Mandelate Pathway of *Pseudomonas putida*: Sequence Relationships Involving Mandelate Racemase, (*S*)-Mandelate Dehydrogenase, and Benzoylformate Decarboxylase and Expression of Benzoylformate Decarboxylase in *Escherichia coli*[†,‡]

Amy Y. Tsou, Stephen C. Ransom, and John A. Gerlt*
Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742
Douglas D. Buechter, Patricia C. Babbitt,* and George L. Kenyon*
Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, California 94143
Received March 9, 1990; Revised Manuscript Received July 11, 1990

ABSTRACT: The genes that encode the five known enzymes of the mandelate pathway of *Pseudomonas putida* (ATCC 12633), mandelate racemase (*mdl*A), (*S*)-mandelate dehydrogenase (*mdl*B), benzoylformate decarboxylase (*mdl*C), NAD^+-dependent benzaldehyde dehydrogenase (*mdl*D), and $NADP^+$-dependent benzaldehyde dehydrogenase (*mdl*E), have been cloned. The genes for (*S*)-mandelate dehydrogenase and benzoylformate decarboxylase have been sequenced; these genes and that for mandelate racemase [Ransom, S. C., Gerlt, J. A., Powers, V. M., & Kenyon, G. L. (1988) *Biochemistry 27*, 540] are organized in an operon (*mdl*CBA). Mandelate racemase has regions of sequence similarity to muconate lactonizing enzymes I and II from *P. putida*. (*S*)-Mandelate dehydrogenase is predicted to be 393 amino acids in length and to have a molecular weight of 43 352; it has regions of sequence similarity to glycolate oxidase from spinach and ferricytochrome b_2 lactate dehydrogenase from yeast. Benzoylformate decarboxylase is predicted to be 499 amino acids in length and to have a molecular weight of 53 621; it has regions of sequence similarity to enzymes that decarboxylate pyruvate with thiamin pyrophosphate as cofactor. These observations support the hypothesis that the mandelate pathway evolved by recruitment of enzymes from preexisting metabolic pathways. The gene for benzoylformate decarboxylase has been expressed in *Escherichia coli* with the *trc* promoter, and homogeneous enzyme has been isolated from induced cells.

The enzymes of the mandelate metabolic pathway (Figure 1) allow various pseudomonads to utilize one or both of the enantiomers of mandelate as a sole carbon source (Fewson, 1988). Usually, the catabolism of (*R*)-mandelate requires the presence of mandelate racemase to equilibrate the enantiomers and generate (*S*)-mandelate, which can then be oxidatively degraded to benzoate by the remaining enzymes in the pathway. Benzoate is subsequently converted to acetyl-CoA and succinyl-CoA by the enzymes of the β-ketoadipate pathway. Our interests in the mechanisms of the reactions catalyzed by enzymes in the mandelate pathway found in *Pseudomonas putida* require that the genes for the various enzymes be cloned and sequenced so that large amounts of wild-type and mutant enzymes can be obtained. We have previously reported the cloning (Ransom et al., 1988) and expression (Tsou et al., 1989) of the gene for mandelate racemase (*mdl*A), the first enzyme in the pathway.

In this article we describe the cloning of a single restriction fragment that encodes the five enzymes of the pathway. The DNA sequence of part of this fragment reveals that the five genes do not form a single operon: the genes for benzoylformate decarboxylase (BFD),[1] (*S*)-mandelate dehydrogenase (MDH), and mandelate racemase (MR) are arranged in an operon (*mdl*CBA); the genes for the benzaldehyde dehydrogenases are independently transcribed. The predicted amino acid sequences of MDH and BFD reveal that each of these enzymes shows sequence similarities to enzymes catalyzing similar reactions from other sources. MR and muconate lactonizing enzyme I (MLE I, also from *P. putida*) recently were discovered to be structurally related (Neidhart et al., 1990). The available sequence information for MDH and BFD now suggests that these enzymes also evolved by recruitment of preexisting enzymes. Finally, the gene for BFD has been expressed in *Escherichia coli* JM105 with the *trc* promoter; homogeneous enzyme has been isolated from induced cells.

Materials and Methods

The plasmid pSCR1 that contains the sequenced gene for MR as well as approximately 3 kb of DNA upstream of the initiation codon for MR has been described (Ransom et al., 1988). The preparation of chromosomal DNA from *P. putida* ATCC 12633, vectors, bacteriophage, most bacterial strains, and sources for commercially available enzymes and reagents have also been previously described (Ransom et al., 1988; Tsou et al., 1989).

Conjugative transfers of plasmids between *E. coli* HB101 and *P. putida* ATCC 17453 were performed in the presence of *E. coli* MM294 transformed with pRK2013 (obtained from Dr. Jeffrey Karns, USDA Laboratories, Beltsville, MD).

[†] This is paper 12 in a series on mandelate racemase; for paper 11, see Lin et al. (1990). This research was supported by Grants GM-34572, GM-40570, and AR-17323 from the National Institutes of Health.
[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J05293.

[1] Abbreviations: ALS, acetolactate synthase; BFD, benzoylformate decarboxylase; FPLC, fast protein liquid chromatography; GOX, glycolate oxidase; LB, Luria broth; MDH, (*S*)-mandelate dehydrogenase; MLE, muconate lactonizing enzyme; MR, mandelate racemase; PDC, pyruvate decarboxylase; POX, pyruvate oxidase; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis.

0006-2960/90/0429-9856$02.50/0

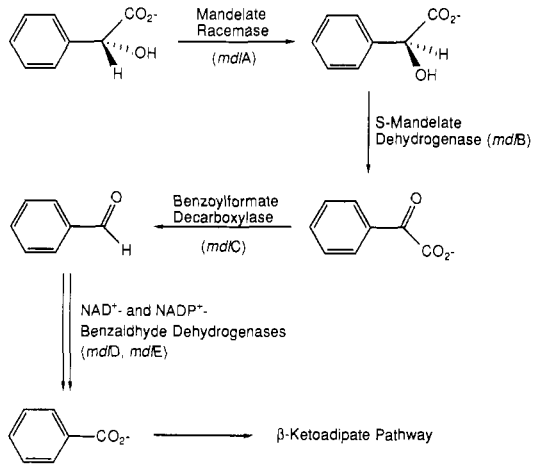

FIGURE 1: Mandelate pathway in *P. putida* ATCC 12633.

