

Supporting Information

Engineering the ‘Missing Link’ in Biosynthetic (-)-Menthol Production: Bacterial Isopulegone Isomerase

Andrew Currin,[†] Mark S. Dunstan,[†] Linus O. Johannissen,[†] Katherine A. Hollywood,[†]
Maria Vinaixa,[†] Adrian J. Jervis,[†] Neil Swainston,[†] Nicholas J. W. Rattray,[†] John M.
Gardiner,[†] Douglas B. Kell,[†] Eriko Takano,^{†‡} Helen S. Toogood,[‡] and Nigel S.
Scrutton^{†‡*}

[†]Manchester Centre for Fine and Speciality Chemicals (SYNBIOCHEM) and [‡]School of
Chemistry, Manchester Institute of Biotechnology, University of Manchester, Manchester M1 7DN,
United Kingdom

* Corresponding Author: Professor Nigel S Scrutton: nigel.scrutton@manchester.ac.uk

Table of Contents

METHODS

- S1. General Materials and Reagents
- S2. Gene Synthesis, Mutagenesis and Cloning
 - S2.1 Ketosteroid isomerase protein sequence
 - S2.2 GeneGenie design for KSI synthesis
 - S2.3 Mutagenesis rounds
 - Table S1. KSI variants generated in the three rounds of mutagenesis
 - S2.4 Ambiguous codons used for KSI mutagenesis
 - Table S2. Ambiguous codon design for KSI mutagenesis
 - S2.5 SpeedyGenes Protocol
- S3. Biotransformations and Analytical Techniques
 - S3.1 Extraction and purification of enzymes
 - S3.2 GC/GC-MS analysis of *Mentha* monoterpenoids
 - S3.3 Steady state reactions
 - Figure S1. Pulegone (**2**) standard curve.
- S4. Construction of the (+)-*cis*-Isopulegone (**3**) to (-)-Menthol (**1**) Enzyme Cascade Construct
 - Table S3. Design of the ribosomal binding site (RBS) sequences.
- S5. KSI Crystallisation methods
 - Table S4. KSI data collection and refinement statistics.
- S6. KSI Molecular Simulation methods
 - Figure S2. Overlay of energy minimised structures.
- S7. Chiral Analysis of KSI Biotransformations
 - Table S5. Product enantiomeric identity of wild-type and variant KSI with **3** by chiral GC.
- S8. Robotic Directed Evolution Strategy
 - Figure S3. The robotics-driven directed evolution workflow.
- S9. Crystal Structures and Molecular Simulation Discussion
 - Figure S4. Superimposition of the crystal structures of wild type and four variant KSI enzymes.
 - Figure S5. Distances restrained during MD simulations of wild type KSI.
- S10. Cascade Biocatalysis Products Identification
 - Figure S6. Isopulegol isomer identification by GC-MS.
 - Figure S7. Separation of **1** and precursors by GC-MS.

S1. General Materials and Reagents

All chemicals, reagents, media, organic solvents and other materials used in this study were obtained from Sigma-Aldrich Co Limited, Fisher Scientific UK Limited, Agilent Technologies UK Limited, Roche Applied Sciences, and VWR International Limited, unless otherwise stated. The expression vector pET21b was obtained from Merck (Novagen). Competent *Escherichia coli* cells NEB5 α and T7 Express were obtained from New England Biolabs. Gene sequencing and oligonucleotide synthesis were performed by GATC Biotech and IDT DNA, respectively.

S2 Gene Synthesis, Mutagenesis and Cloning

S2.1 Ketosteroid isomerase protein sequence

```
>KSI_Pp
MNLPTAQEVQGLMARYIELVDVGDIEAIVQMYADDATVEDPFGQPPIHGREQIAAFYRQGLGGGKVRACLTG
PVRASHNGCGAMPFRVEMVWNGQPCALDVIDVMRFDEHGRIQTMQAYWSEVNLSVREPQLLEHHHHHH
```

S2.2 GeneGenie design for KSI synthesis

URL: <http://g.gene-genie.appspot.com/?jobId=97F4103C-ACEA-4EBD-8FD5-F06F73F41EB1>

