JOURNAL OF BACTERIOLOGY, Nov. 1989, p. 6084–6092
0021-9193/89/116084-09$02.00/0

Vol. 171, No. 11

Divergent Transcription of *pdxB* and Homology between the *pdxB* and *serA* Gene Products in *Escherichia coli* K-12

PATRICIA V. SCHOENLEIN,† BENJAMIN B. ROA, AND MALCOLM E. WINKLER*
Department of Molecular Biology, Northwestern University Medical School, Chicago, Illinois 60611

Received 19 May 1989/Accepted 10 August 1989

We report the DNA sequence and in vivo transcription start of *pdxB*, which encodes a protein required for de novo biosynthesis of pyridoxine (vitamin B_6). The DNA sequence confirms results from previous minicell experiments showing that *pdxB* encodes a 41-kilodalton polypeptide. RNase T2 mapping of in vivo transcripts and corroborating experiments with promoter expression vector pKK232-8 demonstrated that the *pdxB* promoter shares its −10 region with an overlapping, divergent promoter. Thus, *pdxB* must be the first gene in the complex *pdxB-hisT* operon. The steady-state transcription level from these divergent promoters, which probably occlude each other, is approximately equal in bacteria growing in rich medium at 37°C. The divergent transcript could encode a polypeptide whose amino-terminal domain is rich in proline and glutamine residues. Similarity searches of protein data bases revealed a significant number of amino acid matches between the *pdxB* gene product and D-3-phosphoglycerate dehydrogenase, which is encoded by *serA* and catalyzes the first step in the phosphorylated pathway of serine biosynthesis. FASTA and alignment score analyses indicated that PdxB and SerA are indeed homologs and share a common ancestor. The amino acid alignment between PdxB and SerA implies that PdxB is a 2-hydroxyacid dehydrogenase and suggests possible NAD^+, substrate binding, and active sites of both enzymes. Furthermore, the fact that 4-hydroxythreonine, a probable intermediate in pyridoxine biosynthesis, is structurally related to serine strongly suggests that the *pdxB* gene product is erythronate-4-phosphate dehydrogenase. The homology between PdxB and SerA provides considerable support for Jensen's model of enzyme recruitment as the basis for the evolution of different biosynthetic pathways.

Two different models have been proposed for the evolution of bacterial genes that mediate the diverse pathways of intermediary metabolism. Horowitz's model of retrograde evolution suggests that the genes involved in each metabolic pathway arose by duplication of a single ancestral gene followed by divergence to form a reaction chain with overlapping specificities (28). This stepwise, sequential evolution of new enzymes in reverse order allows the first enzyme in a pathway to carry a "memory of its origin," which would account for feedback inhibition. By contrast, Jensen proposed that substrate ambiguity is the basis of pathway evolution (29). According to his model of gene recruitment, accumulation of minor amounts of erroneous enzymatic products favored gene duplications, which ultimately mutated and gave rise to fixed, evolutionary homologs encoding enzymes with different activities.

Recent molecular biological comparisons tend to support gene recruitment much more strongly than retrograde evolution. The incisive evolutionary analysis completed by Nichols and co-workers of the related genes that mediate the biosynthesis of anthranilate (*o*-aminobenzoate) and *p*-aminobenzoate strongly favors evolution by recruitment of enzymatic functions (23, 24, 30, 31). In this case, *trpE* and *pabB* or *trp(G)D* and *pabA* arose from common ancestral genes, and *p*-aminobenzoate synthase (PabB-PabA) and anthranilate synthase (TrpE-TrpG domain) carry out related reactions on the same starting substrates, chorismate and glutamine. In addition, glutamine amidotransferases, such as *TrpG* and *PabA*, are members of a broader enzyme family that uses the amide group of glutamine in the biosynthesis of various compounds (48).

More support for Jensen's model of gene recruitment comes from several other experimental systems. The common ancestry of the IlvI and IlvG isozymes suggests a transposition duplication followed by divergence (42). The notion that gene products with related functions arose by recruitment of activity can explain the origin of the LysR family of activator proteins, which bind to DNA in response to different effectors (25), the two-component regulatory systems, which respond to different environmental stimuli (36), and the cryptic evolved β-galactosidase (*ebg*) genes (44), which are functionally and evolutionarily related to *lacI*, *lacY*, and *lacZ* (45). Moreover, the finding that two consecutive steps in methionine biosynthesis are catalyzed by enzymes (MetB and MetC) that share a common ancestor, which was originally interpreted to support retrograde evolution (7), can be reinterpreted in terms of Jensen's recruitment model. Two consecutive steps in isoleucine biosynthesis are also catalyzed by enzymes (ThrC and IlvA) that might be homologous (38). In addition, ThrC and IlvA may be evolutionarily related to DsdA, which is a catabolic deaminase for D-serine (38). Finally, acquisition of novel pathways for the utilization of xylitol, 1,2-propanediol, or amides is consistent with recruitment of enzymatic functions and changes in metabolic regulation as bases for the evolution of new catabolic pathways (reviewed in reference 32).

Despite these advances, relatively little is known about the evolutionary processes that gave rise to biosynthetic pathways that use different starting substrates or produce different products. In this paper, we report that the *pdxB* gene product, which is required for de novo biosynthesis of pyridoxine (vitamin B_6), is an evolutionary homolog of

* Corresponding author.
† Present address: National Cancer Institute, Bethesda, MD 20892.

D-3-phosphoglycerate dehydrogenase, which is encoded by *serA* and catalyzes the first step in the major phosphorylated pathway of serine biosynthesis (43). Pyridoxine is the pyridine-ringed precursor of pyridoxal phosphate (21), a coenzyme essential to every branch of amino acid metabolism (8). This homology implies that PdxB is a 2-hydroxyacid dehydrogenase, indicates possible NAD^+ binding, substrate binding, and active sites of both enzymes, and predicts the substrate used by the PdxB enzyme. The results presented here also demonstrate that *pdxB* is the first gene in the complex *pdxB-hisT* operon and is transcribed from a divergent promoter. The evolutionary relationships between PdxB and SerA are considered in terms of Jensen's model, and they lend strong support to enzyme recruitment as the basis for the evolution of different biosynthetic pathways.

MATERIALS AND METHODS

Materials. Restriction endonucleases, T4 DNA ligase, T4 DNA polymerase, and M13mp18 and M13mp19 phage vectors (47) were purchased from New England BioLabs, Inc. (Beverly, Mass.). Plasmid pKK232-8 (13), deoxynucleotide triphosphates, dideoxynucleotide triphosphates, and single-stranded 17-mer sequencing primer were bought from Pharmacia Fine Chemicals (Piscataway, N.J.). Culture media were from Difco Laboratories (Detroit, Mich.) and Fisher Scientific Co. (Fair Lawn, N.J.). Antibiotics and other biochemical reagents were purchased from Sigma Chemical Co. (St. Louis, Mo.), and inorganic chemicals were from Fisher. [α-^{32}P]dCTP (>800 Ci/mmol), [α-^{32}P]dATP (>3,000 Ci/mmol), and [α-^{32}P]CTP (>400 Ci/mmol) were bought from Amersham Corp. (Arlington Heights, Ill.). Deaza-dGTP was obtained from Boehringer-Mannheim Biochemicals (Indianapolis, Ind.). DNA polymerase I large (Klenow) fragment and Sequenase kits were purchased from U.S. Biochemical Corp. (Cleveland, Ohio). Reagents in gels used to resolve DNA and RNA included ammonium persulfate and acrylamide (Bio-Rad Laboratories, Richmond, Calif.), formamide (Fisher), and urea (Sigma). The Riboprobe Gemini system, pGEM-4Z cloning vector, and RQ1 DNase (RNase free) were purchased from Promega Biotec (Madison, Wis.). RNase T2 was from Bethesda Research Laboratories, Inc. (Gaithersburg, Md.).

