

Supporting Information

Bacterial biosynthesis and maturation of the didemnin anticancer agents

Ying Xu[†], Roland D. Kersten[‡], Sang-Jip Nam[‡], Liang Lu[†], Abdulaziz M. Al-Suwailem[§], Huajun Zheng^{||}, William Fenical[‡], Pieter C. Dorrestein^{‡,⊥}, Bradley S. Moore^{‡,⊥,*}, Pei-Yuan Qian^{†,*}

[†]KAUST Global Collaborative Research, Division of Life Science, School of Science, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China;

[‡]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093, USA;

[§]The Coastal and Marine Resources Core Lab, Red Sea Research Center, 4700 King Abdullah University of Science and Technology, Thuwal, Makkah 23955-6900, Kingdom of Saudi Arabia;

^{||}Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, 250 Bi Bo Road, Shanghai 201203, China;

[⊥]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, California 92093, USA

*Corresponding Authors: Pei-Yuan Qian, Phone: 852-2358-7331, Fax: 852-2358-1559, Email: boqianpy@ust.hk and

Bradley S. Moore, Phone: 858-822-6650, Fax: 858-534-1305, Email: bsmoore@ucsd.edu.

Supporting Information

Supplementary Figures

a FTMS - HF26-30

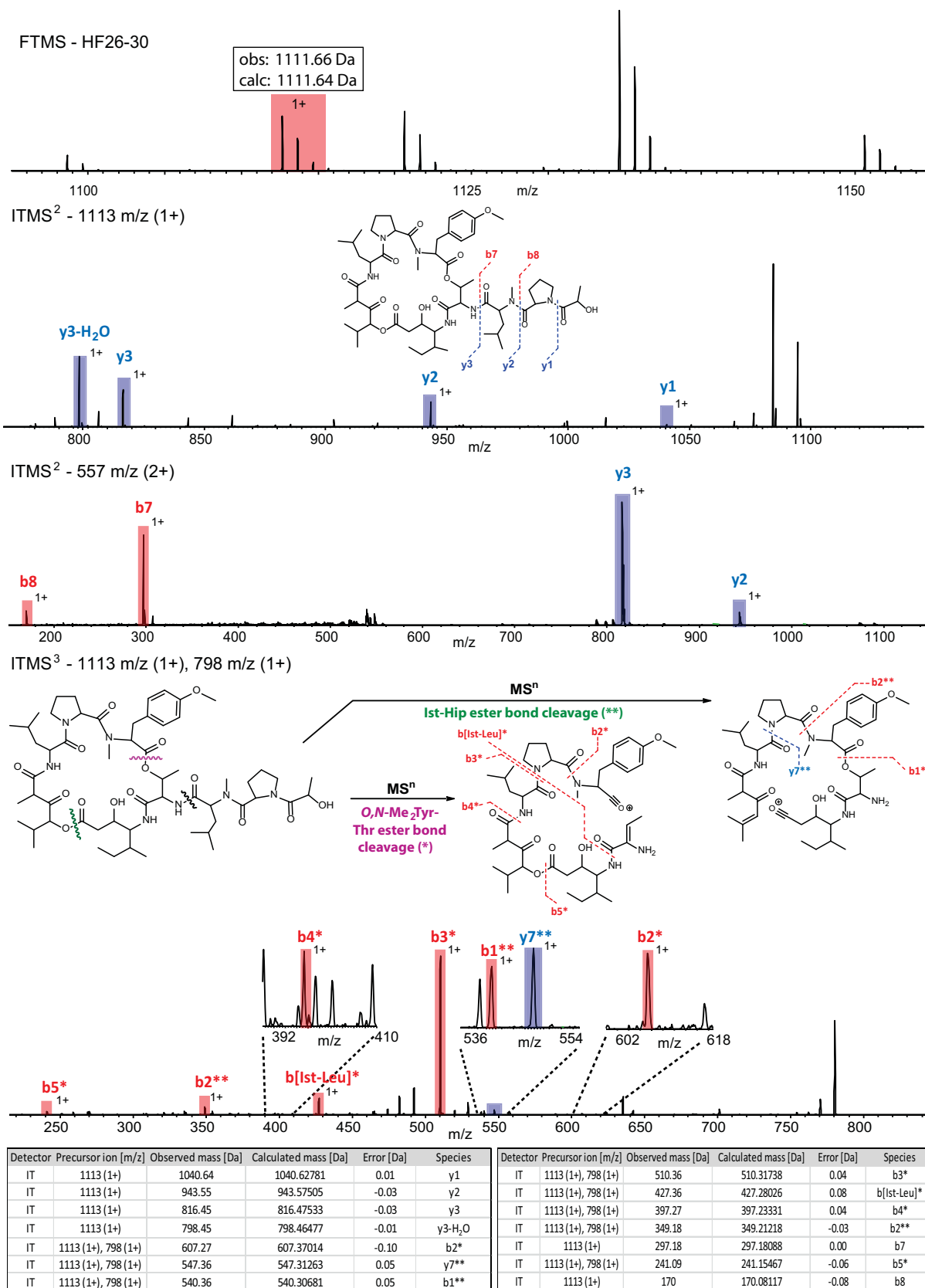


Fig. S1. Characterization of didemnin B. (a) MSⁿ analysis.