Oligonucleotides for KSI gene synthesis:

```
>KSI11F; Forward; 75 bp
AAGGAGATATACATATGAATCTGCCGACCGCCAGGAAGTGCAGGGTCTGATGGCGCGTTATATTGAACTGGTGG
>KSI12R; Reverse; 75 bp
TCCACGGTCGCATCATCCGCATACATCTGCACAATCGCTTCAATATCGCCACATCCACCAGTTCAATATAACGC
>KSI13F; Forward; 75 bp
TGATGCGACCGTGGAAGATCCGTTTGGCCAGCCTCCGATTCACGGCCGGAACAGATCGCGGCGTTTTACCGCCA
>KSI14R; Reverse; 74 bp
TATGGCTGGCACGCACCGGACCGGTGAGACACGCGCGCACTTTACCGCCGCCAGGCCCTGGCGGTAAAACGCC
>KSI15F; Forward; 76 bp
TGCGTGCCAGCCATAATGGCTGCGGCGGATGCCGTTTCGCGTGGAATGGTGTGGAATGGTCAACCATGTGCGCT
>KSI16R; Reverse; 75 bp
ACGCTTGCCATGGTTTGAATGCGGCCATGTTTCATCAAAGCGCATCACATCAATCACATCCAGCGCACATGGTTGAC
>KSI17F; Forward; 73 bp
ATTCAAACCATGCAAGCGTATTGGAGCGAAGTGAATCTGAGCGTGCGGAACCGCAGCTCGAGCACCACCACC
>KSI18R; Reverse; 15 bp
GGTGGTGGTGCTCGA
```

S2.3 Mutagenesis rounds

Table S1. KSI variants generated in the three rounds of mutagenesis.

Round	Mutation type	Amino acids mutated	Comb. ^a	Total variants
1	Combinatorial ambiguous codons	[V88DTK L99DTK]	36	72
		[Y16WVC, Y57WVC]	36	
	Single ambiguous codons	[V20WKS]	8	86
		[F56WKS]	8	
		[G60RSC]	4	
		[L61WKS]	8	
		[V66WKS]	8	
		[F86WKS]	8	
		[M90WKS]	8	
		[V101WKS]	8	
		[D103WVC]	6	
		[M116WKS]	8	
		[A118SYG]	4	
		[W120WKS]	8	
2	Point mutation	[L99I D103S] [L99V D103S] [V88I L99V D103S]	-	3
3	Point mutation	[V88I L99V V101A D103S]	-	1

^aGenetic combinations tested.

S2.4 Ambiguous codons used for KSI mutagenesis

Table S2. Ambiguous codon design for KSI mutagenesis.

Amino Acid	Codon ^a	Mixed bases	Comb. ^b	Encoded res. ^c	Type
V88, L99	DTK	(G,A,T) T (G,T)	6	F, L, I, M, V	Apolar
Y16, Y57, D103	WVC	(A,T) (A,C,G) C	6	S, Y, C, T, N	Polar/H-bond
V20, F56, L61, V66, F86, M90, V101, M116, W120	WKS	(T,A) (T,G) (G,C)	8	F, L, I, M, C, W, S, R	Larger side chains, apolar and polar
G60	RSC	(A,G) (G,C) C	4	G, A, T, S	Small side chains
A118	SYG	(G,C) (T,C) G	4	A, V, L, P	Small side chains, apolar

^aAmbiguous codon used; ^bgenetic combinations tested; ^camino acids encoded.

S2.5 SpeedyGenes Protocol

Oligonucleotides one and eight were utilised as forward and reverse primers, respectively (600 nM each), while the remaining oligonucleotides were mixed together to form the template (30 nM). The reaction (50 μ L) also contained 0.2mM dNTPs, Q5 reaction buffer and 0.02 U/ μ L Q5 hot-start polymerase (New England Biolabs). The reaction constituted an initial denaturation at 98 °C for 90 s, followed by 35 cycles of 98 °C for 20 s, 60 °C for 20 s and 72 °C for 30 s. Full-length genes were then purified by electrophoresis and gel extraction kit (Macherey-Nagel).

