Electronic Supplementary Material (ESI) for ChemComm.

SUPPORTING INFORMATION

Chemoselective Probe for Detailed Analysis of Ketones and Aldehydes Produced by Gut Microbiota in Human Samples

Louis P. Conway[a], Neeraj Garg[a], Weifeng Lin[a], Miroslav Vujasinovic[b], J.-Matthias Löhr[b,c], and Daniel Globisch[a,*]

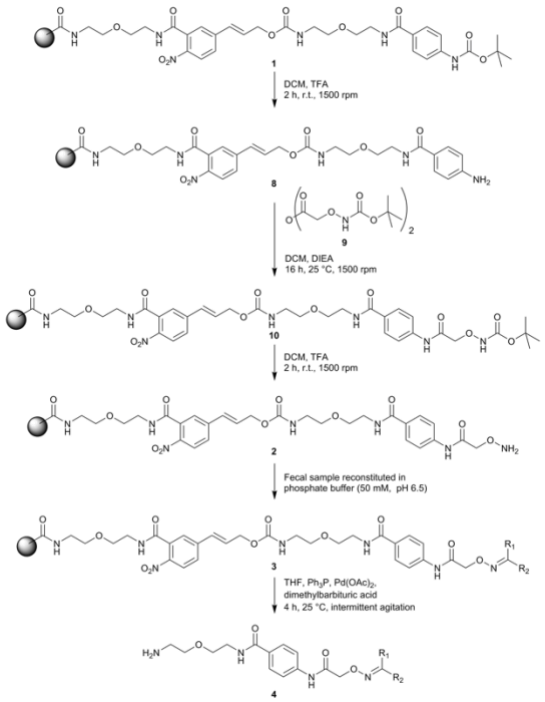
[a] Department of Medicinal Chemistry, Science for Life Laboratory
Uppsala University
Box 574, SE-75123 Uppsala, Sweden

[b] Department for Digestive Diseases, Karolinska University Hospital, Stockholm, Sweden

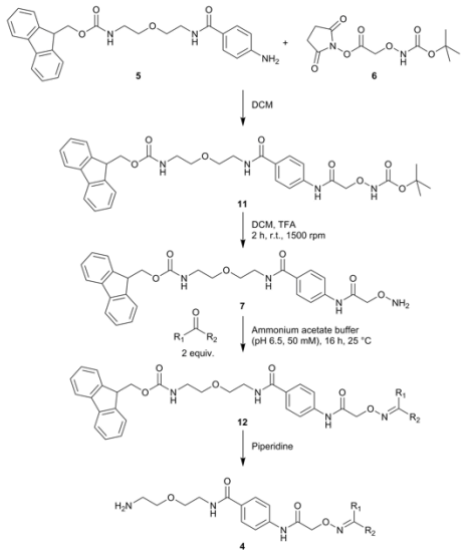
[c] Department of Clinical Science, Intervention and Technology (CLINTEC), Karolinska Institute, Stockholm, Sweden

Contents

1. Supporting schemes



Scheme S1: Preparation of chemical probe activated for carbonyl conjugation.



Scheme S2: Preparation of simplified chemical probe activated for carbonyl conjugation.

2. Supporting figures

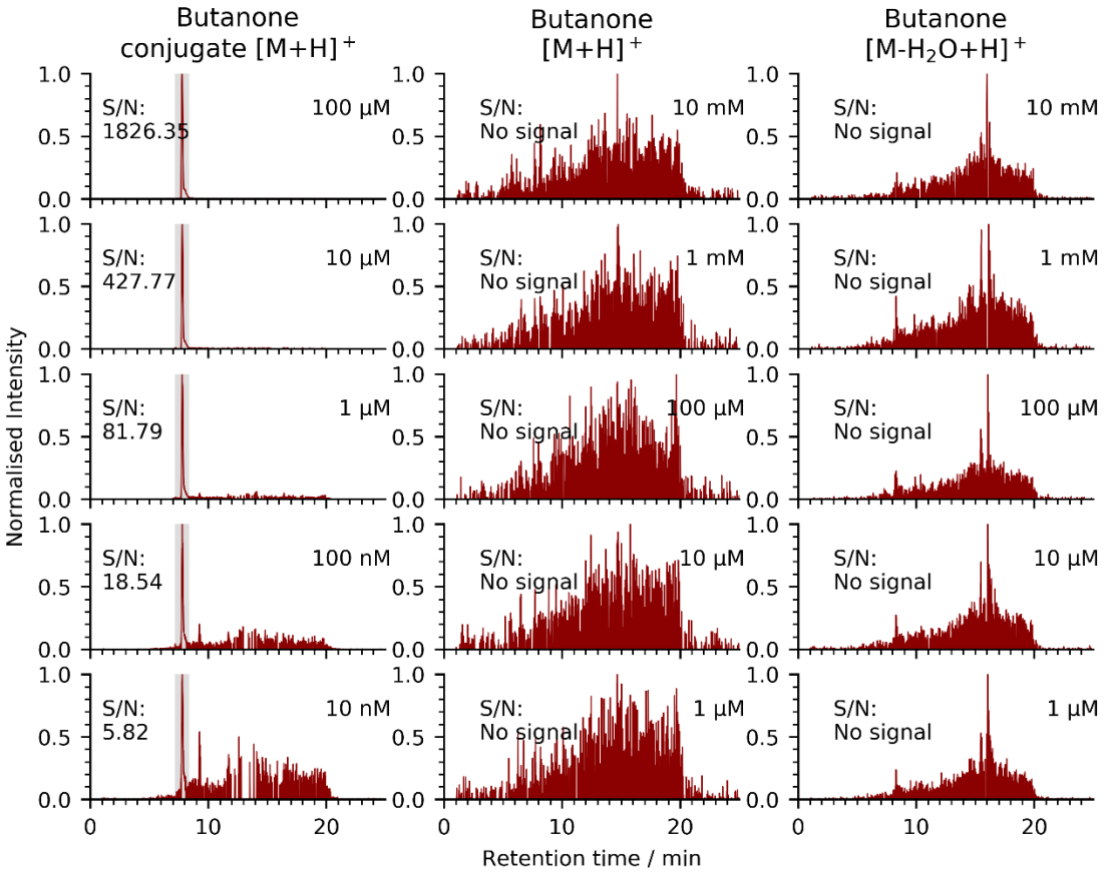


Figure S1: Extracted ion chromatograms (30 ppm) resulting from injections of butanone and its conjugate **4a** at a range of concentrations.

