Supporting Information

Isolation and Biosynthesis of Phenazine–Polyketide Hybrids from *Streptomyces* sp. KIB-H483

Zhiyan Wang,[†,‡,⊥] Feng-Xian Yang,[†,⊥] Chongxi Liu,[†] Li Wang,[†] Yuxin Qi,[†,§] Minghang Cao,[†] Xiaowei Guo,[†] Jie Li,[†,‡] Xueshuang Huang,[§,*] Jing Yang,[†,*] Sheng-Xiong Huang[†,*]

[†]State Key Laboratory of Phytochemistry and Plant Resources in West China, and CAS Center for Excellence in Molecular Plant Sciences, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, People's Republic of China

[§]Hunan Provincial Key Laboratory for Synthetic Biology of Traditional Chinese Medicine, Hunan University of Medicine, Huaihua 418000, People's Republic of China

[‡]University of the Chinese Academy of Sciences, Beijing 100049, People's Republic of China

[⊥]These authors contributed equally to this paper.

* Correspondences:

Prof. Dr. Sheng-Xiong Huang, E-mail: sxhuang@mail.kib.ac.cn

Prof. Dr. Jing Yang, E-mail: yangjingc@mail.kib.ac.cn

Prof. Dr. Xueshuang Huang, E-mail: xueshuanghuang@126.com

List of Supporting Information

Fermentation, extraction and isolation of compounds from Δ*npzH*……..5

Scheme S1. 4-Bromobenzoylation of **1**……..6

Table S1. Deduced function of ORFs in the *npz* biosynthetic gene cluster……....7

Table S2. Deduced function of genes involved in the L-rhamnose biosynthesis……..8

Table S3. Strains and plasmids used and generated in this study……..……..9

Table S4. Crystal data and structure refinement for **1a**…….……..10

Table S5. Crystal data and structure refinement for **6**…….....11

Table S6. Primers used in this study……...….12

Figure S1. Partial sequence alignment of the KS, AT and ACP in type I PKS genes.....13

Figrue S2. Phylogenetic tree of *npzL* gene and partial amino acid sequence alignment of NpzL. 14

Figure S3. Sequence alignment of the NpzI to other proteins in *S.* sp. KIB-H483……..15

Figure S4. Schematic view of *rha* gene cluster and proposed biosynthetic pathway for L-rhamnose……....16

Figure S5. X-ray packing diagram of **1a**……..…...17

Figure S6. X-ray packing diagram of **6**……...17

Figure S7. Construction and agarose gel electrophoresis analysis……..…18

Figure S8. SDS-PAGE analysis of protein NpzI…….……18

Figure S9. ^{1}H NMR spectrum of **1** (600 MHz, $CDCl_3$)……..…19

Figure S10. ^{13}C NMR spectrum of **1** (150 MHz, $CDCl_3$)…….…...19

Figure S11. HSQC spectrum of **1** (600 MHz, $CDCl_3$)……20

Figure S12. HMBC spectrum of **1** (600 MHz, $CDCl_3$)……...20

Figure S13. COSY spectrum of **1** (600 MHz, $CDCl_3$)…… 21

Figure S14. ROESY spectrum of **1** (600 MHz, $CDCl_3$)……..21

Figure S15. ^{1}H NMR spectrum of **4** (600 MHz, $CDCl_3$)……..22

Figure S16. ^{13}C NMR spectrum of **4** (150 MHz, $CDCl_3$)……22

Figure S17. HSQC spectrum of **4** (600 MHz, $CDCl_3$)……23

Figure S18. HMBC spectrum of **4** (600 MHz, $CDCl_3$)……23

Figure S19. COSY spectrum of **4** (600 MHz, $CDCl_3$)……24

Figure S20. ^{1}H NMR spectrum of **5** (600 MHz, $CDCl_3$)……24

Figure S21. ^{13}C NMR spectrum of **5** (150 MHz, $CDCl_3$)……25

Figure S22. HSQC spectrum of **5** (600 MHz, $CDCl_3$)……25

Figure S23. HMBC spectrum of **5** (600 MHz, $CDCl_3$)……26

Figure S24. COSY spectrum of **5** (600 MHz, $CDCl_3$)……26

Figure S25. ^{1}H spectrum of **6** (800 MHz, $CDCl_3$)……...27

Figure S26. ^{13}C spectrum of **6** (200 MHz, $CDCl_3$)……..27

Figure S27. HSQC spectrum of **6** (800 MHz, $CDCl_3$)……28

Figure S28. HMBC spectrum of **6** (800 MHz, $CDCl_3$)……28

Figure S29. COSY spectrum of **6** (800 MHz, $CDCl_3$)…….29

Figure S30. ROESY spectrum of **6** (800 MHz, $CDCl_3$)……..29

Figure S31. ^{1}H spectrum of **7** (800 MHz, $CDCl_3$)……30

Figure S32. ^{13}C spectrum of **7** (200 MHz, $CDCl_3$)……..30

Figure S33. HSQC spectrum of **7** (800 MHz, $CDCl_3$)…….31

Figure S34. HMBC spectrum of **7** (800 MHz, $CDCl_3$)……31

Figure S35. COSY spectrum of **7** (800 MHz, $CDCl_3$)……32

Figure S36. ROESY spectrum of **7** (800 MHz, $CDCl_3$). ……..…….…….32

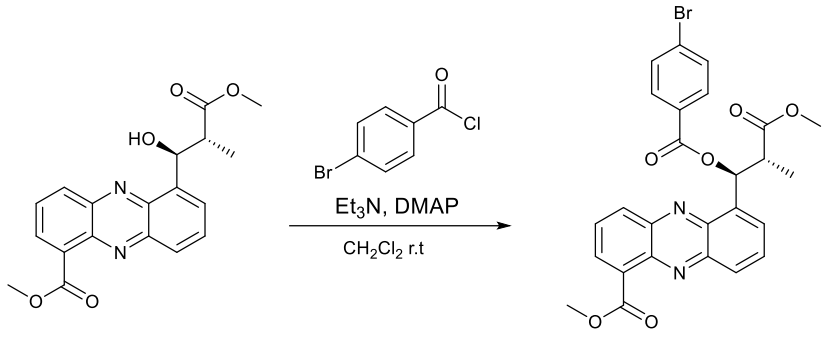
References ……. 33

Fermentation, extraction and isolation of compounds from *ΔnpzH*.

Applying the same method as fermentation of *ΔnpzI*, mutant *ΔnpzH* was cultured with 10 L medium and 1.8 g crude extracts was obtained. 3.5 mg of compound **2** and 1.0 mg of compound **3** were purified from Fr. L subsection by semipreparative HPLC (MeOH–H_2O, 40%, v/v).

Compound **2** (1-carbomethoxyphenazine): yellow solid; ESIMS *m/z* 239 $[M + H]^+$; 261 $[M + Na]^+$; 277 $[M + K]^+$; ^{1}H NMR (600 MHz, $CDCl_3$) δ_H 8.39 (1H, d, J = 8.6 Hz, H-4), 8.34 (1H, m, H-2), 8.24 (2H, d, J = 6.8 Hz, H-7, H-8), 7.87 (3H, overlapped, H-6, H-9, H-3), 4.12 (3H, s, 1-$COOCH_3$). These data were consistent with the structure of 1-carbomethoxyphenazine reported before.[1]

Compound **3** (1,6-dicarbomethoxyphenazine): yellow solid; HRESIMS *m/z* 297.0875 $[M + H]^+$ (calcd for $C_{16}H_{12}N_2O_4$, 297.0875); ^{1}H NMR (800 MHz, $CDCl_3$) δ_H 8.48 (2H, dd, J = 7.0 Hz), 8.29 (2H, dd, J = 7.0 Hz), 7.89 (2H, dd, J = 7.0 Hz), 4.11 (6H, s). These data were consistent with the structure of phenazine-1-carboxylic acid reported before.[2]



Scheme S1. 4-Bromobenzoylation of **1**.

Table S1. Deduced function of ORFs in the *npz* biosynthetic gene cluster.

No.	aa	Proposed function by BLAST search	Identity/ positive	Ortholog identified by BLAST search	Accession number
orf1	293	LysR family transcriptional regulator	76/84	*Streptomyces silaceus*	WP_055701861.1
npzH	770	type I polyketide synthase	67/74	*Streptomyces silaceus*	WP_055701862.1
npzI	351	aldo/keto reductase	36/51	*Aspergillus flavus* NRRL3357	Q00049.2
npzJ	257	thioesterase	31/46	*Aneurinibacillus migulanus*	P14686.1
npzK	928	type I polyketide synthase	53/63	*Streptomyces* sp.	APD71703.1
npzL	117	phosphopantetheine binding protein	65/82	*Streptomyces* sp. JHA26	WP_080406758.1
npzM	447	aldehyde dehydrogenase	80/87	*Streptomyces* sp. JHA26	WP_077801572.1
npzN	366	PDC adenylase	71/80	*Streptomyces caatingaensis*	WP_049716025.1
orf2	168	flavin reductase	73/77	*Streptomyces*	WP_055701869.1
orf3	394	N-methyl-L-tryptophan oxidase	73/82	*Streptomyces* sp. JHA26	WP_077801575.1
orf4	266	SARP family regulatory protein	49/56	*Streptomyces jeddahensis*	OAH10891.1
orf5	239	tRNA-dependent cyclodipeptide synthase	61/74	*Streptomyces caatingaensis*	WP_049714856.1
orf6	108	hypothetical protein	69/81	*Streptomyces caatingaensis*	WP_049714855.1
orf7	184	nitroreductase family protein	70/83	*Streptomyces celluloflavus*	WP_052853393.1
orf8	359	flavin-dependent oxidoreductase	85/90	*Streptomyces silaceus*	WP_055701871.1
orf9	415	DegT/DnrJ/EryCl/StrS aminotransferase	85/91	*Streptomyces silaceus*	WP_055701872.1
orf10	61	4-oxalocrotonate tautomerase	60/69	*Streptomyces griseoviridis*	AGN74882.1
orf11	346	3-oxoacyl-ACP synthase PKS-like	77/86	*Streptomyces luteocolor*	WP_069885392.1
orf12	562	2,3-dihydroxybenzoate-AMP ligase	62/71	*Streptomyces*	WP_030058599.1
orf13	82	phosphopantetheine-binding protein	72/80	*Streptomyces*	WP_055701876.1
npzD	226	isochorismatase	56/66	PhzD, *Streptomyces diastaticus* W2	QJS40197.1
npzE	659	phenazine-specific anthranilate synthase component I	57/70	PhzE, *Streptomyces diastaticus* W2	QJS40198.1
npzC	421	3-deoxy-7-phosphoheptulonate synthase	44/55	PhzC, *Streptomyces diastaticus* W2	QJS40196.1
orf14	861	helix-turn-helix transcriptional regulator	65/73	*Streptomyces*	WP_055701878.1
npzF	278	2,3-dihydro-3-hydroxyanthranilate isomerase	65/72	PhzF, *Streptomyces diastaticus* W2	QJS40193.1
npzG	210	pyridoxamine 5'-phosphate oxidase	62/70	PhzG, *Streptomyces diastaticus* W2	QJS40192.1
orf15	113	hypothetical protein	84/92	*Streptomyces*	WP_055701879.1
npzA/B	162	phenazine biosynthesis protein	71/79	PhzA/B, *Streptomyces diastaticus* W2	QJS40195.1
orf16	634	transcriptional regulator, SARP family	40/52	*Streptomyces violaceusniger* Tu 4113	AEM88110.1

Table S2. Deduced function of genes involved in the L-rhamnose biosynthesis.

No.	Size aa	Proposed function by a BLAST search	Identity/ positive	Ortholog identified by BLAST	Accession number
rhaC	199	dTDP-4-dehydrorhamnose 3,5-epimerase	70/79	*Streptomyces olivaceus*	CAP11386.1
rhaA	876	glucose-1-phosphate thymidylyltransferase	76/86	*Streptomyces olivaceus*	CAP11385.1
rhaB	325	dTDP-glucose 4,6-dehydratase	76/85	*Streptomyces olivaceus*	CAP11384.1
rhaD	306	dTDP-4-dehydrorhamnose reductase	69/79	*Streptomyces olivaceus*	CAP11383.1

Table S3. Strains and plasmids used and generated in this study.

Stains/Plasmids	Characteristics	Source
Streptomyces		
S. sp. KIB-H483	Native producer	Isolated
ΔnpzH	*NpzH* inactivation mutant of *S.* sp. KIB-H483	This study
ΔnpzK	*NpzK* inactivation mutant of *S.* sp. KIB-H483	This study
ΔnpzI	*NpzI* inactivation mutant of *S.* sp. KIB-H483	This study
ΔnpzL	*NpzL* inactivation mutant of *S.* sp. KIB-H483	This study
ΔnpzN	*NpzN* inactivation mutant of *S.* sp. KIB-H483	This study
Escherichia coli		
DH10B	Host strain for cloning	Invitrogen
BW25113/pIJ790	Host strain for PCR targeting	Laboratory stock
ET12567/pUZ8002	Donor strain for conjugation	Laboratory stock
XL1-blue	Host strain for genomic library	Agilent Technologies
BL21(DE3)	Host strain for expression of protein	Laboratory stock
Plasmids		
pIJ773	Apr^r, Plasmid containing the apramycin resistance gene	Laboratory stock
pSuperCos I-3G11	Kan^r, Cosmid containing PKS gene cluster used for knochout	This study
p3G11-*ΔnpzH*	Apr^r, gene inactivation clone used for mutant *ΔnpzH*	This study
p3G11-*ΔnpzK*	Apr^r, gene inactivation clone used for mutant *ΔnpzK*	This study
p3G11-*ΔnpzI*	Apr^r, gene inactivation clone used for mutant *ΔnpzI*	This study
pET26b	Kan^r, plasmid for the expression of protein	Laboratory stock
pET26b-NpzI	Kan^r, plasmid for the expression of *npzI*	This study

Abbreviations: Kan^r, kanamycin resistance; Apr^r, apramycin resistance.

