Determination of Residues Responsible for Substrate and Product Specificity of *Solanum habrochaites* Short-Chain cis-Prenyltransferases

Jin-Ho Kang\(^2\), Eliana Gonzales-Vigil\(^2,3\), Yuki Matsuba, Eran Pichersky, and Cornelius S. Barry*  
Department of Horticulture, Michigan State University, East Lansing, Michigan 48824 (J.-H.K., E.G.-V., C.S.B.); and Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, Michigan 48109 (Y.M., E.P.)

Isoprenoids are diverse compounds that have their biosynthetic origin in the initial condensation of isopentenyl diphosphate and dimethylallyl diphosphate to form C10 prenyl diphosphates that can be elongated by the addition of subsequent isopentenyl diphosphate units. These reactions are catalyzed by either cis-prenyltransferases (CPTs) or trans-prenyltransferases. The synthesis of volatile terpenes in plants typically proceeds through either geranyl diphosphate (C10) or trans-farnesyl diphosphate (C15), to yield monoterpenes and sesquiterpenes, respectively. However, terpene biosynthesis in glandular trichomes of tomato (*Solanum lycopersicum*) and related wild relatives also occurs via the cis-substrates neryl diphosphate (NPP) and Z,Z-farnesyl diphosphate (Z,Z-FPP). NPP and Z,Z-FPP are synthesized by neryl diphosphate synthase1 (NDPS1) and Z,Z-farnesyl diphosphate synthase (zFPS), which are encoded by the orthologous CPT1 locus in tomato and *Solanum habrochaites*, respectively. In this study, comparative sequence analysis of NDPS1 and zFPS enzymes from *S. habrochaites* accessions that synthesize either monoterpenes or sesquiterpenes was performed to identify amino acid residues that correlate with the ability to synthesize NPP or Z,Z-FPP. Subsequent structural modeling, coupled with site-directed mutagenesis, highlighted the importance of four amino acids located within conserved domain II of CPT enzymes that form part of the second α-helix, for determining substrate and product specificity of these enzymes. In particular, the relative positioning of aromatic amino acid residues at positions 100 and 107 determines the ability of these enzymes to synthesize NPP or Z,Z-FPP. This study provides insight into the biochemical evolution of terpene biosynthesis in the glandular trichomes of *Solanum* species.

Plant isoprenoids are structurally heterogeneous, giving rise to pigments, hormones, quinones, and sterols, together with a variety of specialized metabolites that are often restricted to specific genera or families (Kirby and Keasling, 2009). Among these specialized metabolites are volatile short-chain terpenoids with chain lengths of C10, C15, and C20 defined as monoterpenes, sesquiterpenes, and diterpenes, respectively. Subsequently, these hydrocarbons can be modified by cleavage, oxidation, acylation, or glycosylation to yield diverse volatile and nonvolatile compounds with varied activities (Richman et al., 2005; Heiling et al., 2010; Lee et al., 2010; Nguyen et al., 2010; Lange and Turner, 2013). Terpenoids possess multiple roles in plants, contributing to floral scents that attract pollinators and aroma compounds that facilitate seed dispersal (Aharoni et al., 2004; Gang, 2005; Dudareva et al., 2013). In addition, terpenoids can be toxic or act as repellants or signaling molecules that deter or limit herbivory (Schnee et al., 2006; Degenhardt et al., 2009; Bleeker et al., 2011; Schmelz et al., 2011). The chemical properties of terpenoids have led to their exploitation by humans as flavorings, fragrances, medicines, and biofuels, with considerable interest focused in engineering their biosynthesis to improve production efficiencies (Wu et al., 2006; Kirby and Keasling, 2009; Niehaus et al., 2011; Zhang et al., 2011; Westfall et al., 2012; Lange and Ahkami, 2013; Yeo et al., 2013).

The C5 compounds isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) serve as the precursors of isoprenoids and can be formed through either the plastid-localized 2-C-methyl-D-erythritol 4-phosphate pathway or the cytosolic mevalonate pathway (Rodríguez-Concepción and Boronat, 2002; Eisenreich et al., 2004; Concepción and Boronat, 2002; Eisenreich et al., 2004; Chung et al., 2011). Condensation of IPP and DMAPP to form either geranyl diphosphate (GPP) in the trans-configuration or neryl diphosphate (NPP) in the cis-configuration represents the initial step of isoprenoid biosynthesis, and these reactions are catalyzed by trans-prenyltransferases (TPTs) or cis-prenyltransferases (CPTs), respectively (Burke et al., 1999; Oh et al., 2000; Schilmiller et al., 2010; Lange and Turner, 2013).

![Web-only data](https://www.plantphysiol.org/cgi/doi/10.1104/pp.113.230466)
et al., 2009; Akhtar et al., 2013). Successive head-to-tail additions of IPP units in either cis- or trans-configuration determine the final chain length of isoprenoids, which typically ranges between C15 and C120 but can be greater than C10,000 in the case of natural rubber (Hevea brasiliensis; Takahashi and Koyama, 2006).

