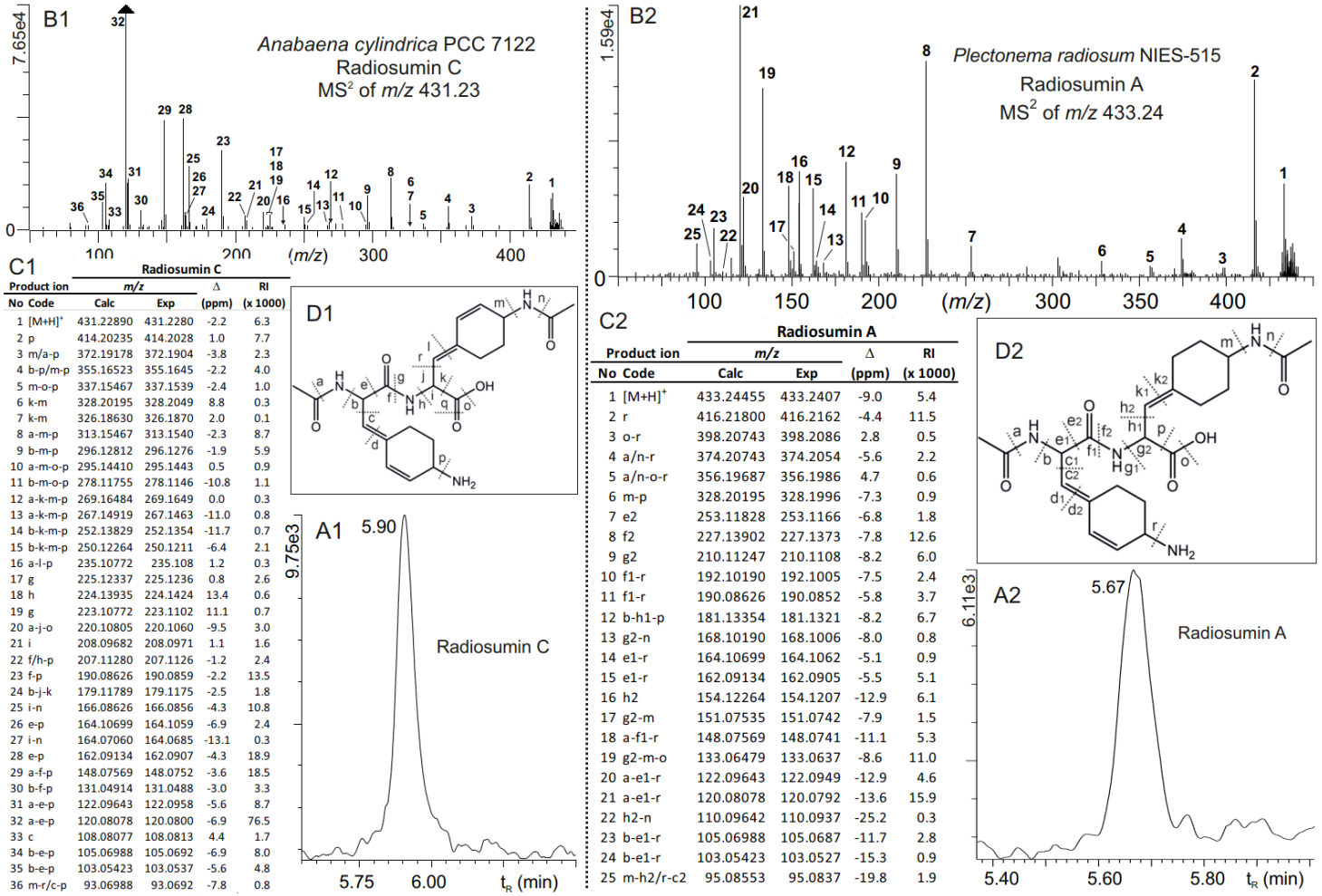
Supplementary information for

Direct pathway cloning and expression of the radiosumin biosynthetic gene cluster

Xiaodan Ouyang,[a] Paul M. D'Agostino,[b] Matti Wahlsten,[a] Endrews Delbaje,[c] Jouni Jokela,[a] Perttu Permi,[d,e] Greta Gaiani,[a] Antti Poso,[f] Piia Bartos,[f] Tobias A. M. Gulder,[b] Hannu Koistinen,[g] David P. Fewer [a,*]
a. Department of Microbiology, University of Helsinki, Viikinkaari 9, FI-00014 Helsinki, Finland
b. Chair of Technical Biochemistry, Technical University of Dresden, Bergstraße 66, 01069 Dresden, Germany
c. Center for Nuclear Energy in Agriculture, University of São Paulo, Avenida Centenário 303, Piracicaba 13400-970, São Paulo, Brazil
d. Department of Chemistry, University of Jyväskylä, Finland
e. Department of Biological and Environmental Science, Nanoscience Center, University of Jyväskylä, Finland
f. Kuopio Campus, School of Pharmacy, University of Eastern Finland
g. Department of Clinical Chemistry and Haematology, Faculty of Medicine, University of Helsinki and Helsinki University Hospital, Helsinki FIN-00014, Finland
* Corresponding author
E-mail address: david.fewer@helsinki.fi



Product ion No	Code	m/z Calc	m/z Exp	Δ (ppm)	RI (x 1000)
1	[M+H]⁺	431.22890	431.2280	-2.2	6.3
2	p	414.20235	414.2028	1.0	7.7
3	m/a-p	372.19178	372.1904	-3.8	2.3
4	b-p/m-p	355.16523	355.1645	-2.2	4.0
5	m-o-p	337.15467	337.1539	-2.4	1.0
6	k-m	328.20195	328.2049	8.8	0.3
7	k-m	326.18630	326.1870	2.0	0.1
8	a-m-p	313.15467	313.1540	-2.3	8.7
9	b-m-p	296.12812	296.1276	-1.9	5.9
10	a-m-o-p	295.14410	295.1443	0.5	0.9
11	b-m-o-p	278.11755	278.1146	-10.8	1.1
12	a-k-m-p	269.16484	269.1649	0.0	0.3
13	a-k-m-p	267.14919	267.1463	-11.0	0.8
14	b-k-m-p	252.13829	252.1354	-11.7	0.7
15	b-k-m-p	250.12264	250.1211	-6.4	2.1
16	a-l-p	235.10772	235.108	1.2	0.3
17	g	225.12337	225.1236	0.8	2.6
18	h	224.13935	224.1424	13.4	0.6
19	g	223.10772	223.1102	11.1	0.7
20	a-j-o	220.10805	220.1060	-9.5	3.0
21	i	208.09682	208.0971	1.1	1.6
22	f/h-p	207.11280	207.1126	-1.2	2.4
23	f-p	190.08626	190.0859	-2.2	13.5
24	b-j-k	179.11789	179.1175	-2.5	1.8
25	i-n	166.08626	166.0856	-4.3	10.8
26	e-p	164.10699	164.1059	-6.9	2.4
27	i-n	164.07060	164.0685	-13.1	0.3
28	e-p	162.09134	162.0907	-4.3	18.9
29	a-f-p	148.07569	148.0752	-3.6	18.5
30	b-f-p	131.04914	131.0488	-3.0	3.3
31	a-e-p	122.09643	122.0958	-5.6	8.7
32	a-e-p	120.08078	120.0800	-6.9	76.5
33	c	108.08077	108.0813	4.4	1.7
34	b-e-p	105.06988	105.0692	-6.9	8.0
35	b-e-p	103.05423	103.0537	-5.6	4.8
36	m-r/c-p	93.06988	93.0692	-7.8	0.8

Product ion No	Code	m/z Calc	m/z Exp	Δ (ppm)	RI (x 1000)
1	[M+H]⁺	433.24455	433.2407	-9.0	5.4
2	r	416.21800	416.2162	-4.4	11.5
3	o-r	398.20743	398.2086	2.8	0.5
4	a/n-r	374.20743	374.2054	-5.6	2.2
5	a/n-o-r	356.19687	356.1986	4.7	0.6
6	m-p	328.20195	328.1996	-7.3	0.9
7	e2	253.11828	253.1166	-6.8	1.8
8	f2	227.13902	227.1373	-7.8	12.6
9	g2	210.11247	210.1108	-8.2	6.0
10	f1-r	192.10190	192.1005	-7.5	2.4
11	f1-r	190.08626	190.0852	-5.8	3.7
12	b-h1-p	181.13354	181.1321	-8.2	6.7
13	g2-n	168.10190	168.1006	-8.0	0.8
14	e1-r	164.10699	164.1062	-5.1	0.9
15	e1-r	162.09134	162.0905	-5.5	5.1
16	h2	154.12264	154.1207	-12.9	6.1
17	g2-m	151.07535	151.0742	-7.9	1.5
18	a-f1-r	148.07569	148.0741	-11.1	5.3
19	g2-m-o	133.06479	133.0637	-8.6	11.0
20	a-e1-r	122.09643	122.0949	-12.9	4.6
21	a-e1-r	120.08078	120.0792	-13.6	15.9
22	h2-n	110.09642	110.0937	-25.2	0.3
23	b-e1-r	105.06988	105.0687	-11.7	2.8
24	b-e1-r	103.05423	103.0527	-15.3	0.9
25	m-h2/r-c2	95.08553	95.0837	-19.8	1.9

Figure S1 Ultra-performance liquid chromatograph with quadrupole time-of-flight (UPLC-QTOF) spectrometry analysis of radiosumin C (**3**) from *Anabaena cylindrica* PCC 7122 and radiosumin A (**1**) from *Plectonema radiosum* NIES-515. (A) Extracted ion chromatogram (*m/z* 431.22 $[M+H]^+$) or (*m/z* 433.24 $[M+H]^+$). (B) Secondary mass (MS^2) spectra (parent ion *m/z* 431.22 or *m/z* 433.24) of the 5.91 min or 5.67 min peak from *m/z* 50 to 450. (C) Table of annotated ions from MS^2 with the difference (Δ) of experimental (Exp) *m/z* value to calculated (Calc) *m/z* value and relative ion intensities (RI). (D) Structure of radiosumin C (**3**) or radiosumin A (**1**).

