JOURNAL OF BACTERIOLOGY, Dec. 1988, p. 5500–5506
0021-9193/88/125500-07$02.00/0

Vol. 170, No. 12

Mutational Analysis of the Catalytic and Feedback Sites of the Tryptophan-Sensitive 3-Deoxy-D-*arabino*-Heptulosonate-7-Phosphate Synthase of *Escherichia coli*

JILL M. RAY,[1] CHARLES YANOFSKY,[2] AND RONALD BAUERLE[1]*

Department of Biology and Molecular Biology Institute, University of Virginia, Charlottesville, Virginia 22901,[1] and Department of Biological Sciences, Stanford University, Stanford, California 94305[2]

Received 16 June 1988/Accepted 29 August 1988

The nucleotide sequence of *aroH*, the structural gene for the tryptophan-sensitive 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase [DAHPS(Trp)], is presented, and the deduced amino acid sequence of AroH is compared with that of the tyrosine-sensitive (AroF) and phenylalanine-sensitive (AroG) DAHPS isoenzymes. The high degree of sequence similarity among the three isoenzymes strongly indicates that they have a common evolutionary origin. In vitro chemical mutagenesis of the cloned *aroH* gene was used to identify residues and regions of the polypeptide essential for catalytic activity and for tryptophan feedback regulation. Missense mutations leading either to loss of catalytic activity or to feedback resistance were found interspersed throughout the polypeptide, suggesting overlapping catalytic and regulatory sites in DAHPS(Trp). We conclude that the specificity of feedback regulation of the isoenzymes was probably acquired by the duplication and divergent evolution of an ancestral gene, rather than by domain recruitment.

In bacteria and plants, 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase (DAHPS) (EC 4.1.2.15) catalyzes the first committed step in the pathway that leads to the biosynthesis of aromatic acids and vitamins. In *Escherichia coli* and other enteric bacteria there are three DAHPS isoenzymes (5). Although catalyzing the same reaction (i.e., the condensation of erythrose-4-phosphate and phosphoenolpyruvate [PEP] to form 3-deoxy-D-*arabino*-heptulosonate 7-phosphate [DAHP]), each isoenzyme is feedback regulated by a different aromatic amino acid. The structural genes *aroF*, *aroG*, and *aroH* encode the tyrosine (Tyr)-, phenylalanine (Phe)-, and tryptophan (Trp)-inhibitable isoenzymes, respectively. The three genes are widely separated on the *E. coli* chromosome, *aroF* mapping at 57 min, *aroG* mapping at 17 min, and *aroH* mapping at 37 min (1). Transcriptional control of *aroF* and *aroG* expression is mediated by the *tyrR* repressor (13), and that of *aroH* is mediated by the *trpR* repressor (6, 23). In wild-type cells grown in minimal medium, the AroG isoenzyme makes up about 80% of the total DAHPS activity, the AroF isoenzyme makes up 20%, and the AroH isoenzyme makes up about 1% (19).

The *aroF* and *aroG* genes have been cloned and have been found to be similar in size and sequence (3, 4, 18). When optimally aligned, 58% of the nucleotides and 53% of the amino acids of the derived polypeptide sequences are identical. The sequence identities are not evenly distributed throughout the two polypeptides, but are found predominantly in several extensive stretches in the interior of the molecules. These observations led Shultz et al. (18) to propose that the highly conserved regions of the polypeptides may be essential for their common catalytic activity, whereas the regions of low or no similarity may serve to enable the proper positioning of the other residues or to bestow regulatory specificity to each isoenzyme. It was also hypothesized that the genes may have evolved through rearrangement of functionally defined domains.

The *aroH* gene has also been cloned, and the sequence of the 5′ regulatory region and of the first 108 and last 354 nucleotides of the coding region has been reported (23). In this report we present the complete nucleotide sequence of *aroH* and compare the amino acid sequences of the three DAHPS isoenzymes. We also describe the results of a mutational analysis undertaken to probe the regions of AroH that are essential for catalytic activity and for feedback regulation in order to test the existence of distinct functional domains in this isoenzyme.

MATERIALS AND METHODS

Bacterial strains and plasmids. *E. coli* AB3248 (*aroF363 aroG365 aroH367 proA2 argE3 ilv-7 his-4 lac gal-2 tsx-358 thi*) was isolated by Wallace and Pittard (20). *E. coli* TB1 [Δ(lac-pro) *strA ara thi* φ80d*lacZ*Δ*M15 hsdR*] was purchased from Bethesda Research Laboratories, Inc., Gaithersburg, Md.

Plasmid pAROH924 (23) consists of a 3.7-kilobase *Bam*HI fragment from the *E. coli* chromosome containing the entire *aroH* gene plus flanking sequences cloned into the *Bam*HI site of vector pACYC177. Plasmid pAHH1 (Fig. 1) was constructed by subcloning the 1.9-kilobase *Bam*HI-*Eco*RI fragment from pAROH924 containing the *aroH* gene and flanking sequences into the *Bam*HI and *Eco*RI sites of plasmid pBR322. Plasmid pAFF1 was constructed by subcloning the 2.2-kilobase *Bam*HI-*Hin*dIII fragment containing the *aroF* gene and flanking sequences from plasmid pKB45 (22) into the *Bam*HI-*Hin*dIII sites of pBR322. Replicative forms of bacteriophages M13mp18 and M13mp19 were purchased from Bethesda Research Laboratories.

