Data visualization approaches for rapid, heuristic interpretation of highly complex mixtures of structurally and compositionally diverse chemical entities associated with pharmaceutically relevant materials

Douglas Kiehl

Research Advisor

Eli Lilly and Company, Indianapolis, IN, USA

Visualization vs Imaging

Visualization is a technique and a process of translating information and large datasets into graphic or pictorial representations in order to promote and/or clarify communication of unique features or concepts.

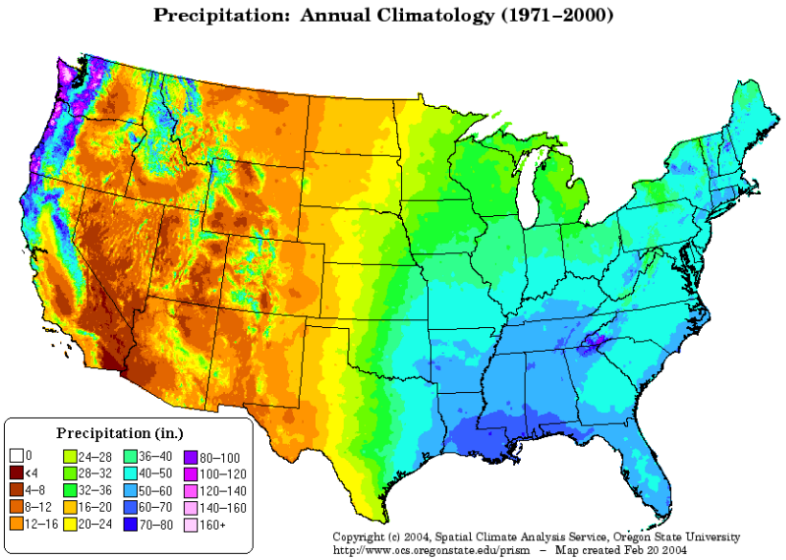
Imaging involves the capture, storage and processing of signals to produce or enhance an image or a visual representation of an object, often in 2D/3D graphical representation (e.g., pixel/voxel rendering).

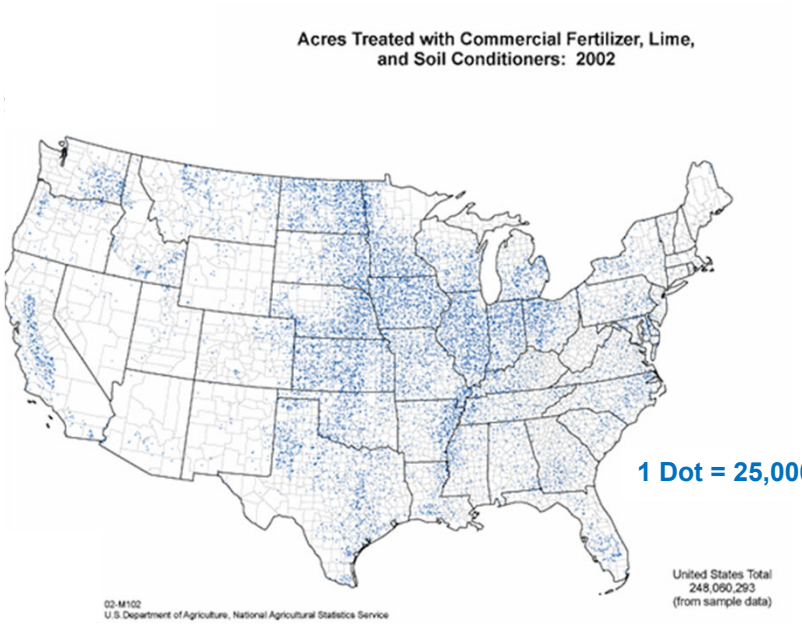
A map is an example of a visualization involving diagrammatic representation or collection of data illustrating relationships, characteristics, distribution, size, number or spatial arrangement of unique features over a given area and according to a chosen scale.

- Maps can be two-dimensional representations of three-dimensional space or can be three-dimensional.

Some examples. . .

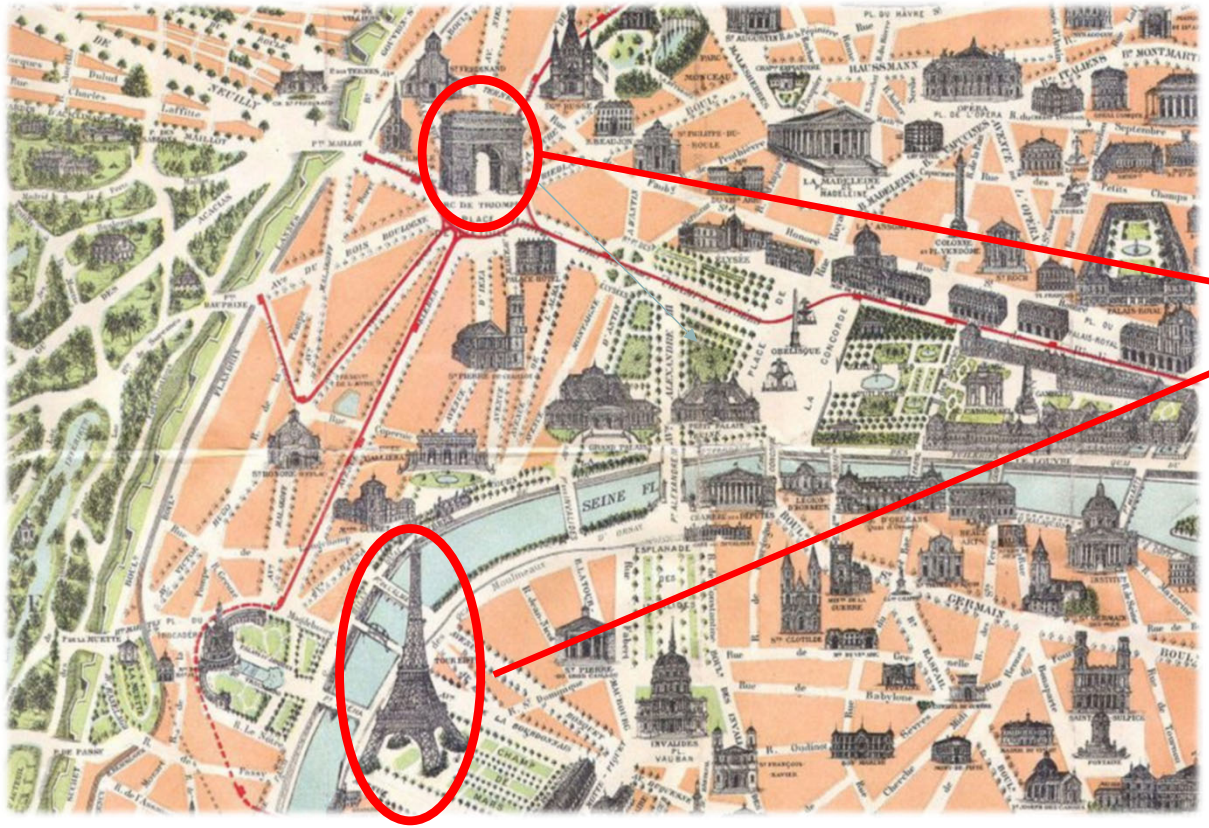
Semi-quantitative …and quantitative maps





Some examples. . .

A qualitative map: Paris, 1900



Intended to bring specific features to the viewer's attention

Complex Molecular Mixtures. . .

Contain many components across a range of concentrations

- **Petroleum products (fuels, polymers, etc.)**
- **Natural organic matter (dissolved marine organic matter)**
- **Biological materials (cell lysates, fermentation media, metabolites, etc.)**
- **Formulated drugs and pharmaceutical materials**




Complex Molecular Mixtures. . .

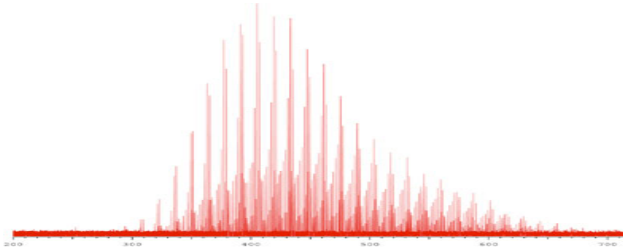
…of pharmaceutical relevance typically include:

- Extractables and leachables
- Related substances in starting materials, intermediates, API and products
- Process reaction impurities, reagents and their reaction products
- Excipients and drug-excipient interaction products
- Degradation products
- Surfactants

Samples frequently contain multiple compounds from multiple categories!!!

Complex Molecular Mixtures. . .

- Present unique challenges to the analyst
  - Multiple, diverse components at various concentrations
  - Abundant additives and formulation components can complicate the search for analytes present at low levels
  - Higher MW analytes = many elemental composition possibilities
  - Timing for results delivery lengthens with sample complexity
- Require advanced technologies and innovative tools
  - Deep expertise
  - Advanced instrumentation
  - Data reduction and visualization



Additives and Processing Aids in Plastic and Rubber Materials

- Anti-oxidants
- Anti-static agents
- UV light protectants
- Plasticizers
- Polymerization initiators
- Lubricants and Slip Agents
- Vulcanization accelerators
- Mold release agents
- Biocides and fungicides
- Catalysts
- Heat stabilizers
- Colorants

Structural Diversity

Compositional Diversity

Extractables can also include polymer fragments, monomers, and oligomers

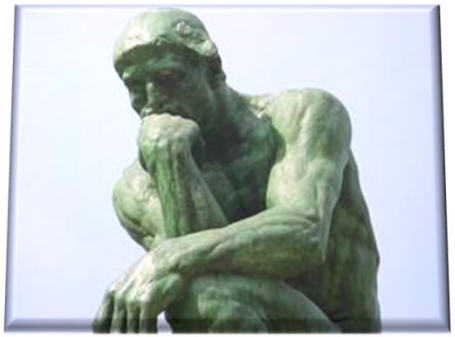
Molecular features contributing to compound diversity

Structural diversity

- *Degree of unsaturation (rings/double bond equivalents)*
- *Substituents*
- *Functional groups*

Isomeric diversity

- *Structural isomers*
- *Positional isomers*
- *Geometric isomers*
- *Stereoisomers*



Compositional diversity

- *Heteroatom content and type*
- *Isotopic contributions*

Polymer Supply Chain for Pharmaceutically Relevant Materials

Converters

Natural Raw Material	Monomer	Polymer	Masterbatcher	End Products & Suppliers
N-4	N-3	N-2	N-1	N
	• Bulk process chemicals • Storage stabilizers	• Initiators/Catalysts • Storage stabilizers • Antioxidants • Processing aids	• Antioxidants • Processing aids • Light stabilizers • Antistatic agents	• Colorants • Processing aids • Lubricants

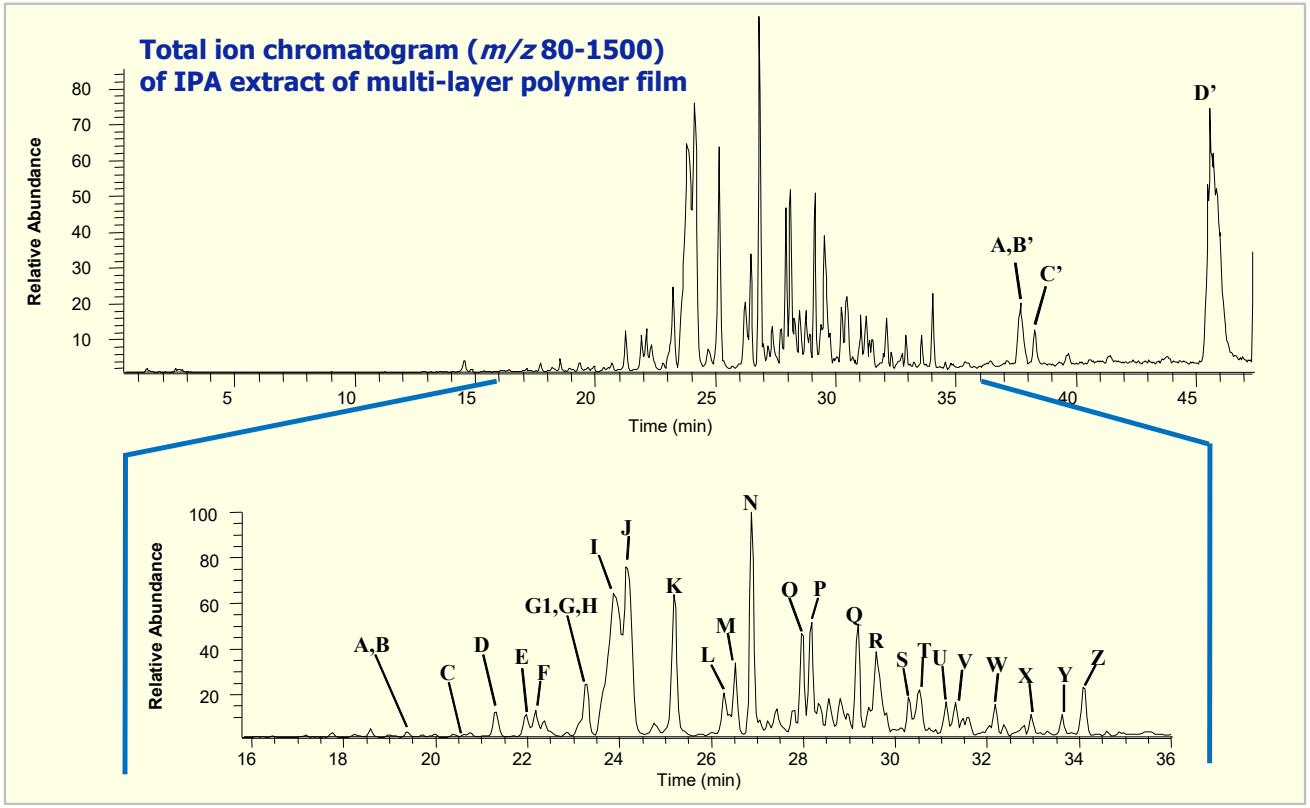
Components for
- Packaging
- Devices
- Container/closure
- Processing

May be cumulative!

