Supporting Information

The Competition and Cooperation of CaeB6-catalyzed C3-Hydroxylation with CaeG1-catalyzed *O*-Methylation Branch the Biosynthetic Pathway of Caerulomycins toward Different 2,2'-Bipyridine Products

Ming Chen[a], Yipeng Zhang[a], Yanan Du[a], Qunfei Zhao[a], Qinglin Zhang[a,b], Jiequn Wu[a,b], and Wen Liu[a,b,c,*]

[a] State Key Laboratory of Bioorganic and Nature Products Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, 345 Lingling Road, Shanghai 200032, China. Email: wliu@mail.sioc.ac.cn

[b] State Key Laboratory of Microbial Metabolism, School of Life Science & Biotechnology, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China.

[c] Huzhou Center of Bio-Synthetic Innovation, 1366 Hongfeng Road, Huzhou 313000, China.

Table of Contents

1. Supplemental Methods

1.1. General Materials and Methods

Materials, Bacterial Strains and Plasmids. Biochemicals and media were purchased from Sinopharm Chemical Reagent Co. Ltd. (China) or Oxoid Ltd. (UK), unless stated otherwise. Enzymes were purchased from Takara Biotechnology Co. Ltd. (China), except for Taq DNA polymerase, which was purchased from Dingguo Co. Ltd. (China). Chemical compounds and reagents were purchased from Sigma-Aldrich Co. (USA), TCI Development Co. Ltd. (China) and J&K Scientific Ltd. (China). The bacterial strains and plasmids used in this study are summarized in Table S1. The primers used in this study are listed in Table S2.

DNA Isolation, Manipulation and Sequencing. DNA isolation and manipulation in *E. coli* and *Streptomyces* strains were carried out according to standard protocols.[1, 2] Polymerase chain reaction (PCR) amplification was performed using an Eppendorf AG 22331 cycler (Eppendorf AG., Germany) with either Taq DNA polymerase for routine genotype verification or KOD FX or PrimeSTAR HS DNA polymerase for high-fidelity amplification. Primer synthesis and DNA sequencing were performed at Shanghai Sangon Biotech Co. Ltd. (China) and Shanghai Majorbio Biotech Co. Ltd. (China).

Sequence Analysis. The corresponding deduced proteins were compared to other known proteins in databases using the available BLAST methods (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Metabolites Analysis. High performance liquid chromatography (HPLC) analysis was conducted on Agilent 1200 and 1260 HPLC systems (Agilent Technologies Inc., USA). Electrospray ionization mass spectrometry (ESI-MS) was performed on a Thermo Fisher LTQ Fleet ESI-MS spectrometer (Thermo Fisher Scientific Inc., USA), and the data were analyzed using Thermo Xcalibur software. High resolution-ESI-MS (HR-ESI-MS) analysis was achieved on a 6230B Accurate-Mass TOF LC/MS System or 6530 Accurate-Mass Q-TOF LC/MS System (Agilent Technologies Inc., USA), and the data were analyzed using Agilent MassHunter Qualitative Analysis software. Nuclear magnetic resonance

(NMR) data were recorded on Bruker_DRX400 and Bruker AV500_spectrometers (Bruker Co. Ltd, Germany) or Agilent 500 MHz PremiumCompact+ NMR spectrometer (Agilent Technologies Inc., USA).

1.2. Construction of gene-inactivation mutants and complementary strains

In-frame Deletion of *caeB6*. The primers caeB6-L-for and caeB6-L-rev were used to amplify a 2.84 kb DNA fragment from *A. cyanogriseus*, which was cloned into pMD19-T to produce pQL1064. The primers caeB6-R-for and caeB6-R-rev were used to obtain a 2.8 kb fragment, and the insertion of this fragment into pMD19-T led to the generation of pQL1065. The 2.84 kb *Eco*RI-*Nde*I fragment from pQL1064 and the 2.8 kb *Nde*I-*Hind*III fragment from pQL1065 were recovered and then cloned into the *Eco*RI-*Hind*III site of pKC1139, producing the recombinant plasmid pQL1066. To transfer the pQL1066 into *A. cyanogriseus*, conjugation between *A. cyanogriseus* and *E. coli* S17-1 harboring pQL1066 was performed. Following a standard procedure[3] except AS-1 medium (yeast extract 1 g, L-alanine 0.2 g, L-arginine 0.5 g, soluble starch 5 g, NaCl 2.5 g, Na_2SO_4 10 g, agar 20 g, pH 8.0, distilled water to 1 L) and 50 μg/mL apramycin was used for maintenance and selection. The pQL1066 was transferred into *A. cyanogriseus*, and the single-crossover mutants were subsequently generated. To induce double crossover, mutants were streaked onto AS-1 plates without antibiotics supplements and cultured at 30 °C until colonies appeared. This process was repeated five times, and then, colonies were selected to test apramycin resistance. The double-crossover mutants QL2011 could be identified from colonies without apramycin resistance through genotype verification (Figure S1).

In-frame Deletion of *caeG1*. The primers caeG1-L-for and caeG1-L-rev were used to amplify a 1.8 kb DNA fragment from *A. cyanogriseus*, which was cloned into pMD19-T to yield pQL1067. The primers caeG1-R-for and caeG1-R-rev were used to obtain a 1.96 kb fragment. The insertion of this fragment into pMD19-T led to the generation of pQL1068. The 1.8 kb *Eco*RI-*Xba*I fragment from pQL1067 and the 1.96 kb *Xba*I-*Hind*III fragment from pQL1068 were recovered and then cloned into the *Eco*RI-*Hind*III site of pKC1139, producing the recombinant plasmid pQL1069. Following the transfer

procedure, pQL1069 was transferred into *A. cyanogriseus* by conjugation, streaked onto AS-1 plates without antibiotics supplementation, and cultured at 30 °C until colonies appeared. This process was repeated five times, and then, colonies were selected to test apramycin resistance. The double-crossover mutants QL2012 could be identified from colonies without apramycin resistance through genotype verification (Figure S1).

1.3. Fermentation and Metabolites Isolation

Fermentation. Cultures of *A. cyanogriseus* maintained on AS-1 medium were chopped and inoculated in a 250-mL Erlenmeyer flask containing 50 mL of tryptic soy broth (TSB) medium (pH 8.0). After incubation at 28 °C and 220 rpm for 2~3 days, 5 mL of broth was further inoculated in 100 mL fermentation medium (soluble starch 20 g, $NaNO_3$ 2 g, $CaCO_3$ 2 g, K_2HPO_4 1 g, yeast extract 1 g, $MgSO_4 \cdot 7H_2O$ 0.5 g, KCl 0.5 g, pH 8.9, distilled water to 1 L), and *A. cyanogriseus* was cultured under same conditions for 7 days. An equal volume of ethyl acetate was used to extract the broth, and the organic phase was transferred and removed by evaporation under vacuum. The crude extract was dissolved in 2 mL of methanol and used for HPLC analysis on a Phenomenex column (Luna 5 μm C18(2), 4.6×250 mm, Phenomenex, USA). The products were eluted at a flow rate of 1 mL/min over a 35-min gradient as follows: T = 0 min, 20% B; T = 5 min, 20% B; T = 15 min, 50% B; T = 20 min, 90% B; T = 25 min, 90% B; T = 30 min, 20% B; T = 35 min, 20% B (solvent A, H_2O + 0.1% HCOOH; solvent B, CH_3CN + 0.1% HCOOH) with monitoring at 315 nm. For HPLC-ESI-MS analysis, the conditions were the same.

