Molecular networking-based for the target discovery of potent antiproliferative polycyclic macrolactam ansamycins from *Streptomyces cacaoi* subsp. *asoensis*

Ling-Li Liu,[a] Zhi-Fan Chen,[a] Yao Liu,[a] Dan Tang,[a] Hua-Hua Gao,[a] Qiang Zhang,[a] Jin-Ming Gao*[a, b]

[a] Shaanxi Key Laboratory of Natural Products & Chemical Biology, College of Chemistry & Pharmacy, Northwest A&F University, 22 Xinong Road, Yangling 712100, Shaanxi, People's Republic of China

[b] State Key Laboratory of Phytochemistry and Plant Resources in West China, Kunming 650201, Yunnan, People's Republic of China

Index

Table S1. Possible clusters in *S. cacaoi* subsp. *asoensis* H2S5 genome predicted by antiSMASH

Cluster	Type	From	To	Most similar known cluster	MIBiG BGC-ID
Cluster 1	T3PKS-NRPS-butyrolactone-Phenazine	45632	153940	Merochlorin_biosynthetic_gene_cluster (70% of genes show similarity)	BGC0001083_c1
Cluster 2	T3PKS	202755	243813	Merochlorin_biosynthetic_gene_cluster (4% of genes show similarity)	BGC0001083_c1
Cluster 3	Amglyccycl	482080	503333	Cetoniacytone_A_biosynthetic_gene_cluster (9% of genes show similarity)	BGC0000283_c1
Cluster 4	NRPS	762110	808516	Paenibactin_biosynthetic_gene_cluster (83% of genes show similarity)	BGC0000401_c1
Cluster 5	Melanin-Terpene	880287	906451	Melanin_biosynthetic_gene_cluster (71% of genes show similarity)	BGC0000908_c1
Cluster 6	T1pks	1275409	1366059	Ansatrienin_(mycotrienin)_biosynthetic_gene_ cluster (57% of genes show similarity)	BGC0000957_c1
Cluster 7	T1pks	1495101	1539207	Maduropeptin_biosynthetic_gene_cluster (3% of genes show similarity)	BGC0001008_c1
Cluster 8	T3pks	1658764	1700026	-	-
Cluster 9	T2pks	1744483	1786998	Spore_pigment_biosynthetic_gene_cluster (83% of genes show similarity)	BGC0000271_c1
Cluster 10	Other	2331721	2375587	-	-
Cluster 11	Ectoine	2588604	2599008	Ectoine_biosynthetic_gene_cluster (100% of genes show similarity)	BGC0000853_c1
Cluster 12	NRPS	3326856	3386298	SCO-2138_biosynthetic_gene_cluster (78% of genes show similarity)	BGC0000595_c1
Cluster 13	Melanin	3798393	3808875	Lactonamycin_biosynthetic_gene_cluster (8% of genes show similarity)	BGC0000238_c1
Cluster 14	Siderophore	3916189	3927958	Desferrioxamine_B_biosynthetic_gene_cluster (100% of genes show similarity)	BGC0000940_c1
Cluster 15	Bacteriocin-Lantipeptide	6472850	6500974	Meilingmycin_biosynthetic_gene_cluster (2% of genes show similarity)	BGC0000093_c1
Cluster 16	Terpene	6614008	6635105	Albaflavenone_biosynthetic_gene_cluster (100% of genes show similarity)	BGC0000660_c1
Cluster 17	Siderophore	7350730	7362670	ND	-
Cluster 18	Bacteriocin	7612594	7624075	ND	-

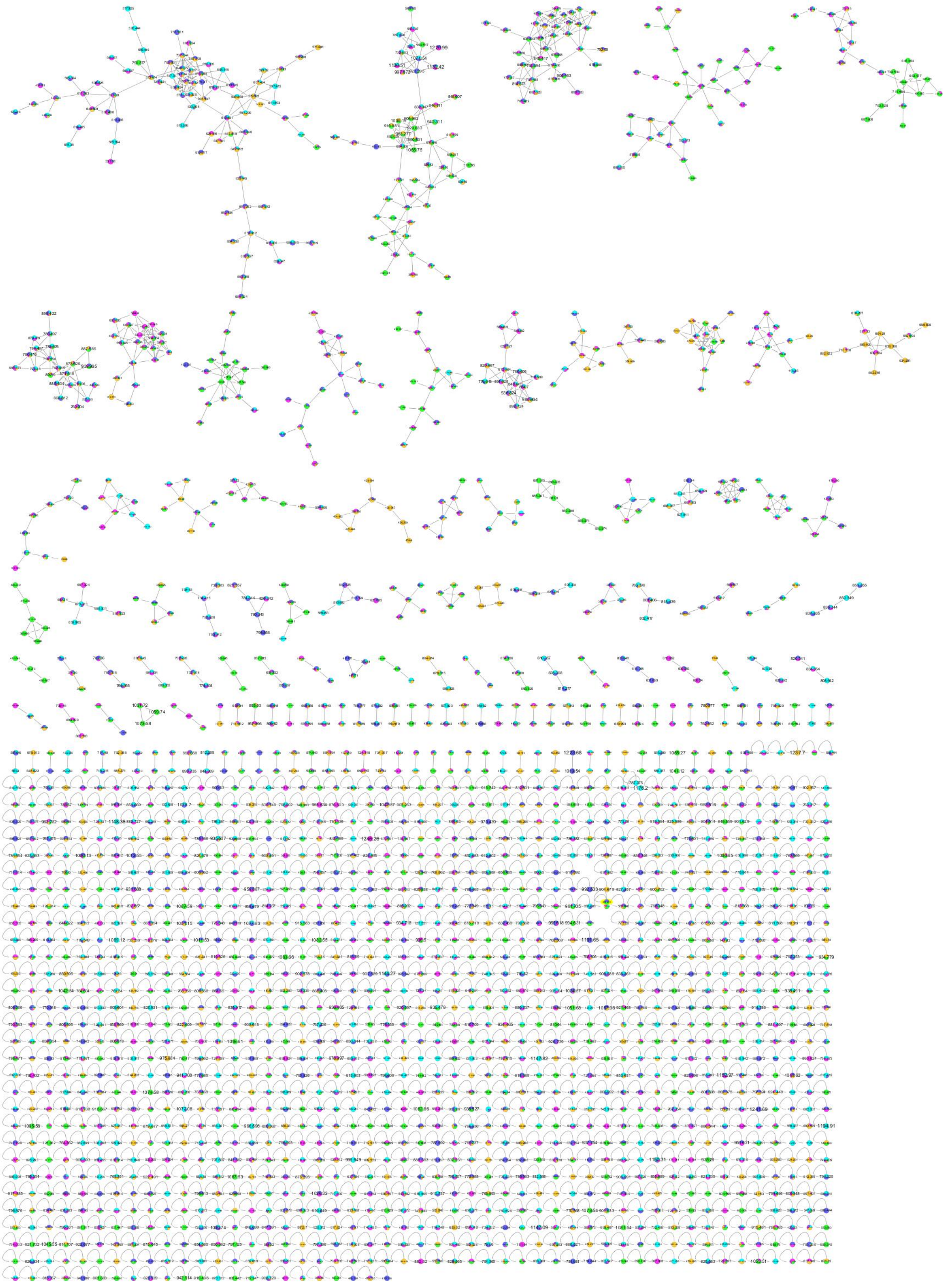
Cluster	**Type**	**From**	**To**	**Most similar known cluster**	**MIBiG BGC-ID**
Cluster 19	Terpene	7663709	7685928	ND	-
Cluster 20	NRPS-Siderophore	7859768	7933278	Scabichelin_biosynthetic_gene_cluster (100% of genes show similarity)	BGC0000423_c1
Cluster 21	Indole	8149476	8170717	Meridamycin_biosynthetic_gene_cluster (5% of genes show similarity)	BGC0001011_c1
Cluster 22	Nucleoside-NRPS	8214006	8269095	Sch47554_/_Sch47555_biosynthetic_gene_cluster (32% of genes show similarity)	BGC0000268_c1
Cluster 23	Betalactone	8340383	8364792	-	-
Cluster 24	Terpene	8451563	8478269	Hopene_biosynthetic_gene_cluster (92% of genes show similarity)	BGC0000663_c1
Cluster 25	Nucleoside-Aminocoumarin-T1PKS-Others-NRPS	8550555	8678658	Rubradirin_biosynthetic_gene_cluster (37% of genes show similarity)	BGC0000141_c1
Cluster 26	Other	8807814	8851755	Herboxidiene_biosynthetic_gene_cluster (2% of genes show similarity)	BGC0001065_c1
Cluster 27	Bacteriocin-Lantipeptide	8992171	9024800	Informatipeptin_biosynthetic_gene_cluster (100% of genes show similarity)	BGC0000518_c1
Cluster 28	T3pks-T1pks	9147535	9201019	Alkylresorcinol_biosynthetic_gene_cluster (100% of genes show similarity)	BGC0000282_c1
Cluster 29	Nrps	9336380	9399107	Myxochelin_biosynthetic_gene_cluster (16% of genes show similarity)	BGC0001345_c1
Cluster 30	Nrps	9696917	9752656	Mitomycin_biosynthetic_gene_cluster (10% of genes show similarity)	BGC0000915_c1
Cluster 31	T3pks-Phenazine-Butyrolactone-Nrps	9890361	9998669	Merochlorin_biosynthetic_gene_cluster (70% of genes show similarity)	BGC0001083_c1

Table S2. Annotation and homologues of genes in *tym*

***tym* gene**	**Size (Aa)**	**Annotation**	***myc* homologue**	**Identity/positive**
tym O1	36	Lipoprotein	XP_021655009	40/53
tym O2	64	Transposase	WP_128382552	80/81
tym R1	695	AraC family transcriptional regulator	KUJ68505	56/68
tym O3	287	Serine-Carboxyl Peptidase	WP_051338324	51/62
tym O4	846	WD-40 repeat-containing protein	WP_124439956	43/55
tym T1	261	ABC-2 type transporter	WP_030715158	97/98
tym O5	323	ABC transporter ATP-binding protein	WP_037806673	89/94
tym R2	245	TetR family transcriptional regulator	WP_063754831	92/94
tym M4	388	FAD dependent oxidoreductase	WP_030948426	88/91
tym R3	242	GntR family transcriptional regulator	WP_079106220	41/56
tym O6	40	hypothetical protein	PSK54999	80/82
tym O7	197	dihydropteridine reductase	*mycH1* (AFG19428)	90/97
tym C1	277	shikimate/quinate dehydrogenase	*mycA5* (AFG19427)	83/87
tym M3	316	O-methyltransferase	*mycF3* (AFG19426)	92/96
tym M2	546	monooxygenase FAD-binding	*mycF2* (AFG19425)	84/88
tym O8	260	N-acetyltransferase	*mycE* (AFG19424)	85/90
tym P5	3301	polyketide synthase	*mycD5* (AFG19423)	83/88
tym P4	3475	polyketide synthase	*mycD4* (AFG19422)	71/79
tym P3	3315	polyketide synthase	*mycD3* (AFG19421)	78/83
tym P2	1870	polyketide synthase	*mycD2* (AFG19420)	90/92
tym P1	3165	polyketide synthase	*mycD1* (AFG19419)	84/89
tym A1	345	3-dehydroquinate synthase	*mycB7* (AFG19418)	91/95
tym A2	295	ROK family protein	*mycB6* (AFG19417)	81/85
tym N	844	AMP-dependent synthetase and ligase	*mycC* (AFG19416)	90/94

***tym* gene**	**Size (Aa)**	**Annotation**	***myc* homologue**	**Identity/positive**
tym M1	308	GCN5-related N-acetyltransferase	*mycF1* (AFG19415)	90/94
tym A3	139	aminoDHQ dehydratase	*mycB5* (AFG19414)	93/97
tym A4	387	DegT/DnrJ/EryC1/StrS_aminotransferase	*mycB4* (AFG19413)	92/96
tym A5	376	oxidoreductase	*mycB3* (AFG19412)	79/84
tym A6	235	HAD-superfamily hydrolase, subfamily IA, variant	*mycB2* (AFG19411)	81/88
tym A7	440	deoxyarabinoheptulosonate-7-phosphate synthase	*mycB1* (AFG19410)	84/89
tym C2	1005	AMP-dependent synthetase and ligase	*mycA4* (AFG19409)	88/92
tym C3	389	acyl-CoA dehydrogenase	*mycA3* (AFG19408)	91/94
tym C4	280	short-chain dehydrogenase/reductase SDR	*mycA2* (AFG19407)	92/96
tym C5	699	NADH: Flavin oxidoreductase/NADH oxidase	*mycA1* (AFG19406)	84/90
tym R1	929	LuxR family transcriptional regulator	*mycG1* (AFG19405)	81/87
tym T1	252	thioesterase	WP_069773350	91/95

Figure S1. The molecular network obtained by combining the LC-MS/MS analyses of five fractions of extracts from strain H2S5.



http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=eb768045e99e4fc2ab4454674fe04b1d

Figure S2. ^{1}H NMR spectrum of strecacansamycin A (1) (800 MHz, CD$_3$OD)

