Volume 14 Number 17 1986 Nucleic Acids Research

# Nucleotide sequence of the gene responsible for D-xylose uptake in *Escherichia coli*

Naotaka Kurose[1], Kunihiko Watanabe and Akira Kimura

Research Institute for Food Science, Kyoto University, Uji, Kyoto 611 and [1]Central Research Laboratories, Takara Shuzo Co., Ltd., Otsu, Shiga 520-21, Japan

Received 19 May 1986; Revised and Accepted 6 August 1986

ABSTRACT

The nucleotide sequence of the cloned DNA, 363 bp in length, has been determined. It can complement the mutation of Escherichia coli having a decreased activity of D-xylose uptake at low temperature. Nucleotide sequence analysis found one possible reading frame coding for a polypeptide consisting of 61 amino acids. Several signal sequences conserved in the promoter regions of E. coli were found in the upstream regions of the open frame. This included the Shine-Dalgarno sequence, the Pribnow box, and the sequence conserved in the "-35 region" with a preferable spacing from each other for an efficient transcription. Downstream from the termination codon, the inverted repeat sequence was present, followed by 3 successive T's.

## INTRODUCTION

The metabolism of D-xylose in Escherichia coli is initiated by the active transport of this pentose into the cells (1-7). D-Xylose once introduced into the cells is further metabolized by sequential enzyme reactions catalyzed by D-xylose isomerase, D-xylulose kinase and the enzymes in the pentose phosphate pathway. On the transport of D-xylose, at least two systems with varying affinities have been considered to be functioning in E. coli K-12 (8). One is a high affinity system involving a periplasmic D-xylose-binding protein (XBP), and the other is a low affinity system facilitated by utilization of the proton-motive force (2,7). However, the detailed mechanisms of these D-xylose transport systems have not yet been reported. There are also no reports describing the cloning of the gene or genes responsible for D-xylose transport, nor the complete nucleotide sequence of the gene(s), although the D-xylose isomerase and D-xylulose kinase structual genes of E. coli have

been isolated and the complete nucleotide sequence of these genes have been determined (9-11).

Recently, we have isolated an E. coli mutant having characteristic properties on D-xylose uptake. The D-xylose uptake activity of the mutant grown on the minimal medium with D-xylose at 25 C was only one-eighth of that of the parental strain grown under the same conditions, although the XBP activities were almost the same for the two strains. An uptake study on sugars at the low temperature (25 °C) indicated that the mutant was deficient in D-xylose uptake activity. Recombinant plasmids, pXT1 and pXT2 derived from pXT1, which enable the mutant to recover from the decreased activity of D-xylose uptake at low temperature, have been described in our previous paper (12). In our present work, we have determined the nucleotide sequence of the DNA which was cloned into pBR322 in our previous work, and this DNA certainly include the gene responsible for D-xylose uptake of E. coli.

## MATERIALS AND METHODS

### Bacterial strains and plasmids used

The strains used were E. coli K-12 strain C600 ($F^-$ hsdR hsdM recA thr leu thi lacY supE tonA) and M3-C (a mutant derived from C600) having an altered D-xylose uptake activity.

The plasmids used were pXT1 (6.1 kb) and pXT2 (4.8 kb) derived from pXT1, which includes the gene capable of complementing the D-xylose uptake mutation of mutant M3-C.

For base-sequencing of the cloned DNA, E. coli strain JM103 and phage vectors M13 mp10 and M13 mp11 were used.

### DNA sequencing procedure

The hybrid plasmid pXT1 was digested with Pst I restriction endonuclease, and the resulting chromosomal DNA fragment, approximately 1.8 kb in length, was separated from the vector pBR322 by electrophoresis in a 1.0% low-melting-temperature agarose gel. The DNA fragment removed from the gel was purified, redigested with Pvu II or the combination of Bam HI and Pvu II restriction endonucleases, and then inserted into the corresponding cloning sites of M13 mp10 and M13 mp11 phage vectors. The resulting recombinant phage DNA was sequenced by

the dideoxy sequencing method of Sangar et al. (13).

The annealing reactions of the primer to the template, the various nucleotide polymerization reactions, and acrylamide gel electrophoresis were carried out by the method described in the text of Takara Shuzo Co., Ltd., Japan.

