

## ***Supplementary Material***

### **Taeanamides A and B, nonribosomal lipo-decapeptides isolated from an intertidal mudflat-derived *Streptomyces* sp.**

**Jinsheng Cui<sup>1</sup>, Eunji Kim<sup>1</sup>, Dong Hyun Moon<sup>1</sup>, Tae Ho Kim<sup>2</sup>, Ilnam Kang<sup>3</sup>, Yeonjung Lim<sup>3</sup>, Daniel Shin<sup>1</sup>, Sunghoon Hwang<sup>1</sup>, Young Eun Du<sup>1</sup>, Myoung Chong Song<sup>1</sup>, Munhyung Bae<sup>4</sup>, Jang-Cheon Cho<sup>3</sup>, Jichan Jang<sup>2</sup>, Sang Kook Lee<sup>1</sup>, Yeo Joon Yoon<sup>\*1</sup>, Dong-Chan Oh<sup>\*1</sup>**

<sup>1</sup> Natural Products Research Institute, College of Pharmacy, Seoul National University, Seoul 08826, Republic of Korea

<sup>2</sup> Molecular Mechanism of Antibiotics, Division of Life Science, Division of Bio & Medical Big Data Department (BK4 Program), Research Institute of Life Science, Gyeongsang National University, Jinju, Gyeongnam 52828, Republic of Korea

<sup>3</sup> Department of Biological Sciences, Inha University, Incheon 22212, Republic of Korea

<sup>4</sup> College of Pharmacy, Gachon University, Incheon 21936, Republic of Korea

**\* Correspondence:**

Yeo Joon Yoon  
yeojoonyoon@snu.ac.kr

Dong-Chan Oh  
dongchanoh@snu.ac.kr

## Table of Contents

**Figure S1.**  $^1\text{H}$  NMR spectrum of taeanamide A (**1**) at 800 MHz in DMSO- $d_6$ .

**Figure S2.**  $^{13}\text{C}$  NMR spectrum of taeanamide A (**1**) at 200 MHz in DMSO- $d_6$ .

**Figure S3.** COSY NMR spectrum of taeanamide A (**1**) at 800 MHz in DMSO- $d_6$ .

**Figure S4.** HSQC NMR spectrum of taeanamide A (**1**) at 800 MHz in DMSO- $d_6$ .

**Figure S5.** HMBC NMR spectrum of taeanamide A (**1**) at 800 MHz in DMSO- $d_6$ .

**Figure S6.** ROESY NMR spectrum of taeanamide A (**1**) at 800 MHz in DMSO- $d_6$ .

**Figure S7.** TOCSY NMR spectrum of taeanamide A (**1**) at 800 MHz in DMSO- $d_6$ .

**Figure S8.**  $^1\text{H}$  NMR spectrum of taeanamide B (**2**) at 800 MHz in pyridine- $d_5$ .

**Figure S9.**  $^{13}\text{C}$  NMR spectrum of taeanamide B (**2**) at 200 MHz in pyridine- $d_5$ .

**Figure S10.** COSY NMR spectrum of taeanamide B (**2**) at 800 MHz in pyridine- $d_5$ .

**Figure S11.** HSQC NMR spectrum of taeanamide B (**2**) at 800 MHz in pyridine- $d_5$ .

**Figure S12.** HMBC NMR spectrum of taeanamide B (**2**) at 800 MHz in pyridine- $d_5$ .

**Figure S13.** ROESY NMR spectrum of taeanamide B (**2**) at 800 MHz in pyridine- $d_5$ .

**Figure S14.** TOCSY NMR spectrum of taeanamide B (**2**) at 800 MHz in pyridine- $d_5$ .

**Figure S15.** CD spectra of taeanamides A (**1**) and B (**2**).

**Figure S16.** HR-MS/MS data of taeanamide A (**1**) and B (**2**) obtained from a Waters XEVO® G2S Q-TOF mass spectrometer.

**Table S1.** LC/MS analysis of L- and D-FDAA derivatives of the amino acids in taeanamide A (**1**).

**Table S2.** LC/MS analysis of L- and D-FDAA derivatives of the amino acids in taeanamide B (**2**).

**Table S3.** LC/MS analysis of L- and D-FDAA derivatives of L-2,4-diamino butanoic acid authentic sample.

**Table S4.** Putative functions of ORFs of the taeanamides biosynthetic gene cluster in *Streptomyces* sp. AMD43.

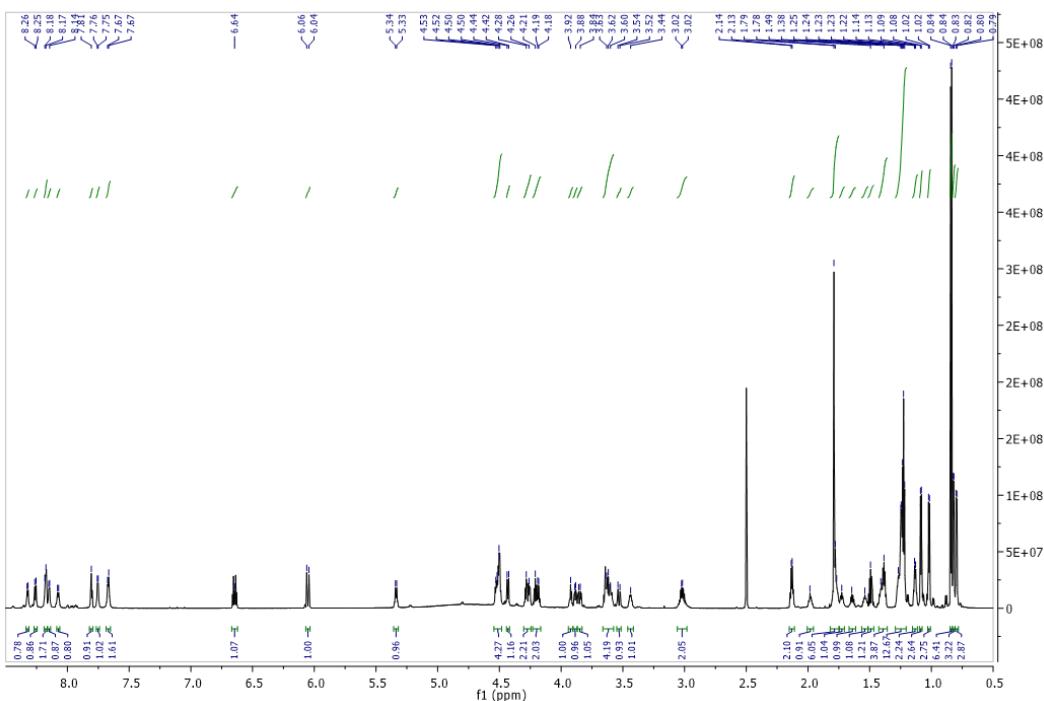
**Table S5.** Adenylation (A) domain substrate specificity predictions of the taeanamide NRPS.

**Figure S17.** Multiple sequence alignment of condensation (C) domains of Taem NRPS and lipopeptide synthetases.

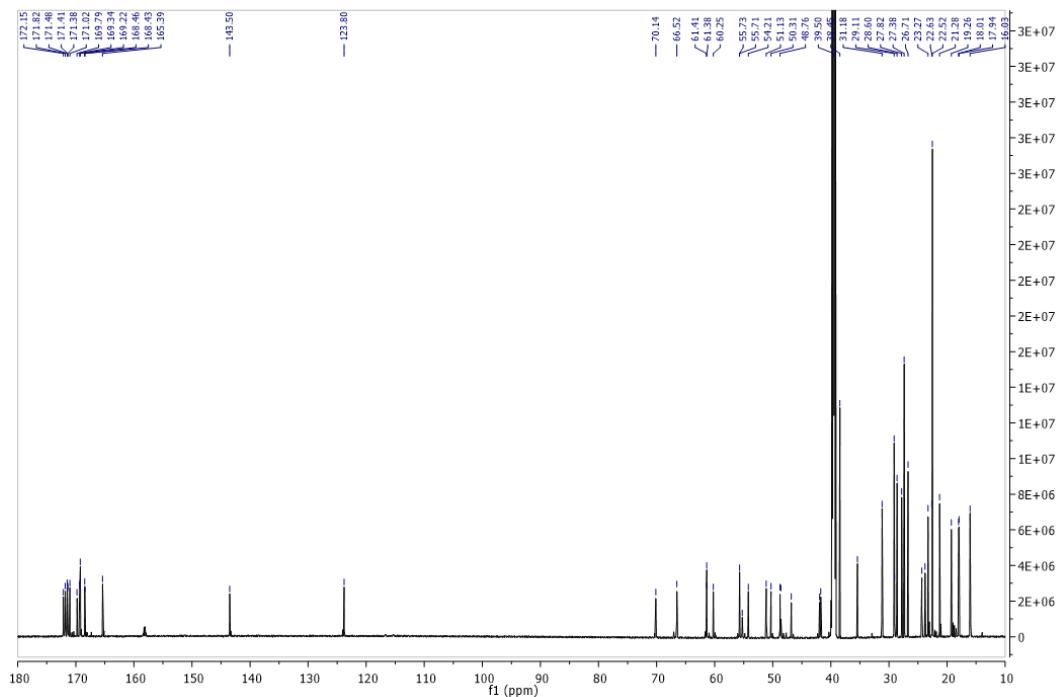
**Figure S18.** Phylogenetic analysis of condensation (C) domains of Taem NRPS and lipopeptide synthetases.

