(1984) *Cell 38*, 317–331.
Saiki, R., Gelffand, D. H., Stoffel, S., Scharf, S., Higuchi, R., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1988) *Science 239*, 487–491.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Schapira, M., Despland, E., Scott, C. F., Boxer, L. A., & Colman, R. W. (1982) *J. Clin. Invest. 69*, 1199–1202.
Schiffman, S., Rapaport, S. I., & Patch, M. J. (1963) *Blood 22*, 733.
Sealy, J. E., Atlas, S. A., Laragh, J. H., Silvergerg, M., & Kaplan, A. P. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 5914–5918.
Seidah, N. G., Cromlish, J. A., Hamelin, J., Thibault, G., & Chrêtien, M. (1986) *Biosci. Rep. 6*, 835–842.
Seidah, N. G., Paquin, J., Hamelin, J., Benjannet, S., & Chrêtien, M. (1988) *Biochimie 70*, 33–46.
Seidah, N. G., Ladenheim, R., Mbikay, M., Hamelin, J., Lutfalla, G., Rougeon, F., Lazure, C., & Chrêtien, M. (1989) *DNA 8*, 563–574.
Seidah, N. G., Sawyer, N., Hamelin, J., Mion, P., Beaubien, G., Brachpapa, L., Rochemont, J., Mbikay, M., & Chrêtien, M. (1990) *DNA Cell Biol. 9*, 737–748.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–507.
Tait, J. F., & Fujikawa, K. (1987) *J. Biol. Chem. 262*, 11651–11656.
Takagaki, Y., Kitamura, N., & Nakanishi, S. (1985) *J. Biol. Chem. 260*, 8601–8609.
Veloso, D., Shilling, J., Shine, J., Fitch, W. M., & Colman, R. W. (1986) *Thromb. Res. 43*, 153–160.
Wachtfogel, Y. T., Kucich, U., James, H. L., Scott, C. F., Schapira, M., Zimmerman, M., Cohen, A. B., & Colman, R. W. (1983) *J. Clin. Invest. 72*, 1672–1677.
Wells, R. D., & Miglietta, J. J., Klysik, J., Larson, J. E., Stirdivant, S. M., & Zacharias, W. (1982) *J. Biol. Chem. 257*, 10166–10171.
Wilson, A. C., Carlson, S. S., & White, T. J. (1977) *Annu. Rev. Biochem. 46*, 573–639.
Wuepper, K. D. (1973) *J. Exp. Med. 138*, 1345–1355.

Nucleotide Sequence Analysis of the *Pseudomonas putida* PpG7 Salicylate Hydroxylase Gene (*nahG*) and Its 3′-Flanking Region[†,‡]

In-Soon You,[*,§] Debabrota Ghosal,[∥] and Irwin C. Gunsalus[§]

Department of Biochemistry, University of Illinois, Urbana, Illinois 61801, and DST Unit on Genetic Engineering, Center for Biotechnology, Jawaharlal Nehru University, New Delhi, India

Received August 8, 1990; Revised Manuscript Received November 2, 1990

ABSTRACT: Gene *nahG* of naphthalene/salicylate catabolic plasmid NAH7 encodes a protein of molecular weight 45 000, salicylate hydroxylase. This enzyme catalyzes the formation of catechol from salicylate, a key intermediate in naphthalene catabolism. DNA sequence analysis of the 3.1-kilobase *Hin*dIII fragment containing the *nahG* locus reveals an open reading frame (ORF) of 1305 base pairs that corresponds to a protein of 434 amino acid residues. The predicted amino acid sequence of salicylate hydroxylase is in agreement with the molecular weight, NH_2-terminal amino acid sequence, and total amino acid composition of the purified salicylate hydroxylase [You, I.-S., Murray, R. I., Jollie, D., & Gunsalus, I. C. (1990) *Biochem. Biophys. Res. Commun. 169*, 1049–1054]. The amino acid sequence between positions 8 and 37 of salicylate hydroxylase shows homology with known ADP binding sites of other FAD-containing oxidoreductases, thus confirming its biochemical function. The sequence of the *Pseudomonas putida* salicylate hydroxylase was compared with those of other similar flavoproteins. A small DNA segment (831 base pairs) disrupts the continuity of the known gene order *nahG* and *nahH*, the latter encoding catechol 2,3-dioxygenase. The complete nucleotide sequence of the intergenic region spanning genes *nahG* and *nahH* has been determined and its biological role proposed.

Pseudomonas putida PpG7 plasmid NAH7 carries all the genetic information necessary for the conversion of naphthalene to pyruvate and acetaldehyde. In plasmid NAH7, the naphthalene catabolic genes are physically and functionally organized in two operons, *nah* and *sal*, on a 25-kb *Eco*RI fragment (Yen & Gunsalus, 1982; Grund & Gunsalus, 1983). The positive regulatory gene *nahR*, which is divergently transcribed from the *sal* operon, is located between the *nah* and *sal* operons (Yen & Gunsalus, 1982, 1985; Schell, 1986). The *nah* operon (*nahABCDE*) encodes the enzymes responsible for the conversion of naphthalene to salicylate while the *sal* operon (*nahGHINLMJK*) specifies enzymes that convert salicylate to pyruvate and acetaldehyde via the meta-cleavage pathway. The gene order of the two operons follows the order of biochemical and enzymatic steps in the naphthalene catabolic pathway (Yen & Gunsalus, 1982; Harayama et al., 1987). The *nahG* gene, the most proximal gene of the *sal* operon, codes for salicylate hydroxylase.

