Tetraterpene Synthase Substrate and Product Specificity in the Green Microalga *Botryococcus braunii* Race L

Hem R. Thapa,†,8 Su Tang,† James C. Sacchettini,†,‡ and Timothy P. Devarenne§,†,‡

†Department of Biochemistry & Biophysics, Texas A&M University, College Station, Texas 77843, United States
‡Department of Chemistry, Texas A&M University, College Station, Texas 77843, United States

**Supporting Information**

**ABSTRACT:** Recently, the biosynthetic pathway for lycopadiene, a C_{40} tetraterpene hydrocarbon, was deciphered from the L race of *Botryococcus braunii*, an alga that produces hydrocarbon oils capable of being converted into combustible fuels. The lycopadiene pathway is initiated by the squalene synthase (SS)-like enzyme lycopaoctaene synthase (LOS), which catalyzes the head-to-head condensation of two C_{20} geranylgeranyl diphosphate (GGPP) molecules to produce C_{40} lycopaoctaene. LOS shows unusual substrate promiscuity for SS or SS-like enzymes by utilizing C_{15} farnesyl diphosphate (FPP) and C_{20} phytyl diphosphate in addition to GGPP as substrates. These three substrates can be combined by LOS individually or in combinations to produce six different hydrocarbons of C_{30}, C_{35}, and C_{40} chain lengths. To understand LOS substrate and product specificity, rational mutagenesis experiments were conducted based on sequence alignment with several SS proteins as well as a structural comparison with the human SS (HSS) crystal structure. Characterization of the LOS mutants *in vitro* identified Ser276 and Ala288 in the LOS active site as key amino acids responsible for controlling substrate binding, and thus the promiscuity of this enzyme. Mutating these residues to those found in HSS largely converted LOS from lycopaoctaene production to C_{30} squalene production. Furthermore, these studies were confirmed *in vivo* by expressing LOS in *E. coli* cells metabolically engineered to produce high FPP and GGPP levels. These studies also offer insights into tetraterpene hydrocarbon metabolism in *B. braunii* and provide a foundation for engineering LOS for robust production of specific hydrocarbons of a desired chain length.

Isoprenoids (a.k.a. terpenes) are one of the largest groups of structurally diverse natural products produced by all domains of life and are biosynthesized via either the mevalonate pathway or the methylerythritol phosphate (MEP) pathway depending on the organism in question. Besides critical roles in primary and secondary metabolism, isoprenoids and derivatives have a wide range of applications including use as medicines, nutraceuticals, agricultural chemicals, fragrances, flavorings, and industrial chemicals. More recently, there has been significant interest in the use and development of isoprenoid-based chemicals as a direct substitute for petroleum-derived fuels since the physicochemical characteristics of some isoprenoids make them ideal for use as combustion fuels. However, low yields in natural hosts have directed isoprenoid-based biofuel research toward metabolic engineering for overproduction of isoprenoid chemicals in industrial microorganisms such as yeast and *E. coli*.

The colony-forming green microalga *Botryococcus braunii* is an exception to low isoprenoid yielding organisms as it naturally produces large amounts of liquid hydrocarbon oils, which are stored in the colony extracellular matrix. These hydrocarbons can be catalytically cracked to produce petroleum-equivalent combustion engine fuels, making *B. braunii* hydrocarbons a promising renewable biofuel feedstock. There are three chemical races of *B. braunii*, races A, B, and L, each producing distinct types of hydrocarbons. Race L, the focus of this study, produces the MEP pathway derived C_{40} tetraterpene hydrocarbon lycopadiene as the predominant hydrocarbon.

The major intermediates and the key enzyme for the lycopadiene biosynthetic pathway have recently been identified in the L race of *B. braunii*. This enzyme, termed lycopaoctaene synthase (LOS), catalyzes a two-step reaction resembling that of squalene synthase (SS). In the first step of the SS reaction, two molecules of C_{15} farnesyl diphosphate (FPP) are condensed head-to-head to generate the intermediate presqualene diphosphate (PSPP), which is reductively rearranged to squalene in the second reaction step (Figure 1). Squalene is then used for sterol production. By comparison, in the LOS reaction two molecules of C_{20} geranylgeranyl diphosphate (GGPP) undergo head-to-head dimerization to produce the prelycopaoctaene diphosphate (PLPP) intermediate in the first step, followed by reductive rearrangement to produce C_{40} lycopaoctaene in the second step (Figure 1). Production of lycopaoctaene represents the committed step in...
RESULTS AND DISCUSSION

LOS Sequence and Structural Comparison to SS Proteins. Initially, LOS was compared to LSS and four other SS enzymes to determine regions of similarity or dissimilarity and to identify mutational targets for understanding LOS substrate specificity. The SS catalyzed reaction has been extensively investigated due to consideration of human SS (HSS) as a drug target for reducing serum cholesterol levels, and the role of important catalytic residues has been verified. From these studies, seven functional SS domains have been identified: domains I–V for substrate binding and catalysis and a FLAP domain and a J-K loop involved in NADPH binding (Figure S2). The alignment of LOS with the five SS proteins showed conservation of all these domains in LOS (Figure S2), making it difficult to use these alignments alone to identify residues conferring LOS substrate specificity. Additionally, residues dictating the promiscuity of LOS may lie outside these conserved regions.

