JOURNAL OF BACTERIOLOGY, Mar. 1995, p. 1627–1629
0021-9193/95/$04.00+0

Vol. 177, No. 6

A Reassessment of the Relationship between *aroK*- and *aroL*-Encoded Shikimate Kinase Enzymes of *Escherichia coli*

M. J. WHIPP AND A. J. PITTARD*

Department of Microbiology, University of Melbourne, Parkville, Victoria 3052, Australia

Received 29 August 1994/Accepted 10 January 1995

In the course of sequencing the *aroK* gene, a number of errors were found in the published sequence. The corrected sequence alters the length of the *aroK* coding region such that the AroK and AroL proteins are now of comparable length and the homology between them extends the entire length of the two enzymes.

Escherichia coli synthesizes two shikimate kinase enzymes whose structures are encoded by the genes *aroL* for shikimate kinase II (3, 5) and *aroK* for shikimate kinase I (11). The *aroL*-encoded enzyme has a K_m of about 200 μM for shikimate and appears to play a major role in the biosynthesis of the aromatic amino acids. Expression of the *aroL* gene is controlled by both the TyrR and TrpR proteins and the amino acids tyrosine and tryptophan as cofactors (3, 5, 6, 8). The roles of *aroK* and shikimate kinase I are less clear. Expression of the gene appears to be constitutive, and the enzyme has a much lower affinity for shikimate, with a reported K_m of greater than 20 mM (4). The *aroK* gene was identified on the basis of the amino acid sequence homology between its putative product and the *aroL*-encoded protein. Its identity was confirmed by the construction of double mutants (*aroK aroL*) which were shown to be aromatic amino acid auxotrophs (11). The sequence of *aroK* published by Løbner-Olesen and Marinus (11) described a protein of either 139 or 144 amino acids, depending on the start point of translation, and showing 34% homology with the AroL protein over a limited sequence of 97 amino acids. The published sequence for the *aroL*-encoded protein describes a protein of 174 amino acids (2).

During some studies of a putative *aroK* mutant (JP1084), we had occasion to sequence the *aroK* gene in this strain and in a related *aroK*$^+$ strain (JP7615). We discovered a number of errors in the published sequence of the *aroK* region (10, 11, 13) which, when corrected, alter the length of the *aroK* coding region to 173 amino acids and extend the homology between the two shikimate kinase enzymes to their entire length. Since we observed no differences in the *aroK* sequences from JP7615 and JP1084, we conclude that JP1084 is not mutated in *aroK* and probably carries a mutation in *aroB* which is in the same transcriptional unit as *aroK* (10).

Cloning and sequencing *aroK*. Chromosomal DNA was obtained from strains JP7615 and JP1084 by the method of Slauch and Silhavy (21) and used as the DNA template in PCRs. The DNA sequence of Løbner-Olesen et al. (10) was used to design oligonucleotides flanking the *aroK* coding sequence. The forward primer used in these reactions corresponded to nucleotides (nt) 1443 to 1460 and contained an *Eco*RI linker at the 5′ end. The reverse primer corresponded to nt 2275 to 2257 and contained a *Bam*HI linker at the 5′ end. The DNA fragments were cloned into M13 vectors, and both strands were sequenced from three independent clones (20). Compressions were resolved by replacing dGTP with dITP in the sequencing reactions by the method of Mills and Kramer (15). When the sequence analysis indicated that the *aroK* reading frame extended further than reported by Løbner-Olesen and Marinus (11), a second reverse primer corresponding to nt 2511 to 2493, again with a *Bam*HI linker at the 5′ end, was used to clone and sequence a longer fragment. (Note that nt 2003 to 2520 corresponds to nt 9 to 523 of the published *aroB* sequence of Millar and Coggins [13]).

A number of differences between our sequence and the published *aroK* sequence were found (Fig. 1). A critical error occurred downstream of the putative translational start codon of the published sequence. The sequence starting at nt 1812 should be changed from TGC CGG GCC T to TGC GGC GCT. This change results in a frameshift which introduces a stop codon two codons further on, making it impossible for either of the reported ATG codons (nt 1792 to 1794 and 1807 to 1809) to function as the translational start of *aroK*. The sequencing gels for this region are shown in Fig. 2. In addition, an extra T was identified at nt 1500, an extra G was identified at nt 2167, an extra T was identified at nt 2222, and an extra C was identified at nt 2294. Analysis of the corrected sequence revealed a putative open reading frame extending from nt 1659 to 2393, just 56 nt before the translational start site of *aroB*.

