Variovoracins: new glycolipopeptide biosurfactants identified by surfactant screening and genome mining.

Bradley Haltli,[†,a,b,c,e,*] Douglas H. Marchbank,[†,a,b] Fabrice Berrué,[d] Hebelin Correa,[a] Nicholas G. McCarville,[a] Noelle Duncan,[a] Joshua Kelly,[a] Alyssa L. Grunwald,[a] Martin Lanteigne,[a] Katherine McQuillan,[a] and Russell G. Kerr,[*,a,b,c]

[a.] Croda Canada Limited, Regis and Joan Duffy Research Centre, 550 University Avenue, Charlottetown, PE, Canada, C1A 4P3

[b.] Department of Biomedical Sciences, University of Prince Edward Island, 550 University Avenue, Charlottetown, PE, Canada, C1A 4P3

[c.] Department of Chemistry, University of Prince Edward Island, 550 University Avenue, Charlottetown, PE, Canada, C1A 4P3

[d.] Aquatic and Crop Resource Development, National Research Council of Canada, Halifax, NS, Canada, B3H 3Z1

[e.] Department of Chemistry, Technical University of Munich, Lichtenbergstraße 4, 85748 Garching, Germany

[†] These two authors contributed equally to this work.

[*] Corresponding author

SUPPORTING INFORMATION

Table of Contents

General Methods

All UHPLC-HRMS analyses were carried out using a Thermo Accela chromatograph with HRMS-ELSD-UV detection~~.~~, which included a Thermo LTQ Exactive equipped with an electrospray ionisation (ESI) source, Sedex 80 LT-evaporative light-scattering detector (ELSD), and a Thermo photodiode array (PDA) detector. Unless otherwise indicated, all UHPLC-HRMS analyses was performed using the following chromatographic conditions: column, Phenomenex Kinetex Core-Shell 100 Å C_{18} column (2.1 × 50 mm, 1.7 µm); mobile phase flow rate, 0.5 mL·min^{-1}; injection volume, 10 µL; linear gradient, H_2O:CH_3CN (95:5, 0.1% formic acid) at 0.2 min to 99% CH_3CN (0.1% formic acid) at 4.8 min, which was held until 8.0 min before returning to H_2O:CH_3CN (95:5, 0.1% formic acid) at 8.5 min and equilibrating for 1.5 min. The following HRMS parameters were used: positive ionization mode; mass resolution, 30,000; mass range, m/z 190 to 2,000; spray voltage, 2.0 kV; capillary temperature, 300°C; S-lens RF voltage, 60.0%, maximum injection time, 10 ms; 1 microscan. The system was controlled by Thermo Xcalibur software. Automated flash chromatography was performed on a Teledyne ISCO CombiFlash® Rf 200 system equipped with a UV detector using pre-packed RediSep® Rf flash columns and the solid sample loading technique to inject samples. Semi-preparative reversed-phase high-performance liquid chromatography (RP-HPLC) was performed on Thermo Accela chromatograph equipped with a Sedex 80 LT- ELSD and a Thermo PDA detector. Nuclear magnetic resonance (NMR) spectra were acquired on a 400 and 600 MHz Bruker Avance III NMR spectrometer equipped with a 5 mm BBO SmartProbe™ and 5 mm TCI CryoProbe™, respectively. All chemical shifts (δ) are reported in ppm and referenced to residual solvent signal of methanol-d_4 (^{1}H, 3.31 ppm; ^{13}C, 49.0 ppm). Coupling constants (*J*) are reported in Hz with the following abbreviations: doublet (d), doublet of doublets (dd), triplet (t), multiplet (m), apparent (app.). Infrared spectra were recorded by attenuated total reflectance with a Thermo Nicolet 6700 FT-IR spectrometer (Smart iTR™). Optical rotations were measured on a Rudolph Autopol® III Polarimeter at 589 nm using a 50 mm microcell. All deionized water was obtained from a~~n~~ Milli-Q® water purification system and had a resistivity of at least 18.2 MΩ·cm. All solvents used for chromatography were of HPLC grade or higher. All other solvents were purchased from commercial sources and used without further purification.

Bacteria Isolation

Strain RKNM0096 was isolated from a soil sample (10 g) collected from Battle Bluffs, northwest of Kamloops, British Columbia, Canada (50.731553, -120.566409) in July, 2011. The sample was placed in sterile bag using a sterile scoopula and maintained at room temperature (20-23 °C) until processing (1 week). RKNM0096 was isolated using a modification of a previously described method for the isolation of motile actinomycetes.[1] A portion of the soil sample (1 g) was ground using a sterilized mortar and pestle and the ground soil was passed through a sterile sieve with 2 mm x 2 mm openings in order to remove large particles. The soil was then spread over the surface of an empty sterile Petri dish and dried in a biosafety cabinet for 7 days at room temperature. The dried sediment (0.5 g) was placed in 25 mL of 10 mM potassium phosphate buffer (pH 7.0) and incubated at 30°C without shaking for 90 min. Without disturbing the settled sediment 8 mL of buffer was removed and centrifuged at room temperature for 20 min at 1,500*g* in a swing bucket rotor. The tube was carefully removed from the centrifuge and allowed to stand for 30 min at room temperature. Without disturbing the pellet, 1 mL of supernatant was removed and serially diluted tenfold in autoclaved tap water. A portion (10 µL) of the diluted sample was distributed on the surface of humic acid vitamin (HAV) agar that was distributed in the wells of 48 well tissue culture plates (VWR). HAV was prepared as previously described, with the modification that noble agar (12 g/L) was used as the solidifying agent and the medium was supplemented with cycloheximide

(50 µg mL^{-1}~~/mL~~), trimethoprim (20 µg~~/mL~~ mL^{-1}) and nalidixic acid (10 µg~~/mL~~ mL^{-1}).[2] The plate was incubated at 30°C and RKNM0096 emerged after 19 days. The strain was purified by serial subculturing on ISP2 agar (Difco).

Characterization of Variovoracins Isolated from *Variovorax paradoxus* RKNM0096

Variovoracin Rha"-Rha'-C$_{10,10,10}$ (1)
Purification of **1** was achieved by semi-preparative RP-HPLC [column, Phenomenex Gemini-NX 110 Å C$_{18}$ column (250 × 10 mm, 5.0 µm); mobile phase flow rate, 3.0 mL·min^{-1}; CH$_3$CN:H$_2$O (80:20, 0.1% formic acid), R$_t$ 22.6 min] and **1** was obtained as a colorless oil (33.6 mg). Rha"-Rha'-C$_{10,10,10}$ (**1**): $[\propto]_D^{25}$ = -45.98° (*c* 1.05 in CH$_3$OH); IR υ_{max} 3300, 2954, 2927, 2857, 1737, 1645, 1546, 1457, 1379, 1317, 1163, 1131, 1068, 1043, 987 cm^{-1}. ^{1}H and ^{13}C NMR spectroscopic data are shown in **Table S1**. All NMR spectra are shown in **Figures S2-S8**. HRMS (ESI) *m/z*: [M + H]$^+$ Calcd for C$_{51}$H$_{95}$N$_2$O$_{17}$ 1007.6625; Found 1007.6635. Assigned fragment ions generated by collision-induced dissociation (CID) were annotated according to fragmentations shown in **Figure S9** shown in **Table S2**. All MS^n spectra are shown in **Figures S10**.

Variovoracin Rha"-OAc$_3$"-Rha'-C$_{10,10,10}$ (2)
Purification of **2** was achieved by semi-preparative RP-HPLC [column, Waters SunFire 100 Å C$_{18}$ column (250 × 10 mm, 5.0 µm); mobile phase flow rate, 3.0 mL·min^{-1}; 100% CH$_3$CN (0.1% formic acid), R$_t$ 15.6 min] and **2** was obtained as a colorless oil (83.2 mg). IR υ_{max} 3312, 2954, 2927, 2857, 1734, 1647, 1539, 1457, 1377, 1318, 1257, 1165, 1131, 1064, 1043, 989 cm^{-1}. ^{1}H and ^{13}C NMR spectroscopic data are shown in **Table S3**. All NMR spectra are shown in **Figures S11-S15**. HRMS (ESI) *m/z*: [M + H]$^+$ Calcd for C$_{53}$H$_{97}$N$_2$O$_{18}$ 1049.6731; Found 1049.6734.

Variovoracins Rha"-Rha'-C$_{8,10,10}$ (3a), Rha"-Rha'-C$_{10,8,10}$ (3b), and Rha"-Rha'-C$_{10,10,8}$ (3c)
Purification of **3a-3c** was achieved by semi-preparative RP-HPLC [column, Phenomenex Gemini-NX 110 Å C$_{18}$ column (250 × 10 mm, 5.0 µm); mobile phase flow rate, 3.0 mL·min^{-1}; CH$_3$CN:H$_2$O (80:20, 0.1% formic acid), R$_t$ 13.3 min] and **3a-3c** was obtained as a colorless oil (10.9 mg). IR υ_{max} 3312, 2955, 2928, 2858, 1737, 1645, 1548, 1460, 1379, 1319, 1200, 1167, 1096, 1068, 1044, 988 cm^{-1}. ^{1}H NMR (CD$_3$OD, 600 MHz, **Figure S16**) δ 5.27 – 5.21 (1H, m), 5.23 – 5.17 (1H, m), 5.01 (1H, d, *J* = 1.5 Hz), 4.79 (1H, d, *J* = 1.7 Hz), 4.40 (1H, t, *J* = 5.6 Hz), 4.06 – 3.99 (1H, m), 4.03 – 3.96 (1H, m), 3.98 (1H, dd, *J* = 3.3, 1.5 Hz), 3.86 (1H, dd, *J* = 3.0, 1.7 Hz), 3.80 – 3.73 (1H, m), 3.79 – 3.75 (1H, m), 3.77 – 3.72 (2H, m), 3.75 – 3.70 (1H, m), 3.73 – 3.68 (1H, m), 3.52 – 3.46 (1H, m), 3.51 – 3.45 (2H, m), 3.40 (1H, app. t, *J* = 9.5 Hz), 2.67 – 2.60 (1H, m), 2.67 – 2.59 (2H, m), 2.57 (1H, dd, *J* = 14.4, 8.0 Hz), 2.54 – 2.48 (2H, m), 1.72 – 1.63 (1H, m), 1.69 – 1.59 (4H, m), 1.59 – 1.53 (2H, m), 1.46 – 1.37 (1H, m), 1.44 – 1.23 (26H, m), 1.38 – 1.30 (1H, m), 1.25 (6H, d, *J* = 6.2 Hz), 0.93 (3H, d, *J* = 6.7 Hz), 0.92 – 0.88 (12H, m). ^{13}C NMR (CD$_3$OD, 150 MHz, **Figure S17**) δ 172.7 (C), 172.3 (C), 172.2 (C), 171.7 (C), 104.0 (CH), 101.5 (CH), 79.4 (CH), 76.6 (CH), 74.1 (CH), 73.2 (CH), 73.0 (CH), 72.4 (CH), 72.23 (CH), 72.18 (CH), 72.16 (CH), 70.6 (CH), 70.0 (CH), 65.5 (CH$_2$), 63.1 (CH$_2$), 56.7 (CH), 51.0 (CH), 41.8 (CH$_2$), 41.6 (CH$_2$), 41.1 (CH$_2$), 40.2 (CH$_2$), 35.3 (CH$_2$), 34.93 (CH$_2$), 34.90 (CH$_2$), 34.86 (CH$_2$), 33.00 (CH$_2$), 32.96 (CH$_2$), 32.7 (CH$_2$), 30.7 (CH$_2$), 30.5 (CH$_2$), 30.44 (CH$_2$), 30.39 (CH$_2$), 30.3 (CH$_2$), 26.3 (CH$_2$), 26.2 (CH$_2$), 26.1 (CH$_2$), 25.9 (CH), 25.8 (CH$_2$), 25.7 (CH$_2$), 23.8 (CH$_3$), 23.73 (CH$_2$), 23.68 (CH$_2$), 23.6 (CH$_2$), 22.4 (CH$_3$), 18.05 (CH$_3$), 18.04 (CH$_3$), 14.47 (CH$_3$), 14.44 (CH$_3$), 14.40 (CH$_3$), 14.37 (CH$_3$). HRMS (ESI) *m/z*: [M + H]$^+$ Calcd for C$_{49}$H$_{91}$N$_2$O$_{17}$ 979.6310; Found 979.6307. All MS^n spectra with annotated fragment ions based on **Figure S9** are shown in **Figure S18**.

Variovoracins Rha"-OAc_3"-Rha'-$C_{8,10,10}$ (4a), Rha"-OAc_3"-Rha'-$C_{10,8,10}$ (4b), and Rha"-OAc_3"-Rha'-$C_{10,10,8}$ (4c)

Purification of **4a-4c** was achieved by semi-preparative RP-HPLC [column, Waters SunFire 100 Å C_{18} column (250 × 10 mm, 5.0 µm); mobile phase flow rate, 3.0 mL·min^{-1}; 100% CH_3CN (0.1% formic acid), R_t 11.1 min] and **4a-4c** was obtained as a colorless oil (16.6 mg). IR υ_{max} 3312, 2955, 2928, 2858, 1737, 1645, 1548, 1460, 1379, 1319, 1200, 1167, 1096, 1068, 1044, 988 cm^{-1}. ^{1}H NMR (CD_3OD, 600 MHz, **Figure S19**) δ 5.27 – 5.21 (1H, m), 5.23 – 5.18 (1H, m), 5.03 (1H, br s), 5.02 (1H, dd, *J* = 9.7, 3.3 Hz), 4.80 (1H, d, *J* = 1.7 Hz), 4.40 (1H, t, *J* = 5.6 Hz), 4.12 (1H, dd, J = 3.3, 1.7 Hz), 4.06 – 4.00 (1H, m), 4.03 – 3.97 (1H, m), 3.91 – 3.86 (1H, m), 3.90 – 3.85 (1H, m), 3.78 – 3.71 (2H, m), 3.76 – 3.72 (1H, m), 3.74 – 3.68 (1H, m), 3.60 (1H, app. t, *J* = 9.7 Hz), 3.54 (1H, app. t, *J* = 9.6 Hz), 3.50 (1H, dd, *J* = 11.1, 5.2 Hz), 3.47 (1H, dd, *J* = 11.1, 5.8 Hz), 2.67 – 2.61 (1H, m), 2.67 – 2.60 (2H, m), 2.57 (1H, dd, *J* = 14.4, 8.0 Hz), 2.54 – 2.48 (2H, m), 2.11 (3H, s), 1.72 – 1.64 (1H, m), 1.68 – 1.59 (4H, m), 1.59 – 1.53 (2H, m), 1.46 – 1.39 (1H, m), 1.44 – 1.23 (26H, m), 1.38 – 1.31 (1H, m), 1.28 (3H, d, *J* = 6.3 Hz), 1.26 (3H, d, *J* = 6.2 Hz), 0.93 (3H, d, *J* = 6.7 Hz), 0.92 – 0.88 (12H, m). ^{13}C NMR (CD_3OD, 150 MHz, **Figure S20**) δ 172.8 (C), 172.7 (C), 172.3 (C), 172.2 (C), 171.7 (C), 103.8 (CH), 101.4 (CH), 79.7 (CH), 76.6 (CH), 75.8 (CH), 73.06 (CH), 72.99 (CH), 72.4 (CH), 72.2 (CH), 71.3 (CH), 70.6 (CH), 70.1 (CH), 70.0 (CH), 65.5 (CH_2), 63.1 (CH_2), 56.7 (CH), 51.0 (CH), 41.8 (CH_2), 41.5 (CH_2), 41.0 (CH_2), 40.2 (CH_2), 35.2 (CH_2), 34.91 (CH_2), 34.88 (CH_2), 34.8 (CH_2), 32.96 (CH_2), 32.92 (CH_2), 32.7 (CH_2), 30.7 (CH_2), 30.43 (CH_2), 30.41 (CH_2), 30.34 (CH_2), 30.30 (CH_2), 26.3 (CH_2), 26.2 (CH_2), 26.0 (CH_2), 25.9 (CH), 25.8 (CH_2), 25.7 (CH_2), 23.8 (CH_3), 23.69 (CH_2), 23.68 (CH_2), 23.64 (CH_2), 23.6 (CH_2), 22.4 (CH_3), 21.1 (CH_3), 18.0 (CH_3), 14.45 (CH_3), 14.43 (CH_3), 14.39 (CH_3), 14.36 (CH_3). HRMS (ESI) *m/z*: $[M + H]^+$ Calcd for $C_{51}H_{93}N_2O_{18}$ 1021.6418; Found 1021.6415. All MS^n spectra with annotated fragment ions based on **Figure S9** are shown in **Figure S23**.

Variovoracins Rha"-OAc_3"-Rha'-$C_{12,10,10}$ (5a), Rha"-OAc_3"-Rha'-$C_{10,12,10}$ (5b), and Rha"-OAc_3"-Rha'-$C_{10,10,12}$ (5c)

Purification of **5a-5c** was achieved by semi-preparative RP-HPLC [column, Waters SunFire 100 Å C_{18} column (250 × 10 mm, 5.0 µm); mobile phase flow rate, 3.0 mL·min^{-1}; 100% CH_3CN (0.1% formic acid), R_t 20.8 min] and **5a-5c** was obtained as a colorless oil (4.7 mg). ^{1}H NMR (CD_3OD, 600 MHz, **Figure S21**) δ 5.28 – 5.21 (1H, m), 5.23 – 5.18 (1H, m), 5.05 – 5.00 (1H, m), 5.03 (1H, d, *J* = 1.5 Hz), 4.80 (1H, d, *J* = 1.4 Hz), 4.40 (1H, t, *J* = 5.6 Hz), 4.13 (1H, dd, *J* = 3.2, 1.8 Hz), 4.06 – 4.00 (1H, m), 4.03 – 3.97 (1H, m), 3.91 – 3.85 (2H, m), 3.78 – 3.71 (2H, m), 3.77 – 3.71 (1H, m), 3.74 – 3.67 (1H, m), 3.61 (1H, app. t, *J* = 9.7 Hz), 3.54 (1H, app.t, *J* = 9.7 Hz), 3.50 (1H, dd, *J* = 11.1, 5.3 Hz), 3.47 (1H, dd, *J* = 11.1, 5.8 Hz), 2.68 – 2.60 (1H, m), 2.67 – 2.59 (2H, m), 2.57 (1H, dd, *J* = 14.3, 7.9 Hz), 2.55 – 2.48 (1H, m), 2.54 – 2.48 (1H, m), 2.12 (3H, s), 1.71 – 1.64 (1H, m), 1.68 – 1.59 (4H, m), 1.59 – 1.53 (2H, m), 1.45 – 1.38 (1H, m), 1.43 – 1.23 (34H, m), 1.37 – 1.31 (1H, m), 1.28 (3H, d, *J* = 6.5 Hz), 1.26 (3H, d, *J* = 6.1 Hz), 0.93 (3H, d, *J* = 6.7 Hz), 0.92 – 0.88 (12H, m). ^{13}C NMR (CD_3OD, 150 MHz, **Figure S22**) δ 172.9 (C), 172.7 (C), 172.3 (C), 172.15 (C), 171.7 (C), 103.7 (CH), 101.4 (CH), 79.6 (CH), 76.6 (CH), 75.8 (CH), 73.05 (CH), 72.99 (CH), 72.4 (CH), 72.1 (CH), 71.3 (CH), 70.6 (CH), 70.1 (CH), 69.9 (CH), 65.5 (CH_2), 63.1 (CH_2), 56.7 (CH), 51.0 (CH), 41.8 (CH_2), 41.5 (CH_2), 41.0 (CH_2), 40.2 (CH_2), 35.2 (CH_2), 34.93 (CH_2), 34.88 (CH_2), 33.03 (CH_2), 33.02 (CH_2), 33.00 (CH_2), 32.95 (CH_2), 32.92 (CH_2), 30.72 (CH_2), 30.68 (CH_2), 30.6 (CH_2), 30.42 (CH_2), 30.40 (CH_2), 30.34 (CH_2), 30.30 (CH_2), 26.3 (CH_2), 26.18 (CH_2), 26.15 (CH_2), 26.0 (CH_2), 25.8 (CH), 23.8 (CH_3), 23.7 (CH_2), 22.4 (CH_3), 21.1 (CH_3), 18.06 (CH_3), 18.03 (CH_3), 14.46 (CH_3), 14.44 (CH_3). HRMS (ESI) *m/z*: $[M + H]^+$ Calcd for $C_{55}H_{101}N_2O_{18}$ 1077.7044; Found 1077.7046. All MS^n spectra with annotated fragment ions based on **Figure S9** are shown in **Figure S24**.

Characterization of Variovoracins Isolated from *Janthinobacterium agaricidamnosum* DSM 9628

Variovoracin Rha'-$C_{10,10,10}$ (6)

Purification of **6** was achieved by semi-preparative RP-HPLC [column, Phenomenex Gemini-NX 110 Å C_{18} column (250 × 10 mm, 5.0 μm); mobile phase flow rate, 3.0 mL·min^{-1}; CH_3CN:H_2O (92:8, 0.1% formic acid), R_t 27.7 min] and **6** was obtained as a colorless oil (17.1 mg). Rha'-$C_{10,10,10}$ (**6**): $[\propto]_D^{25}$ = -35.44° (*c* 0.283 in CH_3OH); IR υ_{max} 3298, 2954, 2926, 2856, 1737, 1643, 1547, 1465, 1379, 1317, 1167, 1067, 1044, 984 cm^{-1}. ^{1}H and ^{13}C NMR spectroscopic data are shown in **Table S6**. All NMR spectra are shown in **Figures S32-S36**. HRMS (ESI) *m/z*: $[M + H]^+$ Calcd for $C_{45}H_{85}N_2O_{13}$ 861.6046; Found 861.6072. Fragment ions generated by collision-induced dissociation (CID) were annotated according to fragment ions based on **Figure S9** and are shown in **Figure S37**.

Variovoracins Rha'-$C_{8,10,10}$ (7a), Rha"-Rha'-$C_{10,8,10}$ (7b), and Rha"-Rha'-$C_{10,10,8}$ (7c)

Purification of **10a-10c** was achieved by semi-preparative RP-HPLC [column, Phenomenex Gemini-NX 110 Å C_{18} column (250 × 10 mm, 5.0 μm); mobile phase flow rate, 3.0 mL·min^{-1}; CH_3CN:H_2O (92:8, 0.1% formic acid), R_t 16.9 min] and **7a-7c** was obtained as a colorless oil (6.4 mg). IR υ_{max} 3300, 2955, 2928, 2858, 1738, 1644, 1548, 1467, 1379, 1317, 1262, 1172, 1129, 1068, 1048, 984 cm^{-1}. ^{1}H NMR (CD_3OD, 600 MHz, **Figure S38**) δ 5.28 – 5.23 (1H, m, H-3B), 5.24 – 5.18 (1H, m, H-3A), 4.81 (1H, d, *J* = 1.6 Hz, H-1'), 4.4 (1H, t, *J* = 5.7 Hz, H-2S), 4.10 – 4.04 (1H, m, H-3C), 4.03 – 3.97 (1H, m, H-2L), 3.76 (1H, dd, *J* = 3.3, 1.6 Hz, H-2'), 3.744 (1H, d, *J* = 5.7 Hz, H-3aS), 3.737 (1H, d, *J* = 5.7 Hz, H-3bS), 3.68 (1H, dq, *J* = 9.5, 6.2 Hz, H-5'), 3.65 (1H, dd, *J* = 9.5, 3.3 Hz, H-3'), 3.49 (1H, dd, *J* = 11.1, 5.2 Hz, H-1aL), 3.46 (1H, dd, *J* = 11.1, 5.7 Hz, H-1bL), 3.38 (1H, app. t, *J* = 9.5 Hz, H-4'), 2.67 – 2.62 (1H, m, H-2aA), 2.63 – 2.59 (1H, m, H-2bA), 2.63 – 2.55 (2H, m, H-2B), 2.62 – 2.55 (1H, m, H-2aC), 2.53 – 2.48 (1H, m, H-2bC), 1.72 – 1.64 (1H, m, H-4L), 1.68 – 1.58 (2H, m, H-4A), 1.65 – 1.58 (2H, m, H-4B), 1.59 – 1.53 (2H, m, H-4C), 1.45 – 1.39 (1H, m, H-3aL), 1.39 – 1.33 (2H, m, H-5B), 1.39 – 1.33 (2H, m, H-5C), 1.37 – 1.32 (1H, m, H-3bL), 1.37 – 1.30 (2H, m, H-5A), 1.26 (3H, d, *J* = 6.2 Hz, H-6'), 0.93 (3H, d, *J* = 6.6 Hz, H-5L), 0.92 – 0.90 (3H, m, H-6L). ^{13}C NMR (CD_3OD, 150 MHz, **Figure S39**) δ 172.7 (C, C-1B), 172.4 (C, C-1C), 172.2 (C, C-1S), 171.7 (C, C-1A), 100.7 (CH, C-1'), 75.51/75.46/75.42 (CH, C-3C), 73.9 (CH, C-4'), 73.0 (CH, C-3B), 72.6 (CH, C-2'), 72.3 (CH, C-3'), 72.1 (CH, C-3A), 70.3 (CH, C-5'), 65.5 (CH_2, C-1L), 63.1 (CH_2, C-3S), 56.8 (CH, C-2S), 51.0 (CH, C-2L), 41.9 (CH_2, C-2B), 41.3 (CH_2, C-2C), 41.1 (CH_2, C-3L), 40.2 (CH_2, C-2A), 35.34 (CH_2, C-4B), 35.29 (CH_2), 34.86 (CH_2, C-4A), 34.83 (CH_2), 34.34 (CH_2), 34.29 (CH_2, C-4C), 33.01 (CH_2), 32.96 (CH_2), 32.73 (CH_2), 32.71 (CH_2), 30.7 (CH_2), 30.49 (CH_2), 30.45 (CH_2), 30.4 (CH_2), 30.3 (CH_2), 26.4 (CH_2, C-5B), 26.3 (CH_2, C-5A), 26.02 (CH_2), 25.96 (CH_2), 25.85 (CH_2), 25.82 (CH, C-4L), 25.5 (CH_2, C-5C), 23.8 (CH_3, C-5L), 23.6 (CH_2), 22.4 (CH_3, C-6L), 18.0 (CH_3, C-6'), 14.47 (CH_3), 14.45 (CH_3), 14.41 (CH_3), 14.39 (CH_3), 14.36 (CH_3). HRMS (ESI) *m/z*: $[M + H]^+$ Calcd for $C_{43}H_{81}N_2O_{13}$ 833.5733; Found 833.5757. All MS^n spectra with annotated fragment ions based on **Figure S9** are shown in **Figure S40**.

Marfey's analysis

The absolute configurations of the leucinol and serine moieties of **1-7** were determined as follows. A crude extract from *V. paradoxus* RKNM0096 was dissolved in THF at a concentration of 10 mg/ mL^{-1}. An aliquot (2.5 mL) was removed and diluted with H_2O (3.5 mL) before adding 2 M $HCl_{(aq)}$ (4.0 mL). The reaction was stirred for 40 h at 60°C before adding sat. $NaHCO_{3(aq)}$ to neutralize the reaction. The mixture was then partitioned with ethyl acetate (10 mL). The organic phase was recovered and evaporated in vacuo to provide **8**, which was purified by automated flash chromatography using the following chromatographic conditions: column, Silicycle 12 g 60 Å C_{18} column (25 x 117 mm, 40-63 µm); mobile phase flow rate, 30 mL·min^{-1}; linear gradient, H_2O:CH_3OH (90:10%) at 0.0 min to 100% CH_3OH at 12 min, which was held until 15 min. Fractions containing **8** were combined and evaporated *in vacuo*. This procedure was repeated for the crude extract from *J. agaricidamnosum* DSM 9628. To complete the Marfey's derivatization, a portion of each reaction product (1.0 mg) was dissolved in H_2O (0.5 mL) and 2 M $HCl_{(aq)}$ (0.5 mL) was added. The reaction vessels were sealed and heated to 95°C overnight. The reaction mixtures were then evaporated *in vacuo* to provide samples containing the free amino acids. These dried solids were dissolved in H_2O (0.5 mL) before adding N_α-(2,4-dinitro-5-fluorophenyl)-L-alaninamide (50 µL of 10 mg/ mL^{-1} stock dissolved in acetone) and 1 M $NaHCO_{3(aq)}$ (25 µL). These reactions were incubated for 45 min at 40°C before quenching with 1 M $HCl_{(aq)}$ (25 µL). Samples were prepared for UHPLC-HRMS analysis by dissolving in CH_3OH (1.0 mL) and removing insoluble precipitate by centrifugation. An aliquot (0.1 mL) of the supernatants was removed and diluted with CH_3OH (0.9 mL). The above Marfey's derivatization were repeated for analytical standards L-serine (1.0 mg, Cat. No.: A11179-14, Thermo Fisher Scientific) and D/L-serine (1.0 mg, Cat. No.: S4375-100G, Sigma-Aldrich). UHPLC-HRMS analyses of the Marfey's derivatives were performed using the following chromatographic conditions: column, Thermo Hypersil Gold 100 Å C_{18} (50 × 2.1 mm, 1.9 µm); mobile phase flow rate, 0.4 mL·min^{-1}; injection volume, 10 µL; linear gradient from H_2O:CH_3CN (95:5, 0.1% formic acid) at 0.0 min to H_2O:CH_3CN (60:40, 0.1% formic acid) at 55.0 min, which was held until 57.0 min before another linear gradient to 100% CH_3CN (0.1% formic acid) at 60.0 min. The following HRMS parameters were used: positive ionization mode; mass resolution, 30,000; mass range, m/z 190 to 2,000; spray voltage, 2.0 kV; capillary temperature, 300°C; S-lens RF voltage, 60.0%, maximum injection time, 10 ms; 1 microscan. Extracted ion chromatograms were generated using the calculated *m/z* of the $[M+H]^+$ adducts of both derivatives. For serine, HRMS (ESI) *m/z*: $[M + H]^+$ Calcd for $C_{12}H_{16}N_5O_8$ 358.0993; Found 358.0992. For leucinol, HRMS (ESI) *m/z*: $[M + H]^+$ Calcd for $C_{15}H_{24}N_5O_6$ 370.1721; Found 370.1721.

