Supporting Information

Biocatalytic Routes to Lactone Monomers for Polymer Production

Hanan L. Messiha,[†‡] Syed T. Ahmed,[‡] Vijaykumar Karuppiah,[†‡] Reynier Suardíaz,[§] Gabriel A. Ascue Avalos,[†‡] Natalie Fey,[§] Stephen Yeates,[‡] Helen S. Toogood,[†‡] Adrian J. Mulholland,[§] and Nigel S. Scrutton*[†‡]

[†]BBSRC/EPSRC Manchester Synthetic Biology Research Centre for Fine and Specialty Chemicals (SYNBIOCHEM), and [‡]School of Chemistry, Faculty of Science and Engineering, University of Manchester, 131 Princess Street, Manchester M1 7DN, U.K.

[§]Centre for Computational Chemistry, School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, U.K.

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

General Information and Methods

Chemicals and Reagents

All chemicals, reagents, media, organic solvents, resins for chromatography 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. The *Escherichia coli* expression cells BL21 (DE3) and C41 (DE3) strains were obtained from Promega UK and Cambridge Bioscience Limited; respectively. Gene sequencing and oligonucleotide synthesis were performed by Eurofins MWG (Ebersberg, Germany). The expression vectors pET21b and pET15b were from Novagen. The substrates (-)-menthone, (+)-dihydrocarvone and (+)-pulegone used in this study were commercially available, and were dissolved in absolute ethanol (0.25 M or 0.5 M). The normal lactones of (-)-menthone, (+)-dihydrocarvone of (+)-pulegone were synthesised by NewChem Technologies Limited to be used as authentic standards for qualitative and quantitative analyses only. The anhydrous organic solvents were of extra dry grade and were stored and opened under strict anaerobic conditions in a glove box. $Mg(BHT)_2(THF)_2$ was synthesised as previously reported,[1, 2] kindly donated by Dr John Morrison, School of Chemistry, University of Manchester and stored in a glove box until used.

Oligonucleotide primer sequences

Site-directed mutagenesis of wild-type (WT) $CHMO_{Phi1}$ was performed using the QuikChange lightning site-directed mutagenesis kit (Agilent) according to the manufacturer's protocols. The sequence of the primer pairs for mutation introduction by PCR can be found in Table S1.

Table S1. Primer sequences for the construction of the $CHMO_{Phi1}$ F249A/F280A/F435A triple mutant (variant $CHMO_{Phi1}$).

Mutation	Direction	Sequence
F249A	Forward	5'-GTGAAAAAAAGCGCAGTTGCCGCGGGTTTTGAAGAAAGCACCCTG-3'
	Reverse	5'-CAGGGTGCTTTCTTCAAAACCCGCGGCAACTGCGCTTTTTTTCAC-3'
F280A	Forward	5'-GCATGGGATCATGGTGGTGGCGCGCGTTTTATGTTTGGCACC-3'
	Reverse	5'-GGTGCCAAACATAAAACGCGCGCCACCACCATGATCCCATGC -3'
F435A	Forward	5'-GGGTCCGAATGGTCCGGCGACCAATCTGCCTCCGAG-3'
	Reverse	5'-CTCGGAGGCAGATTGGTCGCCGGACCATTCGGACCC-3'

The mutated bases are underlined

Protein production and analysis

Cultures of *E. coli* containing BVMOs/CRE-BVMOs were grown in 2xYT medium (500 mL; tryptone 16 g/L, yeast extract 10 g/L and 5 g/L NaCl pH 7.0), containing ampicillin (100 mg/mL) and a 1% inoculum of an overnight pre-culture in the same medium. Cultures were incubated at 37°C until OD_{600nm} reached 0.6, followed by an 18 h induction with isopropyl-β-D-1-thiogalactopyranoside (IPTG; 10 µM for $CHMO_{Phi1,}$ and CRE-$CHMO_{Phi1}$; 500 µM for CPDMO; 100 µM for CRE-CPDMO) at 24 °C. Cells were harvested by centrifugation at 5000*g* for 10 min at 4°C. Cell-free extracts (CFE) for biotransformations were generated by resuspending the cell pellet in PBS buffer (10 mM pH 7.4) containing phenylmethylsufonylfluoride and lysing by ultrasonication (4 °C). The extracts were clarified by centrifugation at 26600*g*, diluted to a protein

concentration of 15 mg/mL, stored at -20 °C in 1 mL aliquots and either used directly or stored at -20 °C.

Protein expressions were assessed by SDS-PAGE, using 12% Mini-PROTEAN TGX Stain-Free gels and the Precision Plus protein unstained markers (BioRad; Figure S1). Protein concentrations were measured by Bradford method using the Bio-Rad protein assay kit according to manufacturer's instructions. Table S2 shows a representative example of purified protein yields, which varied from batch to batch.

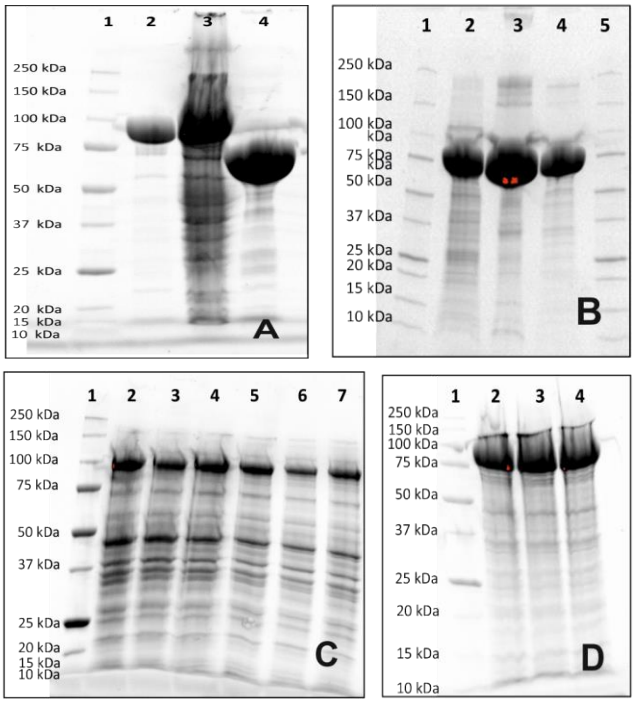


Figure S1. SDS-PAGE analysis of the BVMOs showing the protein expression levels in cell pellets/lysates. A) CPDMO from *Pseudomonas* sp. HI-70, lane 1: protein marker, lane 2: CFE of CRE-CPDMO, lane 3: cell pellet of CRE-CPDMO, lane 4: cell lysate of CPDMO, B) The WT CHMO from *Rhodococcus* sp. $_{Phi1}$ ($CHMO_{Phi1}$) and its F299A/F330A/F485A variant of, lane 1: protein marker, lanes 2, 3 CFE of WT $CHMO_{Phi1}$, lane 4: CFE of the variant of $CHMO_{Phi1}$. C) CRE-WT $CHMO_{Phi1}$, lane 1: protein marker, lanes 2-7 CFEs from various batches. D) CRE-$CHMO_{Phi1}$ variant, lane1: protein marker, lanes 2-4 CFEs from various batches.

Table S2. Total protein concentrations of the produced BVMOs in cell pellets.

Enzyme	Protein Yield (mg/g cell pellet)
CPDMO	60.2
CRE-CPDMO	70.0
Wild type (WT) $CHMO_{Phi1}$	59.2
WT CRE-$CHMO_{Phi1}$	86.0
$CHMO_{Phi1}$ variant	54.6
CRE-$CHMO_{Phi1}$ variant	69.0

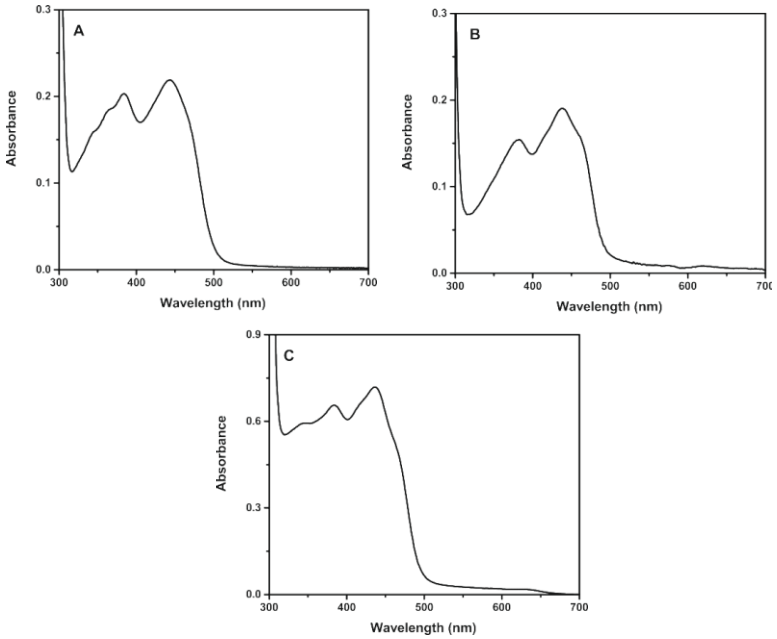


Figure S2. UV-Visible spectra of the purified BVMOs. A) CPDMO from *Pseudomonas* sp. HI-70 (20 μM), B) WT $CHMO_{Phi1}$ (16 μM) and C) The variant of $CHMO_{Phi1}$ (63 μM). Spectra were obtained in 50 mM Tris buffer, pH 7.0 at 25 °C.

Gas Chromatography/Mass Spectroscopy (GC/MS) Analysis

Biotransformation extracts were analysed on an Agilent Technologies 7890B GC system with a 5977A MSD extractor EI source detector using a DB-WAX column (30 m x 0.32 mm x 0.25 μM film thickness; Agilent Technologies). In this method the injector temperature was set at 240°C with a split ratio of 20:1 (1 μL injection). The carrier gas was helium with a flow rate of 2 mL/min and a pressure of 4.6 psi. The program began at 50 °C (1 min hold), ramped to 230 °C at 8°C/min, with a hold at 230 °C (1 min). The ion source temperature of the mass spectrometer (MS) was set to 230°C and spectra were recorded from *m*/*z* 50 to *m*/*z* 250. The mass spectra fragmentation patterns were entered into the NIST/EPA/NIH 11 mass spectral library for identification of a potential match. Compound identification was carried out using authentic standards and also by comparison to reference spectra in the NIST library of MS spectra and fragmentation patterns. Quantitative analysis using authentic standards was performed using experimentally determined relative response factors in relation to the internal standard used (0.1% *sec*-butyl-benzene).

NMR Spectroscopy

NMR analyses were performed on a Bruker 400 MHz NMR spectrometer. ^{1}H NMR and ^{13}C NMR spectra were recorded in deuterated chloroform ($CDCl_3$) and the chemical shifts were referenced to internal tetramethylsilane (TMS).

GC/MS Chromatograms and Data Analysis

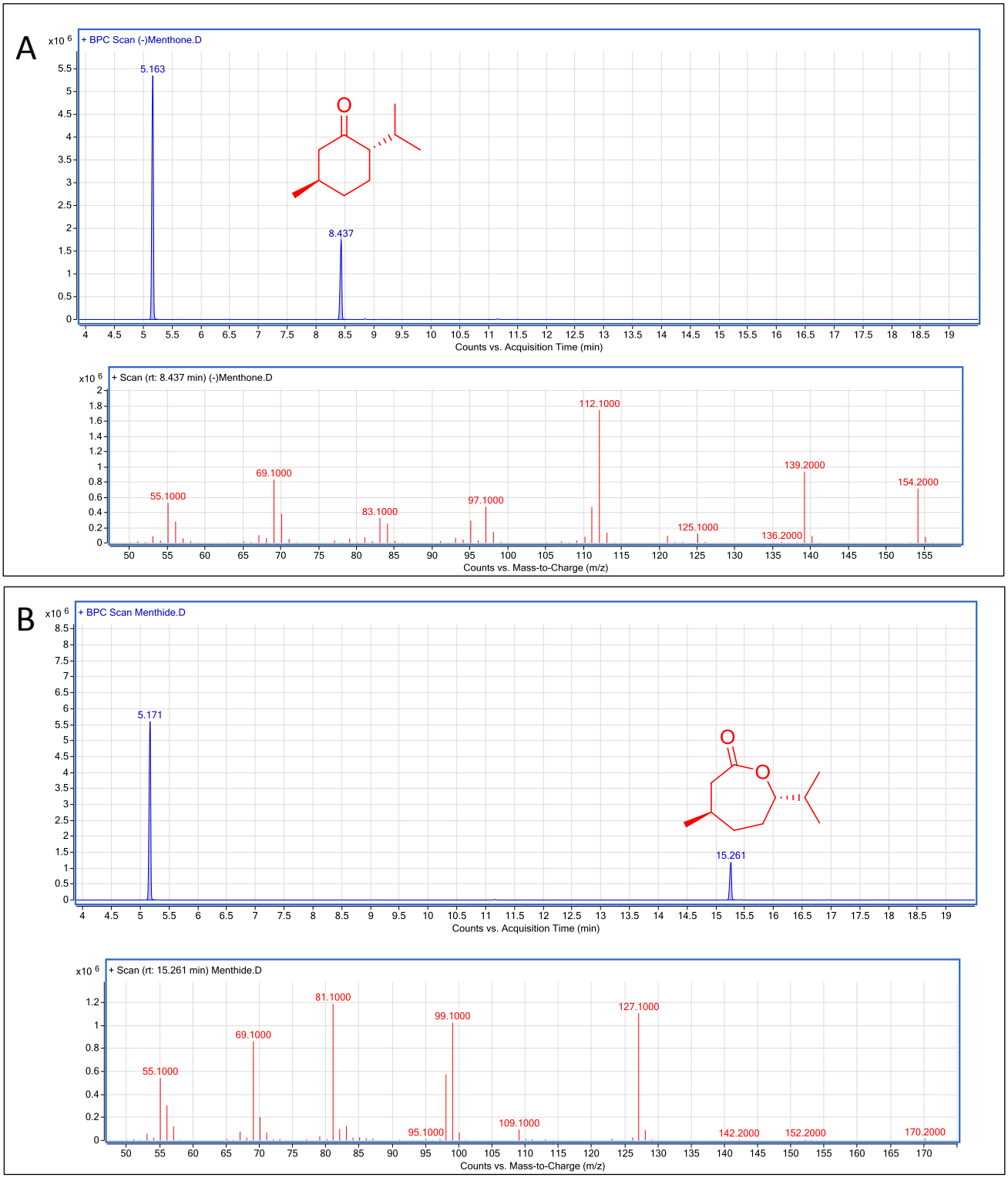


Figure S3. GC/MS chromatograms of A) (-)-menthone and B) menthide compound standards. Each compound (5 mM) was dissolved in ethyl acetate containing 0.1% *sec*-butyl-benzene. The retention times (5.1 min for *sec*-butyl-benzene here and throughout) and mass fragmentation pattern for each compound are shown in the upper and lower panels of each part, respectively.

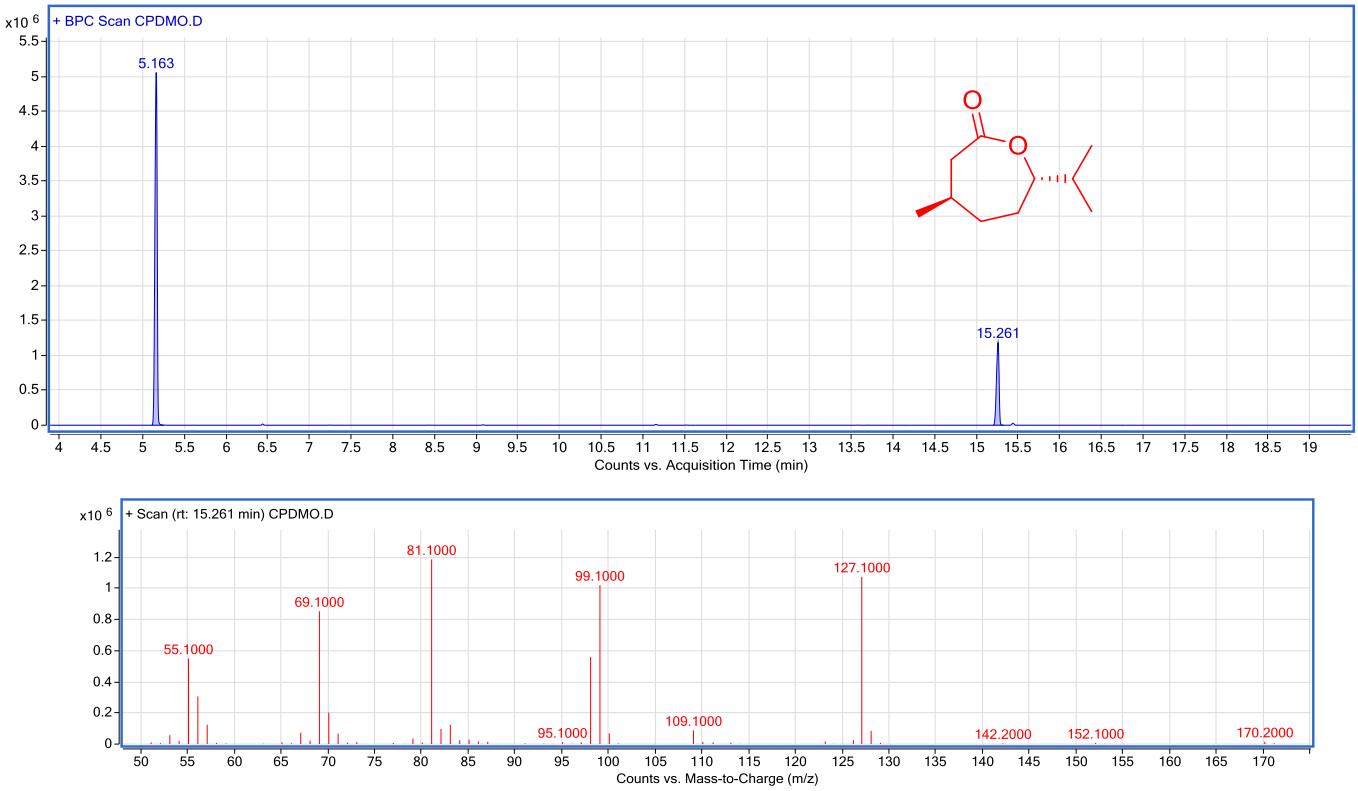
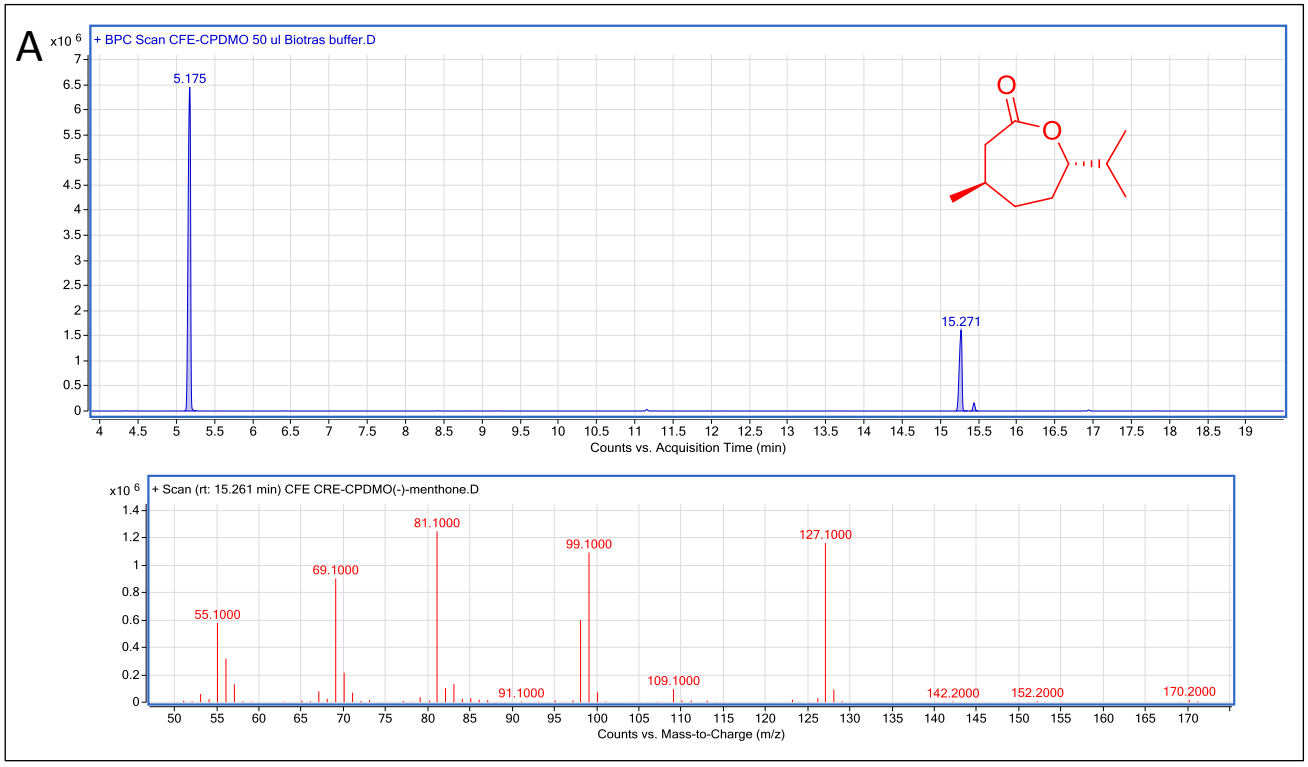


