

NOTES

A Cluster of Three Genes (*dapA*, *orf2*, and *dapB*) of *Brevibacterium lactofermentum* Encodes Dihydrodipicolinate Synthase, Dihydrodipicolinate Reductase, and a Third Polypeptide of Unknown Function

AGUSTÍN PISABARRO, MARCOS MALUMBRES, LUIS M. MATEOS, JOSÉ A. OGUIZA,
AND JUAN F. MARTÍN*

*Area of Microbiology, Department of Ecology, Genetics and Microbiology, Faculty of Biology,
University of León, 24071 León, Spain*

Received 10 December 1992/Accepted 18 February 1993

The *dapA* and *dapB* genes, encoding, respectively, dihydrodipicolinate synthase and dihydrodipicolinate reductase, the two first enzymes of the lysine branch of the aspartic amino acid family, were cloned from the DNA of the amino acid-producing bacterium *Brevibacterium lactofermentum*. The two genes were clustered in a 3.5-kb *Sau3AI*-*Bam*HI fragment but were separated by an open reading frame of 750 nucleotides. The protein encoded by this open reading frame had little similarity to any protein in the data banks, and its function remains unknown. The three genes were translated in *Escherichia coli*, giving the corresponding polypeptides.

The lysine biosynthetic pathway in corynebacteria is still poorly understood despite the use of *Corynebacterium glutamicum* and *Brevibacterium lactofermentum* for the industrial production of lysine. This pathway shares the first two steps with the threonine and methionine biosynthetic pathways, which involve the activation of aspartic acid to aspartyl phosphate and the reduction of the activated amino acid to aspartyl semialdehyde. In the lysine branch, the aspartyl semialdehyde is converted to dihydrodipicolinate by condensation with pyruvate. This first "branching" reaction, which involves the condensation of a C₄ with a C₃ fragment to form the seven-carbon dihydrodipicolinate, is carried out by dihydrodipicolinate synthase, encoded in *Escherichia coli* by the *dapA* gene. The condensation step is followed by a cyclization reaction which results in the formation of 2,3-dihydrodipicolinate and may involve an enzyme-bound C₇ linear intermediate (Fig. 1). Whether this reaction requires more than one protein is uncertain. The dihydrodipicolinate formed is later converted to Δ¹-piperidine-2,6-dicarboxylate by dihydrodipicolinate reductase, which in *E. coli* is encoded by the *dapB* gene. In the last steps of the pathway the Δ¹-piperidine-2,6-dicarboxylate is converted to D,L-diaminopimelic acid by at least two alternative pathways (6, 20). In addition to the succinylase and acetylase variants common to other procaryotes, a shunt pathway in which Δ¹-piperidine-2,6-dicarboxylate is converted in a single step to diaminopimelate by the enzyme D-diaminopimelate dehydrogenase has been described in corynebacteria (11). Finally, diaminopimelic acid is decarboxylated to form lysine by diaminopimelate decarboxylase (*lysA* gene).

Only preliminary information is available on the nature of the *dapA* and *dapB* genes of *C. glutamicum* (1, 5), and no information has been published on that in *B. lactofermen-*

tum. We describe here that the *dapA* and *dapB* genes of *B. lactofermentum* are clustered together with a third open reading frame (ORF) located between them. The three ORFs are translated into polypeptides.

Cloning the *dapB* gene and deletion mapping. Total DNA of *B. lactofermentum* ATCC 13869 (Table 1) was isolated as described previously (7) and partially digested with *Sau3AI* or totally digested with *Hind*III in different experiments. DNA fragments of 4 to 10 kb were ligated with pUC13, and the gene library was amplified in *E. coli* DH5α. Plasmid DNA from *E. coli* was used to transform competent cells (4) of *E. coli* CGSC 4549 *dapB* (Table 1). Several clones were found that complemented the *dapB* auxotrophy of *E. coli* CGSC 4549 in VB minimal medium (21) supplemented with thiamine (3 mM), *m*-diaminopimelic acid (0.1 mM), and ampicillin (50 μg/ml). Rapid screening of plasmid DNA (9) revealed the presence in all prototrophic transformants of an 8.2-kb plasmid, pULDAP1, containing a *Sau3AI* insert of 5.5 kb (Fig. 2) or a 9.0-kb plasmid, pULDAP2, carrying a 6.3-kb *Hind*III insert. Both plasmids were used to retransform *E. coli* CGSC 4549, and complementation of *dapB* mutation was confirmed.

Hybridization analysis of the *B. lactofermentum* chromosome region carrying the *dapA-dapB* cluster was carried out as described by Southern (22) with the modifications introduced by Hopwood et al. (10). pULDAP2 was labelled with [³²P]CTP by nick translation. The hybridizing bands in total DNA corresponded to the expected sizes by the restriction map of pULDAP2: *Hind*III (6.3 kb), *Pst*I (2.4 and 2.7 kb), and *Bam*HI (about 10 kb; data not shown). Hybridization with total *C. glutamicum* DNA showed that both corynebacteria differ in their *Hind*III or *Pst*I restriction fragments.

An internal 3.5-kb *Bam*HI fragment (one of the *Bam*HI sites was constructed, during ligation, from a *Sau3AI* site) was isolated from pULDAP1 and subcloned into pUC119 in

* Corresponding author.