Chiral Derivatization of Rhamnopyranose

The absolute configurations of the rhamnose moieties of **1-7** were determined as follows. A crude extract from *V. paradoxus* RKNM0096 was dissolved in acetone at a concentration of 10 mg/mL mL^{-1}. A 0.25 mL aliquot of this solution was added to H_2O (2.125 mL) followed by the addition of 2 M $NaOH_{(aq)}$ (0.125 mL). The reaction was stirred for 40 h at room temperature before adding ethyl acetate to partition the mixture. The H_2O layer was recovered and 1 M $HCl_{(aq)}$ (0.25 mL) was added before evaporating it in vacuo to provide **9**. This procedure was repeated for the crude extract from *J. agaricidamnosum* DSM 9628 to provide **11**. These reaction products were dissolved in H_2O (0.5 mL) and 1 M $HCl_{(aq)}$ (0.5 mL) was added. The reaction vessels were sealed and heated to 95°C overnight. The vessels were allowed to cool to room temperature before adding 2 M $NaOH_{(aq)}$ (0.25 mL) to neutralize the reactions, which were subsequently evaporated in vacuo to provide samples containing free rhamnose sugars. Pyridine (0.5 mL) and L-cysteine methyl ester hydrochloride (0.5 mL of 5 mg/mL mL^{-1} stock dissolved in pyridine) were added. The reactions were agitated gently for 1 h at 60°C, after which *o*-tolyl isothiocyanate (0.5 mL of 5 mg/mL mL^{-1} stock dissolved in pyridine) was then added. The reactions were agitated gently once more

for 1 h at 60°C before evaporation *in vacuo*. Samples were prepared for UHPLC-HRMS analysis by dissolving in CH_3OH (1.5 mL) and removing insoluble precipitate by centrifugation. An aliquot (0.1 mL) of the supernatants was removed and diluted with CH_3OH (0.9 mL). The above rhamnose derivatizations with L-cysteine methyl ester hydrochloride and *o*-tolyl isothiocyanate were repeated for analytical standards L-rhamnose monohydrate (1.0 mg, Cat. No.: R3875-5G, Sigma-Aldrich) and D-rhamnose (1.0 mg, Cat. No.: MR02762, Carbosynth). UHPLC-HRMS analyses of the thiocarbamoyl-thiazolidine derivatives were performed using the following chromatographic conditions: column, Phenomenex Kinetex Core-Shell 100 Å C_{18} (2.1 × 50 mm, 1.7 µm); mobile phase flow rate, 0.4 mL·min^{-1}; injection volume, 10 µL; linear gradient, H_2O:CH_3CN (87:13, 0.1% formic acid) at 0.5 min to H_2O:CH_3CN (70:30, 0.1% formic acid) at 13.0 min, which was held until 15.0 min before another linear gradient to 100% CH_3CN (0.1% formic acid) at 17.0 min, which was held until 20.0 min before returning to H_2O:CH_3CN (87:13, 0.1% formic acid) at 20.1 min and equilibrating for 5.0 min. The following HRMS parameters were used: positive ionization mode; mass resolution, 30,000; mass range, m/z 190 to 2,000; spray voltage, 2.0 kV; capillary temperature, 300°C; S-lens RF voltage, 60.0%, maximum injection time, 10 ms; 1 microscan. Extracted ion chromatograms were generated using the calculated *m/z* of the $[M+H]^+$ adduct. HRMS (ESI) *m/z*: $[M + H]^+$ Calcd for $C_{18}H_{27}N_2O_6S_2$ 431.1304; Found 431.1298.

Chiral Derivatization of 3-Hydroxydecanoic Acid with Naproxen Chloride

The absolute configurations of the 3-hydroxydecanoate moieties of **1-7** were determined as follows. A crude extract (10.0 mg) from *V. paradoxus* RKNM0096 was dissolved in CH_3CN (1 mL) before adding 2 M $HCl_{(aq)}$ (1 mL). After incubating overnight at 95°C, the reaction was cooled and partitioned with hexanes (2 mL). The hexanes layer was removed and the H_2O layer was partitioned once more with hexanes (1 mL). The hexanes layers were combined, dried with $MgSO_4$, and filtered. The filtrate was evaporated *in vacuo* to provide the released 3-hydroxydecanoic acid (**8**). This acid hydrolysis product was dissolved in anhydrous CH_2Cl_2 (1.5 mL) before adding (*S*)-naproxen chloride (4.0 mg, 16.1 µmol). The reaction mixture was stirred for 1 h at room temperature before it was evaporated in vacuo and dissolved in CH_3OH at a concentration of 0.5 mg~~/mL~~ mL^{-1} in preparation for UHPLC-HRMS. This procedure was also completed for a crude extract (10.0 mg) from *J. agaricidamnosum* DSM 9628 and for UHPLC-HRMS analytical standards (±)-3-hydroxydecanoic acid (Cat. No.: H3648-5MG, Sigma-Aldrich) and (*R*)-3-hydroxydecanoic acid. The latter was obtained from (*R*)-3-hydroxydecanoic acid methyl ester (Cat. No.: FH24326, Carbosynth), which was dissolved in THF/H_2O (1:1, 2 mL) containing LiOH (0.182 mg, 7.59 µmol). The reaction mixture was stirred for 8 h at room temperature. The pH of the mixture was adjusted to 2-3 by adding 3 M $HCl_{(aq)}$ before extracting with ethyl acetate (2 mL). The ethyl acetate layer was removed and evaporated in vacuo to provide (*R*)-3-hydroxydecanoic acid as a standard for the above derivatization. UHPLC-HRMS analyses of the naproxen ester derivatives were performed using the following chromatographic conditions: column, Phenomenex Kinetex Core-Shell 100 Å C_{18} (2.1 × 50 mm, 1.7 µm); mobile phase flow rate, 0.5 mL·min^{-1}; injection volume, 10 µL; linear gradient, H_2O:CH_3CN (95:5, 0.1% formic acid) at 0.2 min to H_2O:CH_3CN (10:90, 0.1% formic acid) at 17 min, which was held until 19.0 min. The following HRMS parameters were used: positive ionization mode; mass resolution, 30,000; mass range, m/z 190 to 2,000; spray voltage, 2.0 kV; capillary temperature, 300°C; S-lens RF voltage, 60.0%, maximum injection time, 10 ms; 1 microscan. Extracted ion chromatograms were generated using the calculated *m/z* of the $[M+Na]^+$ adduct. HRMS (ESI) *m/z*: $[M + H]^+$ Calcd for $C_{24}H_{32}O_5Na$ 423.2142; Found 423.2139.

Purification of Rhamnolipid (RL) Standards

The mono- and di-RL standards were obtained by semi-preparative RP-HPLC separation [column, Phenomenex Gemini-NX 110 Å C_{18} column (250 × 10 mm, 5.0 µm); mobile phase flow rate, 3.0 mL·min^{-1}; CH_3CN:H_2O (58:42, 0.1% formic acid); R_t (di-RL): 19.1 min, R_t (mono-RL): 38.1 min] of a

commercially available 90% rhamnolipid mixture (AGAE Technologies, P/N: R90-10G). The samples were evaporated *in vacuo* to provide 6.9 and 5.2 mg of mono- and di-RL, respectively.

Determination of the Critical Micelle Concentration (CMC)

The CMC values were determined by surface tension measurements on a Kibron Delta-8 multichannel microtensiometer based on the du Noüy-Padday method.[3] Variovoracin samples and rhamnolipids A and B were prepared at 2.0 mM in deionized water. Samples were serially diluted in 96-well plates by diluting 180 µL of sample in 60 µL deionized water. This process was repeated across 24 wells for each sample over two 96 well plates to provide ¾ serial dilutions. The sample dilutions were prepared in duplicate and transferred to high throughput surface tensiometer 96-well plates (PP Dyneplate, Kibron). The plates were left at room temperature to equilibrate for 90 min prior to measurement. Surface tension measurements were performed using the du Noüy maximum pull force technique. Sample temperatures varied between 24-25 °C during these measurements. The CMC values were determined from the change in slope in plots of surface tension vs. surfactant concentration.

Determination of the Characteristic Curvature (Cc)

Equations below represent hydrophilic-lipophilic difference (HLD) equations for ionic/non-ionic surfactant mixtures and derivation of equation describing *F(S)* as a function of Cc_{Var} where x_{SDHS} is the molar fraction of SDHS, x_{Var} is the molar fraction of variovoracin **1**, k is a surfactant constant and equal to 0.17, S is the salinity, S^* is the optimal salinity when $HLD_{mix} = 0$, $EACN$ is the effective alkane carbon number ($EACN = 1$ for toluene), Cc_{SDHS} is the characteristic curvature of SDHS, and Cc_{Var} is the characteristic curvature of **1**.

General HLD equation for surfactant mixture:

$$HLD_{mix} = x_{SDHS} \times HLD_{SDHS} + x_{Var} \times HLD_{Var} \quad (1)$$

where the HLD_{SDHS} for ionic surfactants is:

$$HLD_{SDHS} = \ln S - k \times EACN + Cc_{SDHS} \quad (2)$$

and the HLD_{Var} for non-ionic surfactants is:

$$HLD_{Var} = b \times S - k \times EACN + Cc_{Var} \quad (3)$$

Therefore, HLD_{mix} can be written as:

$$HLD_{mix} = F(S) - k \times EACN + x_{SDHS} \times Cc_{SDHS} + x_{Var} \times \quad (4)$$

where $F(S)$ is:

$$F(S) = x_{SDHS} \times \ln S + x_{Var} \times b \times S \quad (5)$$

When $HLD_{mix} = 0$, $F(S^*)$ is :

$$F(S^*) = k \times EACN - x_{SDHS} \times Cc_{SDHS} - x_{Var} \times Cc_{Var} \quad (6)$$

Cytotoxicity Testing

The cytotoxicity of variovoracins **1**, **2**, and **6** were evaluated against adult human keratinocytes (Gibco C-005-5c) and normal human fibroblasts (ATCC CRL-2522) as described previously.[4] Fluorescence was measured using a Thermo Scientific Varioskan® Flash plate reader at 560/12 nm excitation, 590 nm emission both at time zero and 4 h after alamarBlue™ (ThermoFisher Scientific) addition. Growth inhibition was expressed as a percentage and plotted against the logarithm of concentration. The variable slope dose-response curves were fitted to these data using the variable slope model in GraphPad Prism 8.1.2.

Antimicrobial Testing

Variovoracins **1**, **2**, and **6** were tested for antimicrobial activity in triplicate according to the Clinical Laboratory Standards Institute testing standards using a 96-well plate microbroth dilution assay as previously described.[4] The compound was tested against methicillin-resistant *Staphylococcus aureus* ATCC 33591, *S. warneri* ATCC 17917, vancomycin-resistant *Enterococcus faecium* EF379, *Pseudomonas aeruginosa* ATCC 14210, *Proteus vulgaris* ATCC 12454, and *Candida albicans* ATCC 14035 at a maximum test concentration of 128 µg~~/mL~~ mL^{-1}. Optical density was recorded using a Thermo Scientific Varioskan® Flash plate reader at 600 nm, recording at time zero and then again after incubation for 22 h at 37 °C to determine percent growth inhibition relative to vehicle treated controls.

Genome Sequencing and Bioinformatic Analysis

The *V. paradoxus* RKNM0096 genome was sequenced at the McGill University and Genome Quebec Innovation Centre (Montreal, QC, CA) using 2 SMRT Cells in a PacBio RSII sequencer (Pacific Biosciences, Menlo Park, CA, USA). A total of 140, 476 raw subreads with an average length of 11,269 bp were generated and genome assembly was achieved using a HGAP workflow.[5] Raw subreads were generated from raw .bas.h5 PacBio data files. A subread length cutoff value (30X) was extracted from subreads and used in the preassembly (BLASR) step, which consists of aligning short subreads on long subreads.[6] Since errors in PacBio reads are random, the alignment of multiple short reads on longer reads enables correction of sequencing errors on long reads. These long corrected reads were then used as seeds in a subsequent assembly prepared using the Celera assembler, which generates contigs.[7] These contigs were then 'polished' by aligning raw reads on contigs (BLASR) which were then processed through a variant calling algorithm (Quiver) that generates high quality consensus sequences using local realignments and PacBio quality scores.[5] Over 161,717,463 bp of corrected long subreads were obtained and resulted in the assembly of two contigs. One contig contained 7,193,071 bp while the other contained 1,767 bp. BlastN[8] analysis of the 1.7 kb contig revealed that the contig was a 100% match to the PacBio DNA sequencing internal control sequence (MG495226.1) and was excluded from subsequent analyses. The ends of the 7.2 Mb contig were compared to the linear contig sequence using Geneious (v.10.0; Dotmatics) to identify internal regions of homology. This analysis determined that the ends of the linear sequence overlapped within the contiguous DNA sequence. To confirm that the genome was circular the region containing the predicted overlap was amplified using the primers NMcircCONF-F (5′-CCTCGCCTCGGGTCTCATG) and NMcircCONF-R (5′- CGAGTAGCCGACGACCTTG) which were complementary to nucleotides 7,168,659 – 7,168,677 and 7,169,660 – 7,169,678 of the linear contig sequence, respectively. PCR amplification yielded a product with an apparent molecular weight of approximately 1 kb, which was in good agreement with the predicted amplicon size of (1,019 bp). The PCR amplicon was sequenced by McGill University and Genome Quebec Innovation Centre (Montreal, QC, CA) using the amplification primers and the sequence was compared to the predicted sequence using Geneious and found to be a 100% match. This data confirmed that the chromosome was circular and 7,168,245 bp. The genome sequence~~s~~ was deposited in GenBank under accession number CP046508.1

and annotated using the NCBI Prokaryotic Genome Annotation Pipeline (v.6.7).[9-11] Annotation identified 6,706 total genes, encompassing 6,610 protein coding sequences, 29 pseudogenes, 58 tRNAs, and two complete rRNA (5S, 16S and 23S) operons.

Taxonomic identification was conducted by analyzing the 16S rRNA sequences extracted from the genome using the EZ BioCloud identification tool.[12] Biosynthetic gene cluster (BGC) analyses were conducted using antiSMASH (v.7.1.0) with detection strictness set to relaxed.[13] Non-ribosomal peptide synthetase (NRPS) condensation domain analysis was conducted using the NaPDoS program using default settings.[14] Evolutionary history of glycosyltransferases from the variovoracin BGC (RlpC and RlpE) and rhamnolipid (RhlB and RhlC) BGCs from *Burkholderia thailandensis, B. pseudomallei* and *Pseudomonas aeruginosa* were conducted using MEGA X.[15, 16] Parameters used for the analysis are detailed in the legend of **Figure S29**. To identify BGCs related to the variovoracin locus the RlpC sequence was used to search the GeneBank database for related sequences using the BlastP program.[17, 18] The genomic context of hits identified in the BlastP search were inspected manually to determine if they were in proximity to a NRPS. In cases where the RlpC homolog was co-located with an NRPS the BGC was extracted from the genome sequence. BGCs representing the taxonomic distribution of variovoracin-like pathways (~~23~~24) were aligned using cblaster and the BGC alignment was graphed using clinker using default settings.[19, 20] The NRPS module composition and domain substrate specificity were predicted by antiSMASH (v.7.1.0).[13]

RlpE expression and Characterization

The *rlpE* gene was amplified from *V. paradoxus* RKNM0096 genomic DNA using the primers rlpE_F_NdeI (5′- TTTCATATGTCCTTCCCGTTCG) and rlpE_F-NcoI (5′- AACTCGAGTCAATCGC TGCGGCCCAG). Amplification was performed using KOD Hot Start Master Mix (Millipore Sigma) at a 1X concentration, 5% molecular grade DMSO (5% v/v), 0.4 µM of each primer and 50 ng of genomic DNA. Thermal cycling parameters were 95 °C for 2 min followed by 35 cycles of 95 °C for 30 s, 59 °C for 30s and 72 °C for 1 min, followed by a final extension step at 72 °C for 5 min. Amplification was assessed by agarose gel electrophoresis, confirming that an amplicon with an size of approximately 1 kb was amplified, which was consistent with the expected size of 983 bp. The amplicon was purified from the reaction mixture using the GeneJet PCR Purification Kit (ThermoFisher). The purified amplicon and pET28a (Novagen) were each cut with *Nde*I and *Nco*I (New England Biolabs) and purified using the GeneJET PCR Purification Kit. The amplicon was ligated to pET28a using the NEB Quick Ligation Kit (New England Biolab). Ligations were used to transform *E. coli* NEB5α cells and transformants were selected on LB agar (EMD Millipore) containing 50 µg/mL kanamycin (Sigma) and incubated at 37 °C overnight. Colonies were screened by colony PCR using the same amplification conditions described above. Template DNA was prepared by suspending a portion of a colony in 25 µL of DMSO and 2.5 µ~~u~~L was added to each PCR reaction (50 µL). Plasmid DNA was prepared from three positive colonies identified by colony PCR using the QIAprep Spin Miniprep Kit (Qiagen) and the sequence of the cloned rlpE gene in pET28a with an N-terminal hexa-histidine tag was confirm by sequencing with the T7 promoter (5′ - TAATACGACTCACTATAGGG) and T7 terminator (5′ - GCTAGTTATTGCTCAGCGG) primers that flank the multicloning site in pET28a. Sequencing was performed by Eurofins Genomics to confirm error-free cloning. The sequence of the cloned gene was deposited in GenBank under accession number PX521120.

Due to the high GC content of *rlpE*, *E. coli* Rossetta DE3 pLysS (Novagen) was chosen as the expression host as this strain expresses tRNAs for rare GC-rich codons (AGG, CCA, GGA). A single colony was used to inoculate 50 mL of LB Miller (EMD Millipore) supplemented with 50 µg~~/~~ mL^{-1}~~mL~~ of kanamycin

(Sigma-Aldrich) and 34 µg~~/ mL~~ mL^{-1} of chloramphenicol (Sigma-Aldrich) and the flask was incubated at 37°C with shaking at 250 rpm overnight. Expression cultures (50 mL) were performed in LB Miller (EMD Millipore) supplemented with kanamycin and chloramphenicol. These cultures were inoculated with 0.5 mL of the overnight culture and cultured at 37°C and 250 rpm until the optical density (600 nm) reached 0.5. Isopropyl-beta-D-thiogalactopyranoside (IPTG; Millipore) was then added to a final concentration of 1.0 mM to induce protein expression and the cultures were incubated at 15-°C for 24 h. Cells were harvested by centrifugation (6,000 x *g* for 5 min) and washed once with 20 mM Tris-HCl (pH 8.0). The cell pellet was frozen at -80°C until purification could be performed. To purify His-tagged RlpE, the cells were thawed, suspended in lysis buffer (500 mM NaCl, 5% glycerol, 1% Triton X-100, 25 mM Tris-HCl, pH 8.0) and then lysed via sonication. Cell debris and insoluble protein ~~was~~were removed by centrifugation at 15,000~~ x~~ *g* for 30 min. The supernatant was mixed with 0.5 mL of HisPur Ni-NTA resin (ThermoFisher). The resin was washed six times with 1.0 mL of 75 mM imidazole (Sigma-Aldrich). His-tagged RlpE was eluted with 1.0 mL of 250 mM imidazole. Four batch elutions were performed and pooled. The imidazole elution buffer was exchanged with enzyme buffer (25 mM Tris-HCl, 10% glycerol) and concentrated by centrifugal filtration using a Macrosep 3 kDa spin filter (Pall). Following concentration, the enzyme was aliquoted and stored at -80°C. The purity of the enzyme was analyzed by denaturing polyacrylamide gel electrophoresis (4-15% Mini-PROTEAN precast gel, 160 V, 30 min; Bio-Rad). The calculated molecular weight of His-tagged RlpE was 38.2 kDa. The apparent molecular weight of the purified protein was 33.0 kDa, which was in good agreement with the expected molecular weight of 38.3 kDa (**Figure S44**).

The activity of RlpE was established by incubating the enzyme (0.1 µM) in reaction buffer (25 mM Tris-HCl pH 8.0, 2.5 mM $MgCl_2$) with 1 mM of TDP-L-rhamnose (Carbosynth) and 0.5 mM **6**. Reactions (200 µL) were incubated at 30°C for 4 h. A portion (25 µL) of the reaction was removed at 15 s, 1 min, 5 min, 20 min, 1 h and 4 h and stopped by the addition of two volumes of methanol followed by flash freezing. Quenched reactions were separated by UPLC (Accela^{TM}, Thermo Fisher Scientific Mississauga, ON, Canada) and the eluates analyzed by HRESIMS (LTQ Orbitrap Velos; Thermo Fisher Scientific) (positive mode, monitoring m/z 200-2000). Chromatographic separation was achieved with a Hypersil Gold 1.9 µm C18 175 Å 50 x 2.1 mm column (Thermo Fisher Scientific) and a linear gradient from 50% H_2O/0.1% formic acid (FA) (solvent A) and 50% acetonitrile (CH_3CN)/0.1% FA (solvent B) to 100% solvent B over 5 min followed by a hold of 100% solvent B for 3 min with a flow rate of 300 µL mL^{-1}~~/min~~. Reactions conducted with boiled enzyme showed no conversion of **~~NB-RLP8606~~** to **~~NB-RLP10061~~**. In contrast, enzyme reactions containing intact 6xHis-RlpE resulted in the complete conversion of **~~NB-RLP8606~~** to **~~NB-RLP10061~~** after 4 h (**Figure S44**). To determine the affinity of RlpE for **6** steady state kinetics were determined using reactions consisting of 0.1 µM His-tagged RlpE, 25 mM Tris-HCl (pH 8.0, 2.5 mM $MgCl_2$, 10 mM TDP-L-rhamnose. The concentration of **6** was varied from 0.125 – 8 mM. Reactions were quenched with 2 volumes of~~methanol~~ CH_3OH and flash frozen after 30 s. Six replicates of each ~~rea~~were~~ction~~ was performed and analyzed by UPLC-HRESIMS ~~as described above. Peak~~and peak areas for the reaction product (**1**) were calculated using the Xcalibur software (Thermo Fisher Scientific) as described above. The peak areas were used to calculate initial reaction rates which were used to calculate the K_m using the Lineweaver Burke plot.

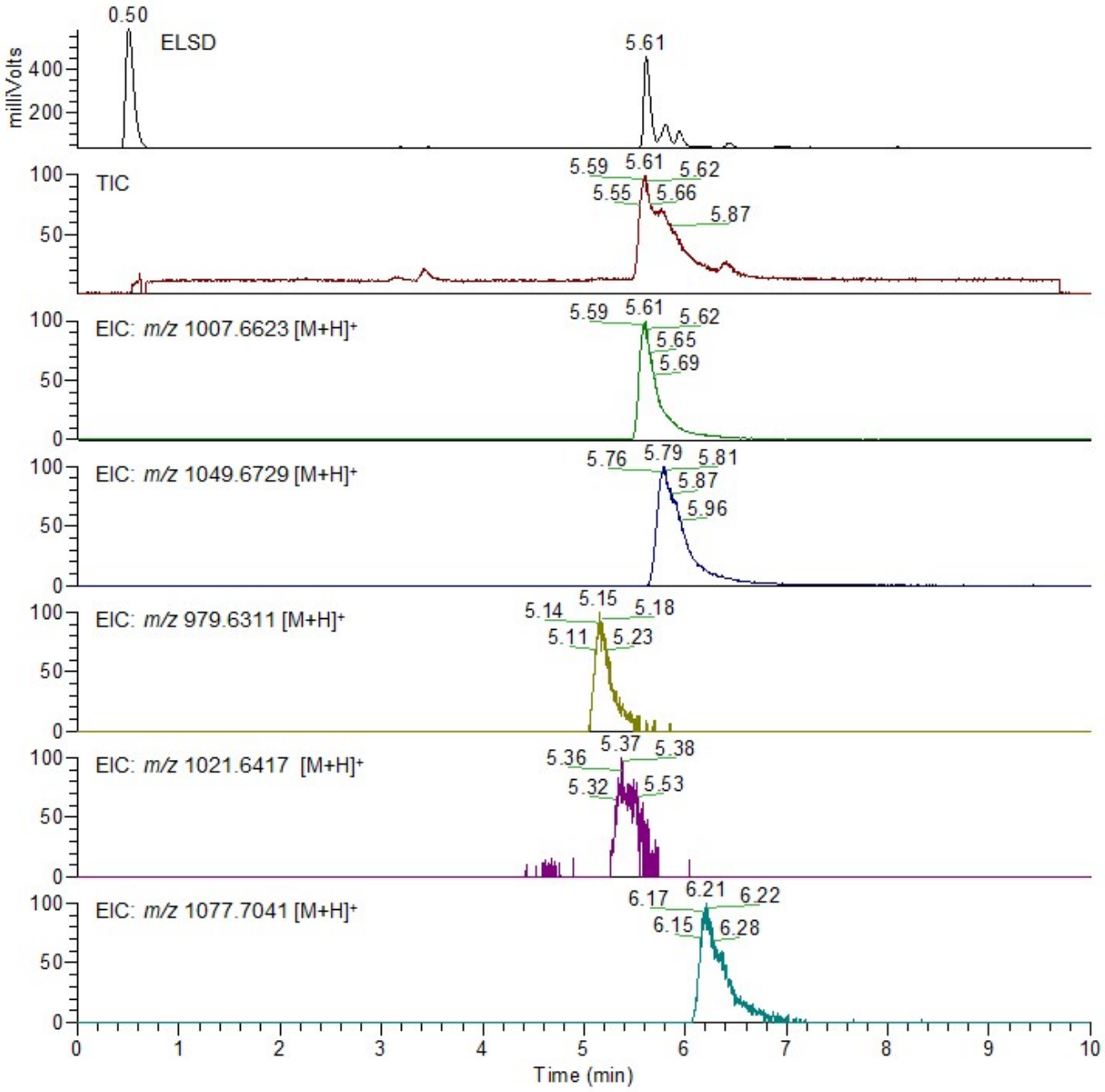


Fig. S1 UHPLC-HRMS-ELSD analysis of *V. paradoxus* RKNM0096 crude extract showing evaporative light scattering detector (ELSD) chromatogram, total ion chromatogram (TIC) and extracted ion chromatograms (EIC) for predicted *m/z* values of protonated ions for variovoracins **1**-**5**.

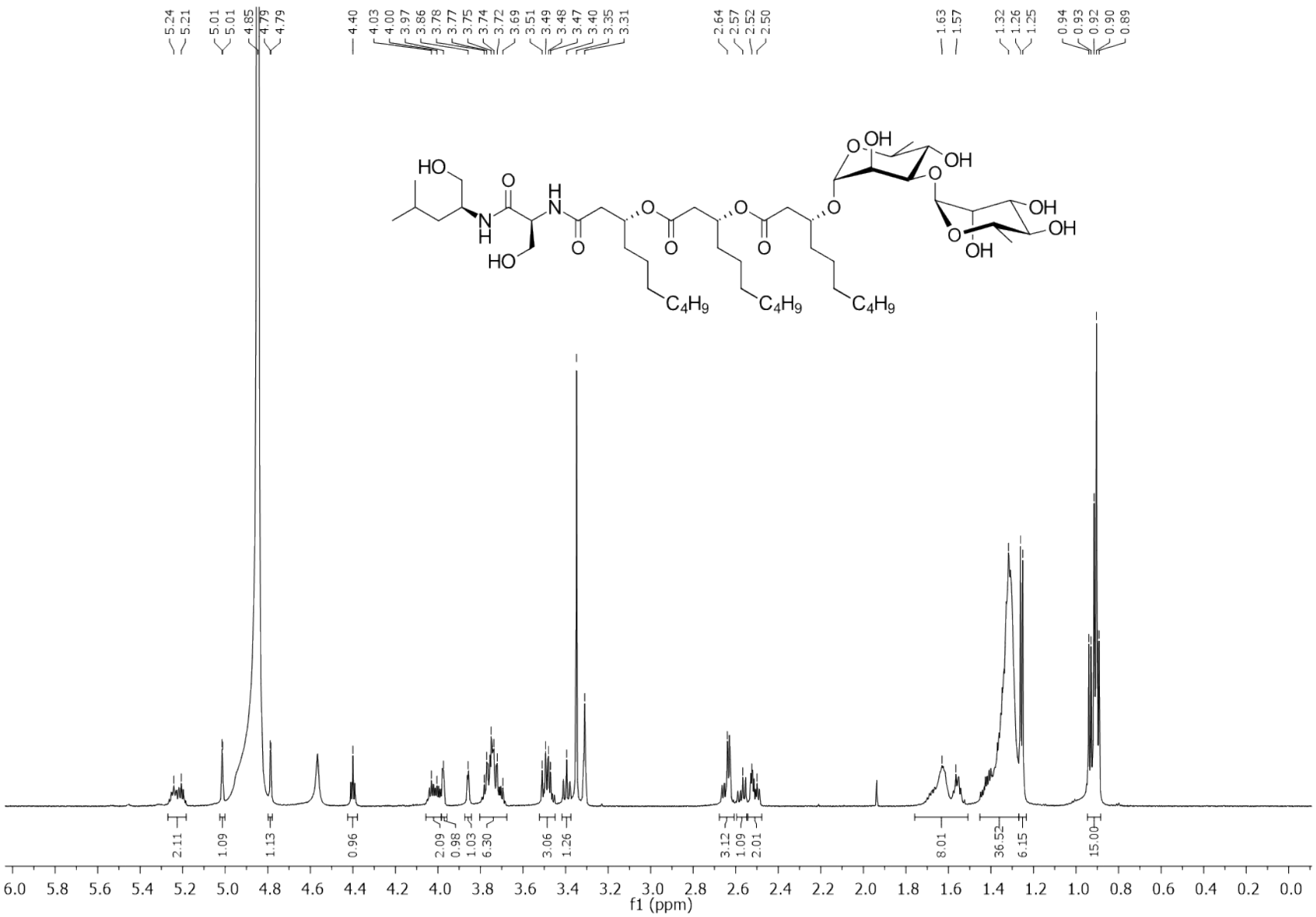


Fig. S2 ^{1}H NMR spectrum of Variovoracin Rha”-Rha’-$C_{10,10,10}$ (**1**) (600 MHz, CD_3OD).

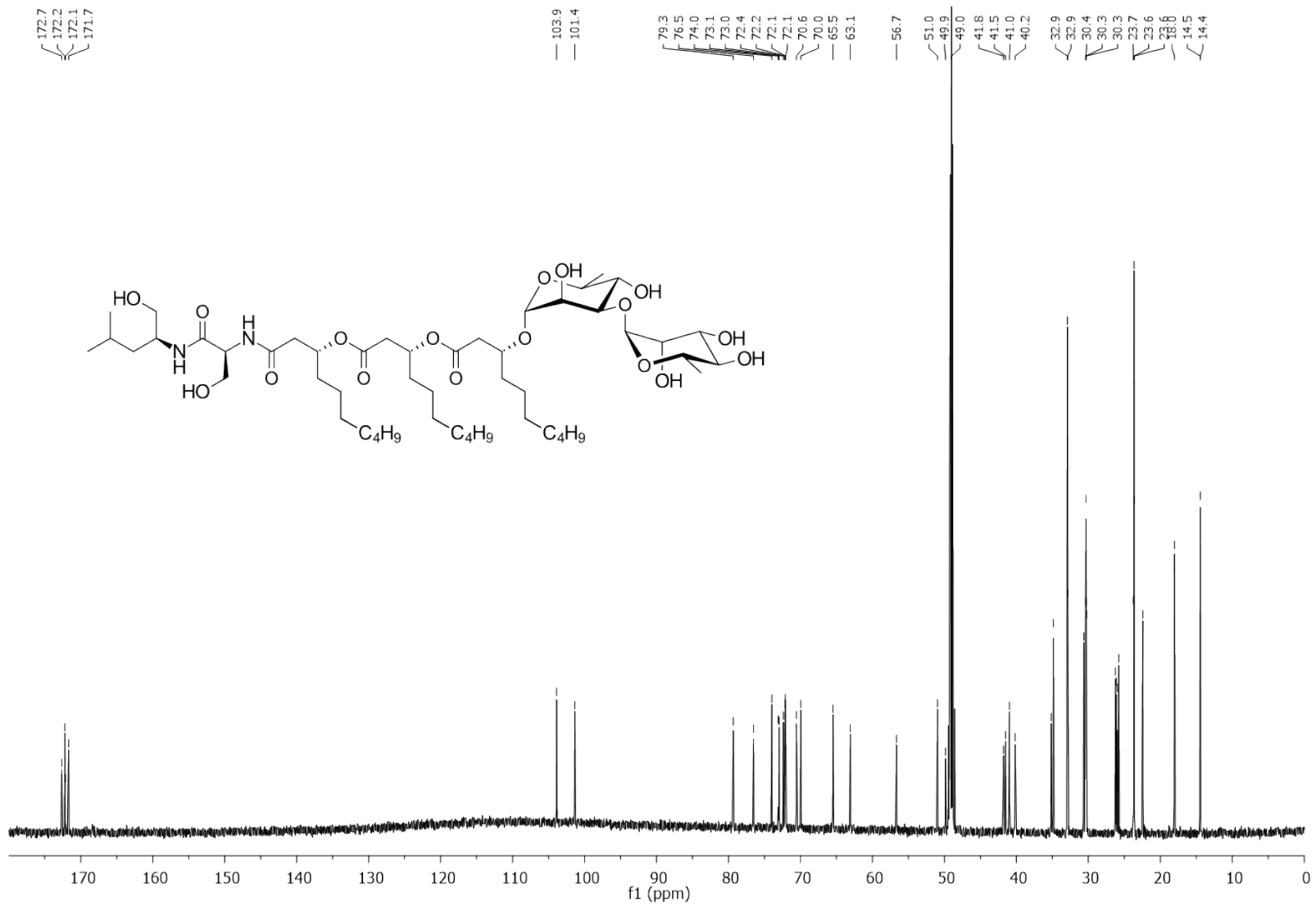


Fig. S3 ^{13}C NMR spectrum of Variovoracin Rha”-Rha’-$C_{10,10,10}$ (1) (150 MHz, CD_3OD).