S3 Biotransformations and Analytical Techniques

S3.1 Extraction and purification of enzymes

Cultures in deep well plates were pelleted by centrifugation (3000 x g), resuspended in lysis buffer (50mM Tris pH 8.0 containing 50% Bugbuster, 0.1 mg/mL lysozyme, protease inhibitors cocktail and 0.1 mg/mL DNase) and agitated at 30 °C 20 min at 1000 rpm. Insoluble material was pelleted by centrifugation and the soluble protein was bound to Ni-NTA resin suspension (Qiagen) (50 µL per variant). The resin was washed, and KSI was eluted with buffer (50 mM Tris pH 7.5 + 250 mM NaCl) containing 10 mM and 250 mM imidazole, respectively. Purified enzyme eluates were buffer exchanged into 50 mM Tris-HCl pH 7.0 and concentrated using Vivaspin 500 (Sartorius) prior to enzyme assays.

S3.2 GC/GC-MS analysis of *Mentha* monoterpenoids

GC-MS achiral quantitative analysis was conducted on a 7890B GC coupled to a 5975 series MSD quadrupole mass spectrometer and equipped with a 7693 autosampler (Agilent, Technologies, UK). Samples (1 µL) were injected onto a DB-WAX column (30 m x 0.320 mm x 0.25 µm; Agilent Technologies) with an inlet temperature of 240 °C and a split ratio of 20:1. Helium was used as the carrier gas with a flow rate of 1.5 mL/min and a pressure of 1.5603 psi. The chromatography was programmed to begin at 40 °C with a hold time of 1 minute, followed by an increase to 150 °C at a rate of 10 °C/min, then a subsequent increase to 210 °C at a rate of 80°C/min and a final hold time of 1 min. The total run time for the analysis was 13.75 min. The MS was equipped with an electron impact ion source using 70eV ionisation and a fixed emission of 35 µA. The mass spectrum was collected for the range 35-550 m/z with a scan speed of 3,125 (N=1).

Chiral product analysis was performed by analysing reactions by GC using an Agilent Technologies 7890A GC system with an FID detector and a Chirasil-DEX-CB column (Agilent; 25 m, 0.32 mm, 0.25 µm). In this method the injector temperature was at 180 °C with a split-less injection. The carrier gas was helium with a flow rate of 1 mL/min and a pressure of 5.8 psi. The program began at 70 °C with an increase of temperature to 150 °C at a rate of 20 °C/minute and a hold for 3 min. This was followed by an increase of temperature to 190 °C at a rate of 2 °C/minute and a hold for 3 min.

For all GC-MS analyses, *sec*-butylbenzene was used as an internal standard to allow for accurate quantification. *sec*-Butylbenzene (0.01%) was added to all samples and the quantification of members of the menthol pathway was calculated relative to the peak area of this internal standard. Calibration curves were constructed for accurate quantification and calibration standards were analysed in a random sequence in the same analytical run as the standards. Vendor binary files were converted to open mzXML data format¹ using ProteoWizard msConvert.² Automated peak profiling and quantification was conducted using in-house scripts written in R.

Analysis of the isopulegol isomers was conducted on a 7890B GC coupled to a 5975 series MSD quadrupole mass spectrometer and equipped with a 7693 autosampler (Agilent, Technologies, UK). The sample (2 µL) was injected onto a DB-5 column (20 m x 0.1 mm x 0.1 µm; Agilent Technologies) with an inlet temperature of 250 °C and a split ratio of 2:1. Helium was used as the carrier gas with a flow rate of 0.1 mL/min and a pressure of 22.303 psi. The chromatography was programmed to begin at 60 °C with a hold time of 2 minutes, followed by an increase to 130 °C at a rate of 40 °C/min, a subsequent increase to 150 °C at a rate of 2 °C/min followed by an increase to 200 °C at a rate of 40 °C/min and a final hold time of 2 min. The total run time for the analysis was

17 min. The MS was equipped with an electron impact ion source using 70 eV ionisation and a fixed emission of 34.6 μA . The mass spectrum was collected for the range 50-550 m/z with a scan speed of 3,125 (N=1).

