Exploring Natural Diversity of Limonene Synthases and Molecular Determinants Involved in Substrate Specificity in *Escherichia coli*

Clement P.M. Scipion[1,2], Jérémy Esque[3], Shreyash Borkar[2], Cristalle Seah[1], Sophie Bozonnet[3], Magali Remaud-Siméon[3], Bo Xue[4,5,6], Wen Shan Yew[4,5,6], Isabelle André[3,], Xixian Chen[2,*]*

1. CNRS@CREATE, 1 Create Way, #08-01 Create Tower, Singapore 138602
2. Singapore Institute of Food and Biotechnology Innovation (SIFBI), Agency for Science, Technology and Research (A*STAR), 31 Biopolis Way, Nanos, Singapore 138669, Singapore
3. Toulouse Biotechnology Institute, TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France. 135, avenue de Rangueil, F-31077 Toulouse Cedex 04, France
4. Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore (NUS), 8 Medical Drive, Singapore 117597, Singapore
5. NUS Synthetic Biology for Clinical and Technological Innovation, 28 Medical Drive, Singapore 117456, Singapore

6. Synthetic Biology Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, 14 Medical Drive, Singapore 117599, Singapore

Corresponding Authors

* Corresponding authors: isabelle.andre@insa-toulouse.fr; xixian_chen@sifbi.a-star.edu.sg

Abstract

Limonene is a chiral, high-demand monoterpene that has wide applications in therapeutics, cosmetics, biofuels, agri-food, biomaterials, and solvent industries. However, its biosynthesis by microbial cell factories is often limited by the poor activity of limonene synthases (LS). Optimization of the rate-limiting enzyme is thus crucial for boosting limonene production. Here, we report the identification of ten LS homologues from sequence data mining and their testing in cells accumulating geranyl pyrophosphate (GPP) or neryl pyrophosphate (NPP) for limonene production. The selectivity of these enzymes towards GPP or NPP was investigated, leading to the identification of a limonene synthase from *Agastache rugosa* displaying a clear substrate preference towards NPP over GPP *in vivo*. This enzyme was selected as a template for engineering. Using *in silico* analyses and mutagenesis, several mutants were engineered that revealed differences in substrate specificity. Among them, a combination of mutations (S8K/I265V/E276P/P277R/A281K/N282T/I285Q/I286L) improved limonene production by 4.8 and 1.9-fold with the GPP and NPP pathways, respectively. The mutant predominantly produced (+)-limonene from GPP and a mixture of limonene from NPP, with ~85–90% of (+)-limonene. It decreased the selectivity for NPP by 2.4-fold. This supports improved biological production of limonene enantiomers from renewable carbon sources.

KEYWORDS: monoterpene, terpenoid, limonene, geranyl pyrophosphate, neryl pyrophosphate, enzyme engineering, limonene synthase, *Agastache rugosa*, enantioselectivity

1. Introduction

Limonene is a monoterpene that is present in lemon and other citrus fruits as well as in 300 other plant species and a few bacterial species[1]. It is optically active and is found naturally in both enantiomer forms, R-(+)-limonene (D-limonene) and S-(-)-limonene (L-limonene). Limonene and many other minor compounds contribute towards the difference in the smell from lemons and oranges[2]. The (+) enantiomer has a fresh smell with a citrusy note, while the (-) one has a turpentine note[2]. Limonene is categorized as generally regarded as safe (GRAS) by the U.S. Food and Drug Administration (FDA), opening interest in its use in eco-friendly cleaning agents and household products. Limonene is currently used in the food, fragrance and cosmetic industry as a preservative, flavor and scent[3–5]. Therapeutically, limonene has been used to solubilize cholesterol-containing gallstones[6] and it has also been widely tested for its anti-inflammatory, antioxidant, anticancer, antiviral, and gastroprotective properties, with emerging applications in many new health sectors like neuroprotection[7,8]. Additionally, potential applications in the agri-food industry as an antimicrobial, herbicidal, and antioxidant agent are being investigated[9]. Furthermore, its derivatives like menthol, perillyl alcohol, caverol and α-terpineol were also proven interesting for food, pharmaceuticals, biomaterials, cosmetics and biofuels due to their aroma and bio-physiological properties[10–13]. In particular, the recent demonstration of limonene as a precursor to valuable jet fuels[10] or its use as a solvent or polymer building blocks[14,15], together with a growing interest in the use of natural components for food and fragrance, have increased the limonene demand in the market. According to Grand View Research, the global limonene market was valued at USD 312.1 million in 2022 and is projected to reach USD 497.6 million by 2030, growing at a Compound Annual Growth Rate (CAGR) of 6.0% (https://www.grandviewresearch.com/industry-analysis/limonene-market-report).

Currently, extraction from citrus rinds, which are by-products of juice industries, is the primary source of limonene production[1]. Citrus fruits are produced in abundance; however, major producers are interested in the whole fruit market, e.g. 95% of total orange production in China is consumed as whole fruits[16]. Furthermore, seasonal fluctuations and diseases like Huanglongbing[17] strongly affect the supply and quality of fruits for limonene extraction. On top of that, part of the limonene produced from citrus fruits is not food-grade as it contains traces of pesticides used in agricultural fields[18], limiting application in household and food products, and thus providing avenues to explore alternative biobased production. Thanks to the rapid development of biotechnological tools, microbial production is emerging as an alternative solution with many advantages, ranging from the use of renewable substrates to greater product specificity, and generally involving gentle processing conditions to produce chemicals that can be considered natural[1,19].

Bacteria and yeast have been successfully employed as microbial limonene production hosts[20], *E. coli* and *S. cerevisiae* being the most studied. Limonene is biosynthesized from the universal terpene precursors, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), which are mainly produced via the mevalonate (MVA) pathway or the methylerythritol phosphate (MEP) pathway. Condensation of IPP and DMAPP leads to geranyl pyrophosphate (GPP) which is further converted to limonene via limonene synthases (LS). GPP is the canonical substrate of limonene synthases; however, studies in tomato species have found a small group of enzymes that uses the cis-isomer of GPP, neryl pyrophosphate (NPP), as a substrate to produce a limited number of monoterpenes[21]. A dedicated NPP synthase was found in *S. lycopersicum* (slNPPS), which can condense IPP and DMAPP to yield NPP[21]. An orthogonal approach utilizing NPP as a LS substrate for limonene synthesis has been recently reported in *S. cerevisiae*[22] and *E. coli*[23] with comparable

production levels to systems using GPP as a substrate. NPP can be advantageously accumulated in microbes and utilized towards limonene production, as it is not a natural substrate of microbial FPP synthase, thereby minimizing flux loss towards *E. coli*'s native terpene biosynthesis pathway and directing flux toward limonene. However, catalytic activity and substrate specificity of LSs towards the non-native substrate, NPP, also need to be optimized to improve limonene yields. Hence, it is promising to explore NPP as an orthogonal substrate for limonene synthases and optimize its catalytic activity and substrate specificity towards NPP through enzyme engineering to improve microbial limonene production.

To date, very few characterized Limonene synthases have been reported and they are often studied in different conditions and lacking details regarding their product selectivity. Among them, the two most studied are the limonene synthase from *Mentha spicata*, which has been identified as highly specific for (-)-limonene production[25], and the limonene synthase from *Citrus limon* and *Citrus sinensis* that are the preferred enzymes for (+)-limonene synthesis[24–26]. In spite of many attempts to mutate amino acid residues located near the active site[27,28] or use the NPP orthogonal pathway[22,28,29], The current highest limonene production titer was achieved using a codon-optimized and truncated version of the wild-type *Mentha spicata* limonene synthase[30]. Though there are other reported enantiospecific LSs, they have not been used as a basis for enzyme engineering[31–36].

In this study, we explored public sequence databases to select a set of 10 LS homologues that were further tested for limonene production in *E. coli*, using a 2-plasmid system developed to screen in parallel the MVA-GPP and MVA-NPP pathways. A limonene synthase from *Agastache rugosa*[37,38] was found to have a preference towards NPP as a substrate, with a limonene titer 11.4-fold higher compared to the GPP pathway. Guided by sequence and structure comparisons of

native limonene synthases, we aimed to better understand the molecular determinants involved in the substrate specificity, as well as to identify some critical residues for limonene production. These positions were targeted by mutagenesis, leading to the construction of 7 mutants that were tested for NPP- and GPP-pathways. The best mutant, containing 8 mutations in the catalytic domain, was found to improve limonene titers by 4.8 and 1.9-fold through the GPP and NPP pathways, respectively. Interestingly, this mutant still displays a preference for the NPP pathway, although to a reduced extent compared to its parental wild-type enzyme.

Materials and Methods

The list of plasmids and strains can be found in **Tables S1 and S2**. All structure related figures were generated using the PyMOL Molecular Graphics System, Version 2.5.5 Schrödinger, LLC. Graphs and statistical analyses were performed using GraphPad Prism version 10.0.1 for Windows, GraphPad Software, Boston, Massachusetts USA, www.graphpad.com. All P values showed results from unpaired t tests assuming Gaussian distribution, with ns, *, **, *** and **** correspond to $P > 0.05$, $P \leq 0.05$, $P \leq 0.01$, $P \leq 0.001$ and $P \leq 0.0001$, respectively.

2.1 Preparation of pEcoCTs-CPMS6a,6b,5 and psY7 strains

To generate pEcoCTs-CPMS6a, the first step was to delete the tcr promoter followed by trGPPS, LS and rrnB T1 and T2 terminators from the pJBEI-6409[39] plasmid. Using restriction-free cloning[40,41], pJBEI-6409 backbone without the deletion area was amplified by polymerase chain reaction (PCR) using iProof High-Fidelity DNA Polymerase (Bio-Rad), in which overlapping overhangs were created by primer design. 1 µL of *Dpn*I-treated PCR product was transformed by heat shock in *E. coli* DH5α, plated on a chloramphenicol (34 µg/mL) selective LB agar plate and incubated overnight at 37 °C. Three colonies were picked and inoculated overnight in 2 mL LB supplemented with chloramphenicol (34 µg/mL) at 37 °C, 300 rpm. Plasmids were extracted using E.Z.N.A. Plasmid Mini Kit I (Omega Bio-tek) and sequenced using Sanger sequencing (Bio Basic) to confirm the deletion. The plasmid was confirmed by performing restriction enzyme digestion (RE) and compared with *in silico* agarose gel simulation on SnapGene software (www.snapgene.com).

This intermediate plasmid was used as a backbone for the insertion of the tcr promoter, trGPPS, LS and rrnB T1 and T2 terminators in-between the p15A origin of replication and the chloramphenicol acetyltransferase expression site. For 2-piece cloning, primers containing

overhangs such that the backbone and inserts shared a 15 bp overlap region at the site of integration were designed. The backbone was PCR-amplified from the intermediate plasmid and insert was amplified from pJBEI-6409. Expected sizes were confirmed on an agarose gel and the PCR product was subsequently treated with *Dpn*I overnight at 37 °C; the E.Z.N.A. gel extraction kit (Omega Bio-tek) was used to purify the *Dpn*I-treated PCR products. 50 ng of each purified DNA backbone and inserts were added to a PCR tube with In-Fusion Snap Assembly Master Mix (Takara Bio). The mixture was incubated at 50 °C for 15 mins, and 1 µL was transformed by heat shock in *E. coli* DH5α, plated on a chloramphenicol (34 µg/mL) selective LB agar plate and incubated overnight at 37 °C. Colonies were analyzed by colony PCR, and their plasmid amplified, extracted, sequenced and RE digested as described above. Finally, LS was deleted from this second intermediate plasmid using restriction-free cloning, to obtain pEcoCTs-CPMS6a (**Figure 1b**).

pEcoCTs-CPMS6b was created by using pEcoCTs-CPMS6a as a backbone. The GPPS gene was first deleted, and 2-piece cloning was performed using In-Fusion assembly as described above, to assemble the resulting plasmid and a truncated NPPS from *S. lycopersicum* as an insert. pEcoCTs-CPMS5 was created by deleting GPPS from the pJBEI-3101[39] plasmid using adapted restriction free cloning described in the previous paragraph, and by changing the antibiotic to ampicillin (100 µg/mL). Full-length limonene synthase genes were obtained from Bio Basic in pEcoCTs-CPMS5. All LS truncations and mutants were obtained by designing primers with overhangs that contained the desired truncation or mutation in the overlap region and restriction free cloning.

For protein overexpression and purification, the genes of interest were cloned in psY7, a vector derived from pET-21d (Novagen). It incorporates an N-terminal maltose binding protein (MBP) tag, a His-tag, followed by an HRV 3C protease site and the protein of interest. Plasmids psY7-tArLS and -tNtLS were obtained by 2-piece cloning, using psY7 without the protein of interest as

a backbone, and tArLS and tNtLS as inserts. Then, the different mutants were obtained by restriction-free cloning.