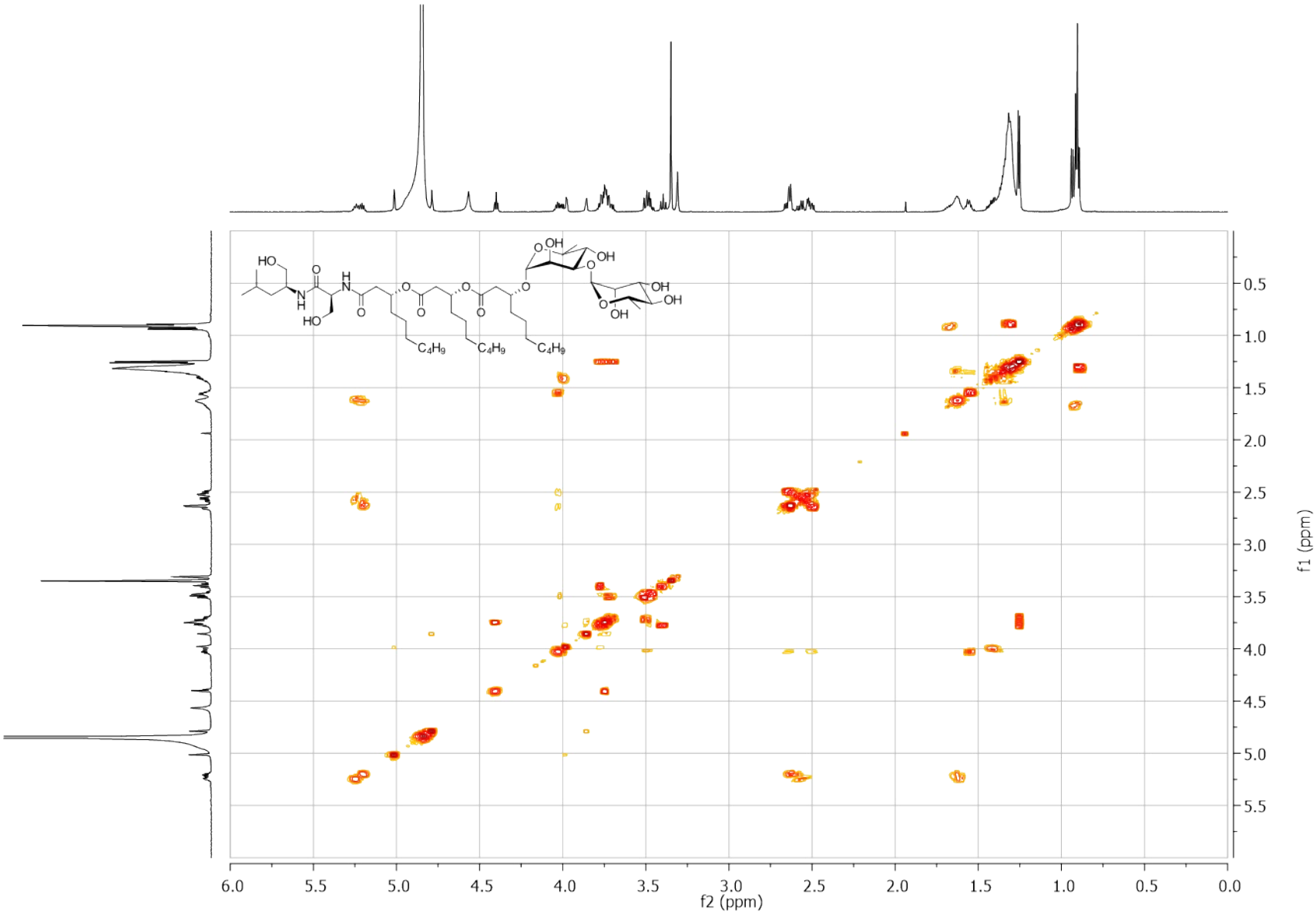


Fig. S4 COSY spectrum of Variovoracin Rha"-Rha'-$C_{10,10,10}$ (**1**).

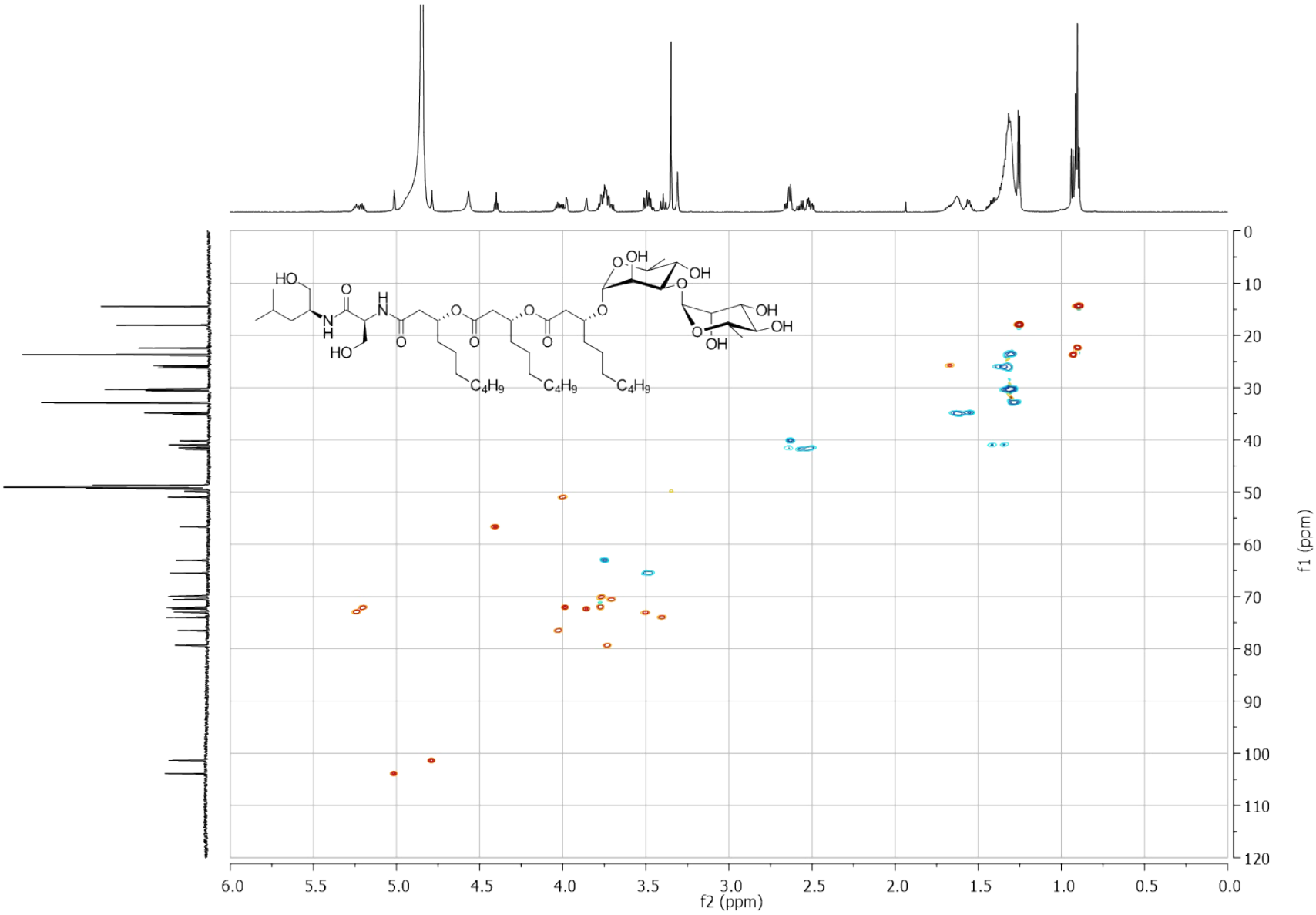


Fig. S5 HSQC spectrum of Variovoracin Rha"-Rha'-$C_{10,10,10}$ (**1**).

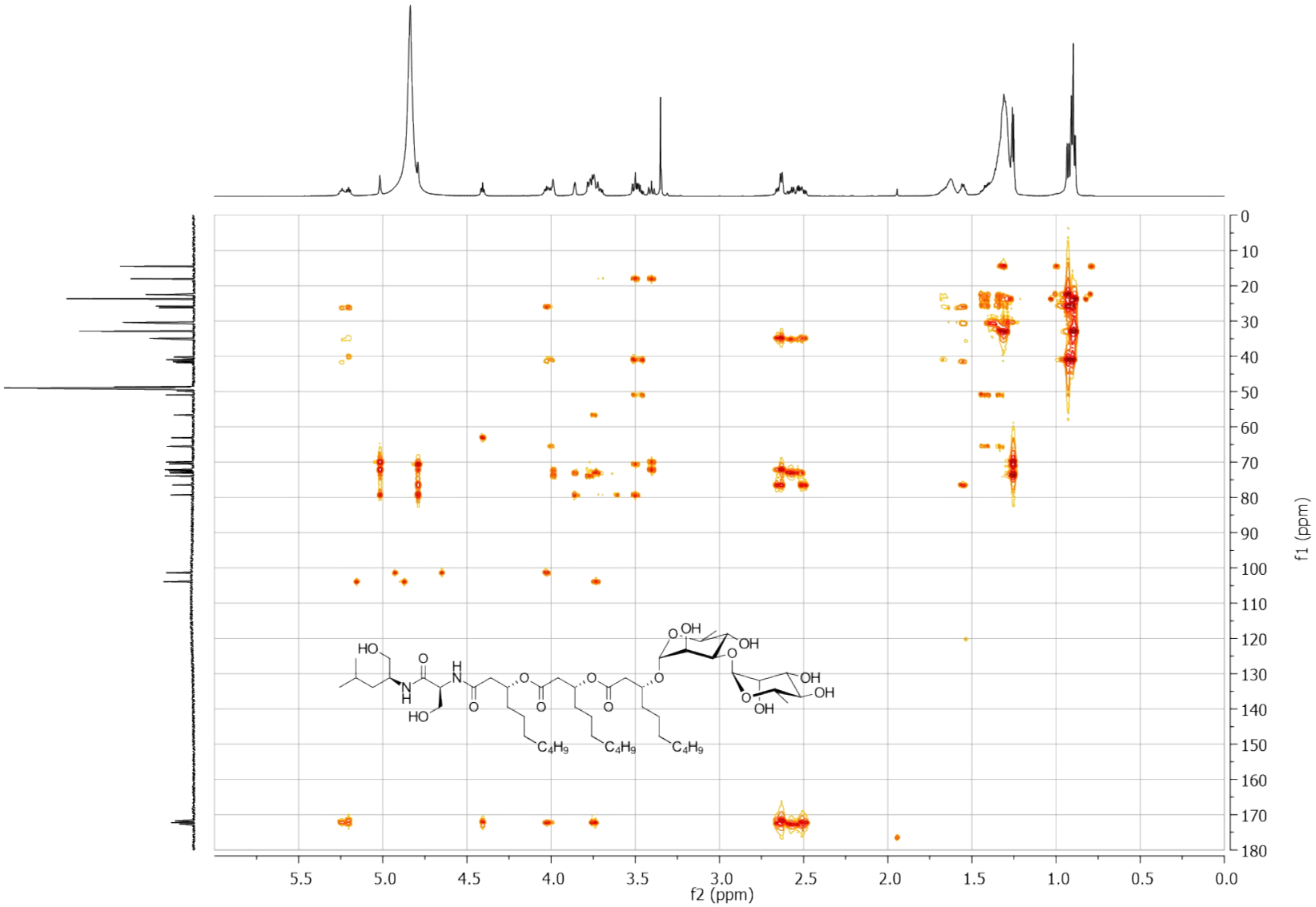


Fig. S6 HMBC spectrum of Variovoracin Rha”-Rha’-$C_{10,10,10}$ (**1**).

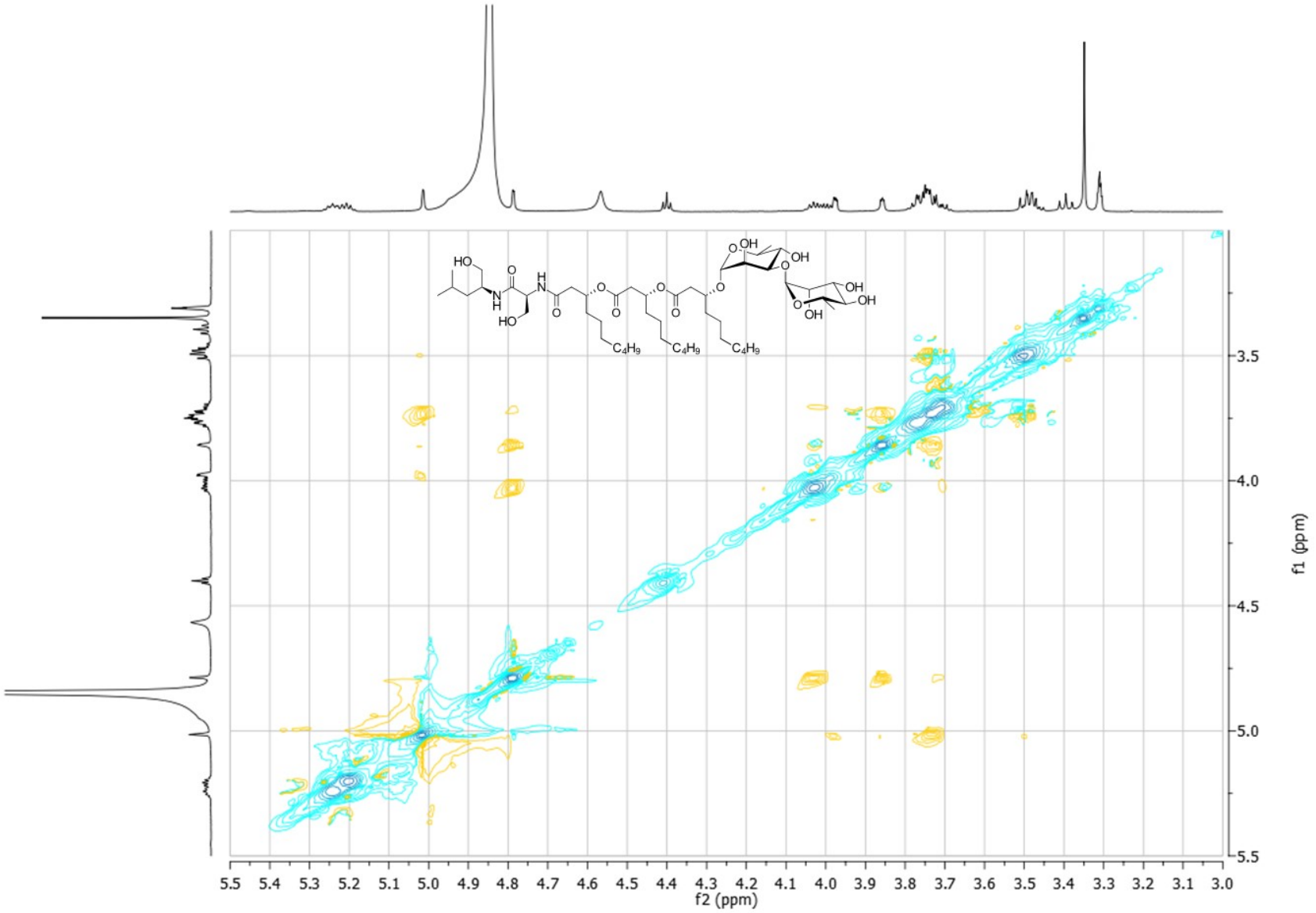


Fig. S7 ROESY spectrum of Variovoracin Rha”-Rha’-$C_{10,10,10}$ (**1**).

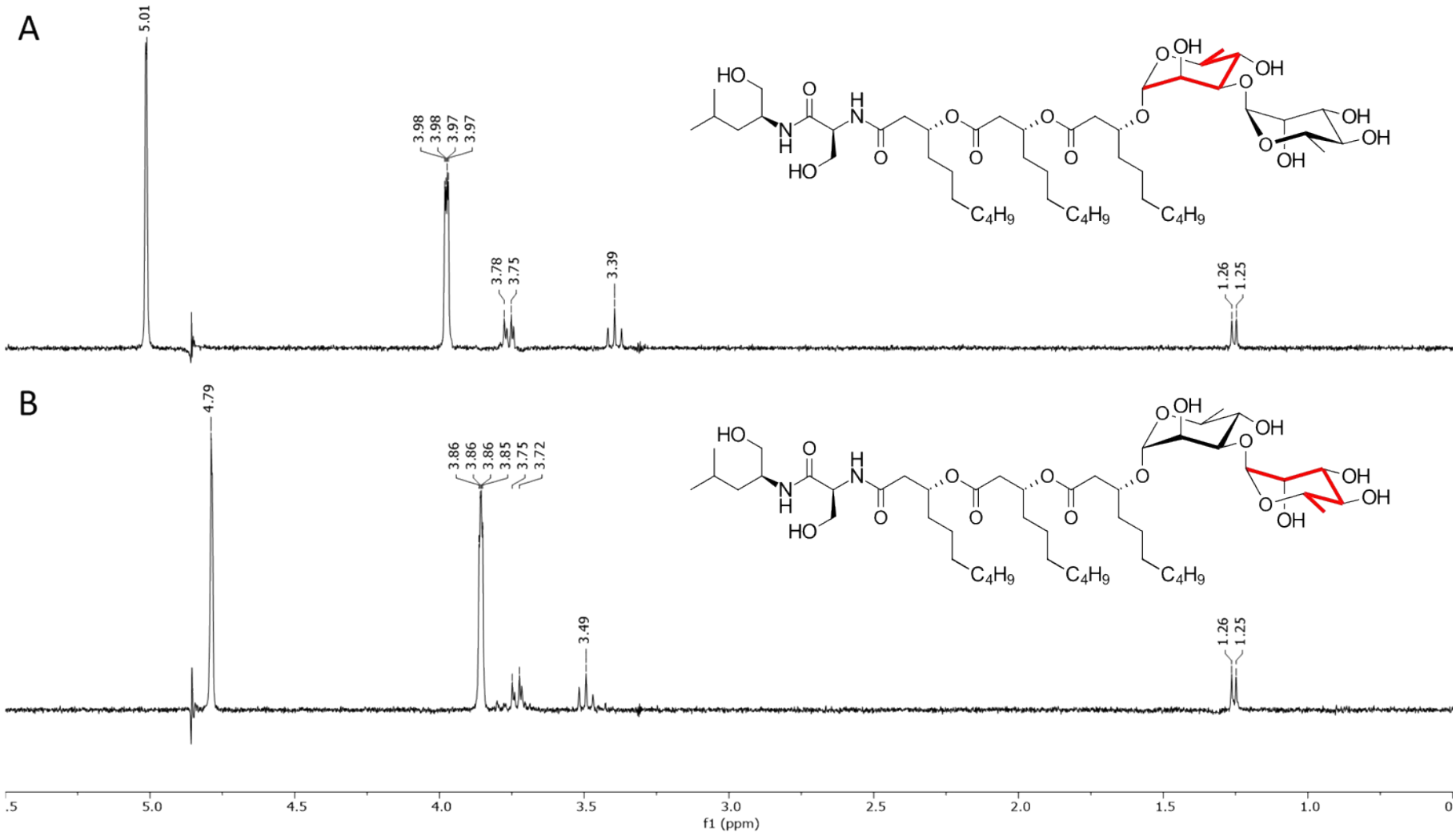


Fig. S8 1D TOCSY spectra of Variovoracin Rha"-Rha'-$C_{10,10,10}$ (**1**) from irradiation of (A) H-1' (δ_H 5.01; mixing time, d9: 200 ms) and (B) H-1" (δ_H 4.79; mixing time, d9: 200 ms). Red bonds show observed TOCSY correlations in each α-rhamnopyranose spin system.

Table S1 NMR spectroscopic data (600 MHz, CD_3OD) for Variovoracin Rha"-Rha'-$C_{10,10,10}$ (**1**).

Position	δ_C, type	δ_H, multiplicity (*J* in Hz)	COSY[a]	HMBC (H→C)[b]
Leucinol				
1L	65.5, CH_2	3.52 – 3.46, m	2L	2L, 3L
2L	51.0, CH	4.03 – 3.97, m	1L, 3aL	1L, 3L, 4L, 1S
3aL	41.0, CH_2	1.46 – 1.39, m	2L, 4L	1L, 2L, 4L
3bL		1.38 – 1.33, m		
4L	25.8, CH	1.72 – 1.65, m	3aL, 5L, 6L	
5L	23.7, CH_3	0.93, d (6.7)	4L	3L, 6L
6L	22.4, CH_3	0.92 – 0.89, m	4L	3L, 5L
Serine				
1S	172.1, C			
2S	56.7, CH	4.40, t (5.7)	3S	1S, 3S, 1A
3S	63.1, CH_2	3.78 – 3.72, m	2S	1S
Decanoic Acid (A)				
1A	172.7, C			
2aA	41.8, CH_2	2.58, dd (14.3, 8.0)	3A	1A, 3A, 4A
2bA		2.55 – 2.48, m		
3A	73.0, CH	5.27 – 5.21, m	2aA, 4A	1A, 2A, 4A, 5A, 1B
4A	35.2, CH_2	1.64 – 1.58, m	3A	
5A	26.3, CH_2	1.41 – 1.29, m		
6A, 7A	n.a.* CH_2	1.38 – 1.28, m		
8A	32.90[c], CH_2	1.32 – 1.24, m		
9A	23.64[c], CH_2	1.36 – 1.27, m	10A	
10A	14.44[c], CH_3	0.92 – 0.88, m	9A	8A, 9A
Decanoic Acid (B)				
1B	171.7, C			
2B	40.2, CH_2	2.66 – 2.61, m	3B	1B, 3B
3B	72.2, CH	5.23 – 5.18, m	2B, 4B	1B, 2B, 4B, 5B, 1C
4B	34.8, CH_2	1.67 – 1.61, m	3B	
5B	26.1, CH_2	1.41 – 1.29, m		
6B, 7B	n.a.* CH_2	1.38 – 1.28, m		
8B	32.90[c], CH_2	1.32 – 1.24, m		
9B	23.64[c], CH_2	1.36 – 1.27, m	10B	
10B	14.44[c], CH_3	0.92 – 0.88, m	9B	8B, 9B
Decanoic Acid (C)				
1C	172.2, C			
2aC	41.5, CH_2	2.68 – 2.61, m	3C	3C
2bC		2.53 – 2.48, m	3C	1C, 3C, 4C
3C	76.5, CH	4.06 – 4.00, m	2aC, 2bC, 4C	1C, 2C, 1'
4C	34.8, CH_2	1.59 – 1.53, m	3C	2C, 3C, 5C
5C	26.0, CH_2	1.41 – 1.29, m		
6C, 7C	n.a.* CH_2	1.38 – 1.28, m		
8C	32.87[c], CH_2	1.32 – 1.24, m		
9C	23.63[c], CH_2	1.36 – 1.27, m	10C	
10C	14.46[c], CH_3	0.92 – 0.88, m	9C	8C, 9C
α-rhamnose (I)				
1'	101.4, CH	4.79, d (1.6)	2'	2', 3', 5', 3C
2'	72.4, CH	3.86, dd (3.3, 1.6)	1', 3'	3', 4'
3'	79.3, CH	3.73, dd (9.6, 3.3)	2'	4', 1''
4'	73.1, CH	3.49, app. t (9.6)		3', 5', 6'
5'	70.6, CH	3.74 – 3.68, m	6'	
6'	18.0, CH_3	1.26, d (6.2)	5'	4', 5'
α-rhamnose (II)				
1"	103.9, CH	5.01, d (1.6)	2"	5", 3'
2"	72.10, CH	3.98, dd (3.4, 1.6)	1", 3"	4"
3"	72.06, CH	3.76, dd (9.5, 3.4)	2"	
4"	74.0, CH	3.40, app. t (9.5)		2", 5", 6"
5"	70.0, CH	3.80 – 3.74, m	6"	
6"	18.0, CH_3	1.26, d (6.2)	5"	4", 5"

[a] COSY correlations are from proton(s) stated to the indicated proton. [b] HMBC correlations (H→C) are from proton(s) stated to the indicated carbon. [c] Assignment of ^{13}C NMR resonances to specific β-hydroxydecanoic acid moieties was not possible due to signal overlap. * The following overlapping ^{13}C NMR resonances could not be assigned (n.a.) unambiguously: 30.6, 30.38, 30.35, 30.35, 30.29, 30.2.

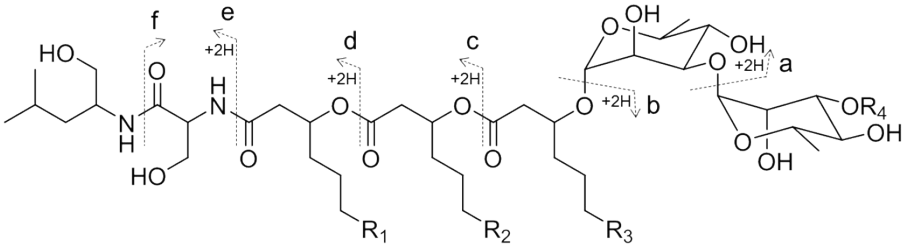


Fig. S9 Variovoracin fragmentations and general scheme for fragment ion annotation.

Table S2 Assignment of fragment ions observed from MS^n spectra of Variovoracin Rha"-Rha'-$C_{10,10,10}$ (**1**) (precursor ion *m*/*z* 1007.66350 $[M+H]^+$).

Fragment Type	Observed *m/z*	Theoretical *m/z*	Proposed Formula	Error (ppm)
$[M + H]^+$	1007.66350	1007.66253	$C_{51}H_{95}N_2O_{17}^+$	0.96
$[M + H – H_2O]^+$	989.64993	989.65196	$C_{51}H_{93}N_2O_{16}^+$	-2.05
$[M + H – H_4O_2]^+$	971.63919	971.64140	$C_{51}H_{91}N_2O_{15}^+$	-2.27
$[M + H – H_6O_3]^+$	953.62857	953.63083	$C_{51}H_{89}N_2O_{14}^+$	-2.37
a	861.60362	861.60462	$C_{45}H_{85}N_2O_{13}^+$	-1.16
a – H_2O	843.59233	843.59405	$C_{45}H_{83}N_2O_{12}^+$	-2.04
a – H_4O_2	825.58181	825.58349	$C_{45}H_{81}N_2O_{11}^+$	-2.03
a – H_6O_3	807.57124	807.57292	$C_{45}H_{79}N_2O_{10}^+$	-2.08
af	744.48785	744.48925	$C_{39}H_{70}NO_{12}^+$	-1.88
b	715.54563	715.54671	$C_{39}H_{75}N_2O_9^+$	-1.52
b – H_2O	697.53514	697.53614	$C_{39}H_{73}N_2O_8^+$	-1.43
b – H_4O_2	679.52453	679.52558	$C_{39}H_{71}N_2O_7^+$	-1.56
b – H_6O_3	661.51375	661.51501	$C_{39}H_{69}N_2O_6^+$	-1.90
bf	598.43073	598.43134	$C_{33}H_{60}NO_8^+$	-1.04
c – H_4O_2	509.39399	509.39490	$C_{29}H_{53}N_2O_5^+$	-1.79
c – H_6O_3	491.38349	491.38433	$C_{29}H_{51}N_2O_4^+$	-1.71
c – H_8O_4	473.37303	473.37377	$C_{29}H_{49}N_2O_3^+$	-1.56
d – H_4O_2	339.26374	339.26422	$C_{19}H_{35}N_2O_3^+$	-1.41
d – H_6O_3	321.25325	321.25365	$C_{19}H_{33}N_2O_2^+$	-1.25
df – H_2O	240.15907	240.15942	$C_{13}H_{22}NO_3^+$	-1.46
f	890.54532	890.54716	$C_{45}H_{80}NO_{16}^+$	-2.07

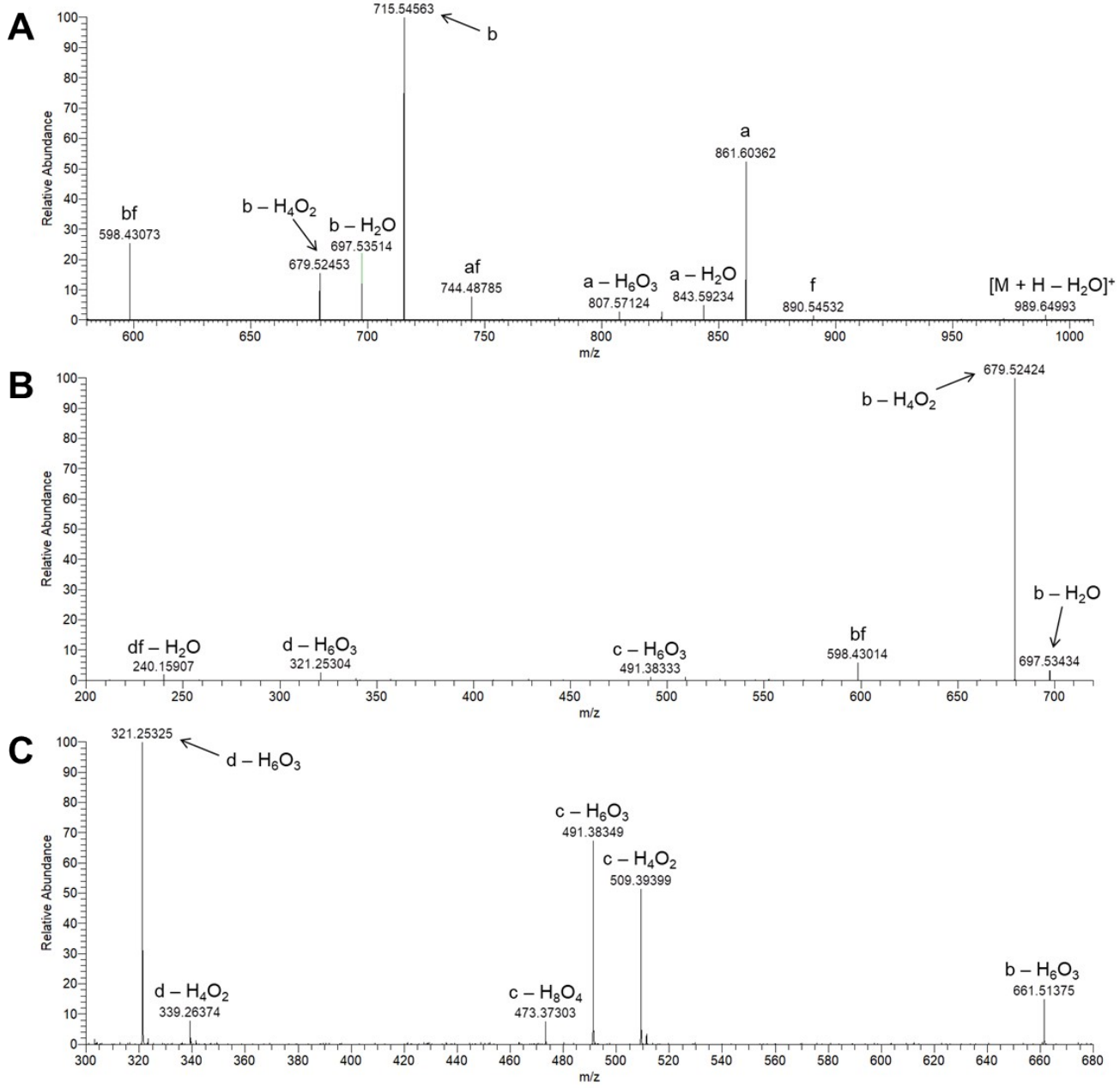


Fig. S10 MS^n spectra of Variovoracin Rha"-Rha'-$C_{10,10,10}$ (**1**) and assignment of observed fragment ions. (A) ESI-MS^2 spectrum of precursor ion *m*/*z* 1007.66 $[M + H]^+$. (B) ESI-MS^3 spectrum of fragment ion *m*/*z* 715.54 (fragment: b). (C) ESI-MS^4 spectrum of fragment ion *m*/*z* 679.52 (fragment: b – H_4O_2).

