Molecular cloning of a cytochrome P450 taxane 10β-hydroxylase cDNA from *Taxus* and functional expression in yeast

Anne Schoendorf*, Christopher D. Rithner†, Robert M. Williams†, and Rodney B. Croteau*‡

*Institute of Biological Chemistry, Washington State University, Pullman, WA 99164-6340; and †Department of Chemistry, Colorado State University, Fort Collins, CO 80523

Contributed by Rodney B. Croteau, December 4, 2000

The early steps in the biosynthesis of Taxol involve the cyclization of geranylgeranyl diphosphate to taxa-4(5),11(12)-diene followed by cytochrome P450-mediated hydroxylation at C5, acetylation of this intermediate, and a second cytochrome P450-dependent hydroxylation at C10 to yield taxadien-5α-acetoxy-10β-ol. Subsequent steps of the pathway involve additional cytochrome P450 catalyzed oxygenations and CoA-dependent acylations. The limited feasibility of reverse genetic cloning of cytochrome P450 oxygenases led to the use of *Taxus* cell cultures induced for Taxol production and the development of an approach based on differential display of mRNA-reverse transcription-PCR, which ultimately provided full-length forms of 13 unique but closely related cytochrome P450 sequences. Functional expression of these enzymes in yeast was monitored by *in situ* spectrophotometry coupled to *in vivo* screening of oxygenase activity by feeding taxoid substrates. This strategy yielded a family of taxoid-metabolizing enzymes and revealed the taxane 10β-hydroxylase as a 1494-bp cDNA that encodes a 498-residue cytochrome P450 capable of transforming taxadienyl acetate to the 10β-hydroxy derivative; the identity of this latter pathway intermediate was confirmed by chromatographic and spectrometric means. The 10β-hydroxylase represents the initial cytochrome P450 gene of Taxol biosynthesis to be isolated by an approach that should provide access to the remaining oxygenases of the pathway.

Taxol biosynthesis | paclitaxel | taxadien-5α-yl acetate | taxadien-5α-acetoxy-10β-ol | *Taxus cuspidata*

The complex diterpenoid Taxol[§] (paclitaxel) (1) is a potent antimitotic agent with excellent activity against a range of cancers (2). The only source of the drug, and of its precursors for semisynthetic preparation (3–5), is yew (*Taxus*) species (6). The supply of Taxol has been limited since the discovery of this natural product (6), and, with increasing applications in chemotherapy, the availability and cost of the drug will remain important issues (7). For the foreseeable future, the supply of Taxol and its synthetically useful precursors must continue to rely on biological methods of production, either in *Taxus* species or, potentially, by cell cultures derived from these plants (7). Improving the biological production yields of the drug depends on a detailed understanding of the biosynthetic pathway, the enzymes catalyzing the sequence of reactions (especially the slow steps), and the genes encoding these enzymes.

The biosynthesis of Taxol (Fig. 1) involves the initial cyclization of geranylgeranyl diphosphate, the universal precursor of diterpenoids, to taxa-4(5),11(12)-diene (8), followed by extensive oxidative modification of this parent olefin and the elaboration of side-chains (9–11), although the precise order of intermediates of the pathway is largely unknown. The dearth of naturally occurring, lightly functionalized taxoids (i.e., taxoids bearing up to three oxygen functional groups) (12) provides no assistance in deciphering the early oxygenation steps of the pathway. However, consideration of more functionalized metabolites, in particular the relative abundances of taxoids bearing oxygen functions at different positions of the taxane core, suggests that the order of oxygenation is at C5 then C10, followed by oxygenation at C2 and C9, then at C13 (9, 13). Oxygenations at C7 and C1 of the taxane nucleus are considered to be very late introductions.

A combination of *in vivo* feeding studies and investigations with cell-free enzyme systems, using yew stem tissue or suspension cultured cells as source material, has recently indicated (14–16) that the early steps of the Taxol biosynthetic pathway proceed in sequence from the parent olefin by cytochrome P450-mediated hydroxylation (with allylic rearrangement) of taxa-4(5),11(12)-diene to taxa-4(20),11(12)-dien-5α-ol, followed by acetylation, and then another cytochrome P450-mediated hydroxylation at C10 to yield taxa-4(20),11(12)-dien-5α-acetoxy-10β-ol (Fig. 1). Although the reaction products are not yet fully defined, subsequent microsomal oxygenations of the taxane nucleus also appear to be catalyzed by distinct cytochrome P450 species (10, 16). Genes encoding several of these Taxol biosynthetic enzymes have been isolated, including those for geranylgeranyl diphosphate synthase (17), taxadiene synthase (18) and several acyltransferases (19, 20), such as that responsible for acetylation of taxadienol (21); however, no cDNAs encoding relevant cytochrome P450 oxygenases of the pathway have been described as yet.

Purification of individual microsomal enzymes for microsequencing, as the basis for cDNA cloning, can be arduous, particularly for the Taxol pathway in which many cytochrome P450 oxygenases participate (16), because these enzymes would be expected to be very similar in physical properties (22) and thus difficult to separate from each other. As an alternative to reverse genetic cloning of the cytochrome P450 oxygenases involved in Taxol biosynthesis, an approach was taken based on differential display of mRNA-reverse transcription-PCR (DD-RT-PCR), using cultured *Taxus* cells induced for Taxol production as the source and with sorting of clones by expression in yeast to define function. This strategy has yielded a family of taxoid oxygenases, including a cDNA encoding the cytochrome P450 taxane 10β-hydroxylase that is described in this paper.

