JOURNAL OF BACTERIOLOGY, Oct. 1989, p. 5687–5693
0021-9193/89/105687-07$02.00/0

Vol. 171, No. 10

Identification and Sequence of the Gene for Abequose Synthase, Which Confers Antigenic Specificity on Group B Salmonellae: Homology with Galactose Epimerase

PAUL WYK[1] AND PETER REEVES[2]*

Department of Microbiology and Immunology, University of Adelaide, Adelaide,[1] and Department of Microbiology, University of Sydney, NSW 2006[2] Australia

Received 26 January 1989/Accepted 19 June 1989

The O antigen of *Salmonella* group B strains contains the sugar abequose, whereas those from group A and D strains contain paratose or tyvelose in its place. This is the essential difference between these *Salmonella* groups. Only the final step in the biosynthesis of abequose differs from that of paratose, and the abequose confers on group B strains their specific O4 antigen. The gene, *rfbJ*, encoding the enzyme abequose synthase for this last specific step has been cloned, identified, and sequenced and has been shown to function in group A and D strains to make them O4$^+$. This one gene thus differentiates group B from group A or group D salmonellae. The enzyme abequose synthase appears to be related to galactose epimerase, and the significance of this is discussed. The *rfbJ* gene and adjacent DNA is of much lower G+C content than is usual for salmonellae, indicating that the region did not originate in a salmonella but was transferred from outside.

We are studying the genetics of a major polymorphism in bacteria and present here the genetic base for the O4 antigenic epitope, which characterizes one form of the O antigen polymorphism.

The O antigen is a polysaccharide which covers much of the surface of gram-negative bacteria, although it may be masked by a capsule external to the cell proper. The O antigen is typically a polymer with a repeating oligosaccharide of 3 to 6 sugar residues, which is linked through an oligosaccharide core to lipid A, the whole comprising the lipopolysaccharide (LPS), which is the major lipid of the outer leaflet of the outer membrane characteristic of gram-negative bacteria. The lipid A and core do not vary greatly within a genus and are thought to be invariable within salmonellae. The O antigen, however, is extremely polymorphic and in salmonellae, about 40 major forms are currently recognized. It should be noted that the genus *Salmonella* comprises a single species, *S. enterica*; unfortunately, the many serovars are generally given full species rank, although DNA hybridization and other factors show that the general level of variation between these serovars is that expected for a single bacterial species (4, 22).

We have studied three *Salmonella* groups (A, B, and D), all of which have an O antigen with a repeat unit of four sugars. Of these sugars, three form a mannosyl-rhamnosyl-galactose backbone common to the three groups, which differ only in the fourth sugar, a dideoxyhexose substituted on the mannosyl residue. The dideoxy sugar is paratose in group A, abequose in group B, and tyvelose in group D (see reference 19 for a recent review). These three dideoxysugars in their (α-1,3) linkage to mannose confer the O2, O4, and O9 antigenic epitope specificities, respectively.

Variation in the O antigen is thought to be largely the result of variation in the *rfb* gene cluster (25), which maps at 42 minutes on the *Salmonella* chromosome (36). The genes for the O antigen-specific parts of the biosynthetic pathways of the component sugars and the genes for the transferases which build up the repeat oligomer on the carrier lipid, undecaprenol phosphate, are all in this gene cluster. Five of these genes are involved in the synthesis of CDP-abequose (Fig. 1), the nucleotide sugar from which abequose is transferred to mannosyl-rhamnosyl-galactosyl-undecaprenolphosphate to complete the oligomer. The genes for the first two steps, *rfbF* and *rfbG*, are well documented (29), but the other genes were not individually named, since the overall reaction carried out by the encoded enzymes was assayed as one in that analysis. However, it is known that there are three steps involved (27, 33), and we propose that the genes be named *rfbH*, *rfbI*, and *rfbJ*, as shown in Fig. 1. A point mutant in *rfbH* has been described previously (48).

It is only at the very last step that the pathways to CDP-

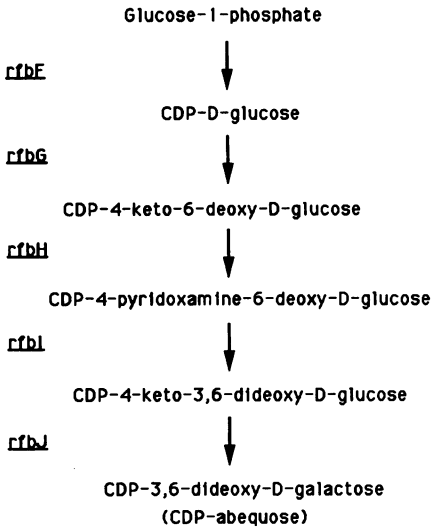


FIG. 1. Pathway for the biosynthesis of CDP-abequose (taken from reference 25).

* Corresponding author.

TABLE 1. Strains used

Strain	Our stock no.[a]	Characteristics	Reference or source
E. coli DS410	P2039	*azi tonA lacY minA minB xyl mtl thi ara*	12
Salmonella			
M6		IMVS isolate J84/0551, serovar dublin, group D	IMVS[b]
M22		IMVS isolate J84/5174, serovar dublin, group D	IMVS
Ty21a	M18	*galE* mutant of Ty2, serovar typhi, group D	13
IMVS1316	M8	Serovar paratyphi A, group A	IMVS
CL4419	P9003	*hsdL trpB2 nml flaA66 rpsL xylT404 ilv-452 metE554 metA22 hsdA*	31

[a] Given only when different from the strain name.
[b] IMVS, Institute of Medical and Veterinary Sciences, Adelaide, Australia.

abequose and CDP-paratose diverge, and abequose synthase, the *rfbJ* product, thus determines the 3,6-dideoxyhexose specificity (27). Note that in group A and group D strains, abequose synthase is replaced by paratose synthase, which reduces the same substrate to the glucose instead of the galactose isomeric state at carbon 4, and we would expect *rfbJ* to be replaced. Group D strains have, in addition, an epimerase to convert CDP-paratose to CDP-tyvelose.

