Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 6869–6874, July 1996
Biochemistry

Ferulate-5-hydroxylase from *Arabidopsis thaliana* defines a new family of cytochrome P450-dependent monooxygenases

(phenylpropanoid/T-DNA tagging)

KNUT MEYER*, JOANNE C. CUSUMANO*, CHRIS SOMERVILLE†, AND CLINT C. S. CHAPPLE*‡

*Department of Biochemistry, Purdue University, West Lafayette, IN 47907-1153; and †Plant Biology Department, Carnegie Institution of Washington, Stanford, CA 94305-4101

Communicated by Eric Conn, University of California, Davis, CA, April 9, 1996 (received for review January 2, 1996)

ABSTRACT **The *fah1* mutant of *Arabidopsis* is defective in the accumulation of sinapic acid-derived metabolites, including the guaiacyl–syringyl lignin typical of angiosperms. Earlier results indicated that the FAH1 locus encodes ferulate-5-hydroxylase (F5H), a cytochrome P450-dependent monooxygenase (P450) of the general phenylpropanoid pathway. We have cloned the gene encoding this P450 by T-DNA tagging and have confirmed the identity of the cloned gene by complementation of the mutant phenotype. F5H shows 34% amino acid sequence identity with the avocado ripening-induced P450 CYP71A1 and 32% identity with the flavonoid-3′,5′-hydroxylases of *Petunia hybrida*. In contrast, it shares much less homology with cinnamate-4-hydroxylase, a P450 that catalyzes the hydroxylation of cinnamic acid three steps earlier in the general phenylpropanoid pathway. Since the highest degree of identity between F5H and previously sequenced P450s is only 34%, F5H identifies a new P450 subfamily that has been designated CYP84.**

Cytochrome P450-dependent monooxygenases (P450s) are a large group of heme-containing enzymes, most of which catalyze NADPH- and O_2-dependent hydroxylation reactions. In plants, two P450s catalyze reactions of the general phenylpropanoid pathway: cinnamate-4-hydroxylase (C4H) and ferulate-5-hydroxylase (F5H). The products of the general phenylpropanoid pathway are critical to plant survival and include UV-absorptive plant secondary metabolites such as flavonoids and hydroxycinnamic acid esters, as well as lignin, a major structural component of the plant secondary cell wall (1) (Fig. 1). Lignin often comprises over 20% of the dry weight of plant cell walls and is, after cellulose, the second most abundant biopolymer in Nature (2).

Several genes encoding plant P450s have been cloned by protein purification-based strategies (3–9); however, the instability, low abundance, and membrane-bound nature of these enzymes makes this approach problematic (10). Because of these difficulties, P450 genes are ideal candidates for cloning by alternative strategies such as T-DNA tagging. For example, F5H has been reported to be a P450, but its detergent instability has precluded its purification (11). We have previously reported that the *fah1* mutant of *Arabidopsis thaliana* appears to be defective in F5H activity (12). The *fah1* mutant fails to accumulate sinapoyl malate, a major phenolic secondary metabolite of *A. thaliana*, and is defective in the synthesis of the guaiacyl-syringyl lignin typical of angiosperms. In mutant screens, *fah1* mutants can be readily identified by their characteristic red fluorescence when observed under UV light. This phenomenon is due to the fact that sinapoyl malate is a fluorescent secondary metabolite that is accumulated in the adaxial leaf epidermis. As a result, wild-type *Arabidopsis* exhibits a pale blue fluorescence under UV, while *fah1* mutants appear dark red because of the fluorescence of chlorophyll in the subtending mesophyll (12). Based upon this visual screen, we have identified a T-DNA-tagged allele of *fah1* and have used the tagged line to clone the gene encoding F5H. These results indicate that T-DNA tagging of plant P450 genes is a useful approach for the cloning of the genes encoding these enzymes that are generally recalcitrant to purification.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Plant Material. *A. thaliana* was grown under a 16-h light/8-h dark photoperiod at 100 μE m^{-2} s^{-1} at 24°C, cultivated in Metromix 2000 potting mixture. Mutant lines *fah1-1* through *fah1-5* were identified by thin-layer chromatography (TLC) as described previously (12). Based upon the red fluorescence of *fah1* mutants under UV light, mutant lines *fah1-6*, *fah1-7*, and *fah1-8* were selected from ethyl methanesulfonate (*fah1-6*, *fah1-7*) or fast neutron (*fah1-8*) mutagenized populations of Landsberg *erecta* M2 seed. The T-DNA tagged line 3590 (*fah1-9*) was similarly identified in the DuPont T-DNA tagged population (13). All lines were backcrossed to wild type at least twice prior to experimental use to remove unlinked background mutations.

Secondary Metabolite Analysis. Leaf extracts were prepared from 100-mg samples of fresh leaf tissue suspended in 1 ml of 50% methanol. Samples were vortexed briefly and then frozen at −70°C. Samples were thawed, vortexed, and centrifuged at 12,000 × *g* for 5 min. Sinapoyl malate content was qualitatively determined by UV fluorescence following chromatography of extracts on silica gel TLC plates in a mobile phase of *n*-butanol/acetic acid/water (5:2:3).

