JOURNAL OF BACTERIOLOGY, Mar. 1993, p. 1524–1527
0021-9193/93/051524-04$02.00/0

Vol. 175, No. 5

NOTES

The *rfaS* Gene, Which Is Involved in Production of a Rough Form of Lipopolysaccharide Core in *Escherichia coli* K-12, Is Not Present in the *rfa* Cluster of *Salmonella typhimurium* LT2

JOHN D. KLENA,† ELIZABETH PRADEL,‡ AND CARL A. SCHNAITMAN*

Department of Microbiology, Arizona State University, Tempe, Arizona 85287-2701

Received 14 October 1992/Accepted 14 December 1992

Partial sequencing of the *rfa* cluster of *Salmonella typhimurium* LT2 indicated a region of 336 bp between *rfaP* and *rfaB* in the site occupied by the *rfaS* gene in *Escherichia coli* K-12. This region does not contain a functional *rfaS* gene, although DNA analysis suggests that the region may have contained an ancestral gene. This conclusion that *S. typhimurium* LT2 lacks *rfaS* is supported by its lipopolysaccharide (LPS) gel phenotype, since LT2 does not make the lipooligosaccharide band characteristic of LPS from smooth strains of *E. coli* K-12.

On the basis of DNA sequence analysis, it has been proposed that the *rfb* gene cluster, which is involved in synthesis of the O-specific polysaccharide portion of lipopolysaccharide (LPS), has evolved by the joining together of segments or blocks of genes which have different origins or histories (5). This is suggested by the presence of discrete blocks of genes which differ in A+T content from the flanking DNA. In broad terms, these blocks contain genes which are not conserved within the genus *Salmonella* and which are involved in synthesizing and assembling the specific sugars which define the different O serovars (1, 2, 5, 9, 15).

A similar phenomenon has been found in a comparison of the *rfaK* and *rfaL* genes of *Escherichia coli* K-12 and *Salmonella typhimurium* LT2 (8). In both organisms, the segment of the *rfa* gene cluster containing *rfaK* and *rfaL* has a high A+T content in comparison with that of flanking regions. In contrast to many *rfa* genes, *rfaK* and *rfaL* encode proteins whose primary amino acid sequences are not significantly conserved between the two organisms. The RfaK and RfaL proteins are involved in modifying the core so that O-specific polysaccharides can be attached (7), and thus they act in concert with functions encoded by the *rfb* gene cluster.

The segment of the *E. coli* K-12 *rfa* cluster between *rfaP* and *rfaB* which contains the *rfaS* gene has an exceptionally high A+T content compared with that of its flanking regions (11). Like *rfaK* and *rfaL*, the *E. coli* K-12 *rfaS* gene encodes a protein which acts in concert with *rfb* functions. In *E. coli* K-12 strain backgrounds which are derived from strain Y10, such as AB1133, *rfaS* mutations exhibit a significant phenotype only when the strain also contains a plasmid carrying a functional *rfb* cluster (7).

The *rfaS* phenotype involves the loss of a distinct form of rough (R) LPS. Rfa$^+$ *E. coli* K-12 strains transformed with a plasmid carrying the *rfb* and *rfp* genes of *Shigella dysenteriae* 1 produce a large amount of a new R LPS gel band in addition to a ladder of bands containing O antigen, and it is this R LPS band which is absent in an *rfaS* mutant (7). Since this *rfaS*- and *rfb*-dependent R LPS may be analogous in structure and function to the R LPS produced by nonenteric genera such as *Neisseria* and *Haemophilus*, we have termed it lipooligosaccharide (LOS) to distinguish it from the *rfaS*-independent R LPS species which are the predominant forms of LPS in *E. coli* K-12 strains without O-antigen plasmids (7).

