Unbiased Evaluation of Bioactive Secondary Metabolites in Complex Matrices

Taichi Inui[a,b,1], Yuehong Wang[a], Samuel M. Pro[a,2], Scott G. Franzblau[a], Guido F. Pauli[a,b,*]

Supplementary Material

[a]Institute for Tuberculosis Research and [b]Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, Chicago, Illinois 60612, USA
*Corresponding author at: Institute for Tuberculosis Research, College of Pharmacy, University of Illinois at Chicago, 833 S. Wood St., Chicago, Illinois 60612, USA, tel.: +1-312-355-1949, fax.: +1-312-355-2693, e-mail: gfp@uic.edu
[1]T. Inui (current address): Wm. Wrigley Jr. Company, Chicago, Illinois 60642, USA
[2]S. Pro (current address): Cherry Instruments, Chicago, Illinois 60614, USA

Table S1. Subgroups of GC Chromatograms. Each of the seven subgroups consisted of several fractions as shown. The number of fractions contained in each subgroup impacted the threshold used to determine positive matches (assignment of bioactives) in the next Stage 4 (See Table S2). The Target fraction was the fraction chosen to generate the reference chromatogram for the GC peak alignment in any given subgroup.

Subgroup	Fractions	# of fractions	Biopeaks	Target fraction#
1	1-12	12	1-3	7
2	13-18	6	4, 5	17
3	19-23	5	6	21
4	25-32	9	7-9	27
5	34-39	6	10-12	37
6	40-47	8	13, 14	42
7	49-64	16	15-19	56

Table S2. Thresholds for Positive Correlation. The number n stands for the number of data points that make up the each of the biopeaks. The threshold was set to satisfy the lower limit of confidence interval for the population coefficient (ρ, 95 %) to be 0.5.

Biopeak	1-3	4, 5	6	7-9	10-12	13, 14	15-19
n	12	6	5	9	6	8	16
Threshold	0.834	0.933	0.959	0.874	0.933	0.891	0.798

Table S3. Bioactive Secondary Metabolites of *O. horridus* identified through MS database matching. From left: compound number; compound name; the *K* value of the CCC fraction in which the compound eluted; GC retention time in minutes; correlation coefficient; mass spectral matching factor (MF) of rank 1, δMF_1 as MF_1-MF_2; the probability that the candidate from the database is the unknown compound assuming the unknown is included in the database (Cp), GC peak area of the compound relative to the entire GC peak area of the run, in percent.

#	Name	K_D	GC rt (min)	r	MF_1	δMF_1	Probability %	Area%
1	Oplopandiol	0.60	20.109	0.96	870			15.7
2	Falcarindiol	0.60	19.916	0.91	867		-	1.0
3	3,10-epoxy-3,7,11-trimethyldodeca-1,6-dien-11-ol	2.15	16.902	0.53	857		-	7.6
4	7,10-epoxy-3,7,11-trimethyldodec-1-ene-3,6,11-triol	2.33	13.295	0.47	871		-	0.3
5	Sesamin	0.44	31.903	0.32	910		-	9.0
6	2-Octenoic acid	0.78	8.565	0.97	906	31	53.3	0.4
7	*n*-Decanoic acid	0.21	10.771	0.99	862	38	91.9	1.4
8	2,4-Undecadien-1-ol	0.44	10.967	0.99	833	2	40.5	0.2
9	9-Oxononanoic acid	1.48	12.305	0.97	874	30	53.5	0.8
10	9-Acetoxynonanal	2.51	12.770	0.94	845	43	81.1	2.9
11	Z,Z-10.12-Hexadecadien-1-ol acetate	0.44	20.063	0.97	868	11	42.6	1.3
12	Falcarinol	0.60	18.919	0.88	819	10	40.4	0.6
13	1-Monoolein	0.17	23.970	0.97	866	28	67.6	4.2
14	1-(1-Methylcyclopentyl)-ethanone	0.17	9.076	0.99	859	19	54.3	0.0
15	2-Hexylfuran	0.17	10.276	0.99	829	22	53.7	0.0
16	7-(2-Furyl)heptanoate	0.60	14.470	0.90	803	19	51.5	0.9
17	1-Cyclopentoxycarbonyl-3-cyclopentylpropionic acid	0.64	14.474	0.93	811	25	59.4	1.8
18	1,3-Dimethyl-(3,7-dimethyloctyl)cyclohexane	0.99	16.907	0.89	884	16	44.4	0.9
19	Cadinol	0.17	12.926	0.98	883	12	43.6	0.3
20	5,9-Dimethyl-2-(1-methylethyldene)-1-cyclodecanone	0.99	13.097	0.88	849	7	38.2	0.8
21	3,9(11)-Diene-10-peroxy-murolan	1.48	17.109	0.98	866	12	48.6	0.9
22	CAS 109897-57-6	2.33	13.286	0.88	863	10	41.5	1.0
23	3,3,8,8-Tetramethyl-tricyclo[5,1,0,0(2,4)]oct-5-ene-5-propenoic acid	1.88	17.612	0.88	847	13	38.5	1.3
24	3-Hydroxy-4,4-dimethyl-3-(1-methyl-3-oxobut-1-enyl)cycloheptanone	0.44	15.640	0.98	840	9	37.2	0.6
25	Ledol	0.78	13.866	0.97	833	23	52.4	0.9
26	Glaucyl alcohol	0.99	17.350	0.87	859	46	76.7	1.7
27	Spathulenol	2.33	15.613	0.88	895	28	65.8	5.8
28	Isoaromadendrene epoxide	1.48	15.915	0.99	826	30	63.3	7.7
29	Dihydroartemisinin	2.33	18.145	0.88	807	27	64.5	2.5
30	Un-identified polyketide diyne	2.33	24.090	0.90	-	-	-	7.8