Electronic Supplementary Material (ESI) for ChemComm.

Supporting information

Pyrroloindoline cyclization in tryptophan-containing cyclodipeptides mediated by an unprecedented indole C3 methyltransferase from *Streptomyces* sp. HPH0547

Hongxia Li,[†,‡] Yan Qiu,[†,‡] Canxiong Guo,[†,‡] Meng Han,[§] Yuyang Zhou,[†,‡] Yue Feng,[†,‡] Shizhong Luo,[†,‡] Yigang Tong,[†,‡] Guojun Zheng[†,‡,*] and Shaozhou Zhu[†,‡,*]

[†] State Key Laboratory of Chemical Resources Engineering, Beijing University of Chemical Technology, Beijing, People's Republic of China, 100029

[‡] College of Life Science and Technology, Beijing University of Chemical Technology, Beijing, People's Republic of China, 100029

[§] MOE Key Laboratory of Bioinformatics, School of Life Sciences, Tsinghua University, Beijing, People's Republic of China, 100084

Table of Contents

Experimental Procedures	**Page**
Gene synthesis and materials	S4
Plasmids, bacterial strains and culture conditions	S4
Generation of expression constructs for truncated stspM1 and full length	S4
Gene expression	S4
Structural modeling of stspM1	S5
In vitro assays	S5
LC-MS analysis	S6
Biochemical characterization of stspM1	S6
Isolation and NMR structural elucidation of stspM1 products	S7
Substrate specificity of stspM1	S8
Bioinformatics studies	S8
Supplementary Tables	
Table S1. Proposed functions of the ORFs in the putative methyltransferase containing CDPS biosynthetic gene clusters.	S9
Table S2. Cloning and expression constructs used in this study.	S9
Table S3. ^{1}H (600 MHz) and ^{13}C NMR (150 MHz) data of the isolated compound 2 in DMSO-*d*6	S10
Supplementary Figures	
Fig. S1. Organization of the CDPS gene clusters from *Streptomyces* sp. HPH0547.	S11
Fig. S2. Alignments of methyltransferase from this study and other methyltransferases from bacteria.	S11
Fig. S3. Size-exclusion chromatography of stspM1.	S12
Fig. S4. Purification and structural models of stspM1	S12
Fig. S5. HPLC analysis after in vitro assays with cWW as the substrate.	S13
Fig.S6. MS2 fragmentation spectrum of cWW (m/z 373).	S13
Fig.S7. MS2 fragmentation spectrum of compounds 2 (m/z 387).	S14
Fig.S8. Biochemical characterization of stspM1	S14
Fig. S9. Chromatograms of the preparative HPLC purification of compounds 2 after large scale in vitro assays.	S15
Fig. S10. Chromatograms of purified compounds 2 after preparative HPLC.	S15
Fig. S11. ^{1}H NMR spectrum of compound 2 (DMSO-*d*6, 600 MHz).	S15
Fig. S12. ^{13}C spectrum of compound 2 (DMSO-*d*6, 150 MHz).	S16
Fig. S13. ^{13}C dept135 spectrum of compound 2 (DMSO-*d*6, 150 MHz).	S16

Fig. S14. ^{1}H-^{1}H-COSY spectrum of compound 2 (DMSO-*d*6, 600 MHz, 600 MHz). S17
Fig.S15. HSQC spectrum of compound 2 (DMSO-*d*6, 600 MHz, 150MHz). S17
Fig. S16. HMBC spectrum of compound 2 (DMSO-*d*6, 600 MHz, 150MHz). S18
Fig. S17. NOESY spectrum of compound 2 (DMSO-*d*6, 600 MHz, 600 MHz). S18
Fig. S18. ROESY spectrum of compound 2 (DMSO-*d*6, 600 MHz, 600 MHz). S19
Fig. S19. HPLC analysis after in vitro assays with cWY as the substrate S19
Fig. S20. MS2 fragmentation spectrum of methylated cWY (m/z 364) S20
Fig. S21. Identified potential CDPS gene clusters with diverse tailoring enzymes. S20

Gene synthesis and materials

The gene for Truncated stspM1 (EPD89497) was optimized by GeneArt (Thermo Fisher Scientific, US) and synthesized by Synbio Technologies® (Suzhou). The optimized gene for truncated stspM1 was deposited in Genebank as MK584931. Meanwhile, the full length stspM1 gene was download from NCBI databases (http://www.ncbi.nlm.nih.gov, Genebank number: EPD89497 and EPD89498) and optimized for *Escherichia coli* heterologous expression by GeneArt and deposited in Genebank as MK573553. Both genes were cloned into pUC57 vector as templates for further cloning. Primers for cloning were ordered from Sangon Biotech® (Shanghai) as listed in Table S2. S-adenosyl-L-methionine (SAM) was obtained from Sigma Aldrich. Diverse cyclodipeptides were purchased from Spec-Chem Industry Inc. Q5 High-Fidelity DNA Polymerase and Gibson Assembly® Master Mix were bought from New England Biolabs (NEB). DNA manipulation and purification kits were bought from Tiangen (China). Kanamycin, ampicillin and other chemicals were obtained from commercial sources.

Plasmids, bacterial strains and culture conditions

pET30a plasmid were obtained from Invitrogen (USA). *E. coli* TOP10 and *E. coli* BL21(DE3) were obtained from Shanghai Weidi Biotechnology and used as a general host for gene cloning, plasmid propagation and protein expression. Recombinant *E. coli* strains with different plasmids were cultivated in liquid/solid LB (pH 7.0) suppling with corresponding antibiotic (100 μg/mL ampicillin or 50 μg/mL kanamycin). Solid agar LB plates were prepared by adding 1.5% (m/w) agar.

Generation of expression constructs for truncated stspM1 and full length stspM1

To clone the stspM1 gene, the truncated and the full length stspM1 gene were amplified from the synthesized genes using primers list in Table S2. pET30a was also amplified and the linear plasmids were purified. The resultant amplicons were then directly ligated into the linear pET30a by Gibson Assembling according to the manufacturer's protocols. The resultant plasmids harboring stspM1 were then introduced in *E. coli* TOP10 by electroporation and transformants were inoculated on LB agar plates containing kanamycin (50 μg/mL). After sequencing, the right transformants were verified and used to prepare the plasmids for further expression.

Gene expression

The resulted constructs having His6-tag at the N terminus were then retransformed into *E. coli* BL21 (DE3) cells by heat shock at 42 °C. Single colony was picked from agar plate, inoculated on

LB medium (50 mL) supplemented with kanamycin (50 μg/mL) and kept at 37 °C overnight. 5 L of liquid culture containing 50 μg/mL kanamycin were then inoculated with the overnight pre-culture and shaken at 37 °C until the OD reached to 0.7-1.0. The flasks were then cooled on ice for 0.5 h to slow down the growth of *E. coli*. The production of recombinant proteins was then initiated with the addition of 0.2 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The cells were growth for overnight at 16 °C and harvested by centrifugation (4000 rpm, 15 min, 4 °C). The obtained pellets were re-suspended in 50 mL lysis buffer (50 mM Tris, 300 mM NaCl, 10 mM imidazole, pH 8.0) and the cells were lysed by sonication on ice. Insoluble parts of the lysate was then removed by centrifugation (10000 rpm, 1 h, 4 °C). The supernatants containing the recombinant proteins were then filtered and loaded on a Ni–NTA gravity flow column at 4 °C. The unbound proteins were removed by washing buffer (50 mM Tris, 300 mM NaCl, 30 mM imidazole, pH 8.0) and recombinant stspM1 were then eluted from the Ni–NTA column by elution buffer containing high concentration of imidazole (50 mM Tris, 300 mM NaCl, 300 mM imidazole, pH 8.0). The roughly purified proteins were then concentrated to 5 mL and buffer exchanged into Buffer S (10 mM Tris, 200 mM NaCl, 5 mM DTT, pH 7.8) using Amicon Ultra filters. The concentrated proteins were further purified by a Superdex 200 column with the same Buffer S. After purified by the size exclusion chromatography, StspM1-containing fractions were selected, concentrated by using Amicon Ultra filters and directly flash-frozen using liquid N2. The concentration of the purified proteins were calculated by using a Thermo NanoDrop 2000. The approximate protein yields were about 27.5 mg/L.

Structural modeling of stspM1

Protein Homology/analogY Recognition Engine V 2.0 (Phyre2) server (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) was selected to build the homology modeling structure of stspM1. Homology modeling studies showed that an S-adenosylmethionine-dependent methyltransferase (PDB: 1Y8C) from *Clostridium acetobutylicum* ATCC 824 was the most suitable template structure for modeling. The top ranking model has 100.0% confidence based on the structure of 1Y8C.

In vitro assays

Chemically synthesized cyclic peptides cWW (cyclo-L-Trp-L-Trp) were redissolved in reaction buffer (pH 7.5, 50 mM Tris, 100 mM NaCl, 1mM SAM, 1 mM cWW) for in vitro assays.

