

SUPPORTING INFORMATION

Supporting Information

High Throughput Screening with SAMDI Mass Spectrometry for Directed Evolution

Adam J. Pluchinsky, Daniel J. Wackelin, Xiongyi Huang, Frances H. Arnold and Milan Mrksich

TABLE OF CONTENTS

I. General Methods	S3–S6
II. Supplementary Figs. S1–S8 and Table S1	S7–S16
III. Screening of enzymes for C–H alkylation activity	S17
IV. Directed evolution of C–H alkylation enzymes	S17
V. Nucleotide and amino acid sequences	S17–S20
VI. Chemical synthesis and characterization	S21–S23
VII. Small scale enzymatic reactions and product calibration curves	S24–S25
VIII. Determination of enantioselectivity	S26
IX. NMR Spectra	S27–S28
X. Supplemental references	S29

I. General methods

(A) General. Unless otherwise noted, all chemicals and reagents were obtained from commercial suppliers (Sigma-Aldrich, VWR, Alfa Aesar, Combi-Blocks) and used without further purification. Disulfides used to form self-assembled monolayers were purchased from ProChimia Surfaces. Silica gel chromatography was carried out using AMD Silica Gel 60, 230-400 mesh. ^1H and ^{13}C NMR spectra were recorded on a Varian Inova 300 MHz, Varian Inova 500 MHz, or Bruker Prodigy 400 MHz instrument in CDCl_3 and are referenced to residual protio solvent signals. Data for ^1H NMR are reported as follows: chemical shift (δ ppm), multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, p = pentet, sext = sextet, m = multiplet, dd = doublet of doublets, dt = doublet of triplets, ddd = doublet of doublet of doublets), coupling constant (Hz), integration. Sonication was performed using a Qsonica Q500 sonicator. High-resolution mass spectra were obtained at the California Institute of Technology Mass Spectral Facility. Synthetic reactions were monitored by thin layer chromatography (TLC, Merck 60 gel plates) using a UV-lamp or an appropriate TLC stain for visualization. *E. coli* cells were grown using Luria-Bertani medium (LB) or Hyperbroth (AthenaES) (HB) with 0.1 mg/mL ampicillin (LBamp or HBamp). Primer sequences are available upon request. T5 exonuclease, Phusion polymerase, and Taq ligase were purchased from New England Biolabs (NEB, Ipswich, MA). M9-N minimal medium (abbreviated as M9-N buffer; pH 7.4) was used as a buffering system for whole cells and lysates, unless otherwise specified. M9-N buffer was used without a nitrogen source; it contains 47.7 mM Na_2HPO_4 , 22.0 mM KH_2PO_4 , 8.6 mM NaCl, 2.0 mM MgSO_4 , and 0.1 mM CaCl_2 .

E. coli cells were grown using Luria-Bertani medium (LB) or Hyperbroth (AthenaES) (HB) with 0.1 mg/mL ampicillin (LBamp or HBamp). Primer sequences are available upon request. T5 exonuclease, Phusion polymerase, and Taq ligase were purchased from New England Biolabs (NEB, Ipswich, MA). M9-N minimal medium (abbreviated as M9-N buffer; pH 7.4) was used as a buffering system for whole cells and lysates, unless otherwise specified. M9-N buffer was used without a nitrogen source; it contains 47.7 mM Na_2HPO_4 , 22.0 mM KH_2PO_4 , 8.6 mM NaCl, 2.0 mM MgSO_4 , and 0.1 mM CaCl_2 .

(B) Chromatography. Chiral GC was conducted using an Agilent 7820A instrument (FID) and an Agilent CycloSil-B column (30 m \times 0.32 mm, 0.25 μm film). Gas chromatography-mass spectrometry (GC-MS) analyses were carried out using a Shimadzu GCMS-QP2010SE system and J&W HP-5ms column (30 m \times 0.25 mm, 0.25 μm film).

(C) Cloning and error prone mutagenesis. pET22b(+) was used as a cloning and expression vector for all enzymes described in this study. Error prone mutagenesis was performed using Taq DNA Polymerase. The PCR products were digested with DpnI, gel purified, and ligated using Gibson MixTM.¹ The ligation mixture was used to directly transform electrocompetent *E. coli* strain E. cloni BL21(DE3) cells (Lucigen).

(D) Expression of P411 variants in 96-well plates. Single colonies from LBamp agar plates were picked using sterile toothpicks and cultured in deep-well 96-well plates containing 3 LBamp (300 μL /well) at 37 $^\circ\text{C}$, 220 rpm shaking, and 80% relative humidity overnight. After, HBamp (950 μL /well) in a deep-well 96-well plate was inoculated with an aliquot (50 μL /well) of these overnight cultures and allowed to shake for 2.75 hours at 37 $^\circ\text{C}$, 220 rpm, and 80% relative humidity. The plates were cooled on ice for 30 minutes and the cultures were induced with 0.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) and 1.0 mM 5-aminolevulinic acid (final concentrations). Expression was conducted at 20 $^\circ\text{C}$, 150 rpm for 16–20 hours.

(E) Reaction screening in 96-well plate format. *E. coli* (*E. coli* BL21(DE3)) cells in deep-well 96-well plates were pelleted ($5,000 \times g$, 3 min, RT) and resuspended in M9-N buffer (20 $\mu\text{L}/\text{well}$) by gentle vortexing. The 96-well plate was transferred into an anaerobic chamber. In the anaerobic chamber, reaction buffer (33 mM glucose in M9-N, 380 $\mu\text{L}/\text{well}$) was added, followed by alkane substrate (10 $\mu\text{L}/\text{well}$, 400 mM in EtOH) and ethyl diazoacetate (10 $\mu\text{L}/\text{well}$, 400 mM in EtOH). In some cases, the substrates and reaction buffer were mixed together prior to addition to the plate. The plate was sealed with an aluminum foil and shaken at room temperature, 500 rpm in the anaerobic chamber. After 24 hours, the seal was removed and the reactions were worked up following the appropriate method below.

Product formation screening using GC-MS.

After 24 hours, a solution of 20 mM 1,3,5-trimethoxybenzene (internal standard) in a mixed solvent system (cyclohexane/ ethyl acetate = 1:1, 20 μL), followed by 580 μL pure solvent, was added. The plate was tightly sealed with a reusable silicone mat, vortexed ($15 \text{ s} \times 3$) and centrifuged ($3,000 \times g$, 5 min) to completely separate the organic and aqueous layers. The organic layers (200 $\mu\text{L}/\text{well}$) were transferred to 300 μL vial inserts, which were then placed in 2 mL vials and analyzed by GC.

Product formation screening by SAMDI Mass Spectrometry

After 24 hours, 190 μL crude reaction mixture was aliquoted to PCR plates and frozen on dry ice. Plates were shipped overnight on dry ice.

(1) Preparation of SAMDI Array Plate – Gold-evaporated array plates were made in-house from outsourced polished stainless-steel plates (Ace Metal Crafts). Specifically, purchased steel plates with dimensions matching standard MALDI plates (12.3cm \times 8cm) were washed with hexanes, ethanol and water. Using an electron beam evaporator (Thermionics Laboratory Inc., Hayward, CA), the plates were coated at 0.02nm/s with a 5nm layer of titanium followed by a 35nm layer of gold spots (2.8 mm diameter) at 0.05nm/s in a 384-array fashion. Before use, plates were stored under vacuum. At least two days prior to screening, the plates were soaked in a 0.5 mM ethanolic solution of alkyl disulfides (Prochimea Surfaces) at 4°C. The solution consisted of disulfide alkanes either symmetrically terminated with tri(ethylene glycol) groups or asymmetrically terminated with tri(ethylene glycol) and a maleimide functional group in a stoichiometric ratio of 2:3 expecting to yield a self-assembled monolayer of 20% maleimide surface density¹. The plates were removed from solution, rinsed with ethanol and dried with $\text{N}_{2(\text{g})}$ immediately before use.

(2) Deprotection of Reaction Constituents – 25 μL of each reaction was aliquoted, organizing four 96-well plates into one 384 plate. Liquid handling was performed by a Tecan Freedom EVO 200 Robot. To each well, 10 μL of 100mM HCl was added using an automatic reagent dispenser (Multidrop Combi, Thermo Scientific). Plates were sealed with aluminum sealing tape, mixed briefly, and centrifuged at 5000g for 1 min. After 30 min, 3 equivalents of NaHCO_3 from saturated solution was added and the plates were sealed, mixed and centrifuged at 5000g for 1 min. The plates were either frozen or analyzed using the steps below.

(3) Activity Assay by SAMDI Mass Spectrometry – Using a TECAN liquid-handling robot, 3 μL from each well were spotted to SAMDI plates. The plates were allowed to incubate for 1 hour at 37°C in a humidity chamber, providing sufficient time for immobilization of the thiolated reaction constituents to react with the maleimide. The plates were rinsed with 1% Alconox detergent, deionized ultrafiltered (DIUF) water and ethanol with sufficient pressure to remove all noticeable debris, dried with a stream of $\text{N}_{2(\text{g})}$, and coated with matrix (2',4',6'-trihydroxyacetophenone (THAP) in acetone (18mg/ml). Each spot

was analyzed in positive reflector mode by MALDI-TOF (AB Sciex 5800 series). Captured reaction constituents from whole cell reactions were identified by their mass shifts relative to the unreacted maleimide. Analysis of more than 5100 spectra collected was performed using the analysis tools built into the TOF/TOF™ Series Explorer™ Software, which works to integrate the area under the curve (AUC) of the peak corresponding to a specified mass/charge ratio. For every spectrum in this study, the following peaks were integrated: 1033 Da (thiolated substrate + monolayer), 1119 Da (thiolated C-H insertion product + monolayer). Peaks below a signal to noise threshold set at 20 were not accepted. Each mass spectra generated for each well was used to determine the relative conversion for each variant by dividing the AUC of the product peak by the AUC of both the unreacted substrate and C-H insertion product peaks. The conversion for the three parent enzymes on each 96-well plate were averaged and all other values normalized to this value to acquire relative fold improvement for each variant. The variants with the highest fold improvement from each 96-well library selected for product formation screening using GC and GC-MS detailed above.

(F) Expression of P411 variants. *E. coli* (E. cloni BL21(DE3)) cells carrying plasmid encoding the appropriate P411 variant were grown overnight in 5 mL LBamp. Preculture (5 mL) was used to inoculate 45 mL of HBamp in a 125 mL Erlenmeyer flask; this culture was incubated at 37 °C, 240 rpm for 2.25 hours. The culture was then cooled on ice (20–30 min) and induced with 0.5 mM IPTG and 1.0 mM 5-aminolevulinic acid (final concentrations). Expression was conducted at 24 °C, 140 rpm, for 20 hours. Following, *E. coli* cells were pelleted by centrifugation (2,600 × g, 10 min, 4 °C or 3,000 × g, 5 min, 4 °C). Media was removed and the resulting cell pellet was resuspended in M9-N buffer to OD600 = 60. An aliquot of this cell suspension (3 mL) was taken to determine P411 concentration using the hemochrome assay after lysis by sonication. When applicable, remaining cell suspension was further diluted with M9-N buffer to the OD600 used for the biotransformation (described in (H)) and the concentration of P411 protein in the biotransformation was calculated accordingly.

(G) Hemochrome assay for the determination of heme protein concentration. *E. coli* cells expressing heme protein and resuspended in M9-N buffer were lysed by sonication using a Qsonica Q500 sonicator equipped with a microtip (1 min, 1 second on, 1 second off, 35% amplitude); samples were kept on wet ice for this process. The resulting lysed solution was centrifuged (20,000 × g, 10 min, 4 °C) to remove cell debris. The supernatant (clarified lysate) was separated from the pellet and kept on ice until use. In a falcon tube, a solution of 0.2 M NaOH, 40% (v/v) pyridine, 0.5 mM K₃Fe(CN)₆ was prepared (pyridine-NaOH-K₃Fe(CN)₆ solution). Separately, a solution of 0.5 M Na₂S₂O₄ (sodium dithionite) was prepared in 0.1 M NaOH. To an Eppendorf tube containing 500 μL of clarified lysate in M9-N buffer was added 500 μL of the pyridine-NaOH-K₃Fe(CN)₆ solution, mixed, and transferred to a cuvette; the UV-Vis spectrum of the oxidized Fe(III) state was recorded immediately. To the cuvette was then added 10 μL of the sodium dithionite solution. The cuvette was sealed with parafilm and the UV-Vis spectrum of the reduced Fe(II) state was recorded immediately. A cuvette containing 500 μL of M9-N, 100 μL 1 M NaOH, 200 μL pyridine, and 200 μL water (complete mixture without protein and K₃Fe(CN)₆) was used as a reference for all absorbance measurements. Concentrations of cytochromes P411 were determined using a published extinction coefficient for heme b, $\epsilon_{556(\text{reduced})-540(\text{oxidized})} = 23.98 \text{ mM}^{-1} \text{ cm}^{-1}$.²

(H) Biotransformations using whole *E. coli* cells. Suspensions of *E. coli* (E. cloni BL21(DE3)) cells expressing the appropriate heme protein variant in M9-N buffer (typically OD600 = 60) were degassed by bubbling with argon in sealed vials for at least 40 minutes; the cells were kept on ice during this time. Separately, a solution of D-glucose (250 mM in M9-N) was degassed by sparging with argon for at least 30 minutes. All solutions were then transferred into an anaerobic chamber for reaction set up. To a 2 mL vial were added D-glucose (40 μL of 250 mM stock solution in M9-N buffer), M9-N buffer (typically 343.33 μL), suspension of *E. coli* expressing P411 (OD600 = 1, 6.67 μL), alkane substrate (5

SUPPORTING INFORMATION

μL of 400 mM stock solution in EtOH), and diazo compound (5 μL of 400 mM stock solution in EtOH) in the listed order. Final reaction volume was 400 μL ; final concentrations were 5 mM alkane substrate, 5 mM diazo compound, and 30 mM D-glucose. Note: reaction performed with *E. coli* cells resuspended to OD600 = 1 indicates that 6.67 μL of OD600 = 60 cells were added, and likewise for other reaction OD600 descriptions. The vials were sealed and shaken at room temperature and 500 rpm for 18 hours in the anaerobic chamber. The reactions were worked up and analyzed by GC; the reaction workup procedures are outlined in detail in Section VII. The expression of heme protein was measured using the hemochrome assay (Section I(G)), and the concentration of heme protein in the biotransformation was calculated accordingly.

