Published in final edited form as:

J Inorg Biochem. 2018 November ; 188: 96–101. doi:10.1016/j.jinorgbio.2018.08.006.

A new approach to understanding structure-function relationships in cytochromes P450 by targeting terpene metabolism in the wild

Nathan Wong^a, Xinyue Liu^b, Hannah Lloyd^c, Allison Colthart^a, Alex Ferrazolli^c, Deani Cooper^c, Yihao Zhuang^a, Phillix Esquea^b, Jeffrey Futcher^b, Theodore M. Pochapsky^d, Jeffrey Matthews^d, and Thomas C. Pochapsky^{*,a,b}

^a)Department of Biochemistry, Brandeis University, 415 South St., Waltham, MA 02454-9110

^b)Department of Chemistry, Brandeis University, 415 South St., Waltham, MA 02454-9110

^c)Department of Biology, Brandeis University, 415 South St., Waltham, MA 02454-9110

^d)Malden Catholic High School, 90 Crystal St. Malden, MA 02148

Abstract

A strategy for elucidating sequence determinants of function in the class of cytochrome P450 (CYP) enzymes that catalyze the first steps of terpene metabolism in wild microbiomes is described. Wild organisms that can use camphor, terpineol, pinene and limonene were isolated from soils rich in coniferous waste. Cell free extracts and growth beers were analyzed by gas chromatography/mass spectrometry to identify primary oxidative metabolites. For one organism, *Pseudomonas nitroreducens* TPJM, a cytochrome P450 (CYP108B1) isolated from cell free extracts was demonstrated to catalyze the oxidation of α -terpineol in assays combining the native ferredoxin and putidaredoxin reductase, and the resulting oxidation products identified by gas chromatography/mass spectrometry. Shotgun sequencing of *PnTPJM* identified four candidate P450 genes, including an apparently fragmentary gene with a high degree of homology with the known enzyme CYP108 (P450_{terp}).

Introduction

The cytochrome P450 superfamily of enzymes is widespread throughout the kingdoms of living organisms. Thought to have arisen even before the “Great Oxidation” 2.5 billion years B.P. (3), the P450s have evolved to selectively bind and oxidize myriad substrates, each with multiple possible sites for oxidation. Thus, a selective P450 must not only bind the correct substrate, but must also orient that substrate in the active site so that the appropriate site is presented to the reactive oxygen species generated.

*To whom correspondence should be addressed: pochapsk@brandeis.edu.

Supplementary material

Supplementary material includes GC traces of *PnTPJM* growth beer and CYP108B1 turnover assays, the mass spectrum of primary turnover product **6**, SDS-PAGE gel comparison of CYP108B1 with CYP2D6 and CYP101A1 and sequences of P450 open reading frames identified in *PnTPJM*. Results of contig assembly from shotgun sequencing of *PnTPJM*, as well as open reading frames from *PnTPJM* not found in the template genome, are available from TCP upon request.

Despite the plethora of possible substrate/product combinations for P450s, the P450 fold is unique and also highly conserved. All P450s for which structures have been determined exhibit identical folding topologies, with recognizably homologous secondary structural features in almost all of them (Figure 1) (5). Clearly, nature settled on the P450 fold early in evolution as a safe means of performing a dangerous reaction, the activation of molecular oxygen. Also clear is that the P450 fold provides structural and dynamic elements that must be conserved in order for the enzyme to accomplish the basic task of oxygen activation without self-destructing. On the other hand, there must also be regions that are sufficiently variable that the enzyme can evolve to occupy a specific functional niche. We have recently shown by NMR-directed molecular dynamics that two bacterial P450s with highly divergent functions, CYP101A1 and MycG, exhibit similar conformational changes upon substrate binding (7). CYP101A1 (cytochrome P450_{cam}) catalyzes the 5-*exo* hydroxylation of camphor in the first step of camphor catabolism by *Pseudomonas putida*, while MycG, a bifunctional P450 from *Micromonospora griseorubida*, catalyzes the stereospecific hydroxylation and epoxidation of mycinamicin IV to yield the macrolactone antibiotic mycinamicin II. Interestingly, the regions that show the least displacement relative to the heme upon substrate binding in both enzymes are also the most conserved: While overall sequence identity between CYP101A1 and MycG is only 23.9% based on structural alignment, within those regions that show little or no displacement upon substrate binding, including the axial cysteine binding loop and the J, K and L helices, homology is much higher, yielding 63% identity, 78% similarity and no gaps. This suggests that these regions form the “core” structural elements required for safe activation of O₂.