Cloning of the Restriction Fragment Encoding the Mandelate Pathway. Chromosomal DNA from *P. putida* ATCC 12633 was restricted with *Eco*RI and subjected to Southern blot analysis. Two oligonucleotide primers were synthesized, 5′-end labeled, and used as probes: the mixed 38-mer TAGCC(T/C)TGGCGGCG(C/G)AA(C/G)AA-(T/C)TC(G/A)TA(G/C)TG(G/C)GTGCC(G/A)TG that is the *complement* of the predicted DNA sequence of codons 5–17 of BFD and, given the high GC content of *P. putida* DNA, utilizes G and/or C in the wobble positions; the 17-mer AATGTCCAGCCACCTGT incorporating the exact DNA sequence of codons 273–279 of the gene for MR. Both probes were observed to hybridize to *Eco*RI fragments approximately 10.5 kb in length. Fragments between 7 and 12 kb in size were electroeluted from an agarose gel and ligated in the *Eco*RI site of pKT230. The ligation mixture was used to transform *E. coli* HB101. The plasmids contained in transformants were directly conjugated into *P. putida* ATCC 17453; conjugants containing the 10.5-kb *Eco*RI fragment that encodes the mandelate pathway were selected on benzoylformate minimal agar plates containing 150 μg/mL kanamycin (Hegeman, 1966).

DNA Sequence Analysis. DNA sequence analysis was performed with the Sequence kit and procedure obtained from U.S. Biochemicals. Nested deletions were generated with the kit and procedure obtained from Pharmacia. Regions of severe band compression were resolved by use of either dITP or 2′-deoxy-7-deazaguanosine 5′-triphosphate in place of dGTP in the sequencing reactions.

Comparison of Protein Sequences with Data Bases. The deduced amino acid sequences of MDH and BFD were compared with the most recent NBRF data base (release 23) by use of the Eugene sequence analysis package (Altschul & Erickson, 1986). Alignments were generated with the algorithm developed by Smith and Smith (1990). The statistical significance of similarities between any two sequences was tested by the Monte Carlo program in Eugene. The significance of multiple sequence alignments was tested by serially randomizing each of the probe sequences (BFD, MDH, or MR) ten times and then aligning them with the related sequences. The percent identity of either the authentic or randomized probe sequence with a related sequence was calculated and compared.

Expression of the Gene for BFD in E. coli. The DNA sequence of the gene for BFD revealed the presence of a unique *Cla*I site in codons 17 and 18 and a unique *Hin*dIII site downstream of the gene. A heteroduplex was prepared from two synthetic oligonucleotides (one 71-mer and one 75-mer)

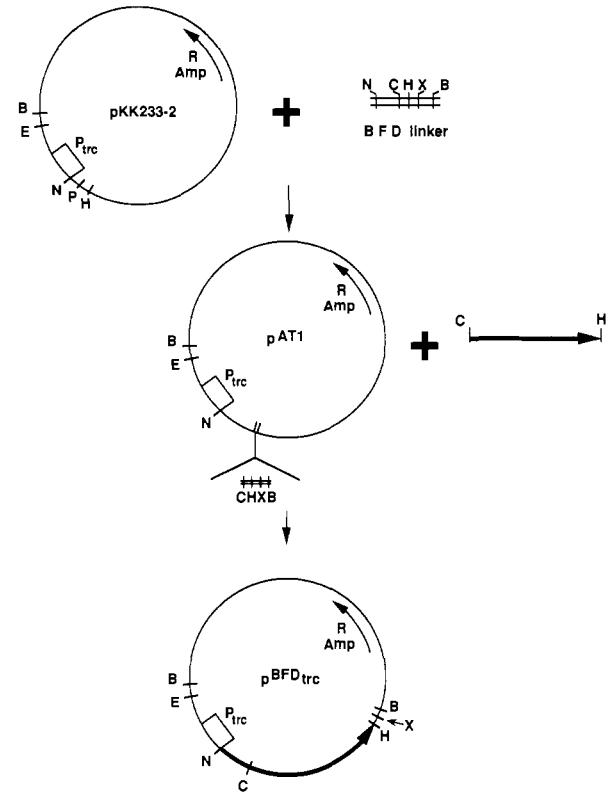

FIGURE 2: Strategy for the construction of pBFDtrc. Restriction sites: B, *Bam*HI; C, *Cla*I; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; X, *Xba*I. pKK233-2 (4593 bp) was digested with *Hin*dIII, made blunt ended with the Klenow fragment and dNTPs, and digested with *Nco*I prior to ligation with the synthetic linker described in the text; the resulting plasmid was designated pAT1 (4654 bp). pAT1 was digested with *Cla*I and *Hin*dIII and ligated with the *Cla*I–*Hin*dIII fragment containing the gene for BFD to obtain pBFDtrc (6225 bp).

with an *Nco*I sticky end at one end followed by the N-terminal coding region of BFD to the unique *Cla*I site followed by *Hin*dIII, *Xba*I, and *Bam*HI sites and finally additional bases to generate a blunt end. As detailed in the next paragraph and summarized in Figure 2, the *Cla*I and *Hin*dIII sites allow placement of the gene for BFD in pKK233-2; the *Bam*HI and *Xba*I sites permit the excision and insertion of the gene in other expression vectors.

The expression vector pKK233-2 has a unique *Nco*I site followed by a unique *Hin*dIII site. The vector was restricted with *Hin*dIII, made blunt by the action of the Klenow fragment in the presence of deoxynucleoside triphosphates, restricted with *Nco*I, and ligated with the synthetic heteroduplex. Plasmids were isolated from transformants and analyzed for the presence of *Bam*HI, *Cla*I, *Hin*dIII, *Nco*I, and *Xba*I sites; a plasmid having these sites (pAT1) was used for further manipulations. This plasmid was restricted with *Cla*I and *Hin*dIII and ligated with the 1510-bp restriction fragment obtained from the intact gene with the same enzymes. Plasmids were isolated from transformants, screened by restriction analysis, and sequenced in the region of the construction. One such plasmid having the correct sequence was designated pBFDtrc and used for isolation of enzyme from transformed *E. coli* cells.

