The biosynthesis of thymol, carvacrol, and thymohydroquinone in Lamiaceae proceeds via cytochrome P450s and a short-chain dehydrogenase

Sandra T. Krausea,1, Pan Liaoa, b, Christoph Crocollb,1, Benoit Boachonb, d, e, Christiane Förstera, b, c, Franziska Leideckera, b, Natalie Wiesea, Dongyan Zhaoa, b, c, Joshua C. Wooda, c, R. Robin Buelle, c, f, Jonathan Gershenzonc, Natalia Dudarevb, h, j, and Jörg Degenhardtia, b, e, f

a Institute of Pharmacy, Department of Pharmaceutical Biotechnology, Martin Luther University Halle-Wittenberg, 06120 Halle, Germany; b Department of Biochemistry, Purdue University, West Lafayette, IN 47907-2063; c Department of Biochemistry, Max Planck Institute for Chemical Ecology, 07745 Jena, Germany; d University of Lyon, UJM-Saint-Etienne, CNRS, BVPam Formation de Recherche en Evolution 3727, 42000 Saint-Etienne, France; e Department of Plant Biology, Purdue University, West Lafayette, IN 47907; f Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, IN 47907-2010

Edited by Rodney Croteau, Washington State University, Pullman, WA; received June 2, 2021; accepted October 31, 2021

Thymol and carvacrol are phenolic monoterpenes found in thyme, oregano, and several other species of the Lamiaceae. Long valued for their smell and taste, these substances also have antibacterial and anti-spasmodic properties. They are also suggested to be precursors of thymohydroquinone and thymoquinone, monoterpenes with anti-inflammatory, antioxidant, and anti-tumor activities. Thymol and carvacrol biosynthesis has been proposed to proceed by the cyclization of geranyl diphosphate to γ-terpinene, followed by a series of oxidations via p-cymene. Here, we show that γ-terpinene is oxidized by cytochrome P450 monoxygenases (P450s) of the CYP71D subfamily to produce unstable cyclohexadienol intermediates, which are then dehydrogenated by a short-chain dehydrogenase/reductase (SDR) to the corresponding ketones. The subsequent formation of the aromatic compounds occurs via keto–enol tautomerisms. Combining these enzymes with γ-terpinene in vitro assays or in vivo in Nicotiana benthamiana yielded thymol and carvacrol as products. In the absence of the SDRs, only p-cymene was formed by rearrangement of the cyclohexadienol intermediates. The nature of these unstable intermediates was inferred from reactions with the γ-terpinene isomer limonene and by analogy to reactions catalyzed by related enzymes. We also identified and characterized two P450s of the CYP76S and CYP736A subfamilies that catalyze the hydroxylation of thymol and carvacrol to thymohydroquinone when heterologously expressed in yeast and N. benthamiana. Our findings alter previous views of thymol and carvacrol formation, identify the enzymes involved in the biosynthesis of these phenolic monoterpenes and thymohydroquinone in the Lamiaceae, and provide targets for metabolic engineering of high-value terpenes in plants.

aromatic monoterpenes | Lamiaceae | carvacrol | thymol | thymohydroquinone

The phenolic monoterpenes of the Lamiaceae are widely used constituents of pharmaceuticals, cosmetics, and food products (1). Exports of plants containing thymol or carvacrol are employed in medicine for their antibacterial, anti-spasmodic, antioxidant, and anti-cancer properties. Because of their pungent, warm, and aromatic odors, they also serve as additives to cosmetics and are used in aromatherapy. Thymol and carvacrol are best known as the aroma compounds of oregano and thyme, in which they provide the herbal, pizza-like tastes that are traditionally used in Mediterranean cuisine and food preservation (2). The occurrence of phenolic monoterpenes is restricted to a few genera in the Lamiaceae (Thymus, Origanum, Satureja, and Thymbra), Apiaceae (Trachyspermum), and Verbenaceae (Lippia). Of these, the essential oils of Thymus are the most important commercial source of phenolic monoterpenes (3). Thymus vulgaris L. and Origanum species also produce the structurally related monoterpenes thymohydroquinone and thymoquinone, which were first described in the essential oil of Nigella sativa L. black seed (4). Thymohydroquinone was shown to exhibit anti-inflammatory, antioxidant, cytotoxic, and anti-cancer activities (5). To date, only a few biosynthetic pathways to pharmaceutically valuable, oxidized terpenes have been completely elucidated, such as those leading to artemisinin, paclitaxel, and the phenolic, labdane-type diterpenes of sage and rosemary (7–9). For monoterpenes, a complex biosynthetic pathway has only been described for menthol and its derivatives in Mentha (10).

Significance

The monoterpane alcohols thymol, carvacrol, and thymohydroquinone are characteristic flavor compounds of thyme, oregano, and other Lamiaceae. These specialized metabolites are also valuable for their antibacterial, anti-spasmodic, and antitumor activities. We elucidated the complete biosynthetic pathway of these compounds, which starts with the formation of γ-terpinene from geranyl diphosphate. The aromatic backbone of thymol and carvacrol is formed by P450 monoxygenases in combination with a dehydrogenase via an unstable intermediate. Additional P450s hydroxylate thymol and carvacrol to form thymohydroquinone. Our findings demonstrate a mechanism for the formation of phenolic monoterpenes that differs from previous predictions and provides targets for metabolic engineering of high-value terpenes in plants.


