JOURNAL OF BACTERIOLOGY, Aug. 1992, p. 5228–5236
0021-9193/92/165228-09$02.00/0

Vol. 174, No. 16

# Nucleotide Sequences of the Genes Regulating O-Polysaccharide Antigen Chain Length (*rol*) from *Escherichia coli* and *Salmonella typhimurium*: Protein Homology and Functional Complementation

ROGER A. BATCHELOR,[1]* PIETRO ALIFANO,[2] ELIO BIFFALI,[2]† SHEILA I. HULL,[1] AND RICHARD A. HULL[1]

*Department of Microbiology and Immunology, Baylor College of Medicine, Texas Medical Center, Houston, Texas 77071,[1] and Centro di Endocrinologia ed Oncologia Sperimentale del Consiglio Nazionale Delle Ricerche and Dipartimento di Biologia e Patologia Cellulare e Molecolare, Università di Napoli, via S. Pansini 5, 80131 Naples, Italy[2]*

Received 18 March 1992/Accepted 27 May 1992

**In this article, we report on the nucleotide sequences of the *rol* genes of *Escherichia coli* O75 and *Salmonella typhimurium* LT2. The *rol* gene in *E. coli* was previously shown to encode a 36-kDa protein that regulates size distribution of the O-antigen moiety of lipopolysaccharide. The *E. coli* and *S. typhimurium rol* gene sequences consist of 978 and 984 nucleotides, respectively. The homology between the nucleotide sequences of these two genes was found to be 68.9%. Both the *E. coli rol* and *S. typhimurium rol* genes are transcribed counter to the histidine operon and code for deduced polypeptides of 325 and 327 amino acids, respectively. The *S. typhimurium rol* gene was previously identified to encode a protein of unknown function and to share a transcription termination region with *his*. The homology between these deduced polypeptide sequences was observed to be 72%. A complementation test was performed in which the *S. typhimurium rol* gene was placed in *trans* with an *E. coli* plasmid (pRAB3) which encodes the O75 *rfb* gene cluster and not *rol*. The protein expressed from the *S. typhimurium rol* gene was found to regulate the distribution of the O75 O polysaccharide on the lipopolysaccharide of the host strain, *E. coli* SØ874. The mechanism of Rol action may be independent of O antigen subunit structure, and its presence may be conserved in members of the family *Enterobacteriaceae* and other gram-negative bacilli that express O polysaccharides on their surface membrane.**

The O polysaccharides, or somatic antigen structures, of members of the family *Enterobacteriaceae* and species of the genera *Pseudomonas* and *Vibrio* consist of repeated oligosaccharide subunits. These subunits may be composed of one to seven sugar residues linked in a variety of arrangements and are frequently chemically modified to establish serotypic specificity among these bacteria. These units may form a polymer or chain greater than 40 O units in length (25). The O polysaccharide is covalently linked to the core oligosaccharide/lipid A structure to form lipopolysaccharide (LPS). LPS appears as a series of bands with a ladderlike pattern on silver-stained or radiolabeled sodium dodecyl sulfate (SDS)-polyacrylamide gels. Each band on the gel corresponds to a lipid A and core oligosaccharide molecule attached to an O chain of a certain size. In LPS preparations of normal or wild-type bacteria, these bands have a bimodal or multimodal appearance due to the expression of preferred sizes of O antigen on the core/lipid A structure (18, 20, 24, 42, 49, 51).

The O-polysaccharide moieties of LPS have been demonstrated to mediate many biological effects, such as resistance to killing by normal, nonspecific serum (43, 53, 62), resistance to phagocytosis by monocytes, and resistance to killing by cationic peptides (54, 64). O-chain antigens are also important for attachment of bacteriophages and may be involved in the interaction with eucaryotic cell membranes (48, 65). In *Escherichia coli*, *Salmonella* spp., *Klebsiella pneumoniae*, *Serratia marcescens*, and *Pseudomonas aeruginosa*, the biological activities of the O chain have been correlated to its size and distribution on the surface of the bacterial outer membrane (19, 21, 26, 28, 35, 39, 52). The longer O chains are more important in mediating the biological effects of these gram-negative bacteria.

In *Salmonella typhimurium* and *E. coli*, the *rfb* genetic region encodes the enzymes needed for O-specific monosaccharide synthesis and for the transferases necessary for incorporation of these sugar residues into the O-antigen subunit structure. The *rfb* gene cluster maps near the histidine operon in both *S. typhimurium* and *E. coli*, at map positions 42 and 44, respectively (3, 38, 57). Other important genetic regions involved in O-antigen biosynthesis include a polymerase (*rfc*) at map position 31 in *S. typhimurium* and a ligase enzyme possibly consisting of two proteins, encoded by the *rfbT* and *rfaL* genes, located at map positions 42 and 79, respectively, of *S. typhimurium* (57). The polymerase is thought to link the O subunits together into a chain, and the ligase may join the polymerized O chain onto the core oligosaccharide (25).

Recently, we described a new protein or enzyme to be included in this group of O-antigen biosynthesis genes. This protein was found encoded in a region adjacent to the O75 *rfb* region of *E. coli* and appears to regulate the size distribution of the O-antigen chain onto the core oligosaccharide (7). We named this gene *rol* for regulator of O length. In this article, we report on the DNA sequence of this gene

* Corresponding author.
† Present address: Dipartimento di Biochemica e Biologia Molecolare, Stazione Zoologia "A. Dohrn," Villa Comunale, 80121, Naples, Italy.

TABLE 1. Plasmids and *E. coli* strains used in this study

| Plasmid or strain | Description | Reference or source |
|---|---|---|
| Plasmids | | |
| pACYC184 | Cloning vector | 14 |
| pBluescript KSII | Cloning vector | Stratagene |
| pBR322 | Cloning vector | 9 |
| pRAB1 | 17.5-kb *Sal*I fragment (*rol*$^+$ *rfb*$^+$) in pACYC184 | 7 |
| pRAB3 | 13.4-kb *Eco*RI-*Xho*I fragment (*rol rfb*$^+$) in pBluescript KSII | 7 |
| pRAB15 | 1.77-kb *Ava*I-*Afl*II fragment (*rol*$^+$ *rfb*) in pBluescript KSII | This study |
| pHS8500 | 8.5-kb *Eco*RI fragment (*his*$^+$ *rol*$^+$ *rfb*) in pBR322 | 12, 15 |
| pRAB20 | 6.5-kb *Pst*I fragment from pHS8500 in pBR322 | This study |
| Strains | | |
| SK2881 | *hsdR4 recA1 endA sbcB15* | 29 |
| SØ874 | Δ(*sbc-rfb*), O75$^-$ | 47 |
| HU1124 | A29 Rif$^r$ F′ Kan$^r$, O75$^+$ | 34 |
| RAB1 | SØ874(pRAB1), Cm$^r$ O75$^+$ | 7 |
| RAB3 | SØ874(pRAB3), Ap$^r$ O75$^+$ | 7 |
| RAB15 | SØ874(pRAB15), Ap$^r$ O75$^-$ | This study |
| RAB20 | SØ874(pRAB20), Tc$^r$ O75$^-$ | This study |
| RAB23 | SØ874(pRAB20, pRAB3), Ap$^r$ Tc$^r$ O75$^+$ | This study |