Media. Minimal agar for the growth of strain AB3248 consisted of minimal salts medium [K_2HPO_4, 10.5 g/liter; KH_2PO_4, 4.5 g/liter; $(NH_4)_2SO_4$, 1.0 g/liter; $MgSO_4$, 0.1 g/liter] containing Bacto-Agar (15 g/liter; Difco Laboratories, Detroit, Mich.) and supplemented with glucose (2.5 g/liter); arginine, histidine, isoleucine, and valine (each at 0.1 g/liter); proline (0.2 g/liter); and thiamine, *p*-aminobenzoic acid, and *p*-hydroxybenzoic acid (each at 0.02 g/liter). Where indicated, Trp, Tyr, and Phe were added at 0.1 g/liter. Cultures

* Corresponding author.

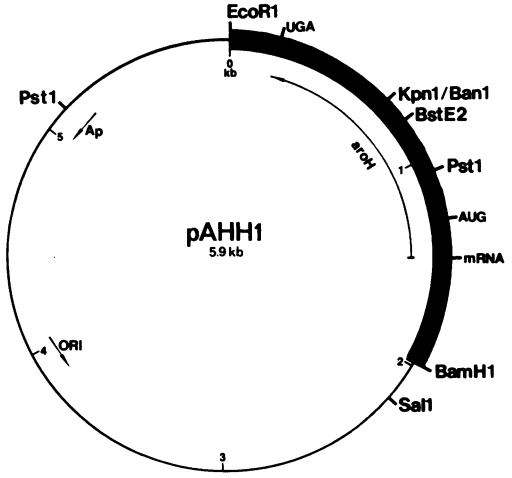

FIG. 1. Physical map of plasmid pAHH1. The pBR322 ampicillin resistance gene (Ap) and the origin of replication (ORI) are indicated. Arrows denote the direction of transcription. kb, Kilobase pairs.

for enzymatic analysis were grown in minimal medium lacking agar, except that glucose was present at 5 g/liter and the medium was further supplemented with casein hydrolysate (2 g/liter; U.S. Biochemical Corp., Cleveland, Ohio) and yeast extract (0.2 g/liter). LB broth contained tryptone (10 g/liter), yeast extract (5 g/liter), and NaCl (10 g/liter). Ampicillin was used at 50 μg/ml in LB broth and at 25 μg/ml in minimal agar.

Enzymes and reagents. DNA restriction and modification enzymes were obtained commercially and used as recommended by the supplier. PEP (monocyclohexylammonium salt), 5-methyl-DL-tryptophan (5-MT), 3-β-indoleacrylic acid, D-erythrose-4-phosphate sodium salt, and (1,3-bis[tris(hydroxymethyl)methylamino]propane) (BTP) were purchased from Sigma Chemical Co., St. Louis, Mo.

Hydroxylamine and nitrous acid mutagenesis. pAHH1 DNA was isolated by alkaline sodium dodecyl sulfate lysis of cells from amplified cultures of TB1(pAHH1) and was purified by gel sieve chromatography with Bio-Gel AcA34 (Bio-Rad Laboratories, Richmond, Calif.), as described by Just et al. (7).

For hydroxylamine mutagenesis, 1 μg of purified plasmid DNA was incubated in a total volume of 1 ml with 800 mM hydroxylamine hydrochloride in 200 mM potassium phosphate buffer (pH 6.0)–2 mM EDTA at 37°C for 24 h. The treated DNA was dialyzed for 15 h at 4°C with three changes of TE buffer (50 mM Tris hydrochloride [pH 7.5], 10 mM EDTA). The DNA was then precipitated by the addition of 0.1 volume of 3 M sodium acetate (pH 4.8) and 2 volumes of ethanol. The DNA was collected by centrifugation, washed with 75% ethanol, dried in vacuo, and redissolved in TE buffer.

For nitrous acid mutagenesis, 1 μg of purified plasmid DNA was incubated in 100 mM sodium acetate buffer (pH 4.6) containing 50 mM sodium nitrite (freshly prepared) at room temperature for 30 min in a final volume of 600 μl. The reaction was stopped by the addition of 100 μl of 2.5 M sodium acetate buffer (pH 7.0) and 200 μl of water. The DNA was precipitated by the addition of 2 volumes of ethanol, dissolved in TE buffer, precipitated twice more, dried in vacuo, and redissolved in TE buffer.

Mutagenized DNA was transformed into strain AB3248 with selection on LB agar containing ampicillin. Transformants were screened for the Aro$^-$ phenotype (i.e., auxotrophy for Phe, Trp, and Tyr) on minimal agar containing ampicillin and for the MTR phenotype (i.e., resistance to 5-MT) on minimal agar containing ampicillin and 5-MT (50 μg/ml). Normally, strains with only the *aroH* gene intact (i.e., *aroF aroG* mutants) are prototrophic, but are subject to growth inhibition by Trp, since Trp-mediated repression of *aroH* expression and feedback inhibition of DAHPS(Trp) starves the cell for Phe and Tyr (14). However, this Trp sensitivity is alleviated in *aroF aroG* strains with multicopy *aroH*, as used here, thereby necessitating the use of the analog, 5-MT, for the selection of *aroH* regulatory mutants. Putative auxotrophic and MTR mutant clones were purified and retested. To eliminate mutations in the vector DNA, we excised the *Bam*HI-*Eco*RI insert of each mutant plasmid and recloned it into untreated pBR322.