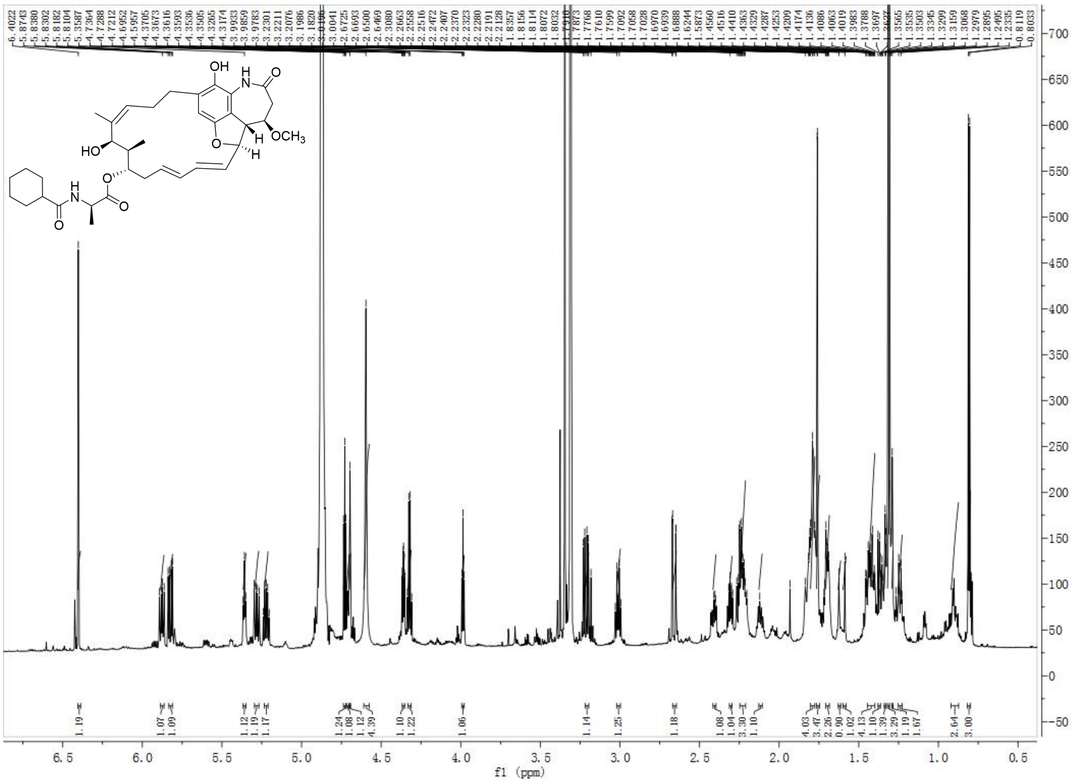


Figure S3. ^{13}C NMR spectrum of strecacansamycin A (1) (800 MHz, CD$_3$OD)

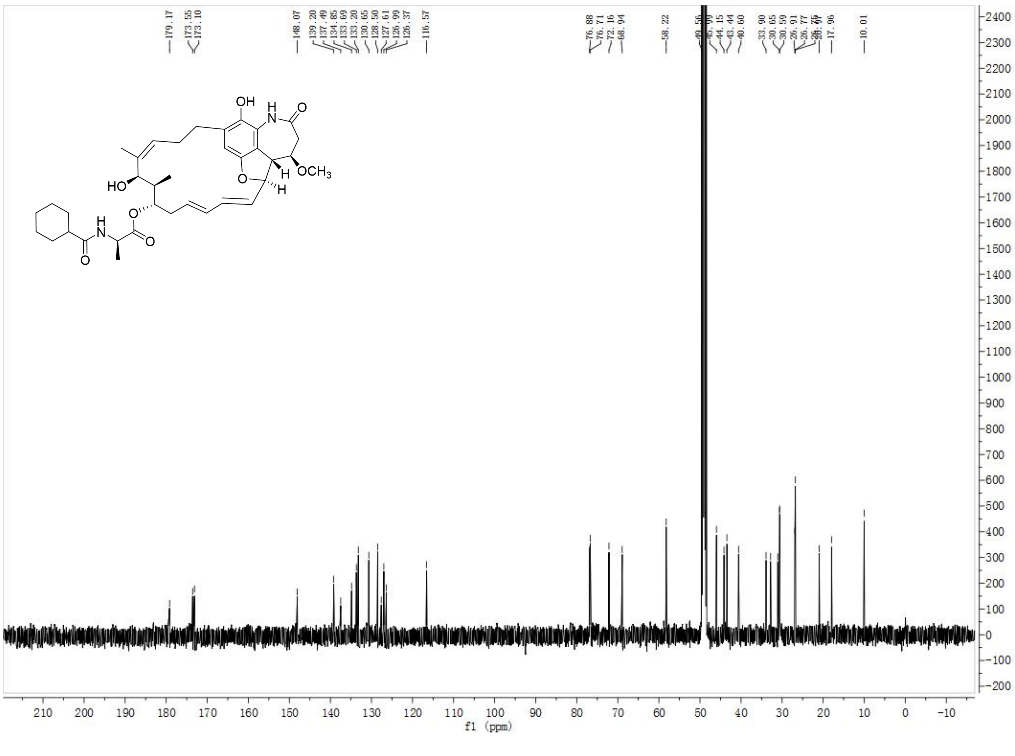


Figure S4: HSQC spectrum of strecacansamycin A (1) (CD_3OD).

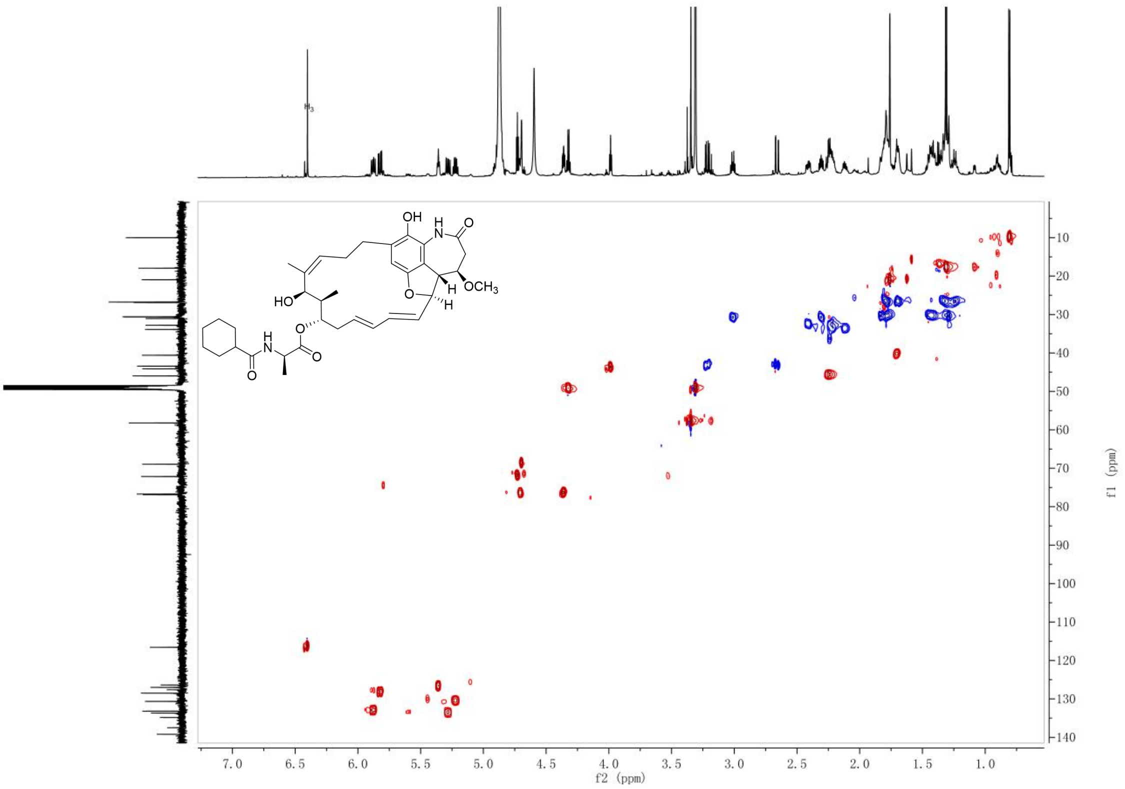


Figure S5: ^{1}H-^{1}H COSY spectrum of strecacansamycin A (1) (CD_3OD).

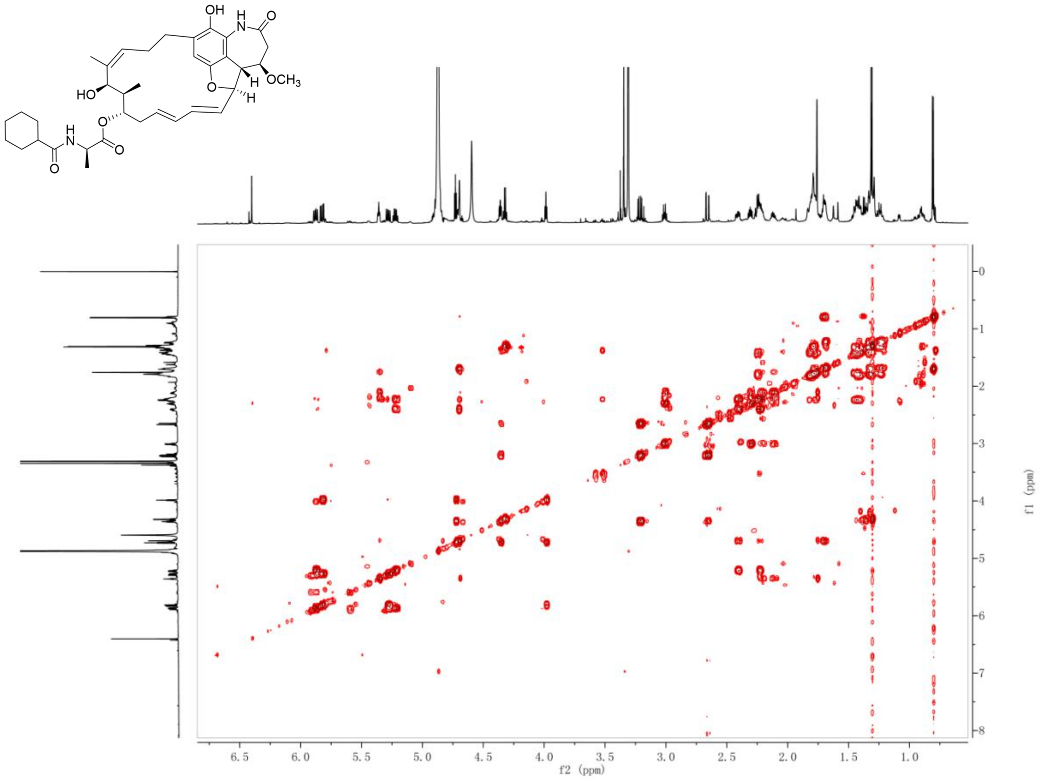


Figure S6: HMBC spectrum of strecacansamycin A (1) (CD_3OD).

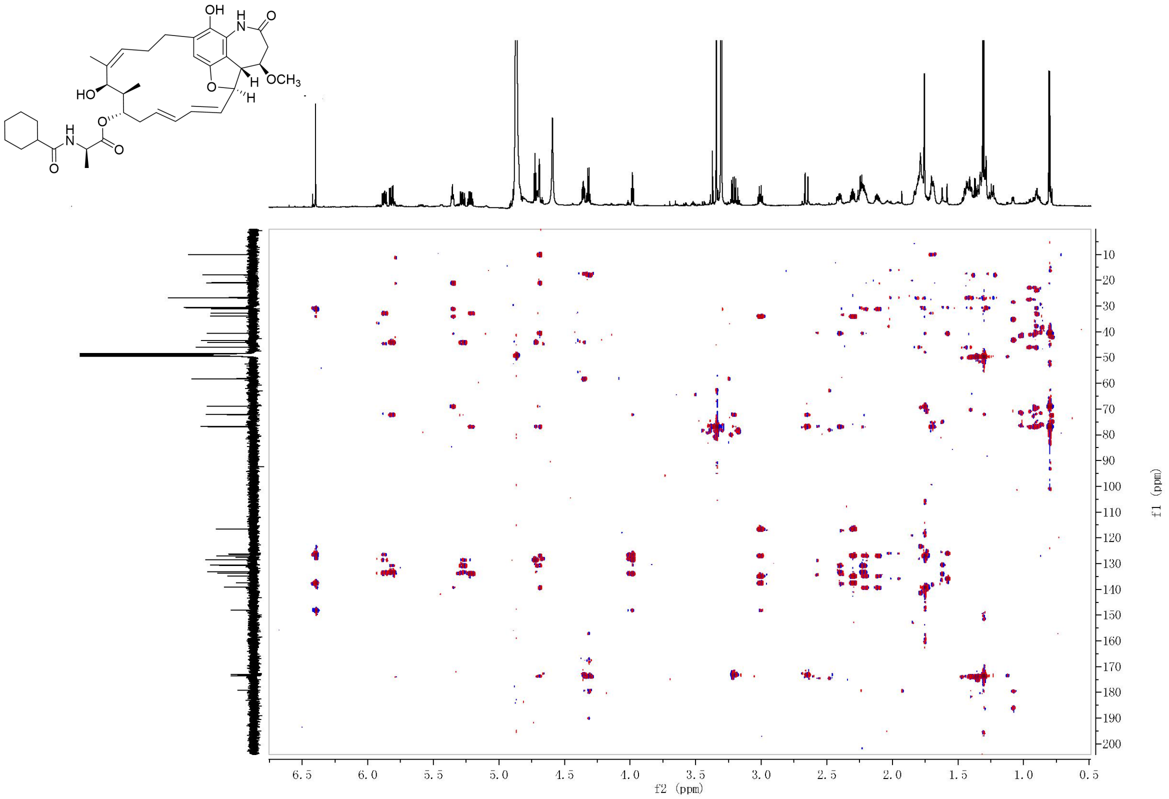


Figure S7: The enlargement (part 1) of HMBC spectrum of compound 1.

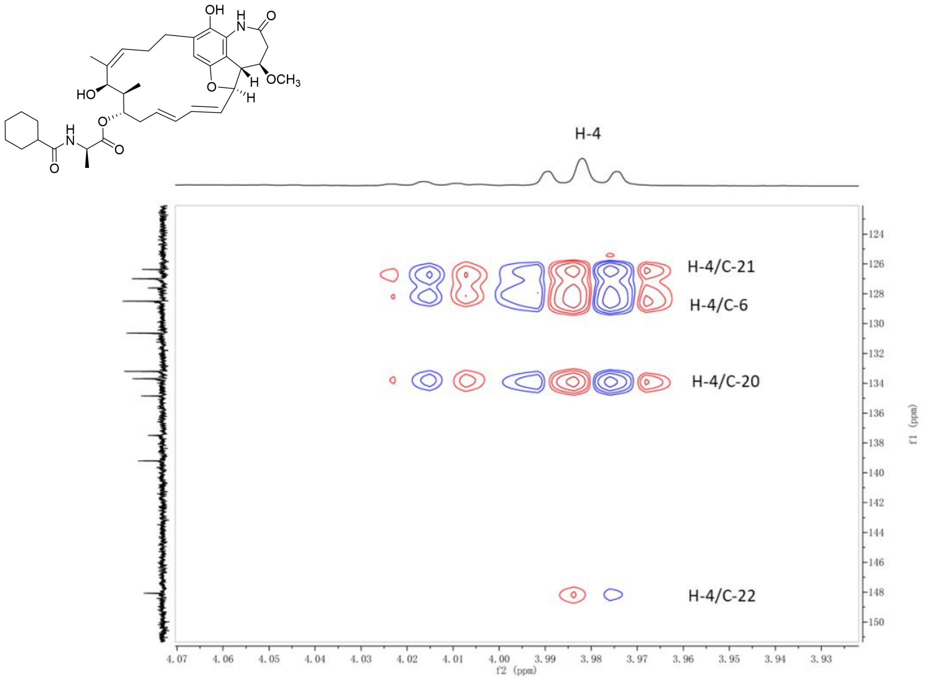


Figure S8: The enlargement (part 2) of HMBC spectrum of compound 1.