Assays were carried out in 100 μL reaction buffer containing 40 μM stspM1. The reaction mixtures were incubated at 30 °C for 120 min and quenched by the addition of 2 μL of TFA. The reaction solution was then extracted with 200 μL of ethyl acetate. After removing precipitate by centrifuging at 12,000 × g for 1 min, the extracts were then analyzed by analytical HPLC.

10 μL of the sample was then subjected to HPLC (controler; Shimadzu CMB-20A, diode array detector; Shimadzu SPD-20A, pumps; Shimadzu LC-20A) with a reverse-phase HPLC column (SilGreen, particle size 5 μm, 12 nm, 4.6 × 250 mm) operating at 0.8 mL/min. The following gradient was used: solvent A; water containing 0.1% TFA, solvent B; acetonitrile containing 0.1% TFA, 10% B (2min) to 50% B (in 13 min) to 95% B (in 10 min), 95%B for 5 min and down to 10% (in 10 min).

LC-MS analysis

Mass spectrometric analysis of the *in vitro* studies were conducted by LC-MS. For the structural analysis of the cWW and reaction products, collision-induced dissociation fragmentation was carried out using online LC-MS. To reveal the methylation site, a Thermo Scientific Q Exactive mass spectrometer connected with a Thermo-Dionex (Ultimate 3000) HPLC system were employed to further MS/MS analysis. The substrates and the products were separated with a home-made fused silica capillary column (75 μm ID, 150 mm length; Upchurch, Oak Harbor, WA) packed with C-18 resin (300 Å, 5 μm, Varian, Lexington, MA). The following gradient was used: solvent A; 20% water (0.1% formic acid) and 80% solvent B; acetonitrile (0.1% formic acid) at flow rate of 0.3 ml/min.

Xcalibur 2.1.2 software were used to control the Q Exactive mass spectrometer. A single full-scan MS in the orbitrap (100-1800 m/z, 70,000 resolution) followed by 20 data-dependent MS/MS scans at 27% normalized collision energy (HCD) were performed. Target cWW and enzymatic products were identified and analyzed.

Biochemical characterization of stspM1

For time-dependent conversion of cWW mediated by stspM1, the typical in vitro assays (100 μL) in reaction buffer (pH 8.0, 50 mM Tris, 100 mM NaCl, 1 mM SAM, 1 mM cWW) were used. 40 μM stspM1 was added to initiate the reaction. 1 mM cWW was monitored for varying durations (5, 10, 15, 20, 25, 30, 40, 50, 60, 80, 100, 120 min). Reactions were stopped at each point by addition of 2 μL TFA. The reaction mixtures were then extracted with 200 μL of ethyl acetate, centrifuged

(12,000 × g, 10 min), and resultant extracts were then analyzed by analytical HPLC. All the experiments were independently conducted in triplicates. The obtained data was mentioned in mean SD.

The effects of temperature on stspM1 were investigated using cWW as substrates. Series of reactions were carried out at different temperature ranging from 20 to 60 °C with 5 °C increments. 40 μM stspM1 was added in 100 μL reaction buffer (pH 7.5, 50 mM Tris, 100 mM NaCl, 1 mM SAM, 1 mM cWW) and shaken for 30 min. The reactions were quenched by addition of 2 μL TFA and then analyzed by HPLC. The enzymatic activities were compared and the highest activity was set to 100%.

The effects of pH on stspM1were probed with reaction buffer (50 mM Tris, 100 mM NaCl, 1 mM SAM, 1 mM cWW, pH was adjusted from 4.0 to 10.0 by NaOH or HCl). Each reaction mixture having different pH was performed at 45 °C for 20 min and quenched by addition of 2 μL TFA.

The kinetic constants of stspM1 were measured by keeping the stspM1 and SAM concentrations constant. 13.64 μM stspM1 was added in 100 μL reaction buffer with varying concentrations (0.025 mM, 0.05 mM, 0.1 mM, 0.15 mM, 0.2 mM, 0.5 mM, and 1 mM of cWW). The linear conversion range of stspM1 was first investigated and 3 min was selected as a proper reaction time for further studies. All the reactions were performed at 45 °C for 3 min. Each reaction was initiated by the addition of stspM1 and stopped by adding TCA (final concentration of 1% (v/v)) after 3 min. The reaction rate was then determined by analytical HPLC. The velocity of reaction was calculated from three independent experiments and the error bar was added for data analysis.

Isolation and NMR structural elucidation of stspM1 product

For isolation of the stspM1 enzymatic products, the in vitro assays were scaled up to 250 mL containing 5% DMSO. A typical scaled-up assay reaction contained 1 mM cWW (93 mg) and 28 μM of recombinant stspM1. The reaction was carried out at 30 °C with constant shaking (180 rpm) for 48 h. The reaction was then evaporated to 25 mL. The extracts were further purified by preparative HPLC (controler; Shimadzu CMB-20A, diode array detector; Shimadzu SPD-20A, pumps; Shimadzu LC-6AD) with a preparative reverse-phase HPLC column (SunFire C18 OBD Prep Column, 100Å, 5 μm, 19 mm × 250 mm, 1/pkg) operating at 6 mL/min. The following gradient was used: solvent A; water containing 0.1% TFA, solvent B; acetonitrile containing 0.1% TFA, 10% B (2min) to 50% B (in 13 min) to 95% B (in 15 min), 95% B for 5 min and down to 10% (in 5 min).

Finally, 25 mg compounds 2 was obtained as white powder and subjected to NMR analysis.

NMR structural studies on compound 2 were performed on a Bruker Avance 600 spectrometer. All the NMR data and spectra of the product 2 were provided in Table S3 and Figures S12-S18, respectively.

Substrate specificity

Five different cyclodipeptides were synthesized and used to probe the substrate specificity of stspM1. The standard in vitro assays (pH 7.5, 50 mM Tris, 100 mM NaCl, 1 mM SAM, 1 mM substrates and 40 μM stspM1) were performed except the substrate used in each experiments (Table. 1). Reactions were conducted at 30 °C for 4 h and quenched by addition of 2 μL TFA. The reaction solution was then extracted with 200 μL of ethyl acetate and analyzed by HPLC. The conversion was calculated from three independent experiments and the error bar was added for data analysis.

Bioinformatics studies

Nucleotide and protein sequences were downloaded from NCBI databases (http://www.ncbi.nlm.nih.gov) and analyzed by Clone Manager. Genome mining studies were performed by using the web-based bioinformatic tools PSI-BLAST. The StspCDPS gene (EPD89497) was used as a query sequence. The algorithm parameters were set as follows: Maximum target sequences to display was set to 100; expect value threshold was set to 10; BLOSUM62 was chosen as the matrix for scoring parameters. The newly identified CDPS gene clusters were analyzed manually and the function of each identified tailoring enzymes were predicted based on the conserved domains obtained from BLAST searches. Multiple sequence alignments were performed by using the program ClustalW and visualized with ESPript 3.0 (http://endscript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi) to identify strictly conserved amino acid residues (Fig. S2).

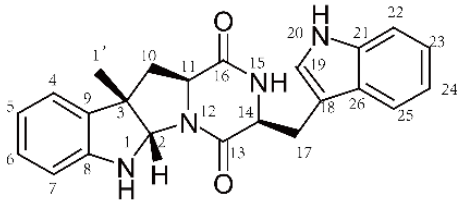
Table S1. Proposed functions of the ORFs in the putative methyltransferase containing CDPS biosynthetic gene clusters.

