Available online at www.sciencedirect.com
SCIENCE DIRECT®
Carbohydrate Research 338 (2003) 2503–2519
CARBOHYDRATE RESEARCH
www.elsevier.com/locate/carres

Perspective/Review

Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly

Gabrielle Samuel, Peter Reeves*

School of Molecular and Microbial Biosciences, University of Sydney, Sydney, NSW 2006, Australia

Received 5 May 2003; accepted 13 July 2003

Abstract

The O-antigen is an important component of the outer membrane of Gram-negative bacteria. It is a repeat unit polysaccharide and consists of a number of repeats of an oligosaccharide, the O-unit, which generally has between two and six sugar residues. O-Antigens are extremely variable, the variation lying in the nature, order and linkage of the different sugars within the polysaccharide. The genes involved in O-antigen biosynthesis are generally found on the chromosome as an O-antigen gene cluster, and the structural variation of O-antigens is mirrored by genetic variation seen in these clusters. The genes within the cluster fall into three major groups. The first group is involved in nucleotide sugar biosynthesis. These genes are often found together in the cluster and have a high level of identity. The genes coding for a significant number of nucleotide sugar biosynthesis pathways have been identified and these pathways seem to be conserved in different O-antigen clusters and across a wide range of species. The second group, the glycosyl transferases, is involved in sugar transfer. They are often dispersed throughout the cluster and have low levels of similarity. The third group is the O-antigen processing genes. This review is a summary of the current knowledge on these three groups of genes that comprise the O-antigen gene clusters, focusing on the most extensively studied *E. coli* and *S. enterica* gene clusters.

Keywords: Bacterial repeat unit polysaccharide; O-Antigen gene cluster; Nucleotide sugar biosynthesis; Glycosyl transferase; Recombination; *E. coli*; *S. enterica*

Contents

Abbreviations: abequose: Abe; 4-acetamido-4,6-dideoxy-D-mannose: PerNAc; 2-acetamido-2,3-dideoxy-D-mannosaminuronic acid: Man(2NAc3N)A; 2-acetamidogalacturonic acid: GalNAcA; 2-acetamidoquinovose: QuiNAc; 2-acetamidofucose: FucNAc; 3-acetamidofucose: Fuc3NAc; 4-acetamidofucose: Fuc4NAc; 2-acetamidomannuronic acid: ManNAcA; *N*-acetyl-viosamine: VioNAc or DD-Qui4N; 3-amino-3-deoxy-quinovose: D-Qui3N; Ascarylose: Asc; Bacterial repeat-unit polysaccharides: RUPs; Colitose: Col; 6-deoxy-D-*manno*-heptose: D-*man*-6dHep; 6-deoxy-D-talose: D-6dTal; 6-deoxy-L-talose: L-6dTal; insertion sequence: IS; paratose: Par; perosamine: Per; transmembrane: TM; tyvelose: Tyv.

* Corresponding author. Tel.: +61-519-661-3433.
E-mail address: p.reeves@angis.usyd.edu.au (P. Reeves).

0008-6215/03/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2003.07.009

1. Introduction

Bacterial repeat-unit polysaccharides (RUPs) are extremely diverse and occur in multiple forms, with substantial variation within a species. They include capsules, exopolysaccharides and O-antigens. This review will focus on the O-antigens of a few species, but we will also include comments on some other RUPs.

The O-antigen is an important component of the outer membrane of Gram-negative bacteria. It acts as a receptor for bacteriophages and is also important in the host immune response. Like all RUPs, it consists of a number of repeats of an oligosaccharide, in this case the O unit, which generally has between two and six sugar residues.[1] The O-antigen is extremely variable, the variation lying in the nature, order and linkage of the different sugars within the polysaccharide. Indeed, there are 186 and 54 O-antigens documented in the *Escherichia coli* (including *Shigella*) and *Salmonella enterica* typing schemes, respectively,[2–4] and only three of these have identical O-antigen structures in the two species, namely *E. coli* O111/*S. enterica* O35,[5] *E. coli* O55/*S. enterica* O50,[5,6] and *E. coli* O157/*S. enterica* O30.[7] There are also 193 different serogroups of *Vibrio cholerae*, each expressing a distinct O-antigen, and 20 different *Pseudomonas aeruginosa* O-antigen structures have been recognised.[8] Several other species that have been studied have 40 or more serogroups defined by O-antigens, with examples given by Jansson (1999),[9] clearly demonstrating the enormous diversity of O-antigens in Gram-negative bacteria. Nearly all *E. coli* and *S. enterica* O-antigens are heteropolymers. Exceptions include *E. coli* O8,[10] *E. coli* O9,[11] *E. coli* O9a[12] and *S. enterica* O54.[13] In contrast, almost all *Klebsiella pneumoniae* O-antigens are homopolymers. *V. cholerae* O-antigens can be either hetero- or homopolymers, and all *P. aeruginosa* O-antigens are heteropolymers.

In this review, we will focus on *E. coli* and *S. enterica* O-antigen gene clusters as they are the most extensively studied, with some comparisons to those from, for example, *V. cholerae* and *P. aeruginosa*. Further, the organisation of gene clusters is essentially the same in all RUPs, and much of what we say about O-antigens also applies to capsules, secreted RUPs, and also the polysaccharide components of S-layer glycoprotein, and where necessary we will use non-O-antigen examples.

2. O-Antigen gene clusters

The genes involved in O-antigen biosynthesis are generally found on the chromosome as an O-antigen gene cluster. In *E. coli* and *S. enterica,* these gene clusters generally lie between the *galF* and *gnd* genes.[14] In *P. aeruginosa,* the O-antigen gene cluster is situated between *himD* and *tyrB*,[8] in *V. cholerae* between *gmhD* and *rjg*,[15] and in *Yersinia spp*. between *hemH* and *gsk*.[16,17] There are some exceptions to these 'rules'. For example the *E. coli* O8, O9 and O9a gene clusters are between *gnd* and the *his* operon. The *E. coli* O8 and O9 O-antigens are identical to the *K. pneumoniae* O5 and O3 O-antigens, respectively,[18] and the corresponding O-antigen gene clusters are also between *gnd* and the *his* operon, as are other O-antigen gene clusters in *K. pneumoniae*. The atypical chromosomal position of the *E. coli* O8 and O9 O-antigen gene clusters suggests that they may have arisen by lateral transfer from the respective *K. pneumoniae* O-antigen clusters.[19] It is interesting that whereas most *E. coli* O-antigens are heteropolymers, many *K. pneumoniae* O-antigens are homopolymers, including those shared with *E. coli*. A similar situation is seen in *Yersinia enterocolitica* O3. The *Y. enterocolitica* O3 O-antigen is also a homopolymer whose corresponding O-antigen gene cluster maps to a locus different to where the O-antigen gene clusters corresponding to the heteropolymer O-antigens are found.[20] In this case, there is a short oligosaccharide whose synthesis genes are located independently of the gene cluster between *hemH* and *gsk*.[21] Although this oligosaccharide is treated as an outer core component, it maps between *hemH* and *gsk* and is equivalent to a single O-unit of a typical *Y. enterocolitica* O-antigen.

In general O-antigen genes are very close to each other, often overlapping in their reading frames, and are thought to be transcribed as a unit. A 39 bp JUMPStart sequence is present just upstream of a number of O-antigen gene clusters, and within this sequence are two smaller sub-sequences named ops1 and ops2. Although the promoter can normally initiate transcription alone,[22] ops1 and ops2 have been demonstrated to be involved in the recruitment of elongation factors.[23–25] Overall, O-antigen gene clusters will typically have a GC content lower than the genome average. This atypical GC content seen in so many O-antigen gene clusters provides strong evidence that the clusters have been recently transferred, by inter-species lateral transfer, from a different bacterial species. This switching of O-antigen genes between different bacterial species explains, at least in part, how the diversity of O-antigen structures arose in Gram-negative bacteria.

To date, *E. coli* O-antigen gene clusters that have been sequenced (excluding *Shigella*) include O113, O111, O104, O157, O55, O75, O91, O26, O7, O8 and O9a. It is now known that the previously identified species of *Shigella*; *S. dysenteriae, S. boydii, S. sonnei and S. flexneri* are in fact clones of *E. coli*[26] and, in total, seven Shigella O-antigen gene clusters have been sequenced. These include Flexneri 2a, Sonnei, Dysenteriae O1 and Boydii O4, O5, O6 and O9. O-Antigens of *S. enterica* A, B, C2, D1, D2, D3 and E1 have related structures and most genes in their O-antigen gene clusters have also been identified[27–30] (see for example Fig. 1), as has the O-antigen cluster of *S. enterica* O54.[31] In addition, all 20 *P. aeruginosa* O-antigen clusters and four *V. cholerae* O-antigen gene clusters have been sequenced, as well as a number of clusters from other Gram-negative bacteria, such as *Vibrio anguillarum* O1 and *K. pneumoniae* O1, O5 and O3 O-antigen gene clusters.

As one would expect, the structural variation of O-antigens is mirrored by genetic variation in the O-antigen gene clusters. The genes encode proteins that mostly fall into three major groups.[1] The first set of proteins are involved in the biosynthesis of the nucleotide sugar precursors of the O-antigen. The second group, the glycosyl transferase proteins, then sequentially transfer the various precursor sugars to form an oligosaccharide on a carrier lipid, undecaprenyl phosphate (UndP), which is situated in the inner membrane facing the cytoplasmic side.[14] The proteins of the third group, the O-antigen processing proteins, are involved in translocation across the membrane, and polymerisation (Fig. 2). For most O-antigens of *E. coli* and *S. enterica*, once completed, the O-unit on UndPP is translocated (flipped) across the membrane to be exposed on the periplasmic face where it is polymerised to form a long chain O-antigen. However, the scenario is slightly different for the homopolymer *V. cholerae* O1, *K. pneumoniae* O3 and O5, and *E. coli* O8, O9 and O9a homopolymers. These O-antigens are synthesised directly as a polymer within the cytoplasm before a different set of proteins transport the polysaccharide across the membrane.[32] Some O-antigens include non-sugar components, and genes related to their synthesis or transfer may also be present in the O-antigen gene cluster. In addition, remnant genes, insertion sequences (IS) and H-repeats have been identified in some gene clusters.

3. Gene nomenclature

The nomenclature for RUP genes distinguishes between the three major groups discussed above. For the genes of a specific sugar pathway, the three letter component of the Demerec name is used, followed by a 4th letter, which is given in order of function in the pathway. For example, *rmlA*, *rmlB*, *rmlC* and *rmlD* for the four gene dTDP-L-Rha pathway. Such genes are quite common and once the biochemistry and genetics of a pathway is determined, the genes can usually be identified by sequence alone. For other genes, names beginning with 'w' have been used. These are mostly sugar transferase genes, genes for repeat unit processing, or genes of unknown function. Repeat unit processing genes are often conserved across a wide range of gene clusters and have been given names beginning with '*wz*', for example '*wzy*' being used for all O-antigen polymerase genes. This system of bacterial polysaccharide gene nomenclature (BPGN) was introduced in 1996,[1] but other systems of nomenclature are still in use and there are genes with two (or more) names. We refer those interested to our web site: http://www.angis.usyd.edu.au/BacPolGenes/welcome.html for further information on genes, including synonyms, gene clusters and biosynthetic pathways. We maintain a database of gene names allocated and it is most helpful if those planning to apply the system ask us (BPGDadministration@microbio.usyd.edu.au) to reserve names that are then held until the work is released.

