UNIVERSITY OF ALBERTA

TMIC
The Metabolomics Innovation Centre

Leveraging Local Libraries for Positively Parsimonious Peak Tables

James J. Harynuk
Department of Chemistry University of Alberta, Edmonton, Canada

Where do we use GC×GC-MS?

Petroleum

Foods / Flavours

Environment

Omics

Are there any compounds in my sample that are…

… harmful?

… interesting?

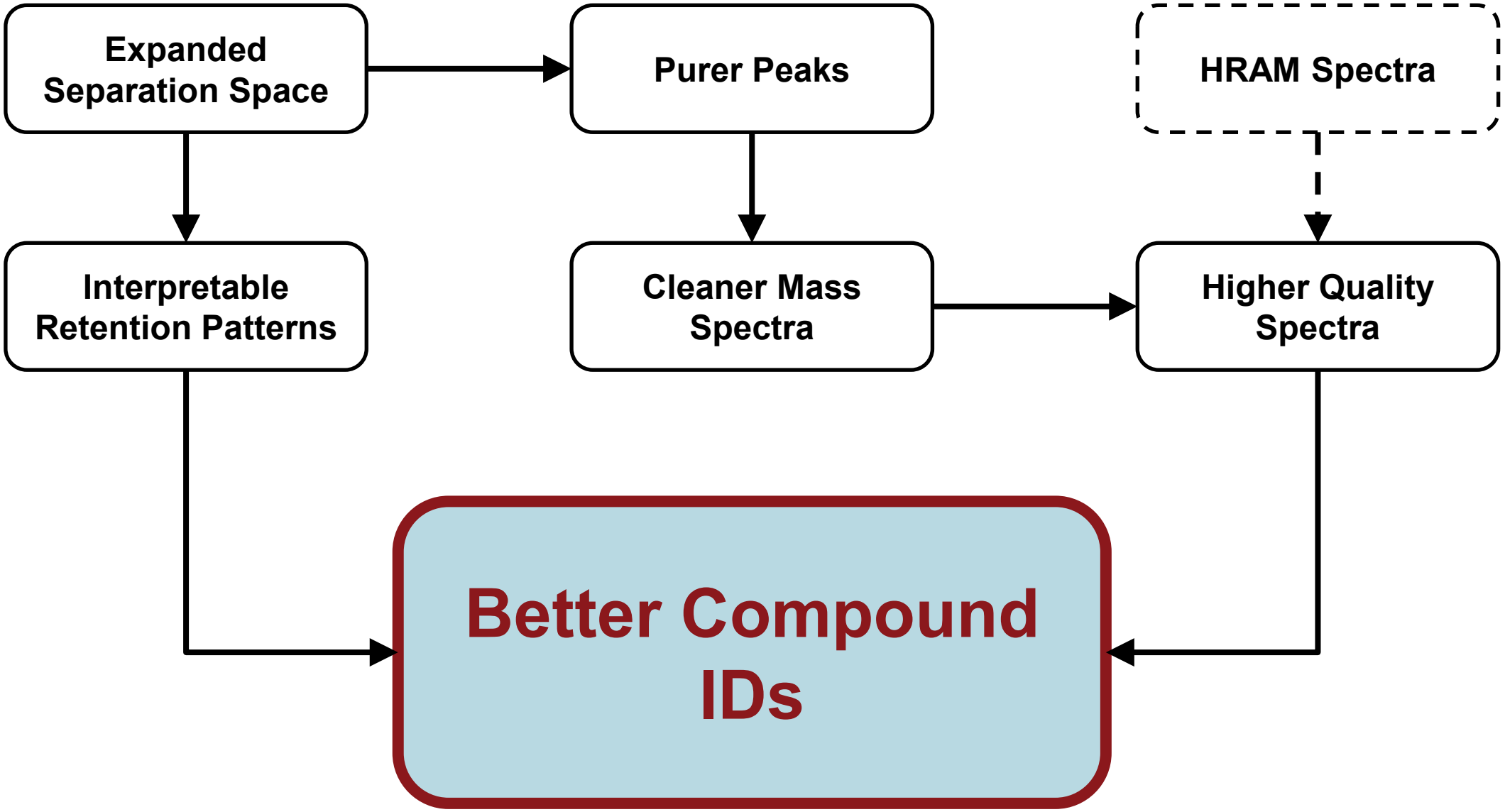
… indicative of exposure?

… important?

… regulated?

Images by Szabolcs Molnar, Devon Breen, fernando zhiminaicela from Pixabay

GC×GC-(HR)MS is perfect for this



An old story…

Have:

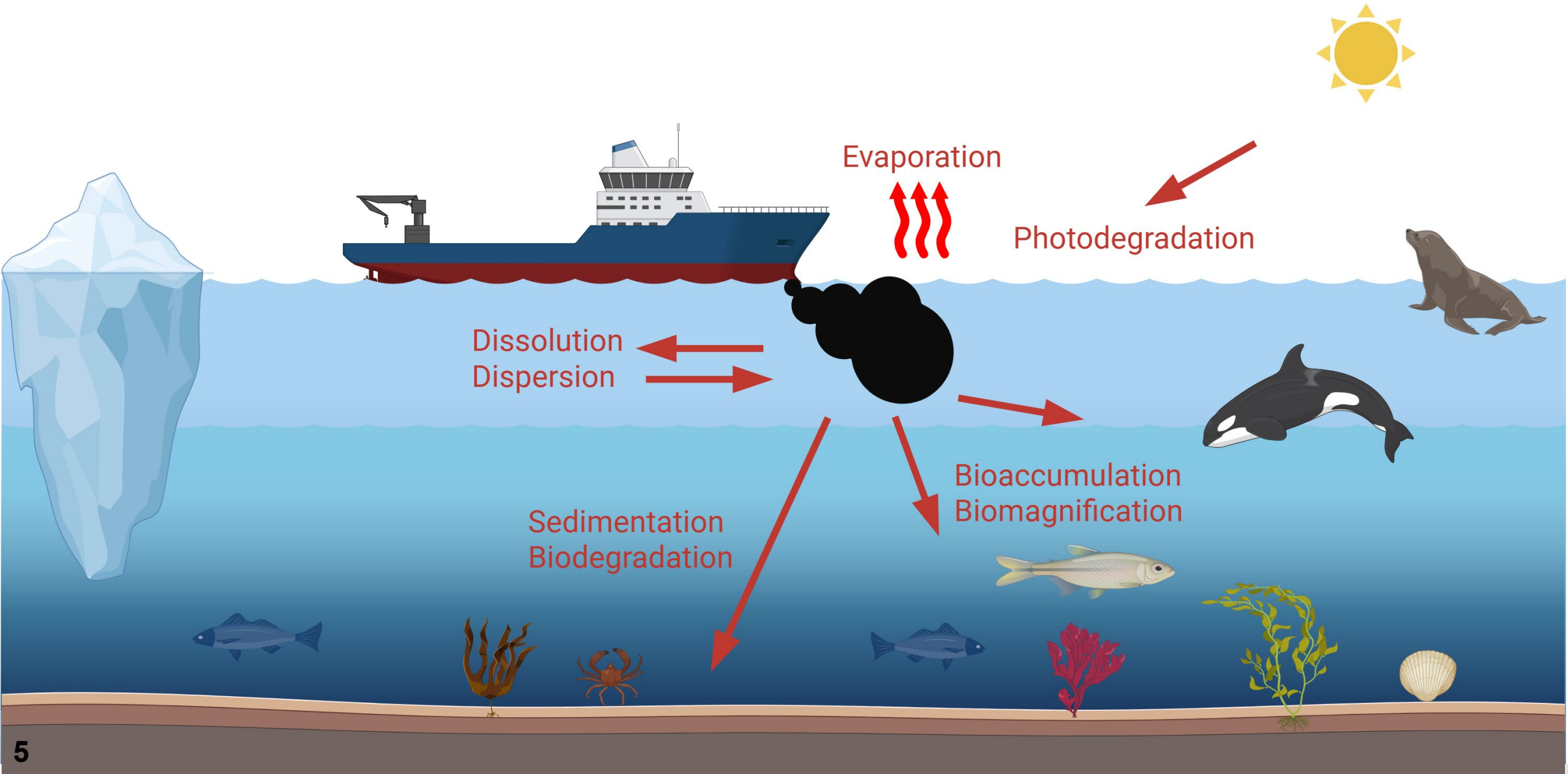
- 100s – 1000s of peaks/sample
- Many samples
- Not many standards
- High quality spectra
- Powerful retention patterns

Need:

- Coherent, aligned peak tables
- Consistent peak IDs

How to meet needs quickly, reliably, without losing our minds?

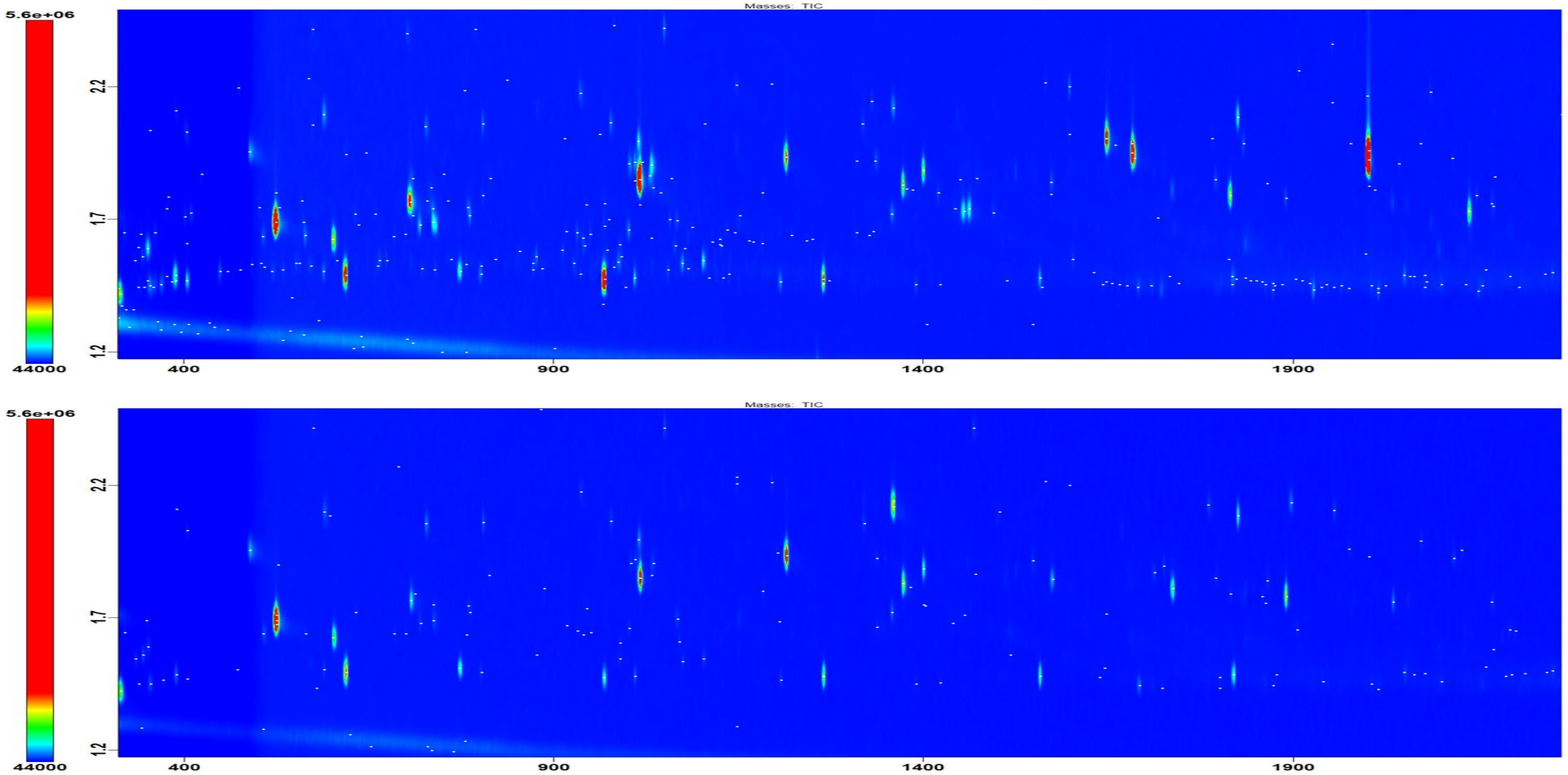
Transport and Environmental Fate of Potential Spilled Oils in the Arctic



