil/water emulsions. As a powerful emulsion stabilizer (Desai & Banat, 1997) emulsan has become a commercial product and has found a number of industrial applications (Gutnick, 1987; Gutnick *et al.*, 1991; Shabtai *et al.*, 1985). The bioemulsifier consists of a polysaccharide backbone made up of *N*-acyl D-galactosamine, *N*-acyl L-galactosamine uronic acid and 2,4-diamino-6-deoxy-D-glucosamine (Belsky *et al.*, 1979; Zuckerberg *et al.*, 1979; Gutnick, 1987) (Fig. 1). The amphipathic properties of emulsan are due in part to the presence of fatty acids linked to the polysaccharide backbone in both ester and amide linkages (Belsky *et al.*, 1979). The biopolymer accumulates on the cell surface of exponential-phase RAG-1 cells as a minicapsule, and is released into the medium as a protein–polysaccharide complex as the cells approach stationary phase (Goldman *et al.*, 1982; Pines *et al.*, 1983). This release

Abbreviations: Cm, chloramphenicol; Km, kanamycin; TLU, translucent; TMR, transmembrane helical region.

The GenBank/EMBL accession number for the sequence analysis of the eight fragments determined in this work is AJ243431.

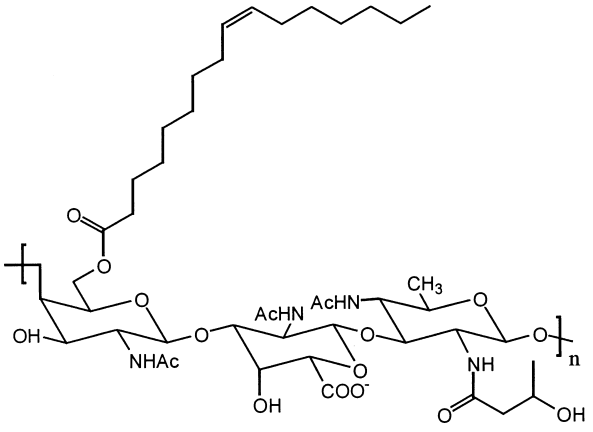

Fig. 1. Structure of the emulsan-like polymer.

from the cell surface involves the participation of an exocellular esterase which has been cloned, sequenced and overexpressed in *Escherichia coli* (Alon & Gutnick, 1993; Reddy *et al*., 1989; Shabtai & Gutnick, 1985). Removal of the protein yields a polymer termed apoemulsan, which retains much of the emulsifying activity towards certain hydrocarbon substrates but is inactive in the emulsification of relatively non-polar, hydrophobic, aliphatic materials (Zosim *et al*., 1986; Shabtai & Gutnick, 1985). While much information has been obtained on the chemistry and physiology of emulsan production as well as on its application, little is known about the genes involved in emulsan biosynthesis. In this report we present results of analysis of a 27 kb gene cluster, termed *wee*, encoding the genes required for biosynthesis of the emulsan biopolymer.

METHODS

Bacterial strains and growth conditions. *A. lwoffii* RAG-1 was grown on ethanol-minimal salts medium (Shabtai & Gutnick, 1985) supplemented with Hutner's Metals 44 (Cohen-Bazire *et al*., 1957) at 30 °C. *E. coli* was grown on Luria–Bertani broth (LB) at 37 °C. For growth on plates, liquid media were solidified with 1·5 % Difco agar. Where appropriate, media were supplemented with antibiotics to the following final concentrations: chloramphenicol (Cm), 25 μg ml^{-1}; kanamycin (Km), 75 μg ml^{-1}; and ampicillin, 100 or 150 μg ml^{-1}.

Isolation of translucent (TLU) mutants. Emulsan-defective *A. lwoffii* RAG-1 mutants were obtained following mini-Tn*10*Km transposon mutagenesis (Leahy *et al*., 1993). TLU mutants were selected by plating the cell suspension onto ethanol-minimal salts medium containing Km and Cm (to counterselect the donor). TLU mutants are visually less opaque then the wild-type and have previously been shown to be emulsan-defective (Bayer *et al*., 1983; Pines & Gutnick, 1981).

Emulsifying activity assay. Emulsan was assayed functionally on the basis of its ability to form an oil/water emulsion from a mixture of hexadecane and 2-methylnaphthalene in 7·5 ml 20 mM Tris buffer (pH 7·0) containing 10 mM $MgSO_4$ as previously described (Rosenberg *et al*., 1979). One unit of emulsan is that amount which gives rise to a turbidity of 100 Klett units in the standard assay. Typically, purified emulsan exhibits a specific activity of 150–180 U (mg biopolymer)$^{-1}$.

Analysis of TLU mutants. TLU mutants were analysed by Southern hybridization (Southern, 1975). The gene encoding Km resistance was excised from the plasmid pLOFKm with *Not*I, labelled using the DIG system (DIG-High prime DNA labelling kit; Roche Molecular Biochemicals) and used as a probe in Southern hybridization assays. Genomic DNA was isolated as described by Hopwood *et al*. (1985). *Eco*RI-digested genomic DNA was used as template for Southern hybridization. To obtain the flanking sequence of the insertion, an *Eco*RI mini-library of the approximate size of the fragment that reacted with the Km probe was inserted into pUC18. The clones containing the Km resistance gene were selected on plates and the isolated plasmids sequenced using an ABI 377 DNA sequencing apparatus (Perkin-Elmer). The orientation of the fragments with respect to each other was predicted according to the sequence analysis and confirmed by PCR (Eppendorf Master Cycler model 5330 plus) using appropriate primers. Sequence databases were searched with the National Center for Biotechnology Information BLAST network server (Altschul *et al*., 1990). The TMHMM server was used to predict transmembrane helices of the proteins (Sonnhammer *et al*., 1998).