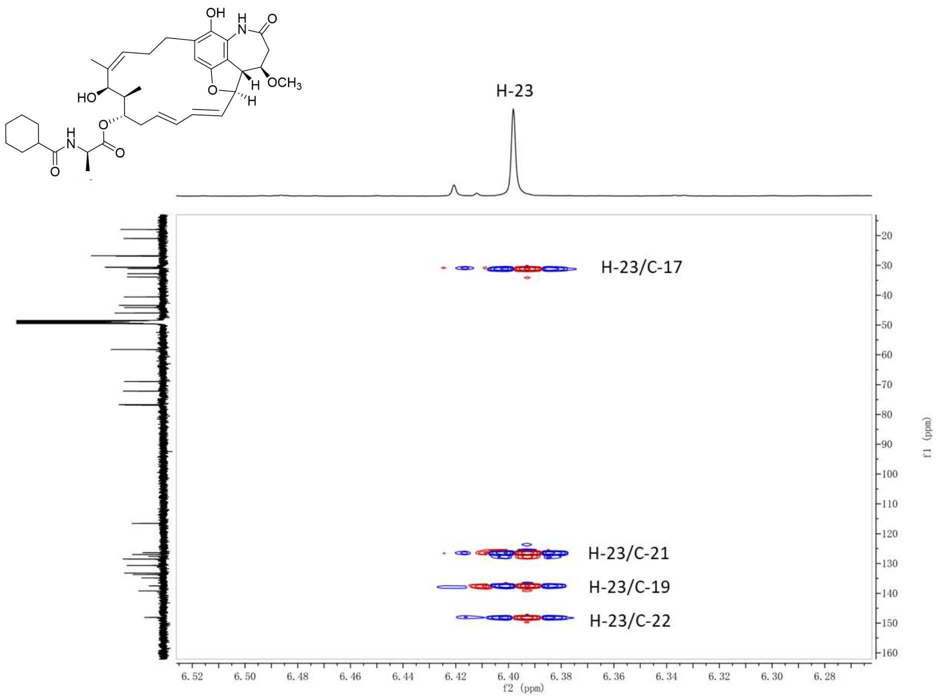


Figure S9: The enlargement (part 3) of HMBC spectrum of compound 1.

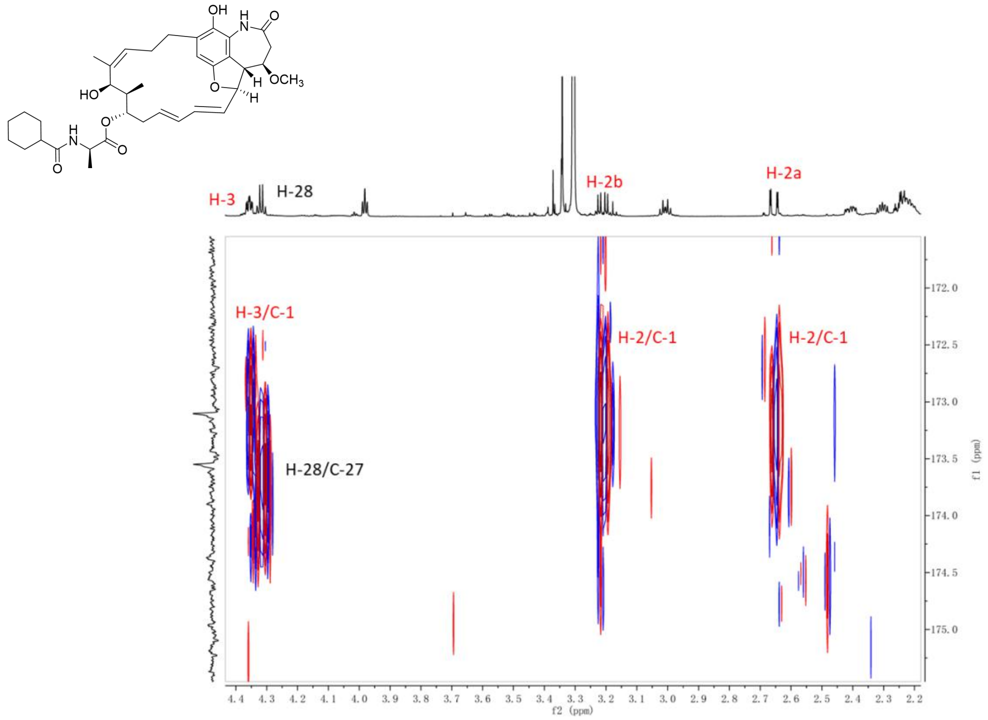


Figure S10: NOESY spectrum of strecacansamycin A (1) (CD_3OD).

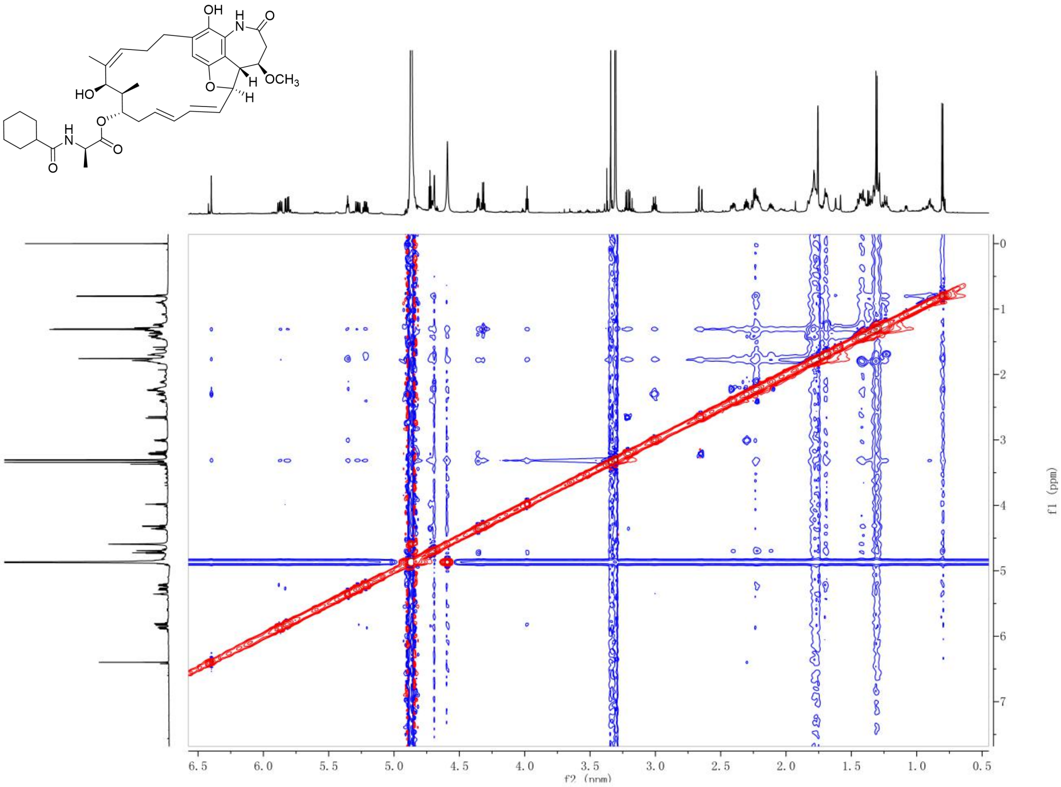


Figure S11: HRESIMS spectrum of strecacansamycin A (1).

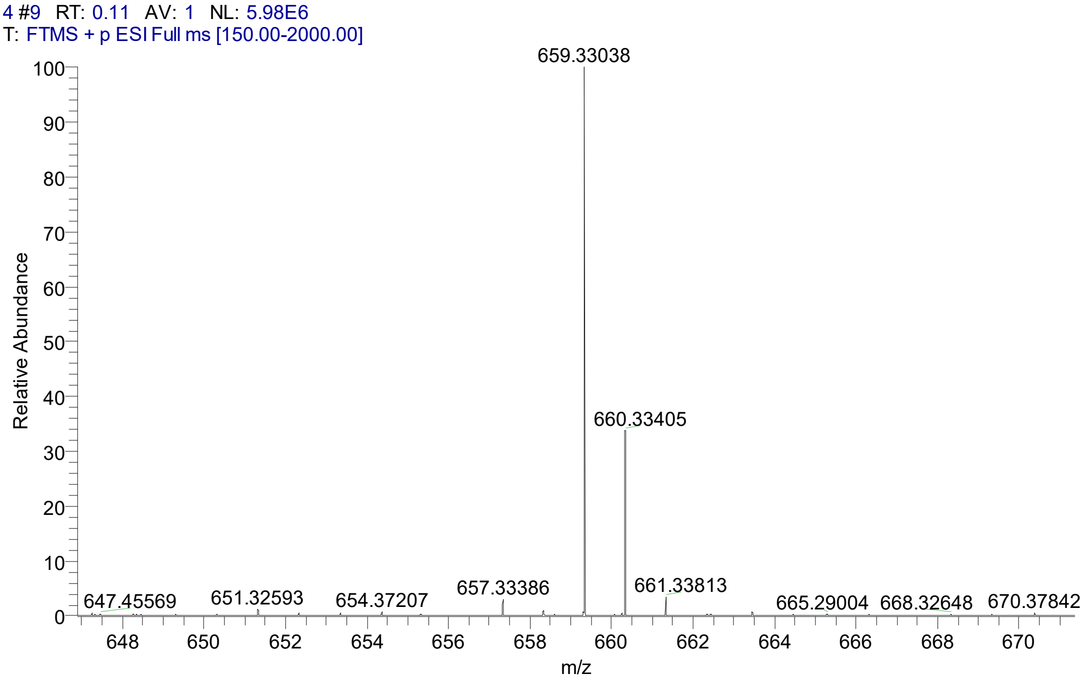


Figure S12: MS/MS spectrum of strecacansamycin A (1).

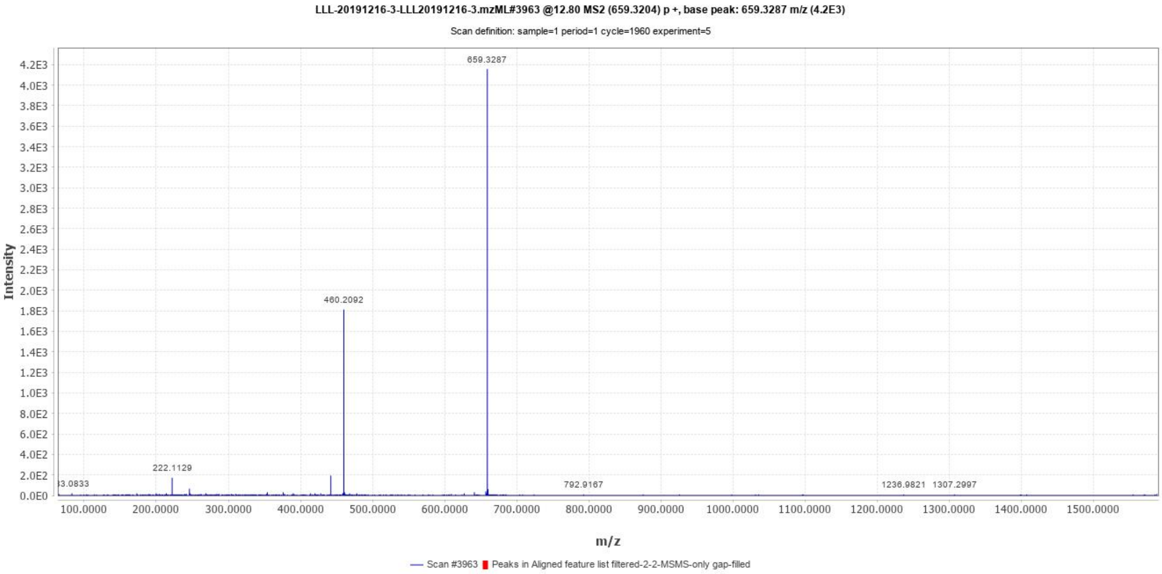


Figure S13: UV spectrum of strecacansamycin A (1).

