Supplementary information

Directed evolution of C-methyltransferase PsmD for enantioselective pyrroloindole derivative production

Diana A. Amariei,[a] Julia Tenhaef,[b] Thomas Classen,[b] Benoit David,[c] Tobias M. Rosch,[b] Holger Gohlke,[c,d] Stephan Noack[b] and Jörg Pietruszka* [a,b]

Table of content

Methods

Automated Screening Process

The protein expression module and enzymatic reaction module were run on an automated platform using a Spinnaker2 BT (Thermo Scientific, Canada) for all material transfer between devices and the process was orchestrated by the Momentum Software (ThermoScientific, Canada). The activity assay module was performed using an OT-2 liquid handler (Opentrons, USA).

Protein expression module

Per library, 66 colonies resulting from the mutant library transformation were picked using a liquid handler (Tecan Fluent,Tecan, Switzerland) with integrated colony picker (Pickolo, SciRobotics, Israel) and used to inoculate 900 µL of LB media in a square well deep well plate (sqDWP). Additionally, controls were carried on each plate. Cells transformed with WT PsmD were used as positive and cells transformed with the empty pET21a(+) vector as negative control. After incubation for 16 h at 800 rpm and 37°C in an automated incubator (Cytomat2, Thermo Scientific, Germany), 10 µL of the precultures were transferred to a fresh sqDWP to inoculate 900 µL TB-autoinduction media and were incubated for another 24 h before being passed down to the enzymatic reaction module. The TB media contained additionally 0.5% (w/v) glycerol, 0.05% (w/v) glucose and 0.2 (w/v) lactose for auto-induction.

Enzymatic reaction module

After cultivation, the cells were separated by centrifugation for 5 min at 3500 rpm (Rotanda 450R, Hettich, Germany) and the TB medium was removed using the liquid handler. The cells were resuspended in 200 µL freshly prepared reaction mixture containing 1 mM substrate (100 mM stocks in DMSO were used, leading to 1% v/v DMSO content in the final mixture) and 2 mM SAM (98% purity, BLD Pharm, cat no. BD20061) in KPi buffer (50 mM, pH 7.5). The reaction was performed in a sqDWP at 40°C for 24, 7 or 3 h in an automated incubator (Cytomat 10, ThermoScientific, Germany). After the reaction, the cells were quenched by addition of 20 µL 5% TFA and separated via centrifugation. 100 µL of the supernatant was transferred to a fresh microtiter plate (MTP) and stored at -20°C until further use.

Activity assay module

The activity assay was performed using an Opentrons® OT-2 robot. For the activity assay 50 µL of the supernatant resulting from the enzymatic reaction module were mixed with 50 µL H_2S0_4 98% and incubated for 30 min at room temperature. After, 50 µL DMAB solution (300 mM in isopropanol) was added to each well, mixing thoroughly. After another 1 h incubation step at room temperature, the absorbance at 580 nm was measured using a plate reader. Heat maps were generated based on absorbance values and the cells providing the lowest absorbance at 580 nm were subjected to plasmid isolation and sequencing.

Mutant library generation

The amino acids in the selected positions were modified by iterative saturation mutagenesis. The mutant libraries were generated using the 22c-trick. This reduces codon redundancy, leading to reduced libraries.[1] Degenerated primer mixtures were used in a whole plasmid amplification PCR approach. The codon harmonized PsmD_*Sa* gene in pET21a(+) was purchased from Genscript and used as template. The reaction mixtures contained 1x PrimeStar GXL buffer, 200 µM (each) dNTPs, 300 nM (each) forward and reverse primers, 0.2 ng/µL template and 1 µL PrimeStar DNA polymerase. The initial denaturation was performed at 98 °C for 30 s, after which the program was continued with 17 cycles of denaturation (98 °C, 10 s), annealing (touchdown 72 °C to 56 °C 15 s) and elongation (68 °C, 3 min). After, 25 cycles of denaturation (98 °C, 10 s), annealing (55 °C 15 s) and elongation (68 °C, 3 min) were performed. The final annealing was performed at 68°C for 5 min. The PCR products were subjected to digestion with DpnI then analyzed by agarose (1%) gel electrophoresis, using Midori Green as a stain. After purification from the gel, the products were subjected to ligation with T4 DNA ligase in the presence of 6% (v/v) PEG 4000. After the inactivation of the ligase, 10 µL of the PCR product was used to transform chemically competent *E.coli* DH5α cells using heat shock. All the obtained colonies were collected and mixed and the plasmid mixtures were isolated using a plasmid isolation kit (Innuprep, Analytik Jena). The pure primer mixtures were analyzed by sequencing and used to transform via heat shock chemically competent *E. coli* BL21 Gold (DE3) for protein expression.

Protein expression

E. coli BL21 Gold (DE3) precultures were incubated with LB and ampicillin for 16 h. After, TB autoinduction media containing additionally 0.5% (w/v) glycerol, 0.05% (w/v) glucose and 0.2 (w/v) lactose was inoculated and incubated at 35 °C for 24 h. For the large-scale protein expression, regular TB + ampicillin was used and the induction was started by adding 100 µM IPTG (final concentration) at OD_{600} of 0.6-0.8.

Protein purification

The isolation of PsmD_Sa mutants was performed by immobilized metal affinity chromatography. After protein expression, the cells were resuspended in lysis buffer (50 mM KPi, pH 7.5) to the final concentration of 0.2 mg/mL. The cell disruption was carried out by sonication (Branson Sonifier II "Modell W-250", Heinemann) and the cell debris was separated by centrifugation (10000 rpm, 4 °C, 30 min). A 5 mL Ni-NTA column (Superflow Cartridge, QIAGEN GmbH, Hilden, Germany) was used for purification. The column was equilibrated with 3 CV of equilibration buffer (50 mM KPi, 500 mM NaCl, pH 7.5) and the filtered cell lysate was loaded on the column. The column was washed with 5 CV of washing buffer (50 mM KPi, 500 mM NaCl, 100 mM imidazole, pH 7.5), and for elution 3 CV of elution buffer (50 mM KPi, 500 mM NaCl, 250 mM imidazole, pH 7.5) were used. The final protein solution was concentrated using Vivaspin® 20 centrifugal concentrators (MWCO 10 kDa, Sartorius, Germany). To remove the imidazole, the protein solution was washed three times with 20 mL of lysis buffer. The purified protein was stored at -20 °C with 25% v/v glycerol.

Activity determination using isolated enzymes

The specific activity of the selected hits was determined using the MTase-Glo™ assay (Promega, USA).[2] The reaction mixture contained 20 µM substrate, 30 µM SAM (Promega, provided in the MTase-Glo™ assay kit) and 1 µg enzyme in 1x MTase-Glo™ buffer (20mM Tris buffer, pH 8, 50 mM NaCl, 1 mM EDTA, 3 mM $MgCl_2$, 0.1 mg/ml BSA, 1 mM dithiothreitol) in a final volume of 20 µL. The mixtures were incubated at 40 °C for 15 min in Nunc™ flat-bottom white 96-well plates, after which the assay was performed according to the manufacturer's protocol. A SAH calibration curve was produced for the quantification of SAH final concentration. The luminescence was measured using a Tecan Infinite M1000Pro microplate reader. The measurements were performed at 23 °C, with no attenuation and an integration time of 1000 ms.

Determination of enantioselectivity

For the analysis of enantioselectivity, normal phase HPLC analysis was performed on a Lux Amylose-1, 250 x 4.6 mM, 5 µM chiral column. 1 mM substrate **2a**, 2 mM SAM (98% purity, BLD Pharm, cat no. BD20061), and 30 µg enzyme were mixed in Kpi buffer (50 mM, pH 7.5) in a total volume of 200 µL. The reaction mixture was incubated at 40 °C and 800 rpm for 20 h. After incubation, the substrate and product were extracted with 3 x 200 µL ethyl acetate. After evaporation of the extraction solvent, the samples were solved in 200 ml elution solvent mixture (90:10 n-heptane:2-propanol). The samples were run through the column at a rate of 1 mL/min, using 20 µL injection volume, with detection at 205 nm for a total of 90 min per run.

Enzyme immobilization

After protein expression, 32 mL lysate was prepared using 0.2 g *E. coli* BL21 Gold cells containing W166C mutant/mL buffer (KPi 50 mM, pH 7.5). 3.2 mL (10% v/v) Ni-NTA agarose resin (Protino, Germany) slurry was added to the lysate and the mixture was incubated on ice for 1 h, mixing gently by rotation. The mixture was then centrifuged for 2 min at 3000 rpm, the supernatant removed and the resin was resuspended in 30 mL washing buffer (KPi 50 mM, 50 mM imidazole, pH 7.5) and incubated on ice for 5 min. The mixture was centrifuged for 2 min at 3000 rpm and the previous washing step was repeated. The resin was then washed once with reaction buffer (KPi 50 mM, pH 7.5) and was separated by centrifugation to be used in the preparative reaction. The immobilization of CtHMT was performed using the same procedure from 22 mL lysate, using 2.2 mL Ni-NTA resin slurry.

Preparative enzymatic reaction

The preparative enzymatic methylation of substrate **2a** was performed using the mutant W166C and CtHMT immobilized on Ni-NTA resin. The resin containing immobilized W166C and the resin containing CtHMT were mixed and resuspended in 79 mL reaction mixture containing 50 mg substrate **2a** (2 mM final concentration, added from 200 mM stock in DMSO), 0.2 mM SAH, 10 mM MeI (added from a 10 M stock

in DMSO) in KPi buffer (50 mM, pH 7.5). The mixture was incubated at 35 °C and 300 rpm for 15 h. After, the Ni-NTA resin was separated by centrifugation and the product was extracted from the supernatant with 3 x 100 mL ethyl acetate. The organic phase was then washed with brine and dried with $MgSO_4$. After filtration and solvent evaporation, the product was isolated by column chromatography using a DCM:MeOH (9:1) mixture for elution. The product was obtained as a white powder with a 60% yield.

NMR measurements

^{1}H and ^{13}C NMR measurements were performed using an Advance/DRX 600 nuclear magnetic resonance spectrometer (Bruker, USA) in $CDCl_3$, at 600 and 151 MHz respectively. The chemical shifts were determined relative to the solvent.

Homology model generation

The coordinates of the homology model previously described in the literature for WT PsmD_*Sa* were used, initially generated using Modeller and the crystal structure of PsmD from *S. griseofuscus* in the closed form (PDB: 7ZKG) as a template.[3-5] The mutant models were generated by replacing residue W166 using the rotamer function in UCSF Chimera.[6] The conformation of the new residue was chosen from the Dunbrack 2010 library, based on the highest probability.[7]

Molecular docking

A massive probing of the active site of PsmD_*Sa* wt, W166P, and W166C was performed using molecular docking of the substrate **2a** and **3a** with *AutoDock Vina* (V1.2.3) employing the *Vinardo* scoring function.[8-10] The configuration parameters were:

center_x = 38.45; center_y = 41.65; center_z = 43.62; size_x = 20.90; size_y = 25.30; size_z = 32.27; scoring = vinardo; energy_range = 3; exhaustiveness = 8; num_modes = 50; cpu = 11

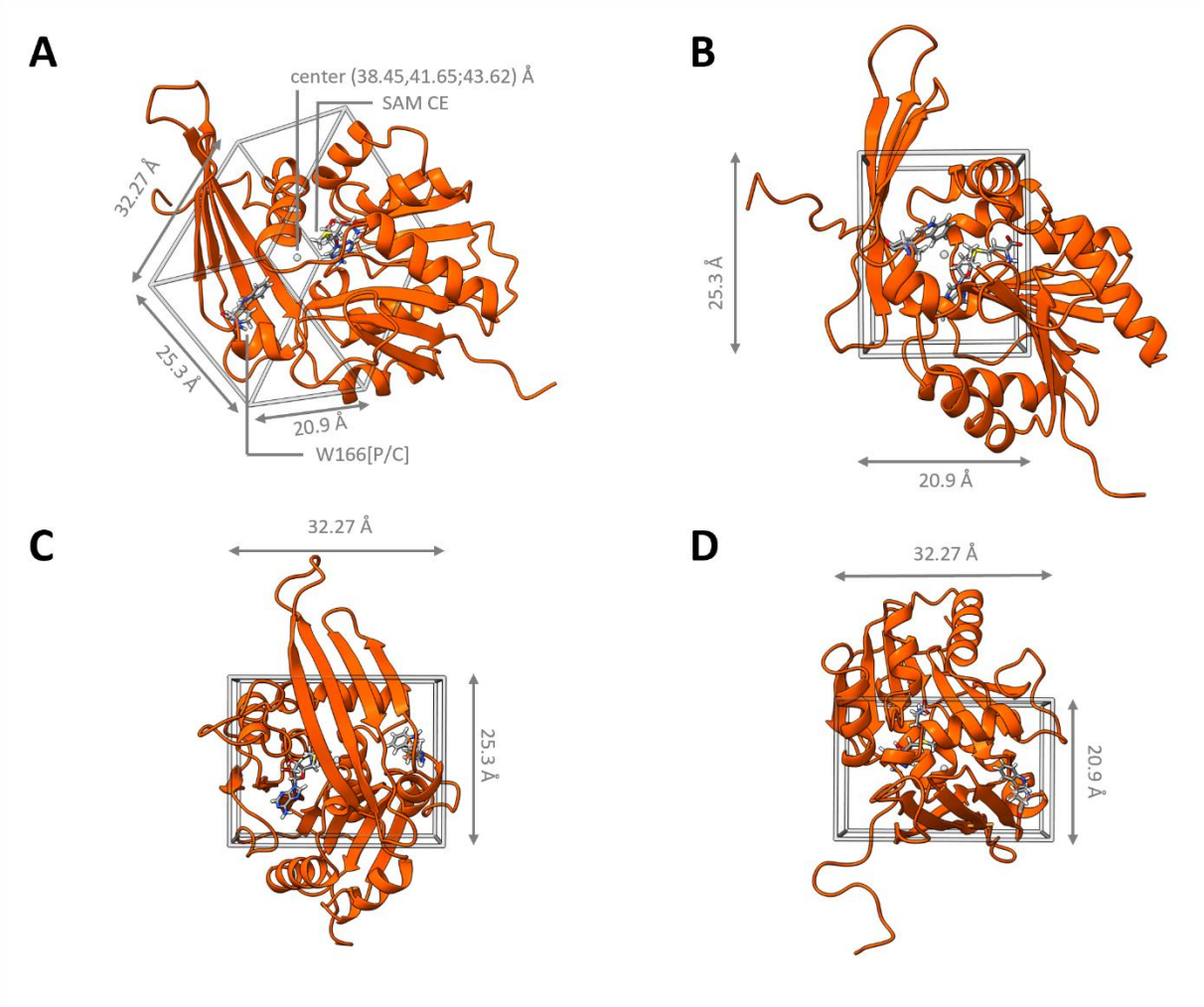


Figure S1. Docking box size and position within the structure of PsmD_*Sa*. A (*vide supra* homology model generation) – Perspective overview. The center of the box is shown as a grey sphere juxtaposed with the reactive methyl group of SAM (atom ID: CE). Residue 166 is shown explicitly, superimposing tryptophan, proline, and cysteine. B-C – Orthogonal, axial perspectives.

