JOURNAL OF BACTERIOLOGY, May 1991, p. 3231–3234
0021-9193/91/103231-04$02.00/0

Vol. 173, No. 10

A New Family of Integral Membrane Proteins Involved in Transport of Aromatic Amino Acids in *Escherichia coli*

J. P. SARSERO,[1] P. J. WOOKEY,[1] P. GOLLNICK,[2]† C. YANOFSKY,[2]* AND A. J. PITTARD[1]

Department of Microbiology, University of Melbourne, Parkville, Victoria 3052, Australia,[1] and Department of Biological Sciences, Stanford University, Stanford, California 94305-5020[2]

Received 26 December 1990/Accepted 8 March 1991

The nucleotide sequence of *tnaB* of the tryptophanase operon of *Escherichia coli* is presented. TnaB is a tryptophan-specific permease that is homologous to Mtr, a second tryptophan-specific permease, and to TyrP, a tyrosine-specific permease. Each member of this family appears to contain 11 membrane-spanning domains.

Transport of aromatic amino acids in *Escherichia coli* is carried out by at least five distinct systems. Four of these were first described by Brown (4), following description of the analogous systems in *Salmonella typhimurium* by Giovanna Ames (1). One of these is called the general aromatic transport system and transports phenylalanine, tyrosine, and tryptophan. The other three are specific for individual amino acids. The K_m for the general transport system is about 4×10^{-7} M, and for the three specific systems it is about fivefold higher. The genes for all of these systems have been mapped. The general transport system is coded for by *aroP* (4), the tyrosine- and phenylalanine-specific systems are coded for by *tyrP* and *pheP* (25), and the tryptophan-specific system is coded for by *mtr* (C. Yanofsky cited by Oxender [20]). The gene products of *tyrP* and *aroP* have been shown to be integral membrane proteins (7, 28). Studies with uncouplers and with strains deficient in Mg-ATPase indicate that transport via the AroP, TyrP, and PheP systems is driven by the proton motive force (24a, 28). Each of these systems appears to involve a single component analogous to the lactose permease. The *aroP* gene has been cloned (7, 14) and sequenced (14), as has *tyrP* (27, 28). The cloning and sequencing of *pheP* and *mtr* have been completed (13a, 20a, 21a).

The fifth gene involved in transport of the aromatic amino acids is *tnaB* (9), also designated *trpP* (2). This gene, which is part of a transcription unit containing the structural gene for the enzyme tryptophanase, codes for a tryptophan-specific transport system. This system, referred to as a low-affinity system, is reported to have a K_m of about 7×10^{-5} M (5) and to play a role in transporting tryptophan which is to be used catabolically. Transport by this system is inhibited by glucose, presumably as a result of direct inhibition by dephosphorylated enzyme III^{Glc} (5, 19). However, transcription of the *tnaAB* operon is subject to catabolite repression and tryptophan induction (8, 12, 22). Preliminary results of the cloning and partial sequencing of this gene have been reported (8).

The purposes of this note are to report the complete DNA sequence of *tnaB* and the amino acid sequence of the protein for which it codes and to compare the amino acid sequence of TnaB with those of the Mtr and TyrP proteins. The proteins encoded by *pheP* and *aroP*, although closely homologous to each other, bear little overall similarity to the three proteins described in this report (14, 20a).

tnaB is in the tryptophanase operon, 90 bp downstream from *tnaA*, the structural gene for tryptophanase (8). The sequences of *tnaA* and the *tnaA-tnaB* intercistronic region and the first 13 codons of *tnaB* have been reported previously (8). Figure 1 shows the completed nucleotide sequence of *tnaB*, including the previously reported upstream region. *tnaB* was sequenced on both strands by using the dideoxy-chain termination method (21). Sequencing was performed with single-stranded DNA prepared from discrete restriction fragments cloned from pMD6 (8) into pUC118 or pUC119 (23). Single-stranded DNA was prepared by using M13K07 helper bacteriophage (23). The sequence of the 300 bp downstream of *tnaB* is also presented in Fig. 1. There are no additional open reading frames in this segment of the sequence.

A single open reading frame codes for a strongly hydrophobic polypeptide of 415 amino acid residues (Fig. 1). This sequence is compared with the sequences of the Mtr and TyrP polypeptides in Fig. 2. As can be seen, there is striking similarity between TnaB and Mtr, with 52% of the residues being identical. Whereas there is 31% identity between TnaB and TyrP and 33% identity between Mtr and TyrP, only 23% of the residues are conserved in all three proteins. On this basis, it appears that an ancestral gene duplicated and one copy evolved into the current tyrosine transporter, with the second copy serving as the precursor for the two genes currently involved in tryptophan transport.

