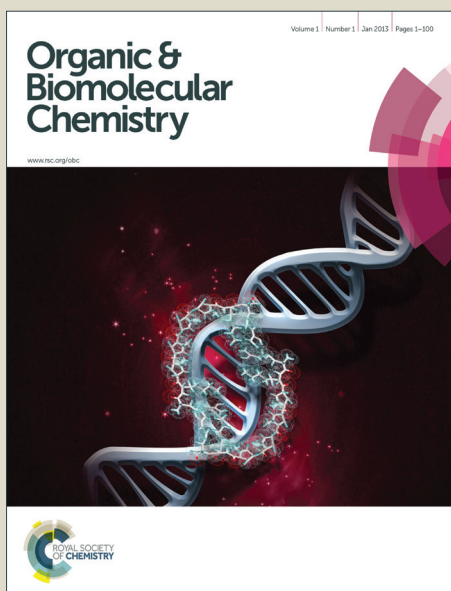


Organic & Biomolecular Chemistry

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

Improved selectivity of an engineered multi-product terpene synthase

Ryan Lauchli,*^{†a,b} Julia Pitzer,^{†a} Rebekah Z. Kitto,^a Karolina Z. Kalbarczyk,^a and Kersten S. Rabe*^{a,c}

^a California Institute of Technology, Division of Chemistry & Chemical Engineering, 1200 E. California Blvd., Pasadena CA 91125.

^b Current address: Sustainable Chemistry Solutions, Inc., 1052 W. Alameda Avenue, #110, Burbank, California USA 91506. Email: ryan@bio-catalyst.com

^c Current address: KIT-Campus North, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany. Email: kersten.rabe@kit.edu

[†] RL and JP contributed equally to this work.

*Authors to whom correspondence can be addressed

Electronic Supplementary Information (ESI) available: Product distribution plot, kinetic plots, nucleic acid and amino acid sequences. See DOI: 10.1039/b000000x/

Abstract

Mutation of the sesquiterpene synthase Cop2 was conducted with a high-throughput screen for cyclization activity using a non-natural substrate. A mutant of Cop2 was identified that contained three amino acid substitutions. This mutant, 17H2, converted the natural substrate FPP into germacrene D-4-ol with 77% selectivity. This selectivity contrasts with that of the parent enzyme in which germacrene D-4-ol is produced as 29% and α -cadinol as 46% of the product mixture. The mutations were shown to each contribute to this selectivity, and a homology model suggested that the mutations lie near to the active site though would be unlikely to be targeted for mutation by rational methods. Kinetic comparisons show that 17H2 maintains a $k_{\text{cat}}/K_{\text{M}}$ of $0.62 \text{ mM}^{-1}/\text{s}^{-1}$, which is nearly identical to that of the parent Cop2, which had a $k_{\text{cat}}/K_{\text{M}}$ of $0.58 \text{ mM}^{-1}/\text{s}^{-1}$.

1.

Introduction

Terpene synthase enzymes exhibit exceptional control over the formation of a wide variety of products via highly complex cationic cyclizations.¹⁻³ For example, from a single 15-carbon substrate, farnesyl diphosphate (FPP), terpene synthases can generate over 300 distinct products. In many cases, a terpene synthase can guide the reaction of the substrate to just one of these hundreds of possible products with high product selectivity and form multiple chiral centers in the process. Thus, terpene synthases control some of the most complex reactions in nature. Understanding and controlling these reactions to predict their products, to alter their product distributions, and to increase activity has proven to be difficult.

Several groups have recently taken on the challenge of altering and controlling the course of terpene synthase reactions using mutagenesis.⁴⁻¹² These studies have largely focused on mutations either surrounding the active site or at locations identified in close homologues. This limitation exists because there has been a lack of an efficient system with which to screen the large numbers of mutants generated via random mutagenesis, many of which may be misfolded. Furthermore, even when focusing on mutations in the active site or those identified using homology or structural information, the resulting mutants frequently suffer from decreased activity. For example, mutation to the H- α 1 loop of sesquiterpene synthases Cop4 and Cop6 severely impeded the kinetic behavior of these enzymes.⁹ In another example, thermostabilization of the sesquiterpene tobacco epi-aristolochene synthase caused lowered activity in the thermostabilized mutant.⁸ It is a

general problem in terpene synthase engineering that mutations in these enzymes often lead to decreased activity.

We are aware of only one previous study using a high-throughput assay to screen terpene synthase mutants that were engineered for altered product formation.⁴ The assay identified terpene synthase mutants with an increased likelihood of correct folding, yet it provided no information into the cyclization activity of these mutants. Therefore, significant effort was expended in a secondary screen to determine which mutants maintained activity.

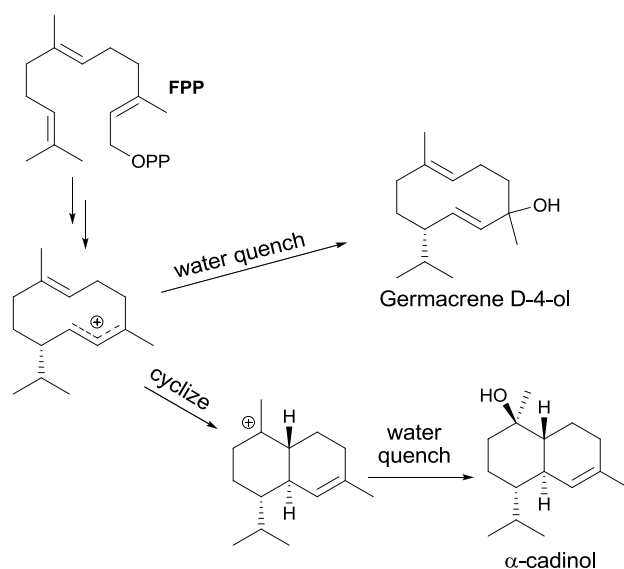
We recently disclosed a high-throughput assay for cyclization activity in sesquiterpene synthases and subjected a highly selective fungal terpene synthase to two rounds of directed evolution.¹³ In the course of these two rounds, in which the enzyme was evolved for thermostability, the high selectivity of the enzyme for its native product was maintained and there was no apparent loss of activity. In the current study, we applied random mutagenesis to a multi-product terpene synthase using this high-throughput assay to identify a mutant that maintained wild-type activity while concurrently changing and improving the product selectivity of the enzyme.

