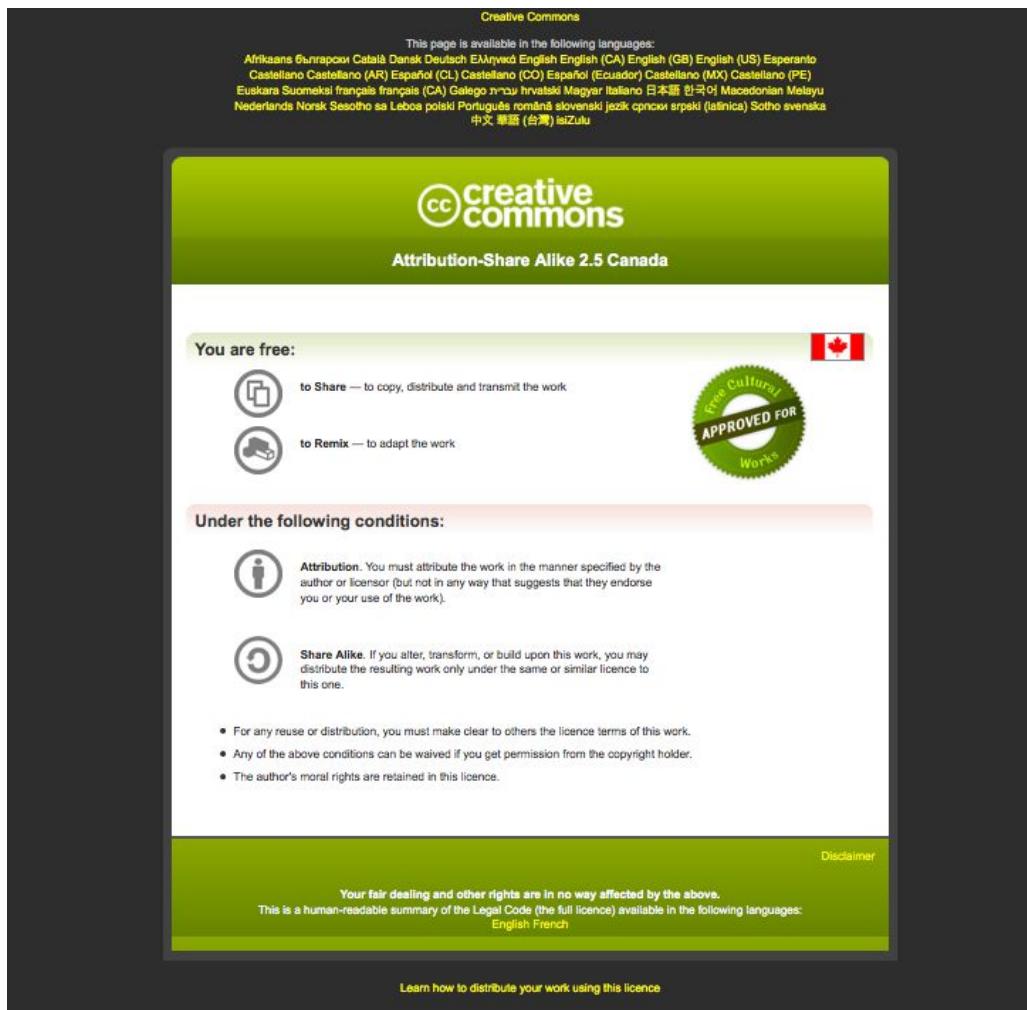




Canadian Bioinformatics Workshops

www.bioinformatics.ca

bioinformaticsdotca.github.io



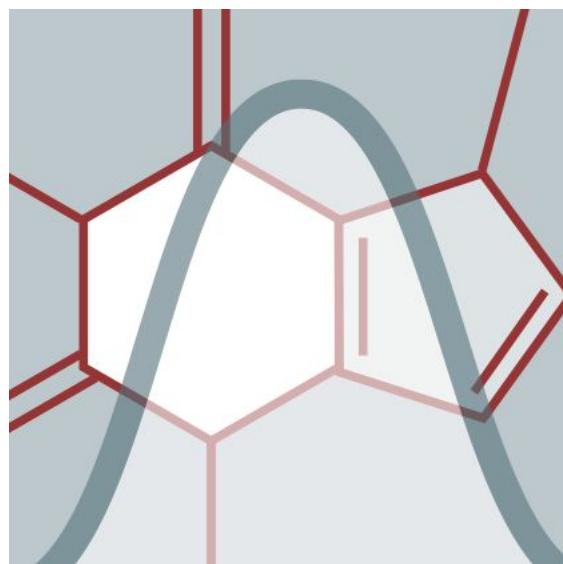
Introduction to Metabolomics



David Wishart

Informatics and Statistics for Metabolomics

July 6-7, 2023



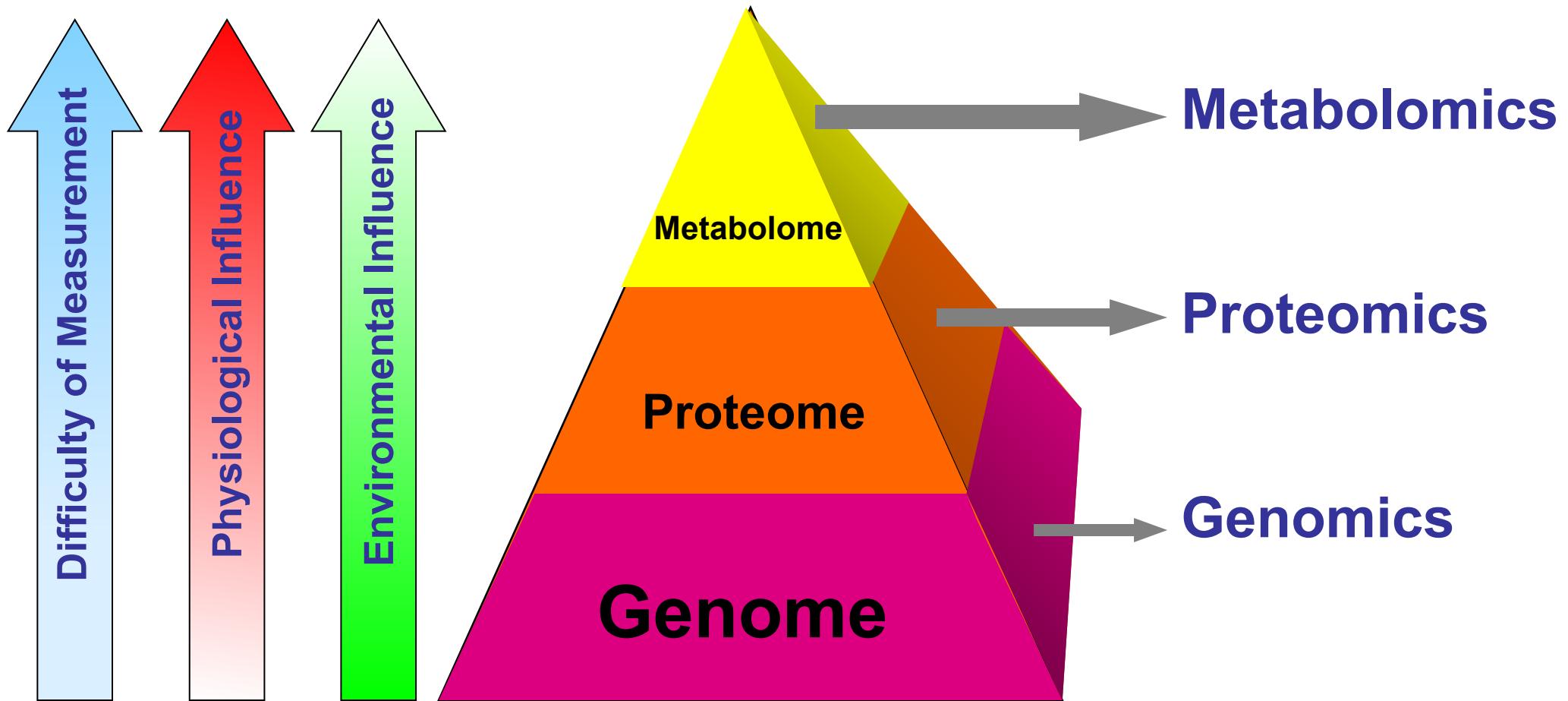
Schedule For July 6, 2023

Time	Module
8:00 (MST)/10:00 (EST)	Arrival & Check-in
8:30 (MST)/10:30 (EST)	Welcome (Nia Hughes)
9:00 (MST)/11:00 (EST)	Module 1: Introduction to Metabolomics (David Wishart)
10:30 (MST)/12:30 (EST)	Break/Lunch (45 min)
11:15 (MST)/13:15 (EST)	Module 2: Targeted, Quantitative Metabolomics (David Wishart)
12:15 (MST)/14:15 (EST)	Lunch/Break (45 min)
13:00 (MST)/15:00 (EST)	Module 3 (Lab): Quantitative Metabolomics (David Wishart)
15:00 (MST)/17:00 (EST)	Break (30 min)
15:30 (MST)/17:30 (EST)	Module 4: Databases for Biological Interpretation (David Wishart)
17:00 (MST)/19:00 (EST)	Finish

Learning Objectives

- To define metabolomics and the size of the metabolome(s)
- To appreciate the importance and potential applications of metabolomics
- To understand the operational principles of key metabolomics technologies (LC, GC, MS and NMR)
- To understand the difference between targeted and untargeted metabolomics

The Pyramid of Life



What is Metabolomics?

- **Genomics** - A field of life science research that uses High Throughput (HT) technologies to identify and/or characterize all the *genes* in a given cell, tissue or organism (i.e., the genome).
- **Metabolomics** - A field of life science research that uses High Throughput (HT) technologies to identify and/or characterize all the *small molecules or metabolites* in a given cell, tissue or organism (i.e., the metabolome).

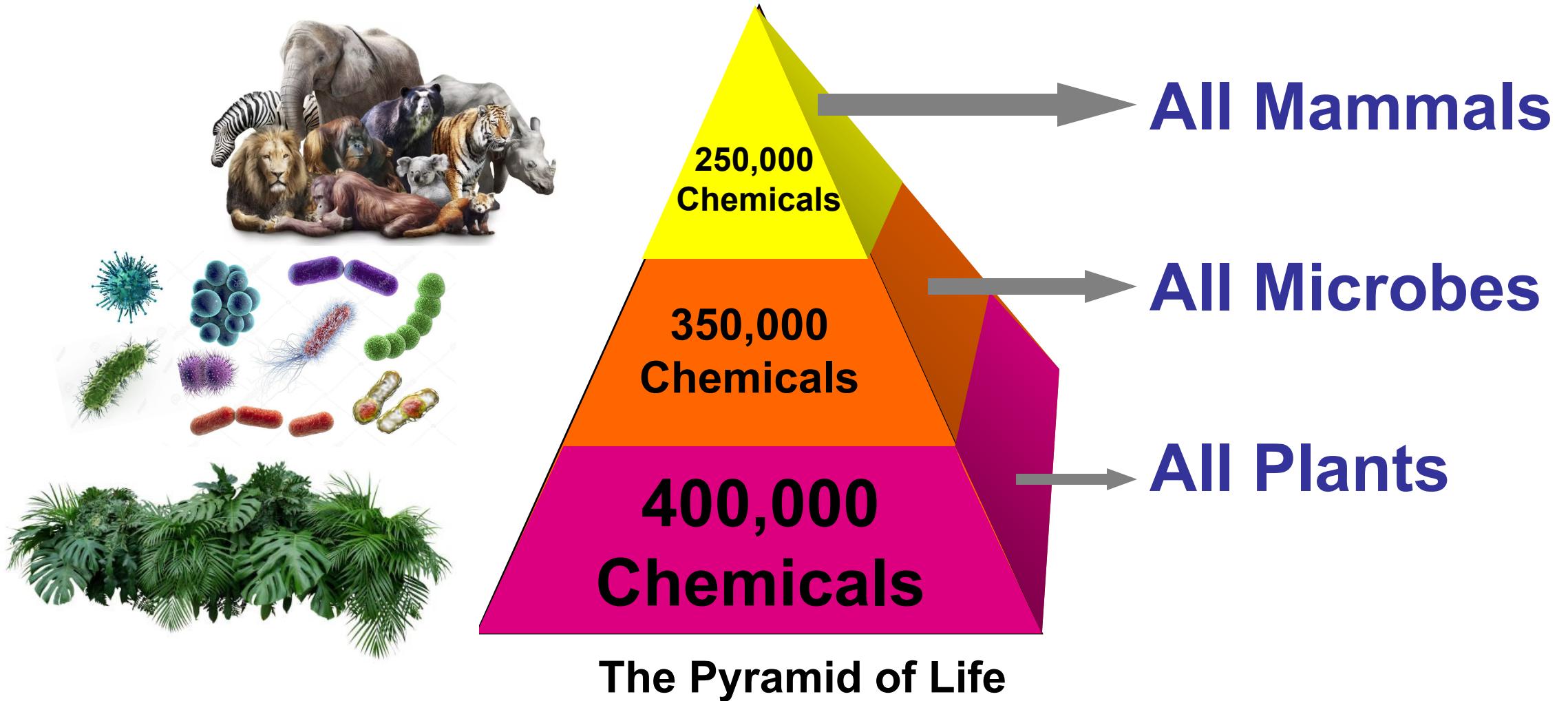
What is a Metabolite?

- Any organic molecule detectable in the body with a MW < 1500 Da
- Includes peptides, oligonucleotides, sugars, nucleosides, organic acids, ketones, aldehydes, amines, amino acids, lipids, steroids, alkaloids, foods, food additives, toxins, pollutants, drugs and drug metabolites
- Includes human & microbial products
- Concentration > detectable (1 pM)

What is a Metabolome?

- The complete collection of small molecule metabolites in a cell, organ, tissue or organism
- Includes endogenous and exogenous molecules as well as transient or even theoretical molecules
- Defined by the detection technology (NMR, MS)
- Metabolome size is always ill-defined, depends on the organism and is constantly evolving (2500 compounds in 2005, 250,000 compounds in 2023, 5 million compounds by 2025??)