The hydrophobicity profiles of the three proteins derived by using the algorithm of Engelman et al. (10) and a window of 19 amino acids are shown in Fig. 3. It appears that for at least two and probably three of these proteins there are 11 hydrophobic spans linked with hydrophilic loops in the cytoplasm and the periplasm. Although 12 membrane spans have been postulated for many of the sugar transport proteins (17) and for the two aromatic amino acid transport proteins AroP and PheP (20a), two factors argue for the presence of only 11 spans for these three proteins. (i) With a size range of 403 to 415 amino acids, they are smaller than other proteins postulated to form 12 spans, such as GlpT, UhpT, AroP, and PheP (452, 463, 458, and 457 residues, respectively) (11, 13, 16). This is not a sufficient argument in itself, since the lactose permease which is proposed to occupy 12 spans has only 417 amino acids (15). However, all of these proteins have clusters of the positively charged amino acids, arginine and lysine, in many of their hydrophilic cytoplasmic loops. In particular, the cytoplasmic loop

* Corresponding author.
† Present address: Department of Biological Sciences, State University of New York at Buffalo, Buffalo, NY 14260.

```
   1 CACTTCACCGCAAAACTTAAAGAAGTTTAATTAATACTACAGAGTGGCTATAAGGATGTT
     HisPheThrAlaLysLeuLysGluValEnd

  61 AGCCACTCTCTTACCCTACATCCTCAATAACAAAAATAGCCTTCCTCTAAAGGTGGCATC

 121 ATGACTGATCAAGCTGAAAAAAAGCACTCTGCATTTTGGGGTGTTATGGTTATAGCAGGT
     MetThrAspGlnAlaGluLysLysHisSerAlaPheTrpGlyValMetValIleAlaGly

 181 ACAGTAATTGGTGGAGGTATGTTTGCTTTACCTGTTGATCTTGCCGGTGCCTGGTTTTTC
     ThrValIleGlyGlyGlyMetPheAlaLeuProValAspLeuAlaGlyAlaTrpPhePhe

 241 TGGGGTGCCTTTATCCTTATCATTGCCTGGTTTTCAATGCTTCATTCCGGGTTATTGTTA
     TrpGlyAlaPheIleLeuIleIleAlaTrpPheSerMetLeuHisSerGlyLeuLeuLeu

 301 TTAGAAGCAAATTTAAATTATCCCGTCGGCTCCAGTTTTAACACCATCACCAAAGATTTA
     LeuGluAlaAsnLeuAsnTyrProValGlySerSerPheAsnThrIleThrLysAspLeu

 361 ATCGGTAACACCTGGAACATTATCAGCGGTATTACCGTTGCCTTCGTTCTCTATATCCTC
     IleGlyAsnThrTrpAsnIleIleSerGlyIleThrValAlaPheValLeuTyrIleLeu

 421 ACTTATGCCTATATCTCTGCTAATGGTGCGATCATTAGTGAAACGATATCAATGAATTTG
     ThrTyrAlaTyrIleSerAlaAsnGlyAlaIleIleSerGluThrIleSerMetAsnLeu

 481 GGTTATCACGCTAATCCACGTATTGTCGGGATCTGCACAGCCATTTTCGTTGCCAGCGTA
     GlyTyrHisAlaAsnProArgIleValGlyIleCysThrAlaIlePheValAlaSerVal

 541 TTGTGGTTAAGTTCGTTAGCCGCCAGTCGTATTACCTCATTGTTCCTCGGGCTGAAGATT
     LeuTrpLeuSerSerLeuAlaAlaSerArgIleThrSerLeuPheLeuGlyLeuLysIle

 601 ATCTCCTTTGTGATCGTGTTTGGTTCTTTTTTCTTCCAGGTCGATTACTCCATTCTGCGC
     IleSerPheValIleValPheGlySerPhePhePheGlnValAspTyrSerIleLeuArg

 661 GACGCCACCAGCTCCACTGCGGGAACGTCTTACTTCCCGTATATCTTTATGGCTTTGCCG
     AspAlaThrSerSerThrAlaGlyThrSerTyrPheProTyrIlePheMetAlaLeuPro

 721 GTGTGTCTGGCGTCATTTGGTTTCCACGGCAATATTCCCAGCCTGATTATTTGCTATGGA
     ValCysLeuAlaSerPheGlyPheHisGlyAsnIleProSerLeuIleIleCysTyrGly

 781 AAACGCAAAGATAAGTTAATCAAAAGCGTGGTATTTGGTTCGCTGCTGGCGCTGGTGATT