Thus, a structural comparison was used to identify the amino acids that control LOS substrate binding. To do this, a predicted LOS structure was generated that showed LOS as an α-helical protein with several regions matching previous SS structures such as a large catalytic channel running through the center of the protein (dashed box in Figure S3A), an N-terminal flexible region (green in Figure S3A), and a C-terminal transmembrane domain (magenta in Figure S3A).

While there have been several SS and SS-like structures determined, the HSS crystal structure, specifically a truncated HSS protein in complex with the FPP analog farnesyl thiophosphate (FSP), was chosen for comparison to the predicted LOS structure. This HSS protein retains full SS activity and was generated by deleting the N-terminal flexible hydrophobic region and the C-terminal transmembrane domain. As we previously reported, a similar LOS C-terminal truncation, producing LOS, maintained full enzyme activity (Figure S3B). In order to match the HSS structure, N- and C-terminal truncations were made to LOS to generate LOS (Figure S3C), which had a marginal reduction in enzyme activity (Figure S3B). This suggests a comparison of LOS to HSS should offer valuable information regarding potential LOS substrate binding sites as both structures superimpose very well (Figure S3C).

Identification of LOS Residues for Mutagenesis. Based on the HSS–LOS structural comparison, five LOS amino acids appear to be important for substrate interaction: Thr65, Met180, Ser276, Ala288, and Val289 (Figure 2A). Except for Ser276, all residues are located within the conserved functional SS domains I, III, and IV, and all five LOS residues are different from the corresponding amino acids in the other SS proteins (Figure 2A). For instance, the corresponding HSS amino acids are Val69, Leu183, Tyr276, Phe288, and Cys289 (Figure 2A), all of which have been shown to form a hydrophobic floor in the central cavity of the HSS active site for interaction with the nonpolar FSPP tails. The hydrophobic side chains of Val69, Leu183, and Phe288 form the hydrophobic floor for the substrate 1 (S1) pocket, whereas the hydrophobic portions of the Tyr276 and Cys289 side chains form the hydrophobic floor for the substrate 2 (S2) pocket (Figure 2B). The corresponding five LOS residues are located in the same hydrophobic region of the LOS active site, with the Thr65, Met180, and Ala288 side chains contributing to the S1 pocket.

Figure 1. Squalene production and lycopadiene biosynthesis in Botryococcus braunii race L. Two molecules of farnesyl diphosphate (FPP) are condensed by squalene synthase (SS) to produce the presqualene diphosphate (PSPP) intermediate followed by reductive rearrangement to produce squalene. Lycopadiene synthase (LOS) catalyzes the condensation of two molecules of geranylgeranyl diphosphate (GGPP) to produce the prelycopaeca diphosphate (PLPP) intermediate followed by reductive rearrangement to produce lycopadiene, which is subsequently reduced to produce lycopadiene. This discovery provides a unique opportunity to engineer the LOS enzyme for robust production of individual hydrocarbons with specific chain lengths for possible industrial uses such as fuel production.

In order to accomplish this, the mechanism underlying LOS substrate promiscuity needs to be understood. It has been proposed that LOS and the B. braunii L race SS (LSS) arose from an ancient SS gene duplication, with LSS maintaining SS activity, while LOS evolved the new catalytic function to produce lycopadiene leading to hydrocarbon production. The promiscuous LOS activity is likely due to preservation of the original SS activity following the gene duplication event. Thus, conversion of LOS to an enzyme producing only one hydrocarbon with specific chain length for possible industrial uses such as fuel production could be accomplished by understanding how LOS binds substrate for product formation. In this study, we use protein alignments between LOS and SS proteins, as well as a comparison of a predicted LOS 3D structure with the human SS crystal structure to conduct rational site-directed mutagenesis in order to identify the critical LOS residues important for conferring substrate and product specificity.
substrate pocket hydrophobic floor, and Val289 and Ser276 contributing to the S2 substrate pocket hydrophobic floor (Figure 2C). Replacement of HSS Phe288 with Ala288 in LOS and HSS Tyr276 with Ser276 in LOS appears to enlarge the S1 and S2 substrate binding pockets, respectively, to accommodate the larger GGPP substrate (Figure 2C). We therefore hypothesized LOS Ala288 and Ser276 were key mutations that arose during gene duplication to allow GGPP binding while retaining FPP binding. Thus, Ala288 and Ser276 plus the three other residues, Thr65, Met180, and Val289, were chosen for mutational studies to decipher the roles of these amino acids in substrate binding and product specificity.