RESULTS AND DISCUSSION

Identification of the gene cluster responsible for emulsan biosynthesis

We have previously reported that emulsan-defective mutants can readily be isolated on the basis of their TLU colonial morphology (Bayer *et al*., 1983) as well as their resistance to a specific bacteriophage ap3 (Pines & Gutnick, 1981), which uses either cell-associated emulsan or emulsan localized at an oil/water interface as a phage receptor. After mini-Tn*10*Km transposon mutagenesis, 126 *A. lwoffii* RAG-1 TLU mutants were isolated on ethanol-minimal salts medium containing Km and Cm. These mutants were resistant to phage ap3, but retained their sensitivity to a second *A. lwoffii* RAG-1 specific phage, ap2 (Pines & Gutnick, 1981). We have named these TLU Km-resistant mutant strains, *A. lwoffii*

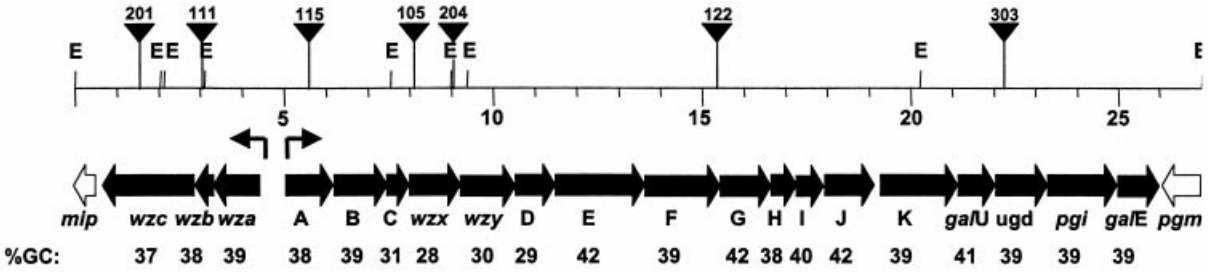

Fig. 2. Genetic organization of the *wee* cluster responsible for emulsan production. The scale of the cluster size is in kilobases. The filled triangles indicate the Tn*10* transposon inserts. The numbers above these triangles indicate the corresponding *A. lwoffii* LN mutant. The filled arrows represent completely sequenced ORFs; open arrows represent partially sequenced ORFs. Sites marked with an E indicate the *Eco*RI restriction sites. Putative rightward and leftward promoter sites are indicated by the thin black arrows. The names of the genes are shown below the corresponding ORFs. ORFs labelled solely with capital letters refer to the putative pathway-specific proteins, Wee A–K. The G+C content (mol %) of each of the ORFs is presented below the name. Position numbers of the *A. lwoffii* LN mutants were (in bp): 1520 (LN201), 3071 (LN111), 5601 (LN115), 8171 (LN105), 9025 (LN204), 15331 (LN122) and 22225 (LN303).

LN strains. Forty-six emulsan-defective mutants were analysed by Southern analysis using the Km resistance marker as probe (data not shown). Seven unique *Eco*RI fragments of various sizes carrying the Km resistance marker were obtained and analysed by sequence analysis. The corresponding mutants were *A. lwoffii* LN105, LN111, LN115, LN122, LN201, LN204 and LN303 (Fig. 2). The orientation of these fragments with respect to each other was predicted from the sequence and then confirmed by PCR using primers from adjacent fragments. These PCR products were sequenced, thereby confirming the orientation.

As shown in Table 1, the seven transposon mutants exhibited decreased emulsan activities, between 5 and 14% of the parental values. Of interest was the finding that insertions in the two clusters apparently transcribed in opposite directions led to similar decreases in emulsan activity (compare, for example, the emulsifying activity in mutants LN111 and LN204 (Fig. 2). The results strongly suggest that genes associated with both clusters are essential for producing high levels of active emulsan. It should be noted, however, that the decrease in emulsifying activity associated with the mutant phenotypes may not necessarily be due to disruption of the gene interrupted by the transposon (Fig. 2), but might also result from polar effects reducing expression of downstream genes.

Sequence analysis of these eight fragments revealed a 26953 bp region (Fig. 2), that was deposited in the GenBank/EMBL database with the accession number AJ243431. Twenty-two putative ORFs were predicted using the Clone Manager program (Scientific & Educational Software, Durham). Of these, 20 sequences were complete while partial (approx. 65–75%) sequences for two additional ORFs were also obtained (Fig. 2). Two clusters, transcribed in opposite directions were found to be separated by a non-translated sequence of 607 bp. Computer analysis revealed putative σ^{70} promoter sequences within this non-translated region in both upstream and downstream directions. Interestingly, no sequence was found which corresponded to the 39 bp JUMPstart element generally located upstream of polysaccharide biosynthetic gene clusters (Hobbs & Reeves, 1994) nor for the 8 bp *ops* elements normally present within the JUMPstart sequences which are thought to function in antitermination of transcription (Bailey *et al.*, 1997; Nieto *et al.*, 1996). One region, transcribed in the upstream direction, appears to encode 3 complete ORFs, while the second region is transcribed in the opposite direction and encodes 17 complete ORFs (Fig. 2).

Table 1. Emulsan activities of *A. lwoffii* insertion mutants

Cells were grown on ethanol-minimal salts medium as described in Methods until maximum emulsifying activity was obtained in the cell-free culture broth (typically, 72 h). Emulsan activity was assayed as described in Methods. The sites of insertions associated with each mutation are shown in Fig. 2.

Strain	Emulsan activity (U ml^{-1})
RAG-1	220
LN105	22
LN111	11
LN115	24
LN122	23
LN201	18
LN204	13
LN303	31

Predicting the biosynthetic route of emulsan based on homologies to the genes involved in emulsan biosynthesis

Using computer-assisted BLAST searches, putative homologues were identified which suggested an involvement in (a) production of nucleotide sugar precursors, (b) transglycosylation reactions for synthesis of the membrane-bound trisaccharide repeat unit, (c) transacetylation, (d) repeat unit polymerization and polysaccharide transport. We employ the BPGN (bacterial polysaccharide gene nomenclature) scheme (Reeves *et al.*, 1996) for naming polysaccharide biosynthetic genes according to the three letter *w*** system and refer to the emulsan biosynthetic cluster(s) as the *wee* (*e*xopolysaccharide; *e*mulsan) region. In accordance with this system, genes which we consider are likely to be pathway-specific have been labelled *weeA–weeK*, respectively. Other genes (*wza*, *wzb*, *wzc*, *wzx*, *wzy*) in the emulsan *wee* cluster were found to share sequence homology with a variety of more general saccharide-processing genes encoding functions common to a variety of capsular biosynthetic clusters, and were therefore assigned the same commonly used name (Reeves *et al.*, 1996). In addition, sequence analysis revealed five ORFs exhibiting similarities to sugar pathway gene products, Pgi, Pgm (partially sequenced), GalU, GalE and Ugd. The putative proteins encoded by the *wee* cluster along with the relevant homologues and their sequence identity and similarity values are shown in Table 2. Based on these homologies and their functions we propose a hypothetical biosynthetic pathway (Fig. 3). The subsequent verification of these proposed reactions by suitable *in vitro* analyses is currently in progress. The basis for this proposal is presented below. Reference is made both to Fig. 2 and to Table 2, which can be used for clarification.

