Volume 15 Number 20 1987 Nucleic Acids Research

Sequence of the complete P protein gene and part of the M protein gene from the histidine transport operon of *Escherichia coli* compared to that of *Salmonella typhimurium*

Robert Kraft[1] and Leslie A.Leinwand[1,2]

Departments of [1]Genetics and [2]Microbiology and Immunology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA

Submitted August 12, 1987 Accession no.Y00455

An EcoRI fragment of the histidine permease operon from E. coli strain Y1090 was cloned into pGEM1 and sequenced by the dideoxy method (Sanger, F. et al., 1977, PNAS 74, 5463-5467). It is presented below with the potential polypeptides encoded and the identical region from S. typhimurium (Higgins, C.F. et al., 1982, Nature 298, 723-727).

```
     M Protein ->     10        20        30        40        50        60        70        80        90       100       110
                 glupheLeuAsnAlaPhePheArgSerGlyLeuAsnCysThrValLeuAlaLeuThrLeuAsnThrCysAlaGlyAlaIleArgSerVal
E. coli          GAATTCCTTAACGCTTTTCTCCGCAGTGGCCTGAACTGTACCGTGCTGGCGCTGACGCTTAACACCTGCGCTTACACTACCGAGATTTTTGCTGGGGCAATCCGTTCGGTT
                 ::  : :::::::: :: :::::::: :: :: :::::::::: :::::::::: :: :: :::::::: :::: :: :: :::: :: :::: :: :::: :: :::::
S. typhimurium   GATCTTCTTAACGCGTTTTTCCGCAGCGGACTAAACTGTACCGTACTGGCGCTGACGCTGAATACCTGCGCGTACACCACGGAAATTTTCGCCGGGGCGATTCGCTCGGTA
                 aspleuLeuAsnAlaPhePheArgSerGlyLeuAsnCysThrValLeuAlaLeuThrLeuAsnThrCysAlaTyrThrThrGluIlePheAlaGlyAlaIleArgSerVal

     120       130       140       150       160       170       180       190       200       210       220       230
ProHisGlyGluIleGluAlaAlaArgAlaTyrGlyPheSerthrPheLysMetTyrArgCysIleIleLeuProSerAlaLeuArgIleAlaLeuProAlaTyrSerAsnGluValIleLeuMet
CCGCATGGGGAAATTGAAGCCGCCAGAGCCTATGGCTTCTCGACTTTTAAAATGTATCGCTGCATTATTTTGCCTTCTGCGCTGCATTGCGTTACCGGCATACAGCAACGAAGTGATCCTGATG
::::: :: :::::::::: ::  : :::::::::::::: : :::::::::::::::::::::::::::::: :: ::  : :: :: :: :: :: :: :: :::::::::::::::::  : :::
CCGCACGGTGAAATTGAAGCGGCACGGGCCTATGGCTTCTCGTCATTTAAAATGTATCGCTGCATTATTTTGCCGTCGGCATTACGCATCGCCTTGCCTGCCTACAGCAACGAAGTGATTTTAATG
ProHisGlyGluIleGluAlaAlaArgAlaTyrGlyPheSerserPheLysMetTyrArgCysIleIleLeuProSerAlaLeuArgIleAlaLeuProAlaTyrSerAsnGluValIleLeuMet

240       250       260       270       280       290       300       310       320       330       340       350       360
LeuHisSerThrAlaLeuAlaPheThrAlaThrValProAspLeuLeuLysIleAlaArgAspIleAsnalaAlaThrTyrGlnProPheThrAlaPheGlyIleAlaAlaValLeuTyrLeuile
CTGCACTCTACTGCGTTGGCATTTACTGCCACGGTGCCGGATCTGCTGAAAATAGCCCGCGATATTAACGCCGCCACGTATCAACCTTTTACCGCCTTCGGCATTGCCGCCGGTGCTCTATTTAATC
:::::::::: :: :::::: :::::: :: :::::::::: :: :: :: :: :: :: :: :::::::::: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
CTGCACTCTACCGCCGCTGGCGTTACCGCGACGGTGCCGGATTTGTTGAAAATCGCCCGCGATATTAACTCAGCGACTTACCAGCCGTTTGGCATTGCCGCCGGTGTTGTATTTACTC