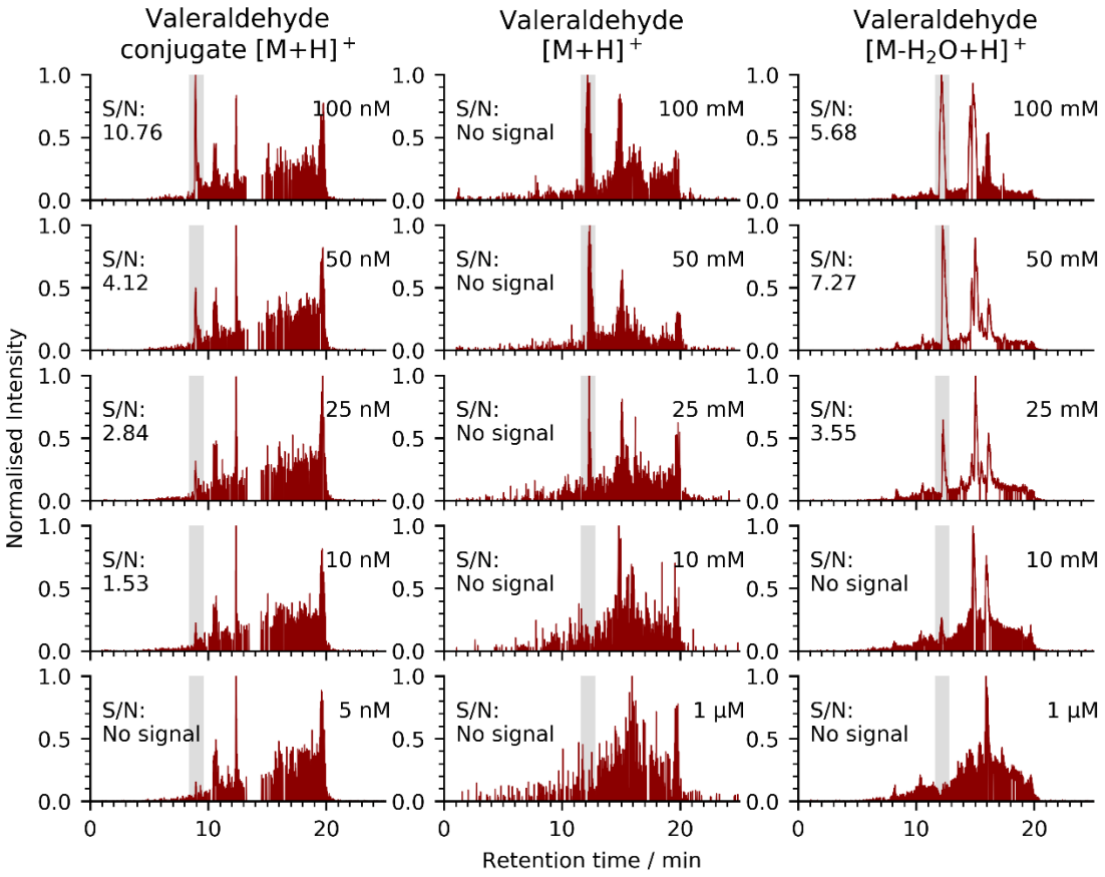


Figure S2: Extracted ion chromatograms (30 ppm) resulting from injections of valeraldehyde and its conjugate **4b** at a range of concentrations.

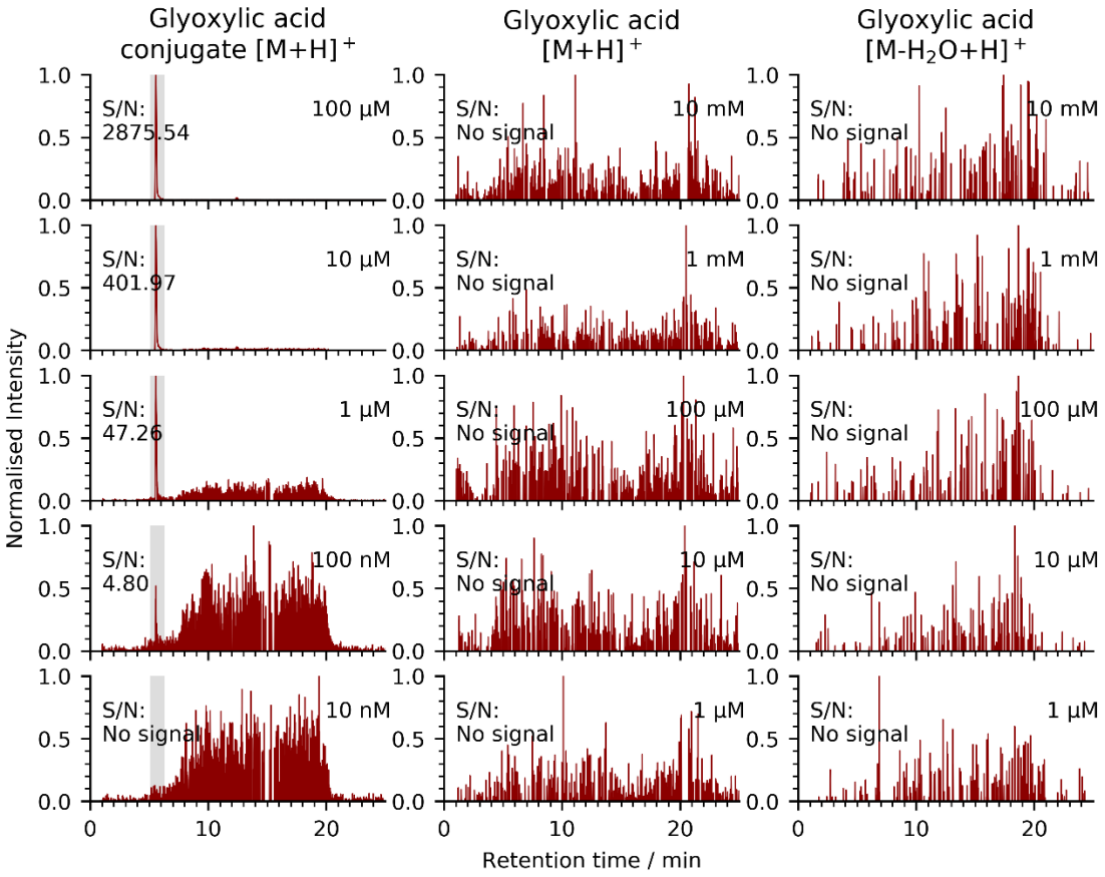


Figure S3: Extracted ion chromatograms (30 ppm) resulting from injections of glyoxylic acid and its conjugate **4d** at a range of concentrations.

