Escherichia coli O123 O antigen genes and polysaccharide structure are conserved in some *Salmonella enterica* serogroups

Clifford G. Clark,[1,2] Andrew M. Kropinski,[3] Haralambos Parolis,[4] Christopher C. R. Grant,[1] Keri M. Trout-Yakel,[1] Kristyn Franklin,[3] Lai-King Ng,[1,2] Nikolay A. Paramonov,[4]† Lesley A. S. Parolis,[4] Kris Rahn[3] and Helen Tabor[1]

Correspondence
Clifford G. Clark
clifford_clark@phac-aspc.gc.ca

[1]Enteric Diseases Program, Bacteriology and Enteric Diseases Program, National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, MB R3E 3R2, Canada

[2]Department of Medical Microbiology, 510 Medical Sciences Building, University of Manitoba, 730 William Avenue, Winnipeg, MB R3T 2N2, Canada

[3]Laboratory for Foodborne Zoonoses, Public Health Agency of Canada, Guelph, ON N1G 3W4, Canada

[4]School of Pharmaceutical Sciences, Rhodes University, Grahamstown 6140, South Africa

The serotyping of O and H antigens is an important first step in the characterization of *Salmonella enterica*. However, serotyping has become increasingly technically demanding and expensive to perform. We have therefore sequenced additional *S. enterica* O antigen gene clusters to provide information for the development of DNA-based serotyping methods. Three *S. enterica* isolates had O antigen gene clusters with homology to the *Escherichia coli* O123 O antigen region. O antigen clusters from two serogroup O58 *S. enterica* strains had approximately 85 % identity with the *E. coli* O123 O antigen region over their entire length, suggesting that these *Salmonella* and *E. coli* O antigen regions evolved from a common ancestor. The O antigen cluster of a *Salmonella* serogroup O41 isolate had a lower level of identity with *E. coli* O123 over only part of its O antigen DNA cluster sequence, suggesting a different and more complex evolution of this gene cluster than those in the O58 strains. A large part of the *Salmonella* O41 O antigen DNA cluster had very close identity with the O antigen cluster of an O62 strain. This region of DNA homology included the *wzx* and *wzy* genes. Therefore, molecular serotyping tests using only the O41 or O62 *wzx* and *wzy* genes would not differentiate between the two serogroups. The *E. coli* O123 O-antigenic polysaccharide and its repeating unit were characterized, and the chemical structure for *E. coli* O123 was entirely consistent with the O antigen gene cluster sequences of *E. coli* O123 and the *Salmonella* O58 isolates. An understanding of both the genetic and structural composition of *Salmonella* and *E. coli* O antigens is necessary for the development of novel molecular methods for serotyping these organisms.

Received 6 October 2008
Accepted 23 March 2009

†Present address: Centre for Infectious Disease, Institute of Cell and Molecular Science, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, 4 Newark Street, London E1 2AT, UK.

Abbreviations: 2D, two-dimensional; CDC, Centers for Disease Control and Prevention; COSY, correlated spectroscopy; DEPT, distortionless enhanced polarization transfer; HMBC, heteronuclear multiple bond correlation; HMQC, heteronuclear multiple quantum correlation; HOHAHA, homonuclear Hartman–Hahn; NML, National Microbiology Laboratory; NOESY, nuclear Overhauser effect spectroscopy; PS, polysaccharide.

The GenBank/EMBL/DDBJ accession numbers for the *Salmonella enterica* isolates S897 (serotype 58:l,z_{13},z_{28}:1,5), S1669 (serotype 58:r:z_{53},z_{47}:z_{73}) and 07-0209 (serotype 41:z_4,z_{23}:-) sequences are EU825756, EU825757 and EU825758, respectively.

A sequence alignment figure and tables of HMBC spectroscopy and NOESY data are available as supplementary material with the online version of this paper.

INTRODUCTION

Serotyping of the oligosaccharide (O) antigen and the flagellar (H) antigens of *Salmonella* spp. is a key step in the characterization of these organisms (Threlfall & Frost, 1990). The comprehensive database of *Salmonella* serotypes (Lindberg & Le Minor, 1984) is almost 80 years old; as a consequence, the significance of *Salmonella* serotypes is well understood by physicians, epidemiologists, and laboratory scientists and technicians. Serogroup classifications often indicate genetic relatedness (Porwollik *et al.*, 2004). Because of the high cost and labour involved in serotyping, it is usually performed fully only by reference laboratories. Consequently, it is necessary to develop molecular methods that will allow the rapid and inexpensive determination of the most common of the more than 2500 *Salmonella* serotypes (Fitzgerald *et al.*, 2007).

The O antigen of Gram-negative bacteria is the highly variable, surface exposed, component of the LPS. It consists of repeated 3–6 monosaccharide units, with the variation in O antigens resulting from differences in the composition of the monosaccharide units and sugar linkages. Variations among O antigens provide the structural basis for both the *Salmonella* Kauffmann–White (Lindberg & Le Minor, 1984) and *Escherichia coli* O antigen serotyping schemes (Ørskov *et al.*, 1977). There are currently 46 O antigens recognized in the 2579 serovars comprising *Salmonella* (Grimont & Weill, 2007) and 180 types of O antigen in *E. coli* (Ballmer *et al.*, 2007).

The genes required for synthesis of the O antigen in *Salmonella enterica* and many *E. coli* strains are found in a cluster between the *galF* and *gnd* genes on the chromosome (Samuel & Reeves, 2003). The protein products of the genes within O antigen clusters can be generally divided into three groups: (i) those required for the synthesis of the sugars, (ii) those involved in the transfer and modification of the O units, and (iii) those necessary for the polymerization and transport of the O units (Samuel & Reeves, 2003). The sugar biosynthetic genes are relatively homogeneous among *S. enterica* strains (Xiang *et al.*, 1993), in contrast to the highly heterogeneous transferase/flippase and polymerization genes encoded by the *wzx* and *wzy* loci. This heterogeneity can be used as the basis for the development of novel molecular serotyping methods (Ballmer *et al.*, 2007).

In the present study, the DNA sequence of the O antigen region from three *S. enterica* strains was determined as the first step in developing probes for molecular serotyping. Two of these strains were O58 serogroup and one was serogroup O41. All three *S. enterica* O antigen clusters shared DNA sequence similarity with the *E. coli* O123 serogroup O antigen region. In complementary research we determined the structure of the *E. coli* O123 O-polysaccharide (PS) and showed that it was consistent with the DNA sequence of the corresponding O antigen DNA clusters.

METHODS

Strains and culture conditions. *S. enterica* isolates S897 and S1669 were from the Enterics Strain Collection at the National Microbiology Laboratory (NML), Winnipeg, MB, Canada. Isolate S897 was originally obtained from Centers for Disease Control and Prevention (CDC), Atlanta, GA, USA, in 1964 (CDC isolate number 923). S1669 was originally obtained from the Institut Pasteur, Paris, France, in 2004 (Pasteur Institute isolate number 7095-89). S897 was subspecies II, serotype 58 : l,z_{13},z_{28}:1,5, while isolate S1669 was subspecies IIIb, serotype 58 : r : z_{53},z_{47} : z_{73}. *S. enterica* IIIa O41 (07-0209) had the full serotype designation 41 : z_4,z_{23} : -, and was isolated from a reptile (type unknown) in 2007. The *E. coli* O123 : H- strain abc was obtained from Dr Fleming Scheutz (Statens Serum Institut, Copenhagen, Denmark).