TPTs and CPTs are distinct and operate through different catalytic mechanisms (Takahashi and Koyama, 2006; Liang, 2009). Several plant TPTs have been characterized, including those involved in the synthesis of GPP, 2E,6E-farnesyl diphosphate, geranylgeranyl diphosphate (C20), and solanesyl diphosphate (C45), that serve as the precursors of terpenes, sterols, hormones, and carotenoids (Burke et al., 1999; Hirooka et al., 2003; Ducluzeau et al., 2012). Similarly, plant CPTs involved in polyisoprenoid biosynthesis have been identified, including those involved in the synthesis of rubber (Cunillera et al., 2000; Asawatreratanakul et al., 2003; Schmidt et al., 2010; Surmacz and Swiezewska, 2011; Kera et al., 2012; Post et al., 2012). Recently, characterization of the seven-member CPT family of tomato (Solanum lycopersicum) led to the identification of four SICPT enzymes involved in the synthesis of polyisoprenoids, including SICPT3 that rescues a dolichol-deficient mutant of yeast (Saccharomyces cerevisiae) (Akhtar et al., 2013). A further three enzymes, SICPT1 (neryl diphosphate synthase1 [NDFS1]), SICPT2, and SICPT6, synthesize the short chain cis-prenyl diphosphates NPP, nerylneryl diphosphate, and 2Z,6Z-farnesyl diphosphate (Z,Z-FPP), respectively, and each gene is associated with a terpene synthase gene cluster (Schilmiller et al., 2009; Falara et al., 2011; Akhtar et al., 2013; Matsuba et al., 2013).

Plants typically synthesize volatile terpenes from either GPP or 2E,6E-farnesyl diphosphate, giving rise to diverse monoterpenes and sesquiterpenes, respectively (Chen et al., 2011). However, type VI glandular trichomes of tomato and closely related species synthesize monoterpenes, and each gene is associated with a terpene synthase locus (Gonzales-Vigil et al., 2012; Matsuba et al., 2013). The presence of monoterpene synthases that use NPP as their substrate suggested the existence in within this cluster, suggesting that their different activities arose through sequence divergence alone (van der Hoeven et al., 2000; Sallaud et al., 2009; Matsuba et al., 2013).

Metabolite profiling of 79 S. habrochaites accessions revealed considerable intraspecific qualitative and quantitative variation in terpene synthesis, leading to the subsequent isolation of a second sesquiterpene synthase, zingiberene synthase (ShZIS), which is an allele of the TPS45 locus and encodes an enzyme that also utilizes Z,Z-FPP as a substrate (Gonzales-Vigil et al., 2012; Matsuba et al., 2013). In addition, three sequences that encode monoterpene synthases (ShPHS1, pinene synthase, and limonene synthase) that utilize NPP to synthesize β-phellandrene, α-pinene, and limonene, respectively, were also identified and shown to be alleles of the TPS20 locus (Gonzales-Vigil et al., 2012; Matsuba et al., 2013). The presence of monoterpene synthases that use NPP as their substrate suggested the existence in S. habrochaites accessions of an NDPS1 enzyme rather than zFPS. In this study, complementary DNAs (cDNAs) corresponding to the CPTI locus were isolated from the trichomes of chemically diverse S. habrochaites accessions, and comparative sequence analysis together with homology modeling were employed to identify specific amino acid residues that correlate with either NDPS1 or

---

**Figure 1.** Pathway for the synthesis of short-chain cis-prenyl diphosphates in the trichomes of Solanum species. The synthesis of NPP occurs through the single head-to-tail condensation of DMAPP and IPP and is catalyzed by NDPS1. The synthesis of Z,Z-FPP is catalyzed by zFPS and can proceed through two routes, either directly via DMAPP through successive head-to-tail condensations with two molecules of IPP or through the single condensation of NPP and IPP. OPP refers to the pyrophosphate group on each molecule.
zFPS activity. The role of these residues in determining substrate and product specificity of short-chain CPTs was confirmed through site-directed mutagenesis, revealing an essential role for the relative positions of aromatic amino acids within a hydrophobic cleft between helices II and III. These data provide insight into the biochemical evolution of trichome-derived specialized metabolites that function as insect deterrents.

RESULTS

Identification and Characterization of NDPS1 from *S. habrochaites*

To identify sequences encoding NDPS1 enzymes from *S. habrochaites* and to begin to address the evolutionary relationship of NDPS1 and zFPS, cDNAs from the CPT1 locus were amplified and sequenced from multiple accessions of *S. habrochaites* that predominantly synthesize either monoterpenes or sesquiterpenes from the cis-pathway. In total, 25 cDNAs were recovered from 17 accessions that share between 97% and 100% identity at the nucleotide level. Phylogenetic analysis revealed that these sequences group into two major clades (Fig. 2). One clade is defined by the presence of the previously characterized zFPS gene from *S. habrochaites* accession LA1777 (Sallaud et al., 2009), and the second, containing fewer sequences, is more closely related to NDPS1 from tomato (Schilmiller et al., 2009).

