

Supporting Information

Modulating chemoselectivity in a Fe(II)/ α -ketoglutarate dependent dioxygenase for the oxidative modification of a non-proteinogenic amino acid

Fabian Meyer,^{+ [a]} Raphael Frey,^{+ [a][c]} Mathieu Ligibel,^[b] Emine Sager,^[b] Kirsten Schroer,^[b] Radka Snajdrova,^[b] and Rebecca Buller * ^[a]

^[a] Competence Center for Biocatalysis, Institute of Chemistry and Biotechnology, Zurich University of Applied Sciences, Einsiedlerstrasse 31, 8820 Wädenswil, Switzerland

^[b] Novartis Institutes for BioMedical Research, Global Discovery Chemistry, 4056 Basel, Switzerland

^[c] present address: Bioconjugates R&D, Lonza Ltd, Rottenstrasse 6, 3930 Visp, Switzerland

+ These authors contributed equally to the work.

*rebecca.buller@zhaw.ch

Table of contents

I.	Material & Methods	S2-S44
II.	Supporting figures and tables	S45-S64
III.	References	S65-S68

I. Materials & Methods

Materials:

All chemical and biological reagents were purchased from commercial sources and used without further purification. LB medium tryptone and yeast extract were purchased from Carl Roth (Karlsruhe, Germany). Buffer salts were purchased from Sigma-Aldrich (MO, USA). The substrate L-homophenylalanine (L-hPhe) was obtained from abcr GmbH (Karlsruhe, Germany). The analytical standards for (S)-2-amino-4-oxo-4-phenylbutanoic HCl and 2-amino-4-hydroxy-4-phenylbutanoic acid HCl were purchased from Sigma-Aldrich (MO, USA) and Enamine Ltd (Kiev, Ukraine), respectively. The cofactors sodium α -ketoglutarate, sodium ascorbic acid and ammonium iron sulfate hexahydrate, DNase I as well as derivatization agent dansyl chloride were all purchased from Sigma-Aldrich (MO, USA). Lysozyme and polymyxin B were obtained from Carl Roth (Karlsruhe, Germany). Phusion high-fidelity DNA polymerase, T4 DNA ligase and all restriction enzymes used in this study were purchased from New England BioLabs (MA, USA). Oligonucleotides were synthesized by Microsynth AG (Balgach, Switzerland). The gene for the wild type Smp4H and other α -KGD enzymes were synthesized by Twist Bioscience (CA, USA)

Experimental methods:

Plasmid construction and protein expression: Genes encoding alpha-ketoglutarate dependent dioxygenases in pET28b+ were purchased from Twist Bioscience, for detailed information see table below. Each plasmid was transformed into BL21(DE3) *E. coli* and the cells were plated on a LB agar plate containing 50 μ g/ml kanamycin. A single colony of freshly transformed cells was cultured overnight in 3 ml of LB medium containing 50 μ g/ml kanamycin. 0.1 ml of the culture was used to inoculate 0.9 mL of Zymo5052 auto-induction medium supplemented with 50 μ g/ml kanamycin in a 96-well deep well plate. Expression was carried out for 24 h at 20 °C, 300 rpm (5-cm shaking diameter) using a Duetz system (Kuhner AG, Basel, Switzerland). The cells were pelleted by centrifugation at 4000g, 4 °C, for 15 min, and the supernatant was discarded. The cell pellet was stored in -80°C freezer prior to biotransformation reactions.

DNA sequence of Smp4H (insert of plasmid pET28_Smp4H):

```
ATGAGCACCCATTTTCTGGGCAAAGTGAAATTTGATGAAGCACGTCTGGCAGAAGATCT
GAGCACCTGGAAGTTGCAGAATTTAGCAGCGCATATAGCGATTTTGCATGTGGTAAATG
GGAAGCATGTGTTCTGCGTAATCGTACCGGTATGCAAGAAGAAGATATTGTTGTTAGCCA
TAATGCACCGGCACTGGCAACACCGCTGAGCAAAAGCCTGCCGTATCTGAATGAACTGG
```

TTGAAACCCATTTTATTGATTGTAGCGCAGTTCGTTATACCCGTATTGTTTCGTGTTAGCGAAAA
TGCCTGTATTATTCCGCATAGCGATTATCTGGAAGTGGATGAGACATTTACCCGTCTGCA
TCTGGTTCTGGATACCAATAGCGGTTGTGCAAATACCGAAGAGGATAAAAATCTTTCACAT
GGGTCTGGGCGAAATTTGGTTTCTGGATGCAATGCTGCCGCATAGTGCAGCATGTTTTAG
CAAACACCCGCGTCTGCACCTGATGATTGATTTTGAAGCAACCGCATTTCGGGAATCATT
TCTGCGCAATGTTGAACAGCCGTTACCACACGTGATATGGTTGATCCGCGTAAAGAACT
GACCGATGAAGTTATTGAAGGTATTCTGGGCTTTAGCATCATTATTAGCGAAGCCAATTA
TCGCGAAATCGTTAGCATTCTGGCAAACACTGCACTTCTTCTATAAAGCAGATTGCCGCAG
CATGTATGATTGGCTGAAAGAAATTTGTAAACGTCGTGGTGATCCGGCACTGATTGAAAA
AACCGCAAGCCTGGAACGTTTTTTTTCTGGGTCATCGTGAACGTGGTGAAGTTATGACCTA
T

Amino acid sequence of SmP4H:

MSTHFLGKVKFDEARLAEDLSTLEVAEFSSAYSDFACGKWEACVLRNRTGMQEEDIVVSHN
APALATPLSKSLPYLNELVETHFDCSAVRYTRIVRVSENAIIPHS DYLELDETFTRLHLVLDL
NSGCANTEEDKIFHMGLGEIWFLDAMLPHSAACFSKTPRLHLMIDFEATAFPESFLRNVEQP
V
TTRDMVDPRKELTDEVIEGILGFSIISEANYREIVSILAKLHFFYKADCRSMYDWLKEICKRR
GDPALIEKTASLERFFLGHRERGERGEVMTY

Site-saturation mutagenesis library construction: NNK libraries were generated by overlap extension PCR using pET28_SmP4H as a template. Two overlapping fragments were generated by PCR using the primer pairs pET28(+)/aaX(-) and aaXNNK(+) and pET28(-) or pET28_outer(-); the fragment R272NNK was cloned using primer pET28_outer(-) instead of pET28(-) (see Table S2). The resulting fragments were purified by agarose gel electrophoresis (AGE) and subjected in equimolar amounts to assembly PCR using the outer primers pET28(+)/pET28(-). The PCR reaction was purified by a PCR clean-up kit and the full-length mutated fragment was digested using *Xba*I and *Xho*I for 2h at 37°C. The digested fragment was ligated with correspondingly cut vector pET28 at a 5:1 ratio at 16°C overnight. Subsequently, the ligation mixture was incubated at 65°C for 10 min to inactivate the ligase.

The NNK libraries were transformed into *E.coli* NEB-10 cells by adding 10 µL of the ligation mixture to ca. 100 µL of chemically competent cells, followed by a short heat-shock (42°C for 40 s). The transformation mixtures were plated on Agar plates supplemented with kanamycin and incubated overnight at 37°C. In order to analyze library quality, the resulting colonies were pooled, the plasmid library isolated and sequenced.

The isolated NNK libraries were then re-transformed into *E. coli* BL21(DE3) using ca. 50 ng of DNA for 100 µL of competent cells. 89 colonies of each library were picked and used to inoculated 400 µL

of LB medium (supplemented with 50 µg/mL kanamycin) in a 2-mL 96-deep-wellplate. Additionally, four wells were inoculated with WT transformants and two wells with an empty pET28-vector controls. The plate was incubated at 37°C overnight. 10 µL of the overnight cultures were used to inoculate 1 mL Zymo-5052 auto-induction medium supplemented with 50 µg/mL kanamycin in 96-deep-wellplate. Expression was carried out for 60 – 70 h at 20 °C at 300 rpm (5-cm shaking diameter) using a Duetz system (Kuhner AG, Basel, Switzerland). The cells were harvested by centrifugation (4000 g, 20 min, 4°C). The supernatant was discarded and the pellets stored at -80°C until use.

Library construction for 2-site randomization at W40 and I103: Beneficial mutations at positions W40 and I103 identified in the site-saturation mutagenesis libraries were recombined in 2-site randomization libraries. Additional amino acids that improved the desired enzymatic activity > 2-fold were identified as G, C and T at position W40 and A and S at position I103. As a consequence, degenerate codon DBK¹ at position W40, encoding amino acids A, R, C, G, I, L, M, F, S, T, W and V, and degenerate codon DYA at position I103, encoding amino acids A, I, L, S, T, V, were chosen to build a two-site combinatorial library in a second round of engineering (for screening results, see Figure S4). First, three PCR fragments were generated using the primer pairs pET28(+)/W40(-), W40DBK(+)/I103(-) and I103DYA(+)/pET28(-) (see Table S2). The fragments were assembled by assembly PCR and cloned into pET-28 vector as described above. Again, the library were first transformed into *E.coli* NEB-10 to analyze the quality of the pooled library by sequencing before transformation into BL21 (DE3) for expression. Colony picking, inoculation of overnight culture and expression were carried out as described for the site-saturation mutagenesis libraries, whereas three wells were inoculated with W40M transformants as control.

Crude-lysate screening assay: Crude lysate screening assays were carried out according to a previously published patent with some slight modifications². In brief, the 96-deep-wellplates containing the expressed enzyme libraries were thawed at RT and 100 µL of lysis buffer (50 mM Na-phosphate, pH 6.3 containing 1 mg/mL lysozyme, 0.5 mg/mL Polymyxin B and spatula tip of DNase I) was added to each well. The plates were sealed with breathable membranes and incubated at 25°C in an Eppendorf plate shaker at 850 rpm for 30-60 min until the lysates became homogenous. A 2x concentrated co-factor mixture was prepared, containing sodium ascorbate (250 mM) and sodium α -ketoglutarate (270 mM) in Na-phosphate buffer, pH 6.3. Immediately before use, ammonium iron(II) sulfate (300 mM in deionized water) and L-hPhe (400 mM in 1 M NaOH) were diluted 100-fold into the co-factor stock. The biotransformation was initiated by adding 100 µL of the co-factor/substrate mixture to each well. The plates were sealed with a breathable membrane and incubated at 20°C, shaking at 900 rpm for 16-20 h.

For analysis by RP-HPLC-MS, the biotransformations were subjected to derivatization with dansyl chloride. Briefly, 440 μ L of a solution of NaHCO₃ (5% in deionized water) was distributed into a new 96-deep-wellplate. 40 μ L of the biotransformation mixtures were transferred to each well, followed by 320 μ L of a freshly prepared solution of dansyl chloride (6 mg/mL in MeCN). The plates were tightly sealed with aluminium foil covers and incubated at 55°C at 800 rpm for 3-4 h. Subsequently, the plates were centrifuged at 4400g for 20 min and the supernatants were analysed on an Agilent 1260 HPLC system equipped with a single quadrupole MSD using an Agilent Poroshell 120 EC-C18 column (2.7 μ m 3.0 x 50 mm plus 3.0 x 5 mm) heated at 40 °C. Water/MeCN (95/5% + 0.2% formic acid) and MeCN (+ 0.2% formic acid) were used as solvent A and B, respectively. The following method was used for screening: 0-1 min, B = 5%; 1-6 min, B = 5-95%; 6-8 min, B = 95%; 8-10 min, B = 5%. Fold improvement over parent was calculated monitoring the different products in the selective ionization mode and normalizing the peak area to the averaged value obtained for the WT or W40M in case of the W40 I103 site directed mutagenesis library. Measurement of controls were included on each plate. Improved variants were re-inoculated from the library masterplates and analyzed by sequencing.

Preparative scale biotransformation and derivatization: For structural characterization of the enzymatic products, the crude lysate biotransformation was scaled-up in order to obtain enough material for isolation by preparative RP-HPLC and subsequent NMR analysis. *E. coli* BL21 (DE3) carrying the genes for engineered variants W40M and W40Y was streaked-out from a glycerol stock and a single colony was used to inoculate 5 mL of LB medium (supplemented with kanamycin). The overnight culture was used to inoculate 0.5 L Zymo-5052 autoinduction medium in a 2 L culture flask. The cultures were incubated at 25°C and 160 rpm for 20 h. The cells were harvest by centrifugation upon reaching an OD600 of approx. 14 and the pellets were stored at -20°C until use. For the preparative scale biotransformation, the expression pellets (ca. 7-8 g wet cell mass) were resuspended in 30 mL lysis buffer. The mixtures (ca. 35 mL) were shaken slightly at room temperature and then transferred to a 250 mL Erlenmeyer flask without baffles. 40 mL of the 2x concentrated cofactor mixture were supplemented with 0.4 mL of a solution of ammonium iron(II) sulfate (300 mM in water). Then, 35 mL of this Fe(II)-supplemented co-factor mixture were immediately added to the lysates in the Erlenmeyer flask, followed by addition of 0.7 mL L-hPhe (400 mM in 1 M NaOH). The flasks were sealed with a breathable membrane and incubated in at 20°C at 140 rpm for 3 h.

The biotransformation with W40Y was directly subjected to extraction without prior derivatization. The reaction mixture was first washed with EtOAc (80 mL) and then the aqueous phase was acidified with 50% HCl to pH 0.9. The mixture was saturated with NaCl and extracted 3x with THF (100 mL). The organic layer was washed with brine, dried over Na₂SO₄ and concentrated *in vacuo* to give compound 5 as a crude oil.

Due to the higher hydrophilicity of the hydroxylated compounds, a different procedure was followed for the isolation of the products from the biotransformation with W40M, which involved a derivatization with dansyl-chloride prior extraction. The biotransformation was quenched by adding acetonitrile (70 mL). The resulting mixture was distributed into 50 mL falcon tubes and centrifuged 20 min at maximum speed to remove precipitated material. The supernatant was transferred to a 250 mL round-bottom flask equipped with a magnetic stir bar and the pH was carefully adjusted to 8.5 with 1 M NaOH. The flask was protected from light using aluminium foil and 378 mg of dansyl-chloride powder was added, corresponding to 5 equivalents with respect to the amount of L-hPhe used in the biotransformation. The flask was loosely capped with a stopper and the reaction was stirred overnight at room temperature.

For work-up and extraction, acetonitrile was first removed on the rotary evaporator to facilitate phase separation during extraction. The remaining mixture was acidified to pH 2.0 using 50% HCl upon which the reaction mixture turned yellow. The mixture was transferred to a separation funnel and extracted once with ethylacetate (80 mL). The organic phase was washed with brine, dried over Na₂SO₄ and concentrated *in vacuo* to give a brown oil. The crude extract can be stored in the fridge but should be further purified as soon as possible.

For purification by preparative RP-HPLC, the crude extract was taken up in 3 mL acetonitrile. Some insoluble material was removed by filtration through a 0.45 µm filter, followed by injection onto preparative C-18 column (Kromasil 100-10-C18, 30 x 250 mm) using a Gilson PLC 2020 chromatography system. Water/acetonitrile (95/5% + 0.2% acetic acid) and Water/Acetonitrile (5/95% + 0.2% acetic acid) were used as mobile phases A and B and standard gradient of 10-90% was used. Fractions were analysed by LC-MS and product-containing fractions pooled and lyophilized. Yellow residues were obtained for the derivatized compounds 2,4 and 6, which were subjected to further purification.

The following procedures were performed prior to NMR analysis:

For compound 5, carboxybenzyl derivatization was performed instead of dansyl chloride derivatization. To a glass vial equipped with a magnetic stirrer the raw isolated compound 5, 1 mL THF, 10 mL of a saturated NaHCO₃ solution and 89 µL of benzyl chloroformate (2 eq. considering the starting material) were added. The reaction mixture was stirred at room temperature for 30 minutes, acidified to pH = 1 with 25% HCl and extracted with 100 mL ethyl acetate. The organic layer was dried over MgSO₄, concentrated to dryness and directly injected on a preparative HPLC.

On an Interchim PuriFlash 4250 HPLC equipped with DAD and with a VP 250/21 Nucleodur 100-10 C18ec column (5 µm 21 x 250 mm) using water and acetonitrile containing 0.1 % formic acid as solvent A and B, respectively. The following LC methods were used to purify the substrate and derivatives: 0-5 min, B = 0 %, 5-30 min, B = 0-70 % and 30-35 min, B=70 %).

For compounds 2 and 4, purification was carried out on a Waters system equipped with a Photodiode Array Detector 2998, a SQ Detector 2, a Sampler Manager 2767, a Binary Gradient Module 2545 and with a XBridge Shield column (5 μ m 19 x 250 mm) using water and acetonitrile containing 0.1 % formic acid as solvent A and B, respectively. The following LC methods were used to purify the substrate and derivatives: 0-0.5 min, B = 25 %, 0.5-24.1 min, B = 25-55 %, 24.1-27.5 min, B=55-100%.

For compound 6, purification was carried out on a Waters system equipped with a Photodiode Array Detector 2998, a SQ Detector 2, a Sampler Manager 2767, a Binary Gradient Module 2545 and with a XBridge Shield column (5 μ m 19 x 250 mm) using water and acetonitrile containing 0.1 % formic acid as solvent A and B, respectively. The following LC methods were used to purify the substrate and derivatives: 0-0.5 min, B = 35 %, 0.5-24.1 min, B = 35-55 %, 24.1-27.5 min, B=55-100%.

NMR measurements: All NMR measurements (1 H-1D, HSQC, COSY, ROESY and HMBC) for the assignment of all compounds in [D₆]DMSO were recorded at 300K on a Bruker 600 MHz Avance III spectrometer (600.13 MHz for 1 H and 150.90 MHz for 13 C) equipped with a 1.7 mm TCI CryoProbeTM with actively shielded z-gradients. All spectra were referenced according to the internal solvent signal (1 H: DMSO = 2.5 ppm and 13 C: DMSO = 39.5 ppm).

Data for 1 H NMR are reported as follows: chemical shift (δ ppm), multiplicity (d = doublet, t = triplet, q = quartet), coupling constant (Hz), integration. Data for 13 C NMR are reported as follows: chemical shift (δ ppm).

Cloning, expression and purification of His-tagged enzyme variants: The genes for the Smp4H variants (WT, W40M, W40M-I103L and W40Y) were amplified by PCR using the primer pair pET28(+)/pET28(-) and digested with *Nde*I and *Xho*I. A N-terminal His-tag was introduced by ligating the fragments into a pET28b-vector digested with the same restriction enzymes. The resulting gene product was confirmed by sequencing analysis and transformed into expression strain BL21(DE3) using a heat-shock protocol. 5 mL of an overnight culture grown at 37°C was used to inoculate 0.5 L of TB medium (supplemented with TB-salts and 50 μ g/mL kanamycin) in a 2 L culture flask. The culture was first grown at 37°C until an OD₆₀₀ of 0.6 was reached. The temperature was lowered to 25°C and expression was induced with 1 mM IPTG. The culture was incubated at 25°C overnight for 16-20 h shaking at 180 rpm. The cells were harvested by centrifugation (4200g, 4°C) and the pellet was stored at -20°C.

For protein purification, an expression pellet from a 0.5 L culture was thawed at room temperature. 30 mL of binding buffer (50 mM Tris-HCl pH 7.4, containing 500 mM NaCl, 20 mM imidazole and 10 mM β -mercaptoethanol) supplemented with 1 mg/mL lysozyme and a spatula tip of DNase I was added. The pellet was resuspended by shaking slightly at room temperature for ca. 20 min. Once

homogeneous, the cells were further lysed by 3 cycles of sonication at 50% amplitude for 1 min while keeping the lysate cold on ice. The lysate was cleared by centrifugation (8500 g) for 1 h at 4°C. The supernatant was filtered through a 0.45- μ m filter and subjected to Ni-NTA purification on an ÄKTA pure FPLC system using a HisTrap HP 5 ml (GE Healthcare, Massachusetts, USA), pre-equilibrated with binding buffer. After extensive washing with binding buffer, the protein was eluted by increasing the amount of elution buffer (50 mM Tris-HCl pH 7.4, containing 500 mM NaCl, 250 mM imidazole and 10 mM β -mercaptoethanol). The fractions containing the His-tagged protein were combined and concentrated to < 5 mL using Amicon Ultra-15 concentration units with a 10 kDa cut-off (Merck Millipore, MA). The concentrated protein fractions were subjected to buffer exchange on a HiTrap desalting column (3 x 5 mL columns, serially connected) using a 10 mM sodium phosphate buffer at pH 7.5 as the running buffer. For storage, the protein solution was concentrated to ca. 500 μ M, aliquoted, snap-frozen in liquid nitrogen and stored at -80°C. Protein purity was assessed by reducing SDS-PAGE analysis.

In-vitro activity assay with purified enzyme: *In vitro* activity assays were carried out in 200 μ l of 50 mM sodium phosphate pH 6.3, containing 20 μ M purified enzyme, 0.5 mM substrate, 50 mM α -ketoglutaric acid sodium salt, 10 mM sodium ascorbate, and 0.1 mM ammonium iron(II) sulfate. Enzyme assay was incubated 60 min at 20 °C on a shaking incubator at 900 rpm. Reaction was quenched by transferring 20 μ L of reaction mixture into 80 μ L acetonitrile. Prior to analysis by RP-HPLC-MS, the quenched reaction mixture was derivatized with dansyl chlorid by adding 220 μ L of a NaHCO₃ solution (5% in deionized water) followed by 80 μ L of a freshly prepared solution of dansyl chloride (12 mg/mL in acetonitrile) and incubation at 55°C on a shaking incubator at 900 rpm. Subsequently, the samples were centrifuged at 10'000g for 3 min at room temperature and the supernatants were analysed by LC-MS using extracted ion monitoring as described above. Conversion was quantified comparing to a negative control containing no enzyme, assuming same ionization efficiency of all products and substrates.

To determine the apparent k_{cat} , K_M and k_{cat}/K_M values (where possible), assays were carried out in an identical manner but with less enzyme (1 μ M) as described above except that reactions were performed at 6 to 7 different substrate concentrations (0.25, 0.5, 1.0, 1.5, 2.0, 3.0 and 4.0 mM). At indicated time points (1, 2, 3, 4, and 5 min for W40M and W40M I103L and 5, 10, 15 and 20 min for wild type SmP4H), 20 μ L of reaction mixture was transferred into 80 μ L of acetonitrile to quench the reaction followed by derivatization with dansyl chloride as explained above. The product formation was monitored by LC-MS, and the peak area was plotted over time, which was then fitted by linear regression using excel. Product formation was calculated using calibration curves, prepared from known concentrations of the products (purchased standards for compound 2 and 3), whereas product formation of 4 was estimated using a calibration curve of compound 3, assuming same ionization efficiency as for

compound 3. For all variants and reactions, the observed initial rates were fitted to the Michaelis–Menten equation using R.

