APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Jan. 1984, p. 15-21
0099-2240/84/010015-07$02.00/0

Vol. 47, No. 1

Cloning and Sequencing of the Xylose Isomerase and Xylulose Kinase Genes of *Escherichia coli*

V. BRYAN LAWLIS,[1]†* MARK S. DENNIS,[1] ELLSON Y. CHEN,[2] DOUGLAS H. SMITH,[2] AND DENNIS J. HENNER[2]

Departments of Biocatalysis[1] and Molecular Biology,[2] Genentech, Inc., South San Francisco, California 94080

Received 27 May 1983/Accepted 10 October 1983

A 4.2-kilobase-pair fragment of the *Escherichia coli* chromosome which contains the genes for xylose isomerase and xylulose kinase was cloned into plasmid pBR322. The hybrid plasmid (designated pECX14) complements strains deficient in either or both of the two enzymes. Deletion derivatives of pECX14 were used to localize the two genes on the cloned fragment. The entire nucleotide sequence of the cloned fragment was determined. Open reading frames which, if translated, would encode proteins of molecular weights 54,000 and 52,000 were found. These were identified as the isomerase and kinase structural genes, respectively.

Catabolism of D-(+)-xylose (xylose) involves the transport of xylose into the cell, the isomerization of the xylose to D-xylulose, and the phosphorylation of the xylulose to D-xylulose-5-phosphate. This phosphorylated intermediate is then metabolized by the pentose phosphate and Embden-Meyerhoff pathways. In *Aerobacter aerogenes* (23), *Escherichia coli* (6), and *Salmonella typhimurium* (20, 21), the xylose transport, isomerase, and kinase are all induced by growth on xylose. In *S. typhimurium*, the genes for these activities are organized in a cluster which appears to be regulated by a positive control mechanism (21). In *E. coli*, the isomerase, kinase, and a regulatory gene appear to be clustered within a cloned 9.7-kilobase pair fragment (12, 24). In that investigation, no sequence data and no assignment of the genes for xylose utilization were reported.

This study describes the cloning of the genes for xylose isomerase and xylulose kinase from *E. coli* by complementation of a mutant deficient in xylose utilization. The identification of each gene by subcloning experiments and enzyme assays is described, and the complete nucleotide sequence of both genes is also reported.

MATERIALS AND METHODS

Bacterial strains and plasmids. Strain MM294 (F^- *supE44 endA1 thi-1 hsdR4*) was used for plasmid constructions (1). A library of random Tn*10* insertions in strain N99 (*galK strR*) was provided by H. Miller. Strain JC1553 (*argG6 metB1 his-1 leu-6 recA1 mtl-2 xyl-7 gal-6 lacY1 supE44 malA1 tonA2 rpsL-104 lamR lambda*$^-$ F^-) was obtained from the Northern Regional Research Laboratory culture, accession number 11391 (depositor, Merle Woucha). Strain JC1553 appears to lack xylose isomerase activity, xylulose kinase activity, and xylose transport (unpublished data; M. Woucha, personal communication). Plasmid pBR322 was used as the cloning vector (3).

Enzymes and chemicals. Restriction enzymes, T4 DNA ligase, *Bal*31, and the Klenow fragment of DNA polymerase I were obtained through commercial suppliers and used according to specifications provided by the supplier. Xylulose, xylose, and ribulose were purchased from Sigma Chemical Co. Transcription-translation kits were purchased from Codon Laboratories and used according to the instructions supplied with the kits.

Growth and transformation liquid media used in this study were Luria broth and minimal A. Solid media used were Luria broth, eosin methylene blue (EMB)-xylose, and minimal A (15). Additives to minimal A media were required amino acids (40 μg/ml) and xylose (0.2%). Ampicillin (20 μg/ml) and tetracycline (5 μg/ml) were added at the indicated concentrations in selective media. Cells were made competent for transformation, using the $CaCl_2$ shock method (8). Selection for loss of the Tn*10* transposon was done on media containing fusaric acid as described before (13). P1 transduction, ampicillin enrichment, and ethyl methanesulfonate mutagenesis were as described previously (15).

Enzyme assays. Cells were grown to mid-log phase in medium A containing 0.2% Casamino Acids with or without 0.2% xylose. For strains carrying a plasmid, 20 μg of ampicillin per ml was also added. Cultures were harvested by centrifugation at 7,000 rpm for 10 min in a Sorvall GSA rotor. Cellular extracts were prepared by resuspending 50 to 100 mg of frozen cells in 0.5 ml of 10 mM Tris-hydrochloride–1 mM EDTA, pH 7.5, containing 5 mg of lysozyme per ml. After 10 min on ice the samples were briefly sonicated. Cellular debris was removed by centrifugation for 3 min in a Fisher model 235A microcentrifuge. Protein concentrations were measured by either the biuret method (9) or the Bio-Rad kit (Bio-Rad Laboratories) (4).

The activity of xylose isomerase was measured at 60°C by the rate of xylulose formation. In 0.5 ml, the reaction contained 25 μmol of Tris-hydrochloride (pH 7.5), 9.0 μmol of $MgSO_4$, 1.4 μmol of $CaCl_2$, and 90 μmol of xylose. The reaction was initiated by the addition of enzyme to the prewarmed reaction mixture. Samples, 0.02 ml, were taken at 0, 2, 5, and 10 min and mixed with 0.18 ml of 0.1 N HCl to quench the reaction. Xylulose was measured by the cysteine-carbazol method and compared with a standard curve of ribulose (7).