A partial gene order for the *rfb* cluster of strain LT2 was established by using a series of mutants with deletions extending from *his* into *rfb* (23, 29), and we have recently cloned the whole region from strains LT2 (group B, serovar typhimurium), IMVS1316 (group A, serovar paratyphi A), and Ty2 (group D, serovar typhi) and located the known genes, approximately, on a restriction map of the region (2, 3, 44).

In this study, we located and sequenced the *rfbJ* gene and showed that when cloned into group A or group D strains, it confers on them the ability to make abequose, which is then incorporated into O antigen.

MATERIALS AND METHODS

Strains are listed in Table 1. Strains marked IMVS were obtained from the Institute of Medical and Veterinary Sciences, Adelaide, Australia. All derived plasmids contain *rfb* DNA from strain CL4419 (an LT2 strain). Map coordinates are as in reference 3. Plasmids pPR301 and pPR302 contain the *Eco*RI fragment of *rfb* from positions 0 to 9.57, inserted in opposite orientations in the *Eco*RI site of pEMBL8 (9). All other plasmids contain *rfb* DNA from LT2 cloned into pUC18 (47). Plasmid pPR303 is an *Eco*RI to *Bgl*II fragment cloned between the *Eco*RI and *Bam*HI sites of pUC18; pPR586 and pPR587 contain an *Sph*I fragment cloned into the *Sph*I site; and pPR874 contains an *Hpa*I fragment cloned into the *Sma*I site. Plasmids pPR877 and pPR878 are two of a family made from pPR303 by BAL 31 digestion from the *Pst*I polylinker site, followed by addition of *Pst*I linkers and ligation. The plasmids were used in experiments presented in Fig. 2, and the map locations of restriction sites, the ends of BAL 31 digestion, and orientation of inserts can be seen there.

Analysis of O antigen and LPS. The O2, O4, and O9 O antigen epitopes on cells were detected by slide agglutination as described by Leinonen (20). Sera were generously given by the Institute of Medical and Veterinary Science, Adelaide, South Australia.

The procedure of Hitchcock and Brown (16) was used as a rapid means of preparing samples for electrophoresis on acrylamide gels to detect LPS. The Western blot (immunoblot) procedure of Burnette (6) was followed to locate O4 antigen in LPS by using specific anti-O4 serum; the procedures of De Joungh-Leuvenink et al. (8) were used for adsorption of secondary antibodies and for developing the horseradish peroxidase assay.

SDS-PAGE. Proteins were analyzed essentially as previously described (24) on 11 to 20% gradient sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gels. Samples were heated in a boiling water bath for 5 minutes before loading, and gels were stained by the method of Achtman et al. (1).

Minicell methods. Minicells were purified on sucrose step gradients, and [^{35}S]methionine was incorporated into pro-

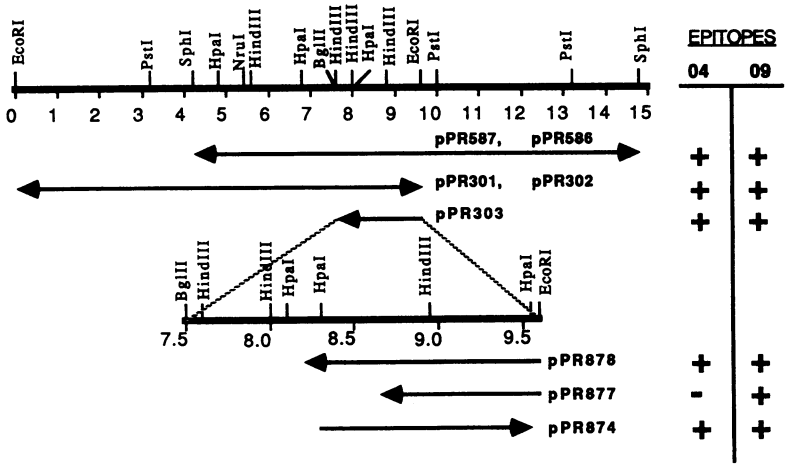


FIG. 2. Expression of the antigenic epitope O4 in strain M6 carrying plasmids with LT2 *rfb* DNA. Arrowheads on lines representing clones indicate the direction of transcription from the external *lac* promoter. Arrowheads on both ends of a line indicates that the insert was cloned in both orientations, in which case the name to the left represents the plasmid in which the external *lac* promoter reads into the insert material from right to left. Map positions (in kilobases) are from Brahmbhatt et al. (3).

teins as described by Dougan and Kehoe (10). Extracts of labeled minicells were electrophoresed on 11 to 20% gradient SDS-PAGE, dried down at 80°C onto Whatman 3MM filter paper, and autoradiographed at room temperature.

DNA methods. Routine methods were as described by Maniatis et al. (26). Sequencing was carried out by the dideoxy method of Sanger et al. (37, 38). Fragments from an *Hpa*I-*Hin*dIII double digest of the *Eco*RI-*Pst*I insert from pPR303 were treated with DNA polymerase to end fill the *Hin*dIII- and *Eco*RI-cut ends. They were separated on an agarose gel, and individual bands were cut out and cloned into M13mp18 (30) for sequencing. Further sequence was obtained from some of the BAL 31 deletions of pPR303 described above, after transferring the *Eco*RI-*Pst*I fragments to M13mp18. Remaining gaps were filled in by using synthetic oligonucleotides as primers and a suitable M13mp18 clone from those described above, to give sequence in both directions from the *Hpa*I site at 8.31 to the *Eco*RI site at 9.57.