Synthesis of amino sugar precursors

Since all three nucleotide sugar precursors are amino sugars (Fig. 1) we assume that the initial step in the biosynthesis of these precursors would be the formation of UDP-*N*-acetyl-D-glucosamine from L-fructose-6-phosphate analogous to the case in *E. coli* (Fig. 3). The biosynthesis of this intermediate involves three genes, *glmS*, *glmM* and *glmU*, which are not located in a single

Table 2. Putative proteins encoded by the *wee* cluster and similarity to database sequences

Protein (amino acids)	Molecular mass (kDa)	Putative function	Homologue	Organism	% Identity (similarity)	Amino acids	Accession no.
Mip (178)*	–	Macrophage infectivity potentiator	FkpA	*Aeromonas hydrophila*	47 (62)	268	AAC45362
			Mip	*Coxiella burnetii*	36 (52)	230	P51752
			Mip	*Legionella rubrilucens*	37 (57)	248	AAC45704
Wzc (726)	81·6	Protein tyrosine kinase	Ptk	*Acinetobacter johnsonnii*	73 (87)	733	CAA75431
			EpsB	*Pseudomonas solanacearum*	34 (56)	750	Q45409
			Wzc	*Escherichia coli* (O9a:K30)	37 (57)	721	AAD21564
			Etk	*Escherichia coli* (O127:H6)	34 (54)	726	CAB43868
			Orf6	*Klebsiella pneumoniae* K2	33 (57)	722	Q48452
			AmsA	*Erwinia amylovora*	34 (57)	726	Q46631
Wzb (142)	16·4	Low-molecular-mass protein tyrosine phosphatase	Ptp	*Acinetobacter johnsonnii*	79 (86)	142	CAA75430
			EpsP	*Pseudomonas solanacearum*	42 (59)	145	Q45408
			Wzb	*Escherichia coli* (O9a:K30)	37 (56)	148	AAD21563
			Orf5	*Klebsiella pneumoniae* K2	37 (55)	144	Q48451
			AmsI	*Erwinia amylovora*	35 (56)	144	Q46630
Wza (366)	38·7†	Outer-membrane lipoprotein	EpsA	*Pseudomonas solanacearum*	32 (52)	377	Q45407
			AmsH	*Erwinia amylovora*	28 (49)	377	Q46629
			Orf4	*Klebsiella pneumoniae* K2	29 (51)	378	Q48450
			Wza	*Escherichia coli* (O9a:K30)	28 (50)	379	AAD21562
WeeA (379)	41·9	UDP-*N*-acetylglucosamine 2-epimerase	WecB	*Escherichia coli*	64 (77)	376	P27828
			EpsC	*Pseudomonas solanacearum*	59 (76)	396	P52641
			Orf1	*Neisseria meningitidis*	56 (74)	372	AAC38285
WeeB (417)	45·8	UDP-*N*-acetyl-D-galactosamine uronic acid dehydrogenase	WecC	*Escherichia coli*	64 (77)	420	P27829
			EpsD	*Pseudomonas solanacearum*	63 (77)	423	Q45410
			RffD	*Erwinia carotovora*	64 (78)	287	AAC12870
WeeC (184)	20·6	*O*/*N*-Acetylgalactosamine acetyltransferase	WbbJ	*Escherichia coli* K-12	28 (44)	169	P37750
			EpsH	*Streptococcus thermophilus*	31 (47)	159	AAC44015
			WcfD	*Bacteroides fragilis*	30 (49)	218	AAD40715
Wzx (401)	46·8	Flippase (*wzx*)	WbpF	*Pseudomonas aeruginosa*	20 (42)	411	AAC45858
			CapF	*Staphylococcus aureus*	21 (38)	396	P39855
Wzy (436)	50·0	Polymerase (*wzy*)	CapE	*Staphylococcus aureus*	23 (42)	440	P39854
			Rfc	*Salmonella typhimurium*	20 (40)	407	P26479
WeeD (320)	36·7	Mannosyltransferase C	ManC	*Aquifex aeolicus*	24 (37)	368	AAC06753

Table 2 (*cont.*)