This article is a PNAS Direct Submission. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

1T.S.K. and P.L. contributed equally to this work.

2Present address: DynaMo Center, Department of Plant and Environmental Science, University of Copenhagen, 1871 Frederiksberg C, Denmark.

3Present address: Department of Biochemistry, Max Planck Institute for Chemical Ecology, 07745 Jena, Germany.

4Present address: Biotechnology Center, Cornell University, Ithaca, NY 14853.

5To whom correspondence may be addressed. Email: joerg.degenhardt@pharmazie.uni-halle.de.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2110092118/-/DCSupplemental.

Published December 20, 2021.
However, the biosynthetic pathways for phenolic monoterpenes like thymol or carvacrol remain uncharacterized. Most monoterpenes are biosynthesized by fusion of the ubiquitous C5 intermediates, isopentenyl diphosphate, and its isomer dimethylallyl diphosphate, resulting in the formation of a C10 compound, geranyl diphosphate (GDP). This acyclic intermediate is the substrate for the large enzyme family of monoterpene synthases that form cyclic or acyclic products with an enormous variety of carbon skeletons (11, 12). In previous studies in thyme and oregano, the cyclic monoterpene olefin γ-terpinene was proposed as a precursor of thymol and carvacrol (17). Essential oil composition of the T. vulgaris chemotypes dominated by carvacrol (C type), thymol (T type), and geraniol (G type). Terpenes were extracted with hexane and analyzed by GC-MS. The following terpenes were identified: 1, α-thujene; 2, α-pinene; 3, myrcene; 4, α-terpinene; 5, p-cymene; 6, γ-terpinene; 7, cis-sabinene hydrate; 8, linalool; 9, nerol; 10, neral; 11, thymoquinone; 12, geraniol; 13, geranial; 14, thymol; 15, carvacrol; 16, geranyl acetate; 17, (E)-β-caryophyllene; 18, thymohydroquinone; and 19, germacrene D. Nonyl acetate (10 μL/mL) was added as internal standard (IS) for quantification.

In this study, we investigated the biosynthetic pathway leading to the formation of thymol, carvacrol, and thymohydroquinone in Lamiaceae. The biosynthesis of thymol, carvacrol, and thymohydroquinone in Lamiaceae proceeds via cytochrome P450s and a short-chain dehydrogenase.
subfamily from thyme and oregano accessions producing high levels of thymol and carvacrol. When these CYP genes were heterologously expressed and combined with a short-chain dehydrogenase from thyme in vitro or coexpressed in vivo in Nicotiana benthamiana, thymol or carvacrol were formed. Based on the characteristics of the expressed enzymes and their reaction with other substrates, we constructed the biosynthetic pathway leading to thymol and inferred the nature of unstable intermediates. Furthermore, we identified and characterized two P450s of the CYP76S and CYP736A subfamilies that hydroxylate thymol and carvacrol to thymohydroquinone when expressed in vivo in yeast and in N. benthamiana.

**Results**

 Isolation of a γ-Terpinene Synthase and P450 Monoxygenases of the CYP71D Subfamily from Thyme and Oregano Varieties with High Levels of Thymol and Carvacrol. To investigate the biosynthetic route leading to thymol and carvacrol, we first identified plants rich in these phenolic monoterpenes. Three *T. vulgaris* accessions were chosen from the originally described monoterpene chemotypes (3), two of which contained high levels of either thymol (T type) or carvacrol (C type) in their leaves, while the third contained high levels of geraniol and geranyl acetate (G type) and only low amounts of phenolic monoterpenes (Fig. 1B). These chemotypes were used to identify the first step of the pathway, the formation of the monocular diene γ-terpinene from GDP (Fig. 1L). In analogy to previously identified γ-terpinene synthases from oregano (14, 17) and related thyme species (15, 16, 18, 19), we isolated a gene with >90% amino acid identity, *TvTPS2*. Transcript levels correlated positively with the high-carvacrol and -thymol levels in the C and T chemotypes but were very low in plants of the G chemotype (Fig. 2A). Expression of *TvTPS2* in a bacterial system confirmed that this gene indeed encodes a functional γ-terpinene synthase (Fig. 2B).

