JOURNAL OF BACTERIOLOGY, Apr. 1997, p. 2777–2782
0021-9193/97/$04.00+0
Copyright © 1997, American Society for Microbiology
Vol. 179, No. 8

Cloning of the *dapB* Gene, Encoding Dihydrodipicolinate Reductase, from *Mycobacterium tuberculosis*

MARTIN S. PAVELKA, JR.,[1,2] TORIN R. WEISBROD,[2,3] AND WILLIAM R. JACOBS, JR.[2,3]*

Howard Hughes Medical Institute,[3] Department of Microbiology and Immunology,[1] and Department of Biochemistry,[2] Albert Einstein College of Medicine, Bronx, New York 10461

Received 9 December 1996/Accepted 6 February 1997

Diaminopimelate (DAP) is used by bacteria for the synthesis of lysine. In many species of bacteria, including mycobacteria, DAP is also used for peptidoglycan biosynthesis. In this report we describe the cloning of the *dapB* gene encoding dihydrodipicolinate reductase (DHPR), which catalyzes a key branch point reaction in the bacterial DAP biosynthetic pathway, from *Mycobacterium tuberculosis*. Analyses of the DapB proteins from different bacterial species suggest that two different classes of DHPR enzymes may exist in bacteria.

The aspartate amino acid family includes those amino acids with carbon skeletons derived primarily from aspartate and includes methionine, threonine, isoleucine, and lysine (Fig. 1). Diaminopimelate (DAP), an intermediate from this pathway, is the direct precursor to lysine in all bacteria and is also an important component of the peptidoglycan in many bacteria, including mycobacteria (Fig. 1) (28). We recently demonstrated that the aspartate family pathway is essential to *Mycobacterium smegmatis*; i.e., mutants with a disruption in *ask*, the gene encoding aspartokinase, the first enzyme of this pathway, are nonviable, even when all the products of the aspartate family are present in the growth medium (18). The essential product of the pathway appears to be DAP, as we were able to construct an *ask* mutant of *M. smegmatis* using a strain unable to convert DAP to lysine due to a mutation in the *lysA* gene (18). Thus, DAP starvation is a bacteriocidal event in mycobacteria.

We are interested in the DAP biosynthetic pathway of mycobacteria because DAP is not produced or required by humans. The study of mycobacterial DAP biosynthetic enzymes would allow the design of specific enzyme inhibitors that could have potential as new antimycobacterial agents. Furthermore, DAP auxotrophic mutants of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG could be useful as new vaccine strains and tools for the study of tuberculosis pathogenesis. Recently, a DAP auxotrophic mutant of *Shigella flexneri* was proven useful in the delivery of naked DNA into epithelial cells in preliminary DNA immunization experiments (26). In other work, complementation of DAP auxotrophy has been utilized as a method to maintain plasmids in vivo in animal experiments with *Salmonella typhimurium* (7).

Three variant pathways exist for the synthesis of DAP in bacteria (Fig. 1). The succinylase pathway is present in *Escherichia coli* (13), while the acetylase and/or the dehydrogenase pathway can be found among the members of the genus *Bacillus* (30). *Corynebacterium glutamicum* has both the succinylase pathway and the dehydrogenase pathway (24). It is not entirely clear which DAP pathways are used by mycobacteria. They probably utilize, at the very least, the succinylase pathway since *Mycobacterium leprae* is known to have the *dapE* and *dapF* genes, encoding the enzymes responsible for the last two steps of the pathway (27). All three DAP pathways share two enzymes, dihydrodipicolinate synthase, encoded by the *dapA* gene, and dihydrodipicolinate reductase (DHPR), encoded by the *dapB* gene. The latter enzyme, dihydrodipicolinate reductase, catalyzes the conversion of dihydrodipicolinate (DHP) to tetrahydrodipicolinate in a pyridine nucleotide-dependent reaction (shown in bold in Fig. 1) (9). The DHPR protein of *E. coli* functions as a tetramer and shows an unusual ability to utilize either NADH or NADPH as a cofactor for catalysis, with the former nucleotide only slightly preferred over the latter (20). We do not know which variant DAP pathway is used by *M. tuberculosis*; therefore, we directed our attention to the *dapB* gene, since the DHPR-catalyzed reaction yields the substrate for all three variant DAP biosynthetic pathways (Fig. 1).

Cloning of the *M. tuberculosis dapB* gene. To clone the *dapB* gene of *M. tuberculosis*, we took advantage of the previously released sequence of the *dapB* gene of *M. leprae* (27). Oligonucleotide primers (BJ23891, 5′GTGAACACTATGCGAGTAGG3′; BJ23892, 5′TCATTGCAGGTTGAGTAAGG3′) were designed against the *M. leprae dapB* open reading frame and synthesized by the Albert Einstein College of Medicine (AECOM) oligonucleotide facility. Standard PCR reactions were done (18), using the GeneAmp PCR kit from Perkin Elmer (Norwalk, Conn.) with these primers and *M. leprae* genomic DNA (*M. leprae* genomic DNA was generously supplied by L. P. Miller, AECOM). The 748-bp *M. leprae dapB* PCR product was used as a probe in colony hybridizations with a *Bam*HI-digested *M. tuberculosis* H37Rv genomic library constructed in pUC19 (kindly supplied by L. P. Miller). A hybridization-positive clone was obtained and single colony purified. The plasmid from this clone was confirmed to contain *dapB* by Southern analysis using the original *M. leprae dapB* gene PCR product as the probe (data not shown). The plasmid, pYUB701, contains two *Bam*HI fragments, 7 and 1.5 kb in size, with the *dapB* gene present on the 7-kb fragment (Fig. 2). It is not known if the two *Bam*HI fragments represent contiguous genomic fragments.