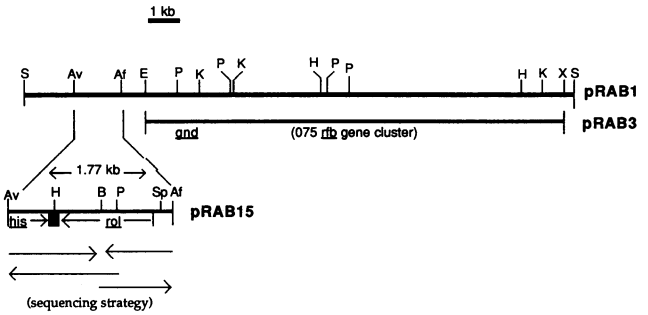



FIG. 1. Physical map of the insert DNA of pRAB1 and its deletion derivatives, pRAB3, encoding the *gnd* gene and O75 *rfb* region, and pRAB15, encoding the *rol* gene. The *rol*-encoding DNA and its flanking regions were sequenced from pRAB15 and/or its deletion subclones as indicated by the arrows. Arrows adjacent to the genes indicate the direction of transcription. The black box between *rol* and *his* represents the region of potential transcription termination. The position of the *gnd* gene was determined as described in Results. Restriction endonuclease sites: S, *Sal*I; E, *Eco*RI; K, *Kpn*I; X, *Xho*I; Av, *Ava*I; Af, *Afl*II; Sp, *Spe*I; H, *Hin*dIII; B, *Bgl*II; P, *Pst*I. All *Pst*I, *Hin*dIII, *Kpn*I, *Xho*I, *Eco*RI, and *Sal*I sites are shown. The *Bgl*II, *Spe*I, *Ava*I, and *Afl*II sites have been mapped only on the pRAB15 plasmid.

and its deduced polypeptide sequence characteristics and discuss its structural and functional homology to a similar protein in *S. typhimurium*.

## MATERIALS AND METHODS

**Bacterial strains and plasmids.** The plasmids and *E. coli* strains used in this study are listed in Table 1. Plasmid preparations were made by a large- or small-scale cleared-lysate method (60). Transformations were performed by the cold $CaCl_2$ method as described by Brown et al. (10). *E. coli* SK2881 was used as an initial cloning host for the plasmids encoding *rfb* and *rol* DNA (29). Subsequently, the DNA was moved into an *E. coli* SØ874 strain (47). This strain carries a deletion of the *his-rfb* region and is very useful for characterization of LPS expression. Bacteria were grown on L agar or L broth with 0.5% glucose and supplemented with ampicillin (100 μg/ml), tetracycline (20 μg/ml), or chloramphenicol (20 μg/ml). The cloning vectors used were pACYC184 (14), pBluescript KSII (Stratagene Cloning Systems, La Jolla, Calif.), and pBR322 (9).

**Enzymes and buffers.** Restriction endonucleases were obtained from Bethesda Research Laboratories, Inc. (Gaithersburg, Md.), Boehringer Mannheim (Indianapolis, Ind.), and New England Biolabs, Inc. (Beverly, Mass.). Klenow enzyme was purchased from Boehringer Mannheim, and T4 DNA ligase was purchased from Bethesda Research Laboratories. Avian myeloblastosis virus reverse transcriptase and RNasin were obtained from Promega Inc. (Madison, Wis.). Sequenase version 2.0, dideoxynucleotides, and sequencing reagents and buffers were obtained from the U.S. Biochemical Corporation (Cleveland, Ohio). The conditions and buffers used were those recommended by the manufacturers.

**Sequencing the *E. coli rol* and *S. typhimurium rol* genes.** As shown in Fig. 1, a 1.77-kb *Ava*I-*Afl*II fragment which was previously shown to encode the *rol* gene product was removed from pRAB1; the ends were filled in with Klenow fragment, and the fragment was ligated into the *Eco*RV site of the pBluescript KSII vector to produce pRAB15. Either single- or double-stranded insert DNA from pRAB15 and its deletion clones was sequenced by the dideoxynucleotide method (58) with the strategy shown in Fig. 1. The sequencing of the *S. typhimurium* DNA was done as previously described (11–13, 15).

**DNA sequence analysis.** DNA and protein sequence analysis was performed with the EuGene software package developed by the Molecular Biology Information Resource at Baylor College of Medicine (33). The Lawrence and Goldman program was used for DNA homology and alignment (32). Another program was used to align deduced polypeptide sequences (2, 46). GenBank searches for DNA and peptide sequence homology were conducted with the BLAST and FASTA programs (2, 8, 17, 50).

**RNA preparation.** Total cell RNA was obtained from late-log-phase cells of strains RAB15 and HU1124. HU1124 is the parental strain from which the pRAB plasmids encoding the O75 *rfb* and *rol* genes were cloned. One hundred milliliters of cell suspension was spun down, and the pellets were washed once with TES (50 mM Tris [pH 8.0], 5 mM EDTA, 50 mM NaCl), resuspended in 100 μl of TES, and transferred to a 1.5-ml microcentrifuge tube. The RNA suspension was mixed with 40 μl of 1% lysozyme (in 0.25 M Tris [pH 8.0]) and 30 μl of 0.5 M EDTA and then incubated for 10 min on ice. Then, 200 μl of RNAzol B solution (Cinna/Biotecx Laboratories, Inc., Houston, Tex.), containing guanidine thiocyanate, 2-mercaptoethanol, and phenol, was added, and the solution was mixed by pipetting up and down five times. Chloroform (100 μl) was added, mixed by shaking rapidly up and down for 15 s, and then placed on ice for 5 min. The tubes were spun for 15 min in a microcentrifuge. The supernatant was drawn off, and the RNAzol B-chloroform treatment was repeated once more to completely clear the solution. After being spun for 15 min, the supernatant was precipitated by addition of an equal volume of isopropanol and stored at −70°C for 30 min. Following centrifugation, the pellet was dried briefly, and the RNA was

resuspended in 100 μl of TE (Tris-EDTA) containing 10 mM vanadyl ribonucleoside complex (Bethesda Research Laboratories). This solution was mixed with 20 μl of an RNase-free DNase I (Boehringer Mannheim) mixture (100 mM $MgCl_2$, 100 mM dithiothreitol, 4 μl of DNase I) for 1 h. This reaction was stopped by addition of 25 μl of 50 mM EDTA–1.5 M NaCl–1% SDS, and the reaction mix was subjected to phenol extraction, ethanol precipitation, and resuspension in 100 μl of a diethyl pyrocarbonate-treated solution of 0.5 mM EDTA in distilled water. The RNA concentration and purity were determined by the optical density ratio at 260 and 280 nm and evaluation on an agarose gel stained with ethidium bromide.