PCR and cloning of O antigen regions. The O antigen gene clusters between the JUMPStart sequence (Hobbs & Reeves, 1994) and *gnd* from isolates S897, S1669 and 08-0209 were amplified by long-range PCR using primers 412 and 482 (Li & Reeves, 2000) with an Expand long range dNTPack kit (Roche Diagnostics) according to the methods of the manufacturer. Template DNA was prepared in plugs using the PulseNet standard protocol for the preparation of *Salmonella* genomic DNA for PFGE (Ribot *et al.*, 2006), but replacing SeaKem Gold agarose with low-melting-point agarose. The proportion of DMSO used in each PCR was optimized to 5 % (v/v). The amplification conditions were: 92 °C for 2 min; 10 cycles of 92 °C for 10 s, 65 °C for 15 s and 68 °C for 15 min; 20 cycles of 92 °C for 10 s, 65 °C for 15 s and 68 °C for 15 min plus an additional 20 s added to each cycle; and a final extension at 68 °C for 7 min. Following PCR amplification, amplicons were visualized in 1 % agarose (Invitrogen) gels after staining with 0.005 % ethidium bromide (Fisher Scientific). Amplicons from several PCRs were pooled and sheared in a nebulizer (Invitrogen) for 3 min at 20 p.s.i. to obtain fragments of approximately 500 bp to 4 kb long. The pooled fragments were purified using Montage PCR centrifugal filter devices (Millipore) and cloned into the pCR4-TOPO vector using the TOPO TA cloning kit according to the manufacturer's instructions (Invitrogen). Transformants were screened on Luria–Bertani agar plates containing 100 μg ampicillin ml^{-1} and 100 μl commercially prepared X-Gal–IPTG solution (USB) for the identification of clones that carried a *Salmonella* DNA insert in the vector. DNA was isolated from positive clones by the 'boiled cell' method (Sambrook & Russell, 2001) and *Salmonella* DNA inserted into pCR4-TOPO was amplified by PCR using the FastStart *Taq* DNA polymerase kit (Roche Diagnostics) with primers M13 (5′-GTAAAACGACGGCCAGT-3′) and T7 (5′-GTAATACGACTCACTATAG-3′) specific to plasmid sequences flanking the insertion site. The amplification conditions were: 94 °C for 5 min; 35 cycles of 94 °C for 30 s, 50 °C for 30 s and 72 °C for 45 s; followed by a final extension at 72 °C for 5 min. Amplicons were visualized in agarose gels as above and purified at the DNA Core Facility at NML using the AMPure PCR purification system (Agencourt Bioscience). The amplicons were then sequenced as described below using the M13 and T7 primers.

DNA sequencing and annotation of O antigen regions. DNA sequencing was performed at the DNA Core Facility at the NML using BigDye terminator 3.1 cycle sequencing kits (Applied Biosystems) according to the manufacturer's instructions. DNA sequence data were generated using either an ABI 3100 or a 3730 DNA analyser (Applied Biosystems). Lasergene DNASTAR (DNASTAR) and Kodon (Applied Maths) software was used for editing, and annotation of DNA sequences. BLAST searches for homologues were conducted using the NCBI nucleotide collection (nr/nt) database.

Serotyping. Serotyping of *S. enterica* isolates S1669, S897, 07-0209 and *E. coli* O123 was performed by bacterial agglutination assay on

glass slides at the NML using *Salmonella*- and *E. coli*-specific antisera prepared in house, and as described by Ewing (1986), LeMinor (1997), LeMinor & Popoff (1997) and Ørskov & Ørskov (1984).

Isolation of *E. coli* O123 LPS and O-PS. *E. coli* O123 was grown on trays of Mueller–Hinton agar at 37 °C for 48 h and crude LPS was extracted using a modification of the hot phenol technique of Westphal & Jann (1965). The LPS was further purified by ultracentrifugation at 105 000 ***g*** for 16 h in a Beckman model L8-80M ultracentrifuge. The pellet was collected, resuspended in 50 ml H_2O, and centrifuged a second time, yielding 924 mg from approximately 20 g dry weight of cells. A portion of the pellet (214 mg) was suspended in 1 % acetic acid (100 ml) and heated at 90 °C for 90 min. The solid material was removed by centrifugation and the supernatant was dialysed against running water (3500 M_r cut-off) for 24 h and then lyophilized (yield 128 mg). This material was then purified by gel permeation chromatography on a Sephacryl S-200 SF column (90 × 2.6 cm; Pharmacia) irrigated with 0.1 M sodium acetate (pH 5.0, 20 ml h^{-1}). Column effluents were monitored using a differential refractometer (Waters). The O-PS fraction was lyophilized (yield 75 mg). A portion of the PS (25 mg) was O-deacetylated by treatment with 12 % aqueous ammonia for 3 h at 60 °C. The ammonia was removed under vacuum and the resulting O-deacetylated PS was designated PS-1.

Component analyses and methylation. The PS was hydrolysed with 2 M trifluoroacetic acid at 121 °C for 1.5 h. After reduction with $NaBH_4$ and acetylation with acetic anhydride/pyridine (1 : 1 v/v, 100 °C, 1 h) the samples were analysed by GLC. This was performed on a Hewlett Packard model 6890 gas chromatograph fitted with a flame-ionization detector and using helium as carrier gas. A fused-silica DB-17 bonded-phase capillary column (30 m × 0.25 mm, film thickness 0.25 mm; J&W Scientific) was used for separating alditol acetates and partially methylated alditol acetates (programme I: 180 °C for 2 min, then 3 °C min^{-1} to 240 °C), and acetylated octylglycosides (programme II: 180 °C for 2 min, then 2 °C min^{-1} to 240 °C). The identification of all the derivatives was determined by comparison with authentic standards and confirmed by GLC-MS on a Hewlett Packard 5988A instrument, using the appropriate column. Spectra were recorded at 70 eV and an ion-source temperature of 200 °C.

The absolute configuration of the sugars was determined essentially as devised by Leontein *et al.* (1978). A hydrolysate of the PS was re-*N*-acetylated with a mixture of acetic anhydride in methanol and pyridine (1 : 5 : 1 v/v, room temperature, 30 min) and then treated with (−)-2-octanol and acetylated. The chirality of the alanine and 3-hydroxybutyric acid was determined similarly by GLC of the trifluoroacetylated (−)-2-octyl esters. These derivatives were prepared by the treatment of a re-*N*-acetylated hydrolysate of the PS (2 mg) with (−)-2-octanol (0.5 ml) in the presence of concentrated (12.85 M) trifluoroacetic acid (100 µl) for 16 h at 130 °C, followed by trifluoroacetylation with trifluoroacetic anhydride and pyridine (2 : 1 v/v, 25 °C, 6 h). The PS was methylated according to a modified Hakamori procedure (Phillips & Fraser, 1981) using potassium dimethylsulfinylmethanide in DMSO. The methylated PS was hydrolysed and converted to alditol acetates as above.

NMR spectroscopy. Samples were deuterium-exchanged by lyophilizing several times from D_2O and then examined as solutions in 99.99 % D_2O containing a trace of acetone as an internal standard (δ 2.230 for ^1H and 31.07 p.p.m. for ^{13}C). Spectra were recorded at 30 °C on a Bruker AMX-400 spectrometer equipped with an X32 computer using standard UXNMR software and pulse programmes. Data were processed using Bruker XWINNMR software and spectra were overlaid for ease of interpretation using the PRONTO programme (Kjaer *et al.*, 1994). Signals were assigned using one-dimensional ^1H and ^{13}C experiments, including ^{13}C DEPT and two-dimensional (2D) experiments such as ^1H-^1H correlated spectroscopy (COSY) (Bax & Freeman, 1981), homonuclear Hartman–Hahn (HOHAHA) spectroscopy with two different mixing times (Bax & Davis, 1985), and ^1H-^{13}C heteronuclear multiple quantum correlation (HMQC) spectroscopy (Bax & Subramanian, 1986; Bax & Summers, 1986). The sequence was determined using nuclear Overhauser effect spectroscopy (NOESY) (Baumann *et al.*, 1981) and heteronuclear multiple bond correlation (HMBC) experiments. The mixing times for the HOHAHA spectroscopy were 45 and 90 ms, and for the NOESY was 200 ms. A delay of 60 ms was used for the evolution of long-range couplings in the HMBC experiment.

