Computer-based analyses of the protein constituents of transport systems catalysing export of complex carbohydrates in bacteria

Ian T. Paulsen, Alison M. Beness and Milton H. Saier, Jr

Author for correspondence: Milton H. Saier, Jr. Tel: +1 619 534 4084. Fax: +1 619 534 7108.
e-mail: msaier@ucsd.edu

Department of Biology,
University of California at
San Diego, La Jolla, CA
92093-0116, USA

Bacteria synthesize and secrete an array of complex carbohydrates including exopolysaccharides (EPSs), capsular polysaccharides (CPSs), lipopolysaccharides (LPSs), lipo-oligosaccharides (LOSs) and teichoic acids (TCAs). We have analysed the families of homologous proteins that appear to mediate excretion of complex carbohydrates into or across the bacterial cell envelope. Two principal families of cytoplasmic-membrane transport systems appear to drive polysaccharide export: polysaccharide-specific transport (PST) systems and ATP-binding cassette-2 (ABC-2) systems. We present evidence that the secretion of CPSs and EPSs, but not of LPSs, LOSs or TCAs via a PST or ABC-2 system requires the presence of a cytoplasmic-membrane-periplasmic auxiliary protein (MPA1 or MPA2, respectively) in both Gram-negative and Gram-positive bacteria as well as an outer-membrane auxiliary (OMA) protein in Gram-negative bacteria. While all OMA proteins are included within a single family, MPA1 and MPA2 family proteins are not demonstrably homologous to each other, even though they share common topological features. Moreover, MPA1 family proteins (which function with PST systems), but not MPA2 family proteins (which function with ABC-2 systems), possess cytoplasmic ATP-binding domains that may either exist as separate polypeptide chains (for those from Gram-positive bacteria) or constitute the C-terminal domain of the MPA1 polypeptide chain (for those from Gram-negative bacteria). The sizes, substrate specificities and regions of relative conservation and hydrophobicity are defined allowing functional and structural predictions as well as delineation of family-specific sequence motifs. Each family is characterized phylogenetically.

Keywords: transport, exopolysaccharides, capsular polysaccharides, lipopolysaccharides, lipo-oligosaccharides, teichoic acids

INTRODUCTION

All living organisms display on their cell surfaces an array of complex carbohydrates which serve as species-, tissue- and cell-type-specific recognition constituents (Saier & Jacobson, 1984). In addition to recognition functions, carbohydrates provide protective functions to the cell that produces them (Jann & Jann, 1990; Reeves,

1992; Bik *et al.*, 1995; Roberts, 1995). The presence of cytoplasmically synthesized hydrophilic macro-molecules within or external to the cell envelope implies the existence of transport mechanisms that allow export of biosynthetic intermediates or the final products of biosynthesis across the cytoplasmic membrane (Kroll *et al.*, 1990; Smith *et al.*, 1990; Manning *et al.*, 1995; Rosenow *et al.*, 1995). In the case of Gram-positive bacteria, transport across the plasma membrane may be sufficient to allow the polysaccharide product to reach its final destination, but in the case of Gram-negative bacteria, mechanisms must exist that allow transport across both membranes of the envelope (Frosch *et al.*, 1992; Dinh *et al.*, 1994; Guidolin *et al.*, 1994; Lin *et al.*, 1994; Lazarevic & Karamata, 1995).

Extracellular bacterial polysaccharides include (a) the lipopolysaccharides (LPSs) embedded in the outer membranes of Gram-negative prokaryotes, (b) teichoic acids (TCAs) found associated with the membranes and walls of Gram-positive bacteria, (c) capsular poly-saccharides (CPSs) of both Gram-positive and Gram-negative bacteria, (d) exopolysaccharides (EPSs), also of both Gram-positive and Gram-negative organisms, and (e) lipo-oligosaccharides (LOSs) secreted as signalling molecules by nodulating bacteria when in symbiotic association with leguminous plants. The various carbohydrate-containing constituents localized to the cell surface are frequently essential virulence factors of pathogenic bacteria as they provide protective functions that allow the organism to combat host defence mechanisms (Hornick et al., 1970; Robbins & Robbins, 1984; Pearce & Roberts, 1995). These macromolecules often exhibit the characteristic of phase variation, allowing the bacteria to successfully evade the immune system of the host (Saier & Jacobson, 1984; Bartlett et al., 1988; Seifert & So, 1988; Dybvig, 1993).

While synthesis of the various classes of bacterial polysaccharides noted above has been extensively stud-ied, relatively little is known about the proteins and mechanisms responsible for their export. In the case of cell-surface LPSs, the two structural constituents (core and O-antigen chain) may be assembled independently. Following completion of the core and O-antigen chain, these moieties may be exported to the periplasmic surface of the cytoplasmic membrane by energy-de-pendent processes where they are joined (Mulford & Osborn, 1983; McGrath & Osborn, 1991; Reeves, 1993; Kido et al., 1995). In other cases, it seems that the entire LPS is assembled at the cytoplasmic face of the cyto-plasmic membrane and then translocated intact to its final destination in the outer leaflet of the outer membrane (Jann & Jann, 1984; Jann et al., 1985). These facts imply that more than one transport system may be required for export of LPSs, and five plausible mechanisms have been proposed (Schnaitman & Klena, 1993). While the molecular details of these processes are poorly understood for LPSs, even less is known about the corresponding processes for CPSs and EPSs. These external polysaccharides may be exported by a one-, two- or multi-step process (Boulnois & Jann, 1989; Kröncke et al., 1990; see Discussion).

In this paper we have analysed the sequences of transport proteins concerned with the export of various poly-saccharides. We have identified and characterized three novel families of auxiliary proteins associated with two evolutionarily distinct families of cytoplasmic-membrane-localized carbohydrate export systems.

METHODS

Computer methods. The FASTA (Pearson & Lipman, 1988) and BLAST (Altschul et al., 1990) programs were used to screen the peptide and translated nucleotide databases. The statistical significance of sequence similarities between putative members of the various families was established using the

RDF2 (Pearson & Lipman, 1988) and GAP (Devereux et al., 1984) programs with 200 random shuffles. Binary comparison scores are expressed as SD (Dayhoff et al., 1983). A value of 10 SD was deemed sufficient to establish homology (Saier, 1994).

Multiple sequence alignments were constructed using the TREE (Feng & Doolittle, 1990) and PILEUP (Devereux et al., 1984) programs. Phylogenetic analyses were performed using TREE.

Mean hydropathy and mean similarity analyses were cond-ucted for all protein families analysed. They were based on the complete multiple alignments, only small portions of which are presented in this report. The hydropathy analyses were conducted essentially as described by Kyte & Doolittle (1982) with a sliding window of 20 residues except for the analyses conducted for the OMA family where a sliding window of 7 residues was used. Similarly, a sliding window of 7 residues for the OMA family and of 20 residues for all other families was used to generate the mean similarity plots. Most of these analyses are not presented and are merely described in the text. However, these plots will be provided to the interested reader upon request from I. T. Paulsen.

Charge bias analysis of membrane protein topology was performed using the program TOP PRED (von Heijne, 1992). Signature sequences were defined according to Bairoch (1992). The programs MEME and MAST (Bailey & Elkan, 1994) were also utilized to help identify conserved motifs within the protein families. All of the methods used have been applied to numerous transport proteins and evaluated (see Saier, 1994, 1996, for recent reviews).

RESULTS

Polysaccharide transport proteins

It has been noted previously that Gram-negative bac-terial CPS and EPS export systems may require the presence of system-specific auxiliary proteins (Boulnois & Roberts, 1990; Kroll et al., 1990; Frosch et al., 1991; Hashimoto et al., 1993; Pearce & Roberts, 1995). These auxiliary proteins are typically encoded within operons or gene clusters concerned with complex carbohydrate biosynthesis and export. Only a few of these proteins have been functionally characterized. In an attempt to understand the constituents required for polysaccharide transport, we initially identified all homologues of the putative outer-membrane auxiliary proteins and sub-sequently searched for potential cytoplasmic-membrane transport proteins as well as additional auxiliary con-stituents. The proteins identified, their putative sub-cellular localization and their family associations are summarized in Tables 1, 2 and 3. Table 1 lists the bacterial sources of the auxiliary proteins as well as the abbreviations used throughout this analysis; Tables 2 and 3 provide corresponding data for the cytoplasmic-membrane transport constituent proteins of the polysaccharide-specific transport (PST) and ATP-bind-ing cassette-2 (ABC-2) permeases, respectively.

We identified 10 putative outer-membrane homologues that compose a family which we designated the outer-membrane auxiliary (OMA) family (Table 1). These proteins derive exclusively from Gram-negative bac-

Table 1. Auxiliary proteins of the OMA, MPA2 and MPA1 families

Protein	Organism	Function	Accession no.	Database*	Size (aa)
OMA family					
GumB	*Xanthomonas campestris*	EPS export	U22511	GP	213
OtnA	*Vibrio cholerae*	CPS export	X79822	GP	>260
KpsD	*Escherichia coli*	CPS export	Q03961	SP	558
ExoF	*Sinorhizobium meliloti*	EPS export	Q02728	SP	421
Orf4	*Klebsiella pneumoniae*	CPS export	D21242	GP	378
AmsH	*Erwinia amylovora*	EPS export	X77921	GP	377
EpsA	*Pseudomonas solanacearum*	EPS export	U17898	GP	377
BexD	*Haemophilus influenzae*	CPS export	P22236	SP	394
CtrA	*Neisseria meningitidis*	CPS export	A43845	PIR	387
VexA	*Salmonella typhi*	CPS export	D14156	GP	355
MPA2 family					
KpsE	*Escherichia coli*	CPS export	L19929	GP	382
BexC	*Haemophilus influenzae*	CPS export	P22930	SP	377
CtrB	*Neisseria meningitidis*	CPS export	M57677	GP	387
VexD	*Salmonella typhi*	CPS export	D14156	GP	434
MPA1 family					
ExoP	*Sinorhizobium meliloti*	EPS export	P33698	SP	786
EpsB	*Pseudomonas solanacearum*	EPS export	U17898	GP	750
AmsA	*Erwinia amylovora*	EPS export	X77921	GP	726
OtnB	*Vibrio cholerae*	CPS export	X79822	GP	335
Orf6	*Klebsiella pneumoniae*	CPS export	D21242	GP	722
GumC	*Xanthomonas campestris*	EPS export	U22511	GP	377
CapAB	*Staphylococcus aureus*	CPS export	U10927	GP	221/228†
CpsCD	*Streptococcus pneumoniae*	CPS export	U09239	GP	230/227†
CpsBC	*Streptococcus agalactiae*	CPS export	L09116	GP	230/183†

* GP, GenBank protein database; SP, SWISS-PROT.