S3.3 Steady state reactions

Product concentrations from steady state reactions were determined by comparison to a pulegone standard curve. Reactions were monitored over a 1 hour time period, and corrected for losses over time. The standard curve shows the absorbance at the start of the incubation.

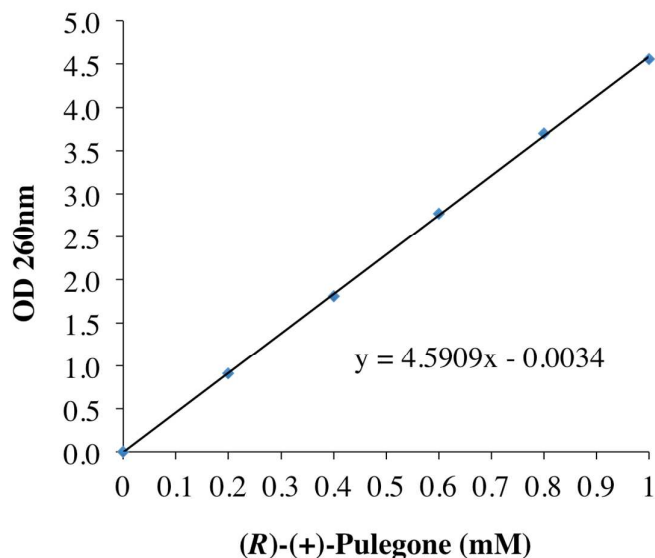


Figure S1. Pulegone standard curve. Standards (100 μL) were composed of 50 mM Tris pH 7.0 containing 0-1 mM (*R*)-pulegone (*R*-2) within UV-Star microplates (Greiner Bio-One) and covered with a ClearVue sealing sheet. The absorbance was determined at 260 nm.

S4. Enzyme Cascade Construct from (+)-*cis*-Isopulegone (3) to Menthol (1)

The three biocatalytic enzymes were combined into a functional operon with one promoter (T7). The ribosomal binding sites were calculated using the Salis RBS calculator (REF).

Table S3. Design of the ribosomal binding site (RBS) sequences for the cascading enzyme construct.



Gene	RBS sequence	Predicted TIR ^a
KSI	TTTGTTTAACTTTAAGAAGGAGA	6689
MpPGR	TTTGTTTAACTTTAAGAAGGAGA	11325
MMR	CCCATTAATAAATTAAGCTTTATATAC	918
	GACTCATCAATAAAGGTTACTTCCT	535
	ATTTTCACATACGTCTACTT	199

^aTIR = translation initiation rate. KSI = ketosteroid isomerase from *P. putida*; MpPGR = pulegone reductase from *Mentha piperita*; MMR = (-)-menthone:(-)-menthol reductase from *M. piperita*; Op = *lac* operator; SD = RBS sequences. RBS sequences were designed with the Salis RBS calculator (<https://salislab.net/software/>).

S5. KSI Crystallisation Methods

Table S4. KSI data collection and refinement statistics of the four variant KSI crystal structures

	KSI Variant											
	D103S			V88I/L99V			V88I/L99V/D103S			V88I/L99V/V101A/D103S		
PDB Code	6F4Y			6F50			6F54			6F53		
Data collection												
Wavelength	1.00			1.00			1.00			1.00		
Space group	P 21 21 21			P 21 21 21			C2			C 2 2 21		
<i>a, b, c</i> (Å)	36.17	73.53	94.82	35.69	73.54	95.41	86.61	71.54	50.63	35.740	94.720	73.810
α, β, γ (°)	90.00	90.00	90.00	90.00	90.00	90.00	90.00	90.37	90.00	90.00	90.00	90.00
Resolution (Å) ^a	47.7 (1.9)			95.4 (2.0)			43.3 (1.08)			29.1 (1.49)		
R_{sym} or R_{merge} ^a	0.74 (0.10)			0.8 (0.91)			0.93 (0.09)			0.43 (2.7)		
$I / \sigma I$ ^a	11.3 (2.2)			11.6 (2.1)			8.9 (1.2)			23.1 (2.7)		
Completeness (%) ^a	99.7 (97.1)			99.9 (97.98)			87.4 (42.9)			99.6 (98.9)		
Redundancy ^a	6.3 (6.4)			6.5 (6.5)			3.3 (2.7)			5.8 (4.7)		
CC half	0.7			0.9			0.5			0.9		
total observations	19954			17276			115237			20641		
Refinement												
Resolution (Å)	1.92			95.4 (2.0)			43.3 (1.08)			29.1 (1.49)		
$R_{\text{work}} / R_{\text{free}}$ ^a	0.204 (0.252)			0.183 (0.227)			0.18 (0.198)			0.183 (0.202)		
No. atoms	2122			2088			2346			1046		
R.m.s. deviations												
Bond lengths (Å)	0.01			0.009			0.006			0.007		
Bond angles (°)	1.19			1.05			1.01			1.08		