4. Nucleotide sugar biosynthesis pathways

By far the majority of the components of O-antigens are sugars or sugar derivatives, and most are transferred from nucleotide sugar precursors. The precursors of several of the sugars commonly found in RUPs are part of other generally present pathways. These include UDP-Glc, UDP-Gal, and UDP-GlcNAc. The genes involved in the synthesis of these nucleotide sugars, and the genes involved in the early steps of a pathway that has a housekeeping function, are not usually duplicated in the O-antigen gene cluster.[14] Other sugars are found only in RUPs and as a given form is present in only some strains, the genes for synthesis of its precursor are only required by such strains. These genes are generally found in the O-antigen or another RUP gene cluster. Each O-antigen gene cluster is thus expected to have nucleotide sugar biosynthesis pathway genes corresponding to each of the non-housekeeping sugars present in their respective O-antigens. The nucleotide sugar biosynthesis pathways seem to be conserved in different RUP gene clusters across a wide range of species. The genes for each such pathway are typically found as a block within the cluster. Many of the pathways include a nucleotide-diphospho-6-deoxy-4-hexulose intermediate, which acts as a substrate for various epimerases and reductases prior to reduction of the 4-keto group to give a nucleotide-diphospho-6-deoxy-sugar.

In the following sections, we will look at the pathways for 30 nucleotide sugars that are present in various O-antigen structures. We will consider such pathways, grouping them according to which NDP is present in the NDP-precursor sugar. Our focus is on O-antigens but the NDP-sugars are also donors for transfer of sugars to other substrates, and as the pathways are the same we will take examples from non-O-antigen contexts if that is where the work has been done.

4.1. dTDP-Sugar biosynthesis pathways

Several 6-deoxyhexose sugars are synthesised from glucose-1-phosphate as dTDP-sugars. These include L-Rha, L-Tal, VioNAc, D-Fuc, D-Fuc4NAc and D-Fuc3-NAc.

The first dTDP pathway to be described was that for dTDP-L-Rha and the four genes were named *rmlABCD* in the BPGN system. Genes *rmlA* and *rmlB* encode a glucose-1-phosphate thymidyl-transferase and a dTDP-D-glucose-4,6-dehydratase, respectively, that act to convert glucose-1-phosphate into the dTDP-6-deoxy-D-*xylo*-4-hexulose intermediate which is the branch point for most of the pathways that follow. There are known structures for both enzymes as well as functional data.[33–36]

4.1.1. L-Rhamnose. L-Rha is widely distributed not only in *E. coli* and *S. enterica* O-antigens, but also in those of other species, and other RUPs. As stated above, synthesis of L-Rha requires RmlC and RmlD in addition to RmlAB. RmlC and RmlD have been well characterised both genetically and biochemically,[37,38] and encode a TDP-6-deoxy-D-glucose-3,5-epimerase and dTDP-6-deoxy-L-mannose dehydrogenase, respectively, which act sequentially on dTDP-6-deoxy-D-*xylo*-4-hexulose to give dTDP-L-Rha. All four L-Rha biosynthesis genes have been reported for many *E. coli* and *S. enterica* O-antigen gene clusters. They have also been identified in, for example, the *V. anguillarum* O1 O-antigen gene cluster and the *Burkholderia mallei* O-antigen gene cluster. In *E. coli* and *S. enterica*, except for *E. coli* Boydii 9 where four extra genes are situated between *rmlA* and *rmlC*,[39] the *rmlABCD* genes are found together at the 5′ end of the O-antigen gene cluster, in the order *rmlBDAC*. They are generally clustered in other species but the order can vary, being for example *rmlBADC* in *Actinobacillus actinomycetemcomitans*.[40]

4.1.2. 6-Deoxy-L-talose (also known as L-pneumose). This sugar has been reported in *E. coli* and *P.*

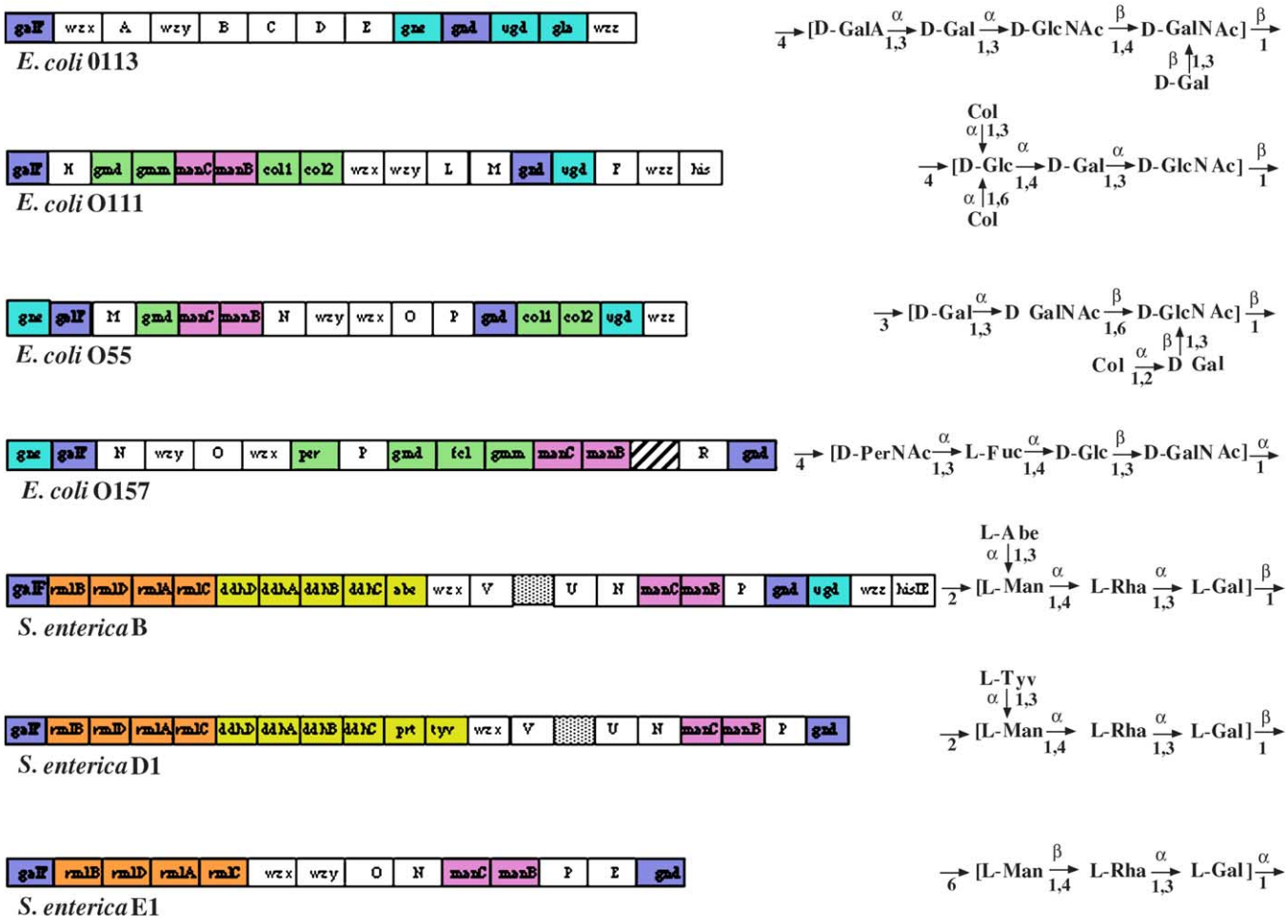

Fig. 1. Representative *E. coli* and *S. enterica* O-antigen gene clusters and their corresponding O-antigen structures. The gene clusters reside between the housekeeping genes *galF* and *gnd* (purple). Genes involved in synthesising UDP-sugar precursors are shown in turquoise. Genes involved in synthesising CDP-sugar precursors are shown in yellow. Genes involved in synthesising dTDP-sugar precursors are shown in orange. Genes involved in synthesising GDP-sugar precursors are shown in pink (for GDP-D-Man synthesis) or green (for synthesis of other GDP-sugars). Remnant *wzy* gene in *S. enterica* B and D1 shown as dotted box. Transferase genes and O-antigen processing genes are not coloured. Transferase genes are indicated by a single letter. See abbreviation list for complete sugar names. Not to scale.

aeruginosa, but the pathway has only been described and the genes identified in *A. actinomycetemcomitans*.[41] dTDP-L-Tal synthesis requires *rmlABC* and *tll*. Tll encodes a C-4 reductase that, like RmlD, acts on the product of *rmlC*, giving dTDP-L-6dTal, the epimer of dTDP-L-Rha.

4.1.3. Viosamine and *N*-acetylviosamine. The *E. coli* O7 structure contains VioNAc and, in addition to *rmlA* and *rmlB*, the *E. coli* O7 O-antigen gene cluster contains *vioA* and *vioB*. VioA is presumed to carry out the transamination of dTDP-6-deoxy-D-*xylo*-4-hexulose to dTDP-4-amino-4,6-dideoxy-D-glucose (VioN), and VioB is thought to subsequently *N*-acetylate VioN to dTDP-VioNAc.[42] VioN also exists in O-antigens and, as dTDP-VioN is an intermediate in the dTDP-VioNAc pathway, one can expect dTDP-VioN to be the substrate for transfer of VioN in such cases.

4.1.4. D-Fucose. dTDP-D-Fuc synthesis was characterised in *A. actinomycetemcomitans* and involves the reduction of dTDP-6-deoxy-D-*xylo*-4-hexulose by Fcd.[43]

4.1.5. 4-Acetamido-4,6-dideoxy-D-galactose (*N*-acetylthomosamine). This sugar is found in the enterobacterial common antigen (ECA) where its synthesis pathway has been elucidated. It involves two enzymes encoded by *fdfA* and *fdfB*, that respectively aminate dTDP-6-deoxy-D-*xylo*-4-hexulose to give dTDP-4-amino-4,6-dideoxy-D-galactose, and acetylate the product to dTDP-4- acetamido-4,6-dideoxy-D-galactose.