The next step of the predicted pathway for phenolic monoterpenes requires the oxidation of γ-terpinene. Thus, we searched Expressed Sequence Tag (EST) databases (14) and available RNA sequencing (RNAseq) data of oregano and thyme for cytochrome P450s (CYPs) of the CYP71D subfamily, which were previously implicated in monoterpene oxidation. The limonene-6-hydroxylase from *Mentha spicata* (CYP71D18) (21) was used as a query in a blast search of an EST database generated from peltate trichomes of *Origanum vulgare* (14). Completion of matching EST sequences by rapid amplification of complementary DNA (cDNA) ends–PCR in two oregano accessions led to the identification of a gene designated as *CYP71D178* (22). A search for transcripts with similarities to *CYP71D178* in the C and T chemotypes of *T. vulgaris* as well as other accessions of thyme and oregano (SI Appendix, Table S1) resulted in 14 different sequences belonging to the CYP71D subfamily. Assuming that CYP sequences with amino acid identities of 97% or higher represent alleles of the same gene (23), these sequences were assigned to five additional gene names and designated as *CYP71D179*, *CYP71D180*, *CYP71D181*, *CYP71D182*, (22), and *CYP71D507*, according to the nomenclature of D. R. Nelson (24) (SI Appendix, Table S1).

All of the isolated CYP71D sequences encoded proteins with the typical amino acid motifs responsible for CYP catalysis, including a proline-rich hinge, a lipophilic membrane anchor at the N-terminal end, and a heme-binding motif in the C-terminal region, as evident from an amino acid alignment of the CYP71D enzymes, including CYP71D18 from *M. spicata* shown in [SI Appendix, Fig. S1](https://doi.org/10.1073/pnas.2110092118). A dendrogram analysis including members of the CYP71D subfamily from other plants showed that all identified thyme and oregano enzymes grouped with monoterpene-metabolizing enzymes from other Lamiaceae species (SI Appendix, Fig. S2). One cluster contained CYP71D178, CYP71D179, and CYP71D182, while a second was formed by CYP71D180 and CYP71D181. Both clusters showed a close relationship to the limonene-hydroxylating enzymes CYP71D18 and CYP71D13 from *Menha*. CYP71D507 did not cluster with the other monoterpene-metabolizing enzymes and displayed only around 58% amino acid similarity to them.

**The Activity of the CYP71D Enzymes Suggests the Oxidation of γ-Terpinene to a Cyclohexadienol Intermediate.** To examine the potential function(s) of these CYPs, we chose the CYP71D enzymes present in the T, C, and G chemotypes of *T. vulgaris*. Expression of *TvCYP71D179-T* was highest in the T type and lowest in the C type, with intermediate expression levels in the G type despite the lack of phenolic monoterpenes in this chemotype (Fig. 2C). Conversely, *TvCYP71D180-C* was almost exclusively expressed in the C type, while only trace transcript amounts were detected in the G and T types. The transcript levels of *TvCYP71D507-T* were similar in all three chemotypes. Moreover, a positive correlation between the expression of these CYPs and the levels of thymol and carvacrol was also found to be present in oregano (Fig. 3C). Metabolic profiling and transcriptomic analysis of different tissues of oregano (accession “USA”) lead to the isolation of *OvCYP71D507*, which exhibits 90.1% amino acid identity with *TvCYP71D507-T* and was well correlated with thymol levels. In addition, an oregano *OvCYP71D180* gene encoding a protein with 91% amino acid identity to *TvCYP71D180-C* appeared to be highly expressed in tissues producing high amounts of carvacrol (Fig. 3 A and B).

After expression of *TvCYP71D179-T, TvCYP71D180-C, and TvCYP71D507-T* in yeast and incubation of the microsomal fractions with γ-terpinene, the only monoterpenic product detected was p-cymene (Fig. 2D), an aromatic hydrocarbon previously proposed as an intermediate in phenolic monoterpene biosynthesis (20). However, this product was not accepted as a substrate for further oxidations by these CYPs, as would have been expected for the predicted pathway (Fig. 1A).

Because of their similarity with limonene-6-hydroxylase from *M. spicata*, we tested the thyme enzymes with (+)-limonene as a substrate. Limonene is a cyclohexanoid diene monoterpenes similar to γ-terpinene but with one of its double bonds in the isopropyl side chain instead of the ring. Both *TvCYP71D179* and *TvCYP71D507* hydroxylated (+)-limonene at an allylic position, C-3, to form (+)-trans-isopiperitenol (SI Appendix, Fig. S3 A and B). In addition, *TvCYP71D179* produced minor amounts of the allylic C-6 hydroxylation products, (+)-trans-carveol and (+)-cis-carveol. In contrast, the major product of *TvCYP71D180* was the 6-hydroxylated (+)-cis-carveol, in addition to minor amounts of (+)-trans-carveol and (+)-trans-isopiperitenol (SI Appendix, Fig. S3 A and B). Based on the conversion of limonene to either a C-3 or C-6 hydroxylated product, and the fact that the CYP71D enzymes characterized previously carry out hydroxylations at either an allylic position or on an aromatic ring, we hypothesized that *TvCYP71D179, TvCYP71D180, and TvCYP71D507* produce aliphatic alcohols from γ-terpinene. If these cyclohexadienol intermediates were unstable under the in vitro assay conditions employed, they could have been converted to p-cymene. In this interpretation, p-cymene is an artifact rather than a true intermediate of thymol and carvacrol biosynthesis. Thus, we further hypothesized that the cyclohexadienol intermediate might be converted directly to thymol or carvacrol via oxidation to the respective ketone, followed by aromatization via keto–enol tautomerism (Fig. 4A). This proposed pathway has some similarities to that of menthol biosynthesis in *Mentha × piperita*, in which the CYP71D enzyme catalyzes the allylic hydroxylation of limonene at C-3 to form (−)-trans-isopiperitenol (Fig. 4B), which is