Gene	length(aa)	GenBank accession no.	proposed function	conserved domain analysis
StspCDPS	248	EPD89497.1	cWW cyclodipeptide synthase	tRNA-dependent cyclodipeptide synthase
StalCDPS	-	-		
5917CDPS	-	-		
StflCDPS	-	-		
StspM1	205	EPD89498.1	Methylation and cyclization	Methyltransferase domain;
StalM1	205	WP_078845769.1		
5917CDPS-M	498	WP_030408990.1		
StflCDPS-M	498	WP_055497289.1		
StspPS	333	EPD89499.1	Penylation and cyclization	Squalene/phytoene synthase
StalPS	333	WP_051803778.1		
5917PS	333	WP_051698706.1		
StflPS	333	WP_055497265.1		
StspM2	278	EPD89500.1	N-Methylation	Methyltransferase domain
StalM2	278	WP_051803776.1		
5917M2	278	WP_016473259.1		
StflM2	278	WP_055497264.1		

Table S2. Cloning and expression constructs used in this study.

construct	name	sequence
pET30a-Truncated stspM1	TstspM1 FP	CTTTAAGAAGGAGATATACATATG CTG GGT GAA CTG GCA CGT CGT GG
	TstspM1 RP	AGTGGTGGTGGTGGTGGTGCTCGAGGCT GCT GCT ACG AAC CAT GGT CCA GG
	pET41-FP	CTCGAGCACCACCACCACCACCACTGAGAT
	pET41-RP	ATG TAT ATC TCC TTC TTA AAG TTA AAC AAA ATT ATT TCT AGA GGG G
pET30a-stspM1	StspM1 FP	CTTTAAGAAGGAGATATACATATG AGT TCG GAA ACA GCC ACA CCC GCC
	StspM1 RP	AGTGGTGGTGGTGGTGGTGCTCGAGCGA GCT GCT CCG CAC CAT GGT CC
	pET41-FP	CTCGAGCACCACCACCACCACCACTGAGAT
	pET41-RP	ATG TAT ATC TCC TTC TTA AAG TTA AAC AAA ATT ATT TCT AGA GGG G

Table S3. ^{1}H (600 MHz) and ^{13}C NMR (150 MHz) data of the isolated compound **2** in DMSO-*d*6

compound			
Pos.	δ_H, multi., J[Hz]	δ_C	Group
1	2.54, S		NH
1’	1.39, S	22.8	CH3
2	5.26, S	82.8	CH
3		50.7	C
4	7.33, M, 7.5	113.3	CH
5	7.07, M, 7.6	120.9	CH
6	6.5, D, 7.8	108.9	CH
7	6.98, M, 7.5	127.8	CH
8		148.1	C
9		133.9	C
10a	1.91, Dd, 12.4, 10.7	40.5	CH2
10b	2.21, Dd, 12.4, 7.0	40.5	CH2
11	4.55, Dd, 10.6, 7.1	57.8	CH
12			N
13		167.6	C=O
14	4.38， Dd, 6.7, 4.4	55.1	CH
15	7.67		NH
16		169.8	C=O
17a	3.01, Dd, 15.0, 6.6	24.5	CH2
17b	3.38, Dd, 15.0, 4.5	24.5	CH2
18		109.5	C
19	7.24, D, 2.0	124.2	CH
20	10.85		NH
21		136.0	C
22	6.60， T, 7.3	117.5	CH
23	7.07, M, 7.6	121.9	CH
24	6.98, M, 7.5	118.3	CH
25	7.58, D, 7.9	118.4	CH
26		127.2	C

Truncated stspM1 :

MLGELARRGYAVTGLDRSAAMLERARRRLGEETTLIHAALPHIPAEAGPFDAVVSAAGGL
NYLPEEQISATFAAVARALPAGGTFTFDVFGRGFFRKFFDSSAPRVMALELDDIAYIWTF
TASPEAPFVDMAYTQFTPAPAADGGEPPFLRTRDLHRYYPLPHTTVRRLAAEHGFTDTKA
YDNYSTDPSGPDSLYDTWTMVRSSS

Artificial full length stspM1 :

MSSETATPADPYTNLADSYDRLAEWAVTCQKESPRDRVADFLQTFWQSQQRPVRTVLEIC
CGTGLMLGELARRGYAVTGLDRSAAMLERARRRLGEETTLIHAALPHIPAEAGPFDAVVS
AAGGLNYLPEEQISATFAAVARALPAGGTFTFDVFGRGFFRKFFDSSAPRVMALELDDIA
YIWTFTASPEAPFVDMAYTQFTPAPAADGGEPPFLRTRDLHRYYPLPHTTVRRLAAEHGF
TDTKAYDNYSTDPSGPDSLYDTWTMVRSSS

Figure S1. Organization of the CDPS gene clusters from *Streptomyces* sp. HPH0547. Sequence of the optimized stspM1 from *Streptomyces* sp. HPH0547. The sequence of truncated and full length stspM1 have been submitted to GeneBank as an accession number: MK584931 and MK573553

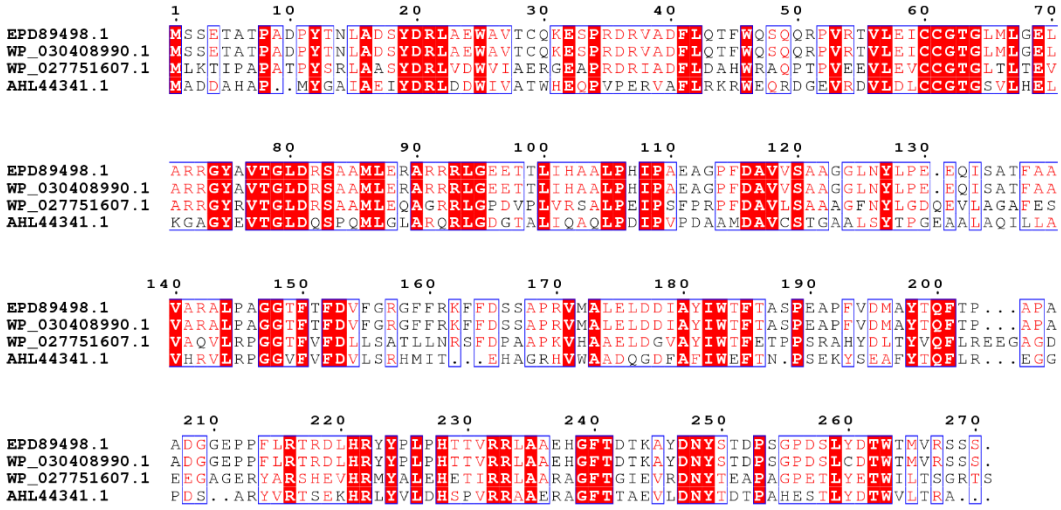


Figure S2. Alignments of methyltransferase from this study and other methyltransferases from bacteria. WP_030408990.1 is the class I SAM-dependent methyltransferase from *Streptomyces* sp. NRRL F-5917. WP_027751607.1 is class I SAM-dependent methyltransferase from *Streptomyces* sp. CNH287. AHL44341.1 is the C3 methyltransferase from *Streptomyces griseofuscus*

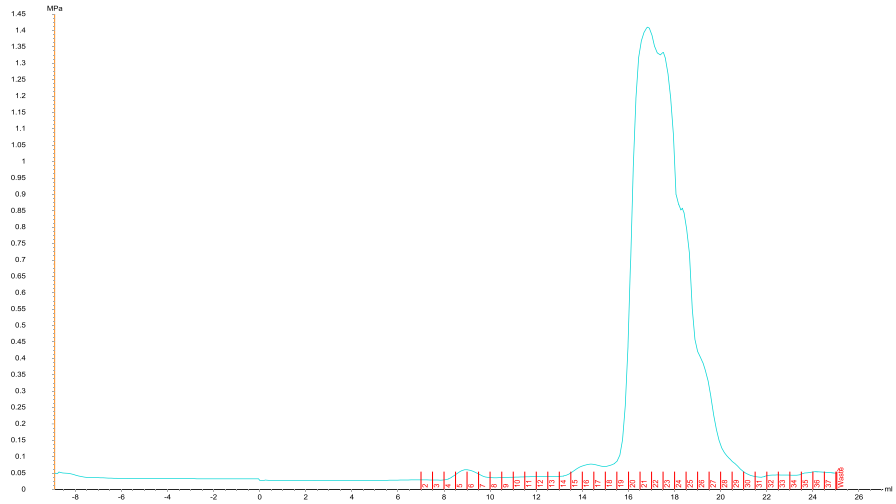


Figure S3. Size-exclusion chromatography of stspM. The elution fractions was further analyzed by SDS-PAGE.

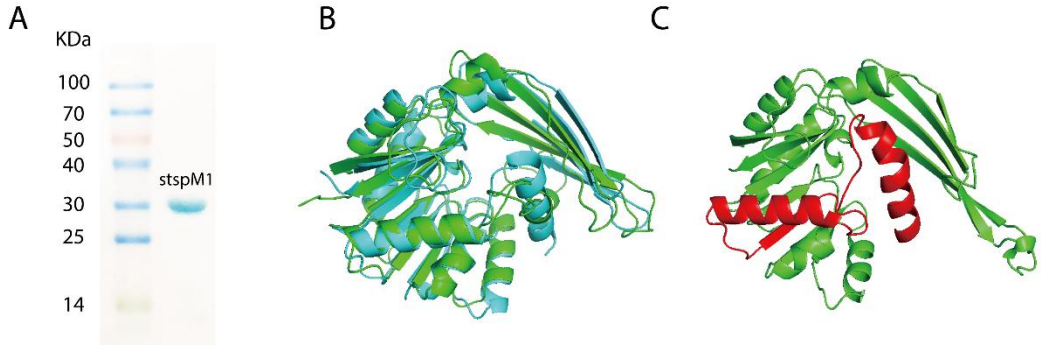


Figure S4. Purification and structural models of stspM1. (A) Analysis of the purified His6-StspM1 using SDS-PAGE. Lane 1, Marker; lane 2, purified stspM1 after two step purification. (B) Homology model of stspM1 (sky blue) overlaid with the studied structure (1Y8C) of the SAM methyltransferase from Clostridium acetobutylicum ATCC 824. (C) The N terminus SAM binding site is shown in red.