**Figures S19.** LC/MS profiles of (a) EtOAc extract of strain AMD43 showing the existence of both taeanamides A and B (**1** and **2**); (b) taeanamide A (**1**) after purification; (c) taeanamide A (**1**) after 10 days in MeOH (room temperature).

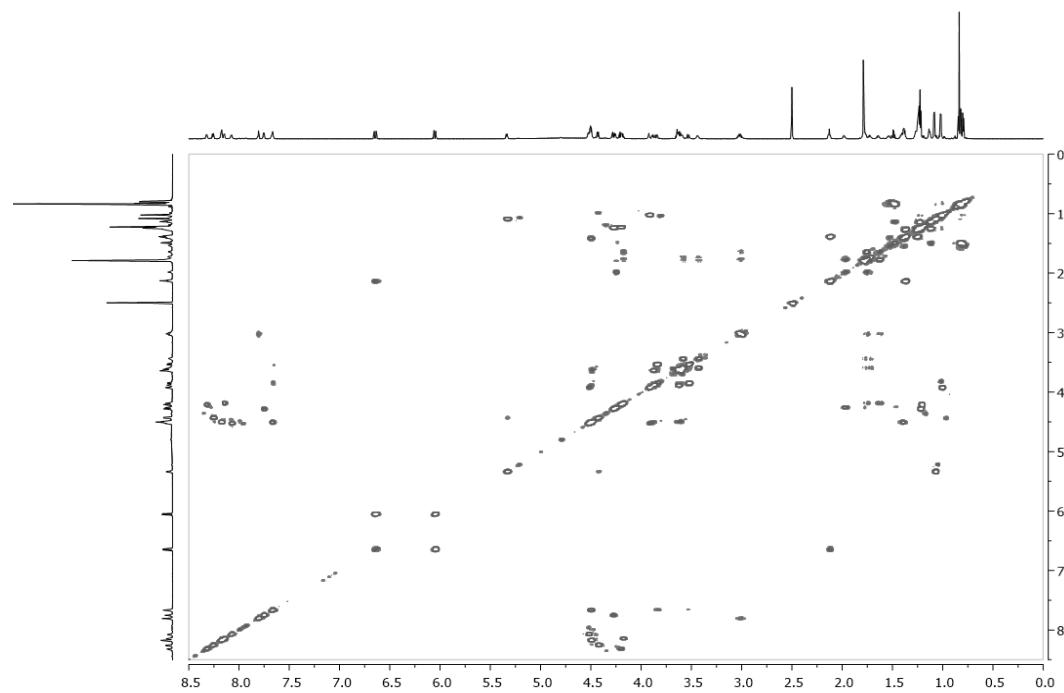
**Figure S1.**  $^1\text{H}$  NMR spectrum of taeanamide A (**1**) at 800 MHz in  $\text{DMSO}-d_6$ .



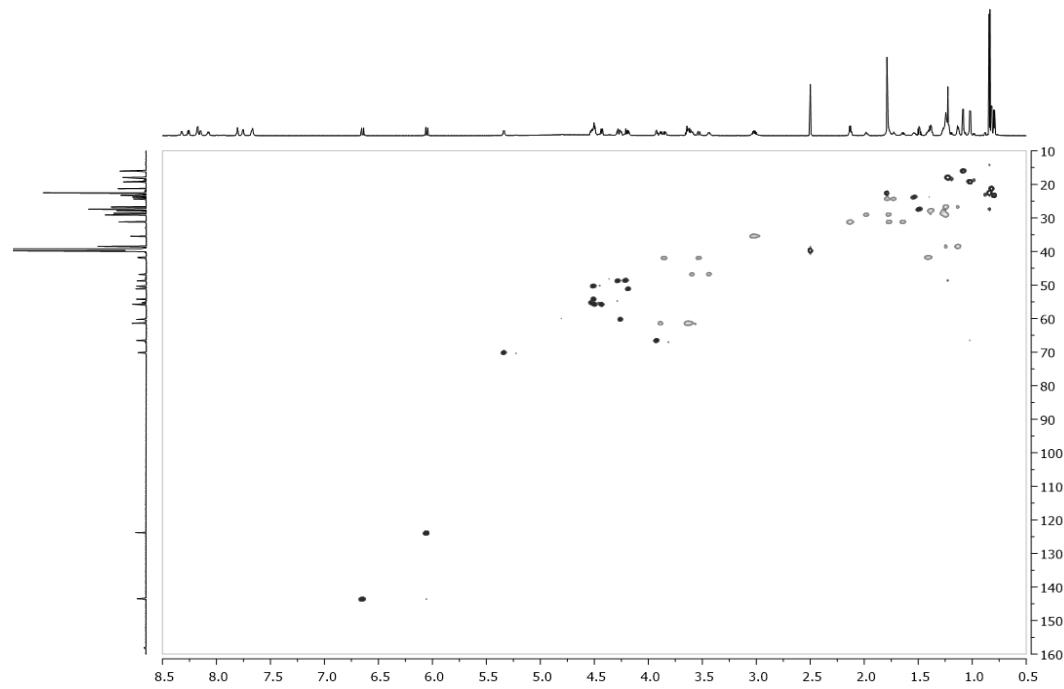
**Figure S2.**  $^{13}\text{C}$  NMR spectrum of taeanamide A (**1**) at 200 MHz in  $\text{DMSO}-d_6$ .



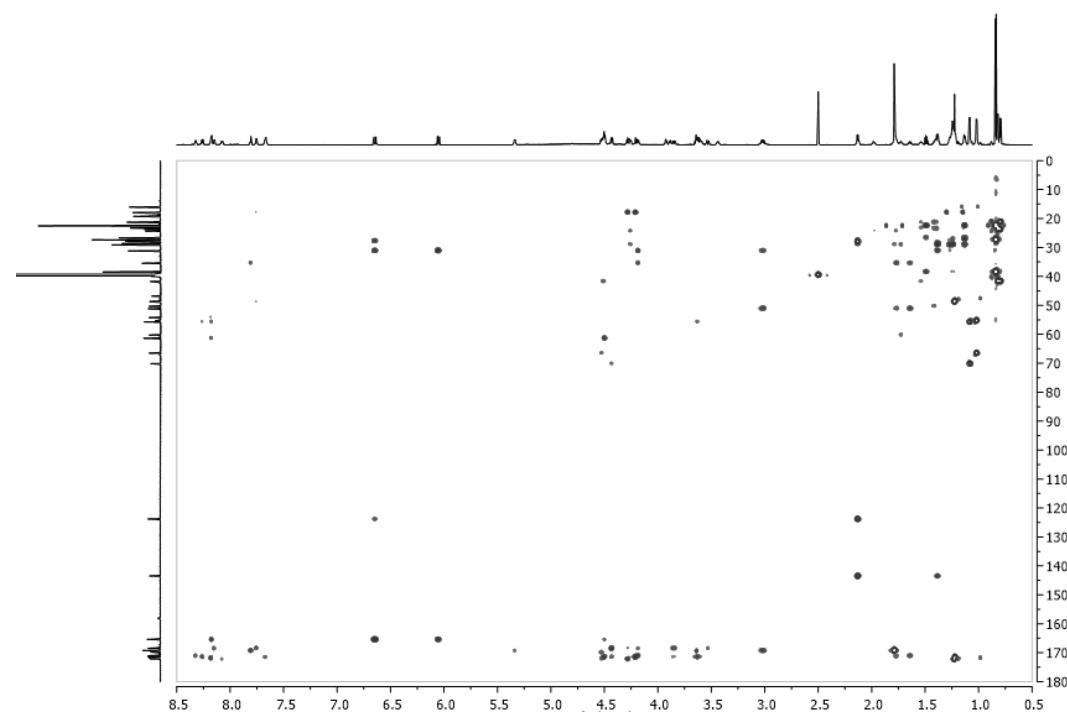
**Figure S3.** COSY NMR spectrum of taeanamide A (**1**) at 800 MHz in DMSO-*d*<sub>6</sub>.