Materials and Methods

Plant Materials, Substrate and Reagents. Initiation, propagation, and induction (with methyl jasmonate) of *Taxus* sp. cell cultures

Abbreviations: DD-RT-PCR, differential display of mRNA-reverse transcription-PCR; RACE, rapid amplification of cDNA ends; NOE, nuclear Overhauser effect; HSQC, heteronuclear single quantum coherence; 2D-TOCSY, two-dimensional homonuclear total correlation spectroscopy; 2D-ROESY, two-dimensional rotating frame Overhauser enhancement spectroscopy.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. AF318211).

‡To whom reprint requests should be addressed. E-mail: croteau@mail.wsu.edu.

§Paclitaxel is the generic name for Taxol, which is now a registered trademark of Bristol-Myers Squibb Company. Because of the greater familiarity of the word Taxol, we use it in this paper.

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.

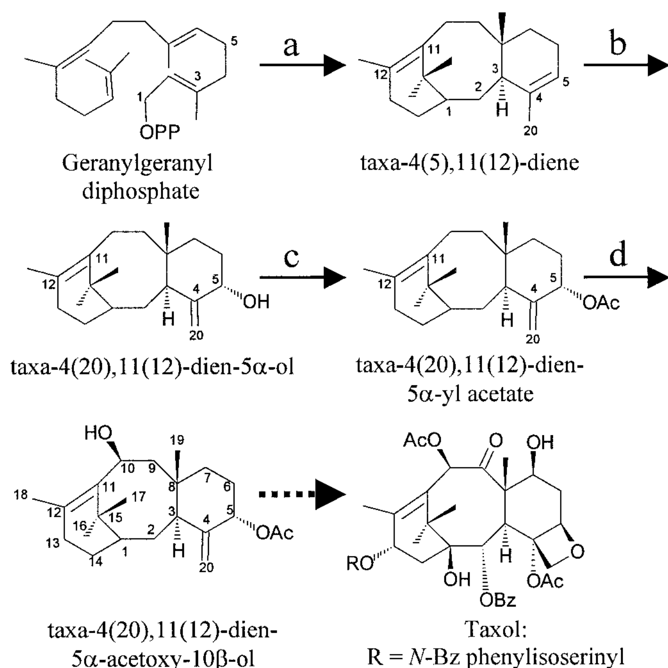

Fig. 1. Outline of early steps of the Taxol biosynthetic pathway illustrating cyclization of geranylgeranyl diphosphate to taxadiene by taxadiene synthase (*a*), hydroxylation and rearrangement of the parent olefin to taxadien-5α-ol by taxadiene 5α-hydroxylase (*b*), acetylation by taxadienol-*O*-acetyltransferase (*c*), and hydroxylation to taxadien-5α-acetoxy-10β-ol by the taxane 10β-hydroxylase (*d*). The broken arrow indicates several as yet undefined steps.

have been previously described (17, 23). Enzymes and reagents were obtained from United States Biochemical, GIBCO/BRL, Promega, and New England BioLabs and used according to the manufacturers' instructions. Chemicals were purchased from Sigma. The preparation of [20-^{3}H]taxa-4(20),11(12)-dien-5α-yl acetate (6 Ci/mol) has been described (16).

Vectors and DNA Manipulation. Unless otherwise stated, all routine DNA manipulations and cloning procedures were performed by standard methods (24), and PCR amplifications were performed by established protocols (25). DNA was sequenced by using Amplitaq DNA polymerase and fluorescence cycle sequencing on an Applied Biosystems Prism 373 DNA Sequencer. The *Saccharomyces cerevisiae* expression vector pYeDP60 was a generous gift from Dr. Denis Pompon (26).

***Escherichia coli* and Yeast Strains.** The *E. coli* strains XLI-Blue MRF′ (Stratagene) and TOP10F′ (Invitrogen) were used for routine cloning and cloning of PCR products, respectively. Yeast strains used for expression were gifts from Dr. Denis Pompon. These *S. cerevisiae* strains, each expressing one of two different *Arabidopsis thaliana* cytochrome P450 reductases, are designated WAT11 and WAT21 (26).

cDNA Library Construction. A cDNA library was prepared from mRNA isolated from *Taxus cuspidata* suspension cell cultures that had been induced to maximal Taxol production with methyl jasmonate for 16 h. A method for isolation of total RNA from 1.5 g *T. cuspidata* cells was empirically developed by using a buffer containing 4 M guanidine thiocyanate, 25 mM EDTA, 14 mM 2-mercaptoethanol, and 100 mM Tris·HCl, pH 7.5. Cells were homogenized on ice by using a polytron (4 × 15-sec bursts at setting 7). The homogenate was then adjusted to 2% (vol/vol) Triton X-100 and allowed to stand 15 min on ice, after which an equal volume of 3 M sodium acetate, pH 6.0, was added. The mixed solution was incubated on ice for 15 min, followed by centrifugation at 15,000 × *g* for 30 min at 4°C. The supernatant was then mixed with 0.8 vol of isopropanol and left to stand on ice for 5 min. After centrifugation at 15,000 × *g* for 30 min at 4°C, the resulting pellet was redissolved in 8 ml of 20 mM Tris·HCl, pH 8.0, containing 1 mM EDTA, then adjusted to pH 7.0 by addition of 2 ml of 2 M NaCl in 250 mM Mops buffer at pH 7.0. Total RNA was recovered by passing this solution over a nucleic acid isolation column (Qiagen, Valencia, CA) following the manufacturer's instructions. Poly(A)$^+$ RNA was purified by using the Oligotex mRNA kit following the manufacturer's instructions (Qiagen). Messenger RNA prepared in this fashion was used to construct a library by using a λZAPII-cDNA synthesis kit and ZAP-cDNA gigapack III gold packaging kit (Stratagene) following the manufacturer's instructions. The isolated mRNA was also used to construct a RACE (rapid amplification of cDNA ends) library by using a Marathon cDNA amplification kit (CLONTECH, Palo Alto, CA).