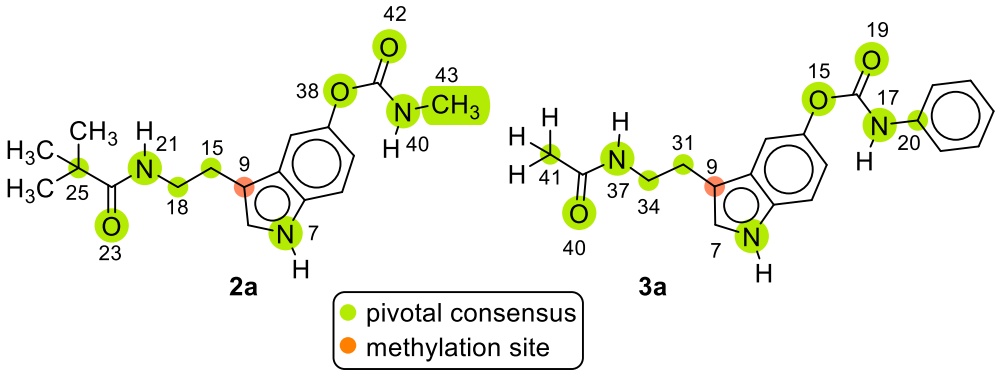
The energy range was set to be restrictive (3 kcal mol^{-1}) to obtain binding with a highest affinity only. Each docking may yield 50 poses. To probe the conformational space allowing for quantitative analyses, the docking for each combination ([**2a**,**3a**] x [wt,W166C,W166P]) was repeated 40 times. Thus, 2000 docking poses could have been obtained for each combination. Due to the set energy restraint, only 11,567 binding poses were obtained for the six combinations.

Scripting/coding in bash, python, Windows PowerShell, and UCSF Chimera were supported by ChatGPT as implemented in the search engine Ecosia.

Substrate **2a** has 23 non-hydrogen atoms. Thus, the Cartesian coordinate set has 69 dimensions. For substrate **3a** these are 75. For the comparative analysis, the dimensionality of the data set was reduced by selecting pivotal atoms only, representing the entire degrees of freedom of the configuration. E.g., the

entire orientation of the indole ring and directly attached atoms can be represented by the atom numbers 7, 15, and 38 for substrate **2a**. The selected pivotal atoms are shown in Scheme S1. Additionally, the nucleophile carbon number 9 was extracted from the binding modes, too. In total, the configuration was represented by 11 atoms leading to a 33-dimensional data set.

This reduced data set of all 11,567 poses was subjected to principal component analysis, independent of the substrate type or enzyme docked into, extracting three principal components representing 67% (PC1), 9% (PC2), and 6% (PC3) of the entire variance. This was done only to make the data set manageable as a three-dimensional data set representing the obtained configurational space.



Scheme S1. Pivotal atoms representing the configuration (green). The nucleophile carbon is highlighted in orange. The numbers are atom ids from the docking files (PDBQT).

This three-dimensional configurational data set (principal component) was clustered by distance into 10 clusters. These clusters represent different binding poses. The putative productive binding pose was selected based on the structural and mechanistic information previously acquired for the homolog PsmD_*Sg*.[11] Within the present data set, the productive binding pose corresponds to cluster 6.

Detailed information on, e.g., substrate type, enzyme environment, binding energy, distance between nucleophile and electrophile, were projected onto the data set (*cf.* Figures S2 and S3).

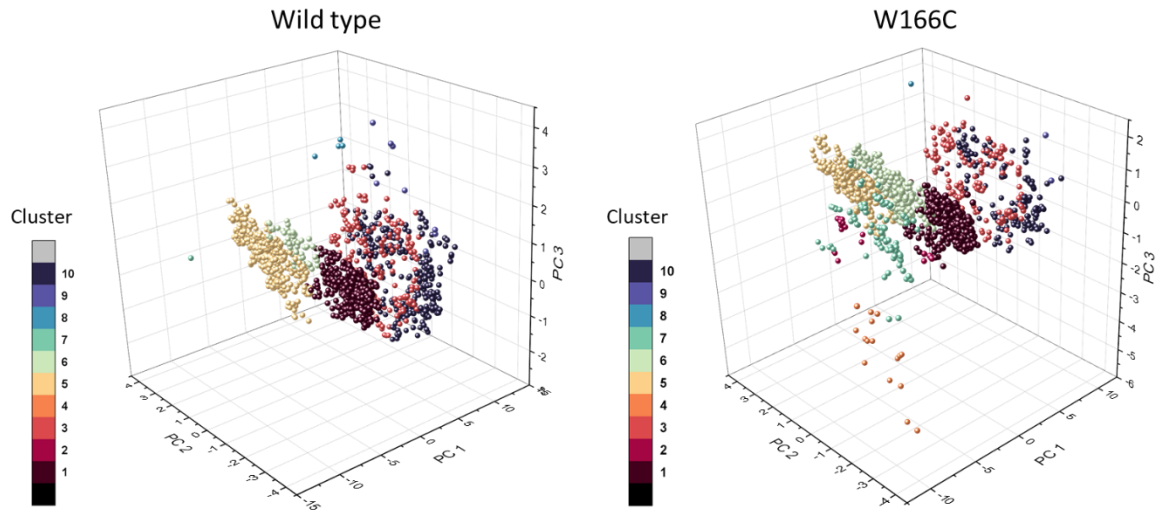
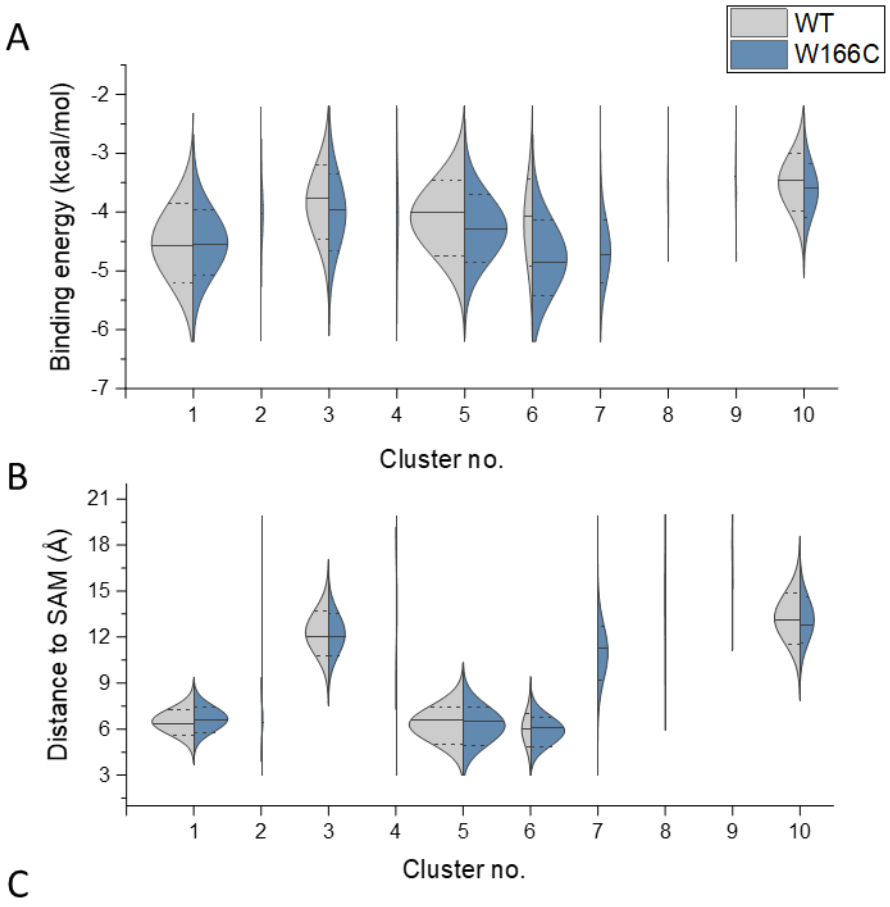


Figure S2. Cluster representation after principal component analysis of the obtained docking poses of **2a** in the catalytic site of WT PsmD_*Sa* and mutant W166C.



C

Cluster	1	2	3	4	5	6	7	8	9	10
N [WT]	541	0	297	0	701	114	1	7	8	330
N [W166C]	441	23	216	16	543	442	130	2	2	184

Figure S3. Clustering of substrate **2a** poses in WT PsmD_*Sa* and the mutant W166C, and the normal distribution of calculated parameters in each cluster. The median is represented in a solid line; the dashed lines indicate the standard deviation. Cluster 6 is considered the active binding mode, according to previous studies on PsmD_*Sa*. **A.** Distribution of calculated binding energies per pose cluster. **B.** Distribution of the distance between the SAM methyl group and the methylation site per pose cluster. **C.** Number of samples assigned to each cluster in the principal component analysis.

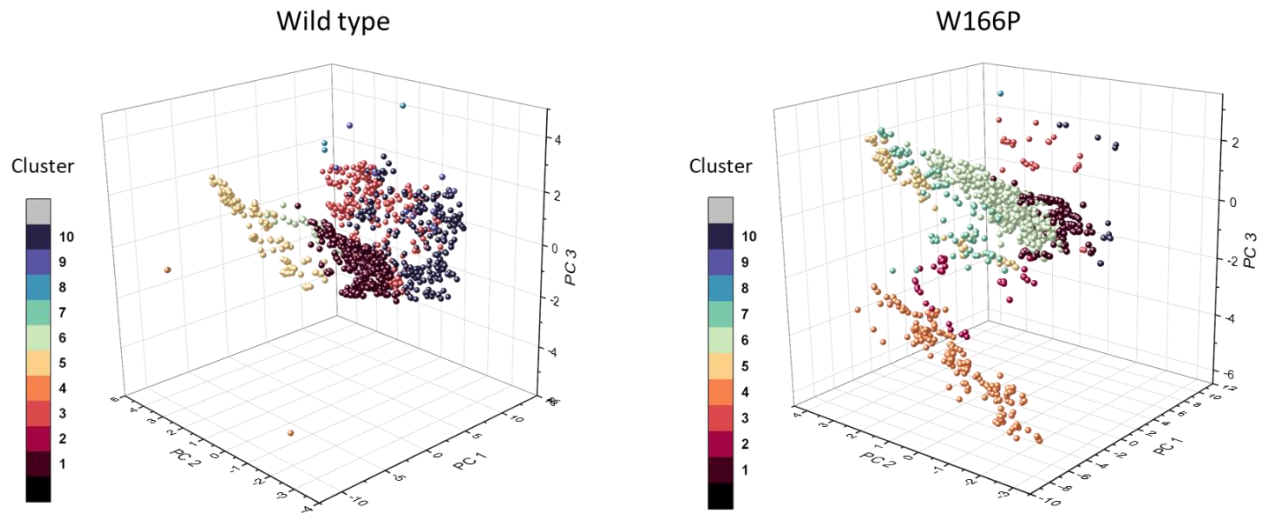
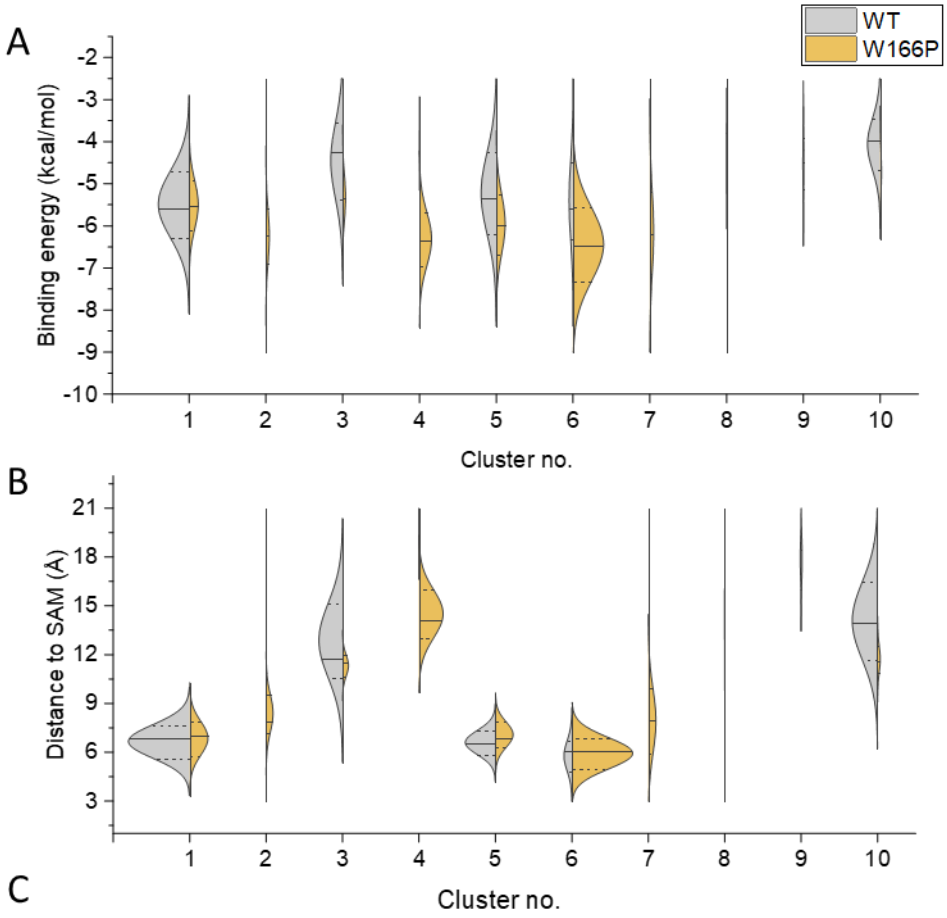


Figure S4. Cluster representation after principal component analysis of the obtained docking poses of **3a** in the catalytic site of WT PsmD_*Sa* and mutant W166P.