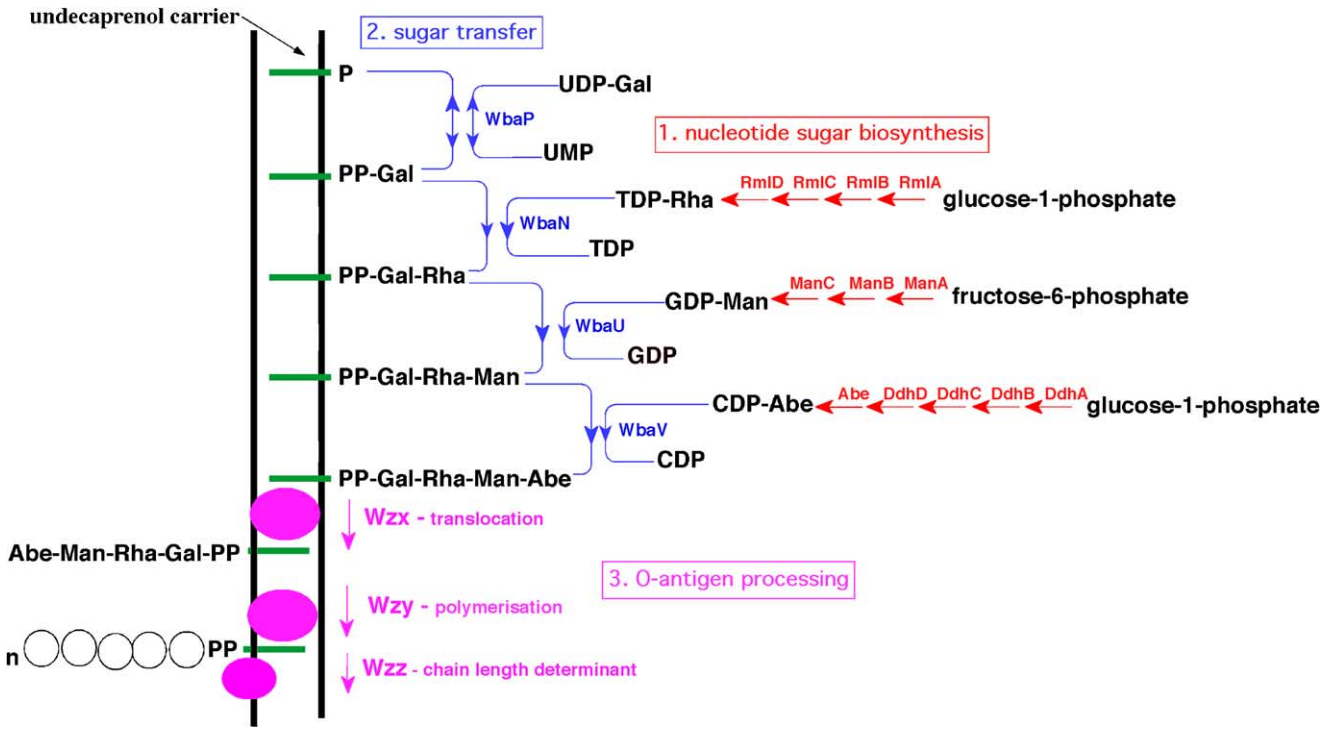

Fig. 2. Synthesis of O-antigens. O-Antigen synthesis involves three groups of proteins i. the proteins involved in the biosynthesis of nucleotide sugar precursors (genes shown in red). This stage occurs in the cytoplasm. ii. The sugar transferases (genes shown in blue), which transfer the nucleotide sugars onto a UndP lipid carrier (green) to form the O-unit. This stage occurs on the cytoplasmic face of the inner membrane. iii. Proteins involved in O-antigen assembly and export (genes shown in purple). This involves both the translocation of the O-unit onto the periplasmic face of the inner membrane, and its polymerisation into an O-antigen

4.1.6. 3-Acetamido-3,6-dideoxy-D-galactose (3-acetamidofucose). This sugar is found in the glycan component of the *Aneurinibacillus thermoaerophilus* S-layer glycoprotein, where its synthesis pathway has been elucidated. It involves three enzymes encoded by *fdtA, fdtB* and *fdtC*, that respectively isomerise dTDP-6-deoxy-D-*xylo*-4-hexulose to dTDP-3-keto-6-deoxyglucose, aminate this product to give dTDP-3-amino-3,6-dideoxy-D-galactose, and finally acetylate it to give the precursor to 3-acetamidofucose (dTDP-3-acetamido-3,6-dideoxy-D-galactose).[44]

4.2. CDP-Sugar biosynthesis pathways

4.2.1. The (dideoxyhexose) *ddh* genes. The sugars synthesised from glucose-1-phosphate as CDP-sugars are all rare, but those known thus far are all present in *Yersinia pseudotuberculosis*, with some also found in other species. The pathway has been characterised for 3,6-dideoxyhexose sugars. The first two genes are known as *ddhA* and *ddhB*. These genes encode proteins that catalyse the conversion of glucose-1-phosphate to CDP-6-deoxy-D-*xylo*-4-hexulose, a pathway that parallels the *rmlA* and *rmlB* dependent synthesis of dTDP-6-deoxy-D-*xylo*-4-hexulose. Similar to dTDP-6-deoxy-D-*xylo*-4-hexulose, CDP-6-deoxy-D-*xylo*-4-hexulose also acts as a branch point, however so far, besides 3,6-dideoxyhexoses, CDP-6-deoxy-D-gulose is the only sugar known to be synthesised by this pathway. In addition to *ddhAB*, 3,6-dideoxyhexoses also require *ddhC* and *ddhD* for their synthesis. DdhC is a pyridoxamine 5'-phosphate dependent Fe-S protein while DdhD is an Fe-S flavoprotein. DdhC and DdhD interact to carry out the C-3 deoxygenation, converting CDP-6-deoxy-D-*xylo*-4-hexulose to CDP-3,6-dideoxyhexoses. *ddhABCD* genes have been identified in *S. enterica* A, B, C2, D1, D2 and D3, which contain paratose, abequose or tyvelose. In addition, they are found in most O-antigen gene clusters of *Y. pseudotuberculosis* and in *V. anguillarum* O1. All four enzymes have been studied over several years by Liu and colleagues, with their work recently reviewed.[45] Note that colitose, another 3,6-dideoxyhexose, is synthesised as GDP-colitose, as discussed below.

4.2.2. 3,6-Dideoxy-D-*ribo*-hexose (paratose), 3,6-dideoxy-D-*xylo*-hexose (abequose), 3,6-dideoxy-D-*arabino*-hexose (tyvelose), and 3,6-dideoxy-L-*arabino*-hexose (ascarylose). 3,6-Dideoxyhexose synthesis pathways diverge after DdhD. CDP-abequose synthase and CDP-paratose synthase reduce the keto group at C-4 to give either the galactose stereochemistry for CDP-Abe, or the glucose stereochemistry for CDP-Par. Both

synthases are homologues of GalE and the CDP-paratose synthase has recently been characterised.[46] CDP-Par 2-epimerase is responsible for the synthesis of CDP-Tyv and is also a GalE homologue.[47] Its structure has recently been determined, and predictions made on its reaction mechanism.[48] Formation of CDP-Asc requires epimerisation at C-5 by AscE followed by reduction of the keto group at C-4 by AscF.[49]

4.2.3. 6-Deoxy-D-gulose (antiariose). The pathway for 6-deoxy-D-gulose is based on genetic data for *Yersinia enterocolitica* O8.[50,51] The *ddhAB* genes and two genes, named *cdgA* and *cdgB,* are present in the *Y. enterocolitica* O8 gene cluster. CdgA and CdgB have 55% and 52% identity to AscE and AscF, which function as an epimerase and C-4 reductase, respectively. Indeed, CdgA and CdgB are presumed to carry out parallel reactions to AscE and AscF, differing only in the use of a 6-deoxy substrate instead of the 3,6-dideoxy intermediate. The genes are immediately downstream of the *ddh* genes, similar to the additional genes of the 3,6-dideoxyhexose pathways in *Y. pseudotuberculosis* and *S. enterica* that were discussed above.

4.3. GDP-Sugar biosynthesis pathways

4.3.1. D-Mannose and GDP-6-deoxy-D-*lyxo*-4-hexulose. The GDP set of pathways starts with a 3-step conversion of fructose-6-phosphate to D-Man, a hexose sugar that is frequently found in O-antigens of *E. coli*, *S. enterica* and other Gram-negative bacteria. The conversion involves a type I phosphomannose isomerase (PMI), a phosphomannomutase and a GDP-mannose pyrophosphorylase (GMP), encoded by *manA*, *manB* and *manC*, respectively. The PMI function is required both for mannose metabolism and biosynthesis of GDP-D-Man. In the *Enterobacteriaceae* the *manA* gene is generally present as part of the mannose metabolism pathway, but is absent in other groups that do not metabolise mannose. The *manB* and *manC* genes are specific to GDP-D-Man synthesis and found in gene clusters for mannose-containing RUPs. In some instances, for example the alginate gene cluster of *P. aeruginosa*, ManC is bifunctional, having both PMI and GMP functions.

The monofunctional and bifunctional forms of ManC are of similar size, and the similarity extends throughout their length. As yet, there is no means of distinguishing them by sequence and as most are identified by homology one can only speculate that ManC is generally monofunctional in species that also have a *manA* gene. *Helicobacter pylori* has a bifunctional ManC and it has been shown that GDP-D-Man, the end product of the pathway, inhibits the PMI activity of this enzyme, probably by binding to the active site of the GMP domain. This provides an explanation for the presence of a bifunctional protein with activity for both the first and last steps of the pathway.[52] The presence of bifunctional and monofunctional ManC enzymes has delayed full adoption of the ManC name, but we think it preferable to retain one name for both forms if only for the pragmatic reason that in most cases the enzyme has not been examined for PMI activity. ManC seems an appropriate name as the common function is GMP activity.

Gmd catalyses the conversion of GDP-D-Man into GDP-6-deoxy-D-*lyxo*-4-hexulose, which then functions as a precursor for other 6-deoxyhexose GDP-sugar nucleotides such as GDP-L-Fuc, GDP-perosamine (GDP-D-Rha4N), GDP-D-Rha, and the 3,6 dideoxyhexose nucleotide sugar GDP-colitose. The *gmd* gene was identified in 1996,[53] and Gmd was later well characterised.[54,55]

4.3.2. L-Fucose and 4,6-dideoxy-D-mannose (perosamine). The *fcl* gene for conversion of 6-deoxy-D-*lyxo*-4-hexulose to GDP-L-Fuc was identified in the colanic acid (CA) gene cluster,[56] and Fcl, a 3,5-epimerase/4-reductase has since been well characterised.[55,57] The *per* gene was identified as the only gene present, together with *manAB* and *gmd*, in gene clusters for RUPs containing perosamine and again has now been characterised biochemically.[58] Both *fcl* and *per* are present in *E. coli* O157,[59] *V. cholerae* O1[60] and *Brucella melitensis*.[61]

4.3.3. 3,6-Dideoxy-L-*xylo*-hexose (colitose). In the case of the colitose containing *S. enterica* O35/*E. coli* O111 O-antigens and *E. coli* O55, the corresponding O-antigen clusters contain the colitose synthesis genes, *col1* and *col2,* which convert 6-deoxy-D-*lyxo*-4-hexulose to GDP-colitose.[62,63] Gene assignment is based on occurrence of the genes in *E. coli*, as above, *Y. pseudotuberculosis* O6, and *V. cholerae* O139, all with colitose-containing O-antigens. Colitose has long been known to be synthesised as the GDP derivative,[64] but the enzymology has still not been reported. The gene names, *col1* and *col2* are temporary, and when the functional order of the enzymes has been determined, the genes will be designated *colAB*. Interestingly, in *E. coli* O55 the *col1* and *col2* genes are found outside of the main O-antigen gene cluster, downstream of *gnd*,[63] in a region that can be thought of as an extension of the O-antigen gene cluster. This arrangement may reflect a relatively recent assembly of the genes for colitose synthesis in this case.

