Accurate Mass Spectral Database: Harnessing the Power of High Performance Mass Spectrometry at Long Last

LORNE FELL[1]; Viatcheslav Artaev[1]; Kevin McNitt[1]; Steve Robles[1]; Albert Lebedev[2] | [1]LECO Corporation, Saint Joseph, MI; [2]Moscow State University, Moscow, Russian Federation

Introduction

The use of mass spectral databases have been a foundation of identification tools for mass spectrometry. Their use and development has enabled GC/MS to be a dominant and routine technique in many applications of hyphenated MS technologies. Over the past several years the use of high resolution (accurate mass) mass spectrometry and its data expanded into all areas of applications, which could take advantage of spectral databases specifically developed to exploit novel instrumentation capabilities. In this presentation we describe the development and use of an accurate mass-based spectral database, its curation, and application of novel algorithms resulting in similarity searches which are improved by the information obtained by high performance GC/MS.

Methods

Standard mixtures comprising of alkanes, PAHs, semivolatiles, and pesticides were analyzed using a high resolution time-of-flight (HRTOF) mass spectrometer—Pegasus GC-HRT (LECO Corporation, Saint Joseph, MI)—at 10 spectra-per-second (m/z 40–300) in high resolution mode (25,000 at FWHH). The resulting chromatographic peaks were automatically found, deconvoluted, and curated into an Accurate Mass Library (AML). Typical mass accuracies were <1 ppm. In order to take advantage of this high mass accuracy data, a novel matching algorithm (AML Rank) was developed to improve the identification process. The AML and its ranking algorithm was compared to standard nominal mass libraries for a number of environmental and food samples. Many MS vendors have begun to develop accurate mass libraries (some of them being called databases) in many different forms. In fact, even the highly influential NIST has modified its programs to incorporate aspects of accurate mass in its search engine, but most uses of this high performance data are done solely with formulae assigned to the structure. Formula use is helpful and correct for all cases of non-targeted screening, however, accurate mass data can also be used when the chemical formula is unknown, by answering the question, "How close to the accurate mass is the candidate ion?". Herein we describe the use of a novel algorithm which ranks deconvoluted accurate mass spectra against a library of curated accurate spectra. The complex samples used in this study were not always allowed to reach sufficient separation to exclude any possible ion interferences from the coeluted analytes. Such interferences could significantly reduce similarity scores of the analytes of interest and put them outside of the top hits in the library search results if applying a standard NIST algorithm. High performance data compared to low performance libraries will not yield the best results. However, this accurate mass algorithm, using a ranking system to sort the hit list, dramatically improved the results and filtered the correct hits to the top of the search results even in the presence of heavily interfering ions. Examples of the results with various forms of interferences are presented.

AML Ranking

Accurate mass library rank is a measure of how close the masses align between two spectra. The spectral masses are paired if they have overlapping mass confidence intervals which are based on the acquisition resolution. Each pair is scored by the mass difference relative to a mass tolerance value. The final score for the spectra is the sum of the scores for each pair, weighted by the sum of the abundance of both masses in the pair. The score is not affected by the difference in abundance between the matching masses.

The NIST similarity score is based on the relative abundances of the matched pairs of masses, and weighs them based on MW and the abundance ratios of adjacent matching peaks. There is no comparison of accurate masses because the masses are nominal. Therefore AML rank and NIST similarity are independent metrics, each scoring based on different spectral characteristics.

Results

Megamix

New AML was applied to several standard mixture solutions for validating. In case of volatile compounds (8260B MegaMix Calibration Mix, Restek, USA) 46 compounds of the mixture were in the library at the time of the analysis. Among them 27 had 1st position (hit 1–59%) in the hit list, 6 in the 2nd position (hit 2–13%), and 3 in the 3rd position (hit 3–6%). In total, 78% were in the top 3 hits. Ten compounds were in top 10 hits of the library search hit list (see Table 1).

For the semivolatile compounds the results were comparable, while some results were extraordinary. For benz[a]anthracene the score using AML reached 1000 points, which is quite extraordinary for any library in general (Table 2). Very good identification results were obtained for isomeric compounds. Table 3 shows an example of identification of isomeric anthracene and phenanthrene.

Table 1. Library search results of MegaMix using AML.