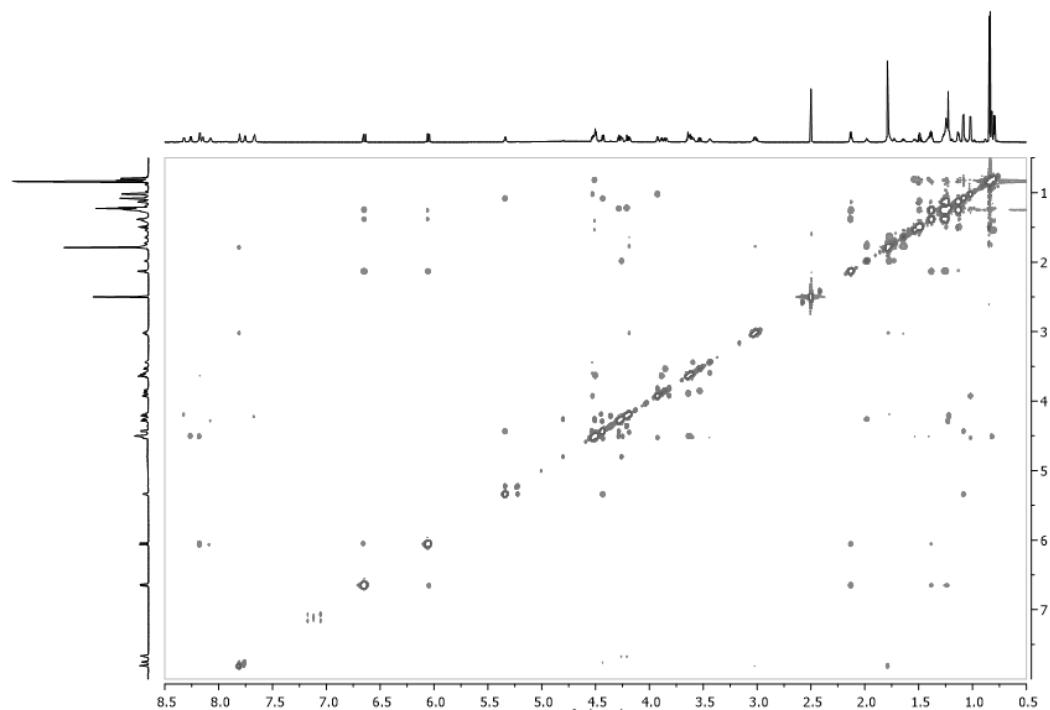
**Figure S4.** HSQC NMR spectrum of taeanamide A (**1**) at 800 MHz in DMSO-*d*<sub>6</sub>.



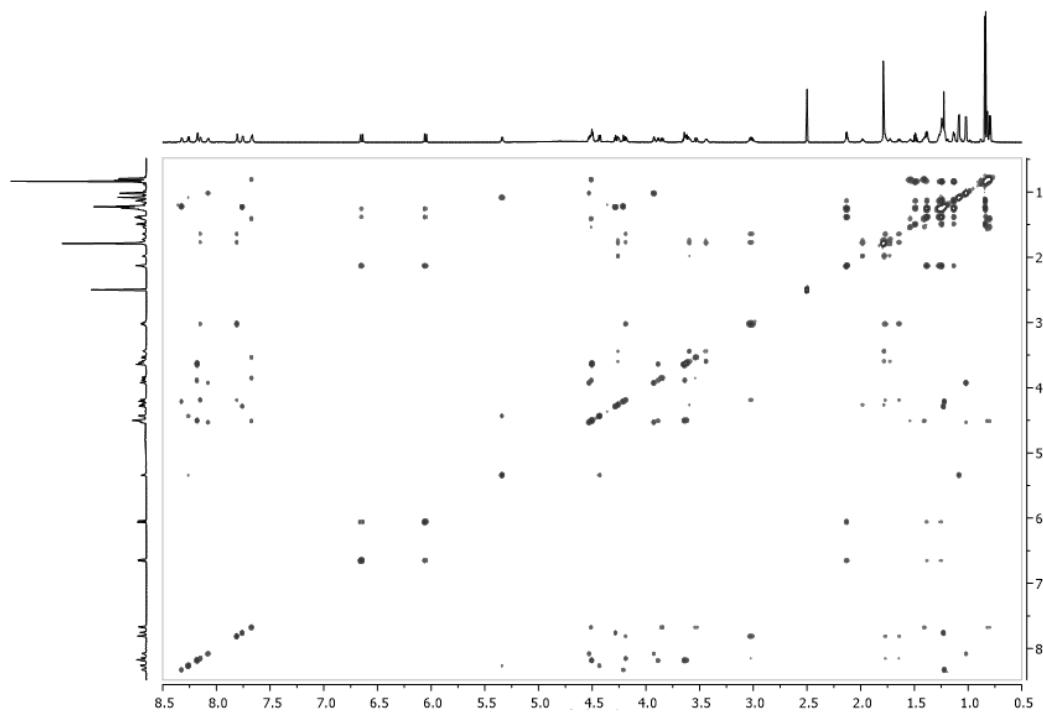
**Figure S5.** HMBC NMR spectrum of taeanamide A (**1**) at 800 MHz in DMSO-*d*<sub>6</sub>.



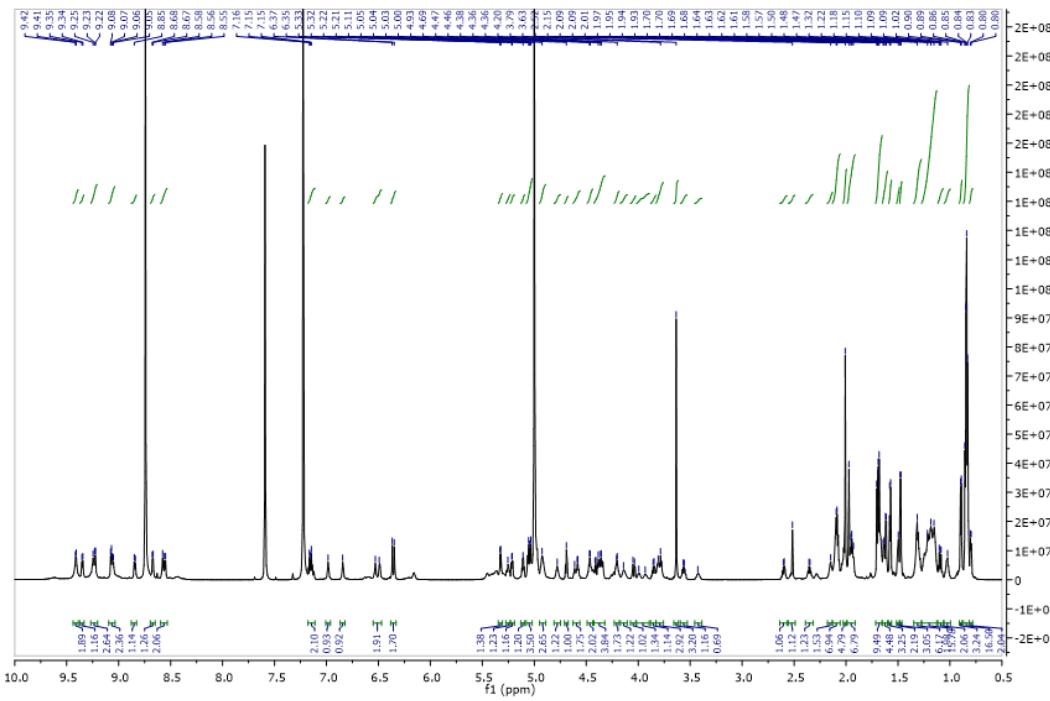
**Figure S6.** ROESY NMR spectrum of taeanamide A (**1**) at 800 MHz in DMSO-*d*<sub>6</sub>.



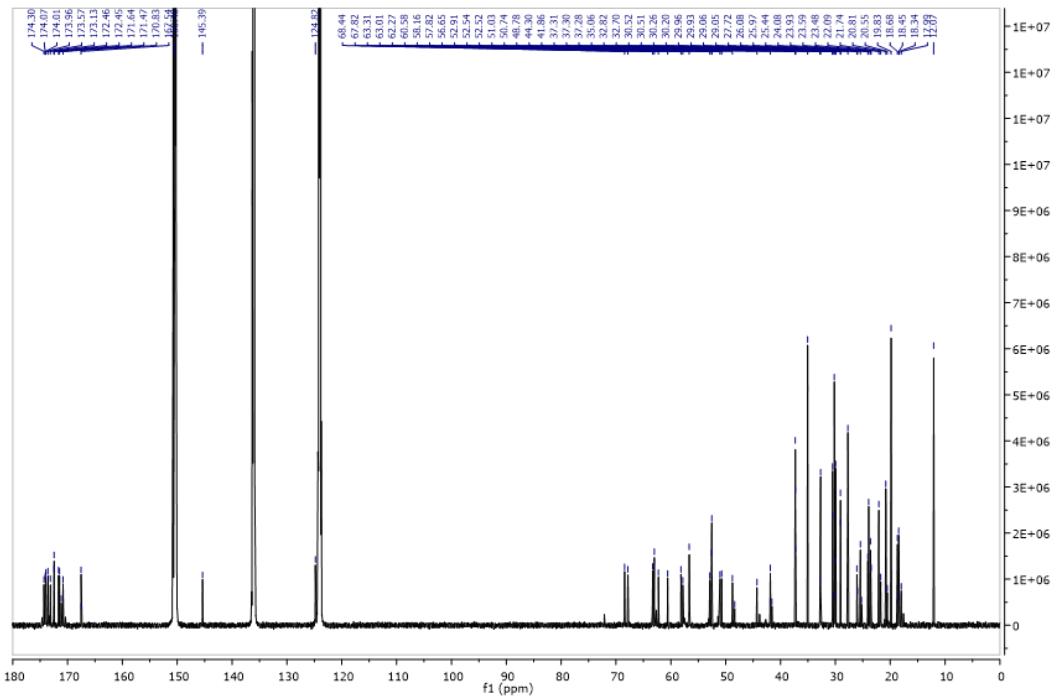
**Figure S7.** TOCSY NMR spectrum of taeanamide A (**1**) at 800 MHz in DMSO-*d*<sub>6</sub>.