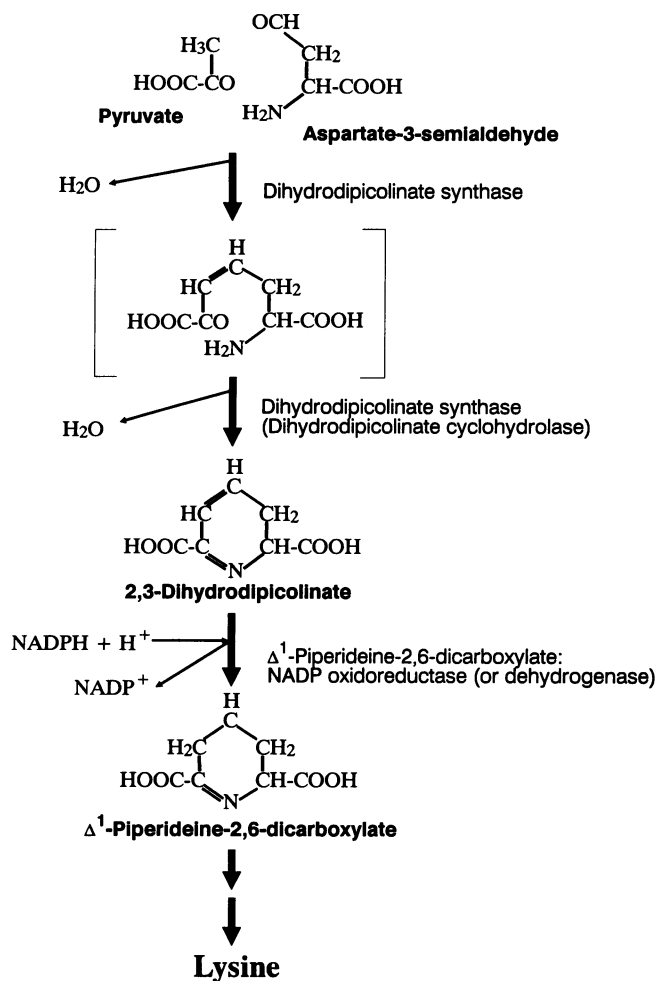


FIG. 1. Conversion of aspartyl semialdehyde into 2,3-dihydrodipicolinate and Δ^1 -piperideine-2,6-dicarboxylate by the dihydrodipicolinate synthase (*dapA*) and dihydrodipicolinate reductase (*dapB*) enzymes.

both orientations, giving rise to plasmids pULAP301 and pULAP302 (Fig. 2). Both of these plasmids complemented the *dapB* auxotrophy of *E. coli* CGSC 4549. Additional deletion mapping studies (Fig. 2) showed that the *dapB* gene was located in the 1.9-kb *Cla*I fragment (pULAP05) internal

to the initial inserts of *B. lactofermentum* DNA cloned in pULDAP1 and pULDAP2.

Complementation studies showing linkage of *dapB* and *dapA*. Since some of the genes of the lysine biosynthetic pathway are clustered in certain microorganisms, several *E. coli* lysine auxotrophs blocked in different steps of the pathway (Table 1) were transformed with plasmids pULDAP1 and pULDAP2. Three different *dapA* mutants (*E. coli* CGSC 4547, CGSC 4548, and RDA8) were complemented by pULDAP1, but no complementation was observed upon transformations with pULDAP2. Deletion mapping studies for *dapA* complementation lead to the conclusion that the *dapA* gene of *B. lactofermentum* is linked to the *dapB* gene, as in an initial report describing *C. glutamicum* (5). This result contrasts with the location in *E. coli* where *dapA* (min 53) maps opposite to the *dapB* gene (min 0.5) on the *E. coli* chromosome (3). Linkage of amino acid biosynthetic genes is known in some gram-positive bacteria (14) and may explain the more efficient biosynthesis of lysine by the corynebacteria. Since both *dapA* and *dapB* genes are expressed in both orientations, the presence of promoter sequences in each of them is likely. Lysine auxotrophs of *E. coli* which were altered in other genes of the lysine pathway (Table 1) were not complemented by pULDAP1 or pULDAP2.

Three ORFs exist in the *dapA-dapB* cluster. A 3.5-kb *Sau*3AI-*Bam*HI fragment containing the *dapA* and *dapB* genes (Fig. 2) was used for sequence analysis in plasmids pUC119 and pUC118. Single-stranded plasmid DNA was isolated after transformation of *E. coli* WK6*mutS* (Table 1) with the recombinant phagemids and the helper virus M13K07 (16). Sequencing was performed by the dideoxy method (19), using the Sequenase system (United States Biochemicals Co.) and the TaqTrack Sequencing system (Promega). α -³⁵S-dATP (600 Ci mmol⁻¹) was purchased from Amersham Corp.

Three ORFs were found within the DNA fragment (Fig. 3). ORF1 extends from nucleotides 729 to 1472 and corresponds to the DNA fragment that complements the *dapB* auxotrophy in *E. coli* by deletion mapping studies. It encodes a protein of 248 amino acids with a deduced *M_r* of 26,055. ORF3 is located in the DNA fragments that complement the *dapA* mutations in *E. coli*; it extends from nucleotides 2480 to 3382 and encodes a protein of 301 amino acids with a deduced *M_r* of 31,310. An additional ORF, ORF2, has been found between the *dapB* and *dapA* genes. One methionine and four valines are present in the 5'-terminal region of ORF2, any of which may be translation initiation sites, and

TABLE 1. *E. coli* strains used in this work

Strain ^a	Genotype	Source
DH5 α	F ⁻ <i>recA1 endA1 gyrA96 thi-1 hsdR17</i> ($r_K^- m_K^+$) <i>sup44 relA1</i> $\lambda^- \phi 80dlacZ\Delta M15 \Delta(lacZYA-argF)U169$	Laboratory stock
WK6 <i>mutS</i>	$\Delta(lac-proAB)$ <i>galE strA mutS215::Tn10/F' lacI^qZ\Delta M15 proA⁺B⁺</i>	R. Zell
K38	HfrC (λ)	S. Tabor
CGSC 4549	<i>thi-1 relA1 dapB17::Mu</i> $\lambda^- spoT1$	B. Bachmann
CGSC 4548	<i>thi-1 relA1 spoT1 dapA16</i>	B. Bachmann
CGSC 4547 ^b	<i>thi-1 relA11 dapA</i> $\lambda^- spoT1$	B. Bachmann
RDA8	<i>araD139 lacU169 rpsL thi-A dapA::Mu</i>	C. Richaud
CGSC 4557	<i>dapD8 relA1 spoT1 thi-1</i> λ^-	B. Bachmann
CGSC 5080	<i>hisA323 \Delta(bioH-asd)29</i> λ^-	B. Bachmann
RLA11	<i>araD139 \Delta(lacIPOZYA)U169 rpsL thiA lysA::Mu</i>	C. Richaud

^a CGSC, *E. coli* General Stock Center.