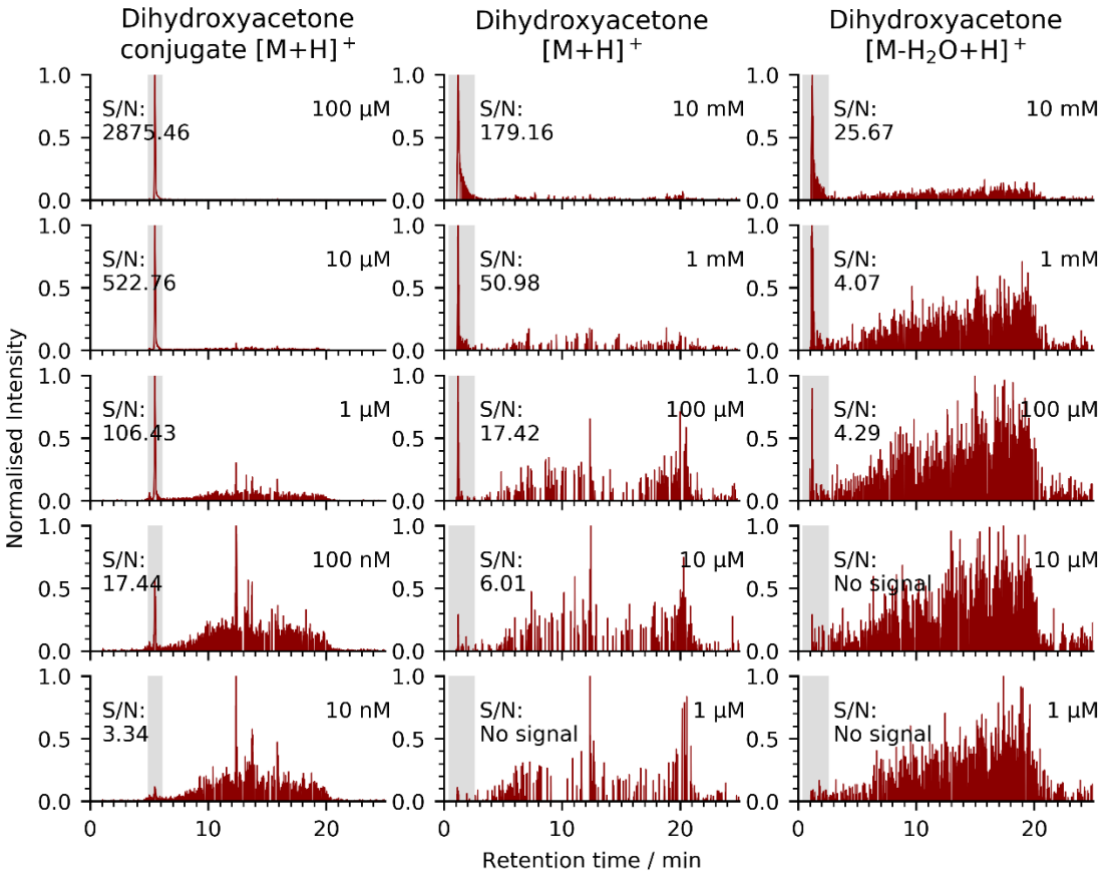


Figure S4: Extracted ion chromatograms (30 ppm) resulting from injections of dihydroxyacetone and its conjugate **4e** at a range of concentrations.

Figure S5: The first half of the total 112 metabolites detected after analysis of fecal samples using the carbonyl-specific chemical probe. Red indicates no detection, yellow denotes metabolites which are annotated on the basis of mass, while green represents metabolites which have been validated with commercial standards.

Figure S6: The second half of the total 112 metabolites detected after analysis of fecal samples using the carbonyl-specific chemical probe. Red indicates no detection, yellow denotes metabolites which are annotated on the basis of mass, while green represents metabolites which have been validated with commercial standards.

3. Supporting tables

Table S1: A list of the standards synthesized from commercial aldehydes and ketones in this study.

Substrate	Monoisotopic mass:	Predicted m/z:	RT / min:
Galactose	180.0634	459.2085	4.78
Glucose	180.0634	459.2085	4.81
Sodium mesoxalate monohydrate	117.9902	397.1353	4.89
Ribose	150.0528	429.1979	4.98
β-Hydroxypyruvic acid	104.0110	383.1561	5.17 / 5.71
L-(+)-Erythrulose	120.0423	399.1874	5.34
Glycolaldehyde dimer	60.0211	339.1662	5.35
2-Deoxy-D-ribose	134.0579	413.2030	5.35
Dihydroxyacetone	90.0317	369.1768	5.45
Glyoxylic acid	74.0004	353.1455	5.56
Formaldehyde	30.01057	309.1557	5.7
3,3-Diethoxy-1-propanol	74.0368	353.1819	5.78 / 6.00
Noroxymorphone hydrochloride solution	287.1158	566.2609	5.79
DL-Lactaldehyde solution (1 M)	74.0368	353.1819	5.92
Hydroxyacetone	74.0368	353.1819	5.95
Acetaldehyde	44.02622	323.1713	6.22
Pyruvic acid	88.0160	367.1611	6.35
Acetoin	88.0524	367.1975	6.49
Lithium acetoacetate	102.0317	381.1768	6.82
Glutaraldehyde solution	100.0524	379.1975	6.82
2,3-Pentanedione	100.0524	379.1975	6.85
Acetone	58.0419	337.1870	6.88
Acetylacetone	100.0524	379.1975	6.88
2-Ketobutyric acid	102.0317	381.1768	6.89 / 7.18
Levulinic acid	116.0473	395.1924	6.89
Propanal	58.04187	337.1870	6.92
4-Hydroxybenzaldehyde	122.0368	401.1819	7.68
Butanone	72.05752	351.2026	7.80
Cyclopentanone	84.05752	363.2026	7.91
2-Pentanone	86.0732	365.2183	8.45
Isovaleraldehyde	86.0732	365.2183	8.59
2-Methylbutyraldehyde	86.0732	365.2183	8.61
Salicylaldehyde	122.0368	401.1819	8.70
Valeraldehyde	86.0732	365.2183	8.71
2-Methylbutyraldehyde	86.0732	365.2183	8.83
Benzaldehyde	106.0419	385.1870	9.00
Acetophenone	120.0575	399.2026	9.25
2-Hexanone	100.0888	379.2339	9.29
2-Ethylbutyraldehyde	100.0888	379.2339	9.31
3-Hexanone	100.0888	379.2339	9.58
Hexanal	100.0888	379.2339	9.58
2-Methyl-3-pentanone	100.0888	379.2339	9.60
4-Methyl-2-pentanone	100.0888	379.2339	9.62
o-Tolualdehyde	120.0575	399.2026	9.66
p-Tolualdehyde	120.0575	399.2026	9.76
m-Tolualdehyde	120.0575	399.2026	9.78
(±)-Jasmonic acid	210.1256	489.2707	9.87
Safranal	150.1045	429.2496	10.90
Octanal	128.1201	407.2652	11.00
2-Octanone	128.1201	407.2652	11.01
(1R)-(−)-Myrtenal	150.1045	429.2496	11.02
R-Carvone	150.1045	429.2496	15.52