Plasmids, bacterial strains, and media. Plasmid pNU93 was the source of the DNA used in this study (3). The *Escherichia coli* K-12 chromosomal insert in plasmid pNU93 was originally from Clarke-Carbon plasmid pLC28-44 (35). Subcloning fragments from pNU93 into M13 RFII vectors and plasmids pKK232-8 and pGEM-4Z was accomplished by standard techniques (4, 33). When necessary, restriction fragments with noncompatible ends were filled in with Klenow enzyme (5′ overhangs) or chewed back with T4 DNA polymerase (3′ overhangs) to give blunt-ended fragments suitable for cloning (4). M13 phage clones used in DNA sequencing and the *Hin*dIII-*Taq*I fragment subclone in pGEM-4Z, which was designated pNU200 and used as a template for RNA probes, are depicted in Fig. 1. Derivatives of the promoter expression vector, pKK232-8, are listed and drawn in Fig. 4.

E. coli K-12 W3110 prototrophic strain NU426 (3) was used as the source of total RNA in transcript mapping experiments. Derivatives of plasmid pKK232-8 were contained in strain NU637 (*E. coli* K-12 W3110 *srl*::Tn*10 recA1*). Plasmid pNU200 was propagated in strain JM109 [*recA1 endA1 gyrA96 thi hsdR17 supE44 relA1* Δ(*lac proAB*) (F′ *traD36 proAB*$^+$ *lacI*^{q}*ZΔM15*)] (47).

Bacterial strains were grown in LB medium (18) supplemented with 30 μg of cysteine per ml (LB-Cys) at 37°C. Ampicillin was added at 50 μg/ml. Resistance to chloramphenicol (Cm^r) was tested on a series of LB-Cys plates containing 3.1, 6.3, 12.5, 25, 50, 75, and 100 μg of chloramphenicol per ml. Cm^r was scored after 24 and 72 h of incubation at 37°C.

Molecular biological methods. DNA sequences were determined by a modification of the Sanger method, which included 7-deaza-dGTP in elongation reaction mixes, and by a variation of the Sequenase kit reactions, which included dITP and single-stranded binding protein in reaction mixes, as described previously (15, 40). The DNA sequence of each subcloned fragment was determined independently by both methods. DNA products from sequencing reactions were resolved on 80-cm 6% polyacrylamide gels containing TBE buffer, 8 M urea, and 20% (vol/vol) formamide (15, 40). DNA sequences were analyzed by using the UWGCG (University of Wisconsin), PCGene (Intelligenetics, Inc., Mountain View, CA), and BIONET computer programs. Similarity searches of the GenBank and EMBL nucleotide and NBRF/PIR and SWISS-PROT protein data bases were performed by using the FASTA program (39) on the BIONET system. Alignment score analysis (19, 22) was performed by using the PCGene PCOMPARE program.

RNase T2 mapping of in vivo transcripts was completed essentially as described previously (15, 40). Total cellular RNA was prepared by a variation of the sodium dodecyl sulfate-hot phenol method and treated with DNase (RNase free). RNA probes corresponding to both DNA strands were synthesized by using SP6 or T7 RNA polymerase according to instructions provided with the Riboprobe Gemini system. Hybridization reactions and RNase T2 treatments were performed as described previously (16), with the following changes: (i) 50 μg of total RNA and 250,000 cpm (Cerenkov) of RNA probe were used per hybridization reaction; (ii) hybridization reactions were incubated overnight at 50°C; and (iii) 300 μl of an RNase T2 stock solution (120 U/ml) was added to each hybridization reaction. Protected segments of the probes were resolved on 40-cm gels as described above.

RESULTS

DNA sequence of *pdxB*. Previously, we established that the *pdxB* biosynthetic gene is located in the same complex operon upstream from *hisT*, which encodes the important modification enzyme tRNA pseudouridine synthase I (Fig. 1) (2, 3). The *pdxB-hisT* operon contains at least three other unusual genes (Fig. 1), including the following: *asd′*(*usg-1*), which is located between *pdxB* and *hisT* and encodes a polypeptide related to *Streptococcus mutans* aspartate semialdehyde dehydrogenase (1, 26); *dsg1*(*dedA*), which is immediately downstream from *hisT* (2, 37) and encodes an integral membrane protein (P. J. Arps and M. Winkler, unpublished data); and *dedB*, which is immediately downstream from *dsg-1*(*dedA*) (12, 37) and encodes a polypeptide that is evolutionarily related to certain chloroplast polypeptides (26), contains a zinc finger DNA-binding motif (12), is probably membrane associated (12), seems to act as a pleiotropic regulator (D. M. Connolly and M. Winkler, unpublished), and may play a role in maximum expression of the folate biosynthetic gene, *folC* (11, 12). In fact, data from plasmid constructs suggest that *folC* may be the last gene in the *pdxB-hisT* operon (11).

We wanted to learn the role of *pdxB* in pyridoxal phosphate biosynthesis and whether *pdxB* is the first gene in the