Field Pea Project

- How does adding N to soil influence pea metabolism and volatiles?
  - Water: no extra N (control)
  - Nitrogen: extra N added to soil
  - Wild Type: N-fixing bacteria
  - Mutant: non-N- fixing bacteria
- VOCs from pea plants collected on TD tubes.

Example Chromatograms - Peas

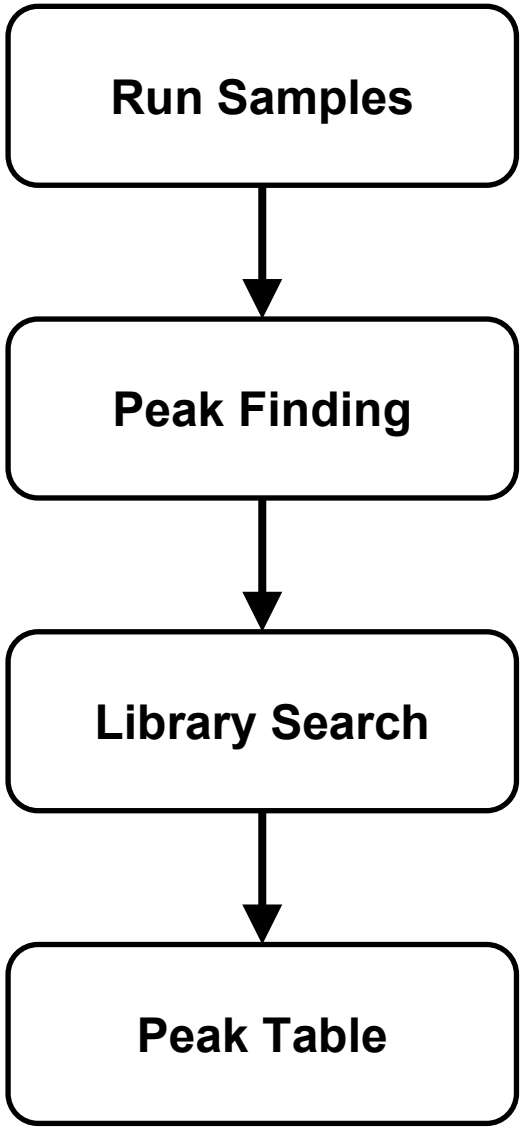


Basic Workflow

Peaks close to S/N cutoff?

Poor library matches?