b

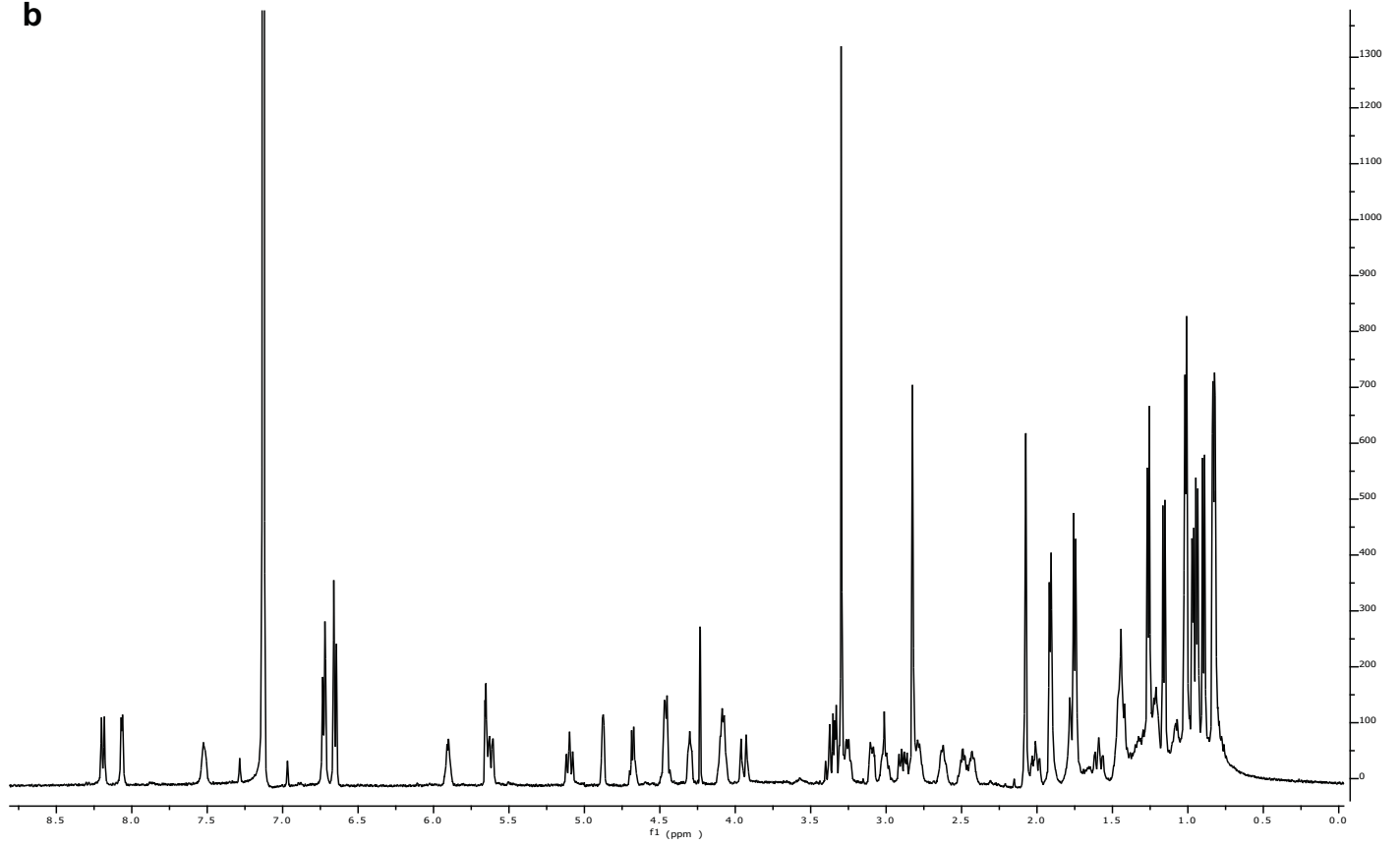
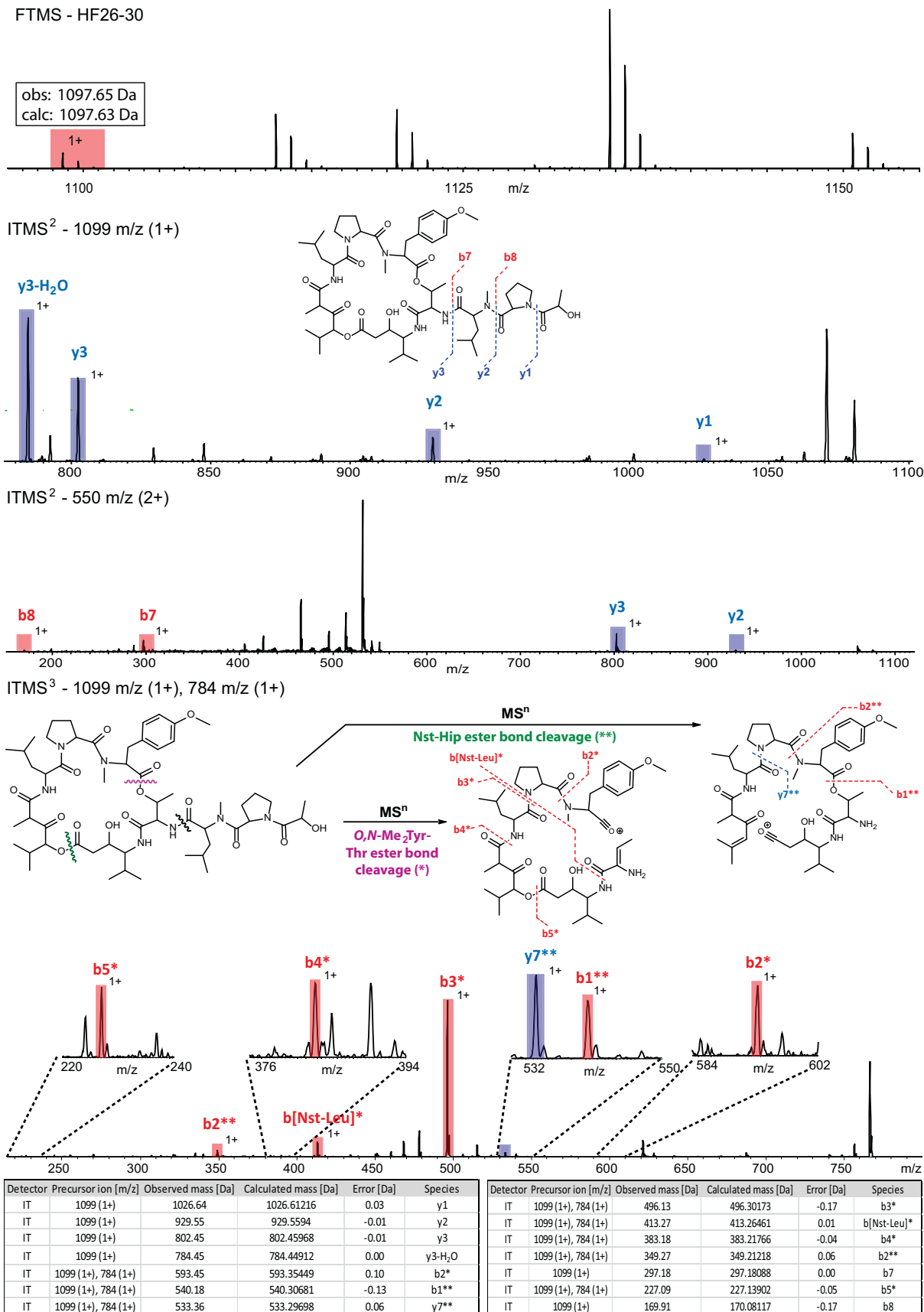


Fig. S1. Characterization of didemnin B. (b) ^1H NMR spectrum.

a

FTMS - HF26-30

Fig. S2. Characterization of nordidemnin B. (a) MSⁿ analysis.

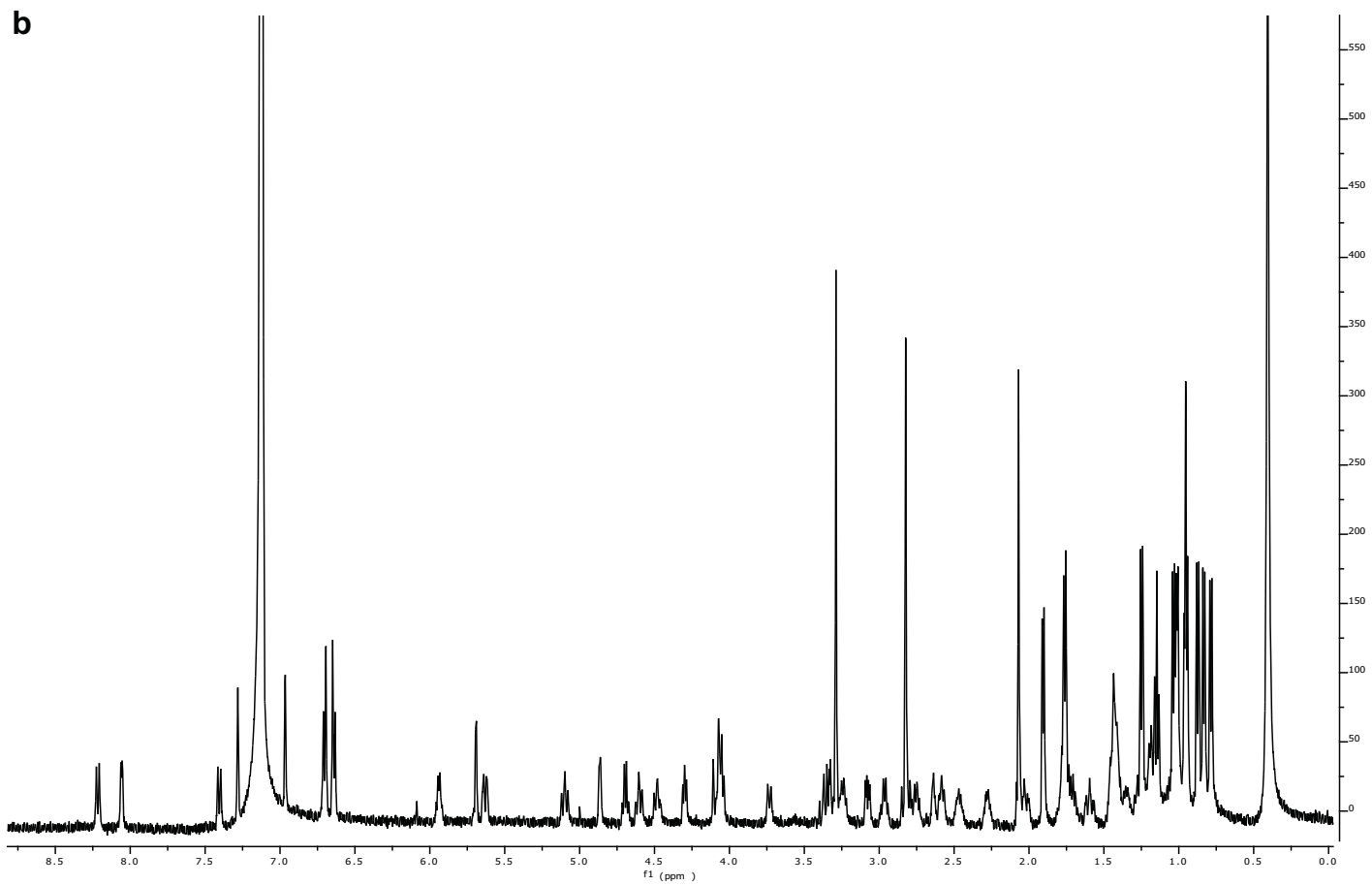


Fig. S2. Characterization of nordidemnin B. (b) ^1H NMR spectrum.