Restriction fragment exchange mapping. The plasmid-borne mutations were localized within *aroH* by restriction fragment exchange mapping. Specific restriction fragments were removed from wild-type pAHH1 and replaced with the corresponding fragment from the mutant plasmids. The DNA fragments of the wild-type and mutant plasmid digests were separated in low-melting-temperature agarose and ligated in the presence of the agarose by using T4 DNA ligase (8). The ligation mixture was used to transform AB3248 with selection on LB agar containing ampicillin. Transformants were tested for either the MTR or Aro$^-$ phenotype, as appropriate.

DNA sequencing. The wild-type *aroH* gene was first sequenced by the method of Maxam and Gilbert (10), using overlapping restriction fragments of plasmid pAROH924. The sequence was verified by the dideoxy-chain termination method (16), using restriction fragments of pAHH1 cloned into the replicative form of the M13mp18 and M13mp19 bacteriophage sequencing vectors. Aro$^-$ and MTR mutations were identified by dideoxy sequencing of the appropriate restriction fragment, as indicated by the results of the fragment exchange mapping.

Preparation of crude cell extracts. Cells were grown at 37°C in 100 ml of supplemented minimal medium (see above). The expression of *aroH* was derepressed by the addition of 100 μg of 2-β-indoleacrylic acid per ml (12) when the cultures reached an A_{550} of 0.5. Incubation was continued for 6 h, after which the cells were harvested and washed in 1 volume of 0.9% saline. The drained cell pellet was kept at −70°C overnight, thawed on ice, and then suspended in 5 ml of cold SPP (0.05 M potassium phosphate [pH 7.0] containing 100 μg of PEP per ml). All subsequent steps were carried out at 4°C. The cells were disrupted by sonication for 35 s at 70 W, a Branson sonifier (Branson Sonic Power Co., Danbury, Conn.). Cell debris was removed by centrifugation at 48,000 × *g* for 30 min. The supernatant was passed through a column of Sephadex G-25 (bed volume, 11 ml) equilibrated with SPP to remove residual 3-β-indoleacrylic acid and other small molecules.

Assay of DAHPS. DAHPS activity was assayed by a modification of the continuous spectrophotometric method described by Schoner and Herrman (17). This method can be reliably used with crude extracts of strains carrying the *aroH* gene on a multicopy plasmid. The complete reaction mixture contained 10 mM BTP (pH 7.0), 100 μM PEP, and 300 μM erythrose-4-phosphate in a total volume of 1 ml. The mixture was equilibrated to room temperature, other additives, such as inhibitors, were added, and the reaction was started by the addition of enzyme extract. Enzyme activity is measured by monitoring the rate of disappearance of PEP at 232 nm by

```
                                        30                                  60
ATG AAC AGA ACT GAC GAA CTC CGT ACT GCG CGT ATT GAG AGC CTG GTA ACG CCC GCC GAA CTC GCG
MET Asn Arg Thr Asp Glu Leu Arg Thr Ala Arg Ile Glu Ser Leu Val Thr Pro Ala Glu Leu Ala

                                    90                                  120
CTA CGG TAT CCC GTA ACG CCT GGC GTC GCC ACC CAT GTC ACC GAC TCC CGC CGC AGA ATT GAA AAA
Leu Arg Tyr Pro Val Thr Pro Gly Val Ala Thr His Val Thr Asp Ser Arg Arg Arg Ile Glu Lys

                        150                                 180
ATA CTG AAT GGT GAA GAT AAG CGA CTG TTG GTC ATT ATT GGC CCC TGC TCG ATC CAC GAT CTC ACC
Ile Leu Asn Gly Glu Asp Lys Arg Leu Leu Val Ile Ile Gly Pro Cys Ser Ile His Asp Leu Thr

            210                                 240
GCT GCA ATG GAG TAC GCC ACC CGT CTG CAG TCG CTG CGC AAC CAG TAC CAG TCA CGG CTG GAA ATC
Ala Ala Met Glu Tyr Ala Thr Arg Leu Gln Ser Leu Arg Asn Gln Tyr Gln Ser Arg Leu Glu Ile

    270                                 300                                 330
GTA ATG CGC ACC TAT TTT GAA AAA CCA CGA ACT GTT GTC GGC TGG AAA GGA CTA ATC TCC GAT CCA
Val Met Arg Thr Tyr Phe Glu Lys Pro Arg Thr Val Val Gly Trp Lys Gly Leu Ile Ser Asp Pro

                                    360                                 390
GAT TTA AAC GGC AGC TAT CGG GTA AAT CAC GGT CTG GCG CTG GCG CGC AAA TTA CTT TTA CAG GTA
Asp Leu Asn Gly Ser Tyr Arg Val Asn His Gly Leu Glu Leu Ala Arg Lys Leu Leu Leu Gln Val

                                420                                 450
AAT GAG CTG GGC GTC CCA ACC GCG ACC GAG TTC CTC GAT ATG GTG ACC GGT CAG TTT ATT GCT GAT
Asn Glu Leu Gly Val Pro Thr Ala Thr Glu Phe Leu Asp Met Val Thr Gly Gln Phe Ile Ala Asp

                        480                                 510
TTA ATC AGT TGG GGC GCG ATT GGC GCA CGT ACT ACC GAA AGT CAG ATC CAC CGC GAA ATG GCT TCG
Leu Ile Ser Trp Gly Ala Ile Gly Ala Arg Thr Thr Glu Ser Gln Ile His Arg Glu Met Ala Ser

            540                                 570
GCA CTC TCC TGT CCG GTA GGT TTT AAA AAT GGT ACC GAT GGC AAT ACG CGG ATT GCT GTG GAT GCT
Ala Leu Ser Cys Pro Val Gly Phe Lys Asn Gly Thr Asp Gly Asn Thr Arg Ile Ala Val Asp Ala

    600                                 630                                 660
ATC CGC GCA GCC GCG CAG CAT ATG TTC CTC TCG CCA GAC AAA AAT GGT CAG ATG ACC ATC TAT CAG
Ile Arg Ala Ala Ala Gln His Met Phe Leu Ser Pro Asp Lys Asn Gly Gln Met Thr Ile Tyr Gln

                                    690                                 720
ACC AGC GGC AAC CCG TAT GGC CAC ATT ATT ATG CGT GGC GGC AAA AAA CCG AAT TAT CAT GCC GAT
Thr Ser Gly Asn Pro Tyr Gly His Ile Ile Met Arg Gly Gly Lys Lys Pro Asn Tyr His Ala Asp

                                750                                 780
GAT ATC GCC GCA GCC TGC GAT ACG CTG CAC GAG TTT GAT TTA CCT GAA CAT CTG GTG GTG GAT TTC
Asp Ile Ala Ala Ala Cys Asp Thr Leu His Glu Phe Asp Leu Pro Glu His Leu Val Val Asp Phe

                        810                                 840
AGC CAC GGT AAC TGC CAG AAG CAG CAC CGT CGC CAG TTA GAA GTT TGT GAG GAT ATT TGT CAG CAA
Ser His Gly Asn Cys Gln Lys Gln His Arg Arg Gln Leu Glu Val Cys Glu Asp Ile Cys Gln Gln

            870                                 900
ATC CGC AAT GGC TCT ACG GCG ATT GCT GGA ATT ATG GCG GAA AGT TTC CTG CGC GAA GGA ACG CAA
Ile Arg Asn Gly Ser Thr Ala Ile Ala Gly Ile Met Ala Glu Ser Phe Leu Arg Glu Gly Thr Gln

    930                                 960                                 990
AAA ATC GTC GGC AGT CAG CCG CTC ACT TAC GGT CAA TCC ATT ACC GAC CCG TGT CTG GGC TGG GAG
Lys Ile Val Gly Ser Gln Pro Leu Thr Tyr Gly Gln Ser Ile Thr Asp Pro Cys Leu Gly Trp Glu

                                1020
GAT AGG CAA CGC CTG GTC GAA AAA CTC GCC TCT GCG GTA GAT ACC CGC TTC TGA
Asp Arg Gln Arg Leu Val Glu Lys Leu Ala Ser Ala Val Asp Thr Arg Phe TER
```