Figure S4. GC/MS chromatograms of the biotransformation reaction product of the purified CPDMO with (-)-menthone. Biotransformation reaction conditions: purified CPDMO (2 μM), (-)-menthone (5 mM), $NADP^+$ (15 μM), glucose (15 mM), glucose dehydrogenase (GDH from *Pseudomonas* sp., 10U) in 1 mL 50 mM Tris buffer, pH 7.0 at 28 °C. The reaction products were extracted in ethyl acetate containing 0.1% *sec*-butyl-benzene as internal standard for GC/MS analysis. The retention times and mass fragmentation pattern for each compound are shown in the upper and lower panels, respectively.



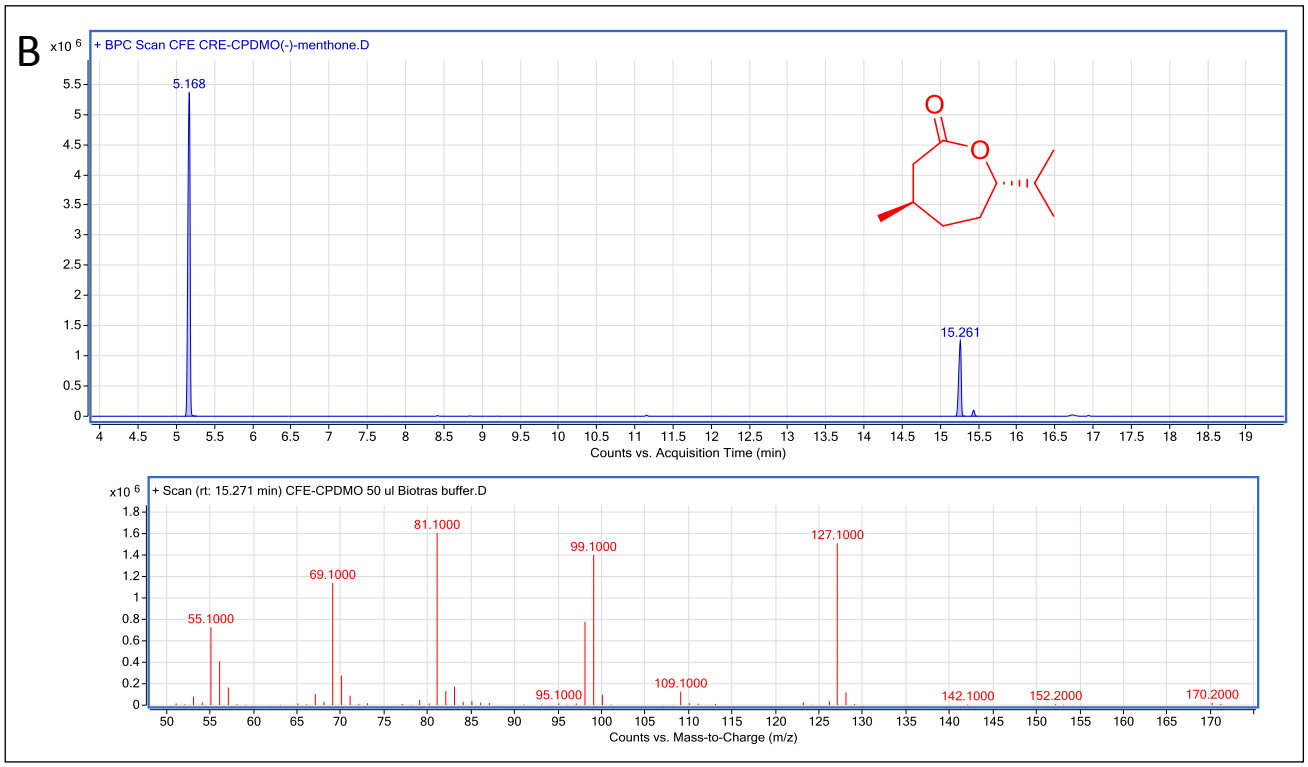


Figure S5. GC/MS chromatograms of the biotransformation reaction products of CFE of CPDMO (A) and CFE of CRE-CPDMO (B) with (-)-menthone. Reaction conditions: A) CFE of CPDMO (50 µL), (-)-menthone (5 mM), $NADP^+$ (15 µM), glucose (15 mM) and glucose dehydrogenase (GDH from *Pseudomonas* sp., 10U) in 1 mL 50 mM Tris buffer, pH 7.0 at 28 °C. B) CFE of CRE-CPDMO (50 µL), (-)-menthone (5 mM) and 100 mM sodium phosphite in 1 mL 50 mM Tris buffer, pH 7.0 at 28 °C The reaction products were extracted in ethyl acetate containing 0.1% *sec*-butyl-benzene as internal standards for GC/MS analysis. The retention times and mass fragmentation pattern for each compound are shown in the upper and lower panels of each part, respectively. Additional tiny peaks at RT of 15.432 was also identified.

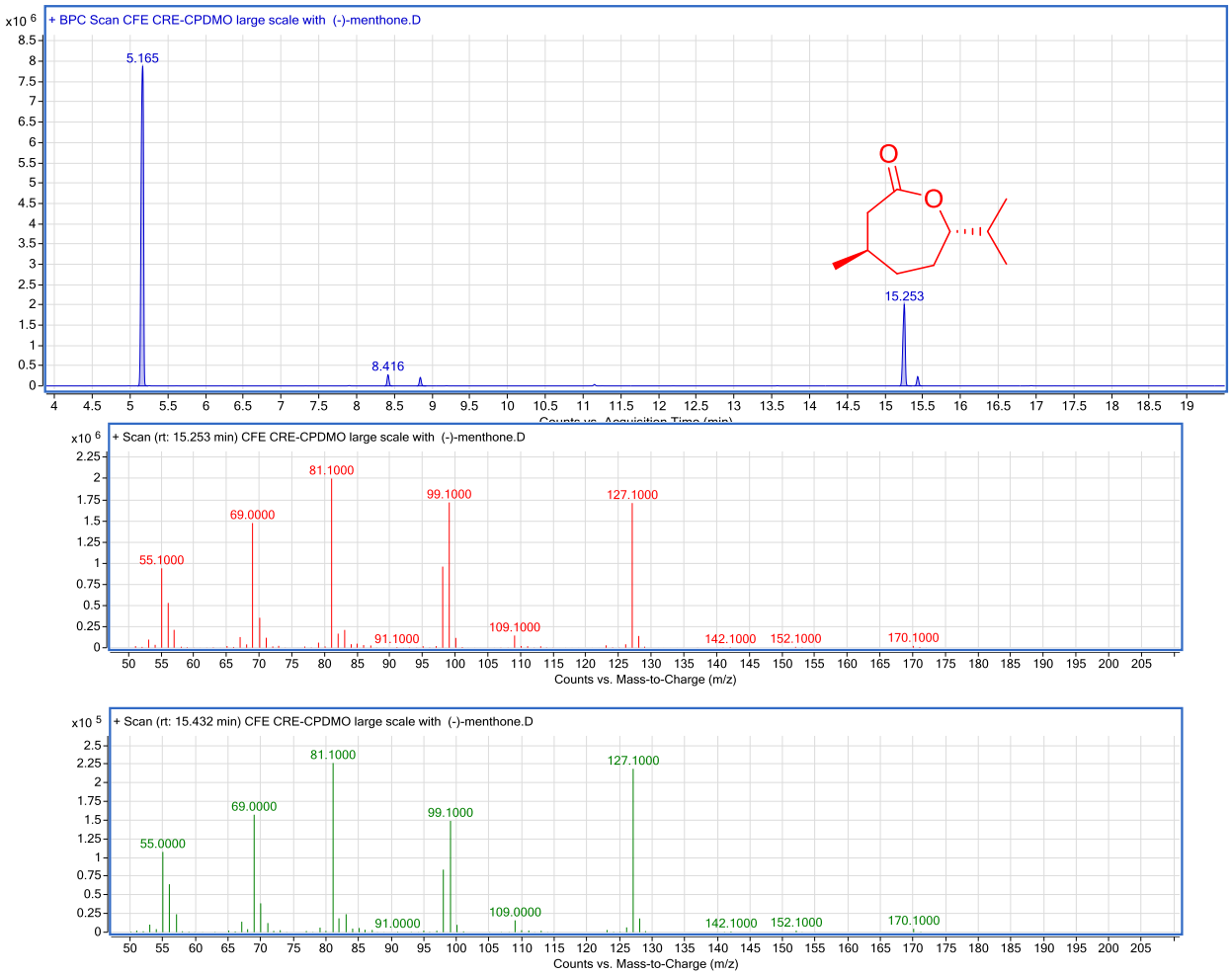


Figure S6. GC/MS chromatograms of the biotransformation reaction products of the CFE of CRE-CPDMO with 5 mM (-)-menthone in a 500 mL reaction volume. Reaction conditions: CFE of CRE-CPDMO (50 µL/mL), (-)-menthone (5 mM) and 100 mM sodium phosphite in 500 mL 50 mM Tris buffer, pH 7.0 at 28 °C. The reaction products were extracted in ethyl acetate containing 0.1% *sec*-butyl-benzene as internal standards for GC/MS analysis. The retention times and mass fragmentation pattern for each compound are shown in the upper and lower panels, respectively. The retention times for (-)-menthone, menthide product and a possible additional lactone isomer are 8.416, 15.253 and 15.432, respectively. Additional tiny peaks at RT of 8.86 was also identified.

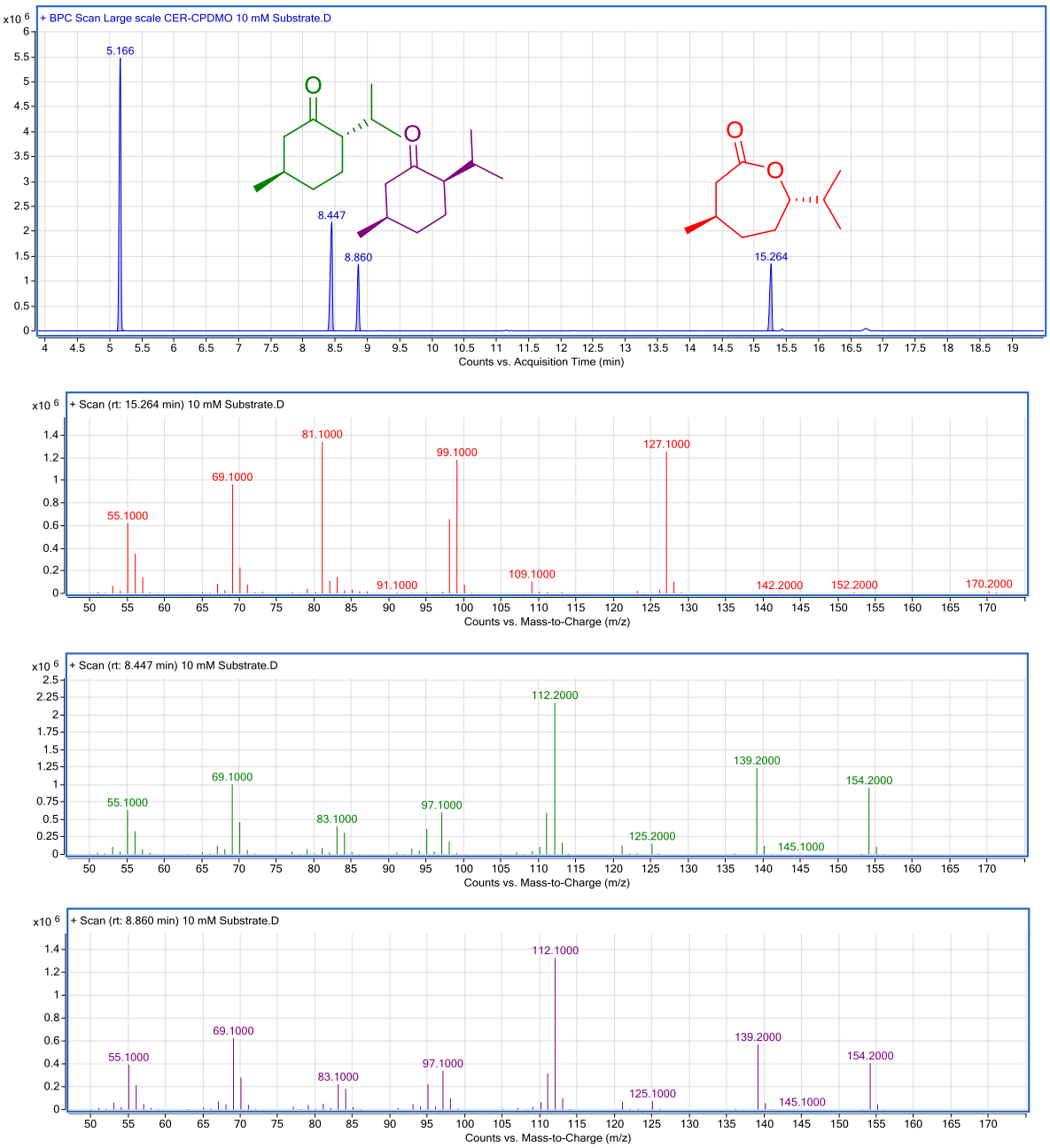
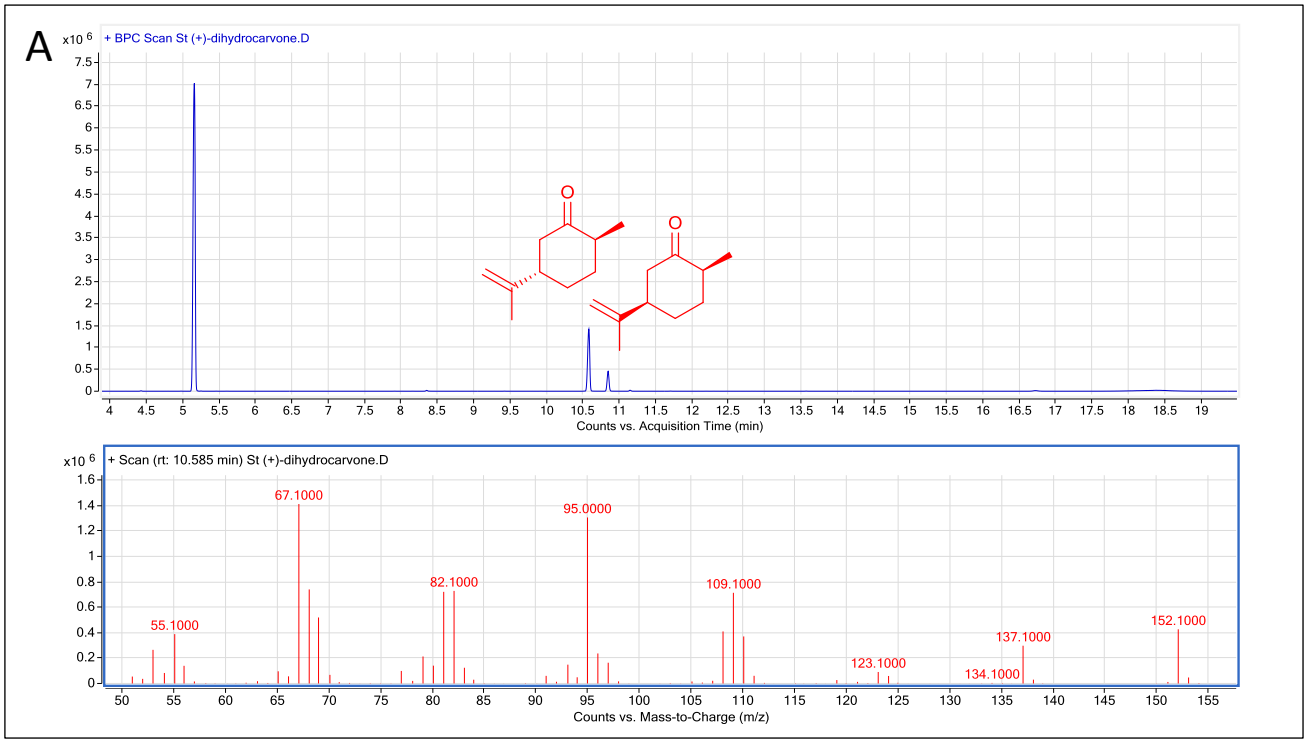


Figure S7. GC/MS chromatograms of the biotransformation reaction products of the CFE of CRE-CPDMO with 10 mM (-)-menthone in 500 mL reaction volume. Biotransformation reaction conditions: CFE of CRE-CPDMO (50 µL/mL), (-)-menthone (10 mM) and 100 mM sodium phosphite in 500 mL 50 mM Tris, pH 7.0 at 28 °C. The reaction products were extracted in ethyl acetate containing 0.1% *sec*-butyl-benzene as internal standards for GC/MS analysis. The retention times and mass fragmentation pattern for each compound are shown in the upper and lower panels, respectively. The retention times for (-)-menthone, menthide product and isomenthone are 8.447, 15.264 and 8.86, respectively. Additional tiny peaks at RT of 8.86 was also identified.