Salicylate hydroxylase (E.C. 1.14.13.1), a flavoprotein, converts salicylate to catechol via decarboxylative hydroxylation (Figure 1). This enzyme was first purified from *P. putida* (Yamamoto et al., 1965) and later from *Pseudomonas cepacia* (Tu et al., 1981) and other soil microorganisms (White-Stevens & Kamin, 1972). Although the biochemistry of salicylate hydroxylase is known in detail (Wang & Tu, 1984;

[†] This work was supported in part by NIH Grant DK00562-33 and a grant in aid from Du Pont.
[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J05317.
[*] Correspondence should be addressed to this author. Present address: Department of Biology, California State University, Fresno, CA 93740-0073.
[§] University of Illinois.
[∥] Jawaharlal Nehru University.

0006-2960/91/0430-1635$02.50/0

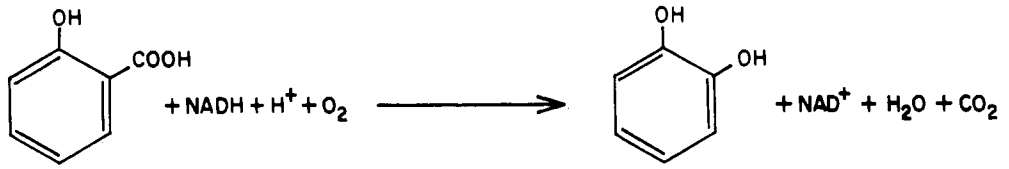

FIGURE 1: Conversion of salicylate to catechol by salicylate hydroxylase.

Wang et al., 1984; Einarsdottir et al., 1988), the genes specifying the enzyme are largely uncharacterized. Mechanistic studies with the salicylate hydroxylase isolated from *P. cepacia* suggest that in the first step salicylate and NADH bind to salicylate hydroxylase randomly, resulting in a reduced enzyme–substrate complex, which in turn binds to oxygen leading to the production of hydroxylated substrate, CO_2, and water (Wang et al., 1984). Recently, the *nahG*-encoded salicylate hydroxylase (SH)[1] has been purified from *P. putida* PpG7 (You et al., 1990).

The *nahG* gene of plasmid NAH7 was mapped earlier by restriction analysis of plasmid DNA prepared from several nahG::Tn5 mutants (Yen & Gunsalus, 1982, 1985). The *nahG* mutations were further mapped on the 1.5-kilobase (kb) *Sal*I fragment within a 3.1-kb *Hin*dIII fragment (Figure 2) containing the complete *nahG* gene (You & Gunsalus, 1986). The promoter and the transcription start site for the *sal* operon were shown to be located between the first *Sal*I and *Dde*I sites on the 3.1-kb fragment (Schell, 1986). It is of interest to study the primary structure of SH. In order to study the structure and function relationship of the salicylate hydroxylase encoded by plasmid NAH7, it is essential to obtain a functional *nahG* clone of minimum size. In this paper, we report the cloning and nucleotide sequence of the salicylate hydroxylase gene, *nahG*, from *P. putida* PpG7. Unusual features revealed by the DNA sequence analysis of the intergenic region, which spans the termination codon of *nahG* and the initiation codon of *nahH*, are discussed.

Materials and Methods

Bacterial Strains, Vector Plasmids, and Culture Conditions. Bacterial strains used were *P. putida* strains PpG7 and PpG1901 containing a wild-type NAH7 plasmid, PpG1900 (Yen & Gunsalus, 1982), and *Escherichia coli* JM83 (Vieira & Messing, 1982). *E. coli* strain CSR603 was used as the host for maxicell analysis of the plasmid-encoded proteins (Sancar et al., 1979). The cloning vectors were pKT240 (Bagdasarian et al., 1983) and pUC13 (Vieira & Messing, 1982), respectively. *E. coli* was grown in L-broth under appropriate antibiotic selective pressure (Maniatis et al., 1982), whereas *P. putida* strains were grown according to the procedure described by Yen and Gunsalus (1982, 1985). The clones containing *nahG* were tested either for the SH activity (Yamamoto et al., 1965; You et al., 1990) or for growth on salicylate as the sole source of carbon and energy.

DNA Isolation, Analysis, and Manipulation. Plasmid DNA was routinely prepared by the alkaline lysis method of Maniatis et al. and purified by centrifugation on two CsCl gradients. Restriction digestion, ligation, transformation, and gel electrophoresis of DNA were carried out according to the procedures described by Maniatis et al. (1982). Recombinant clones were screened by the rapid-analysis method of Birnboim and Doly (1979).

