Sequence and structural analysis of the *rfb* (O antigen) gene cluster from a group C1 *Salmonella enterica* strain

SANG J. LEE,[†] LAJWANT K. ROMANA[‡] and PETER R. REEVES^{*}

Department of Microbiology, The University of Sydney, NSW 2006, Australia

(Received 15 January 1992; revised 22 May 1992; accepted 8 June 1992)

The *rfb* (O antigen) gene cluster of a group C1 *Salmonella enterica* strain was sequenced; it comprised seven open reading frames which precisely replaced the 16 open reading frames of a group B strain. Two genes of the mannose biosynthetic pathway were present: *rfbK* (phosphomannomutase) had a G + C content of 0.61 and had only 40% identity to *rfbK* of group B but was very similar to *cpsG* of the capsular polysaccharide pathway with 96% identity, whereas *rfbM* [guanosine diphosphomannose (GDP-Man) pyrophosphorylase] had a G + C content of 0.39. Other genes had G + C contents ranging from 0.24 to 0.28. *rfbM*(C1) and *rfbM*(B) had 60% identity, which is much less than expected within a species, but nonetheless indicates a much more recent common ancestor than for *rfbK*. The other genes showed much lower or no similarity to *rfb* genes of other *S. enterica* strains. It appears that the gene cluster evolved outside of *Salmonella* in a species with low G + C content: the *rfbM* gene presumably derives from that period whereas the *rfbK* gene appears to have arisen after transfer of the cluster to *S. enterica* by duplication of the *S. enterica cpsG* gene, presumably replacing an *rfbK* gene of low G + C content.

Introduction

The O antigen of Gram-negative bacteria is a cell surface polysaccharide which comprises many repeats of an oligosaccharide. This repeat unit varies among strains and in *Salmonella enterica* there are about 60 different forms (Ewing, 1986) which have been used to divide the species serologically into groups A, B, C1 etc. In group C1 strains the repeat unit contains four mannose residues, one *N*-acetylglucosamine residue and a glucose branch (Lindberg *et al.*, 1988). The genes responsible for biosynthesis and assembly of the O antigen repeat unit are located within the *rfb* gene cluster which maps at 42 min on the *S. enterica* chromosome (Sanderson & Roth, 1989). The genetics of the *rfb* region has been reviewed by Mäkelä & Stocker (1984), and more recently we have cloned the *rfb* regions from groups B (Brahmbhatt *et al.*, 1988; Jiang *et al.*, 1991; Wyk & Reeves,

1989), A and D (Liu *et al.*, 1991; Verma *et al.*, 1988; Verma & Reeves, 1989), C1 (Lee *et al.*, 1992), C2 (Brown *et al.*, 1991, 1992) and E1 (Wang *et al.*, 1992).

In *S. enterica* of groups A, B and D polymerization of the repeat units to give complete O antigen is dependent on the *rfe* gene which maps at 32 min (Mäkelä & Stocker, 1984; Sanderson & Roth, 1989). The O antigens of *E. coli* groups O8 and O9 have been shown to be synthesized by single sugar chain extension and not by synthesis and subsequent polymerization of an oligomeric O unit. It has been suggested that group C1 *S. enterica* resembles groups O8 and O9 *E. coli* in this regard as all three have in common dependence on a functional *rfe* gene (Jann & Jann, 1984).

In group C2 the *rfe* gene is within the *rfb* region (Brown *et al.*, 1991, 1992); in group E1 the polymerization function is again within the *rfb* cluster and a potential *rfe* gene has been identified (Wang *et al.*, 1992). Some studies have indicated that in group C1 the polymerization function is located within the *rfb* region (Mäkelä, 1966); however, if group C1 O antigen synthesis, like that of *E. coli* groups O8 and O9, is by single sugar extension, then there will not be an O antigen polymerase at all.

In our report on the cloning of the group C1 *rfb* gene cluster we localized within the cluster the *rfbM* and *rfbK* genes (Lee *et al.*, 1992). The *rfb* region of group C1 showed no significant similarity on Southern hybridization to the *rfb* regions of groups A, B, D, C2 or E1.

* Author for correspondence. Tel. 2 692 2536; fax 2 692 4571.

[†] Present address: Department of Biochemistry, University of Medicine & Dentistry of New Jersey, Piscataway, New Jersey 08854-5635, USA.

[‡] Present address: Department of Biochemistry, University of Queensland, Royal Brisbane Hospital, Herston, Queensland 4029, Australia.

The nucleotide sequence data reported in this paper have been submitted to GenBank and have been assigned the accession number M84642.

However, one of the mannose pathway genes (*rfbK*) hybridized strongly to the *cps* region of strain LT2. To further understand the evolutionary origin of the *rfb* gene cluster of group C1, we have sequenced the whole region, studied its expression in *E. coli* K12 and compared the sequence with those of other *rfb* gene cluster sequences.

Methods

Media, antibiotics, DNA techniques, strains, plasmids and bacteriophages. Media, antibiotics and DNA techniques were reported previously (Lee *et al.*, 1992), except for 2 × YT broth (1.6%, w/v, Bactotryptone, 1%, w/v, yeast extract, 0.5% NaCl) used for M13 phage propagation. *S. enterica* strain M40 was the source of all *rfb* region DNA used in the study. All *E. coli* strains used were derivatives of strain K12: NM522 (Gough & Murray, 1983) was used as host for pGEM-7zf(+) and pT7T3 19U plasmids; JM101 (Messing *et al.*, 1981) was used for propagation of M13mp19; C600 (Maniatis *et al.*, 1982) and Sφ874 [F⁻ *trp lacZ rpsL thi upp Δ(his-rfb)*] (Neuhard & Thomasen, 1976) was used for study of gene expression. All plasmids made in this study are shown in Figs 1 or 3. Bacteriophages M13mp19 and M13KO7 were supplied by Boehringer Mannheim and Pharmacia respectively.

Serology. The presence of O antigen was detected by slide agglutination using antibody to antigen O7, which is one of the group C1 antigens (Ewing, 1986).

Sequencing. The clones used for sequencing the *rfb* region of strain M40 are shown in Fig. 1. The vector used was pT7T3 19U or pGEM-7zf(+) and nested deletion sets were generated with *ExoIII* and *S1* nuclease (Henikoff, 1987); selected deletion derivatives were sequenced from packaged cosmid obtained using M13KO7 as helper

phage (Vieira & Messing, 1987). The sequencing direction was determined by the orientation of the insert in the original clone and is indicated in Fig. 1.

All the cloned fragments overlapped other fragments except at the *XhoI* site at position 7160; the *BglII-PstI* (6726–8631) fragment was cloned into M13mp19 for sequencing across this site using single stranded DNA prepared from phage.

DNA sequencing was done by the dideoxy chain termination method (Sanger *et al.*, 1977, 1980) using dye-labelled M13 standard primer (Promega), or other oligonucleotide primers used together with dye-labelled dideoxy terminators. The methods in both cases were as described by Applied Biosystems for the 373A automatic sequencer, which was used to analyse the products. Specific primers for dye dideoxy termination sequencing were synthesized on an Applied Biosystems DNA synthesizer.

Computer analysis. Computer analysis of sequence data was done on the MBIS system at the CSIRO Biotechnology Laboratory, Sydney, Australia (Bucholtz & Reisner, 1986) or (after June 1990) on the Australian National Genomic Information Service (ANGIS) at Sydney University: both systems incorporate several sets of programs.

Sequence data was assembled using the *SAP* program of Staden (1982, 1986). We used the program *NIP* (Staden, 1984) to calculate G + C content, molecular weight and a hydrophobicity index. The *RNY* preference method was used to look for open reading frames (Shepherd, 1981). The programs *ALOM* (GCG package; Devereux *et al.*, 1984) based on the method of Klein *et al.* (1985) and *PepPlot* (GCG package) based on the method of Kyte & Doolittle (1982) were used to identify potential transmembrane segments, and the transmembrane segments were then further checked using the algorithm described by Eisenberg *et al.* (1984).

The programs *BESTFIT* and *GAP* in the GCG package were used to do sequence comparisons. The parameters for running *BESTFIT* were gap weight, 3.0; gap length weight, 0.10; the parameters for running *GAP* were gap weight, 3.0; gap length weight, 0.1. The program *CLUSTAL*

Fig. 1. Map of the *rfb* region of strain M40 and sequencing strategy. The restriction sites and the scale are as in Lee *et al.* (1992), but the precise map positions given here and elsewhere in this paper are based on the sequence data reported in this paper, as there were minor corrections made after sequencing. Distances from position 0 shown in kb. The ORFs inferred from the sequence are shown as boxes and the location of the *gnd* gene is also shown. The plasmids used for sequencing are shown with arrows showing direction of sequencing. The *PstI-PstI* (2.965–5.025), *PstI-PstI* (5.025–2.965), *XhoI-PstI* (7.153–5.025), *SacI-XhoI* (9.850–7.153), *SalI-PstI* (9.90–14.26) and *EcoRI-PstI* (10.316–8.847) fragments were cloned into pT7T3 19U and designated pPR1251, pPR1252, pPR1254, pPR1256, pPR1257 and pPR1259, respectively. The *SalI-XhoI* (4.590–7.153), *XhoI-EcoRI* (7.153–10.316) and *PvuII-SacI* (12.445–9.850) fragments were cloned into pGEM-7zf(+), and designated pPR1253, pPR1255 and pPR1258, respectively. Part of pPR1257 was not sequenced and is indicated by lighter shading. B, *BglII*; C, *ClaI*; E, *EcoRI*; P, *PstI*; Pv, *PvuII*; Sa, *SacI*; S, *SalI*; X, *XhoI*.