Table S4. Crystal data and structure refinement for **1a**.

Compound	**1a**
Identification code	global
Empirical formula	$C_{26}H_{21}BrN_2O_6$
Formula weight	537.36
Temperature	100(2) K
Wavelength	0.71073 Å
Crystal system	Orthorhombic
Space group	$P2_12_12_1$
Unit cell dimensions	a = 5.8283(9) Å α= 90º.
	b = 17.667(3) Å β= 90º.
	c = 22.751(3) Å γ= 90º.
Volume	2342.6(6) $Å^3$
Z	4
Density (calculated)	1.524 mg/m^3
Absorption coefficient	1.801 mm^{-1}
F(000)	1096
Crystal size	0.900 × 0.070 × 0.040 mm^3
Theta range for data collection	1.459 to 31.223º.
Index ranges	-8<=h<=8, -24<=k<=25, -32<=l<=28
Reflections collected	26423
Independent reflections	7012 [R(int) = 0.0673]
Completeness to theta = 25.242º	100.0 %
Absorption correction	Semi-empirical from equivalents
Refinement method	Full-matrix least-squares on F^2
Data/restraints/parameters	7012/651/456
Goodness-of-fit on F^2	0.997
Final R indices [I>2sigma(I)]	R1 = 0.0436, wR2 = 0.0855
R indices (all data)	R1 = 0.0751, wR2 = 0.0958
Flack parameter	0.013(6)
Extinction coefficient	n/a
Largest diff. peak and hole	0.494 and -0.501 $e.Å^{-3}$

Table S5. Crystal data and structure refinement for **6**.

Compound	**6**
Identification code	global
Empirical formula	$C_{23}H_{26}N_2O7$
Formula weight	442.46
Temperature	100(2) K
Wavelength	1.541780.71073 Å
Crystal system	Monoclinic
Space group	C 1 2 1
Unit cell dimensions	a = 32.3728(12) Å α= 90º.
	b = 5.1558(2) Å β= 104.4460(10)º.
	c = 13.2082(5) Å γ= 90º.
Volume	2134.85(14) $Å^3$
Z	4
Density (calculated)	1.377 mg/m^3
Absorption coefficient	0.854 mm^{-1}
F(000)	936
Crystal size	1.000 × 0.060 × 0.030 mm^3
Theta range for data collection	2.82 to 72.09º
Index ranges	-39<=h<=39, -6<=k<=5, -16<=l<=16
Reflections collected	29925
Independent reflections	4163 [R(int) = 0.0525]
Completeness to theta = 72.09º	99.9 %
Absorption correction	Semi-empirical from equivalents
Max. and min. transmission	0.97 and 0.84
Refinement method	Full-matrix least-squares on F^2
Data/restraints/parameters	4163/1/295
Goodness-of-fit on F^2	1.028
Final R indices [I>2sigma(I)]	R1 = 0.0273, wR2 = 0.0682
R indices (all data)	R1 = 0.0286, wR2 = 0.0689
Flack parameter	0.06(6)
Largest diff. peak and hole	0.185 and -0.175 e.$Å^{-3}$

Table S6. Primers used in this study.

Target gene	Primer names	Sequences	Product length (bp)
Primers for screening libraries			
483SL-1	483SL-1F	gatcgacgccgaggacacc	542
	483SL-1R	tgagccgttcgaggatgcc	
482SL-2	483SL-2F	ggctcccgtccgtgaacc	522
	483SL-2R	gactcccagaactccaggcag	
Primer pairs used for mutant producer strain construction			
ΔnpzH	NPH-UP	tcagcgggccgccacggcgtccaggtcgaacccggtgcgattccggggatccgtcgacc	1448
	NPH-DOWN	atgagcacgatggacccggccgacgacggatccgccgtctgtaggctggagctgcttc	
ΔnpzI	NPI-UP	tgacgggcactacggcggtggccgccgtcgcggcccgggattccggggatccgtcgacc	1448
	NPI-DOWN	cgcaacaacggcctgcgggtctcggagctcgccctgggatgtaggctggagctgcttc	
ΔnpzK	NPK-UP	atggcaggccgctttcccggcgcccgggacaccgacgagattccggggatccgtcgacc	1448
	NPK-DOWN	acggaccgcggccgactgcccctcgacgctgggcgcggttgtaggctggagctgcttc	
ΔnpzL	NPL-UP	atgacgcattggagtggcaccgtgacatcccagtcccccattccggggatccgtcgacc	1448
	NPL-DOWN	gtcgagggcccggggcctggcctgcgaggaccggccgcctgtaggctggagctgcttc	
ΔnpzN	NPN-UP	tggccgggtcgacgacccggaaggtcacgtacggcgagaattccggggatccgtcgacc	1448
	NPN-DOWN	gtgtcacccgtagaacaatccgtgctggacctgccgttctgtaggctggagctgcttc	
Primer pairs used for PCR verification of the double cross-over mutants			
ΔnpzH	NPH-VF	ctgcgcaccatacggatcg	2486 bp for wild type strain
	NPH-VR	gagcgaacgaaggaggagg	1621 bp for mutant strain
ΔnpzI	NPI-VF	tacgggcggtcgtgcgaca	1055 bp for wild type strain
	NPI-VR	ggaggcgcaccatgcgcta	1682 bp for mutant strain
ΔnpzK	NPK-VF	gtggccgtcatcggcatgg	1020 bp for wild type strain
	NPK-VR	tccgatggcgcagtagccc	1616 bp for mutant strain
ΔnpzL	NPL-VF	gcctgccatgccgatgacg	618 bp for wild type strain
	NPL-VR	cgcagcaagacggtgctgc	1766 bp for mutant strain
ΔnpzN	NPN-VF	tcatcacgacctgcccgcg	987 bp for wild type strain
	NPN-VR	gcggtcaccgtttcctcgc	1524 bp for mutant strain
Primer pairs used for protein expression			
NpzI	26b-NPBF	aagaaggagatatacatatgcgctatacgaccttcggacg	1090
	26b-NPBR	tcgagtgcggccgcaagcttcgccaccggtacgggcggtc	

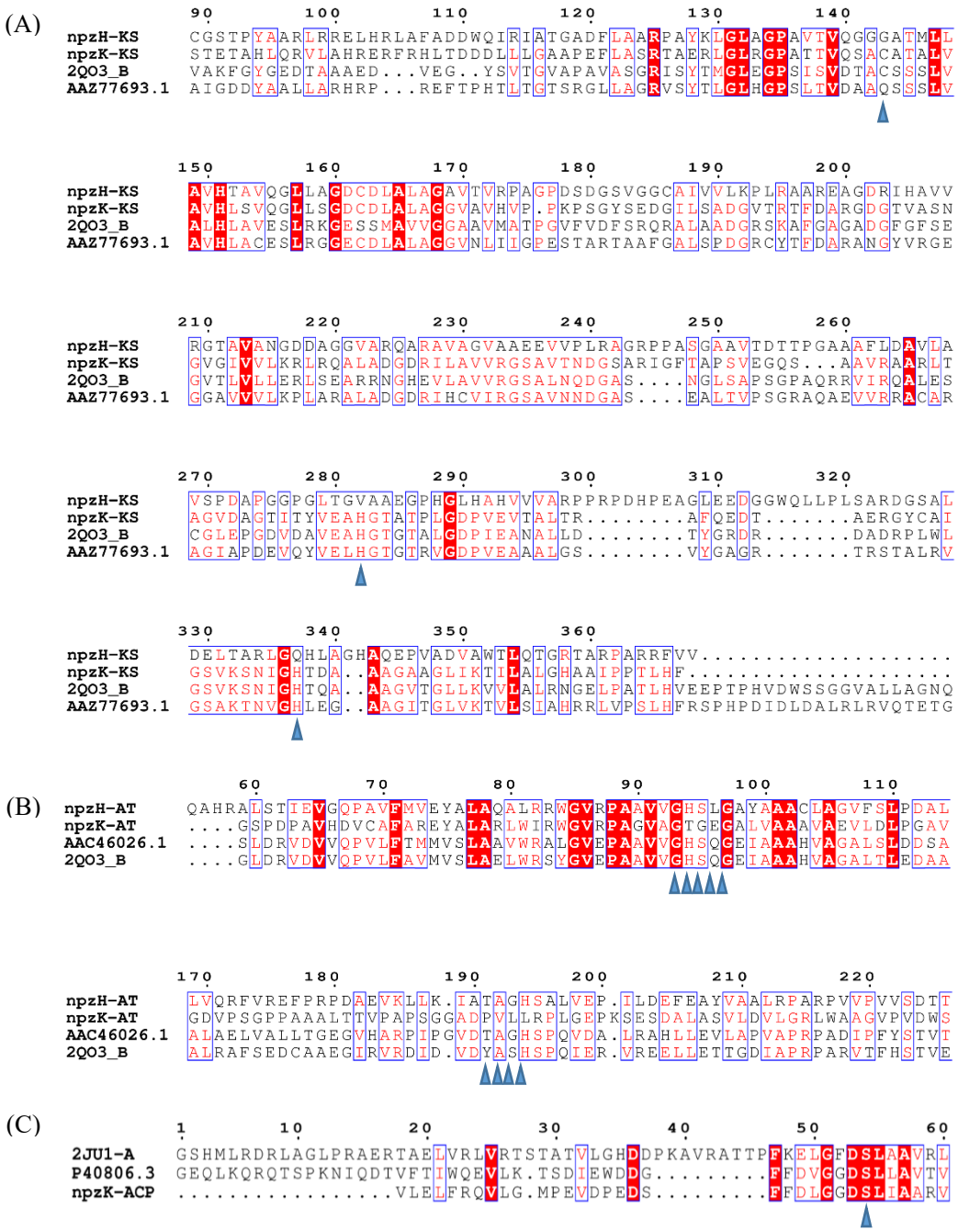
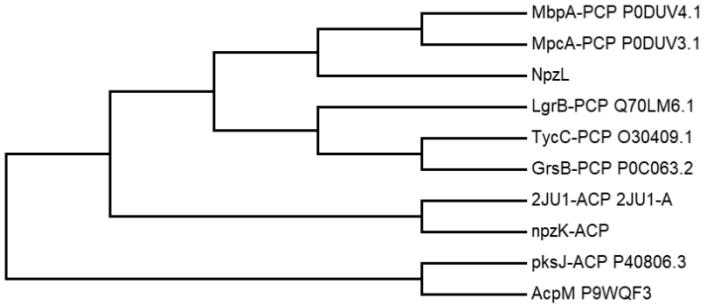
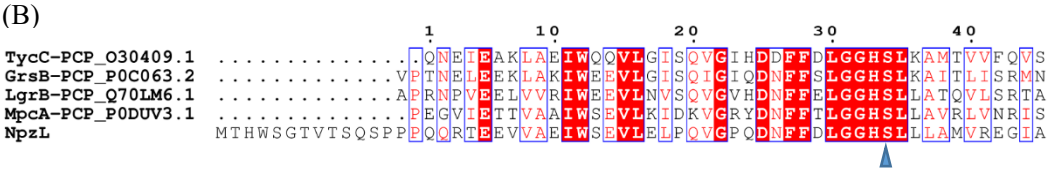


Figure S1. Partial sequence alignment of the KS, AT and ACP in type I PKS genes in *npz* cluster. The alignment was proceeded with other related proteins by ClustalX.[3] (A) KS domains of NpzH and NpzK; KS domain of erythromycin polyketide synthase module 3 (2QO3); KS domain of chlorothricin polyketide synthase ChlA1 (AAZ77693.1). (B) AT domains of NpzH and NpzK; AT domain of erythromycin polyketide synthase module 3 (2QO3); AT domain of niddamycin polyketide synthase module 5 (AAC46026.1). (C) ACP domain of 6-deoxyerythronolide B synthase module 2 (2JU1); ACP domain of PksJ (P40806.3); ACP domain of NpzK. Blue triangles refer to active sites of proteins.

(A)



(B)



Figrue S2. Phylogenetic tree of *npzL* gene (A) and partial amino acid sequence alignment of NpzL (B). (A) Maximum-likelihood tree showing the NpzL was shown to group in the PCP clade. (B) Blue triangles refer to active sites of proteins.