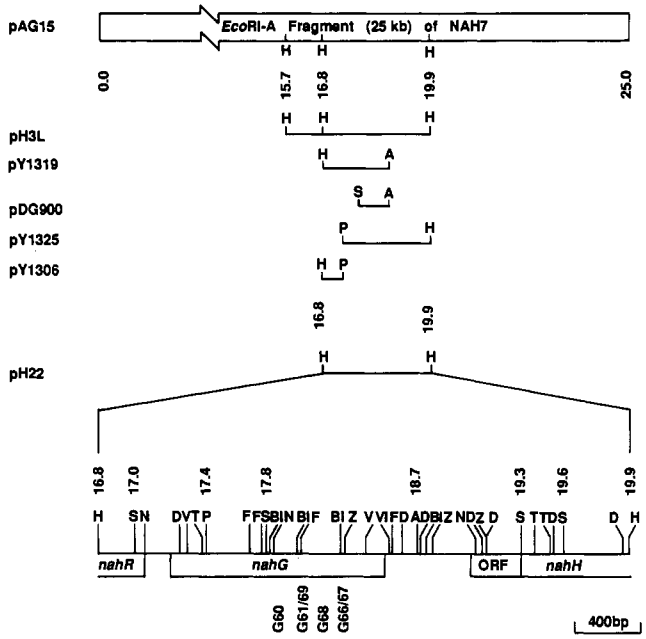

FIGURE 2: Physical and genetic map of the 3.1-kb *Hin*dIII fragment containing *nahG*. Clones and subclones of the *nahG* gene region are shown. Plasmids pY1319, pY1306, pY1325, and pDG900 were used for sequence determination. Protein coding regions are indicated by boxed areas. Coordinates are based on an *Eco*RI fragment (0–25 kb) containing naphthalene catabolic genes (Yen & Gunsalus, 1982; Grund & Gunsalus, 1983). The *nahG* mutations (G60, G61/69, G68, and G66/67) were previously described (Yen & Gunsalus, 1985; You & Gunsalus, 1986). Restriction sites are the following: A, *Hpa*I; Bl, *Bgl*I; D, *Dde*I; F, *Hin*fI; H, *Hin*dIII; N, *Nco*I; P, *Pst*I; S, *Sal*I; T, *Stu*I; V, *Pvu*II; VI, *Pvu*I; Z, *Sph*I.

Cloning and Subcloning of nahG. Plasmid pAG15 was constructed earlier by Grund and Gunsalus (1983). Construction of plasmids pH22 and pH3L was described by You and Gunsalus (1986) and that of pY1325 by Ghosal et al. (1987). A 1.9-kb *Hin*dIII–*Hpa*I fragment containing the *nahG* gene was cloned from NAH7 into pUC13 that had been cleaved with *Hin*dIII and *Sma*I, to give rise to pY1319. Plasmid pY1306 was derived from pH22 after digestion with *Hin*dIII and *Pst*I and subsequent ligation into pUC13. For the single-stranded dideoxy sequencing, a 900-bp *Sal*I–*Hpa*I fragment (from pY1319) containing the right half of the *nahG* gene (coordinates 17.8–18.7, Figure 2) was subcloned in a phasmid vector pMa5 (Friedrich et al., unpublished results). The recombinant plasmid thus obtained was designated pDG900.

Sequencing of nahG and Its Downstream Region. Nucleotide sequences were initially determined by the chemical method of Maxam and Gilbert (1980) and later by the chain-termination procedure of Sanger et al. (1977). Radiolabeled DNA fragments were prepared by using either [γ-^{32}P]ATP and T4 polynucleotide kinase or [α-^{32}P]dATP or [α-^{32}P]dCTP and the Klenow enzyme, the large fragment of *E. coli* DNA polymerase I. For the dideoxy DNA sequencing,

[1] Abbreviations: aa, amino acid(s); ADP, adenine dinucleotide phosphate; bp, base pair(s); FAD, flavin adenine dinucleotide; kb, kilobase(s); NAD(P)H, reduced nicotine adenine dinucleotide (phosphate); NAH7, naphthalene-degrading plasmid; *nahG*, a gene that encodes salicylate hydroxylase; nt, nucleotides; ORF, open reading frame; PAGE, polyacrylamide gel electrophoresis; PHBH, *p*-hydroxybenzoate hydroxylase; RBS, ribosome-binding site; SDS, sodium dodecyl sulfate; SH, salicylate hydroxylase.

