JOURNAL OF BACTERIOLOGY, Aug. 1990, p. 4652–4660
0021-9193/90/084652-09$02.00/0

Vol. 172, No. 8

Cloning, Expression, and Characterization of the *Escherichia coli* K-12 *rfaD* Gene

JOYCE C. PEGUES, LISHI CHEN, ALFRED W. GORDON, LI DING, AND WILLIAM G. COLEMAN, JR.*

Section on Pharmacology, Laboratory of Biochemical Pharmacology, National Institute of Diabetes and Digestive and Kidney Diseases, Building 8, Room 2A-03, Bethesda, Maryland 20892

Received 18 December 1989/Accepted 25 May 1990

The *rfaD* gene encodes ADP-L-glycero-D-mannoheptose-6-epimerase, an enzyme required for the biosynthesis of the lipopolysaccharide precursor ADP-L-glycerol-D-mannoheptose. The precise localization of the *rfaD* gene on a 1.3-kilobase *SspI-HpaI* fragment is reported. The *rfaD* gene and the flanking regions were completely sequenced. The location of the *rfaD* gene on the physical map of the *Escherichia coli* chromosome was determined. Primer extension studies were used to define the regulatory region of the *rfaD* gene. The cloned *rfaD* gene directed the synthesis of a 37,000-dalton polypeptide in several in vivo and in vitro expression systems. N-terminal analysis of purified ADP-L-glycero-D-mannoheptose-6-epimerase confirmed the first 34-amino-acid sequence deduced from the nucleotide sequence of the *rfaD* gene coding region. The primary structure of the *rfaD* protein contains the sequence fingerprint for the ADP-binding βαβ fold at the N terminus.

In previous reports, we characterized lipopolysaccharide and barrier function mutants of *Escherichia coli* K-12 (10, 11) genetically, chemically, and biochemically. One of these mutants was shown to have a genetic lesion at 81 minutes on the *E. coli* K-12 chromosome, designated *rfaD*. Its phenotype included altered heptose (L-glycero-D-mannoheptose) and lipopolysaccharide biosynthesis and increased outer membrane permeability. These phenotypic changes were shown to be directly related to the loss of the enzyme ADP-L-glycero-D-mannoheptose-6-epimerase, required to catalyze the conversion (Fig. 1) of ADP-D-glycero-D-mannoheptose to ADP-L-glycero-D-mannoheptose (9). Although we have previously described the enzymatic activity of the *rfaD* gene product and the preliminary cloning of a large RfaD$^+$ DNA fragment into pBR322 (9, 10), the present communication is the first report which clearly defines and precisely locates the *rfaD* gene in terms of its nucleotide sequence and its gene product in any gram-negative bacterium.

MATERIALS AND METHODS

Materials. Restriction enzymes, T4 ligase, calf intestinal phosphatase, phenylmethylsulfonyl fluoride, and pepstatin A were obtained from commercial sources.

Bacterial strains, media, and plasmids. Bacterial strains and plasmids used in this study are described in Table 1. Bacteria were grown aerobically at 37°C in LB medium (27) or in the minimal medium of Davis and Mingioli (12). For solid medium, 1.5% agar was added. Strains containing the temperature-inducible plasmid pGP1-3 were grown at 30°C.

When necessary, antibiotics were used at the following concentrations (per milliliter): 50 μg of ampicillin, 20 μg of tetracycline, 400 μg of cycloserine, 200 μg of rifampin, and 30 and 60 μg of novobiocin.

Analytical and quantitative techniques. Plasmid DNA was prepared by the procedure of Birnboim and Doly (3) followed by purification by density gradient centrifugation (28). DNA fragments were purified by agarose gel electrophoresis followed by electroelution (37). RNA was purified from strains CL510 and CL529 by the procedure described by Hyman et al. (20). DNA was estimated by A_{260}, and protein concentrations were determined by the Bradford procedures (4, 32). λ clones from the genomic library were screened as indicated below. Samples (10 to 15 μl) of bacteriophage lysates (10^4 to 10^7 PFU/ml) of selected phage clones from the λ-*E. coli* gene library of Kohara et al. (22) were applied to seeded lawns and incubated at 37°C for 16 h. The plaques were transferred to the Colony/Plaquescreen membrane (E. I. du Pont de Nemours & Co., Inc., Wilmington, Del.) and screened by hybridization. Nick translation, prehybridization, hybridization (65°C), and washing were performed according to the instructions of the manufacturer (E. I. du Pont de Nemours). The procedures for sodium dodecyl sulfate-polyacrylamide gel electrophoresis (25, 26), the T7 RNA polymerase-expression system (40, 41), and cell-free coupled in vitro transcription and translation (5) were performed as previously described. (Cells were radiolabeled [exclusively the *rfaD* gene product)] as described by Tabor and Richardson [41].) Amino acid composition was determined after hydrolysis of the protein sample (49 μg) in 6 N HCl (16). The phenyl isothiocyanate-derivatized amino acids and the Pierce amino acid standard H were analyzed by reverse-phase column chromatography (4.6 by 150 mm; IBM/Jones) in a 1090 Hewlett-Packard high-pressure liquid chromatography system.

Assay conditions for the epimerase in crude extracts were described by Coleman and Deshpande (9, 10) and by Kontrohr and Kocsis (24). High-pressure liquid chromatography data were acquired and analyzed by the data collection system described by Minton and Attri (30).

Plasmid constructions. The RfaD$^+$ plasmid pCG50 (10) (Fig. 2) was constructed by inserting a 9-kilobase (kb) *PstI* fragment from the colicin E1-*E. coli* DNA hybrid plasmid pLC32-45 (8) into pBR322. Restriction analyses of pLC32-45 showed that it contained one *Eco*RI and two *Sma*I sites. Taking advantage of the single *Eco*RI site, we ligated *Eco*RI-digested pLC32-45 with similarly restricted pBR322. The resulting plasmid, pCK4 (Fig. 2) conferred colicin E1, ampicillin, and tetracycline resistance to transformants. The pCK4 plasmid complemented strains carrying the *rfaD* mutation. pCK4 was digested with *PstI*, resulting in four DNA fragments designated A, B, C, and D (Fig. 2). *PstI* fragment

* Corresponding author.

