Supporting Information

Substrate Pre-Folding and Water Molecule Organization Matters for Terpene Cyclase Catalyzed Conversion of Unnatural Substrates

Stephan C. Hammer[+], Per-Olof Syrén[+], and Bernhard Hauer*

Table of Contents

SUPPLEMENTARY NOTES

Chemicals. All chemicals and solvents were purchased from Sigma Aldrich (Steinheim, Germany), Alfa Aesar (Karlsruhe, Germany), Acros Organics (Nidderau, Germany) or Carl-Roth (Karlsruhe, Germany): Geraniol (Alfa Aesar, A13736, 97%), Triton X-100 (Sigma-Aldrich, X100).

Analytics. The ^{1}H and ^{13}C NMR spectra were recorded on an Avance 500 spectrometer (Bruker) operating at 500.15 MHz and 125.76 MHz, respectively. All spectra were recorded at room temperature in $CDCl_3$ Chemical shifts (δ) are expressed in ppm and referenced to tetramethylsilane (TMS, δ = 0 ppm). GC analyses were performed on a Shimadzu GC-2010 equipped with a flame ionization detector using a DB-5 capillary column (Agilent, 30 m × 250 μm × 0.25 μm) and hydrogen as carrier gas (linear velocity 30 cm/s). Further, using an Agilent 7890A GC system equipped with a flame ionization as well as a mass spectrometer combined by flow splitting, using a HP-5MS capillary column (Agilent, 30 m × 250 μm × 0.25 μm) and helium as carrier gas (constant pressure 14.937 psi). Injections were performed in split mode (split ratio 5:1). Squalene was measured with a temperature program starting at 120 °C for 3 min, followed by an increase of 30 °C/min to 320 °C, which was kept for 15 min. All other molecules were analyzed with the same method, starting at 50 °C for 3 min, then 6 °C/min to 120 °C, 10 °C/min to 150 °C, 15 °C/min to 170 °C, 20 °C/min to 200 °C, 25 °C/min to 250 °C and finally 30 °C/min to 310 °C which was kept for 5 min.

Expression and purification of *Aac*SHC. The gene encoding for SHC from *Alicyclobacillus acidocaldarius* (NCBI no. BAA25185.1) was amplified via PCR and cloned into a pET-22b(+)-vector system (Merck, Darmstadt, Germany). *Sac*I and *Nde*I were used as restriction sites. The plasmid was transformed into *Escherichia coli* DH5α and successful cloning was verified by DNA sequencing of the entire gene. For protein expression the vector construct was transformed in *E. coli* BL21(*DE3*) and glycerol stocks were prepared. The glycerol stock was used to inoculate a 5 mL LB medium pre-culture (100 mg/mL ampicillin), which was incubated overnight at 37 °C and 180 rpm. These cultures were used to inoculate 500 mL of TB medium (12g/L tryptone, 24g/L yeast extract, 0.4% glycerol, 89 mM potassium phosphate, pH 7.0) in 2 L Erlenmeyer flasks. Auto-induction[54] was used while incubating for 18 h at 37 °C (180 rpm). Cells were harvested (16.900 × g, 14 min, 4 °C), frozen in liquid nitrogen and stored at -80 °C.

Biotransformations. The frozen cell pellet was resuspended in lysis buffer (3 mL per gram pellet, 200 mM citrate, 1 mM EDTA, pH 6.0) and homogenized using a tissue homogenizer. After addition of phenylmethanesulfonylfluoride (1 mM) and DNAse (ca. 10 μg/mL), the resulting suspension was disrupted by either passing through a high-pressure homogenizer (EmulsiFlex-C5, Avestin, 100 MPa) twice, or by ultrasonic treatment (Branson Sonifier 250, duty cycle 35%, output control 4, 3 × 1 min). The suspension was centrifuged (38.700 × g, 45 min, 4 °C) and the pellet was washed with wash buffer (3 mL per gram pellet, 60 mM citrate, pH 6.0). After centrifugation (38.700 × g, 45 min, 4 °C), the pellet was resuspended in solubilisation buffer (1 mL per gram pellet, 1% detergent, 60 mM citrate, pH 6.0) and incubated overnight by gently mixing at 4 °C. Triton X-100 was used as detergent. After centrifugation (38.700 × g, 45 min, 4 °C) the pellet was discarded, the solution was subsequently heat shocked at 50 °C for 15 min and precipitated *E. coli* proteins were removed via centrifugation. To obtain SHC solutions with higher purity the Triton X-100-solubilized SHC variants were further purified by diluting with water (1:5, Milli-Q water) and loading onto a column with DEAE-Sepharose (Sigma Aldrich, diethylaminoethyl-sephacel). The bound proteins were washed with equilibration buffer (0.2% detergent, 12 mM citrate, pH 6.0) and eluted with elution buffer (0.2% detergent, 12 mM citrate, 200 mM NaCl, pH 6.0). The