Characterization of LOS Mutants in Vitro. In order to determine the roles of LOS Thr65, Met180, Ser276, Ala288, and Val289 in substrate binding and product formation, these amino acids were mutated to the corresponding HSS residues individually and in combinations (Table 1). The mutated proteins were purified, enzyme activity characterized in a mixed substrate assay (Figure 3A) containing equimolar $^3$H-FPP and $^3$H-GGPP, and the enzymatic activity quantified based on $^3$H incorporation into the squalene, C$_{35}$H$_{58}$, and lycopaoctaene products (Figure 3B). The FPP/GGPP mixed substrate assay was chosen over the FPP/PPP or GGPP/PPP mixed substrate assays because the FPP/GGPP assay is more robust in terms of product formation.

For the M1 to M5 single amino acid mutations the results indicated M1, M2, and M5 retained the ability to produce lycopaoctaene, whereas replacement of Ser276 and Ala288 with the larger aromatic amino acids Tyr and Phe in M3 and M4, respectively, lead to a significant loss of lycopaoctaene production when compared to wild-type (WT; Figure 3B). In the case of C$_{35}$H$_{58}$ production, all single mutants retained at least WT activity (Figure 3B). Interestingly, the M3 mutant showed a significant increase in C$_{35}$H$_{58}$ production to 183% of WT LOS (Figure 3B). Except for M3, all single mutants retained at least WT activity levels for squalene production (Figure 3B). The observed activity for M3 and M4 suggests the increased residue size in these mutants blocks GGPP binding in the S1 pocket by the M4 A288F mutation and in the S2 pocket by the M3 S276Y mutation, indicating the importance of LOS Ser276 and Ala288 in GGPP binding. Further support for this is seen by the increased C$_{35}$H$_{58}$ production in M3 (Figure 3B), suggesting favorable binding of one GGPP in the S1 pocket and one FPP in the S2 pocket. Additionally, the decreased lycopaoctaene production, but no change in squalene and

### Table 1. List of Lycopaoctaene Synthase (LOS) Mutants Characterized in This Study

<table>
<thead>
<tr>
<th>mutant name</th>
<th>mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>T65V</td>
</tr>
<tr>
<td>M2</td>
<td>M180L</td>
</tr>
<tr>
<td>M3</td>
<td>S276Y</td>
</tr>
<tr>
<td>M4</td>
<td>A288F</td>
</tr>
<tr>
<td>M5</td>
<td>V289C</td>
</tr>
<tr>
<td>M6</td>
<td>T65V M180L</td>
</tr>
<tr>
<td>M7</td>
<td>T65V A288F</td>
</tr>
<tr>
<td>M8</td>
<td>V289C S276Y</td>
</tr>
<tr>
<td>M9</td>
<td>S276Y A288F</td>
</tr>
<tr>
<td>M10</td>
<td>T65V M180L A288F</td>
</tr>
<tr>
<td>M11</td>
<td>T65V M180L S276Y A288F</td>
</tr>
<tr>
<td>M12</td>
<td>T65V M180L S276Y A288F V289C</td>
</tr>
</tbody>
</table>

ACS Chemical Biology

DOI: 10.1021/acschembio.7b00457
ACS Chem. Biol. 2017, 12, 2408−2416
C_{35}H_{58} production in M4 (Figure 3B) suggests limited binding of two GGPP molecules and preferable binding of two FPP molecules, or one FPP in the S1 pocket and one GGPP in the S2 pocket (Figure 3A,B).

Next, the LOS double mutants M6 to M9 (Table 1) were characterized for enzymatic activity using the same mixed substrate assay (Figure 3A). The enzyme activity of M6 is similar to that of WT (Figure 3B). When comparing the M7 double mutant to the parent M4 background, M7 did not show a difference in lycopaoctaene production, but C_{35}H_{58} production was slightly increased and squalene formation was significantly reduced (Figure 3B). For the M8 double mutant, in comparison to the M3 parent background, a small increase in C_{35}H_{58} production was observed, whereas the activities for lycopaoctaene and squalene production remained the same (Figure 3B). Most interestingly, the M9 double mutant, which has both the M3 S276Y and the M4 A288F mutations, showed a drastic reduction in the ability to produce lycopaoctaene and C_{35}H_{58} compared to WT, while the SS activity is substantially increased more than 3-fold over the WT enzyme (Figure 3B). These results for M9 further support our hypothesis that Ser276 and Ala288 are important for controlling LOS substrate binding, and replacing these residues with bulky aromatic amino acids blocks the binding of GGPP in both substrate binding pockets. This is further supported by a previous study where the corresponding Phe288 was shown to be essential in the HSS reaction. The HSS F288A crystal structure showed mutation to the smaller Ala resulted in a deeper cleft in the S1 substrate pocket. Thus, having Ala at position 288 in LOS would expand the S1 substrate pocket to allow GGPP binding. Additionally, Phe288 was proposed to be involved in PSPP diphosphate ionization and orientation. Accordingly, the F288A mutation in HSS significantly reduced enzyme activity to 22.7% of WT activity. Since the WT LOS with Ala288 is fully active, this would suggest additional LOS residues that are different from SS allow for proper PLPP diphosphate orientation and ionization.