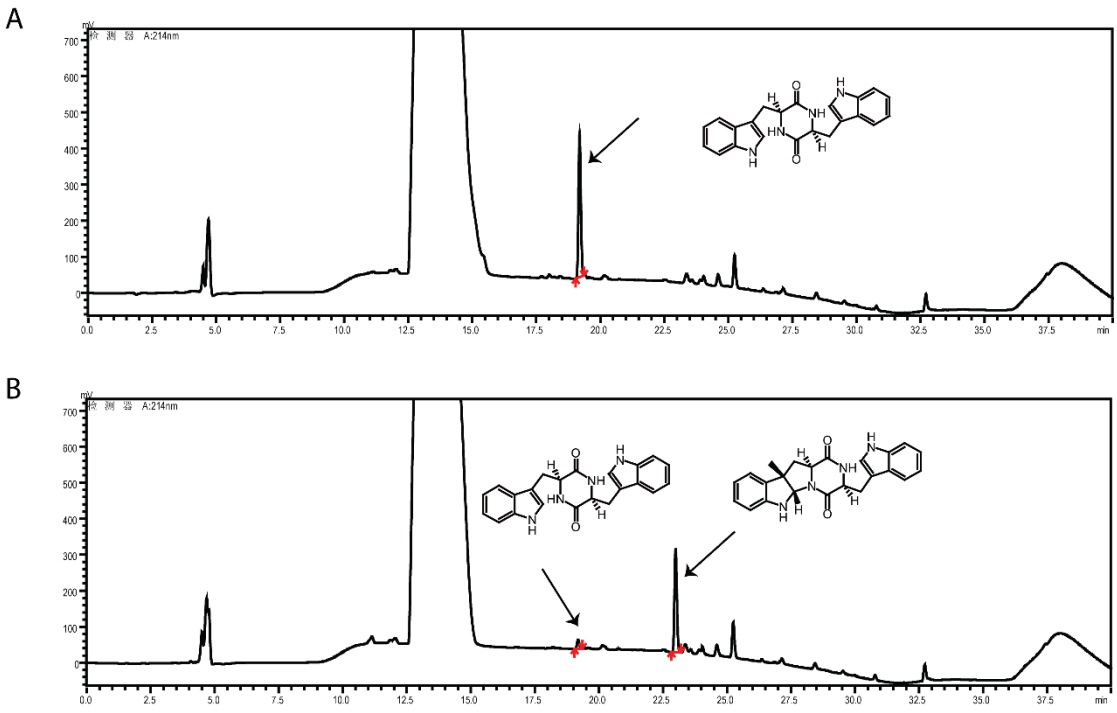


Figure S5. HPLC analysis after in vitro assays with cWW as the substrate. A, Blank; B, in vitro assays with StspM1

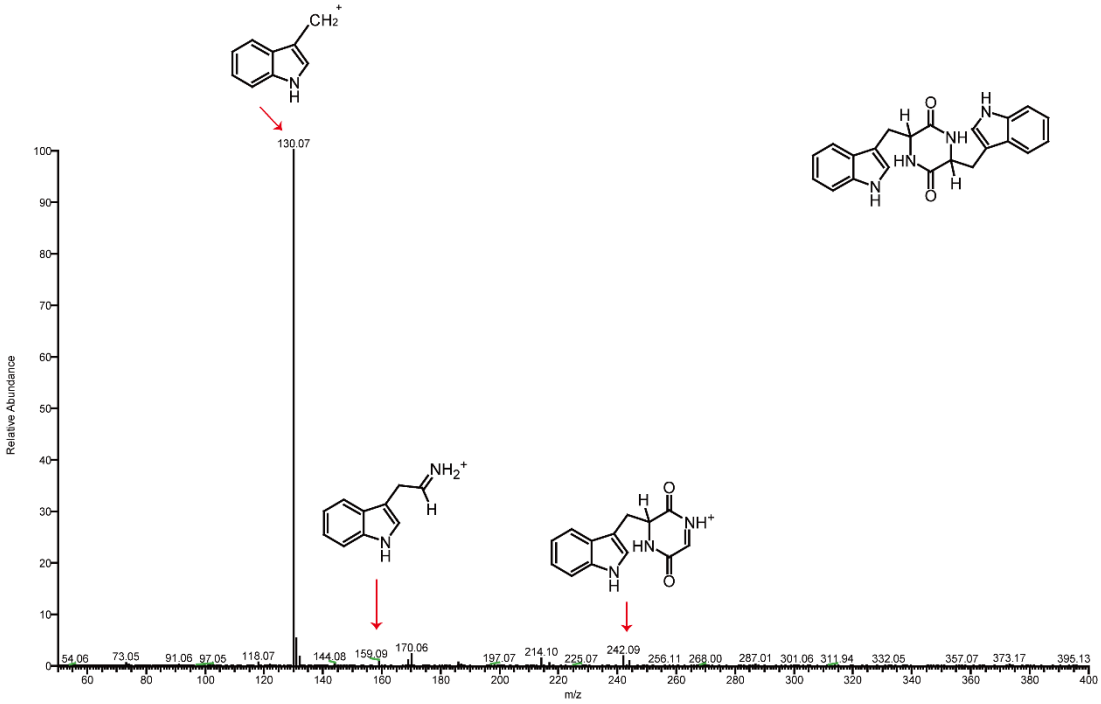


Figure S6. MS2 fragmentation spectrum of cWW (m/z 373). Relevant fragment ions and neutral losses used to identify the substrate are shown.

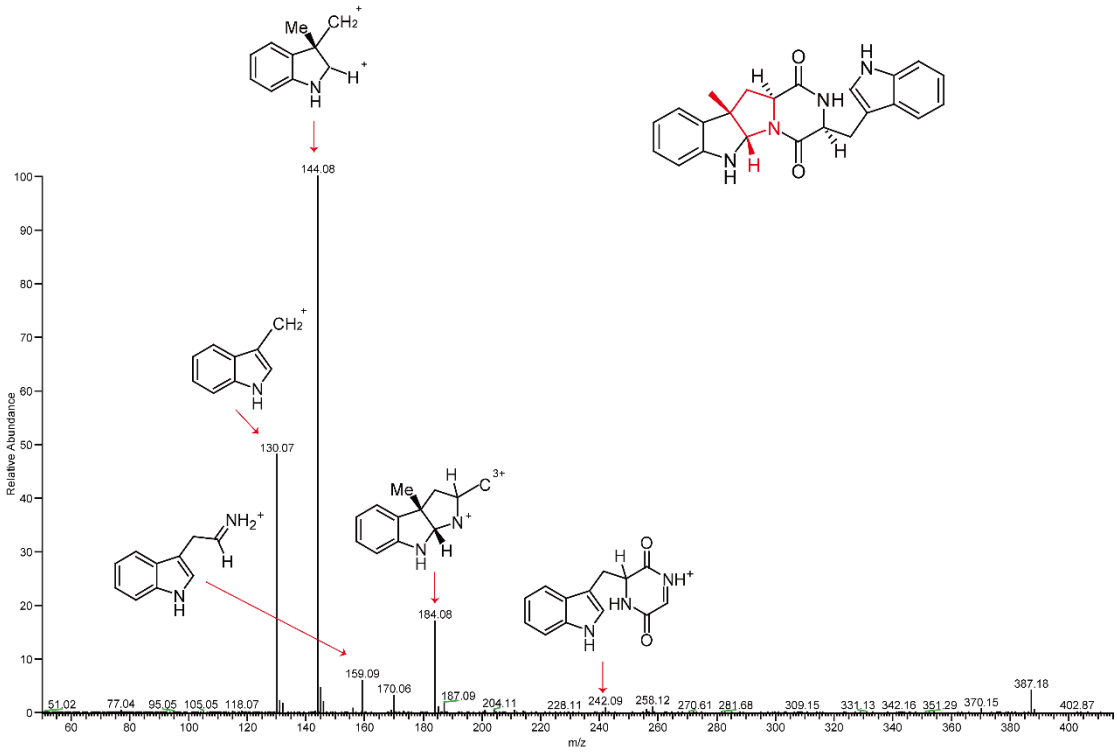


Figure S7. MS2 fragmentation spectrum of compounds 2 (m/z 387). Relevant fragment ions and neutral losses used to identify the in vitro reaction product are shown.