RESULTS AND DISCUSSION

Salmonella O58 and *E. coli* O123 O-antigen biosynthetic clusters are highly related

S. enterica isolates S1669 (subsp. IIIb 58 : r : z_{53},z_{74} : z_{73}) and S897 (subsp. II 58 : l,z_{13},z_{28} : 1,5) were both serovar O58 isolates, though their genetic backgrounds differed. DNA sequences of the O antigen regions of these two isolates were extremely similar, differing only by 143 nucleotide substitutions (0.9 %) along the entire 15 808 bp of overlapping sequence (Fig. 1). Genetic differences between the O antigen DNA sequences of isolates S897 and S1669 included a region of 14 single nucleotide polymorphisms (nt 217 to 257) in the *rmlB* gene that resulted in a difference of 9 amino acids in the proteins from these two isolates (Fig. 1). The O antigen DNA sequence of S1669 contained a 6 nt insertion in the *wfbF* gene, corresponding to position 10 346 of the aligned S1669 and S897 O antigen sequence. Other than these two regions, there did not appear to be any clustering of nucleotide differences in the O antigen sequences of the two isolates. It is not clear how the O58 O antigen cluster came to reside in both subspecies II (S897) and IIIb (S1669) *S. enterica* isolates. These two subspecies are quite well separated phylogenetically using MLST comparing five housekeeping or seven invasion genes (Boyd *et al.*, 1996). The O58 O antigen cluster may have been transferred from the ancestor of one of the strains to the ancestor of the other, or the two strains may have diverged while maintaining almost identical O antigen gene clusters.

The O antigen gene clusters from both *Salmonella* O58 isolates showed approximately 85 % identity in BLASTN comparisons with the homologous clusters in two *E. coli* O123 isolates, 43W (GenBank accession number DQ676933.1) and CB9827 (GenBank accession number DQ676934.1). This was very similar to the mean identity of *Salmonella* and *E. coli* housekeeping genes in general (Sharp, 1991), and is most consistent with the evolution of the cluster from a common ancestor of both *Salmonella* O58 and *E. coli* O123. In addition, the order of genes was strictly conserved in the O antigen sequences from all the isolates (Table 1). This is also consistent with evolution from a common ancestor. Comparison of *S. enterica* O58 and *E. coli* O123 O antigen clusters did not indicate clusters

Fig. 1. Comparison of sequence differences between the O antigen gene clusters of the two *Salmonella* O58 strains, S897 and S1669. Numbers above nucleotides represent the position of the nucleotide in the sequence, beginning with position 41 and ending with position 15 678. Boxes highlight the region of higher variability between nt 217 and 257, as well as the 6 nt insertion within S1669 corresponding to nt 10 321–10 326 of the consensus sequence. The program used to create this figure arbitrarily assigned one of each pair of divergent nucleotides as the consensus.

Table 1. Properties of the O58 antigen sequences from *Salmonella* isolates S897 and S1669

ORF	Gene name	Position in S1669 O antigen sequence	Position in S897 O antigen sequence	Length S1669/ S897 (bp)	Mol% G+C S1669	Mol% G+C S897	Amino acids different between S1669 and S897	Percentage amino acid identity with *E. coli* O123	
								S1669	S897
1	*rmlB*	101–1171	132–1202	1171	46.59	45.85	9	87	87
2	*rmlA*	1175–2044	1206–2075	870	36.32	36.32	3	95	95
3	*wzx*	2170–3621	2201–3652	1452	32.85	32.71	4	89	89
4	*vioA*	3626–4756	3657–4787	1131	32.63	32.80	4	89	89
5	*wfbA*	4753–5460	4784–5491	708	28.95	28.39	3	85	85
6	*wfbB*	5464–6441	5495–6472	978	30.47	30.37	3	80	80
7	*wfbC*	6438–7004	6469–7035	567	31.75	32.45	1	94	94
8	*wfbD*	7004–7402	7035–7433	399	35.59	36.09	4	90	90
9	*wfbE*	7480–8499	7511–8530	1020	34.90	35.00	3	82	81
10	*wzy*	8483–9745	8514–9779	1263/1266	31.75	31.60	7	82	81
11	*wfbF*	9738–10862	9772–10890	1125/1119	29.16	29.22	11	81	80
12	*fnlA*	10876–11907	10904–11935	1032	37.31	37.40	4	96	96
13	*qnlA*	11900–12769	11928–12797	870	32.41	32.53	1	89	89
14	*qnlB*	12744–13901	12772–13929	1158	34.97	34.97	3	93	92
15	*wbwH*	13901–15094	13929–15122	1194	32.75	32.91	2	80	79
16	*wbuC*	15102–15611	15130–15639	510	37.25	37.25	1	85	86

of nucleotide substitutions or amino acid differences; nucleotide differences appeared to be dispersed fairly evenly along the sequence. The sequence identity between both the O58 clusters and the O123 cluster ranged from 79 to 96 % local identity (Table 1; Supplementary Fig. S1 available with the online journal). These results were consistent with divergence of the *Salmonella* O58 and *E. coli* O123 O antigen regions by mutation, as opposed to recombination. Similar conclusions were drawn for the evolution of the *rmlA*, *rmlB* and *rmlD* genes of the dTDP-L-rhamnose pathway (Li & Reeves, 2000).

Strong identity has been found between the *S. enterica* and *E. coli wzx* (89 %) and *wzy* (81–82 %) genes. PCR assays have previously been developed to identify the *E. coli* O123 *wzx* and *wzy* genes (Beutin *et al.*, 2007). Use of these *E. coli* primer sets also produced amplicons of the correct size from the two *Salmonella* O58 isolates (data not shown). Very different *wzx* and *wzy* genes are found in strains with different O antigens (Samuel & Reeves, 2003). The fact that the *wzx* and *wzy* genes were very similar in both *E. coli* and the *S. enterica* serovar O58 isolates indicated that the oligosaccharide structures recognized by the products of these genes were very similar in both genera of bacteria. Both *Salmonella* O58 isolates were found to produce strong agglutination in *E. coli* O123 antiserum. An *E. coli* O123 isolate agglutinated strongly in *Salmonella* O58 antiserum. These results confirmed that the *E. coli* O123 and *Salmonella* O58 O antigen oligosaccharides are identical or very similar.