DNA sequence analysis were performed with DNASIS (Hitachi SK Co., Japan).

Hydrophilicity of a polypeptide

Hydrophilicity values of amino acids constructed of a polypeptide were calculated with DNASIS.

Enzymes and biochemicals

Restriction endonucleases, T4 DNA ligase and M13 sequencing kit were all obtained from Takara Shuzo Co., Ltd., Kyoto, Japan. [$\alpha$-$^{32}$P]dCTP was purchased from Amersham Corp. All other chemicals were reagent grade obtained from commercial sources.

## RESULTS

### Subcloning of the gene

Recombinant plasmid pXT1 was about 6.1 kb in length, and contained the about 1.8 kb chromosomal DNA fragment involving the gene responsible for D-xylose uptake of E. coli inserted at the Pst I site of vector plasmid pBR322 (12). To obtain a shorter DNA fragment containing the active region, the cloned DNA (about 1.8 kb) obtained from pXT1 was digested with Sau 3AI restriction endonuclease and recloned into the Bam HI site of pBR322. The recombinant plasmids obtained were used for transformation of E. coli M3-C, a mutant having a drasticly decreased D-xylose uptake activity at a temperature below 37 C compared to that of the wild type. The resulting transformants were selected as well-grown colonies at 25 C on Davis-Mingioli minimal agar plates containing 0.2% of D-xylose and ampicillin (4 µg/ml). Seven transformants which were able to recover from the M3-C's characteristic poor growth rate on D-xylose were shown to harbour various plasmids smaller than pXT1. Of the 7 transformants, the smallest plasmid (designated as pXT2, 4.8 kb) included the active region of about 360 bp in length. Restriction mapping of pXT2 has indicated that the active region of the subcloned DNA includes the Bam HI restriction site and is

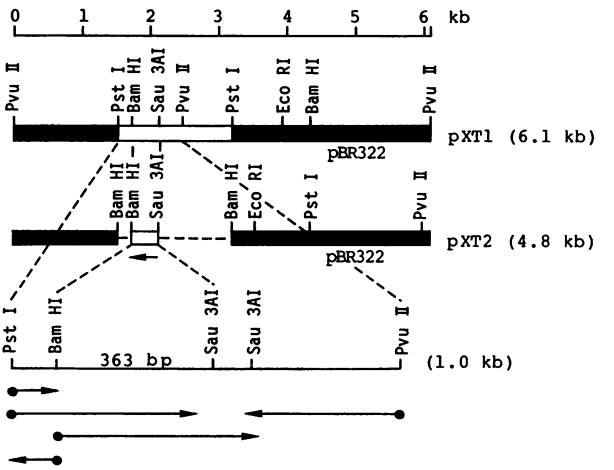


Figure 1. Physical map of the recombinant plasmids pXT1 and PXT2 (Kurose et al., 1985) and the strategy for the DNA sequencing of a gene responsible for D-xylose uptake. The restriction sites of pXT1 DNA used for sequencing are indicated at the map coordinated on the top column. At the bottom column in the expanding scale, the extent and direction of DNA sequencing are indicated by horizontal arrows. The cloned circles represent the 5' termini of the cloned fragments generated by restriction enzyme digestion.

located within a Pst I-Pvu II fragment (approximately 1.0 kb) which was contained in the cloned 1.8 kb DNA fragment of pXT1 (Fig. 1).