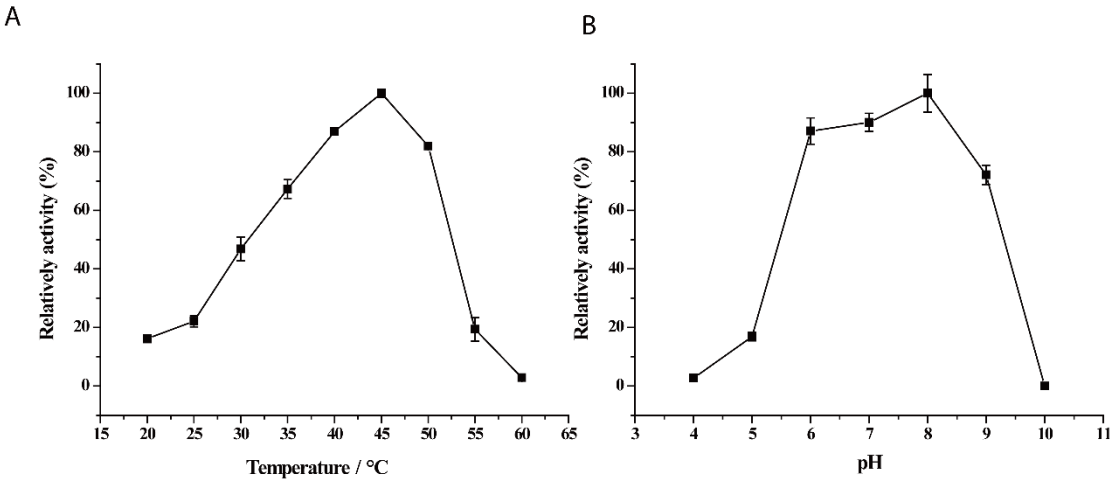


Figure S8. Biochemical characterization of stspM1. (A). Effects of the thermoactivity of stspM1 from *Streptomyces* sp. HPH0547. The activity of the enzyme at 45 °C was taken as 100%. (B). Effects of pH on the activity of the purified stspM1. The activity of the enzyme at pH 8.0 was taken as 100%. Each point represents the mean of three independent experiments

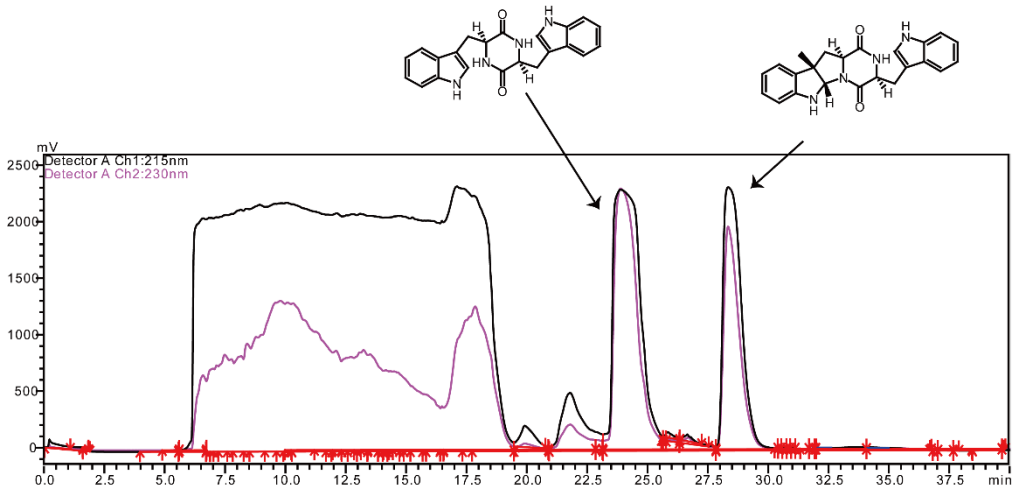


Figure S9. Chromatograms of the preparative HPLC purification of compounds 2 after large scale in vitro assays.

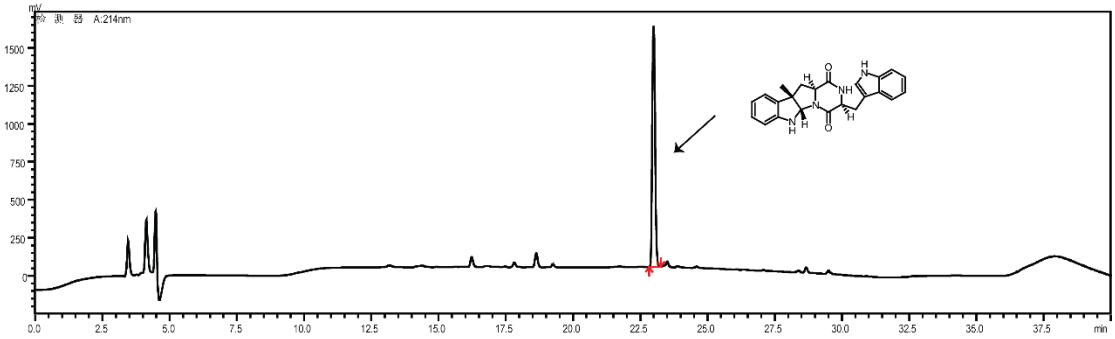


Figure S10. Chromatograms of purified compounds 2 after preparative HPLC.

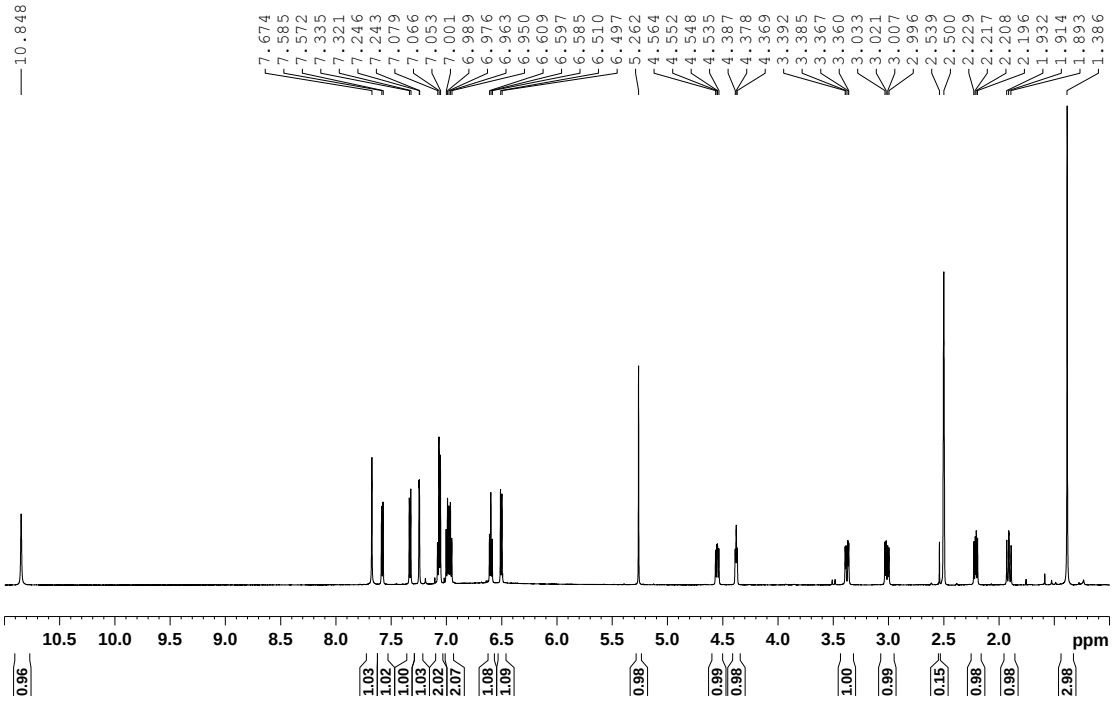


Figure S11. [1] H NMR spectrum of compound 2 (DMSO-*d*6, 600 MHz).