Table S2: Previously reported carbonyl-containing metabolites.

Carbonyl metabolite	HMDB ID
D-Glucose	HMDB0000122
D-Ribose	HMDB0000283
Deoxyribose	HMDB0003224
Dihydroxyacetone	HMDB0003224
Glyoxylic acid	HMDB0000119
Acetoin	HMDB0003243
Propanal	HMDB0003366
4-Hydroxybenzaldehyde	HMDB0011718
Isovaleraldehyde	HMDB0006478
Valeraldehyde	HMDB0006478
2-Methylbutanal	HMDB0031526
Benzaldehyde	HMDB0006115
Methyl isobutyl ketone	HMDB0002939
Octanal	HMDB0001140

Table S3: Identified carbonyl metabolites linked to disease development.

Carbonyl metabolite	Link to disease
Octanal	Lung injury and disease
4-Hydroxynonenal	Neurodegenerative diseases
Glyoxylic acid	Bladder infections; Primary hyperoxaluria
Isovaleraldehyde	Hepatic encephalopathy
Ketoleucine	Maple syrup urine disease (MSUD); colorectal cancer; Rheumatoid arthritis
4-Hydroxybenzaldehyde	Colorectal cancer; obesity
Acrolein	Spinal cord injury, multiple sclerosis, Alzheimer's disease, cardiovascular disease, diabetes mellitus
Malondialdehyde	Oxidative damage related diseases
Dopamine quinone	Parkinson's disease
Erythrose	Schizophrenia

4. General

All non-aqueous reactions were performed using flame- or oven dried glassware under an atmosphere of dry nitrogen. All reagents and solvents were purchased from Sigma-Aldrich or Fischer Scientific and were used without further purification. HPLC grade solvents were used for HPLC purification and mass spectrometry grade for UHPLC-ESI-MS analysis. Solutions were concentrated *in vacuo* on a Heidolph or a IKA rotary evaporator. Thin Layer Chromatography (TLC) was performed on silica gel 60 F-254 plates. Visualization of the developed chromatogram was performed using fluorescence quenching or staining with CAM (cerium ammonium molybdate), ninhydrin, Ehrlich reagent (4-(dimethylamino)benzaldehyde) or vanillin. Chromatographic purification of products was accomplished using flash column chromatography on Merck silica gel 60 (40−63 μm) or preparative reverse phase HPLC on an Agilent HPLC-1100 series system equipped with a Symmetry Prep C18 column (19 × 150 mm,

7 µm) at a 2.5 or 4.0 mL/min flow rate. All synthesized compounds were ≥95 % pure as determined by NMR. NMR spectra were recorded on a Bruker 600 MHz spectrometer (^{1}H NMR: 600.18 MHz, ^{13}C NMR: 150.92 MHz) or Agilent 400 MHz spectrometer (^{1}H NMR: 399.97 MHz, ^{13}C NMR: 100.58 MHz) or Varian 300 MHz spectrometer (^{13}C NMR: 75.43 MHz). Chemical shifts are reported in parts per million (ppm) on the δ scale from an internal standard. Multiplicities are abbreviated as follows: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet. Glass vials used for handling magnetic beads were microwave vials from Biotage (0.2-0.5 mL or 0.5-2.0 mL). High-resolution mass spectra were acquired on a SYNAPT G2-S High Definition Mass Spectrometer (HDMS) using an electrospray ionization (ESI) source with an AQCUITY UPLC I-class system and equipped with a Waters ACQUITY UPLC BEH C18 column (2.1 × 75 mm, 1.7 µm particle size) or Waters ACQUITY UPLC HSS T3 column (1.8 × 100 mm, 2.1 µm particle size). HPLC samples were analyzed by analytical reverse phase HPLC on an Agilent HPLC-1100 series system equipped with a Symmetry Prep C18 column (19 × 100 mm, 3.5 µm) at a 0.25 mL/min flow rate.

5. Ethical approval

Patient fecal samples were obtained in accordance with the World Medical Association Declaration of Helsinki and all patients gave written informed consent. Approval for the study was obtained from the ethical committee at Karolinska Institutet Hospital (Ethical approval number: Dnr 2017/290-31). Fecal samples were collected using routine clinical collection protocols and all patient codes have been removed in this publication. All samples were stored at -80 °C.

6. Description of procedures

6.1 UHPLC-MS analysis

Mass spectrometric analysis was performed on an Acquity UPLC system connected to a Synapt G2 Q-TOF mass spectrometer, both from Waters Corporation (Milford, MA, USA). The system was controlled using the MassLynx software package v 4.1, also from Waters. The separation was performed on an Acquity UPLC® BEH C18 column (1.7 µm, 100×2.1 mm) from Waters Corporation. The mobile phase consisted of a combination of 0.1% formic acid in MilliQ water (A) and 0.1% formic acid in LC-MS grade methanol (B). The column temperature was 40 °C and the mobile phase gradient applied was as follows: 0-2 min, 0% B; 2-15 min, 0-100% B; 15-18 min, 100% B; 18-20 min, 100-0% B; 20-25 min, 0% B, with a flow rate of 0.3 ml/min.