Complex Molecular Mixtures. . .

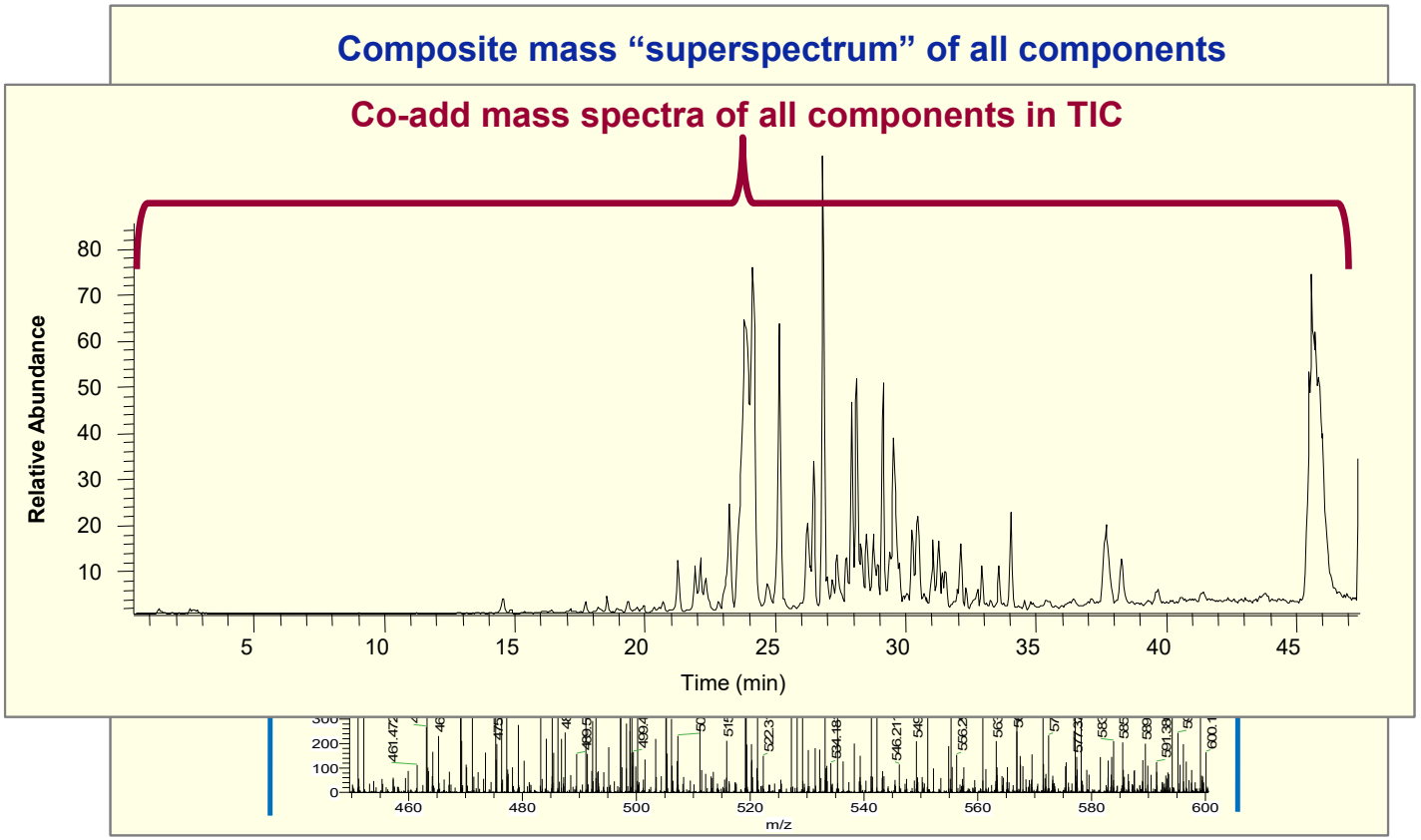
How Complex is "Complex?"

Chromatographic complexity



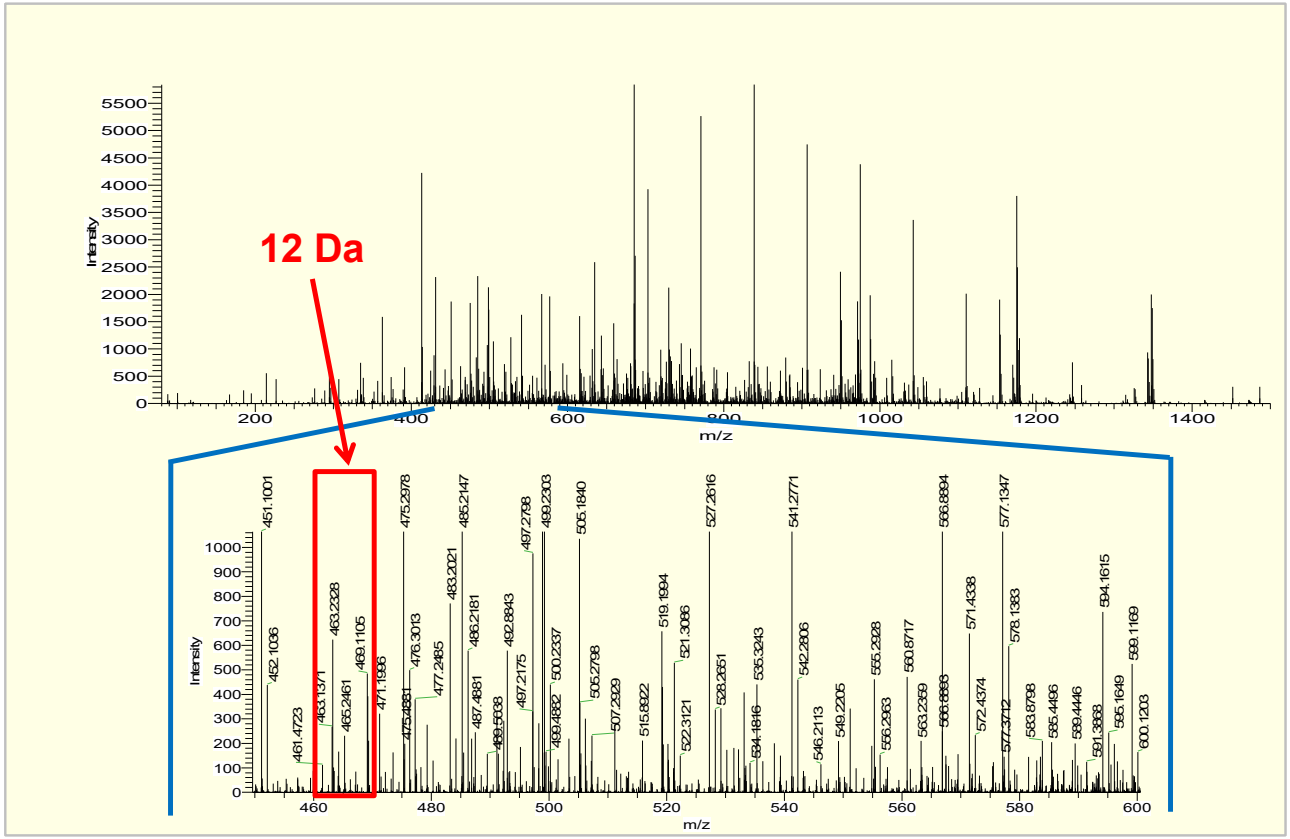
How Complex is "Complex?"

Mass spectral complexity



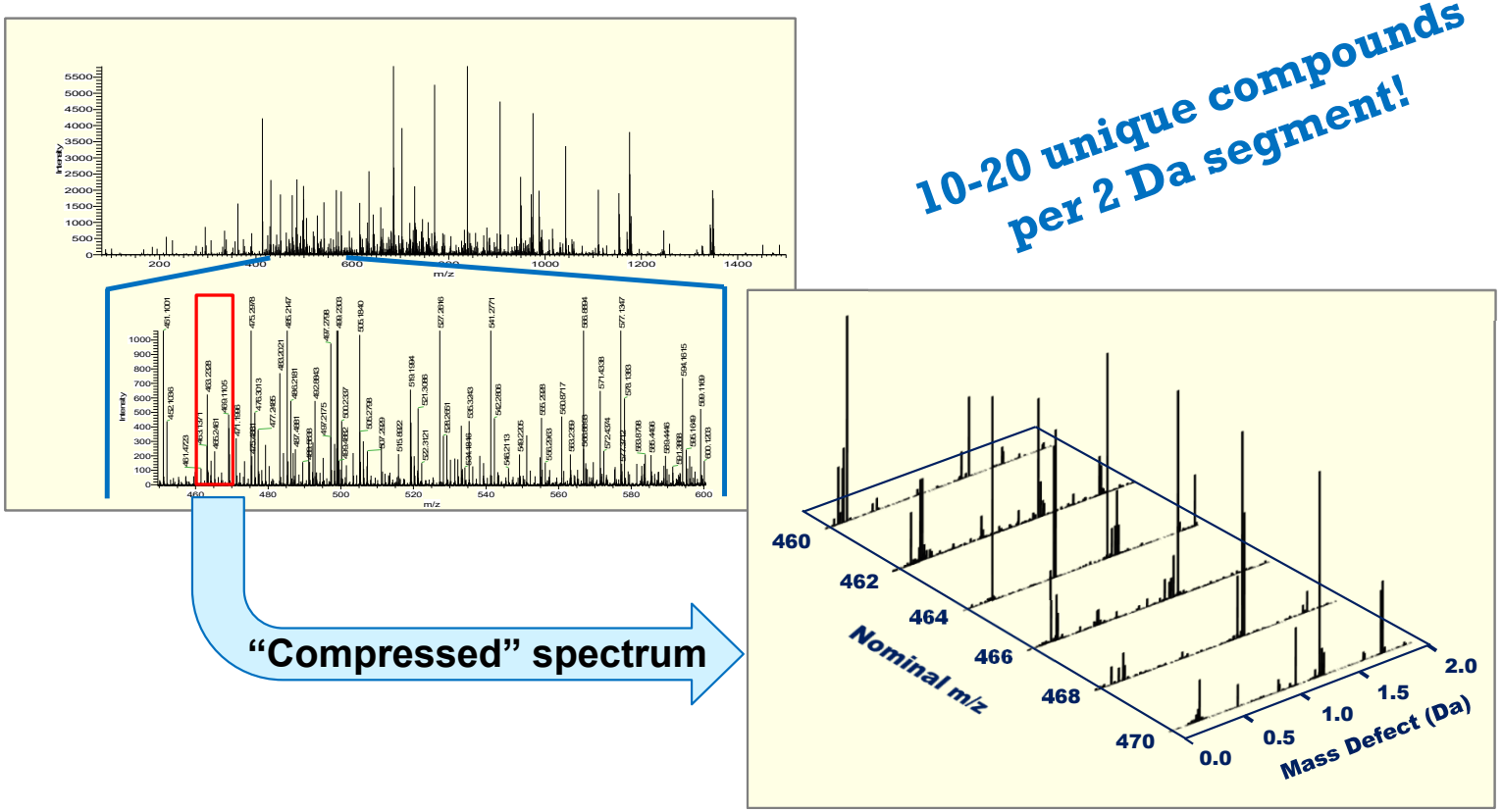
How Complex is "Complex?"

Mass spectral complexity

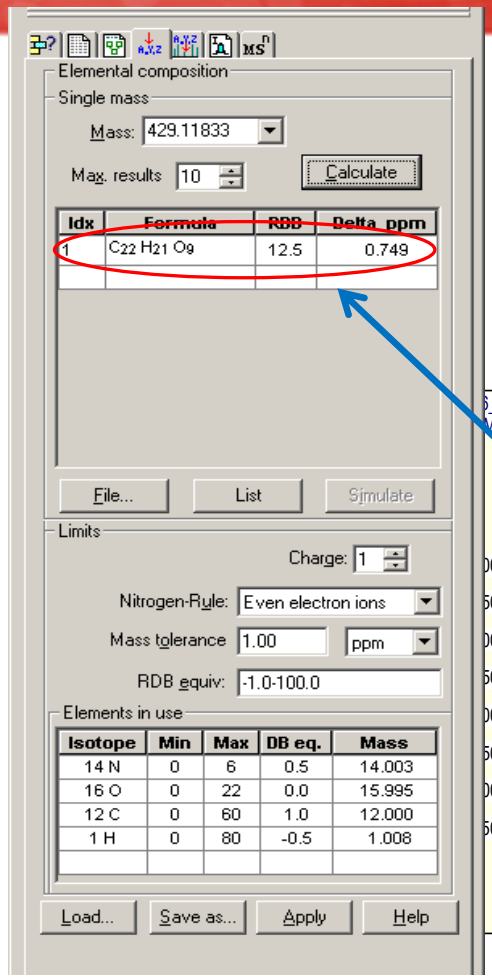


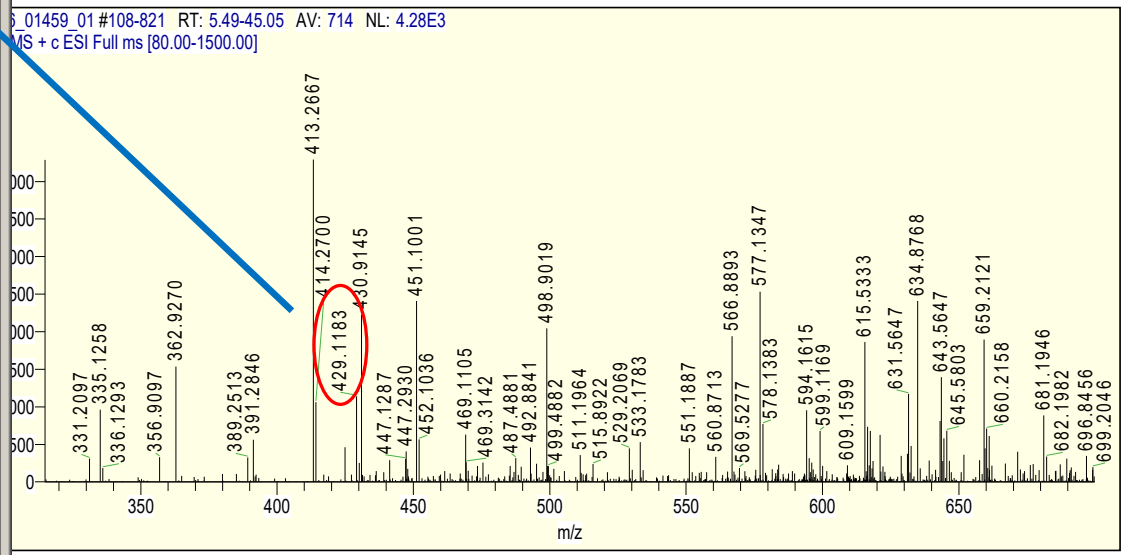
How Complex is "Complex?"