FIG. 2. Nucleotide coding sequence and deduced amino acid sequence of the *E. coli aroH* DAHPS gene and protein. The sequence has been submitted to the EMBL, GenBank, and DDBJ nucleotide sequence data bases and has been assigned the accession number JO4ZZ1.

using a recording spectrophotometer. Any activity detected in a parallel reaction lacking erythrose-4-phosphate was subtracted. This nonspecific activity is negligible in derepressed extracts of cells with multicopy *aroH* plasmids. One unit of activity is equivalent to the removal of 1 nmol of PEP per min at 25°C. The protein content of the extracts was determined by using the Bio-Rad protein reagent (Bio-Rad Laboratories) with bovine serum albumin as the standard.

RESULTS

Sequence analysis of *aroH*. The nucleotide sequence of the *aroH* gene is presented in Fig. 2, together with the inferred amino acid sequence of its protein product. The gene contains 1,041 nucleotides, encoding a polypeptide of 347 residues.

A comparison of the primary sequence of AroH with that of the other two isoenzymes, AroG and AroF, is shown in Fig. 3. The three polypeptides are similar in size, AroG having 350 residues and AroF having 356. The sequence alignment was readily achieved by visual inspection because of the high degree of sequence similarity among the three isoenzymes. The few gaps required in the AroH and AroG sequences were positioned to maximize identities in the overall alignment.

Pairwise comparisons of the three sequences reveal that AroH and AroG have 196 (57%) amino acid residues in common, AroG and AroF have 185 (53%), and AroH and AroF have 165 (48%). When all three polypeptides are compared, 142 (41%) residues are identical. In addition, there are 46 conservative differences among the three molecules. The sequence identities are found predominantly in clusters rather than distributed randomly throughout the sequence. The region of greatest similarity is the stretch between residues 52 and 190, which includes several blocks

FIG. 3. Comparison of the amino acid sequences of the AroH, AroG, and AroF polypeptides. Numbering of residues is that of the AroH sequence. Each dash indicates a gap of one residue. AroG and AroF residues that are identical to those of AroH are indicated with a dot. AroH and AroG residues that are considered conservative changes relative to those of AroH are represented in capital letters. Conservative changes are defined as follows: G = A, D = E, N = Q, S = T, H = K = R, I = L = M = V, F = Y = W. Mutational changes in AroH are indicated with arrows above the AroH sequence. Auxotrophic mutations are designated by lower-case letters, and MTR mutations are shown in capital letters. oc and op represent ochre and opal nonsense mutations, respectively. The sequence of AroF is taken from Schultz et al. (18); that of AroG is from Davies and Davidson (3).

of nearly identical sequence (i.e., residues 52 to 64, 86 to 110, 121 to 130, and 150 to 190). There is significantly less similarity in the carboxyl termini of the isoenzymes (46 identities in the last 157 residues) and little or none in the amino termini (6 identities in the first 41 residues).