II. Supplementary Figs. S1–S8 and Table S1

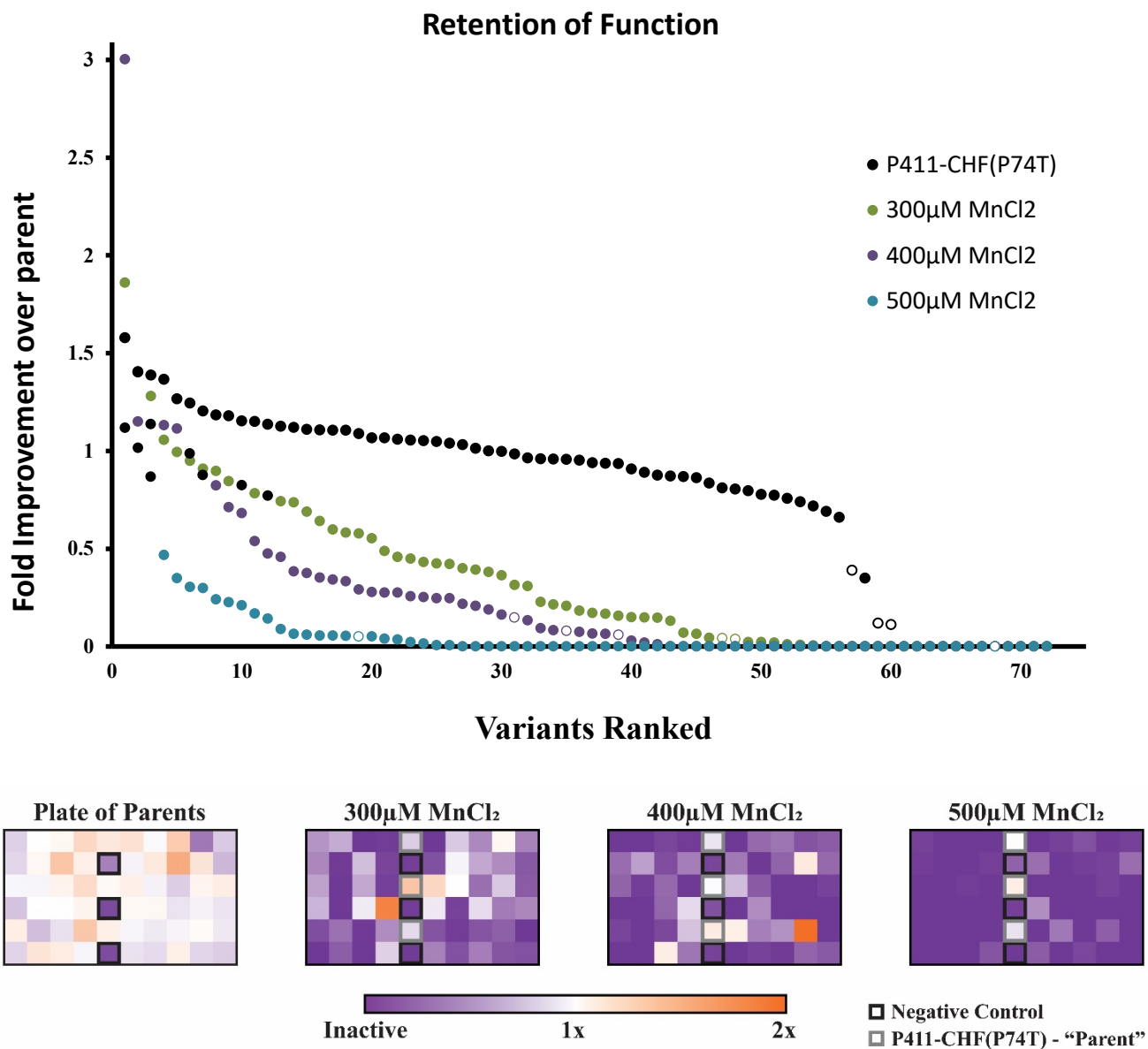


Figure S1: Results from 3 mutational loads compared to a plate of parent enzymes. Following the procedure in the experimental section, random libraries of P411-CHF(P74T) were generated using error-prone PCR and various concentrations of MnCl₂. A mutation rate using 400 μ M MnCl₂ was adequate to proceed with this enzyme. All parents are shown in black.

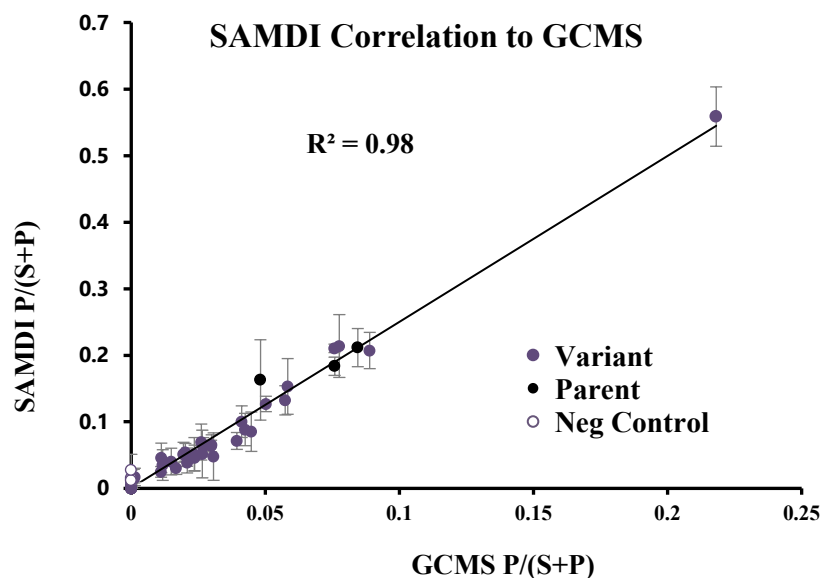


Figure S2 a: Scatterplot of screening results from Figure 2 inclusive of an additional variant that was subsequently identified as a false positive by GCMS (Figure S2 b). Data was collected via GCMS and SAMDI from a library of 70 mutants using P411-CHF(P74T) as the parent and 400 μM MnCl_2 during epPCR comparing product to remaining substrate. Values were calculated as a fraction of product over the total of the remaining starting material and product formed. Correlation was determined using least squares linear regression. Here a slope of 2.5 shows that the SAMDI and GC methods do not give the same uncalibrated measurement of yield, and is likely due to different ionization efficiencies of the reaction products. We have, however, verified that the SAMDI method does provide a quantitative measure of yield and allows identification of top variants and perform evolution.

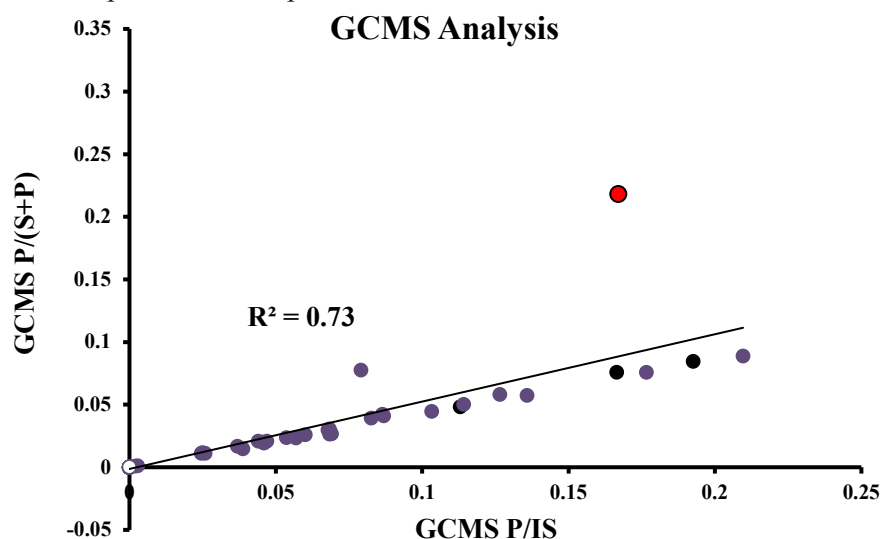


Figure S2 b: Correlations of variants collected on GCMS from the random mutagenesis in Figure S2a library using P411-CHF(P74T) as the parent and 400 μM MnCl_2 during epPCR comparing product to remaining substrate with product concentration corrected by an internal standard. Data showed good correlation for most points; however, this direct comparison reveals a point (red) to be a false positive hit when only considering product and remaining substrate. Parent controls are shown in black and negative controls in white.

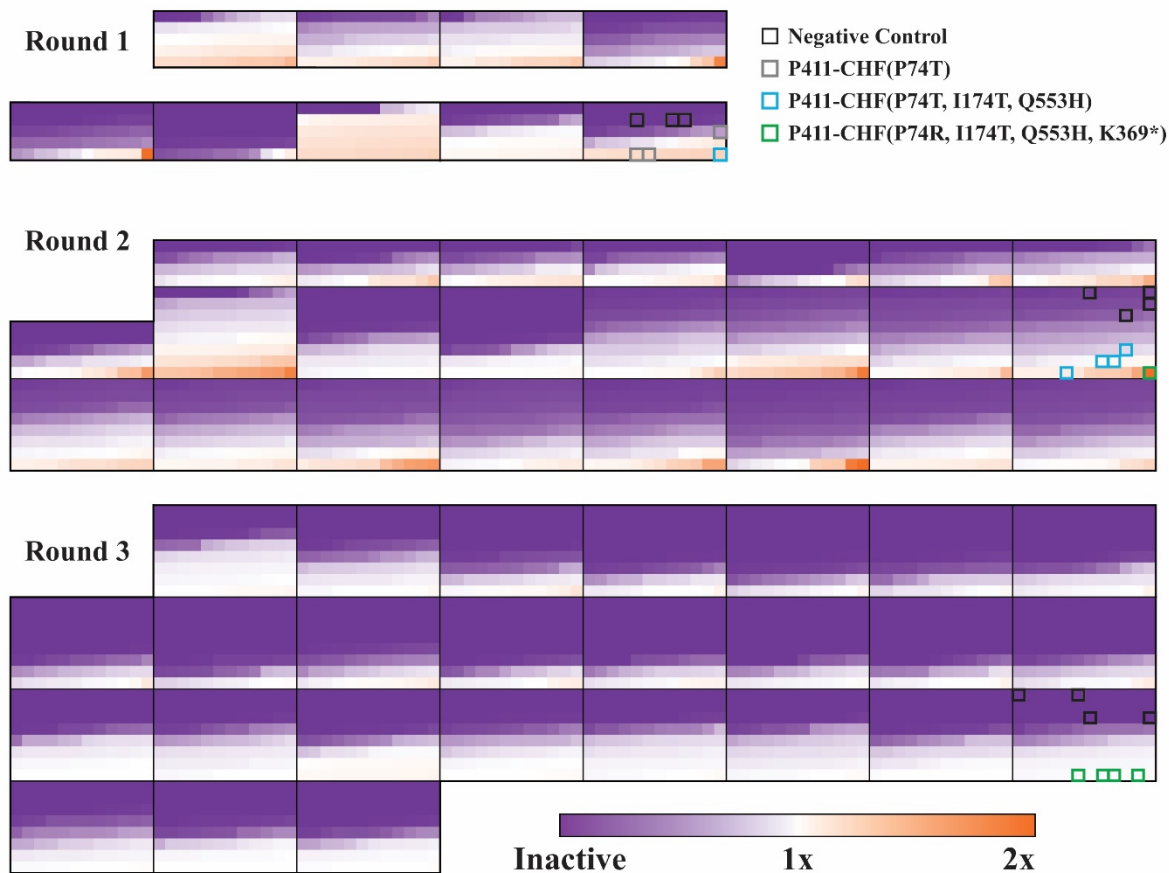


Figure S3: Heatmaps display the relative activity of 4,956 samples measured via SAMDI. Mutants are shaded and organized by relative fold improvement normalized by the average of parent controls on each respective 96-well plate. Visible differences in the amount of active versus inactive variants in each heat map can be attributed to the number of mutations in each variant. Variants identified by SAMDI as having a fold improvement of more than two times the relative standard deviation on the average of parent activity from each library were selected to be validated by GCMS. The mutant confirmed to be the most active was then chosen for the next round of evolution. The best confirmed variant from each round (boxed in blue for round 1 and in green for round 2) were used as the parent in each subsequent round of evolution. In the third round of experiments, none of the variants were verified either by SAMDI or via GCMS to have significantly improved activity over the previous parent. In the third round, we did not identify a variant with significantly improved activity out of the 2,496 variants screened and decided to end the campaign. We chose the best variant from round 2 as the final variant.