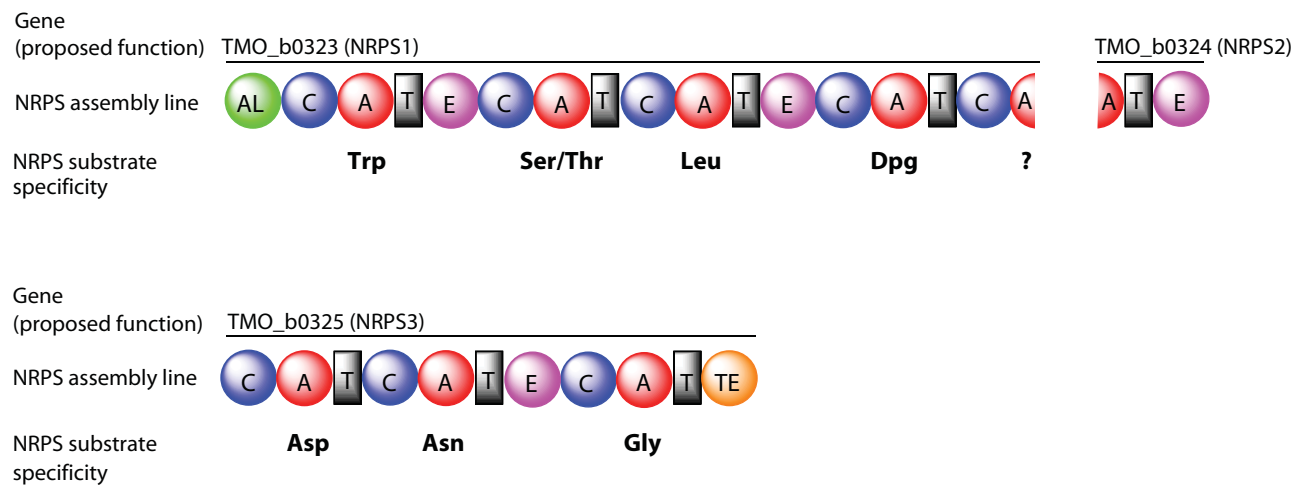


Fig. S3. Bioinformatic analysis of uncharacterized *nrs1* gene cluster on plasmid 2 in the *T. mobilis* KA081020-065 genome. Domain notation: AL, acyl ligase; C, condensation domain; A, adenylation domain; T, thiolation domain; E, epimerization; TE, thioesterase. Substrate abbreviation: Dpg, dihydroxyphenylglycine.

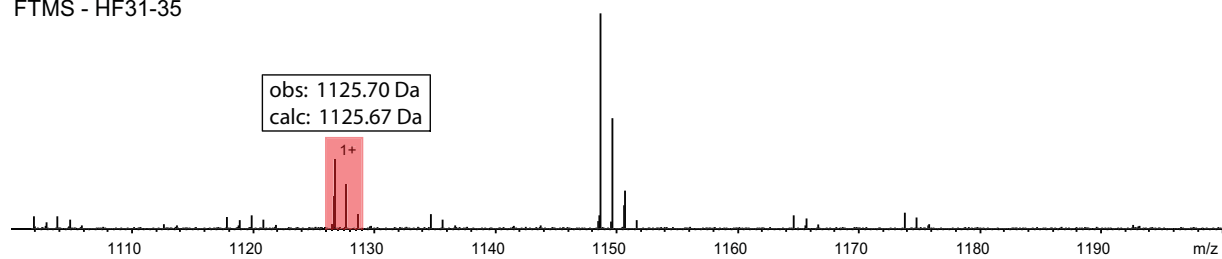


Gene	Size [aa]	Sequence similarity/organism	Proposed function	Identity/similarity	GenBank accession no.
<i>orf1</i>	261	Thioesterase, <i>Actinomadura kijaniata</i>	Thioesterase	41%, 54%	ACB46473
<i>orf2</i>	483	Membrane protease subunit stomatin/prohibitin-like protein, α -proteobacterium BAL199	Putative protease	42%, 59%	ZP_02189733
<i>orf3</i>	560	Cyclic peptide transporter, <i>Methylobacter tundripaludum</i> SV96	Transporter	48%, 68%	ZP_07653047
<i>orf4</i>	70	None	Hypothetical		
<i>orf5</i>	190	GTPase domain-containing protein, <i>Methylobacter tundripaludum</i> SV96	Hypothetical	39%, 57%	ZP_07653048
<i>orf6</i>	987	Hydrophobic/amphiphilic exporter-1, <i>Azospirillum</i> sp. B510	Resistance	38%, 56%	YP_003450508
<i>orf7</i>	377	Secretion protein, <i>Azospirillum</i> sp. B510	Resistance	28%, 47%	YP_003450507
<i>didA</i>	2123	OciA protein, <i>Planktothrix rubescens</i> NIVA-CYA 98	NRPS	29%, 40%	CAQ48254
<i>didB</i>	1796	Linear gramicidin synthetase subunit D, <i>Stigmatella aurantiaca</i> DW4/3-1	NRPS	36%, 48%	ZP_01459555
<i>didC</i>	1330	NRPS, <i>Myxococcus xanthus</i> DK 1622	NRPS	41%, 52%	YP_632257
<i>didD</i>	3853	Amino acid adenylation domain protein, <i>Streptomyces violaceusniger</i> Tu 4113	NRPS	39%, 50%	ZP_07603194
<i>didE</i>	1705	Amino acid adenylation domain protein, <i>Acetivibrio cellulolyticus</i> CD2	NRPS/PKS	36%, 53%	ZP_07325073
<i>didF</i>	1613	HctF, <i>Lyngbya majuscula</i>	NRPS	35%, 52%	AAV42398
<i>didG</i>	1413	NRPS/PKS, <i>Amycolatopsis mediterranei</i> U32	PKS	46%, 56%	YP_003765866
<i>didH</i>	1286	NRPS/PKS, <i>Myxococcus xanthus</i> DK 1622	NRPS	39%, 53%	YP_631961
<i>didI</i>	873	NRPS, <i>Myxococcus xanthus</i> DK 1622	NRPS	39%, 52%	YP_632257
<i>didJ</i>	2163	Amino acid adenylation domain protein, <i>Lyngbya majuscula</i> 3L	NRPS	36%, 53%	ZP_08431746

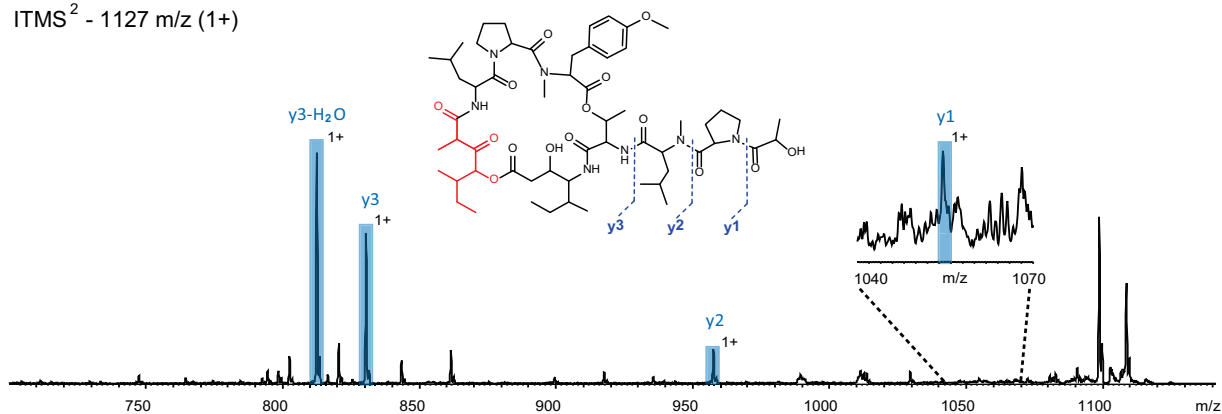
Fig. S4. Organization and deduced functions of the open reading frames within and flanking the didemnin biosynthetic gene cluster.