To determine the total turnover number for the different substrate-enzyme combinations, assays were carried out as described above using 20 μM purified enzyme and 2 mM of substrate and incubation time of 2 hours before derivatization. Product formation was calculated using calibration curves, prepared from purchased standards of compound 2 with known concentrations.

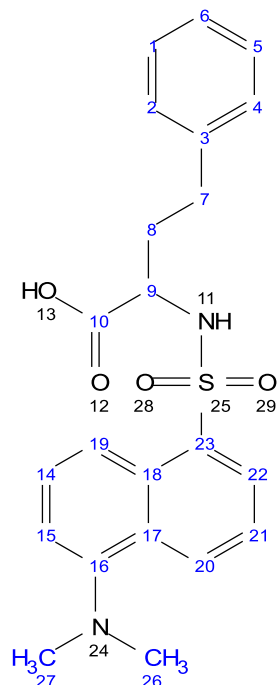
In case of substrate scope screen (Figure 5), *in vitro* activity assay with purified enzyme were carried out as described above using 20 μM purified enzyme (wild type and W40M I103L) and 1 mM of substrate, due to low solubility of (S)- α -amino-3,4-dichlorobenzenebutanoic acid. The reactions were incubated for 2 hours before derivatization and further analysis.

Lactonization: Lactonization of compound 2 was performed starting from derivatized reactions obtained from *in vitro* activity assay with purified enzyme KGOH8 W40M I103L as described above. In brief, *in vitro* assays were carried out in triplicates with 20 μM purified enzyme and 0.5 mM L-homophenylalanine as substrates followed by derivatization with dansyl chlorid. Derivatized reaction mix was centrifuged at 10'000 g for 3 min at RT and 300 μl supernatant was transferred into a round bottom flask evaporated under vacuum. Obtained residue was dissolved in 500 μl methanol and centrifuged at 10'000 g for 3min at RT. Supernatant was transferred into a new tube and 50 μl 37% HCl was added and incubated at room temperature for 2 hours at 900 rpm. Samples were then measured via LCMS to evaluate conversion into lactone 6.

Molecular docking analysis: The homology model of SmP4H and its respective variants have been obtained using SwissModel³. All docking simulations were based on this homology model. The protein structure was prepared by Chimera Dock Prep tool⁴, prior to molecular docking by SwissDock⁵. Energy minimization of the ligand has been performed prior to docking using MM2 energy minimization using Chem3D. The docking process was performed using default parameters. In case of the docking simulations with SmP4H WT and variant W40Y, the search grid for docking analysis was confined to the active site. Each docking result was visually inspected using UCSF Chimera software⁴.

Characterization of compounds:

Compound 1:

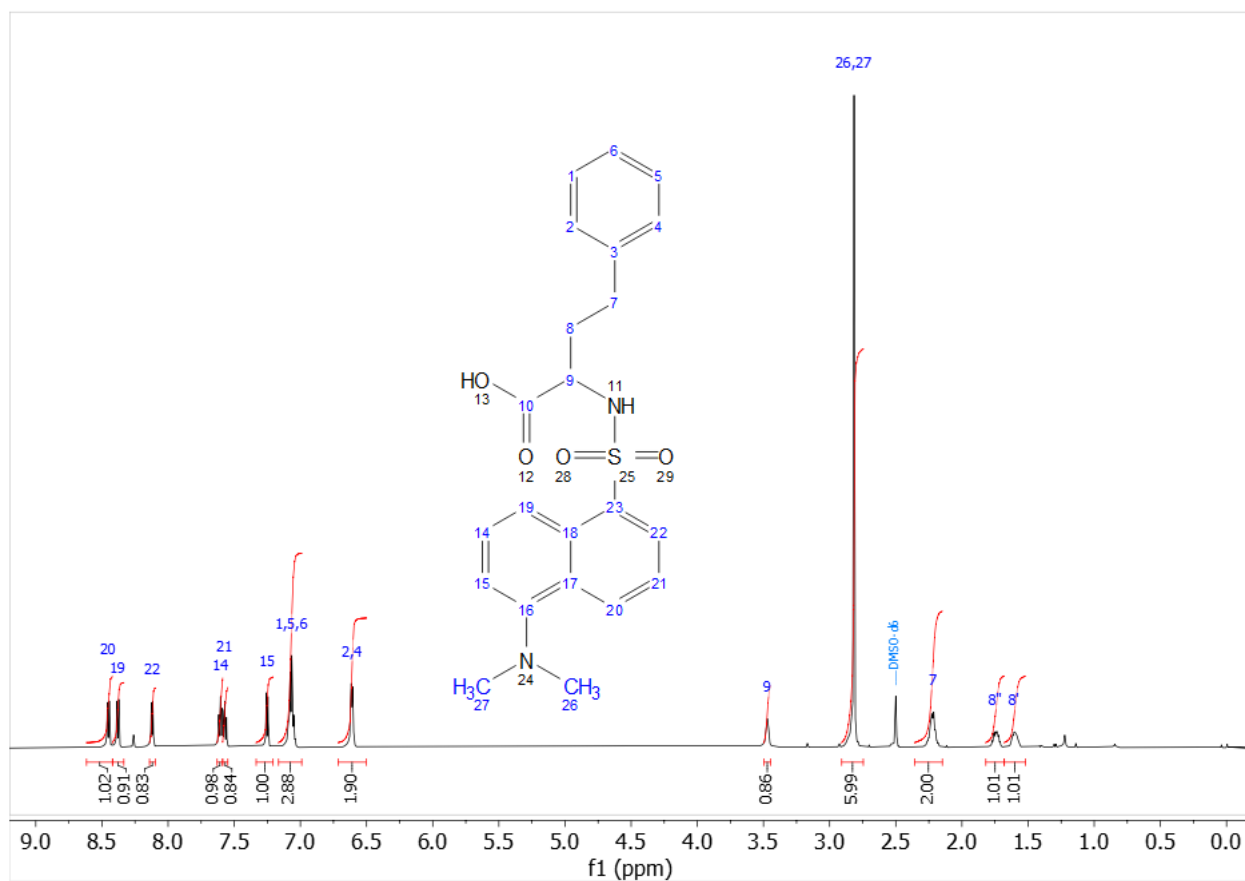


¹H NMR (600 MHz, DMSO) δ 8.45 (d, J = 8.5 Hz, 1H, 20), 8.38 (d, J = 8.7 Hz, 1H, 19), 8.12 (d, J = 7.1 Hz, 1H, 22), 7.60 (t, J = 8.0 Hz, 1H, 14), 7.57 (t, J = 8.0 Hz, 1H, 21), 7.25 (d, J = 7.5 Hz, 1H, 15), 7.16 – 6.99 (m, 3H, 1, 5, 6), 6.61 (d, J = 7.2 Hz, 2H, 4), 3.47 (t, J = 5.3 Hz, 1H, 9), 2.82 (s, 6H, 26), 2.36 – 2.15 (m, 2H, 7), 1.74 (ddt, J = 12.6, 9.8, 6.6 Hz, 1H, 8''), 1.60 (ddd, J = 14.4, 9.4, 5.3 Hz, 1H, 8').

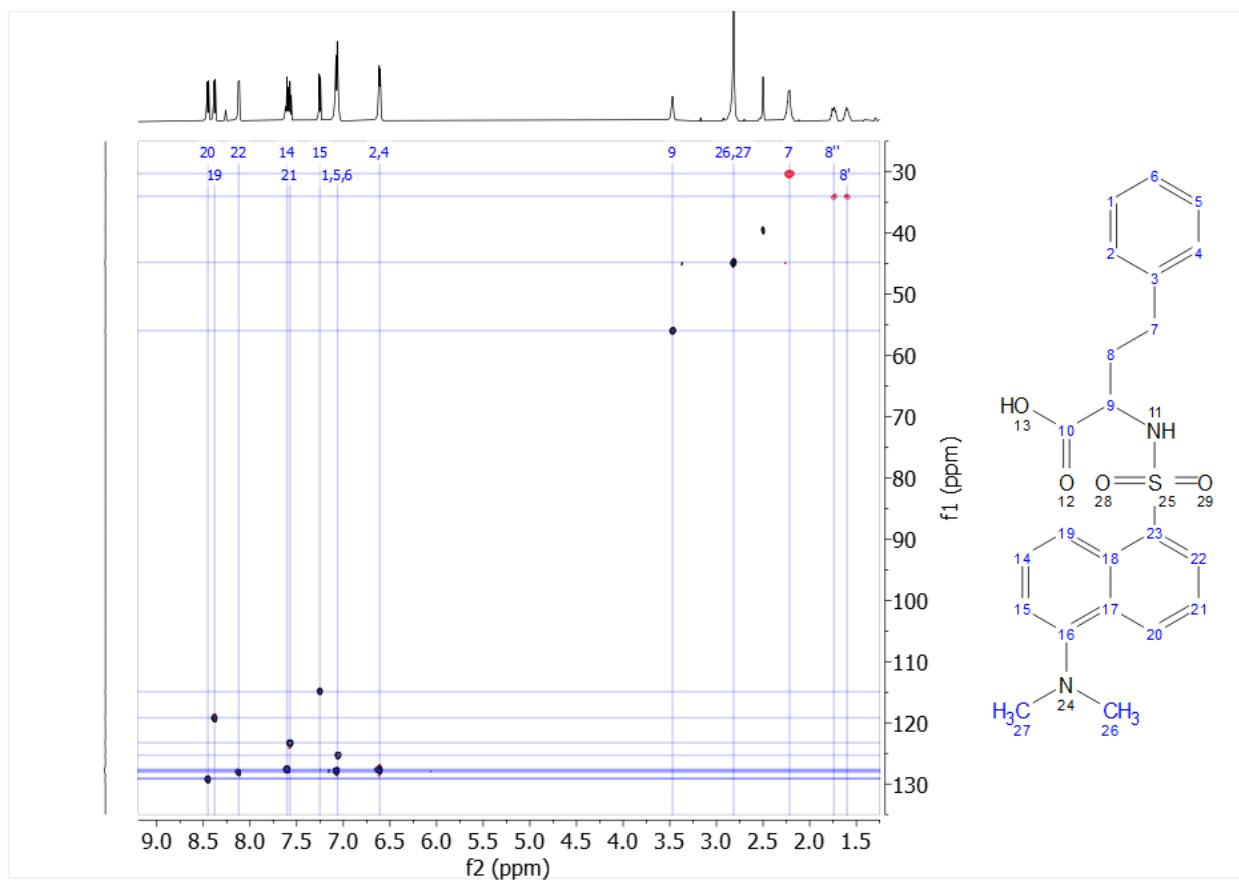
2D NMR assignment:

Assignment	H- δ (ppm)	C- δ (ppm)	splitting	J (Hz)	^1H - ^1H COSY	^1H - ^1H ROESY	^1H - ^{13}C HMBC	# H
CH-1/5	7.06	127.8	m		H-6, H-2/4		C-3, C-1/5	2
CH-2/4	6.61	127.7	d	7.2	H-1/5	H-7, H-8	C-7, C-6, C-2/4	2
C-3		141.4						
CH-6	7.06	125.3	m		H-1/5		C-2/4	1
CH ₂ -7	2.22	30.3	m		H-8	H-2/4	C-8, C-9, C-2/4, C-3	2
CH ₂ -8'	1.60	34.0	ddd	14.5, 9.4, 5.3	H-7, H-9	H-2/4	C-7, C-9, C-3, C-10	1
CH ₂ -8''	1.74	34.0	ddt	12.6, 9.8, 6.6				1
CH-9	3.47	55.9	t	5.3	H-8		C-7, C-8, C-10	1
C-10		172.8						
CH-14	7.60	128.0	T	8.0	H-19, H-15		C-18, C-16	1
CH-15	7.25	114.9	d	7.5	H-14	H-26/27	C-19, C-20, C-16	1
C-16		151.2						
C-17		136.4						
C-18		129.2						
CH-19	8.38	119.2	d	8.7	H-14		C-15, C-23, C-17	1
CH-20	8.45	129.1	d	8.5	H-21	H-26/27	C-18, C-16, C-22	1
CH-21	7.57	123.3	t	8	H-20, H-22		C-23, C-17	1
CH-22	8.12	128.1	d	7.1	H-21		C-20, C-18	1
C-23		129.0						
CH ₃ -26/27	2.82	44.78	s			H-20, H-15	C-16	6

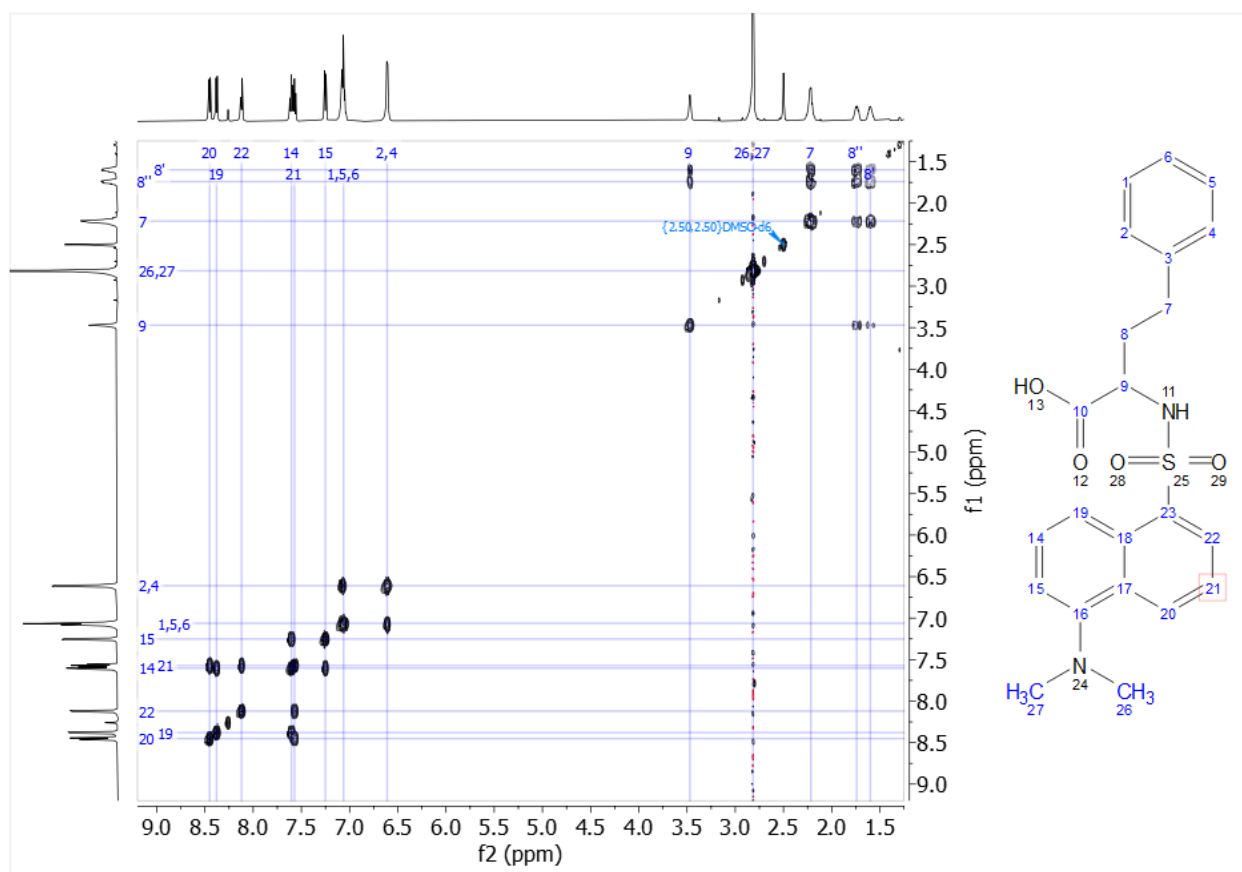
- 1D NMR



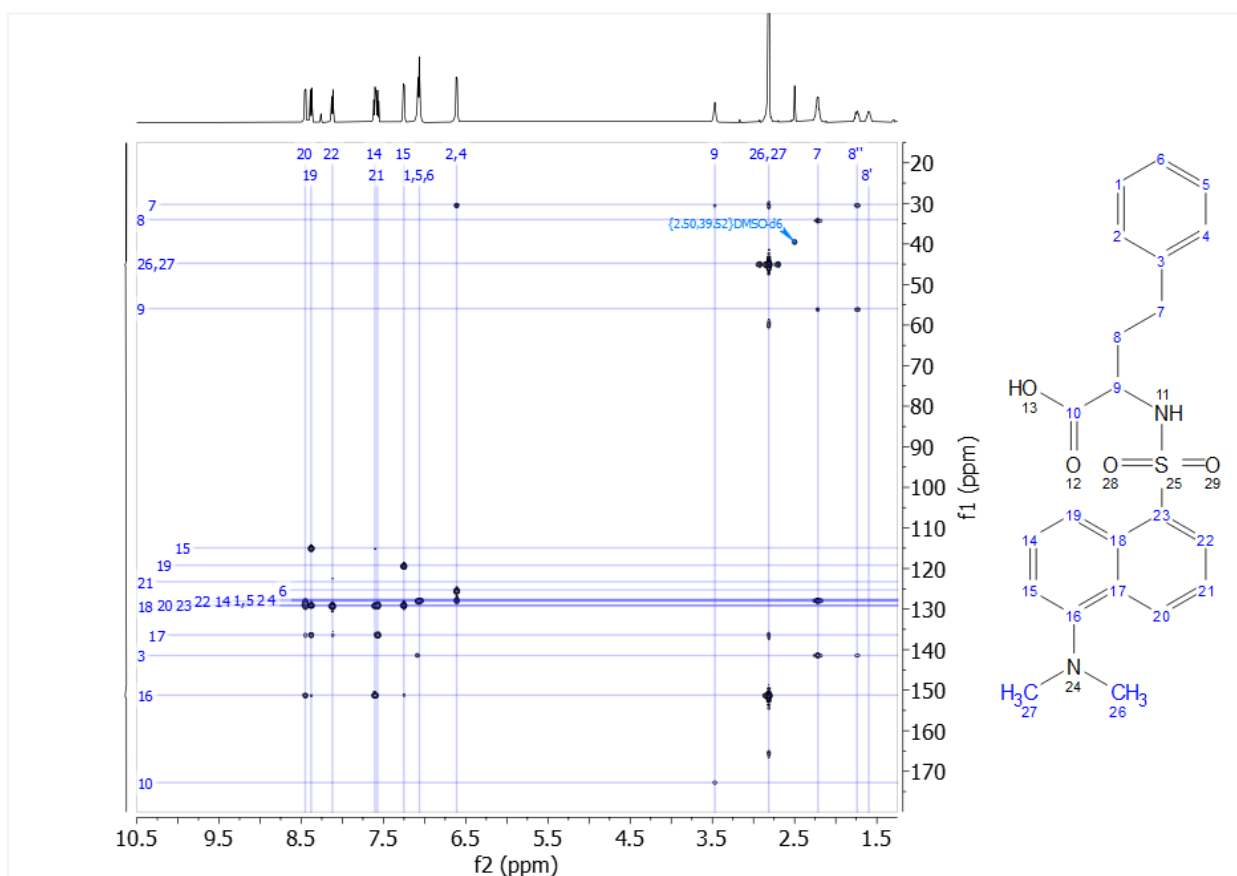
- HSQC



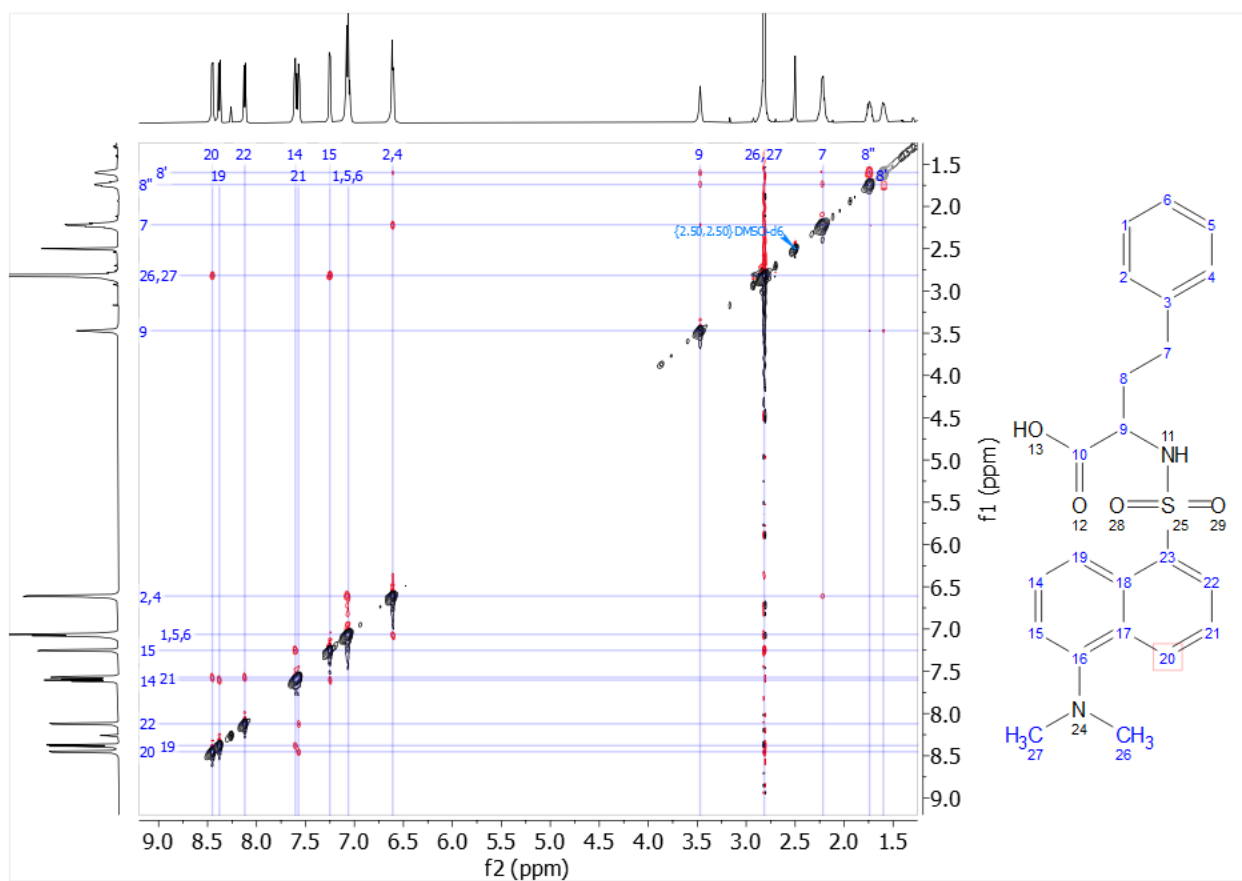
- COSY



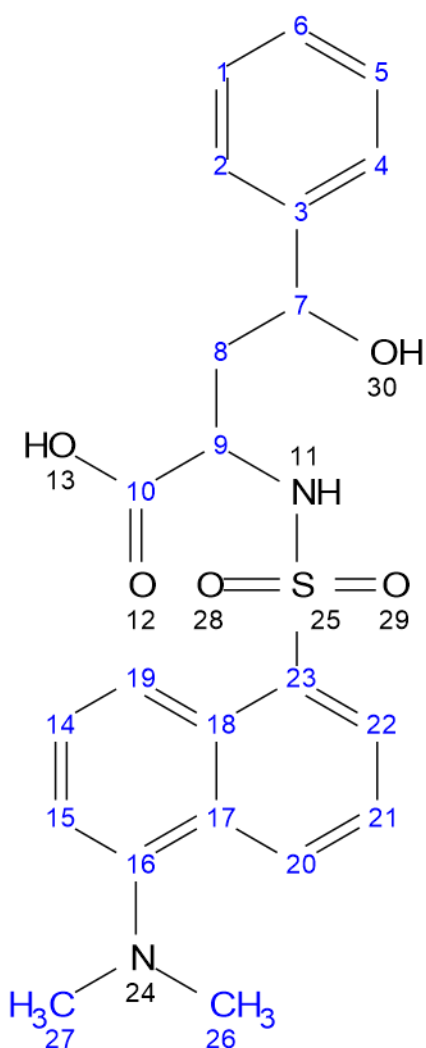
- HMBC



- ROESY



Compound 2:



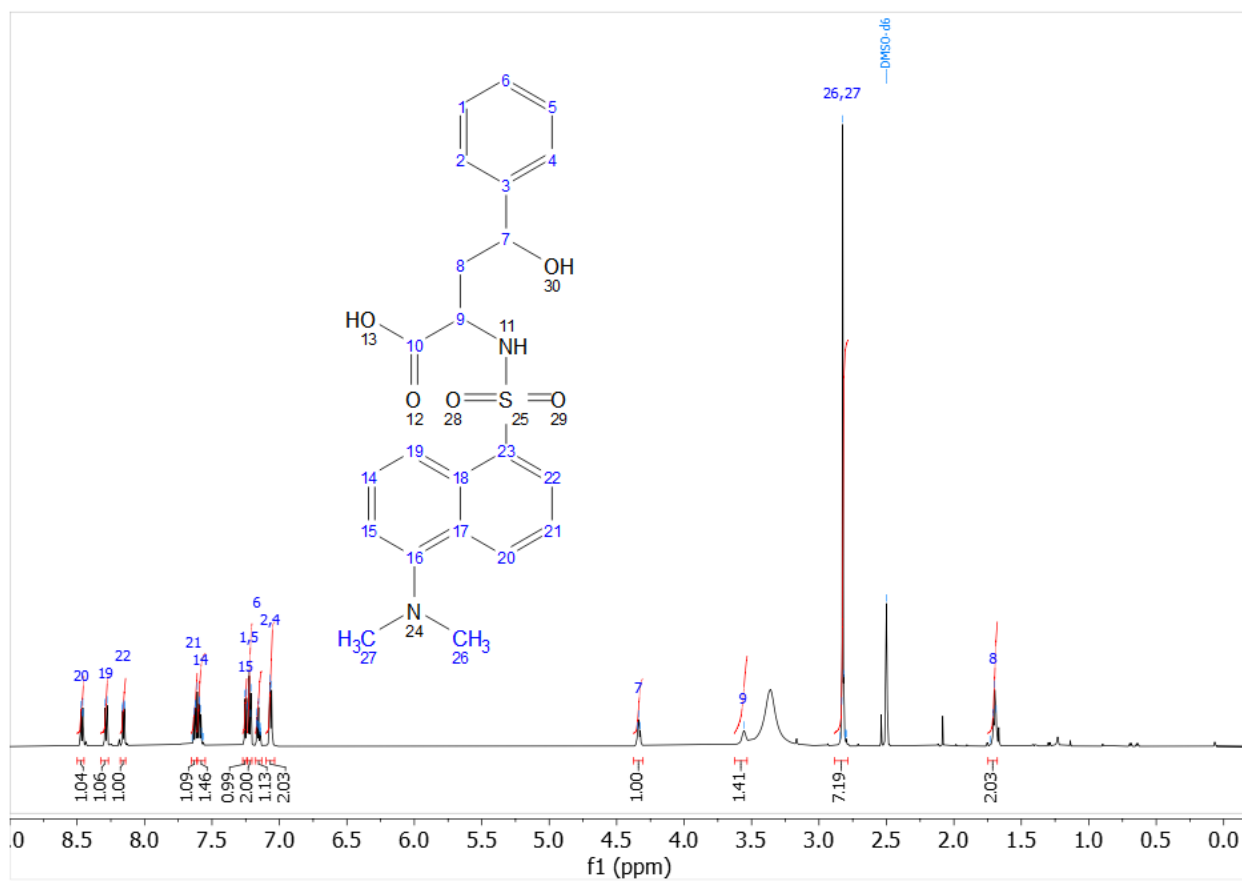
Thanks to the fact that the substrate used in the biotransformations was enantiopure L-homophenylalanine, deduction of the stereochemistry of compound 2 (Figure 1, main text) was possible from knowledge of the relative stereochemistry of compound 6 (*vide infra*).

^1H NMR (600 MHz, DMSO) δ 8.46 (d, $J = 8.4$ Hz, 1H, 20), 8.29 (d, $J = 8.6$ Hz, 1H, 19), 8.16 (dd, $J = 7.4, 1.3$ Hz, 1H, 22), 7.63 (t, $J = 7.8$ Hz, 1H, 21), 7.59 (t, $J = 8.4, 7.8$ Hz, 1H, 14), 7.25 (d, $J = 7.5$ Hz, 1H, 15), 7.22 (t, $J = 7.5$ Hz, 2H, 1, 5), 7.15 (t, $J = 7.4$ Hz, 1H, 6), 7.07 (d, $J = 7.4$ Hz, 2H, 2, 4), 4.34 (t, $J = 6.4$ Hz, 1H, 7), 3.56 (s, 1H, 9), 2.82 (s, 7H, 26, 27), 1.70 (t, $J = 6.5$ Hz, 2H, 8).

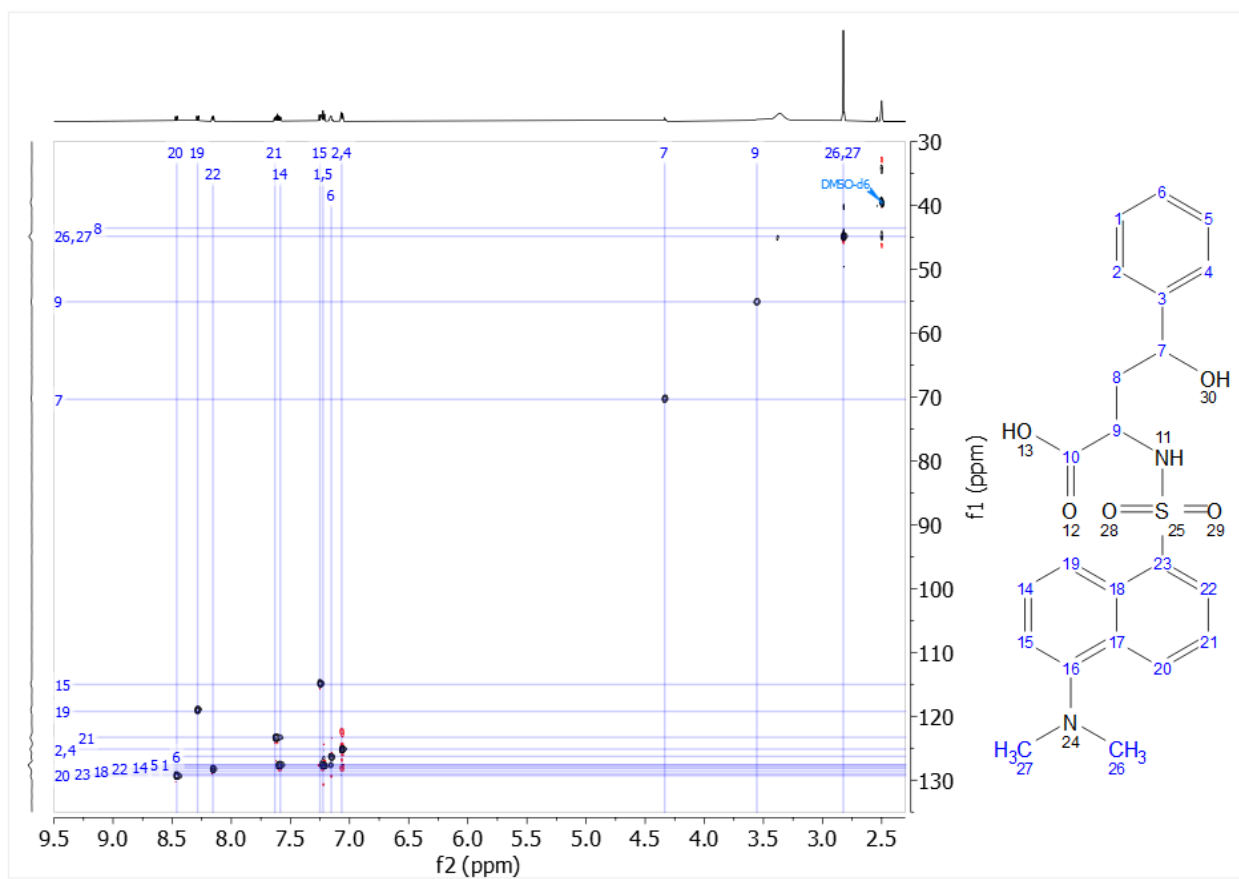
2D NMR assignment:

Assignment	H- δ (ppm)	C- δ (ppm)	splitting	<i>J</i> (Hz)	¹ H- ¹ H COSY	¹ H- ¹ H ROESY	¹ H- ¹³ C HMBC	# H
CH-1/5	7.22	127.6	t	7.5	H-6, H-2/4		C-3, C-1/5	2
CH-2/4	7.07	125.2	d	7.5	H-1/5	H-8	C-7, C-6, C-2/4	2
C-3		146.1						
CH-6	7.15	126.3	t	7.4	H-1/5		C-2/4	1
CH-7	4.34	70.4	t	6.4			C-8, C-9, C-3, C-2/4	1
CH ₂ -8	1.70	43.6	t	6.5	H-9	H-9, H-2/4	C-10, C-9, C7, C3	2
CH-9	3.56	55.1	s		H-8	H-8, H-22, H-2/4		1
C-10		172.8						
CH-14	7.59	128.0	t	8.4, 7.8	H-19, H-15		C-16, C-18	1
CH-15	7.25	115.0	d	7.5	H-14	H-26/27	C-19, C-20, C-16	1
C-16		151.3						
C-17		135.8						
C-18		128.6						
CH-19	8.29	119.2	d	8.6	H-14		C-15, C-23, C-17	1
CH-20	8.46	129.3	d	8.4	H-21	H-26/27	C-18, C-16, C-22	1
CH-21	7.63	123.3	t	7.8	H-20, H-22		C-23, C-17	1
CH-22	8.16	128.2	dd	7.4, 1.3	H-21	H-9	C-20, C-18	1
C-23		129.1						
CH ₃ -26/27	2.82	44.8	s			H-20, H-15	C-16	6

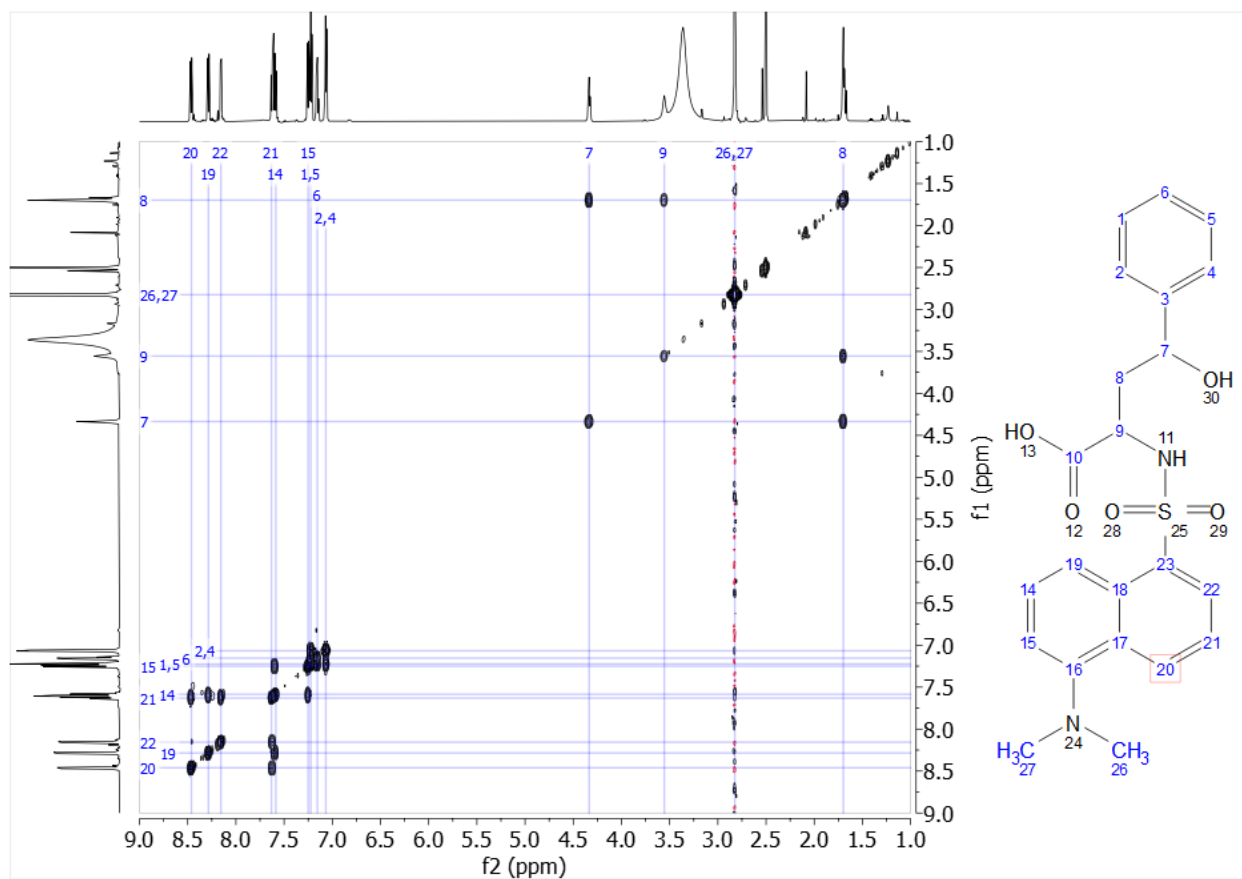
- 1D NMR



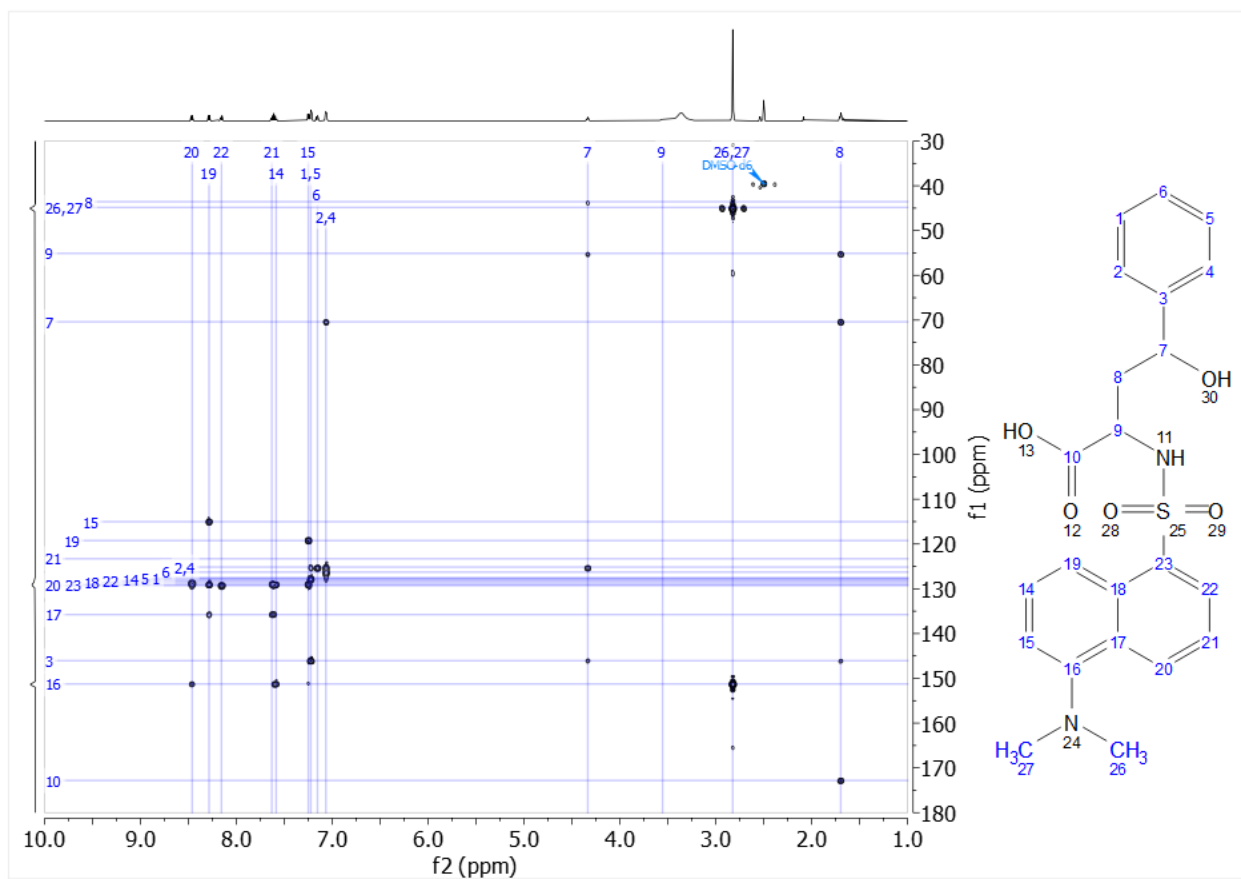
- HSQC



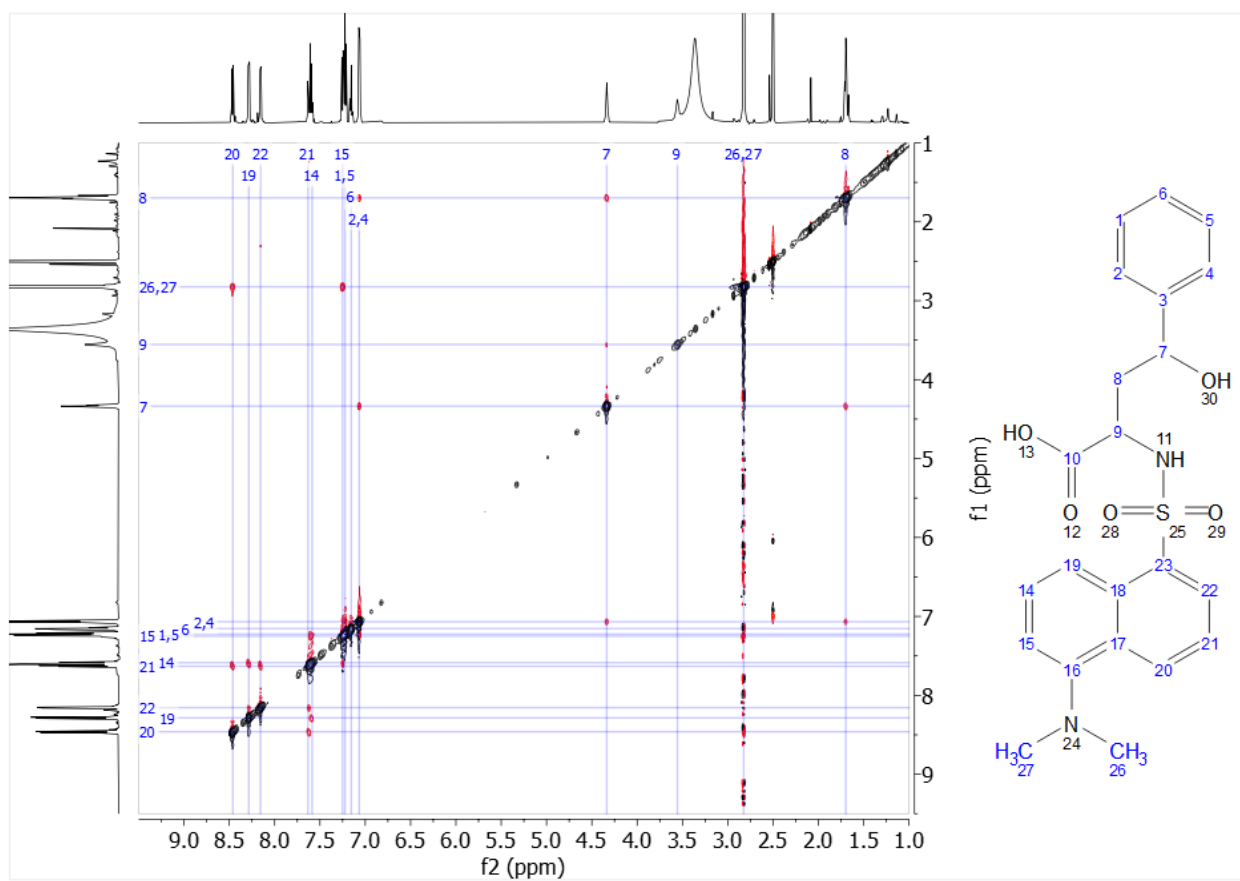
- COSY



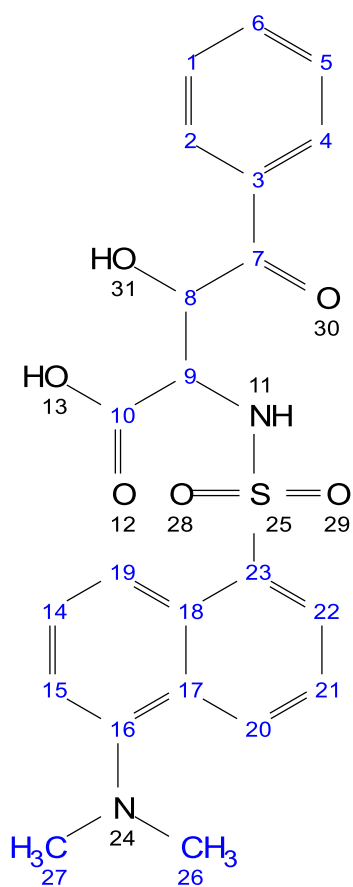
- HMBC



- ROESY



Compound 4:

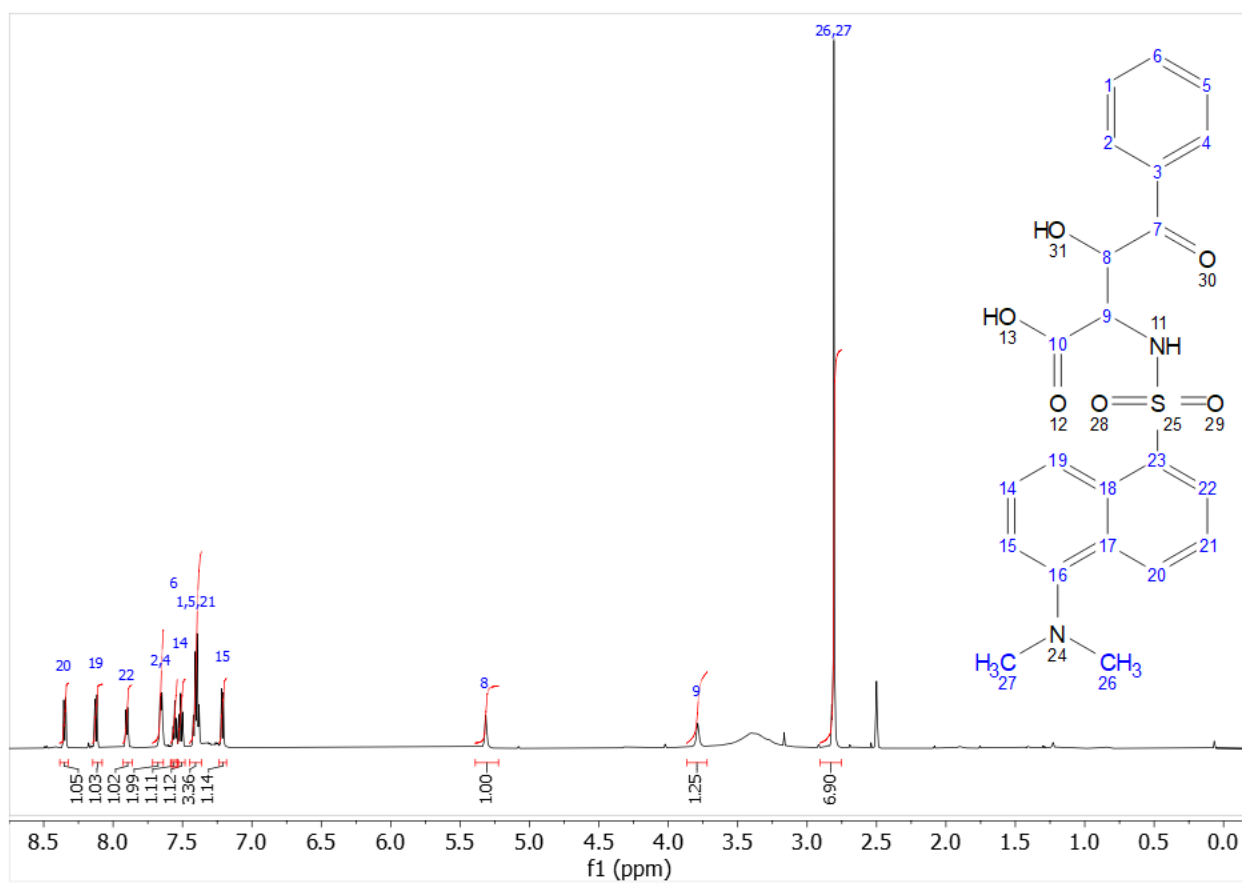


^1H NMR (600 MHz, DMSO) δ 8.35 (d, $J = 8.5$ Hz, 1H, 20), 8.12 (d, $J = 8.6$ Hz, 1H, 19), 7.90 (d, $J = 7.3$ Hz, 1H, 22), 7.65 (d, $J = 7.7$ Hz, 2H, 2, 4), 7.56 (t, $J = 7.4$ Hz, 1H, 6), 7.51 (t, $J = 8.1$ Hz, 1H, 14), 7.45 – 7.36 (m, 3H, 1, 5, 21), 7.21 (d, $J = 7.5$ Hz, 1H, 15), 5.39 – 5.22 (m, 1H, 8), 3.79 (s, 1H, 9), 2.81 (s, 7H, 26, 27).

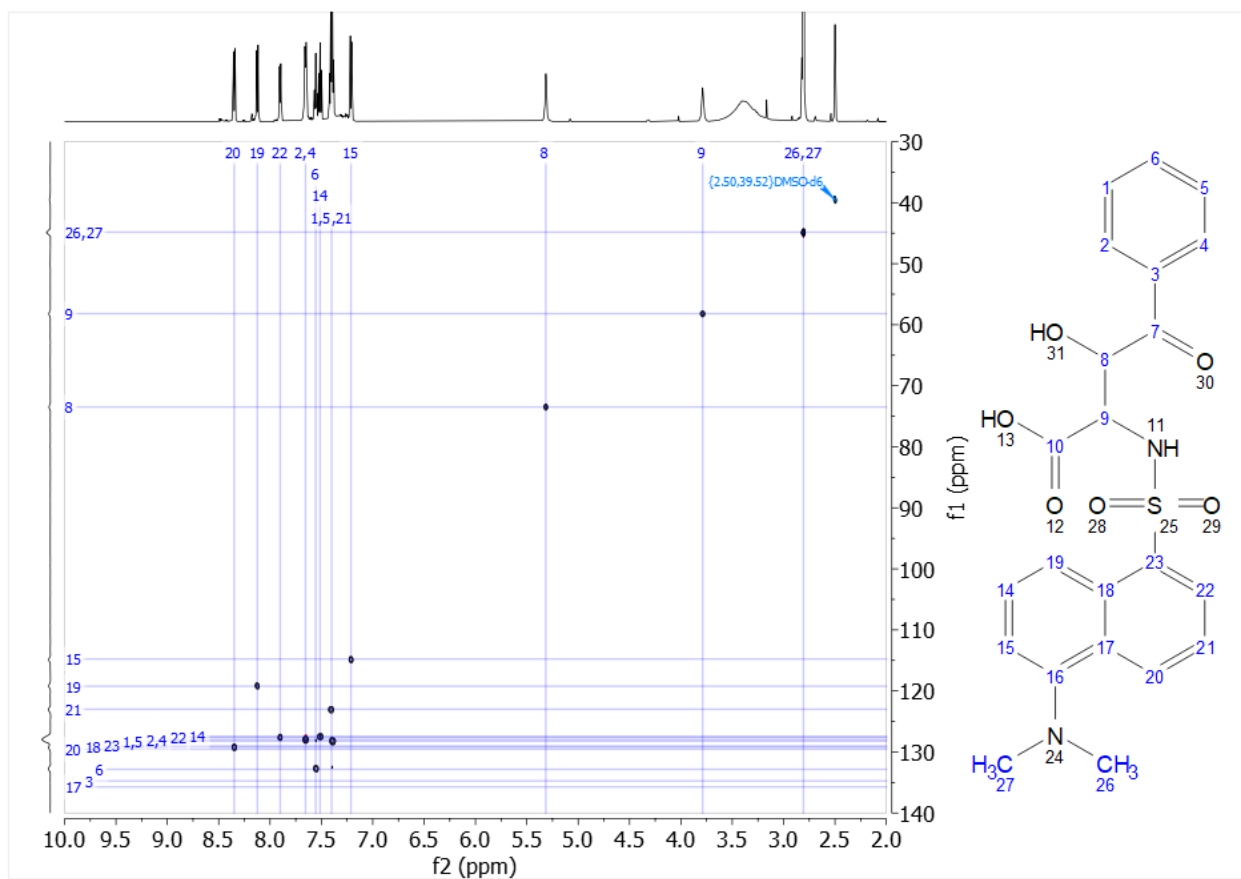
2D NMR assignment:

Assignment	H- δ (ppm)	C- δ (ppm)	splitting	<i>J</i> (Hz)	¹ H- ¹ H COSY	¹ H- ¹ H ROESY	¹ H- ¹³ C HMBC	# H
CH-1/5	7.40	128.2	m		H-6, H-2/4		C-3, C-1/5	2
CH-2/4	7.65	128.0	d	7.7	H-1/5	H-8	C-7, C-6, C-2/4	2
C-3		134.7						
CH-6	7.56	132.8	t	7.4	H-1/5		C-2/4	1
C-7		198.1						
CH-8	5.32	73.5	m		H-9	H-9, H-2/4	C-10, C-9	1
CH-9	3.79	58.2	s		H-8	H-8, H-22, H-2/4		1
C-10		170.0						
CH-14	7.51	127.5	t	8.1	H-19, H-15		C-16, C-18	1
CH-15	7.21	114.8	d	7.5	H-14	H-26/27	C-19, C-20, C-16	1
C-16		151.1						
C-17		135.7						
C-18		129.2						
CH-19	8.12	119.2	d	8.6	H-14		C-15, C-23, C-17	1
CH-20	8.35	129.6	d	8.5	H-21	H-26/27	C-18, C-16, C-22	1
CH-21	7.40	123.0	m		H-20, H-22		C-23, C-17	1
CH-22	7.90	127.6	d	7.3	H-21	H-9	C-20, C-18	1
C-23		129.0						
CH ₃ -26/27	2.81	45.2	s			H-20, H-15	C-16	6

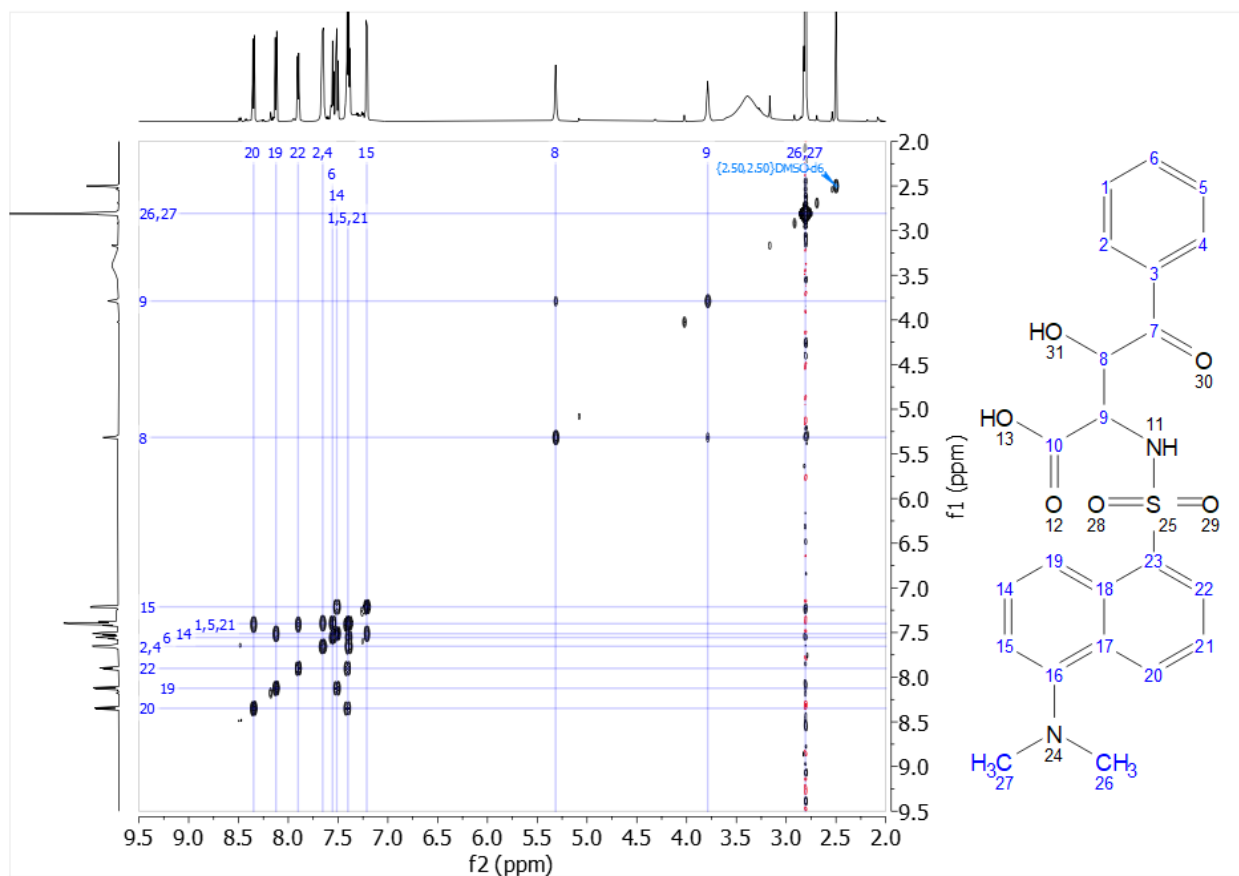
- 1D NMR



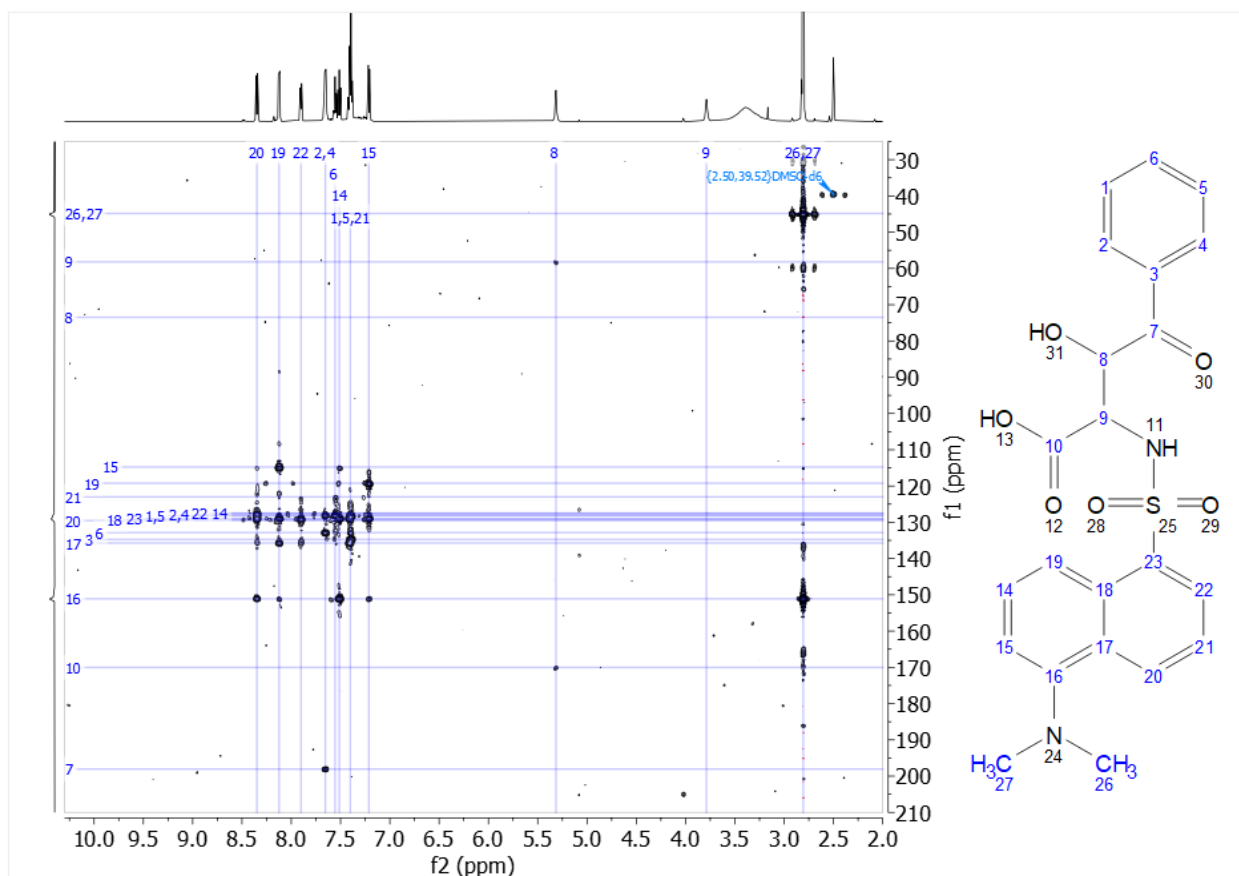
- HSQC



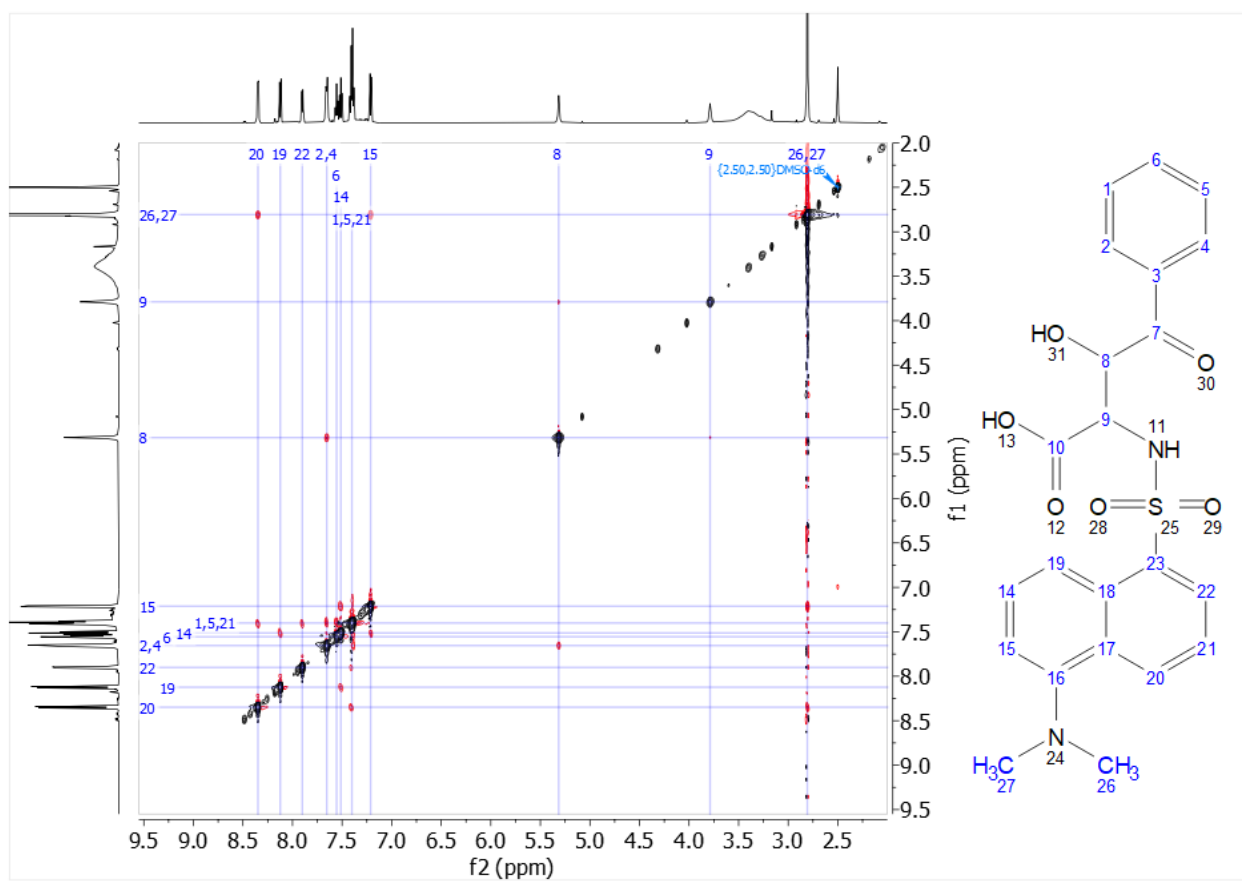
- COSY



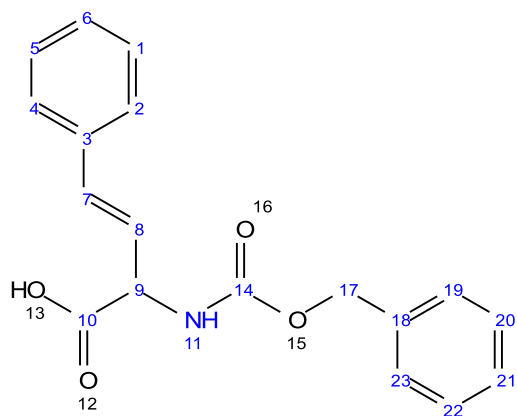
- HMBC



- ROESY



Compound 5

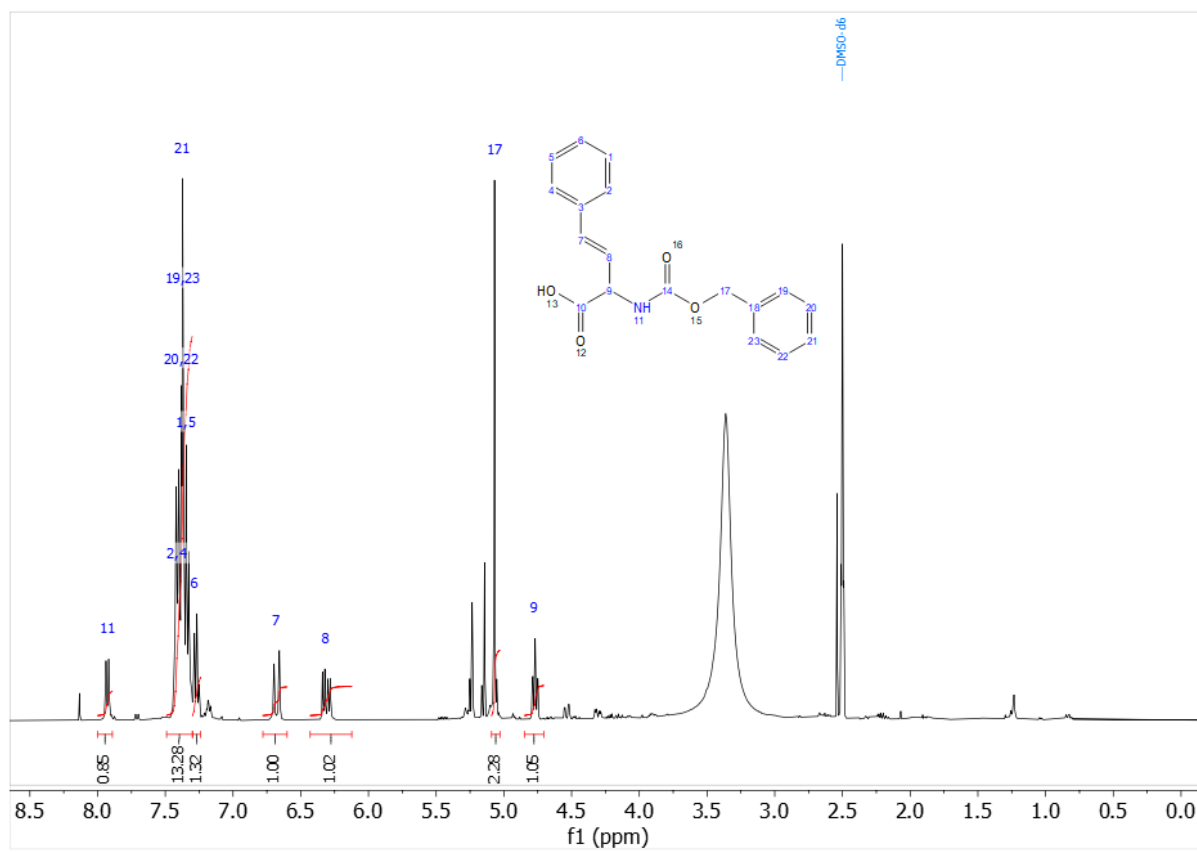


$^1\text{H NMR}$ (400 MHz, DMSO) δ 7.93 (d, $J = 7.9$ Hz, 1H, 11), 7.49 – 7.30 (m, 13H, 1, 5, 19, 20, 22, 23), 7.30 – 7.24 (m, 1H, 6), 6.68 (dd, $J = 16.1, 1.4$ Hz, 1H, 7), 6.31 (dd, $J = 16.0, 6.9$ Hz, 1H, 8), 5.07 (s, 2H), 4.77 (ddd, $J = 8.2, 6.8, 1.5$ Hz, 1H, 9).