Complementation analysis. For complementation experiments, *E. coli* strains were electroporated with plasmid DNA and transformants were selected on Luria-Bertani (LB) agar containing 50 μg of ampicillin per ml and 100 μg of DAP (a mixture of the L,L-, D,D-, and *meso*-DAP isomers; Sigma Chemical Co., St. Louis, Mo.) per ml. Colonies were picked and patched onto LB agar plates containing ampicillin with or without DAP and scored for the ability to grow in the absence

* Corresponding author. Mailing address: Howard Hughes Medical Institute, Albert Einstein College of Medicine, 1300 Morris Park Ave., Bronx, NY 10461. Phone: (718) 430-2888. Fax: (718) 518-0366. E-mail: jacobs@aecom.yu.edu.

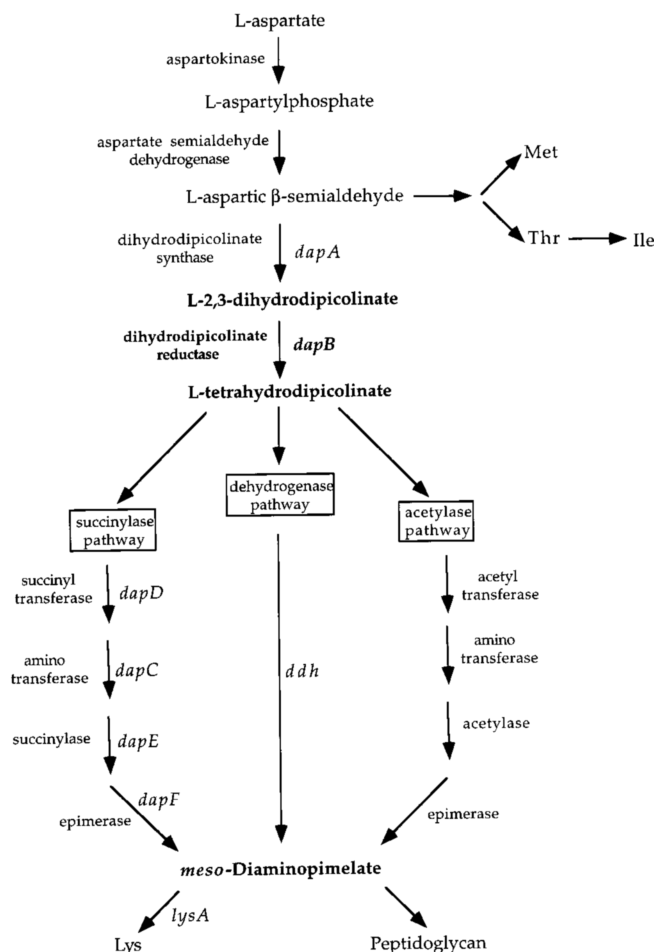

FIG. 1. The aspartate amino acid family pathway and the three variant pathways for DAP biosynthesis in bacteria. Adapted from reference 28.

of DAP. We determined that pYUB701 expressed a functional DHPR activity by its ability to complement an *E. coli dapB* mutant (Table 1). Plasmid pYUB701 does not complement other *E. coli* DAP auxotrophs with mutations in *dapA*, *dapD*, or *dapE* (Table 1), suggesting that these genes are not present on this plasmid. In *E. coli*, the *dap* genes are not found together in an operon but instead are scattered throughout the chromosome (2). However, the *dapB* and *dapA* genes of *Brevibacterium lactofermentum* are located within the same operon, with the *dapB* gene promoter proximal, followed by an unidentified open reading frame and then *dapA* (19). The *dapB* and *dapA* genes of *M. leprae* appear to be organized in a similar fashion (27). It is possible that a similar *dapB dapA* operon arrangement also exists in *M. tuberculosis*. The insert within pYUB701 is large enough to encode both genes, and the inability of pYUB701 to complement an *E. coli dapA* mutant may indicate that the *M. tuberculosis dapA* gene is not efficiently expressed in *E. coli.*

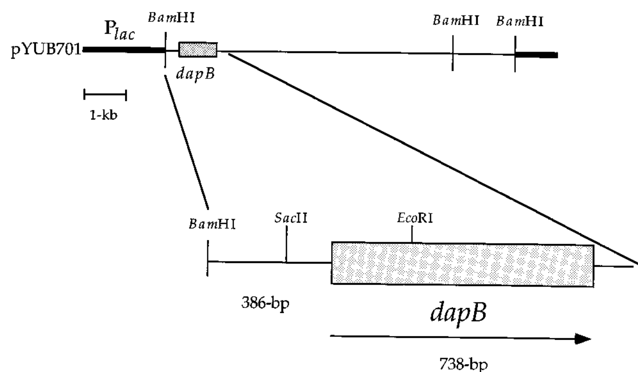

FIG. 2. Map of insert within plasmid pYUB701 encoding *dapB* from the *M. tuberculosis* H37Rv::pUC19 library.

