Supporting Information

Opportunities and Limitations for Untargeted Mass Spectrometry Metabolomics to Identify Biologically Active Constituents in Complex Natural Product Mixtures

Lindsay K. Caesar, [†] Joshua J. Kellogg, [†] Olav M. Kvalheim, [‡] and Nadja B. Cech*[†]

[†] Department of Chemistry & Biochemistry, University of North Carolina Greensboro, Greensboro, NC 27402, United States

[‡] Department of Chemistry, University of Bergen, Bergen, Norway

* corresponding author, email: nadja_cech@uncg.edu, voice 336-324-5011, fax 336-334-5402

Content

- Plant extraction, production of simplified *A. keiskei* fraction, and randainal (**5**) NMR data
- Figure S1: Calibration curves of standard compounds **1-4** (berberine, magnolol, cryptotanshinone, and α-mangostin)
- Figure S2: Dose-response curves of standard compounds **1-4** (berberine, magnolol, cryptotanshinone, and α-mangostin)
- Table S1: Effect of data acquisition protocols on selectivity ratio analyses
- Table S3: False positives and their distribution in selectivity ratio models
- Table S4: Effect of data processing protocols on selectivity ratio analyses
- Figure S3: Biological activity data of second stage fractions
- Figure S4: Fractionation Scheme
- Table S5: Effect of round of fractionation on selectivity ratio analyses
- Figure S5: Example chromatograms of active fractions from first and second stages of fractionation
- Table S6: Comparison of stage-one models and their identification of randainal among the top contributors to biological activity
- Figure S6: MS^2 spectrum (negative mode) of randainal (compound **5**)
- Figure S7: ^1H NMR spectrum (700 MHz, CD_3OD) of randainal (compound **5**)
- Figure S8: HSQC spectrum (700 MHz, CD_3OD) of randainal (compound **5**)
- Figure S9: HMBC spectrum (700 MHz, CD_3OD) of randainal (compound **5**)
- Figure S10: ^1H NMR spectrum (500 MHz, acetone-d_6) of randainal (compound **5**)

Note: Table S2 is provided as a separate Excel document, and includes the complete peak lists (post MZMine processing) for all models produced.
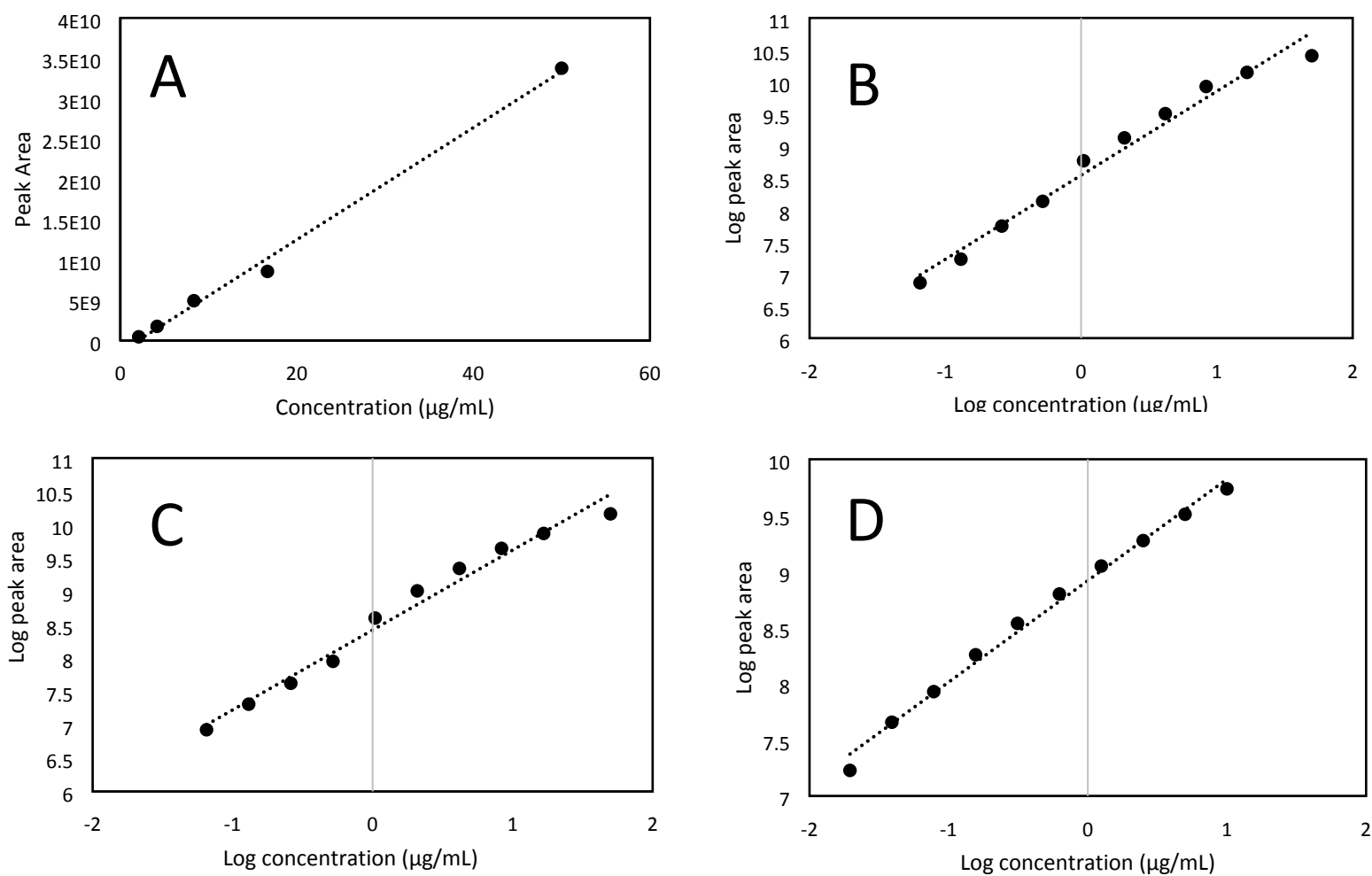
Plant Material and Extraction

Fresh *Angelica keiskei* Koidz. (Apiaceae) roots were collected on November 14, 2015 in Williams, Oregon from Strictly Medicinal Seeds ® (Sample # 12444, N 42°12'17.211", W 123°19'34.60"). The identity of the sample was confirmed by Richard A. Cech and a voucher specimen was deposited at the University of North Carolina Chapel Hill Herbarium (NCU627665). Fresh root material was dried at 40°C for 24 h in a single-wall transite oven (Blue M Electric Company, Blue Island, IL, USA), yielding 138.9 g of dried root material. Roots were then ground to a powder using a Wiley Mill Standard Model No. 3 (Arthur Thomas Col, Philadelphia, PA, USA). Powdered root was submerged in MeOH at 160 g/L for 24 h, then filtered from the solvent. This process was repeated using the same root material every 24 h for 72 h. The resulting methanol extract was then subjected to liquid-liquid partitioning. Fats were separated from the mixture by partitioning 10% aqueous methanol and hexane 1:1). The aqueous/methanol layer was partitioned again using EtOAc/MeOH/H_2O (4:5:1). Lastly, hydrosoluble tannins were separated from the EtOAc layer by washing it with a 1% NaCl aqueous solution (1:1). The resulting EtOAc extract was dried under nitrogen, yielding 3,650.32 mg of material.