(Higgins & Sharp, 1989) provided a good multiple alignment program for either protein or nucleic acid sequences.

Promoter search programs were generously provided by Dr M. C. O'Neill: AUTOST22 used the criteria published by O'Neill & Chifari (1989), ALIGNIC2 searched for promoters on the basis of information content as described by O'Neill (1989a), and MOLBVH3 used the Berg-von Hippel modified information content function described by O'Neill (1989b). The program VONHIPIC was used to calculate the index for each presumptive promoter, as described by O'Neill (1989b). A training back-propagation neural networks program (O'Neill, 1991) to define the promoters was run for us by Dr O'Neill on his computer. Promoter positions are given as the first base of the potential mRNA.

Results

Sequencing the rfb region and identification of open reading frames

The *rfb* region of strain M40 (group C1) has been located previously (Lee *et al.*, 1992). The regions upstream and downstream of the *rfb* gene cluster have similar restriction maps to those of corresponding regions of strain LT2. The start position for our analysis of the *rfb* gene cluster of strain M40 is that used for strain LT2, as it is in the conserved region upstream of *rfb*. We sequenced from a *Pst*I site at position 2965 (present in both M40 and LT2) to position 13280 (Fig. 1), which is within the *gnd* gene downstream of the *rfb* gene cluster. The sequence is presented in Fig. 2.

The sequence from bases 2965 to 4026 was very similar to the corresponding sequence of strain LT2 (Jiang *et al.*, 1991) and the sequence from bases 13086 to 13280 was very similar to the LT2 sequence from positions 21 887 to 22 081. These regions flank the *rfb* gene cluster and had less than 2% difference in nucleotide sequence between the two groups B and C1. The 9059 bp sequence from positions 4027 to 13085 of M40 thus replaces the *rfb* cluster of strain LT2 and presumably is the *rfb* cluster of strain M40: it contains seven ORFs which were named *orf4-11*, *orf5-19*, *orf6-17*, *orf7-17*, *orf8-36*, *orf9-79* and *orf11-45* (Figs 1 and 2) based on their start position (kb) on the *rfb* map (e.g. *orf4-11* starts at approximately base 4100 on the M40 *rfb* map). All ORFs read in the same direction from *orf2-80* to *gnd* as in group B (Jiang *et al.*, 1991).

All the ORFs started with an ATG codon and finished with a TAA codon except for *orf7-17*, which ended with a TGA codon. The sites that are the best approximations for potential Shine-Dalgarno ribosome binding sites are shown in Fig. 3. There are non-coding sequences of 384, 14 and 285 bp, respectively, before the three ORFs *orf4-11*, *orf9-79* and *orf11-45*, whereas all other junctions have overlapping termination and initiation codons (Fig. 2 and Fig. 3). Overlaps of one base (as in *orf4-11/orf5-19*) and four bases (as in *orf7-17/orf8-36*) have already been

described for polycistronic operons (Kozak, 1983). The other three ORFs show more extensive overlaps of 17 bp (*orf5-19/orf6-17*) and 14 bp (*orf6-17/orf7-17*); this kind of overlap is rather common in bacterial virus genomes (Shaw *et al.*, 1978), but rarely found in bacterial genomes (Carlomagno *et al.*, 1988).

Complementation of O7 epitope expression in E. coli K12

We previously located the *rfb* gene cluster of strain M40 by looking for clones which conferred O antigen expression on an *rfb* delete *E. coli* strain (Lee *et al.*, 1992).

E. coli does not produce O antigen as it has a mutation in *rfb*. However some *rfb* genes remain and in order to determine if the *E. coli* K12 *rfb* region could complement plasmids missing part of the M40 *rfb* region, we tested strain C600 carrying various plasmids. As can be seen (Fig. 4) the region required for group C1 O antigen expression in the presence of the K12 *rfb* region could be narrowed down to the 4.58 kb regions between the *Sal*I site at base 4590 (left end of pPR1262) and the *Bgl*II site at base 9318 (right end of pPR1040). This is presumably the region required for group specificity. There are three complete ORFs (*orf5-19*, *orf6-17* and *orf7-17*) in this region (Fig. 2), and to determine if all were required for group C1 O antigen synthesis, several subclones were constructed, transformed into *E. coli* K12 strain C600 and tested for O antigen synthesis using O7 specific antiserum (Fig. 4).

To check the requirement for *orf5-19*, we constructed pPR1250 by inserting the *Sac*I-*Xba*I fragment (5498–8901) between the *Sac*I and *Xba*I sites of pT7T3 19U, thereby removing 313 bp of *orf5-19* with the remainder in frame from the polylinker. This clone did agglutinate with O7 antibody. The 14 bp *Eco*RI-*Sac*I polylinker region of pPR1250 was then replaced with the 40 bp *Eco*RI-*Sac*I polylinker of pYUM1118 (Futo *et al.*, 1989), so that *orf5-19* was read out of frame. C600 carrying the resultant plasmid pPR1289 did not agglutinate with the antibody, and as ORFs *orf6-17* and *orf7-17* have very good Shine-Dalgarno sequences and presumably were expressed we conclude both that *orf5-19* cannot be complemented by *E. coli* K12 and that *orf5-19* is functional even in the absence of the first 313 bp.

pPR1263 was constructed to remove *orf7-17* by cloning the *Sal*I-*Hind*III fragment (4590–7720) into the *Sal*I-*Sma*I site of pGEM-7zf(+). That this clone failed to confer O7 antigen synthesis on C600 shows that *orf7-17* is required. We were unable to construct a derivative lacking only *orf6-17*, but clearly, when complemented by the *rfb* region of *E. coli* K12, at least two and probably three genes are required for and are sufficient for expression of the group C1 O antigen as indicated by presence of epitope O7.

orf2.80
CTGCAGGGATCAAAAGAAATCGTACTGGTACTCACGCGTCTAAAAACGCCGTTGAGAACCACCTTCGACACCTCTTATGAACCTTGAATCACTTCTTGAGCAGCGCGTTAAGCGTCAGCTTT 3084
PstI G

TGGCGGAAGTGCAATCTATCTGCCCACCGGGCGTAACGATTATGAACGTGCGCCAGGCGCAGCGTTAGGGCTGGGGCAGCTCTATTCTGTGCGCGCTCCGGTTGTGGGCGATAACCCTTT 3204
G T T C

TCATTGTGGTACTCCCGGATATTATTATTTGATGATGCTACCGCGATCCGCTGCGCTATAACCTTGGCGCGATGGTGGCGCGTTTCAATGAAACGGGTGCGAGCCAGGTGCTGGCGAAAC 3324
C G

GCATGAAAGGTGATTATCCGGAGTATCCGTTATCCAGACGAAAGAACCTCTGGATAATGAAGGCAAGTCAGCCGGATTGTGGAGTTTATCGAAAAACCGGATCAGCCGAGACGCTGG 3444

ATTCCGATTTGATGGCGGTAGGCGCTTATGTGCTTTCAGCCGACATCTGGGCGAACTGGAAGAACCGAACCGGGCGCTGGGGCCGTATCCAGCTCACCGATGCCATTGCGAGAACTGG 3564
C T

CGAAAAACAGTCGGTTGACGCGATGCTAATGACGGGTGACAGCTATGACTGCGGTAAAAAATGGGCTACATGCAGGCATTGTGAAGTACGGGTGCGCAACCTGAAAGAGAGCGCA 3684
C

AGTTCCGTAAGAGCATAGACGAGCTTCTGCATGAATAAGTATTAAACAACCGTGATAAATGGTTGGTGATAACATAATAACGGCAGTGAACATTCGAAGCGCAAGTTGGCTGAAACGAG 3804
T

TGTTGACTGCCGTTTTAGTTTTGTATAAAGGGCTTAAGTAACAAGGGTTATCTGGAGCATTTTAACTGCTGATTTTATAAGATTAACTCTGTTTCCGGATGCAATTAATAAGACAATTA 3924
-35 -10

3928 end homology
GCGTTTAACTTTTGTAGTGGCTTTGCCCTGCTGGCGAGGTTTGTAAACAAATCGATATGACGAGTGCAGCTGATGAGCCAGGGGCGGTAGCGTGTCTGAATAATTATAATAT 4044
C G

Start orf4.11
M I Y Y I F I V I F P P F F S F V K N K

CGTATTTATAAGATTATTTTGTAGAAACAATAAGAAATTTTATTTGAGAGACTTAGCATGATATATTACATTTTATTGTAATTTTCCGTTCTTTTCATTGTTTAAAGATAAAA 4164

N I K I Y A L M L S F L F L V S F C S L R W Q T G T D W L P Y Y D D F M S P G N
ATATAAAATTTACGCTTGTATGCTAAGTTTTTATTCTTGTCTCGTTTGTCTTACGTTGCGAGACTGCTACTGATTGGCTCCCTATTACGATGATTTCATGAGTCCAGGGAATC 4284

R H D F E I G Y V L Y V K L I R Y L T D N Y T L F L F T T S I I P I A L I F W G
GACATGACTTTGAAATCGGGTATGTTATATACGTCAAATTAATAAGATATTGACAGATAATTATACATTGTTCTATTACTACATCAATTATCTATCGGTTAATATTCTGGGGAT 4404