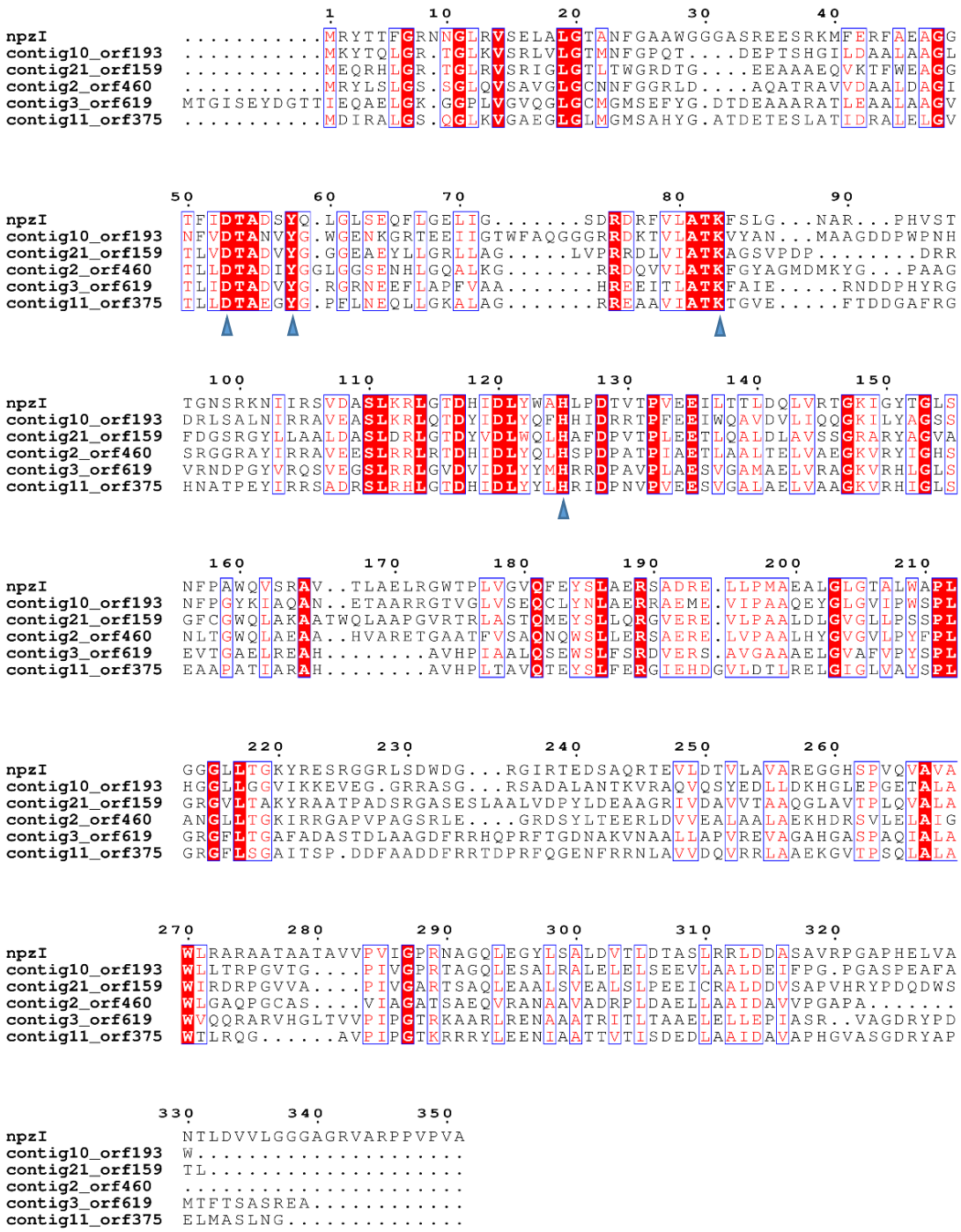


Figure S3. Sequence alignment of the NpzI to other proteins in *S.* sp. KIB-H483 with more than 33% identity. Contig10_orf193, contig21_orf159, contig2_orf460, contig3_orf619 and contig11_orf375 separately showed 40.1%, 39.1%, 37.9%, 36.3% and 34.0% identity to NpzI. All of them are predicted to be aldo-keto reductases (AKRs) and contained the conserved tetrad of active site residues (marked by blue triangles) of AKRs.

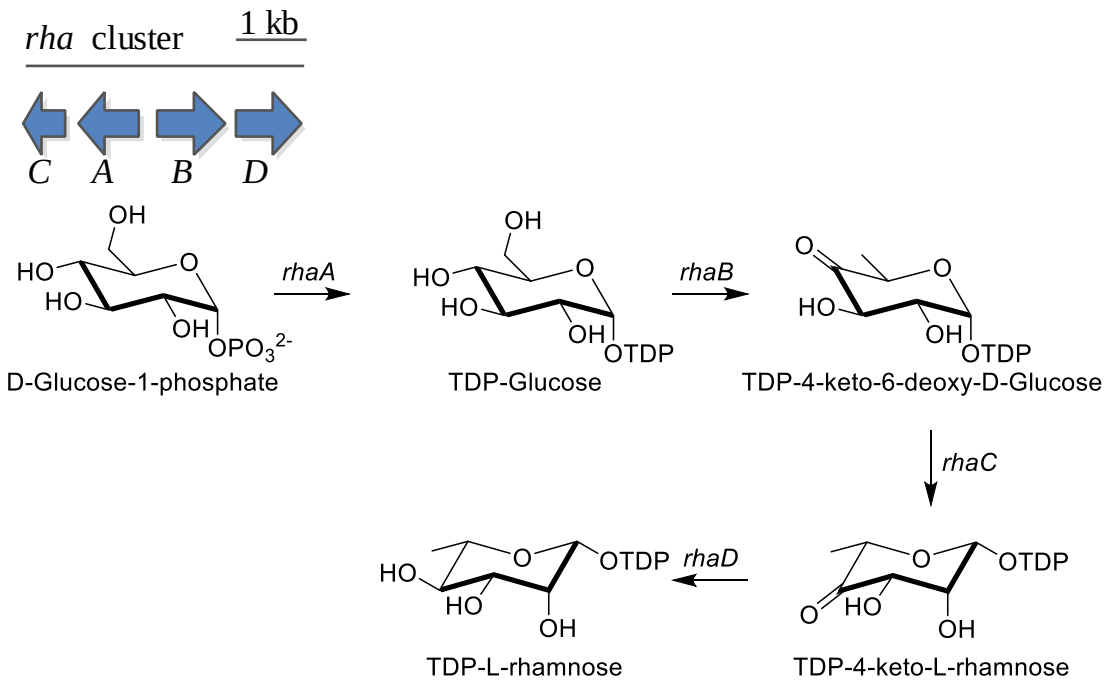


Figure S4. Schematic view of *rha* gene cluster of *S.* sp. KIB-H483 and proposed biosynthetic pathway for L-rhamnose.

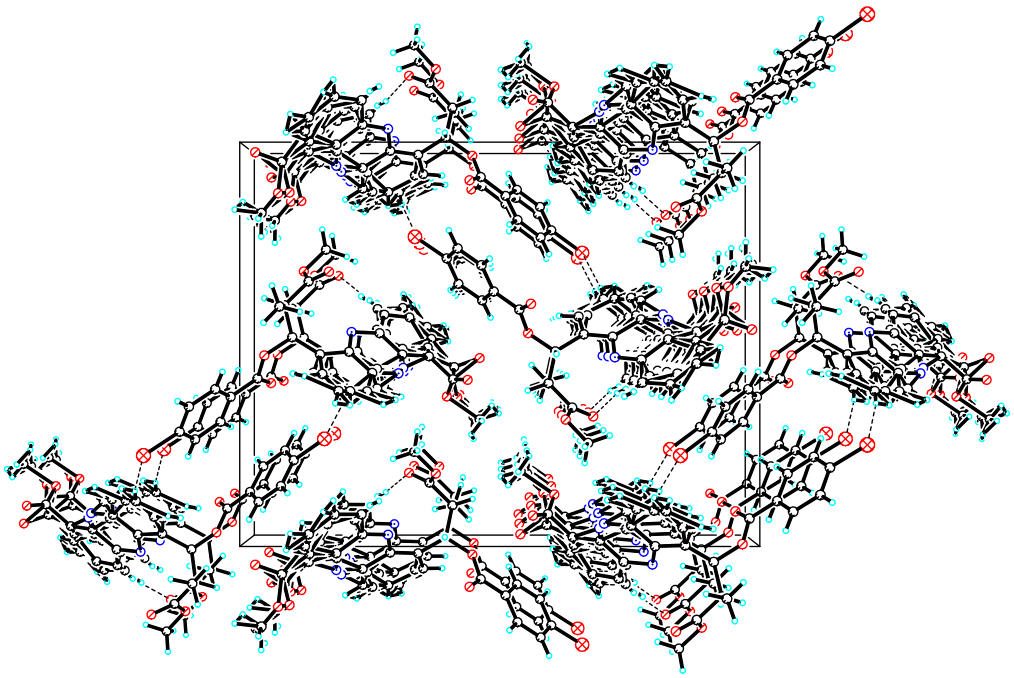


Figure S5. X-ray packing diagram of **1a** (carbon, gray; oxygen, red; nitrogen, blue; hydrogen, cyan). Hydrogen-bonds are shown as dashed lines.

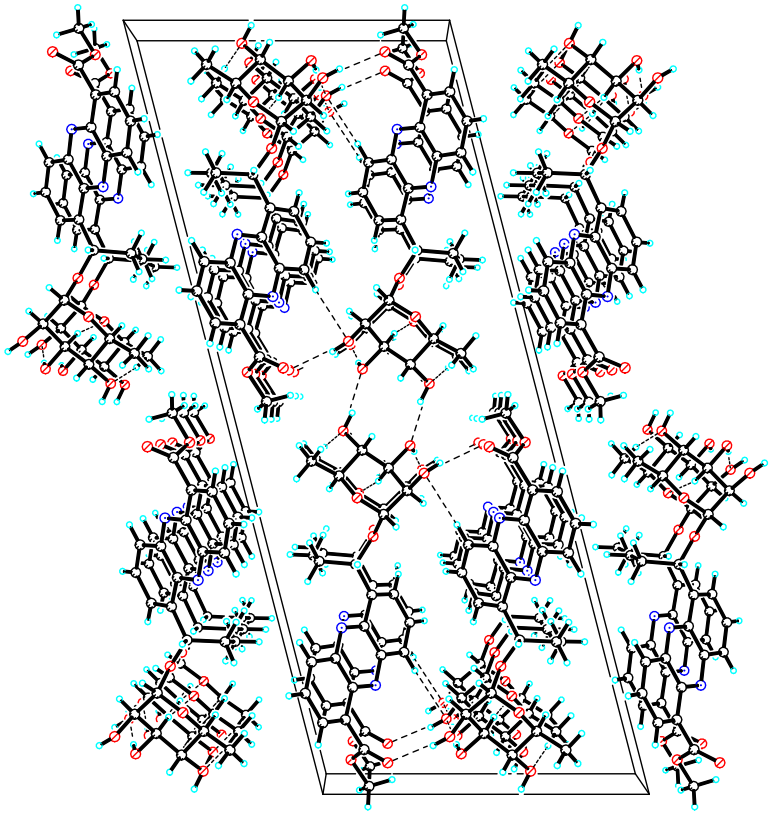


Figure S6. X-ray packing diagram of **6** (carbon, gray; oxygen, red; nitrogen, blue; hydrogen, cyan). Hydrogen-bonds are shown as dashed lines.

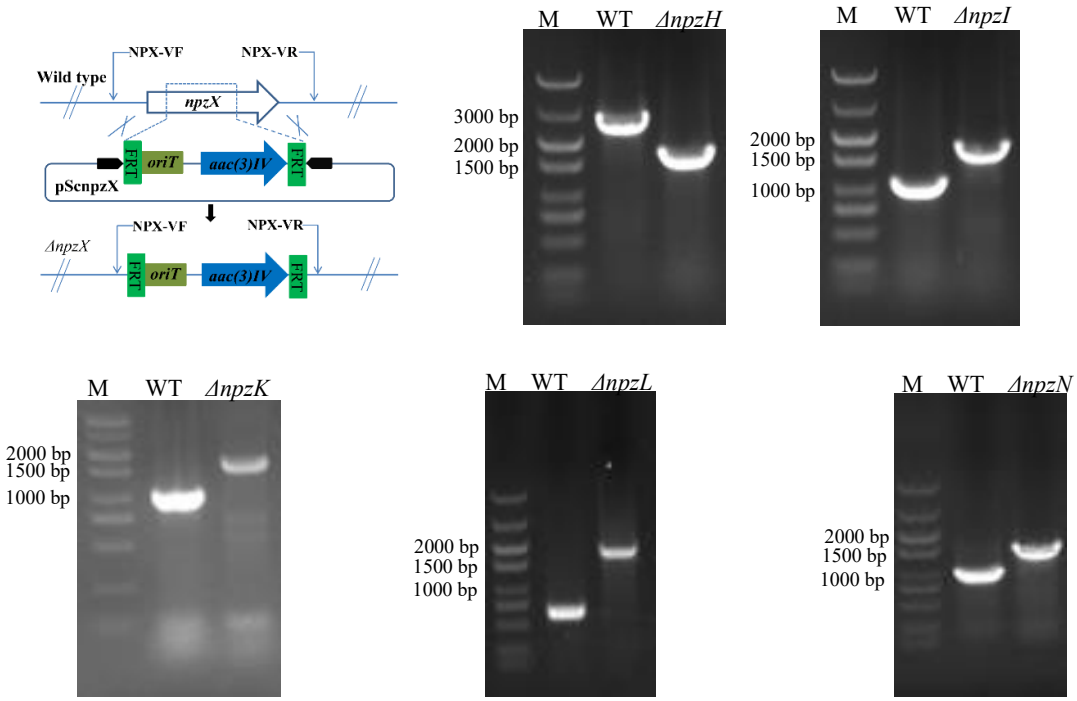


Figure S7. Construction and agarose gel electrophoresis analysis of mutants *ΔnpzH*, *ΔnpzI*, *ΔnpzK*, *ΔnpzL* and *ΔnpzN*. WT, Wild type; M, DNA marker.

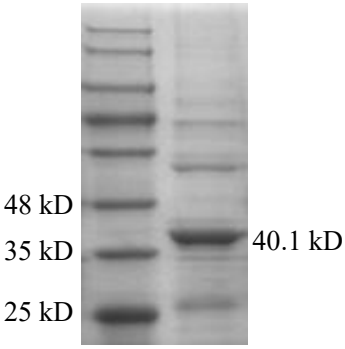


Figure S8. SDS-PAGE analysis of His-tagged NpzI protein from *E. coli* BL21(DE3)-pET26b-*npzI* using nickel-NTA column.