Results

Cop2 is a multi-product sesquiterpene synthase from *Coprinus cinereus*, and no crystal structure is known for this enzyme.¹⁴ The product composition formed from Cop2 includes products believed to result from a number of distinct cyclization chemistries.

Error-prone PCR (epPCR) was used to create random point mutations in the Cop2 gene. The Cop2 mutants thus generated were expressed as a library in *E. coli*. The library consisted of

about 3000 clones with an average error rate of 1.4 base pair (bp) mutations per gene. This library was screened for cyclization activity using the high-throughput screen we previously developed. The screen relies upon the usage of a surrogate substrate (S1), which contains a vinyl methyl ether moiety at the head of the molecule in place of the dimethyl-alkene moiety found in the natural substrate, FPP. Terpene synthase-mediated cyclization of S1 is measured by methanol release using an enzyme-coupled colorimetric readout. An active mutant denoted 17H2 was identified in this library with comparable activity to the parent as measured by this screen.



Scheme 1 Proposed formation of the two major products from FPP catalyzed by Cop2

The mutant of Cop2 denoted 17H2 was shown to be a triple mutant containing mutations L59H, T65A and S310Y. The reactions of both Cop2 and 17H2 with the natural substrate FPP were investigated by GC-MS measurements. Reactions of purified enzyme were run in buffer in vials with an overlayer of hexanes. Though previous headspace analysis of Cop2 reactions in cell culture led to the proposal that Cop2 is a germacrene A synthase,¹⁴ analysis of the terpenes

extracted into overlayers showed that this enzyme generates a mixture of products with α -cadinol and germacrene D-4-ol present in the largest amounts. Because α -cadinol and germacrene D-4-ol are alcohols rather than hydrocarbons (like germacrene A), they were not proportionally represented by headspace analysis because they have both lower vapor pressure and increased solubility in the water layer. Scheme 1 shows a portion of the proposed reaction mechanism for the formation of these two products. No single product comprises more than 50% of the product mixture, and Cop2 is therefore a multi-product sesquiterpene synthase. In contrast, the triple mutant 17H2 generates a much more selective product mixture as shown in Figure 1. This mixture is comprised of 77% of germacrene D-4-ol, and no α -cadinol is produced. Thus, product selectivity was significantly improved.

We measured the activity of 17H2 on FPP using GC measurements. Kinetic studies on both 17H2 and Cop2 were estimated using FPP as the substrate, and K_m values were determined assuming Michaelis-Menten kinetics to be 25.5 μM for Cop2 and 22.6 μM for 17H2. The k_{cat} values of 0.015 s^{-1} and 0.014 s^{-1} for Cop2 and 17H2, respectively, were also similar. Consequently, the k_{cat}/K_m values of 0.58 $\text{mM}^{-1}/\text{s}^{-1}$ and 0.62 $\text{mM}^{-1}/\text{s}^{-1}$ were nearly identical and show that the mutant retained the activity of the parent (see Table 1). Therefore, screening with a surrogate substrate for active enzymes gave a useful approximation of activity on the native substrate, FPP.

Table 1 Summary of the kinetic parameters of Cop2 and 17H2 with FPP.

Enzyme	K_m [μM]	+/-	v_{max} [$\mu\text{M}/\text{h}$]	+/-	Enzyme concentration [μM]	k_{cat} [s^{-1}]	k_{cat}/K_m [$\text{mM}^{-1}/\text{s}^{-1}$]
Cop2	25.5	6.7	10.7	1.0	0.2	0.015	0.58
17H2	22.6	7.1	25.3	2.0	0.5	0.014	0.62

Because 17H2 contained three point mutations, we sought to elucidate if one of the single point mutations caused the high selectivity or if the combination of the three mutations was responsible for this effect. The three single mutants L59H, T65A and S310Y were therefore constructed, and the reactions of FPP catalyzed by these mutants were analyzed.

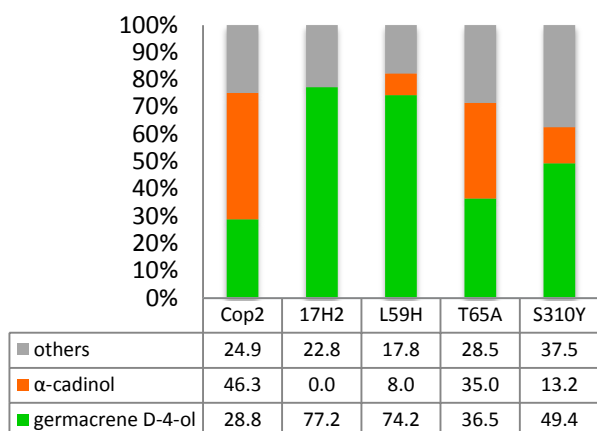


Figure 1 Product distribution in % of Cop2, the triple mutant 17H2 and the single mutants L59H, T65A and S310Y analyzed by GC.

Analysis of the three single mutants showed that, in comparison to Cop2, L59H and S310Y produce elevated germacrene D-4-ol amounts and reduced α -cadinol amounts from FPP. The mutant T65A also showed a small improvement in germacrene D-4-ol selectivity. None of the single mutants completely abolished α -cadinol production. These results are shown in Figure 1 and suggest that a synergistic effect of the three mutations is responsible for the improvement in the product selectivity of 17H2 towards germacrene D-4-ol.

Though a triple mutant was identified from the Cop2 mutant library, any of the three single mutants would be found using our screening system, and the mutants all give comparable activities when measured with the assay (data not shown).