Krause et al.
The biosynthesis of thymol, carvacrol, and thymohydroquinone in Lamiaceae proceeds via cytochrome P450s and a short-chain dehydrogenase
Fig. 2. Expression of four genes involved in the production of phenolic monoterpenes. (A) Relative transcript levels of the γ-terpene synthase gene TvTPS2 in leaves of the T. vulgaris chemotypes geraniol (G), carvacrol (C), and thymol (T). (B) GC-MS chromatogram of products of TvTPS2 after heterologous expression and incubation with the substrate GDP. (C) Relative transcript levels of the CYPs involved in phenolic monoterpene biosynthesis in leaves of the T. vulgaris chemotypes C, G, and T. (D) GC-MS chromatograms of CYP products after heterologous expression in yeast in the presence of the substrate γ-terpinene. (E) GC-MS chromatograms of products after coincubation of CYP and SDR enzymes in the presence of the substrate γ-terpinene. (F) GC-MS chromatograms of products after transient transformation of CYP71Ds and TvSDR1 in N. benthamiana in the presence of the substrate γ-terpinene. Enzyme products were identified by comparison of their retention times and mass spectra with those of authentic standards. EIC, extracted ion chromatogram; EV, empty vector control; TIC, total ion chromatogram. The following terpenes were identified: 1, α-thujene; 2, α-pinene; 3, myrcene; 5, p-cymene; 6, γ-terpinene; 12, geraniol; 14, thymol; and 15, carvacrol. In the quantitative real-time PCR experiments, each bar represents the mean value of three biological replicates ± SE (n = 3). Mean values were tested by one-way ANOVA for significant differences (P < 0.05). Tukey’s post hoc test was used for pairwise comparisons; different letters indicate significant differences.
The Short-Chain Dehydrogenase \( \text{TvSDR1} \) Catalyzes the Oxidation of Allylic Monoterpenes Alcohols to Their Corresponding Ketones. To identify genes with similarity to the \( M. \times \text{piperita} \) trans-isopiperitenol dehydrogenase (ISPD), we searched RNAseq data of the \( C, T \) and \( G \) chemotypes of \( T. \ vulturis \) and \( O. \ vulgare \). The sequences \( \text{TvSDR1-C}, \ \text{TvSDR1-T}, \) and \( \text{TvSDR1-G} \) were identified, which encoded the same deduced amino acid sequence with 79% amino acid identity to the trans-ISPD from mint. The deduced amino acid sequence of \( \text{TvSDR1} \) contained the motifs typical for this enzyme class, including TGXXXGXG, (C)NAG, YXXXK, and the catalytic tetrad Asn-Ser-Tyr-Lys (\( SI \) Appendix, Fig. S4). While it would have once been considered a “classical” SDR, this enzyme can now be assigned to the family SDR110C, according to the latest hidden Markov model–based nomenclature system (26, 27). The Asp residue at position 52 in the protein indicated a preference for NAD(H) over NADP(H) as a cofactor. No signal peptides were predicted by SignalP 4.1 (28) or iPSORT (29). Analysis of \( \text{TvSDR1} \) expression revealed no differences in transcript levels in leaves of the \( C, T \) and \( G \) chemotypes of \( T. \ vulturis \) (\( SI \) Appendix, Fig. S5). A homologous gene was found in transcriptomic data of oregano (accession “USA”). \( \text{OvSDR1} \), which exhibits 93.6% amino acid identity with thyme \( \text{TvSDR1} \), was well correlated with thymol levels (Fig. 3).

The recombinant protein \( \text{TvSDR1} \) accepted the alcohols trans- and cis-(-)-isopiperitenol and converted them to the corresponding ketone isopiperitenone (\( SI \) Appendix, Fig. S64). It also converted trans- and cis-(-)-carvone to the ketone carvone. p-Cymene was not accepted as a substrate (\( SI \) Appendix, Fig. S68).