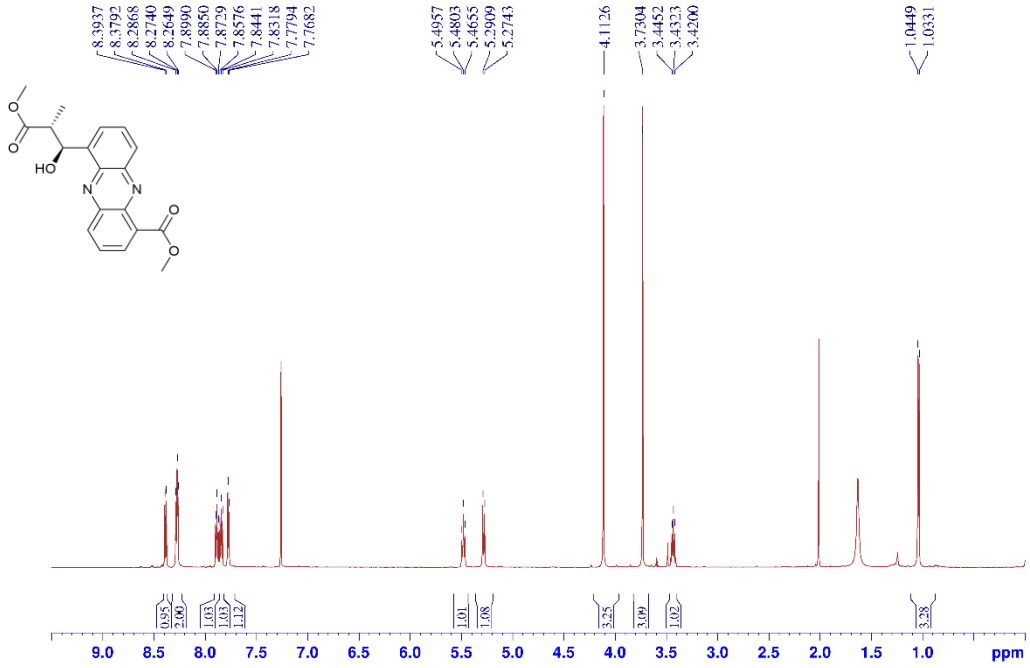


Figure S9. ^{1}H NMR spectrum of **1** (600 MHz, $CDCl_3$).

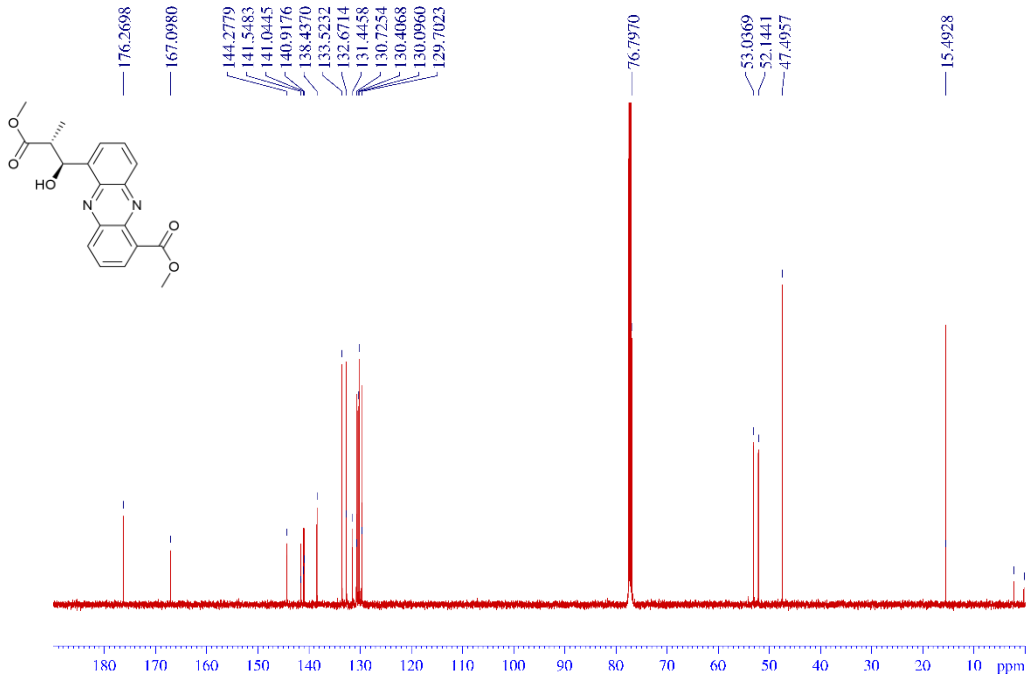


Figure S10. ^{13}C NMR spectrum of **1** (150 MHz, $CDCl_3$).

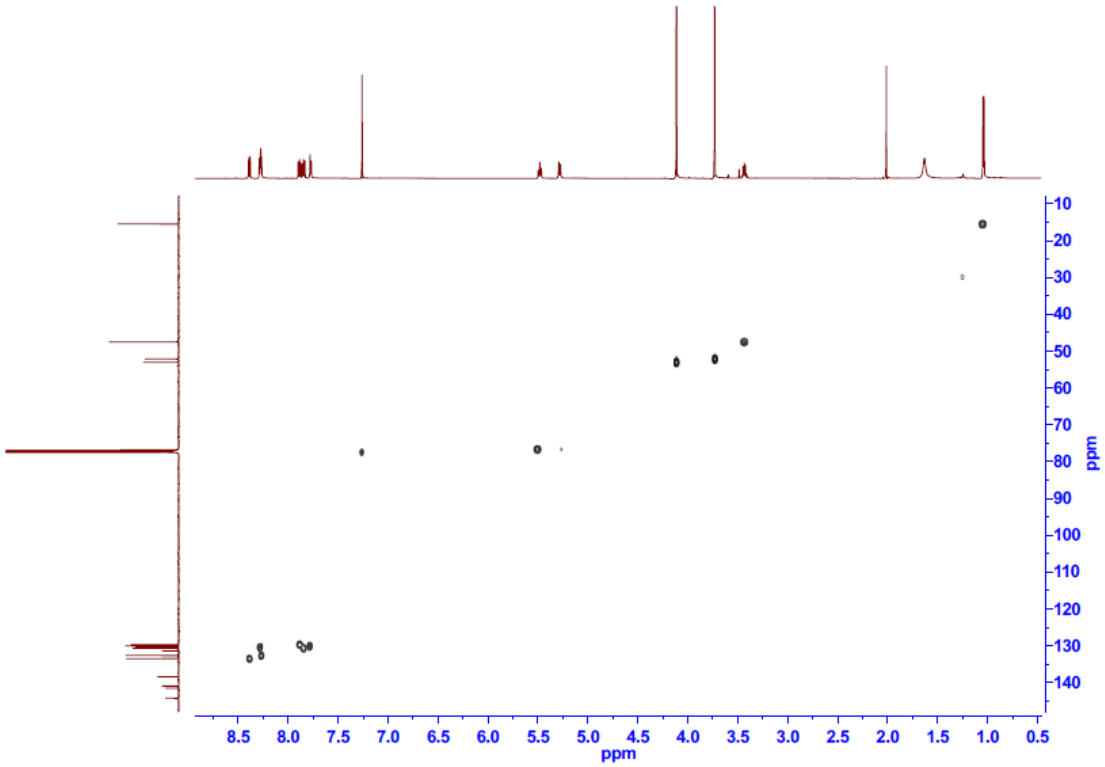


Figure S11. HSQC spectrum of **1** (600 MHz, $CDCl_3$).

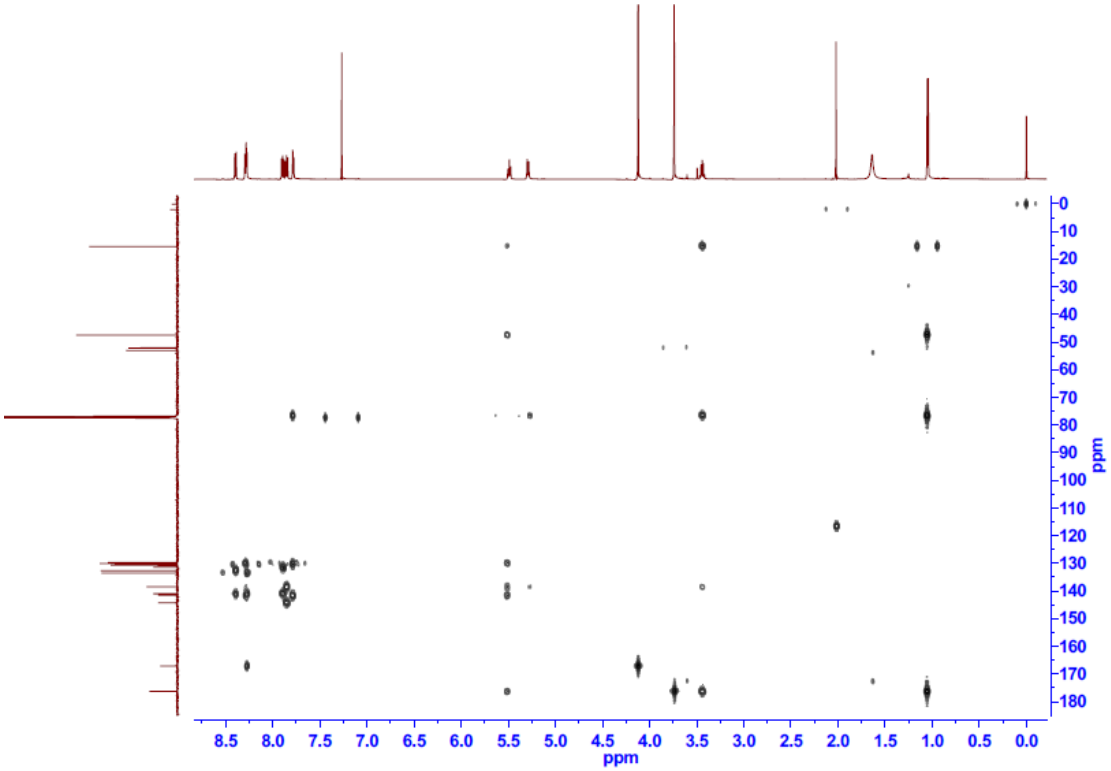


Figure S12. HMBC spectrum of **1** (600 MHz, $CDCl_3$).

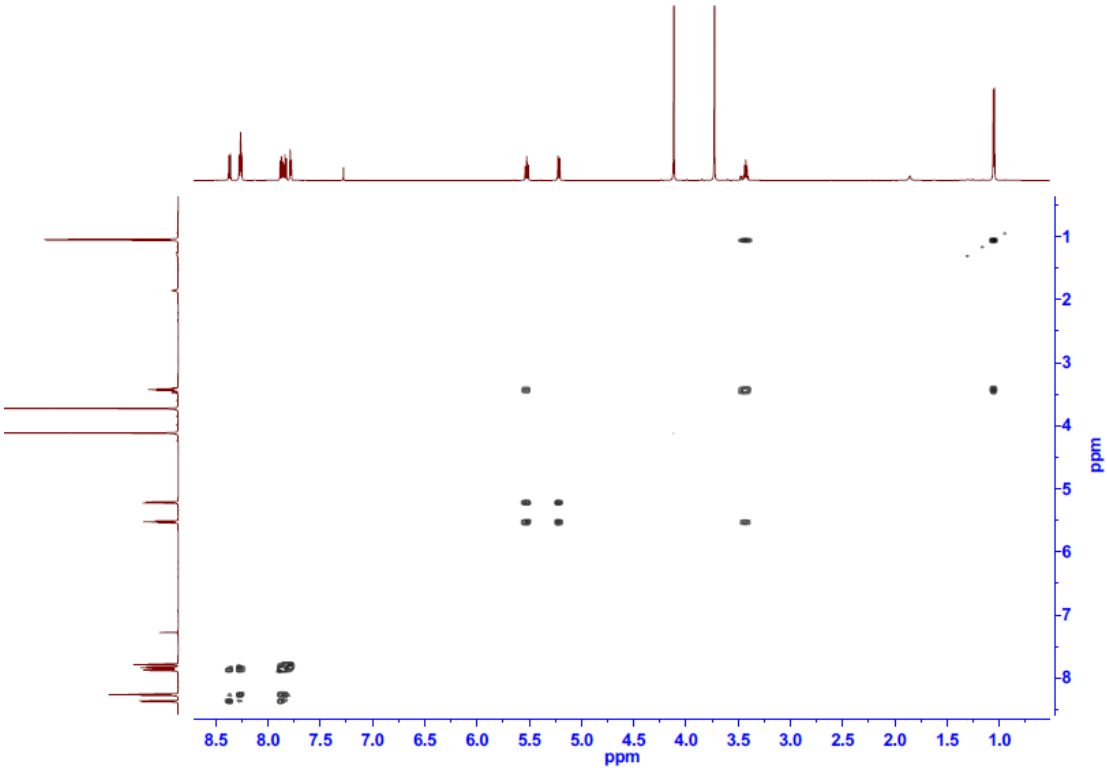


Figure S13. COSY spectrum of **1** (600 MHz, $CDCl_3$).

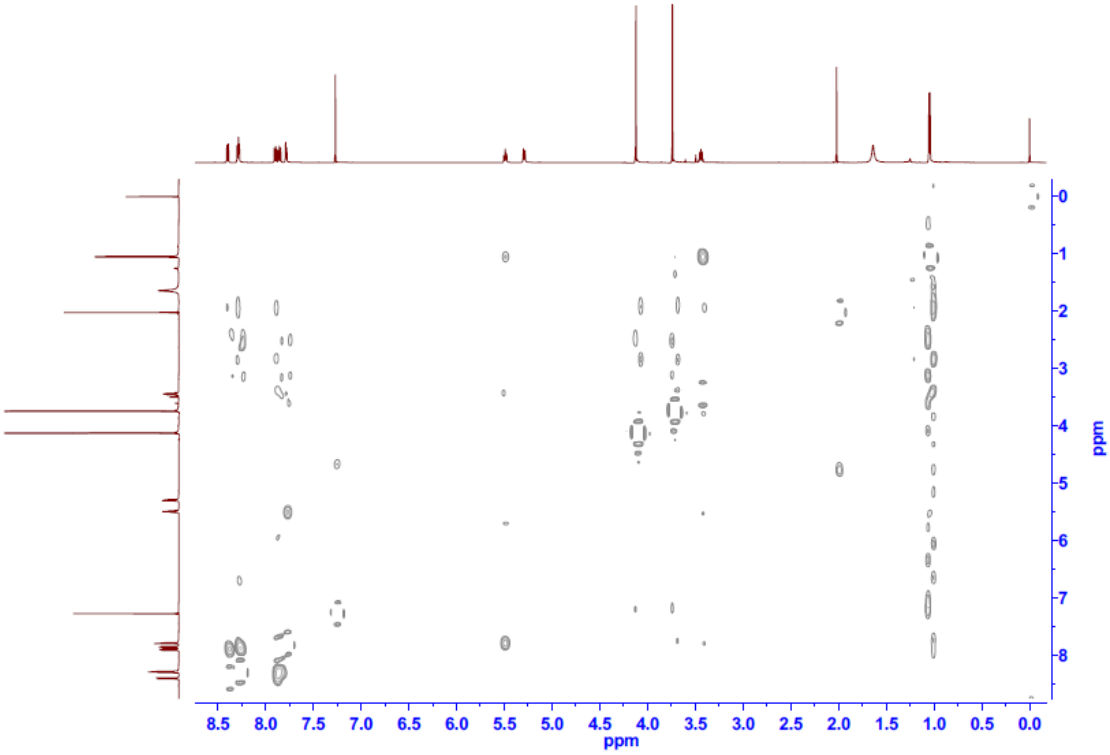


Figure S14. ROESY spectrum of **1** (600 MHz, $CDCl_3$).