An alignment of Cop2 with aristolochene synthase and pentalenene synthase was constructed (Figure 2).¹⁵⁻¹⁷ The structures and corresponding active sites of aristolochene synthase and pentalenene synthase are known,^{16, 17} and comparison to Cop2 reveals that residues which are mutated in 17H2 (as highlighted in green) are directly adjacent to active site residues for T65 and S301, and are five residues removed from an active site residue in the case of L59. Despite L59 being the mutated residue furthest from the predicted active site cavity, it causes the largest change in product selectivity.

```

Feature 1
1DGP_A      4 PTQWSYLCHPRVKEVQDEVDGYFlenwkFPSPFKAVRTFLdakFSEVTCLYFPPLaLDDRIHFACRIILTVLFLDDVLEIm- 82
Cop2       26 dCPFPLRVNPLCDEVGRLSEQWFLryanYSPPRAV-AFMaIkAGELTAACYPDadAFHLRVSDDFMNFLENDDWLDIf- 103
1PS1_A     8 HIPLPGRQSPDHARAEAEQLAWPRslglIRSDDAAERHLrggYADLASRFYPHaTGADLDLGVDILMSWFFLEDDLDFDgpr 87

Feature 1
1DGP_A     228 CSAVKVLAEEESKLGIPATKRVLWSMTREWETVHDEIVAEEKiaspd-----gcseaAKAYMK-GLEYQMSGNEQNSKTTR 300
Cop2       248 FNMVSI VMHQKGFALQEAVNFV GELCKKAMERFQADKRNLPswgp-----edgeVAMYVD-GIQNWIVGSLNWSidgt 320
1PS1_A     234 NNMVMI LRREHGWSKRSVSHMQNEVRARLEQYLLLESCLpkvgeiyqldtaereaLERYRTdAVRTVIRGSYDWRHSSG 313

```

Figure 2 Alignment of the amino acid sequences of Cop2 with other sesquiterpene synthases. The residues involved in substrate binding are marked in yellow. The blue boxes highlight the conserved aspartate rich motif and the N(S/T)(E/D) triad, necessary for the binding of Mg²⁺ ions. Cop2, 1DGP_A: aristolochene synthase from *Penicillium roqueforti*, 1PS1_A pentalenene synthase.

The locations of the mutations in 17H2 were further elucidated by a homology model of Cop2 built using SwissModel upon the crystal structure of the bacterial sesquiterpene synthase,

pentalene synthase (Figure 3).¹⁸ The two enzymes shared 21% sequence homology, and pentalene synthase had the highest homology to Cop2 of any terpene synthase crystal structure. The model shows that all three mutations are located near the active site yet do not line the active site cavity. Both L59H and T65A mutations are located within the C-helix (orange in Figure 3), and the S301Y mutation is located at the end of the J-helix (red in Figure 3). The location of these mutations is non-obvious—they would likely not have been targeted for site-directed or site-saturation mutagenesis.

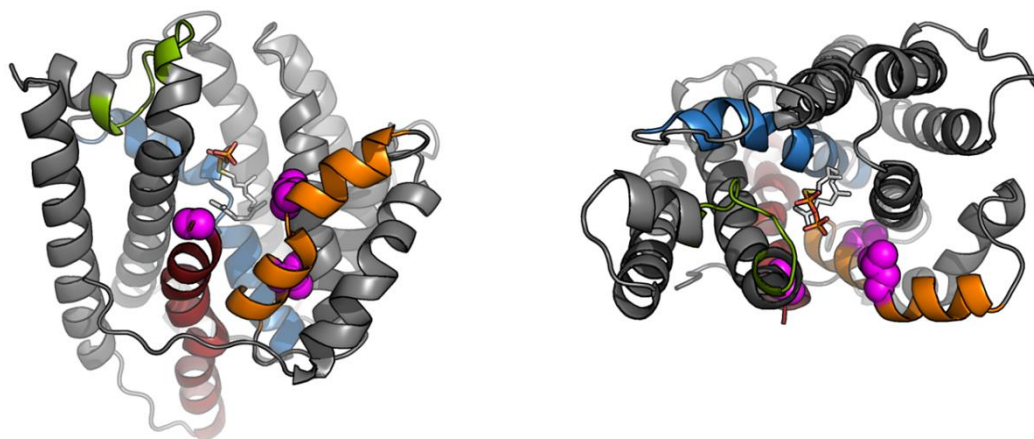


Figure 3 Homology model of Cop2 based on pentalenene synthase structure 1HM7.¹⁵⁻¹⁷ The FFP analog from structure 4KUX is shown in order to indicate the active site. The three residues mutated in 17H2 are shown in pink as follows: S310 is at the end of the J-helix (in red), L59 is in the upper portion of the C-helix (in orange) and T65 is in the lower portion of the C-helix. Also shown are the G-helix in light blue and the H- α 1 loop in green. The image on the right is a view from the top.

Discussion

Enzymatic terpene cyclization is initiated by the formation of a reactive carbocation.¹⁻³ Class I enzymes, which include sesquiterpene synthases like Cop2, generate the carbocation by the release of the pyrophosphate (PP_i) group of the substrate. Enzymes of this class contain characteristic DDxxD and NSE/DTE motifs which facilitate the binding and subsequent cleavage of the PP_i induced by three Mg²⁺ ions as cofactors.¹⁹⁻²¹ Additionally, a basic motif (RY) is also involved in the coordination of the Mg²⁺ and PP_i complex. Mutation of the amino acids involved in Mg²⁺ and PP_i binding generally results in severely attenuated enzyme activity. Mutations in this region in aristolochene synthase, for example, were reported to result in altered product distributions along with a 100-6500 fold decrease in the observed k_{cat}/K_m .⁶ Furthermore, mutations in the Mg²⁺ coordinating motifs of a fungal trichodiene synthase reduced kinetic parameters as well as product specificity.⁷ Consistent with these previous results, the present work identified no mutations in this area of the protein that maintained activity.

