Supporting Information

Chow et al. 10.1073/pnas.1505127112

SI Methods

Computational Methods.

Sequence similarity network of TPS. The sequence similarity network of the terpene synthase-like 2 subgroup was downloaded from the SFLD website and visualized with Cytoscape v2.8.3 using the organic yFiles layout.

Homology modeling and substrate docking. The crystal structure of epi-isozizaene synthase (PDB ID code 3LG5) was used as the template for building homology models of B5GLM7 and pentalenene synthase. The crystal structure was processed using the Schrödinger Protein Preparation Wizard, and all water of crystallization was removed. Homology modeling procedures are similar to our previous work (1). Briefly, we performed a multisequence alignment using PROMALS3D (2) and then created homology models using Schrödinger Prime (3). All of the structures were energy minimized by using restrained minimizations in the presence of their cocrystallized ligands (RMSD tolerance, 0.35 Å). The OPLS 2005 force field was used throughout this study (4).

Schrödinger Glide docking was used for all of the docking calculations (3, 5, 6). Glide XP scores, which are empirical energy functions for estimating binding free energies of the ligands, were used for substrate prediction. To ensure the ligands docked into the correct position, position restraints were added. The diphosphate moiety of the crystal structure 3LG5 is used as a reference with an RMSD tolerance of 1.5 Å. The position of the C1 carbon is consistent with the other related crystal structures of "terpene synthase-like 2 subgroup" such as 4OKZ, 4KUX, and 3V1V.

DFT calculations of geometries, ECD spectrum, and ^{13}C chemical shifts for cucumene. DFT calculations were carried out using Gaussian 09 software. Geometries were optimized at the B3LYP/6–31+G(d,p) level, followed by frequency calculations to ensure these were stationary points on the potential energy surface. For ECD and ^{13}C NMR calculations, geometries were further optimized with inclusion of the integral equation formalism polarizable continuum model implicit solvent model (the solvent for ECD was n-hexane and for NMR was benzene or $CHCl_3$). For ECD calculations, 30 excited states were solved (NStates = 30).

Experimental Methods.

Cloning, expression, and purification of B5GLM7. B5GLM7 was cloned into a pNIC28-Bsa4–based vector as previously described (7). Expression of B5GLM7 was carried out using *Escherichia coli* strain BL21(DE3). An overnight culture of *E.coli* harboring the expression vector was inoculated into 1 L Luria-Bertani (LB) medium containing 35 μg/mL kanamycin. The cells were allowed to grow to an OD_{600nm} of 0.6 before inducing with 0.1 mM isopropyl β-D-1-thiogalactopyranoside. Expression of B5GLM7 was then carried out at 25 °C for 24 h. The cells were harvested by centrifugation, resuspended in lysis buffer (50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl, and 10 mM Imidazole), and lysed by sonication. The lysate was clarified by centrifugation at 27,000 × *g* for 30 min, and the supernatant was loaded onto a column containing 5 mL Ni-NTA agarose (Qiagen). The column was washed with wash buffer (50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl, and 20 mM imidazole) and eluted with five column volumes of elution buffer (50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl, and 250 mM imidazole). The His-tag from B5GLM7 was cleaved using ProTEV Plus (Promega) and removed by passing the sample through a 120-mL Superdex 200 column (GE Healthcare). The protein was eluted in the presence of 50 mM Hepes, pH 7.6, 300 mM NaCl, and 10 mM β-mercaptoethanol. Five percent glycerol was added to the sample before storage at −80 °C. Each liter of LB routinely produced about 2–5 mg of purified B5GLM7.

TLC assay. Terpene synthase activity of B5GLM7 was assayed in 40 μL of 35 mM Hepes buffer, pH 7.6, containing 10 mM $MgCl_2$, 100 μM, 5 μCi/μmol of [^{14}C] GPP, [^{14}C] FPP, or [^{14}C] GGPP, and 160 μg B5GLM7. Incubations were conducted at 25 °C for 16 h. The diphosphate group was cleaved from the substrates using 3 mg acid phosphatase (MP Biomedicals) in the presence of 50 mM sodium acetate, pH 4.5, 0.1% Triton X-100, and 20% (vol/vol) n-propanol in a 300-μL reaction. The reaction was incubated at 30 °C for 4 h. The sample was extracted three times with methyl *tert*-butyl ether, dried on a SpeedVac, and spotted on a reverse-phase RP-18 $F_{254}S$ TLC plate (EMD). The plate was developed with 19:1 acetone:water, and radioactivity was detected by phosphorimaging.