The xylulose kinase assay was performed essentially as described by Simpson (22). The assay is a coupled reaction in which the rate of xylulose-dependent ADP formation is measured. In 1 ml, the reaction contained 50 μmol of Tris-hydrochloride (pH 7.0), 5.0 μmol of $MgCl_2$, 1.0 μmol of EDTA, 1.0 μmol of phosphoenolpyruvate, 0.5 μmol of ATP, 1.0 μmol of xylulose, 0.01 μmol of NADH, and 80 U each of

* Corresponding author.
† Present address: Genencor, Inc., South San Francisco, CA 94080.

lactate dehydrogenase and pyruvate kinase. The reaction was initiated by addition of the cell extract. The rate of NADH disappearance was monitored at 340 nm ($\epsilon = 6.22 \times 10^3$/mol per cm) on a Gilford 2600 spectrophotometer equipped with a microprocessor. High NADH oxidase activity was observed; therefore, the machine was programmed to automatically subtract the rate of a control reaction without xylulose from a reaction containing xylulose.

Transport was measured by harvesting 4 ml of a mid-log culture grown as above, washing with minimal A medium (with no carbon source), and resuspending in 4 ml of minimal A medium. D-[*U*-^{14}C]xylose (76 Ci/mmol; Amersham Corp.) was added at 1.8 μM and incubation was at 37°C. A 1-ml sample was taken at 10, 30, and 60 s, filtered on a 0.45-μm filter (Amicon), and washed with 5 ml of minimal A medium. Filters were dried and counted after the addition of scintillation fluid (Scintisol; Isolab).

DNA sequencing. DNA sequencing was by the dideoxy chain termination method (19). The single-stranded M13 cloning vectors mp8 and mp9 were used (14). Specific restriction fragments were purified by elution from acrylamide gels and ligated into appropriately digested M13 vectors. Certain fragments were shortened by using the processive nuclease *Bal*31 befor cloning into M13 vectors (11). For fragments in which compressions made sequence determination difficult, dITP was substituted for dGTP in the reaction mix (16).

DNA isolation. Large-scale plasmid preparations were made by the cleared lysate procedure and further purified by Bio-Gel A6 chromatography (5). Small-scale plasmid preparations were made by the sodium dodecyl sulfate-NaOH precipitation method (2). *E. coli* chromosomal DNA was prepared as described previously (18), with an additional step of purification by banding on a CsCl gradient.

Gel electrophoresis and electroelution. Agarose gels were 1% Seakem agarose in 40 mM Tris-hydrochloride–5 mM sodium acetate–1 mM EDTA, pH 7.9, and electrophoresis was in the same buffer. Acrylamide gels were 5% acrylamide in 90 mM Tris–90 mM boric acid–3 mM EDTA, pH 8.0 (TBE), and electrophoresis was in the same buffer. Fragments were purified from gels by electroelution into a dialysis bag in 0.1× TBE as previously described (10).

RESULTS

Selection of mutants deficient in xylose utilization. Mutants deficient in xylose utilization were isolated by two methods. The first method entailed screening a pool of random Tn*10* insertions in strain N99 for colonies unable to utilize xylose. Cells from the pool were plated on EMB-xylose-tetracycline plates, and white colonies, indicating an inability to utilize xylose, were picked. After two rounds of purification for single colonies, xylose isomerase and xylulose kinase assays were performed. One such mutant was determined to be lacking both the isomerase and kinase functions. P1 transduction was used to move the transposon from that mutant into strain MM294 by selection for tetracycline resistance on EMB-xylose plates. The majority of the tetracycline-resistant transductants were Xyl$^-$, and characterization showed that these lacked both the isomerase and kinase functions. The Tn*10* transposon was eliminated by selection on fusaric acid plates. One of the Xyl$^-$, fusaric-acid resistant, tetracycline-sensitive colonies was chosen and characterized. Enzyme assays showed that this derivative was also lacking both the kinase and isomerase functions. This particular mutation was designated *xyl-22*, and the strain was referred to as BL002. Transport assays showed that this strain was able to transport xylose (data not shown).

The second method used to isolate Xyl$^-$ mutants was mutagenesis of strain MM294 with ethyl methanesulfonate. After two rounds of ampicillin enrichment for cells unable to grow with xylose as a carbon source, cultures were screened for xylose utilization mutants on EMB-xylose plates. One particular mutant was found to be lacking the isomerase function and had both the kinase activity and transport activities; this mutation was designated *xyl-20*, and the strain was referred to as BL001.

Selection for plasmids complementing Xyl$^-$ mutants. A random partial *Sau*3A bank of *E. coli* DNA was constructed in the following manner. *E. coli* DNA was partially digested with *Sau*3A, using conditions that gave DNA fragments ranging from 4 to 9 kilobase pairs in length. This partially digested DNA was electrophoresed on an acrylamide gel, and fragments >3 kilobase pairs in length were eluted from the gel. These eluted fragments were ligated into pBR322, which had been digested with *Bam*HI and treated with alkaline phosphatase. The ligation mixture was transformed into strain MM294, and transformants were selected on Luria broth-ampicillin plates. Approximately 5,000 transformants were pooled by scraping the colonies from the plates, and plasmid DNA was prepared from the pooled transformants. The pooled plasmid DNA was transformed into strain BL001, and transformants were selected on minimal agar plates containing ampicillin and xylose as the sole carbon source. Colonies that appeared after 2 days were picked and plasmid DNA was prepared. These plasmids were transformed into BL002 with selection on EMB plates containing xylose and ampicillin. Those plasmids that were able to transform BL001 Xyl$^+$, as indicated by red colonies on EMB-xylose plates, were examined for their restriction pattern, using restriction endonuclease *Hae*III. Since all eight of the isolates had identical *Hae*III restriction patterns, one isolate, designated pECX14, was chosen for further characterization.