^b *E. coli* CGSC 4547 was described previously as a *dapC* mutant.

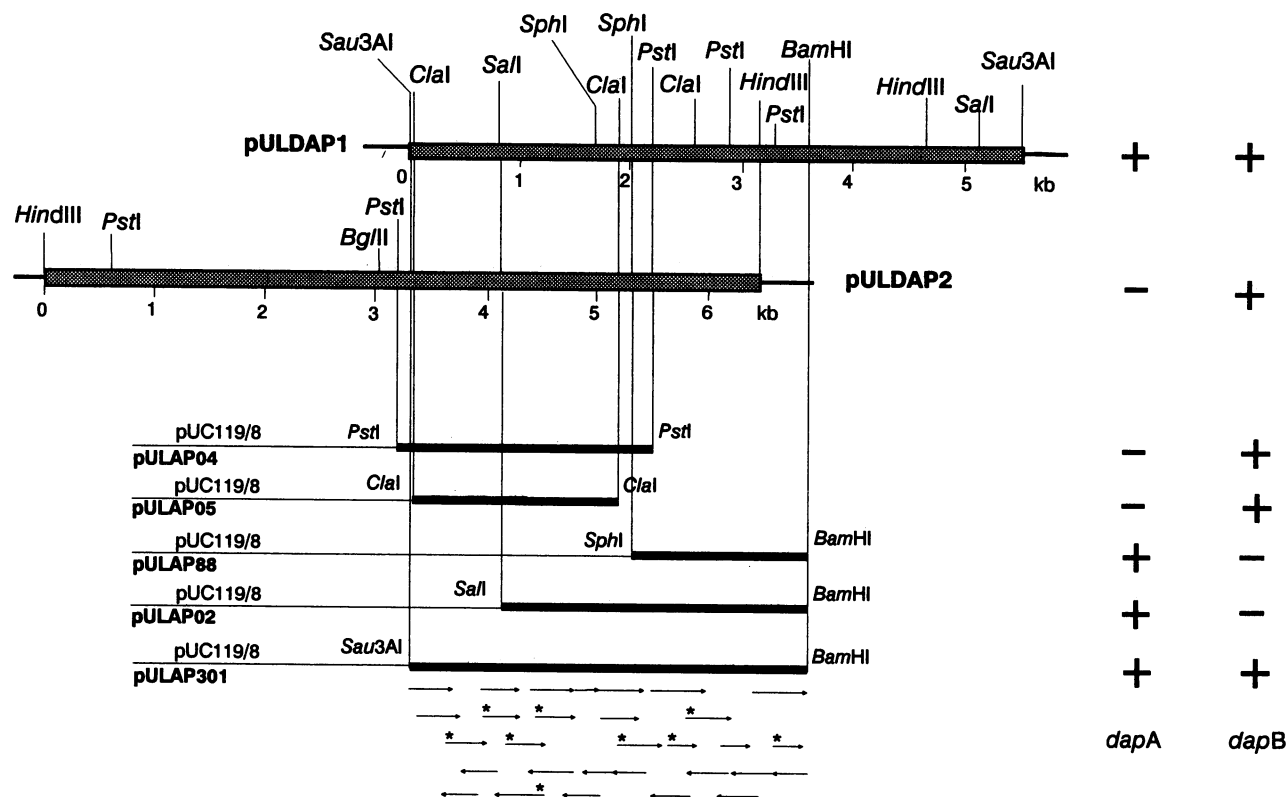


FIG. 2. Restriction and deletion mapping of pULDAP1 and pULDAP2 and subcloning of internal fragments. Complementation by the different plasmids of *dapA* and *dapB* mutations was carried out in *E. coli* mutants (Table 1). The DNA insert of pULAP301 and pULAP302 was trimmed down by restriction endonucleases or unidirectional deletions (*) (Erase-a-Base system, Promega), and the sequencing strategy is shown. pULAP302 contains the same insert as pULAP301 but in the opposite orientation.

the ORF2 extends to nucleotide 2276 (Fig. 3). Assuming the first valine (position 1527) to be the N-terminal amino acid, ORF2 encodes a putative protein of 250 amino acids corresponding to an M_r of about 27,962, whereas if the methionine (position 1719) is considered, a protein of 186 amino acids and an M_r of 21,008 would be expected. Each ORF is preceded by a standard GAAGG ribosome-binding sequence which occurs in most genes of corynebacteria (15).

Analysis of the protein-coding regions by the Gribskov algorithm (8) revealed that ORF1 displays the normal codon usage found in corynebacteria (14a, 15), whereas some rare codons were found in ORF2 and ORF3. Comparison of the codon usage for the three sequences shows differences between *dapB* and the other genes. Thus, *dapB* does not use AGA, AGG, and CGA (Ser), GGG (Gly), AUA (Ile), and UUA (Leu), which are rare codons in corynebacteria (14a). However, ORF2 and *dapA* use most of these rare codons. In addition, ORF2 and *dapA* show a preference for other codons that are poorly used in most corynebacterial genes: UGU (Cys), GGA (Gly), UAU (Tyr), and GUA (Val). The G+C content of the three ORFs fits the G+C content of the chromosome of *B. lactofermentum* (0.54 mol%). The G+C content of the third position of the codons (GC3s) is lower in ORF2 and *dapA* than the average GC3s of corynebacteria (between 0.5 and 0.7 mol%).