All computer sequence analysis was done on the MBIS System at the CSIRO Biotechnology Laboratory, Sydney, Australia (5), which incorporates several sets of programs. The sequence was assembled by using the DB system of Staden (41, 42), and the molecular weight and a hydrophobicity index were calculated by the program ANALYSEQ (41). We used the program PREDICT, which combines eight different methods to produce a joint protein secondary structure prediction (11). The program MTX (34) was used with the substitution matrix of Dayhoff et al. (7) to give a dot matrix sequence comparison, and the program BESTFIT in the GCG package, which uses the "local homology" algorithm of Smith and Waterman (40) was used to align proteins.

RESULTS

Identification of *rfbJ*. The O antigens of groups A, B, and D are chemically very closely related, and as a first step in unraveling the genetic basis of O antigen variation, we have looked for the genetic basis for the specificity of group B strains relative to group A or D, essentially because strain LT2, a group B strain (serovar typhimurium) has been well characterized genetically. Group B salmonellae are distinguished from group A and D strains by having abequose in their O antigen and are expected to have the gene *rfbJ*, which encodes the one step specific to the abequose biosynthetic pathway.

Our earlier studies (3, 44) had shown that the *rfb* regions of *Salmonella* strains LT2 (group B) and Ty2 (group D) were almost identical over much of their lengths, with only about 4.37 kilobases of LT2 DNA not showing homology with Ty2 DNA by restriction mapping, by Southern blotting, or heteroduplex analysis. We have since shown by Southern blotting at low stringency that there is some homology over much of this latter region (unpublished data) and now refer to it as a region of low homology. The deletion which defined, collectively, the terminal three genes of the abequose pathway (*rfbH*, *rfbI*, and *rfbJ*) also ended within this region of low homology, giving an approximate location for *rfbJ* (although it should be noted that the data showed only that at least one of the three genes lay in this area). We therefore transferred plasmids pPR586, pPR587, and pPR301 (Fig. 2), which cover this approximate area, into representative strains of group A and group D (Ty21a, M6, and M22). The transformants all produced epitope O4, in addition to the O2 or O9 epitopes of the parent strains.

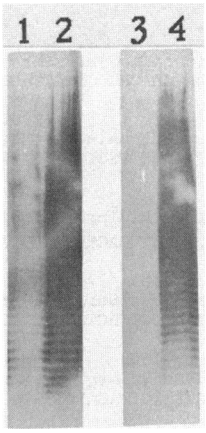


FIG. 3. Epitope O4 of group B expressed in cells of strain M6, a group D strain. Cell lysates were run on an 11 to 20% SDS-PAGE gradient gel and transferred to nitrocellulose paper. This was probed with anti-O4 antisera as the primary antiserum and developed with a commercially available goat anti-rabbit horseradish peroxidase conjugate. Strains are M6 carrying pPR874 (lane 1), pPR303 (lane 2), or no plasmid (lane 3) and CL4419 (lane 4).

A family of BAL 31 deletion mutants was then made from pPR303, and a selection of these clones was transformed into strain M6 and checked for O4 antigen. M6 carrying pPR878, which had an *rfb* DNA insert of 1.40 kilobases (from 8.17 to 9.57), was able to produce O4 antigen, whereas M6 carrying pPR877, which had an *rfb* fragment of 1.13 (from 8.44 to 9.57), did not (Fig. 2).

The DNA necessary for this complementation was further defined by using pPR874, which contains an *Hpa*I fragment bounded by 8.31 and 9.54, and which, when transformed into strain M6, gave a strain which produced O4 antigen, as well as O9 antigen. Thus, the DNA necessary for this complementation has been localized to a 1.23-kilobase DNA fragment, which we assume must carry *rfbJ*.

LPS from M6 strains carrying plasmids which express *rfbJ* showed the typical ladder (data not shown) when it was electrophoresed on polyacrylamide gels and silver stained (15, 18, 32). When such gels were developed as a Western blot with anti-O4 sera, the typical LPS ladder was again seen (Fig. 3), showing that the O4 determinant was polymerized to give LPS of normal length.

Visualization of *rfbJp* and the sequence. Plasmids carrying *rfbJ* were transformed into DS410, and minicells were prepared and labeled, and their proteins analyzed by SDS-PAGE. The result is shown in Fig. 4. Strains carrying *rfbJ* produce one extra band that corresponds to a protein of M_r 33,000.

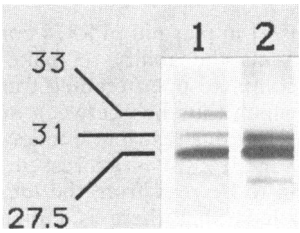


FIG. 4. Proteins produced by minicells carrying a plasmid with *rfbJ*. Strain DS410 carrying pPR874 (lane 1) or pUC18 (lane 2) was grown with [^{35}S]methionine, minicells were harvested and lysed, and whole cell extracts were run on SDS-PAGE. An autoradiograph is shown, and an additional band of 33 kilodaltons can be seen when pPR874 is present.