Peak #	Name	R.T. (s)	position	Base Mass
1	2,2-dichloropropane	133.528	1	77.0179
2	2-Propenoic acid, methyl ester	134.368	absent	55.0179
3	Methane, bromochloro-	135.04	absent	129.9
6	Tetrahydrofuran	152.848	1-caliper (3)	96.9606
7	Ethane, 1,1,1-trichloro-	152.848	1	96.9606
8	Ethane, 1,2-dichloro-	155.62	1	61.9918
9	1-Propene, 1,1-dichloro-	160.492	absent	74.9996
12	Carbon Tetrachloride	165.196	1	116.906
22	Propane, 1,2-dichloro-	200.896	1	62.9997
23	Trichloroethylene	201.484	1	129.9138
24	Methane, dibromo-	204.172	1	173.8496
25	Methane, bromodichloro-	210.388	Absent	82.945
28	Methyl methacrylate	220.048	2	69.0336
30	1-Propene, 1,3-dichloro-, (Z)-	261.208	Absent	74.9997
31	Toluene	310.516	3	91.0542
32	1-Propene, 1,3-dichloro-, (E)-	311.44	8	74.9996
34	Ethane, 1,1,2-trichloro-	322.024	1	96.9607
35	Propane, 1,3-dichloro-	348.148	Absent	76.0074
36	Methacrylic acid, ethyl ester	355.96	1	69.0335
38	Methane, dibromochloro-	364.696	absent	128.8922
40	Ethane, 1,2-dibromo-	383.26	1	106.9491
41	Tetrachloroethylene	389.476	2	165.8719
44	Benzene, chloro-	456.624	5(1)	112.0074
46	Ethane, 1,1,1,2-tetrachloro-	456.844	1	130.9216
48	Ethylbenzene	472.972	1	91.0542
53	p-Xylene	483.22	absent	91.0542
54	Methane, tribromo-	500.692	2	172.8418
57	Styrene	508.42	2	104.0621
58	p-Xylene	510.436	absent	91.0542
61	2-Butene, 1,4-dichloro- (Z)-	524.632	1	53.0387
63	Ethane, 1,1,2,2-tetrachloro-	531.856	1	82.9449
66	Propane, 1,2,3-trichloro-	538.96	1	74.9996
68	Benzene, (1-methylethyl)-	543.112	1	105.0698
71	2-Butene, 1,4-dichloro-, (E)-	547.144	1	74.9996
72	Benzene, bromo-	548.068	5	77.0386
75	Benzene, 1-chloro-2-methyl-	566.8	9	91.0542
78	Benzene, propyl-	569.068	1	91.0542
79	Benzene, 1-chloro-4-methyl-	571.588	2(1)	91.0542
82	Mesitylene	581.08	1	105.0699
83	Ethane, pentachloro-	586.708	1	166.8796
84	Benzene, (1-methylethyl)-	590.32	4	105.0701
85	Benzene, tert-butyl-	599.812	1	119.0854
86	Benzene, 1,2,4-trimethyl-	600.484	1	105.0699
88	Benzene, 1,3-dichloro-	609.136	10	145.9683
92	Benzene, (1-methylpropyl)-	612.832	1	105.0698
94	Benzene, 1,4-dichloro-	613.924	10	145.9684
96	p-Cymene	622.24	1	119.0855
97	Benzene, 1,2-dichloro-	629.716	10	145.9683
98	Benzene, butyl-	642.736	2	91.0542
100	Ethane, hexachloro-	654.916	1	165.8721
101	Propane, 1,2-dibromo-3-chloro-	660.124	3	156.9235
103	Benzene, nitro-	662.728	1	77.0387
114	Benzene, 1,2,3-trichloro-	712.624	3(9)	179.9294
117	Naphthalene	716.632	8	128.0619
118	1,3-Butadiene, 1,1,2,3,4,4-hexachloro-	720.016	1	224.8406
124	Benzene, 1,2,3-trichloro-	731.02	2	179.9294
125	1,3-Butadiene, 1,1,2,3,4,4-hexachloro-	731.776	3	224.8407

Table 2. Library search results for benz[a]anthracene using AML.