Protein (amino acids)	Molecular mass (kDa)	Putative function	Homologue	Organism	% Identity (similarity)	Amino acids	Accession no.
WeeE (712)	77·3	Unknown	Xdh‡	*Galactocandida mastotermitis*	26 (42)	353	AAC24597
			YdjJ‡	*Escherichia coli*	24 (42)	347	P77280
			Sdh‡	*Saccharomyces cerevisiae*	29 (42)	346	P35497
			IdhA§	*Sinorhizobium meliloti*	30 (45)	330	AAC70005
			OphB§	*Burkholderia cepacia*	24 (44)	391	AAD03557
			Gfo§	*Zymomonas mobilis*	23 (42)	381	2098323
WeeF (604)	69·4	No identities found					
WeeG (405)	44·5	Glycosyltransferase	plE	*Bordetella pertussis*	28 (46)	403	S70676
			bpJ	*Pseudomonas aeruginosa*	24 (41)	413	AAC45864
WeeH (203)	23·3	UDP-galactose phosphate transferase	YvfC	*Bacillus subtilis*	53 (72)	202	CAA96480
			OrfC	*Campylobacter jejuni*	54 (69)	200	AAD09295
			WbaP‖	*Escherichia coli* (O9a:K30)	40 (62)	476	AAD21565
			WbaP‖	*Salmonella enterica*	38 (58)	476	P26406
WeeI (228)	23·1	Acetyltransferase	WbdR	*Escherichia coli* O157	50 (71)	221	AAC32350
			WlaI	*Campylobacter jejuni* 81116	25 (47)	203	CAA72358
			CapG	*Staphylococcus aureus*	31 (45)	172	P39856
WeeJ (366)	43·7	Perosamine synthetase	Per	*Escherichia coli* O157	41 (59)	366	AAC32343
			WbeE	*Vibrio cholerae*	36 (54)	367	1586120
			PerA	*Brucella melitensis*	34 (52)	367	AAC98613
WeeK (624)	69·7	dTDP-glucose 4,6-dehydratase/UDP-glucose 4-epimerase	LpsB	*Rhizobium etli*	46 (63)	683	AAB66669
			Orf10	*Vibrio cholerae* O139	48 (67)	646	AAC46251
			WbcP	*Yersinia enterocolitica*	46 (65)	638	S51266
			WbpM	*Pseudomonas aeruginosa*	43 (62)	665	AAD45269
GalU (291)	31·9	UTP-glucose-1-phosphate uridylyltransferase	GalU	*Haemophilus influenzae*	57 (74)	295	P44878
			GalU	*Escherichia coli*	56 (73)	302	P25520
			HasC	*Streptococcus pyogenes*	42 (64)	304	Q54713
Ugd (416)	47·4	UDP-glucose dehydrogenase	RkpK	*Sinorhizobium meliloti*	28 (51)	437	CAA10918
			TuaD	*Bacillus subtilis*	26 (48)	461	AAB94865
			AceM	*Acetobacter xylinus*	29 (49)	449	CAA72087
Pgi (557)	62·7	Glucose-6-phosphate isomerase	Pgi	*Acinetobacter calcoaceticus*	74 (84)	557	Q59088
			Pgi	*Escherichia coli*	45 (63)	549	P11537
			Pgi	*Haemophilus influenzae*	43 (62)	563	P44312
GalE (338)	37·3	UDP-glucose 4-epimerase	GalE	*Escherichia coli*	53 (70)	338	P09147
			GalE	*Brucella abortus*	56 (70)	335	AAC46054
			GalE	*Neisseria meningitidis*	51 (71)	339	AAA63156
Pgm (305)*	–	Phosphoglucomutase	XanA	*Xanthomonas campestris*	60 (75)	448	P29955
			ManB	*Vibrio cholerae*	61 (76)	463	Q06951
			AlgC	*Pseudomonas aeruginosa*	33 (50)	463	P26276

* Only partial sequence available.

† Size of the mature protein.

‡ N-terminal half of ORF.

§ C-terminal half of ORF.

‖ ORF homologous to C-terminal portion of protein.

region of the chromosome (Trefzer *et al*., 1999). This is consistent with the fact that homologues to these genes were not found within the *wee* cluster.

Homologues of Pgm have previously been shown to be bifunctional enzymes catalysing both phosphoglucomutase and phosphomannomutase activity (Ye *et al*., 1994). GalU homologues show UTP-glucose-1-phosphate uridylyltransferase activity (Crater *et al*., 1995; Weissborn *et al*., 1994), whilst GalE homologues exhibit UDP-glucose 4-epimerase activity (Lee *et al*., 1995; Scupham & Triplett, 1997). To obtain one of the nucleotide sugar precursors of emulsan, UDP-*N*-acetyl-D-galactosamine, from UDP-*N*-acetyl-D-glucosamine, the 4-epimerase activity is needed. We propose that GalE is responsible for this conversion (Fig. 3).

RkpK of *Sinorhizobium meliloti*, a Ugd homologue, exhibits UDP-glucose dehydrogenase activity (Kereszt *et al*., 1998). It should be noted that the similarity of Ugd to its homologues is significantly lower than the similarity of the other sugar pathway genes to their homologues. It is possible, therefore, that in *A. lwoffii* RAG-1 the Ugd homologue may use a different substrate. No putative roles were assigned for Pgi, Pgm GalU and Ugd in the biosynthetic pathway of emulsan. The reason *A. lwoffii* LN303 is emulsan-negative could be explained by a polar effect on *galE* downstream of *ugd*.

WeeA and WeeB are closely related to WecB and WecC of the enterobacterial common antigen (ECA) cluster of *E. coli* (Table 2). WecB is an UDP-*N*-acetylglucosamine 2-epimerase (Meier-Dieter *et al*., 1990; Sala *et al*., 1996) while WecC was found to exhibit UDP-*N*-acetyl-mannosaminuronic acid dehydrogenase activity (Meier-Dieter *et al*., 1990). On the basis of the backbone structure of emulsan and in analogy with known enzymic interconversions of nucleoside-diphosphate-linked sugars, we propose that WeeA and WeeB are involved in the biosynthesis of UDP-*N*-acetyl-L-galactosaminuronic acid (Singh *et al*., 1990). Accordingly, WeeA would convert UDP-*N*-acetyl-D-glucosamine into UDP-*N*-acetylmannosamine. Subsequently, WeeB would oxidize the UDP-*N*-acetylmannosamine into UDP-*N*-acetylmannosaminuronic acid (Fig. 3). The third enzyme, which would be needed to obtain UDP-*N*-acetyl-L-galactosaminuronic acid, is a 3,5-epimerase and could be WeeE or WeeF, for which no putative functions have been assigned.

WeeJ is similar to several putative perosamine synthetases (Table 2), which show high similarity to a large number of pyridoxal-binding proteins (Stroeher *et al*., 1995). The WeeK protein (Table 2) is similar to several dTDP-glucose 4,6-dehydratases and UDP-glucose 4-epimerases (Comstock *et al*., 1996; Dean *et al*., 1999; Skurnik *et al*., 1995). Analysis of the protein sequence of WeeK predicted four transmembrane helical regions (TMRs) within the N-terminal portion of the protein. If WeeK is similar to a dTDP-glucose 4,6-dehydratase, it may be responsible for the conversion of UDP-D-glucosamine into UDP-4-keto-6-deoxy-D-glucosamine. WeeJ could subsequently catalyse the formation of diamino 2,4-diamino-6-deoxy-D-glucosamine, a component of the repeat unit, from UDP-4-keto-6-deoxy-D-glucosamine (Fig. 3).