2.2 Preparation of MG1655 *ispA* S80F by CRISPR

To perform the S80F mutation of *ispA* gene from K12 *E. coli* MG1655 strain, CRISPR Cas9-assisted recombineering method was used[42]. gRNA sequences were identified for wild-type Cas9 of 20 bp length in the target gene with reference to the MG1655 genome - ASM584v2 (*Escherichia coli* str. K-12 substr. MG1655) for SpCas9 protospacer adjacent motif (PAM), NGG on 3' side. A guide RNA sequence covering the site of mutagenesis was selected. p15A origin pTarget plasmid used in our previous work was used[42]. Plasmid pTarget-J23119-ispA-MG1655-HA-ispA-S80F was assembled using In-Fusion plasmid DNA assembly with pTarget plasmid, gRNA and homologous arm harboring mutation. CRISPR Cas9-assisted recombineering was then performed as previously described[42]. The mutation was checked by sequencing. The strains were stored in 40% glycerol at -80 °C before further use.

2.3 Limonene production and analysis

For limonene production experiments, competent cells of the *E. coli* K-12 MG1655 or MG1655 *ispA S80F* were prepared using the Mix & Go! *E. coli* Transformation Kit (Zymo Chem) following the manufacturer's specifications. Plasmids were then co-transformed by heat shock and plated on a LB agar plate supplemented with chloramphenicol (34 µg/mL) and ampicillin (100 µg/mL) and incubated overnight at 37 °C. Three freshly grown colonies were picked and inoculated in 1 mL LB media (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) with chloramphenicol (34 µg/mL) and ampicillin (100 µg/mL), and incubated at 37 °C, 300 rpm overnight. Each pre-culture was inoculated in 1 mL R-media (2 g/L ammonium sulphate, 4.2 g/L KH_2PO_4, 11.24 g/L K_2HPO_4, 1.86 g/L citric acid, and 10 mL/L of 100x trace element solution. The 100x trace element solution

consists of $CoCl_2 \cdot 6H_2O$ (0.25 g/L), $MnSO_4 \cdot 4H_2O$ (1.5 g/L), $CuSO_4 \cdot 2H_2O$ (0.15 g/L), H_3BO_3 (0.3 g/L), $Na_2MoO_4 \cdot 2H_2O$ (0.25 g/L), $Zn(CH_3COO)_2$ (0.8 g/L), Fe(III) citrate (5 g/L), and EDTA (0.84 g/L) at pH 8). Magnesium sulphate solution containing magnesium sulphate (0.5 g/L) and thiamine (0.0045 g/L) was supplemented to the media. The carbon sources were added (1 g/L glucose, 10 g/L glycerol), an inducer (2.5 mM lactose), and two times the usual working concentration of chloramphenicol (68 µg/mL) and ampicillin (200 µg/mL) to a starting optical density at 600 nm (OD_{600}) of 0.1 in a 14 mL-snap cap tube. A 200 µL dodecane overlay was added to each tube. No dodecane was used for the chiral column cultures, which were produced in solid phase microextraction (SPME) vials. Tubes were incubated at 30 °C, 300 rpm. After 72 hrs, the OD of the solution was measured, and the dodecane layer was extracted and diluted in hexane, for limonene to be quantified by liquid-injection gas chromatography/mass spectrometry (GC-MS, DB-5 column, for limonene quantification) or SPME GC-MS (Supelco 50/30µm DVB/CAR/PDMS, StableFlex (1 cm), Agilent 112-6632 Cyclosil-B column, for chiral analysis). (+) or (-) limonene (Sigma-Aldrich) were serial-diluted and used as standards for titer quantification. Limonene concentration in mg/L was calculated in the dodecane layer, then to the corresponding media volume. It was then divided by the OD_{600}, to obtain specific limonene (mg/L/OD).

For the expression analysis of limonene synthases in the lysate, the same protocol was followed, with single transformation of the pEcoCTs-CPMS5 plasmid of interest in MG1655, and the media were supplemented with ampicillin only. After 72 hours, the cultures were spun down, the cell pellet was resuspended in B-PER (ThermoFisher), vortexed for 10 minutes at room temperature, and 13 µL of the lysate was loaded into an SDS-PAGE.

2.4 Protein purification

For LS purification, psY7 containing the gene of interest (**Table S1**) was transformed in *E. coli* BL21-Gold (DE3) competent cells by heat shock. Proteins were overexpressed in ZYM media (9.6 g/L tryptone, 4.8 g/L yeast extract, 20 mM sodium phosphate dibasic, 25 mM potassium dihydrogen phosphate, 50 mM ammonium chloride, 5 mM sodium sulphate, 0.05% glucose, 0.5% glycerol, 2 mM magnesium sulphate, 0.02 mM iron (III) chloride). Briefly, a single colony was picked and incubated overnight in LB supplemented with 100 µg/mL ampicillin. 50 mL of ZYM media supplemented with 30 mM lactose and 100 µg/mL ampicillin was added to a 250 mL glass conical flask. Pre-culture was added to the media to a starting OD_{600} of 0.1. The flask was incubated at 30°C, 300 rpm for 24 hrs. Cells were harvested by centrifuging the culture at 3,900 rpm, 4 °C for 45 mins. The cell pellet was suspended in 2.5 mL of His-lysis buffer (50 mM Tris/HCl pH 8.0, 500 mM NaCl, 20 mM imidazole, 1 mg/mL lysozyme (BioBasic), 20 U/mL DNaseI (NEB) and 5% glycerol), and subsequently lyzed by freeze-thaw cycles. The mixture was frozen at -80 °C overnight. It was thawed at room temperature for 30 minutes and re-frozen at -80 °C for 1-2 hours, and the process was repeated twice, before thawing at room temperature to obtain the cell lysate. The lysate was centrifuged at 15,000 rpm at 4 °C for 45 mins. The supernatant was added to 1 mL of Ni-NTA beads (Cube Biotech), which were equilibrated in His-binding buffer (50 mM Tris/HCl pH 8.0, 500 mM NaCl, 20 mM imidazole, 5% glycerol), and incubated at 4 °C overnight with end-to-end mixing. The beads were washed with 40 mL His-binding buffer. An equal amount of protease buffer (50 mM Tris/HCl pH 8.0, 20 mM imidazole, 2 mM DTT, 5% glycerol), was added to the solution, as well as 30 µL (18 IU) of 3C protease, His, Human (Genscript). The solution was incubated at 4°C overnight with end-to-end mixing. This allowed to cleave out the protein of interest, while the MBP and His tags were still bound to the beads. The cleaved proteins were eluted with 3 x 0.5 mL His-binding buffer, assessed for purity with SDS-

PAGE, and concentration was checked using a spectrophotometer, using calculated $E_{1\%}$. Freshly purified enzymes were kept at 4 °C for a maximum of 4 days for kinetics experiments.

2.5 Kinetics experiments

The activities of purified limonene synthases were characterized using the EnzChek Pyrophosphate Assay Kit (ThermoFisher). GPP and NPP (Echelon Biosciences) were used as substrates. Each reaction was set up in a 200 µL system in a 96-well flat bottom plate (Greiner). Reaction consisted of 1x reaction buffer, 0.2 mM 2-amino-6-mercapto-7-methylpurine ribonucleoside, 1 U/µL purine nucleoside phosphorylase, 0.03 U/µL inorganic pyrophosphatase, 20 mM $MgCl_2$, to which was added either pyrophosphate standards (0, 10, 20, 30, 40 µM) or a fixed concentration of limonene synthase (0.016, 0.047, 0.124, 0.124, 0.047 µM for tArLS, tArLS-S8K, tArLS-loop1, tArLS-S8K-I265V-loop1 and tNtLS, respectively) and different concentrations of substrate (serial dilution from 80 µM to 5 µM, plus one well without substrate). It is worth noting $MgCl_2$ was supplemented as magnesium is required for limonene synthase activity, and 20 mM was reported to be optimal[25]. For each substrate, a negative control containing no enzyme was performed. All the components were incubated at 26 °C for 10 minutes before adding the pyrophosphate or substrate. Then, the absorbance at 360 nm (A_{360}) was immediately measured, every 30 seconds for 1 hour. For standard curve, A_{360} versus pyrophosphate concentration was plotted and slope of the linear fit was computed for the conversion of A_{360} to pyrophosphate produced in an enzymatic reaction. For enzymatic reactions, the background A_{360} values observed for controls without limonene synthase were subtracted to obtain net absorbance A_{360net}. A_{360net} values were converted to pyrophosphate produced using the slope of the calibration curve. Each experiment was performed in duplicate. Then, initial velocity or reaction rate (V in µM/min), for each substrate concentration, was determined by computing the slope of the linear

portion when plotting the produced pyrophosphate concentration as a function of reaction time. Finally, kinetics parameters (K_M, V_{max} and k_{cat}) were obtained by fitting the Michaelis Menten kinetic equation to the plot corresponding to substrate concentration ([Substrate]) over the initial Velocity (*V*). For each enzyme and substrate (GPP/NPP), the fitting was performed after merging all points (replicates) and using Gnuplot 5.2[43]. The software provided the mean and standard deviation for K_M and V_{max}. The uncertainty of k_{cat} was computed by dividing the protein concentration as done for the mean. However, the uncertainty of the catalytic efficiency (dcatEff) was calculated by using the error propagation formula: $dcatEff = catEff \times \sqrt{((dK_M/K_M)^2 + (dkcat/kcat)^2)}$, where dK_M and dkcat correspond to the uncertainties of Km and kcat respectively, and catEff is obtained by dividing kcat over Km.

2.6 Computational methods

Sequence analyses

Limonene synthase (LS) sequences were retrieved from the PFAM database by searching the domain architecture corresponding to PF1397-PF03936, as shown in both characterized LS, *Mentha spicata* (Uniprot ID: Q40322) and *Citrus sinensis* (Uniprot ID: A0A1C9J6A7). Altogether, 10840 sequences were recovered (December 2022). Among them, a majority was annotated as terpene synthases/cyclases. We selected only the ones reported as Limonene Synthase (LS) in their names, leading to 44 unreviewed and 8 reviewed sequences. Redundancy was removed using CD-HIT[44,45] leading to a dataset to 10 representative sequences. Among them, 4 reviewed sequences were characterized as (-) or 4S Limonene Synthase (LS), 2 reviewed sequences characterized and 2 unreviewed annotated as (+) or 4R LS, and two sequences without enantiomer characterization (**Table S3**). To analyze the residue conservation, these 10 sequences were aligned using MAFFT webserver[46] (https://mafft.cbrc.jp/alignment/server/index.html) and

Figure S1 was produced using ESPript / ENDscript 3.0 server[47], by providing the Multiple Sequence Alignement (MSA) from MAFFT and the AF2 model of *Mentha spicata* for secondary structure representation. AF2 model was preferred to the X-ray structures (PDB IDs: 2ONH or 2ONG[48]), because the whole sequence was modelled with a high confidence (PLDDT > 90 on average), except the N-terminal region (1-56 AA) corresponding to a disordered target peptide. Moreover, the two terpene synthase domains (PF1397 and PF03936) were well structurally aligned compared to the crystallographic structures (Cα-RMSD value around 0.4Å). The AF2 models of the 10 sequences were downloaded from the Uniprot database and were well superimposed on the known domains. Only N-ter regions were found to be disordered with low pLDDT values, leading us to experimentally test the truncated forms of the 10 sequences. Among them, both truncated sequences from *Agastache rugosa* and *Nepeta tenuifolia* were aligned for residue conservation analysis. Finally, AF2 model of *Agastache rugosa* was used as structural template for secondary structure representation (**Figure 3**). It is worth noting that the numbering of residues given for the mutants of tArLS (wild-type truncated form of *Agastache rugosa*), correspond to the new sequence without first 68 amino acids. Thus, the mutant S8K should correspond to S76K with the Uniprot numbering, for example.

MD simulations

Both tArLS and tArLS-S8K-I265V-loop1 (monomeric form) in complex with GPP were investigated using MD simulations. For the wild-type (tArLS), 3D coordinates were taken from the Uniprot database (ID: Q940E7), whereas the mutant tArLS-S8K-I265V-loop1 was modelled using colabfold webserver (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb) with default parameters. Three magnesium (Mg2+) ions in the active site and GPP were placed

after superimposition of X-ray structure from *Citrus sinensis* (PDB ID: 5UV1[26]) onto AF2 model. GPP coordinates were generated after structural alignment and replacement of the fluorine of the OFV ligand by a hydrogen atom. Protonation states of titrable residues were assigned using PROPKA webserver (https://server.poissonboltzmann.org/pdb2pqr), with a pH value of 7.5. These complexes (LS, Mg^{2+} and GPP) were solvated using gmx solvate tool of GROMACS version 2021.3[49] and then 20 mM of MgCl2 were added to neutralize the solvated systems. Charmm36m ForceField[50] (charmm36-jul2021) was used for modelling protein, ions and modified TIP3P water model, whereas CGENFF was used for GPP parameters. All MD steps, including minimization, equilibration (NVT, NPT) and production were run using GROMACS version 2021.3. Energy minimization was first performed using steepest descent algorithm with a maximum step number of 50,000 or a maximum force of 1000 for convergence. Then, heating (*NVT* ensemble) and equilibration (*NPT* ensemble) were run during 100 ps. Temperature coupling (300 K) and pressure (1 bar) were controlled using a velocity rescaling thermostat and Parrinello–Rahman algorithm, respectively. Finally, MD simulation was performed for 110 ns with a time step of 2 fs and snapshots were saved every 10 ps. For all stages, the following parameters were set up. LINCS algorithm was used to constrain the covalent bonds with hydrogen atoms. Short-range of non-bonded interactions (Lennard-Jones and Electrostatics) were treated with a cut-off of 0.12 nm. Long-range Lennard-Jones interactions were computed using a potential switching function from 0.1 to 0.12 nm, whereas electrostatic interactions were calculated using the particle mesh Ewald method on an approximately 0.1 nm grid with a fourth-order spline. The same protocol was used to investigate the complex of the mutant with NPP.