Hit Table - 8270-2D-DCM-Modulation5s-split100-DCM_3 | Data Process Monitor | Chromatogram | Peak Table - 8270-2D-DCM-Modulation5s-split100-DCM_3

Hit	Name	Similarity	Mass	Reverse	Probability	CAS	Library	Id
>1*	Benz[a]anthracene	1000		1000	0.0	56-55-3	Lebedev-HRT-Library	170
2	Chrysene	931		937	0.0	218-01-9	Lebedev-HRT-Library	171
3	Benz[a]anthracene	901		903	41.3	56-55-3	mainlib	99998
4	Benz[a]anthracene	886		887	30.9	56-55-3	replib	30277
5	Triphenylene	884		885	22.5	217-59-4	mainlib	99999
6	Triphenylene	884		884	28.5	217-59-4	replib	30272
7	Triphenylene	884		886	28.5	217-59-4	replib	30271
8	Naphthacene	879		882	23.0	92-24-0	replib	30256
9	Triphenylene	878		880	28.5	217-59-4	replib	30270
10	Naphthacene	873		873	15.4	92-24-0	mainlib	99821

Table 3. Library search results for anthracene using AML.

Hit Table - 8270-2D-DCM-Modulation5s-split100-DCM_3 | Data Process Monitor | Chromatogram | Peak Table - 8270-2D-DCM-Modulation5s-split100-DCM_3

Hit	Name	Similarity	Mass	Reverse	Probability	CAS	Library	Id
>1*	Anthracene	988		988	0.0	120-12-7	Lebedev-HRT-Library	167
2	Phenanthrene	951		957	0.0	85-01-8	Lebedev-HRT-Library	166
3	Phenanthrene	939		942	57.2	85-01-8	mainlib	65969
4	Anthracene	935		938	48.6	120-12-7	replib	26710
5	Phenanthrene	917		918	25.1	85-01-8	replib	26708
6	Anthracene	915		920	48.6	120-12-7	replib	26704
7	Phenanthrene	914		915	25.1	85-01-8	replib	26716
8	9H-Fluorene, 9-methylene-	906		958	17.2	4425-82-5	replib	26706
9	Anthracene	904		906	48.6	120-12-7	replib	26705
10	Anthracene	902		903	13.9	120-12-7	mainlib	66037

Water Analysis

Surprisingly good results were obtained for aliphatic compounds identified in complex environmental samples. Usually, a search using standard mass spectra of aliphatic compounds results in a hit list including homologous compounds (aldehydes, alkanes, naphthenes, alcohols), where the correct compound may have rather low ranking. AML library search often provides the correct analyte as the best hit. For example, the AML library search results of hexadecane and tetradecanol-1 show those compounds as #1 hits (Table 4,5).

Table 4. Library search results for hexadecane using AML (water sample).

Hit	Name	Expected Ion m/z	Similarity	Reverse	CAS	Library	Formula
1	Hexadecane	226.2655	935	945	544-76-3	Lebedev-HRT-Library	$C_{16}H_{34}$
2	Hexadecane	226.2655	919	921	544-76-3	replib	$C_{16}H_{34}$
3	Hexadecane	226.2655	911	911	544-76-3	replib	$C_{16}H_{34}$
4	Heptadecane	240.2812	909	912	629-78-7	replib	$C_{17}H_{36}$
5	Hexadecane	226.2655	908	910	544-76-3	replib	$C_{16}H_{34}$
6	Pentadecane	212.2499	901	904	629-62-9	replib	$C_{15}H_{32}$
7	Tetradecane	198.2342	899	908	629-59-4	replib	$C_{14}H_{30}$
8	Hexadecane	226.2655	896	896	544-76-3	mainlib	$C_{16}H_{34}$
9	Tetradecane	198.2342	896	900	629-59-4	replib	$C_{14}H_{30}$
10	Nonadecane	268.3125	895	901	629-92-5	replib	$C_{19}H_{40}$