Different Metabolomes



Human Metabolomes (2023)

3670 (T3DB)

Toxins/Env. Chemicals

1250 (DrugBank)

Drug metabolites

70,926 (FooDB)

Food additives/Phytochemicals

2634 (DrugBank)

Drugs

255,207 (HMDB)

Endogenous metabolites



Theoretical Human Metabolomes

400,000 (Lipidome)

Lipids/Lipid derivatives

10,000 (Drug metabolome)

Secondary drug metabolites

400,000 (Food metabolome)

Secondary food metabolites

4,000,000 (Secondome)

Secondary endogenous metabolites

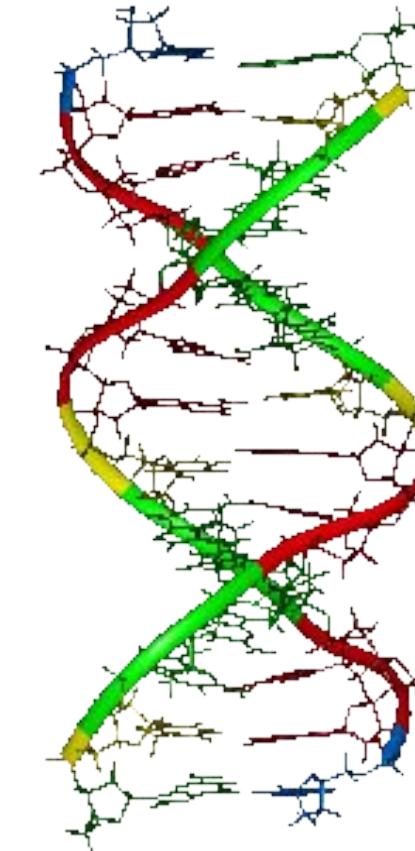
M mM μ M nM pM fM

Why is Metabolomics Important?

Small Molecules Count...

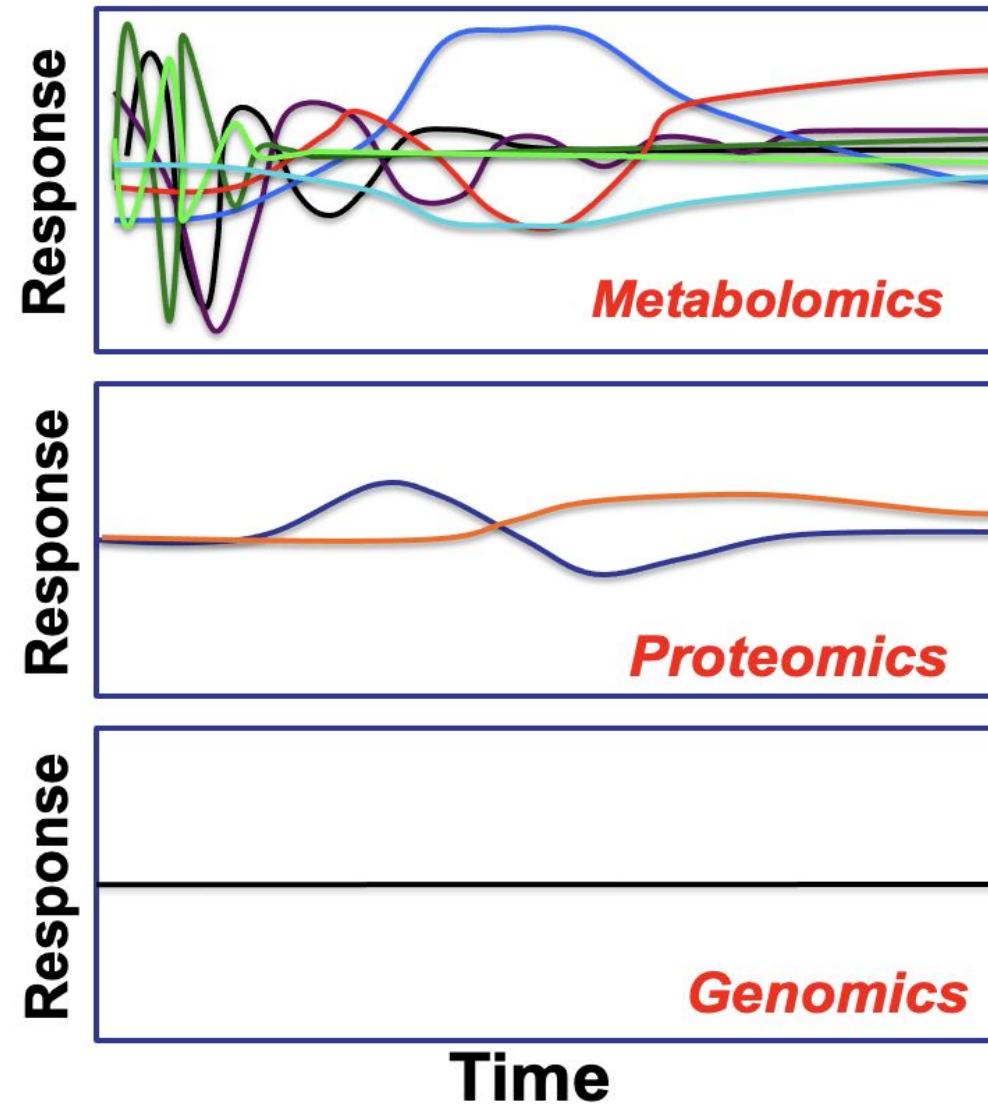
- **>95% of all diagnostic clinical assays test for small molecules**
- **89% of all known drugs are small molecules**
- **50% of all drugs are derived from pre-existing metabolites**
- **30% of identified genetic disorders involve diseases of small molecule metabolism**
- **Small molecules serve as cofactors and signaling molecules to 1000's of proteins**

Metabolites Are the Canaries of the Genome

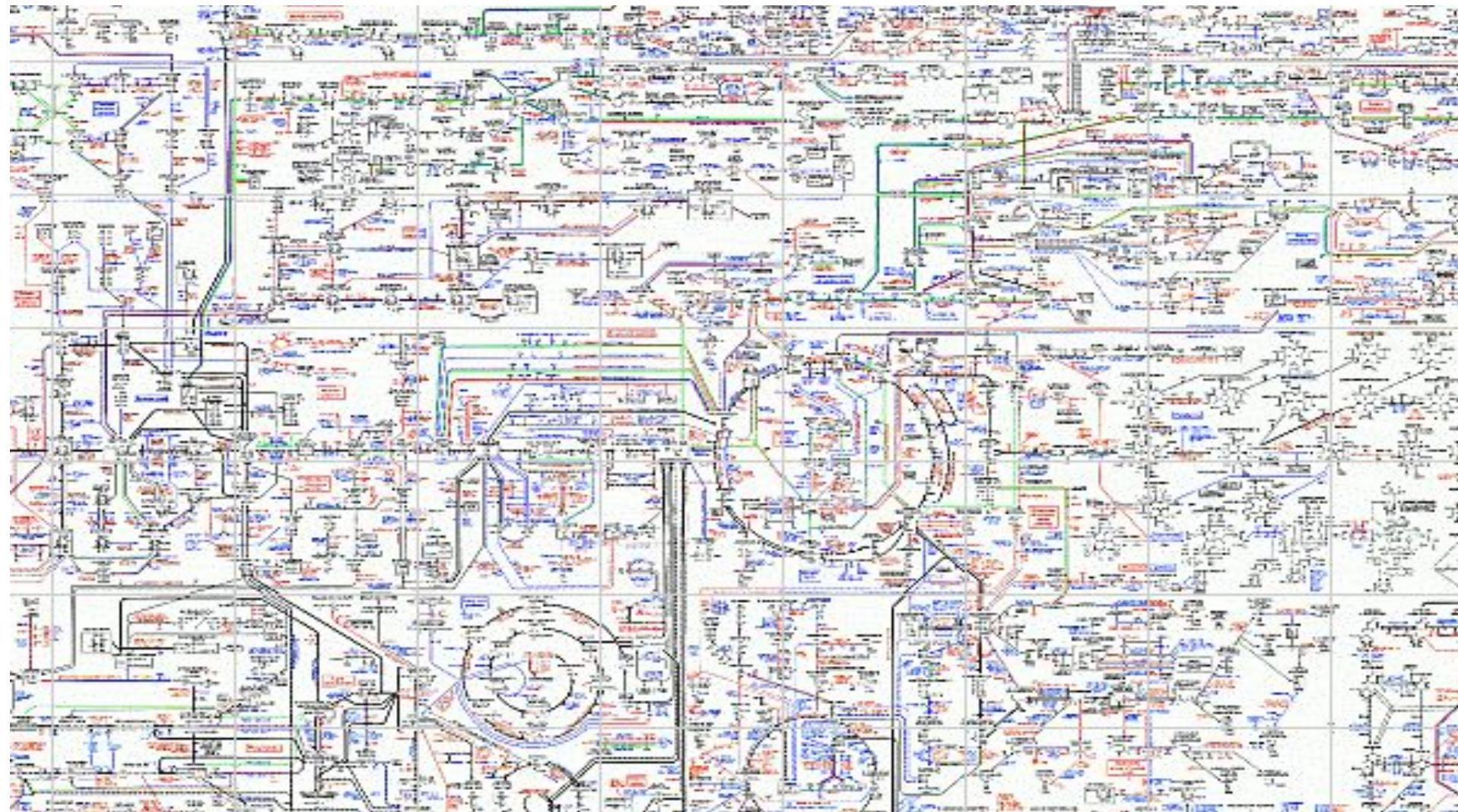


A single base change can lead to a 10,000X change in metabolite levels

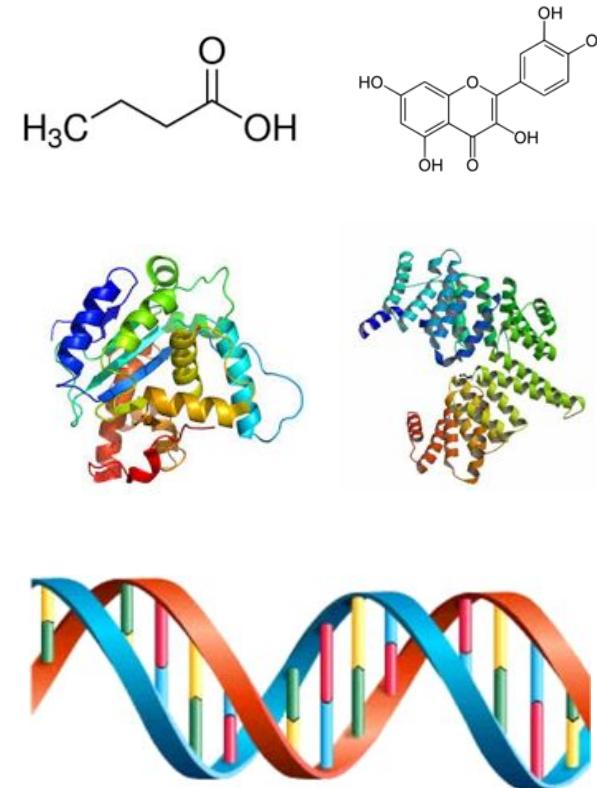
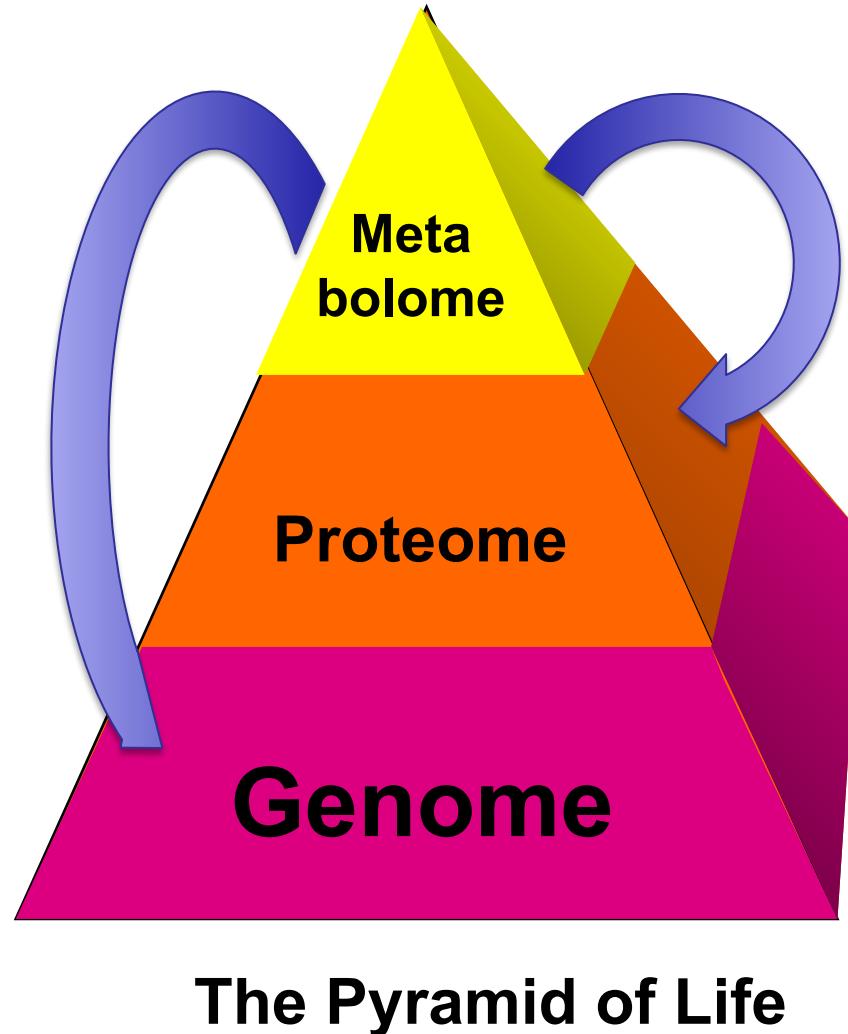
Metabolomics is More Time Sensitive Than Other “Omics”



Metabolism is “Understood”



The Metabolome is Connected to all other “Omes”



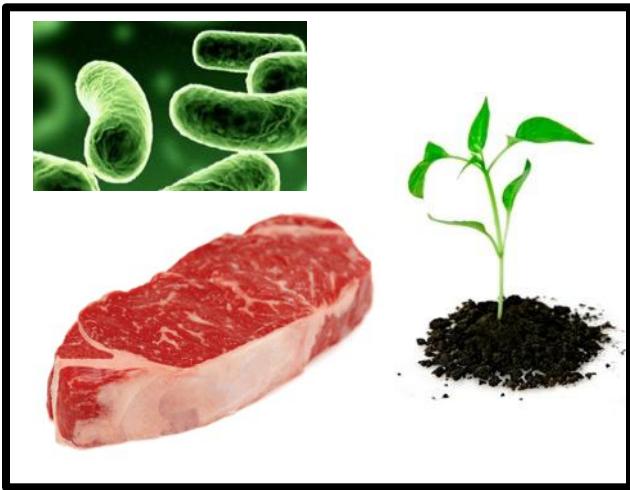
The Metabolome is Connected to All Other “Omes”