Mass spectral complexity



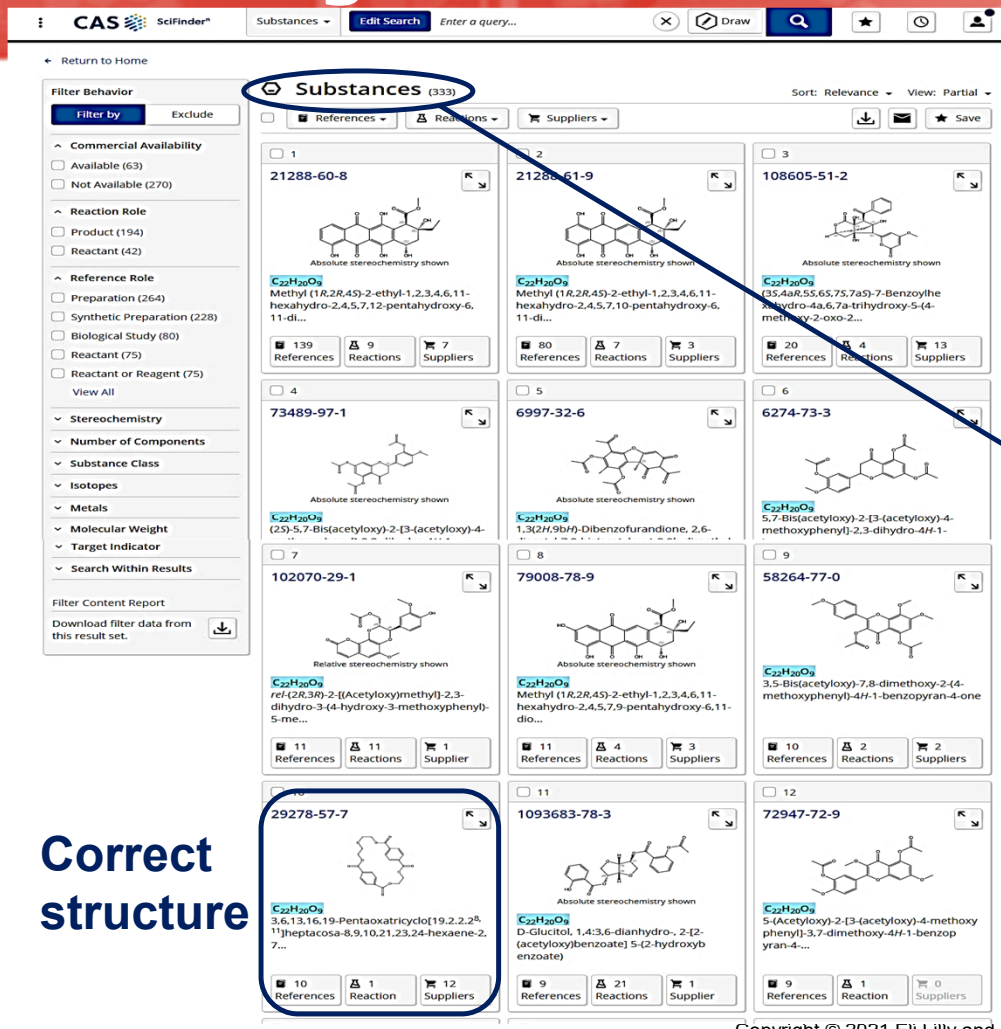
To Extol the Virtues of Accurate Mass Measurement…





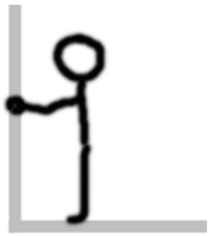
<1 ppm mass accuracy narrows possible choices for elemental composition…

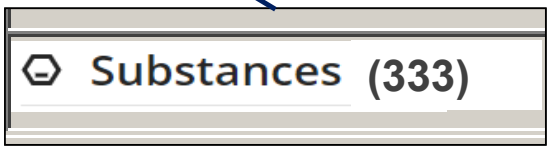
Correlating Structure with Formula



Correct structure

...but many structures are possible...





(All $C_{22}H_{20}O_9$!!)

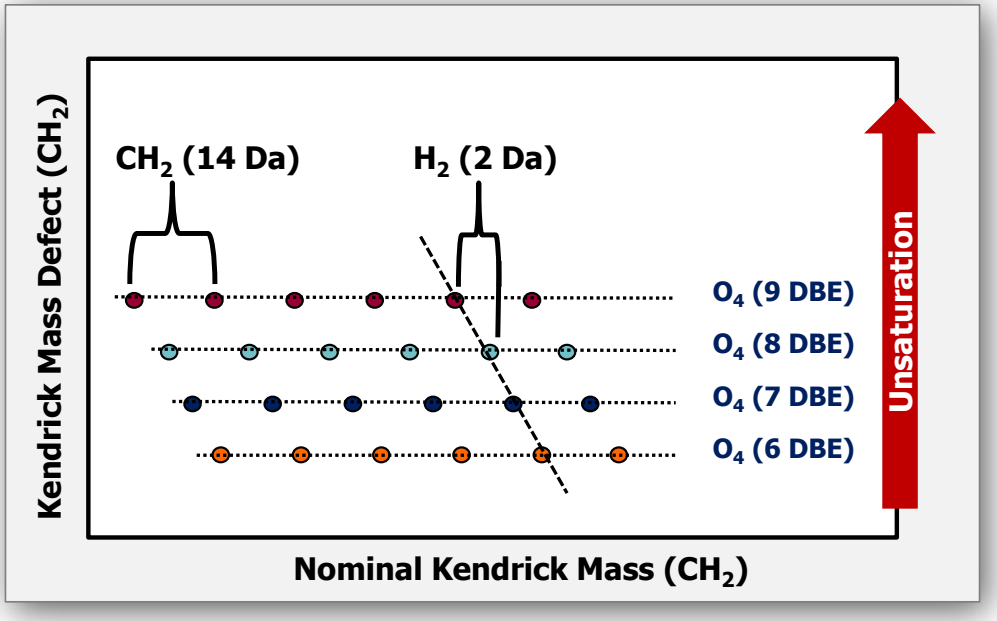
Some Tools for Visualizing Complex Molecular Mixtures: Kendrick Mass Defect Diagram

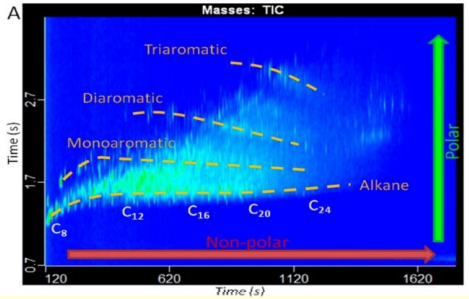
Kendrick mass defect diagram (2D or 3D modified)

– Rescales IUPAC mass scale (based on 12C mass = 12.000 00 Da) to "Kendrick" mass scale for CH2 (e.g., converts CH2 mass from 14.015 65 to 14.000 00 Da)

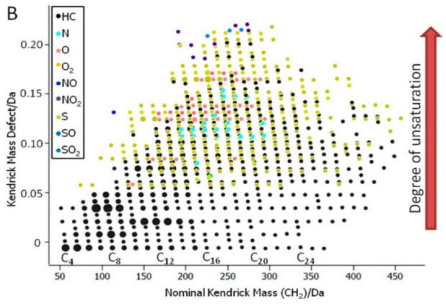
– Sorts compounds into homologous series by compound "class" (numbers of O, N, other heteroatom)

– Compound type (rings and double bonds) and degree of alkylation (CH_2 groups)





GC x GC ToFMS, heavy crude oil



Jonathan D. Byer; Kevin Siek; Karl Jobst; *Anal. Chem.* **2016,** 88, 6101-6104.

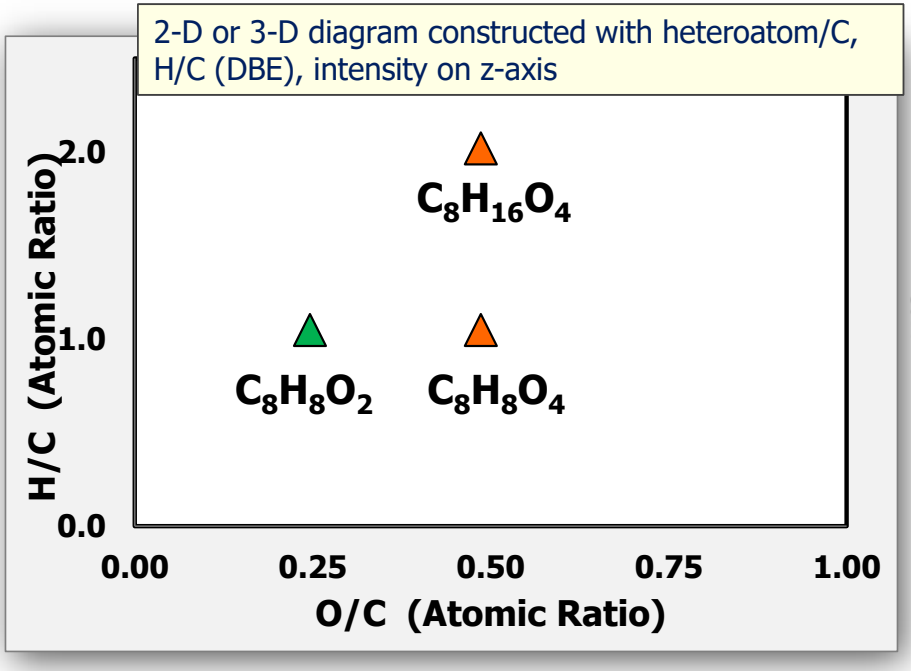
Some Tools for Visualizing Complex Molecular Mixtures: van Krevelen Diagram

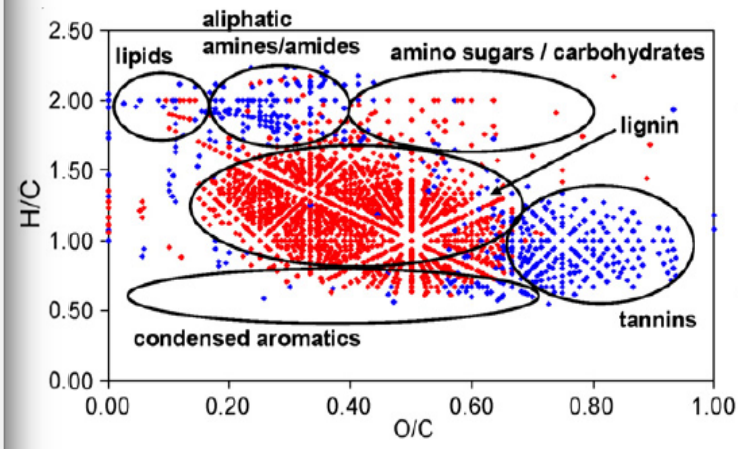
van Krevelen diagram (2D or 3D modified)

– Projects elemental composition according to H/C and heteroatom/C atomic ratios

– Compounds are classified according to degree of saturation, alkylation, heteroatoms, dehydration, deamination

– Compositional differences are amplified permitting classification, prediction of composition and origin