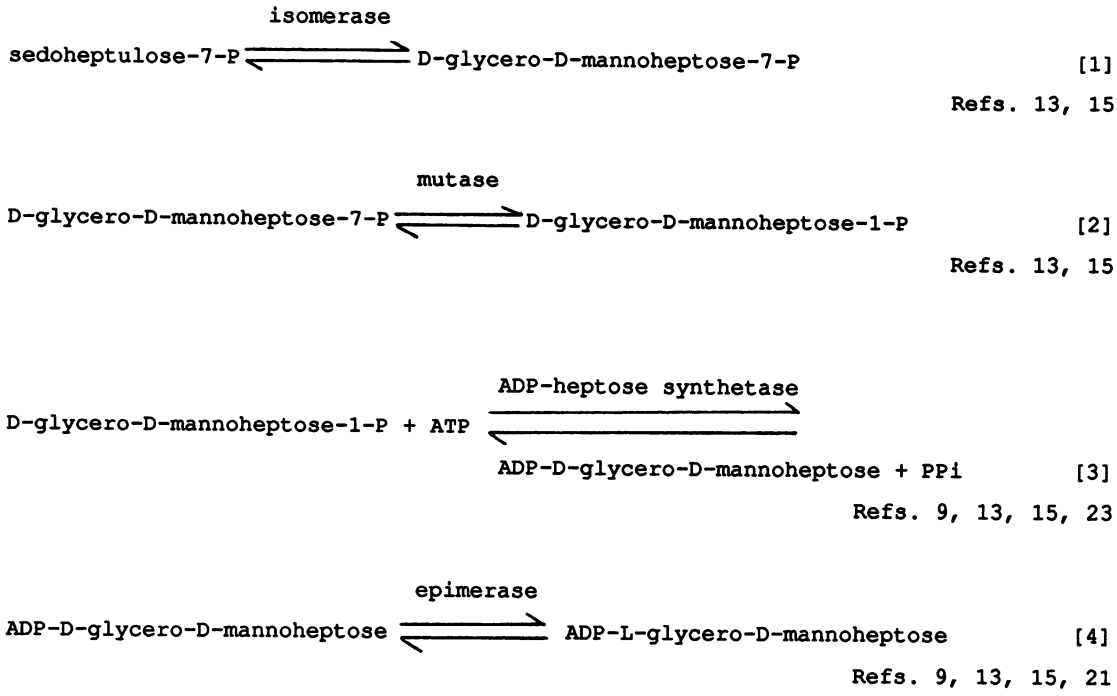

FIG. 1. Formation of ADP-L-glycero-D-mannoheptose.

A (9 kb) was inserted into *Pst*I-restricted pBR322, resulting in pCG50. Purified fragment A was subcloned in the vectors pT7-1, pT7-5, and pUC19 (Fig. 3B and Table 1). Deletions (Fig. 3B) were achieved by restriction endonuclease digestion of selected regions of the fragment A insert.

Purification of ADP-L-glycero-D-mannoheptose-6-epimerase. The *rfaD* gene product was purified from strain CL627 (an *E. coli* K38 strain containing both pGP1-3 [40] and pCG6 [Table 1]) that exclusively expressed, after thermal induction (42), the plasmid-borne *rfaD* gene. Unlabeled cells (25 g) and 2 g of [^{35}S]methionine-radiolabeled cells were suspended in 25 ml of TEM buffer (10 mM Tris hydrochloride [pH 8.0], 10

TABLE 1. Bacterial strains and plasmids

Strains or plasmid	Genotype or characteristic	Reference or source
E. coli K-12		
JM109	F′ *traD36 proAB*$^+$ *lacI*^{q} *lacZ*ΔM15/*recA1* Δ(*lac-proAB*) *endA1 gyrA96 thi-1 hsdR17 supE44 relA1 mcrA*	38
CL510	F$^-$ *proA2 leuB6 thi-1 lacY1 rpsL20 ara hsdS20*	This study
CL515	F$^-$ *proA2 leuB6 thi-1 lacY1 rpsL20 ara rfaD70 hsdS20*	9
CL529	CL515(pCG50)	This laboratory
K38	HfrC *phoA4 tonA22 garB10 ompF627 relA1 pit10 spo10* T2 (λ)	31
K38(pGP1-3)	Contains T7 polymerase gene, Tet^r	35, 36
Plasmids		
pBR322	Cloning vector, Amp^r Tet^r	Bethesda Research Laboratories, Inc., Gaithersburg, Md.
pT7-1	Cloning and expression vector, T7 promoter, Amp^r	36
pGP1-3	T7 gene 1, λ promoter, *c*I857, Tet^r	29
pT7-5[a]	Cloning and expression vector, T7 promoter, Amp^r	S. Tabor
pUC19	Cloning and expression vector, Amp^r	38
pLC32-45	Colicin E1-*E. coli* DNA hybrid plasmid, ColE1^r	8, 9
pCK4	pLC32-45 in pBR322	10
pCG50	9.0-kb *Pst* fragment from pCG4 in pBR322, Amp^r Tet^r	9, 10
pCG19	Deletion plasmid of pCG50	This study
pCG3	2.5-kb *Eco*RI-*Sma*I fragment from pCG50 in pT7-1	This study
pCG4	6.5-kb *Eco*RI-*Sma*I fragment from pCG50 in pT7-1	This study
pCG5	1.8-kb *Eco*RI-*Nru*I fragment from pCG50 in pT7-1	This study
pCG6	2.5-kb *Eco*RI-*Sma*I fragment from pCG50 in pT7-5	This study
pCG13	1.3-kb *Ssp*I-*Hpa*I fragment from pCG50 in pUC19	This study

[a] pT7-5 is similar to pT7-1 except that the β-lactamase gene is in an opposite orientation and is not under the control of the T7 promoter (S. Tabor, personal communication).