- Small molecules (i.e., AMP, CMP, GMP, TMP) are the primary constituents of the genome & transcriptome
- Small molecules (i.e., the 20 amino acids) are the primary constituents of the proteome
- Small molecules (i.e., lipids and glycolipids) give cells their shape, form, integrity and structure
- Small molecules (sugars, lipids, AAs, ATP) are the source of all cellular energy
- Small molecules serve as cofactors and signaling molecules for both the proteome and the genome
- *The genome & proteome largely evolved to catalyze the chemistry of small molecules*

Metabolomics Applications

- Genetic Disease Tests
- Nutritional Analysis
- Clinical Blood Analysis
- Clinical Urinalysis
- Cholesterol Testing
- Drug Compliance
- Transplant Monitoring
- MRS and CS imaging
- Toxicology Testing
- Clinical Trial Testing
- Fermentation Monitoring
- Food & Beverage Tests
- Nutraceutical Analysis
- Drug Phenotyping
- Water Quality Testing
- Petrochemical Analysis

Metabolomics Methods



Biological or Tissue Samples



Extraction



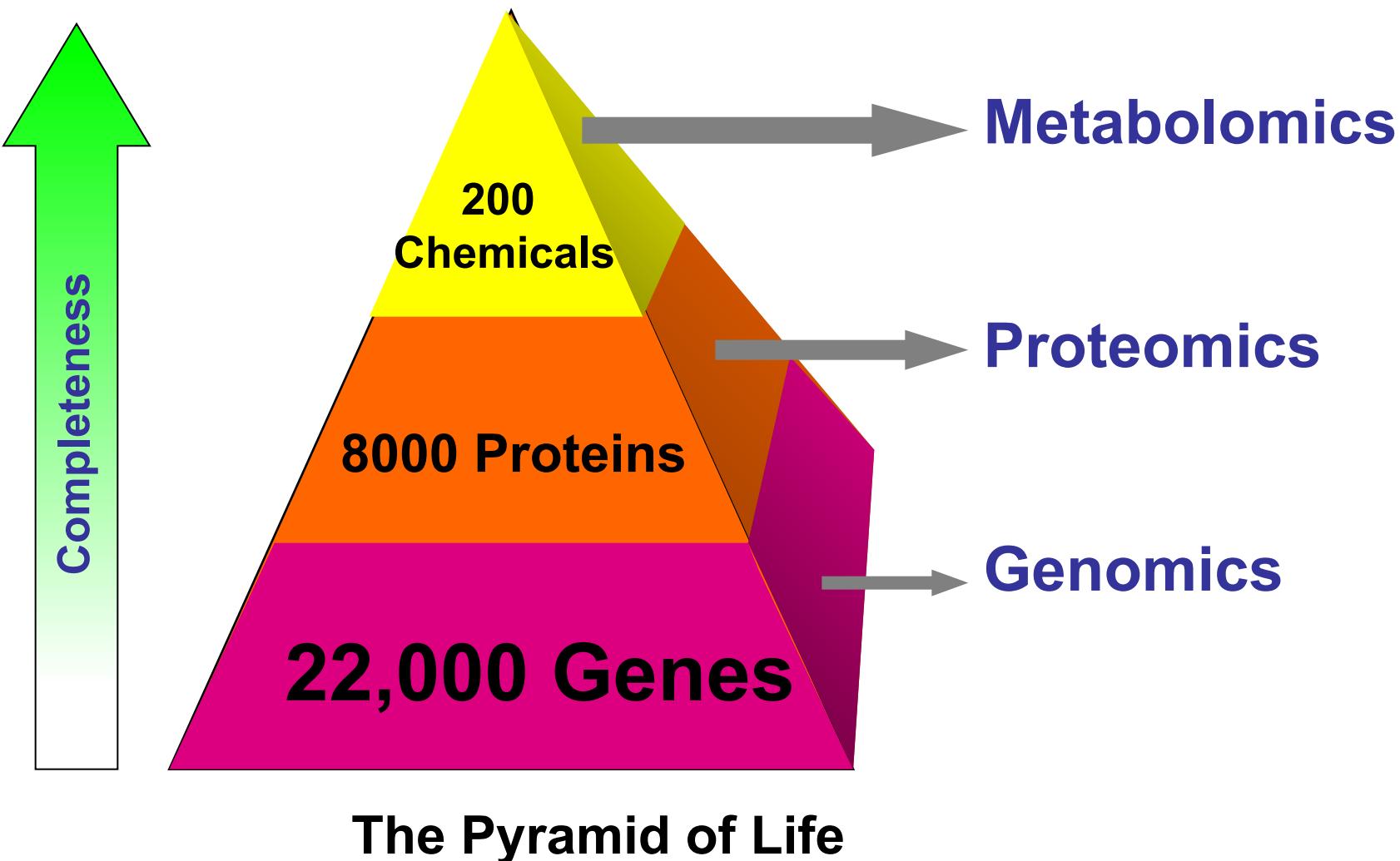
Biofluids or Extracts

Data Analysis

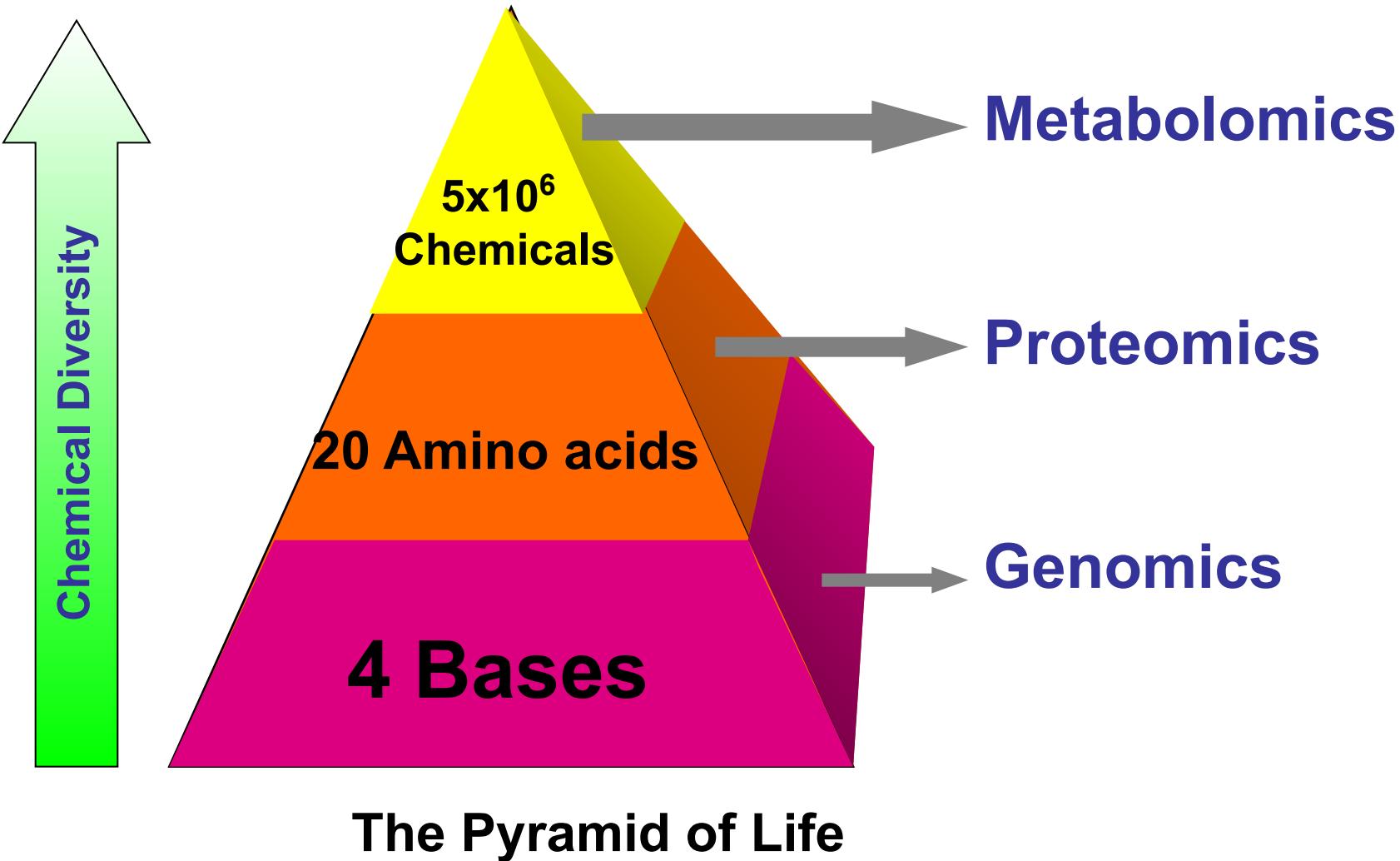


Chemical Analysis

Comparing “Omics” Coverage



Why Metabolomics is Difficult

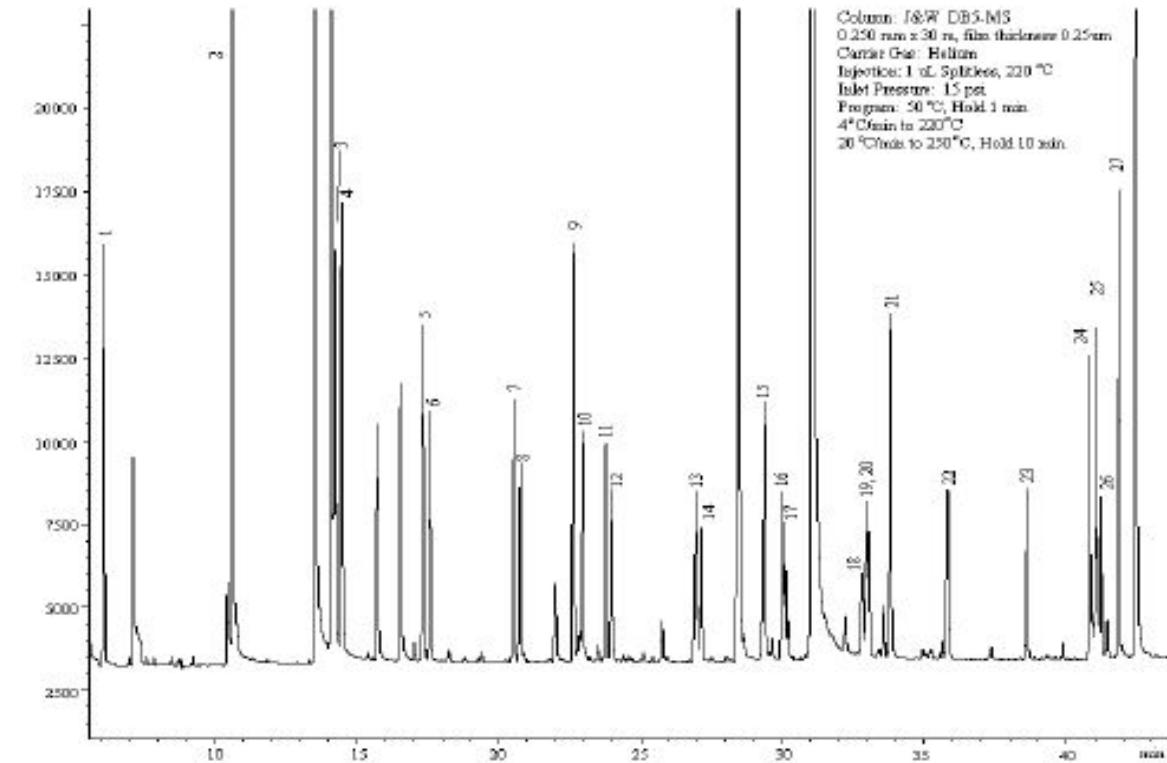
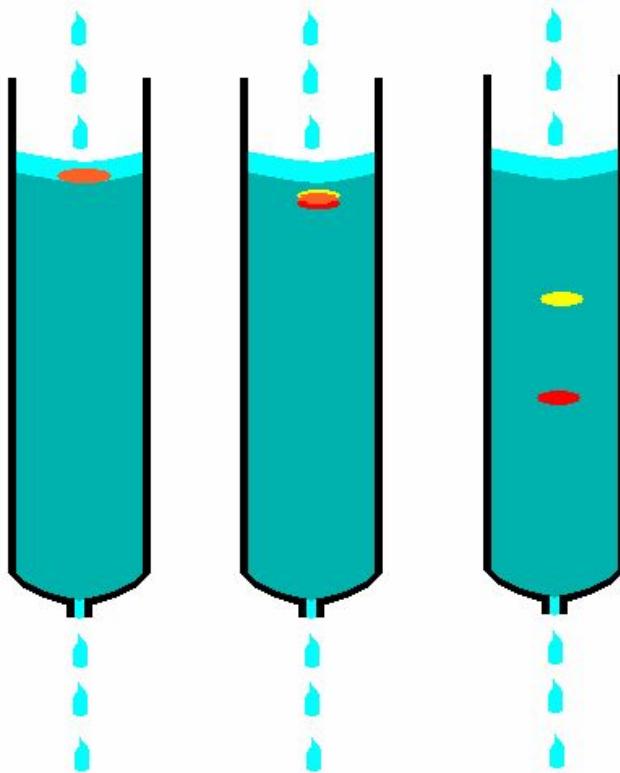


Metabolomics Technologies



- UPLC, HPLC
- CE/microfluidics
- LC-MS
- FT-MS
- QqQ-MS
- NMR spectroscopy
- X-ray crystallography
- GC-MS
- FTIR

Chromatography

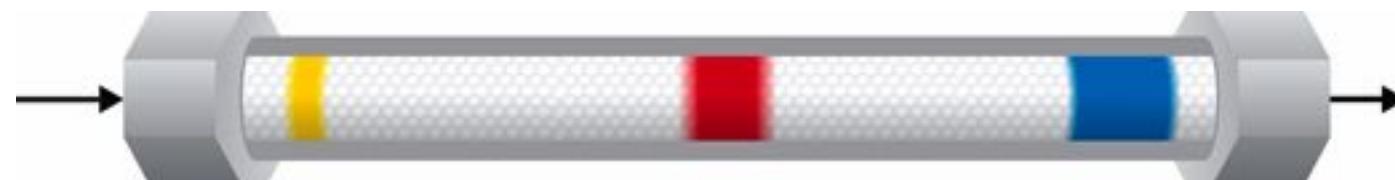


Chromatography

- The separation of components in a mixture that involves passing the mixture dissolved in a "mobile phase" through a *stationary phase*, which separates the analyte to be measured from other molecules in the mixture based on differential partitioning between the mobile and stationary phases
- Column, thin layer, liquid, gas, affinity, ion exchange, size exclusion, reverse phase, normal phase, gravity, high pressure