Analogy of the *B. lactofermentum* dihydrodipicolinate synthase and dihydrodipicolinate reductase with other proteins. The deduced amino acid sequence of *B. lactofermentum* dihydrodipicolinate synthase (EC 4.2.1.52) is very similar to

that of the corresponding *E. coli* enzyme, which contains 272 amino acids and an M_r of 31,375 (32.6% identical residues) (Fig. 4). It is interesting to note the high amino acid similarity despite a lower similarity in the nucleotide sequences, due to the low G+C content of *E. coli*. The dihydrodipicolinate synthase is also similar to the two reported isoenzymes of *Triticum* sp. (26.4 and 27% identical amino acid residues). The deduced enzyme of *B. lactofermentum* is very similar to that of *C. glutamicum* since the only two nucleotide changes alter a single amino acid.

Analysis of similarity to other proteins of the PIR data bank revealed a high similarity (28.9% identical amino acids) to the *N*-acetylneuraminic lyase of *E. coli* (encoded by the *nalE* gene [13, 17]), an enzyme which recognizes pyruvate as substrate, as occurs with dihydrodipicolinate synthase, and condenses it with a different substrate: *N*-acetylmannosamine. The similarity between this enzyme and the dihydrodipicolinate synthase is striking, since seven different amino acid stretches in the protein (*a* to *g* in Fig. 4) are conserved. This probably indicates a conservation of the mechanism of biosynthesis of dihydrodipicolinate synthase with reactions which use pyruvate to form C-C bonds through C-3 of this keto acid.

Dihydrodipicolinate synthase condenses pyruvate and aspartyl semialdehyde to form the seven-carbon compound 2,3-dihydrodipicolinate. The reaction may proceed with the formation of a linear 6-keto amino acid intermediate, which is cyclized to 2,3-dihydrodipicolinate by the dihydrodipicolinate cyclohydrolase (Fig. 1). Yugari and Gilvarg (25) indi-