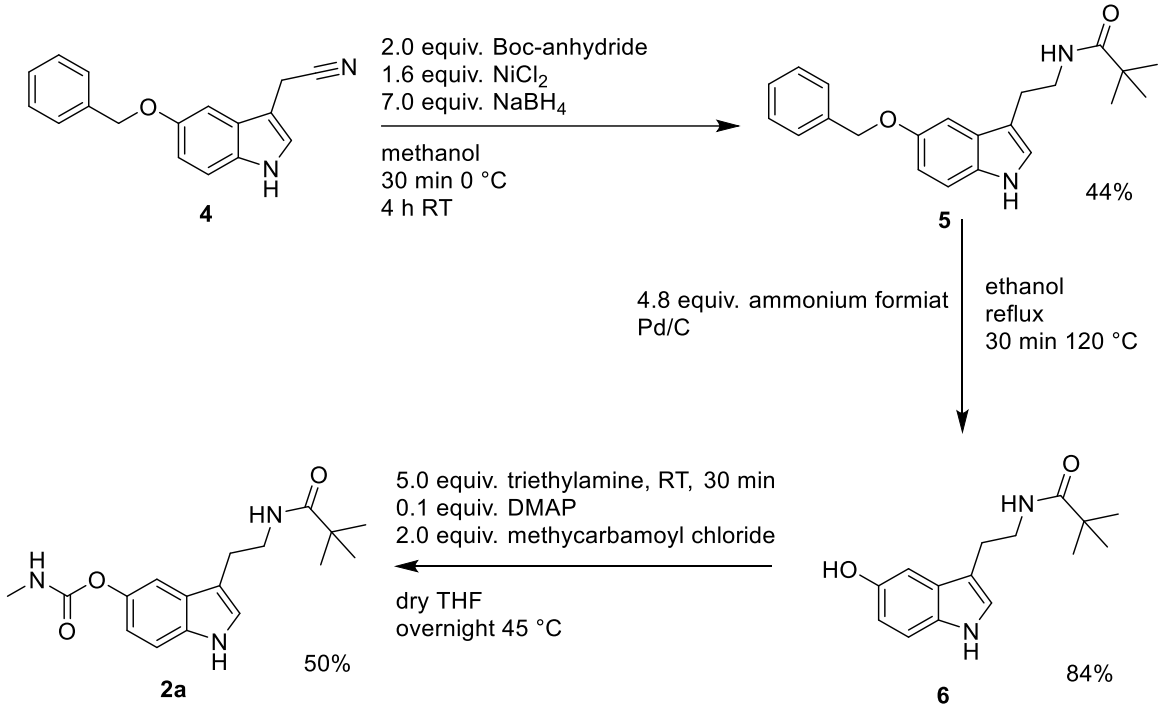


Cluster	1	2	3	4	5	6	7	8	9	10
N [WT]	762	0	294	2	375	98	1	3	11	315
N [W166P]	227	86	73	300	220	757	95	1	0	32

Figure S5. Clustering of substrate **3a** poses in WT PsmD_*Sa* and the mutant W166P and the normal distribution of calculated parameters in each cluster. The median is represented in a solid line; the dashed lines indicate the standard deviation. Cluster 6 is considered the active binding mode, according to previous studies on PsmD_*Sa*. **A.** Distribution of calculated binding energies per pose cluster. **B.** Distribution of distances between the SAM methyl group and the methylation site per pose cluster. **C.** Number of samples assigned to each cluster in the principal component analysis.

Substrate synthesis

Substrates **2a** and **3a** were synthesized according to literature.[12]



Scheme S2. Synthesis route towards **2a**.

Synthesis of *N*-(2-(5-(benzyloxy)-1*H*-indol-3-yl)ethyl)pivalamide (5)

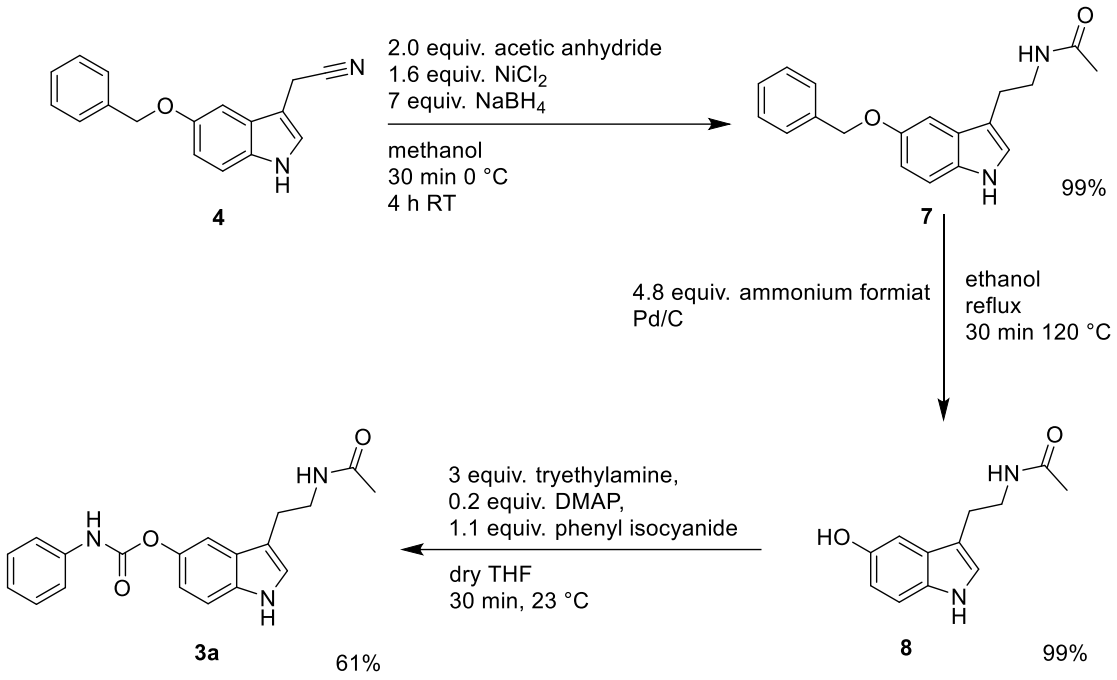
500 mg 2-(5-(benzyloxy)-1*H*-indol-3-yl)acetonitrile (**4**) were solved in 20 mL methanol under stirring at 0 °C on an ice bath. Afterward, the anhydride and the $NiCl_2$ were added. Finally the $NaBH_4$ was added slowly. The reaction was incubated with stirring at 0 °C for 30 min. After, the reaction mix was incubated for 4 h at room temperature. The reaction was monitored by TLC (solvent: EtOAc). After, the solvent was evaporated and the remaining solid was solved in 20 mL EtOAc and the mixture was filtered. The filtrate was washed with 2 x 10 mL saturated $NaHCO_3$ solution and brine. The two phases were separated and the organic phase was dried with $MgSO_4$. After filtration, the solvent was evaporated. The product was separated by flash chromatography (solvent: EtOAc).

Synthesis of *N*-(2-(5-hydroxy-1*H*-indol-3-yl)ethyl)pivalamide (6)

Ammonium formate and Pd/C were added to a round bottom flask. **5** was solved in 20 mL ethanol and added to the mixture. The reaction mix was stirred for 30 min at 120 °C, at reflux. After, the mixture was cooled down and filtered, washing the filtrate with EtOAc. The solvent was then removed by evaporation. The crude mixture was used in the next step.

Synthesis of 3-(2-pivalamidoethyl)-1*H*-indol-5-yl methylcarbamate (2a)

All steps were performed under inert conditions. The crude **6** product was solved in 15 mL dry THF under N_2 atmosphere. Triethylamine was added to the mixture and stirred at room temperature for 30 min. DMAP was then added, followed by the methylcarbamoyl chloride. The reaction mixture was stirred and incubated at 45 ˚C under inert atmosphere overnight. Samples from the reaction were analysed via TLC, using petrol ether:EtOAc (1:1) as elution system. Afterwards, the reaction was quenched by addition of KPi buffer (50 mM, pH 7.5) until the mixture clarified. The product was extracted 3 times with EtOAc and then dried on $MgSO_4$. The solid was filtered, and the solvent evaporated. The product and remaining substrate were isolated via flash chromatography using EtOAc containing 2,5% (v%) MeOH as elution system. 105 mg of product **2a** were obtained, as a white solid.



Scheme S3. Synthesis route towards **3a**.

Synthesis of *N*-(2-(5-(benzyloxy)-1*H*-indol-3-yl)ethyl)acetamide (7)

500 mg 2-(5-(benzyloxy)-1*H*-indol-3-yl)acetonitrile (**4**) were solved in 20 mL methanol under stirring at 0 °C on an ice bath. After, the anhydride, and the $NiCl_2$ were added. Finally the $NaBH_4$ was added slowly. The reaction was incubated with stirring at 0 °C for 30 min. After, the reaction mix was incubated overnight at room temperature. The reaction was monitored by TLC (solvent: EtOAc). After the reaction, the solvent was evaporated and the remaining solid was solved in 20 mL EtOAc, and the mixture was filtered. The filtrate was washed with 2 x 10 mL saturated $NaHCO_3$ solution and brine. The two phases were separated and the organic phase was dried with $MgSO_4$. After filtration, the solvent was evaporated, obtaining a

brown oil. Full conversion to product **7** was achieved and the crude product was used directly in the next step.

Synthesis of *N*-(2-(5-hydroxy-1*H*-indol-3-yl)ethyl)acetamide (8)

Ammonium formate and Pd/C were added to a round bottom flask. Crude **7** was solved in 20 mL ethanol and added to the mixture. The reaction mix was stirred for 30 min at 120 °C, at reflux. After, the mixture was cooled down and filtered, washing the filtrate with EtOAc. The solvent was then removed by evaporation. The product was purified by flash chromatography, using EtOAc with 7% (v%) MeOH as elution system. 354 mg of product were obtained as a white powder.

Synthesis of 3-(2-acetamidoethyl)-1*H*-indol-5-yl phenylcarbamate (3a)

All steps were performed under inert conditions. Triethylamine and 8 were solved in 4 mL dry THF and were stirred under N_2 atmosphere for 15 min. Phenyl isocyanate was added dropwise over 5 s and then the DMAP was added. After stirring for 30 min at 23 °C, the reaction was quenched with saturated aqueous ammonium chloride solution (20 mL). The product was extracted with EtOAc (3 × 15 mL). The combined organic layers were washed with brine, dried over Na2SO4 and concentrated under reduced pressure. The resulting residue was purified by flash chromatography (9:1 DCM/MeOH). 151 mg 3a were obtained as a white powder.

Protein sequences

Chloracidobacterium thermophilum halide methyltransferase (*Ct*HMT)

MGHHHHHHAENLYFQGSGLGMDADTASFWEEKYRADLTAWDRGGVSPALEHWLAEGALKPGRILIPGCGYGHEVL
ALARRGFEVWGLDIALTPVRRLQEKLAQAGLTAHVVEGDVRTWQPEQPFDAVYEQTCLCALSPEDWPRYEAQLCRWL
RPGGRLFALWMQTDRPGGPPYHCGLEAMRVLFALERWRWV EPPQRTVPHPTGFFEYAAILERLV

Streptomyces albulus PsmD methyltransferase (PsmD_*Sa*)

MQGQPHQDAGMPEPYAATADVYDRLVAYAIAQWGESPRPRMADFIEQAWKARGQRVRRVLELCCGTGLMTEELV
RRGYEVTAVDRSETMLALAKKRVGGAADFRQIELPAPLPGDTDAVVCTAAAFNYQSSAHSLGETLHAVATVLPAGATF
VFDIETAALLKGHWGNRMWAADEGDLAFIWNFTSQPDTTYCDVHYTQFTRSEAGPDTYTGTREVHRLYAFDHDTVR
AQARAAGFARAEVFDNYTERPATDATHY ETWFLTRDESLEHHHHHH

Gene sequences

Codon harmonized PsmD_*Sa* gene:

ATGCAGGGACAGCCGCACCAGGATGCGGGTATGCCCGAGCCATACGCCGCGACGGCCGATGTGTACGACCGGCTCGTCGCGT
ATGCCATAGCCCAATGGGGAGAGTCTCCGCGGCCTCGGATGGCCGACTTCATCGAGCAGGCATGGAAGGCTCGTGGGCAGCG
CGTGCGCCGGGTGCTGGAGCTGTGTTGCGGCACCGGTCTGATGACTGAGGAACTGGTGCGGCGCGGTTACGAGGTGACAGC
TGTAGACCGTTCCGAGACCATGTTGGCCCTCGCGAAGAAGCGGGTCGGCGGTGCAGCCGACTTCCGACAGATCGAGCTCCCC
GCCCCGCTGCCCGGCGACACGGACGCAGTGGTATGCACCGCAGCCGCGTTCAACTACCAGTCCAGTGCACACTCACTGGGAG
AAACCCTACACGCCGTAGCCACGGTGCTGCCAGCCGGTGCAACGTTCGTCTTCGACATCGAGACTGCAGCGCTCCTCAAGGGA
CATTGGGGCAATCGCATGTGGGCCGCCGACGAGGGCGACCTGGCGTTCATCTGGAACTTCACTAGTCAGCCGGACACCACCTA
CTGCGACGTGCATTACACACAGTTCACGCGCTCTGAAGCGGGACCGGACACCTACACCGGCACCCGCGAGGTACACCGGCTGT
ACGCGTTCGACCACGACACCGTCCGCGCCCAGGCACGCGCCGCCGGATTCGCACGTGCGGAAGTGTTCGACAACTACACCGA
ACGCCCCGCCACCGACGCCACCCACTACGAGACGTGGTTCCTCACCCGGGACGAGAGC

Primers

Table S1. List of saturated mutagenesis targets and the respective primers used for the generation of mutant libraries.