High Pressure (Performance) Liquid Chromatography - HPLC

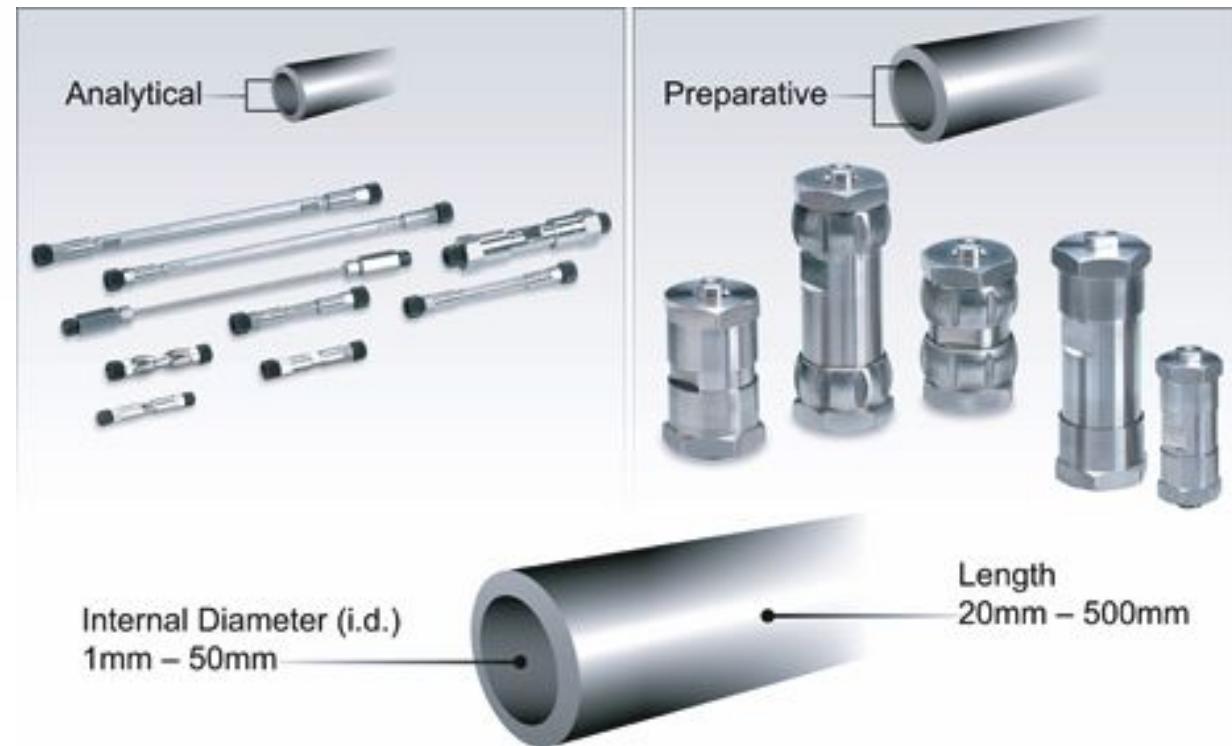
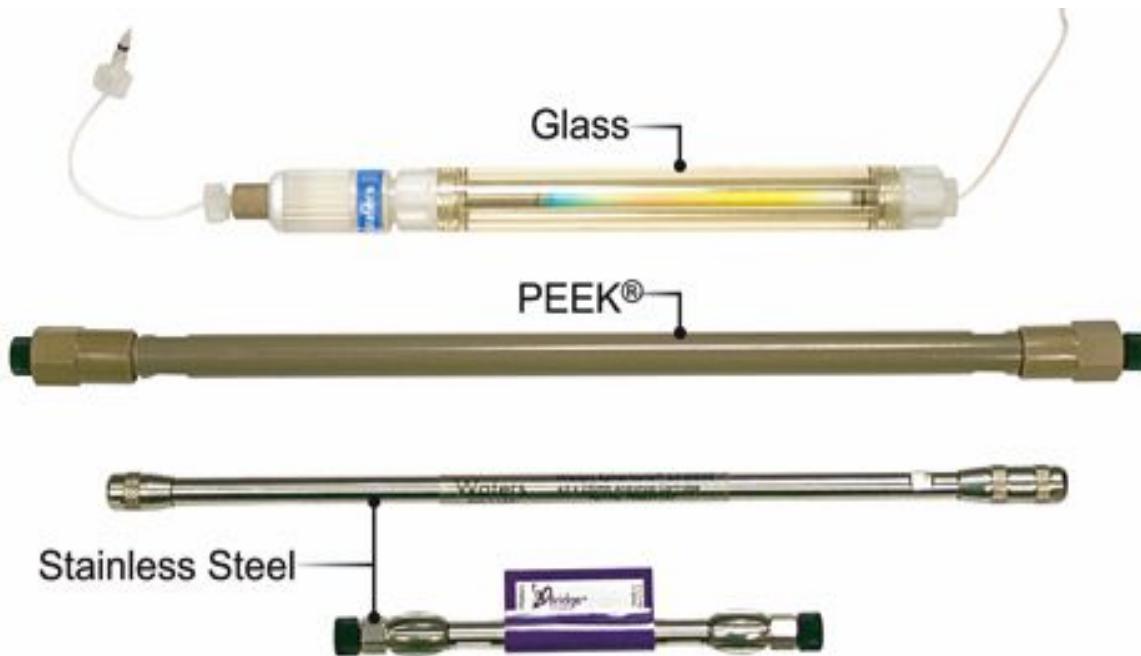
- Developed in 1970's
- Uses high pressures (6000 psi) and smaller (5 μm), pressure-stable particles
- Allows compounds to be detected at ppt (parts per trillion) level
- Allows separation of many types of polar and nonpolar compounds



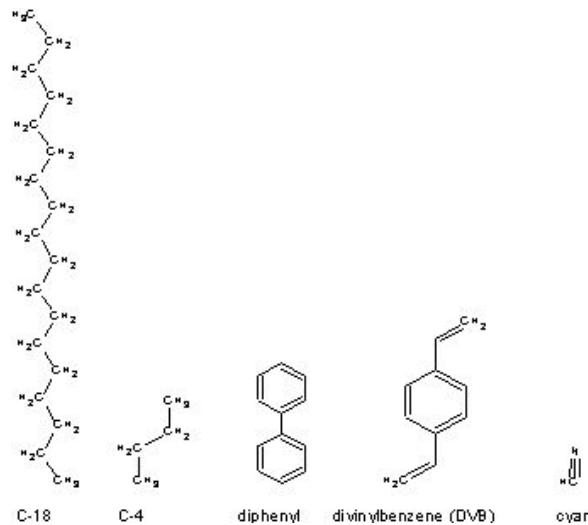
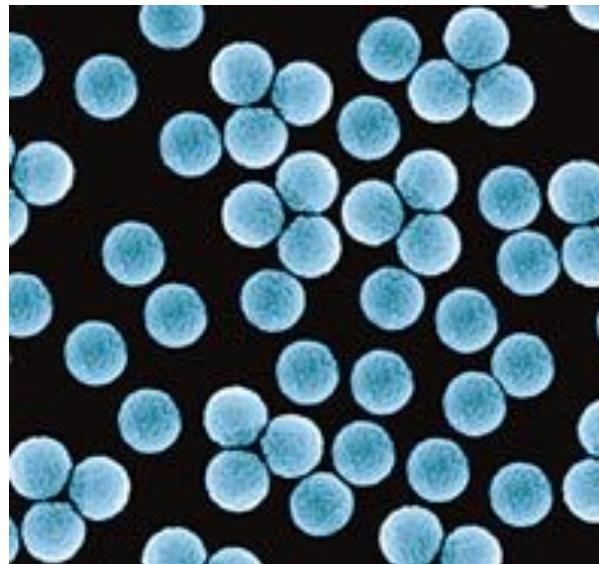
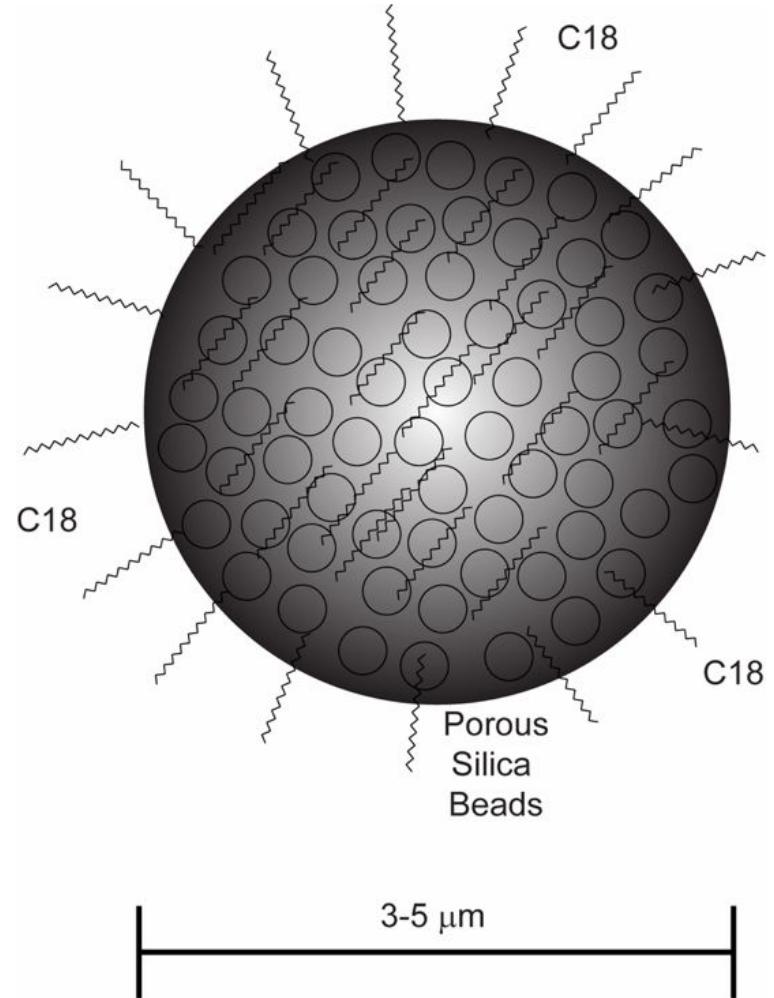
HPLC Modalities

- **Reversed phase** – for separation of non-polar molecules (non-polar stationary phase, polar mobile phase)
- **Normal phase** – for separation of non-polar molecules (polar stationary phase, non-polar/organic mobile phase)
- **HILIC** – hydrophilic interaction liquid chromatography for separation of polar molecules (polar stationary phase, mixed polar/nonpolar mobile phase)

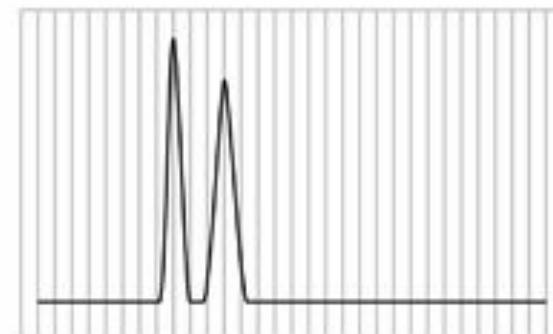
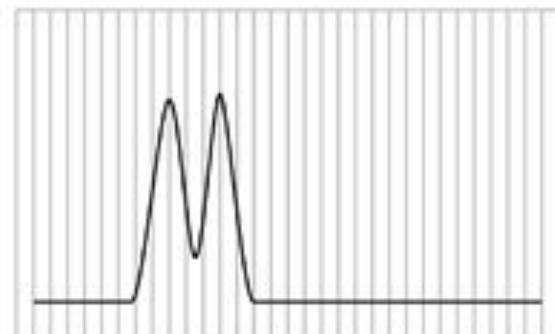
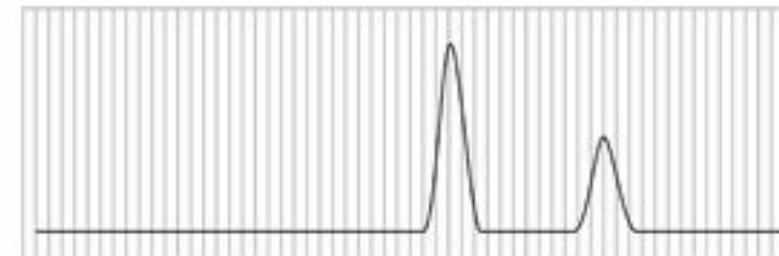
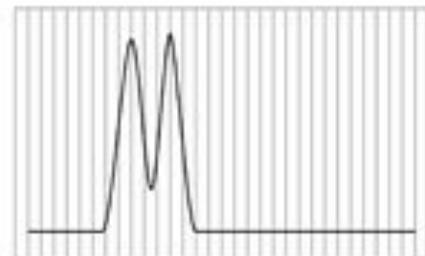
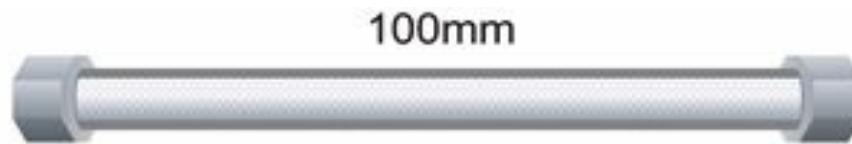
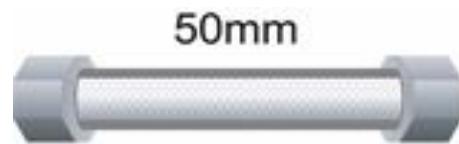
HPLC Columns