**Mapping the *E. coli rol* transcriptional start position.** A primer extension method modified from that of Sambrook et al. (56) was used to map the 5′ transcriptional start position of the *rol* gene. Hybridization of the primer to RNA was done by the procedure of Hu and Davidson (23). Twenty or 150 μg of total cell RNA from strain RAB15 and RH1124, respectively, was resuspended in 2.8 μl of hybridization buffer (360 mM PIPES [piperazine-*N*,*N*′-bis(2-ethanesulfonic acid), pH 6.4], 3.6 M NaCl, 9 mM EDTA) and heated to 65°C for 10 min. A 16-mer oligonucleotide (5′-CCACAAC TGCACTAGT-3′) was labeled at its 5′ end by T4 polynucleotide kinase (New England Biolabs) in the presence of [γ-$^{32}$P]ATP (NEN Research Products, Boston, Mass.). This primer is countersense to the coding region of the *rol* DNA sequence and is 3′ to the beginning of the open reading frame (ORF) shown in Fig. 2. One microliter of primer ($10^5$ cpm) and 20 μl of deionized formamide were added to the RNA suspension, and the mixture was heated to 80°C for 10 min and then incubated at 37°C overnight. Following hybridization, the suspension was precipitated and resuspended in reverse transcriptase buffer containing 1 mM each of the four nucleotide triphosphates, as described by Sambrook et al. (56). Avian myeloblastosis virus reverse transcriptase (21 U; Promega Inc.) was added, and the solution was incubated for 2 h at 37°C. The reaction was stopped by addition of EDTA to a concentration of 10 mM. The mixture was ethanol precipitated and resuspended in 6 ml of distilled water, after which 4 μl of sequencing-gel loading buffer was added. After being heated for 3 min at 95°C, 2 μl of this solution was loaded on an 8% polyacrylamide sequencing gel. Adjacent lanes were loaded with dideoxynucleotide sequencing reaction mixes with pRAB15.

**Complementation analysis of *S. typhimurium* Rol function.** The plasmid pHS8500 encodes the *S. typhimurium rol* gene and much of the histidine operon, as shown in Fig. 5 (12, 15). This plasmid was digested with *Pst*I to remove a portion of the ampicillin resistance gene and thereby inactivate it. The larger, remaining *Pst*I fragment was circularized by T4 DNA ligase and designated pRAB20. *E. coli* SØ874 was transformed with this plasmid to generate strain RAB20. This *rol*$^+$ strain was then transformed with pRAB3 (Fig. 1), which encodes the O75 *rfb* region but not *rol*. Transformants that were resistant to tetracycline and ampicillin and were positive for O75 expression, as detected by slide agglutination in O75 antiserum, were isolated and checked for the presence of both plasmids by a small-scale cleared-lysate method (60). The ability of *S. typhimurium* Rol to regulate the size distribution of the O polysaccharides expressed from the *E. coli* O75 *rfb* region encoded on pRAB3 was tested in this double transformant (RAB23).

**LPS profiles on acrylamide gels.** LPS was isolated from the bacterial strains shown in Fig. 6 as described by Hitchcock and Brown (22). LPS samples were placed on an SDS–15% polyacrylamide resolving gel with a 6% polyacrylamide stacking gel (0.8%:30%, bisacrylamide-acrylamide) as described by Laemmli (31) and electrophoresed at 30 mA for approximately 4 h. The gels were fixed and silver stained by the method of Tsai and Frasch (63).

**Nucleotide sequence accession numbers.** The *E. coli* and *S. typhimurium rol* DNA sequences presented in this article have been assigned GenBank accession numbers M89934 and M89933, respectively.

## RESULTS

**Nucleotide sequence of the *E. coli rol* gene.** Dideoxynucleotide sequencing of the insert DNA of pRAB15 revealed that the region consisted of 1,769 nucleotides. Of these nucleotides, 1,330 are shown in Fig. 2. Only one large ORF (bp 101 to 1114) encoding a deduced polypeptide of 37.8 kDa was identified. The actual start of translation is probably initiated from the third AUG of this sequence (bp 137). This conclusion is supported by the following observations. The molecular mass of the deduced polypeptide from this start position is 36 kDa. This size is in good agreement with our previously published in vitro translation results of a 35.5-kDa protein encoded within this region (7). A potential ribosome-binding site is observed at bp 128 to 131, five nucleotides upstream from the third AUG, indicating a possible translational start at bp 137. Finally, the transcriptional start position of this gene was mapped by primer extension, and the 5′ end of the mRNA was found to start at bp 107 (Fig. 3). This position is downstream of the first AUG position. Putative promoter-binding sites are as shown in Fig. 2. A long palindromic sequence was found at the 3′ end of the ORF, suggestive of a transcription termination stem-loop structure.

**Searching GenBank.** An initial search of the GenBank data base failed to identify any DNAs or peptides with homology to the putative *rol* gene or deduced Rol amino acid sequence. However, a region of 595 nucleotides downstream from the *E. coli rol* translational stop codon (bp 1114) was found to have substantial homology to the histidine operon and its reported transcriptional terminator structures in both *E. coli* K-12 and *S. typhimurium* LT2 (12, 13). The homology of the O75 DNA in this region to the *S. typhimurium hisIE* DNA was calculated to be 80 to 81%, while the homology to the *E. coli* K-12 *hisIE* DNA was 92 to 94%.