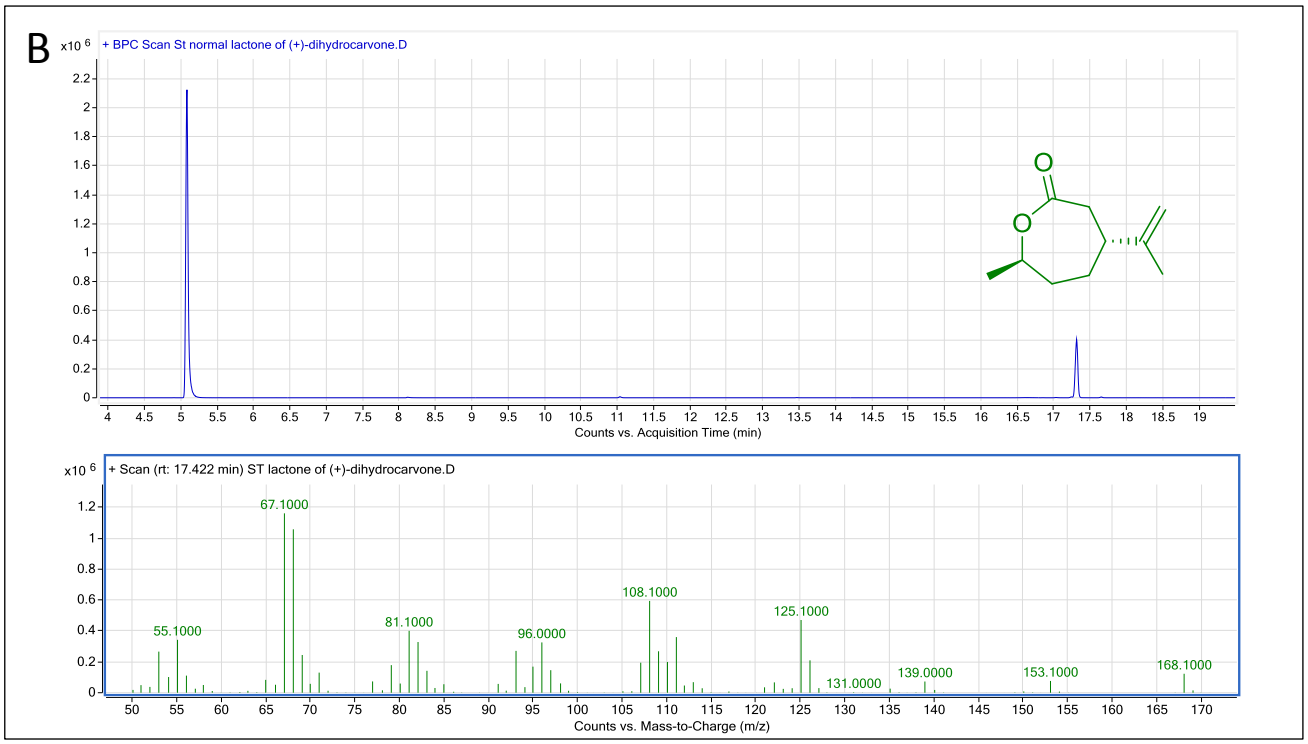


Figure S8. GC/MS chromatograms of the (+)-dihydrocarvone and dihydrocarvide (the normal lactone of (2R,5R)-(+)-dihydrocarvone) standards. A) Spectrum of (+)-dihydrocarvone (5 mM) which is a mixture of 77% of the (*2R,5R)* isomer (RT = 10.585 min) and 22% of the (*2R,5S*) isomer (RT = 10.845 min). Both isomers have identical mass spectra as shown in red. B) Spectrum of the normal lactone (5 mM) in green (RT = 17.422). Both standards were prepared in ethyl acetate containing 0.1 % *sec*-butyl-benzene as internal standard for GC/MS analysis.

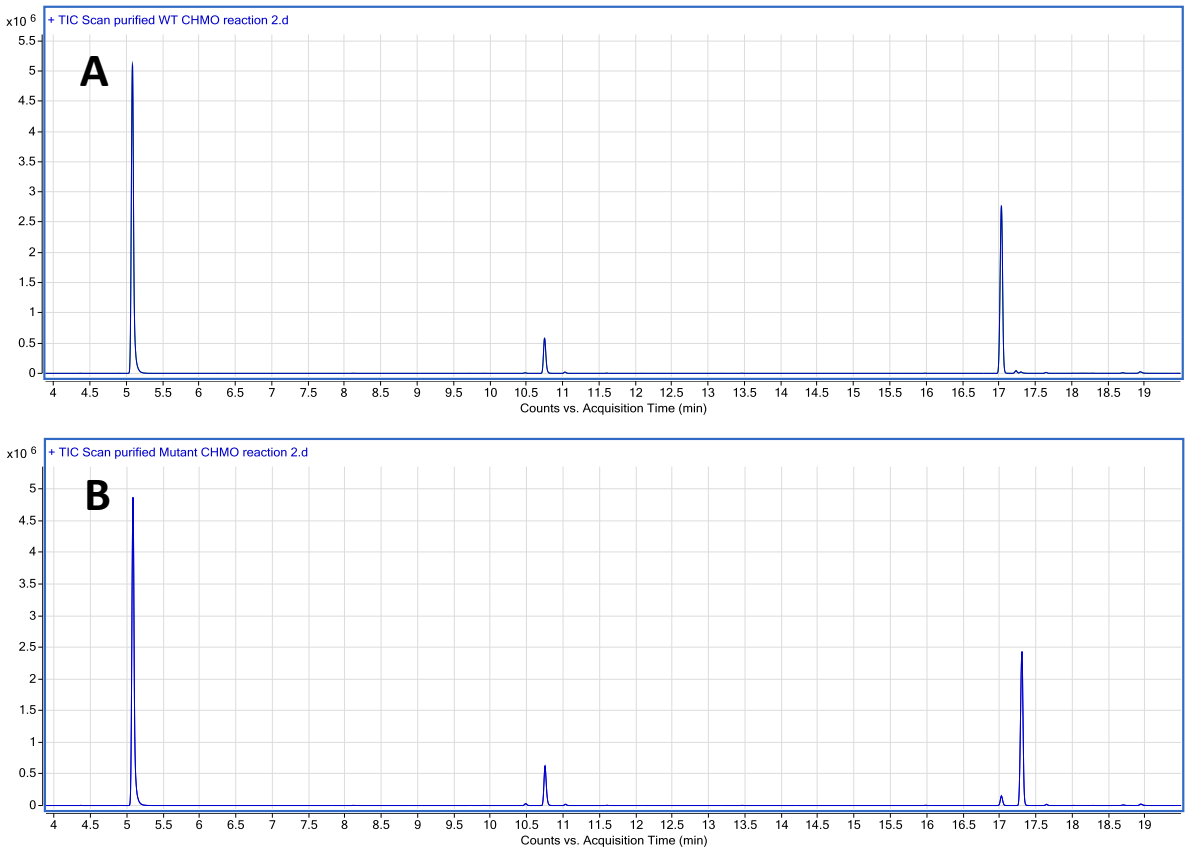


Figure S9. GC/MS chromatograms of the reaction products of purified WT CHMO$_{Phi1}$ A) and variant CHMO$_{Phi1}$ B) with (+)-dihydrocarvone. Biotransformation reaction conditions: WT CHMO$_{Phi1}$ or its variant (2 µM), (+)-dihydrocarvone (5 mM), $NADP^+$ (15 µM), glucose (15 mM), glucose dehydrogenase (GDH from *Pseudomonas* sp., 10U) in 1 mL 50 mM Tris buffer, pH 7.0 at 28 °C. The reaction products were extracted in ethyl acetate containing 0.1% *sec*-butyl-benzene as internal standard for GC/MS analysis. The retention times for the normal lactone, abnormal lactone and unreacted (*2R,5S*)-(+)-dihydrocarvide are 17.42, 17.147 and 10.7, respectively.

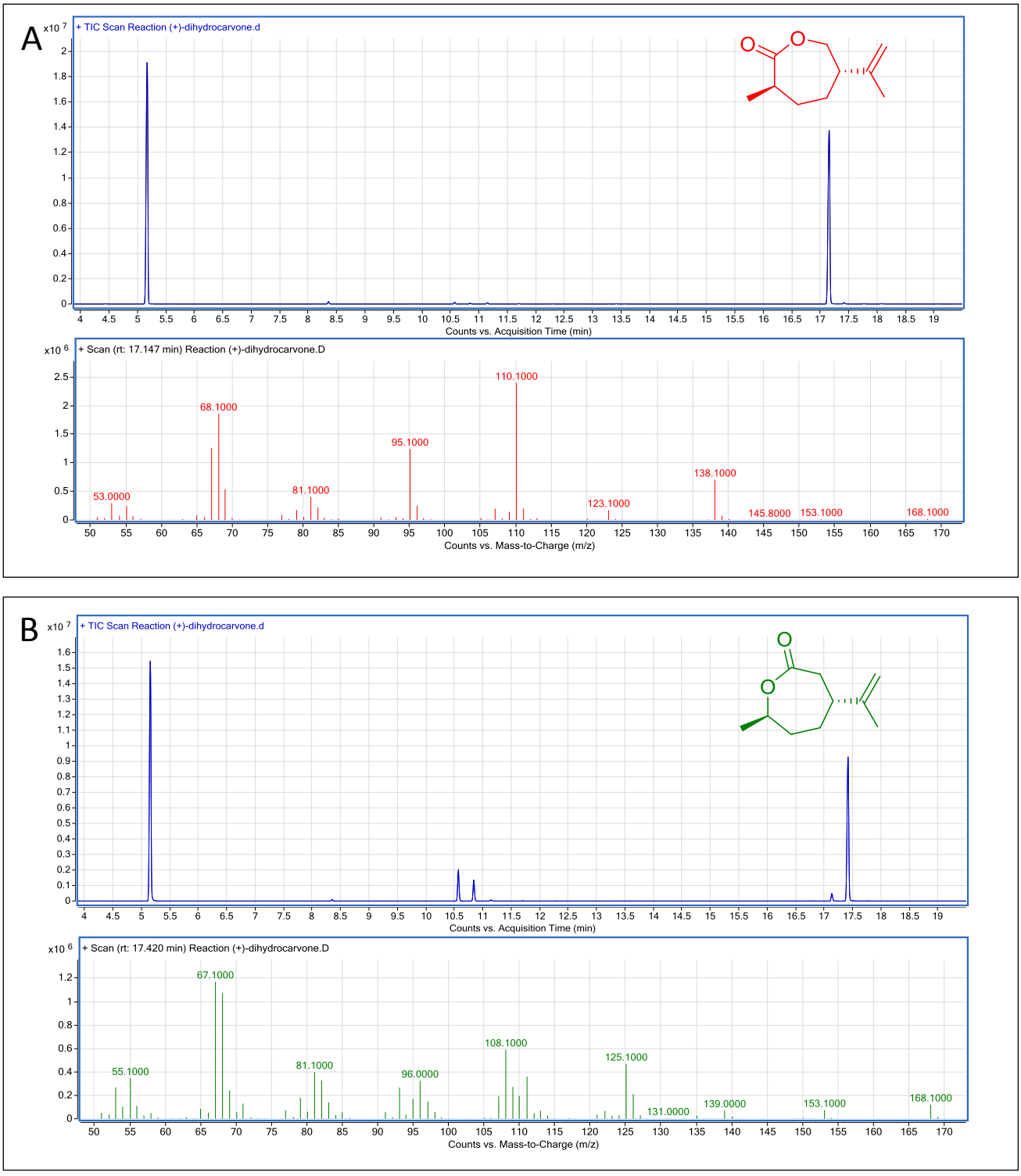


Figure S10. GC/MS chromatograms of the reaction products of the CFE of WT CRE CHMO$_{Phi1}$ in A) and its variant in B) with (+)-dihydrocarvone. Reaction conditions: CFE of WT CRE-CHMO$_{Phi1}$ or its variant (50 µL/mL), (+)-dihydrocarvone (5 mM) and 100 mM sodium phosphite in 500 mL 50 mM Tris, pH 7.0 at 28 °C. The reaction products were extracted in ethyl acetate containing 0.1% *sec*-butyl-benzene as internal standard for GC/MS analysis. The retention times and mass fragmentation pattern for each compound are shown in the upper and lower panels of each part, respectively. The retention times for (*2R,5R*)-(+)-dihydrocarvone, (*2R,5S*)-(+)-dihydrocarvone, abnormal lactone and normal lactone are 10.573, 10.845, 17.147 and 17.42, respectively.

^{1}H and ^{13}C NMR Spectra of Lactone monomers

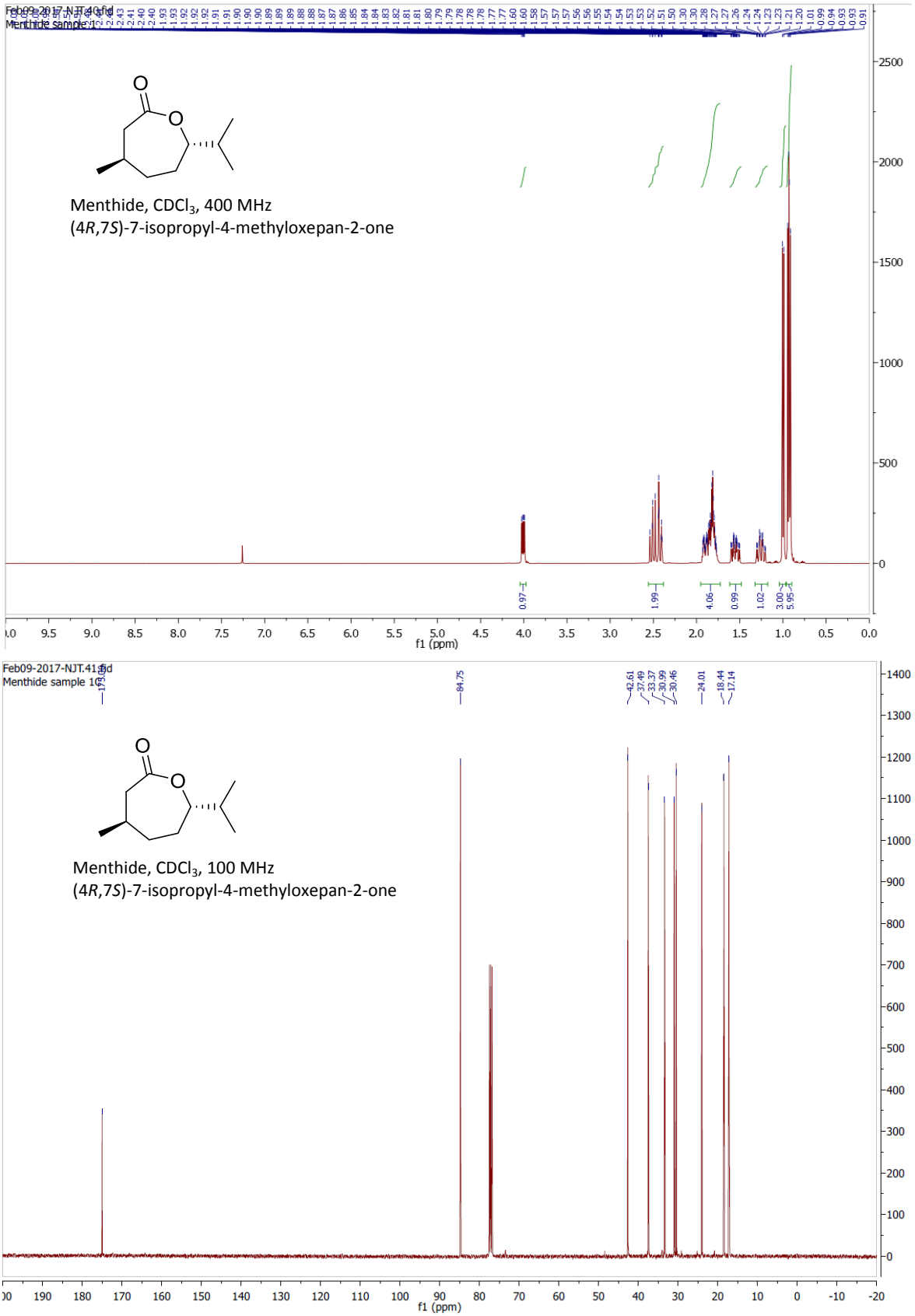
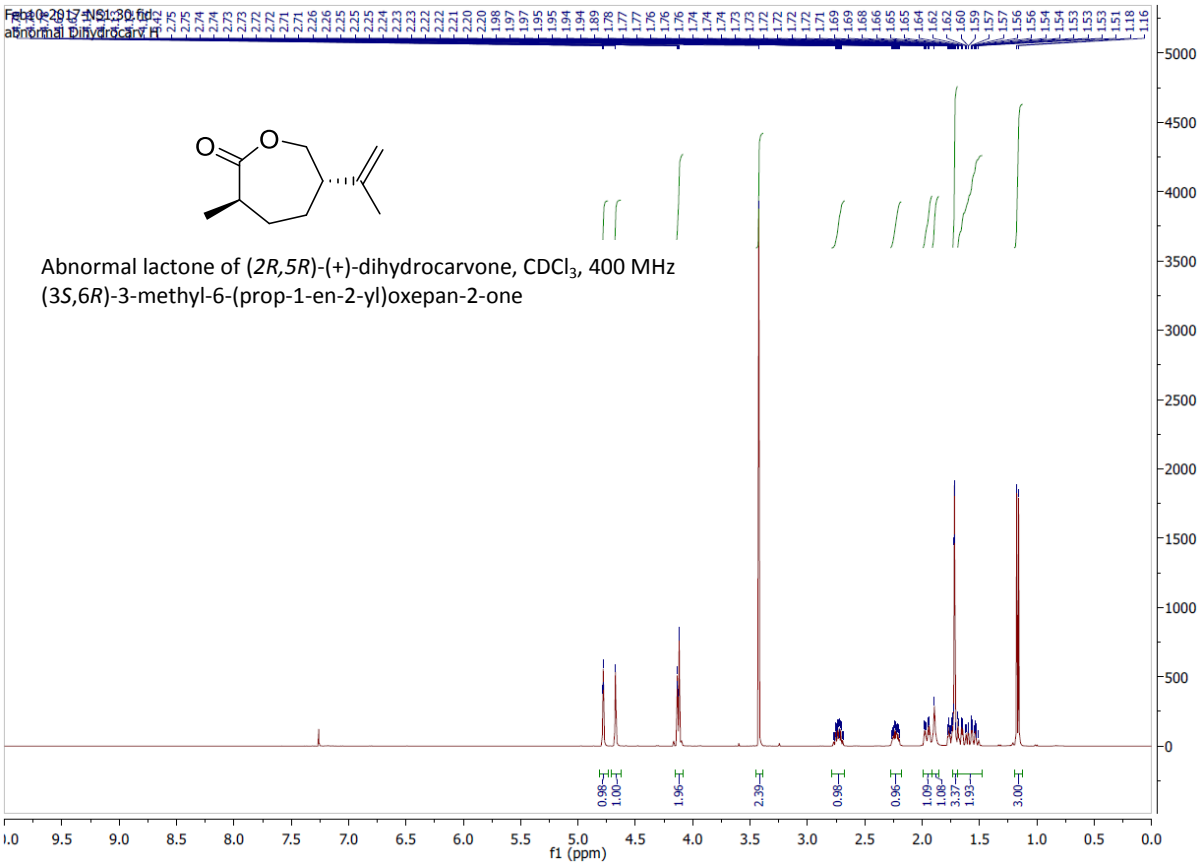


Figure S11. ^{1}H and ^{13}C Spectra of menthide. NMR analyses were performed on a Bruker 400 MHz NMR spectrometer.