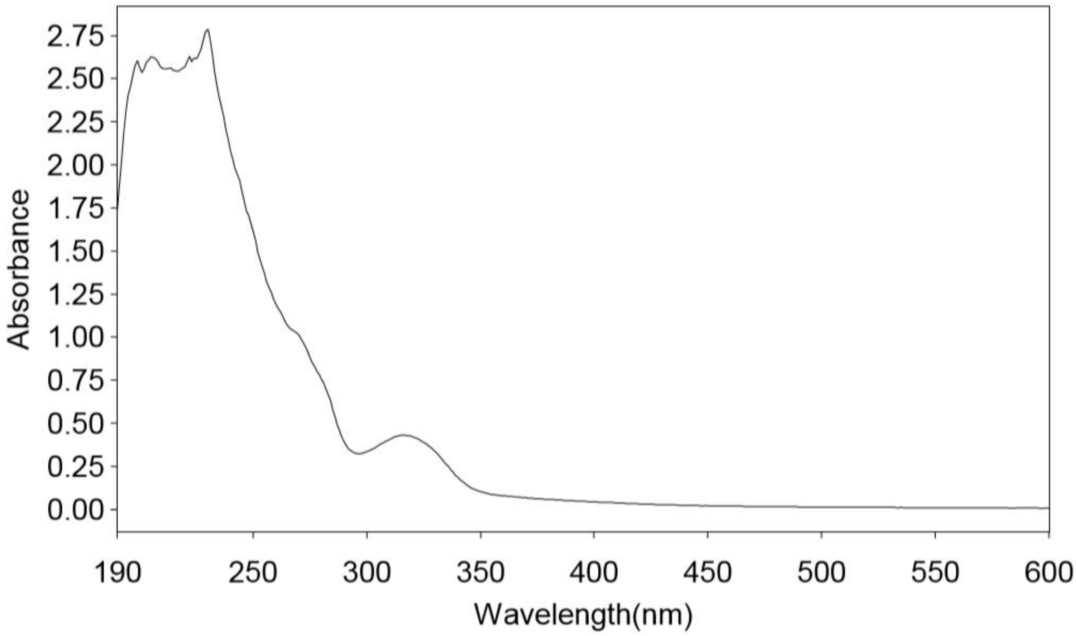


Figure S14: ^{1}H NMR spectrum of strecacansamycin B (2) (800 MHz, CD_3OD)

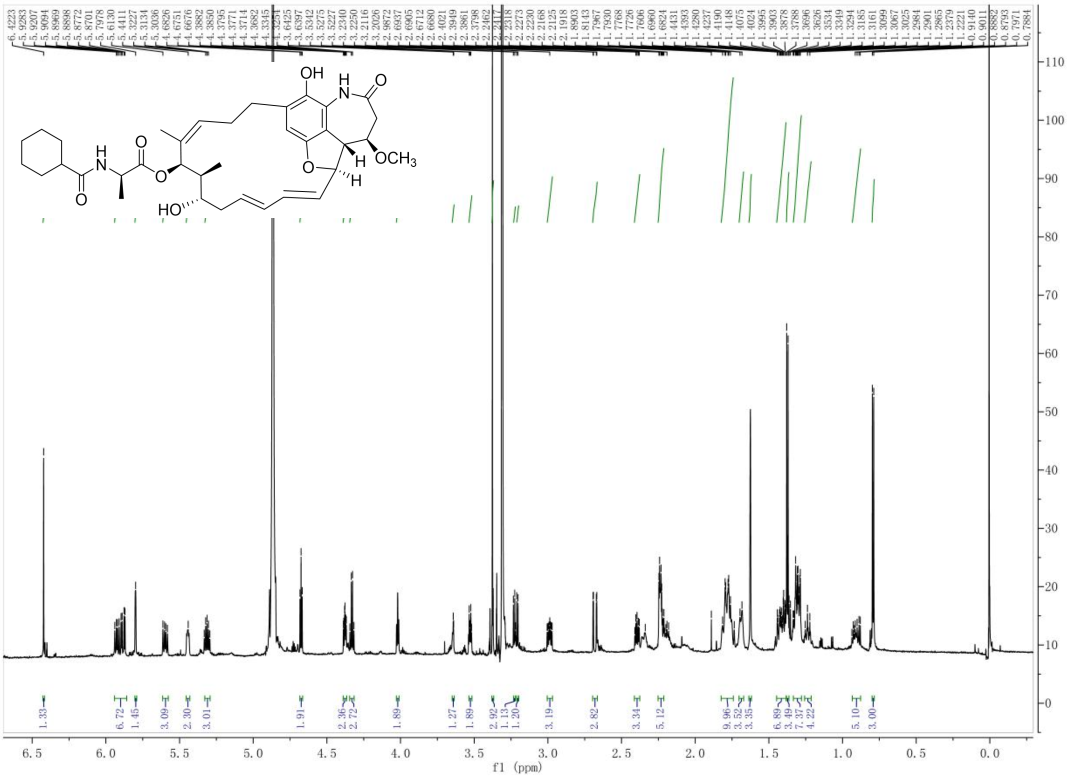


Figure S15: ^{13}C NMR spectrum of strecacansamycin B (2) (800 MHz, CD_3OD)

Figure S16: HSQC spectrum of strecacansamycin B (2) (CD_3OD).

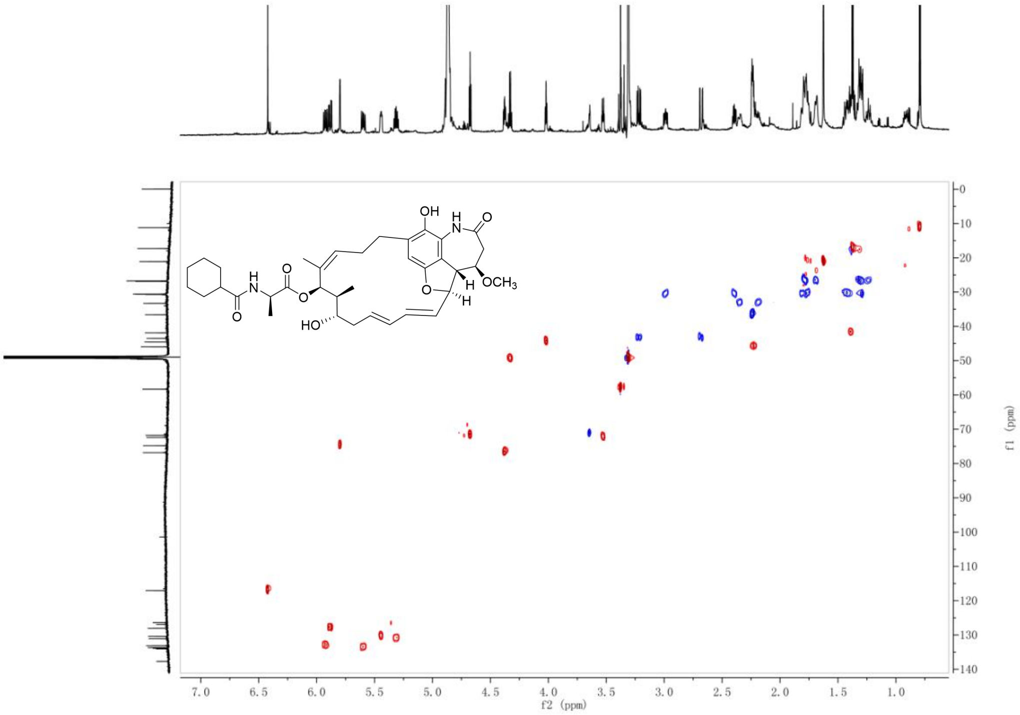


Figure S17: ^1H-^1H COSY spectrum of strecacansamycin B (2) (CD_3OD).

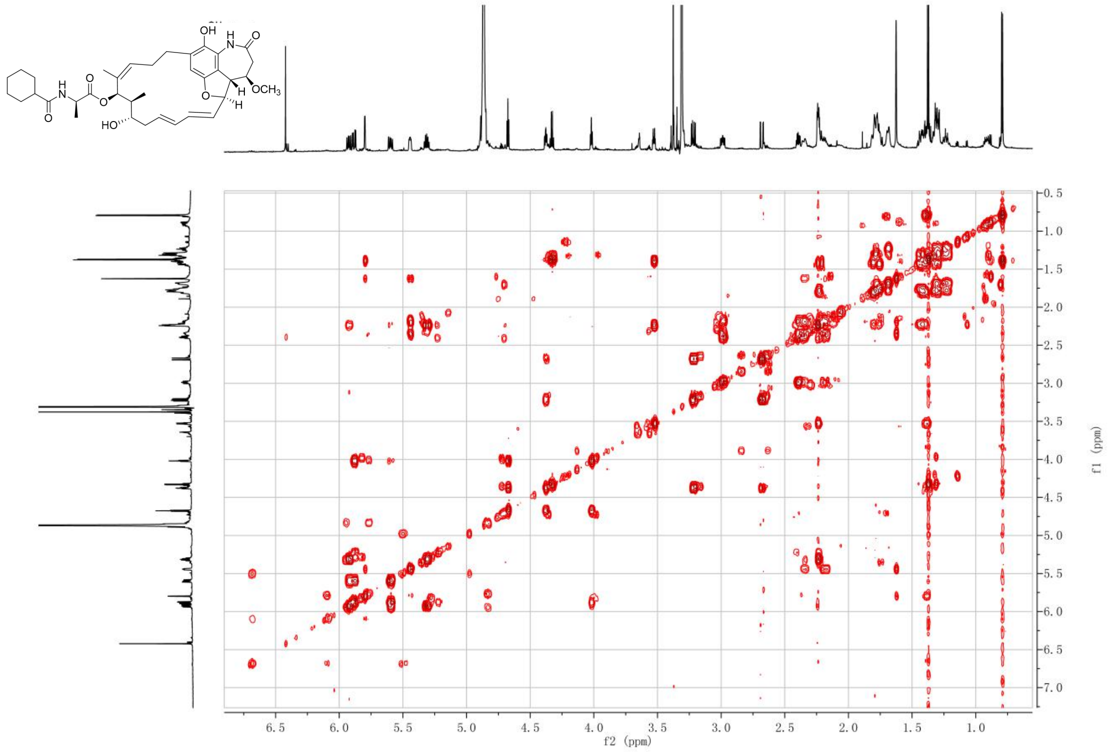


Figure S18: HMBC spectrum of strecacansamycin B (2) (CD_3OD).

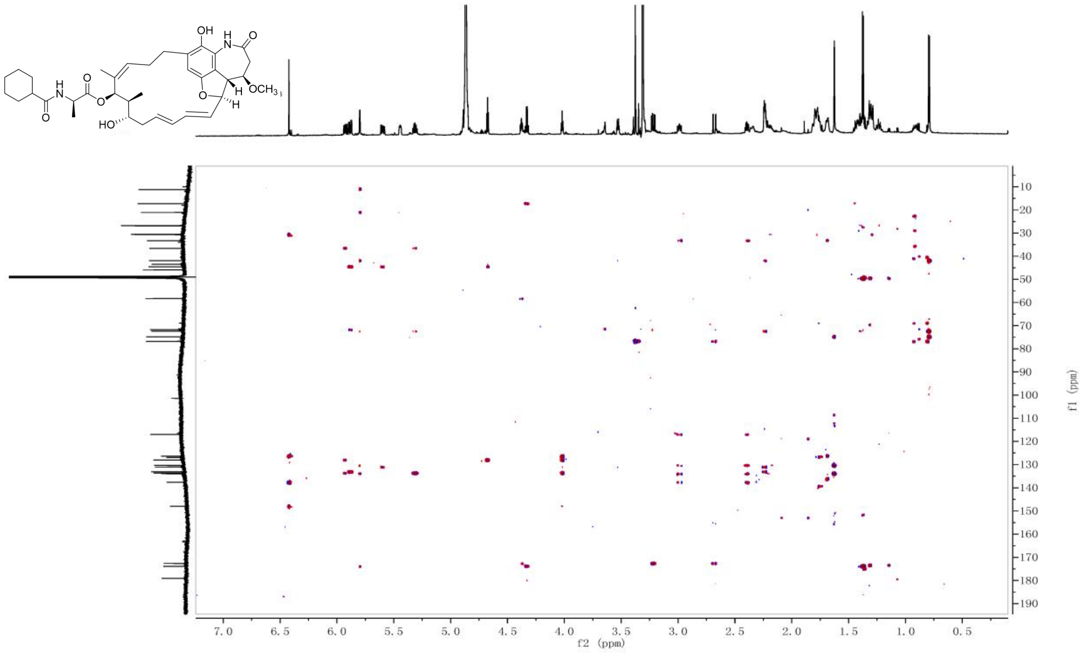


Figure S19: NOESY spectrum of strecacansamycin B (2) (CD_3OD).

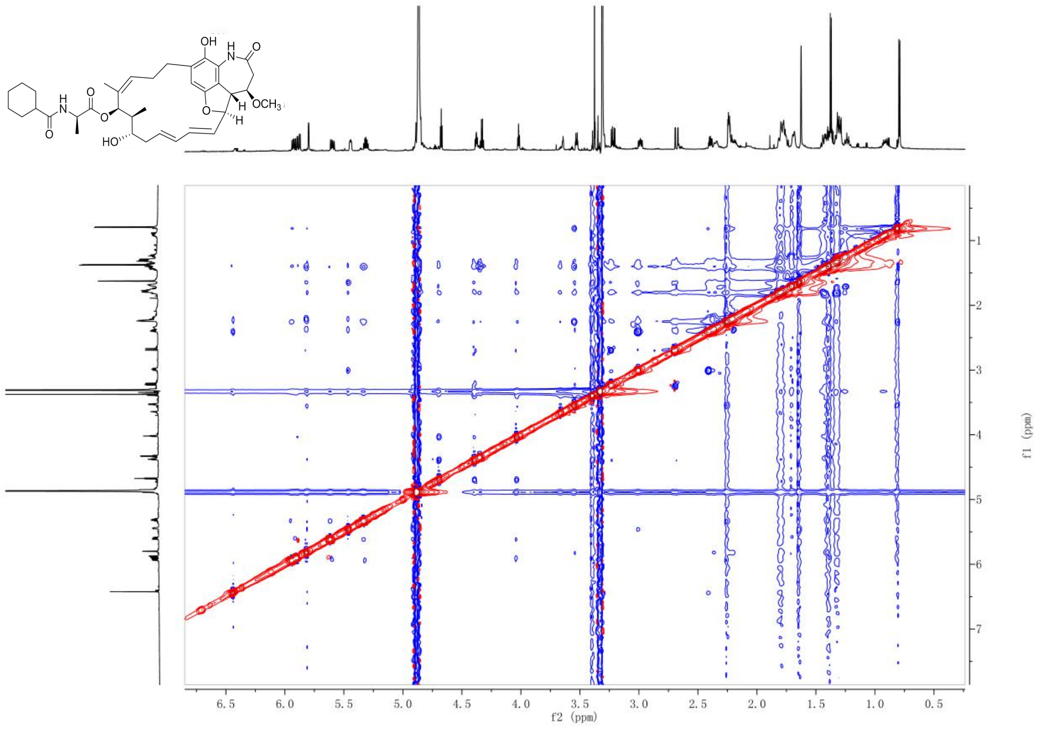


Figure S20: HRESIMS spectrum of strecacansamycin B (2).

