The nucleotide sequence of *Escherichia coli* genes for L-fucose dissimilation

Zhe Lu and E.C.C.Lin

Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA 02115, USA

Submitted May 15, 1989

EMBL accession no. X15025

L-Fucose is dissimilated by *Escherichia coli* through an inducible pathway mediated by a permease encoded by *fucP*, an isomerase encoded by *fucI*, a kinase encoded by *fucK*, and an aldolase encoded by *fucA*. The aldolase cleaves fuculose 1-phosphate to dihydroxyacetone phosphate and L-lactaldehyde. Under anaerobic conditions the L-lactaldehyde is reduced to L-1,2-propanediol by an oxidoreductase encoded by *fucO*. These structural genes, together with the regulatory gene *fucR* encoding an activator protein, cluster at min 60 of the chromosomal map (1). Cloning and deletion analysis of the *fuc* region showed that the structural genes constitute two divergently transcribed operons *fucAO* (counterclockwise) and *fucPIK* (clockwise). The *fucR* gene is located downstream of *fucK* (2,3).

The entire *fuc* region (8.9 kb) has been sequenced. The sequence of the *fucO* and *fucA* genes was presented in a separate paper describing the effects of an insertion (IS5) mutation on the expression of the *fucAO* and *fucPIK* operons (3). We report here the sequence from positions 2001 to 8901, which includes the remaining *fuc* genes. This part of the sequence reveals 6 long open reading frames (ORFs), all of them have good Shine-Dalgarno (SD) ribosome binding sequences which are properly placed. The 4 ORFs assigned to the *fucPIK* operon and *fucR* are based on their positions and complementation tests (2). The functions of the 2 other ORFs, if any, are not known. One is located between *fucK* and *fucR* encoding a peptide of 134 amino acids. The other is located at the right end of the sequenced region, down stream of *fucR*, encoding at least 63 amino acids. The region between positions 2075 to 2595 contains controlling elements for the divergently transcribed *fucAO* and *fucPIK* operons (3). There are 4 or 5 potential cAMP receptor protein (CRP) binding sites scattered in this region based on homologies to the reported consensus CRP binding sequence (4). Further study is needed to identify the functional sites.

```

2001 CCAGCGGGGTCATTCCAGCGAAGTGTCAA   TAACTGACGAGCAAGTATTATTCGTTCCA   TAGCTACCTCTCTCTGATTCAAAACAGGG   CAATAATGTGTGTCCTTCACACTATTGAA
                                     fucA <-----\
2121 TTAGCGGTTAATTACCCACCACTTCTCTTC   CTGATTAAACAGAAAGAAATTCACAAGCTT   ATATTTTGTGACCTGGTTCACATAACACA   GTAATAACTGCAAGTCTCTTTTATAAAC
2241 CCCATTAAATGACCGCTCTTAAAAATAT     TTATCAAAACGGTCATTTTCTATTCTCTCC   AAGCCGGAATGACCGTTTTCGGCACAAAC   AATTAACTGCGTCATCTGATTGTGTTTTT
                                     C R P
2361 TATGATTATTTTCTGAACCGGCGATGAAG   TTTCGATTATTAAGTGATGGTAGTCACAT   AAAGTCACTTCTAGCTAATAAGTGTGACC   GCGTCATATTACAGAGCGGTTTTTATTTG
2481 AAAATGATCCATGAGTTCATTTCAGACAG   GCAAATATTCACTGATATGAAGCCGAAGT   CGCTGGTTTGTGCATTTTGAAAACATAACC   GATTACGTGCTTAAGCTCTGAACATAAGA
                                     /-----> fucP
2601 GATCTCTGAGAAACACATCAATCAAAAC     GCAGAGTTACCGTGGCGTAGATAAAGATGC   AGGCCAAAGCAGAAGTATACATATTCATT   CCGCGTCTGTGCTCACTGTTTTTCTTTG