The high A+T content of the *E. coli* K-12 *rfaS* gene and its involvement with an *rfb*-encoded function suggested that RfaS protein might not be conserved across species lines. To test this, we determined the sequence of the comparable region from the *rfa* cluster of *S. typhimurium* LT2. A 1.5-kb *Eco*RI-*Nru*I fragment (Fig. 1) was subcloned from plasmid pKZ26 (6), which has a 7.7-kb *Hin*dIII fragment from *S. typhimurium* LT2 carrying *rfaGBIJ*$^+$ (6). Translation of the sequence at the *Eco*RI end of the fragment gave a partial open reading frame (ORF) which showed regions of homology to the *E. coli* K-12 RfaB protein, while sequence from the *Nru*I end gave a partial ORF with homology to RfaP protein. This indicated that the 1.5-kb fragment spanned the junction between *rfaP* and *rfaB*.

The DNA sequence of an 815-bp region containing this junction is shown in Fig. 2. The distance between the *rfaP* termination codon and the initiation codon of *rfaB* is much shorter in *S. typhimurium* LT2 (333 bp) than in *E. coli* K-12 (1,015 bp). As for *rfaS* of *E. coli* K-12, the region between *rfaP* and *rfaB* from *S. typhimurium* LT2 is rich in A+T (71% A+T in LT2 compared with 73% in K-12). The 333-bp region between *rfaP* and *rfaB* contains one significant ORF of 76 amino acids which is read in the same direction as *rfaP* and *rfaB*. The initiation codon of this ORF is located almost the same distance from the termination codon of *rfaP* as the *rfaS* initiation codon in *E. coli* K-12 (33 bp compared with 38 bp in *E. coli*), and it is preceded by a reasonable ribosome binding site. This ORF shows short regions of partial homol-

* Corresponding author.
† Present address: Rocky Mountain Laboratories, National Institutes of Health, Hamilton, MT 59840.
‡ Present address: Laboratoire CEA-Jean Maetz, BP 68, 06230 Villefranche-sur-Mer, France.

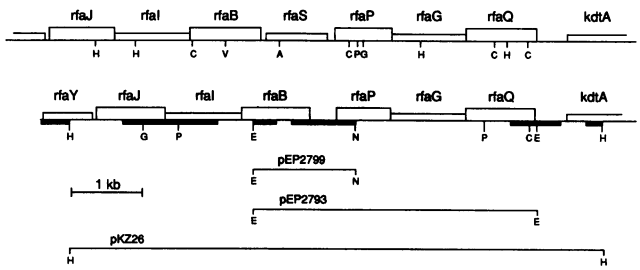

FIG. 1. Physical map of the *pyrE* end of the *rfa* clusters of *E. coli* K-12 and *S. typhimurium* LT2 (14). The top line shows a partial restriction map of the *kdtA-rfaY* region of *E. coli*. The open boxes are drawn to scale and show the coding regions of the genes deduced from the DNA sequence (7). The second line shows the corresponding region from *S. typhimurium*. The heavy lines indicate DNA regions which have been sequenced. The coding regions have been drawn the same size as their *E. coli* counterparts since the complete sequence is not known for any of the genes shown. The lower portion of the figure shows the inset of *S. typhimurium rfa* DNA from pKZ26 (6) and two fragments which were subcloned from this plasmid into pGEM vectors to generate plasmids pEP2793 and pEP2799 for DNA sequencing. The size of the region between the 3′ end of *rfaP* and the 5′ end of *rfaB* was the same in the chromosomal *rfa* cluster of *S. typhimurium* LT2 strain SL3770 as in plasmid pEP2799 as determined by polymerase chain reaction analysis. Restriction sites: A, *Acc*I; C, *Cla*I; E, *Eco*RI; G, *Bgl*II; H, *Hin*dIII; P, *Pst*I; N, *Nru*I; V, *Ava*I.

ogy to the *E. coli* K-12 RfaS protein, but alignment of these regions required the introduction of large gaps into the RfaS amino acid sequence (Fig. 3). There was no significant homology between the *E. coli* K-12 RfaS protein and the shorter ORFs from this region (not shown). In contrast, the C-terminal end of RfaP protein and the N-terminal end of RfaB showed substantial regions of homology between the two organisms (77% amino acid identity for RfaP and 83% for RfaB for the regions shown in Fig. 3).

