Nucleotide sequence of the *aroP* gene encoding the general aromatic amino acid transport protein of *Escherichia coli* K-12: homology with yeast transport proteins

Nadine Honoré and Stewart T.Cole*
Laboratoire de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France

Submitted January 9, 1990

The *aroP* gene is situated between the *ampDE* operon (1) and the *gA-aceEF* gene cluster at 2.6 min. on the *E. coli* linkage map (2). Its complete nucleotide sequence has been determined and used to deduce the primary structure of AroP, a transport protein that takes up tryptophan, phenylalanine and tyrosine. Consistent with its membrane location (3) AroP is highly hydrophobic and apparently comprises two equivalent domains, each composed of six alpha-helical segments with membrane-spanning potential. In common with many transport proteins the observed molecular weight of AroP, 37 kD (3), is significantly lower than that predicted from the sequence, 49.6 kD. Although the hydropathy profile and general organisation of AroP resemble those of other bacterial transporters (4) there is no homology at the primary structure level. In contrast, AroP is highly homologous to two transport proteins from *Saccharomyces cerevisiae* involved in arginine and histidine uptake (5, 6). AroP and the yeast histidine permease HIP1 show 30% identity over 405 amino acid residues, suggesting a common evolutionary origin.

REFERENCES

1. Honoré,N., Nicolas,M.-H. and Cole,S.T. (1989) *Mol. Microbiol.* **3**, 1121–1130.
2. Bachmann,B.J. (1987) In *Escherichia coli* and *Salmonella typhimurium*, American Society for Microbiology (Neidhardt,F.C., ed.), pp. 807–876.
3. Chye,M.-L., Guest,J.R. and Pittard,J. (1986) *J. Bacteriol.* **167**, 749–753.
4. Maiden,M.C.J., Davids,E.O., Baldwin,S.A., Moore,D.C.M. and Henderson,P.J.F. (1987) *Nature* **325**, 641–643.
5. Tanaka,J. and Fink,G.R. (1985) *Gene* **38**, 205–214.
6. Hoffmann,W. (1985) *J. Biol. Chem.* **260**, 11831–11837.

```
                                   M  M  E  G  Q  Q  H  G  E  Q  L  K  R  G  L  K  N  R  H  I  Q  L  I  A  L  G  G  S  I
GTAGATCAAAAAAACAACCACCGCACGAGGTTTCATGATGGAAGGTCAACAGCACGGCGAGCAGCTAAAGCGCGGCCTTAAAAACCGCCATATTCAGCTTATCGCGCTGGGTGGCTCGAT
        10        20        30        40        50        60        70        80        90       100       110       120
  G  T  G  L  F  L  G  S  A  S  V  I  Q  S  A  G  P  G  I  I  L  G  Y  A  I  A  G  F  I  A  F  L  I  M  R  Q  L  G  E  M
AGGGACCGGGTTATTCCTGGGTAGCGCCTCCGTAATACAGTCCGCAGGCCCAGGGATTATCCTGGGTTACGCCATTGCTGGTTTTATCGCCTTTCTGATCATGCGTCAGCTGGGTGAAAT
       130       140       150       160       170       180       190       200       210       220       230       240
  V  V  E  E  P  V  A  G  S  F  S  H  F  A  Y  K  Y  W  G  S  F  A  G  F  A  S  G  W  N  Y  W  V  L  Y  V  L  V  A  M  A
GGTGGTCGAAGAACCTGTCGCAGGCTCCTTTAGCCACTTTGCTTATAAATACTGGGGCAGTTTTGCCGGTTTCGCCTCTGGCTGGAACTACTGGGTACTGTACGTTTTAGTTGCCATGGC
       250       260       270       280       290       300       310       320       330       340       350       360
  E  L  T  A  V  G  K  Y  I  Q  F  W  Y  P  E  I  P  T  W  V  S  A  A  V  F  F  V  V  I  N  A  I  N  L  T  N  V  K  V  F