C L K T Q K N I S L T I L S V C V F Y S Y Y Y L G S F F G A E R R I A I G L S
GCTGAAAACACAAAAAATATATTAATTAATCTGCTGTTTGTGTTTTTATAGTTATTATTAATCTGGGGCTTTTTTTTGGGGCCGAAAGAAATATTGCAATTGGGTTATCTT 4524

F F A L I Q Y K S N K K V Q S L I L I L C A S T F H I S S L V T L S V F L I N K
TTTTGCTTTAATCCAATATAAAAGTAATAAAGGTGCAATCGCTTATCTTATCTATGTCGCTGACATTTTCATATCTCGTCTTTAGTTACCTTATCTGTTTTTTTAACTCAATAAGT 4644

Sall
L S L N L Y K I L L V L G A I L S L P L S H Y L S D I I S S V I S L I P V E I V
TATCTTAAACTTATATAAAATATTACTTGTGCTTGGTGCCATCTATCATTGCCTTTGTCACATATTATTATCAGATATAAATTTCTCCGTATTAGTCTAATTCCTGTCGAAATTTGTC 4764

R Y K L T V Y T Q N A Q E Y G S I S I S G I L K R V V I S A I F I Y T L S F D I
GCTATAAAGTCTGTTTATCTCAAAATGCACAGAATATGTTCCATAAGTATATCTGGAATTTTAAAAAGGTGGTAATTTCTGCTATTTTATTACACATTATCTTTTGATATAA 4884

K N N K A N L F L V K T Y L F G T I I Y L F L S P I S A M F S V I S I Y F T I V
AAAATAAAGGCAAAATTTATTTTGGTAAAAACATATTATTGGAACATATCATATATTATTATTATCTCGGATCTCAGCCATGTTCTCAGTAATTAGTATATATTACTATAGTCG 5004

E I L L I P A V L V R V G I F T R I P A L I F I V I F Y F G Y Q V Y S I L G S Y
AAATATGTTGATCTGCTGAGTATTAGTGGCTGAGGAATTTTACACGGAATCTGCTCTGATTTTATAGTGATTTTATTTTGGCTATCAAGTTACTCTATAGTGGGCTTATC 5124

PstI **Start orf5.19**
P E L F Y P Y I S V F S E I O R Q G I Y * M L L I Y T G S Y P D D K C G V G D Y V
CTGAACCTTTCTATCTTACATATCGGTATTTCAGAAATTCAGCGTCAAGGTATTACTAATGTTATTATATATACGGGTTCTGATCCAGATGATAAATGTGGAGTAGGTGATTATGT 5224

Y N L N Q E I K K N Y T V N V V K L S L F E L I Y K I V S N R K I I K L I N I Q
TTATACTTAAACCAAGAAATTAAGAAATTAACAGTTAAGTGTCAAACTATCTTATTCGAATTAATCTATAAAATAGTTTCTAATAGAAAATTAATCAAACTTATTAATATCA 5364

Y P S I G F S T N K I A A F K P H V A F I L A K L V G L K T S I T L H E F S S L
ATATCCATCGATTGGATTTTCGACAAACAAATTCGCGCTTTTAAACCTCATGTGCAATTATCTTAGCTAAGTTAGTTGGCTTAAAAACATCGATAACGCTACACGAATTTAGTAGTTT 5484

S K R A Q Y F L K I F K L A D Y I I F T T Q Y E K N I G E K T L F N S A K T R L
ATCAAGAGAGCTCAATATTTCCTAAAAATTTTAAAGCTGCTGATTATATAATCTTTACAACCTCAGTATGAGAAGAATATAGGTGAAAAACATTATTATCTGCAAAACAGACT 5604

SacI
I P I A S N I P S C Q K K E K K F D L V H F G I I S K G K G I E E F I W I I K E
AATTCCTATCGCATCTAACATACCTCTTGGCAAAAGAAAGAAAAAATTTGATCTTGTTCATTTTGGTATAATTTCCAAAGGAAAGGGAATAGAAGAAATTTATCTGGATTATTAAAGA 5724

L N K K N I K F R S A L I G Y V P S A D N S Y A N L I E Q C T E Q K C E L L L
ACTTAATAAGAAAAATTAATAATTCGCTCTGCTTGTGTTATGTTCTAGTGTGACAAATGTTATGCAATTTAATTATAGAACAGTGTACAGAACAAAAAGTCCGAGTTACTTTT 5884

N K S A A E I S N L L A E T D M A I L P Y P D G I S E R R G T A L A A M I N R V
AAATAAATCAGCAGCGAAATATCAATCTGCTAGCTGAACTGACATGGCTATATTACCATATCCCGATGGCATCTCAGAACGGAGGAGTACTGCCCTTGTGCGATGATTATCGAGT 5964

L V F S L R G Q F S S E F E N I A V L G N D K N D L L S K V L Y Y I N N T D S F
ATTAGTATTTTCCCTAAGGGGCGAGTTTTCGCTGAAATTTGAAAAATATGCCGTATTAGGGAACGATAAAAAATGATCTTTTGTAGTAAAGTTTATATATATAAATAATACAGATAGCTT 6084

A K I N D S A Y E Y S E K R N W Q S I A Q L Y I G L I N N D S N K *
Start orf6.17 M I V I N N Y F S G V
TGCCAAATTAATGATAGTGCTATGAGTACTCTGAAAAAGAAATTTGGCAGTCGATCGCTCAACTGTATATAGGACTAATAAACAATGATAGTAATAAATAATTTATTAGTGGTGT 6204

L K R G I P I Y T E E L V L Q M K K D S M Q V C E L T C P K V L Y P L P A F I H
TAAAGAGGAATACCGATTATATCTGAAAGAGTTAGTTCTTCAATGAAAAAGGACTCTATGCAAGCTGTGAAATTAACCTGTGCAAAAGTCTTACCTTTTACCTGCTTTATACATA 6324

N F L F I F Y E O I L T P L I G L L L K S K F N I Y P Y N S T S I D A Y L G K
ATTTTTTATTCATATCTATGAACAATACTTACTCCGCTTATTGGACTTATATTAAAGTCAAAATTTAATATCTATCCATATAATAGTACATCAATTATAGACGCTACTTGGGTAAT 6444

S V V I I H D L I S L R K K N H S L S A K Y V S Y C M L K A S Q L K A D Y I Y I
CAGTGGTTATTATTCATGATCTAATAAGCTTAAGAAAAAGAACCTTCPTTATCCGCGAAATATGTGTCATATTGTATGCTCAAAGCATCACAATTAAGCCGATTATATATATAT 6564