\( \gamma \)-Terpinene Is Converted to Thymol or Carvacrol by the Combined Activity of a \( \text{CYP71D} \) Monoxygenase and the Short-Chain Dehydrogenase \( \text{TvSDR1} \). To test whether the combination of the CYPs with \( \text{TvSDR1} \) produces thymol and carvacrol, microsomes containing \( \text{TvCYP71D179}, \text{TvCYP71D180}, \) or \( \text{TvCYP71D507} \) were coincubated with recombinant \( \text{TvSDR1} \) in vitro. In assays with \( \text{TvCYP71D179} \) or \( \text{TvCYP71D507} \), the fed \( \gamma \)-terpinene substrate was converted to thymol, along with minor amounts of carvacrol in case of \( \text{TvCYP71D179} \) (Fig. 2E). Both minor amounts of thymol and carvacrol are produced by \( \text{TvCYP71D507} \). Conversely, carvacrol was produced as the major product when \( \text{TvCYP71D180} \) was used instead of \( \text{TvCYP71D179} \) in assays. All other CYPs identified in \( T. \ vulturis \) and \( O. \ vulgare \) also formed either thymol (\( \text{OvCYP71D178} \)) or carvacrol (\( \text{OvCYP71D181} \) and \( \text{TvCYP71D182} \)) after coincubation in the presence of the substrate \( \gamma \)-terpinene (\( SI \) Appendix, Fig. S7). All assays including the empty vector control contained some p-cymene, which is formed spontaneously from the \( \gamma \)-terpinene substrate. However, the concentration of p-cymene is increased approximately fivefold in the presence of the \( \text{CYP71D} \) enzymes, suggesting that p-cymene might also be formed via the degradation of the unstable dienol intermediate. To test this hypothesis, the \( \gamma \)-terpinene substrate was preincubated with the \( \text{CYP71D} \) enzymes for 15 min before the addition of \( \text{TvSDR1} \). Preincubation increased formation of p-cymene (\( SI \) Appendix, Fig. S8), indicating the instability of the \( \text{CYP71D} \) products under these assay conditions.

To determine whether the \( \text{CYP71D} \) enzymes and the identified SDR cooperate to produce phenolic monoterpenes in planta, as they do in an in vitro assay, the corresponding genes were coexpressed in \( N. \ benthamiana \), a species that does not naturally produce any phenolic monoterpenes. After infiltration with the \( \gamma \)-terpinene substrate, leaves transiently expressing \( \text{TvCYP71D179-T} \) and \( \text{TvSDR1} \) produced thymol with small amounts of carvacrol (Fig. 2F). Likewise, leaves expressing both \( \text{TvCYP71D180-C} \) and \( \text{TvSDR1} \) formed carvacrol (Fig. 2F). The combination of \( \text{TvCYP71D507} \) and \( \text{TvSDR1} \) produced both carvacrol and low levels of thymol in the tobacco expression system and thereby differed from the production in yeast. To reconstitute the complete thymol biosynthetic pathway in \( N. \ benthamiana \), we transiently expressed the \( \gamma \)-terpinene synthase \( OvTPS2 \) from \( O. \ vulgare \) (14) together with \( \text{TvCYP71D179-T} \) and \( \text{TvSDR1} \). \( OvTPS2 \) converts the endogenous \( N. \ benthamiana \) GDP to \( \gamma \)-terpine, a substrate for the CYP enzymes. Five d after infiltration, thymol was detected in case of \( \text{TvCYP71D179} \) (Fig. 2F). This suggested that \( \text{TvSDR1} \) may contribute to thymol formation in planta but is not sufficient to produce thymol in transient expression systems.
monoterpenes in the Lamiaceae and other plant families commonly occur in glandular trichomes, structures that may promote biosynthesis while minimizing autotoxicity and further metabolism (30). Only traces of monoterpenes were detected in empty vector control transformed *N. benthamiana* leaves.