Analysis of Nucleic Acids. For Southern analysis, DNA was extracted from leaf material (14), digested with restriction endonucleases, electrophoretically separated, transferred to Hybond N^+ membrane (Amersham), and hybridized with cDNA probes according to standard protocols (15). RNA was extracted from leaf material (16), electrophoretically separated, transferred to Hybond N^+ membrane (Amersham), and probed with radiolabeled probes prepared from cDNA or genomic clones according to standard protocols. Sequence analysis was performed on plasmid DNA using the U.S. Biochemical Sequenase kit version 2.0 using standard vector-based sequencing oligonucleotides or custom-synthesized oligonucleotides as appropriate.

Plasmid Rescue. Plasmid rescue was conducted using DNA isolated from the *fah1-9* mutant according to the protocol of Behringer and Medford (17) with minor modifications. Five μg of *Eco*RI-digested genomic DNA was incubated with 125 units

Abbreviations: P450, cytochrome P450-dependent monooxygenase; F5H, ferulate-5-hydroxylase; C4H, cinnamate-4-hydroxylase; T-DNA, transferred DNA; TLC, thin-layer chromatography.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. U38416).

‡To whom reprint requests should be addressed.

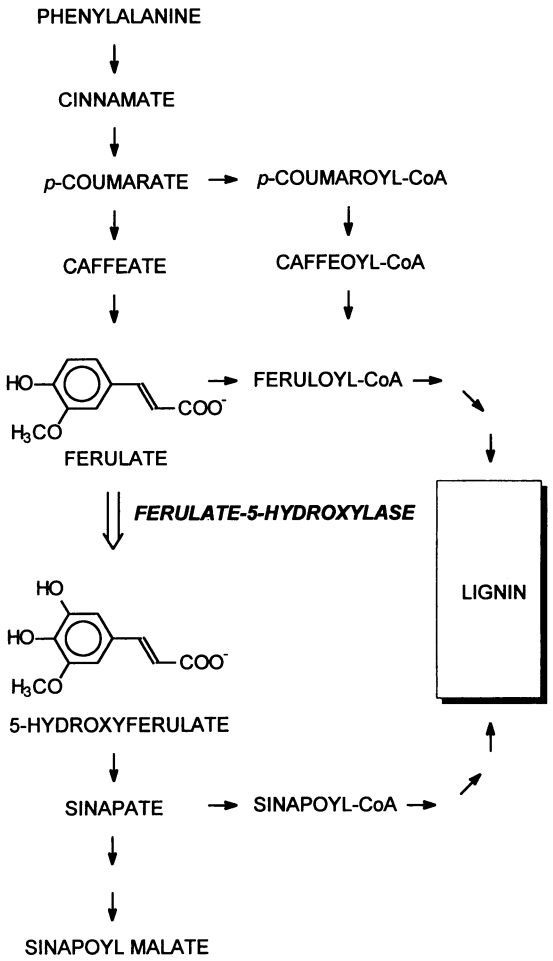

FIG. 1. The biosynthetic pathway for lignin and sinapic acid-derived secondary metabolites in *Arabidopsis*.

of T4 DNA ligase overnight at 14°C in a final volume of 1 ml. The ligation mixture was concentrated approximately 4-fold by two extractions with equal volumes of 2-butanol and was then ethanol-precipitated and electroporated into competent DH5-α cells.

Identification of cDNA and Genomic Clones. cDNA and genomic clones for F5H were identified by standard techniques (15) using a 2.3-kb *Sac*II/*Eco*RI fragment from the rescued plasmid (pCC1) as a probe. The cDNA clone pCC30 was identified in the λ-PRL2 library (18) kindly provided by Dr. Thomas Newman (U.S. Department of Energy Plant Research Laboratory, Michigan State University, East Lansing). A genomic cosmid library of *A. thaliana* (ecotype Landsberg *erecta*) generated in the binary cosmid vector pBIC20 (19) was screened with the radiolabeled cDNA insert derived from pCC30. Genomic inserts in the pBIC20 T-DNA are flanked by the neomycin phosphotransferase gene for kanamycin selection adjacent to the T-DNA right border sequence and the β-glucuronidase gene for histochemical selection adjacent to the left border. One positive clone (pBIC20-F5H) was characterized by restriction digestion and Southern analysis in comparison with *Arabidopsis* genomic DNA.

Complementation of the *fah1* Mutant. The pBIC20-F5H cosmid and the parental pBIC20 vector were introduced into *Agrobacterium tumefaciens* C58 pGV3850 (20) by electroporation (21). Stability of the complementation construct was confirmed by restriction analysis of plasmid DNA isolated from *A. tumefaciens*. Cultures harboring the binary vectors were used to transform the *fah1-2* mutant by vacuum infiltration (22). Kanamycin-resistant seedlings derived from independent infiltration experiments were grown in soil for 3 weeks. The sinapoyl malate content of leaf material from the transgenics was qualitatively determined by TLC as described above.