fractions with triterpene cyclase (assessed by SDS-PAGE analysis) were pooled and the concentration was determined with Bradford Ultra (Expedeon) using bovine serum albumin as standard. The reactions were performed with 0.2% detergent, 12 mM citrate buffer pH 6.0, 1.0 mol% catalyst and 2 mM substrate. To set up the reactions, squalene was emulsified in buffer (0.2% Triton X-100, 60 mM citrate, pH 6.0) to a final concentration of 2 mM. For the reaction 100 μL of this solution were mixed with SHC solution and buffer (0.2% Triton-X100, 60 mM citrate, pH 6.0). The other substrates were used to generate a DMSO stock solution (200 mM). For the reactions 10 μL of the substrate DMSO stock solution was mixed with SHC solution and buffer (1.0 mol% enzyme). The substrates were converted at 50°C for 40 h. Reactions in buffer and with the D376C mutant (knockout of the Brønsted acid) were used as negative controls. After the conversion, the reaction mixture was extracted with ethyl acetate (2×800 μL), the combined organic phase was dried with sodium sulfate and analyzed by gas chromatography.

Synthesis of geranyl alkyl ethers. Geranyl alkyl ethers were synthesized by Williamson ether synthesis, similar to a protocol described previously.[1] The corresponding alcohol (8.0 mmol, 1.0 equiv) was dissolved in 5 mL dry THF and added dropwise to a suspension of sodium hydride (1.1 equiv) in 20 mL dry THF. The mixture was stirred at room temperature for 2 h. Geranyl bromide (0.8 equiv) was dissolved in 15 mL DMSO, the solution was added to the reaction mixture and stirred for 22 h. The mixture was poured onto ice water and extracted with diethyl ether for three times. The combined organic layers were dried over Na_2SO_4 concentrated in vacuo and purified by flash column chromatography using silica gel 60.

(*E*)-1-Ethoxy-3,7-dimethylocta-2,6-diene (Geranyl ethyl ether): *n*-hexane and dichloromethane was used for column chromatography. Geranyl ethyl ether was isolated as colorless oil (0.29 g, 1.6 mmol, 25 % yield). R_f = 0.4 (*n*-hexane and dichloromethane 1:1). ^{1}H-NMR (500 MHz, $CDCl_3$): δ (ppm) 5.36 (t, J = 6.45 Hz, 1H), 5.10 (t, J = 6.74 Hz, 1H), 3.98 (d, J = 6.50 Hz, 2H), 3.48 (q, J = 6.90 Hz, 2H), 2.10 (m, 2H), 2.03 (m, 2H), 1.68 (s, 3H), 1.67 (s, 3H), 1.60 (s, 3H), 1.21 (t, J = 7.0 Hz, 3H). ^{13}C-NMR (125 MHz, $CDCl_3$) δ (ppm) 139.8, 131.6, 124.1, 121.0, 67.1, 65.4, 39.6, 26.4, 25.7, 17.7, 16.4, 15.3. HREIMS m/z 182.1666 (calculated for $C_{12}H_{22}O^+$ 182.1665).

(*E*)-1-Butoxy-3,7-dimethylocta-2,6-diene (Geranyl butyl ether): *n*-hexane and dichloromethane was used for column chromatography. Geranyl butyl ether was isolated as colorless oil (0.79 g, 3.5 mmol, 39 % yield). R_f = 0.35 (*n*-hexane and dichloromethane 1.5:1). ^{1}H-NMR (500 MHz, $CDCl_3$): δ (ppm) 5.35 (t, J = 6.47 Hz, 1H), 5.10 (t, J = 6.47 Hz, 1H), 3.97 (d, J = 6.70 Hz, 2H), 3.41 (t, J = 6.71 Hz, 2H), 2.10 (m, 2H), 2.03 (m, 2H), 1.68 (s, 3H), 1.67 (s, 3H), 1.60 (s, 3H), 1.57 (quin, J = 6.85 Hz, 2H), 1.38 (sex, J = 6.75 Hz, 2H), 0.91 (t, J = 7.39 Hz, 3H). ^{13}C-NMR (125 MHz, $CDCl_3$) δ (ppm) 139.8, 131.6, 124.1, 121.1, 70.0, 67.3, 39.6, 31.9, 26.4, 25.7, 19.4, 17.7, 16.5, 14.0.