S K T T K R V I D S I E L F K N C K G Y Y F P N T F F R F E E I A K K N T I L D
 C A A A A C A A C T A A G C G T G T T A T A G A T A G C A T T G A A C T C T T T A A A A A T T G T A A A G C G T A T T A T T C C C A A C A C A T T T T T C G T T C G A G A A A T T G C G A A A A A A A T A C G A T T C T T G A T C 6684
 L G Y I L D L V T G G V G D N K D L D G A L K L Y S I S K D E R L P L K I L G C G
 T T G A T A T A T T A T T G G T A A C A G G G T G G T G A T A A A A G A T C T C G A T G T G C T T T G A A A T T A T T A T C A T C T A T T A G C A A G G A T G A A C G A T T G C C G C T T A A A A T A T T A G C G T G T G G A A 6804
 BglII
 N A I E R V K K I I D D H R V Q S E I E V I P F L D L D D V V S L Y C N S T F I
 A T G C C A T A G A G A G G T A A A A A A A A T A A T T G A T G A C C A T A G A G T C C A A A G T G A A A T T G A A G T A T T C C T T T C T T G G A T C T T G A T G A T G T G T A A G C T T T A T T G T A A C T C A C C T T T A T T 6924
 W A H S K Y E G Y G R A V A E A K L S H K K I L C T Q I S A F M E Q K D E N V Y
 G G G C A C A T T C G A A A T A T G A G G T T A T G G A A G A G C T A G C A G A G C T A A G T T G C C A T A A A A A G A T T T A T G T A C A C A A A T A T C A G C C T T A T G G A A C A A A A G A T G A A A T G T A T A T C 7044
 L Y T N Q N D F H I Q Y K A V L A S K Y K D I H G Q L I E H E I F K S Q L E F L
 T T T A T A C T A A T C A A A A T G A C T T T C A C A T C C A A T A A A G C T G T G T T A G C T A G C A A G T A T A A G A T A T C C A T G G G C A A T T G A T A G A C A T G A A A T T T T A A A C C C A A C T C G A G T T T T T A T 7164Y
 K K * XhoI
 Start orf7.17
 M E K S K I L I L T P R F P Y P V V G G D R L R I Y R I C K E L S K Y Y T L D L
 A T G G A A A A A G T A A A A A T A C T C C T C A C C C A G A T T T C C G T A T C C A G T G G T A G G A G G G A T A G T T A A G A T A T A T C G C A T T T G T A A G G A A T T G T C T A A A T A C T A C A C A T T A G A T T G 7284
 L S L C D S I E D L N F I V K N D H V F D K I F R I Y H P K I K S Y F N V L K A
 C T A G T C T A T G C G A T A G T A T C G A G A C T T A A A T T T A T A G T A A A A A T A T C A G C G T C T T G A T A A A A T T T C C G G A T C T A T C A T C C T A A A A T A A A A G T A T T T T A A T G T A T T A A A A G C A 7404
 L P G R K P L Q I A Y Y K N T E F E N K L N E I I G N Y D L T L S H L I R V G D
 T T A C C T G G T A G A A A C C A C T T C A G A T A G C T T A T T A T A A A A A T A C T G A A T T T G A A A A T A A A C T T A A T G A G A T A A T T G G A A T T A T G A T T T G A C T C T T T C A T C A T C A T T C G T G T T G T G A T 7524
 Y T L N K P G L H I L E M T D A I S L N Y S R I K K E A P K N S L K S I I Y S I
 T A C A C T T T A A A T A A G C C G C T T G C A T A T T C T G A A T G A C T G A T G C A A T T C C T T A A A T T A T T C C G T A T A A A A A A A A G C T C G A A A A A T T C T T T A A A A G C A C A T T A C T A C T A T C 7644
 E Q E R L L K Y E K E V Y G R Y S L I S L I S E V D K K F L F G N R N D N I L V
 G A A C A A G A G C G A T T A T A A A A T A T G A A A A A A A G T T A T G C C G C T A T T C A C T A T A T C G T A A T A T C A G A G T T G A C A A A A A T C C T T T T T G G A A T A G G A A T G A T A A C A T A T A T T A C A A G T T T T G A A 7764
 C N N G V D L E D Y P F T K R V I E N T N I I N L I F I G N L C S F Q N F D G V
 T G T A A T A A T G G A G T G A T C T T G A G A T T A T C C T T T A C A A A G A G A G C A T T G A A A A T A C C A A T A T A A T A A T C T T A T T T T A T T G G A A C T T G T G T C A T T C C A A A A T T T T G A C G G A G T G 7884
 K W F V K N I L P S I N K N N N G K Q Y I F H I L G K I N K S D K L Y L Q G F E
 A A T T G G T T C G T A A A A A T A T T C T A C C A T C A A T A A A A A A T A A T A A C G T A A G C A A T A T A T T T T C C A C A T T T A G G T A A A A T A A C A A A A G T A T A A C A T A T A T T A C A A G T T T T G A A 8004
 N V V V S G S V T N M A D A A K I G H I G I C P V R F G A G V Q N K I L E Y M A
 A A T G T T G T G T G A T G C G A G C T T A C A A A T A T G C T G A T G C G G C T A A A A T A G G A C A T A T C G A A T T T G C C A G T T C G G T T T G T G C A G T G T A C A A A A T A A A A T A T A G A A T A T A T G C T 8124
 L G L P T I T S R M G Y E G I E A N I G E I L I A D N S D E Y L K S L E T L S
 C T T G A C T T C C A C A T A T C A T C C C G A A T G G G T A T A G A G G A A A T A T G T G A A G A G A T A T A A T T G C T G A T A A T A G T G A T G A C T T G A A G C G T T A G A A C T T T G A T 8244
 E N S V Y Q M I A K N A R N F A E K F N W S T R L S V L V K N I E R L T G K *
 Start rfbM
 G A A A A T C T G T A T A T C A A A T G A T T G C T A A G A A T G C G A A A T T T T G T C G G A G A A A T T A A T T G G T C A A C G C G T T A T C A G T T T T G T G A A A A A T A T G A G A G G T A A C A G G A A A A T G 8364
 I T P I I M A G G N G S R L W P L S R T L Y P K Q F L C L D G S Q S M L Q T T I
 T T A C C A C A A T A A T A T G G C T G G G G A A A T G G A A G T C G G T A T G G C A C T T T C T C G A A C G T A T A T C C T A A C A G T T T T A T G T T A G A T G C T C T C A A G A T A T G C T C A G A C C A C A A T A 8484
 T R I N N L N A S D P I V I C N E Q H R F I V A E Q L K E L S K S S G D I I L E
 C G C G A A T A A T A A T T G A A T G C T C A G A T C C T A T T G T A T T T G T A A T G A G C A G C A G G T T A T T G T T G C A A C A G C T T A A A G A A T A T C T A A A T A T C T G C G A T A T A A T T T A G A A C 8604
 P V G R N T A P A V A L A A L K L C K K N A L L V L A A D H I I K D E E T F C
 C T G T T G A C G T T A A C A G C G C C T T G C A G C T T A A A A T G C C T A A G A A C C C T T A C T G T T A G T A T T G C T G C T G A C C A T A T A A T A A G A C G A A G A A A C G A A A T T T T G T A 8724
 PstI
 K T I Q D A R K Y A E A G K L V T F G I V P T M P E T G Y G Y I R R G Q A L F S
 A G A C T A T A C A A G C G T A G A A A A T A T G C C A A G C A G A A A A T T A G T A C G T T T G G T A T T G T A C T A C C A T C C A G A C A G A G T T A T G T T A T A T A C G C G T G G C A A G C T T T T C T C T G 8844
 PstI
 A E N D C S L A F R V A E F V E K P N L E T A Q S Y L D S G E Y Y W N S G M F L F
 C A G A A A A T G A T T G T T C T T T G C G T T A G G T G G T G A A T T G T C G A A A A C C A A T C T A G A A A C G C T C A G A G T A T A T T G A T T C I G G T A G A T A T T A C T G G A C A C G G T A T G T T T T A T 8964
 F R A D R Y I E E L K I R P D I Y K A C S L A M E S A V T D L D F I R V D E D S
 T T A G G C A G A T C G A T A T A T T A A G A G T A A A A A T T G C C C T G A T A T T A T A A A G C T T G C T A T T A G C G A T G G A A T C T G C T T A C G G A T C T C G A T T T C A T T C A G T G G A T G A A G A T T C A T 9084
 F C A C P D E S I D Y A V M E K T N D A V V V P L N A G W S D V G S W S S L W E
 T T T G C G T T G C C G G A T G A A T C A A T T G A T T A T G C A G T A T G G A A A A G A C G A A T A C G C G G T T G C T A C C T T A A A T G C G G T T G G A T G A T G C G G T T C T G G T C T T C T T T G G G A G 9204
 I S D K D S N G N V T K G D V L S H N A D N C Y L H A E T G L V T A V G V K D L
 T A A G T A T A A G A T A T A A T G T A T G T A C T A A A G C G A T G T A A T A G C C A C A A T G C A G A T A A T T G C T A T C C C A T G C A G A A C C G G T T T G T A A C C G C T G T T G T G T C A A A G A T C T T A 9324
 BglII
 I V V Q T K D A V L V A N T N C V Q D V K K I V E K I K L E N R H E H I T H K E
 T A G T C G T G C A A C G A A A G A T G C C G T T T A G T T G C A A C A C A A C T G T T C A A G A T G T A A G A A A A T T G T G A C A A A T T A A A T T A G A A A A C G T A G C A T A T A A C T A C T C G G G A A G 9444
 V Y R P W G K A Y D S I D F G E R Y Q V K R I T V K A P G E G I S Q Q H Y H R A E
 T T T A T A G A C C A T G G G T A A A T A C G A T T C A A T C G A T T C G G T G A G C G T T A T C A G T A A A G A G A A T C A C I G T A A A A C C C G G A A G A A T T C A G A C A C A C A T A T C A T T C G C G G A A C 9564
 H W I I V A G T A K I T I K G E V K I L T E N E S V Y I P V G V K H C L E N P G
 A C T G G A T T A C G T T G C A G T A C C G A A A A A T A A C C A A A A G G T A A G T G A A A A T T T A A C T G A A A A T G A A T C C G T A T A T A T C T G T C G G T G T T A A G A C T G C C T T G A A A C C A G G G A 9684
 Start rfbK
 K I A L E L I E V R S G A Y L G E D D I V R F S D K Y G R N * M N N L T
 A A A T T G C A C T T G A A C T A T T A G A T A A G A T C C G C C A T A T T A G G G A G G A T G A T A T T G T T C G T T T T C A G A T A A G A T G T C G T A A T T A A G A A T A A C G T A A T A A T G A A T A A T T A A C 9804
 C F K A Y D I R G R L G E E L N E D I A W R I G R A Y G E Y L K P K T I V L G G
 T T G T T T A A G C C T A C G A C A T T C G C G C A G G C T G G C G A A G A G C T A A T G A A G A T A T C G C T G G C G A T T G C C G C C T A C G C G A A T A T T T A A A C C G A A A C C A T C G T G T G G C G G 9924
 SacI
 D V R L T S E A L K L A L A K G L Q D A G V D V L D I G M S G T E E I Y F A T F
 C A G C G C G C C T G A C A G C A G G C G T A A A G C T G G C G C T G G C A A A A G G T T A C A G A C G A C G C G G C C T C T G G A T A T G C G T A T G T C C G C A C C G A A G A G A T T A C T T C C G C A C C T T 10044
 SalI
 H L G V D G G I E V T A S H N P M D Y N G M K L V R E G A R P I S G D T G L R D
 C C A C C T C G C G T G A C G C G T A T T A G A G T A C C G C A G C C A T A A C C G A T G A C T A C A C G G T A G A A G C T G G T C G C G A G G C C C C G C C G A T A A G C G G C A C A C C G T C T G C C G A 10164

V Q R L A E A G D F P P V N D A A R G S Y R Q I S L R D A Y I D H L L A Y I S V
CGTTCAGCGTCTGGCGAAGCGGGCGATTTCGCCGCGTGAACGACGCCGCGCGGACGCTACAGACAAATCAGCCTGCGCGACGCTACATCGACCACCTGCTCGCTACATCAGTGT 10284

N N L T P L K L V V N S G N G A A G P V I D A I E A R L K A L G A P V E F I K I
TAACAATCTCACCCGCTGAAGCTGGTGGTGAATTCGGCAACGGGGCGGGCGGGCGGTCATTGACGCCATCGAAGCGGCTGAAGCGCTGGGGCGCGCGGTGGAGTTCATCAAAAT 10404