BamHI
GGATCCCAATCGATACCTGGAACGACAACCTGATCAGGATATCCAATGCCTTGAATATTGACGTTGAGGAAGGAATCACCAGCCATCTCAACTGGAAGA 100
CCTGACGCCTGCTGAATTGGATCAGTGGCCCAATCGACCACCAACAGGTTGGCTATTACCGGGGATATCAAAAACAACTCGCGTGAACGTTTCGTGCT 200
CGGCAACCGGATGCCAGCGATCGACATATCGGAGTCACCAACTTGAGCCTGCTGCTTCTGATCCATCGACGGGAACCAACGGCGGCAAGCAGTGGG 300
GGAAGGGGAGTTGGTGGACTCTGAACAGTGGGCTCTGAAGTGGTAGGCGACGGGCGAGCATCTGAAGCGTGCAGTGTGGTACCAGGTTAGCGGTT 400
TCAGTTTCTGTACAACTGGAGCAGGACTAGCAGAGGTTGTAGCGGTTGAGCGCTTCCATCACAAGCACTTAAAAGTAAAGAGGCGGAAACCACAAGCGC 500
CAAGGAACTACCTGCGGAACGGGCGGTGAAGGGCAACTTAAGTCTCATATTTCAAACATAGTTCACCTGTGTGATTAATCTCCAGAACGGAACAAACTG 600
ATGAACAATCGTTAAACAACAGACCAAACGGTCAGTTAGGTATGGATATCAGCACCTTCGAATGGGTACGCTAGACTGGTGGGCGTTTGAAAAACT 700
CTTCGCCCCACGAAATGAAAGGAGCATAATGGGAATCAAGGTTGGCGTCTCGGAGCCAAAGCGGTGTGGTCAAATATTGTGGCAGCAGTCAATGAG 800
SD m g i k v g v l g a k g r v g q t i v a a v n e
TCCGACGATCTGGAGCTTGTTCAGAGATCGGCGTCGACGATGATTGAGCCTCTGTAGTACAAACGGGCGTGAAGTGTGCTGACTTACCACCTCCTA 900
s d d l e l v a e i g v d d d l s l l v d n g a e v v v d f t t p n
ACGCTGTGATGGGAACCTGGAGTCTGCATCAACACGGCATTCTTCGGGTTGTGGAAACACGGGCTTCGATGATGCTCGTTTGGAGCAGGTTCCGCGC 1000
a v m g n l e f c i n g i s a v v g t t g f d d a r l e q v r a
CTGGCTTGAAGGAAAAGACAATGCGGTGTTCTGATCGCACCTAACTTTGCTATCTCTCGGTGTTGACCATGGTCTTTTCCAAGCAGGCTGCCGCTTC 1100
w l e g k d n v g v l i a p n f a i s a v l t m v f s k q a a r f
TTCGAATCAGCTGAAGTATTGAGCTGCACCACCCCAACAGCTGGATGCACCTTCAGCACCAGGATCCACACTGCTCAGGGCATTGCTCGGGCAGCA 1200
f e s a e v i e l h h p n k l d a p s g t a i h t a q g i a a a r k
AAGAAGCAGGATGGACGCACGACAGATGCGACCGAGCAGCCTTGAGGTTCCCGTGGCGCAAGCGTAGATGGAATCCAGTTCACGCAGTCCGCA 1300
e a g m d a q p d a t e g a l e g s r g a s v d g i p v h a v r m
GTCCGCGATGGTTCACGAGCAAGTTATCTTTGGCACCCAGGCTCAGACCTTGACCATCAAGCAGGACTCCTATGATCGCAACTCATTTGACCAGGT 1400
s g m v a h e q v i f g t q g q t l t i k q d s y d r n s f a p g
GTCTGGTGGTGTGCGCAACATTCACAGCACCAGGCTAGTCTGAGACTTGAACATTACCTAGGCTGTAAGGCTCATTTTCAGCAGCGGGTGGAA 1500
v l v g v r n i a q h p g l v v g l e h y l g l .
TTTTTTAAAGGAGCGTTTAAAGGCTGTGGCCGAACAAGTTAAATGAGCGTGGAGTTGATAGCGTGCAGTCTTTACTCCACCGCTGATGTTGAGTG 1600
SD v a e q v k l s v l i a c s s f t p p a d v e w
GTCAACTGATGTTGAGGGCGCGAAGCACTCGTCGAGTTTGGGGTTCGCTGCTACGAACTTTTGATAAGCCGAACCCCGAACTGCTTCCAATGCT 1700
s t d v e g a e a l v e f a g r a c y e t f d k p n p r t a s h a
CGGTATCTCGCCACATCATGAAGTGGGGCACACTGCTTGGCTTGGATGCAATGCCAGATGATATCCGAGGCATTCTCGGTCGCGGACCCATG 1800
a y l r h i m e v g h t a l l e h a n a t m y i r g i s r s a t h e
AATGGTCCGACACCGCCATTTTTCCTCTCAACTGCTCAGCCTTTCGTGCACAGCGGAGAATCGGAAGTAGTGGTGCACCTCTCATCGATGAAGA 1900
l v r h r h f s f s q l s q r f v h s g e s e v v v p t l i d e d
TCCGAGTTCGGTGAACCTTTCAATGACCGCATGGATGAGTCTCGGTTGCTTCAATGAGCTGTTAATGCGTGGAAAGAAAACCTGGCGATGAACCG 2000
p q l r e l f m h a m d e s r f a f n e l l n a l e e k l g d e p
AATGCACTTTAAAGAAAAGCAGGCTCGTCAAGCAGCTCGCGTGTGTCGCCAACCGTACAGATCCAGAACTCGTGGTCTGGAACCTTCGCACT 2100
h a l l r k k q a r a v l g h a t e s r i v s g h f r t w
GGAGGCATTTTCAATGGCATGCGAGCCAGTGAACATGCAGACGTCGAAATCCCGGAAGTAGCGGTAGGATGTTAAGAAAGCTGCAGGTAGCAGCGCAAC 2200
r h f i g m r a s e h a d v e i r e v a v g c l r k l q v a a p t
TGTTTTCGGTGATTTGAGATTGAAACTTTGGCAGACGGATCGCAAATGGCAACAGCCCGTATGTCATGGACTTTTAAACGAAAGCTCACACCCACGAG 2300
v f g d f e i e t l a d g s q m a t s p y v m d f .
CTAAAAATTCATATAGTTAAGACAACATTTTGGCTGTAAGACAGCCGTAACCACTTCTTGCTCATGTCAATTGTTCTTATCGGAATGTGGCTTGGC 2400
GATTGTTATGAAAAGTGTGTAGGTTTTTGGCGGGTGTTTAACCCCAATGAGGAAAGGTAACCTTGAACCTATGAGCAGAGTTTAAACAGCT 2500
SD m s t g l t a
AAGACCGGAGTAGAGCACTTCGGCACCGTTGGAGTAGCAATGGTTACTCCATTACCGGAATCCGGAGACATCGATATCGCTGCTGGCCGGAAGTCCGCG 2600
k t g v e h f g t v g v a m v t p f t e s g d i d i a a g r e v a a
CTTATTTGGTGTGATAAGGGCTTGGATCTTTGGTTCGCGGGCACCCTGGTGAATCCCAACGCAACCCGCGCTGAAAAACTAGAACTGCTCAAGG 2700
y l v l d k g l a g l a g t g e s p t t t a a e k l e l k a
CGTTCGTGAGGAAGTTGGGATCGGGCAAGCTCATCGCGGTGTGCGAACAACACCGGGACATCTGTGGAACCTGCGGAAGCTGCTGCTTCTGCT 2800
v r e e v g d r a k l i a g v g t n n t r t s v e l a e a a a s a
GGCGCAGCGGCTTTTAGTTGTAACCTTATTACTCCAAGCCGACCAAGAGGATGCTGGCGCACTCGGTGCAATTGCTGCAGCAACAGAGGTTTC 2900
g a d g l l v v t p y y s k p s q e g l l a h f g a i a a a t e v p
CAATTTGCTCTATGACATTCCTGGTCCGTCAGGATTTCCAATGAGTCTGATACCATGAGACCGCTGAGTGAATTACCTACGATTTTGGCGGTCAAGGA 3000
i c l y d i p g r s g i p i e s d t m r r l s e l p t i l a v k d
CGCCAAGGGTGACCTCGTTGCAGCCAGTCAATGATCAAGAAACGGGACTTGCCCTGGTATTACAGCGATGACCCACTAAACCTTGTTTGGCTTCTTTG 3100
a k g d l v a a t s l i k e t g l a w y s g d d p l n l v w l a l
GGCGGATCAGGTTTCATTTCCGTAATGGACATGCAGCCCCACAGCATTAAGTGAAGTTGACACAAGCTTCGAGGAAGGCGACCTCGTCCGTGGCGGG 3200
g g s g f i s v i g h a a p t a c a r e l y t s f e e g d l v r a r e
AAATCAACGCAAACTATCACCGCTGGTACCGTCCCAAGTCCGCTGGTGGAGTCAAGTGGCAAAAGCTGCTTGGCTGACAGGCAACAGTATGAG 3300
i n a k l s p l v a a q g r l g g v s l a k a a l r l q g i n v g
AGATCTCGACTTCCAATTAAGCTCCAATGAGCAGAACTTGGGCTCTCCGAGAAGACATGAAAAAGCTGGAGTTCTATAAATATGAATGATCCC 3400
d p r l p i m a p n e q e l e a l r e d m k k a g v l .
GAAATCGCGCCGGAAGGTTACCCGCAAGCGGGCCACCAGAAGCTGGTCAGGAAAACATCTGGATACCCCTGTCTTCAGGCACCAGATGCTTCCTC 3500
TAACGAGAGCGCTGTAAGGCTGAGACCGCGGAAACGACAATCGGGATCGTGGCAAGTGTCAAGGATCC 3573
BamHI

ORF1

ORF2

ORF3

FIG. 3. Nucleotide and deduced amino acid sequences of the *Sau3AI-BamHI* insert of pULAP301. The amino acid sequence deduced from ORF2 is shown from the first valine. The Shine-Dalgarno (SD) sequences are underlined. The nucleotide sequences of the three ORFs have been deposited in the EMBL Data Library under accession number Z21502.