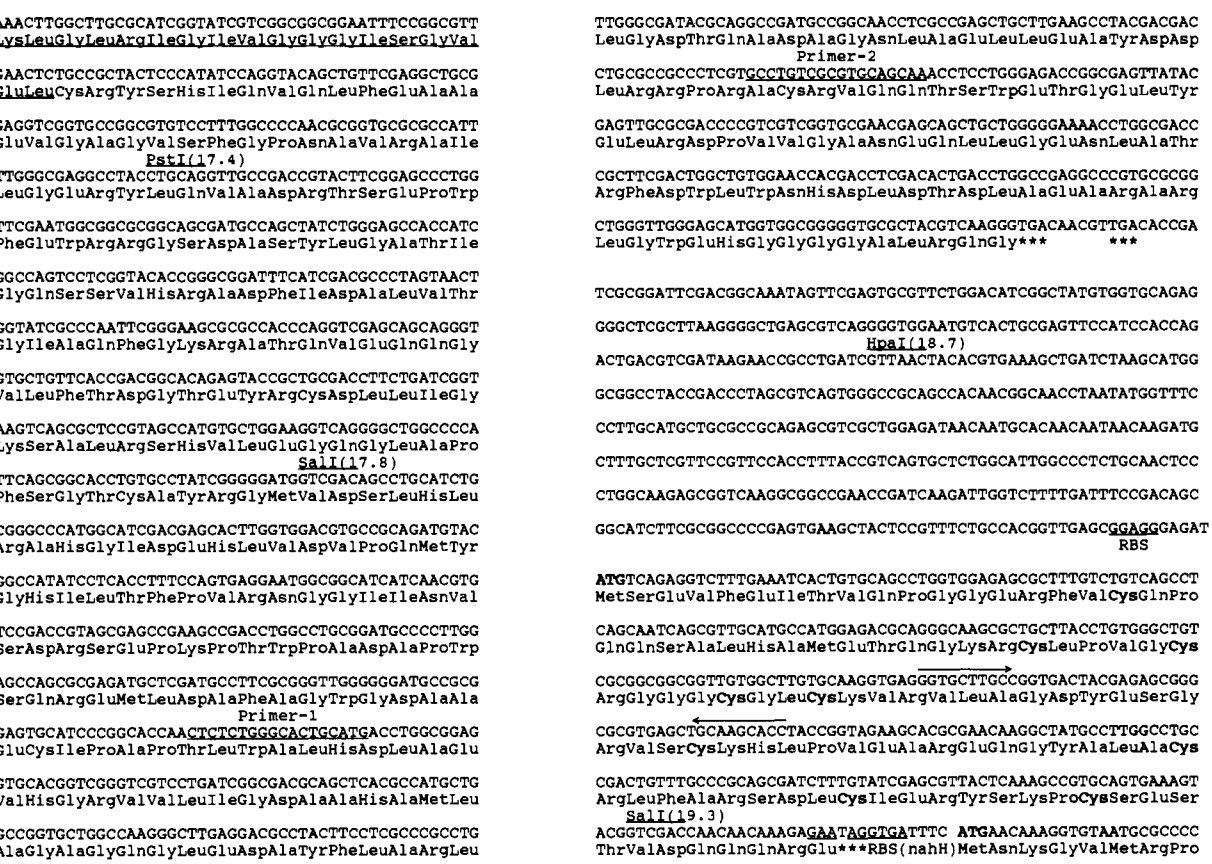

FIGURE 3: DNA sequence of salicylate hydroxylase and its 3′-flanking region. The given sequence is that of the coding strand. The deduced amino acid (aa) sequences for the proposed reading frames are shown in their three-letter abbreviations. The underlined aa's were confirmed by the NH_2-terminal aa sequencing of the purified salicylate hydroxylase from *P. putida* PpG7. RBS denotes a putative ribosome binding site. The underlined nucleotide sequences (two 18-mers) are the synthetic oligomers that were used as primers for the plasmid dideoxy sequencing. Also shown are key restriction sites with their coordinates. Termination codons are indicated by asterisks. Arrows show repeated sequences. Cysteine residues in the putative polypeptide encoded by the 327-bp ORF are shown in boldface.

a Sequenase (Tabor & Richardson, 1985) kit purchased from U.S. Biochemical Corp. was used. Two (forward and reverse) M13 sequencing primers from Boehringer Mannheim Biochemicals were used to check the sequences surrounding the *Pst*I site (coordinate 17.4 in Figures 2 and 3) within the *nahG* gene. Also, two octadecadeoxyoligonucleotides (see Figure 3) designed on the basis of predetermined sequences were prepared by a DNA synthesizer and used for sequencing pDG900 in order to confirm the nucleotide sequences that had been obtained by the chemical method.

Enzymes and Chemicals. Restriction endonucleases, DNA modifying enzymes, isopropyl β-D-thiogalactopyranoside, 5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside, and antibiotics were purchased from New England Biolabs, Boehringer Mannheim Biochemicals, or Sigma Chemical Co. Radiochemicals [γ-^{32}P]ATP, [α-^{32}P]dCTP, [α-^{32}P]dATP, and ^{35}S-dATP were obtained from New England Nuclear or Amersham. All other chemicals were of molecular biology or of reagent grade.

Results

Cloning and Expression of the nahG Gene. Summarized in Figure 2 are clones and subclones used in this study, which contain various regions of *nahG*. Also shown is the genetic organization of the 3.1-kb *Hin*dIII fragment containing genes *nahR*, *nahG*, and *nahH*. Two adjacent 4.2-kb *Hin*dIII fragments when cloned into the *Hin*dIII site of the broad host range vector pKT240 (i.e., pH3L in Figure 2) enabled *P. putida* PpG1900 to grow on salicylate as the sole source of carbon and energy. Therefore, plasmid pH3L contains a functional *nahG* gene in addition to the regulator *nahR* that was shown to be present (You & Gunsalus, 1986). Earlier studies (Ghosal et al., 1987; You et al., 1988) showed that the NH_2 terminus of the *nahH* protein lies next to the *Sal*I site (19.3), while that of *nahR* is at coordinate 17.1. Based on these data, plasmid pY1319 containing a 1.9-kb *Hin*dIII–*Hpa*I fragment was constructed and analyzed for the *nahG* gene.

E. coli JM83 containing pY1319 showed SH activity: 2 and 8 milliunits in the absence and presence of a gratuitous inducer, anthranilate, respectively. One milliunit corresponds to 1 nmol of NADH oxidized per minute per milligram of protein. *E. coli* CSR603 containing pY1319 was subjected to maxicell analysis for visualization of the *nahG* gene product. Subsequent sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) of the radiolabeled proteins by the Laemmli method (1970) failed to show a band corresponding to the *nahG* polypeptide.