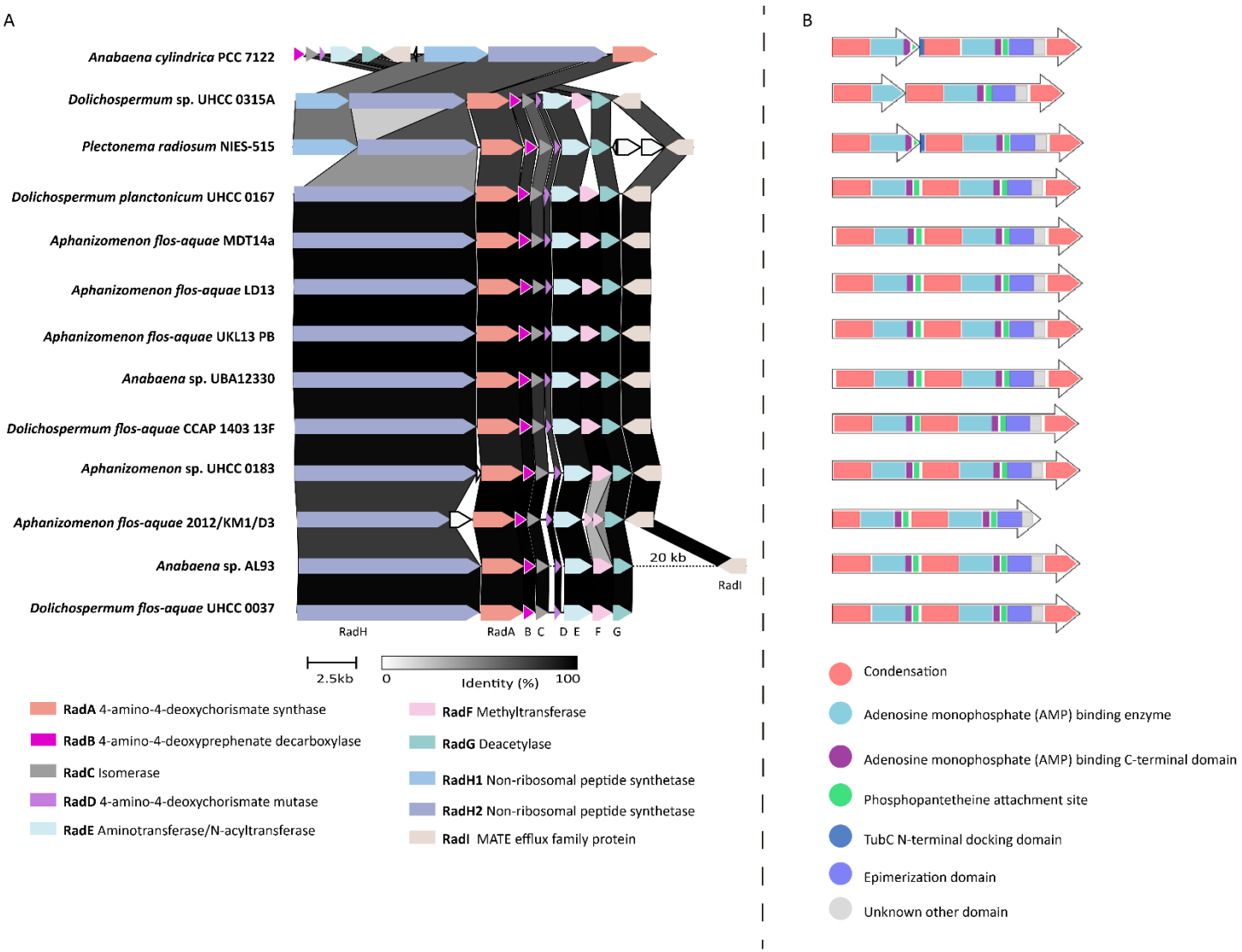


Figure S2 Bioinformatic analysis of radiosumin biosynthetic pathways. (A) Identification and comparison of the radiosumin (*rad*) biosynthetic gene cluster from 13 strains of cyanobacteria. (B) Organization of catalytic domains in the RadH non-ribosomal peptide synthetase.

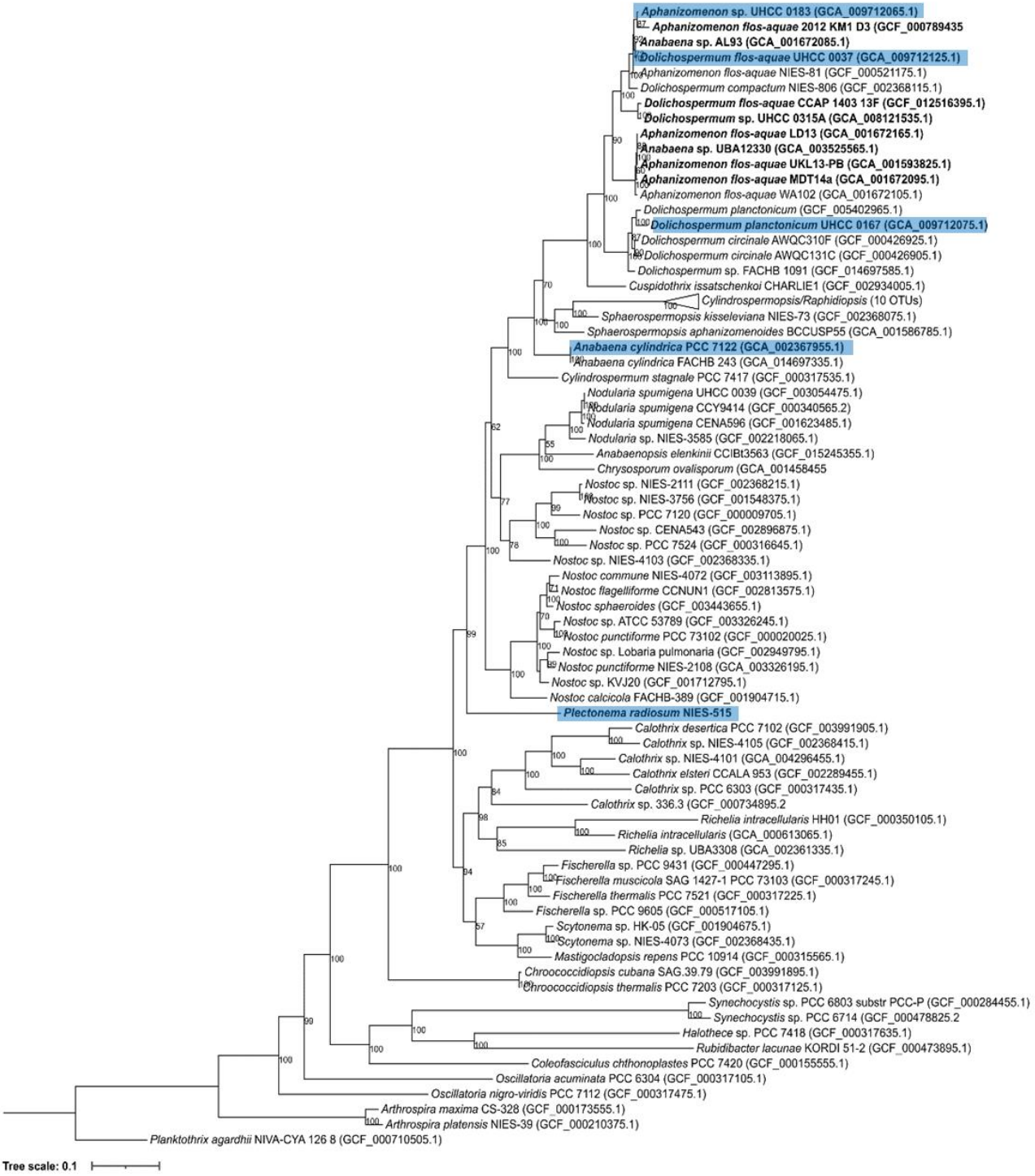


Figure S3 Maximum likelihood phylogenomic tree based on concatenated alignment of 120 bacterial single-copy conserved marker genes from 75 cyanobacterial genomes. Cyanobacteria that encode a radiosumin biosynthetic gene cluster are marked in bold. Radiosumin producers are marked in bold and with a blue box.