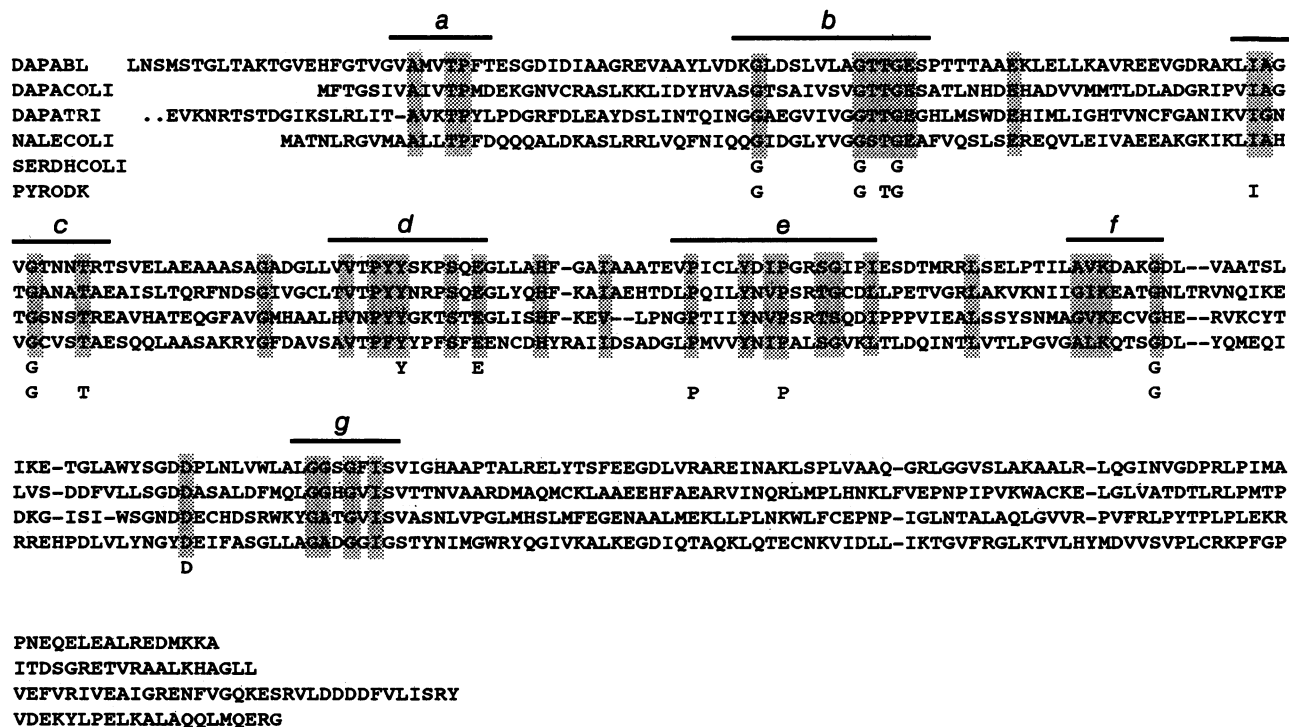


FIG. 4. Amino acid sequence comparison of the dihydrodipicolinate synthase of *B. lactofermentum* (DAPABL) to other dihydrodipicolinate synthases of *E. coli* (DAPACOLI [18]) and *Triticum* sp. (DAPATRI [12]) and *N*-acetylneuraminic acid lyase of *E. coli* (NALECOLI [13, 17]). *a* to *g* indicate the conserved domains. The amino acid sequences of the serine dehydrogenase of *E. coli* and the pyruvate-orthophosphate dikinase of *Clostridium symbiosum* (PIR data bank) were compared, and only the conserved amino acids of the domains are shown. Identical or functionally related amino acids are shaded. The families of functionally homologous amino acids were as follows: K-R, T-S, G-A, I-L-V-M, Y-F, and D-E.

cated that it was not clear whether a single enzyme carries out the carbon-carbon bond formation and the elimination of the oxygen at C-6 by an internal condensation. The molecular mechanism of *N*-acetylneuraminic acid synthesis (or hydrolysis) is extremely similar to that of dihydrodipicolinate synthase in that it involves condensation of the C-3 of pyruvate with the aldehyde group at C-1 of *N*-acetylmannosamine. Both genes are evolutionarily related and might have some common activities. A comparison of the amino acids conserved in the two types of enzymes (*N*-acetylneuraminic acid lyases and dihydrodipicolinate synthases) suggests that seven specific domains (*a* to *g* in Fig. 4) are probably involved in pyruvate recognition and condensation.

Several pyruvate-recognizing enzymes have been sequenced. These all show some degree of amino acid similarity in domains *b* and *c*, but no sequence of any other enzyme that carries out a C-C condensation involving pyruvate, similar to dihydrodipicolinate synthase, is available. The nucleotide sequence of the *dapA* gene (but not the *dapB* gene) of *C. glutamicum* was reported (1), and it differs from the *B. lactofermentum* gene in two nucleotides at positions 3275 to 3276.

Less information is available about dihydrodipicolinate reductases (EC 1.3.1.26), encoded by the *dapB* gene. Only the *E. coli* gene has been reported (2), and no similar genes have been sequenced from corynebacteria or other gram-positive bacteria. The *dapB* gene of *E. coli* encodes a protein of 273 amino acids with a deduced M_r of 28,798 and showed 28.1% identical residues with the *B. lactofermentum* enzyme. Several other proteins showed a consensus NADH

binding region with the dihydrodipicolinate reductase (see below), but little overall similarity. A molecular weight of 34,000 was proposed for the dihydrodipicolinate reductase of *C. glutamicum*, but this was not confirmed by sequencing the gene (5). The enzyme is strongly similar to that of *E. coli* and shows a domain which is conserved in several dehydrogenases, including the aspartic semialdehyde dehydrogenases (*asd*) of *E. coli*, *Saccharomyces cerevisiae*, and *C. glutamicum*; *N*-acetylglutamyl phosphate reductases (*argC*) of *Streptomyces clavuligerus*, *E. coli*, and *Bacillus subtilis*; glyceraldehyde-3-phosphate dehydrogenases of *C. glutamicum*, white mustard, and *Mesembryanthemum crystallinum*; and malate dehydrogenases of *E. coli* and pig mitochondria.