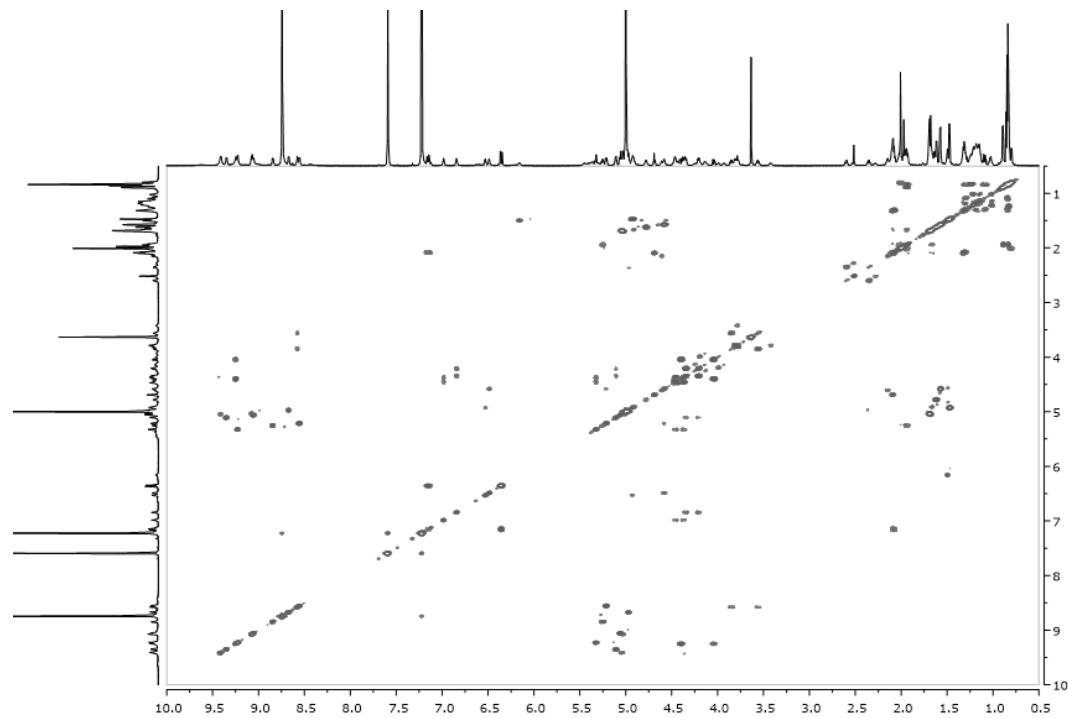
**Figure S8.** <sup>1</sup>H NMR spectrum of taeanamide B (**2**) at 800 MHz in pyridine-*d*<sub>5</sub>.



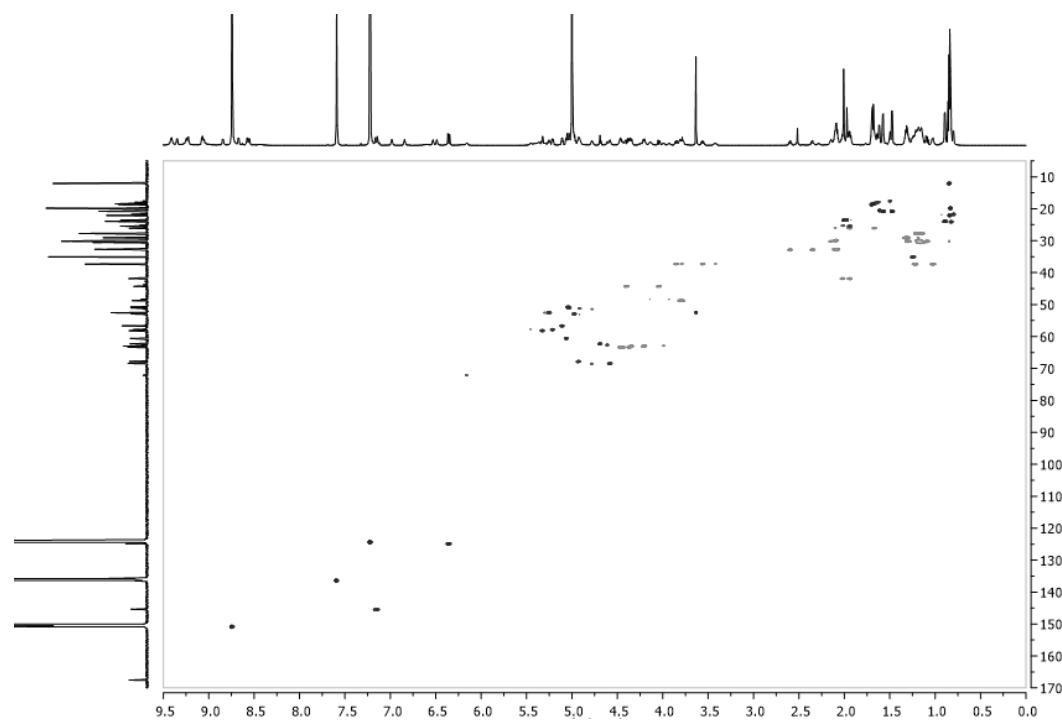
**Figure S9.** <sup>13</sup>C NMR spectrum of taeanamide B (**2**) at 200 MHz in pyridine-*d*<sub>5</sub>.



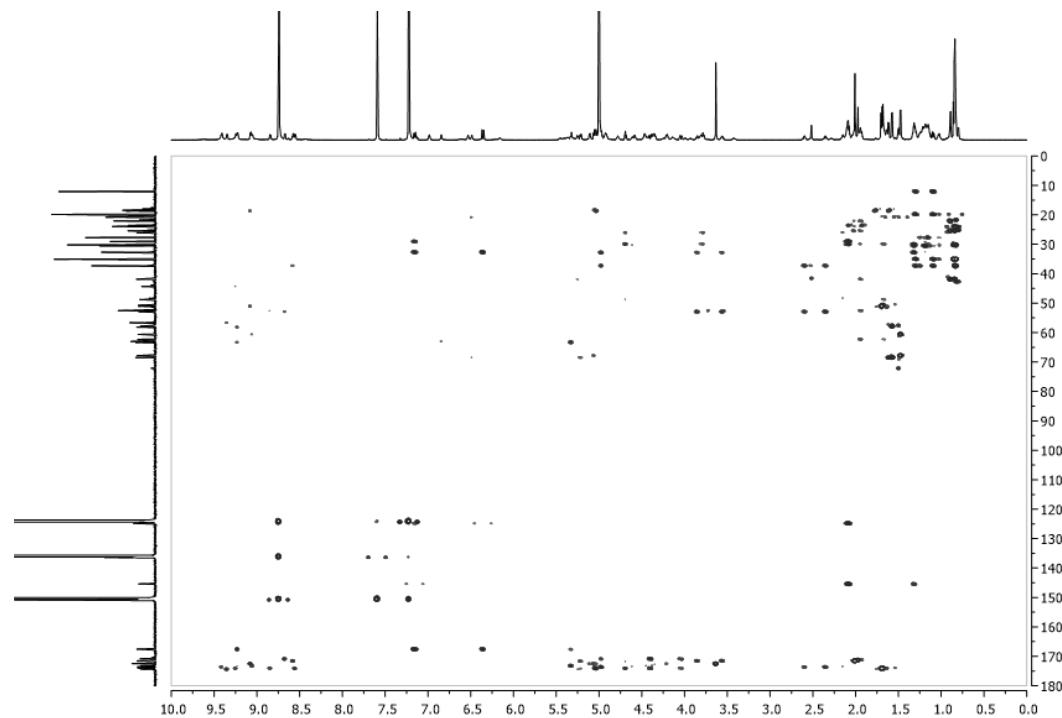
**Figure S10.** COSY NMR spectrum of taeanamide B (**2**) at 800 MHz in pyridine-*d*<sub>5</sub>.



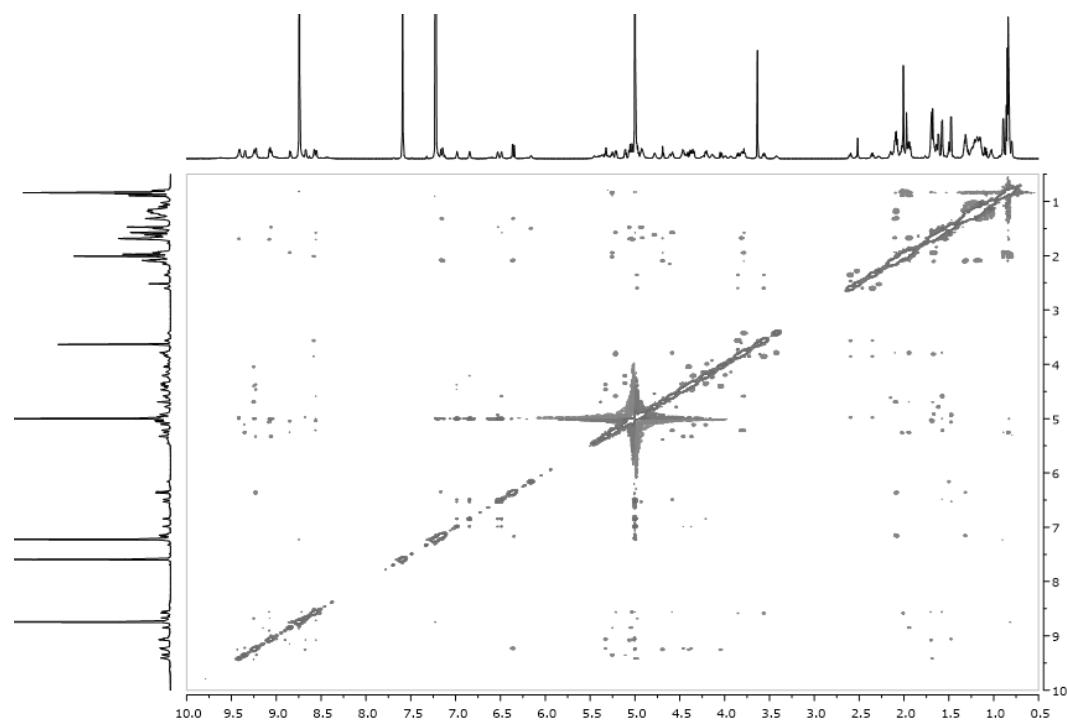
**Figure S11.** HSQC NMR spectrum of taeanamide B (**2**) at 800 MHz in pyridine-*d*<sub>5</sub>.