Analysis of sequence of the *rfaP-rfaB* region and partial sequences of the *rfaQ-kdtA* region of *S. typhimurium* LT2 (data not shown) allowed a more precise placement of the genes at the *pyrE* end of the *S. typhimurium* LT2 *rfa* cluster in comparison with *E. coli* K-12 *rfa* (3, 8, 10, 11, 14). A composite map is shown in Fig. 1. Partial sequences of both the RfaQ and the KdtA proteins exhibited substantial regions of homology between the two organisms (more than 70% amino acid identity). In addition, we observed four blocks ranging from 15 to 30 bp within the noncoding region separating *kdtA* and *rfaQ* (3, 10) and one 30-bp block 18 amino acids downstream from the *rfaQ* initiation codon in which the DNA sequences from the two organisms were nearly identical (data not shown). This region is thought to contain the major promoters and regulatory regions for the divergent transcription of the long *rfa* hexose operon and the *kdtA* operon (12). The lengths of the *rfaQ*, *-G*, and *-P* genes must be nearly identical in both organisms, since the physical distance from the conserved *Cla*I site near the 5′ end of *rfaQ* to the 3′ end of *rfaP* appears to be identical. *rfaQ*, *-G*, and *-P* all exhibit efficient cross-species complementation (7, 10), suggesting that the proteins they encode are of similar structure in both organisms. The size of the *Eco*RI-*Hin*dIII fragment containing part of *rfaB*, *rfaIJ*, and part of *rfaY* of *S. typhimurium* LT2 agrees exactly with the sizes of these genes in *E. coli* K-12 (8, 11). There are no remaining gaps in the *S. typhimurium* LT2 *rfa* cluster which could accommodate an analog of *E. coli* K-12 *rfaS*.

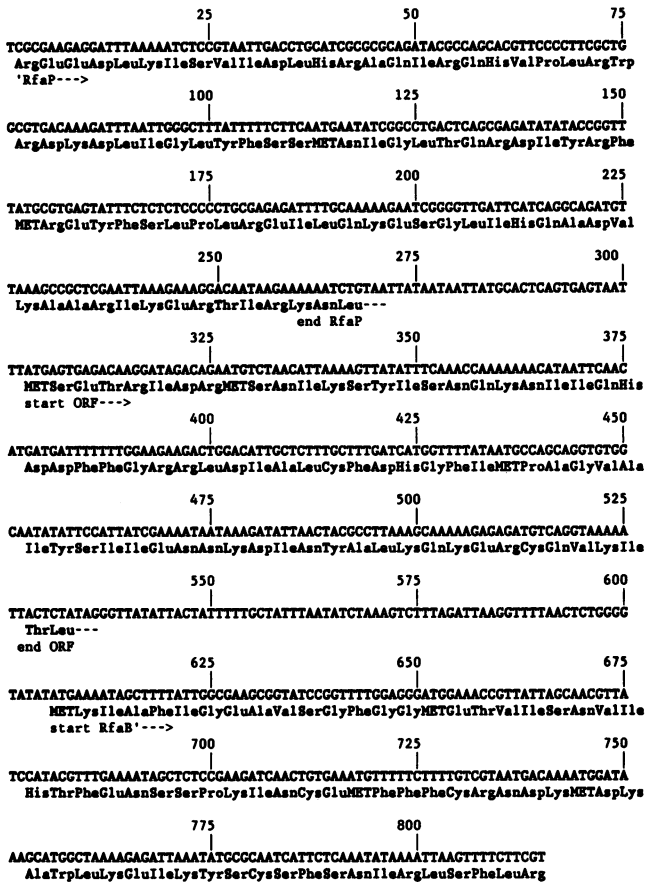

FIG. 2. Sequence of an 815-bp region from the *rfa* gene cluster of *S. typhimurium* LT2 extending from the 3′ end of the *rfaP* coding region into the 5′ end of the coding region of *rfaB*. The derived amino acid sequences of the C-terminal end of RfaP, the 76-amino-acid ORF, and the N-terminal end of RfaB are shown. Short ORFs upstream from the probable initiation codons of the ORF and RfaB are not shown. DNA sequencing and analysis using PC/GENE were done as described previously (11).