The gene content of the *S. enterica* O58 and *E. coli* O123 O antigen gene clusters allowed a tentative assignment of some of the sugars that may comprise the O antigen oligosaccharide (Beutin *et al.*, 2007; Samuel & Reeves, 2003). In the absence of the *rmlC* and *rmlD* genes, RmlA and RmlB should instead convert glucose-1-phosphate to dTDP-6-deoxy-D-xylo-4-hexulose, which could then be used as a substrate by FnlA, QnlA and QnlB to produce *N*-acetyl-L-quinovosamine and/or by VioA to produce viosamine (4-amino-4,6-dideoxy-D-glucose; Fig. 2). Genes encoding enzymes for both pathways are present in all three O antigen clusters (Table 1). L-Quinovosamine and viosamine, with additional modifications, were found in the *E. coli* O123 O antigen oligosaccharide (see below). The pathways for the production of both sugars utilize the same substrate (Fig. 2); these data suggest that there is no regulation of the two pathways except through the availability of the substrate. Of the other genes in the cluster, the genes identified as *wfbE* and *wfbF* have previously been classified as glycosyltransferases, *wfbC* as an acetyltransferase, and *wbwH* as an L-*N*-acetyl-quinovosamine transferase (Beutin *et al.*, 2007). The acetyltransferase is likely involved in the transfer of the 3-hydroxybutyramido to the L-Qui*p*NAcyl residue. The authors tentatively identified the function of the *wbfA* gene product as an amino acid (glycine) transferase. This gene product is likely responsible for the transfer of alanine to viosamine, resulting in the synthesis of D-Qui*p*4NAlaAc. The *wfbD* gene had homology with protein family PF01575, which contains monoamine oxidase C hydratase, and other hydratases and dehydratases. Products of the remaining two genes were designated either non-functional or having currently unknown functions (Beutin *et al.*, 2007). Overall, the content of the O antigen gene cluster is consistent with the tetrasaccharide composition of the O antigen repeat unit (see below).

Recently a group of *E. coli* with O serogroups O17, O44, O73, O77 and O106 were found to share a common four-sugar O antigen backbone unit with *Salmonella* O6 and O14 (Fitzgerald *et al.*, 2003; Wang *et al.*, 2007). These strains also possessed almost identical O antigen gene clusters. The authors hypothesized that the common O antigen backbone might be important for pathogenicity of the organisms or for survival in the environment (Wang *et al.*, 2007). It is clear that the very high homology between *S. enterica* O58 and *E. coli* O123 O-antigen gene clusters is not unique.

Salmonella O41 and O62 O antigen gene clusters share regions of high identity that have homology with the *E. coli* O123 O antigen gene cluster

A *Salmonella* IIIa O41 isolate, 07-0209, was also found upon DNA sequencing to contain a region with homology

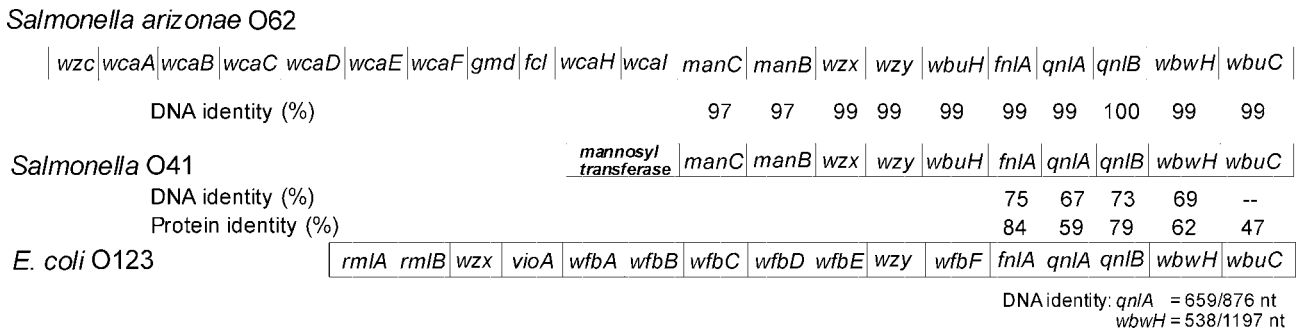

Fig. 2. Pathways for synthesis of the *Salmonella* O58 O antigen based on genes in the O antigen cluster.

to the *E. coli* O123 O antigen sequence (Table 2, Fig. 3). The O antigen DNA sequence of this isolate was also homologous over part of its length with the sequence comprising (approximately) genes SARI 00778 to SARI 00803 (*gnd*) of a *S. enterica* subsp. *arizonae* group IIIa expressing serotype O62 : z_4,z_{23} : – (SARC 5, CDC 346-86, RSK2980, GenBank sequence accession number CP000880). The SARC 5 sequence was identified in a BLAST search using the *Salmonella* O42 O antigen gene cluster, and was characterized as the O antigen DNA cluster of SARC 5 by the presence of the *gnd* gene, though *galF* did not appear to be present. Some of the SARC 5 O antigen genes were not present in isolate 07-0209, which had a mannosyltransferase as the first gene in the O antigen cluster instead. There may have been recombination between a DNA fragment carrying the mannosyltransferase gene and an M286-like *manC* gene into an O antigen region similar to the SARC 5 strain, removing genes associated with colonic acid biosynthesis but retaining all subsequent genes. The very low mol% G+C of the mannosyltransferase gene and its homology to *Yersinia pestis* genes (Table 2), compared with the very high mol% G+C of the *manC* and *manB* genes, support this interpretation of the data. The second gene in the 07-0209 O antigen cluster was *manC*. This gene showed a 93 % DNA identity with the *manC* gene of *S. enterica* subsp. I O41 strain M286 over 1433 nt; it was 97 % identical to a region encompassing 1068 nt of the SARC 5 strain. Only a portion of the total sequence, comprising the *fnlA* to *wbuC* genes, was homologous to the *E. coli* O123 sequence in both 07-0209 and SARC 5, and the degree of homology exhibited by strain 07-0209 was much less than that seen with isolates S1669 and S8987. Each gene within this region was most closely related to the homologous gene in SARC 5, but had a different level of homology to the *E. coli* O123 O antigen sequence. However, the fact that these genes had 31–35 mol% G+C and showed the same overlap or close spacing as the *Salmonella* O58 genes supported the hypothesis that they were acquired as a unit and have diverged subsequently.

The O antigen oligosaccharide structure of a *S. enterica* subsp. *arizonae* O62 isolate has been determined (Vinogradov *et al.*, 1994). It was found to consist of a backbone containing four L-rhamnose residues plus one each of residues β-D-GlcNAc and α-D-GalANAc. No mannose was detected in spite of the fact that the only O62 O antigen gene cluster for which the DNA sequence is available has homologues of the *manB* and *manC* genes (GenBank accession number CP000880). However, the O62 cluster also contains a homologue of the *gmd* gene; Gmd catalyses the conversion of D-mannose into a sugar that then serves as a precursor for the synthesis of other sugars, including rhamnose (Samuel & Reeves, 2003). The D-mannose is likely to be present in the serogroup O41 oligosaccharide, as evidenced by the absence of *gmd* and the presence of a gene encoding a protein with homology to a mannosyltransferase. Consequently, the serogroup O41 oligosaccharide is also predicted to lack rhamnose. As noted above and in Fig. 2, the genes within the cluster exhibiting homology to genes of the *E. coli* O123 O antigen cluster are responsible for the production of L-QuiNAc. However, *S. enterica* O41 does not have homologues of either the *rmlA* and *rmlB* genes or the *gmd* gene associated with production of the dTDP-6-deoxy-D-xylo-4-hexulose intermediate that serves as the substrate for production of L-QuiNAc. It is not clear whether this sugar would actually be synthesized. The serogroup O62 O antigen does not contain L-QuiNAc (Vinogradov *et al.*, 1994), so it appears

Table 2. Properties of the O antigen sequence from *Salmonella* O41 isolate 07-0209

ORF	Position in sequence	Mol% G+C	Similar protein	Percentage DNA identity with SARC 5	Next closest homologue	Percentage protein identity with next closest homologue	GenBank accession no.
1	52–1197	28.45	Mannosyl-transferase	NF	*Y. pestis* CO 92*	41	NP-406588.1
2	1343–2779	55.39	*manC*	97	*Salmonella* O41 strain M286	97	AAG41701.1
3	2924–4291	60.31	*manB*	97	*S. enterica* subsp. *arizonae*	97	AAG41714.1
4	4288–5574	31.78	*wzx*	99	NF	NF	–
5	5630–6781	28.21	*wzy*	99	*Geobacter* sp.FRC-32	34	EAT62104.1
6	6855–7949	29.77	*wbuH*	99	*E. coli* O4	60	AAT85654.1
7	7962–8993	35.47	*fnlA*	99	*E. coli* O123	84	ABG81792.1
8	8986–9861	31.74	*qnlA*	99	*E. coli* O123	59	ABG81793.1
9	9845–10990	31.86	*qnlB*	100	*E. coli* O123	79	ABG81794.1
10	10990–12186	34.42	*wbwH*	99	*E. coli* O123	62	ABG81795.1
11	12210–12707	32.93	*wbuC*	99	*E. coli* O123	47	ABG81796.1

NF, Not found.

