JOURNAL OF BACTERIOLOGY, Dec. 1991, p. 7410–7411
0021-9193/91/237410-02$02.00/0

Vol. 173, No. 23

Physical Maps of the *rfa* Loci of *Escherichia coli* K-12 and *Salmonella typhimurium*

CARL A. SCHNAITMAN,[1]* CRAIG T. PARKER,[1] JOHN D. KLENA,[1] ELIZABETH L. PRADEL,[1] NANCY B. PEARSON,[1] KENNETH E. SANDERSON,[2] AND P. RONALD MACLACHLAN[2]†

Department of Microbiology, Arizona State University, Tempe, Arizona, 85287,[1] and Department of Biological Science, University of Calgary, Calgary, Alberta, Canada T2N 1N4[2]

The *rfa* loci at min 81 of the *Escherichia coli* K-12 map (3) and min 79 of the *Salmonella typhimurium* map (14) are clusters of genes involved in synthesis of the core region of lipopolysaccharide. The functions of genes at this locus have been extensively studied (for reviews, see references 11 and 13). The *rfa* loci of the two organisms have been cloned and partially sequenced, and there is enough overlap of sequenced regions from both organisms to allow us to assemble physical maps for the loci (Fig. 1). In *E. coli* K-12, the sequences of the *kbl* (1) gene, which flanks the locus to the left, and the *fpg* (4) and *rpmBG* (10) genes, which flank the locus to the right, have been published.

The *Salmonella* genes *rfaD*, *-F*, *-C*, *-L*, *-K*, *-J*, and *-I* were assigned to open reading frames by the ability of plasmids bearing *rfa* restriction fragments to complement standard *Salmonella* mutations, and both the direction of transcription and the limits of these genes were additionally defined by transposon mutagenesis of the plasmid inserts. The sequence and assignment to an open reading frame of *E. coli* K-12 *rfaD* has been described (12). *E. coli* genes *rfaF*, *-C*, *-L*, *-K*, *-Z*, and *-Y* were identified by the homology of the derived protein sequences to those of their *Salmonella* counterparts. *E. coli* genes *rfaK*, *-B*, *-P*, and *-G* were assigned by complementation of standard *Salmonella* mutations by *E. coli* plasmids introduced into *S. typhimurium* by electroporation. These assignments have also been confirmed by the phenotypes of mutants obtained by crossing mutations constructed in vitro into the *E. coli* chromosome. Experimental details of the cloning, sequencing, sequence comparison, and complementation studies will be described elsewhere.

Assignment of the *E. coli* K-12 counterparts of *Salmonella* genes *rfaI* and *-J* presented a problem because of the difference in structure of the hexose region between the two organisms (for a discussion of this problem, see reference 2, p. 5315 to 5316). Since these genes are thought to encode different sugar transferases in the two organisms, it was proposed that the *E. coli* K-12 genes be designated *rfaM* and *-N* (2, 7). However, in *E. coli* K-12 the two genes which occupy the same positions with respect to flanking genes as *rfaI* and *-J* of *S. typhimurium* show considerable homology at the protein level to *rfaI* and *-J* from *S. typhimurium* (5), and restriction fragments from *E. coli* K-12 which contained both of these genes complemented either *rfaI* or *rfaJ* mutants of *S. typhimurium*. Thus, these pairs of genes appear to be very similar in structure and function even though they exhibit different sugar specificites, and on this basis we propose that they be termed *rfaI* and *-J* for both organisms.

A number of open reading frames which did not correspond to known genes were found. We have assigned gene designations to those which appear to have a function in lipopolysaccharide biosynthesis or a distinctive relationship to other *rfa* genes. The two open reading frames located between *rfaJ* and *-K* have been designated *rfaY* and *-Z*. The protein sequences of these open reading frames are strongly conserved between the two organisms, and preliminary analysis indicates that lack of these genes results in a slightly truncated core lipopolysaccharide as indicated by gel migration. The protein encoded by the open reading frame lying between *rfaG* and *kdtA* exhibits significant homology to a family of inner core biosynthesis genes including *rfaC*, *rfaF*, and *kdtA*, and this gene has been designated *rfaQ*. It is likely that the gene for the 38-kDa polypeptide between *rfaB* and *rfaP* and the gene for the 18-kDa polypeptide located adjacent to *kdtA* (6) will also have *rfa* functions. The situation is less clear with respect to the open reading frame for a 26-kDa polypeptide (sequence given in references 1 and 12) located between *rfaD* and *kbl*. If all of these are *rfa* genes, the locus contains at least 17 genes.