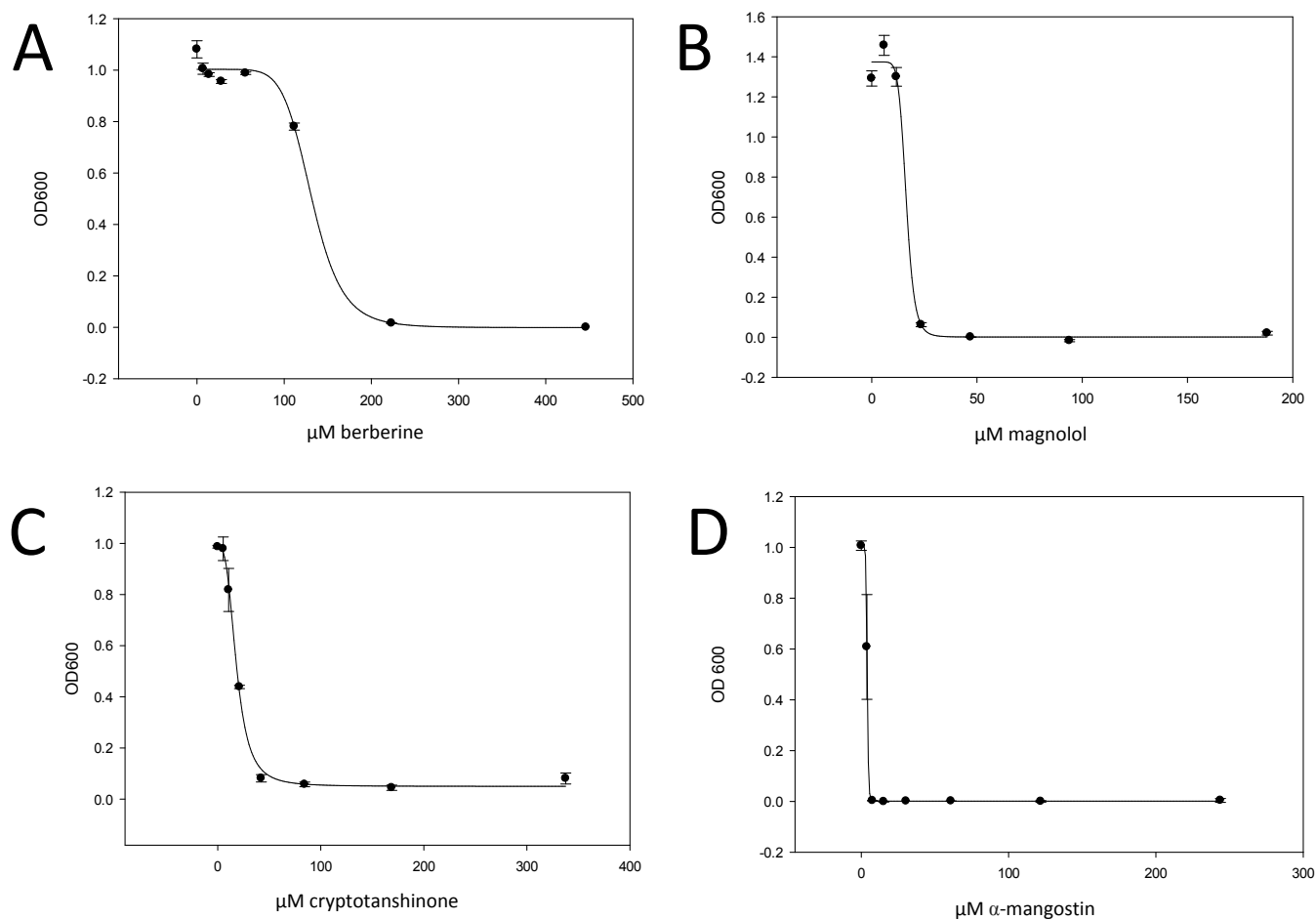## Production of Simplified *A. keiskei* Fraction

The EtOAc extract was separated using a 40 min normal-phase gradient conducted on a Combiflash RF instrument (Teledyne ISCO, Lincoln, NE, USA). The gradient began with a three min hold at 100% hexane, after which it was increased to 100% chloroform over the next 20 min. It was then held at 100% chloroform for nine min, after which the gradient was increased to 20:80 MeOH:CHCl$_3$ over three min. These conditions were held for five min, after which they were increased to 100% methanol over two min. The gradient was held at 100% methanol for one min. The resulting tubes were separated into nine fractions and subjected to biological activity testing. The ninth fraction was collected from 20-100% methanol, and was used for the remainder of the experimental procedures, due to its lack of antimicrobial activity (<15% inhibition at 100 µg/mL against a laboratory strain of *Staphylococcus aureus*, SA1199[34]).

NMR Data for Randainal

Randainal (5): yellow, amorphous powder; HRESIMS *m/z* 279.1028 [M-H]$^-$ (calculated for $C_{18}H_{15}O_3^-$, 279.1021). Fragmentation patterns matched predicted patterns as well as previously reported fragments from the literature[64] (Figure S6, Supporting Information). ^{1}H NMR (700 MHz, CD$_3$OD) δ = 3.34 (2H, d, J = 6.3 Hz, H$_2$-7'), 4.57 (2H, s, OH), 5.00 (1H, ddd, J = 10, 2, 1 Hz, H-9'$_a$), 5.06 (1H, ddd, J = 17, 2, 1 Hz, H-9'$_b$), 5.98 (1H, ddt, J = 17, 10, 6.7 Hz, H-8'), 6.58 (1H, dd, J = 15.6, 8 Hz, H-8), 6.81 (1H, d, J = 8.2 Hz, H-5'), 6.86 (1H, d, J=8.4 Hz, H-5), 7.01 (1H, dd, J=8.2, 2.2 Hz, H-6'), 7.11 (1H, d, J=2 Hz, H-2'), 7.52 (1H, dd, J=8.5, 2.2 Hz, H-6), 7.55 (1H, d, J = 2.3 Hz, H-2), 7.63 (1H, d, J = 15.6 Hz, H-7), 9.52 (1H, d, J = 7.9 Hz, H-9) (Figure S7, Supporting Information). To assign shifts corresponding to protons in the aromatic rings, HSQC data (700 HMz, CD$_3$OD) were used to identify the correlation between H-2 and C-2 (Figure S8, Supporting Information) and HMBC data (700 HMz, CD$_3$OD) were used to identity correlations between C-2 and H-7 and H-6 (Figure S9, Supporting Information). Previous literature reports on this compound were completed in acetone-d_6.[65] To confirm the identity of this compound, an additional ^{1}H NMR (500 MHz, acetone-d_6) was run and chemical shifts matched literature values (Figure S10, Supporting Information).[65]