*Seventeen other *Y. pestis* isolates had an identical or very similar gene encoding mannosyltransferase.

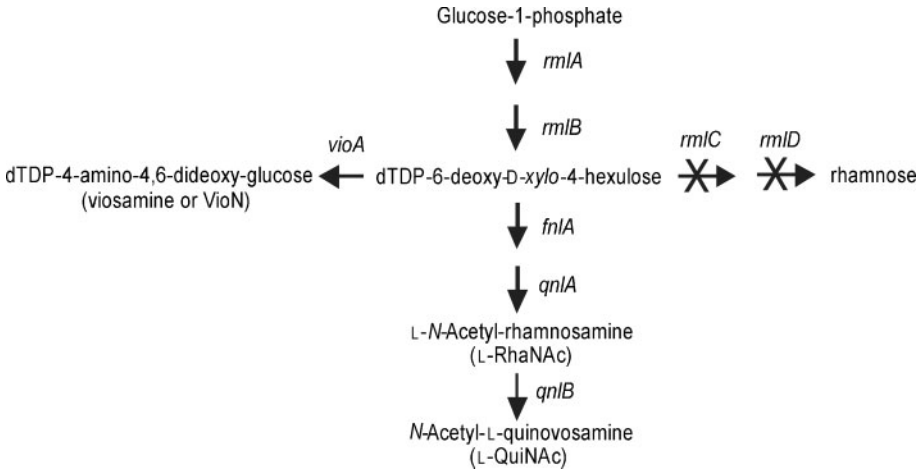

Fig. 3. Schematic comparison of the O antigen clusters of *S. enterica* O41, *S. enterica* O62 and *E. coli* O123.

possible that the biosynthetic pathways determined by the homologues of the *fnlA*, *qnlA* and *qnlB* genes are not active in strains of this serogroup in spite of the fact that dTDP-6-deoxy-D-xylo-4-hexulose is likely produced by the activity of the *gmd* gene product. The serogroup O62 O antigen cluster also contains a homologue of the *fcl* gene, which utilizes dTDP-6-deoxy-D-xylo-4-hexulose to produce GDP-L-fucose (Samuel & Reeves, 2003). This sugar is also not present in the serogroup O62 O antigenic oligosaccharide (Vinogradov *et al.*, 1994). It is clear that the pathways that lead to synthesis of the O antigenic oligosaccharide are complex and that the structure of this oligosaccharide cannot always be accurately predicted from the genes present in the O antigen gene cluster. Structural and chemical studies will be required for a better understanding of the composition of the O41 O antigen oligosaccharide.

S. enterica isolate 07-0109 produces the O41 antigen when tested serologically. However, the putative *wzx* and *wzy* were almost identical to the *wzx* and *wzy* genes of the O62 SARC 5 strain, though they appeared to be otherwise unique (Table 2). PCR assays, Luminex probes or other oligonucleotide reagents designed to detect these genes in the O41 serotype will also detect them in the O62 serovar. To unambiguously distinguish these two serovars using molecular methods it will be necessary to provide primers or probes for additional genes within the O antigen clusters of one or both serovars. The only gene available in O41 strains would be the mannosyltransferase; any of the non-homologous O62 genes could be used, with a preference for those that are unique.

Serotyping demonstrated no cross-reactivity between the S. *enterica* O41 and O62 strains. There was also no reactivity of these strains with *E. coli* O123 antiserum, nor of *E. coli* O123 isolates with the *Salmonella* O41- and O62-specific antiserum. Analysis of the O123 O antigen oligosaccharide was required to elucidate the molecular basis for these observations.

Determination of the structure of the *E. coli* O123 O antigen oligosaccharide

^{1}H and ^{13}C NMR analysis of the PS fraction. Mild acid hydrolysis of the LPS isolated from *E. coli* O123 bacteria with 2 % acetic acid yielded a PS, which was purified by gel permeation chromatography on Sephacryl S-200. The ^{1}H NMR spectrum was complex and showed signals of different intensities owing to partial O-acetylation. The PS was thus O-deacetylated with aqueous ammonia to give PS-1 whose ^{1}H NMR spectrum was typical of a regular polymer and showed four signals in the anomeric region (δ 4.5–5.5) at δ 5.03 ($J_{1,2}$ 3.2 Hz), 4.99 ($J_{1,2}$ 2.4 Hz), 4.95 ($J_{1,2}$ 3.6 Hz) and 4.46 ($J_{1,2}$ 7.8 Hz), indicating a tetrasaccharide repeating unit. The spectrum showed signals for three CH_3 groups of *N*-acetyl substituents (δ 1.96, 1.95 and 2.04), two for C-6 of deoxyhexoses (δ 1.25 and 1.18), and signals at δ 1.26, 1.23 and 2.42, which were assigned to alanine (Zubkov *et al.*, 1995) and 3-hydroxybutyryl (Knirel *et al.*, 1986) substituents. The ^{13}C spectrum confirmed the concept of a tetrasaccharide repeating unit and showed, *inter alia*, signals for five carbons bearing nitrogen, one for an unlinked C-6, five signals for C=O, three for CH_3 of *N*-acetyl groups and a signal at 45.22 p.p.m. for C-2 of a 3-hydroxybutyryl group. The ^{13}C-DEPT spectrum indicated that the signals at 60.98, 68.52 and 45.22 belonged to CH_2 groups.

Composition and methylation analysis. Acid hydrolysis of the PS and analysis of the sugars by GLC-MS gave 2-amino-2-deoxy-glucose (GlcN), 2-amino-2,6-dideoxy-glucose (quinovosamine, QuiN), 2,6-dideoxy-2-(3-hydroxybutyramido)hexose and 4-(*N*-alanyl)amino-4,6-dideoxy-hexose in the ratios 22 : 10 : 1:4. The presence of alanine (7.3 %) in the hydrolysate was shown by amino acid analysis and confirmed by GLC-MS. In order to ascertain the positions of substitution of the sugars the PS was methylated and the derived partially methylated alditol

acetates were analysed by GLC-MS. The following sugar derivatives were detected: 2-deoxy-2-(*N*-methylacetamido)-3,4-di-O-methylglucose, 2-deoxy-2-(*N*-methylacetamido)-4,6-di-O-methylglucose and 2,6-dideoxy-2-(*N*-methylacetamido)-4-O-methylhexose. The fourth sugar was not detected, probably due to degradation. The results above indicate that the polymer contains 6-substituted GlcN, 3-substituted GlcN and 3-substituted QuiN.