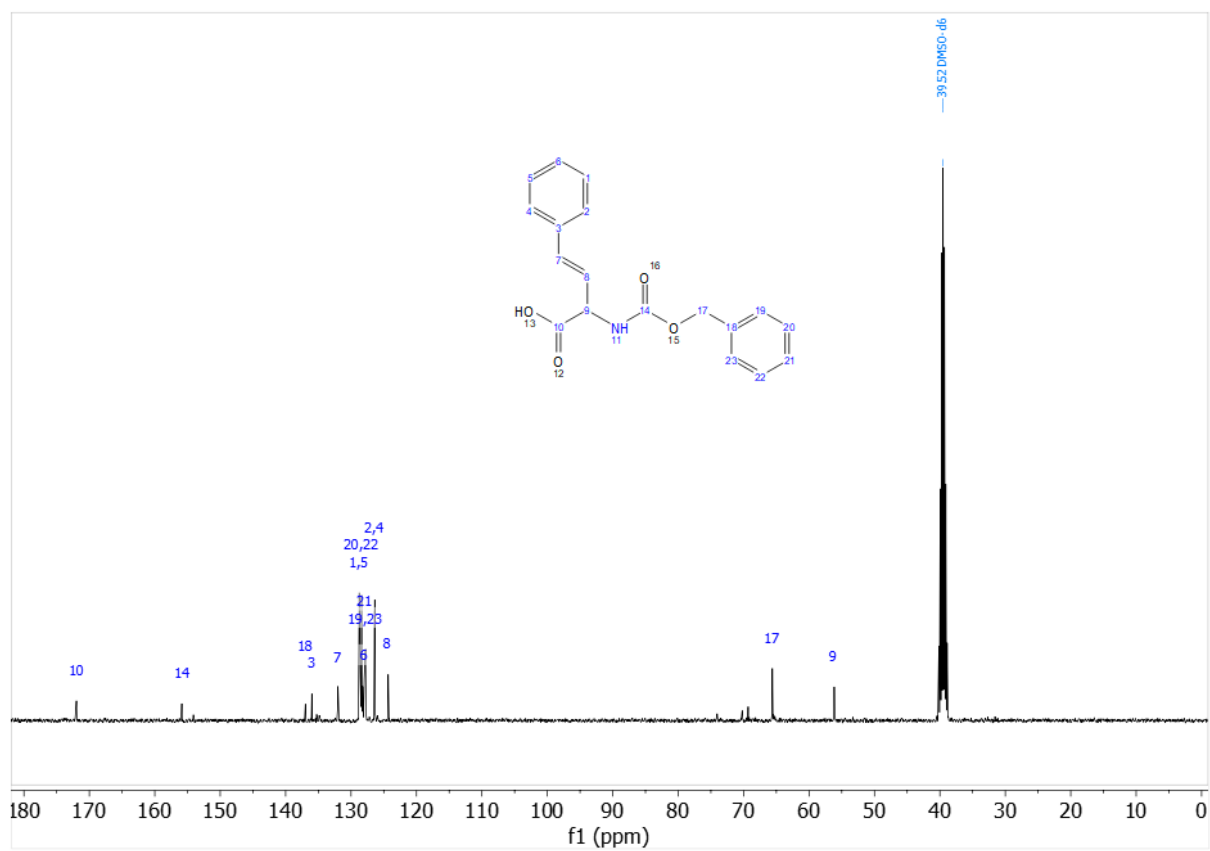
2D NMR assignment:

Assignment	H- δ (ppm)	C- δ (ppm)	splitting	<i>J</i> (Hz)	¹ H- ¹ H COSY	¹ H- ¹ H ROESY	¹ H- ¹³ C HMBC	# H
CH-1/5	7.34	128.8	m		H-6, H-2/4		C-3, C-1/5	2
CH-2/4	7.41	126.4	m		H-1/5	H-7, H-8'	C-7, C-6, C-2/4	2
C-3		136.0						
CH-6	7.27	128.0	m		H-1/5		C-2/4	1
CH-7	6.68	132.0	dd	16.1, 1.4	H-8	H-9, H-2/4	C-2/4, C-9, C-3	1
CH-8	6.31	124.3	dd	16.0 6.9	H-7, H-9	H-2/4, H-9	C-10, C-9, C-3	1
CH-9	4.77	56.2	ddd	8.2, 6.8, 1.5	H-8, H-11	H-7, H-8, H-11	C,8, C-7, C-14, C-10	1
C-10		172.0						
NH-11	7.93		d	7.9	H-9		C-8, C-9, C-14	1
C-14		155.9						
CH ₂ -17	5.07	65.6	s			H-19/23	C19/23, C-18, C-14	2
C-18		137.0						
CH-19/23	7.37	127.8	m		H-20/22	H-17		2
CH-20/22	7.38	128.4	m		H-19/23, H21			2
CH-21	7.38	127.9	m		H-20/22			1

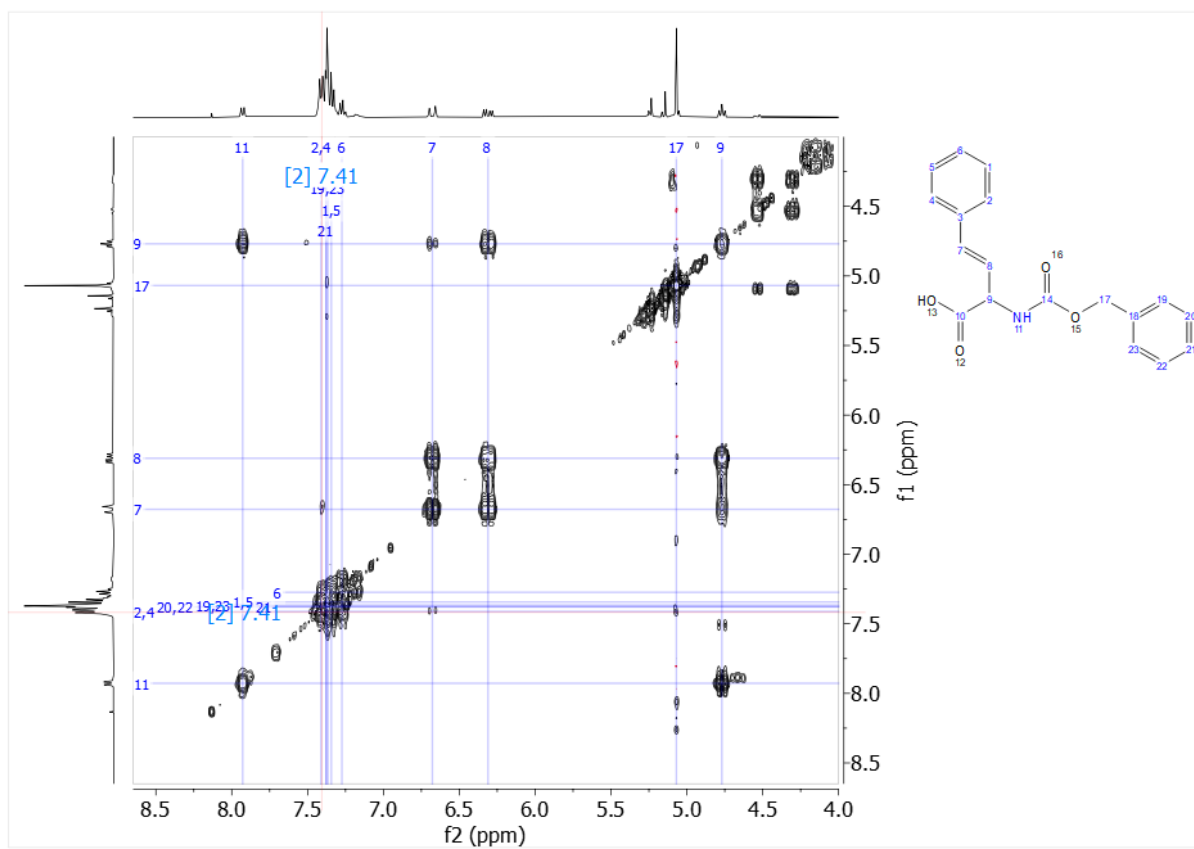
• 1D NMR



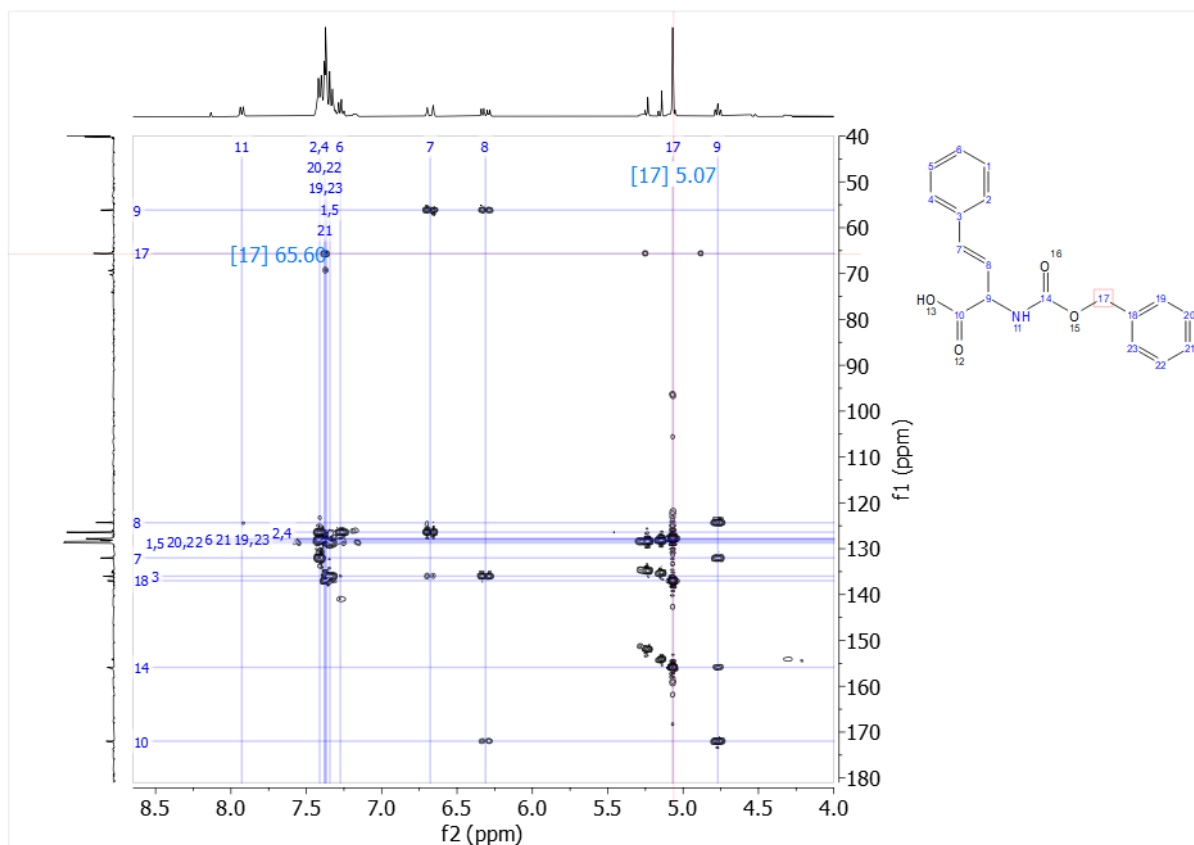
- 1D ^{13}C NMR



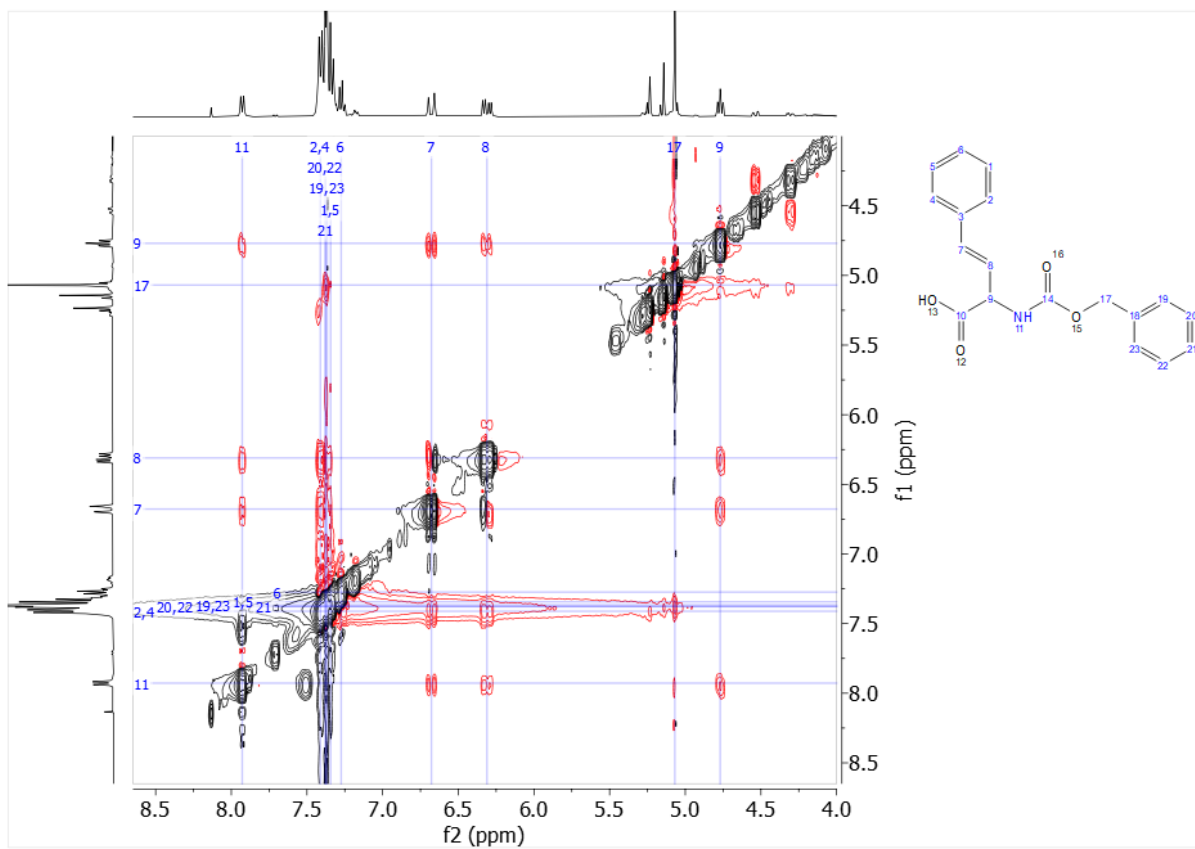
- COSY



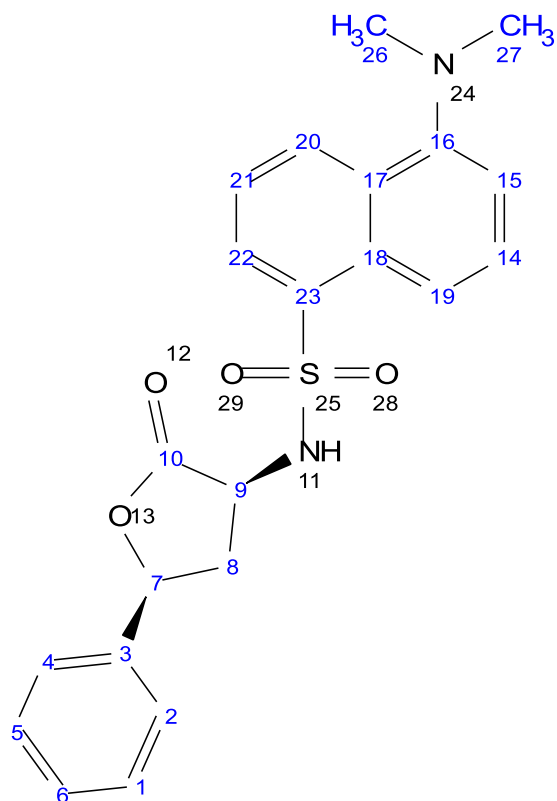
- HMBC



- ROESY



Compound 6



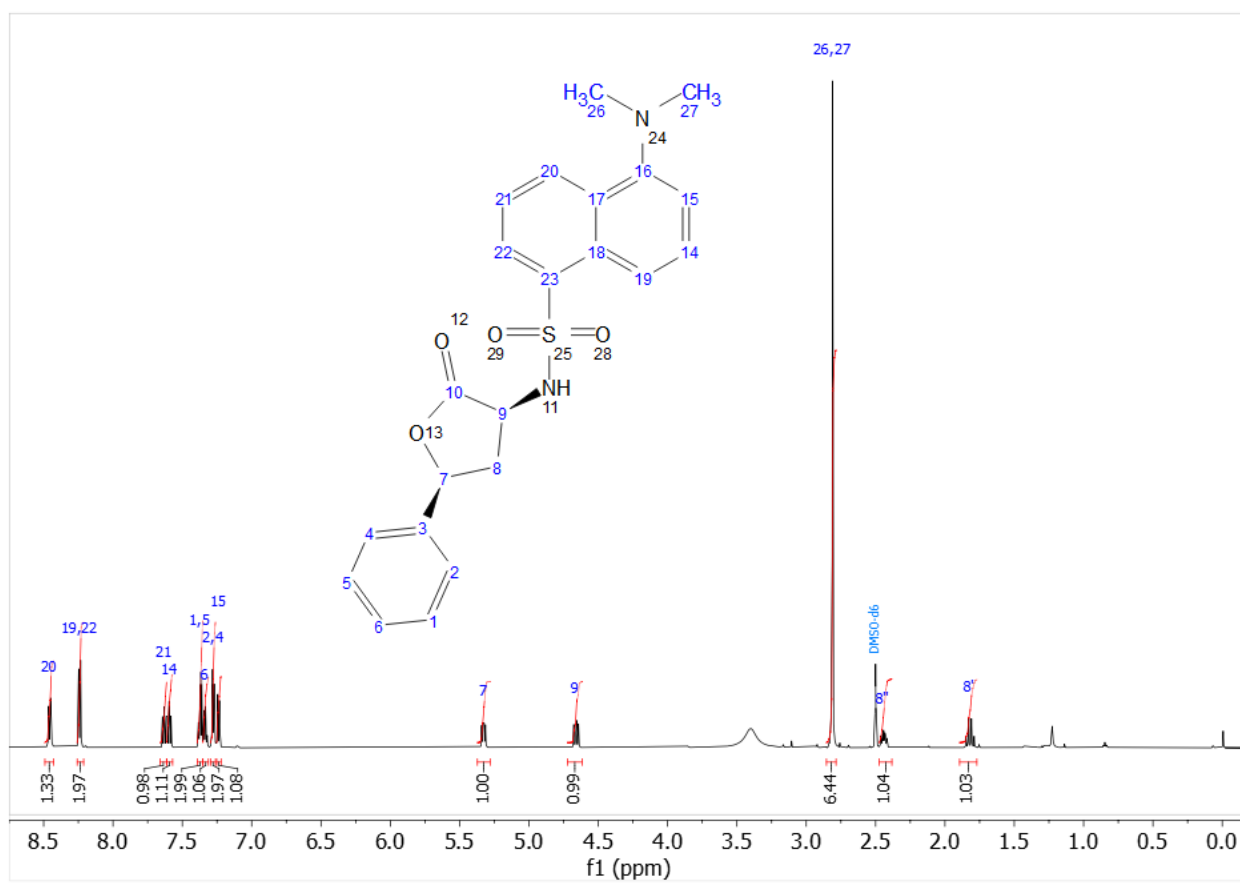
For compound 6 the relative stereochemistry was determined by NMR, the absolute stereochemistry is given randomly.

^1H NMR (600 MHz, DMSO) δ 8.46 (dt, $J = 8.5, 1.1$ Hz, 1H, 20), 8.24 (d, $J = 8.3$ Hz, 2H, 19, 22), 7.63 (dd, $J = 8.5, 7.3$ Hz, 1H, 21), 7.60 (dd, $J = 8.7, 7.5$ Hz, 1H, 14), 7.37 (dd, $J = 8.1, 6.3$ Hz, 2H, 1, 5), 7.33 (t, $J = 7.3$ Hz, 1H, 6), 7.28 (d, $J = 7.0$ Hz, 2H, 2, 4), 7.24 (d, $J = 7.4$ Hz, 1H, 15), 5.33 (dd, $J = 11.0, 5.5$ Hz, 1H, 7), 4.66 (dd, $J = 11.8, 8.4$ Hz, 1H, 9), 2.81 (s, 6H, 26, 27), 2.44 (ddd, $J = 12.2, 8.4, 5.5$ Hz, 1H, 8"), 1.82 (q, $J = 11.7$ Hz, 1H, 8').