Table S3 NMR spectroscopic data (600 MHz, CD_3OD) for Variovoracin Rha"-$OAc_{3"}$-Rha'-$C_{10,10,10}$ (**2**).

Position	δ_C, type	δ_H, multiplicity (*J* in Hz)	COSY[a]	HMBC[b]
Leucinol				
1aL	65.5, CH_2	3.50, dd (11.1, 5.2)	1bL, 2L	2L, 3L
1bL		3.47, dd (11.1, 5.8)	1aL, 2L	2L, 3L
2L	51.0, CH	4.03 – 3.97, m	1L, 3aL	1L, 3L, 4L, 1S
3aL	41.0, CH_2	1.45 – 1.38, m	2L	1L, 2L, 4L, 5L, 6L
3bL		1.39 – 1.31, m		
4L	25.8, CH	1.71 – 1.64, m	5L, 6L	2L, 3L, 5L, 6L
5L	23.8, CH_3	0.93, d (6.6)	4L	3L, 4L, 6L
6L	22.4, CH_3	0.92 – 0.89, m	4L	3L, 4L, 5L
Serine				
1S	172.1, C			
2S	56.7, CH	4.40, t (5.6)	3S	3S, 1A
3S	63.1, CH_2	3.77 – 3.71, m	2S	1S, 2S
Decanoic Acid (A)				
1A	171.7, C			
2A	40.2, CH_2	2.66 – 2.61, m	3A	1A, 3A
3A	72.1, CH	5.23 – 5.18, m	2A, 4A	2A, 4A, 5A
4A	34.90, CH_2	1.66 – 1.59, m	3A, 5A	
5A	26.2, CH_2	1.41 – 1.30, m	4A	
6A, 7A	n.a. CH_2	1.37 – 1.27, m		
8A	32.95*, CH_2	1.33 – 1.25, m		
9A	23.69*, CH_2	1.35 – 1.28, m	10A	
10A	14.44*, CH_3	0.92 – 0.88, m	9A	8A, 9A
Decanoic Acid (B)				
1B	172.7, C			
2aB	41.8, CH_2	2.57, dd (14.3, 7.9)	3B	1B, 3B, 4B
2bB		2.54 – 2.48, m		
3B	72.98, CH	5.27 – 5.22, m	2aB, 4B	2B, 4B, 5B
4B	35.2, CH_2	1.66 – 1.59, m	3B, 5B	
5B	26.3, CH_2	1.41 – 1.30, m	4B	
6B, 7B	n.a. CH_2	1.37 – 1.27, m		
8B	32.95*, CH_2	1.33 – 1.25, m		
9B	23.69*, CH_2	1.35 – 1.28, m	10B	
10B	14.44*, CH_3	0.92 – 0.88, m	9B	8B, 9B
Decanoic Acid (C)				
1C	172.2, C			
2aC	41.5, CH_2	2.67 – 2.61, m	2bC, 3C	1C, 3C, 4C
2bC		2.54 – 2.48, m	3aC, 3C	
3C	76.6, CH	4.06 – 4.00, m	2aC, 2bC, 4C	2C, 4C, 5C, 1'
4C	34.88, CH_2	1.59 – 1.53, m	3C, 5C	
5C	26.0, CH_2	1.41 – 1.30, m	4C	
6C, 7C	n.a. CH_2	1.37 – 1.27, m		
8C	32.92*, CH_2	1.33 – 1.25, m		
9C	23.68*, CH_2	1.35 – 1.28, m	10C	
10C	14.46*, CH_3	0.92 – 0.88, m	9C	8C, 9C
α-rhamnose (I)				
1'	101.4, CH	4.80, d (1.3)	2'	3C, 2', 3', 5'
2'	72.4, CH	3.90 – 3.84, m	1', 3'	3', 4'
3'	79.8, CH	3.77 – 3.71, m	2'	4', 1''
4'	73.05, CH	3.54, app. t (9.6)	5'	3', 5', 6'
5'	70.6, CH	3.74 – 3.68, m	4', 6'	3', 4', 6'
6'	18.02, CH_3	1.26, d (6.2)	5'	4', 5'
α-rhamnose (II)				
1"	103.8, CH	5.03, d (1.3)	2"	3', 3", 5"
2"	69.9, CH	4.13, dd (3.0, 1.3)	1", 3"	1", 3", 4"
3"	75.8, CH	5.00 – 5.05, m	2", 4"	4", 5", O*C*OCH_3
4"	71.3, CH	3.61, app. t (9.7)	3", 5"	2", 3", 5", 6"
5"	70.1, CH	3.90 – 3.86, m	4", 6"	
6"	18.05, CH_3	1.28, d (6.6)	5"	4", 5"
O*C*OCH_3	172.9, C			
OCO*CH*$_3$	21.1, CH_3	2.12, s		O*C*OCH_3

[a] COSY correlations are from proton(s) stated to the indicated proton. [b] HMBC correlations (H→C) are from proton(s) stated to the indicated carbon. [c] Assignment of ^{13}C NMR resonances to specific β-hydroxydecanoic acid moieties was not possible due to signal overlap. * The following overlapping ^{13}C NMR resonances could not be assigned (n.a.) unambiguously: 30.7, 30.42, 30.40, 30.40, 30.34, 30.30.

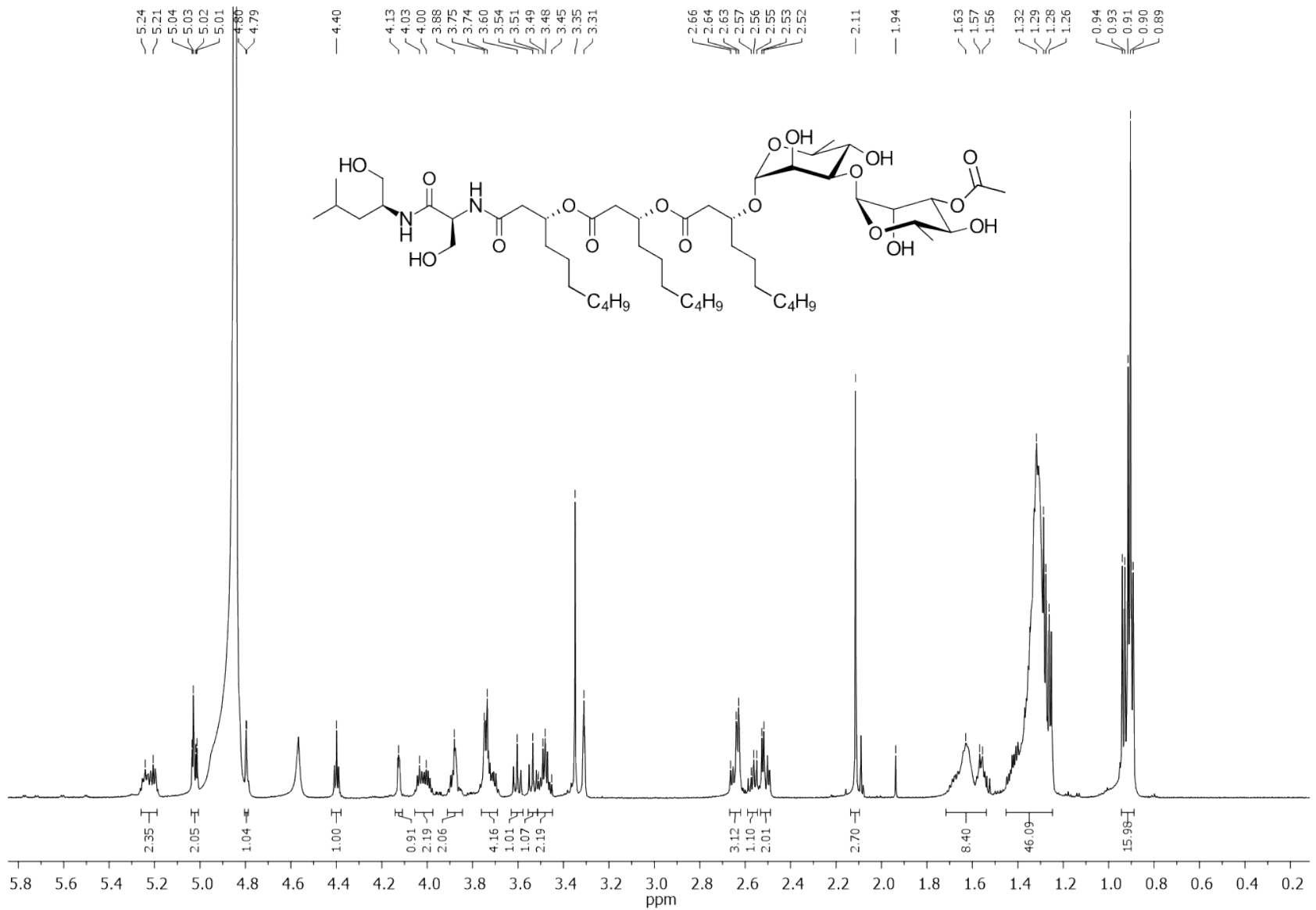


Fig. S11 ^{1}H NMR spectrum of Variovoracin Rha"-OAc_{3}-Rha'-$C_{10,10,10}$ (**2**) (600 MHz, CD_3OD).

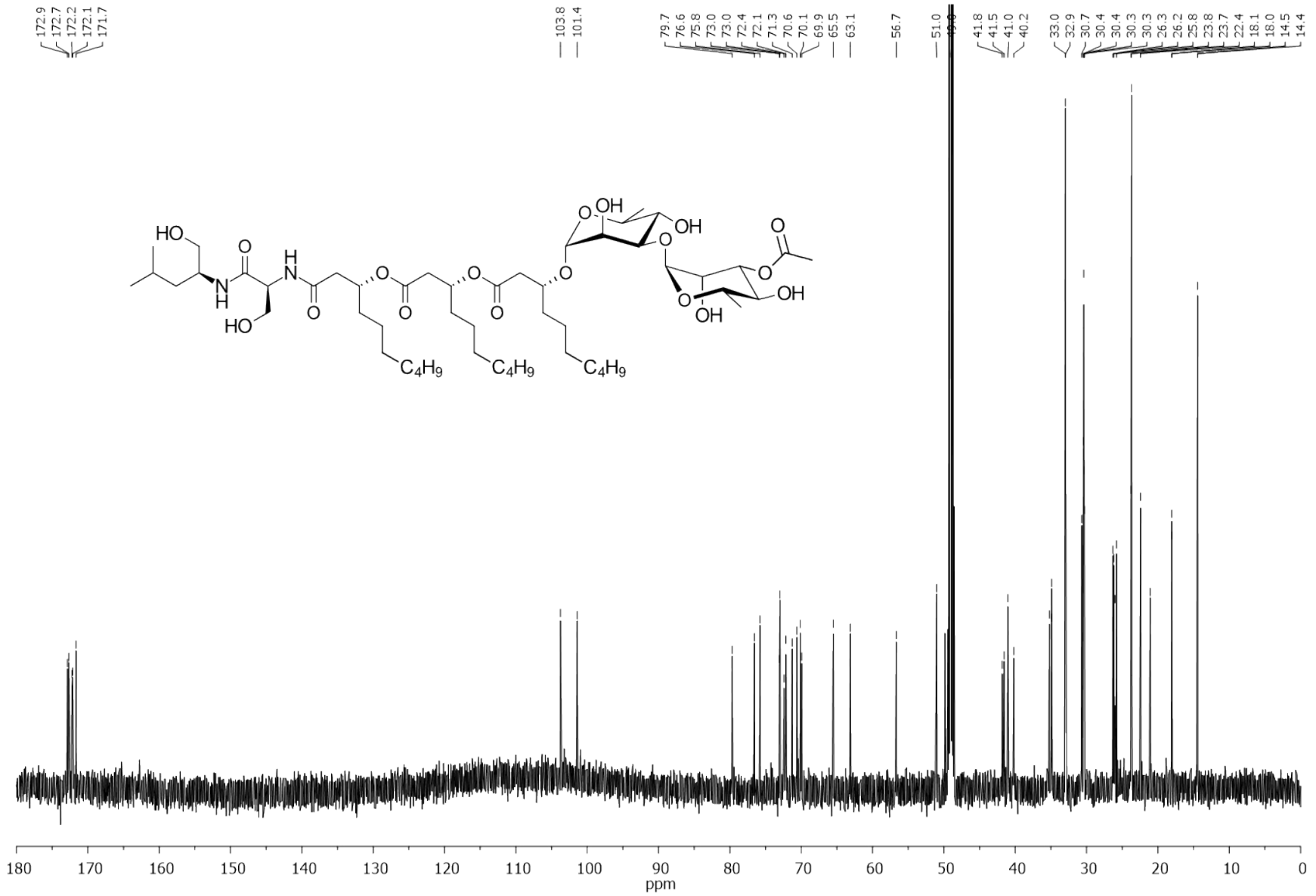


Fig. S12 ^{13}C NMR spectrum of Variovoracin Rha"-OAc_3"-Rha'-$C_{10,10,10}$ (**2**) (150 MHz, CD_3OD).

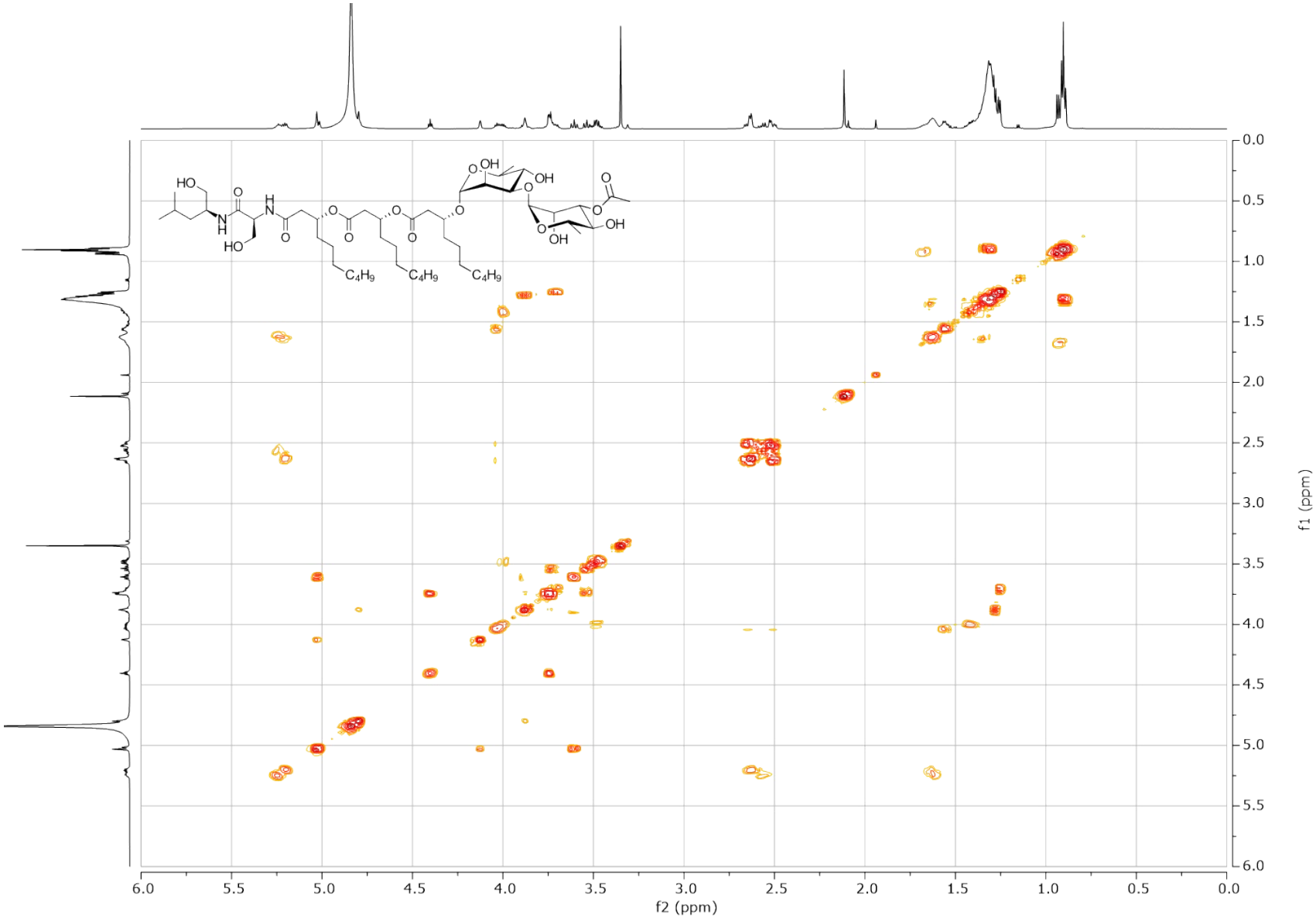


Fig. S13 COSY spectrum of Variovoracin Rha"-OAc_3"-Rha'-$C_{10,10,10}$ (**2**).

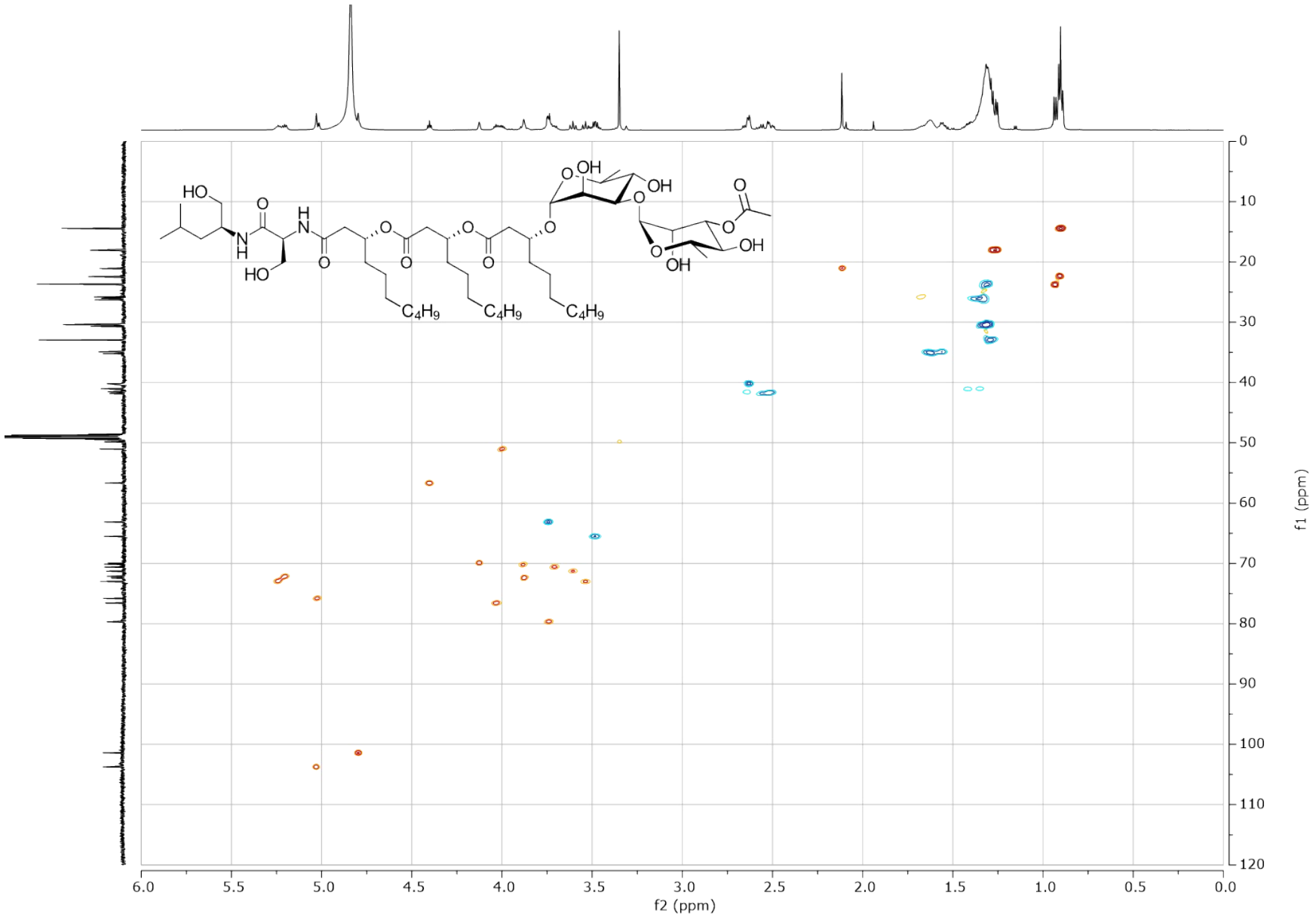


Fig. S14 HSQC spectrum of Variovoracin Rha”-$OAc_{3”}$-Rha’-$C_{10,10,10}$ (**2**).

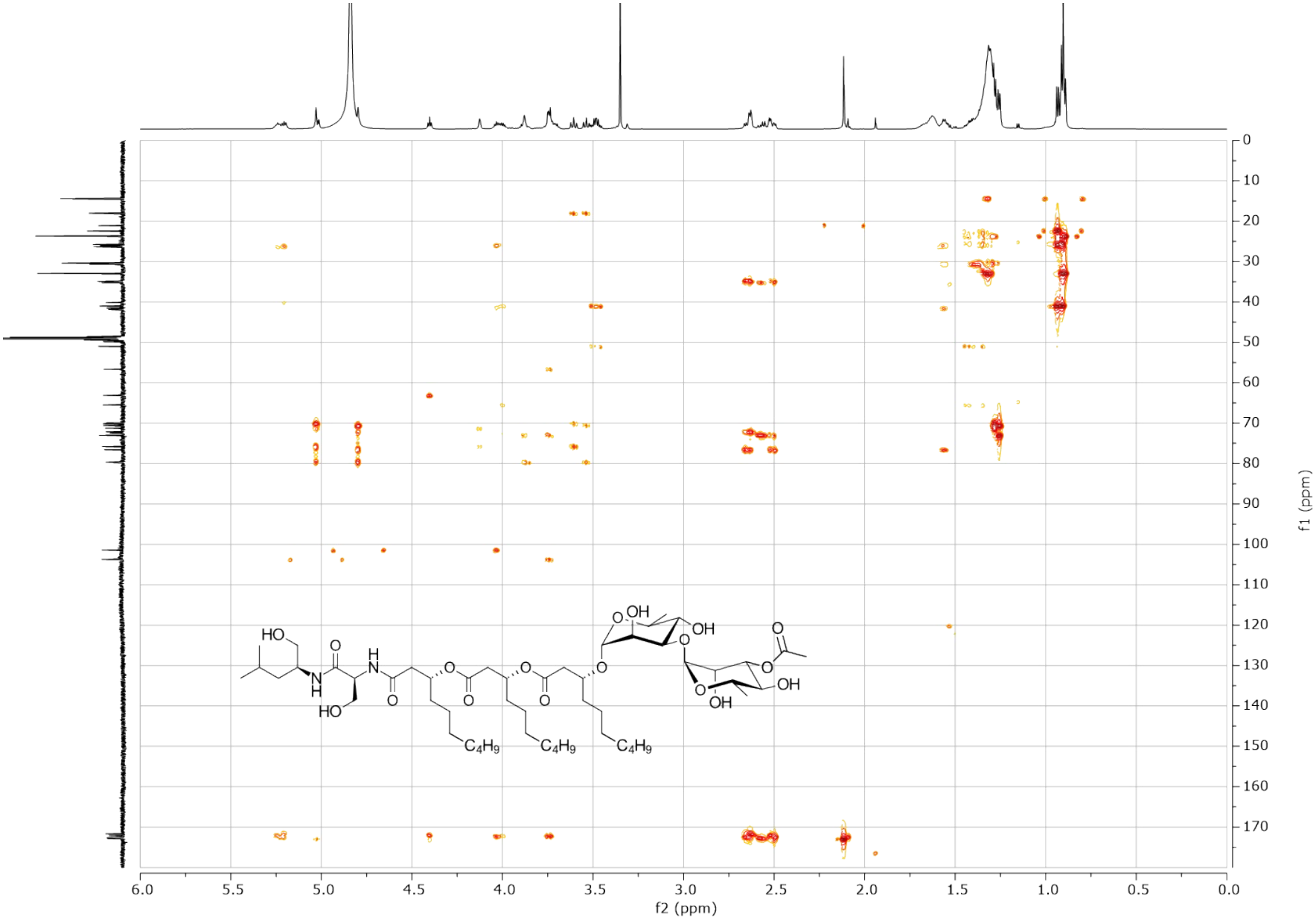


Fig. S15 HMBC spectrum of Variovoracin Rha"-OAc_3-Rha'-$C_{10,10,10}$ (**2**).

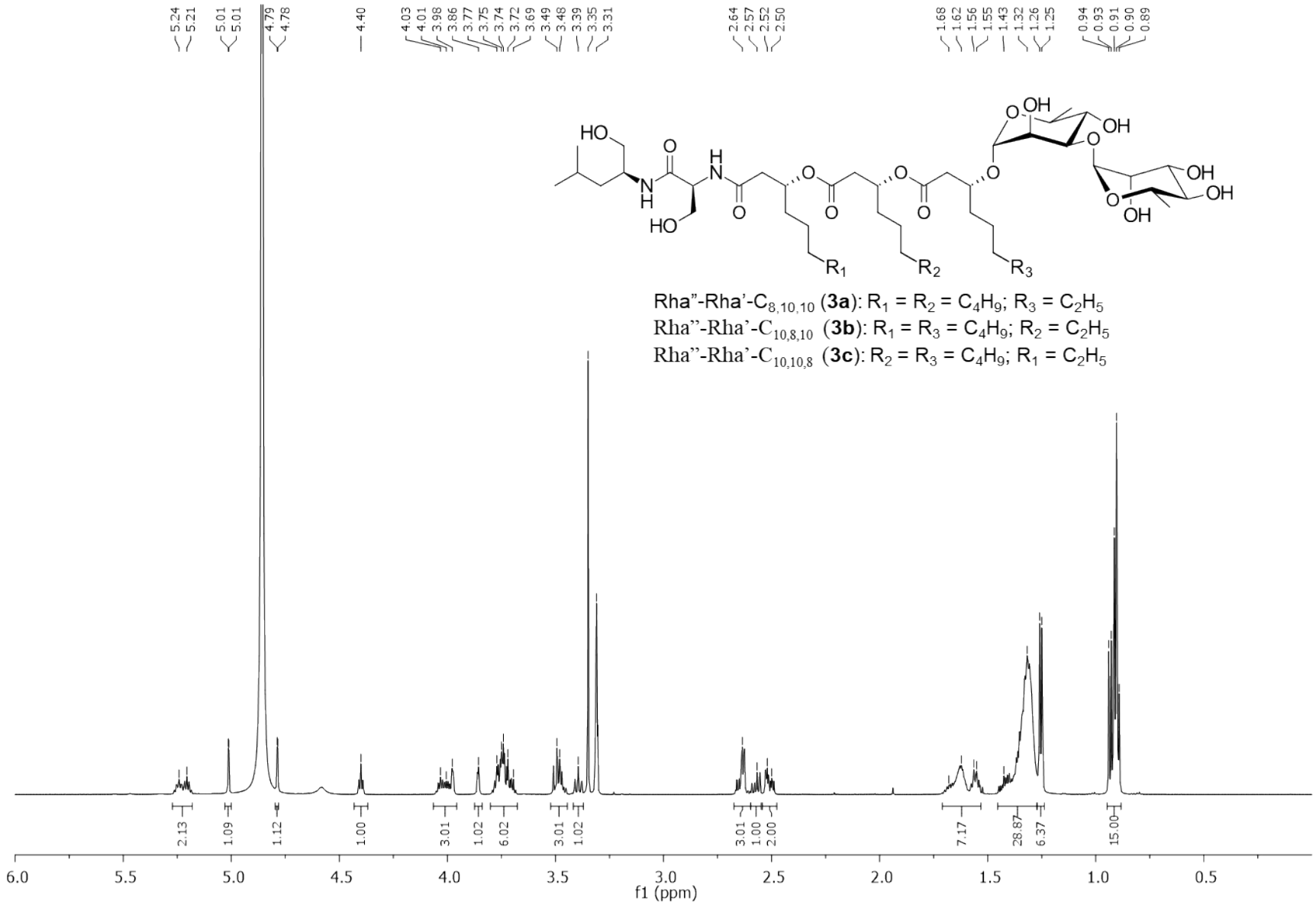


Fig. S16 ^{1}H NMR spectrum of Variovoracins Rha”-Rha’-$C_{8,10,10}$ (**3a**), Rha”-Rha’-$C_{10,8,10}$ (**3b**), and Rha”-Rha’-$C_{10,10,8}$ (**3c**) (600 MHz, CD_3OD).

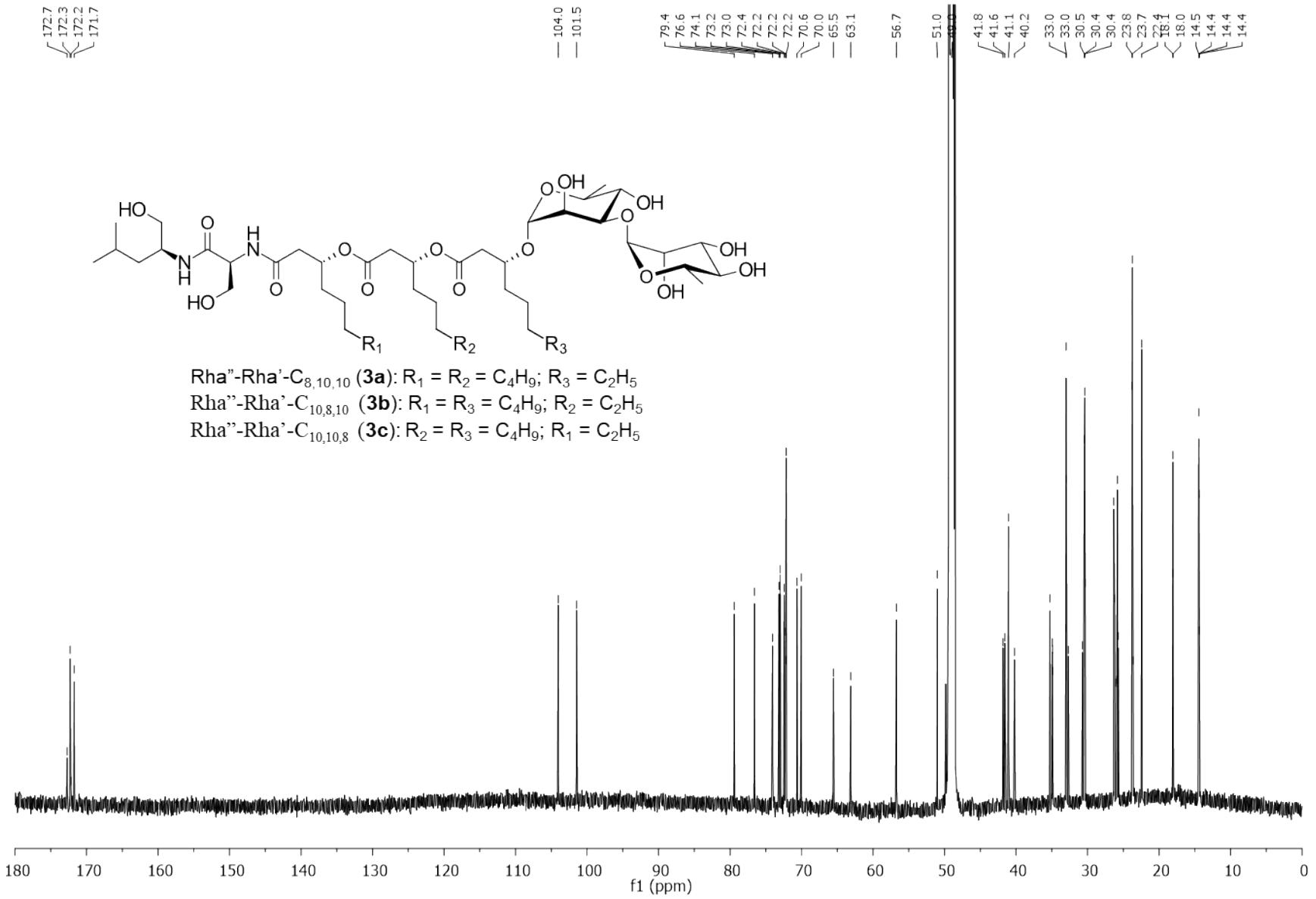


Fig. S17 ^{13}C NMR spectrum of Variovoracins Rha"-Rha'-$C_{8,10,10}$ (**3a**), Rha"-Rha'-$C_{10,8,10}$ (**3b**), and Rha"-Rha'-$C_{10,10,8}$ (**3c**) (150 MHz, CD_3OD).