2D NMR spectroscopy. The identities of the 6-deoxyaminohexoses, identification of the anomeric configurations of the sugars, confirmation of the positions of substitution and the sequence of the sugars in the repeating unit of the O123 O-specific PS were obtained from 2D NMR studies on the O-deacetylated material PS-1. The sugar residues were labelled a–d in order of the decreasing chemical shift of their anomeric protons. In the COSY experiment the resonances for H-1 through to H-4 could be traced for residues a and c, for H-1 through to H-5 for b, and for H-1 through to H-3 for d. The entire spin systems for residues c and d (H-1 to H-6) were clearly seen in the respective H-1 and H-6 tracks of the HOHAHA spectrum and indicated that these were the 6-deoxy sugars. For units a and b, H-1 to H-4 could be traced in the respective H-1 tracks of the HOHAHA spectrum. The assignment of the ^{13}C signals came from analysis of the ^{1}H-^{13}C correlations in the HMQC spectrum. The cross-peaks observed in the HMBC spectrum for C-5/H-3 of a and C-6/H-4 of b then allowed the remaining sets of C-H signals for these residues to be assigned. Thus residue a is the 6-substituted sugar and residue b is 3-substituted. The spin systems for the alanyl and 3-hydroxybutyryl substituents could be traced in the COSY and HOHAHA spectra, and the ^{13}C signals then followed from the HMQC spectrum. The assignments were confirmed by the two- and three-bond correlations observed in the HMBC spectrum. The assignments are shown in Table 3 and the HMBC data in Supplementary Table S1 available with the online journal. The chemical shifts indicate that all sugar components are in the pyranose conformation and the sugars can be identified by comparison of the chemical shifts with those of model compounds (Bock & Thøgersen, 1982; Knirel *et al.*, 1986, 1988, 1993; Lipkind *et al.*, 1988). The linkage positions of the sugar residues were clearly identified from the large downfield shifts of the carbons involved in the glycosidic linkages. The results from the hydrolysis, methylation analysis and NMR spectroscopy indicate that the PS is

Table 3. NMR data for *E. coli* O123 PS-1

The data show the chemical shifts in p.p.m. relative to acetone at δ 2.23 for ^{1}H and 31.07 for ^{13}C, at 30 °C. Additional signals were assigned to CH_3 of NAc at δ 1.96, 1.95 and 2.04, and 23.02 (2C) and 22.96 p.p.m., and to CO of NAc at 173.97, 174.12 and 174.98 p.p.m. Linkage carbons are indicated in bold.

Residue	Repeating unit		Hydrogen or carbon						
			1	2	3	4	5	6a	6b
a	→6)-α-D-GlcNAc	H	5.03	3.91	3.71	4.24	3.70	4.08	3.90
		C	97.98	54.37	71.82	71.28	69.89	**68.52**	
		^{3}J	3.2	10.5					
b	→3)-α-D-GlcNAc	H	4.99	4.04	3.87	3.59	4.25	3.82	3.80
		C	97.29	54.06	**75.44**	68.64	72.35	60.98	
		^{3}J	2.4	10.4					
				(−1.1)*	(+3.7)	(−2.6)			
c	→3)-α-L-QuiNAcyl	H	4.95	4.07	3.79	3.28	4.26	1.25	
		C	97.29	54.50	**76.33**	74.48	68.46	17.32	
		^{3}J	3.6	10.6	9.6	9.3	6.6		
				(−1.1)	(+4.6)	(−2.4)			
d	→3)-β-D-Qui4NAlaAc	H	4.46	3.46	3.79	3.80	3.55	1.18	
		C	103.35	73.24	**76.81**	57.63	72.01	17.34	
		^{3}J	7.8	10.0			6.2		
				(−2.7)	(+2.3)	(−0.5)			
e	(*S*)-3-hydroxybutyryl	H		2.42	4.18	1.23			
		C	174.50	45.22	65.89	22.80			
		^{3}J		6.8		6.3			
f	L-AlaNAc	H		4.22	1.26				
		C	175.21	50.56	16.88				
		^{3}J			7.1				

^{3}J, coupling constant (Hz).

*$\Delta\delta$, Values in parentheses are α and β glycosylation shifts.

composed of a tetrasaccharide repeating unit containing 6-substituted *N*-acetylglucosamine [GlcNAc (a)], 3-substituted *N*-acetylglucosamine [GlcNAc (b)], 3-substituted 2,6-dideoxy-2-(3-hydroxybutyramido)-glucose [QuiNAcyl (c)] and 3-substituted 4-(*N*-acetylalanyl)amido-4,6-dideoxyglucose [Qui4NAlaAc (d)]. The sequence of the residues in the repeating unit was determined from the inter-residue nuclear Overhauser effects observed in the NOESY spectrum and the data are shown in Supplementary Table S2 available with the online journal. The three-bond correlation between H-3 of c and C-1 of a, and between H-1 of d and C-6 of a, in the HMBC spectrum confirmed the NOESY data and the following sequence can be written for the sugars in the repeating unit: 6a1-3c1-3b1-3d1. Confirmation of the location of the *N*-acetylalanyl group on the Qui4N residue came from observation of a nuclear Overhauser effect between H-1 of residue b and the CH_3 of the alanine (Zubkov *et al.*, 1995).

The determination of the absolute configurations of the sugars and the acyl substituents was achieved by a combination of chromatographic and spectroscopic techniques. The GlcNAc residues were found to have the D configuration by GLC analysis of the acetylated (−)-2-octyl glycosides derived from a hydrolysate of the PS according to the method of Leontein *et al.* (1978). The alanine and 3-hydroxybutyric acid in the hydrolysate were analysed by GLC as *N*-acetylated O-trifluoroacetylated (−)-2-octyl esters. The retention times were compared with those of derivatives of authentic optical isomers and in this way the alanine was found to be an L isomer and the 3-hydroxybutyric acid to be an *S* isomer. The configuration of the aminodeoxy sugars was determined spectroscopically by an analysis of the β-effects of glycosylation on the chemical shifts of the carbons adjacent to the linkage carbons (Kochetkov *et al.*, 1984; Lipkind *et al.*, 1988). The C-4 resonance of unit b was shifted upfield by 2.4 p.p.m. by the glycosylation of c at C-3 (shifts relative to GlcNAc, Table 3) (Lipkind *et al.*, 1988). This relatively large value indicates that the glycosylating sugar must have a different absolute configuration, i.e. unit c must be an L isomer. Conversely C-2 of unit b has only a small negative β-shift ($\Delta\delta$ −1.1) whereas a larger one would be expected for glycosylation by a D sugar. The same magnitudes for the β-effects of glycosylation are seen for unit c (shifts compared to QuiNAc) (Knirel *et al.*, 1993) confirming glycosylation by a sugar of different absolute configuration i.e. unit a (determined as D, above). The glycosylation shifts for unit d are opposite to those for b and c, i.e. a large negative value for C-2 and a small negative value for C-4 (shifts compared to Qui4NAc) (Knirel *et al.*, 1988), indicating that the glycosylating sugar has the same configuration, i.e. D. Thus residues a, b and d have the D configuration and unit c has the L configuration.

Once the spectra of the PS-1 were assigned it was a much simpler task to analyse the spectra of the O-deacetylated polymer PS. Most of the resonances were assigned using COSY, HOHAHA and HMQC experiments as before. The positions of the O-acetyl groups in the PS were determined by comparison of the chemical shifts of the PS with those of the PS-1. It was clear from the downfield shifts of H-6 ($\Delta\delta$ 0.55) and H-5 ($\Delta\delta$ 0.20) of residue b (→3)-α-D-Glc*p*NAc) that the site of O-acetylation was at O-6 of this residue. The ^{13}C shifts confirmed this finding with the resonance for C-6 of residue b in the PS being deshielded by 2.74 p.p.m. and that for C-5 showing an upfield shift of 2.06 p.p.m. The amount of O-acetylation was estimated at 30 % from the 1D ^{1}H spectrum.