(*E*)-3,7-dimethyl-1-(octyloxy)octa-2,6-diene (Geranyl octyl ether): *n*-hexane and dichloromethane was used for column chromatography. Geranyl octyl ether was isolated as colorless oil (0.81 g, 3.0 mmol, 47 % yield). R_f = 0.3 (*n*-hexane and dichloromethane 2:1). ^{1}H-NMR (500 MHz, $CDCl_3$): δ (ppm) 5.35 (t, J = 6.40 Hz, 1H), 5.10 (t, J = 6.70 Hz, 1H), 3.97 (d, J = 6.67 Hz, 2H), 3.40 (t, J = 6.67 Hz, 2H), 2.10 (m, 2H), 2.03 (m, 2H), 1.68 (s, 3H), 1.66 (s, 3H), 1.60 (s, 3H), 1.56 (m, 2H), 1.28 (m, 10H), 0.88 (t, J = 6.94 Hz, 3H). ^{13}C-NMR (125 MHz, $CDCl_3$) δ (ppm) 139.7, 131.6, 124.1, 121.1, 70.3, 67.3, 39.6, 31.8, 29.8, 29.5, 29.3, 26.4, 26.3, 25.7, 22.7, 17.7, 16.5, 14.4. HREIMS m/z 266.2614 (calculated for $C_{18}H_{34}O^+$ 266.2604).

Molecular dynamics (MD) simulation. MD simulations on the triterpene cyclase from *Alicyclobacillus acidocaldarius* was based on the PDB file 1UMP[2] using YASARA[3] (Yet Another Scientific Artificial Reality Application) version 14.5.21. All missing hydrogens were added to the starting enzyme structure and the hydrogen network was optimized using the Amber99 force field. The aza-group in the prefolded C30 substrate analogue was replaced by a terminal isoprene unit to construct the natural substrate squalene. For geraniol and geranyl octyl ether the prefolded aza-analog was used as a starting template to construct the corresponding substrate. Crystallographic water molecules and the presumably stabilizing detergent molecule were kept. The structure was minimized through repeated steps of short molecular dynamics and energy minimizations, initially on all hydrogens, then by releasing fixed waters and finally on all atoms. Force field parameterization for substrates was obtained by the AUTOSMILES[4] methodology. The Y420W/G600F double mutant was constructed using the refined wild-type structure and by performing the corresponding amino acid alterations and subsequent minimization. All simulations were performed in a waterbox that contained approximately 13000 explicit water molecules. The pH was set to 6 with the adequate protonation states of enzyme side chains predicted by the built-in empirical method[5] in YASARA. The simulation cell was neutralized through the addition of 0.9% NaCl. The correct protonation state of active site residues was subsequently verified by visual inspection. Hence the catalytic acid (D376) was protonated as well as D374 of the DxDD motif. The MD-simulations were performed under the canonical ensemble at 298 K using a Berendsen thermostat (and the Amber99 force field). PME accounted for long-range electrostatics[6] during the MD simulations of the variant and wild-type enzyme. A cut-off of 7.86 Å for Van der Waals interactions was used.

After a short equilibration, 20 ns MD-trajectories were analyzed to assess the relative abundance of productive transition state structures through investigation of distances relevant for protonation and the first ring formation. A productive transition state structure was defined as i) having a distance between the acidic *anti* proton sitting on D376 and the substrate carbon within van der Waals distance ($d_{prot} \leq 2.9$ Å, see Fig. S2) ii) having a distance between the two substrate isoprene carbons involved in the first ring closure in the range 3.9 Å-4.1 Å and centered around the value of 4.01 Å obtained through high-level quantum mechanical calculations on a model transition state of squalene cyclization in the gas phase[7] iii) having an angle for proton transfer larger than 135 ° ($\Theta_{prot} > 135°$, Fig. S1).