† The two numbers separated by a solidus indicate the number of amino acids in the two constituents (MPA1 and C) of Gram-positive bacterial MPA1 family proteins.

Table 2. Cytoplasmic-membrane transporters of the PST family concerned with complex carbohydrate transport

PST subfamily	Protein	Organism	Function	Accession no.	Database*	Size (aa)
PST(1)	GumJ	*Xanthomonas campestris*	EPS export	U22511	GP	498
	Orf11	*Klebsiella pneumoniae*	CPS export	D21242	GP	485
	ExoT	*Sinorhizobium meliloti*	EPS export	L20758	GP	582
	AmsL	*Erwinia amylovora*	EPS export	X77921	GP	446
	CapF	*Staphylococcus aureus*	CPS export	P39855	SP	396
PST(2)	RfbE	*Shigella flexneri*	LPS export	P37781	SP	418
	RfbX1	*Escherichia coli*	LPS export	P37746	SP	415
	RfbX2	*Shigella dysenteriae*	LPS export	Q03583	SP	396
	RfbX3	*Yersinia pseudotuberculosis*	LPS export	H47070	PIR	437
	RfbX4	*Salmonella enterica*	LPS export	A42476	PIR	432
	TrsA	*Yersinia enterocolitica*	LPS export	S30030	PIR	418
	YifJ	*Escherichia coli*	Unknown	P27834	SP	416
	HI0867	*Haemophilus influenzae*	Unknown	U32768	GP	404
	HI1700	*Haemophilus influenzae*	Unknown	U32842	GP	404

* GP, GenBank protein database; SP, SWISS-PROT.

Table 3. Cytoplasmic-membrane transporters of the ABC-2 family concerned with complex carbohydrate and drug transport

Columns 1–4 provide details for the integral-membrane constituents of the permeases; columns 5–8 provide details for the corresponding ATP-binding constituents.

Membrane protein	Accession no.	Database	Size (aa)	ABC-2 protein	Accession no.	Database	Size (aa)	Organism	Function
KpsM	P24584	SP	258	KpsT	P24586	SP	218	*Escherichia coli*	CPS export
BexB	P19391	SP	265	BexA	P10640	SP	217	*Haemophilus influenzae*	CPS export
CtrC	P32015	SP	265	CtrD	P32016	SP	216	*Neisseria meningitidis*	CPS export
VexB	D14156	GP	264	VexC	D14156	GP	246	*Salmonella typhi*	CPS export
NodJ1	X87578	GP	168	NodI1	X87578	GP	347	*Rhizobium galegae*	LOS export
NodJ2	P26025	SP	262	NodI2	P26050	SP	306	*Bradyrhizobium japonicum*	LOS export
NodJ3	P06755	SP	259	NodI3	P08720	SP	311	*Rhizobium leguminosarum*	LOS export
NodJ4	Q07757	SP	254	NodI4	Q07756	SP	320	*Azorhizobium caulinodans*	LOS export
DrrB1	U00023	GP	288	DrrA1	U00023	GP	331	*Mycobacterium leprae*	Daunorubicin resistance
DrrC	U00023	GP	244	DrrA1	U00023	GP	331	*Mycobacterium leprae*	Daunorubicin resistance
DrrB2	M73758	GP	283	DrrA2	P32010	SP	330	*Streptomyces peucetius*	Daunorubicin resistance
RfbA1	L41518	GP	259	RfbB1	L41518	GP	246	*Klebsiella pneumoniae*	LPS export
RfbA2	L34166	GP	255	RfbB2	L31466	GP	246	*Serratia marcescens*	LPS export
RfbD	U24571	GP	259	RfbE	S28581	PIR	239	*Yersinia enterocolitica*	LPS export
RfbH	X59554	GP	257	RfbI	S28474	PIR	250	*Vibrio cholerae*	LPS export
Orf261	D43637	GP	256	Orf431	D43637	GP	431	*Escherichia coli*	LPS export
Sll0760	D64000	GP	294	Sll0759	D64000	GP	347	*Synechocystis* sp.	Unknown
OleC5	S32909	PIR	273	OleC4	S32904	PIR	325	*Streptomyces antibioticus*	Oleandomycin resistance
TagG	U13832	GP	275	TagH	P42954	SP	527	*Bacillus subtilis*	TCA export
YhhJ	U00039	GP	374	YhiG	U00039	GP	308	*Escherichia coli*	Unknown
YadH	P36880	SP	203	YadG	P36879	SP	308	*Escherichia coli*	Unknown
Ycf38	U30821	GP	290	–	–	–	–	*Cyanophora paradoxa*	Unknown

teria. All appear to be involved in the transport of CPSs or EPSs (Kroll *et al.*, 1990; Frosch *et al.*, 1991; Hashimoto *et al.*, 1993). Structural analyses (see below) suggest that these proteins are outer-membrane constituents rather than periplasmic proteins (Dorman, 1995).

In each operon or gene cluster encoding an OMA protein a gene encoding an additional auxiliary protein of distinctive topology was found. Each of these latter proteins was found to possess two putative transmembrane-spanning segments (TMSs) separated by a large hydrophilic loop. On the basis of sequence analyses, these proteins fall into two distinct families that do not exhibit significant sequence similarity with each other (using several programs, including RDF2, GAP, MAST, etc.; see Methods). We have designated these two families the cytoplasmic-membrane-periplasmic auxiliary (MPA) 1 and 2 families. The MPA2 family proteins derive exclusively from Gram-negative bacteria, but the MPA1 family includes both Gram-positive and Gram-negative bacterial members. The MPA1 proteins from Gram-negative bacteria generally exhibit large C-terminal hydrophilic domains (the C-domains)

which include an ATP-binding motif (Walker *et al.*, 1982; Higgins, 1992). The corresponding Gram-positive bacterial MPA1 proteins lack the C-domain, but a homologous C-protein is encoded within each of the gene clusters that encodes a MPA1 protein, suggesting that the C-proteins of Gram-positive bacteria serve as the functional analogue of the C-domains in the Gram-negative bacterial MPA1 proteins. We therefore designated the Gram-negative bacterial proteins MPA1-C and the corresponding Gram-positive bacterial proteins MPA1 + C.

Each fully sequenced polysaccharide production gene cluster encoding both an OMA protein and an MPA protein also encodes a putative cytoplasmic-membrane transporter (Tables 2 and 3). The latter proteins could be divided into two families. The larger of these two families is the ABC-2 subfamily of the ABC superfamily with an integral-membrane 6-spanner protein acting in conjunction with an ABC protein (Table 3; Higgins, 1992; Reizer *et al.*, 1992; Fath & Kolter, 1993; Tam & Saier, 1993; Kuan *et al.*, 1995). The other family represents a novel family of putative transporters with 12 TMSs. This family, which we have designated the

PST family, includes CPS-, EPS- and LPS-specific transport proteins (Table 2) (Boulnois & Roberts, 1990; Huang & Schell, 1995).

A summary of the findings concerning auxiliary protein-dependent transport systems, to be reported in more detail below, is provided in Tables 4 and 5, and a schematic depiction of the relevant proteins is presented in Fig. 1. As summarized in Tables 4 and 5, each gene cluster encoding an OMA protein also encodes either an MPA1 or an MPA2 protein, but not both. Also found within gene clusters encoding OMA–MPA2 proteins are pairs of genes encoding 2-component ABC-2 systems. In most cases, gene clusters encoding an OMA–MPA1 protein pair also possess a gene encoding a PST protein. In one Gram-positive bacterium which possesses an MPA1 homologue, a PST protein is encoded within the same gene cluster although no OMA homologue encoded within the same cluster could be identified. Thus, in Gram-negative bacteria, each pair of OMA–MPA2 auxiliary proteins appears to function with an ABC-2 transport system while each pair of OMA–MPA1 proteins probably functions with a PST system.

Fig. 1 provides a schematic description of the three types of auxiliary protein-dependent systems revealed in our study, all of which transport CPSs or EPSs. In Fig. 1(a) an ABC-2 transporter is postulated to function in conjunction with an MPA2 protein and an OMA protein. In Fig. 1(b) a PST transporter acts in conjunction with an MPA1-C protein and an OMA protein. The Gram-negative bacterial MPA1-C protein possesses a C-terminal cytoplasmic C-domain that bears an ATP-binding site. In Fig. 1(c), depicting the generalized homologous Gram-positive bacterial system, two polypeptide chains (MPA1 + C) are encoded by two distinct genes. One of these proteins (MPA1) bears two TMSs and an abbreviated extracytoplasmic loop while the other (C) is a soluble protein, bearing the ATP-binding site. Together these two proteins constitute the near equivalent of a complete Gram-negative bacterial MPA1-C protein. These Gram-positive bacterial proteins probably function in conjunction with PST transporters.

