Integrated genome and metabolome mining unveiled structure and biosynthesis of novel lipopeptides from a deep-sea *Rhodococcus*

Table of contents

Table S1. Putative genes identified in the *rhp* gene cluster.

protein	amino acids	proposed function	closest homolog (protein, origin)	identities/ positives (%)	accession #
ORF1	391	regulatory enzyme	histidine kinase [*Rhodococcus*]	100/100	WP_230558200.1
ORF2	219	regulatory enzyme	response regulator transcription factor [*Rhodococcus yunnanensis*]	99/100	WP_269419590.1
ORF3	165		FxsA family protein [*Rhodococcus*]	100/100	WP_230558198.1
ORF4	537		amidohydrolase family protein [*Rhodococcus yunnanensis*]	99/99	WP_269419591.1
ORF5	531	*N*-acyltransferase	apolipoprotein *N*-acyltransferase [*Rhodococcus* sp. KRD197]	98/99	WP_206511629.1
rhp	8919	NRPS	non-ribosomal peptide synthase [uncultured Rhodococcus sp.]	98/98	WP_300399371.1
ORF7	199		hypothetical protein [*Rhodococcus*]	100/100	WP_230558194.1
ORF8	686		aminodeoxychorismate synthase [*Rhodococcus* sp. KRD197]	99/100	WP_206511506.1
ORF9	104		chorismate mutase [*Rhodococcus*]	100/100	WP_075839937.1
ORF10	112		RNA polymerase-binding protein RbpA [*Rhodococcus*]	100/100	WP_075839938.1
ORF11	66		hypothetical protein BJF84_24140 [*Rhodococcus* sp. CUA-806]	98/100	OLT33093.1
ORF12	196		sugar O-acetyltransferase [*uncultured Rhodococcus* sp.]	100/100	WP_230558192.1
ORF13	342		linear amide C-N hydrolase [*Rhodococcus*]	100/100	WP_230558191.1
ORF14	337		GGDEF domain-containing protein [*Rhodococcus*]	99/100	WP_230558190.1
ORF15	381	regulatory enzyme	histidine kinase [*Rhodococcus*]	100/100	WP_230558189.1

Table S2. Putative genes identified in the *rmd* gene cluster.

protein	amino acids	proposed function	closest homolog (protein, origin)	identities/ positives (%)	accession #
rmdE	540	acyl-CoA ligase	AMP-binding protein [*Rhodococcus yunnanensis*]	99/99	WP_269419124.1
rmdF	662	acyl-CoA ligase	acetoacetate-CoA ligase, [*Rhodococcus yunnanensis*]	91/94	WP_269419125.1
ORF3	144		thioesterase family protein [*Rhodococcus*]	100/100	WP_269419126.1
ORF4	482		amidohydrolase family protein [*Rhodococcus*]	100/100	WP_230555287.1
ORF5	209	regulatory enzyme	TetR/AcrR family transcriptional regulator [*Rhodococcus*]	100/100	WP_230555286.1
ORF6	141		PPOX class F420-dependent oxidoreductase [*Rhodococcus*]	100/100	WP_206511370.1
rmdA	9858	NRPS	non-ribosomal peptide synthetase, [uncultured *Rhodococcus* sp.]	100/100	WP_300396890.1
rmdB	3211	NRPS	non-ribosomal peptide synthetase, [uncultured *Rhodococcus* sp.]	100/100	WP_294638938.1
rmdC	821	NRPS	non-ribosomal peptide synthetase, [uncultured *Rhodococcus* sp.]	99/100	WP_300500358.1
rmdD	1961	NRPS	non-ribosomal peptide synthetase, [uncultured *Rhodococcus* sp.]	99/99	WP_294639008.1
ORF11	354		esterase [*Rhodococcus* sp. CUA-806]	99/100	OLT31464.1
ORF12	2888		non-ribosomal peptide synthetase, [*Rhodococcus* sp. KRD197]	99/99	WP_307786731.1

ORF13	1202		non-ribosomal peptide synthetase, [uncultured *Rhodococcus* sp.]	99/100	WP_300499683.1
ORF14	322		sulfotransferase [*Rhodococcus* sp. KRD197]	99/100	WP_075838817.1
ORF15	323		alpha/beta fold hydrolase [*Rhodococcus* sp. KRD197]	99/99	WP_307786732.1
ORF16	282	transporter	ABC transporter permease [*Rhodococcus*]	99/100	WP_230557628.1
ORF17	281	transporter	ABC transporter permease [*Rhodococcus*]	100/100	WP_206509620.1
ORF18	323	transporter	ATP-binding cassette domain-containing protein [*Rhodococcus*]	100/100	WP_206509621.1

Table S3. MZMINE2 parameters for MS raw data processing.

Feature detection	
MS level 1	1.0e5
MS level 2	1.0e3
ADAP chromatogram builder	• group intensity threshold: 5.0e4 • minimum highest intensity: 5.0e4 • *m/z* tolerance: 0.01 *m/z* or 10 ppm
Chromatogram deconvolution	
baseline cut-off	• minimum peak height: 9.0e4 • peak duration range: 0.0 – 10.0 min • baseline level: 5.0e4 • m/z range for MS2 scan pairing: 0.01 Da • RT range for MS2 scan pairing: 0.2 min
Alignment	
Join aligner	• *m/z* tolerance: 0.002 *m/z* or 5 ppm • retention time tolerance: 0.1 min
Identification	
Adduct search ([M+Na–H], [M+K–H], [M+Mg−2H], [M+NH_3], [M-Na+NH_4], [M+1, ^{13}C])	• retention time tolerance: 0.1 min • *m/z* tolerance: 0.002 *m/z* or 5 ppm • max relative adduct peak height: 100%

Table S4. Relative abundance of rhodoheptins A-U in the purified rhodoheptin-enriched fraction.

compound	**$[M+H]^+$**	***m/z***	**R_t (min)**	**relative abundance (%)**
rhodoheptin A (**1**)	$C_{52}H_{96}N_7O_{11}$	994.7152	22.3	0.8
rhodoheptin B (**2**)	$C_{53}H_{98}N_7O_{11}$	1008.7305	26.0	0.6
rhodoheptin C (**3**)	$C_{53}H_{98}N_7O_{11}$	1008.7311	27.3	0.5
rhodoheptin D (**4**)	$C_{54}H_{98}N_7O_{11}$	1020.7308	23.2	2.4
rhodoheptin E (**5**)	$C_{54}H_{100}N_7O_{11}$	1022.7470	30.1	13.6
rhodoheptin F (**6**)	$C_{55}H_{94}N_7O_{11}$	1028.7001	21.4	0.9
rhodoheptin G (**7**)	$C_{56}H_{96}N_7O_{11}$	1042.7154	26.6	0.7
rhodoheptin H (**8**)	$C_{56}H_{102}N_7O_{11}$	1048.7637	29.8	1.5
rhodoheptin I (**9**)	$C_{56}H_{104}N_7O_{11}$	1050.7787	19.1	28.7
rhodoheptin J (**10**)	$C_{57}H_{96}N_7O_{11}$	1054.7154	22.3	4.3

rhodoheptin K (**11**)	$C_{53}H_{98}N_7O_{12}S$	1056.6981	21.8	0.4
rhodoheptin L (**12**)	$C_{57}H_{98}N_7O_{11}$	1056.7317	28.8	21.9
rhodoheptin M (**13**)	$C_{57}H_{106}N_7O_{11}$	1064.7940	22.1	1.7
rhodoheptin N (**14**)	$C_{58}H_{100}N_7O_{11}$	1070.7470	30.4	0.9
rhodoheptin O (**15**)	$C_{58}H_{106}N7O_{11}$	1076.7948	14.7	0.6
rhodoheptin P (**16**)	$C_{59}H_{100}N_7O_{11}$	1082.7472	28.6	3.0
rhodoheptin Q (**17**)	$C_{55}H_{102}N_7O_{12}S$	1084.7290	25.8	1.0
rhodoheptin R (**18**)[b]	$C_{59}H_{102}N_7O_{11}$	1084.7633	15.3	4.9
rhodoheptin S (**19**)	$C_{60}H_{104}N_7O_{11}$	1098.7781	18.9	1.7
rhodoheptin T (**20**)	$C_{60}H_{110}N_7O_{11}$	1104.8262	34.8	3.3
rhodoheptin U (**21**)	$C_{63}H_{108}N_7O_{11}$	1138.8105	32.3	7.0

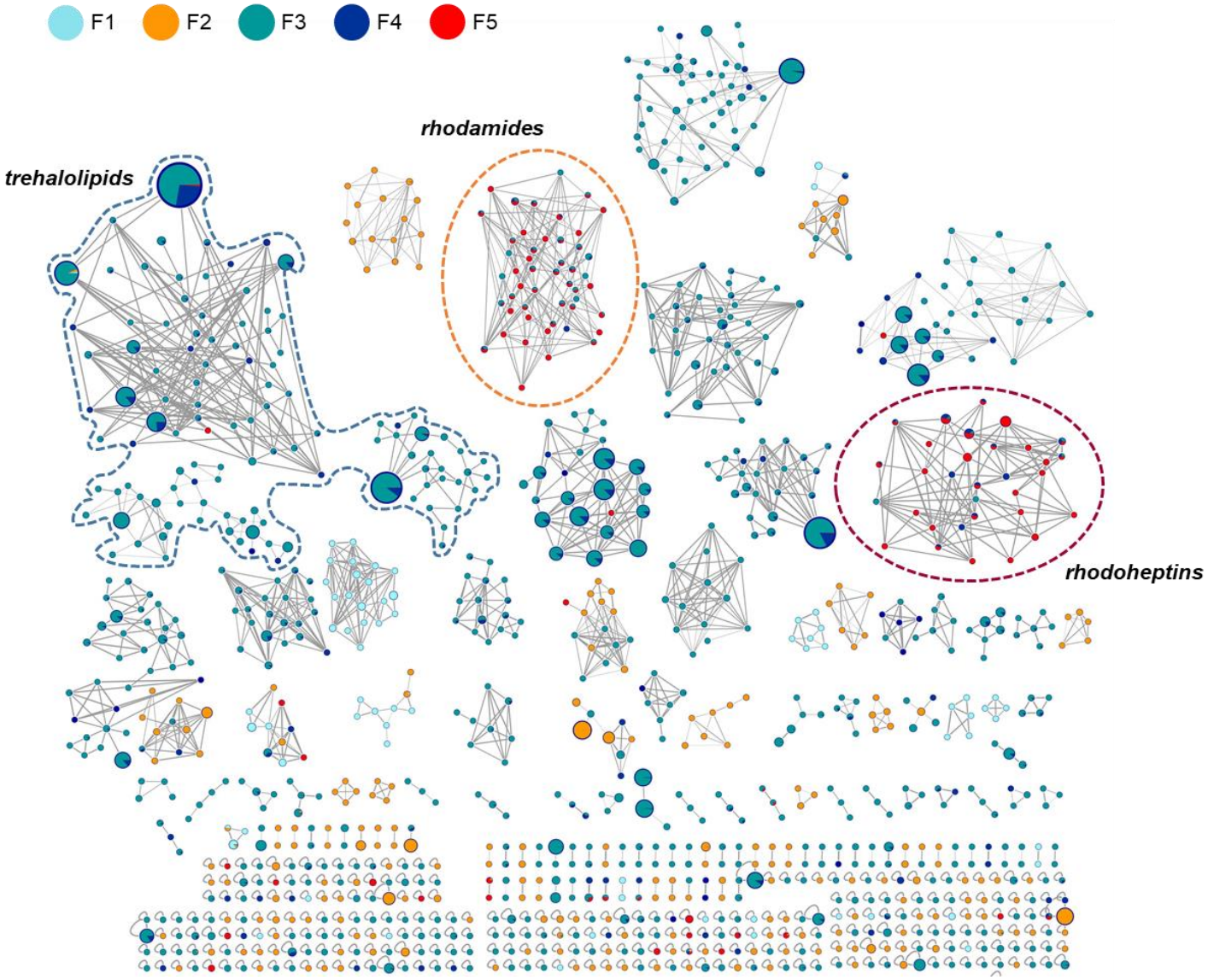

Figure S1. Global molecular network of the SPE fractions (F1-F5) from the organic extract of *Rhodococcus* sp. I2R, highlighting biosurfactants produced by the bacterium, i.e. rhodoheptins (dashed red line), rhodamides (dashed orange line), and trehalolipids (dashed blue line). Nodes are visualized as pie charts showing metabolite distribution in SPE fractions. Edge thickness reflects cosine score similarity and node size is related to metabolite amounts (peak area).

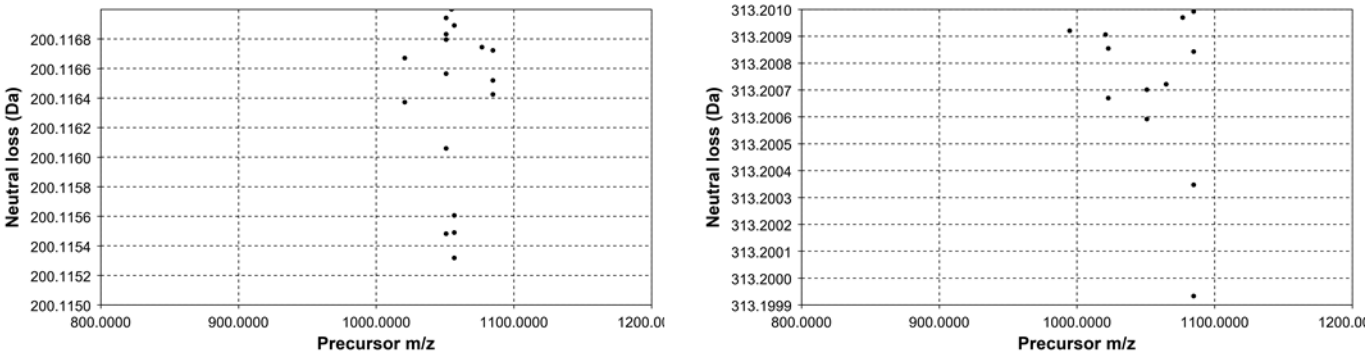

Figure S2. Neutral loss/precursor *m/z* plots from HR-MS/MS data of the F5 fraction from *Rhodococcus* sp. I2R organic extract. These plots highlight rhodoheptin precursor ions displaying neutral losses of 200.12 amu ($C_9H_{16}N_2O_3$, left panel) and 313.20 amu ($C_{15}H_{27}N_3O_4$, right panel) corresponding, respectively, to the Thr-Val and Leu-Ser-Ile/Leu motifs, predicted from genomic analysis of the *rhd* gene cluster.