All MD analyses and structural visualization were performed using Gromacs tools and PyMOL 2.6 respectively. Binding free energies were estimated using gmx_MMPBSA[51]. The receptor was

defined as enzyme and Mg^{2+} ions, whereas the ligand corresponded to GPP or NPP. Some parameters for PB calculations were fine-tuned, such as PBradii set to 7 (due to CHARMM force field), and the internal constant dielectric (indi) set to 10 (due to metal ions). 100 snapshots were equally picked up during the last 50 ns of MD, for which the stability of the system was reached.

3. Results and discussion

3.1 Screening of limonene synthases

To facilitate LS screening and mutagenesis, two different plasmids were constructed. Both contain the mevalonate pathway genes, but one plasmid (pEcoCTs-CPMS6a) includes a truncated form of GPPS to accumulate GPP, whereas the second (pEcoCTs-CPMS6b) incorporates the gene encoding neryl pyrophosphate synthase (NPPS) to accumulate NPP. Each plasmid was co-transformed with a second plasmid only encoding LS (pEcoCTs-CPMS5) for limonene production. The key enzymes and reactions involved in MVA pathway and the production of limonene or other terpenes are given in **Figure 1a**, whereas the plasmids are detailed in **Figure 1b** and **Table S1**, and the strains in **Table S2**. Plasmids pEcoCTs-CPMS6a and 6b were cloned from the previously reported 1-plasmid system pJBEI-6409[39]. The repetition of the trc promoters of the pJBEI-6409 plasmid make it unstable, as plasmid recombination can happen during bacterial replication[52]. When such recombination occurs, half of the mevalonate pathway is not expressed, thus leading to very low limonene level, solely based on the endogenous GPP produced by *E. coli*. To tackle this issue, pEcoCTs-CPMS6a was built by deleting the tcr promoter, trGPPS, LS and the rrnB T1 and T2 terminators from the pJBEI-6409 plasmid, then re-introducing them, without LS, in-between the p15A origin of replication and the chloramphenicol acetyltransferase expression site. This way, if recombination occurs, either the origin of replication or the antibiotic resistance is lost, which makes the strain unable to survive. pEcoCTs-CPMS6b was then obtained by

substituting trGPPS by trNPPS in pEcoCTs-CPMS6a. pEcoCTs-CPMS5 was created by deleting GPPS from the pJBEI-3101[39] plasmid. pEcoCTs-CPMS6a and 6b can be used in parallel to screen for limonene production from either GPP or NPP, which provides a versatile platform to screen for LS from different organisms, as well as LS mutants (**Figure 1b**).

For bacterial production of (-)-limonene, limonene synthase from *Mentha spicata* is generally used[10,27,29,30,39,53–56], and *Citrus sinensis* or *Citrus limon* for (+)-limonene[22,28,57]. Here, we aimed to investigate LSs reported in the literature and compare them in the same conditions in terms of limonene production. First, we searched in the Pfam database[58] for terpene synthase domain, selected LS-annotated enzymes that were clustered to reach a total of 10 LSs (**Table S3**). Among them, 6 reviewed sequences were described in the literature to be enantioselective, either for (+)-limonene (MsLS, AbLS[59], PsLS[35], CsaLS) or (-)-limonene (CsLS, LaLS[60,61]) production (**Table S3**). Two unreviewed sequences (ArLS[37] and NtLS[62]) were annotated at transcript level. Finally, the last two sequences (CaLS[34] and PfLS) were annotated as reviewed and unreviewed LS, respectively, but without mentioning their enantioselectivity.

To confirm or characterize the enantioselectivity of the selected sequences, the 10 full-length LS encoding genes were cloned into the pEcoCTs-CPMS5 plasmid and co-transformed with pEcoCTs-CPMS6a, to consider the GPP route (natural substrate of LS) first. Limonene produced after cell growth was analyzed using a chiral column (**Figure 2a and Figure S2, left)**. For the high producers (MsLS, CsLS, ArLS and NtLS), the enantioselectivity was found to match those reported in the literature (**Table S3**). Unfortunately, for the low producers, low signal and background noise prevented from concluding on the enantioselectivity.

To improve the expression of plant LS and limonene production in microbes, truncation of LS was assessed in parallel. It was previously reported that truncating LS right before the pair of

arginine residues (MsLS Arg58-59) located after the N-terminal signal peptide improved limonene production, while truncating after Arg58 was detrimental to production[63]. Removing the N-terminal plastid-targeting sequence of limonene synthases improves solubility and catalytic efficiency in *E. coli* by preventing inclusion body formation, reducing chaperone interference, and eliminating kinetic impairment, thereby enhancing limonene production[63]. Furthermore, this region was predicted as very disordered by Alphafold2 for all modelled LSs. The 10 homologues were aligned using both sequence and structure alignments with MsLS as a sequence and structure (PDB 2ONG[48]) of reference (**Figure S1**), and they were truncated before the residue corresponding to MsLS Arg58. The 10 truncated LS coding genes were cloned into the pEcoCTs-CPMS5 plasmid. The 20 plasmids were individually co-transformed with pEcoCTs-CPMS6a to assess limonene production from each of them via the GPP pathway. As can be seen from **Figure 2b**, for all LSs – except LaLS – truncating the N-terminal peptide significantly improved limonene production. This was the case for low limonene producers (AbLS, CsaLS, LaLS, CaLS and PfLS) as well as higher producers (MsLS, PsLS, CsLS, ArLS and NtLS). This emphasizes the rate-limiting effect of LS in the limonene production pathway. Moreover, the truncation improved production regardless of the limonene enantiomer produced. We found that truncated MsLS (tMsLS) was producing (-)-limonene with a 354-fold increase compared to its full-length counterpart, and tCsLS had a 305-fold (+)-limonene production compared to CsLS (**Figures 2b**). This screening did not reveal any LS capable of producing limonene with a titer higher than that obtained with tMsLS and tCsLS for (-)- and (+)-limonene, respectively. Limonene produced with truncated LSs was analyzed by a chiral column, which allowed to confirm the enantioselectivity ((+)-limonene) of two more LSs, PsLS and CsaLS, which are predominantly producing (-)-limonene (**Figure S2, center**).

Then, using only the truncated limonene synthases, we compared the production of limonene through the GPP and NPP pathways, by co-transforming pEcoCTs-CPMS5 with pEcoCTs-CPMS6a or 6b in parallel (**Figure 2c**). Again, tMsLS was the highest producer for (-)-limonene, and tCsLS for (+)-limonene. Interestingly, the production was higher *via* the GPP pathway for tMsLS, whereas we found no significant difference for tCsLS. For the low limonene producers (tAbLS, tCsaLS, tLaLS, tCaLS and tPfLS), a significant improvement of the limonene production from NPP was observed. This could be due to the absence of competition for substrate NPP from endogenous *E. coli* enzymes, or a faster production due to the straightforward production of the neryl cation, essential for limonene cyclization as suggested by Kumar and colleagues[26]. Limonene obtained from NPP was analyzed using a chiral column, which led to a noteworthy observation for all tested sequences: a LS specific for the production of one limonene enantiomer from GPP produced 10 to 15% of the opposite enantiomer from NPP (**Figure S2, right**). The production of the opposite enantiomer when using NPP as a substrate could be due to subtle differences in the active site conformation, which may allow for alternative binding orientations or transition states that allow the formation of both enantiomers, unlike the more rigid or specific interactions with GPP that lead to higher enantiopurity. Indeed, it has been experimentally shown that the fluorinated analogue of NPP was more flexible inside (+)-limonene synthase *Citrus sinensis*, leading to the presence of two main conformers (PDB ID: 5UV2[26]). However, in this work, they also showed that the addition of Mn^{2+} and NPP do not change the enantioselectivity. So, another explanation for this enantioselectivity ratio change may come from the presence of Mg^{2+}, in our work. In the literature, it has already been shown that the type of metal ion can change the enantiomer ratio[64]. Thus, NPP route improves limonene production but may slightly affect

enantioselectivity, which can be detrimental when aiming for industrial production of one of the two limonene enantiomers.

3.2 Exploring molecular determinants involved in substrate specificity

Notably, the NPP pathway did not improve limonene production to the same extent for all 10 LSs, indicating there are substrate preferences for each LS. In particular, tArLS[39] and tNtLS[62], which share a high sequence identity (88%) (**Figure 3a**), showed a 11.5-fold and 2.0-fold improvement in limonene production levels, respectively, when using the NPP pathway (**Figure 2c**). Thus, both sequences were selected as model enzymes to further investigate the possible determinants at the origin of substrate preference.

Structural analysis was performed by comparing 3D-models of both LSs. Surprisingly, the active site residues (first shell around GPP) are found to be fully conserved between the two homologues, indicating that production performances and selectivity could come from residues beyond the active site (**Figure 3b**). We thus analyzed further the surrounding catalytic domain to search for differences in terms of amino acid composition. In particular, we identified 4 residues and 1 loop that differed between the 2 enzymes, as highlighted in **Figure 3a and c**: S8, I17, I265, E276-I286 (hereafter referred to as "loop1") and P530. These key residues are either located in the C-terminal RR(X_8)W region[65] (S8, I17), or in the N-terminal catalytic domain (I265, loop1, P530).

[6626]To validate the role of highlighted key residues between tArLS and tNtLS, single or multiple mutants were built by introducing tNtLS' residues in tArLS: S8K, I17M, I265V, loop1 (E276P, P277R, A281K, N282T, I285Q, I286L) and P530A. **Figures 4a and b** display the limonene production for each mutant using GPP or NPP as substrate. Two single mutations (S8K and I265V), as well as the loop1 substitutions, increased limonene production from GPP and NPP. Mutants S8K-loop1 and S8K-I265V-loop1, which combine the previous beneficial mutations,

showed improved ability to produce limonene. The mutant S8K-I265V-loop1 led to a limonene titer enhanced by 4.8-fold from GPP and 1.9-fold from NPP. Nevertheless, limonene titers obtained in *E. coli* whole cell production with mutant S8K-I265V-loop1 did not reach that of tNtLS which was still 1.4-fold higher from substrate NPP (**Figure 4b and Figure 4c)**. Interestingly, the limonene production gap between tNtLS and tArLS limonene production from GPP was greatly reduced by the introduction of the mutations into wild-type tArLS, from 14.8 - to 3.1-fold. (**Figure 4a and Figure 4c**). These mutations in tArLS also reduced the difference between limonene titer obtained from NPP and GPP (**Figure 4d**). Indeed, the ratio of limonene produced from NPP and that from GPP (NPP/GPP ratio), reflecting substrate specificity *in vivo*, dropped from 11.2 (wild-type tArLS) to 4.6 (tArLS- S8K-I265V-loop1), reaching a value close to the one of tNtLS (2.0). Thus, these results indicate that selected residues, surrounding the conserved active site, participate in the limonene production using either GPP or NPP as substrate. To confirm the role of the loop1 in limonene production, we swapped this region from tArLS to tNtLS (P280E-R281P-K285A-T286N-Q289I-L290I). The corresponding tNtLS mutant showed a significant decrease of limonene production from both GPP and NPP (**Figure 4c**). Interestingly, it also increased the NPP/GPP ratio to 3.4 (**Figure 4d**), showing that the swap of this region is key to recover the specificity of either tArLS or NtLS, but it is not sufficient.

To go further in the characterization of both enzymes (tArLS and tNtLS), purification (**Figure S3**) and enzymatic assays of wild-type and selected tArLS mutants were performed, to determine their kinetics parameters (K_M, k_{cat}, K_M/k_{cat}) (**Table 1** and **Figure S4**). For tArLS, the binding efficiency was found to be higher toward GPP than NPP, but with a lower activity, resulting in a catalytic efficiency (k_{cat}/K_M) 2.4-fold higher for NPP than GPP, in agreement with the results obtained *in vivo*. tNtLS also showed a catalytic efficiency increased by 1.8-fold and 1.3-fold

compared to that of tArLS. This appears mainly due to a reduced K_M for both GPP and NPP and is also in agreement with limonene production levels obtained *in vivo*. Of note, the *in vivo* NPP/GPP ratio of tArLS (11.2) is much higher than the ratio between the catalytic efficiency with NPP and that with GPP (2.39). These differences between *in vitro* characterization and *in vivo* production could be due to the low concentration of GPP present *in vivo*. Indeed, GPP is also a substrate for FPP synthase competing with LS, whereas NPP can be metabolized only by LS, and the 2 substrates can also be produced at different levels. To test this hypothesis, we constructed the MG1655 *ispA* S80F strain using CRISPR-Cas9[67] asymmetric homology arms recombineering[42] to weaken the GPP-to-FPP activity of the FPP synthase, and hence allowing GPP to accumulate[10,68]. We hypothesized that this would allow tArLS to produce more limonene. Using either wild-type MG1655 or MG1655 *ispA* S80F as hosts for the pEcoCTs-CPMS6a and pEcoCTs-CPMS5-tArLS plasmids, we compared the limonene production and found a 1.6-fold improvement in limonene production in the latter strain (**Figure 4e**). This supports our hypothesis that the competing activity of FPP synthase contributes to diminishing GPP availability and widens the difference in NPP/GPP ratio obtained with tArLS, which seems applicable to tNtLS as well.