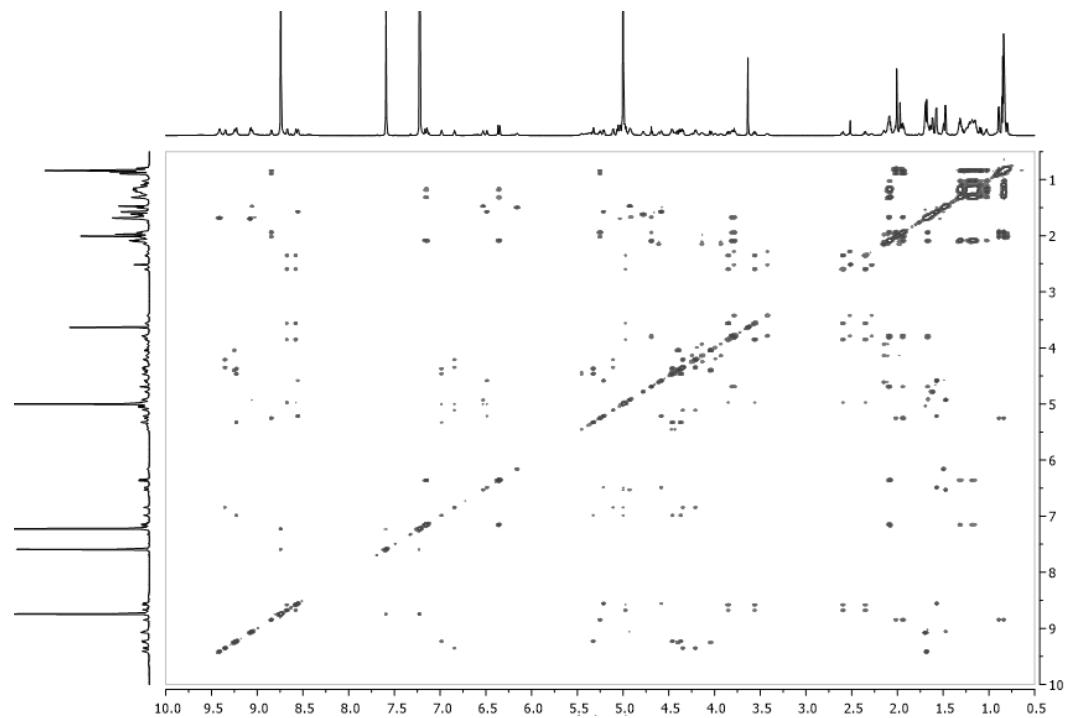
**Figure S12.** HMBC NMR spectrum of taeanamide B (**2**) at 800 MHz in pyridine-*d*<sub>5</sub>.



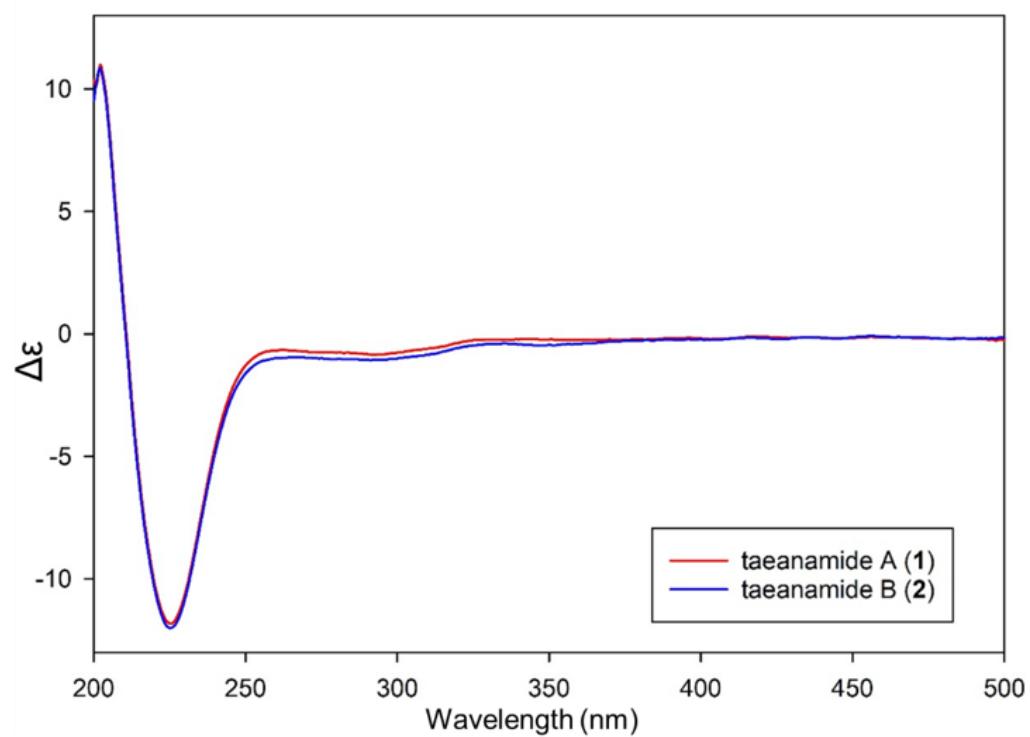
**Figure S13.** ROESY NMR spectrum of taeanamide B (**2**) at 800 MHz in pyridine-*d*<sub>5</sub>.