The phylogeny of the *S. habrochaites* CPT1 sequences suggests that several may encode enzymes with NDPS1 activity, and this hypothesis was tested by expressing a codon-optimized synthetic version of CPT1 from *S. habrochaites* accession LA2409, a monoterpene-producing accession (Gonzales-Vigil et al., 2012), in *Escherichia coli* and assaying the activity of the purified recombinant enzyme. Upon incubation with DMAPP and IPP at equimolar concentrations, the recombinant enzyme synthesized the C10 isoprenoid NPP that was detected as nerol following dephosphorylation, indicating that the enzyme possesses NDPS1 activity (Fig. 3A). Furthermore, as the ratio of IPP to DMAPP was increased, the product specificity of the enzyme remained unchanged. In contrast, purified recombinant enzyme derived from a codon-optimized version of CPT1 from *S. habrochaites* accession LA1393, a sesquiterpene-producing accession that is identical to the previously characterized zFPS enzyme from *S. habrochaites* accession LA1777 (Sallaud et al., 2009; Gonzales-Vigil et al., 2012), preferentially synthesized the C15 isoprenoid NPP that was detected as *Z*,*Z*-FPP, indicating that the enzyme possesses NDPS1 activity (Fig. 3A). Furthermore, as the ratio of IPP to DMAPP was increased, the product specificity of the enzyme remained unchanged. In contrast, purified recombinant enzyme derived from a codon-optimized version of CPT1 from *S. habrochaites* accession LA1393, a sesquiterpene-producing accession that is identical to the previously characterized zFPS enzyme from *S. habrochaites* accession LA1777 (Sallaud et al., 2009; Gonzales-Vigil et al., 2012), preferentially synthesized the C15 product, *Z*,*Z*-FPP (Fig. 3A). The activities of CPT1 from LA2409 and LA1393 were also determined utilizing NPP and IPP as substrates. As previously documented, zFPS is able to synthesize *Z*,*Z*-FPP from NPP and IPP (Sallaud et al., 2009), although CPT1 from LA2409 is unable to utilize this substrate combination (Fig. 3B). These data indicate that CPT1 from *S. habrochaites* accession LA2409 encodes an enzyme with NDPS1 activity.

Identification of Amino Acid Residues Important for the Activity of NDPS1 and zFPS

The predicted protein sequences of NDPS1 encoded by CPT1 from accession LA2409 and zFPS encoded by CPT1 from accession LA1393 are highly similar to one another, sharing 94% identity and 97% similarity over the entire length of their sequences, which corresponds to 18 amino acid differences. A comparative sequence approach was utilized to further refine the identity of the amino acids responsible for specifying either NDPS1 or zFPS activity. Amino acid alignments of putative NDPS1 and zFPS proteins of *S. habrochaites* were investigated for conserved residues that correlate with the ability of the parent accession to synthesize either monoterpenes or sesquiterpenes (Fig. 2) and, in the
case of putative *S. habrochaites* NDPS1 proteins, that also possess an identical residue at the corresponding position in tomato NDPS1. In total, 11 residues that reside within the predicted mature protein of NDPS1 and zFPS matched these criteria, and these were generally spaced throughout the protein sequence, but several were located within conserved domains II, III, and V of CPTs (Table I; Supplemental Fig. S1).

This analysis was followed by changing 10 residues that correlate with changes in substrate in NDPS1 from accession LA2409 to those found in zFPS from accession LA1393 (Pro/Ser-45, which corresponds to the second residue in the predicted mature protein, was excluded), and we made the reciprocal changes in zFPS (LA1393). A, The synthesis of nerol (C10) and farnesol (C15) using various ratios of IPP and DMAPP. B, The synthesis of farnesol (C15) using various ratios of IPP and NPP. Experimental details are provided in “Materials and Methods.” Each data point represents the mean ± SD of two or three replicates.

![Graph A](image1.png)  
**Figure 3.** In vitro enzymatic assay of CPT1 from *S. habrochaites* accessions LA2409 and LA1393. A, The synthesis of nerol (C10) and farnesol (C15) using various ratios of IPP and DMAPP. B, The synthesis of farnesol (C15) using various ratios of IPP and NPP. Experimental details are provided in “Materials and Methods.” Each data point represents the mean ± SD of two or three replicates.

The two resulting recombinant proteins were named NDPS1-M10 variants were determined and compared with the activities of the two parent enzymes. An IPP:DMAPP ratio of 5:1 was utilized for these experiments, as it corresponds to the typical ratio synthesized by 4-hydroxy-3-methyl-2-(E)-butenyl-diphosphate reductase (Rohdich et al., 2002). The NDPS1-M10 variant was able to utilize DMAPP and IPP to synthesize Z,Z-FPP at levels that are similar to that of the zFPS enzyme from accession LA1393 (Fig. 4A). The ability of NDPS1-M10 to synthesize Z,Z-FPP coincided with a greatly reduced ability to synthesize NPP. Similarly, the zFPS-M10 recombinant protein lost the ability to synthesize Z,Z-FPP but gained NDPS1 activity and synthesized NPP at levels similar to NDPS1 from accession LA2409 (Fig. 4A). Furthermore, substitution of the M10 residues also changed the substrate specificity of NDPS1 and zFPS with respect to their ability to utilize NPP. For example, the NDPS1-M10 variant gained the ability to synthesize Z,Z-FPP from NPP and IPP, whereas this activity was lost in the zFPS-M10 variant (Fig. 4B). Together, these data illustrate that substitution of the 10 conserved amino acids influenced the substrate and product specificities of trichome-expressed short-chain CPTs and were sufficient to interchange NDPS1 and zFPS activities.

**Homology Modeling of NDPS1 and zFPS**

To attempt to refine which subset of the M10 amino acids determine the substrate and product specificities of NDPS1 and zFPS, homology models were generated to visualize the relative positions of these amino acids within three-dimensional space (Fig. 5). Currently, there are no available crystal structures of plant CPTs. However, BLAST searches of the Protein Data Bank (PDB) database revealed that NDPS1 and zFPS, without the chloroplast transit peptide, each shares approximately 40% amino acid identity with an *E. coli* undecaprenyl pyrophosphate synthase (PDB no. 1X07A), which catalyzes the formation of a C55 undecaprenyl pyrophosphate involved in peptidoglycan synthesis in the bacterial cell wall (Guo et al., 2005). Homology models of NDPS1 and zFPS were obtained using the SWISS-MODEL Workspace with QMEAN z-score values of −2.41 and −2.20 for NDPS1 and zFPS, respectively, which represent acceptable models.