Nucleotide Sequence of the nahG Gene. The complete nucleotide (nt) sequence of *nahG* is presented in Figure 3. The sequence reveals a long open reading frame (ORF) of 1305 bp capable of encoding 434 amino acids (aa). The ORF includes the initiation codon ATG at position 1 and the termination codon TGA at position 1305. SDS–PAGE of the purified *nahG* gene product (SH) showed a molecular weight of 45 000, which corresponds to approximately 450 aa residues. Since the NH_2-terminal aa sequence of the first 25 residues (underlined) of the purified SH protein was known (You et al., 1990), the 5′-end of the gene was placed accordingly (i.e., 416-bp downstream of the *Hin*dIII site at 16.8). The 5′-end of the SH gene delineates the 3′-end to be ca. 1800 bp away from the *Hin*dIII site (16.8), which is consistent with the DNA

	x y - y - G - G - - G - - - y - - y (loop)	y - y - z	
AlkT	A I V V V G A G T A G V N A A F W L 2 19	I F S R E 30 34	
CamA	N V V I V G T G L A G V E V A F G L 6 23	R L V G D 33 37	
AlkT	S V V V L G G G V I G L E V A S A A 144 161	V T V I E 168 172	
TodA	R L L I V G G G L I G C E V A T T A 144 161	V T I L E 168 172	
NahG	R I G I V G G G I S G V A L A L E L 8 25	V Q L F E 33 37	
TodA	H V A I I G N G V G G F T T A Q A L 3 20	S L I G D 30 34	
PHBH	Q V A I I G A G P S G L L L G Q L L 4 21	N V I L E 28 32	
TfdB	D V L V V G T G P A G A S A G A L L 8 25	T M L I N 32 36	

FIGURE 4: Homology between the NH_2-terminal portion of SH and other ADP binding sites. A proposed ADP binding region of SH is shown in comparison with other aa sequences involved in ADP binding. Shown at the top is an aa sequence "fingerprint" with ADP-binding properties (Wierenga et al., 1986). Amino acids that are in agreement with the fingerprint are shown in boldface. The length of loop varies from 7 to 11. Abbreviations: (x) K, R, H, S, T, Q, N; (y) A, I, L, V, M, C; (z) D, E. Alignment of AlkT with CamA was published by Eggink et al. (1990) and the PHBH sequence was from Wierenga et al. (1983). ADP binding sites for TfdB (Perkins et al., 1990) and TodA (Zylstra & Gibson, 1989) are proposed on the basis of their sequence similarity to the ADP binding fingerprint. Abbreviations: AlkT, rubredoxin reductase; CamA, putidaredoxin reductase; NahG, salicylate hydroxylase; PHBH, *p*-hydroxybenzoate hydroxylase; TfdB, 2,4-dichlorophenol hydroxylase; TodA, toluene reductase.

sequence data (Figures 2 and 3).

The aa sequence deduced from the *nahG* DNA sequence presented in Figure 3 is supported by the aa composition of the purified SH (Table I). The calculated molecular weight of the SH from *P. putida* PpG7 is 46818, which is also in agreement with the SDS–PAGE analysis of salicylate hydroxylase. The NH_2-terminal aa methionine is processed off. The G + C content of the *nahG* coding region is 65 mol %, which is close to that (63%) of *P. putida* (Normore, 1976) and suggests its origin from the genus *Pseudomonas*. The preferential usage (76%) of C- and G-ending codons was also observed. There are 35 basic and 57 acidic amino acids, indicating an acidic character of the SH protein.

Nucleotide Sequence of the 3'-Flanking Region of the nahG Gene. The nt sequence of the intergenic region spanning the termination codon of *nahG* and the initiation codon of *nahH* has been determined. DNA sequencing revealed a 831-bp gap to which no function had been assigned. The intergenic region contains several short ORFs. Of these, an ORF of 327 bp starts with an ATG codon and is preceded by a ribosome binding site (RBS) like sequence, e.g., GGAGG (Figure 3). No mutation has been documented in this intergenic region nor in the adjacent *nahH* locus. The latter encodes catechol 2,3-dioxygenase and is well characterized by cloning, sequencing, and functional analysis (Ghosal et al., 1987).

The 327-bp ORF possesses the following characteristics: It is capable of coding for a putative polypeptide of 108 aa with a molecular weight of 11913. The predicted polypeptide is rich in cysteine residues, a characteristic of many redox proteins. No similarity was found between the deduced aa sequence of the 327-bp ORF and any sequence contained in the NIH bacterial protein data base. The cysteine cluster that spans residues 40–48 in the putative protein, however, resembles a motif CXXXXCXXC found in the iron–sulfur binding region of plant and cyanobacterial ferredoxins (Cammack, 1983). The motif provides an insight into the function of the 327-bp ORF-encoded protein. The 327-bp ORF has a G + C content of 57 mol %. The last two codons and the termination codon of the 327-bp ORF overlap with the RBS of the distal gene, *nahH*, thus implying a translational coupling of the two genes. An inverted repeat (9 bp) of unknown significance exists within the 327-bp ORF, which can potentially form a stable stem–loop structure (Figure 3).

Table I: Amino Acid Composition of Salicylate Hydroxylase

amino acid	predicted[a]	direct anal.[c]
Ala	53	51
Arg	30	29
Asn	9	
Asp	28 (37)[b]	37
Cys	5	6
Gln	20	
Glu	29 (49)[b]	45
Gly	53	60
His	14	12
Ile	17	17
Leu	51	55
Lys	5	6
Met	4	5
Phe	13	14
Pro	19	19
Ser	16	19
Thr	15	15
Trp	11	
Tyr	11	11
Val	30	31
total	433	432
MW	46818	45000