Figure S1: Calibration curves of standard compounds of berberine (**A**), magnolol (**B**), cryptotanshinone (**C**), and α-mangostin (**D**) using a Thermo-Fisher Q-Exactive Plus Orbitrap mass spectrometer (Thermo Fisher Scientific, MA, USA) connected to an Acquity UPLC system (Waters Corporation, Milford, MA, USA). Separations were completed by using a reversed phase UPLC column (BEH C$_{18}$, 1.7 μm, 2.1 x 50 mm, Waters Corporation, Milford, MA, USA). Calibration curves in B-D were log-log transformed to improve linearity.

Figure S2: Dose-response curves of berberine (**A**), magnolol (**B**), cryptotanshinone (**C**), and α-mangostin (**D**) against *S. aureus* SA1199.[36] Turbidimetric data were obtained by comparing OD_{600} values of test wells relative to vehicle control following 18 hours of incubation at 37 °C. Models were constructed using untransformed triplicate data and fit using four-parameter logistic curves, represented as the mean ± SEM.

Table S1: Effect of data acquisition protocols on selectivity ratio analyses. We assessed the impact of pool number, bioassay concentration, and mass spectral concentration on final biochemometric results by evaluating changes in the selectivity ratio ranking of berberine and magnolol, as well as the impact on false positives identified in the models.

Subset	# Fractions	Conc. tested in bioassay (µg/mL)	Conc. analyzed in MS (mg/mL)	Number of ions included in model (m/z / R_t pairs)	Model Produced? (Y/N)	Number of model components	% independent, % dependent	SR ranking berberine	SR ranking magnolol	# false positive co-varying with berberine	# false positives co-varying with magnolol	Number of false positive not co-varying
1[a]	3	100	0.1	870	Y	4	99.99, 99.92	1	20	2	16	1
2	3	50	0.1	870	N	N/A	N/A	N/A	N/A	N/A	N/A	N/A
3	3	25	0.1	870	Y	5	99.99, 99.95	N/A	14	0	17	0
4	5	100	0.1	870	Y	2	99.38, 84.98	1	14	1	15	0
5	5	50	0.1	870	Y	2	99.37, 86.40	1	12	1	15	0
6	5	25	0.1	870	Y	2	99.38, 84.82	1	14	1	15	0
7	10	100	0.1	870	Y	5	99.79, 98.55	1	8	2	22	0
8	10	50	0.1	870	Y	5	99.79, 82.00	22	4	0	22	0
9	10	25	0.1	870	Y	5	99.81, 88.07	1	13	2	25	8
10	3	100	0.01	370	Y	5	99.98, 100	7	27	0	18	7
11	3	50	0.01	370	N	N/A	N/A	N/A	N/A	N/A	N/A	N/A
12	3	25	0.01	370	N	N/A	N/A	N/A	N/A	N/A	N/A	N/A
13	5	100	0.01	370	Y	4	99.71, 99.83	7	20	0	19	4
14	5	50	0.01	370	Y	3	99.57, 99.76	20	16	0	19	4
15	5	25	0.01	370	Y	3	99.57, 99.73	1	20	0	19	4
16	10	100	0.01	370	N	N/A	N/A	N/A	N/A	N/A	N/A	N/A
17	10	50	0.01	370	Y	2	94.96, 49.17	33	18	0	30	11
18	10	25	0.01	370	Y	3	62.28, 79.11	1	36	0	28	28

[a]Cryptotanshinone correctly identified as contributing to activity (19th). Cryptotanshinone only contributed to activity in the three-pool set.

Table S3: False positives and their distribution in selectivity ratio models.

# Fractions	Concentration tested in bioassay (µg/mL)	Concentration analyzed in MS	Number of ions included in model[a]	Number of ions with selectivity ratio > 0 (% total[b])	% associated with berberine and magnolol[c]	% co-varying false positives[c]	% non-co-varying false positives[c]
3	100	0.1	870	26 (3%)	27%	69%	4%
3	50	0.1	--	--	--	--	--
3	25	0.1	870	20 (2%)	15%	85%	0%
5	100	0.1	870	22 (3%)	28%	72%	0%
5	50	0.1	870	22 (3%)	28%	72%	0%
5	25	0.1	870	22 (3%)	28%	72%	0%
10	100	0.1	870	30 (3%)	20%	80%	0%
10	50	0.1	870	28 (3%)	21%	79%	0%
10	25	0.1	870	41 (5%)	34%	66%	0%
3	100	0.01	370	33 (9%)	25%	55%	20%
3	50	0.01	--	--	--	--	--
3	25	0.01	--	--	--	--	--
5	100	0.01	370	32 (9%)	28%	59%	13%
5	50	0.01	370	32 (9%)	28%	59%	13%
5	25	0.01	370	32 (9%)	28%	59%	13%
10	100	0.01	--	--	--	--	--
10	50	0.01	370	50 (14%)	18%	60%	22%
10	25	0.01	370	65 (18%)	14%	43%	43%