EcoRI
H N T P D G T F P N G I P N P L L P E C R D D T R K A V I E H G A D M G I A F D
CCACAACACCCCGGACGGCACCTTCCGCAACGGGATTCACCAACCGGTGCTGCGGAGTGTGCGGACGACACCCGCAAGGCGGTATTGAACACGGGGCGGACATGGGCATCGCCTTTGA 10524

G D F D R C F L F D E K G Q F I E G Y Y I V G L L A E A F L E K H P G A K I I H
CGCGACCTTCGACCGCTGCTTCCGTGTCATGAGAAAGGCGASTTTATCGAGGGGTACTACATCGTGGGCTGCTGGCGAAGCGTTTCTGGAAAAACACCCGGGGCGGAAGATTATCCA 10644

D P R L T W N T E A V V T A A G G T P V M S K T G H A F I K E R M R T E D A I Y
CGACCCGCGCTTACCTGGAAACGAGGACGTTGTTGACGGCGCGGGCGGAGCGCGGTGATGTCAAAAACCCGGGACGCGCTTTATCAAGGAGCGGATGCGTACGGAGGATGCCATTTA 10764

G G E M S A H H Y F R D F A Y C D S G M I P W L L V A E L V C L K G Q S L G E L
TGGCGGGAAATGAGCGCGCATCACTATTTCGTGATTTGCTTACTGTGACAGCGGAATGATCCCGTGGCTGCTGGTGGCGGAGCTGGTGTGCTGAAGGGACAGTGGTGGGTGAAGT 10884

V R D R M A A F P A S G E I N S R L A E P A A A I A R V E A H F A E E A Q A V D
GGTCCGCGACCGGATGGCGCGTTCGCCGCGAGCGGGAAATCAACAGCAGGCTGGCGAGCGCGGACGCGCCATCGCGCGGTGGAAGCGCACTTCGCGGAGAGGCGGACGCGGTGGA 11004

R T D G L S M S F A D W R F N L R S S N T E P V V R L N V E S R G D I P L M E A
CGCGACGGAGGACTCAGCATGCTGCTTCGCCGCTGCGCTGCTGCAACACCGAAGCGGTGGTGGCGCTGAATGTGGAATCAGCGGGCATATCCGCTGATGGAAGC 11124

R T K E I L Q L L N S *
GCGAACGAAAGAAATCTTCAGCTACTCAATTCATAAAAAGCAGCCCTCTTTATGGAGGGCTAATCTTTTCGGATATTCTCTAACCAATTCGTAGTATAAGGCCAATAAATATTAAGTGT 11244

—35 —10 11286 —35
GTTTGATTGTTGGTGGTTCCTATTCTAATAGATATAAGGCTTCGCTATGTTCTTTTCGACAGTATAAAATATCATGATGATTATTTTACCTTCTGGTAATATTAATAT 11364

—10 11382 —35 —10 11434 Start *orf11.45*
TATGGGCTAAATAATGATGTAATCTGTGAATTCCTGTAACCTCCACGAATATTTATTTACTTTTACATCTCGGAAAGTAAAAATCGCAACTATAAAAGCGAATTAACCACTTTAAAGTTGA 11484

L F I T F F L I A L G F Y L R T I F V S K M G A D L T G V M L L F T O L T A Y L
TTTATTTATTTACTTTTCTTATTTGCGTAGGATTTTATCTTGAACATAATTTGCTCTCAAGATGGGCGCTGATTTAAGAGTGTGATGTGTTATTTACAACTGACTGCTTATTT 11604

N L A E L G I G V A A A S L L Y K P L S E G D Y A K I K Y L T L L L S T I Y R Y
AAATCTCGCGAGCTTGGCATAGGGGTAGCTGCGCAAGCCTTTTATACAGCGCTGAGTGAGGGGATTTATGCAAAAAATAAAATACTTAACCTTTATGCTGTCTACAATCTATAGATA 11724

I S F L V L L I G I V I G F G I Y F F I D S V N A V S H V F I Y W A F F V I N T
TATCTCGTTTATTTATGATAGGTATAGTAATFGGCTTTGGTATTTACTTTTTCATCGATCTGTTAATGTCAGTAAGTACGTTTATTTATTTATGGGCAATTTTGTAAATAAATAC 11844

S L T Y S Y A K H S T L T A N Q Q Y S V V R K I Q G G K I L I I A L O L L L
ATCTTTAATCTTATCTTATGCAAAAGCATCAACTCTATTAAGTCAAAATCAGCAATATTCAGTAGTCCGTAATAACAGGGTGGCGGAAAAATATTAATTTATGGCGTCAAAATATTGCT 11964

L V T T H N F L L Y L L V E T I G V I V Q Y F I F K N I I N N D I H F K V V P Q
ATAGTACCACACATAATTTCTTACTTATTTGCTCGTGAACATATTGGCGGTGATAGTACAGTATTTTATATTTAAAAATATTTAATAACGACATCCATTTTAAGGTAGTCCACA 12084

S I S D D E K T T L K N E L K I K I K N M F F H K I G G V L V L N T D Y L L V S
ATCTATTAGCGATGATGAAAAACAACATTAATAAATGAATTAATAAATAAATAAATAAATATGTTTTTCATAAGATGGTGGTGTGTTAGTTTAAATAGTATTATTTGCTAGTTTC 12204

K F L N L S Y G T I Y G S Y M M V F O V V T V L M S S F V N A I T A S V G N F L
AAATCTTCTTAATTTAAGTCACGATTTATGGAAGTTATATGATGGCTTCCAGGTGTGACTGTGGCTATGTCAGTTTGTGTAATGCTATTACAGCATAGTGTGGGTAATTTT 12324

I N Q N D D E V T S I A K Q F N T V F I A L A T E I S L N M Y F L V N D F I T S
AATCAATCAAAATGATGATGAAGTTACAGTATGCAAGCAATTTAACACAGTATTTATAGCATTGGCTACCTTCACTTTTGAATATGATTTTGTGTTAATGACTTTATTACCAG 12444

W I G E K F I L G N G I V I L M L V N V F I S V I R I P C D I F K N A T G F F G
CTGGATAGCGGAAAAATTCATACTAGGTATGTTATATTAATGCTTGTATGATTTATAAGTGTCAITAGAAATTCCTTGATATTTTAAAAACGCAACTGGTTTTGTGG 12564

EcoRI
D V Y Y P L L E G V V N L F F S A L L A F Y I G L P G I I I G T I I S N V L I T
GGATGATATATTCTTTGCTTGGAGGAGTGGTAAATTTATTTTTCGCGCTCTGGCTTTTATATTGGTCTCCCGGTATTATTATGGCAGGATTATATCTAATGTATTAATTAC 12684

L I A K P L Y L Y G K M F G R F N A L K K Y L S F V L K P L I F S F V I F A V F
ATTAATTGCAAGCCATTATATCTATATGTTAAATGTTTGGTAGATTTAATGCAITAAAAAGTACTTGCATTGTAAGTAAACCTTTAATATTCTCATTTTGTGTTATTTTGGCTGTTT 12804

Y F T R E Q J I F F K V S N W F D F I S K L T I V S L V S M I I V F A V F Y A D
TTATTTTACAGAGAACAAATATTTTCTTTAAAGTTTCAACTGGTTTGACTTTTATAAGTAAGCTTACTATTGTATCGCTGGTAAGTATGATTATAGTTTTCGGGTTTTTATGCAGA 12924

A N F R S F V K R I L R V V F *
TGCTAATCTTAGGCTTTTGTAAAGAAATTTTACGCGTGGTATTTAAGTTAACCAAGCGGAAAAATATCCGCCGAACCAATGTGCTATTAAATGGACAAGTATTTAATATGCTTTAC 13044

Start *gnd*
GCTGTAATAATTTAGTACACATAATGGTGAATATATGTAATCTCCACACCTGACAGGAGTATGTAATGTCACAGCAAGATCGGCGTCTGCGTATGGCAGTATGGGGCGTAACCTCG 13164

A L N I E S R G Y T V S V F N R S R E K T E E V I A E N P G K K L V P Y Y T V
CGCTCAACATCGAAAGCGTGGTTATACGCTCTCCGTTTTCAACCGCTCCCGTGAAGAGACCGAAGTATGTCGAGAAATCCGCGCAAAAGCTGGTGCCTTATTATACGGTG 13280

Fig. 2. Sequence of part of the *rfb* gene cluster of *S. enterica* (M40). The *Pst*I site at the start of the sequence presented corresponds to the *Pst*I site between positions 2965 and 2970 in group B *rfb* DNA (Jiang *et al.*, 1991). Two regions are very similar to group B sequence (from 2965 to 4026, and from 13086 to 13280) and the group B bases are shown below where they vary. ORFs within the region from 4027 to 13085 are shown with their predicted amino acid sequences. Potential transmembrane segments are underlined. Potential promoters are shown by bars above the —35 and —10 region in the DNA sequence, and the start of the potential transcript is indicated. Some important restriction enzyme sites are shown by double underlining of the DNA sequence.

Gene	Sequence near start codon
<i>orf4-11</i>	ATTGAGAGACTTAGC ATG
<i>orf5-19</i>	GTCAGGTATTCTACT ATG
<i>orf6-17</i>	ATAGGACTAATAACA ATG ATAGTAATAAATAA
<i>orf7-17</i>	CAACTCGAGTTTTTAT ATG GAAAAAGTAA
<i>rfbM</i>	TGAGGTAAACAGGAAA ATGA
<i>rfbK</i>	AAGGAATAACGTAATA ATG
<i>orf11-45</i>	CATCCGGAAGTAAAA ATG

Fig. 3. Predicted translation initiation elements of the *rfb* genes of a group C1 *S. enterica* strain. The sequences are aligned with respect to the start ATG codons (shown in **bold**). Potential Shine-Dalgarno sequences are underlined and the termination codon of the preceding cistron is indicated by a bar above if it is in the sequence shown.