^aValues in the parentheses refer to the outer resolution shell.

S6. KSI Molecular Simulation Methods

Molecular dynamics simulations were carried out using Gromacs 4.6.1 with the Gromos 53A6 force field and periodic boundary conditions.³⁻⁴ The enzyme was placed in a solvation box of minimum 10 Å around the protein. The simulation protocol was as follows: after energy minimization the system was thermalized 300 K for 100 ps using NVT dynamics, and the pressure was then equilibrated for 100 ps using NPT dynamics; the protein and substrates were constrained during these steps. All constraints and pressure couplings were then switched off and harmonic constraints of 100 kJ/mol/Å² were applied to keep the (+)-*cis*-isopulegone (**3**) near a reactive conformation, as defined by the distances in the energy minimised DFT model for the variant: 2.0 Å for the distance between the carbonyl oxygen of **3** and the carboxylic/hydroxyl H of D/S103 and 3.5 Å for the distance between substrate C and carboxylate O of D40 (Figure S5). The system was then relaxed at 250 K, 280 K, 290 K and 300 K for 1 ns each, before the 50 ns production MD runs at 300 K. Representative structures for illustrative purposes were selected as those with the lowest RMSD for the substrate and residues 40 and 103 relative to the average, following structural alignment to the protein backbone.

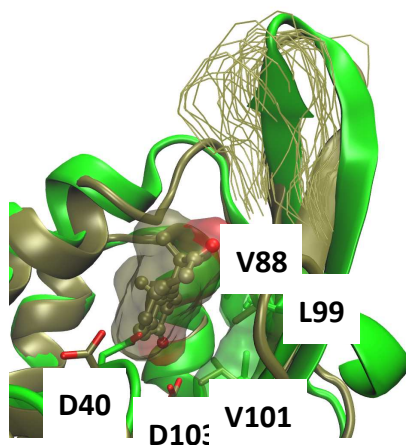
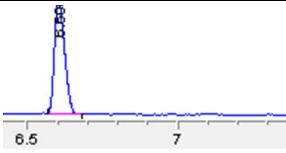
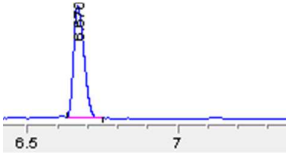
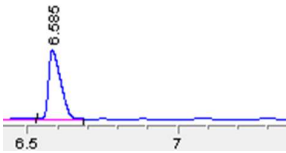
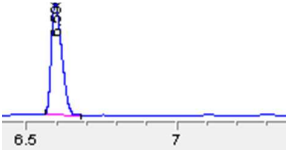


Figure S2. Overlay of energy minimised structure (green cartoon) and representative structure from unrestrained molecular dynamics (MD) simulations (brown cartoon) for WT KSI with bound androstenedione (**6**; brown sticks). The solvent accessible surface areas for the substrate and residues 88, 99 and 101 from the representative and minimised structures, respectively are shown as transparent surfaces. The conformation of the β -hairpin over the course of the simulation is shown using a backbone trace of residues 90-97 every 2 ns after thermalization. The MD simulations were performed using Gaussian09 revision D.01.⁵

S7. Chiral Analysis of KSI Biotransformations

Table S5. Product enantiomeric identity of wild-type and variant KSI with **3** by chiral GC.