By comparing the first domain of the dihydrodipicolinate reductase with the conserved region in a variety of dehydrogenases, it is possible to propose the conservation of the sequence, (V/I)(A/G)(V/I)XGXXGXXG, that is maintained in all of these dehydrogenases. NADP-dependent oxidoreductases constitute a large collection of enzymes that vary widely in size and substrate specificity. All of them have a similar β - α - β fold which forms a pocket to accommodate the ADP moiety of the coenzyme (24). This sequence is universally conserved in different NADP-dependent dehydrogenases, including dihydrodipicolinate reductase.

No homology was found when the deduced protein encoded by ORF2 was compared with the PIR data bank.

Three proteins are formed in *E. coli* from the *dapA*-ORF2-*dapB* cluster. To establish whether ORF2 was actually being translated into proteins, the three ORFs were expressed in

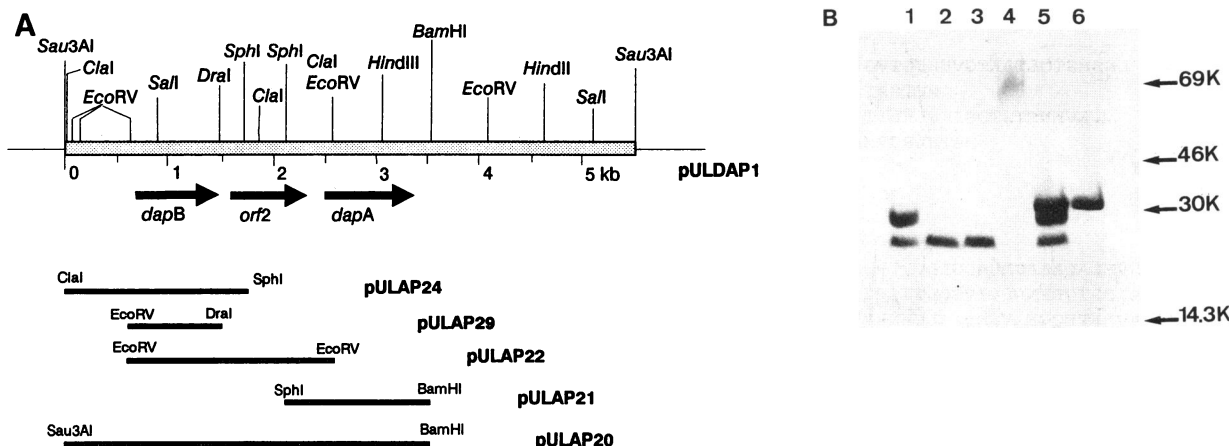


FIG. 5. Expression of *B. lactofermentum* cloned genes in *E. coli*. (A) Physical map of pULDAP1 and subcloning of several restriction fragments into the pT7 plasmids (23). (B) Autoradiograph of a sodium dodecyl sulfate-polyacrylamide gel with cell extracts of *E. coli* K38 cultures carrying different plasmids: lane 1, pULAP22; lane 2, pULAP24; lane 3, pULAP29; lane 4, control pT7-5 without insert; lane 5, pULAP20; lane 6, pULAP21.

E. coli from the T7 phage promoter by the T7 RNA polymerase (23). Several constructions carrying *dapA*, *dapB*, ORF2, and combinations of them were cloned downstream of the strong T7 ϕ 10 promoter in pT7-5 and pT7-6 (23).

The corresponding α -³⁵S-methionine-labelled proteins were observed in sodium dodecyl sulfate-polyacrylamide gels (Fig. 5). No proteins were obtained in control clones transformed with the pT7-5 plasmid without insert. Three proteins were observed in *E. coli* K38(pGP1-2) cells transformed with pT7-5 plus the 3.4-kb *Bam*HI insert (pULAP20) harboring the three ORFs. Two constructions (pULAP24 and pULAP29) were translated into a protein of about 26 kDa, corresponding exactly to the dihydrodipicolinate reductase. pULAP21, which carried the *dapA* gene exclusively, showed a single band of 31 kDa, corresponding exactly to the expected molecular weight of the dihydrodipicolinate synthase. Construction of pULAP22 carrying ORF2, in addition to the *dapB* gene, produced a protein of 29 kDa that seems to originate from ORF2 and also the 31-kDa dihydrodipicolinate synthase.

The presence of a third ORF (ORF2) between *dapA* and *dapB* was unexpected. Expression studies in *E. coli* showed that ORF2 appeared to encode a protein of about 29 kDa. This result agrees with the deduced protein of ORF2 (28 kDa) when we used the first valine (position 1527 [Fig. 3]) as the first translated amino acid. A study of codon usage in this ORF showed some codons which are used at low frequency in corynebacteria. However, codon usage of ORF2 is to a large extent similar to that of *dapA*, and it is efficiently translated in *E. coli* with the T7 phage promoter.

It is likely that ORF2 is also translated in corynebacteria; studies are in progress to establish whether ORF2 plays a role in lysine biosynthesis. It is tempting to speculate that ORF2 may encode a polypeptide involved in cyclization of the linear seven-carbon 6-keto amino acid. Such an activity would not be required for *N*-acetylneuraminic acid synthesis since in that case the condensation product is not cyclized. This hypothesis is consistent with the fact that *nalE*- and *dapA*-encoded polypeptides are very similar, which suggests that a separate cyclizing polypeptide may be required in the formation of dihydrodipicolinate but not in the synthesis of *N*-acetylneuraminic acid.