While comparisons of P450s with very different functions are useful for identifying common features, they are less helpful when considering the subtle differences that control substrate selectivity and regio- and stereospecificity of the reaction that is catalyzed. We have recently observed that the entire P450 structure is sensitive to substrate orientation in the CYP101A1 active site (8). This suggests that regions remote from the active site are critically involved in determining where precisely oxidation will occur on the bound substrate. For this reason, we have begun to investigate a class of cytochromes P450 that catalyze similar reactions within the same environment, those involved in the catabolism of monoterpenes by bacterial communities (microbiomes) in pine forest soils (9,10). These microbiomes are critical members of forest ecosystems, remediating terpene components of pine resins and oils that would poison the soil if allowed to build up. It is our expectation that careful comparative functional, structural and genetic analysis of the P450 enzymes that catalyze the initial steps of terpene metabolism will lead to a more complete understanding of the roles of particular structural features and even individual residues in determining the preferred substrate for a given P450 as well as the regiospecificity of the oxidation(s) it catalyzes. We have chosen to work with organisms isolated from the same environment in the hope that horizontal gene transfers would minimize sequence divergence due to random mutation, allowing us to focus on those sequence differences that occur in regions we have previously found to be important in substrate selection and product specificity (SS/PS) in other P450s that we have examined.

The current report describes the approaches that we are developing for this effort, which includes isolation of target organisms, chemical characterization of the metabolites resulting

from their use of various simple terpenes found in coniferous tissue and nearby soils, genetic analysis via high-throughput genome sequencing, genome alignment and homology searching to identify gene loci that are likely involved in terpene metabolism, and (ultimately) gene synthesis, expression and characterization of any cytochrome P450s and associated proteins that are identified.

Results

Strain isolation and identification.

Soil samples were collected, extracted and plated onto minimal media containing one of four common components of pine oil: camphor **1**, α -terpineol **2a**, limonene **3** or α -pinene **4**. The plates were also supplemented with 0.1% glucose, sufficient to allow colonies to become established. Those colonies that showed vigorous growth on these plates were selected for further investigation. Extraction of genomic DNA followed by sequencing of the 16S rRNA genes allowed the strains to be typed by genus (Table 1). Of the strains isolated from soil samples, *Streptomyces* are the most common. This is in contrast with the results of direct extraction of pine resin samples, in which *Pseudomonas* species are largely predominant (9).

We next plated each of the isolated strains in serial dilution to determine the specificity of each strain for a given terpene as carbon source (Table 1). Many of these strains proved to be capable of using more than one terpene, suggesting that more than one operon involved in primary catabolic processes is present, either chromosomally or encoded on plasmids.

Isolation and identification of *Pseudomonas nitroreducens* TPJM.

In a preliminary effort, we isolated from soil obtained from a pine swamp in Whipple Hill Park (Lexington, MA) a strain of bacteria that is highly efficient in the use of α -terpineol as a carbon source. 16S rRNA sequencing identified this organism as *Pseudomonas nitroreducens*. Peterson and co-workers had previously identified another *Pseudomonas* strain capable of metabolizing α -terpineol and characterized the cytochrome P450 from that strain (P450_{terp}, or CYP108) (1,4), providing us with a potentially useful comparison. Given that *Streptomyces* genomes typically contain multiple P450 genes, while pseudomonads often contain only one (or none), we chose to use this strain (*PnTPJM*) as a test case for approaches to identifying P450s from wild organisms.

Isolation of an active P450 from *PnTPJM*.