PsmD_*Sa* variant	**Primer sequence (5'-3')**
Y197X	Fw: Phos-NDTACACAGTTCACGCGCTCTGAAGCGG Fw: Phos-VHGACACAGTTCACGCGCTCTGAAGCGG Fw: Phos-TGGACACAGTTCACGCGCTCTGAAGCGG Rev: Phos-ATGCACGTCGCAGTAGGTGGTGTCCGG
Y15X	Fw: Phos-CCCGAGCCANDTGCCGCGACGGCCG Fw: Phos-CCCGAGCCAVHGGCCGCGACGGCCG Fw: Phos-CCCGAGCCATGGGCCGCGACGGCCG Rev: Phos-CATACCCGCATCCTGGTGCGGCTG
W33X	Fw: Phos- CATAGCCCAANDTGGAGAGTCTCCGC Fw: Phos- CATAGCCCAAVHGGGAGAGTCTCCGC Rev: Phos- GCATACGCGACGAGCCGG
W166X	Fw: Phos-GGGACATNDTGGCAATCGCATGTGG Fw: Phos-GGGACATVHGGGCAATCGCATGTGG Rev: Phos-TTGAGGAGCGCTGCAGTCTCGATG

Supplementary figures

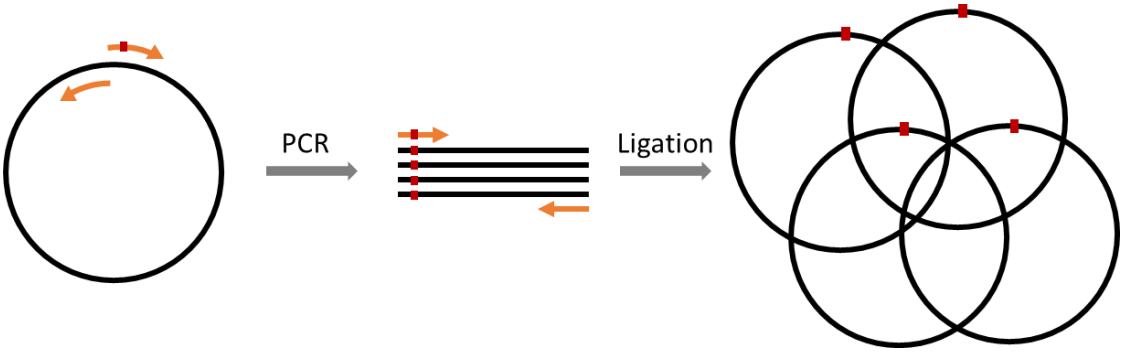


Figure S6. Procedure scheme for the genetic mutant library generation. The primers containing the mutation are degenerated and designed according to the 22c trick to produce all amino acids in the respective position.

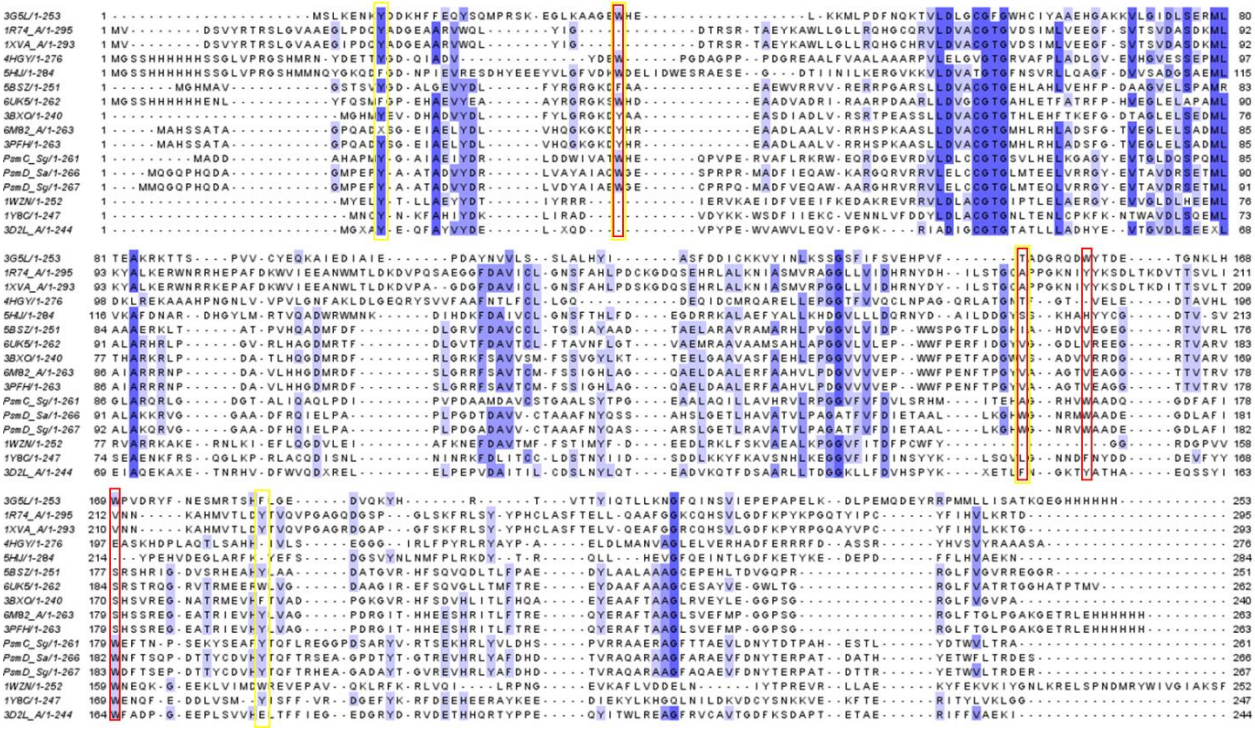


Figure S7. Sequence alignment of PsmD_*Sa* and other small molecule methyltransferases with similar structures. The blue color represents the percentage of the residues that agree with the consensus sequence. The residues in the "tryptophan cluster" positions from the catalytic site are marked with a red frame. The positions used as targets for saturation mutagenesis were marked with a yellow frame. The MTase structures were identified using Foldseek.[13] Only structures available in PDB were selected and the sequence labels contain the respective PDB codes. The sequence alignment was obtained using Clustal Omega and the figure was generated using Jalview.[14, 15]

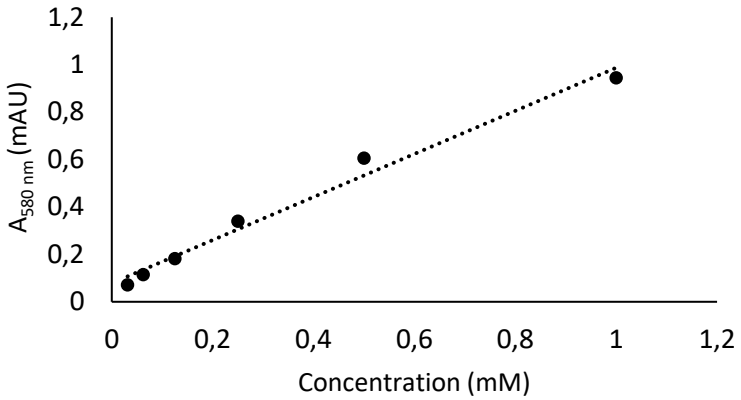


Figure S8. Colorimetric assay calibration of **2a** in KPi buffer (50 mM, pH 7.5)

<>	1	2	3	4	5	6	7	8	9	
A	0,30219999	0,1489	0,1881	0,2014	0,1287	0,23630001	0,2119	0,1321	0,14229999	
B	0,1139	0,2097	0,1934	0,12720001	0,1565	0,1664	0,14830001	0,15350001	0,1375	
C	0,1119	0,1441	0,1752	0,1276	0,1021	0,1181	0,1235	0,1514	0,2027	WT
D	0,1189	0,1761	0,1247	0,14030001	0,1202	0,1036	0,1222	0,1211	0,2026	
E	0,1602	0,1219	0,1118	0,182	0,1128	0,1663	0,1507	0,1705	0,23899999	EV
F	0,13850001	0,1595	0,1608	0,1595	0,19400001	0,17389999	0,1688	0,12819999	0,24240001	
G	0,15790001	0,19850001	0,1742	0,1267	0,1441	0,1619	0,1216	0,1313		
H	0,25440001	0,14830001	0,169	0,2484	0,1409	0,1575	0,115	0,1345		

Figure S9. Example of a heatmap obtained after a library screening round. The marked wells were selected for sequencing

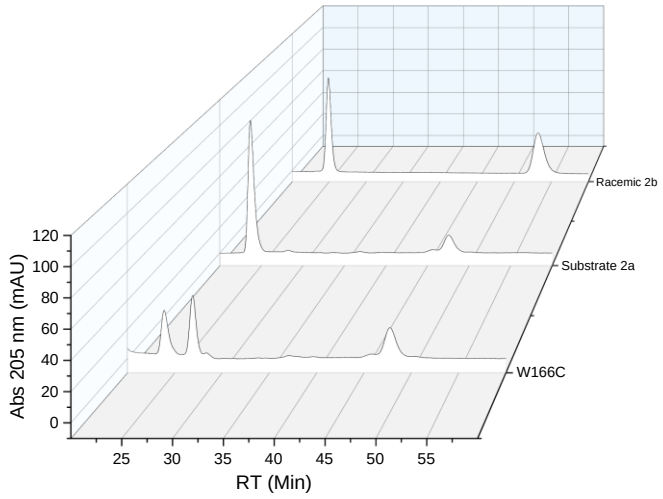


Figure S10. Chiral chromatography results comparing **2a** (peak at 24 min), the racemic **2b** (peaks at 25 and 53 min) and the reaction mixture of W166C and **2a** after 20 h incubation. The peak at 47 min corresponds to an unknown impurity.

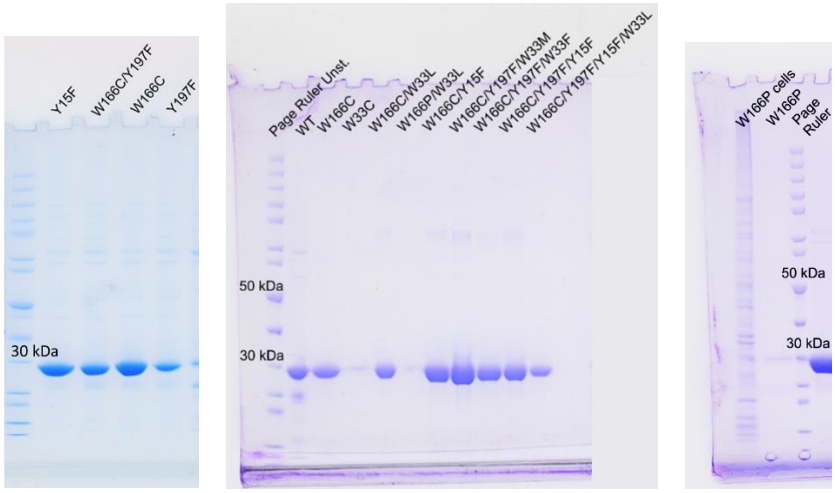


Figure S11. SDS Page gels of the isolated mutants. The expected protein band is at 30 kDa. The gel pictures were cropped for clarity, including only the relevant bands.

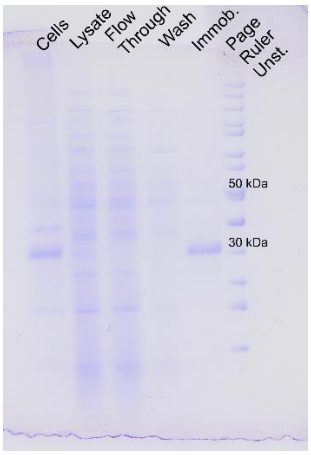


Figure S12. SDS Page gels of the immobilization steps for PsmD W166C. The expected protein band is at 30 kDa. The gel image was cropped for clarity.