Complementation of xylose-deficient mutants. Plasmid pECX14 was transformed into strains JC1553, BL001, and BL002 with selection on EMB-ampicillin-xylose plates and minimal-xylose plates. The presence of the plasmid in all three strains enabled them to utilize xylose as indicated by red colonies on the EMB-xylose plates and growth on the minimal-xylose plates. Transformation of the parent plasmid, pBR322, into the same three strains did not allow utilization of xylose. Since strain BL001 is deficient in xylose isomerase and JC1553 and BL002 are deficient in both xylose isomerase and xylulose kinase, the plasmid probably contains genes coding for both of these functions. However, there is the possibility that these mutants are regulatory mutants and that pECX14 carries a regulatory element which complements the mutations in one or more of these xylose-deficient strains. To differentiate between these possibilities and to localize the coding regions for the different functions, various deletion and insertion derivatives of pECX14 were constructed. These derivatives were tested for complementation of Xyl$^-$ strains, and strains carrying these derivatives were assayed for expression of xylose isomerase and kinase.

Deletion derivatives of pECX14. Figure 1 is a restriction map of the insert in plasmid pECX14 and the surrounding region of pBR322. Deletion derivatives were constructed by ligation of purified fragments of pECX14 and pBR322. For instance, derivative pECX14-A was constructed by ligation of the 2,000-base pair (bp) *Eco*RI-*Bgl*II fragment of pECX14

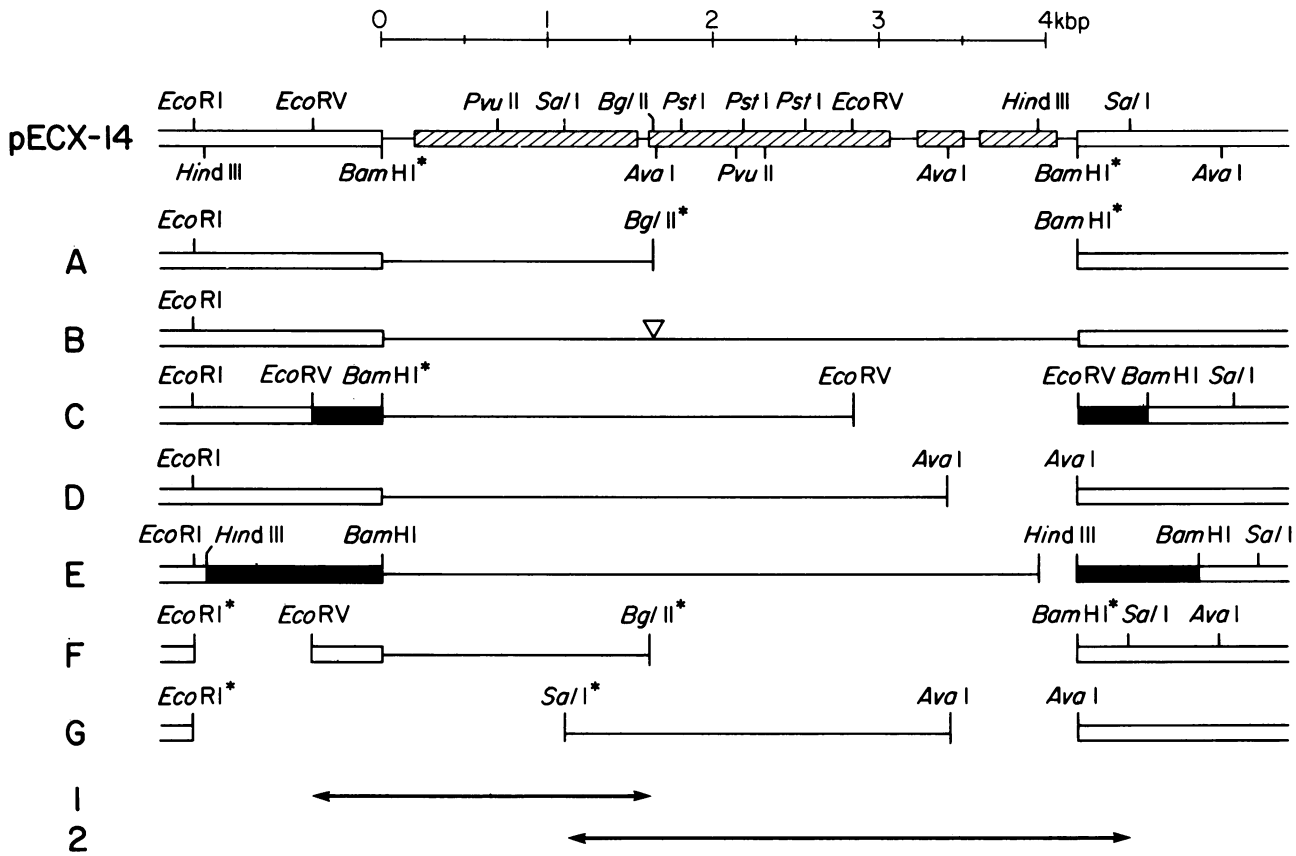

FIG. 1. Map of pECX-14 and derivatives. The physical map of plasmid pECX-14 and each of its derivatives described in the text is shown. Restriction sites which were used for construction, but which were destroyed in the construction process, are indicated by a *. The bottom two lines indicate the restriction fragments used to program the in vitro transcription-translation reactions as described in the text. Symbols: ▽, insertion of 4 bp; □, portion of pBR322; ■, portion of pBR322 duplicated during the construction process; ▨, open reading frames.