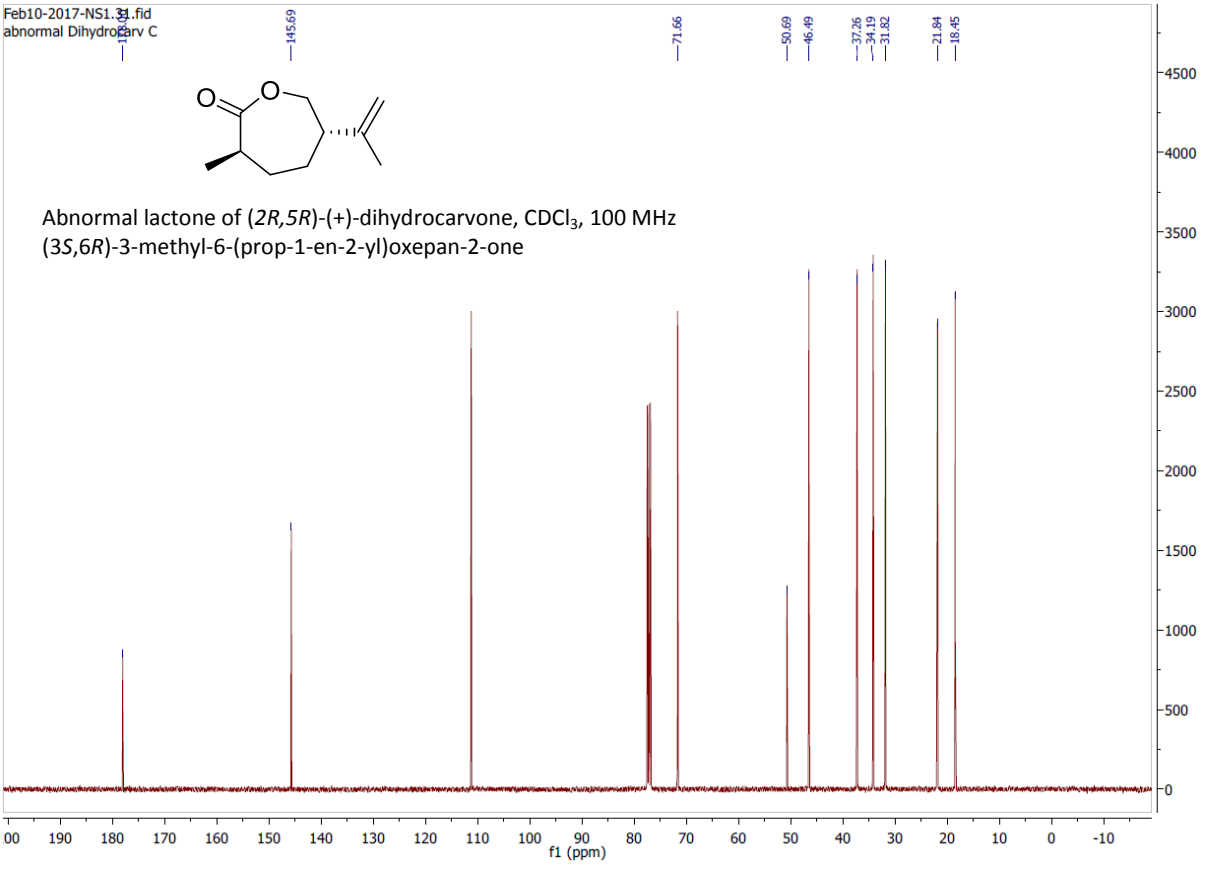


Figure S12. ^{1}H and ^{13}C Spectra of the abnormal lactone of (*2R,5R*)-(+)-dihydrocarvone. NMR analyses were performed on a Bruker 400 MHz NMR spectrometer.

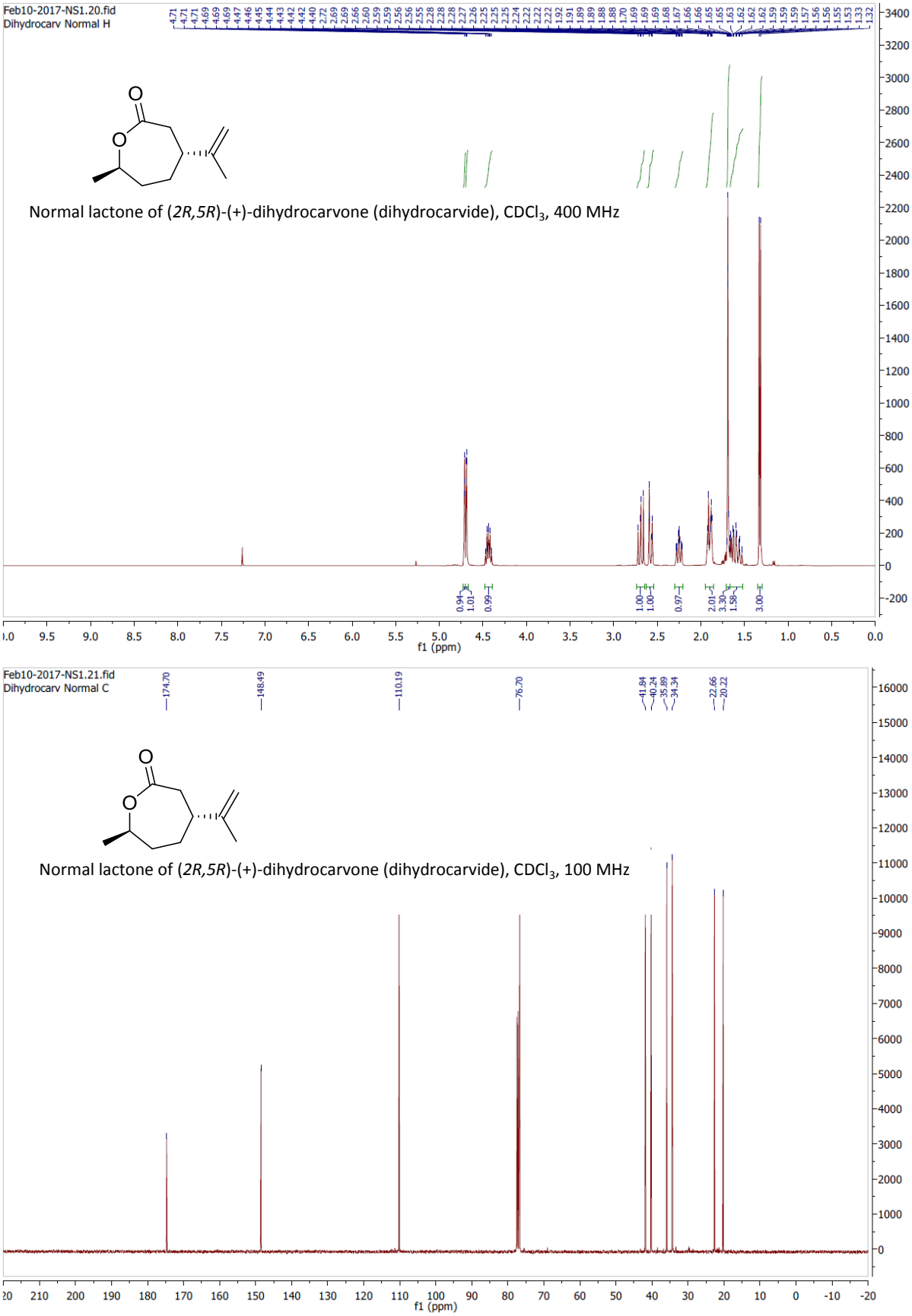


Figure S13. ^{1}H and ^{13}C Spectra of the normal lactone of (*2R,5R*)-(+)-dihydrocarvone (dihydrocarvide). NMR analyses were performed on a Bruker 400 MHz NMR spectrometer.

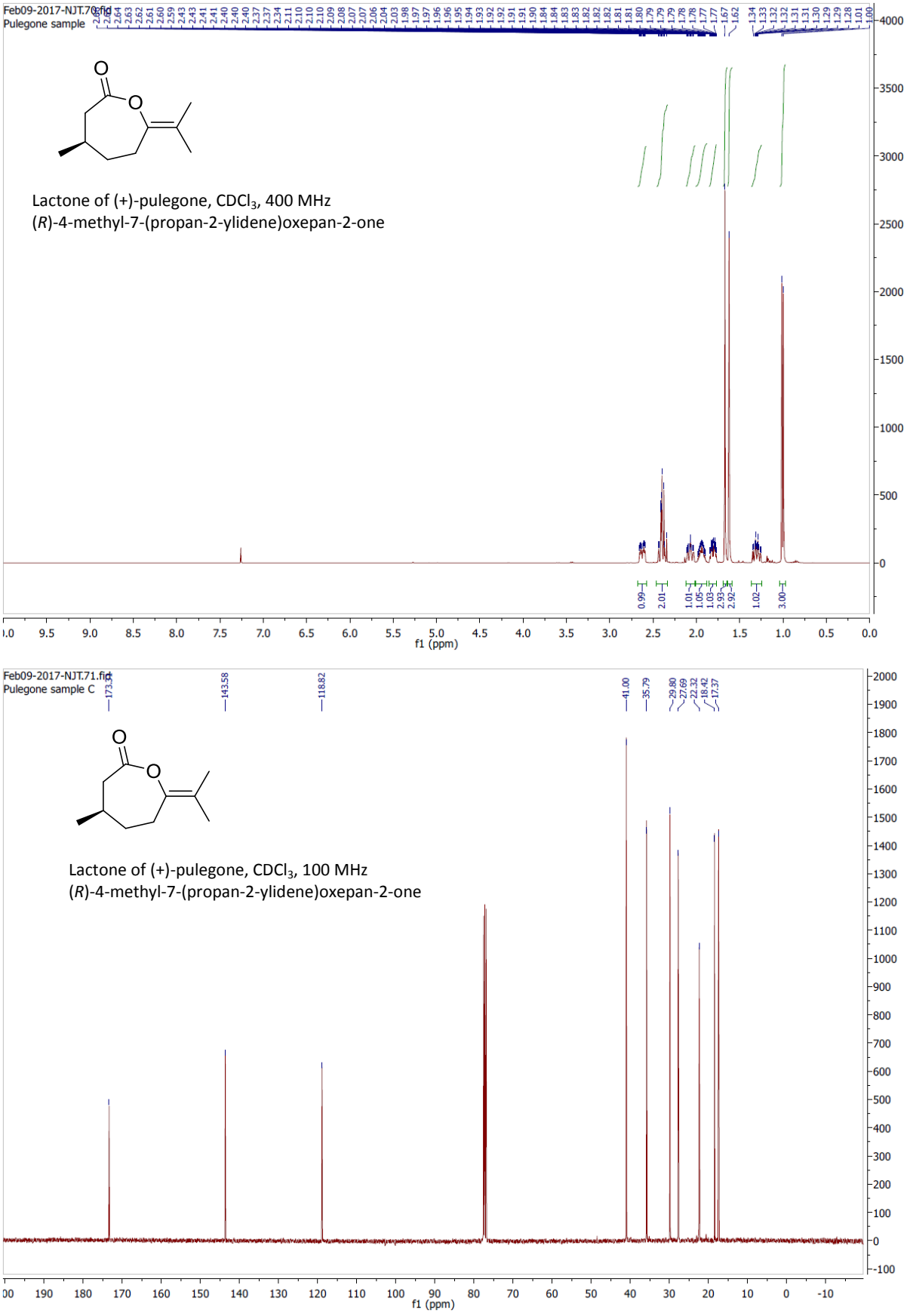
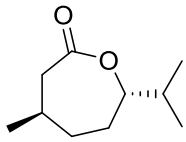


Figure S14. ^{1}H and ^{13}C Spectra of the lactone of (+)-pulegone. NMR analyses were performed on a Bruker 400 MHz NMR spectrometer.

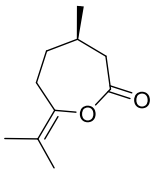
Characterisation Data for the Lactone monomers

(4*R*,7*S*)-7-isopropyl-4-methyloxepan-2-one [Menthide]



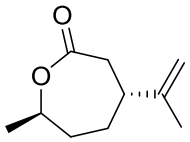
Clear oil (402 mg) **^{1}H NMR** (400 MHz, $CDCl_3$) 3.99-4.02 (1H, dd, J = 4, 4 Hz), 2.40-2.54 (2H, m), 3.38-3.42 (1H, m), 1.77-1.93 (4H, m), 1.50-1.60 (1H, m), 1.20-1.30 (1H, m), 1.00 (3H, d, J = 8 Hz), 0.91-0.94 (6H, t, J = 8 Hz) **^{13}C NMR** (100 MHz, $CDCl_3$) 175.00, 84.75, 42.61, 37.49, 33.37, 30.99, 30.46, 24.01, 18.44, 17.14.

(*R*)-4-methyl-7-(propan-2-ylidene)oxepan-2-one [lactone of (+)-Pulegone]



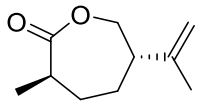
Clear oil (121 mg) **^{1}H NMR** (400 MHz, $CDCl_3$) 2.59-2.66 (1H, m), 2.34-2.43 (2H, m), 2.03-2.11 (1H, m), 1.90-1.98 (1H, m), 1.77-1.84 (1H, m), 1.67 (1H, s), 1.62 (1H, s), 1.26-1.35 (1H, s), 1.05 (3H, d, J = 4 Hz) **^{13}C NMR** (100 MHz, $CDCl_3$) 173.31, 143.58, 18.82, 41.00, 35.79, 29.80, 27.69, 22.32, 18.42, 17.37.

(4*R*,7*R*)-7-methyl-4-(prop-1-en-2-yl)oxepan-2-one (Normal lactone of (*2R*,*5R*)-(+)-dihydrocarvone= dihydrocarvide)



Clear oil (211 mg) **^{1}H NMR** (400 MHz, $CDCl_3$) 4.71 (1H, brs), 4.69 (1H, brs), 4.40-4.47 (1H, m), 2.69 (1H, t, J = 12 Hz), 2.55-2.60 (1H, m), 2.25 (1H, t, J = 12 Hz), 1.88-1.92 (2H, m), 1.69 (3H, s), 1.53-1.66 (2H, m), 1.32 (3H, d, J = 4 Hz) **^{13}C NMR** (100 MHz, $CDCl_3$) 174.70, 148.49, 110.19, 76.70, 41.84, 40.24, 35.89, 34.34, 22.66, 20.22.

(3*R*,6*S*)-3-methyl-6-(prop-1-en-2-yl)oxepan-2-one (Abnormal lactone of (*2R*,*5R*)-(+)-dihydrocarvone)



Clear oil (190 mg) **^{1}H NMR** (400 MHz, $CDCl_3$) 4.78 (1H, brs), 4.67 (1H, brs), 4.12-4.14 (2H, m), 3.42 (3H, s), 2.69-2.77 (1H, m), 2.20-2.27 (1H, m), 1.72 (3H, brs), 1.51-1.69 (1H, m), 1.17 (3H, d, J = 8 Hz) **^{13}C NMR** (100 MHz, $CDCl_3$) 178.00, 145.69, 111.20, 71.66, 50.69, 46.49, 37.26, 34.19, 31.82, 21.84, 18.45.

Polymerization Methods

Ring Opening Polymerization

The composition of the polymerization reactions is described in Table S3.

Table S3. Ring-opening polymerization (ROP) reaction setup.