Purification of BFD from Transformed and Induced E. coli JM105. After inoculation, two 1500-mL cultures of *E. coli* JM105 transformed with pBFDtrc in LB containing 50 μg/mL ampicillin were grown with vigorous aeration at 37 °C until the OD_{590nm} reached 1.0, at which time production of BFD was induced by the addition of isopropyl β-thiogalactopyranoside to a final concentration of 0.2 mM. After further

incubation at 37 °C for 2 h, the cells were harvested by centrifugation at 5000*g* for 20 min. The wet cells were weighed and stored overnight at −20 °C.

All of the following steps were performed at 4 °C unless otherwise indicated, and all of the buffers, except those used in the final chromatography step, contained 0.1 mM phenylmethanesulfonyl fluoride.

The cells (12.4 g) were thawed and suspended in 75 mL of 50 mM sodium phosphate, pH 7.0. A solution of lysozyme dissolved in the same sodium phosphate buffer (1.24 mL of a 10 mg/mL solution) was added to the suspended cells. After being stirred for 2 h, the resulting suspension was sonicated for eight cycles of 3 min each. The sonicated suspension was diluted to 130 mL with additional buffer and centrifuged at 100000*g* for 60 min. The supernatant was decanted and used for purification of BFD.

Solid ammonium sulfate (39.5 g, 50% saturation) was added to the supernatant over a 10-min period. After being stirred for an additional 60 min, the suspension was centrifuged at 10000*g* for 20 min. An additional 24.6 g of solid ammonium sulfate (75% saturation) was added over 15 min to the supernatant (140 mL). This mixture was stirred for 90 min and centrifuged at 10000*g* for 20 min. The pellet was resuspended in 28 mL of 100 mM sodium phosphate, pH 6.0, and heated in a water bath at 55 °C for 10 min. The mixture was cooled in an ice bath and centrifuged at 15000*g* for 15 min. The pellet was discarded, and the supernatant was dialyzed overnight against two changes of 2 L each of 50 mM sodium phosphate, pH 6.0.

The contents of the dialysis bag were clarified by centrifugation at 15000*g* for 15 min and applied to a column of DEAE-Sephacel (24 cm × 2.6 cm) that had been previously equilibrated in 50 mM sodium phosphate, pH 6.0. The column was eluted at a flow rate of 1 mL/min with a 1-L linear gradient from 0 to 0.5 M NaCl in 50 mM sodium phosphate, pH 6.0. Fractions (7 mL) containing greater than 3 units/mL BFD were combined and concentrated to a final volume of 2.1 mL by ultrafiltration (Amicon XM-50 membrane). The concentrated solution was applied to a column of Sephacryl S-200 (91 cm × 2.6 cm) equilibrated in 50 mM sodium phosphate, pH 6.0, and eluted at a flow rate of 14 mL/h. Fractions (4.5 mL) containing greater than 1.5 units/mL BFD were combined and concentrated to a final volume of 3.5 mL by ultrafiltration.

Following clarification by filtration, 38 μL of the enzyme-containing solution was further purified by FPLC on an analytical Mono Q HR 5/5 column (Pharmacia). BFD was eluted from the column with a linear gradient of 150–400 mM NaCl in sodium phosphate buffer, pH 6.0, over 110 min at a flow rate of 1 mL/min. Fractions containing activity were analyzed by SDS–PAGE, and those were electrophoretically homogeneous were combined and concentrated by ultrafiltration to a final concentration of 1 mg/mL. The purified enzyme was stored at 4 °C.

N-Terminal Sequence Analyses of BFD. One nanomole of BFD isolated by Dr. George Garcia from *P. putida* ATCC 12633 (Garcia, 1987) and 200 pmol of enzyme purified from *E. coli* JM105 were dried by vacuum centrifugation (Savant Speedvac) and subjected to automated N-terminal sequence analyses with an Applied Biosystems Model 470A gas-phase sequenator located in the Biomolecular Resource Center at the University of California, San Francisco.

Assay for BFD. The coupled assay using horse liver alcohol dehydrogenase was previously described (Weiss et al., 1988). The assay was initiated by the addition of benzoylformate so that the rate of benzoylformate-independent oxidation of NADH could be measured. The background oxidation of NADH was significant only in assays of crude cell extracts.

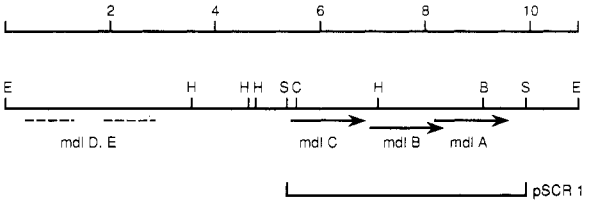

FIGURE 3: Limited restriction map of the 10.5-kb *Eco*RI restriction fragment of *P. putida* DNA cloned in pSCR4 encoding the mandelate pathway. Restriction sites: B, *Bam*HI; C, *Cla*I; E, *Eco*RI; H, *Hin*dIII; S, *Sau*3A. The position of the 4.6-kb *Sau*3A restriction fragment of *P. putida* DNA cloned in pSCR1 (Ransom et al., 1988) is also given.

Results and Discussion

Cloning of the Genes for the Mandelate Pathway. Earlier experiments reported by Wheelis and Stanier (1970) and our recent experiments (Tsou et al., 1989) suggest that the genes for the five enzymes in the mandelate pathway are closely clustered on the chromosome. We decided to synthesize a degenerate 38-mer probe based upon the N-terminal sequence of BFD and a 17-mer probe from the sequence of MR to identify a chromosomal fragment that might encode all five activities.

Both probes hybridized to *Eco*RI restriction fragments of identical size, approximately 10.5 kb. A library of *Eco*RI restriction fragments was constructed in pKT230, and by selecting for the plasmid with kanamycin and for the genes of the mandelate pathway with benzoylformate as sole carbon source, several colonies were obtained. Restriction analysis of the plasmids isolated from these colonies revealed that the same 10.5-kb *Eco*RI restriction fragment was present in each and that both orientations of the restriction fragment in the vector had been obtained. Two plasmids with the *Eco*RI fragment in opposite orientations were designated pSCR4 and pSCR5. A limited restriction map of the *Eco*RI fragment is shown in Figure 3; the region in common with the DNA we originally cloned in pSCR1 is indicated. The relative positions of *mdl*D and *mdl*E were established by deletion analysis (data not shown); these are separated from *mdl*C by at least 2 kb of DNA. The relative positions of *mdl*A, *mdl*B, and *mdl*C were established by deletion analysis (data not shown), and the exact positions by the results of the DNA sequence analysis summarized in this article.