**Gene Expression–Metabolite Correlation Analysis Identified P450 Candidates for Thymohydroquinone Biosynthesis.** For the further conversion of thymol and carvacrol to thymohydroquinone, an additional hydroxylation reaction was proposed to be catalyzed by a cytochrome CYP450 hydroxylase. Next, the conversion of thymohydroquinone to thymoquinone involves an oxidation reaction that could be catalyzed by an alcohol dehydrogenase (31). As *O. vulgare* produces both thymohydroquinone and thymoquinone (32), transcriptomics analysis (Dataset S1, Tabs 1 through 3) and metabolic profiling (Dataset S1, Tab 4) of different tissues, including young leaves, mature leaves, stems, and inflorescences, were performed on oregano accession “USA” to discover the biosynthetic genes involved. Gas chromatography–mass spectrometry (GC-MS) analysis revealed that thymohydroquinone and thymoquinone biosynthesis in thyme and oregano in comparison to the first steps of menthol biosynthesis in *Mentha × piperita*. (A) In thymol, carvacrol, *p*-cymene and thymohydroquinone biosynthesis, geranyl diphosphate (GDP) is first cyclized to γ-terpinene by the terpene synthase (TvTPS2). P450s hydroxylate γ-terpinene either at C-3 or C-6, and the dienol intermediates are converted by a short-chain dehydrogenase (SDR) to the corresponding ketones. These allylic ketones undergo keto-enol tautomerisms to form thymol and carvacrol. The formation of *p*-cymene results from the spontaneous rearrangement of the dienol intermediates due to their instability in aqueous conditions. Thymol and carvacrol are converted to thymohydroquinone by CYP76S40 or CYP736A300 and, subsequently, to thymoquinone by spontaneous conversion or enzymatic action. (B) First steps of the pathway from (-)-limonene to menthol in *Mentha × piperita*, which also starts with an allylic P450 oxidation. The resulting alcohol is converted to the corresponding ketone by a short-chain dehydrogenase (10). Enzyme abbreviations are the following: L3OH: (-)-limonene-3-hydroxylase (CYP71D13) and ISPDL3OH: (-)-(trans)-isopiperitenol dehydrogenase.
CYP76S40 and CYP736A300 Catalyze the Formation of Thymohydroquinone from Thymol and Carvacrol in Plants. To test whether the seven selected OvCYP candidates convert thymol and carvacrol to thymohydroquinone, the genes were transiently expressed in N. benthamiana. Five d after infiltration, leaf discs were fed with thymol or carvacrol for 24 h, and terpenoids were extracted for GC-MS analysis. Only leaves transformed with OvCYP1 (designated as OvCYP76S40) and OvCYP2 (designated as OvCYP736A300) produced significantly higher levels of thymohydroquinone after feeding with thymol and carvacrol than the empty vector–transformed control plants (SI Appendix, Fig. S12). The small amounts of thymohydroquinone detected in empty vector control leaves are likely the result of unspecific native N. benthamiana enzymes that hydroxylate carvacrol and thymol. Analysis of OvCYP76S40 and OvCYP736A300 messenger RNA transcripts by RT-qPCR in different oregano tissues further confirmed their expression in thymohydroquinone-producing tissues (SI Appendix, Fig. S13), which was consistent with the transcriptomic data. To validate the functions of OvCYP76S40 and OvCYP736A300 independently of a plant background, the proteins were expressed in yeast, followed by feeding with thymol or carvacrol. In yeast, both enzymes were also able to convert thymol and carvacrol into thymohydroquinone. A small amount of thymoquinone was also detected (Fig. 5B). Similar alleles of these genes (TvCYP1 and TvCYP2) were also identified in the RNAseq datasets from the T. vulgaris “English” accession (SI Appendix, Fig. S14). When TvCYP1 and TvCYP2 genes were transiently expressed in N. benthamiana, leaf discs produced significantly higher levels of thymohydroquinone after feeding with thymol and carvacrol, relative to the empty vector–transformed control plants (SI Appendix, Fig. S12 A and B).

Discussion
A CYP71D Monooxygenase and a Short-Chain Dehydrogenase Are Required for Thymol and Carvacrol Formation from γ-Terpinene.
Thymol, carvacrol, thymohydroquinone, and thymoquinone are valuable pharmaceuticals and flavor ingredients that are almost exclusively produced by species of the Lamiaceae. Previous studies suggested that the biosynthetic pathway to these compounds begins with the cyclization of GDP to γ-terpinene by a terpene synthase, and γ-terpinene synthases have been identified in various thyme (15, 16, 18, 19) and oregano (14, 17) species. However, little was known about the later steps, except from the pioneering work of Poulose and Croteau (20), who proposed p-cymene as an intermediate from γ-terpinene en route to aromatic monoterpenes. Here, we show that thymol...
and carvacrol formation proceeds via hydroxylation of γ-terpinene at either the 3- (to thymol) or 6- (to carvacrol) position catalyzed by CYPs of the 71D subfamily to give unstable products that are likely to be cyclohexadienols. These intermediates are next converted by an SDR, TvSDR1, to their corresponding ketones, which can then aromatize via keto–enol tautomerism, forming thymol and carvacrol (Fig. 4).

Although none of the apparent TvCYP71D179, TvCYP71D180, TvCYP71D181, TvCYP71D182, and TvCYP71D507 products were stable enough for chemical characterization, the formation of cyclohexadienols after allylic oxidation is supported by the fact that these enzymes convert the monoterpene limonene to allylic hydroxylated products (SI Appendix, Fig. S3). Enzymes of the CYP71D subfamily from other plants also produce exclusively hydroxylated products (33–37). In the in vitro assays, we detected only the aromatic hydrocarbon, p-cymene, a dehydrogenation product of the unstable α,α,β,β-unsaturated dienel enzyme product (Fig. 2D). The other unstable intermediates of the phenolic monoterpene pathway are the TvSDR1 products, which we propose as α,α,β,β-cross-conjugated cyclohexadienol, since TvSDR1 was found to convert other allylically hydroxylated terpenes to their corresponding ketones (SI Appendix, Fig. S6). Following their formation, these ketones readily aromatize to phenols.

CYP enzymes play a major role in the formation of aromatic rings in other terpenes as well. For example, the formation of gossypol, a well-studied sesquiterpene dimer from cotton, requires CYP enzymes of the 71BE, 82, and 706 subfamilies (38). These enzymes all catalyze allylic hydroxylation itself facilitates the eventual formation of a naphthalene ring system with three phenolic hydroxyl groups. In the biosynthesis of the abietane diterpene ferrugiol, a member of the CYP76 family carries out both a hydroxylation and a dehydrogenation leading to aromatization (39).