TGAGCTGACTGCCGTGGGTAAATACATTCAGTTCTGGTATCCGGAAATCCCCACCTGGGTTTCTGCCGCCGTATTCTTTGTGGTGATTAACGCCATCAACCTGACCAACGTTAAAGTGTT
       370       380       390       400       410       420       430       440       450       460       470       480
  G  E  M  E  F  W  F  A  I  I  K  V  I  A  V  V  A  M  I  I  F  G  G  W  L  L  F  S  G  N  G  G  P  Q  A  T  V  S  N  L
TGGCGAGATGGAGTTCTGGTTTGCCATTATCAAAGTTATCGCGGTGGTAGCGATGATCATCTTCGGCGGCTGGCTGCTATTCAGTGGCAACGGCGGCCCGCAGGCGACCGTTAGCAACCT
       490       500       510       520       530       540       550       560       570       580       590       600
  W  D  Q  G  G  F  L  P  H  A  F  T  G  L  V  M  M  M  A  I  I  M  F  S  F  G  G  L  E  L  V  G  I  T  A  A  E  A  D  N
GTGGGATCAGGGTGGTTTCCTGCCGCACGCGTTCACCGGGCTGGTGATGATGATGGCGATTATCATGTTCTCGTTCGGTGGTCTGGAACTGGTGGGGATCACCGCAGCAGAAGCTGATAA
       610       620       630       640       650       660       670       680       690       700       710       720
  P  E  Q  S  I  P  K  A  T  N  Q  V  I  Y  R  I  L  I  F  Y  I  G  S  L  A  V  L  L  S  L  M  P  W  T  R  V  T  A  D  T
CCCGGAGCAAAGTATACCGAAAGCAACTAACCAGGTTATCTACCGCATCCTGATTTTCTATATTGGTTCGTTAGCCGTTCTGCTCTCACTGATGCCGTGGACCCGCGTTACCGCCGATAC
       730       740       750       760       770       780       790       800       810       820       830       840
  S  P  F  V  L  I  F  H  E  L  G  D  T  F  V  A  N  A  L  N  I  V  V  L  T  A  A  L  S  V  Y  N  S  C  V  Y  C  N  S  R
CAGTCCGTTTGTGCTGATCTTCCACGAGTTAGGCGATACCTTTGTGGCGAATGCGCTGAACATCGTGGTACTGACTGCGGCGCTCTCCGTGTACAACAGCTGCGTATATTGCAACAGCCG
       850       860       870       880       890       900       910       920       930       940       950       960
  M  L  F  G  L  A  Q  Q  G  N  A  P  K  A  L  A  S  V  D  K  R  G  V  P  V  N  T  I  L  V  S  A  L  V  T  A  L  C  V  L
TATGCTGTTTGGTCTGGCACAACAGGGTAATGCGCCAAAAGCGCTGGCGTCTGTCGATAAACGTGGTGTACCAGTAAATACCATTCTGGTGTCTGCACTGGTAACGGCGTTGTGCGTACT
       970       980       990      1000      1010      1020      1030      1040      1050      1060      1070      1080
  I  N  Y  L  A  P  E  S  A  F  G  L  L  M  A  L  V  V  S  A  L  V  I  N  W  A  M  I  S  L  A  Ḥ  M  K  F  R  R  A  K  Q
GATTAACTACCTTGCCCCAGAGTCCGCTTTCGGACTGTTAATGGCGCTGGTGGTATCTGCACTGGTAATCAACTGGGCGATGATTAGCCTGGCGCATATGAAATTCCGTCGCGCCAAGCA
      1090      1100      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
  E  Q  G  V  V  T  R  F  L  L  L  Y  P  L  G  N  W  I  C  L  L  F  M  A  A  V  L  V  I  M  L  M  T  P  G  M  A  I  S  V
GGAACAAGGCGTGGTAACTCGCTTCCTGCTGCTTTATCCGCTGGGTAACTGGATCTGCCTGCTGTTTATGGCGGCGGTACTGGTGATTATGCTGATGACCCCAGGAATGGCGATTTCGGT
      1210      1220      1230      1240      1250      1260      1270      1280      1290      1300      1310      1320
  Y  L  I  P  V  W  L  I  V  L  G  I  G  Y  L  F  K  E  K  T  A  K  A  V  K  A  H  *
ATACCTGATCCCGGTATGGCTGATCGTGTTAGGTATCGGCTATCTGTTTAAAGAGAAAACCGCCAAAGCCGTAAAAGCGCATTAATCTCTCTACGCCCTCACCCGTACAGGGTGAGGGCA
      1330      1340      1350      1360      1370      1380      1390      1400      1410      1420      1430      1440
```

---

* To whom correspondence should be addressed