The high A+T content and the sequence and location of the ORF in the region between *rfaP* and *rfaB* of *S. typhimurium* LT2 suggest that this region may once have contained a gene analogous to *rfaS*. The gap between *rfaP* and *rfaB* is what has been termed a gray hole, a region containing what appears to be remnants of an ancestral gene but encoding no functional protein (4). To our knowledge, this is the first time evidence linking a gray hole to a functional gene in a related organism has been obtained. A distinct reduction in conservation of amino acid sequence was observed at the ends of the *rfaZ* and *rfaC* coding regions, which are adjacent to the nonconserved *rfaKL* block, and this was suggested to indicate ancestral sites of recombination (8). No similar regions of reduced amino acid sequence conservation between the two species were found at the N-terminal end of RfaB or the C-terminal end of RfaP, but this may reflect the presence of noncoding regions flanking the *E. coli* K-12 *rfaS* gene (11) and lying between the C-terminal end of *rfaP* and the 76-amino-acid ORF in *S. typhimurium* which are almost as large as the regions of reduced conservation which flank *rfaKL*. The question of whether RfaS protein is conserved across species lines cannot be answered definitively until its sequence in another organism is available.

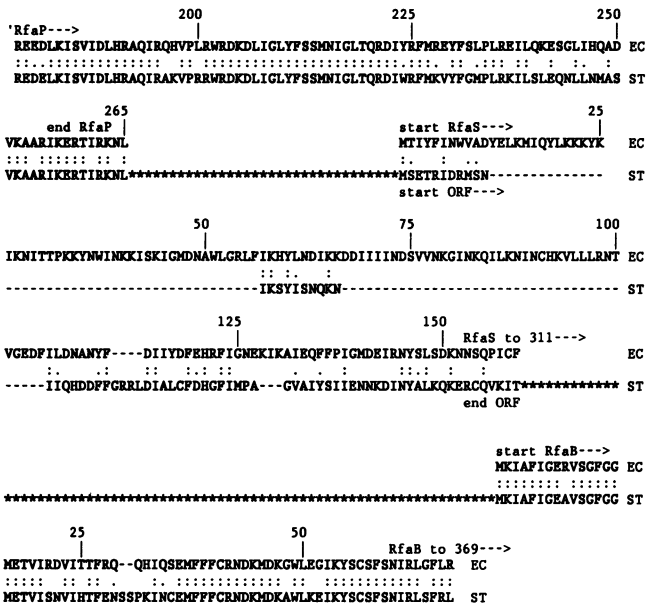

FIG. 3. Comparison of the derived amino acid sequences of RfaP, RfaS, and RfaB proteins of *E. coli* K-12 (EC) with the corresponding proteins and the 76-amino-acid ORF from *S. typhimurium* LT2 (ST). The numbering indicates the amino acid residues of the *E. coli* proteins (10, 11). Asterisks indicate regions in the *S. typhimurium* DNA sequence which do not contain significant coding regions, with each asterisk indicating 3 bp, and the dashed regions represent gaps in the amino acid sequence introduced to permit alignment. Colons indicate identical residues, and periods indicate conservative substitutions.