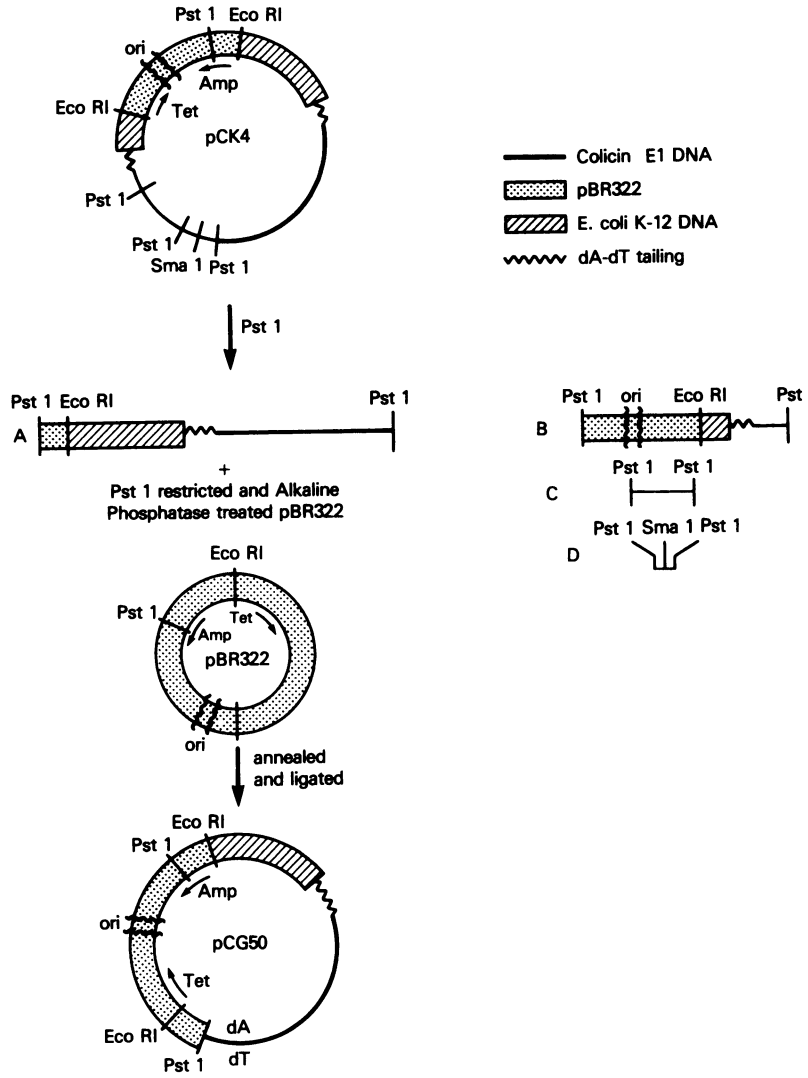

FIG. 2. Construction of pCG50. Plasmid pLC32-45 was identified from the Clarke-Carbon (8) colony bank by its ability to complement *rfaD* mutants. pLC32-45 and pBR322 were restricted with *Eco*RI and ligated. The resulting plasmid, pCK4, was Amp^r Tet^r Col^r $rfaD^+$. pCK4 was restricted with *Pst*I to generate fragment A; this fragment was cloned into pBR322, and the recombinant was designated pCG50. Cells transformed with fragment B were Tet^r but $RfaD^-$. Fragments C and D are colicin E1 fragments of 85 and 1,900 bp, respectively.

mM EDTA, 0.1 mM β-mercaptoethanol, 57 μM phenylmethylsulfonyl flouride, 14 μM pepstatin A). Crude extract, prepared as previously described (9), was treated with 0.6% streptomycin sulfate and Whatman Cell Debris Remover. The clarified cell extracts were subjected sequentially to 0 to 20% and 20 to 60% ammonium sulfate fractionation. The majority of the enzyme activity was present in the 20 to 60% ammonium sulfate fraction. This fraction was suspended in 50 ml of TEM buffer and desalted by Sephadex G-25 chromatography (5- by 50-cm column equilibrated with TEM buffer). The eluate was applied to a Beckman DEAE 5PW high-pressure liquid chromatography column (21.5 by 150 mm) previously equilibrated in TEM buffer. The proteins were eluted stepwise with a series of 0.3 M steps from 0 to 0.9 M KCl in TEM buffer. The fraction between 0.3 and 0.6 M KCl containing radioactivity and epimerase activity was rechromatographed with the DEAE 5PW column. Proteins were eluted with a linear gradient of 0.25 to 0.5 M KCl in TEM buffer. Fractions with enzyme activity and radioactivity were pooled and chromatographed on a Beckman TSK 3000 SW gel filtration column (21.5 by 300 mm) with TEM buffer (pH 7.4) containing 0.2 M NaCl. The final procedure of gel filtration on the TSK 3000 SW column yielded a homogeneous fraction. Although the recovery of the epimerase was only 0.1% by these procedures, sufficient protein was recovered for the analyses reported here. Samples were analyzed for purity by polyacrylamide gel electrophoresis. The purified protein was stored in TEM buffer containing 25% ethylene glycol in liquid nitrogen.

Nucleotide sequence determination and N-terminal amino acid analysis. For nucleotide sequence analysis of the *rfaD* gene, the modified M13 phages, M13mp18 and M13mp19 (29, 44) were utilized as cloning vectors for selected restriction fragments shown in Fig. 4. Nucleotide sequences were determined by the dideoxy-chain termination method of Sanger et al. (36) as modified (42) for use with the Sequenase enzyme (United States Biochemical Corp., Cleveland, Ohio). Oligonucleotide primers (17-mers) for sequence analysis were synthesized with a DNA synthesizer (model 381A; Applied Biosystems, Inc., Foster City, Calif.).

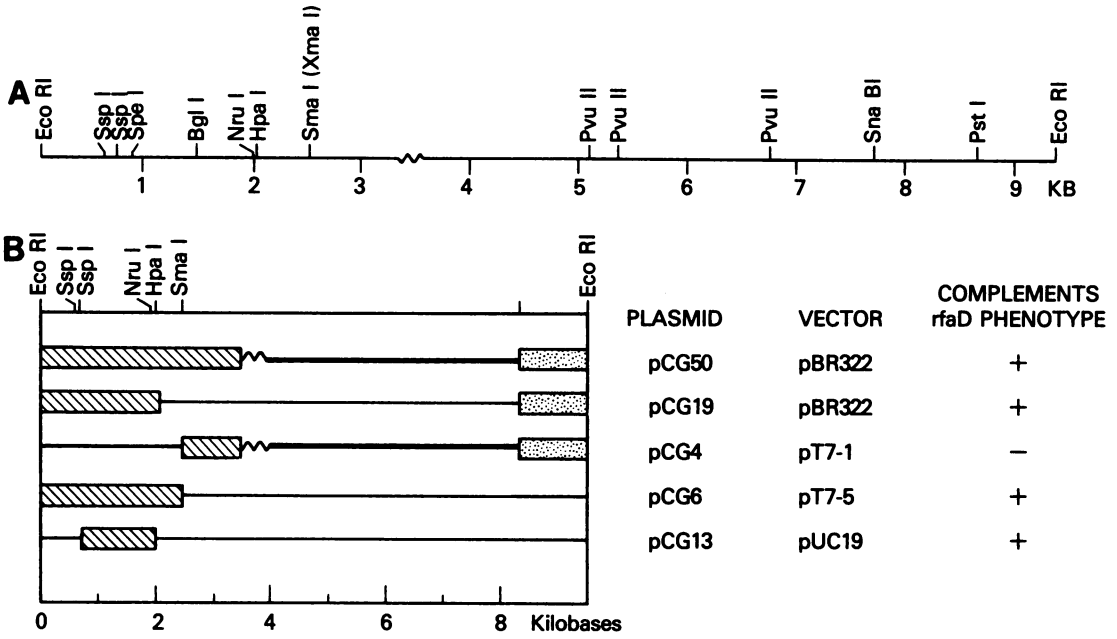

FIG. 3. Schematic illustration of restriction endonuclease sites and *rfaD* plasmid constructs. (A) Restriction map of an RfaD$^+$ *Eco*RI restriction fragment (9.0 kb), designated fragment A. (B) Cloned fragments, vectors, and complementation phenotypes. Only restriction sites relevant to plasmid construction are shown. Symbols: ▨, *E. coli* K-12 DNA; ▩, pBR322 DNA; ▬, colicin E1 DNA; ——, deletions; and ∿∿, dA-dT homopolymer.