The above data allow the structure to be written for the repeating unit of *E. coli* O123 O-specific PS as shown in Fig. 4. The structure of the repeating unit is unique and contains some dideoxysugars not previously found in the *E. coli* series of PSs. Like most of the O-PSs in this group it is neutral. L-Alanine has only been reported in the PS of one other strain of *E. coli* viz. O167 where it is amidically linked to C-6 of galacturonic acid (Linnerborg *et al.*, 1997). QuiNAc in the form of its L isomer is not very common in bacterial PSs, and has been reported in the O-PSs of *Pseudomonas fluorescens* biovar A (Knirel *et al.*, 1993), of several *Proteus* strains (Shashkov *et al.*, 1998; Vinogradov *et al.*, 1991; Ziolkowski *et al.*, 1997) and of *Shigella boydii* type 13 (Dmitriev *et al.*, 1973), and in the capsular PS of *Bacteroides fragilis* (Kasper *et al.*, 1983). D-Qui4N carrying the *N*-acetyl-L-alanyl substituent has only been reported once before, in the capsular PS from *Alteromonas* sp. CMM 155 (Zubkov *et al.*, 1995). A related compound carrying an L-seryl substituent was identified in the LPS of *E. coli* O114 (Dmitriev *et al.*, 1983).

The two *S. enterica* O58 isolates, the O41 isolate and the O62 sequenced strain carry all or part of the O antigen gene cluster also found in *E. coli* O123. The known oligosaccharide structure of *E. coli* O123 is consistent with the DNA sequence of this isolate and the two *S. enterica* O58 isolates. Some of the sugars that may be expected to be

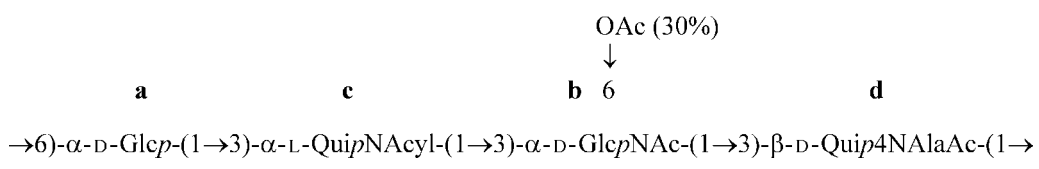

Fig. 4. Structure of the repeating unit of the *E. coli* O123 O-antigenic PS. L-QuiN, 2-amino-2,6-dideoxy-L-glucose; D-Qui4N, 4-amino-4,6-dideoxy-D-glucose; acyl, (*S*)-3-hydroxybutyryl; AlaAc, *N*-acetyl-L-alanyl.

present in the *S. enterica* O41 O antigen oligosaccharide from the presence of genes homologous to the *E. coli* O123 O antigen gene cluster were not actually present. A detailed knowledge of both O antigen gene clusters and O antigen saccharide structure is required for the appropriate interpretation and use of any DNA-based molecular serotyping method using *wzx* and *wzy* genes.

ACKNOWLEDGEMENTS

We acknowledge the contributions of Lorea Baker, Katherine Brown and Kristine Cruz of the Identification and Serotyping Laboratory, Enteric Diseases Program, Bacteriology and Enteric Diseases Program, NML, for performing all the serological assays. We wish to thank Brynn Kaplen, Kimberly Melnychuk, Shari Tyson and Travis Murphy of the DNA Core Facility, NML, for DNA sequencing and oligonucleotide synthesis. We acknowledge the contributions of Gary van Domselaar, Head of the Bioinformatics group at NML, who created for us a program that provided invaluable assistance in the creation of the Supplementary Fig. S1. We offer many thanks to Matt Gilmour for a critical reading of the manuscript and suggestions for revisions. Finally, we acknowledge the suggestions, advice and technical support of Patti Fields and Collette Fitzgerald, especially in choosing strains to sequence and in methodology for amplification of O antigen regions.

REFERENCES

Ballmer, K., Korczak, B. M., Kuhnert, P., Slickers, P., Ehricht, R. & Hächler, H. (2007). Fast DNA serotyping of *Escherichia coli* by use of an oligonucleotide microarray. *J Clin Microbiol* **45**, 370–379.

Baumann, R., Wider, G., Ernst, R. R. & Wüthrich, K. (1981). Improvement of 2D NOE and 2D correlated spectra by symmetrization. *J Magn Reson* **44**, 402–406.

Bax, A. & Davis, D. G. (1985). MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J Magn Reson* **65**, 355–360.

Bax, A. & Freeman, R. (1981). Investigation of complex networks of spin-spin coupling by two-dimensional NMR. *J Magn Reson* **44**, 542–561.

Bax, A. & Subramanian, S. (1986). Sensitivity-enhanced two-dimensional heteronuclear shift correlation NMR spectroscopy. *J Magn Reson* **67**, 565–569.

Bax, A. & Summers, M. F. (1986). ^{1}H and ^{13}C assignments from sensitivity-enhanced detection of heteronuclear multiple-bond connectivity by 2D multiple quantum NMR. *J Am Chem Soc* **108**, 2093–2094.

Beutin, L., Wang, Q., Naumann, D., Han, W., Krause, G., Leomil, L., Wang, L. & Feng, L. (2007). Relationship between O-antigen subtypes, bacterial surface structures and O-antigen gene clusters in *Escherichia coli* O123 strains carrying genes for Shiga toxin and intimin. *J Med Microbiol* **56**, 177–184.

Bock, K. & Thøgersen, H. (1982). Nuclear magnetic resonance spectroscopy in the study of mono- and oligosaccharides. *Annu Rep NMR Spectrosc* **13**, 1–57.

Boyd, E. F., Wang, F. S., Whittam, T. S. & Selander, R. K. (1996). Molecular genetic relationships of the salmonellae. *Appl Environ Microbiol* **62**, 804–808.

Dmitriev, B. A., Backinovsky, L. V., Kochetkov, N. K. & Khomenko, N. A. (1973). Immunochemical studies on *Shigella boydii* lipopolysaccharides. *Eur J Biochem* **34**, 513–518.

Dmitriev, B. A., Lvov, V., Tochtamysheva, N. V., Shashkov, A. S., Kochetkov, N. K., Jann, B. & Jann, K. (1983). Cell-wall lipopolysaccharide of *Escherichia coli* 0114 : H2. Structure of the polysaccharide chain. *Eur J Biochem* **134**, 517–521.

Ewing, W. H. (1986). *Edwards & Ewing's Identification of Enterobacteriaceae*, 4th edn. New York: Elsevier.

Fitzgerald, C., Sherwood, R., Gheesling, L., Brenner, F. W. & Fields, P. I. (2003). Molecular analysis of the *rfb* O antigen gene cluster of *Salmonella enterica* serogroup O : 6,14 and development of a serogroup-specific PCR assay. *Appl Environ Microbiol* **69**, 6099–6105.

Fitzgerald, C., Collins, M., van Duyne, S., Mikoleit, M., Brown, T. & Fields, P. (2007). Multiplex, bead-based suspension array for molecular determination of common *Salmonella* serogroups. *J Clin Microbiol* **45**, 3323–3334.

Grimont, P. A. D. & Weill, F.-X. (2007). *Antigenic Formulae of the Salmonella Serovars*, 9th edn. Paris: Institut Pasteur.