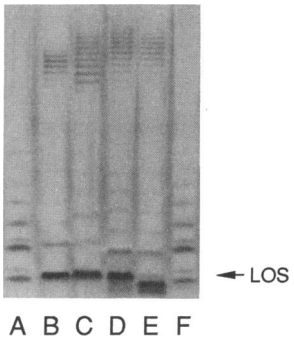

FIG. 4. Tricine-sodium dodecyl sulfate gel analysis of LPS from wild-type *S. typhimurium* LT2 (lanes A and F), two S isolates of *E. coli* (lane B, serotype O55:B5:H3; lane C, serotype O111:B4), *E. coli* K-12 expressing the O antigen of *S. dysenteriae* 1 (lane D), and a similar strain of *E. coli* K-12 with an *rfaS* mutation (lane E). The arrow indicates the LOS band in lane D and the similar R LPS band in lanes B and C. To enhance resolution of the lower bands, the gels were underloaded and a short development time was used. Under these conditions, the larger *Salmonella* bands (20 to 30 O units per molecule) were faint and are not visible in the photograph. Cultures were grown to mid-log phase at 30°C in Luria-Bertani broth, and LPS extraction and electrophoresis were as described previously (7). *S. typhimurium* LT2 strain SL3770 was from the Salmonella Genetic Stock Center, Calgary, Alberta, Canada; *E. coli* O55:B5:H3 ATCC 12014 was obtained from the American Type Culture Collection; and *E. coli* O111:B4 was from E. Heath. The *E. coli* K-12 strains used were CS1861, which is *rfa*$^+$ and carries plasmid pSS37 containing the *rfp* gene and the *rfb* gene cluster of *S. dysenteriae* 1 (7), and CS2684, which is the same as CS1861 except Δ*lac* and *rfaS2007*::Tn*lac* (7). The *E. coli* K-12 background strain CS180 is derived from AB1133.

LPS from *S. typhimurium* LT2 (Fig. 4, lanes A and F) does not have an abundant R band which resembles the LOS band of *E. coli* K-12. While this is consistent with a loss of *rfaS* function, it also does not have an abundant faster-migrating R band like that of an *rfaS* mutant of *E. coli* K-12 (Fig. 4, lane E). Instead, the core has a doublet of two much less abundant R bands which, on the basis of spacing, appear to be related to the doublets seen in the O-antigen ladder. The reason for the doublet bands in the O-antigen ladder of *S. typhimurium* is not known, but their regular spacing indicates that they result from the production of two different forms of LPS acceptor to which identical O units are attached and not from heterogeneity of the O units. These two R bands are not observed in gels of LPS from *S. typhimurium* LT2 R mutants (7), so their chemical structure in relation to the LPS produced by R mutants is unknown.

The LPS from two smooth (S) isolates of *E. coli* (Fig. 4, lanes B and C) has an abundant R band resembling the LOS band in LPS from *E. coli* K-12 (Fig. 4, lane D). Although not all of the S strains in our collection produce an abundant R band like the *E. coli* K-12 LOS band (data not shown), production of such a band appears to be a common feature among enteric bacteria. One of these S isolates, *E. coli* O111:B4, is the parent of J-5, a *galE* strain which has been widely used to study LPS biosynthesis and as a source of LPS for studies on the immunologic and pharmacologic effects of endotoxin.

For the strains which produce LOS, it is not known whether LOS represents a constant proportion of the LPS in all cells or whether individual cells have the ability to shift back and forth between the production of LOS and S LPS. Such a shift is feasible, since the production of LOS is controlled by elements in both the *rfa* and the *rfb* gene clusters (7). A heterogeneous population in which some members produce primarily LOS might have an advantage in colonizing some environmental niches by virtue of the exposure of surface proteins, such as adhesins or lectins, which might otherwise be inaccessible if the surface is covered by long chains of O-specific polysaccharide. Can a gene such as *rfaS* which is involved in LOS production contribute to the virulence of an organism or to the range of environments in which it can thrive? Might the lack of *rfaS* be one of the factors which contributes to the apparently mild virulence (13) of *S. typhimurium* LT2? It is interesting that *S. typhimurium* LT2, which has long been one of the "gold standards" in LPS research, may be atypical in terms of the range of LPS molecules it can produce.

We thank P. R. MacLachlan and K. Sanderson for providing plasmids and DNA sequence.

This research was supported by Public Health Services grant GM-39087.