Enzyme/Standard	GC Trace	Product
(<i>R</i>)-2 standard		N/A
(<i>S</i>)-2 standard		N/A
Wild type KSI		(<i>R</i>)-2
KSI variant V88I/L99V/V101A/D103S		(<i>R</i>)-2

Duplicate reactions (200 μ L) were performed in 50 mM Tris pH 7.0 containing 1 mM **3** and 10 μ M KSI. After a 24 h incubation at 30 $^{\circ}$ C (180 rpm), reactions were extracted with 180 μ L ethyl acetate containing 0.01% *sec*-butylbenzene and dried with anhydrous $MgSO_4$. Chiral product analysis was performed by analysing reactions by GC using an Agilent Technologies 7890A GC system with an FID detector and a Chirasil-DEX-CB column (Agilent; 25 m, 0.32 mm, 0.25 μ m).

S8. Robotic Directed Evolution Strategy

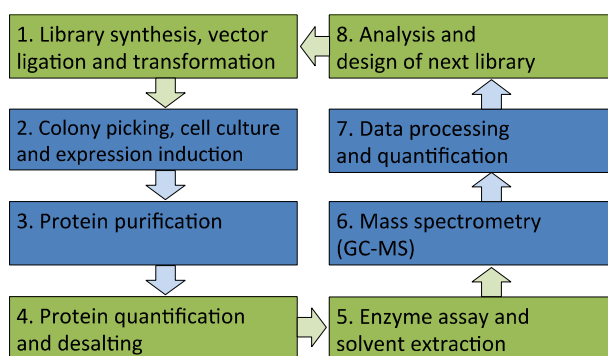


Figure S3. The robotics-driven directed evolution workflow. The process included manual (green) and automated (blue) procedures. Following library synthesis, vector ligation and transformation into *E. coli* (1), colonies expressing single KSI variants were picked, grown and the expression induced (2). His-tagged KSI proteins were purified by IMAC (3), followed by quantification and desalting (4). Reactions were processed for GC-MS analysis by solvent extraction (5 & 6), followed by automated data processing (7) and analysis for the design of the next library (8). This data, together with structural information, was utilized to design subsequent variants for screening.