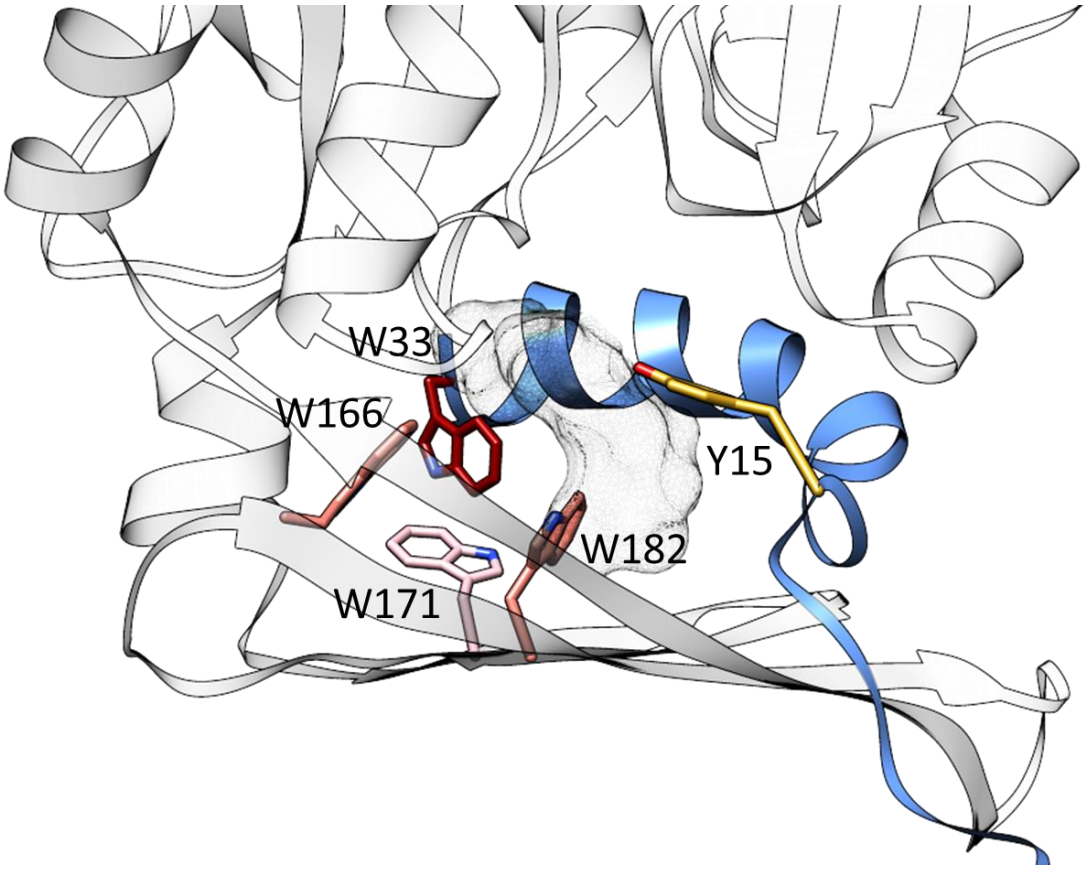


Figure S13. Homology model of PsmD_*Sa* with the docked natural substrate **1a** (surface represented as mesh). The tryptophan cluster is shown in relation to the catalytic site of the enzyme. The intensity of the color reflects the conservation of the tryptophan residues across MTase homologs (the intensity of the color increases with the conservation score of the residue) as calculated by the ConSurf web server.[16-18] The tryptophan cluster presumably anchors the α-helix supporting the N-terminal lid (here represented in blue). Our results show that substituting W33 leads to a loss of activity. On a structural level, we suggest that this change could alter the motion of the lid, thereby affecting the position of Y15 within the catalytic site (represented here in yellow).

Table S2. Site-specific mutagenesis results for PsmD_*Sa*, performed to assess the importance of the positions targeted by saturation mutagenesis. The site-specific mutagenesis was performed as described in our earlier work.[11] The reactions were performed using whole-cell biocatalysts, 1 mM substrate, and 2 mM SAM in KPi buffer (50 mM, pH 7.5), incubated at 35 °C for 16 h. The conversion was determined by RP-HPLC analysis.

Variant	Conversion (%)
Y15A	-
Y15F	>99%
W33A	14.8 ± 5.32
W166F	60.5 ± 4.39
Y197A	60.5 ± 9.66
WT PsmD_*Sa*	>99%

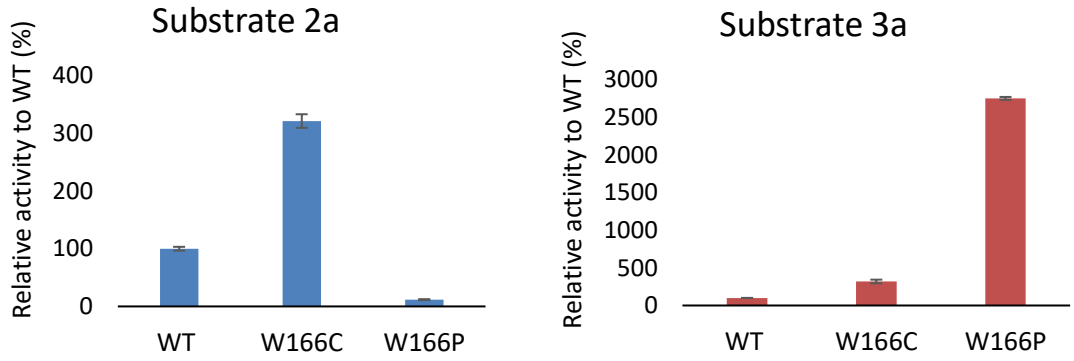
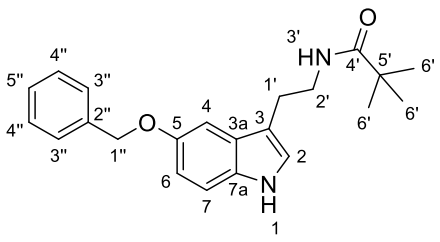


Figure S14. Comparison of the relative activities of W166C and W166P with WT PsmD towards substrates **2a** and **3a**. The activities were determined using the MTase Glo assay, as described in the experimental section, using 20 µM substrate and 30 µM SAM.

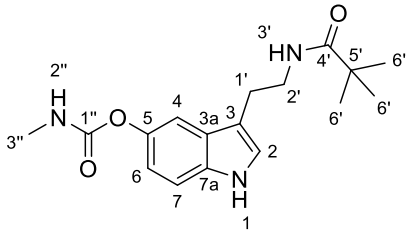
NMR data

***N*-(2-(5-(benzyloxy)-1*H*-indol-3-yl)ethyl)pivalamide (5)**



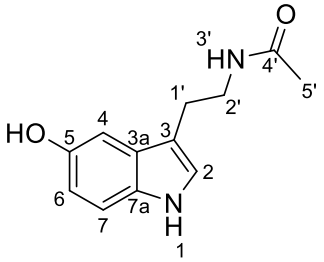
1H-NMR (600 MHz, methanol-d4): δ[ppm] = 7.48 (d, *J* = 7.3 Hz, 2H, 3''-H), 7.37 (t, *J* = 7.6 Hz, 2H, 4''-H), 7.30 (t, *J* = 8.0 Hz, 1H, 5''-H), 7.22 (d, *J* = 8.7 Hz, 1H, 7-H), 7.17 (s, 1H, 4-H), 7.03 (s, 1H, 2-H), 6.83 (d, *J* = 11.2 Hz, 1H, 6-H), 5.10 (s, 2H, 1''-H), 3.29 (t, *J* = 7.5 Hz, 2H, 2'-H), 2.85 (t, *J* = 7.4 Hz, 2H, 1'-H), 1.43 (s, 9H, 6'-H). **13C NMR** (151 MHz, methanol-d4): δ[ppm] = 158.54 (C-4'), 154.00 (C-5), 139.46 (C-2''), 133.63 (C-7a), 129.41 (C-4''), 129.16 (C-5''), 128.77 (C-3a), 128.70 (C-3''), 124.24 (C-2), 113.26 (C-6), 112.82 (C-7), 103.33 (C-4), 79.89(C-5'), 72.16 (C-1''), 42.45 (C-2'), 28.80 (C-6'), 26.94 (C-1'). **MS** (ESI): m/z calculated for $C_{17}H_{23}N_3O_3+K^+$: 389.163, $[M+K]^+$ found: 389.3.

3-(2-pivalamidoethyl)-1H-indol-5-yl methylcarbamate (2a)



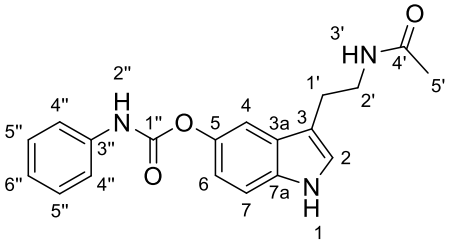
1H-NMR (600 MHz, methanol-d4): δ[ppm] = 7.31 – 7.24 (m, 2H, 7-H, 4-H), 7.09 (s, 1H, 2-H), 6.83 (d, *J* = 8.7 Hz, 1H, 6-H), 3.29 (t, *J* = 6.8 Hz, 2H, 2'-H), 2.86 (t, *J* = 7.3 Hz, 2H, 1'-H), 2.79 (s, 3H, 3''-H), 1.43 (s, 9H, 6'-H) **13C NMR** (151 MHz,methanol-d4): δ[ppm] = 158.96 (C-1''), 158.37 (C-4'), 145.34 (C-5), 135.64 (C-7a), 128.89 (C-3a), 124.82 (C-2), 116.69 (C-6), 113.67 (C-3), 112.24(C-7), 111.67 (C-4), 79.75 (C-5'), 42.35 (C-2'), 28.64(C-6'), 27.49 (C-3''), 26.63 (C-1'). **MS** (ESI): m/z calculated for $C_{17}H_{23}N_3O_3+K^+$: 356.137 $[M+K]^+$; found: 356.3.

***N*-(2-(5-hydroxy-1*H*-indol-3-yl)ethyl)acetamide (8)**



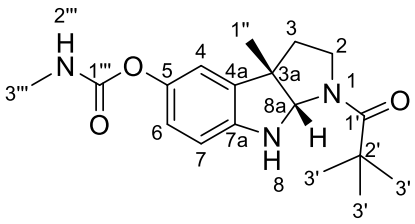
1H-NMR (600 MHz, methanol-d4): δ[ppm] = 7.15 (d, *J* = 8.6 Hz, 1H, 4-H), 6.99 (s, 1H, 2-H), 6.93 (d, *J* = 2.4 Hz, 1H, 7-H), 6.67 (dd, *J* = 8.6, 2.5 Hz, 1H, 6-H), 3.42 (t, *J* = 7.4 Hz, 2H, 2'-H), 2.85 (t, *J* = 7.4 Hz, 2H, 1'-H), 1.91 (s, 3H, 5'-H) **13C NMR** (151 MHz,methanol-d4): δ[ppm] = 173.26 (C-4'), 151.09 (C-5), 133.07 (C-7a), 129.48 (C-3a), 124.20 (C-2), 112.66 (C-3), 112.49 (C-4), 112.35 (C-6), 103.47 (C-7), 41.43 (C-2'), 26.22 (C-1'), 22.59 (C-5') **MS** (ESI): m/z calculated for $C_{17}H_{23}N_3O_3$+H^+: 219.1055 $[M+H]^+$; found: 219.2.

3-(2-acetamidoethyl)-1H-indol-5-yl phenylcarbamate (3a)



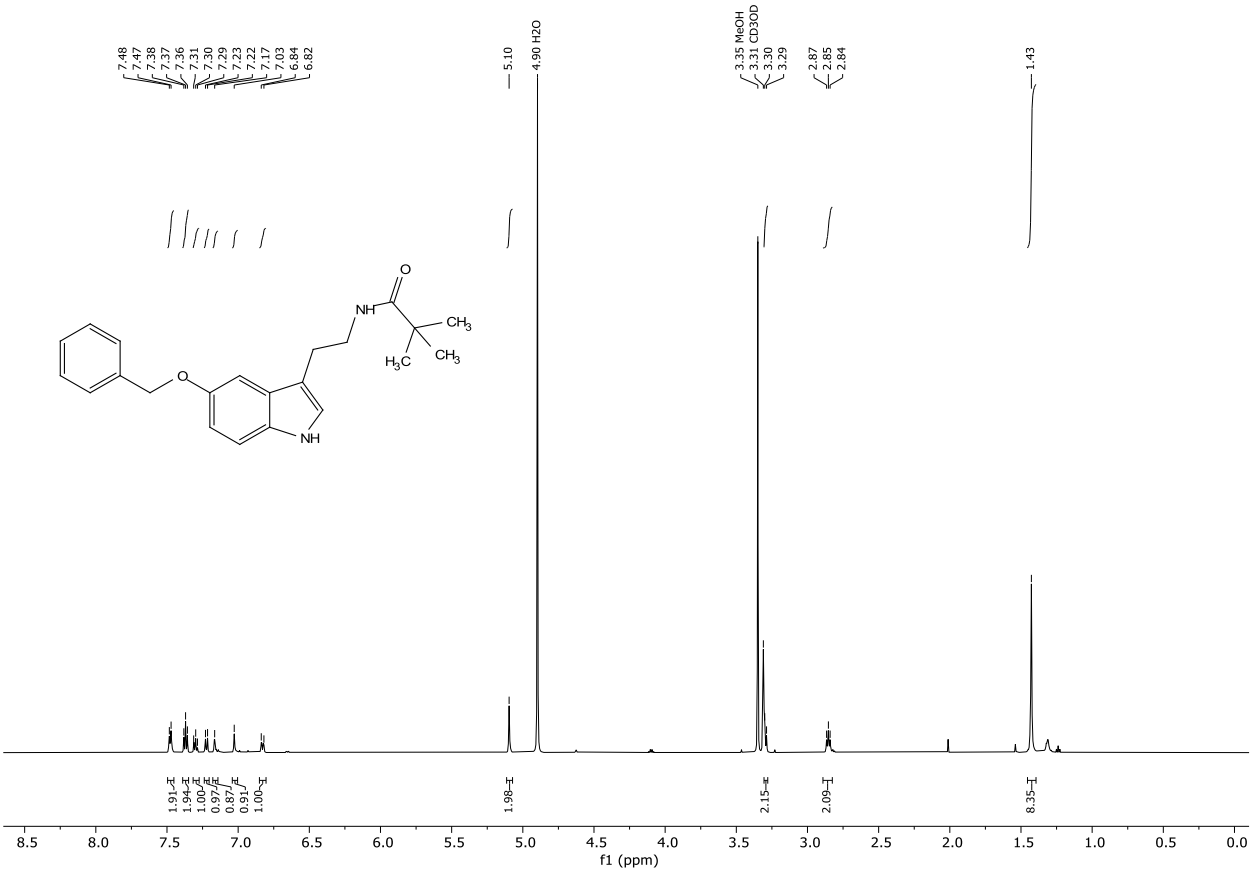
^{1}H-NMR (600 MHz, methanol-d4): δ[ppm] = 7.50 (d, *J* = 7.7 Hz, 2H, 4''-H), 7.37 – 7.27 (m, 4H, 7-H, 4-H, 5''-H), 7.14 (s, 1H, 2-H), 7.05 (t, *J* = 7.4 Hz, 1H, 6''-H), 6.92 (dd, *J* = 8.6, 2.3 Hz, 1H, 6-H), 3.45 (t, *J* = 7.3 Hz, 2H, 2'-H), 2.92 (t, *J* = 7.3 Hz, 2H, 1'-H), 1.91 (s, 3H, 5'-H). **^{13}C NMR** (151 MHz, methanol-d4): δ[ppm] = 173.34 (C-4'), 155.55 (C-1''), 145.14 (C-5), 140.07 (C-6''), 135.92 (C-7a), 129.90 (C-4''), 129.07 (C-3a), 125.05 (C-2), 124.31 (C-3''), 119.94 (C-5''), 116.87 (C-6), 113.73 (C-3), 112.5 (C-4), 111.8 (C-7), 41.5 (C-2'), 26.1 (C-1'), 22.6 (C-5') **MS** (ESI): m/z calculated for $C_{19}H_{19}N_3O_3$+H^+: 338.1426 $[M+H]^+$; found: 338.3.