GC/MS analysis. Samples for GC/MS analysis were prepared by procedures similar to those for the TLC assay with the exception that the reaction volume was scaled up to 1 mL. After incubation at 25 °C for 16 h, the sample was extracted three times with pentane, and the extracts were concentrated to ∼100 μL with a gentle stream of nitrogen. A 5-μL portion of the sample was loaded onto the Agilent DB-5ms capillary column (30 m × 0.25 mm, 0.25-μm film) and eluted with a temperature program of 60 °C for 2 min, a gradient of 2 °C/min from 60 °C to 120 °C, a gradient of 10 °C/min from 120 °C to 230 °C, and 230 °C for 5 min. High-resolution mass spectrometry analysis of purified cucumene by GC/MS was carried out by the Metabolomics Core at the University of Utah HSC Cores Research Facility.

Kinetic studies. Kinetic constants for B5GLM7 were determined in 200 μL of 35 mM Hepes buffer, pH 7.6, containing 10 mM $MgCl_2$, 89 nM B5GLM7, and 1–25 μM of 5 μCi/μmol [^{14}C]FPP. The assay mixture was incubated at 25 °C for 10 min before being quenched with 50 μL of 0.5 M EDTA, pH 8.0. The mixture was extracted three times with hexane. The extracts were mixed with 10 mL Ultima Gold mixture (PerkinElmer), and radioactivity was determined on a Tri-Carb 2910TR liquid scintillation spectrometer (PerkinElmer). Three replicates were obtained for each substrate concentration, and kinetic parameters were determined by fitting the initial rates to the Michaelis–Menten equation using GraFit 5 (Erithacus Software).

NMR characterization of cucumene. To characterize the product from B5GLM7, the kinetic assay conditions were scaled up to 400 mL. After completion, the mixture was extracted with pentane. The extracts were dried and reconstituted in 400 μL DMSO. The product was purified by HPLC using an XBridge BEH C18 column (4.6 × 50 mm; Waters).The product was eluted with a gradient of increasing acetonitrile from 1% to 100% over 30 min, followed by 100% acetonitrile for 15 min. The flow rate was maintained at 2 mL/min, and the product was detected at 214 nm. HPLC-purified material was re-extracted with pentane, dried with a gentle stream of nitrogen, and dissolved in $CDCl_3$ or C_6D_6. The samples were transferred to a 3-mm Norell NMR tube (Sigma), and NMR experiments were carried out at 25 °C on an Inova 600 spectrometer equipped with an HCN cryo-genic probe. Spectra were processed with VnmrJ 3.2 and visualized using MestReNova 7.1 (for 1D spectra) or Sparky 3.114 (for 2D spectra).

ECD spectroscopy. ECD spectra of monoterpene standards and cucumene were obtained on a Jasco J-815 CD spectrometer using parameters described by Tedesco et al. (8). Samples were dissolved using n-hexane within a concentration range of 0.5 to 30 mM. Quartz cells (1-mm path length) were used, and data points were collected between 180 and 250 nm.

1. Tian BX, et al. (2014) Predicting the functions and specificity of triterpenoid synthases: a mechanism-based multi-intermediate docking approach. *PLOS Comput Biol* 10(10): e1003874.
2. Pei J, Kim BH, Grishin NV (2008) PROMALS3D: A tool for multiple protein sequence and structure alignments. *Nucleic Acids Res* 36(7):2295–2300.
3. Anonymous (2014) Schrödinger Suite. Impact Version 6.5; Prime Version 3.8; MacroModel, Version 10.5 (Schrödinger, Inc., New York).
4. Banks JL, et al. (2005) Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J Comput Chem* 26(16):1752–1780.
5. Halgren TA, et al. (2004) Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem* 47(7):1750–1759.
6. Friesner RA, et al. (2004) Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 47(7):1739–1749.
7. Sauder MJ, et al. (2008) High throughput protein production and crystallization at NYSGXRC. *Methods Mol Biol* 426:561–575.
8. Tedesco D, Zanasi R, Wainer IW, Bertucci C (2014) Stereochemical and conformational study on fenoterol by ECD spectroscopy and TD-DFT calculations. *J Pharm Biomed Anal* 91:92–96.