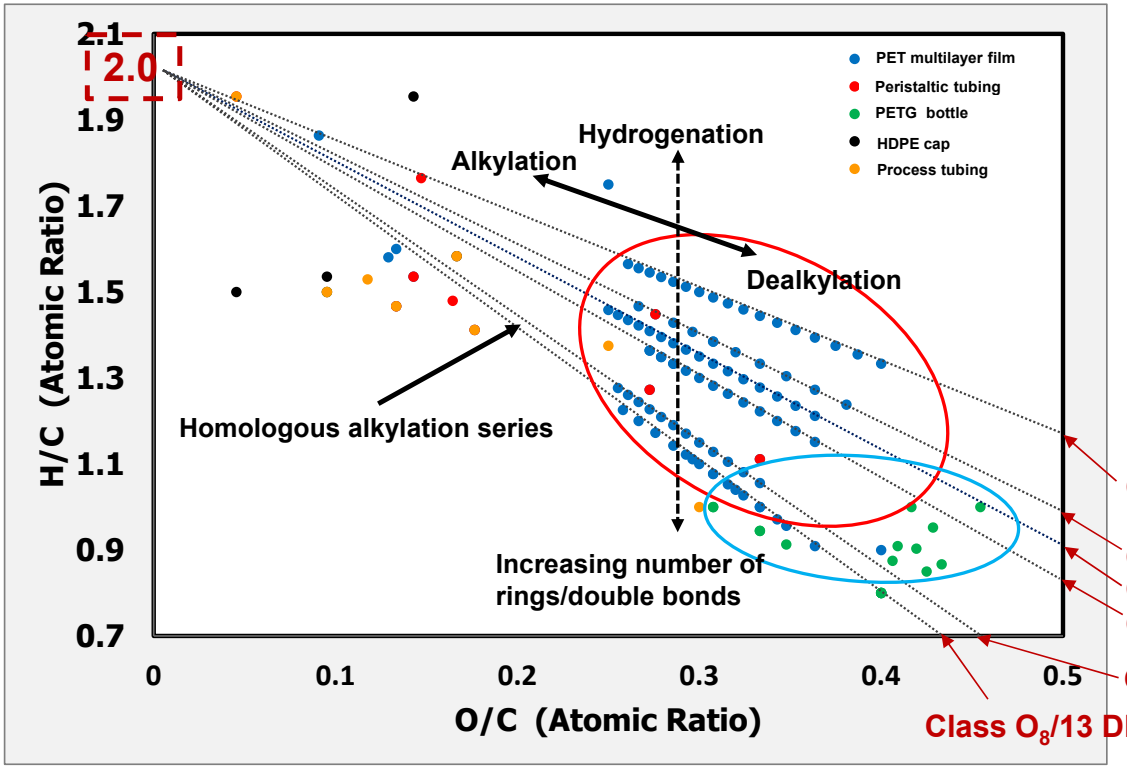


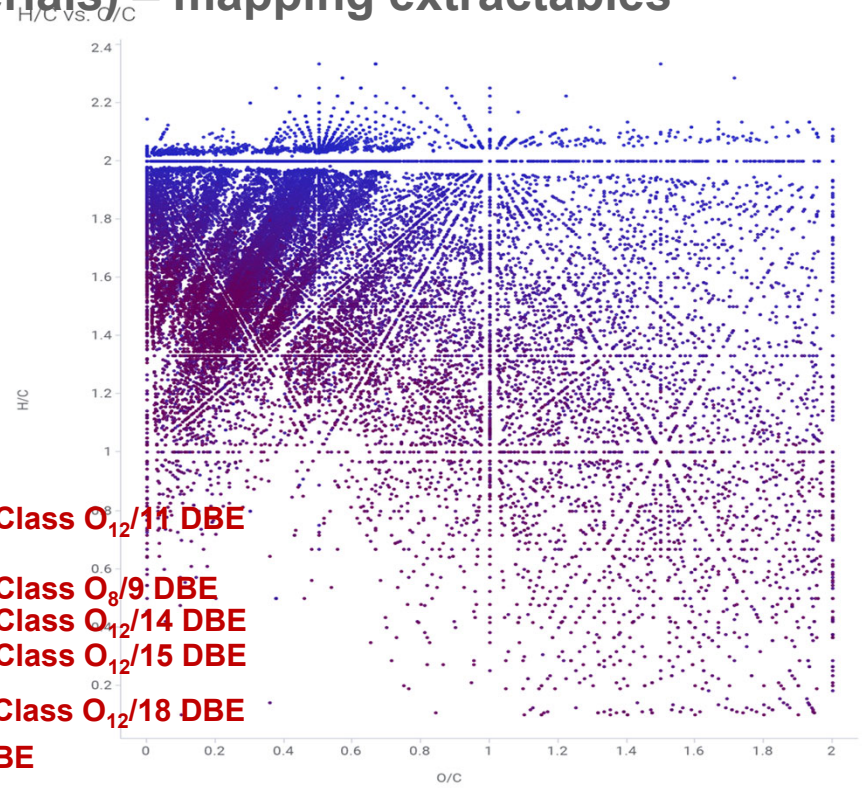
Dissolved organic matter

R.L. Sleighter, P.G. Hatcher / Marine Chemistry 110 (2008) 140–152

van Krevelen Diagram: Visualizing Relationships

FT-ICR LC-MS analysis of IPA extracts (5 materials) – mapping extractables



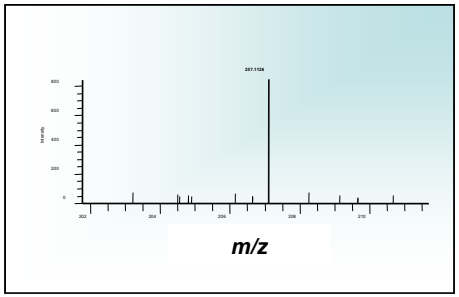


For any compound series $C_{c\text{-}n}H_{h\text{-}2n}O_o$: $y = (h\text{-}2n)/c\text{-}2n$ and $x = o/(c\text{-}n)$ ⇨ $y=2+(h\text{-}2c)/ox$

Spectral Map Libraries

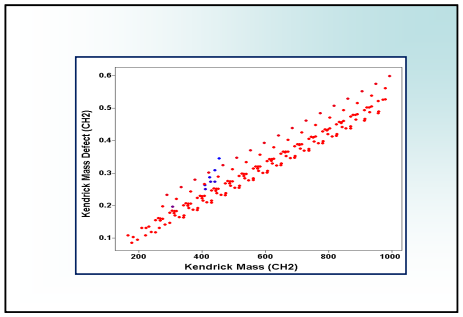
- **A traditional discrete spectral library is a collection of *individual* spectra representing individual chemical entities**

Each library entry represents a single, ***unique compound***



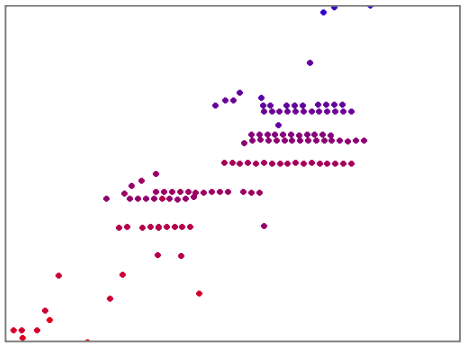
- **A spectral map library is a collection of *sets* of spectra representing visualizations of complex mixtures of structurally and compositionally diverse chemical entities**

Each library entry represents a ***unique extractables profile*** corresponding to a component or material

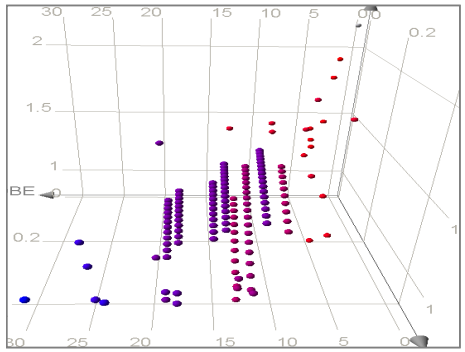


Spectral Map Libraries

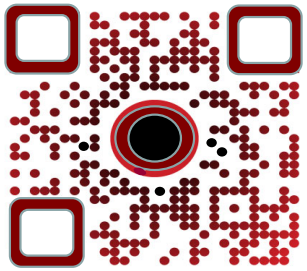
Standardized visualizations saved as QR Codes represent unique, highly complex mixture and material "fingerprints" that can be compared both qualitatively and quantitatively



2-D

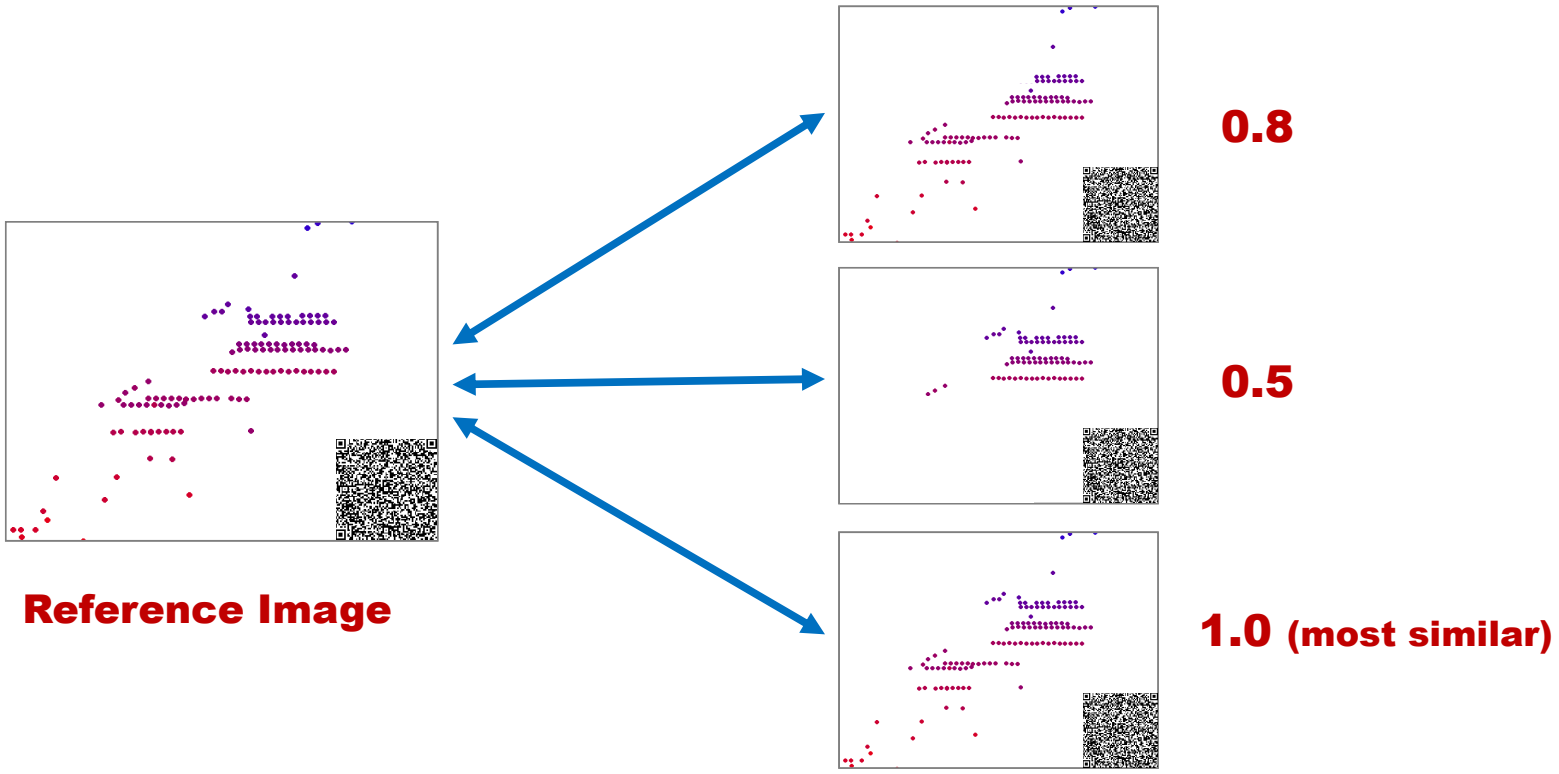


3-D



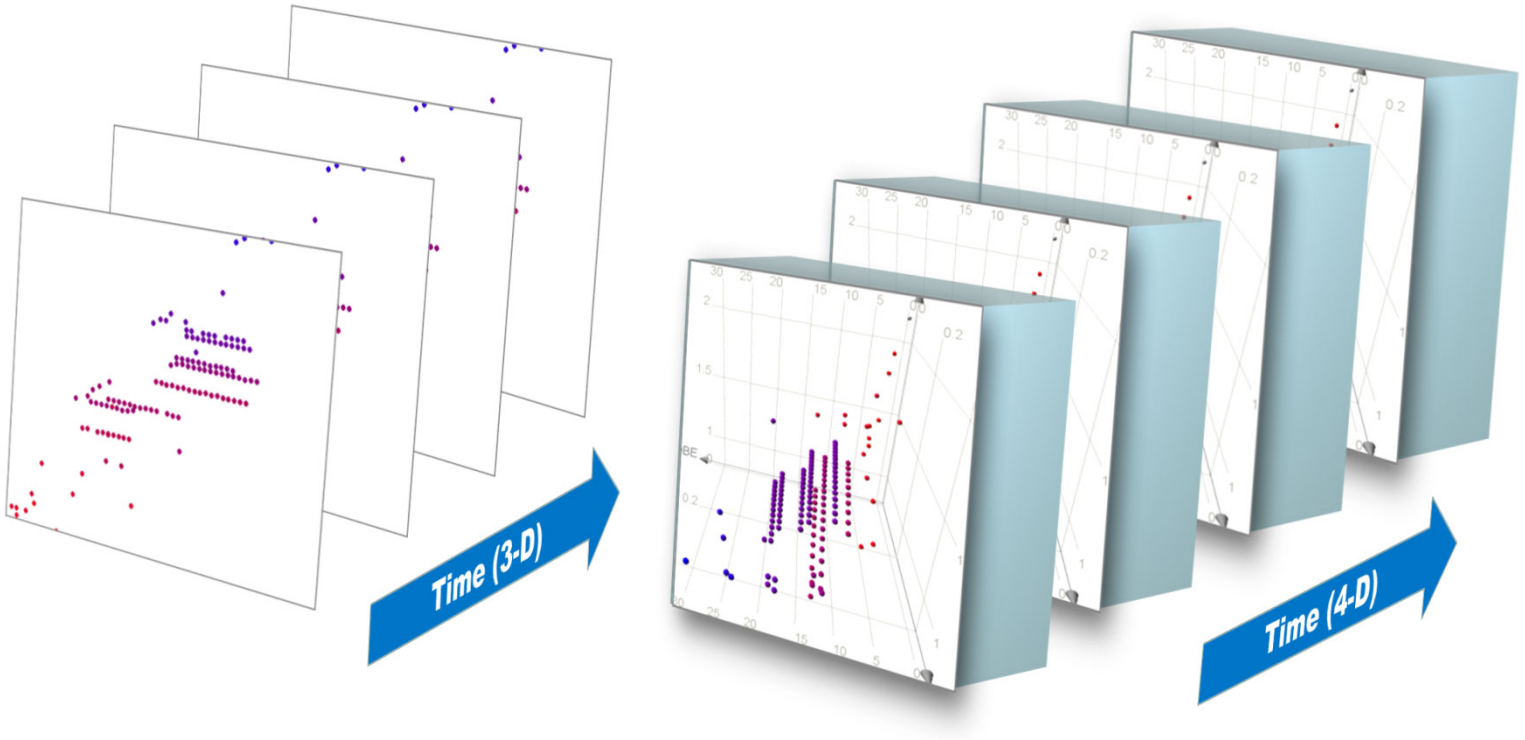
Spectral Map Libraries

Pairwise or batch comparison with a reference material and index of similarity / dissimilarity with anomaly detection