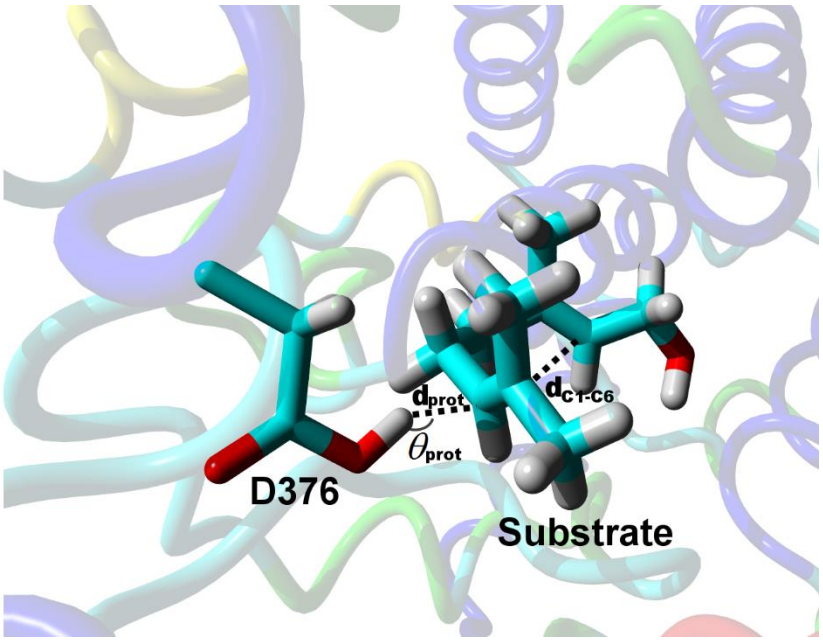


Figure S1: Key parameters analyzed from the MD-simulations. Θ_{prot} and d_{prot} refer to the OH---C angle and distance associated with proton transfer between the catalytic acid (D376) and the terminal isoprene group of the substrate. The parameter d_{C1-C6} refers to the intramolecular carbon-carbon distance of relevance for the initial cyclization step. The substrate, exemplified by geraniol, is shown in bold sticks.

SUPPLEMENTARY RESULTS

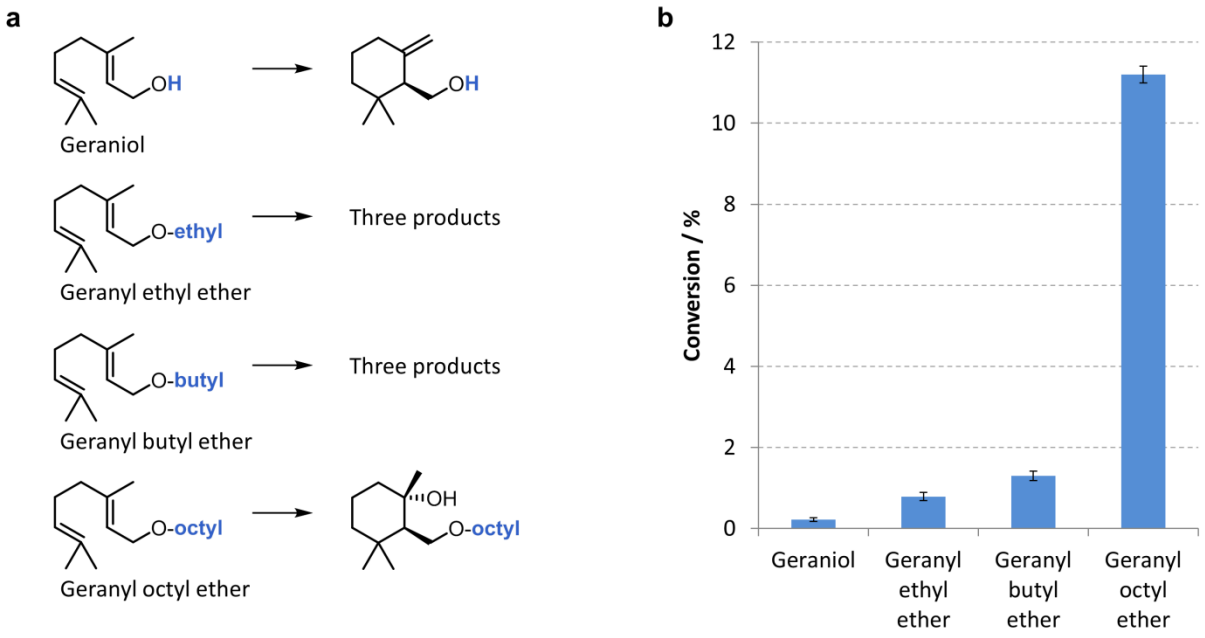


Figure S2: Conversion of geraniol and geranyl alkyl ethers with *Aac*SHC wild type. The conversion data for geraniol and geranyl alkyl ethers with *Aac*SHC wild type enzyme are shown. (**a**) Geraniol is converted to γ-cyclogeraniol as recently reported (confirmed by authentic standard).[8] Geranyl ethyl ether and Geranyl butyl ether generated several products, which were not further analyzed. Geranyl octyl ether selectively yielded the monocyclic alcohol (confirmed by authentic standard).[9] (**b**) The conversion data is plotted for every substrate. The error bars indicate the standard deviation from triplicate measurements.