to the 4,000-bp *Eco*RI-*Bam*HI fragment of pBR322. In some instances, three-part ligations were needed: for derivative pECX14-C the 2,000-bp *Eco*RI-*Bgl*II fragment of pECX14, the 1,200-bp *Bgl*II-*Eco*RV fragment of pECX14, and the 4,200-bp *Eco*RI-*Eco*RV fragment of pBR322 were ligated. In all instances, restriction analysis was performed to check that the plasmid had been properly constructed. Plasmid pECX14-B was constructed by opening pECX14 at the unique *Bgl*II site, repairing to flush ends with the Klenow fragment of *Pol*I, and religating the flush ends. This should create a 4-bp insert at the *Bgl*II site. Plasmid pECX14-F was constructed by restricting pECX14-A with *Eco*RI and *Eco*RV, repairing to flush ends with the Klenow fragment of *Pol*I, and ligating the flush ends. Plasmid pECX14-G was constructed by restricting pECX14-D with *Eco*RI and *Sal*I, repairing to flush ends with the Klenow fragment of *Pol*I, and ligating the flush ends. Figure 1 shows a schematic diagram of all of the derivatives used in this study, indicating which portions of the pECX14 insert DNA are contained on each and how they are inserted into the pBR322 vector. The various derivatives of pECX14 were transformed into strain JC1553, and transformants were selected on EMB-xylose-ampicillin plates. Plasmid pECX14 and derivatives D and E gave red colonies on the indicator plates, indicating that these plasmids were able to complement the xylose utilization deficiency in this strain. All other derivatives were white, indicating that they were unable to complement at least one of the enzymatic deficiencies in this strain. To determine which enzymatic activities were deficient in the various derivatives, enzyme assays of the isomerase and kinase activities were performed. The results of these assays are depicted in Table 1. It is clear from the results shown in Table 1 that derivatives D and E provide both the isomerase and kinase activities. Derivatives A and F both provide only the isomerase activities, whereas derivative G provides only the kinase activity. Comparing these data with the restriction map indicates that the isomerase gene must reside on the left-hand portion of the insert DNA (with respect to Fig. 1) and preceding the *Bgl*II site, whereas the kinase gene must reside on the right-hand portion of the insert and preceding the *Ava*I site. Since the F and G derivatives are missing the tetracycline promoter from pBR322, these data also indicate that these genes are being transcribed from their own promoter(s).

In vitro transcription-translation. In vitro transcription-translation studies were done with pECX14, several of the derivatives, and isolated DNA fragments from the pECX14 insert DNA. These studies were designed to allow us to determine the size of the protein products and to help localize the proteins on the DNA insert. Figure 2 shows a fluorogram of the products of these in vitro transcription-translation reactions. A reaction programmed with pECX14 (lane 4) showed three new protein products that were not seen in a reaction programmed with pBR322 (lane 5). These

TABLE 1. Measurement of isomerase and kinase levels

Strain	Xylose genotype	Plasmid	Phenotype (EMB plates)	Kinase (U/mg of protein)[a]		Isomerase (U/mg of protein)[a]	
				Induced	Uninduced	Induced	Uninduced
MM294	*xyl*$^+$		Red	0.06	<0.01	0.373	0.014
BL001	*xyl-20*		White	0.072	<0.01	<0.005	<0.005
	xyl-20	pECX14	Red	ND[b]	ND	0.325	0.100
	xyl-20	pECX14-A	White	ND	ND	1.040	ND
BL002	*xyl-22*		White	<0.01	<0.01	0.006	0.005
	xyl-22	pECX14	Red	ND	ND	0.510	0.230
	xyl-22	pECX14-A	White	ND	ND	0.917	ND
JC1553	*xyl-7*		White	<0.01	<0.01	0.042	0.037
	xyl-7	pECX14	Red	1.88	0.79	0.418	0.191
	xyl-7	pECX14-A	White	<0.01	<0.01	0.500	0.100
	xyl-7	pECX14-B	White	ND	ND	0.325	ND
	xyl-7	pECX14-C	White	ND	ND	0.530	ND
	xyl-7	pECX14-D	Red	0.75	0.91	0.260	0.086
	xyl-7	pECX14-E	Red	0.58	1.01	0.290	0.098
	xyl-7	pECX14-F	White	<0.01	<0.01	0.350	0.150
	xyl-7	pECX14-G	White	0.50	0.20	0.025	0.025

[a] Units = micromoles of product formed per minute.
[b] ND, Not determined.

appear to migrate at approximately 56,000, 50,000, and 19,000 molecular weight and are indicated by arrows. When the reaction was programmed with pECX14-A (lane 3) or pECX14-C (lane 2), the 56,000- and 19,000-molecular-weight bands were not seen; however, the 50,000-molecular-weight band was still present. When the reaction was programmed with pECX14-B (lane 1), only the 56,000-molecular-weight band was missing. When the reactions were programmed with purified DNA fragments, the fragment containing the left half of the insert DNA (fragment 1 in Fig. 1) synthesized the 50,000-molecular-weight band (lane 7), whereas the fragment containing the right half of the insert (fragment 2 in Fig. 1) synthesized the 56,000- and 19,000-molecular-weight bands (lane 8). The complementation data above suggest that the 50,000-molecular-weight band encoded on the left half of the fragment is the xylose isomerase, and the 56,000-molecular-weight band is probably the xylulose kinase.