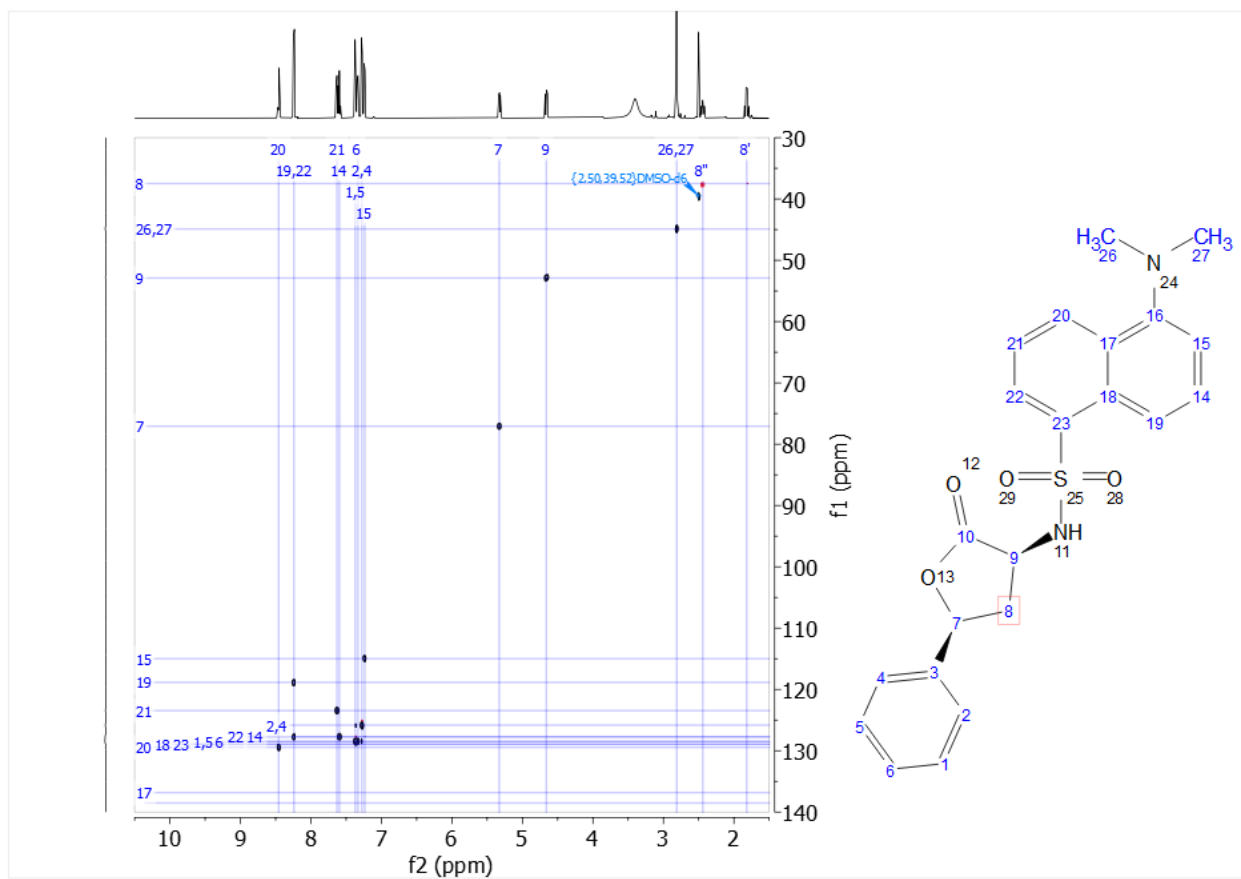
2D NMR assignment:

Assignment	H- δ (ppm)	C- δ (ppm)	splitting	<i>J</i> (Hz)	¹ H- ¹ H COSY	¹ H- ¹ H ROESY	¹ H- ¹³ C HMBC	# H
CH-1/5	7.37	128.5	dd	8.1, 6.3	H-6, H-2/4		C-3, C-1/5	2
CH-2/4	7.28	125.8	d	7.0	H-1/5	H-7, H-8'	C-7, C-6, C-2/4	2
C-3		138.5						
CH-6	7.33	128.5	t	7.3	H-1/5		C-2/4	1
CH-7	5.33	77.1	dd	11.0, 5.5	H-8	H-9, H-2/4, H-8''	C-2/4, C-3	1
CH ₂ -8'	1.82	37.5	q	11.7	H-7, H-9	H-2/4	C-7, C-9, C-3	1
CH ₂ -8''	2.44	37.5	ddd	12.2, 8.4, 5.5		H-7, H-9, H-19/22	C-9, C-10	1
CH-9	4.66	52.9	dd	11.8, 8.4	H-8	H-7, H-8'', H-19/22	C,8, C-10	1
C-10		173.8						
CH-14	7.60	127.7	dd	8.6, 7.5	H-19, H-15		C-16, C-18	1
CH-15	7.24	115.0	d	7.4	H-14	H-26/27	C-19, C-20	1
C-16		151.3						
C-17		136.8						
C-18		129.0						
CH-19	8.24	118.8	d	8.3	H-14	H-8''	C-15, C-23, C-17	1
CH-20	8.46	129.5	dt	8.5, 1.1	H-21	H-26/27	C-18, C-16, C-22	1
CH-21	7.63	123.5	dd	8.5, 7.3	H-20, H-22		C-23, C-17	1
CH-22	8.24	127.8	d	8.3	H-21	H-9, H-8''	C-20, C-18	1
C-23		128.9						
CH ₃ -26/27	2.81	44.9	s			H-20, H-15	C-16	6

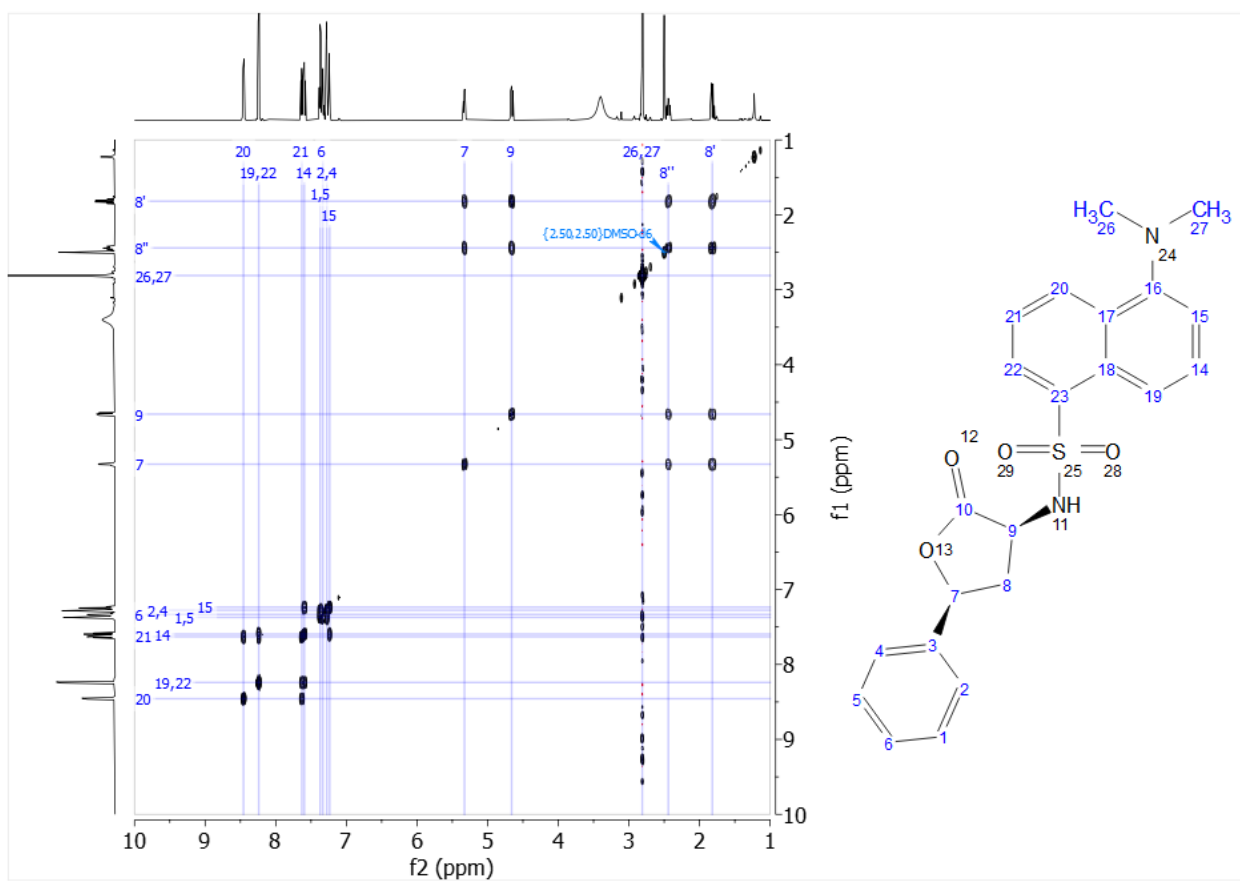
- 1D NMR



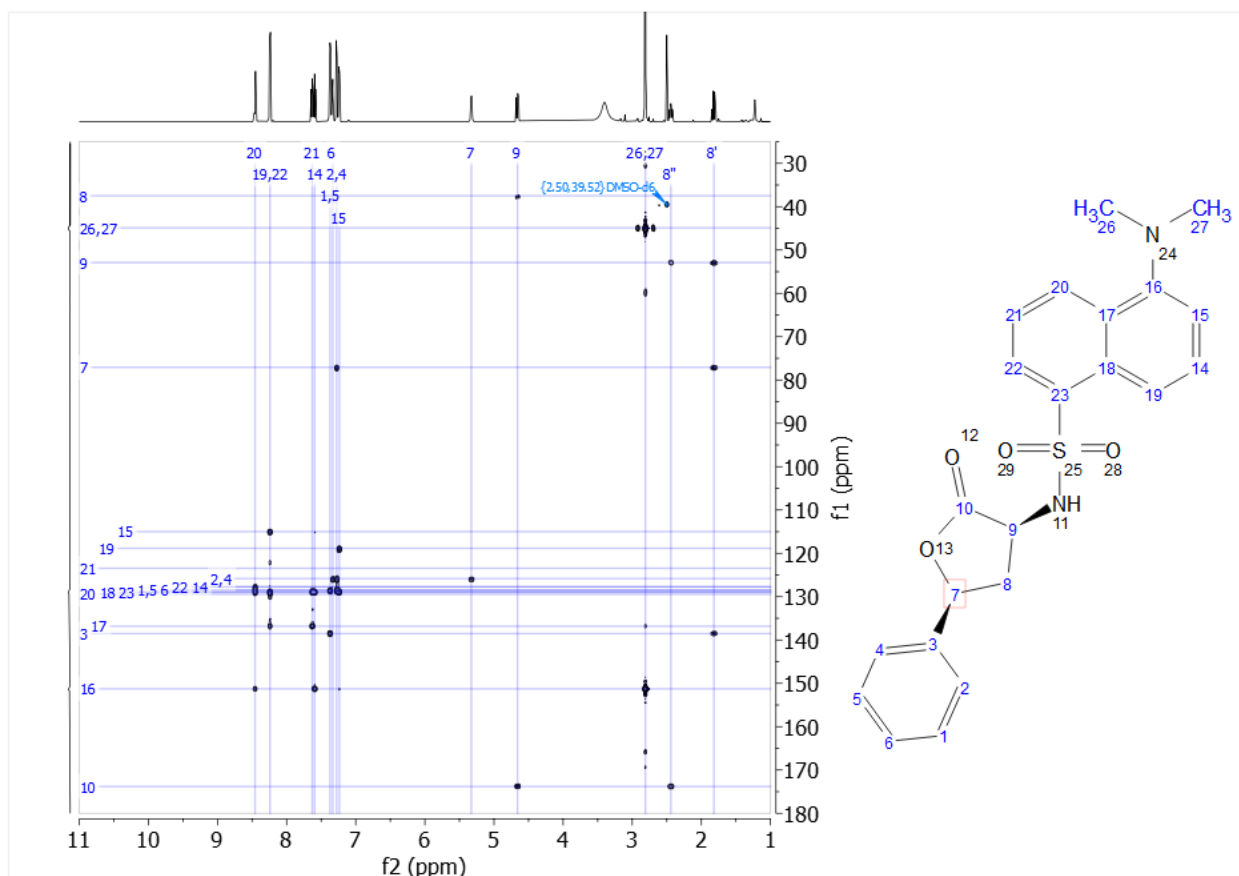
- HSQC



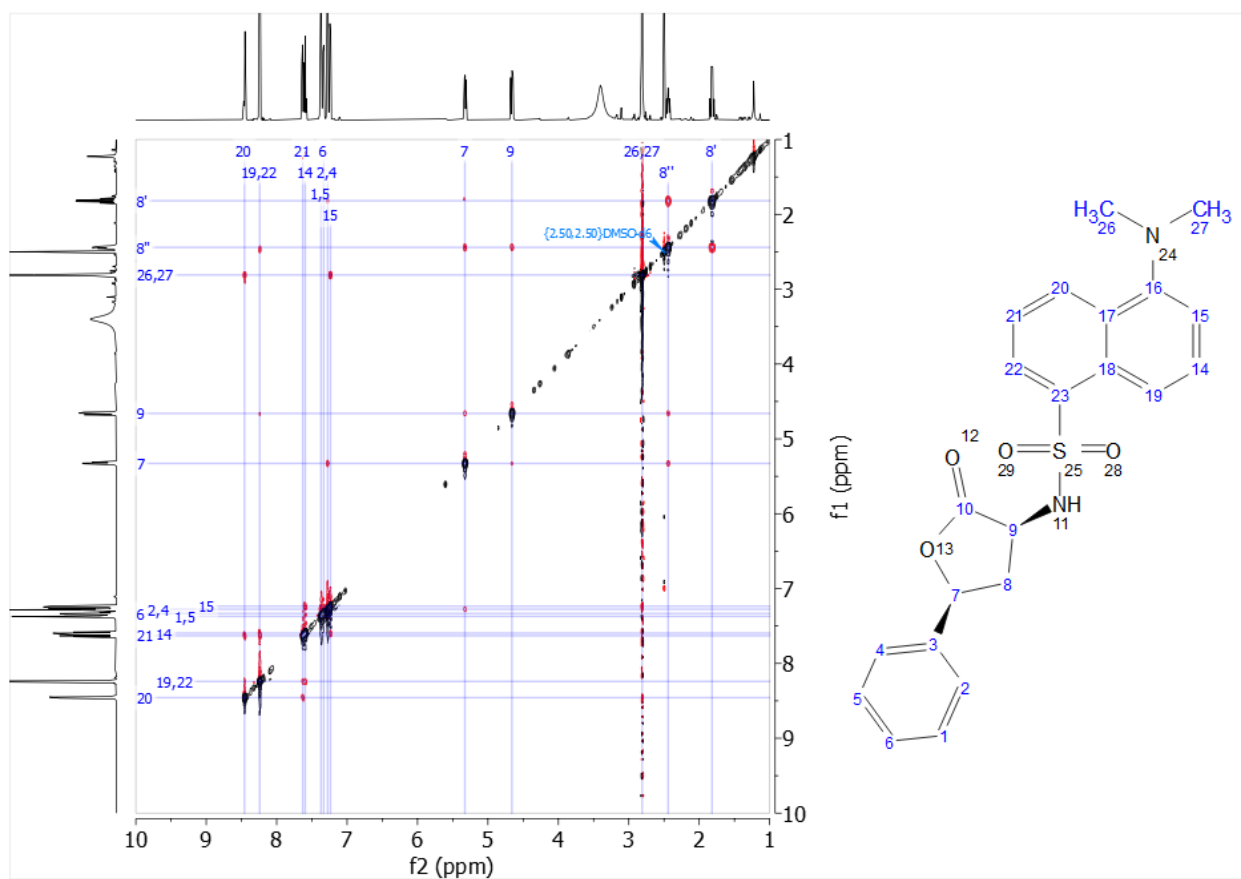
- COSY



- HMBC



- ROESY



II. Supporting Figures and Tables

Table S1 Panel of Fe(II)/ α -KG-dependent dioxygenases used to identify a starting scaffold with hydroxylase activity towards L-homophenylalanine.

Natural diversity					
	Name	Uniprot nr.		Name	Uniprot nr.
1	YcfDRM ⁶	D0MK34	19	KDO1 ⁷	C7QJ42
2	VioC ⁸	Q6WZB0	20	KDO3 ⁷	A5FF23
3	Ido ⁹	E2GIN1	21	St16DOX ¹⁰	M1AUM2
4	LdoA ¹¹	B2J686	22	TauD ¹²	P37610
5	SruPH ¹³	E5XT07	23	YLL057c ¹⁴	Q12358
6	GetF ¹⁵	E0WFM0	24	AsqJ ¹⁶	Q5AR53
7	XdPH ^{13,17}	A0A068QPC1	25	H7-1 ¹⁸	P18548*
8	SmPH4 ²	Q92LF6	26	GloF ¹⁹	S3D784
9	P3H ²⁰	O09345	27	DAO ²¹	Q7XKU5
10	DACSP ²²	O06499	28	PAHX ²³	O14832
11	AsnO ²⁴	Q9Z4Z5	29	CtrZ ²⁵	P21688
12	FzmG ²⁶	U5YQY6	30	LpxO ²⁷	A0A0H3NVJ7
13	SadA ²⁸	Q0B2N4	31	BcmC ²⁹	A0A2G1XAR5
14	vCPH ³⁰	Q84406	32	PtlH ³¹	Q82IZ1
15	TMLH ³²	Q9NVH6	33	AvLDO ³³	Q3MCP9
16	XanA ³⁴	Q4QZZ9	34	AusE ³⁵	Q5AR34
17	FrbJ ³⁶	Q0ZQ39	35	RdpA ³⁷	Q8KSC8
18	GriE ³⁸	A0A0E3URV8	36	Phex-KGD ³⁹	A0A0N9HQ36

*instead of wild type H7-1 mutant was used¹⁸.

Table S2 Primer sequences

Entry	Name	Sequence (5'-3')
1	pET28(+)	GTG AGC GGA TAA CAA TTC CCC TCT AG
2	pET28(-)	GCT TTG TTA GCA GCC GGA TCT CAG
3	pET28 outer(-)	CAC CCG CCG CGC TTA ATG C
4	Y32NNK(+)	AGT TGC AGA ATT TAG CAG CGC ANN KAG TGA TTT TGC ATG TGG TAA ATG G
5	Y32(-)	TGC GCT GCT AAA TTC TGC AAC TTC CAG G CAG AAT TTA GCA GCG CAT ATA GCG ATN NKG CCT GTG GTA AAT GGG AAG C
6	F35NNK(+)	
7	F35(-)	ATC GCT ATA TGC GCT GCT AAA TTC TGC AAC TTC C
8	W40NNK(+)	ATA TAG CGA TTT TGC ATG TGG TAA ANN KGA GGC ATG TGT TCT GCG
9	W40(-)	TTT ACC ACA TGC AAA ATC GCT ATA TGC GCT GC
10	V57NNK(+)	TAT GCA AGA AGA AGA TAT TNN KGT GAG CCA TAA TGC ACC G
11	V57(-)	AAT ATC TTC TTC TTG CAT ACC GGT ACG
12	R93NNK(+)	GAT TGT AGC GCA GTT CGT TAT ACC NNK ATC GTT CGT GTT AGC
13	R93(-)	GGT ATA ACG AAC TGC GCT ACA ATC AAA ATG G
14	V95NNK(+)	AGC GCA GTT CGT TAT ACC CGT ATT NNK CGC GTT AGC GAA AAT GC
15	V95(-)	AAT ACG GGT ATA ACG AAC TGC GCT ACA ATC AAA ATG G
16	I103NNK(+)	GTT CGT GTT AGC GAA AAT GCC TGT NNK ATC CCG CAT AGC
17	I103(-)	ACA GGC ATT TTC GCT AAC ACG AAC AAT ACG
18	Y109NNK(+)	CTG TAT TAT TCC GCA TAG CGA TNN KCT TGA ACT GGA TGA G
19	Y109(-)	ATC GCT ATG CGG AAT AAT ACA GGC ATT TTC G
20	R118NNK(+)	GGA ACT GGA TGA GAC ATT TAC CNN KCT TCA TCT GGT TCT G
21	R118(-)	GGT AAA TGT CTC ATC CAG TTC CAG ATA ATC G
22	M168NNK(+)	AGC AAA ACA CCG CGT CTG CAC CTG NNK ATC GAT TTT GAA GC
23	M168(-)	CAG GTG CAG ACG CGG TGT TTT GCT AAA ACA TGC
24	R272NNK(+)	TGG AAC GTT TTT TTC TGG GTC ATN NKG AGC GTG GTG AAG TTA TGA C
25	R272(-)	ATG ACC CAG AAA AAA ACG TTC CAG GCT TGC G
26	W40DBK(+)	ATA TAG CGA TTT TGC ATG TGG TAA ADB KGA GGC ATG TGT TCT GCG
27	I103DYA(+)	GTT CGT GTT AGC GAA AAT GCC TGT DYA ATC CCG CAT AGC

Table S3 Distribution of oxidative modified L-homophenylalanine products from wild type and evolved variants based on extracted ion chromatograms. Reaction conditions: 1mM L-hPhe, 120 mM ascorbic acid, 20 μ M enzyme, 60 min at 20°C. Mean values (n = 3 biological replicates) are shown. R= NH₂-CH-COOH.

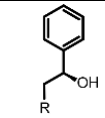
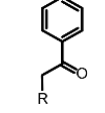
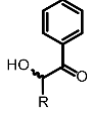
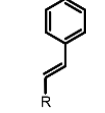
Product	WT	W40M	W40M I103L
2 	1.5%	31.3%	38.9%
3 	3.5%	20.3%	6.1%
4 	0.2%	11.9%	30.4%
5 	0.2%	2.8%	1.1%

Table S4: Sequencing results for the two hotspots, W40 and I103, which were identified via site directed saturation mutagenesis approach in the first round of evolution. Notably, smaller residues

W40NNK	fold increase over wild type	count of variants on plate
Met	4.4	1
Gly	2.3-3.5	6
Cys	2.3-2.8	4
Thr	2.3-2.5	4

W40NNK	fold increase over wild type	count of variants on plate
Ala	1.9-3.7	3
Ser	3.1-3.7	6

Table S5 Conversion of L-homotyrosine, L-homophenylalanine and (*S*)- α -amino-3,4-dichlorobenzenebutanoic acid to different oxidative products with wildtype Smp4H (see Figure 5a). Conversions are estimated through comparison with the negative control omitting the enzyme, in which the detected substrate peak intensity was set to 100 %. The calculations of product conversions assume similar ionization efficiency for the products compared to the respective substrates.

Substrates/Products	7	8	9
R ₁ = H, R ₂ = OH	0.5%	0.0%	0.0%
R ₁ = H, R ₂ = H	3.8%	10.7%	3.0%
R ₁ = Cl, R ₂ = Cl	32.2%	30.2%	4.0%

Table S6 Conversion of L-homotyrosine, L-homophenylalanine and (*S*)- α -amino-3,4-dichlorobenzenebutanoic acid to different oxidative products with engineered variant Smp4H W40M I103L (see Figure 5b). Conversions are estimated through comparison with the negative control omitting the enzyme, in which the detected substrate peak intensity was set to 100 %. The calculations of product conversions assume similar ionization efficiency for the products compared to the respective substrates.