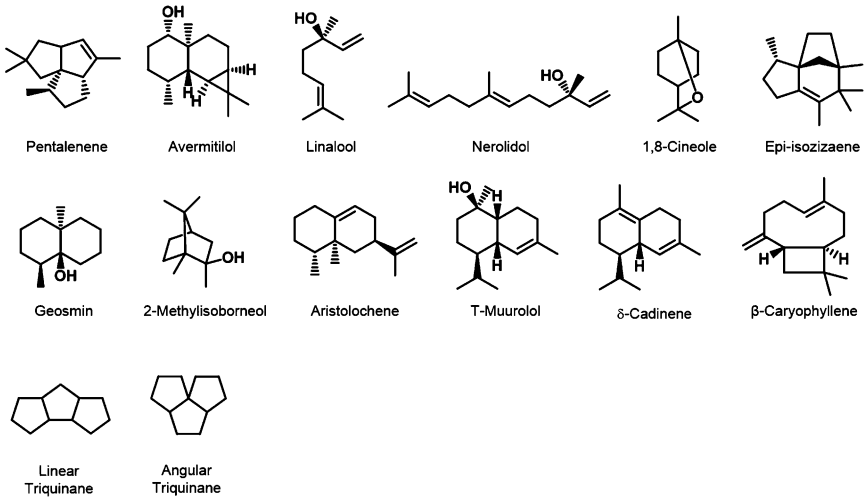


Fig. S1. Structures for compounds described in this study.