Within the terpene synthase active site, amino acids (primarily aliphatic and aromatic) direct the propagation of the carbocation through complex reaction cascades.^{3, 19, 20} Additionally, reaction steps involving acid-base catalysis are thought to occur during sesquiterpene cyclization. The termination of the reaction cascade can occur either via a nucleophilic attack by water, leading to the formation of terpene alcohols or via proton abstraction to form double bonds. The active site of terpene synthases acts not only as a chaperone for the correct folding of the substrate but it also rearranges the carbocation and protects it from premature quenching by solvent molecules. Most previous mutational studies of sesquiterpene synthases have targeted residues within the active site.

Schmidt-Dannert and coworkers, for example, investigated the effects of mutation in the H- α 1 loop (which caps the active site) of several sesquiterpene synthases from *Coprinus cinereus*.⁹ These synthases include the α -cuprenene synthase Cop6, and the multi-product terpene synthases Cop3 and Cop4. Though mutation of residues in the Cop6 H- α 1 loop caused little change in product selectivity, both Cop4 and Cop3 showed significant changes in product compositions upon mutation. The mutants were much less active than the parent enzymes with at least 3-fold slower k_{cat} values for Cop6 and two orders of magnitude slower for Cop4.

Weiss investigated the effects of structure guided mutations in tobacco epi-aristochene synthase (TEAS).⁸ The crystal structure for this enzyme was previously solved, and the structure was used to introduce thermostabilizing mutations that were distant from the active site. Despite this distance, the thermostabilized mutant identified in this work suffered from losses in product selectivity and activity.

Though some studies have indeed identified mutations that maintain or even improve upon wild-type activity, these studies generally rely on structures and enzymes with close homologs. For example, Chen and Zhang solved the crystal structure of a bisabolene synthase, and used it to guide the mutation of the homolog amorphadiene synthase (82% amino acid identity).¹¹ Mutation of T399 to serine increased the k_{cat} of the enzyme two-fold without a decline in product selectivity. In another study, Noel and Chappell utilized the structure of TEAS to guide the inter-conversion of the product distribution of this enzyme and the close homolog henbane premnaspirodiene synthase (HPS) (72% amino acid identity) without negatively affecting activity.⁵

Keasling *et al.* used a humulene synthase homology model to identify active site residues.¹⁰ For each of 19 amino acids lining the active site, site-saturation mutagenesis was constructed to generate libraries of mutants. Six mutants containing up to 5 mutations each demonstrated impressive selectivities for products up to 85.7%, and several of these mutants demonstrated activity comparable to that of the parent. The method unfortunately required extensive and tedious use of low-throughput GC-MS measurements. Keasling and co-workers conducted a separate study on the selective *Gossypium arboreum* (+)- δ -cadinine synthase.⁴ Random mutagenesis by epPCR and a high-throughput solubility screen based on fusion to chloramphenicol acetyltransferase (CAT) were used to identify soluble mutants. From 1000 clones of the epPCR library, 100 mutants were re-screened for activity using GC-MS, for each of two rounds of directed evolution. A total of 20 clones were identified that, in addition to the native product (+)- δ -cadinene, also produced the new product germacrene D-4-ol. The best enzyme in this study was a double mutant that demonstrated 53% selectivity for germacrene D-4-ol with only slight decreases in kinetic properties.²²

The double mutant from Keasling's study and 17H2 from the current study both produce germacrene D-4-ol.⁴ Keasling's hypothesis for the increased production of germacrene D-4-ol was that the mutations they found allowed water into the active site. When water can reach the active site it leads to the termination of the reaction cascade by hydroxylation. In absence of water, however, another cyclization occurs before the reaction is quenched.

In the case of Cop2, water quenching is responsible for formation of both of the major products— α -cadinol and germacrene D-4-ol. This fact suggests that simple ingress of water into the active site is not responsible for the increase in selectivity for germacrene D-4-ol

mediated by 17H2. Instead, the three mutations may remove non-bonding interactions that would favor the second cyclization to form α -cadinol.

Residues that cause an alteration of product selectivity upon mutation in sesquiterpene synthases have been called plasticity residues.¹⁰⁻¹¹ The search for plasticity residues has primarily been limited to active site residues regardless of whether active site residues are ideal for this purpose. In Cop2, residues L59, T65, and S310 are not predicted to be active site residues, yet they are plasticity residues that both alter product distributions and maintain activity. The ideal plasticity residues may in fact be located remote from the active site, and this appears to be the case for Cop2. Further application of random mutagenesis and the high-throughput cyclization activity screen used here may reveal other instances of plasticity residues located in positions not currently amenable to structural prediction.

Experimental

Instrumentation

Gas chromatography (GC) was performed on a Shimadzu GC-17A instrument equipped with an FID detector. The column was an HP-5 column (Agilent, 30 m x 0.32 mm x 0.25 μ m film). For kinetic determinations, the temperature program was an initial temp of 90 °C for 1.0 min, followed by a temp ramp of 9.0 °C per min to 174 °C, followed by a ramp of 45 °C per min to 260 °C, and the temp was held at 260 °C for 2.0 min. The split ratio was 1.0, column flow was 1.68 mL/min, the injector temp was 250 °C, and the detector temp was 300 °C. Absorbance data from 96-well microtitre plates was collected on a Tecan Infinite[®] M200. Proteins were purified on an ÄKTAexpress with 1 ml HisTrap columns (GE Healthcare Life Sciences). GC-MS

measurements were conducted on a Shimadzu GCMS-QP2010 SE instrument in electron ionization (EI) mode. The column used was a Shimadzu SHRXI-5MS (30 m x 0.25 mm x 0.25 μ m film). The temperature program was an initial temp of 75 °C for 6.0 min, followed by a temp ramp of 9.0 °C per min to 180 °C, followed by a temp ramp of 35 °C per min to 280 °C, and the temp was held at 280 °C for 3.0 min. The split ratio was 20, column flow was 1.5 mL/min and the injector temp was 250 °C.