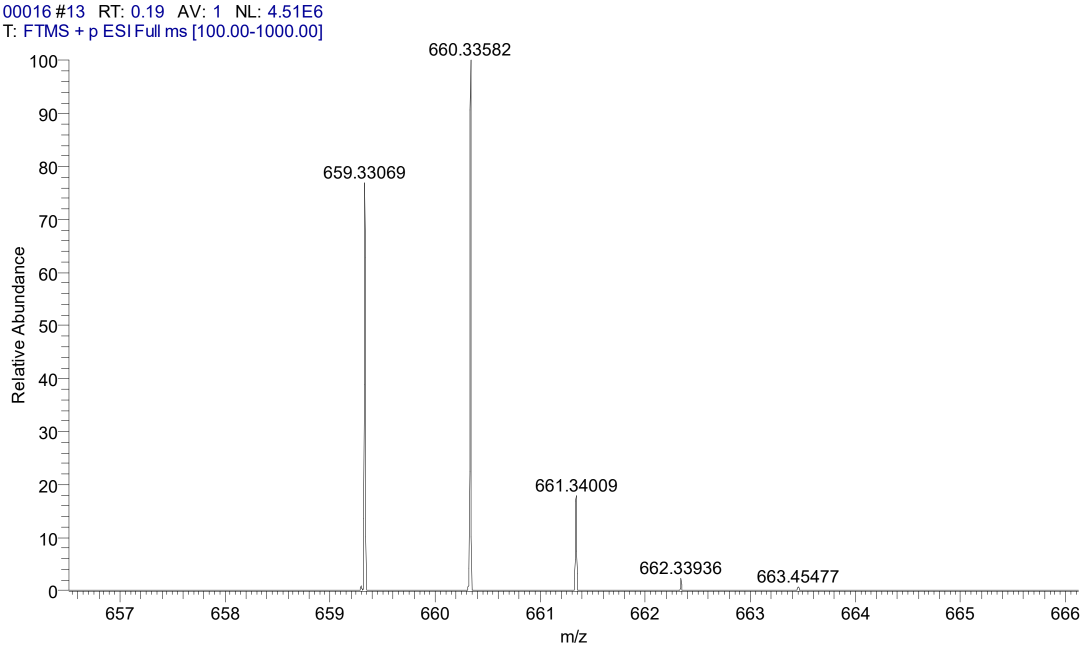


Figure S21: MS/MS spectrum of strecacansamycin B (2).

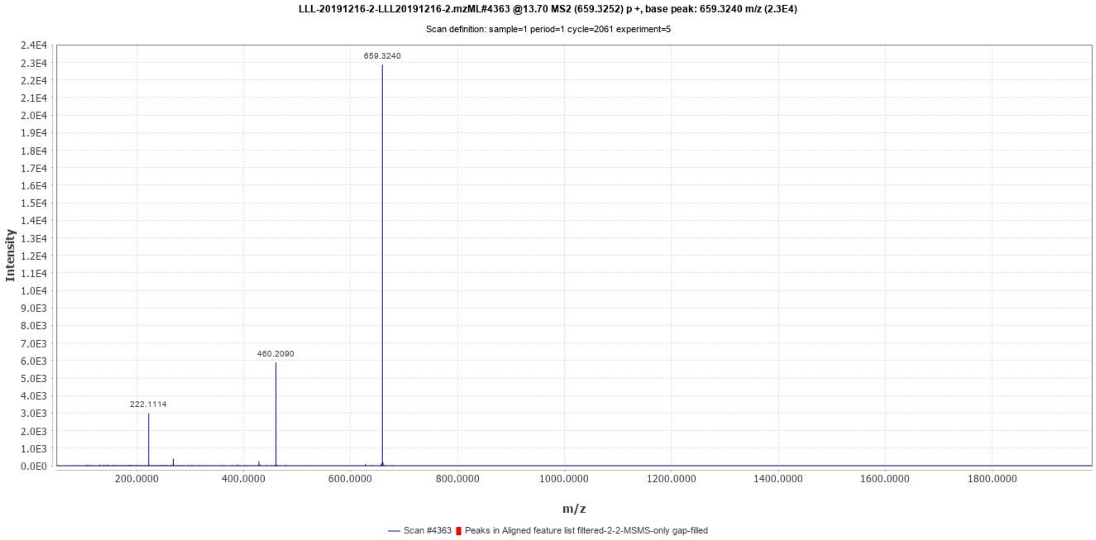


Figure S22: UV spectrum of strecacansamycin B (2).

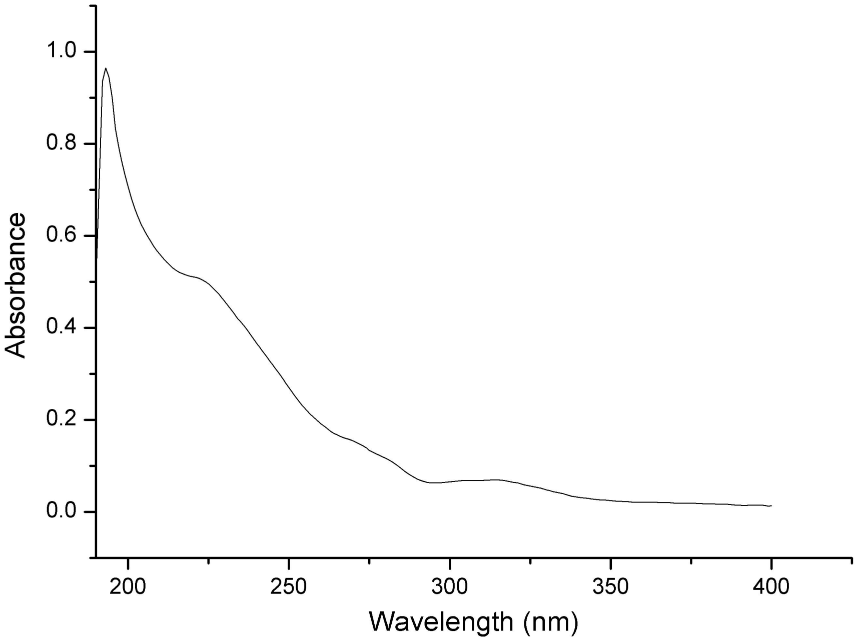
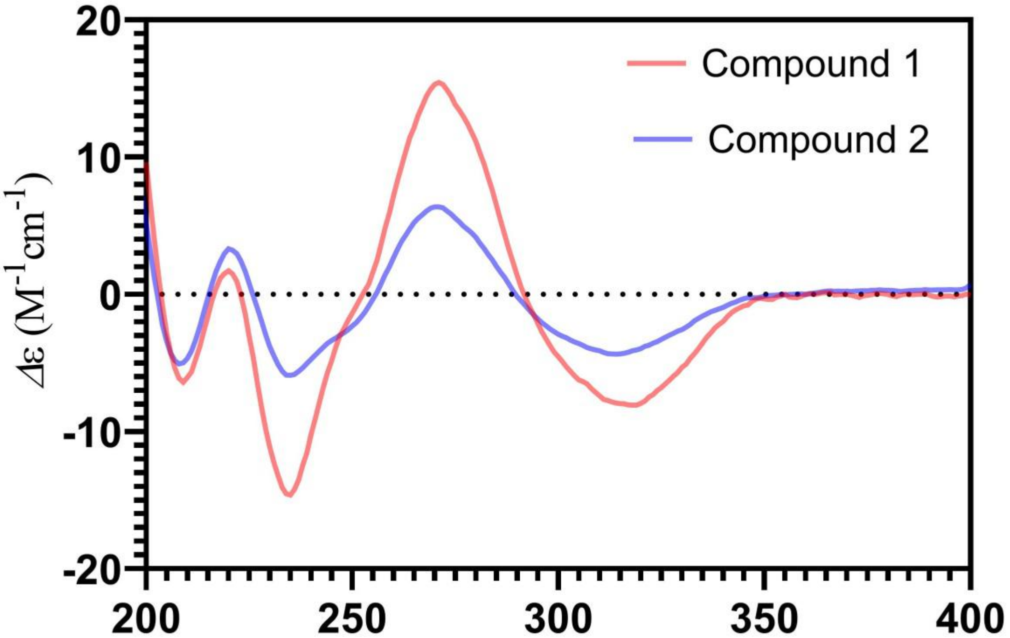


Figure S23: CD spectra of strecacansamycin A (1) and B (2).



The method of the ECD calculation.

The calculation was performed according to our previously reported protocol. A preliminary conformational search was performed in Conflex 6.7 using MMFF94s forcefield. Conformers were saved and further optimized using the density functional theory (DFT) method and CPCM solvent model at B3LYP/6-31+G(d,p) level in Gaussian 09 software package.[1] The frequency was calculated at the same level of theory to check optimized results. The stable conformers with populations greater than 1% and without imaginary frequencies were submitted to ECD calculation by the TDDFT using cam-B3LYP/TZVP method associated with CPCM solvent model in Methanol. The excitation energies (E), oscillator strength (f), rotatory strength in velocity form (Rvel), and rotatory strength in length form (Rlen) of the lowest 32 excited states were calculated. ECD spectra of different conformers were summated in SpecD is according to their Boltzmann-calculated distributions.

[1] Gaussian 09, Revision D.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2013.

Figure S24: ^{1}H NMR spectrum of strecacansamycin C (3) (500 MHz, CD_3OD).

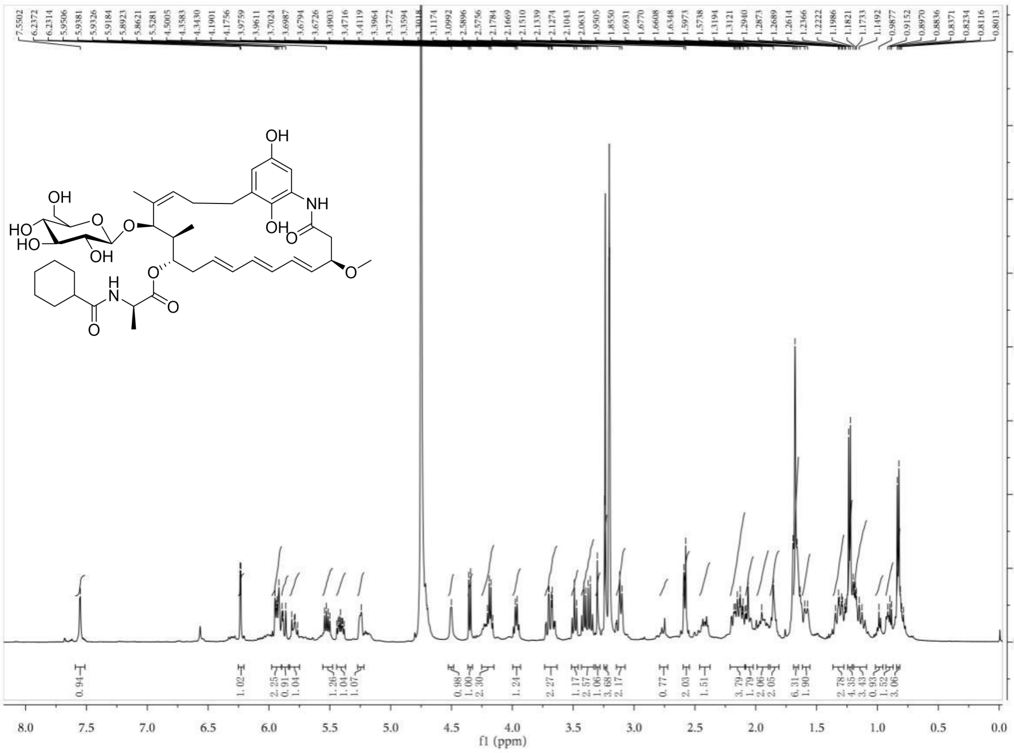


Figure S25: ^{13}C NMR spectrum of strecacansamycin C (3) (125 MHz, CD_3OD).

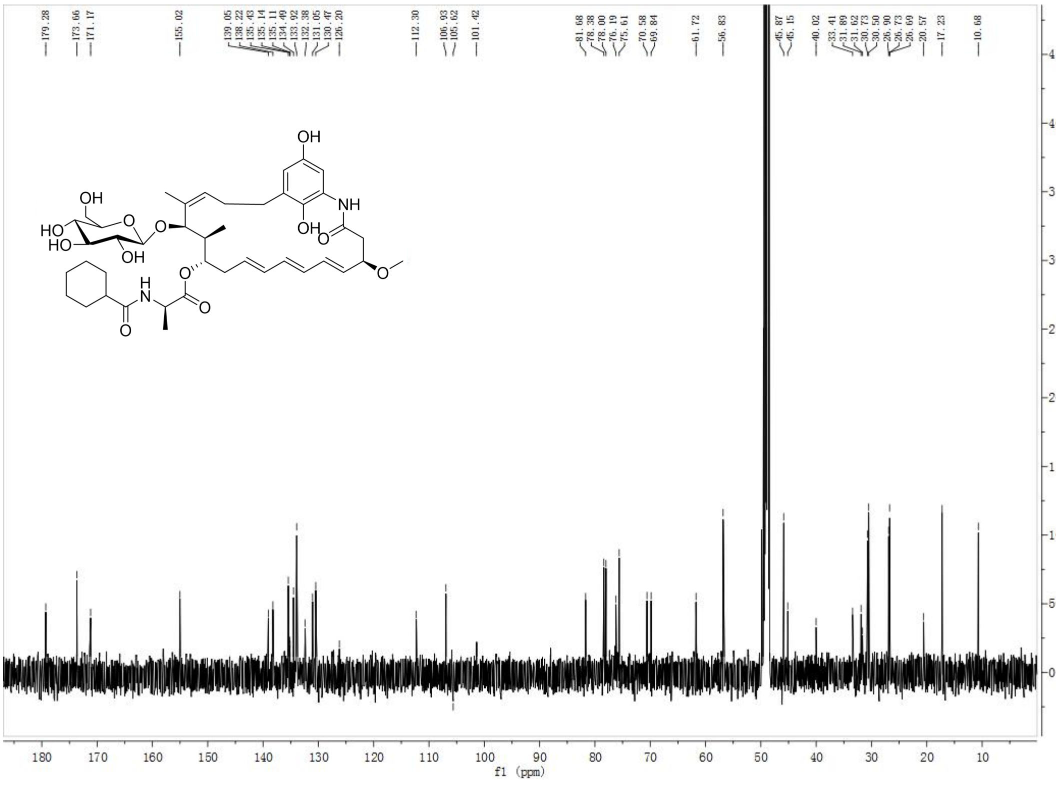


Figure S26: DEPT spectrum of strecacansamycin C (3) (125 MHz, CD_3OD).