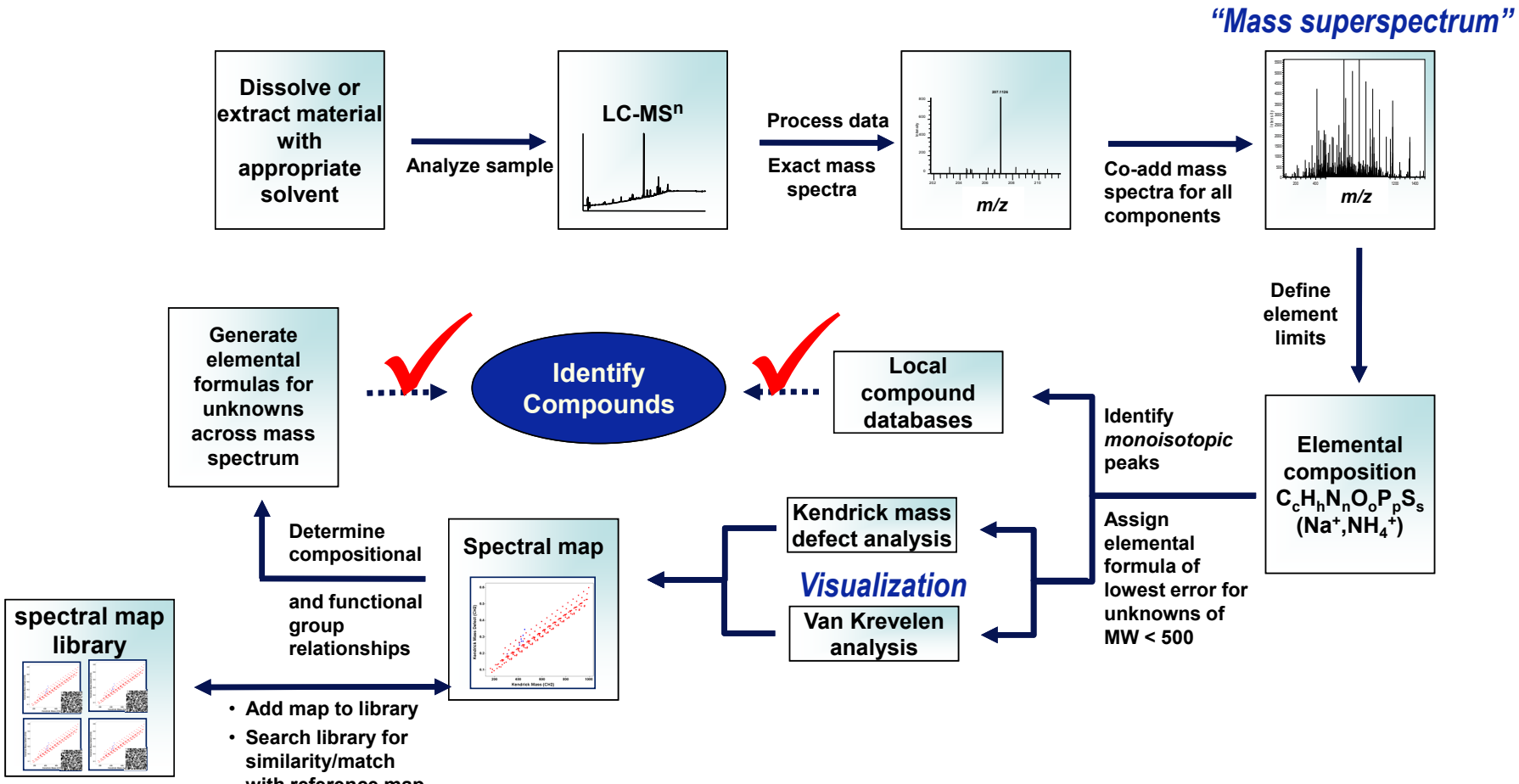


Spectral Map Libraries

High resolution digitized image grouping to evaluate change over time (pixelated 2-D or voxelated 3-D image)

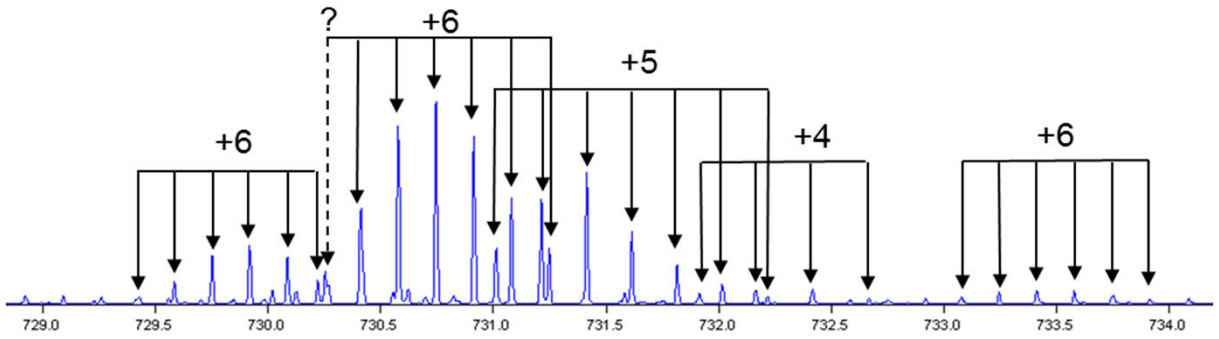


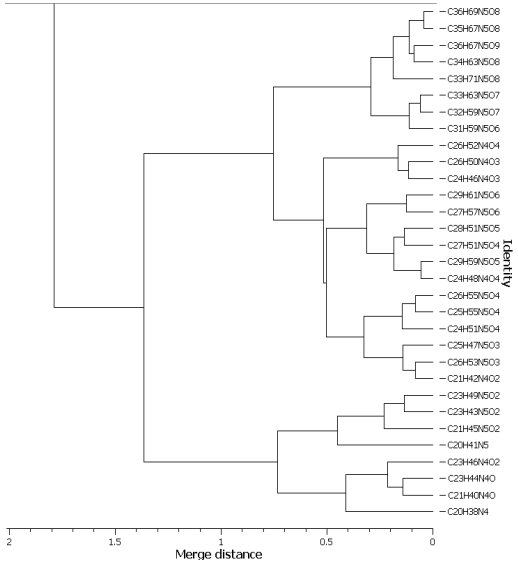
LC-MS Analytical Workflow for Extractables Assessment



Difference Correlation and Self-Calibrating Mass Maps

- Detect monoisotopic peaks
- Determine charge state
- Generate an ***autocorrelation spectrum*** using Fourier transform methods
- Autocorrelation spectrum has peaks corresponding to ***m/z differences*** source spectrum
- Apply Kendrick mass defect methods to classify peaks based on formula relationships
- Group related compositions using hierarchical clustering or automated van Krevelen plot analysis



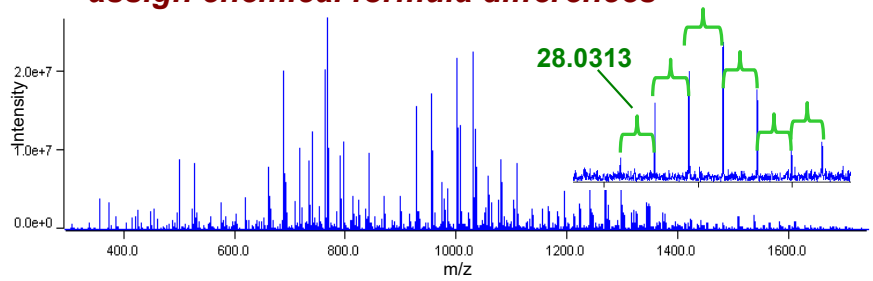


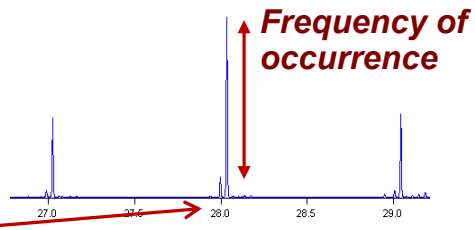
Difference Correlation and Self-Calibrating Mass Maps

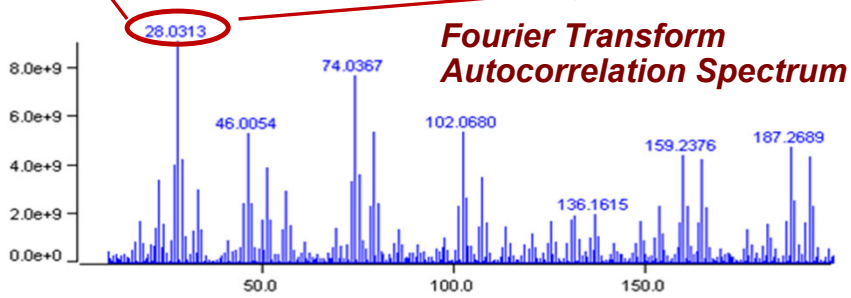
accurate m/z *difference* correlation

	Formula	.oss	Calc Mass	Expt Mass	Mono Inty	Abs Error	mDa Error	ppm Error
1	H2O		18.0106	18.0106	675963072	0.000055	0.06	3.06
2	CO		27.9949	27.9949	988126208	0.000013	-0.01	-0.47
3	C2H4		28.0313	28.0313	8958516224	0.000010	-0.01	-0.37
4	CH2O2		46.0055	46.0054	5211857920	0.000071	-0.07	-1.55
5	C3H4O		56.0262	56.0261	1176259584	0.000088	-0.09	-1.57
6	C3H6O2		74.0368	74.0367	7576710656	0.000065	-0.07	-0.88
7	C5H8O		84.0575	84.0574	799050624	0.000101	-0.10	-1.20
8	C4H6O3		102.0317	102.0316	552958464	0.000080	-0.08	-0.79
9	C5H10O2		102.0681	102.0680	5281296384	0.000079	-0.08	-0.78
10	C4H8O4		120.0423	120.0421	1193746816	0.000126	-0.13	-1.05
11	C6H10O3		130.0630	130.0629	842935104	0.000121	-0.12	-0.93
12	C7H14O2		130.0994	130.0993	1729864064	0.000091	-0.09	-0.70
13	C8H14O3		158.0943	158.0941	606668480	0.000169	-0.17	-1.07

Identify monoisotopic peaks, charge state and assign chemical formula differences