DNA Sequence Analysis of the HindIII–BamHI Fragment of the P. putida DNA Cloned in pSCR1. The previously unsequenced portion of the *Hin*dIII–*Bam*HI restriction fragment of pSCR1 that is upstream of the 5′-end of the gene for MR (*mdl*A) (Figure 3) was sequenced by subcloning smaller restriction fragments in M13 and use of the M13 universal primer to determine sequence; when necessary, synthetic primers were used to extend the data obtained with the universal primer. Both strands were completely sequenced. These new data provide all but the first 13 residues of MDH.

DNA Sequence Analysis of the 2.3-kb HindIII Fragment of the P. putida DNA Cloned in pSCR4. The 2.3-kb *Hin*dIII fragment of pSCR4 (immediately to the left of the *Hin*dIII–*Bam*HI fragment described in the previous section; see Figure 3) was subcloned in pUC19 in both orientations. Nested deletions were constructed in each with exonuclease III and nuclease S_1. Following selection of appropriate deletions by restriction analysis, sequence was determined with the universal primer. Both strands were completely sequenced. This

```
LDH   mlkykpllkisknceaailrasktrlntiraygstvpksksfeqdsrkrtqswtalrvgailaatssvaylnwhngqidnepkldmnkqk   90
GOX   ------------------------------------------------------------------------------------------   0

MDH   ------------------------------------------------------------------------------------------   0

LDH   ispaevakhnkpddcwvvingyvydltrflpnhpggqdvikfnagkdvtaifeplhapnvidkyiapekklgplqgsmppelvcppyapg   180
GOX   ------------------------------------------------------------------------------------------   0

MDH   ------------------------------------------------------------------------------------------   0

LDH   etkediarkeqlksllppldNIiNlyDfEyLAsQtLtKqawaYYsSGAnDEvThrENhNAyhRIfFKPKILVDVrkvDisTdmLGshvdv   270
GOX   ------------------meItNVneYEaiAkQkLPKMVYDYYaSGAEDqwTlaENRNAFsRIlFrPrILiDVtniDmtTtiLGfkiSM    71
                        *  ** **  *    ********   *****     **  *    **** ****          **   **
MDH   ------------------msqNlfNVeDYrkLrqkrLPKMVYDYlegGAEDEygvkhNRdvFqqwrFKPKrLVDVsrrslqaevLGkrqSM    73

LDH   PfyvsaTALcKlgnPlEGEkdvARgcgqgvTkvpqmIsTLaScSpeeIiEaApsdkqIqWyQLYVnsDRkItddLVknvEklGvKALfvT   360
GOX   PimIaPTAmqKmahP-EGEyAtARAAsaAGT-----ImTLSSwatsSvEEvAstgpGIrfFQLYVykDRnvvaqLVrrAEraGfKAiaLT   155
      *  * ** *     *  *  * ****  **         **  *  ***  *    *  ******   * *    *  *   *   * **
MDH   PllIgPTgLngalwP-kGdlAlARAAtkAGi-----pfvLStaSnmSIEdlArqcdGdlWFQLYV-ihReIaqgmVlkAlhtGyttLvLT   156

LDH   VDaPsLGqREkDmKlkF--------------------SntkagpkamkKtnveeSQGaS-----ralSkfIDpSLtWKDieeLkkkTkLP   425
GOX   VDtPrLGrREaDiKNRFvlPpfl--------------tLknfegidlgKmdkandsGlS-----syvagQIDrSLsWKDvaWLqtiTsLP   226
       *    * ** *  ***  *                  **                **  *        * * * *  *    **
MDH   tDvavnGyRErDlHNRFkiPmsysakvvldgclhprwSLdfvrhgmpqlanfvsSQtsSlemqaalmSrQmDaSfnWealrWLrdlwphk   246

LDH   IviKGVqrtEDvikAaeiGvsGVvLSNHGGRQLDfsrAPIEVLAEtmpileqrnlKdkleVFvDGGVRRGTDVlKALcLGAkGVgLGRPf   515
GOX   ILVKGVitAEDArlAvqhGAaGiIvSNHGaRQLDyvpAtImaLeEvV-----kaAqgriPVFlDGGVRRGTDVfKALALGAaGVfiGRPv   311
       ****   ****      ** *************    * ****  *       **   **  * * *** *  *******  * ***
MDH   lLVKGllsAEDAdrciaeGAdGVILSNHGGRQLDcaisPmEVLAqsV-------AKtgkPVliDsGfRRGsDivKALALGAeaVlLGRat   329

LDH   LYanscyGrnGVeKaieiLRDEIEmsMrLlGvtSIaELkPDlLdlstlkarTvgVpndvlynevyegptltefeda                591
GOX   vfsLAAeGEaGVkKVLqmmRDEfElTMALsGCrSlkEiSrshiaadwdgpssraVarL-----------------                369
      ** *** ** **  **  *   *  * *  **  *  **** *        *  *  *
MDH   LYgLAArGEtGVdeVLtlLkadIdrTlAqiGCpdItsLSPDyLq-negvtnTapVdhLigkgtha-----------                393
```

FIGURE 4: Alignment of MDH with LDH and GOX.

fragment encodes all of BFD and the first 13 residues of MDH.

DNA and Protein Sequence Compilation. The DNA sequence determined in the present study and that previously reported for the gene for MR (Ransom et al., 1988) have been combined and are included in the supplementary material (see paragraph at end of paper regarding supplementary material). The deduced amino acid sequences of MR, MDH, and BFD, along with sequence comparisons with related enzymes, are given in Figures 4–6.

The predicted N-terminal sequence of BFD agrees exactly with the first 19 residues that were determined by Edman degradation of protein purified from *P. putida*. The experimentally determined amino acid composition and that deduced from the gene sequence are in generally good agreement (data not shown). The predicted molecular weight is 53 621 and is in reasonable agreement with the polypeptide molecular weight of ca. 57 500 determined by SDS–PAGE.