The OMA family

The OMA family consists of 10 homologues, all of which probably function in polysaccharide export. Several members of the family have been shown to be essential for the production of EPSs, and in a few cases, evidence for a translocation function is available, e.g. CtrA (Frosch et al., 1989, 1991) and KpsD (Bronner et al., 1993a,b). Representative proteins of this family have been found localized to the outer membrane, e.g. CtrA (Frosch et al., 1991) and EpsA (Huang & Schell, 1995), although KpsD has been reported to be a periplasmic protein (Silver et al., 1987; Dorman, 1995). All members of the family gave comparison scores of at least 10 SD with other established members of the family, thereby establishing homology.
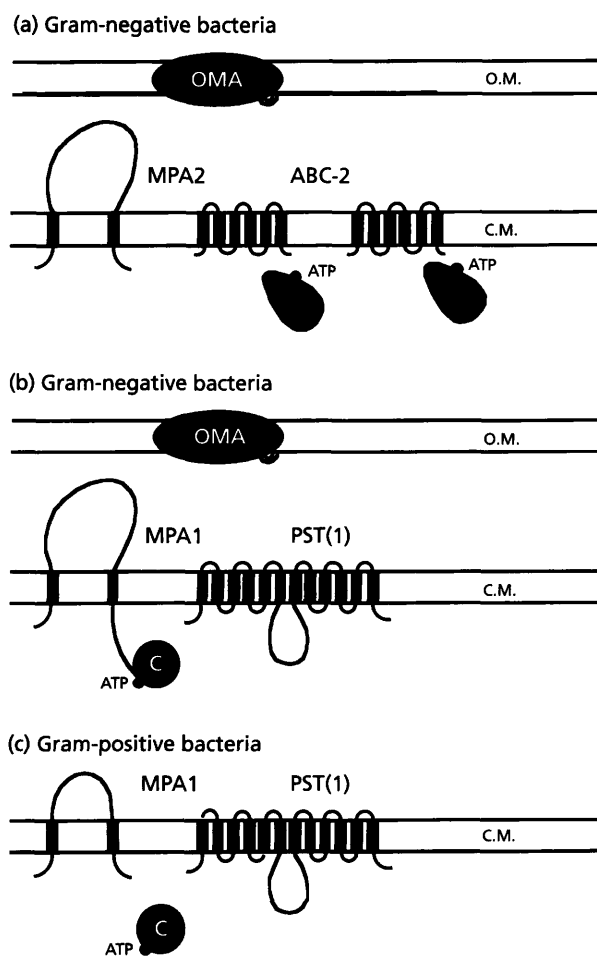

Fig. 1. Schematic representation of three types of carbohydrate export systems in bacteria that utilize auxiliary proteins to facilitate the export process. (a) Prototype CPS export system in Gram-negative bacteria consisting of four dissimilar proteins. These constituents are (1) a membrane-embedded channel-forming constituent of an ABC-2-type transport system, probably a homodimeric six-TMS protein, (2) an ATP-binding cassette (ABC) constituent of the same system, probably present in two copies, (3) an MPA2 family member and (4) an OMA family member. (b) Prototypical CPS or EPS export system in Gram-negative bacteria consisting of three dissimilar proteins. These constituents are (1) a cytoplasmic membrane PST(1) transport protein, (2) an MPA1 family member with a cytoplasmic domain (C) that bears an ATP-binding motif and (3) an OMA family member. (c) Prototypical CPS export system in Gram-positive bacteria consisting of three dissimilar proteins. These constituents are (1) a cytoplasmic-membrane PST(1) family member, (2) an MPA1 family protein and (3) a cytoplasmic protein (C) bearing an ATP-binding motif. Note that the MPA1+C proteins of Gram-positive bacteria depicted in (c) constitute the structural equivalent of the MPA1-C proteins of Gram-negative bacteria depicted in (b). C.M., cytoplasmic membrane; O.M., outer LPS-containing membrane. The small black spheres labelled ATP on the ABC and C proteins/domains indicate the putative ATP-binding sites present in these proteins.

Mean hydropathy and similarity plots for the OMA protein family were derived from the multiple alignment of these protein sequences (data not shown). Each OMA protein exhibited an N-terminal hydrophobic segment.
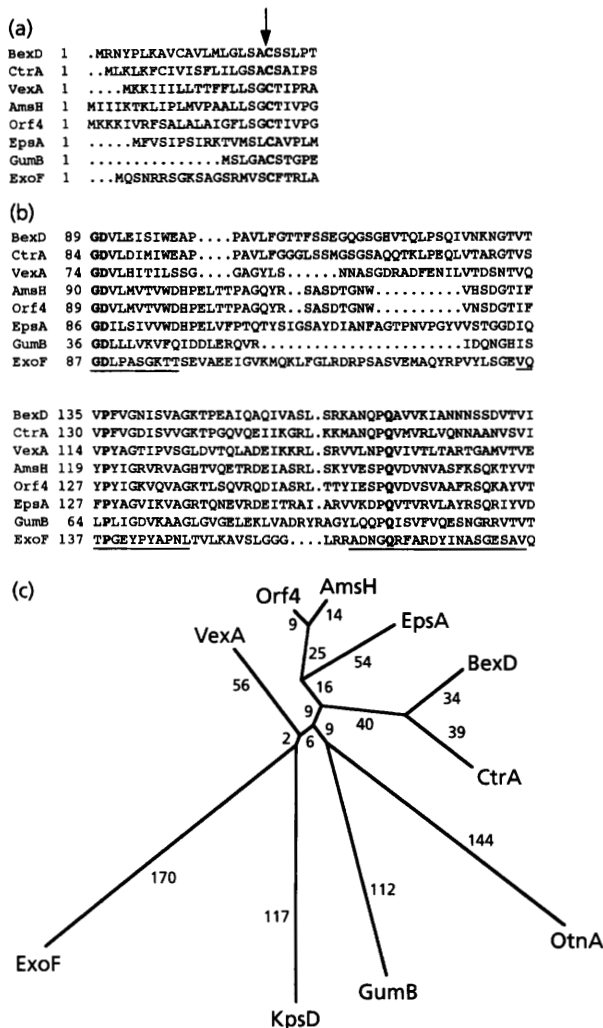
(a)

```
BexD   1  .MRNYPLKAVCAVLMLGLSACSSLPT
CtrA   1  ..MLKLKFCIVISFLILGSACSAIPS
VexA   1  ....MKKIIILLTTFFLLSGCTIPRA
AmsH   1  MIIIKTKLIPLMVPAALLSGCTIVPG
Orf4   1  MKKKIVRFSALALAIGFLSGCTIVPG
EpsA   1  .....MFVSIPSIRKTVMSLCAVPLM
GumB   1  ...............MSLGACSTGPE
ExoF   1  ...MQSNRRSGKSAGSRMVSCFTRLA
```

(b)

```
BexD   89  GDVLEISIWEAP....PAVLFGTTFSSEGQGSGEVTQLPSQIVNKNGTVT
CtrA   84  GDVLDIMIWEAP....PAVLFGGGLSSMGSGSAQQTKLPEQLVTARGTVS
VexA   74  GDVLHITILSSG....GAGYLS.....NNASGDRADFENILVTDSNTVQ
AmsH   90  GDVLMVTVWDHPELTTPAGQYR..SASDTGNW..........VHSDGTIF
Orf4   89  GDVLMVTVWDHPELTTPAGQYR..SASDTGNW..........VNSDGTIF
EpsA   86  GDILSIVVWDHPELVFPTQTYSIGSAYDIANFAGTPNVPGYVVSTGGDIQ
GumB   36  GDLLLVKVFQIDDLERQVR.......................IDQNGHIS
ExoF   87  GDLPASGKTTSEVAEEIGVKMQKLFGLRDRPSASVEMAQYRPVYLSGEVQ
```

```
BexD  135  VPFVGNISVAGKTPEAIQAQIVASL.SRKANQPQAVVKIANNNSSDVTVI
CtrA  130  VPFVGDISVVGKTPGQVQEIIKGRL.KKMANQPQVMVRLVQNNAANVSVI
VexA  114  VPYAGTIPVSGLDVTQLADEIKKRL.SRVVLNPQVIVTLTARTGAMVTVE
AmsH  119  YPYIGRVRVAGHTVQETRDEIASRL.SKYVESPQVDVNVASFKSQKTYVT
Orf4  127  YPYIGKVQVAGKTLSQVRQDIASRL.TTYIESPQVDVSVAAFRSQKAYVT
EpsA  127  FPYAGVIKVAGRTQNEVRDEITRAI.ARVVKDPQVTVRVLAYRSQRIYVD
GumB   64  LPLIGDVKAAGLGVGELEKLVADRYRAGYLQQPQISVFVQESNGRRVTVT
ExoF  137  TPGEYPYAPNLTVLKAVSLGGG....LRRADNGQRFARDYINASGESAVQ
```

(c)

Fig. 2. Alignment of selected regions and phylogenetic tree of the proteins of the OMA family. (a) Alignment of the N-terminal regions of the OMA family proteins displaying the conserved signal peptidase cleavage site (as indicated by the arrow). Note that GumB may be N-terminally truncated due to a sequencing error. (b) Alignment of a well conserved region of OMA family proteins. In both (a) and (b), residue numbers in each protein are displayed to the left of the alignment and fully conserved residues are highlighted in bold type. Lines underneath the alignment shown in (b) indicate those regions of greatest sequence similarity selected for analyses (see Results). The OMA family protein OtnA has not been included in the alignments presented as the complete sequence of this protein is not currently available. (c) Phylogenetic tree of the OMA family prepared using the TREE program of Feng & Doolittle (1990). Numerical values for the branch lengths (drawn to scale and expressed in arbitrary units) are provided adjacent to the branches.

Examination of these sequences revealed that following this leader sequence, the majority of the OMA proteins possess a reasonable match to the consensus sequence for signal peptidase II cleavage (see arrow in Fig. 2a), including the invariant cysteine (marked in bold print in Fig. 2a), which, following cleavage, becomes covalently modified to yield a lipoprotein. Interestingly, the *Escherichia coli* KpsD protein is a possible exception to

this observation as it contains only a partial match with a signal peptidase II cleavage consensus sequence, suggesting that it may not be modified. This divergence may correlate with the suggestion that KpsD is a periplasmic protein (Silver *et al.*, 1987; Dorman, 1995).

The remaining portions of the proteins contain many short, moderately hydrophobic regions that are likely to form membrane-embedded β-strands. This last conclusion is supported in part by secondary structural analyses (Chou & Fasman, 1978; Garnier *et al.*, 1978) that predict that these proteins are predominantly of β-structure. The N-terminal hydrophobic peaks display poor conservation with the exception of the lipoprotein cleavage site.