For N-terminal amino acid analysis, 10 μg of the purified enzyme was subjected to automated Edman analysis (18) with the Applied Biosystems model 477A protein sequenator.

Primer extension analysis. A 37-mer oligonucleotide designed to anneal to nucleotides 1029 to 1066 (Fig. 5) was synthesized and used as primer in the primer extension method described by Shelness and Williams (38). Resulting primer extension products were applied on a DNA sequencing gel, and product size was determined by comparison with an adjacent sequencing ladder (36) after autoradiography.

RESULTS

Localization of *rfaD* gene on fragment A and complementation studies. A physical map of the 9-kb *Eco*RI-restricted fragment A is shown in Fig. 3A. Restriction of pCG50 with *Hpa*I-*Sna*BI removed the middle portion (5.2 kb) of fragment A, leaving approximately 2.0 kb on each end (Fig. 3A and B). After ligation, the resultant plasmid, pCG19, was tested for its ability to complement *rfaD* mutants, and complementation was observed. When the *Eco*RI-*Sma*I portion of fragment A was removed (pCG4, Fig. 3B), complementation activity was abolished. Plasmid pCG6 containing the 0- to 2.5-kb *Eco*RI-*Sma*I portion of fragment A and pCG13 containing a 1.3-kb *Ssp*I-*Hpa*I fragment both complemented *rfaD* mutants. The complementation analyses summarized in Fig. 3B indicated that the *rfaD* gene is located between the *Ssp*I and *Hpa*I restriction sites.

Identification of *rfaD* gene product. The *rfaD* gene product, encoded by plasmids that complement the *rfaD* mutation (Fig. 3B), was expressed in several in vivo and in vitro expression systems in the presence of [^{35}S]methionine. The

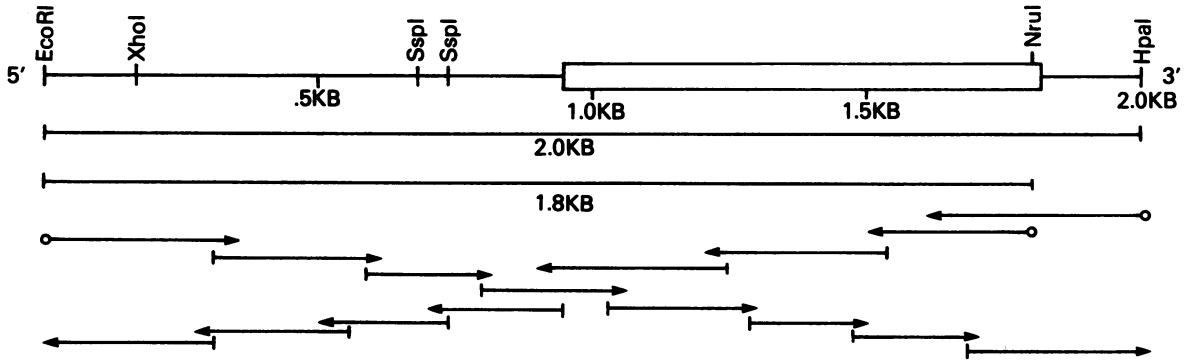

FIG. 4. Sequencing strategy for the *rfaD* gene. Two DNA fragments (a 2,037-bp *Eco*RI-*Hpa*I fragment and a 1,800-bp *Eco*RI-*Nru*I fragment) are shown schematically. The *rfaD* coding region is depicted by the boxed region. The arrows denote primers and sequencing directions (o⟶, universal primer; ⊢⟶, specific oligonucleotide primers). Relevant restriction endonuclease sites are indicated.