Using an adaptor-based cloning strategy, regulatory sequences upstream of the translation initiation site of the F5H gene isolated from the pBIC20-F5H cosmid were replaced with the strong constitutive CaMV 35S promoter (23). The resulting construct carries 2719 bp of the F5H genomic sequence driven by the CaMV 35S promoter fused 50 bp upstream of the inferred ATG start codon. As a result, the 35S promoter drives the expression of the F5H gene using the transcription start site of the viral promoter and the termination signal present on the F5H genomic sequence. This expression cassette was inserted into the T-DNA of the binary vector pGA482 (24) to give pGA482-35S:F5H, and this plasmid was introduced into *A. tumefaciens* by electroporation. Transformation of the *fah1-2* mutant with pGA482-35S:F5H and the empty pGA482 vector, isolation of kanamycin-resistant lines, and TLC assay for sinapoyl malate content was conducted as described above.

RESULTS

Identification of the T-DNA-Tagged Allele of *fah1*. A putatively T-DNA-tagged *fah1* mutant was identified in the DuPont collection of T-DNA-tagged lines (13) by screening adult plants under long wave UV light. A red fluorescent line (line 3590) was selected, and its progeny were assayed for sinapoyl malate content by TLC. The analyses indicated that line 3590 did not accumulate sinapoyl malate. Reciprocal crosses of line 3590 to a *fah1-2* homozygote, followed by analysis of the F1 generation for sinapoyl malate content, demonstrated that line 3590 was a new allele of *fah1*, and it was designated *fah1-9*.

Preliminary experiments indicated co-segregation of the kanamycin-resistant phenotype of the T-DNA-tagged mutant with the *fah1* phenotype. Selfed seed from seven kanamycin-resistant (*fah1-9* × *FAH1*) F1 plants segregated 1:3 for kanamycin resistance ($kan^{sensitive}/kan^{resistant}$), and 3:1 for sinapoyl malate deficiency (*FAH1/fah1*). From these lines, *fah1* plants gave rise to only $kan^{resistant}$, *fah1* progeny. To determine the genetic distance between the T-DNA insertion and the FAH1 locus, multiple test crosses were performed between a (*fah1-9* × *FAH1*) F1 and a *fah1-2* homozygote. The distance between the FAH1 locus and the T-DNA insertion was evaluated by determining the frequency at which $FAH1/kan^{resistant}$ progeny were recovered in the test cross F1. In the absence of crossover events, all kanamycin-resistant F1 progeny would be unable to accumulate sinapoyl malate and would thus fluoresce red under UV light. In 682 $kan^{resistant}$ F1 progeny examined, no sinapoyl malate proficient plants were identified, indicating a very tight linkage between the T-DNA insertion site and the FAH1 locus.

Plasmid Rescue and cDNA Cloning. Plasmid rescue was conducted using *Eco*RI-digested DNA prepared from homozygous *fah1-9* plants (17). DNA from rescued plasmids was double digested with *Eco*RI and *Sal*I. Plasmids generated from internal T-DNA sequences were identified by the presence of triplet bands at 3.8, 2.4, and 1.2 kb and were discarded. One plasmid (pCC1) giving rise to the expected 3.8 kb band plus a novel 5.6 kb band was identified as a putative external right border plasmid. Using a *Sac*II/*Eco*RI fragment of pCC1 that appeared to represent *Arabidopsis* DNA, a putative full-length cDNA clone (pCC30) for F5H was identified. The putative F5H clone carried a 1.9-kb *Sal*I/*Not*I insert, the sequence of which was determined. Blastx analysis (25) indicated that this cDNA encodes a cytochrome P450-dependent monooxygenase, consistent with earlier reports that (i) the *fah1* mutant is

defective in F5H (12) and (ii) F5H is a cytochrome P450-dependent monooxygenase (11).

Southern and Northern Analysis. To determine whether the putative F5H cDNA actually represented the gene that was disrupted in the *fah1* mutant lines, Southern analysis of DNA from nine independently-derived *fah1* alleles and three wild-type ecotypes was performed using the putative F5H cDNA as a probe (Fig. 2). These data indicated the presence of a restriction fragment length polymorphism between the tagged line and the wild types. Similarly, a restriction fragment length polymorphism was detected using DNA from the *fah1-8* mutant line (Fig. 2). The *fah1-8* line was generated with fast neutrons, a technique reported to cause deletion mutations. These data indicate that the genomic DNA of the *fah1-8* and *fah1-9* (the T-DNA-tagged line) alleles is disrupted in the region corresponding to the putative F5H cDNA.