The mutants S8K, loop1 and S8K-I265V-loop1 show a gradually increasing binding affinity (lower K_M). Nevertheless, their k_{cat} and therefore, their catalytic efficiency values were decreased compared to the wild-type enzyme. One can also note that these mutations dropped down their NPP/GPP catalytic efficiency ratio (1.74, 1.15 and 1.36 respectively) to the level of tNtLS ratio (1.66).

The expression levels of the different mutants expressed from pEcoCTs-CPMS5 plasmid were subsequently checked by SDS-PAGE (**Figure S3b**). The tArLS wild-type enzyme and its mutants cannot be seen on the SDS-PAGE and are poorly expressed compared to tNtLS. However,

quantification of purified protein revealed measurable yields: 27 µg of wild-type tArLS was obtained, while the S8K, loop1, and S8K-I265V-loop1 mutants yielded 311 µg, 206 µg, and 290 µg, respectively, under identical conditions. This suggests that the mutant variants may be expressed at higher levels than the wild-type enzyme, despite remaining below the detection limit of SDS-PAGE. It is interesting to note that tNtLS-loop1 mutant also shows a lower expression than wild-type tNtLS. Differences in enzyme solubility and stability, mRNA stability and transcription, translation differences may explain these results. The difference in expression levels *in vivo* contributes to the disparity in limonene yield observed between tArLS and tNtLS. Improving expression could certainly further enhance limonene production with tArLS.

3.3 Exploring the dynamics of tArLS and tArLS-S8K-I265V-loop1

To better understand the molecular determinants involved in the differences between tArLs and the tArLS-S8K-I265V-loop1 mutant, MD simulations of the two monomeric enzymes in complex with Mg^{2+} and GPP were performed during 110 ns. To be consistent with results of native gel (**Figure S5**), monomeric forms were also modelled and simulated. Protein and GPP dynamics were further investigated. The stability of the studied enzymes was monitored by plotting Root Mean Square Deviation (RMSD) of backbone atoms over MD time (**Figure S6a**). Both enzymes were stable after 110 ns of MD simulation, with a plateau of RMSD values around 2 Å, indicating that no significant conformational change was observed. Wild-type tArLS was found to reach the plateau faster than the mutant, 50 ns versus 80 ns respectively. Moreover, the fluctuations of RMSD values seem to be higher for tArLS compared to the mutant. Altogether, these results indicated a higher flexibility for the wild type compared to the mutant. This latter was investigated by computing Root Mean Square Fluctuations (RMSF) of backbone atoms, which were averaged per residue (**Figures S6b**). The results show that wild-type tArLS is more flexible than the mutant.

The main differences were observed in the terpene synthase N-terminal domain (PFAM ID: PF01397), located at the bottom of the active site. In addition, the highest difference was seen around the residue I325 in the terpene synthase metal binding domain (PFAM ID: PF03936). Regarding the effect of the mutations, all of them were found to enhance flexibility of the mutant. Indeed, the mutations S8K and I265V induced a higher flexibility on the upstream residues, R2-Y7 and F252-P258, respectively. Interestingly, loop1 showed a higher mobility, especially for residues H280 and A281K, with RMSF values of 0.7 Å (WT) versus 1.5 Å (mutant) and 0.6 Å (WT) versus 1.7 Å (mutant), respectively. To better localize these key regions, B-factor values were mapped onto both enzyme structures (**Figure S6c**).

Moreover, the dynamics of ligands (Mg^{2+} and GPP) were investigated for both enzymes, tArLs and the mutant tArLS-S8K-I265V-loop1. All pairwise distances between the three Mg^{2+} of the active site were computed and plotted over MD time (**Figure S7a**). Each distance corresponds to two Mg^{2+} ions coordinated by the following residues: D302/D298/E376/D455 for d1, D302/D455/E450/T446/D442 for d2, and E376/E450/T446/D442 for d3. The results showed that all distances were stabilized after 40 ns for both enzymes, meaning that Mg^{2+} ions are well coordinated to stay in the active site. Both distances d1 and d3 displayed a similar profile between wild-type and mutant, with an increase of the distance until a plateau around 4.3 Å and 5.9 Å for d1 and d3, respectively. Regarding the fluctuations of the distance d2, the values were stabilized around 4.75 Å and 5 Å for wild-type and mutant, respectively. Interestingly, the decrease of d2 observed in wild type seems to be correlated with the jump of d1 (around 25 ns), however this behavior was not seen in the mutant. These results indicate that the dynamics of Mg^{2+} ions is slightly different between both enzymes, what could play a role in the efficiency of the catalysis but would require in-depth mechanism investigation. The molecular catalytic process is also driven

by the conformational sampling of ligand, here the geranyl diphosphate (GPP). Thus, the dynamics of GPP was investigated for both tArLS and tArLS-S8K-I265V-loop1 enzymes. To verify the anchoring of the pyrophosphate head of GPP, distances between each phosphorus atom (PA or PB) and the center of gravity (COG) of Mg^{2+} ions were computed (**Figure S7b**). For both enzymes, these distances were stabilized after 40 ns with values around 0.8-0.65 Å (COG-PB) and 2.6-2.5 Å (COG-PA) for wild-type and mutant, respectively. Thus, these results show that the pyrophosphate head is well anchored between the three Mg^{2+} ions, and the flexibility of GPP should mostly come from its polyisoprenoid tail. To investigate the mobility of the polyisoprenoid tail, the distance between C1 (beginning of the tail) and C8 (end of the tail) was computed and monitored over MD time (**Figure S8**). Interestingly, GPP was able to adopt several conformations ranging from extended (> 7 Å) to cyclic (< 5 Å) for both enzymes. However, the wild type showed higher fluctuations compared to the mutant, indicating that the flexibility of the GPP and therefore the flexibility of the enzyme seems to be higher. This observation is in agreement with the B-factor analysis described in **Figure S6c**.

Finally, molecular interactions were studied between GPP and both enzymes. Binding free energy calculations were performed on the last 50 ns of MD for the stability of the systems and using MMPBSA approach[51]. It resulted in a binding affinity ($\Delta G^{binding}$) of -43.27 kcal/mol ± 9.02 and -38.37 kcal/mol ± 7.19 for mutant and wild-type, respectively. These results are in agreement with experimental data where the mutant has a higher binding affinity compared to the wild-type, i.e. K_M of 8.0 µM ($\Delta G^{binding}$=-7 kcal/mol for the mutant) versus K_M of 33.32 µM ($\Delta G^{binding}$=-6.15 kcal/mol for the wild-type). Of note, MMPBSA value corresponds only to the enthalpy of $\Delta G^{binding}$. To go further in the molecular understanding, the binding free energy was decomposed per residue and plotted onto 3D structure for both enzymes (**Figure S9a**). The mapping showed that almost

all residues contributed to the binding of GPP in a favorable (blue) or unfavorable (red) way, for both enzymes. The long-range effect (residues far from the active site) is probably due to the electrostatic of Mg^{2+} metal ions, for which the contributions are the highest. As expected, the mutant showed on average more favorable or neutral contributions leading to a lower total binding free energy. Strikingly, for both enzymes, residues lining the active site were unfavorable, with higher values for the wild type. Regarding the effect of mutations on energy contribution, the decomposition per residue located in loop1 was investigated (**Figure S9b**). Interestingly, almost all mutations changed into favorable energy contributions, showing the importance of this region. This result is in agreement with the observed increase of binding affinity of GPP, when comparing the wild-type (K_M = 33.32 µM) with the mutant tArLS-loop1 (K_M = 10.8 µM). Lastly, we investigated the interactions of GPP in the active site for the wild-type and the mutant. To illustrate the main interactions in the active site, one example of each GPP conformation (extended or cyclic) is shown in **Figure 5** for both enzymes. In spite of the flexibility of GPP, same key residues highlighted in the literature[57] were observed in contact (distance threshold <5 Å) with GPP for both enzymes, i.e. W270, T295, Y519, H525. The three amino acids W270, T295 and H525 are organized as a platform to stabilize the polyisoprenoid tail of GPP through van der Waals interactions, on one side of the active site. A hydrogen bond network can be observed in the triad as shown with blue dashed lines in **Figure 5**. It is worth noting that the other side of the active site is stabilized by hydrophobic residues such as I291, I294, V399, V404, L438 and L511. Xu *et al.* showed that Y573 and D496 in *Mentha spicata* were crucial in early stage (before cyclization) of limonene production[56]. The importance of potential key hydrogen bond between these two residues was reported, corresponding to equivalent residues D442 and Y519 in tArLS (wild-type or mutant). In our MD simulations, these two residues were found in contact (< 5 Å), but not forming

direct hydrogen bond. One explanation could come from the fact that the Mg^{2+} ions were modelled explicitly in our work, unlike in Xu *et al*.[56]. The residue Y519 was observed to be shared between D442 and the oxygen O1 of GPP, with distances around 4 Å from its hydroxyl. From this observation, the hydroxyl of the conserved tyrosine (Y519) could stabilize the pyrophosphate after cleavage. Surprisingly, amongst these key residues, only H525 was shown as contributing favorably to the GPP binding (see MPBSA tables in **Figure 5**). Indeed, Mg^{2+} ions contributed the most favorably with very low energy values (< -600 kcal/mol). However, the mutant showed less favorable energy values compared to the wild-type, especially for MG3. Thus, this result could explain the decrease of k_{cat} in the mutant for GPP (**Table 1**). Regarding NPP binding, similar trend to that of GPP was observed with an enhanced binding affinity ($\Delta G^{binding}$) of -51.93 kcal/mol ± 6.29 for NPP in the mutant, versus -38.37 kcal/mol ± 7.19 for GPP (**Figure S10**).

[56] This study aimed at identifying and characterizing LSs of potential interest for microbial limonene production in *E. coli*. In particular, our aim was to explore the main molecular determinants responsible for their substrate specificity in order to promote the use of NPP rather than GPP as substrate, to avoid i) loss in GPP precursor through the action of enzymes competing with LS and ii) perturbation of cell metabolism due to the consumption of GPP by LS. Implementing an orthogonal pathway leading to NPP and including a LS highly specific for NPP would be of prime interest. This approach could help to remove some of the obstacles to increasing limonene production levels. Our strategy relied on the exploration of natural diversity using sequence-based mining that enabled to identify of 10 LS candidates, that were further screened *in vivo* using newly developed screening assays allowing to evaluate their preference for either GPP or NPP precursors. Several enzymes were found to display a preference for NPP: tAbLS, tCsaLS, tArLS, tNtLS, tLaLS, tCaLS and tPfLS. To the best of our knowledge, *Agastache rugosa*

limonene synthase (tArLS) is the first LS reported with a pronounced preference for NPP, with 1.7 and 19.6 mg/L/OD limonene produced from the GPP and NPP pathways, respectively. Sequence, structure and dynamics investigation of tArLS allowed to highlight key amino acid residues potentially important for substrate binding affinity and catalytic activity of the enzyme, that were further evaluated by mutagenesis. Some of these promising positions could be considered in the future to increase preference for NPP precursor of high limonene producing enzymes, such as those from *Mentha spicata* or *Citrus sinensis.* This work sets the ground for the development of more efficient and specific LSs to improve the bioproduction of limonene.

[4748]AUTHOR INFORMATION

Author Contributions

C.P.M.S., S.B. and C.S. performed the experiments; C.P.M.S. and X.C. analyzed the data; J.E. and I.A. performed MD and bioinformatics analysis; X.C. and I.A designed the study; All authors contributed to the manuscript writing and review; All authors approved the final version of the manuscript.

Funding Sources

This work is funded by the EcoCTs project with the support of the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme. We would like to acknowledge the funding support from Singapore Institute of Food and Biotechnology Innovation (SIFBI), and Central Research Fund (Applied / Translational Research) from A*STAR.

Notes

Conflict of interest: Singapore Application No.10202402813S, filed at the Intellectual Property Office of Singapore.

ACKNOWLEDGMENT

We thank Taek Soon Lee for the pJBEI-6409 plasmid, through Addgene plasmids #47048 and #50603. Authors would like to thank the regional computing meso center CALMIP (Grant P19013 and P22027) for granting access to High Performance Computing resources.

ASSOCIATED CONTENT

Supporting Information.