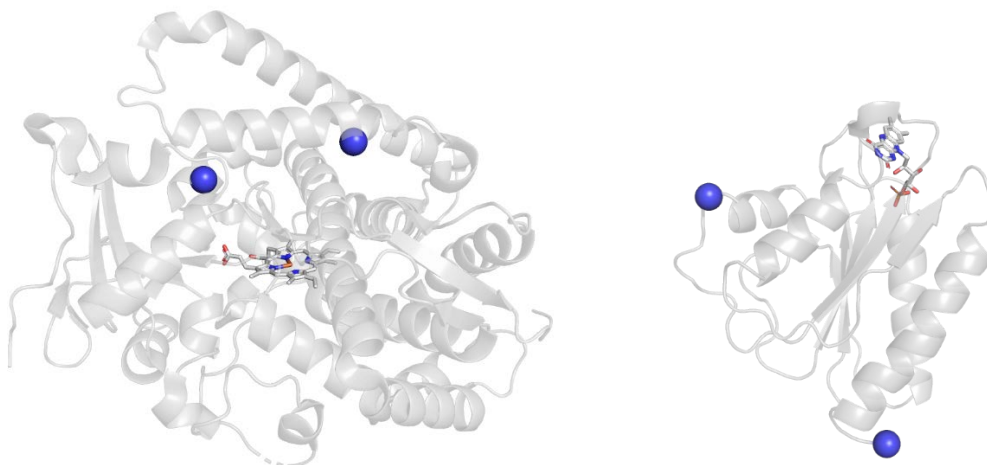


Figure S4: Structural visualization of amino acids mutated during evolution of P411-CHF to our final variant. The final variant was identified to contain 4 amino acid mutations from the initial parent with 2 in the heme domain and 2 in the FMN domain. The following mutations were accumulated (from left to right): In the heme domain: P74T, I174T. In the FMN domain: Both a Q553H mutation and a stop codon at K639. While the P74T mutation identified in the initial parent is located in the active site, none of the mutated residues identified by SAMDI are more than 15 Å from the iron atom in the center of the heme cofactor.

Note: The structure of the heme domain was modeled using the crystal structure of a related P411 variant (PDB 5UCW), which contains nine additional mutations. The structure of the FMN domain was modeled using the crystal structure from a related variant (PDB 1BVY).

SAMDI Reproducibility

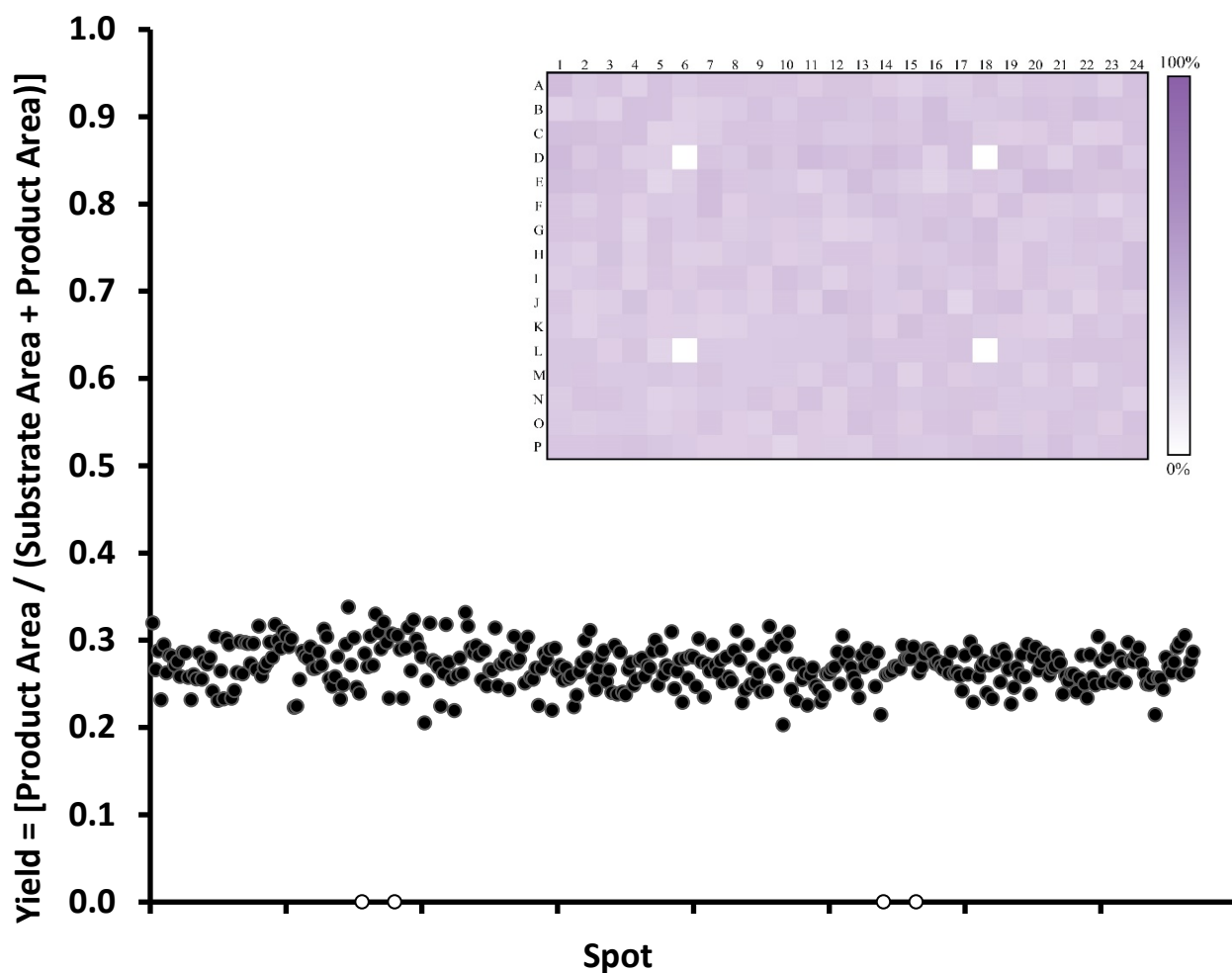


Figure S5: Reproducibility in the SAMDI method. Using SAMDI, a whole-cell reaction with thioacetate-tagged 1-methoxyhept-2-ene was deprotected and spotted 380 times on a SAMDI plate, without prior purification or centrifugation. 10 mL of reaction was conducted from 1 colony and 10 mM substrate at 25 °C for 24 hours. Negative controls using P411-TrpB are located in rows D and L, columns 6 and 18. A standard deviation of 2.3% was found on an average yield of 27%, indicating the assay is able to discern active wells from non-active wells sufficiently. From these data, we subsequently calculated the resolving power of the technique using the m/z values for each peak and the average resolution output by the instrument. Here we found a value of 0.1 m/z , indicating the method's applicability to transformations of minute changes in mass.

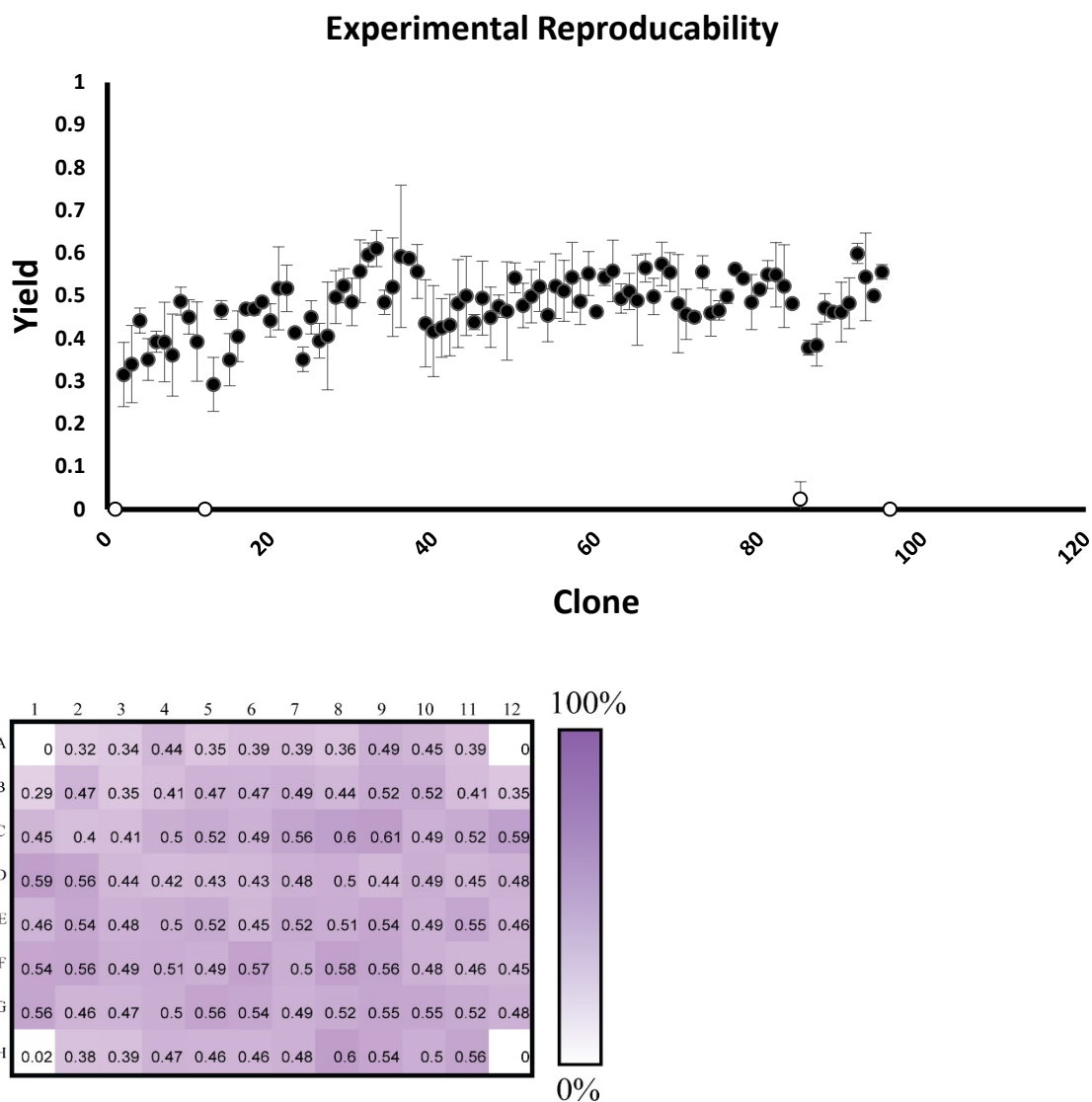


Figure S6: Results of screening a plate of the same clone to show the assay's variability. Measurement of P411-CHF(P74T) activity with thioacetate-tagged 1-methoxyhept-2-ene in a 96-deepwell format. 1 mL of expressed cells was centrifuged and resuspended in 0.4 mL M9-N with 10 mM substrate at 22 °C for 24 hours. Negative controls with no cells are located in rows A and H, columns 1 and 12. Each data point represents a separate reaction using a different clone from the same enzyme. Each reaction was deprotected and spotted in quadruplicates to SAMDI plates, without prior purification or centrifugation and averaged to give each data point. All values are then averaged. A coefficient of variation of 14% was found on an average yield of 48%, indicating the assay's reproducibility in screening. An increase in variability outside of the SAMDI technique is expected here as a result of differences in expression and all experimental errors.