TABLE 1. pYUB701 complementation in *E. coli*

E. coli strain	Relevant genotype	Source or reference	Plasmid	Growth on LB containing[a]:	
				AP^{50}	$AP^{50}DAP^{100}$
ec[2]904	*dapA*::Cm	C. Richaud	pYUB701	0/10	10/10
AT999	*dapB*::Mu	6	pKS^+	0/10	10/10
			pYUB701	10/10	10/10
AT986	*dapD8*	6	pYUB701	0/10	10/10
AT987	*dapE6*	6	pYUB701	0/10	10/10

[a] Transformants were selected on LB medium containing an antibiotic (ampicillin [AP]) without or with DAP. Ten clones from each transformation experiment were picked and patched onto antibiotic-containing medium without or with DAP. AP^{50}, 50 μg of ampicillin per ml; DAP^{100}, 100 μg of DAP per ml.

Sequence analysis of the *dapB* gene. One end of the genomic DNA insert within plasmid pYUB701 was sequenced using the Applied Biosystems Prism Dye Terminator Cycle Sequencing Core kit with AmpliTaq DNA polymerase (Perkin Elmer) and an Applied Biosystems 377 automated DNA sequencer. Sequence data for both strands were obtained by primer walking. The nucleotide sequence of the *M. tuberculosis dapB* gene with the translated amino acid sequence is shown in Fig. 3. The *dapB* open reading frame is 738 bp in length and has a guanosine plus cytosine (G+C) content of 63%, which is within the range reported for *M. tuberculosis* genomic DNA (29). The *dapB* genes of *M. tuberculosis* and *M. leprae* are the same size, with a nucleotide identity of 82% and an identity of 88% at the amino acid level for the translated products. We confirmed that the cloned *dapB* gene in pYUB701 was from *M. tuberculosis* H37Rv by Southern analysis using probes made from either restriction endonuclease-generated fragments from pYUB701 or a PCR product from pYUB701, generated using primers which flank the gene. No signal was seen for *M. smegmatis*, while a strong signal was seen for *M. tuberculosis* H37Rv and *M. bovis* BCG (data not shown). The lack of hybridization to *M. smegmatis* genomic DNA is likely due to sequence divergence between the species.

Recently, the crystal structure of the *E. coli* DapB protein complexes with NADPH was solved to a 2.2-Å resolution (23). It was demonstrated that the *E. coli* enzyme has two domains, an NAD(P)H binding domain in the N terminus (residues 2 to 130 and 240 to 273) and a proposed DHP substrate binding domain, located towards the C terminus (residues 131 to 239) (23). Within the DHPR-NADPH structure, these two sites are situated approximately 12 Å away from each other, and it was proposed that a large conformational change must occur in the protein in order to bring the two domains close enough together for catalysis (23).