S9. Crystal Structures and Molecular Simulations

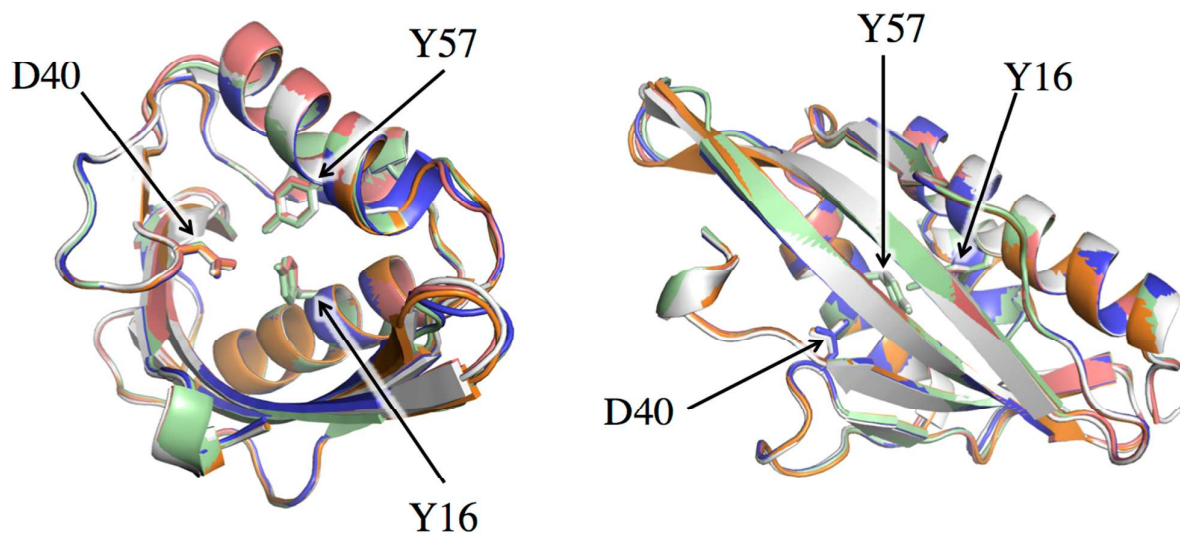


Figure S4. Superimposition of the crystal structures of wild type and four variant KSI enzymes. The structures are shown as grey, green, red, orange and blue cartoons for wild-type (PDB 1OPY)⁶ and variants D103S, V88I/L99V, V88I/L99V/D103S and V88I/L99V/V101A/D103S, respectively. Variant residues are shown as sticks in the same respective colours.

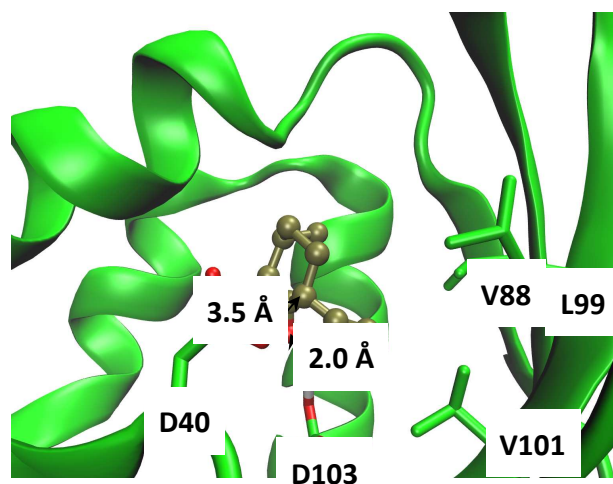


Figure S5. Distances restrained during MD simulations of wild type KSI. The energy-minimised structure is shown as a green cartoon, with key residues as atom coloured sticks with green carbons. The bound substrate 3 is shown as brown balls and sticks.

S10. Cascade Biocatalysis Products Identification

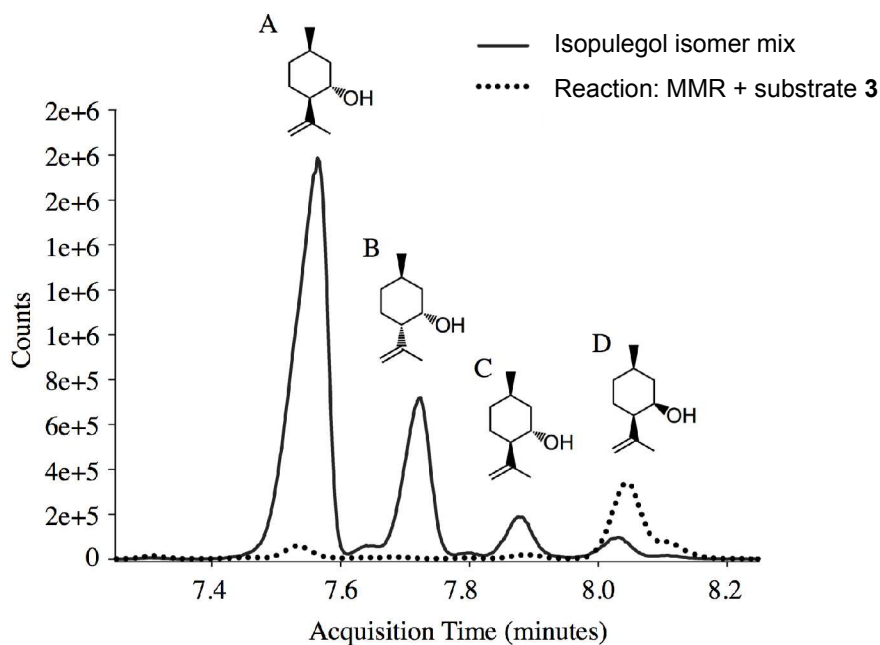


Figure S6. Isopulegol isomer identification by GC-MS. Analysis of isopulegol isomers was performed by GC-MS using a DB-5 column. Compounds: A = (-)-isopulegol; B = (+)-neoisopulegol; C = (+)-iso-isopulegol and D = (+)-neoisopulegol (**10**).