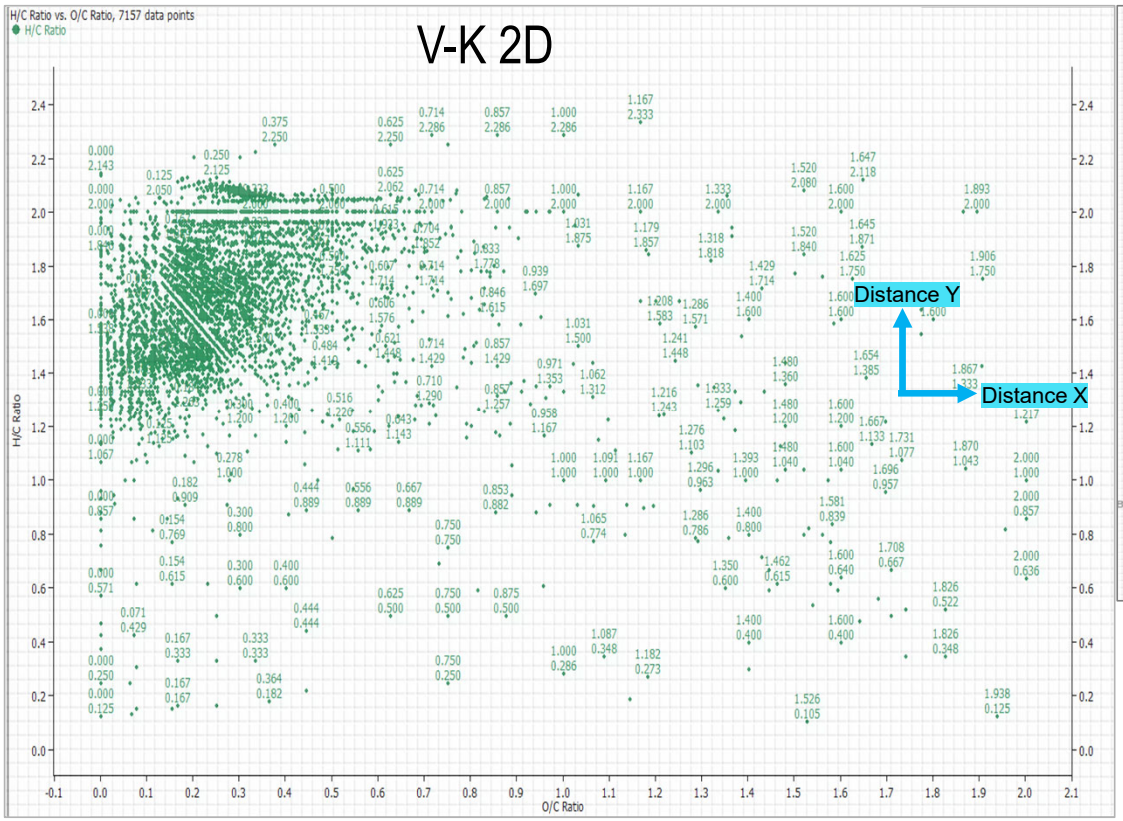


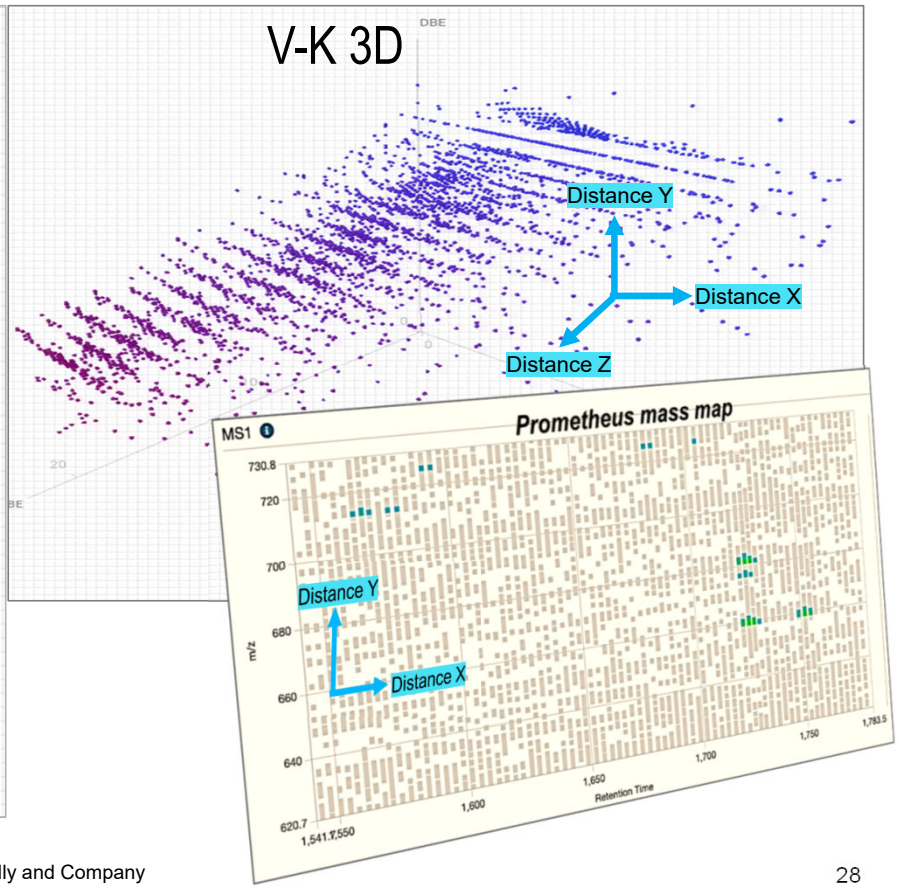


Basis for a *self-calibrating* mass map

Difference Correlation and Self-Calibrating Mass Maps

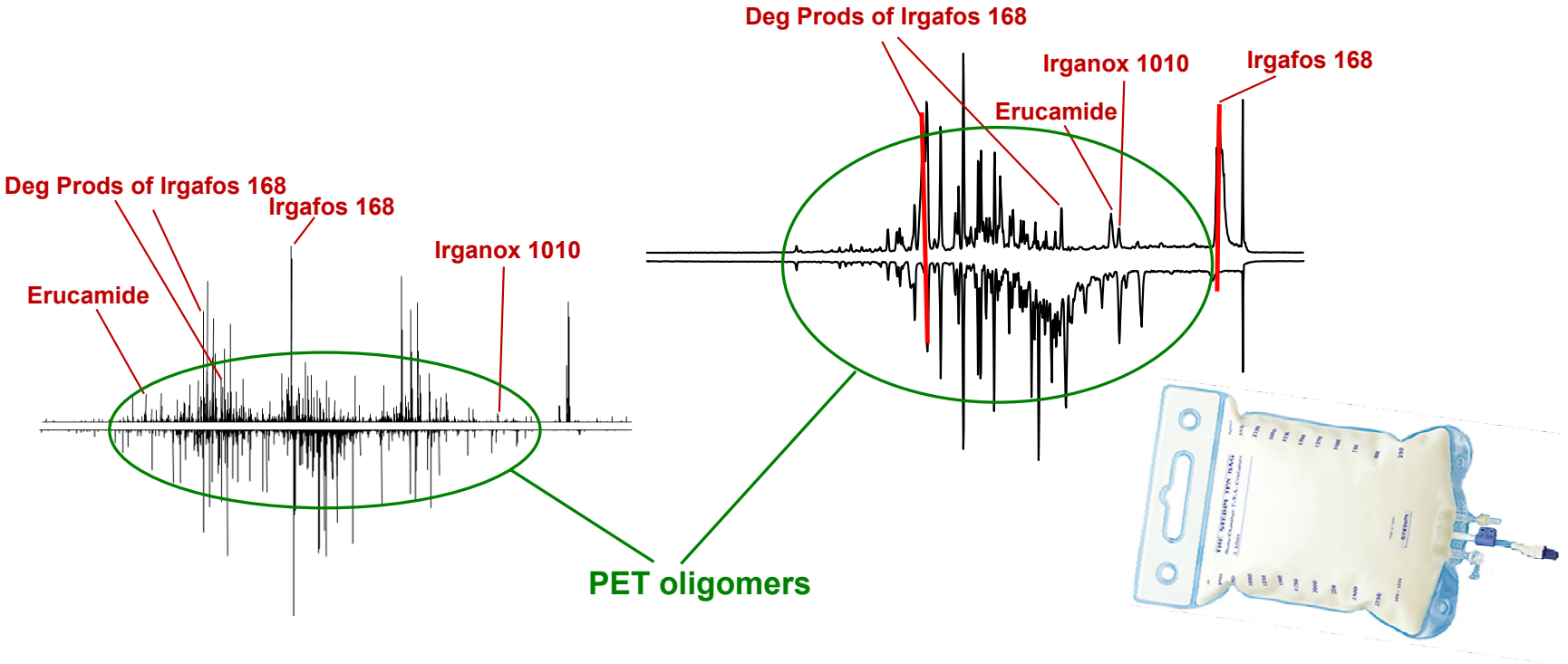
N-dimensional difference correlation and self-calibrating mass maps



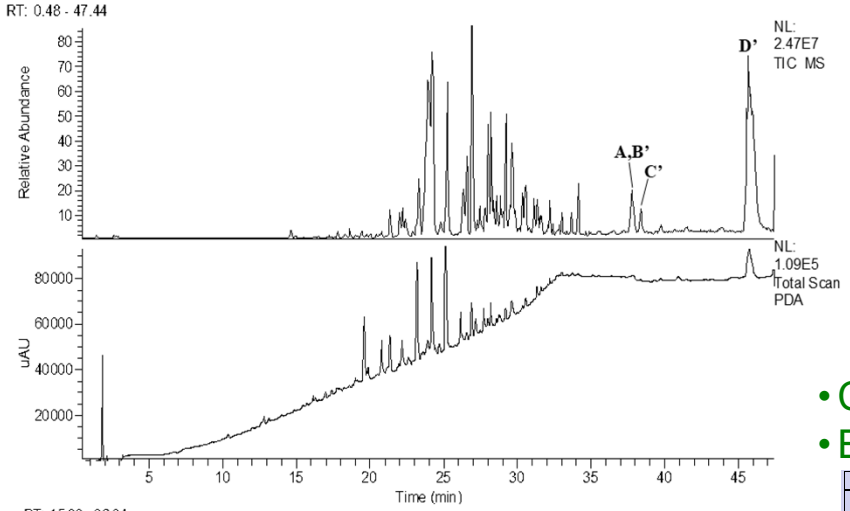


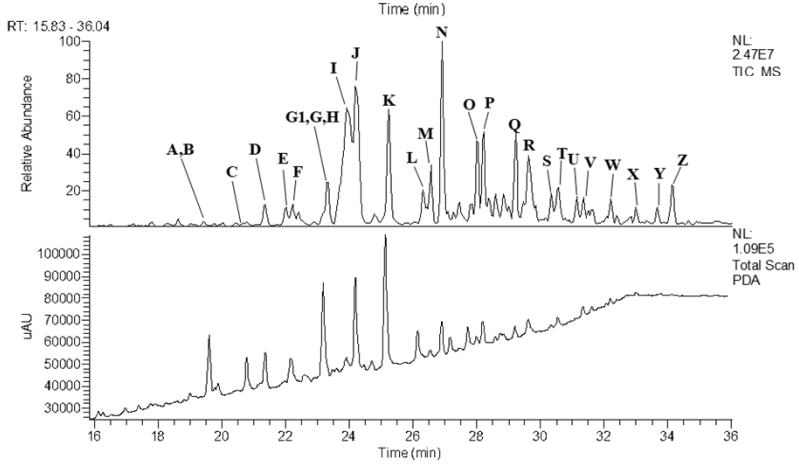
Visualization in Practice: Multi-layer Film Biopharmaceutical Bag Extractables

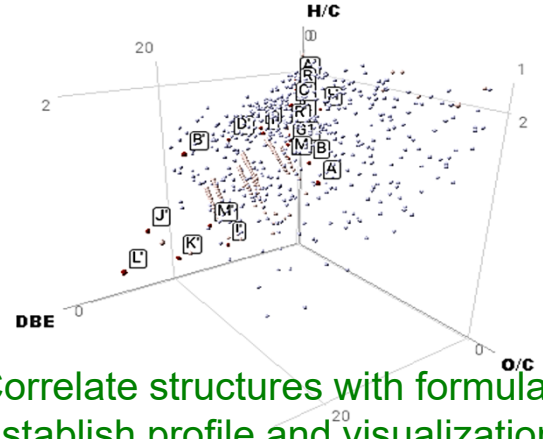
FT-ICR LC-MS total ion chromatograms (right) and mass spectra (left) of IPA extracts of multi-layer biopharmaceutical bags from different sources



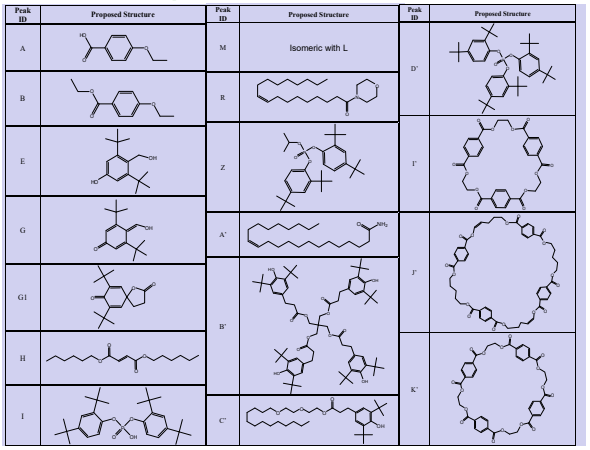
Visualization in Practice: Multi-layer Film Biopharmaceutical Bag Extractables