Finally, the M10 triple mutant, the M11 quadruple mutant, and the M12 quintuple mutant (Table 1) were analyzed for enzymatic activity using the mixed substrate assay. The M10 mutant showed activity similar to that of the M4 parent background (Figure 3B). For the M11 and M12 mutants, the activities are very similar to each other. But in comparison to the M9 parent background, both M11 and M12 showed a further reduction in C_{35}H_{58} production and no significant change in lycopaoctaene and squalene formation (Figure 3B).

Figure 3. Characterization of LOS mutants in vitro. (A) Reaction scheme showing LOS products from the FPP/GGPP mixed substrate assay. (B) Enzyme activities of the LOS mutants in the FPP/GGPP mixed substrate assay showing production of squalene, C_{35}H_{58}, and lycopaoctaene. Numbers on top of each bar indicate the activity percentage based on WT LOS activity as 100%. Data shown are from three independent experiments (n = 3). LOS, lycopaoctaene synthase; FPP, farnesyl diphosphate; GGPP, geranylgeranyl diphosphate.
These results suggest mutation of the five LOS amino acids studied here to the corresponding HSS residues is sufficient to change LOS substrate binding specificity to mainly FPP binding, and thus product specificity to squalene production (compare WT and M12 in Figure 3B). It should be noted the M12 mutant retained residual activity for lycopaoctaene and C35H58 production (Figure 3B), indicating low levels of GGPP binding and suggesting further modification is required in the M12 background to completely abolish GGPP binding (Figure 3B).

Analysis of LOS and LOS Mutants in Vivo. To further support the in vitro assay data, the activity of WT LOS and the LOS mutants was characterized using Escherichia coli as an in vivo expression system since the native host, B. braunii, cannot be transformed. Like B. braunii, E. coli utilizes the MEP pathway for isoprenoid biosynthesis.21 Most importantly, E. coli does not naturally produce lycopaoctaene, C35H58, or squalene (Figure 4, panel I). Thus, the E. coli expression system provides a clean platform on which to characterize LOS and the LOS mutants in vivo.

The in vitro mixed substrate assays used above for LOS mutant characterization contained equal FPP and GGPP concentrations. Thus, an E. coli line generating equal and high amounts of FPP and GGPP is needed for in vivo LOS and LOS mutant characterization. However, E. coli has limited FPP amounts and even lower GGPP levels.22 This is consistent with only small amounts of squalene and no C35H58 or lycopaoctaene produced from LOS overexpression in E. coli (Figure 4, panel II). To overcome this limitation, FPP synthase from E. coli (a.k.a. IspA, referred to here as EcFPS) was overexpressed to increase FPP levels. Overexpression of EcFPS with LOS showed squalene production almost doubled, a roughly equal amount of C35H58 was detected, and no lycopaoctaene was produced (Figure 4, panel III).
suggests EcFPS overexpression slightly increased intracellular FPP levels, and GGPP levels were high enough to contribute to \( \text{C}_{35}\text{H}_{58} \) production but not high enough for LOS-mediated lycopentaene production.

To direct more carbon into the MEP pathway and further boost FPP and GGPP production, deoxyxylulose-5-phosphate synthase from *E. coli* (*EcDXS*), which catalyzes the first and rate limiting MEP pathway step,\(^{23}\) was overexpressed with EcFPS and LOS. This expression generated high levels of squalene and increased the amount of \( \text{C}_{35}\text{H}_{58} \) (Figure 4, IV), suggesting a large increase in carbon flux through the MEP pathway to amplify FPP production. However, lycopentaene production was still not detected (Figure 4, panel IV), arguing EcDXS overexpression did not increase GGPP pools to a level required for lycopentaene biosynthesis.

In order to enhance the GGPP pool, GGPP synthase from *Arabidopsis thaliana* (*AtGGPPS11*) was overexpressed along with EcFPS and LOS. Hydrocarbon production analysis in this line showed mainly ample lycopentaene production, indicating high levels of GGPP generated by *AtGGPPS11* (Figure 4, panel V). Trace \( \text{C}_{35}\text{H}_{58} \) amounts and no detection of squalene in this line (Figure 4, panel V) may have resulted from an FPP pool reduction by *AtGGPPS11*. Additionally, the 2-fold higher LOS affinity for GGPP over FPP\(^{12}\) likely contributes to preferred lycopentaene production when GGPP levels are high. To further boost FPP and GGPP production, we overexpressed EcDXS, EcFPS, and *AtGGPPS11* in one line with LOS. Analysis showed lycopentaene production was further enhanced and remained the predominant product, while trace amounts of squalene were detected and \( \text{C}_{35}\text{H}_{58} \) production slightly increased (Figure 4, panel VI). These data suggest both the FPP and GGPP pools increased as a result of EcDXS overexpression in this line.