Reverse Phase Column

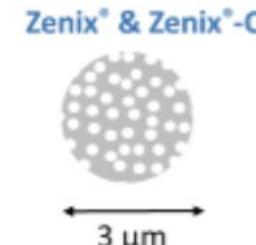


HPLC Separation Efficiency



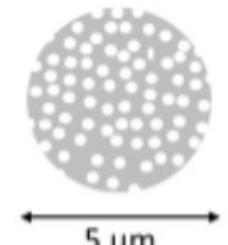
Unix™

$1.8 \mu\text{m}$



Zenix® & Zenix®-C

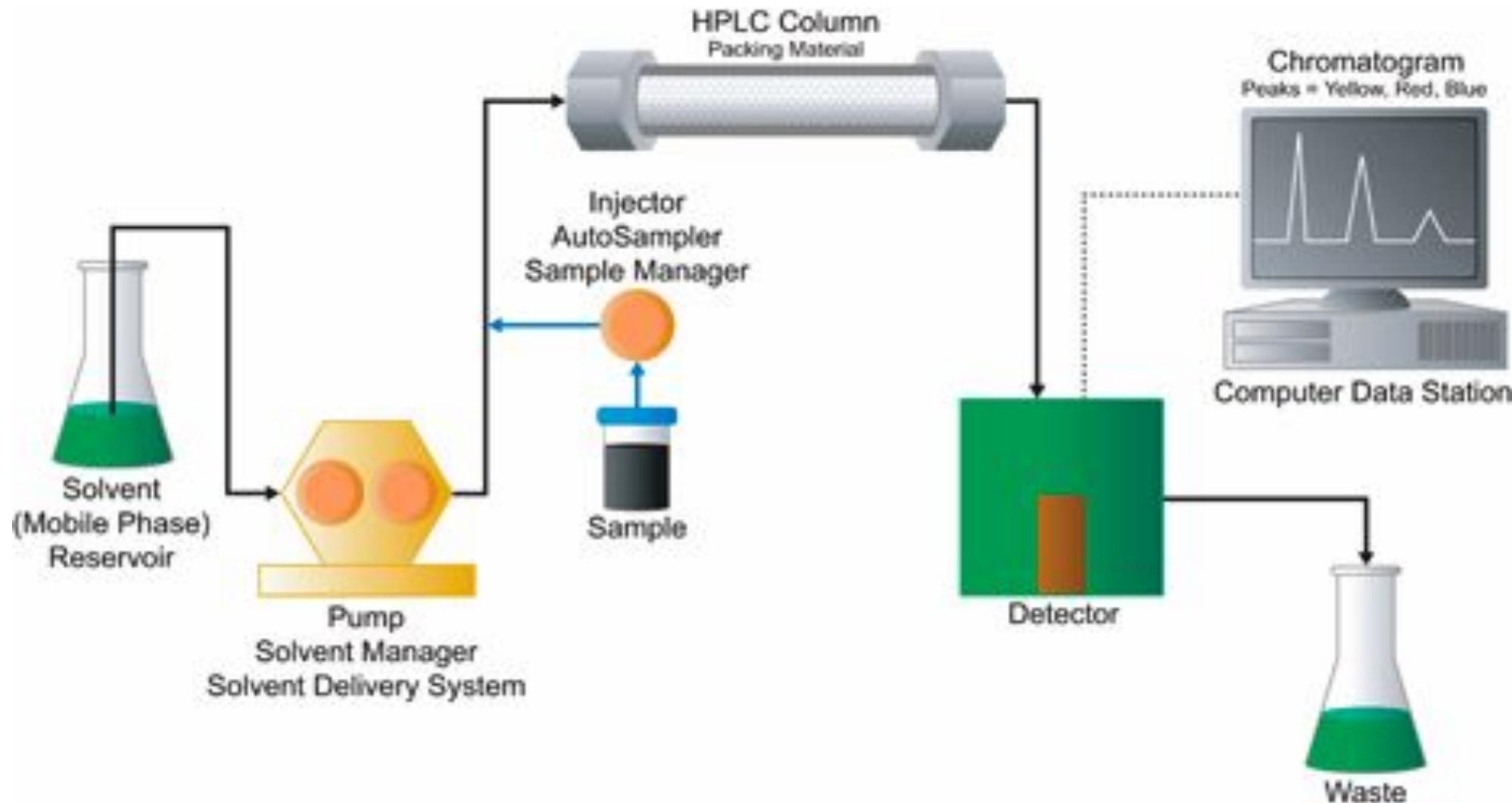
$3 \mu\text{m}$



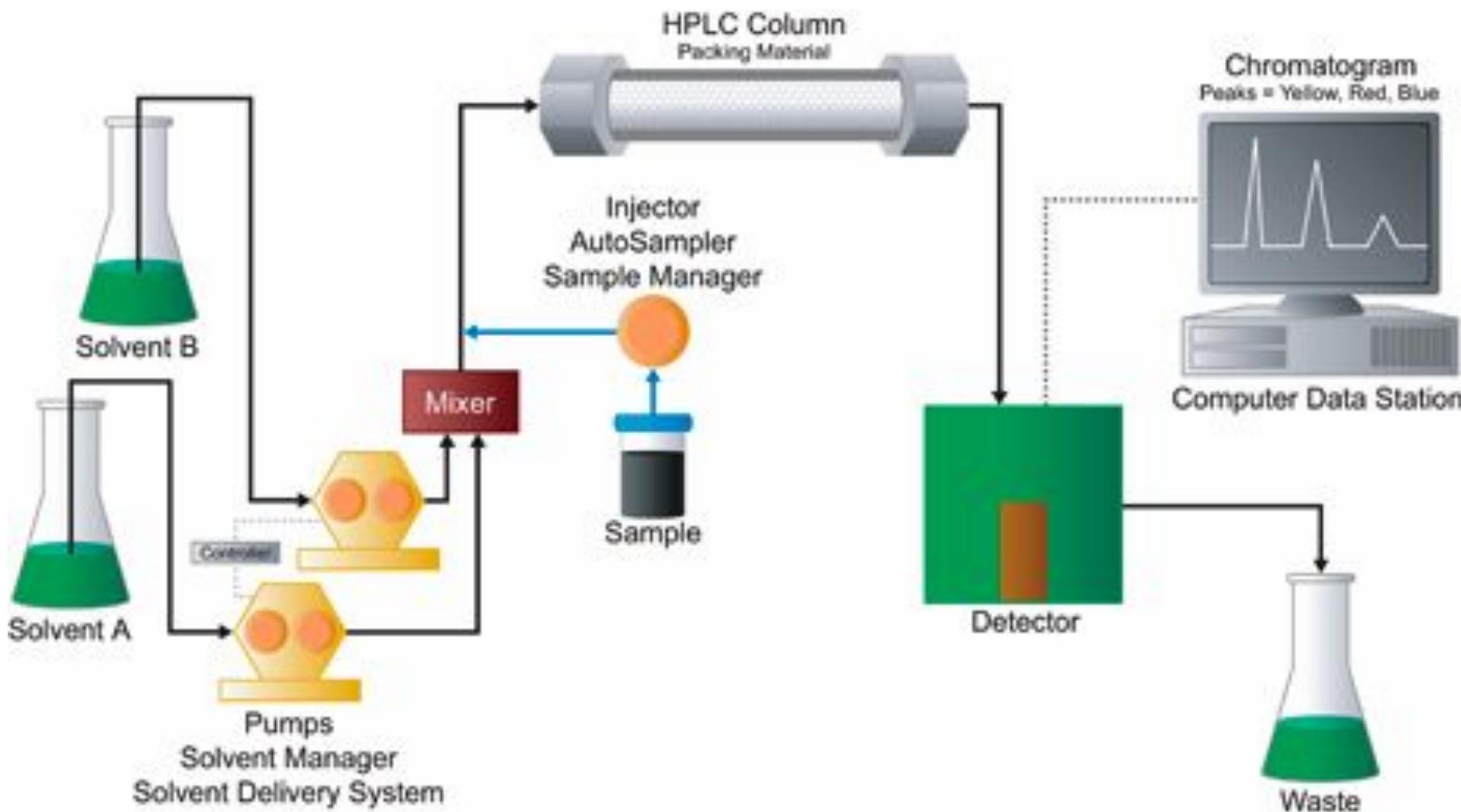
SRT® & SRT®-C

$5 \mu\text{m}$

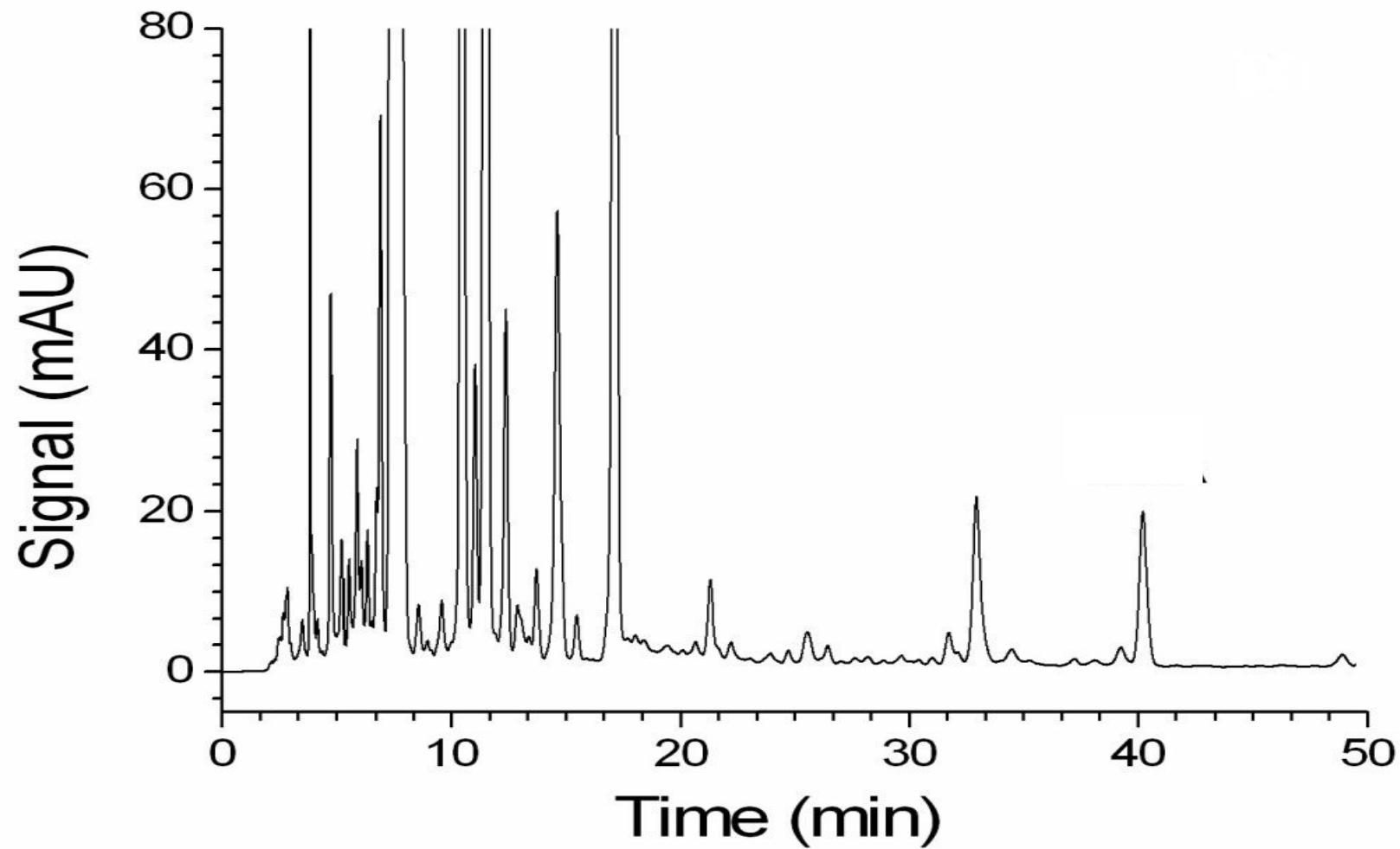
HPLC Schematic



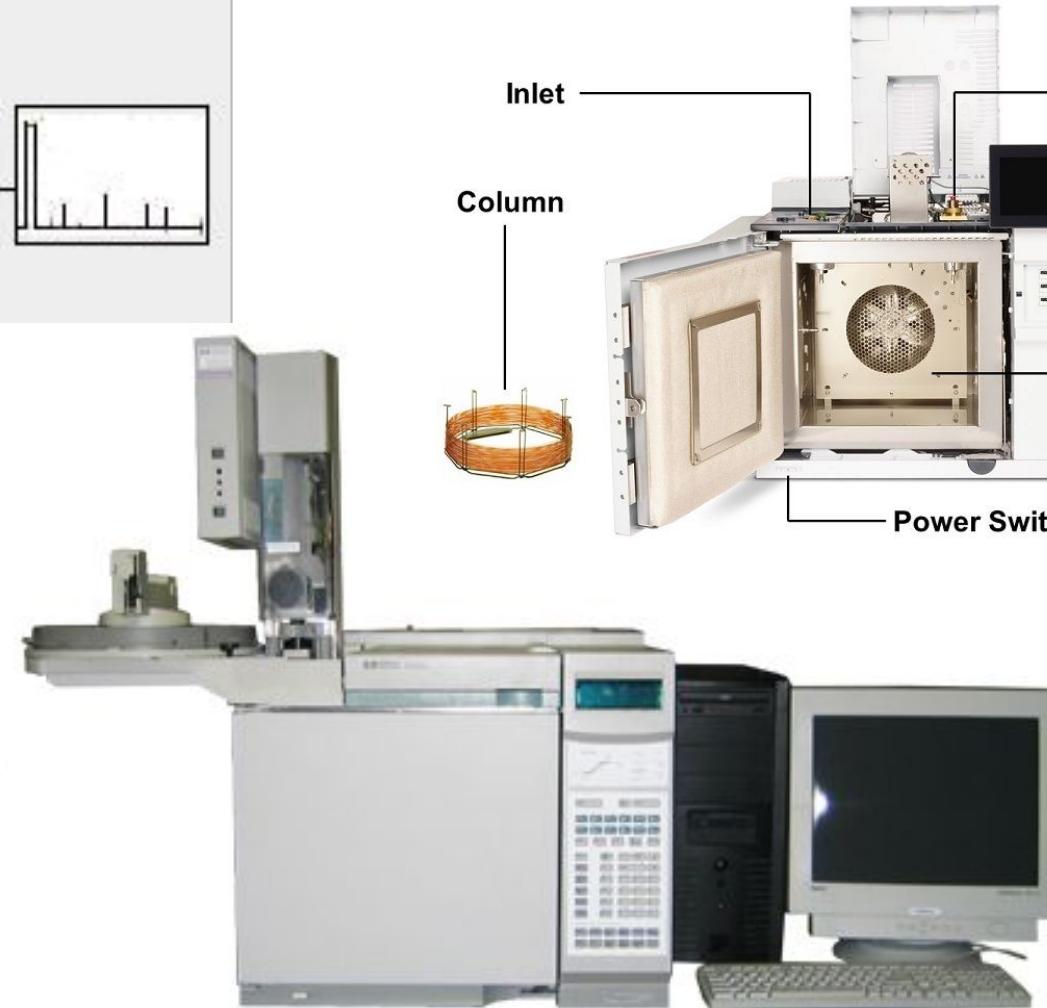
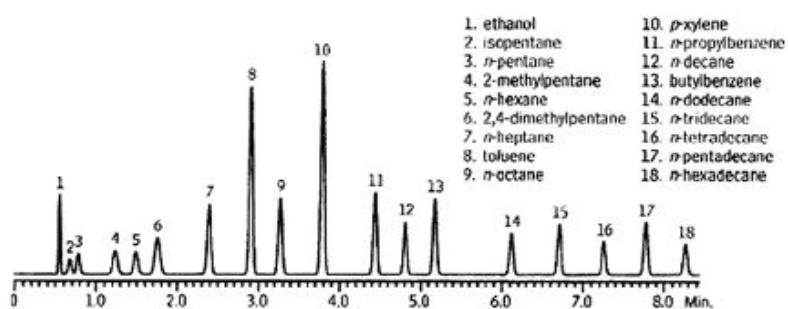
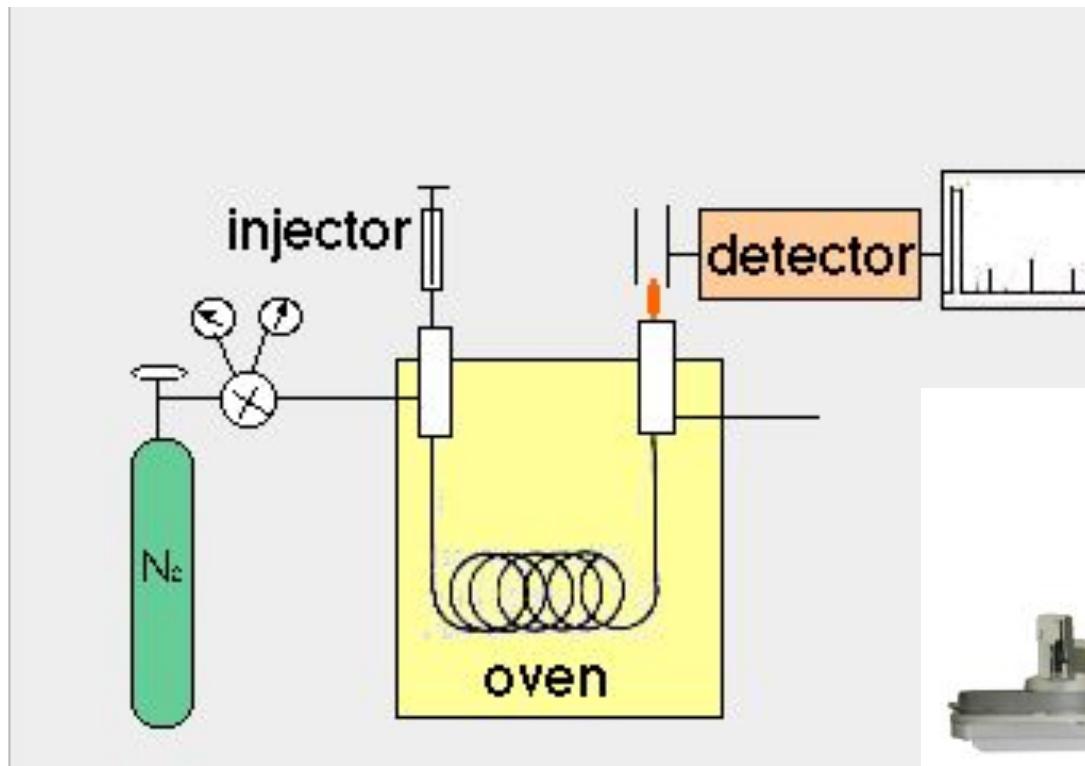
Gradient HPLC Schematic



HPLC of a Biological Mixture



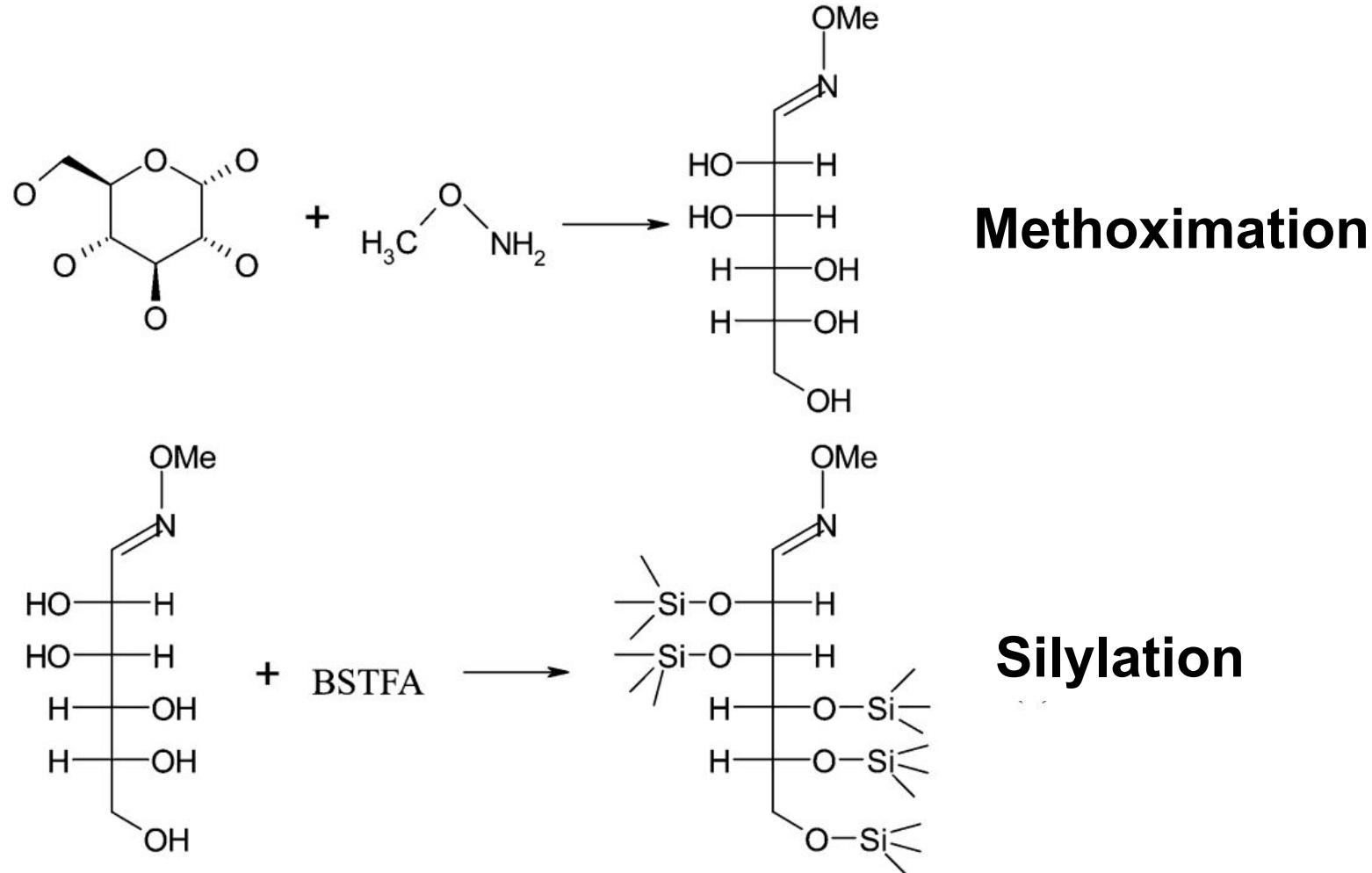
Gas Chromatography