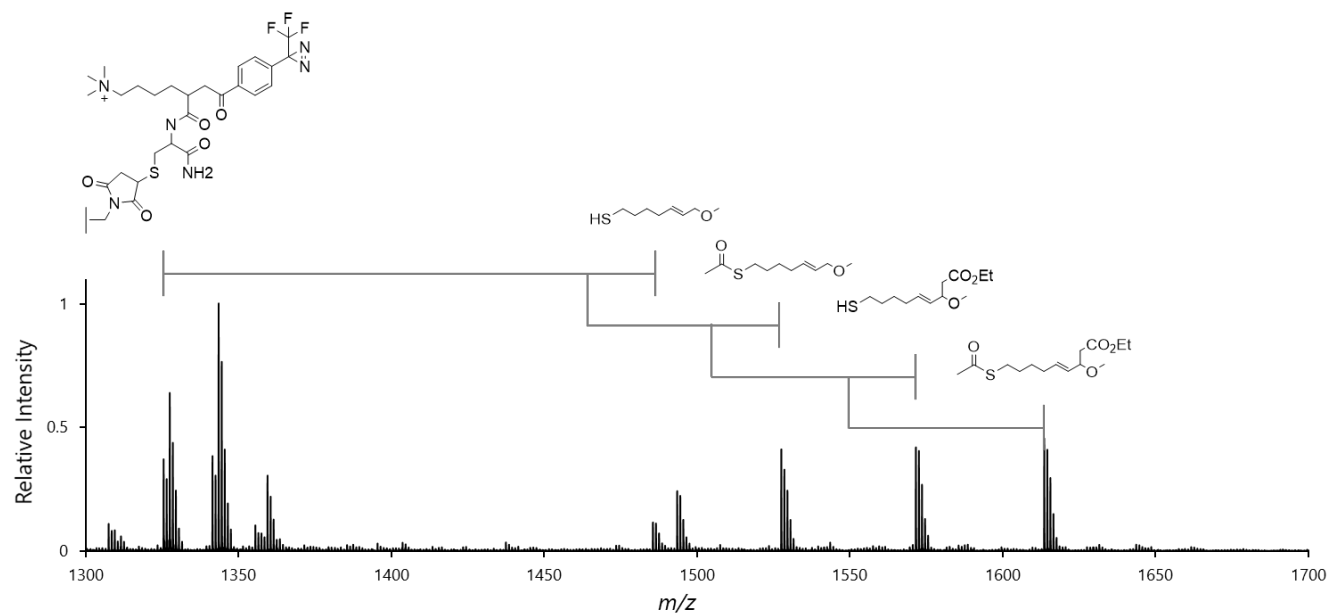
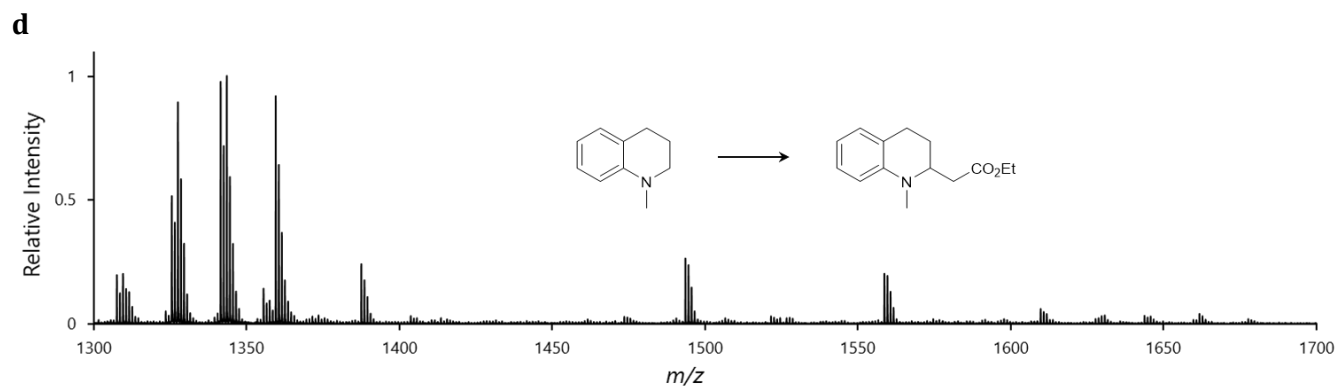
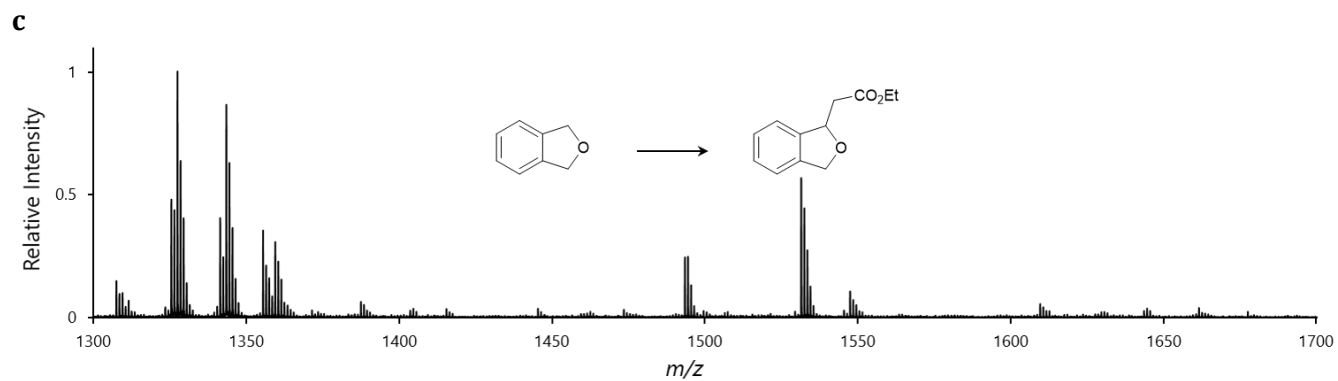
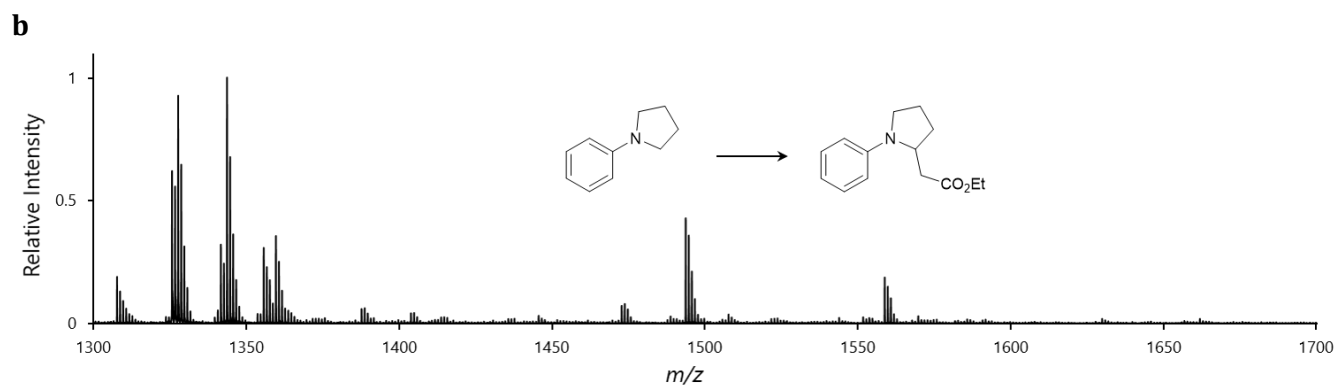
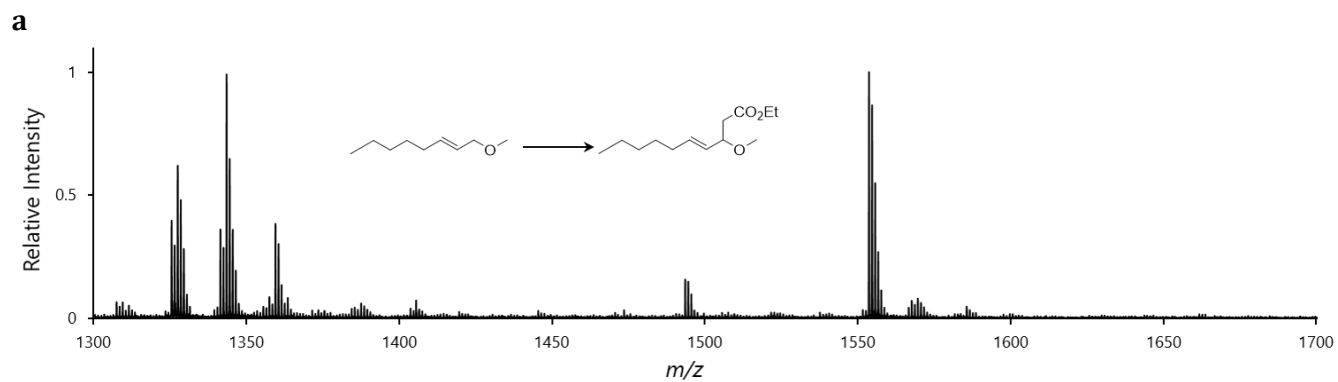


Figure S7: Using methods previously reported elsewhere,⁷ a diazirine-terminated monolayer was used to analyze the reaction products for the reaction from Figure S5, demonstrating that analysis of this reaction was not reliant on thiol-maleimide chemistry. Here, SAMDI-MS spectra displays the diazirine monolayer at m/z 1325 (after loss of nitrogen and conversion to carbene) and clear peaks corresponding to both the deprotected and remaining acetate-protected reaction products. Crude lysates were briefly deprotected in 28mM HCl for 30minutes and quenched with saturated sodium bicarbonate in order to show the ability of this monolayer to immobilize all 4 species. No extraction or centrifugation was performed prior to analysis. The deprotected and acetate-protected substrate appear at m/z 1485 and m/z 1527, respectively, and their corresponding insertion product are seen + 86 Daltons at m/z 1571 and m/z 1613, respectively. Primary byproducts are seen due to reaction with water (m/z 1343) and 2,4,6-trihydroxyacetophenone (m/z 1493), the MALDI matrix.

SUPPORTING INFORMATION



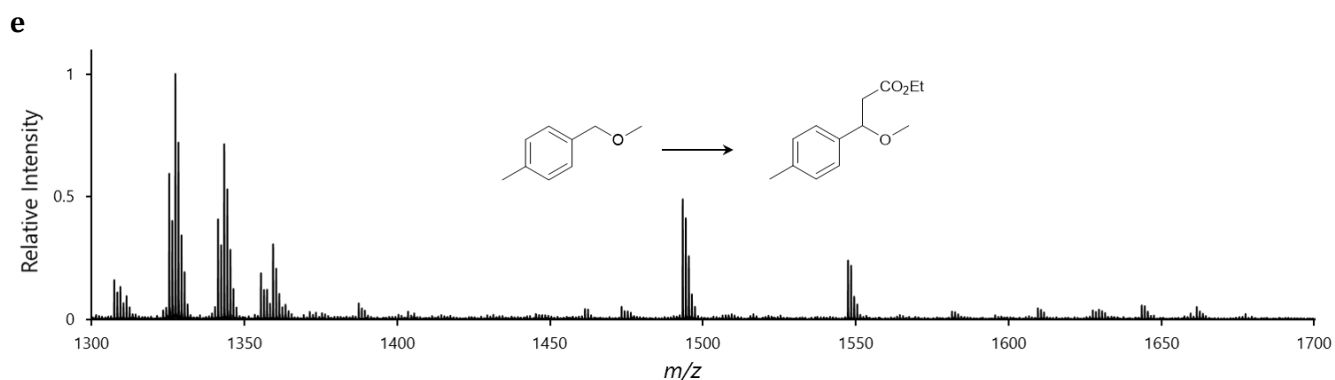


Figure S8 (a-e): SAMDI-MS data that demonstrates the technique's generalizability by its capacity to record product formation for a variety of reactions within the substrate scope of the enzyme. Because the expected reaction products do not contain functional groups that may be exploited for straight-forward immobilization, reactions were analyzed using a diazirine monolayer and conditions reported in Figure S7. The evolved P411 variant was used to catalyze the reaction of ethyl diazoacetate and **a.** (E)-1-Methoxyoct-2-ene, **b.** Ethyl 2-(1-phenylpyrrolidin-2-yl)acetate **c.** Ethyl 2-(1,3-dihydroisobenzofuran-1-yl)acetate **d.** Ethyl 2-(1-methyl-1,2,3,4-tetrahydroquinolin-2-yl)acetate and **e.** 1-(Methoxymethyl)-4-methylbenzene. The peaks corresponding to the substrate and product for these reactions are as follows: **a.** m/z 1467 and 1553, **b.** m/z 1472 and 1558, **c.** m/z 1445 and 1531, **d.** 1472 and 1558, and **e.** m/z 1461 and 1547, respectively. As expected, all product peaks are shifted +86 Da. Further knowledge of the differences in ionization and immobilization efficiency between the substrate and their product is required to quantitate yields; however, it is clear that the evolved P411 variant remains active for a number of substrates.