Substrates/Products	7	8	9
R ₁ = H, R ₂ = OH	7.9%	0.0%	0.0%
R ₁ = H, R ₂ = H	33.5%	5.1%	28.8%
R ₁ = Cl, R ₂ = Cl	8.2%	3.1%	50.7%

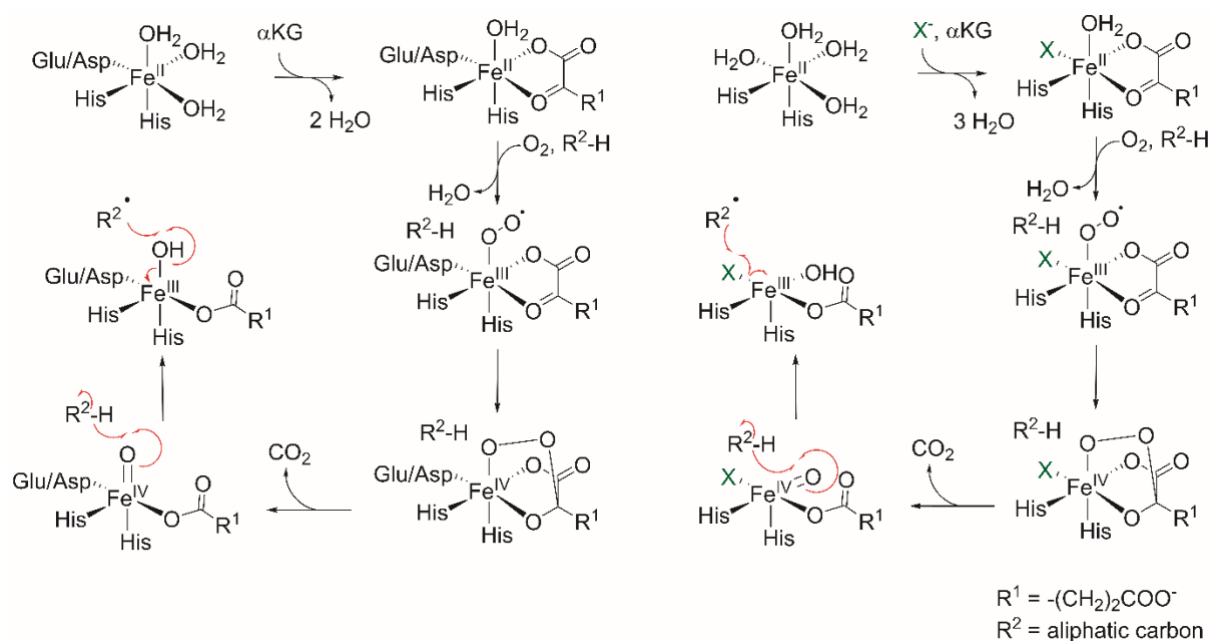


Figure S1 Proposed reaction mechanism of hydroxylation (left) and halogenation (right) by Fe(II)/2OG oxygenases. In case of halogenases, the Fe-coordinating Glutamate/Aspartate residues is replaced by a Glycine allowing for the coordination of a halide. After formation of the high-valent Fe(IV) intermediate, a hydrogen atom is abstracted from the substrate molecule. The substrate radical consequently reacts with either the hydroxy-group (hydroxylases) or the halide (halogenases) bound to the metal center to form the final product. Mechanism adapted from Voss et al.⁴⁰

```

KGOH_08_SmPH 1 -----MSTHFLGKVKFDEARLAEDLSTLEVAEFSSAYSDFACGWEACVLRNR
cis-P4H      1 MASWSHPQFEKGATTRILGVVQLDQRRLTDDLAVLAKSNFSSEYSDFACGWEFCMLRNQ

KGOH_08_SmPH 49 IGMQEEDIVVSHNAPALATPLSKSLPYLNELVETHFDCSAVRYTRIVRSENACIIPHSD
cis-P4H      61 SGKQEEQRVVVHETPALATPLGQSLPYLNELLDNHFDRDSRYARIIRSENACIIPHRD

KGOH_08_SmPH 109 YLELDETFTRHLVLDTNSGCANTEEDKIFHMGLGEIWFLDAMLPHSACFSKTPRLHLM
cis-P4H      121 YLELEGKFIRHLVLDTNEKCSNTEENNIFHMGRGEIWFLDASLPHSACFSPTPRLHLM

KGOH_08_SmPH 169 IDEEATAFPESFLRNVEQVTTRDMDPRKELTDEVIEGILGFSIIISEANYREIVSILA
cis-P4H      181 VDIEGTRSLEVAINVEQSARNATVDTRKEWTDETIESVLGFSEIISEANYREIVAILA

KGOH_08_SmPH 229 KLHFFYKADCRSMYDWLKEICKRRGPALIEKTASLERFFLGHRERGEVMTY
cis-P4H      241 KLHFFHKVHCVDMYGWLKEICKRRGPALIEKANSLERFYLIDRAAGEVMTY

```

Figure S2 Sequence alignment of KGOH8_SmP4H and the homologous proline hydroxylase (MIP4H) used to obtain a homology model.

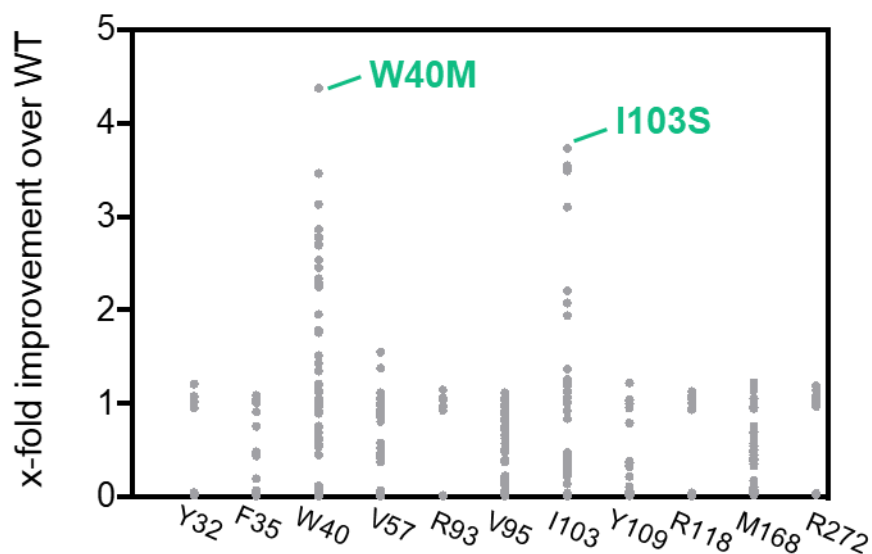


Figure S3: Results of the first evolution round of Smp4H towards improved hydroxylation of L-homophenylalanine. Two hotspots, W40 and I103, were identified via a site directed saturation mutagenesis approach leading to variants with up to 4.5-fold improved activity over the wildtype enzyme Smp4H. Additional amino acids that improved the desired enzymatic activity > 2-fold were identified as M, G, C and T at position W40 and A and S at position I103. As a consequence, degenerate codon DBK, encoding amino acids A, R, C, G, I, L, M, F, S, T, W and V, and degenerate codon DYA, encoding amino acids A, I, L, S, T, V were chosen to build a two-site combinatorial library in a second round of engineering (for screening results, see Figure S4). Enzymatic assays were carried out as described in crude-lysate screening assay section.

Multi-site randomization

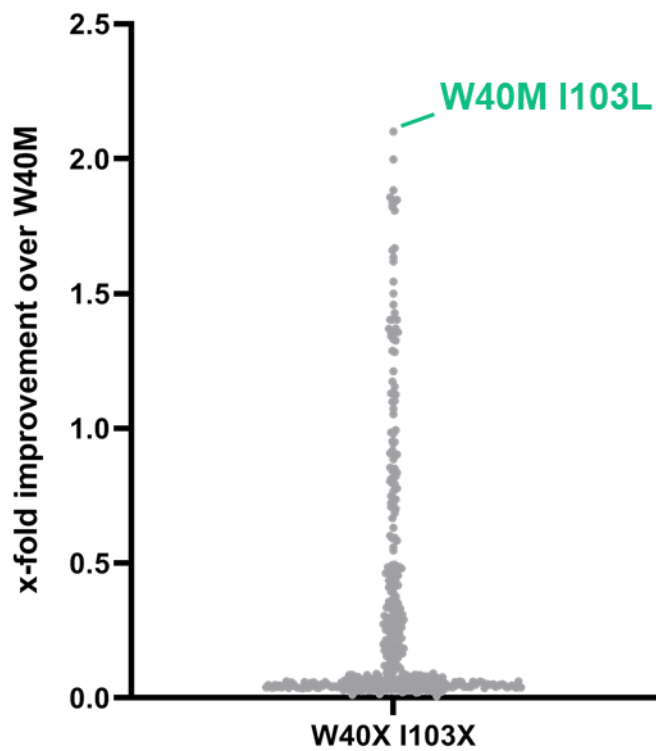


Figure S4 2nd evolution round of SmP4H towards improved hydroxylation of L-homophenylalanine based on multi-site randomization at identified hotspots W40 and I103. Using DBK codons at position W40 and DYA codons at position I103 resulting in 108 combinations. Identification of W40M I103L with 2-fold improved activity over SmP4H W40M. Enzymatic assays were carried out as described in crude-lysate screening assay section.

Variant	K_M L-hPhe (1) [mM]	k_{cat} min ⁻¹	Rel. k_{cat}
WT	1.10 ± 0.24	0.015 ± 0.001	1
W40M	0.24 ± 0.04	0.124 ± 0.003	8
W40M I103L	0.40 ± 0.08	1.68 ± 0.07	112

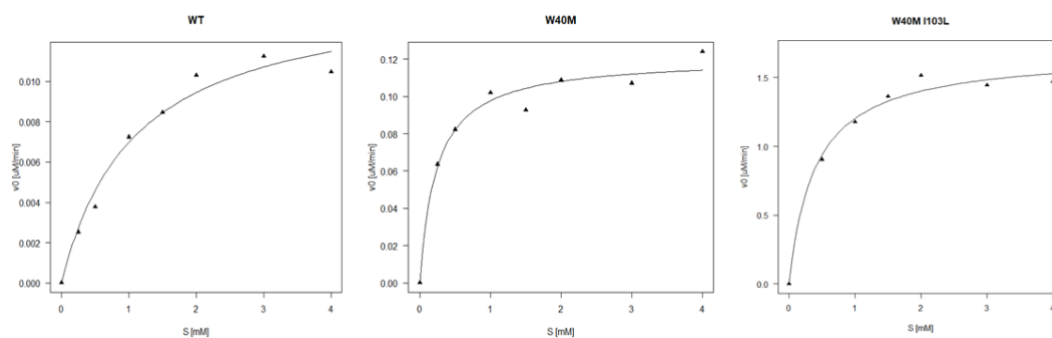
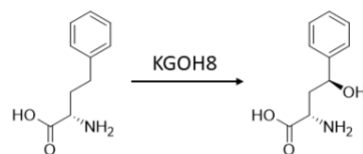


Figure S5 Michaelis-Menten kinetics of SmP4H wildtype, W40M and W40M I103L variant evaluated by monitoring production of γ -hydroxylated L-hPhe from L-homophenylalanine. Calibration curve was calculated based on a standard curve from γ -hydroxylated L-hPhe. Given are apparent k_{cat} and K_m values.

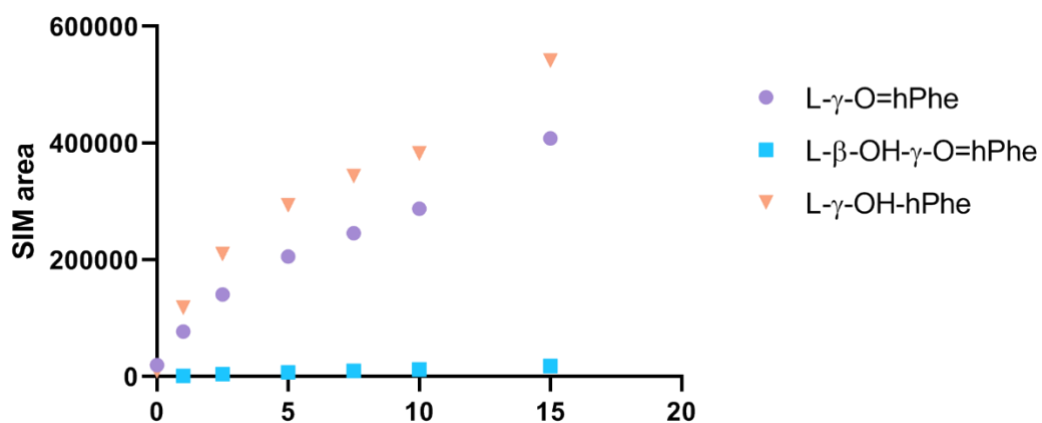


Figure S6 Time course of purified W40M I103L shows γ -hydroxylation of L-homophenylalanine is performed first followed by oxidation to result in L- γ -O=hPhe and second hydroxylation at the β -position.

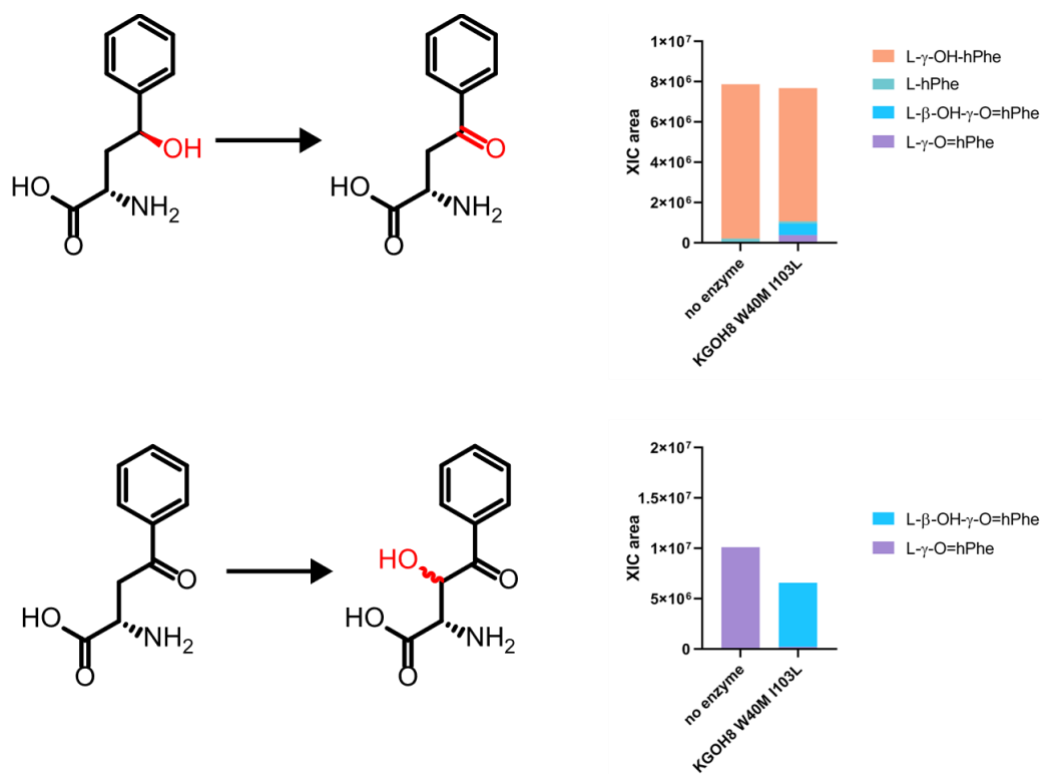


Figure S7 Incubation of intermediates L-γ-OH-hPhe and L-γ-O=hPhe with KGOH8 W40M I103L compared to no enzyme controls resulted in formation of subsequent products indicating an enzymatic cascade oxidation cascade to compounds 3 and 4 from the γ-hydroxylated product 2.

Variant	K_M L- γ -O=hPhe (3) [mM]	k_{cat} min ⁻¹	Rel. k_{cat}
WT	0.42 ± 0.07	0.22 ± 0.01	1
W40M	0.54 ± 0.12	2.4 ± 0.14	11
W40M I103L	0.71 ± 0.12	23.6 ± 1.3	107

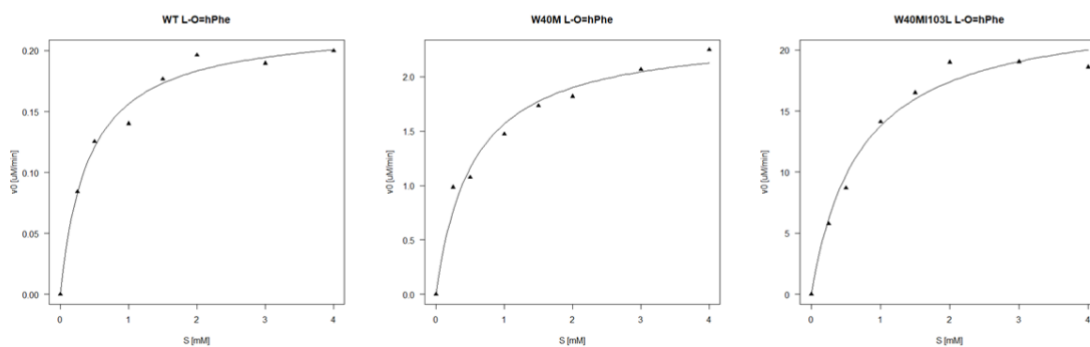
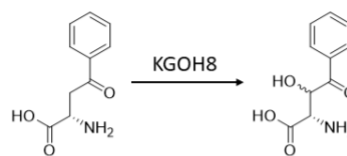


Figure S8 Michaelis-Menten kinetics of SmP4H wildtype, W40M and W40M I103L variant evaluated by monitoring production of β -hydroxy- γ -ketone L-hPhe from compound **3**. Calibration curve was calculated based on a standard curve from compound **3**. Given are apparent k_{cat} and K_m values.

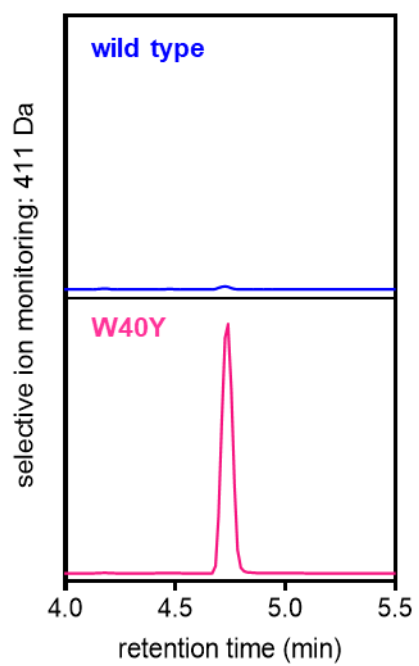


Figure S9 Selective ion monitoring revealed production of 3,4-desaturated L-homophenylalanine in case of SmP4H W40Y (red) variant compared with wild type SmP4H in blue.

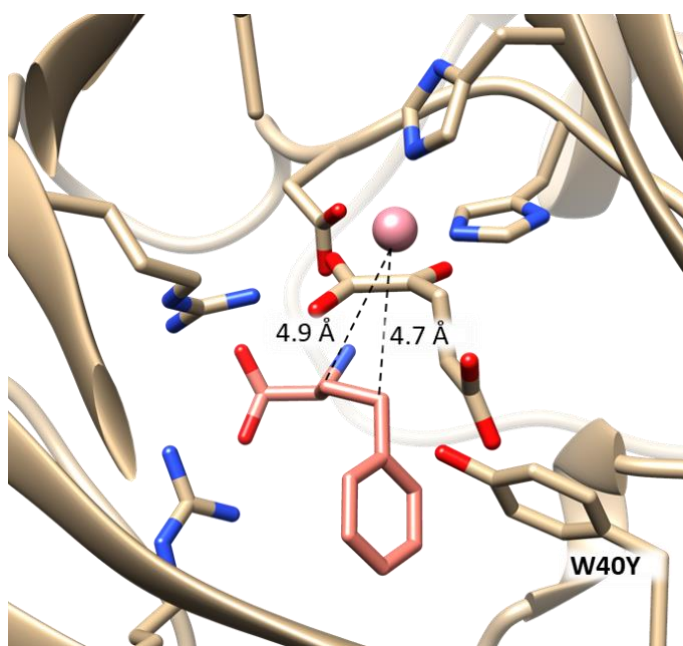


Figure S10: Docking solution found for W40Y with L-hPhe (docking solution ranked 3rd, $\Delta G = -6.0$ kcal/mol). Tyr-40 may cause the aromatic ring of the substrate to shift away slightly leading to a conformation that exposes both β - and γ -H atoms to the Fe(II) center in similar distances. In such a conformation, two subsequent H-atom transfers followed by radical dimerization may be favoured over rebound of a hydroxyl group. However, other docking solutions in which only the H-atom in γ -position is reasonably exposed for abstraction cannot be excluded. Alternatively, the proximal Tyr-40 may be directly involved in a radical mechanism leading to desaturation instead of hydroxylation (see main text).

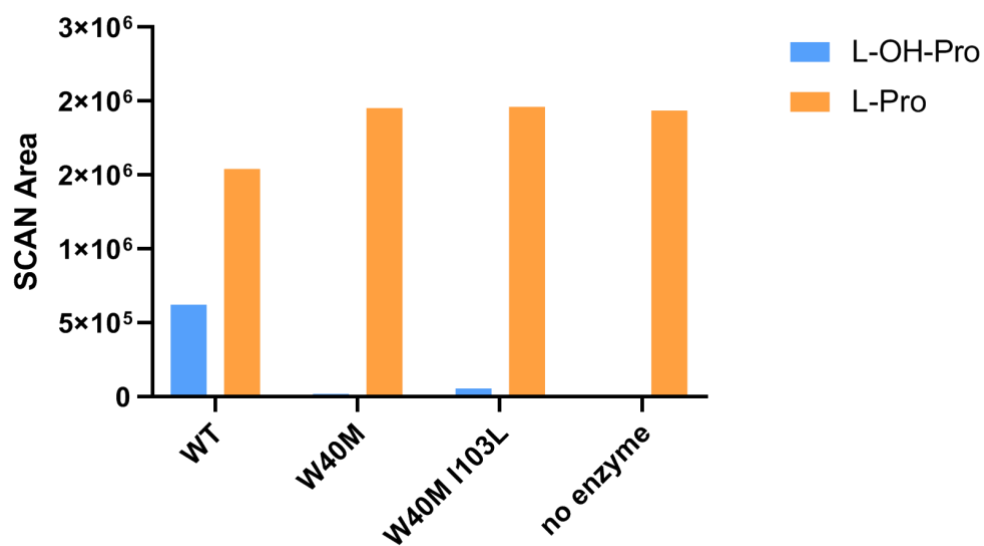


Figure S11 Biocatalysis with different SmP4H variants with L-proline as substrate. Reaction conditions: 2mM L-proline, 10 mM ascorbic acid, 1 μ M enzyme, 2h at 20°C. In case of the engineered SmP4H variants hydroxylation activity towards the native substrate L-proline is abolished.