**Nucleotide sequence of the *S. typhimurium rol* gene.** The *his* transcriptional terminator structure of *S. typhimurium* was previously shown to be bifunctional and to be shared with a gene encoding a protein of unknown function (13). The DNA sequence which encodes this *S. typhimurium* protein is shown in Fig. 2 below the *E. coli rol* sequence. The homology of these two DNA sequences within the encoding regions is 68.9%. The nucleotides in the regulatory regions of the *E. coli rol* gene (the putative promoter, transcriptional start position, Shine-Dalgarno sequence [59], and translational start position) appear to be similar and conserved in the *S. typhimurium* sequence. The ORFs of these two genetic regions match closely. The second AUG of the *E. coli rol* sequence is not found in the *S. typhimurium* sequence, further supporting the conclusion that translation is initiated from the third AUG. The potential termination region of the *E. coli* O75 *rol* gene is very similar (88%) to the *his* termination region published for *E. coli* K-12 (12). The *E. coli* DNA sequence between bp 1199 and 1298 has no counterpart in the *S. typhimurium* DNA sequence. The DNA sequence from bp 1299 to 1746 showed good homology to the *hisIE* region in both genera (only sequence to bp 1330

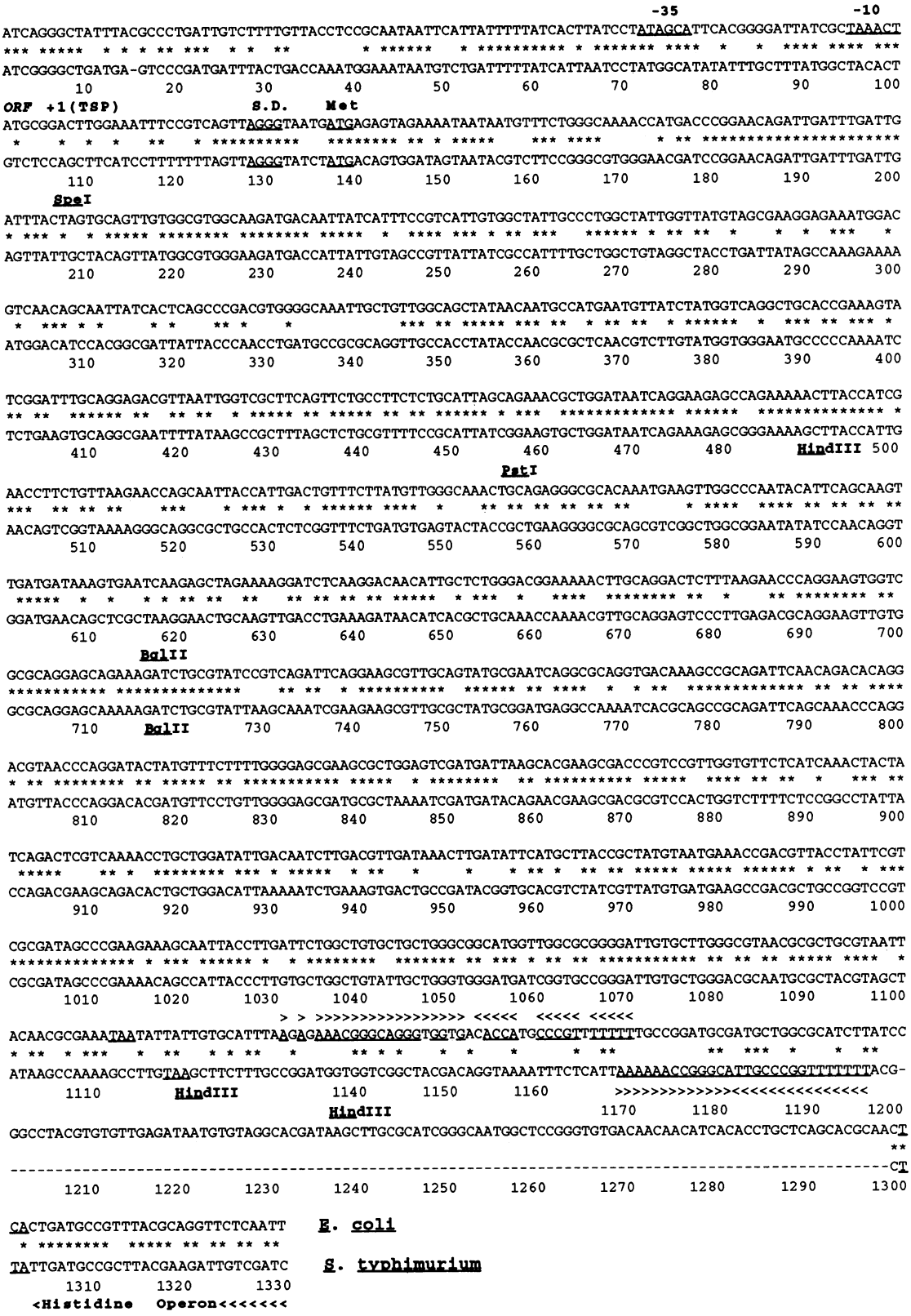



FIG. 2. Nucleotide sequence of the *E. coli* and *S. typhimurium rol* coding strands with flanking regions. The asterisks between the sequences signify homologous nucleotides between the two sequences. The putative regulatory regions are marked as follows: S.D., potential ribosome-binding site; TSP, transcriptional start position; −10 and −35, potential promoter-binding regions. Restriction enzyme sites corresponding to those on the physical maps in Fig. 1 and 5 are shown above the line for the *E. coli* sequence and below the line for the *S. typhimurium* sequence. Only the *Bgl*II sites are common to both sequences. The other restriction sites are shown to facilitate orientation to the physical maps. The translational stop codons are underlined. The nucleotides of actual or potential transcriptional terminator structures for both the *rol* genes are denoted by a line below the sequence, with arrowheads indicating the stem-loop orientation. The underlined nucleotides of the *E. coli rol* termination structure are those having homology to the previously reported *E. coli* K-12 *his* potential transcriptional terminator.

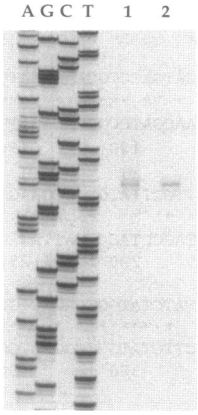



FIG. 3. Mapping of the transcriptional start position of *E. coli rol*. The first four lanes (A, G, C, and T) represent the dideoxynucleotide sequencing reactions of pRAB15 with a 16-mer oligonucleotide primer. Lanes 1 and 2 were loaded with cDNA generated off mRNA from total cell RNA isolated from strains HU1124 and RAB15, respectively. A primer extension method was used, as described in the text. The same primer was used for both the dideoxynucleotide sequencing and the primer extension. The most intense bands in lanes 1 and 2 represent the 5′ end of the *rol* mRNA.

is shown in Fig. 2). The termination codons of the complementary strands of the *hisIE* genes are underlined in Fig. 2.