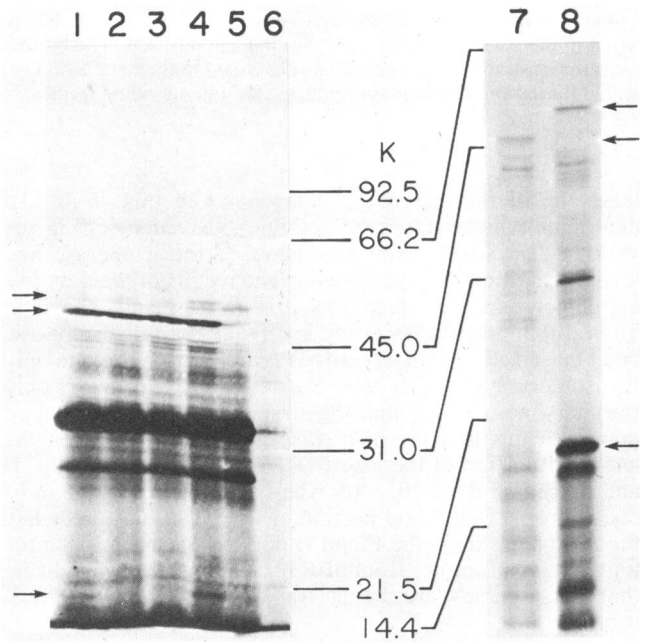

FIG. 2. In vitro transcription-translation. An in vitro transcription-translation system was programmed with plasmid DNA (lanes 1 to 6) or purified DNA fragments (lanes 7 and 8). Bands appearing in reactions programmed with pECX-14 or its derivatives, but not with pBR322, are designated with arrows. Lanes 1 to 6 were electrophoresed on a 10% acrylamide gel; lanes 7 and 8 were electrophoresed on a 12.5% acrylamide gel. The positions of stained molecular-weight markers are indicated. The DNAs used to program the reaction in each lane are: 1, pECX14-B; 2, pECX14-C; 3, pECX14-A; 4, pECX14; 5, pBR322; 6, no DNA; 7, fragment 1; 8, fragment 2.

Nucleotide sequence. The entire nucleotide sequence of the insert DNA in pECX14 was determined by the dideoxy chain termination method after cloning specific restriction fragments in M13 cloning vectors mp8 and mp9 (14). All restriction sites used for cloning were sequenced on another fragment to avoid errors due to closely spaced identical restriction sites. The entire nucleotide sequence of the DNA insert of pECX14 is shown in Fig. 3. An examination of the nucleotide sequence showed that there were four open reading frames with translated molecular weights of 52,000, 54,000, 10,000, and 16,000. The initiation and termination codons of these open reading frames are indicated in Fig. 3. The sizes of these open frames agree well with the data from the in vitro transcription-translation, which showed three proteins of similar molecular weights. The enzyme assay data with the various deletion derivatives of pECX14 suggest that the first open reading frame encodes the xylose isomerase gene and the second encodes the xylulose kinase gene.