Monomer	$[\underline{M}]_0$ (mg)	Catalyst (mg)	Initiator (μL)	Toluene (mL)	Polymer
ε-Caprolactone	400	21.3	3.63	1.4	Polycarprolactone (PCL)
Menthide	280.6	20	3.4	1.714	Polymenthide (PM)
Dihydrocarvide	277.28	20	3.4	1.714	Polydihydrocarvide (PDC)

$[M]_0$ is the lactone monomer concentration of 1 M. both the catalyst$[Mg(BHT)_2(THF)_2]_0$ and the initiator $[BnOH]_0$ are at 0.02 M concentration. *i.e.* $[M]_0$:$[Mg(BHT)2(THF)2]_o$:$[BnOH]_o$ is at a ratio of 50:1:1. The concentrations of the reactants are similar to the method reported previously.[3]

Gel Permeation Chromatography (GPC)

For GPC analysis, conventional measurements were performed using a Marlvern GPC (Viscotek GPC max model VE2001 GPC) with a Viscotek VE 3580 differential refractometer detector calibrated with polystyrene standards. Samples were dissolved in THF (1 mg/mL). The data were acquired and analysed on Viscotek OMNISEC software. The polymers were further characterised at Malvern Laboratories, UK using the OMNISEC system by advanced Mutli-Detection GPC from Malvern Instruments (Table 3 in the manuscript, Figure S17) . The GPC was coupled with refractive index (RI), Right-Angle Light Scattering (RALS), Low-Angle Light Scattering (LALS) and viscometer detection. Samples were prepared in THF (3 mg/mL). The three polymers were compared using the Mark-Houwink plot/equation (Equation 1). The temperature and solvent were fixed to investigate the relationship between the intrinsic viscosity $[\eta]_i$ and the molecular weight M_i of the fractionated polymer samples.

$$[\eta]_i = KM_i^a \qquad \textbf{Equation 1}$$

Table S4. Molecular weight and dispersitiesof the crude polymers obtained by conventional GPC analysis.

Polymer	M_n (*k*Da)	M_w (*k*Da)	$Đ_M$
Polymenthide (PM)	9.5	11.5	1.217
Polydihydrocarvide (PDC)	6.8	8.4	1.235
Polycaprolactone (PCL)	7.4	20.9	2.8

Data were obtained by analysing the polymer samples in a Viscotek GPC max (model VE2001) with differential refractometer detector (Viscotek VE 3580) using narrow distribution polystyrene standards. M_n = number average molecular weight (*k*Da); M*w* = weight average molecular weight (Da); Đ = sample dispersity (M_w/M_n). Samples were prepared in THF (1 mg/mL of solid polymer or 30 μL of liquid polymer/mL), with fractions (200 μL) separated on a set of standard ViscoGel columns. The columns and detectors were maintained at 35 °C.

NMR Spectra of the Polymers

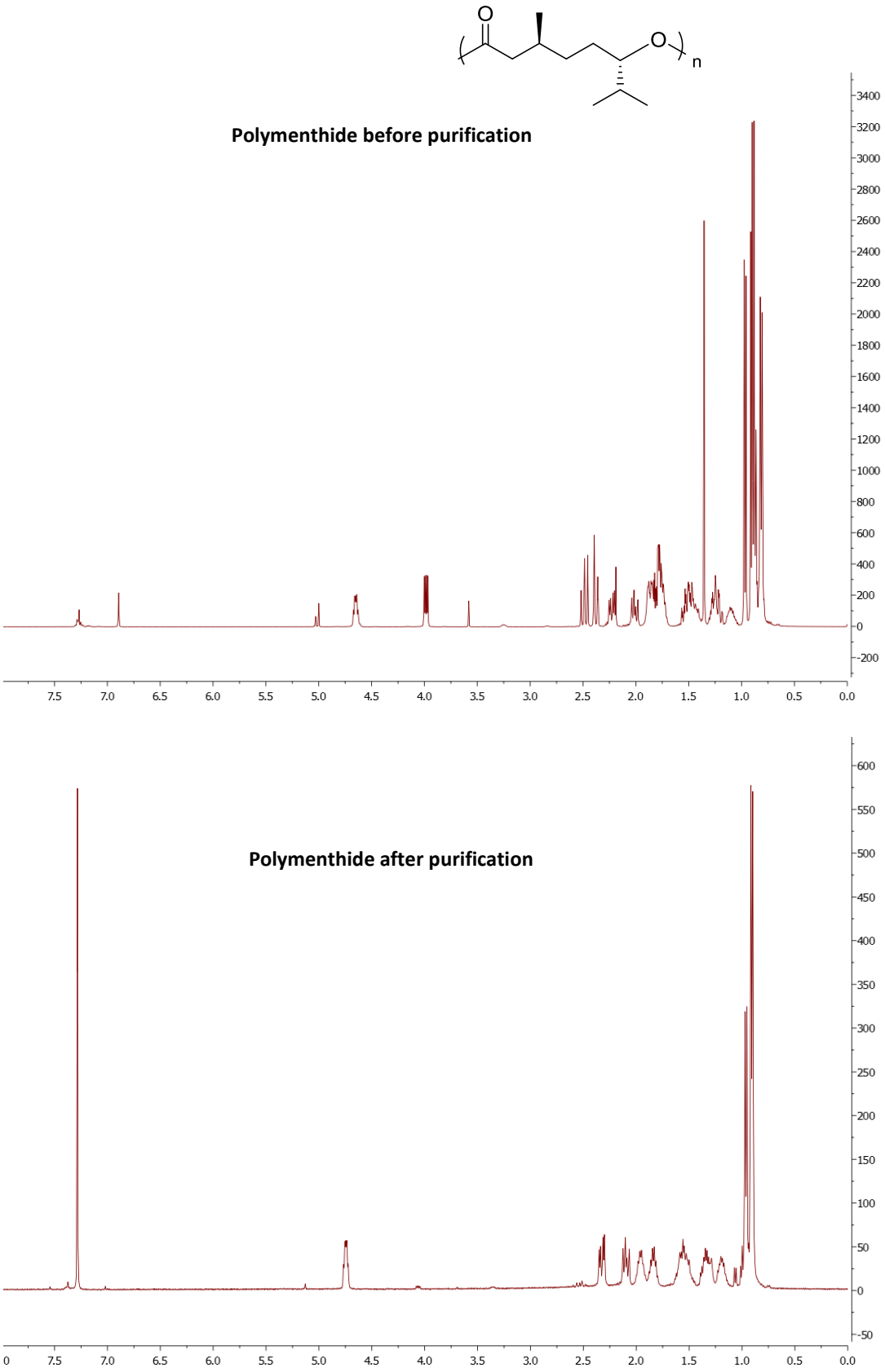


Figure S15. ^{1}H and ^{13}C Spectra of polymenthide (PM). NMR analyses were performed on a Bruker 400 MHz NMR spectrometer.

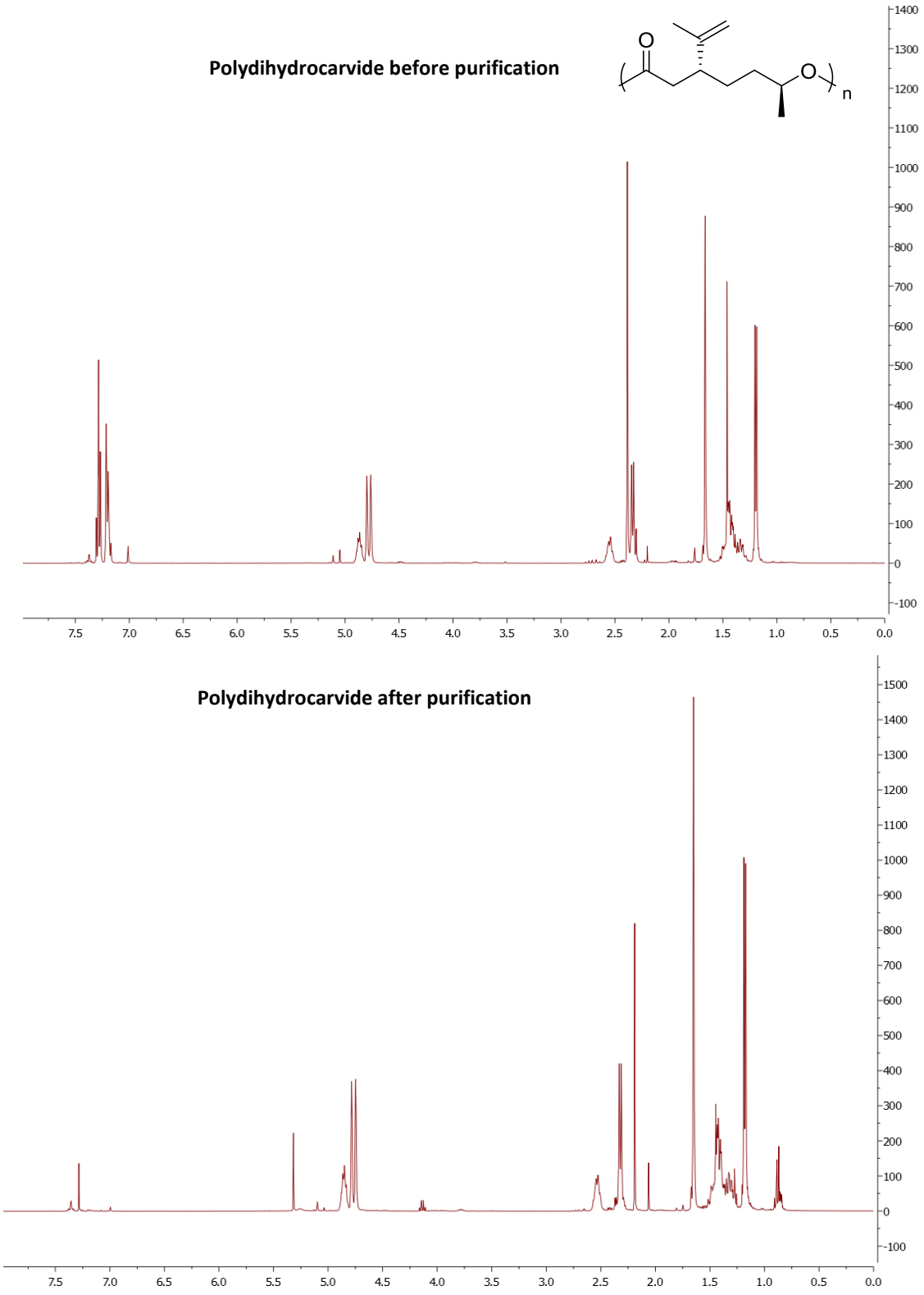


Figure S16. ^{1}H and ^{13}C Spectra of polydihydrocarvide (PDC). NMR analyses were performed on a Bruker 400 MHz NMR spectrometer.

Polymer Analysis

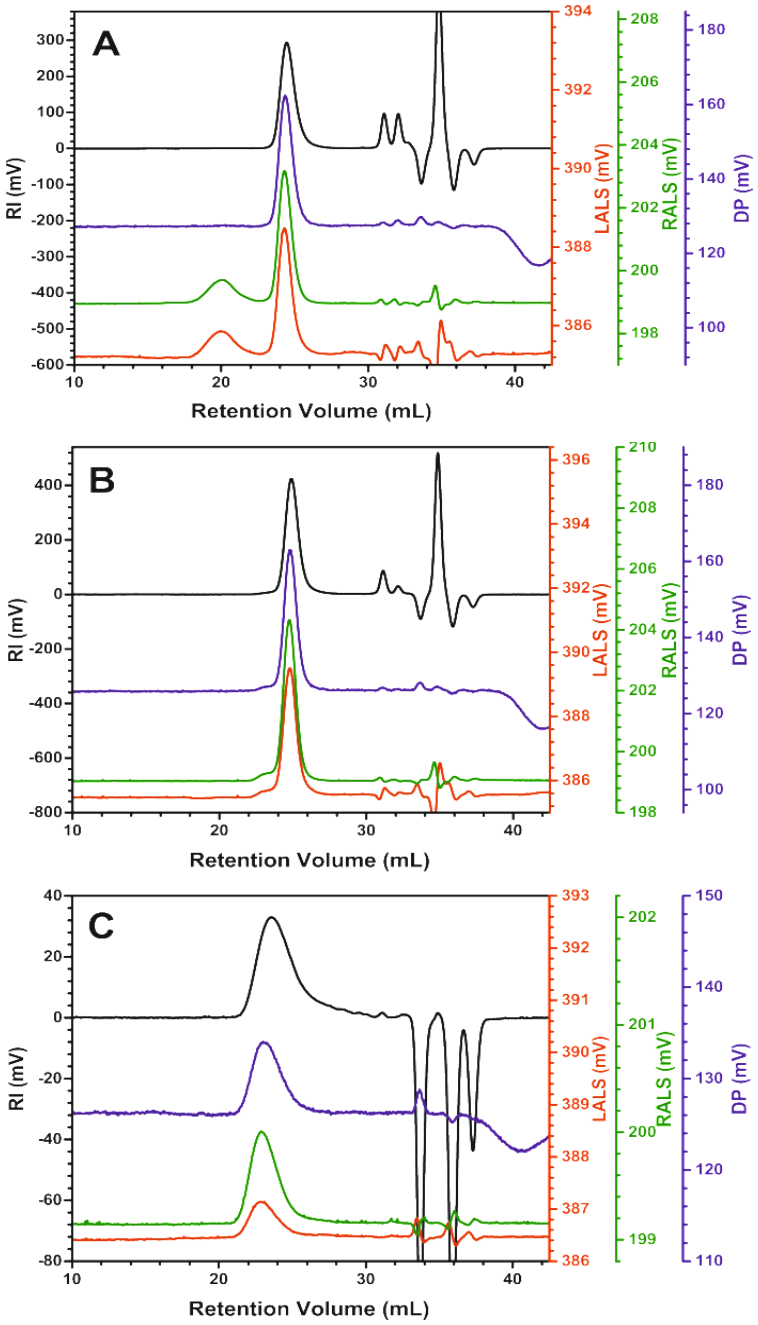


Figure S17. Triple detection GP chromatograms of the produced polymers PM (in A), PDC (in B) and PCL (in C) using the OMISEC™ triple detection method. The data were acquired and analysed by OMISEC™ Multi-Detection GPC system from Malvern Instruments. The GPC was coupled with refractive index (RI in black), Right-Angle Light Scattering (RALS in green), Low-Angle Light Scattering (LALS in red) and viscometer detection (differential pressure, DP in blue). Samples were prepared in THF (3 mg/mL of solid polymer or 20 μl of liquid polymer/mL) and 50 μl injections of samples were separated isocratically on a set of standard ViscoGel columns. The columns and detectors were maintained at 35 °C. The detectors recorded the elution of the samples from the columns and the analysis was performed using the OMNISEC™ V10.20 software.

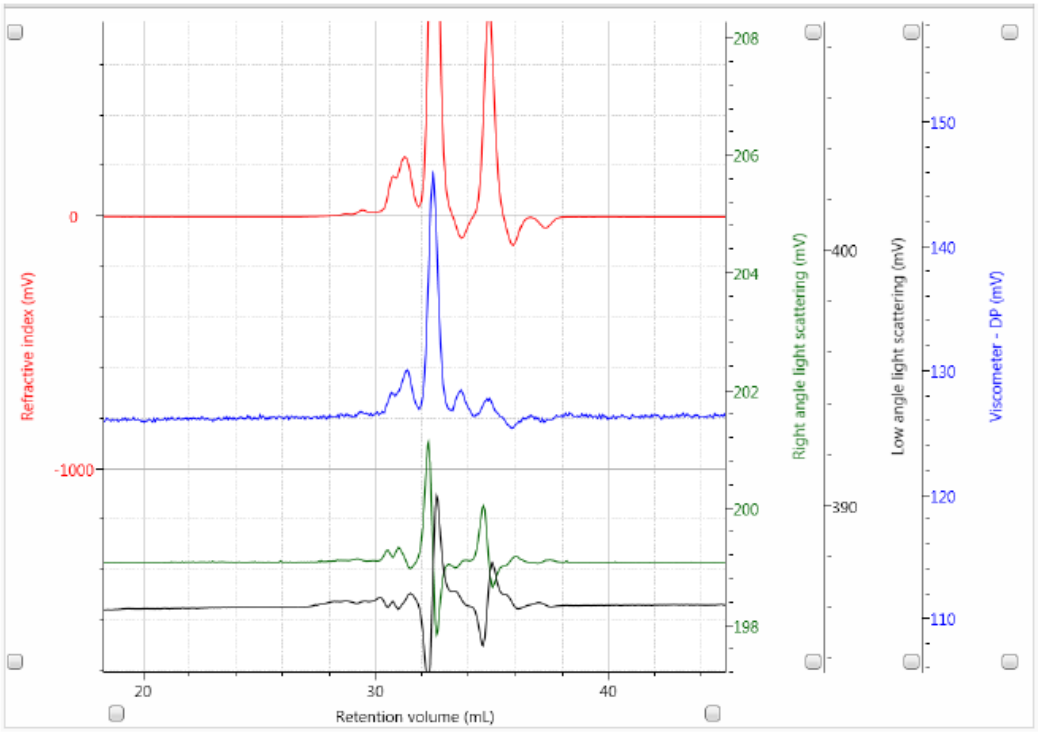


Figure S18. GP chromatograms for the polymer of the lactone of (+)-pulegone using the OMISECTM triple detection method. The data were acquired and analysed by the advanced OMISECTM Multi-Detection GPC system (Malvern Instruments). The GPC coupled with refractive index (RI in red), Right-Angle Light Scattering (RALS in green), Low-Angle Light Scattering (LALS in black) and viscometer detection (differential pressure, DP in blue). Samples were prepared in THF (3 mg/mL or 20 µl of liquid polymers/mL), 50 µl injections of samples were separated isocratically on a set of standard ViscoGel columns. The columns and detectors were maintained at 35 °C. The detectors recorded the elution of the samples from the columns and the analysis was performed using the OMNISECTM V10.20 software. The analyte peak eluted very close to the permeation peak, at approximately 32 mL and the light scattering showed both positive and negative peaks

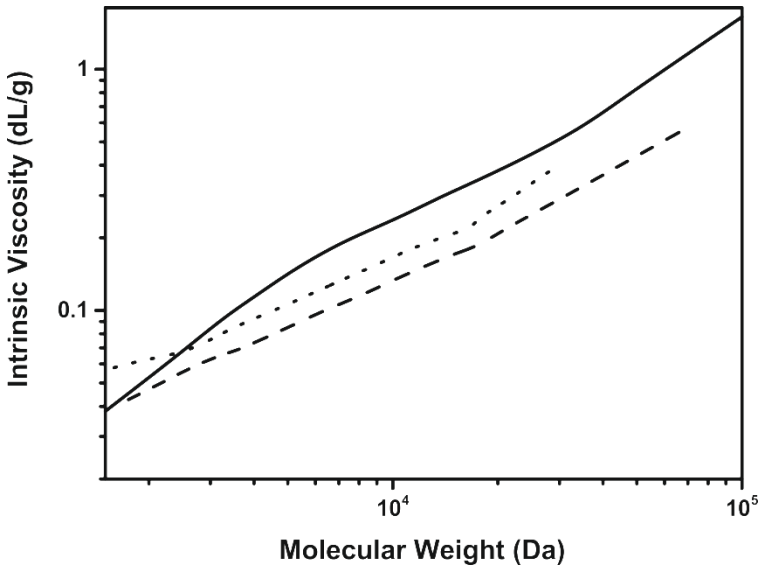


Figure S19. Mark-Houwink plot for the generated polymers PM (dotted line), PDC (dashed line) and PCL (solid line). The plot was produced by plotting the M_w from the light scattering detector against the intrinsic viscosity on a log-log graph (Equation S1). Samples were prepared in THF (3 mg/mL for solid polymer or 20 μL of liquid polymer/mL) and 50 μL injections of samples were separated isocratically on a set of standard ViscoGel columns. The columns and detectors were maintained at 35 °C. The detectors recorded the elution of the samples from the columns and the analysis was performed using the OMNISEC^{TM} V10.20 software.