Gene	Size [aa]	Sequence similarity/organism	Proposed function	Identity/similarity	GenBank accession no.
<i>orf8</i>	77	MbtH domain-containing protein, <i>Herpetosiphon aurantiacus</i> ATCC 23779	MbtH-like protein	80%, 89%	YP_001542806
<i>orf9</i>	68	Hypothetical protein, <i>Acidovorax</i> sp. JS42	Hypothetical	79%, 91%	YP_986866
<i>orf10</i>	45	None	Hypothetical		
<i>orf11</i>	60	None	Hypothetical		
<i>orf12</i>	190	Hypothetical protein, <i>Acidovorax</i> sp. JS42	Hypothetical	94%, 97%	YP_986861
<i>orf13</i>	75	Hypothetical protein, <i>Acidovorax</i> sp. JS42	Hypothetical	98%, 100%	YP_004387524
<i>orf14</i>	324	CAAX amino terminal protease family, <i>Synechococcus</i> sp. PCC 7335	Putative protease	30%, 49%	ZP_05035401
<i>orf15</i>	398	Cyanate transport system protein, <i>Pseudomonas syringae</i> pv. <i>syringae</i> 642	Transport	40%, 52%	ZP_07265073
<i>orf16</i>	255	GntR family transcriptional regulator, <i>Chromobacterium violaceum</i> ATCC 12472	Regulation	39%, 55%	NP_903400

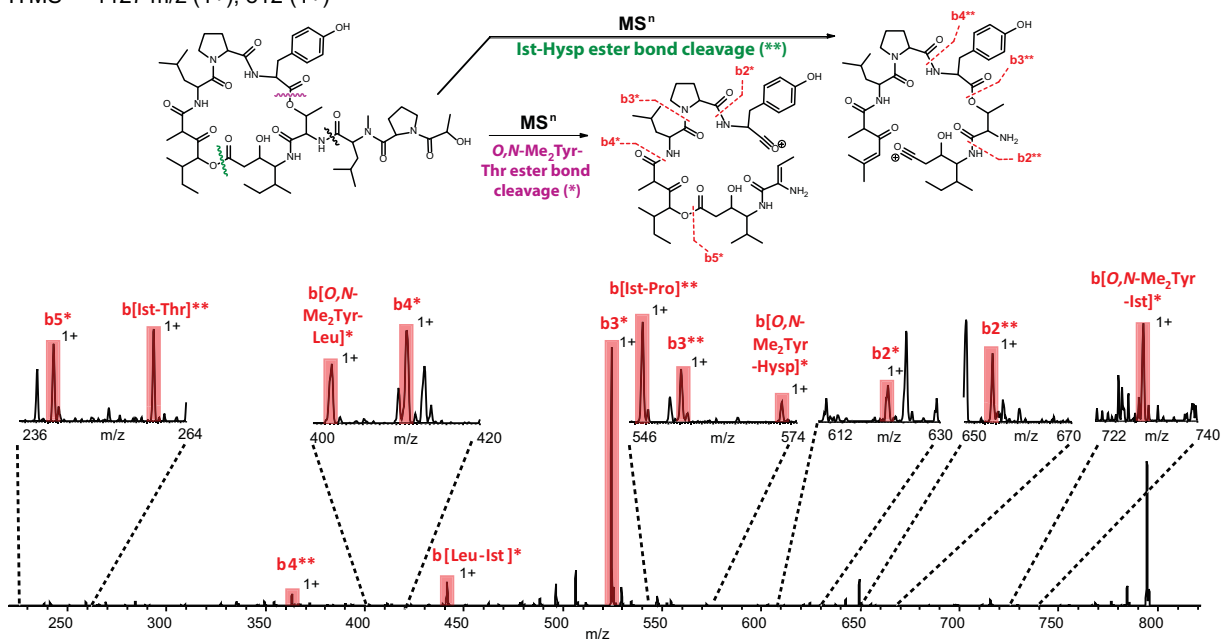
Fig. S4, continued. Organization and deduced functions of the open reading frames within and flanking the didemnins biosynthetic gene cluster.



ITMS² - 1127 m/z (1+)



ITMS³ - 1127 m/z (1+), 812 (1+)



Detector	Precursor ion [m/z]	Observed mass [Da]	Calculated mass [Da]	Error [Da]	Species
IT	1127 (1+)	1054.64	1054.64346	0.00	y1
IT	1127 (1+)	957.55	957.59070	-0.04	y2
IT	1127 (1+)	830.45	830.49098	-0.04	y3
IT	1127 (1+)	812.45	812.48042	-0.03	y3-H ₂ O
IT	1127 (1+), 812 (1+)	729.45	729.44331	0.01	b[O,N-Me ₂ Tyr-Ist]*
IT	1127 (1+), 812 (1+)	655.36	655.37069	-0.01	b2**
IT	1127 (1+), 812 (1+)	621.45	621.38634	0.06	b2*
IT	1127 (1+), 812 (1+)	572.36	572.33303	0.03	b[O,N-Me ₂ Tyr-Hysp]*

Detector	Precursor ion [m/z]	Observed mass [Da]	Calculated mass [Da]	Error [Da]	Species
IT	1127 (1+), 812 (1+)	554.27	554.32301	-0.05	b3**
IT	1127 (1+), 812 (1+)	547.36	547.31263	0.05	b[Ist-Pro]**
IT	1127 (1+), 812 (1+)	524.36	524.33358	0.03	b3*
IT	1127 (1+), 812 (1+)	441.36	441.29591	0.06	b[Leu-Ist]*
IT	1127 (1+), 812 (1+)	411.27	411.24951	0.02	b4*
IT	1127 (1+), 812 (1+)	402.24	402.23873	0.00	b[O,N-Me ₂ Tyr-Leu]*
IT	1127 (1+), 812 (1+)	363.18	363.22838	-0.05	b4**
IT	1127 (1+), 812 (1+)	259.18	259.16523	0.01	b[Ist-Thr]**
IT	1127 (1+), 812 (1+)	241.18	241.15522	0.02	b5*

Fig. S5. Characterization of didemnin B derivative [Hysp]didemnin B produced by *Tistrella mobilis* KA081020-065 due to putative NRPS substrate promiscuity in DidF A8 domain. MSⁿ analysis of [Hysp]didemnin B. Hysp, α-(α-hydroxy sec-butylacetyl) propionic acid.

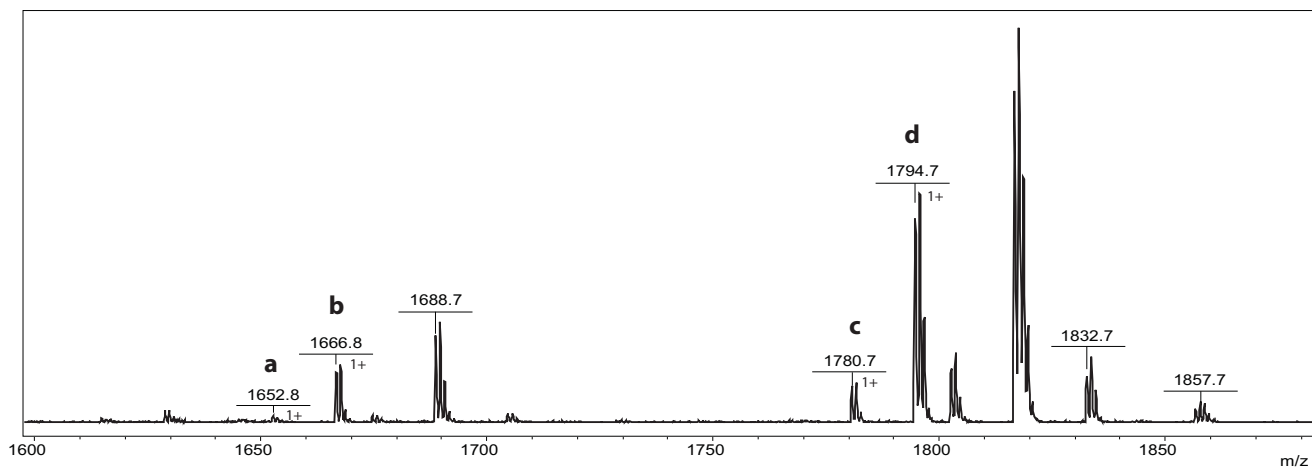


Fig. S6. Detection of didemnin precursors in the extract of *Tistrella mobilis* KA081020-065 by MALDI-TOF MS. a – Nordidemnin X ($[M+H]^+$, obs 1651.8 Da, calc 1651.9 Da), b – Didemnin X ($[M+H]^+$, obs 1665.8 Da, calc 1666.0), c – Nordidemnin Y ($[M+H]^+$, obs 1779.7 Da, calc 1780.0 Da), d – Didemnin Y ($[M+H]^+$, obs 1793.7 Da, calc 1794.0 Da).