SUPPORTING INFORMATION

Table S1: Beneficial mutations discovered by SAMDI

Gen.	Percent Yield	Fold Improvement	Mutation	TTN
0	2.44%	--	--	637
1	4.46%	1.83	I174T Q553H	1107
2	6.28%	1.4	K639*	1310

Note: Due to the large number of variants screened, other improved variants were identified in the course of this study with lower fold-improvement, however only the top variant from each round was chosen and characterized more rigorously.

III. Screening of enzymes for C–H alkylation activity

Testing heme proteins for reaction discovery

A composite plate of 92 heme proteins and their variants from *Bacillus megaterium* and *Rhodothermus marinus* were screened for formation of ethyl (E)-9-(acetylthio)-3-methoxynon-4-enoate from substrates (E)-S-(7-methoxyhept-5-en-1-yl) ethanethioate and ethyl diazoacetate. These proteins were cloned and used in other studies, including carbene Si–H insertion,³ alkene cyclopropanation,^{4,5} and alkane C-H functionalization.⁶ Expression of these proteins followed the procedures as described by the prior studies and testing for initial activity was carried out with whole *E. coli* cells. The general procedure for reaction screening in 96-well plate format (Section I (E)) was employed and the reactions were analyzed by GC-MS. Note: Since this was an initial test, proper expression of the indicated proteins was not verified and experimental replicates were not performed. A variant closely related to P411-CHF from prior alkane C-H functionalization studies,⁶ P411-CHF +P74T, was identified as the most promising variant for evolution.

IV. Directed evolution of C–H alkylation enzymes

Error-prone libraries were generated using *Taq* Polymerase and differing concentrations of manganese chloride, and screened in 96-well plates. Following the general screening in 96-well plate procedure (Section I (E)), variants which exhibited higher formation of C–H alkylation product were identified. A summary of the beneficial mutation(s) is presented in Section II Supplementary Table S1. The locations of the selected beneficial mutations are displayed on a structural model of the P411 enzyme (Section II, Supplementary Fig. 4).

Variants which were identified to show higher activity during screening were streaked out on LB_{amp} agar plates. A single colony was selected, sequenced, and the TTN measured for our product using whole *E. coli* cells overexpressing the desired protein (the cell density was normalized such that each reaction contained the same cell density, general procedure Section I). The results are summarized in Supplementary Table 1.

V. Nucleotide and amino acid sequences

All heme proteins disclosed below were cloned into a pET22b(+) vector.

DNA and amino acid sequence of P411-CHF(P74T), a previously reported cytochrome P411 variant which was used as the starting point for evolution:

DNA Sequence:

```
ATGACAATTAAGAAATGCCTCAGCCAAAACGTTTGGAGAGCTTAAAAATTTACCGTTAT
TAAACACAGATAAACCGGTTCAAGCTTTGATGAAAATTGCGGATGAATTAGGAGAAATCTT
TAAATTCGAGGCGCCTGGTCGTGTAACGCGCTACTTATCAAGTCAGCGTCTAATTAAGAA
GCATGCGATGAATCACGCTTTGATAAAGAGTTAAGTCAAACGCTGAAATTTCTGCGTGATT
TTCTTGGAGACGGGTTAGCCACAAGCTGGACGCATGAAAAAAATTGGAAAAAAGCGCATA
ATATCTTACTTCCAAGCTTTAGTCAGCAGGCAATGAAAGGCTATCATGCGATGATGGTCGA
TATCGCCGTGCAGCTTGTTCAAAGTGGGAGCGTCTAAATGCAGATGAGCATATTGAAGTA
TCGGAAGACATGACACGTTTAAACGCTTGATACAATTGGTCTTTGCGGCTTTAACTATCGCTT
```

SUPPORTING INFORMATION

TAACAGCTTTTACCGAGATCAGCCTCATCCATTTATTATAAGTCTGGTCCGTGCACTGGATG
AAGTAATGAACAAGCTGCAGCGAGCAAATCCAGACGACCCAGCTTATGATGAAAACAAGC
GCCAGTTTCAAGAAGATATCAAGGTGATGAACGACCTAGTAGATAAAATTATTGCAGATC
GCAAAGCAAGGGGTGAACAAAGCGATGATTTATTAACGCAGATGCTAAACGGAAAAGATC
CAGAAACGGGTGAGCCGCTTGATGACGGGAACATTCGCTATCAAATTATTACATTCTTATA
TGCGGGAGTTGAAGGTACAAGTGGTCTTTTATCATTGCGCTGTATTTCTTAGTGAAAAATC
CACATGTATTACAAAAAGTAGCAGAAGAAGCAGCACGAGTTCTAGTAGATCCTGTTCCAA
GCTACAAACAAGTCAAACAGCTTAAATATGTCGGCATGGTCTTAAACGAAGCGCTGCGCTT
ATGGCCAACGGTTCCTTATTTTTCCCTATATGCAAAAAGAAGATACGGTGCTTGGAGGAGAA
TATCCTTTAGAAAAAGGCGACGAAGTAATGGTTCTGATTCCTCAGCTTCACCGTGATAAAA
CAGTTTGGGGAGACGATGTGGAGGAGTTCCGTCCAGAGCGTTTTGAAAATCCAAGTGCGAT
TCCGCAGCATGCGTTTAAACCGTTTGGAAACGGTCAGCGTGCGTCTATCGGTCAGCAGTTC
GCTCTTCATGAAGCAACGCTGGTACTTGGTATGATGCTAAAACACTTTGACTTTGAAGATC
ATACAAACTACGAGCTCGATATTAAGAAGTCTTACGTTAAAACCTAAAGGCTTTGTGGT
AAAAGCAAAATCGAAAAAATTCCGCTTGGCGGTATTCCTTCACCTAGCACTGAACAGTCT
GCTAAAAAAGTACGCAAAAAGGCAGAAAACGCTCATAATACGCCGCTGCTTGTGCTATAC
GGTTCAAATATGGGTACCGCTGAAGGAACGGCGCGTGATTTAGCAGATATTGCAATGAGC
AAAGGATTTGCACCGCAGGTCGCAACGCTTGATTCACACGCCGGAATCTTCCGCGCGAAG
GAGCTGTATTAATTGTAACGGCGTCTTATAACGGTCATCCGCCTGATAACGCAAAGCAATT
TGTCGACTGGTTAGACCAAGCGTCTGCTGATGAAGTAAAAGGCGTTCGCTACTCCGATTT
GGATGCGGCGATAAAAAGTGGGCTACTACGTATCAAAAAGTGCCTGCTTTTATCGATGAAA
CGCTTGCCGCTAAAGGGGCAGAAAACATCGCTGACCGCGGTGAAGCAGATGCAAGCGACG
ACTTTGAAGGCACATATGAAGAATGGCGTGAACATATGTGGAGTGACGTAGCAGCCTACTT
TAACCTCGACATTGAAAACAGTGAAGATAATAAATCTACTCTTCACTTCAATTTGTCGAC
AGCGCCGCGGATATGCCGCTTGCGAAAATGCACGGTGCGTTTTCAACGCTCGAGCACCACC
ACCACCACCACTGA

Amino Acid Sequence:

MTIKEMPQPKTFGELKNLPLLNTDKPVQALMKIADELGEIFKFEAPGRVTRYLSSQRLIKEACD
ESRFDKELSQTLKFLRDFLDGLATSWTHEKNWKKAHNILLPSFSQQAMKGYHAMMVDIAVQ
LVQKWERLNADEHIEVSEDMTRLTLDTIGLCGFNYRFNSFYRDQPHPFIIISLVRALDEVMNKLQ
RANPDDPAYDENKRQFQEDIKVMNDLVDKIIADRKARGEQSDDLQMLNGKDPETGEPLDD
GNIRYQIITFLYAGVEGTSGLLSFALYFLVKNPHVLQKVAEEAARVLVDPVPSYKQVKQLKYV

SUPPORTING INFORMATION

GMVLNEALRLWPTVPYFSLYAKEDTVLGGEYPLEKGDEVMLIPQLHRDKTVWGDDVEEFRP
ERFENPSAIPQHAFKPFNGQRASIGQQFALHEATLVLGMMLKHFDHFEDHTNYELDIKELLTLK
PKGFVVKAKSKKIPLGGIPSPSTEQSAKKVRKKAENAHNTPLLVL YGSNMGTAEGTARDLADI
AMSKGFAPQVATLDSHAGNLPREGAVLIVTASYNGHPPDNAKQFVDWLDQASADEVKGVRY
SVFGCGDKNWATTYQKVPAFIDETLAAKGAENIADRGEADASDDFEGTYEEWREHMWSDVA
AYFNLDIENSEDNKSTLSLQFVDSAADMPLAKMHGAFSTLEHHHHHH*

DNA and amino acid sequence of our final variant, an evolved C–H alkylation enzyme:

DNA Sequence:

ATGACAATTAAGAAATGCCTCAGCCAAAACGTTTGGAGAGCTTAAAAATTTACCGTTAT
TAAACACAGATAAACCGGTTCAAGCTTTGATGAAAATTGCGGATGAATTAGGAGAAATCTT
TAAATTCGAGGCGCCTGGTCGTGTAACGCGCTACTTATCAAGTCAGCGTCTAATTAAGAA
GCATGCGATGAATCACGCTTTGATAAAGAGTTAAGTCAAACGCTGAAATTTCTGCGTGATT
TTCTTGGAGACGGGTTAGCCACAAGCTGGACGCATGAAAAAAATTGGAAAAAAGCGCATA
ATATCTTACTTCCAAGCTTTAGTCAGCAGGCAATGAAAGGCTATCATGCGATGATGGTCGA
TATCGCCGTGCAGCTTGTTCAAAGTGGGAGCGTCTAAATGCAGATGAGCACATTGAAGTA
TCGGAAGACATGACACGCTTAACGCTTGATACAATTGGTCTTTGCGGCTTTAACTATCGCTT
TAACAGCTTTTACCGAGATCAGCCTCATCCATTTACTATAAGTCTGGTCCGTGCACTGGATG
AAGTAATGAACAAGCTGCAGCGAGCAAATCCAGACGACCCAGCTTATGATGAAAACAAGC
GCCAGTTTCAAGAAGATATCAAGGTGATGAACGACCTAGTAGATAAAATTATTGCAGATC
GCAAAGCAAGGGGTGAACAAAGCGATGATTTATTAACGCAGATGCTAAACGGAAAAGATC
CAGAAACGGGTGAGCCGCTTGATGACGGGAACATTCGCTATCAAATTATTACATTCTTATA
TGCGGGAGTTGAAGGTACAAGTGGTCTTTTATCATTGCGCTGTATTTCTTAGTGAAAAATC
CACATGTATTACAAAAGTAGCAGAAGAAGCAGCACGAGTTCTAGTAGATCCTGTTCCAA
GCTACAAACAAGTCAAACAGCTTAAATATGTCGGCATGGTCTTAAACGAAGCGCTGCGCTT
ATGGCCAACGGTTCCTTATTTTTCCCTATATGCGAAAGAAGATACGGTGCTTGGAGGAGAA
TATCCTTTAGAAAAAGGCGACGAAGTAATGGTCTGATTCTCAGCTTCACCGTGATAAAA
CAGTTTGGGGAGACGATGTGGAGGAGTCCGTCCAGAGCGTTTTGAAAATCCAAGTGCAT
TCCGCAGCATGCGTTTAAACCGTTTGGAAACGGTCAGCGTGCGTCTATCGGTCAGCAGTTC
GCTCTTCATGAAGCAACGCTGGTACTTGGTATGATGCTAAAACACTTTGACTTTGAAGATC
ATACAAACTACGAGCTCGATATTAAGAAGTACTTACGTTAAAACCTAAAGGCTTTGTGGT
AAAAGCAAATCGAAAAAATTCCGCTTGGCGGTATTCCTTCACCTAGCACTGAACAGTCT
GCTAAAAAAGTACGCAAAAAGGCAGAAAACGCTCATAATACGCCGCTGCTTGTGCTATAC

SUPPORTING INFORMATION

GGTTCAAATATGGGTACCGCTGAAGGAACGGCGCGTGATTTAGCAGATATTGCAATGAGC
AAAGGATTTGCACCGCAGGTCGCAACGCTTGATTCACACGCCGAAATCTTCCGCGCGAAG
GAGCTGTATTAATTGTAACGGCGTCTTATAACGGTCATCCGCCTGATAACGCAAAGCAATT
TGTCGACTGGTTAGACCATGCGTCTGCTGATGAAGTAAAAGGCGTTCGCTACTCCGTATTT
GGATGCGGGCGATAAAAACTGGGCTACTACGTATCAAAAAGTGCCTGCTTTTATCGATGAAA
CGCTTGCCGCTAAAGGGGCAGAAAACATCGCTGACCGCGGTGAAGCAGATGCAAGCGACG
ACTTTGAAGGCACATATGAAGAATGGCGTGAACATATGTGGAGTGACGTAGCAGCCTACTT
TAACCTCGACATTGAAAACAGTGAAGATAATTAA

Amino Acid Sequence:

MTIKEMPQPKTFGELKNLPLLNTDKPVQALMKIADELGEIFKFEAPGRVTRYLSSQRLIKEACD
ESRFDKELSQTLKFLRDFLGDGLATSWTHEKNWKKKAHNILLPSFSQQAMKGYHAMMVDIAVQ
LVQKWERLNADEHIEVSEDMTRLTLDTIGLCGFNYRFNSFYRDQPHPFTISLVRALDEVMNKL
QRANPDDPAYDENKRQFQEDIKVMNDLVDKIIADRKARGEQSDDLTLQMLNGKDPETGEPLD
DGNIRYQIITFLYAGVEGTSGLLSFALYFLVKNPHVLQKVAEEAARVLVDPVPSYKQVKQLKY
VGMVLNEALRLWPTVPYFSLYAKEDTVLGGEYPLEKGDEVMLIPQLHRDKTVWGDDVEEFR
PERFENPSAIPQHAFKPFNGQRASIGQQFALHEATLVLGMMMLKHFDHFEDHTNYELDIKELLTL
KPKGFFVKAKSCKIPLGGIPSPSTEQSAKKVRKKAENAHNTPLLVLVYGSNMGTAEGTARDLAD
IAMSKGFAPQVATLDSHAGNLPREGAVLIVTASYNGHPPDNAKQFVDWLDHASADEVKGVRY
SVFGCGDKNWATTYQKVPAFIDETLAAKGAENIADRGEADASDDFEGTYEEWREHMWSDVA
AYFNLDIENSEDN*

VI. Chemical synthesis and characterization

General Procedure A: Methylation of alcohols

To a 250 mL round bottom flask was added NaH (60% dispersion in mineral oil, 15–30 mmol, 1.2–1.5 equiv.). The flask was evacuated and filled with argon (3 times). Anhydrous THF (45–80 mL) was added by syringe and the reaction mixture was cooled to 0 °C in an ice bath. Alcohol (10–20 mmol, 1.0 equiv.) in THF (5–10 mL) was added dropwise and the reaction mixture was allowed to warm to room temperature and stirred for 30 minutes. Following, iodomethane (20–40 mmol, 2.0 equiv.) in THF (10 mL) was added and the reaction was stirred at room temperature (8–15 hours). The reaction was quenched by the addition of brine (60 mL) or NH₄Cl (sat. aq., 60 mL) and the phases were separated. The aqueous layer was extracted with diethyl ether (3 × 60 mL); the combined organics were washed with aq. sodium thiosulfate (10% w/v, 50 mL, when necessary), dried over Na₂SO₄ and concentrated under reduced pressure. Purification by silica column chromatography with hexanes / ethyl acetate or pentane / diethyl ether afforded compounds the desired products in 45–60% yield.

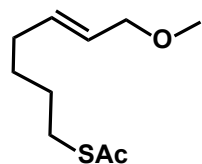
General Procedure B: Thioacetate Tagging of alkyl bromines

To a 100 mL round bottom flask was added alkyl bromide (1–10 mmol, 1 equiv.) in acetone (10–50 mL). Potassium thioacetate (6–60 mmol, 6.0 equiv.) in acetone was added and the reaction mixture was stirred for up to two days until the bromide was fully consumed as indicated by TLC. Solvent was removed under reduced pressure and crude product was purified by silica column chromatography (hexanes / ethyl acetate) afforded the desired products in 95–99% yield.

(E)-S-(7-methoxyhept-5-en-1-yl) ethanethioate

To a 100 mL flamed dried flask was added Grubbs' catalyst 2nd generation (85 mg, 1 mol%). The flask was then evacuated and backfilled with argon three times. Under argon, a dry CH₂Cl₂ solution containing 6-bromo-1-hexene (1.63 g, 10 mmol, 1.0 equiv.) and crotonaldehyde (3.50 g, 50 mmol, 5.0 equiv.) was added to the flask. The mixture was stirred under reflux for 20 hours and then cooled to room temperature and filtered through a silica plug. The solvent was removed under reduced pressure and the crude product was purified by flash chromatography (hexanes / ethyl acetate) to give (E)-7-bromohept-2-enal (1.6 g, 84% yield). This product was then dissolved in 10 mL dry THF and then added slowly to a suspension of NaBH₄ (375 mg, 10 mmol, 1.0 equiv.) in dry THF (10 mL) at 0 °C. To this reaction mixture, iodine (1.27 g, 5 mmol, 0.5 equiv.) in 10 mL of THF was slowly added at 0 °C. Reaction was stirred until the aldehyde was fully reduced as indicated by TLC. The reaction was quenched with NH₄Cl (sat. aq.), the phases were separated, and the aqueous phase was extracted with ethyl acetate (3 × 20 mL). The combined organic

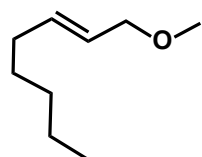
layers were washed with brine and dried over Na₂SO₄. The solvent was removed under reduced pressure and the crude alcohol product was used directly without purification. **General Procedure A** was used for the methylation step and **General Procedure B** was used for the thioacetate tagging. The final product was obtained with 50% overall yield (1.01g, 5 mmol).



¹H NMR (400 MHz, Chloroform-*d*) δ 5.67 (dtt, *J* = 15.4, 6.5, 1.1 Hz, 1H), 5.54 (dtt, *J* = 15.4, 6.1, 1.3 Hz, 1H), 3.85 (dq, *J* = 6.0, 1.0 Hz, 2H), 3.31 (s, 3H), 2.86 (t, *J* = 7.2 Hz, 2H), 2.32 (s, 3H), 2.06 (tdq, *J* = 7.5, 6.6, 1.1 Hz, 2H), 1.64 – 1.53 (m, 2H), 1.53 – 1.39 (m, 2H). ¹³C NMR (101 MHz, CDCl₃) δ 195.98, 134.03, 126.66, 77.36, 77.04, 76.72, 73.18, 57.75, 31.75, 30.65, 29.02, 28.95, 28.17. HRMS (EI) *m/z*: 202.1031 (M⁺); calc. for C₁₀H₁₈SO₂: 202.1028.

(E)-1-Methoxyoct-2-ene

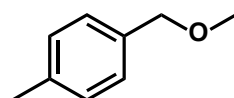
Prepared from (E)-oct-2-en-1-ol using **General Procedure A**. This compound is known in the literature⁶.



¹H NMR (400 MHz, CDCl₃) δ 5.70 (dtt, *J* = 15.6, 6.6, 1.2 Hz, 1H), 5.54 (dtt, *J* = 15.3, 6.2, 1.4 Hz, 1H), 3.86 (dq, *J* = 6.2, 1.0 Hz, 2H), 3.31 (s, 3H), 2.08 – 1.99 (m, 2H), 1.43 – 1.34 (m, 2H), 1.34 – 1.21 (m, 4H), 0.88 (t, *J* = 7.0 Hz, 3H). ¹³C NMR (101 MHz, CDCl₃) δ 135.2, 126.1, 73.5, 57.8, 32.4, 31.5, 28.9, 22.7, 14.2.

1-(Methoxymethyl)-4-methylbenzene

Prepared from *p*-tolylmethanol using **General Procedure A**. This compound is known in the literature⁶.



¹H NMR (400 MHz, CDCl₃) δ 7.23 (d, *J* = 8.0 Hz, 2H), 7.17 (d, *J* = 7.9 Hz, 2H), 4.43 (s, 2H), 3.37 (s, 3H), 2.35 (s, 3H).

Ethyl (E)-9-(acetylthio)-3-methoxynon-4-enoate

To a 100 mL flamed dried flask was added Grubbs' catalyst 2nd generation (85 mg, 1 mol%). The flask was then evacuated and backfilled with argon three times. Under argon, a dry CH₂Cl₂ solution containing 6-bromo-1-hexene (1.63 g, 10 mmol, 1.0 equiv.) and crotonaldehyde (3.50 g, 50 mmol, 5.0 equiv.) was added to the flask. The mixture was stirred under reflux for 20 hours and then cooled to room temperature and filtered through a silica plug. The solvent was removed under reduced pressure and the crude product was purified by flash chromatography (hexanes / ethyl acetate) to give (E)-7-bromohept-2-enal (1.6 g, 84% yield).

In a dry 100 mL round bottom flask, under argon, a solution of diisopropylamine (3.3 mmol, 1.1 equiv.) in THF (15 mL) was cooled to -78 °C and kept at this temperature for the remainder of the reaction. *n*-Butyllithium (3.3 mmol, 1.1 equiv., 1.6 or 2.5 M in hexanes) was added dropwise and the resulting mixture

SUPPORTING INFORMATION

was stirred for 15-30 min. Ethyl acetate (3 mmol, 1.0 equiv.) was added dropwise and the mixture was stirred for an additional 30-45 min. Then, neat aldehyde (3.3 mmol, 1.1 equiv.) was added slowly and the solution was stirred for a further 3 hours. The reaction mixture was quenched at -78 °C by the addition of NH₄Cl (sat. aq., 30 mL) and allowed to thaw to room temperature. Phases were separated and the aqueous phase was extracted with ethyl acetate or diethyl ether (3 × 20–30 mL). The combined organics were washed with NH₄Cl (sat. aq., 2 × 10–15 mL), brine (10 mL), dried over Na₂SO₄ and concentrated under reduced pressure. Purification by silica column chromatography with hexanes / ethyl acetate afforded the desired aldol adducts in 75% yield. **General Procedure A** was used for the methylation step and **General Procedure B** was used for the thioacetate tagging. The final product was obtained with 40% overall yield (1.15g, 4 mmol).

¹H NMR (400 MHz, Chloroform-d) δ 5.68 (dtd, J = 15.4, 6.8, 0.8 Hz, 1H), 5.31 (ddt, J = 15.4, 8.1, 1.4 Hz, 1H), 4.14 (qd, J = 7.1, 0.7 Hz, 2H), 3.97 (tdd, J = 8.1, 5.5, 0.8 Hz, 1H), 3.25 (s, 3H), 2.86 (t, J = 7.2 Hz, 2H), 2.59 (dd, J = 14.9, 8.1 Hz, 1H), 2.42 (dd, J = 15.0, 5.5 Hz, 1H), 2.32 (s, 3H), 2.16 – 2.00 (m, 2H), 1.64 – 1.53 (m, 3H), 1.53 – 1.39 (m, 2H), 1.25 (t, J = 7.1 Hz, 3H). **¹³C NMR** (101 MHz, CDCl₃) δ 195.93, 171.04, 134.44, 129.21, 78.73, 77.34, 77.03, 76.71, 60.46, 56.16, 41.34, 31.61, 30.66, 28.99, 28.90, 28.16, 14.25. **HRMS** (EI) m/z: 288.1388 (M⁺); calc. for C₁₄H₂₄SO₄: 288.1395.

VII. Small scale enzymatic reactions and product calibration curve

Enzymatic reactions performed on analytical scale were conducted following the general procedure described below, also described in Section I (H). Product formation was quantified by GC based on the calibration curve of the chemically synthesized compound. TTN is defined as the amount of product divided by total heme protein as measured by the hemochrome assay (Section I (G)). Analysis data presented in this section are for results shown in Fig. 3 of the main text.

General procedure for biotransformations using whole *E. coli* cells

Suspensions of *E. coli* expressing the appropriate heme protein variant in M9-N buffer (OD600 = 30) were degassed by bubbling with argon in sealed vials for at least 40 minutes; the cells were kept on ice during this time. Separately, a solution of D-glucose (250 mM in M9-N) was degassed by sparging with argon for at least 30 minutes. To a 2 mL vial was added a suspension of *E. coli* expressing heme protein (OD600 = 1, 390 μ L) in M9-N with glucose. In the anaerobic chamber, alkane substrate (5 μ L of 400 mM stock solution in EtOH), and ethyl diazoacetate (5 μ L of 400 mM stock solution in EtOH) were added in the listed order. Final reaction volume was 400 μ L; final concentrations were 5 mM alkane substrate, 5 mM ethyl diazoacetate, and 30 mM D-glucose. The vials were sealed and shaken at room temperature and 500 rpm for 18 hours in the anaerobic chamber. The expression of heme protein was measured using the hemochrome assay (Section I (G)), and the concentration of heme protein in the biotransformation was calculated accordingly. Reactions in figure S7 were run at OD30 with 10 mM of each reactant, and the volumes of each part were adjusted to maintain the concentrations of each other reaction constituent.