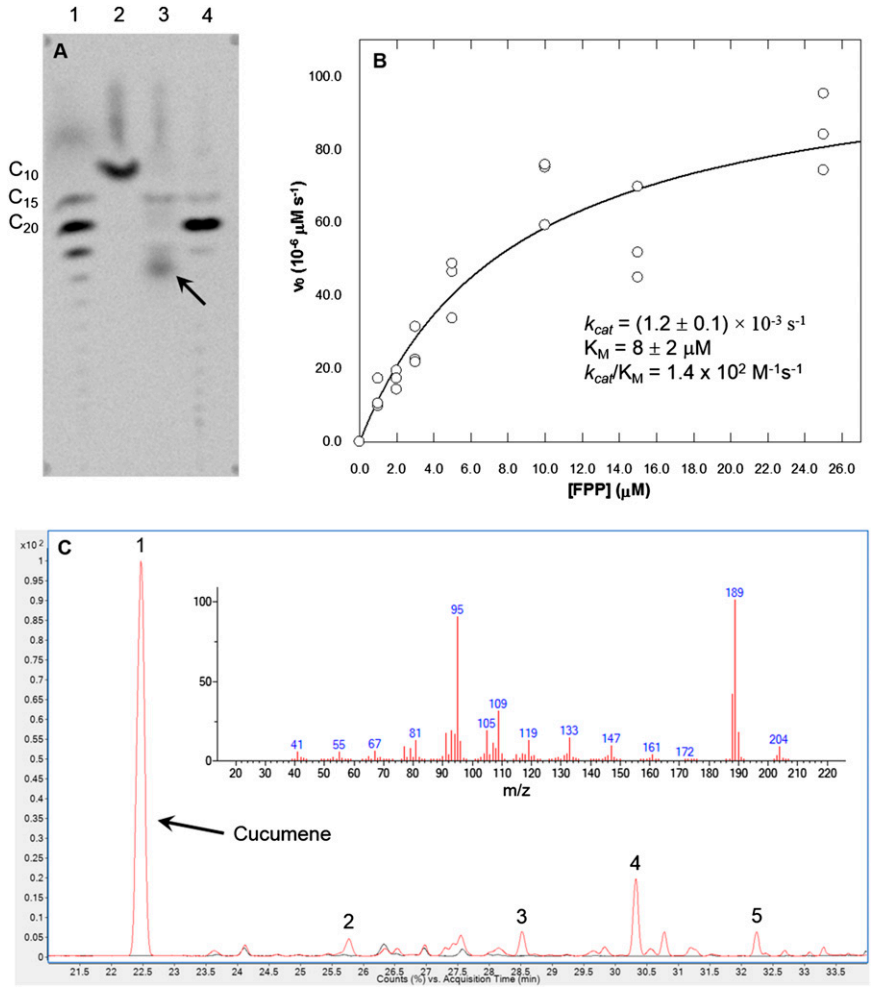


Fig. S2. Biochemical characterization of terpene synthase B5GLM7. (*A*) Product analysis using GPP, FPP, and GGPP. Lane 1, FPP synthase standard; lane 2, GPP; lane 3, FPP; lane 4, GGPP. (*B*) Michaelis-Menten plot of the reaction rate of B5GLM7 against different concentrations of FPP. (*C*) GC/MS analysis of the reaction catalyzed by B5GLM7, including the mass spectrum of cucumene (1), unknown sesquiterpenes (2–4), and β-Caryophyllene (5). Black, control reaction without enzyme; red, reaction with B5GLM7.