Synthesis and acylation of the repeat unit

Three ORFs, WeeD, WeeG and WeeH, show significant similarity to bacterial glycosyltransferases (Table 2). The WeeH protein is similar to the C-terminal portion of WbaP, which is the galactosyl transferase responsible for transferring the first galactose 1-phosphate from GDP-galactose to the undecaprenyl phosphate (UndPP) during the biosynthesis of the O-antigen of *Salmonella enterica* (Wang *et al*., 1996). In this regard, analysis of the protein sequence of WeeH predicts one TMR, which is similar to the C-terminal portion of WbaP (Wang *et al*., 1996). These three gene products, *weeH*, *D* and *G*, might be involved in the transfer of the three activated nucleotide sugars of emulsan to UndPP (Fig. 3).

WeeC and WeeI show significant similarity to acetyltransferases from other bacteria. These proteins all contain a sequence of 50 amino acids exhibiting similarity to a conserved region of the NodL-LacA-CysE acetyltransferase family (Lin *et al*., 1994). These proteins might also be involved in transacylation of the polysaccharide backbone with other longer-chain-length fatty acids, thereby conferring amphipathic characteristics on the water-soluble emulsan polysaccharide (Fig. 3). Thus far we have found no homologues which might be involved in the transamidation reactions which might generate amide linkages with longer-chain-length fatty acids on the amino sugars.

Polymerization and polymer transport

The Wzx protein exhibits similarity to polysaccharide exporter proteins (PST), which are highly divergent (Table 2), but share a similar predicted topology containing 12 TMRs (Paulsen *et al*., 1997). Wzx is considered to catalyse the translocation of the membrane-bound repeat unit so that rather than facing the cytoplasm, the repeat unit now faces the periplasm (Liu *et al*., 1996). A second integral membrane protein with 12 predicted TMRs, Wzy, shows similarity to several proteins which have been implicated in polymerization of the repeat unit on the periplasmic side of the cytoplasmic membrane (Whitfield, 1995). Three ORFs (*wza*, *wzb* and *wzc*) comprise that region of the *wee* cluster, which is transcribed in the opposite direction. This conserved region of genes is prevalent in clusters for expression of *E. coli* group I K antigens and in capsular clusters of several other bacteria (Rahn *et al*., 1999). In *E. coli* strains carrying mutations in *wza* and *wzc*, capsular polysaccharide is still partially polymerized but not assembled on the cell surface, suggesting that these proteins are involved in translocation of the polysaccharide out of the periplasm into the environment (Drummelsmith & Whitfield, 1999). Homologues of Wza (Table 2) are putative outer-membrane lipoproteins with a predicted β-barrel structure and are members of the outer-membrane auxiliary (OMA)

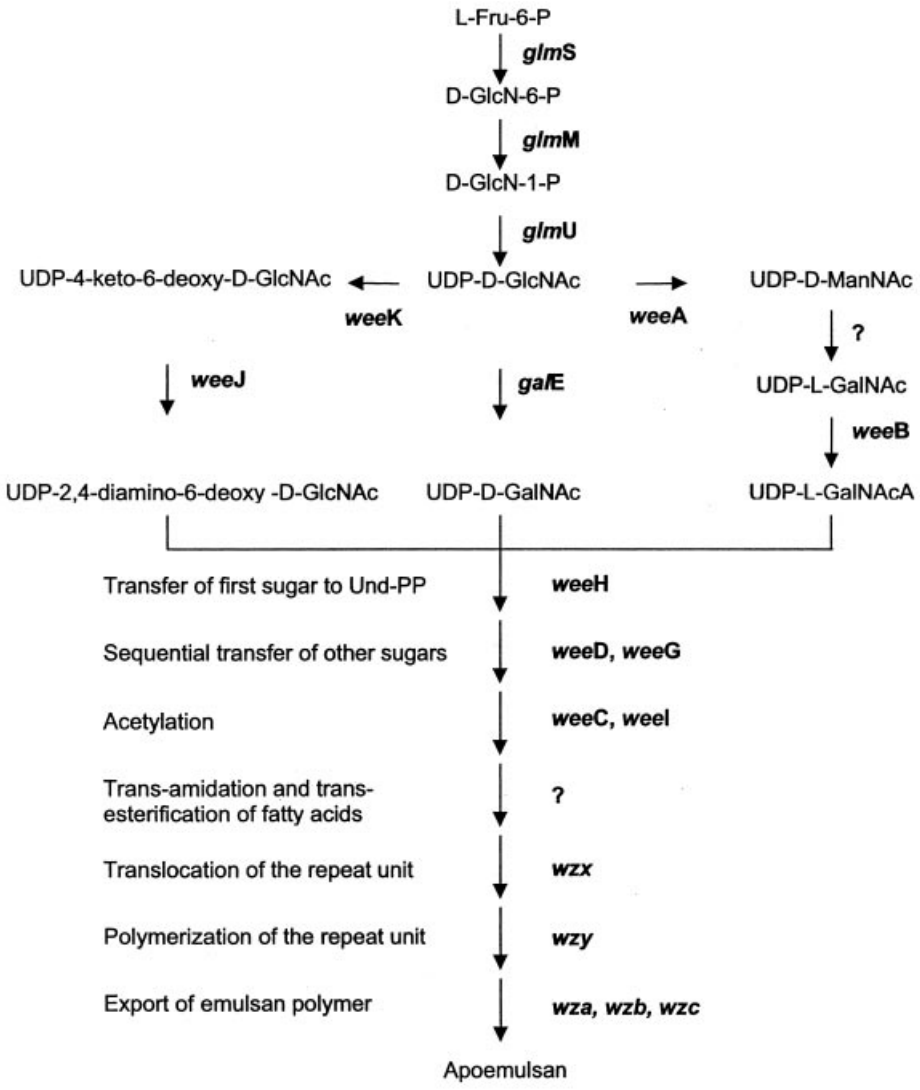

Fig. 3. Proposed biosynthetic pathway of emulsan. Fru, Fructose; GlcN, glucosamine; GlcNAc, *N*-acetylglucosamine; ManNAc, *N*-acetylmannosamine; GalNAc, *N*-acetylgalactosamine; GalNAcA, *N*-acetylgalactosamine uronic acid.