**Positioning the *gnd* gene with respect to *rol* and *rfb*.** The *gnd* gene for 6-phosphogluconate dehydrogenase has been mapped to the distal end of the *his* operon at map position 42 in *S. typhimurium* (57). Previously published restriction endonuclease maps of cloned DNA from *S. typhimurium* LT2 show the genetic regions from *his* to *gnd* to be in the approximate positions shown in Fig. 5 (11, 27, 55). As established in this study, the *rol* gene is contiguous with the *his* operon. In *E. coli* K-12, the *gnd* gene has also been positioned at the distal end of the *his* operon (3). In *E. coli* O75, the position of the *gnd* gene was determined by hybridization of two synthetic 20-mer end-labeled probes, specific for the amino and carboxy ends of *gnd*, respectively, with exonuclease-generated deletion subclones of pRAB1 (not shown) and by comparison of the restriction map of pRAB1 with that of *gnd* (4, 44). This *gnd* gene is positioned near the *Eco*RI side of pRAB3, as shown in Fig. 1. The approximately 0.7-kb *Pst*I-*Kpn*I fragments found in this position in both *E. coli* and *S. typhimurium* (Fig. 1 and 5) are in good agreement with reports of a conserved *Pst*I-*Kpn*I site in *gnd* (4, 55). The *gnd* regions in both *E. coli* and *S. typhimurium* have been reported to be tightly linked with the *rfb* regions (27, 45).

**Homology of *E. coli* and *S. typhimurium* Rol proteins.** Kyte and Doolittle (30) hydropathy plots of deduced amino acid sequences for *E. coli* and *S. typhimurium* Rol are shown in Fig. 4A. The two amino acid sequences upon which these plots were prepared are shown aligned in Fig. 4B. The *E. coli rol* and *S. typhimurium rol* deduced peptides consist of 325 and 327 amino acid residues, respectively. These two proteins have a very similar secondary structure. Especially evident are the two hydrophobic regions of 20 amino acids on either end of the molecule. The net hydropathy values for these regions exceed a 1.9 index value and are suggestive of membrane-spanning domains. Overall, the molecule is relatively hydrophilic, with a net hydropathy value of −0.7. When conservative substitutions in amino acid sequence are considered, the homology between these proteins is 93.5%.

**Functional complementation of *E. coli* O75 *rfb* region with *S. typhimurium* Rol.** Such a high degree of homology in protein structure suggests a similar function. To determine whether the *S. typhimurium* Rol protein could regulate expression of the O-polysaccharide antigen of *E. coli* O75, a complementation experiment was conducted. The pRAB20 plasmid (Fig. 5), which encodes *S. typhimurium rol* and part of the *his* operon, was placed in *trans* with pRAB3, expressing the unregulated O75 antigen in *E. coli* SØ874. This double transformant strain was named RAB23. The double transformant RAB23 agglutinated in rabbit anti-O75 serum, as did the control strains RAB1 and RAB3. The negative controls, *E. coli* SØ874 with no plasmid and with pRAB20, did not agglutinate in this serum. As shown in the silver-stained polyacrylamide gel in Fig. 6, the *E. coli* SØ874 strain harboring the pRAB20 plasmid (RAB20) did not express an O chain (no ladder present), whereas RAB3 (*E. coli* SØ874 carrying pRAB3) and RAB1 (*E. coli* SØ874 carrying pRAB1) did produce LPS ladders. The LPS ladder from RAB3 is not bimodal and appears to be unregulated. In contrast, RAB1 and the double transformant RAB23 both showed the normal, bimodal display of LPS on this gel. The LPSs produced by both of these strains appear to be of equal molecular weight, judging by the identical or equal spacing of the bands on the gel. From these results, we conclude that the *S. typhimurium* Rol protein encoded on pRAB20 was able to regulate the size distribution of the O75 O-polysaccharide antigen of *E. coli* onto the core/lipid A structure of the host *E. coli* SØ874.

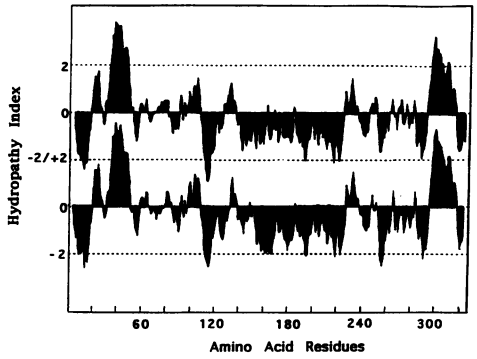
While the manuscript of this article was being prepared, a protein of unknown function, encoded on DNA sequence adjacent to the putative *rfe* gene (*rfe* ORF2; GenBank accession no. M76129), was found to have 26% homology to Rol over several hundred amino acid residues (41). The *rfe* gene is thought to encode UDP-GlcNAc:undecaprenoylphosphate transferase, responsible for lipid I synthesis and necessary for the biosynthesis of the enterobacterial common antigen and some O antigens (25, 41, 61). A Monte Carlo Simulation analysis, comparing the *E. coli* Rol with this *rfe* ORF2 protein, found the level of homology to be significant (1, 5).

## DISCUSSION

We have shown that the *rol* genes in *E. coli* O75 and *S. typhimurium* LT2 code for proteins that are structural and functional homologs. The *rol* genes are located between the *his* and *gnd* loci in these strains. Based on previous work by Carlomagno et al. (13) and the results presented herein, we may also conclude that transcription of the *S. typhimurium rol* gene proceeds counter to that of the *his* region and terminates at a bifunctional transcriptional terminator. It has not been established whether the potential *E. coli rol* terminator is also bifunctional. However, we have shown that the *E. coli rol* gene is also transcribed counter to the *his* operon. The *E. coli rol* gene also appears to have its own promoter region and may be independent of the *rfb* gene cluster.

The Rol protein appears to regulate the distribution of the O-antigen subunits into a chain of the preferred size. The mechanism of action of Rol appears to be independent of the O-antigen subunit structure. The *E. coli* O75 O-antigen subunit structure consists of a tetrasaccharide, a glucosamine-galactose-rhamnose backbone with mannose attached to the galactose. The O-antigen subunit structure of *S. typhimurium* is composed of a pentasaccharide, a mannose-rhamnose-galactose backbone with an acetylated abe-

A.




B.

```
                                                    2.17 average
ECROL :1    MRVENNNVSGQNHDPEQIDLIDLLVQLWRGKMTIIISVIVAIALAIGYLAVAKEKWTSTA
STROL :1    -T-DS-TS--RGN--------E--L----------VA--I--L--V---II---------

ECROL :61   IITQPDVGQIAGYNNAMNVIYGQAAPKVSDLQETLIGRFSSAFSALAETLDNQEEPEKLT
STROL :61   ------AA-V-T-T--L--L--GN---I-EV-ANF-S---------S-V----K-R----

ECROL :121  IEPSVKNQQLPLTVSYVGQTAEGAQMKLAQYIQQVDDKVNQELEKDLKDNIALGRKNLQD
STROL :121  --Q---G-A---S--D-ST------RR--E------EQLAK--QV------T-QT-T--E

ECROL :181  SLRTQEVVAQEQKDLRIRQIQEALQYANQAQVTKPQIQQTQDVTQDTMFLLGSEALESMI
STROL :181  --E--------------K--E---R--DE-KI-Q-------------------D--K---
                                                                  1.94
ECROL :241  KHEATRPLVFSSNYYQTRQNLLDIDNLDVDKLDIHAYRYVMKPTLPIRRDSPKKAITLIL
STROL :241  QN---------PA----K-T----K--K-TADTV-V----------V------T----V-
            average
ECROL :301  AVLLGGMVGAGIVLGRNALRNYNAK
STROL :301  -------I-----------S-KP-AL
```