PnTPJM was grown in 1 L batches in minimal media regularly supplemented with α -terpineol in order to maintain growth and stabilize any plasmids required for terpeneol metabolism. Cells were harvested soon after the last addition of terpeneol, to ensure that all protein components needed for terpeneol metabolism were present in the cell paste. Cell lysis followed by ion exchange (DEAE) and size exclusion chromatography in terpeneol-supplemented buffers yielded a hemoprotein with an absorbance maximum at 393 nm, as expected for a high-spin cytochrome P450, that upon reduction and exposure to carbon monoxide yielded the typical spectrum for a CO-bound ferrocyclochrome P450 with a $\lambda_{\max}$ at 450 nm (Figure 1). A later fraction eluting from the DEAE column yielded spectroscopic evidence for a highly expressed ferredoxin ($\lambda_{\max} = 417$ nm). SDS-PAGE gel electrophoresis

comparing CYP101A1 (414 residues) and CYP2D6 (480 residues) indicate that the P450 isolated from *PnTPJM* (hereafter referred to as CYP108B1) is ~450–490 residues in length.

Combinations of the ferredoxin and CYP108B1-containing fractions did not yield evidence of α -terpineol turnover. However, addition of putidaredoxin reductase (PdR) to the P450-ferredoxin mixture resulted in NADH consumption, suggesting that the P450 requires both a ferredoxin and a reductase for activity, as was observed for CYP108 (1). The same assay (that is, NADH consumption in the presence of terpineol) on raw cell-free extracts of *PnTPJM* also showed activity. The assays were run on a sufficient scale to permit extraction of organic materials into methylene chloride, which were then analyzed by gas chromatography/mass spectrometry to identify potential metabolites (Fig. 2). Extraction of growth beers of *PnTPJM* show the primary metabolite is terpineol hydroxylation product **5a**, along with the acetylated form **5b** as well as further oxidized ketone **6** (Figure S4). On the other hand, the assay consisting of the purified CYP108B1, partially pure ferredoxin and PdR yielded almost exclusively ketone **6** (Figures S1–S3).

Shotgun sequencing of the *PnTPJM* genome and identification of P450 genes.

The evidence that a P450 was present and responsible for the primary oxidation of α -terpineol was sufficient to warrant genomic characterization of *PnTPJM*. Attempts to isolate plasmid DNA from the strain were unsuccessful; whether this was the result of plasmid fragility or the absence of any plasmid remains unclear. We therefore prepared whole genomic DNA from *PnTPJM* in sufficient purity for shotgun sequencing. The sequencing and assembly of the genome relative to the reference *Pseudomonas* genome is described more completely in the Experimental section. Our analysis yielded four open reading frames (ORFs) from *PnTPJM* that show significant homology to known P450s (see Supplementary Material). While none of the four genes as sequenced encoded a protein that matched with the gel-estimated size of CYP108B1, a gene encoding a 379-residue protein shows a high degree of homology with a RefSeq-annotated *Rhodococcus* P450 (GenBank entry [WP_080726552.1](#)) that encodes a P450 that is 455 residues in length, and also shows significant homology with CYP108 (See Figure 2). A Met residue at position 80 in the *Rhodococcus* sequence matches with the presumed initiator Met in the *PnTPJM* gene. Within the range of comparison, the two genes show 95% identity, 98% similarity, and 0% gaps. Given that there are multiple incomplete reads in the region of the genome where the ORF is located, and nearby open reading frames in the vicinity do not match with recognizable sequences, we suspect that the sequenced gene is fragmentary. Based on this assumption, we used the intact 455-residue *Rhodococcus* protein sequence to model the structure shown in Figure 3. The SwissModel program selected the crystal structure of CYP108 (P450_{terp}, PDB entry 1CPT) as the most suitable template for modeling, further supporting the likelihood that the 379-residue gene from *PnTPJM* is a fragment of the gene encoding CYP108B1. We are currently attempting to confirm this relationship by mass spectrometric sequencing of tryptic digests of the P450.

Metabolic profiles of beers and cell-free extracts from growths of terpene-metabolizing strains.