Isolation and purification of compounds 2 (from the wild-type strain), 3 and 4 (from the QL2011 mutant strain). The fermentation broth was extracted three times using half volumes of ethyl acetate. The organic phase was transferred and evaporated under vacuum to obtain the crude extract. Compounds **2**, **3** and **4** were recrystallized with ethanol from the crude extract, and then purified through semipreparative HPLC. HPLC isolation was conducted on an Agilent Zorbax column (SB-C18, 5 μm,

9.4×250 mm, Agilent Technologies Inc., USA) and eluted at a flow rate of 3 mL/min over a 35-min gradient as follows: T = 0 min, 20% B; T = 5 min, 20% B; T = 15 min, 50% B; T = 20 min, 90% B; T = 25 min, 90% B; T = 30 min, 20% B; T = 35 min, 20% B (solvent A, H_2O + 0.1% HCOOH; solvent B, CH_3CN + 0.1% HCOOH) with monitoring at 315 nm. Compounds **2** and **3** were white powders, and **4** was a yellow powder.

1.4. Protein Expression and Purification

CaeB6. The primers caeB6-for and caeB6-rev were used to amplify the 1116 bp DNA fragment containing *caeB6*. The PCR product was cloned into pMD19-T to produce pQL1087 with an *Nde*I-*Hind*III insert that was subsequently cloned into the same sites of pET-37b. The resulting plasmid pQL1088 was transferred into *E. coli* BL21 (DE3) which harboring pGro7 (Takara) for expression. A 50-mL culture of *E. coli* was grown overnight in lysogeny broth (LB) medium contain 25 μg/mL chloromycetin and 50 μg/mL kanamycin, diluted 100-fold in fresh medium and then incubated at 37 °C and 220 rpm for 3 h, followed by the addition of 0.5 mg/L L-arabinose (added when the optical density at 600 nm (OD_{600}) reached 0.6). To induce protein expression, 100 μM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the culture 2 h later, and then, the culture was further incubated at 25 °C for 20 h. The cells were harvested by centrifugation and stored at -80 °C before lysis. The thawed cells were re-suspended in lysis buffer (50 mM Tris-HCl, 300 mM NaCl, 2.5 mM imidazole, 10% (v/v) glycerol, pH 8.0). After disruption by an ultrahigh pressure cells crush homogenizer (FB-110X, Shanghai Litu Mechanical Equipment Engineering Co., Ltd, China or JN-02HC, JNBIO, China), the soluble fraction was collected, subjected to purification by using a HisTrap FF column (GE Healthcare, USA), dialyzed against buffer (50 mM Tris-HCl, 100 mM NaCl, 1mM DTT, 10% (v/v) glycerol, pH 7.5) to remove salts using a PD-10 Desalting Column (GE Healthcare, USA) according to the manufacturers' protocols, concentrated to 53.4 μM by Vivaspin (GE healthcare), and stored at -80°C for *in vitro* assays. The purity of CaeB6 (41 kD) was examined by 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis (Figure S2).

CaeG1. The primers caeG1-for and caeG1-rev were used to amplify the 1062 bp DNA fragment containing *caeG1*. The PCR product was cloned into pMD19-T to yield pQL1090, which had an *EcoR*I-*XholI* insert that was subsequently cloned into the same sites of pGEX-6p-1. The resulting plasmid pQL1091 was transferred into *E. coli* BL21(DE3) for expression. A 50 mL culture of *E. coli* was grown 7 h in LB medium with 50 μg/mL ampicillin, diluted 100-fold in fresh medium, and then incubated at 37 °C and 220 rpm for 3 h, followed by the addition of 100 μM IPTG (added when OD_{600} reached 0.6). CaeG1 with a molecular weight of 39 kD (66.8 kD with tag) was expressed at 16 °C and 220 rpm for 14 h. The cells were then collected by centrifugation and stored at -80 °C before lysis. The thawed cells were re-suspended in lysis buffer (50 mM Tris-HCl, 300 mM NaCl and 2.5 mM imidazole, pH 8.0). After disruption by an ultrahigh pressure cells crushed homogenizer, the soluble fraction was collected and purified using a GSTrap HP column (GE Healthcare, USA), dialyzed against buffer (50 mM Tris-HCl, 100 mM NaCl, 1mM DTT, 10% (v/v) glycerol, pH 7.5) to remove salts using a PD-10 Desalting Column, concentrated to 123 μM by Vivaspin and stored at -80 °C for *in vitro* assays. The purity of the proteins was tested using 10% SDS-PAGE analysis (Figure S2).

1.5. Determination of the Flavin cofactor of CaeB6

The protein solution of CaeB6 at a concentration of 1 mg/ml was incubated at 100 °C for 5 min for denaturation and then subjected to HPLC-DAD analysis on a Phenomenex column (Luna 5 μm C18(2), 4.6×250 mm, Phenomenex, USA) through gradient elution of solvent A (H_2O containing 20 mM ammonium acetate) and solvent B (CH_3CN) at a flow rate of 1 mL/min over a 35-min period with the following steps: T = 0 min, 5% B; T = 2 min, 5% B; T = 20 min, 90% B; T = 25 min, 90% B; T = 30 min, 5% B and T = 35, 5% B (mAU at 448 nm), standard FAD and FMN were used as controls.

1.6. Biochemical Assay of CaeB6

The reaction of CaeB6 was carried out at 30 °C in 50 mM Tris-HCl buffer (pH 8.0) containing 0.2 mM compound **3** in the presence of 10 μM FAD, 1 mM NADPH, and 5 μM CaeB6. After reacted for 2 h, an

equal volume of MeCN was added to the assay to quench the reaction. After centrifugation (10 min at 12000 rpm), the supernatant was analyzed by HPLC or HPLC-ESI-MS on a Phenomenex column (Luna 5μm C18(2), 4.6×250 mm, Phenomenex, USA) and eluted under the same conditions of metabolite analysis.

1.7. Biochemical Assay of CaeG1

Compound 3 as the substrate. The reaction of CaeG1 was carried out at 30 °C in 50 mM Tris-HCl buffer (pH 8.0) containing 0.2 mM compound **3** in the presence of 1 mM SAM and 5 μM CaeG1. After reacted for 2 h, an equal volume of MeOH was added to the assay to quench the reaction. After centrifugation (10 min at 12000 rpm), the supernatant was analyzed by HPLC or HPLC-ESI-MS under the same conditions as those used in the CaeB6 assays.