The *rfa* locus consists of three distinct blocks of genes as defined by the direction of transcription. At the left end is a block of genes consisting of *rfaDFCL* which is transcribed rightward. There is no evidence that these are transcribed as a single operon, and there may be a gap between *rfaD* and *rfaF*. In the middle of the locus is a large block of about 10 genes consisting of *rfaQGP--BIJYZK* which are transcribed leftward, and evidence has been presented that some or all of these are organized into a complex operon with internal promoters (2). At the right end, the *kdtA* gene involved in attachment of ketodeoxyoctulosonic acid and the 18-kDa polypeptide gene are transcribed rightward, and the close positioning of these genes suggests that they might constitute an operon (6). There is an AT-rich gap of about 400 bp between *rfaQ* and *kdtA* which could accomodate a complex promoter region.

The most recent *E. coli* K-12 linkage map (3) lists six genes between *rpmGB* and *rfa*. Of these, *envC* is now known to be located closer to *mtl* (8). The phenotype of *pcsA* mutants includes cold-sensitive inhibition of cell division and sensitivity to detergents. We have found a similar phenotype associated with some *rfaG* mutants, indicating that this mutation may be allelic to *rfaG*. The status of genes *gadRS*, *radC*, and *mutM* with respect to this region is unknown. We wish to point out that the direction of transcription of *kbl* and *tdh* is incorrect in the current map (3).

ACKNOWLEDGMENTS

We acknowledge our colleagues who have generously provided DNA sequences prior to publication. S. Raina and C. P. Georgopoulos provided the sequences of the *E. coli* 26-kDa protein and *rfaD*. A. Wright provided the sequence of the *E. coli* gene *rfaC* and partial sequences of *rfaF* and *rfaL*. C. R. Raetz provided the sequences of *kdtA* and the 18-kDa protein from *E. coli*.

* Corresponding author.
† Present address: Department of Microbiology, University of Guelph, Guelph, Ontario, Canada M1G 2W1.