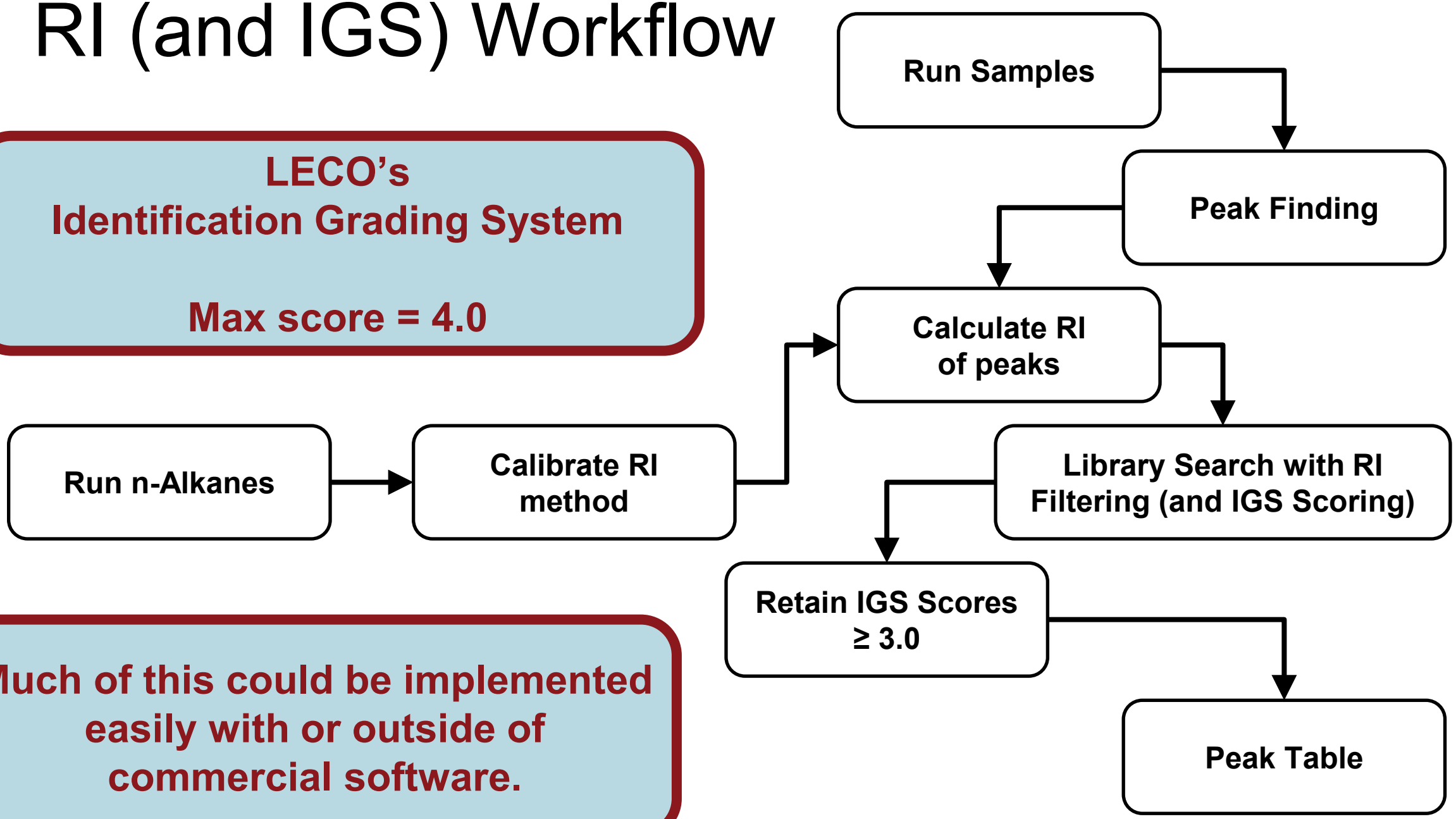
Comparing across samples?



RI (and IGS) Workflow

LECO's Identification Grading System

Max score = 4.0



Much of this could be implemented easily with or outside of commercial software.

Peak Table Comparisons - Peas

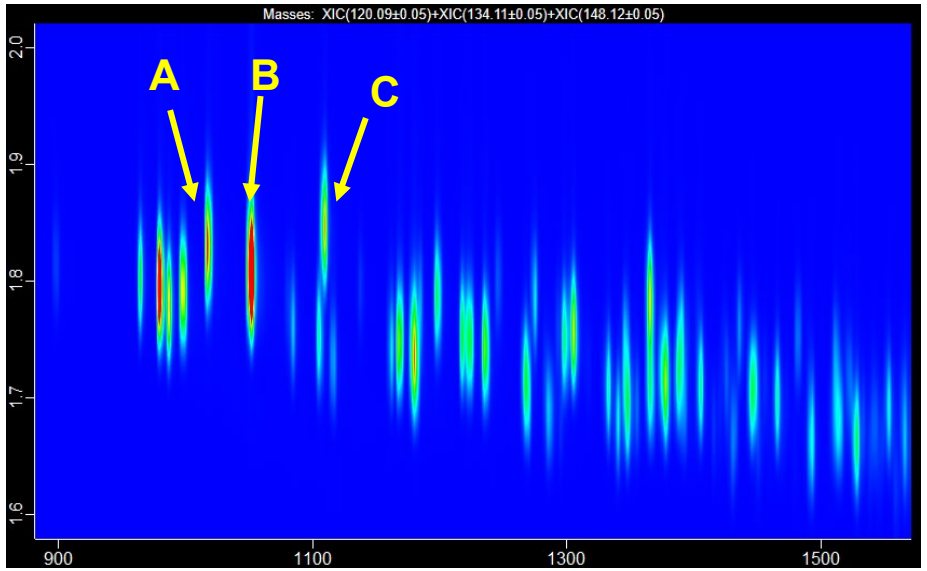
Name	Area	R.T. (s)	Similarity	Reverse	Mass Accuracy (ppm)
Heptane	1013486	405, 1.470	919	919	0.35
Peak 74	79224	405, 1.610			
Peak 75	697597	405, 2.030			
Peak 77	33730	410, 1.303			
Propane, 1,2-dichloro-	267776	410, 1.725	843	860	-1.74
1,4-Dioxane	97313	425, 1.871	816	816	-0.85
Cyclohexane, methyl-	436418	450, 1.505	909	909	-1.03
Pentane, 2,2,3,4-tetramethyl-	115901	460, 1.503	859	859	N/A
Methyl Isobutyl Ketone	218347	475, 2.195	906	906	-0.43
Cyclopentane, 1,2,4-trimethyl-	47517	477.5, 1.509	821	821	N/A
Pyridine	4901691	490, 1.955	953	956	-0.93
Pentane, 3-ethyl-	111738	492.5, 1.530	833	836	N/A
Peak 95	144311	502.5, 1.744			
Pentane, 2,3,3-trimethyl-	50593	505, 1.535	817	817	N/A
Peak 97	366567	507.5, 1.635			
Hexane, 2,3-dimethyl-	59605	510, 1.520	858	933	N/A
Heptane, 2-methyl-	105485	520, 1.505	913	913	N/A
Peak 101	401686	522.5, 1.620			
Toluene	46188951	525, 1.690	930	935	-0.08
Peak 104	25906	527.5, 1.236			
Peak 105	33147	527.5, 1.700			
1-Pentanol	56076	530, 1.744	858	858	N/A
Peak 108	107854	535, 1.510			
Peak 111	48988	545, 1.278			
Peak 113	119932	552.5, 1.535			
Cyclohexane, 1,3-dimethyl-, cis-	77277	557.5, 1.534	842	842	-2.28
(2-Fluorophenyl) methanol, 2-methylpropyl ether	120057	560, 1.770	847	847	N/A
Peak 116	48172	562.5, 1.264			
Butane, 1-iodo-	197913	562.5, 1.690	806	806	-1.84
Peak 118	631025	565, 1.640			
Peak 119	33180	570, 1.232			
Peak 120	76564	570, 2.229			
Peak 121	114551	572.5, 1.526			
Peak 122	34995	575, 1.236			
3-Pentanone, 2,4-dimethyl-	99104	575, 2.055	814	816	-1.1
Cyclopentanone	151209	575, 2.415	856	856	-1.26
Octane	387671	590, 1.505	891	891	-0.74