A similar comparison was made among the nine *fah1* alleles for the abundance of transcript corresponding to the putative F5H cDNA. Northern blot analysis (Fig. 2) indicated that the putative F5H mRNA was represented at similar levels in leaf tissue of Columbia, Landsberg *erecta*, and Wassilewskija ecotypes and in the ethyl methanesulfonate-induced *fah1-1*, *fah1-4*, *fah1-5*, and *fah1-7*. Transcript abundance was substantially reduced in leaves from plants homozygous for the *fah1-2*, *fah1-3*, and *fah1-6*, all of which were ethyl methanesulfonate-induced, the fast neutron-induced mutant *fah1-8* and in the tagged line *fah1-9*. The mRNA in the *fah1-8* mutant also appears to be truncated. Taken together, these data provided strong correlative evidence that the cDNA clone that had been identified is encoded by the FAH1 locus.

Complementation of the *fah1* Mutation. To demonstrate the identity of the F5H gene at the functional level, the transformation-competent pBIC20 cosmid library (19) was screened for corresponding genomic clones using the full-length F5H cDNA as a probe. A clone (pBIC20-F5H) carrying a genomic insert of 17 kb that contains 2.2 kb of sequence upstream of the putative F5H start codon and 12.5 kb of sequence downstream of the stop codon of the F5H gene was transformed into the *fah1-2* mutant by vacuum infiltration. Thirty independent infiltration experiments were performed, and 167 kanamycin-resistant seedlings, representing at least three transformants from each infiltration, were transferred to soil and were analyzed with respect to sinapic acid-derived secondary metabolites. Of these plants, 164 accumulated sinapoyl malate in their leaf tissue as determined by TLC (Fig. 3). These complementation data indicate that the gene defective in the *fah1* mutant is present on the binary cosmid pBIC20-F5H.

To delimit the region of DNA on the pBIC20-F5H cosmid responsible for complementation of the mutant phenotype, a 2.7-kb fragment of the F5H genomic sequence was fused downstream of the CaMV 35S promoter in the binary plasmid pGA482, and this construct (pGA482-35S:F5H) was transformed into the *fah1-2* mutant. The presence of sinapoyl malate in 109 of 110 transgenic lines analyzed by TLC or by *in vivo* fluorescence under UV light indicated that the *fah1* mutant phenotype had been complemented (Fig. 3).

Sequence of the F5H cDNA. The F5H cDNA contains a 1560-bp open reading frame that encodes a protein with a molecular weight of 58,728. The putative ATG initiation codon is flanked by an A at −3 and a G at +4, in keeping with the nucleotides commonly found flanking the initiator methionine in plant mRNAs (26). Immediately following the inferred initiator methionine is a 17-amino acid sequence containing nine hydroxy amino acids (Fig. 4). The subsequent 15-amino acid sequence is rich in hydrophobic amino acids, including 11 hydrophobic residues comprised of phenylalanine, isoleucine, leucine, and valine residues. This hydrophobic stretch is immediately followed by an RRRR putative stop transfer sequence. F5H also shares significant sequence identity with other P450s. Following the stop transfer sequence near the amino terminus is the sequence PPGPRGWP, which obeys the consensus for the proline-rich sequence found in many P450s (27). Also notable is the stretch between Pro-450 and Gly-460. This region contains eight residues that comprise the heme-binding domain and are highly conserved among most P450s, one exception being allene oxide synthase from *Linum usitatissimum* (6). The Pro-450 to Gly-460 region contains Cys-458 in F5H, which by analogy is most likely the heme-binding ligand in this enzyme.

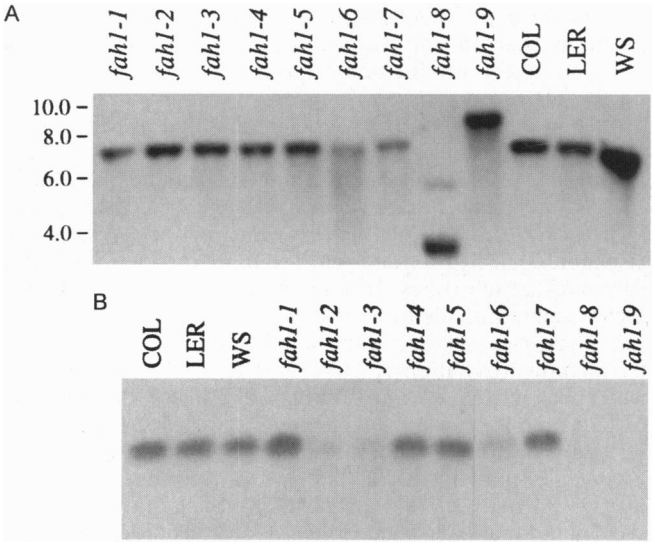

FIG. 2. (*A*) Southern blot analysis comparing hybridization of the F5H cDNA to *Eco*RI-digested genomic DNA isolated from wild type [ecotypes Columbia (COL), Landsberg *erecta* (LER), and Wassilewskija (WS)] and the nine *fah1* alleles. (*B*) Northern blot analysis comparing hybridization of the F5H cDNA to RNA isolated from wild type [ecotypes Columbia (COL), Landsberg *erecta* (LER), and Wassilewskija (WS)] and the nine available *fah1* alleles.