[a] Predicted from the DNA sequence by using lysine as the NH_2-terminal residue. [b] Represents the sum of Asn + Asp and Gln + Glu, respectively. [c] Based on direct analysis of the purified salicylate hydroxylase and those shown are the average values of two independent determinations (You et al., 1990).

```
                    20                                                  50
SH      MKNNKLGLRIGIVGGGISGVALALELCRYSHIQVQLFEAAPAF--GEVGAGVSFGPNAVR
             :  :::I:G:G SG: L:  L : :  :V L   :P :  G : AGV :    V
PHBH         MKTQVAIIGAGPSGLLLGQLLHKAGIDNVILERQTPDYVLGRIRAGV-LEQGMVD
                                                   (40)    ***

                          80                                        110
SH      AIVGLGLGEAYLQVADRTSEPWEDVWF--EWRRGSDASYLGATIAPGV-GQSSVHRADFI
         :   G::   :   :   E   :: F  ::RR  D   L::  :  V GQ: V R D::
PHBH    LLREAGVDRR-MARDGLVHEG-VEIAFAGQRRR-IDLKRLSGGKTVTVYGQTEVTR-DLM
                                                  (100)

                        140                                         170
SH      DALVTHLPEGIAQFGKRATQVEQQGGEVQ-VLFT-DGTEYR--CDLLIGADGIKSALRSH
        :A   :   : : :      :::: GE   V F  DG   R  CD : G DG: :  R :
PHBH    EA--REACGATTVYQAAEVRLHDLQGERPYVTFERDGERLRLDCDYIAGCDGFHGISR-Q
                                          (150)

                    190                                     220
SH      VLEGQGLAPQVPRFSGTCAYRGMVDSLHLREAYRAHGIDEHLV--DVPQMYLGLDGHILT
         : :: L  :V  F    :: G :: L L ::      : ::L:  : P: : :L ::
PHBH    SIPAERL--KV--FERVYPF-GWLG-L-LADT---PPVSHELIYANHPRGF-ALCSQ--R
                                          (200)

                    250                                     280
SH      FPVRNGGIINVVAFISDRSEPKPTWPADAPWVREASQREMLDAFAGWGDAARALLECIPA
         : R:   ::V :: ::: E    W: D  W  E   R :    A    :: :L   I A
PHBH    SATRSRYYVQV-PL-TEKVE---DWSDDRFWT-ELKAR-LPAEVAEKLVTGPSLEKSI-A
                                          (250)

                    310                                     340
SH      PTLWALHDLAELPGYVHGRVVLIGDAAHAMLPHQGA-GAGQGLEDAYFLARLLGDTQADA
        P L:::  : E P   HGR: L GDAAH ::P  GA G   :  D   L RLL  :  :
PHBH    P-LRSF--VVE-P-MQHGRLFLAGDAAH-IVPPTGAKGLNLAASDVSTLYRLLLKAYRE-
                                     *  **         (310)

                          370                                       400
SH      GNLAELLEAYDD--LRRPRACRVQQTS-WETGELYELRDPVVGANEQLLGENLATRFDWL
        G  :ELL  Y    LRR :  : :: S W T: L  : D   : ::::   :L   :  L
PHBH    GR-GELLGRYSAICLRRIW--KAERFSYWMTSVLHRFPDTD-AFSQRIQQTEL--EY-YL
                                    (350)

                                430
SH      -WNHDLDTDLAEARARLGWEHGGGGALRQG
          : :L T :AE    L  E: G
PHBH    GSEAGLAT-IAENYVGLPYEEIG
                  (390)
```

FIGURE 5: Amino acid sequence comparison between the salicylate hydroxylase from *P. putida* and the *p*-hydroxybenzoate hydroxylase from *P. fluorescens*. The SH aa sequence was aligned with the published PHBH sequence (Weijer et al., 1982). Coordinates at the top and the bottom (in parentheses) are for SH and PHBH, respectively. The colons represent conservative replacements. Asterisks indicate conserved amino acid residues between the two flavin hydroxylases, which are of known structural and functional importance in PHBH. Underlined sequences are the proposed second ADP binding sites for SH (this study) and PHBH (Eggink et al., 1990), respectively. Dashes have been introduced to obtain optimum alignment between the two hydroxylases.

Comparison of the nahG Protein with Other Oxidoreductases. The aa sequence between residues 8 and 37 of salicylate hydroxylase shows homology with a consensus sequence ("fingerprint") that is involved in ADP binding (Wierenga et al., 1986; Eggink et al., 1990). The potential ADP binding site of SH is shown in Figure 4. The ADP binding fingerprint forms a characteristic secondary structure called a $\beta\alpha\beta$-fold. The ADP-binding $\beta\alpha\beta$-fold is known to contact the bottom of the AMP portion of FAD in the FAD binding domain (Wierenga et al., 1986).

In comparison of the aa sequence predicted from the *nahG* gene to the known *p*-hydroxybenzoate hydroxylase (PHBH) aa sequence from *Pseudomonas fluorescens* (Weijer et al., 1982, 1983), the aa sequences were 25% homologous (Figure 5). The deduced *nahG* aa sequence was also compared with that of a plasmid-encoded phenol hydroxylase (Liu & Chapman, 1984; Perkins et al., 1990). SH shows 20% aa identity with phenol hydroxylase encoded by gene *tfdB* (data not shown). The SH gene shows no significant overall similarity with any other sequences in the GenBank.

Discussion

The biochemistry and genetics of several dioxygeneses that catalyze fission of the aromatic nucleus have been studied in detail. Amino acid sequences derived from the corresponding gene sequences were reported for dioxygenase genes *xylE* (Nakai et al., 1983), *nahH* (Ghosal et al., 1987; Harayama et al., 1987), *bphC* (Taira et al., 1988), and others. High and low homologies were observed among them at the nt and aa sequence level. The sequence divergence may reflect individual differences in their functions. Such detailed genetic studies are lacking in the case of monooxygenases, though their biochemistry is rather well-known (Yamamoto et al., 1965; Tu et al., 1981; Weijer et al.; 1983; Wierenga et al., 1979, 1983, 1986).