No RI/IGS: 789 peaks

Name	Area	R.T. (s)	Similarity	Reverse	Mass Accuracy (ppm)	RI	Lib. RI	IGS Score	Concerns
Heptane	1013486	405, 1.470	919	919	0.35	708.4	700	4	
Propane, 1,2-dichloro-	267776	410, 1.725	843	860	-1.74	710.8	701	3	M+:0
1,4-Dioxane	97313	425, 1.871	816	816	-0.85	718.1	673	4	
Cyclohexane, methyl-	436418	450, 1.505	909	909	-1.03	730.1	727	4	
Hexane, 2,4-dimethyl-	115901	460, 1.503	856	856	N/A	734.9	732	3	M+:0
Methyl Isobutyl Ketone	218347	475, 2.195	906	906	-0.43	742.2	735	4	
Cyclopentane, 1,2,4-trimethyl-	47517	477.5, 1.509	821	821	N/A	743.4	776	3	M+:0
Pyridine	4901691	490, 1.955	953	956	-0.93	749.4	746	4	
Pentane, 2,3,4-trimethyl-	111738	492.5, 1.530	829	848	N/A	750.6	753	3	M+:0
Pentane, 2,3,3-trimethyl-	50593	505, 1.535	817	817	N/A	756.6	760	3	M+:0
Hexane, 2,3-dimethyl-	59605	510, 1.520	858	933	N/A	759	760	3	M+:0
Heptane, 2-methyl-	105485	520, 1.505	913	913	N/A	763.9	765	3	M+:0
Toluene	46188951	525, 1.690	930	935	-0.08	766.3	763	4	Saturated
1-Pentanol	56076	530, 1.744	858	858	N/A	768.7	765	3	M+:0
Cyclohexane, 1,3-dimethyl-, cis-	77277	557.5, 1.534	842	842	-2.28	781.9	785	4	
Benzene, 1-fluoro-3-methyl-	120057	560, 1.770	828	828	-2.42	783.1	767	4	
Butane, 1-iodo-	197913	562.5, 1.690	806	806	-1.84	784.3	819	4	
3-Pentanone, 2,4-dimethyl-	99104	575, 2.055	814	816	-1.1	790.4	781	4	
Cyclopentanone	151209	575, 2.415	856	856	-1.26	790.4	791	4	
Octane	387671	590, 1.505	891	891	-0.74	797.6	800	4	

With RI/IGS: 146 peaks

Alkyl benzenes



Peak	Sample	Name	Match	RI	Lib RI	IGS
A	CD	Benzene, 1-ethyl-4-methyl-	938	980.5	954 ± 6(55)	4
	MFO	Benzene, 1-ethyl-4-methyl-	937	980.5	954 ± 6(55)	4
	VLSD	Benzene, 1-ethyl-2-methyl-	936	980.5	971 ± 7(62)	4
	VLSO	Benzene, 1-ethyl-4-methyl-	934	980.5	954 ± 6(55)	4
	D	Benzene, 1-ethyl-3-methyl-	899	980.5	957 ± 7(69)	4
B	CD	1,3,5-Trimethylbenzene	958	996.6	972 ± 9(109)	4
	MFO	1,3,5-Trimethylbenzene	961	995.4	972 ± 9(109)	4
	VLSD	Benzene, 1,2,4-trimethyl-	949	996.6	990 ± 6(90)	4
	VLSO	Benzene, 1,2,3-trimethyl-	953	996.6	1013 ± 7(75)	4
	D	1,3,5-Trimethylbenzene	958	996.6	972 ± 9(109)	4
C	CD	Benzene, 1,2,3-trimethyl-	946	1024.7	1013 ± 7(75)	4
	MFO	Benzene, 1,2,4-trimethyl-	956	1024.7	990 ± 6(90)	4
	VLSD	Benzene, 1,2,4-trimethyl-	956	1024.7	990 ± 6(90)	4
	VLSO	Benzene, 1,2,4-trimethyl-	953	1024.7	990 ± 6(90)	4
	D	Benzene, 1,2,3-trimethyl-	953	1024.7	1013 ± 7(75)	4

Can We Do Better?