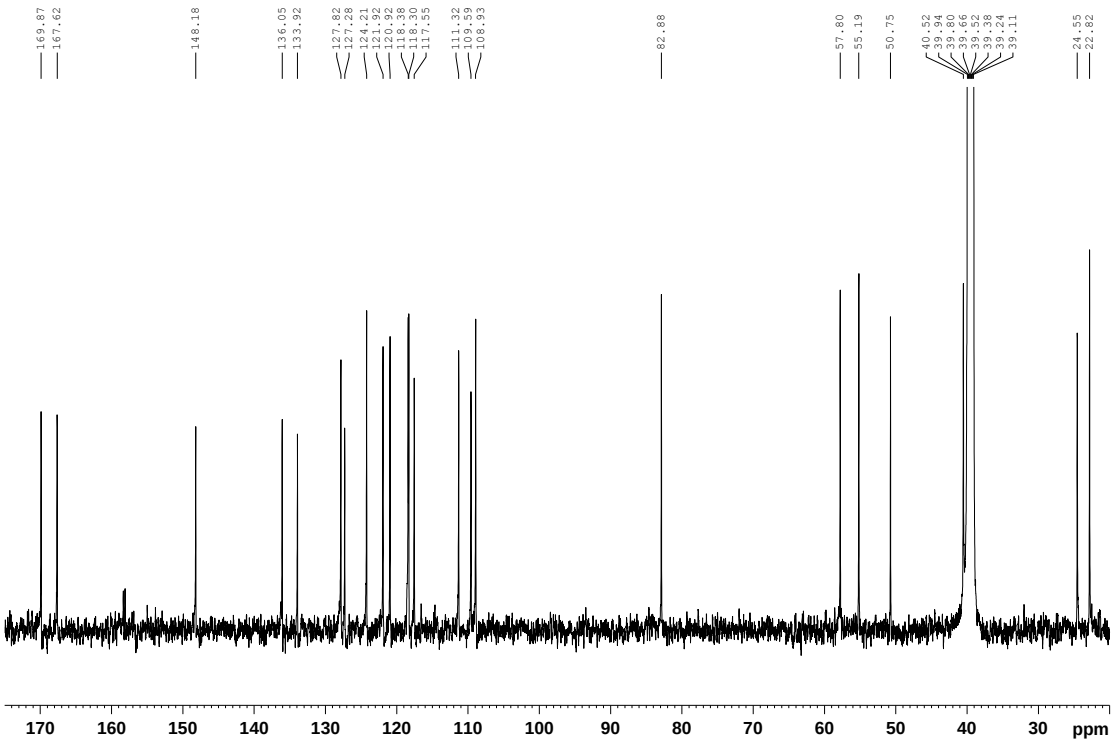


Figure S12. ^{13}C spectrum of compound 2 (DMSO-*d*6, 150 MHz).

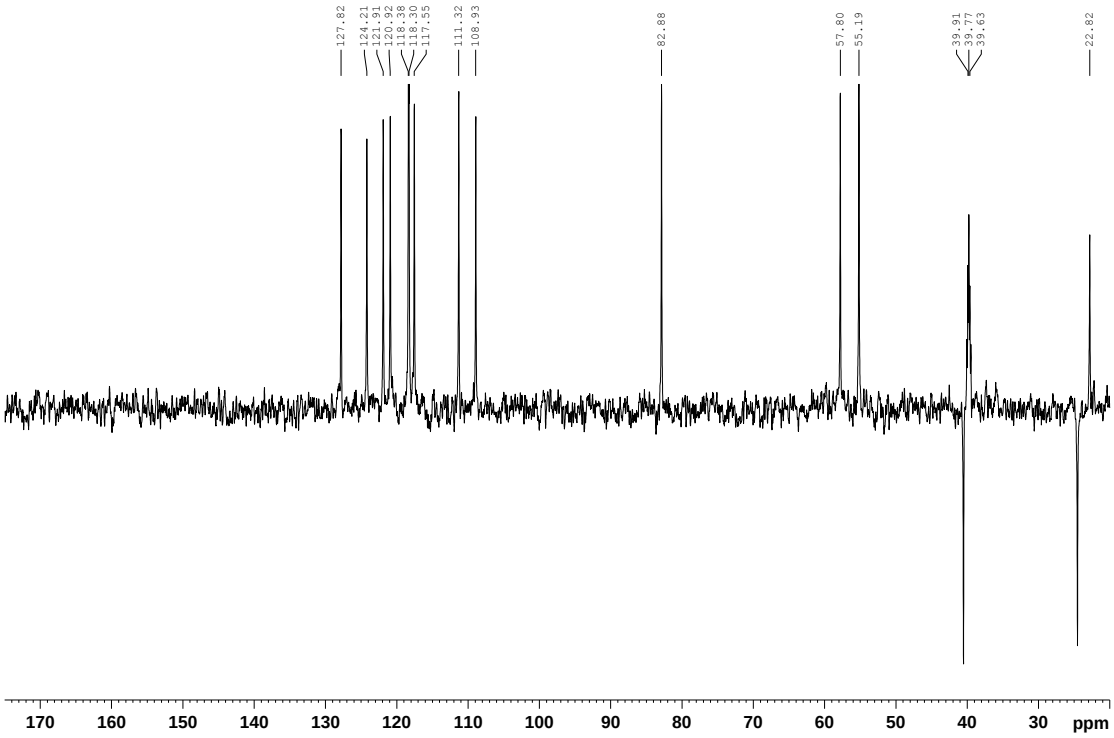


Figure S13.^{13}C dept135 spectrum of compound 2 (DMSO-*d*6, 150 MHz).

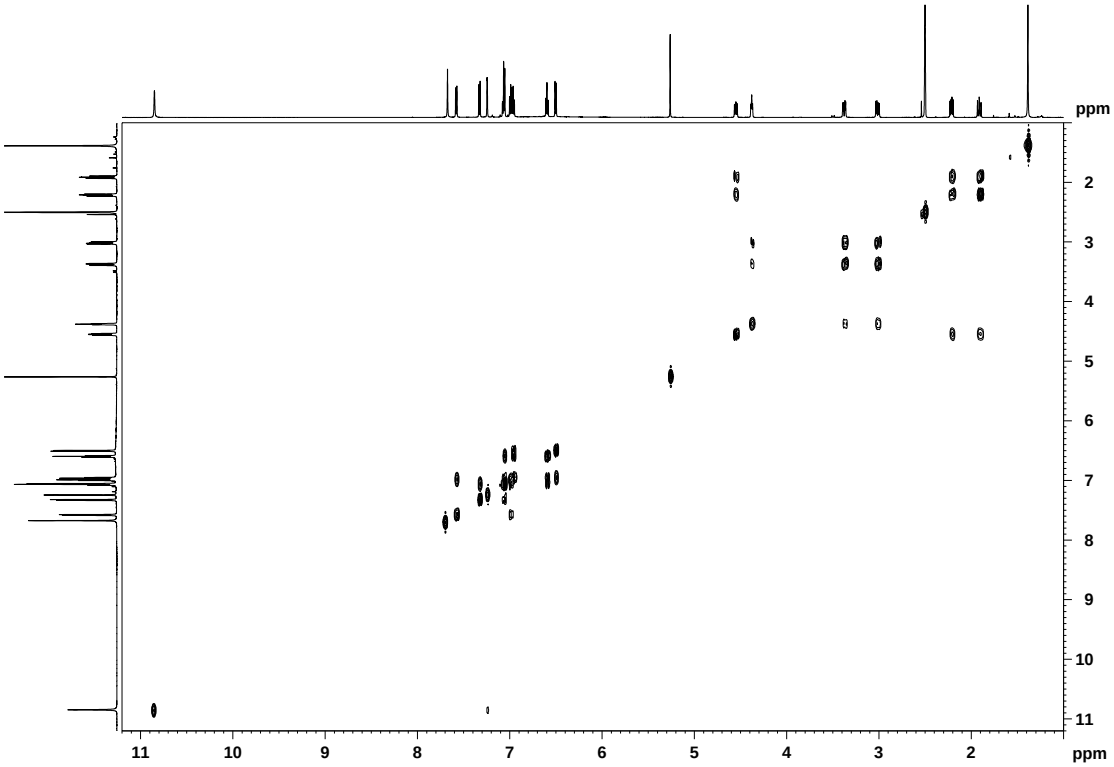


Figure S14. ^{1}H-^{1}H-COSY spectrum of compound 2 (DMSO-*d*6, 600 MHz, 600 MHz).