The samples were introduced into the q-TOF using positive electrospray ionization. The capillary voltage was set to -2.50 kV and the cone voltage was 40 V. The source temperature was 100 °C, the cone gas flow 50 L/min and the desolvation gas flow 600 L/h. The instrument was operated in MSE mode, the scan range was m/z = 50-1200, and the scan time was 0.3 s. A solution of sodium formate (0.5 mM in 2-propanol: water, 90:10, v/v) was used to calibrate the instrument and a solution of leucine-encephalin (2 ng/µl in acetonitrile: 0.1% formic acid in water, 50:50, v/v) was used for the lock mass correction at an injection rate of 30 s.

6.2 Preparation of bead-bound, unactivated probe 8

The bead-bound, unactivated probe was prepared according to the method previously reported.[1]

6.3 Activation of carbonyl-specific chemoselective probe 2

The bead-bound, unactivated probe **8** (320 nmol, carboxylate basis) were suspended in a solution of crude (Boc-aminooxy)acetic anhydride (2.5 mg, 6.9 µmol) in DCM (200 µL). The suspension was agitated at 25 °C in a ThermoMixer (1,500 rpm) for 16 h. After the reaction was complete, the supernatant was removed and the beads were washed with THF (3 × 200 µL). The beads were then suspended in a solution of DCM (200 µL) and TFA (100 µL) and agitated in a ThermoMixer (1,500 rpm) for 2 h. The beads were then washed with phosphate buffer (2 × 200 µL, 50 mM, pH 6.5).

6.4 Preparation of fecal metabolite extracts

A scalpel was used to collect approximately 100 mg of the frozen fecal sample (stored at -80 °C). The sample was freeze-dried overnight. The correct amount of water was added (100 µL ultrapure water for every 60 mg of dried fecal sample). The mixture was vortexed and subsequently homogenized by a FastPrep 24 homogenizer (3 cycles, 6 m/s, 40 s, MP Biomedicals) using specialized tube D (MP Biomedicals). The mixture was taken out from tube D into Eppendorf tubes, then centrifuged (18,620 g, 5 min, 4 °C) and the supernatant was collected. A portion of the supernatant (100 µL) was combined with LC-MS grade methanol (400 µL) and stored at -20 °C for at least 1 h. The suspension was vortexed, centrifuged (12,000 g, 5 min, 20 °C) and the solvents were removed from the supernatant through vacuum centrifugation. The residue was redissolved in phosphate buffer (200 µL, pH 6.5, 50 mM).

6.5 Treatment of fecal metabolite extracts

The activated beads **2** were used to treat the fecal extract (200 µL in pH 6.5 phosphate buffer, 50 mM, derived from 100 µL of supernatant). The mixture was shaken for 16 h at 1500 rpm and 25 °C. The fecal extract solution was removed from the beads and the beads were washed with THF (2 x 200 µL) before being resuspended in THF (300 µL).

6.6 Cleavage of the bead-bound chemical probe 3

The suspension of beads was transferred to a glass vial. Triphenylphosphine (97.0 µL, 12.9 mM in THF, 1.25 µmol) and dimethylbarbituric acid (90.0 µL, 30.7 mM in THF, 2.76 µmol) solutions were added to the vial, followed by palladium (II) acetate solution (84.0 µL, 6.53 mM in THF, 549 nmol). The vial was quickly sealed and a stream of nitrogen was passed through until approximately half the volume of the suspension remained. The vial was agitated at intervals on a vortexer and the reaction was allowed to continue 5 h. In parallel, a sample of unmodified beads was treated with the same cleavage conditions as the activated beads treated with fecal extract and used as control sample. The supernatant was removed from the beads using magnetic separation and the solvent removed using a vacuum centrifuge. The residues were redissolved in MeOH (30 µL each) and triphenylphosphine and triphenylphosphine oxide were precipitated through the addition of water (120 µL each). The suspension was centrifuged (benchtop centrifuge, 12,000 g, 5 min), the supernatant removed,

and the solvent was again removed with the vacuum centrifuge. The residues were redissolved in water/acetonitrile solution (95:5 v/v) and submitted for LC-MS analysis.

6.7 Synthesis of Fmoc-protected simplified probe 11

Compound **5** (40 mg, 89.8 µmol) was combined with (Boc-aminooxy)acetic acid NHS ester (129 mg, 92 µmol, prepared according to a literature procedure)[2] and DCM (2 ml). The reaction was stirred at room temperature for 18 h, before the solvents were removed under reduced pressure and the residue was purified by column chromatography (5:3 EtOAc/hexane rising to 3:1 EtOAc/hexane, followed by 100% EtOAc) to yield the product as a white solid (2.8 mg, 5 %). δ_H (500 MHz, MeOD) 1.47 (s, 9H, Boc), 3.31 (m, 2H, obscured by residual solvent peak, FmocNHC*H*$_2$), 3.55 (4H, m, ArCONHC*H*$_2$, FmocNHCH$_2$C*H*$_2$), 3.64 (2H, t, *J* 5.3, ArCONHCH$_2$C*H*$_2$), 4.14 (1H, t, *J* 6.7, FmocC*H*), 4.29 (2H, d, *J* 7.0 Hz, FmocC*H*$_2$), 4.39 (2H. s, BocNHC(O)C*H*$_2$), 7.37 (2H, app. t, *J* 7.5, Fmoc *Ar*), 7.28 (2H, t, *J* 7.5, Fmoc *Ar*), 7.59 (2H, dd, *J* 7.5, 1.0, Fmoc *Ar*), 7.73 (2H, d, *J* 8.7, aminobenzoate), 7.77 (2H, d, *J* 7.5, Fmoc *Ar*), 7.81 (2H, d, *J* 8.7, *Ph*); δ_C NMR (125.8 MHz, MeOD) 28.4, 41.0, 41.7, 49.6, 67.7, 70.5, 71.0, 77.0, 83.4, 120.2, 120.9, 126.1, 128.1, 128.8, 129.3, 131.1, 142.2, 142.6, 145.3, 159.0, 169.7, 170.0; *m/z* 619.2775 ([M+H]$^+$, 100%) requires 619.2762.