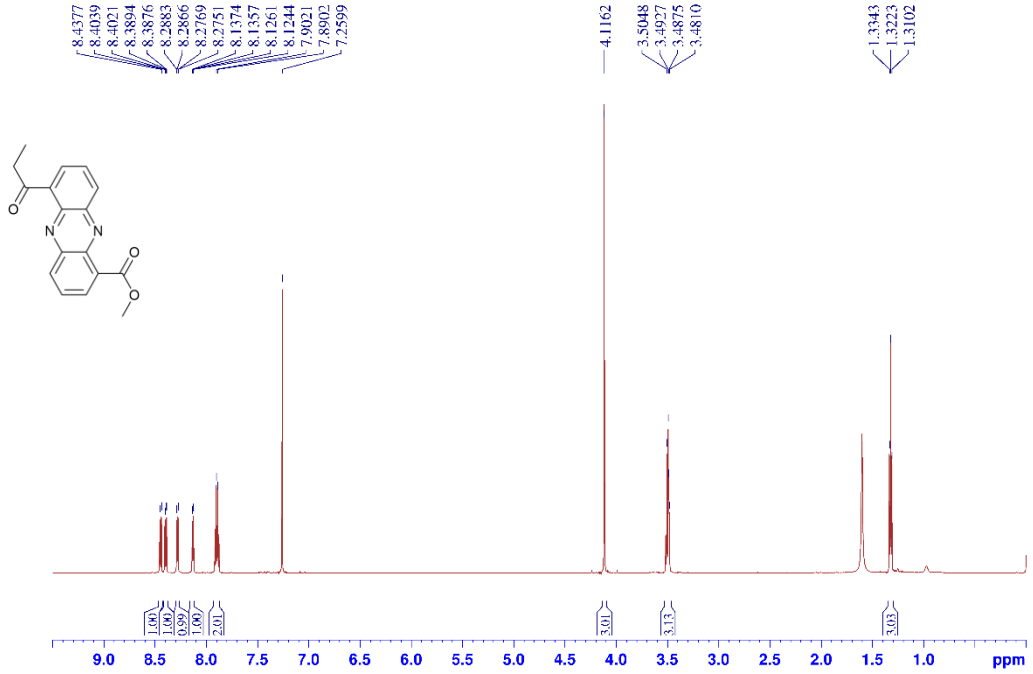


Figure S15. ^{1}H NMR spectrum of **4** (600 MHz, $CDCl_3$).

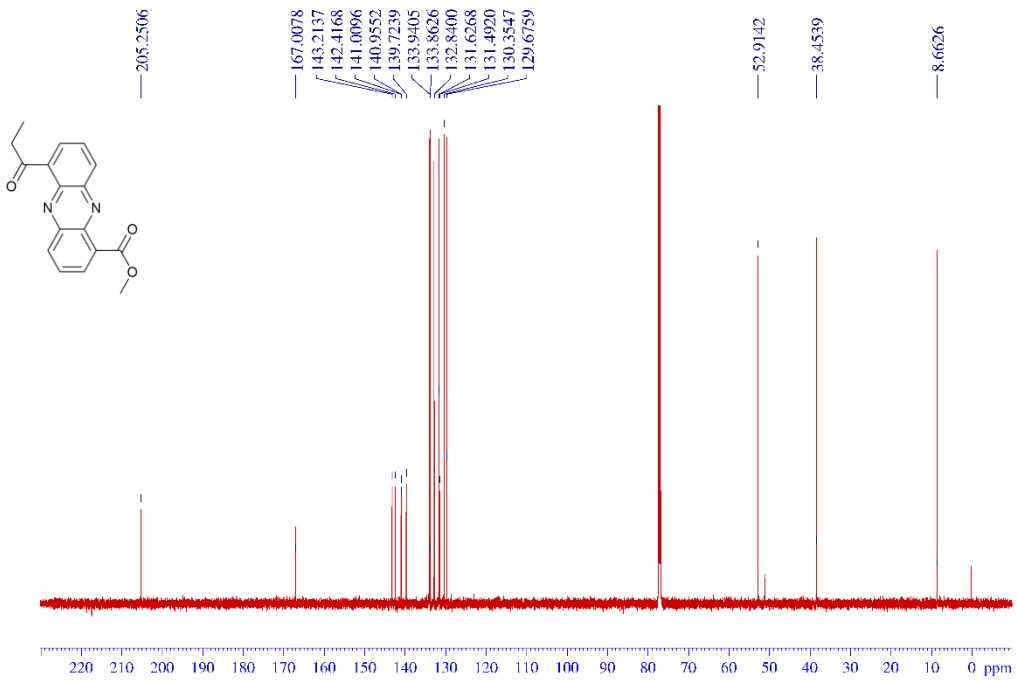


Figure S16. ^{13}C NMR spectrum of **4** (150 MHz, $CDCl_3$).

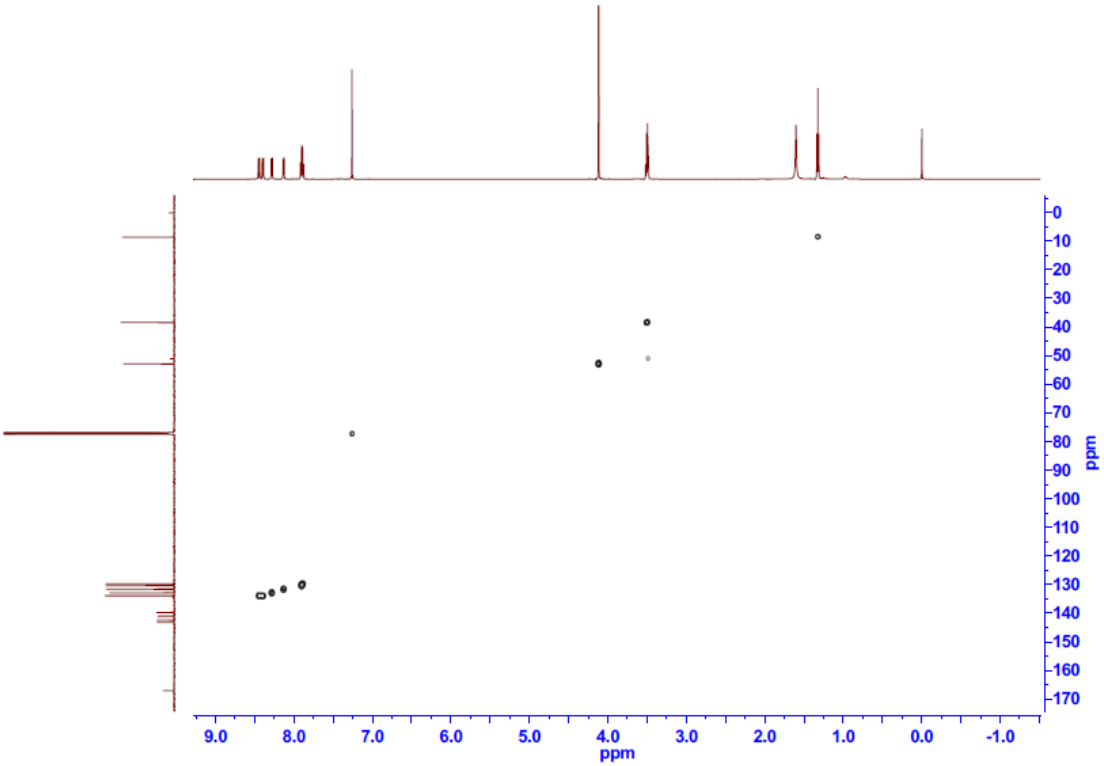


Figure S17. HSQC spectrum of **4** (600 MHz, $CDCl_3$).

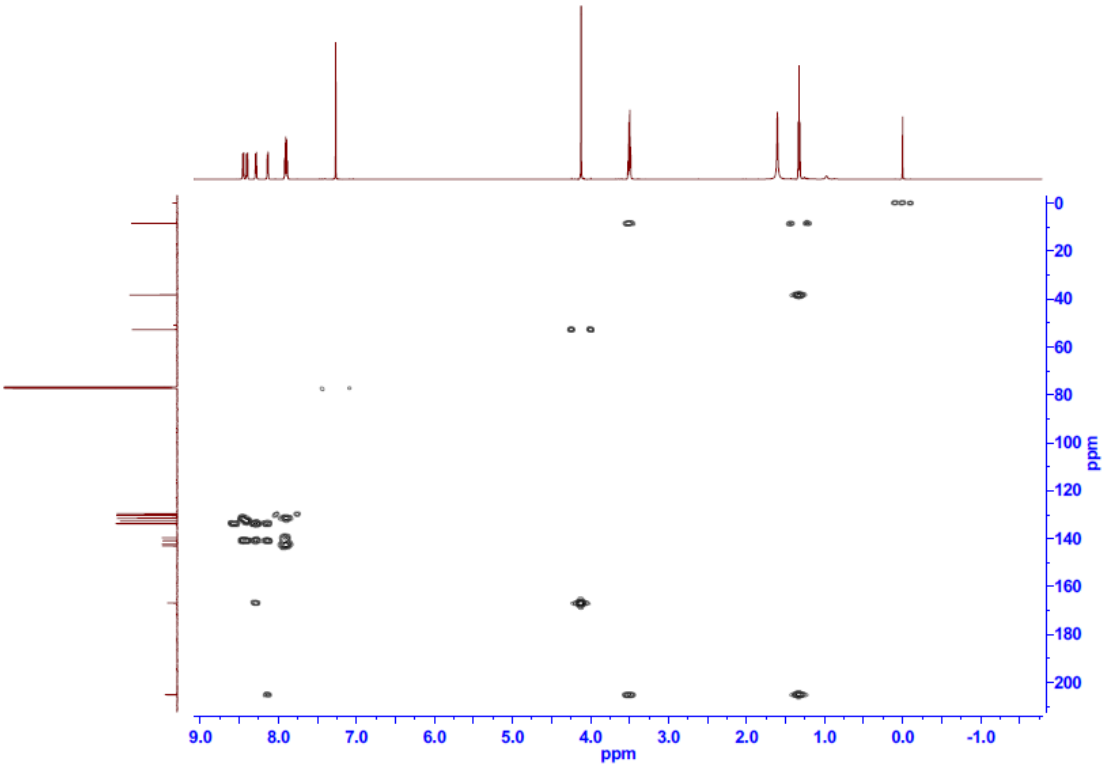


Figure S18. HMBC spectrum of **4** (600 MHz, $CDCl_3$).

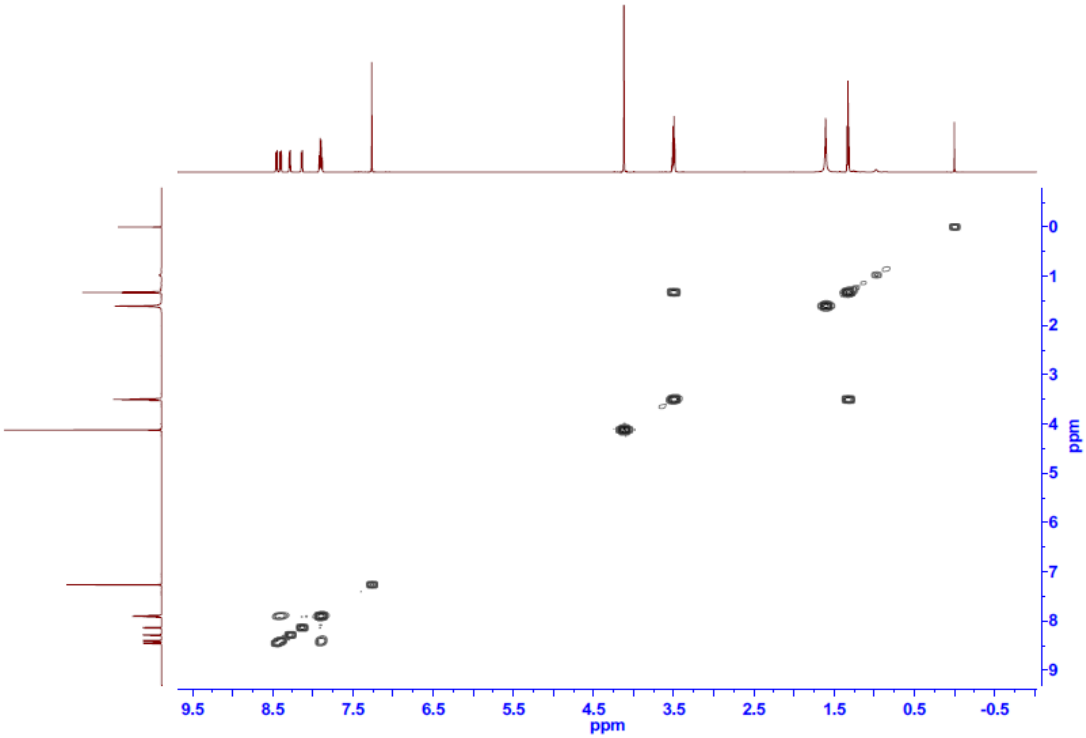


Figure S19. COSY spectrum of **4** (600 MHz, $CDCl_3$).