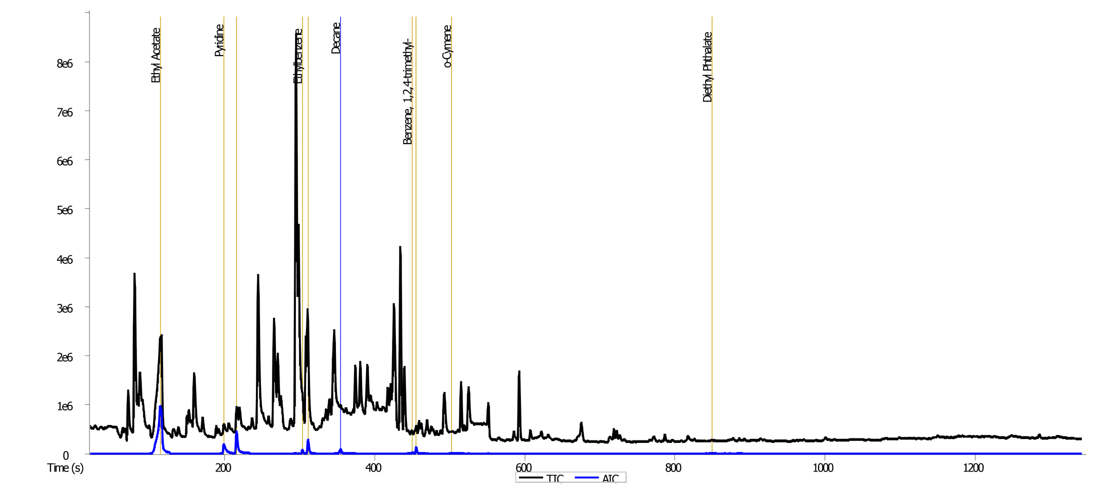
Table 5. Library search results for tetradecanol-1 using AML (water sample).

Hit	Name	Expected Ion m/z	Similarity	Reverse	Library	CAS	Formula
1	1-Tetradecanol	214.2291	897	920	Lebedev-HRT-Library	112-72-1	$C_{14}H_{30}O$
2	1-Dodecanol	186.1978	890	927	replib	112-53-8	$C_{12}H_{26}O$
3	1-Hexadecanol	242.2604	888	914	mainlib	36653-82-4	$C_{16}H_{34}O$
4	1-Tetradecanol	214.2291	888	929	replib	112-72-1	$C_{14}H_{30}O$
5	1-Tetradecanol	214.2291	877	889	replib	112-72-1	$C_{14}H_{30}O$
6	1-Hexadecanol	242.2604	876	910	replib	36653-82-4	$C_{16}H_{34}O$
7	1-Tetradecanol	214.2291	876	886	replib	112-72-1	$C_{14}H_{30}O$
8	Cyclotetradecane	196.2186	875	913	replib	295-17-0	$C_{14}H_{28}$
9	1-Dodecanol	186.1978	871	879	replib	112-53-8	$C_{12}H_{26}O$
10	1-Undecanol	172.1822	870	913	mainlib	112-42-5	$C_{11}H_{24}O$

Food and Fragrance Analysis

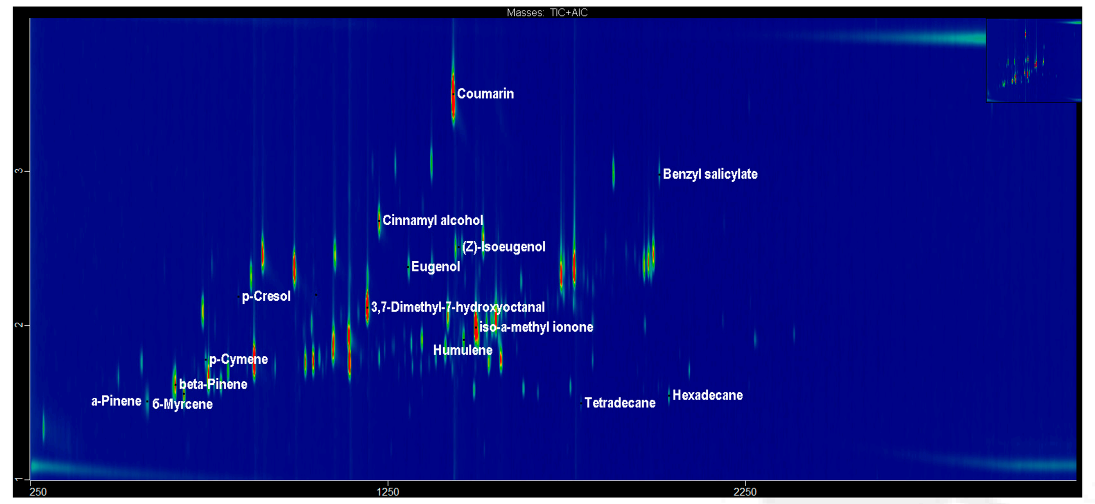
The AML search was successfully implemented to the food and fragrance samples obtained using GC-HRT and GCxGC-HRT.