Differential Display of mRNA. mRNA differential display was performed by using the Delta Differential Display Kit (CLONTECH) by following the manufacturer's instructions except where noted. mRNA, for reverse transcription to template for PCR amplification, was isolated as described above from two different *T. cuspidata* suspension cell cultures, one that had been induced with methyl jasmonate 16 h before RNA isolation and the other that had not been treated (i.e., uninduced). Cytochrome P450-specific forward primers [CC(T/A/G/C) TT(C/T) GG], instead of random primers, were used in combination with reverse-anchor-(dT)$_9$N$_{-1}$N$_{-1}$ primers (where N$_{-1}$ = A, G, or C) provided in the kit. The anchor designed by CLONTECH was added to each P450-specific primer to increase the annealing temperature after the fourth low stringency PCR cycle; this strategy led to a significant reduction of the background signal. Each cytochrome P450-specific primer was used with the three anchored oligo(dT) primers terminated by each nucleotide. PCR reactions were performed with a RoboCycler 96 Temperature Cycler (Stratagene), using one cycle at 94°C for 5 min, 40°C for 5 min, and 68°C for 5 min, following by three cycles at 94°C for 30 sec, 40°C for 30 sec, and 68°C for 5 min, and then 32 cycles at 94°C for 20 sec, 60°C for 30 sec, and 68°C for 2 min; finally, the reactions were heated at 68°C for 7 min. The resulting amplicons were separated on a 6% denaturing polyacrylamide gel (HR-100, Genomyx, Foster City, CA) by using the LR DNA Sequencer Electrophoresis System (Genomyx).

Differential display bands of interest were cut from the dried gel and eluted with 100 μl of 10 mM Tris·HCl buffer, pH 8.0, containing 1 mM EDTA, by incubation overnight at 4°C. A 5-μl aliquot of the extract was used to reamplify the cDNA fragment by PCR using the same primers as in the original amplification. The reactions were initially heated to 94°C for 2 min, then underwent 30 cycles at 94°C for 1 min, 60°C for 1 min, and 68°C for 2 min. Finally, to facilitate cloning of the PCR product, the reactions were heated at 68°C for 7 min. Amplicons were analyzed by agarose gel electrophoresis, bands were excised from the gel, and the DNA was extracted from the agarose. This gel-purified DNA was then ligated into the T/A cloning vector pCR2.1-TOPO (Invitrogen).

The DD-RT-PCR-based screening revealed about 100 clearly differentially expressed bands, all of which were sequenced and analyzed. Of these, 39 represented PCR products containing cytochrome P450-like sequences. The nucleotide and deduced peptide sequences of these 39 amplicons were compiled by using the GCG fragment assembly programs and the sequence alignment program PILEUP (27). This comparison of cloned sequences revealed that C-terminal fragments from 21 different cyto-

chrome P450 genes had been isolated. These cytochrome P450 sequences were used to prepare hybridization probes to isolate the corresponding full-length clones by screening the cDNA library.

cDNA Library Screening. After initial selection by sequence comparison, 12 probes were randomly labeled by using the Ready-To-Go kit (Amersham Pharmacia) following the manufacturer's instructions. Plaque lifts of the *T. cuspidata* phage library were made on nylon membranes and were screened by using mixtures of two radiolabeled probes each. Phage DNA was cross-linked to the nylon membranes by autoclaving on fast exhaust cycle for 3 min at 120°C. After cooling, the membranes were washed for 5 min in 2× SSC. Prehybridization was performed for 1 to 2 h at 65°C in 6× SSC, containing 0.5% SDS, and 5× Denhardt's reagent. Hybridization was preformed in the same buffer for 20 h at 65°C. The nylon membranes were then washed twice for 5 min each in 2× SSC with 0.1% SDS at room temperature, and twice for 1 h each in 1× SSC with 0.1% SDS at 65°C. After washing, the membranes were exposed for 17 h to Kodak (Rochester, NY) XAR film at −70°C. Positive plaques were purified through one additional round of hybridization. Purified λZAPII clones were *in vivo* excised as pBluescript II SK(−) phagemids and transformed into *E. coli* SOLR cells (Stratagene). The size of each cDNA insert was determined by PCR using T3 and T7 promoter primers. Inserts (>1.6 kb, of a size necessary to encode a typical cytochrome P450 of 50–60 kDa) were sequenced and sorted into groups based on sequence similarity/identity by using the GCG fragment assembly programs (27). Each unique sequence was used as a query in database searching by using either BLAST or FASTA programs (27) to define sequences with significant homology to plant cytochrome P450 sequences. These clones were also compared at both the nucleic acid and amino acid levels by using the PILEUP and GAP programs (27).