Other than *PnTPJM*, the other organisms were isolated from soil samples obtained at the same location. While the organisms on which we are focusing our current efforts are all capable of primary metabolism of the four terpenes used for selection, we observed large numbers of colonies formed on the primary culture plates that were resistant to the bacteriostatic action of the terpene. Although they could not use the terpene for growth, they could abide its presence, and presumably did not flourish due to the lack of sufficient glucose in the solid media.

With the exception of *PnTPJM*, extracts of growth beers from liquid cultures of the terpene-metabolizing strains typically showed only the remaining primary terpene nutrient, along with impurities observed by GC/MS analysis of the commercial terpene used. This indicates that the uptake and metabolism in these strains is selective for the particular terpene nutrient added, and that there is little or no excretion of metabolic products into the environment (that is, in the absence of competitors or symbiotes the organisms are efficient at using the terpene to support life processes).

For metabolic screening, cell-free extracts (CFEs) from the organism were taken from growths using the terpene that supported that strain's growth at maximum dilution of cells (Table 1). CFEs from *Streptomyces* strain 704 grown on terpineol **2a** contain significant amounts of unmetabolized terpineol after 24 hours. Furthermore, cell free extracts of strains 716, 717, 718 and 720 grown on limonene media all show relatively large amounts of terpineol in the extracts, suggesting that hydration of the vinyl group of limonene to form **2a** is an early step in metabolism.

Although *Pseudomonas* strain 715 metabolized all four terpenes tested, pinene **4** was selected for metabolic characterization of this strain. Again, CFEs from strain 715 was found to contain significant amounts of terpineol **2a**. The conversion of **4** to **2a** requires the cleavage of a carbon-carbon bond in the four-membered fused ring of **4**, as well as either a 1,2 methyl or 1,2 vinyl shift. This could proceed through oxidation of **4** to a primary alcohol, followed by generation of a non-classical carbonium ion leading to **2a**. However, considerably more information is required to confirm this possibility.

Discussion

Perhaps the most important observation that we have made thus far is that many (most) of the organisms isolated from a single environment are capable of using more than one terpene for their carbon and energy needs. Whether this is due to the presence of multiple primary monooxygenases, each capable of initiating catabolism of a particular terpene, or perhaps fewer less selective enzymes that can hydroxylate more than one substrate remains to be seen. Furthermore, colonies that establish themselves on solid media in the presence of terpene and small amounts of glucose but failed to thrive may represent symbiotes that are resistant to the bacteriostatic effects of terpenes but rely on primary metabolism by other organisms to produce carbon sources that they can use. As noted above, while most of the strains discussed here did not excrete significant amounts of processed terpene into the

growth medium, the beers examined were for the most part from log phase growth cultures with rapidly dividing cells, rather than the steady state situation one might expect in the wild microbiome. In situ in the wild, these organisms may well process more terpenes than they require for their own needs, excreting preliminary processed terpenoids that can be used by other organisms. Indeed, by far the most common genes (> 50) we have identified in the *PrTPJM* genome thus far are identified as coding for ABC transporters, likely responsible for the import/export of raw materials for metabolism and export of processed materials into the environment.

The research described here represents a comprehensive approach to a detailed understanding of sequence/structure/function relationships in the cytochrome P450 superfamily. While current sequence homology and threading routines are sufficient to identify likely P450s in a newly sequenced genome, the annotation available in genomic databases typically provides little more information. Given the importance of P450s in synthetic biology and bio-engineering efforts (11–13), the ability to rapidly identify enzymes with a desired activity based on sequence alone seems a worthy goal. We ultimately aim to couple the results of this project with the understanding of structure, dynamics and function of P450 enzymes that we are obtaining from multinuclear solution nuclear magnetic resonance (NMR) characterization of substrate binding and recognition in the P450 superfamily.

Experimental

Isolation of strains.

Soil samples were collected from local coniferous environments, including Whipple Hill Park, Lexington MA. Soils were extracted with distilled water, and the extracts plated on minimal media agar plates containing 0.1 % glucose and 0.1% of the appropriate terpene (terpineol, limonene, camphor or pinene). The terpene was also dissolved in a minimum amount of ethanol, placed in the plate lids, and after evaporation of the ethanol, the lids placed over the culture plates. All of the terpenes used are sufficiently volatile that strains capable of using the terpene as carbon and energy source were able to grow. While most plates showed multiple small colonies (suggesting that the organisms were resistant to excess terpene), only those colonies that continued to grow over several days were chosen for further characterization.