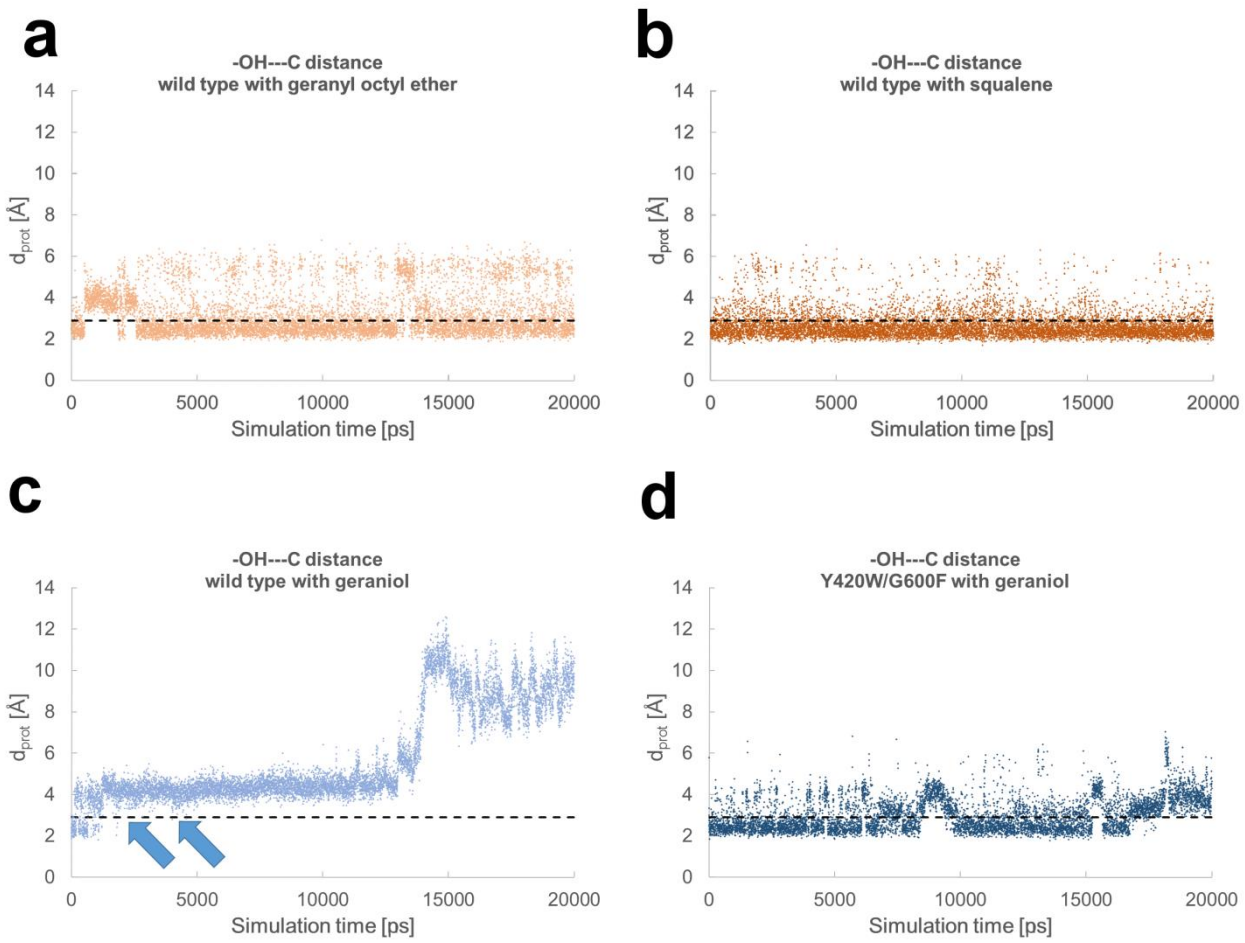


Figure S3: Distribution of protonation distances from 20 ns MD simulations. The dashed line corresponds to the van der Waals distance between hydrogen and carbon (2.9 Å) and defines snapshots as unproductive (over the line) or productive (below the line) for proton transfer. A) Data for wild type with geranyl octyl ether. B) Data for wild type with the natural substrate squalene. C) Data for wild type with geraniol. For clarity, the arrows point at snapshots with a productive -OH---C distance. D) Data for Y420W/G600F with geraniol. The -OH---C distance (d_{prot}) is defined in Fig. S1.