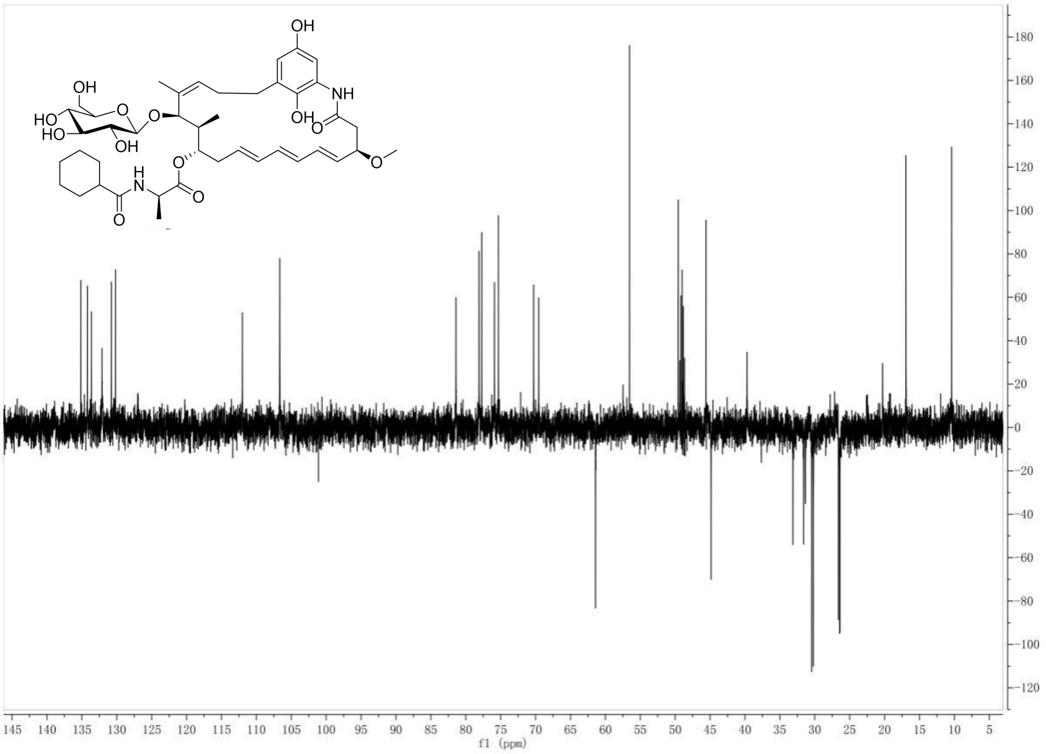


Figure S27: HSQC spectrum of strecacansamycin C (3) (CD_3OD).

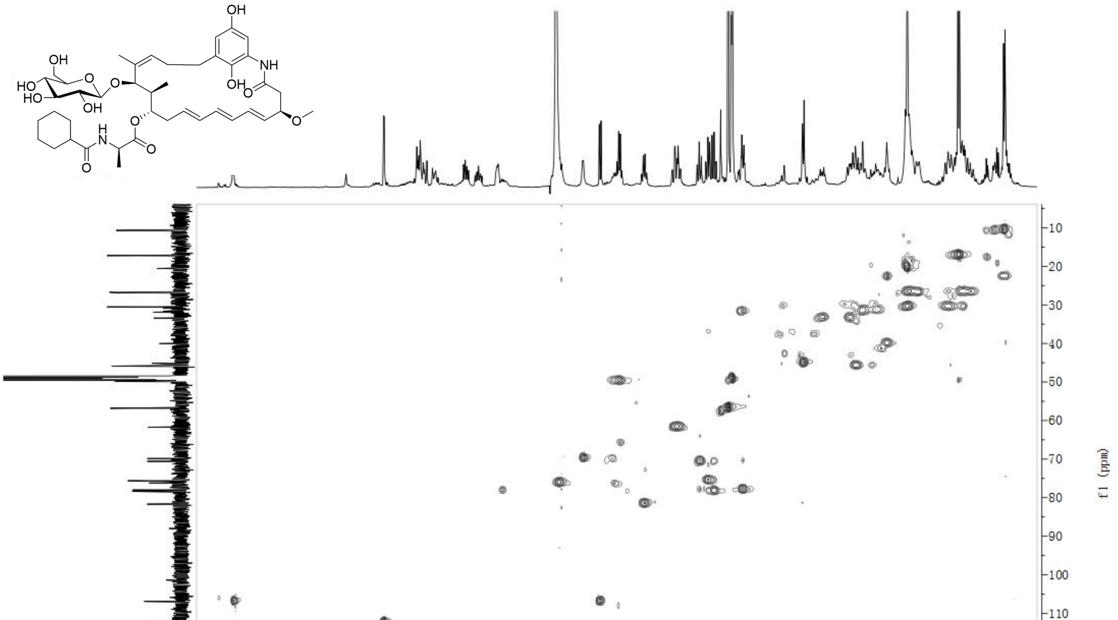


Figure S28: ^1H-^1H COSY spectrum of strecacansamycin C (3) (CD_3OD).

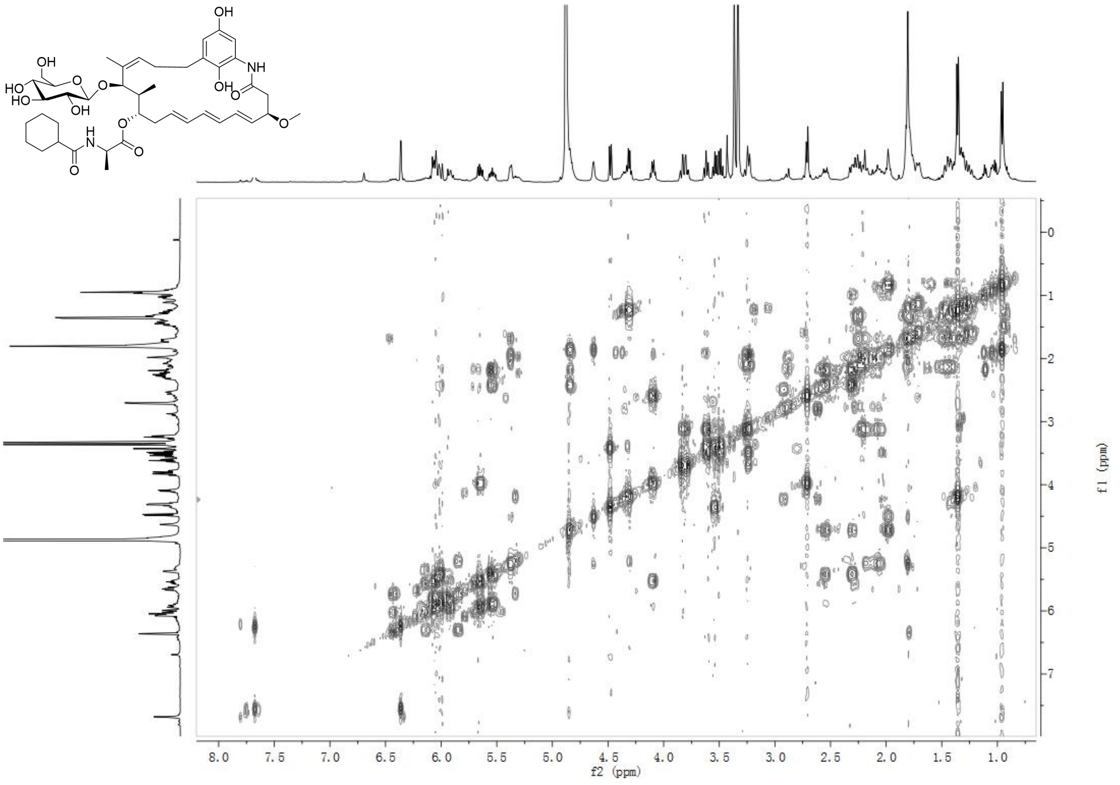


Figure S29: HMBC spectrum of strecacansamycin C (3) (CD_3OD).

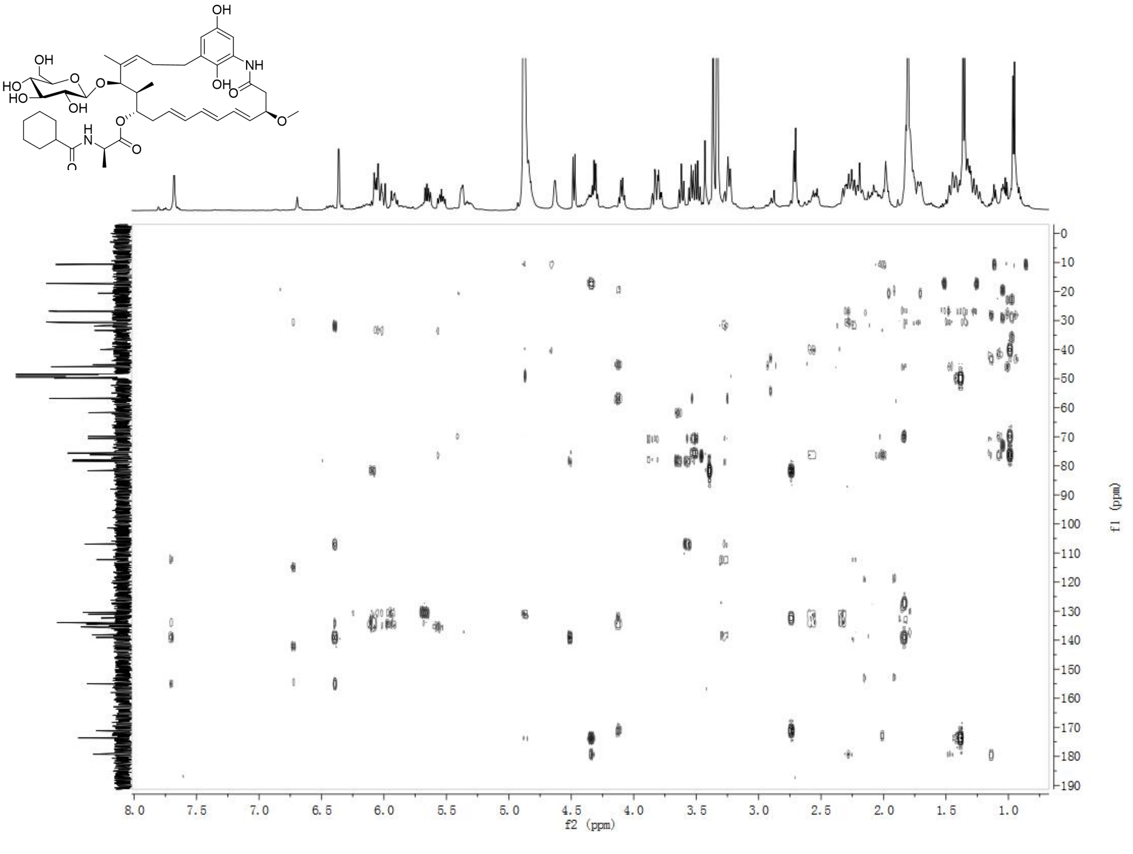


Figure S30: NOSEY spectrum of strecacansamycin C (3).

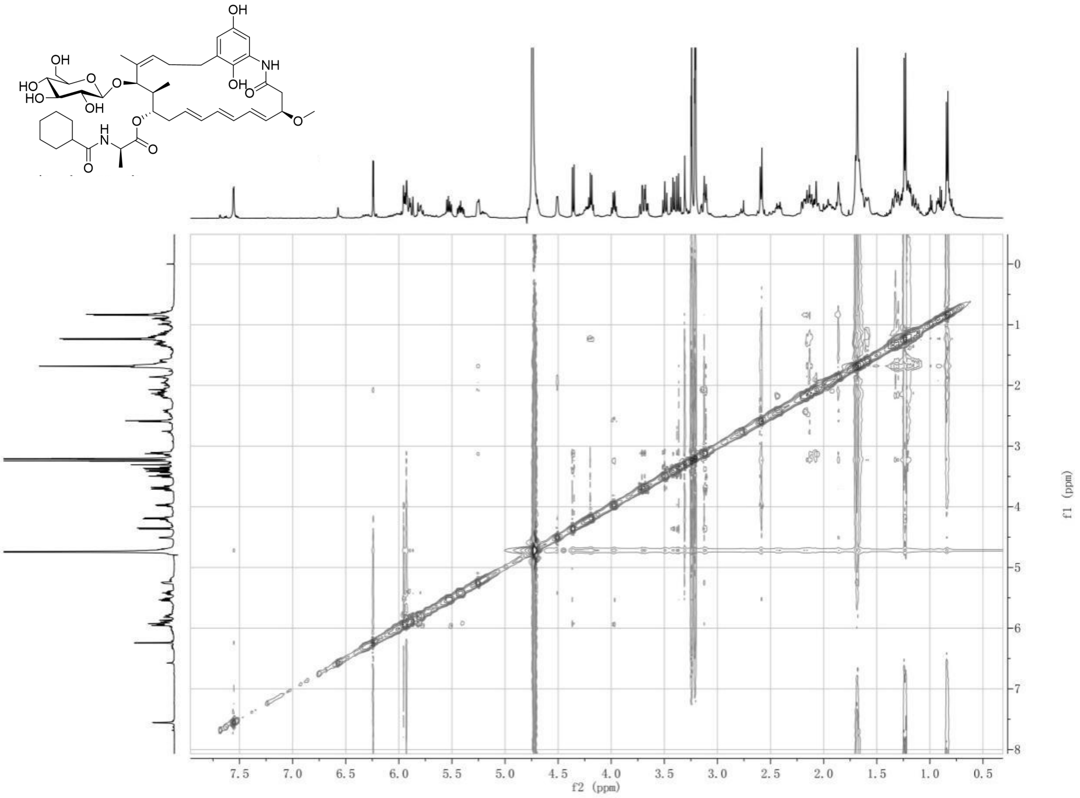


Figure S31: HRESIMS spectrum of strecacansamycin C (3).