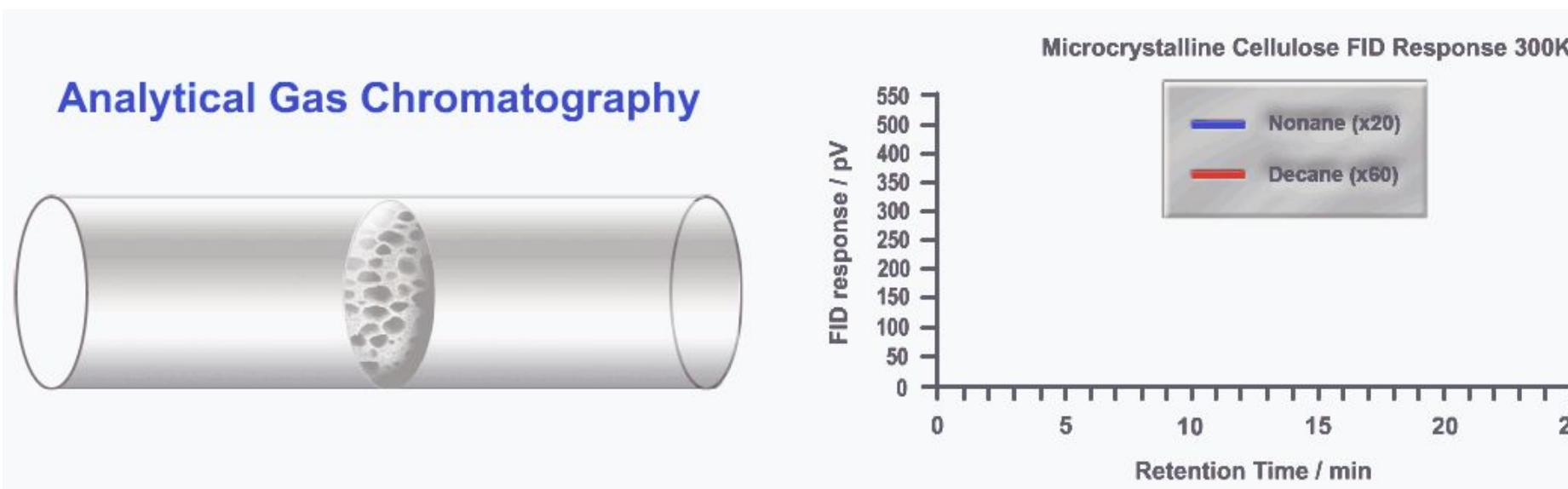
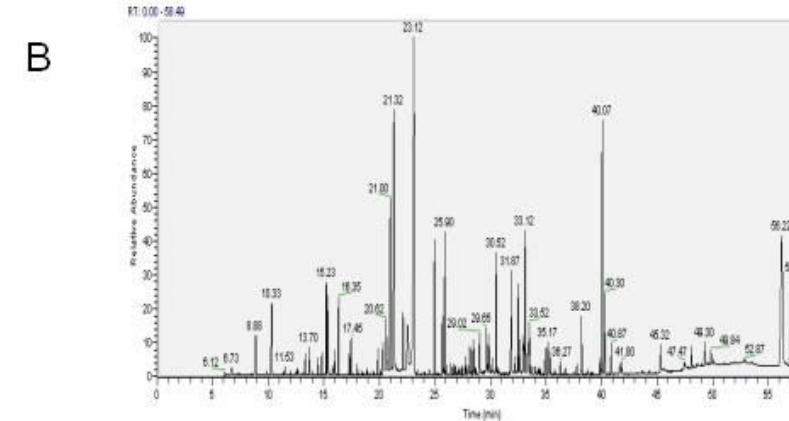
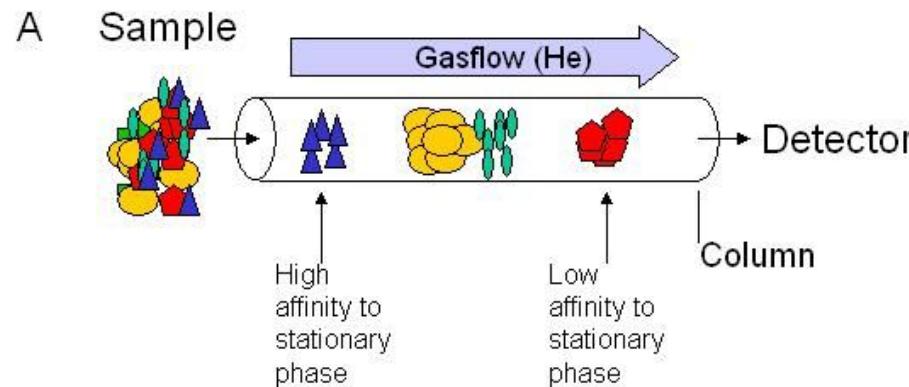
Gas Chromatography

- Involves a sample being vaporized to a gas and injected into a column
- Sample is transported through the column by an inert gas mobile phase
- Column has a liquid or polymer stationary phase that is adsorbed to the surface of a metal tube
- Columns are 1.5-10 metres in length and 2-4 mm in internal diameter
- Samples are usually derivatized with TMS (trimethylsilane) to make them volatile

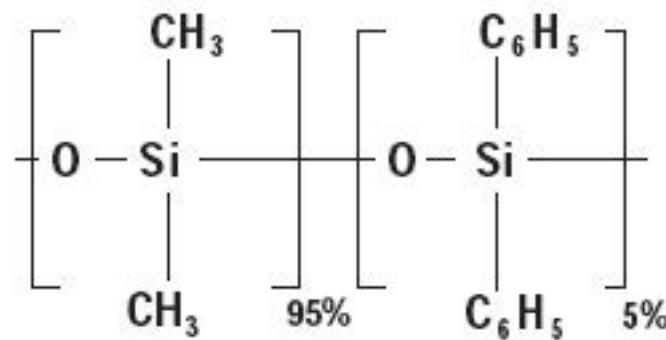
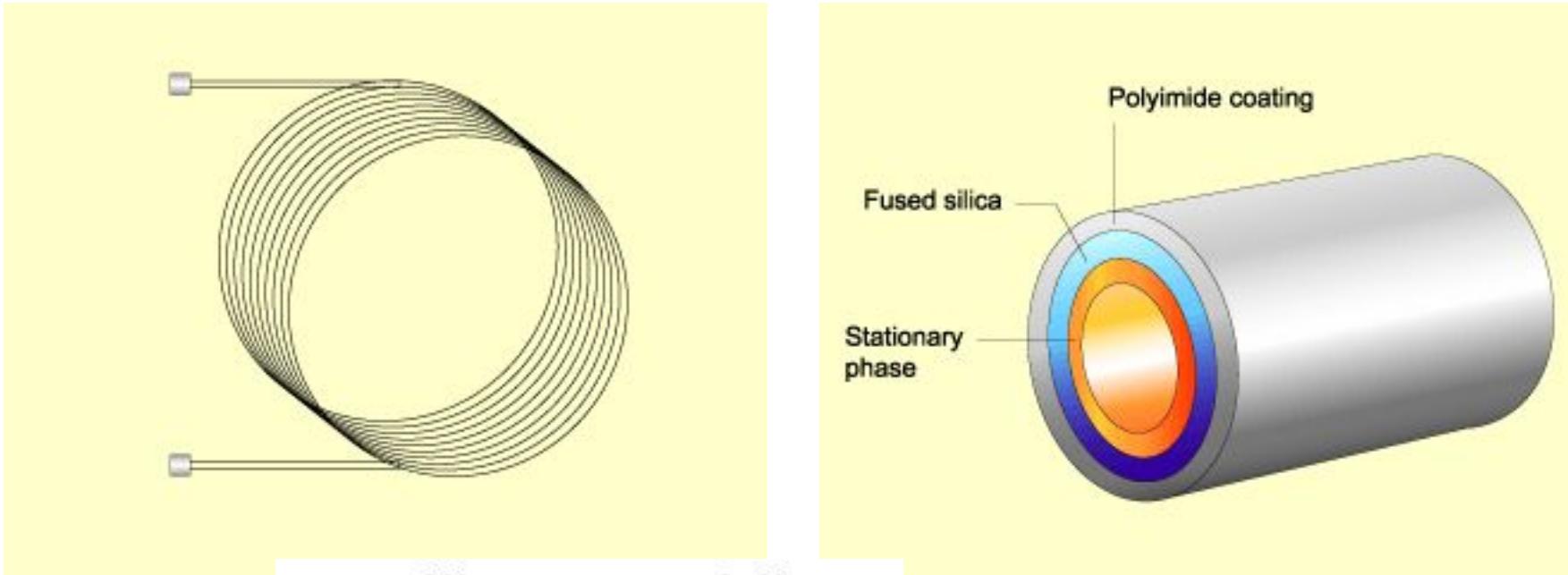
Derivatization