     LysArgLysAspLysLeuIleLysSerValValPheGlySerLeuLeuAlaLeuValIle

 841 TATCTCTTCTGGCTCTATTGCACCATGGGGAATATTCCGCGAGAAAGCTTTAAGGCGATT
     TyrLeuPheTrpLeuTyrCysThrMetGlyAsnIleProArgGluSerPheLysAlaIle

 901 ATCTCCTCAGGCGGCAACGTTGATTCGCTGGTGAAATCGTTCCTCGGCACCAAACAGCAC
     IleSerSerGlyGlyAsnValAspSerLeuValLysSerPheLeuGlyThrLysGlnHis

 961 GGCATTATCGAGTTTTGCCTGCTGGTGTTCTCTAACTTAGCTGTTGCCAGTTCGTTCTTT
     GlyIleIleGluPheCysLeuLeuValPheSerAsnLeuAlaValAlaSerSerPhePhe

1021 GGTGTCACGCTGGGGTTGTTCGATTATCTGGCGGACCTGTTTAAGATTGATAACTCCCAC
     GlyValThrLeuGlyLeuPheAspTyrLeuAlaAspLeuPheLysIleAspAsnSerHis

1081 GGCGGGCGTTTCAAAACCGTGCTGTTAACCTTCCTGCCACCTGCGTTGTTGTATCTGATC
     GlyGlyArgPheLysThrValLeuLeuThrPheLeuProProAlaLeuLeuTyrLeuIle

1141 TTCCCGAACGGCTTTATTTACGGGATCGGCGGTGCCGGGCTGTGCGCCACCATCTGGGCG
     PheProAsnGlyPheIleTyrGlyIleGlyGlyAlaGlyLeuCysAlaThrIleTrpAla

1201 GTCATTATTCCCGCAGTGCTTGCAATCAAAGCTCGCAAGAAGTTTCCCAATCAGATGTTC
     ValIleIleProAlaValLeuAlaIleLysAlaArgLysLysPheProAsnGlnMetPhe

1261 ACGGTCTGGGGCGGCAATCTTATTCCGGCGATTGTCATTCTCTTTGGTATAACCGTGATT
     ThrValTrpGlyGlyAsnLeuIleProAlaIleValIleLeuPheGlyIleThrValIle

1321 TTGTGCTGGTTCGGCAACGTCTTTAACGTGTTACCTAAATTTGGCTAAATCCTTCAAGAA
     LeuCysTrpPheGlyAsnValPheAsnValLeuProLysPheGlyEnd

1381 GCCAGCCATTCGCTGGCTTCTTGCCTCTCAGGAAATCACTTATGTCCAAATGGCAACTCG

1441 CCTGATCCTCCTTCACCACGTATGCTTTGCGTCACCTTACTATCAGGACGCTTTAGCCCA

1501 TGTCCCGCTTTTTGATTTGTAGTTTTGCCCTGGTTTTACTTTATCCGCCGGGATTGATAT

1561 GTACCTCGTTGGTTTACCGCGCATCGCCGCCGATCTCAATGCCAGCGAAGCGCAGTTGCA

1621 TATTGCGTTCTCCGTATATCTGGCGGGGATGGCAGATCAAGCTTATCGATACCGTCGACC

1681 TCGATCGA
```

FIG. 1. Nucleotide sequence of the *E. coli tnaB* gene. The derived amino acid sequence of TnaB is indicated, and the proposed ribosome-binding site for *tnaB* is underlined. The sequences of the carboxyl terminus of tryptophanase and the 90-bp intercistronic region are taken from reference 8.

between putative spans X and XI contains a number of such charged residues. In LacY, UhpT, GlpT, AroP, and PheP, there are at least 53 amino acids between the last of these charged residues and the carboxyl terminus. It is possible to fold these sequences into two hydrophobic spans with the additional positive charges which occur at the carboxyl ends of all of these proteins remaining in the cytoplasm. Mtr and TyrP have only 28 and 29 residues, respectively, between these positively charged amino acids and the carboxyl terminus. Also, neither of these proteins possesses positively charged residues at its carboxyl terminus. Figure 4 compares the carboxyl termini of these various proteins. Both of these observations indicate that span XI is the last span, with the carboxyl terminus in the periplasm. (Alternatively, the carboxyl end could be embedded in the membrane but would fail to constitute a complete 12th span.) The situation with TnaB is not so clear, as there are 41 residues after a strong group of charged amino acids and a single lysine is found as the third residue from the carboxyl end.