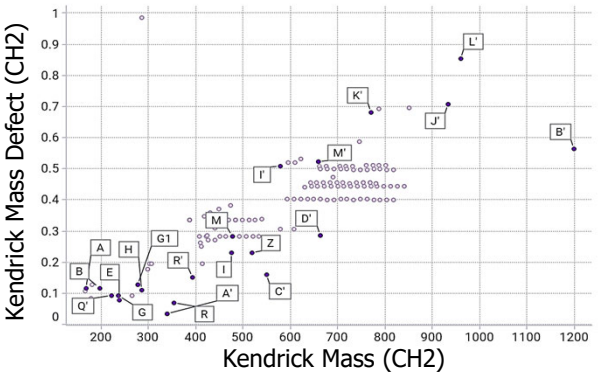


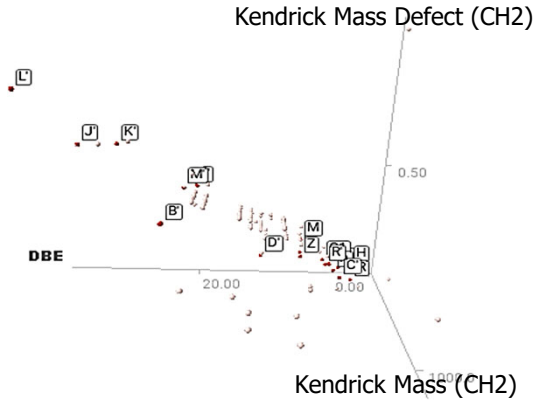




- Correlate structures with formulas
- Establish profile and visualizations

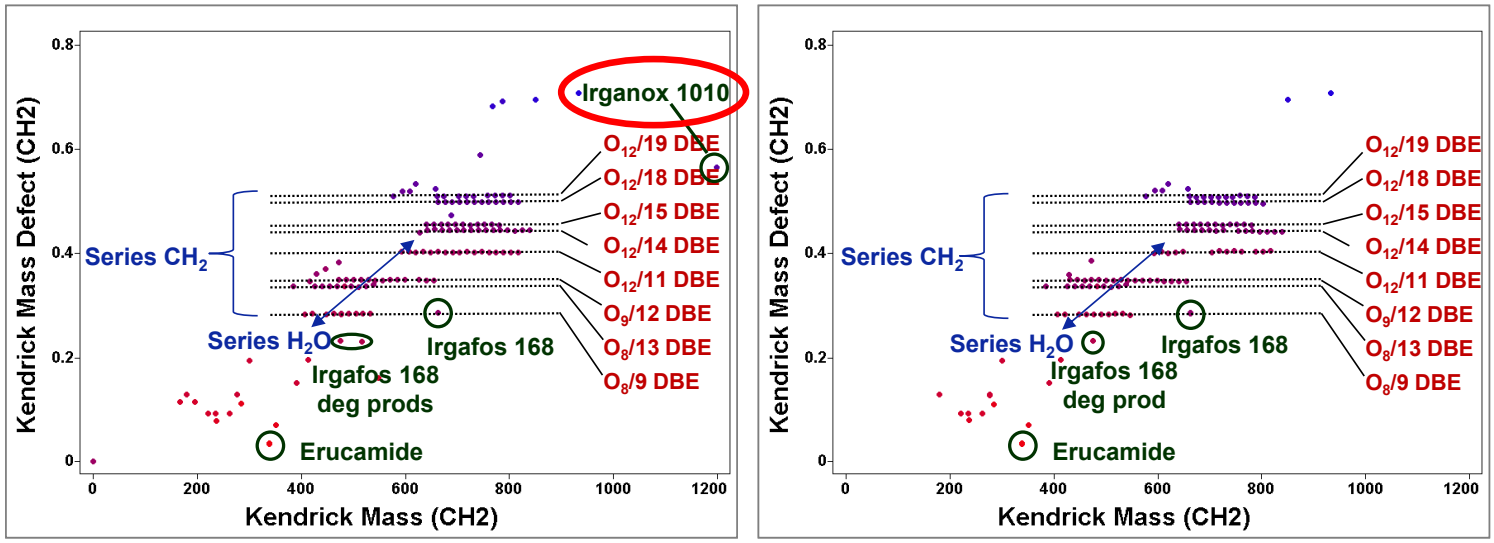






Visualization in Practice: Multi-layer Film Biopharmaceutical Bag Extractables

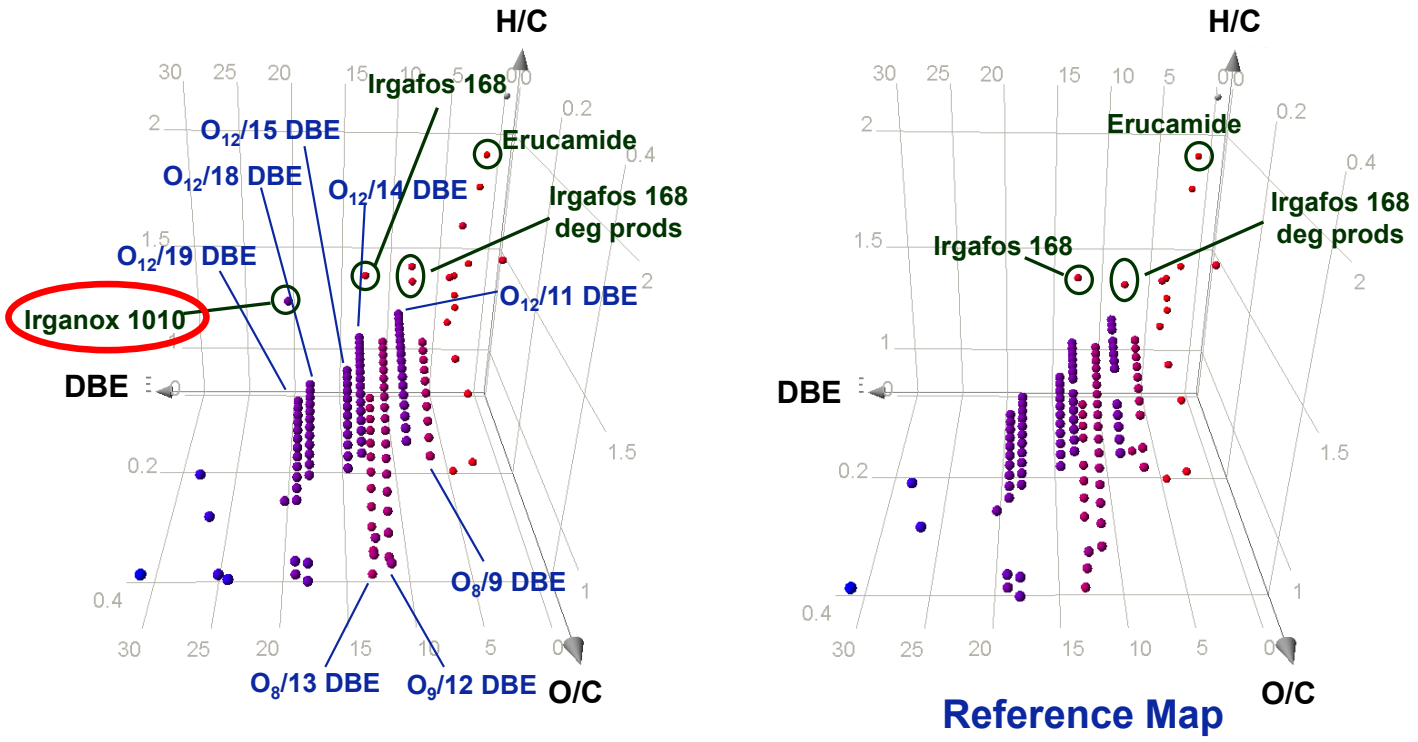
Comparative Kendrick mass defect diagrams



Reference Map

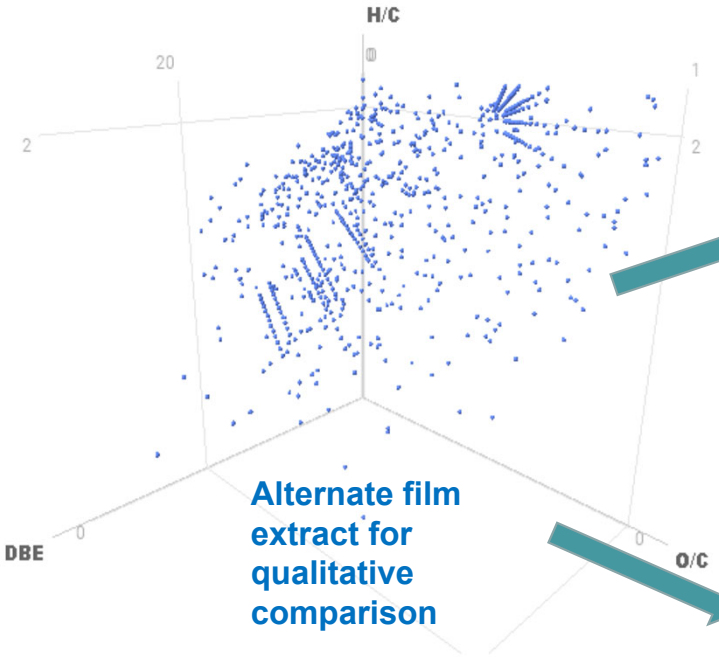
Visualization in Practice: Multi-layer Film Biopharmaceutical Bag Extractables

Comparative 3D van Krevelen diagrams

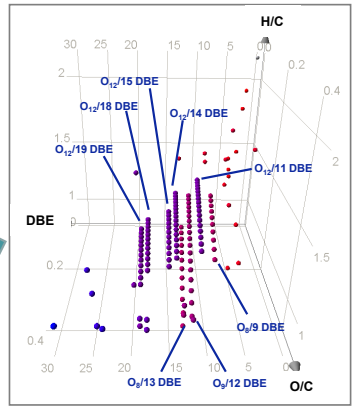


Visualization in Practice: Multi-layer Film Biopharmaceutical Bag Extractables

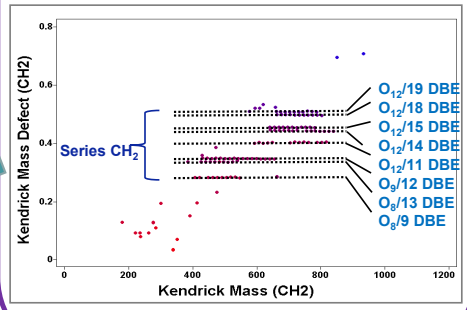
Visualized similarity and anomaly detection

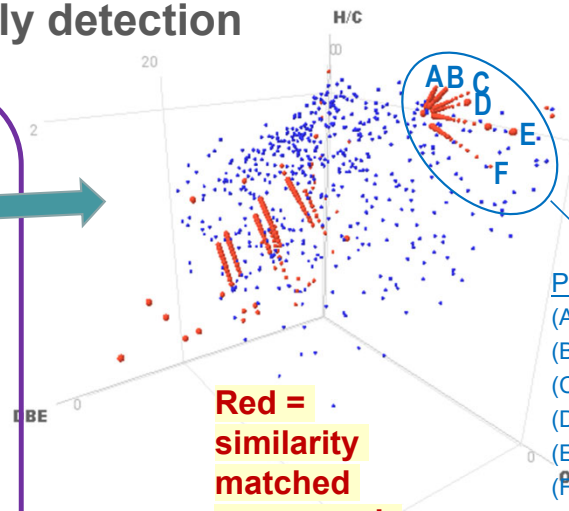


Alternate film extract for qualitative comparison



Library spectral maps

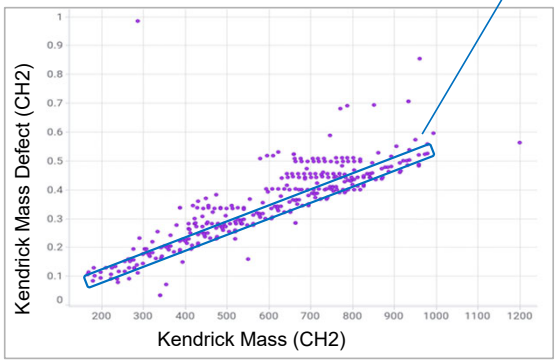




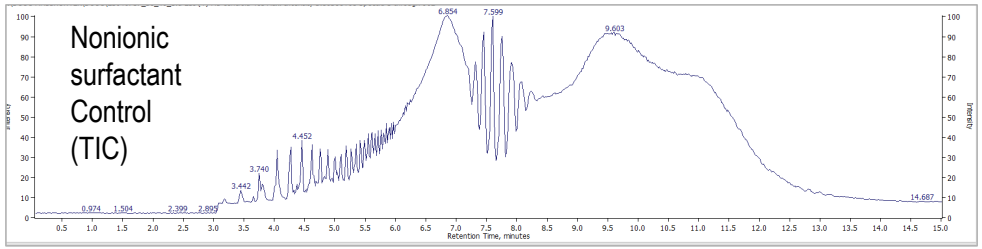
Red = similarity matched compounds

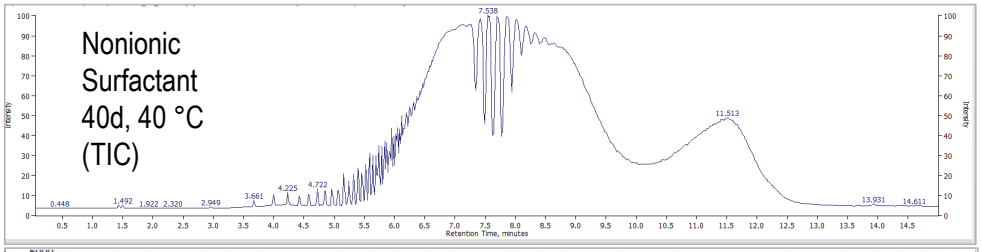
PEG series

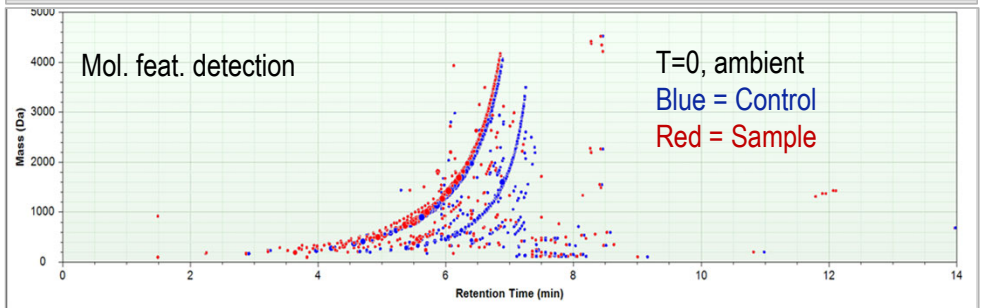
(A) $[(C_2H_4O)_n + H_2O + CH_2 + H]^+$
(B) $[(C_2H_4O)_n + H_2O + H]^+$
(C) $[(C_2H_4O)_n + CH_2O_2 + H]^+$
(D) $[(C_2H_4O)_n + O_2 + H, Na]^+$
(E) $[(C_2H_4O)_n + O_2 + H, NH_4]^+$
(F) $[(C_2H_4O)_n + C_2H_2O_4 + Na]^+$

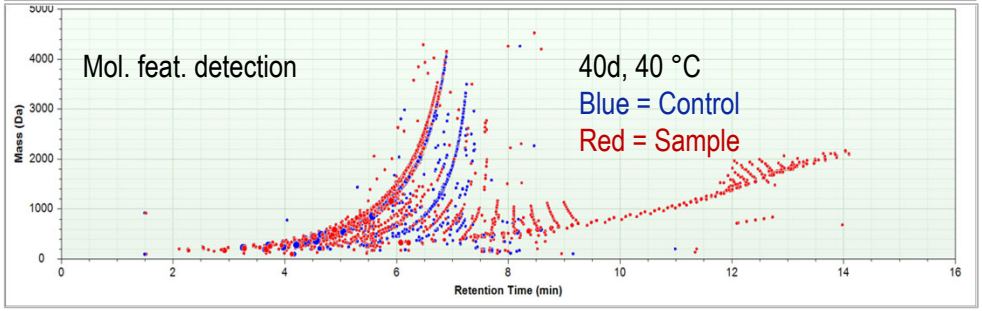


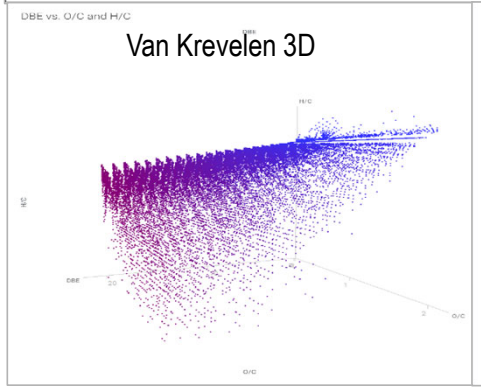
Nonionic Surfactants: Comparative Control and Stress Degradation Samples