Sequence alignment of 10 limonene synthase homologues; chiral chromatograms of limonene synthases; enzyme purification and expression analysis; kinetics analysis of tArLS mutants and tNtLS; native gel of purified limonene synthases; molecular dynamics analysis of protein stability; Mg^{2+} stability in active site; GPP conformational sampling; binding free energy calculations of GPP for tArLS and tArLS-S8K-I265V-loop1; key interactions for GPP and NPP binding to wild-type and mutant tArLS; tables of plasmids, strains, and limonene synthases used.

Figure S1. Limonene synthase 10 homologues alignment (ESPript / ENDscript 3.0 server[47]), with MsLS as a sequence and structure (PDB 2ONG[48]) of reference.

Figure S2. Chiral column chromatograms of different limonene synthases. For each LS, the full-length sequence was analyzed through the GPP pathway (left), and the truncated enzymes were analyzed through the GPP (middle) or NPP pathway (right). (+) and (-) limonene peaks are indicated.

Figure S3. Enzyme purification and expression analysis (a) SDS-PAGE of the purified tArLS mutants, pooled after affinity chromatography. (b) SDS-PAGE of purified tArLS and NtLS after affinity chromatography (1, 11). Expression analysis in expression conditions of the WT and mutant enzymes (2-9). Lanes 12 is an empty-plasmid control and 13 a no-induction control. For lanes 2 to 8, there is not band detected for the expression of tArLS, which explains low limonene production, where as tNtLS can be observed on lanes 9 and 10.

Figure S4. Kinetics analysis of tArLS mutants and tNtLS. V (µM/min) was obtained from the measurement of pyrophosphate production over time and plotted against the substrate concentration [S], either GPP (left) or NPP (right). Michaelis Menten kinetic equation was used to

calculate K_M and k_{cat} of each enzyme. Enzyme concentrations are 0.016, 0.047, 0.124, 0.124, 0.047 μM for tArLS, tArLS-S8K, tArLS-loop1, tArLS-S8K-I265V-loop1 and tNtLS, respectively.

Figure S5. Native gel of purified limonene synthases. Purified tArLS and tNtLS were run on a NativePAGE Bis-Tris (Thermofisher) native gel. They were found to be monomers (MW of tArLS and tNtLS are 64.2 and 64.3 kDa, respectively).

Figure S6. Protein stability. (a) RMSD on backbone atoms plotted as a function of MD time for the two enzymes, tArLS-WT (violet) and the mutant tArLS-S8K-I265V-loop1 (green). (b) RMSF per residue computed from backbone atoms for both enzymes. The red arrows and transparent rectangle highlight the mutated regions, whereas orange arrows show the regions with high differences of flexibility in favor of the wild-type (WT). PFAM domains are given as blue (PF01397) and yellow (PF03936) rectangles spanning the residue range as defined by PFAM database. (c) B-factors mapped on 3D structures of both enzymes, tArLS-WT and the mutant tArLS-S8K-I265V-loop1. The enzymes are shown as b-factor putty, where higher flexibility is displayed as larger radius and red cartoon representation. Key regions shown in (b) are also highlighted on the 3D structures.

Figure S7. Magnesium (Mg^{2+}) stability in the active site. (a) On the top, Mg^{2+} coordination is displayed with ions in green spheres, amino acid residues in cyan sticks (only side chain) and GPP in violet sticks. Coordinated residues and GPP are labelled, and the computed distances between Mg^{2+} ions are defined as d1, d2 and d3. On the bottom, the three distances are plotted as function of MD time. (b) On the top, the anchoring of the pyrophosphate head of GPP is illustrated with ions in transparent green spheres and GPP in violet sticks. The small lightblue sphere represents the center of gravity (COG) of the three Mg^{2+}ions and the phosphorus atoms (PA and PB) of the

pyrophosphate are shown in spheres and labelled. On the bottom, the distance between COG of Mg^{2+}ions and PA (or PB) is plotted as a function of MD time.

Figure S8. GPP conformational sampling. Distance between C1 and C8 of GPP is plotted as a function of MD time. Example of an extended and cyclic conformation is shown on the right with GPP represented as sticks.

Figure S9. Binding free energy calculation of GPP for both enzymes, tArLS and tArLS-S8K-I265V-loop1. (a) MMPBSA results are mapped on each 3D structure of wild-type (on the top) and the mutant (on the bottom). The color scale ranges from blue (lower than -20kcal/mol) to red (upper than 20kcal/mol). The enzyme is shown in cartoon, whereas GPP is in sticks and Magnesium ions in balls. (b) Highlight of the loop1 contribution in the binding free energy for wild-type (on the top) and the mutant (on the bottom). The enzymes are shown in grey cartoon, except for the loop1 colored as in (a). GPP is displayed in magenta sticks, whereas Mg^{2+}ions are in green spheres.

Figure S10. Key interactions for GPP and NPP binding to wild-type tArLS and tArLS-S8K-I265V-loop1 mutant, respectively. On the left, one snapshot with an extended conformation of GPP or NPP is represented, whereas cyclic conformations are displayed on the right. The Mg^{2+}ions are shown in green spheres and GPP or NPP in magenta/palecyan sticks with labelling of O1 (pyrophosphate head) and C3 (polyisoprenoid tail). Key residues(Xu et al., 2018) are labelled for each snapshot. Blue dashed lines on left (a) represent potential hydrogen bonds. On the right, tables provide MMPBSA results of Mg^{2+}ions and key residues for both enzymes.

Table S1. Plasmids used in this study

Table S2. Strains used in this study

Table S3. Limonene synthases used in study

ABBREVIATIONS

AbLS, *Abies grandis* (Grand fir) limonene synthase; AcAc-CoA, acetoaceyl-CoA; ArLS, *Agastache rugosa* (Korean mint) limonene synthase; AtoB, Acetyl-CoA acetyltransferase; CsaLS, *Cannabis sativa* (Hemp) limonene synthase; CsLS, *Citrus sinensis* (Sweet orange) limonene synthase, CaLS, *Coffea arabica* (Arabian coffee) limonene synthase; DMAPP, dimethylallyl pyrophosphate; FPPS, farnesyl pyrophosphate synthase; GGPP, geranylgeranyl pyrophosphate; GGPPS, GGPP synthase; GPP, geranyl pyrophosphate; GPPS, geranyl pyrophosphate synthase; HMG-CoA, 3-hydroxy-3-methylglutaryl-CoA; HMGR, HMG-CoA reductase; HMGS, HMG-CoA synthase; IDI, isopentenyl-diphosphate isomerase; IPP, isopentenyl pyrophosphate; LaLS, *Lavandula angustifolia* (Lavender) limonene synthase; LS, limonene synthase; MBP, maltose binding protein; MsLS, *Mentha spicata* (Spearmint) limonene synthase; MEP, methylerythritol phosphate; MK, MVA kinase; MVA, mevalonate, MVA-5-P, mevalonate-5-phopshate; NtLS, *Nepeta tenuifolia* limonene synthase; NPP, neryl pyrophosphate; NPPS, neryl pyrophosphate synthase; PfLS, *Perilla frutescens* var. frutescens limonene synthase; PsLS, *Picea sitchensis* (Sitka spruce) limonene synthase; PMD, pyrophosphomevalonate decarboxylase; SPME, solid phase microextraction; STS, sesquiterpene synthase; PCR, polymerase chain reaction

REFERENCES

(1) Jongedijk, E.; Cankar, K.; Buchhaupt, M.; Schrader, J.; Bouwmeester, H.; Beekwilder, J. Biotechnological Production of Limonene in Microorganisms. *Appl Microbiol Biotechnol* **2016**, *100* (7), 2927–2938. https://doi.org/10.1007/s00253-016-7337-7.

(2) Kvittingen, L.; Sjursnes, B. J.; Schmid, R. Limonene in Citrus: A String of Unchecked Literature Citings? *J Chem Educ* **2021**, *98* (11), 3600–3607. https://doi.org/10.1021/acs.jchemed.1c00363.

(3) Espina, L.; Gelaw, T. K.; de Lamo-Castellví, S.; Pagán, R.; García-Gonzalo, D. Mechanism of Bacterial Inactivation by (+)-Limonene and Its Potential Use in Food Preservation Combined Processes. *PLoS One* **2013**, *8* (2). https://doi.org/10.1371/journal.pone.0056769.

(4) Ravichandran, C.; Badgujar, P. C.; Gundev, P.; Upadhyay, A. Review of Toxicological Assessment of D-Limonene, a Food and Cosmetics Additive. *Food and Chemical Toxicology* **2018**, *120*, 668–680. https://doi.org/https://doi.org/10.1016/j.fct.2018.07.052.

(5) Vatankhah Lotfabadi, S.; Mortazavi, S. A.; Yeganehzad, S. Study on the Release and Sensory Perception of Encapsulated D-Limonene Flavor in Crystal Rock Candy Using the Time–Intensity Analysis and HS-GC/MS Spectrometry. *Food Sci Nutr* **2020**, *8* (2), 933–941. https://doi.org/https://doi.org/10.1002/fsn3.1372.

(6) Sun, J. *D-Limonene: Safety and Clinical Applications*; 2007; Vol. 12.

(7) Vieira, A. J.; Beserra, F. P.; Souza, M. C.; Totti, B. M.; Rozza, A. L. Limonene: Aroma of Innovation in Health and Disease. *Chem Biol Interact* **2018**, *283*, 97–106. https://doi.org/https://doi.org/10.1016/j.cbi.2018.02.007.

(8) Eddin, L. B.; Jha, N. K.; Meeran, M. F. N.; Kesari, K. K.; Beiram, R.; Ojha, S. Neuroprotective Potential of Limonene and Limonene Containing Natural Products. *Molecules*. MDPI AG August 1, 2021. https://doi.org/10.3390/molecules26154535.

(9) Ibáñez, M. D.; Sanchez-Ballester, N. M.; Blázquez, M. A. Encapsulated Limonene: A Pleasant Lemon-like Aroma with Promising Application in the Agri-Food Industry. A Review. *Molecules*. MDPI AG June 1, 2020. https://doi.org/10.3390/molecules25112598.

(10) Mendez-Perez, D.; Alonso-Gutierrez, J.; Hu, Q.; Molinas, M.; Baidoo, E. E. K.; Wang, G.; Chan, L. J. G.; Adams, P. D.; Petzold, C. J.; Keasling, J. D.; Lee, T. S. Production of Jet Fuel Precursor Monoterpenoids from Engineered Escherichia Coli. *Biotechnol Bioeng* **2017**, *114* (8), 1703–1712. https://doi.org/10.1002/bit.26296.

(11) Ciriminna, R.; Lomeli-Rodriguez, M.; Demma Carà, P.; Lopez-Sanchez, J. A.; Pagliaro, M. Limonene: A Versatile Chemical of the Bioeconomy. *Chemical Communications* **2014**, *50* (97), 15288–15296. https://doi.org/10.1039/C4CC06147K.

(12) Johnson, T. J.; Jahandideh, A.; Johnson, M. D.; Fields, K. H.; Richardson, J. W.; Muthukumarappan, K.; Cao, Y.; Gu, Z.; Halfmann, C.; Zhou, R.; Gibbons, W. R. Producing Next-Generation Biofuels from Filamentous Cyanobacteria: An Economic Feasibility Analysis. *Algal Res* **2016**, *20*, 218–228. https://doi.org/https://doi.org/10.1016/j.algal.2016.10.020.

(13) Thomsett, M. R.; Moore, J. C.; Buchard, A.; Stockman, R. A.; Howdle, S. M. New Renewably-Sourced Polyesters from Limonene-Derived Monomers. *Green Chemistry* **2019**, *21* (1), 149–156. https://doi.org/10.1039/C8GC02957A.

(14) Kleybolte, M. M.; Zainer, L.; Liu, J. Y.; Stockmann, P. N.; Winnacker, M. (+)-Limonene-Lactam: Synthesis of a Sustainable Monomer for Ring-Opening Polymerization to Novel, Biobased Polyamides. *Macromol Rapid Commun* **2022**, *43* (17), 2200185. https://doi.org/https://doi.org/10.1002/marc.202200185.

(15) Durkin, A.; Taptygin, I.; Kong, Q.; Gunam Resul, M. F. M.; Rehman, A.; Fernández, A. M. L.; Harvey, A. P.; Shah, N.; Guo, M. Scale-up and Sustainability Evaluation of Biopolymer Production from Citrus Waste Offering Carbon Capture and Utilisation Pathway. *ChemistryOpen* **2019**, *8* (6), 668–688. https://doi.org/https://doi.org/10.1002/open.201900015.

(16) Paggiola, G.; Stempvoort, S. Van; Bustamante, J.; Barbero, J. M. V.; Hunt, A. J.; Clark, J. H. Can Bio-Based Chemicals Meet Demand? Global and Regional Case-Study around Citrus Waste-Derived Limonene as a Solvent for Cleaning Applications. *Biofuels, Bioproducts and Biorefining* **2016**, *10* (6), 686–698. https://doi.org/https://doi.org/10.1002/bbb.1677.