The sharp change in LOS activity from squalene production (Figure 4, panel IV) to lycopentaene production in the presence of *AtGGPPS11* (Figure 4, panel V, VI) suggests the
FPP pools were too low to support LOS squalene production due to FPP conversion to GGPP by AtGGPPS11. This is supported by LSS mediated squalene production in these engineered E. coli lines (Figure S4), which is used as a proxy for FPP levels. Maximal squalene levels were seen when LSS was expressed with EcDXS and EcFPS (Figure S4, panel IV), which was 4.2 times higher than LSS expression alone (Figure S4, panel II). This suggests maximal FPP levels in the EcDXS and EcFPS expressing E. coli line. The inclusion of AtGGPPS11 (with or without EcDXS) reduced squalene production to approximately 1.6 times that of LSS expression alone (Figure S4, panel V, VI), suggesting FPP pools are reduced close to WT E. coli levels due to AtGGPPS11 expression. This lower FPP level in combination with the previously mentioned higher affinity of LOS for GGPP may explain the preference for lycopoaetnae production in the presence of moderate FPP pools and high GGPP pools.

Interestingly, this product specificity of LOS for lycopoaetnae over C_{35}H_{58} and squalene as observed in E. coli (Figure 4, panel VI) could mirror hydrocarbon metabolism in B. braunii race L. Our past studies have shown that while lycopoaetnae is rapidly converted to lycopadiene, there is no detectable squalene production and low levels of C_{35}H_{58}, a reduced product of C_{35}H_{58}.

This would suggest in the L race LOS has access to high GGPP levels, leading to the product specificity for lycopoaetnae over C_{35}H_{58} or squalene. In B. braunii race L, this may come about by metabolic channeling of the MEP pathway toward GGPP, and thus lycopadiene biosynthesis via lycopoaetnae production by LOS.

Next, we investigated the in vitro activity of several select LOS mutants by coexpression in the EcDXS, EcFPS, and AtGGPPS11 expressing E. coli line, which should have the highest levels of both FPP and GGPP. Although expression of LOS with EcDXS, EcFPS, and AtGGPPS11 in E. coli did not result in a hydrocarbon production profile similar to WT LOS in the in vitro mixed substrate assay, expression of the LOS mutants in this E. coli line allowed us to generate valuable information as discussed below that further supports the in vitro data.

The mutants selected for expression in the EcDXS, EcFPS, and AtGGPPS11 expressing E. coli line were M3, M4, and M9 (Table 1). These mutants are focused on the LOS Ser276 and Ala288 residues, which appear to have the biggest in vitro impact on lycopoaetnae production and a large increase in squalene production compared to WT LOS (Figure S5, panel IV). As intended, the M12 mutations drastically reduced squalene levels compared to WT LOS (Figure S5, panel V). However, the C_{35}H_{58} levels were higher than WT LOS in vivo (Figure S5, panel V), while in vitro the C_{35}H_{58} levels decreased for M12 compared to WT LOS (Figure 3). Again, this could be due to high levels of GGPP in vivo. As intended, the M12 mutations changed LOS product specificity from lycopoaetnae to squalene both in vitro and in vivo. Overall, while the absolute changes in product formation for the mutants in vivo (Figure S5) do not match that seen for the in vitro data (Figure 3), the trends in product formation for the M3, M4, M9, and M12 mutants in vivo are consistent with the enzymatic activities observed in the in vitro assays.

Structural Insights into the Role of LOS Ser276 and Ala288 in GGPP Binding. The studies presented above indicate LOS residues Ser276 and Ala288 play a key role in substrate binding and product specificity. To gain more insight into the role of these two amino acids in controlling substrate binding, we compared the HSS structure in complex with FSPP to the predicted LOS structure focusing on HSS Tyr276 and Phe288 and LOS Ser276 and Ala288 (Figure 6A). The analysis showed the distances in HSS from the Phe288 benzene ring and the Tyr276 phenol group to the corresponding S1 and S2 FSPP tails are 3.54 and 4.35 Å, respectively (Figure 6A). Given that GGPP is 4.00–4.80 Å longer than FPP, depending on which branch of the terminal isoprene unit is used for measurement, this analysis suggests HSS cannot bind GGPP because the bulky Phe288 in the S1 binding site would clash with GGPP, and taking into account atomic radii and repulsion between atoms Tyr276 in the S2 binding site would interfere with GGPP binding. This notion is further supported by the observed LOS M9 double mutant in vitro and in vivo enzyme activities, which showed substantial losses of lycopoaetnae and C_{35}H_{58} formation, and a large increase in squalene production as a result of replacing the small Ser276 and Ala288 residues with Tyr and Phe, respectively (Figure 3B and Figure 5, panel IV). Further support of Ser276 and Ala288 controlling GGPP binding in LOS is the 7.85 Å between the Ala288 side chain and the S1 FSPP tail, and the 7.06 Å between the Ser276 hydroxyl group and the S2 FSPP tail (Figure 6A). These distances should provide enough room in the two LOS substrate binding sites to accommodate the larger GGPP hydrophobic tail.