16S rRNA sequence identification of strains.

Strains of the sixteen isolated bacterial strains capable of using at least one of the four terpenes as sole carbon and energy source were transferred into 5 mL liquid cultures containing minimal media supplemented with 0.1% glucose and ~5 mg of the appropriate terpene. Cultures were aerated by shaking and grown overnight at 27°C. After centrifugation, the genomic DNA was extracted by a standard kit protocol, and PCR was performed to amplify the 16S ribosomal RNA gene. After purification of the PCR product, the rRNA genes were sequenced by standard methods, and the resulting genes compared with known genotypes to identify the genus of the bacterial isolates.

Shotgun sequencing of *Pseudomonas nitroreducens* TPJM.

This strain (collected at Whipple Hill, Lexington MA) is particularly efficient in the use of terpineol **2a** as a sole carbon and energy source, and, given the relative simplicity of the *Pseudomonas* genome, was chosen to optimize the data collection and analysis of genomic DNA that will be used for other strains. The 16S rRNA sequence was >99 % match for both *P. fluorescens* and *P. nitroreducens*. An Illumina HiSeq sequencer was used to sequence genomic DNA fragments ~200 to 400 bps in length (paired indexing), and the reads were assembled after index tag removal into contiguous DNA strands (contigs) using the program Velvet (14). 1.34×10^6 short-paired reads from a total of 1.58×10^6 were assembled using a 35-base kmer and a minimum contig length of 400. 2311885 unordered contigs > 400 bases in length were generated. These contigs were matched to the known genome of *Pseudomonas fluorescens* SBW25 (GenBank entry [NC_012660.1](#), ref. (15)) using the alignment program Mauve (16). Open reading frames in the assembled genome were identified using GenMarkS (17), and potential P450 genes identified by BLAST searching against a library of known or suspected pseudomonad cytochrome P450 genes.

Purification of CYP108B1 from *P. nitroreducens* TPJM.

1L of a minimal media supplemented with 0.1% glucose was inoculated with an overnight 5 mL culture of *Pn*TPJM to which 5 μ L of α -terpineol was added. After inoculation, 20 mg of α -terpineol was added and the flask shaken vigorously at 27° C and monitored by odor to supplement α -terpineol as needed. Cells were harvested near the end of log phase growth, while the culture still contained sufficient α -terpineol to be detected by odor. After freezing at –80° C overnight, cells from 5 L of growth were lysed by sonication in the presence of toluene sulfonyl chloride protease inhibitor, DNase 1 and RNase A, the pellets separated by centrifugation and discarded. The supernatant was filtered and loaded onto an anion exchange DEAE column equilibrated with buffer containing 20 mM Tris HCl pH 8.0, 10 mM β -mercapatoethanol and 1.5 mM α -terpineol. After washing with 6 column volumes of the same buffer, a salt gradient (0 – 600 mM KCl) was used to elute bound proteins. The fractions were monitored by UV-visible spectroscopy. The fraction that elutes at ~300 mM KCl showed a strong absorption band at 393 nm, indicating the presence of a high-spin cytochrome P450. An overlapping fraction eluting at ~400 mM KCl showed spectroscopic evidence for a ferredoxin, and was saved separately. Further purification of the P450-containing band was accomplished by centrifugal concentration and size exclusion chromatography (100 kDa cutoff gel) run with buffer containing 100 mM KCl, 50 mM Tris HCl pH 7.4 and 2 mM α -terpineol. The purified P450 gave a single band on SDS-PAGE gel, using CYP101A1 (414 residues) and CYP2D6 (480 residues) as reference markers. The isolated P450 ran almost identically with CYP2D6 and somewhat behind CYP101A1.

Activity assays.