(3a*S*,8a*S*)-3a-methyl-1-pivaloyl-1,2,3,3a,8,8a-hexahydropyrrolo[2,3-*b*]indol-5-yl methylcarbamate (2b)



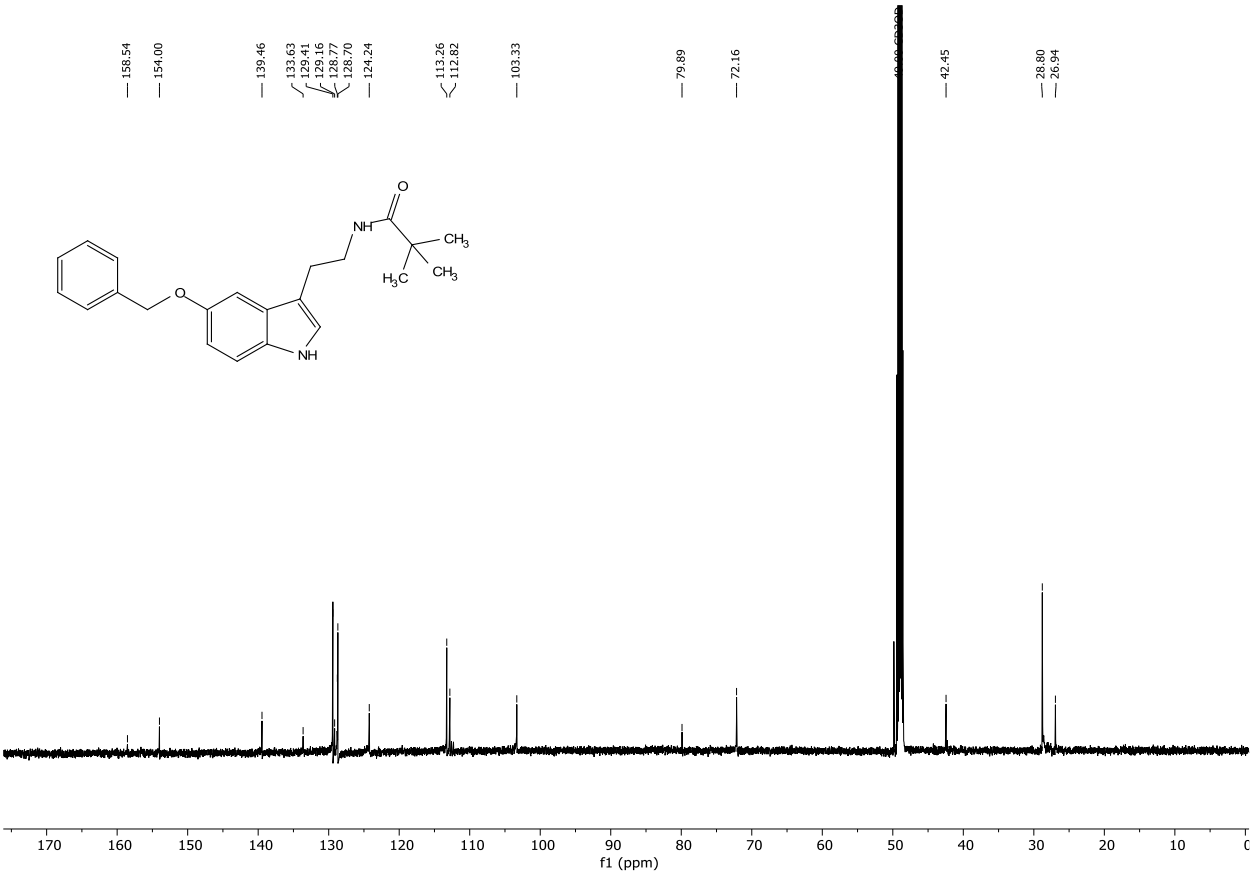
1H-NMR (600 MHz, methanol-d4): δ[ppm] = 6.83 (s, 1H, 6-H), 6.73 (d, *J* = 8.3 Hz, 1H, 7-H), 6.65 – 6.55 (m, 1H, 4-H), 5.02 (s, 1H, 8a-H), 3.65 – 3.56 (m, 1H), 3.05 – 2.97 (m, 1H), 2.76 (s, 3H, 3'''-H), 2.20 (q, *J* = 6.1 Hz, 1H), 2.08 – 1.99 (m, 1H), 1.53 (s, 5H), 1.46 (s, 4H), 1.39 (s, 3H, 1''-H) **13C NMR** (151 MHz,methanol-d4): δ[ppm] = 158.63 (C-1'''), 155.67 (C-1'), 147.72 (C-7a), 145.49 (C-5), 136.14 (C-4a), 122.28 (C-7), 117.50 (C-6), 110.44 (C-4), 83.84 (C-8a), 81.65 (C-2'), 55.20 (C-3a), 46.56 (C-2), 37.93 (C-3), 28.72 (C-3'), 27.73 (C-3'''), 24.92 (C-1'') **MS** (ESI): m/z calculated for $C_{18}H_{25}N_3O_3$+K^+: 370.153 $[M+K]^+$; found: 370.3.

The ^{1}H and ^{13}C NMR data is in accordance with the data previously described in literature.[12]



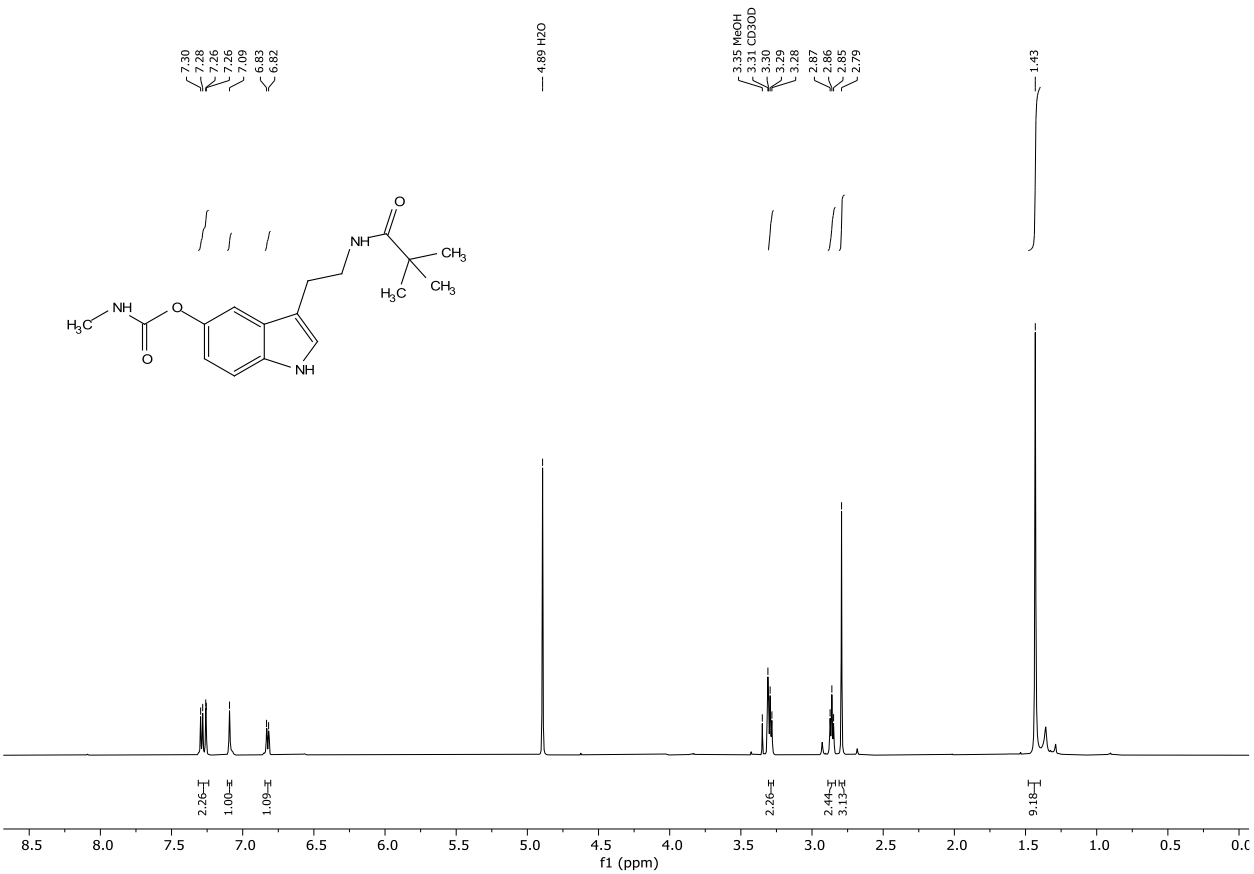
^{1}H NMR (CD_3OD, 600 MHz)

Figure S15. ^{1}H NMR spectrum of **5**.



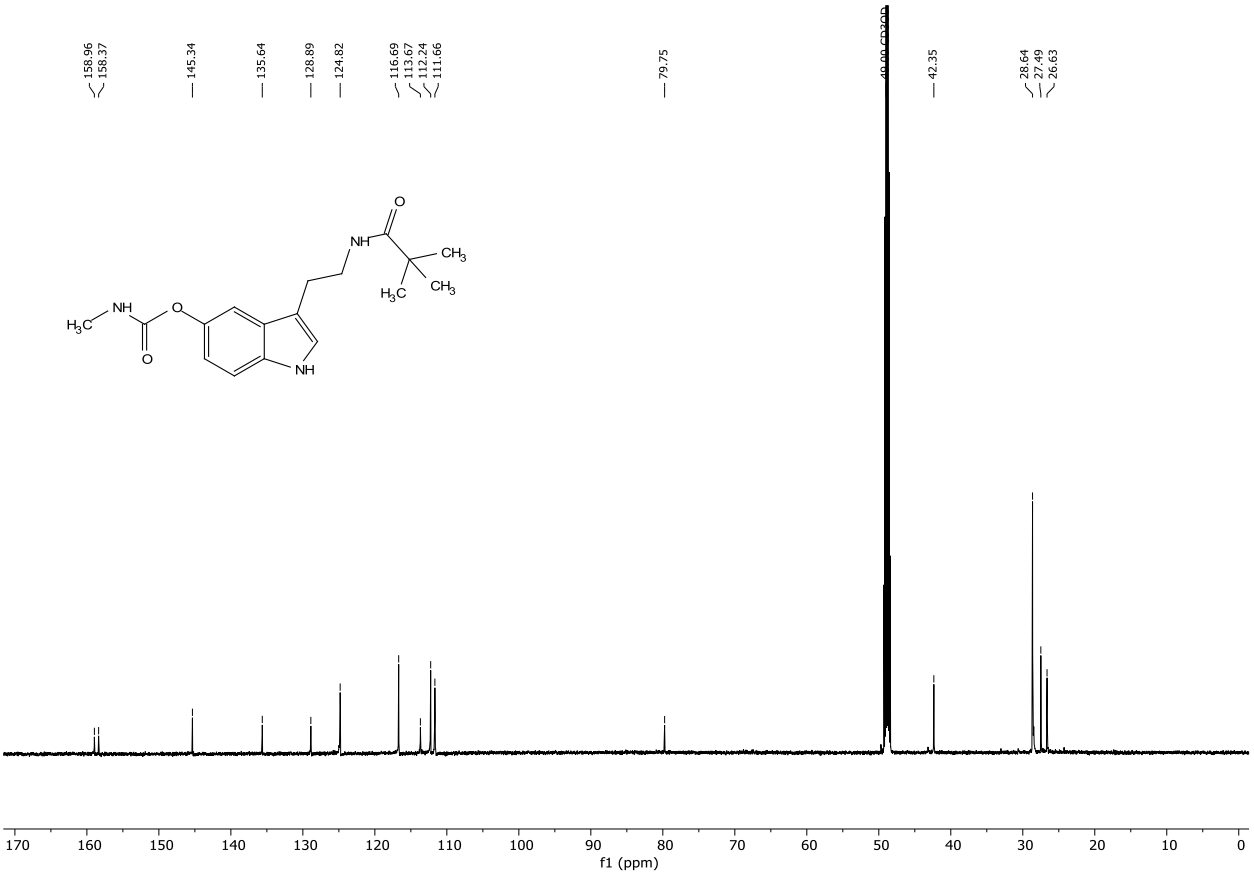
^{13}C NMR (CD_3OD, 151 MHz)

Figure S16. ^{13}C NMR spectrum of **5**.