General procedures

Alcohol oxidase (AOX) was obtained from MP Biomedicals. Surrogate 1 (**S1**) was synthesized as described.¹³ Farnesyl diphosphate (FPP) was synthesized from farnesol according to the method of Poulter.²³ Primers were purchased from Integrated DNA Technologies (IDT). *E. coli* Express electrocompetent *E. coli* BL21 (DE3) was purchased from Lucigen. Alkaline phosphatase, Taq polymerase, T4 ligase, and restriction enzymes DpnI, XhoI, and NdeI were purchased from New England Biolabs (NEB). DNA sequencing was performed by Laragene (Culver City, CA, USA). TB refers to terrific broth from RPI, and LB refers to Luria broth from the same vendor. The Cop2 gene was a gift from Professor Claudia Schmidt-Dannert and was supplied as pUC-Cop2. The terpene synthase gene 03688 in pET-28a(+) was a gift from Professor David E. Cane. Standard Buffer is 50 mM PIPES, 10 mM MgCl₂, 100 mM NaCl, 2 mM DTT at pH 7.6.

Cloning of Cop2 into pET-22b

The gene for Cop2 was amplified from the vector pUC-Cop2 with the primers Cop2-for
TTAACTTTAAGAAGGAGATATACATATGCACCACCACCACCACCCAAGTCCCGCGG
GAGC and Cop2-rev

GGGCTTTGTTAGCAGCCGGATCTCATTAGGCCAAAACCCGCACAGC using Phusion polymerase (NEB) with HF buffer and 60°C as the annealing temperature. The part of the sequence that is homologous to the vector backbone is in italics; the start codon is underlined. The vector backbone was amplified from pET-22b using the primers pET22-for *GCTTTGTTAGCAGCCGGATCTCA* and pET22b-rev *CCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACAT* and the LongRange PCR Kit from Qiagen with annealing temperature of 50°C. After separation on a 1% agarose gel the PCR products were excised and purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research) following the procedures suggested by the manufacturer. 3 µl of the gene and 2 µl of the vector backbone were combined with 15 µl of the mix described for the one-step isothermal DNA assembly by Gibson et al.²⁴ and the assembly was performed for 60 min at 50 °C. An aliquot of 5 µl of the resulting solution was transformed into chemically competent *E. coli* DH5α. After plasmid purification with the Zyppy™ Plasmid Miniprep Kit (Zymo Research), correct assembly was verified by sequencing (Laragen). However a point mutation resulting in a frame shift and premature abortion of the translation was identified. Therefore the gene sequence was fixed in a site-directed mutagenesis using the primers Cop2mt-for *GACGAAATCATATTTTTCTCGATTCTGAAGACCGCTG* and Cop2mt-rev *GAGAAAATATGATTTTCGTCATTAGTCCTGCGCGTTTATC* and Phusion polymerase (NEB) with HF buffer and 60°C as the annealing temperature. The base pair to be inserted is underlined. After digestion for 120 minutes at 37°C with DpnI (Neb) 5 µl were transformed into chemically competent *E. coli* DH5α. After plasmid purification with the Zyppy™ Plasmid Miniprep Kit (Zymo Research) the success of the base insertion was verified by sequencing (Laragen). The plasmid was then transformed into BL21(DE3) for protein overexpression.

Cop2 library formation

A library of Cop2 mutants was produced through error-prone PCR using the primers pETCop2Fwd2 (AAAAAACCATATGCCAAGTCCCGCGGGAG) and pETCop2Rev2 (TTTTTCTCGAGTTTAGGCCAAAACCCGCAC). The template was the Cop2 gene in pET-22b, using Taq (NEB) along with 200 μ M MnCl₂ and 0.2 mM of each dNTP (Roche). These conditions gave an approximate mutation rate of 1.4 base pairs per gene. The inserts produced by these epPCR reactions were double digested with XhoI and NdeI (NEB). A pET-28a(+) plasmid with a low-activity terpene synthase gene (sscg_03688, from Professor David E. Cane) was double digested with XhoI and NdeI (NEB) to remove the gene and then dephosphorylated using calf intestinal alkaline phosphatase (NEB). Both digestions were followed by gel purification and extraction using a kit (Promega) to produce purified DNA. The insert and backbone were ligated using T4 ligase (NEB) at 16 °C overnight and the DNA was cleaned and concentrated using a kit (Zymo Research). Electro-competent BL21 (DE3) cells were transformed with 1 μ L of cleaned ligation product and the resulting colonies were hand-picked individually into 96-well shallow plates (1 mL volume per well) using autoclaved toothpicks. The 96-well plates contained 300 μ L of LB media per well and 0.05 g/L of kanamycin. Cultures were grown overnight at 37 °C while shaking at 225 rpm and 80% rh. Next, 50 μ L aliquots of the cultures were transferred into 96-well deep well plates (2 mL volume per well) containing 800 μ L of TB media as well as 0.05 g/L of kanamycin. Cryostocks of the libraries were made during transfer by transferring 120 μ L per well of the cultures into 96-well microtiter plates in 20% glycerol. These cryostocks were flash-frozen on dry ice and stored in the -80 °C freezer. The deep-well expression plate cultures were grown at 37 °C for 4 h shaking at 225 rpm and 80% rh. Protein production was then induced by adding 50 μ L per well of IPTG in LB media for a total IPTG concentration of 0.5 mM. The

plates were then returned to the shaker to sit at 25 °C with 225 rpm and 0% rh for 24 h. The cells were then pelleted by centrifugation at 3,000 g for 10 min at 4 °C. The plates containing the cell pellets were frozen at -20 °C overnight or until screening protocol began.