**Figure S14.** TOCSY NMR spectrum of taeanamide B (**2**) at 800 MHz in pyridine-*d*<sub>5</sub>.

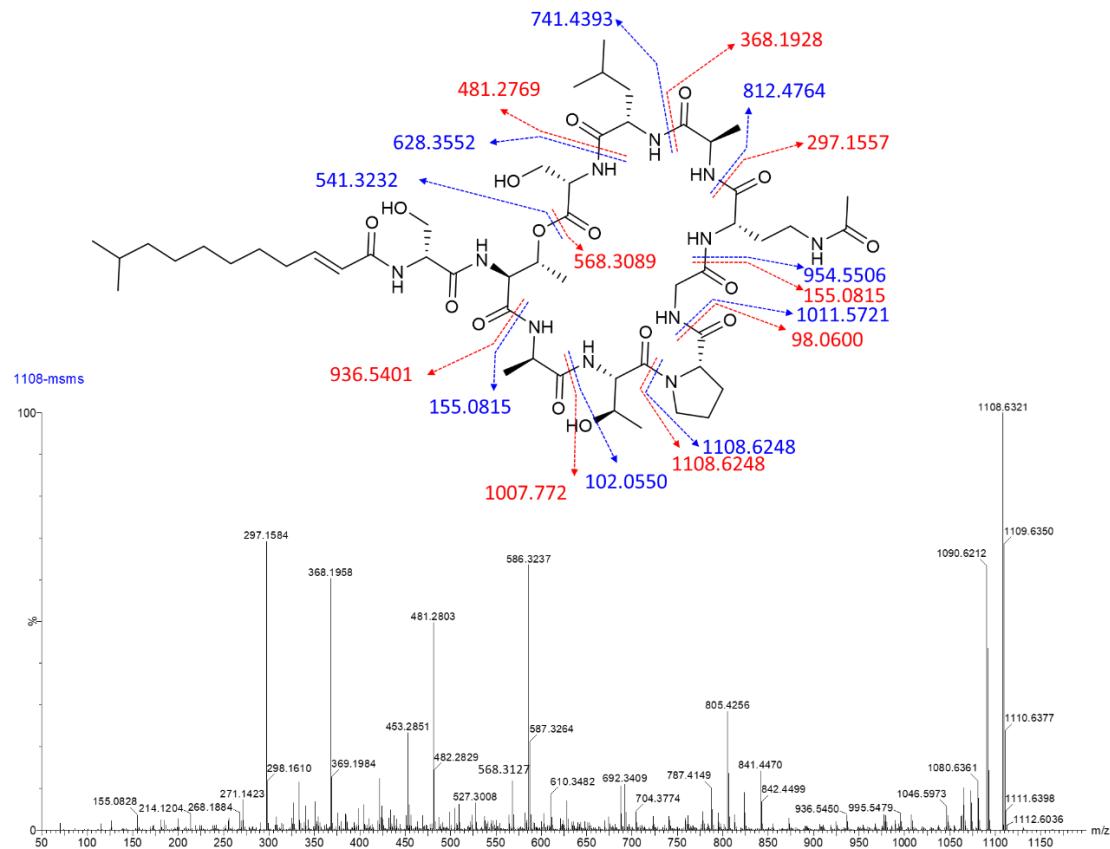


**Figure S15.** CD spectra of taeanamides A (**1**) and B (**2**).

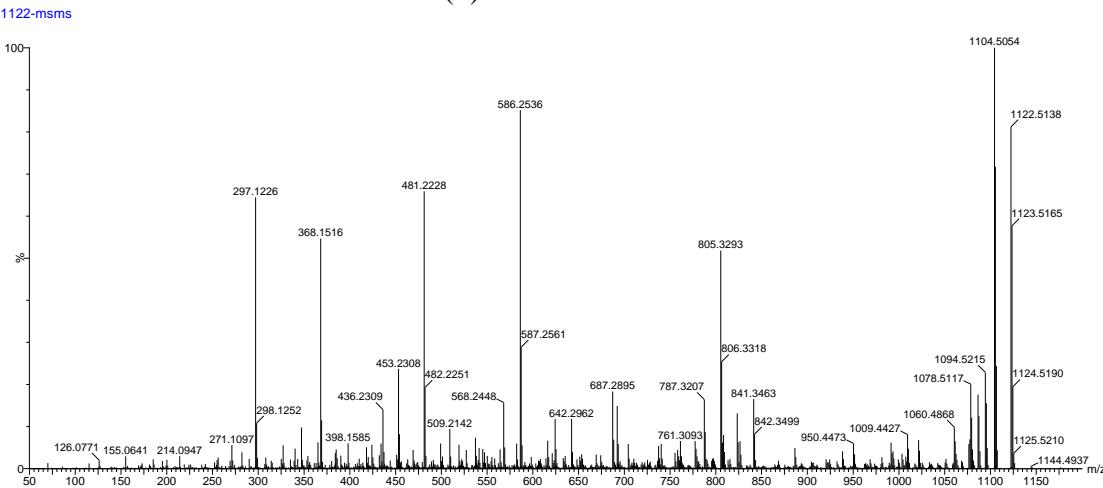


**Figure S16.** HR-MS/MS data of taeanamide A (**1**) and B (**2**) obtained from a Waters XEVO® G2S Q-TOF mass spectrometer.

(a) HR-MS/MS data of taeanamide A (**1**).



(b) HR-MS/MS data of taeanamide B (**2**).



**Table S1.** LC/MS analysis of L- and D-FDAA derivatives of the amino acids in taeanamide A (**1**).

Taeanamide A						
	Ala	Ser	Thr	Leu	Pro	<i>N</i> (4)-Acetyl-2,4-diaminobutyric acid → 2,4-diamino-butanoic acid
[M+H] <sup>+</sup> ( <i>m/z</i> )	342	358	372	384	368	623
Retention time (min)						
L-FDAA	26.5	18.5; 19.0	19.5	34.5	24.5	33.8
D-FDAA	23.6	18.5; 19.0	23.0	38.5	25.8	35.3
Elution order	D→L	N	L→D	L→D	L→D	L→D
Configuration	D	Not determined	L	L	L	L

**Table S2.** LC/MS analysis of L- and D-FDAA derivatives of the amino acids in taeanamide B (**2**).

Taeanamide B						
	Ala	Ser	Thr	Leu	Pro	<i>N</i> (4)-Acetyl-2,4-diaminobutyric acid → 2,4-diamino-butanoic acid
[M+H] <sup>+</sup> ( <i>m/z</i> )	342	358	372	384	368	623
Retention time (min)						
L-FDAA	26.1	17.9; 18.5	19.1	34.0	24.0	33.4
D-FDAA	23.0	18.0; 18.6	22.6	38.1	25.3	35.3
Elution order	D→L	N	L→D	L→D	L→D	L→D
Configuration	D	Not determined	L	L	L	L

**Table S3.** LC/MS analysis of L- and D-FDAA derivatives of L-2,4-diamino butanoic acid authentic sample.

	L-2,4-diamino butanoic acid authentic sample
+ L-FDAA	33.6
+ D-FDAA	35.3
Elution order	L→D

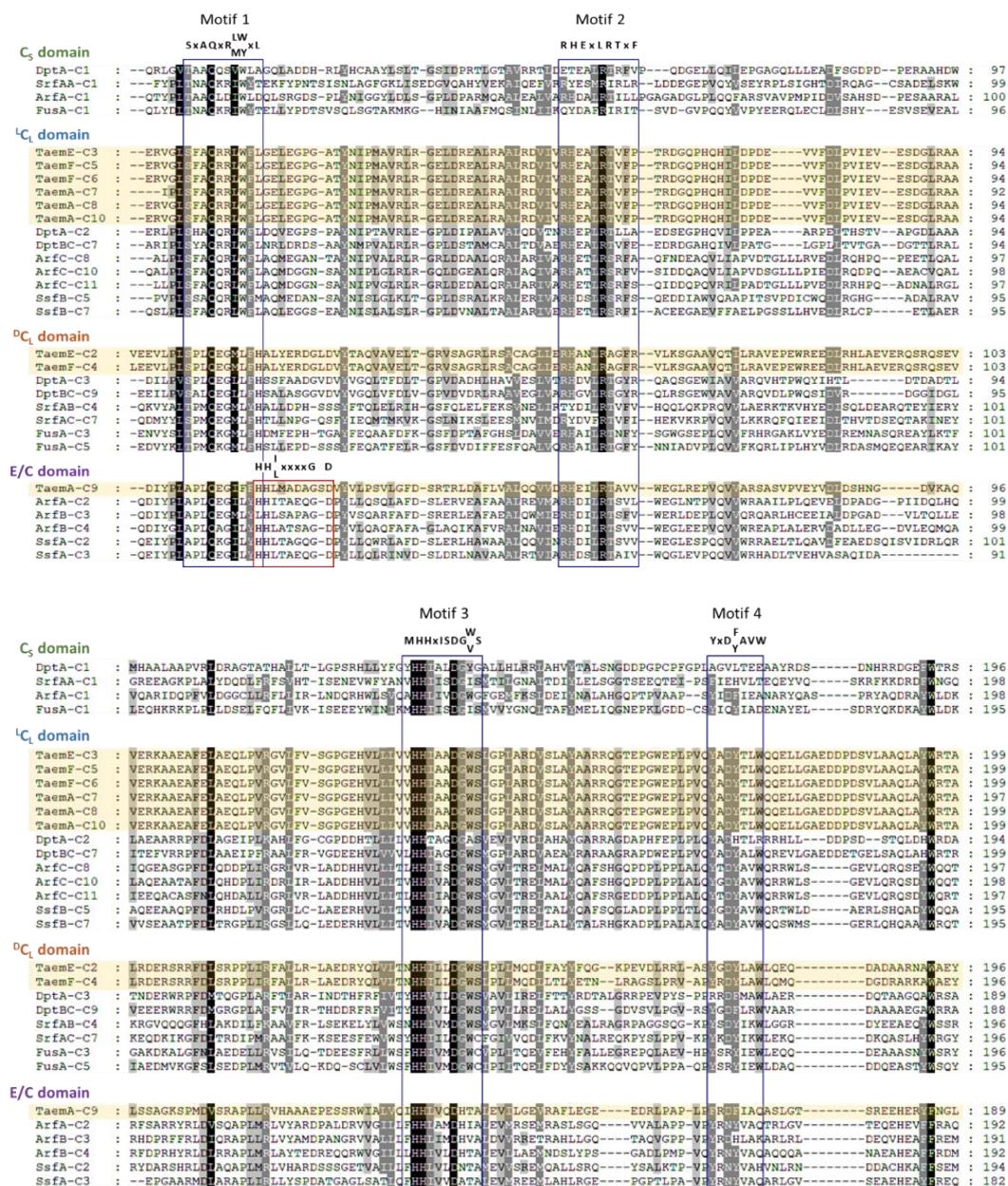
**Table S4.** Putative functions of ORFs of the taeanamides biosynthetic gene cluster in *Streptomyces* sp. AMD43.

ORF	Size (aa)	Putative function	Best match	
			Organism/GenBank (residue)	Identity/ Similarity (%)
Orf1	389	Hypothetical protein	<i>Streptomyces</i> sp. Ag109_O5-10 / WP_143063641.1 (389)	95/97
Orf2	435	N-6 DNA methylase	<i>Streptomyces flavoviridis</i> / WP_189596001.1 (435)	98/98
TaemA	4512	NRPS		
TaemB	486	Diaminobutyrate aminotransferase	<i>Streptomyces</i> sp. SAI-083 / WP_123988349.1 (423)	91/95
TaemC	265	Thioesterase	<i>Streptomyces</i> sp. S063 / WP_128818897.1 (256)	56/70
TaemD	70	MbtH protein	<i>Streptomyces chrestomyceticus</i> / WP_125043567.1 (76)	67/84
Orf3	92	Transposase	<i>Streptomyces</i> sp. Ag82_G6-1 / WP_097221223.1 (383)	75/86
Orf4	85	Transposase	<i>Streptomyces</i> sp. CB01580 / OKJ25938.1 (241)	75/79
Orf5	187	Transposase	<i>Streptomyces</i> sp. Ag109_O5-1 / WP_123981086.1 (374)	81/82
Orf6	725	Integral membrane protein	<i>Streptomyces tsukubensis</i> / WP_077968499.1 (736)	62/73
Orf7	67			
TaemE	3786	NRPS		
TaemF	3211	NRPS		
Orf8	394	Maturase	<i>Streptomyces</i> sp. F001 / WP_129803435.1 (608)	83/89
Orf9	570	Recombinase family protein	<i>Streptomyces</i> sp. AS58/ WP_079001718.1 (641)	97/98

**Table S5.** Adenylation (A) domain substrate specificity predictions of the taeanamide NRPS. The underlined amino acids in the substrate specific code are residues that do not match the Stachelhaus codes [1].