In general, the models of NDPS1 and zFPS aligned to the template with reasonable overlap of the predicted helices observed (Fig. 5, A and B). However, the third α-helix of both NDPS1 and zFPS is considerably shorter than that observed in the *E. coli* undecaprenyl pyrophosphate synthase, and instead of an extended helix structure, the sequences of both NDPS1 and zFPS form disordered loops. The dimension of the hydrophobic cleft formed between helices II and III is known to influence the product chain length of CPTs (Kharel et al., 2006; Noike et al., 2008). Recently, six additional tomato CPTs (SlCPT2–SlCPT7) were identified and characterized that synthesize prenyl diphosphates ranging between C15 and C65 in length (Akhtar et al., 2013). The enzymes that synthesize the shorter chain length products lack amino acid residues within a variable domain downstream of the conserved domain III. Homology models with acceptable QMEAN z-score values ranging between −1.52 and −3.44 were generated between five additional tomato CPT enzymes (SlCPT2 and SlCPT4–SlCPT7) and various CPT templates (Supplemental Fig. S2). A reliable model was not obtained for SlCPT3. Structure homology models of SlCPT4 and SlCPT5, which synthesize C55 and C60 products (Akhtar et al., 2013), possess a complete third α-helix...
that resembles that of the *E. coli* undecaprenyl pyrophosphate synthase, whereas SlCPT2, SlCPT6, and SlCPT7, which synthesize short-chain prenyl diphosphates of between C15 and C35 in length, lack a complete third \( \alpha \)-helix (Supplemental Fig. S2). The incomplete \( \alpha \)-helix in NDPS1, SlCPT2, SlCPT6, and SlCPT7 is the result of a shorter intervening sequence downstream of CPT conserved domain III (Akhtar et al., 2013; Fig. 5; Supplemental Fig. S2). Together, these data suggest a correlation between the length of the third \( \alpha \)-helix in the tomato CPTs, the size of the resulting hydrophobic cleft between the second and third helices, and the product chain length.

The structural models of NDPS1 and zFPS are similar to each other, with no difference in the length of the third \( \alpha \)-helix predicted (Fig. 5, A and B), suggesting that the different catalytic activities of these enzymes is unlikely to be caused by major structural changes within this region. However, several of the divergent amino acids between NDPS1 and zFPS are located within CPT conserved domains II and III, which lie within or adjacent to either the second or third \( \alpha \)-helix (Fig. 5; Table I). In particular, the relative positions of the aromatic amino acids corresponding to Tyr-100 in NDPS1 and Phe-107 in zFPS, which are located within the second \( \alpha \)-helix within CPT conserved domain II, differ substantially between the two proteins (Fig. 5; Table I). Notably, the presence of bulky amino acids within the hydrophobic cleft between helices II and III is known to influence the product chain length of CPTs (Kharel et al., 2006; Noike et al., 2008), suggesting that the relative positions of NDPS1 Tyr-100 and zFPS Phe-107 may impact the substrate and product specificities of these enzymes.

### The Relative Positions of Residues within Region II Determine the Substrate and Product Specificities of NDPS1 and zFPS

As the domain II residues, and particularly the relative positions of Tyr-100 and Phe-107, constitute the major differences between the NDPS1 and zFPS structural models (Fig. 5), we converted the four conserved amino acids within domain II, Glu-98, Tyr-100, Ile-106, and Ile-107, of NDPS1 to those present in zFPS, generating NDPS1-M4 (Fig. 6A). We also converted these four domain II amino acids in zFPS-M4 to the four residues from NDPS1. Incubation of zFPS-M4 with DMAPP and IPP led to the synthesis of NPP at a similar amount to that observed for NDPS1 from LA2409 (Fig. 6B).

![Figure 4](https://www.plantphysiol.org/)

Figure 4. In vitro enzymatic assay of M10 mutant variants of NDPS1 and zFPS from *S. habrochaites*. A, The synthesis of nerol (C10) and farnesol (C15) by the wild type (WT) and M10 mutant variants of NDPS1 and zFPS at a 5:1 ratio of IPP to DMAPP. B, The synthesis of farnesol (C15) by the wild type and M10 mutant variants of NDPS1 and zFPS at a 5:1 ratio of IPP to NPP. Experimental details are provided in “Materials and Methods.” Each data point represents the mean ± SD of two or three replicates.
Furthermore, recombinant zfPS-M4 protein was unable to utilize NPP as a substrate (Fig. 6C). Conversely, the NDPS1-M4 recombinant enzyme displays catalytic properties that are similar to zfPS and is able to utilize both DMAPP + IPP and NPP + IPP to synthesize Z,Z-FPP. Together, these data indicate that the residues within conserved domain II of NDPS1 and zfPS, which lie within the second α-helix, are important for determining the substrate and product specificity of these short-chain CPTs.