We compared the amino acid sequences of the mycobacterial DapB proteins with that of *E. coli* and five other bacterial DapB sequences (Fig. 4). All of these proteins share several areas of homology, in particular the motif (V/I)(A/G)(V/I)XGXXGXXG, located in the extreme N terminus of each

```
GGATCCCTGC GGTTTGGCCG TTGGGGAGCT TCGTTGTGCA CCAGCCGTTT     50
BamHI
CCGACCAACC TGTGGTGGGA TGTGCCGGTT GCCCACTGGT CGATATGGTG    100

GACTTCGCAC CGGTCACCGG GTGCGTGGCA GTTGGGTCGA CCATGGGTGC    150

GGTGCCCGCG TTGCTGCGGG TCCGTTTCCC GTGGCCGCCC TTCGAACCGG    200

ACGTGCGACT TTCACCGTAT CTGGCTCTTC ACGGAATCTG CCGATGGGGT    250

GGTTCAGACA GCTGCGACAG AACTACCGTG CAGGTCTTTC ACCTCCACTC    300

GATCAACAAG CGCCTCACGG CGCACGCCGG GTTTGGTGCA GCCGCGGTCG    350
                         SD
TTGGCCTCGA GGACGGCCGG TCT[AGGG]TGA AGGCCATGCG GGTAGGCGTC    400
                                       M  R   V  G  V
CTTGGAGCCA AAGGCAAGGT CGGAACGACA ATGGTGCGGG CGGTGGCCGC    450
L  G  A  K   G  K  V   G  T  T   M  V  R  A   V  A  A
CGCCGACGAC CTGACCCTAT CCGCCGAGCT GGATGCCGGC GATCCGCTGA    500
 A  D  D   L  T  L  S   A  E  L   D  A  G   D  P  L  S
GCCTGCTAAC GGACGGTAAC ACCGAGGTCG TCATCGACTT CACCCACCCG    550
  L  L  T   D  G  N   T  E  V  V   I  D  F   T  H  P
GACGTGGTGA TGGGCAATCT GGAGTTCCTC ATCGACAACG GAATTCACGC    600
D  V  V  M   G  N  L   E  F  L   I  D  N  G   I  H  A
CGTGGTCGGT ACCACGGGGT TCACCGCCGA GCGGTTTCAA CAAGTCGAAT    650
 V  V  G   T  T  G  F   T  A  E   R  F  Q   Q  V  E  S
CGTGGCTCGT CGCAAAACCC AACACATCGG TGTTGATAGC GCCAAACTTC    700
  W  L  V   A  K  P   N  T  S  V   L  I  A   P  N  F
GCGATCGGAG CGGTGCTGTC CATGCATTTC GCCAAGCAGG CCGCACGGTT    750
A  I  G  A   V  L  S   M  H  F   A  K  Q  A   A  R  F
TTTCGACTCG GCCGAGGTCA TTGAGCTGCA TCATCCGCAC AAGGCTGACG    800
 F  D  S   A  E  V  I   E  L  H   H  P  H   K  A  D  A
CGCCGTCAGG CACGGCCGCG CGTACCGCGA AGCTGATCGC CGAGGCCCGA    850
  P  S  G   T  A  A   R  T  A  K   L  I  A   E  A  R
AAAGGCTTGC CGCCCAATCC CGATGCCACC AGTACCAGCC TGCCGGGCGC    900
K  G  L  P   P  N  P   D  A  T   S  T  S  L   P  G  A
GCGTGGTGCC GACGTCGACG GCATACCGGT GCACGCGGTG CGGCTGGCCG    950
 R  G  A   D  V  D  G   I  P  V   H  A  V   R  L  A  G
GACTGGTCGC CCACCAGGAA GTGCTGTTCG GGACCGAGGG GGAGACTCTG   1000
  L  V  A   H  Q  E   V  L  F  G   T  E  G   E  T  L
ACCATCCGCC ACGATAGCCT CGATCGCACA TCGTTTGTGC CCGGTGTGCT   1050
T  I  R  H   D  S  L   D  R  T   S  F  V  P   G  V  L
GTTGGCGGTG CGCCGCATCG CCGAACGCCC TGGTCTCACC GTAGGTCTTG   1100
 L  A  V   R  R  I  A   E  R  P   G  L  T   V  G  L  E
AGCCCCTACT CGATCTGCAC TGAACGAACA CTGTGATGAC CAAGCGCACC   1150
  P  L  L   D  L  H   .
```

FIG. 3. Nucleotide sequence from pYUB701 containing *dapB*. The *Bam*HI site used for cloning is shown, along with the putative Shine-Dalgarno (SD) site (boxed) (25). The translated DapB amino acid sequence is shown below the nucleotide sequence.