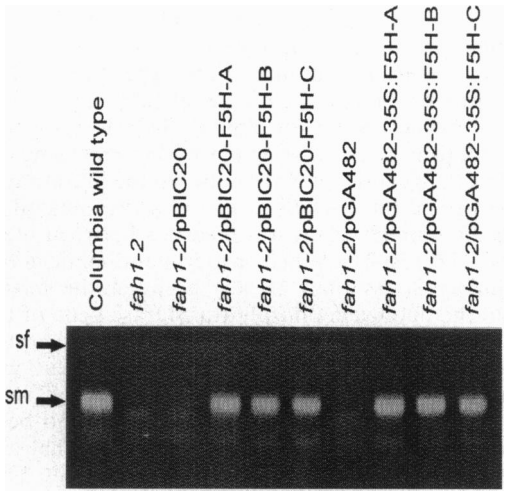

FIG. 3. Analysis of sinapic acid-derived secondary metabolites in wild type, *fah1-2*, and independently-derived transgenic *fah1-2* plants carrying the T-DNA derived from either the pBIC20-F5H cosmid or the pGA482-35S:F5H vector. Extracts prepared from plants transformed with the corresponding vectors lacking the F5H-derived sequences are included as controls. Extracts were prepared from leaf material as described in *Materials and Methods*, analyzed by TLC, and photographed under ultraviolet light. o, origin; sm, sinapoyl malate; sf, solvent front.

```
MESSISQTLSKLSDPTTSLVIVVSLFIFISFITRRRRPPY--PPGPRGWPIIGN-MLMMDQLTHRGLANL  F5H
MV-----LLSELAAAT--LIFLTTH IFISTLLSITNGRR-LPPGPRGWPVIGA-LPLLGAMPHVSLAKM  F3.5H
M---------DLLLIEKTLVALFAAIIGAILISKLRGKKFKLPPGPIPVPIFGNWLQVGDDLNHRNLTDL  C4H

AKKYGGLCHLRMGFLHMYAVSSPEVARQVLQVQDSVFSNRPATIAISYLTYDRADMAFAHYGPFWRQMRK  F5H
AKKYGAIMYLKVGTCGMVVASTPDAAKAFLKTLDLNFSNRPPNAGATHLAYGAQDMVFAHYGPRWKLLRK  F3.5H
AKRFGEILLLRMGQRNLVVSSPELAKEVLHTQGVEFGSRTRNVVFDIFTGKGQDMVFTVYGEHWRKMRR  C4H

VCVMKVFSRKRAE----SWASVR-DEVDKMVRSV--SCNVGKPINVGEQI-FALTRNIT----YRAAFGS  F5H
LSNLHMLGGKALE----NWANVRANELGHMLKSMFDMSREGERVVVAEMLTFAMANMIGQVILSKRVFVN  F3.5H
IMTVPFFTNKVVQQYRYGWEAEAAAVVDDVKK---NPAAATEGIVIRRRLQLMMYNNMFRIMFDRR-FES  C4H

ACEKGQDEFIRILQEFSKLFGAF--NVADFIPYFGWIDPQGINKRLVKARNDLDGFIDDIIDEHMKKKEN  F5H
K-GVEVNEFKDMVVELMTTAGYF--NIGDFIPCLAWMDLQGIEKGMKRLHKKFDALLTKMFDEHKATSYE  F3.5H
EDDPLFLKLKALNGERSRLAQSFEYNYGDFIPI-----LRPFLRNYLKLCKEVKDKRIQLFKDYFVD--E  C4H

QNAVDDGDVVDTDMVDDLLAFYSEEAKLVSETADLQNSIKLTRDNIKAIIMDVMFGGTETVASAIEWALT  F5H
RKG-------KPDFLDCVM-----------ENRDNSEGERLSTTNIKALLLNLFTAGTDTSSSAIEWALA  F3.5H
RKKIGSTKKMDNNQLKCA----------IDHILEAKEKGEINEDNVLYIVENINVAAIETTLWSIEWGIA  C4H

ELLRSPEDLKRVQQELAEVVGLDRRVEESDIEKLTYLKCTLKETLRMHPPIPLLLHETAEDTSI-DGFFI  F5H
EMMKNPAILKKAQGEMDQVIGNNRRLLESDIPNLPYLRAICKETFRKHPSTPLNLPRISNEPCIVDGYYI  F3.5H
ELVNHPEIQAKLRHELDTKLGPGVQITEPDVQNLPYLQAVVKETLRLRMAIPLLVPHMNLHDAKLGGFDI  C4H

PKKSRVMINAFAIGRDPTSWTDPDTFRPSRFLEPGVPDF--KGSNFEFIPFGSGRRSCPGMQLGLYALDL  F5H
PKNTRLSVNIWAIGRDPEVWENPLEFYPERFLSGRNSKIDPRGNDFELIPFGAGRRICAGTRMGIVMVEY  F3.5H
PAESKILVNAWWLANNPDQWKKPEEFRPERFLEEE-AKVEANGNDFRYLPFGVGRRSCPGIILALPILGI  C4H

AVAHILHCFTWKLPDGMKPSELDMNDV-FGLTAPKATRLFAVPTTRLICAL.                    F5H
ILGTLVHSFDWKLPSEV--IELNMEEA-FGLALQKAVPLEAMVTPRLPIDVYAPLA.               F3.5H
TIGRLVQNFELLPPPGQSKIDTDEKGGQFSLHILKHSTIVAKPRS------F.                    C4H
```