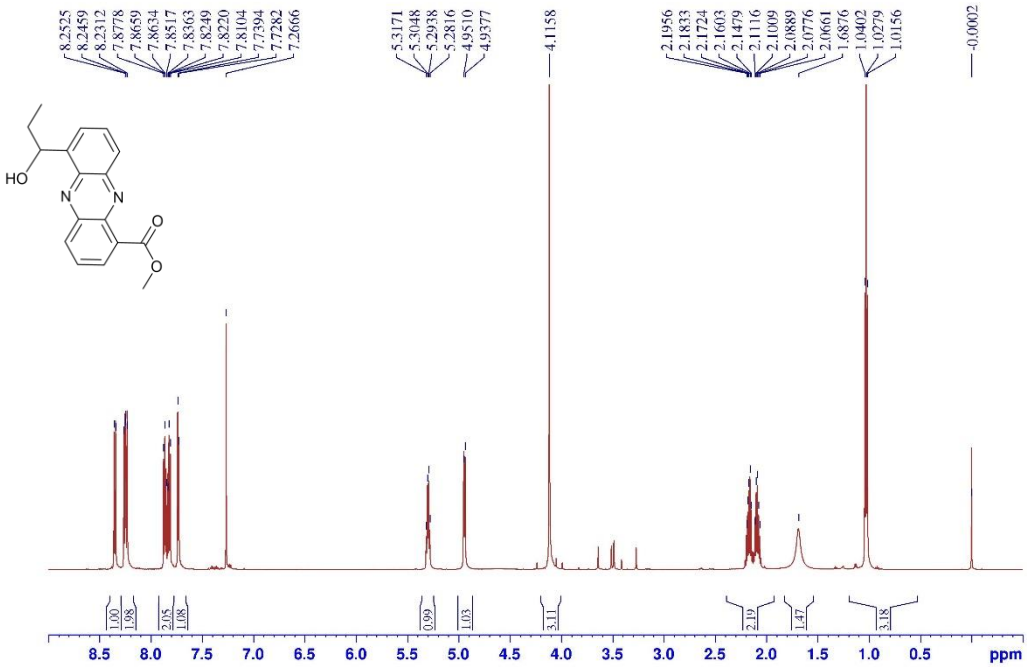


Figure S20. ^{1}H NMR spectrum of **5** (600 MHz, $CDCl_3$).

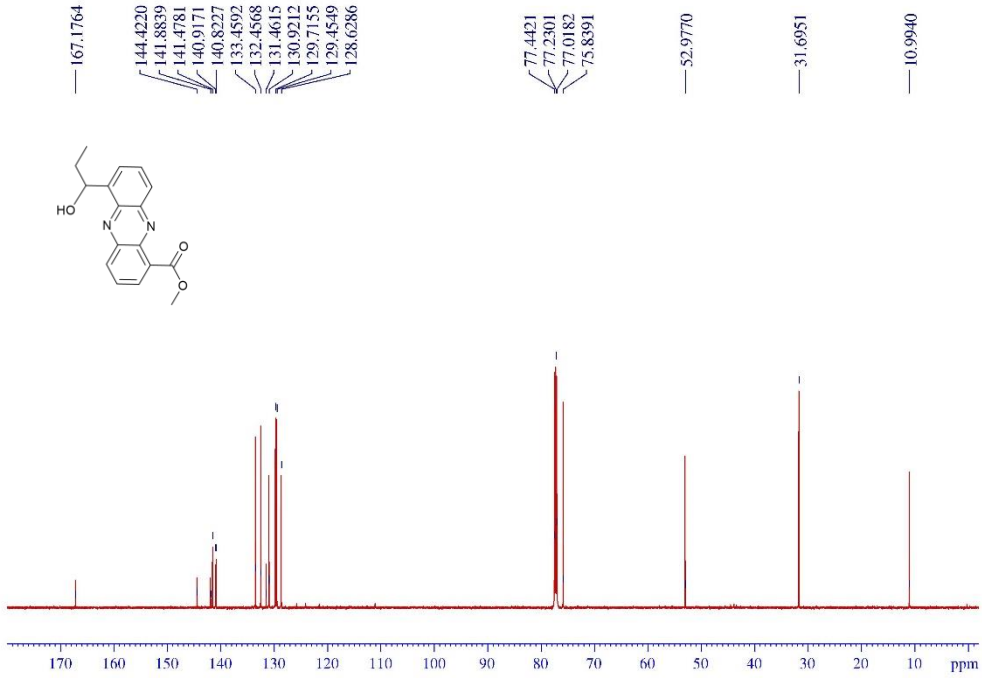


Figure S21. ^{13}C NMR spectrum of **5** (150 MHz, $CDCl_3$).

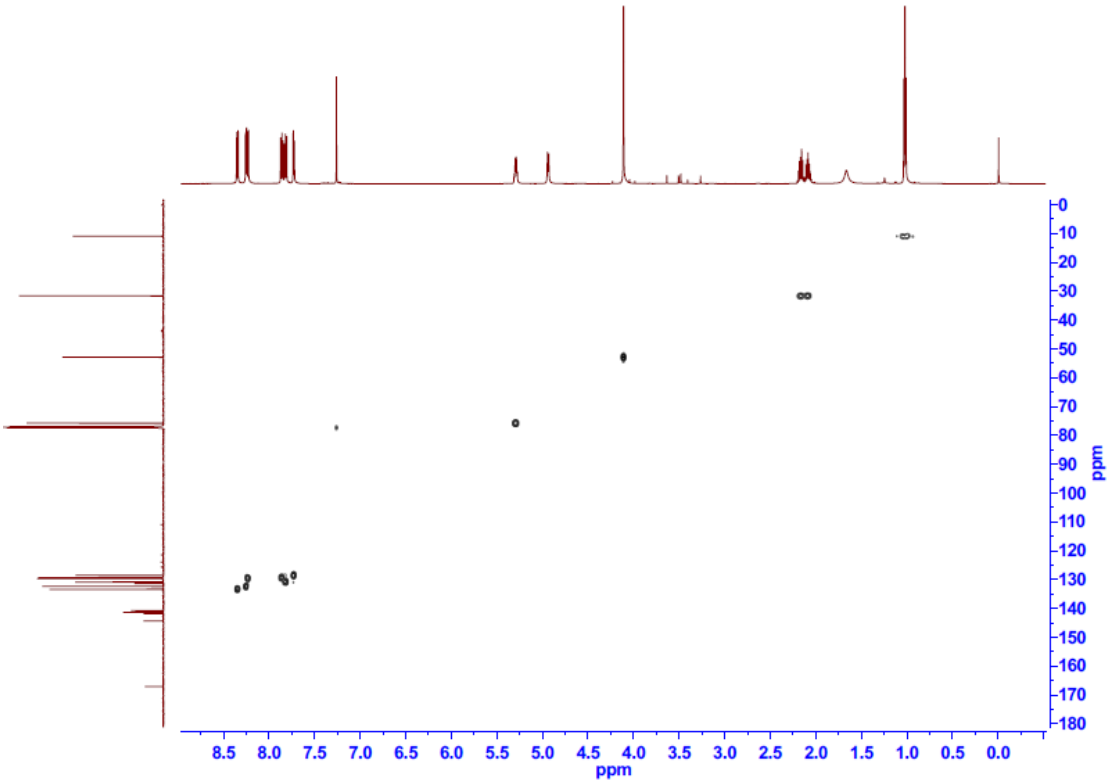


Figure S22. HSQC spectrum of **5** (600 MHz, $CDCl_3$).

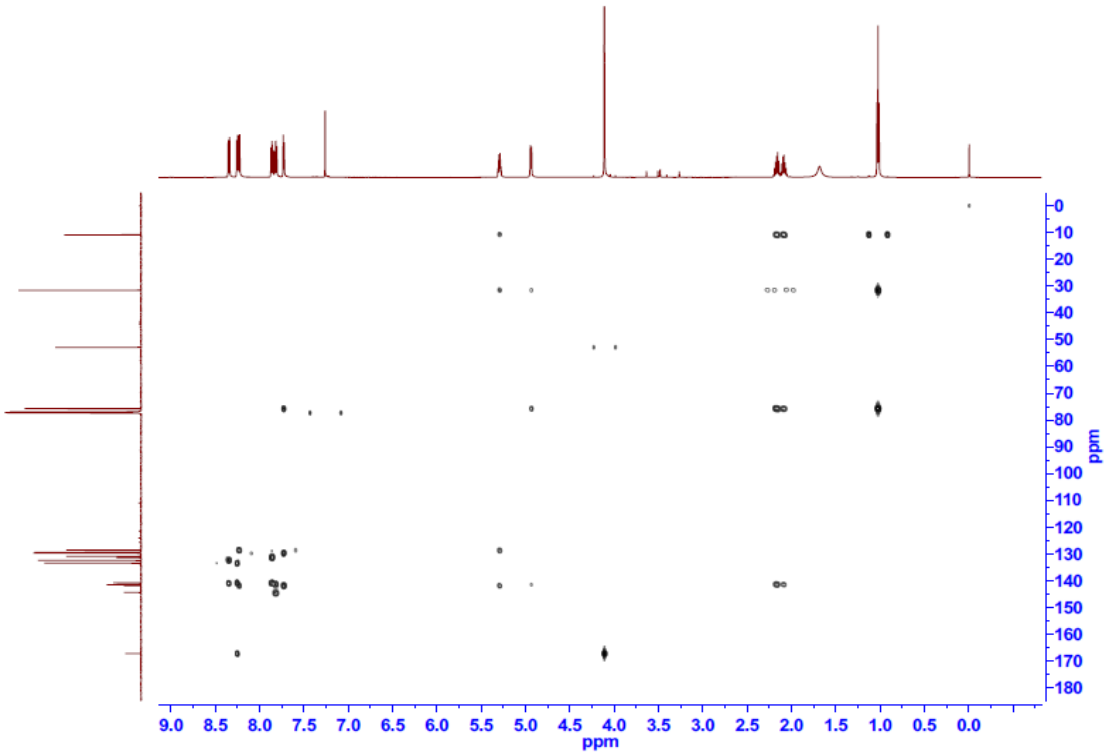


Figure S23. HMBC spectrum of **5** (600 MHz, $CDCl_3$).

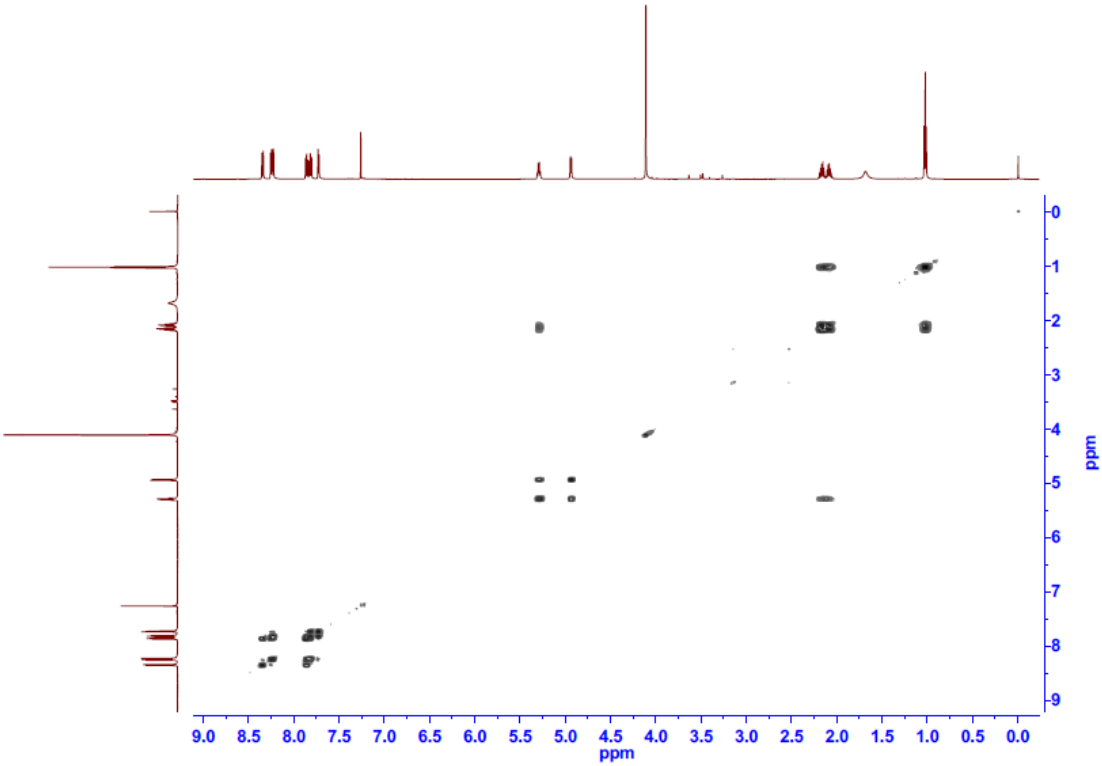


Figure S24. COSY spectrum of **5** (600 MHz, $CDCl_3$).