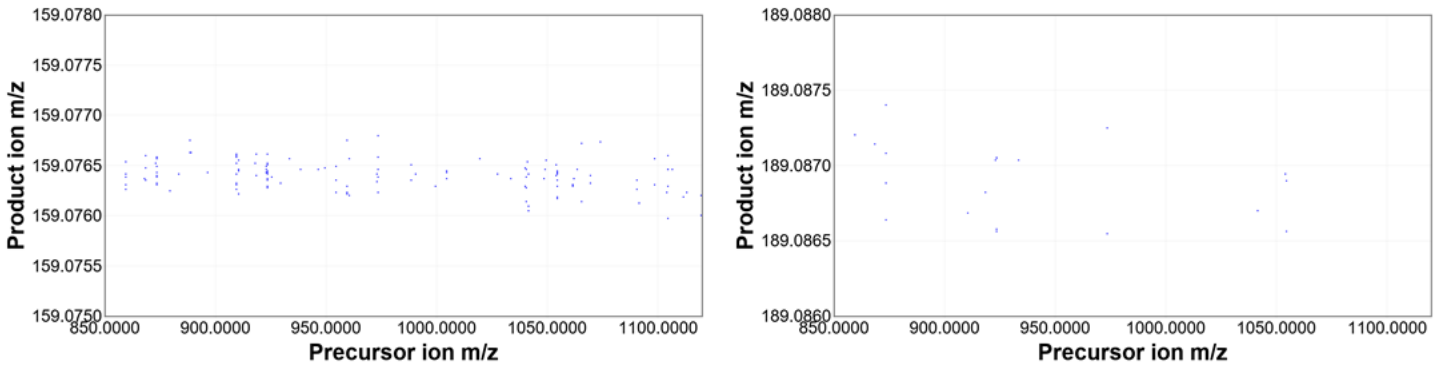

Figure S3. Product ion/precursor *m/z* plots from HR-MS/MS data of the F5 fraction from *Rhodococcus* sp. I2R organic extract. These plots highlight rhodamide precursor ions displaying fragment ions at *m/z* 159.076 ($C_6H_{11}N_2O_3^+$, left panel) and 189.087 ($C_7H_{13}N_2O_4^+$, right panel) corresponding, respectively, to the Ser-Ala and Thr-Ser motifs predicted from genomic analysis of the *rmd* gene cluster.

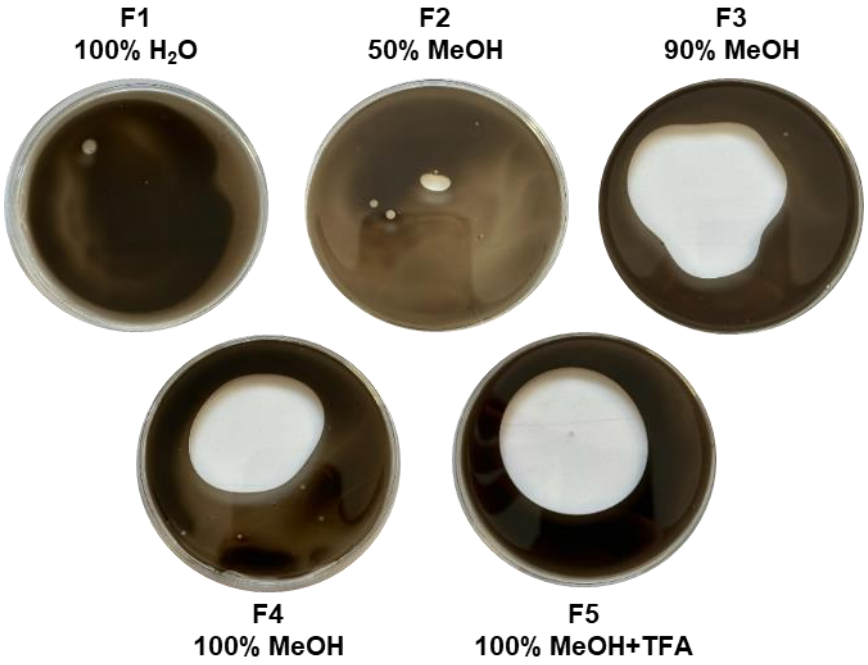

Figure S4. Oil-spreading assay to assess biosurfactant activity of the SPE fractions F1-F5 from the organic extract of *Rhodococcus* sp. I2R. 1 µL of each fraction (50 mg/mL) was placed onto the center of the oil layer. The presence of biosurfactants was indicated by formation of a clear zone in the oil layer.

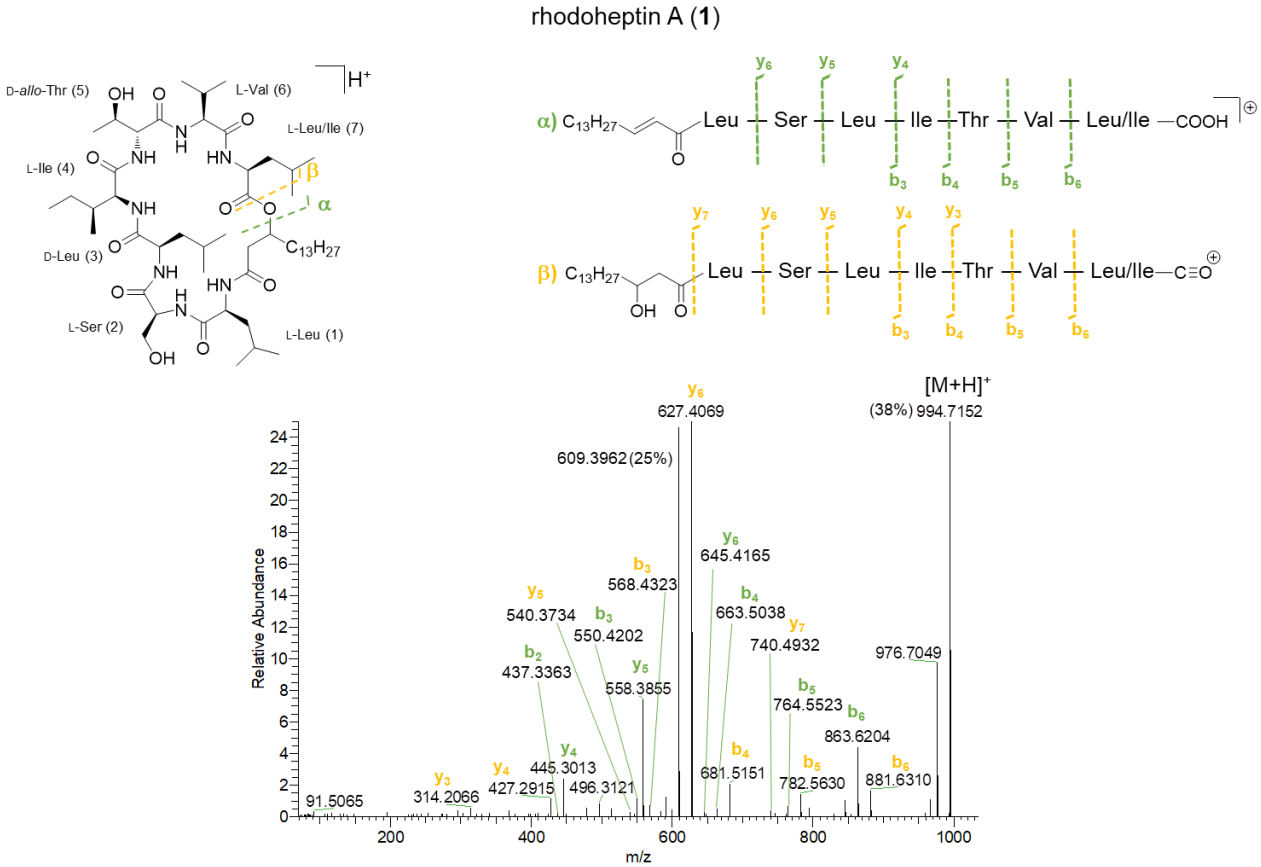

Figure S5. HR-MS^2 spectra of the [M+H]$^+$ pseudomolecular ion of rhodoheptin A (**1**).

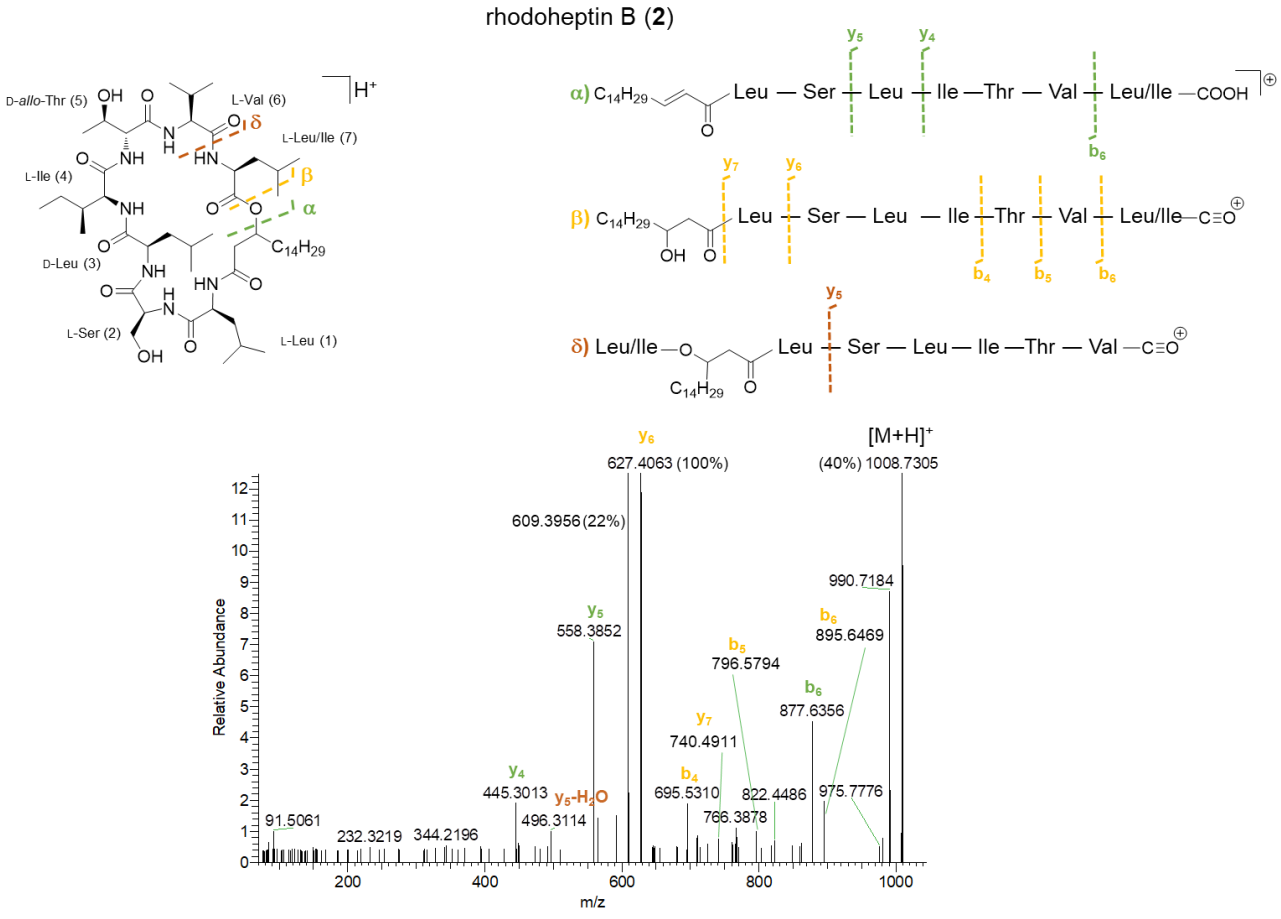

Figure S6. HR-MS^2 spectra of the [M+H]$^+$ pseudomolecular ion of rhodoheptin B (**2**).