4.3.4. D-Rhamnose. D-Rha is widely present in O-antigens including the *P. aeruginosa* A-band O-antigen, which contains a repeat unit of three D-Rha residues. GDP-D-Rha is synthesised by reduction of GDP-6-

deoxy-D-*lyxo*-4-hexulose. The enzyme responsible, Rmd, has been studied in several labs.[54,65]

4.3.5. 6-Deoxy-D-talose. 6-Deoxy-D-Tal is rare and as yet not reported in O-antigens but is found in a surface polysaccharide of *Actinobacillus actinomycetemcomitans*, for which the pathway has been described. GDP-D-6dTal is synthesised by reduction of GDP-6-deoxy-D-*lyxo*-4-hexulose. The enzyme responsible, Tld, has been studied in two laboratories.[66,67]

4.3.6. GDP-D-Mannose mannosyl hydrolase. The *E. coli* and *S. enterica* gene clusters that include *gmd* generally also have a *gmm* gene, which encodes a GDP-D-Man mannosyl hydrolase. This hydrolase removes GDP-D-Man from the sugar biosynthesis pathway and is thought to regulate the GDP-sugar pathways past GDP-D-Man.[62] An exception to this is found in *E. coli* O55. However, the *E. coli* O55 O-antigen gene cluster does have a remnant *gmm* gene, and Gmm has 57% amino acid identity to Gmm of O157.[63]

4.4. UDP-Glucose based pathways

UDP-Glc has a housekeeping role, and also acts as a precursor for RUP specific sugar pathways

4.4.1. Glucuronic acid and galacturonic acid. UDP-Glc acts as a precursor for UDP-GlcA. The dehydrogenase gene *ugd*, for this reaction, is a housekeeping gene in *E. coli* and *S. enterica* as it is involved in CA synthesis.[53] It is located downstream of *gnd* near the O-antigen gene cluster. The *E. coli* Boydii O4, O5, O6 and O9 O-antigens, which all contain GlcA, do not have an additional copy of *ugd*, but RUP gene clusters in species that lack CA do have the *ugd* gene, identified examples being in the *V. cholerae* O139 gene cluster and *Streptococcus pneumoniae* capsule clusters.[68] UDP-GlcA is itself a precursor for UDP-GalA. The gene involved encodes an epimerase, and was first identified in the *S. pneumoniae* type I capsule gene cluster.[69] We propose that this gene be named *gla*. Homologues of this gene have since been identified in the *E. coli* O113 and the *V. cholerae* O139 O-antigen gene clusters. As both the *E. coli* O113 and the *V. cholerae* O139 O-antigen contain GalA, these genes are presumed to be *gla* genes.

4.4.2. Galactose. We can also mention here the well-studied C-4 epimerase, GalE, which converts UDP-Glc into UDP-Gal, found in many O-antigen structures. UDP-Gal is used for synthesis of essential structures such as LPS core, and in some species GalE also acts (in the reverse direction) in catabolism of Gal. Thus, *galE* is not found in the O-antigen gene cluster but, for example in *E. coli, galE* is in an operon of genes involved in Gal uptake and catabolism.

4.5. UDP-*N*-Acetylglucosamine based pathways

Similar to UDP-Glc, UDP-GlcNAc also has a housekeeping role, as well as acting as a precursor for RUP specific sugar pathways.

4.5.1. *N*-Acetylgalactosamine. Until quite recently it was assumed that GalE, discussed above, was able to use UDP-GlcNAc as well as UDP-Glc as a substrate, but it has now been shown that in some cases at least a separate UDP-GlcNAc epimerase is required for synthesis of UDP-GalNAc. The gene was identified for *E. coli* O113[70] and named *gne*, the name first used for a UDP-GlcNAc epimerase involved in teichoic acid synthesis.[71] *gne* genes were then identified for *P. aeruginosa* O6,[72] *E. coli* O55,[63] *Y. enterolitica* O8[73] and *Plesiomonas shigelloides*.[74]

The *Y. enterolitica* O8 Gne enzyme was particularly well characterised, and compared with GalE enzymes. Gne could carry out epimerisation of both UDP-GlcNAc and UDP-Glc, although the conversion of UDP-Glc to UDP-Gal was less efficient.[73] *E. coli* GalE, however, can only perform the UDP-Glc epimerisation.[73] Molecular modelling revealed the reason for this. The active sites of the proteins differ in having Tyr136 and Tyr299 in GalE compared to Leu136 and Cys297 in Gne. The former residues are much bulkier than the latter and a model of the Gne structure based on the known GalE structure indicated that these residues would hinder access to the site for the bulkier UDP-GlcNAc.[73] This was supported by mutational studies.

An analysis of the two amino acid residues at positions 136 and 297 in a number of published GalE proteins has revealed a possible mis-designation of function, whereby several of these are likely to be Gne proteins.[73] However, Gne and GalE are very similar and it is difficult to designate a particular gene as *gn*e or *galE* based on sequence alone, even if particular attention is made to the two amino acid residues discussed above. For example, in *Y. pseudotuberculosis* a *gne* gene was identified by its presence only in serotypes that had GalNAc in the O-antigen[75] as it does not fully fit the pattern set by the *Y. enterolitica* O8 *gne* gene.

4.5.2. *N*-Acetylgalactosaminuronic acid. A *P. aeruginosa* O6 gene encodes an enzyme that catalyses the conversion of UDP-GalNAc to UDP-GalNAcA.[76] A mutation in this gene could be complemented by a gene in the *S. enterica* Vi gene cluster, thought to perform the same function, and both are now given the name *gna*. Interestingly the *P. aeruginosa* enzyme could also use

UDP-GlcNAc as a substrate, although both the K_m and K_{cat} values indicated that UDP-GalNAc is the preferred substrate.

4.5.3. *N*-Acetyl-L-fucosamine and *N*-acetyl-L-quinovosamine. UDP-L-FucNAc and UDP-L-QuiNAc are derived from UDP-GlcNAc in overlapping pathways. UDP-L-FucNAc is a 2-acetamido-2,6-dideoxyhexose epimerase. The three genes involved in its biosynthesis, *fnlA*, *fnlB* and *fnlC,* were identified in both the *E. coli* O26 O-antigen gene cluster[77] and the *P. aeruginosa* O11 O-antigen gene cluster,[78] but the functions were only recently determined for *P. aeruginosa* O11 and a *Staphylococcus aureus* capsule gene cluster.[79] FnlA, FnlB and FnlC catalyse a 5-step reaction cascade encoding, respectively, a trifunctional 4,6-dehydratase/5-epimerase/3-epimerase, a reductase and a C-2 epimerase.[79]

UDP-L-QuiNAc, also a 2-acetamido-2,6-dideoxyhexose epimerase, is present in the *V. cholerae* O37 O-antigen where the genes were identified.[80] The pathway is initiated by FnlA, and QnlA and QnlB then carry out reactions similar to those of FnlB and FnlC. FnlB and QnlA reduce the 4-keto group to give either UDP-L-PneNAc or UDP-L-RhaNAc, respectively. FnlC and QnlB are C-2 epimerases that act on the respective products to give either UDP-L-Fuc2NAc or UDP-L-Qui2NAc, respectively. The specificity lies in the reduction step as QnlB is capable of carrying out C-2 epimerisation of either substrate, although FnlC can act only on UDP-L-PneNAc. Note, however, that only one example of each has been studied.

4.5.4. *N*-Acetyl-D-quinovosamine and *N*-acetyl-D-fucosamine. D-QuiNAc is present in *P. aeruginosa* O6 and O10, and D-FucNAc in *P. aeruginosa* O5 and O11. The O5 and O6 gene clusters have been sequenced and studied in considerable detail.[81–85] The two sugars are clearly synthesised as UDP-sugars from UDP-GlcNAc in a combined pathway. WbpM initiates the pathway and WbpK is also involved in the synthesis of UDP- D-FucNAc. *P. aeruginosa* O11, which also contains D-FucNAc has, as expected, a homologue of *wbpK.* However, while much has been resolved in a very interesting pathway, it is perhaps premature to allocate gene names until all functions are determined, and consequently gene names have not been included in Fig. 3.

4.5.5. *N*-Acetyl-D-mannosamine and *N*-acetyl-D-mannosaminuronic acid. UDP-ManNAc is synthesised by C-2 epimerisation of UDP-GlcNAc, and UDP-D-ManNAcA is synthesised from UDP-D-ManNAc by UDP-D-ManNAc dehydrogenase. The pathway was first established for the ECA but the genes identified as being involved are present in other gene clusters where expected, and we now call them *mnaA* and *mnaB*. The *mnaA* gene has been found in a range of bacterial gene clusters including that for the *S. enterica* O54 O-antigen gene cluster. There have been several enzymatic studies and the structure has recently been described.[86] The *mnaA* and *mnaB* genes are both present in gene clusters for RUPs that contain ManNAcA, including the *E. coli* ECA gene cluster, the *S. enterica* Vi gene cluster and *S. aureus* capsule gene clusters. *S. aureus* MnaB has recently been biochemically characterised.[87]

4.5.6. 2,3-Diacetamido-2,3-dideoxy-D-mannuronic acid. The *P. aeruginosa* O5 O-antigen gene cluster has homologues of *mnaA* and *mnaB*, that are presumed to be involved in synthesis of the Man(2NAc3N)A, which is found in the O-antigen. One might have expected them to be *mnaA* and *mnaB* genes, with the pathway proceeding via UDP-D-ManNAcA. However the similarity is low and the genes do not complement[88] suggesting that the UDP-D-GlcNAc precursor may be *N*-diacetylated before further modification.[88]

4.6. Other nucleotide sugar synthesis genes

4.6.1. *N*-Acetylneuraminic acid (Neu5Ac, sialic acid). The *E. coli* O104 O-antigen is one of the few *E. coli* O-antigen structures that contains Neu5Ac.[89] The synthesis of Neu5Ac is unusual in that, unlike the nucleotide-sugar biosynthesis pathways discussed above, the CMP nucleotide is not added to the sugar until the final step of Neu5Ac synthesis. Both *Neisseria meningitidis* and *E. coli* group II capsules also contain Neu5Ac. In both cases three genes are required for the synthesis of CMP-Neu5Ac from GlcNA[90,91] and the three homologous genes in the *E. coli* O104 O-antigen gene cluster are assumed to carry out the same reactions. Three sets of names have been used in the past and, to meet the aim of having pathway genes named in pathway order, a new 3-letter base is now employed. The genes encode a putative *N*-acetylmannosamine synthetase (*nnaA*), a NeuNAc synthetase (*nnaB*) and a CMP–NeuNAc synthetase (*nnaC*),[92] all of which are likely to function in Neu5Ac synthesis. Another gene present in the *E. coli* O104 O-antigen gene cluster, *nnaD*, is also homologous to a gene in *N. meningitidis*, the gene product of which was recently shown to interact with the respective Neu5Ac synthetase. Therefore, in *E. coli* O104, NnaD is likely to interact with NnaB.[92]

4.6.2. LD-*manno*-Heptose and DD-*manno*-heptose. ADP-L-*glycero*-D-*manno*-heptose (LD-*man*-Hep) and GDP-D-*glycero*-D-*manno*-heptose (DD-*man*-Hep) are synthesised in parallel interacting pathways, in which the nucleotide component is again atypically added late. LD-*man*-Hep is present in the LPS inner core of many species of Gram-negative bacteria, while DD-*man*-Hep

is less common but found, for example, in an S-layer glycoprotein of *Aneurinibacillus thermoaerophilus.* The pathways for ADP-LD-*man*-Hep and GDP-DD-*man*-Hep were determined recently and have been summarised by Valvano et al.[93] They are shown in Fig. 3 The *hld* and *hdd* genes are specific to the ADP-LD-*man*-Hep and GDP-DD-*man*-Hep pathways, respectively, while the *gmh* genes are common to both pathways. This is an unusual occurrence with some interesting features. The common enzyme GmhA converts D-sedoheptulose 7-P to racemic D-*glycero*-α,β-D-*manno*-heptose 7-P. HddA and HldA are C-1 kinases that phosphorylate the α anomer or β anomer, respectively. This distinction between the two pathways is retained through to formation of DD-*man*-Hep or LD-*man*-Hep. The next two steps convert these 1,7 diphosphate intermediates to GDP-D-*glycero*-α-D-*manno*-heptose or ADP-D-*glycero*-β-D-*manno*-heptose, respectively. In *E. coli* and many other species, this is achieved by the activity of the bifunctional HldE enzyme, which combines the functions of HldA and HldC. The LD-*man*-Hep pathway has a final step carried out by HldD, converting ADP-D-*glycero*-β-D-*manno*-heptose to ADP-L-*glycero*-β-D-*manno*-heptose.