High-throughput screening

Screening for all libraries and substrates was initiated by thawing deep well plates with cell pellets at room temperature for 30 min. Then, to lyse the cells, 300 µL of lysis buffer were added to each well. The lysis consisted of Standard Buffer (pH: 7.6, 50 mM PIPES, 10 mM MgCl₂, 100 mM NaCl, 2 mM DTT) containing 10% betaine, 0.5 mg/mL lysozyme and 0.02 mg/mL DNase I. The plates were vortexed to resuspend the pellets, and then were incubated at 37 °C for 1 h. Cell debris was pelleted by centrifugation at 5,000 g at 4 °C for 15 min. Using a liquid-handling robot, 100 µL of lysate from each well was transferred to 96-well microtiter plates, and then 100 µL of 0.50 mM **S1** in Standard Buffer was added.

After the lysate and **S1** were incubated for 1 h at 25 °C, 10 µL of alcohol oxidase (50 µL of AOX stock solution dissolved in 950 µL of 0.1 M KPi, pH 8.0) were added to each well, and the plates were shaken on a table top orbital shaker at 600 rpm for 10 minutes at room temperature. Next 16 µL of 0.5 M ethylenediaminetetraacetic acid (EDTA, pH 8.0) were added to each well, and the plates were shaken at 600 rpm for 1 min at room temperature. Finally, 50 µL of 0.16 M Purpald in 2 M NaOH was added to each well. The plates were shaken at 600 rpm for 25 min at room temperature, after which a microtiter plate reader was used to record the absorbance of the wells at 550 nm.

Construction of the single mutants of 17H2 via fragment assembly PCR

In order to analyze the three mutations of 17H2 individually, the corresponding single mutants L59H, T65A and S310Y were prepared. The forward and reverse primers with the respective mutations were designed and the PCR reactions were prepared with 65 ng Cop2 DNA template, 0.5 μ L Phusion Polymerase, 1 μ L of forward and reverse primer [10 μ M], HF Buffer 5x, and 1 μ L dNTPs [10 mM].

The following temperature program was used for the PCR for 30 cycles.

Temperature	Time
98 °C	30 sec
98 °C	10 sec
58 °C	20 sec
72 °C	15 sec/1 kb
72 °C	10 min
10 °C	until end

The samples were loaded on a 1% agarose gel and purified using the Promega Kit for DNA extraction and purification. The DNA was eluted in 50 μ L dH₂O and its concentration was determined with a spectrophotometer (Nanodrop). The corresponding fragments were combined via fragment assembly PCR and the amount of the large and the small fragments were set according to their relative sizes. After 15 cycles with an extension time of 20 sec, the primers

Cop2fwd2 and Cop2rev2 were added and another 20 cycles were started to amplify the full construct.

The following primers were used for the preparation of the single mutants of Cop2:

Primer	Nucleotide sequence
pETCop2Fwd2	AAAAAACCATATGCCAAGTCCCGCGGGAG
pETCop2Rev2	TTTTTCTCGAGTTTAGGCCAAAACCCGCAC
Cop2L59HrevCor	CTCGCCCGCTTTGTGTGCCATAAAC
Cop2L59Hfwd	GTTTATGGCACACAAAGCGGGCGAG
Cop2T65AfwAg	CGAGCTCGCTGCTGCATGC
Cop2T65ArevAg	GCATGCAGCAGCGAGCTCG
Cop2S310Yfwd	GGGTCTCTAAATTGGTACATCGACGGGACA
Cop2S310Yrev	TGTCCCGTCGATGTACCAATTTAGAGACCC

The DNA inserts were gel extracted and purified and 1 μ g DNA was digested with 1 μ L XhoI and 1 μ L NdeI for 1 h at 37 °C. The same applies for the vector DNA. The pET28a(+) vector backbone was cut with XhoI and NdeI for 1h and dephosphorylated with 1 μ L Alkaline Phosphatase (Calf Intestine, NEB) for 30 min.

Ligation and transformation

After another purification step by gel extraction, ligation of the cut insert (10 μL) and the digested and dephosphorylated vector backbone (2.5 μL) was done with T4 DNA Ligase (1 μL) over night at 16 $^{\circ}\text{C}$. The ligations were worked up with the DNA Clean & Concentrator kit (Zymo Research) and eluted in 6 μL dH_2O . For the transformation 1 μL of the concentrated ligation was added to 50 μL of electrocompetent *E. coli* BL21 (DE3) cells. The cells were transferred into a cooled electroporation cuvette and pulsed with a voltage of 2.5 kV. Immediately afterwards 950 μL SOC media was added and the cells were incubated for 1 h at 37 $^{\circ}\text{C}$ and 225 rpm. After recovery 30 μL und 200 μL of the cell suspension were plated on LB_{Kan} agar plates with a kanamycin sulfate concentration of 50 mg/L. The plates were incubated over night at 37 $^{\circ}\text{C}$.

In order to identify positive clones, overnight cultures were made and Minipreps were prepared with the QIAprep Spin Miniprep Kit and sent for sequencing to Laragen with the primers T7 Promoter and T7 Terminator. Glycerol stocks with a final glycerol concentration of 20 % were made from clones with the correct sequences, frozen on dry ice and stored at -80 $^{\circ}\text{C}$.

Cultivation and protein expression

The proteins were overexpressed in *E. coli* BL21. Two overnight cultures were made from the strain of interest by inoculating 5 mL of LB_{Kan} medium with cells from a cryostock or from a plate. The cultures were grown at 37 $^{\circ}\text{C}$ and 225 rpm overnight. For the main cultures, 1 L TB media was prepared in 2 L Erlenmeyer flasks and autoclaved before the antibiotic kanamycin sulfate was added to a final concentration of 50 mg/L. The TB_{Kan} media were inoculated with the overnight cultures and incubated at 37 $^{\circ}\text{C}$ and 200 rpm. The ODs at 600 nm were measured with a spectrophotometer. After ~3-4 h $\text{OD}_{600\text{s}}$ of 0.5-1.0 were reached and protein expression was

induced by the addition of IPTG to an end concentration of 0.5 mM (0.12 g IPTG were dissolved in 5 mL dH₂O and added to 1 L culture media). After the induction, the temperature was decreased to 25 °C and the cultures were incubated for another 24 h at 200 rpm. Then the cells were harvested by centrifugation at 4,000 x g and 4 °C for 15 min. The pellets were resuspended in ~20 mL Buffer A (pH: 7.0, 50 mM Tris/HCl, 5 mM MgCl₂, 300 mM NaCl, 2 mM DTT, 10 % (v/v) glycerol, 10 mM imidazole for Buffer A and 500 mM for Buffer B) and frozen at -20 °C.