Preliminary assays were performed in a 96-well plate reader, with wells containing 25 μ L each of 100 mM KCl, 50 mM Tris HCl pH 7.4, 2 mM α -terpineol and 10 mM NADH. The loss of NADH absorption at 340 nm was monitored. To each well, 2 μ L the purified *Pn*TPJM P450 and the partially purified ferredoxin was added. No NADH oxidation was observed until 2 μ L of putidaredoxin reductase (PdR) was added. After scale-up to 1 mL, the

functional reaction mixture was allowed to proceed until NADH was consumed, and then was extracted with methylene chloride for analysis as described below.

Analysis of cell free extracts and beers from terpene-metabolizing strains.

Growth beers from overnight growths of the strains discussed in Table 1 were extracted with methylene chloride, passed over anhydrous magnesium sulfate to remove water and the resulting extracts analyzed by gas chromatography/mass spectrometry (GC/MS) on an Agilent 7890A gas chromatograph equipped with a 5975C electron impact ionization quadrupole mass spectrometer. Peaks were identified by matching mass spectra with the library supplied with the ChemStation software.

In the case of cell free extracts, cells were grown in the presence of the terpene of interest, lysed by sonication and the cell-free extracts extracted with methylene chloride. If an emulsion formed, the emulsions were broken up by freezing prior to separation of the organic and aqueous layers. The crude methylene chloride extracts were passed over anhydrous magnesium sulfate prior to GC/MS analysis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

TCP thanks Prof. Susan Lovett (Dept. of Biology, Brandeis University) for many helpful discussions and gracious use of laboratory equipment during his sabbatical in her laboratory. HL acknowledges support from a T. Helaine Allen Summer Fellowship (Brandeis University). NRW and AMC acknowledge support from NIH training grant T32GM007596.

References cited

1. Peterson JA, Lu JY, Geisselsoder J, Graham-Lorence S, Carmona C, Witney F, and Lorence MC (1992) Cytochrome P-450terp - Isolation and purification of the protein and cloning and sequencing of its operon. *Journal of Biological Chemistry* 267, 14193–14203 [PubMed: 1629218]
2. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, and Schwede T (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research* 42, W252–W258 [PubMed: 24782522]
3. Danielson PB (2002) The Cytochrome P450 Superfamily: Biochemistry, Evolution and Drug Metabolism in Humans. *Current Drug Metabolism* 3, 561–597 [PubMed: 12369887]
4. Hasemann CA, Ravichandran KG, Peterson JA, and Deisenhofer J (1994) Crystal structure and refinement of cytochrome P450terp at 2.3 Å resolution. *J Mol Biol* 236, 1169–1185 [PubMed: 8120894]
5. Pochapsky TC, Kazanis S, and Dang M (2010) Conformational plasticity and structure/function relationships in cytochromes P450. *Antioxidants & Redox Signaling* 13, 1273–1296 [PubMed: 20446763]
6. Poulos TL, Finzel BC, and Howard AJ (1987) High-resolution crystal structure of cytochrome P450cam. *J Mol Biol* 195, 687–700 [PubMed: 3656428]
7. Tietz DR, Colthart AM, Pochapsky SS, and Pochapsky TC (2017) Substrate recognition by two different P450s: Evidence for conserved roles in a common fold. *Scientific Reports* 7, 13581 [PubMed: 29051575]