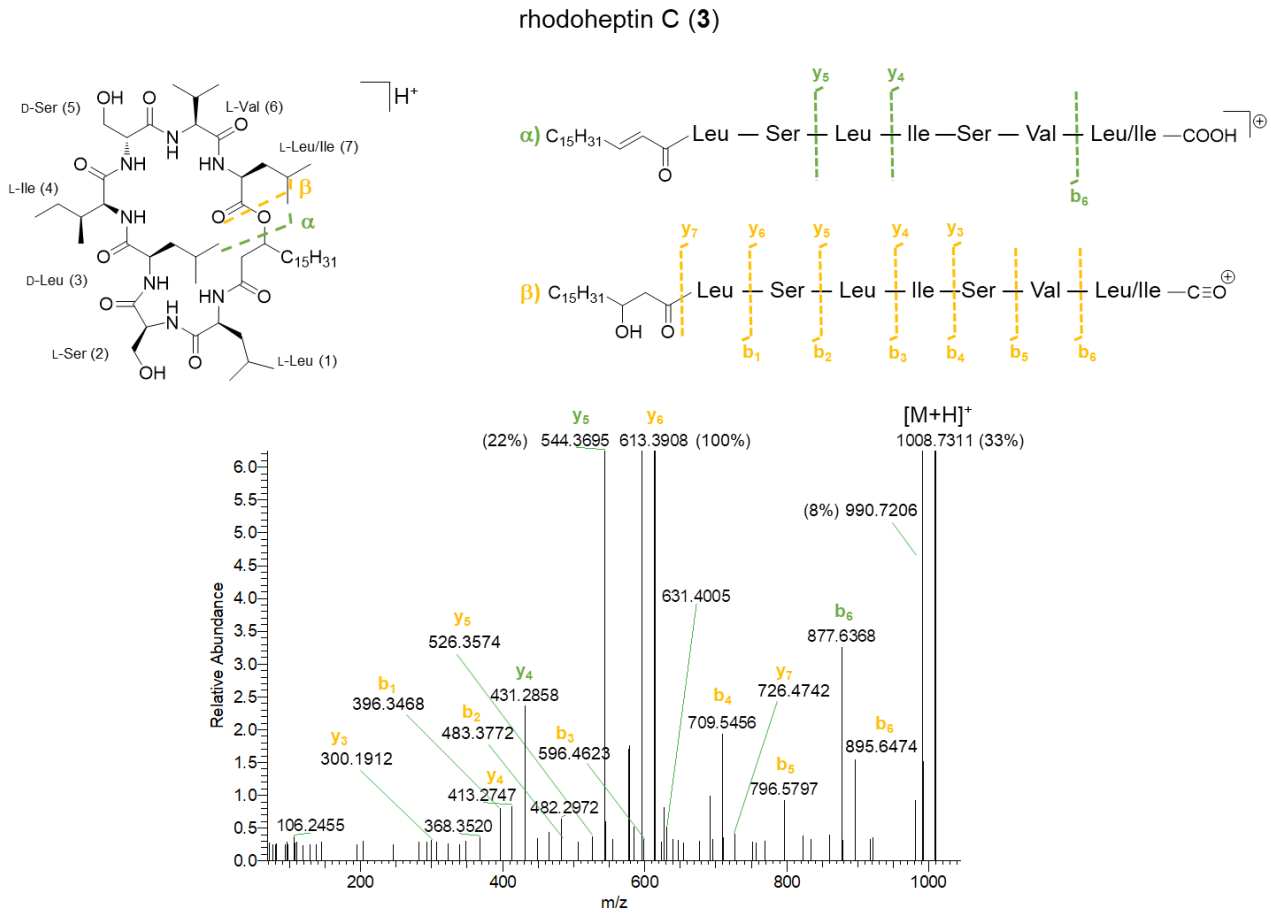

Figure S7. HR-MS^2 spectra of the [M+H]$^+$ pseudomolecular ion of rhodoheptin C (**3**).

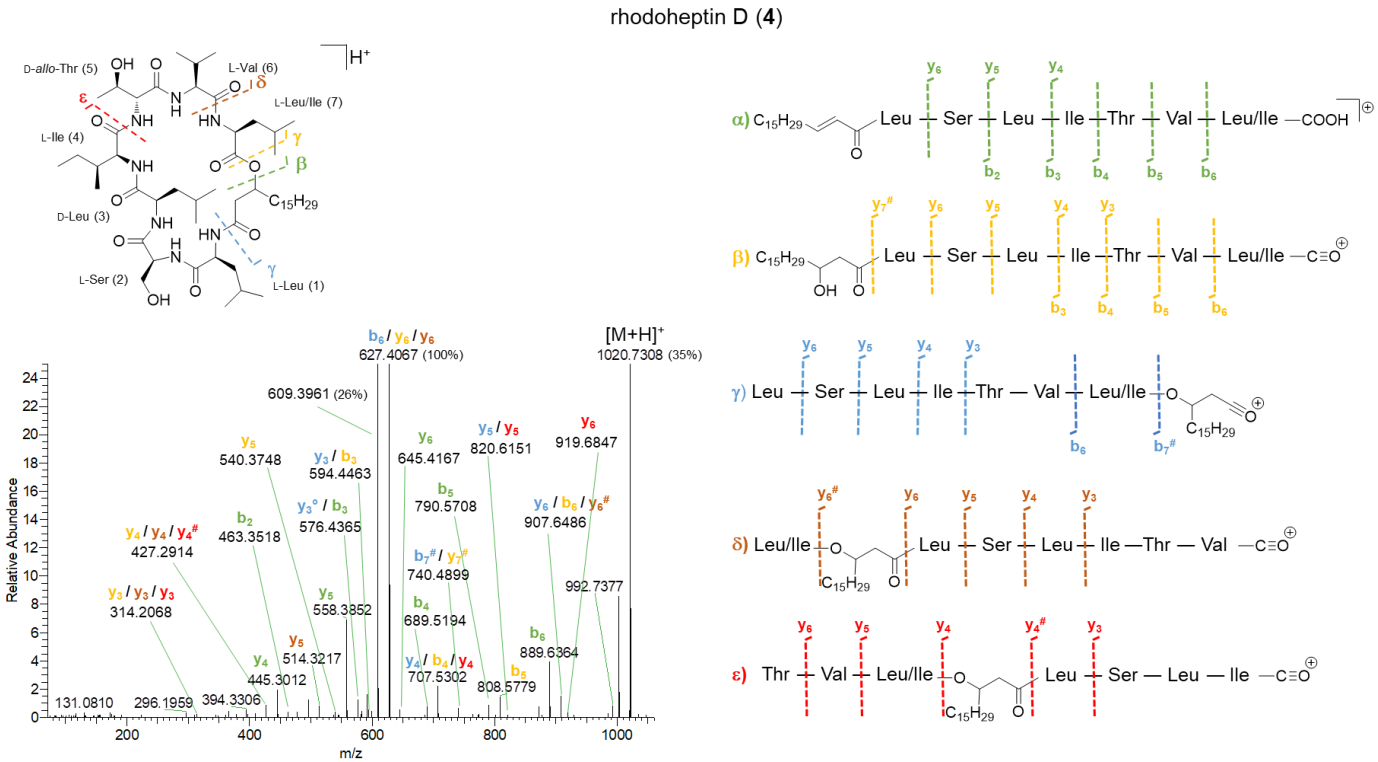

Figure S8. HR-MS^2 spectra of the [M+H]$^+$ pseudomolecular ion of rhodoheptin D (**4**).

rhodoheptin F (**6**)

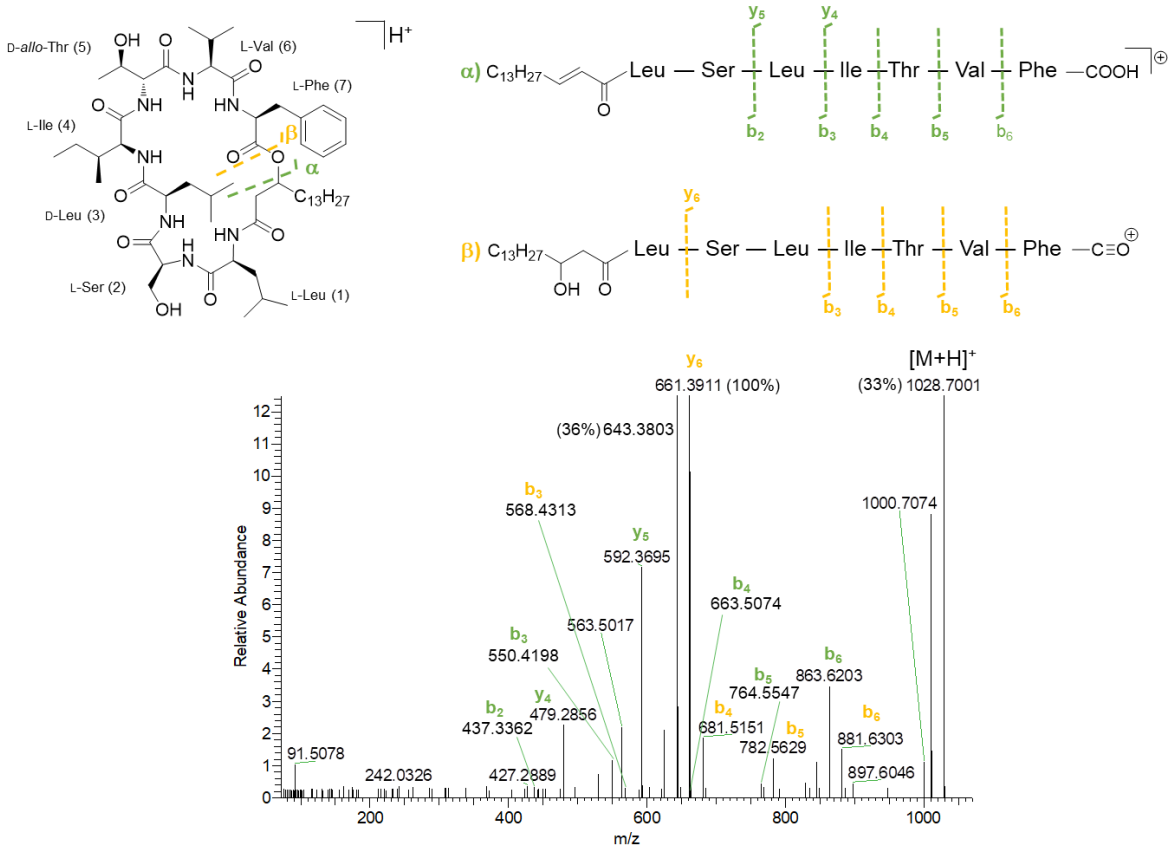

Figure S9. HR-MS^2 spectra of the [M+H]$^+$ pseudomolecular ion of rhodoheptin F (**6**).

rhodoheptin G (**7**)

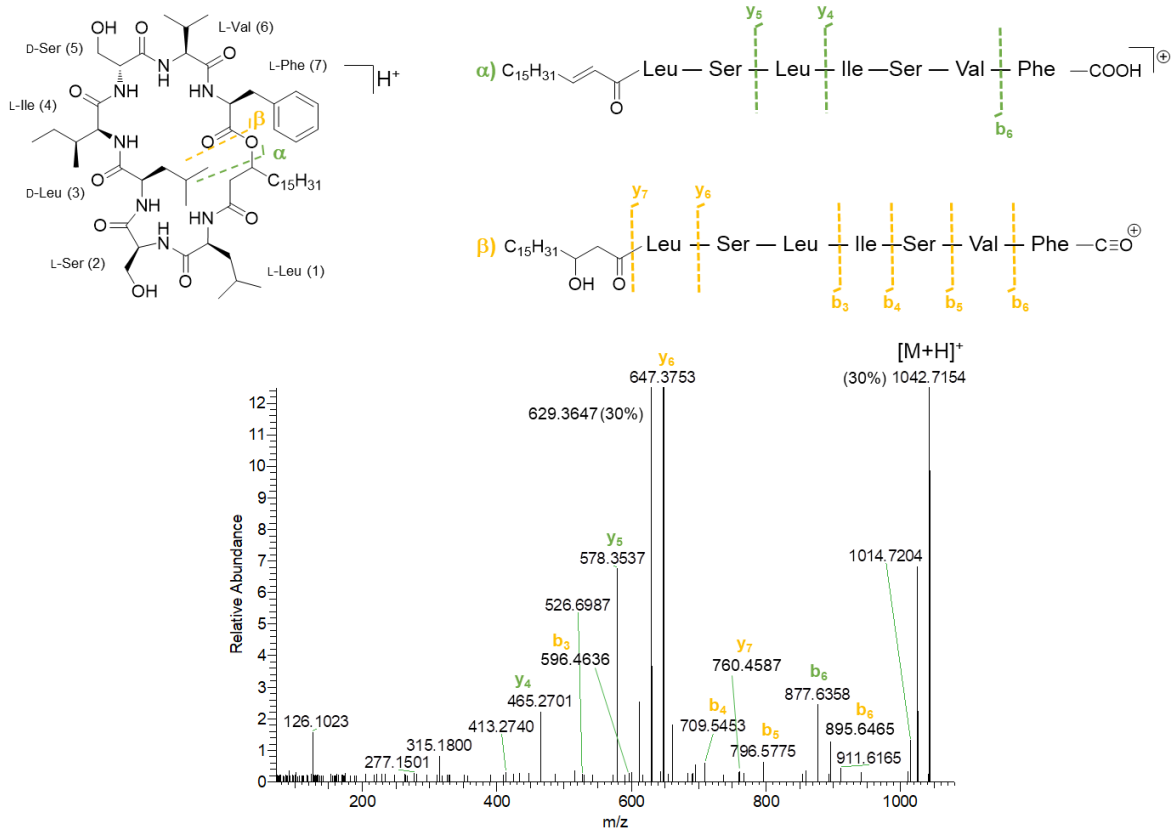

Figure S10. HR-MS^2 spectra of the [M+H]$^+$ pseudomolecular ion of rhodoheptin G (**7**).

rhodoheptin H (**8**)