REFERENCES

1. **Brown, P. K., L. K. Romana, and P. R. Reeves.** 1991. Cloning of the *rfb* gene cluster of a group C2 *Salmonella* strain: comparison with the *rfb* regions of groups B and D. Mol. Microbiol. **5**:1873–1881.
2. **Brown, P. K., L. K. Romana, and P. R. Reeves.** 1992. Molecular analysis of the *rfb* gene cluster of *Salmonella* serovar muenchen (strain M67): the genetic basis of the polymorphism between groups C2 and B. Mol. Microbiol. **6**:1385–1394.
3. **Clementz, T., and C. R. H. Raetz.** 1991. A gene coding for 3-deoxy-D-manno-octulosonic-acid transferase in *Escherichia*

coli. J. Biol. Chem. **266**:9687–9696.

4. **Daniels, D. L., G. Plunkett III, V. Burland, and F. R. Blattner.** 1992. Analysis of the *Escherichia coli* genome: DNA sequence of the region from 84.5 to 86.5 minutes. Science **257**:771–778.
5. **Jiang, X.-M., B. Neal, F. Santiago, S. J. Lee, L. K. Romana, and P. R. Reeves.** 1991. Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2). Mol. Microbiol. **5**:695–713.
6. **Kadam, S. K., A. Rehemtulla, and K. E. Sanderson.** 1985. Cloning of *rfaG*, *B*, *I*, and *J* genes for glycosyltransferase enzymes for synthesis of the lipopolysaccharide core of *Salmonella typhimurium*. J. Bacteriol. **161**:277–284.
7. **Klena, J. D., R. S. Ashford II, and C. A. Schnaitman.** 1992. Role of *Escherichia coli* K-12 *rfa* genes and the *rfp* gene of *Shigella dysenteriae* 1 in generation of lipopolysaccharide core heterogeneity and attachment of O antigen. J. Bacteriol. **174**:7297–7307.
8. **Klena, J. D., E. Pradel, and C. A. Schnaitman.** 1992. Comparison of the lipopolysaccharide biosynthesis genes *rfaK*, *rfaL*, *rfaY*, and *rfaZ* of *Escherichia coli* K-12 and *Salmonella typhimurium*. J. Bacteriol. **174**:4746–4752.
9. **Liu, D., N. K. Verma, L. K. Romana, and P. R. Reeves.** 1991. Relationships among the *rfb* regions of *Salmonella* serovars A, B, and D. J. Bacteriol. **173**:4814–4819.
10. **Parker, C. T., E. Pradel, and C. A. Schnaitman.** 1992. Identification and sequences of the lipopolysaccharide core biosynthetic genes *rfaQ*, *rfaP*, and *rfaG* of *Escherichia coli* K-12. J. Bacteriol. **174**:930–934.
11. **Pradel, E., C. T. Parker, and C. A. Schnaitman.** 1992. Structures of the *rfaB*, *rfaI*, *rfaJ*, and *rfaS* genes of *Escherichia coli* K-12 and their roles in the assembly of the lipopolysaccharide core. J. Bacteriol. **174**:4736–4745.
12. **Roncero, C., and M. J. Casadaban.** 1992. Genetic analysis of the genes involved in the synthesis of the lipopolysaccharide core in *Escherichia coli* K-12: three operons in the *rfa* locus. J. Bacteriol. **174**:3250–3260.
13. **Sanderson, K. E., and B. A. D. Stocker.** 1987. *Salmonella typhimurium* strains used in genetic analysis, p. 1220–1224. *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.
14. **Schnaitman, C. A., C. T. Parker, J. D. Klena, E. L. Pradel, N. B. Pearson, K. E. Sanderson, and P. R. MacLachlan.** 1991. Physical maps of the *rfa* loci of *Escherichia coli* K-12 and *Salmonella typhimurium*. J. Bacteriol. **173**:7410–7411.
15. **Wang, L., L. K. Romana, and P. R. Reeves.** 1992. Molecular analysis of a *Salmonella enterica* group E1 *rfb* gene cluster: O antigen and the genetic basis of the major polymorphism. Genetics **130**:429–443.