Volume 17 Number 2 1989 Nucleic Acids Research

Nucleotide sequence of the *Escherichia coli* porin thermoregulatory gene *envY*

M.D.Lundrigan, M.J.Friedrich and R.J.Kadner

Department of Microbiology, School of Medicine and Molecular Biology Institute, University of Virginia,
Charlottesville, VA 22908, USA

Submitted November 28, 1988 Accession no. X13548

The *envY* gene influences temperature-dependent expression of several *E. coli* envelope proteins, most notably the porins, OmpF and OmpC, and the lambda receptor, LamB (1). *envY* has been mapped, cloned, and its product identified (1). Sequencing of this gene was performed by the Sanger chain termination method (2) on a 945 bp *Eco*RI-*Pst*I fragment derived from the *envY* complementing plasmid pML22 and cloned into M13. The sequence revealed an open reading frame which codes for a polypeptide of 253 amino acids (MW 28,642). A large number of cysteine residues were indicated consistent with previous studies showing a requirement for reducing agents in denaturing the EnvY protein. Analysis of the amino acid sequence suggests that EnvY is a soluble protein and that codon usage is like that of other weakly expressed *E. coli* proteins. A region of dyad symmetry was observed in the 5′ non-coding region of this gene.

```
        10        20        30        40        50        60        70        80        90       100       110       120
GAATTCAAGACTTAATTGCTAAAAAGCGTGCTAAGAAAGAAGATCGTTATCGGCGAACGGGCCGATCACTCATGCAGGAAGCATAGATCCTTTGAGCGGAGTTTCTGATATCTGAGGAGTGC

       130          140           150          160          170           180          190          200           210
GAA ATG CAA TTG AGC AGC AGT GAA CCT TGC GTG GTG ATC CTG ACC GAA AAA GAG GTA GAG GTA AGC GTC AAT AAC CAT GCT ACG TTT ACC CTT
     M   Q   L   S   S   S   E   P   C   V   V   I   L   T   E   K   E   V   E   V   S   V   N   N   H   A   T   F   T   L

   220          230           240          250          260           270          280          290           300
CCG AAA AAC TAC CTG GCC GCC TTC GCG TGC AAC AAT AAC GTC ATT GAA CTC TCA ACG TTA AAT CAC GTA TTA ATC ACC CAC ATC AAC CGT AAC
 P   K   N   Y   L   A   A   F   A   C   N   N   N   V   I   E   L   S   T   L   N   H   V   L   I   T   H   I   N   R   N

310         320           330          340          350           360          370          380           390          400
CGT ATC ATC AAC GAT TAT CTG TTG TTT TTA AAT AAG AAC TTA ACC TGT GTA AAG CCC TGG TCG CGG CTG GCA ACC CCG GTT ATC GCT TGT CAT
 R   I   I   N   D   Y   L   L   F   L   N   K   N   L   T   C   V   K   P   W   S   R   L   A   T   P   V   I   A   C   H

        410           420          430          440           450          460          470           480          490
AGC ACA CCG GAA GTG TTC CCG CTA GCC GCC AAC CAC AGC AAG CAG CAA CCC AGC AGA CCC TGC GAA GCG GAG TTG ACG CGC GCA TTG CTT TTT
 S   T   P   E   V   F   P   L   A   A   N   H   S   K   Q   Q   P   S   R   P   C   E   A   E   L   T   R   A   L   L   F

    500           510          520          530           540          550          560           570          580
ACC GTA TTG TCT AAC TTT CTT GAG CAA TCG CGG TTT ATT GCC CTA CTG ATG TAT ATC TTA CGC AGC AGC GTC CGC GAC ACG GTC TGC CGC ATT
 T   V   L   S   N   F   L   E   Q   S   R   F   I   A   L   L   M   Y   I   L   R   S   S   V   R   D   T   V   C   R   I

590           600          610          620           630          640          650           660          670          680
ATT CAA AGC GAT ATT CAG CAT TAC TGG AAT CTG CGA ATT GTC GCC AGT TCG CTA TGT TTA AGC CCC AGC CTG CTC AAA AAG AAA TTA AAA AAC
 I   Q   S   D   I   Q   H   Y   W   N   L   R   I   V   A   S   S   L   C   L   S   P   S   L   L   K   K   K   L   K   N

          690          700          710           720          730          740           750          760          770
GAA AAT ACC AGC TAT AGC CAG ATT GTC ACA GAG TGT CGT ATG CGT TAC GCC GTA CAG ATG TTA TTG ATG GAT AAC AAA AAT ATC ACT CAG GTG
 E   N   T   S   Y   S   Q   I   V   T   E   C   R   M   R   Y   A   V   Q   M   L   L   M   D   N   K   N   I   T   Q   V

      780          790          800           810          820          830           840          850          860
GCG CAA TTA TGT GGC TAT AGC AGC ACG TCG TAC TTT ATC TCT GTT TTT AAG GCG TTT TAC GGC CTG ACA CCG TTG AAT TAT CTC GCC AAA CAG
 A   Q   L   C   G   Y   S   S   T   S   Y   F   I   S   V   F   K   A   F   Y   G   L   T   P   L   N   Y   L   A   K   Q

  870          880          890       900       910       920       930       940
CGA CAA AAA GTG ATG TGG TGA AGGGCAAAGCGGAAACGGATAAGACGGGCATAAATGAGGAAGAAATGCGCCCCTGCAG
 R   Q   K   V   M   W   *
```

Acknowledgements: This work was supported by NIH Grant GM 19078.

References: 1) Lundrigan, M.D. and Earhart, C.F. (1984) J. Bacteriology 157, 262-268. 2) Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467.