Discussion

We have sequenced the *rfb* (O antigen) gene cluster of a group C1 strain of *S. enterica* and find that there is only limited similarity to the cluster of group B (or groups A, D, E1 and C2 which are related to group B).

GDP-Man pathway genes

We have shown previously that genes of the mannose biosynthetic pathway lie between positions 8.86 and 11.96 kb on the *rfb* gene cluster of strain M40 (group C1) (Lee *et al.*, 1992). *orf8-36* and *orf9-79* are at the positions where *rfbM* and *rfbK* respectively had been located. The genes were compared with those of the mannose pathway of the *rfb* regions of group B (Jiang *et al.*, 1991) and group C2 (Brown *et al.*, 1991, 1992), and those of the *cps* gene cluster of strain LT2 (Stevenson *et al.*, 1991). The similarities in the sequences of the deduced polypeptides of the encoded phosphomannomutases and guanosine diphosphomannose (GDP-Man) pyrophosphorylases are shown in Table 1(a,b) and the alignments of the

Fig. 4. Complementation of antigen O7 expression in *E. coli* K12. In (a) and (b) are represented clones used in the localization of DNA required for the expression of antigen factor O7 in C600. pPR756 is a cosmid clone and pPR941, pPR1039, pPR1040, and pPR1041 are deletion derivatives of pPR756; all have been described previously (Lee *et al.*, 1992). pPR1261 was constructed by cloning the *EcoRI*-*KpnI* fragment (0.000-9.587) into pYUM1118 and removing the *SacI* fragment (0.638-2.927). pPR1262, in addition, lacked the *SacI* fragment (2.927-4.590). pPR1250 was constructed by inserting the *SacI*-*XbaI* fragment (5.498-8.901) into pT7T3 19U. pPR1289 was made by replacing the 14 bp *EcoRI*-*SacI* fragment of the pT7T3 19U polylinker region of pPR1250 with a 40 bp *EcoRI*-*SacI* fragment of the pYUM1118 polylinker region. pPR1263 has the *SacI*-*HincII* fragment (4.590-7.720) cloned into pGEM-7zf(+). Dotted lines represent DNA deleted and solid lines the DNA present in the clones. The triangle represents DNA inserted. All clones read from left to right.

Table 1. *Percentage amino acid and nucleotide identity among GDP-mannose pyrophosphorylases and phosphomannomutases*

The program BESTFIT was used to determine the alignments. We used gap weight and gap length weight values of 3.0 and 0.1, respectively. The percentage amino acid identity is shown, and where meaningful nucleotide base identity is shown in parentheses.

(a) GDP-mannose pyrophosphorylases			
	<i>rfbM</i> (B)	<i>cpsB</i> (B)	<i>rfbM</i> (C2)
<i>rfbM</i> (C1)	58.0	70.1 (63.3)	58.7
<i>rfbM</i> (B)		59.1	77.6 (73.6)
<i>cpsB</i> (B)			59.4
(b) Phosphomannomutases			
	<i>rfbK</i> (B)	<i>cpsG</i> (B)	<i>rfbK</i> (C2)
<i>rfbK</i> (C1)	23.7	96.9 (96.4)	22.2
<i>rfbK</i> (B)		22.3	68.3 (69.4)
<i>cpsG</i> (B)			21.9

sequences are shown in Fig. 5(a,b) respectively. Dendrograms showing the relationships are given in Fig. 6. DNA sequence similarities were barely detectable except for those with amino acid identities above about 70%.

The low levels of identity show why *rfbM*(C1) did not hybridize with *rfbM*(B) or *cpsB*(B) by Southern hybridization (Lee *et al.*, 1992). These data also clearly show that *rfbK* of group C1 is very closely related to the *cpsG* gene as suggested earlier by Southern hybridization and restriction enzyme data (Lee *et al.*, 1992). Of the differences which are present, many occur in the 28 bp at the end from 11131 to 11159, and the sequence from bases 9791 to 11130 shows 97.1% similarity to *cpsG*(B). *rfbM*(C1) stands apart from the other *rfbM* sequences but has the best fit with *cpsB*(B).

The overall homologies were clear although the levels of similarity in GDP-Man pyrophosphorylases were much higher than in phosphomannomutases (Fig. 6). This was similar to the observation made previously for GDP-Man pyrophosphorylase and phosphomannomutase of the *rfb* and *cps* gene clusters of strain LT2 (Stevenson *et al.*, 1991).

orf5-19, orf6-17 and orf7-17

These three ORFs together are sufficient when complemented by the *rfb* region of *E. coli* K12 to confer O7 specificity. *E. coli* K12 is an *rfb* mutant but may retain some *rfb* functions. The group C1 O antigen contains mannose, *N*-acetylglucosamine (GlcNAc) and glucose. We know that there are mannose pathway genes functionally equivalent to *rfbK* and *rfbM* in the *cps* region of *E. coli* K12 (unpublished data) as in *S. enterica* LT2 (Stevenson *et al.*, 1991), and these could substitute for

rfbK and *rfbM*. GlcNAc is found in peptidoglycan and enterobacterial common antigen so we expect genes for synthesis of UDP-GlcNAc to be elsewhere on the chromosome, and synthesis of UDP-glucose is known to be encoded by housekeeping genes. Thus all expected biosynthetic pathway genes are accounted for in *E. coli* K12, and *orf5-19*, *orf6-17* and *orf7-17* presumably encode the transferases for assembly of the group C1 O antigen as these are expected to be specific for the *S. enterica* group C1 structure.

We can estimate how many transferases are required as the main chain of the group C1 O antigen contains two α (1-2) mannosyl mannose linkages, one β (1,2) mannosyl mannose linkage, one β (1-3) mannosyl GlcNAc linkage and one α (1-2) GlcNAc mannose linkage (Lindberg *et al.*, 1988). One would expect *a priori* that there would be five transferases, one transferring the first sugar of the biological repeat unit and the others making oligomer linkages (Jann & Jann, 1984). We have assumed for the sake of economy that the transferase for the side chain glucose is not essential for O antigen polymerization and is encoded by a gene outside of *rfb* as this is often the case for side chain glucose (Mäkelä & Stocker, 1984).

We are left then with an expectation of five transferases but only three ORFs being sufficient to confer O antigen specificity. Unlike the situation for the other *rfb* gene clusters we have studied (Jiang *et al.*, 1991; Brown *et al.*, 1991, 1992; Wang *et al.*, 1992), in the case of group C1 we cannot account for all expected functions with the number of ORFs present in the *rfb* cluster.

One possible explanation is that the polymerases are not as specific as we expect and that for example a single α (1-2) mannose transferase makes both α (1-2) mannosyl mannose linkages, although it would be difficult to explain how the one transferase could specify the addition of precisely two residues to give the generally accepted O antigen structure. Alternatively, perhaps the O antigen synthesized in the *rfb* delete strain with the three group C1 genes is incomplete but includes the O7 epitope: there are precedents for this in a mutant of the *E. coli* O9 *rfb* region which synthesizes and polymerizes a shorter repeat unit (Kido *et al.*, 1989) and an *E. coli* O4 clone which expresses the O4 epitope but lacks a sugar in its O antigen (Hariguchi *et al.*, 1991). Unfortunately, the clones we have used are unstable in S ϕ 874 and it was not possible to obtain sufficient O antigen for SDS-PAGE analysis to resolve this matter. Finally, it is possible that not all genes for group C1 synthesis are in the *rfb* cluster as for the other groups, but that some functions are provided by genes elsewhere on the chromosome, and present in both *S. enterica* and *E. coli*.

The O antigen sugar transferases are said to be membrane bound, although the sequence of the group B *rfb* gene cluster gave no indication of this except for the

Fig. 5. Relationships of the amino acid sequence of protein encoded by (a) GDP-Man pyrophosphorylases and (b) phosphomannomutases. The sequences were aligned using the computer program CLUSTAL. In these comparisons, the gap penalty was set at 3. Asterisks indicate amino acids identical in all four sequences and dots those similar in all four.

Fig. 6. Evolutionary trees depicting the divergence of (a) phosphomannomutase and (b) GDP-Man pyrophosphorylases from the *rfb* gene clusters of *S. enterica* groups B, C1 and C2 and the *S. enterica* *cps* gene cluster. The proteins are referred to by their gene symbols and serogroup. The scale indicates percentage identity of amino acids after alignment using BESTFIT.

galactose transferase (Jiang *et al.*, 1991). However, in this case analysis of the three ORFs revealed that *orf5-19* and *orf6-17* have each one potential transmembrane segment, but *orf7-17* has none. The sequences of the three ORFs were used to search for any similarities with the other sequenced genes in the *rfb* region of *S. enterica* group B (Jiang *et al.*, 1991) and also with the sequences in GenBank (October 1991); no obvious similarities were detected with any of the known transferases or indeed other proteins.

orf11-45

The amino acid sequence of the protein encoded by *orf11-45* has 12 possible transmembrane segments distributed throughout its length. We have seen such a protein in *S. enterica* *rfb* gene clusters of strains of groups B (Jiang *et al.*, 1991), D (Liu *et al.*, 1991), C2 (Brown *et al.*, 1992) and E1 (Wang *et al.*, 1991). Each contains 12 predicted transmembrane segments with the transmembrane segments in approximately similar positions in *orf11-45* as in the other proteins (data not shown). Comparison, using BESTFIT, gave a particularly good alignment of *orf11-45p* with *orf7-9p*(C1) with only six small gaps and 33% identity. The similarity with the others was at the 20% identity level, which is not much above the level with the other 12 transmembrane proteins.