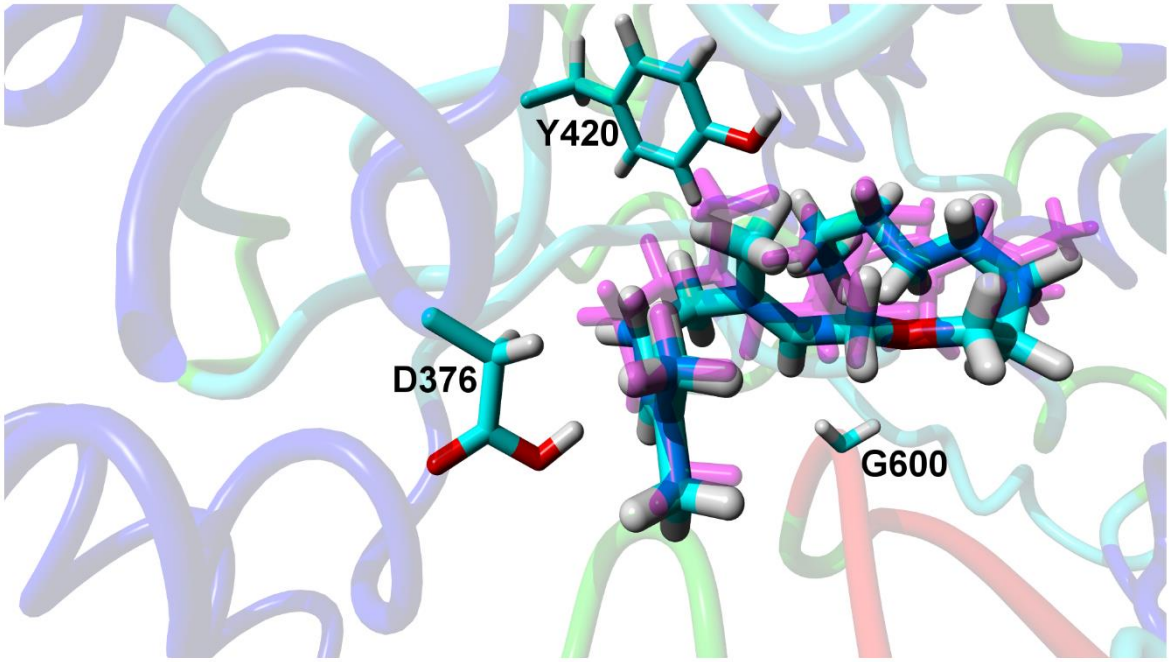


Figure S4: Superposition of an MD snapshot Superposition of an MD snapshot of *AacSHC* wild type complexed with octyl geranylether on an energy minimized structure of *AacSHC* wild type complexed with squalene. The catalytic acid (D376) and the positions targeted for mutagenesis are labeled. The geranyl octylether substrate is shown in bold sticks and squalene in magenta.