Gene *nahG* encoding salicylate hydroxylase, an FAD-de-

pendent monooxygenase, has been well characterized genetically, whereas little was known about the gene product until recently. We have cloned a SH gene, and the complete nt sequence has been determined. The NH_2-terminal amino acid of the SH enzyme is lysine. The SH protein was expressed poorly from the cloned *nahG* gene, most likely due to the nature of positive regulation of the *sal* operon by regulator *nahR*. A similar result was obtained when an in vitro coupled transcription/translation system was used (data not known). The predicted aa sequence of SH is in good agreement with the data obtained from the direct analysis of the purified SH as shown in Table I. Conforming with a high G + C content of the *nahG* gene (65%), 76% of the codons used in *nahG* ends with C or G, as noted with other *Pseudomonas* genes (Nakai et al., 1983). The comparison of the *nahG*-encoded SH with a known *P. putida* SH indicated that they were different in terms of the size, composition, and the NH_2-terminal aa (You et al., 1990). It is not known whether the latter is also plasmid-encoded.

Earlier studies by us and others showed remarkable homology between genes of the TOL plasmid pWWO and the NAH7 plasmid, which specify aromatic ring fission enzymes (e.g., catechol 2,3-dioxygenase coding genes, *xylE* and *nahH*). The above observation led to the following proposal. A cluster of genes coding for the enzymes involved in catechol catabolism may have moved as a discrete segment into a region downstream of the salicylate hydroxylase coding gene (*nahG*) to form an operon structure (i.e., *sal* operon) for the existing coordinated expression of the salicylate pathway enzymes. The presence of a 831-bp intergenic region between *nahG* and *nahH* supports the above notion. The intergenic region shows sequence homology with the corresponding region of the TOL plasmid (Harayama et al., 1987; S. Harayama, personal communication).

Unlike the *nahG* coding region, the proposed ORF (327 bp) found in the 3′-flanking region of *nahG* is relatively low both in total G + C content and in G- or C-ending codons. Sharp deviation of the spacing of cysteinyl residues in the putative polypeptide (encoded by the 327-bp ORF) from the conserved common structural features of other bacterial ferredoxins (Tanaka et al., 1974; Mulligan et al., 1988) makes it difficult to propose the putative gene as the one encoding a redox protein. The presence of a cysteine-rich core sequence (CXXXXCXXC) in this small polypeptide, however, suggests its presumptive role as an electron carrier in salicylate catabolism. Oligonucleotide site-directed mutagenesis and biochemical characterization of the proposed protein will undoubtedly answer the question.

Among the flavin hydroxylases involved in aromatic catabolism, SH has been investigated most extensively with respect to the enzymatic mechanism (Kamin et al., 1976; Wang & Tu, 1984; Einarsdottir et al., 1988). Although the primary structure, functional domains, and tertiary structure of PHBH had long been determined by using amino acid sequencing and X-ray crystallography (Weijer et al., 1979; Wierenga et al., 1983), only recently has the primary structure of a plasmid-encoded phenol hydroxylase become available (Perkins et al., 1990). Unlike most dioxygenases involved in aromatic metabolism, the above enzymes use NAD(P)H directly for the production of a reduced enzyme–FAD complex.

It appears that ADP binding sites of known oxidoreductases are conserved through evolution; they may have originated from a single primordial dinucleotide binding domain by gene fusion with one or more other domains. This view is reflected in the observation that the overall homology between SH and PHBH is only 25% at the level of amino acids while that between SH and phenol hydroxylase is 20%. The limited homologies observed are, however, not uniformly distributed throughout the sequences. Rather, the homologies are localized in patches, e.g., in FAD binding sites (Figure 4) and in the known catalytic regions of PHBH (Figure 5). On the basis of the comparison of the *nahG* aa sequence with PHBH, a tentative location for a second ADP binding $\beta\alpha\beta$-fold could be proposed to be residues 304–314 of SH (Weijer et al., 1983; Eggink et al., 1990). The putative second binding site for ADP includes Gly 313, Asp 314, and four hydrophobic residues of SH from 310 to 313 (underlined sequences in Figure 5).

Although the substrate binding and catalytic active sites of SH remain to be elucidated, the aa sequence comparison between SH and PHBH led to two interesting observations. One of them is that the aa sequence of SH between residues 48 and 50 is very similar to that of PHBH between residues 45 and 47 (Figure 5). The same is true for the residues of SH from 319 to 325 and those from 290 to 296 of PHBH. Also, Gly 49 and Gly 324 of SH nicely align with Gly 46 and Gly 295 of PHBH, respectively. The above two sites of PHBH (denoted by asterisks in Figure 5) are known to be involved in keeping the substrate *p*-hydroxybenzoate in proper position with respect to the carboxyl group of the substrate and the isoalloxazine ring of FAD (Weijer et al., 1983).

The second observation is that the aa residues of PHBH that bury the substrate in a hydrophobic environment are also conserved in SH, e.g., residues at 47 (Val), 199 (Leu), 210 (Leu), 292 (Pro), and 296 (Ala) of PHBH (Weijer et al., 1983). In spite of low aa homology (25%) between SH and PHBH, therefore, it is reasonable to propose their emergence from a common ancestor through divergent evolution. Higher homology (50%) between SH and PHBH was observed when the aa were grouped on the basis of their charge characteristics. It is of note that both benzoate and *p*-hydroxybenzoate bind to SH with almost equal affinity, though they remain unmodified unlike its normal substrate, salicylate (Wang et al., 1984; Einasdottir et al., 1988). Detailed biochemical and genetic studies of catalytically and structurally important residues in SH will aid the further understanding of the enzymatic mechanism.