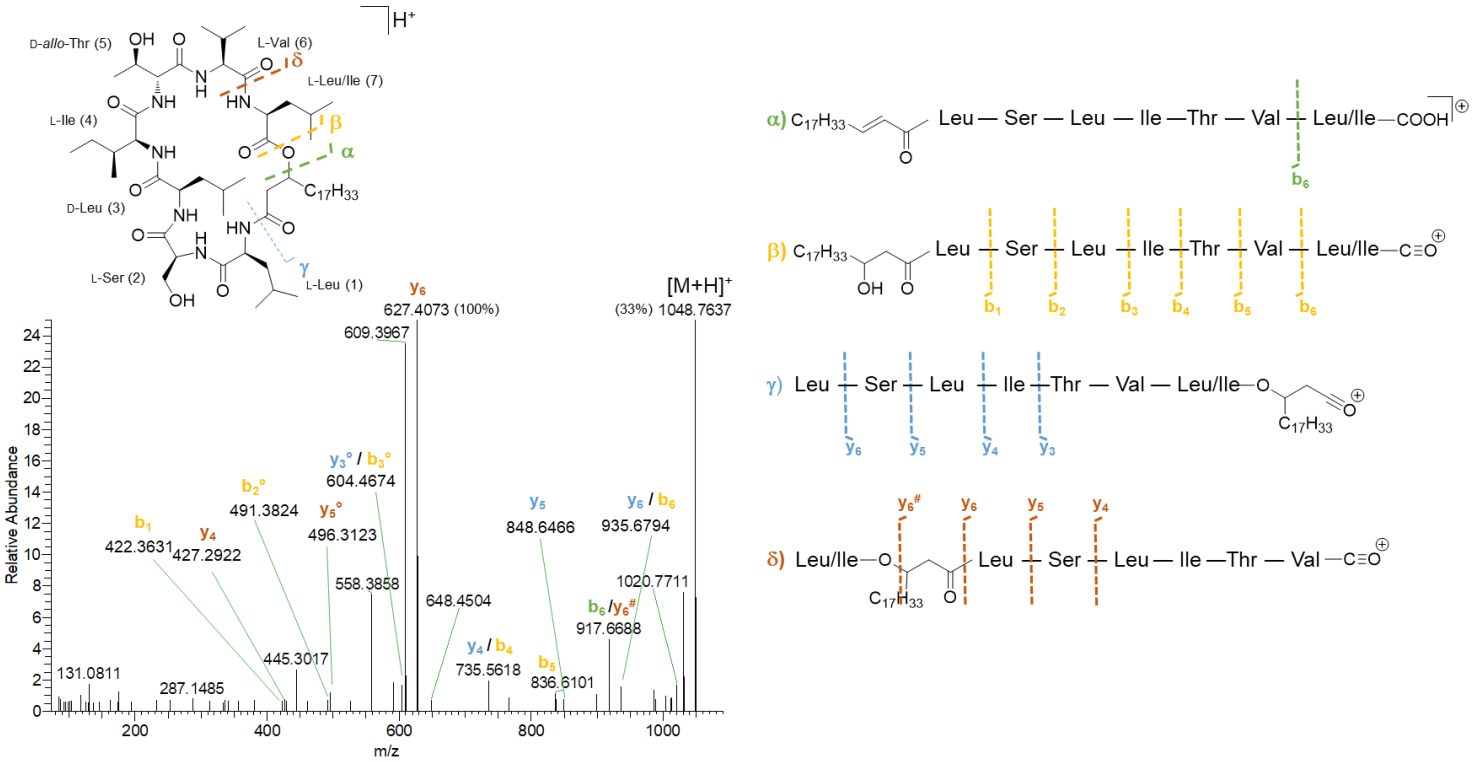

Figure S11. HR-MS^2 spectra of the [M+H]$^+$ pseudomolecular ion of rhodoheptin H (**8**).

rhodoheptin I (**9**)

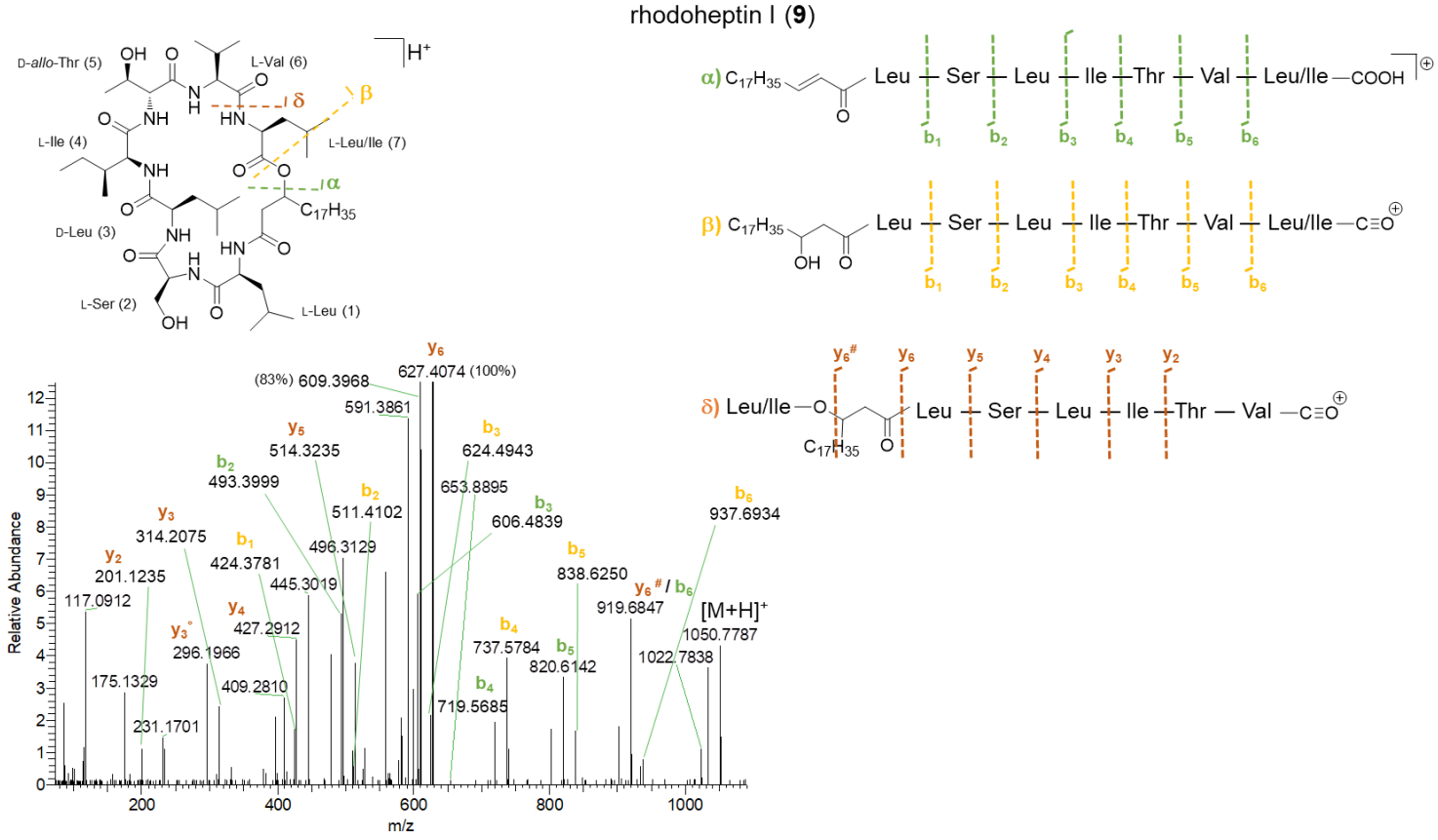

Figure S12. HR-MS^2 spectra of the [M+H]$^+$ pseudomolecular ion of rhodoheptin I (**9**).

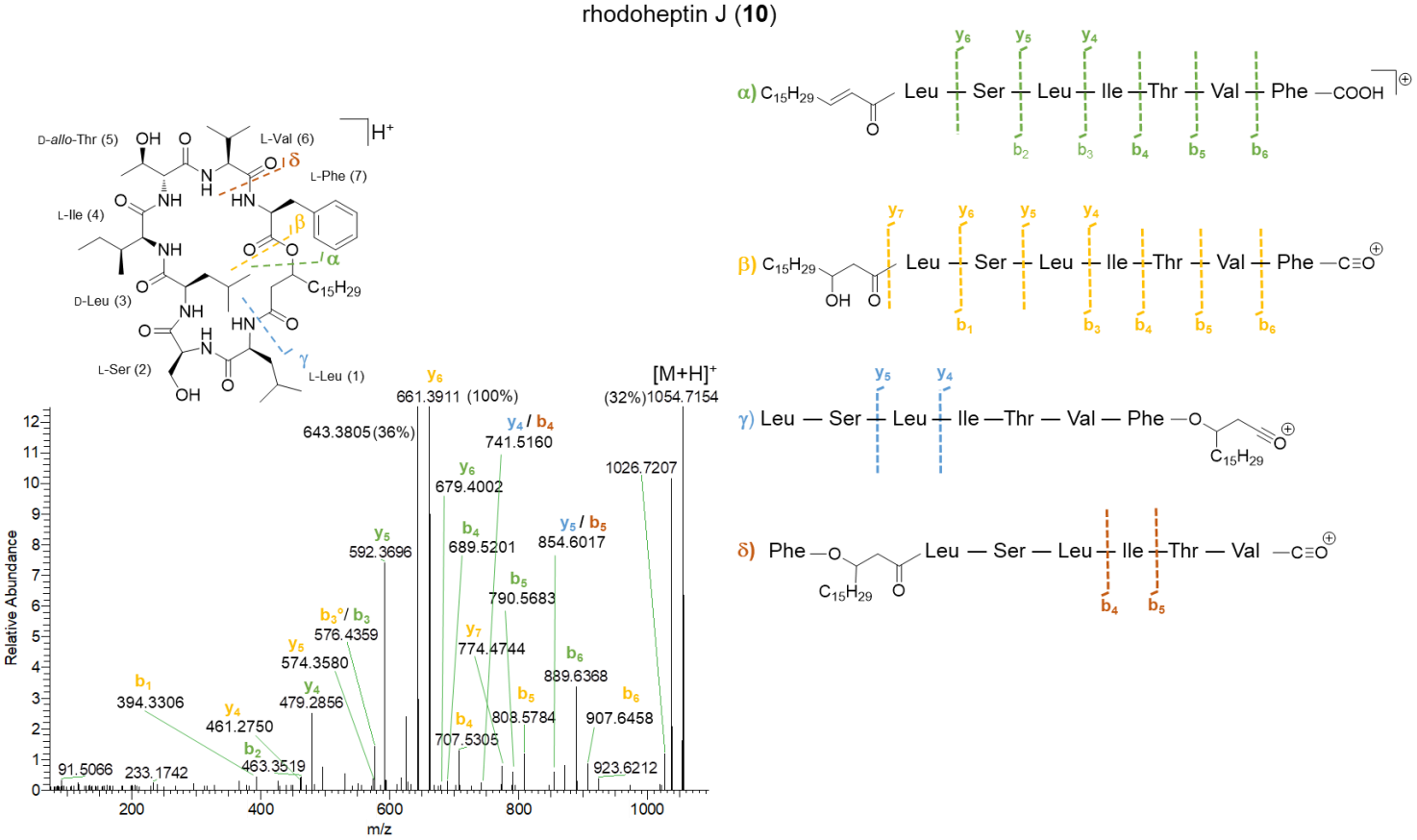

Figure S13. HR-MS^2 spectra of the [M+H]$^+$ pseudomolecular ion of rhodoheptin J (**10**).

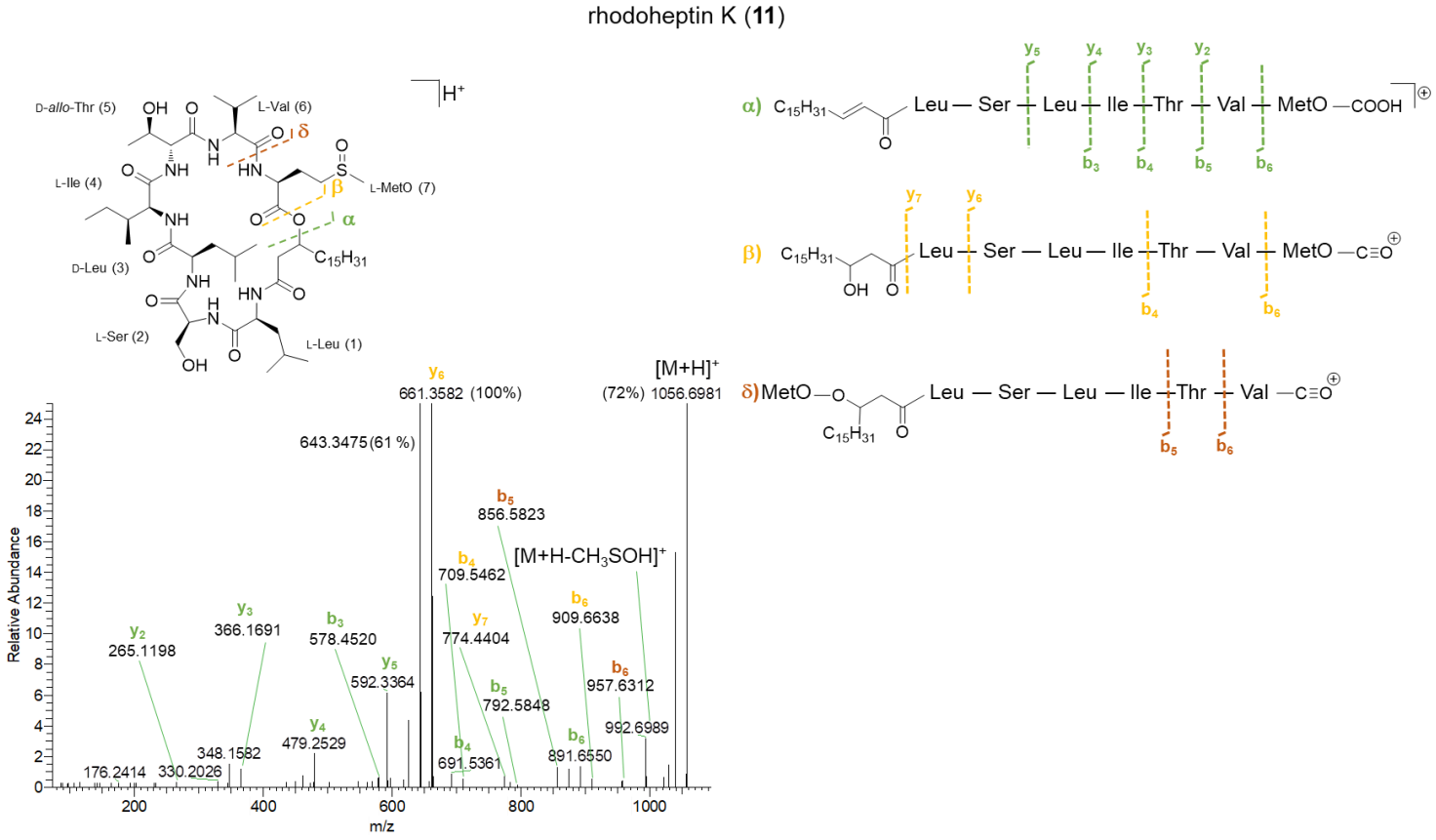

Figure S14. HR-MS^2 spectra of the [M+H]$^+$ pseudomolecular ion of rhodoheptin K (**11**).