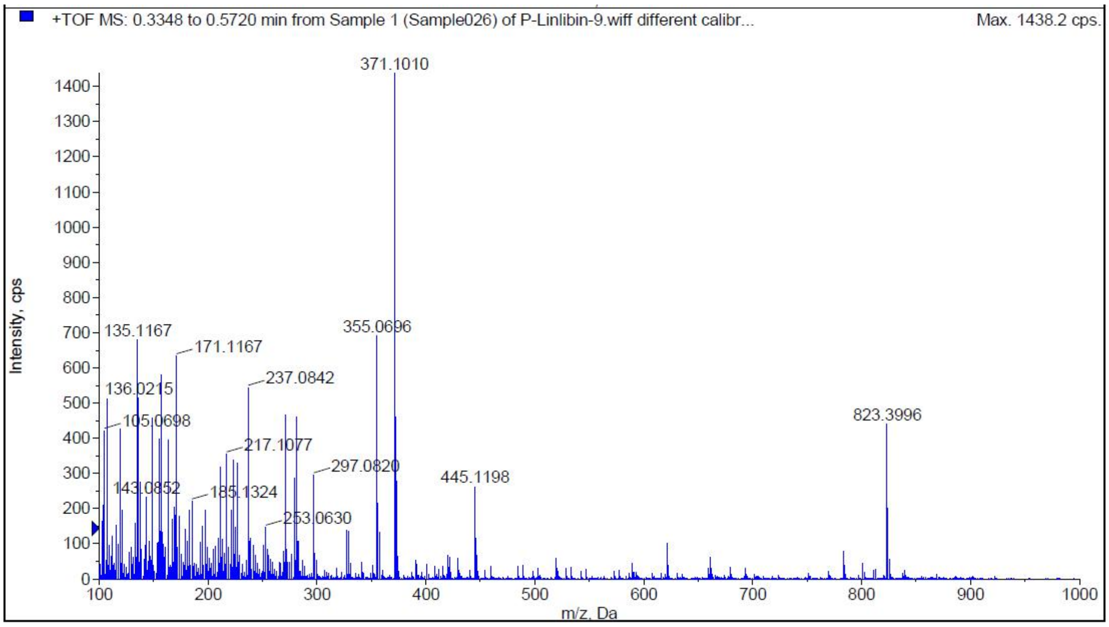


Figure S32: MS/MS spectrum of strecacansamycin C (3).

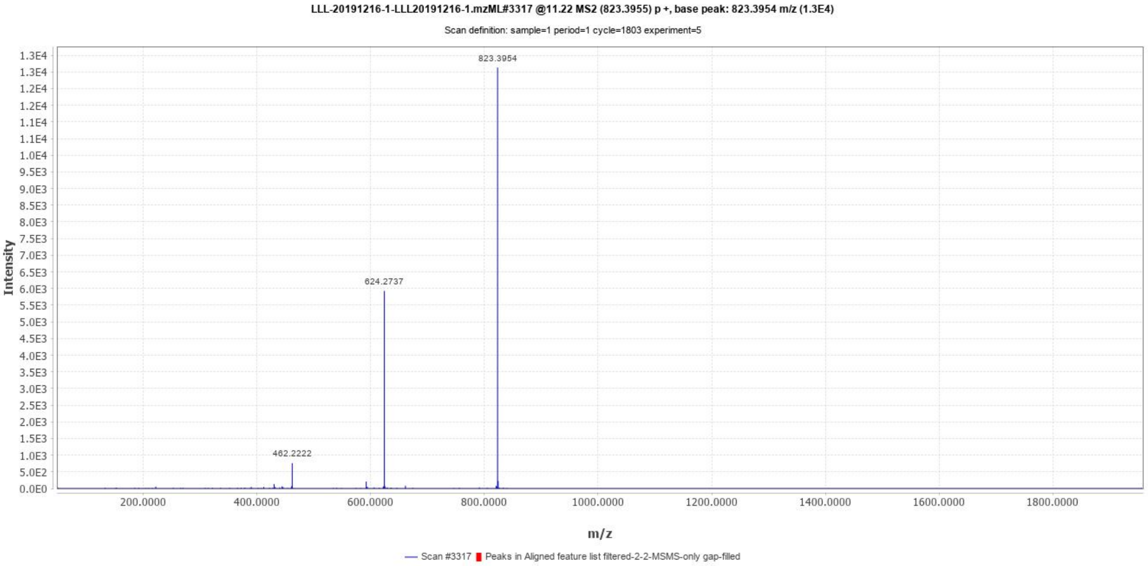


Figure S33: IR spectrum of strecacansamycin C (3).

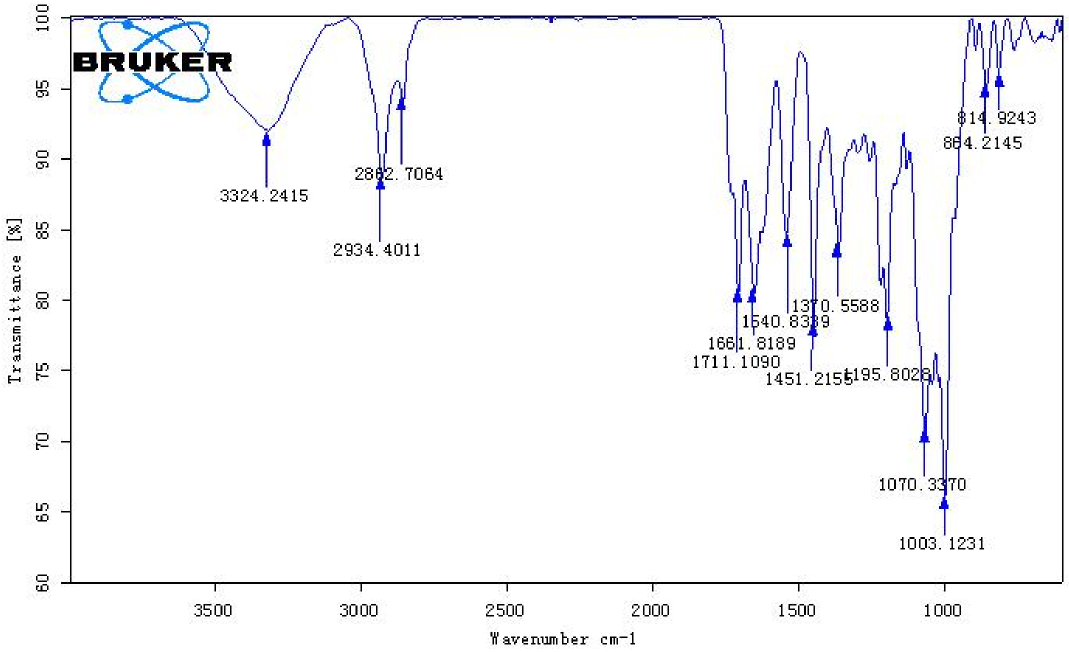


Figure S34: UV spectrum of strecacansamycin C (3).

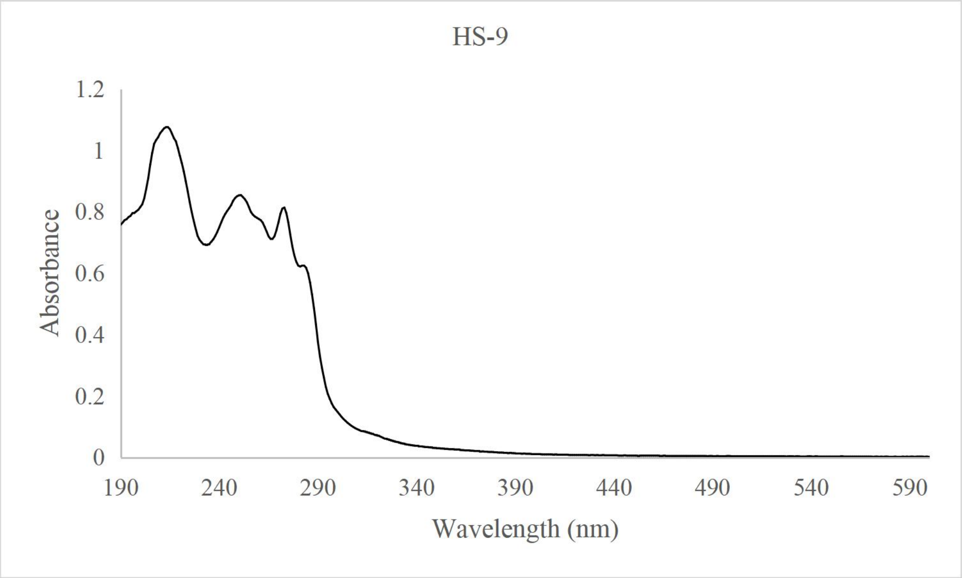


Figure S35: Hydrolysis of compound 3 and GC-MS analysis of the hydrolyzates and trimethylsilyl etherification. Blackline, compound 3. Purple line, D-(+)-Glucose. Blue line, L-(-)-Glucose.

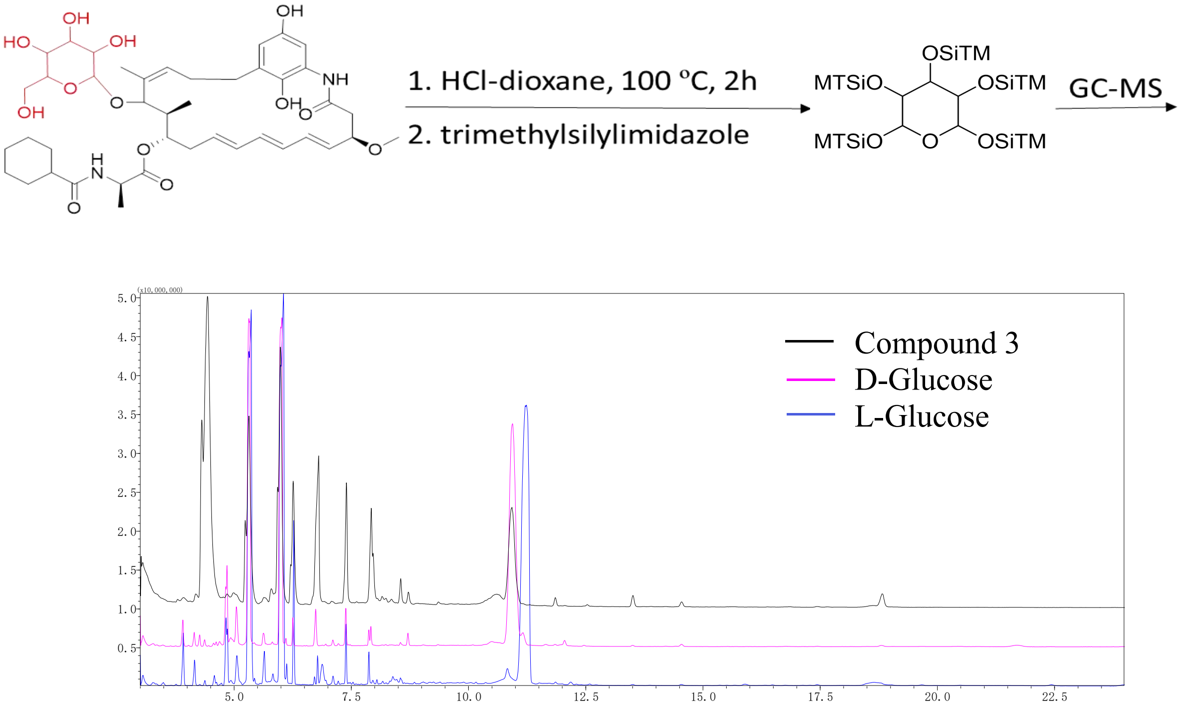


Figure S36. Graphic illustration of the *S. cacaoi* subsp. *asoensis* H2S5 genome.

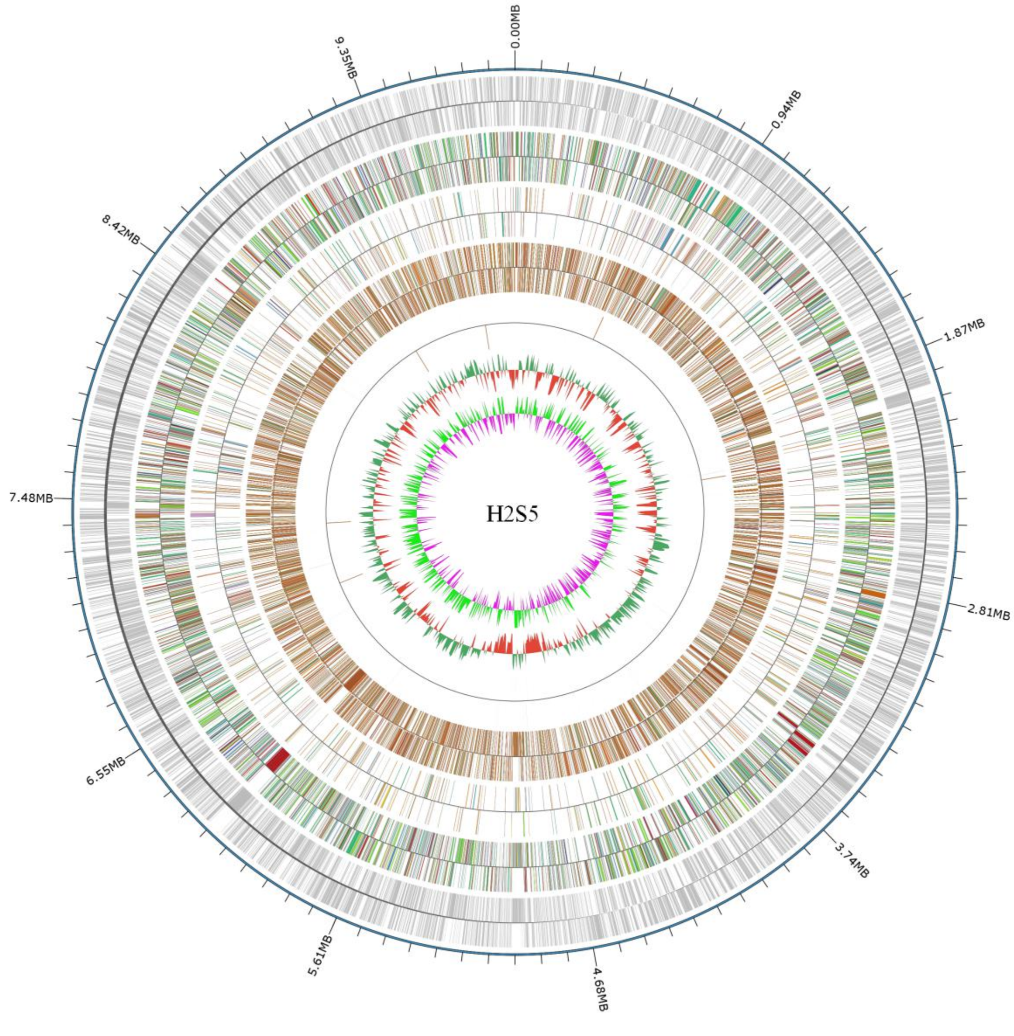


Figure S37. Amino acid sequence alignments of the domain active sites. The red box indicated the LDD motif in KR domain and the missing of the active site in DH domain.