6.8 Preparation of probe-conjugated standards

A solution of Fmoc-protected probe **11** (50 µl, 1.0 mM in MeOH, 50 nmol, Scheme S2) was evaporated under reduced pressure. The residue was combined with DCM (50 µl) and TFA (100 µl). The solution was shaken at 1500 rpm for 2 h, before the solvents were removed under reduced pressure. The residue was then combined with a solution of either four aldehyde/ketone standards (0.5 equiv. each in 400 µl, 50 mM, pH 6.5 ammonium acetate buffer) or, for the LOD measurements, a single aldehyde/ketone standard (10 equiv. in 400 µl, 50 mM, pH 6.5 ammonium acetate buffer). The resulting solution was then shaken at 1500 rpm for 18 h at 25 °C. The solvents were then removed under reduced pressure, and the residues were treated with piperidine (80 µl) and shaken at 1500 rpm for 4 h at 25 °C. The piperidine was then removed under reduced pressure, and the residue was redissolved in MeOH (100 µl) followed by water (400 µl). The solution was diluted as necessary in a solution of water and acetonitrile (95:5 *v/v*) before being submitted for UPLC-MS analysis (see section 6.1).

6.9 LC-MS analysis

Six injections were performed for fecal extract-treated bead cleavage product and six injections for the control sample. See section 6.1 for details of the UHPLC-MS analysis. For the first 90 s of the analysis, the output of the UHPLC system was diverted to waste and did not enter the mass spectrometer.

6.10 Data analysis

Data files from the LC-MS analysis were converted into the NetCDF file format using MassLynx 4.1 (Waters). The XCMS library was used to perform peak detection and align the chromatograms.[3] The feature list was reduced by eliminating those features with an *m/z* value less than 279.1451 (the *m/z* value corresponding to the monoprotonated probe with no captured metabolite). More abundant features in the

control sample and less than five-fold higher abundance in the feces sample set were eliminated from the data analysis. Mass values of each feature with 279.1451 Da subtracted (corresponding to the mass of the probe) were compared to the human metabolome database in order to find plausible candidates for the parent metabolites. Commercial or synthetic standards (section 6.3) were then used to confirm the identity of the metabolites and identification of the correct regioisomers.

6.11 Signal to noise calculation

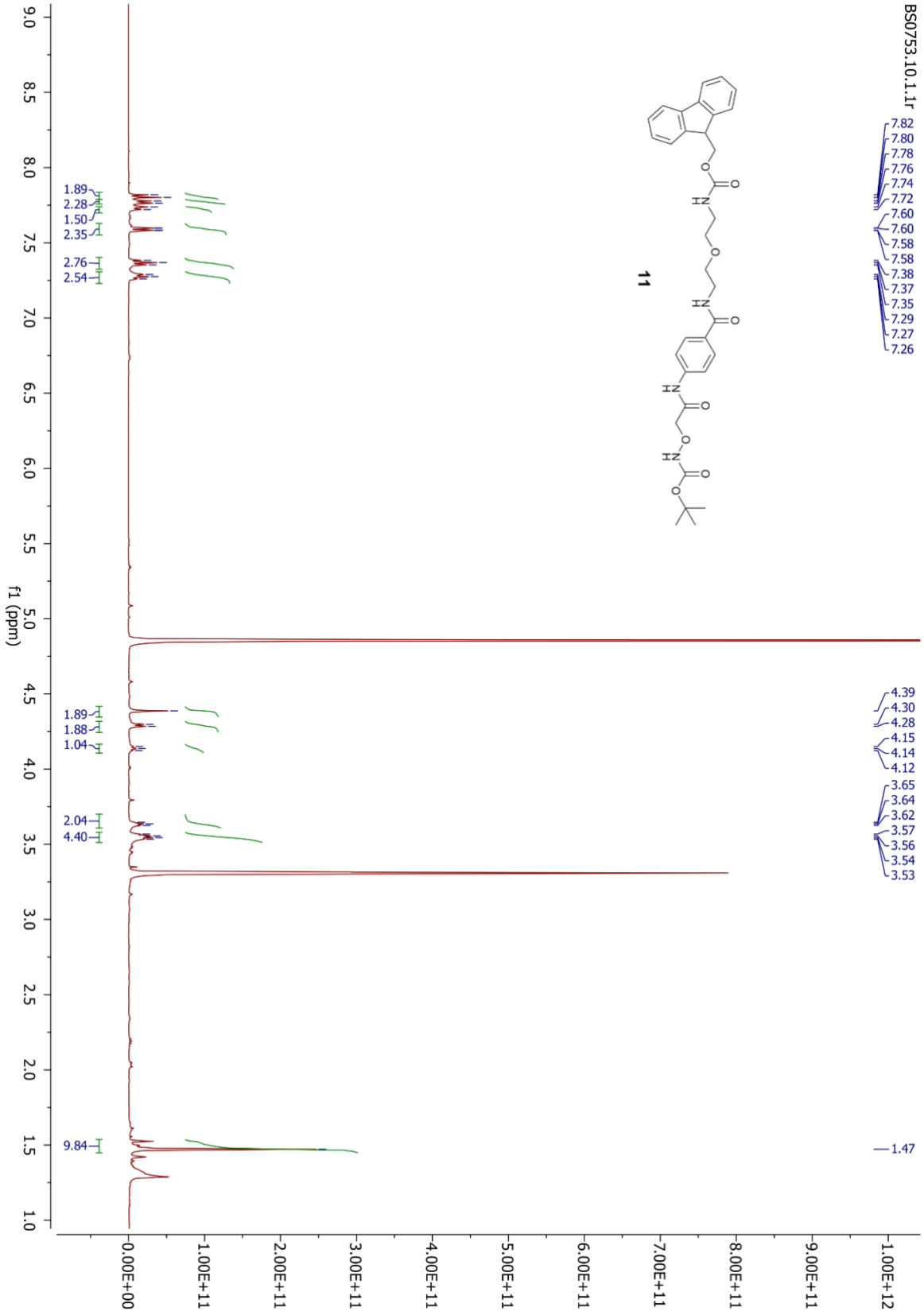
Signal to noise ratios were calculated according to European Pharmacopoeia guidelines.[4]