Generation of Full-Length Clones by 5′-RACE. Of 18 clones thus far examined, full-length sequences of nine were obtained by screening of the *T. cuspidata* λ-phage library with the corresponding probe mixtures. To obtain the 5′-terminal sequence of the remaining nine truncated clones, 5′-RACE was performed by using the Marathon cDNA amplification kit (CLONTECH) according to the manufacturer's instructions with sequence-specific reverse primers. Using the 5′ sequences thus acquired, and the previously obtained 3′-sequence information, primers corresponding to these terminal regions were designed, and the full-length versions of each clone were obtained by amplification with *Pfu* polymerase (Stratagene) using library cDNA as target. These primers were also designed to contain nucleotide sequences encoding restriction sites that were used to facilitate cloning into the yeast expression vector.

cDNA Expression of Cytochrome P450 Enzymes in Yeast. Appropriate restriction sites were introduced by standard PCR methods (25) immediately upstream of the ATG start codon and downstream of the stop codon of all full-length cytochrome P450 clones. These modified amplicons were gel purified, digested with the corresponding restriction enzymes, and then ligated into the expression vector pYeDP60. The sequence-verified clones were then transformed into yeast by using the lithium acetate method (28). Selected transformants were grown to stationary phase in SGI medium, then used as inoculum for a large-scale expression culture grown in YPL medium (26). Approximately 24 h after induction of cytochrome P450 expression with galactose (to 10% final concentration), a portion of the yeast cell culture was harvested by centrifugation. One-half was treated with carbon monoxide, and the cytochrome P450 CO-difference spectrum (29) was recorded directly (untreated cells as control) to assess the presence of functional, recombinant cytochrome P450 and to estimate the quantity of competent enzyme.

Assay of Recombinant Cytochrome P450 Oxygenases. Isolated transformants for each full-length cytochrome P450 clone shown to express a functional enzyme by CO-difference spectrum (eight clones) were grown to stationary phase in 2 ml of SGI medium at 30°C and used to inoculate a 10-ml expression culture (in YPL medium). Approximately 8 h after induction, cells were harvested by centrifugation (10 min at 1500 rpm), and the pellet was then resuspended in 2 ml of fresh YPL medium.

To eliminate the additional complication and uncertainty associated with microsome isolation for *in vitro* assay, [20-^{3}H]taxa-4(20),11(12)-dien-5α-yl acetate (10^6 dpm, 40 μM) was added directly to the cell suspension to assay conversion *in vivo*. After 12 h of incubation at 30°C with agitation (250 rpm), the mixture was treated for 15 min in a sonication bath and extracted three times with 2 ml of diethyl ether to ensure isolation of the biosynthetic product(s). These ether extracts, containing residual substrate and derived product(s), were concentrated to dryness, resuspended in 200 μl of CH_3CN, and filtered. These samples were then analyzed by reversed phase radio-HPLC using a previously established method (14) capable of separating taxoids ranging in polarity from taxadiene to taxadien-hexaol. To identify the biosynthetic product, the material isolated by HPLC (>97% purity by GC) was subjected to GC-MS and NMR analysis using established procedures for taxoid structural elucidation (14, 16, 30) (see also *Experimental Methods*, which are published as supplemental data on the PNAS web site, www.pnas.org).

Results and Discussion

Cloning of Cytochrome P450 Hydroxylases from *Taxus*. Biochemical studies have indicated that the second and fourth specific steps of Taxol biosynthesis diverging from primary metabolism (i.e., from geranylgeranyl diphosphate) involve the cytochrome P450-mediated hydroxylation of taxadiene to taxadien-5α-ol (14) and the cytochrome P450-dependent hydroxylation of taxadien-5α-yl acetate to the 5α,10β-taxadien-diol monoacetate (16) (Fig. 1). The next three oxygenation steps of the Taxol pathway also appear to be catalyzed by cytochrome P450 hydroxylases, and the remaining three oxygenations are likely catalyzed by cytochrome P450 enzymes as well (10, 14). The early oxygenations are slow steps of the reaction pathway and, thus, important candidates for cDNA isolation for the purpose of overexpression in relevant producing organisms to increase Taxol yields (11, 31). Protein purification of cytochrome P450 enzymes from *Taxus* microsomes (32) as the basis for cDNA cloning was eliminated as not feasible because the number of P450 species involved, and their predicted similarity in physical properties (22), would almost certainly prevent bringing the individual proteins to homogeneity for amino acid microsequencing.

As an alternative approach to cytochrome P450 cloning by reverse genetics, a strategy was used that was based on DD-RT-PCR for the isolation of transcriptionally active cytochrome P450s in *Taxus* cells, which previous biochemical studies had shown to undergo substantial up-regulation of the Taxol pathway 16 h after induction with methyl jasmonate (17, 23). Because an increase in the relevant enzyme activities was shown to result from induction (indicating *de novo* protein synthesis), mRNA from an untreated cell culture was compared with that of a culture that had been so induced for 16 h. To obtain predominantly induced cytochrome P450 sequences, forward primers were designed based on a highly conserved motif in plant cytochrome P450 genes. Related strategies have been used with other plants (33). We chose for this purpose the [PFG] motif, which is a very well-conserved region of the heme-binding domain (34). The corresponding codons of this region contain

BIOCHEMISTRY

only two degenerate positions; thus, a set of only eight nondegenerate primers was necessary to encompass all sequence possibilities. This [PFG] motif is located 200–250 bp upstream of the stop codon, and the length of the 3′-untranslated region should range between 100 and 300 bp. Thus, the length of the expected PCR fragments would be in the 300- to 550-bp range. This DD-RT-PCR-based strategy revealed roughly 100 differentially expressed species, and the sequences of all were obtained and analyzed. Of these, 39 represented PCR products containing a cytochrome P450-type sequence. Analysis of these sequences revealed that the 3′-terminus from 21 unique cytochrome P450 genes had been isolated. Based on sequence similarity (see below), 12 of these DNA fragments were selected and used as labeled hybridization probes to screen the methyl jasmonate-induced *T. cuspidata* cell cDNA library. By this means, nine clones were obtained in full-length form. Nine additional clones, which were truncated at the 5′-terminus, were obtained in full-length form by using a 5′-RACE method to acquire the missing 5′ sequences.