We have argued previously that *orf12-8*(B) is not likely to be a transferase, but to be involved in some other O antigen related function, perhaps O antigen transport,

and the same arguments would apply to *orf11-45*. *orf11-45* is required for O antigen expression in strain Sφ874 but not in strain C600 so must be complemented in strain C600 by the K12 form.

orf4-11

The group C1 O antigen is polymerized in C600 or Sφ874 carrying pPR1042 (data not shown), which indicates that if group C1 has an O antigen polymerase (see introduction), then the *rfc* gene is in the *rfb* cluster.

The *rfc* genes of groups B and C2, and the putative *rfc* gene of group E1 have similar predicted structures with 11 transmembrane segments (Brown *et al.*, 1992; Collins & Hackett, 1991; Wang *et al.*, 1992). *orf4-11* has a similar predicted structure and is the only ORF to do so. Thus if group C1 does have an *rfc* gene it is most likely *orf4-11*. The level of sequence similarity between presumed *rfc* genes is low in all cases and *orf4-11* is not out of line. We constructed clones either with or without *orf4-11* but instability has thus far prevented us from obtaining convincing evidence for *rfc* function of *orf4-11*.

Note that if group C1 O antigen is extended by single sugar extension, without involvement of an O antigen polymerase, this will not affect the total number of genes expected as in this case the O antigen chain will be started by addition of the first sugar to a carrier, which will require a specific enzyme as this step is different to subsequent additions of that sugar to partially completed O antigen by single sugar extension. There will thus be no effect on the argument given above on the expected number of genes.

Promoters

Our earlier studies (Jiang *et al.*, 1991) indicated potential promoters in the intergenic regions before *orf2-80* and *rfbB* respectively of group B, but the modified Berg-von Hippel indices (O'Neill, 1989b) were such that the promoters were not likely to function without help. This region is also present in strain M40 (Lee *et al.*, 1992) and the sequence of the promoter before *rfbB* at position 3928 is the same in group C1 as in group B (Fig. 2).

Computer searches using the three programs provided by Dr. M. C. O'Neill (O'Neill & Chiafari, 1989; O'Neill, 1989a,b) and a training back-propagation neural networks program (O'Neill, 1991) revealed a large number of potential promoters but the best had a modified Berg-von Hippel index of 3.88, and again none are expected to function without help.

The genes from *orf4-11* to *rfbK* are overlapping such that translational readthrough is possible, except for the gap of 14 bp before *rfbK*, which includes a good potential Shine-Dalgarno sequence (Fig. 3). It appears that these

Table 2. G + C content and P1, P2 and P3 values for the genes and ORFs of the *rfb* gene cluster

The values for each position are indicated as a percentage of the base content.

	P1	P2	P3	G + C
<i>orf4-11</i>	0.32	0.30	0.24	0.29
<i>orf5-19</i>	0.37	0.30	0.23	0.30
<i>orf6-17</i>	0.39	0.27	0.24	0.30
<i>orf7-17</i>	0.39	0.28	0.25	0.31
<i>rfbM</i>	0.51	0.37	0.28	0.39
<i>rfbK</i>	0.63	0.42	0.78	0.61
<i>orf11-45</i>	0.36	0.27	0.25	0.30

genes comprise a single operon. There is a 285 bp gap between *rfbK* and *orf11-45* and *orf11-45* may or may not be in this operon. There are candidate promoters at positions 11 286, 11 382 and 11 434 (Fig. 2) with modified Berg-von Hippel indices of 8.23, 8.34 and 5.28 respectively, but none of them would be expected to function without positive control, and identification of a promoter will require direct analysis of promoter function.

G + C content and codon usage

The G + C content of the *rfb* region of strain M40 is generally low (Fig. 7 and Table 2), with discrete regions of different G + C content: *orf4-11*, *orf5-19*, *orf6-17*, *orf7-17* and *orf11-45* have a G + C content of about 0.30, *rfbM* a G + C content of 0.39 and *rfbK* a G + C content of 0.61 (Table 2 and Fig. 7). The regions flanking the *rfb* region in strain LT2 are of typical *S. enterica* G + C content and codon usage (Jiang *et al.*, 1991) and these regions are similar in strain M40. As none of the *rfb* genes of strain M40 have a G + C content appropriate for *S. enterica* genes, they must all have been acquired by lateral transfer, as proposed for the *rfb* genes of group B strain LT2 (Jiang *et al.*, 1991). The properties of *orf4-11*, *orf5-19*, *orf6-17*, *orf7-17* and *orf11-45* are typical of those of a low G + C content organism, with average P1, P2 and P3 values (Sueoka, 1988) of 0.37, 0.28 and 0.24 respectively. The low value of P3 in particular suggests that this region has not been in *S. enterica* for long. These properties are similar to those of the central region of the *rfb* gene cluster of strain LT2 (Jiang *et al.*, 1991), and suggest a similar source for both sets of genes.

rfbK arose by gene duplication of *rfbM*

rfbM(C1) has a G + C content of 0.39 and P1, P2 and P3 values of 0.51, 0.37 and 0.28 respectively, which is similar to values for *rfbM* and *rfbK* of strain LT2 with 0.4 and 0.409 G + C content and P1, P2 and P3 of 0.55 and 0.517,

Fig. 7. G + C content of the *rfb* region of *S. enterica* group C1. The G + C content for the *rfb* genes and a few hundred bases of flanking DNA is from sequence presented in this paper (Fig. 2), and for the remainder of the flanking region is based on the sequence of a group B strain thought to have essentially the same sequence in these regions [Jiang *et al.* (1991) for *orf0-0* to *orf2-8* and Reeves & Stevenson (1991) for *gnd*].

0.37 and 0.431, and 0.30 and 0.295 respectively (Jiang *et al.*, 1991). The *rfbM* gene of strain M40 is the gene of the *rfb* gene cluster most similar to a gene in the *rfb* region of strain LT2, but it has diverged substantially. Both *rfbM* genes may have been derived originally from the same low G + C content organism, and must have undergone most of their divergence in such a low G + C content species.

rfbK(C1) has a G + C content of 0.61 and P1, P2 and P3 values of 0.63, 0.42 and 0.78 respectively, and is very similar (96.4%) to the *cpsG* gene of strain LT2. The tree for these *rfb* and *cps* genes is shown in Fig. 6: the phosphomannomutase genes seem to evolve significantly faster than the GDP-Man pyrophosphorylase genes, based on the differences between the genes of groups C2 and B, and between that pair of strains and the *cpsG/rfbK*(C1) pair. However, the 3% amino acid difference between *rfbK*(C1) and *cpsG* is very anomalous compared with the 17% difference between *rfbM*(C1), and *cpsB*, *cpsG* and *rfbK*(C1) are almost identical except for the last 28 bp and presumably arose by gene duplication. *cpsG* is part of the *cps* gene cluster for the widespread M antigen and was presumably the parent gene, with *rfbK* of strain M40 arising via gene duplication in a relatively recent event. Strain M40 retains its *cpsG* gene as shown by Southern blotting but we have not checked to see if it is expressed.

The pair of *cps* genes must have been present over a very long period in a high G + C species as their DNA is typical of 0.61 G + C content DNA (Stevenson *et al.*, 1991). The evolution of the group B and group C2 *rfbM* and *rfbK* genes, and the group C1 *rfbM* gene appears to have been in a 0.4 G + C content species, with relatively recent transfer to *S. enterica*. It seems most likely that the group C1 gene cluster at the time of transfer to *S. enterica* would have included both *rfbM* and *rfbK* genes, and we can only speculate as to how the phosphomannomutase gene of the *cps* gene cluster came to replace the original

rfbK gene, and thus become the *rfbK* gene for the C1 group. One possibility is that on transfer, or perhaps later, the *rfbK* gene suffered mutation or loss but that the gene cluster nonetheless functioned because the *cpsG* gene complemented the defect: gene duplication of *cpsG* and translocation of one copy by recombination could then have given the arrangement currently observed. If this indeed is what occurred, then recombination events to effect the replacement must have been very close to the ends of the genes. The terminal 28 bp of *rfbK* is of 0.3 G+C content and this may represent the end of the original low G+C content *rfbK* gene.

In summary, the group C1 *rfb* region has seven genes, arranged in an operon like manner. Of these genes, two are mannose biosynthetic pathway genes, one encodes a presumptive O antigen polymerase and four remain unidentified. One of the unidentified genes appears to encode a membrane protein and resembles genes found in all other *S. enterica* *rfb* gene clusters: the remaining three may well encode specific transferases. The whole cluster is derived from a low G+C content organism except for *rfbK* which appears to be a recent copy of the *cpsG* gene.

We would like to thank Dr O'Neill for carrying out promoter searches for us. The work was supported by a grant from the Australian Research Council.