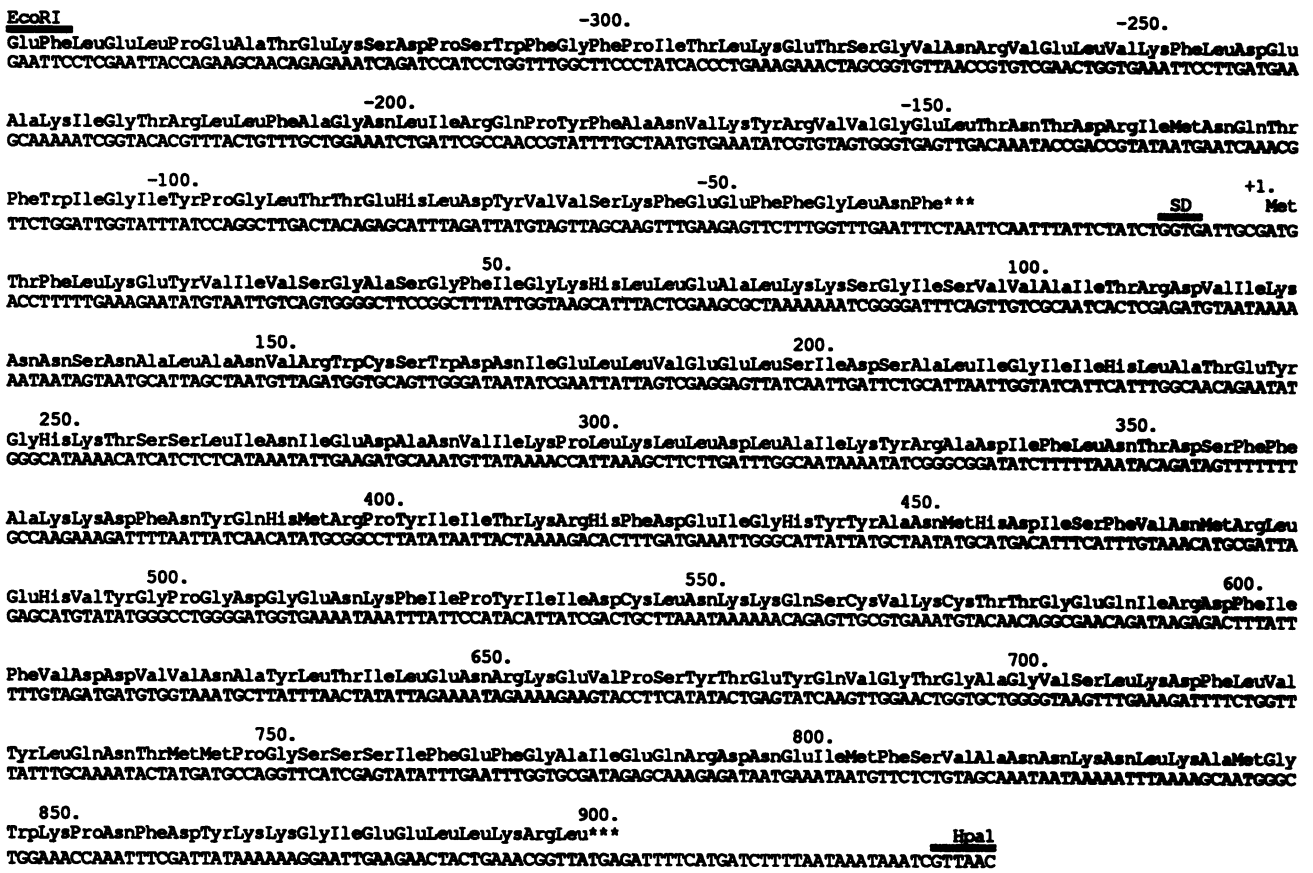


FIG. 5. DNA sequence of the *rfb* region from the *Eco*RI site at 9.57 to the *Hpa*I site at 8.35. An incomplete open reading frame and a complete one assumed to represent *rfbJ* are shown by their derived amino acid sequences. A possible Shine-Dalgarno site is indicated (SD). See Fig. 2 for map positions.

The region from the *Eco*RI site at 9.57 to the *Hpa*I site at 8.31 was sequenced, and the sequence is shown in Fig. 5. It contains a single complete open reading frame which must thus represent the *rfbJ* gene. There is also the 3′ end of an open reading frame comprising 109 codons, which reads through the *Eco*RI site and terminates 27 bases from an ATG codon which we assume to be the start of the *rfbJ* gene. The protein encoded is composed of 299 amino acids and has a predicted molecular mass of 34,106 daltons, which agrees well with the estimate based on the protein seen on PAGE gels. The protein was given a hydrophobicity index of 16.18 by ANALYSEQ.

DISCUSSION

The 1.23-kb insert in plasmid pPR874 confers on group A and group D strains the ability to make abequose, and, hence, the one complete open reading frame present must represent *rfbJ*, which thus lies between approximate positions 8.31 and 9.21 of our restriction map.

The *rfbJ* gene in pPR874 is expressed, although in the wrong orientation to be read from the *lac* promoter of the vector. We conclude that there is either another vector promoter reading into the insert from the other end or there is a promoter in the insert; however, there is no good promoter sequence apparent upstream of *rfbJ* and resolution of this matter requires further study. The upstream open reading frame terminates 27 bases before the start of *rfbJ*. There is a weak potential Shine-Dalgarno sequence which is marked in Fig. 5 and may function or there may be some readthrough from the upstream open reading frame, although the 27-base interval is not within the range reported for efficient readthrough (14). Alternatively, *rfbJ* may start at the GTG codon three codons upstream of the ATG codon and may not have a Shine-Dalgarno sequence.