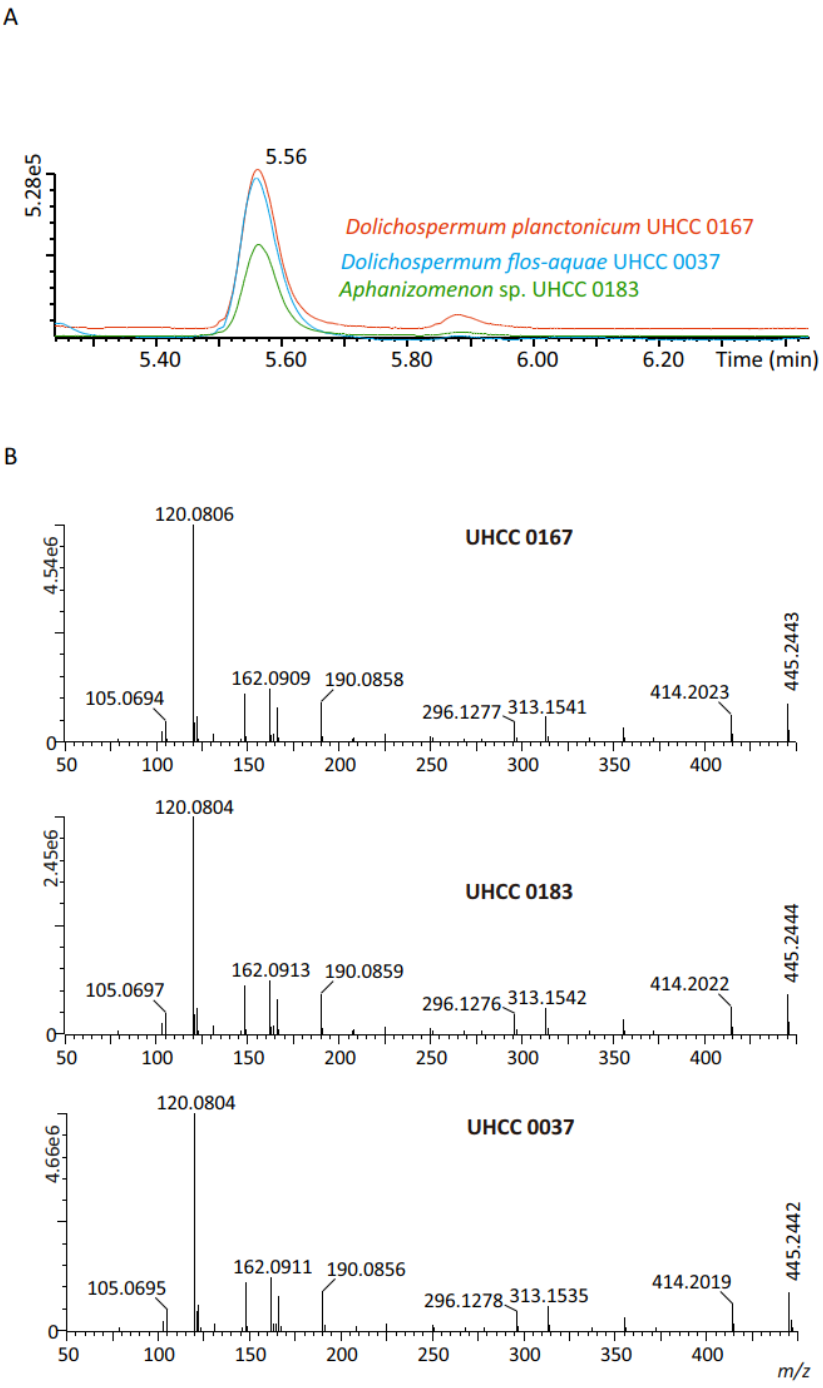
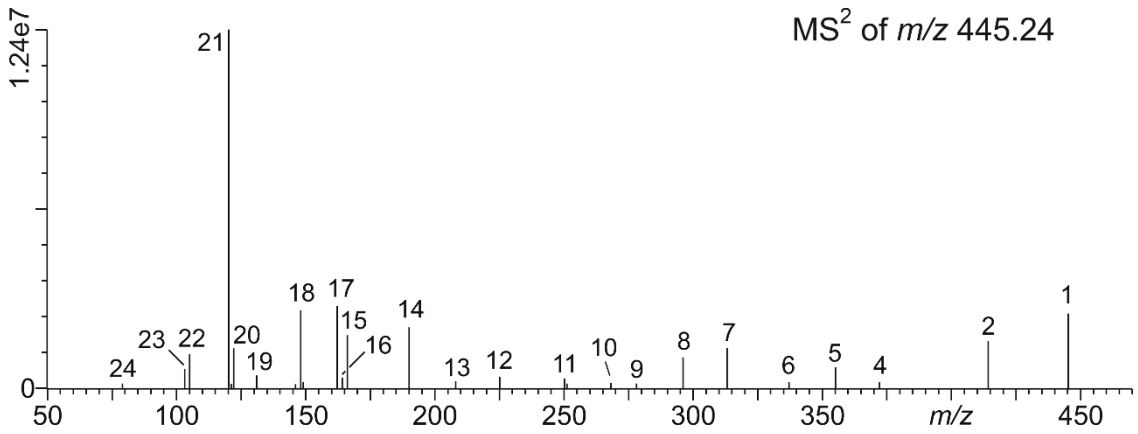
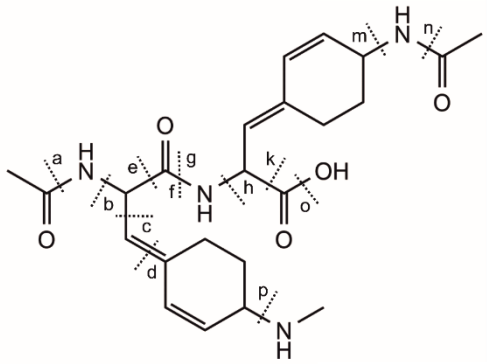


Figure S4 UPLC-QTOF spectrometry analysis of molecule *m/z* 445.24 from three strains. (A) MS^2 (*m/z* 445.24) chromatograms of methanol extracts from *Dolichospermum planctonicum* UHCC 0167, *Dolichospermum flos-aquae* UHCC 0037, and *Aphanizomenon* sp. UHCC 0183. (C) MS^2 spectra of the 5.56 min peaks (*m/z* 445.24) from 50 to 450 of extracts from *Dolichospermum planctonicum* UHCC 0167, *Dolichospermum flos-aquae* UHCC 0037 and *Aphanizomenon* sp. UHCC 0183.



Product ion		molecule 445.24 *m/z*		Δ
No	Code	Calc	Exp	(ppm)
1	$[M+H]^+$	445.24455	445.2440	-1.3
2	p	414.20235	414.2017	-1.7
3	m	386.20743	386.2061	-3.6
4	a-p	372.19178	372.1908	-2.8
5	m-p	355.16523	355.1645	-2.2
6	m-o-p	337.15467	337.1541	-1.8
7	a-m-p	313.15467	313.1539	-2.6
8	b-m-p	296.12812	296.1278	-1.2
9	b-m-o-p	278.11755	278.1167	-3.2
10	a-k-m-p	267.14919	267.1487	-2.0
11	b-k-m-p	250.12264	250.1221	-2.4
12	g	225.12337	225.1232	-1.0
13	h	208.09682	208.0969	0.1
14	f-p	190.08626	190.0864	0.5
15	h-n	166.08626	166.0867	2.3
16	e-p	164.10699	164.1073	1.6
17	e-p	162.09134	162.0916	1.3
18	a-f-p	148.07569	148.0764	4.5
19	b-f-p	131.04914	131.0494	1.6
20	a-e-p / c	122.09643	122.0966	1.0
21	a-e-p / c	120.08078	120.0813	3.9
22	b-e-p	105.06988	105.0702	2.6
23	b-e-p	103.05423	103.0544	1.2
24	d-p	79.05423	79.0541	-2.3



4

Figure S5 MS^2 spectra of protonated molecules *m/z* 445.24 and table of annotated product ions with differences (Δ) of experimental (Exp) *m/z* values to calculated (Calc) *m/z* values).