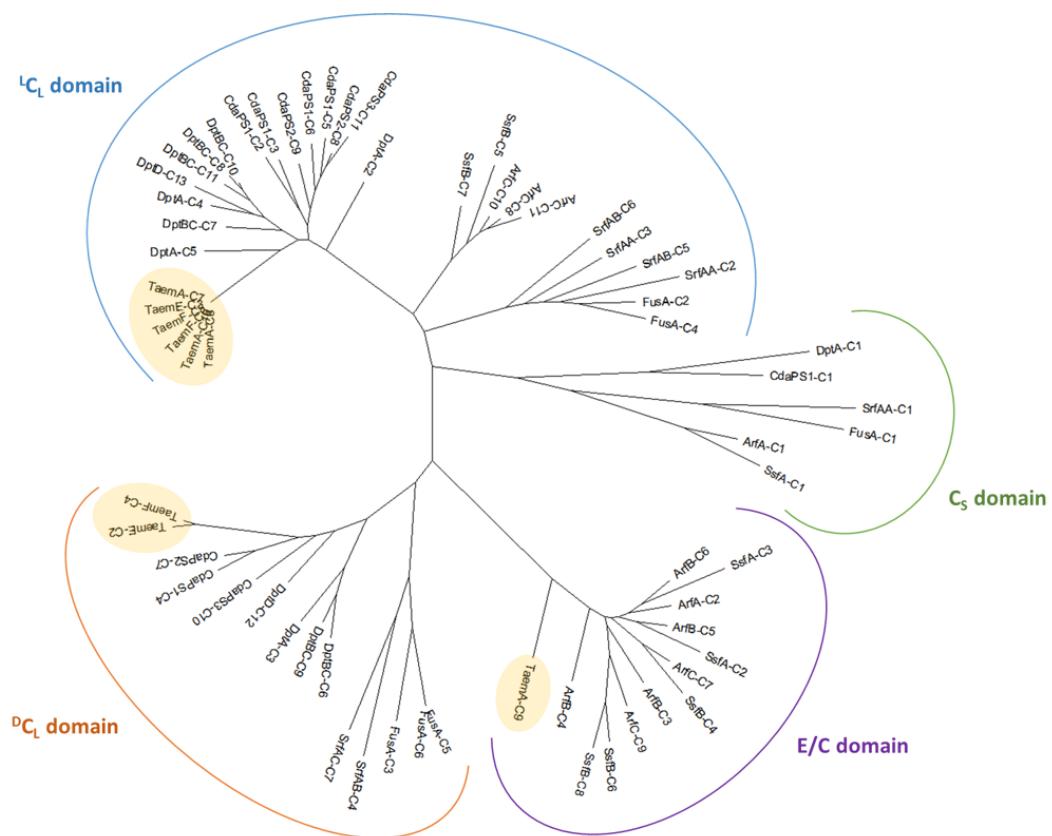
A domain	Substrate specific code	Stachelhaus code match	Predicted amino acid	Actual amino acid in taeanamide
A1	D V W H I S L V D K	100%	Ser	D-Ser
A2	D F W S V G M V H K	100%	Thr	L-Thr
A3	D V <u>R</u> H I S L V <u>E</u> K	80%	Ser	D-Ala
A4	D F W S V G M V H K	100%	Thr	L-Thr
A5	D V Q Y <u>A</u> A H V V K	90%	Pro	L-Pro
A6	D I L Q L G L I W K	100%	Gly	Gly
A7	D A <u>R</u> Q <u>I</u> G L <u>V</u> D K	70%	Gln	<i>N</i> -Ac-L-Dab
A8	D V <u>R</u> H I S L V <u>E</u> K	80%	Ser	D-Ala
A9	D A <u>L</u> L I G A V <u>A</u> K	80%	Val	L-Leu
A10	D V W H I S L V D K	100%	Ser	L-Ser

**Figure S17.** Multiple sequence alignment of condensation (C) domains of Taem NRPS and lipopeptide synthetases. Blue boxes indicate core motifs 1 through 5 of C domains [2]. It includes specific conserved residues at each motif according to C domain subtype. E/C domains contain a characteristic HH[I/L]xxxxGD motif at their N-terminus (Red box). Amino acid sequences of comparative C domains are extracted from daptomycin synthetase (*Streptomyces filamentosus* NRRL 11379, GenBank AY787762), surfactin synthetase (*Bacillus velezensis* FZB42, GenBank AJ575642), arthrobactin synthetase (*Pseudomonas* sp. MIS38, GenBank AB107223), fusaricidin synthetase (*Paenibacillus polymyxa*, GenBank EF451155), and syringafactin synthetase (*Pseudomonas syringae* pv. tomato str. DC3000, GenBank AE016853).