Compounds 4 and 5 as the substrates. The reaction of CaeG1 was performed at 30 °C in 50 mM Tris-HCl buffer (pH 8.0) containing 0.2 mM compound **4** or **5** in the presence of 1 mM SAM (SAM was added in batches at 0 h, 4 h, 8 h and 12 h, if the reaction time exceeded 4 h) and 5 μM CaeG1. An equal volume of MeOH was added to the assay to quench the reaction at different times. After centrifugation (10 min at 12000 rpm), the supernatant was analyzed by HPLC or HPLC-ESI-MS under the same conditions as those used in CaeB6 assays.

1.8. Preparation and Isolation of compound 5

Dozens of 100-μL reaction mixtures containing 1 mM compound **4**, 5 mM SAM (added in batches at 0 h, 4 h, 8 h and 12 h) and 10 μM CaeG1 in 50 mM Tris-HCl (pH 8.0) were incubated at 30 °C for 24 h. The reaction was quenched and extracted with 200 μL of ethyl acetate three times. The combined organic layer was concentrated for HR-MS and NMR analysis. Compound **5** was a yellow powder.

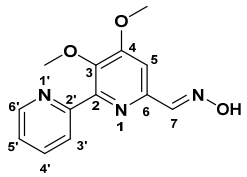
2. Supplemental Tables

Table S1. Bacterial strains, plasmids and primers used in this study.

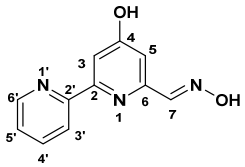
Strain/Plasmid/Primers	Characteristic(s)/Sequence	Source/Use/Reference
E. coli		
DH5α	Host for general cloning	Invitrogen
BL21 (DE3)	Host for protein expression	NEB
Actinoalloteichus		
NRRL B-2194	Wild type strain, CAE-producing	NRRL
QL2011	*caeB6* inframe deletion mutant, CAE-C non-producing	This study
QL2012	*caeG1* inframe deletion mutant, CAE-A non-producing	This study
Plasmids		
pMD19-T	*E. coli* subcloning vector	Takara
pKC1139	*E. coli-Streptomyces* shuttle vector for gene inactivation and complementary, temperature sensitive replication in *Streptomyces*	[2]
pET-37a	Protein expression vector used in *E. coli*, encoding His tag in the C-terminal, kanamycin resistance	Novagen
pGro7	Protein coexpression vector used to increase recovery of target proteins in the soluble fraction in *E. coli*, chloromycetin resistance	Takara
pGEX-6p-1	Protein expression vector used in *E. coli*, encoding GST tag in the N-terminal, ampicillin resistance	Novagen
pQL1064	pMD19-T derivative containing partial *caeB6* fragment	This study
pQL1065	pMD19-T derivative containing partial *caeb6* fragment	This study
pQL1066	pKC1139 derivative for *caeB6* in-frame deletion	This study
pQL1067	pMD19-T derivative containing partial *caeG1* fragment	This study
pQL1068	pMD19-T derivative containing partial *caeG1* fragment	This study
pQL1069	pKC1139 derivative for *caeG1* in-frame deletion	This study
pQL1087	pMD-19T derivative containing *caeB6* fragment	This study
pQL1088	pET-37b derivative containing *caeB6* fragment	This study
pQL1089	pET-37 derivative containing *caeB6*-R205A fragment	This study
pQL1090	pMD-19T derivative containing *caeG1* fragment	This study
pQL1091	pGEX-6p-1 derivative containing *caeG1* fragment	This study
Primers		
caeB6-L-for	TAAGAATTCGCTGCACTACCTGGTGCTGCTGC	*caeB6* inframe deletion
caeB6-L-rev	TAACATATGATCGGCGGATTCGTGGACCTGG	
caeB6-R-for	TAACATATGCCGATGGTGGACCTGACCCTGG	
caeB6-R-rev	TAAAAGCTTACCACCTGCGGACAGAGTCTGG	

caeB6-gt-for	GATGGCGTTGTAGCGGTCGTCAC	Genotype
caeB6-gt-rev	GCTGCCCTACCTGGGCATCGG	verification
caeG1-L-for	TTGAATTCTACACCTCCCGCACGTCTTGA	*caeG1*
caeG1-L-rev	GCTCTAGAGGCCAGGTCGTCGACGTGTG	inframe
caeG1-R-for	GCTCTAGAGAACGGACCCGCACGGACT	deletion
caeG1-R-rev	CCAAGCTTTCGGCACGGCGGGGAATGGT	
caeG1-gt-for	GGAGGTGTTGCCGTTGGAGG	Genotype
caeG1-gt-rev	CTCCACCAGGCTGAAGTCCA	verification
caeB6-for	ATACATATGCGAAAATCAGCGGAAATCGCGG	CaeB6
caeB6-rev	TATAAGCTTCCCGCCGCGACGCGGTGGCACC	expression
caeB6-R205A-for	CCCGGTCCCGGCGGGCAGTGTTGTACGTGC	CaeB6-R205A
caeB6-R205A-rev	GCACGTACAACACTGCCCGCCGGGACCGGG	expression
caeG1-for	AAAGAATTCATGTCGGAAATGACCGCCGCGGAG	CaeG1
caeG1-rev	TTTCTCGAGTCAGACCGGGCGGGCCTCCACCAG	expression

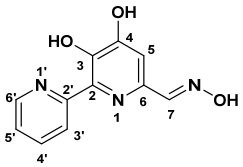
Table S2. HR-ESI-MS and ^{1}H and ^{13}C NMR data for CAEs (in deuterated dimethyl sulfoxide (DMSO-*d6*); 500 MHz for ^{1}H and 125 MHz for ^{13}C NMR; chemical shifts are reported in ppm).