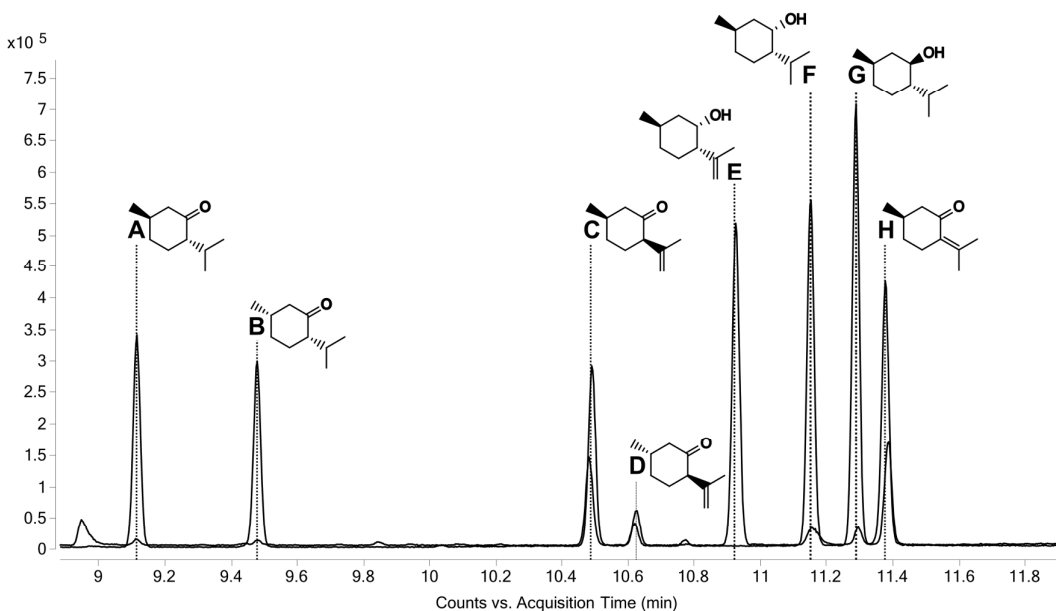


Figure S7. Separation of (-)-menthol and precursors by GC-MS. Analysis of monoterpenoids was performed by GC-MS using a DB-WAX column. Compounds: A = (-)-menthone (**8**); B = (+)-isomenthone (**9**); C = (+)-*cis*-isopulegone (**3**); D = (+)-*trans*-isopulegone (**7**); E = isopulegol; F = (+)-neoisomenthol (**5**); G = (-)-menthol (**1**) and H = (*R*)-pulegone ((*R*)-**2**).

References

- (1) Pedrioli, P. G. A.; Eng, J. K.; Hubley, R.; Vogelzang, M.; Deutsch, E. W.; Raught, B.; Pratt, B.; Nilsson, E.; Angeletti, R. H.; Apweiler, R.; Cheung, K.; Costello, C. E.; Hermjakob, H.; Huang, S.; Julian, R. K.; Kapp, E.; McComb, M. E.; Oliver, S. G.; Omenn, G.; Paton, N. W.; Simpson, R.; Smith, R.; Taylor, C. F.; Zhu, W.; Aebersold, R., *Nat. Biotechnol.* **2004**, *22*, 1459–1466.
- (2) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M.-Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P., *Nat. Biotechnol.* **2012**, *30*, 918–920.
- (3) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F., *J. Comput. Chem.* **2004**, *25*, 1656-1676.
- (4) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C., *J. Comput. Chem.* **2005**, *26*, 1701-1718.
- (5) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; J. A. Montgomery, J.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Lyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 09 D.01*, Gaussian, Inc.: Wallingford, CT, 2016.
- (6) Kim, S. W.; Cha, S.-S.; Cho, H.-S.; Kim, J.-S.; Ha, N.-C.; Cho, M.-J.; Joo, S.; Kim, K. K.; Choi, K. Y.; Oh, B.-H., *Biochemistry* **1997**, *36*, 14030-14036.