LeuHisSerThrAlaLeuAlaPheThrAlaThrValProAspLeuLeuLysIleAlaArgAspIleAsnsnsrAlaThrTyrGlnProPheThrAlaPheGlyIleAlaAlaValLeuTyrLeuleu

    370       380       390       400       410       420       430       440       P Protein ->      470       480
IleSerTyrValLeuIleSerLeuPheArgArgAlaGlulysArgTrpLeuGlnHisVallysproSerserthrhisSTOP      METSerGluAsnLysLeuasnValIleAspLeu
ATCTCTTATGTCCTGATCAGCCTCTTTCGCAGAGCGGAAAAACGCTGGTTGCAGCATGTGAAACCTTCTTCAACGCACTGAGAATACGATGTCCGAGAATAAATTAAACGTTATCGATTTG
:: :: :::::::::: ::::: :: ::  : :::::: :::::::::::::: :::   : ::: :  :  :  :: : :::::: :: :::::::: :::::::::::::::::::::::
ATTTCGTATGTCCTGATAAGCCTGTTCCGTCGGGCGGAAAGACGCTGGTTGCAGCACGTG---TCCTCTAAATAATTCGAGTAACATGATGTCAGAAAATAAATTACACGTTATCGATTTG
IleSerTyrValLeuIleSerLeuPheArgArgAlaGluargArgTrpLeuGlnHisVal---serSerlysSTOP          METMETSerGluAsnLysLeuhisValIleAspLeu

   490       500       510       520       530       540       550       560       570       580       590       600
HisLysArgTyrGlygluHisGluValLeuLysGlyValSerLeuGlnAlaasnAlaGlyAspValIleSerIleIleGlySerSerGlySerGlyLysSerThrPheLeuArgCysIle
CACAAACGCTACGGCGAACATGAAGTGCTGAAAGGGGTATCACTGCAAGCGAATGCCGGAGATGTAATAAGCATCATCGGATCGTCGGGATCGGGGAAAAGTACCTTTCTGCGCTGCATT
:::::::::::::::::::: :::::::::::::::::::: ::::: ::  :::::::::::: :: :::::::::: :: :::::::::: :: :: :: :: :: :: :: :: :: :: :: :
CACAAACGCTACGGCGGTCATGAAGTGCTGAAAGGGGTATCGCTGCAGGCCCGCGCCGGAGATGTGATTAGCATCATCGGCTCGTCCGGCTCCGGTAAAAGCACTTTTTTGCGCTGTATT
HisLysArgTyrGlyglyHisGluValLeuLysGlyValSerLeuGlnAlaargAlaGlyAspValIleSerIleIleGlySerSerGlySerGlyLysSerThrPheLeuArgCysIle

   610       620       630       640       650       660       670       680       690       700       710       720
AsnPheLeuGluLysProSerGluGlyserIlevalValAsnGlyGlnthrIleAsnLeuValArgAspLysAspGlyGlnLeuLysValAlaAspLysAsnGlnLeuArgLeuLeuArg
AACTTCCTCGAAAAACCGAGTGAAGGGTCGATCGTGGTCAATGGCCAGACGATCAATCTGGTGCGCGACAAAGACGGTCAACTCAAAGTCGCCGATAAAAATCAACTGCGCTCTACTGCGC
:::::::::::::::::::: :::::  ::::  : :: :: :: :: :::  :: :::::::::::::::::: :: :: :::::::: :: :::::::::: :: :::  :::::  :::
AACTTCCTCGAAAAACCGAGCGAAGGCGCGATTATCGTGAACGGTCAGAACATTAATCTGGTGCGCGACAAAGATGGGCAGCTCAAAGTGGCGGATAAAAATCAGCTACGCTTGTTGCGT
AsnPheLeuGluLysProSerGluGlyalaIleileValAsnGlyGlnasnIleAsnLeuValArgAspLysAspGlyGlnLeuLysValAlaAspLysAsnGlnLeuArgLeuLeuArg

   730       740       750       760       770       780       790       800       810       820       830       840       850
ThrArgLeuThrMetValPheGlnHisPheAsnLeuTrpSerHisMetThrValLeuGluAsnValMetGluAlaProIleGlnValLeuGlyLeuSerLysglngluAlaArgGluArgAlaval
ACACGCCTGACGATGGTATTCCAGCACTTCAATCTCTGGAGCCATATGACGGTGCTGGAAAACGTCATGGAAGCGCCGATTCAGGTGTTGGGCCTGAGCAAGCAGGAAGCGCGCGAGCGGGCGGTG
:: :::::::::::::: :: :::::::::::: :::::::::::: :::::::::::::::::: :: :::::::::::::::::::::  ::::  : :::::::: :: :::::::::::::::: ::
ACCCGCCTGACGATGGTGTTTCAGCACTTCAACCTCTGGAGCCACATGACGGTGCTGGAAAATGTGATGGAAGCGCCGATTCAGGTACTGGGATTAAGCAAGCACGACGCGCGCGAGCGGGCGTTG