DISCUSSION

The results of the above studies show clearly that we have cloned a segment of the *E. coli* chromosome that encodes at least a part of the xylose utilization pathway. Also, these results have enabled us to identify the coding sequences for the isomerase and kinase genes. The first open reading frame (52,000 molecular weight) in the cloned segment has been identified as the xylose isomerase gene, and the second

```
GATCTTACTT TTGTTGCGCA ATTGTACTTA TTGCATTTTT CTCTTCGAGG AATTACCCAG TTTCATCATT CCATTTTATT TTGCGAGCGA GCGACACTTG  100

                                                                                                        m
TGAATTATCT CAATAGCAGT GTGAAATAAC ATAATTGAGC AACTGAAAGG GAGTGCCCAA TATTACGACA TCATCCATCA CCCGCGGCAT TACCTGATTA  200

etl
TGGAGTTCAA TATGCAAGCC TATTTTGACC AGCTCGATCG CGTTCGTTAT GAAGGCTCAA AATCCTCAAA CCCGTTAGCA TTCCGTCACT ACAATCCCGA  300

CGAACTGGTG TTGGGTAAGC GTATGGAAGA GCACTTGCGT TTTGCCGCCT GCTACTGGCA CACCTTCTGC TGGAACGGGG CGGATATGTT TGGTGTGGGG  400

GCGTTTAATC GTCCGTGGCA GCAGCCTGGT GAGGCACTGG CGTTGGCGAA GCGTAAAGCA GATGTCGCAT TTGAGTTTTT CCACAAGTTA CATGTGCCAT  500

TTTATTGCTT CCACGATGTG GATGTTTCCC CTGAGGGCGC GTCGTTAAAA GAGTACATCA ATAATTTTGC GCAAATGGTT GATGTCCTGG CAGGCAAGCA  600

AGAAGAGAGC GGCGTGAAGC TGCTGTGGGG AACGGCCAAC TGCTTTACAA ACCCTCGCTA CGGCGCGGGT GCGGCGACGA ACCCAGATCC TGAAGTCTTC  700

AGCTGGGCGG CAACGCAAGT TGTTACAGCG ATGGAAGCAA CCCATAAATT GGGCGGTGAA AACTATGTCC TGTGGGGCGG TCGTGAAGGT TACGAAACGC  800

TGTTAAATAC CGACTTGCGT CAGGAGCGTG AACAACTGGG CCGCTTTATG CAGATGGTGG TTGAGCATAA ACATAAAATC GGTTTCCAGG GCACGTTGCT  900

TATCGAACCG AAACCGCAAG AACCGACCAA ACATCAATAT GATTACGATG CCGCGACGGT CTATGGCTTC CTGAAACAGT TTGGTCTGGA AAAAGAGATT 1000

AAACTGAACA TTGAAGCTAA CCACGCGACG CTGGCAGGTC ACTCTTTCCA TCATGAAATA GCCACCGCCA TTGCGCTTGG CCTGTTCGGT TCTGTCGACG 1100

CCAACCGTGG CGATGCGCAA CTGGGCTGGG ACACCGACCA GTTCCCGAAC AGTGTGGAAG AGAATGCGCT GGTGATGTAT GAAATTCTCA AAGCAGGCGG 1200

TTTCACCACC GGTGGTCTGA ACTTCGATGC CAAAGTACGT CGTCAAAGTA CTGATAAATA TGATCTGTTT TACGGTCATA TCGGCGCGAT GGATACGATG 1300

GCACTGGCGC TGAAAATTGC AGCGCGCATG ATTGAAGATG GCGAGCTGGA TAAACGCATC GCGCAGCGTT ATTCCGGCTG GAATAGCGAA TTGGGCCAGC 1400

AAATCCTGAA AGGCCAAATG TCACTGGCAG ATTTAGCCAA ATATGCTCAG GAACATCATT TGTCTCCGGT GCATCAGAGT GGTCGCCAGG AACAACTGGA 1500

                                    end1                        >>>> >>> >>>>     <<<<<<<< <<<
AAATCTGGTA AACCATTATC TGTTCGACAA ATAACGGCTA ACTGTGCAGT CCGTTGGCCC GGTTATCGGT AGCGATACCG GGCATTTTTT TAAGGAACGA 1600

     met2
TCGATATGTA TATCGGGATA GATCTTGGCA CCTCGGGCGT AAAAGTTATT TTGCTCAACG AGCAGGGTGA GGTGGTTGCT GCGCAAACGG AAAAGCTGAC 1700

CGTTTCGCGC CCGCATCCAC TCTGGTCGGA ACAAGACCCG GAACAGTGGT GGCAGGCAAC TGATCGCGCA ATGAAAGCTC TGGGCGATCA GCATTCTCTG 1800

CAGGACGTTA AAGCATTGGG TATTGCCGGC CAGATGCACG GAGCAACCTT GCTGGATGCT CAGCAACGGG TGTTACGCCC TGCCATTTTG TGGAACGACG 1900

GGCGCTGTGC GCAAGAGTGC ACTTTGCTGG AAGCGCGAGT TCCGCAATCG CGGGTGATTA CCGGCAACCT GATGATGCCC GGATTTACTG CGCCTAAATT 2000

GCTATGGGTT CAGCGGCATG AGCCGGAGAT ATTCCGTCAA ATCGACAAAG TATTATTACC GAAAGATTAC TTGCGTCTGC GTATGACGGG GGAGTTTGCC 2100

AGCGATATGT CTGACGCAGC TGGCACCATG TGGCTGGATG TCGCAAAGCG TGACTGGAGT GACGTCATGC TGCAGGCTTG CGACTTATCT CGTGACCAGA 2200

TGCCCGCATT ATACGAAGGC AGCGAAATTA CTGGTGCTTT GTTACCTGAA GTTGCGAAAG CGTGGGGTAT GGCGACGGTG CCAGTTGTCG CAGGCGGTGG 2300

CGACAATGCA GCTGGTGCAG TTGGTGTGGG AATGGTTGAT GCTAATCAGG CAATGTTATC GCTGGGGACG TCGGGGGTCT ATTTTGCTGT CAGCGAAGGG 2400

TTCTTAAGCA AGCCAGAAAG CGCCGTACAT AGCTTTTGCC ATGCGCTACC GCAACGTTGG CATTTAATGT CTGTGATGCT GAGTGCAGCG TCGTGTCTGG 2500

ATTGGGCCGC GAAATTAACC GGCCTGAGCA ATGTCCCAGC TTTAATCGCT GCAGCTCAAC AGGCTGATGA AAGTGCCGAG CCAGTTTGGT TTCTGCCTTA 2600

TCTTTCCGGC GAGCGTACGC CACACAATAA TCCCCAGGCG AAGGGGGTTT TCTTTGGTTT GACTCATCAA CATGGCCCCA ATGAACTGGC GCGAGCAGTG 2700

CTGGAAGGCG TGGGTTATGC GCTGGCAGAT GGCATGGATG TCGTGCATGC CTGCGGTATT AAACCGCAAA GTGTTACGTT GATTGGGGGC GGGGCGCGTA 2800

GTGAGTACTG GCGTCAGATG CTGGCGGATA TCAGCGGTCA GCAGCTCGAT TACCGTACGG GGGGGGATGT GGGGCCAGCA CTGGGCGCAG CAAGGCTGGC 2900

GCAGATCGCG GCGAATCCAG AGAAATCGCT CATTGAATTG TTGCCGCAAC TACCGTTAGA ACAGTCGCAT CTACCAGATG CGCAGCGTTA TGCCGCTTAT 3000

                                                        end2
CAGCCACGAC GAGAAACGTT CCGTCGCCTC TATCAGCAAC TTCTGCCATT AATGGCGTAA ACGTTATCCC CTGCCTGACC GGGTGGGGGA TAATTCACAT 3100

CTATATATCT CAGTAATTAA TTAATATTTA GTATGAATTT ATTCTGAAAA TCATTTGTTA ATGGCATTTT TCAGTTTTGT CTTTCGTTGG TTACTCGTAA 3200

               met3
TGTATCGCTG GTAGATATGG AGATCGTTAT GAAAACCTCA AAGACTGTGG CAAAACTATT ATTTGTTGTC GGGGCGCTGG TTTATCTGGT TGGGCTATGG 3300

ATCTCATGCC CATTGTTAAG TGGAAAAGGC TATTTTCTTG GCGTGTTAAT GACAGCAACT TTTGGCAACT ATGCATATCT TCGCGCAGAA AAACTCGGGC 3400

AACTGGATGA TTTTTTTACC CATATCTGCC AGTTAGTTGC GTTAATCACT ATCGGTGTCT TGTTTATCGG TGTTTTAAAC GACCTATCAA TACTTATGAA 3500

   end3
ATGGTGATCT ATCCCATCGC CTTTTTTGTC TGCTTGTTTG GTCAAATGCG TTTGTTTCGC TCGGCATGAG CAACATAAAG CTCTTACATA TTCAGGAATG 3600

             met4
AAAGGAATAC TGTGATGGAC AACAAAATAT CAACCTATTC ACCGGCCTTT AGTATTGTGT CATGGATAGC TCTCGTTGGT GGTATCGTTA CCTATCTGTT 3700

AGGGCTATGG AATGCAGAGA TGCAGTTAAA TGAAAAAGGA TATTATTTTG CCGTACTGGT ATTAGGACTG TTTTCTGCGG CGTCTTATCA AAAGACCGTT 3800

CGGGACAAGT ATGAAGGCAT ACCGACCACT TCCATTTATT ATATGACCTG CCTGACTGTC TTTATTATCT CTGTTGCATT ACTGATGGTA GGTCTGTGGA 3900

ATGCGACATT ATTACTCAGC GAAAAAGGTT TTTATGGACT GGCTTTCTTC TTAAGCTTGT TTGGTGCAGT AGCGGTGCAG AAAAATATTC GTGATGCCGG 4000

                                                    end4
AATAAACCCA CCAAAAGAAA CACAGGTTAC CCAGGAAGAA TACAGCGAAT AACTCACGTA AGCCCGGTCA GTCCAATGTG ACCGGGCTTT TACTTAACTC 4100

ACTAATCTGT TTCTGTCGAT TCGTTGTACC AGCATAGAAA GTAACAAACT CGCTGCCAAC GTCGCGCAAA AGATC                          4175
```