[a] representing unique m/z / R_t pairs
[b] expressed as a percentage of the total number of ions included in model
[c] expressed as a percentage of the total number of ions with selectivity ratio > 0.
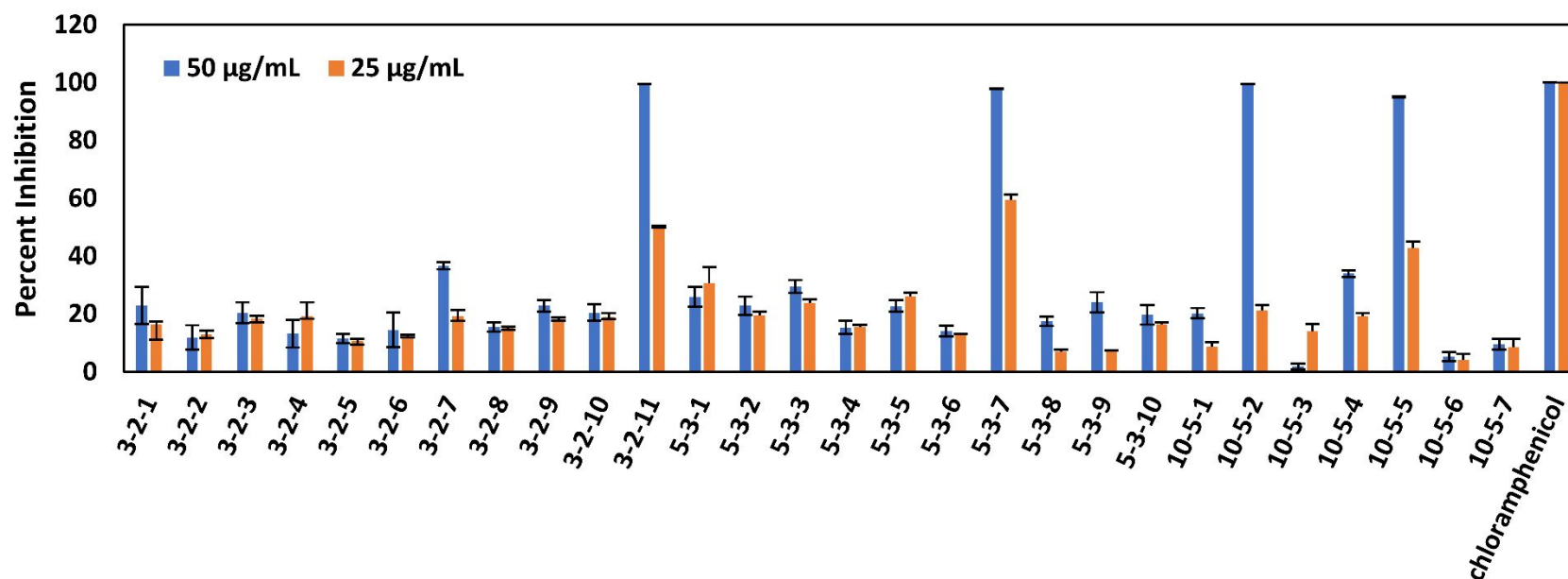
Table S4: Effect of data processing protocols on selectivity ratio analyses. All models contained 870 unique mass/retention time pairs and were produced using data acquired from the ten-pool set analyzed at 100 µg/mL in both the biological assay and during mass spectral analysis.

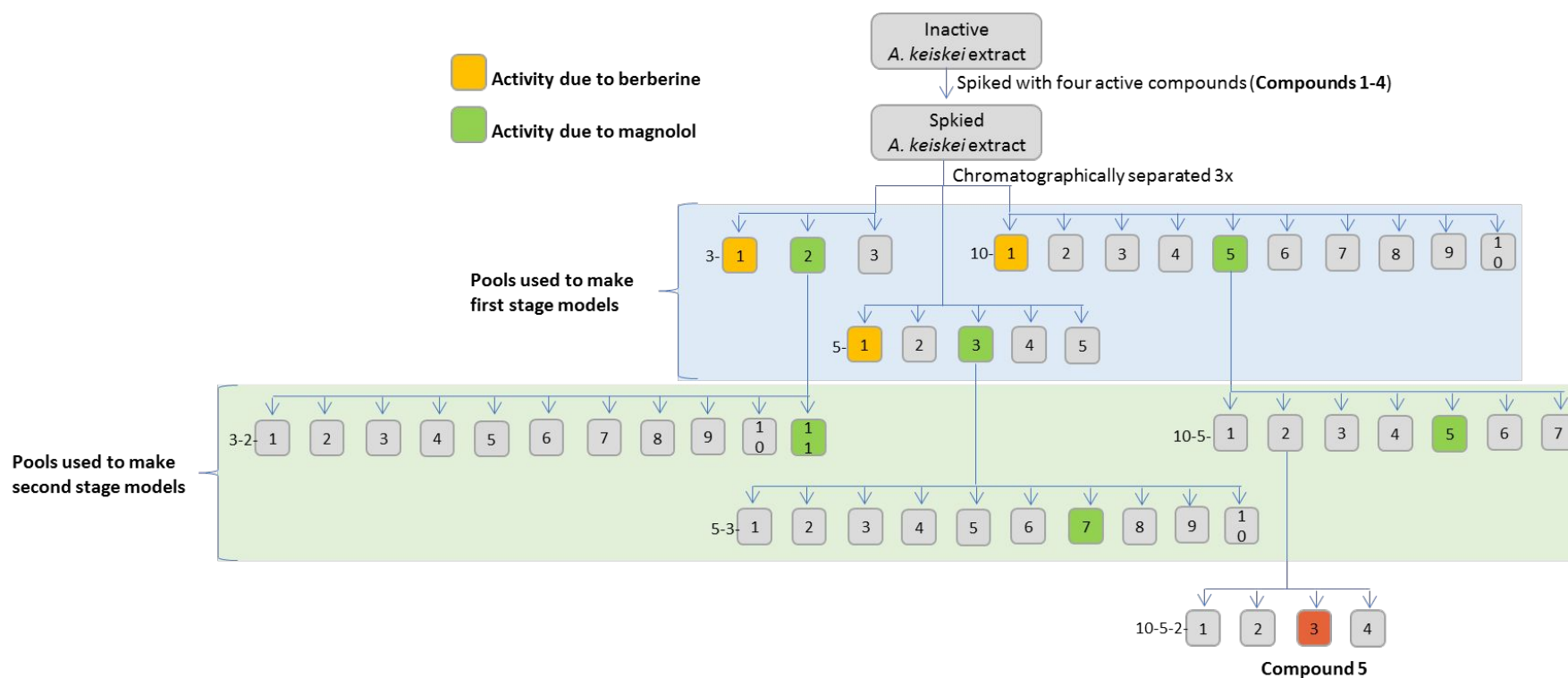| Data Transformation?	Dendrogram Filtering?	Percent Variance Cutoff?	Number of model component	% independent, % dependent	SR ranking berberine	SR ranking magnolol	# false positives co-varying with berberine[a]	Number of false positives co-varying with magnolol[a]	Number of contaminants identified with dendrogram analysis in model[a,b]	Number of false positive not co-varying[a]
N	N	N	5	99.77, 98.71	23	120	13	0	4	27
N	N	Y	5	99.77, 98.71	2	9	3	21	1[c]	0
N	Y	N	5	99.79, 98.55	17	110	20	1	N/A	25
N	Y	Y	5	99.79, 98.55	1	8	2	22	N/A	0
Y	N	N	5	79.90, 99.77	17	213	17	3	2	21
Y	N	Y	5	79.90, 99.77	17	205	17	3	2	21
Y	Y	N	5	81.10,99.75	19	200	18	3	N/A	22
Y	Y	Y	5	81.10,99.75	19	192	19	3	N/A	21