FIG. 4. Amino acid sequence alignment of phenylpropanoid pathway P450s: F5H (*Arabidopsis thaliana*), flavonoid-3′,5′-hydroxylase (*Petunia hybrida*) [F3,5H] and C4H (*Helianthus tuberosus*). Shaded residues represent amino acids that are identical to those of F5H.

DISCUSSION

The cloning of plant P450s has been hampered by the fact that these membrane-localized proteins are present in very low abundance and are often unstable to purification. While the genes encoding several plant P450s have been identified by the application of conventional biochemical approaches, many of the other cloned plant P450 genes have been identified by PCR-based approaches and represent genes of unknown function. The availability of mutants that are blocked in P450-catalyzed reactions provides an alternative method for the cloning of plant P450 genes and permits an unambiguous association of these genes with the function of their encoded proteins (28–30).

A substantial percentage of mutant alleles in T-DNA-tagged populations are not the direct result of a T-DNA insertional event (31). In mutants of this type, the T-DNA and the mutation of interest are not genetically linked, thus these mutants are not useful for cloning the affected gene. For this reason, prior to attempting plasmid rescue using the *fah1-9* mutant, the genetic distance between the mutation in the FAH1 locus and the T-DNA insertion was estimated by determining the number of crossover events between these two markers. The results of these experiments indicated a very tight linkage (less than 15 kb) between the *fah1-9* mutation and the neomycin phosphotransferase gene of the T-DNA, suggesting that the T-DNA may have integrated directly in the FAH1 locus. Following plasmid rescue and cDNA cloning, Southern and Northern analyses (Fig. 2) were then used to determine whether the cDNA that had been cloned represented F5H. These data indicated that the structure of the genomic DNA corresponding to the F5H cDNA was disrupted in two of the *fah1* alleles, and expression of the corresponding mRNA was suppressed in five of nine *fah1* alleles. These data provide strong correlative evidence that the cDNA encodes F5H, data which is further strengthened by the complementation of the *fah1-2* mutant with the F5H genomic clone (Fig. 3). Since Northern analysis reveals that two additional transcripts are at least partially encoded by the genomic insert of pBIC20-F5H (data not shown), mutant complementation was also carried out using a construct carrying 2719 bp of the F5H genomic sequence driven by the cauliflower mosaic virus 35S promoter fused 50 bp upstream of the inferred ATG start codon. The fact that this construct also complemented the mutant phenotype (Fig. 3) provides unequivocal proof that the F5H gene has been identified.

The deduced amino acid sequence of F5H was compared with the motifs known to be conserved in other P450s (Fig. 4). The most obvious of these conserved regions is the heme-binding domain between residues Pro-450 and Gly-460, including Cys-458, the presumed heme-binding cysteine. The F5H sequence also encodes the necessary features for targeting to the endoplasmic reticulum: a hydroxy amino acid-rich N-terminal peptide, followed by a hydrophobic domain and positively-charged stop transfer sequence. This region is also followed by the proline-rich sequence found in many other P450s (27).

Among the P450s that have been identified in plants, F5H shares more sequence homology with the flavonoid-3′,5′-hydroxylases (CYP75) than it does with C4H (CYP73) (Fig. 5). This is interesting because both C4H and F5H catalyze the hydroxylation of C6-C3 (hydroxy)cinnamic acids, while the flavonoid-3′,5′-hydroxylases use the more complex flavonoid aglycones as substrates. It is tempting to speculate that the higher degree of identity between F5H and the flavonoid-3′,5′-hydroxylases is due to the fact that both enzymes catalyze the *meta*-hydroxylation of a *para*-hydroxylated aromatic ring, while C4H introduces a hydroxyl group into the aromatic ring of cinnamic acid, a nonhydroxylated molecule, and does so *para* to the C3 side chain. On the other hand, the observed homologies may not be related to substrate specificities and may simply represent the fortuitous grouping of the few plant P450 genes that have been cloned to date. The P450 with the closest overall amino acid sequence identity to F5H is the avocado ripening-related P450 CYP71A1 (3), the endogenous function of which is unknown. The degree of identity between these two proteins is 34%, less than the cutoff value of 40% that delineates P450 families (32), thus F5H from *Arabidopsis thaliana* identifies a new P450 family that has been designated CYP84.