4.6.3. 6-Deoxy-D-*manno*-heptose. 6-Deoxy-D-*manno*-heptose (D-*man*-6dHep) is found in the *Y. pseudotuberculosis* 1a, 2a and 4b O-antigens and also as a homopolymer in a *Burkholderia mallei* capsule. All gene clusters include the genes for synthesis of GDP-DD-*man*-Hep as well as two additional genes that have been named *dmhA* and *dmhB*. Based on homology, *dmhA* is assumed to encode a GDP-DD-*man*-Hep 4,6-dehydratase, which converts GDP-DD-*man*-Hep into GDP-6-deoxy-D-*lyxo*-4-hexulo-DD-*man*-Hep. This intermediate is then assumed to be reduced by DmhB, a C-4 reductase, to GDP-D-*man*-6dHep.[16]

4.7. Overview of biosynthetic pathway genes

We can summarise the state of knowledge of biosynthetic pathways by saying that of the 69 sugars listed as components of RUPs by Knirel and Kochetkov in 1994,[94] the biosynthesis of about half is now understood, such that one can give function based names for the pathway genes. About five sugars, including the very widespread Glc and Gal, are found as NDP sugars in housekeeping functions. The genes for biosynthesis of these NDP sugars are housekeeping genes and not in the O-antigen gene clusters. They have been named in the usual manner for genes in major metabolic pathways, such that in general the same name is used across all bacteria. There are also an additional 30 nucleotide sugars for which pathway genes have been identified in RUP gene clusters and have now been given pathway names (Fig. 3). The 30 nucleotide sugars include several acetamido sugars. Those acetamido sugars that have the *N*-acetyl group at C-2 are generally synthesised from UDP-GlcNAc, and the corresponding amino sugars are very rare or not reported in O-antigens. Other acetamido sugars are synthesised by *N*-acetylation of the corresponding amino sugar. Those included in this review are dTDP-Fuc4NAc, dTDP-Fuc3NAc and dTDP-VioNAc; GDP-PerNAc is also thought to be synthesised in a similar way.[59] This method of synthesis would imply that the respective amino sugar is also available for incorporation into the O-antigen. Indeed, there are several examples of C-3 and C-4 amino sugars in major reviews of O-antigen structures.[9,94]

In general, once specific pathway genes have been identified, they can be recognised with reasonable certainty from sequence alone. However, this approach is far from ideal because although we now have many cases where there is sequence data, there is little or no biochemistry. As for whole genome sequences, it is necessary both to infer what one can from such data and to use names that help us comprehend the inferences. There may be ongoing difficulty in identifying genes from sequence in cases like *gne,* which are closely related to a more common gene, in this case *galE*, and there appears to be an independent change of function in different groups of RUPs or species.

5. Glycosyl transferases and O-unit formation

Glycosyl transferases sequentially add the sugars to the growing sugar chain, until the O-unit/O-antigen has been synthesised.[1] The extensive range of sugars found in O-antigens allows for numerous combinations of donor sugar, acceptor sugar and acceptor carbon atom for the glycosidic linkages. In turn, this provides for a very large number of linkage specificities and hence of glycosyl transferase specificities. It is therefore not surprising that the analysis of O-antigen clusters has led to the identification of an enormous number of different glycosyl transferase genes. The diverse function of the transferases is reflected in the heterogeneity of their DNA sequences. It thus becomes challenging, when analysing individual O-antigen clusters, to identify the specificities of each transferase. Further, only a small proportion of the putative transferases have been studied biochemically, and there are only a few cases where all the transferases in a particular O-antigen cluster have been identified.

The glycosyl transferase genes are commonly found dispersed throughout the O-antigen cluster. In several O-antigen clusters, whose transferase genes have been assigned specific functions, for example in the O-antigen clusters *S. enterica* A, B, C2, D1, D2 and E1,[95] *E. coli* O104,[92] *E. coli* Dysenteriae 1[96] and *P. aeruginosa* O5,[81] O6[82] and O11,[78] it has been observed that the order of

transferase genes within the cluster corresponds to the inverse order of the reaction sequence of the products. Whether or not this represents an important factor in O-antigen biosynthesis is still unknown.

5.1. Classification of glycosyl transferases

Much effort has been placed into creating a classification system for glycosyl transferases. This work initially involved grouping glycosyl transferases depending on their function. They can, for example, be grouped on the basis of the stereochemistry of reaction substrates and products: they either retain or invert the stereochemistry of the glycosyl donor during linkage formation.[97] They were also grouped according to the specific substrate that they transferred.[98] When sequence data became available, it was thought that it too could be used to classify glycosyl transferases, but as discussed above, the sequence similarity of glycosyl transferases is often very low, making this a challenging task.[99,100] Recently, however, hydrophobic cluster analysis (HCA)[101,102] has been used to develop a comprehensive classification of nucleotide-diphospho-sugar glycosyl transferases.[103] Twenty six families were recognised initially and this has now risen to 65 (http://afmb.cnrs-mrs.fr/CAZY/index.html) with just over half including prokaryote examples. Enzymes within a family may have different substrates and transfer different sugars, but are consistent in being either all inverting or all retaining.

5.2. Sugar phosphate transferase genes

O-Unit synthesis is initiated by the transfer of a sugar phosphate to UndP, and only a limited number of sugar phosphates are known to function in this step. The initial transferases catalysing these reactions are distinct in that they have several predicted transmembrane segments, which are uncommon in the simple sugar transferases.[29,104] They correspond to one of two protein families, the polyisoprenyl-phosphate *N*-acetylhexosamine-1-phosphate transferase (PNPT) family or the polyisoprenyl-phosphate hexose-1-phosphate transferase (PHPT) family.[105] Both families have many members but most have not been characterised.

Members of the PHPT family are only found in bacteria, and include WbaP, first identified in *S. enterica* where it acts as a galactose-1-P initial transferase.[1,30] *wbaP* is situated at the 3′ end of the *S. enterica* A, B, C2, D1, D2, D3 and E1 O-antigen gene clusters.[47] The *E. coli* K30 capsule has galactose as its first sugar and *wbaP* has also been identified in the *E. coli* K30 capsule cluster. Other initial transferases in the PHPT family are WbgY, a putative 4N-D-Fuc2NAc-P transferase for *P. shigelloides* and *E. coli* Sonnei O-antigens,[106] and WblG of *Bordetella pertussis*.[107] WcaJ, a glucose-P transferase for CA of *E. coli*,[53] is also in this family.

The PNPT family is found in both Eucaryotes and Procaryotes and includes WecA, which transfers either GalNAc-P or GlcNAc-P to UndP.[108,109] In *Enterobacteriaceae*, *wecA* is situated outside of the O-antigen gene cluster and is part of the ECA gene cluster:[108,110] it acts as the initial transferase for both ECA and those O-antigens that are initiated with either GlcNAc or GalNAc. WbpL also belongs to the PNPT of initial transferases, and has been shown to be required for synthesis of the *P. aeruginosa* O5, O6 and O11 O-antigens,[81] where it is thought to transfer D-FucNAc-P to UndP.[82] For a more in depth review of initial transferases, we refer the reader to Valvano (2003).[105]

5.3. Overview of transferase genes

All O-antigens and other surface polysaccharides appear to be synthesised on UndPP, and as we have seen, there is a clear distinction between the initiating sugar phosphate transferases and the transferases that add the remaining sugars. There seems to be a limited number of specificities among the former group, and one can predict that, as for WecA and WbpA, once characterised in one case, the genes will be relatively easily identified in other gene clusters. In contrast, the simple sugar transferases, which are responsible for most of the variation in the structures, are much more heterogeneous and, as we noted, in general the proteins can be identified in the generic sense as sugar transferases and put into specific HCA families, but one still cannot infer the specific function with confidence from sequence alone. Further, although each transferase can now be placed into a specific HCA family, most of the transferases incorporated in this scheme are from genome sequences and of unknown function. As more transferases are characterised biochemically, and we learn the range of specificities found in each family, we can hope that a combination of gene cluster sequence, HCA family allocation for each putative transferase, and O-antigen structure, may enable identification of most of the transferase genes.

There are already examples of specific transferases being identified in unrelated gene clusters. The *wbaP* and *wbaZ* genes of the *E. coli* K30 capsule gene cluster were named by homology to the characterised *wbaP* and *wbaZ* genes of the *S. enterica* C2 gene cluster. Both clusters have Gal and α-(1→3)-linked D-Man as the first and second sugars, respectively, and given the levels of similarity the genes were considered to have the same function in both. A second example is the D-Gal α-(1→3) GlcNAc linkage which is present in several O-antigen structures. The transferase catalysing this linkage was initially identified in *E. coli* Dysenteriae O1 and named WbbP. In each case where we find other D-Gal α-(1→3)

GlcNAc linkages, we also find a *wbbP* homolog. With such a circumstance, we can be reasonably confident on identification of these genes as the expected D-Gal α-(1→3) transferase. There are other examples but as yet they are few, and the rarity of these cases may reflect the considerable number of specific linkages in RUPs, such that we do not have enough examples to be confident of specific identifications. It is important to identify many more transferase genes as only then can we move more widely to identification by homology.

6. O-Antigen processing genes

There are essentially two major methods of O-antigen processing that have been described, those requiring the *wzx, wzy* and (sometimes) *wzz* genes, and those requiring the *wzm* and *wzt* genes. Typical *E. coli* and *S. enterica* heteropolymer O-antigens require the functions of Wzx, Wzy and Wzz for normal O-antigen assembly. For a more detailed look at the biochemical mechanisms of the O-antigen processing proteins, we refer you to a recent review by Valvano.[105]

6.1. Wzx-flippase

Once the UndPP-linked O-unit is completed, it is thought to be 'flipped' across the inner membrane by a flippase encoded by the *wzx* gene,[111] although the evidence is indirect. Wzx proteins are highly hydrophobic membrane proteins generally having 12 transmembrane (TM) segments. They reveal very little sequence conservation, however groups of Wzx proteins do share similar motifs and *wzx* genes that are not identified by similarity to known *wzx* genes can generally be identified using a few Wzx sequences and programs which look for conserved blocks in a group of two or more unaligned protein sequences, for example by using the Block Maker program (www.blocks.fhcrc.org/blockmkr/make_blocks.html).[68]

Complementation experiments have shown that *E. coli* O7, O16 and *S. enterica* B O-antigen clusters can complement a *wzx* gene defect in *E. coli* O16 or O7.[112] Each Wzx protein acts as a flippase for a specific UndPP-linked oligosaccharide, yet these results demonstrate that Wzx action is not highly specific. How Wzx functions, and the reason for its high level of variation, is not known.