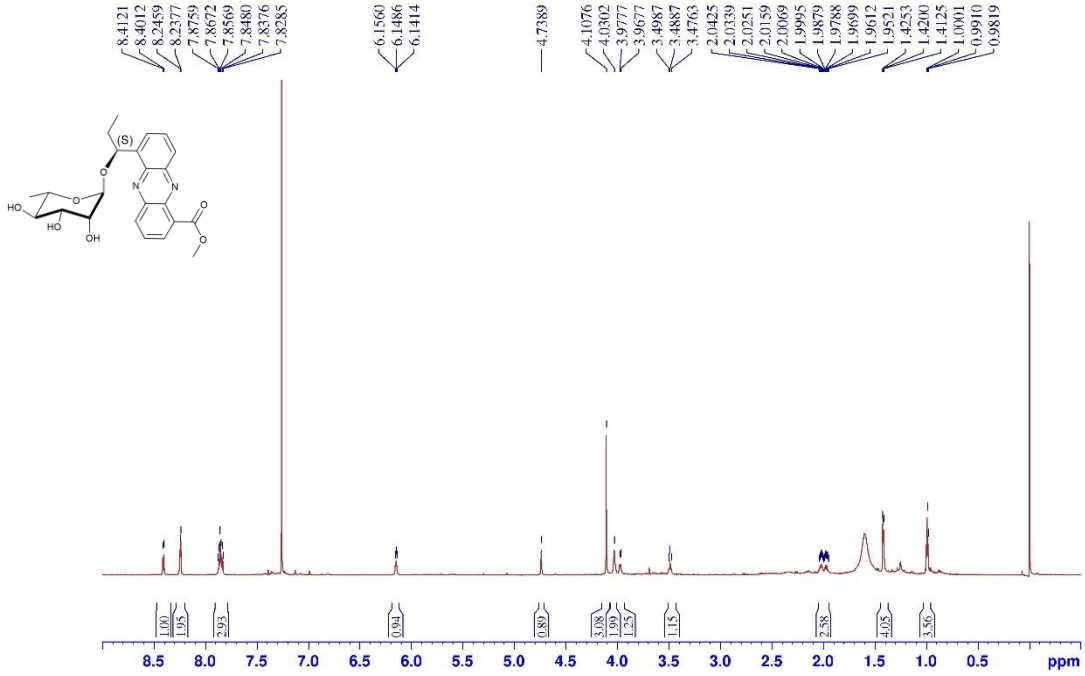


Figure S25. ^{1}H spectrum of **6** (800 MHz, $CDCl_3$).

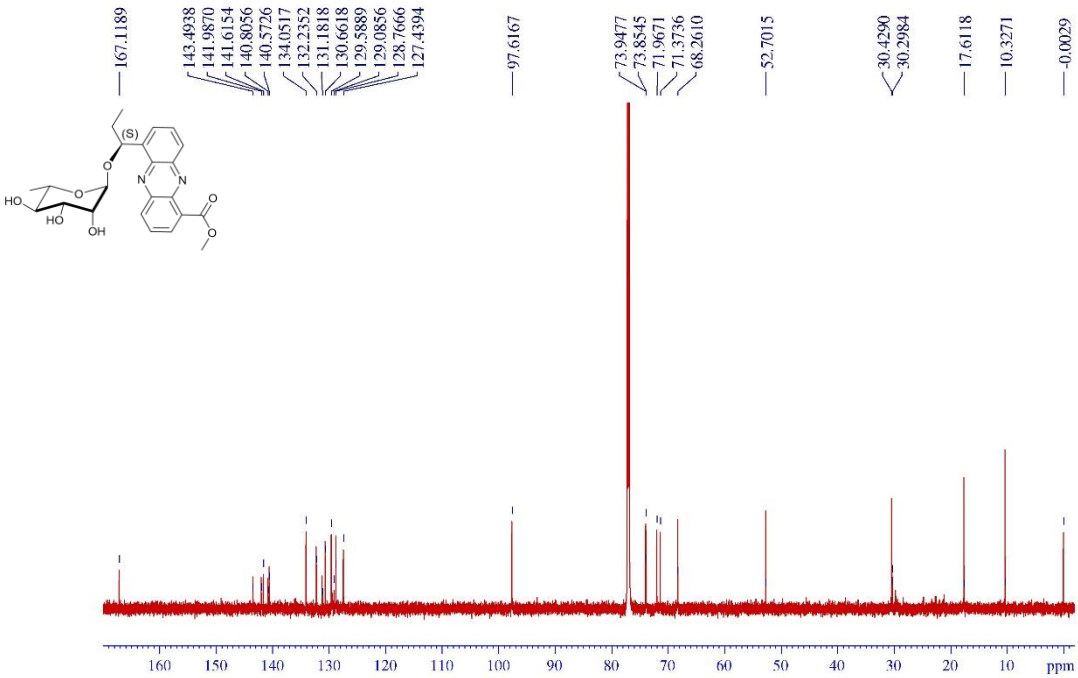


Figure S26. ^{13}C spectrum of **6** (200 MHz, $CDCl_3$).

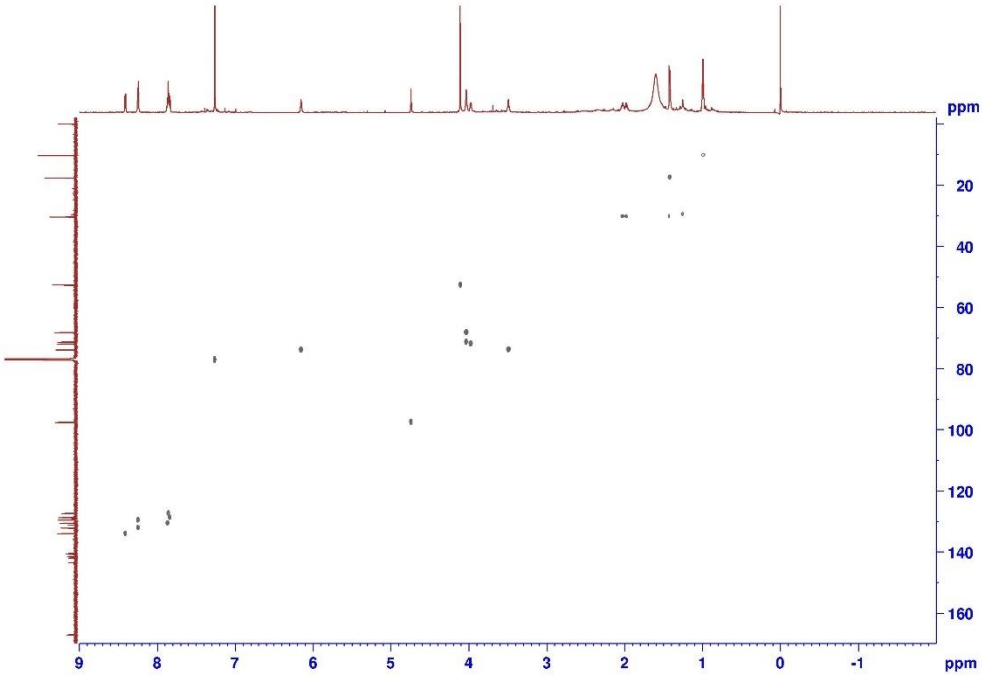


Figure S27. HSQC spectrum of **6** (800 MHz, $CDCl_3$).

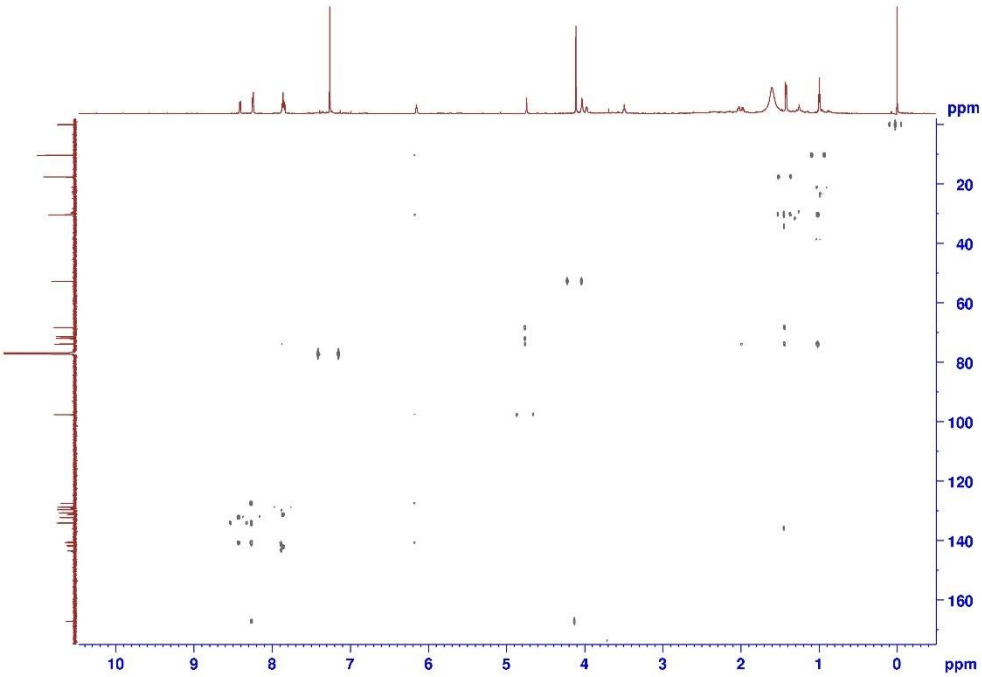


Figure S28. HMBC spectrum of **6** (800 MHz, $CDCl_3$).

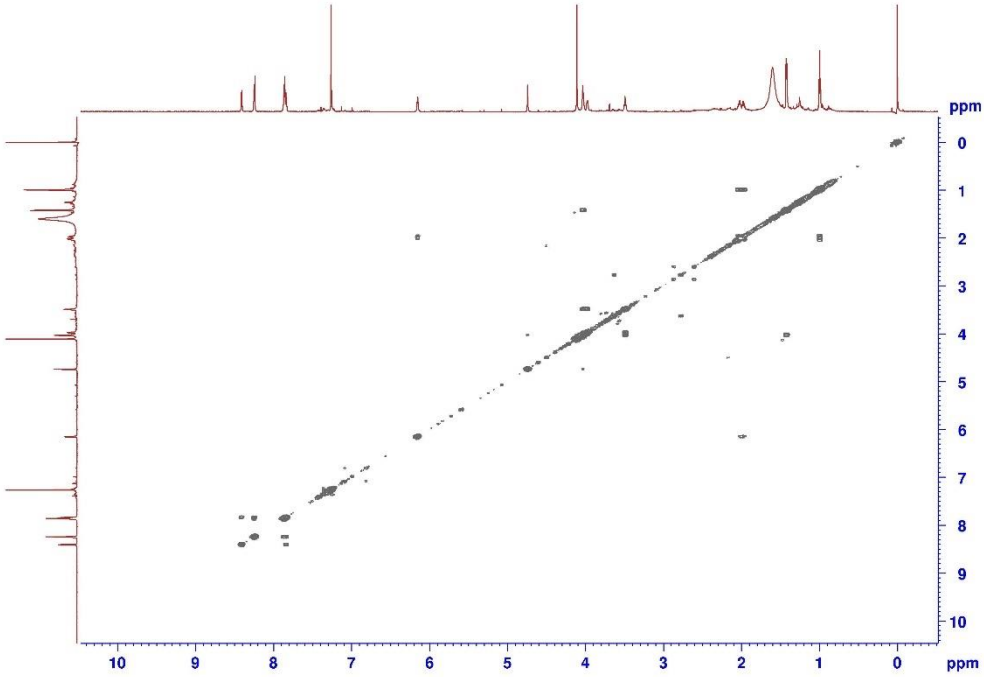


Figure S29. COSY spectrum of **6** (800 MHz, $CDCl_3$).

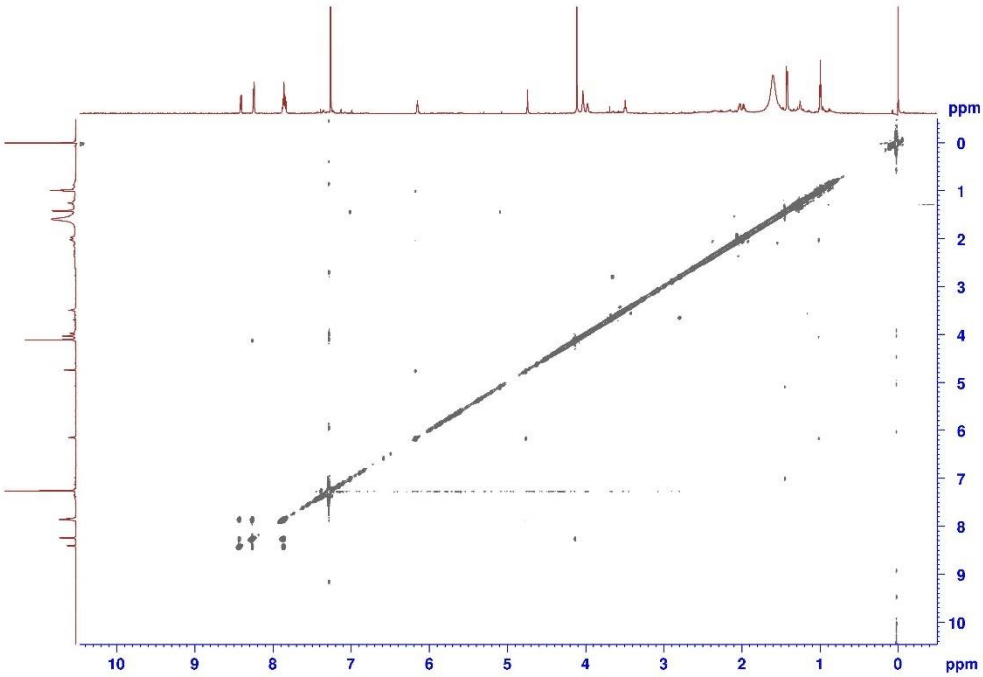


Figure S30. ROESY spectrum of **6** (800 MHz, $CDCl_3$).