To test the hypothesis that the relative positions of Tyr-100 in NDPS1 and Phe-107 in zfPS are responsible for determining their activities, site-directed mutagenesis was performed to generate mutant versions of each enzyme in which the positions of these aromatic amino acids were switched and replaced with those present in the other enzyme (Fig. 6A). The resultant NDPS1 Tyr-100Ser, Ile-107Phe and zfPS Ser-100Tyr, Phe-107Ile constructs were designated NDPS1-M2 and zfPS-M2, respectively. The activity of each of the M2 constructs was determined using both DMAPP + IPP and NPP + IPP as substrates. Whereas the NDPS1-M4 construct gained the activity typically associated with the NDPS1-M2 and zfPS-M2 constructs, the activity of each of the M2 constructs was determined using both DMAPP + IPP and NPP + IPP as substrates. Whereas the NDPS1-M4 construct gained the activity typically associated with zfPS, the M2 construct synthesized a mixture of C10 and C15 prenyl diphosphates when using DMAPP and IPP as substrates (Fig. 6B). In addition, while the NDPS1-M2 enzyme was able to utilize NPP as a substrate to synthesize Z,Z-FPP, this reaction was less efficient than was observed with the NDPS1-M4 and zfPS enzymes (Fig. 6C). In contrast, the activity of the zfPS-M2 enzyme was similar to that of zfPS-M4 and mirrored that of the wild-type NDPS1 enzyme (Fig. 6). However, the zfPS-M2 enzyme also retained a residual level of activity when supplied with NPP and IPP, although this was minor when compared with the activity of zfPS from LA1393 (Fig. 6C). In combination, these data indicate that while the relative positioning of the bulky aromatic amino acids between the predicted second and third helices of these short-chain CPTs are sufficient to convert the substrate and product specificities of zfPS to those of NDPS1, they are insufficient to confer the complete reciprocal change to zfPS.

The dimensions and structure of the hydrophobic cleft of CPTs may be particularly important for specifying enzyme activity when multiple rounds of chain elongation occur, as is the case when Z,Z-FPP is synthesized from DMAPP and IPP. As such, the amino acids immediately flanking the aromatic residues within conserved domain II could be important for determining the CPT activity through modifying their spatial orientation. To examine the role of the additional residues within the hydrophobic cleft in more detail, two additional mutants of NDPS1, NDPS1-M3V1 (Tyr-100Ser, Ile-106Leu, Ile-107Phe) and NDPS1-M3V2 (Glu-98Asp, Tyr-100Ser, Ile-107Phe), were constructed (Fig. 6A). In particular, we were interested to determine whether the Ile-106Leu substitution, which lies immediately adjacent to Phe-107 in the NDPS1-M4 and NDPS1-M2 constructs, influences the activity of these recombinant enzymes.

The activity of the resulting NDPS1-M3V1 recombinant enzyme more closely resembled zfPS and produced predominantly Z,Z-FPP, but it still produced NPP as approximately 10% of the total product, whereas the NDPS1-M3V2 enzyme produced NPP and Z,Z-FPP in approximately equal proportions (Fig. 6B). Similarly, the NDPS1-M3V1 and NDPS1-M3V2 enzymes utilized NPP and IPP as substrates at similar relative proportions to those observed when using DMAPP and IPP as substrates (Fig. 6C). Furthermore, as anticipated from the activities of the zfPS-M4 and zfPS-M2 enzymes, both zfPS-M3V1 (Asp-98Asp, Ser-100Tyr, Phe-107Ile) and zfPS-M3V2 (Ser-100Tyr, Leu-106Ile, Phe-107Ile) enzymes synthesized NPP when supplied with DMAPP and IPP as substrates and essentially lost the ability to efficiently utilize NPP as a substrate (Fig. 6, B and C). Overall, these data indicate that substitution of the aromatic amino acids at positions 100 and 107 is sufficient to convert zfPS into NDPS1, but the complete reciprocal change in...
activity from NDPS1 to zFPS requires additional amino acid substitutions at positions 98 and 106.

**Kinetic Analysis of Mutant Variants of NDPS1 and zFPS**

The activities of mutant variants of NDPS1 and zFPS revealed qualitative changes in the synthesis of prenyl diphosphates that are consistent with a role of amino acid residues within CPT conserved region II as determinants of substrate and product specificity (Figs. 4 and 6). Furthermore, the majority of the mutant enzymes synthesize prenyl diphosphates at abundances similar to those of the corresponding parent enzyme. For example, NDPS1 and zFPS-M10 synthesize approximately the same amounts of NPP when supplied with DMAPP and IPP as substrates. However, it is possible that the kinetic properties of the mutant enzyme variants differ from the wild-type enzymes, and these differences may not be apparent if substrate concentrations are limiting. Therefore, kinetic analysis was performed on the wild type and select mutant variants of both NDPS1 and zFPS.

Kinetic analysis of NDPS1 from LA2409 revealed $K_m$ values for DMAPP and IPP of 112 and 140 μM, respectively (Table II), which are similar to the values of 177 and 152 μM previously reported for tomato NDPS1 (Schilmiller et al., 2009). The $K_m$ values of the M10 and M2 variants of zFPS using DMAPP as a substrate are similar to that obtained for NDPS1, although the turnover number ($k_{cat}$) and subsequent catalytic efficiency of the zFPS-M2 enzyme are slightly lower than those observed for NDPS1. The $K_m$ for IPP of zFPS-M2 is lower than those measured for the NDPS1 and zFPS-M10 enzymes, although the catalytic efficiencies of each enzyme are similar (Table II). The $K_m$ of zFPS from LA1393 varied slightly when compared with previously published data (Sallaud et al., 2009). The $K_m$ values for DMAPP and NPP were 76 ± 12 and 51 ± 9 μM (Table II), respectively, compared with 35 ± 5 and 10 ± 2 μM from the previous study. In addition, the $K_m$ for IPP, when measured using NPP as a substrate, was 147 ± 59 μM. The kinetic properties of the NDPS1-M10 and NDPS1-M4 enzymes were also determined and compared with those of the wild-type zFPS enzyme. The $K_m$ for each substrate was comparable for all three enzymes, and although the catalytic efficiencies of the NDPS1-M10 enzyme are slightly lower than those calculated for the wild-type zFPS enzyme from LA1393, the values for the NDPS1-M4 enzyme are very similar to those of the wild-type zFPS enzyme (Table II).