The product of *rfbJ*. Abequose synthase is a dehydrogenase converting CDP-4-keto-3,6-dideoxy-D-glucose to CDP-3,6-dideoxy-D-galactose, (CDP-abequose), transferring two hydrogen atoms from NADPH to the keto moiety on carbon 4 (27). NAD-linked dehydrogenases, as a class of proteins, have been extensively studied by X-ray crystallography. The NAD-binding domain, common to NAD-linked dehydrogenases, consists of a six-stranded parallel beta sheet (strands named βA to βF) connecting alpha helices.

The PREDICT program (11) predicts several beta strands within the first 140 amino acids of abequose synthase. The amino acids of the first two show a reasonable fit to the pattern for βA and βB, particularly with regard to amino acids thought to be necessary for structural stability and function in the NADH-binding domain (35) (Fig. 6).

We conclude that abequose synthase probably has the six-beta-strand structure characteristic of NAD-linked dehydrogenases and that βA and βB occupy residues 6 to 13 and 31 to 37, respectively, with potential βC and βD segments at residues 48 to 54 and 74 to 79.

Comparison of the *rfbJ* and *galE* genes and their products. The enzyme UDP-glucose-4-epimerase encoded by the gene

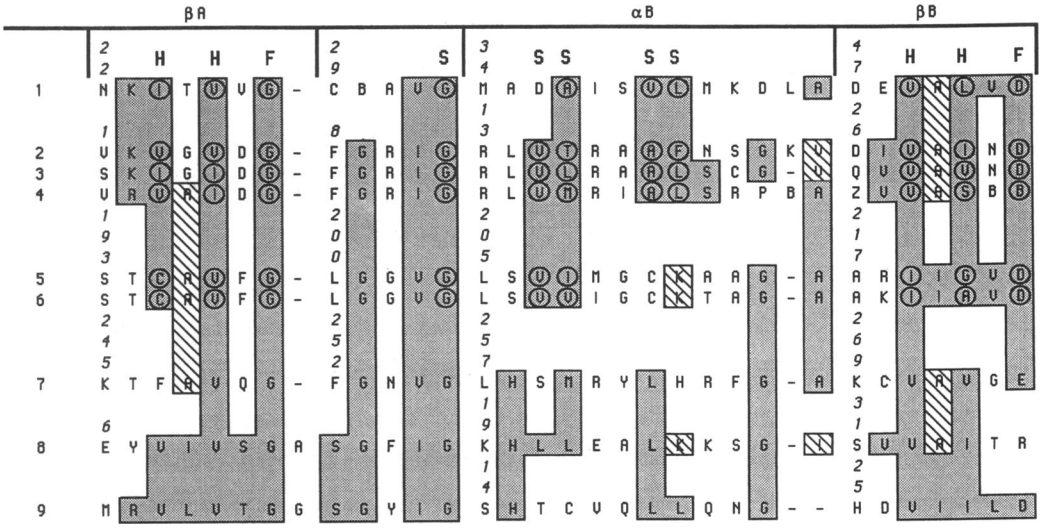


FIG. 6. Comparison of abequose synthase and UDP-glucose-4-epimerase with dehydrogenases of known tertiary structure. Sequences shown are: 1, dogfish lactate dehydrogenase; 2, pig glyceraldehyde-3-phosphate dehydrogenase; 3, lobster glyceraldehyde-3-phosphate dehydrogenase; 4, yeast glyceraldehyde-3-phosphate dehydrogenase; 5, horse liver alcohol dehydrogenase; 6, rat liver alcohol dehydrogenase; 7, bovine glutamate dehydrogenase; 8, abequose synthase; and 9, UDP-glucose-4-epimerase. The alignment of sequences 1 to 7 is taken from Table 4 of Rossman et al. (35) and is based on known tertiary structures. The functions indicated on the top line are β sheet residues in hydrophobic pockets facing α helices (H) or residues in the helices in the pocket and facing the β sheet (S), or they are functional in binding the AMP component of the coenzyme (F). Residues with these functions were identified by Rossman and are circled. Sequence 8 is from this study, and sequence 9 is from reference 21. These two bacterial sequences have been aligned by eye with the other seven sequences. The residues shown under βA and βB comprise essentially the first two β sheets predicted (see Fig. 6). We have shaded homologous residues by using three criteria: (i) all residues circled as indicated above; (ii) all residues in sequences 7 to 9 which are identical or very similar to circled residues; (iii) all blocks of 2 or more identical or very similar residues at a given position, unless the homology involves only 2 of the 4 enzyme specificities represented by sequences 1 to 7. In most positions, only one set of identical or similar residues was found; in the other positions, distinctive shading was used for one of the 2 sets.

galE interconverts UDP-galactose and UDP-glucose by a mechanism which has similarities to that of abequose synthase (see below).

The sequence of the *Escherichia coli galE* gene has been published (21), and we have analyzed the predicted amino acid sequence by the methods discussed above. We find beta strand and alpha helix predictions that fit with the model for the NAD^+-binding domain, and there is good homology with well characterized NAD-linked dehydrogenases for the first two beta segments of that domain (Fig. 6).