protein. This motif is found in a variety of dinucleotide (NAD or NADP)-dependent dehydrogenases and is part of the dinucleotide binding domain (19). The *E. coli* DHPR nucleotide binding domain is composed of four α-helices (A1 to A4) and seven β-strands (B1 to B7), making up a Rossman dinucleotide binding fold (23). This nucleotide binding fold can be separated into two halves, the first containing B1, B2, B3, and A1, with β strands B2 and B3 connected by a loop spanning positions Glu40 through Gly59. The second part of the fold is comprised of B4, B5, B6, A2, and A3. The two halves of the nucleotide binding domain are connected by the Ser65 to Asp74 region (23). Interestingly, this part of the protein (the loop connecting B2 with B3 in the first part of the nucleotide binding domain and the region connecting the two halves of the nucleotide binding domain together) varies among the DapB sequences of the various species (indicated by the bold lines in Fig. 4). This region is maintained within the DapB proteins of *E. coli*, *Haemophilus influenzae*, *Pseudomonas syringae*, and *Bacillus subtilis*, while most of it is absent from the DapB proteins of *M. tuberculosis*, *M. leprae*, *B. lactofermentum*, and *C. glutamicum* (Fig. 4). The gapped region in the *M. tuberculosis* DapB protein is from Ala34 to Thr43. These sequence differences are within a flexible region of the Rossman fold which is generally tolerant of such changes (23).

As mentioned previously, the *E. coli* DHPR enzyme displays an unusual dual affinity for either NADH or NADPH as a cofactor. A set of definitive predictive rules for the nucleotide specificity of dehydrogenases is largely unknown (21). In the structure of the *E. coli* DHPR-NADPH complex, it was found that the Arg39 residue of the protein has an electrostatic interaction with the 2′-phosphate of NADPH, and thus, it was proposed that the Arg39 residue may play a role in defining the NADPH affinity of the *E. coli* enzyme (23). In contrast, the recently reported structure of the *E. coli* DHPR-NADH complex showed that NADH binding is dependent upon Glu38, which makes hydrogen bonds with the 2′ and 3′ hydroxyls of the adenosyl ribose (21). Strikingly, the DapB proteins of *E. coli*, *H. influenzae*, and *P. syringae* all have an arginine residue in the position equivalent to Arg39 of the *E. coli* enzyme (located at the beginning of the divergent loop region and indicated in bold type in Fig. 4). However, the corresponding residue in the DapB proteins of *M. tuberculosis*, *M. leprae*, *B. lactofermentum*, and *C. glutamicum* is not arginine, but rather an alanine or a valine. All of the proteins have a negatively charged residue near the Arg38 equivalent position, corresponding to the Glu38 of the *E. coli* protein. If the presence of a positively charged Arg39 residue and a nearby negatively charged (Glu38) residue in the *E. coli* DHPR determines the dual specificity for NADPH and NADH, then one would predict that the DHPR enzymes of *H. influenzae* and *P. syringae* would also display dual nucleotide specificity. Logically, it follows that the mycobacterial DHPR proteins, and the others like them (Fig. 4), lacking the positively charged Arg residue at the equivalent position and having a negatively charged (Glu or Asp) residue nearby, would show a preference for NADH. In the case of the *B. subtilis* enzyme (Fig. 4), the side chain of a histidine residue replacing the Arg residue at that position may not carry a positive charge in the microenvironment of the protein or may not be long enough to interact with the 2′-phosphate of the NADPH; thus, this DHPR enzyme would be predicted to use only NADH as a cofactor.