a FTMS - HF28

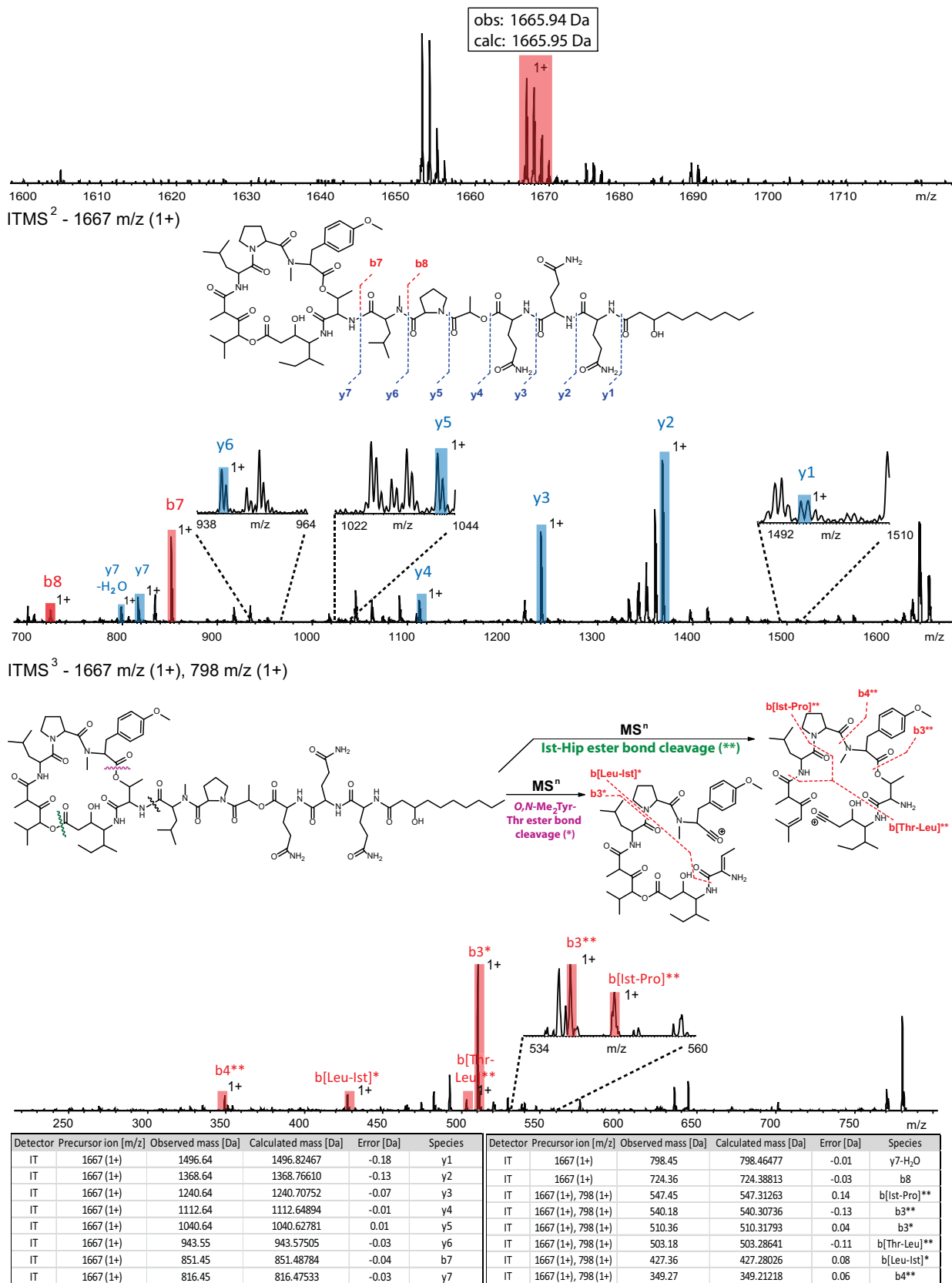


Fig. S7. Characterization of didemnin precursors by MSⁿ. (a) MSⁿ analysis of didemnin X.

b FTMS - HF28

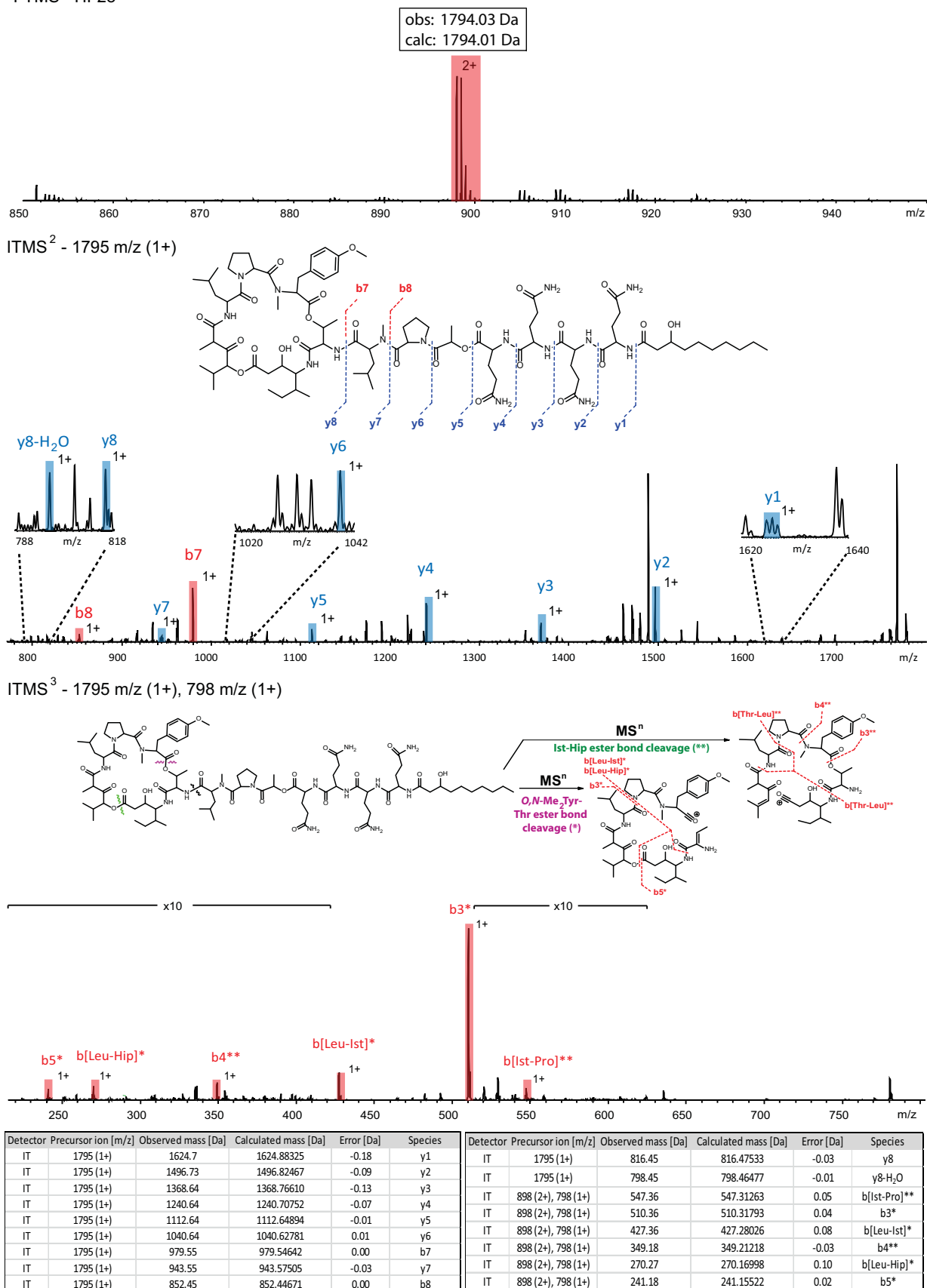


Fig. S7. Characterization of didemnin precursors by MSⁿ. (b) MSⁿ analysis of didemnin Y.