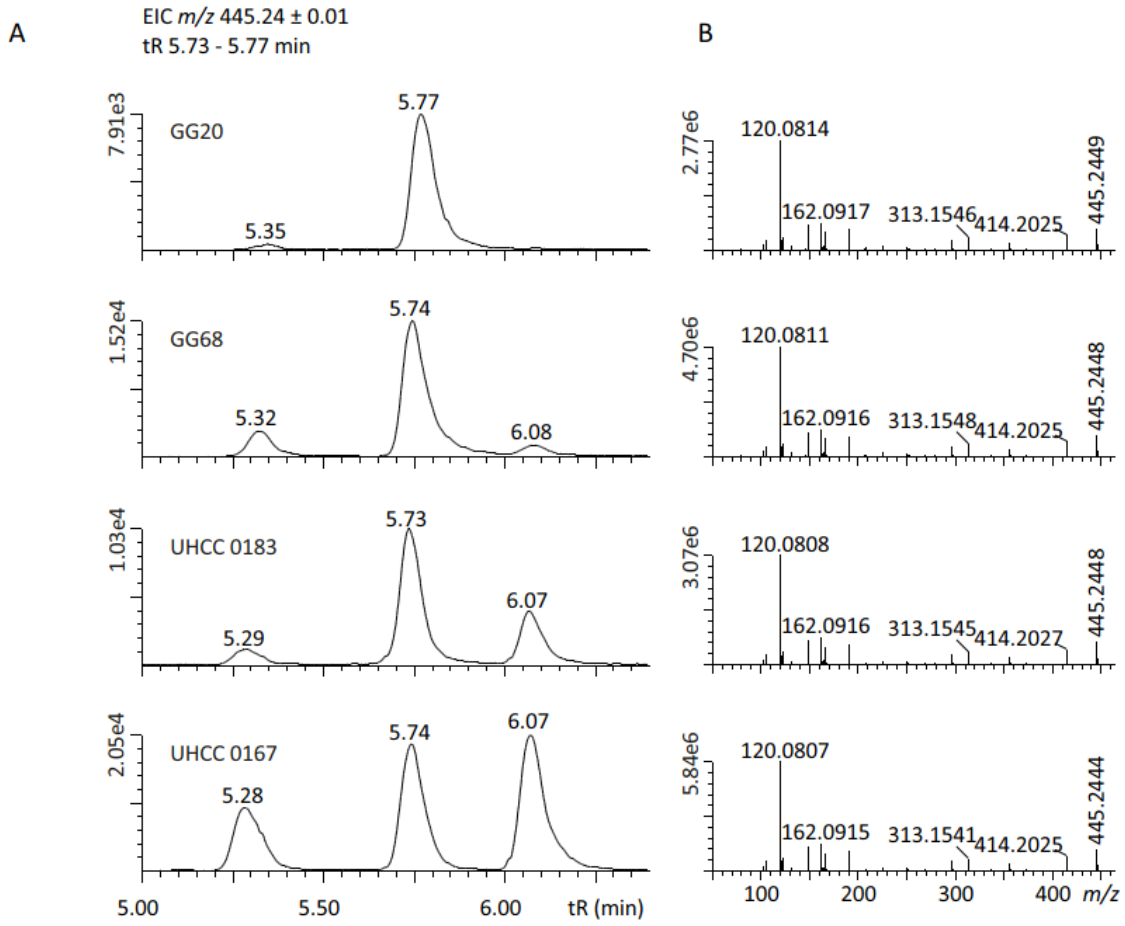


Figure S6 UPLC-QTOF spectrometry analysis of compound **4** in methanol extracts from blooms GG20 and GG68, using *Aphanizomenon* sp. UHCC 0183 and *Dolichospermum planctonicum* UHCC 0167 as controls. (A) Extracted ion chromatograms of methanol extracts from blooms GG20, GG68 and *Aphanizomenon* sp. UHCC 0183 and *Dolichospermum planctonicum* UHCC 0167. (B) MS^2 spectra from *m/z* 50 to 465 of peaks eluting from 5.77 min showed diagnostic product ions for **4** (Figure S5).

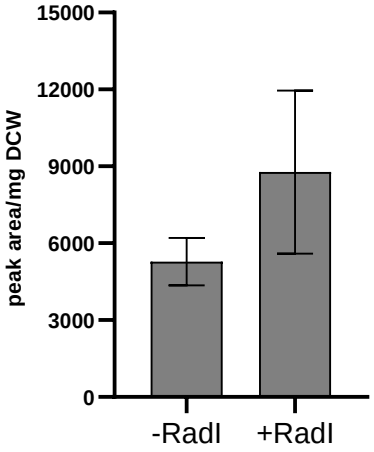


Figure S7 Estimated production levels of compound **4** in *E. coli* BAP1strains, harbouring pET28b-ptetO-*radHABCDEFG-GFP* (-RadI) or pET28b-ptetO-*radHABCDEFGI-GFP* (+RadI). Production levels were calculated by using peak area divided by dry cell weight. Data are representative of triplicate determinations. Error bars indicate standard deviations.

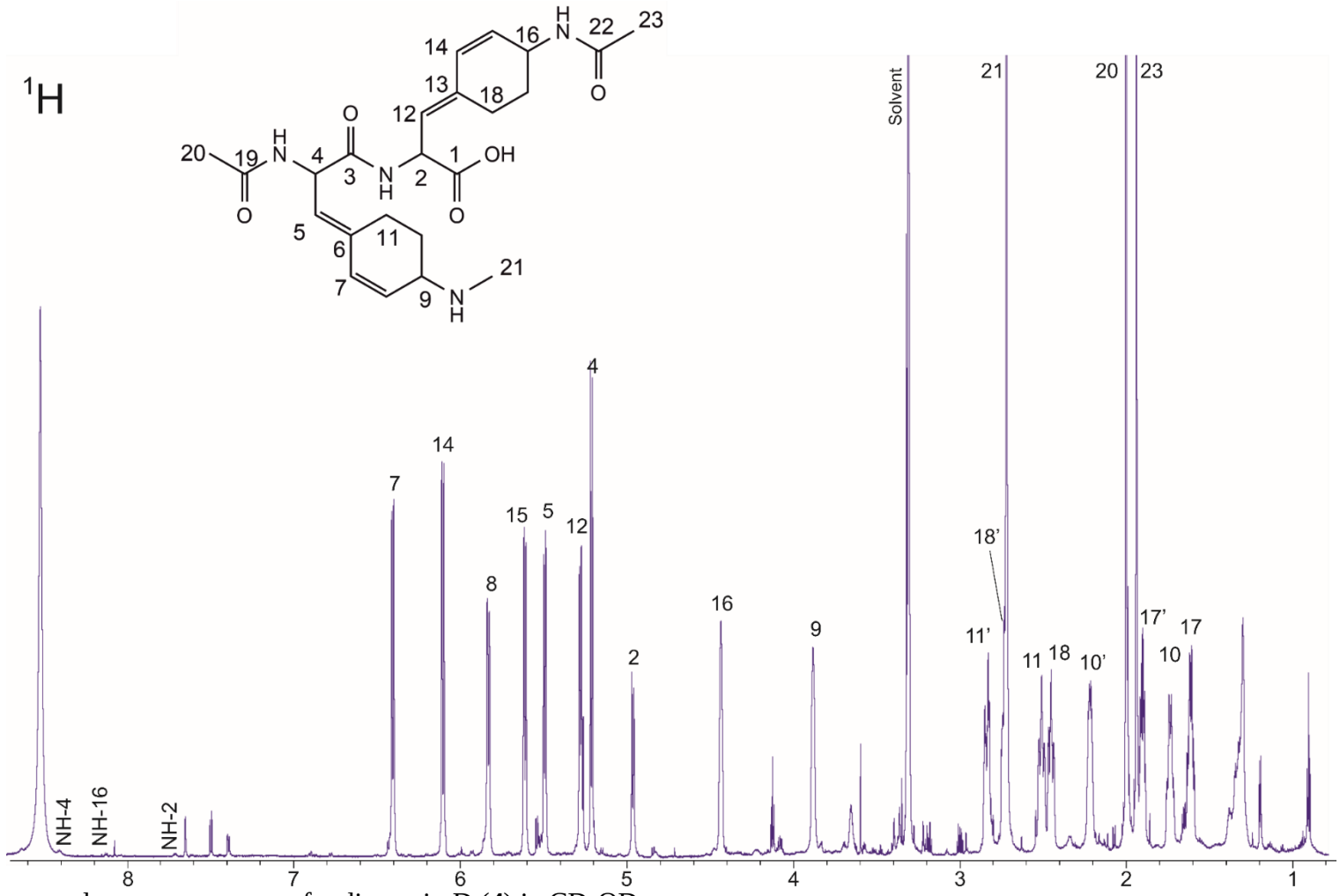


Figure S8 Annotated proton spectrum of radiosumin D (**4**) in CD_3OD.