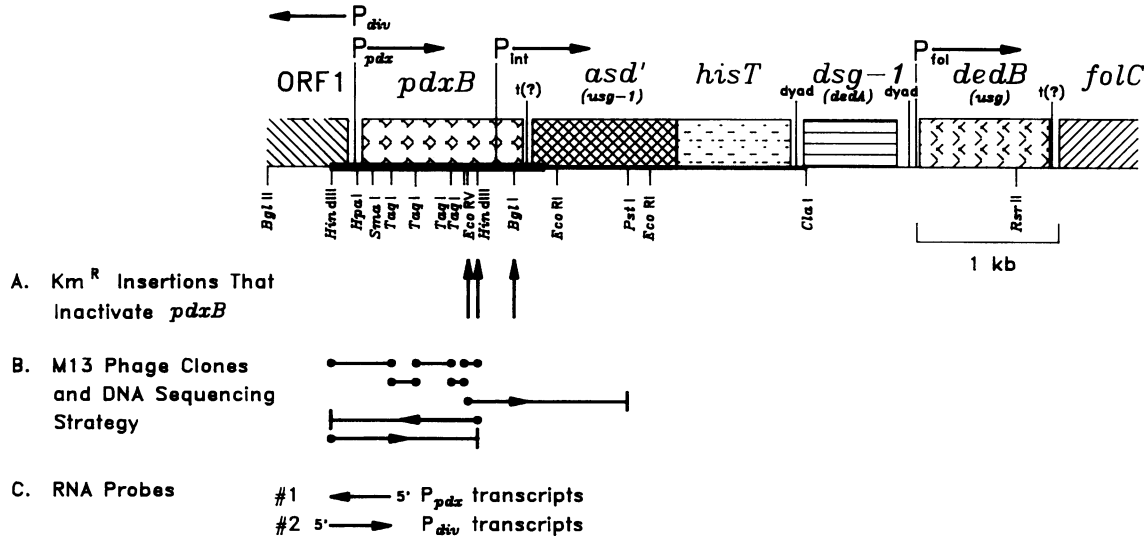

FIG. 1. Structure of the complex *pdxB-hisT* operon of *E. coli* K-12. The region analyzed in this paper is marked by the thick black line under *pdxB*. Sequences to the right of *pdxB* were reported previously in the following references, where names in parentheses indicate alternate designations for genes: *asd'*(*usg-1*) and *hisT* (1); dsg-1(*dedA*) (3, 37); and *dedB*(*usg*) and *folC* (12, 37). The figure is drawn to scale. Reading frames of genes are marked by shaded rectangles, and their functions are described in the text. The full extent of the ORF1 reading frame is based on preliminary data not reported here. The promoters upstream from *pdxB* (P_{pdx}), divergent and overlapping P_{pdx} (P_{div}), and internal to *pdxB* (P_{int}) were located in the chromosome by transcript mapping (Fig. 2 through 5) (1, 3), whereas the promoter upstream from *dedB*(*usg*) and *folC* (P_{fol}) was located on plasmid constructs (11). Strong dyad symmetries and possible weak rho-independent terminators are marked in intercistronic regions by dyad and t(?), respectively. The level of readthrough past these putative transcription terminators is unknown in the bacterial chromosome. (A) Positions of chromosomal kanamycin resistance cassette (Km^R) insertions that inactivate *pdxB* function (2, 3). (B) M13 phage clones used in the DNA sequencing strategy. Fragments sequenced completely on both DNA strands are indicated (●——●). Fragments sequenced on one strand from the closed circles to beyond the arrowheads were used to overlap cloning sites. (C) RNA probes used in RNase T2 mapping of in vivo transcripts (see Fig. 5). The *Hin*dIII-*Taq*I fragment indicated by the extent of the arrows was ligated into plasmid pGEM-4Z that had been cut with *Hin*dIII and *Hin*cII, and RNA probes 1 and 2 were synthesized from linearized plasmid using SP6 and T7 RNA polymerase, respectively, in an in vitro transcription reaction containing [α-^{32}P]CTP (see Materials and Methods).

unusual *pdxB-hisT* operon. Therefore, we determined the sequence of *pdxB* on both DNA strands according to the strategy in Fig. 1B, with overlapping of all restriction sites used for subcloning. The complete sequence of the region corresponding to the blackened thick line in Fig. 1 is presented in Fig. 2. For consistency with our earlier work (3), the *Hin*dIII site at the start of the sequence is numbered 440. Assignment of the probable translation start of *asd'*(*usg-1*) at position 1857, description of the strong dyad symmetry centered at position 1819 in the *pdxB-asd'*(*usg-1*) intercistronic region, and localization of the relatively strong P_{int} internal promoter were presented previously (1, 3).

Identification of the long open reading frame between nucleotides 657 and 1793 as the *pdxB* gene product is based on four considerations. First, this is the only long open reading frame in the region limited by the transcription analysis presented below. Second, the reading frame is the only one that can account for loss of *pdxB* function by chromosomal insertions into the *Eco*RV, *Hin*dIII, or *Bgl*I sites marked in Fig. 1A. Consistent with this interpretation, mini-Mu dl(Km^r) insertions into the bacterial chromosome near the base numbered 840 in Fig. 2 also inactivate *pdxB* and disrupt the same long open reading frame as the Km^R cassette insertions (H.-M. Lam and M. Winkler, unpublished result). Third, this is the only reading frame that encodes a polypeptide with a molecular mass of about 42 kilodaltons, which matches the size of PdxB polypeptide synthesized in minicells (2, 3). Fourth, as described below, the PdxB polypeptide shown in Fig. 2 contains amino acid identities over its entire length with the *serA* gene product, an enzyme of known function. Taken together, these observations prove that the open reading frame indicated in Fig. 2 must correspond to the *pdxB* gene product.

The assignment of the translation start of PdxB cannot be made with as great a level of certainty. Figure 3 presents an expanded view of the region at the start of PdxB. The most likely translation starts are the GUG at position 657 or the AUG at position 681, which result in polypeptides with molecular masses of 41,335 (the maximal possible size) or 40,424 daltons, respectively. This range matches the 42-kilodalton molecular mass of PdxB on sodium dodecyl sulfate denaturing gels. There are two additional reasons to favor these possible translation start sites instead of sites further downstream, such as GUG (position 720) (Fig. 3). First, the alignment between PdxB and SerA seems to begin before the Val residue coded by GUG (position 720), although the alignment in this region is somewhat weak (see below). Second, there is a sequence and a direct repeat [designated by (*pdx*) and arrows, respectively, the Fig. 3] that is also found near the likely translation start of *pdxA* (40). The *pdxA* gene encodes a polypeptide involved in a different branch of pyridoxine biosynthesis that PdxB (40). Since the *pdx* sequence is not common in the bacterial database of GenBank, it is possible that this sequence and the direct repeat play a role in translational regulation of pyridoxine biosynthesis; however, this model remains to be tested. At any rate, together these considerations suggest that translation of PdxB begins at GUG (position 657) or AUG (position 681). None of the possible translation start codons is preceded by a particularly good match to the

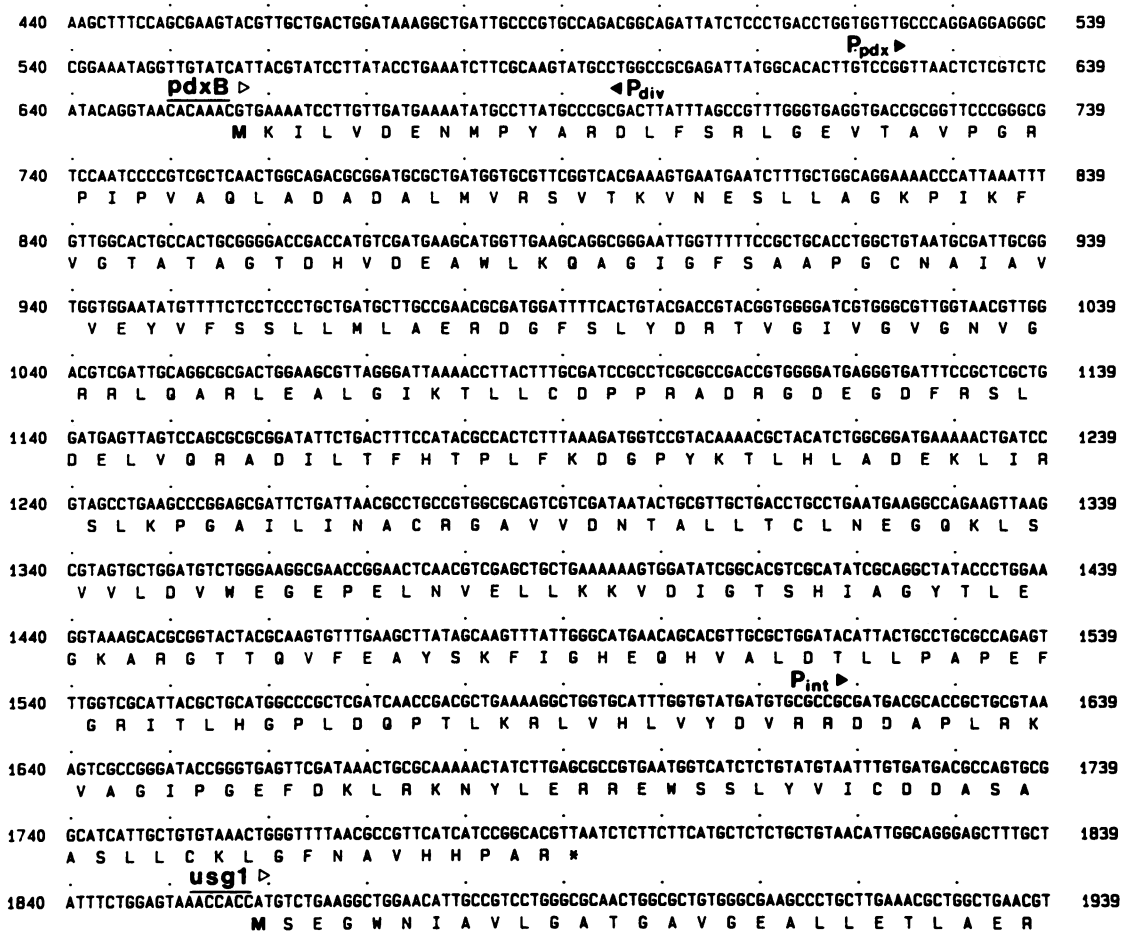

FIG. 2. DNA sequence of *pdxB* and predicted amino acid sequence of the *pdxB* gene product. The DNA sequence was determined by the strategy in the legend to Fig. 1A, and only the DNA strand containing the *pdxB*-coding strand is shown. The *Hin*dIII site at the start of the sequence is numbered 440 for consistency with our previous restriction map (3). Evidence that the reading frame from positions 657 through 1790 is *pdxB* is presented in Results. The P_{int} transcription start site at position 1618, the *pdxB-asd'(usg-1)* intercistronic region, and the start of *asd'(usg-1)* were identified earlier (1, 3). The P_{pdx} and P_{div} transcription start sites at positions 624 and 596 on the opposite strand, respectively, are based on data presented in Fig. 4 and 5 and are shown in detail in Fig. 3.