FIG. 4. (A) Two overlapped Kyte and Doolittle (30) hydropathy plots of the deduced amino acid sequences of *E. coli* Rol (top) and *S. typhimurium* Rol (bottom). Positive values represent regions of potential hydrophobic secondary structure, while negative values represent potential hydrophilic secondary structure. The values shown are averaged over a window of nine amino acid residues. The average hydropathic index for each of these proteins is −0.7. (B) Alignment of *E. coli* Rol (ECROL) and *S. typhimurium* Rol (STROL) deduced amino acid sequences. Homologous amino acid residues are denoted by a dash. The two regions with lines drawn over them correspond to the two hydrophobic peaks seen in panel A and show the average hydropathic value of these two 20-amino-acid sequences.

quose attached to the mannose and glucose branching off the galactose (25). Complementation analysis with other genetic regions in *Salmonella* spp. and between *Salmonella* and *E. coli* strains suggests that the action of the ligase (*rfaL*), and in some cases the polymerase enzyme (*rfc*), may also be independent of O-antigen subunit structure (25, 36, 38, 61). However, more work needs to be done in this area to substantiate these generalizations or explain the exceptions to them.

Polymerization and assembly of the O polysaccharide onto the core/lipid A structure may take place in or on the inner membrane in *S. typhimurium* (40). Several of the genetic regions thought to be involved in *S. typhimurium* O-antigen biosynthesis (*rfb*, *rfaL*, and *rfc*) have been cloned and sequenced (16, 27, 36). Information obtained from the deduced polypeptide sequences of the *rfb* gene cluster suggests that some of these proteins have hydrophobic membrane-spanning domains (27). Similarly, the amino acid sequences of the polymerase and ligase proteins indicate that they may also be integral or peripheral membrane proteins (16, 36). The reported hydropathic index of both the putative

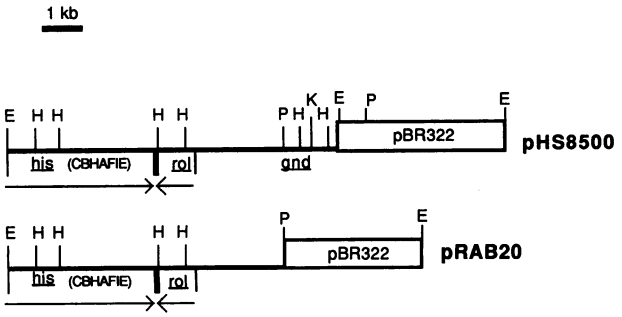



FIG. 5. Physical map of plasmid pHS8500, which encodes the *S. typhimurium gnd*, *rol*, and part of the *his* operon genes, and its subclone pRAB20, which has the *Pst*I fragment deleted. The arrows beneath the genes indicate the direction of transcription. The black box between *rol* and *his* represents the bifunctional transcriptional terminator. Restriction endonuclease sites: E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; K, *Kpn*I. The *Pst*I and *Kpn*I sites in the *gnd* gene are similarly located in the *E. coli gnd* gene, shown in Fig. 1.

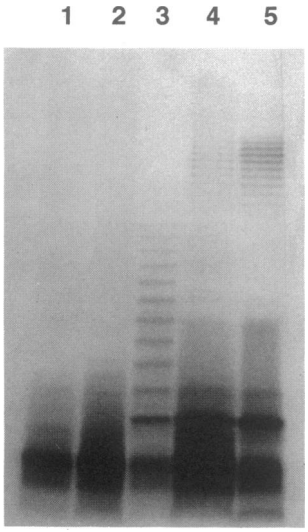



FIG. 6. Silver-stained LPS preparations on a 15% polyacrylamide gel from strains harboring the plasmids shown in Fig. 1 and 5 or from control strains. A complementation test of pRAB20, encoding the *S. typhimurium rol* gene, with pRAB3, encoding the *E. coli* O75 *rfb* region, in *E. coli* SØ874 (lane 4) demonstrates that *S. typhimurium* Rol is functionally equivalent to *E. coli* Rol. Control LPS preparations include (lane 1) *E. coli* SØ874, (lane 2) pRAB20 in SØ874, (lane 3) pRAB3 in SØ874, and (lane 5) pRAB1 in SØ874.

polymerase and ligase proteins is approximately +0.7, indicating that these proteins are very hydrophobic in nature. While overall Rol is a hydrophilic protein, its hydrophobic ends suggest that it may also be membrane associated. These results are all in agreement with the proposed model for LPS biosynthesis: the cytoplasmic location of O-antigen monomer synthesis and the membrane assembly of the O subunit structure, polymerization and ligation onto the lipid A/core structure, and subsequent translocation to the outer membrane (25, 38). It will be interesting to see whether the Rol protein is a membrane-associated protein which interacts with these other biosynthetic enzymes or if its location and mechanism of action lie elsewhere.

The GC content of the *E. coli* and *S. typhimurium rol* genes is 46.2 and 50.4%, respectively. This GC content is above the average GC content reported for the *rfb*, *rfaL*, and *rfc* gene sequences of *S. typhimurium*. The low GC content of these genes has fueled the hypothesis that parts of the *rfb* gene cluster and the *rfaL* and *rfc* genes may have originated outside of these bacteria and may have been passed through horizontal gene transfer (16, 27, 36).

Recent reports of complementation experiments in *E. coli* O111 and *Shigella flexneri* indicate that similar *rol*-like genes are present in these bacteria as well (6, 37). We previously reported that our pRAB3 plasmid was complemented by genetic regions encoded on *E. coli* DH1 (7). Also, DNA in the termination regions of *E. coli rol* sequence had homology to a similar region in *E. coli* K-12, suggesting that the *rol* gene is also present in this strain. These results indicate that *rol*-like genes are present and conserved in members of the family *Enterobacteriaceae* and other gram-negative bacteria that express different sizes of O-polysaccharide antigens on the surface. Furthermore, the significant homology of Rol to a protein encoded near the *rfe* gene suggests that other proteins with a similar structure and possibly a similar mechanism of action may be involved in regulating the length of other surface polysaccharide antigens.

## ACKNOWLEDGMENTS

P.A. and E.B. are indebted to M. S. Carlomagno and C. B. Bruni for constant encouragement and advice.

This work was supported by Public Health Service research grants AI-18462 and AI-21009 and by the Edward J. and Josephine Hudson Scholars Fund.