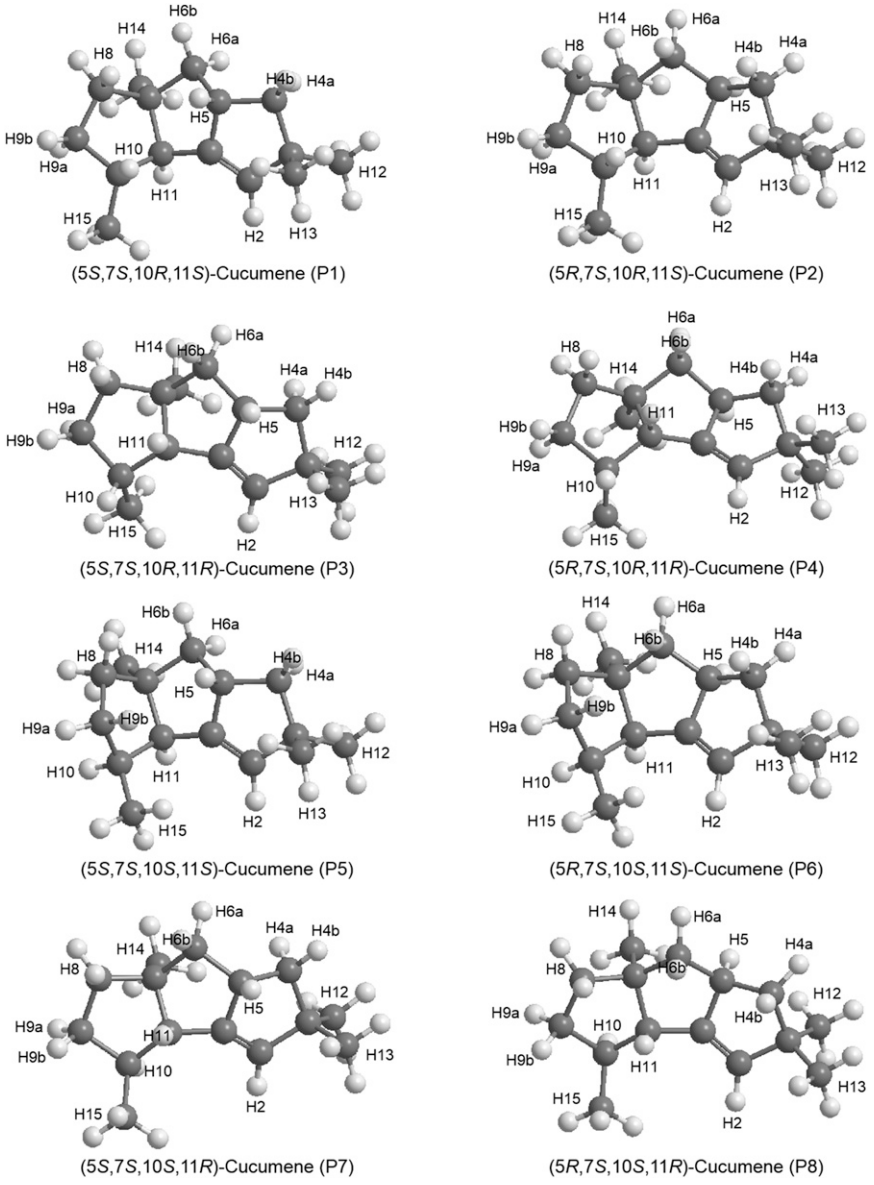


Fig. S3. DFT-optimized geometry of the eight diastereomers of cucumene.

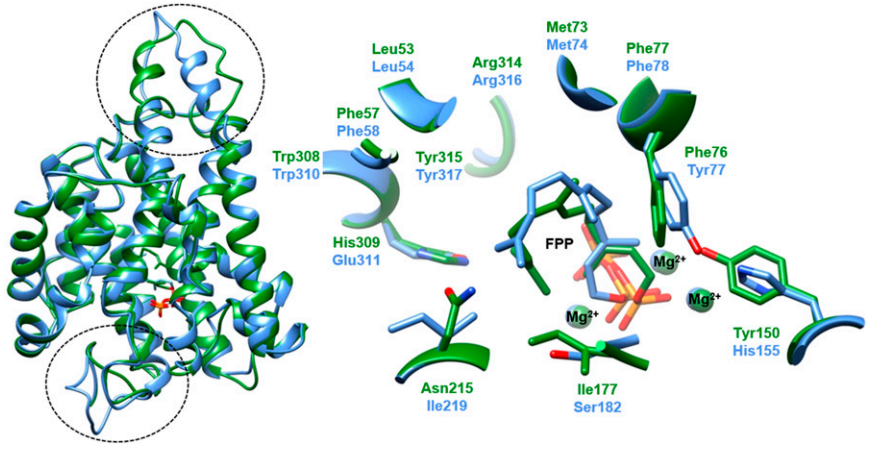


Fig. S4. Structural alignment of the homology model of pentalenene synthase (Q55012, green) and cucumene synthase (B5GLM7, blue).

Table S1. Sequence identity and similarity of selected terpene synthases

	Identity (similarity), %						
Identity (similarity), %	B5GLM7	Q55012	Q82IY4	Q82RR7	B5H117	B5GMG2	Q9K499
B5GLM7	100 (100)	33 (50)	35 (51)	37 (49)	34 (46)	35 (50)	27 (38)
Q55012		100 (100)	76 (86)	34 (49)	36 (51)	34 (53)	26 (39)
Q82RR7			100 (100)	36 (51)	36 (51)	36 (54)	27 (39)
Q82RR7				100 (100)	35 (47)	33 (48)	29 (41)
B5H117					100 (100)	34 (47)	23 (35)
B5GMG2						100 (100)	26 (39)
Q9K499							100 (100)

EFI target 509499 (B5GLM7). Pentalenene synthase from *Streptomyces* exfoliatus (Q55012). Pentalenene synthase from *Streptomyces* avermitilis (Q82IY4). Avermitilol synthase (Q82RR7). Linalool/Nerolidol synthase (B5H117). 1,8-Cineole synthase (B5GMG2). Epi-isozizaene synthase (Q9K499).