FIG. 3. Sequence of the insert DNA of pECX14. The nucleotide sequence of the insert DNA of pECX14 is presented. The beginning and end of each of the four open reading frames are indicated with "met" and "end." The extent of each frame is as follows: frame 1, 200 to 1,532; frame 2, 1,606 to 3,058; frame 3, 3,217 to 3,505; frame 4, 3,615 to 4,050. The position of the inverted repeat discussed in the text is designated by >>>>> <<<<<.

(54,000 molecular weight) has been identified as the xylulose kinase gene. The rationale for these assignments is straightforward. When pECX-14 or any of the derivatives containing the first open reading (derivative A, B, C, D, E, or F) was present in strain JC1553, isomerase enzymatic activity could be detected (Table 1). Since derivatives pECX14-A and pECX14-F contain only the first open reading frame, other portions of the cloned region are not required for the isomerase activity, and since pECX14-A can complement the isomerase defect in three phenotypically different mutant strains, it is unlikely that the first reading frame could be a regulatory gene. Therefore, the first reading frame must encode the structural gene for the xylose isomerase gene. The assignment of the second open reading frame (54,000 molecular weight) in the sequence to the kinase gene is based on similar arguments. When pECX14-G, which contains only the second reading frame, is present in strain JC1553, kinase activity, but not isomerase activity, can be detected, and at levels approximately 10 times higher than in an induced wild-type strain (MM294). The higher levels of kinase activity provide one piece of supporting evidence that pECX14-G carries the structural gene for the kinase activity, since it is on a multicopy plasmid.

The identity of the third (10,000 molecular weight) and fourth (16,000 molecular weight) open reading frames in the sequence is uncertain. The fourth open reading frame definitely encodes a protein, as shown by the in vitro transcription-translation data; however, a small 10,000-molecular-weight protein corresponding to the third reading frame was not detected. Genetic evidence in *S. typhimurium* suggests that the *xyl* operon is organized in the order *xylA*(isomerase)-*xylB*(kinase)-*xylR*(regulatory)-*xylT*(transport) (21). If the gene order is the same in *E. coli*, the third and fourth proteins might be a regulatory protein and a transport protein (or two regulatory proteins). Further studies will be necessary to determine whether these genes have any function in xylose utilization in *E. coli*.

Of interest was the finding that the isomerase and kinase genes seem to be separate transcriptional units. Both the in vitro and in vivo evidence suggests that the isomerase gene appears to have its own promoter in the 200-bp region of the insert that precedes its initiation codon. When the upstream tetracycline gene promoter of pBR322 is removed (pECX14-F), there is no decrease in the levels of isomerase activity (Table 1). In vitro transcription-translation of a DNA fragment (lane 7, Fig. 2) containing the isomerase gene also shows the isomerase protein, indicating that some promoter activity resides on that fragment. Strains carrying the isomerase gene on a plasmid show an increase in isomerase activity upon xylose induction, suggesting that the natural regulatory regions are intact, although the induction ratio is much less than that seen for strain MM294 without the plasmid. The evidence for the conclusion that the xylulose kinase gene has its own promoter is that derivative pECX14-G, which has a deletion extending from the *Eco*RI site of pBR322 to the *Sal*I site in the isomerase gene (Fig. 1), shows kinase enzymatic activity in strain JC1553 (Table 1). This deletion removes the tetracycline gene promoter of pBR322 and any possible promoter preceding the isomerase gene, suggesting that the kinase gene has its own promoter somewhere in the region between the *Sal*I site and the start of the kinase gene. Also, the transcription-translation data show that, when a DNA fragment extending from that same *Sal*I site to the *Sal*I site in pBR322 (see Fig. 1) is used to program the reaction, a protein band of the correct size for the kinase gene is produced (lane 8, Fig. 2). When the DNA sequence between the end of the isomerase gene and the beginning of the kinase gene is examined, a stem and loop structure followed by a string of seven T's (shown in Fig. 3), which looks very much like a transcription termination signal, is present (17).