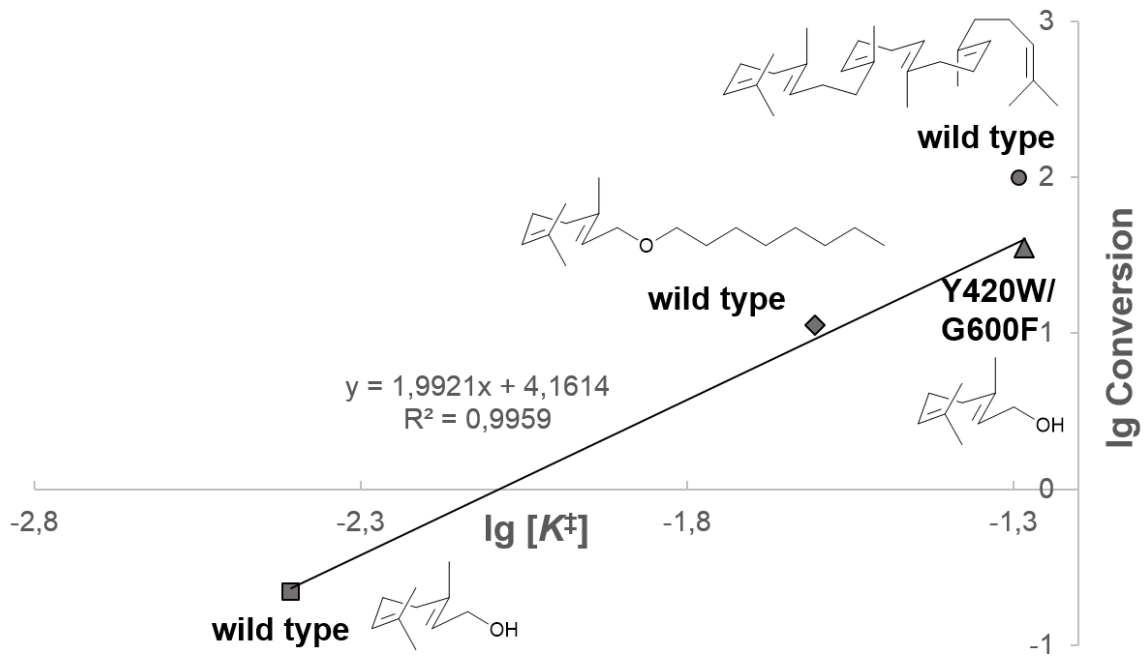


Figure S5: LFER-analysis. LFER-analysis based on the experimental and computational results stresses the prerequisite of productive substrate pre-folding for efficient catalysis. $K^{\ddagger}$ is the equilibrium constant for productive transition state formation and is calculated from the MD-simulations (see Table 1). The corresponding data for wild type (100% conversion) is shown by the circle.

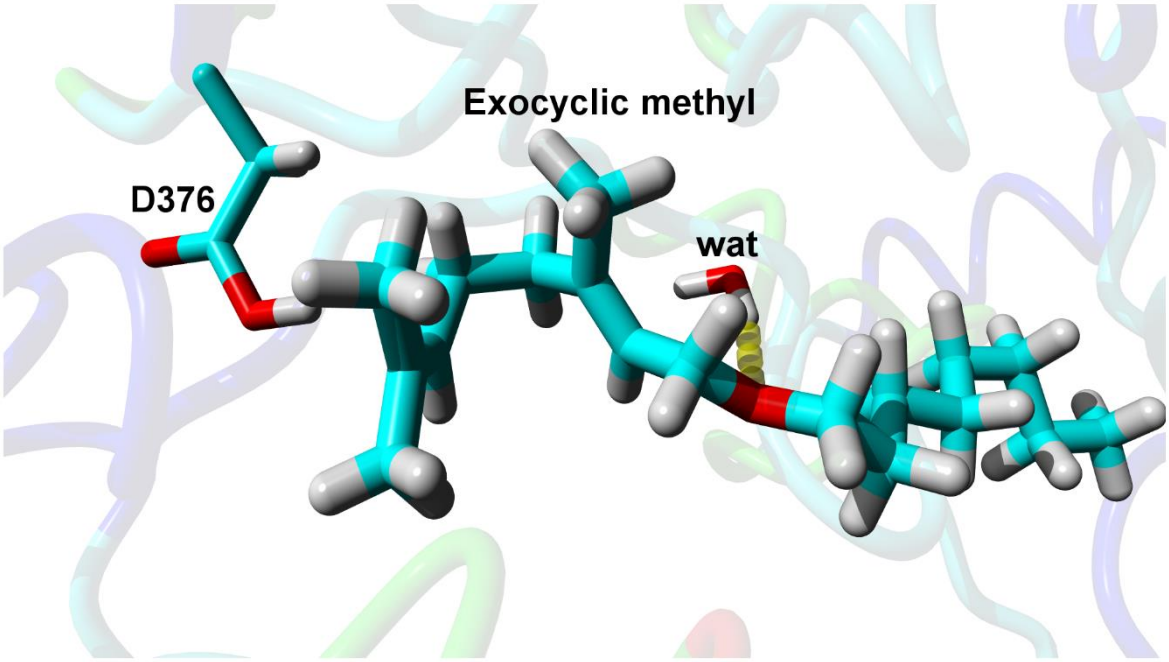
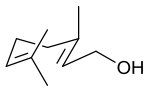
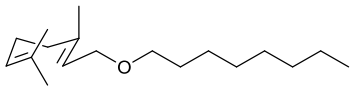
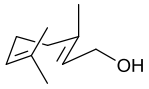