MDH recently was purified from *P. putida* ATCC 12633 in the laboratory of Professor Charles A. Fewson, University of Glasgow, Glasgow, Scotland (personal communication). Comparison of the N-terminal sequence of this protein with that deduced from the DNA sequence revealed that 28 of 30 residues matched exactly, thereby confirming that *mdl*B encodes MDH. The molecular weight of MDH determined by SDS–PAGE, 44 000, is in excellent agreement with the value of 43 352 predicted by our DNA sequence analysis.

Proteins Related to BFD, MDH, and MR. We recently discovered that MR and MLE I, also from *P. putida* ATCC 12633, must be evolutionarily related (Neidhart et al., 1990). Although the primary sequences exhibit only 25.6% sequence identity (see below), the secondary, tertiary, and quaternary structures are nearly superimposable. Whether MR evolved from MLE I, since MLE I is also necessary for the metabolism of mandelate, or the two enzymes evolved from a common ancestor is currently unknown. We are interested in determining whether evolutionary relationships might involve other enzymes in the mandelate pathway. If detected, these would support the notion that enzymes have been recruited from preexisting pathways to evolve the mandelate pathway. Such evolution should be especially important in a soil bacterium such as *P. putida* since it would be advantageous to utilize as many potential carbon and energy sources as possible within its ecological niche. Since our observations regarding MR and MLE I suggest that even modest levels of sequence identity can occur in proteins with striking amounts of conserved three-dimensional structure, searching the available data bases for proteins with sequence identity should be instructive in understanding the evolution of the mandelate pathway.

Significant sequence similarities were found among MDH and both ferricytochrome b_2 lactate dehydrogenase from yeast (LDH; Guiard, 1985) and glycolate oxidase from spinach (GOX; Volokita & Somerville, 1987; Cederlund et al., 1988) (Figure 4). The overall percent identity of MDH with either LDH or GOX is 42.5%; the mean percent identity for 10 randomizations of MDH aligned with the other two α-hydroxy acid dehydrogenases is 15.6%, with the score for the authentic sequence being 10.3 standard deviation units from that mean. Pairwise alignments of MDH with LDH and GOX produced sequence identities of 30.3% and 33.3%, respectively. Both pairwise alignments yield Monte Carlo scores well into the range predictive of "probable homology". Crystallographic

```
ALS   MeMlSgAemvvrsLidQGVKqVFGyPGgavLdiyDaLhtVGGIdhvLvRHEqaAVhmADGlARAtGEvgVvlvTSGPGatNAItGIAtAY    90
POX   -mkqTVAaYiaktLesaGVKriwGvtGDslngLsDsLnrmGtIeWmstRHEevAAfAAgaeAQlSGElAVcagscGPGnLhlINGLfdch    89
PDC   MseiTlgkYlfErLkqvnVnTVFGlPGDfnLsLLDkiyeVEGmRWagnanElnAAyAADGYAR-ikgmscIitTfGvGeLsAlNGIAgsY    89
        * **     * *  **  **** **   *  *      *  *  *   *   •  ********   *     *  *  **   *  *
BFD   --MaSVhgttyElLrrQGidTVFGnPGsneLpfLkdfp--EdfRyiLalqEacvVgiADGYAQASrkpAfInlhSaaGtgNAmgaLsnAw    86

ALS   mdsiPLvVlsGQvatSlIGydAF----QEcDmvgisRPvVKhSfLVkqtEDI--PQVLkkAfwlAasgRpGPVVVdLPkDi---LnPAnk    171
POX   rnHVPVLaiAahipsSeIGsgyF----QEthpqeLfRecshycELVsSpEqI--PQVLaiAmrkAvlnRgvsVVV-LPgDva--LKPA--    168
PDC   aeHVgVLhvvGvpsiSsqakqlLLhtlgNgDftvfhRmsaniSEttAmitDIctPQAeidrcirttyvtQrPVYLgLPanlvdlnvPA--    177
        * ** * ***     **  ***    * *   * ** ** *   **      * *   *   *    ******  * *     * *
BFD   nsHsPLiVtAGQqtramIGveALL---tNvDaanLpRPlVKwSyepASaaev--PhAmsrAihmAsmapQGPVYLsvPyDdw--dKdA--    167

ALS   lpyvwPESvsMrSYnPttt--ghkgQikRaLqsvVavkKpvVyVgGgait-AGcHqqLKEtvEaLnlPVVcsLMGlGafpaTHrqvlGML    258
POX   -----PEgatMhwYHapqpvVtpEeEeLRkLaqLLrysSNiAlmcGsgC--AGAHkELVEfAgkiKAPiVhALrGKehveydnPydvGMt    251
PDC   ---kllQtpidmSlkPnda--esEkEviDtilvLVkdAKNPVIladacCsrhdvkAEtKkLidltqfPafVtPMGKGsiseqHPrygGvy    262
           ***      *     *    * ** *   *  ******* *     * * *  * *** ***** ***         *       *
BFD   ----dPQShhlfdrHvsss-VrlndQdLDiLvkaLnsASNPAIVlGpdvdaAnAnAdcVmLAErLKAPVwVAPsaprcpfpTrhpcfrgL    252

ALS   Gmh-GtyEanmTMhNADVIfaVGvrFdDRttnnl-AkYcPnAtvlhIDIDPtSIsktvtaDipIVGDarqvLeqmLeLlsqeSahqpLDe    346
POX   Gli-GfssgfhTMmNADtlvllGtqFpyR------AfYptdAkIIqIDInPaSIgahskvDmAlVGDIkstLrALLpLVEEkadrkfLDk    334
PDC   vgtlskpEvkeavEsADlILsVGAllsDf-ntgsfsYsyktknIvefhsDhmkIrnAtfpg---VqmkfvlqklLtNiadaakg------    342
         *        *  ** *  **          *  *    *    **      *   * *** **      ** ***** *
BFD   mpa-GiaaisqllEghDVvLviGApvfryhqydpgqYlkPgtrlIsvtcDPleaarApmgD-AIVaDIgamasALaNLVEEsSr------    334

ALS   irdwwqqieqwrARqcLK-yDthSEK-IKPQaVieTLwrltkgDAyVTsDVGqhqmfAAlYypfdkpRRwinSgg---lGSMGFgLPAAL    431
POX   aledyr-----dARkgLddlAkPSEKaIHPQylaqQishFAadDAIfTcDVGTpTvwAArYlkmNgkRRllgSfn---hGSManAmPqAL    416
PDC   -ykpva----vpARtPAn-aAvPAstpLKqEwmwnQLgnFLqEgdvViaEtGTsafginqttfpNntygisqvlwgsigfttGatLgAAf    426
                 *  ***  *  *   **** *  **   * * **   *    *    *              *  *******
BFD   --qlpt-----aApePAK-vDqdAgr-LHPEtVfdTLndmApEnAIylnEststTgkcAalnmrNpgsyyfca-----aGglGFALPAAi    410

ALS   GVkmAlPEetVVcVtGDGSiQMnIQELSTAlQYeLPVlvVnLNNryLGmVKqwqdMIysGrhSqsymqslpdFvRrg-AyGhvGIqishp    520
POX   GAQatEPERQVVAmcGDGgfsMlmgdflsvvQmkLPVkiVVfNNsvLGfVA--meMkaGGyltdgtelhdtnFaRiaeAcGitGIrvEkA    504
PDC   aAeeidPkkrVIlfIGDGSlQltvQEiSTmirwgLkpylfVLNNdgYtieK----lIhGpkaqyneiqgwdhlslLP---tfgakdyEth    509
      *** ***************    *  * ** **  *   *  **  **  *    *      *            ***      *     *
BFD   GVQlAEPERQVIAVIGDGSanysIsaLwTAaQYniPtifVimNNgtYGrfA----MvcrrsrSrkrswagcardRLP--rtrqGlwcpsA    494

ALS   hgwkanLaRrwnrcaiiawcllmlps-mAaststrcrfAGaewmkcg----------                                    566
POX   sevdeaLqRafsiDgpvlvdvvvakeElAiPpqikleqAkgfslymlraiisgrgdevielaktnwlr                          572
PDC   rvattgewdkltqDksfndnskirmiEvmlPvfdcstklG----------                                          549

BFD   esrqp----------                                                                             499
```