family (Paulsen *et al*., 1997). It has been suggested that Wza is involved in forming a channel for surface polysaccharide secretion (Drummelsmith & Whitfield, 2000). Proteins similar to Wzc of *A. lwoffii* RAG-1 are grouped in the cytoplasmic membrane periplasmic auxiliary (MPA1) family (Paulsen *et al*., 1997). Sequence analysis of Wzc of *A. lwoffii* RAG-1 predicts two TMRs and an ATP-binding motif in the C-terminal tail, which is similar to members of the MPA1 protein family. Additionally, Wzc of *A. lwoffii* RAG-1 shows high similarity to Ptk of *Acinetobacter johnsonnii*, Etk of *E. coli*, and Wzc of *E. coli*, all autophosphorylating protein tyrosine kinases (Grangeasse *et al*., 1997; Ofir *et al*., 1999; Vincent *et al*., 1999). Proteins similar to Wzb are predicted to be acid phosphatases. Dephosphorylation of the phosphorylated protein tyrosine kinase was demonstrated with Ptp of *A. johnsonnii* and Wzb of *E. coli* (Grangeasse *et al*., 1997; Vincent *et al*., 1999). These activities have recently been determined in *A. lwoffii* RAG-1 and the results will be presented separately.

No putative functions have been assigned to ORFs WeeE and WeeF. In the case of WeeF no homologues were identified. ORF WeeE encodes a putative 712 aa protein. The first N-terminal half of the translated sequence shows low similarity to sugar dehydrogenases, while the C-terminal half shows low similarity to three different enzymes of no obvious function for emulsan biosynthesis.

Interestingly, the G + C content of four ORFs, *weeD* (a glycosyltransferase), *wzx*, *wzy* and *weeC*, is lower than that of the remainder of the cluster (Fig. 1). Sequence analysis revealed a preferential usage of the A + T-rich codons for leucine, isoleucine, lysine, asparagine and

phenylalanine. This has previously been reported for several genes involved in polysaccharide biosynthesis, including that of *S. enterica* (Reeves, 1993). This unusually low G+C content supports the hypothesis that the polysaccharide gene clusters were likely to have been assembled from several sources and that these genes have evolved in species with accordingly low G+C content (Reeves, 1993).

Thus far, *A. lwoffii* RAG-1 is the only natural isolate that produces the emulsan biopolymer. It is somewhat surprising, therefore, that almost all of the biosynthetic genes are at least partially homologous to known proteins involved in polysaccharide biosynthesis in other organisms. Capsular gene clusters have been cloned from a number of Gram-negative bacteria although the most studied to date are those of *E. coli*, which may be regarded as a paradigm for Gram-negative bacteria (Roberts, 1996). A new system for classification recognizes four groups of *E. coli* polysaccharide capsules, based on genetic and biosynthetic criteria (Whitfield & Roberts, 1999). The occurrence of homologues in the *wee* cluster to proteins involved in biosynthesis of group 1 K antigens in *E. coli*, coupled with the fact that the emulsan biopolymer is a high molecular mass (apparent molecular mass $\sim 10^6$) surface polysaccharide during exponential growth suggests that emulsan is a member of the group 1 family (Paulsen *et al*., 1997; Whitfield & Roberts, 1999). In contrast, the *wee* cluster apparently contains neither JUMPstart nor *ops* sequences, which are conserved in group 1 clusters (Rahn *et al*., 1999). In addition, group 1 capsules are generally not considered to contain amino sugars (Roberts, 1996). According to this criterion, emulsan might be thought of as a member of the group 4 family of polysaccharides. The two biosynthetic pathways (group 1 and group 4) appear to differ by the presence of a WbaP homologue, which functions in the initial transglycosylation reaction in the biosynthesis of group 1 K antigens (Wang *et al*., 1996).

As a working hypothesis, and in analogy to the pathway for biosynthesis of type I *E. coli* capsules (Whitfield & Roberts, 1999), we propose that emulsan biosynthesis starts with the conversion of L-fructose-6-phosphate into the three nucleotide sugar precursors (Fig. 3). The first amino sugar is then transferred to undecaprenyl phosphate by a WbaP homologue. Subsequently, other glycosyltransferases add the remaining two amino sugars of the repeat unit. The membrane-bound repeat unit is then transferred across the plasma membrane by the Wzx homologue. Polymerization is believed to occur at the periplasmic face of the plasma membrane and is catalysed by the polymerase Wzy. Accordingly, the nascent polymer grows at the reducing terminus, one repeat unit at a time. Wza and Wzc are putatively involved in the translocation of the mature polysaccharide through the outer membrane. Despite the strong similarity of the organization of the *wee* cluster to clusters of other capsular biosynthetic genes, the functional emulsan bioemulsifier is relatively unusual. This may be due to the amphipathicity conferred on the apoemulsan backbone by specific transacetylases (or transacylases) within the cluster. For example, cleavage of about 40% of the ester linkages resulted in a dramatic reduction of emulsan activity and hydrocarbon substrate specificity (Shabtai & Gutnick, 1985). In addition, hydrocarbon substrate specificity and full emulsifying activity have also been shown to depend on non-covalent protein(s) associated with the amphipathic polysaccharide during its release into the culture broth. Nonetheless, removal of the emulsan-associated protein, yielding apoemulsan, retains many of the original properties of emulsan itself.

Experiments to characterize the gene products and their enzymic activities *in vitro* are currently in progress, along with studies on the regulation of pathway expression.

ACKNOWLEDGEMENTS

This work was supported in part by a grant from the USAID CDR Project Grant Number 94-A14-043. We gratefully acknowledge the excellent technical assistance of Rina Avigad. We thank Professor Victor de Lorenzo for providing the pLOFKm plasmid.

REFERENCES

Alon, R. N. & Gutnick, D. L. (1993). Esterase from the oil-degrading *Acinetobacter lwoffii* RAG-1: sequence analysis and overexpression in *Escherichia coli*. *FEMS Microbiol Lett* **112**, 275–280.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

Bailey, M. J., Hughes, C. & Koronakis, V. (1997). RfaH and the *ops* element, components of a novel system controlling bacterial transcription elongation. *Mol Microbiol* **26**, 845–851.

Bayer, E. A., Skutelsky, E., Goldman, S., Rosenberg, E. & Gutnick, D. L. (1983). Immunochemical identification of the major cell surface agglutinogen of *Acinetobacter calcoaceticus* RAG-92. *J Gen Microbiol* **129**, 1109–1119.