Similar ROP reactions were performed with the lactone of pulegone, with ^{1}H NMR analysis showing a hint of polymer formation. However, GPC analysis did not show any evidence of polymer formations, with the potential analyte peak eluting very close to the permeation peak and light scattering showed both positive and negative peaks (Figure S24). This may be due to refractive index changes within the sample, due to low molecular weight compound presence (e.g. impurities). The lack of polymer formation with the lactone of pulegone may be due to steric hindrance of the orientation of the isobutene of the monomer. Further studies are required to determine suitable ROP conditions to enable polymers of pulegone lactone formation to be successful.

Crystal Structure Discussion

Sequence and Structural Holomogy

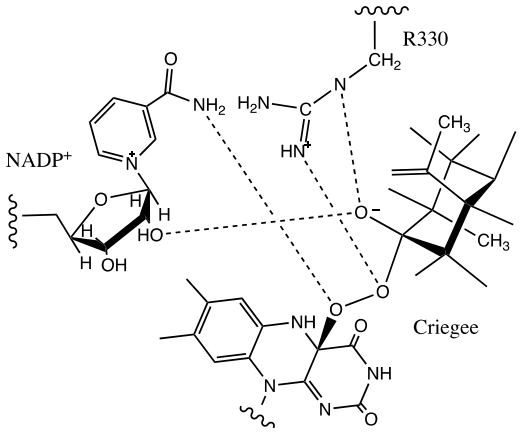
Several structures of BVMO proteins were available in the protein data bank. The closest homologues to $CHMO_{Phi1}$ are CHMO structures from *Rhodococcus sp.* HI-31[4-6], phenylacetone monooxygenase (PAMO) from *Thermobifida fusca*[7, 8], OTEMO from *Pseudomonas putida*[9], steroid monooxygenase (STMO) from *Rhodococcus rhodochrous*[10] and more recently $BVMO_{AFL838}$ from *Aspergillus flavus*[11] and CHMO structures from the thermophile *Thermocrispum municipall.*[12] Analysis using the DALI server[13] identified CHMO structures from *Rhodococcus sp.* HI-31, namely $CHMO_{open}$ (PDB 3GWF), $CHMO_{closed}$ (PDB 3GWD), $CHMO_{rotated}$ in complex with cyclohexanone (PDB 3UCL), and $CHMO_{Tight}$ (PDB 4RG3) and $CHMO_{Loose}$ (PDB 4RG4) in complex with ε-caprolactone, have the highest similarity with $CHMO_{Phi1}$, with RMSD values ranging from 0.6-0.9 Å. $CHMO_{Phi1}$ and CHMO from *Rhodococcus sp.* HI-31 share a very high sequence identity of 90%.

Computational Methodologies

Molecular Dynamics Simulations

The crystal structures for the $CHMO_{Phi1}$ (WT and its variant) reported here have been used as starting coordinates for MD simulations. Six sets of different MD simulations were performed: WT and variant including the NADPH and FAD cofactor and no substrate (2 sets of MD simulations), plus WT and variant plus Criegee intermediate in eq, eq or ax, ax conformation at the starting geometry (4 sets of MD simulations). Missing residues (496-504) on the mutant were modelled based on the WT geometry.

The Criegee intermediate was built from the crystal structure coordinates of the FAD cofactor, the axial or equatorial conformations refers to the (2*R*,5*R*)-(+)-dihydrocarvone moiety of the molecule. *Ad hoc* parameters were produced for the (2*R*,5*R*)-(+)-dihydrocarvone substrate and Criegee intermediates using the CHARMM General Force Field program (CGenFF)[14, 15] via the ParamChem service. As CGenFF produces only a quick by-analogy parametrization we have no reason to expect good parameters for the Criegee intermediate, especially for the peroxide bridge moiety. For this reason, a dihedral restraint of 180° for the O-O-C-N dihedral (where the oxygen atoms correspond to the peroxide bridge and the N to the FAD moiety) was imposed based on the results of the DFT calculations, see below. Additionally, four distance restraints of 2 Å for the Criegee intermediate simulations were imposed (Scheme S1). These restraints were based on the distances reported in earlier work by Polyak *et. al.*[16] Please notice the distances reported by Polyak *et. al* involve H atoms and the distances used by us involve the corresponding H-bonded heavy atom. Thus, the distance CYHN:O1-Arg329:Hε (Polyak et al.'s nomenclature) = 1.60 Å becomes larger when we use the distance CYHN:O1-Arg329:Nε. As the Nε - Hε bond distance is approximately 1 Å but the angle CYHN:O1-Arg329:Hε-Arg329:Nε is not necessarily 180°, the distance CYHN:O1-Arg329:Nε can range from approximately 1.60 Å to 2.60 Å, depending on the angle adopted. We have restrained the CYHN:O1-Arg329:Nε distance to a medium 2.0 Å distance; the same applies to the other distance restraints used. The simulations were set up using standard simulation protocols with the CHARMM-GUI.[17, 18] Hydrogen coordinates were generated using standard protonation states for all ionizable residues using CHARMM.[19] The system was solvated with a pre-equilibrated TIP3P truncated octahedral water box. Water molecules were randomly replaced by ions to ensure the neutralization of the system, and an additional KCl salt concentration corresponding to 0.15 M. We ran production trajectories of 100 ns after 10 ns of equilibration using NAMD, and for each MD starting configuration, we performed three replicas.[20] Temperature and pressure were held constant at 300 K and 1 atm, respectively. We used the CHARMM36[21] force field with periodic boundary conditions and the particle mesh Ewald method[22] for the long-range electrostatics in combination with a 12 Å cutoff for the evaluation of the nonbonded interactions.



Scheme S1. Distance restraints (dashed lines) used during MD simulations of Criegee intermediate. All distance restraints were set to 2Å. Notice that the distances are between heavy (non-hydrogen) atoms.

DFT Cluster Models

All geometry optimizations used Jaguar (v. 8.5),[23] and the Zhao and Truhlar modified mPW91 + modified B95 (PWB6K) functional.[24] This functional was used in small-molecule studies of the Baeyer-Villiger rearrangement,[25, 26] where it performed well. Single point energies of the enzyme active site (M(med), see below for discussion of active site models used) have also been calculated with the standard B3LYP functional[27-33] as implemented in Jaguar, as well as the dispersion-corrected version (D3[34, 35]) of this functional (B3LYP-D3). In addition, loose convergence criteria (5 times larger than default criteria) were used in the enzyme active site models. Test calculations on other complexes using the more stringent default convergence criteria did not lead to significant changes in energies, bond lengths, or angles, but were much more time-consuming.

All calculations relating to the protein environment used the double zeta quality 6-31+G* basis set to save on computational cost. For the active site model of BVMO, energies including CDM solvent effects were determined from single point calculations on these geometries, incorporating the Poisson–Boltzmann finite-element model of solvation[36, 37] as implemented in Jaguar. The dielectric constant was set to 4, as commonly used in the cluster model approach[38] and the probe radius to 2 Å (labelled as CDM_protein).

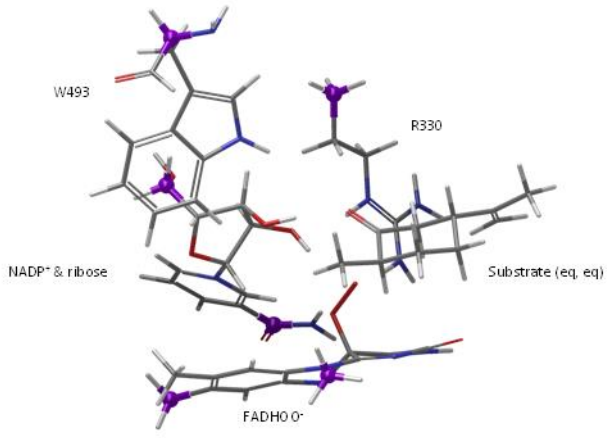


Figure S20. Cluster model (M(med_W) set-up. The carbon atoms constrained to crystallographically-observed positions are shown as purple spheres (2 in $FADHOO^-$, 2 in $NADP^+$ & ribose, 1 in R330, 1 in W493).

We have considered an active site model including substrates, $NADP^+$ and $FADHOO^-$ versions of the cofactors, as well as the side chain of an arginine residue R330 (R329 in Polyak's work), with a protonated guanidium, which, together with the $NADP^+$ ribose hydroxyl groups, contributes to the hydrogen-bonding with the substrates (labelled M(med), 5 Cartesian constraints of residues to crystallographically observed positions). As discussed in the manuscript, this model was extended by inclusion of a tryptophan residue W493 (M(med_W), 6 constraints, Figure S20) and simplified by omission of the R330 residue (M(med_noR), 4 constraints). In addition, a cut-down version of the active site including only the $FADHOO^-$ cofactor (labelled M(sml), 2 constraints, -1 charge) was explored. The overall charge of these active site models is +1 for all apart from M(sml) where it is -1. The discussion below will also consider our attempts at optimising an extended model of the active site, including the residues identified for mutation (phenylalanines F249, F280, F435), as well as the conserved tryptophan (W493), threonine (T436) and leucine (L438) (labelled M(extWT), 11 constraints), and the corresponding variant (phenylalanines to alanines, F249A, F280A, F435A) (labelled M(extMU), 11 constraints).

Unfortunately, we were unable to complete frequency calculations on the DFT models discussed here; structural similarity and the eigenvectors followed during optimizations strongly suggest that these geometries are the correct transition states. Unless otherwise stated, vibrational frequencies were not computed for other stationary points, and so the energetic data do not include a correction for zero-point energy, although we note that this would be expected to be quite small.

The calculations reported here focussed on the active site, rather than the wider protein environment, and we used a small number of constraints to hold residues in the crystallographically-observed positions. This "cluster" approach has been used for the exploration of selectivity in enzyme catalysis with success (see, for example, reference [39]), not in the least because, as long as the key properties of the active site are captured, all isomers considered are affected similarly by the constraints, leading to some helpful cancellation of errors. The effect of the wider protein environment can be introduced by using a continuum dielectric model of solvation as described above, the impact on calculated relative energies was relatively small in the present case.

To position the (2*R*,5*R*)-(+)-dihydrocarvone substrate in the cluster model (M(med)) for the reactant, Criegee intermediate and product geometry, we started with distance constraints as reported by Polyak and co-workers for QM/MM optimizations;[16] once these constrained optimizations had converged, the constraints were removed and the structures re-optimised. Other optimizations started from these geometries, with addition/removal of residues as appropriate. Some structural adjustments were used, due to the different steric properties of the (2*R*,5*R*)-(+)-dihydrocarvone substrate compared to the cyclohexanones considered by Polyak *et al.*,[16] however, these were relatively minor, as shown by the structural data summarised below.

Computational Results and Discussion

Active Site Models

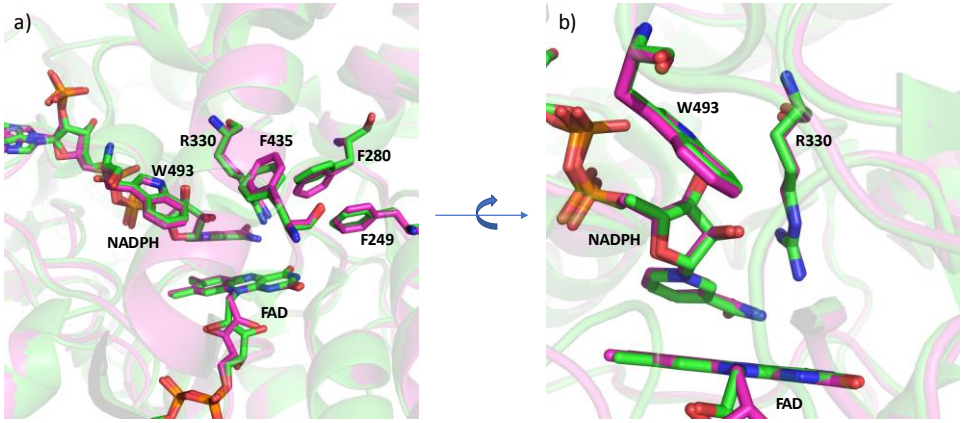


Figure S21. Superposition of $CHMO_{rho}$ structure used in Polyak's QM/MM calculations[16] 3GWD[40] (green) and $CHMO_{Phi1}$ reported in this work (pink). A) FAD, NADPH cofactors, W493, R330 as well as the mutated phenilananines residues (249, 280, 435). b) superposition of residues used in DFT cluster calculations. Notice the geometries of both reported PDB entries are very similar.

Enzyme active site models were generated from the same structure determined by X-ray crystallography described by Polyak and co-workers[16] ($CHMO_{rho}$, *Rhodococcus sp.*, strain HI-31, PDB code 3GWD).[40] An overlay of the active site with the structure of $CHMO_{Phi1}$ was shown to be very similar (Figure S21).

MD Simulation Results

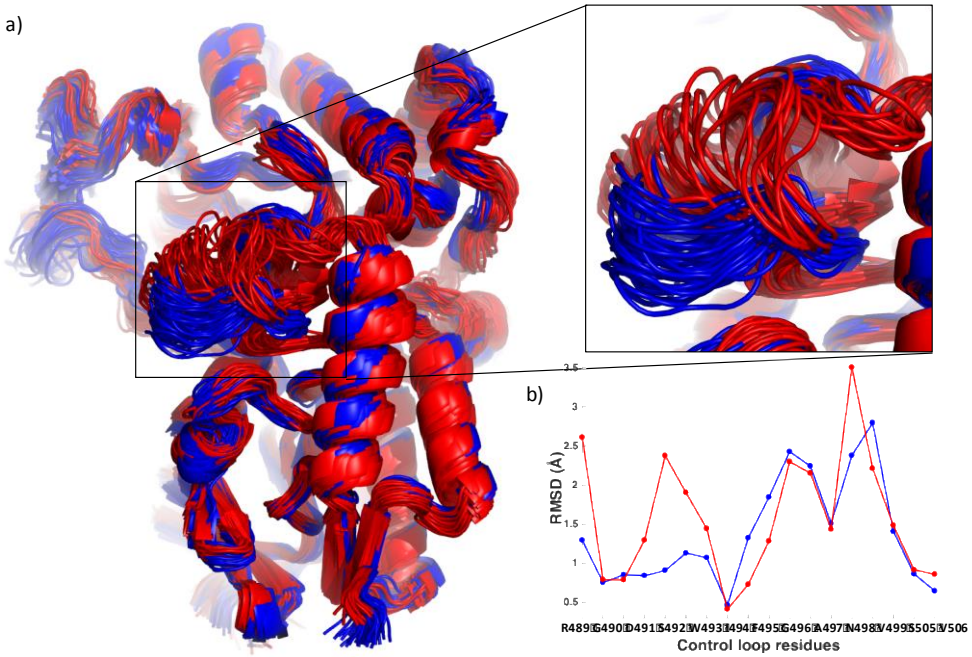


Figure S22. Superposition of MD simulation snapshots collected at uniform intervals showing the WT (blue), variant (red) $CHMO_{Phi1}$**.** a) Zoomed control loop (residues 489 to 506). b) RMSD during the MD simulations of control loop residues.

The MD simulations of the enzyme without substrate (Figure S22) show segments 486 to 506 are more disordered in the variant (RMSD of each amino acid in Figure S22b). A disordered control loop is characteristic of an open conformation which is consistent with the experimental crystallographic data. Figure S23 shows a superposition of the frames along the MD simulations of the protein with the Criegee intermediate. In the case of the WT, the eq, eq conformation is stable during the whole MD simulation (Figure S23a) while the ax, ax conformation changes to eq, eq after 40 ns (Figure S23b; this inversion was observed after 16 ns and after 21 ns in replicas 2 and 3 respectively). This change is mainly due to the strong steric interaction with W493 that is very close to the substrates putative position in the WT. In the case of the variant, both eq, eq and ax, ax conformation can be accommodated in the binding pocket during the whole MD (Figure S23c,d). This is because the active site in the variant is considerably larger and so able to better accommodate the ax, ax conformation. Additionally the control loop is more flexible and the W493 moves away to a different position.