Figure S6: Snapshot from the MD simulations. The positioning of water molecules and their associated dynamics impact the catalytic outcome of the polycyclization cascade through either deprotonation of the exocyclic methyl group or nucleophilic attack. For geranyl octyl ether, the substrate backbone accommodates a water that preferably attacks (Table S3) the isoprene carbon corresponding to the generated tertiary carbocation. The catalytic acid (D376) and the productive water molecule for nucleophilic attack (wat) are labeled.

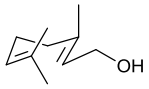
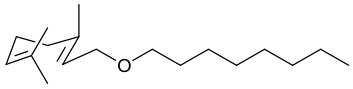
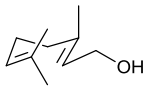
Table S1. Average distances and angles for the first ring closure obtained from 20 ns MD[a].

Variant	Substrate	d_{prot} [Å]	Θ_{prot} [°]	$d_{C1\text{-}C6}$ [Å]
wild type		5.9	138	5.0
wild type		3.2	115	3.6
Y420W/G600F		3.1	122	3.9
wild type		2.7	132	3.6

[a] See Fig. S1 for definitions of angles and distances.

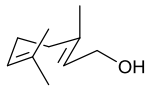
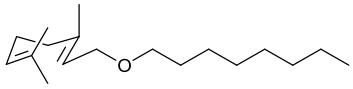
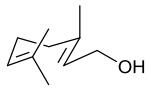
.

Table S2. The modulation of substrate shape and active site organization impact the placement and availability of water molecules for deprotonation.[a]

Variant	Substrate	Average number of water molecules within 5 Å from the exocyclic methyl group	Average CH---O_{wat} distance[b] [Å]	Average CH---O_{wat} angle[c] [°]
wild type		2.43	2.93	134
wild type		1.59	2.99	136
Y420W/G600F		1.15	3.57	131

[a] Based on 20 ns MD and analysis of the productive transition state structures according to the definition in Fig. S1. See also Fig. S6. [b] Refers to the average distance between an exocyclic methylene proton of the substrate and the closest water molecule. [c] Refers to the average angle between the exocyclic CH of the substrate and the most proximal water molecule.

Table S3. The modulation of substrate shape and active site organization impact the feasibility of water addition.

Variant	Substrate	Average number of water molecules within 5 Å[a]	Average C---O_{wat} distance[b] [Å]	Probability of water addition[c]
wild type		2.55	4.21	0.2
wild type		1.35	4.38	0.7
Y420W/G600F		0.70	4.64	0.15

[a] Of the electrophilic isoprene carbon of the substrate. [b] Refers to the average distance between the electrophilic isoprene carbon and a water molecule. [c] Based on manual analysis of 20 randomly distributed snapshots from the MD-simulations and corresponding to productive transition state structures for ring closure. Water addition was estimated to be feasible if the water molecule closest to the isoprene carbon had a lone pair pointing towards the electrophilic carbon.

SUPPLEMENTARY REFERENCES

1 S. Nakamura, K. Ishihara and H. Yamamoto, *J. Am. Chem. Soc.*, 2000, **122**, 8131–8140.

2 K. U. Wendt, K. Poralla and G. E. Schulz, *Science*, 1997, **277**, 1811–1815.

3 E. Krieger, T. Darden, S. B. Nabuurs, A. Finkelstein and G. Vriend, *Proteins*, 2004, **57**, 678–683.

4 A. Jakalian, D. B. Jack and C. I. Bayly, *J. Comput. Chem.*, 2002, **23**, 1623–1641.

5 E. Krieger, J. E. Nielsen, C. A. E. M. Spronk and G. Vriend, *J. Mol. Graph. Model.*, 2006, **25**, 481–486.

6 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

7 L. Smentek and B. A. Hess, *J. Am. Chem. Soc.*, 2010, **132**, 17111–17117.

8 S. C. Hammer, A. Marjanovic, J. M. Dominicus, B. M. Nestl and B. Hauer, *Nat. Chem. Biol.*, 2015, **11**, 121–126.

9 P.-O. Syrén, S. C. Hammer, B. Claasen and B. Hauer, *Angew. Chem. Int. Ed.*, 2014, **53**, 4845–4849.