A phylogenetic analysis of these proteins groups them into two clusters which reflect the sequence differences noted above (Fig. 5). The gram-negative organisms, *E. coli*, *H. influenzae*, and *P. syringae*, cluster together, as do the high G+C content gram-positive organisms (*M. tuberculosis*, *M. leprae*, *B. lactofermentum*, and *C. glutamicum*) and the low G+C content gram-positive *B. subtilis*. These groupings correlate with differences seen in Fig. 4 for the loop region in the nucleotide binding domain of these proteins. We propose that two classes of DapB proteins exist, one class found in gram-negative organisms (having the Glu38/Arg39 equivalent residues and able to use NADH or NADPH as a cofactor) and the other class found in gram-positive organisms (lacking the Arg39 equivalent residue and able to use only NADH as a cofactor). In this scenario, the *B. subtilis* enzyme would function the same as the enzymes from the high G+C content gram-positive bacteria. Not shown in our analyses in Fig. 4 and Fig. 5 were the partial sequences of the *dapB* genes of two other gram-negative organisms, *Pseudomonas aeruginosa* (14) and *Klebsiella pneumoniae* (4). Both of these proteins also have an Arg residue at a position equivalent to the *E. coli* residue position and are very similar to the *E. coli* class of proteins, particularly in the loop region; thus, we would predict that these two proteins also belong to the first class of DHPR enzymes described above.

```
                       NAD-binding motif
E.coli            MHDANIRVAI AGAGGRMGRQ LIQAALALEG VQLGAALERE GSSLLGSDAG   50
H.influenzae      M---TLKIAI AGAGGRMGCQ LIQAVHSAEG VELGAAFERK GSSLVGTDAG   47
P.syringae        M----RRIAV VGAAGRMGKT LIEAVQQAPG AGLTAAIDRP DSTLVGADAG   46
■ M.tuberculosis  M-----RVGV LGAKGKVGTT MVRAVAAADD LTLSAELDA- ------GDPL   38
B.subtilis        MSNETIKLVI AGPRGRMGQE AVKLAERTPH FDLVGAIDHT YDQQKLSDVM   50
B.lactofermentum  M---GIKVGV LGAKGRVGQT IVAAVNESDD LELVAEIGV- ------DDDL   40
C.glutamicum      M---GIKVGV LGAKGRVGQT IVAAVNESDD LELVAEIGV- ------DDDL   40
M.leprae          -----MRVGV LGAKGKVGST MVRAVQAVED LTLSAEVEA- ------GDPL   38
Consensus         M---...V.V .GA.GR.G.T .V.AV..... ..L.A..... .......D..   50

ELAGAGKTGV TVQSSLDAVK DDFDVFIDFT RPEGTLNHLA FCRQHGKGMV  100
ELAGIGHLGV AVSDDLESQK DKFDLLIDFT RPEGTLEHIA FCVANNKKMV   97
ELAALGRIGV PLSGDLAKVA DEFDVLIDFT HPSVTLKNLA FCRKAGKAMI   96
SL-------- -----L--TD GNTEVVIDFT HPDVVMGNLE FLIDNGIHAV   73
PVESDAFIYT DIHACF--TE TQPDVLIDLT IPEIGKVHTK IALEHGVRPV   98
SL-------- -----L--VD NGAEVVVDFT IPNAVMGNLE FCINNGISAV   75
SL-------- -----L--VD NGAEVVVDFT IPNAVMGNLE FCINNGISAV   75
SL-------- -----L--TE GNTEVVIDFT HPDVVMGNLE FLIGNGIHAV   73
.L........ .....L--.. ....V.IDFT .P.....NL. FC..NG...V  100

IGTTGFDEAG KQAIRD---A AADIAIVFAA NFSVGVNVML KLLEKAAKVM  147
IGTTGFDENG KVAIKA---A SDKIAIVFAS NFSVGVNLVF KLLEKAAKVM  144
IGTTGFSAEE KQRLVE---A GKDIPIVFAA NFSIGVNLCL KLLDTAARVL  143
VGTTGFTAER FQQVESWLVA KPNTSVLIAP NFAIGAVLSM HFAKQAARFF  123
VGTTGFSEAD LKELTS-LTE EKGIGAIIAP NFALGAILMM KFSKMAANYF  147
VGTTGFDDAR LEQVRAWLEG KDNVGVLIAP NFAISAVLTM VFSKQAARFF  125
VGTTGFDDAR LEQVRDWLEG KDNVGVLIAP NFAISAVLTM VFSKQAARFF  125
VGTTGFTAER LQQVQSWLFN QPNTSVLIAP SFAISAVLSV HFARQAAPFF  123
VGTTGF.... .......L.. .......IAP NFAIGA.L.. .F...AA..F  150
                  "DHP- binding motif"
GDYTDIEIIE AHHRHKVDAP SGTALAMGEA IAHALDKDLK DCAVYSREGH  197
GDYCDIEVIE AHHRHKVDAP SGTALSMGEH IAKTLGRDLK THGVFCREGI  194
GDEVDIEITE AHHRHKVDAP SGTALRMGEV VASALGRDLE KVAVYGREGQ  193
DS--A-EVIE LHHPHKADAP SGTAARTAKL IAEA--R--K G--LPPNPDA  164
ED--V-EIIE LHHDQKLDAP SGTALKTAEM ISEV--R--K E--KQQGHPD  188
ES--A-EVIE LHHPNKLDAP SGTAIHTAQG IAAA--R--K EAGMDAQPDA  168
ES--A-EVIE LHHPNKLDAP SGTAIHTAQG IAAA--R--K EAGMDAQPDA  168
DS--SVEVIE LHHPHKADAP SGTATRTATL IAEA--R--K --GLPLKPDA  165
..--..EVIE LHH.HK.DAP SGTA..TA.. IA.A--R--K ..........  200

TGERVPGTIG FATVRAGDIV GEHTAMFADI GERLEIT-HK ASSRMTFANG  246
TGERKRDEIG FSTIRASDVV GEHSVWFADI GERVEIS-HK ASSRMTFANG  243
TGARDRQTIG FATIRAGDVV GDHTVLFAAD GERVEIT-HK ASSRMTFAKG  242
TSTSLPGARG -ADV---DGI PVHAVRLAGL VAHQEVLFGT EGETLTIRHD  210
EKEILPGARG -AEQ---NGI RLHSVRLPGL IAHQEVMFGM DGQTLQIRHD  234
TEQALEGSRG -ASV---DGI PVHAVRMSGM VAHEQVIFGT QGQTLTIKQD  214
TEQALEGSRG -ASV---DGI PVHAVRMSGM VAHEQVIFGT QGQTLTIKQD  214
TSTSLPGARG -ADV---DGI PVHSVRLAGL VAHQEVLFGT AGETLTIRHD  211
T...L.G.RG -A.V---DGI ..H.VR.AG. .AH.EV.FG. .G.TLTI..D  250

AVRSALWLSG KESGLFDMRD ----VLDLNN L-----  273
AVRAGKWLEN KANGLFDMTD ----VLDLNN L-----  270
AVRAAMWLDG KAPGLYDMQD ----VLGLH- ------  267
SLDRTSFVPG VLLAVRRIAE RPGLTVGLEP LLDLH-  246
SYNRASFMSG VKLSVEQVMK IDQLVYGLEN IID---  267
SYDRNSFAPG VLVGVRNIAQ HPGLVVGLEH YLGL--  248
SYDRNSFAPG VLVGVRNIAQ HPGLVVGLEH YLGL--  248
SLDRTSFVPG VLLAVRRITE RPGLTVGLEP LLNLQL  247
S..R.SF..G V..GV..... ...LV.GLE. ....--  286
```