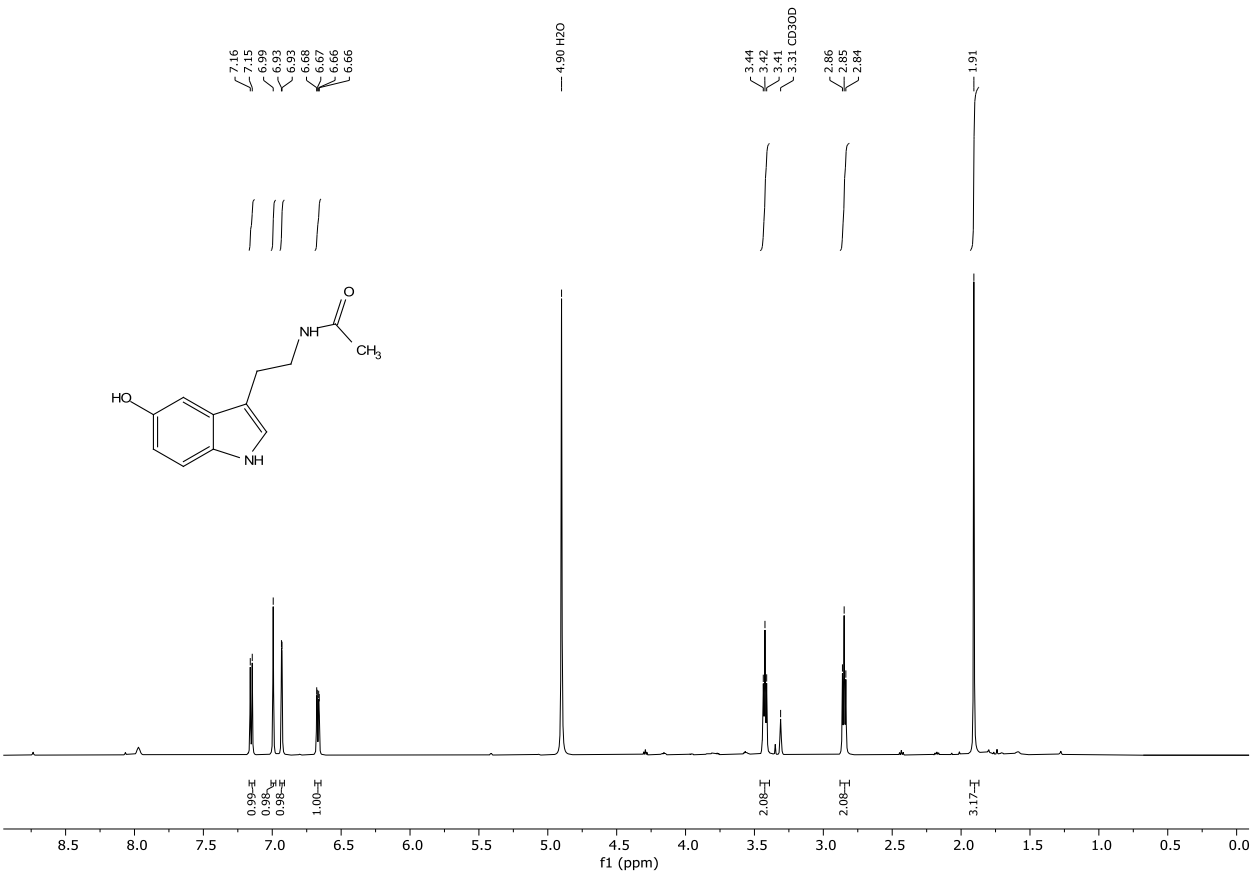
^{1}H NMR (CD_3OD, 600 MHz)

Figure S17. ^{1}H NMR spectrum of **2a**.



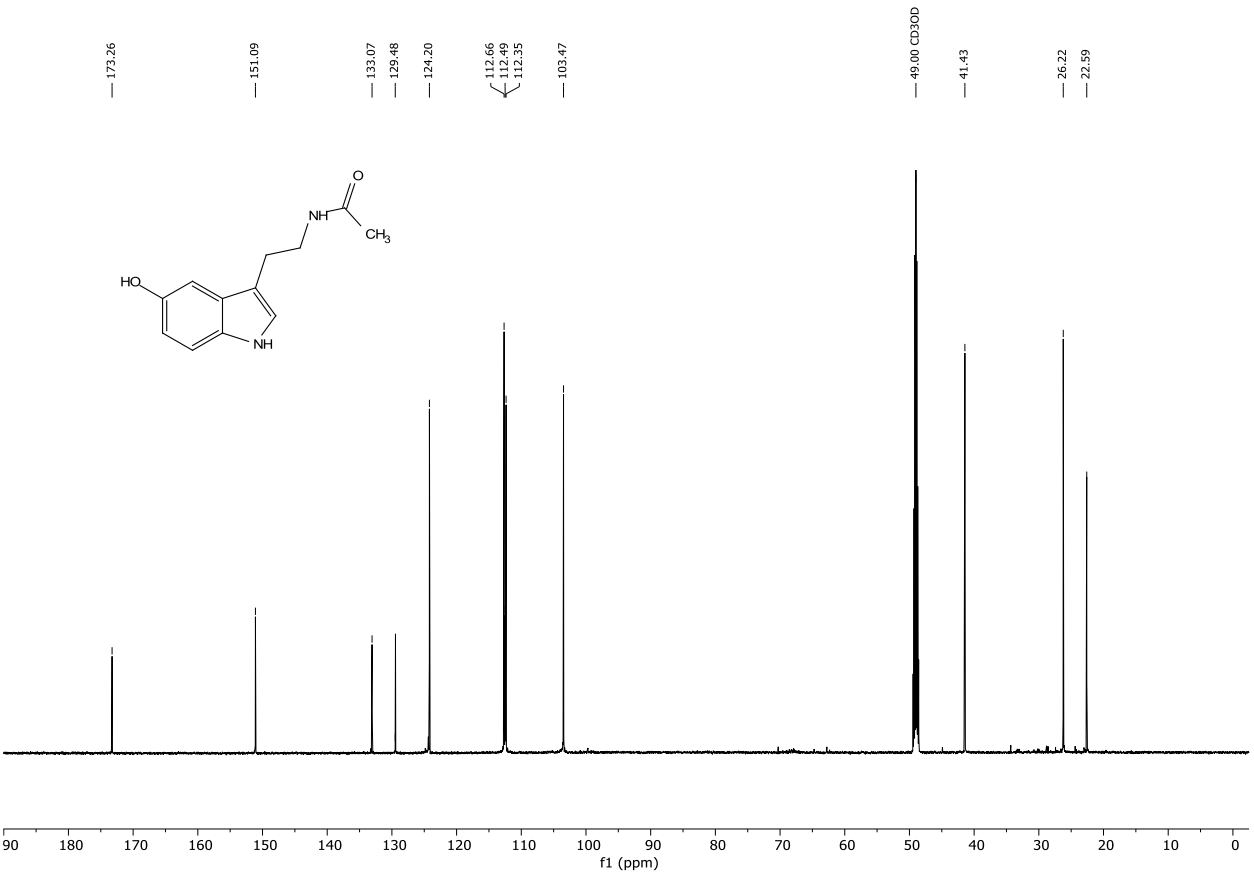
^{13}C NMR (CD_3OD, 151 MHz)

Figure S18. ^{13}C NMR spectrum of **2a**.



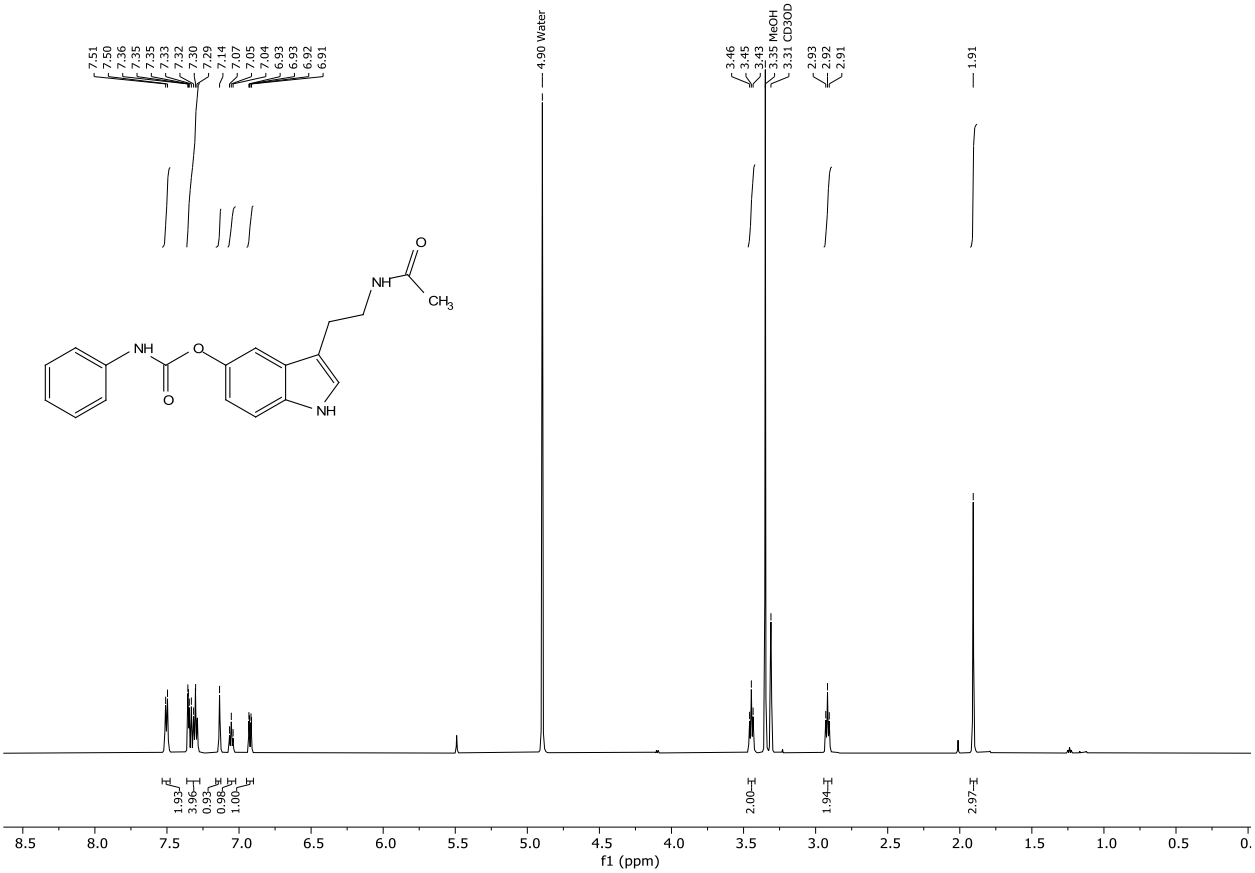
^{1}H NMR (CD_3OD, 600 MHz)

Figure S19. ^{1}H NMR spectrum of **8**.



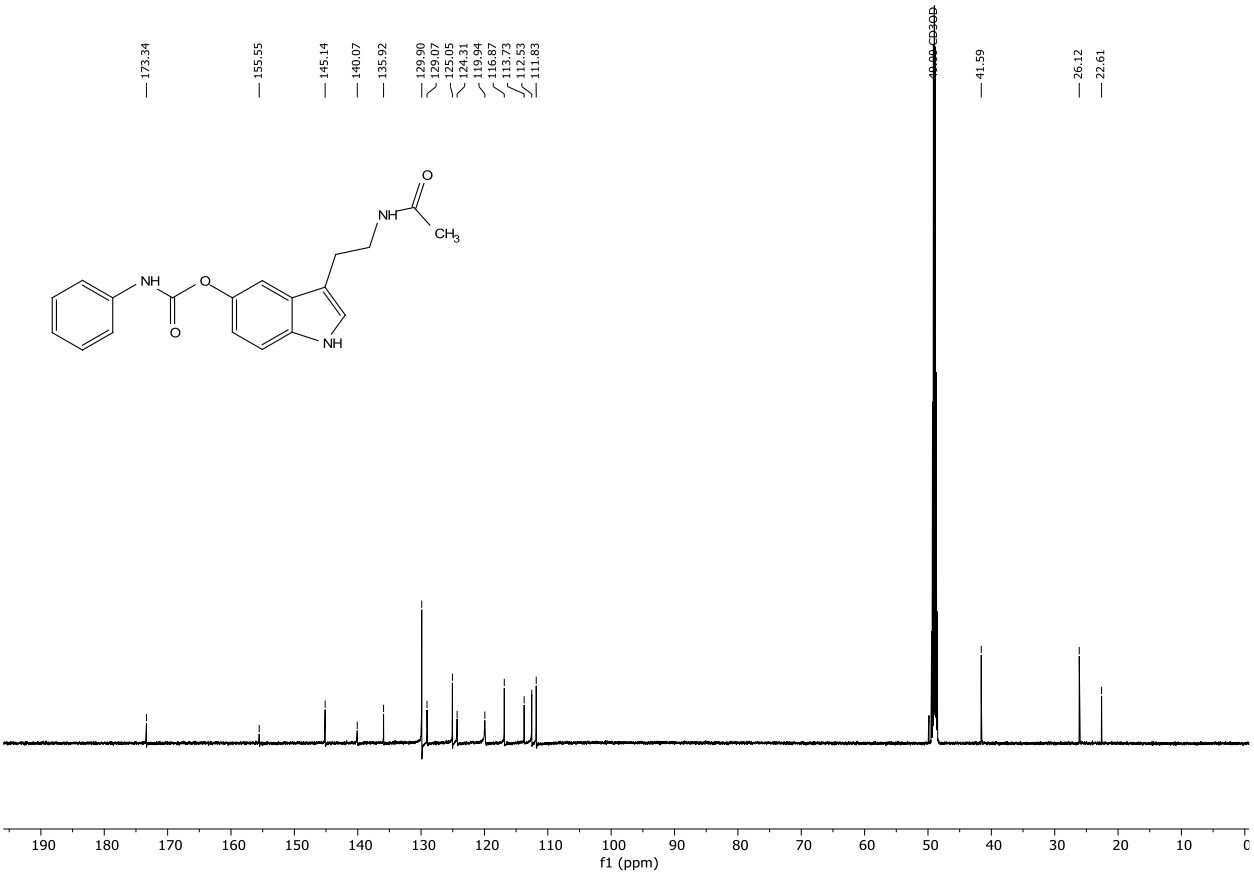
^{13}C NMR (CD_3OD, 151 MHz)

Figure S20. ^{13}C NMR of **8**.