Identifying the translational start site. To determine the translational start site for *aroK*, we first constructed an *aroK-lacZ* translational fusion and then used oligonucleotide site-directed mutagenesis to localize the actual site. PCR was used to generate a suitable fragment. The forward primer was that used previously, while the reverse primer corresponded to nt 1966 to 1947 and contained a *Bam*HI linker at its 5′ end. The fragment generated by PCR was cloned into an M13 vector, and its sequence was checked for any misincorporated nucleotides. The translational fusion was constructed by inserting the *Eco*RI-*Bam*HI *aroK* fragment into *Eco*RI-and-*Bam*HI-cleaved pMU2386 (8, 18). In this plasmid, pMU3258, *aroK* is fused in phase with codon 8 of *lacZ* so that β-galactosidase activity is dependent on transcription and translation from *aroK*.

The translational start site would be expected to lie upstream of the conserved site A of the putative ATP-binding sequence (Fig. 3). To approximately localize the start position in this upstream sequence, a stop codon was introduced at nt 1791 to 1793 (TAT to TAA) by site-directed mutagenesis (using a commercially available kit from United States Biochemical Corp.). The sequence was verified, and the fragment was inserted into pMU2386 to give plasmid pMU3259. The β-galactosidase activity expressed from this fusion is the same as that from the wild-type fusion (Table 1). This result indicated that the translational start was downstream of nt 1793.

The most likely site was considered to be the ATG codon at

* Corresponding author.

```
                    ⇒              ⇒        * *
1779 CTCGCAATATCTTATGAGGTTTCAGTTCATGTCCTGCGGCGCTCTCTGAGCGAAGCGGGT

1839 TTATCATTAACGAATAGTCTTAGTAGTACCGAAAAAATGGCAGAGAAACGCAATATCTTT
                                        M  A  E  K  R  N  I  F

1899 CTGGTTGGGCCTATGGGTGCCGGAAAAAGCACTATTGGGCGCCAGTTAGCTCAACAACTC
     L  V  G  P  M  G  A  G  K  S  T  I  G  R  Q  L  A  Q  Q  L

1959 AATATGGAATTTTACGATTCCGATCAAGAGATTGAGAAACGAACCGGAGCTGATGTGGGC
     N  M  E  F  Y  D  S  D  Q  E  I  E  K  R  T  G  A  D  V  G

2019 TGGGTTTTCGATTTAGAAGGCGAAGAAGGCTTCCGCGATCGCGAAGAAAAGGTCATCAAT
     W  V  F  D  L  E  G  E  E  G  F  R  D  R  E  E  K  V  I  N

2079 GAGTTGACCGAGAAACAGGGTATTGTGCTGGCTACTGGCGGCGGCTCTGTGAAATCCCGT
     E  L  T  E  K  Q  G  I  V  L  A  T  G  G  G  S  V  K  S  R
                                    *
2139 GAAACGCGTAACCGTCTTTCCGCTCGTGGCGTTGTCGTTTATCTTGAAACGACCATCGAA
     E  T  R  N  R  L  S  A  R  G  V  V  V  Y  L  E  T  T  I  E
                        * ⇓
2199 AAGCAACTTGCACGCACGCAGCGTGATAAAAAACGCCCGTTGCTGCACGTTGAAACACCG
     K  Q  L  A  R  T  Q  R  D  K  K  R  P  L  L  H  V  E  T  P
                                    *
2259 CCGCGTGAAGTTCTGGAAGCGTTGGCCAATGAACGCAATCCGCTGTATGAAGAGATTGCC
     P  R  E  V  L  E  A  L  A  N  E  R  N  P  L  Y  E  E  I  A

2319 GACGTGACCATTCGTACTGATGATCAAAGCGCTAAAGTGGTTGCAAACCAGATTATTCAC
     D  V  T  I  R  T  D  D  Q  S  A  K  V  V  A  N  Q  I  I  H

2379 ATGCTGGAAAGCAACTAA
     M  L  E  S  N  *
```