6.2. Wzy-polymerase

Once the UndPP-linked O-unit has been translocated across the membrane, Wzy, the O-antigen polymerase, functions to link the O-units via a glycosidic linkage, polymerising them into a long chain O-antigen. As each O-unit varies, Wzy proteins will generally differ in specificity. Thus, although these proteins all share a common name, one would not expect a particular *wzy* to complement mutations in other *wzy* genes. The specificity of *wzy* genes is nicely illustrated in *S. agalactiae* where capsule types 1a and 3 differ only in their *wzy* genes. Each Wzy polymerises a common oligosaccharide to form either a 3-sugar main chain with 2-sugar side branch or a 2-sugar main chain with 3-sugar side branch.[113]

Very little in vitro work on Wzy has been carried out due to experimental difficulties, and consequently Wzy has not been well characterised. Sequence analysis shows that Wzy is highly hydrophobic with 8-12 TM segments. A cytoplasmic loop of approximately 30 amino residues is also a characteristic of the protein structure.[114] As there is often little similarity in either the TM regions or the cytoplasmic loop of different Wzy proteins, identification based on sequence alone is often difficult, but use of the program Block Maker (see above) is usually more successful. If not, then advantage can be taken of the fact that most gene clusters containing a *wzx* gene have only one other gene encoding a highly hydrophobic protein that has the appropriate topology.

6.3. Wzz–chain length determinant

The third O-antigen processing protein, Wzz, is the O-antigen chain length determinant. Wzz is situated in the inner membrane and the protein probably functions, via an interaction with Wzy, to control the chain length of the O-antigen,[115,116] although how it carries out its role is still unknown. Wzz is characterised by two TM segments, located in the amino and carboxyl ends, with a large periplasmic hydrophilic domain.[117] Unlike Wzx and Wzy, Wzz is not excessively variable and easily identified by sequence. In *E. coli* and *S. enterica*, *wzz* is not normally found within the O-antigen gene cluster, but is situated just downstream, between *gnd* and the *his* operon.[116,118] In other O-antigen gene clusters, *wzz* is found within the major gene cluster, being at the 5′ end of the in *P. aeruginosa*,[8] and at the 3′ end of the gene cluster in *Y. pseudotuberculosis*.[75]

6.4. The Wzm/Wzt export system

The Wzx/Wzy assembly and export pathway predominates among the better-studied O-antigens, but many RUPs have the Wzm/Wzt system. For *E. coli* and *S. enterica* O-antigens, it is only recorded for homopolymers, but it is more common in other species as well as in *E. coli* and *S. enterica* capsules. RUPs with the Wzm/Wzt assembly and export pathway are synthesised entirely within the cytoplasm, before being transported across the membrane by an ABC transporter.[119,120] The genes encoding this ABC transporter are *wzm* and *wzt*

and they have been identified in *E. coli* O8, O9 and O9a O-antigen gene clusters.[120] This system of O-antigen processing can also be seen in *V. cholerae* O1, O22 and O139, and *V. anguillarum* O1 O-antigens.

7. Origins and maintenance of O-antigen gene clusters

Most O-antigen gene clusters contain those genes that are necessary for O-antigen synthesis, namely, biosynthetic pathway genes, transferase genes and genes involved in export and assembly. Lawrence[121] has proposed that bacterial operons arose because having the genes clustered facilitated lateral gene transfer. This applies very strongly to O-antigen and other RUP gene clusters. There are many cases of substitution of one O-antigen form by another by replacement of the gene cluster, well known examples being the gain of the O157 gene cluster by an O55:H7 strain to generate the infamous O157:H7 clone of *E. coli*, and the replacement of the O1 O-antigen by the O139 O-antigen in 7th pandemic *V. cholerae*. Transfer between species also occurs and the O-antigen complement of a species changes over time such that *E. coli* and *S. enterica* have only three forms in common. These events are only possible if the genes are clustered. We have discussed the population and evolutionary issues more fully elsewhere.[26,122,123]

7.1. Diversity of O-antigen gene clusters is extended by genetic exchange

The study of gene clusters often gives us clues as to their evolutionary origins. There are many instances where several O-antigens of a species have related structures. For example most *Y. pseudotuberculsis* O-antigens have related structures and gene clusters.[16,75] These gene clusters include near identical genes for common components and quite different genes for the components that differentiate the O-antigens. In some cases one can see how one gene cluster could be derived from another by recombination between two others, such as *Y. pseudotuberculosis* 1a through reassortment by recombination of the genes in the 1b and 4b gene clusters.[16,75] Another example is of the probable origin of *S. enterica* D2 from D1 and E1 clusters[124] with a mutant H-repeat transposable element at the putative recombination site (Fig. 2).

In the above cases, the new gene clusters are typical in containing all the genes for the new structure. However some gene clusters have non-functional remnants of genes. A well-documented example of a remnant gene involves the *S. enterica* A, D1 and B O-antigens, in which *wzy*, which is responsible for an α-(1→6)-linkage between O-units, is usually far from the main gene cluster. However some group B strains with an α-(1→2) linkage between O-units have an α-(1→2)-specific *wzy* gene within the gene cluster, and there is good evidence for this being the ancestral situation. Those B strains, including the well documented strain LT2, with the distant α-(1→6)- *wzy* gene, have recently been shown to have a remnant α-(1→2)-specific *wzy* gene between *wbaV* and *wbaU*,[125] and several stages in reduction of this remnant were observed. A remnant of the α-(1→2)-specific *wzy* gene is also found in the *S. enterica* D1 gene cluster between *wbaV* and *wbaO*. This suggests that the α-(1→6)-linkage replaced the α-(1→2)-linkage, and that this newly acquired α-(1→6) *wzy* gene has not yet migrated into the main cluster.

Non-housekeeping genes essential for the O-antigen synthesis may also be found outside of the gene cluster. For example in the *E. coli* O55 O-antigen gene cluster the colitose synthesis genes, *col1* and *col2*, are found just downstream of the O-antigen gene cluster.[63] The *manB*, *manC* and *gmd* genes that precede *col1* and *col2* in function are in the main cluster, and presumably were involved in synthesis of a different GDP-sugar until the two *col* genes were gained, after which there was a subsequent loss of the other pathway. Although *col1* and *col2* are close enough to the major gene cluster for co-transfer, this is perhaps an intermediate stage, and one can predict that selection pressure will eventually bring them within the cluster for easier lateral transfer of all of the genes.

7.2. Late modification of the O-unit

Several acetyl transferase and glucosyl transferase genes involved in O-antigen synthesis are also found outside of the gene clusters for a different reason. These include transferase genes whose products modify O-antigens after polymerisation. For example, the *S. enterica* B O-antigen can be modified by addition of glucose to the galactose residue. This is either via a (1→4) linkage, determined by chromosomal genes, or a (1→6) linkage, determined by the gene on the P22 lysogenic phage.[126–128] The glycosyl transfer occurs after the O-unit has been translocated to the periplasmic face of the CM. The transfer is from Und-P-Glc, which is synthesised and translocated by GtrB and GtrA respectively, common to all such pathways. Note that these are not in function order but these order names are now well established. The third gene encodes the transferase and is variable, as is to be expected as it determines linkage specificity. We suggest each should have a specific four-letter name although this has not always been done. *O*-Acetyl transferases may also be important in the synthesis and modification of particular O-antigens. The genes encoding these enzymes can sometimes be found outside the O-antigen cluster, for example *oafA* of *S. enterica* that encodes an O-acetyl transferase that modifies the group B antigen by acetylation of the Abe

residue,[129] and *gtrA*, *gtrB* and *wgtJ* in the *Y. pseudotuberculosis* O7 gene cluster.[75]

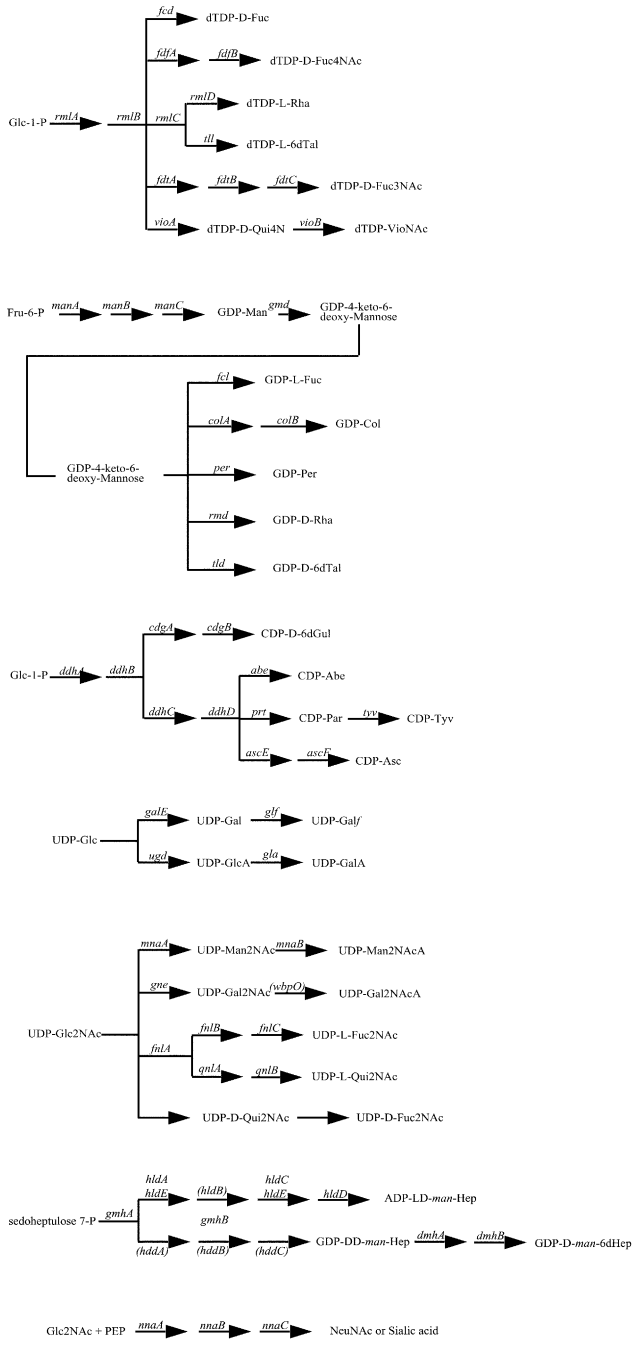

Fig. 3. Nucleotide-sugar biosynthetic pathways. The nucleotide-sugar biosynthetic pathways can be grouped on the basis of their sugar-linked nucleotide: either CDP, UDP, dTDP or GDP. For each nucleotide pathway the initial nucleotide-sugar precursor is shown on the left hand of the diagram. The Neu5Ac and heptose pathways are exceptions in that the nucleotide is added to the sugar precursor at a late stage of the pathway. Names in parentheses (*hldB, hddB*) were not applied to the genes found, but held for single pathway enzymes should they occur in other gene clusters. The genes involved in each reaction step of the pathways are shown above the arrows. See abbreviation list for complete sugar names.

Such modifications are found in most well-studied systems so are probably very widespread. A remarkable case is found in *E. coli* Flexneri, recently reviewed by Allison and Verma[130] in which 12 distinct antigenic forms have the same main chain but vary in the presence of phage encoded O-acetyl and glucose side-branches.

8. Concluding remarks

O-Antigens are an important component of the outer membrane of Gram-negative bacteria. They are highly variable and occur in multiple forms, with substantial variation within and between different species. The genes involved in O-antigen biosynthesis are generally found on the chromosome as an O-antigen gene cluster and the structural variation of O-antigens is mirrored by genetic variation seen in these clusters. In this review, we have summarised the current knowledge on O-antigen gene clusters, focusing our attention on the most extensively studied *E. coli* and *S. enterica* clusters.