Protein purification

After thawing the cell pellets, the lysis of the resuspended cells was performed by ultrasonication for 2 min (pulse on 5 sec, pulse off 5 sec). The cell lysates were constantly kept on ice and after sonication they were centrifuged at 20,000 x g for 30 min at 4 °C. The supernatants were filtered through 0.2 µm cellulose filters and if necessary 0.02 mg/mL DNase was added. The purification of the proteins was performed with an ÄKTAexpress system from GE Healthcare. Since the proteins were hexahistidine-tagged, 1 mL HisTrap Ni-NTA columns were used for the purification and an imidazole gradient was applied for elution of the protein. The fractions containing protein were identified by monitoring the absorbance at 280 nm, and then pooled. The buffer was exchanged to Standard Buffer containing 10 % betaine (w/v) using Amicon® Ultra-15 Ultracel® cellulose 10,000 MW-cutoff centrifugal filters. The purified proteins were aliquoted and stored at -80 °C until they were further used. The protein concentration was determined by measuring the absorbance at 280 nm. Therefore the protein was denatured in 500 µL 6 M guanidinium chloride and the average of three different measurements was calculated. To check the purity and molecular weights of the proteins, SDS-PAGE analysis was performed.

GC analysis

To test the conversion of the natural substrate FPP, GC-scale reactions were performed using an enzyme concentration of 2 μM and a substrate concentration of 300 μM . The reactions were performed in 2 mL GC vials in a total reaction volume of 500 μL . First, Standard Buffer containing 10% betaine was added to the vials. Enzyme and substrate were added and the reactions were overlaid with 500 μL hexanes containing 200 μM dodecane as internal standard (4.6 μL dodecane per 100 mL hexanes). The samples were closed with a septum cap and in the case of Cop2 they were incubated at 30°C for 2 h. The incubation time and temperature were chosen depending on the current enzyme and experiment. Afterwards the reactions were stopped by adding 40 μL of EDTA (0.5 M, pH 8) and vortexing. Sodium sulfate was added to glass pipettes that were stuffed with small pieces of KimwipesTM tissue and 250 μL of the organic layer were then filtered through the sodium sulfate. The filtrates were collected in a GC vial with a low-volume insert and used for GC analysis. Relative quantifications were determined using ion abundances.⁴ Reactions were run in triplicate.

Kinetic measurements

In order to estimate the kinetic parameters of the parent Cop2 and the triple mutant 17H2, GC analysis was conducted as above with different substrate concentrations. All reactions were performed in 2 mL GC vials in 500 μL Standard Buffer with 10% betaine. In the case of the wild type Cop2, an enzyme concentration of 0.2 μM was used, whereas for the mutant 17H2 a concentration of 0.5 μM was used. FPP concentrations from 6.25-100 μM and from 10-200 μM were used for the wild type and mutant enzymes, respectively. All samples were closed with a septum cap and incubated at room temperature. The reactions were stopped after 5, 10, 15, 20, 30, 40, 50 and 60 min by the addition of 40 μL of EDTA (0.5 M, pH 8) and vortexing. The organic layer was filtered through sodium sulfate as described above and used for GC

measurements. Double determinations were made for all samples. The curve fitting was done with Prism. The same results were obtained with the program Mathematica. The product formation increased linearly for the first 30-40 min. The peak areas were converted into estimated product concentrations with the help of the internal standard dodecane with a known concentration of 200 μM . Therefore, the products are quantified as dodecane equivalent μM units. Though this method is only an estimation of product concentrations, the use of the same comparison for both Cop2 and 17H2 gives a useful approximation of relative kinetic behavior.

Identification of germacrene D-4-ol and α -cadinol

Compounds were identified by comparison of their retention time, mass fragmentation patterns, and Kovat's indices to known values. Fragmentation patterns were searched for similarity in the NIST library in the Shimadzu 2010 GC-MS software. Kovat's indices were determined by running a standard sample containing linear alkanes from C10 to C16. Germacrene D-4-ol was readily determined using these methods. The retention times and fragmentation patterns of α -cadinol, however, are quite similar to other diastereomers. To aid in product identification, T-muurolol synthase sscg_03688 was expressed in *E. coli* containing the gene in pET-28a(+) as provided by David E. Cane.²⁵ The enzyme was purified as described in the literature.²⁵ GC-scale reactions were conducted using 2 μM of enzyme for Cop2, 17H2, and sscg_03688 and 300 μL of FPP. Cop2 and 17H2 reactions were run in standard buffer with 10% w/v betaine. The sscg_03688 reaction was run in standard buffer with 20% glycerol. The reactions were incubated overnight at room temp and worked up and analyzed by GCMS as described above. Analysis of the product by GCMS showed that the product from Cop2 had a slightly longer

retention time than T-muurolol, and a similar mass spectral fragmentation pattern, and was therefore identified as α -cadinol.²⁶⁻³³

Conclusions

In this study, a multi-product terpene synthase, Cop2, was converted to a more selective mutant by random mutagenesis. The results suggest that random mutagenesis and screening with an assay for cyclization activity may be an effective strategy to identify residues that influence product selectivity without imposing activity penalties. No structural information was used to guide the introduction of mutations, and the mutations were found at residues that would be non-obvious targets for mutagenesis in structure-based approaches.

Acknowledgements

We are grateful to Frances H. Arnold, Claudia Schmidt-Dannert, David Cane, Scott Virgil, Robert M. Coates, and David Christianson for their pioneering studies, invaluable advice, encouragement, and materials. We thank Sabine Brinkmann-Chen, John McIntosh, Thomas Heel, and Chris Farwell for advice and assistance. RL acknowledges the support of NIH fellowship 1F32GM095061. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. KSR thanks the Deutscher Akademischer Austauschdienst (DAAD) for a postdoctoral fellowship. KZK and RZK acknowledge the support of Summer Undergraduate Research Fellowships from the California Institute of Technology.