## REFERENCES

1. **Altschul, S. F., and B. W. Erickson.** 1986. Optimal sequence alignment using affine gap costs. Bull. Math. Biol. **48:**603–616.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
3. **Bachman, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54:**130–197.
4. **Barcak, G. J., and R. E. Wolf, Jr.** 1988. Growth-rate-dependent expression and cloning of *gnd* alleles from natural isolates of *Escherichia coli*. J. Bacteriol. **170:**365–371.
5. **Barker, W. C., and M. O. Dayhoff.** 1972. Detecting distant relationships: computer methods and results, p. 101–110. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure. National Biomedical Research Foundation, Washington, D.C.
6. **Bastin, D. A., L. K. Romana, and P. R. Reeves.** 1991. Molecular cloning and expression in *Escherichia coli* K-12 of the *rfb* gene cluster determining the O antigen of an *E. coli* O111 strain. Mol. Microbiol. **5:**2223–2231.
7. **Batchelor, R. A., G. E. Haraguchi, R. A. Hull, and S. I. Hull.** 1991. Regulation by a novel protein of the bimodal distribution of lipopolysaccharide in the outer membrane of *Escherichia coli*. J. Bacteriol. **173:**5699–5704.
8. **Bilofsky, H. S., and W. C. Burks.** 1988. The GenBank (R) genetic sequence data bank. Nucleic Acids Res. **16:**1861–1864.
9. **Bolivar, F., R. L. Rogriguez, P. J. Green, M. C. Betlach, H. L. Heyneker, H. W. Boyer, J. H. Crosa, and S. Falkow.** 1977. Construction and characterization of new cloning vehicles. II. A multi-purpose cloning system. Gene **2:**95–113.
10. **Brown, M. G., A. Weston, J. R. Saunders, and G. O. Humphreys.** 1979. Transformation of *Escherichia coli* C600 by plasmid DNA at different phases of growth. FEMS Microbiol. Lett. **5:**219–222.
11. **Carlomagno, M. S., F. Blasi, and C. B. Bruni.** 1983. Gene organization in the distal part of the *Salmonella typhimurium* histidine operon and determination and sequence of the operon transcriptional terminator. Mol. Gen. Genet. **191:**413–420.
12. **Carlomagno, M. S., L. Chiariotti, P. Alifano, A. G. Nappo, and C. B. Bruni.** 1988. Structure and function of the *Salmonella typhimurium* and *Escherichia coli* K-12 histidine operons. J. Mol. Biol. **203:**585–606.
13. **Carlomagno, M. S., A. Riccio, and C. B. Bruni.** 1985. Convergently functional, rho-independent terminator in *Salmonella typhimurium*. J. Bacteriol. **163:**362–368.
14. **Chang, A. C. Y., and S. N. Cohen.** 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J. Bacteriol. **134:** 1141–1156.
15. **Chiariotti, L., P. Alifano, M. S. Carlomagno, and C. B. Bruni.** 1986. Nucleotide sequence of the *Escherichia coli hisD* gene and of the *Escherichia coli* and *Salmonella typhimurium hisIE* gene. Mol. Gen. Genet. **203:**382–388.
16. **Collins, L. V., and J. Hackett.** 1991. Molecular cloning, characterization, and nucleotide sequence of the *rfc* gene, which encodes an O-antigen polymerase of *Salmonella typhimurium*. J. Bacteriol. **173:**2521–2529.
17. **George, D. G., W. C. Barker, and L. T. Hunt.** 1976. The Protein Identification Resource (PIR). Nucleic Acids Res. **14:**11–16.
18. **Goldman, R. C., and F. Hunt.** 1990. Mechanism of O-antigen distribution in lipopolysaccharide. J. Bacteriol. **172:**5352–5359.
19. **Goldman, R. C., K. Joiner, and L. Leive.** 1984. Serum-resistant

mutants of *Escherichia coli* O111 contain capsule and cover more of their lipid A core with O antigen. J. Bacteriol. **159:**877–882.