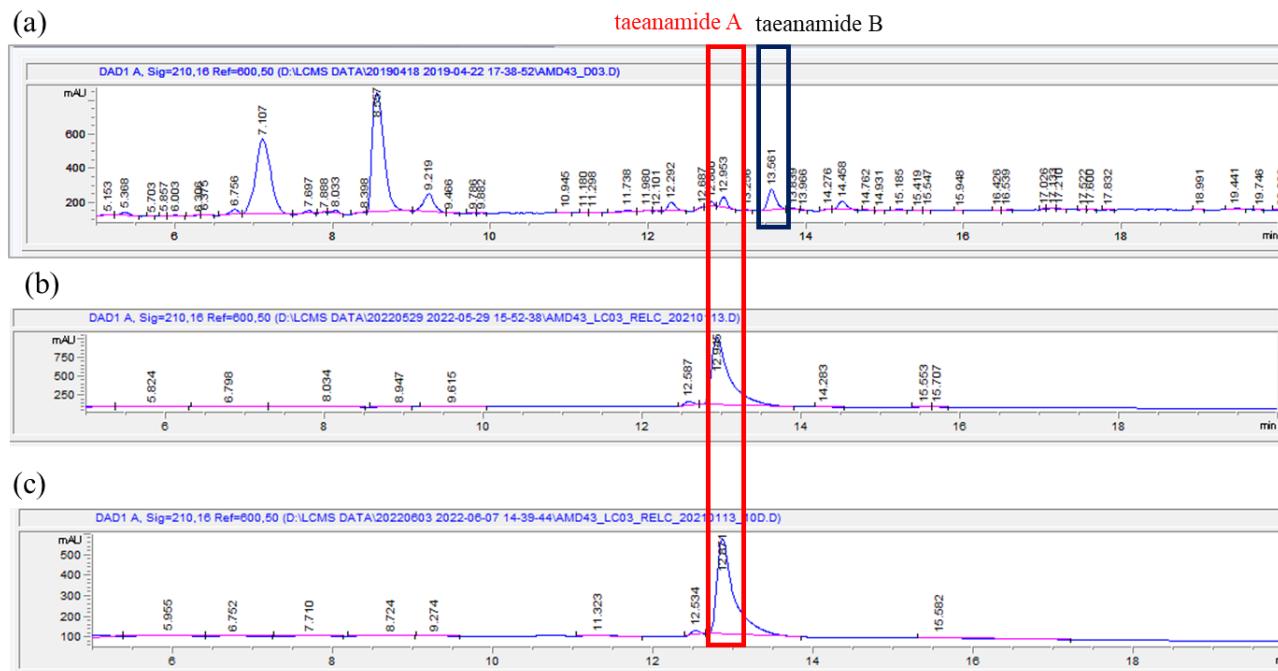


		Motif 5
		I Gx FVN I QC V L A * R
<b>C<sub>5</sub> domain</b>		
DptA-C1	: IAGADEAPGCLSE-----REAGALAVPLRPTVPSGERTKEDAAACATTGARWSSLIVPNTTAAPIREHAAADETVILPLTPATLG--PALRPGCNMLDNPFLDL-	: 294
SrfAA-C1	: FETVPELVSEK-----RSRADAGLDAKRFSQSDPHDLYGRHSFCEEHKVVLSPQSLLITYYKVTRGDDVIGCIPGNQINA--KEKQIEMMDSPLPRTSV	: 297
Arfa-C1	: YRCQFPEPLLEPRHR-QQSFLAGQTACTQVYAGCPAIIHENQCVRAFGAAFHFLLA-LHVYFSTTAQREWVQLPLMSQGA--RFRSFLIAQVSARNG-	: 299
Fusa-C1	: FSDLEPELTGWKS---YNPLSLSTHAVREHTFVPEVYHEHQAFCCQNRIEFLFOFFMGCMYIIHHMTNQFDVIGCISFANGQNK--KERQIGENESLAAFR	-- : 292
<b>C<sub>1</sub> domain</b>		
TaemE-C3	: IDGLPEELCLPTD--RPRPAIAATHHGGATTDLT[PACI]HOD[VEI]RRQNSL[FMV]QS[LAVI]SRLGAGED[PVG]P[AGCT]DD--ALDDVVFVNITIVLRRTDL	: 301
TaemF-C5	: IDGLPEELCLPTD--RPRPAIAATHHGGATTDLT[PACI]HOD[VEI]RRQNSL[FMV]QS[LAVI]SRLGAGED[PVG]P[AGCT]DD--ALDDVVFVNITIVLRRTDL	: 301
TaemF-C6	: IDGLPEELCLPTD--RPRPAIAATHHGGATTDLT[PACI]HOD[VEI]RRQNSL[FMV]QS[LAVI]SRLGAGED[PVG]P[AGCT]DD--ALDDVVFVNITIVLRRTDL	: 301
TaemA-C7	: IDGLPEELCLPTD--RPRPAIAATHHGGATTDLT[PACI]HOD[VEI]RRQNSL[FMV]QS[LAVI]SRLGAGED[PVG]P[AGCT]DD--ALDDVVFVNITIVLRRTDL	: 299
TaemA-C8	: IDGLPEELCLPTD--RPRPAIAATHHGGATTDLT[PACI]HOD[VEI]RRQNSL[FMV]QS[LAVI]SRLGAGED[PVG]P[AGCT]DD--ALDDVVFVNITIVLRRTDL	: 301
TaemA-C10	: IDGLPEELCLPTD--RPRPAIAATHHGGATTDLT[PACI]HOD[VEI]RRQNSL[FMV]QS[LAVI]SRLGAGED[PVG]P[AGCT]DD--ALDDVVFVNITIVLRRTDL	: 301
DptA-C2	: IAGLPEOLELPPTD--HTRPAVTRGEALIAFTVPEHTHTH[RAM]CAHGV[VFMV]Q[LAAT]SRHGAGHD[PIC]P[AGCT]DD--GTEDDVVFVNITIVLRNDV	: 296
DptBC-C5	: IAGAPAELTLPTD--RPPRPAVASTAGERVAFVTPAGHC[ADL]RAHGA[VFMV]Q[LAAT]SRLGAGGD[PIC]P[AGCT]DD--ATEEDVVFVNITIVLRCDV	: 301
Arfc-C8	: IAGAPALLMLPTD--RPPRPAQDYGASSVAVVDERISAGK[KALC]QRGRAVLYNTMS[WAM]SRLSGQSD[VIGC]P[AN]GTRA--EIEGIDVVFVNITIA[R]RIDT	: 299
Arfc-C10	: IAGAPALLTLPTD--RPPRPAQDYGASSVAVVDERISAGK[KALC]QRGRAVLYNTMS[WAM]SRLSGQSD[VIGC]P[AN]GTRA--EIEGIDVVFVNITIA[R]RIDT	: 300
Arfc-C11	: IAGAPALLTLPTD--RPPRPAQDYGSSRVVLDCHISAGK[KALC]QRGRAVLYNTMS[WAM]SRLSGQSD[VIGC]P[AN]GTRA--EIEGIDVVFVNITIA[R]RIDT	: 299
SsfB-C5	: TGAPVPLLTLPPTD--RPPRPAHGDYSGASVALDARSTDRTFCQAQSVPPMIFMG[WAVI]SRLSGQEE[VIGC]P[AN]GTRA--EIEGIDVVFVNITIA[R]RIDT	: 297
SsfB-C7	: TGAPVPLLTLPPTD--RPPRPAQCDFFAGASLAVRNENSO[TAG]RAT[QCGV]LYNTMS[WAM]SRLSGQEE[VIGC]P[AN]GTRA--EIEGIDVVFVNITIA[R]RIDT	: 297
<b>D<sub>C</sub> domain</b>		
TaemE-C2	: IAGVVEPTRMAE----ASGARTLLPQEIQHTAEVESVVRLPEGV[RARGV]VNTE[QG]WAVI[GAVTGRED]VIGCIT[VSC]P[EEIAGIES]V[G]FVNITIVP[R]RDL	: 298
TaemF-C4	: IADICEPTRLAS----TSEARALAAPGSMDTTPTATAVRPEGV[RARGV]VNTE[QG]WAVI[GAVTGRED]VIGCIT[VSC]P[EEIAGIES]V[G]FVNITIVP[R]RDL	: 298
DptA-C3	: IAGLAEPTRVIAL----GTEGSSGVIP-EVLEEE[SEPT]SE[TSER]VAV[RGRGV]VAVSVQ[EW]WALVIGRIVGRD[VPG]LTV[GCG]P[AEVG]V[E]FVNITIVP[R]RDL	: 290
DptBC-C9	: TGLEEPSLWAP----GVSRDGVVP-AAFHGA[DGL]ISK[VAV]RGRGV[VAV]V[Q]WALVIGRIVGRD[VPG]VTV[GCG]P[AEVG]V[E]FVNITIVP[R]RDL	: 289
SrfB-C4	: IADFECPSPSLDPG---RLASEKKDYQNEEYSFVNDEP[VVAQ]QOT[NRHCV]GPN[FOAWGV]VSKNYTD[VPG]V[GCG]P[SEING]TTA[L]FVINIIPURIKI	: 299
SrfAC-C7	: IEGFEGCOTTFAE---QRKKQKDGYEPKELLFSP[PEAETKA]FTEL[KSOHT]LSTA[VAV]WSV[LSRY]QCGSD[IIGC]V[GCG]P[AEIKG]V[E]FVINIAPRVKL	: 299
Fusa-C3	: IAGYEQQTLP--CVGGASKGEGYVAEKLYN[VSR]TER[EV]RDAHV[MMNI]QSLWCIA[ORYNGSKD]VIGC[VSC]P[AEI]PGIDR[L]FVINIIPRWT	: 300
Fusa-C5	: IEDYDGNTVLEPKTKSQAREAGYVLEKHVL[GAS]TGRQDV[V]EKRNV[V]NT[V]Q[EW]GLI[ORYNASSD]V[S]G[VSC]P[AEIAGIEN]V[G]FVNITIVP[R]QSV	: 301
<b>E/C domain</b>		
TaemA-C9	: IADVEEPTAPPFG---LLDVVRGDGTDVSEAQCFDAV[AA]R[EK]CRRRL[VE]TATI[FH]WARI[TV]IASRDI[V]C[V]F[G]NHA[GAGADR]L[F]FINTIP[R]V--	: 290
Arfa-C2	: IAEVDEPTLPFG---LQE[VQ]GDDGRGIDEAQCFDAV[YQR]RTO[RQAGV]V[AS]H[W]WARI[LAATSGC]R[V]C[V]F[G]NQGGEGADR]L[F]FINSIP[R]IDV	: 295
Arfb-C3	: IADIDEPTLPCG---LQDVGGDGQGIEEALM[DTDFSR]R[G]CRRL[VE]V[AS]H[L]LAR[L]G[LS]GRTA[V]C[V]F[G]NMEAGEGEGQAL[F]FINTIP[R]DVL	: 294
Arfb-C4	: IGDIDEPTVAFG---VQDVHGHDGSTIVDSE[Q]DSSV[AQR]REG[RRL]CIAASLY[H]Q[WQ]VIAQVSGRE[V]C[V]F[G]NQGGEGADR]L[F]FINTIP[R]VSV	: 295
SsfA-C2	: IAEVVEPTLPFG---IHDVPA[DGSG]TIEEDR[RL]ND[ALR]R[EQ]CRRL[VE]V[AS]H[V]WARI[V]SLANR[D]V[G]C[V]F[G]NSG-EGAER[L]F]FINTIP[R]VD-	: 295
SsfA-C3	: IGDIDAPTL[PF]D---LIRDVQQGDSRTIEEAR[Q]PD[PA]L[RG]R[EQ]CRRL[VE]V[AS]H[H]LWGRV[IAAAT]GNPDR[V]C[V]F[G]LQGGAGADR]L[F]FINTIP[R]IDV	: 285

**Figure S18.** Phylogenetic analysis of condensation (C) domains of Taem NRPS and lipopeptide synthetases. Amino acid sequences of comparative C domains are extracted from daptomycin synthetase (*Streptomyces filamentosus* NRRL 11379, GenBank AY787762), surfactin synthetase (*Bacillus velezensis* FZB42, GenBank AJ575642), arthrobactin synthetase (*Pseudomonas* sp. MIS38, GenBank AB107223), fusaricidin synthetase (*Paenibacillus polymyxa*, GenBank EF451155), syringafactin synthetase (*Pseudomonas syringae* pv. tomato str. DC3000, GenBank AE016853), and CDA synthetase (*Streptomyces coelicolor* A3(2), GenBank AL645882).



**Figures S19.** LC/MS profiles of (a) EtOAc extract of strain AMD43 showing the existence of both taeanamides A and B (**1** and **2**); (b) taeanamide A (**1**) after purification; (c) taeanamide A (**1**) after 10 days in MeOH (room temperature).



## References

1. Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **1999**, *6*, 493-505.
  2. Rausch et al. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC. Evol. Biol.* **2007**, *7*, 78.