This work was supported by grants from the CE, Brussels, Belgium (BIOT-CT91-0264 RZJE), and the CICYT, Madrid, Spain (BIO92-0708). M. Malumbres and J. A. Oguiza were supported by fellowships from the University of Le3n and the Ministry of Education and Science.

REFERENCES

- Bonnassie, S., J. Oreglia, and A. M. Sicard. 1990. Nucleotide sequence of the *dapA* gene from *Corynebacterium glutamicum*. *Nucleic Acids Res.* **18**:6421.
- Bouvier, J., C. Richaud, F. Richaud, J. C. Patte, and P. Stragier. 1984. Nucleotide sequence and expression of the *Escherichia coli* *dapB* gene. *J. Biol. Chem.* **259**:14829-14834.
- Bukhari, A., and A. L. Taylor. 1971. Genetic analysis of diaminopimelic acid- and lysine-requiring mutants of *Escherichia coli*. *J. Bacteriol.* **105**:844-854.
- Chung, C. T., S. L. Niemela, and R. H. Miller. 1989. One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. USA* **86**:2172-2175.
- Cremer, J., L. Eggeling, and H. Sahm. 1990. Cloning the *dapA-dapB* cluster of the lysine-secreting bacterium *Corynebacterium glutamicum*. *Mol. Gen. Genet.* **220**:478-480.
- Cremer, J., L. Eggeling, and H. Sahm. 1991. Control of the lysine biosynthesis sequence in *Corynebacterium glutamicum* as analyzed by overexpression of the individual corresponding genes. *Appl. Environ. Microbiol.* **57**:1746-1752.
- del Real, G., A. Aguilar, and J. F. Mart3n. 1985. Cloning and expression of tryptophan genes from *Brevibacterium lactofermentum*. *Biochem. Biophys. Res. Commun.* **133**:1013-1019.
- Gribskov, M., J. Devereux, and R. R. Burgess. 1984. The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression. *Nucleic Acids Res.* **12**:539-549.
- Holmes, D. S., and M. Quigley. 1981. A rapid boiling method for the preparation of bacteria plasmids. *Anal. Biochem.* **114**:193-197.
- Hopwood, D. A., M. J. Bibb, K. F. Chater, T. Kieser, C. J. Bruton, H. M. Kieser, and D. J. Lydiate. 1985. Genetic manipulation of *Streptomyces*. A laboratory manual. John Innes Foundation, Norwich, United Kingdom.
- Ishino, S., K. Yamaguchi, K. Shirahata, and K. Araki. 1984. Involvement of meso- α,ϵ -diaminopimelate D-dehydrogenase in lysine biosynthesis in *Corynebacterium glutamicum*. *Agric. Biol. Chem.* **48**:2557-2560.
- Kaneko, T., T. Hashimoto, R. Kumpaisal, and Y. Yamada. 1990.

- Molecular cloning of wheat dihydrodipicolinate synthase. *J. Biol. Chem.* **265**:17451–17455.
13. Kawakami, B., T. Kudo, Y. Narahashi, and K. Horikoshi. 1986. Nucleotide sequence of the N-acetylneuraminate lyase gene of *Escherichia coli*. *Agric. Biol. Chem.* **50**:2155–2158.
 14. Ludovice, M., J. F. Martín, P. Carrachas, and P. Liras. 1992. Characterization of the *Streptomyces clavuligerus argC* gene encoding N-acetylglutamyl-phosphate reductase: expression in *Streptomyces lividans* and effect on clavulanic acid production. *J. Bacteriol.* **174**:4606–4613.
 - 14a. Malumbres, M., et al. Unpublished data.
 15. Martín, J. F. 1989. Molecular genetics of amino acid-producing corynebacteria, p. 25–59. In S. Baumberg, I. Hunter, and M. Rhodes (ed.), *Microbial products. A practical approach*. Society for General Microbiology Symposium 44. Cambridge University Press, Cambridge.
 16. Mead, D. A., and B. Kemper. 1988. Chimeric single-stranded DNA phage-plasmid cloning vectors, p. 85–102. In R. L. Rodriguez and D. T. Denhardt (ed.), *Vectors: a survey of molecular cloning and their uses*. Butterworths, Stoneham, Mass.
 17. Ohta, Y., K. Watanabe, and A. Kimura. 1985. Complete nucleotide sequence of the *E. coli* N-acetylneuraminate lyase. *Nucleic Acids Res.* **13**:8843–8852.
 18. Richaud, F., C. Richaud, P. Ratet, and J. C. Patte. 1986. Chromosomal location and nucleotide sequence of the *Escherichia coli dapA* gene. *J. Bacteriol.* **166**:297–300.
 19. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.
 20. Schrupf, B., A. Schwarzer, J. Kalinowski, A. Pühler, L. Eggeling, and H. Sahm. 1991. A functionally split pathway for lysine synthesis in *Corynebacterium glutamicum*. *J. Bacteriol.* **173**:4510–4516.
 21. Smith, O. H., and C. Yanofsky. 1962. Enzymes involved in the biosynthesis of tryptophan. *Methods Enzymol.* **5**:794–806.
 22. Southern, E. M. 1979. Gel electrophoresis of restriction fragments. *Methods Enzymol.* **68**:152–176.
 23. Tabor, S., and C. C. Richardson. 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc. Natl. Acad. Sci. USA* **82**:1074–1078.
 24. Wierenga, R. K., P. Terpstra, and G. J. Hol. 1986. Prediction of the occurrence of the ADP-binding $\beta\alpha\beta$ -fold in proteins using an amino acid sequence fingerprint. *J. Mol. Biol.* **187**:101–107.
 25. Yugar, Y., and C. Gilvarg. 1965. The condensation step in diaminopimelate synthesis. *J. Biol. Chem.* **240**:4710–4716.