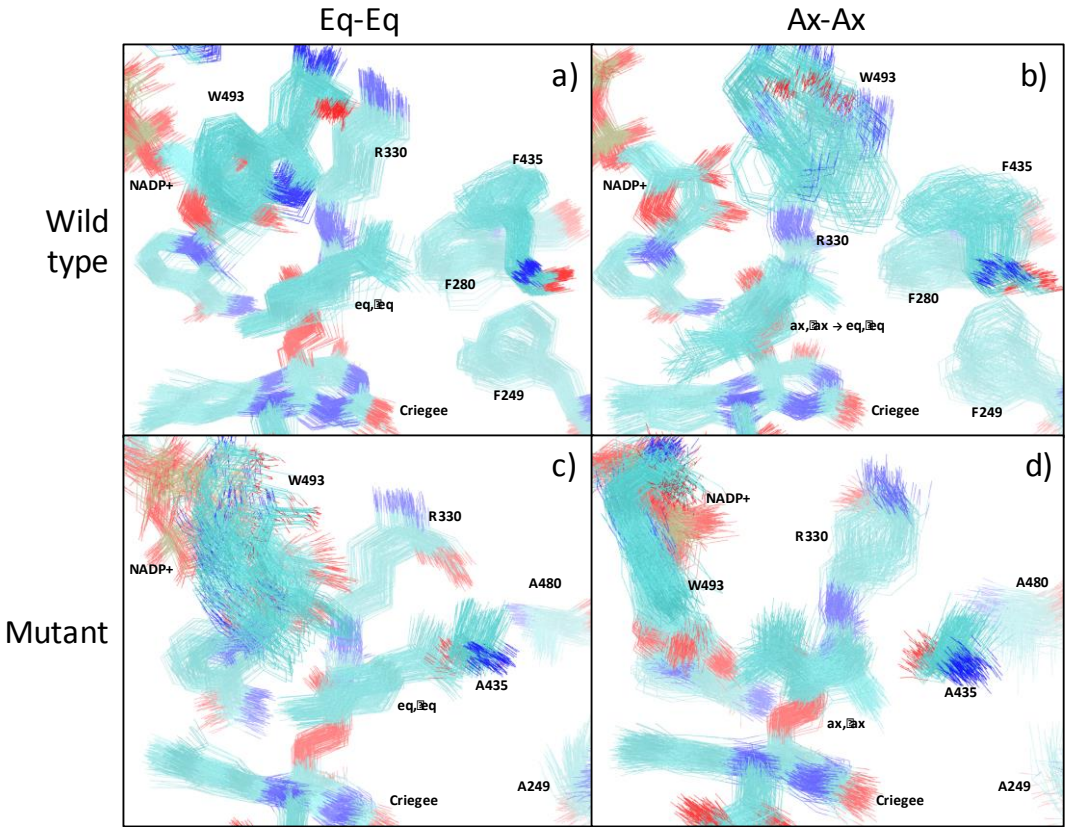


Figure S23. Superposition of MD simulation snapshots collected at uniform intervals showing part of the Criegee intermediate and NADP$^+$, residues 249, 280, R330, 435 and W493 in $CHMO_{Phi1}$**.** A) WT with Criegee intermediate in eq, eq conformation; B) WT with Criegee intermediate in ax, ax conformation; C) variant with Criegee intermediate in eq, eq conformation and D) variantwith Criegee intermediate in ax, ax conformation.

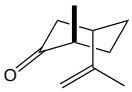
DFT Calculations

Cyclohexanone Monooxgenase (CHMO) Catalysis

For our initial studies of the enzyme-catalysed Baeyer-Villiger oxidation, we based our calculations on the active site model as shown in the MD, QM/MM and QM studies reported by Polyak and co-workers.[16, 41] This model (M(med)) includes the relevant co-factors, including the ribose side-chain of $NADP^+$, as well as an arginine residue adjacent to the site found to hydrogen bond with the substrate (denoted R330 here, but R329 in Polyak et al's work). Both the dihydrocarvone substrate (2*R*,5*R isomer*) and a 2-methyl-substituted cyclohexanone were introduced to the active site and reactant, Criegee intermediate, migration transition state and product complexes were optimised with five cluster model constraints as described above, at the PWB6K/6-31+G* level of theory. The relevant equatorial and axial conformers of the substrates were considered and Table S5 shows the calculated potential energies.

Table S5. Relative potential energies for conformers of (2*R*)-methyl-cyclohexanone and (2*R*,5*R*)-(+)-dihydrocarvone (M(med), PWB6K/6-31+G*, *k*cal mol^{-1}).

eq, eq



ax, ax

	(2*R*)-Me, eq	(2*R*)-Me, ax	(2*R*,5*R*)-(+)-dihydrocarvone, eq, eq	(2*R*,5*R*)-(+)-dihydrocarvone, ax, ax
Reactant	0.0	1.1	0.0	8.5
Criegee Intermediate	7.2	2.1	8.0	9.1
Transition State[a]	25.8 (AL)	24.4 (NL)	25.9 (AL)	28.4 (NL)
Product	-68.2 (AL), -68.9 (NL)	-65.5 (NL), -62.4 (AL)	-67.3 (AL), -69.2 (NL)	-64.0 (NL) , -60.1 (AL)

[a]Only one TS could be located for each conformer.

The energy differences between conformers are much better defined for the (2*R*,5*R*)-(+)-dihydrocarvone substrate than for 2-methyl-cyclohexanone. Note that interactions between the substrate and the active site are stronger in the Criegee intermediate than for the reactant, such that the calculated energy differences between conformers are less reliable/more noisy in the reactant. The eq, eq conformer shows short contacts between the substrate and the ribose attached to the $NADP^+$ co-factor, which destabilise the Criegee intermediate. These results suggest that the overall barrier to reaction is lower for the axial conformer, leading to the normal lactone product. In addition, the active site shows a preference for the normal lactone product in both conformers, although again this may be prone to conformational noise as interactions are weaker than for the Criegee intermediate and transition state.

For the (2*R*,5*R*)-(+)-dihydrocarvone substrate, reactant, Criegee intermediate and transition state energies favour the eq, eq conformer which leads to the abnormal lactone product. For the ax, ax conformer, reactant and Criegee intermediate are again close in energy, while the selectivity here favours the normal lactone. For all product complexes, the normal lactone is energetically favoured in this model.

For the (2*R*,5*R*)-(+)-dihydrocarvone in M(med), method effects were also explored, considering dispersion, free energy and solvation corrections to model the protein environment, as well as exploring the effect of different density functionals. These results are summarised in Table S6. We note that the PWB6K functional was optimised to capture thermochemistry, thermochemical kinetics and nonbonded interactions[24] reasonably well, which should include dispersion effects to some extent.

Table S6. Method effects on calculated energies (*k*cal moL^{-1}) for M(med) cluster model.

	PWB6K ΔE (opt)	PWB6K ΔE, sl (eps=4) (SP[a])	B3LYP ΔE (SP[a])	B3LYP-D3 ΔE (SP[a])	B3lYP-D3 ΔE (opt)
eq, eq					
Reactant	0.0	0.0	0.0	0.0	0.0
Criegee IM	8.0	5.5	18.2	9.5	9.6
Transition State[b]	25.9	24.2	23.5	14.5	15.5
Reactant	-67.3 (AL), -69.2 (NL)	-67.1 (AL), -66.2 (NL)	-65.4 (AL), -65.5 (NL)	-62.7 (AL), -61.6 (NL)	-61.5 (AL), -63.4 (NL)
ax, ax					
Reactant	8.5	8.4	10.5	7.9	7.8
Criegee IM	9.1	8.8	20.3	10.8	11.1
Transition State[a]	28.4 (19.9)[c]	27.7 (19.3) [c]	25.6 (15.1) [c]	18.8 (10.9) [c]	18.6 (10.8)[c]
Reactant	-64.0 (NL) , -60.9 (AL)	-63.0 (NL), -61.4 (AL)	-60.9 (NL), -58.4 (AL)	-60.0 (NL), -57.3 (AL)	-58.8 (NL), -56.0 (AL)

[a]Using PWB6K optimised geometry. [b]Only one TS could be located for each conformer. [c] Barrier relative to ax, ax reactant.

While changes to the computational approach affect the calculated barriers, especially when dispersion corrections are used with the B3LYP functional, they do not substantially alter the predicted selectivities in this case. The active site reactant model is consistently favouring the eq, eq conformer, and once the substrate is in the active site, the Criegee intermediate also favours the eq, eq conformer whereas the migration barrier is lower for the ax, ax conformer. Both product geometries can be accommodated in the active site, but only transition states leading to the experimentally observed lactone product could be located.

Barriers to this reaction are lower for the dispersion-corrected B3LYP-D3 functional, and these are also in better agreement with the QM/MM data reported by Polyak and co-workers[16] for the cyclohexanone substrate. While dispersion corrections have been found unlikely to affect relative energies significantly in Polyak et al.'s work,[16] here they seem to improve on results obtained with other functionals, even though the PWB6K functional is designed to capture nonbonded interactions well. Note, however, that the attractive interactions within the active sites modelled could perhaps also be exaggerated in cluster models such as this when a dispersion correction is included, as the "balancing interactions" with the wider protein and solvent environment are not likely to be fully captured by thecontinuum model of solvation, so the two sets of optimizations with different

functionals may provide a guide to the likely range of values. Most importantly, our interpretation looks to be robust to changes in the DFT approach used.

Residue W493 is important for interactions between the substrate and the residues in the active site, and has thus been included in M(med_W), with M(med) a possible model for the variant where this residue moves away from the reacting site. A number of different changes can be observed in the crystal structure geometry and MD simulations for the variant and we have explored some of these by modifying the cluster models in DFT calculations. Thus, the increased conformational freedom of R330 might lead it to move away from the reacting site, this has been probed with M(med_noR), which does not include W493 or R330. The $NADP^+$ co-factor might also become more mobile, with M(sml) considering interaction with only the $FADHOO^-$ moiety. Table S7 compares the calculated potential energies for these variations with the data for M(med).

Table S7. Relative potential energies for conformers of (*2R,5R*)-(+)-dihydrocarvone, PWB6K/6-31+G*, *k*cal moL^{-1}

	M(sml)	M(med_noR)	M(med)	M(med_W)
eq, eq				
Reactant	0.0	0.0	0.0	0.0
Criegee IM	8.4	-1.2	8.0	6.8
Migration TS[a] (AL)	26.7	19.3	25.9	25.6
Product (AL)	-62.7	-69.8	-67.3	-66.7
ax, ax				
Reactant	5.0	1.9[b]	8.5	3.5
Criegee IM	10.2	0.9	9.1	11.6
Migration TS[a] (NL)	29.5 (24.5)[c]	24.4 (22.5)[c]	28.4 (19.9)[c]	35.3 (31.8)[c]
Product (NL)	-72.7	-67.7	-64.0	-61.7

[a]Only one TS could be located for each conformer. [b]Change in O2 orientation due to loss of H-bonding. [c]Barrier relative to ax, ax reactant.

Comparison of M(med) and M(med_W) showed interactions with the tryptophan residue probably destabilise the eq, eq conformer somewhat, reducing the energy difference between the two conformers in the reactant complexes. For these cluster models, substrate and product lactone are held quite loosely by a range of hydrogen bonds in the protein active site, making these energies more prone to be affected by computational noise, slight differences in starting geometries and the consideration of additional residues (see below) than the Criegee intermediates and migration transition states. This could affect M(sml) and M(med_noR) energetics here and Table S8 summarises the barriers to reaction from each Criegee intermediate to the migration transition state. Tables S9-S14 and Figure S24 show the values for the structural parameters.

Table S8. Barriers to reaction from Criegee intermediates as indicated for conformers of (2*R*,5*R*)-(+)-dihydrocarvone, PWB6K/6-31+G*, kcal mol^{-1}.

	M(sml)	**M(med_noR)**	**M(med)**	**M(med_W)**
eq, eq	18.3	20.5	17.9	18.8
ax, ax	19.3	23.5	19.3	23.7

These results suggest that FADHOO$^-$ alone is catalytically competent once a Criegee intermediate has been formed; there may be other factors such as substrate interactions with the active site which are compromised by these simplifications, but the barriers to reaction are relatively consistent, if a bit high with the DFT approach used here. There are some differences between the reactivity of the two conformers, but these are small and may stretch the reliability of both the cluster approach and the computational methodology and software used here. In addition, structural analysis of M(sml) for the ax, ax conformers shows substrate rotation into an area normally occupied by the NADP$^+$ co-factor, while for M(med_noR), some movement into the site of the arginine residue occurs. These changes suggest that more extensive sampling to fully understand the dynamic behaviour of the active site in variant and WT are necessary before such energy differences can be considered reliable; this lies outside of the scope of the present study.

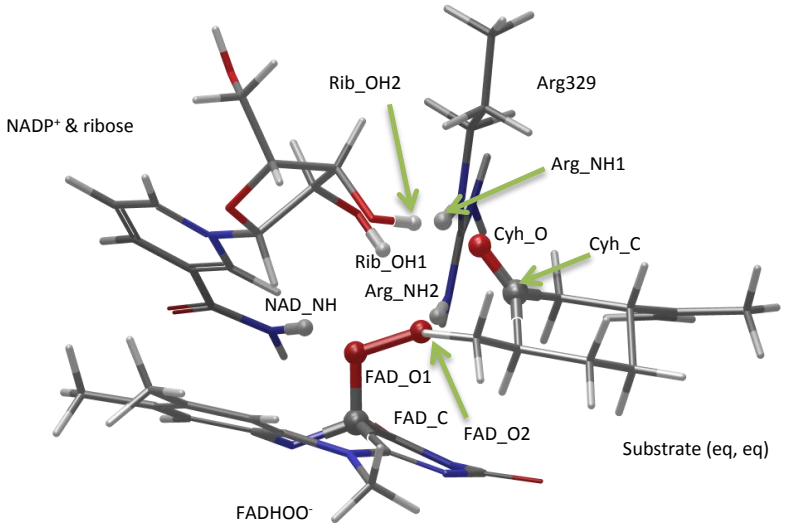


Figure S24. Modelled structure, indicating atom numbering for Tables S9- S14.

Table S9. Key structural parameters from calculated geometries for PWB6K with substrate (2*R*,5*R*)-(+)-dihydrocarvone . A) M(med_W)

Interaction	**Reactant**	**Criegee IM**	**TS**	**Product**
eq, eq			(AL)	(AL)
rel. E (kcal mol^{-1})	0.0	6.8	25.6	-66.7
FAD_C – FAD_O1	1.411	1.438	1.384	1.363
FAD_O1 – FAD_O2	1.429	1.426	1.837	4.726
FAD_O2 – Cyh_C	2.919	1.514	1.341	1.316
Cyh_C – Cyh_O	1.217	1.326	1.299	1.217
NAD_NH – FAD_O1	1.975	2.604	2.362	1.925
Arg_NH1 – Cyh_O	3.167	1.800	1.823	2.061
Arg_NH2 – FAD_O2	1.593	1.879	1.935	4.128
Rib_OH1 – Cyh_O	1.820	1.824	1.898	1.986
Rib_OH2 – Cyh_O	2.527	1.702	1.668	2.468
Rib_OH2 – Arg_NH1	2.227	2.005	2.149	2.616
shortest contact with W493	3.524 (NH-MeH)	2.482 (CH-CyhH)	2.678 (CH-MeH)	3.638 (CH-MeH)
ax, ax			(NL)	(NL)
rel. E (kcal mol^{-1})	3.5	11.6	35.3	-61.7
FAD_C – FAD_O1	1.402	1.435	1.376	1.361
FAD_O1 – FAD_O2	1.434	1.425	1.826	5.113
FAD_O2 – Cyh_C	3.212	1.522	1.337	1.314
Cyh_C – Cyh_O	1.219	1.325	1.290	1.218
NAD_NH – FAD_O1	2.194	2.474	2.266	1.843
Arg_NH1 – Cyh_O	3.152	1.767	1.726	1.846
Arg_NH2 – FAD_O2	1.566	2.450	1.860	3.008
Rib_OH1 – Cyh_O	1.757	1.936	2.582	2.678
Rib_OH2 – Cyh_O	2.596	1.589	1.704	2.212
Rib_OH2 – Arg_NH1	2.232	2.057	2.216	2.641
shortest contact with W493	2.409 (CH-propMeH)	2.238 (CH-propMeH)	2.294 (CH-propMeH)	2.223 (NH_propCH)

Table S10. Key structural parameters from calculated geometries for PWB6K with substrate (2*R*,5*R*)-(+)-dihydrocarvone. B) M(med)

Interaction	Reactant	Criegee IM	TS	Product (NL)	Product (AL)
eq, eq			(AL)	(NL)	(AL)
rel. E (kcal mol^{-1})	0.0	8.0	25.9	-69.2	-67.3
FAD_C – FAD_O1	1.406	1.436	1.379	1.357	1.360
FAD_O1 – FAD_O2	1.433	1.426	1.866	4.551	4.223
FAD_O2 – Cyh_C	2.940	1.510	1.333	1.316	1.318
Cyh_C – Cyh_O	1.216	1.326	1.295	1.217	1.216
NAD_NH – FAD_O1	2.003	2.425	2.303	1.847	1.857
Arg_NH1 – Cyh_O	2.918	1.789	1.823	1.918	2.400
Arg_NH2 – FAD_O2	1.555	1.971	1.939	4.936	3.516
Rib_OH1 – Cyh_O	1.827	1.876	1.902	2.528	1.894
Rib_OH2 – Cyh_O	2.215	1.641	1.672	1.960	1.980
Rib_OH2 – Arg_NH1	2.287	2.155	2.779	2.477	2.560
ax, ax			(NL)	(NL)	(AL)
rel. E (kcal mol^{-1})	8.5	9.1	28.4	-64.0	-60.1
FAD_C – FAD_O1	1.401	1.432	1.380	1.360	1.358
FAD_O1 – FAD_O2	1.433	1.426	1.811	5.188	4.054
FAD_O2 – Cyh_C	2.820	1.518	1.339	1.315	1.301
Cyh_C – Cyh_O	1.220	1.323	1.299	1.217	1.223
NAD_NH – FAD_O1	1.944	2.352	2.209	1.810	1.850
Arg_NH1 – Cyh_O	2.108	1.719	1.715	1.843	1.957
Arg_NH2 – FAD_O2	1.500	1.891	1.892	4.443	4.351
Rib_OH1 – Cyh_O	2.294	2.116	2.232	3.037	2.224
Rib_OH2 – Cyh_O	1.789	1.591	1.620	2.076	1.932
Rib_OH2 – Arg_NH1	2.448	2.128	2.120	2.618	2.353

Table S11. Key structural parameters from calculated geometries for PWB6K with substrate (2*R*,5*R*)-(+)-dihydrocarvone. C) M(med_noR)

Interaction	Reactant	Criegee IM	TS	Product
eq, eq			(AL)	(AL)
rel. E (kcal mol^{-1})	0.0	-1.2	19.3	-69.8
FAD_C – FAD_O1	1.398	1.421	1.371	1.360
FAD_O1 – FAD_O2	1.439	1.422	1.823	4.223
FAD_O2 – Cyh_C	2.847	1.520	1.339	1.318
Cyh_C – Cyh_O	1.214	1.316	1.287	1.216
NAD_NH – FAD_O1	2.167	2.237	2.470	1.857
Rib_OH1 – Cyh_O	1.773	1.761	1.679	1.894
Rib_OH2 – Cyh_O	2.765	1.618	1.614	1.980
Rib_OH2 – FAD_O2	1.642	3.592	3.203	3.662
ax, ax			(NL)	(NL)
rel. E (kcal mol^{-1})	1.9	0.9	24.4	-67.7
FAD_C – FAD_O1	1.395	1.419	1.371	1.360
FAD_O1 – FAD_O2	1.438	1.416	1.807	5.188
FAD_O2 – Cyh_C	3.263[a]	1.511	1.341	1.315
Cyh_C – Cyh_O	1.221	1.320	1.286	1.217
NAD_NH – FAD_O1	1.758	2.168	2.283	1.810
Rib_OH1 – Cyh_O	1.849	1.848	1.817	3.037
Rib_OH2 – Cyh_O	1.836	1.565	1.585	2.076
Rib_OH2 – FAD_O2	3.187	3.198	3.020	4.186

[a]O_2 rotation away from substrate.