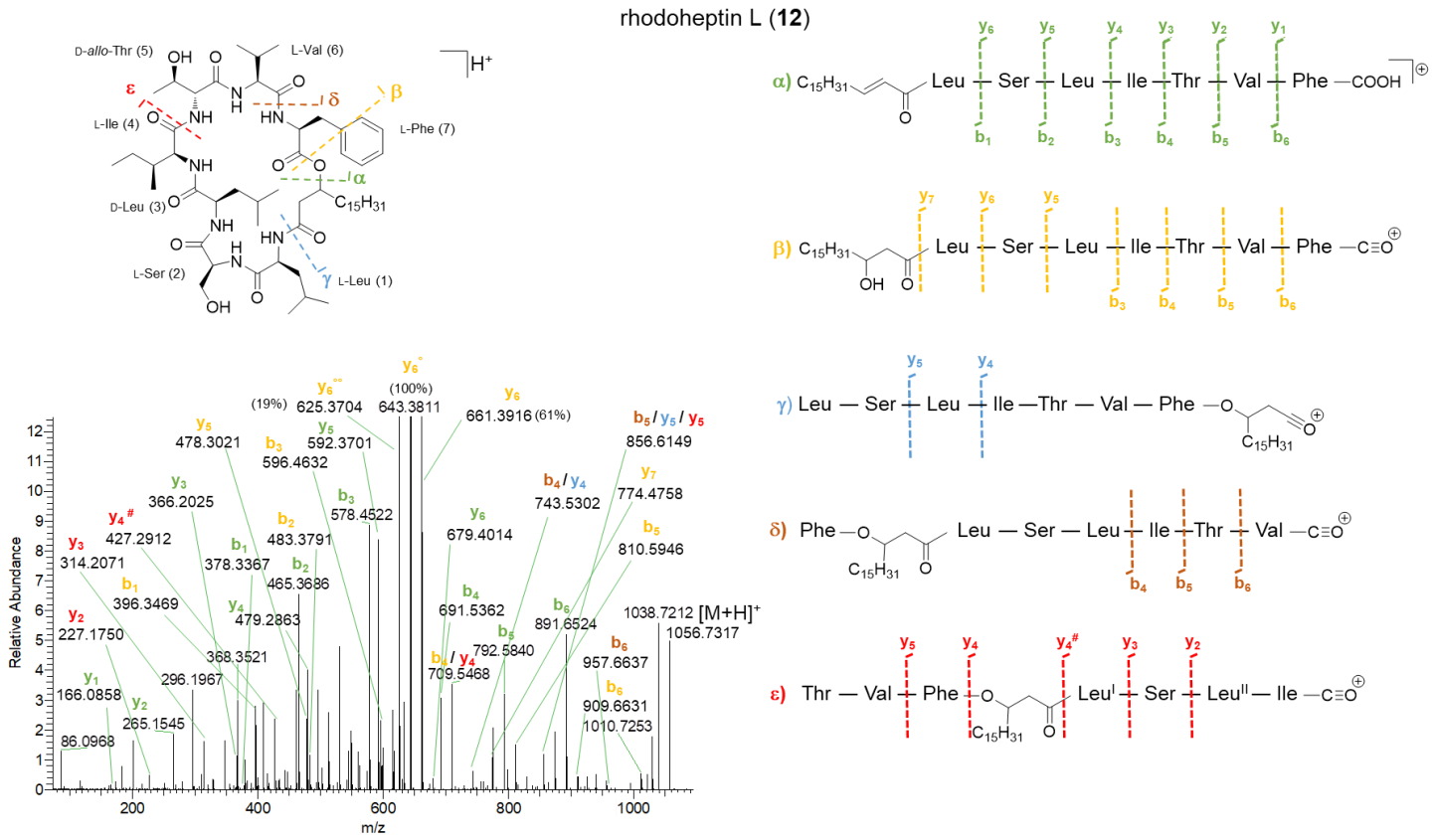

Figure S15. HR-MS^2 spectra of the [M+H]$^+$ pseudomolecular ion of rhodoheptin L (**12**).

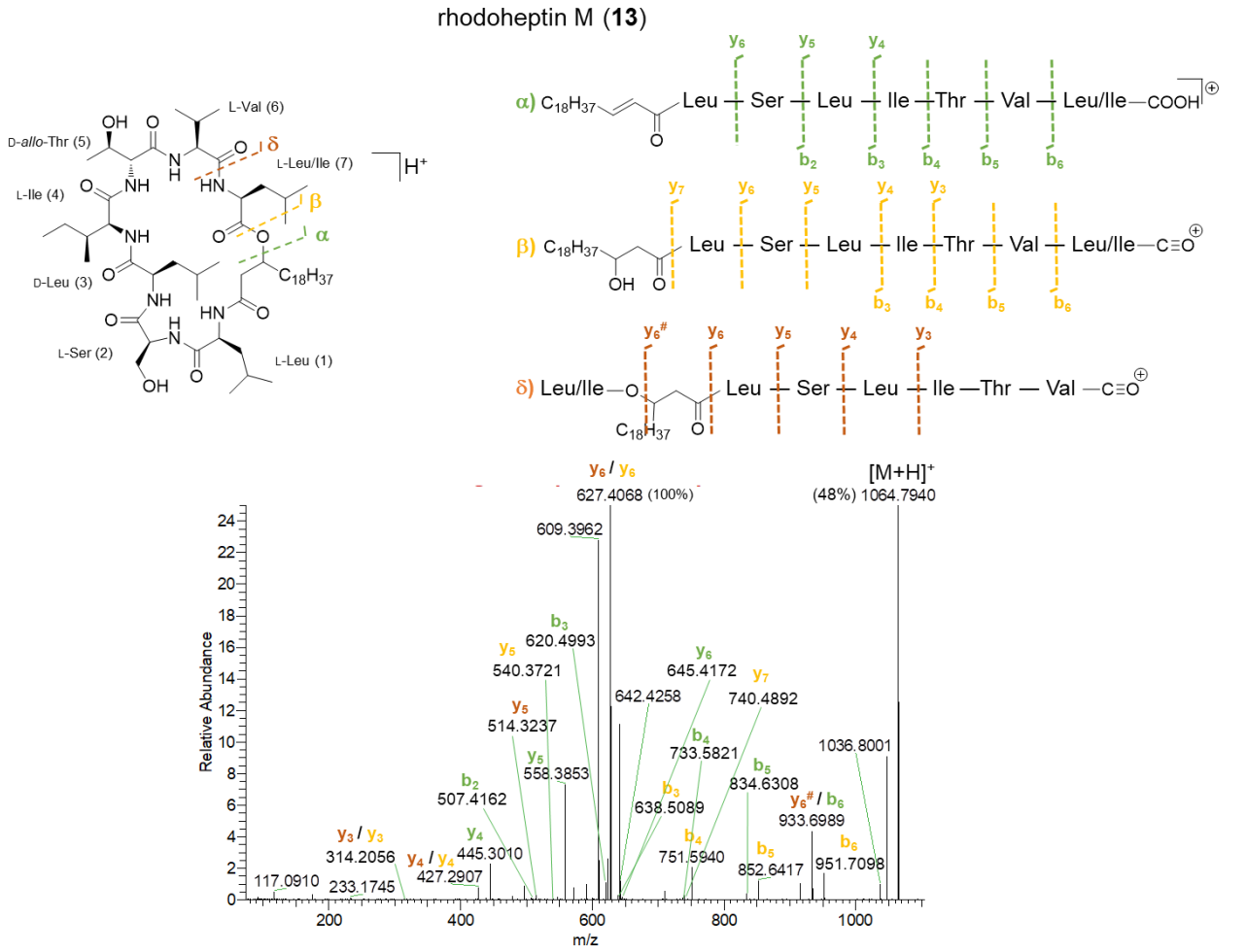

Figure S16. HR-MS^2 spectra of the [M+H]$^+$ pseudomolecular ion of rhodoheptin M (**13**).

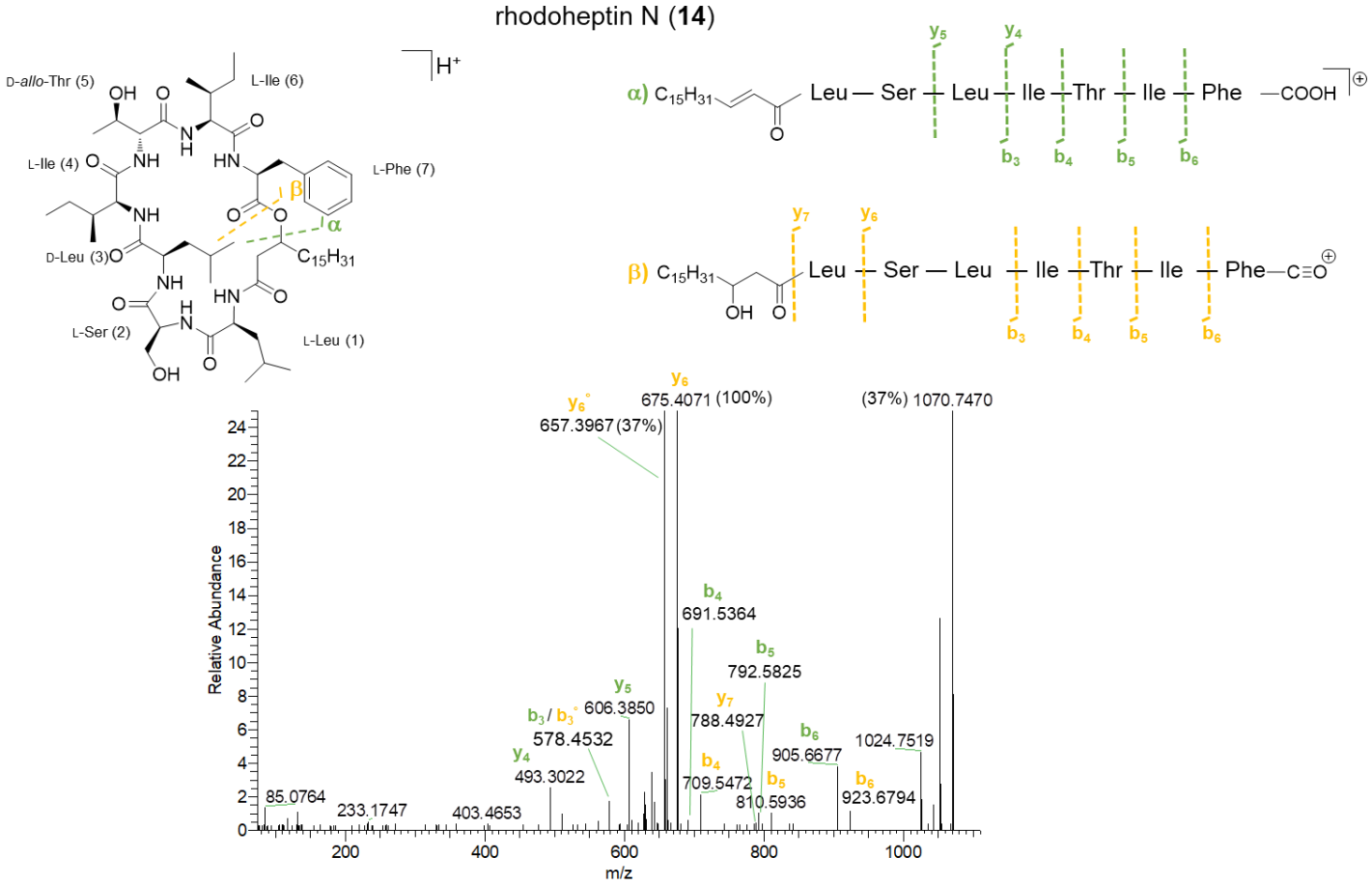

Figure S17. HR-MS^2 spectra of the [M+H]$^+$ pseudomolecular ion of rhodoheptin N (**14**).

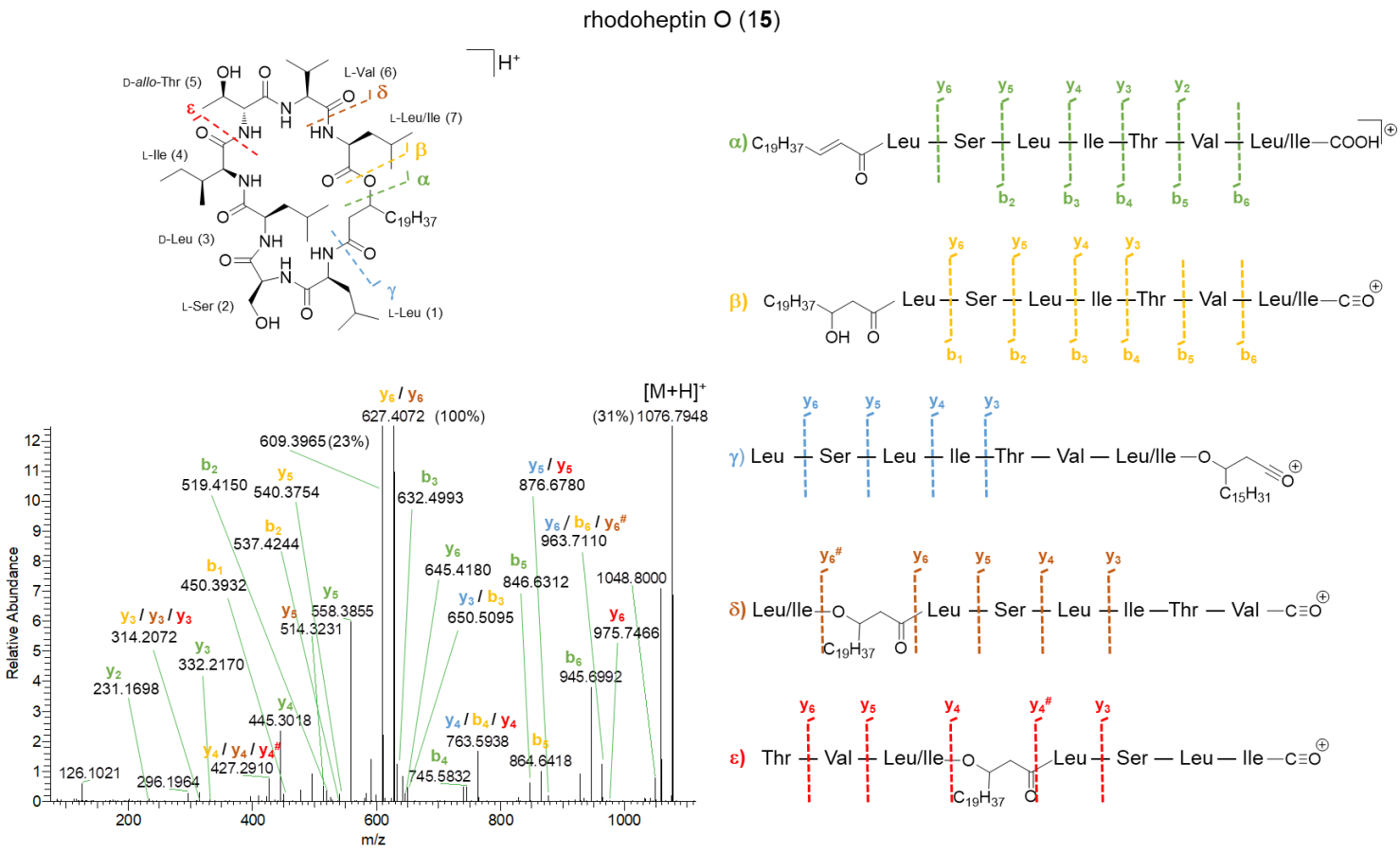

Figure S18. HR-MS^2 spectra of the [M+H]$^+$ pseudomolecular ion of rhodoheptin O (**15**).