In vitro construction of hybrid genes. The high degree of sequence similarity among the AroH, AroG, and AroF polypeptides strongly indicates that they have a common evolutionary origin. The clustered distribution of the sequence identities in the three homologs adds support to the speculation, made earlier by Shultz et al. (18) after comparing the AroF and AroG sequences, that the most highly conserved regions of the polypeptides may constitute homologous catalytic domains for substrate binding, whereas the less-conserved areas (for example, the carboxyl-terminal segment) may constitute nonhomologous regulatory domains for the binding of each specific feedback inhibitor.

This idea was tested by the in vitro construction of in-frame hybrid forms of the *aroH* and *aroF* genes. This was made possible by the presence in the two genes of a conserved *Ban*I restriction site located within the codon for Gly-187, which lies near the end of the highly conserved amino-terminal segment of the polypeptides (Fig. 1 and 3). It was reasoned that if the hybrid enzymes were functional, it might be possible to correlate the specificity of feedback inhibition in the DAHPS isoenzymes with either the amino-terminal or carboxyl-terminal half of the polypeptide.

The left and right arms of the *aroH* and *aroF* inserts of recombinant plasmids pAHH1 and pAFF1, separated by the *Ban*I cleavage, were isolated and reciprocally exchanged by ligation in vitro. The hybrid inserts were then ligated into plasmid pBR322. The desired AroH-AroF and AroF-AroH hybrid plasmids were recovered by transformation into strain AB3248, an *aroH aroF aroG* triple mutant, with selection for Ap^r, and the structure of each hybrid plasmid was verified by restriction analysis. Subsequent testing showed that the transformants containing both the AroH-AroF and AroF-AroH hybrid plasmids remained auxotrophic. The control reconstructions of the parental *aroH* and *aroF* genes yielded prototrophic transformants, as expected. Thus, we conclude that the two hybrid DAHP synthases either are catalytically inactive or are hyperlabile

TABLE 1. Sequence changes in *aroH* mutants

Mutant	Fragment location	Codon change[a]	Amino acid change
MTR prototrophs			
MTR39	*Bam*HI-*Pst*I	CCC→**CTC**	Pro-18→Leu
MTR20	*Pst*I-*Kpn*I	GTG→**ATG**	Val-147→Met
MTR16	*Pst*I-*Kpn*I	GGT→**GAT**	Gly-149→Asp
MTR22	*Pst*I-*Kpn*I	GGT→**AGT**	Gly-149→Ser
MTR36	*Pst*I-*Kpn*I	GGT→**TGT**	Gly-149→Cys
MTR23	*Pst*I-*Kpn*I	GCA→**ACA**	Ala-177→Thr
MTR37	*Kpn*I-*Eco*RI	CAG→**CGG**	Gln-272→Arg
MTR17	*Kpn*I-*Eco*RI	GCG→**ACG**	Ala-293→Thr
Missense auxotrophs			
aroH15	*Bam*HI-*Pst*I	CGC→**TGC**	Arg-91→Cys
aroH3	*Pst*I-*Kpn*I	GGA→**AAA**	Gly-105→Lys
aroH13	*Pst*I-*Kpn*I	GAG→**AAG**	Glu-142→Lys
aroH11	*Pst*I-*Kpn*I	CCG→**ACG**	Gly-162→Thr
aroH24	*Pst*I-*Kpn*I	GCA→**GTA**	Ala-163→Val
aroH12	*Pst*I-*Kpn*I	CGT→**TGT**	Arg-164→Cys
aroH7	*Kpn*I-*Eco*RI	GCT→**ACT**	Ala-198→Thr
aroH19	*Kpn*I-*Eco*RI	GCT→**GTT**	Ala-198→Val
aroH33	*Kpn*I-*Eco*RI	GGC→**GAC**	Gly-223→Asp
aroH8	*Kpn*I-*Eco*RI	CGT→**TGT**	Arg-232→Cys
aroH5	*Kpn*I-*Eco*RI	GGA→**AGA**	Gly-296→Arg
aroH4	*Kpn*I-*Eco*RI	GAA→**AAA**	Glu-300→Lys
aroH10	*Kpn*I-*Eco*RI	TGT→**TAT**	Cys-326→Tyr
Nonsense auxotrophs			
aroH34	*Bam*HI-*Pst*I	CGA→**TGA**	Arg-52→opal
aroH6	*Kpn*I-*Eco*RI	CAA→**TAA**	Gln-308→ochre

[a] Changes are shown in bold type.

in vivo, thereby precluding the in vitro analysis of feedback specificity. Because of this, further analysis of the hybrid enzymes was abandoned.

Mutational analysis of the AroH protein. In vitro mutagenesis of the *aroH* gene was then used to probe the regions of the polypeptide important for catalytic and regulatory function. pAHH1 plasmid DNA was mutagenized with either nitrous acid or hydroxylamine, and auxotrophic (Aro^-) and 5MT-resistant (MTR) mutants were isolated as described in Materials and Methods. Approximately 1 in every 100 transformant colonies screened was found to be a mutant. Ten MTR and nineteen Aro^- isolates were recovered for subsequent analysis.