ACKNOWLEDGMENTS

We thank Dr. K. H. Friedrich (Max-Planck-Institut für Biochemie, Germany) for synthetic oligonucleotides and phasmid vector pMa5, Dr. S. Harayama (University of Geneva, Switzerland) for helpful discussion, and Dr. H. Ono (California State University, Fresno) for the critical reading of the manuscript.

REFERENCES

Bagdasarian, M. M., Amann, E., Lurz, R., Ruckert, B., & Bagdasarian, M. (1983) *Gene 26*, 278–282.

Birnboim, H. C., & Doly, J. (1979) *Nucleic Acids Res. 7*, 1513–1523.

Cammack, R. (1983) *Chem. Scr. 21*, 87–95.

Eggink, G., Engel, H., Vriend, G., Terpstra, P., & Witholt, B. (1990) *J. Mol. Biol. 212*, 135–142.

Einarsdottir, G. H., Stankovich, M. T., & Tu, S.-C. (1988) *Biochemistry 27*, 3277–3285.

Ghosal, D., You, I.-S., & Gunsalus, I. C. (1987) *Gene 55*, 19–28.

Grund, A. D., & Gunsalus, I. C. (1983) *J. Bacteriol. 156*, 89–94.

Harayama, S., Rekik, M., Wasserfallen, A., & Baroch, A. (1987) *Mol. Gen. Genet. 210*, 241–247.

Kamin, H., White-Stevens, R. H., & Presswood, R. P. (1976) *Methods Enzymol. 53*, 527–543.
Laemmli, U. K. (1970) *Nature (London) 227*, 680–685.
Liu, T., & Chapman, P. J. (1984) *FEBS Lett. 173*, 314–318.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.
Morris, C. M., & Barnsley, E. A. (1982) *Can. J. Microbiol. 28*, 73–79.
Mulligan, M. E., Buikema, W. J., & Haselkorn, R. (1988) *J. Bacteriol. 170*, 4406–4410.
Nakai, C., Kagamiyama, H., Nozaki, M., Nakazawa, T., Inouye, S., Ebina, Y., & Nakazawa, A. (1983) *J. Biol. Chem. 258*, 2923–2928.
Normore, W. M. (1976) in *Handbook of Biochemistry and Molecular Biology* (Fasman, G. D., Ed.) 3rd ed., pp 65–235, The Chemical Rubber Co., Cleveland, OH.
Perkins, E. J., Gordon, M. P., Caceres, O., & Lurquin, P. F. (1990) *J. Bacteriol. 172*, 2351–2359.
Sancar, A., Hack, A. M., & Rupp, W. D. (1979) *J. Bacteriol. 137*, 692–693.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Schell, M. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 79*, 369–373.
Shine, J., & Dalgarno, L. (1975) *Nature (London) 254*, 34–38.
Tanaka, M., Haniu, M., Yasunobu, K., Dus, K., & Gunsalus, I. C. (1974) *J. Biol. Chem. 249*, 3689–3701.
Tu, S.-C., Romero, F. A., & Wang, L.-H. (1981) *Arch. Biochem. Biophys. 209*, 423–432.
Vieira, J., & Messing, J. (1982) *Gene 19*, 259–268.
Wang, L.-H., & Tu, S.-C. (1984) *J. Biol. Chem. 259*, 10682–10688.
Wang, L., Tu, S., & Lusk, R. C. (1984) *J. Biol. Chem. 259*, 1136–1142.
Weijer, W. J., Hofsteenge, J., Vereijken, J. M., Jekel, P. A., & Beintema, J. J. (1982) *Biochim. Biophys. Acta 704*, 385–388.
Weijer, W. J., Hofsteenge, J., Beintema, J. J., Wierenga, K., & Drenth, J. (1983) *Eur. J. Biochem. 133*, 109–118.
White-Stevens, R. H., & Kamin, H. (1972) *J. Biol. Chem. 247*, 2358–2370.
Wierenga, R. K., & Drenth, J. (1983) *J. Mol. Biol. 167*, 725–739.
Wierenga, R. K., De Jong, R. J., Kalk, K. H., Hol, W. G. J., & Drenth, J. (1979) *J. Mol. Biol. 131*, 55–73.
Wierenga, R. K., Terpstra, P., & Hol, W. G. J. (1986) *J. Mol. Biol. 187*, 101–107.
Yamamoto, S., Katagiri, M., Maeno, H., & Hayaishi, O. (1965) *J. Biol. Chem. 240*, 3408–3413.
Yen, K.-M., & Gunsalus, I. C. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 874–878.
Yen, K.-M. & Gunsalus, I. C. (1985) *J. Bacteriol. 162*, 1008–1013.
You, I.-S., & Gunsalus, I. C. (1986) *Biochem. Biophys. Res. Commun. 141*, 986–92.
You, I.-S., Ghosal, D., & Gunsalus, I. C. (1988) *J. Bacteriol. 170*, 5409–5415.
You, I.-S., Murray, R. I., Jollie, D., & Gunsalus, I. C. (1990) *Biochem. Biophys. Res. Commun. 169*, 1049–1054.
Zylstra, G. J., & Gibson, D. T. (1989) *J. Biol. Chem. 264*, 14940–14948.