FIGURE 5: Alignment of BFD with ALS, POX, and PDC.

```
MLEI     MtsaLIErIDaiiVDlPtiRPhklAmhTmQqQtLVvlrvRcSdGVeGIGEattiGGlAygyEspEgIKanIdahL    75
MLEII    ---mkIEaIDvtlVDVPasRPiqmsftTVQkQsyaivqiR-agGlcGIGEgssvGGptwssEcaEtIKviIetyL    71
         *   **       * **   *   *  **    **      * ** *          *
MR       MsevLItglrtraVnVPlayPvhtAvgTVgtapLVlidlatSaGVvGhsylfaytpvAlk-slkqllddmaamiv    74

MLEI     APaLIGlaAdNinAamlkLdklakGNtfAKsGIesALleAqGKrLgLPvSeLLGGgvRdslavAWTLASGDTaRD   150
MLEII    APlLIGkdAtNlrelqhlmeravtGNysAKAaIDvALhDlkahsLnLPLSdLiGGaiqqgipiAWTLASGDTqRD   146
            *         *    *         * **** *  ** **    **  ***   *     *    * *
MR       nepLapvsleamlAkrfcLagytglirmAaAGIDmAawDAlGKvhetPLvkLLGanaRp--vqAydshSlDgvkl   147

MLEI     IAEArhMlEiRRHRvFKlKIGadpveQDLkHVvtIkrelGDsAsVRVDvNQyWDEsqAIracQvLgdnGidLIEQ   225
MLEII    IAiAeeMiErRRHnrFKiKlGvrspadDLrHiekIierVGDrAaVRVDiNQaWDEntAsvwiprLeaaGVeLvEQ   221
           *          *  * ***     ***  *  *   ***     ** **  *   **   * *   **  **
MR       atEravtaaelgfRavKtKIGypaldQDLavVrsIrqaVGDdfgimVDyNQslDvpaAIkrsQaLqqeGVtwIEe   222

MLEI     PisRiNrgGqvRLnqrspaPImADESieSvEdAFsLAadGAasiFaLKiaKnGGpravLRtAqIAEAaGIglYGG   300
MLEII    PvaRsNfdalrRLsadngVaIlADESlsSlasAFeLArhhcvdaFsLKlcnmGGVantLkvAAIAEAsGIasYGG   296
         *       *  *       **     *    *  *      **          * ***    *   * *   **
MR       PtlqhdyeGhqRiqsklnVPvqmgEnwlgpEemFkalsiGAcrlampdamKiGGVtgwiRasAlAqqfGIp--ms   295

MLEI     TMLegaIGTlAsaHaflTLrqlTwgtELfGPLlLteeivneppqyrDFqlhIPrtPGLGltlDEqrLarFaRr-   373
MLEII    TMLdssIGTaAalHvyaTLptmpfecELLGPwvLAdtltqtqLeikDFeIrlPsgPGLGvdiDpdkLrhFtRag   370
           *   *          *    *     *  * **          *          **  ** *    *
MR       shLfqeIsa--------hLlaaTptahwLerLdLAgsvieptLtfeggnavIPdlPGvGiiwrEkeigkylv--   359
```

FIGURE 6: Alignment of MR with MLE I and MLE II.

studies of both GOX (Lindqvist & Braenden, 1985) and LDH (Xia et al., 1987) reveal that these have structurally similar α/β barrel domains that contain the binding site for FMN, the cofactor known to bind to GOX, LDH, and MDH from *Acinetobacter calcoaceticus* (Hoey et al., 1987). Of the 13 residues known to be in the active site of GOX (Lindqvist &

Bränden, 1985, 1989; Lindqvist, 1989), 9 are conserved in all three α-hydroxy acid dehydrogenases, including Tyr 24, Tyr 129, Lys 230, and Arg 257 (using the GOX numbering system). Therefore, we believe that it is reasonable to hypothesize that MDH evolved by recruitment of the duplicated gene for an α-hydroxy acid dehydrogenase in another metabolic pathway in *P. putida*.