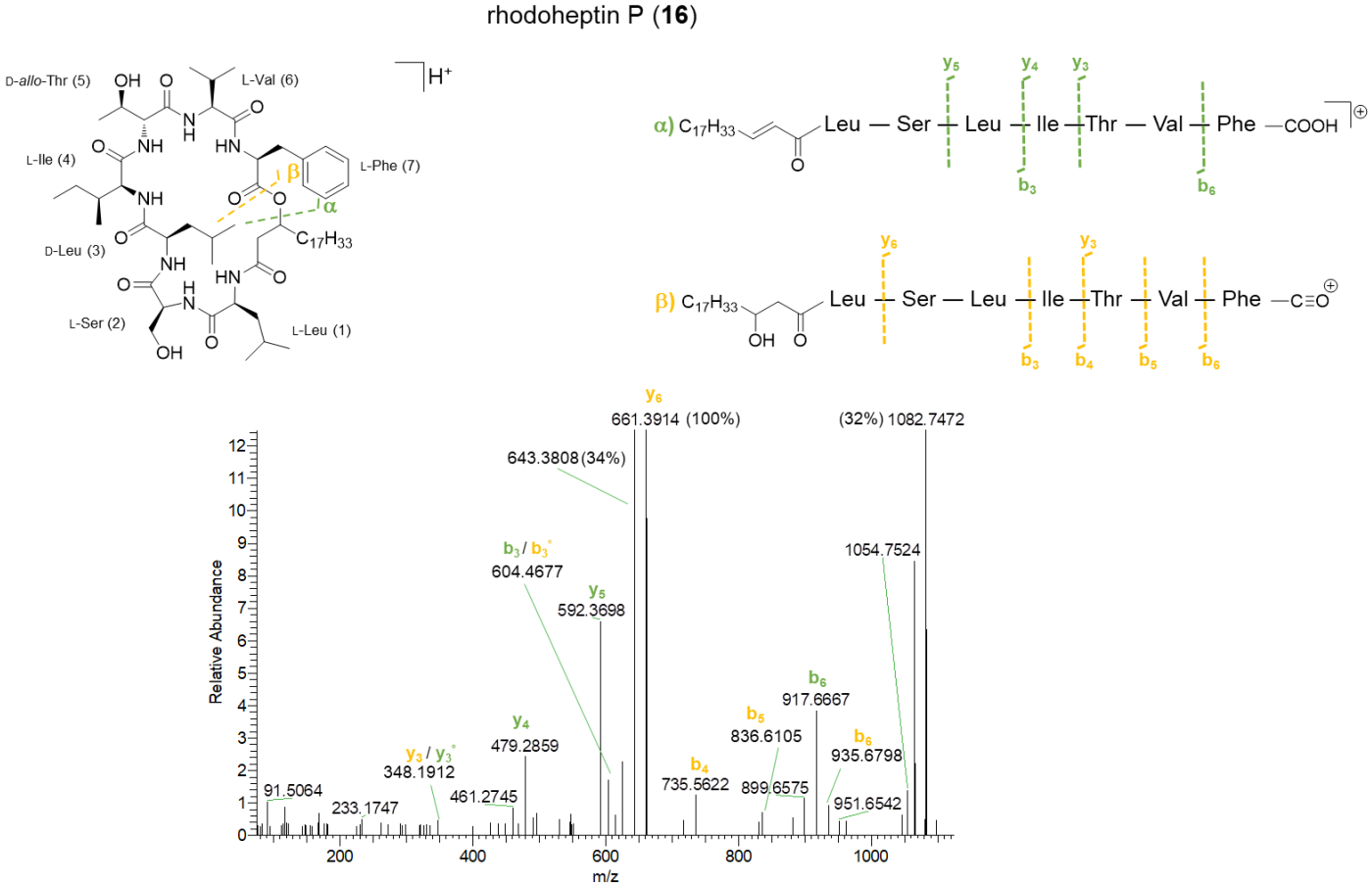

Figure S19. HR-MS^2 spectra of the [M+H]$^+$ pseudomolecular ion of rhodoheptin P (**16**).

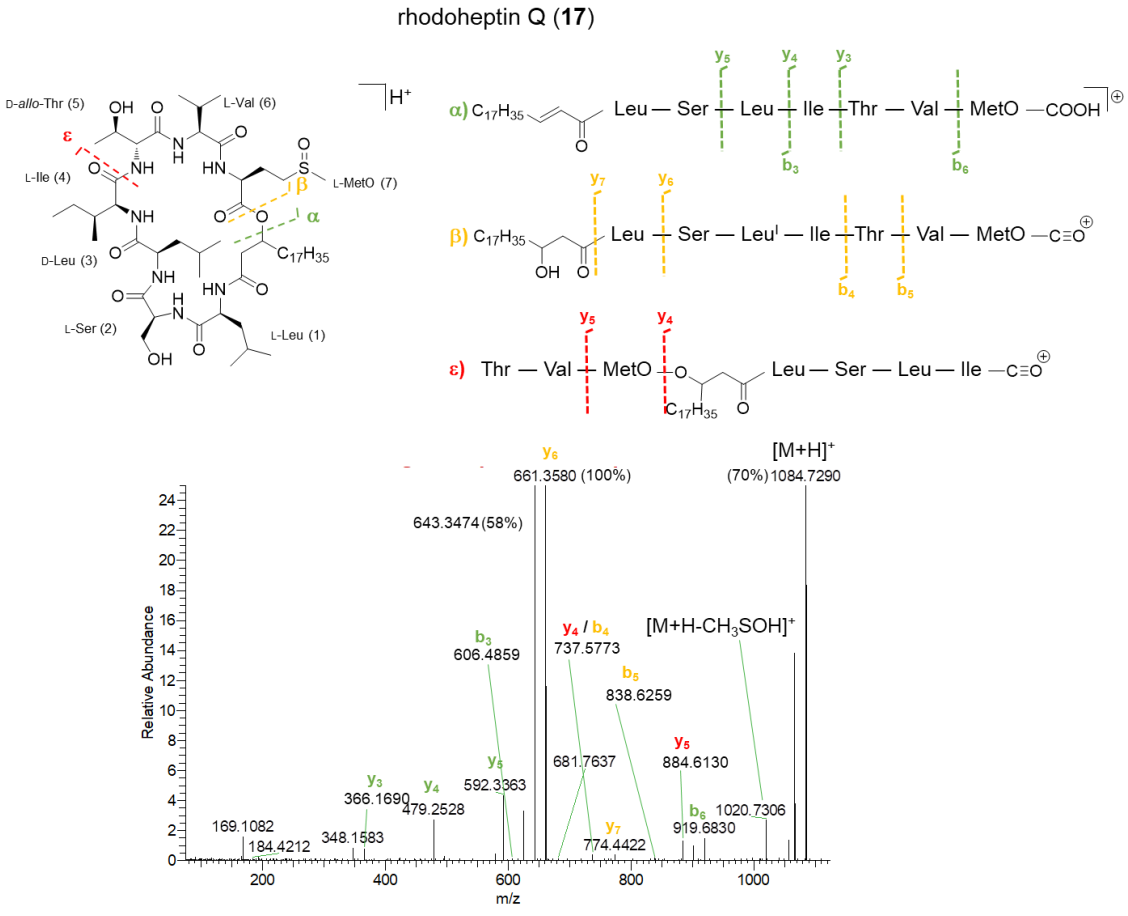

Figure S20. HR-MS^2 spectra of the [M+H]$^+$ pseudomolecular ion of rhodoheptin Q (**17**).

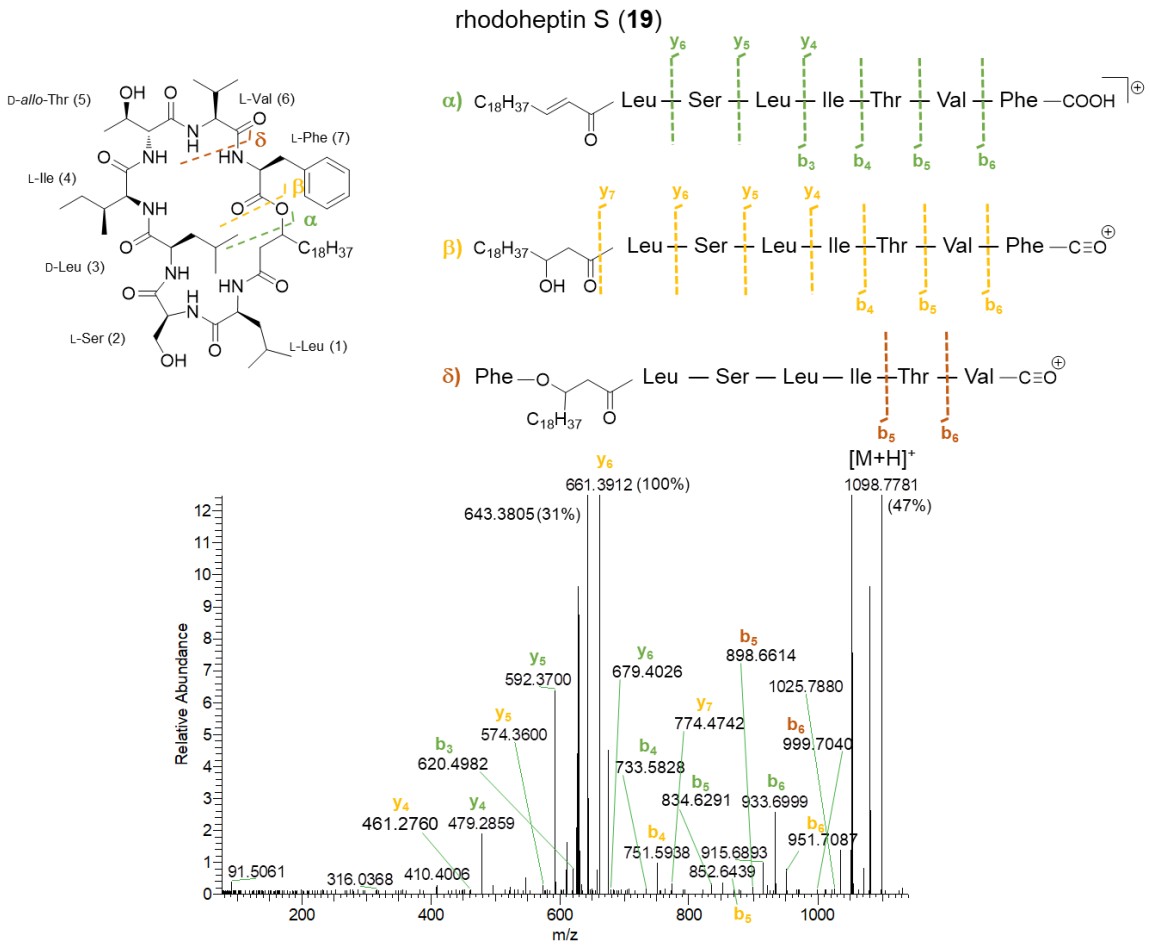

Figure S21. HR-MS^2 spectra of the [M+H]$^+$ pseudomolecular ion of rhodoheptin S (**19**).