[a] Only top 50 ions were included in this summary

[b] These contaminants were identified and removed using dendrogram filtering, so models that went through dendrogram filtering will not have this type of contaminant in the model

[c] polysiloxane contaminant peak identified as top contributor to bioactivity

Figure S3. Biological activity data of sub-pools resulting from chromatographic separation of pools 3-2, 5-3, and 10-5, which contained active concentrations of magnolol. Growth inhibition of *Staphylococcus aureus* (SA1199)[36] relative to vehicle control was measured turbidimetrically using OD_{600} values. Data presented are the results of triplicate analyses ± SEM. The positive control chloramphenicol was tested at concentrations of 100 and 10 µg/mL.
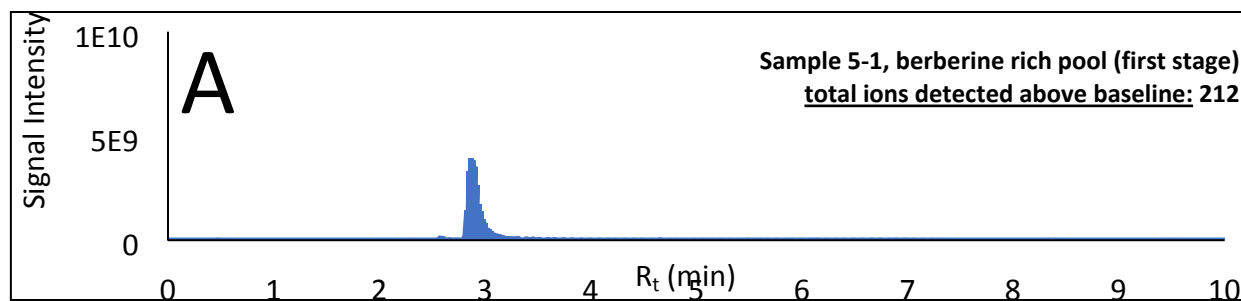
Figure S4: Fractionation Scheme illustrating chromatographic separation of spiked *A. keiskei* extract and subsequent fractionations. Pools used to produce first- and second-stage models have been identified in brackets. Although two additional antimicrobial compounds were spiked into the inactive *A. keiskei* extract (cryptotanshinone and α-mangostin), only berberine and magnolol were concentrated sufficiently to contribute to the activity of chromatographically separated fractions.
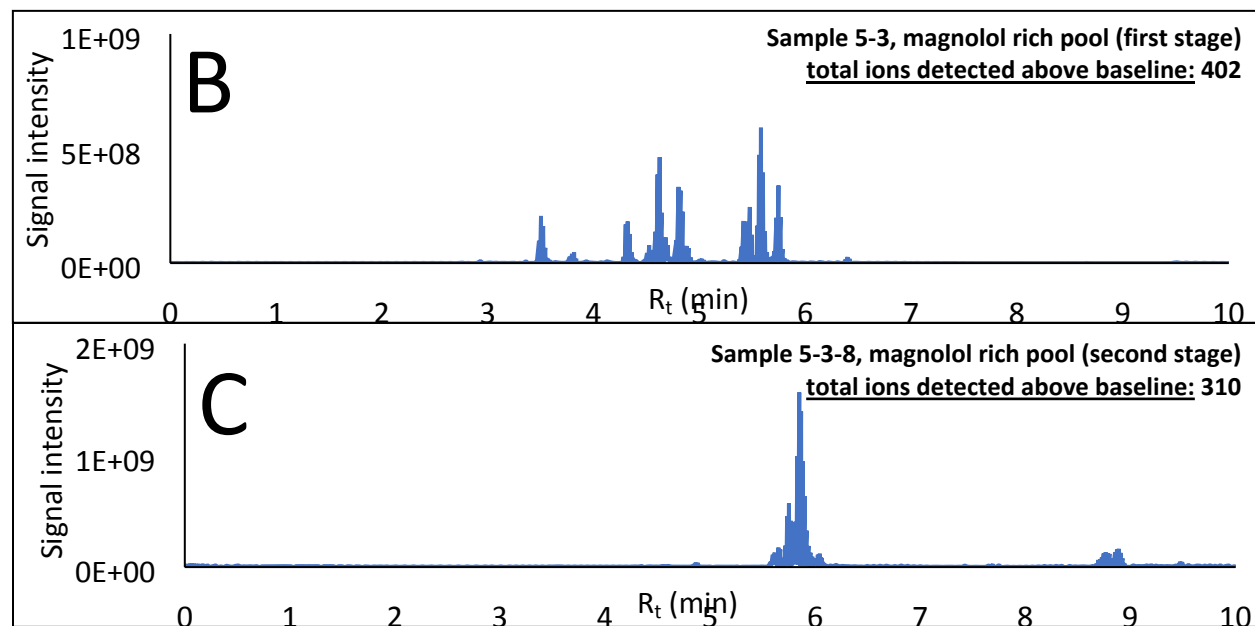
Table S5: Effect of round of fractionation on selectivity ratio analyses.

Round of Fractionation	# Fractions	Concentration tested in bioassay (ug/mL)	Model Produced? (Y/N)	Number of model components	% independent, % dependent	SR ranking magnolol	# false positives co-varying with magnolol [a]	Number of false positive not co-varying [a]
1	3	50	N	N/A	N/A	N/A	N/A	N/A
2	11	50	Y	1	32.62, 86.52	1	18	0
1	3	25	Y	5	99.99, 99.95	14	17	0
2	11	25	Y	1	31.39, 88.97	6	18	0
1	5	50	Y	2	99.37, 86.40	12	13	0
2	10	50	Y	1	43.68, 91.27	1	15	1
1	5	25	Y	2	99.38, 84.82	14	13	0
2	10	25	Y	1	42.97, 72.03	2	16	1
1	10	50	Y	5	99.79, 82.00	4	18	0
2	7	50	Y	2	61.92, 94.10	N/A	6	12 [b]
1	10	25	Y	5	99.81, 88.07	13	10	2
2	7	25	Y	1	36.95, 76.91	4	16	0