ThrArgLeuThrMetValPheGlnHisPheAsnLeuTrpSerHisMetThrValLeuGluAsnValMetGluAlaProIleGlnValLeuGlyLeuSerLyshisaspAlaArgGluArgAlaleu

        860       870       880       890       900       910       920       930       940       950       960       970
LysTyrLeuAlaLysValGlyIleAspGluArgAlaGlnGlyLysTyrProValHisLeuSerGlyGlyGlnGlnGlnArgValSerIleAlaArgAlaLeuAlaMetGluProgluValLeuLeu
AAGTATCTGGCAAAAGTCGGGATAGACGAACGTGCGCAGGGGAAATATCCGGTGCATCTTTCGGCGGTCAACAACAACGTGTTTCTATCGCGCGGCGCTGGCGATGGAACCGGAAGTTTTACTG
:: :::::::: ::: :  :::::  :: :  :: :: :::::  :::::::: :: :::::  :::::  :::::  :: :: :::::::: :::::  :::::::::::::::::  :: :::::::
AAATATCTGGCGAAGGTGGGGATTGATGAGCGCGCTCAGGGCAAATATCCCGTCCATCTCTCCGGCGGCCAACAGCAGCGCGTTTCTATTGCGCGCGCGCTGGCGATGGAACCGGAAGTTTTACTG
LysTyrLeuAlaLysValGlyIleAspGluArgAlaGlnGlyLysTyrProValHisLeuSerGlyGlyGlnGlnGlnArgValSerIleAlaArgAlaLeuAlaMetGluProasspValLeuLeu

 980       990      1000      1010      1020      1030      1040      1050      1060      1070      1080      1090      1100
PheAspGluProThrSerAlaLeuAspProGluLeuValGlyGluleuLeuArgIleMetGlnGlnLeuAlaGluGluGlyLysThrMetValValValThrHisGluMetGlyPheAlaArgHis
TTTGATGAACCTACCTCGGCGCTCGATCCTGAACTGGTAGGCGAACTGTTGCGTATTATGCAGCAACTGGCAGAAGAGGGCAAAACCATGGTGGTAGTGACTCACGAAATGGGCTTTGCTCGCCAT
:: :::::::: :: :::::::::::::::::::::::: ::::: :::::::: :: :::::::: :::::: :::::: :::: :::::::: :::::: :: :: :: :: :::::::::::::
TTCGATGAACCCACATCGGCGCTCGATCCTGAACTGGTCGGCGAAGTGTTGCGCATCATGCAACAACTGGCGGAAGAAGGGCAAAACGATGGTGGTGGTCACGCATGAAATGGGCTTCGCTCGCCAT
PheAspGluProThrSerAlaLeuAspProGluLeuValGlyGluvalLeuArgIleMetGlnGlnLeuAlaGluGluGlyLysThrMetValValValThrHisGluMetGlyPheAlaArgHis

    1110      1120      1130      1140      1150      1160      1170      1180      1190      1200      1210      1220
ValSerthrHisValIlePheProProSerGlyGluAsnGluGluGluGlyalaProGluGlnleuPheGlyAsnProGlnSerProArgLeuGluGlnargPheLeuLysGlySerLeuLysSTOP
GTTTCTACTCATGTCATTTTCCTCCATCAGGGGAAAATAGAAGAAGAGGGCGCGCCGGAGCAGTTATTTGGCAACCCGCAAAGCCCTCGTCTGCAACCGTTCCTTAAGGGATCGCTGAAATAAT
:: ::: :: :: :: :::::  :: :::::::::::::  :::::::::::  :::::  :::::  :::::::::::  :::: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :
GTCTCTTCGCACGTTATTTTTCTGCATCAGGGGAAAATTGAAGAAGAGGGTGATCCGGAGCAGGTATTCGGCAATCCGCAAAGCCCGCGTTTACAGCAATTCCTGAAAGGCTCGCTGAAATAAA
ValSerserHisValIlePheProProSerGlyGluAsnGluGluGluGlyasgProGluGlnvalPheGlyAsnProGlnSerProArgLeuGlnglnPheLeuLysGlySerLeuLysSTOP

1230      1240      1250      1260      1270      1280      1290      1300      1310      1320      1330
TTCCTCTCCGGTGGCTAACGCTGCCGGAGTATTTTCATGGGGAATTTGCGAGTCAGCTTCCGCTGAAATGTAAAAGTTACAGTCCGCTAAAAATACTTCTGAATTC    E. coli
 : : :  :  :  :  :  :  :  :::::  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
AGAAGCCCGGCGGCGCATCGCTCACCGGGCAGGATACGGCGCTTTGGGAACTAGCGAATCACATCCGCCAGCGCCTCTTCTAAATCATACCAGCGAAATGCGAATC    S. typhimurium
```


© IRL Press Limited, Oxford, England.