The roles of the enzymes involved in this pathway to phenolic monoterpene were also determined by coexpression of the γ-terpinene synthase OvTPS2 with TvCYP71D179-T and TvSDR1 in N. benthamiana in vivo (SI Appendix, Fig. S9), as well as the combination of the CYP71Ds with TvSDR1 in vitro (Fig. 2E). These experiments resulted in the production of thymol and carvacrol, respectively. The proposed pathway is analogous to that converting limonene to isopiperitenone in M. × piperita en route to menthol (Fig. 4P). Thyme and oregano CYP71D178-182 share 73 to 78% amino acid identity to M. × piperita limonene-6-hydroxylase (CYP71D18) (SI Appendix, Figs. S1 and S2). Similarly, TvSDR1 shares 79% amino acid identity to the (−)-trans-ISPD from M. × piperita.

Formation of either Thymol or Carvacrol Depends on the Regiospecificity of the CYP71D-Catalyzed Reaction. The pathways to thymol and carvacrol diverge at the CYP71D-catalyzed step, in which some of the characterized enzymes produce a C-3-oxygenated product as an intermediate to thymol, while others produce a C-6-oxygenated product as an intermediate to carvacrol. The regiospecificity of the enzymes was supported by gene expression–metabolite correlations, sequence alignment, and the biochemical characterization of the heterologously expressed proteins. The genes encoding CYP71D178-182 clustered in two groups. The group with CYP71D179 had higher transcript levels in thyme of the C chemotype and the “English” accession and was associated with thymol production (Fig. 2C and SI Appendix, Fig. S14). On the other hand, the second group, which contained CYP71D180 and CYP71D181, was correlated with both high transcript levels in thyme of the C chemotype and high-carvacrol production in O. vulgare (Figs. 2C and 3). Moreover, TvCYP71D179 produced the C-3-oxidized thymol when assayed in combination with TvSDR1, while TvCYP71D180 produced the C-6-oxygenated carvacrol under the same conditions (Fig. 2 E and F). TvCYP71D507 catalyzed a C-3 oxygenation in the yeast expression system (Fig. 2E and SI Appendix, Fig. S3) and both C-3 and C-6 oxygenations in after infiltration in tobacco leaves (Fig. 2F).

The position specificity of the CYP71D enzymes was also evident when the monoterpene (+)-limonene was used as a substrate in the reaction. The thymol-associated TvCYP71D179 formed chiefly the C-3-oxygenated product (+)-trans-isopiperitenone, while the carvacrol-associated TvCYP71D180 produced mainly the C-6-oxygenated product (+)-cis-carveol (SI Appendix, Fig. S3). Because of the high-sequence similarity among these enzymes, only a few amino acid residues may be crucial for the product outcome. Among the Mentha CYP71D enzymes, one amino acid position was shown to determine the regiospecificity of CYP71D13 (C-3) and CYP71D18 (C-6). In CYP71D18, limonene hydroxylation was altered from C-6 to C-3 when phenylalanine 363 was substituted by isoleucine in the substrate recognition site 5 (21). However, in the thyme and oregano CYP71D enzymes (CYP71D178-182), a phenylalanine was always present at this position with an isoleucine substitution in TvCYP71D507, which may be responsible for variability of hydroxylation among expression systems (Fig. 2 E and F and SI Appendix, Figs. S1 and S3A). Thus, phenylalanine 363 is not crucial for regiospecificity in this group. Further sequence analysis and site-directed mutagenesis are required to determine which amino acid(s) is responsible for the position-specific hydroxylation of γ-terpinene en route to thymol and carvacrol.

Interaction of a Short-Chain Dehydrogenase with the CYP71D Enzymes Drives Phenolic Monoterpene Formation. The short-chain dehydrogenase TvSDR1 shares an amino acid sequence identity of 79% to trans-ISPD from mint, which catalyzes the oxidation of trans-isopiperitenone to the respective ketone (25). SDRs have diverse sequences often sharing amino acid sequence identities as low as 15 to 30% (40, 41). Thus, with a sequence identity of 79% between TvSDR1 and the M. × piperita dehydrogenase, it is not surprising that both enzymes are associated with monoterpene metabolism and catalyze very similar reactions (Fig. 4). Short-chain dehydrogenases are also known to participate in the biosynthesis of other terpenoids, including iridoids and cardenolides (42). Our pathway reconstruction demonstrated the ability of CYP71D and SDR enzymes to work together in planta (Fig. 2 F and SI Appendix, Fig. S9). Similar interactions of a CYP enzyme and a dehydrogenase in terpene metabolism were also reported from other biosynthetic pathways. For example, the formation of perillaldehyde in different Perilla species was shown to involve a P450-mediated C-7 hydroxylation of limonene followed by the oxidation of perilla alcohol to perillaldehyde by alcohol dehydrogenases (43). In caraway, the biosynthesis of the oxygenated monoterpene carvone is realized by the hydroxylation of limonene to trans-carveol by a limonene-6-hydroxylase and the subsequent oxidation by a SDR (44). In S. officinalis, the pathway to sabinil is conducted via a P450-mediated hydroxylation of sabiniene to sabinol, which is subsequently dehydrogenated to sabinone (45, 46). A similar pathway for sabinone formation en route to thujone was described for western redcedar (47). The biosynthetic pathway to camphor is realized by hydroxylation of bornyl diphosphate to bornol and subsequent oxidation to camphor in sage, tansy, and ginger (Zingiber zerumbet) (46, 48–50). These examples illustrate that P450-mediated hydroxylation are commonly followed by ketone formation through dehydrogenation in biosynthetic pathways leading to derivatives of oxygenated monoterpene, among various plant species.