Additional characterization of the two substrate binding sites was conducted by comparing the HSS structure and the predicted LOS structure with the Staphylococcus aureus dehydroquinase synthase (CrtM) structure bound to the GGPP analog geranylgeranyl thiophosphate (GGSP; Figure 6B).19 CrtM catalyzes an SS-like reaction, except the second reaction step is an NADPH-independent rearrangement of PSP to produce dehydroquinale,19 which contains a cis double bond at the linkage between the two FPP molecules (Figure S5). Except for the NADPH binding residues, CrtM shares sequence similarity with the conserved SS domains,25 and the substrate binding sites in CrtM and HSS are structurally homologous.3,5 Although the natural substrate of CrtM is FPP, CrtM can also utilize GGPP to produce a C_{35} version of dehydroquinale in the presence of elevated GGPP
levels. This is presumably due to GGPP binding in the S2 site and FPP in the S1 site. This is supported by mutation of the bulky Tyr residue to Ala and the structure of the CrtM protein, which showed binding of the longer GGPP substrate at the S1 site and production of a C40 version of dehydro-squalene.

The structural comparison of HSS, LOS, and CrtM in relation to the FSPP (A) and GGSP (B) substrate analogs. (A) Superposition of the HSS crystal structure bound to two FSPP molecules and the predicted LOS 3D structure. (B) Superposition of the CrtM crystal structure bound to two GGSP molecules, the HSS crystal structure (FSPP removed), and the predicted LOS 3D structure. HSS, human squalene synthase; LOS, lycopaoctaene synthase; FSPP, farnesyl thiodiphosphate; GGSP, geranylgeranyl thiodiphosphate.

Figure 6. Position comparison of the key residues controlling substrate binding from HSS, LOS, and CrtM in relation to the FSPP (A) and GGSP (B) substrate analogs. (A) Superposition of the HSS crystal structure bound to two FSPP molecules and the predicted LOS 3D structure. (B) Superposition of the CrtM crystal structure bound to two GGSP molecules, the HSS crystal structure (FSPP removed), and the predicted LOS 3D structure. HSS, human squalene synthase; LOS, lycopaoctaene synthase; FSPP, farnesyl thiodiphosphate; GGSP, geranylgeranyl thiodiphosphate.

This work was supported by grant NSF-EFRI-PSBR #1240478 to T.P.D.
Supporting Information for:

Substrate and product specificity for a tetraterpenoid synthase from the green microalga *Botryococcus braunii* race L

Hem Raj Thapa,† Su Tang,† James C. Sacchettini,†‡ Timothy P. Devarenne*,†

†Department of Biochemistry & Biophysics, Texas A&M University, College Station TX 77843, USA
‡Department of Chemistry, Texas A&M University, College Station TX 77843, USA
Figure S1. LOS reactions products when incubated with equimolar amounts of (A) FPP and PPP, or (B) GGPP and PPP.
**Figure S2.** Amino acid sequence alignment of LOS with squalene synthase (SS) proteins from different organisms using Clustal Omega. Conserved domains and regions found in a typical squalene synthases are highlighted indicated above the sequence. The FPP/PSSPP diphosphate binding binding motifs are shown in blue, NADPH binding residues are highlighted in green, and the residues of LOS chosen for mutagenesis and the corresponding residues in the SS proteins are shown in red. Transmembrane domain predicted by TMpred for each protein are underlined. LOS, lycopoaetaene synthase (AC# KT388101); LSS, SS from *Botryococcus braunii* race L (AC# KT388100); BSS, SS from *Botryococcus braunii* race B (AC# AH009227); AtSS, SS from *Arabidopsis thaliana* (AC# D29017.1); YSS, SS from *Saccharomyces cerevisae* (AC# M63979.1); and HSS, Human SS (AC# Q6IA1X1).
Figure S3. 3D-Structures of lycopaoctaene synthase (LOS) and human squalene synthase (HSS). (A) A model of the LOS structure as predicted by the I-TASSER program. The N-terminal flexible region is shown in green and a double pass transmembrane domain predicted by TMPred software is shown in magenta. (B) LOS enzyme assay for lycopaoctaene production using LOS\textsuperscript{1-391} and LOS\textsuperscript{33-381} with geranylgeranyl diphosphate (GGPP) given as a substrate. (C) Comparison of the crystal structure of HSS\textsuperscript{31-370} (light blue) bound to two molecules of farnesylthiol diphosphate (FSPP) and a computationally truncated model of LOS\textsuperscript{33-381} (light brown). (D) Enzyme activities of LOS\textsuperscript{1-391} and LOS\textsuperscript{33-381} in the FPP/GGPP mixed substrate assay showing production of squalene, C\textsubscript{35}H\textsubscript{58}, and lycopaoctaene.
Figure S4. GC-MS profiles and quantitation of n-hexane extractable metabolites from *E. coli* cells expressing LSS with MEP pathway gene constructs. (I) Empty vector expression. (II) LSS expression. (III) *Ec* FPS and LSS expression. (IV) *Ec* DXS, *Ec* FPS, and LSS expression. (V) *Ec* FPS, *At* GGPPS11, and LSS expression. (VI) *Ec* DXS, *Ec* FPS, *At* GGPPS11, and LSS expression. (VII) Squalene standard. Shown to the right of the GC-MS profiles is quantitation of squalene based on the GC-MS data and expressed in ng/mL. The GC-MS data shown are representatives from three independent experiments (n = 3) and each experiment was used for quantitation. Numbers shown in the quantitation are the average value for squalene.
**Figure S5.** The two step reaction carried out by *S. aureus* CrtM to produce dehydrosqualene.
METHODS