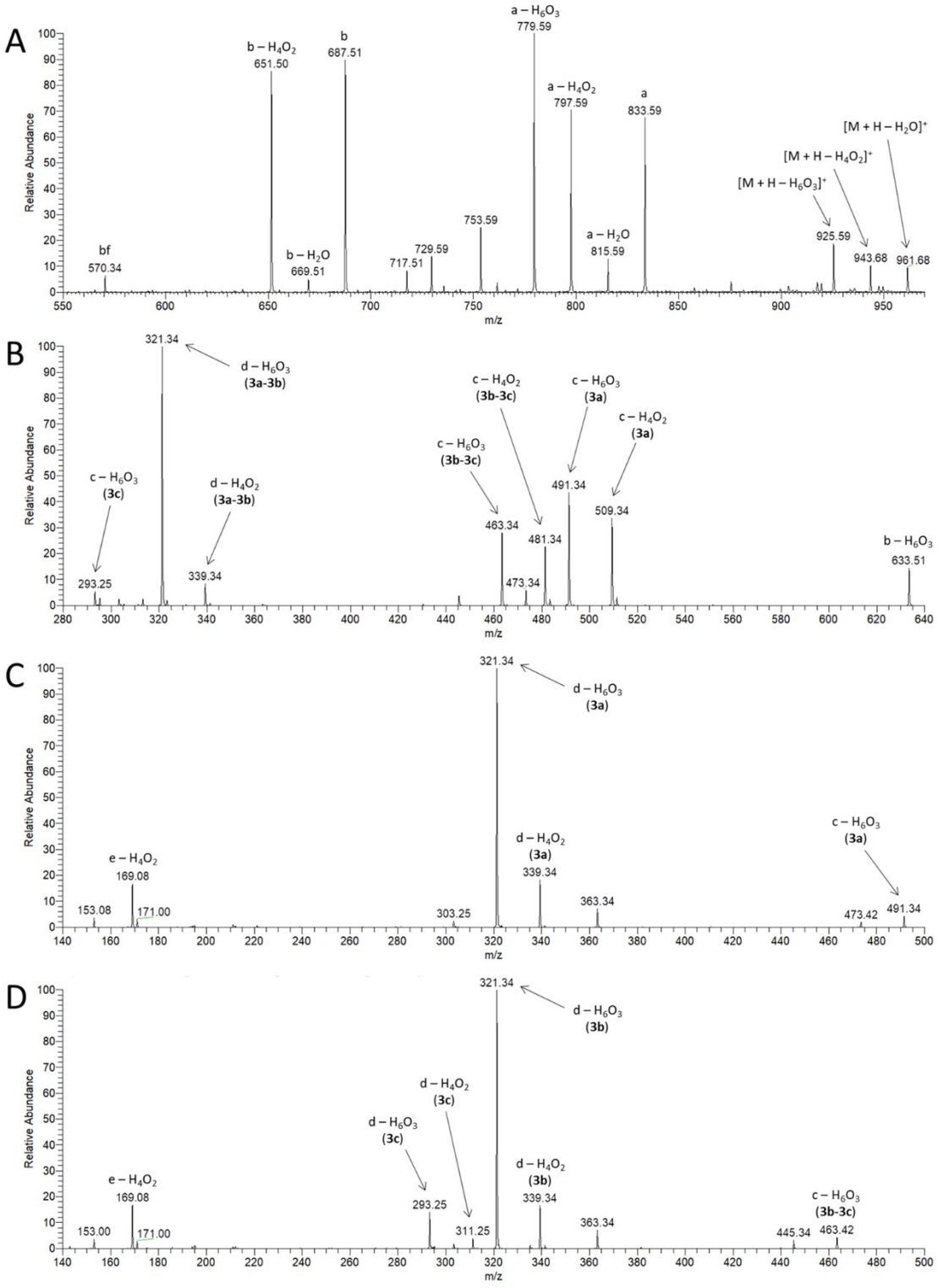


Fig. S18 Analysis of Variovoracins Rha"-Rha'-$C_{8,10,10}$ (**3a**), Rha"-Rha'-$C_{10,8,10}$ (**3b**), and Rha"-Rha'-$C_{10,10,8}$ (**3c**) by ESI-MS/MS and assignment of observed fragment ions. (A) ESI-MS^2 scan of precursor ion *m/z* 979.7 $[M + H]^+$. (B) ESI-MS^3 scan of fragment ion b – H_4O_2 *m/z* 651.5. (C) ESI-MS^4 scan of fragment ion c – H_4O_2 *m/z* 509.3 from Rha"-Rha'-$C_{8,10,10}$ (**3a**). (D) ESI-MS^4 scan of fragment ion c – H_4O_2 *m/z* 481.3 from Rha"-Rha'-$C_{10,8,10}$ (**3b**) and Rha"-Rha'-$C_{10,10,8}$ (**3c**).

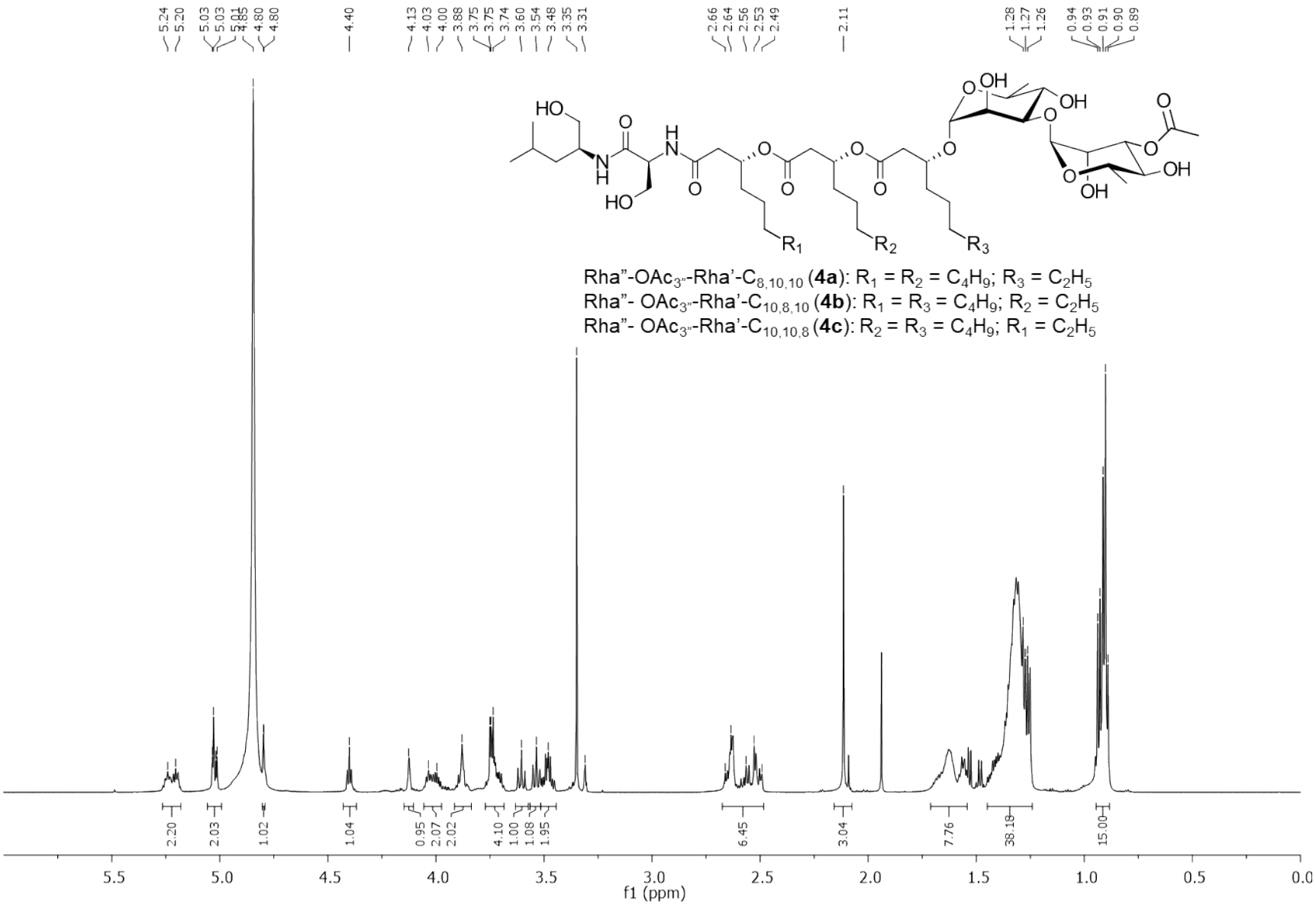


Fig. S19 ^{1}H NMR spectrum of Variovoracins Rha"-$OAc_{3"}$-Rha'-$C_{8,10,10}$ (**4a**), Rha"-$OAc_{3"}$-Rha'-$C_{10,8,10}$ (**4b**), and Rha"-$OAc_{3"}$-Rha'-$C_{10,10,8}$ (**4c**) (600 MHz, CD_3OD).

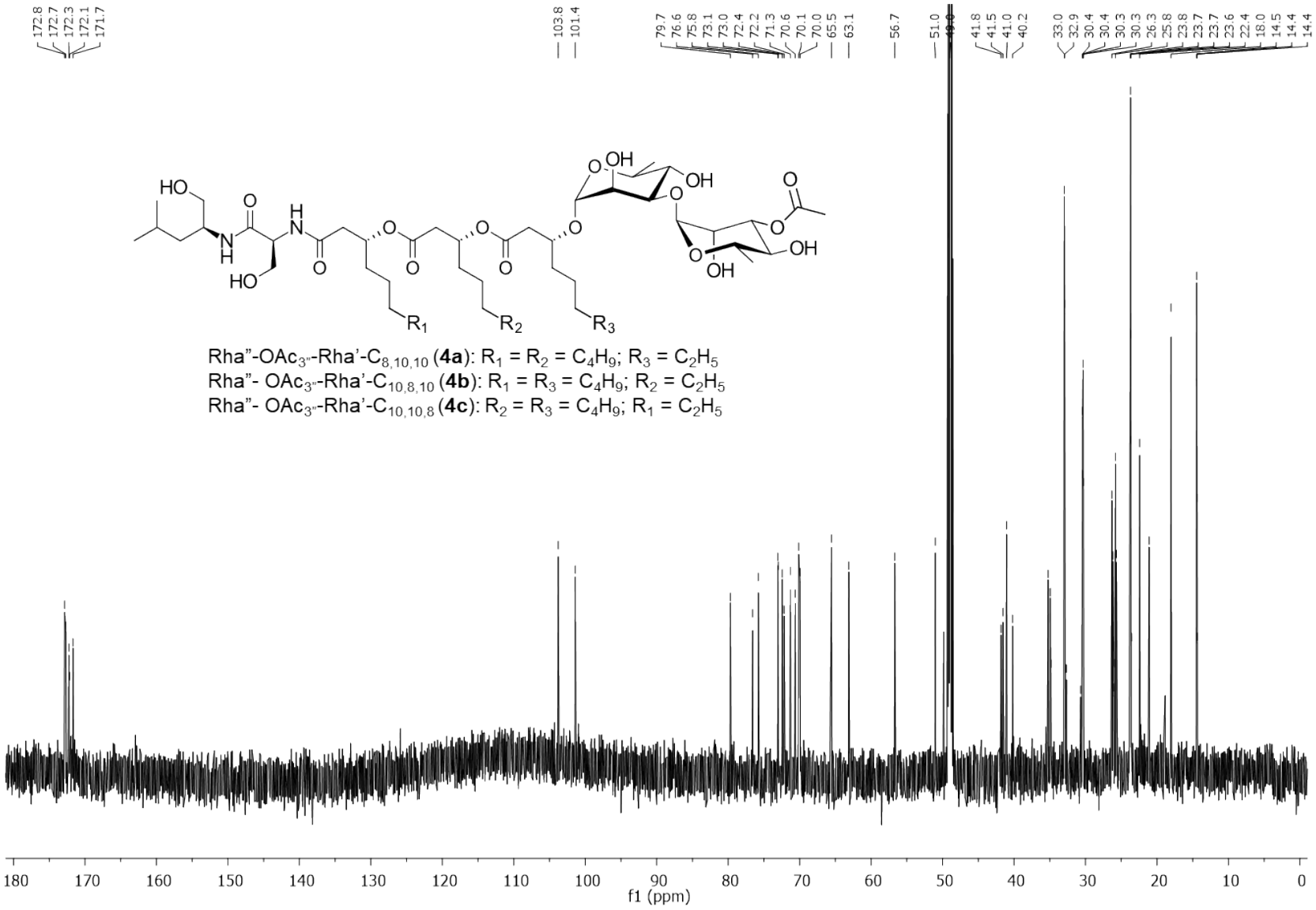


Fig. S20 ^{13}C NMR spectrum of Variovoracins Rha"-$OAc_{3"}$-Rha'-$C_{8,10,10}$ (**4a**), Rha"-$OAc_{3"}$-Rha'-$C_{10,8,10}$ (**4b**), and Rha"-$OAc_{3"}$-Rha'-$C_{10,10,8}$ (**4c**) (150 MHz, CD_3OD).

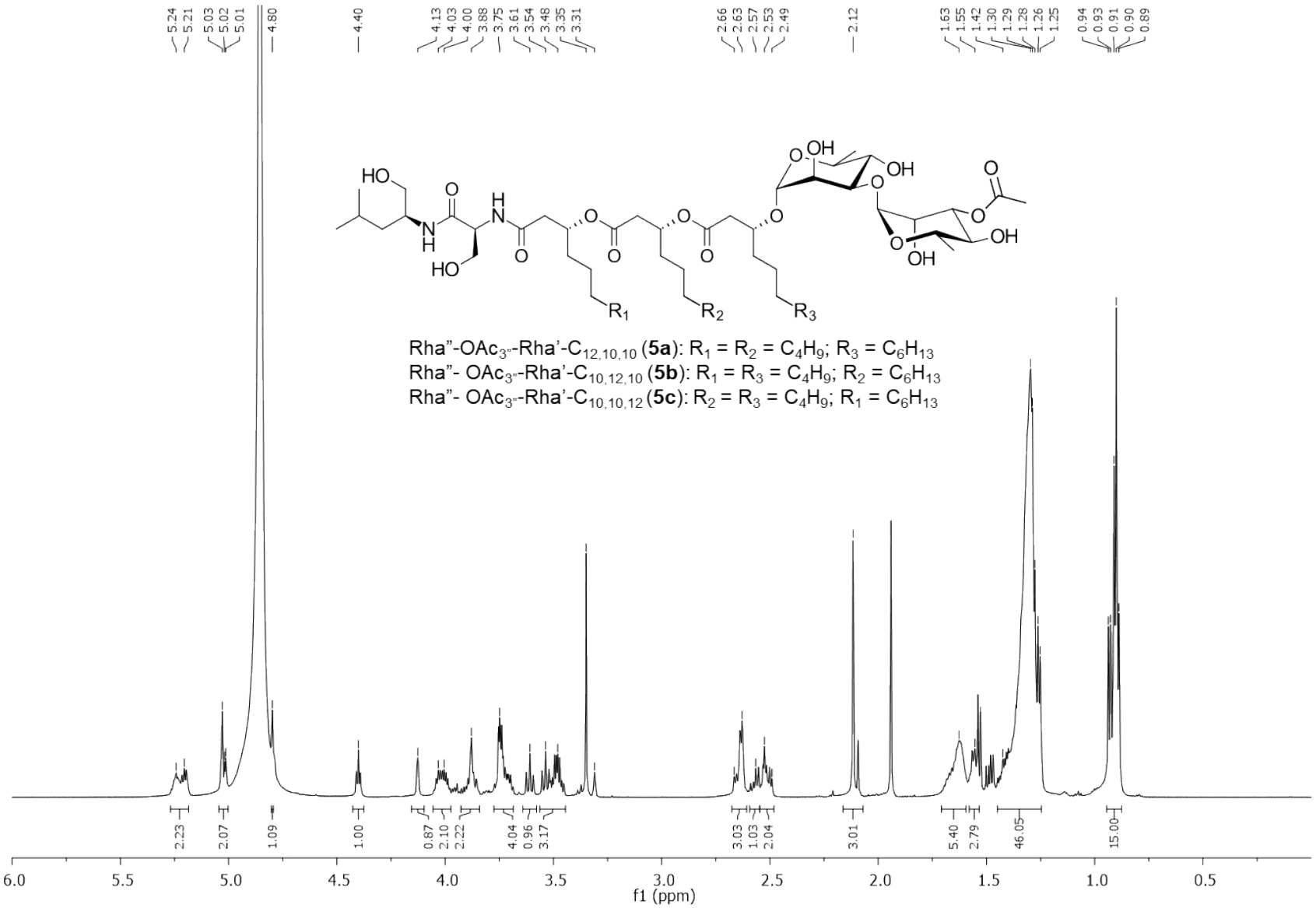


Fig. S21 ^{1}H NMR spectrum of Variovoracins Rha"-$OAc_{3"}$-Rha'-$C_{12,10,10}$ (**5a**), Rha"-$OAc_{3"}$-Rha'-$C_{10,12,10}$ (**5b**), and Rha"-$OAc_{3"}$-Rha'-$C_{10,10,12}$ (**5c**) (600 MHz, CD_3OD).

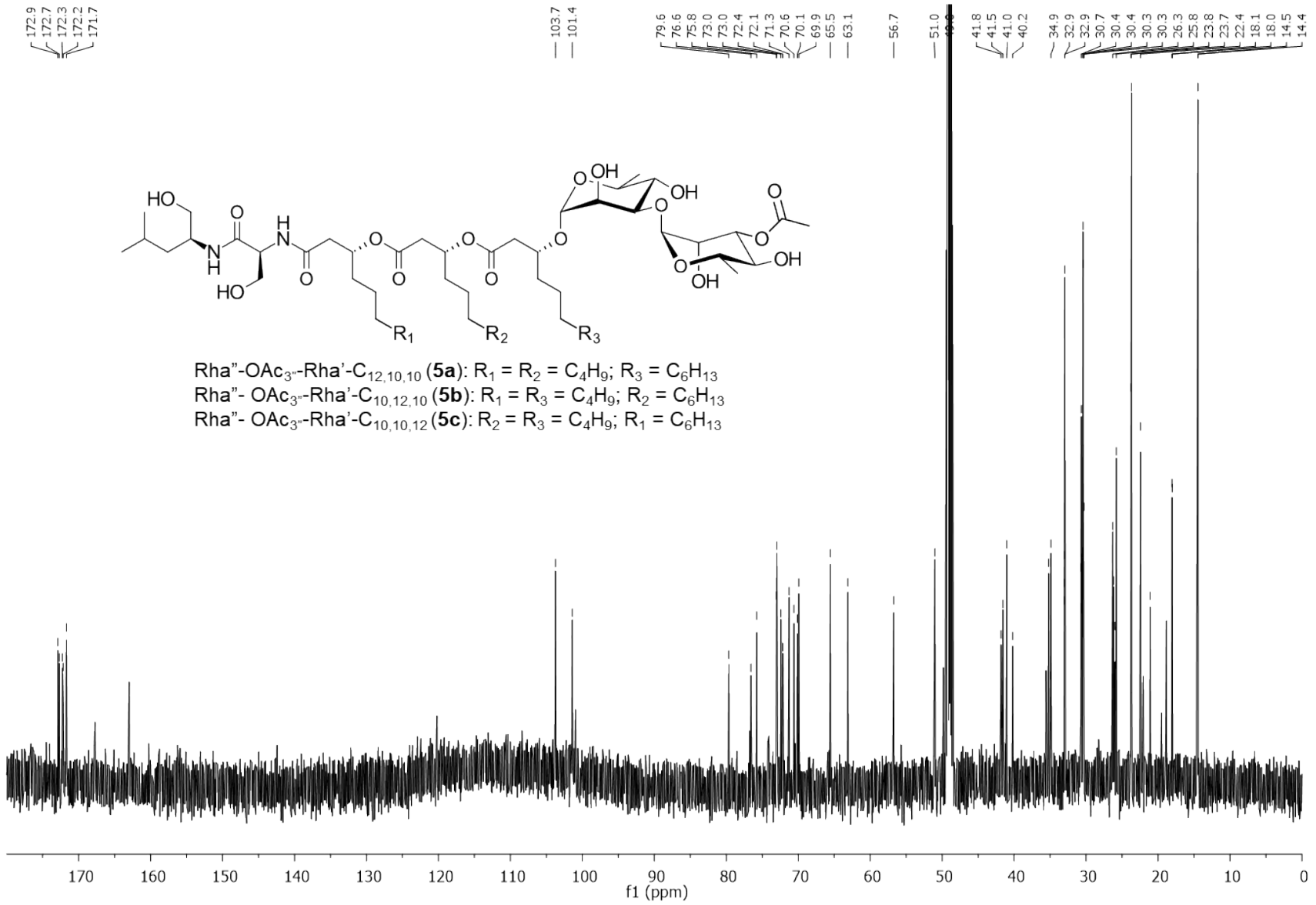


Fig. S22 ^{13}C NMR spectrum of Variovoracins Rha"-OAc$_{3"}$-Rha'-C$_{12,10,10}$ (**5a**), Rha"-OAc$_{3"}$-Rha'-C$_{10,12,10}$ (**5b**), and Rha"-OAc$_{3"}$-Rha'-C$_{10,10,12}$ (**5c**) (150 MHz, CD_3OD).

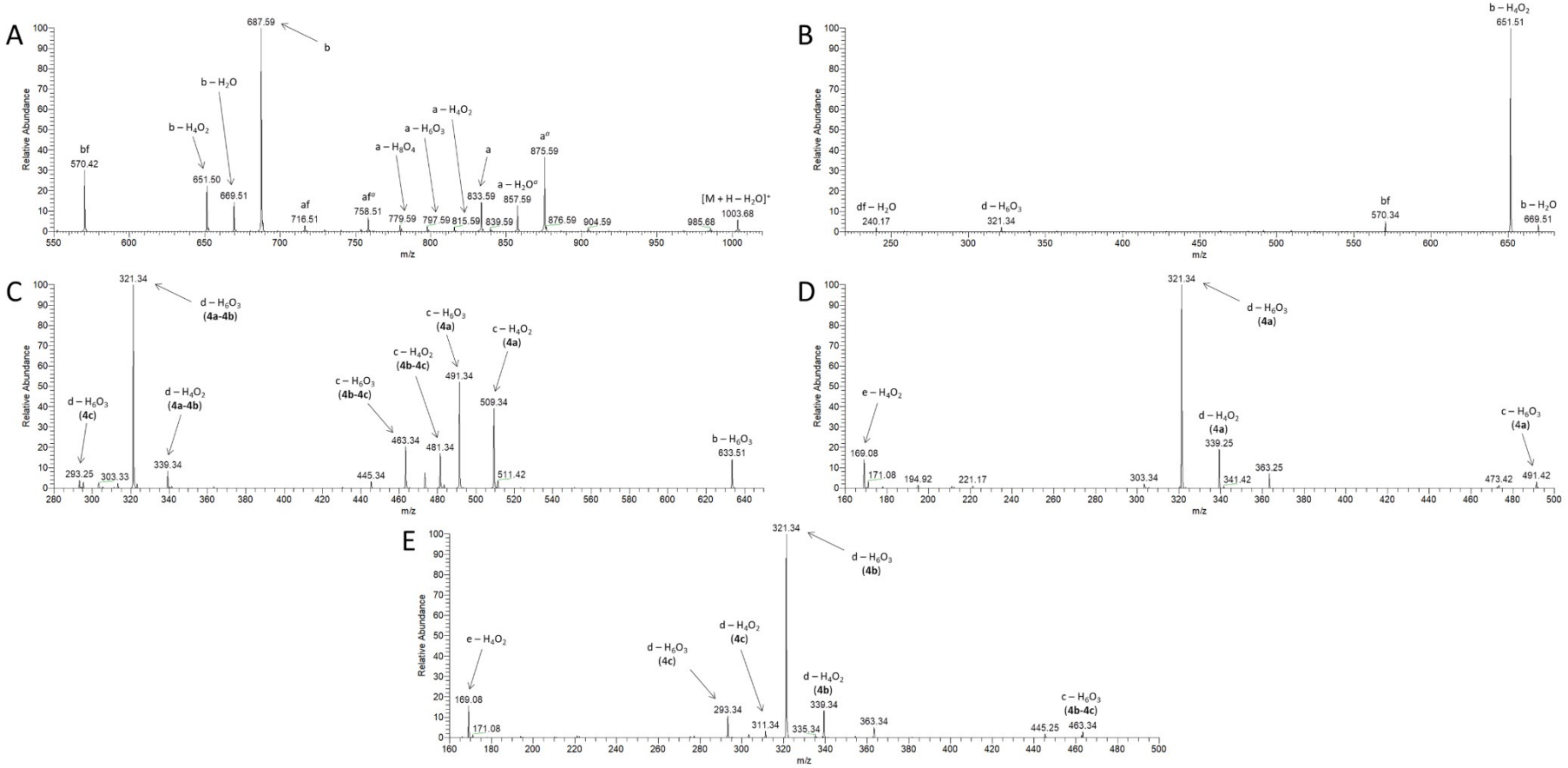


Fig. S23 Analysis of Variovoracins Rha"-OAc_3"-Rha'-$C_{8,10,10}$ (**4a**), Rha"-OAc_3"-Rha'-$C_{10,8,10}$ (**4b**), and Rha"-OAc_3"-Rha'-$C_{10,10,8}$ (**4c**) by ESI-MS/MS and assignment of observed fragment ions. (A) ESI-MS^2 scan of precursor ion *m/z* 1021.7 $[M + H]^+$; (B) ESI-MS^3 scan of fragment ion b *m/z* 687.6; (C) ESI-MS^4 scan of fragment ion b – H_4O_2 *m/z* 651.5 from Rha"-OAc_3"-Rha'-$C_{8,10,10}$ (**4a**); (D) ESI-MS^5 scan of fragment ion c – H_4O_2 *m/z* 509.3 from Rha"-OAc_3"-Rha'-$C_{8,10,10}$ (**4a**); (E) ESI-MS^5 scan of fragment ion c – H_4O_2 *m/z* 481.3 from Rha"-OAc_3"-Rha'-$C_{10,8,10}$ (**4b**) and Rha"-OAc_3"-Rha'-$C_{10,10,8}$ (**4c**).

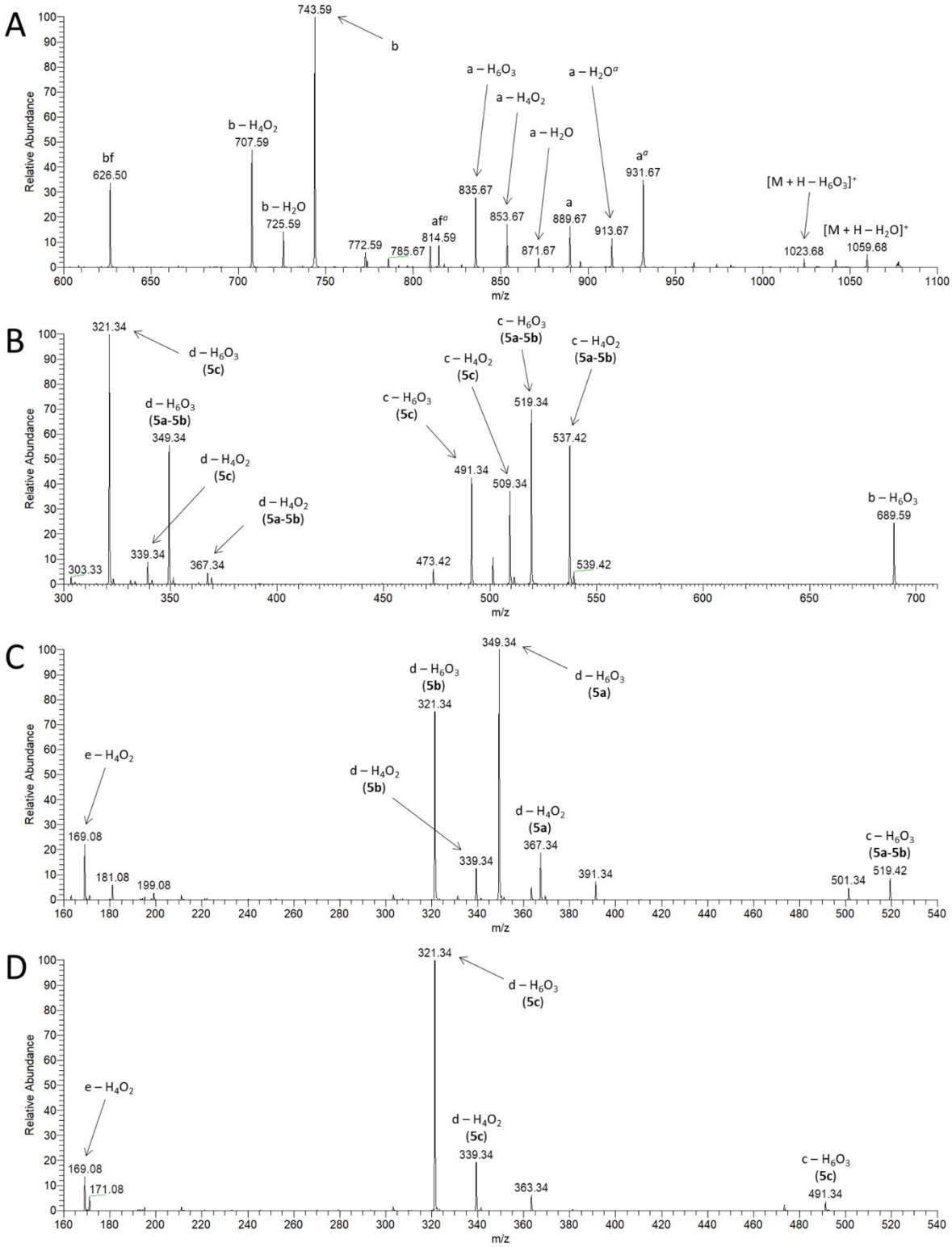


Fig. S24 Analysis of Variovoracins Rha"-OAc_3"-Rha'-$C_{12,10,10}$ (**5a**), Rha"-OAc_3"-Rha'-$C_{10,12,10}$ (**5b**), and Rha"-OAc_3"-Rha'-$C_{10,10,12}$ (**5c**) by ESI-MS/MS and assignment of observed fragment ions. (A) ESI-MS^2 scan of precursor ion *m/z* 1077.7 $[M + H]^+$; (B) ESI-MS^3 scan of fragment ion b – H_4O_2 *m/z* 707.6; (C) ESI-MS^4 scan of fragment ion c – H_4O_2 *m/z* 537.4 from Rha"-OAc_3"-Rha'-$C_{12,10,10}$ (**5a**) and Rha"-OAc_3"-Rha'-$C_{10,12,10}$ (**5b**); (D) ESI-MS^4 scan of fragment ion c – H_4O_2 *m/z* 509.4 from Rha"-OAc_3"-Rha'-$C_{10,10,12}$ (**5c**).