The mutations were localized to one of three segments of the *aroH* insert of pAHH1 by restriction fragment exchange mapping as follows. Wild-type pAHH1 DNA (Fig. 1) was doubly digested with either *Bam*HI and *Pst*I, *Pst*I and *Kpn*I, or *Kpn*I and *Eco*RI, thereby excising specific segments of the *aroH* insert. The large vector fragments were isolated and ligated with the appropriate restriction fragments prepared from each mutant plasmid, reconstructing the *aroH* gene; the ligated DNA was transformed into strain AB3248 with selection for Ap^r. Transformants were then tested for the Aro^- or MTR phenotype, as appropriate. By this method, mutations were mapped to the 890-base-pair *Bam*HI-*Pst*I fragment, the 355-base-pair *Pst*I-*Kpn*I fragment, or the 702-base-pair *Kpn*I-*Eco*RI fragment of the insert DNA of pAHH1. Both Aro^- and MTR mutations were found in all three segments of the *aroH* insert (Table 1).

To differentiate MTR mutants that are the desired AroH feedback-resistant types from those that have alterations of the *aroH* operator, we analyzed the repressibility of *aroH* expression in each MTR strain. Cultures were grown in supplemented minimal medium and supplemented minimal medium containing Trp, crude extracts were prepared, and DAHPS specific activity was assayed. Of the 10 mutants, 2 possessed constitutive levels of DAHPS activity; in both, the activity displayed wild-type sensitivity to feedback inhibition by Trp. The mutations of both strains mapped to the *Bam*HI-*Pst*I fragment, which contains the upstream regulatory sequences (Fig. 1). It was therefore assumed that these are *aroH* operator mutants; they were not characterized further.

Appropriate restriction fragments containing the Aro^- and MTR mutations were excised from the plasmids and cloned into bacteriophage vector M13mp18 or M13mp19 to identify the mutational changes by DNA sequencing. The nature and location of each change are presented in Table 1 and Fig. 3. All but one of the mutations resulted from a change of a single nucleotide. The exception (*aroH3*) has a change of two adjacent nucleotides; however, this results in the change of only one amino acid. Of the 15 Aro^- mutations, 13 are missense and 2 are nonsense types. Two mutants, E142K and R232C, were recovered more than once.

It is noteworthy that auxotrophic and feedback-resistant mutations are found interspersed rather than segregated and are scattered throughout the length of the AroH polypeptide (Fig. 3). It is also striking that all of the Aro^- missense changes are in residues that are invariant among the three isoenzymes. In contrast, the MTR mutations are found predominantly in residues that are nonconserved.

Feedback inhibition of the DAHPS of MTR mutants. The DAHPS activity of crude extracts of 2-β-indoleacrylic acid-induced cultures of AB3248 carrying pAHH1 and the various MTR plasmids was assayed for sensitivity to feedback inhibition by Trp (Fig. 4). The wild-type enzyme was inhibited to a maximum of about 50%, as has been previously reported (2, 14). The DAHPS activity of five of the MTR mutants (those with changes P18L, V147M, G149D, G149C, and A177T) was completely resistant to Trp inhibition,

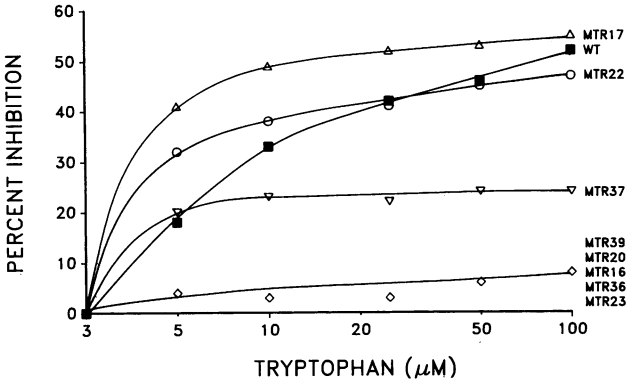

FIG. 4. Feedback inhibition of the DAHPS of MTR mutants by L-tryptophan.

whereas that of another (Q272R) was partially resistant. The other two mutants (G149S and A293T) showed anomalous behavior. The DAHPS activity of the former was about as Trp sensitive as the wild-type enzyme, whereas that of the latter appeared to have enhanced sensitivity to the inhibitor. The possibility that the MTR phenotype of these two strains is the result of a specific insensitivity of their DAHPS to inhibition by 5-MT has been excluded (data not shown). Thus, this anomaly appears to be a manifestation of differences in the feedback properties of the two enzymes in vivo and in vitro.

The wild-type and mutant enzymes were also assayed for feedback sensitivity to Tyr and Phe, the specific inhibitors of the AroF and AroG isoenzymes, respectively, and to chorismic acid, which regulates DAHPS in other bacterial species (21). Neither the wild-type nor any of the mutant enzymes were found to exhibit significant sensitivity (i.e., greater than 20% inhibition) to any of these metabolites, even at concentrations as high as 300 μM.