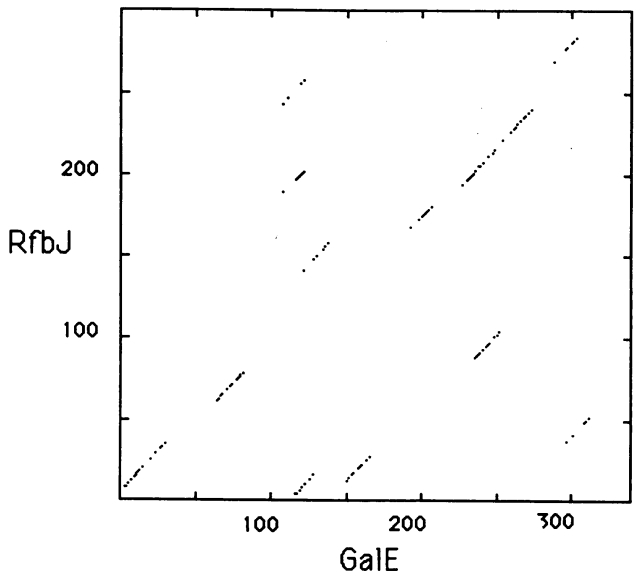
An MTX dot plot comparison (Fig. 7) of the amino acid sequences of abequose synthase and the epimerase shows homology at the amino-terminal end and in the carboxy-terminal half. Alignments for these two regions are shown in Fig. 7: the sequences assigned above to βA are very strongly conserved between the two, with eight identical and three similar amino acids within and immediately beyond βA. The carboxy-terminal half has a more moderate but extended amino acid homology.

The homology between abequose synthase and UDP-glucose-4-epimerase is extremely interesting. The epimerase is a dimer with a single tightly bound NAD^+, which is used to oxidize carbon 4 of UDP-glucose or UDP-galactose to UDP-4-keto-glucose (46). However, the UDP-4-keto-glucose and NADH are not released, but the keto group can be reduced back to a hydroxyl group in either the glucose or galactose isomeric form. Thus, the enzyme can interconvert UDP-glucose and UDP-galactose through its unusual property of being able to attack the keto group from either side.

Abequose synthase uses NADPH as a cofactor and reduces the 4-keto group of CDP-4-keto-3,6-dideoxy-glucose to the galactose isomeric form. We propose that it is structurally similar to UDP-glucose-4-epimerase but can attack its substrate from one side only, to give CDP-abequose (or, alternatively, the isomer CDP-paratose may be too tightly bound to the active site for it to be released at detectable levels). Abequose synthase also differs from UDP-glucose-4-epimerase in releasing the NAD^+, thus acting as a typical dehydrogenase.

The corresponding enzyme of group A (and also group D) strains, paratose synthase, reduces the 4-keto group of the same substrate to the glucose isomeric form and differs from abequose synthase in allowing attack from the other side only (or in allowing release only of CDP-paratose). In the accompanying paper we give the sequence of the paratose synthase gene of a group D salmonella.

The low G+C content. The *rfbJ* gene has a very low guanosine-plus-cytosine (G+C) content of 0.32, whereas *Salmonella* genes normally have about 0.51 G+C content. The variation in G+C content in different bacterial species is thought to be due to differences in the mutation rates (A.T)→(G.C) and (G.C)→(A.T), (where (A.T) refers to an AT or TA base pair and (G.C) to a GC or CG base pair). This gives rise to a bias in genetic drift (a directional mutation pressure) to give the whole genome a G+C content characteristic for a given species or group (43). The *rfbJ* gene, and perhaps more or even all of the *rfb* gene cluster, must thus have developed in an organism with a low G+C content and then have been transferred to the genus *Salmonella*. The genus *Salmonella* is in the family *Enterobacteriaceae*, with G+C contents ranging from 0.38 to 0.60 (4), and the *rfbJ* gene is presumed to derive from a source outside of this family. The G+C content of first, second, and third codon bases (P_1, P_2, and P_3) is 0.41, 0.30, and 0.24, respectively, showing the descending content characteristic of low-G+C species (28, 43).




```
RfbJ   8 VIVSGASGFIGKHLLEALKKSGISVVAITRDVIKNNSNALANVRWCSWDN
         ||| ||||||| |    |  |  || |       |      || |
GalE   3 VLVTGGSGYIGSHTCVQLLQNGHDVIILDNLCNSKRSVLPVIERLGGKHP

      58 I........ELLVEELSIDSALIGIIHLAT.............EYGH
                 | || ||  | ||  |||||                | ||
      53 TFVEGDIRNEALMTEILHDHAIDTVIHFAGLKAVGESVQKPLEYYDNNDN

      84 KTSSLINIEDANVIKPLKLLDLAIKYRADIFLNTDSFFAKKDFNYQHMRP
          | ||    | || | |   | |  | |    ||          |
     103 GTLRLISAMRAANVKNFIFSSSATVYGDQPKIPYVESFPTGTPQSPYGKS

     134 YIITKRHFDEIGHYYANMHDISFVNMRLEHVYGPGDGE.......NKFIP
          ||   | ||    ||     |  |        || |       | |||
     153 KLMVEQILTDLQKAQPDWSIALLRYFNPVGAHPSGDMGEDPQGIPNNLMP

     177 YIIDCLNKKQSCV.......KCTTGEQIRDFIFVDDVVNAYLTILENRKE
          || |   |  ||          |  ||||| | || || |  ||   |
     203 YIAQVAVGRRDSLAIFGNDYPTEDGTGVRDYIHVMDLADGHVVAMEKLAN

     220 VPSYTEYQVGTGAGVSLKDFLVYLQNTMMPGSSSIFEFGAIEQRDNEIM.
          ||   | || | | || | |    | |  |||   ||     ||  |
     253 KPGVHMYNLGAGVGNSVLDVV....NAFSKACGKPVNYHFAPRREGAFRP

     269 FSVANNKNLKAMGWKPNFDYKKGIEELLKR 298
          |    |   | ||         || | |
     299 TGRTPAKPTVNLNWRVTRTLDEMAQDTWHW 328
```

FIG. 7. Relationships of the amino acid sequences of abequose synthase (RfbJ) and UDP-glucose-4-epimerase (GalE). The diagram shows the output from the program MTX, with the span for comparisons set at 16 residues. The sequence alignment was obtained with the program BESTFIT by using the default parameters of gap weight = 5.0 and gap length weight = 0.3.