Gas Chromatography



GC-Columns

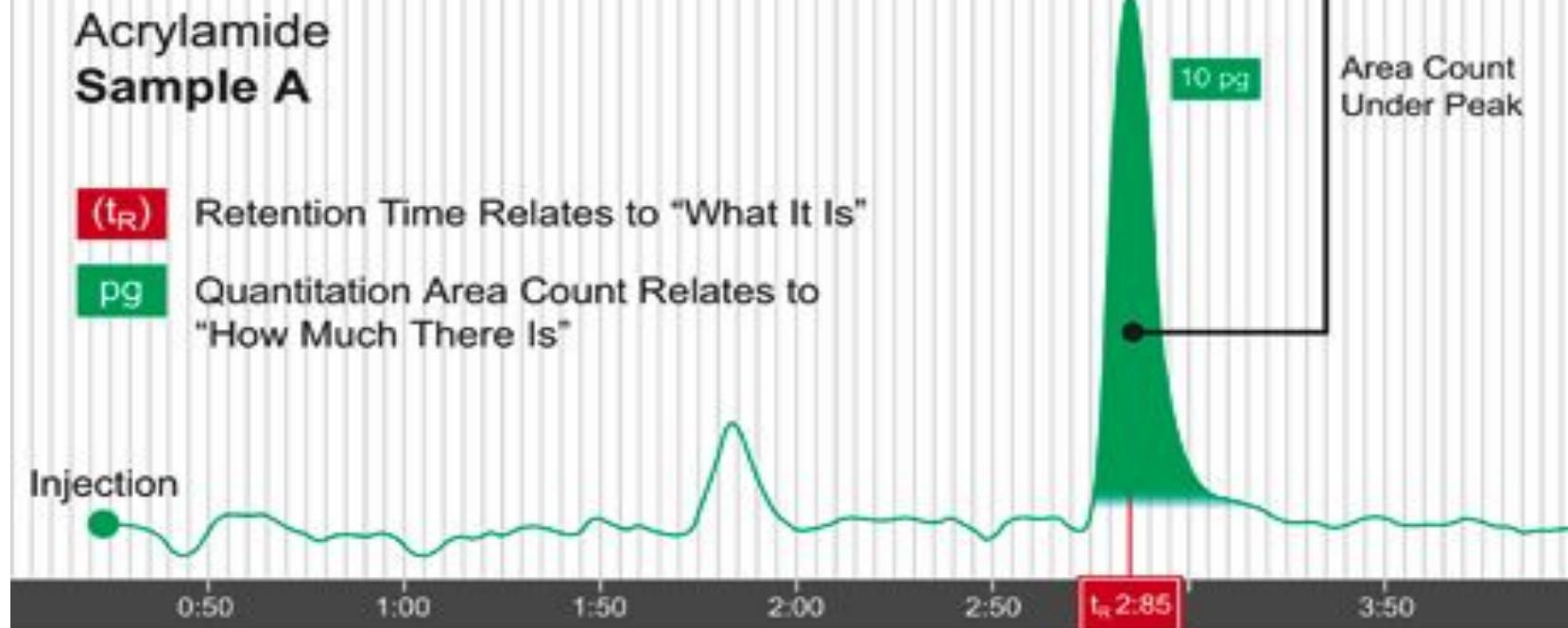
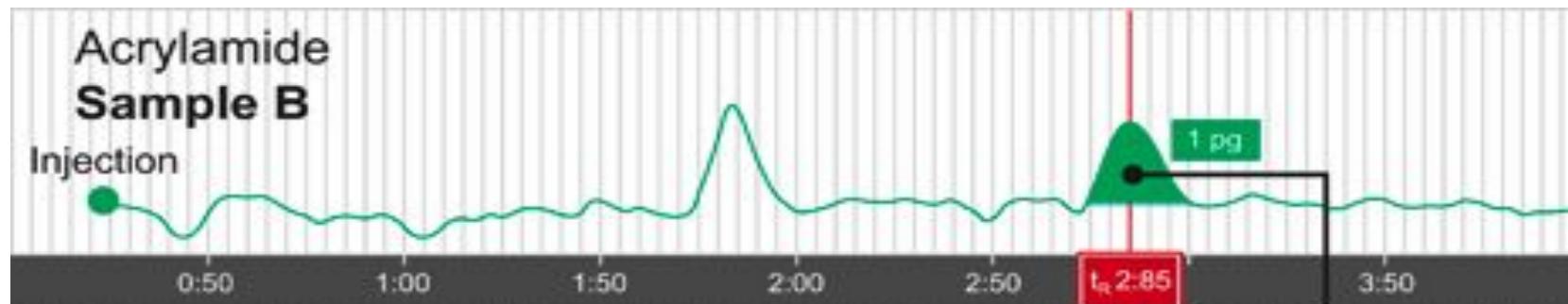


Polysiloxane

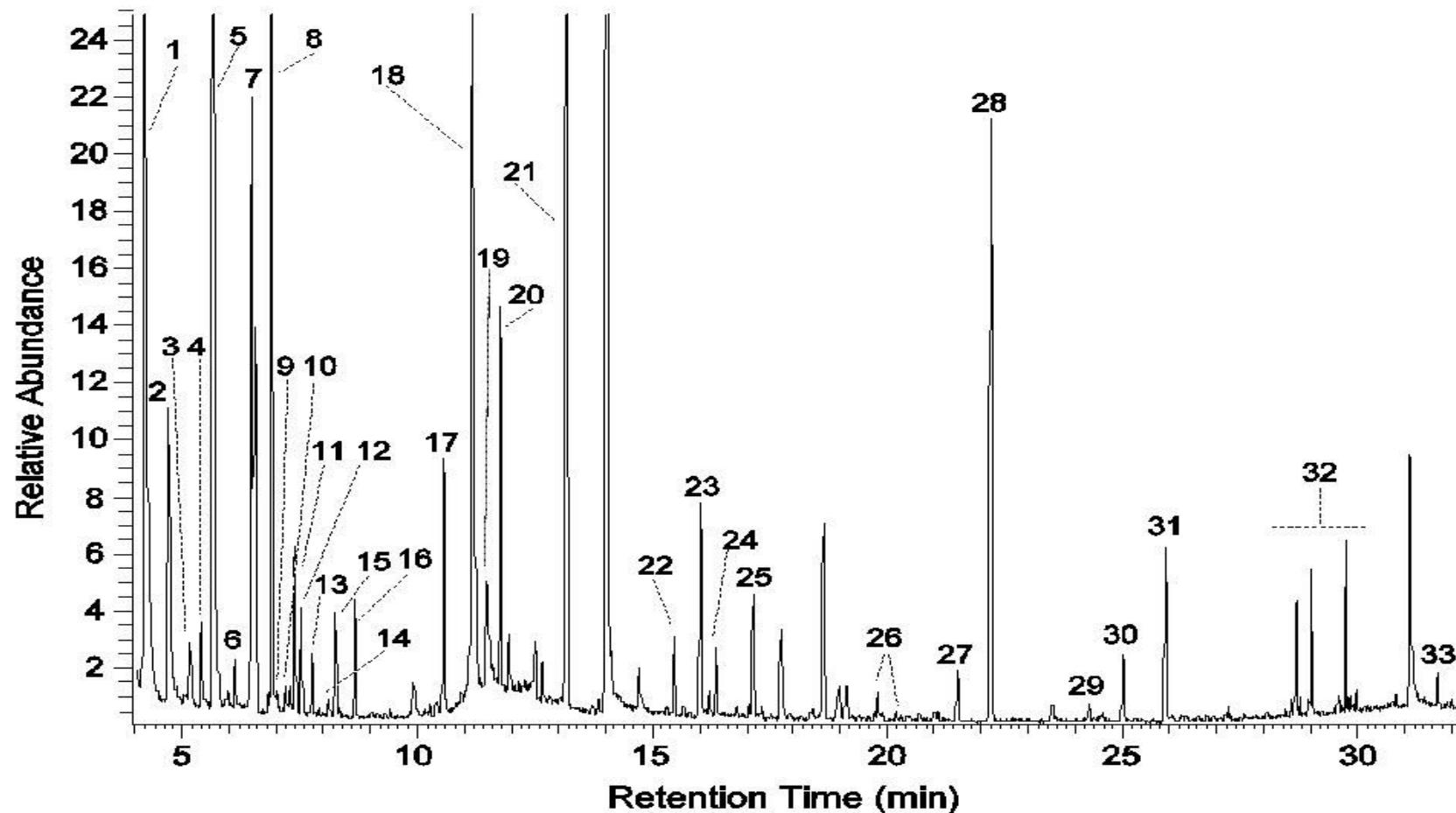
Retention Time/Retention Index

- **Retention time (RT) is the time taken by an analyte to pass through a column**
- **RT is affected by compound, column (dimensions and stationary phase), flow rate, pressure, carrier, temp.**
- **Comparing RT from a standard sample to an unknown allows compound ID**
- **Retention index (RI) is the retention time normalized to the retention times of adjacently eluting n-alkanes**