(17) Liu, X.; Gmitter, F. G.; Grosser, J. W.; Wang, Y. Effects of Rootstocks on the Flavor Quality of Huanglongbing-Affected Sweet Orange Juices Using Targeted Flavoromics Strategy. *RSC Adv* **2023**, *13* (8), 5590–5599. https://doi.org/10.1039/d2ra08182b.

(18) Nichkova, M.; Fu, X.; Yang, Z.; Zhong, P.; Sanborn, J. R.; Chang, D.; Gee, S. J.; Hammock, B. D. Immunochemical Screening of Pesticides (Simazine and Cypermethrin) in Orange Oil. *J Agric Food Chem* **2009**, *57* (13), 5673–5679. https://doi.org/10.1021/jf900652a.

(19) Chandran, S. S.; Kealey, J. T.; Reeves, C. D. Microbial Production of Isoprenoids. *Process Biochemistry* **2011**, *46* (9), 1703–1710. https://doi.org/https://doi.org/10.1016/j.procbio.2011.05.012.

(20) Moser, S.; Pichler, H. Identifying and Engineering the Ideal Microbial Terpenoid Production Host. *Applied Microbiology and Biotechnology*. Springer Verlag July 20, 2019, pp 5501–5516. https://doi.org/10.1007/s00253-019-09892-y.

(21) Schilmiller, A. L.; Schauvinhold, I.; Larson, M.; Xu, R.; Charbonneau, A. L.; Schmidt, A.; Wilkerson, C.; Last, R. L.; Pichersky, E. *Monoterpenes in the Glandular Trichomes of Tomato Are Synthesized from a Neryl Diphosphate Precursor Rather than Geranyl Diphosphate*. www.pnas.org/cgi/content/full/.

(22) Ignea, C.; Raadam, M. H.; Motawia, M. S.; Makris, A. M.; Vickers, C. E.; Kampranis, S. C. Orthogonal Monoterpenoid Biosynthesis in Yeast Constructed on an Isomeric Substrate. *Nat Commun* **2019**, *10* (1). https://doi.org/10.1038/s41467-019-11290-x.

(23) Wu, J.; Cheng, S.; Cao, J.; Qiao, J.; Zhao, G.-R. Systematic Optimization of Limonene Production in Engineered Escherichia Coli. *J Agric Food Chem* **2019**, *67* (25), 7087–7097. https://doi.org/10.1021/acs.jafc.9b01427.

(24) Lücker, J.; El Tamer, M. K.; Schwab, W.; Verstappen, F. W. A.; van der Plas, L. H. W.; Bouwmeester, H. J.; Verhoeven, H. A. Monoterpene Biosynthesis in Lemon (Citrus Limon). *Eur J Biochem* **2002**, *269* (13), 3160–3171. https://doi.org/https://doi.org/10.1046/j.1432-1033.2002.02985.x.

(25) Morehouse, B. R.; Kumar, R. P.; Matos, J. O.; Olsen, S. N.; Entova, S.; Oprian, D. D. Functional and Structural Characterization of a (+)-Limonene Synthase from Citrus Sinensis. *Biochemistry* **2017**, *56* (12), 1706–1715. https://doi.org/10.1021/acs.biochem.7b00143.

(26) Kumar, R. P.; Morehouse, B. R.; Matos, J. O.; Malik, K.; Lin, H.; Krauss, I. J.; Oprian, D. D. Structural Characterization of Early Michaelis Complexes in the Reaction Catalyzed by (+)-Limonene Synthase from Citrus Sinensis Using Fluorinated Substrate Analogues. *Biochemistry* **2017**, *56* (12), 1716–1725. https://doi.org/10.1021/acs.biochem.7b00144.

(27) Leferink, N. G. H.; Ranaghan, K. E.; Karuppiah, V.; Currin, A.; van der Kamp, M. W.; Mulholland, A. J.; Scrutton, N. S. Experiment and Simulation Reveal How Mutations in Functional Plasticity Regions Guide Plant Monoterpene Synthase Product Outcome. *ACS Catal* **2018**, *8* (5), 3780–3791. https://doi.org/10.1021/acscatal.8b00692.

(28) Ignea, C.; Pontini, M.; Motawia, M. S.; Maffei, M. E.; Makris, A. M.; Kampranis, S. C. Synthesis of 11-Carbon Terpenoids in Yeast Using Protein and Metabolic Engineering. *Nat Chem Biol* **2018**, *14* (12), 1090–1098. https://doi.org/10.1038/s41589-018-0166-5.

(29) Wu, J.; Cheng, S.; Cao, J.; Qiao, J.; Zhao, G.-R. Systematic Optimization of Limonene Production in Engineered Escherichia Coli. *J Agric Food Chem* **2019**, *67* (25), 7087–7097. https://doi.org/10.1021/acs.jafc.9b01427.

(30) Rolf, J.; Julsing, M. K.; Rosenthal, K.; Lütz, S. A Gram-Scale Limonene Production Process with Engineered Escherichia Coli. *Molecules* **2020**, *25* (8). https://doi.org/10.3390/molecules25081881.

(31) Bohlmann, J.; Phillips, M.; Ramachandiran, V.; Katoh, S.; Croteau, R. CDNA Cloning, Characterization, and Functional Expression of Four New Monoterpene Synthase Members of the Tpsd Gene Family from Grand Fir (Abies Grandis). *Arch Biochem Biophys* **1999**, *368* (2), 232–243. https://doi.org/https://doi.org/10.1006/abbi.1999.1332.

(32) Trapp, S. C.; Croteau, R. B. *Genomic Organization of Plant Terpene Synthases and Molecular Evolutionary Implications*; 2001.

(33) Jo¨rg Bohlmann, J.; Meyer-Gauen, G.; Croteau, R. *This Contribution Is Part of the Special Series of Inaugural Articles by Members of the Plant Terpenoid Synthases: Molecular Biology and Phylogenetic Analysis (Terpene Cyclaseisoprenoidsplant Defensegenetic Engineeringsecondary Metabolism)*; 1998; Vol. 95. www.pnas.org.

(34) Del Terra, L.; Lonzarich, V.; Asquini, E.; Navarini, L.; Graziosi, G.; Suggi Liverani, F.; Pallavicini, A. Functional Characterization of Three Coffea Arabica L. Monoterpene Synthases: Insights into the Enzymatic Machinery of Coffee Aroma. *Phytochemistry* **2013**, *89*, 6–14. https://doi.org/https://doi.org/10.1016/j.phytochem.2013.01.005.

(35) Byun-McKay, A.; Godard, K. A.; Toudefallah, M.; Martin, D. M.; Alfaro, R.; King, J.; Bohlmann, J.; Plant, A. L. Wound-Induced Terpene Synthase Gene Expression in Sitka Spruce That Exhibit Resistance or Susceptibility to Attack by the White Pine Weevil. *Plant Physiol* **2006**, *140* (3), 1009–1021. https://doi.org/10.1104/pp.105.071803.

(36) Landmann, C.; Fink, B.; Festner, M.; Dregus, M.; Engel, K.-H.; Schwab, W. Cloning and Functional Characterization of Three Terpene Synthases from Lavender (Lavandula Angustifolia). *Arch Biochem Biophys* **2007**, *465* (2), 417–429. https://doi.org/https://doi.org/10.1016/j.abb.2007.06.011.

(37) Maruyama, T.; Saeki, D.; Ito, M.; Honda, G. *Molecular Cloning, Functional Expression and Characterization of d-Limonene Synthase from Agastache Rugosa*; 2002.

(38) Dang, J.; Lin, G.; Liu, L.; Zhou, P.; Shao, Y.; Dai, S.; Sang, M.; Jiang, Z.; Liu, C.; Wu, Q. Comparison of Pulegone and Estragole Chemotypes Provides New Insight Into Volatile Oil Biosynthesis of Agastache Rugosa. *Front Plant Sci* **2022**, *13*. https://doi.org/10.3389/fpls.2022.850130.

(39) Alonso-Gutierrez, J.; Chan, R.; Batth, T. S.; Adams, P. D.; Keasling, J. D.; Petzold, C. J.; Lee, T. S. Metabolic Engineering of Escherichia Coli for Limonene and Perillyl Alcohol Production. *Metab Eng* **2013**, *19*, 33–41. https://doi.org/10.1016/j.ymben.2013.05.004.

(40) Unger, T.; Jacobovitch, Y.; Dantes, A.; Bernheim, R.; Peleg, Y. Applications of the Restriction Free (RF) Cloning Procedure for Molecular Manipulations and Protein Expression. *J Struct Biol* **2010**, *172* (1). https://doi.org/10.1016/j.jsb.2010.06.016.

(41) Chen, X.; T, R.; Esque, J.; Zhang, C.; Shukal, S.; Lim, C. C.; Ong, L.; Smith, D.; André, I. Total Enzymatic Synthesis of Cis-α-Irone from a Simple Carbon Source. *Nat Commun* **2022**, *13* (1), 7421. https://doi.org/10.1038/s41467-022-35232-2.

(42) Shukal, S.; Lim, X. H.; Zhang, C.; Chen, X. Metabolic Engineering of Escherichia Coli BL21 Strain Using Simplified CRISPR-Cas9 and Asymmetric Homology Arms Recombineering. *Microb Cell Fact* **2022**, *21* (1). https://doi.org/10.1186/s12934-022-01746-z.

(43) Williams, T.; Kelley, C. Gnuplot 4.4: An Interactive Plotting Program. ... *Documentation, Http://Sourceforge. Net/Projects/Gnuplot* **2010**, No. C.

(44) Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data. *Bioinformatics* **2012**, *28* (23). https://doi.org/10.1093/bioinformatics/bts565.

(45) Li, W.; Godzik, A. Cd-Hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences. *Bioinformatics* **2006**, *22* (13). https://doi.org/10.1093/bioinformatics/btl158.

(46) Katoh, K.; Rozewicki, J.; Yamada, K. D. MAFFT Online Service: Multiple Sequence Alignment, Interactive Sequence Choice and Visualization. *Brief Bioinform* **2018**, *20* (4). https://doi.org/10.1093/bib/bbx108.

(47) Robert, X.; Gouet, P. Deciphering Key Features in Protein Structures with the New ENDscript Server. *Nucleic Acids Res* **2014**, *42* (W1), W320–W324. https://doi.org/10.1093/nar/gku316.

(48) Hyatt, D. C.; Youn, B.; Zhao, Y.; Santhamma, B.; Coates, R. M.; Croteau, R. B.; Kang, C. *Structure of Limonene Synthase, a Simple Model for Terpenoid Cyclase Catalysis*; 2007. www.pnas.orgcgidoi10.1073pnas.0700915104.

(49) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *Journal of Computational Chemistry*. 2005. https://doi.org/10.1002/jcc.20291.

(50) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat Methods* **2016**, *14* (1). https://doi.org/10.1038/nmeth.4067.

(51) Valdés-Tresanco, M. S.; Valdés-Tresanco, M. E.; Valiente, P. A.; Moreno, E. Gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. *J Chem Theory Comput* **2021**, *17* (10). https://doi.org/10.1021/acs.jctc.1c00645.

(52) Gelain, L.; Yeoh, J. W.; Hossain, G. S.; Alfenore, S.; Guillouet, S.; Ling, H.; Poh, C. L.; Gorret, N.; Foo, J. L. Identification and Monitoring of Cell Heterogeneity from Plasmid Recombination during Limonene Production. *Appl Microbiol Biotechnol* **2025**, *109* (1). https://doi.org/10.1007/s00253-024-13273-5.

(53) Alonso-Gutierrez, J.; Kim, E. M.; Batth, T. S.; Cho, N.; Hu, Q.; Chan, L. J. G.; Petzold, C. J.; Hillson, N. J.; Adams, P. D.; Keasling, J. D.; Garcia Martin, H.; Lee, T. S. Principal Component Analysis of Proteomics (PCAP) as a Tool to Direct Metabolic Engineering. *Metab Eng* **2015**, *28*, 123–133. https://doi.org/10.1016/j.ymben.2014.11.011.

(54) Srividya, N.; Davis, E. M.; Croteau, R. B.; Lange, B. M. Functional Analysis of (4S)-Limonene Synthase Mutants Reveals Determinants of Catalytic Outcome in a Model Monoterpene Synthase. *Proc Natl Acad Sci U S A* **2015**, *112* (11), 3332–3337. https://doi.org/10.1073/pnas.1501203112.

(55) Xu, J.; Ai, Y.; Wang, J.; Xu, J.; Zhang, Y.; Yang, D. Converting S-Limonene Synthase to Pinene or Phellandrene Synthases Reveals the Plasticity of the Active Site. *Phytochemistry* **2017**, *137*, 34–41. https://doi.org/10.1016/j.phytochem.2017.02.017.

(56) Xu, J.; Xu, J.; Ai, Y.; Farid, R. A.; Tong, L.; Yang, D. Mutational Analysis and Dynamic Simulation of S-Limonene Synthase Reveal the Importance of Y573: Insight into the Cyclization Mechanism in Monoterpene Synthases. *Arch Biochem Biophys* **2018**, *638*, 27–34. https://doi.org/https://doi.org/10.1016/j.abb.2017.12.007.