[a] only top twenty contributors were considered for this metric

[b] in this case, an unexpected active compound (randainal) was identified as the fifth top contributor to activity. Likely, the activity of this compound was masked by antagonists until this round of fractionation. Nine of the 12 "non-co-varying false positives" actually co-varied with randainal, and only three represented actual false positives that did not co-vary with an active compound.
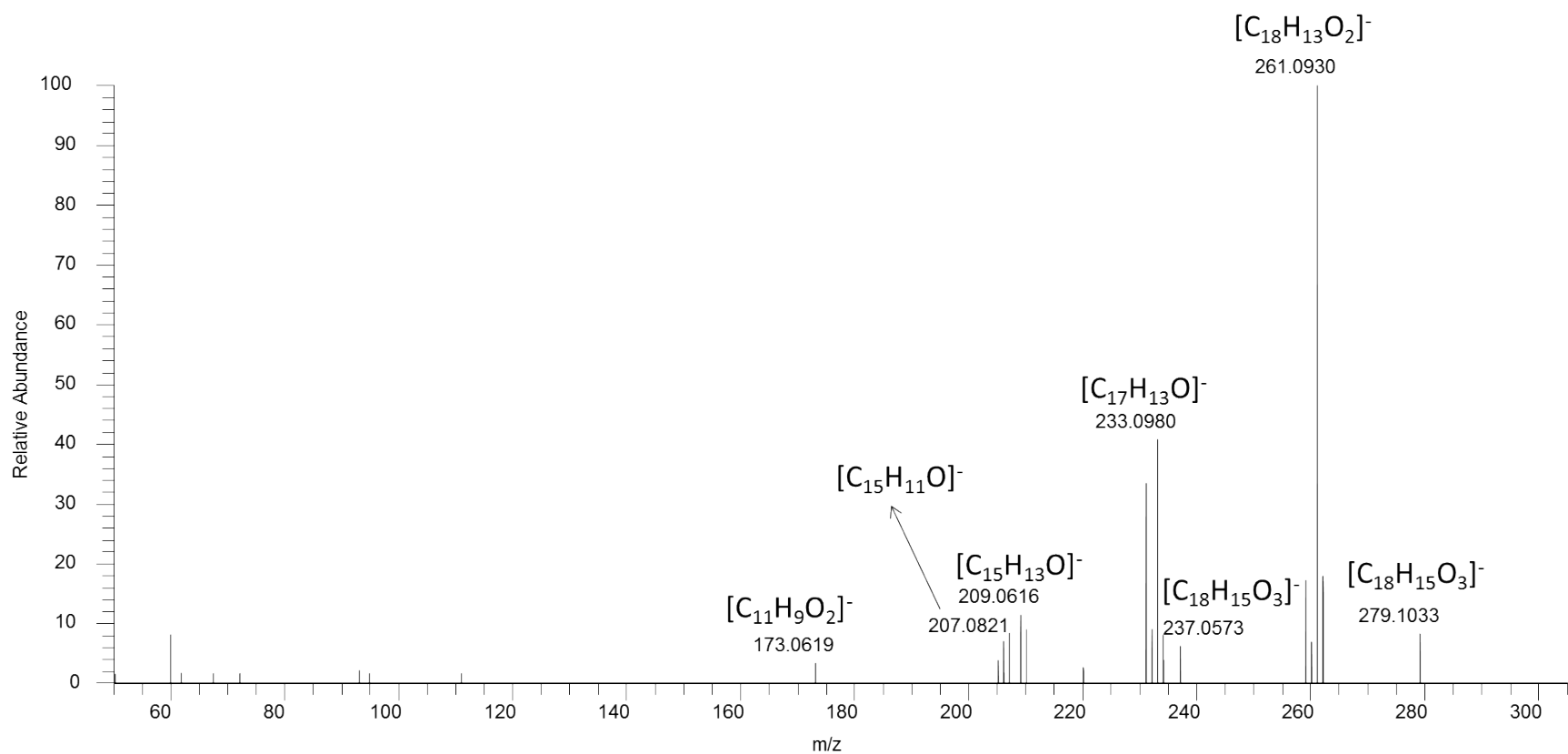
Figure S5: Example chromatograms incorporating both positive- and negative-mode data of selected active pools belonging to the five-pool set analyzed at 0.1 mg/mL in the mass spectrometer. For first-stage pools, baseline cutoffs were set to 2.0×10^6 for positive mode and 1.0×10^6 for negative mode. For second-stage pools, baseline cutoffs were set to 2.0×10^6 for both positive mode and negative mode. **S5A**. Berberine-rich pool from the first round of chromatographic separation. **S5B**. Magnolol-rich pool from first round of chromatographic separation. **S5C**. Magnolol-rich pool from second round of chromatographic separation.