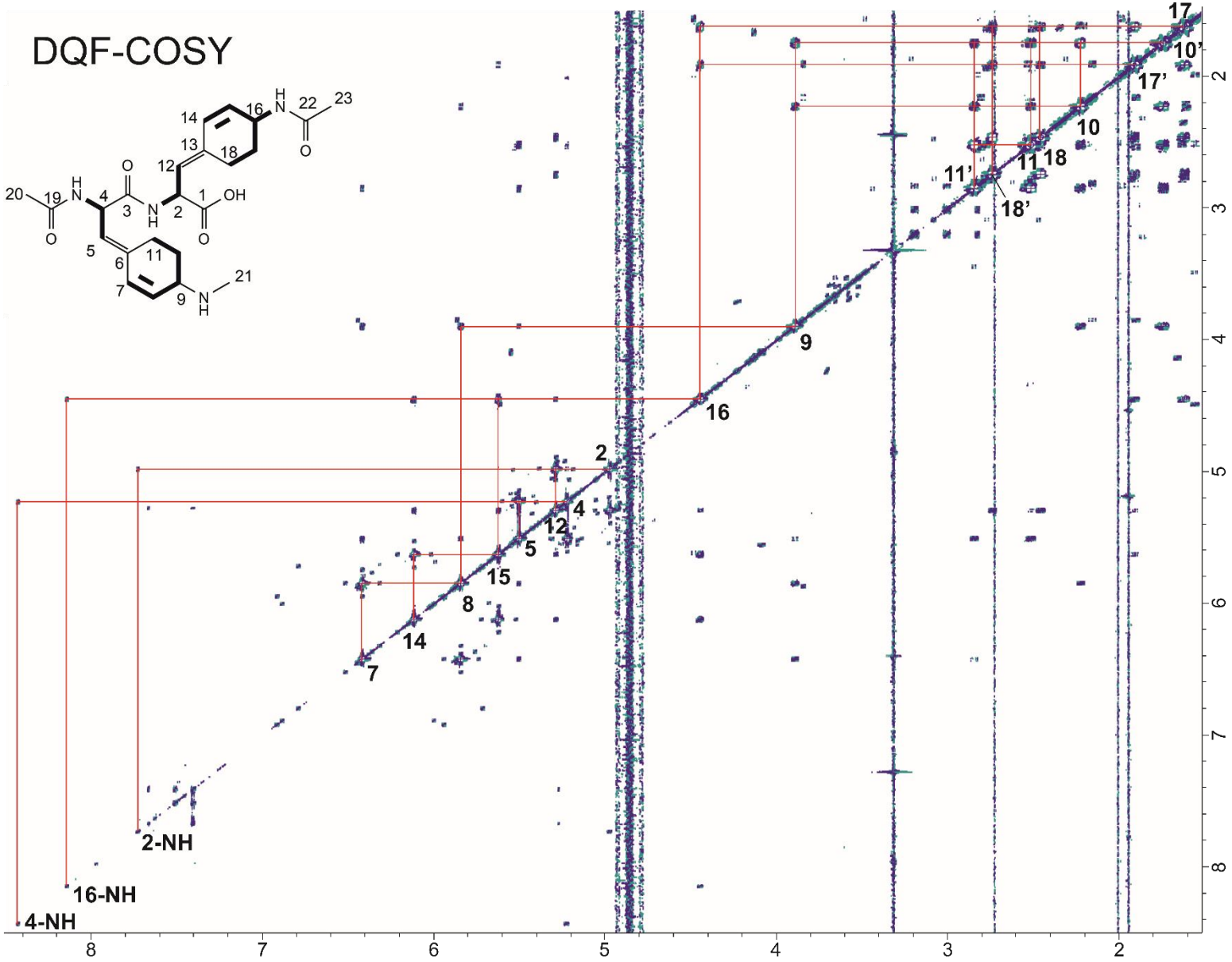


Figure S9 DQF-COSY spectrum of radiosumin D (**4**) in CD_3OD.

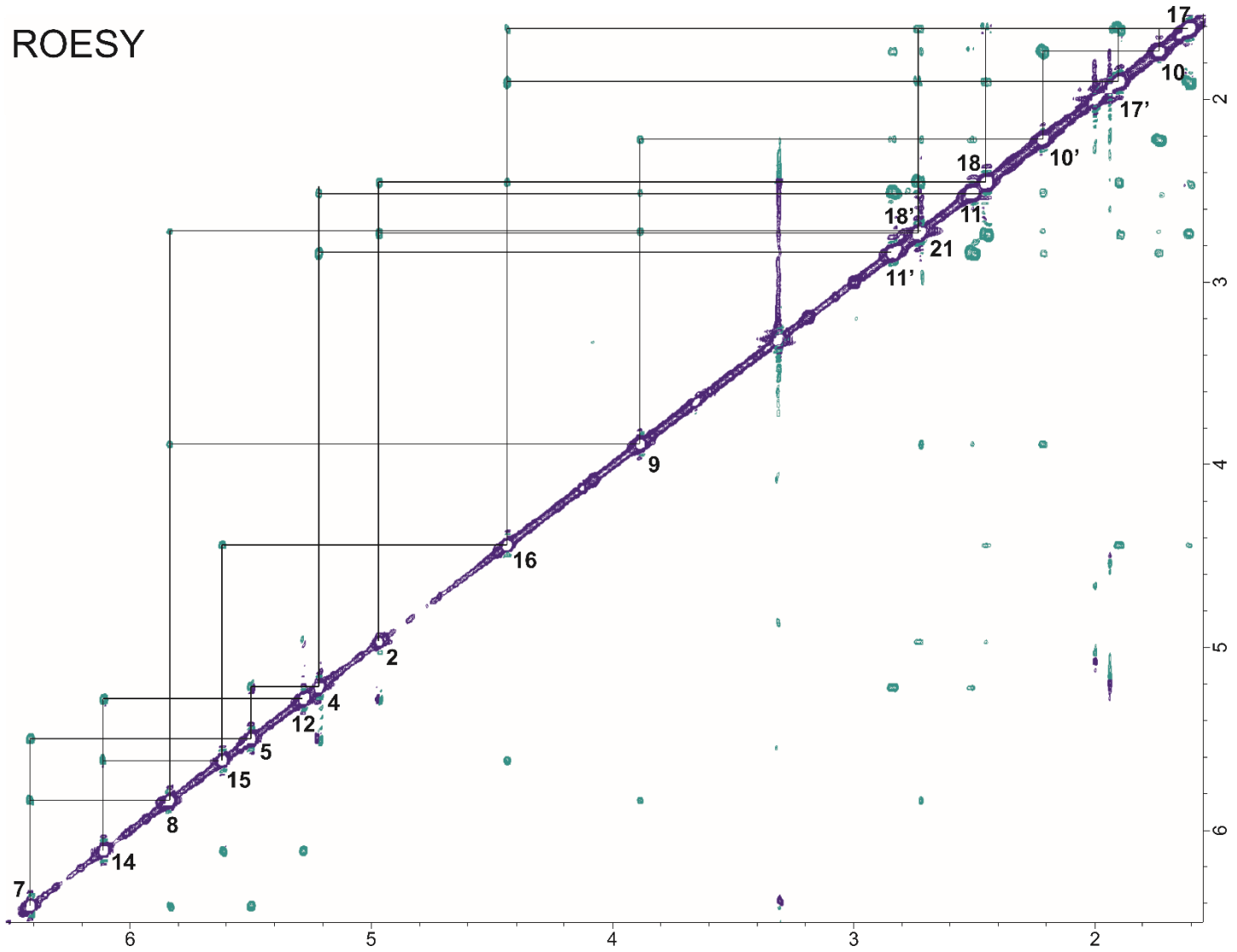


Figure S10 ROESY spectrum of radiosumin D (**4**) in CD_3OD.