Shine-Dalgarno sequence, although one candidate is marked for the putative GUG (position 657) start (Fig. 3).

Transcription start of *pdxB*. To determine whether *pdxB* is the first gene in the operon, we cloned the fragments depicted by lines in Fig. 4 into the promoter expression vector pKK232-8 (13). We transformed these plasmid constructs into a chloramphenicol-sensitive (Cm^s) strain, W3110 *srl*::Tn*10 recA1*, and selected for the presence of plasmids by resistance to ampicillin on LB medium at 37°C. The resulting strains were then streaked onto LB plates or inoculated into liquid LB medium containing concentrations of chloramphenicol that ranged between 3 and 100 μg/ml. Plates or cultures were incubated for 24 to 72 h at 37°C before being scored for resistance to chloramphenicol, which indicates the presence of a promoter in the cloned insert.

Results of resistance to 25 μg of chloramphenicol per ml are summarized in Fig. 4, although consistent results were obtained at other antibiotic concentrations. A promoter is likely present in the 187-base-pair *Hin*dIII (position 440)-*Hpa*I (position 627) fragment upstream from *pdxB* (Fig. 4). Internal deletion of the *Hpa*I (position 627)-*Sma*I (position 734) fragment seemed to decrease, but did not eliminate, the activity of this promoter (Fig. 4); however, rigorous controls for changes in plasmid copy numbers were not performed. In addition, Fig. 4 clearly shows the presence of a divergent promoter in the *Hin*dIII (position 440)-*Sma*I (position 734) fragment. In this case, deletion up to the *Hpa*I (position 627) site decreased or eliminated transcription from the divergent promoter (Fig. 4). Thus, the *Hin*dIII (position 440)-*Hpa*I (position 627) fragment contains closely spaced or overlapping divergent promoters. This finding strongly implies that *pdxB* is the first gene in the *pdxB-hisT* operon.

We next completed RNase T2 mapping of chromosomal transcripts to locate the transcription start sites from these divergent promoters. In this experiment, we isolated total RNA from a W3110 prototroph strain growing exponentially in LB medium at 37°C, hybridized total RNA to the ^{32}P-labeled RNA probes depicted in Fig. 1C, treated hybrids with RNase T2, and resolved protected segments on urea-formamide gels as described in Materials and Methods. Results of a typical experiment are presented in Fig. 5. We used RNA size standards instead of DNA sequencing ladders, because we noticed discrepancies between sizes based on DNA standards and known lengths of certain longer RNA molecules. We detected a single 252-nucleotide protected segment (lane 5) and a closely spaced 157-nucleotide protected doublet (lane 4) corresponding to *pdxB* and divergent (*div*) chromosomal transcription, respectively (Fig. 5). This

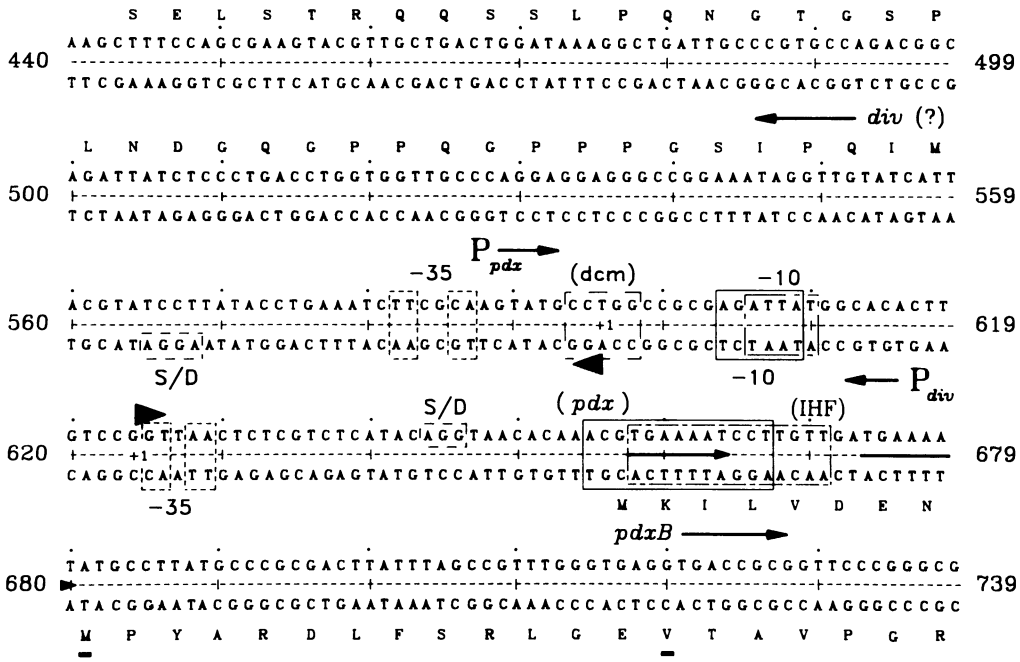

FIG. 3. Detail of the region containing the divergent P_{pdx} and P_{div} promoters. The start sites of transcription from P_{pdx} and P_{div} are indicated by black triangles at positions 624 and 596 on the opposite DNA strand, respectively. Probable −35 regions and the shared −10 regions of P_{pdx} and P_{div} are boxed. The most probable translation start site in *pdxB* at position 657 is preceded by a partial Shine-Dalgarno sequence (S/D). Less likely translation start sites at positions 681 and 720 are underlined (see Results). The ORF1 reading frame of a possible divergent (*div*) gene product starts on the opposite DNA strand at position 558, and preliminary sequence data suggest that it can extend for at least 189 amino acids (data not shown). The conserved (*pdx*) sequence (bases 655 through 667) and direct repeat (arrows starting at position 658) found near the translation starts of both *pdxB* and *pdxA* (40) may play roles in coordination or control of *pdx* gene expression at the translation level (see the text). A *dcm* DNA methylation site in P_{pdx} (positions 594 through 598) (34) and a sequence resembling an integration host factor (IHF)-binding site (positions 658 through 671) (17) are also marked, although the significance, if any, of these sites to *pdxB* or *div* expression is unknown.

result proves that *pdxB* is indeed the first gene in the complex *pdxB-hisT* operon.

The in vivo transcription start sites indicated by the lengths of the protected segments in Fig. 5 are marked in Fig. 2 and 3. We estimate that the error in this experiment is ±2 nucleotides. Figure 3 shows that P_{pdx} and P_{div} overlap and share the same −10 region (positions 604 to 610). This placement of P_{pdx} and P_{div} is consistent with the promoter expression vector experiment shown in Fig. 4. Deletion of the *Hpa*I (position 627)-*Sma*I (position 734) fragment occurs immediately after the transcription start and removes the translation start of *pdxB*. This deletion may destabilize the fused *pdxB-cat* transcript, since it is well known that untranslated stretches of mRNA are often relatively unstable (14). In the case of the divergent transcript, the *Hpa*I (nucleotide 627) site is in the middle of the −35 region. Therefore, constructions at this site may reduce or eliminate transcription from P_{div}, depending on whether new adjoining sequences can restore the −35 region. Finally, densitometer tracings of lanes 4 and 5 in Fig. 5 indicate that the steady-state levels of transcription from P_{pdx} or P_{div} were nearly equal in bacteria growing in rich medium at 37°C. This observation again matches results from promoter expression vector experiments, in which similar antibiotic resistance levels were imparted by the presence of P_{pdx} or P_{div} in pKK232-8 (Fig. 4; data not shown). Other features of P_{pdx} and P_{div} and the region transcribed divergently from *pdxB* (designated *div* or ORF1 in Fig. 2, 3, and 4) are considered in the Discussion.