Table S6: Comparison of stage-one models and their identification of randainal among the top contributors to biological activity.
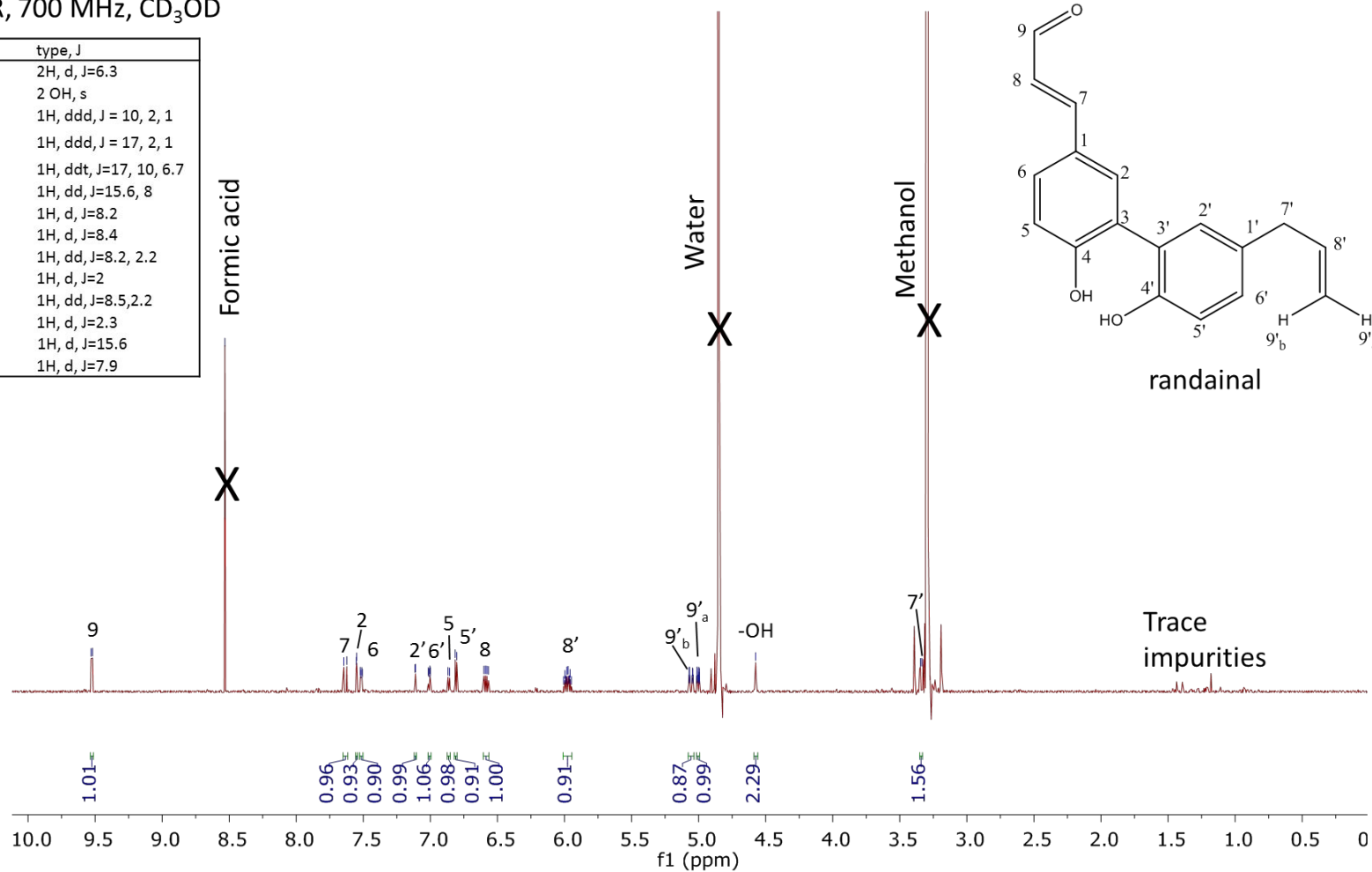
Subset	Round of Fractionation	# Fractions	Conc. tested in bioassay (µg/mL)	Conc. analyzed in MS (mg/mL)	Model Produced? (Y/N)	Number of model components	% independent, % dependent	Did model identify randainal?	SR ranking of randainal
1	1	3	100	0.1	Y	4	99.99, 99.92	Y	23
2	1	3	50	0.1	N	N/A	N/A	N/A	N/A
3	1	3	25	0.1	Y	5	99.99, 99.95	Y	17
4	1	5	100	0.1	Y	2	99.38, 84.98	N	N/A
5	1	5	50	0.1	Y	2	99.37, 86.40	N	N/A
6	1	5	25	0.1	Y	2	99.38, 84.82	N	N/A
7	1	10	100	0.1	Y	5	99.79, 98.55	Y	19
8	1	10	50	0.1	Y	5	99.79, 82.00	Y	14
9	1	10	25	0.1	Y	5	99.81, 88.07	Y	25
10	1	3	100	0.01	Y	5	99.98, 100	N	N/A
11	1	3	50	0.01	N	N/A	N/A	N/A	N/A
12	1	3	25	0.01	N	N/A	N/A	N/A	N/A
13	1	5	100	0.01	Y	4	99.71, 99.83	N	N/A
14	1	5	50	0.01	Y	3	99.57, 99.76	N	N/A
15	1	5	25	0.01	Y	3	99.57, 99.73	N	N/A
16	1	10	100	0.01	N	N/A	N/A	N/A	N/A
17	1	10	50	0.01	Y	2	94.96, 49.17	N	N/A
18	1	10	25	0.01	Y	3	62.28, 79.11	N	N/A
19	2	11	50	0.1	Y	1	32.62, 86.52	Y	50
20	2	11	25	0.1	Y	1	31.39, 88.97	Y	49
21	2	10	50	0.1	Y	1	43.68, 91.27	N	N/A
22	2	10	25	0.1	Y	1	42.97, 72.03	N	N/A
23	2	7	50	0.1	Y	2	61.92, 94.10	Y	5
24	2	7	25	0.1	Y	2	62.68, 86.41	N	N/A