FIG. 1. The nucleotide sequence of the *aroK* gene and the predicted amino acid sequence. The possible start sites (⇒) and the stop site predicted previously (11) (⇓) are indicated. The putative ribosome-binding site is underlined twice. The sites of corrections to the nucleotide sequence (*) are indicated. The sites of introduced mutations are shown in bold type.

nt 1875 to 1877. This codon is preceded by a possible Shine-Dalgarno sequence, TtAGtAGt, which coincides with five of the eight possible bases of the consensus sequence (TAAGG AGG) of Rinquist et al. (19). In order to test the role of this ATG codon, site-directed mutagenesis was used to change it to ACG, a codon not known to act as an initiation codon, and the effect of this mutation was tested in the *lacZ* translational fusion. The effect of this mutation in plasmid pMU3260 (Table 1) was to reduce β-galactosidase levels more than 1,000-fold, supporting the hypothesis that this ATG is the initiation codon.

The *aroK* gene is the first gene in a transcription unit including *aroB*, an unidentified reading frame and the *dam* gene (10). Insertional mutations in *aroK* such as *aroK*::Cam^r, in addition to destroying *aroK* activity also have a polarity effect on *aroB* expression (11). Since *aroB* encodes the second enzyme in the

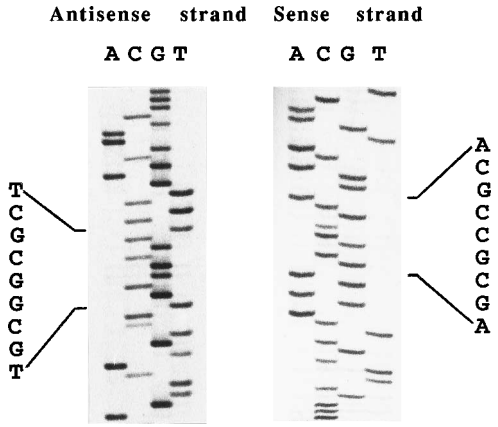

FIG. 2. Autoradiographs of sequencing gels showing antisense and sense strands for the region covering nt 1813 to 1821.

```
                10        20        30        40        50        60
AroK  MAEKRNIFLVGPMGAGKSTIGRQLAQQLNMEFYDSDQEIEKRTGADVGWVFDLEGEEGFR
      :..  .::.::.: ::.:.:..::. ::. : :.:: .... . .:. . . :. .:::
AroL  MTQ--PLFLIGPRGCGKTTVGMALADSLNRRFVDTDQWLQSQLNMTVAEIVEREEWAGFR
                10        20        30        40        50

              70        80        90       100       110       120
AroK  DREEKVINELTEKQGIVLATGGGSVKSRETRNRLSARGVVVYLETTIEKQLARTQRDKKR
      .::. .....:. ...:.::::: . .  .:. . ..:.:::: ....  ..: : . .
AroL  ARETAALEAVTA-PSTVIATGGGIILTEFNRHFMQNNGIVVYLCAPVSVLVNRLQAAPEE
      60        70        80        90       100       110

             130       140       150       160       170
AroK  PLLHVET--PPREVLEALANERNPLYEEIADVTIRTDDQSAKVVANQIIHMLESN
        : .. :  : .: .... .::.:: :.:...: .......:...    . ..
AroL  DLRPTLTGKPLSEEVQEVLEERDALYREVAHIIIDATNEPSQVISEIRSALAQTI
      120       130       140       150       160       170
```

FIG. 3. Alignment of the AroK and AroL proteins. The protein sequences were aligned with the FASTA programs of Lipman and Pearson (9). The putative site A of the ATP-binding site is underlined in both proteins.

common pathway of aromatic amino acid biosynthesis, such effects could confuse nutritional tests for an AroK phenotype. The tests were therefore carried out with strains carrying the *aroB351* mutation. The tests measured growth on minimal medium (16) supplemented with shikimic acid, phenylalanine, and tyrosine. On this medium, an *aroB aroL aroK*$^+$ strain will grow whereas an *aroB aroL aroK* strain will not.