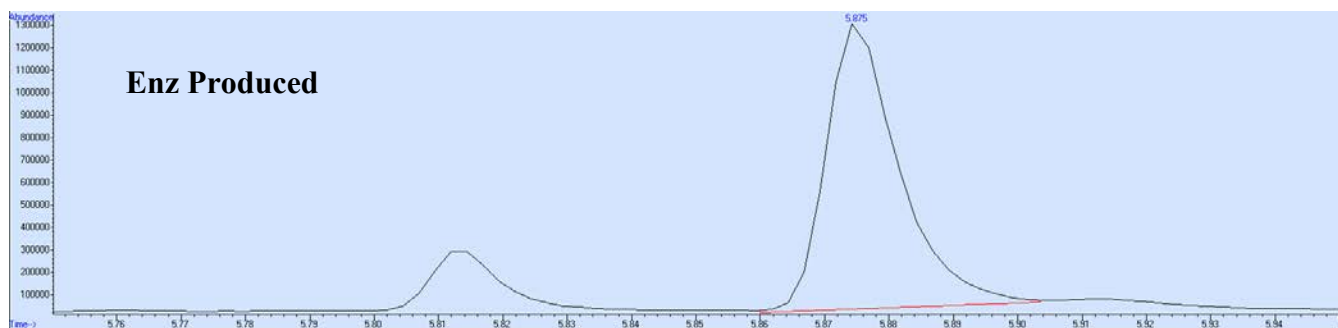
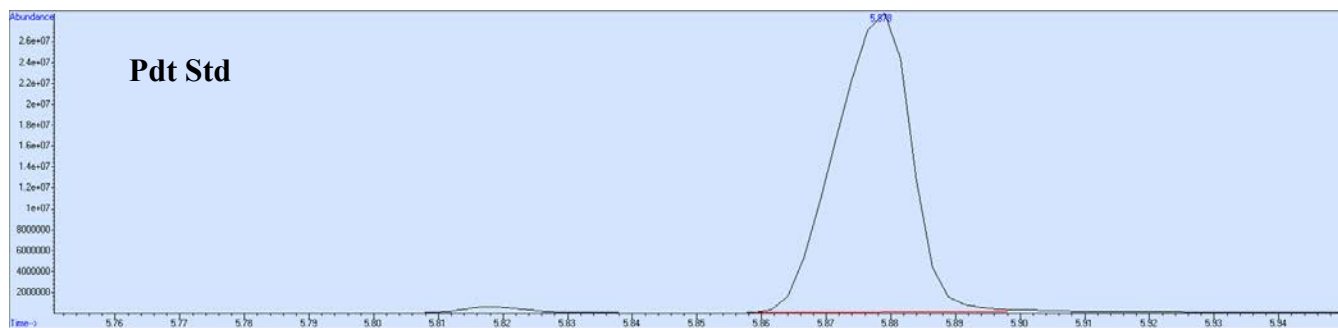
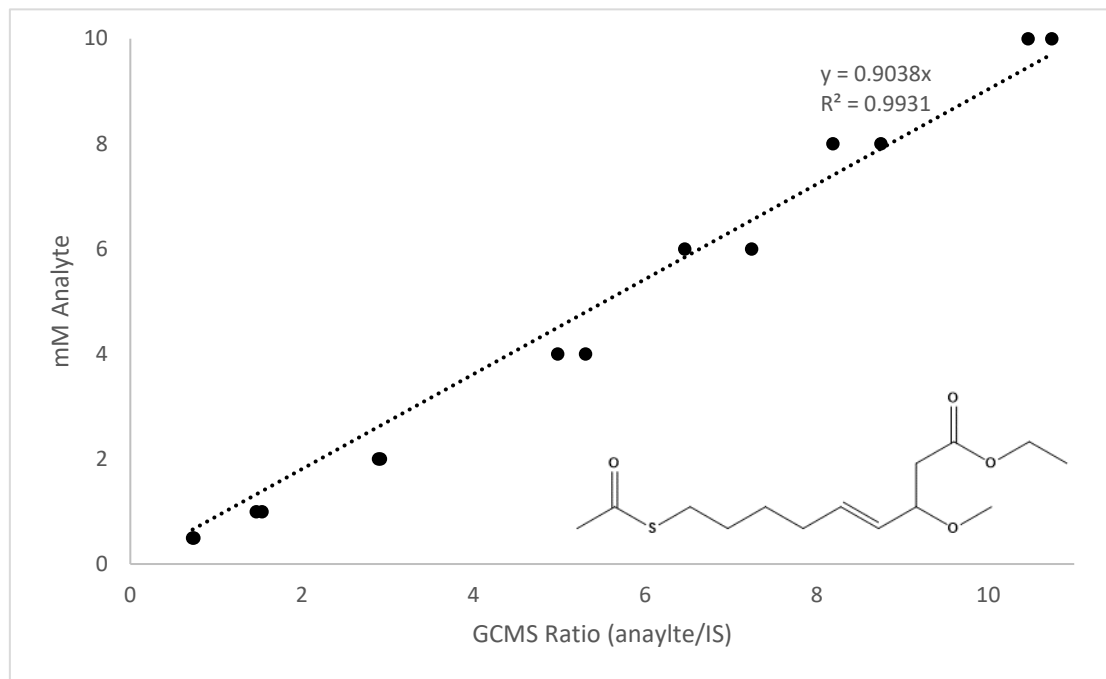
Reaction workup for quantitative GC analysis. Internal standard 1,3,5-trimethoxybenzene (20 μ L of 20 mM stock solution in cyclohexane) was added to the reaction vial followed by mixed solvent (cyclohexane / ethyl acetate = 1 : 1, 600 μ L). The mixture was transferred to a 1.5 mL microcentrifuge tube, vortexed (10 seconds, 3 times), and centrifuged (20,000 \times g, 5 minutes) to completely separate the organic and aqueous layers. The organic layer was taken for GC analysis.

GC calibration curve preparation

Stock solutions of chemically synthesized products at various concentrations (0.2 to 200 mM in EtOH) were prepared. To a microcentrifuge tube were added 360 μ L M9-N buffer, 20 μ L product stock solution, 20 μ L internal standard (20 mM 1,3,5-trimethoxybenzene in cyclohexane), and 600 μ L mixed solvent system (cyclohexane : ethyl acetate = 1:1). The mixture was vortexed (10 seconds, 3 times) then centrifuged (20,000 \times g, 5 min) to completely separate the organic and aqueous layers. The organic layer was removed for GC analysis. The standard curves plot product concentration in mM (y-axis) against the ratio of product area to internal standard area on the GC (x-axis).

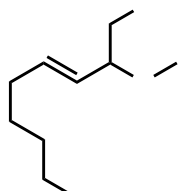
Ethyl (E)-9-(acetylthio)-3-methoxynon-4-enoate

GC calibration curve with 1,3,5-trimethoxybenzene as internal standard (IS)

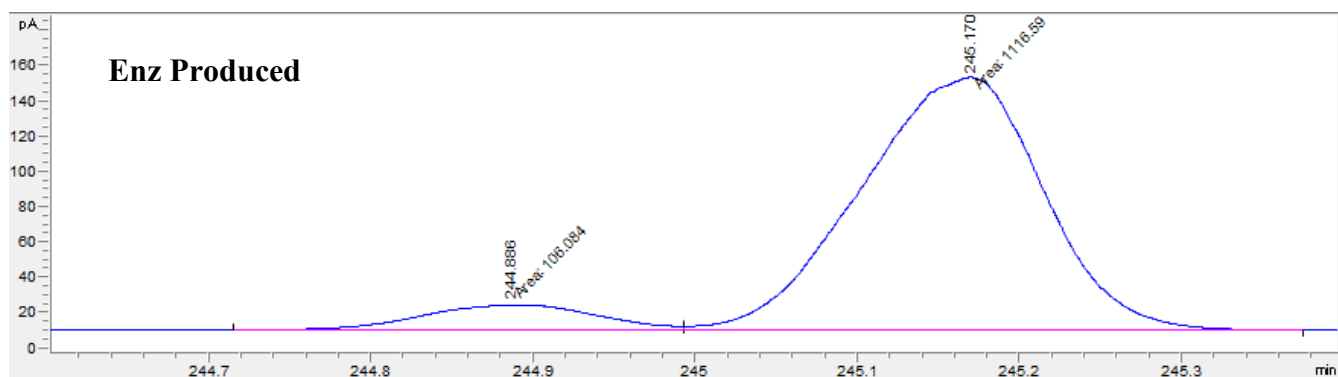
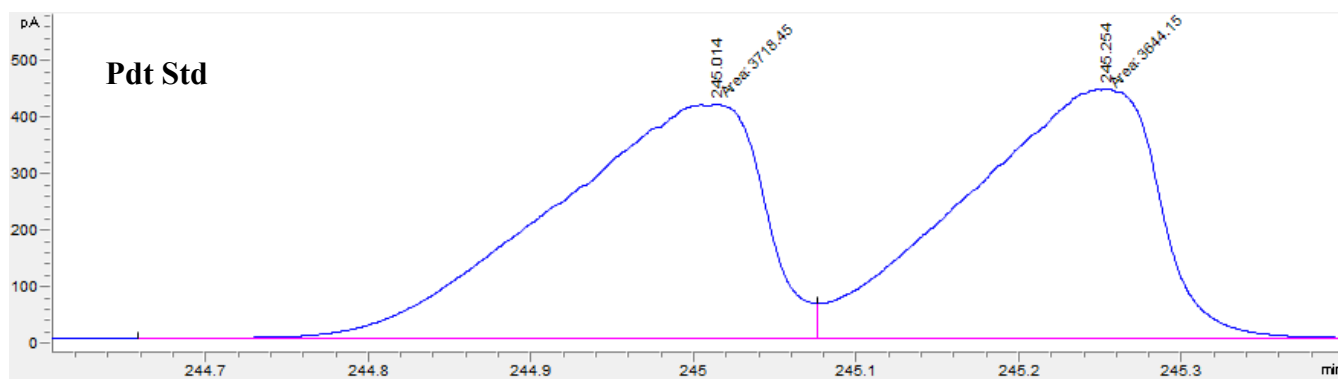


VIII. Determination of enantioselectivity

Enantioselectivity of the enzymatic C–H alkylation product was determined by chiral GC analysis. A representative trace, and conditions, are shown below. The absolute configuration of the synthesized product was assigned to be (S) by analogy to the chiral GC separation of Ethyl (E)-3-methoxydec-4-enoate reported in the literature.⁴

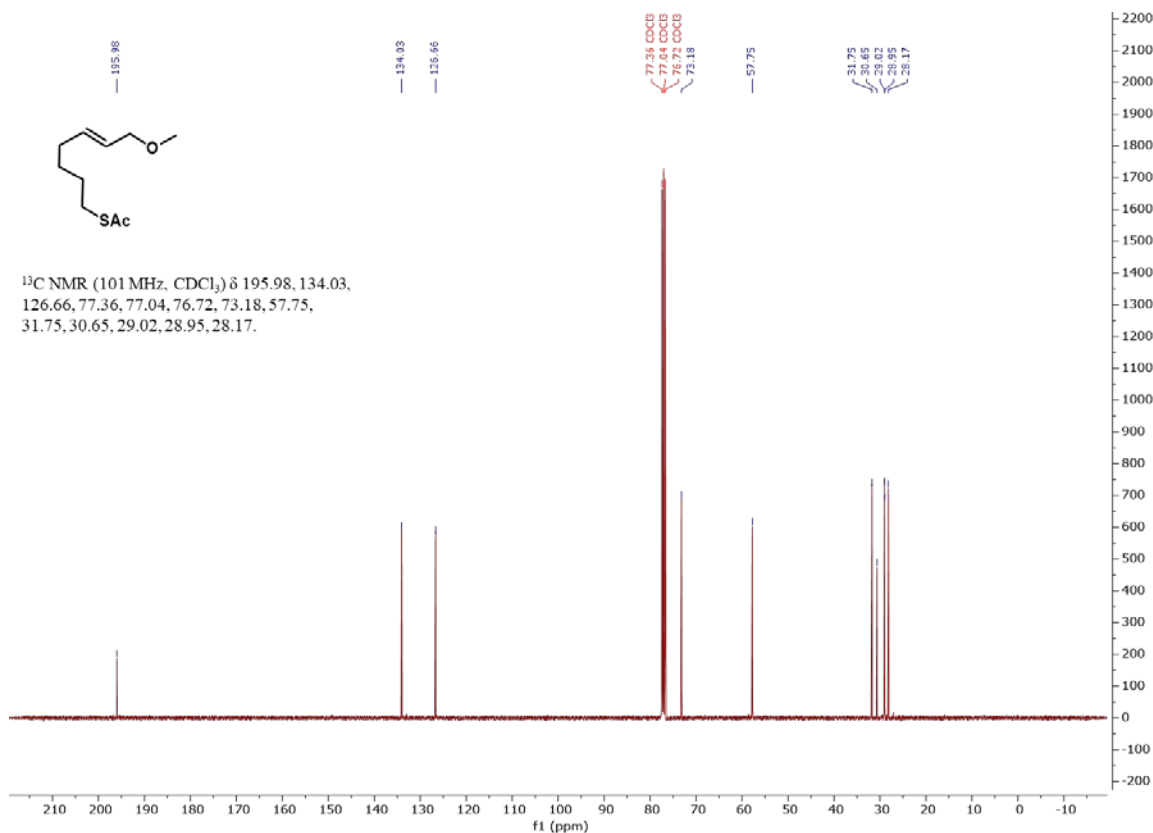
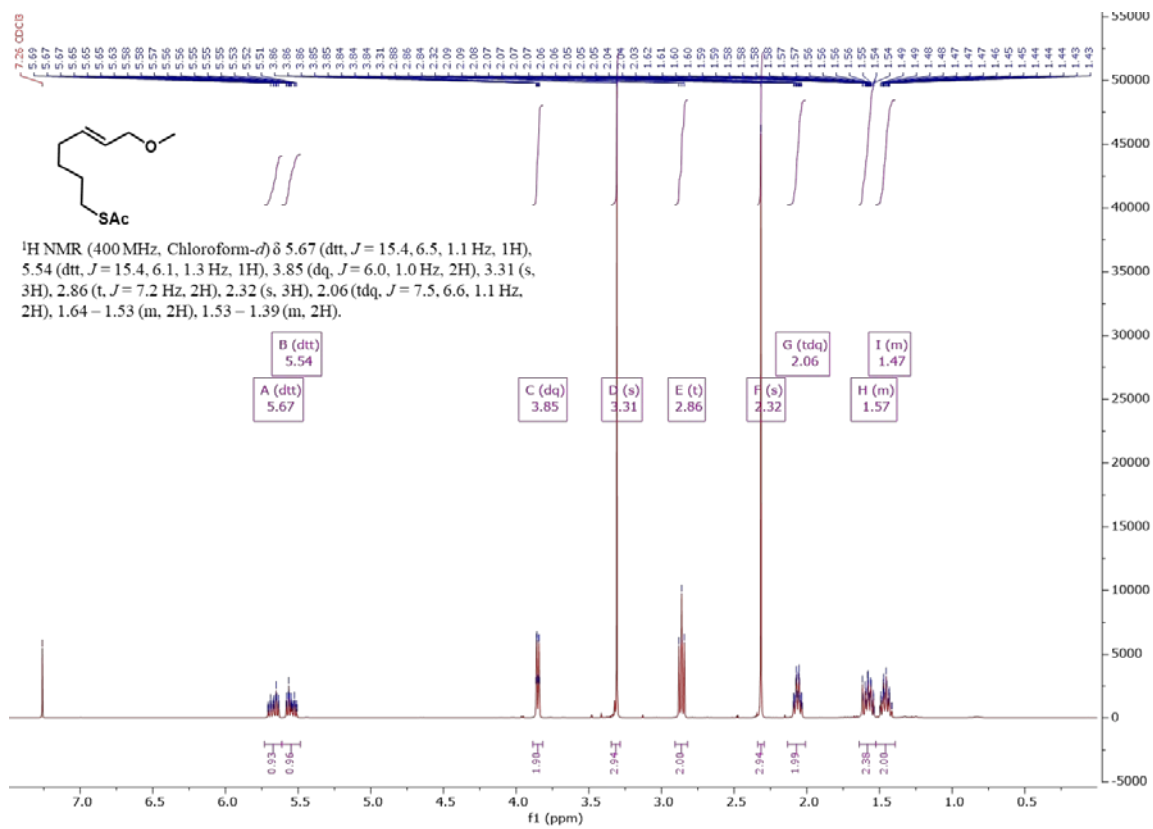
**Ethyl (E)-9-(acetylthio)-3-methoxynon-4-enoate**