Reagents. Radiolabeled substrates [1-\(^{3}\)H]-FPP (specific activity, 18.2 Ci mmol\(^{-1}\)) and [1-\(^{3}\)H]-GGPP (specific activity, 20.0 Ci mmol\(^{-1}\)) were purchased from PerkinElmer, and non-labeled FPP and GGPP were purchased from Sigma. All other chemicals were purchased from VWR unless otherwise noted.

Protein sequences for amino acid comparison. The following proteins were used for sequence comparison to LOS (AC# KT388101): LSS, *B. braunii* L race SS (AC# KT388100); BSS, *B. braunii* B race SS (AC# AH009227); AtSS, *Arabidopsis thaliana* SS (AC# D29017.1); YSS, yeast (*Saccharomyces cerevisiae*) SS (AC# M63979.1); HSS, human SS (AC# Q6IAAX1).

Structure comparison. The LOS protein 3-D model was generated using the Iterative Threading ASSEmbly Refinement (I-TASSER) bioinformatics program\(^{1-3}\). Among five models of LOS predicted by I-TASSER, the model with a high confidence score (C-score = 0.19) was used for structural analysis in this study. The molecular graphics and structural analyses of the predicted LOS structure, the HSS crystal structure (PDB ID; 3WEF), and the CrtM crystal structure (PDB ID; 3AE0) were done using the UCSF Chimera package\(^4\).

LOS expression vectors and LOS site-directed mutagenesis. The 6xHis-LOS\(^{1-391}\) and 6xHis-LOS\(^{23-381}\) expression constructs were generated by cloning LOS cDNA bases 1 to 1,173 and 97 to 1,143, respectively, into the *NheI* and *HindIII* restriction sites of pET28a\(^5\). The desired LOS point mutations were created using the 6xHis-LOS\(^{1-391}\) background and primers containing the altered nucleotide sequence(s). Conditions for PCR-mediated site directed mutagenesis reaction using Pfu Turbo DNA Polymerase (Agilent Technologies) were as follows: 98°C initial denaturation for 2 min, 17 cycles of denaturation, annealing, and extension (98°C for 30 sec, 55°C
for 30 sec, 72°C for 8 min, respectively), and a final extension at 72°C for 10 min. All mutants were verified by DNA sequencing.

**Protein expression and purification.** The desired expression construct was transformed into *E.coli* BL21(DE3), and an overnight culture from a single colony was used to inoculate (1.7% v/v) 100 mL of TB medium. The culture was grown at 37°C to OD_{600} = 0.8, and protein expression induced by adding 1 mM isopropyl β-D-thiogalactoside. The induced cultures were then grown for an additional 6 hours at 25°C, cells harvested by centrifugation, and the pellets stored at -80°C for future use. Purification of the 6xHis tagged proteins was conducted using Ni-NTA agarose resin (QIAGEN) following the manufacturer’s recommendations with some modifications. All steps of purification were done either on ice or in a 4°C cold room. In a typical purification, the 100 mL culture pellet was suspended in 10 mL of lysis buffer (50 mM MOPS, pH=7.8, 300 mM NaCl, 10 mM imidazole, 1x general protease inhibitor cocktail (Sigma), 20 mM MgCl₂, 2.5 mM DTT, 1% glycerol (v/v)) followed by cell lysis using four successive probe sonications at 70% maximum power for 20 sec. The lysed samples were centrifuged at 16,000 x g for 30 min at 4°C, the supernatant incubated with 1 mL of pre-equilibrated Ni-NTA resin for 60 min at 4°C using a rocking table, the sample loaded into a gravity column, and the flow-through was discarded. The Ni-NTA column was washed with 10 mL of wash buffer (50 mM MOPS, pH=7.8, 300 mM NaCl, 60 mM imidazole, 20 mM MgCl₂, 2.5 mM DTT, 1% glycerol (v/v)), and the protein bound to the nickel resin was eluted with 2 mL of elution buffer (50 mM MOPS, pH=7.8, 300 mM NaCl, 400 mM imidazole, 20 mM MgCl₂, 2.5 mM DTT, 1% glycerol (v/v)). The eluted protein sample was dialyzed against storage buffer (50 mM MOPS, pH=7.8, 300 mM NaCl, 20 mM MgCl₂ and 5 mM DTT, 1% glycerol (v/v)), concentrated using an Amicon Ultra centrifugal filter (30kDa cutoff;
EMD Millipore) to the desired concentration, an equal volume of glycerol added, and the protein sample stored at -20°C for short term or at -80°C for future use.