By reference to the hydrophobicity profiles (10, 26), the incidence and strength of turns (6), and the distribution of

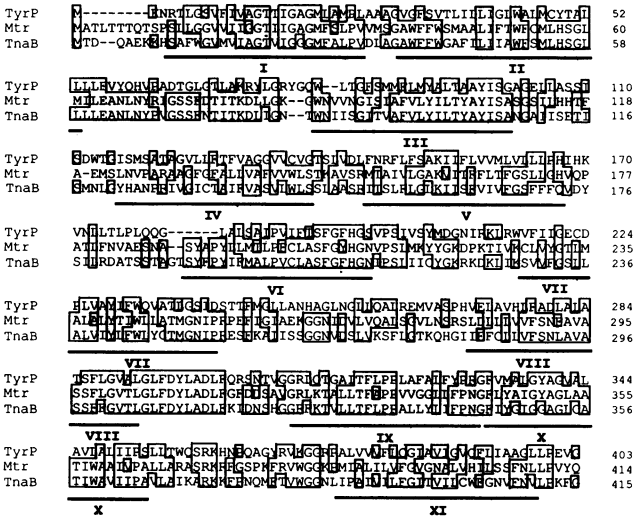

FIG. 2. Comparison of the amino acid sequences of proteins Mtr, TnaB, and TyrP. Identical residues are boxed. The putative hydrophobic spans shown for Mtr in Fig. 5 are indicated by underlining. The TyrP sequence differs from the one previously published (27) in that the sequence from position 131 is AGGVV and not RRVANL. We thank Susan Howitt for this correction.

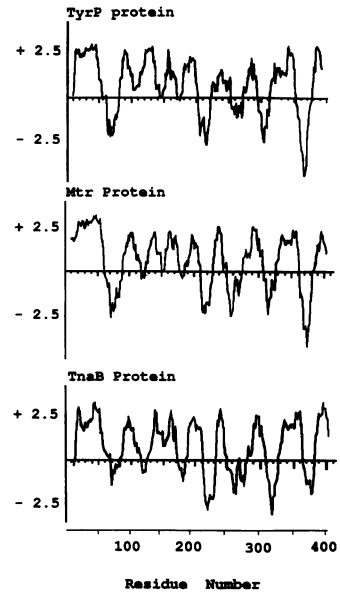

FIG. 3. Comparison of the hydropathy profiles of the three proteins derived by using the algorithm of Engelman et al. (10) with a window of 19 amino acids.

UhpT	E^- E^- R^+ K^+ I R^+ R^+ E K^+ K^+ I Q Q L T V A
GlpT	E^- K^+ R^+ R^+ H E Q L L Q E^- R^+ N G G
LacY	R^+ R^+ Q V N E^- V A
PheP	K^+ T L R^+ R^+ K^+
AroP	K^+ E^- K^+ T A K^+ A V K^+ A H
TyrP	L L P E^- V G
Mtr	L L P V Y Q
TrpP	V L P K^+ F G

FIG. 4. Carboxyl termini of a number of transport proteins, showing the clusters of positive charges in those for which 12 spans have been proposed.

charged amino acids (18, 24), it is possible to propose a model for the secondary structure of these proteins. Although this is speculative, it does simplify further discussion of these proteins and forms the basis for future experiments. Figure 5 shows one such proposal for the structure of Mtr. The residues conserved in all three proteins are in boldface. According to this model, approximately two-thirds of the conserved residues are found in the regions spanning the membrane. However, the distribution is not even, with spans IV, V, and XI showing very few conserved residues. Of particular interest is a block of conserved residues which extend through the latter half of span VIII into the following cytoplasmic loop. Among the individual amino acids, glycine (20 of 41) and proline (8 of 17) show the greatest conservation. A relative abundance of glycines is seen in span I of these proteins and also in span I of AroP and PheP. The lysine residue present within span V is conserved in all five of the transport proteins, TnaB, Mtr, TyrP, AroP, and PheP. Proline residues are distributed about equally between the hydrophobic core and the hydrophilic loops. Brandl and Deben (3) have previously noted the presence of membrane-buried proline residues in transport proteins and proposed a possible role for them in the regulation of transport channels. In agreement with the observations and predictions of others (18, 24), the cytoplasmic loops of the three proteins carry a net positive charge whereas the periplasmic loops carry a net negative charge. With reference to the proposed model for Mtr, the cytoplasmic loops have a net positive charge of +9 and the periplasmic loops have a net negative charge of −4.