**DISCUSSION**

Tomato and closely related wild species exhibit both qualitative and quantitative variation in trichome-derived volatile terpenes, several of which are synthesized by enzymes encoded within a rapidly evolving gene cluster on the short arm of chromosome 8 that is characterized by gene duplications, deletions, rearrangements, and conversion (Matsuba et al., 2013). The TPS genes within the chromosome 8 terpene cluster are atypical in that they encode enzymes that use cis-prenyl diphosphates of C10, C15, or C20 chain lengths that are synthesized by CPTs encoded by genes also present within this cluster (Sallaud et al., 2009; Schilmiller et al., 2009; Gonzalez-Vigil et al., 2012; Matsuba et al., 2013).

Within genus *Solanum*, the capacity to synthesize monoterpenes from NPP appears to be present in multiple
species and accessions, including some *S. habrochaites* accessions, and is associated with the CPT1 locus that encodes NDPS1 and the *TPS19* and *TPS20* loci that encode monoterpene synthases that use NPP (Schilmiller et al., 2009; Falara et al., 2011; Gonzales-Vigil et al., 2012; Matsuba et al., 2013). However, the ability to synthesize sesquiterpenes from *Z,Z*-FPP is thus far restricted to specific accessions of *S. habrochaites* and involves both the evolution of the *TPS45* locus (whose alleles encode either ShSBS or ShZIS) and the evolution of an allele at the CPT1 locus that encodes zFPS activity (Sallaud et al., 2009; Gonzales-Vigil et al., 2007; Xu et al., 2007; Keeling et al., 2008; Green et al., 2011).

The chain length of isoprenoids varies dramatically, from C10 in the case of monoterpene to more than C10,000 for natural rubber, and this wide range of diverse products is catalyzed by prenyltransferases with different substrate and product specificities (Takahashi and Koyama, 2006). In the case of the prenyltransferases, which are structurally and mechanistically distinct from the cis-prenyltransferases, bulky aromatic residues at the interface between two α-helices are known to determine product chain length, and a combination of crystallography, structural modeling, bioinformatics, and biochemical analyses has led to the development and validation of predictive models of their functional properties (Ohnuma et al., 1996, 1997, 1998; Tarshis et al., 1996; Wallrapp et al., 2013).

Although less well defined, determinants of substrate and product specificity of cis-prenyltransferases have also been reported. These studies, in which mutagenesis of specific residues and the incorporation of additional residues at the flexible loop adjacent to domain III allowed products of longer chain length to be synthesized, highlighted the importance of the number of amino acid residues within the third α-helix and residues within conserved domain III for the determination of product chain length (Kharel et al., 2006; Noike et al., 2008). For example, Ala substitution of bulky amino acids such as Leu-84 in the *Z,E*-farnesyl diphosphate synthase of *Mycobacterium tuberculosis* allows an increase in product chain length over the wild-type enzyme (Noike et al., 2008). In the structural models of CPTs (Fig. 5), domain III residues form the third α-helix and the preceding disordered loop (Fig. 5C), which together with the second α-helix forms the hydrophobic cleft close to the substrate-binding sites where chain elongation takes place.

### Table II. Kinetic analysis of NDPS1 and zFPS mutants

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Substrate</th>
<th>$k_m$</th>
<th>$k_{cat}$</th>
<th>$k_{cat}/k_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDPS1 (LA2409)</td>
<td>IPP (DMPAPP, 150 μM)</td>
<td>140 ± 11 (4.7 ± 0.5) × 10⁻³</td>
<td>(3.3 ± 0.1) × 10⁻⁴</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DMAPP (IPP, 50 μM)</td>
<td>112 ± 4 (2.6 ± 0.3) × 10⁻³</td>
<td>(2.3 ± 0.1) × 10⁻⁴</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DMAPP (IPP, 150 μM)</td>
<td>228 ± 43 (6.6 ± 1.2) × 10⁻³</td>
<td>(3.0 ± 0.7) × 10⁻⁴</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DMAPP (IPP, 50 μM)</td>
<td>160 ± 17 (2.2 ± 0.2) × 10⁻³</td>
<td>(1.4 ± 0.1) × 10⁻⁴</td>
<td></td>
</tr>
<tr>
<td>zFPS-M10</td>
<td>IPP (DMPAPP, 150 μM)</td>
<td>37 ± 5 (1.5 ± 0.6) × 10⁻³</td>
<td>(3.1 ± 0.6) × 10⁻⁴</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DMAPP (IPP, 50 μM)</td>
<td>108 ± 25 (1.0 ± 0.2) × 10⁻²</td>
<td>(1.0 ± 0.1) × 10⁻⁴</td>
<td></td>
</tr>
<tr>
<td>zFPS (LA1393)</td>
<td>IPP (NPP, 150 μM)</td>
<td>147 ± 59 (1.9 ± 1.1) × 10⁻¹</td>
<td>(1.3 ± 0.6) × 10⁻³</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NPP (IPP, 50 μM)</td>
<td>51 ± 9 (6.9 ± 1.6) × 10⁻⁴</td>
<td>(1.3 ± 0.1) × 10⁻³</td>
<td></td>
</tr>
<tr>
<td>NDPS1-M10</td>
<td>DMAPP (IPP, 50 μM)</td>
<td>47 ± 9 (6.4 ± 0.7) × 10⁻³</td>
<td>(1.5 ± 0.1) × 10⁻⁴</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IPP (NPP, 150 μM)</td>
<td>157 ± 16 (1.1 ± 0.4) × 10⁻⁴</td>
<td>(7.3 ± 0.5) × 10⁻³</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NPP (IPP, 50 μM)</td>
<td>14 ± 2 (3.7 ± 0.8) × 10⁻⁳</td>
<td>(2.8 ± 0.9) × 10⁻⁴</td>
<td></td>
</tr>
<tr>
<td>NDPS1-M4</td>
<td>DMAPP (IPP, 50 μM)</td>
<td>54 ± 3 (8.8 ± 3.0) × 10⁻⁷</td>
<td>(1.6 ± 0.6) × 10⁻⁴</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IPP (NPP, 150 μM)</td>
<td>111 ± 50 (1.6 ± 0.9) × 10⁻¹</td>
<td>(1.8 ± 1.3) × 10⁻⁴</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NPP (IPP, 50 μM)</td>
<td>97 ± 9 (1.8 ± 0.7) × 10⁻⁴</td>
<td>(1.8 ± 0.1) × 10⁻⁴</td>
<td></td>
</tr>
</tbody>
</table>
occurs (Guo et al., 2005; Noike et al., 2008). The Solanum short-chain CPTs encoded by CPT1, CPT2, and CPT6 lack amino acid residues downstream of CPT domain III (Akhtar et al., 2013), which shortens the length of the third α-helix in these proteins, compared with those of CPT4, CPT5, and CPT7 that catalyze the formation of longer chain cis-prenyl diphosphates (Akhtar et al., 2013; Supplemental Fig. S2). Therefore, a correlation exists between the length of the third α-helix and the product chain length. However, this correlation does not explain the difference in the product chain length of NDPS1 and zFPS, and our data indicate a heretofore unknown role for the relative positioning of aromatic residues within conserved domain II, which is part of the second α-helix (Fig. 5), as important for specifying the relative activities of these short-chain CPTs. Further experiments are required to determine whether this region has broad utility across diverse CPTs.