Cyclosil-B column: 140 °C

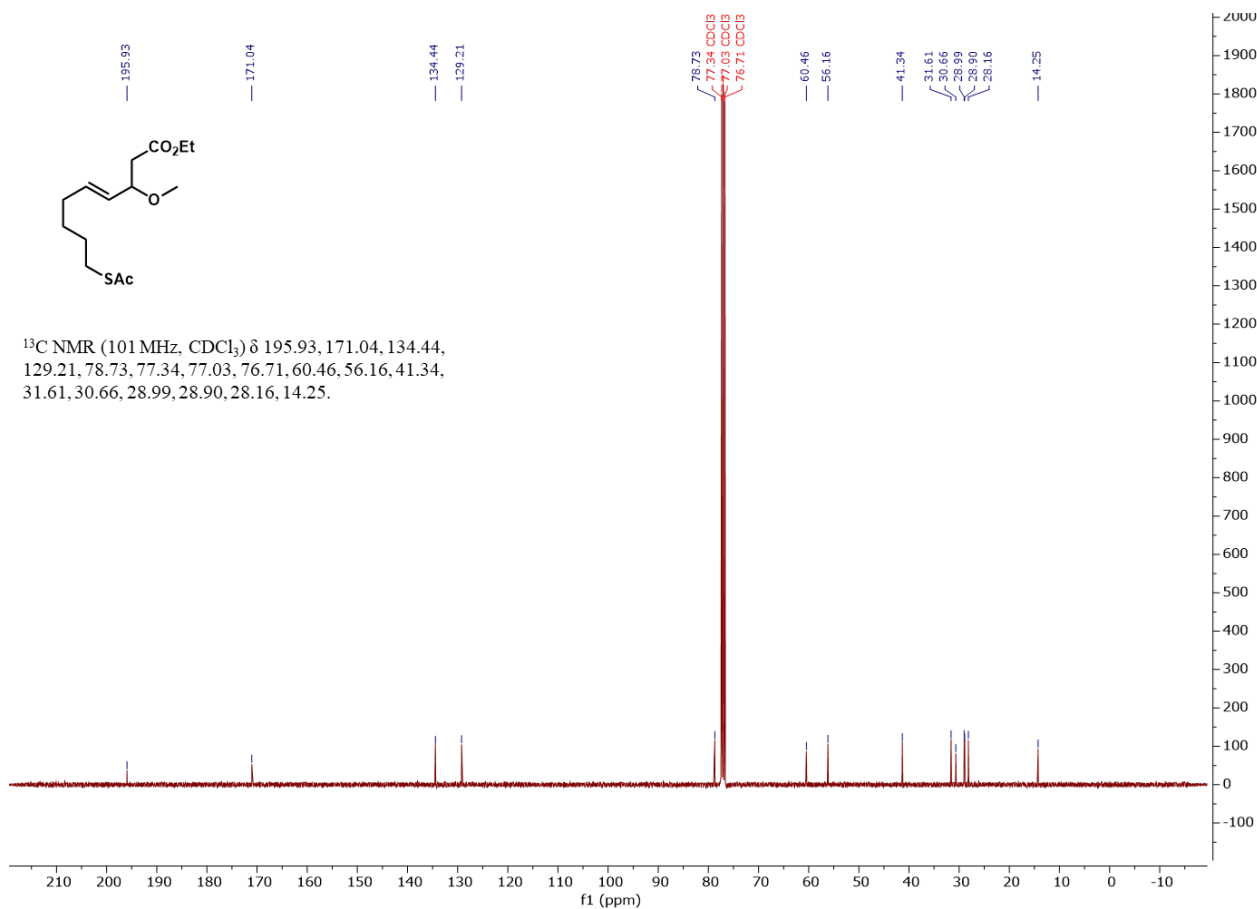
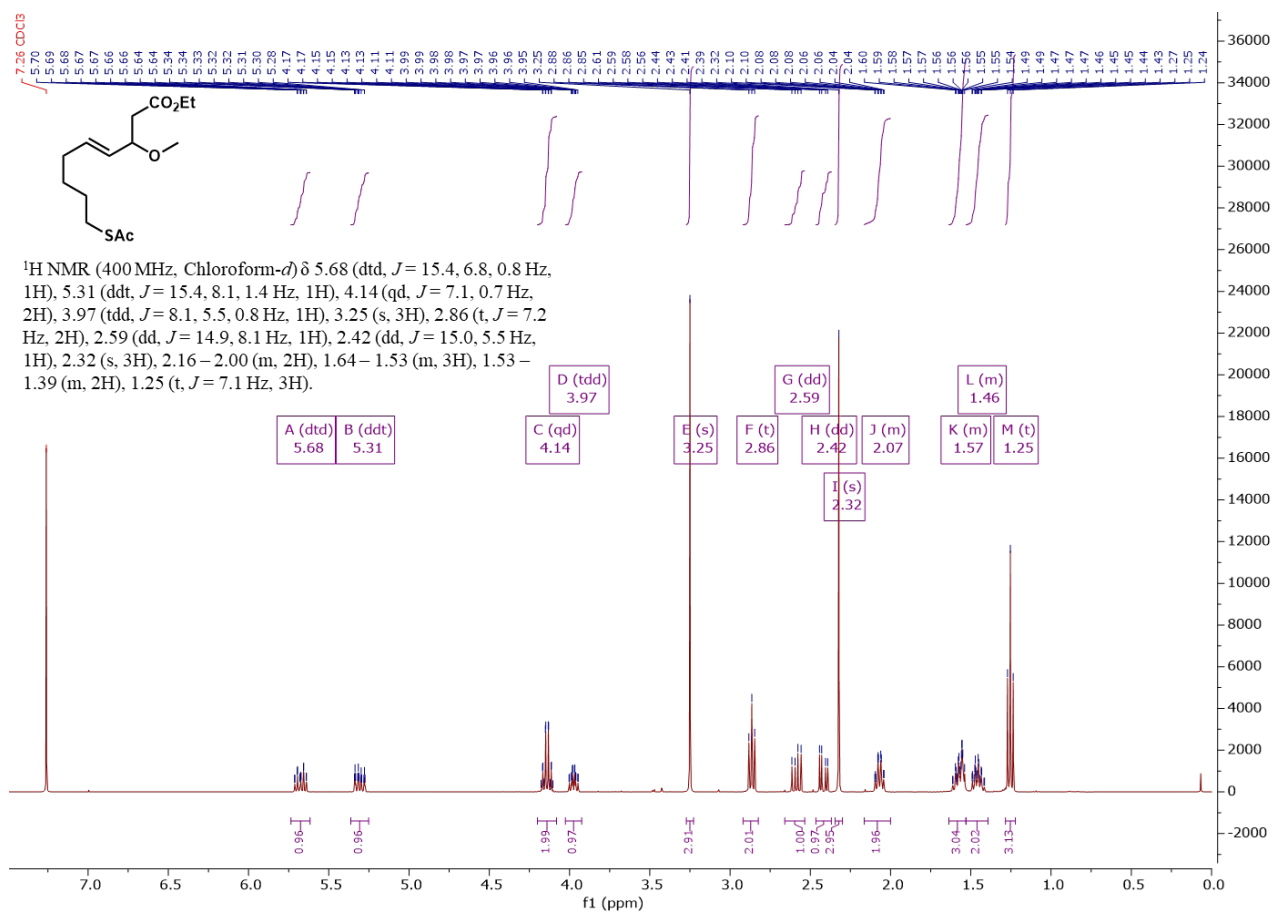


<i>rac</i>			P411-CHF +P74T		
Retention Time (min)	Area (mAu*s)	Area %	Retention Time (min)	Area (mAu*s)	Area %
245.014	3718.4	50.505	244.894	83.8	9.253
245.254	3644.2	49.495	245.168	821.4	90.747
Total	7362.6	100	Total	905.2	100
P411-CHF +P74T + I174T +Q553H			P411-CHF +P74T + I174T +Q553H +K639*		
Retention Time (min)	Area (mAu*s)	Area %	Retention Time (min)	Area (mAu*s)	Area %
244.886	106.1	8.676	244.889	152.4	8.739
245.17	1116.6	91.324	245.171	1591.1	91.261
Total	1222.7	100	Total	1743.5	100

IX. NMR Spectra



SUPPORTING INFORMATION



X. Supplemental references

1. Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchinson III, C. A. & Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345 (2009).
2. Berry, E. A. & Trumpower, B. L. Simultaneous determination of hemes a, b, and c from pyridine hemochrome spectra. *Anal. Biochem.* 161, 1–15 (1987).
3. Kan, S. B. J., Lewis, R. D., Chen, K. & Arnold, F. H. Directed evolution of cytochrome c for carbon–silicon bond formation: Bringing silicon to life. *Science* 354, 1048–1051 (2016).
4. Knight, A. M., Kan, S. B. J., Lewis, R. D., Brandenberg, O. F., Chen, K. & Arnold, F. H. Diverse engineered heme proteins enable stereodivergent cyclopropanation of unactivated alkenes. *ACS Cent. Sci.* 4, 372–377 (2018).
5. Hernandez, K. E., Renata, H., Lewis, R. D., Kan, S. B. J., Zhang, C., Forte, J., Rozzell, D., McIntosh, J. A. & Arnold, F. H. Highly stereoselective biocatalytic synthesis of key Cyclopropane Intermediate to Ticagrelor. *ACS Catal.* 6, 7810–7813 (2016).
6. Zhang, R. K.; Chen, K.; Huang, X.; Wohlschlager, L.; Renata, H.; Arnold, F. H. Enzymatic Assembly of Carbon–Carbon Bonds via Iron-Catalysed Sp³ C–H Functionalization. *Nature* 2019, 565 (7737), 67–72. <https://doi.org/10.1038/s41586-018-0808-5>.
7. Helal, K. Y.; Alamgir, A.; Berns, E. J.; Mrksich, M. Traceless Immobilization of Analytes for High-Throughput Experiments with SAMDI Mass Spectrometry. *J Am Chem Soc* 2018, 140 (26), 8060–8063. <https://doi.org/10.1021/jacs.8b02918>.