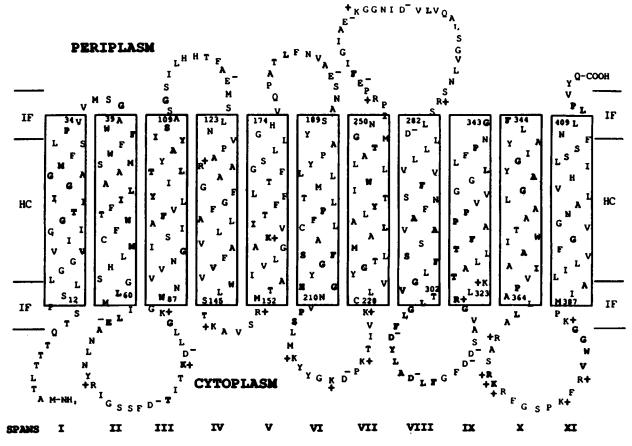

FIG. 5. Model of the arrangement of the Mtr protein across the membrane. Amino acid residues conserved in Mtr, TnaB, and TyrP are in boldface. IF, interface; HC, hydrocarbon core (3 nm, corresponding to the length of a 20-residue helix).

Of particular interest for future experiments on structure and function are the considerable K_m differences reported for TnaB and Mtr, the fact that one of these appears to be inhibited by enzyme III^{Glc}, and the different specificities of Mtr and TnaB versus TyrP.

While preparing this report, we became aware of the results of Heatwole and Somerville (13a), who have also cloned and sequenced *mtr*. The two sequences for the coding region are identical, but those researchers propose the likelihood of 12 membrane spans.

This work was supported by grants from the Australian Research Council and the National Science Foundation (grant DMB 8703685). J. P. Sarsero was the recipient of a Commonwealth Post Graduate Award.

We thank Frank Gibson for assistance with the hydrophobicity profiles and Frank Gibson, Graeme Cox, and Pi Jing for many discussions on possible structures. We also thank Michael Deeley and Valley Stewart for preliminary partial sequence data for *tnaB*.

REFERENCES

1. **Ames, G. F.** 1964. Uptake of amino acids by *Salmonella typhimurium*. Arch. Biochem. Biophys. **104**:1–18.
2. **Bachmann, B. J.** 1983. Linkage map of *Escherichia coli* K-12, edition 7. Microbiol. Rev. **47**:180–230.
3. **Brandl, C. J., and C. M. Deben.** 1986. Hypothesis about the function of membrane-buried proline residues in transport proteins. Proc. Natl. Acad. Sci. USA **83**:917–921.
4. **Brown, K. D.** 1970. Formation of aromatic amino acid pools in *Escherichia coli*. J. Bacteriol. **104**:177–188.
5. **Burrous, S., and R. D. DeMoss.** 1963. Studies of tryptophan permease in *Escherichia coli*. Biochem. Biophys. Acta **73**:623–637.
6. **Chou, P. Y., and S. D. Fasman.** 1974. Conformational parameters for amino acids in helical, β-sheet and random coil regions calculated from proteins. Biochemistry **13**:211–221.
7. **Chye, M. L., J. R. Guest, and J. Pittard.** 1986. Cloning of the *aroP* gene and the identification of its product in *Escherichia coli* K-12. J. Bacteriol. **167**:749–753.
8. **Deeley, M. C., and C. Yanofsky.** 1981. Nucleotide sequence of the structural gene for tryptophanase of *Escherichia coli* K-12. J. Bacteriol. **147**:787–796.
9. **Edwards, R. M., and M. D. Yudkin.** 1982. Location of the gene for the low-affinity tryptophan-specific permease of *Escherichia coli*. Biochem. J. **204**:617–619.
10. **Engelman, D. M., T. A. Steitz, and A. Goldman.** 1986. Identifying nonpolar trans bilayer helices in amino acid sequences of membrane proteins. Annu. Rev. Biophys. Biophys. Chem. **15**:321–353.
11. **Friedrich, M. J., and R. F. Kadner.** 1987. Nucleotide sequence of the *uhp* region of *Escherichia coli*. J. Bacteriol. **169**:3556–3563.
12. **Gollnick, P., and C. Yanofsky.** 1990. $tRNA^{Trp}$ translation of leader peptide codon 12 and other factors that regulate expression of the tryptophanase operon. J. Bacteriol. **172**:3100–3107.
13. **Gott, P., and W. Boos.** 1988. The transmembrane topology of the Sn-glycerol-3-phosphate permease of *Escherichia coli*, analyzed by PhoA and LacZ protein fusions. Mol. Microbiol. **2**:655–663.