FIG. 4. Comparison of eight bacterial DapB proteins, from *E. coli* (5), *H. influenzae* (10), *P. syringae* (16), *B. subtilis* (12), *B. lactofermentum* (19), *C. glutamicum* (8), and *M. leprae* (27). The alignment was done with Geneworks software (11) using pairwise alignment (15). The consensus sequence is shown at the bottom, and the invariant residues are boxed. The NAD-binding motif (V/I)(A/G)(V/I)XGXXGXXG and the putative DHP-binding motif E(L/A)HHXXKXDAPSGTA are both indicated above the sequences. The major gap region in the N termini of the proteins is indicated by bold lines above the sequences, while the *E. coli* Arg39 residue and the corresponding residues of the other proteins are shown in bold type.

DHP, the substrate of the DHPR enzyme, is negatively charged; therefore, the substrate binding domain of the protein would be expected to have positively charged residues (23). In the structure of the *E. coli* DHPR-NADH complex, the putative DHP-binding domain was positioned in the protein in the His159 and His160 to Lys163 region (23). All of the proteins examined in this study have a striking conservation of sequence in this region, including a "DAP box" with the motif E(L/A)HHXXKXDAPSGTA (Fig. 4). This consensus could likely be considered a DHP-binding motif.

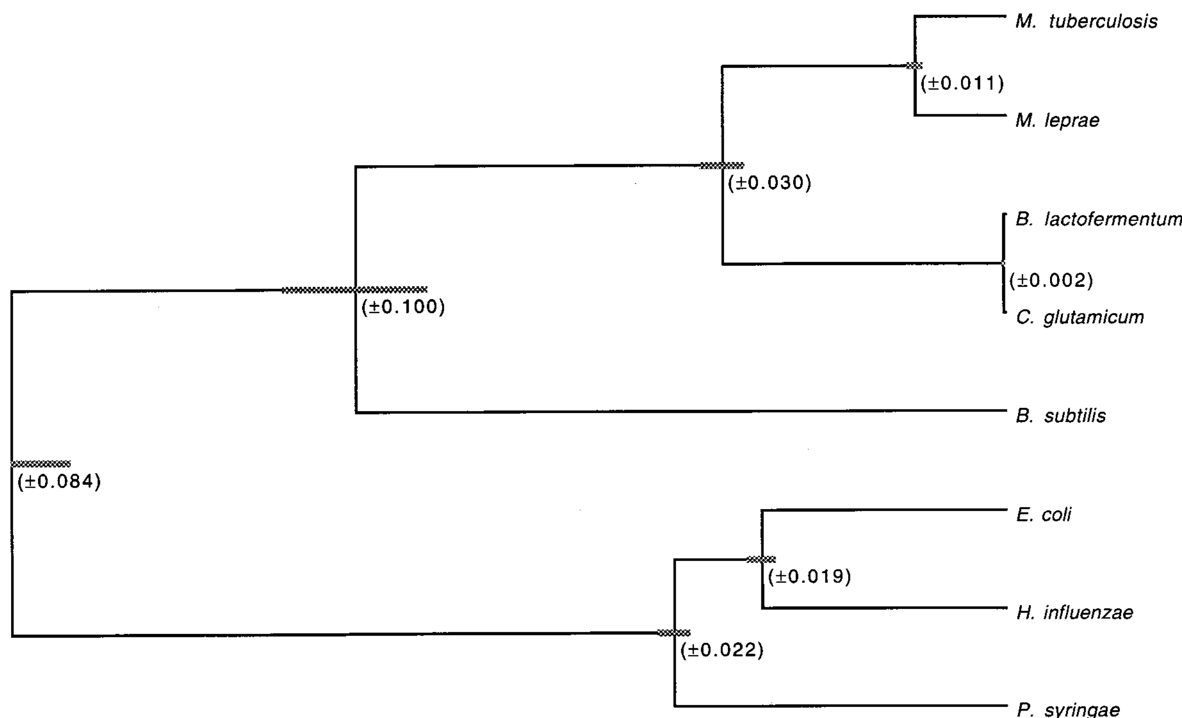

FIG. 5. Relatedness of eight bacterial DapB proteins. This evolutionary tree was constructed with Geneworks software (11), using the unweighted pair group method with arithmetic mean (UPGMA) algorithm (17). The estimated genetic distance between sequences is proportional to the length of the horizontal lines connecting one sequence to another. The hatched lines and corresponding values at the branch points indicate the standard error for each branch position.

We have cloned the first *dap* gene from the pathogenic mycobacterial species *M. tuberculosis*. The recent demonstrations of allelic exchange in slow-growing mycobacteria such as *M. tuberculosis* (3) and *M. bovis* BCG (1, 22) and our recent demonstration of a way to construct DAP auxotrophs of *M. smegmatis* (18) should allow for the construction of DAP auxotrophs of *M. tuberculosis*. Studies of the DHPR enzyme from *M. tuberculosis* and comparison with the DHPR enzyme of *E. coli* should reveal additional information about the cofactor selectivity of pyridine nucleotide-requiring enzymes and determine approaches for the design of inhibitors of the mycobacterial DHPR enzyme.

Nucleotide sequence accession number. The DNA sequence encoding the *M. tuberculosis dapB* gene was submitted to GenBank and assigned the accession number U66101.

This work was supported by National Institutes of Health grants AI26170 and AI33696.

We gratefully thank L. P. Miller for *M. leprae* genomic DNA and the *M. tuberculosis* H37Rv library and G. Scapin and J. Blanchard for discussions and critical reading of the manuscript.