Evolutionary relationship between PdxB and SerA. Searches of the NBRF/PIR and SWISS-PROT protein data

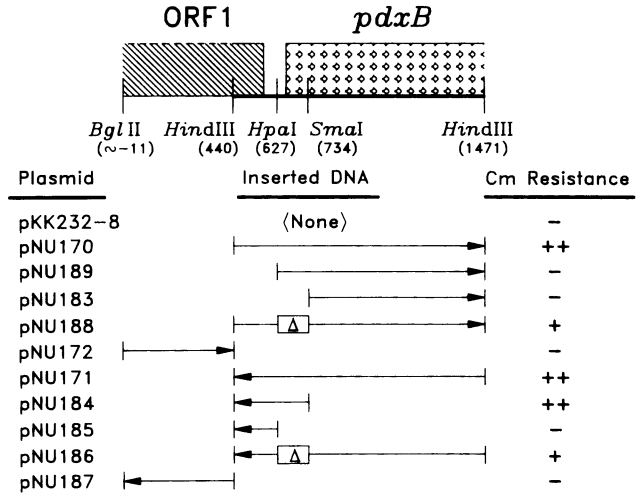

FIG. 4. Localization of promoters upstream from *pdxB* by expression in vector pKK232-8. The arrows indicate the DNA fragments and their orientations upstream of the promoterless *cat* gene in pKK232-8 (13). Internal deletions in plasmids pNU188 and pNU186 are indicated by rectangles. The plasmid constructs were transformed into the *recA1* mutant NU637, and the resulting strains were streaked onto LB-Cys plates containing 25 μg of chloramphenicol per ml. Chloramphenicol resistance was scored after incubating the plates for 24 and 72 h at 37°C. The experiment was repeated several times. Similar qualitative results were obtained at chloramphenicol concentrations between 12.5 and 75 μg/ml in LB-Cys plates and in liquid cultures (data not shown). −, No growth; +, small single colonies; ++, large single colonies.

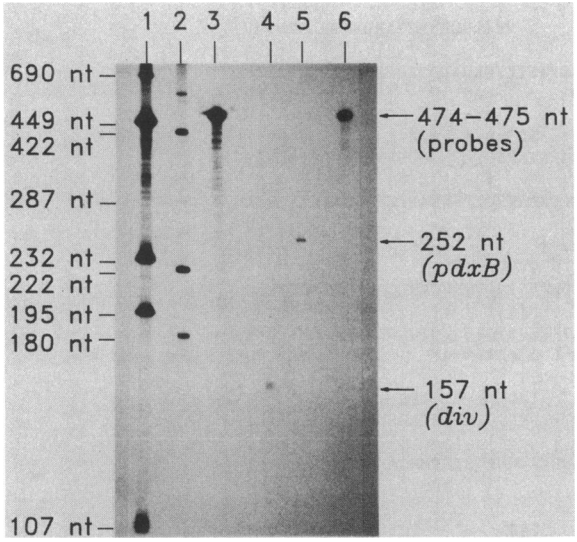

FIG. 5. RNase T2 mapping of transcripts initiated from P_{pdx} and P_{div} in the *E. coli* chromosome. Total RNA, isolated from the W3110 prototroph NU426, grown in LB-Cys liquid medium with shaking at 37°C, was hybridized with RNA probe 1 or 2 (Fig. 1C). RNA-RNA hybrids were treated with RNase T2 to digest unpaired regions, and protected segments were resolved on urea-formamide-polyacrylamide gels as described in Materials and Methods. Lanes: 1 and 2, RNA size standards; 3 and 6, starting probes 2 (475 nucleotides) and 1 (474 nucleotides), respectively; 4, protected segment of probe 2, corresponding to transcription from P_{div}; 5, protected segment of probe 1, corresponding to transcription from P_{pdx}. The experiment was performed twice with different RNA preparations.

bases revealed a remarkable number of amino acid identities and conservative matches between the *pdxB* gene product and D-3-phosphoglycerate dehydrogenase (Fig. 6), which is encoded by *serA* (43, 46). SerA is a 2-hydroxyacid dehydrogenase that uses NAD^+ in the conversion of 3-phosphoglycerate to 3-phosphohydroxypyruvate (Fig. 7). This reaction is the first step of the major phosphorylated pathway of serine biosynthesis, and bacterial SerA enzymatic activity is subject to allosteric feedback inhibition by L-serine (43, 46).

We used two criteria to verify that the amino acid matches between PdxB and SerA are statistically significant enough to indicate shared ancestry. First, the FASTA algorithm (39), which flagged the similarity in the first place, gave initial and optimized match scores of 161 and 177, respectively, for the PdxB and SerA polypeptides, whereas the mean initial match score for PdxB and all polypeptides in the NBRF/PIR data base was 34 ± 9. Second, we applied the alignment score method, developed originally by Dayhoff and co-workers (19). In our calculation, the PdxB and SerA sequences were jumbled 100 times, and the comparisons were performed by using the Dayhoff MDM-78 matrix, a bias of 60, and a gap penalty of 3. The alignment score for PdxB and SerA was 7.305 (a score of 5 or 3 indicates significant or probable homology, respectively) (22). In contrast, we compared the sequence of PdxB or SerA to that of histidinol dehydrogenase (HisD), which Jensen speculated might be evolutionarily related to SerA (29). The alignment scores between PdxB and HisD or SerA and HisD were 0.98 or 0.44, respectively, which does not support Jensen's contention. In summary, these analyses show that PdxB and SerA are evolutionarily related and share a common ancestor. Additional comparisons between the PdxB and SerA enzymes and the *pdxB* and *serA* DNA sequences are made in the Discussion.

DISCUSSION

In this paper, we show three interesting features of *pdxB*, which encodes a protein required for pyridoxine (vitamin B_6) biosynthesis. First, *pdxB* is the first gene in the complex *pdxB-hisT* operon (Fig. 1 and 2). Second, the promoter for *pdxB* shares its −10 region with an overlapping divergent promoter (Fig. 2 through 5). In vivo, P_{pdx} and P_{div} seem to be about equal in strength in bacteria growing exponentially in LB medium (Fig. 4 and 5). Third, the *pdxB* gene product is evolutionarily related to the major serine biosynthetic enzyme, D-3-phosphoglycerate dehydrogenase, which is encoded by *serA* (Fig. 6).

Divergent, overlapping promoters are often used by pairs of genes that are related by regulation or function (6). For example, one divergent transcript may encode an enzyme, whereas the other encodes a regulatory protein that acts at the divergent promoter itself. Based on preliminary sequence information that extends about 450 base pairs to the left of the *Hin*dIII (position 440) site, the divergent transcript initiated in vivo at P_{div} contains a long open reading frame (ORF1, Fig. 1 and 4). The putative ORF1 polypeptide has an unusual amino terminus that is rich in Pro(P) and Gln(Q) residues (Fig. 3). However, we presently do not know whether the divergent transcript is actually translated or whether disruption of ORF1 affects expression of the *pdxB-hisT* operon. Ongoing experiments should address these issues.

Examination of the P_{pdx} and P_{div} promoters suggests that they probably function in a mutually exclusive manner, since both promoters use the same −10 region (Fig. 3). The promoter region lacks striking dyad symmetries indicative of protein binding sites, although imperfect dyads are centered at positions 603, 618, and 690. The promoter region also contains a *dcm* DNA methylation site (34) and a sequence similar to an integration host factor binding site (17) (positions 594 through 598 and 658 through 671, respectively; Fig. 3), although the significance of these sites in promoter function is unknown.