Nucleotide sequence of the cloned gene

To determine the nucleotide sequence of the gene containing the active region of pXT1, we first digested pXT1 with Pst I restriction endonuclease to isolated the 1.8 kb DNA fragment. Furthermore, the 1.8 kb fragment was digested with Pvu II or the combination of Bam HI and Pvu II, and then the Pst I-Pvu II fragment generated was inserted into the corresponding cloning sites of the M13 mp10 and M13 mp11 phage vectors which were treated with the combination of Pst I and Sma I. In the same manner, the Pst I-Bam HI fragment generated was inserted into M13 phage vectors treated with Pst I and Bam HI; the Bam HI-Pvu II fragment generated inserted into M13 phage vectors treated with Bam HI and Sma I. After transfection of E. coli strain JM103, the recombinant phages were propagated and the resulting

```
  1                                                                  5'  G

  2 ATCGCCCACTTCCTTCAGTTCATGGGCTGGTTTGTTCCGCAGCTCCACCAGCGCCTGGC
                                                -35 SQ
 61 AATATTTATTACTCATTAAGCCCCCACGTAATTCCCTGAGAGATACCACTCTTCACCTG
    Pribnow          starting of mRNA            SD
120 ATG CAG CCC GCT TAC TGC TTT TCC GTA AAC ACC GTT CAC GAC GCG
  1 Met Gln Pro Ala Tyr Cys Phe Ser Val Asn Thr Val His Asp Ala

165 CCA GAA AAT TGT TTC GTT CTG GCT GGG AGT GGC TTT CAC GGA ATG
 16 Pro Glu Asn Cys Phe Val Leu Ala Gly Ser Gly Phe His Gly Met

210 CCG CCA TCC ACA CCG TTG CAG CAC GAC GGT ATA AGC CCC TGG ACT
 31 Pro Pro Ser Thr Pro Leu Gln His Asp Gly Ile Ser Pro Trp Thr

255 CCA GTT CTT CCG CCT GGC GGG TCA GGC ACA AAA TCA CCC GCG GGT
 46 Pro Val Leu Pro Pro Gly Gly Ser Gly Thr Lys Ser Pro Ala Gly

300 CGT TAG TGCCGACATAGAAATTGCGCACAGGTCTGGTTTCACGAACTGGTTGTAATT
 61 Arg Ter      IR                  IR    termination of mRNA

357 CCGGATC  3'
```

Figure 2. Nucleotide sequence of D-xylose uptake gene and the flanking regulatory unit sequence. The nucleotide sequence of the gene is indicated and nucleotides are numberd from the Sau 3AI site. The amino acid sequence of the expected polypeptide for D-xylose uptake predicted from the sequences is given below the sequence. Several regulatory sequences flanking the gene are indicated with underlines. They include the Shine-Dalgarno (SD) seqences, the Pribnow box and the conserved sequence (-35 SQ) located at -10 and -35 nucleotides respectively, upstream from the starting point of mRNA synthesis which was tentatively assined to position 74 as judged from the topology of the regulatory signal sequence described above. Downstream from the presumed termination codon of the gene at positions 303-305, the inverted repeat sequences (IR) are presented at positions 308-316 and 327-335, followed by 3 succesive T's which is a preferable site for the termination of mRNA synthesis.

phage DNA was subjected to base sequencing (Fig. 1). The nucleotide sequence of the DNA fragment, about 360 bp long containing the active region responsible for D-xylose uptake is shown in Fig. 2. Examination of the nucleotide sequence of the subcloned DNA onto pXT2 shows one possible open reading frame, possibly coding for a polypeptide. The region starting from position 120 and ending at position 305 is capable of coding for a polypeptide containing 61 amino acids. The expected polypeptide is shown in Fig. 2.

Transcriptional signals

Prokaryotic consensus sequences for transcriptional

Table 1. Codon usage in the gene for D-xylose uptake

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Phe | UUU | 2 | Ser | UCU | 0 | Tyr | UAU | 0 | Cys | UGU | 1 |
| | UUC | 1 | | UCC | 2 | | UAC | 1 | | UGC | 1 |
| Leu | UUA | 0 | | UCA | 2 | End | UAA | 0 | End | UGA | 0 |
| | UUG | 1 | | UCG | 0 | End | UAG | 1 | Trp | UGG | 1 |
| Leu | CUU | 1 | Pro | CCU | 1 | His | CAU | 0 | Arg | CGU | 1 |
| | CUC | 0 | | CCC | 3 | | CAC | 3 | | CGC | 0 |
| | CUA | 0 | | CCA | 3 | Gln | CAA | 0 | | CGA | 0 |
| | CUG | 1 | | CCG | 3 | | CAG | 2 | | CGG | 0 |
| Ile | AUU | 0 | Thr | ACU | 1 | Asn | AAU | 1 | Ser | AGU | 1 |
| | AUC | 0 | | ACC | 1 | | AAC | 1 | | AGC | 1 |
| | AUA | 1 | | ACA | 2 | Lys | AAA | 1 | Arg | AGA | 0 |
| Met | AUG | 2 | | ACG | 0 | | AAG | 0 | | AGG | 0 |
| Val | GUU | 3 | Ala | GCU | 2 | Asp | GAU | 0 | Gly | GGU | 2 |
| | GUC | 0 | | GCC | 0 | | GAC | 2 | | GGC | 3 |
| | GUA | 1 | | GCA | 0 | Glu | GAA | 1 | | GGA | 1 |
| | GUG | 0 | | GCG | 2 | | GAG | 0 | | GGG | 2 |