20. **Goldman, 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.
21. **Grossman, N., M. A. Schmetz, J. Foulds, E. N. Klima, V. Jiminez, L. L. Leive, and K. Joiner.** 1987. Lipopolysaccharide size and distribution determine serum resistance in *Salmonella montevideo*. J. Bacteriol. **169:**856–863.
22. **Hitchcock, P. J., and T. M. Brown.** 1983. Morphological heterogeneity among *Salmonella* lipopolysaccharide chemotypes in silver-stained polyacrylamide gels. J. Bacteriol. **154:**269–277.
23. **Hu, M. C.-T., and N. Davidson.** 1986. Mapping transcriptional start points on cloned genomic DNA with T4 polymerase: a precise and convenient technique. Gene **42:**21–29.
24. **Jann, B., K. Reske, and K. Jann.** 1975. Heterogeneity of lipopolysaccharides. Analysis of chain lengths by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Eur. J. Biochem. **60:**239–246.
25. **Jann, K., and B. Jann.** 1984. Structure and biosynthesis of O-antigens, p. 138–186. *In* E. T. Rietschel (ed.), The handbook of endotoxin. Elsevier Biomedical Press, Amsterdam.
26. **Jessop, H. L., and P. A. Lambert.** 1986. The role of surface polysaccharide in determining the resistance of *Serratia marcescens* to surface killing. J. Gen. Microbiol. **52:**76–84.
27. **Jiang, S.-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.
28. **Joiner, K. A., N. Grossman, M. Schmetz, and L. Leive.** 1986. $C_3$ binds preferentially to long chain lipopolysaccharide during alternate pathway activation by *Salmonella montevideo*. J. Immunol. **136:**710–715.
29. **Kushner, S.** 1987. Useful host strains and techniques for recombinant DNA experiments, p. 1190–1219. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaecter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
30. **Kyte, J., and R. F. Doolittle.** 1982. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **157:**105–132.
31. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
32. **Lawrence, C. B., and D. A. Goldman.** 1988. Definition and identification of homology domains. Comput. Appl. Biosci. **4:**25–33.
33. **Lawrence, C. B., T. Y. Shalom, and S. Honda.** 1989. EuGene: a software package for nucleotide and protein sequence analysis for UNIX systems. Molecular Biology Information Resources, Department of Cell Biology, Baylor College of Medicine, Houston, Tex.
34. **Lomberg, H., M. Hellstrom, U. Jodal, H. Leffler, K. Lincoln, and C. Svanborg-Eden.** 1984. Virulence-associated traits in *Escherichia coli* causing first and recurrent episodes of urinary tract infections in children with or without vesicoureteral reflux. J. Infect. Dis. **150:**561–569.
35. **MacIntyre, S., R. Lucken, and P. Owen.** 1986. Smooth lipopolysaccharide is the major protective antigen for mice in the surface extract from IATS serotype 6 contributing to the polyvalent *Pseudomonas aeruginosa* vaccine PEV. Infect. Immun. **52:**76–84.
36. **MacLachlan, P. R., S. K. Kadam, and K. E. Sanderson.** 1991. Cloning, characterization, and DNA sequence of the *rfaLK* region for lipopolysaccharide synthesis in *Salmonella typhimurium* LT2. J. Bacteriol. **173:**7151–7163.
37. **Macpherson, D. F., R. Morona, D. W. Beger, K.-C. Cheah, and P. A. Manning.** 1991. Genetic analysis of the rfb region of *Shigella flexneri* encoding the Y serotype O-antigen specificity. Mol. Microbiol. **5:**1491–1499.
38. **Mäkelä, P. H., and B. A. D. Stocker.** 1984. Genetics of lipopolysaccharide, p. 59–137. *In* E. T. Rietschel (ed.), The handbook of endotoxin. Elsevier Biomedical Press, Amsterdam.
39. **McCallum, K. L., G. Schoenhais, D. Laakso, B. Clarke, and C. Whitfield.** 1989. A high-molecular-weight fraction of smooth lipopolysaccharide in *Klebsiella pneumoniae* serotype O1:K20 contains O-antigen epitope and determines resistance to nonspecific serum killing. Infect. Immun. **57:**3816–3822.
40. **McGrath, B. C., and M. J. Osborn.** 1991. Localization of the terminal steps in O-antigen synthesis in *Salmonella typhimurium*. J. Bacteriol. **173:**649–654.
41. **Meier-Dieter, U., K. Barr, R. Starman, L. Hatch, and P. Rick.** 1992. Nucleotide sequence of the *Escherichia coli rfe* gene involved in the synthesis of enterobacterial common antigen. J. Biol. Chem. **267:**746–753.
42. **Munford, R. S., C. L. Hall, and P. D. Rick.** 1980. Size heterogeneity of *Salmonella typhimurium* lipopolysaccharide in outer membranes and culture supernatant membrane fragments. J. Bacteriol. **144:**630–640.
43. **Muschel, L. H., and L. J. Larsen.** 1970. The sensitivity of smooth and rough gram-negative bacteria to the immune bacteriocidal reaction. Proc. Soc. Exp. Biol. Med. **133:**345–352.
44. **Nasoff, M. S., H. V. Baker II, and R. E. Wolf, Jr.** 1984. DNA sequence of the *Escherichia coli* gene, *gnd*, for 6-phosphogluconate dehydrogenase. Gene **27:**253–264.
45. **Neal, B. L., G. C. Tsiolis, M. W. Heuzenroeder, P. A. Manning, and P. R. Reeves.** 1991. Molecular cloning and expression in *Escherichia coli* K-12 of chromosomal genes determining the O antigen of an *E. coli* O2:K1 strain. FEMS Microbiol. Lett. **82:**345–352.
46. **Needleman, S. B., and C. D. Wunsch.** 1970. A general method applicable to the search for similarities in the amino acid sequences of two proteins. J. Mol. Biol. **48:**443–453.
47. **Neuhard, J., and E. Thomassen.** 1976. Altered deoxyribonucleotide pools in P2 eductants of *Escherichia coli* K-12 due to deletion of the *dcd* gene. J. Bacteriol. **126:**999–1001.
48. **Nnalue, N. A., S. Newton, and B. A. D. Stocker.** 1990. Lysogenization of *Salmonella choleraesuis* by phage 14 increases average length of O-antigen chains, serum resistance and intraperitoneal mouse virulence. Microb. Pathogen. **8:**393–402.
49. **Palva, E. T., and P. H. Mäkelä.** 1980. Lipopolysaccharide heterogeneity in *Salmonella typhimurium* analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Eur. J. Biochem. **107:**137–143.
50. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:**2444–2448.
51. **Peterson, A. A., A. Haug, and E. J. McGroarty.** 1986. Physical properties of short- and long-O-antigen-containing fractions of lipopolysaccharide from *Escherichia coli* O111:B4. J. Bacteriol. **165:**116–122.
52. **Peterson, A. A., and E. J. McGroarty.** 1985. High-molecular-weight components in lipopolysaccharides of *Salmonella typhimurium*, *Salmonella minnesota*, and *Escherichia coli*. J. Bacteriol. **162:**738–745.
53. **Porat, R., M. A. Johns, and W. R. McCabe.** 1987. Selective pressures and lipopolysaccharide subunits as determinants of resistance of clinical isolates of gram-negative bacilli to human serum. Infect. Immun. **55:**320–328.
54. **Rana, F. R., E. A. Macias, C. M. Sultany, M. C. Modzrakowski, and J. Blazyk.** 1991. Interactions between magainin 2 and *Salmonella typhimurium* outer membranes. Biochemistry **30:**5858–5866.
55. **Reeves, P., and G. Stevenson.** 1989. Cloning and nucleotide sequence of the *Salmonella typhimurium* LT2 *gnd* gene and its homology with the corresponding sequence of *Escherichia coli* K-12. Mol. Gen. Genet. **217:**182–184.
56. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed., p. 7.37–7.83. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
57. **Sanderson, K. E., and J. R. Roth.** 1988. Linkage map of *Salmonella typhimurium*, edition 7. Microbiol. Rev. **52:**485–532.
58. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequenc-

ing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.

59. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16S rRNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
60. **So, M., W. S. Dallas, and S. Falkow.** 1978. Characterization of an *Escherichia coli* plasmid encoding for synthesis of heat-labile toxin: molecular cloning of the toxin determinant. Infect. Immun. **21**:405–411.
61. **Sugiyama, T., N. Kido, T. Komatsu, M. Ohta, and N. Kato.** 1991. Expression of the cloned *Escherichia coli* O9 *rfb* gene in various mutant strains of *Salmonella typhimurium*. J. Bacteriol. **173**:55–58.
62. **Tomas, J. M., B. Ciurana, V. J. Benedi, and A. Juarez.** 1988. Role of lipopolysaccharide and complement in susceptibility of *Escherichia coli* and *Salmonella typhimurium* to non-immune serum. J. Gen. Microbiol. **134**:1009–1016.
63. **Tsai, C. M., and C. E. Frasch.** 1982. A sensitive silver stain for detecting lipopolysaccharide in polyacrylamide gels. Anal. Biochem. **119**:115–119.
64. **Weiss, J., Q. Beckerdite, and P. Elsbach.** 1980. Resistance of gram-negative bacteria to purified bacterial leukocyte proteins. J. Clin. Invest. **65**:619–628.
65. **Yeh, H.-Y., and D. M. Jacobs.** 1992. Characterization of lipopolysaccharide fractions and their interactions with cells and model membranes. J. Bacteriol. **174**:336–341.