Sequence Analysis. The full-length cytochrome P450 sequences obtained (using the initial 12 probes) were compared, and it was shown that a total of 18 unique sequences (showing <85% similarity) had been defined. Two of these acquisitions were not identical to any of the 21 3′-terminal fragments originally found by the DD-RT-PCR cloning strategy, bringing the total number of unique cytochrome P450 genes, and gene fragments, to 23. The clones obtained were also compared with all known plant cytochrome P450 sequences in the databases to provide a dendrogram of these relationships (data not shown), which revealed that 13 of the *Taxus* clones sorted into one cytochrome P450 family. This large group of related clones seems to most closely resemble the CYP90, CYP85, and CYP88 cytochrome P450 families, some members of which are known to be involved in terpenoid metabolism [e.g., gibberellin (diterpene, C_{20}) and brassinosteroid (triterpene, C_{30}) biosynthesis; refs. 35–37], suggesting that the cytochrome P450 clones obtained from *Taxus* could be involved in the biosynthesis of the diterpenoid Taxol.

Demonstration of Functional Cytochrome P450 Expression in Yeast. Functional cytochrome P450 expression can be obtained by using the pYeDP60 plasmid in yeast (*S. cerevisiae*) engineered to coexpress one of the two cytochrome P450 reductases from *A. thaliana*; the plant-derived reductase is important for efficient electron transfer to the cytochrome (26). Because a functional P450 cytochrome, in the appropriately reduced form, will competently bind to carbon monoxide and give a characteristic CO-difference spectrum with absorbance maximum near 450 nm (29), we developed this spectrophotometric means for assessing, and quantitatively estimating, the presence of functional heme-containing and properly folded recombinant cytochrome P450 in transformed yeast cells by *in situ* (*in vivo*) measurement. Of the 13 related, full-length cytochrome P450 clones from *Taxus* thus far obtained, eight yield detectable CO-difference spectra when expressed in *S. cerevisiae* and measured by this *in situ* assay (Fig. 2).

Assay of Transformed Yeast Cells with Taxadienyl Acetate. Transformed yeast cells that functionally expressed a recombinant cytochrome P450 gene from *Taxus* (by CO-difference spectrum) were tested *in vivo* for their ability to oxygenate (hydroxylate or epoxidize) taxoid substrates fed exogenously, thereby eliminating the need for microsome isolation for such preliminary *in vitro* assay. Accordingly, all 8 functional clones of the relevant 13 full-length forms expressed in the induced yeast host were fed [20-$^3H_2$]taxa-4(20),11(12)-dien-5α-yl acetate in separate incubations and compared with untransformed controls similarly fed (and that were shown to be inactive with taxoid substrates). The

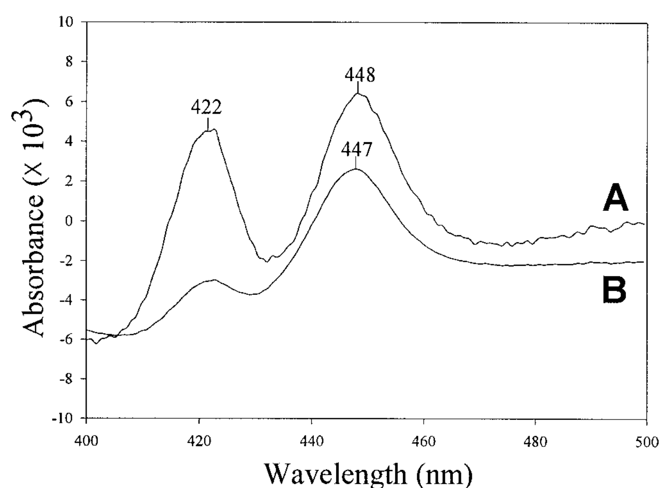

Fig. 2. Reduced CO-difference spectrum of intact yeast cells harboring cDNA clone F14 (*A*). The smooth lower spectrum (*B*) illustrates the CO-difference spectrum of the corresponding isolated microsomes to demonstrate that the absorbance at 422 nm in (*A*) was due to an endogenous soluble yeast pigment. The microsomal preparation was catalytically functional in the NADPH and oxygen-dependent taxane 10β-hydroxylase reaction.

ether extracts resulting from these incubations were analyzed by radio-HPLC, and one of these (derived from the clone designated F14) revealed the nearly complete conversion of the taxadien-5α-yl acetate substrate to a more polar product (Fig. 3). The metabolite isolated by HPLC was subjected to GC-MS analysis and shown to possess a retention time (compared with the starting material) and mass spectrum consistent with that of a taxadien-diol monoacetate {the parent ion [P^+] was observed at m/z 346 [taxadienyl acetate (M_r = 330) plus *O*] with diagnostic ions at m/z 328 [P^+-H_2O], 313 [P^+-H_2O-CH_3], 286 [P^+-CH_3COOH], 271 [P^+-CH_3COOH-CH_3], 268 [P^+-CH_3COOH-H_2O], and 253 [P^+-CH_3COOH-CH_3-H_2O]}.