References

1. E. Breitmaier, *Terpenes; Flavors, Fragrances, Pharmaca, Pheromones*, Weinheim: Wiley-VCH, 2006.
2. Dickschat, J. S. *Nat. Prod. Rep.* 2011, **28**, 1917.
3. R. K. Allemann, *Pure Appl. Chem.*, 2008, **80**, 1791.
4. Y. Yoshikuni, V. J. Martin, T. E. Ferrin, and J. D. Keasling, *Chem. Biol.*, 2006, **13**, 91.
5. B. T. Greenhagen, P. E. O'Maille, J. P. Noel and J. Chappell, *Proc. Nat. Acad. Sci.*, 2006, **103**, 9826.
6. B. Felicetti and D. E. Cane, *J. Am. Chem. Soc.*, 2004, **126**, 7212.
7. L. S. Vedula, J. Jiang, T. Zakharian, D. E. Cane and D. W. Christianson, *Arch. Biochem. Biophys.*, 2008, **469**, 184.
8. J. E. Diaz, C.-S. Lin, K. Kunishiro, B. K. Feld, S. K. Avrantinis, J. Bronson, J. Greaves, J. G. Saven and G. A. Weiss, *Prot. Sci.*, 2011, **20**, 1597.
9. F. Lopez-Gallego, G. T. Wawrzyn and C. Schmidt-Dannert, *Appl. Environ. Microbiol.*, 2010, **76**, 7723.
10. Y. Yoshikuni, T. E. Ferrin and J. D. Keasling, *Nature*, 2006, **440**, 1078.
11. J.-X. Li, X. Fang, Q. Zhao, J.-X. Ruan, C.-Q. Yang, L.-J. Wang, D. J. Miller, J. A. Faraldos, R. K. Allemann, X.-Y. Chen and P. Zhang, *Biochem. J.*, 2013, **451**, 417.
12. D. B. Little and R. B. Croteau, *Arch. Biochem. Biophys.*, 2002, **402**, 120.
13. R. Lauchli, K. S. Rabe, K. Z. Kalbarczyk, A. Tata, T. Heel, R. Z. Kitto and F. H. Arnold, *Angew. Chem. Int. Ed.*, 2013, **52**, 5571.
14. S. Agger, F. Lopez-Gallego and C. Schmidt-Dannert, *Mol. Microbiol.*, 2009, **72**, 1181.

15. A. Marchler-Bauer, C. Zheng, F. Chitsaz, M. K. Derbyshire, L. Y. Geer, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, C. J. Lanczycki, F. Lu, S. Lu, G. H. Marchler, J. S. Song, N. Thanki, R. A. Yamashita, D. Zhang, S. H. Bryant, *Nucleic Acids Res.*, 2013, **41**, D348.
16. J. M. Caruthers, I. Kang, M. J. Rynkiewicz, D. E. Cane and D. W. Christianson, *J. Biol. Chem.*, 2000, **275**, 25533.
17. C. A. Lesburg, G. Zhai, D. E. Cane and D. W. Christianson, *Science*, 1997, **277**, 1820.
18. SWISS-MODEL, Protein Structure Bioinformatics group at the Swiss Institute of Bioinformatics and the Biozentrum University of Basel <http://swissmodel.expasy.org/>
19. Y. Gao, R. B. Honzatko and R. J. Peters, *Nat. Prod. Rep.*, 2012, **29**, 1153.
20. D. W. Christianson, *Chem. Rev.*, 2006, **106**, 3412.
21. D. J. Miller and R. K. Allemann, *Nat. Prod. Rep.*, 2012, **29**, 60.
22. Keasling reports higher selectivity and productivity when no organic overlayer is used to sequester products. We do not use these data for comparison because the more volatile (+)- δ -cadinene is removed into the vapor phase and thereby biases both selectivity and productivity in favor of germacrene D-4-ol.
23. V. J. Davisson, A. B. Woodside, T. R. Neal, K. E. Stremmer, M. Muehlbacher and C. D. Poulter, *J. Org. Chem.* **2006**, *5*, 4768.
24. D. G. Gibson, L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison III and H. O. Smith, *Nat Methods*. 2009 May;**6**:343.
25. Y. Hu, W. K. W. Chou, R. Hopson and D. E. Cane, *Chem. Biol.* 2011, *18*, 32.
26. B. Pickel, D. P. Drew, T. Manczak, C. Weitzel, H. T. Simonsen and D.-K. Ro, *Biochem. J.* 2012, **448**, 261.
27. M. Hazzit, A. Baaliouamer, M. L. Faleiro, and M. G. Miguel, *J. Agric. Food Chem.*, 2006, **54**, 6314.

28. M. Jalali-Heravi, B. Zekavat and H. Sereshti, *J. Chromatogr. A*, 2006, **1114**, 154.
29. W. A. Asuming, P. S. Beauchamp, J. T. Descalzo, B. C. Dev, V. Dev, S. Frost and C. W. Ma, *Biochem. Syst. Ecol.*, 2005, **33**, 17.
30. R. Baranauskiene, P. R. Venskutonis, P. Viskelis and E. Dambrauskiene, *J. Agric. Food Chem.*, 2003, **51**, 7751.
31. G. Gkinis, O. Tzakou, D. Iliopoulou and V. Roussis, *Z. Naturforsch. C.*, 2003, **58**, 681.
32. V. Roussis, M. Tsoukatou, P. V. Petrakis, I. Chinou, M. Skoula and J. B. Harborne, *Biochem. Syst. Ecol.*, 2000, **28**, 163.
33. J. P. Mevy, A. Bousquet-Melou, S. Greff, J. Millogo and C. Fernandez, *Biochem. Syst. Ecol.*, 2006, **34**, 815.