Plasmid pMU3251 is derived from a low-copy-number cloning vector. It carries a DNA fragment which extends 48 bp beyond the putative stop codon of the Løbner-Olesen et al. (11) sequence but is 121 bp short of the putative stop codon for the corrected sequence. Plasmid pMU3255 was similarly derived and carries a DNA fragment which extends 115 bp beyond the putative stop codon for the corrected sequence. These plasmids were introduced into the strain JP7837 (*aroB351 aroL478*::Tn*10 aroK17*::Cam^r), and the resulting strains were tested for their growth on minimal medium as described above. Only the strain carrying the large *aroK* fragment on pMU3255 was able to grow on minimal medium supplemented with shikimic acid, phenylalanine, and tyrosine. These results provide further confirmation that the extended sequence is correct.

A comparison of the amino acid sequences of AroL and AroK. The amino acid sequence predicted from the corrected *aroK* sequence corresponds to a protein of 173 amino acids with a calculated molecular mass of 19,526 Da. This value is in agreement with the value of Ely and Pittard of 20,000 Da (5). The AroL protein is 173 amino acids (or 174, if one counts the first methionine which is apparently removed after translation) (4). The alignment of the two proteins is shown in Fig. 3, where it can be seen that they share 30% identity over their entire length. The putative site A of the ATP-binding site is underlined in both proteins and shows good agreement with the general consensus sequence described by Walker et al. (22). Some proteins which bind and hydrolyze ATP have a second conserved site (site B) about 90 (22) or 120 (17) amino acids

TABLE 1. Effects of mutations on β-galactosidase activity of *aroK-lacZ* translational fusions in strain JP8042

Plasmid	Mutation in *aroK-lacZ* fusion	β-Galactosidase activity[a]
pMU3258	None	2,089
pMU3259	TAT (1793)→TAA	2,066
pMU3260	ATG (1876)→ACG	1.5

[a] β-Galactosidase activities were measured by the method of Miller (14). Each value is the average of at least three independent assays. Strain JP8042 has the genotype Δ(*argF-lac*)*U169 tsx recA56 tyrR366 srl* IN(*rmD-rmE*)*1*.

further toward the carboxyl end of the protein. The major characteristic of this site is a string of four hydrophobic residues followed by an aspartate and preceded eight residues earlier by an arginine. Shikimate kinase II, the *aroL*-encoded protein, has the sequence R--------HIIID which is some 131 residues downstream of site A. No corresponding sequence is found in shikimate kinase I, the *aroK*-encoded protein. If this sequence is important for ATP hydrolysis, its absence in shikimate kinase I may contribute to its weaker activity as a shikimate kinase, in addition to its much lower affinity for shikimate (4). When the *aroL*- and *aroK*-encoded sequences were analyzed for the presence of an α-helix and β-sheet by the method of Chou and Fasman (1) and for hydrophobicity by the method of Kyte and Doolittle (7), they showed little structural similarity (data not shown). Their similar size and the 30% homology, on the other hand, suggest a common origin. The only phenotypic change detected in *aroK* mutants is that in the presence of an *aroL* mutation, the double mutant behaves as an aromatic amino acid auxotroph. Although it is still possible that *aroK* encodes a protein whose major function is unrelated to the biosynthesis of the aromatic amino acids, in the light of current evidence, it seems more likely that it is a gene coding for shikimate kinase whose importance to the organism has been displaced by the more catalytically active and finely regulated gene *aroL*. Although the DNA sequences upstream of the coding sequences of both genes show 40% homology, the three TyrR box sequences of *aroL* are disrupted or nonfunctional in *aroK* (data not shown).

ADDENDUM

During the course of this work, Martin et al. (12) reported on a homologous region of the chromosome in *Pseudomonas aeruginosa* and sequenced part of this organism's *aroK* gene. From these studies and an analysis of the published *E. coli* sequences for *aroK* and *aroL*, Mattick (12a) suggested the possibility of a number of frameshift errors in the published *aroK* sequence, which in fact corresponded to those identified during the course of this work.