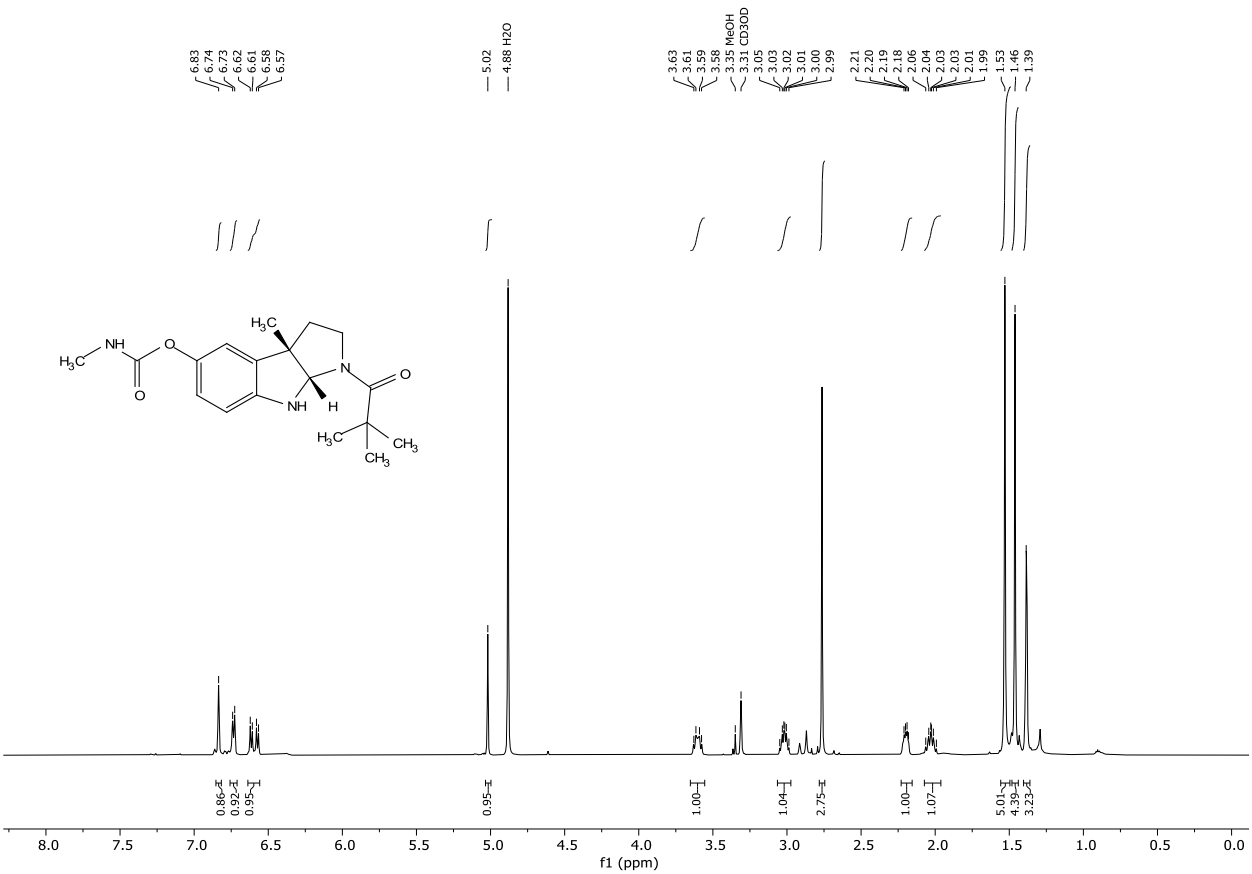
^{1}H NMR (CD_3OD, 600 MHz)

Figure S21. ^{1}H NMR spectrum of **3a**.



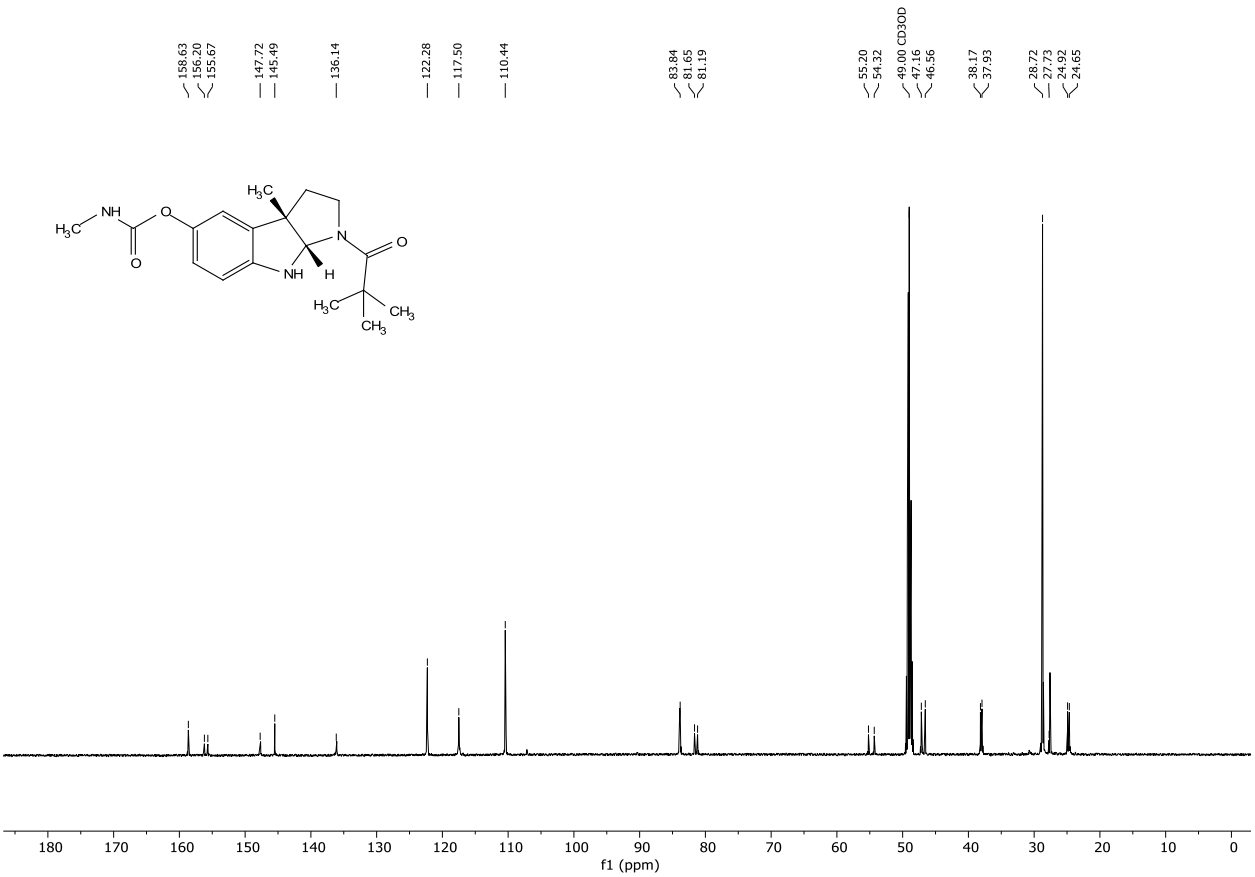
^{13}C NMR (CD_3OD, 151 MHz)

Figure S22. ^{13}C NMR spectrum of **3a**.



^{1}H NMR (CD_3OD, 600 MHz)

Figure S23. ^{1}H NMR spectrum of **2b**.



^{13}C NMR (CD_3OD, 151 MHz)

Figure S24. ^{13}C NMR spectrum of **2b**.

Python scripts

The activity assay was performed on an Opentrons® OT-2 robot, using the following scripts:

Script 1:

```
# number of complete columns

cols = 9

# takes 6 min for 10 cols


#-----------------

# number of samples

n = cols*8


### Start of actual script ###


from opentrons import protocol_api


# metadata

metadata = {

    'protocolName': 'Methyltransferase_part1',

    'author': 'Tobias Rosch <t.rosch@fz-juelich.de>',

    'description': 'This script is part of the methyltransferase assay
which includes handling of fuming sulfuric acid.',

    'apiLevel': '2.11'

}


# protocol run function. the part after the colon lets your editor know

# where to look for autocomplete suggestions
```

```
def run(protocol: protocol_api.ProtocolContext):


    # labware

    sample = protocol.load_labware('nest_96_wellplate_200ul_flat', '1')
# samples

    sulf = protocol.load_labware('nest_96_wellplate_200ul_flat', '2') #
plate with sulfuric acid

    tiprack1 = protocol.load_labware('opentrons_96_tiprack_300ul', '5')
# 300 uL tips for left pipette


    # pipettes

    multi300_pipette = protocol.load_instrument(

         'p300_multi_gen2', 'left', tip_racks=[tiprack1])


    # pipette 50 uL sample into plate with sulfuric acid

    s1 = [x.bottom(0.5) for x in sample.wells()[0:n]]

    d1 = [x.bottom(1) for x in sulf.wells()[0:n]]

    multi300_pipette.transfer(50, s1, d1, mix_before = (3, 50), new_tip
= 'always')
```

Script 2:

```
# number of complete columns

cols = 10

# takes 4 min for 10 cols



#-----------------

# number of samples
```

```
n = cols*8


### Start of actual script ###


from opentrons import protocol_api


# metadata

metadata = {

    'protocolName': 'Methyltransferase_part2',

    'author': 'Tobias Rosch <t.rosch@fz-juelich.de>',

    'description': 'This script is part of the methyltransferase assay
which includes handling of fuming sulfuric acid.',

    'apiLevel': '2.11'

}


# protocol run function. the part after the colon lets your editor know

# where to look for autocomplete suggestions

def run(protocol: protocol_api.ProtocolContext):


    # labware

    sulf = protocol.load_labware('nest_96_wellplate_200ul_flat', '2') #
plate with sulfuric acid

    dmab = protocol.load_labware('ibg_12well_reservoir', '4') # plate
with dmab

    tiprack1 = protocol.load_labware('opentrons_96_tiprack_300ul', '5')
# 300 µL tips for left pipette


    # pipettes
```

```
    multi_300_pipette = protocol.load_instrument(
         'p300_multi_gen2', 'left', tip_racks=[tiprack1])


    # Add 50 uL DMAB
    d3 = [x.bottom(1) for x in sulf.wells()[0:n]]
    multi_300_pipette.transfer(50,
dmab.wells_by_name()['A1'].bottom(2), d3, new_tip = 'always')
```

References

1. S. Kille, C. G. Acevedo-Rocha, L. P. Parra, Z.-G. Zhang, D. J. Opperman, M. T. Reetz and J. P. Acevedo, *ACS Synth. Biol.*, 2013, **2**, 83-92.
2. K. Hsiao, H. Zegzouti and S. A. Goueli, *Epigenomics*, 2016, **8**, 321-339.
3. B. Webb and A. Sali, *Curr. Protoc. Bioinformatics*, 2016, **54**, 5.6.1-5.6.37.
4. D. A. Amariei, M. Haase, M. K. T. Klischan, M. Wäscher and J. Pietruszka, *ChemCatChem*, 2024, e202400052.
5. A. C. S. Marc A. Martí-Renom, András Fiser, Roberto Sánchez, Francisco Melo, and Andrej Šali, *Annu. Rev. Biophys. Biomol. Struct.*, 2000, **29**, 291-325.
6. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605-1612.
7. Maxim V. Shapovalov and Roland L. Dunbrack, *Structure*, 2011, **19**, 844-858.
8. J. Eberhardt, D. Santos-Martins, A. F. Tillack and S. Forli, *J. Chem. Inf. Model.*, 2021, **61**, 3891-3898.
9. O. Trott and A. J. Olson, *J. Comput. Chem.*, 2009, **31**, 455-461.
10. H. Sticht, R. Quiroga and M. A. Villarreal, *Plos One*, 2016, **11**.
11. D. A. Amariei, N. Pozhydaieva, B. David, P. Schneider, T. Classen, H. Gohlke, O. H. Weiergräber and J. Pietruszka, *ACS Catal.*, 2022, **12**, 14130-14139.
12. P. Schneider, B. Henßen, B. Paschold, B. P. Chapple, M. Schatton, F. P. Seebeck, T. Classen and J. Pietruszka, *Angew. Chem. Int. Ed.*, 2021, **60**, 23412-23418.
13. M. Van Kempen, S. S. Kim, C. Tumescheit, M. Mirdita, J. Lee, C. L. M. Gilchrist, J. Söding and M. Steinegger, *Nat. Biotechnol.*, 2023, **42**, 243-246.
14. F. Madeira, M. Pearce, A. R. Tivey, P. Basutkar, J. Lee, O. Edbali, N. Madhusoodanan, A. Kolesnikov and R. Lopez, *Nucleic Acids Res.*, 2022, **50**, W276-W279.
15. A. M. Waterhouse, J. B. Procter, D. M. Martin, M. Clamp and G. J. Barton, *Bioinformatics*, 2009, **25**, 1189-1191.
16. B. Yariv, E. Yariv, A. Kessel, G. Masrati, A. B. Chorin, E. Martz, I. Mayrose, T. Pupko and N. Ben-Tal, *Protein Sci.*, 2023, **32**, e4582.
17. H. Ashkenazy, S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko and N. Ben-Tal, *Nucleic Acids Res.*, 2016, **44**, W344-W350.
18. G. Celniker, G. Nimrod, H. Ashkenazy, F. Glaser, E. Martz, I. Mayrose, T. Pupko and N. Ben-Tal, *Isr. J. Chem.*, 2013, **53**, 199-206.