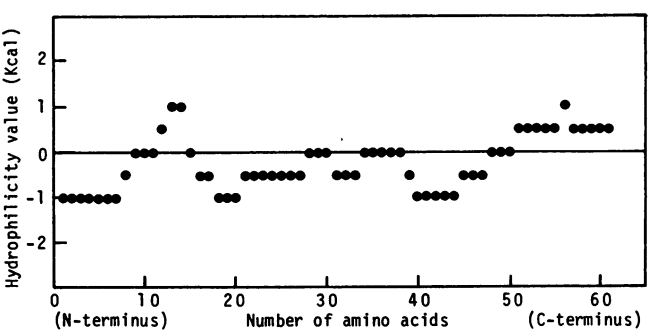


Figure 3. Hydrophilicity values of amino acids constructed of the expected polypeptide for D-xylose uptake. Closed circle presents one amino acid residue. The N-terminal region of this polypeptide was rich in unhydrophilic amino acids.

initiation have well been documented (14-16). In the precise DNA sequence shown in Fig. 2, we find two hexanucleotides, TCCGCA and AATATT at positions 37 and 61, respectively, preceding the initiation codon ATG at position 120. The former sequence matches in three out of six positions to the consensus sequence TTGACA conserved in the "-35 region" upstream from the initiation site (14) and is identical to that reported for the purf gene (17). The latter sequence also agrees in four out six

nucleotides with the Pribnow box TATAAT (14). Furthermore, these consensus sequences are separated from each other by 18 bp, presumably being the most preferrable spacing for a efficient transcription (15). Regarding the transcriptional termination signal, further extensive studies are still necessary for it to be precisely defined. The inverted repeat sequences can be located at position 308 and 307, which can be shown to form a hairpin loop structure. These inverted repeats are immediately followed by a T-rich sequence. These sequence arrangements are often found in the prokaryotic terminal region of mRNA (14).

Analysis of an expected polypeptide

As shown in Fig. 2, an expected polypeptide was constructed of 61 amino acid residues, and was estimated to have about 6,000 daltons in molecular weight. Codon usage for this peptide derived from DNA sequence data is shown in Table 1. The N-terminus of this polypeptide is found to be rich in unhydrophilic amino acids (Fig. 3).

## DISCUSSIONS

In our previous study, the D-xylose uptake activities of M3-C were investigated at 25 and 30 °C and compared with those of the parental strain (C600). As for the uptake of other sugars at 25 C by M3-C, among the compounds tested, only the D-xylose uptake activity was found to be lower than that of the parental strain. On the other hand, the activity of XBP of M3-C was slightly higher than that of the parental strain. From the results presented above, the mutant M3-C is considered to be deficient in the factor involving D-xylose uptake other than XBP. Furthermore, we cloned a gene able to complement the mutation of M3-C. However, the protein product(s) encoded from the cloned gene have not been characterized yet. Therefore, to obtain the information for the product(s), we have established the nucleotide sequence of the cloned gene in a smaller plasmid pXT2. As a result, it was found to include four open reading frames containing two shorter open frames, two in a normal and the other two in a reverse sequence (data not shown). One open reading frame was expected in only one of the reverted sequence

DNA with the transcriptional signals upstream from the initiation codon ATG (Fig. 2).