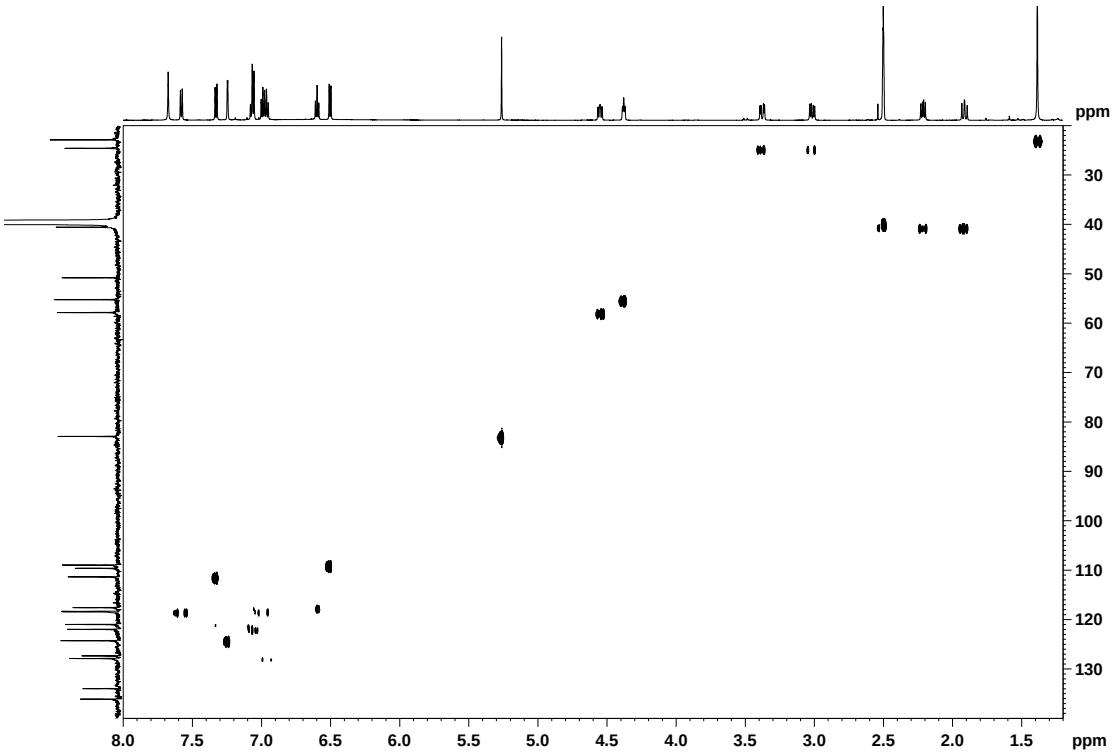


Figure S15. HSQC spectrum of compound 2 (DMSO-*d*6, 600 MHz, 150MHz).

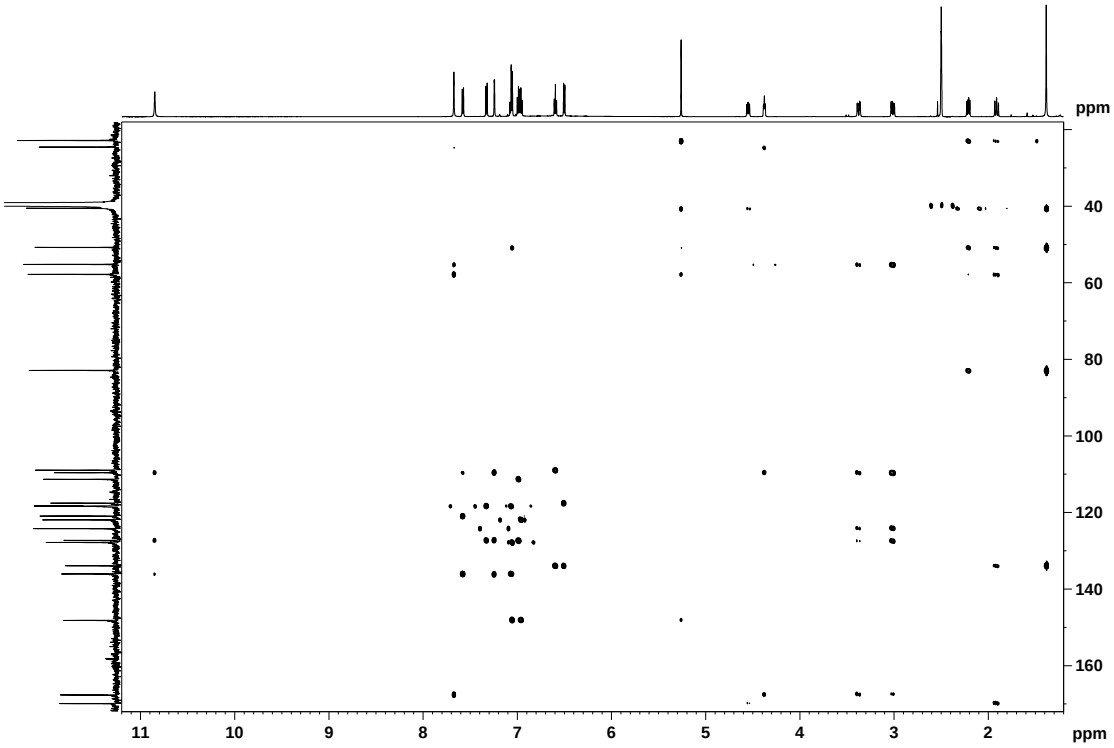


Figure S16. HMBC spectrum of compound 2 (DMSO-*d*6, 600 MHz, 150MHz).

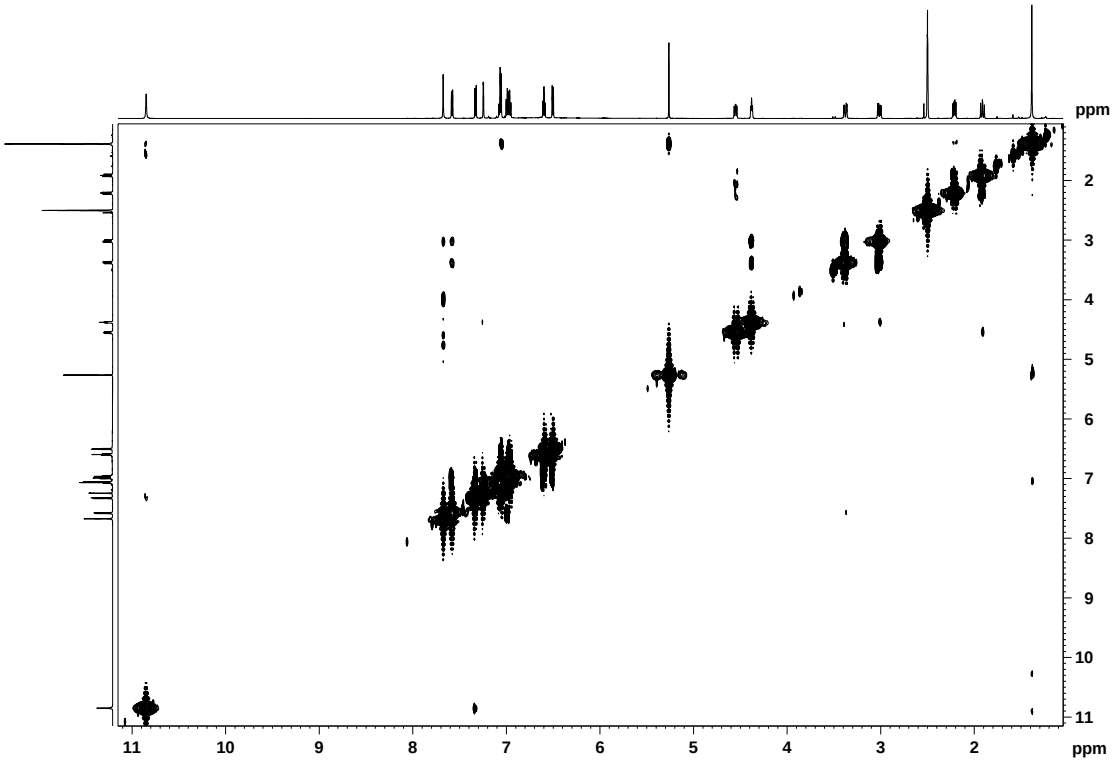


Figure. S17. NOESY spectrum of compound 2 (DMSO-d6, 600 MHz, 600 MHz).

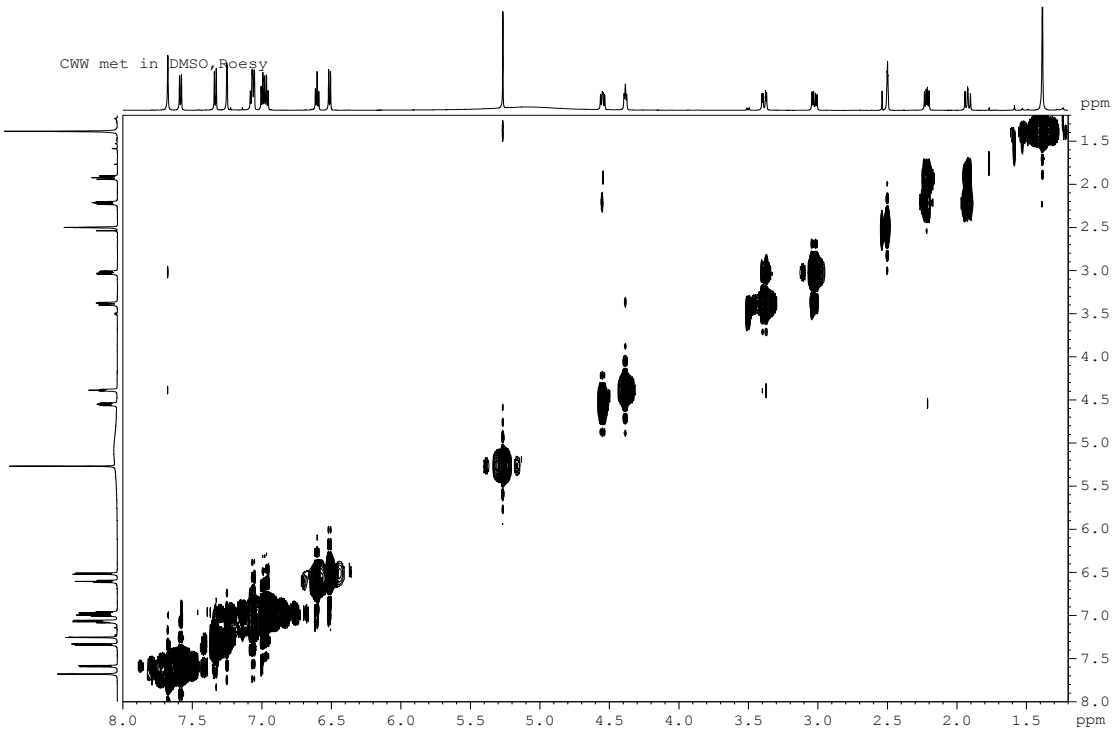


Figure. S18. ROESY spectrum of compound 2 (DMSO-d6, 600 MHz, 600 MHz).