P_3 can range from 0.09 to 0.95 and reflects quite well the directional mutation pressure (43). P_1 and P_2 vary over a smaller range, due to constraints related to amino acid choice, and presumably drift much more slowly. The values of P_1 and P_2 for *rfbJ* are only a little lower than those for a typical gene with a P_3 of 0.24. The value of P_3 of *rfbJ* thus still reflects the equilibrium established in its low-G+C source species and has not had time to drift significantly since transfer to the genus *Salmonella*. Given that synonomous substitution (base changes which do not affect the amino acid coded and occur mostly in base 3) since the divergence of the genera *Escherichia* and *Salmonella* averages 0.90 per synynomous site (39) and such substitution would bring P_3 close to equilibrium for the species, we conclude that transfer of *rfbJ* must have occurred a short time ago relative to the time since the two genera diverged. Several genes encoding virulence factors in *E. coli* and *Vibrio cholerae* have low G+C contents, but they are all factors present in only some pathogenic strains or species and in the case of *E. coli*, the genes are often on plasmids. Transfer from other species is easily explained. However the O antigen is characteristic of gram-negative bacteria and is very widely present. It is, therefore, surprising to find that genes concerned with the specificity of this polymorphic but generally present character have been transferred from a distantly related species.

ACKNOWLEDGMENTS

This work was supported by a grant from the Australian Research Grants Scheme, now incorporated into the Australian Research Council.

LITERATURE CITED

1. **Achtman, M., S. Schwuchow, R. Helmuth, G. Morelli, and P. A. Manning.** 1978. Cell-cell interactions in conjugating *Escherichia coli*: Con$^-$ mutants and stabilization of mating aggregates. Mol. Gen. Genet. **164:**171–183.
2. **Brahmbhatt, H. N., N. B. Quigley, and P. R. Reeves.** 1986. Cloning part of the region encoding biosynthetic enzymes for surface antigen (O-antigen) of *Salmonella typhimurium*. Mol. Gen. Genet. **203:**172–176.
3. **Brahmbhatt, H. N., P. J. Wyk, N. B. Quigley, and P. R. Reeves.** 1988. A complete physical map of the *rfb* locus encoding biosynthetic enzymes for the O-antigen of *Salmonella typhimurium* LT2. J. Bacteriol. **170:**98–102.
4. **Brenner, D. J.** 1984. Enterobacteriaceae, p. 408–516. *In* N. R. Krieg and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 1. The Williams & Wilkins Co., Baltimore.
5. **Bucholz, C. A., and A. H. Reisner.** 1986. MBIS—an integrated system for the retrieval and analyses of sequence data from nucleic acids and proteins. Nucleic Acids Res. **14:**265–272.
6. **Burnette, S.** 1982. "Western blotting": electrophoresis of protein from sodium dodecyl sulphate polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. Anal. Biochem. **112:**195–203.
7. **Dayhoff, M. O., R. M. Schwartz, and B. C. Orcutt.** 1978. A model of evolutionary changes in proteins, p. 345–352. M. O. Dayhoff (ed.), Atlas of protein sequence and structure, vol. 5, suppl. 3. National Biomedical Research Foundation, Washington, D.C.
8. **de Jongh-Leuvenink, J., R. W. Vreede, J. H. Marcelis, M. de Vos, and J. Verhoef.** 1985. Detection of antibodies against lipopolysaccharides of *Escherichia coli* and *Salmonella* R and S strains by immunoblotting. Infect. Immun. **50:**716–720.
9. **Dente, L., G. Cesareni, and R. Cortese.** 1984. pEMBL: a new family of single stranded plasmids. Nucleic Acids Res. **11:** 1645–1655.
10. **Dougan, G., and M. Kehoe.** 1984. The minicell system as a method for studying expression from plasmid DNA. Methods Microbiol. **17:**233–258.
11. **Eliopoulos, E. E., A. J. Geddes, M. Brett, D. J. C. Pappin, and J. B. C. Findlay.** 1982. A structural model for the chromophore-binding domain of ovine rhodopsin. Int. J. Biol. Macromol. **4:**263–273.
12. **Frazer, A. C., and R. Curtis III.** 1975. Production, properties and utility of bacterial minicells. Curr. Top. Microbiol. Immunol. **69:**1–84.
13. **Germanier, R., and R. Furer.** 1975. Isolation and characterization of galE mutant Ty21a of Salmonella typhi: a candidate strain for a live, oral typhoid vaccine. J. Infect. Dis. **131:** 553–558.
14. **Gold, L., and G. Stormo.** 1987. Translation initiation, p. 1302–1307. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.