Fig. 2(b) shows those regions of the OMA family proteins that exhibit the highest degree of residue conservation. Within the segment of the multiple alignment shown, there are three regions exhibiting striking conservation. With the exception of the ExoF protein, these regions (underlined in Fig. 2b) exhibit generalized consensus sequences as follows (alternative residues at a particular position are in parentheses; X = any residue):

(1) GD(LIV)L*X(LIV)*X(LIV)*(WFLIV)(EDQNST)**

(2) (LIV)*X(LIVFY)*P(LIVFY)*(LIVA)G*X(LIV)*X (LIVA)*XG

(3) (LIVA)X(EDQNST)P*Q(LIVA)*XV*X(LIV)*X (EDQNSTA)**X$_2$(GSA)X$_3$(STY)V

Each of these generalized consensus sequences displays regions of at least three alternating hydrophobic residues (indicated by single asterisks in the three sequences shown above), usually separated by a variable residue (X). The fourth alternating position in two of these sequences (1 and 3) is hydrophilic and of essentially the same composition (indicated by double asterisks). By contrast, in consensus sequence (2), the alternating hydrophobic–X sequence proceeds without interruption for 12 residues. This general pattern, common to all three sequences, undoubtedly reflects the β-strand secondary structures typical of outer-membrane proteins. None of these consensus sequences proved to be specific to the OMA family; therefore, they are not suitable for use as signature sequences.

The phylogenetic tree for the OMA family is shown in Fig. 2(c). As can be seen, six of the OMA family members cluster together while the remaining four are far more divergent. The six proteins that cluster together proved to be of fairly uniform size (355–394 residues), while the remaining proteins may be either much shorter (GumB and OtnA) or much longer (ExoF and KpsD). The most distant member of the family, ExoF, is from *Sinorhizobium meliloti* which is the most divergent bacterium among those represented (Table 1).

The MPA2 family

The MPA2 family has only four sequenced protein members. As noted above and in Table 4, these proteins appear to function in conjunction with an OMA protein and an ABC-2-type transport system. Each set of these

Table 4. Organization of bacterial CPS and EPS export systems

A dash indicates the apparent absence of a protein of the indicated family encoded in this particular gene cluster. Unknown indicates uncertainty with respect to the presence of such a protein encoded in the gene cluster.

OMA	MPA2	MPA1	PST	ABC	Other (?)*
GumB	–	GumC	GumJ	–	–
OtnA	–	OtnB	Unknown	Unknown	–
KpsD	KpsE	–	–	KpsT/KpsM	–
ExoF	–	ExoP	ExoT	–	–
Orf4	–	Orf6	Orf11	–	–
AmsH	–	AmsA	AmsL	–	–
EpsA	–	EpsB	Unknown	Unknown	EpsE/EpsF
VexA	VexD	–	–	VexB/VexC	–
BexD	BexC	–	–	BexA/BexB	–
CtrA	CtrB	–	–	CtrC/CtrD	–
–	–	CapA/CapB	CapF	–	–
–	–	CpsB/CpsC	Unknown	Unknown	–
–	–	CpsC/CpsD	Unknown	Unknown	–

* The sequenced portion of the *eps* gene cluster does not include a gene encoding a PST or ABC permease. The *epsE* and *epsF* genes possibly encode transport proteins based on putative membrane topology, but they lack statistically significant similarity to known transport proteins or to other sequenced proteins in the current databases.

proteins presumably constitutes a polysaccharide export system. MPA2 proteins are all derived from Gram-negative bacteria as was observed for the OMA proteins. They are homologous, exhibiting comparison scores with each other in excess of 30 SD. Experimental evidence has suggested that KpsE (Bronner *et al.*, 1993a,b), VexD (Hashimoto *et al.*, 1993) and CtrB (Frosch *et al.*, 1989) are involved in polysaccharide translocation.

Mean hydropathy and similarity plots for the aligned MPA2 family members were derived (data not shown). The hydropathy plot revealed two putative TMSs, one near the N termini and a second at the C termini of the MPA2 proteins. Charge-bias analysis (von Heijne, 1992) indicated that the N and C termini are localized to the cytoplasm while the large central loop is localized to the periplasm. Supporting this proposed topological model, β-lactamase or alkaline phosphatase fusion studies with BexC (Kroll *et al.*, 1990), CtrB (Frosch *et al.*, 1991) and KpsE (Rosenow *et al.*, 1995) have indicated that the central regions of these proteins are situated in the periplasmic space. The proteins were found to exhibit a comparable degree of similarity throughout their lengths except for the cytoplasmic N termini which are both divergent in sequence and of variable length (see Table 1). Secondary structural analysis (Chou & Fasman, 1978; Garnier *et al.*, 1978) predicted that all four proteins exhibited a pronounced α-helical structure in the central 50 residues of their periplasmic loops.

Figs 3(a) and (b) show the multiple alignments of two regions of the MPA2 proteins that exhibit marked sequence conservation. The first of these is derived from a region near the centre of the periplasmic loop corresponding to the region of predicted α-helical structure. The second is derived from the C-terminal portion of this loop, just preceding the second TMS. Two potential signature sequences were derived from these regions (see underlined regions in Figs 3a and b, respectively). These sequences are:

(1) $(EQ)X(LIV)LX_3EX_2(LIV)NX_4RX_3(DE)X(LIV)X$
$(FYW)XEX_2(LIVM)X_2A$

(2) $(EQ)AXRX_4LX(LIV)ISXP$

The N-terminal 20 residues of signature sequence (1) form a strongly amphipathic α-helix with hydrophilic residues localized almost exclusively to one side of the helix and hydrophobic residues localized almost exclusively to the other. Residues 21–33 in this signature sequence similarly form a strongly amphipathic α-helix, but it is out of phase with that formed by residues 1–20. The spacing of regions of conservation, corresponding roughly to one helical turn, is particularly noteworthy in the first half of signature sequence (1). By contrast with signature sequence (1), sequence (2) cannot form an amphipathic α-helix. Both of these potential signature sequences were screened against the SWISS-PROT database. Only members of the MPA2 family were retrieved with sequence (1), but other proteins were retrieved with sequence (2). Sequence (1) is therefore a valid signature sequence, characteristic of the MPA2 family.

The phylogenetic tree for the MPA2 family is shown in Fig. 3(c). While BexC and CtrB cluster closely together, KpsE is more distant and VexD is even more distant from the other three members. Comparing this tree with that for the OMA family proteins (Fig. 2c), it should be
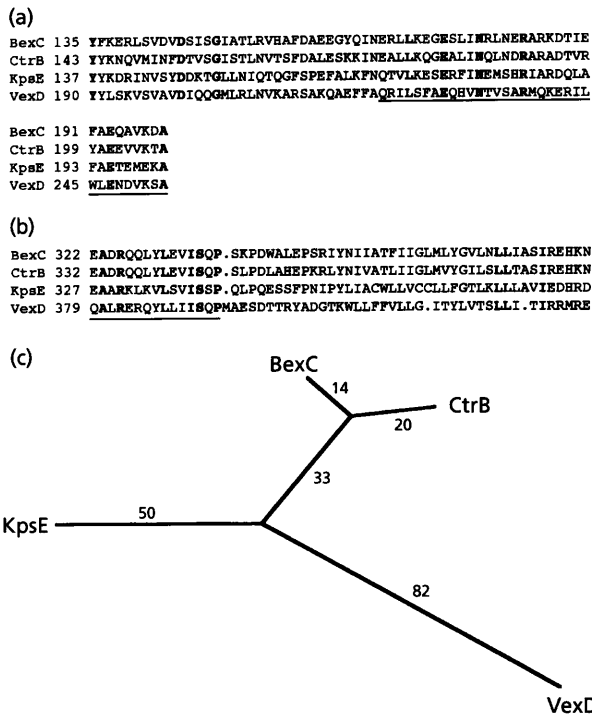
(a)

```
BexC 135 YFKERLSVDVDSISGIATLRVEAFDAEEGYQINERLLKEGESLIMRLNERARKDTIE
CtrB 143 YYKNQVMINFDTVSGISTLNVTSFDALESKKINEALLKQGEALIMQLNDRARADTVR
KpsE 137 YYKDRINVSYDDKTGLLNIQTQGFSPEFALKFNQTVLKESERFIMEMSHRIARDQLA
VexD 190 YYLSKVSVAVDIQQGMLRLNVKARSAKQAEFFAQRILSFAEQBVMTVSARMQKERIL
```

```
BexC 191 FAEQAVKDA
CtrB 199 YAEEVVKTA
KpsE 193 FAETEMEKA
VexD 245 WLENDVKSA
```

(b)

```
BexC 322 EADRQQLYLEVISQP.SKPDWALEPSRIYNIIATFIIGLMLYGVLNLLIASIREHKN
CtrB 332 EADRQQLYLEVISQP.SLPDLAHEPKRLYNIVATLIIGLMVYGILSLLTASIREHKN
KpsE 327 EAARKLKVLSVISSP.QLPQESSFPNIPYLIACWLLVCCLLFGTLKLLLAVIEDHRD
VexD 379 QALRERQYLLIISQPMAESDTTRYADGTKWLLFFVLLG.ITYLVTSLLI.TIRRMRE
```

(c)

Fig. 3. Alignment of selected regions and phylogenetic tree of MPA2 family proteins. In (a) and (b) the multiple alignments of two regions of high sequence similarity are presented. The underlined regions are those which exhibit greatest sequence similarity. The underlined region in (a) was used to derive a signature sequence for the MPA2 family. In (c) the phylogenetic tree for the MPA2 family was prepared as described in the legend to Fig. 2.

(a)

(b)

Fig. 4. (a) Schematic depiction of the organization of MPA1 family proteins. The top four proteins are from Gram-negative bacteria while the bottom three pairs of proteins are from Gram-positive bacteria. The latter proteins are each 'split' into two polypeptide chains (MPA1 and C) encoded by two distinct genes. The rectangular boxes represent the putative TMSs while the ovals represent the ATP-binding motifs in the C proteins/domains. (b) Phylogenetic tree for the MPA1 family. The Gram-positive MPA1 and C proteins were artificially fused to generate the tree which was constructed as outlined in the legend to Fig. 2.

noted that both trees show that the Bex and Ctr proteins are closely allied. However, the OMA tree reveals that KpsD is far more distant from the other related members of the family than is VexA. The reverse is true for the Kps and Vex members of the MPA2 family.