**Radioactive in vitro enzyme assays.** The enzyme assays were conducted using the protocol described previously\(^5\). In brief, a 50 μL reaction was initiated by adding 1 μg of purified protein to a reaction buffer containing 50 mM MOPS, pH 6.8, 2.5 mM β-mercaptoethanol, 20 mM MgCl\(_2\), 2 mM NADPH, and 10 μM of \(^3\)H-prenyl-PP (0.125 μCi; specific activity = 0.25 Ci mmol\(^{-1}\)). For the mixed substrate assay 10 μM of both \(^3\)H-FPP and \(^3\)H-GGPP was used (0.125 μCi for each substrate; specific activity = 0.25 Ci mmol\(^{-1}\)). The specific activity for each substrate was adjusted by adding non-labeled FPP and/or GGPP. The reactions were incubated for 60 min at 37°C and terminated by adding 60μL of n-hexane, followed by brief vortexing and centrifugation. Thirty microliters of the supernatant was then analyzed on Silica gel 60 TLC plates using n-hexane as the mobile phase. \(^3\)H incorporation into reaction products was determined by scraping the spots corresponding to authentic standards of lycopaoctaene (\(R_f = 0.09\)), C\(_{35}\)H\(_{58}\) (\(R_f = 0.12\)) and squalene (\(R_f = 0.17\)) followed by an analysis on a liquid scintillation counter.

**EcDXS and EcFPS cloning.** The cDNA sequences of *EcDXS* (AC# NP_414954) and *EcFPS* (a.k.a. ispA; AC# NP_414955) were obtained from the NCBI nucleotide database\(^6\). Genomic DNA from *E.coli* K-12 strain MG1655 and gene specific primers were used to amplify the PCR product using Phusion DNA polymerase followed by cloning into the pGEM-T vector (Promega). Gene specific primers were as follows: For *EcDXS*, forward primer 5’-ATGAGTTTTGATATTGCCAAATACCCG-3’ and reverse primer 5’-TTATGCCAGCCAGCCTTGATTTTG-3’; for *EcFPS*, forward primer 5’-ATGGACTTTTCGCAGCAACTC-3’ and reverse primer 5’-TTATTTATTACGCTGGATGATGTTAGTCCGC-3’.
**EcDXS, EcFPS, and AtGGPPS11 expression constructs.** The DNA template for AtGGPPS11 was obtained from our previous study\(^5\) and the DNA sequence encoding the 56 N-terminal chloroplast targeting signal amino acids was deleted to yield AtGGPPS11\(^{57-371}\). The expression constructs were made using the following restriction sites and vectors: EcDXS in NdeI and XhoI of pET22b, EcFPS in BamHI and SalI of pACYDuet-1, and AtGGPPS11\(^{57-371}\) in NdeI and XhoI of pACYDuet-1. For experiments involving expression of EcFPS and AtGGPPS11\(^{57-371}\), the two cDNAs were expressed using a single expression construct, EcFPS + AtGGPPS11\(^{57-371}\) in pACYDuet-1, by cloning EcFPS into the second multiple cloning site of AtGGPPS11\(^{57-371}\):pACYDuet-1 using the BamHI and SalI restriction sites.

**LOS in vivo mutant analysis in E.coli.** The expression construct(s) described above were either transformed individually or in combination into E.coli BL21 (DE3) cells and the positive transformants were selected using standard procedures. The E.coli line harboring the desired construct(s) was grown in 60 mL of TB medium at 37°C to OD\(_{600}\) = 0.8, and gene expression was induced by adding 1mM isopropyl β-D-thiogalactoside. After induction, cells were grown for 6 hrs at 25°C, 50 mL of cells were harvested by centrifugation, the pellet snap frozen with liquid nitrogen, and the pellet lyophilized. Freeze-dried cells were transferred to a glass vial and extracted with n-hexane by vortexing vigorously at room temperature. The organic extracts were centrifuged at 1,000 x g, the supernatant dried, resuspended in 200 μL n-hexane, and a 5 μL aliquot analyzed by GC-MS.

**GC-MS Analysis.** A Bruker 436-GC-SCION SQ Premium GC-MS system was used to evaluate hydrocarbon production. GC-MS analyses were conducted using a 5% Phenyl BR-5ms capillary column (30 m × 0.25 mm, film thickness: 0.25 μm) in electron ionization (70 eV) mode using helium as a carrier gas at a flow rate of 2.58 mL min\(^{-1}\). The initial oven temperature was held
at 220°C for 1 min followed by an increase in temperature at a rate of 5°C min⁻¹ to 280°C, ramped again to 300°C at a rate of 2°C min⁻¹, and held for 20 min. The temperature for the injection port, the interface, and ion source were set at 280°C, 250°C, and 200°C, respectively. Quantitation for hydrocarbon production in *E. coli* was done based on the calibration curve generated with commercially available squalene standard (Sigma).

**References:**