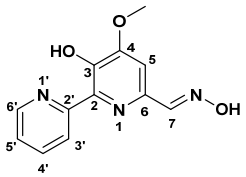
2
CAE-C



3
CAE-H



4



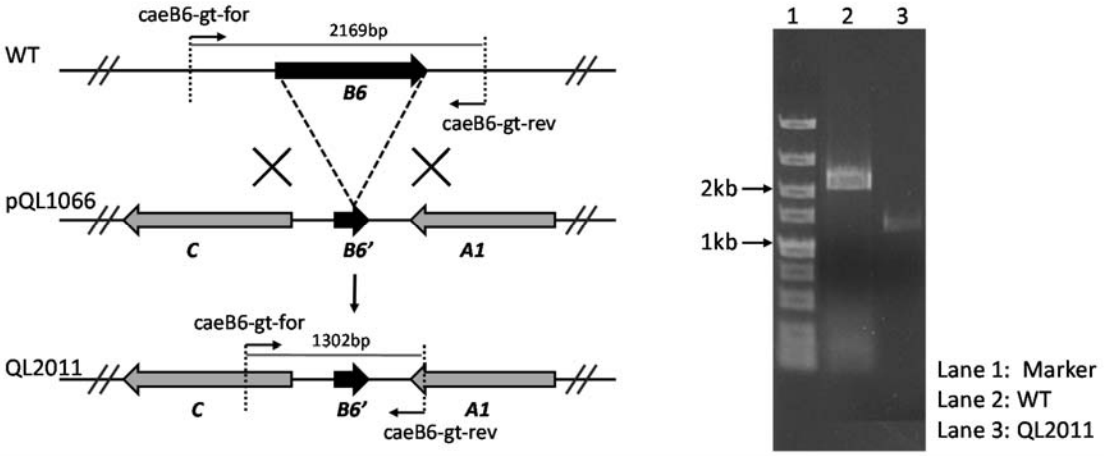
5
CAE-B

Compound		**2**		**3**		**4**		**5**	
HR-ESI-MS	Mol. Formula	$C_{13}H_{13}N_3O_3$		$C_{11}H_9N_3O_2$		$C_{11}H_9N_3O_3$		$C_{12}H_{11}N_3O_3$	
	Obs. $[M+H]^+$	260.1030		216.0768		232.0713		246.0873	
	Cacl. $[M+H]^+$	260.1025		216.0722		232.0717		246.0877	
NMR		δ_C	δ_H (J)	δ_C	δ_H (J)	δ_C	δ_H (J)	δ_C	δ_H (J)
	2	151.0	--	163.0	--	153.7	--	155.0	--
	3	--	--	109.0	7.80 d	146.7	--	134.1	--
	3-OH	--	--	--	--	--	14.46 s	--	--
	3-OCH_3	61.0	3.74 d	--	--	--			--
	4	159.4	--	166.4	--	134.6	--	157.1	--
	4-OH	--	--	--	--	--	--	--	--
	4-OCH_3	56.0	3.97 s	--	--	--	--	55.7	3.93 s
	5	103.5	7.47 s	107.7	7.27 s	107.4	7.29 s	103.5	7.38 s
	6	148.2	--	152.2	--	143.3	--	145.8	--
	7	148.5	8.02 s	148.0	8.17 s	148.8	7.99 s	147.1	8.05 s
	7-NOH	--	11.60 s	--	11.85 s	--	11.35 s	--	11.44 s
	2’	155.9	--	153.4	--	157.3	--	155.2	--
	3’	124.0	7.67 d	121.1	8.38 d	120.5	8.51 d	123.7	8.52 d
	4’	136.4	7.90 t	137.8	8.00 d	138.8	8.10 m	138.9	8.10 m
	5’	123.2	7.43 m	124.8	7.53 m	123.6	7.55 m	120.6	7.56 m
	6’	148.8	8.66 d	149.0	8.71 d	145.8	8.66 d	148.8	8.66 d

3. Supplemental Figures

Figure S1. Verification of mutant genotypes. **a**) Construction of the *caeB6* in-frame deletion mutant QL2011. The primer pair caeB6-gt-for and caeB6-gt-rev was used for genotype verification. **b**) Construction of the *caeG1* in-frame deletion mutant QL2012. The primer pair of caeG1-gt-for and caeG1-gt-rev was used for genotype verification.

a)



b)

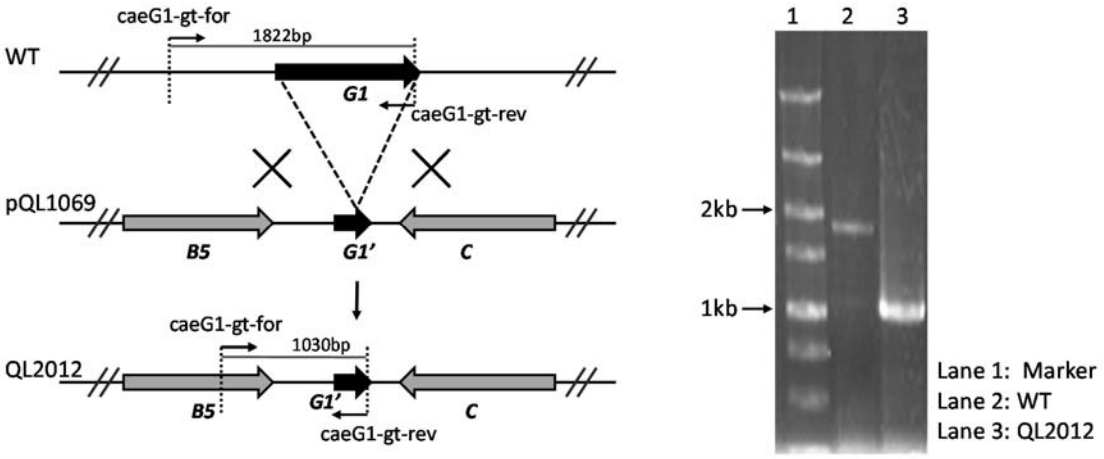


Figure S2. SDS-PAGE analysis of purified recombinant proteins. Each gel contains the protein ladder (left) and the purified recombinant protein (right): a) CaeB6 (41 kD) and b) CaeG1 (66.8 kD with GST tag).

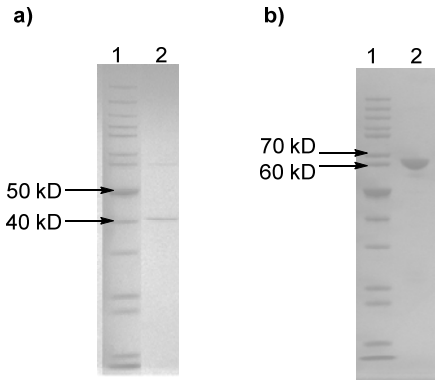


Figure S3. Examination of the protein nature of CaeB6. **a**) UV spectrum of the CaeB1 supernatant. **b**) HPLC-DAD analysis of (i) FMN, (ii) FAD, and (iii) the supernatant of CaeB6.