The MPA1 family

The MPA1 family contains nine members, six from Gram-negative bacteria and three from Gram-positive bacteria. Most of these proteins have been shown to be essential for production of EPSs (Guidolin *et al.*, 1994; Lin *et al.*, 1994; Bugert & Geider, 1995; Huang & Schell, 1995), and one of these, ExoP, has been experimentally implicated in membrane translocation of the polysaccharide (Long *et al.*, 1988; Reuber *et al.*, 1991; Reuber & Walker, 1993). The topological similarities between the MPA1 and MPA2 families (Fig. 1), and the close genetic association between MPA1 protein-encoding genes and genes encoding OMA and/or PST proteins (Table 4), supports the notion that they are involved in polysaccharide export.

The comparative linear structures of seven MPA1 family members are depicted schematically in Fig. 4(a). These proposed structures are based on the sequence comparisons and hydropathy analyses described below. The
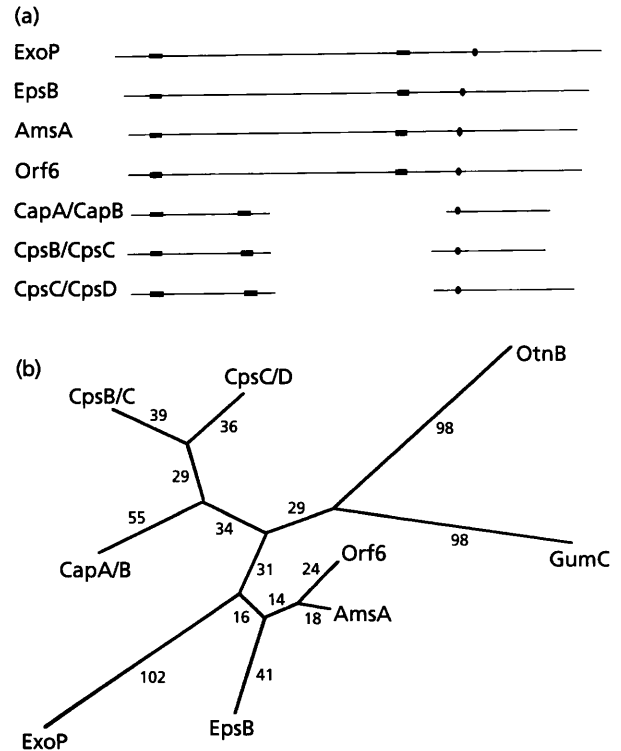
two proteins not shown, OtnB and GumC, appear to lack the C-terminal regions that are characteristic of the other MPA1 family members. The four Gram-negative bacterial proteins shown at the top of Fig. 4(a) are similar in length and topology, and except for the C-terminal tails in the MPA1 proteins, they resemble the MPA2 proteins with respect to topology and size, although they are not demonstrably homologous with them. Each of these MPA1 proteins possesses a short hydrophilic N terminus followed by a TMS, a large periplasmic loop of about 400 residues, a second TMS and finally a C-terminal tail of about 200–250 residues (C) that bears an ATP-binding motif. In the case of the EpsB and ExoP proteins, analysis of alkaline phosphatase fusions has supported the notion that their central 400-residue loops are located periplasmically (Reuber *et al.*, 1991; Glucksmann *et al.*, 1993; Huang & Schell, 1995). By contrast, each of the Gram-positive bacterial MPA1 + C family members consists of two polypeptide chains. The integral-membrane constituent in each of these protein pairs bears the two TMSs found in the Gram-negative bacterial proteins and an extra-
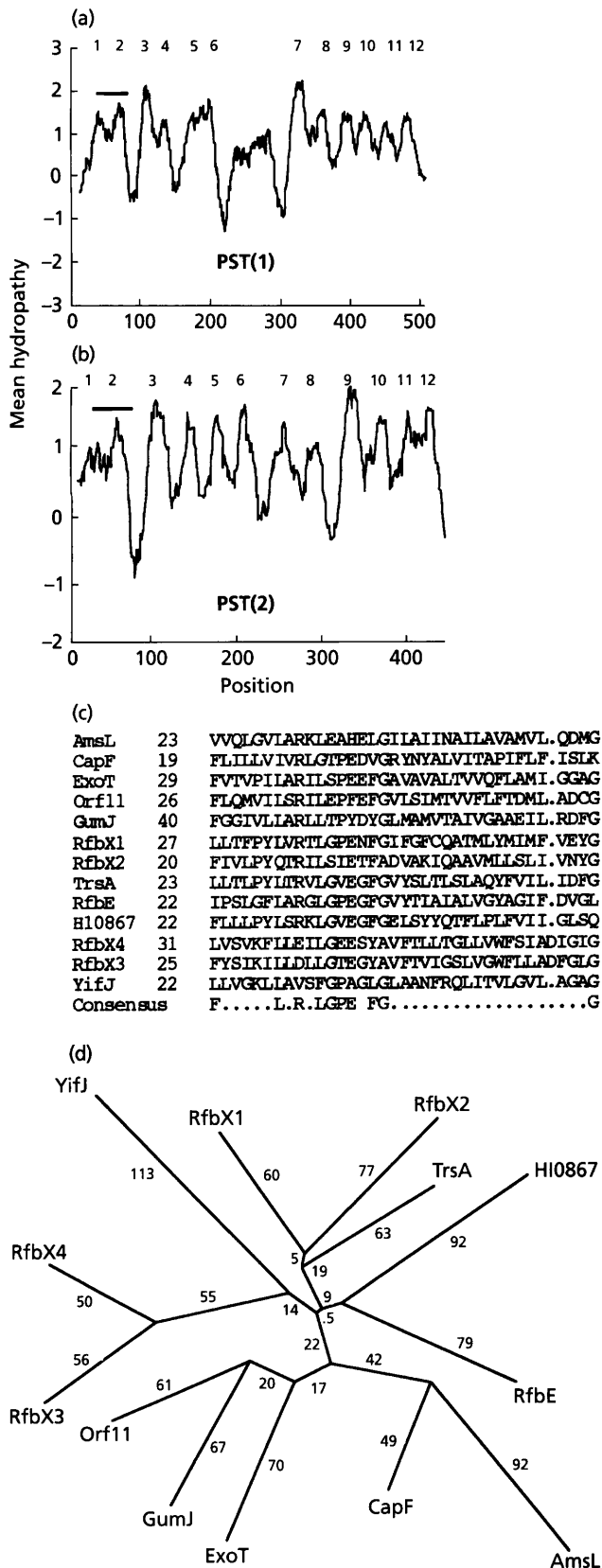
cytoplasmic loop that is much shorter than those found in the Gram-negative bacterial MPA1 proteins. This loop is about 100–130 residues in the Gram-positive bacterial proteins. The soluble protein, C, corresponding to the C-terminal domain of a Gram-negative bacterial member of the MPA1 family (MPA1-C), bears the ATP-binding motif. These membrane and cytoplasmic Gram-positive bacterial constituents are designated MPA1 and C, respectively, in Fig. 1(c).

Mean hydropathy and mean similarity plots were derived for the MPA1 family members (data not shown). Two peaks of hydrophobicity were prominent, and the intervening regions, the extracytoplasmic loop regions, proved to be strongly hydrophilic. The C-terminal tails bearing the ATP-binding motif in the Gram-negative bacterial proteins proved to exhibit two moderately hydrophobic peaks. The first of these two peaks just follows the ATP-binding motif while the second occurs near the C termini of the represented proteins.

The MPA1 proteins are divergent in sequence with comparable degrees of sequence similarity throughout their lengths. They exhibit three regions of maximal similarity, but none of these were found to exhibit sufficient sequence similarity to allow derivation of a generalized signature sequence for the entire MPA1 family. The third region, bearing the ATP-binding Walker motif A, was found to share sequence similarity with a wide range of ATP-hydrolysing enzymes that are not demonstrably homologous to the MPA1 family proteins.

The phylogenetic tree for the MPA1 family members is shown in Fig. 4(b). The three Gram-positive bacterial members of the MPA1 family (CapA/CapB, CpsB/CpsC and CpsC/CpsD) cluster together, the four full-length Gram-negative bacterial members (ExoP, EpsB, AmsA and Orf6) also cluster together and the two remaining members that appear to lack the C-terminal tails (OtnB and GumC) are the most divergent members of the MPA1 family.

In separate analyses (data not shown), the Gram-negative bacterial proteins were artificially spliced and divided into two segments corresponding to regions exhibiting sequence similarity with the two polypeptide chains (the MPA1- and C-domains) of the Gram-positive bacterial members. The trees resulting from analyses of each of these segments resembled those shown in Fig.

proteins in the two subfamilies. The bars above the plots correspond to the region of the multiple alignment shown in (c). 'Position' refers to the alignment position in the multiple alignment and does not correspond to the residue number in any one member of the family. The hydrophobicity scale of Kyte & Doolittle (1982) was used with a sliding window of 20 residues. (c) Alignment of the N-terminal segments depicted by the bars in (a) and (b) of members of the PST family. (d) Phylogenetic tree of the proteins of the PST family. The formats of presentation in (c) and (d) are as described in the legend to Fig. 2 except that residues at any one position conserved in greater than 50 % of the proteins are indicated at the bottom of the alignment shown in (c) (Consensus).

Fig. 5. Characteristics of the PST family proteins. Mean hydropathy plots for the PST(1) and PST(2) subfamilies of the PST family are shown in (a) and (b), respectively. The numbers above the hydropathy plots indicate the putative TMSs of the

4(b), with the exception that GumC and OtnB were absent from the tree derived from the C-domains and GumC clustered with ExoP rather than OtnB in the tree derived from the MPA1 domains. Statistical analyses indicated that the clustering pattern shown in Fig. 4(b) for GumC is artifactual due to the lack of the C-terminal domains in both GumC and OtnB. Otherwise, the clustering pattern shown appears to reflect the phylogenies of the organisms from which these proteins were derived rather than the deletions/insertions that have occurred during evolution of this protein family.

Cytoplasmic-membrane transporters of the PST superfamily

As summarized in Tables 4 and 5, each Gram-negative bacterial OMA–MPA1 pair and each Gram-positive bacterial MPA1 family member probably functions together with a PST protein. Potential exceptions to this generalization can be observed in Table 4. In the cases of OtnA/OtnB from *Vibrio cholerae*, CpsB/CpsC from *Streptococcus agalactiae* and CpsC/CpsD from *Streptococcus pneumoniae* where cytoplasmic transport systems have not been identified, the gene clusters have not been fully sequenced. In the case of EpsA/EpsB, two integral-membrane proteins (EpsE/EpsF) are encoded within the same gene cluster. These proteins both possess 12 putative TMSs and do not share sequence similarity with each other or any protein in the current databases. As the *eps* gene cluster may not have been fully sequenced, it is not clear whether one, both or neither of these two proteins functions in cytoplasmic-membrane transport.