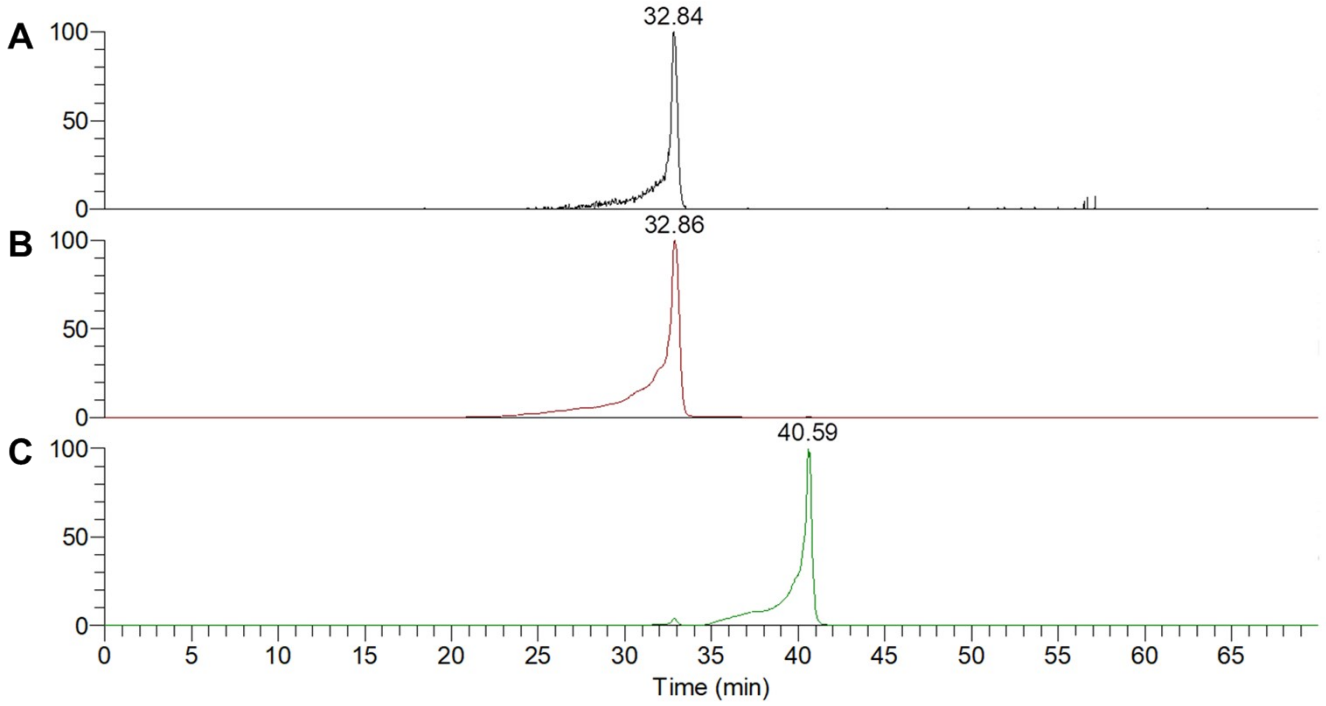


Fig. S25 Configuration assignment of the leucinol residue of variovoracins **1-5** from *V. paradoxus* RKNM0096 using Marfey's method. Extracted ion chromatograms (*m/z* 370.1721 $[M+H]^+$, ±5.0 ppm) represent products from derivatization reactions of (A) lipopeptide **8** from variovoracins **1-5**; (B) L-leucinol; (C) D-leucinol.

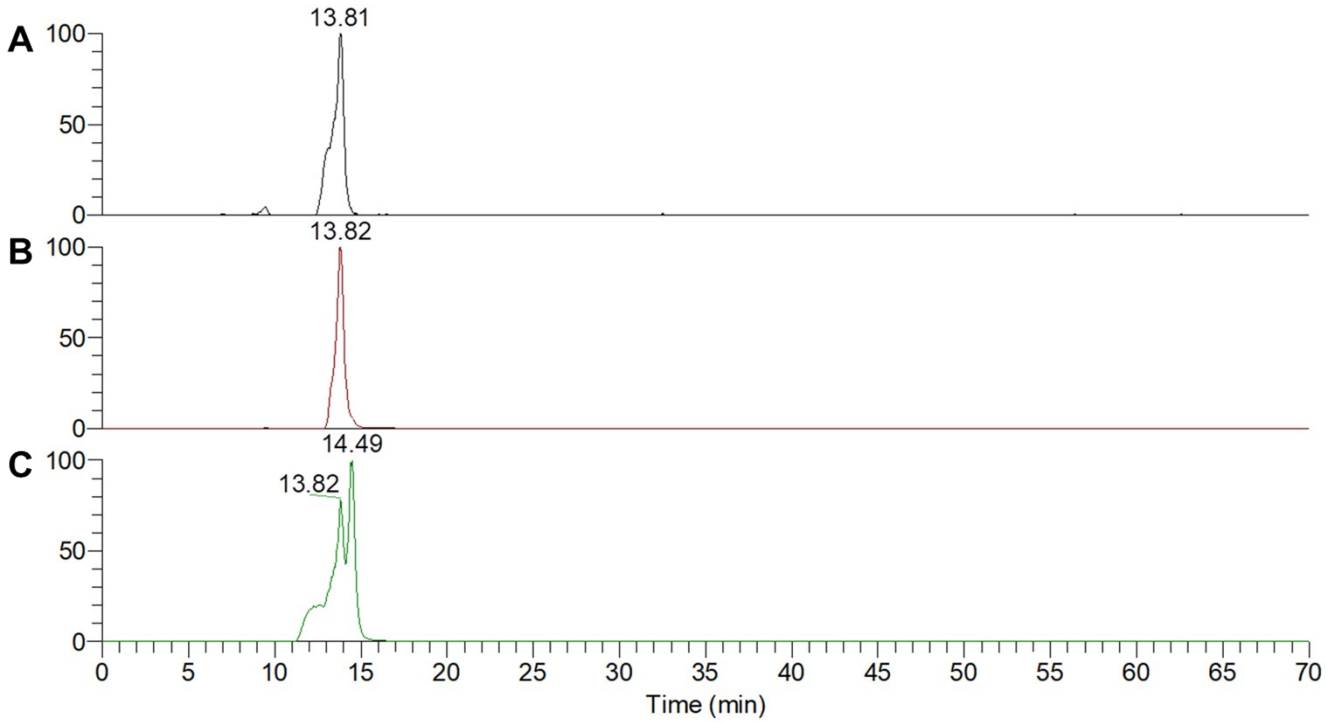


Fig. S26 Configuration assignment of the serine residue of variovoracins **1-5** from V. paradoxus RKNM0096 using Marfey's method. Extracted ion chromatograms (m/z 358.0993 [M+H]+, ±5.0 ppm) represent products from derivatization reactions of (A) lipopeptide **8** from variovoracins **1-5**; (B) L-serine; (C) DL-serine.

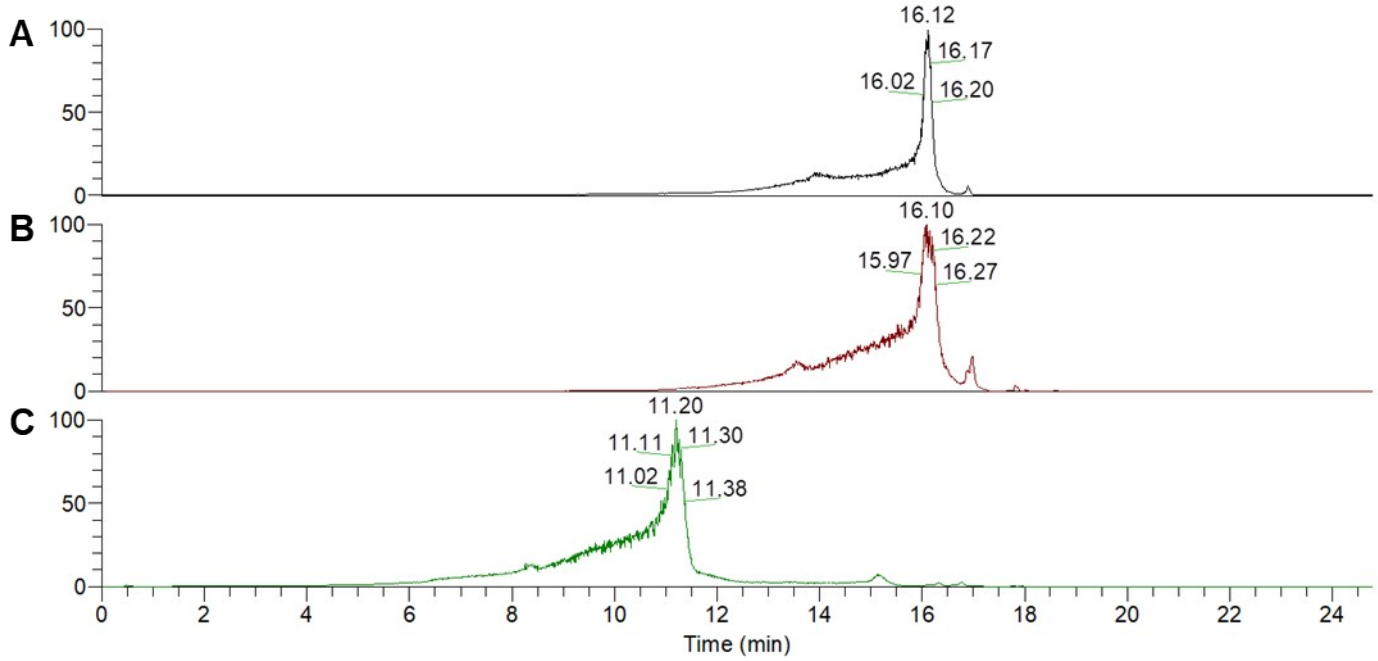


Fig. S27 Elucidation of the absolute configuration of the α-rhamnopyranose moieties of variovoracins **1-5** isolated from *V. paradoxus* RKNM0096 using a RP-HPLC method previously reported by Tanaka et al.[21] Reactions were performed on glycolipid **9** (**Scheme 1**) and the thiazolidine derivatives were analyzed by UHPLC-HRMS. Extracted ion chromatograms (*m*/*z* 431.1305 $[M + H]^+$, ±5.0 ppm) of the anticipated thiazolidine derivatives from: (A) glycolipid **9** from *V. paradoxus* (RKNM0096). (B) L-rhamnose standard. (C) D-rhamnose standard.

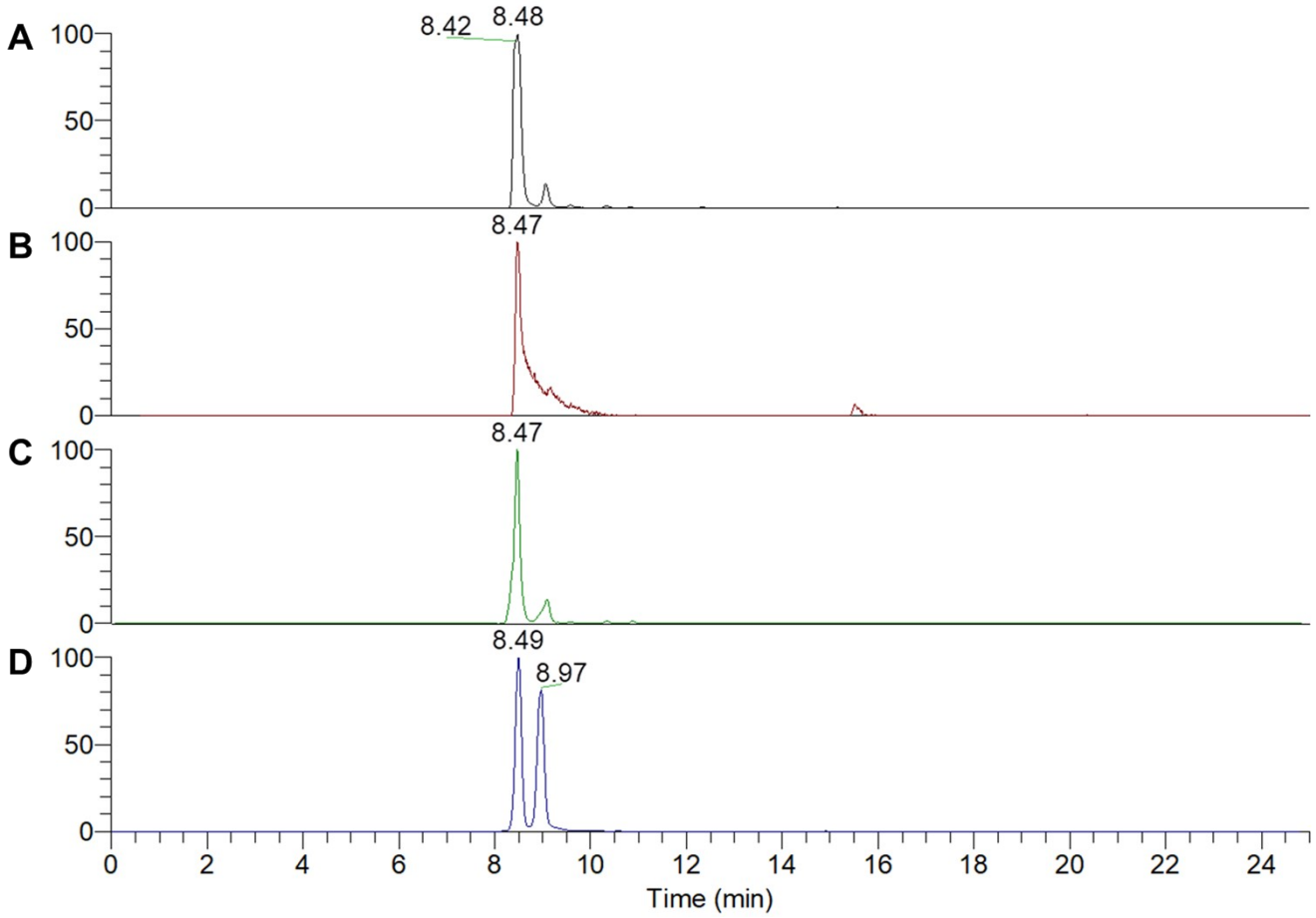


Fig. S28 Configuration assignment of the 3-hydroxydecanoic acid moieties of variovoracins produced by *V. paradoxus* RKNM0096 and *J. agaricidamnosum* DSM 9628. Extracted ion chromatograms (*m/z* 423.2142 $[M+Na]^+$, ±5.0 ppm) representing products of naproxen chloride derivatization reactions from (A) 3-hydroxydecanoic acid derived from variovoracins (**1**-**5**) produced by *V. paradoxus* RKNM0096; (B) 3-hydroxydecanoic acid derived from variovoracins (**6**-**7**) produced by *J. agaricidamnosum* DSM 9628; (C) (*R*)-3-hydroxydecanoic acid; and (D) (*R/S*)-3-hydroxydecanoic acid.

Table S4 Biosynthetic gene clusters (BGCs) in the *V. paradoxus* RKNM0096 genome sequence (NZ_CP046508.1) identified by antiSmash (v7.1.0).[13]

Anti-SMASH Region	BGC Type	Genome Coordinates
1	Arylpolyene	299,970 - 341,175
2	Hydrogen-cyanide	377,412 - 390,307
3	Redox-cofactor	753,781 - 775,949
4	Resorcinol, arylpolyene	2,429,037 - 2,475,063
5	Terpene	3,100,809 - 3,122,559
6	NRPS, T1PKS	3,565,335 - 3,617,467
7	Resorcinol	4,135,551 - 4,117,467
8	NRPS, TIPKS	5,022,601 - 5,112,765
9	NRPS, ectoine	5,337,932 - 5,387,782

Table S5 Annotation of the region of the *Variovorax* RKNM0096 genome corresponding to biosynthetic gene cluster 9 identified by antiSMASH (V7.1.0).[13]

	Varivoracin	Gene	Protein	BlastP Closest Match		
Locus tag	Gene	length (bp)	length (aa)	Description	Accession no.	Identity (%)
GNX71_24850	*orf1*	2769	922	*Variovorax boronicumulans* magnesium-translocating P-type ATPase	WP_307648753.1	98.59
GNX71_24845	*orf2*	759	252	*Variovorax* sp. ZS18.2.2 hypothetical protein	WP_257861347.1	92.46
GNX71_24840	*orf3*	210	69	*Variovorax* sp. ZS18.2.2 hypothetical protein	WP_257861346.1	92.31
GNX71_24835	*orf4*	384	127	*Variovorax* sp. ZS18.2.2 ectoine synthase	WP_257861344.1	96.85
GNX71_24830	*orf5*	1233	410	*Variovorax* sp. ZS18.2.2 MFS transporter	WP_257861343.1	95.61
GNX71_24825	*orf6*	1572	523	*Variovorax* sp. ZS18.2.2 class I adenylate-forming enzyme family protein	WP_257861342.1	92.73
GNX71_24820	*orf7*	987	328	*Variovorax* sp. ZS18.2.2 class I SAM-dependent methyltransferase	WP_257861340.1	95.73
GNX71_24815	*orf8*	1278	425	*Variovorax* sp. ZS18.2.2 acyl-CoA dehydrogenase family protein	WP_257861339.1	94.54
GNX71_24810	*orf9*	471	156	*Variovorax* sp. ZS18.2.2 low molecular weight protein-tyrosine-phosphatase	WP_257861338.1	93.55
GNX71_24805	*orf10*	420	139	*Variovorax boronicumulans* acyl-CoA thioesterase	WP_307686539.1	97.84
GNX71_24800	*orf11*	1302	433	*Variovorax* sp. IB41 TRAP transporter permease	WP_198786829.1	99.31
GNX71_24795	*orf12*	501	166	*Variovorax paradoxus* TRAP transporter small permease	WP_015866690.1	100.00
GNX71_24790	*orf13*	1029	342	*Variovorax* sp. 770b2 TRAP transporter substrate-binding protein	WP_256219280.1	98.11
GNX71_24785	*orf14*	996	331	*Variovorax boronicumulans* tripartite tricarboxylate transporter substrate binding protein	WP_307648759.1	96.08
GNX71_24780	*orf15*	1617	538	*Variovorax* sp. ZS18.2.2 indolepyruvate oxidoreductase beta subunit	WP_257861331.1	96.65
GNX71_24775	*orf16*	2160	719	*Variovorax* sp. ZS18.2.2 indolepyruvate ferredoxin oxidoreductase alpha subunit	WP_206174887.1	98.33
GNX71_24770	*orf17*	651	216	*Variovorax* sp. DXTD-1 c-type cytochrome	RST49610.1	91.62
GNX71_24765	*orf18*	1041	346	*Variovorax boronicumulans* PQQ-dependent sugar dehydrogenase	WP_176659682.1	98.27
GNX71_24760	*rlpA*	915	304	*Variovorax paradoxus* EPS LysR family transcriptional regulator	ADU38687.1	98.68
GNX71_24755	*rlpB*	7476	2491	*Variovorax* sp. ZS18.2.2 non-ribosomal peptide synthetase	WP_257861326.1	96.59
GNX71_24750	*rlpC*	1320	439	*Variovorax boronicumulans* nucleotide disphospho-sugar-binding domain-containing protein	WP_307648765.1	97.49
GNX71_24745	*rlpD*	213	70	*Variovorax* sp. 770b2 MbtH family protein	WP_093437515.1	98.44
GNX71_24740	*rlpE*	969	322	*Variovorax boronicumulans* glycosyltransferase family 2 protein	WP_307606394.1	92.14
GNX71_24735	*rlpF*	1260	419	*Variovorax* sp. ZS18.2.2 TCR/Tet family MFS transporter	WP_257861322.1	95.47
GNX71_24730	*orf19*	1611	536	*Variovorax paradoxus* xanthine dehydrogenase small subunit	WP_013542891.1	95.19
GNX71_24725	*orf20*	2469	822	*Variovorax paradoxus* xanthine dehydrogenase molybdopterin binding subunit	WP_042582286.1	98.65
GNX71_24720	*orf21*	993	330	*Variovorax paradoxus* LysR family transcriptional regulator	WP_042582285.1	98.79
GNX71_24715	*orf22*	1401	466	*Variovorax paradoxus* NCS2 family permease	WP_013542888.1	99.36
GNX71_24710	*orf23*	1536	511	*Variovorax boronicumulans* ABC transporter ATP-binding protein	WP_042582283.1	98.24
GNX71_24705	*orf24*	1143	380	*Variovorax* sp. KBW07 ABC transporter permease	WP_124465102.1	99.47
GNX71_24700	*orf25*	921	306	*Variovorax* sp. KBW07 ABC transporter permease	WP_124460136.1	98.69
GNX71_24695	*orf26*	1173	390	*Variovorax boronicumulans* BMP family ABC transporter substrate-binding protein	WP_280808640.1	99.74
GNX71_24690	*orf27*	381	126	*Variovorax paradoxus* hypothetical protein	WP_042582279.1	93.60
GNX71_24685	*orf28*	819	272	*Variovorax* sp. ZS18.2.2 hypothetical protein	WP_257861312.1	84.73
GNX71_24680	*orf29*	750	294	*Variovorax boronicumulans* EAL domain-containing protein	WP_176659697.1	96.79
GNX71_24675	*orf30*	1887	628	*Variovorax boronicumulans* ATPase, T2SS/TP4/T4SS family	WP_307686526.1	96.34

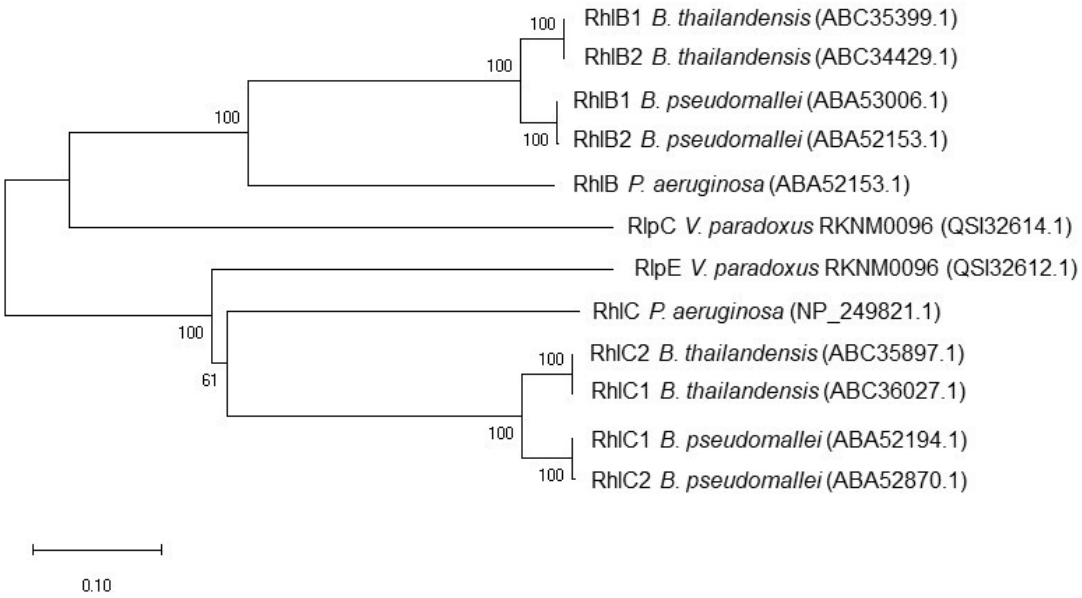


Fig. S29 Evolutionary history of glycosyltransferases from the variovoracin BGC (RlpC and RlpE) and rhamnolipid (RhlB and RhlC) BGCs from *Burkholderia thailandensis, B. pseudomallei* and *Pseudomonas aeruginosa*.[15] The accession number of the protein sequences used in the analysis are provided in brackets. The phylogeny shown was inferred using the Neighbor-Joining method.[22] The optimal tree with the sum of branch length = 2.19434862 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [23] The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method and are in the units of the number of amino acid differences per site.[24] The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). All positions containing gaps and missing data were eliminated (complete deletion option). There were a total of 308 positions in the final dataset. Phylogenies inferred using the UPGMA[25] and maximum likelihood[26] methods returned the same overall tree topology (data not shown). Evolutionary analyses were conducted in MEGA X.[16]

Fig. S30 Comparison of the variovoracin BGC to selected glycolipopeptide BGCs from 23 bacteria representing taxa identified in BlastP[17, 18] searches using RlpC as a query and exhibiting ≥ 50% identity to RlpC. Clusters were aligned using cblaster[19] and the BGC alignment was graphed using clinker.[20] Accession numbers for the nucleotide sequences used in the alignment are provide under the organism's name. Conserved genes are color coded. The NRPS module composition and domain substrate specificity were predicted by antiSMASH (v.7.1.0)[13] and are shown within the NRPS open reading frame arrows: C = condensation domain, A = adenylation domain, T = thiolation/peptidyl carrier domain, R = reductase domain. Predicted A-domain substrates are indicated by a subscript single-letter amino acid code. Core genes involved in variovoracin biosynthesis are labeled at the top of the figure. Taxonomic class of bacteria indicated by the colored circle adjacent to the strain name: Betaproteobacteria – green, Alphaproteobacteria – black, Gammaproteobacteria – orange.

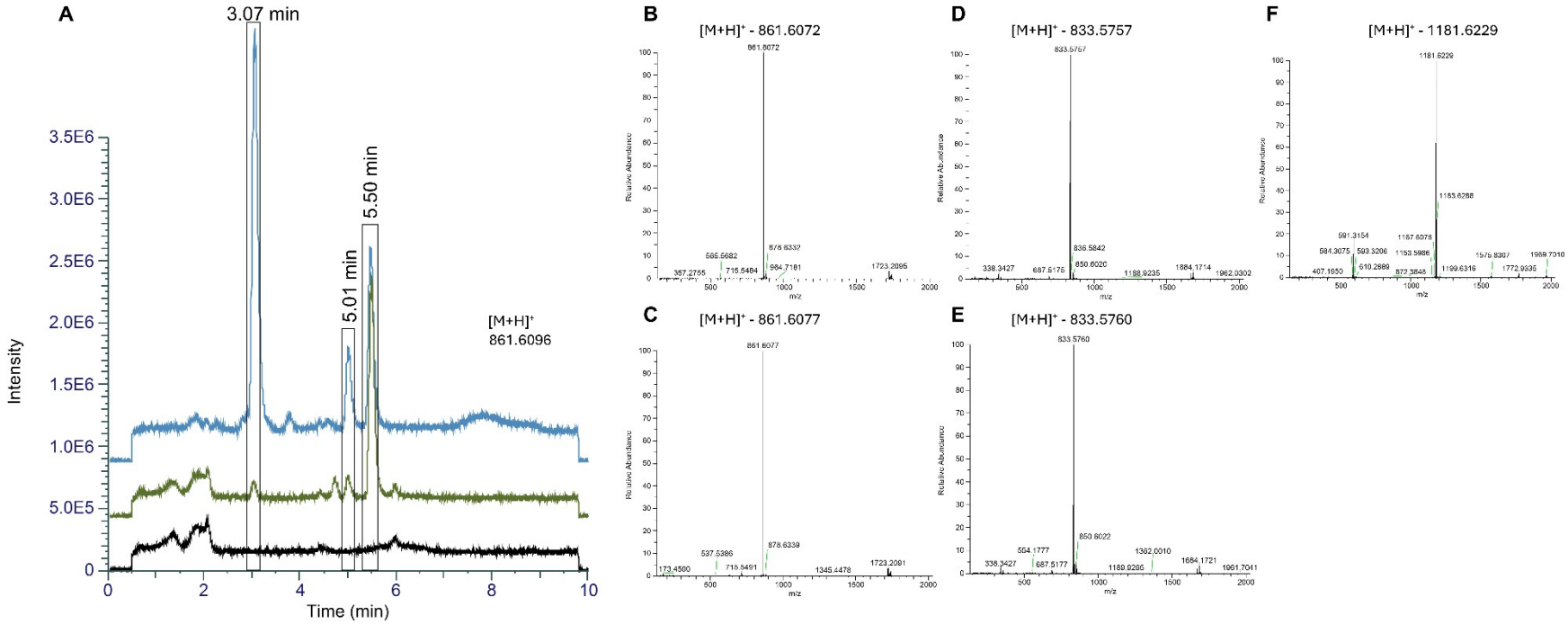


Fig. S31 UHPLC-HRMS analysis of fermentation extracts of *J. agaricidamnosum* DSM 9628 and *V. paradoxus* DSM 21786. Total ion chromatograms [A] of *J. agaricidamnosum* DSM 9628 (blue), *V. paradoxus* DSM 21786 (green) and the media blank (black) extracts. High-resolution mass spectra for peaks eluting at 5.50 min ([B] DSM 9628, [C] DSM 21786), 5.01 min ([D] DSM 9628, [E] DSM 21786) and the jagaricin peak eluting at 3.07 min [F]. The jagarcin-related ions were not observed in the DSM 21786 extract.

Table S6. NMR spectroscopic data (600 MHz, CD_3OD) for Variovoracin Rha'-$C_{10,10,10}$ (**6**).