Table S12. Key structural parameters from calculated geometries for PWB6K with substrate (2*R*,5*R*)-(+)-dihydrocarvone. D) M(sml)

Interaction	Reactant	Criegee IM	TS	Product
eq, eq			(AL)	(AL)
rel. E (kcal mol^{-1})	0.0	8.4	26.7	-62.7
FAD_C – FAD_O1	1.366	1.388	1.351	1.317
FAD_O1 – FAD_O2	1.447	1.429	1.804	3.971
FAD_O2 – Cyh_C	3.725[a]	1.652	1.372	1.336
Cyh_C – Cyh_O	1.212	1.257	1.253	1.201
ax, ax			(NL)	(NL)
rel. E (kcal mol^{-1})	5.0	10.2	29.5	-72.7
FAD_C – FAD_O1	1.368	1.387	1.350	1.318
FAD_O1 – FAD_O2	1.446	1.421	1.785	4.621
FAD_O2 – Cyh_C	4.017[a]	1.621	1.362	1.430
Cyh_C – Cyh_O	1.212	1.265	1.244	1.205

[a] Substrate has moved "above" O_2.

Table S13. Key structural parameters from calculated geometries for B3LYP-D3 with substrate (2*R*,5*R*)-(+)-dihydrocarvone.

Interaction	Reactant	Criegee IM	TS	Product (NL)	Product (AL)
eq, eq			(AL)	(NL)	(AL)
rel. E (kcal mol^{-1})	0.0	9.6	15.5	-63.4	-61.5
FAD_C – FAD_O1	1.440	1.468	1.411	1.377	1.382
FAD_O1 – FAD_O2	1.466	1.476	1.846	4.441	4.120
FAD_O2 – Cyh_C	2.906	1.551	1.371	1.333	1.337
Cyh_C – Cyh_O	1.233	1.338	1.325	1.233	1.230
NAD_NH – FAD_O1	1.951	2.429	2.349	1.828	1.846
Arg_NH1 – Cyh_O	3.147	1.729	1.777	1.859	2.447
Arg_NH2 – FAD_O2	1.577	1.866	1.833	4.841	3.459
Rib_OH1 – Cyh_O	1.776	1.867	1.892	2.526	1.869
Rib_OH2 – Cyh_O	2.493	1.625	1.625	1.905	1.939
Rib_OH2 – Arg_NH1	2.226	2.080	2.135	2.449	2.588
ax, ax			(NL)	(NL)	(AL)
rel. E (kcal mol^{-1})	7.8	11.1	18.6	-58.8	-56.0
FAD_C – FAD_O1	1.434	1.465	1.406	1.378	1.376
FAD_O1 – FAD_O2	1.467	1.472	1.843	5.075	3.987
FAD_O2 – Cyh_C	2.827	1.561	1.361	1.333	1.319
Cyh_C – Cyh_O	1.239	1.335	1.318	1.231	1.239
NAD_NH – FAD_O1	1.996	2.302	2.178	1.799	1.842
Arg_NH1 – Cyh_O	2.380	1.695	1.771	1.787	1.907
Arg_NH2 – FAD_O2	1.421	1.815	1.769	4.327	4.276
Rib_OH1 – Cyh_O	1.960	2.134	2.202	3.249	2.191
Rib_OH2 – Cyh_O	1.807	1.570	1.632	2.030	1.911
Rib_OH2 – Arg_NH1	2.397	2.094	2.115	2.589	2.346

Table S14. Key structural parameters from calculated geometries for PWB6K with substrate 2-methyl-cyclohexanone.

Interaction	Reactant	Criegee IM	TS	Product (NL)	Product (AL)
eq			(AL)	(NL)	(AL)
rel. E (kcal mol^{-1})	0.0	7.2	25.8	-68.9	-68.2
FAD_C – FAD_O1	1.409	1.436	1.379	1.358	1.361
FAD_O1 – FAD_O2	1.432	1.425	1.869	4.580	4.290
FAD_O2 – Cyh_C	2.932	1.508	1.331	1.314	1.316
Cyh_C – Cyh_O	1.218	1.330	1.291	1.217	1.218
NAD_NH – FAD_O1	2.008	2.512	2.321	1.867	1.881
Arg_NH1 – Cyh_O	2.916	1.713	1.831	1.892	2.141
Arg_NH2 – FAD_O2	1.589	1.900	1.935	1.714	1.761
Rib_OH1 – Cyh_O	1.833	1.853	1.902	2.456	1.992
Rib_OH2 – Cyh_O	2.476	1.660	1.674	1.983	2.099
Rib_OH2 – Arg_NH1	2.249	2.023	2.161	2.446	2.492
ax			(NL)	(NL)	(AL)
rel. E (kcal mol^{-1})	1.1	2.1	24.4	-65.5	-62.4
FAD_C – FAD_O1	1.408	1.435	1.381	1.359	1.356
FAD_O1 – FAD_O2	1.431	1.425	1.813	4.974	4.647
FAD_O2 – Cyh_C	3.357	1.495	1.336	1.311	1.317
Cyh_C – Cyh_O	1.220	1.332	1.302	1.221	1.216
NAD_NH – FAD_O1	1.983	2.423	2.214	1.907	1.860
Arg_NH1 – Cyh_O	3.570	1.677	1.725	1.958	1.956
Arg_NH2 – FAD_O2	1.571	1.877	1.885	4.383	4.955
Rib_OH1 – Cyh_O	1.768	1.924	2.030	1.980	2.476
Rib_OH2 – Cyh_O	2.797	1.580	1.623	2.239	2.031
Rib_OH2 – Arg_NH1	2.256	2.100	1.381	1.359	1.356

These data suggest that different versions of the DFT cluster model used are quite well-defined and able to retain a range of short contacts with the substrate/product molecules throughout all steps, supporting the catalysis. The two substrate conformers lead to slight differences in their interactions with the residues modelled, likely dominated by the steric bulk of the propenyl substituent for the dihydrocarvone. These are best illustrated by inspection of Figure S20 which illustrates differences between the key species for M(med).

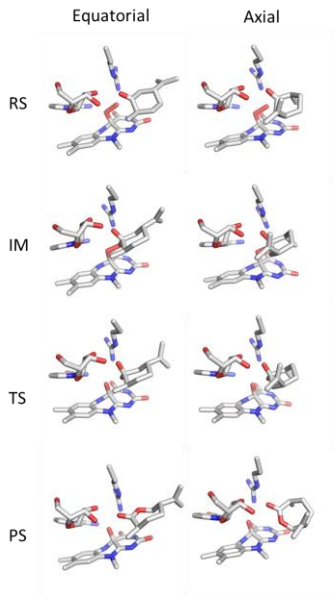


Figure S25. Comparison of conformers in active site region.

Transition states were initially located for a cut-down model (M(sml) containing only the $FADHOO^-$ cofactor and not the $NADP^+$ or the relevant arginine residue. Despite these simplifications, transition states could not be found for the less-favourable product in either case; some transition states involved rotation of the substrate relative to the co-factor's plane, which could not be accommodated in the active site due to close contacts with other residues present and perhaps suggest that selectivity arises because this site is well-defined. For both Criegee intermediate and transition state, the eq, eq conformer is favoured due to fewer short contacts between the methyl substituent and the cofactor, twisting the substrate away from favourable alignment for the migration step.

The structural analysis of WT and variant $CHMO_{Phi1}$ suggested that additional residues "behind" the active site play an important role in determining the experimentally observed selectivity, as discussed in the manuscript. We experimented with calculations on such an extended cluster model, initially including the W493 in both WT and variant, as well as the mutated residues and additional residues from the control loop. This gave a model containing 240 atoms (2384 basis functions) for the WT (M(extWT) and 210 atoms (2036 basis functions) for the variant (M(extMU)), with 11 constraints used in each case. At this stage of our study, we did not have a full variant crystal structure geometry, so used the tryptophan position as observed in the WT for W493.

Optimizations of these extended models proved possible for both reactants and Criegee intermediates, but were found to be very sensitive to starting geometry and the constraints used, as well as very slow to converge. In the WT, the phenylalanine residues were the likely source of this sensitivity, while the interactions between ax, ax substrate and some of the mutated residues appeared important for the variant model. While these calculations showed a switch in energy preference for the Criegee intermediates, in line with experimental observations, the reactant optimizations appeared less consistent, giving quite variable energies and structures in response to minor changes in the starting geometries. In view of the MD simulation results reported here, which also suggest considerable flexibility in this region, we judged our current static cluster models too unreliable for selectivity predictions and have thus not attempted transition state optimizations to evaluate the overall reaction barriers. A more complete exploration of selectivity, combining MD sampling and QM/MM calculations from multiple starting configurations, lies beyond the scope of this study.

References

[1] Fang, H. J., Lai, P. S., Chen, J. Y., Hsu, S. C. N., Peng, W. D., Ou, S. W., Lai, Y. C., Chen, Y. J., Chung, H., Chen, Y., Huang, T. C., Wu, B. S., and Chen, H. Y. (2012) ε-Caprolactone polymerization under air by the biocatalyst: magnesium 2,6-di-tert-butyl-4-methylphenoxide, *J Polym Sci Pol Chem 50*, 2697-2704.

[2] Calabrese, J., Cushing, M. A., and Ittel, S. D. (1988) Sterically hindered magnesium aryloxides, *Inorg Chem 27*, 867-870.

[3] Wilson, J. A., Hopkins, S. A., Wright, P. M., and Dove, A. P. (2015) Synthesis and postpolymerization modification of one-pot ω-Pentadecalactone block-like copolymers, *Biomacromolecules 16*, 3191-3200.

[4] Mirza, I. A., Yachnin, B. J., Wang, S., Grosse, S., Bergeron, H., Imura, A., Iwaki, H., Hasegawa, Y., Lau, P. C., and Berghuis, A. M. (2009) Crystal structures of cyclohexanone monooxygenase reveal complex domain movements and a sliding cofactor, *J Am Chem Soc 131*, 8848-8854.

[5] Yachnin, B. J., Sprules, T., McEvoy, M. B., Lau, P. C., and Berghuis, A. M. (2012) The substrate-bound crystal structure of a Baeyer-Villiger monooxygenase exhibits a Criegee-like conformation, *J Am Chem Soc 134*, 7788-7795.

[6] Yachnin, B. J., McEvoy, M. B., MacCuish, R. J., Morley, K. L., Lau, P. C., and Berghuis, A. M. (2014) Lactone-bound structures of cyclohexanone monooxygenase provide insight into the stereochemistry of catalysis, *ACS Chem. Biol. 9*, 2843-2851.

[7] Malito, E., Alfieri, A., Fraaije, M. W., and Mattevi, A. (2004) Crystal structure of a Baeyer-Villiger monooxygenase, *Proc. Natl. Acad. Sci. U. S. A. 101*, 13157-13162.

[8] Orru, R., Dudek, H. M., Martinoli, C., Torres Pazmino, D. E., Royant, A., Weik, M., Fraaije, M. W., and Mattevi, A. (2011) Snapshots of enzymatic Baeyer-Villiger catalysis: oxygen activation and intermediate stabilization, *J Biol Chem 286*, 29284-29291.

[9] Leisch, H., Shi, R., Grosse, S., Morley, K., Bergeron, H., Cygler, M., Iwaki, H., Hasegawa, Y., and Lau, P. C. (2012) Cloning, Baeyer-Villiger biooxidations, and structures of the camphor pathway 2-oxo-Delta(3)-4,5,5-trimethylcyclopentenylacetyl-coenzyme - a monooxygenase of *Pseudomonas putida* ATCC 17453, *Appl. Environ. Microbiol. 78*, 2200-2212.

[10] Franceschini, S., van Beek, H. L., Pennetta, A., Martinoli, C., Fraaije, M. W., and Mattevi, A. (2012) Exploring the structural basis of substrate preferences in Baeyer-Villiger monooxygenases: insight from steroid monooxygenase, *J. Biol. Chem. 287*, 22626-22634.

[11] Ferroni, F. M., Tolmie, C., Smit, M. S., and Opperman, D. J. (2016) Structural and catalytic characterization of a fungal Baeyer-Villiger monooxygenase, *PLoS One 11*, e0160186.

[12] Romero, E., Castellanos, J. R., Mattevi, A., and Fraaije, M. W. (2016) Characterization and crystal structure of a robust cyclohexanone monooxygenase, *Angew. Chem., Int. Ed. Engl. 55*, 15852-15855.

[13] Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D, *Nucleic Acids Res. 38*, W545-549.

[14] Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., and Mackerell, A. D. (2010) CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields, *J. Comput. Chem. 31*, 671-690.

[15] Gurusamy-Thangavelu, S. A., Emond, S. J., Kulshrestha, A., Hillmyer, M. A., Macosko, C. W., Tolman, W. B., and Hoye, T. R. (2012) Polyurethanes based on renewable polyols from bioderived lactones, *Polym Chem-Uk 3*, 2941-2948.

[16] Polyak, I., Reetz, M. T., and Thiel, W. (2012) Quantum mechanical/molecular mechanical Study on the mechanism of the enzymatic Baeyer–Villiger reaction, *J. Am. Chem. Soc. 134*, 2732-2741.

[17] Jo, S., Kim, T., Iyer, V. G., and Im, W. (2008) CHARMM-GUI: A web-based graphical user interface for CHARMM, *J. Comput. Chem. 29*, 1859-1865.

[18] Balke, K., Schmidt, S., Genz, M., and Bornscheuer, U. T. (2016) Switching the regioselectivity of a cyclohexanone monooxygenase toward (+)-trans-dihydrocarvone by rational protein design, *ACS Chem. Biol. 11*, 38-43.

[19] Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009) CHARMM: The biomolecular simulation program, *J. Comput. Chem. 30*, 1545-1614.

[20] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD, *J. Comput. Chem. 26*, 1781-1802.

[21] Huang, J., and MacKerell, A. D. (2013) CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data, *J. Comput. Chem. 34*, 2135-2145.

[22] Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems, *J. Chem. Phys. 98*, 10089-10092.

[23] Schrödinger-Inc. (2014) Jaguar, 8.5 ed., New York, NY.

[24] Zhao, Y., and Truhlar, D. G. (2005) Design of density functionals that are broadly accurate for thermochemistry, thermochemical kinetics, and nonbonded interactions, *J. Phys. Chem. A 109*, 5656-5667.
[25] Alvarez-Idaboy, J. R., Reyes, L., and Mora-Diez, N. (2007) The mechanism of the Baeyer-Villiger rearrangement: quantum chemistry and TST study supported by experimental kinetic data, *Org. Biomol. Chem. 5*, 3682-3689.
[26] Itoh, Y., Yamanaka, M., and Mikami, K. (2013) Theoretical study on the regioselectivity of Baeyer–Villiger reaction of α-Me-, -F-, -CF3-cyclohexanones, *J. Org. Chem. 78*, 146-153.
[27] Slater, J. C. (1974) *Quantum theory of molecules and solids, Vol. 4: The self-consistent field for molecules and solids*, McGraw-Hill, New York.
[28] Vosko, S. H., Wilk, L., and Nusair, M. (1980) Accurate spin-dependent electron liquid correlation energies for local spin-density calculations - a critical analysis, *Can. J. Phys. 58*, 1200-1211.
[29] Becke, A. D. (1988) Density-functional exchange-energy approximatin with correct asymptotic-behaviour, *Phys. Rev. A 38*, 3098-3100.
[30] Lee, C. T., Yang, W. T., and Parr, R. G. (1988) Development of the Colle-Salvetty correlation-energy formula into a functional of the electron-density, *Phys. Rev. B 37*, 785-789.
[31] Miehlich, B., Savin, A., Stoll, H., and Preuss, H. (1989) Results obtained with the correlation-energy density functionals of Becke and Lee, Yang and Parr, *Chem. Phys. Lett. 157*, 200-206.
[32] Becke, A. D. (1993) Density-functional thermochemistry. 3. The role of exact exchange., *J. Chem. Phys. 98*, 5648-5652.
[33] Stephens, P. J., Devlin, F. J., Chabalowski, C. F., and Frisch, M. J. (1994) Ab-initio calculation of vibrational absorption and circular-dichroism spectra using density-functional force-fields, *J Phys Chem-Us 98*, 11623-11627.
[34] Goerigk, L., and Grimme, S. (2011) A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions, *Phys. Chem. Chem. Phys. 13*, 6670-6688.
[35] Grimme, S., Antony, J., Ehrlich, S., and Krieg, H. (2010) A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu, *J. Chem. Phys. 132*, 154104.
[36] Tannor, D. J., Marten, B., Murphy, R., Friesner, R. A., Sitkoff, D., Nicholls, A., Honig, B., Ringnalda, M., and Goddard, W. A. (1994) Accurate first principles calculation of molecular charge distributions and solvation energies from Ab Initio Quantum Mechanics and Continuum Dielectric theory, *J. Am. Chem. Soc. 116*, 11875-11882.
[37] Marten, B., Kim, K., Cortis, C., Friesner, R. A., Murphy, R. B., Ringnalda, M. N., Sitkoff, D., and Honig, B. (1996) New model for calculation of solvation free energies: correction of self-consistent reaction field Continuum Dielectric theory for short-range hydrogen-bonding effects, *J. Phys. Chem. 100*, 11775-11788.
[38] Himo, F. (2017) Recent trends in quantum chemical modeling of enzymatic reactions, *J. Am. Chem. Soc. 139*, 6780-6786.
[39] Lind, M. E. S., and Himo, F. (2013) Quantum chemistry as a tool in asymmetric biocatalysis: limonene epoxide hydrolase test case, *Angew. Chem., Int. Ed. Engl. 52*, 4563-4567.
[40] Mirza, I. A., Yachnin, B. J., Wang, S., Grosse, S., Bergeron, H., Imura, A., Iwaki, H., Hasegawa, Y., Lau, P. C. K., and Berghuis, A. M. (2009) Crystal structures of cyclohexanone monooxygenase reveal complex domain movements and a sliding cofactor, *J. Am. Chem. Soc. 131*, 8848-8854.
[41] Polyak, I., Reetz, M. T., and Thiel, W. (2013) Quantum mechanical/molecular mechanical study on the enantioselectivity of the enzymatic Baeyer–Villiger reaction of 4-hydroxycyclohexanone, *J. Phys. Chem. B 117*, 4993-5001.