In general, the functions and evolutionary relationships of plant CPTs, together with the structural features that determine their substrate and product specificities, remain poorly understood. By utilizing an integrated approach involving comparative sequence analysis of chemically distinct germplasm, homology modeling, and site-directed mutagenesis, coupled with enzyme activity assays and kinetics, we have identified and characterized allelic diversity at the CPT1 locus within a natural population of S. habrochaites. This allelic diversity causes variation between NDPS1 and zFPS through altering the relative positions of aromatic amino acids within a structurally conserved hydrophobic cleft that is important for the product chain length determination of CPTs. This variation led to the formation of Z,Z-FPP in the chloroplasts of type VI glandular trichomes, which represents an essential step in the evolution of sesquiterpene biosynthesis.

MATERIALS AND METHODS

Plant Material and Growth Conditions

Seeds of Solanum habrochaites accessions were obtained from the C.M. Rick Tomato Genetics Resource Center (http://tgrc.ucdavis.edu/). Seeds were germinated on filter paper in the dark. Following germination, two to three plants per accession were grown on Jiffy-7 Peat Pellets (http://www.jiffy.com/) for 3 weeks as described previously (Schilmiller et al., 2009). Plants were transplanted into peat-based compost supplemented with fertilizer in greenhouses equipped with environmental controls and supplemental lighting at Michigan State University.

Trichome Isolation and Gene Cloning

Trichomes were isolated from the stems and petals of two to three fully grown plants of the same accession, and RNA extraction, cDNA synthesis, and gene cloning were performed as described previously (Gonzales-Vigil et al., 2012) using primers conserved for NDPS1 and zFPS (Sallaud et al., 2009; Schilmiller et al., 2009).

DNA Sequence Analysis, Multiple Sequence Alignments, Phylogenetic Analysis, and Homology Modeling

DNA sequences were edited and assembled using Sequencher software version 4.8 (http://genecodes.com/). Sequence analysis was performed with MEGA version 5 (Tamura et al., 2011) using MUSCLE (Edgar, 2004) for constructing the sequence alignments and the maximum likelihood method based on the model of Tamura and Nei (1993) for phylogenetic analysis. A bootstrap test was used to assess the reliability of the tree. Percentage identity between nucleotide and predicted amino acid sequences of TPS enzymes was calculated using the alignment made by MUSCLE to calculate distance. Homology models were constructed within the SWISS-MODEL Workspace using the automatic alignment algorithm (Arnold et al., 2006), and models were visualized and annotated using PyMOL (www.pymol.org). Models of NDPS1 and zFPS were generated using the crystal structure of Escherichia coli undecaprenyl pyrophosphate synthase chain A (PDB no. 1X07A) as a template (Guo et al., 2005). Models of SICPT4, SICPT5, and SICPT7 were generated using the crystal structure of Staphylococcus aureus undecaprenyl diphosphate synthase chain A (PDB no. 4HBE) as a template (Zhu et al., 2013).

Synthesis of Codon-Optimized Genes and Site-Directed Mutagenesis

Codon-optimized versions of NDPS1 and zFPS, lacking the chloroplast-targeting sequences, together with M10 versions of each sequence (Supplemental Fig. S3) were synthesized by Genscript (http://www.genscript.com/). Each synthetic gene contained BamHI and SalI restriction enzyme recognition sequences at the 5’ and 3’ ends, respectively. The CPT1 inserts were released from the pUC57 cloning vector by digestion with BamHI, and SalI and ligated into the pHS8 vector, previously linearized with the same enzymes. Recombinant clones were transformed into E. coli BL21 cells. Site-directed mutagenesis was performed using codon-optimized versions of NDPS1 and zFPS in the pUC57 vector as templates and utilizing standard methods (Sambrook and Russell, 2001). Primers utilized for site-directed mutagenesis are listed in Supplemental Table S1, and the sequences of the codon-optimized and mutant variants of NDPS1 and zFPS are provided in Supplemental Data Set S1.