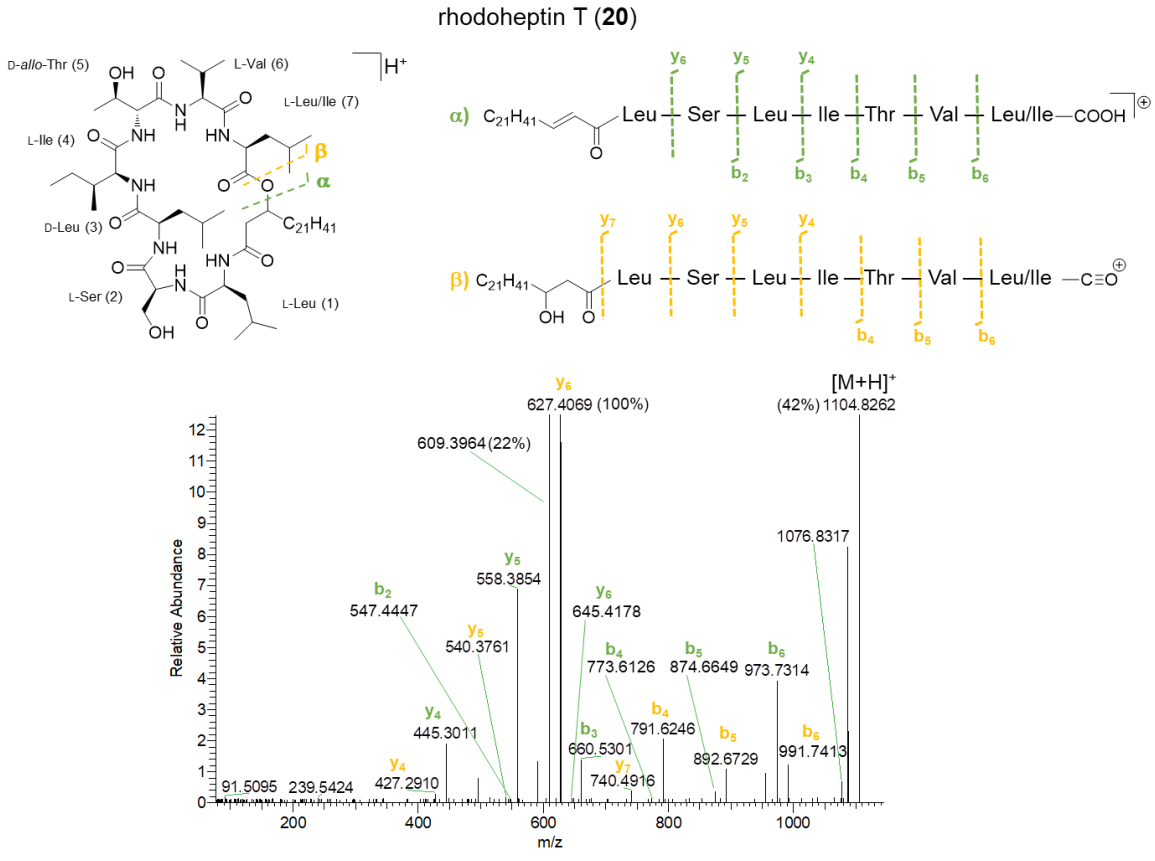

Figure S22. HR-MS^2 spectra of the [M+H]$^+$ pseudomolecular ion of rhodoheptin T (**20**).

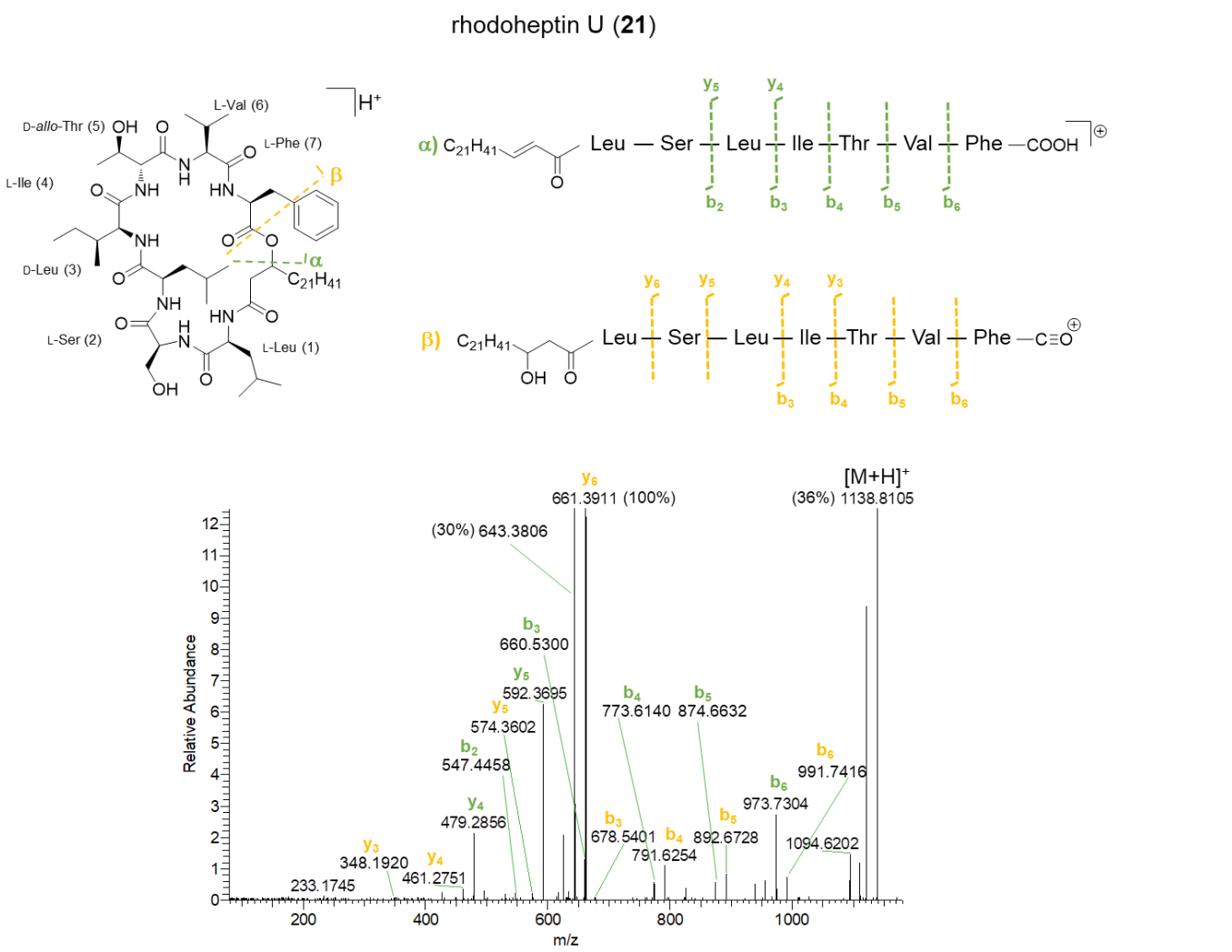

Figure S23. HR-MS^2 spectra of the $[M+H]^+$ pseudomolecular ion of rhodoheptin U (**21**).

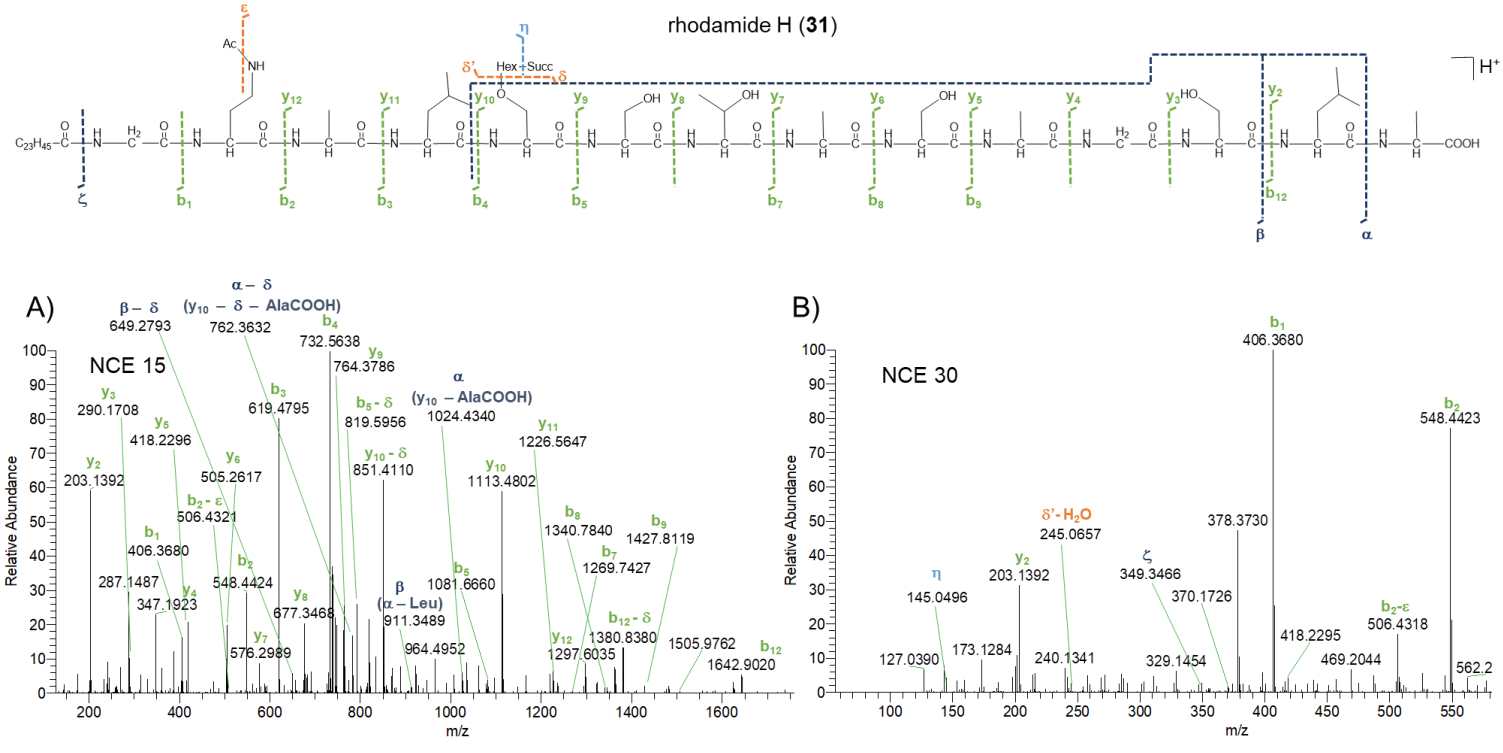

Figure S24. HR-MS^2 spectra of the $[M+2H]^{2+}$ pseudomolecular ion of rhodamide H (**31**) acquired at NCE 15 (A) and 30 (B). Hex = hexosyl; Succ = succinyl; Ac = acetyl.

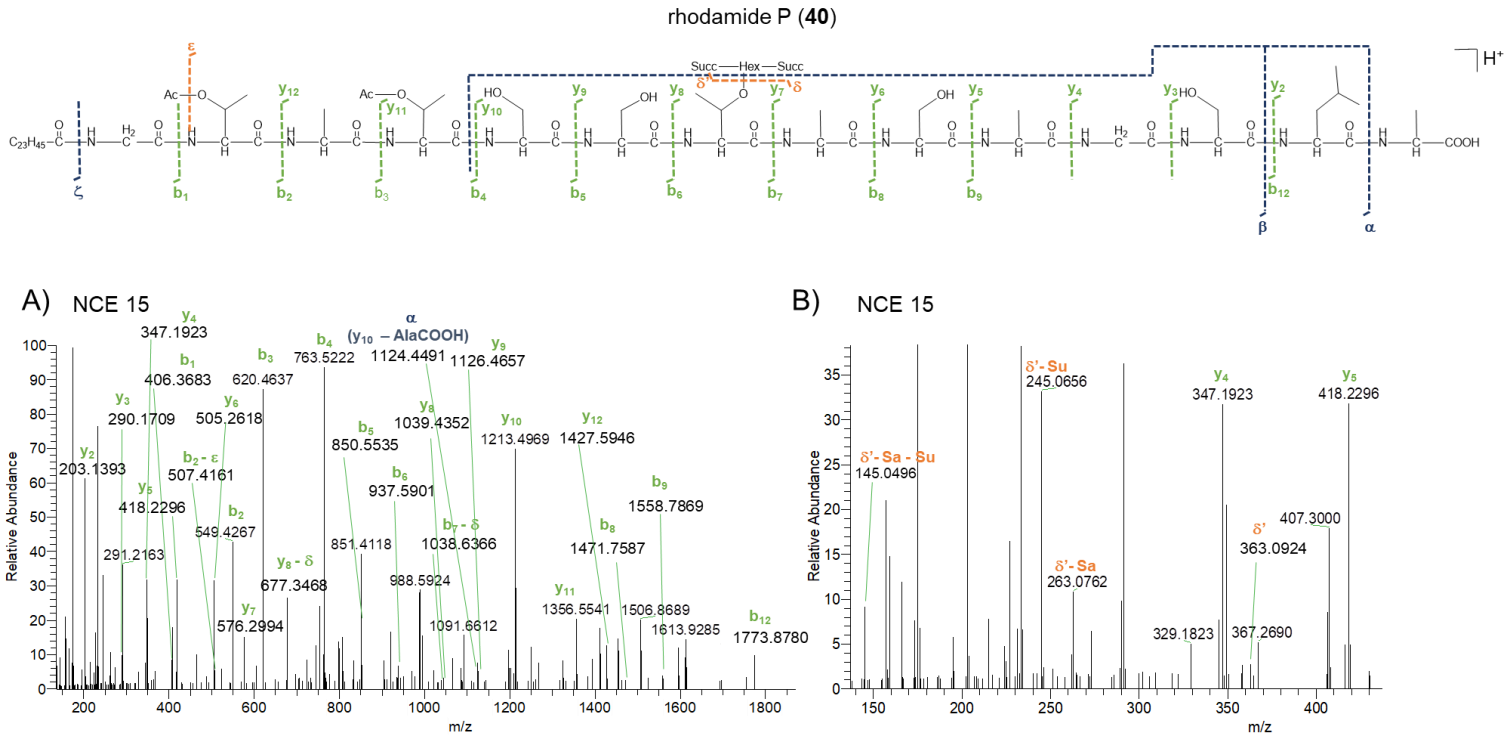

Figure S25. HR-MS^2 spectrum of the $[M+2H]^{2+}$ pseudomolecular ion of rhodamide P (**40**) acquired at NCE 15 (A). Zoom in HR-MS^2 spectrum of the $[M+2H]^{2+}$ pseudomolecular ion of rhodamide P (**40**) acquired at NCE 15 over the *m/z* range 140-430 (B). Hex = hexosyl; Succ = succinyl; Ac = acetyl; Sa, succinic anhydride; Su, succinic acid.