7. References

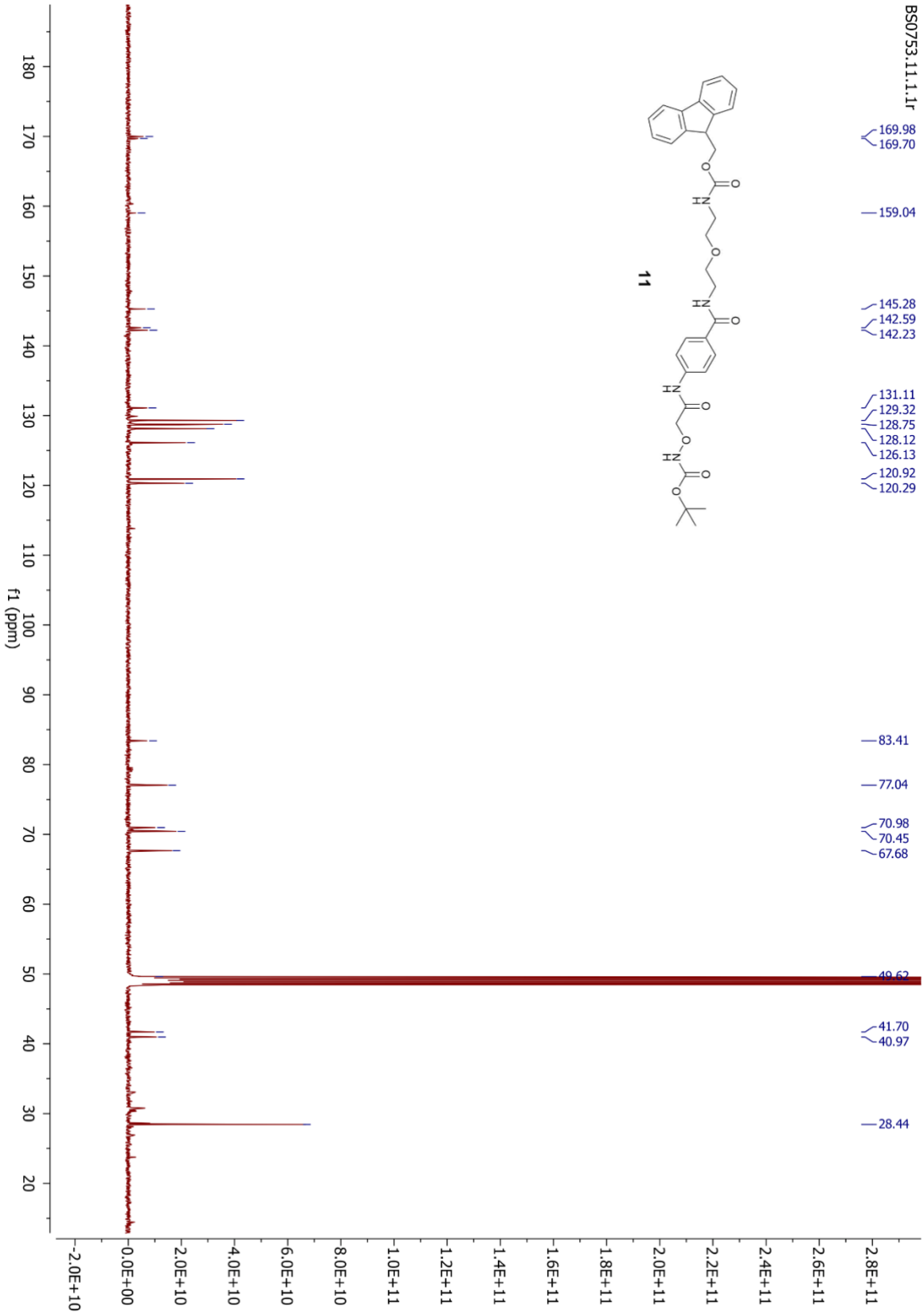
1. N. Garg, L. P. Conway, C. Ballet, M. S. P. Correia, F. K. S. Olsson, M. Vujasinovic, J. M. Lohr and D. Globisch, *Angew Chem Int Edit*, 2018, **57**, 13805-13809.
2. M. Lelle and K. Peneva, *Amino Acids*, 2014, **46**, 1243-1251.
3. C. A. Smith, E. J. Want, G. O'Maille, R. Abagyan and G. Siuzdak, *Anal Chem*, 2006, **78**, 779-787.
4. T. Wenzl, H. Johannes, A. Schaechtele, P. Robouch and J. Stroka, *Guidance Document on the Estimation of LOD and LOQ for Measurements in the Field of Contaminants in Feed and Food*, European Union Reference Laboratory (EURL), DOI: 10.2787/8931, 2016.