2721 AVANLNLDIL LPOFQOAF L T N F Q A G L I Q S A F Y F G Y F I I P
2841 GCGGTAGCAATAACCTTACCAACATTTT     ATTACCTCAATTCAGCAGGCTTTACGCT     GACAAATTCAGGCTGGCGTATCAATC     GCGCTTTACTTTGGTATTTCATTATCCC
2961 I P A G I L N K K L S T X A G I I T G L F L Y A L G A A L L U A A E I H W Y T
2841 AATCCCTGCTGGGATATTGAATAAAAT     CAGTATAAAGCAGGATATTACCGGGTT     ATTTTATATGCTTGGGCTGCAATTAT     CTGGCCCGCGCAGAAATAATGAACACAC
2961 L F L V G L F I I A A G L G C L E T A A N P F V T V L G P E S S G H F R L N L A
2961 CTGTGTTTGTGCTGCTTATTATTATTGCG   AGCCGGATAGGTTGTCTGGAACTGCCGC     AAACCTTTTGTACGGTATTAGGCGCGGA   AAGTAGTGCTCACTTCCGCTTAATCTTCG
3081 Q T F N S F G A T I A V V F G Q S L I L S N V P H Q S Q D V L D K M S P E Q L S
3081 GCAAACATTAACTCGTTTGGCGCAATTT     CCGGGTGTCTTTGGCGAAGCTTATTTT     GTCTAACGTGCCACATCAATCGCAAGACGT   TCTCGATAAATGCTTCCAGAGCAATTGAG
3201 ATYHNSLV L S T X A G I I T G L F L Y A L G A A L L U A A E I H W Y T
3201 TCGGTATATACACAGCTGTGATTATCGET   ACAGACACTTATATGATCATCGTGCTAT     CGTGTACTGTGCGCTGCTGATCATGCT     GACGAAATCCCGGCTTGCAGAGTGATAA
3321 H S D A K Q G S F S A S L S R L A R I R H U R W A V L A Q F C Y V G A Q T A C W
3321 TCACGTGACGCCAAACAGGATCCTTCTC     CGCATCGCTTCTCGCTGGCGGCTATTCG     CCACCTGGCGCTGGCGGCTATTAGCGCAATT   CTGCTATGTCGGCGCAAAACGGCTGCTG
3441 S Y L I R Y A V E E I P G H T A G F A A N Y L T G T M V C F F I G R F T G T W L
3441 GAGCTATTTTATCTACCTACCTGTABAAEA   AATTCAGGTATGACTCGAGGCTTTGCCGC     TAACATTTAACCGGAACCATGCTGTCTT     GTTATTTGCTGTTTACCGGCTACCTGGCT
3561 S P A P S V L A A T A L I A N A L V L L V A F A G S H V L L A L T L S C S
3561 CATCATCTTCTGACGACGACCAAAAGTCT     CGCGCTACCGCATTAATCGTATGCACT     GTGCCGTGATCTCAGCTTCCGTCGCGTCA     TGTGGCTTAATAGCGCTGACTTATGCAG
3681 A F R I I Q Y P T I P S L G I K N L G Q D T K Y G S S F I V N I I G G G I V T
3681 CGCCTTTATGTGATTCAGTACCCAAATAT   CTTCGCTGGGATTAAGAATCTCGGCCA     GGACACCAAAATAGTGTGCTCTTCATGCT   TATGACCATTTATGGCGCGGATTGTGCAC
3801 P V M G P V S D A A G N I P T A E L I P A L C F A V I F I A R F R S Q T A T H
3801 TCCGTCATGCTTTTGTGCTGAGTACGCGGC   GGGCAACATCCCACTGCTGAATGATCCC     CGCACTGCTGCTCGCGGTCACTTTATCTT   TGCCGTTTCCGTTCTCAACAGCGCAACTAA
                                     fucP-->\\
                                     fucI
3921 CTGAACATATTTCCGAATAAGTAGGAA     TCTGTATGAAAAAATCAGCTTACCGGAAA   ATTGGTATCCGCGGTTATTGACGCTGCT   CGCATGTGCTTCCGTGAGTGGCTTGAAGA
4041 Q T M N H A A A A L L T E K L R H A C G A A V E C V I S D T C I A G M A E A
4041 GAAACAATGAATGGCGAAAGCTACGGCC   GCACCTGCTGACGCAAGAACTGCGGCATGCC   TCGGAGCTGGCGTCAAGTGTGATCTTTC   GATACCTGTATCGCGGATGGCTGAAGCC

```

[illegible]

ACKNOWLEDGMENTS: We thank Y.-M. Chen for helpful discussions. This study was supported by Public Health Service grants 1-R01-GM39693, 5-R01-GM11983 and 1-R01-GM40993.

- REFERENCES:** 1. Lin, E.C.C. (1987) In Neidhardt, F.C. (ed. in chief), *Escherichia coli* and *Salmonella typhimurium*-Cellular and Molecular Biology, Am. Soc. Microbiol., Washington, Vol.1, pp. 244-284.
2. Chen, Y.-M., Zhu, Y., and Lin, E.C.C. (1987) Mol. Gen. Genet., 210, 331-337.
3. Chen, Y.-M., Lu, Z., and Lin, E.C.C. (1989) submitted to J. Bact.
4. Crombrughe, B. de, Busby, S., and Buc, H. (1984) Science 224, 831-838.