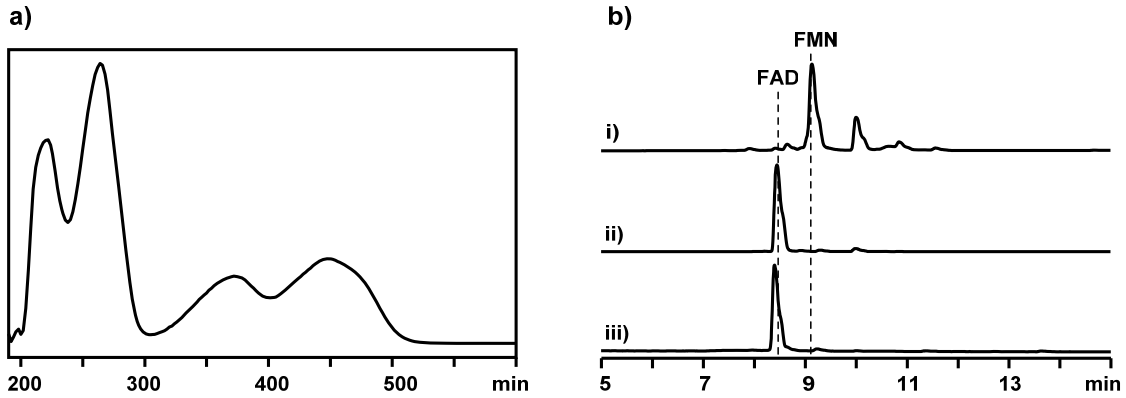


Figure S4. Reconstitution of CaeB6 *in vitro* activity. a) HPLC analysis of (i) standard **1**; and (ii) a complete CaeB6 assay containing **1**, FAD, NAD(P)H and CaeB6. b) HPLC analysis of (i) standard **3**; (ii) standard **4**; (iii) a complete CaeB6 assay containing **3**, FAD, NADPH and CaeB6; (iv) a complete assay without FAD; (v) a complete assay in which FAD was replaced by FMN; (vi) a complete assay without NADPH; and (vii) a complete assay in which NADPH was replaced by NADH.

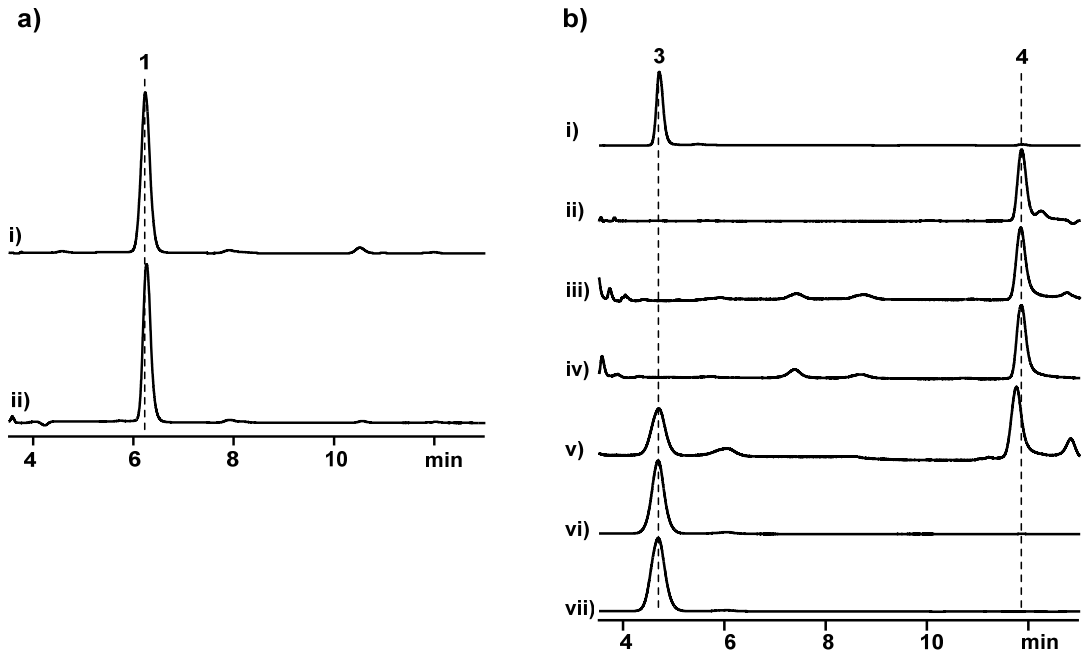


Figure S5. *In vitro* characterization of CaeB6- and CaeG1-catalyzed reactions. a) Assays of CaeG1 activity in the conversion of **3** into **1** analyzed by HPLC, the UV absorbance was monitored at 315 nm. A complete CaeG1 assay containing 4, SAM and CaeG1 with reaction times of (i) 2 h, (ii) **4** h, (iii) 8 h, (iv) 12 h and (v) 24 hr and (vi) standard **4**. b) Assays of CaeG1 activity in the conversion of **5** into **2** analyzed by HPLC-MS, the ESI *m/z* $[M+H]^+$ modes are indicated in the dashed rectangle. A complete CaeG1 assay containing **5**, SAM and CaeG1 with reaction times (i) 1 h, (ii) 2 h, (iii) 4 h, (iv) 8 h and (v) 12 h.

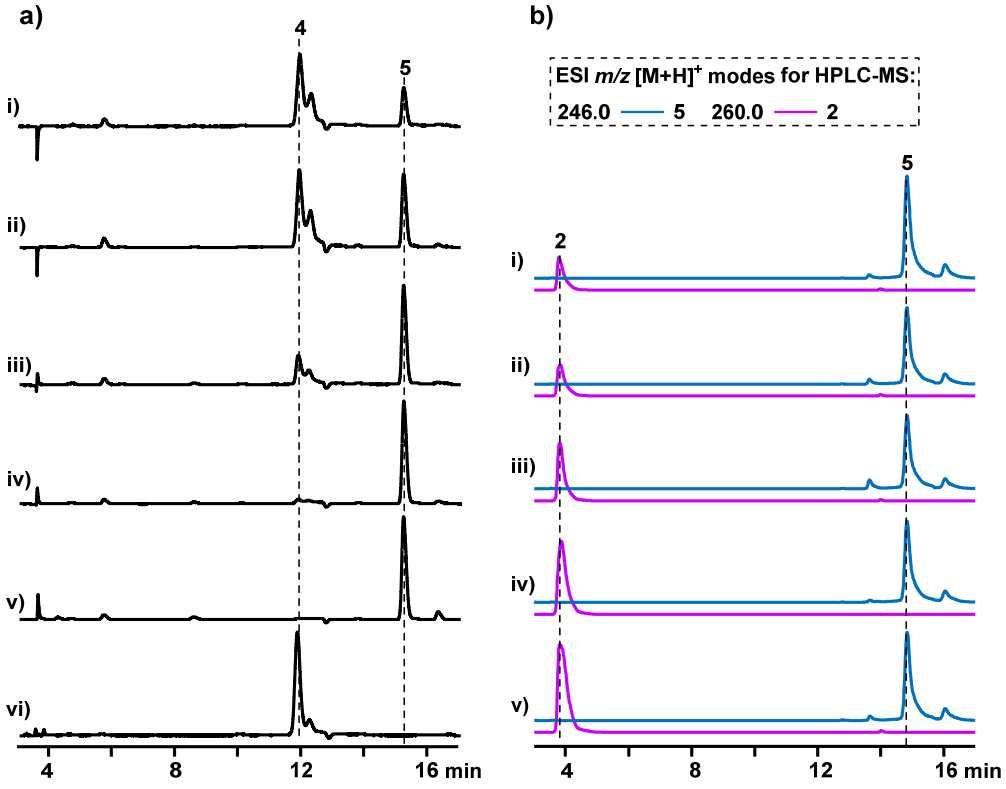


Figure S6. The ^{1}H NMR (DMSO-*d6*, 500 MHz) spectrum of **4**.