Table 2 lists 14 members of the PST family which all exhibit comparison scores in excess of 10 SD with each other. Five of these proteins are involved in CPS or EPS production, and in the cases of ExoT (Long *et al.*, 1988; Reuber & Walker, 1993) and GumJ (Betlach *et al.*, 1987; Capage *et al.*, 1987), evidence is available suggesting that these proteins are involved in polysaccharide transport. The PST family also includes six proteins associated with O-antigen production in Gram-negative bacteria. These proteins have been implicated in LPS transport (Klena & Schnaitman, 1993; Macpherson *et al.*, 1995; Morona *et al.*, 1995; Liu *et al.*, 1996). Three PST family proteins of unknown function have been identified by genome sequencing projects.

Mean hydropathy analyses indicated that the PST family includes proteins exhibiting two distinct topologies (Figs 5a and b). Both exhibit 12 putative TMSs with N and C termini predicted to be localized to the cytoplasm. PST(1) subfamily proteins, which correspond to the proteins that function with OMA and MPA1 proteins and are believed to transport CPSs or EPSs, possess cytoplasmic loops of substantial size (~ 80 residues) between TMS6 and TMS7. By contrast, PST(2) subfamily proteins, which do not apparently function with an OMA or MPA1 family member and transport LPSs, do not possess an appreciable cytoplasmic loop (loop size of about 15 residues).

Mean similarity plots for the PST(1) and PST(2) subfamilies of the PST family (data not shown) revealed that regions of greatest sequence similarity are found in the N-terminal regions encompassing TMSs 1 and 2. The multiple alignment of this region for 13 of the 14 members of the PST family is shown in Fig. 5(c). The consensus sequence below the alignment indicates the regions of greatest similarity. The sequences are highly divergent, and no family-specific signature sequence could be derived.

The phylogenetic tree shown in Fig. 5(d) reveals that the transporters specific for CPSs cluster loosely together [PST(1) family; bottom] as do the proteins concerned with LPS transport (top). Proteins of unknown function cluster with the LPS transporters. This last fact suggests that the proteins of unknown function are concerned with O-antigen transport. The overall topology of the tree shown in Fig. 5(d) suggests that phylogenetic grouping correlates with specificity.

Cytoplasmic-membrane transporters of the ABC superfamily

As summarized in Table 4, each OMA–MPA2 protein pair appears to function in conjunction with an ABC-2-type transporter, a member of a discrete subfamily of the ABC superfamily (Reizer *et al.*, 1992; Tam & Saier, 1993). Transporters of the ABC superfamily can function in the uptake or expulsion of either small solutes or macromolecules via ATP-hydrolysis-energized processes (Higgins, 1992; Kuan *et al.*, 1995). These systems consist of from one to four polypeptide chains, depending on the system.

The ABC-2 systems analysed here all possess two dissimilar subunits, one being a six-TMS integral-membrane protein, the other bearing the ATP-binding motifs as reported for the limited number of these transport systems analysed in a previous report (Table 3; Reizer *et al.*, 1992). They are concerned with the transport of LPSs, TCAs, LOSs and drugs (Table 2). Multiple alignments and hydropathy plots for some of these proteins have been presented elsewhere (Reizer *et al.*, 1992; Manning *et al.*, 1995).

Phylogenetic trees of the ABC and integral-membrane protein constituents of the ABC-2 family were derived, and the latter is presented in Fig. 6. Comparison of the two trees revealed that the clustering patterns are strikingly similar, with proteins exhibiting closest phylogenetic relationships serving similar functions, regardless of the organism of origin. Thus, the proteins transporting CPSs cluster together (cluster A, Fig. 6), those transporting O-antigen of LPSs cluster together (cluster B) and the single TCA transporter is represented on a deep branch off the LPS transporter cluster.

Nod proteins concerned with LOS transport similarly cluster together (cluster C) and all drug transporters (cluster D) cluster together. The ORFs represented with no known function cluster loosely with the LOS/drug transporters. In general, branch lengths in the tree
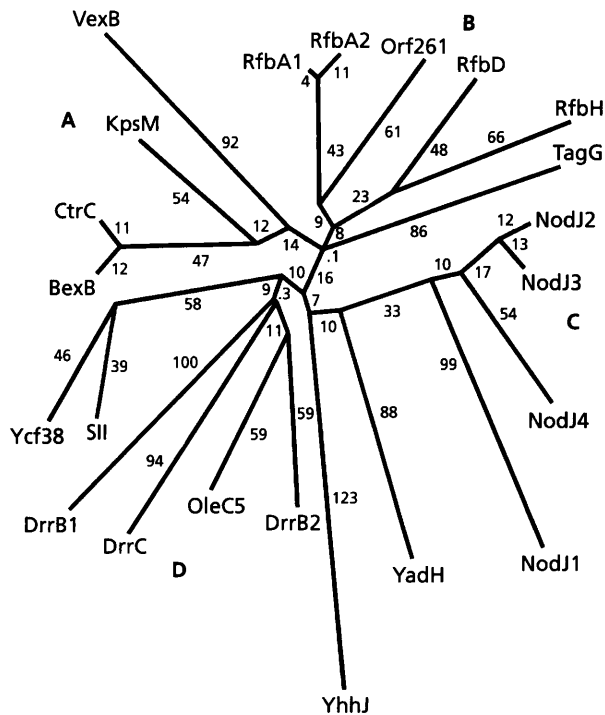
Fig. 6. Phylogenetic tree of the integral-membrane constituents of sequenced ABC-2 family proteins. Protein abbreviations are as indicated in Table 3 and format of presentation is as described in the legend to Fig. 2.

shown in Fig. 6 are much longer than those for the tree derived from the sequences of the ABC proteins, indicating that the membrane proteins have diverged more rapidly from their common origin than have the ABC proteins. This fact is in agreement with a similar conclusion reached for the protein constituents of ABC transporters concerned with the uptake of small molecules into bacteria (Kuan et al., 1995).

DISCUSSION

In this report we have analysed the multicomponent transport systems concerned with the export of complex carbohydrates in bacteria. These systems fall into six export system types as summarized in Table 5. Three of

these types probably utilize PST permeases to translocate their substrates across the cytoplasmic membrane while the other three use ABC-2-type permease systems. Furthermore, three types of systems require auxiliary proteins while three do not. All of the systems that use auxiliary proteins transport CPSs and EPSs, while those that do not transport LPSs, TCAs, LOSs and drugs. The former substrates are generally of higher molecular mass than the latter substrates, and all of the former substrates, but not all of the latter substrates, are exported across the entire cell envelope.

As outlined in Table 5, type 1 systems, found in Gram-negative bacteria, use a PST(1) cytoplasmic-membrane transporter and function with an MPA1-C and an OMA protein to export CPSs or EPSs. Type 2 systems, found in Gram-positive bacteria, use a PST(1) system that functions in conjunction with an MPA1+C protein pair to transport CPSs. Type 3 systems, found in Gram-negative bacteria, use a PST(2) system without the co-operation of a known auxiliary protein to export LPSs.

Type 4 systems, found in Gram-negative bacteria, use an ABC-2 system that functions in conjunction with an MPA2 and an OMA protein to export CPSs. Types 5 and 6, found in Gram-positive and Gram-negative bacteria, respectively, do not utilize known auxiliary proteins and export drugs or TCAs, and LPSs or LOSs, respectively.

As noted above, each of the PST systems that transports CPSs or EPSs in a Gram-negative bacterium functions with an MPA1-C protein, consisting of a single polypeptide chain as well as an OMA protein. By contrast, each PST system specific for CPSs in a Gram-positive bacterium functions together with the structural equivalent of a Gram-negative bacterial MPA1-C protein. However, these homologous Gram-positive bacterial proteins are 'split' proteins encoded by two genes rather than one gene. We have termed these protein pairs MPA1+C. Because the cytoplasmic domain of a Gram-negative bacterial MPA1-C protein or the corresponding C protein of a Gram-positive bacterial MPA1+C protein pair possesses an ATP-binding domain, we propose that this domain may energize polysaccharide export by ATP hydrolysis. This suggestion is reinforced by the fact that MPA2 proteins, which

Table 5. Organization of multicomponent polysaccharide transport systems in bacteria

Type	Bacterial source	Substrate	Cytoplasmic-membrane protein	Auxiliary protein	
				MPA type	OMA type
1	Gram-negative	CPS, EPS	PST(1)	MPA1-C	OMA
2	Gram-positive	CPS	PST(1)	MPA1 + C	–
3	Gram-negative	LPS	PST(2)	–	–
4	Gram-negative	CPS	ABC-2	MPA2	OMA
5	Gram-positive	TCA, drugs	ABC-2	–	–
6	Gram-negative	LPS, LOS	ABC-2	–	–

function in conjunction with an ABC-2-type system for polysaccharide export in Gram-negative bacteria, do not possess an ATP-binding cytoplasmic domain. ABC-2 systems possess their own ABC proteins/domains and are known to energize transport by ATP hydrolysis (type 4 export systems in Table 5) (Berger & Heppel, 1974; Higgins, 1992). Interestingly, PST(2) LPS exporters which do not function with an MPA or an OMA protein apparently lack the C protein/domain. We therefore suggest that LPS export via a PST(2) system uses the proton motive force to energize export. Export systems that can use either ATP or proton motive force to energize transport, depending on the presence of an ATP-binding cytoplasmic protein, have been documented for arsenite export in bacteria (Dey & Rosen, 1995).

It is interesting to note that the cytoplasmic loop domain localized between TMSs 6 and 7 in PST(1) proteins of both Gram-negative and Gram-positive bacteria (types 1 and 2 in Table 5) is lacking in the PST(2) proteins. The latter proteins, in contrast with the former proteins, apparently do not function in conjunction with a C protein/domain. We therefore propose that this central cytoplasmic loop domain in a PST(1) protein functions in the binding of the C domain or protein. In this regard, it is interesting to note that the corresponding central loop of the lactose permease of *E. coli* is believed to function in the allosteric regulation of lactose uptake (Wilson *et al.*, 1990; Hoischen *et al.*, 1996). A role of the central loop of PST(1) proteins as well as the C protein/domain in regulation of PST-mediated polysaccharide export rather than in energization of transport cannot be ruled out.