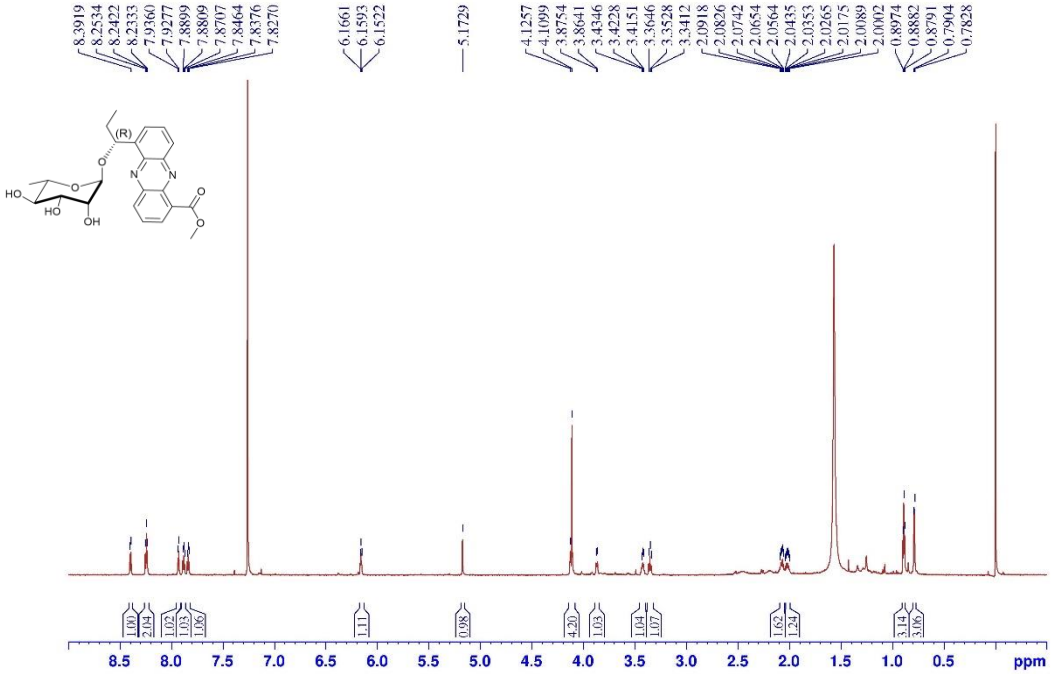


Figure S31. ^{1}H spectrum of **7** (800 MHz, $CDCl_3$).

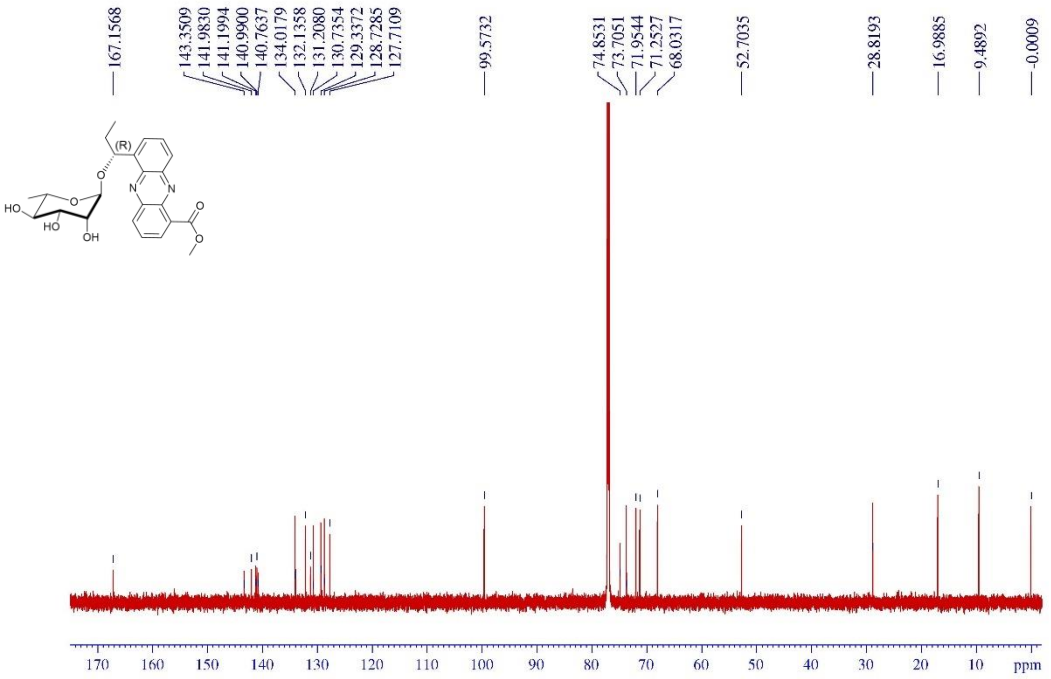


Figure S32. ^{13}C spectrum of **7** (200 MHz, $CDCl_3$).

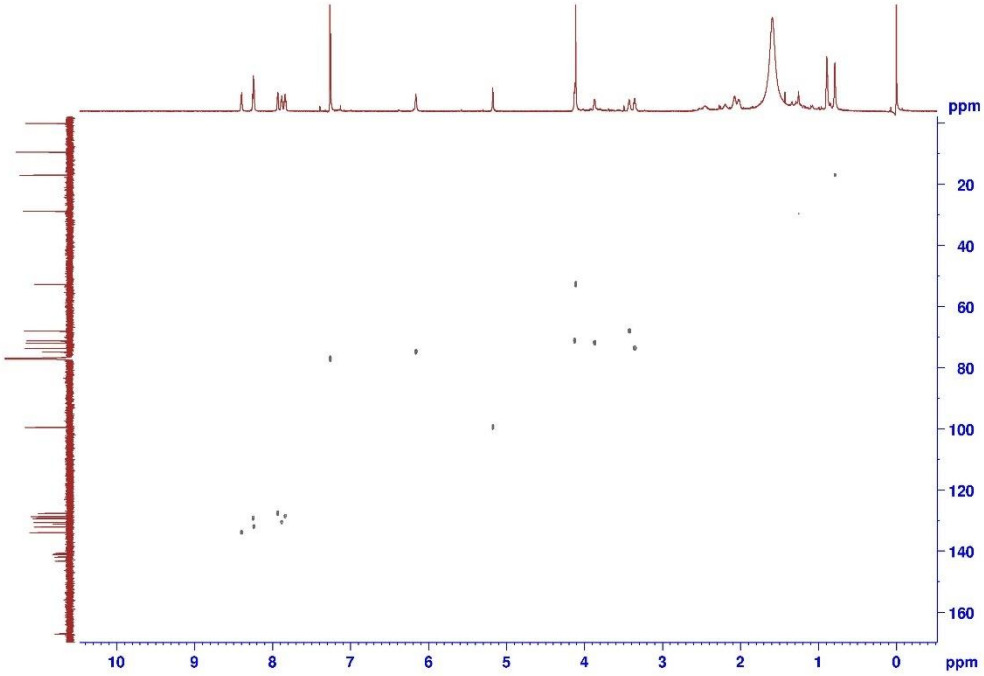


Figure S33. HSQC spectrum of **7** (800 MHz, $CDCl_3$).

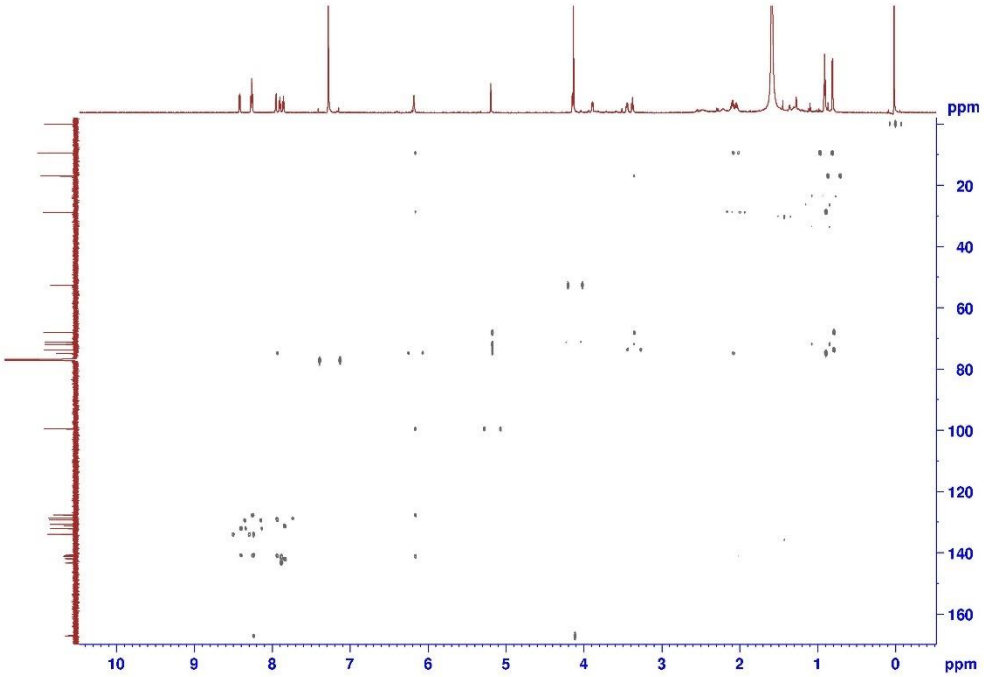


Figure S34. HMBC spectrum of **7** (800 MHz, $CDCl_3$).

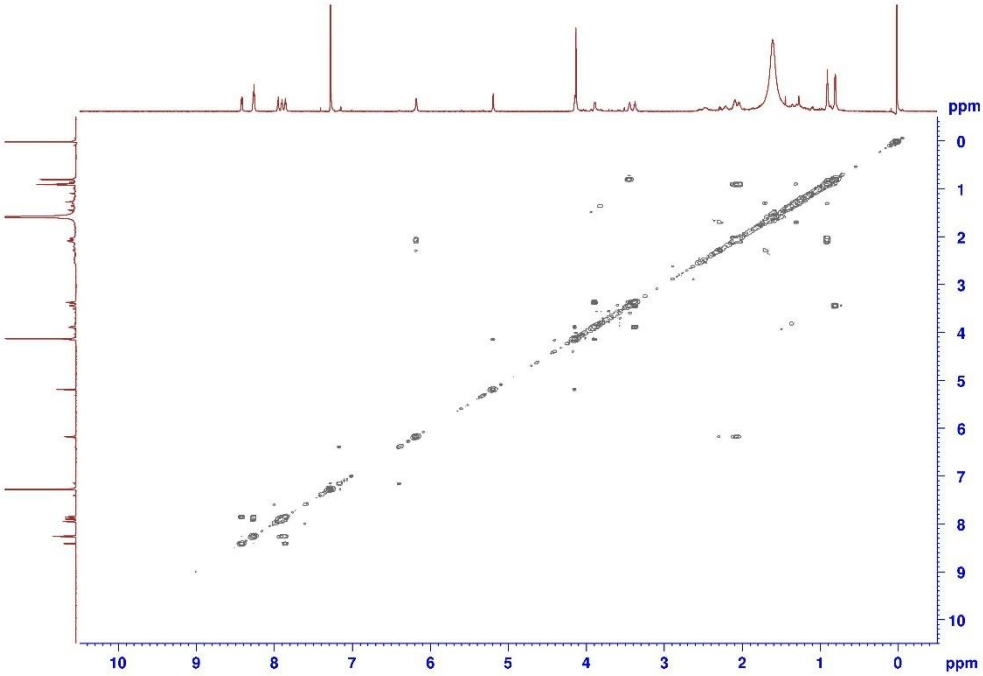


Figure S35. COSY spectrum of **7** (800 MHz, $CDCl_3$).

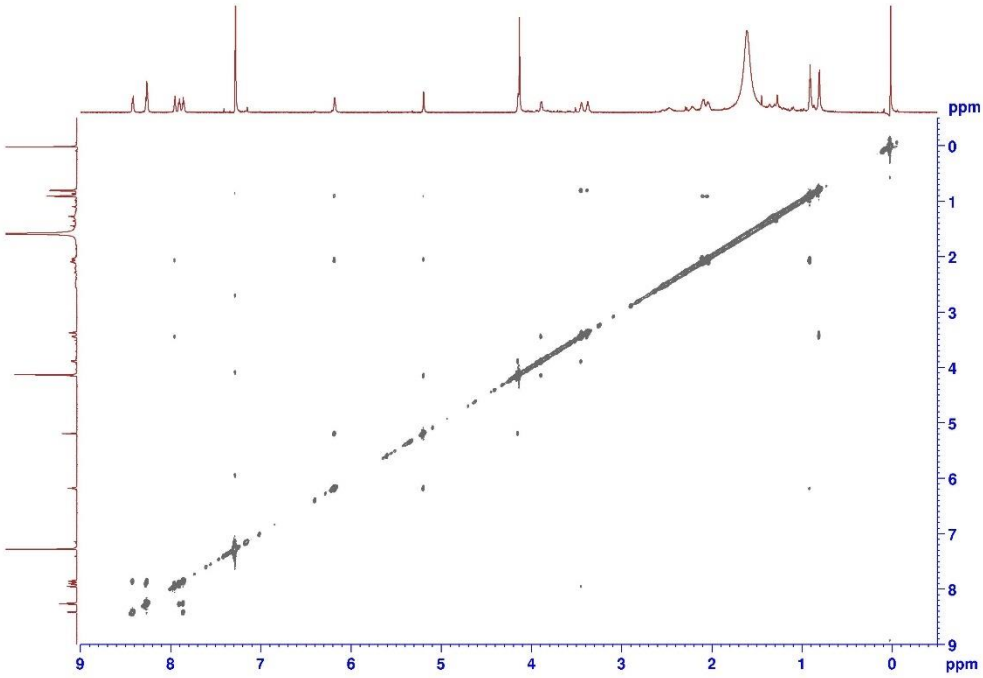


Figure S36. ROESY spectrum of **7** (800 MHz, $CDCl_3$).

REFERENCES

(1) Breitmaier, E.; Hollstein, U. *J. Org. Chem.* **1976**, *41*, 2104-2108.

(2) Pusecker, K.; Laatsch, H.; Helmke, E.; Weyland, H. *J. Antibiot.* **1997**, *50*, 479-483.

(3) Robert, X.; Gouet, P. *Nucleic Acids Res.* **2014**, *42*, 320-324.