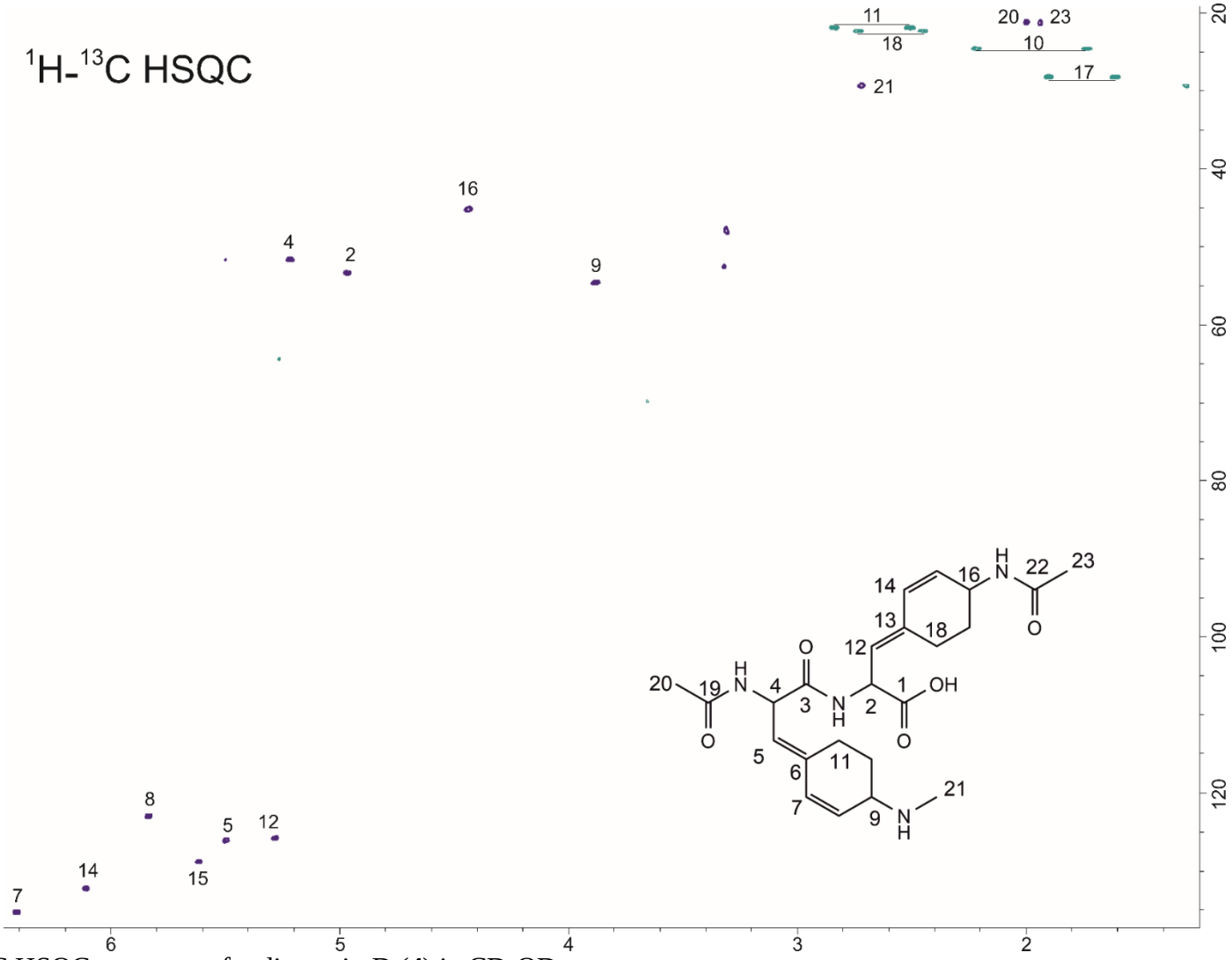


Figure S11 ^{1}H-^{13}C HSQC spectrum of radiosumin D (**4**) in CD_3OD.

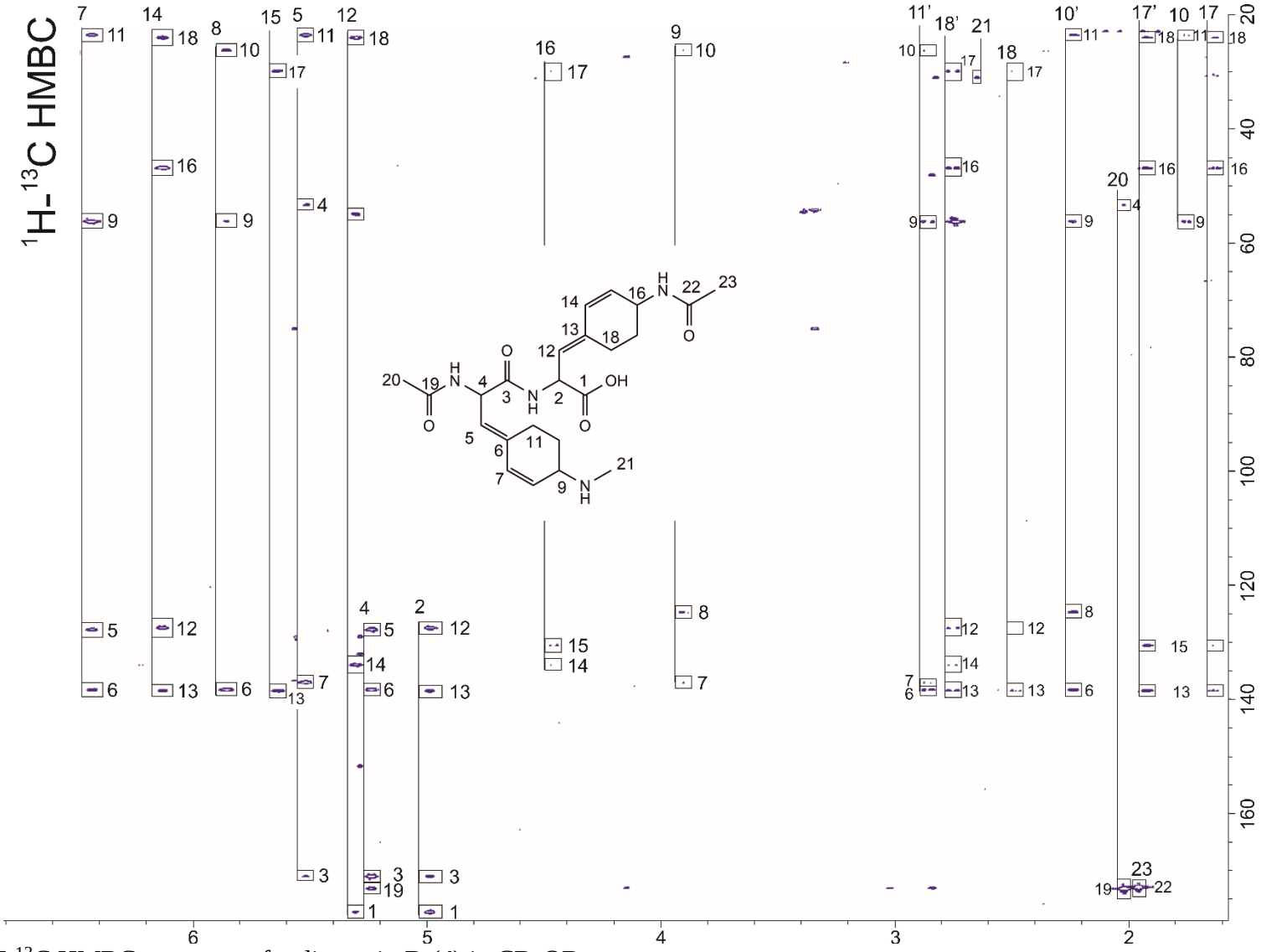


Figure S12 ^{1}H-^{13}C HMBC spectrum of radiosumin D (**4**) in CD_3OD.

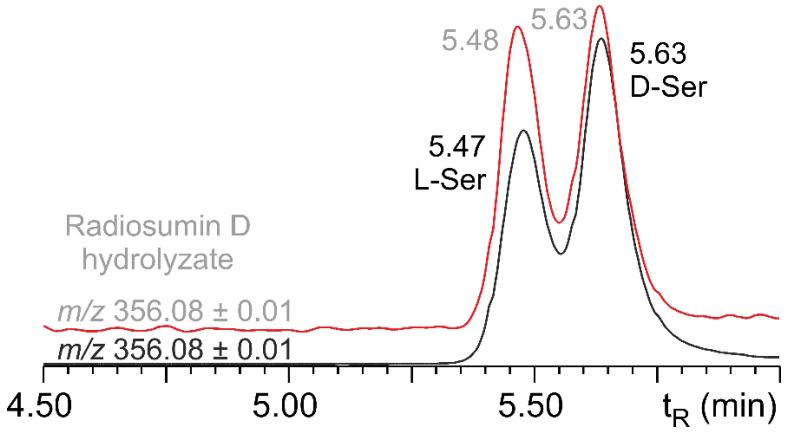
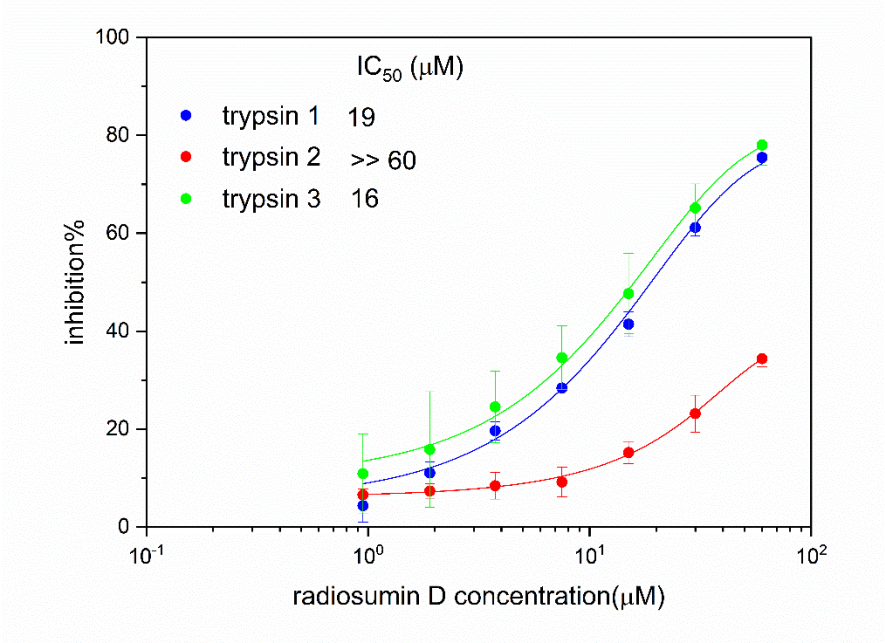