Preparative scale incubations of the transformed yeast harboring clone F14, with the taxadien-5α-yl acetate substrate, allowed the HPLC-based isolation of about 100 μg of the unknown diol monoacetate (>97% purity by GC) for NMR

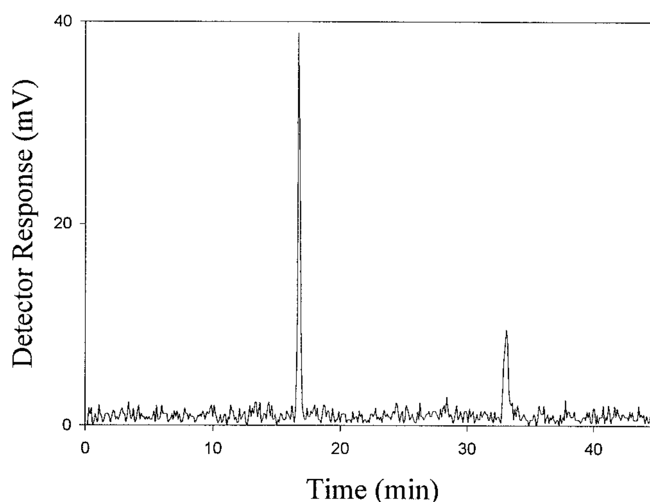

Fig. 3. Reversed phase radio-HPLC analysis of the biosynthetic product (R_t = 16.9 min) generated from exogenous taxadien-5α-yl acetate (R_t = 32.9 min) administered to induced yeast cells harboring cDNA clone F14. The product purified by this means was identified by GC-MS and NMR methods as taxa-4(20),11(12)-dien-5α-acetoxy-10β-ol.

Table 1. Complete ^{1}H-NMR assignments and one-bond correlated ^{13}C assignments (as measured indirectly from HSQC) for the biosynthetic product derived from taxadien-5α-yl acetate by the cytochrome P450 expressed from clone F14

Position number	Carbon δ, ppm	α-Proton δ, ppm	β-Proton δ, ppm
1	43.9		1.59
2	28	1.47	1.53
3	35.9	2.84	
4			
5	75.9		5.47
6	27.9	1.66	1.55
7	33.6	1.84	0.9
8			
9	47.6	1.42	2.21
10	67.2	4.9	
11			
12			
13	30.3	1.8	2.26
14	22.7	1.26	1.96
15			
16	31.8	1.14 (*exo*)	
17	25.3	1.59 (*endo*)	
18	20.7	1.71	
19	21.4		0.56
20	111.6	5.07 (*exo*) 4.67 (*endo*)	
21(acetate)	21	1.66	

For position numbering, see Fig. 1.

analysis. Because all of the ^{1}H resonances of taxadien-4(20),11(12)-dien-5α-ol (and of the acetate ester) had been previously assigned (14) (see also supplementary data), structure elucidation of the unknown diol monoacetate was accomplished by ^{1}H detection experiments (sample size limited direct ^{13}C measurements). The ^{1}H-NMR spectrum is illustrated in supplementary Fig. 4, and Table 1 lists the complete ^{1}H assignments along with their one-carbon correlated ^{13}C assignments as determined indirectly from heteronuclear single quantum coherence (HSQC) (supplementary Fig. 5). The assignments were consistent with those of other known taxadien-monool and -diol derivatives. For example, chemical shifts (in ppm) for C5 (δ 75.9, C5; δ 5.47, H5) and C10 (δ 67.2, C10; δ 4.9 H10) were assigned as oxy-methines. The shifts for C20 (δ 111.6, C20; δ 5.07, H20*exo*; δ 4.67, H20*endo*) were consistent with the exocyclic methylene observed in other taxa-4(20),11(12)-dienes. Other characteristic shifts were observed for H7α (δ 1.84), H19 methyl (δ 0.56), H3 (δ 2.84), and the *gem*-dimethyls H16 (δ 1.14, *exo*), and H17 (δ 1.59, *endo*).

The two-dimensional homonuclear total correlation spectroscopy (2D-TOCSY) spectra (supplementary Figs. 6 and 7) complemented the HSQC data and permitted additional regiochemical assignments. Proton H5 (δ 5.47) (supplementary Fig. 7 *A* and *E*) was correlated strongly with H6 (δ 1.66, δ 1.55) and H7 (δ 1.94, δ 0.9) protons but had no appreciable coupling to either of the H20 signals (δ 5.07, δ 4.67) or to H3 (δ 2.84), which is a common feature observed with taxadiene derivatives. The spin system defined in part by H3 (δ 2.84), H2 (δ 1.47, δ 1.53), H1 (δ 1.59), H13 (δ 1.80, δ 2.26) and H14 (δ 1.26, δ 1.96) was apparent in supplementary Fig. 7 *C* and *E*. The allylic methyl H18 (δ 1.71) also displayed a weak correlation with H13. In contrast to the extended spin correlations noted in supplementary Fig. 7*D*, the H9 (δ 1.42, δ 2.21) and H10 (δ 4.9) signals formed an isolated spin system (see supplementary Fig. 7*B*), which included the C10 hydroxyl proton (δ 0.85). A correlation was also observed between the two *gem*-dimethyl proton signals (δ 1.14 and δ 1.59), which was consistent with the spectra of other taxadiene derivatives.

Two-dimensional rotating frame Overhauser enhancement spectroscopy (2D-ROESY) spectra (supplementary Figs. 6*B* and 8) were used to confirm the regiochemical assignments and to assess relative stereochemistry [several of these nuclear Overhauser effect (NOE) correlations are listed in supplementary Table 2]. The 2D-TOCSY showed that a second heteroatom was introduced into the C9-C10 fragment, but the regiochemistry was ambiguous based on this single measurement. The 2D-ROESY confirmed that oxidation had occurred at C10 and placed the C10 hydroxyl in the β-orientation. This assignment was also supported by an observed NOE between the H10 proton (δ 4.90; supplementary Fig. 8*B*) and the allylic methyl H18 (δ 1.71), which is consistent with an α-configuration for H10. Additional stereochemical assignments were made by noting correlations between H9β (δ 2.21) and H17 (δ 1.59) located on the *endo* C17-methyl (supplementary Fig. 8*E*), H19 (δ 0.56) on the β-oriented C19-methyl, and H2β (δ 1.53). The other H9 signal (δ 1.42) correlated with H19 and H7β (δ 0.90), as well as H10 (δ 4.90) (supplementary Fig. 8 *D* and *B*). It was also noted that $^3J_{HH}$ was large (11.7 Hz) between H9β and H10α, consistent with a nearly anti-arrangement for this pair; a smaller coupling (5.3 Hz) between H9α and H10 was consistent with a *gauche* relationship between these two protons.