Position	δ_C, type	δ_H, multiplicity (*J* in Hz)	COSY[a]	HMBC (H→C)[b]
Leucinol				
1aL	65.5, CH_2	3.49, dd (11.1, 5.2)	1bL, 2L	2L, 3L
1bL		3.46, dd (11.1, 5.8)	1aL, 2L	2L, 3L
2L	51.0, CH	4.03 – 3.97, m	1L, 3L	1L, 3L, 1S
3aL	41.0, CH_2	1.45 – 1.39, m	2L, 4L	1L, 2L, 4L, 5L, 6L
3bL		1.37 – 1.31, m	2L	1L, 2L, 4L, 5L, 6L
4L	25.82, CH	1.71 – 1.64, m	3aL, 5L, 6L	2L, 3L, 5L, 6L
5L	23.78, CH_3	0.93, d (6.7)	4L	3L, 4L, 6L
6L	22.4, CH_3	0.92 – 0.90, m	4L	3L, 4L
Serine				
1S	172.2, C			
2S	56.8, CH	4.40, t (5.7)	3S	1S, 3S, 1A
3aS	63.1, CH_2	3.75, d (3.9)	2S	1S, 2S
3bS		3.74, d (3.9)	2S	1S, 2S
Decanoic Acid (A)				
1A	171.7, C			
2aA	41.9, CH_2	2.60 – 2.54, m	3A	1A, 3A, 4A
2bA		2.53 – 2.48, m	3A	3A
3A	73.0, CH	5.28 – 5.22, m	2aA, 2bA, 4A	1A, 2A, 4A, 5A
4A	35.4, CH_2	1.64 – 1.58, m		
5A	26.4, CH_2	1.39 – 1.30, m		
6A, 7A	n.a.* CH_2	1.38 – 1.27, m		
8A	33.02[c], CH_2	1.34 – 1.25, m		
9A	23.74[c], CH_2	1.36 – 1.28, m		
10A	14.45[c], CH_3	0.92 – 0.89, m	9A	8A, 9A
Decanoic Acid (B)				
1B	172.7, C			
2aB	40.2, CH_2	2.65, dd (15.4, 5.6)	3B	1B, 3B, 4B
2bB		2.63 – 2.58, m	3B	3B
3B	72.1, CH	5.24 – 5.18, m	2aB, 2bB, 4B	1B, 2B, 4B, 5B
4B	34.9, CH_2	1.68 – 1.61, m		
5B	26.3, CH_2	1.39 – 1.30, m		
6B, 7B	n.a.* CH_2	1.38 – 1.27, m		
8B	33.02[c], CH_2	1.34 – 1.25, m		
9B	23.74[c], CH_2	1.36 – 1.28, m		
10B	14.45[c], CH_3	0.92 – 0.89, m	9B	8B, 9B
Decanoic Acid (C)				
1C	172.4, C			
2aC	41.3, CH_2	2.61 – 2.57, m	2bC, 3C	3C, 4C
2bC		2.53 – 2.47, m	2aC, 3C	3C
3C	75.5, CH	4.09 – 4.04, m	2aC, 3bC, 4C	1C, 2C, 4C, 5C, 1'
4C	34.3, CH_2	1.59 – 1.53, m		
5C	25.9, CH_2	1.39 – 1.30, m		
6C, 7C	n.a.* CH_2	1.38 – 1.27, m		
8C	32.97[c], CH_2	1.34 – 1.25, m		
9C	23.72[c], CH_2	1.36 – 1.28, m		
10C	14.47[c], CH_3	0.92 – 0.89, m	9C	8C, 9C
α-rhamnose (I)				
1'	100.7, CH	4.81, d (1.6)	2'	5', 3C
2'	72.6, CH	3.76, dd (1.6, 3.4)	1', 3'	1', 4'
3'	72.3, CH	3.65, dd (9.5, 3.4)	2', 4'	2', 5', 4'
4'	73.9, CH	3.38, app. t (9.5)	3', 5'	5', 6'
5'	70.3, CH	3.68, dq (9.5, 6.2)	4', 6'	1', 6'
6'	18.0, CH_3	1.26, d (6.2)	5'	4', 5'

[a] COSY correlations are from proton(s) stated to the indicated proton. [b] HMBC correlations (H→C) are from proton(s) stated to the indicated carbon. [c] Assignment of ^{13}C NMR resonances is ambiguous due to signal overlap. * The following overlapping ^{13}C NMR resonances could not be assigned (n.a.) unambiguously: 30.8, 30.50, 30.45, 30.40, 30.34.

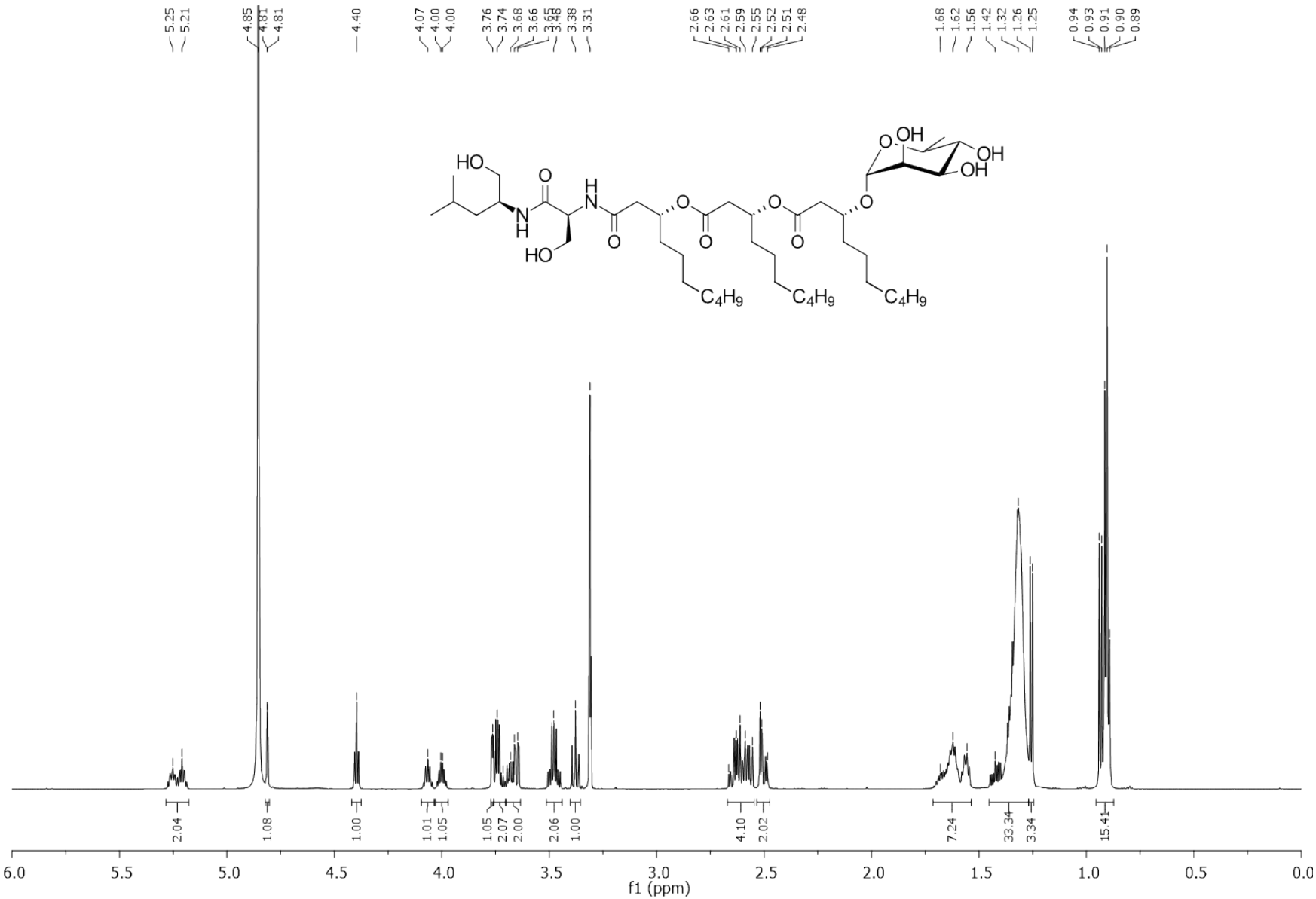


Fig. S32 ^{1}H NMR spectrum of Variovoracin Rha'-$C_{10,10,10}$ (**6**) (600 MHz, CD_3OD).

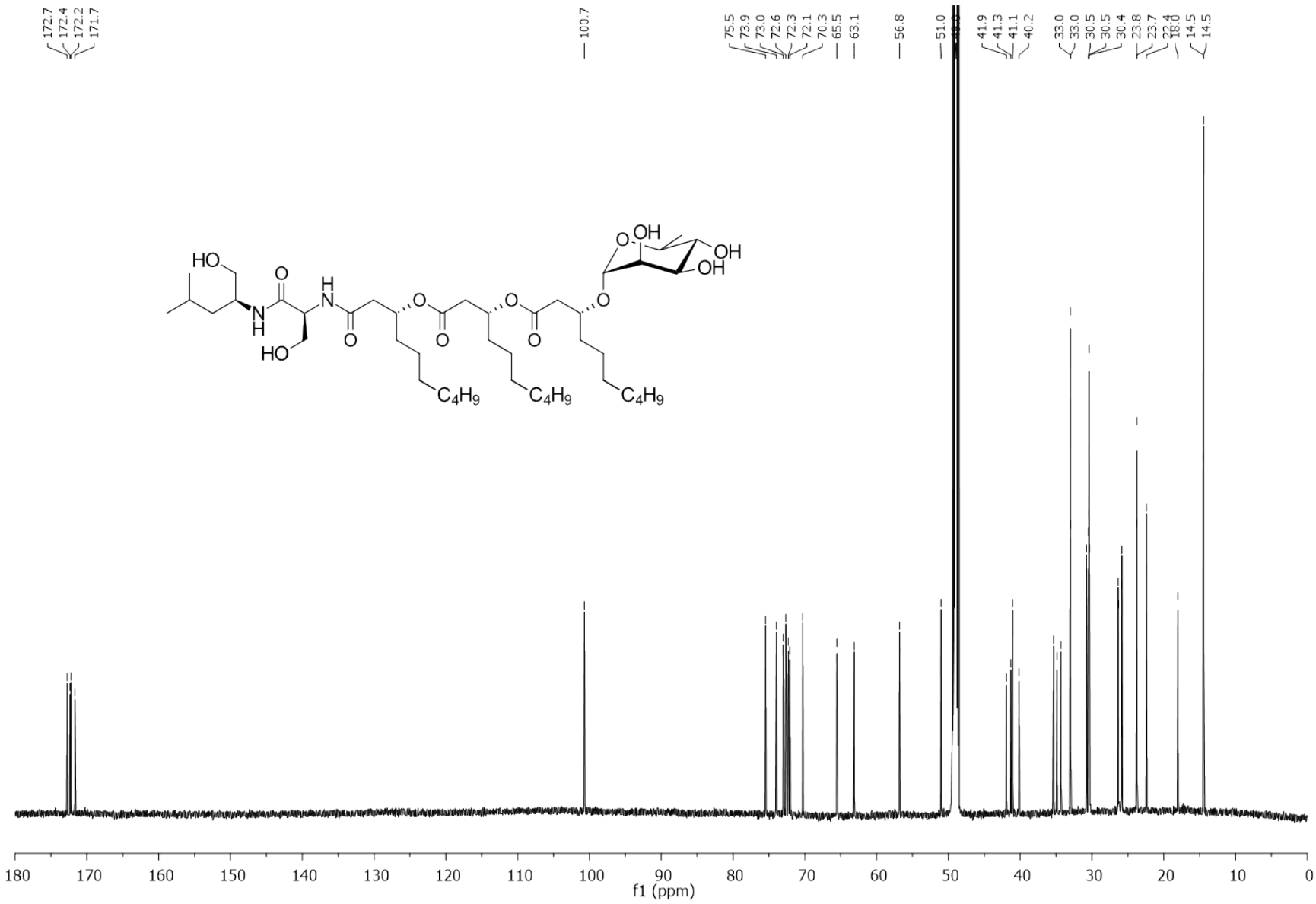


Fig. S33 ^{13}C NMR spectrum of Variovoracin Rha'-$C_{10,10,10}$ (**6**) (150 MHz, CD_3OD).

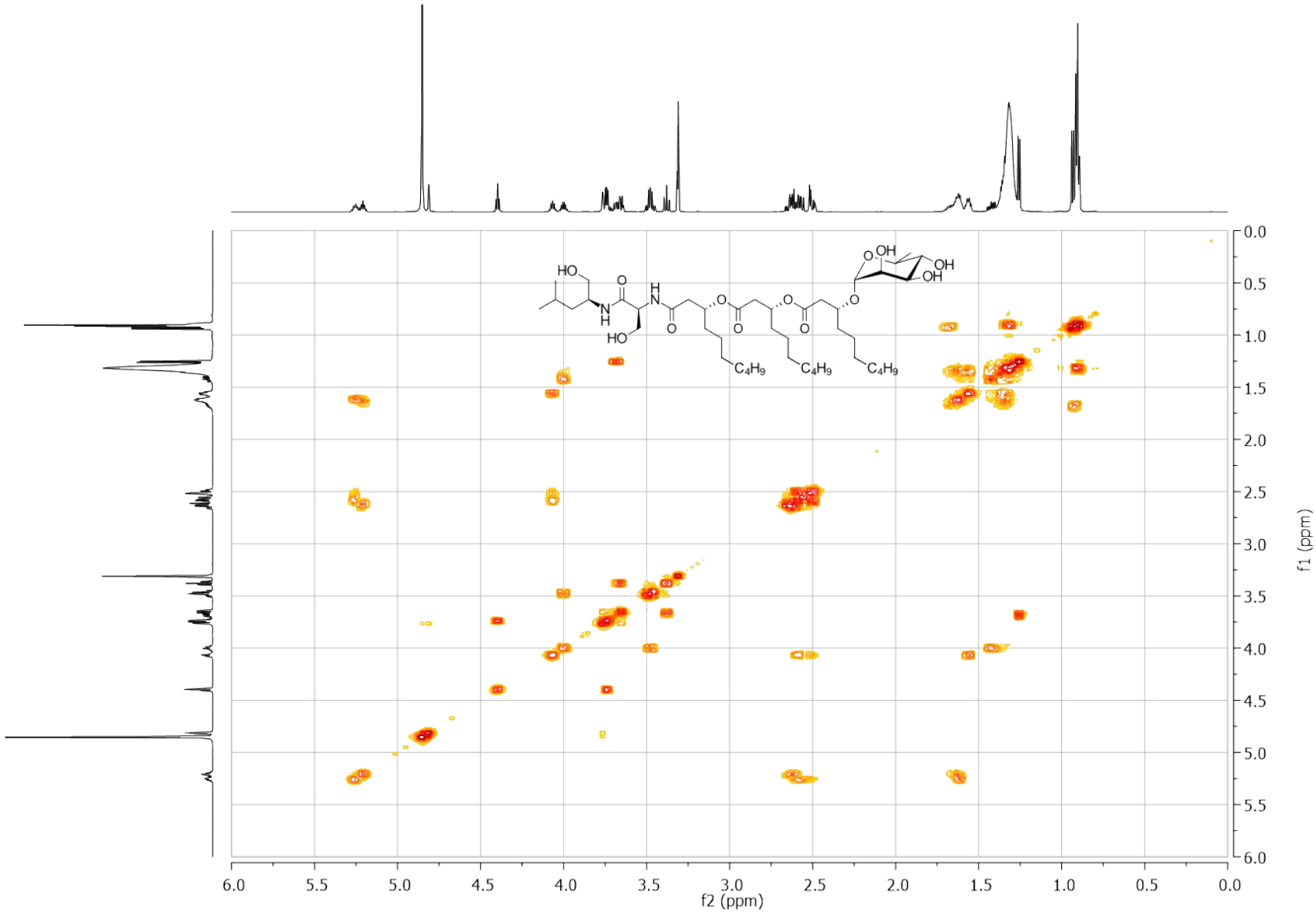


Fig. S34 COSY spectrum of Variovoracin Rha'-$C_{10,10,10}$ (**6**).

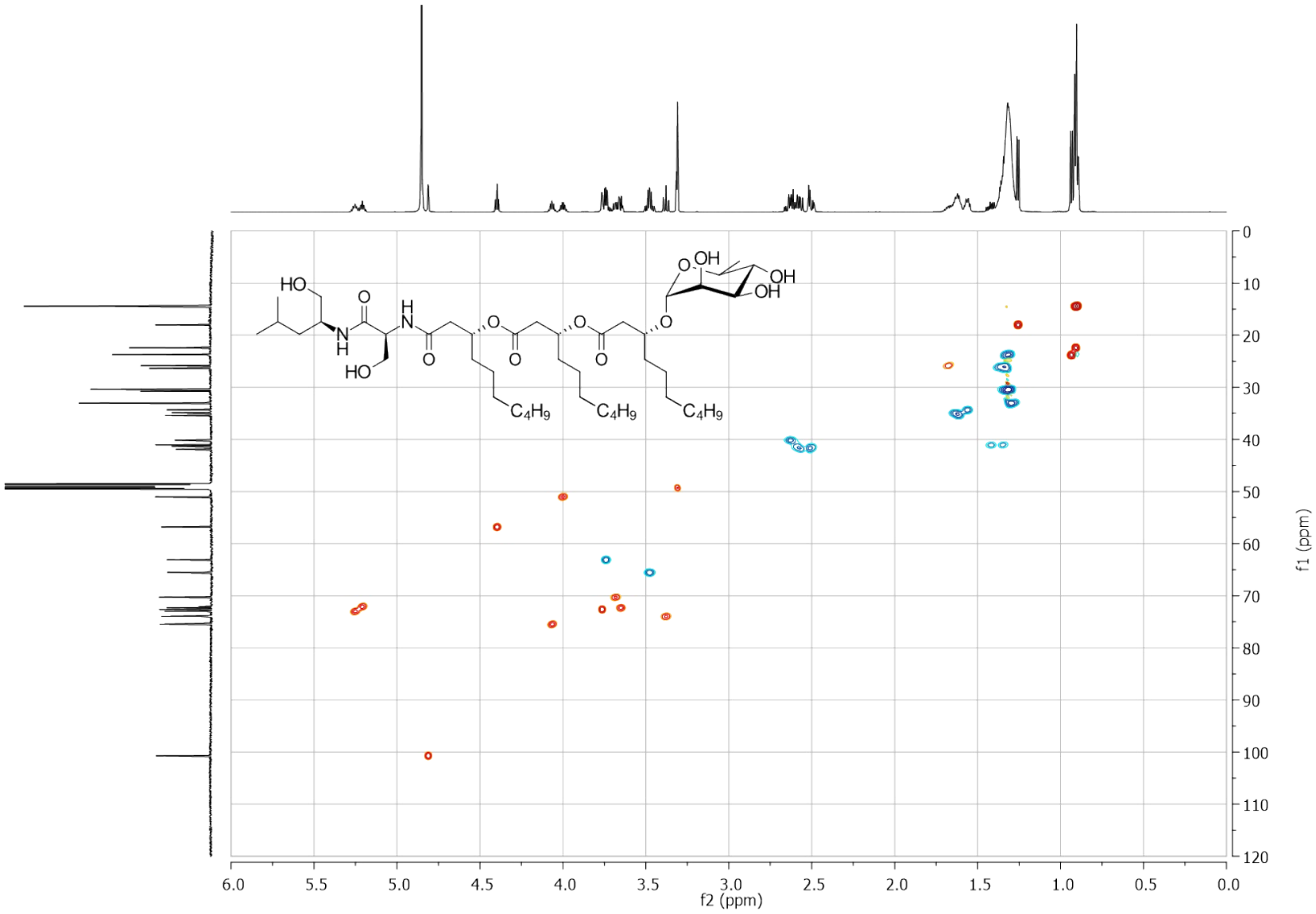


Fig. S35 HSQC spectrum of Variovoracin Rha'-$C_{10,10,10}$ (**6**).

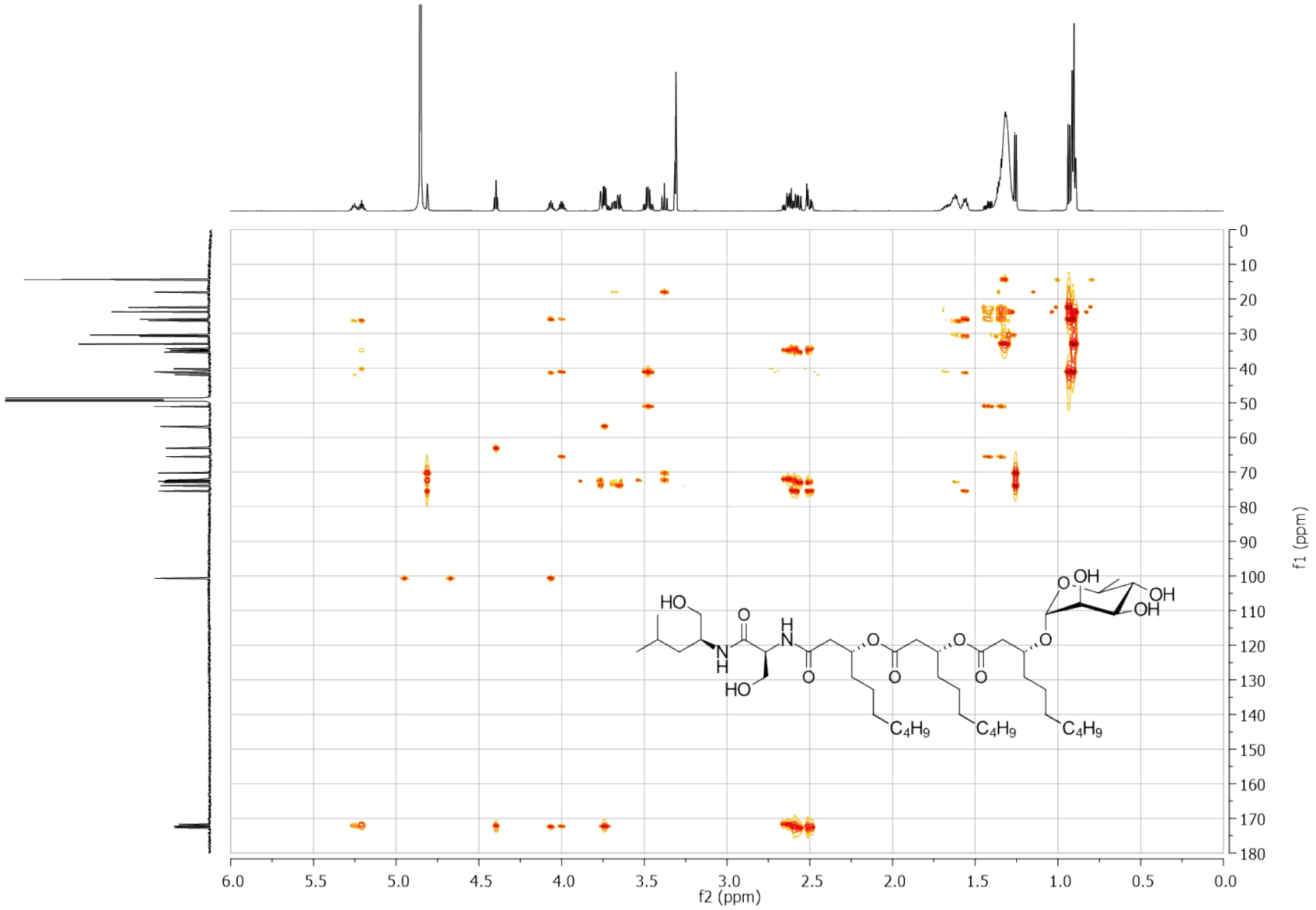


Fig. S36 HMBC spectrum of Variovoracin Rha'-$C_{10,10,10}$ (**6**).

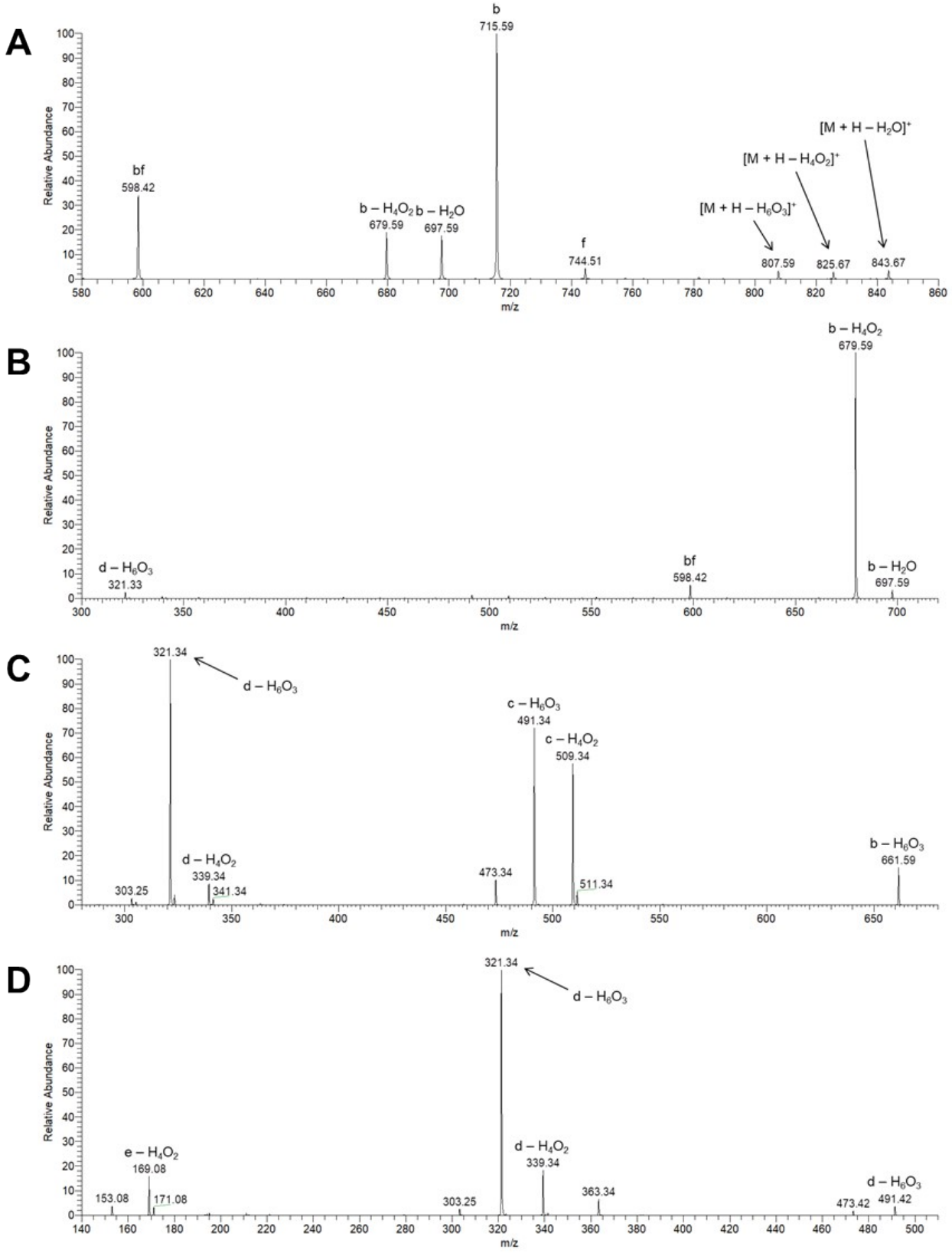


Fig. S37 Analysis of Variovoracin Rha'-$C_{10,10,10}$ (**6**) by ESI-MS/MS and assignment of observed fragment ions. (A) ESI-MS^2 scan of precursor ion *m/z* 861.67 $[M + H]^+$; (B) ESI-MS^3 scan of fragment ion b *m/z* 715.6; (C) ESI-MS^4 scan of fragment ion b – H_4O_2 *m/z* 679.6; (D) ESI-MS^5 scan of fragment ion c – H_4O_2 *m/z* 509.3.

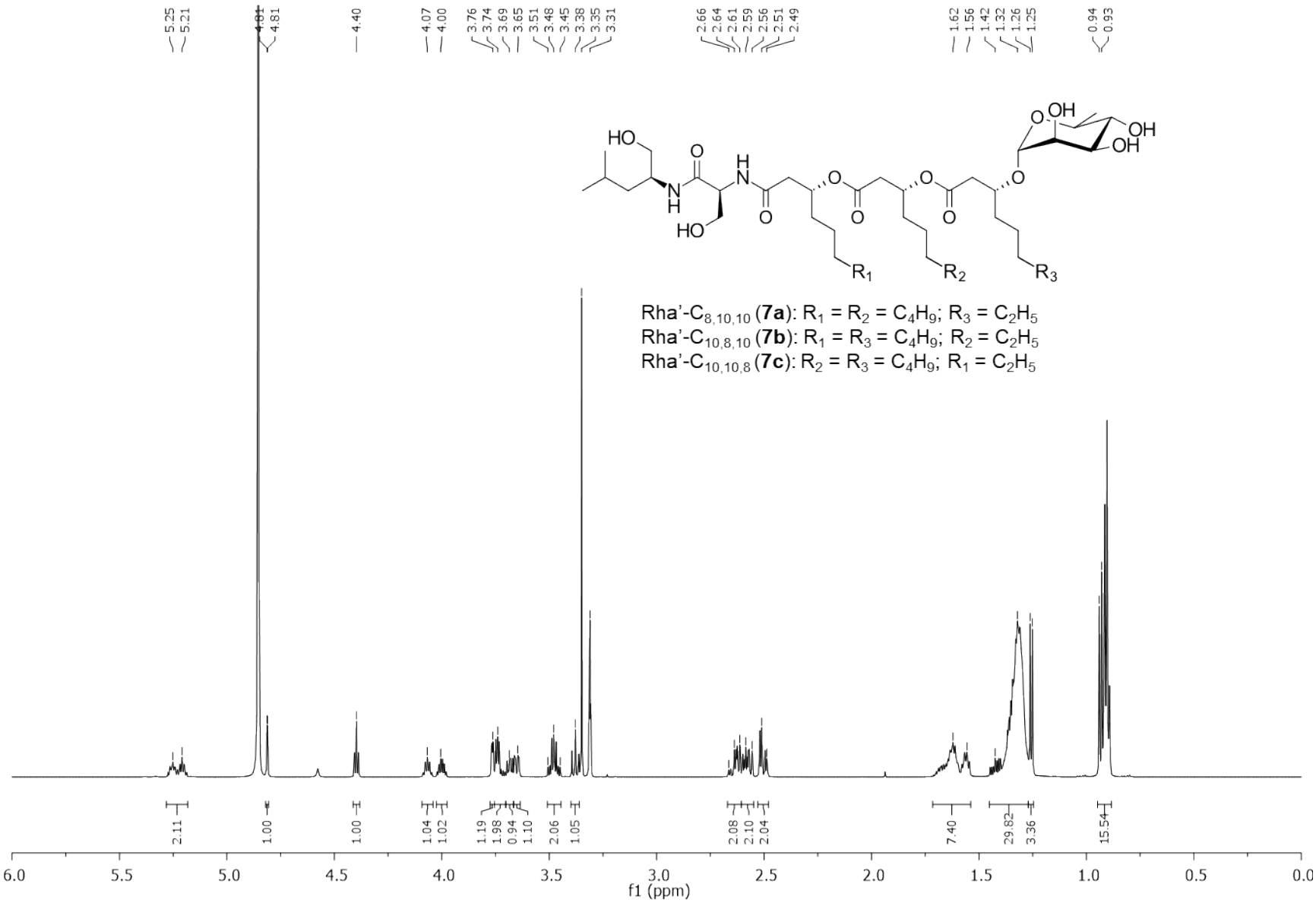


Fig. S38 ^{1}H NMR spectrum of Variovoracins Rha'-$C_{8,10,10}$ (**7a**), Rha"-Rha'-$C_{10,8,10}$ (**7b**), and Rha"-Rha'-$C_{10,10,8}$ (**7c**) (600 MHz, CD_3OD).