The genes in the O-antigen cluster can typically be categorised into three groups. These include the nucleotide sugar pathway genes, sugar glycosyl transferase genes and O-antigen processing genes. Nucleotide sugar biosynthesis genes are often found together in the cluster for a given pathway. A significant number of nucleotide sugar biosynthesis pathways have now been identified and these pathways seem to be conserved in different O-antigen clusters and across a wide range of species such that, in general, once a particular sugar pathway has been identified in one case, it can be identified in other cases with a confidence level comparable to that for other genes identified by sequence in 'genome' studies.

An analysis of the O-antigen processing genes reveals that, although different *wzx* and *wzy* genes have little sequence conservation, they are also normally identifiable in the O-antigen gene cluster, as is the more conserved *wzz*. On the other hand, the glycosyl transferase genes are often dispersed throughout the O-antigen gene cluster, and have only low levels of similarity to other glycosyl transferases. This is sufficient for them to be identified as glycosyl transferases by homology but not to be allocated a specific function. Indeed, the specific function of only a small proportion of putative glycosyl transferases has been elucidated. Their characterisation will prove the most challenging in the future. As additional data is collected, further trends both within and between the different O-antigen clusters from distinct bacterial species will be identified with the aim of discovering the genetic and evolutionary basis for bacteria expressing so many forms of O-antigen on their cell surface.

References

1. Reeves, P. R.; Hobbs, M.; Valvano, M.; Skurnik, M.; Whitfield, C.; Coplin, D.; Kido, N.; Klena, J.; Maskell, D.; Raetz, C.; Rick, P. *Trends Microbiol.* **1996**, *4*, 495–503.
2. CDC. Laboratory methods for the diagnosis of epidemic dysentery and cholera, 1999.
3. Lior, H. In *Classification of Escherichia coli*; Gyles, C. L., Ed.; CAB International, 1994; pp 31–72.
4. Popoff, M. Y.; Minor, L. L. *Antigenic formulas of the Salmonella* serovars; 7th revision; WHO Collaborating Centre for Reference and Research on *Salmonella*. Institut Pasteur, 1997.
5. Kenne, L.; Lindberg, B.; Söderholm, E.; Bundle, D. R.; Griffith, D. W.; Morris, J. G., Jr. *Carbohydr. Res.* **1983**, *111*, 289–296.
6. Lindberg, B.; Lindh, F.; Lönngren, J.; Lindberg, A. A.; Svenson, S. B. *Carbohydr. Res.* **1981**, *97*, 105–112.
7. Perry, M. B.; MacLean, L.; Griffith, D. W. *Biochem. Cell Biol.* **1986**, *64*, 21–28.
8. Raymond, C. K.; Sims, E. H.; Kas, A.; Spencer, D. H.; Kutyavin, T. V.; Ivey, R. G.; Zhou, Y.; Kaul, R.; Clendenning, J. B.; Olson, M. V. *J. Bacteriol.* **2002**, *184*, 3614–3622.
9. Jansson, P.-E. In *The chemistry of O-polysaccharide chains in bacterial lipopolysaccharides*; Brade, H.; Opal, S. M.; Vogel, S. N.; Morrison, D. C., Eds.; Marcel Dekker, 1999; pp 155–178.
10. Jansson, P. E.; Lönngren, J.; Widmalm, G.; Leontein, K.; Slettengren, K.; Svenson, S. B.; Wrangsell, G.; Dell, A.; Tiller, P. R. *Carbohydr. Res.* **1985**, *145*, 59–66.
11. Prehm, P.; Jann, B.; Jann, K. *Eur. J. Biochem.* **1976**, *67*, 41–55.
12. Kido, N.; Kobayashi, K. *J. Bacteriol.* **2000**, *182*, 2567–2573.
13. Keenleyside, W. J.; Perry, M. B.; MacLean, L. L.; Poppe, C.; Whitfield, C. *Mol. Microbiol.* **1994**, *11*, 437–448.
14. Reeves, P. R. In *Biosynthesis and assembly of lipopolysaccharide*; Neuberger, A.; van Deenen, L. L. M., Eds.; Vol. 27; Elsevier Science Publishers, 1994; pp 281–314.
15. Sozhamannan, S.; Deng, Y. K.; Li, M.; Sulakvelidze, A.; Kaper, J. B.; Johnson, J. A.; Nair, G. B. *Infect. Immun.* **1999**, *67*, 5033–5040.
16. Pacinelli, E.; Wang, L.; Reeves, P. R. *Infect. Immun.* **2002**, *70*, 3271–3276.
17. Zhang, L.; Toivanen, P.; Skurnik, M. *Microbiology* **1996**, *142*, 277–288.
18. Jann, K.; Jann, B. In *Structure and biosynthesis of O-antigen*; Rietschel, E. T., Ed.; Vol. 1; Elsevier Scientific Publishers, 1984; pp 138–186.
19. Sugiyama, T.; Kido, N.; Kato, Y.; Koide, N.; Yoshida, T.; Yokochi, T. *Gene* **1997**, *198*, 111–113.
20. Zhang, L.; Al-Hendy, A.; Toivanen, P.; Skurnik, M. *Mol. Microbiol.* **1993**, *9*, 309–321.
21. Skurnik, M.; Venho, R.; Bengoechea, J. A.; Moriyon, I. *Mol. Microbiol.* **1999**, *31*, 1443–1462.
22. Marolda, C. L.; Valvano, M. A. *J. Bacteriol.* **1998**, *180*, 3070–3079.
23. Hobbs, M.; Reeves, P. R. *Mol. Microbiol.* **1994**, *12*, 855–856.
24. Bailey, M. J. A.; Hughes, C.; Koronakis, V. *Mol. Microbiol.* **1997**, *26*, 845–851.
25. Artsimovitch, I.; Landick, R. *Cell* **2002**, *109*, 193–203.
26. Pupo, G. M.; Lan, R.; Reeves, P. R. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 10567–10572.
27. Verma, N. K.; Quigley, N. B.; Reeves, P. R. *J. Bacteriol.* **1988**, *170*, 103–107.
28. Wyk, P.; Reeves, P. R. *J. Bacteriol.* **1989**, *171*, 5687–5693.
29. Jiang, X. M.; Neal, B.; Santiago, F.; Lee, S. J.; Romana, L. K.; Reeves, P. R. *Mol. Microbiol.* **1991**, *5*, 695–713.
30. Curd, H.; Liu, D.; Reeves, P. R. *J. Bacteriol.* **1998**, *180*, 1002–1007.
31. Keenleyside, W. J.; Whitefield, C. *J. Biol. Chem.* **1996**, *271*, 28581–28592.
32. Whitfield, C. *Trends Microbiol.* **1995**, *3*, 178–185.
33. Zuccotti, S.; Zanardi, D.; Rosano, C.; Sturla, L.; Tonetti, M.; Bolognesi, M. *J. Mol. Biol.* **2001**, *313*, 831–843.
34. Blankenfeldt, W.; Giraud, M. F.; Leonard, G. A.; Rahim, R.; Creuzenet, C.; Lam, J. S.; Naismith, J. H. *Acta Crystallogr., D Biol. Crystallogr.* **2000**, *56*, 1501–1504.
35. Allard, S. T. M.; Giraud, M.-F.; Whitfield, C.; Messner, P.; Naismith, J. H. *Acta Crystallogr., Sect. D* **2000**, *56*, 222–225.
36. Allard, S. T.; Giraud, M. F.; Whitfield, C.; Graninger, M.; Messner, P.; Naismith, J. H. *J. Mol. Biol.* **2001**, *307*, 283–295.
37. Giraud, M. F.; Gordon, F. M.; Whitfield, C.; Messner, P.; McMahon, S. A.; Naismith, J. H. *Acta Crystallogr., D Biol. Crystallogr.* **1999**, *55*, 706–708.
38. Giraud, M. F.; McMiken, H. J.; Leonard, G. A.; Messner, P.; Whitfield, C.; Naismith, J. H. *Acta Crystallogr. D, Biol. Crystallogr.* **1999**, *55*, 2043–2046.
39. Wang, L.; Qu, W.; Reeves, P. R. *Infect. Immun.* **2001**, *69*, 6923–6930.
40. Kaplan, J. B.; Perry, M. B.; MacLean, L. L.; Furgang, D.; Wilson, M. E.; Fine, D. H. *Infect. Immun.* **2001**, *69*, 5375–5384.
41. Nakano, Y.; Suzuki, N.; Yoshida, Y.; Nezu, T.; Yamashita, Y.; Koga, T. *J. Biol. Chem.* **2000**, *275*, 6806–6812.
42. Marolda, C. L.; Feldman, M. F.; Valvano, M. A. *Microbiology* **1999**, *145*, 2485–2495.
43. Yoshida, Y.; Nakano, Y.; Nezu, T.; Yamashita, Y.; Koga, T. *J. Biol. Chem.* **1999**, *274*, 16933–16939.
44. Pfoestl, A.; Hofinger, A.; Kosma, P.; Messner, P. *J. Biol. Chem.* **2003**, *278*, 26410–26417.
45. Johnson, D. A.; Liu, H. *Curr. Opin. Chem. Biol.* **1998**, *2*, 642–649.
46. Hallis, T. M.; Lei, Y.; Que, N. L.; Liu, H. *Biochemistry* **1998**, *37*, 4935–4945.
47. Reeves, P. R. *Trends Genet.* **1993**, *9*, 17–22.
48. Koropatkin, N. M.; Liu, H.-W.; Holden, H. M. *J. Biol. Chem.* **2003**, *278*, 20874–20881.
49. Thorson, J. S.; Lo, S. F.; Ploux, O.; He, X.; Liu, H. W. *J. Bacteriol.* **1994**, *176*, 5483–5493.
50. Zhang, L.; Radziejewska-Lebrecht, J.; Krajewska-Pietrasik, D.; Toivanen, P.; Skurnik, M. *Mol. Microbiol.* **1997**, *23*, 63–76.