DISCUSSION

The high degree of sequence similarity among the AroH, AroG, and AroF polypeptides is convincing evidence that the three isoenzymes have a common evolutionary origin. It has been previously suggested that the three DAHPS homologs might have arisen during evolution by the combination of pre-existing parts of different origins (18). For example, it can be speculated that the genetic fusion of a common ancestral catalytic domain with different amino acid (i.e., Trp, Phe or Tyr)-binding domains has led to the attainment of specific feedback regulation in each isoenzyme (21). This possibility is consistent with the great variations in sequence homology in different regions of the polypeptides (Fig. 3). However, the results of the mutational probing of the AroH isoenzyme reported here do not support this idea.

Rather than being localized to discrete regions of the polypeptide, missense mutations that eliminate catalytic activity are found throughout the *aroH* gene and are interspersed with those that alter or remove feedback regulation in the enzyme. This indicates that the functionally important regions of DAHPS(Trp) are not physically separate and probably include overlapping sequences. From these results it now appears certain that the three isoenzymes arose by the duplication and divergent evolution of a common ancestral gene, as has been recently proposed by Pittard (13). Our finding that the AroH-AroF and AroF-AroH hybrid proteins lack catalytic activity is consistent with this, implying the coevolution of the amino-terminal and carboxylterminal segments of each DAHPS isoenzyme.

All of the catalytic *aroH* mutations analyzed thus far alter amino acid residues that are totally conserved in the three isoenzymes (Fig. 3). Furthermore, many of these are located within one of the major blocks of sequence conservation in the amino-terminal half of the polypeptide. It is not yet clear which mutations are identifying residues essential for substrate binding and which might be causing structural instability or alterations in conformation. However, the occurrence of adjacent, inactivating mutations in residues Gly-162, Ala-163, and Arg-164 (Fig. 3) is particularly noteworthy. These residues, being located within one of the highly conserved sequence blocks (Fig. 3), are attractive candidates as components of the active site of the enzyme.

In contrast to the inactivating missense mutations in AroH, seven of the eight mutations that alter feedback regulation of the enzyme change residues that are not rigidly conserved in the three isoenzymes. Three of the residues important for feedback regulation, namely Val-147, Gly-149, and Thr-177, flank the putative active-site residues, Gly-162, Ala-163, and Arg-164, mentioned above. Gly-149 appears to be of particular significance, since MTR mutants with three different changes have been recovered. One of these (G149C) is a G-to-T transversion, which presumably arose spontaneously. Feedback inhibition of DAHPS(Phe) by Phe and DAHPS(Tyr) by Tyr has been shown to be competitive with the substrate PEP (11, 16). Assuming that Trp inhibition of DAHPS(Trp) is also competitive for PEP, it may be that residues 140 to 180 of AroH are structural components of overlapping binding sites for Trp and PEP.

The above speculations are based on the assumption that inherent in each AroH polypeptide is an active site for substrate binding and a feedback site for Trp binding. Since the active form of DAHPS(Trp) is a homodimer (J. M. Ray and R. Bauerle, unpublished results), the possibility exists that the holoenzyme possesses a composite active site(s) and/or feedback site(s), assembled from residues of both of the AroH subunits. However, there is at present no evidence supporting this type of structure in DAHPS(Trp) (2) or in the more extensively studied dimeric DAHPS(Tyr) and tetrameric DAHPS(Phe) isoenzymes (11, 17).

Comparisons of the AroH sequence with those of other proteins have not been informative in the identification of putative active-site or feedback site sequences. Other than the strong homologies that exist among the three DAHPS isoenzymes, AroH does not share significant sequence homology with any other proteins of known structure, as indicated by global searches of the GenBank (edition 56) and National Biomedical Research Foundation (edition 17) sequence data bases with the FASTP program (stringency Ktup2) of Lipman and Pearson (9). Furthermore, pairwise comparisons of AroH with other bacterial proteins that are known to bind one or more of the aromatic amino acids, including TrpE (anthranilate synthase component I), TrpR (tryptophan repressor), TrpS (tyrosyl-tRNA synthetase), TyrR (tyrosine repressor), TyrA (chorismate mutase-prephenate dehydrogenase), PheA (chorismate mutase-prephenate dehydratase), and TnaA (tryptophanase), also revealed no significant sequence homologies. A number of enzymes that bind phosphorylated compounds have been found to have the sequence Gly-X-Gly-X-X-Gly at the amino-terminal end of an alpha helix where the compound binds (15). However, this sequence is not found in any of the DAHPS isoenzymes.

It is interesting that the DAHPS(Trp) of two MTR mutants

(G149S and A293T) display either full or somewhat enhanced sensitivity to Trp inhibition in vitro (Fig. 4), rather than the expected loss of sensitivity. This observation is particularly surprising for the G149S mutant, since two other changes in this residue (G149D and G149C) lead to full feedback resistance under the same conditions of assay. Fragment exchange mapping and DNA sequencing have unequivocally shown that the MTR phenotype of the two strains derives from the single nucleotide change indicated in Table 1. Also, the anomalous behavior is not due to a specific insensitivity to 5-MT, the selective agent in the mutant screen. It is possible that it derives from changes in the conformational state of the mutant enzymes in vitro or perhaps is a reflection of more subtle changes in the kinetic properties of these enzymes relative to those of the other MTR strains. Purification and kinetic characterization of the wild-type and mutant enzymes are currently under way.