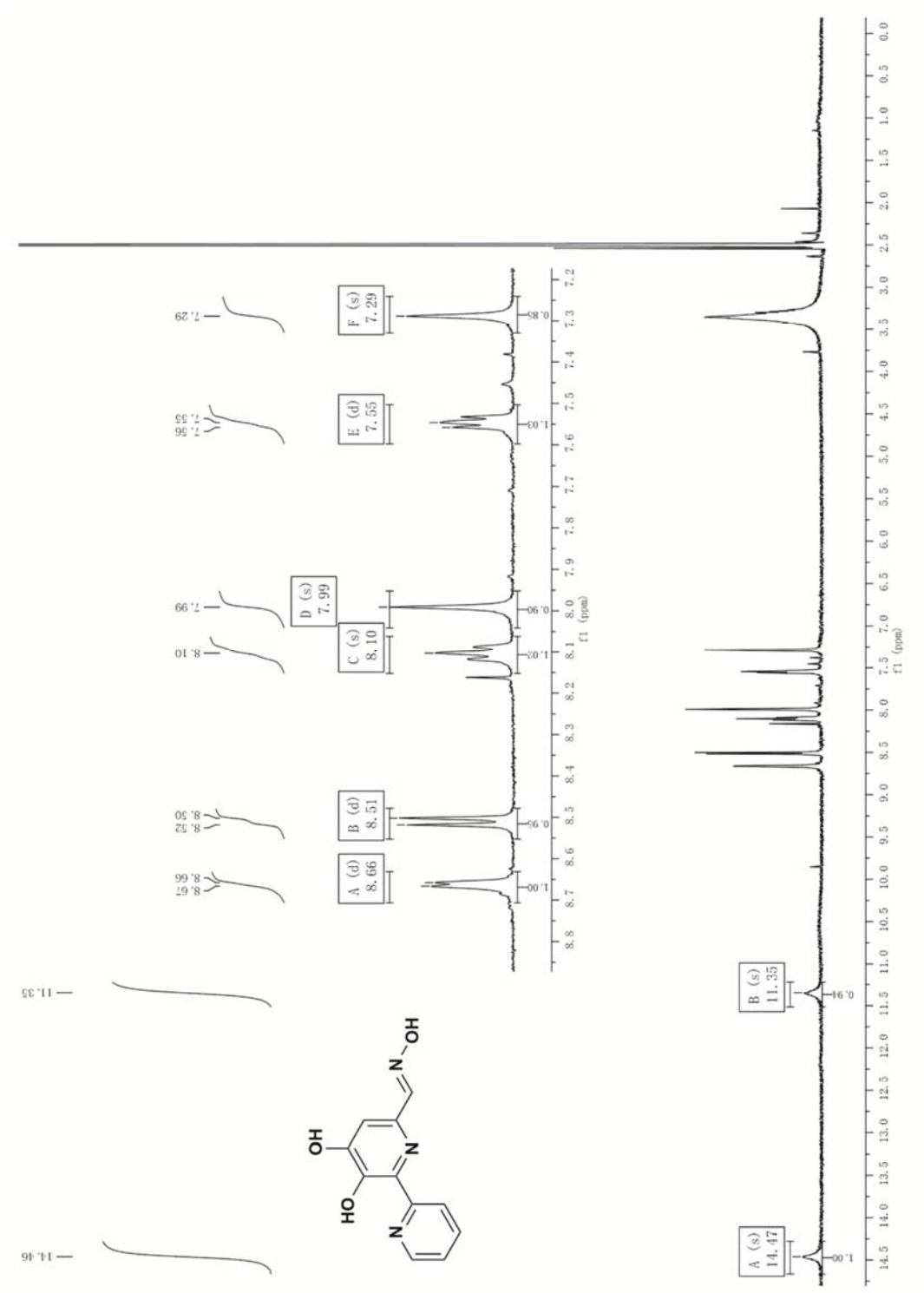


Figure S7. The ^{13}C NMR (DMSO-*d6*, 125 MHz) spectrum of **4**.