51. Skurnik, M. In *Molecular genetics, biochemistry and biological role of Yersinia* lipopolysaccharide; Skurnik, M.; Bengoechea, J. A.; Granfors, K., Eds.; Kluwer Plenum, 2003.
52. Wu, B.; Zhang, Y.; Zheng, R.; Guo, C.; Wang, P. G. *FEBS Lett.* **2002**, *519*, 87–92.
53. Stevenson, G.; Andrianopoulos, K.; Hobbs, H.; Reeves, P. R. *J. Bacteriol.* **1996**, *178*, 4885–4893.
54. Kneidinger, B.; Graninger, M.; Adam, G.; Puchberger, M.; Kosma, P.; Zayni, S.; Messner, P. *J. Biol. Chem.* **2001**, *276*, 5577–5583.
55. Somoza, J. R.; Menon, S.; Schmidt, H.; Joseph-McCarthy, D.; Dessen, A.; Stahl, M. L.; Somers, W. S.; Sullivan, F. X. *Structure Fold Des.* **2000**, *8*, 123–135.
56. Andrianopoulos, K.; Wang, L.; Reeves, P. R. *J. Bacteriol.* **1998**, *180*, 998–1001.
57. Rosano, C.; Bisso, A.; Izzo, G.; Tonetti, M.; Sturla, L.; De Flora, A.; Bolognesi, M. *J. Biol. Chem.* **2000**, *303*, 77–91.
58. Albermann, C.; Piepersberg, W. *Glycobiology* **2001**, 11.
59. Wang, L.; Reeves, P. R. *Infect. Immun.* **1998**, *66*, 3545–3551.
60. Stroeher, U. H.; Karageorgos, L. E.; Brown, M. H.; Morona, R.; Manning, P. A. *Gene* **1995**, *166*, 33–42.
61. Godfroid, F.; Taminiau, B.; Danese, I.; Denoel, P.; Tibor, A.; Weynants, V.; Cloeckaert, A.; Godfroid, J.; Letesson, J. J. *Infect. Immun.* **1998**, *66*, 5485–5493.
62. Wang, L.; Reeves, P. R. *J. Bacteriol.* **2000**, *182*, 5256–5261.
63. Wang, L.; Huskic, S.; Cisterne, A.; Rothemund, D.; Reeves, P. R. *J. Bacteriol.* **2002**, 184.
64. Elbein, A. D.; Heath, E. C. *J. Biol. Chem.* **1965**, *240*, 1926–1931.
65. Maki, M.; Jarvinen, N.; Rabina, J.; Roos, C.; Maaheimo, H.; Renkonen, R. *Eur. J. Biochem.* **2002**, *269*, 593–601.
66. Suzuki, N.; Nakano, Y.; Yoshida, Y.; Nezu, T.; Terada, Y.; Yamashita, Y.; Koga, T. *Eur. J. Biochem.* **2002**, *269*, 5963–5971.
67. Maki, M.; Jarvinen, N.; Rabina, J.; Maaheimo, H.; Mattila, P.; Renkonen, R. *Glycobiology* **2003**, *13*, 295–303.
68. Jiang, S.-M.; Wang, L.; Reeves, P. R. *Infect. Immun.* **2001**, *69*, 1244–1255.
69. Munoz, R.; Lopez, R.; de Frutos, M.; Garcia, E. *Mol. Microbiol.* **1999**, *31*, 703–713.
70. Paton, A. W.; Paton, J. C. *Infect. Immun.* **1999**, *67*, 5930–5937.
71. Estrela, A. I.; Pooley, H. M.; de Lencastre, H.; Karamata, D. *J. Gen. Microbiol.* **1991**, *137*, 943–950.
72. Creuzenet, C.; Belanger, M.; Wakarchuk, W. W.; Lam, J. S. *J. Biol. Chem.* **2000**, *275*, 19060–19067.
73. Bengoechea, J. A.; Pinta, E.; Salminen, T.; Oertelt, C.; Holst, O.; Radziejewska-Lebrecht, J.; Piotrowska-Seget, Z.; Venho, R.; Skurnik, M. *J. Bacteriol.* **2002**, *184*, 4277–4287.
74. Kowal, P.; Wang, P. G. *Biochemistry* **2002**, 15410–15414.
75. Reeves, P. R.; Pacinelli, E.; Wang, L. In *O-antigen gene clusters of Yersinia pseudotuberculosis*; Skurnik, M.; Bengoechea, J. A.; Granfors, K., Eds.; Kluwer Plenum, 2003; pp 199–206.
76. Zhao, X.; Creuzenet, C.; Belanger, M.; Egboshima, E.; Li, J.; Lam, J. S. *J. Biol. Chem.* **2000**, *275*, 33252–33259.
77. D'Souza, J. M.; Wang, L.; Reeves, P. R. *Gene* **2002**, *297*, 123–127.
78. Dean, C. R.; Franklund, C. V.; Retief, J. D.; Coyne, M. J. J. R.; Hatano, K.; Evans, D. J.; Pier, G. B.; Goldberg, J. B. *J. Bacteriol.* **1999**, *181*, 4275–4284.
79. Kneidinger, B.; O'Riordan, K.; Li, J.; Brisson, J.-R.; Lee, J. C.; Lam, J. S. *J. Biol. Chem.* **2003**, *278*, 3615–3627.
80. Kneidinger, B.; Larocque, S.; Brisson, J.-R.; Cadotte, N.; Lam, J. S. *Biochem. J.* **2003**, *371*, 989–995.
81. Burrows, L. L.; Charter, D. F.; Lam, J. S. *Mol. Microbiol.* **1996**, *22*, 481–495.
82. Belanger, M.; Burrows, L. L.; Lam, J. S. *Microbiology* **1999**, *145*, 3505–3521.
83. Burrows, L. L.; Urbanic, R. V.; Lam, J. S. *Infect. Immun.* **2000**, *68*, 931–936.
84. Creuzenet, C.; Lam, J. S. *Mol. Microbiol.* **2001**, *41*, 1295–1310.
85. Creuzenet, C.; Urbanic, R. V.; Lam, J. S. *J. Biol. Chem.* **2002**, *277*, 26729–26778.
86. Campbell, R. E.; Mosimann, S. C.; Tanner, M. E.; Strynadka, N. C. J. *Biochemistry (Communication)* **2000**, *39*, 14993–15001.
87. Portolés, M.; Kiser, K. B.; Bhasin, N.; Chan, K. H. N.; Lee, J. C. *Infect. Immun.* **2001**, *69*, 917–923.
88. Burrows, L. L.; Pigeon, K. E.; Lam, J. S. *FEMS Microbiol. Lett.* **2000**, *189*, 135–141.
89. Gamian, A.; Romanowska, E.; Ulrich, J.; Defaye, J. *Carbohydr. Res.* **1992**, *236*, 195–208.
90. Annunziato, P. W.; Wright, L. F.; Vann, W. F.; Silver, R. P. *J. Bacteriol.* **1995**, *177*, 312–319.
91. Petersen, M.; Fessner, W.; Frosch, M.; Luneberg, E. *FEMS Microbiol. Lett.* **2000**, *184*, 161–164.
92. Wang, L.; Briggs, C. E.; Rothemund, D.; Fratamico, P.; Luchansky, J. B.; Reeves, P. R. *Gene* **2001**, *270*, 231–236.
93. Valvano, M. A.; Marolda, C. L.; Bittner, M.; Glaskin-Clay, M.; Simon, T. L.; Klena, J. D. *J. Bacteriol.* **2000**, *182*, 488–497.
94. Knirel, Y. A.; Kochetkov, N. K. *Biochemistry (Moscow)* **1994**, *59*, 1325–1383.
95. Liu, D.; Haase, A. M.; Lindqvist, L.; Lindberg, A. A.; Reeves, P. R. *J. Bacteriol.* **1993**, *175*, 3408–3413.
96. Klena, J. D.; Schnaitman, C. A. *Mol. Microbiol.* **1993**, *9*, 393–402.
97. Henrissat, B. *Biochem. J.* **1991**, *280*, 309–316.
98. IUBMB. *Enzyme nomenclature: Recommendations of the nomenclature committee of the International Union of Biochemistry and Molecular Biology*; Press, A., Ed.; 1992.
99. Paulson, J. C.; Colley, K. J. *Proc. Natl. Acad. Sci. USA* **1989**, *264*, 17615–17618.
100. Joziasse, D. H. *Glycobiology* **1992**, *2*, 271–277.
101. Gaboriaud, C.; Bissery, V.; Benchetrit, T.; Mornon, J. P. *FEBS Lett.* **1987**, *224*, 149–155.
102. Lemesle-Varloot, L.; Henrissat, B.; Gaboriaud, C.; Bissery, V.; Morgat, A.; Mornon, J. P. *Biochimie* **1990**, *72*, 555–574.
103. Campbell, J. A.; Davies, G. J.; Bulone, V.; Henrissat, B. *Biochem. J.* **1997**, *326*, 929–939.

104. Amer, A. O.; Valvano, M. A. *J. Bacteriol.* **2000**, *182*, 498–503.
105. Valvano, M. A. *Frontiers Biosci.* **2003**, *8*, S452–S471.
106. Xu, D. Q.; Cisar, J. O.; Ambulos Jr, N.; Burr, D. H.; Kopecko, D. J. *Infect. Immun.* **2000**, *78*, 4414–4423.
107. Allen, A. G.; Maskell, D. J. *Mol. Microbiol.* **1996**, *19*, 37–52.
108. Alexander, D. C.; Valvano, M. A. *J. Bacteriol.* **1994**, *176*, 7079–7084.
109. Amor, P. A.; Whitfield, C. *Mol. Microbiol.* **1997**, *26*, 145–161.
110. Meier, U.; Mayer, H. *J. Bacteriol.* **1985**, *163*, 756–762.
111. Liu, D.; Cole, R.; Reeves, P. R. *J. Bacteriol.* **1996**, *178*, 2102–2107.
112. Feldman, M. F.; Marolda, C. L.; Monteiro, M. A.; Perry, M. B.; Parodi, A. J.; Valvano, M. A. *J. Biol. Chem.* **1999**, *274*, 35129–35138.
113. Yamamoto, S.; Miyake, K.; Koike, Y.; Watanabe, M.; Machida, Y.; Ohta, M.; Iijima, S. *J. Bacteriol.* **1999**, *181*, 5176–5184.
114. Morona, R.; Mavris, M.; Fallarino, A.; Manning, P. A. *J. Bacteriol.* **1994**, *176*, 733–747.
115. Batchelor, R. A.; Alifano, P.; Biffali, E.; Hull, S. I.; Hull, R. A. *J. Bacteriol.* **1992**, *174*, 5228–5236.
116. Bastin, D. A.; Brown, P. K.; Haase, A.; Stevenson, G.; Reeves, P. R. *Mol. Microbiol.* **1993**, *7*, 725–734.
117. Morona, R.; Van Den Bosch, L.; Manning, P. *J. Bacteriol.* **1995**, *177*, 1059–1068.
118. Franco, V. A.; Liu, D.; Reeves, P. R. *J. Bacteriol.* **1998**, *180*, 2670–2675.
119. Reizer, J.; Reizer, A.; Saier, M. H., Jr. *Protein Sci.* **1992**, *1*, 1326–1332.
120. Kido, N.; Torgov, V. I.; Sugiyama, T.; Uchiya, K.; Sugihara, H.; Komatsu, T.; Kato, N.; Jann, K. *J. Bacteriol.* **1995**, *177*, 2178–2187.
121. Lawrence, J. G.; Roth, J. R. *Genetics* **1996**, *143*, 1843–1860.
122. Reeves, P. R.; Wang, L. In *Genomic Organization of LPS-specific Loci*; Hacker, J.; Kaper, J. B.; Compans, W.; Cooper, M., Eds.; Springer Verlag, 2002; 264/I, pp. 109–135.
123. Shepherd, J. G.; Wang, L.; Reeves, P. R. *Infect. Immun.* **2000**, *68*, 6056–6061.
124. Xiang, S. H.; Hobbs, M.; Reeves, P. R. *J. Bacteriol.* **1994**, *176*, 4357–4365.
125. Wang, L.; Andrianopoulos, K.; Liu, D.; Popoff, M. Y.; Reeves, P. R. *J. Bacteriol.* **2002**, *184*, 1669–1677.
126. Mäkelä, P. H. *J. Bacteriol.* **1973**, *116*, 847–856.
127. Sanderson, K. E.; Hessel, A.; Rudd, K. E. *Microbiol. Rev.* **1995**, *59*, 241–303.
128. Vander Byl, C.; Kropinski, A. M. *J. Bacteriol.* **2000**, *182*, 6472–6481.
129. Slauch, J. M.; Lee, A. A.; Mahan, M. J.; Mekalanos, J. J. *J. Bacteriol.* **1996**, *178*, 5904–5909.
130. Allison, G. E.; Verma, N. K. *Trends Microbiol.* **2000**, *8*, 17–23.