Figure S13 Extracted negative ion (*m/z* 356.08; $[M-H]^-$) chromatograms of Marfey derivative (1-fluoro-2,4-dinitrophenyl-5-L-alanine) radiosumin D (**4**) hydrolysate and L- and D-Ser reference amino acid Marfey derivatives.

A



B

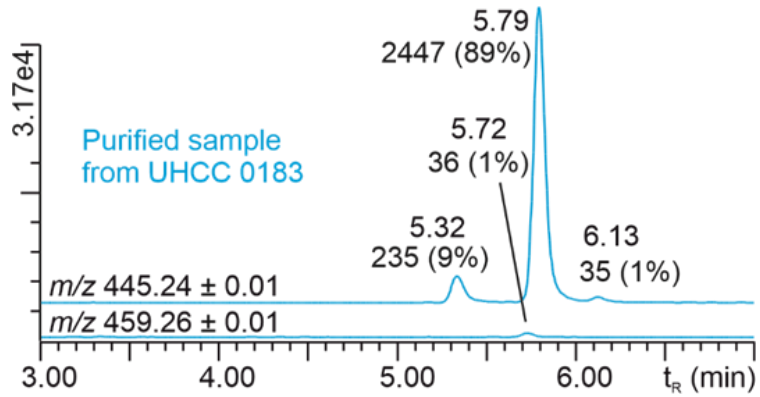


Figure S14 (A) Inhibition of human trypsin-isoenzymes by purified radiosumin D (**4**) sample (B) UPLC-QTOF spectrometry analysis of purified sample from *Aphanizomenon* sp. UHCC 0183.

Table S1. Stachelhaus 10-residue specificity code for RadH adenylation domains predicted by NRPS predicitor [2] as implemented in antiSMASH 6.0. [1]

Strains	Accession number (genome)	Accession number (protein)	Module1	Module2
Dolichospermum planctonicum UHCC 0167	VILE02	WP_265915766	DAEMAGGVLK	DAETSGGVLK
Aphanizomenon sp. UHCC 0183	VILC01	WP_168465319	----------	----------
Dolichospermum flos-aquae UHCC 0037	VILF01	WP_168636709	----------	----------
Aphanizomenon flos-aquae MDT14a	LJOX01	OBQ31167	----------	----------
Aphanizomenon flos-aquae 2012 KM1 D3	JSDP01	QSV71108	----------	----------
Anabaena sp. AL93	LJOU01	OBQ20459	----------	------M----------
Aphanizomenon flos-aquae LD13	LJOY01	OBQ27227	----------	------M----------
Aphanizomenon flos-aquae UKL13 PB	LTEC02	MBO1043034	----------	------M----------
Anabaena sp. UBA12330	DQEB01	HCQ23020	----------	------M----------
Dolichospermum sp. UHCC 0315A	CP043056	WP_246863057 WP_148760648	----------	------M----------
Dolichospermum flos-aquae CCAP 1403 13F	CP051206	WP_168696234	----------	------M----------
Anabaena cylindrica PCC 7122	CP003659	WP_015215721	--------S---------	------M----------
Plectonema radiosum NIES-515	JAOWRF01	WP_015215720 WP_263746364 WP_263746366	--------S---------	--S--IA----------

Table S2. Strains and plasmids used in this study.

Strains	Description	Reference or source
E. coli BAP1	Host strain for expression BL21(DE3) Δ*prp*RBCD ::T7prom-*sfp*,T7prom-*prpE*	[3]
Plasmids	**Description**	
pET28b-ptetO-*GFP*v2	Tetracycline inducible expression plasmid, ColE1, Kan^R, addition of *GFP*	[4]
pET28b-ptetO-*radH-GFP*	Tetracycline inducible expression plasmid, ColE1, Kan^R, addition of *GFP*, habouring bimodular NRPS protein gene *radH*	This study
pET28b-ptetO-*radH-radABCDEFG-GFP*	Built using pET28b-ptetO-*radH-GFP* as the vector and *radABCDEFG* single piece nucleotide insert.	This study
pET28b-ptetO-*radH-radABCDEFGI-GFP*	*radI* inserted into the pET28b-ptetO-*radH-radABCDEFG-GFP* vector	This study

Table S3. List of oligonucleotide primers used for cloning and screening procedures.

Name	Sequence (5'-3')	Description
gib_ptetO_radH1_F	**TCAGTGATAGAGAAGAGGATCGACC**ATGCAGGGCAATTCTTCTTTG	Amplification of *radH* with 25 bp homology sequence
gib_ptetO_radH2_R	**CAGTTCTTCACCTTTGCTAACCATG**CACGTGTTATCCATAAACTCTATCTTTTAAG	Amplification of *radH* with 25 bp homology sequence
C-GFP_for_1	CATGGTTAGCAAAGGTGA	Amplification of pET-28b-ptetO backbone for subsequent cloning of *radH*
spec-ptet-R	GGTCGATCCTCTTCTCTATC	Amplification of pET-28b-ptetO backbone for subsequent cloning of *radH*
screen_0167radH2_F	TCAGGTTTCTGTAGGAGTTATTTC	Colony screening and sequencing primer
screen_GFP_R	TTACCGTTGGTCGCATCACC	Colony screening and sequencing primer
C-GFP_for_2	CATGGTTAGCAAAGGTGAAGAACTGT	Amplification of pET-28b-ptetO-*radH* backbone for subsequent cloning of *radABCDEFG*
spec_radH2-R	GTGTTATCCATAAACTCTATCTTTTAAGATTTCAACAT	Amplification of pET-28b-ptetO-*radH* backbone for subsequent cloning of *radABCDEFG*
gib_radH2_radA-F	**AAAAGATAGAGTTTATGGATAA**CACATGAAAACCCTGATTATTGACAAC	Amplification of *radABCDEFG* with 22 bp homology sequence
gib_GFP_radG-R	**CAGTTCTTCACCTTTGCTAACCATG**CACGTGTCAATGCCTGCCCAATAATC	Amplification of *radABCDEFG* with 25 bp homology sequence

radI-fwd	**AGGATTATGGCAGGCATTGACAC**ATGAACCCAACACTCACAAATAAG	Amplification of *radI* with 25 bp homology sequence
radI-rev	**TTCTTCACCTTTGCTAACCATGCAC**GTGCTATATGCGCCTGTTTTC	Amplification of *radI* with 25 bp homology sequence

Homology arms are marked in bold

Table S4. Root-mean-square deviation (RMSD) values of the ligand (red) and the protein Cα atoms (blue) for each trypsin-ligand pair.