(57) Cheng, S.; Liu, X.; Jiang, G.; Wu, J.; Zhang, J.; Lei, D.; Yuan, Y.-J.; Qiao, J.; Zhao, G.-R. Orthogonal Engineering of Biosynthetic Pathway for Efficient Production of Limonene in Saccharomyces Cerevisiae. *ACS Synth Biol* **2019**, *8* (5), 968–975. https://doi.org/10.1021/acssynbio.9b00135.

(58) Mistry, J.; Chuguransky, S.; Williams, L.; Qureshi, M.; Salazar, G. A.; Sonnhammer, E. L. L.; Tosatto, S. C. E.; Paladin, L.; Raj, S.; Richardson, L. J.; Finn, R. D.; Bateman, A. Pfam: The Protein Families Database in 2021. *Nucleic Acids Res* **2021**, *49* (D1), D412–D419. https://doi.org/10.1093/nar/gkaa913.

(59) Bohlmann, J.; Meyer-Gauen, G.; Croteau, R. Plant Terpenoid Synthases: Molecular Biology and Phylogenetic Analysis (Terpene Cyclase⁻isoprenoids⁻plant Defense⁻genetic Engineering⁻secondary Metabolism). *Biochemistry* **1998**, *95*.

(60) Booth, J. K.; Page, J. E.; Bohlmann, J. Terpene Synthases from Cannabis Sativa. *PLoS One* **2017**, *12* (3). https://doi.org/10.1371/journal.pone.0173911.

(61) Günnewich, N.; Page, J. E.; Köllner, T. G.; Degenhardt, J.; Kutchan, T. M.; Danforth, D. *Functional Expression and Characterization of Trichome-Specific (-)-Limonene Synthase and (+)-α-Pinene Synthase from Cannabis Sativa*; 2007; Vol. 2. www.ebi.ac.uk/clustalw/.

(62) Maruyama, T.; Ito, M.; Kiuchi, F.; Honda, G. *Molecular Cloning, Functional Expression and Characterization of d-Limonene Synthase from Schizonepeta Tenuifolia*; 2001; Vol. 24.

(63) Back, M.; Chappell, K.; Noel, J. *Truncation of Limonene Synthase Preprotein Provides a Fully Active "Pseudomature" Form of This Monoterpene Cyclase and Reveals the Function of the Amino-Terminal Arginine Pair †*; 1997; Vol. 277.

(64) Riehl, P. S.; Richardson, A. D.; Sakamoto, T.; Reid, J. P.; Schindler, C. S. Origin of Enantioselectivity Reversal in Lewis Acid-Catalysed Michael Additions Relying on the Same Chiral Source. *Chem Sci* **2021**, *12* (42). https://doi.org/10.1039/d1sc03741b.

(65) Lei, D.; Qiu, Z.; Qiao, J.; Zhao, G. R. Plasticity Engineering of Plant Monoterpene Synthases and Application for Microbial Production of Monoterpenoids. *Biotechnology for Biofuels*. BioMed Central Ltd December 1, 2021. https://doi.org/10.1186/s13068-021-01998-8.

(66) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl,

S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

(67) Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (1979)* **2012**, *337* (6096), 816–821. https://doi.org/10.1126/science.1225829.

(68) Reiling, K. K.; Yoshikuni, Y.; Martin, V. J. J.; Newman, J.; Bohlmann, J.; Keasling, J. D. Mono and Diterpene Production in Escherichia Coli. *Biotechnol Bioeng* **2004**, *87* (2), 200–212. https://doi.org/https://doi.org/10.1002/bit.20128.

FIGURE LEGENDS

Figure 1. Microbial production of limonene. (a) Biosynthetic pathway of limonene. Terpene precursors isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) are synthesized through the mevalonate (MVA) pathway from acetyl-CoA. These are converted to geranyl pyrophosphate (GPP) by the GPP synthase (GPPS) enzyme. GPP can be further converted by farnesyl pyrophosphate synthase (FPPS) to sesquiterpenes and terpenes with C ≥ 20. Introduction of limonene synthase (LS) converts most of the GPP to either (+)- or (-)-limonene. Alternatively, terpene precursors can be assembled to neryl pyrophosphate (NPP), by the NPP synthase (NPPS), an orthogonal pathway. NPP is converted by LS to limonene, but *E. coli*'s FPP does not use NPP as a substrate for the biosynthesis of its terpenes. (b) Plasmidic systems used for the expression of limonene from acetyl-CoA: pEcoCTs-CPMS6a + pEcoCTs-CPMS5 for limonene production from GPP, pEcoCTs-CPMS6b + pEcoCTs-CPMS5 for limonene production from NPP.

Figure 2. Limonene production by different homologous LSs in *E. coli*. (a) Chiral chromatography analysis of limonene produced with different full-length LS (indicated on top of the limonene peak) via the GPP route. (b) Comparing full-length and truncated LS for limonene production after 72 hours growth. Production levels are reported relatively to final culture OD to take into account growth fluctuations among strains. (c) Comparing GPP and NPP pathway using truncated LS.

Figure 3. tArLS and tNtLS alignment. (a) Sequence alignment with secondary structure annotation, using tArLS AlphaFold2 model[66] as a structural template. Arrows highlight the residues selected for mutagenesis. (b) Superimposition of the tArLS (cyan) and tNtLS (purple)

active sites. (c) AF2 model of tArLS with the location of the mutations highlighted in red. GPP (light gray) and magnesium ions (dark gray spheres) were placed after superimposition of X-ray structure of *Citrus sinensis* (PDB ID: 5UV1[26]).

Figure 4. Limonene production analyses of tArLS and tNtLS mutants. (a) and (b) tArLS mutants using the GPP and NPP pathway, respectively. (c) tNtLS loop1 mutant. (d) limonene titer ratio of NPP over GPP for the tArLS and tNtLS mutants (e) tArLS in wild-type MG1655 or MG1655 *ispA* S80F

Figure 5. Key interactions for GPP binding. The active sites of the wild-type tArLS and the mutant (tArLS-S8K-I265V-loop1) are shown in (a) and (b), respectively. On the left, one snapshot with an extended conformation of GPP is represented, whereas a cyclic conformation is displayed on the right. The Mg^{2+}ions are shown in green spheres and GPP in magenta sticks with labelling of O1 (pyrophosphate head) and C3 (polyisoprenoid tail). Key residues, discussed in the literature[56], are labelled for each snapshot. Blue dashed lines on left (a) represent potential hydrogen bonds. On the right, tables provide MMPBSA results of Mg^{2+}ions and key residues for both enzymes.

TABLES

Table 1. Kinetics of tArLS, tNtLS and tArLS mutants using NPP or GPP as substrates.

	tArLS		**tArLS-S8K**		**tArLS-loop1**		**tArLS-S8K-I265V-loop1**		**tNtLS**	
	GPP	**NPP**	**GPP**	**NPP**	**GPP**	**NPP**	**GPP**	**NPP**	**GPP**	**NPP**
K_M (μM)	33.3 ± 11.4	44.3 ± 5.4	18.1 ± 2.4	20.1 ± 5.2	10.8 ± 2.6	14.8 ± 2.0	8.0 ± 0.9	8.5 ± 0.8	12.7 ± 1.2	16.0 ± 1.5
k_{cat} (sec^{-1})	0.053 ± 0.008	0.169 ± 0.010	0.022 ± 0.001	0.043 ± 0.004	0.004 ± 0.000	0.007 ± 0.000	0.011 ± 0.000	0.015 ± 0.000	0.037 ± 0.001	0.079 ± 0.003
Catalytic efficiency (mM^{-1} sec^{-1})	1.60 ± 0.6	3.81 ± 0.51	1.23 ± 0.18	2.38 ± 0.39	0.39 ± 0.10	0.47 ± 0.07	1.32 ± 0.15	1.78 ± 0.18	2.95 ± 0.30	4.98 ± 0.1
Catalytic efficiency NPP/GPP ratio	2.38		1.93		1.19		1.35		1.69	

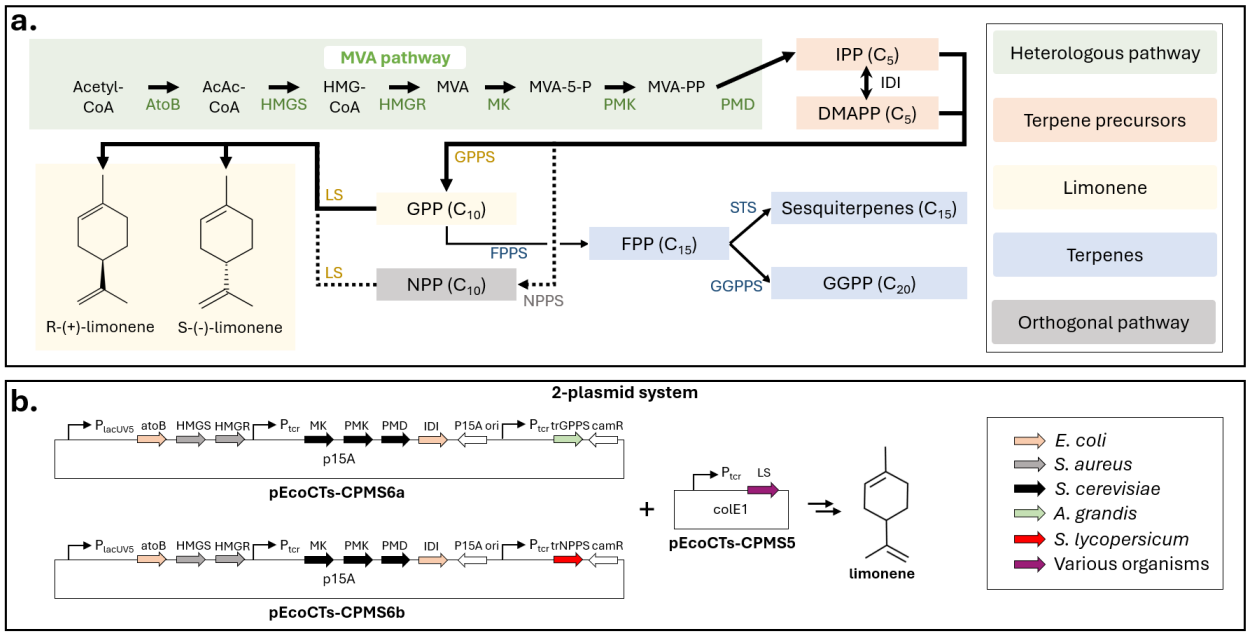

a.
MVA pathway
Acetyl-CoA
AtoB
AcAc-CoA
HMGS
HMG-CoA
HMGR
MVA
MK
MVA-5-P
PMK
MVA-PP
PMD
IPP (C5)
IDI
DMAPP (C5)
GPPS
LS
GPP (C10)
FPPS
FPP (C15)
STS
Sesquiterpenes (C15)
LS
NPP (C10)
NPPS
GGPPS
GGPP (C20)
R-(+)-limonene
S-(-)-limonene
Heterologous pathway
Terpene precursors
Limonene
Terpenes
Orthogonal pathway
b.
2-plasmid system
PlacUV5 atoB HMGS HMGR Ptcr MK PMK PMD IDI P15A ori Ptcr trGPPS camR
p15A
pEcoCTs-CPMS6a
PlacUV5 atoB HMGS HMGR Ptcr MK PMK PMD IDI P15A ori Ptcr trNPPS camR
p15A
pEcoCTs-CPMS6b
+
Ptcr LS
colE1
pEcoCTs-CPMS5
limonene
E. coli
S. aureus
S. cerevisiae
A. grandis
S. lycopersicum
Various organisms