C FTMS - HF28

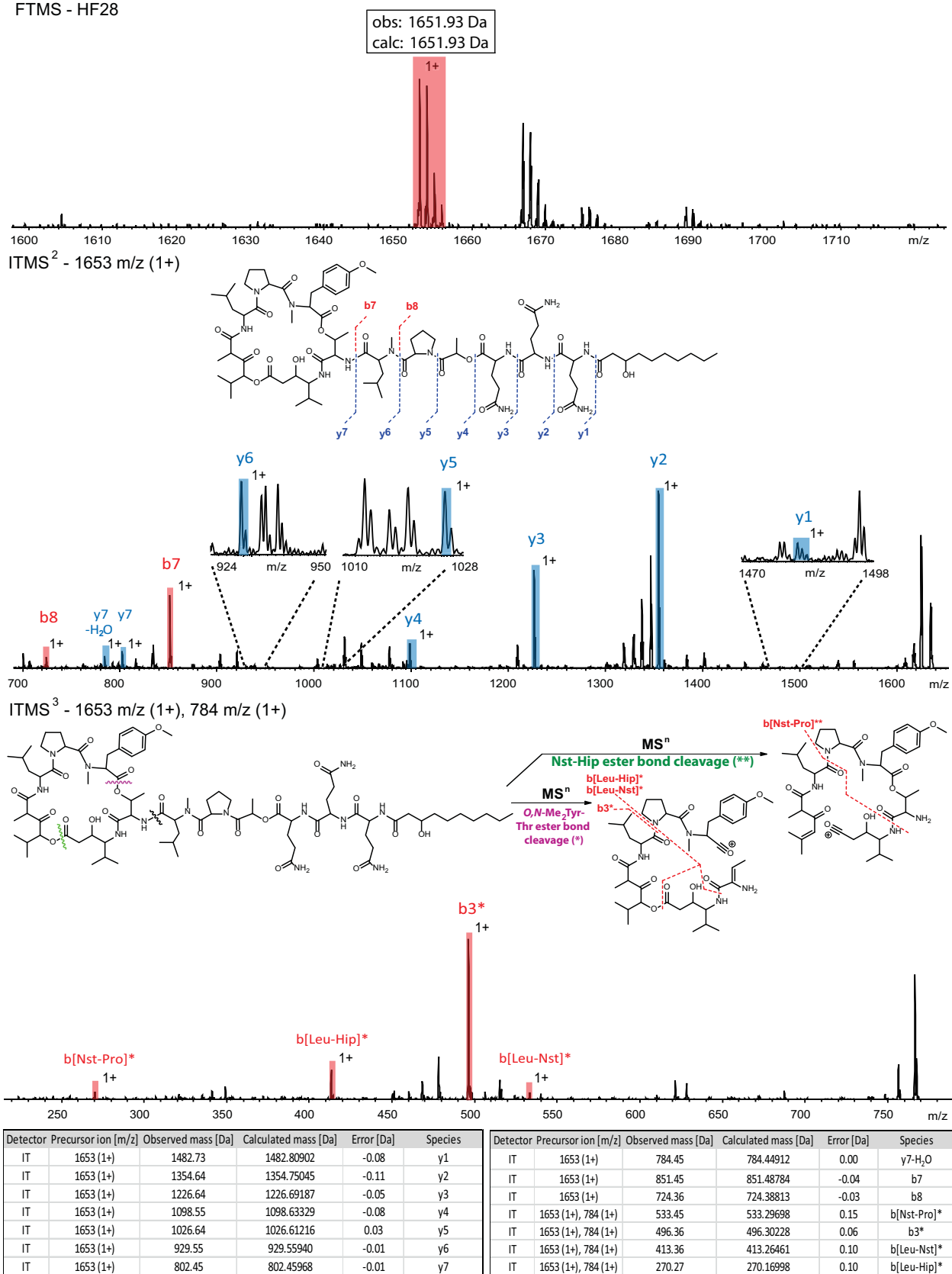
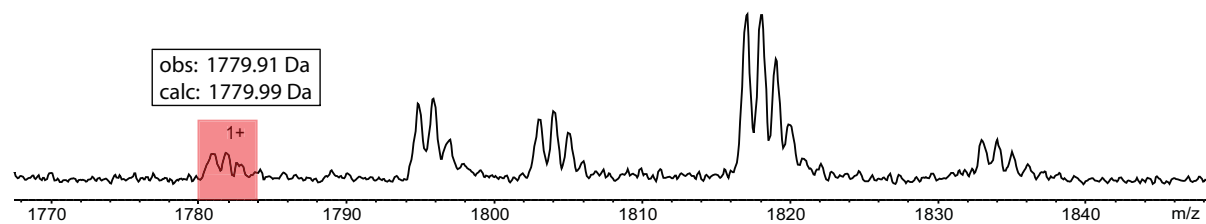
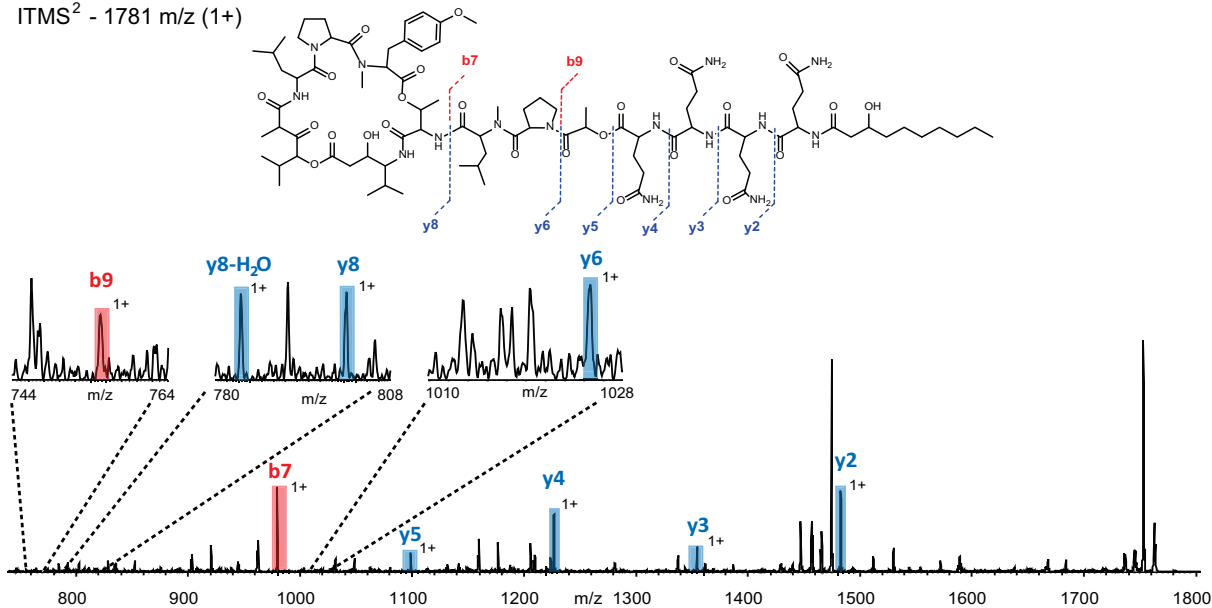


Fig. S7. Characterization of didemnin precursors by MSⁿ. (c) MSⁿ analysis of nordidemnin X.

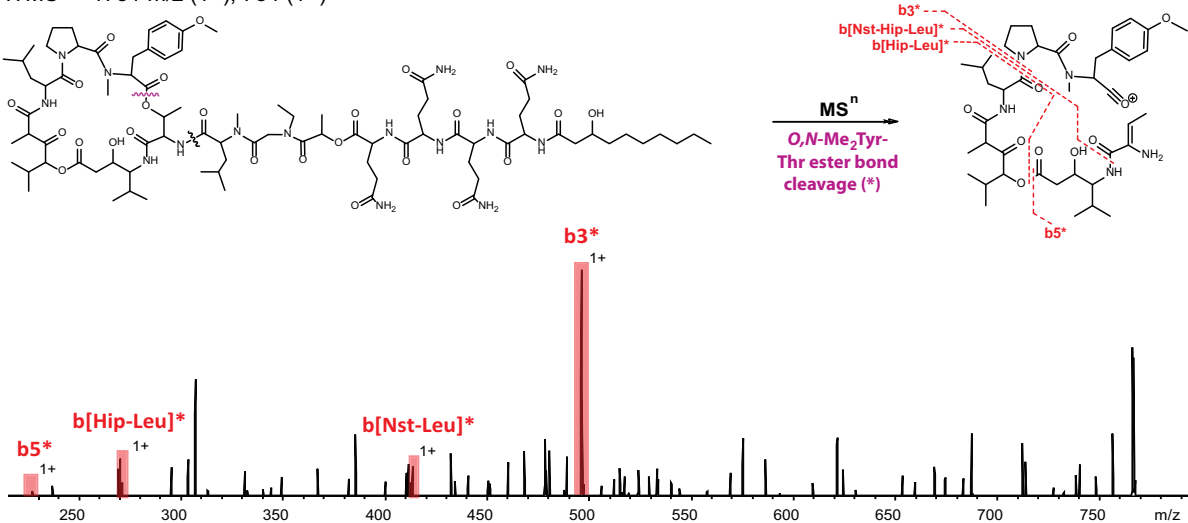
d ITMS - HF26-30



ITMS² - 1781 m/z (1+)



ITMS³ - 1781 m/z (1+), 784 (1+)