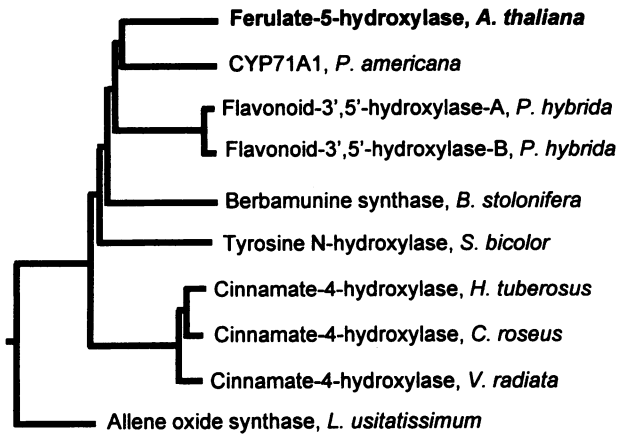

FIG. 5. Phylogenetic tree showing the relationship between the amino acid sequences of F5H, the other plant P450s of known function, and the avocado ripening-related P450 CYP71A1. The tree was constructed using the clustal analysis method of the LaserGene Megalign program. The closest homologues of F5H are the avocado ripening-related P450 (CYP71A1): percent identity 34.2% and the *Petunia hybrida* flavonoid-3',5'-hydroxylases A and B: percent identity 31.6 and 31.8% respectively.

The cloning of the structural gene for F5H is an important step in understanding the general phenylpropanoid pathway and how secondary metabolism and syringyl lignin biosynthesis are regulated in *Arabidopsis* and other higher plants. In *Arabidopsis*, sinapoyl malate accumulation appears to be restricted to the leaf upper epidermis, in keeping with a role for sinapate esters in UV attenuation (33). Similarly, syringyl lignin synthesis occurs only in the sclerified parenchyma between the vascular bundles of the *Arabidopsis* rachis (12). Similar findings have been reported in other species where syringyl lignin deposition is developmentally regulated (34). At this time, it is not clear how these biochemical pathways are regulated, but it is possible that the level of F5H expression in these tissues may regulate the flux of phenylpropanoid monomers toward sinapic acid and its subsequent metabolites. While F5H expression may determine lignin monomer composition, other enzymatic steps may also regulate this process such as the substrate specificity of cinnamyl alcohol dehydrogenase (35). Alternatively, the relative expression and substrate specificities of both caffeate/5-hydroxyferulate *O*-methyltransferase and caffeoyl CoA/5-hydroxyferuloyl CoA *O*-methyltransferase may have an important role in determining lignin monomer composition (36). The overexpression of F5H in *Arabidopsis* will allow us to test the hypothesis that this enzyme catalyzes a rate-limiting step in the biosynthesis of syringyl lignin. A related question concerns why gymnosperms, in general, lack syringyl lignin. Theoretically, this could be due to the lack of a functional F5H gene or due to the substrate specificities of other enzymes in the pathway. If the lack of a functional F5H gene is responsible for determining the lignin monomer composition of gymnosperms, the transformation of economically important species such as loblolly pine with the *Arabidopsis* F5H gene may permit modification of lignin composition for the benefit of the pulp and paper industry (37). Expression of F5H in a heterologous expression system will enable the examination of its substrate specificity. The determination of its relative activity toward ferulic acid and feruloyl coenzyme A will be crucial in elucidating the actual biochemical pathway used by plants to channel phenylpropanoid monomers toward the accumulation of sinapic acid-derived secondary metabolites and syringyl lignin.

We thank Drs. Pablo Scolnik and Tim Caspar for supplying the T-DNA-tagged *fah1-9* mutant. We also thank Dr. David Nelson for assigning the appropriate designation to CYP84, Christopher Brown for excellent technical assistance, and Ruth Wilson for the generous donation of M2 seed stock. This is journal paper number 14605 of the Purdue University Agricultural Experiment Station. This work was supported by grants from the Division of Energy Biosciences, U.S. Department of Energy DE-FG02-94ER20138 to C.C.S.C. and DE-FG02-94ER20133 to C.S., and a Swiss National Science Foundation postdoctoral fellowship to K.M.