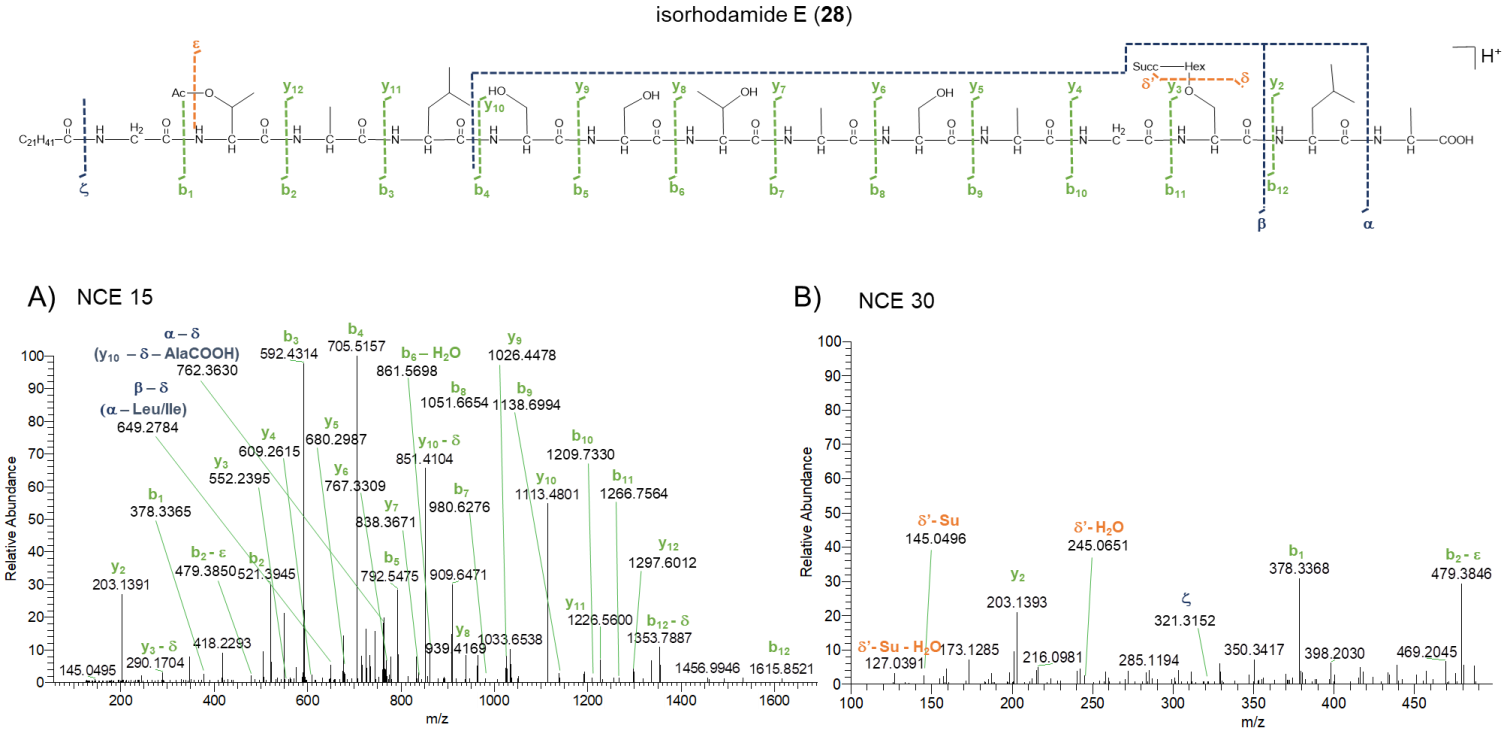

Figure S26. HR-MS^2 spectrum of the $[M+2H]^{2+}$ pseudomolecular ion of isorhodamide E (**28**) acquired at NCE 15 (A). Zoom in HR-MS^2 spectrum of the $[M+2H]^{2+}$ pseudomolecular ion of isorhodamide E (**28**) acquired at NCE 30 over the *m/z* range 100-500 (B). Hex = hexosyl; Succ = succinyl; Ac = acetyl; Su, succinic acid.

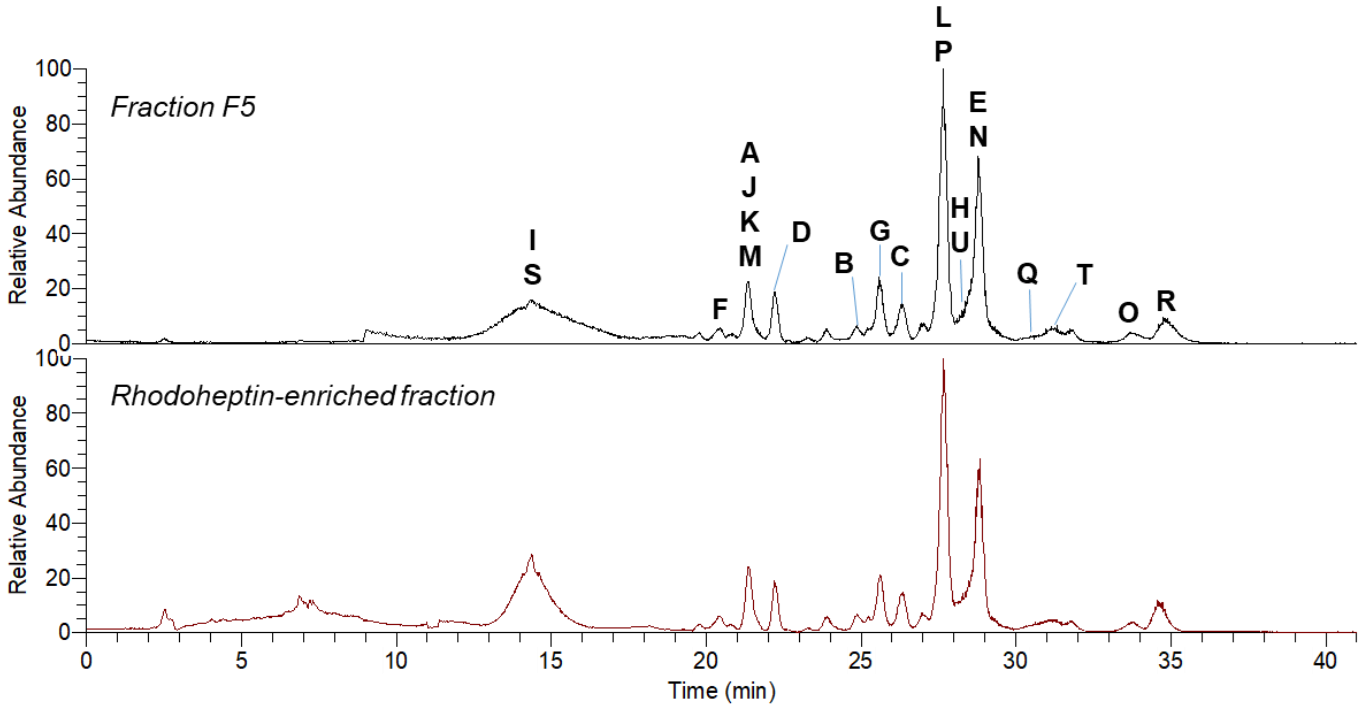

Figure S27. Extracted LC-MS ion chromatograms (using a 5 ppm mass tolerance window) for rhodoheptins A-U from the F5 fraction and the purified rhodoheptin-enriched fraction.

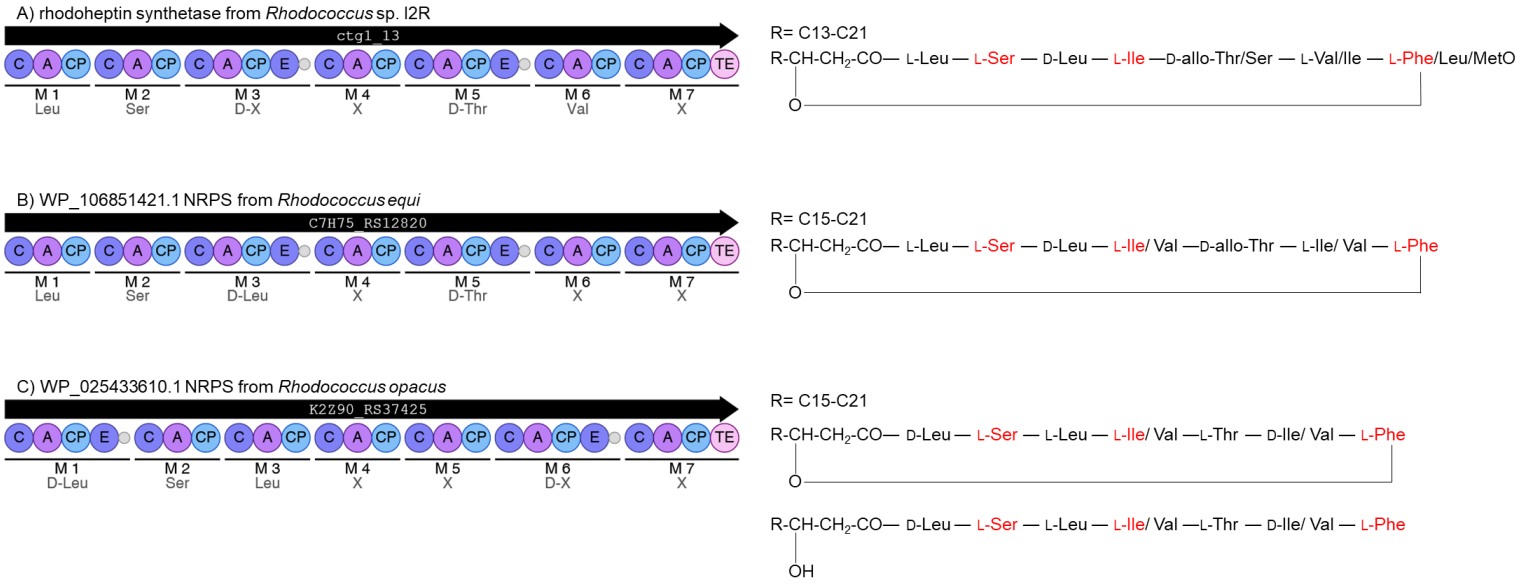

Figure S28. AntiSMASH analysis of *rhp* gene cluster from *Rhodococcus* sp. I2R (A) and rhodoheptin-related biosynthetic gene cluster from *R. equi* (B, WP_106851421.1) and *R. opacus* (C, WP_025433610.1) and general structures of the encoded metabolites. Conserved amino acid residues in the lipopeptide backbone are highlighted in red.

Figure S29. Clinker visualisation of the 50 most (based on cblaster score) similar BGC to the rhodoheptin BGC.

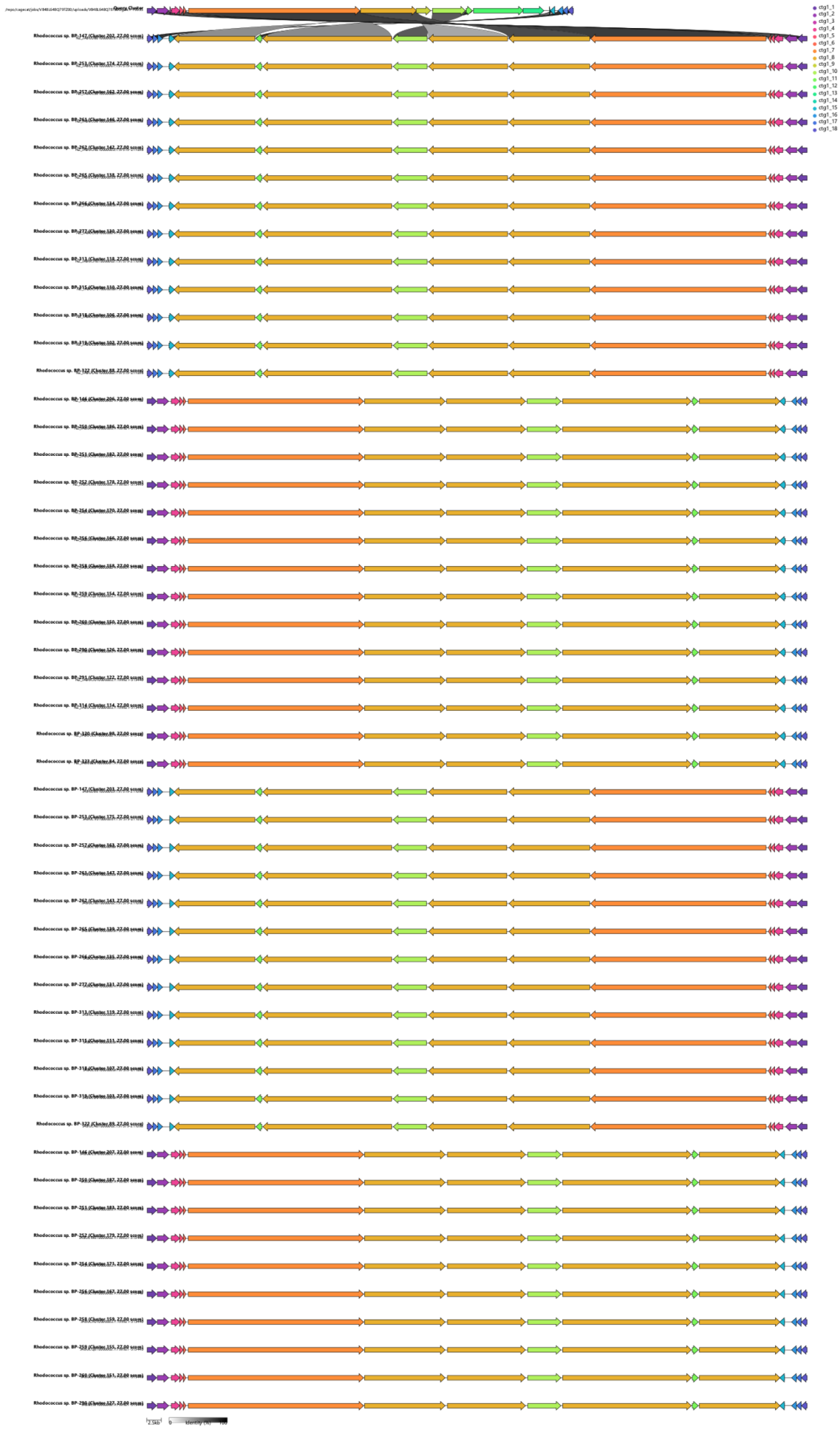

Figure S30. Clinker visualisation of the 50 most (based on cblaster score) similar BGC to the rhodamide BGC.