Belsky, I., Gutnick, D. L. & Rosenberg, E. (1979). Emulsifier of *Arthrobacter* RAG-1: determination of emulsifier-bound fatty acids. *FEBS Lett* **101**, 175–178.

Cohen-Bazire, G., Sistrom, W. R. & Stanier, R. Y. (1957). Kinetic studies of pigment synthesis by non-sulphur purple bacteria. *J Cell Comp Physiol* **49**, 25–68.

Comstock, L. E., Johnson, J. A., Michalski, J. M., Morris, J. G. J. & Kaper, J. B. (1996). Cloning and sequence of a region encoding a surface polysaccharide of *Vibrio cholerae* O139 and characterization of the insertion site in the chromosome of *Vibrio cholerae* O1. *Mol Microbiol* **19**, 815–826.

Crater, D. L., Dougherty, B. A. & van de Rijn, I. (1995). Molecular characterization of *has*C from an operon required for hyaluronic acid synthesis in group A streptococci. Demonstration of UDP-glucose pyrophosphorylase activity. *J Biol Chem* **270**, 28676–28680.

Dean, C. R., Franklund, C. V., Retief, J. D., Coyne, M. J., Hatano, K., Evans, D. J., Pier, G. B. & Goldberg, J. B. (1999). Characterization of the serogroup O11 O-antigen locus of *Pseudomonas aeruginosa* PA103. *J Bacteriol* **181**, 4275–4284.

Desai, J. D. & Banat, I. M. (1997). Microbial production of surfactants and their commercial potential. *Microbiol Mol Biol Rev* **61**, 47–64.

Drummelsmith, J. & Whitfield, C. (1999). Gene products required for surface expression of the capsular form of the group 1 K antigen in *Escherichia coli* (O9a:K30). *Mol Microbiol* **31**, 1321–1332.

Drummelsmith, J. & Whitfield, C. (2000). Translocation of group 1 capsular polysaccharide to the surface of *Escherichia coli* requires a multimeric complex in the outer membrane. *EMBO J* **19**, 57–66.

Goldman, S., Shabtai, Y., Rubinovitz, C., Rosenberg, E. & Gutnick, D. L. (1982). Emulsan in *Acinetobacter calcoaceticus* RAG-1: distribution of cell-free and cell-associated cross-reacting material. *Appl Environ Microbiol* **44**, 165–170.

Grangeasse, C., Doublet, P., Vaganay, E., Vincent, C., Deleage, G., Duclos, B. & Cozzone, A. J. (1997). Characterization of a bacterial gene encoding an autophosphorylating protein tyrosine kinase. *Gene* **204**, 259–265.

Gutnick, D. L. (1987). The emulsan polymer: perspectives on a microbial capsule as an industrial product. *Biopolymers* **26**, S223–S240.

Gutnick, D. L., Allon, R., Levy, C., Petter, R. & Minas, W. (1991). Applications of *Acinetobacter* as an industrial organism. In *The Biology of Acinetobacter*, pp. 411–441. Edited by K. J. Towner, E. Bergogne-Berezin & C. A. Fewson. New York: Plenum.

Hobbs, M. & Reeves, P. R. (1994). The JUMPstart sequence: a 39 bp element common to several polysaccharide gene clusters. *Mol Microbiol* **12**, 855–856.

Hopwood, D. A., Bibb, M. J., Chater, K. F. & 7 other authors (1985). *Genetic Manipulation of Streptomyces: a Laboratory Manual*. Norwich: John Innes Foundation.

Kereszt, A., Kiss, E., Reuhs, B. L., Carlson, R. W., Kondorosi, A. & Putnoky, P. (1998). Novel *rkp* gene clusters of *Sinorhizobium meliloti* involved in capsular polysaccharide production and invasion of the symbiotic nodule: the *rkp*K gene encodes a UDP-glucose dehydrogenase. *J Bacteriol* **180**, 5426–5431.

Leahy, J. G., Jones-Meehan, J. M., Pullias, E. L. & Colwell, R. R. (1993). Transposon mutagenesis in *Acinetobacter calcoaceticus* RAG-1. *J Bacteriol* **175**, 1838–1840.

Lee, F. K., Stephens, D. S., Gibson, B. W., Engstrom, J. J., Zhou, D. & Apicella, M. A. (1995). Microheterogeneity of *Neisseria* lipooligosaccharide: analysis of a UDP-glucose 4-epimerase mutant of *Neisseria meningitidis* NMB. *Infect Immun* **63**, 2508–2515.

Lin, W. S., Cunneen, T. & Lee, C. Y. (1994). Sequence analysis and molecular characterization of genes required for the biosynthesis of type 1 capsular polysaccharide in *Staphylococcus aureus*. *J Bacteriol* **176**, 7005–7016.

Liu, D., Cole, R. A. & Reeves, P. R. (1996). An O-antigen processing function for Wzx (RfbX): a promising candidate for O-unit flippase. *J Bacteriol* **178**, 2102–2107.

Meier-Dieter, U., Starman, R., Barr, K., Mayer, H. & Rick, P. D. (1990). Biosynthesis of enterobacterial common antigen in *Escherichia coli*. Biochemical characterization of Tn*10* insertion mutants defective in enterobacterial common antigen synthesis. *J Biol Chem* **265**, 13490–13497.

Nieto, J. M., Bailey, M. J., Hughes, C. & Koronakis, V. (1996). Suppression of transcription polarity in the *Escherichia coli* haemolysin operon by a short upstream element shared by polysaccharide and DNA transfer determinants. *Mol Microbiol* **19**, 705–713.

Ofir, I., Bloch, Y., Frankel, G., Ullrich, H., Geider, K. & Rosenshine, I. (1999). Protein tyrosine kinases in bacterial pathogens are associated with virulence and production of exopolysaccharide. *EMBO J* **18**, 3241–3248.

Paulsen, I. T., Beness, A. M. & Saier, M. H. J. (1997). Computer-based analyses of the protein constituents of transport systems catalysing export of complex carbohydrates in bacteria. *Microbiology* **143**, 2685–2699.