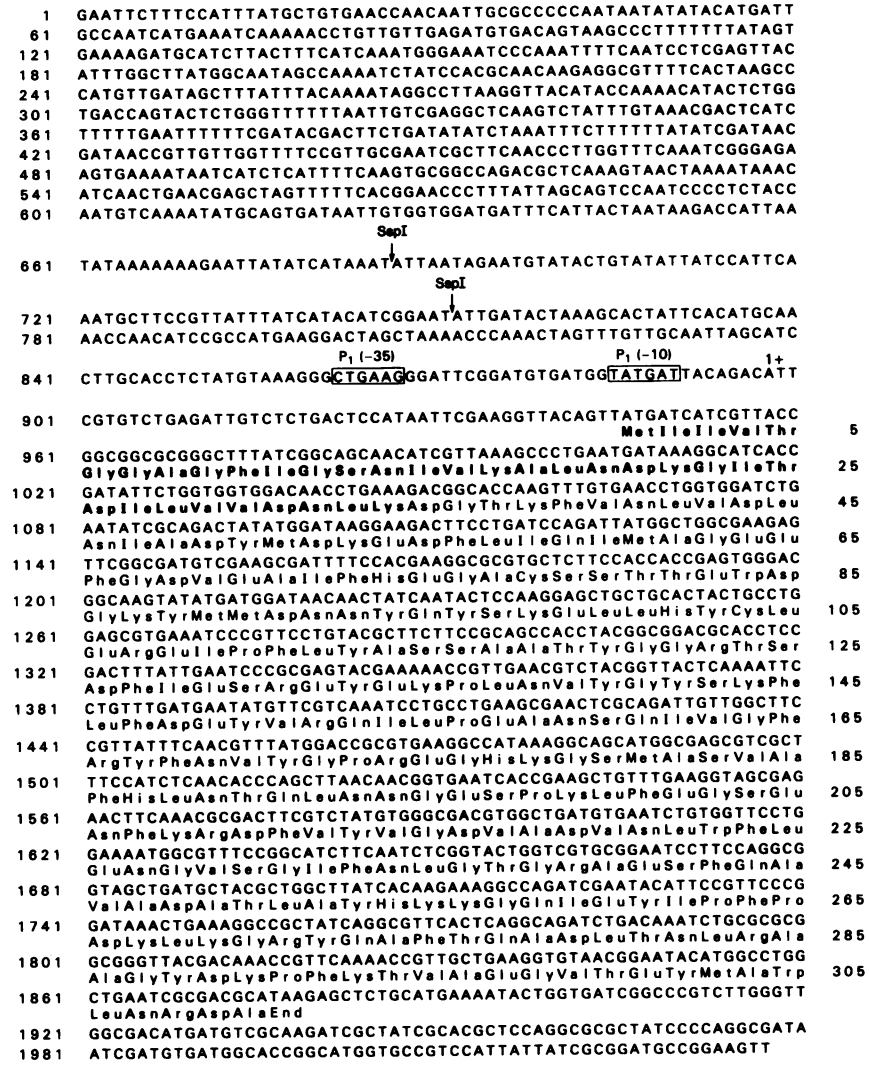

FIG. 5. Nucleotide and deduced amino acid sequences of the *rfaD* and associated reading frames. DNA sequence (GenBank accession no. M33577) of the sense strand of the *rfaD* gene is shown. Nucleotides are numbered starting with the 5′ half of the *Eco*RI site. The complete ORF includes nucleotides 946 to 1876. The amino acid residues determined by microsequence analysis are indicated by bold letters. The tentative *rfaD* promoter sequence is designated P_1.

expressed products were analyzed by polyacrylamide gel electrophoresis and autoradiography. One of the in vivo expression systems used was the T7 RNA polymerase-promoter system described by Tabor and Richardson (41). Typical experimental results are shown in Fig. 6. Lane 1 shows the results of the expression of a 1.8-kb *Eco*RI-*Nru*I fragment cloned into vector pT7-1, designated pCG5. Additional plasmids examined for expression include a 2.5-kb *Eco*RI-*Sma*I *rfaD*$^+$ fragment inserted into vectors pT7-1 and pT7-5; the resulting plasmids were designated pCG3 and pCG6, respectively (Fig. 6, lanes 2 and 5). The results consistently showed that all RfaD$^+$ plasmids expressed a 37,000-dalton polypeptide. Two major proteins, the β-lactamase doublets and a 22,000-dalton protein, were expressed by the pT7-1 vector (see the overexposed autoradiogram in Fig. 6, lane 3). The pT7-5 vector (specially constructed by S. Tabor) did not express the *bla* gene after thermal induction (lane 4). Similar results showing the presence of the 37,000-dalton polypeptide were obtained with RfaD$^+$ plasmids and linear RfaD$^+$ fragments (data not shown) with a cell-free transcription-translation system or RfaD$^+$ plasmids expressed (data not shown) in minicells (1, 7, 31).

ADP-L-glycero-D-mannoheptose-6-epimerase activity. Biochemical evidence that the RfaD$^+$ plasmids contain the sequences for the *rfaD* gene stemmed from the detection of epimerase activity in crude extracts of RfaD$^+$ strains. This activity was restored in RfaD$^-$ strains which have been complemented with RfaD$^+$ plasmids. As previously reported (9), very little if any epimerase was detected in the *rfaD* mutant strain CL515. In contrast, RfaD$^+$ strains and plasmid-complemented RfaD$^-$ strains contained significantly higher levels of epimerase activity (Table 2). Further, it was observed that cell-free protein expression reactions directed by linear RfaD$^+$ DNA likewise produced considerable epimerase activity.

Nucleotide sequence and deduced amino acid sequence of ADP-L-glycero-D-mannoheptose-6-epimerase. The sequences of a 2,037-base-pair (bp) *Eco*RI-*Hpa*I fragment containing the *rfaD* locus were determined for both strands. A detailed restriction map and the specific DNA fragments used in

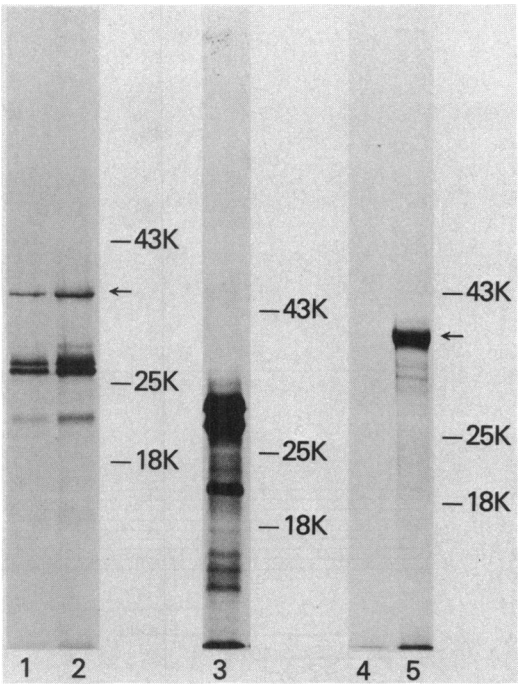

FIG. 6. Autoradiogram of the expression product ([^{35}S]methionine radiolabeled) of the plasmid-borne *rfaD* gene in the T7 promoter-polymerase system. Lanes: 1, pCG5; 2, pCG3; 3, vector pT7-1; 4, vector pT7-5; and 5, pCG6. Molecular weights were estimated with ^{14}C-labeled protein standards (Bethesda Research Laboratories, M_r 14,300 to 200,000). K, $\times 10^3$.

sequencing are shown in Fig. 3 and 4. Analyses of the DNA sequence (Fig. 5) indicated one large opening reading frame (ORF), and putative promoters were also identified upstream of the ORF. This ORF corresponds to the region which expresses the 37,000-dalton protein and that which complements the RfaD$^-$ phenotype. In addition, amino-terminal sequence analysis of the purified ADP-L-glycero-D-mannoheptose-6-epimerase confirmed the first 34 residues of the deduced amino acid sequence (Fig. 5). Further, the preponderance of acidic residues indicated by the deduced amino acid sequence was confirmed by the amino acid composition analyses (Table 3) and the pI (5.85) of the purified *rfaD* gene product. There is significant concordance of the deduced amino acid composition with the experimentally determined composition.

TABLE 2. ADP-L-glycero-D-mannoheptose-6-epimerase activities in various strains with or without the *rfaD* gene[a]

Strain	Sp act (U/mg of protein)[b]
CL510 (*rfaD*$^+$)[c]	9.3
CL515 (*rfaD*)	<1[d]
CL515(pCG50)	23
CL515(pCG6)	43
K38(pCG6, pGP1-3)[e]	290

[a] Assay procedure (see Materials and Methods).
[b] One unit equals 1 nmol of ADP-L-glycero-D-mannoheptose produced in 30 min at 37°C. Samples of 40 μg of crude extracts were assayed except in the reactions with crude extracts (80 μg added per reaction) from CL515(pCG1) and CL515(pCG4).
[c] CL510 is isogenic to CL515 except for the *rfaD* mutation.
[d] The limit of detection of this method is 1 nmol of nucleotide at 0.05 absorbance unit full scale.
[e] The epimerase was expressed in the T7 RNA polymerase expression system (41). Samples of 4 μg of crude extracts were used in these assays.