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

Kasper, D. L., Weintraub, A., Lindberg, A. A. & Lönngren, J. (1983). Capsular polysaccharides and lipopolysaccharides from two *Bacteroides fragilis* reference strains: chemical and immunochemical characterization. *J Bacteriol* **153**, 991–997.

Kjær, M., Andersen, K. V. & Poulsen, F. M. (1994). Automated and semiautomated analysis of heteronuclear and multidimensional nuclear magnetic resonance spectra of proteins: the program PRONTO. *Methods Enzymol* **239**, 288–308.

Knirel, Y. A., Vinogradov, E. V., Shashkov, A. S., Dmitriev, B. A., Kochetkov, N. K., Stanislavsky, E. S. & Mashilova, G. M. (1986). Somatic antigens of *Pseudomonas aeruginosa*. The structure of O-specific polysaccharide chains of *P. aeruginosa* O10 (Lányi) lipopolysaccharide. *Eur J Biochem* **157**, 129–138.

Knirel, Y. A., Dashunin, V. V., Shashkov, A. S., Kochetkov, N. K., Dmitriev, B. A. & Hofman, L. L. (1988). Somatic antigens of *Shigella*: structure of the O-specific polysaccharide chain of the *Shigella dysentery* type 7 lipopolysaccharide. *Carbohydr Res* **179**, 51–60.

Knirel, Y. A., Paramonov, N. A., Shashkov, A. S., Kochetkov, N. K., Zdorovenko, G. M., Veremeychenko, S. N. & Zakharova, I. Y. (1993). Somatic antigens of pseudomonads: structure of the O-specific polysaccharide of *Pseudomonas fluorescens* biovar A strain IMV 1152. *Carbohydr Res* **243**, 205–210.

Kochetkov, N. K., Chizhov, O. S. & Shashkov, A. S. (1984). Dependence of ^{13}C chemical shifts on the spatial interaction of protons, and its application in structural and conformational studies of oligo- and poly-saccharides. *Carbohydr Res* **133**, 173–185.

LeMinor, L. (1997). *Guidelines for the Preparation of Salmonella antisera*. Paris: World Health Organization Collaborating Centre for Reference and Research on Salmonella, Institut Pasteur.

LeMinor, L. & Popoff, M. Y. (1997). *Antigenic Formulas of the Salmonella Serovars*, 7th edn. Paris: World Health Organization Collaborating Centre for Reference and Research on Salmonella, Institut Pasteur.

Leontein, K., Lindberg, B. & Lönngren, J. (1978). Assignment of absolute configuration of sugars by g.l.c. of their acetylated glycosides formed from chiral alcohols. *Carbohydr Res* **62**, 359–362.

Li, Q. & Reeves, P. R. (2000). Genetic variation of the dTDP-L-rhamnose pathway genes in *Salmonella enterica*. *Microbiology* **146**, 2291–2307.

Lindberg, A. A. & Le Minor, L. (1984). Serology of *Salmonella*. *Methods Microbiol* **15**, 1–141.

Linnerborg, M., Wollin, R. & Widmalm, G. (1997). Structural studies of the O-antigenic polysaccharide from *Escherichia coli* O167. *Eur J Biochem* **246**, 565–573.

Lipkind, G. M., Shashkov, A. S., Knirel, Y. A., Vinogradov, E. V. & Kochetkov, N. K. (1988). A computer-assisted structural analysis of regular polysaccharides on the basis of ^{13}C-NMR data. *Carbohydr Res* **175**, 59–75.

Ørskov, F. & Ørskov, I. (1984). Serotyping of *Escherichia coli*. *Methods Microbiol* **14**, 44–112.

Ørskov, I., Ørskov, F., Jann, B. & Jann, K. (1977). Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*. *Bacteriol Rev* **41**, 667–710.

Phillips, L. R. & Fraser, B. A. (1981). Methylation of carbohydrates with dimsyl potassium in dimethyl sulfoxide. *Carbohydr Res* **90**, 149–152.

Porwollik, S., Boyd, E. F., Choy, C., Cheng, P., Florea, L., Proctor, E. & McClelland, M. (2004). Characterization of *Salmonella enterica* subspecies I genovars by use of microarrays. *J Bacteriol* **186**, 5883–5898.

Ribot, E. M., Fair, M. A., Gautom, R., Cameron, D. N., Hunter, S. B., Swaminathan, B. & Barrett, T. J. (2006). Standardization of pulsed-field gel electrophoresis protocols for the subtyping of *Escherichia coli* O157 : H7, *Salmonella*, and *Shigella* for PulseNet. *Foodborne Pathog Dis* **3**, 59–67.

Sambrook, J. & Russell, D. W. (2001). *Molecular Cloning*, 3rd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Samuel, G. & Reeves, P. (2003). Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr Res* **338**, 2503–2519.

Sharp, P. M. (1991). Determinants of DNA sequence divergence between *Escherichia coli* and *Salmonella typhimurium*: codon usage, map position, and concerted evolution. *J Mol Evol* **33**, 23–33.

Shashkov, A. S., Arbatsky, N. P., Widmalm, G., Knirel, Y. A., Zych, K. & Sidorczyk, Z. (1998). Structure and serological specificity of the O-specific polysaccharide of *Proteus penneri* strain 26. *Eur J Biochem* **253**, 730–733.

Threlfall, E. J. & Frost, J. A. (1990). The identification, typing and fingerprinting of *Salmonella*: laboratory aspects and epidemiological applications. *J Appl Bacteriol* **68**, 5–16.

Vinogradov, E. V., Kaca, W., Rozalski, A., Shashkov, A. S., Cedzynski, M., Knirel, Y. A. & Kochetkov, N. K. (1991). Structural and immunochemical studies of O-specific polysaccharide of *Proteus vulgaris* 5/43 belonging to OX19 group (O-variants). *Eur J Biochem* **200**, 195–201.

Vinogradov, E. V., Knirel, Y. A., Kochetkov, N. K., Schlecht, S. & Mayer, H. (1994). The structure of the O-specific polysaccharide of *Salmonella arizonae* O62. *Carbohydr Res* **253**, 101–110.

Wang, W., Perepelov, A. V., Feng, L., Shevelev, S. D., Wang, Q., Senchenkova, S. N., Han, W., Li, Y., Shashkov, A. S. & other authors (2007). A group of *Escherichia coli* and *Salmonella enterica* O antigens sharing a common backbone structure. *Microbiology* **153**, 2159–2167.

Westphal, O. & Jann, K. (1965). Bacterial lipopolysaccharide. Extraction with phenol-water and further applications of the procedure. In *Methods in Carbohydrate Chemistry*, pp. 83–91. Edited by R. Wistler. New York: Academic Press.

Xiang, S.-H., Haase, A. M. & Reeves, P. R. (1993). Variation of the *rfb* gene clusters in *Salmonella enterica*. *J Bacteriol* **175**, 4877–4884.

Ziolkowski, A., Shashkov, A. S., Swierzko, A., Senchenkova, S. N., Toukach, F. V., Cedzynski, M., Amano, K.-I., Kaca, W. & Knirel, Y. A. (1997). Structures of the O-antigens of *Proteus* bacilli belonging to OX group (serogroups O1–O3) used in Weil–Felix test. *FEBS Lett* **411**, 221–224.

Zubkov, V. A., Nazarenko, E. L., Gorshkova, R. P., Ivanova, E. P., Shashkov, A. S., Knirel, Y. A., Paramonov, N. A. & Ovodov, Y. S. (1995). Structure of the capsular polysaccharide from *Alteromonas sp.* CMM 155. *Carbohydr Res* **275**, 147–154.