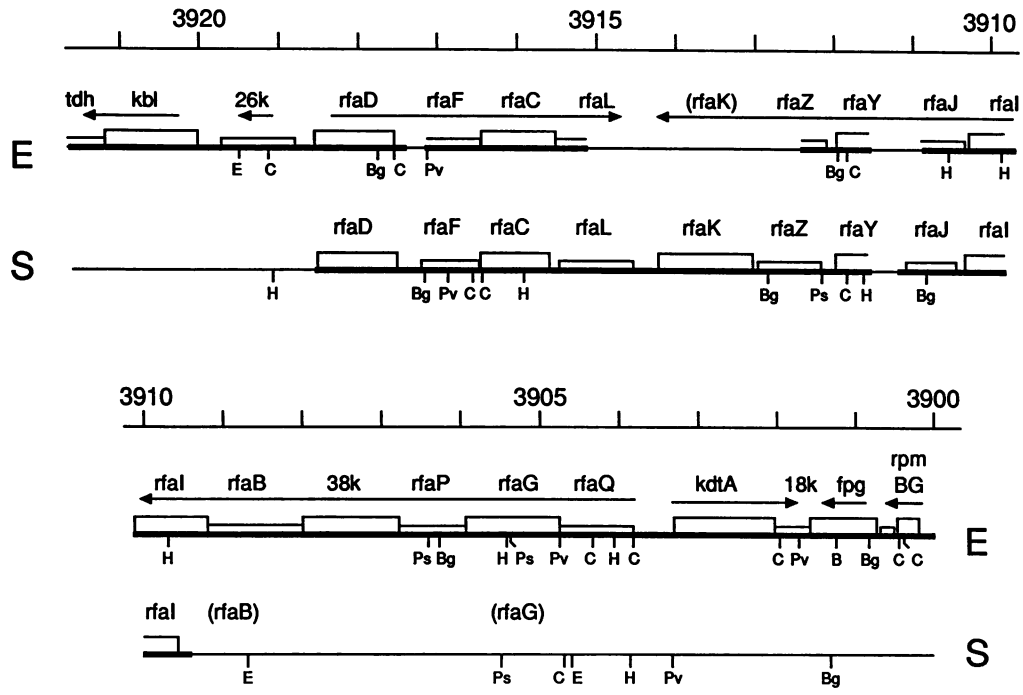

FIG. 1. Physical maps of the *rfa* loci of *E. coli* K-12 and *S. typhimurium*. The loci are about 18 kb (0.4 min) in length. The scale indicates distance in kilobases, and the numbers on the scale are approximate kilobase coordinates of the *E. coli* K-12 restriction map of Kohara et al. (9). It should be noted that *rfa* lies within the inversion in strain W3110 which was used to construct that map (3, 9), and thus the numbering is reversed. The lines marked E and S indicate the physical maps of *E. coli* K-12 and *S. typhimurium*, and the heavy lines indicate regions which have been completely sequenced. The boxes indicate the open reading frames of genes deduced from the DNA sequence, and the genes and gaps between them are drawn approximately to scale. The incomplete boxes indicate open reading frames which terminate where the sequence is not known. Genes are indicated in brackets when the location is known only from assignment to restriction fragments. Arrows below the *E. coli* map indicate the direction of transcription of genes or blocks of genes, which appear to be the same in both organisms. Restriction sites are indicated as follows: H, *Hin*dIII; E, *Eco*RI; Ps, *Pst*; Bg, *Bgl*II; Pv, *Pvu*II; C, *Cla*I; B, *Bam*HI. All of the known sites for these enzymes are shown. The maps are aligned around a *Cla*I site in *rfaY* which the sequence indicates is shared by both organisms. The *E. coli* genes designated *rfaQ*, *-G*, and *-P* correspond, respectively, to the 42-, 39-, and 35-kDa polypeptide genes described previously (2).

We acknowledge research support from an Operating Grant from the Natural Sciences and Engineering Research Council of Canada (to K.E.S.) and grant GM-39087 from the National Institutes of Health (to C.A.S.).

REFERENCES

1. **Aronson, B. D., P. D. Ravnikar, and R. L. Somerville.** 1988. Nucleotide sequence of the 2-amino-3-ketobutyrate coenzyme A ligase (*kbl*) gene of *E. coli*. Nucleic Acids Res. **16**:3586.
2. **Austin, E. A., J. F. Graves, L. A. Hite, C. T. Parker, and C. A. Schnaitman.** 1990. Genetic analysis of lipopolysaccharide core biosynthesis by *Escherichia coli* K-12: insertion mutagenesis of the *rfa* locus. J. Bacteriol. **172**:5312–5325.
3. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54**:130–197.
4. **Boiteux, S., T. R. O'Connor, and J. Laval.** 1987. Formamidopyrimidine-DNA glycosylase of *Escherichia coli*: cloning and sequencing of the *fpg* structural gene and overproduction of the protein. EMBO J. **6**:3177–3183.
5. **Carstenius, P., J.-I. Flock, and A. Lindberg.** 1990. Nucleotide sequence of *rfaI* and *rfaJ* genes encoding lipopolysaccharide glycosyl transferases from *Salmonella typhimurium*. Nucleic Acids Res. **18**:6128.
6. **Clementz, T., and C. R. H. Raetz.** 1991. A gene coding for KDO transferase in *Escherichia coli*: identification, mapping, cloning and sequencing. J. Biol. Chem. **266**:9687–9696.
7. **Creeger, E. S., and L. I. Rothfield.** 1979. Cloning of genes for bacterial glycosyl transferases. I. Selection of hybrid plasmids carrying genes for two glycosyl transferases. J. Biol. Chem. **254**:804–810.
8. **Klein, R. K., B. Henrich, and R. Plapp.** 1990. Molecular cloning of the *envC* gene of *Escherichia coli*. Curr. Microbiol. **21**:341–347.
9. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50**:495–508.
10. **Lee, J. S., G. An, J. D. Friesen, and K. Isono.** 1981. Cloning and the nucleotide sequence of the genes for *Escherichia coli* ribosomal proteins L28 (*rpmB*) and L33 (*rpmG*). Mol. Gen. Genet. **184**:218–223.
11. **Makela, P. H., and B. A. D. Stocker.** 1984. Genetics of lipopolysaccharide, p. 59–137. *In* E. T. Rietschel (ed.), Handbook of endotoxin, vol. 1. Chemistry of endotoxin. Elsevier, Amsterdam.
12. **Pegues, J. C., L. Chen, A. W. Gordon, L. Ding, and W. G. Coleman, Jr.** 1990. Cloning, expression, and characterization of the *Escherichia coli* K-12 *rfaD* gene. J. Bacteriol. **172**:4652–4660.
13. **Rick, P. D.** 1987. Lipopolysaccharide biosynthesis, p. 648–662. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Lowr, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 1. American Society for Microbiology, Washington, D.C.
14. **Sanderson, K. E., and J. R. Roth.** 1988. Linkage map of *Salmonella typhimurium*, edition 7. Microbiol. Rev. **52**:485–532.