References

- BRAHMBHATT, H. N., WYK, P., QUIGLEY, N. B. & REEVES, P. R. (1988). Complete physical map of the *rfb* gene cluster encoding biosynthetic enzymes for the O antigen of *Salmonella typhimurium* LT2. *Journal of Bacteriology* **170**, 98–102.
- BROWN, P. K., ROMANA, L. K. & REEVES, P. R. (1991). Cloning of the *rfb* gene cluster of a group C2 *Salmonella*: comparison with the *rfb* regions of groups B and D. *Molecular Microbiology* **5**, 1873–1881.
- BROWN, P. K., ROMANA, L. K. & REEVES, P. R. (1992). Molecular analysis of the *rfb* gene cluster of *Salmonella* serovar *muenchen* (strain M67): genetic basis of the polymorphism between groups C2 and B. *Molecular Microbiology* (in the Press).
- BUCHOLTZ, C. A. & REISNER, A. H. (1986). MBIS – an integrated system for the retrieval and analysis of sequence data from nucleic acids and proteins. *Nucleic Acids Research* **14**, 265–272.
- CARLOMAGNO, M. S., CHIARIOTTI, L., ALIFANO, P., NAPPO, A. G. & BRUNI, C. B. (1988). Structure and function of the *Salmonella typhimurium* and *Escherichia coli* K-12 histidine operons. *Journal of Molecular Biology* **203**, 585–606.
- COLLINS, L. V. & HACKETT, J. (1991). Molecular cloning, characterisation, and nucleotide sequence of the *rfe* gene encoding an O-antigen polymerase of *Salmonella typhimurium*. *Journal of Bacteriology* **173**, 2521–2529.
- DEVEREUX, J., HAEGERLI, P. & SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387–395.
- EISENBERG, D., SCHWARZ, E., KOMAROMY, M. & WALL, R. (1984). Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *Journal of Molecular Biology* **179**, 125–142.
- EWING, W. H. (1986). *Edwards and Ewings Identification of the Enterobacteriaceae*. Amsterdam: Elsevier Biomedical Press.
- FUTO, S., SETO, Y., MATSUO, T., TONOSAKI, Y. & FUKUCHI, A. (1989). pYUM1118: a new plasmid vector for generating deletions. *Nucleic Acids Research* **17**, 3311.
- GOUGH, J. A. & MURRAY, N. E. (1983). Sequence diversity among related genes for recognition of specific targets in DNA molecules. *Journal of Molecular Biology* **166**, 1–19.
- HARIGUCHI, G. E., ZHRINGER, U., JANN, B., JANN, K., HULL, R. A. & HULL, S. I. (1991). Genetic characterization of the O4 polysaccharide gene cluster from *Escherichia coli*. *Microbiol Pathogenesis* **10**, 351–361.
- HENIKOFF, S. (1987). Unidirectional digestion with exonuclease III in DNA sequence analysis. *Methods in Enzymology* **155**, 156–165.
- HIGGINS, D. G. & SHARP, P. M. (1989). Fast sensitive multiple sequence alignments on a microcomputer. *CABIOS* **5**, 151–153.
- JANN, K. & JANN, B. (1984). Structure and biosynthesis of O-antigen. In *Handbook of Endotoxin*, pp. 138–186. Edited by E. T. Rietschel. Amsterdam: Elsevier Scientific Publishers.
- JIANG, X. M., NEAL, B., SANTIAGO, F., LEE, S. J., ROMANA, L. K. & REEVES, P. R. (1991). Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar *typhimurium* (strain LT2). *Molecular Microbiology* **5**, 695–713.
- KLEIN, P., KANEHISA, M. & DELISI, C. (1985). The detection and classification of membrane spanning proteins. *Biochimica et Biophysica Acta* **815**, 468–476.
- KIDO, N., OHTA, M., IIDA, K.-I., HASEGAWA, T., ITO, H., ARAKAWA, Y., KOMATSU, T. & KATO, N. (1989). Partial deletion of the cloned *rfb* gene of *Escherichia coli* O9 results in synthesis of a new O-antigenic lipopolysaccharide. *Journal of Bacteriology* **171**, 3629–3633.
- KOZAK, M. (1983). Comparison of initiation of protein synthesis in prokaryotes, eukaryotes, and organelles. *Microbiological Reviews* **47**, 1–45.
- KYTE, J. & DOOLITTLE, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**, 105–132.
- LEE, S. J., ROMANA, L. K. & REEVES, P. R. (1992). Cloning and structure of group C1 O antigen (*rfb* gene cluster) from *Salmonella enterica* serovar *montevideo*. *Journal of General Microbiology* **138**, 305–312.
- LINDBERG, B., LEONTEIN, K., LINDQUIST, U., SVENSON, S. B., WRANGSELL, G., DELL, A. & ROGERS, M. (1988). Structural studies of the O-antigen polysaccharide of *Salmonella thompson*, serogroup C1 (6,7). *Carbohydrate Research* **174**, 313–322.
- LIU, D., VERMA, N. K., ROMANA, L. K. & REEVES, P. R. (1991). Relationships of the *rfb* regions of *Salmonella* serovars A, B, and D. *Journal of Bacteriology* **173**, 4814–4819.
- MÄKELÄ, P. H. (1966). Genetic determination of the O antigens of *Salmonella* groups B (4,5,12) and C1 (6,7). *Journal of Bacteriology* **91**, 1115–1126.
- MÄKELÄ, P. H. & STOCKER, B. A. D. (1984). Genetics of lipopolysaccharide. In *Handbook of Endotoxin*, pp. 59–137. Edited by E. T. Rietschel. Amsterdam: Elsevier Science Publishers.
- MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbour, NY: Cold Spring Harbour Laboratory.
- MESSING, J., CREA, R. & SEEBURG, P. H. (1981). A system for shotgun DNA sequencing. *Nucleic Acids Research* **9**, 309–321.
- NEUHARD, J. & THOMASSEN, E. (1976). Altered deoxyribonucleotide pools in P2 educants of *Escherichia coli* K-12 due to deletion of the *dcd* gene. *Journal of Bacteriology* **126**, 999–1001.
- O'NEILL, M. C. (1989a). Consensus methods for finding and ranking DNA binding sites. Application to *Escherichia coli* promoters. *Journal of Molecular Biology* **207**, 301–310.
- O'NEILL, M. C. (1989b). *Escherichia coli* promoters. I. Consensus as it relates to spacing class, specificity, repeat structure, and three-dimensional organization. *Journal of Biological Chemistry* **264**, 5522–5530.
- O'NEILL, M. C. (1991). Training back-propagation neural networks to define and detect DNA-binding sites. *Nucleic Acids Research* **19**, 313–318.
- O'NEILL, M. C. & CHIAFARI, F. (1989). *Escherichia coli* promoters. II. A spacing class-dependent promoter search protocol. *Journal of Biological Chemistry* **264**, 5531–5534.

- REEVES, P. R. & STEVENSON, G. (1989). Cloning and nucleotide sequence of the *Salmonella typhimurium* LT2 *gnd* gene and its homology with the corresponding sequence of *Escherichia coli* K12. *Molecular and General Genetics* **217**, 182–184.
- SANDERSON, K. E. & ROTH, J. R. (1989). Linkage map of *Salmonella typhimurium*. *Microbiological Reviews* **52**, 485–532.
- SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain termination inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5463–5467.
- SANGER, F., COULSON, A. R., BARRELL, B. G., SMITH, A. J. H. & ROE, B. A. (1980). Cloning in single stranded bacteriophage as an aid to rapid DNA sequencing. *Journal of Molecular Biology* **143**, 161–178.
- SHAW, D. C., WALKER, J. E., NORTHPROP, F. D., BARRELL, B. G., GODSON, G. N. & FIDDES, J. C. (1978). Gene K, a new overlapping gene in bacteriophage G4. *Nature, London* **272**, 510–515.
- SHEPHERD, J. C. W. (1981). Method to determine the reading frame of a protein from the purine/pyrimidine genome sequence and if possible evolutionary justification. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 1596–1600.
- STADEN, R. (1982). Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucleic Acids Research* **12**, 505–519.
- STADEN, R. (1984). Computer methods to locate signals in nucleic acid sequences. *Nucleic Acids Research* **12**, 505–519.
- STADEN, R. (1986). The current status and portability of our sequence handling software. *Nucleic Acids Research* **14**, 217–231.
- STEVENSON, G. S., LEE, S. J., ROMANA, L. K. & REEVES, P. R. (1991). The *cps* gene cluster of *Salmonella* strain LT2 includes a second mannose pathway: sequence of two genes and relationship to genes in the *rfb* gene cluster. *Molecular and General Genetics* **227**, 173–180.
- SUEOKA, N. (1988). Directional mutation pressure and neutral molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 2653–2657.
- VERMA, N. K., QUIGLEY, N. B. & REEVES, P. R. (1988). O-Antigen variation in *Salmonella* spp.: *rfb* gene clusters of three strains. *Journal of Bacteriology* **170**, 103–107.
- VERMA, N. & REEVES, P. R. (1989). Identification and sequence of *rfbS* and *rfbE*, which determine antigenic specificity of group A and group D *Salmonellae*. *Journal of Bacteriology* **171**, 5694–5701.
- VIEIRA, J. & MESSING, J. (1987). Production of single-stranded plasmid DNA. *Methods in Enzymology* **153**, 3–11.
- WANG, L., ROMANA, L. & REEVES, P. (1992). Cloning and sequencing of the *rfb* gene cluster determining the O antigen of a group E1 *Salmonella*. *Genetics* **130**, 429–443.
- WYK, P. & REEVES, P. R. (1989). Identification and sequence of the gene for abequose synthase which confers antigenic specificity on group B *Salmonellae*: homology with galactose epimerase. *Journal of Bacteriology* **171**, 5687–5693.