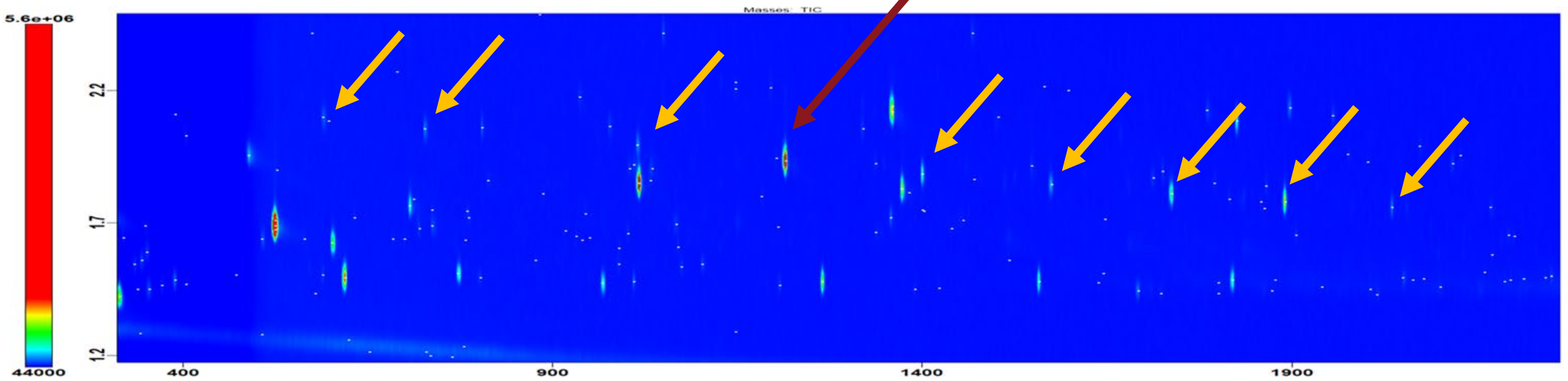
- IGS strategy works well
  - Keep “good” peaks
    - High quality spectra
    - Good library matches
    - ^{1}RI makes sense
  - Catches “known knowns” and “known unknowns”
  - Struggles a bit with isomers
  - Lets the “unknowns” go

- But…
  - What about the “unknown knowns”???
  - New unknowns we learn to expect to find in our samples

Include custom HRMS Library

- What to add to the library?
  - Standards
  - Known compounds

- What about “unknown knowns”
  - We can add these too!



Rules to build library

To add a peak/spectrum, it must…

- Be a “good” peak
  - S/N > 100
  - Chromatographically pure
  - Show up in multiple technical replicates
- Have a RI value
- Give unique ID – (compound name or ID e.g.: JJH – 001)

You don’t need to know exactly what it is

You don’t even need a partial idea of its identity

Uncertain peaks?

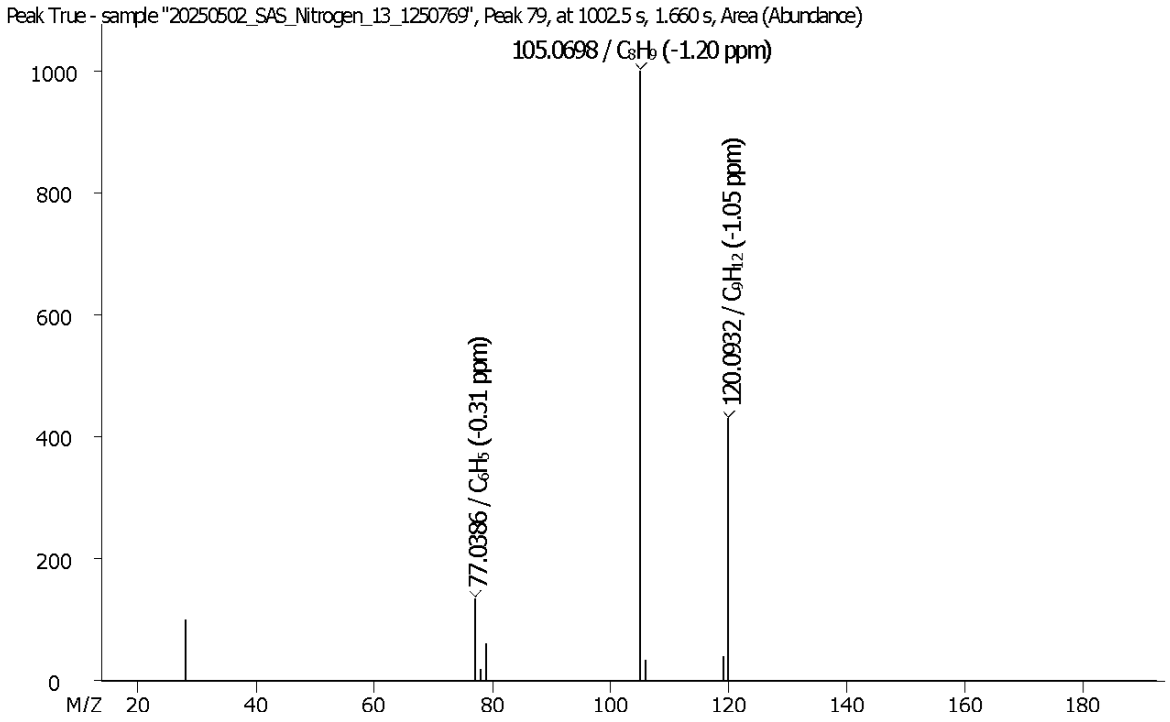
Peaks with

- Multiple isomers
- Only garbage matches

Pick and assign most reasonable hit

- Cannot add to library without hit
- Can always update later

This peak JJH-010 C3 AlkylBz



Chromatogram - 20250502_SAS_Nitrogen_13_1250769 | Hit Table - 20250502_SAS_Nitrogen_13_1250769