The work described in this paper was supported by the Australian Research Council.

We thank S. Lepileo for technical assistance.

REFERENCES

1. **Chou, P. Y., and G. D. Fasman.** 1978. Empirical predictions of protein conformation. Annu. Rev. Biochem. **47:**251–276.
2. **DeFeyter, R. C., B. E. Davidson, and J. Pittard.** 1986. Nucleotide sequence of the transcription unit containing the *aroL* and *aroM* genes from *Escherichia coli* K-12. J. Bacteriol. **165:**233–239.
3. **DeFeyter, R. C., and J. Pittard.** 1986. Genetic and molecular analysis of *aroL*, the gene for shikimate kinase II in *Escherichia coli* K-12. J. Bacteriol. **165:**226–232.
4. **DeFeyter, R. C., and J. Pittard.** 1986. Purification and properties of shikimate kinase II from *Escherichia coli* K-12. J. Bacteriol. **165:**331–333.
5. **Ely, B., and J. Pittard.** 1979. Aromatic amino acid biosynthesis: regulation of shikimate kinase in *Escherichia coli* K-12. J. Bacteriol. **138:**933–943.
6. **Heatwole, V. M., and R. L. Somerville.** 1992. Synergism between the Trp repressor and Tyr repressor in repression of the *aroL* promoter of *Escherichia coli* K-12. J. Bacteriol. **174:**331–335.
7. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.
8. **Lawley, B., and A. J. Pittard.** 1994. Regulation of *aroL* expression by TyrR protein and Trp repressor in *Escherichia coli* K-12. J. Bacteriol. **176:**6921–6930.
9. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227:**1435–1441.
10. **Løbner-Olesen, A., E. Boye, and M. G. Marinus.** 1992. Expression of the *Escherichia coli dam* gene. Mol. Microbiol. **6:**1841–1851.
11. **Løbner-Olesen, A., and M. G. Marinus.** 1992. Identification of the gene (*aroK*) encoding shikimate acid kinase I of *Escherichia coli*. J. Bacteriol. **174:**525–529.
12. **Martin, P. R., M. Hobbs, P. D. Free, Y. Jeske, and J. S. Mattick.** 1993. Characterization of *pilQ*, a new gene required for the biogenesis of type 4 fimbriae in *Pseudomonas aeruginosa*. Mol. Microbiol. **9:**857–868.

12a. **Mattick, J. S.** Personal communication.

13. **Millar, G., and J. R. Coggins.** 1986. The complete amino acid sequence of 3-dehydroquinate synthase of *Escherichia coli* K12. FEBS Lett. **200:**11–17.
14. **Miller, J. H.** 1972. Experiments in molecular genetics, p. 352–355. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
15. **Mills, D. R., and F. R. Kramer.** 1979. Structure-independent nucleotide sequence analysis. Proc. Natl. Acad. Sci. USA **76:**2232–2235.
16. **Monod, J., G. Cohen-Bazire, and M. Cohen.** 1951. Sur la biosynthèse de la β-galactosidase (lactase) chez Escherichia coli. La specificité de l'induction. Biochim. Biophys. Acta **7:**585–599.
17. **Poole, R. K., L. Hatch, M. W. J. Cleeter, F. Gibson, G. Cox, and G. Wu.** 1993. Cytochrome *bd* biosynthesis in *Escherichia coli*: the sequences of the *cydC* and *cydD* genes suggest that they encode the components of an ABC membrane transporter. Mol. Microbiol. **10:**421–430.
18. **Praszkier, J., P. Bird, S. Nikoletti, and J. Pittard.** 1989. Role of countertranscript RNA in the copy number control system of an IncB miniplasmid. J. Bacteriol. **171:**5056–5064.
19. **Ringquist, S., S. Shinedling, D. Barrick, L. Green, J. Binkley, G. D. Stormo, and L. Gold.** 1992. Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site. Mol. Microbiol. **6:**1219–1229.
20. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
21. **Slauch, J. M., and T. J. Silhavy.** 1991. *cis*-acting *ompF* mutations that result in *ompR*-dependent constitutive expression. J. Bacteriol. **173:**4039–4048.
22. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay.** 1982. Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1:**945–951.