Similar reaction patterns have been found for the formation of sesquiterpene derivatives. Artemisinin, a pharmacologically very important sesquiterpenoid lactone, is formed via hydroxyl- ation of amorpha-4,11-diene by CYP71AV1 to artemisinic.
alcohol, which is then oxidized to artemisinc alcohol by an alcohol dehydrogenase (7). The first part of the biosynthesis of costunolide in chichory was found to comprise the hydroxylation of sesquiterpene germacrene A to germacrene alcohol and subsequent oxidation of the hydroxyl group to the respective ketone (germacrene aldehyde). In this case, both oxidation steps were shown to be mediated by the P450 enzyme germacrene A oxidase from the CYP71 family (51).

In the present study, the CYP71D enzymes all produced p-cymene in the in vitro assay in absence of the dehydrogenase, likely through dehydrogenation of the initially formed enzyme product, an unstable cyclohexadienol (Fig. 4). Yet p-cymene is a frequent constituent of plant essential oils, including those of phenolic monoterpene-accumulating plants, such as thyme and oregano. Therefore, the ratio of p-cymene versus thymol or carvacrol in the plant may depend on the relative activities of the CYP71D and SDR enzymes and the physical distance between the two proteins in the cell. If the dehydrogenase is closely associated with the cytochrome P450s, there would be less opportunity for the dienol to rearrange to p-cymene. To facilitate the efficient transport of the unstable dienol intermediate and prevent its release into the cytosol, spatial proximity of the CYP71D and SDR enzymes and the physical distance between the two proteins in the cell. If the dehydrogenase is closely associated with the cytochrome P450s, there would be less opportunity for the dienol to rearrange to p-cymene. To facilitate the efficient transport of the unstable dienol intermediate and prevent its release into the cytosol, spatial proximity of the CYP71D and SDR enzymes and the physical distance between these enzymes. However, p-cymene may not only be formed by spontaneous rearrangement of the CYP71D cyclohexadienol product but also directly by the CYP71D enzymes characterized here, because TvCYP71DS07 and TvCYP1D179 produced higher concentrations of p-cymene in the presence of TvSDR1 than the empty vector (Fig. 2E).

Thymohydroquinone Is Formed by the Subsequent Oxidation of Thymol and Carvacrol. While thymol and carvacrol are formed by the selective oxygenation at C-3 or C-6, we did not observe such a specificity for the subsequent hydroxylation of thymol or carvacrol to thymohydroquinone by OvCYP76S40 or OvCYP736A300. The two enzymes are structurally diverse and fall into different groups of CYPs involved in secondary metabolism (SI Appendix, Fig. S2). Nevertheless, both hydroxylate at C-3 or C-6, the position opposite to the existing hydroxyl group on the aromatic ring, to form the hydroquinone. Although these enzymes catalyze both thymol and carvacrol substrates in vitro and in vivo after heterologous expression in N. benthamiana, we cannot exclude minor differences in substrate preference. When these enzymes were expressed in yeast, thymoquinone was detected as a minor product (Fig. 5B). Thymoquinone could be a genuine side product of the enzyme itself but could also arise from spontaneous conversion of thymohydroquinone, as we observed with a thymohydroquinone standard at different temperatures over times ranging from 1 to 7 d (SI Appendix, Fig. S15). Therefore, we assume that at least some of the thymoquinone observed was generated nonenzymatically from the relatively large quantities of thymohydroquinone produced by the yeast extracts and likely in planta. Alternatively, the conversion to thymoquinone could be catalyzed by alcohol dehydrogenase activities or other CYPs in the plant. The almost 10-fold difference in thymoquinone to thymohydroquinone ratio, observed among the O. vulgare and T. vulgaris varieties (Fig. 3A and SI Appendix, Fig. S14), suggests that the conversion of thymohydroquinone to thymoquinone in the glandular trichomes of these plants depends on more than one mechanism.
D. J. McGarvey, R. Croteau, Terpenoid metabolism.


B. Tohidí, M. Rahimímalek, A. Azarani, H. Trindade, Sequencing and variation of terpene synthase gene (TPS2) as the major gene in biosynthesis of thymol in different Thymus species. Phytochemistry 169, 112162 (2020).


I. Botinick et al., Distribution of primary and specialized metabolites in Nigella sativa seeds, a spice with vast traditional and historical uses. Molecules 17, 10159–10177 (2012).

N. Borróss-Padilla et al., Genome sequencing of four culinary herbs reveals terpenoid genes underlying chemodiversity in the Nepetoideae. DNA Res. 27, dsoa016 (2020).