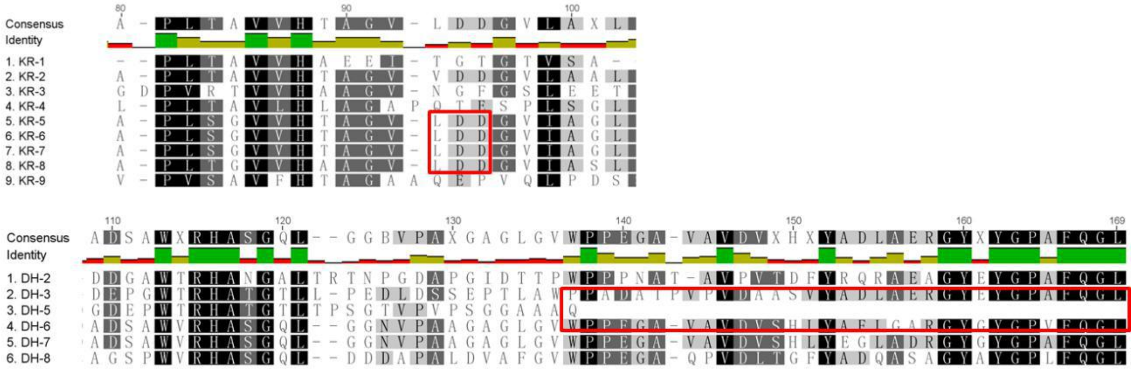
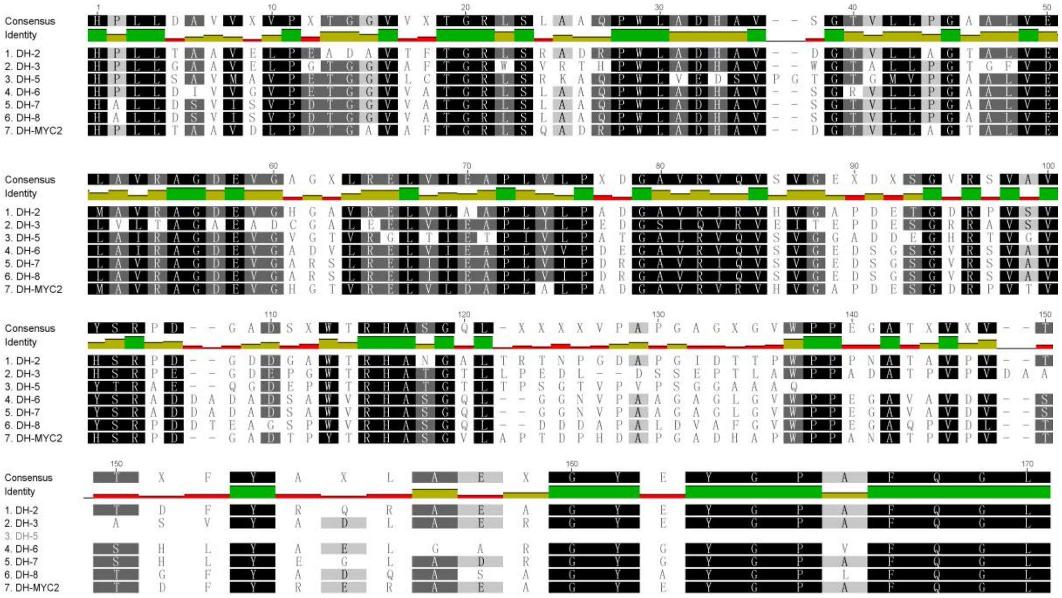


Figure S38. Amino acid alignment of the *tym* BGC AT domains

Figure S39. Amino acid alignment of the *tym* BGC DH domains



```
            1                 10                  20                  30                  40                  50
Consensus   H P L L D A V V X V P X T G G V V X T G R L S L A A Q P W L A D H A V - - S G T V L L P G A A L V E
Identity
1. DH-2     H P L L T A A V E L P E A D A V T F T G R L S R A D R P W L A D H A V - - D G T V L L A G T A L V E
2. DH-3     H P L L G A A V E L P G T G G V A F T G R W S V R T H P W L A D H A V - - W G T A L L P G T G F V D
3. DH-5     H P L L S A V M A V P E T G G V L C T G R L S R K A Q P W L V E D S V P G T G T G M V P G A A L V E
4. DH-6     H P L L D I V V G V P E T G G V V A T G R L S L A A Q P W L A D H A V - - S G R V L L P G A A L V E
5. DH-7     H A L L D S V I S V P D T G G V V A T G R L S L A A Q P W L A D H A V - - S G T V L L P G A A L V E
6. DH-8     H A L L D S V I S V P D T G G V V A T G R L S L A A Q P W L A D H A V - - S G T V L L P G A A L V E
7. DH-MYC2  H P L L T A A V D L P D T G A V A F T G R L S Q A D R P W L A D H A V - - D G T V L L A G T A L V E

                              60                  70                  80                  90                  100
Consensus   L A V R A G D E V G A G X L R E L V I E A P L V L P X D G A V R V Q V S V G E X D X S G V R S V A V
Identity
1. DH-2     M A V R A G D E V G H G A V R E L V L A A P L V L P A D G A V R I R V H V G A P D E T G D R P V S V
2. DH-3     L V L T A G A E A D C G A L E E L V I E A P L I L P E D G S I Q V R V E I T E P D E S G R R A V S V
3. DH-5     L A I R A G D E V G V G T V R G L T I E T P I V L P A T G A L R V Q V S V G G A D D E G H R T V G V
4. DH-6     L A I R A G D E V G A D V L R E L V I E A P L V V P D E G A V R V Q V S V G E D S G S G V R S V A V
5. DH-7     L A V R A G D E V G A R S L R E L I I E A P L V I P D R G A V R I Q V S V G E D S G S G V R S V A V
6. DH-8     L A V R A G D E V G A R S L R E L I I E A P L V I P D R G A V R I Q V S V G E D S G S G V R S V A V
7. DH-MYC2  M A V R A G D E V G H G T V R E L V L D A P L A L P A D G A V R V R V H V G A P D E S G D R P V T V

                              110                 120                 130                 140                 150
Consensus   Y S R P D - - G A D S X W T R H A S G Q L - X X X X V P A P G A G X G V W P P E G A T X V X V - - T
Identity
1. DH-2     H S R P D - - G D D G A W T R H A N G A L T R T N P G D A P G I D T T P W P P P N A T A V P V - - T
2. DH-3     H S R P E - - G D E P G W T R H A T G T L L P E D L - - D S S E P T L A W P P A D A T P V P V D A A
3. DH-5     Y T R A E - - Q G D E P W T R H A T G T L T P S G T V P V P S G G A A A Q
4. DH-6     Y S R A D D A D A D S A W V R H A S G Q L - - G G N V P A A G A G L G V W P P E G A V A V D V - - S
5. DH-7     Y S R A D D A D A D S A W V R H A S G Q L - - G G N V P A A G A G L G V W P P E G A V A V D V - - S
6. DH-8     Y S R P D D T E A G S P W V R H A S G Q L - - D D D A P A L D V A F G V W P P E G A Q P V D L - - T
7. DH-MYC2  H S R P D - - G A D T P Y T R H A S G V L A P T D P H D A P G A D H A P W P P A N A T P V P V - - T

            150                 160                 170
Consensus   T X F Y A X L A E X G Y E Y G P A F Q G L
Identity
1. DH-2     T D F Y R Q R A E A G Y E Y G P A F Q G L
2. DH-3     A S V Y A D L A E R G Y E Y G P A F Q G L
3. DH-5
4. DH-6     S H L Y A E L G A R G Y G Y G P V F Q G L
5. DH-7     S H L Y E G L A D R G Y G Y G P A F Q G L
6. DH-8     T G F Y A D Q A S A G Y A Y G P L F Q G L
7. DH-MYC2  T D F Y R E R A E A G Y E Y G P A F Q G L
```

Figure S40. Amino acid alignment of the *tym* BGC KR domains

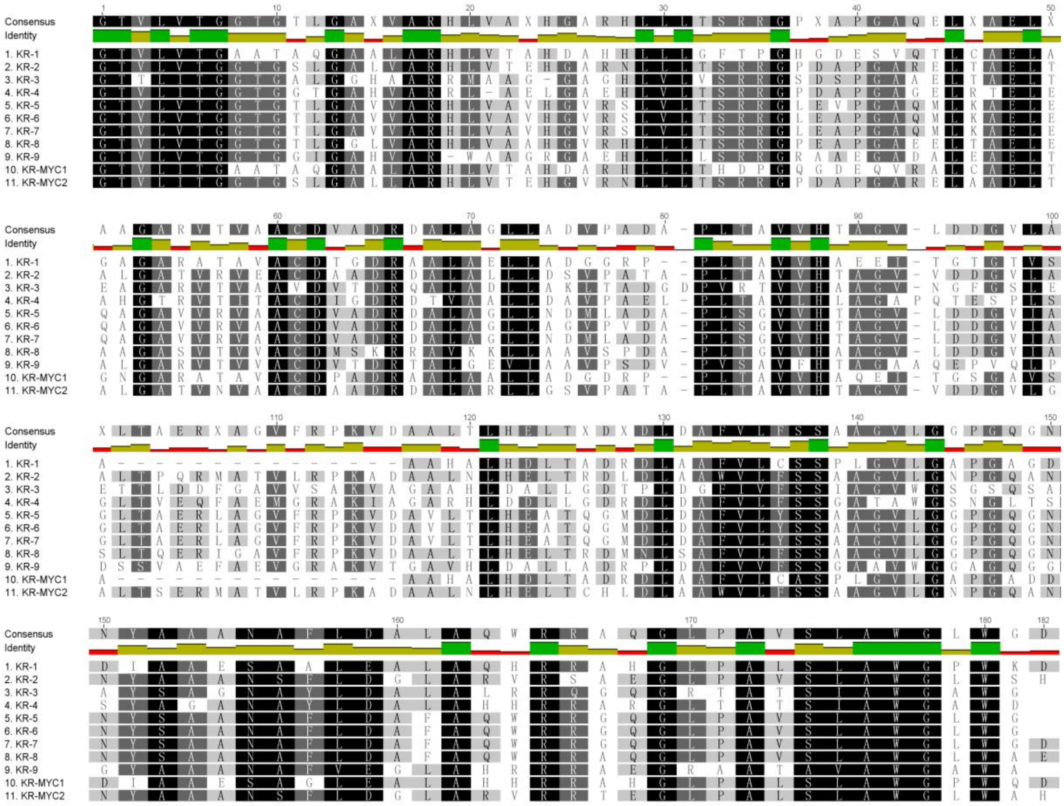


Figure S41. Amino acid alignment of the *tym* BGC ER domains

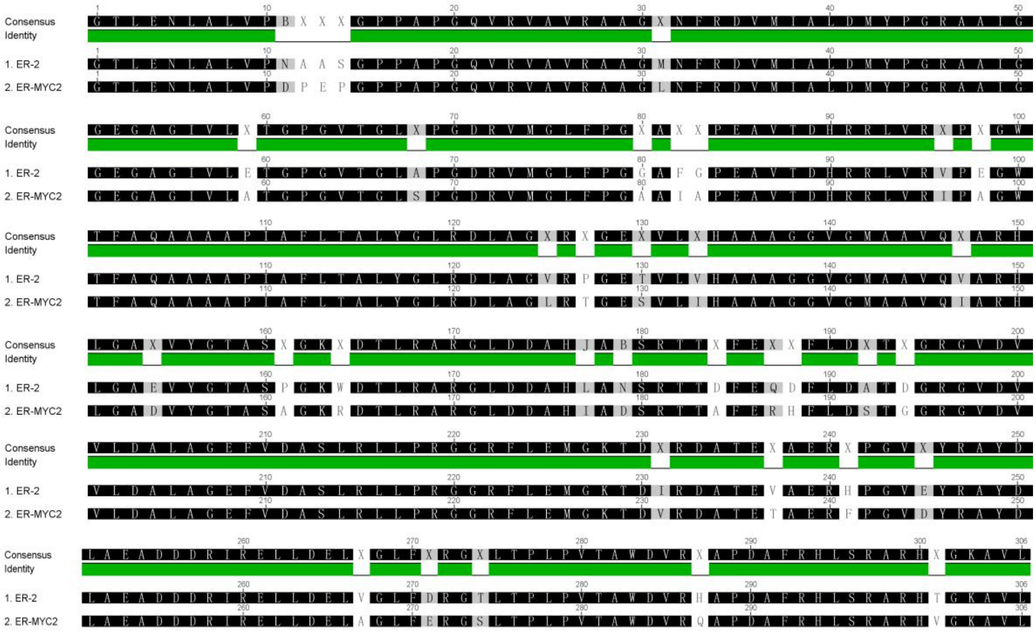


Figure S41. Amino acid alignment of the *tym* BGC KS domains