Figure S6: MS^2 spectrum (negative mode) of randainal. Peaks have been labeled with molecular formulas if they match fragment predictions and/or fragments previously reported in the literature.[64]
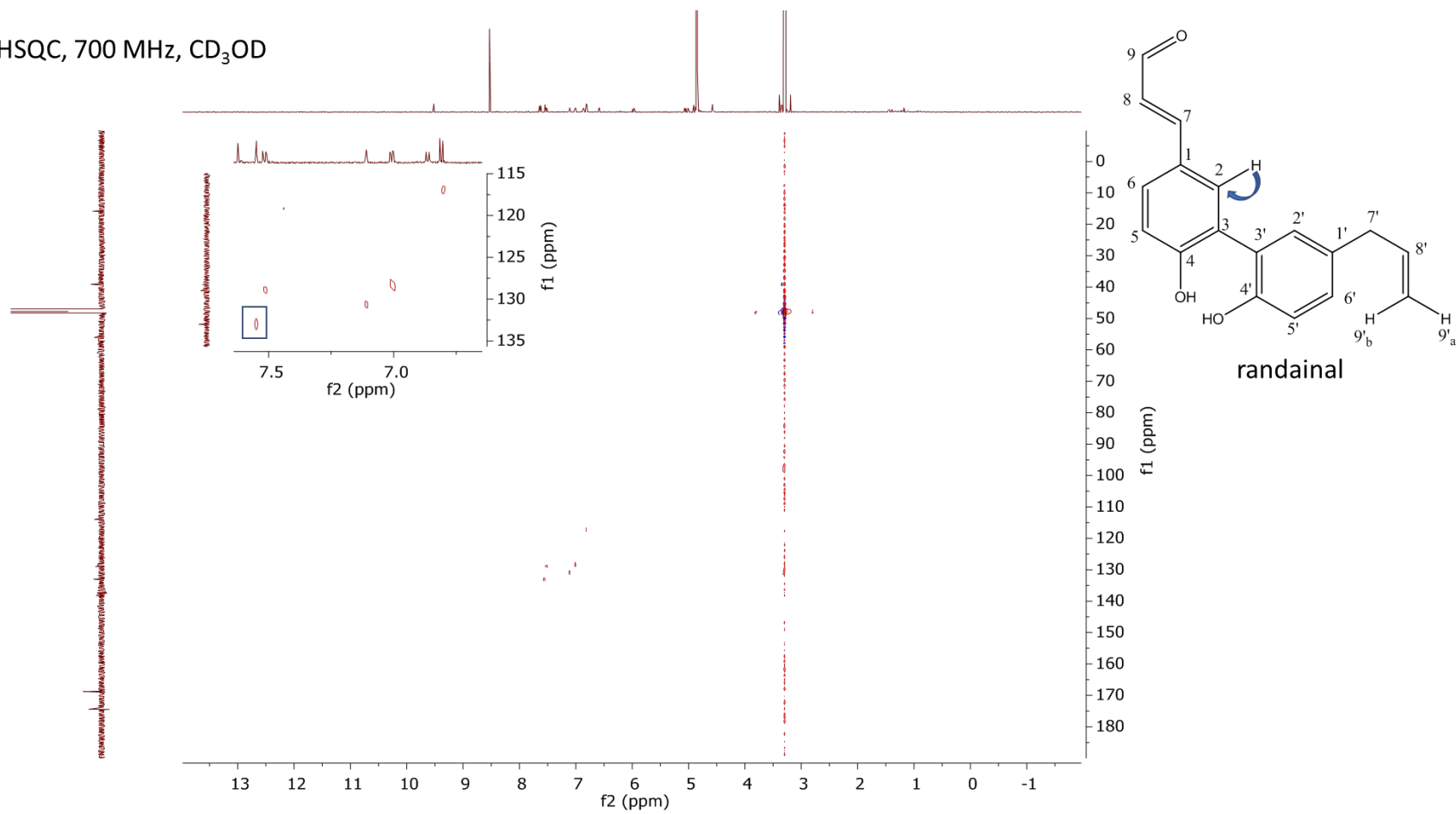
Figure S7: ^{1}H NMR spectrum (700 MHz, CD$_3$OD) of randainal
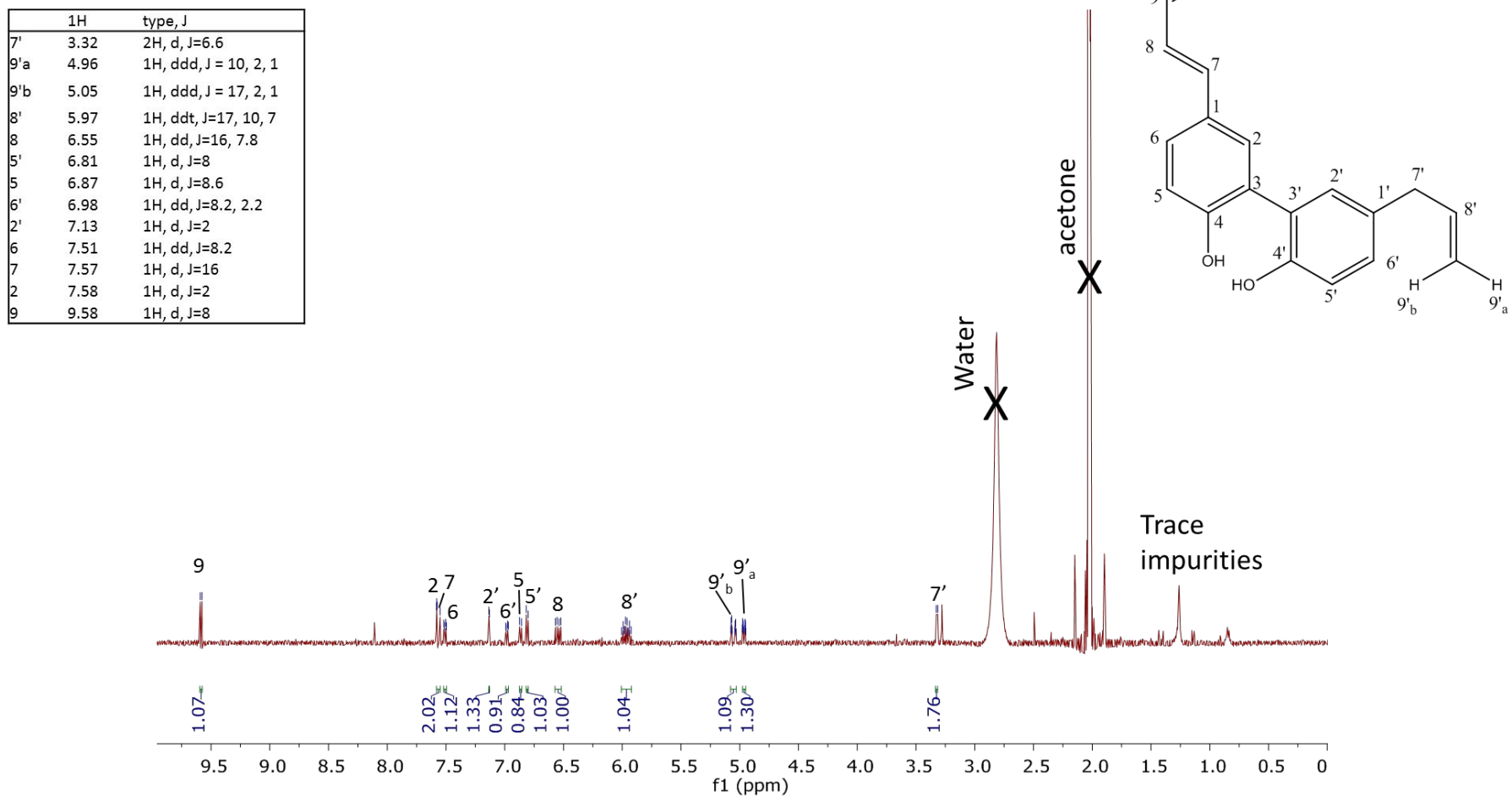
Figure S8: HSQC spectrum (700 MHz, CD$_3$OD) of randainal

Figure S9: HMBC spectrum (700 MHz, CD₃OD) of randainal

¹H NMR, 500 MHz, acetone-d_6

	1H	type, J
7'	3.32	2H, d, J=6.6
9'a	4.96	1H, ddd, J = 10, 2, 1
9'b	5.05	1H, ddd, J = 17, 2, 1
8'	5.97	1H, ddt, J=17, 10, 7
8	6.55	1H, dd, J=16, 7.8
5'	6.81	1H, d, J=8
5	6.87	1H, d, J=8.6
6'	6.98	1H, dd, J=8.2, 2.2
2'	7.13	1H, d, J=2
6	7.51	1H, dd, J=8.2
7	7.57	1H, d, J=16
2	7.58	1H, d, J=2
9	9.58	1H, d, J=8

Figure S10: ¹H NMR spectrum (500 MHz, acetone-d_6) of randainal. Spectra match those previously reported in the literature.[65]