Recombinant Protein Expression and Purification

A single E. coli colony was inoculated into 3 mL of Terrific Broth medium containing 50 μg mL⁻¹ kanamycin, and the culture was grown overnight at 37°C. A 500-μL aliquot of this culture was used to inoculate 100 mL of Terrific Broth medium containing 50 μg mL⁻¹ kanamycin. The culture was grown to log phase (A₆₀₀ = 0.3–0.7) at 37°C, at which point isopropyl-thio-β-D-galactopyranoside (Sigma-Aldrich) was added to a final concentration of 0.1 mM. The culture was grown for an additional 16 h at 20°C. Cells were harvested by centrifugation at 5,000 g for 20 min at 4°C. One-third of the cell pellet was resuspended in 3 mL of lysis buffer (50 mM HEPES, pH 8, 5% glycerol, 100 mM KCl, 7.5 mM MgCl₂, 10 mM imidazole, and 1 mg mL⁻¹ lysozyme) containing 1× Protease Inhibitor Cocktail tablets (www.roche-applied-science.com) and incubated on ice for 30 min prior to sonication. Soluble proteins were recovered from the supernatant following centrifugation at 10,000 g for 20 min at 4°C. A total of 0.5 mL of nickel-nitrilotriacetic acid agarose slurry (www.qiagen.com) was added to the cleared supernatant, and the mixture was incubated for 1 h at 4°C on a rotary shaker prior to passing through a Poly-Prep chromatography column. The column was washed with 10 mL of washing buffer (50 mM HEPES, pH 8, 5% glycerol, 100 mM KCl, 7.5 mM MgCl₂, and 20 mM imidazole). His-tagged proteins were eluted from the column with lysis buffer containing 250 mM imidazole but lacking lysozyme and protease inhibitor. Fractions containing purified proteins were pooled and buffer exchanged to 20 mM HEPES, pH 8, 20% glycerol, and 4 mM dithiothreitol (DTT) with an Amicon Ultra-15 10-kD filter (www.millipore.com). Purified proteins were quantified with the Pierce bicinchoninic acid Protein Assay Kit (www.themecoscientific.com) according to the manufacturer’s instructions. Purified recombinant proteins were approximately 90% pure as determined by SDS-PAGE.

Enzyme Activity Assays

Enzyme assays were modified from a previously published protocol (Thulasiram and Poulter, 2006). The standard enzyme assays were performed in 200 μL of reaction buffer (35 mM HEPES, pH 7.6, 10 mM MgCl₂, 5% glycerol, and 4 mM DTT) containing 1 μg of purified protein and substrates (10 μM each...
Kinetic Analysis

Kinetic analysis was performed using radiolabeled [14C]IPP (60 mCi mmol⁻¹; http://www.perkinelmer.com/) as a substrate with between 1 and 4 μg of purified recombinant enzyme in each reaction. The estimates of Km for DMAPP, IPP, and NPP were determined over a range of substrate concentrations using a fixed concentration of the cosubstrate as follows: NDPS1-LA2409, 30 to 80 μM [14C]IPP at 350 μM DMAPP and 20 to 100 μM DMAPP at 50 μM [14C]IPP; NDPS1-M10, 10 to 100 μM DMAPP at 50 μM[14C]IPP, 20 to 80 μM [14C]IPP at 150 μM NPP, and 10 to 200 μM NPP at 50 μM [14C]IPP; NDPS1-M4, 10 to 100 μM DMAPP at 50 μM [14C]IPP, 20 to 80 μM [14C]IPP at 150 μM NPP, and 10 to 100 μM NPP at 50 μM [14C]IPP; zFPS-LA1395, 10 to 200 μM DMAPP at 50 μM [14C]IPP, 20 to 80 μM [14C]IPP at 150 μM NPP, and 10 to 200 μM NPP at 50 μM [14C]IPP; zFPS-M20, 20 to 80 μM [14C]IPP at 150 μM DMAPP and 10 to 100 μM DMAPP at 50 μM [14C]IPP. All reactions were performed in a 50-μL volume containing 50 mM HEPES, 100 mM KCl, 7.5 mM MgCl₂, 5% (v/v) glycerol, and 5 mM DTT (pH 8.0) and were incubated at 30°C for 10 to 15 min. The reaction was stopped by the addition of [14C]IPP. Apparent Kmax value, the concentration of the enzyme in the assay, and the M, of the monomeric form of each recombinant protein.

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers KF494850 to KF494874.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Amino acid alignment of CPTI proteins from S. habrochaites.

Supplemental Figure S2. Homology models of tomato CPT enzymes.

Supplemental Figure S3. Amino acid alignment of mutant CPTI proteins utilized in enzyme assays.

Supplemental Table S1. Oligonucleotide primers utilized in this study.

Supplemental Data Set S1. Codon-optimized nucleotide and amino acid sequences of mutant variants of NDPS1-M and zFPS utilized in this study.

ACKNOWLEDGMENTS

We thank Dr. Joe Noel (Salk Institute) for providing the PH58 expression vector.

Received October 14, 2013; accepted November 19, 2013; published November 19, 2013.

LITERATURE CITED


Chen F, Tholl D, Bohmann H, Pichersky E (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. Plant J 66: 212–229


Copyright © 2014 American Society of Plant Biologists. All rights reserved.