In Gram-negative bacteria (but not in Gram-positive bacteria) an OMA protein is required for CPS or EPS export. This fact substantiates the suggestion that the OMA proteins are outer-membrane structures that somehow facilitate carbohydrate export across the outer membrane. Interestingly, the system-specific auxiliary proteins (MPA and OMA) are apparently not required for LPS export mediated by either a PST or an ABC-type permease. We therefore propose that the OMA proteins function as outer-membrane 'channel' proteins that allow transport across this bilayer structure and that the periplasmic loops of the MPA proteins interact with the system-specific OMA proteins while the transmembrane regions of the MPA proteins associate with the corresponding PST or ABC transporters. A physical association of the constituents of the transport protein complex in the inner- and outer-membranes is therefore proposed. Whether or not these complexes constitute the Bayer's patches seen by electron microscopy (Bayer, 1968; Mulford & Osborn, 1983; Schnaitman & Klena, 1993) is not known. It is interesting to note that a comparable complex of inner- and outer-membrane transport proteins formed by the so-called 'membrane fusion proteins' which span the periplasm has been proposed (Dinh *et al.*, 1994; Dong & Mergeay, 1994; Saier *et al.*, 1994). These latter proteins are concerned with protein, peptide, metal ion and drug export.

Published experimental evidence, particularly with the Kps and Ctr systems, supports the proposition that auxiliary proteins of the OMA and MPA2 families function in conjunction with an ABC-2 permease (type 4 export systems in Table 5) to facilitate export of CPSs across the two membranes of the Gram-negative bacterial envelope (Frosch *et al.*, 1989, 1991, 1992; Pavelka *et al.*, 1991; Bronner *et al.*, 1993a,b). PST(2) proteins (type 3 export systems in Table 5) (Macpherson *et al.*, 1995; Morona *et al.*, 1995; Liu *et al.*, 1996) have similarly been implicated in LPS export. Furthermore, direct evidence suggests that PST, MPA1 and OMA proteins function in polysaccharide export (types 1 and 2 systems in Table 5) (Long *et al.*, 1988; Reuber *et al.*, 1991; Reuber & Walker, 1993).

The phylogenetic analyses reported here substantiate the suggestion that the proteins of the PST(1), MPA1-C and OMA families function together as polysaccharide exporters (type 1 systems in Table 5) as do the ABC-2, MPA2 and OMA proteins (type 4 systems in Table 5) in Gram-negative bacteria. Similarly, we propose that the PST(1) and MPA1 + C proteins function together as CPS exporters (type 2 systems in Table 5) in Gram-positive bacteria. Our conclusions are based (1) on the occurrence of homologous OMA family members encoded within the gene clusters of type 1 and type 4 systems, (2) the similar topologies of the non-homologous proteins composing the MPA1 and MPA2 families, (3) the presence of cytoplasmic ATP-binding domains or proteins in types 1 and 2 PST-dependent systems, but not in type 4 ABC-2-dependent systems, and (4) the occurrence of either an MPA1 or an MPA2 protein (but not both) in each gene cluster encoding both an OMA protein and a cytoplasmic-membrane exporter.

Our observations lead us to suggest that the PST or ABC-2 systems provide an energy-coupled cytoplasmic transmembrane export function, that an OMA protein provides a porin-like structure in the outer membrane of Gram-negative bacteria, and that an MPA1 or MPA2 protein serves to connect these two transmembrane channels physically and possibly also to facilitate passage through the cell wall. In the case of Gram-positive bacteria, MPA1 proteins may serve only in the latter capacity. However, as MPA1 or MPA2 proteins are always present when CPSs or EPSs are transported, a role in organization and/or function of cytoplasmic-membrane transporters cannot be ruled out.

We have noted an interesting correlation between the type of cytoplasmic-membrane transporter utilized (i.e. PST or ABC-2) with the size and group of the polysaccharides transported. Thus, all ABC-2 transporters that export CPSs (type 4 in Table 5) transport relatively low molecular mass CPSs, while all PST(1) transporters (types 1 and 2 in Table 5) apparently export high molecular mass polysaccharides (our unpublished observations). In this connection it is interesting to note that other ABC-2 type systems (types 5 and 6 in Table 5) also transport lower molecular mass substrates. Size of the substrate transported may also provide an explanation for the fact that an ATP-binding protein/domain

is required for type 1 and type 2 PST-dependent systems but not apparently for type 3 PST-dependent systems. The latter systems transport LPSs rather than EPSs.

The molecular bases for many of the observations and postulates presented here remain to be ascertained. A wealth of confirmatory experiments will now be required to establish the significance of the observations reported. Such experiments should reveal the fundamental mechanistic similarities and differences between the two types of cytoplasmic-membrane transporters and their auxiliary constituents concerned with complex carbohydrate export in bacteria.

ACKNOWLEDGEMENTS

REFERENCES

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, T. L. & Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, pp. 28–36. Menlo Park, CA: AAAI Press.

Bairoch, A. (1992). PROSITE: a dictionary of sites and patterns in proteins. *Nucleic Acids Res* 20, 2013–2018.

Bartlett, D. H., Wright, M. E. & Silverman, M. (1988). Variable expression of extracellular polysaccharide in the marine bacterium *Pseudomonas atlantica* is controlled by genome rearrangement. *Proc Natl Acad Sci USA* 85, 3923–3927.

Bayer, M. E. (1968). Areas of adhesion between wall and membrane of *Escherichia coli*. *J Gen Microbiol* 53, 395–404.

Berger, E. A. & Heppel, L. A. (1974). Different mechanisms of energy coupling for the shock-sensitive and shock-resistant amino acid permeases of *Escherichia coli*. *J Biol Chem* 249, 7747–7755.

Betlach, M. R., Capage, M. A., Doherty, D. H., Hassler, R. A., Henderson, N. M., van der Slice, R. W., Marrelli, J. D. & Ward, M. B. (1987). Genetically engineered polymers: manipulation of xanthan biosynthesis. In *Industrial Polysaccharides: Genetic Engineering, Structure/Property Relations and Applications*, pp. 35–50. Edited by M. Yalpani. Amsterdam: Elsevier Science Publishers.

Bik, E. M., Bunschoten, A. E., Gouw, R. D. & Mooi, F. R. (1995). Genesis of the novel epidemic *Vibrio cholerae* O139 strain: evidence for horizontal transfer of genes involved in polysaccharide synthesis. *EMBO J* 14, 209–216.

Boulnois, G. J. & Jann, K. (1989). Bacterial polysaccharide capsule synthesis, export and evolution of structural diversity. *Mol Microbiol* 3, 1819–1823.

Boulnois, G. J. & Roberts, I. S. (1990). Genetics of capsular polysaccharide production in bacteria. *Curr Top Microbiol Immunol* 150, 1–20.

Bronner, D., Sieberth, V., Pazzani, C., Roberts, I. S., Boulnois, G. J., Jann, B. & Jann, K. (1993a). Expression of the capsular K5 polysaccharide of *Escherichia coli*: biochemical and electron microscopic analyses of mutants with defects in region 1 of the K5 gene cluster. *J Bacteriol* 175, 5984–5992.

Bronner, D., Sieberth, V., Pazzani, C., Smith, A., Boulnois, G., Roberts, I., Jann, B. & Jann, K. (1993b). Synthesis of the K5 (group II) capsular polysaccharide in transport-deficient recombinant *Escherichia coli*. *FEMS Microbiol Lett* 113, 279–284.

Bugert, P. & Geider, K. (1995). Molecular analysis of the *ams* operon required for exopolysaccharide synthesis of *Erwinia amylovora*. *Mol Microbiol* 15, 917–933.

Capage, M. A., Doherty, D. H., Betlach, M. R. & van der Slice, R. (1987). *Recombinant-DNA mediated production of xanthan gum*. Patent International Publication No. WO87/05938.

Chou, P. Y. & Fasman, G. D. (1978). Empirical prediction of protein conformation. *Annu Rev Biochem* 47, 251–276.

Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983). Establishing homologies in protein sequences. *Methods Enzymol* 91, 524–545.

Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* 12, 387–395.

Dey, S. & Rosen, B. P. (1995). Dual mode of energy coupling by the oxyanion-translocating ArsB protein. *J Bacteriol* 177, 385–389.

Dinh, T., Paulsen, I. T. & Saier, M. H., Jr (1994). A family of extracytoplasmic proteins that allow transport of large molecules across the outer membranes of Gram-negative bacteria. *J Bacteriol* 176, 3825–3831.

Dong, Q. & Mergeay, M. (1994). Czc/Cnr efflux: a three-component chemiosmotic antiport pathway with a 12-transmembrane-helix protein. *Mol Microbiol* 14, 185–187.

Dorman, C. J. (1995). DNA topology and the global control of bacterial gene expression: implications for the regulation of virulence gene expression. *Microbiology* 141, 1271–1280.

Dybvig, K. (1993). DNA rearrangements and phenotypic switching in prokaryotes. *Mol Microbiol* 10, 465–471.

Fath, M. J. & Kolter, R. (1993). ABC transporters: bacterial exporters. *Microbiol Rev* 57, 995–1017.

Feng, D.-F. & Doolittle, R. F. (1990). Progressive alignment and phylogenetic tree construction of protein sequences. *Methods Enzymol* 183, 375–387.

Frosch, M., Weisgerber, C. & Meyer, T. F. (1989). Molecular characterization and expression in *Escherichia coli* of the gene complex encoding the polysaccharide capsule of *Neisseria meningitidis* group B. *Proc Natl Acad Sci USA* 86, 1669–1673.

Frosch, M., Edwards, U., Bousset, K., Krauße, B. & Weisgerber, C. (1991). Evidence for a common molecular origin of the capsule gene loci in Gram-negative bacteria expressing group II capsular polysaccharides. *Mol Microbiol* 5, 1251–1263.