The evolutionary relationship between PdxB and SerA has several important implications. Since SerA is a 2-hydroxyacid dehydrogenase (46), then PdxB most likely belongs to this class of enzymes. Several observations support this contention. In their report of the SerA sequence, Tobey and Grant noted that very little sequence similarity has been found among different dehydrogenases (46), despite the fact that dehydrogenases seem to be ancient enzymes with conserved NAD^+ binding, substrate binding, and active sites (9, 10). Lack of amino acid identities prevented these authors from assigning Asp-X-X-Arg and His sites in SerA as probable substrate-binding and active sites (46). In addition, they assigned the Gly-X-(Gly or Ala)-X-X-G-≈20X-Asp sequence found in NAD^+-binding sites to positions 18 through 43 (Fig. 6), even though the His-20 residue does not fit the consensus sequence well (10, 46).

The general lack of sequence similarity among dehydrogenases makes the homology between PdxB and SerA that much more striking (Fig. 6), since amino acid matches extend over the whole length of both enzymes (Fig. 6). In addition, this comparison suggests possible locations of the NAD^+-binding, substrate-binding, and active sites of PdxB and SerA. The sequence suggested by Tobey and Grant as the conserved sequence in the SerA NAD^+-binding site may not be correct, because it occurs in an amino-terminal segment that is not conserved between PdxB and SerA (Fig. 6). Based on our new data, the sequence characteristic of NAD^+-binding sites (10) most likely occurs at the first highly

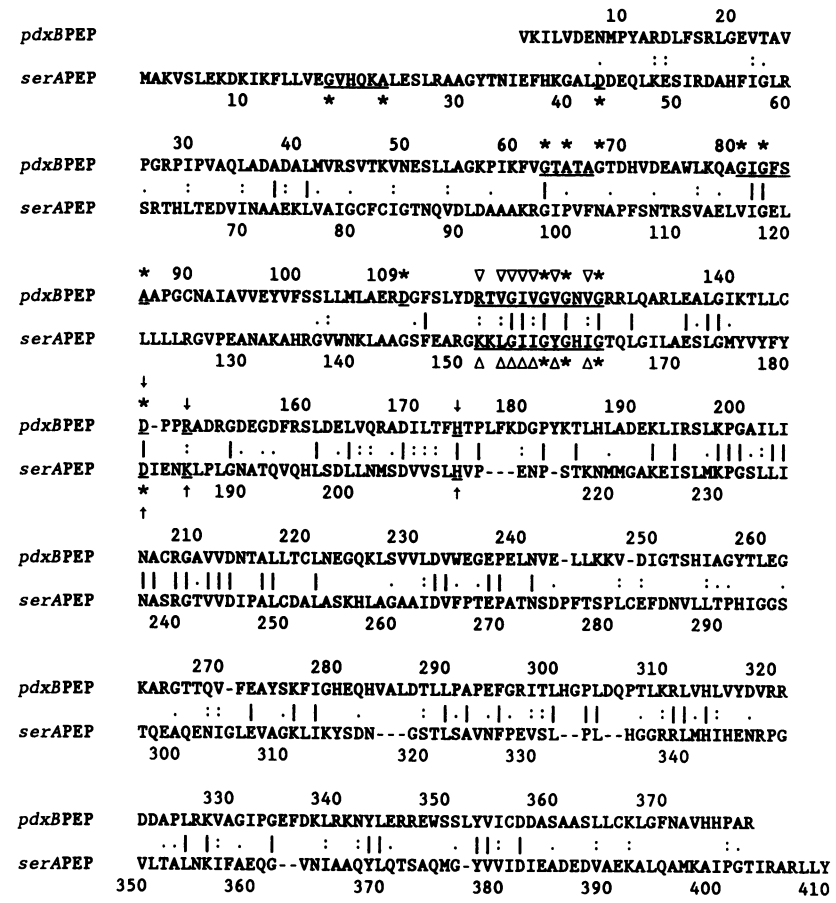

FIG. 6. Amino acid alignment between the *pdxB* gene product and D-3-phosphoglycerate dehydrogenase, which is encoded by *serA*. The maximized alignment was generated by the FASTA program of Pearson and Lipman (39). Amino acid identities are marked by lines (|) and strongly (=) or moderately (≈) conserved matches (T=S≈A≈G≈P; D=E≈N=Q; F=Y≈W; K=R≈H; and I=L=V≈M) are indicated by colons (:) or single dots (.), respectively. The alignment score for the two sequences is greater than 7, which indicates an extremely high probability that PdxB and SerA are evolutionarily related and share a common ancestor (see Discussion). Asterisks (*) below SerA (positions 18 through 43) mark amino acids that Tobey and Grant suggested are conserved in NAD^+-binding domains (10, 46). Based on the alignment, a more likely position for this conserved sequence is positions 117 through 146 in PdxB and 152 through 181 in SerA (* and open triangles for residues found in at least one other dehydrogenase). Other possible but less likely positions for this conserved sequence occur in PdxB between positions 63 and 110 (*). The characteristic His-174–Asp-146 pair and Asp-146–X–X–Arg-149 sequence found in the active site and substrate binding sites, respectively, of 2-hydroxyacid dehydrogenases (9) are indicated above the PdxB sequence by arrows. Corresponding residues in SerA are also marked by arrows (see Discussion).

conserved region of the two enzymes (PdxB at position 117 to SerA at position 152, Fig. 6). Other possible positions for the conserved sequence in the NAD^+ binding site occur in PdxB between residues 63 and 110 (Fig. 6). Unlike SerA, PdxB does contain a His-Asp pair found in the active sites of other 2-hydroxyacid dehydrogenases (positions 146 and 174, Fig. 6) and a characteristic Asp-X-X-Arg substrate binding motif (positions 146 and 149, Fig. 6) (9). In this regard, the sequence comparison shows that SerA shares these features in a slightly modified form; the His-Asp pair occurs at positions 181 and 210, and the Asp-X-X-Arg substrate-binding site is replaced by Asp-X-X-X-Lys at positions 181 to 185 (Fig. 6). Last, secondary structure predictions for the amino-terminal 150 amino acids of PdxB show the appropriate alternating β-strand and α-helical segments characteristic of the NAD^+-binding domain of other oxidoreductases (9, 10). These combined observations strongly suggest that PdxB is a 2-hydroxyacid dehydrogenase.

The latter finding allows us to make a prediction about the reaction catalyzed by PdxB. Despite careful tracer studies by Hill and Spenser and co-workers on *pdx* mutants isolated by Dempsey, the pathway for pyridoxine biosynthesis remains unknown (reviewed in references 20, 21, and 27). However, these combined biochemical and genetic analyses limit the possible intermediates and suggest some likely candidates for the major pyridoxine pathway. Dempsey suggested that 4-hydroxythreonine is a major intermediate, although the steps leading to this compound were unknown (21). It is striking that 4-hydroxythreonine and serine are structurally related in that the former compound contains one additional H-C-OH group. It is also noteworthy that the *serC* gene product, 3-phosphoserine aminotransferase, has long been suspected of playing roles in both serine and pyridoxine biosyntheses (20, 21). Based on these considerations, we suggest that PdxB carries out a reaction analogous to that of SerA, which converts 3-phosphoglycerate to 3-phosphohydroxypyruvate (Fig. 7). We propose that PdxB is erythronate-4-phosphate dehydrogenase and catalyzes the conversion of erythronate-4-phosphate to 3-hydroxy-4-phosphohydroxy-α-ketobutyrate (Fig. 7). The latter compound can then be converted directly into 4-hydroxythreonine by the action of SerC(PdxF) and a phosphatase. Support for this