All of the above evidence suggests that the kinase gene is not produced from an mRNA starting at the isomerase gene. However, the phenotype of the Tn*10* insertion mutant and the *xyl-22* mutation, which lack the kinase and isomerase functions, could most simply be explained by a polarity effect or the loss of a common promoter. The status of the transcriptional unit(s) of the two genes could be resolved by using segments of the cloned genes as probes for mRNA.

One puzzling finding was that derivative pECX14-A could not complement a strain carrying the *xyl-20* mutation (lacking isomerase), as determined by lack of growth on minimal A-xylose plates and white colonies on EMB-xylose plates (Table 1). Indeed, when pECX14-A was transformed into strain MM294, the transformants formed white colonies on EMB-xylose plates (data not shown). The enzymatic activity of the isomerase seems necessary for the effect, as plasmids carrying only the amino-terminal half of the isomerase gene do not show such an effect. Our only tentative explanation is that the partially constitutive expression of the isomerase gene when present on the high-copy plasmid reduces the level of xylose present in the cell below the level needed for the induction of the other enzymes in the pathway. One practical consequence of this finding is that it might be difficult to isolate the xylose isomerase gene from other organisms by direct selection in *E. coli*, unless the xylulokinase gene is linked to the isomerase gene. This problem might be overcome by using a plasmid which contains the *E. coli* xylulokinase gene as the cloning vehicle or, alternatively, a strain which is constitutive for the expression of xylulokinase.

LITERATURE CITED

1. **Backman, K., M. Ptashne, and W. Gilbert.** 1976. Construction of plasmids carrying the cI gene of bacteriophage lambda. Proc. Natl. Acad. Sci. U.S.A. **73**:4174–4178.
2. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
3. **Bolivar, F., R. Rodriguez, P. J. Greene, M. Betlach, H. L. Heyneker, H. W. Boyer, J. Crosa, and S. Falkow.** 1977. Construction and characterization of new cloning vehicles. Gene **2**:95–113.
4. **Bradford, 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. **Clewell, D. B., and D. R. Helinski.** 1972. Effect of growth conditions on the formation of the relaxation complex of supercoiled Col E1 deoxyribonucleic acid and proteins in *Escherichia coli*. J. Bacteriol. **110**:1135–1143.
6. **David, J., and H. Weismeyer.** 1970. Control of xylose metabolism in *Escherichia coli*. Biochim. Biophys. Acta **201**:497–499.
7. **Dische, Z.** 1949. Spectrophotometric method for the determination of free pentose and pentose in nucleotides. J. Biol. Chem. **181**:379–392.
8. **Hershfield, V., H. W. Boyer, C. Yanofsky, M. A. Lovett, and D. R. Helinski.** 1974. Plasmid Col E1 as a molecular vehicle for cloning and amplification of DNA. Proc. Natl. Acad. Sci. U.S.A. **71**:3455–3459.
9. **Lane, E.** 1957. Spectrophotometric and turbidometric methods for measuring proteins. Methods Enzymol. **3**:447–458.
10. **Lawn, R. M., J. Adelman, A. E. Franke, C. M. Houck, M. Gross, R. Najarian, and D. V. Goeddel.** 1981. The sequence of human serum albumin cDNA and its expression in *E. coli*. Nucleic Acids Res. **9**:6103–6114.

11. **Legerski, R. J., J. L. Hodnett, and H. B. Gray, Jr.** 1978. Extracellular nucleases of Pseudomonas *Bal*31. III. Use of the double-stranded deoxyribonuclease activity as the basis of a convenient method for the mapping of fragments of DNA produced by cleavage with restriction enzymes. Nucleic Acids Res. **5**:1445–1464.
12. **Maleszka, R., P. Y. Wang, and H. Schneider.** 1982. A Col E1 hybrid plasmid containing *Escherichia coli* genes complementing D-xylose negative mutants of *Escherichia coli* and *Salmonella typhimurium*. Can. J. Biochem. **60**:144–151.
13. **Maloy, S. R., and W. D. Nunn.** 1981. Selection for loss of tetracycline resistance by *Escherichia coli*. J. Bacteriol. **145**:1110–1112.
14. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double digest restriction fragments. Gene **19**:269–276.
15. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **Mills, D. R., and F. R. Kramer.** 1979. Structure-independent nucleotide sequence analysis. Proc. Natl. Acad. Sci. U.S.A. **76**:2232–2235.
17. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
18. **Saito, H., and K. Miura.** 1963. Preparation of transforming deoxyribonucleic acid by phenol treatment. Biochim. Biophys. Acta **72**:619–629.
19. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
20. **Shamanna, D. K., and K. E. Sanderson.** 1979. Uptake and catabolism of D-xylose in *Salmonella typhimurium* LT2. J. Bacteriol. **139**:64–70.
21. **Shamanna, D. K., and K. E. Sanderson.** 1979. Genetics and regulation of D-xylose utilization in *Salmonella typhimurium* LT2. J. Bacteriol. **139**:71–79.
22. **Simpson, F. J.** 1966. D-Xylulokinase. Methods Enzymol. **9**:454–458.
23. **Wood, W. A.** 1966. Carbohydrate metabolism. Annu. Rev. Biochem. **35**:521–553.
24. **Woucha, M. G., D. L. Steuerwald, and K. E. Brooks.** 1983. Amplification of D-xylose and D-glucose isomerase activities in *Escherichia coli* by gene cloning. Appl. Environ. Microbiol. **45**:1402–1404.