Significant sequence similarities were also found among BFD and pyruvate oxidase (POX), acetolactate synthases (ALS), and pyruvate decarboxylases (PDC); the alignment of the sequence of BFD with POX (Grabau & Cronan, 1986) and the most similar ALS (*ilv*I gene product from *E. coli*; Squires et al., 1983) and PDC (*Saccharomyces cerevisiae*; Kellerman et al., 1986) is reproduced in Figure 5. The overall percent identity between the sequence of BFD and one or more of the sequences for ALS, POX, and PDC is 40.1%. The mean percent identity for ten randomized sequences of BFD is 17.5%, with the score for the authentic sequence being 9.85 standard deviation units from that mean. Monte Carlo scores for comparison of BFD with the related thiamin pyrophosphate requiring enzymes yielded scores in the range predictive for "probable homology" with ALS and "possible homology" with both POX and PDC. Percent identities for pairwise alignments of BFD with the three proteins ranged from 17.2% to 22.2%. Thus, the simultaneous alignment of BFD with all three enzymes suggests greater similarity than do any of the pairwise scores. ALS, POX, and PDC have already been shown to be related (Green, 1989), and it is important to note that the regions of high similarity between BFD and the three related proteins occur in stretches that are highly conserved when they are aligned without BFD. Although no crystallographic information is yet available for any thiamin pyrophosphate requiring enzyme, we suspect that the regions of similarity are associated with amino acid residues in the binding site for the cofactor. BFD is likely to be structurally related to other thiamin pyrophosphate dependent enzymes found in *P. putida* and, therefore, could have evolved by recruitment of the duplicated gene for an enzyme found in a metabolic pathway more widespread in the biosphere.

Considering that MR and MLE I are closely related structurally (Neidhart et al., 1990), it is of interest that MR shows less primary sequence relationship to either MLE I (Aldrich et al., 1987; N. J. Ornston, personal communication; K. Taylor, J. W. Kozarich, and J. A. Gerlt, unpublished observations) or MLE II (Aldrich et al., 1987) (the latter is specific for lactonization of 3-halomuconates) than either MDH or BFD to their homologues; the alignment is reproduced in Figure 6. The overall percent identity of MR with both MLEs is 27.6%. The mean percent identity for alignments with ten randomized sequences is 15.7%, with the score for the authentic sequence being 4.70 standard deviation units from that mean. Pairwise alignment of MR with MLE I and MLE II shows 25.6% and 21.9% identity, respectively. Monte Carlo pairwise analysis of these sequences predicts that MR and MLE I are "probably homologous" and that MR and MLE II are "possibly homologous". Given the significant structural similarity but modest level of sequence similarity between MR and MLE I, the more significant sequence similarities relating MDH and BFD and their homologues can be used as strong evidence for both structural and evolutionary relatedness.

On the basis of this conclusion that MR, MDH, and BFD are likely to be evolutionarily related to enzymes catalyzing chemically similar (MDH and BFD) or dissimilar reactions (MR), we are now attempting to isolate and identify genes for related enzymes in chromosomal DNA from *P. putida* ATCC 12633.

Table I: Summary of Purification of BFD from *E. coli* JM105 Transformed with pBFDtrc

purification step	total vol (mL)	total protein (mg)	total act. (units)	sp act. (units/mg)
crude extract	130	756	4662	6
$(NH_4)_2SO_4$ (75% sat.) pellet	28	280	4200	15
heat treatment supernatant	34	240	4200	17
DEAE-Sephacel	2.1	12	2262	189
Sephacryl S-200	3.5	10	1074	107
Mono Q[a]		3.2	148	46

[a] Calculated from the chromatography of a single 38-μL aliquot of the Sephacryl S-200 purified enzyme.

The Genes for BFD, MDH, and MR Constitute an Operon. The observations that *mdl*D and *mdl*E are separated from *mdl*C by at least 2 kb (Figure 3) and that *mdl*C, *mdl*B, and *mdl*A are transcribed away from the positions of *mdl*D and *mdl*E suggest that the five genes encoding the mandelate pathway are not organized in a single operon. In experiments to be described in detail elsewhere, nuclease S_1 mapping was performed to identify sites for initiation of transcription upstream of *mdl*C and for termination of transcription downstream of *mdl*A.

Initiation of transcription was observed 25 bp upstream of the initiation codon for *mdl*C (base 633 in the sequence in the supplementary material). Termination of transcription was observed approximately 59 bp downstream of the termination codon for *mdl*A (base 4565 in the supplementary material). This termination site was expected on the basis of the stable stem loop sequence AAAAGGCCCGC-TTAT-GCGGGCCTTTT observed downstream of *mdl*A (bases 4516–4541). Taken together, these nuclease S_1 mapping experiments demonstrate that *mdl*C, *mdl*B, and *mdl*A constitute the *mdl*CBA operon.

DNA sequence analysis of the remainder of the *Eco*RI fragment encoding the mandelate pathway is currently underway so that the sequences of both benzaldehyde dehydrogenases might be determined and further details of sequence homologies and transcriptional organization and regulation can be established.

Expression of BFD in E. coli. An *Nco*I restriction site was introduced at the initiation codon and several restriction sites (*Hin*dIII, *Xba*I, and *Bam*HI) that do not occur within the gene for BFD were placed downstream of the gene to allow directional cloning of the gene in a variety of expression vectors. We focused on the expression vector pKK233-2 that utilizes the inducible *trc* promoter and contains a unique *Nco*I site at the initiation codon and a unique downstream *Hin*dIII site, thereby allowing directional cloning of the gene behind the inducible promoter.

Purification of BFD from E. coli JM105 Transformed with pBFDtrc. The purification of BFD was based upon the procedure described by Hegeman (1970) for the enzyme from *P. putida* ATCC 12633. The manganese chloride treatment to remove nucleic acids was found to inactivate all enzymatic activity; this step was omitted without effect. The final chromatographic step, FPLC on a Mono Q column, was particularly effective in removing a contaminant having a similar polypeptide molecular weight. The purification is summarized in Table I, and a photograph of an SDS–PAGE gel after various steps in the purification is reproduced in Figure 7.