Detector	Precursor ion [m/z]	Observed mass [Da]	Calculated mass [Da]	Error [Da]	Species
IT	1781 (1+)	1482.73	1482.80902	0.08	y2
IT	1781 (1+)	1354.64	1354.75045	0.11	y3
IT	1781 (1+)	1226.64	1226.69187	0.05	y4
IT	1781 (1+)	1098.64	1098.63329	-0.01	y5
IT	1781 (1+)	1026.64	1026.61216	-0.03	y6
IT	1781 (1+)	979.45	979.54642	0.10	b7
IT	1781 (1+)	802.45	802.45968	0.01	y8

Detector	Precursor ion [m/z]	Observed mass [Da]	Calculated mass [Da]	Error [Da]	Species
IT	1781 (1+)	784.45	784.44912	0.00	y8-H2O
IT	1781 (1+)	755.27	755.39394	0.12	b9
IT	1781 (1+), 784 (1+)	496.36	496.30228	-0.06	b3*
IT	1781 (1+), 784 (1+)	413.36	413.26461	-0.10	b[Nst-Leu]*
IT	1781 (1+), 784 (1+)	270.18	270.16998	-0.01	b[Hip-Leu]*
IT	1781 (1+), 784 (1+)	227.13	227.13957	0.01	b5*

Fig. S7. Characterization of didemnin precursors by MSⁿ. (d) MSⁿ analysis of nordidemnin Y.

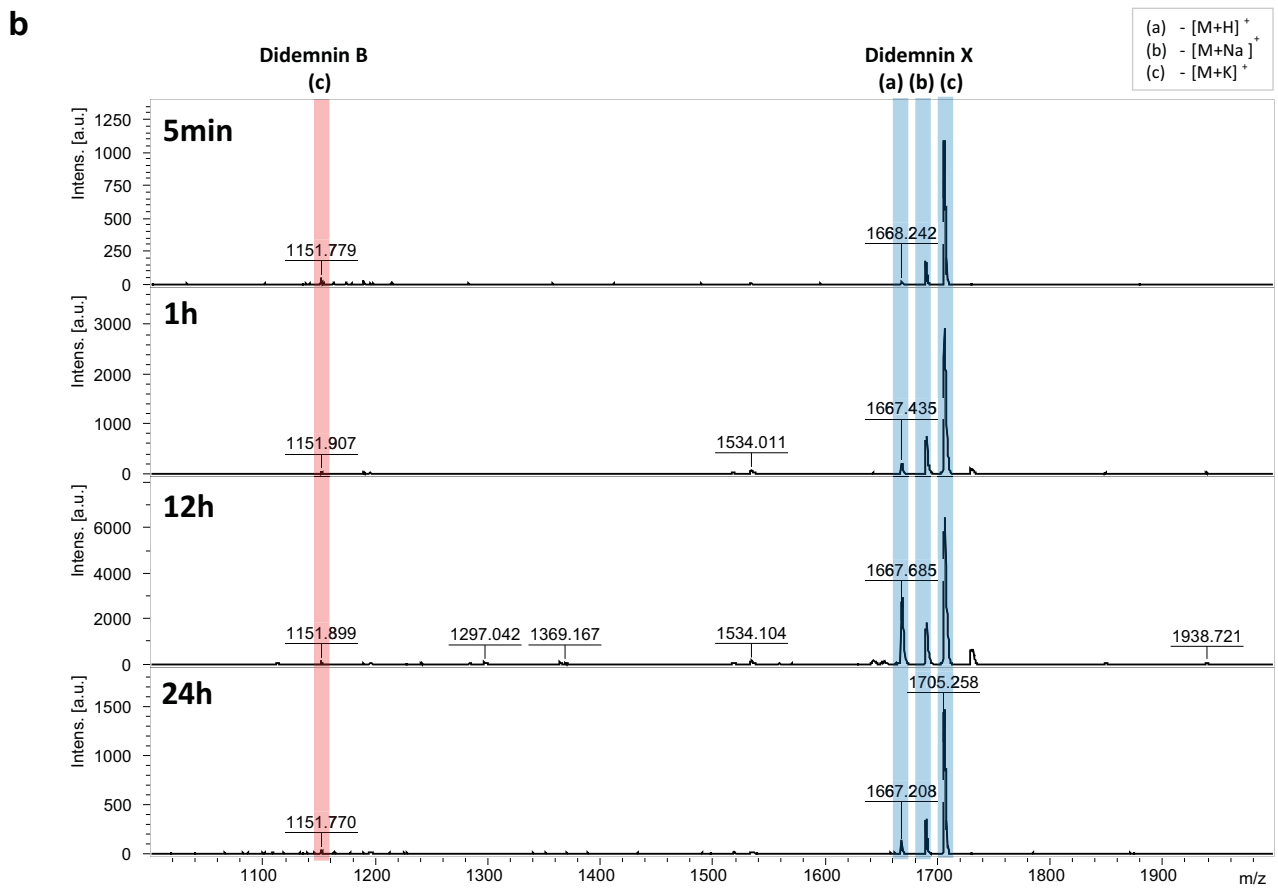
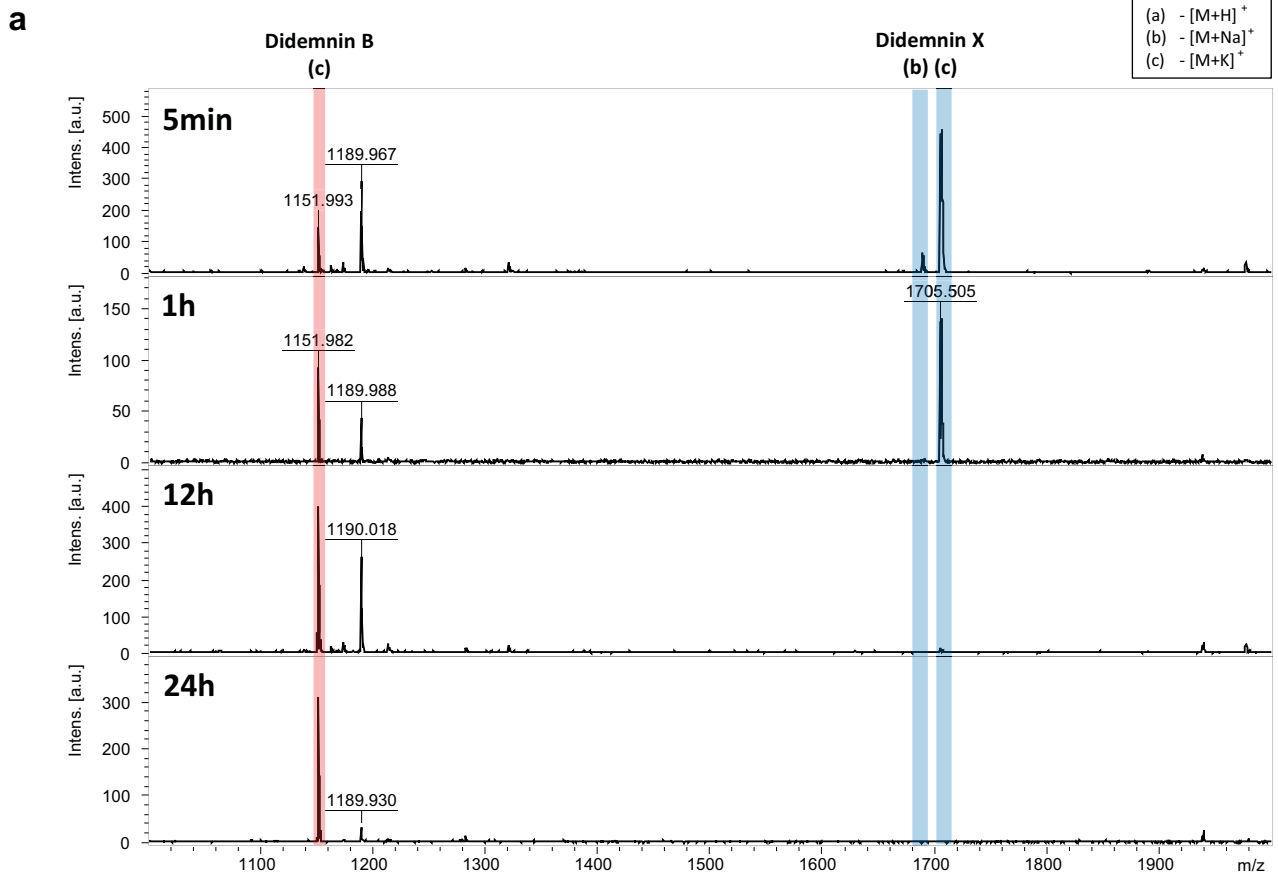


Fig. S8. Didemnin precursor hydrolysis assay. MALDI-TOF MS analysis of didemnin X hydrolysis time course in presence of *Tistrella mobilis* secreted proteome (**(a)**, 1d growth in liquid culture, >10kDa protein cutoff filter) and in presence of sterile GYP medium (**(b)**, negative control).