Pines, O. & Gutnick, D. (1981). Relationship between phage resistance and emulsan production, interaction of phages with the cell-surface of *Acinetobacter calcoaceticus* RAG-1. *Arch Microbiol* **130**, 129–133.

Pines, O., Bayer, E. A. & Gutnick, D. L. (1983). Localization of emulsan-like polymers associated with the cell surface of *Acinetobacter calcoaceticus*. *J Bacteriol* **154**, 893–905.

Rahn, A., Drummelsmith, J. & Whitfield, C. (1999). Conserved organization in the *cps* gene clusters for expression of *Escherichia coli* group 1 K antigens: relationship to the colanic acid biosynthesis locus and the *cps* genes from *Klebsiella pneumoniae*. *J Bacteriol* **181**, 2307–2313.

Reddy, P. G., Allon, R., Mevarech, M., Mendelovitz, S., Sato, Y. & Gutnick, D. L. (1989). Cloning and expression in *Escherichia coli* of an esterase-coding gene from the oil-degrading bacterium *Acinetobacter calcoaceticus* RAG-1. *Gene* **76**, 145–152.

Reeves, P. (1993). Evolution of *Salmonella* O antigen variation by interspecific gene transfer on a large scale. *Trends Genet* **9**, 17–22.

Reeves, P. R., Hobbs, M., Valvano, M. A. & 8 other authors (1996). Bacterial polysaccharide synthesis and gene nomenclature. *Trends Microbiol* **4**, 495–503.

Roberts, I. S. (1996). The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annu Rev Microbiol* **50**, 285–315.

Rosenberg, E., Zuckerberg, A., Rubinovitz, C. & Gutnick, D. L. (1979). Emulsifier of *Arthrobacter* RAG-1: isolation and emulsifying properties. *Appl Environ Microbiol* **37**, 402–408.

Sala, R. F., Morgan, P. M. & Tanner, M. E. (1996). Enzymatic formation and release of a stable glycal intermediate: the mechanism of the reaction catalyzed by UDP-N-acetylglucosamine 2-epimerase. *J Am Chem Soc* **118**, 3033–3034.

Scupham, A. J. & Triplett, E. W. (1997). Isolation and characterization of the UDP-glucose 4′-epimerase-encoding gene, *gal*E, from *Brucella abortus* 2308. *Gene* **202**, 53–59.

Shabtai, Y. & Gutnick, D. L. (1985). Exocellular esterase and emulsan release from the cell surface of *Acinetobacter calcoaceticus*. *J Bacteriol* **161**, 1176–1181.

Shabtai, Y., Pines, O. & Gutnick, D. (1985). Emulsan: a case study of microbial capsules as industrial products. In *Developments in Industrial Microbiology*, pp. 291–307. Amsterdam: Elsevier.

Singh, S., Singh, U., Hogan, S. E. & Feingold, D. S. (1990). Formation of UDP-2-acetamido-2-deoxy-L-galactose and UDP-2-acetamido-2-deoxy-L-galacturonic acid by *Pseudomonas aeruginosa*. *J Bacteriol* **172**, 299–304.

Skurnik, M., Venho, R., Toivanen, P. & al-Hendy, A. (1995). A novel locus of *Yersinia enterocolitica* serotype O:3 involved in lipopolysaccharide outer core biosynthesis. *Mol Microbiol* **17**, 575–594.

Sonnhammer, E. L., von Heijne, G. & Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. In *Proceedings of the International Conference on Intelligent Systems in Molecular Biology*, pp. 175–182. Edited by

J. Glasgow, T. Littlejohn, F. Major, R. Lathrop, D. Sankoff & C. Sensen. Menlo Park, CA: AAAI Press.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**, 503–517.

Stroeher, U. H., Karageorgos, L. E., Brown, M. H., Morona, R. & Manning, P. A. (1995). A putative pathway for perosamine biosynthesis is the first function encoded within the *rfb* region of *Vibrio cholerae* O1. *Gene* **166**, 33–42.

Trefzer, A., Salas, J. A. & Bechthold, A. (1999). Genes and enzymes involved in deoxysugar biosynthesis in bacteria. *Nat Prod Rep* **16**, 283–299.

Vincent, C., Doublet, P., Grangeasse, C., Vaganay, E., Cozzone, A. J. & Duclos, B. (1999). Cells of *Escherichia coli* contain a protein-tyrosine kinase, Wzc, and a phosphotyrosine-protein phosphatase, Wzb. *J Bacteriol* **181**, 3472–3477.

Wang, L., Liu, D. & Reeves, P. R. (1996). C-terminal half of *Salmonella enterica* WbaP (RfbP) is the galactosyl-1-phosphate transferase domain catalyzing the first step of O-antigen synthesis. *J Bacteriol* **178**, 2598–2604.

Weissborn, A. C., Liu, Q., Rumley, M. K. & Kennedy, E. P. (1994). UTP: alpha-D-glucose-1-phosphate uridylyltransferase of *Escherichia coli*: isolation and DNA sequence of the *gal*U gene and purification of the enzyme. *J Bacteriol* **176**, 2611–2618.

Whitfield, C. (1995). Biosynthesis of lipopolysaccharide O antigens. *Trends Microbiol* **3**, 178–185.

Whitfield, C. & Roberts, I. S. (1999). Structure, assembly and regulation of expression of capsules in *Escherichia coli*. *Mol Microbiol* **31**, 1307–1319.

Ye, R. W., Zielinski, N. A. & Chakrabarty, A. M. (1994). Purification and characterization of phosphomannomutase/phosphoglucomutase from *Pseudomonas aeruginosa* involved in biosynthesis of both alginate and lipopolysaccharide. *J Bacteriol* **176**, 4851–4857.

Zosim, Z., Rosenberg, E. & Gutnick, D. L. (1986). Changes in hydrocarbon emulsification specificity of the polymeric bioemulsifier emulsan: effects of alkanols. *Colloid Polym Sci* **264**, 218–223.

Zuckerberg, A., Diver, A., Peeri, Z., Gutnick, D. L. & Rosenberg, E. (1979). Emulsifier of *Arthrobacter* RAG-1: chemical and physical properties. *Appl Environ Microbiol* **37**, 414–420.

Received 30 October 2000; revised 9 March 2001; accepted 23 March 2001.