15. **Goldmann, R. C., and L. Leive.** 1980. Heterogeneity of antigenic side-chain length in lipopolysaccharide from *Escherichia coli* O111 and *Salmonella typhimurium* LT2. Eur. J. Biochem. **107**:145–153.
16. **Hitchcock, P. J., and T. M. Brown.** 1983. Morphological heterogeneity among *Salmonella* lipopolysaccharide chemotypes in silver-stained polyacrylamide gels. J. Bacteriol. **154**:269–277.
17. **Hong, G. F.** 1982. A systematic sequencing strategy. J. Mol. Biol. **158**:539–549.
18. **Jann, B., K. Reske, and K. Jann.** 1975. Heterogeneity of lipopolysaccharides. Analysis of polysaccharide chain length by sodium dodecylsulfate-polyacrylamide gel electrophoresis. Eur. J. Biochem. **60**:239–246.
19. **Jann, K., and B. Jann.** 1984. Structure and biosyntheses of O-antigens, p. 138–186. *In* E. T. Rietschel (ed.), Handbook of endotoxin, vol. 1. Elsevier Science Publishers, Amsterdam.
20. **Leinonen, M.** 1985. Serological methods for the study of bacterial surface antigens, p. 179–206. *In* T. K. Korhonen, E. A. Dawes, and P. H. Makela. (ed.), Enterobacterial surface antigens: methods for molecular characterisation. Elsevier Science Publishers, Amsterdam.
21. **Lemair, H. G., and B. Muller Hill.** 1986. Nucleotide sequences of the galE gene and the galT gene in E. coli. Nucleic Acids Res. **14**:7705–7711.
22. **Le Minor, L., and M. Y. Popoff.** 1987. Designation of *Salmonella enterica* sp. nov., nom. rev., as the type and only species of the genus *Salmonella*. Int. J. Syst. Bacteriol. **37**:465–468.
23. **Levinthal, M., and H. Nikaido.** 1969. Consequences of deletion mutations joining two operons of opposite polarity. J. Mol. Biol. **42**:511–520.
24. **Lugtenberg, B., J. Meijers, R. Peters, P. van der Hoek, and L. van Alphen.** 1975. Electrophoretic resolution of the major outer membrane proteins of *Escherichia coli* K-12 into four bands. FEMS Microbiol. Lett. **58**:254–258.
25. **Makela, P. H., and B. A. D. Stocker.** 1984. Genetics of lipopolysaccharide. Handb. Endotoxin **1**:59–137.
26. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. **Matsuhashi, S., and J. L. Strominger.** 1967. Enzymatic synthesis of cytidine diphosphate-3,6-dideoxyhexoses. IV. Partial resolution of the enzymes required for the reduction of cytidine diphosphate-4-keto-6-deoxy-D-glucose and isolation of a new intermediate. J. Biol. Chem. **242**:3494–3500.
28. **Muto, A., and S. Osawa.** 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. Proc. Natl. Acad. Sci. USA **84**:166–169.
29. **Nikaido, H., M. Levinthal, K. Nikaido, and K. Nakane.** 1967. Extended deletions in the histidine-rough-B region of the *Salmonella* chromosome. Proc. Natl. Acad. Sci. USA **57**:1825–1832.
30. **Norrander, J., T. Kempe, and J. Messing.** 1983. Construction of improved M13 vectors using oligonucleotide-directed mutagensis. Gene **26**:101–106.
31. **Ornellas, E. P., and B. A. D. Stocker.** 1974. Relation of lipopolysaccharide character to P1 sensitivity in *Salmonella typhimurium*. Virology **60**:491–502.
32. **Palva, T., and P. H. Makela.** 1980. Lipopolysaccharide heterogeneity in *Salmonella typhimurium* analyzed by sodium dodecylsulfate-polyacrylamide gel electrophoresis. Eur. J. Biochem. **107**:137–143.
33. **Pape, H., and J. L. Strominger.** 1969. Enzymatic synthesis of cytidine diphosphate 3,6-dideoxyhexoses. V. Partial purification of the two protein components required for introduction of the 3-deoxygroup. J. Biol. Chem. **244**:3598–3604.
34. **Reissner, A. H., and C. A. Bucholtz.** 1986. The MTX package of computer programmes for the comparison of sequences of nucleotides and amino acid residues. Nucleic Acids Res. **14**: 233–238.
35. **Rossman, M. G., A. Liljas, C. I. Brenden, and L. J. Banaszak.** 1975. Evolutionary and structural relationships among dehydrogenases, p. 61–102. *In* P. D. Boyer (ed.), The enzymes, 3rd ed. Academic Press, Inc., New York.
36. **Sanderson, K. E., and J. R. Roth.** 1988. Linkage map of *Salmonella typhimurium*, edition VII. Microbiol. Rev. **52**:485–532.
37. **Sangar, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe.** 1980. Cloning in single stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**:161–178.
38. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
39. **Sharp, P. M., and W. Li.** 1987. The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. J. Mol. Evol. **4**:222–230.
40. **Smith, T. F., and M. S. Waterman.** 1981. Comparison of biosequences. Adv. Appl. Maths **2**:482–489.
41. **Staden, R.** 1982. Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. Nucleic Acids Res. **12**:505–519.
42. **Staden, R.** 1986. The current status and portability of our sequence handling software. Nucleic Acids Res. **14**:217–231.
43. **Sueoka, N.** 1988. Directional mutation pressure and neutral molecular evolution. Proc. Natl. Acad. Sci. USA **85**:2653–2657.
44. **Verma, N. K., N. B. Quigley, and P. R. Reeves.** 1988. O-antigen variation in *Salmonella* spp.: *rfb* gene clusters of three strains. J. Bacteriol. **170**:103–107.
45. **Verma, N. K., and P. R. Reeves.** 1989. Identification and sequence of *rfbS* and *rfbE*, which determine antigenic specificity of group A and group D salmonellae. J. Bacteriol. **171**:5694–5701.
46. **Walsh, C.** 1979. Enzymatic reaction mechanisms, p. 345–351 and 550–553. W. H. Freeman and Co., New York.
47. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.
48. **Yuasa, R., M. Levinthal, and H. Nikaido.** 1969. Biosynthesis of cell wall lipopolysaccharide mutants of *Salmonella*. V. A mutant of *Salmonella typhimurium* defective in the synthesis of cytidine diphosphoabequose. J. Bacteriol. **100**:433–444.