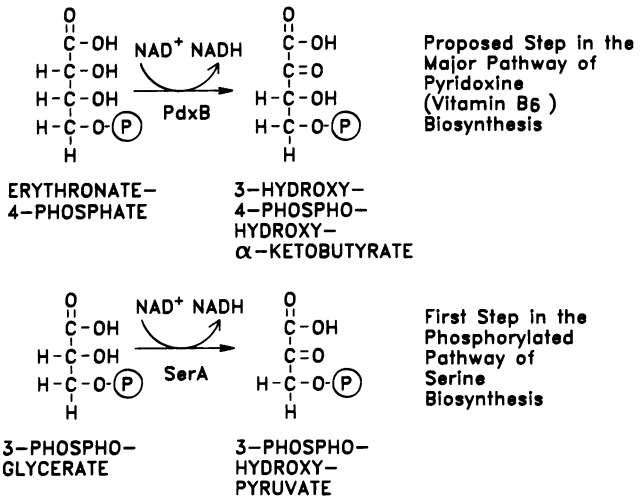

FIG. 7. Proposed step catalyzed by the *pdxB* gene product in the major pyridoxine (vitamin B_6) biosynthetic pathway. The product of this reaction can be converted directly into 4-hydroxythreonine, which is a likely precursor of pyridoxine (see Discussion). If this proposed step is correct, PdxB is erythronate-4-phosphate dehydrogenase.

hypothesis and other relationships between pyridoxine and serine biosynthesis will be reported in another article. (H.-M. Lam and M. Winkler, manuscript in preparation).

The homology between PdxB and SerA is consistent with Jensen's model of gene recruitment as the basis for the evolution of different biosynthetic pathways (see above). To our knowledge, this is one of the only examples that shows such a clear-cut evolutionary relationship between anabolic pathways. We suspect that both pyridoxine and serine were present in the environment of primordial cells. Sometime later, serine and other amino acids became scarce, which favored the evolution of amino acid-biosynthetic enzymes, including those, such as SerC, that use pyridoxal phosphate as a coenzyme. Still, the amount of pyridoxine required for growth was small enough ($\approx 10^{-7}$ M [21]) that the environmental supply was sufficient. Later, as the environmental supply of pyridoxine dwindled, part of the phosphorylated pathway of serine biosynthesis was recruited to synthesize precursors of pyridoxine. According to this scheme, minor activities of SerA and SerC led to the synthesis of the pyridoxine precursor 4-hydroxythreonine. Interestingly, even now, an activity of SerA seems to be required for the alternative pathway of pyridoxine biosynthesis detected in *pdxB* insertion mutants at 30°C (H.-M. Lam and M. Winkler, unpublished data). Eventually, the selective advantage of being able to synthesize pyridoxine de novo led to the duplication of *serA* and divergence of one copy into *pdxB*. Sometime in this process, a transposition also occurred so that *serA* and *pdxB* were located at different chromosomal positions (63 and 50 min, respectively [5]). By contrast, SerC already possessed or refined the activities necessary to function in both pathways, and no duplication of *serC* seems to have occurred.

Several points are noteworthy about this evolutionary scenario. The homology between the *serA*- and *pdxB*-encoded polypeptides is much more pronounced than that between the genes themselves. This apparent lack of homology at the nucleotide level is apparent with both FASTA and dot-matrix analyses (data not shown). Nevertheless, the maximum alignment of the *pdxB and serA* nucleotide sequences is most conserved in the same regions as in the polypeptide alignment (Fig. 6). Lack of homology at the nucleotide level may partly reflect the distinctly different codon usage in the two genes. The *serA* gene contains standard codon usage for *E. coli*, whereas *pdxB* contains a high proportion of infrequent and rare codons, characteristic of proteins expressed at relatively low levels (1). Previous minicell experiments are consistent with this codon usage pattern in that the *pdxB* polypeptide is expressed much less than certain other gene products encoded by the *pdxB-hisT* operon, such as the *asd'(usg-1)* polypeptide (3). In this regard, it is interesting that *pdxA* also contains many rare or infrequently used codons (40), which suggests that expression of some *pdx* gene products might be limited at the translational level by codon usage. Codon usage in bifunctional *serC(pdxF)* is standard and thus is similar to *serA* rather than to *pdxA* and *pdxB*. One last interesting feature of the nucleotide sequence comparison between *pdxB* and *serA* concerns the putative *pdx* translation-regulatory sequence, which is found near the translation starts of both PdxA and PdxB (see Results) (Fig. 3) (40). The *pdx* sequence occurs in a gap region of the comparison and is not present at all in the *serA* DNA sequence (data not shown).

Further enzymological characterization is needed to test the postulated role of PdxB in pyridoxine biosynthesis. The findings presented here should be of considerable significance to structure-function studies being carried out on SerA and other 2-hydroxyacid dehydrogenases (41, 46). In a sense, the PdxB enzyme is a naturally occurring structure-function variant of SerA. As part of these studies, it will be interesting to learn whether PdxB exists as a multimer and is subject to feedback inhibition by serine or other compounds, like its evolutionary homolog, SerA.

ACKNOWLEDGMENTS

We thank D. M. Connolly, W. B. Dempsey, G. A. Grant, H.-M. Lam, and B. P. Nichols for interesting discussions and critical comments.

This work was supported by Public Health Service grant GM37561 from the National Institute of General Medical Sciences. The BIONET computer system is supported by Public Health Service grant P41RR01685 from the National Institutes of Health.

LITERATURE CITED

1. **Arps, P. J., C. C. Marvel, B. C. Rubin, D. A. Tolan, E. E. Penhoet, and M. E. Winkler.** 1985. Structural features of the *hisT* operon of *Escherichia coli* K-12. Nucleic Acids Res. **13**:5297–5315.
2. **Arps, P. J., and M. E. Winkler.** 1987. An unusual genetic link between vitamin B_6 biosynthesis and tRNA pseudouridine modification in *Escherichia coli* K-12. J. Bacteriol. **169**:1071–1079.
3. **Arps, P. J., and M. E. Winkler.** 1987. Structural analysis of the *Escherichia coli* K-12 *hisT* operon by using a kanamycin resistance cassette. J. Bacteriol. **169**:1061–1070.
4. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1987. Current protocols in molecular biology. John Wiley & Sons, Inc., New York.
5. **Bachmann, B. J.** 1987. Linkage map of *Escherichia coli* K-12, p. 395–411. *In* F. C. Neidhardt (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
6. **Beck, C. F., and R. A. J. Warren.** 1988. Divergent promoters, a common form of gene organization. Microbiol. Rev. **52**:318–326.
7. **Belfaiza, J., C. Parsot, A. Martel, C. Bouthier de la Tour, D. Margarita, G. N. Cohen, and I. Saint-Girons.** 1986. Evolution in biosynthetic pathways: two enzymes catalyzing consecutive