Customize Quick Access Toolbar

Hit		Similarity	Reverse	Lib. RI	Mass Differer	Probability	CAS
1	Benzene, 1,2,3-trimethyl-	788	788	1013 ± 7(75)		52.4	526-73-8
2*	Benzene, 1-ethyl-4-methyl-	765	765	954 ± 6(55)		19.1	622-96-8
3	1-Hexanone, 5-methyl-1-phenyl-	764	804			28.1	25552-17-4
4	Benzene, 1-ethyl-2-methyl-	748	748	971 ± 7(62)		16.2	611-14-3
5	Benzene, 1-ethyl-4-methyl-	747	754	954 ± 6(55)		19.1	622-96-8
6	Benzene, 1-ethyl-4-methyl-	743	750	954 ± 6(55)		13.0	622-96-8
7	Benzene, 1-ethyl-3-methyl-	736	736	957 ± 7(69)		10.0	620-14-4
8	1-Butanone, 3-methyl-1-phenyl-	733	762			5.1	582-62-7
9	Mesitylene	731	731	972 ± 9(109)		8.1	108-67-8
10	Benzene, 1,2,4-trimethyl-	726	733	990 ± 6(90)		3.9	95-63-6

Peak Table Comparisons

NIST Library

Sample	# peaks	Num RI	IGS 4	IGS 3	RI err all	RI err 4s	RI err 3s
Water 1	789	149	75	71	18.8	18.0	19.8
Water 9	706	132	71	62	19.2	19.2	19.1
WT 12	586	72	34	37	21.7	21.4	22.3
WT 15	405	76	38	38	19.6	19.3	19.9
N 11	540	75	38	36	20.9	22.4	19.2
N 14	357	57	27	29	20.6	18.7	22.4
Mutant 10	555	79	40	37	20.9	19.0	21.0
Mutant 11	524	84	40	43	18.9	18.1	18.7

$$Err = \sqrt{\frac{\sum(RI_L - RI_E)^2}{n_{RI}}}$$

Custom Library

Sample	# peaks	Num RI	IGS 4	IGS 3	JJH Lib	RI err all	RI err 4s	RI err 3s	RI err JJH
Water 1	789	162	81	78	41	17.9	17.8	18.2	7.3
Water 9	706	141	75	67	28	18.6	19.1	17.8	7.4
WT 12	586	80	39	40	28	19.9	20.5	19.4	8.7
WT 15	405	86	45	41	33	18.4	18.3	18.4	11.0
N 11	540	85	44	40	30	18.8	21.0	15.8	6.1
N 14	357	62	29	32	21	19.7	19.4	20.3	10.4
Mutant 10	555	90	45	43	30	19.4	18.3	18.7	11.3
Mutant 11	524	91	44	43	25	18.7	17.2	19.3	9.5

Deeper dive into peak tables

Name	Retention Index	Lib. RI	RI diffsq	IGS Score	Concerns
3-Pentanone	708.4	708	0.16	4	
2,2-Dimethoxybutane	757.8	759	1.44	3	M+:0
Toluene	766.3	766	0.09	4	Saturated
JJH-004	818.1	818	0.01	4	
3-Hexen-1-ol, (Z)-	854.2	854	0.04	4	
Ethylbenzene	860.2	860	0.04	4	
p-Xylene	868.7	869	0.09	4	
JJH-007	886.7	887	0.09	3	M+:0
a-Pinene	**925.3**	**936**	**114.49**	**4**	
a-Pinene	936.1	936	0.01	4	
Camphene	953	954	1	4	
ß-Myrcene	989.2	990	0.64	3	M+:0
2-Propanol, 1-(2-methoxypropoxy)-	1011.4	1011	0.16	3	M+:0
p-Cymene	1027.8	1028	0.04	4	
D-Limonene	1032.9	1033	0.01	4	
Undecane	1100	1100	0	4	
1,4-Dioxane-2,5-dione, 3,6-dimethyl-, (3S-cis)-	**1127.4**	**1166**	**1489.96**	**4**	
1,4-Dioxane-2,5-dione, 3,6-dimethyl-, (3S-cis)-	1165.8	1166	0.04	4	
Diethyl fumarate	1183.6	1184	0.16	3	M+:0
Naphthalene-D8	1190.4	1190	0.16	4	
JJH-014 (2-Propanol, 1-(2-butoxy-1-methylethoxy)- ??) Peak 1	1237.7	1238	0.09	3	M+:0
JJH-015 (2-Propanol, 1-(2-butoxy-1-methylethoxy)- ??) Peak 2	1242	1242	0	3	M+:0