Table S2. Statistics of the product skeleton prediction and docking scores of terpene synthases

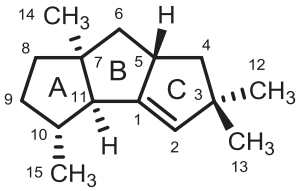
	Statistic of product skeleton prediction		Docking scores		
TPS	True skeleton rank ($R_{[true\text{-}skeleton]}/n_{[total\text{-}skeleton]}$)	True skeleton population ($n_{[true\text{-}total]}/n_{[total\text{-}carbocation]}$)	C_{10}	C_{15}	C_{20}
3LG5	7/224	25/1,927	−14.8	−15.1	NP
Q55012	4/138	110/2,286	−15.0	−15.6	−11.6
B5GLM7	1/163	136/2,342	−12.9	−14.5	−12.5

NP, not predicted.

Table S3. ^{1}H and ^{13}C chemical shift assignments for (5*S*,7*S*,10*R*,11*S*)-cucumene

Position	δ_H,* multiplicity, *J*	δ_C*	δ_H,† multiplicity, *J*	δ_C†	δ_H,‡ multiplicity, *J*	δ_C‡
1	—	154.1	—	154.3	—	154.0
2	5.00, d, 1.9	127.5	5.12, d, 2.0	128.0	5.03, s	127.6
3	—	50.8	—	51.1	—	50.8
4a	1.22, m	47.7	1.38, dd, 12.0, 8.3	48.2	1.26, m	47.8
4b	1.81, dd, 12.0, 7.2		1.91, dd, 12.1, 7.1		1.84, dd, 12.1, 7.1	
5	3.14, p, 8.3	48.0	3.18, p, 8.4	48.4	3.15, p, 8.5	48.0
6a	0.82, t, 11.5	47.8	0.94, t, 11.5	48.3	0.85, t, 11.6	47.9
6b	1.75, dd, 11.9, 7.4		1.8, dd, 12.0, 7.4		1.77, dd, 11.9, 7.5	
7	—	55.3	—	55.6	—	55.3
8	1.50, m	41.3	1.50, m	41.6	1.50, m	41.3
9a	1.19, m	35.3	1.19, m	35.6	1.20, m	35.3
9b	1.66, m		1.64, m		1.66, m	
10	1.53, m	43.8	1.57, m	44.1	1.54, m	43.8
11	1.67, m	57.7	1.77, d, 8.0	58.2	1.70, d, 8.2	57.8
12	1.07, s	30.2	1.17, s	30.4	1.10, s	30.2
13	1.00, s	28.2	1.14, s	28.5	1.04, s	28.3
14	1.06, s	29.8	1.08, s	30.0	1.07, s	29.8
15	1.02, d, 6.7	20.1	1.08, d, 6.6	20.3	1.03, d, 7.5	20.1

Structure

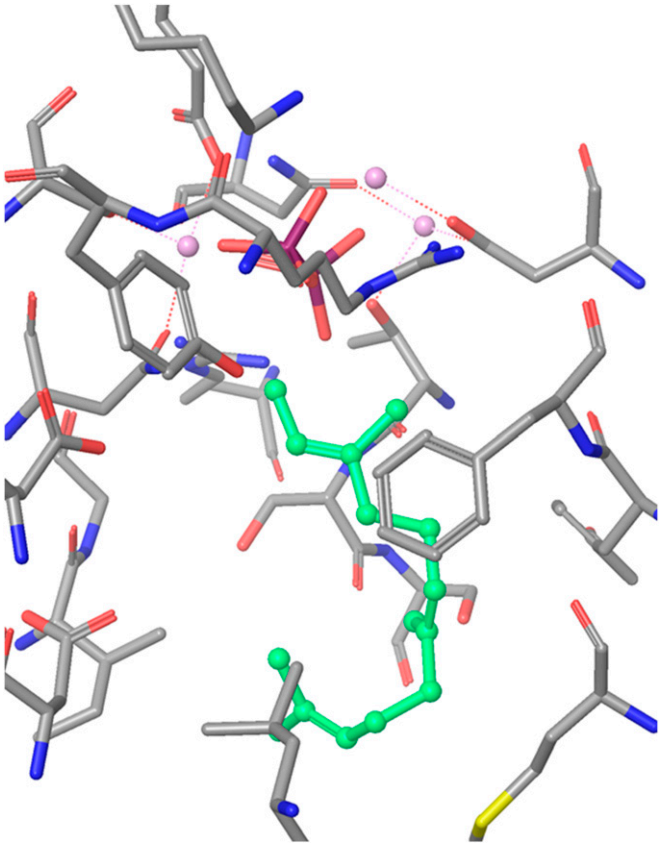


*$CDCl_3$.
†C_6D_6.
‡$CDCl_3 + C_6D_6$ (3:1).