TABLE 3. Amino acid composition of ADP-L-glycero-D-mannoheptose-6-epimerase

Amino acid	Residues	
	Observed[a]	Deduced[b]
Asx[c]	35	40
Asp		(21)
Asn		(19)
Glx[c]	27	32
Glu		(23)
Gln		(9)
Serine	12	16
Glycine	26	28
Histidine	4	5
Threonine	13	14
Alanine	22	25
Arginine	10	11
Proline	6	8
Tyrosine	17	19
Valine	19	20
Methionine	6	7
Isoleucine	15	17
Leucine	23	24
Phenylalanine	17	21
Lysine	15	18
Cysteine		2
Tryptophan		3

[a] The tryptophan and cysteine residues were not determined. Calculated number of residues for M_r = 37,000.
[b] Amino acid residues deduced from the nucleotide sequence.
[c] Asx represents the summation of Asp and Asn, which were individually deduced from the nucleotide sequence. Glx represents the summation of Glu and Gln.

ADP-binding βαβ fold within the RfaD protein. It has been demonstrated in the families of flavin adenine dinucleotide (FAD)-binding and NAD-binding proteins that a particular βαβ unit constituted the binding site for the ADP moiety of the dinucleotides (33, 43). These ADP-binding folds are situated near the amino termini, and all contain the fingerprint sequence Gly-X-Gly-X-X-Gly (43). In Table 4, the region of ADP-L-glycero-D-mannoheptose-6-epimerase containing the putative ADP-binding site is compared with analogous regions of several FAD- and NAD-binding pro-

TABLE 4. ADP-binding βαβ fold of ADP-L-glycero-D-mannoheptose-6-epimerase

Protein (peptide)[a]	Sequence[b]	Reference
NH_2 terminal ADPHE	(2) IIVGAGFIGSNIVK	
LipDH (FAD)	(10) VVLGAGPAGXSAAF	39
NADHDH (FAD)	(8) VIVGGGAGGLEMAT	45
GLPDH (FAD)	(6) IVIGGGINGAGIAA	6
GPDH[c] (NAD)	(5) AINGFGRIGRLVRI	19
GPDH[d] (NAD)	(7) GVNGFGRIGRLVTR	2
LDH (NAD)	(25) TVVGVGAVGMACAI	14
Conserved residues	* * *	

[a] ADPHE, ADP-L-glycero-D-mannoheptose-6-epimerase; LipDH, lipoamide dehydrogenase; NADHDH, NADH dehydrogenase; GLPDH, glycerol-3-phosphate dehydrogenase; GPDH, glyceraldehyde-3-phosphate dehydrogenase; and LDH, lactate dehydrogenase.
[b] The number preceding the sequence is the residue number of the first amino acid shown relative to the amino terminus of the protein.
[c] Source, yeasts.
[d] Source, human.

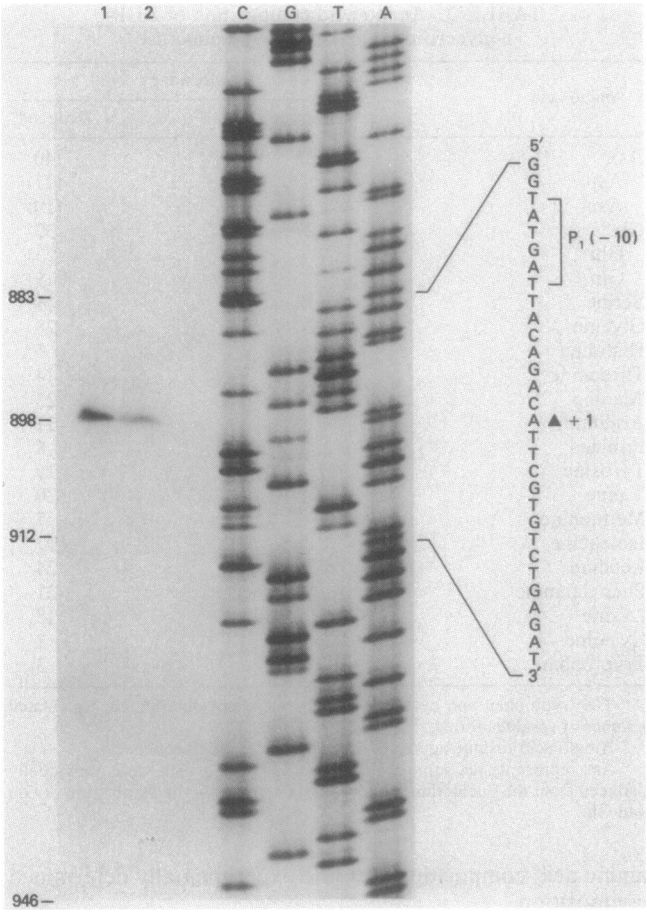

FIG. 7. Primer extension of the *rfaD* transcript. An [α-^{32}P]ATP-end-labeled oligonucleotide primer corresponding to base positions 1029 through 1066 within the *rfaD* coding region was annealed to total cellular RNA from strain CL529. It was extended with reverse transcriptase, and the products (lanes 1 and 2) were analyzed on a DNA sequencing gel. The lanes labeled C, G, T, and A are the dideoxy sequencing reactions of the identical primer and the *rfaD* antisense strand cloned into M13mp18 as a template. The numbers on the left indicate the base position of the *rfaD* sequence. The sequence between bases 912 and 883 is indicated on the right. The symbol ▲ indicates the transcriptional start site, and this position is number +1. The −10 hexamer of the *rfaD* promoter is bracketed.