ROESY spectroscopy was also used to confirm the stereochemistry at H5. Moderately strong correlations were seen between H5 (δ 5.47) (see supplementary Table 2 and Fig. 8*A*) and both C6 signals (δ 1.66, δ 1.55), consistent with an equatorial orientation for H5. The $^3J_{HH}$ coupling was quite small (<3 Hz), indicating a *gauche* relationship between H5 and all other scalar coupled partners, and was further evidence for the adopted equatorial orientation of H5. A moderately strong NOE between H5 and H20*exo* was noted, but there were *no* NOE correlations observed between H5 and other protons on the α-face of the molecule, thereby confirming that H5 was β configured and that the acetate group was α oriented, as in the substrate. One other significant structural motif in taxadiene derivatives is the near occlusion of the H3 proton on the α-face because of the unusual folding of the molecule, thereby making the H3 proton (δ 2.84) a useful probe for this face. Indeed, NOE correlations were observed between H3, H10, H13α, and the allylic methyl H18 (supplementary Table 2 and Fig. 8*C*).

The full assignment of the structure confirms the identity of the biosynthetic product as taxa-4(20),11(12)-dien-5α-acetoxy-10β-ol, and indicates that a cDNA encoding the cytochrome P450 taxane 10β-hydroxylase had been isolated. This 1494-bp cDNA (GenBank accession AF318211) translates a 498-residue deduced protein of molecular weight 56,690 that bears a typical N-terminal membrane anchor (38), with hydrophobic insertion segment (39) and stop-transfer signal (40), and that possesses all of the conserved motifs anticipated for cytochrome P450 oxygenases, including the oxygen-binding domain (41) and highly conserved heme-binding motif (34, 42) with PFG element (amino acids 435–437).

Conclusions

A DD-RT-PCR-based cloning strategy using mRNA isolated from uninduced *Taxus* cells and *Taxus* cells induced for Taxol biosynthesis led to the acquisition of 21 unique cytochrome P450 gene fragments. The use of 12 of these 3′-fragments as hybridization probes for library screening, in combination with 5′-RACE for truncated forms, ultimately yielded 18 distinct full-length clones, 13 of which were shown to be closely related by sequence comparison (60–85% similarity; 52–83% identity) and to resemble CYP families 85, 88, and 90, which are known to contain cytochrome P450 members involved in the biosynthesis of other terpenoids (35–37). Of the 13 closely related clones,

eight were demonstrated to be functionally expressed in yeast by yielding a CO-difference spectrum, and four of these were confirmed to metabolize various taxoid substrates, including the taxane 10β-hydroxylase. It is notable that outlier clones, which did not resemble the other sequences (five in total) but which were functionally expressed, did not yield cytochrome P450 gene products capable of metabolizing taxoid substrates.

These results suggest that the target cytochrome P450 genes responsible for the eight oxygenase steps of the Taxol pathway may reside within the group of 13 related clones now in hand. Furthermore, there are additional evolutionary and enzyme structural considerations that suggest that the cytochrome P450 oxygenases of the Taxol pathway will cluster in a closely related gene family of the type described here. Thus, the genes encoding these sequential oxygenation steps of the Taxol pathway almost certainly arose by duplication and differentiation of an ancestral gene, and so would be expected to be similar. Additionally, the unique size and shape of the taxane ring skeleton would seem to dictate the maintenance of considerable sequence similarity in the substrate binding determinants of these enzymes (and genes), in addition to the expected strong conservation of the universal catalytic elements involved in heme and oxygen binding by cytochrome P450 enzymes. The evidence, based on the cloning strategy directed to acquisition of taxane-metabolizing cytochrome P450s [which should be well represented in the induced *Taxus* cell mRNA pool and cDNA library (i.e., highly up-regulated in Taxol biosynthesis), the demonstrated catalytic capability toward taxoid substrates of several members of this related cytochrome P450 family, and the considerations of sequence similarity constrained by gene origin and enzyme function], suggests that the genes now obtained are relevant cytochrome P450 oxygenases involved in Taxol biosynthesis. Verification of function of the remaining clones, by definition of substrate selectivity and regiochemistry of the encoded oxygenases, and placement of the enzymes in the pathway sequence have a high priority.

We thank J. Hefner, R. E. B. Ketchum, K. Walker, and C. Sanchez for technical assistance, and J. Tamura for preparation of the manuscript. This investigation was supported by Grants CA-55254 and CA-70375 from the National Institutes of Health, by Cytoclonal Pharmaceutics, and by McIntire-Stennis Project 0967 from the Washington State University Agricultural Research Center.