Figure 1

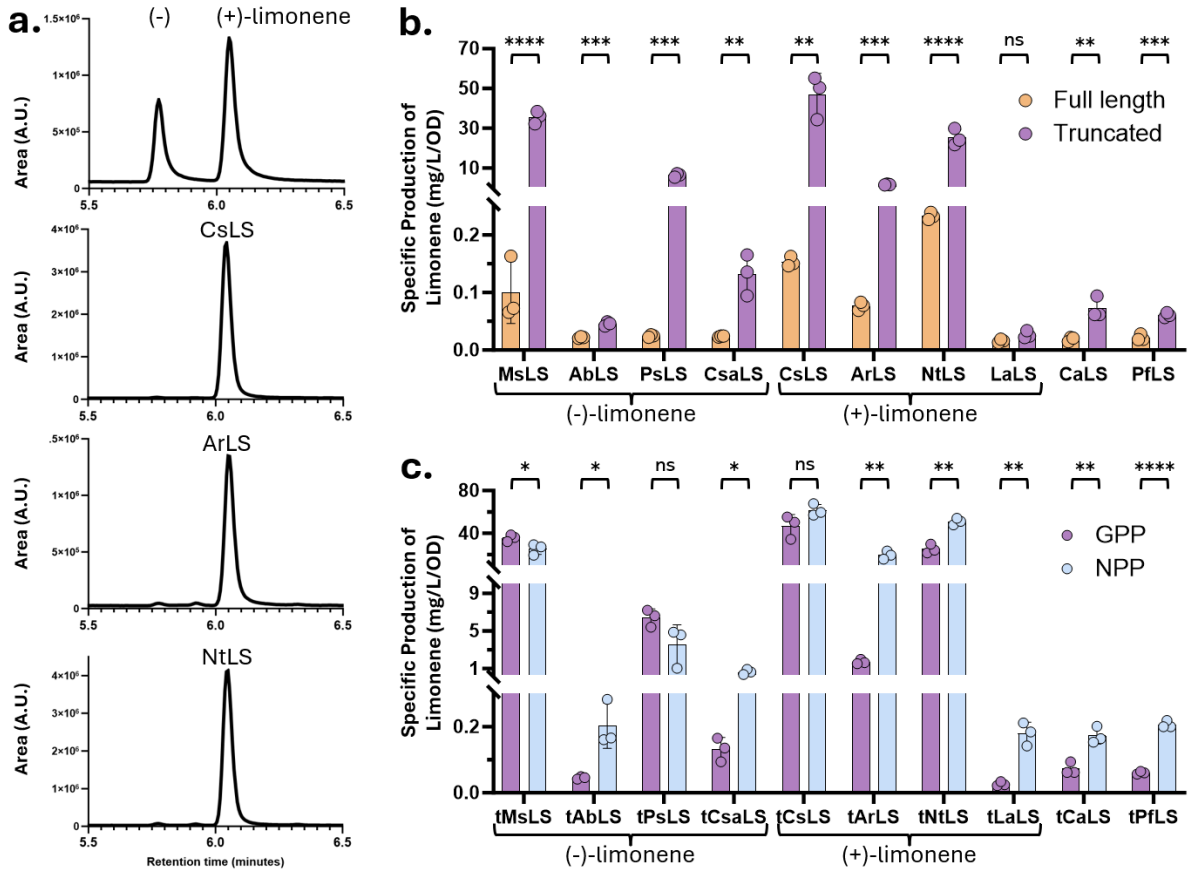

a.
(-)
(+)-limonene
Area (A.U.)
CsLS
ArLS
NtLS
Retention time (minutes)
b.
Specific Production of Limonene (mg/L/OD)
Full length
Truncated
MsLS
AbLS
PsLS
CsaLS
CsLS
ArLS
NtLS
LaLS
CaLS
PfLS
(-)-limonene
(+)-limonene
c.
GPP
NPP
tMsLS
tAbLS
tPsLS
tCsaLS
tCsLS
tArLS
tNtLS
tLaLS
tCaLS
tPfLS

Figure 2

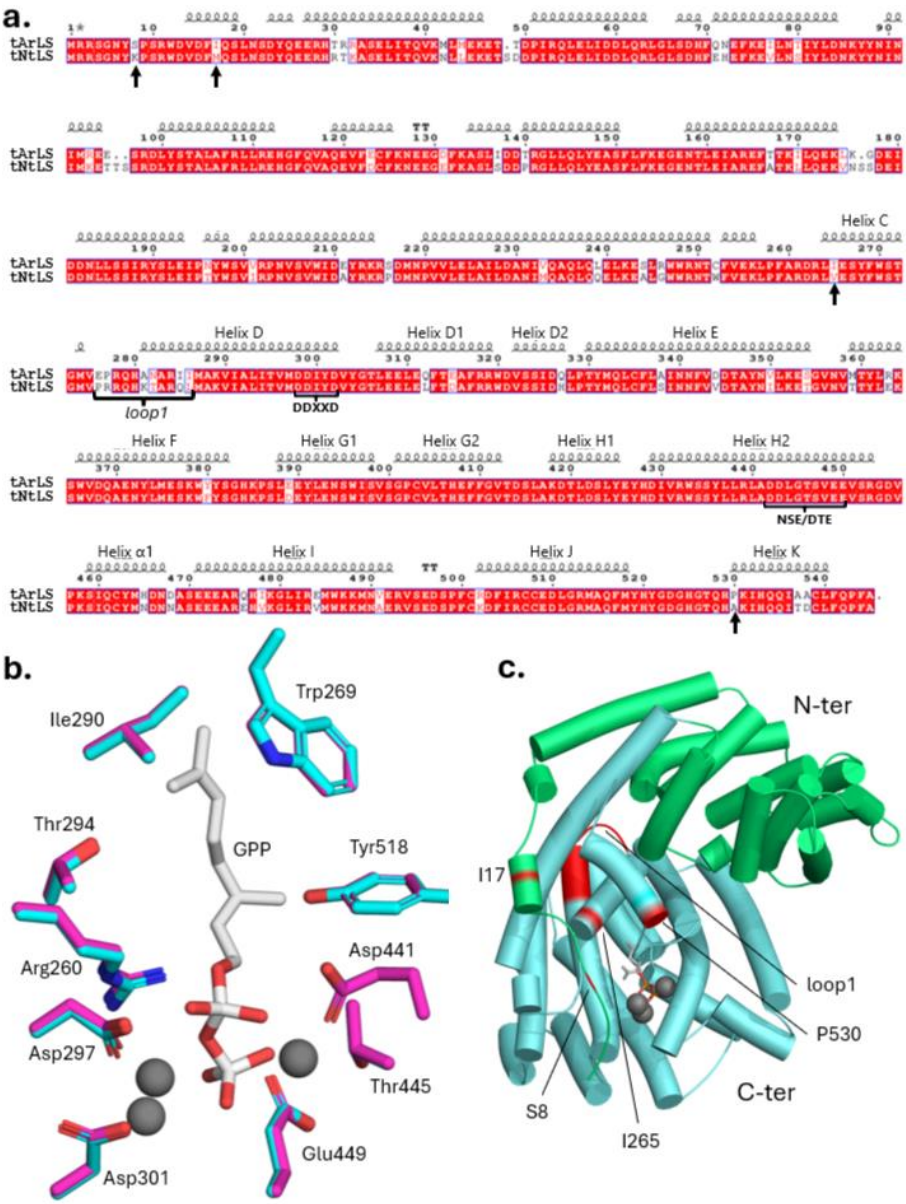

a.
tArLS
tNtLS
Helix C
Helix D
Helix D1
Helix D2
Helix E
loop1
DDXXD
Helix F
Helix G1
Helix G2
Helix H1
Helix H2
NSE/DTE
Helix α1
Helix I
Helix J
Helix K
b.
Trp269
Ile290
Thr294
GPP
Tyr518
Arg260
Asp441
Asp297
Thr445
Glu449
Asp301
c.
N-ter
I17
loop1
P530
C-ter
S8
I265

Figure 3

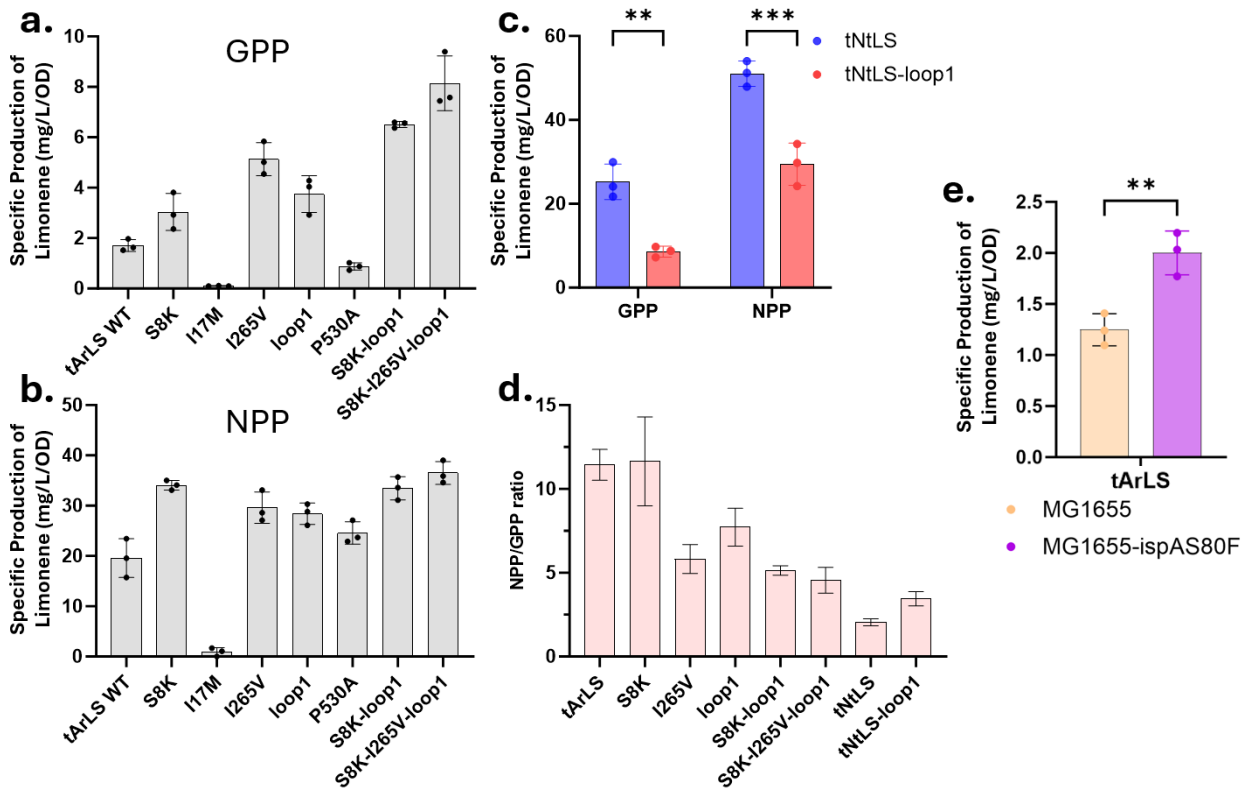
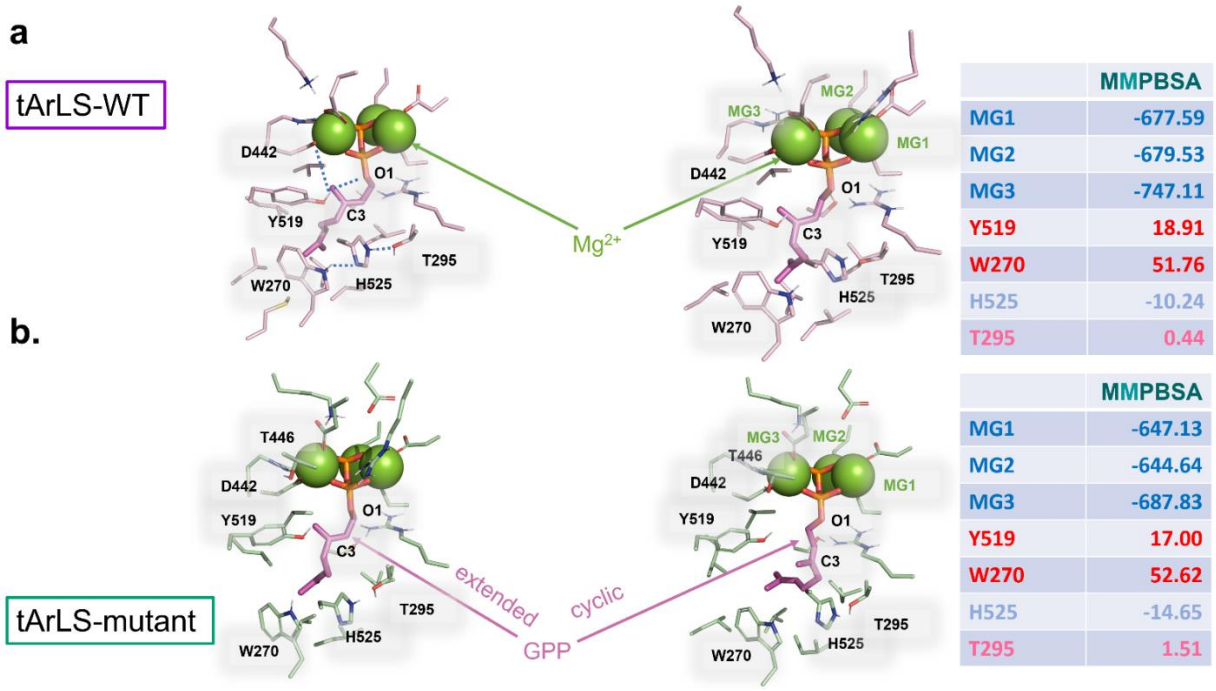

a.
GPP
Specific Production of Limonene (mg/L/OD)
tArLS WT
S8K
I17M
I265V
loop1
P530A
S8K-loop1
S8K-I265V-loop1
b.
NPP
c.
**
***
tNtLS
tNtLS-loop1
GPP
NPP
d.
NPP/GPP ratio
tArLS
tNtLS
tNtLS-loop1
e.
tArLS
MG1655
MG1655-ispAS80F

Figure 4



	MMPBSA
MG1	-677.59
MG2	-679.53
MG3	-747.11
Y519	18.91
W270	51.76
H525	-10.24
T295	0.44

	MMPBSA
MG1	-647.13
MG2	-644.64
MG3	-687.83
Y519	17.00
W270	52.62
H525	-14.65
T295	1.51

Figure 5

For Table of Contents Only