13a. **Heatwole, V. M., and R. L. Somerville.** 1991. Cloning, nucleotide sequence, and characterization of *mtr*, the structural gene for a tryptophan-specific permease of *Escherichia coli* K-12. J. Bacteriol. **173**:108–115.

14. **Honore, N., and S. T. Cole.** 1990. Nucleotide sequence of the *aroP* gene encoding the general aromatic transport protein of *Escherichia coli*. Homology with yeast transport proteins. Nucleic Acids Res. **18**:653.
15. **Kaback, H. R., E. Bibi, and P. D. Roepe.** 1990. β-Galactoside transport in *E. coli*: a functional dissection of Lac permease.

Trends Biol. Sci. **15**:309–314.
16. **Lloyd, A. D., and R. J. Kadner.** 1990. Topology of the *Escherichia coli uhpT* sugar-phosphate transporter analyzed by using Tn*phoA* fusions. J. Bacteriol. **172**:1688–1693.
17. **Maloney, P. C.** 1990. A consensus structure for membrane transport. Res. Microbiol. **141**:374–383.
18. **Michel, H., K. A. Weyer, H. Gruenberg, I. Dunger, D. Oesterhelt, and F. Lottspeich.** 1986. The "light" and "medium" subunits of the photosynthetic reaction centre from *Rhodopseudomonas viridis*: isolation of the genes, nucleotide, and amino acid sequence. EMBO J. **5**:1149–1158.
19. **Nelson, S. O., J. Lengeler, and P. W. Postma.** 1984. Role of III^{Glc} of the phosphoenolpyruvate-glucose phosphotransferase system in inducer exclusion in *Escherichia coli*. J. Bacteriol. **160**:360–364.
20. **Oxender, D. L.** 1975. Genetic approaches to the study of transport systems, p. 214–231. *In* H. N. Christensen (ed.), Biological transport, 2nd ed. W. A. Benjamin, New York.
20a. **Pí, J., P. J. Wookey, A. J. Pittard, and J. Sarsero.** Submitted for publication.
21. **Sanger, F., S. Niklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
21a. **Sarsero, J. P., P. J. Wookey, and A. J. Pittard.** Regulation of expression of the *Escherichia coli mtr* gene by the TyrR protein and the *trp* repressor. J. Bacteriol., in press.
22. **Stewart, V. J., and C. Yanofsky.** 1985. Evidence for transcription antitermination control of tryptophanase operon expression in *Escherichia coli* K-12. J. Bacteriol. **164**:731–740.
23. **Vieira, J., and J. Messing.** 1987. Production of single-stranded plasmid DNA. Methods Enzymol. **153**:3–11.
24. **Von Heinje, G.** 1986. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. EMBO J. **5**:3021–3027.
24a. **Whipp, M.** 1977. Ph.D. thesis. University of Melbourne, Melbourne, Australia.
25. **Whipp, M. J., D. M. Halsall, and A. J. Pittard.** 1980. Isolation and characterization of an *Escherichia coli* K-12 mutant defective in tyrosine- and phenylalanine-specific transport systems. J. Bacteriol. **143**:1–7.
26. **White, S. H., and R. E. Jacobs.** 1990. Observations concerning topology and location of helix ends of membrane proteins of known structure. J. Membr. Biol. **115**:145–158.
27. **Wookey, P. J., and A. J. Pittard.** 1988. DNA sequence of the gene (*tyrP*) encoding the tyrosine-specific transport system of *Escherichia coli*. J. Bacteriol. **170**:4946–4949.
28. **Wookey, P. J., J. Pittard, S. M. Forrest, and B. E. Davidson.** 1984. Cloning of the *tyrP* gene and further characterization of the tyrosine-specific transport system in *Escherichia coli* K-12. J. Bacteriol. **160**:169–174.