1. Wall, M. E. & Wani, M. C. (1995) *Am. Chem. Soc. Symp. Ser.* **583,** 18–33.
2. Goldspiel, B. R. (1997) *Pharmacotherapy* **17,** 1105–1255.
3. Guénard, D., Guéritte-Voegelein, F. & Potier, P. (1993) *Acc. Chem. Res.* **26,** 160–167.
4. Holton, R. A., Biediger, R. J. & Boatman, P. D. (1995) in *Taxol*: *Science and Applications*, ed. Suffness, M. (CRC, Boca Raton), pp. 97–121.
5. Commerçon, A., Bourzat, J. D., Didier, E. & Lavelle, F. (1995) *Am. Chem. Soc. Symp. Ser.* **583,** 223–246.
6. Cragg, G. M., Schepartz, S. A., Suffness, M. & Grever, M. R. (1993) *J. Nat. Prod.* **56,** 1657–1668.
7. Suffness, M. (1995) *Am. Chem. Soc. Symp. Ser.* **583,** 1–17.
8. Koepp, A. E., Hezari, M., Zajicek, J., Stofer Vogel, B., LaFever, R. E., Lewis, N. G. & Croteau, R. (1995) *J. Biol. Chem.* **270,** 8686–8690.
9. Floss, H. G. & Mocek, U. (1995) in *Taxol*: *Science and Applications*, ed. Suffness, M. (CRC, Boca Raton), pp. 191–208.
10. Hezari, M. & Croteau, R. (1997) *Planta Med.* **63,** 291–295.
11. Walker, K. & Croteau, R. (1999) *Rec. Adv. Phytochem.* **33,** 31–50.
12. Baloglu, E. & Kingston, D. G. I. (1999) *J. Nat. Prod.* **62,** 1448–1472.
13. Croteau, R., Hezari, M., Hefner, J., Koepp, A. & Lewis, N. G. (1995) *Am. Chem. Soc. Symp. Ser.* **583,** 72–80.
14. Hefner, J., Rubenstein, S. M., Ketchum, R. E. B., Gibson, D. M., Williams, R. M. & Croteau, R. (1996) *Chem. Biol.* **3,** 479–489.
15. Walker, K., Ketchum, R. E. B., Hezari, M., Gatfield, D., Goleniowski, M., Barthol, A. & Croteau, R. (1999) *Arch. Biochem. Biophys.* **364,** 273–279.
16. Lovy Wheeler, A., Long, R. M., Ketchum, R. E. B., Rithner, C. D., Williams, R. M. & Croteau, R. (2000) *Arch. Biochem. Biophys*., in press.
17. Hefner, J., Ketchum, R. E. & Croteau, R. (1998) *Arch. Biochem. Biophys.* **360,** 62–74.
18. Wildung, M. R. & Croteau, R. (1996) *J. Biol. Chem.* **271,** 9201–9204.
19. Walker, K. & Croteau, R. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 583–587.
20. Walker, K. & Croteau, R. (2000) *Proc. Natl. Acad. Sci. USA*, **97,** 13591–13596. (First Published November 28, 2000; 10.1073/pnas.250491997)
21. Walker, K., Schoendorf, A. & Croteau, R. (2000) *Arch. Biochem. Biophys.* **374,** 371–380.
22. Mihaliak, C. A., Karp, F. & Croteau, R. (1993) *Methods Plant Biochem.* **9,** 261–277.
23. Ketchum, R. E. B., Gibson, D. M., Croteau, R. & Shuler, M. L. (1999) *Biotechnol. Bioeng.* **62,** 97–105.
24. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning*: *A Laboratory Manual*, Cold Spring Harbor Lab. Press, Plainview, NY).
25. Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J. (1990) *PCR Protocols*: *A Guide to Methods and Applications* (Academic, San Diego).
26. Pompon, D., Louerat, B., Bronine, A. & Urban, P. (1996) *Methods Enzymol.* **272,** 51–64.
27. Genetics Computer Group (1994) *Program Manual for the Wisconsin Package* (Genetics Computer Group, Madison, WI), Version 9.0.
28. Ito, H., Fukuda, Y., Murata, K. & Kimura, A. (1983) *J. Bacteriol.* **153,** 163–168.
29. Omura, T. & Sato, R. (1964) *J. Biol. Chem.* **235,** 2379–2385.
30. Williams, D. C., Carroll, B. J., Jin, Q., Rithner, C., Lenger, S. R., Floss, H. G., Coates, R. M., Williams, R. M. & Croteau, R. (2000) *Chem. Biol*. **7,** 969–977.
31. Croteau, R., Hefner, J., Hezari, M. & Lewis, N. G. (1995) *Curr. Top. Plant Physiol.* **15,** 94–104.
32. Hefner, J. & Croteau, R. (1996) *Methods Enzymol.* **272,** 243–250.
33. Schopfer, C. R. & Ebel, J. (1998) *Mol. Gen. Genet.* **258,** 315–322.
34. Durst, F. & Nelson, D. R. (1995) *Drug Metab. Drug Interact.* **12,** 189–206.
35. Winkler, R. G. & Helentjaris, T. (1995) *Plant Cell* **7,** 1307–1317.
36. Bishop, G. J., Harrison, K. & Jones, J. D. (1996) *Plant Cell* **8,** 959–969.
37. Szekeres, M., Nemeth, K., Koncz-Kalman, Z., Mathur, J., Kauschmann, A., Altmann, T., Redei, G. P., Nagy, F., Schell, J. & Koncz, C. (1996) *Cell* **85,** 171–182.
38. Brown, C. A. & Black, S. D. (1989) *J. Biol. Chem.* **264,** 4442–4449.
39. Nelson, D. R. & Strobel, H. W. (1988) *J. Biol. Chem.* **263,** 6038–6050.
40. Sakaguchi, M., Mihara, K. & Sato, R. (1987) *EMBO J.* **6,** 2425–2431.
41. Shimada, H., Sligar, S. G., Yeom, H. & Ishimura, Y. (1997) *in Oxygenases and Model Systems*, ed. Funabiki, T. (Kluwer, Boston), pp. 195–221.
42. von Wachenfeldt, C. & Johnson, E. F. (1995) in *Cytochrome P 450: Structure*, *Mechanism*, *and Biochemistry*, ed. Ortiz de Montellano, P. R. (Plenum, New York), pp. 183–223.