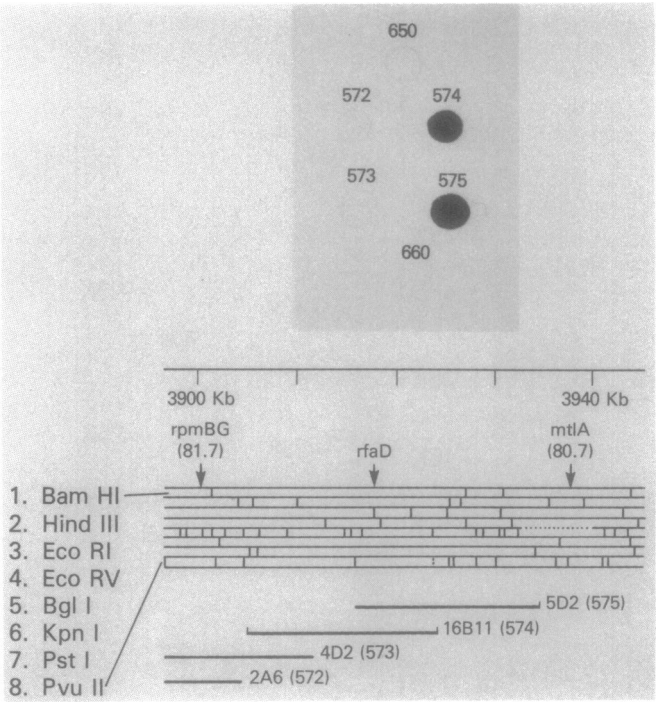

FIG. 8. Location of the *rfaD* gene on the *E. coli* physical map. The upper part of this figure shows the results of DNA-DNA hybridization of the 1.3-kb *Ssp*I-*Hpa*I fragment from pCG13 to selected λ-*E. coli* clones (572 to 575) containing the region of the chromosome located between 3,880 and 3,935 kilobase pairs. Clones 650 and 660 containing DNA from the chromosome kilobase pair regions 4450 to 4460 and 4580 to 4590, respectively, were used as controls. The lower portion of this figure shows the physical map of the region of the *E. coli* map near the *rfaD* locus and the inserts in various λ phage tested in the hybridization experiment.

teins. Comparative analyses show that the highly conserved fingerprint is present in ADP-L-glycero-D-mannoheptose-6-epimerase.

Mapping of 5′ end of *rfaD* mRNA. The *rfaD* transcriptional start site was identified by determining the 5′ end of the *rfaD* mRNA by primer extension (38). A 37-mer primer designed to anneal 83 nucleotides downstream of the ATG start codon was annealed to total RNA extracted from the RfaD$^+$ plasmid-containing strain, CL529. The mapping results of the 5′ end of *rfaD* mRNA are shown in Fig. 7. A single major product was resolved by electrophoresis on a sequencing gel simultaneously with the dideoxy-chain termination ladder (36) produced with the identical primer and the *rfaD* antisense strand of the 1.3-kb *Ssp*I-*Hpa*I RfaD$^+$ plasmid fragment cloned into M13mp18. We concluded that the transcription start site is at nucleotide 898. This result is also consistent with our tentative assignment of the *rfaD* promoter (see Discussion below).

Location of *rfaD* gene on physical map of *E. coli* chromosome. Figure 8 shows the results of hybridizing the 1,300-bp *Ssp*I-*Hpa*I RfaD$^+$ fragment to λ phage clones from the genomic library bank of Kohara et al. (22). Based on our previous chromosomal location of the *rfaD* locus by P1 transductional analysis (10, 11), we chose library clones 572, 573, 574, and 575. These clones span the 3880- to 3935-kb region of the *E. coli* chromosome. As controls, we chose library member clones 650 and 660, which span the 4450- to 4460-kb region. The results shown in Fig. 8 indicate that only clones 574 and 575 hybridized with the *rfaD* sequence probe. These overlapping DNA clones indicate a chromosomal location between 3916 and 3925 kb on the physical map of Kohara et al. (22). In addition, using the new restriction alignment software of Rudd et al. (34), a *Bgl*I restriction site located 526 bp downstream of the *rfaD* ORF was located at 3916.8 kb on the *E. coli* physical map.

DISCUSSION

In the present study, we extended the earlier studies by cloning and sequencing the *rfaD* gene and by purification and initial characterization of the *rfaD* gene product. The *rfaD* gene was located on the *E. coli* physical map by partial analysis of the genomic library bank of Kohara et al. (22). By 5′-end mapping of the *rfaD* mRNA, we defined the transcription start site and thus provided support for our tentative assignment of the *rfaD* promoter region. The motif of the putative *rfaD* promoter sequence shows consensus with an extensive compilation and comparison of promoters for *E.*

coli (17). The *rfaD* gene regulatory and coding regions are located on a 1.3-kb *Ssp*I-*Hpa*I fragment.

The *rfaD* gene encodes a functional enzyme, ADP-L-glycero-D-mannoheptose-6-epimerase. As predicted from the information obtained from the *rfaD* gene nucleotide sequence and amino acid sequence of the N terminus of the purified *rfaD* product, we defined the *rfaD* coding region. Further analysis of the primary structure of the N terminus of the RfaD protein suggested a possible ADP-binding site. The amino acid fingerprint sequence (e.g., G-X-G-X-X-G) and the predicted secondary structure of the N terminus of the *rfaD* protein is consistent with the presence of the ADP-binding βαβ fold (43).

ACKNOWLEDGMENTS

We thank Stanley Tabor (Harvard University) for providing us with information, strains, and plasmids necessary for setting up the T7 polymerase-expression system, Malcolm Moos (Food and Drug Administration) for the amino-terminal analysis of the epimerase, Joe N. Davis (National Institutes of Health) for his assistance with determination of the amino acid composition of the epimerase, Belinda Seto (National Institutes of Health) for helpful suggestions during the development of a purification protocol for the epimerase, and Helen C. Jenerick for her help in typing the manuscript.