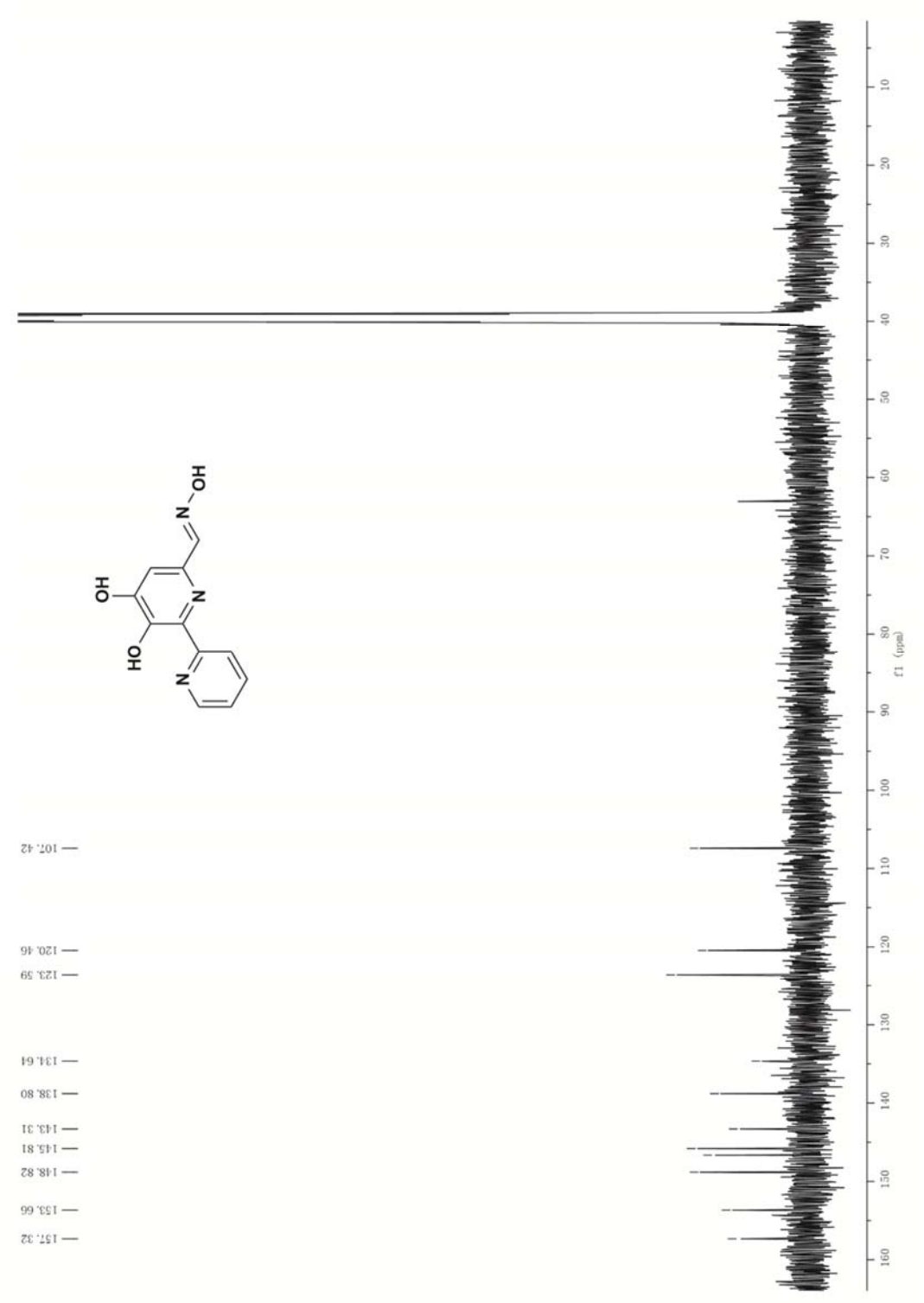


Figure S8. The ^{1}H-^{1}H COSY (DMSO-*d6*, 500 MHz) spectrum of **4**.

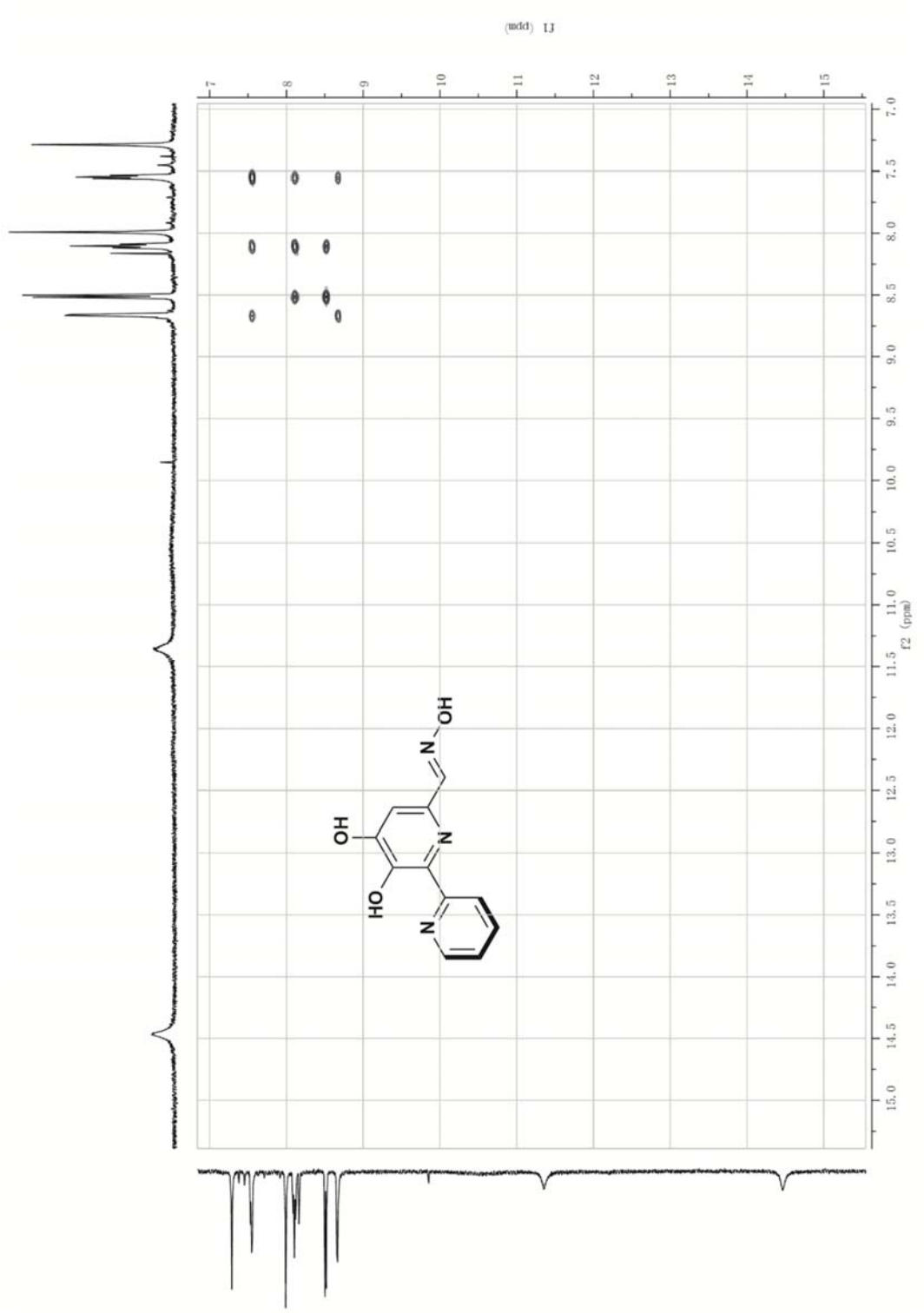


Figure S9. The ^{1}H-^{13}C HSQC (DMSO-*d6*, 125 MHz) spectrum of **4**.