Name	Retention Index	Lib. RI	RI diffsq	IGS Score	Concerns
JJH-006	1297.1	1297	0.01	3	M+:0
JJH-016	1353.1	1353	0.01	3	M+:0
Propanoic acid, 2-methyl-, 3-hydroxy-2,2,4-trimethylpentyl ester	1375	1375	0	3	M+:0
Pentadecane	1500	1500	0	3	M+:0
JJH-017	1690.7	1691	0.09	4	
Benzenesulfonamide, N-butyl-	1792.2	1792	0.04	4	
JJH-018	1842.9	1843	0.01	3	M+:0
Homosalate	1898	1898	0	4	
JJH-013 - Branched C21 HC	2068.2	2068	0.04	3	M+:0
JJH-013 - Branched C21 HC	**2090.9**	**2068**	**524.41**	**3**	**M+:0**
Docosane	2197.7	2200	5.29	3	M+:0
JJH-008.2	2242.5	2243	0.25	3	M+:0
Tricosane	2300	2300	0	4	
Tetracosane	2400	2400	0	4	
(Z)-Docos-9-enenitrile	2492.1	2495	8.41	4	
Pentacosane	2497.4	2500	6.76	4	
Octacosane	2797.7	2800	5.29	4	
Nonacosane	2898	2900	4	4	
Triacontane	2998.3	3000	2.89	3	M+:0

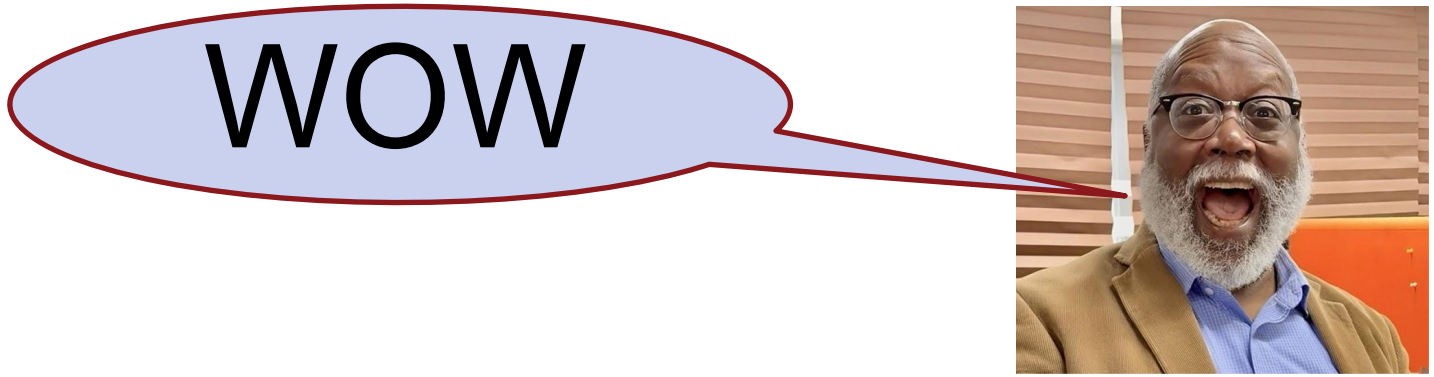
Finding problem peaks is trivial

RI Errors With Custom Library

What happens once we fix the 1-4 problem identifications per peak table?

Custom Library

Sample	# peaks	Num RI	IGS 4	IGS 3	JJH Lib	RI err all	RI err 4s	RI err 3s	RI err JJH
Water 1	789	162	81	78	41	17.9	17.8	18.2	1.9
Water 9	706	141	75	67	28	18.6	19.1	17.8	0.9
WT 12	586	80	39	40	28	19.9	20.5	19.4	0.5
WT 15	405	86	45	41	33	18.4	18.3	18.4	1.2
N 11	540	85	44	40	30	18.8	21.0	15.8	0.5
N 14	357	62	29	32	21	19.7	19.4	20.3	0.9
Mutant 10	555	90	45	43	30	19.4	18.3	18.7	0.6
Mutant 11	524	91	44	43	25	18.7	17.2	19.3	1.9



Individual compounds across samples

With custom library, compounds with IGS ≥ 3.0 are generally assigned the same identification across all samples where they are found

If we just use NIST and IGS, results are not so good

	Sample							
Peak	**Water 1**	**Water 9**	**WT 12**	**WT 15**	**N 11**	**N 14**	**Mut 10**	**Mut 11**
JJH-004	Pk 144	Pk 153	Pk 117	Pk 80	Pk 101	NF	Pk 109	Pk 85
JJH-010	NF	NF	Pk 164	Pk 105	Pk 159	NF	Pk 157	☺ **??**
Acetophenone	☺	☺	☺	Pk 125	Pk 186	☺	Pk 184	☺

NF = not found. ☺ = correct ID

JJH-010 is a C3 alkyl benzene. In Mut 11, peak was identified as a C3 alkyl benzene, but not sure if the correct isomer was assigned.

Concluding Thoughts

Do:

- Use a custom MS library
- Add your “unknown knowns”
- Collect local RI data
- Apply IGS filters

Get better peak tables in less time

Also learn to write Excel macros and/or write simple code

Acknowledgements

LECO
EMPOWERING RESULTS

TMIC
The Metabolomics Innovation Centre

INNOVATION.CA
CANADA FOUNDATION FOR INNOVATION | FONDATION CANADIENNE POUR L'INNOVATION

GenomeAlberta

GenomeCanada

NSERC
CRSNG

UNIVERSITY OF ALBERTA