steps in methionine biosynthesis originate from a common ancestor and possess a similar regulatory region. Proc. Natl. Acad. Sci. USA **83**:867–871.
8. **Bender, D. A.** 1985. Amino acid metabolism. John Wiley & Sons, Inc., New York.
9. **Birktoft, J. J., and L. J. Banaszak.** 1983. The presence of a histidine-aspartic acid pair on the active site of 2-hydroxyacid dehydrogenases. J. Biol. Chem. **258**:472–482.
10. **Birktoft, J. J., and L. J. Banaszak.** 1984. Structure-function relationships among nicotinamide-adenine dinucleotide dependent oxidoreductases. Pept. Prot. Rev. **4**:1–46.
11. **Bognar, A., C. Pyne, M. Yu, and G. Basi.** 1989. Transcription of the *folC* gene encoding folylpolyglutamate synthetase-dihydrofolate synthetase in *Escherichia coli*. J. Bacteriol. **171**:1854–1861.
12. **Bognar, A. L., C. Osborne, and B. Shane.** 1987. Primary structure of the *Escherichia coli folC* gene and its folylpolyglutamate synthetase-dihydrofolate synthetase product and regulation of expression by an upstream gene. J. Biol. Chem. **262**: 12337–12343.
13. **Brosius, J.** 1984. Plasmid vectors for the selection of promoters. Gene **27**:151–160.
14. **Cho, K.-O., and C. Yanofsky.** 1988. Sequence changes preceding a Shine-Dalgarno region influence *trpE* mRNA translation and decay. J. Mol. Biol. **204**:51–60.
15. **Connolly, D. M., and M. E. Winkler.** 1989. Genetic and physiological relationships among the *miaA* gene, 2-methylthio-N^6-(Δ^2-isopentenyl)-adenosine tRNA modification, and spontaneous mutagenesis in *Escherichia coli* K-12. J. Bacteriol. **171**: 3233–3246.
16. **Costa, R. H., E. Lai, and J. E., Darnell, Jr.** 1986. Transcriptional control of the mouse prealbumin (transthyretin) gene: both promoter sequences and a distinct enhancer are cell specific. Mol. Cell. Biol. **6**:4697–4708.
17. **Craig, N. L., and H. A. Nash.** 1984. *E. coli* integration host factor binds to specific sites in DNA. Cell **39**:707–716.
18. **Davis, R. W., D. Botstein, and J. R. Roth.** 1980. Advanced bacterial genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Dayhoff, M. O., W. C. Barker, and L. T. Hunt.** 1983. Establishing homologies in protein sequences. Methods Enzymol. **91**: 524–545.
20. **Dempsey, W. B.** 1980. Biosynthesis and control of vitamin B_6 in *Escherichia coli*, p. 93–111. *In* G. P. Tryfiates (ed.), Vitamin B_6 metabolism and role in growth. Food and Nutrition Press, Westport, Conn.
21. **Dempsey, W. B.** 1987. Synthesis of pyridoxal phosphate, p. 539–543. *In* F. C. Neidhardt (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
22. **Feng, D. F., M. S. Johnson, and R. F. Doolittle.** 1985. Aligning amino acid sequences: comparison of commonly used methods. J. Mol. Evol. **21**:112–125.
23. **Goncharoff, P., and B. P. Nichols.** 1984. Nucleotide sequence of *Escherichia coli pabB* indicates a common evolutionary origin of *p*-aminobenzoate synthetase and anthranilate synthetase. J. Bacteriol. **159**:57–62.
24. **Goncharoff, P., and B. P. Nichols.** 1988. Evolution of aminobenzoate synthases: nucleotide sequences of *Salmonella typhimurium* and *Klebsiella aerogenes pabB*. Mol. Biol. Evol. **5**:531–548.
25. **Henikoff, S., G. W. Haughn, J. M. Calvo, and J. C. Wallace.** 1988. A large family of bacterial activator proteins. Proc. Natl. Acad. Sci. USA **85**:6602–6606.
26. **Henikoff, S., and J. C. Wallace.** 1988. Detection of protein similarities using nucleotide sequence databases. Nucleic Acids Res. **16**:6191–6204.
27. **Hill, R. E., and I. D. Spenser.** 1986. Biosynthesis of vitamin B_6, p. 417–476. *In* D. Dolphin, R. Poulson, and O. Avramovic (ed.), Coenzymes and cofactors, vol. 1. Vitamin B_6-pyridoxal phosphate. John Wiley & Sons, Inc., New York.
28. **Horowitz, N. H.** 1965. The evolution of biochemical syntheses-retrospect and prospect, p. 15–23. *In* V. Byron, and H. J. Vogel (ed.), Evolving genes and proteins. Academic Press, Inc., New York.
29. **Jensen, R. A.** 1976. Enzyme recruitment in evolution of new function. Annu. Rev. Microbiol. **30**:409–425.
30. **Kaplan, J. B., W. K. Merkel, and B. P. Nichols.** 1985. Evolution of glutamine amidotransferase genes: nucleotide sequences of the *pabA* genes from *Salmonella typhimurium*, *Klebsiella aerogenes* and *Serratia marcescens*. J. Mol. Biol. **183**:327–340.
31. **Kaplan, J. B., and B. P. Nichols.** 1983. Nucleotide sequence of *Escherichia coli pabA* and its evolutionary relationship to *trp(G)D*. J. Mol. Biol. **168**:451–468.
32. **Lin, E. C. C., A. J. Hacking, and J. Aguilar.** 1976. Experimental models of acquisitive evolution. BioScience **26**:548–555.
33. **Maniatis, T., E. F. Fitsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
34. **Marinus, M. G.** 1987. Methylation of DNA, p. 697–702. *In* F. C. Neidhardt (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
35. **Marvel, C. C., P. J. Arps, B. C. Rubin, H. O. Kammen, E. E. Penhoet, and M. E. Winkler.** 1985. *hisT* is part of a multigene operon in *Escherichia coli* K-12. J. Bacteriol. **161**:60–71.
36. **Nixon, B. T., C. W. Ronson, and F. M. Ausubel.** 1986. Two-component regulatory systems responsive to environmental stimuli share strongly conserved domains with the nitrogen assimilation regulatory genes *ntrB* and *ntrC*. Proc. Natl. Acad. Sci. USA **83**:7850–7854.
37. **Nonet, M. L., C. C. Marvel, and D. R. Tolan.** 1987. The *hisT-purF* region of the *Escherichia coli* K-12 chromosome. J. Biol. Chem. **262**:12209–12217.
38. **Parsot, C.** 1986. Evolution of biosynthetic pathways: a common ancestor for threonine synthase, threonine dehydratase, and D-serine dehydratase. EMBO J. **5**:3013–3019.
39. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85**:2444–2448.
40. **Roa, B. B., D. M. Connolly, and M. E. Winkler.** 1989. Overlap between *pdxA* and *ksgA* in the complex *pdxA-ksgA-apaG-apaH* operon of *Escherichia coli* K-12. J. Bacteriol. **171**:4767–4777.
41. **Schuller, D. J., C. H. Fetter, L. J. Banaszak, and G. A. Grant.** 1989. Enhanced expression of the *Escherichia coli serA* gene in a plasmid vector: purification, crystallization, and preliminary X-ray data of D-3-phosphoglycerate dehydrogenase. J. Biol. Chem. **264**:2645–2648.
42. **Squires, C. H., M. DeFelice, J. Devereux, and J. M. Calvo.** 1983. Molecular structure of *ilvH* and its evolutionary relationship to *ilvG* in *Escherichia coli* K12. Nucleic Acids Res. **11**:5299–5313.
43. **Stauffer, G. V.** 1987. Biosynthesis of serine and glycine, p. 412–418. *In* F. C. Neidhardt (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
44. **Stokes, H. W., and B. G. Hall.** 1984. Topological repression of gene activity by a transposable element. Proc. Natl. Acad. Sci. USA **81**:6115–6119.
45. **Stokes, H. W., and B. G. Hall.** 1985. Sequence of *ebgR* of *Escherichia coli*: evidence that the EBG and LAC operons are descended from a common ancestor. Mol. Biol. Evol. **2**:478–483.
46. **Tobey, K. L., and G. A. Grant.** 1986. The nucleotide sequence of the *serA* gene of *Escherichia coli* and the amino acid sequence of the encoded protein, D-3-phosphoglycerate dehydrogenase. J. Biol. Chem. **261**:12179–12183.
47. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–109.
48. **Zalkin, H., P. Argos, S. V. L. Narayana, A. A. Tiedeman, and J. M. Smith.** 1985. Identification of a *trpG*-related glutamine amide transfer domain in *Escherichia coli* GMP synthetase. J. Biol. Chem. **260**:3350–3354.