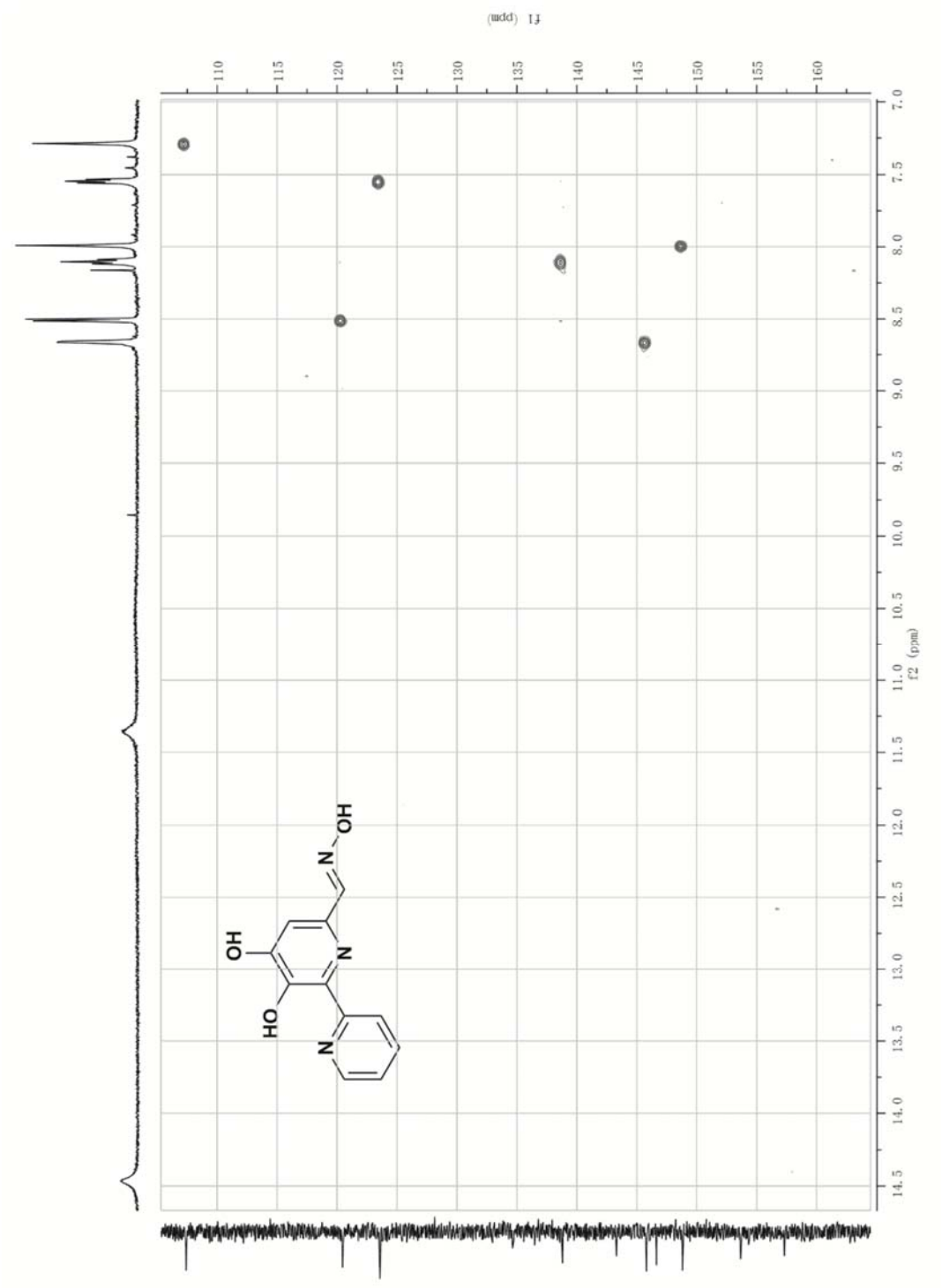
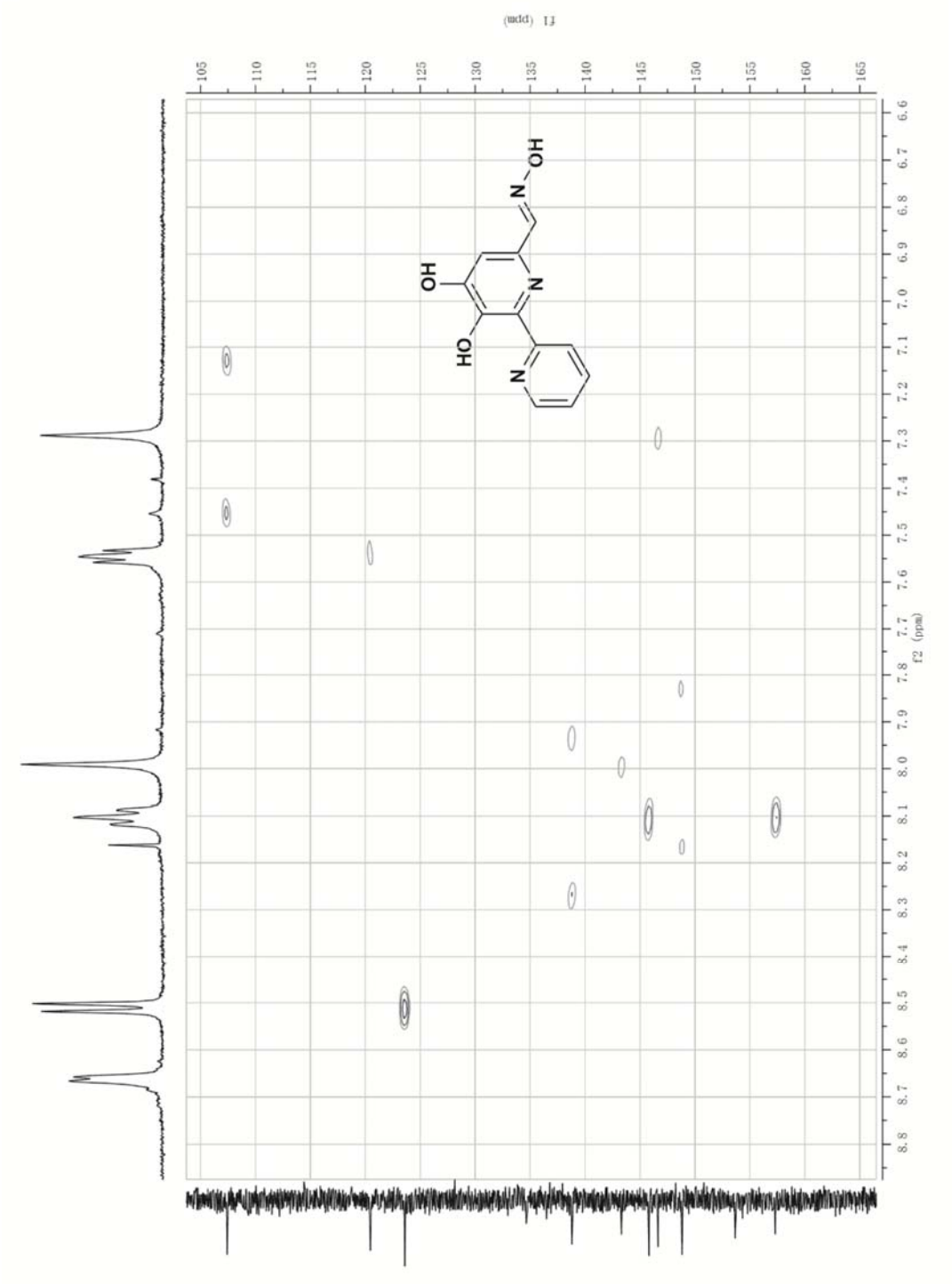


Figure S10. The HMBC (DMSO-*d6*, 125 MHz) spectrum of **4**.



4. Supplemental References

1. S. Joseph and W. R. David, *Molecular cloning: a laboratory manual*, 2001.
2. T. Kieser, *Practical streptomyces genetics*, John Innes Foundation, 2000.
3. X. Qu, N. Jiang, F. Xu, L. Shao, G. Tang, B. Wilkinson and W. Liu, *Mol. Biosyst.*, 2011, **7**, 852-861.