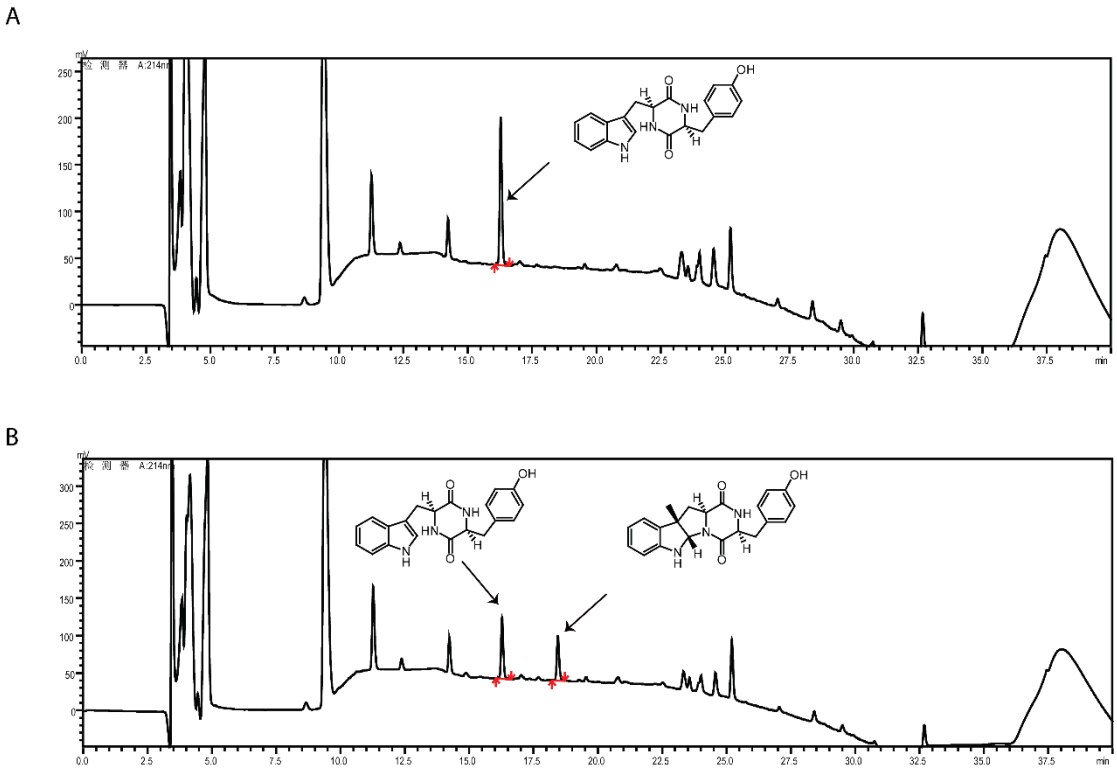


Figure S19. HPLC analysis after in vitro assays with cWY as the substrate. A, Blank; B, in vitro assays with StspM1

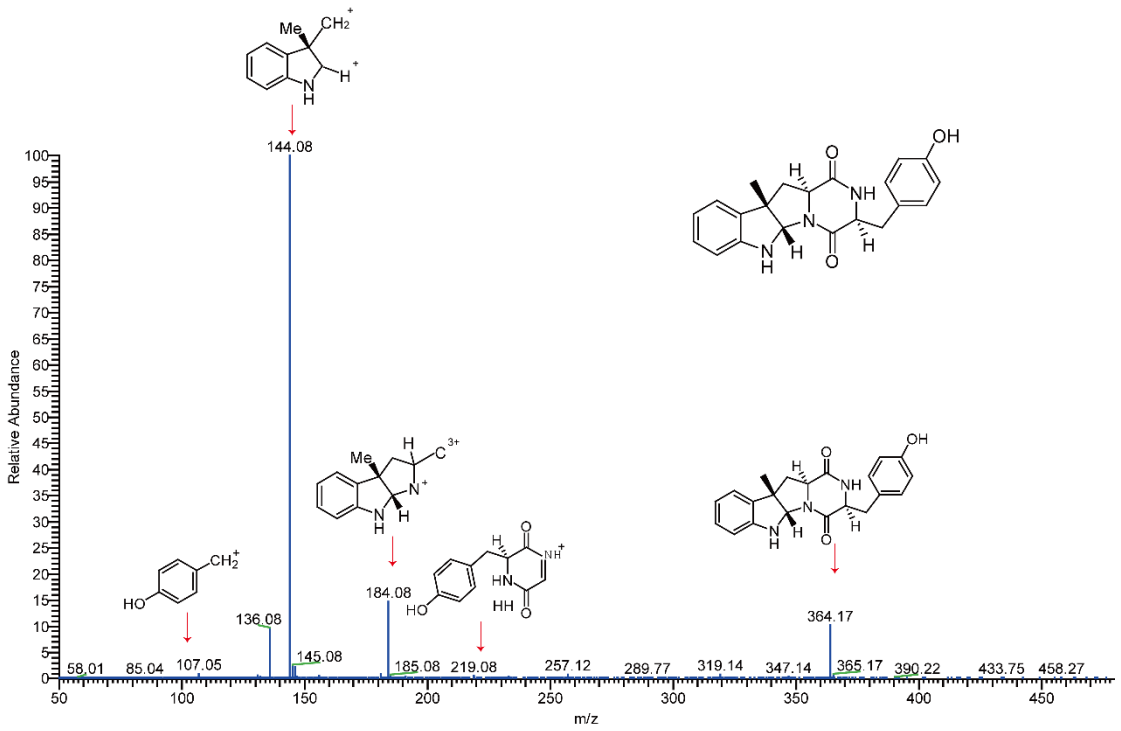


Figure S20. MS2 fragmentation spectrum of methylated cWY (m/z 364). Relevant fragment ions and neutral losses used to identify the in vitro reaction product are shown.

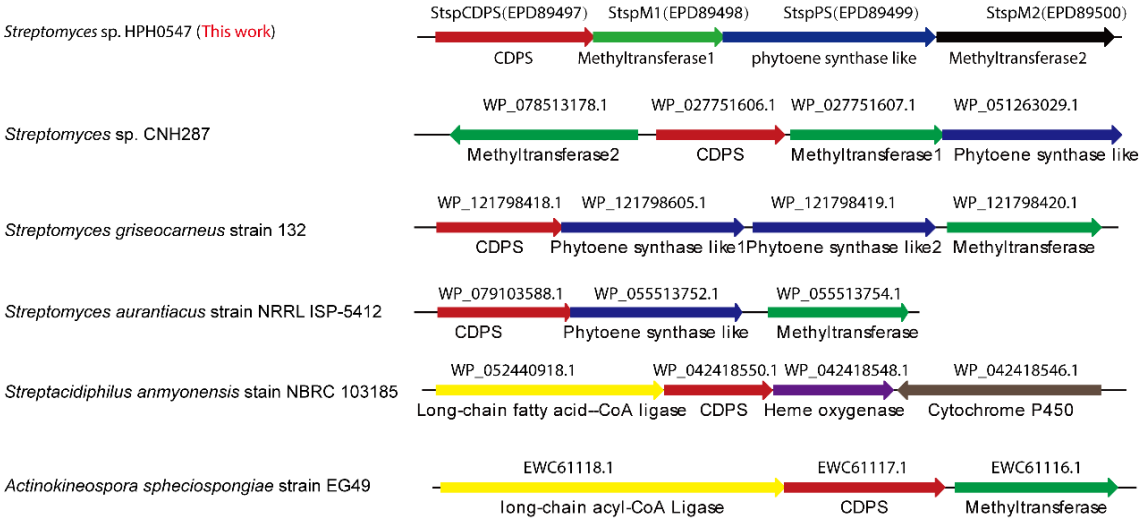


Figure S21. Identified potential CDPS gene clusters with diverse tailoring enzymes.