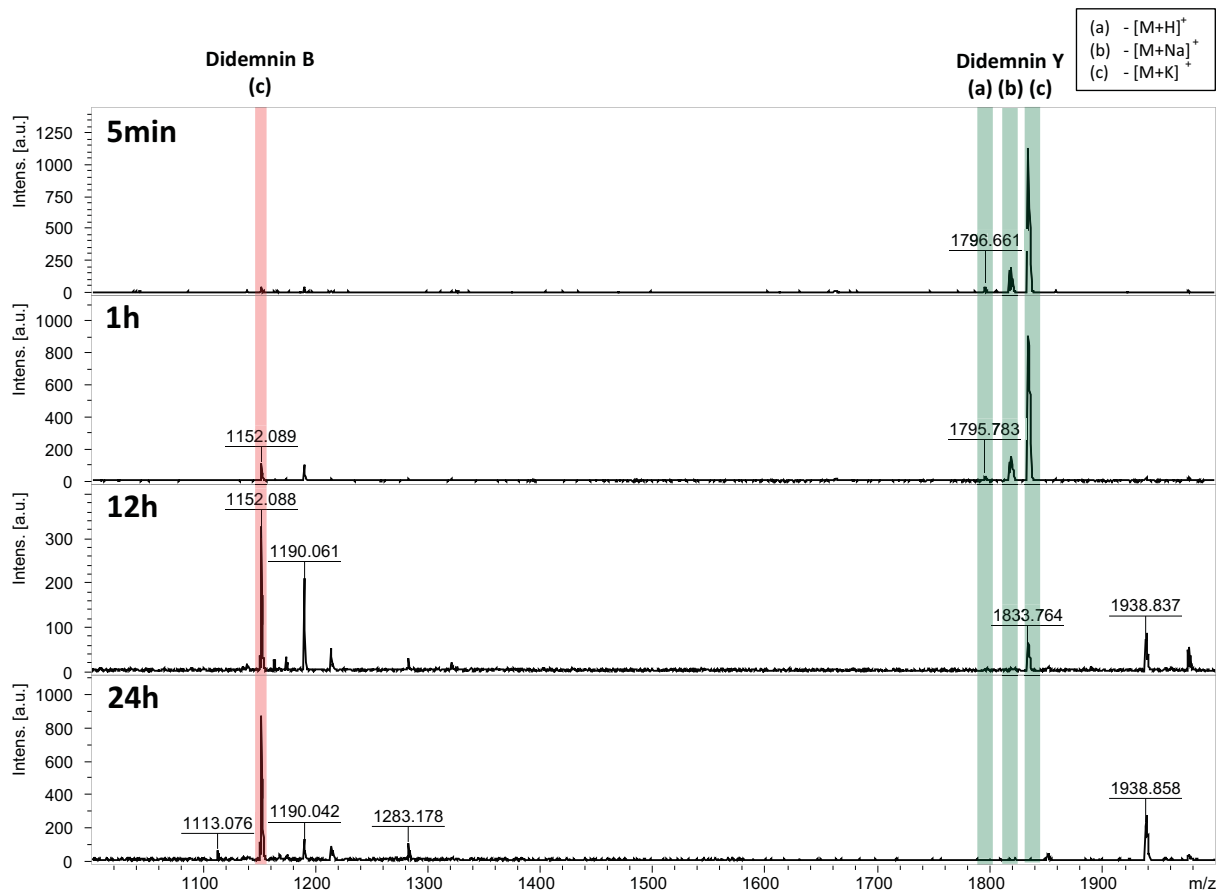
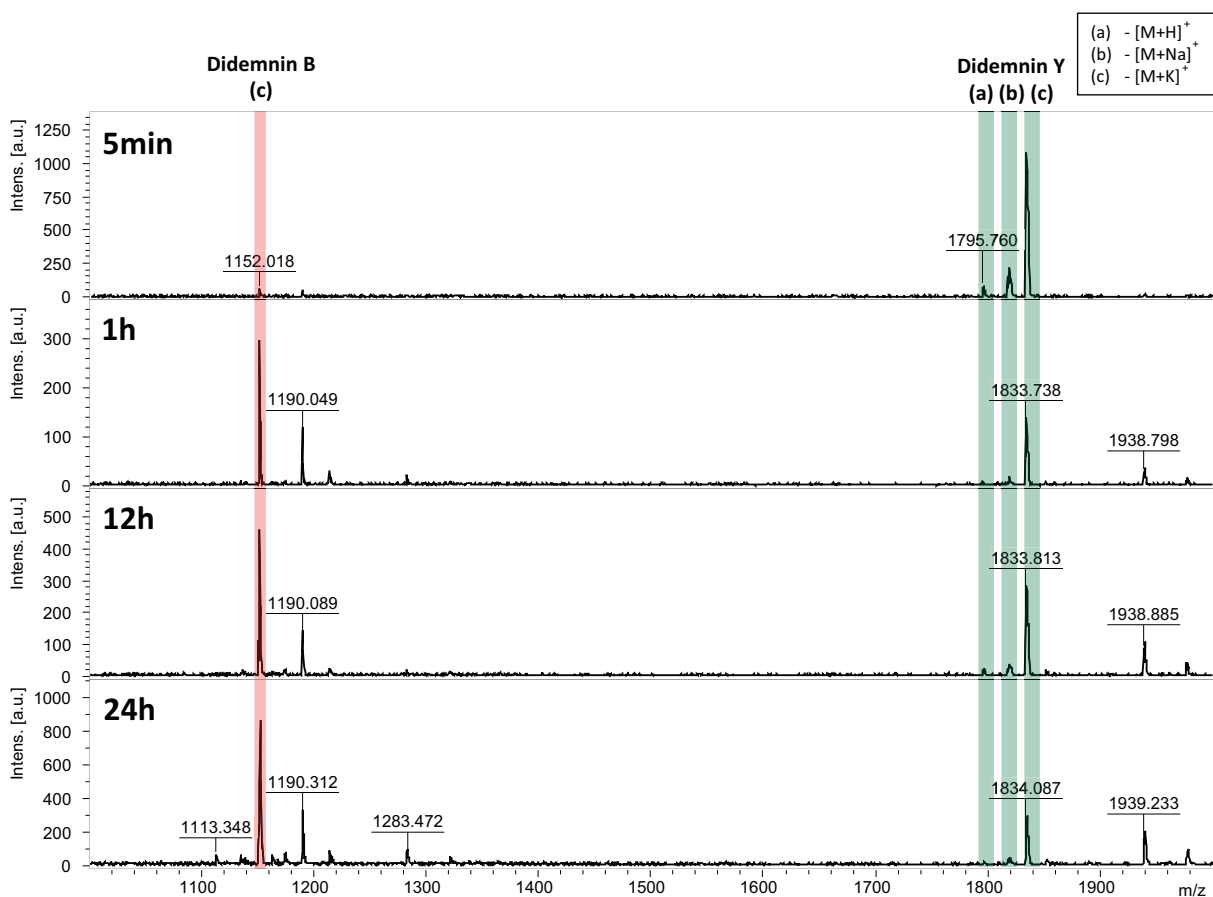
c**d**

Fig. S8. Didemnin precursor hydrolysis assay. MALDI-TOF MS analysis of didemnin Y hydrolysis time course in presence of *Tistrella mobilis* secreted proteome ((c), 1d growth in liquid culture, >10kDa protein cutoff filter) and in presence of sterile GYP medium ((d), negative control).

Table S1. General genome features of *Tistrella mobilis* KA081020-065

Features	Chromosome	pTM1	pTM2	pTM3	pTM4
Topology	Circular	Circular	Circular	Circular	Circular
Genome size (bp)	3919492	692874	690188	1126962	83885
G+C content	68.15%	68.26%	67.61%	68.08%	67.26%
CDs number	3565	605	602	942	72
Coding density	89.08%	90.38%	88.92%	88.55%	85.06%
Average CDs length (bp)	979	1035	1019	1059	991
Assigned function	2852	432	501	726	60
Conserved hypothetical	472	101	71	165	9
Hypothetical	244	72	30	51	3
rRNA operons	2	0	0	1	1
tRNA operons	41	0	1	9	3