The DNA sequence upstream from the initiation codon is slightly A-T rich, where the promoter is expected to be located. The sequence of the putative "-35 region" and the Pribnow box in the upstream regions were both very similar to the known consensus nucleotide sequence (14), and the spacing of 18 nucleotides was in good accordance with the typical spacing of 17 nucleotides (14). However, further biochemical and genetic experiments will be required to confirm our tentative promoter sequence as well as the transcriptional start point. Furthermore, about 20 nucleotides upstream from initiation codon, GAGA, a Shine-Dalgarno like sequence, was located (Fig. 2).

About 3 bp downstream from the translational termination codon TAG, GC-rich inverted repeat sequences were present, followed by a stretch of succesive T's. They presumably constitute the transcriptional termination signal.

The molecular weight of the expected polypeptide was calculated to be 6,079 daltons from the 61 amino acid residues encoded by the possible open frame. The construction of this polypeptide shows that unhydrophilic amino acid residues are rich in the N-terminus. This suggests that the polypeptide may be a membrane protein containing the signal peptide in its N-terminus. Moreover, assuming that this peptide is responsible for D-xylose uptake, it is in good agreement that the expected polypeptide is a protein localized in the membrane.

Nothing is known of the factor involving D-xylose uptake other than XBP, nor the relationship between the factor and the proton-linked system for the transport of D-xylose as mentioned in INTRODUCTION. It is considered that the polypeptide constructed of the 61 amino acid residues can play a important role especially at low temperature in XBP-mediated and/or proton-linked transport system. To elucidate the D-xylose uptake mechanism further, the identification of the product of pXT1 and/or pXT2 and study on the intrinsic function of the product are now in progress.

ACKNOWLEDGMENTS

We wish to thank Dr. K. Murata, and Dr. M. Sakaguchi, Research Institute for Food Science, Kyoto University and Dr. M. Uchida, Dr. S. Hanai, and Dr. A. Obayashi, Central Research Laboratories, Takara Shuzo Co., Ltd., for their valuable suggestions. We also thank Dr. M. Takanami, Professor at Institute of Chemical Research, Kyoto University, for his technical advices.

REFERENCES

1. David, J. D. and Wiesmeyer, H. (1970) Biochim. Biophys. Acta 201, 497-499.
2. Lam, V. M. S., Daruwalla, K. R., Henderson, P. J. F. and Jones-Mortimer, M. C. (1980) J. Bacteriol. 143, 396-402.
3. Shamanna, D. K. and Sanderson, K. E. (1979) J. Bacteriol. 139, 64-70.
4. Shamanna, D. K. and Sanderson, K. E. (1979) J. Bacteriol. 139, 71-79.
5. Stevens, F. J. and Wu, T. T. (1976) J. Gen. Microbiol. 97, 257-265.
6. Wu, T. T. (1976) Biochim. Biophys. Acta 428, 656-663.
7. Davis, E. O., Jones-Mortimer, M. C. and Henderson, P. J. F. (1984) J. Biol. Chem. 259, 1520-1525.
8. Ahlem, C., Huisman, W., Neslund, G. and Dahms, A. S. (1982) J. Biol. Chem. 257, 2926-2931.
9. Lawlis, V. B., Dennis, M. S., Chen, E. Y., Smith, D. H. and Henner, D. J. (1984) Appl. Environ. Microbiol. 47, 15-21.
10. Schellenberg, G. D., Sarthy, A., Larson, A. E., Backer, M. P., Crabb, J. W., Lidstrom, M., Hall, B. D. and Furlong, C. E. (1984) J. Biol. Chem. 259, 6826-6832.
11. Ueng, P. P., Volpp, K. J., Tucker, J. V., Gong, C. S. and Chen, L. F. (1985) Biotechnol. Lett. 7, 153-158.
12. Kurose, N., Murata, K. and Kimura, A. (1985) Agr. Biol. Chem. 49, 2597-2603.
13. Sangar, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA. 74, 5463-5467.
14. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353.
15. Hawley, D. K. and McClure, W. R. (1983) Nucl. Acids Res. 11, 2237-2255.
16. Siebenlist, U., Simpson, R. B. and Gilbert, W. (1980) Cell 20, 269-281.
17. Tso, J. Y., Zalkin, H., Cleemput, M. van, Yanofsky, C. and Smith, J. M. (1982) J. Biol. Chem. 257, 3525-2531.