8. NMR Spectra

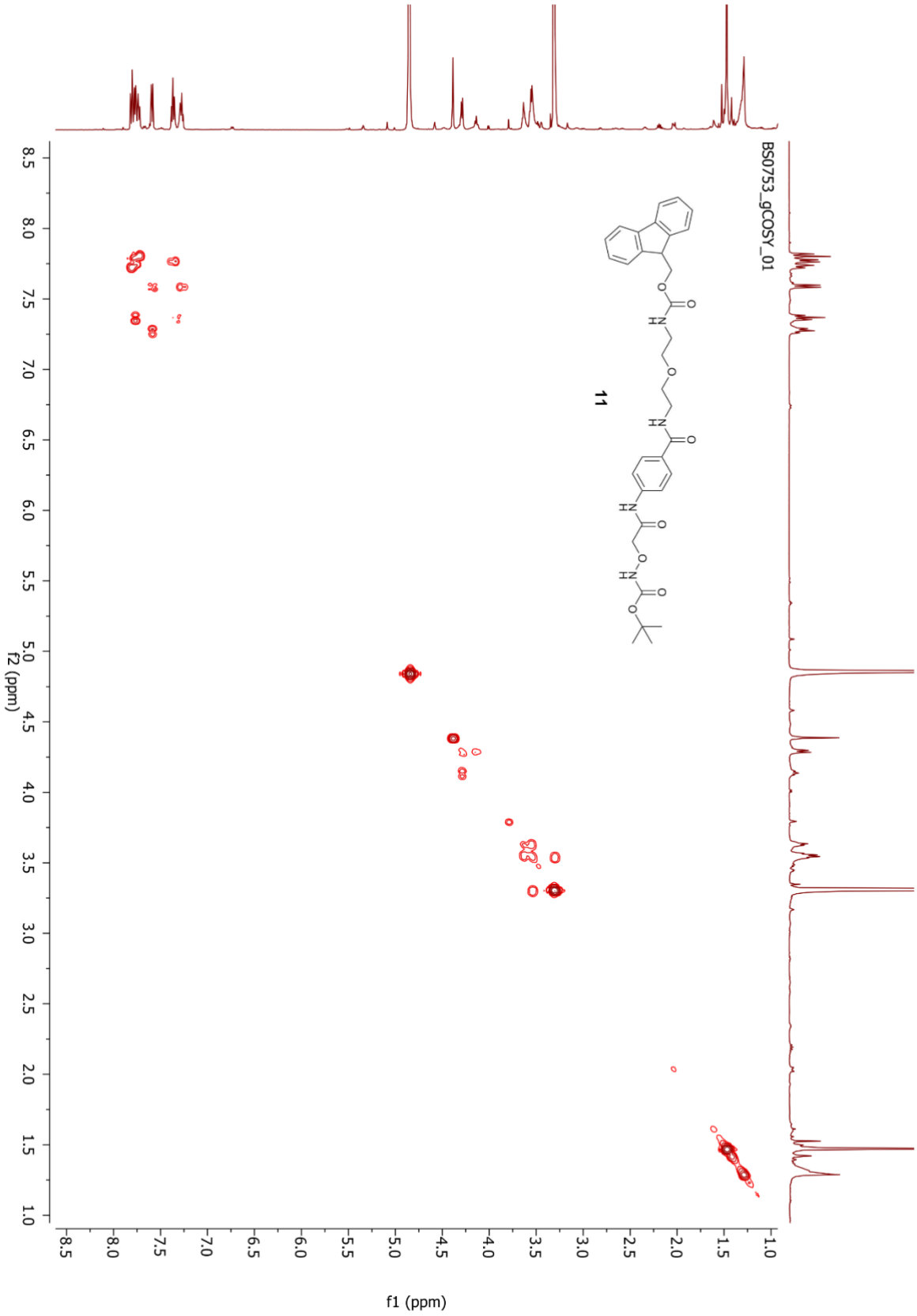
1. Compound 11 ^{1}H NMR Spectrum



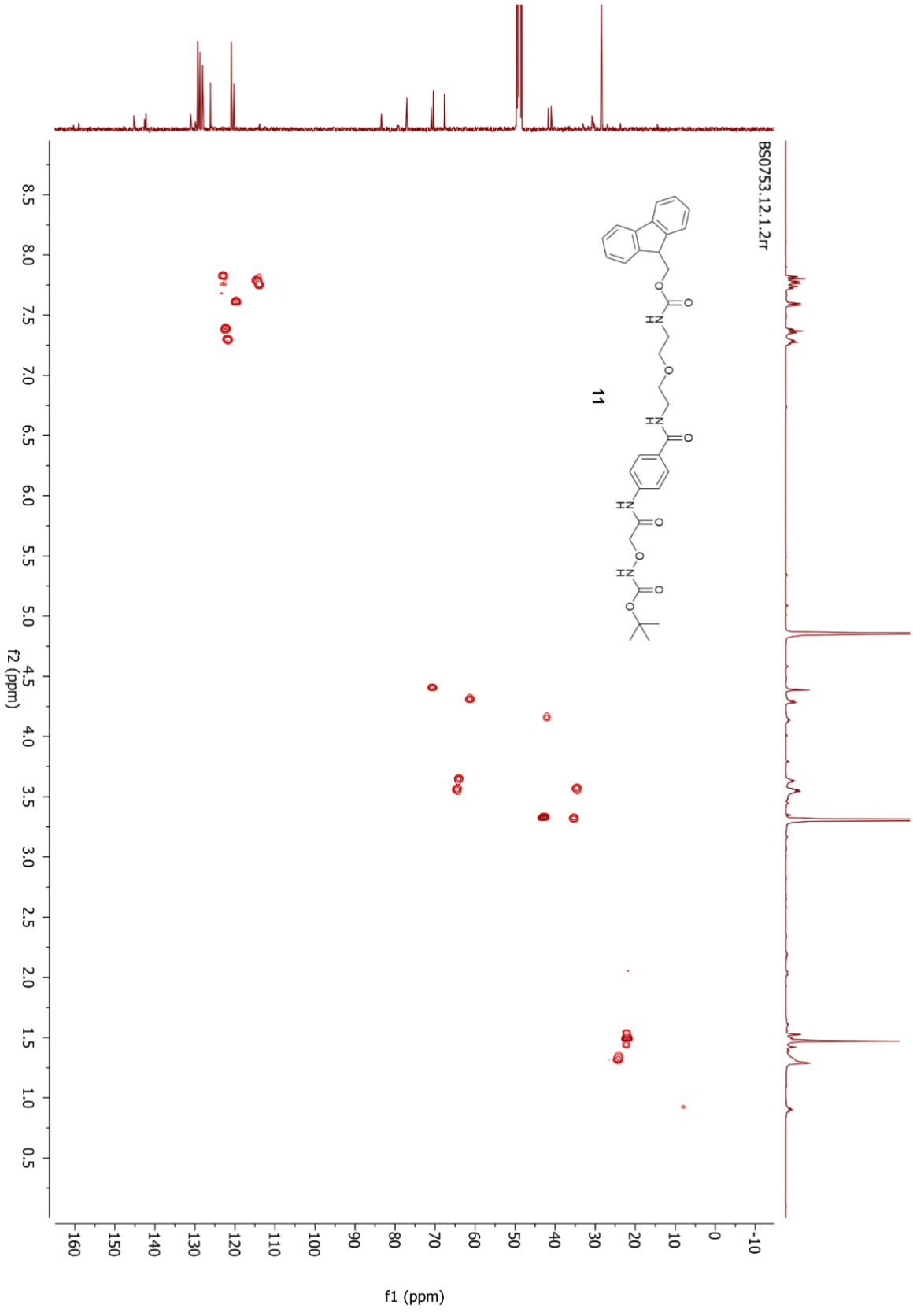
2. Compound 11 ^{13}C NMR Spectrum



3. Compound 11 COSY NMR Spectrum



4. Compound 11 HSQC NMR Spectrum



5. Compound 11 HMBC NMR Spectrum