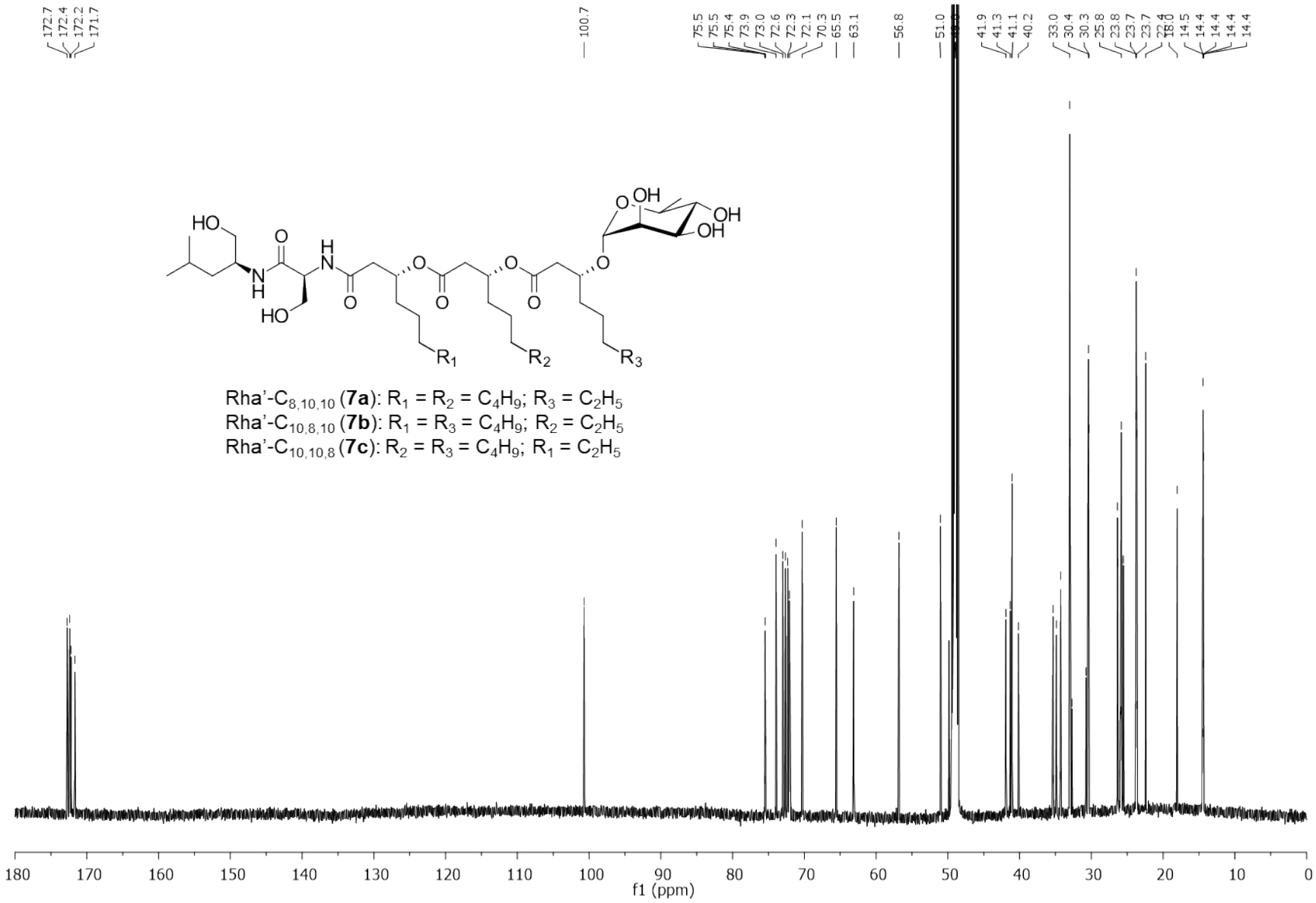


Fig. S39 ^{13}C NMR spectrum of Variovoracins Rha'-$C_{8,10,10}$ (**7a**), Rha"-Rha'-$C_{10,8,10}$ (**7b**), and Rha"-Rha'-$C_{10,10,8}$ (**7c**) (150 MHz, CD_3OD).

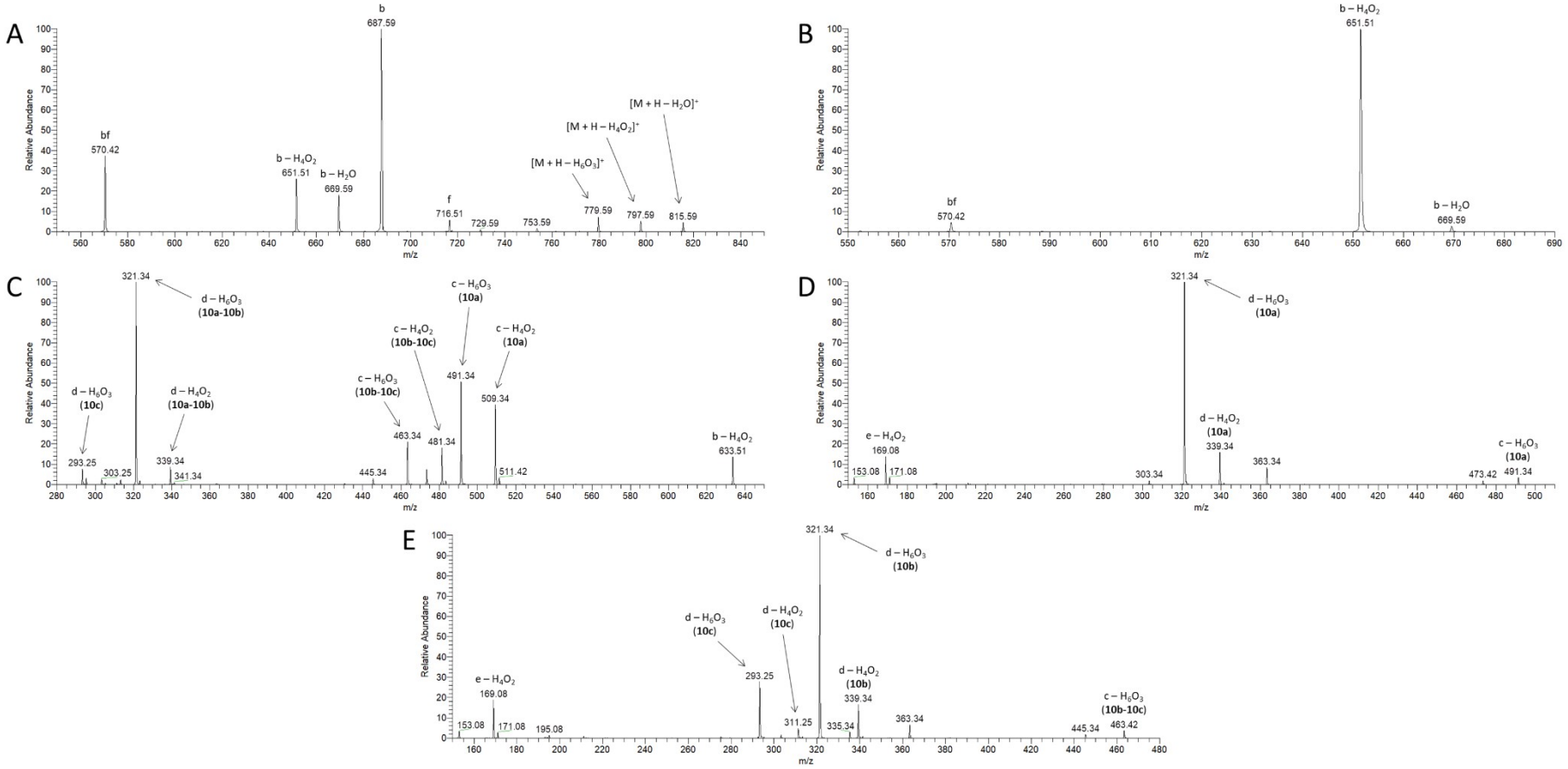
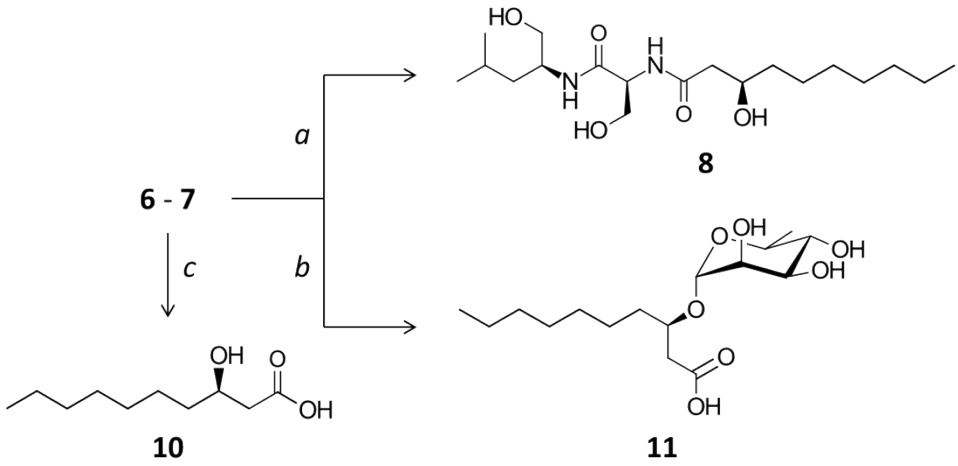


Fig. S40 Analysis of Variovoracins Rha'-$C_{8,10,10}$ (**7a**), Rha"-Rha'-$C_{10,8,10}$ (**7b**), and Rha"-Rha'-$C_{10,10,8}$ (**7c**) by ESI-MS/MS and assignment of observed fragment ions. (A) ESI-MS^2 scan of precursor ion *m/z* 833.7 $[M + H]^+$. (B) ESI-MS^3 scan of fragment ion b *m/z* 687.6. (C) ESI-MS^4 scan of fragment ion b – H_4O_2 *m/z* 651.5. (D) ESI-MS^5 scan of fragment ion c – H_4O_2 *m/z* 509.3. (E) ESI-MS^5 scan of fragment ion c – H_4O_2 *m/z* 481.3.



Scheme S1. Preparation of variovoracins **6-7** isolated from *J. agaricidamnosum* DSM 9628 for Marfey's analysis and other chiral derivatization methods. Reagents and conditions: (a) 0.8 M HCl, H_2O:THF (3:1), 60°C, 18 h. (b) 0.1 M NaOH, H_2O:acetone (9:1), rt, 40 h. (c) 1 M HCl, H_2O:CH_3CN (1:1), 95°C, 18 h**.**

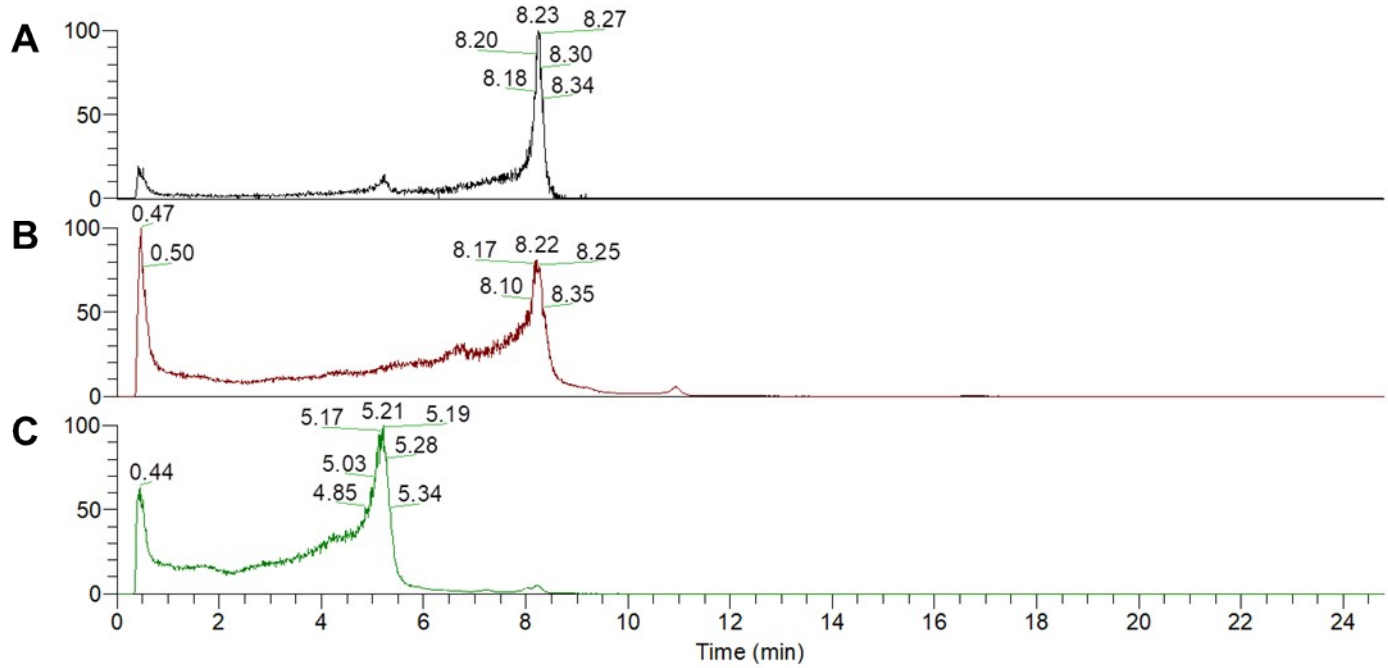


Fig. S41 Elucidation of the absolute configuration of the α-rhamnopyranose moiety of variovoracins **6-7** isolated from *J. agaricidamnosum* DSM 9628 using a RP-HPLC method previously reported by Tanaka et al.[21] Reactions were performed on glycolipid **11** (**Scheme S1**) and the thiazolidine derivatives were analyzed by UHPLC-HRMS. Extracted ion chromatograms (*m/z* 431.1305 $[M + H]^+$, ±5.0 ppm) of the anticipated thiazolidine derivatives from: (A) glycolipid **11** from *J. agaricidamnosum* DSM 9628; (B) L-rhamnose standard; (C) D-rhamnose standard.

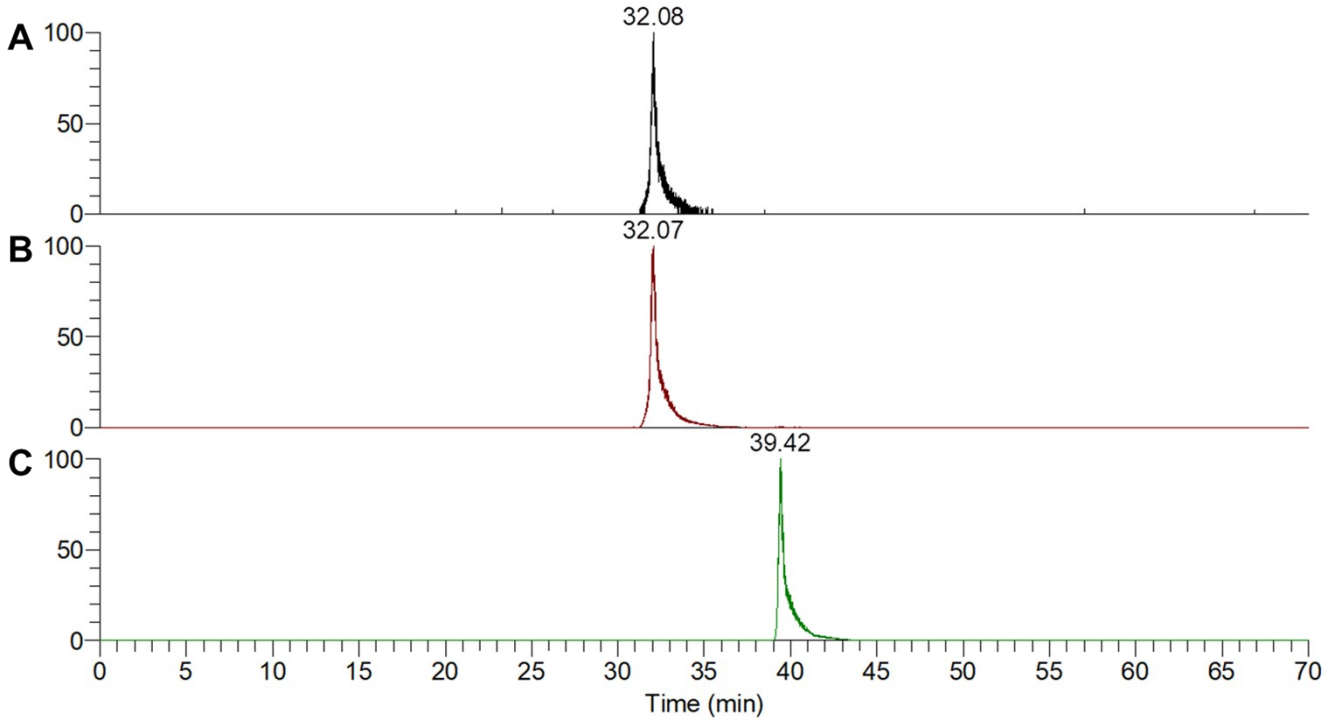


Fig. S42 Configuration assignment of the leucinol residue of variovoracins **6-7** from *J. agaricidamnosum* DSM 9628 using Marfey's method. Extracted ion chromatograms (*m/z* 370.1721 $[M+H]^+$, ±5.0 ppm) represent products from derivatization reactions of (A) lipopeptide **8** from *J. agaricidamnosum* DSM 9628; (B) L-leucinol; (C) D-leucinol.

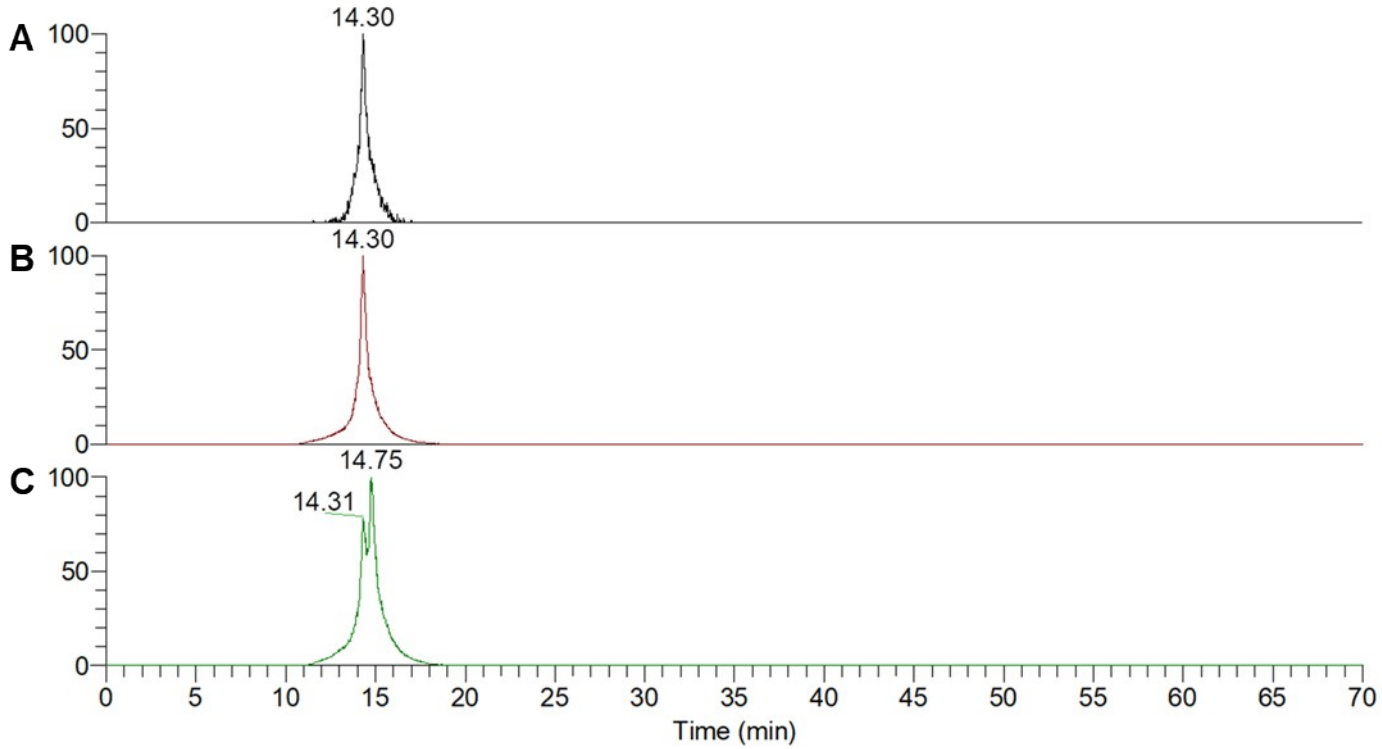


Fig. S43 Configuration assignment of the serine residue of variovoracins **6-7** from *J. agaricidamnosum* DSM 9628 using Marfey's method. Extracted ion chromatograms (*m/z* 358.0993 $[M+H]^+$, ±5.0 ppm) represent products from derivatization reactions of (A) lipopeptide **8** from *J. agaricidamnosum* DSM 9628; (B) L-serine; (C) DL-serine.

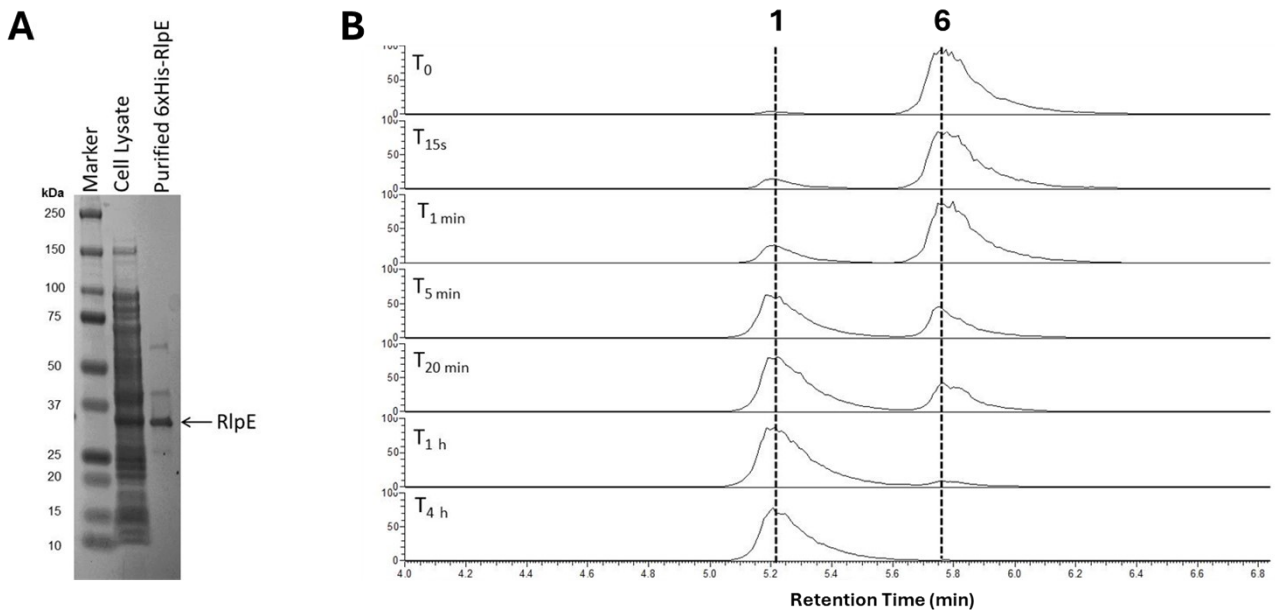
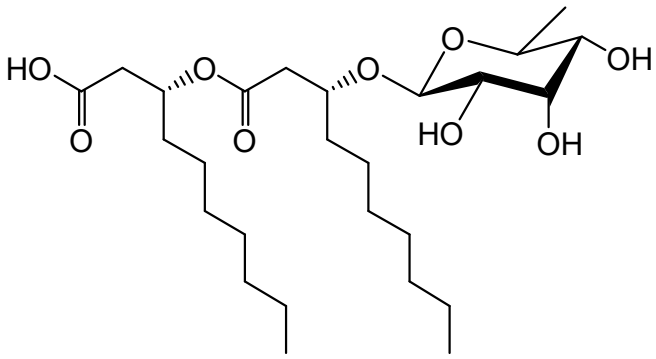
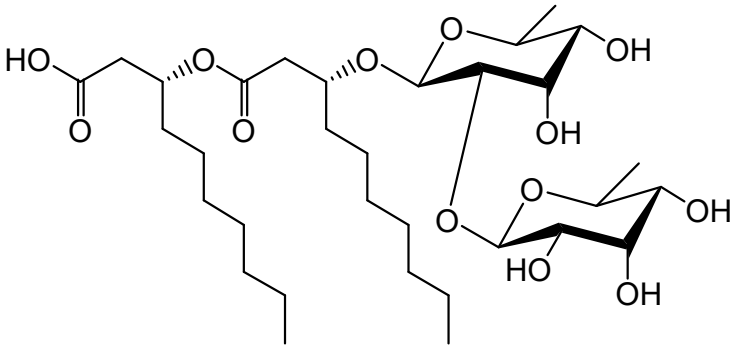


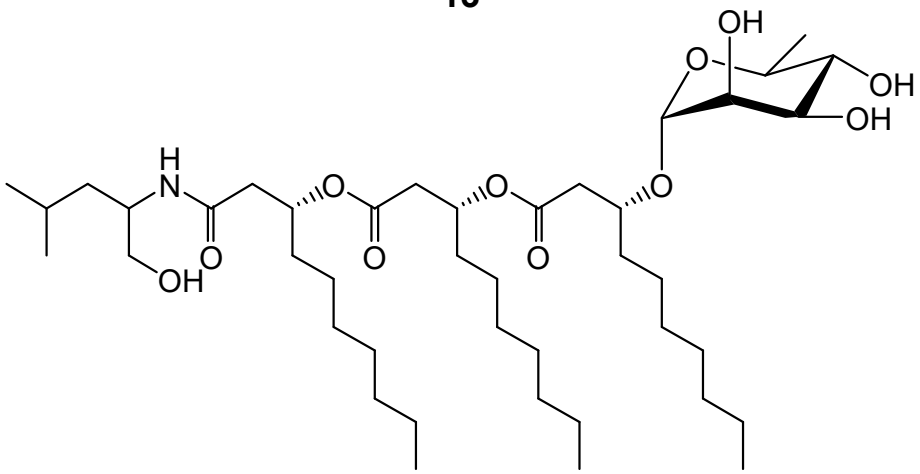
Fig. S44 Characterization of RlpE enzymatic activity. A) SDS-PAGE analysis of purified RlpE. B) UHPLC-HRMS analysis of RlpE reactions containing monorhamnosyl variovoracin **6** undergoing time-dependent conversion to dirhamnosyl variovoracin **1**. The assay mixture consists of dTDP-L-rhamnose (1 mM), variovoracin **6** (0.5 mM), $MgCl_2$ (2.5 mM) in buffer (25 mM Tris-HCl pH 8.0) incubated with 0.1 μM of purified RlpE at 30 °C for 4 h. Reactions were quenched at each time point and subjected to C_{18} UHPLC-HRMS analysis as described in the experimental procedures. Extracted ion chromatograms monitoring the $[M+H]^+$ ions for **1** (*m/z* 1007.66) and **6** (*m/z* 861.60) are shown.



12



13



14

Fig. S45 Natural products with structural similarity to variovoracins. Rhamnolipid A (**12**), rhamnolipid B (**13**) and rhizoleucinoside (**14**).

References

1. M. Hayakawa, M. Otoguro, T. Takeuchi, T. Yamazaki and Y. Iimura, *Antonie van Leeuwenhoek*, 2000, **78**, 171
2. M. Hayakawa and H. Nonomura, *Journal of Fermentation Technology*, 1987, **65**, 501
3. P. Suomalainen, C. Johans, T. Soderlund and P. K. J. Kinnunen, *J. Med. Chem.*, 2004, **47**, 1783
4. D. P. Overy, F. Berrue, H. Correa, N. Hanif, K. Hay, M. Lanteigne, K. Mquilian, S. Duffy, P. Boland, R. Jagannathan, G. S. Carr, M. Vansteeland and R. G. Kerr, *Mycology*, 2014, **5**, 130
5. C. S. Chin, D. H. Alexander, P. Marks, A. A. Klammer, J. Drake, C. Heiner, A. Clum, A. Copeland, J. Huddleston, E. E. Eichler, S. W. Turner and J. Korlach, *Nat Methods*, 2013, **10**, 563
6. M. J. Chaisson and G. Tesler, *BMC Bioinformatics*, 2012, **13**, 238.
7. E. W. Myers, G. G. Sutton, A. L. Delcher, I. M. Dew, D. P. Fasulo, M. J. Flanigan, S. A. Kravitz, C. M. Mobarry, K. H. Reinert, K. A. Remington, E. L. Anson, R. A. Bolanos, H. H. Chou, C. M. Jordan, A. L. Halpern, S. Lonardi, E. M. Beasley, R. C. Brandon, L. Chen, P. J. Dunn, Z. Lai, Y. Liang, D. R. Nusskern, M. Zhan, Q. Zhang, X. Zheng, G. M. Rubin, M. D. Adams and J. C. Venter, *Science*, 2000, **287**, 2196
8. Z. Zhang, S. Schwartz, L. Wagner and W. Miller, *J Comput Biol*, 2000, **7**, 203
9. D. H. Haft, M. DiCuccio, A. Badretdin, V. Brover, V. Chetvernin, K. O'Neill, W. Li, F. Chitsaz, M. K. Derbyshire, N. R. Gonzales, M. Gwadz, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, R. A. Yamashita, C. Zheng, F. Thibaud-Nissen, L. Y. Geer, A. Marchler-Bauer and K. D. Pruitt, *Nucleic Acids Res*, 2018, **46**, D851
10. W. Li, K. R. O'Neill, D. H. Haft, M. DiCuccio, V. Chetvernin, A. Badretdin, G. Coulouris, F. Chitsaz, M. K. Derbyshire, A. S. Durkin, N. R. Gonzales, M. Gwadz, C. J. Lanczycki, J. S. Song, N. Thanki, J. Wang, R. A. Yamashita, M. Yang, C. Zheng, A. Marchler-Bauer and F. Thibaud-Nissen, *Nucleic Acids Res*, 2021, **49**, D1020
11. T. Tatusova, M. DiCuccio, A. Badretdin, V. Chetvernin, E. P. Nawrocki, L. Zaslavsky, A. Lomsadze, K. D. Pruitt, M. Borodovsky and J. Ostell, *Nucleic Acids Res*, 2016, **44**, 6614
12. S. H. Yoon, S. M. Ha, S. Kwon, J. Lim, Y. Kim, H. Seo and J. Chun, *Int J Syst Evol Microbiol*, 2017, **67**, 1613
13. K. Blin, S. Shaw, H. E. Augustijn, Z. L. Reitz, F. Biermann, M. Alanjary, A. Fetter, B. R. Terlouw, W. W. Metcalf, E. J. N. Helfrich, G. P. van Wezel, M. H. Medema and T. Weber, *Nucleic Acids Res*, 2023, **51**, W46
14. N. Ziemert, S. Podell, K. Penn, J. H. Badger, E. Allen and P. R. Jensen, *PLoS One*, 2012, **7**, e34064.
15. D. Dubeau, E. Deziel, D. E. Woods and F. Lepine, *BMC Microbiol*, 2009, **9**, 263.
16. S. Kumar, G. Stecher, M. Li, C. Knyaz and K. Tamura, *Mol Biol Evol*, 2018, **35**, 1547
17. S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipma, *Nucleic Acids Res*, 1997, **25**, 3389
18. S. F. Altschul, J. C. Wootton, E. M. Gertz, R. Agarwala, A. Morgulis, A. A. Schäffer and Y.-K. Yu, *FEBS J.*, 2005, **272**, 5101.
19. C. L. M. Gilchrist, T. J. Booth, B. van Wersch, L. van Grieken, M. H. Medema and Y.-H. Chooi, *Bioinformatics Advances*, 2021, **1**, vbab016.
20. C. L. M. Gilchrist and Y.-H. Chooi, *Bioinformatics*, 2021, **37**, 2473
21. T. Tanaka, T. Nakashima, T. Ueda, K. Tomii and I. Kouno, *Chem. Pharm. Bull.*, 2007, **55**, 899
22. N. Saitou and M. Nei, *Mol Biol Evol*, 1987, **4**, 406
23. J. Felsenstein, *Evolution*, 1985, **39**, 783.
24. M. Nei and S. Kumar, *Molecular Evolution and Phylogenetics*, Oxford University Press, New York, 2000.
25. P. H. A. Sneath and R. R. Sokal, *Numerical Taxonomy*, Freeman, San Francisco, 1973.
26. S. Whelan and N. Goldman, *Mol Biol Evol*, 2001, **18**, 691