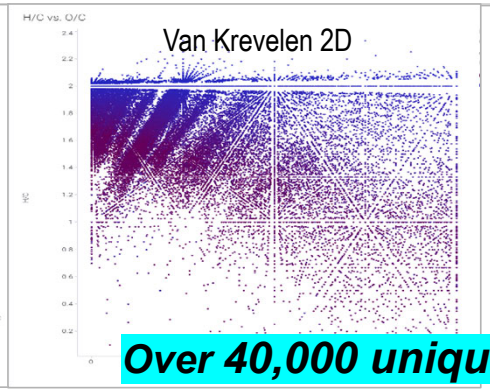


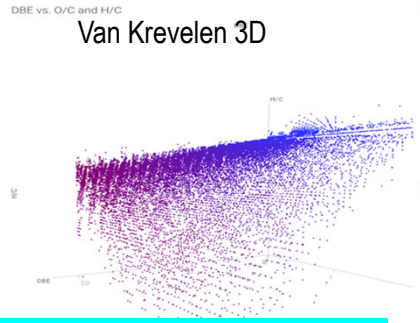


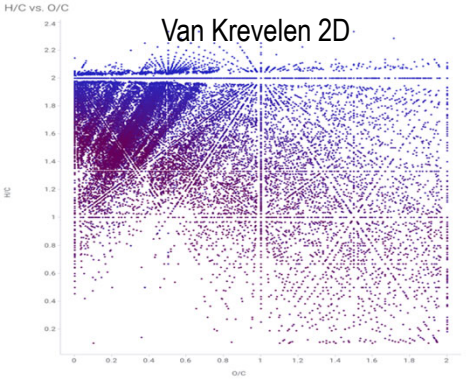






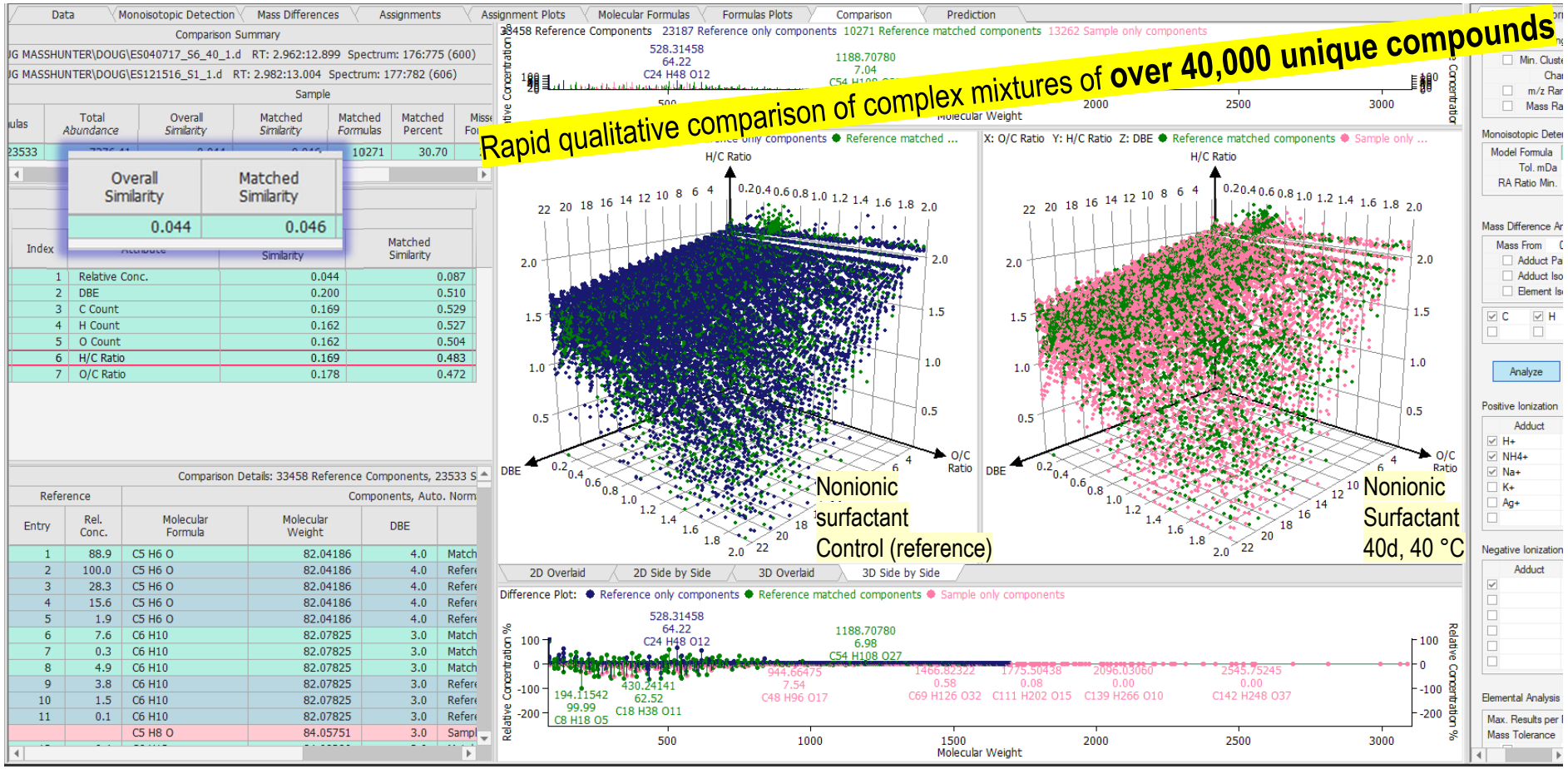




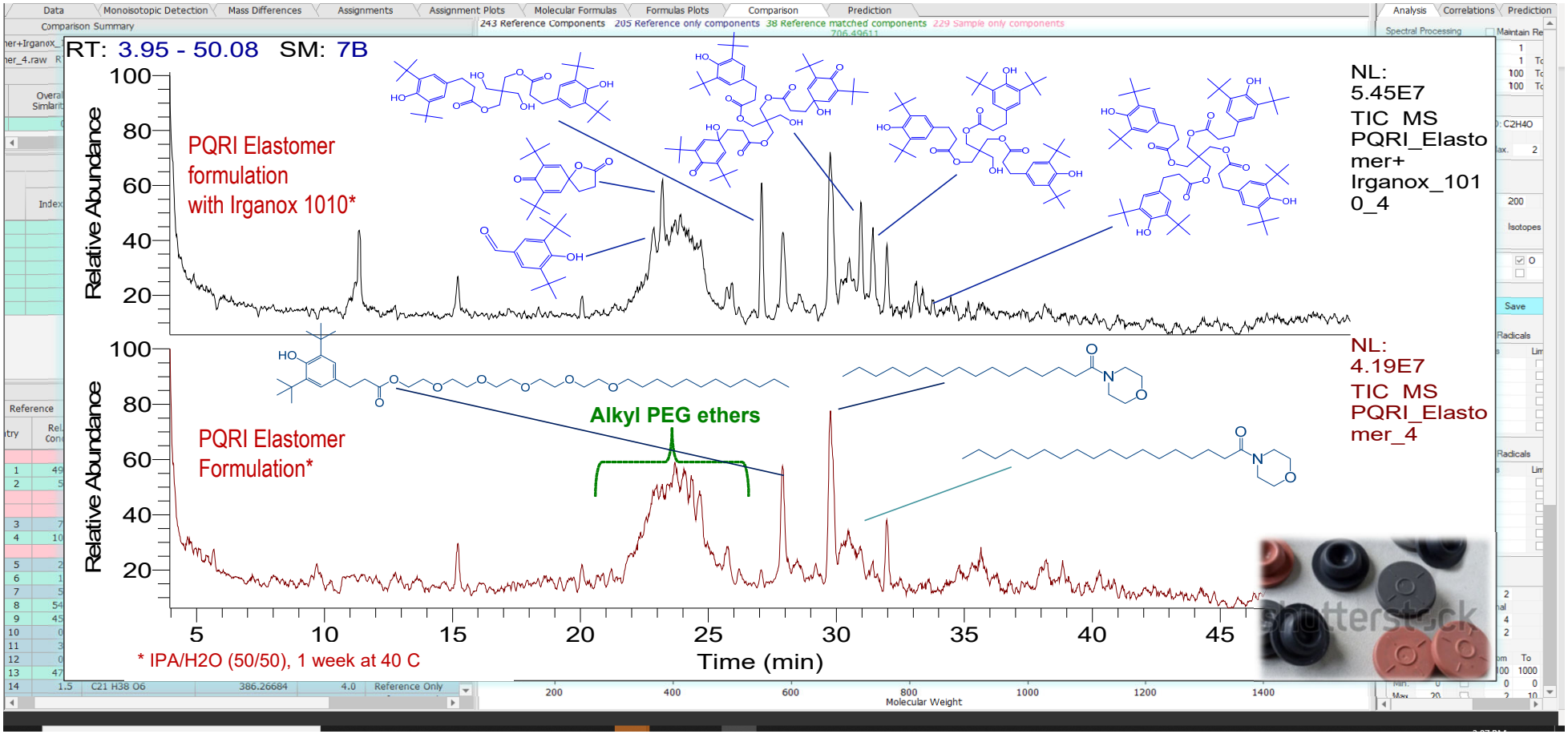


Over 40,000 unique compounds detected

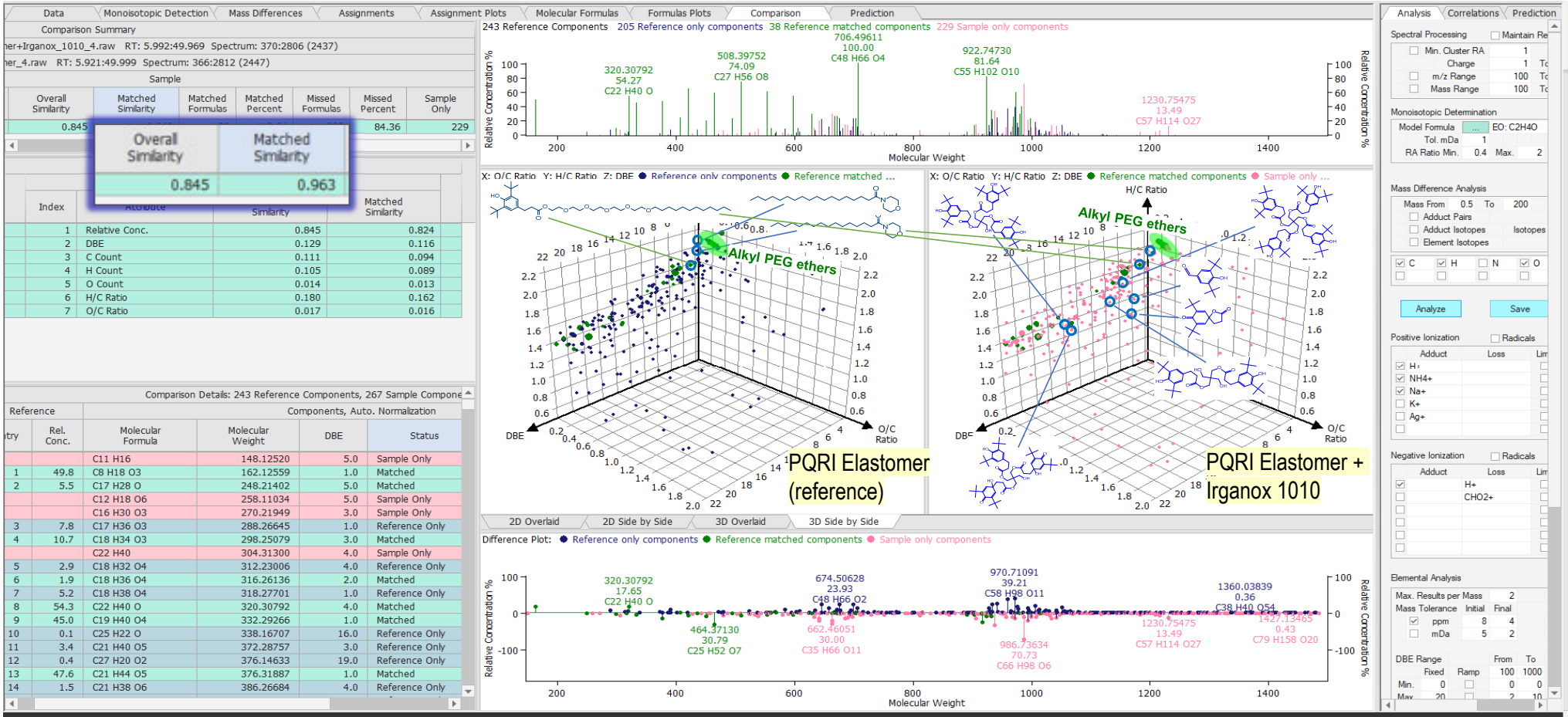
Nonionic Surfactants: Similarity Index and Anomaly Detection



Elastomers: Similarity Index and Anomaly Detection



Elastomers: Similarity Index and Anomaly Detection



In Conclusion, Visualization Approaches Enable…

- Simple, intuitive approach to enabling rapid “at-a-glance” heuristic assessment and visualization of highly complex mixtures of structurally and compositionally diverse chemical entities
- Graphically sorting complex mixture components by series or class based on structural and compositional features
- Confident assignment of elemental composition to higher masses with initial assignment of elemental composition to a limited number of lower MW components
- N-dimensional self-calibrating mass maps based on mass difference correlation
- Rapid and comprehensive pairwise and batch comparison with an index of similarity and anomaly detection
  - Change management
  - Failure mode analysis
  - Material characterization and selection

Acknowledgements

Wolfgang Mueller
Tim Baker (Procter & Gamble)
Bob Strife (Sierra Analytics)
David Stranz (Sierra Analytics)
Scott Campbell (Sierra Analytics)
George Maydwell (Sierra Analytics)
Rob Smith (Prime Labs)
Diane Paskiet (West Pharmaceutical Services)
Cheryl Stults (C & M Technical Consulting)
Ryan Rodgers (NHMFL/FSU)
Alan Marshall (FSU)
Daniel Norwood (Feinberg Norwood & Associates Pharma Consulting)

Thank you!

Many thanks to the PQRI Secretariat and the PQRI/FDA 2021 Conference Organizers!