REFERENCES

1. **Azad, A. K., T. D. Sirakova, L. M. Rogers, and P. E. Kolattukudy.** 1996. Targeted replacement of the mycocerosic acid synthase gene in *Mycobacterium bovis* BCG produces a mutant that lacks mycosides. Proc. Natl. Acad. Sci. USA **93:**4787–4792.
2. **Bachmann, B.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54:**130–197.
3. **Balasubramanian, V., M. S. Pavelka, Jr., S. S. Bardarov, J. Martin, T. R. Weisbrod, R. A. McAdam, B. R. Bloom, and W. R. Jacobs, Jr.** 1996. Allelic exchange in *Mycobacterium tuberculosis* with long linear recombination substrates. J. Bacteriol. **178:**273–279.
4. **Bott, M., and P. Dimroth.** 1994. *Klebsiella pneumoniae* genes for citrate lyase and citrate lyase ligase; localization, sequencing, and expression. Mol. Microbiol. **14:**347–356.
5. **Bouvier, J., C. Richaud, F. V. Richaud, J. C. Patte, and P. Stragier.** 1984. Nucleotide sequence and expression of the *E. coli dapB* gene. J. Biol. Chem. **259:**14829–14834.
6. **Bukhari, A. I., and A. L. Taylor.** 1971. Genetic analysis of diaminopimelic acid- and lysine-requiring mutants of *Escherichia coli*. J. Bacteriol. **105:**844–854.
7. **Curtiss, R., III, K. Nakayama, and S. M. Kelley.** 1989. Recombinant avirulent *Salmonella* vaccine strains with stable maintenance and high level expression of cloned genes *in vivo*. Immunol. Invest. **18:**583–596.
8. **Eikmanns, B. J.** 1992. Direct GenBank submission. Accession number X67737.
9. **Farkas, W., and C. Gilvarg.** 1965. The reduction step in diaminopimelic acid biosynthesis. J. Biol. Chem. **240:**4717–4722.
10. **Fleischmann, R. D., et al.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269:**496–512.
11. **Glynias, M. J.** 1991. Geneworks version 2.0. IntelliGenetics, Inc., Mountain View, Calif.
12. **Henner, D., P. Gollnick, and A. Moir.** 1990. Analysis of an 18-kilobase pair region of the *Bacillus subtilus* chromosome containing the *mtr* and *gerC* operons and the *aro-trp-aro* supraoperon. Proc. Int. Symp. Genet. Ind. Microorg. **6:**657–665.
13. **Kindler, S. H., and C. Gilvarg.** 1960. *N*-Succinyl-L-α,ε-diaminopimelic acid deacylase. J. Biol. Chem. **235:**3532–3535.
14. **Kwon, D., C. Lu, D. Walthall, J. Brown, J. Houghton, and A. Abdelal.** 1994. Structure and regulation of the *carAB* operon in *Pseudomonas aeruginosa* and *Pseudomonas stutzeri*. J. Bacteriol. **176:**2532–2542.
15. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227:**1435–1441.
16. **Liu, L., and P. D. Shaw.** 1997. Characterization of *dapB*, a gene required by *Pseudomonas syringae* pv. tabaci BR2.024 for lysine and tabtoxinine-β-lactam biosynthesis. J. Bacteriol. **179:**507–513.
17. **Nei, M.** 1987. Molecular evolutionary genetics. Columbia University Press, New York, N.Y.
18. **Pavelka, M. S., Jr., and W. R. Jacobs, Jr.** 1996. Biosynthesis of diaminopimelate, the precursor of lysine and a component of peptidoglycan, is an essential function of *Mycobacterium smegmatis*. J. Bacteriol. **178:**6496–6507.
19. **Pisabarro, A., M. Malumbres, L. M. Mateos, J. A. Oguiza, and J. F. Matrin.** 1992. A cluster of three genes (*dapA*, *orf2*, and *dapB*) of *Brevibacterium lactofermentum* encodes dihydrodipicolinate synthase, dihydrodipicolinate reductase, and a third polypeptide of unknown function. J. Bacteriol. **175:** 2743–2749.
20. **Reddy, S. G., J. C. Sacchettini, and J. S. Blanchard.** 1995. Expression, purification, and characterization of *Escherichia coli* dihydrodipicolinate re-

ductase. Biochemistry **34:**3492–3501.
21. **Reddy, S. G., G. Scapin, and J. S. Blanchard.** 1996. Interaction of pyridine nucleotide substrates with *Escherichia coli* dihydrodipicolinate reductase: thermodynamic and structural analysis of binary complexes. Biochemistry **35:**13294–13302.
22. **Reyrat, J.-M., F.-X. Berthet, and B. Gicquel.** 1995. The urease locus of *Mycobacterium tuberculosis* and its utilization for the demonstration of allelic exchange in *Mycobacterium bovis* bacillus Calmette-Guérin. Proc. Natl. Acad. Sci. USA **92:**8768–8772.
23. **Scapin, G., J. S. Blanchard, and J. C. Sacchettini.** 1995. Three-dimensional structure of *Escherichia coli* dihydrodipicolinate reductase. Biochemistry **34:** 3502–3512.
24. **Schrumpf, B., A. Schwarzer, J. Kalinowski, A. Puhler, L. Eggeling, and H. Sahm.** 1991. A functionally split pathway for lysine synthesis in *Corynebacterium glutamicum*. J. Bacteriol. **173:**4510–4516.
25. **Shine, J., and L. Dalgarno.** 1974. The 3′ terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to non-sense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71:**1342–1346.
26. **Sizemore, D. R., A. A. Branstrom, and J. C. Sadoff.** 1995. Attenuated *Shigella* as a DNA delivery vehicle for DNA-mediated immunization. Science **270:** 299–302.
27. **Staffen, B., and S. T. Cole.** 1994. MycDB: an integrated mycobacterial database. Mol. Microbiol. **5:**517–534.
28. **Umbarger, H. E.** 1978. Amino acid biosynthesis and its regulation. Annu. Rev. Biochem. **47:**533–606.
29. **Wayne, L. G., and W. M. Gross.** 1968. Base composition of deoxyribonucleic acid isolated from mycobacteria. J. Bacteriol. **96:**1915–1919.
30. **Weinberger, S., and C. Gilvarg.** 1970. Bacterial distribution of the use of succinyl and acetyl blocking groups in diaminopimelic acid biosynthesis. J. Bacteriol. **101:**323–324.