Frosch, M., Müller, D., Bousset, K. & Müller, A. (1992). Conserved outer membrane protein of *Neisseria meningitidis* involved in capsule expression. *Infect Immun* 60, 798–803.

Garnier, J., Osguthorpe, D. J. & Robson, B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J Mol Biol* 120, 97–120.

Glucksmann, M. A., Reuber, T. L. & Walker, G. C. (1993). Genes needed for the modification, polymerization, export, and processing of succinoglycan by *Rhizobium meliloti*: a model for succinoglycan biosynthesis. *J Bacteriol* 175, 7045–7055.

Guidolin, A., Morona, J. K., Morona, R., Hansman, D. & Paton, J. C. (1994). Nucleotide sequence analysis of genes essential for

capsular polysaccharide biosynthesis in *Streptococcus pneumoniae* Type 19F. *Infect Immun* 62, 5384–5396.

Hashimoto, Y., Li, N., Yokoyama, H. & Ezaki, T. (1993). Complete nucleotide sequence and molecular characterization of ViaB region encoding Vi antigen in *Salmonella typhi*. *J Bacteriol* 175, 4456–4465.

von Heijne, G. (1992). Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J Mol Biol* 225, 487–494.

Higgins, C. F. (1992). ABC transporters: from microorganisms to man. *Annu Rev Cell Biol* 8, 67–113.

Hoischen, C., Pitaknarongphorn, S., Reizer, J., Levin, J., Tomich, J. M. & Saier, M. H., Jr (1996). Involvement of the central loop of the lactose permease of *Escherichia coli* in its allosteric regulation by the glucose-specific Enzyme IIA of the phosphoenolpyruvate-dependent phosphotransferase system. *J Bacteriol* 178, 6082–6086.

Hornick, R. B., Greisman, S. E., Woodward, T. E., DuPont, H. L., Dawkins, A. T. & Snyder, M. J. (1970). Typhoid fever: pathogenesis and immunologic control. *N Engl J Med* 283, 739–746.

Huang, J. & Schell, M. (1995). Molecular characterization of the *eps* gene cluster of *Pseudomonas solanacearum* and its transcriptional regulation at a single promoter. *Mol Microbiol* 16, 977–989.

Jann, B. & Jann, K. (1990). Structure and biosynthesis of the capsular antigens of *Escherichia coli*. *Curr Top Microbiol Immunol* 150, 19–42.

Jann, K. & Jann, B. (1984). Structure and biosynthesis of O-antigens. In *Handbook of Endotoxin, Vol. 1. Chemistry of Endotoxin*, pp. 138–186. Edited by E. T. Rietsche. Amsterdam: Elsevier Science Publishers.

Jann, K., Pillat, M., Weisgerber, C., Shibaev, V. N. & Torgov, V. I. (1985). Biosynthesis of the O9 antigen of *Escherichia coli*. Synthetic glycosyldiphosphomoraprenols as probes for requirement of mannose acceptors. *Eur J Biochem* 151, 393–397.

Kido, N., Torgov, V. I., Sugiyama, T., Uchiya, K., Sugihara, H., Komatsu, T., Kato, N. & Jann, K. (1995). Expression of the O9 polysaccharide of *Escherichia coli*: sequencing of the *E. coli* O9 *rfb* gene cluster, characterization of mannosyl transferases, and evidence for an ATP-binding cassette transport system. *J Bacteriol* 177, 2178–2187.

Klena, J. D. & Schnaitman, C. A. (1993). Function of the *rfb* gene cluster and the *rfe* gene in the synthesis of O antigen by *Shigella dysenteriae* 1. *Mol Microbiol* 9, 393–402.

Kroll, J. S., Loynds, B., Brophy, L. M. & Moxon, E. R. (1990). The *bex* locus in encapsulated *Haemophilus influenzae*: a chromosomal region involved in capsule polysaccharide export. *Mol Microbiol* 4, 1853–1862.

Kröncke, K.-D., Boulnois, G., Roberts, I., Bitter-Suermann, D., Golecki, J. R., Jann, B. & Jann, K. (1990). Expression of the *Escherichia coli* K5 capsular antigen: immunoelectron microscopic and biochemical studies with recombinant *E. coli*. *J Bacteriol* 172, 1085–1091.

Kuan, G., Dassa, E., Saurin, W., Hofnung, M. & Saier, M. H., Jr (1995). Phylogenetic analyses of the ATP-binding constituents of bacterial extracytoplasmic receptor-dependent ABC-type nutrient uptake permeases. *Res Microbiol* 146, 271–278.

Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157, 105–132.

Lazarevic, V. & Karamata, D. (1995). The *tagGH* operon of *Bacillus subtilis* 168 encodes a two-component ABC transporter involved in the metabolism of two wall teichoic acids. *Mol Microbiol* 16, 345–355.

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 function for Wzx (RfbX): a promising candidate for O-unit flippase. *J Bacteriol* 178, 2102–2107.

Long, S., Reed, J. W., Himawan, J. & Walker, G. C. (1988). Genetic analysis of a cluster of genes required for synthesis of the calcofluor-binding exopolysaccharide of *Rhizobium meliloti*. *J Bacteriol* 170, 4239–4248.

McGrath, B. C. & Osborn, M. J. (1991). Evidence for energy-dependent transposition of core lipopolysaccharide across the inner membrane of *Salmonella typhimurium*. *J Bacteriol* 173, 3134–3137.

Macpherson, D. F., Manning, P. A. & Morona, R. (1995). Genetic analysis of the *rfbX* gene of *Shigella flexneri*. *Gene* 155, 9–17.

Manning, P. A., Stroeher, U. H., Karageorgos, L. E. & Morona, R. (1995). Putative O-antigen transport genes within the *rfb* region of *Vibrio cholerae* O1 are homologous to those for capsule transport. *Gene* 158, 1–7.

Morona, R., Macpherson, D. F., van den Bosch, L., Carlin, N. I. A. & Manning, P. A. (1995). Lipopolysaccharide with an altered O-antigen produced in *Escherichia coli* K-12 harbouring mutated, cloned *Shigella flexneri rfb* genes. *Mol Microbiol* 18, 209–223.

Mulford, C. A. & Osborn, M. J. (1983). An intermediate step in translocation of lipopolysaccharide to the outer membrane of *Salmonella typhimurium*. *Proc Natl Acad Sci USA* 80, 1159–1163.

Pavelka, M. S., Jr, Wright, L. F. & Silver, R. P. (1991). Identification of two genes, *kpsM* and *kpsT*, in region 3 of the polysialic acid gene cluster of *Escherichia coli* K12. *J Bacteriol* 173, 4603–4610.

Pearce, R. & Roberts, I. S. (1995). Cloning and analysis of gene clusters for production of the *Escherichia coli* K10 and K54 antigens: identification of a new group of *serA*-linked capsule gene clusters. *J Bacteriol* 177, 3992–3997.

Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* 85, 2444–2448.

Reeves, P. R. (1992). Variation in O-antigens, niche-specific selection and bacterial populations. *FEMS Microbiol Lett* 100, 509–516.

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

Reizer, J., Reizer, A. & Saier, M. H., Jr (1992). A new subfamily of bacterial ABC-type transport systems catalyzing export of drugs and carbohydrates. *Protein Sci* 1, 1326–1332.

Reuber, T. L. & Walker, G. C. (1993). Biosynthesis of succino-glycan, a symbiotically important exopolysaccharide of *Rhizobium meliloti*. *Cell* 74, 269–280.

Reuber, T. L., Long, S. & Walker, G. C. (1991). Regulation of *Rhizobium meliloti exo* genes in free-living cells and *in planta* examined by using Tn*phoA* fusions. *J Bacteriol* 173, 426–434.

Robbins, J. D. & Robbins, J. B. (1984). Re-examination of the protective role of the capsular polysaccharide (Vi antigen) of *Salmonella typhi*. *J Infect Dis* 150, 436–449.

Roberts, I. S. (1995). Bacterial polysaccharides in sickness and in health. *Microbiology* 141, 2023–2031.

Rosenow, C., Esumeh, F., Roberts, I. S. & Jann, K. (1995). Characterization and localization of the KpsE protein of

Escherichia coli K5, which is involved in polysaccharide export. *J Bacteriol* 177, 1137–1143.

Saier, M. H., Jr (1994). Computer-aided analyses of transport protein sequences: gleaning evidence concerning function, structure, biogenesis, and evolution. *Microbiol Rev* 58, 71–93.

Saier, M. H., Jr (1996). Phylogenetic approaches to the identification and characterization of protein families and superfamilies. *Microb Comp Genomics* 1, 129–150.

Saier, M. H., Jr & Jacobson, G. R. (1984). *The Molecular Basis of Sex and Differentiation: a Comparative Study of Evolution, Mechanism and Control in Microorganisms.* New York: Springer.

Saier, M. H., Jr, Tam, R., Reizer, A. & Reizer, J. (1994). Two novel families of bacterial membrane proteins concerned with nodulation, cell division and transport. *Mol Microbiol* 11, 841–847.

Schnaitman, C. A. & Klena, J. D. (1993). Genetics of lipopolysaccharide biosynthesis in enteric bacteria. *Microbiol Rev* 57, 655–682.

Seifert, H. S. & So, M. (1988). Genetic mechanisms of bacterial antigenic variation. *Microbiol Rev* 52, 327–336.

Silver, R. P., Aaronson, W. & Vann, W. F. (1987). Translocation of capsular polysaccharides in pathogenic strains of *Escherichia coli* requires a 60-kilodalton periplasmic protein. *J Bacteriol* 169, 5489–5495.

Smith, A. N., Boulnois, G. J. & Roberts, I. S. (1990). Molecular analysis of the *Escherichia coli* K5 *kps* locus: identification and characterization of an inner-membrane capsular polysaccharide transport system. *Mol Microbiol* 4, 1863–1869.

Tam, R. & Saier, M. H., Jr (1993). Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol Rev* 57, 320–346.

Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. (1982). Distantly related sequences in the α- and β-subunits of ATP synthase myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J* 1, 945–951.

Wilson, T. H., Yunker, P. L. & Hansen, C. L. (1990). Lactose transport of *Escherichia coli* resistant to inhibition by the phosphotransferase system. *Biochim Biophys Acta* 1029, 113–116.