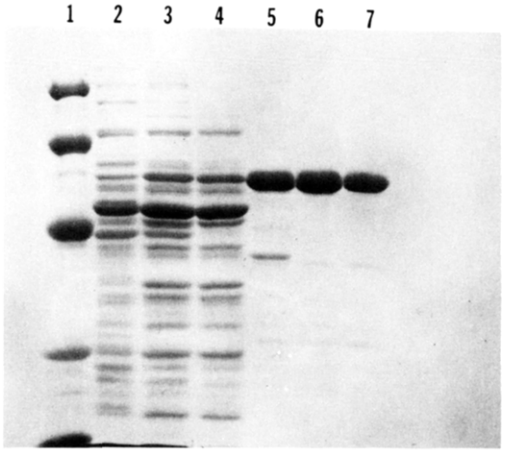

FIGURE 7: Photograph of a 10% SDS–PAGE gel showing the purification of BFD from *E. coli* JM105 transformed with pBFDtrc: lane 1, molecular weight standards (top to bottom: 94K, 67K, 43K, 30K, and 20.1K); lane 2, crude extract; lane 3, after precipitation with 75% $(NH_4)_2SO_4$; lane 4, after heat treatment; lane 5, after DEAE-Sephacel; lane 6, after Sephacryl S-200; lane 7, after Mono Q.

The enzyme so purified is greater than 95% homogeneous as judged by SDS–PAGE (Figure 7) and appears identical with the enzyme isolated from *P. putida*. The molecular weight estimated from SDS–PAGE, 57 000, is in excellent agreement with the value of 57 500 determined by the same procedure for the enzyme isolated from *P. putida* (Reynolds et al., 1988), although both are slightly larger than the value of 53 621 predicted from the DNA sequence. The first eight residues of the N-terminal sequence are identical with those found for enzyme isolated from *P. putida* and are those predicted from the sequence of pBFDtrc if the N-terminal methionine is removed.

The specific activity of the BFD isolated from *E. coli* (46 units/mg) exceeds that found for BFD isolated from *P. putida* (34 units/mg) (Reynolds et al., 1988). In addition, as documented in Table I, the specific activity of the enzyme after the DEAE-Sephacel chromatography is 189 units/mg and continues to decrease as contaminating proteins are removed; this phenomenon was previously noted by Weiss et al. (1988). The explanation for the decrease in specific activity upon purification may be associated with the observation that BFD rapidly loses activity upon dilution (Weiss et al., 1988); dilution of the enzyme during the latter stages of the purification is unavoidable. The enzyme purified by DEAE-Sephacel chromatography (189 units/mg) is sufficiently pure for most mechanistic investigations. Since the decrease in specific activity with purification may cause problems in obtaining crystalline enzyme, procedures to prevent this loss of activity are under investigation.

ACKNOWLEDGMENTS

We thank Professor Charles A. Fewson for the N-terminal sequence of MDH from *P. putida*, Dr. George A. Garcia for purifying BFD from *P. putida* for N-terminal sequence analysis, and Ms. Leslie Taylor of the UCSF Computer Graphics Laboratory for assistance with the sequence analyses. We also thank Dr. Vincent M. Powers for performing assays on cellular extracts, which demonstrated MDH activity encoded by pSCR1.

SUPPLEMENTARY MATERIAL AVAILABLE

The DNA sequence encoding the *mdl*CBA operon, the 5′- and 3′-flanking sequences, and deduced amino acid sequences of BFD (*mdl*C), MDH (*mdl*B), and MR (*mdl*A) (6 pages). Ordering information is given on any current masthead page.

REFERENCES

Aiba, H., Adhya, S., & de Crombrugghe, B. (1981) *J. Biol. Chem. 256*, 11905.
Aldrich, T. L., Frantz, B., Gill, J. F., Kilbane, J. J., & Chakrabarty, A. M. (1987) *Gene 52*, 185.
Altschul, S. F., & Erickson, B. W. (1986) *Bull. Math. Biol. 48*, 603.
Cederlund, E., Lindqvist, Y., Söderlund, G., Bränden, C. I., & Jörnvall, H. (1988) *Eur. J. Biochem. 173*, 523.
Ebina, Y., & Nakazawa, A. (1983) *J. Biol. Chem. 258*, 7072.
Fewson, C. A. (1988) *FEMS Microbiol. Rev. 54*, 85.
Garcia, G. A. (1987) Ph.D. Dissertation, University of California, San Francisco.
Grabau, C., & Cronan, J. E. (1986) *Nucleic Acids Res. 14*, 5449.
Green, J. B. A. (1989) *FEBS Lett. 246*, 1.
Guiard, B. (1985) *EMBO J. 4*, 3265.
Hegeman, G. D. (1966) *J. Bacteriol. 91*, 1140.
Hegeman, G. D. (1970) *Methods Enzymol. 17A*, 674.
Hoey, M. E., Allison, N., Scott, A. J., & Fewson, C. A. (1987) *Biochem. J. 248*, 871.
Kellerman, E., Seeboth, P. G., & Hollenberg, C. P. (1986) *Nucleic Acids Res. 14*, 8963.
Lin, D. T., Powers, V. M., Reynolds, L. J., Whitman, C. P., Kozarich, J. W., & Kenyon, G. L. (1990) *Biochemistry* (in press).
Lindqvist, Y. (1989) *J. Mol. Biol. 209*, 151.
Lindqvist, Y., & Bränden, C.-I. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 6855.
Neidhart, D. C., Kenyon G. L., Gerlt, J. A., & Petsko, G. A. (1990) *Nature* (in press).
Ransom, S. C., Gerlt, J. A., Powers, V. M., & Kenyon, G. L. (1988) *Biochemistry 27*, 540.
Reynolds, L. J., Garcia, G. A., Kozarich, J. W., & Kenyon, G. L. (1988) *Biochemistry 27*, 5530.
Smith, R. F., & Smith, T. F. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 118.
Squires, C. H., DeFelice, M., Devereux, J., & Calvo, J. M. (1983) *Nucleic Acids Res. 11*, 5299.
Tsou, A. Y., Ransom, S. C., Gerlt, J. A., Powers, V. M., & Kenyon, G. L. (1989) *Biochemistry 28*, 969.
Volokita, M., & Somerville, C. R. (1987) *J. Biol. Chem. 262*, 15825.
Weiss, P. M., Garcia, G. A., Kenyon, G. L., Cleland, W. W., & Cook, P. F. (1988) *Biochemistry 27*, 2197.
Wheelis, M. L., & Stanier, R. Y. (1970) *Genetics 66*, 245.
Xia, Z.-X., Shamala, N., Bethge, P. H., Lim, L. W., Bellamy, H. D., Xuong, N. H., Lederer, F., & Mathews, F. S. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 2629.