Table S4. List of observed NOE signals and their respective distances in the DFT-optimized geometry of the diastereomers of cucumene

Observed NOE signals	Distance (Å)							
	P1	P2	P3	P4	P5	P6	P7	P8
H2 → H11	3.0	3.5	3.9	3.4	3.0	3.5	3.9	3.3
H2 → H12	2.8	2.9	2.8	3.0	2.8	2.9	2.8	3.0
H2 → H13	2.9	2.8	3.0	2.8	3.0	2.8	3.0	2.8
H4a → H4b	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8
H4a → H6a	2.4	3.5	2.6	3.5	2.4	3.5	2.6	3.5
H4a → H12	2.4	2.6	2.4	2.6	2.4	2.6	2.4	2.6
H4b → H5	2.4	3.0	2.4	3.0	2.4	3.0	2.4	3.0
H4b → H12	3.0	3.7	3.0	3.7	3.0	3.7	3.0	3.7
H4b → H13	2.6	2.4	2.6	2.4	2.6	2.4	2.6	2.4
H5 → H4a	3.0	2.4	3.0	2.4	3.0	2.4	3.0	2.4
H5 → H6a	3.0	2.5	2.9	2.4	3.1	2.5	2.9	2.4
H5 → H6b	2.5	3.1	2.3	3.1	2.5	3.0	2.3	3.1
H5 → H8	2.6	4.6	4.2	4.8	2.8	4.7	4.2	4.8
H5 → H10	2.7	4.6	4.7	5.1	4.0	5.0	4.9	4.2
H5 → H13	2.8	4.8	2.6	4.7	2.8	4.8	2.6	4.7
H6a → H6b	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8
H6a → H14	2.5	2.5	2.4	2.4	2.4	2.5	2.5	2.5
H6b → H8	2.5	2.2	2.5	2.6	2.3	2.4	2.5	2.6
H6b → H14	3.0	3.6	3.6	3.7	2.9	3.6	3.6	3.7
H8 → H9a	2.5	2.5	2.4	2.4	2.4	2.5	2.4	2.4
H8 → H9b	2.5	2.4	2.4	2.4	2.4	2.5	2.4	2.4
H8 → H14	2.7	2.5	2.8	2.6	2.4	2.5	2.6	2.6
H9a → H9b	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8
H9a → H10	3.1	3.1	2.9	2.9	2.5	2.4	2.4	2.4
H9a → H14	2.5	2.7	2.3	2.3	4.2	4.6	2.6	2.5
H9a → H15	2.6	2.6	2.6	2.6	2.8	2.8	3.1	3.1
H9b → H10	2.4	2.5	2.3	2.3	3.1	3.0	3.0	3.0
H9b → H15	2.9	2.8	3.3	3.3	2.6	2.6	2.5	2.5
H10 → H11	3.0	3.0	2.2	2.2	2.4	2.3	3.0	3.0
H10 → H15	2.5	2.5	2.4	2.5	2.5	2.4	2.5	2.5
H11 → H14	2.5	2.5	3.8	3.8	2.6	2.5	3.8	3.8
H11 → H15	2.7	2.5	3.7	3.7	2.9	3.0	2.5	2.5
H12 → H13	2.6	2.6	2.6	2.6	2.6	2.5	2.6	2.6

Each cell is colored based on the distance between the protons. Red, ≤2.5 Å; orange, ≤3.0 Å; green, ≤3.5 Å; blue, ≤4.0 Å; purple, >4.0 Å.



Movie S1. Enumeration of carbocationic intermediates in the active site of B5GLM7.

Movie S1

Other Supporting Information Files

Dataset S1 (PDF)
Dataset S2 (PDF)
Dataset S3 (PDF)
Dataset S4 (PDF)