8. Asciutto EK, and Pochapsky TC (2018) Some surprising implications of NMR-directed simulations of substrate recognition and binding by cytochrome P450cam (CYP101A1). *J Mol Biol* 430, 1295–1310 [PubMed: 29596916]
9. Vilanova C, Marín M, Baixeras J, Latorre A, and Porcar M (2014) Selecting Microbial Strains from Pine Tree Resin: Biotechnological Applications from a Terpene World. *PLOS ONE* 9, e100740 [PubMed: 24971580]
10. Dorado-Morales P, Vilanova C, Garay CP, Marti JM, and Porcar M (2015) Unveiling Bacterial Interactions through Multidimensional Scaling and Dynamics Modeling. *Scientific Reports* 5, 18396 [PubMed: 26671778]
11. Wang CL, Zada B, Wei GY, and Kim SW (2017) Metabolic engineering and synthetic biology approaches driving isoprenoid production in *Escherichia coli*. *Bioresource Technology* 241, 430–438 [PubMed: 28599221]
12. Behrendorff J, and Gillam EMJ (2017) Prospects for Applying Synthetic Biology to Toxicology: Future Opportunities and Current Limitations for the Repurposing of Cytochrome P450 Systems. *Chemical Research in Toxicology* 30, 453–468 [PubMed: 27957859]
13. Bakkes PJ, Riehm JL, Sagadin T, Ruhlmann A, Schubert P, Biemann S, Girhard M, Hutter MC, Bernhardt R, and Urlacher VB (2017) Engineering of versatile redox partner fusions that support monooxygenase activity of functionally diverse cytochrome P450s. *Scientific Reports* 7, 9570 [PubMed: 28852040]
14. Zerbino DR (2002) Using the Velvet de novo Assembler for Short-Read Sequencing Technologies in *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc. pp
15. Silby MW, Cerdeño-Tárraga AM, Vernikos GS, Giddens SR, Jackson RW, Preston GM, Zhang X-X, Moon CD, Gehrig SM, Godfrey SA, Knight CG, Malone JG, Robinson Z, Spiers AJ, Harris S, Challis GL, Yaxley AM, Harris D, Seeger K, Murphy L, Rutter S, Squares R, Quail MA, Saunders E, Mavromatis K, Brettin TS, Bentley SD, Hothersall J, Stephens E, Thomas CM, Parkhill J, Levy SB, Rainey PB, and Thomson NR (2009) Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*. *Genome Biology* 10, R51 [PubMed: 19432983]
16. Darling ACE, Mau B, Blattner FR, and Perna NT (2004) Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Research* 14, 1394–1403 [PubMed: 15231754]
17. Lomsadze A, Gemayel K, Tang S, and Borodovsky M (2017) Improved Prokaryotic Gene Prediction Yields Insights into Transcription and Translation Mechanisms on Whole Genome Scale. *bioRxiv*, <https://doi.org/10.1101/193490>

Figure 1. (Red, bottom) Ultraviolet-visible spectrum of CYP108B1 isolated from *PnTPJM* with saturating α -terpineol present. λ_{max} is at 393 nm. (Blue, top) same sample after reduction with sodium dithionite and treatment with carbon monoxide. Major peak is at 450 nm, with a shoulder at 420 nm.

Top line P450_{terp}, bottom (subj) TPJM gene fragment (379 residues)

```

Query  61  MQIGKQPGLFSNAEGSE  ----- ILYDQMNEAFMRSISGGCPHVIDSLTSM DPPTHAY  113
          M+I +QP +F++A G+E          IL +Q  + F +SI+ G  ++++LT +DPP HTA
Sbjct  1   MEIERQPDVFTSA -GAENKGSVNPILANQVG DQFTKSINNGSLRILETLTYLDPPEHTAV  59

Query  114 RGLTLNWFQPASIRKLEENIRRIAQASVQRLLD F DGECD FMTDCALYYPLHVMTALGVP  173
          + + L+WF+P ++ + E  IR +A+ +VQRL+ DG+ DF+ D A++YPLHV+M+  GVP
Sbjct  60  KDIALDWFRPGNLSQWEVRIRELAKNA  VQRLLET D GKIDFVKDFAIHYPLHVIMSLFGVP  119

Query  174 EDDEPLMLKLTQDFFGVHEPDEQAVAAPRQSADEAARRFHETIATFYDYFNGFTVD  ---R  230
          E DEP M+ LTQ+FFG +PD          SAD AA+++ TI FY YF  +  R
Sbjct  120 ESDEPRMMLTQEFGTADPD AIRE DVEPLSADAAAKQWVATI QDFYSYFEKL  LLXXXXR  179

Query  231 RSCPKDDVMSLLANS -KLDGNYIDDKYINAYYVAIATAGHDTTSSSSGGAIIGLSRNPEQ  289
          R+ P+DD+ +++A + K DG Y  ++  Y++AIATAGHDTTSS+ G ++ L RNPE
Sbjct  180 RAHPQDDLATIIAQARKEDGEYYPNEVAYGYFIAIATAGHDTTSSSTLSGGMLELGRNPEL  239

Query  290 LALAKSDPALIPRLVDEAVRWTAPVKSFMRTALADTEVRGQNIKRGRIMLSYPSANRDE  349
          LA + DP LIP LV+E++RW +PVK FMR          +RG+ +  GDR ML Y S NRD
Sbjct  240 LARVREDPKLIPNLVNESLRWASPVKHFMRMVTQPYTLRGRELNP GDRFMLLYQSGNRDA  299

Query  350 EVFSNPDEFDITRFPNRHLGFGWGAH  MCLGQHLAKLEMKIFFEELLPKLKSVELSGPPRL  409
          E  NPD FDI R PN+H+ FG+G HMC+GQHLAKLE++I FEELLP ++S+  G  +
Sbjct  300 ETIPNPDVFDIDRRPNKHIAFGYGPHMCIGQHLAKLELRIMFEELLPHIESLSVSVGENKE  359

Query  410 VATNFVGGPKNVPIRF  425
          + TNFVGG KN+P++
Sbjct  360 IQTNFVGGLKNMPVQI  375