1. Hahlbrock, K. & Scheel, D. (1989) *Annu. Rev. Plant Physiol.* **40,** 347–369.
2. Lewis, N. G. & Yamamoto, E. (1990) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **41,** 455–496.
3. Bozak, K. R., Yu, H., Sirevåg, R. & Christoffersen, R. E. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 3904–3908.
4. Mizutani, M., Ward, E., DiMaio, J., Ohta, D., Ryals, J. & Sato, R. (1993) *Biochem. Biophys. Res. Commun.* **190,** 875–880.
5. Teutsch, H. G., Hasenfratz, M. P., Lesot, A., Stoltz, C., Garnier, J.-M., Jeltsch, J.-M., Durst, F. & Werck-Reichhart, D. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 4102–4106.
6. Song, W.-C., Funk, C. D. & Brash, A. R. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 8519–8523.
7. Kraus, P. F. X. & Kutchan, T. M. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 2071–2075.
8. Pan, Z., Durst, F., Werck-Reichhart, D., Gardner, H. W., Camara, B., Cornish, K. & Backhaus, R. A. (1995) *J. Biol. Chem.* **270,** 8487–8494.
9. Koch, B. M., Sibbesen, O., Halkier, B. A., Svendsen, I. & Møller, B. L. (1995) *Arch. Biochem. Biophys.* **323,** 177–186.
10. Durst, F. & O'Keefe, D. P. (1995) *Drug Metab. Drug Interact.* **12,** 171–187.
11. Grand, C. (1984) *FEBS Lett.* **169,** 7–11.
12. Chapple, C. C. S., Vogt, T., Ellis, B. E. & Somerville, C. R. (1992) *Plant Cell* **4,** 1413–1424.
13. Feldmann, K. A. & Marks, M. D. (1987) *Mol. Gen. Genet.* **208,** 1–9.
14. Rogers, S. O. & Bendich, A. J. (1985) *Plant Mol. Biol.* **5,** 69–76.
15. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).
16. Goldsbrough, P. B. & Cullis, C. A. (1981) *Nucleic Acids Res.* **9,** 1301–1309.
17. Behringer, F. J. & Medford, J. I. (1992) *Plant Mol. Biol. Rep.* **10,** 190–198.
18. Newman, T., de Bruijn, F. J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomashow, M., Retzel, E. & Somerville, C. (1994) *Plant Physiol.* **106,** 1241–1255.
19. Meyer K., Benning, G. & Grill, E. (1996) *Genome Mapping in Plants*, ed. Paterson, A. H. (Landes Biomedical Press, Austin, TX), in press.
20. Zambrisky, P., Joos, H., Genetello, C., Leemans, J., van Montagu, M. & Schell, J. (1983) *EMBO J.* **2,** 2143–2150.
21. Meyer K., Leube, M. & Grill, E. (1994) *Science* **264,** 1452–1455.
22. Bent, A. F., Kunkel, B. N., Dahlbeck, D., Brown K. L., Schmidt, R., Giraudat, J., Leung, J. & Staskawicz, B. J. (1994) *Science* **265,** 1856–1860.
23. Odell, J. T., Nagy, F. & Chua, N-H. (1985) *Nature (London)* **313,** 810–812.
24. An, G. (1987) *Methods Enzymol.* **153,** 292–305.
25. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
26. Lütcke, H. A., Chow, K. C., Mickel, F. S., Moss, K. A., Kern, H. F. & Scheele, G. A. (1987) *EMBO J.* **6,** 43–48.
27. Yamazaki, S., Sato, K., Suhara, K., Sakaguchi, M., Mihara, K. & Omura, T. (1993) *J. Biochem. (Tokyo)* **114,** 652–657.
28. Holton, T. A., Brugliera, F., Lester, D. R., Tanaka, Y., Hyland, C. D., Menting, J. G. T., Lu, C.-Y., Farcy, E., Stevenson, T. W. & Cornish, E. (1993) *Nature (London)* **366,** 276–279.
29. Winkler, R. G. & Helentjaris, T. (1995) *Plant Cell* **7,** 1307–1317.
30. Szekeres, M., Németh, K., Koncz-Kálmán, Z., Mathur, J., Kauschmann, A., Altmann, T., Rédei, G. P., Nagy, F., Schell, J. & Koncz, C. (1996) *Cell* **85,** 171–182.
31. Feldmann, K. A., Malmberg, R. L. & Dean, C. (1994) *Arabidopsis*, eds. Meyerowitz, E. M. & Somerville, C. R. (Cold Spring Harbor Press, Cold Spring Harbor, NY), pp 137–172.

32. Nelson, D. R., Kamataki, T., Waxman, D. J., Guengerich, F. P., Estabrook, R. W., Feyereisen, R., Gonzalez, F. J., Coon, M. J., Gunsalus, I. C., Gotoh, O., Okuda, K. & Nebert, D. W. (1993) *DNA Cell Biol.* **12,** 1–51.
33. Landry, L. G., Chapple, C. C. S. & Last, R. L. (1995) *Plant Physiol.* **109,** 1159–1166.
34. Venverloo, C. J. (1971) *Holzforschung* **25,** 18–24.
35. Whetten, R. & Sederoff, R. (1995) *Plant Cell* **7,** 1001–1013.
36. Ye, A.-H., Kneusel, R. E., Matern, U. & Varner, J. E. (1994) *Plant Cell* **6,** 1427–1439.
37. Whetten, R. & Sederoff, R. (1991) *Forest Ecol. Management* **43,** 301–316.