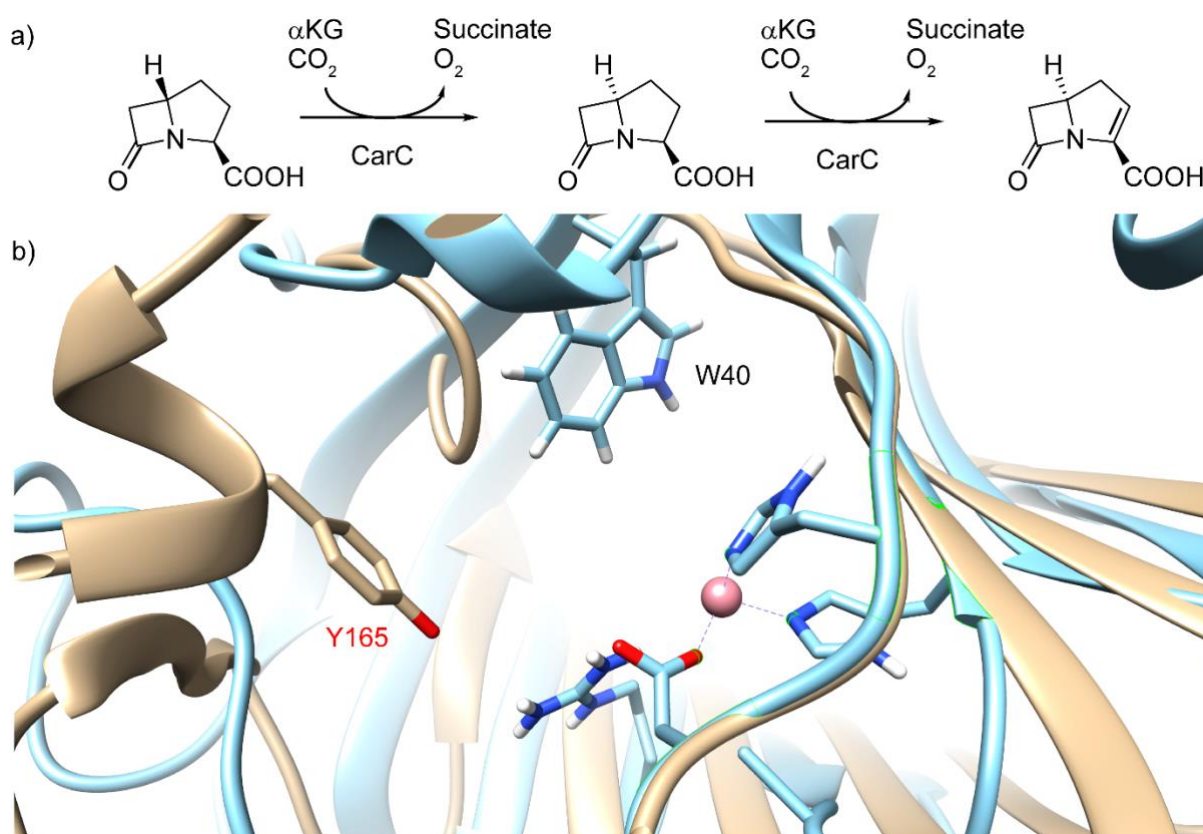


Figure S12 The proposed carbapenam biosynthetic pathway and structural alignment of active site between the homology model of Smp4H and the crystal structure of CarC. a) CarC catalyzes two subsequent reactions in the carbapenam biosynthesis, a stereoinversion reaction turning (3S, 5S)-carbapenam to (3S, 5R)-carbapenam and a subsequent desaturation reaction to (5R)-carbapenam⁴¹. b) Structural alignment of CarC and Smp4H using the catalytic histidine as anchor points. According to literature tyrosine 165 (red) from CarC acts as H-radical donor and is in similar distance to the active site as tryptophan 40 (black) in Smp4H, although on the opposing site.

Variant	K_M L-hPhe (1) [mM]	k_{cat} , min ⁻¹
W40Y	0.12 ± 0.02	0.83 ± 0.02

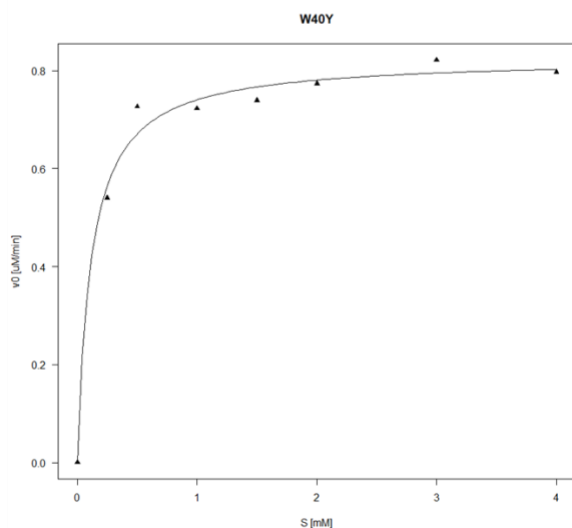
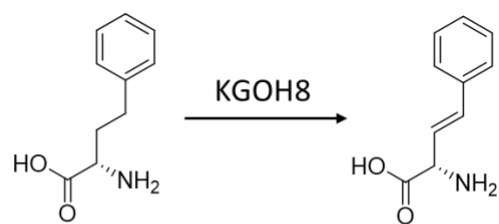


Figure S13 Michaelis-Menten kinetics of Smp4H W40Y variant evaluated by monitoring production of 3,4-desaturated L-hPhe (**5**) from compound **1**. Calibration was based on a standard curve of compound **1** assuming the same ionization efficiency for the product (**5**). The given values are apparent k_{cat} and K_m values.

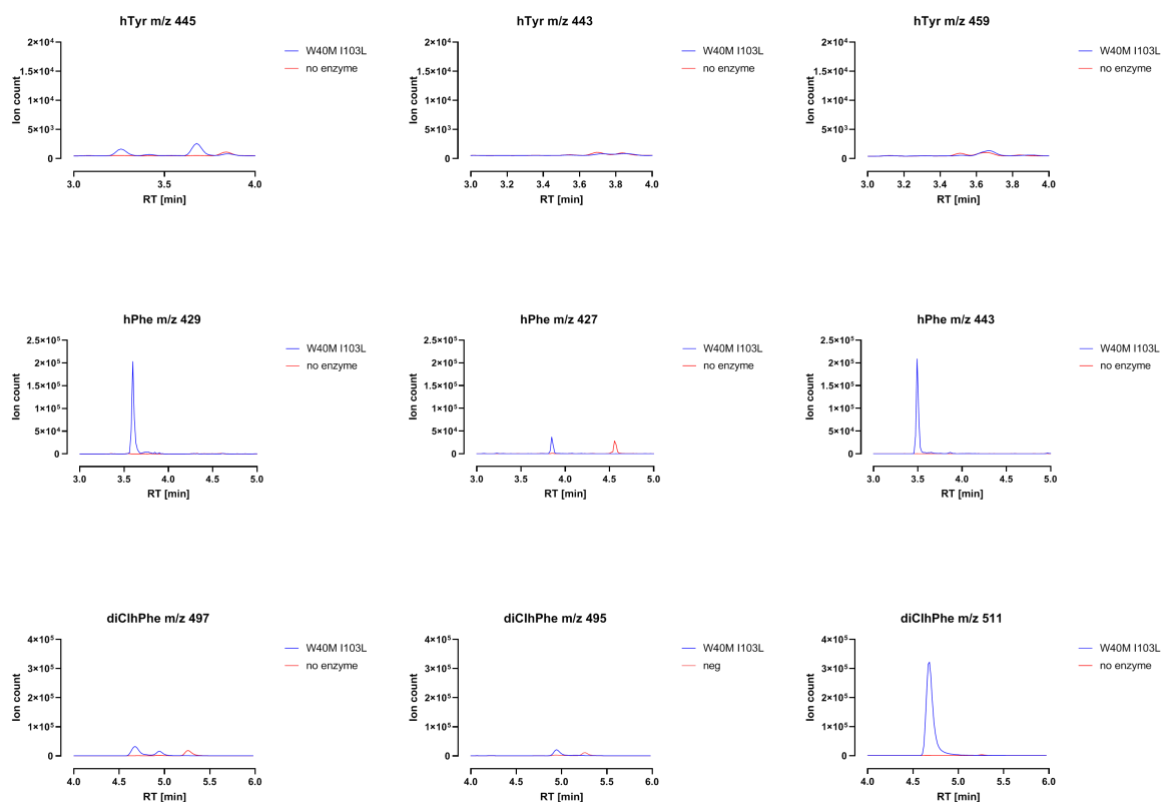


Figure S14 LC-MS chromatograms of Smp4H W40M I103L variant and no enzyme control with different substrates: L-homotyrosine (hTyr), L-homophenylalanine (hPhe) and (S)- α -amino-3,4-dichlorobenzenebutanoic acid (diClhPhe). Shown are single ion monitoring chromatograms with m/z values corresponding to dansylated products with m/z differences of $\Delta m/z = +16$, $\Delta m/z = +14$ and $\Delta m/z = +30$ to the substrate, as displayed in Figure 5. The additional peak detected in the single ion monitoring chromatograms recording $\Delta m/z = +16$ could potentially correspond to the formation of the respective hydroxylated diastereomers.

III. References

- (1) Reetz, M. T.; Carballeira, J. D. Iterative Saturation Mutagenesis (ISM) for Rapid Directed Evolution of Functional Enzymes. *Nat. Protoc.* **2007**, *2*, 891–903.
- (2) Chen H, Bong YK, Cabirol FL, Prafulchandra AG, Li T, Moore JC, Quintanar-Audelo M, Hong Y, Collier SJ, S. D. Biocatalysts and Methods for Hydroxylation of Chemical Compounds. US2017/0121744 A1. Codexis Inc., 2015.
- (3) Bienert, S.; Waterhouse, A.; De Beer, T. A. P.; Tauriello, G.; Studer, G.; Bordoli, L.; Schwede, T. The SWISS-MODEL Repository-New Features and Functionality. *Nucleic Acids Res.* **2017**, *45*, D313–D319.
- (4) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- (5) Grosdidier, A.; Zoete, V.; Michielin, O. SwissDock, a Protein-Small Molecule Docking Web Service Based on EADock DSS. *Nucleic Acids Res.* **2011**, *39*.
- (6) Sekirnik, R.; Wilkins, S. E.; Bush, J.; Tarhonskaya, · Hanna; Münzel, M.; Hussein, A.; Flashman, E.; Mohammed, S.; McDonough, M. A.; Christoph Loenarz, ·; Christopher, ·; Schofield, J. YcfD RM Is a Thermophilic Oxygen-Dependent Ribosomal Protein UL16 Oxygenase. **2018**, *22*, 553–562.
- (7) Baud, D.; Saaidi, P.-L.; Monfleur, A.; Harari, M.; Cuccaro, J.; Fossey, A.; Besnard, M.; Debard, A.; Mariage, A.; Pellouin, V.; Petit, J.-L.; Salanoubat, M.; Weissenbach, J.; de Berardinis, V.; Zapparucha, A. Synthesis of Mono- and Dihydroxylated Amino Acids with New α -Ketoglutarate-Dependent Dioxygenases: Biocatalytic Oxidation of C-H Bonds. *ChemCatChem* **2014**, *6*, 3012–3017.
- (8) Mitchell, A. J.; Dunham, N. P.; Martinie, R. J.; Bergman, J. A.; Pollock, C. J.; Hu, K.; Allen, B. D.; Chang, W.-C.; Silakov, A.; Bollinger, J. M.; Krebs, C.; Boal, A. K. Visualizing the Reaction Cycle in an Iron(II)-and 2-(Oxo)-Glutarate-Dependent Hydroxylase. *J. Am. Chem. Soc.* **2017**, *139*, 13830–13836.
- (9) Kodera, T.; Smirnov, S. V.; Samsonova, N. N.; Kozlov, Y. I.; Koyama, R.; Hibi, M.; Ogawa, J.; Yokozeki, K.; Shimizu, S. A Novel L-Isoleucine Hydroxylating Enzyme, l-Isoleucine Dioxygenase from *Bacillus Thuringiensis*, Produces (2S,3R,4S)-4-Hydroxyisoleucine. *Biochem. Biophys. Res. Commun.* **2009**, *390*, 506–510.
- (10) Nakayasu, M.; Umemoto, N.; Ohyama, K.; Fujimoto, Y.; Lee, H. J.; Watanabe, B.; Muranaka, T.; Saito, K.; Sugimoto, Y.; Mizutani, M. A Dioxygenase Catalyzes Steroid 16 α -Hydroxylation in Steroidal Glycoalkaloid Biosynthesis. *Plant Physiol.* **2017**, *175* (1), 120–133.
- (11) Hibi, M.; Kawashima, T.; Sokolov, P. M.; Smirnov, S. V.; Kodera, T.; Sugiyama, M.;

- Shimizu, S.; Yokozeki, K.; Ogawa, J. L-Leucine 5-Hydroxylase of *Nostoc Punctiforme* Is a Novel Type of Fe(II)/ α -Ketoglutarate-Dependent Dioxygenase That Is Useful as a Biocatalyst. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 2467–2472.
- (12) Eichhorn, E.; Van Der Ploeg, J. R.; Kertesz, M. A.; Leisinger, T. Characterization Of-
Ketoglutarate-Dependent Taurine Dioxygenase from *Escherichia Coli*. *J. Biol. Chem.* **1997**, *272*, 23031–23036.
- (13) 三宅良磨; 出来島康方. Method for Manufacturing Cis-5-Hydroxy-l-Pipecolic Acid. WO2016076159A1, 2016.
- (14) Hogan, D. A.; Auchtung, T. A.; Hausinger, R. P. Cloning and Characterization of a Sulfonate/ α -Ketoglutarate from Dioxygenase from *Saccharomyces Cerevisiae*. *J. Bacteriol.* **1999**, *181*, 5876–5879.
- (15) Mattay, J.; Hüttel, W. Pipecolic Acid Hydroxylases: A Monophyletic Clade among Cis-Selective Bacterial Proline Hydroxylases That Discriminates l-Proline. *ChemBioChem* **2017**, *18*, 1523–1528.
- (16) Bräuer, A.; Beck, P.; Hintermann, L.; Groll, M. Structure of the Dioxygenase AsqJ: Mechanistic Insights into a One-Pot Multistep Quinolone Antibiotic Biosynthesis. *Angew. Chemie Int. Ed.* **2016**, *55*, 422–426.
- (17) Hara, R.; Kino, K. Characterization of Novel 2-Oxoglutarate Dependent Dioxygenases Converting l-Proline to Cis-4-Hydroxy-l-Proline. *Biochem. Biophys. Res. Commun.* **2009**, *379*, 882–886.
- (18) Ji, J.; Fan, K.; Tian, X.; Zhang, X.; Zhang, Y.; Yang, K. Iterative Combinatorial Mutagenesis as an Effective Strategy for Generation of Deacetoxycephalosporin C Synthase with Improved Activity toward Penicillin G. *Appl. Environ. Microbiol.* **2012**, *78* (21), 7809–7812.
- (19) Houwaart, S.; Youssar, L.; Hüttel, W. Pneumocandin Biosynthesis: Involvement of a Trans - Selective Proline Hydroxylase. *ChemBioChem* **2014**, *15*, 2365–2369.
- (20) Shibasaki, T.; Mori, H.; Ozaki, A. Cloning of an Isozyme of Proline 3-Hydroxylase and Its Purification from Recombinant *Escherichia Coli*. *Biotechnol. Lett.* **2000**, *22*, 1967–1973.
- (21) Zhao, Z.; Zhang, Y.; Liu, X.; Zhang, X.; Liu, S.; Yu, X.; Ren, Y.; Zheng, X.; Zhou, K.; Jiang, L.; Guo, X.; Gai, Y.; Wu, C.; Zhai, H.; Wang, H.; Wan, J. Developmental Cell A Role for a Dioxygenase in Auxin Metabolism and Reproductive Development in Rice. *Dev. Cell* **2013**, *27*, 113–122.
- (22) Method for Enhancing Activity of Trans-4-Hydroxyproline Biosynthesis System Containing Recombinant DNA. CN104278047A, 2013.
- (23) McDonough, M. A.; Kavanagh, K. L.; Butler, D.; Searls, T.; Oppermann, U.; Schofield, C. J. Structure of Human Phytanoyl-CoA 2-Hydroxylase Identifies Molecular Mechanisms of Refsum Disease. *J. Biol. Chem.* **2005**, *280*, 41101–41110.

- (24) Strieker, M.; Essen, L.-O.; Walsh, C. T.; Marahiel, M. A. Non-Heme Hydroxylase Engineering For Simple Enzymatic Synthesis Of L-Threo-Hydroxyaspartic Acid. *ChemBioChem* **2008**, *9*, 374–376.
- (25) Fraser, P. D.; Miura, Y.; Misawa, N. In Vitro Characterization of Astaxanthin Biosynthetic Enzymes. *J Biol Chem.* **1997**, *272*, 6128–6135.
- (26) Huang, Z.; Wang, K.-K. A.; Lee, J.; Van Der Donk, W. A. Biosynthesis of Fosfazinomycin Is a Convergent Process. *Chem. Sci.* **2015**, *6*, 1282–1287.
- (27) Gibbons, H. S.; Lin, S.; Cotter, R. J.; Raetz, C. R. H. Oxygen Requirement for the Biosynthesis of the S-2-Hydroxymyristate Moiety in Salmonella Typhimurium Lipid A. Function of LpxO, A New Fe²⁺/Alpha-Ketoglutarate-Dependent Dioxygenase Homologue. *J Biol Chem.* **2000**, *275*, 32940–32949.
- (28) Qin, H.-M.; Miyakawa, T.; Jia, M. Z.; Nakamura, A.; Ohtsuka, J.; Xue, Y.-L.; Kawashima, T.; Kasahara, T.; Hibi, M.; Ogawa, J.; Tanokura, M. Crystal Structure of a Novel N-Substituted L-Amino Acid Dioxygenase from Burkholderia Ambifaria AMMD. *PLoS One* **2013**, *8* (5), e63996.
- (29) Meng, S.; Han, W.; Zhao, J.; Jian, X.; Pan, H.; Tang, G. A Six-Oxidase Cascade for Tandem C–H Bond Activation Revealed by Reconstitution of Bicyclomycin Biosynthesis. *Angew. Chemie Int. Ed.* **2018**, *57*, 719–723.
- (30) Longbotham, J. E.; Levy, C.; Johannissen, L. O.; Tarhonskaya, H.; Jiang, S.; Loenarz, C.; Flashman, E.; Hay, S.; Schofield, C. J.; Scrutton, N. S. Structure and Mechanism of a Viral Collagen Prolyl Hydroxylase. *Biochemistry* **2015**, *54*, 6093–6105.
- (31) You, Z.; Omura, S.; Ikeda, H.; Cane, D. E.; Jogl, G. Crystal Structure of the Non-Heme Iron Dioxygenase PtlH in Pentalenolactone Biosynthesis. *J. Biol. Chem.* **2007**, *282*, 36552–36560.
- (32) Reddy, Y. V.; Al Temimi, A. H. K.; White, P. B.; Mecinovi, J. M. Evidence That Trimethyllysine Hydroxylase Catalyzes the Formation of (2S,3S)-3-Hydroxy-N ε-Trimethyllysine. *Org. Lett.* **2017**, *19*, 400–403.
- (33) Correia Cordeiro, R. S.; Enoki, J.; Busch, F.; Mügge, C.; Kourist, R. Cloning and Characterization of a New Delta-Specific L-Leucine Dioxygenase from Anabaena Variabilis. *J. Biotechnol.* **2018**, *284*, 68–74.
- (34) Montero-Morán, G. M.; Li, M.; Rendón-Huerta, E.; Jourdan, F.; Lowe, D. J.; Stumpff-Kane, A. W.; Feig, M.; Scazzocchio, C.; Hausinger, R. P. Purification and Characterization of the Fe II-and R-Ketoglutarate-Dependent Xanthine Hydroxylase from Aspergillus Nidulans. *Biochemistry* **2007**, *46*, 5293–5304.
- (35) Nakashima, Y.; Mori, T.; Nakamura, H.; Awakawa, T.; Hoshino, S.; Senda, M.; Senda, T.; Abe, I. Structure Function and Engineering of Multifunctional Non-Heme Iron Dependent Oxygenases in Fungal Meroterpenoid Biosynthesis. *Nat. Commun.* **2018**, *9*, 104.
- (36) Desieno, M. A.; Van Der Donk, W. A.; Zhao, H. Characterization and Application of the

- Fe(II) and α -Ketoglutarate Dependent Hydroxylase FrbJ. *Chem. Commun.* **2011**, 47 (36), 10025–10027.
- (37) Müller, T. A.; Fleischmann, T.; Van Der Meer, J. R.; Kohler, H. P. E. Purification and Characterization of Two Enantioselective α -Ketoglutarate-Dependent Dioxygenases, RdpA and SdpA, from *Sphingomonas Herbicidovorans* MH. *Appl. Environ. Microbiol.* **2006**, 72, 4853–4861.
- (38) Zwick, C. R.; Renata, H. Remote C-H Hydroxylation by an α -Ketoglutarate-Dependent Dioxygenase Enables Efficient Chemoenzymatic Synthesis of Manzacidin C and Proline Analogs. *J. Am. Chem. Soc.* **2018**, 140, 1165–1169.
- (39) Lau, W.; Sattely, E. S. Six Enzymes from Mayapple That Complete the Biosynthetic Pathway to the Etoposide Aglycone. *Science* **2015**, 349, 1224–1228.
- (40) Voss, M.; Honda Malca, S.; Buller, R. M. U. Exploring the Biocatalytic Potential of Fe/ α -Ketoglutarate Dependent Halogenases. *Chem. Eur. J.* **2020**, 26, 7336.
- (41) Chang, W. C.; Guo, Y.; Wang, C.; Butch, S. E.; Rosenzweig, A. C.; Boal, A. K.; Krebs, C.; Bollinger, J. M. Mechanism of the C5 Stereoinversion Reaction in the Biosynthesis of Carbapenem Antibiotics. *Science* **2014**, 343, 1140–1144.