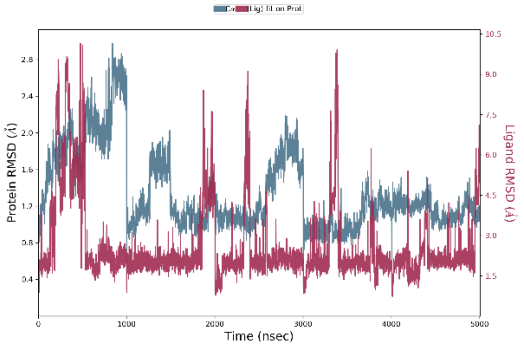
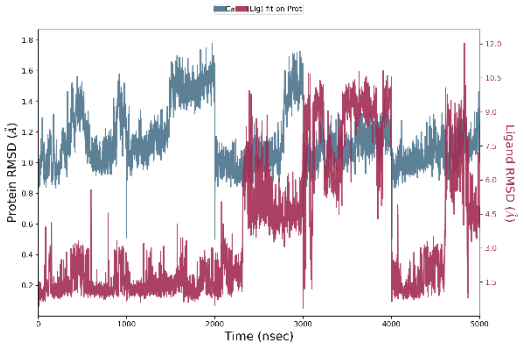
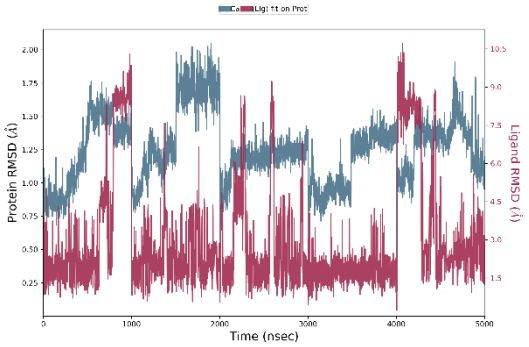
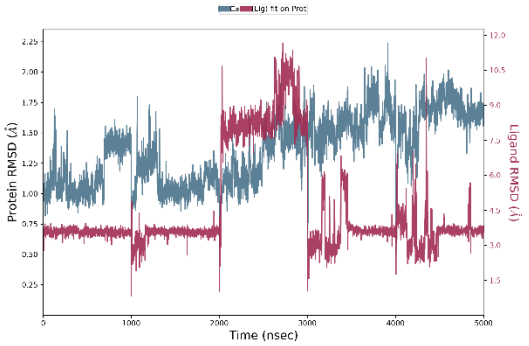
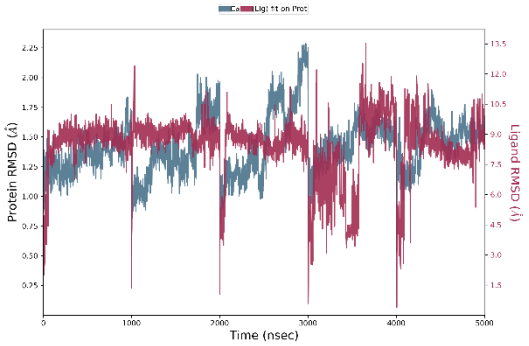
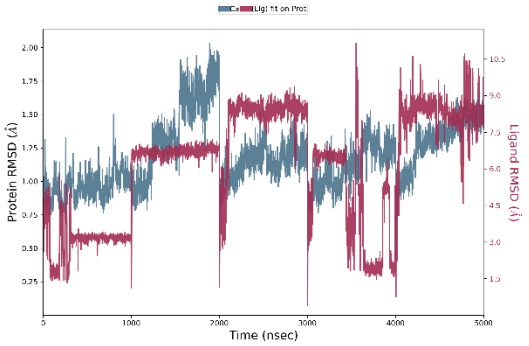
	Trypsin-1	**Trypsin-2**	**Trypsin-3**
Radiosumin-C (3)			
Radiosumin-D (4)			

Table S5. Ligand contact histograms of amino acids of trypsin-1, trypsin-2 and trypsin-3 that are in contact with radiosumin C (**3**) and radiosumin D (**4**) in MD simulation. Red = ionic interaction, green = H bond interaction, blue = water bridge interaction, purple = hydrophobic interaction.

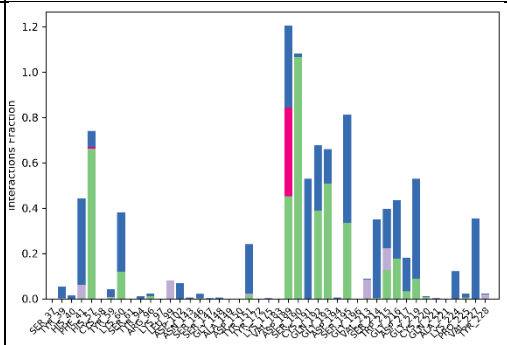
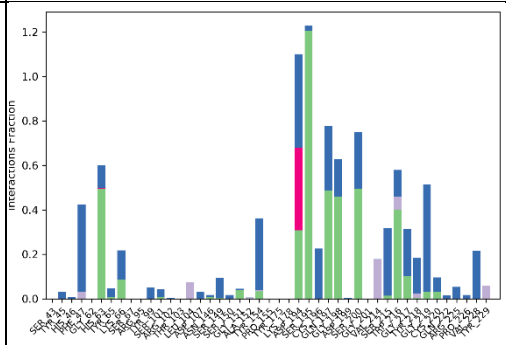
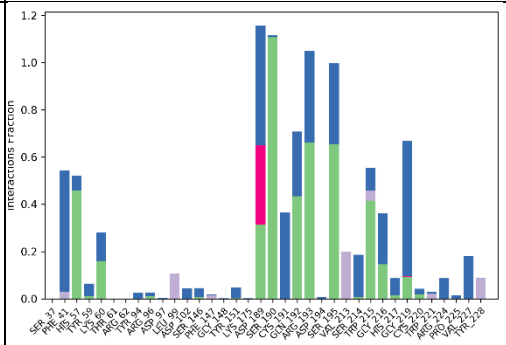
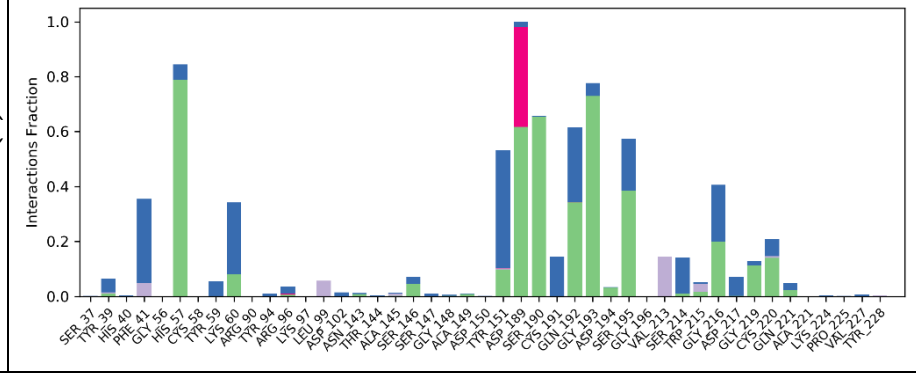
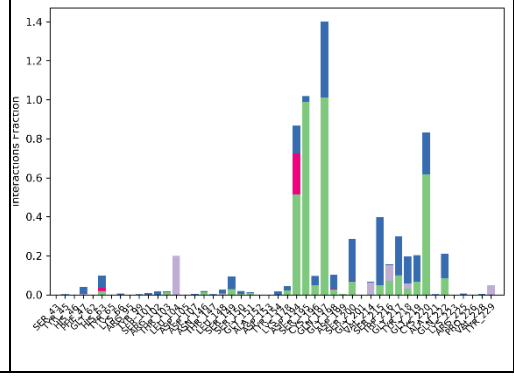
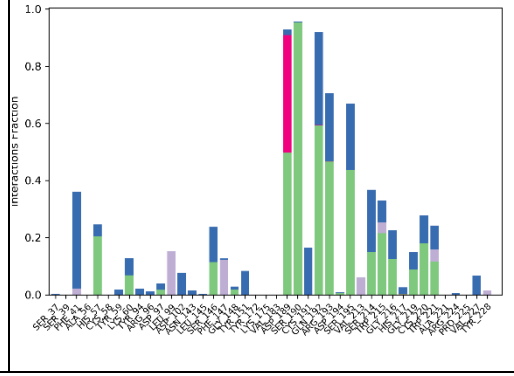
	Trypsin-1	**Trypsin-2**	**Trypsin-3**
Radiosumin-C (3)			
Radiosumin-D (4)			

Table S6. Experimental parameters used in NMR experiments

Experiment	SW F_1 [Hz]	SW F_2 [Hz]	$t_{1,max}$ [ms]	t_2 [ms]	τ_m [ms]	NS
^{1}H with pre-saturation	-	12820	-	51280	-	16
TOCSY	8002	8012	256	2048	90	8
DQF-COSY	8800	8800	1024	4096	-	8
ROESY	8800	8800	200	1024	200	8
^{13}C HSQC	36226	8012	256	1024	-	8
^{13}C HSQC-TOCSY	36226	8012	256	900	90	24
^{13}C HMBC	40256	8012	512	900	-	64
^{15}N HMBC	8112	8012	128	900	-	24

1. K. Blin, S. Shaw, A. M. Kloosterman, Z. Charlop-Powers, G. P. van Wezel, M. H. Medema and T. Weber, *Nucleic Acids Res.*, 2021, **49**, W29-W35.
2. M. Rottig, M. H. Medema, K. Blin, T. Weber, C. Rausch and O. Kohlbacher, *Nucleic Acids Res.* 2011, **39**, W362-367.
3. B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane and C. Khosla, *Sci.*, 2001, **291**, 1790-1792.
4. E. R. Duell, P. M. D'Agostino, N. Shapiro, T. Woyke, T. M. Fuchs and T. A. M. Gulder, *Microb Cell Fact.*, 2019, **18**, 1-11.