```

Figure 2.

BLAST alignment of cytochrome P450_{terp} (query) (1) with 379-residue gene fragment from TPJM contig (subj), yielding 47% identity, 67% similarity and a 3% gap penalty.

Figure 3.

Model of a possible structure for CYP108B1 generated from the 455-residue GenBank entry [WP_0807264552.1](#) (*Rhodococcus*) using SwissModel (2), and templated on the structure of CYP108 (P450_{terp}), PDB entry 1CPT (4). Secondary structures are labeled according to the scheme of Poulos (6). Sequence references are to: Helix **A**, Glu 43-Asn 56; **B**, His 76-Arg 84; **B-B'** loop, Asn 85-Gln 106; **B'**, Gln 107-Asn 117; **C**, Pro 133-Phe 146; **D**, Pro 148-Ala 171; **E**, Phe 178-Phe 195; **F**, Glu 202-Phe 213; **G**, Ala 233-Ala 257; **H**, Leu 263-Gln 269; **I**, Asn 280-Arg 311; **J**, Pro 313-Glu 321; **K**, Pro 323-Ala 337; **β3-β4** sheets, His 342-Leu 367; **K'**, Tyr 368-Arg 373; **L**, Gly 405-His 423; **β5** sheet, Val 428-Gln 450.

Scheme 1.

Terpenes and terpene oxidation products referred to in the text.

Table 1.

Isolates capable of using listed terpenes as sole carbon and energy source. Numbers refer to the maximum dilution of cells for which growth was observed on the given carbon source. For example, *Streptomyces* strain 711 was able to grow on terpineol media up to the highest dilution (6), but noticeable growth was observed on camphor only for the undiluted culture (1), and did not grow at all on limonene or pinene.

Strain #	Genus	Limonene (0.2%)	Camphor (0.2%)	Terpineol (0.1%)	Pinene (0.2%)
705	<i>Dermacoccus</i>	2	3	2	2
706	<i>Streptomyces</i>	0	4	5	0
707	<i>Streptomyces</i>	1	1	0	1
708	<i>Streptomyces</i>	2	1	3	2
709	<i>Streptomyces</i>	2	3	5	2
710	<i>Streptomyces</i>	0	4	5	0
711	<i>Streptomyces</i>	0	1	6	0
714	<i>Streptomyces</i>	4	3	4	3
715	<i>Pseudomonas</i>	6	6	6	6
716	<i>Streptomyces</i>	4	4	3	3
717	<i>Streptomyces</i>	3	1	2	2
718	<i>Pseudomonas</i>	4	5	3	3
719	<i>Streptomyces</i>	0	1	1	0
720	<i>Streptomyces</i>	6	5	6	5
721	<i>Rhodococcus</i>	2	6	5	2
722	<i>Rhodococcus</i>	2	6	4	2