- **Extra Virgin Olive Oil GC-HRT**
  - Shown for only those species that match for AML library.



Peak #	Name	Similarity	AML Rank	Hit #	Library	Mass Accuracy (ppm)	R.T. (s)	Peak S/N
18	Ethyl Acetate	820	979	1	Lebedev-HRT-Library	-0.36	114.0	2904
44	Pyridine	927	926	1	Lebedev-HRT-Library	-0.39	198.7	7770
52	Toluene	910	919	1	Lebedev-HRT-Library	0.47	215.6	2160
85	Ethylbenzene	816	881	1	Lebedev-HRT-Library	1.06	303.4	335
89	o-Xylene	904	861	1	Lebedev-HRT-Library	1.16	310.9	556
110	Decane	768	664	1	Lebedev-HRT-Library	N/A	354.5	132
162	Benzene, 1,2,4-trimethyl-	762	814	1	Lebedev-HRT-Library	0.58	450.1	256
165	D-Limonene	861	851	1	Lebedev-HRT-Library	0.17	455.1	233
186	o-Cymene	651	776	1	Lebedev-HRT-Library	0.67	502.1	52
286	Diethyl Phthalate	802	828	1	Lebedev-HRT-Library	N/A	849.5	168
322	Dodecane	694	761	1	Lebedev-HRT-Library	N/A	1243.6	56

- **Perfume Sample GCxGC HRT**
  - Shown only for those species that match for AML



Peak #	Name	Similarity	AML Rank	Hit #	Library	Mass Accuracy (ppm)	R.T. (s)	Peak S/N
27	a-Pinene	875	893	1	Lebedev-HRT-Library	-0.92	577, 1.510	309
29	beta-Pinene	928	881	1	Lebedev-HRT-Library	0.73	655, 1.620	1808
31	6-Myrcene	890	878	1	Lebedev-HRT-Library	N/A	679, 1.560	946
36	p-Cymene	849	937	1	Lebedev-HRT-Library	0.75	739, 1.780	428
40	(Z)-Ocimene	843	929	1	Lebedev-HRT-Library	N/A	763, 1.630	264
42	(Z)-Ocimene	902	932	1	Lebedev-HRT-Library	N/A	784, 1.630	388
46	p-Cresol	854	882	1	Lebedev-HRT-Library	0.77	832, 2.189	91
69	4-Allylanisole	741	850	1	Lebedev-HRT-Library	-0.77	1048, 2.195	35
91	3,7-Dimethyl-7-hydroxyoctanal	900	926	1	Lebedev-HRT-Library	N/A	1192, 2.120	1942
96	Cinnamyl alcohol	902	916	1	Lebedev-HRT-Library	-0.63	1225, 2.680	682
108	Eugenol	872	910	1	Lebedev-HRT-Library	1.22	1306, 2.380	336
130	Coumarin	943	904	1	Lebedev-HRT-Library	-2.58	1435, 0.500	3323
136	(Z)-Isoeugenol	798	848	1	Lebedev-HRT-Library	1.43	1447, 2.510	134
142	Humulene	857	922	1	Lebedev-HRT-Library	0.59	1462, 1.910	313
153	iso-a-methyl ionone	909	871	1	Lebedev-HRT-Library	-1.83	1495, 1.990	2514
228	Tetradecane	865	823	1	Lebedev-HRT-Library	N/A	1789, 1.500	63
255	Benzyl salicylate	874	921	1	Lebedev-HRT-Library	1.33	2008, 2.985	470
256	Hexadecane	883	817	1	Lebedev-HRT-Library	N/A	2035, 1.550	324

Conclusions

An Accurate Mass Library was created and used with an AML ranking algorithm to search data obtained with high resolution GC-MS and GCxGC-MS instrumentation.

Delivering the Right Results