LITERATURE CITED

1. **Adler, H. I., W. D. Fisher, A. Cohen, and A. A. Hardigree.** 1967. Miniature *E. coli* cells deficient in DNA. Proc. Natl. Acad. Sci. USA **57**:321–326.
2. **Arcuri, P., R. Martinelli, and F. Salvatore.** 1984. The complete sequence of the full length cDNA for human liver glyceraldehyde-3-phosphate dehydrogenase: evidence for multiple mRNA species. Nucleic Acids Res. **12**:9179–9189.
3. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
4. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**:248–254.
5. **Chen, H., and G. Zubay.** 1983. Prokaryotic coupled transcription-translation. Methods Enzymol. **101**:674–690.
6. **Choi, Y. L., S. Kawase, M. Kawamukai, R. Utsumi, H. Sakai, and T. Komano.** 1989. Nucleotide sequence of the glycerol-3-phosphate dehydrogenase gene of *Escherichia coli* and regulation by cAMP-CRP complex. Agric. Biol. Chem. **53**:1135–1143.
7. **Clark-Curtiss, J. E., and R. Curtiss III.** 1983. Analysis of recombinant DNA using *Escherichia coli* minicells. Methods Enzymol. **101**:347–363.
8. **Clarke, L., and J. Carbon.** 1975. Biochemical construction and selection of hybrid plasmids containing specific segments of the *Escherichia coli* genome. Proc. Natl. Acad. Sci. USA **72**: 4361–4365.
9. **Coleman, W. G., Jr.** 1983. The *rfaD* gene codes for ADP-L-glycero-D-mannoheptose-6-epimerase, an enzyme required for lipopolysaccharide core biosynthesis. J. Biol. Chem. **258**: 1985–1990.
10. **Coleman, W. G., Jr., and K. S. Deshpande.** 1985. New cysE-pyrE-linked *rfaD* mutation in *Escherichia coli* K-12 that results in a heptoseless lipopolysaccharide. J. Bacteriol. **161**:1209–1214.
11. **Coleman, W. G., Jr., and L. Leive.** 1979. Two mutations which affect the barrier function of the *Escherichia coli* K-12 outer membrane. J. Bacteriol. **139**:899–910.
12. **Davis, B. D., and E. S. Mingioli.** 1950. Mutants of *Escherichia coli* requiring methionine or vitamin B_{12}. J. Bacteriol. **60**:17–28.
13. **Eidels, L., and J. J. Osborn.** 1974. Phosphoheptose isomerase, first enzyme in the biosynthesis of aldoheptose in *Salmonella typhimurium*. J. Biol. Chem. **249**:5642–5648.
14. **Eventoff, W., M. G. Rossmann, S. S. Taylor, H. J. Torff, H. Meyer, W. Keil, and H. H. Kiltz.** 1977. Structural adaptations of lactate dehydrogenase isozymes. Proc. Natl. Acad. Sci. USA **74**:2677–2681.
15. **Ginsburg, V., P. J. O'Brien, and C. W. Hall.** 1962. Guanosine diphosphate D-glycero-D-mannoheptose. J. Biol. Chem. **237**: 497–499.
16. **Hare, P. E.** 1977. Subnanomole-range amino acid analysis. Methods Enzymol. **47**:3–18.
17. **Harley, C. B., and R. P. Reynolds.** 1987. Analysis of *E. coli* promoter sequences. Nucleic Acids Res. **15**:2343–2361.
18. **Hewick, R. M., M. W. Hunkapiller, L. E. Hood, and W. J. Dreyer.** 1981. A gas-liquid solid phase peptide and protein sequenator. J. Biol. Chem. **256**:7790–7997.
19. **Holland, J. P., and M. J. Holland.** 1979. The primary structure of a glyceraldehyde-3-phosphate dehydrogenase gene from *Saccharomyces cerevisiae*. J. Biochem. **254**:9839–9845.
20. **Hyman, H. C., R. Gafny, G. Glaser, and S. Razin.** 1988. Promoter of the mycoplasma pneumonia rRNA operon. J. Bacteriol. **170**:3262–3268.
21. **Kocsis, B., and T. Kontrohr.** 1984. Isolation of adenosine 5′-diphosphate-L-glycero-D-mannoheptose, the assumed substrate of heptose transferase(s), from *Salmonella minnesota* R595 and *Shigella sonnei* Re mutants. J. Biol. Chem. **259**: 11858–11860.
22. **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.
23. **Kontrohr, T., and B. Kocsis.** 1981. Isolation of adenosine 5′-diphosphate-D-glycero-D-mannoheptose, an intermediate in lipopolysaccharide biosynthesis of *Shigella sonnei*. J. Biol. Chem. **256**:7715–7718.
24. **Kontrohr, T., and B. Kocsis.** 1986. Separation and quantification of ADP-D-glycero-D-mannoheptose and ADP-L-glycero-D-mannoheptose by borate liquid chromatography and its use in the assay of ADP-D-glycero-D-mannoheptose-6-epimerase. J. Chromatogr. **354**:417–423.
25. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
26. **Lugtenberg, B., J. Meijers, R. Peters, P. van der Hoek, and L. Alphen.** 1975. Electrophoretic resolution of the major outer membrane protein of *Escherichia coli* K-12 into four bands. FEBS Lett. **58**:254–259.
27. **Luria, S. E., and J. W. Burrows.** 1957. Hybridization between *Escherichia coli* and *Shigella*. J. Bacteriol. **74**:461–476.
28. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 93. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
29. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:21–78.
30. **Minton, A. P., and A. K. Attri.** 1986. Lap computers in the laboratory. Comput. Appl. Biosci. **2**:167–171.
31. **Reeves, J.** 1979. Use of minicells for bacteriophage-directed polypeptide synthesis. Methods Enzymol. **68**:493–503.
32. **Rodriguez, R. L., and R. C. Tait.** 1983. Recombinant DNA techniques: an introduction, p. 42–43. Addison-Wesley Publishing Co., Reading, Mass.
33. **Rossmann, M. G., D. Moras, and K. W. Olsen.** 1974. Chemical and biological evolution of nucleotide-binding protein. Nature (London) **250**:194–199.
34. **Rudd, K. E., W. Miller, J. Ostell, and D. A. Benson.** 1990. Alignment of *Escherichia coli* K-12 DNA sequences to a genomic restriction map. Nucleic Acids Res. **18**:313–321.
35. **Russell, M., and P. Model.** 1984. Replacement of the *fip* gene of *Escherichia coli* by an inactive gene cloned on a plasmid. J. Bacteriol. **159**:1034–1039.
36. **Sanger, F. S., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
37. **Schleif, R., and Wensink, P.** 1981. Practical methods in molecular biology, p. 123–124. Springer-Verlag, New York.
38. **Shelness, G. S., and D. L. Williams.** 1984. Apo lipoprotein II messenger RNA transcription and splicing heterogeneity yield

six 5′ untranslated leader sequences. J. Biol. Chem. **259:** 9929–9935.

39. **Stephens, P. E., H. M. Lewis, M. G. Darlison, and J. R. Guest.** 1983. Nucleotide sequence of the lipoamide dehydrogenase gene of *Escherichia coli* K-12. Eur. J. Biochem. **135:**519–527.
40. **Tabor, S., H. E. Huber, and C. C. Richardson.** 1987. *Escherichia coli* thioredoxin confers processivity on the DNA polymerase activity of the gene 5 protein of bacteriophage T7. J. Biol. Chem. **262:**16212–16223.
41. **Tabor, S., and C. C. Richardson.** 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA **82:**1074–1078.
42. **Tabor, S., and C. C. Richardson.** 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **84:**4767–4771.
43. **Wierenga, R. K., P. Terpstra, and W. G. J. Hoi.** 1986. Prediction of the occurrence of the ADP-binding βαβ-fold in proteins using an amino acid sequence fingerprint. J. Mol. Biol. **187:**101–107.
44. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 pUC19 vectors. Gene **33:**103–119.
45. **Young, I. G., B. L. Rogers, H. D. Campbell, A. Jaworowski, and D. C. Shaw.** 1981. Nucleotide sequence coding for the respiratory NADH dehydrogenase of *Escherichia coli*. Eur. J. Biochem. **116:**165–170.