ACKNOWLEDGMENTS

This work was supported by Public Health Service research grant GM35889 and predoctoral training grant GM07082 from the National Institutes of Health and by research grant DMB 8703685 from the National Science Foundation. C.Y. is a career investigator of the American Heart Association. R.B. was a visiting scientist of the American Heart Association while on sabbatical leave at Stanford University.

LITERATURE CITED

1. **Bachmann, B. J.** 1987. Linkage map of *Escherichia coli* K-12, edition 7, p. 807–876. *In* J. L. Ingraham, K. Brooks, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
2. **Camakaris, H., and J. Pittard.** 1974. Purification and properties of 3-deoxy-D-arabinoheptulosonic acid-7-phosphate synthetase (trp) from *Escherichia coli*. J. Bacteriol. **120**:590–597.
3. **Davies, W. D., and B. E. Davidson.** 1982. The nucleotide sequence of *aroG*, the gene for 3-deoxy-D-*arabino*heptulosonate-7-phosphate synthetase (phe) in *Escherichia coli* K-12. Nucleic Acids Res. **10**:4045–4058.
4. **Davies, W. D., J. Pittard, and B. E. Davidson.** 1985. Cloning of *aroG*, the gene coding for phospho-2-keto-3-deoxy-heptonate aldolase (phe) in *Escherichia coli*, and subcloning of the *aroG* promoter and operator in a promoter-detecting plasmid. Gene **33**:323–331.
5. **Doy, C. H., and K. D. Brown.** 1965. Control of aromatic biosynthesis: the multiplicity of 7-phospho-2-oxo-3-deoxy-D-*arabino*-heptonate D-erythrose-4-phosphate lyase (pyruvate phosphorylating) in *Escherichia coli W*. Biochim. Biophys. Acta **104**:377–389.
6. **Grove, C. L., and R. P. Gunsalus.** 1987. Regulation of the *aroH* operon of *Escherichia coli* by the tryptophan repressor. J. Bacteriol. **169**:2158–2164.
7. **Just, L., R. Frankis, W. S. Lowery, R. A. Meyer, and G. V. Paddock.** 1983. Rapid high-yield purification of plasmids by alkaline extraction, PEG precipitation and agarose gel chromatography. Biotechniques **1**:136–140.
8. **King, P. V., and R. W. Blakesley.** 1986. Optimizing DNA ligations for transformation. BRL Focus **8**:1–3.
9. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1438.
10. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
11. **McCandliss, R. J., M. D. Poling, and K. M. Herrmann.** 1978. 3-Deoxy-D-*arabino*-heptulosonate 7-phosphate synthase: purification and molecular characterization of the phenylalanine-sensitive isoenzyme from *Escherichia coli*. J. Biol. Chem. **253**:4259–4265.
12. **Morse, D. E., R. D. Mosteller, R. F. Baker, and C. Yanofsky.** 1969. Direction of *in vivo* degradation of tryptophan messenger RNA: a correction. Nature (London) **223**:40–43.
13. **Pittard, A. J.** 1987. Biosynthesis of the aromatic amino acids, p. 368–394. *In* J. L. Ingraham, L. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
14. **Pittard, J., J. Camaskaris, and B. J. Wallace.** 1969. Inhibition of 3-deoxy-D-*arabino* heptulosonic acid-7-phosphate synthetase (trp) in *Escherichia coli*. J. Bacteriol. **97**:1242–1247.
15. **Rogers, N. K.** 1986. The modelling of electrostatic interactions in the function of globular proteins. Prog. Biophys. Mol. Biol. **48**:37–66.
16. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5436–5467.
17. **Schoner, R., and K. M. Herrman.** 1976. 3-Deoxy-D-*arabino*-heptulosonate 7-phosphate synthase. Purification, properties, and kinetics of the tyrosine-sensitive isoenzyme from *Escherichia coli*. J. Biol. Chem. **251**:5440–5447.
18. **Shultz, J., M. A. Hermodson, C. Garner, and K. M. Herrman.** 1984. The nucleotide sequence of the *aroF* gene of *Escherichia coli* and the amino acid sequence of the encoded protein, the tyrosine-sensitive 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase. J. Biol. Chem. **259**:9655–9661.
19. **Tribe, D. E., H. Camakaris, and J. Pittard.** 1976. Constitutive and repressible enzymes of the common pathway of aromatic biosynthesis in *Escherichia coli* K-12: regulation of enzyme synthesis at different growth rates. J. Bacteriol. **127**:1085–1097.
20. **Wallace, B. J., and J. Pittard.** 1967. Genetic and biochemical analysis of the isoenzymes concerned in the first reaction of aromatic biosynthesis in *Escherichia coli*. J. Bacteriol. **93**:234–244.
21. **Whitaker, R. J., A. Berry, G. S. Byng, M. J. Fiske, and R. A. Jensen.** 1985. Clues from *Xanthomonas campestris* about the evolution of aromatic biosynthesis and its regulation. J. Mol. Evol. **21**:139–149.
22. **Zurawski, G., K. D. Brown, D. Killingly, and C. Yanofsky.** 1978. The nucleotide sequence of the leader region of the phenylalanine operon of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **75**:4271–4275.
23. **Zurawski, G., R. P. Gunsalus, K. D. Brown, and C. Yanofsky.** 1981. Structure and regulation of *aroH*, the structural gene for tryptophan repressible 3-deoxy-D-*arabino*-heptulosonic acid-7-phosphate synthetase of *Escherichia coli*. J. Mol. Biol. **145**:47–73.