Compound Identification and Quantification

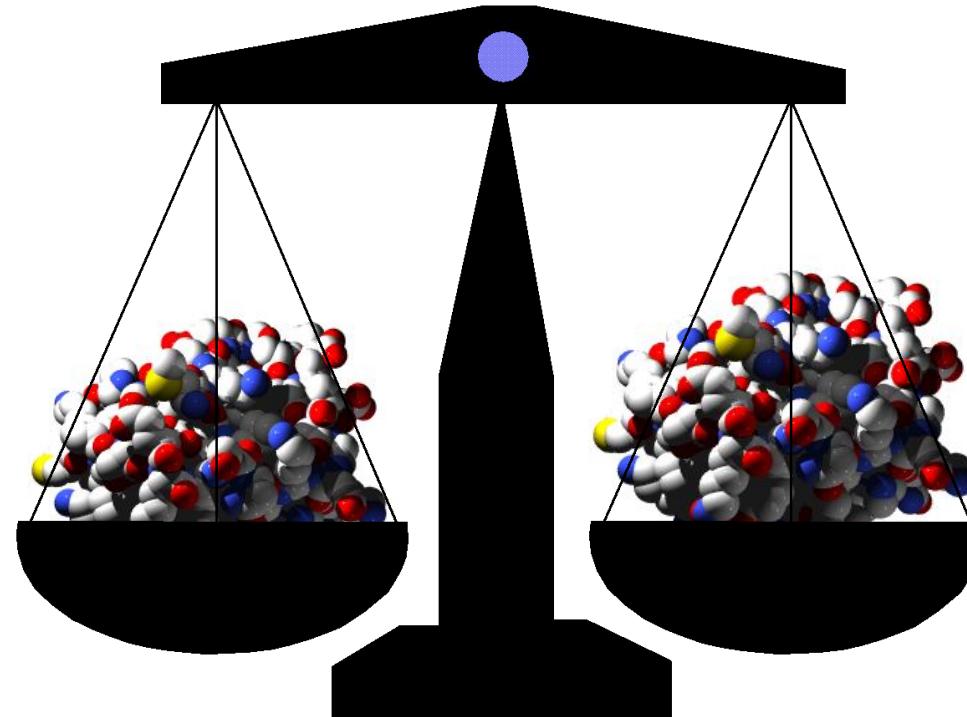


GC-MS Chromatogram of a Biological Mixture



Mass Spectrometry

- Analytical method to measure the molecular or atomic weight of samples



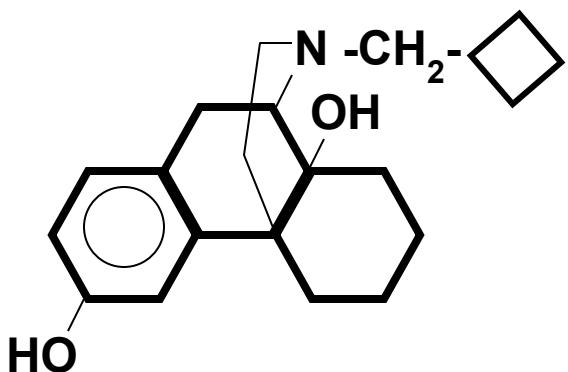
Typical Mass Spectrometer



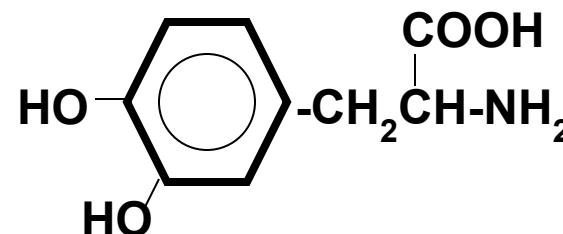
MS Principles

- Different compounds can be uniquely identified by their mass

Butorphanol



L-dopa



Ethanol



MW = 327.1

MW = 197.2

MW = 46.1

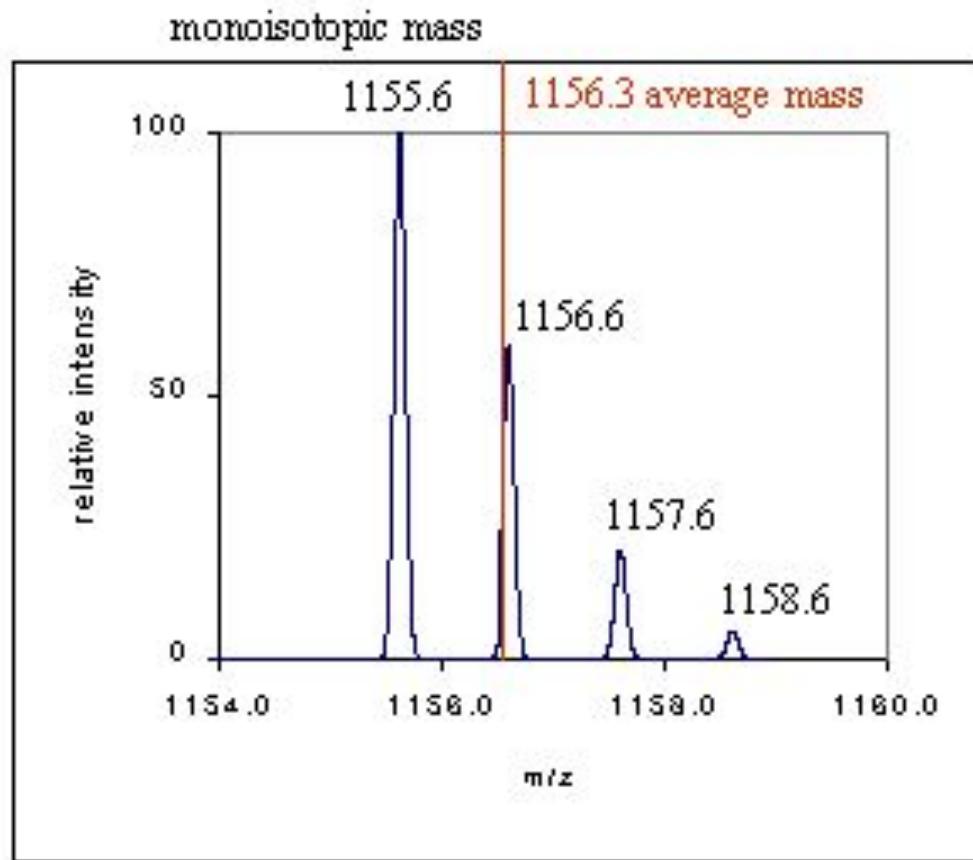
Mass Spectrometry

- For small organic molecules the MW can be determined to within 1 ppm or 0.0001% which is sufficiently accurate to confirm the molecular formula from mass alone
- For large biomolecules the MW can be routinely determined within an accuracy of 0.002% (i.e. within 1 Da for a 40 kD protein)
- Recall 1 dalton = 1 atomic mass unit (1 amu)

Different Types of MS

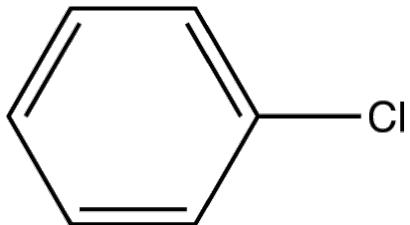
- **GC-MS - Gas Chromatography MS**
 - separates volatile compounds in gas column and ID's by mass
- **LC-MS - Liquid Chromatography MS**
 - separates delicate compounds in HPLC column and ID's by mass
- **MS-MS - Tandem Mass Spectrometry**
 - separates compound fragments by magnetic or electric fields and ID's by mass fragment patterns

Masses in MS



- Monoisotopic mass is the mass determined using the masses of the most abundant isotopes
- Average mass is the abundance weighted mass of all isotopic components

Isotopic Distributions



$^1\text{H} = 99.9\%$
 $^2\text{H} = 0.02\%$

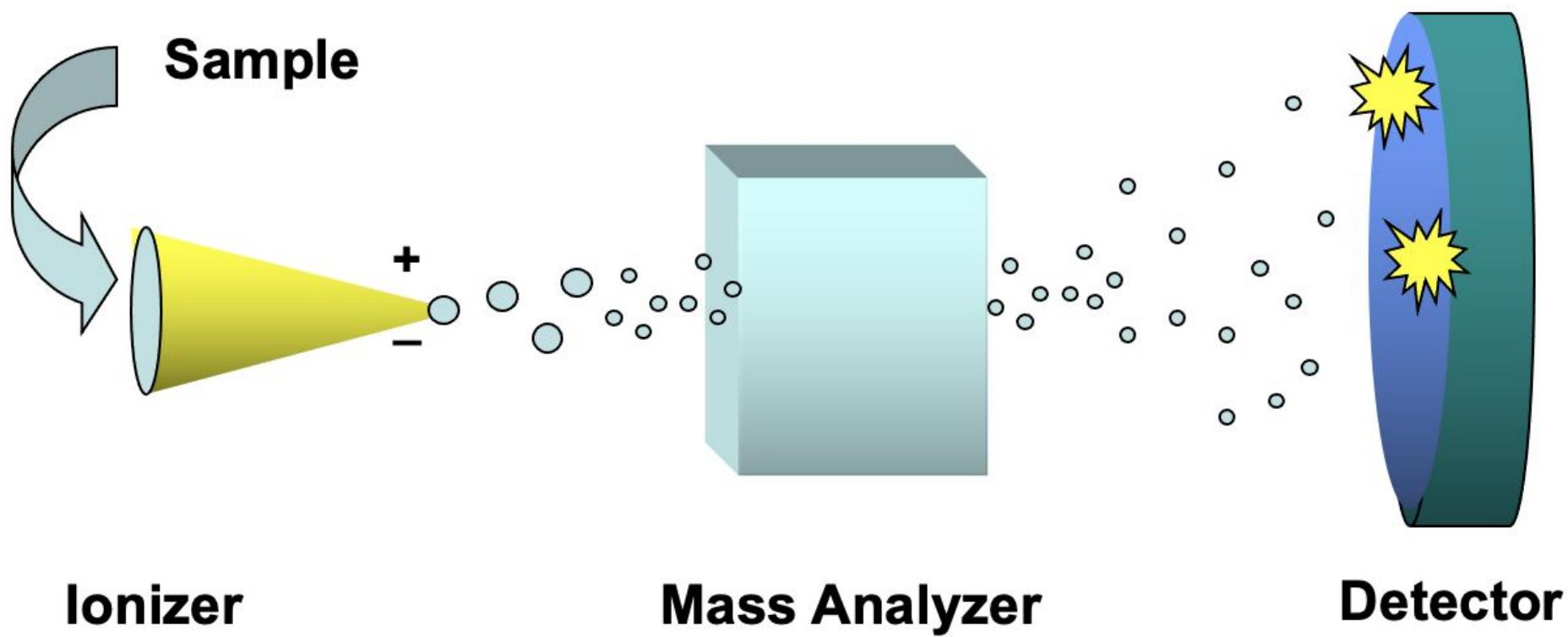
$^{12}\text{C} = 98.9\%$
 $^{13}\text{C} = 1.1\%$

$^{35}\text{Cl} = 68.1\%$
 $^{37}\text{Cl} = 31.9\%$

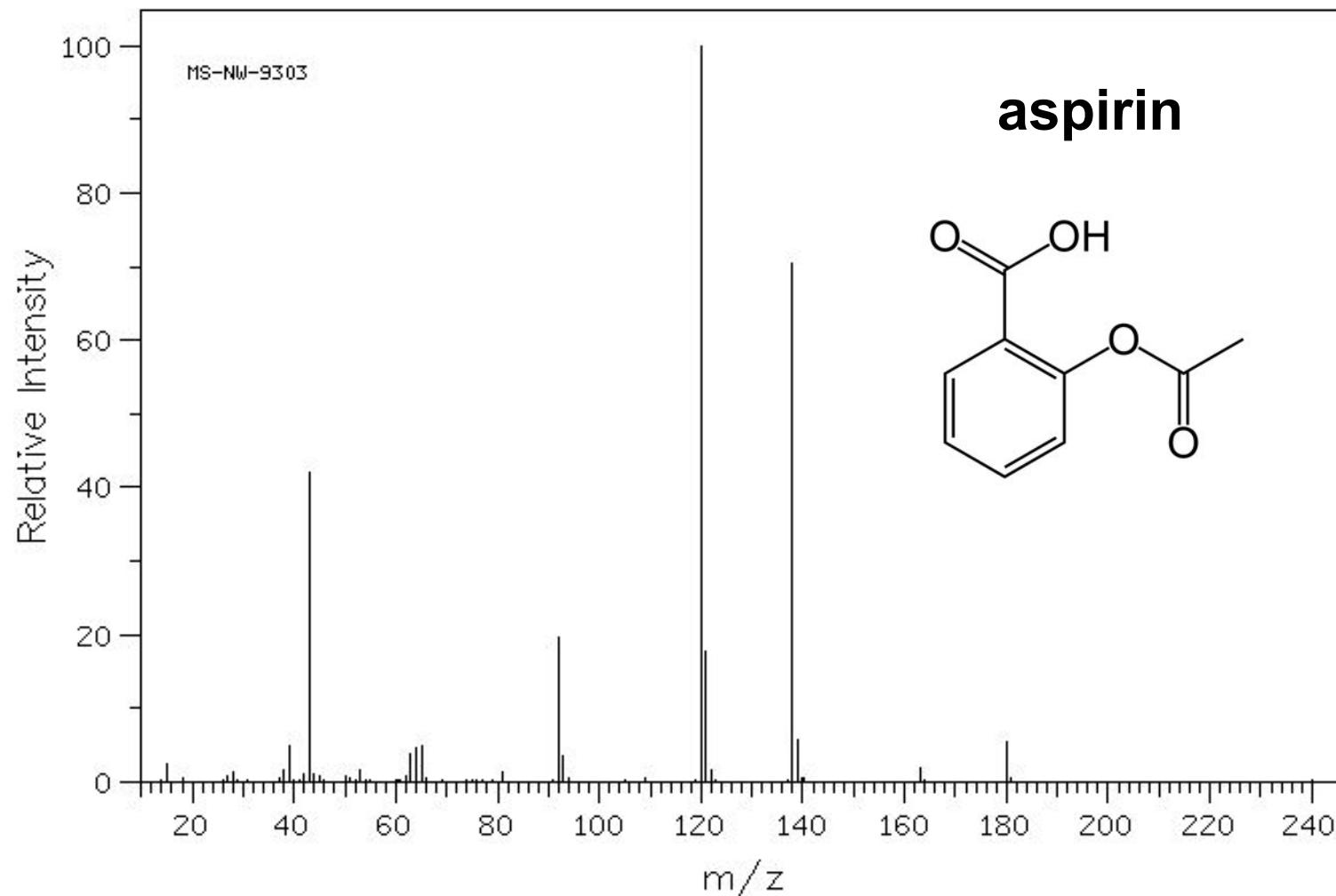
As an example, the molecule $\text{C}_6\text{H}_5\text{Cl}$ weighs 112.00797 Da. Its isotopic distribution is:

Mass	Fraction	Intensity	Potential Combinations of Atoms		
112.00797	0.7098995	100.00	$^{12}\text{C}_6 \ ^1\text{H}_5 \ ^{35}\text{Cl}$		
113.00797	0.0466010	6.56	$^{12}\text{C}_5 \ ^{13}\text{C} \ ^1\text{H}_5 \ ^{35}\text{Cl}$	$^{12}\text{C}_6 \ ^1\text{H}_4 \ ^2\text{H} \ ^{35}\text{Cl}$	
114.00797	0.2281708	32.14	$^{12}\text{C}_4 \ ^{13}\text{C}_2 \ ^1\text{H}_5$	$^{12}\text{C}_6 \ ^1\text{H}_3 \ ^2\text{H}_2 \ ^{35}\text{Cl}$	$^{12}\text{C}_6 \ ^1\text{H}_5 \ ^{37}\text{Cl}$
115.00797	0.0149130	2.10	...		
116.00797	0.0004092	0.06	...		
117.00797	0.0000060	0.00	...		

Mass Spec Principles



Typical Mass Spectrum



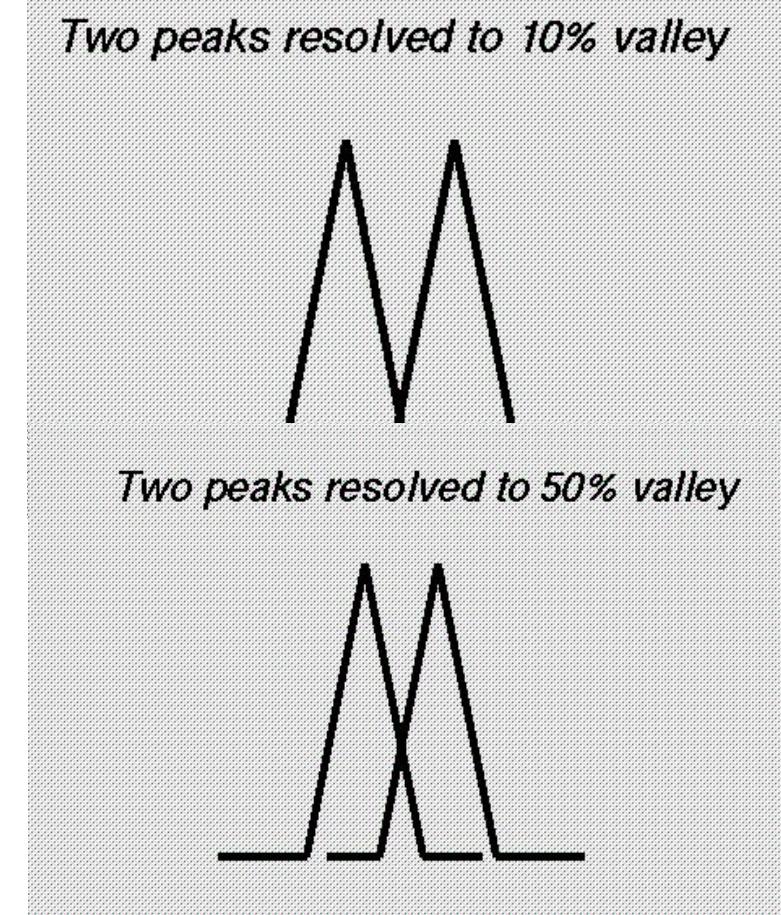
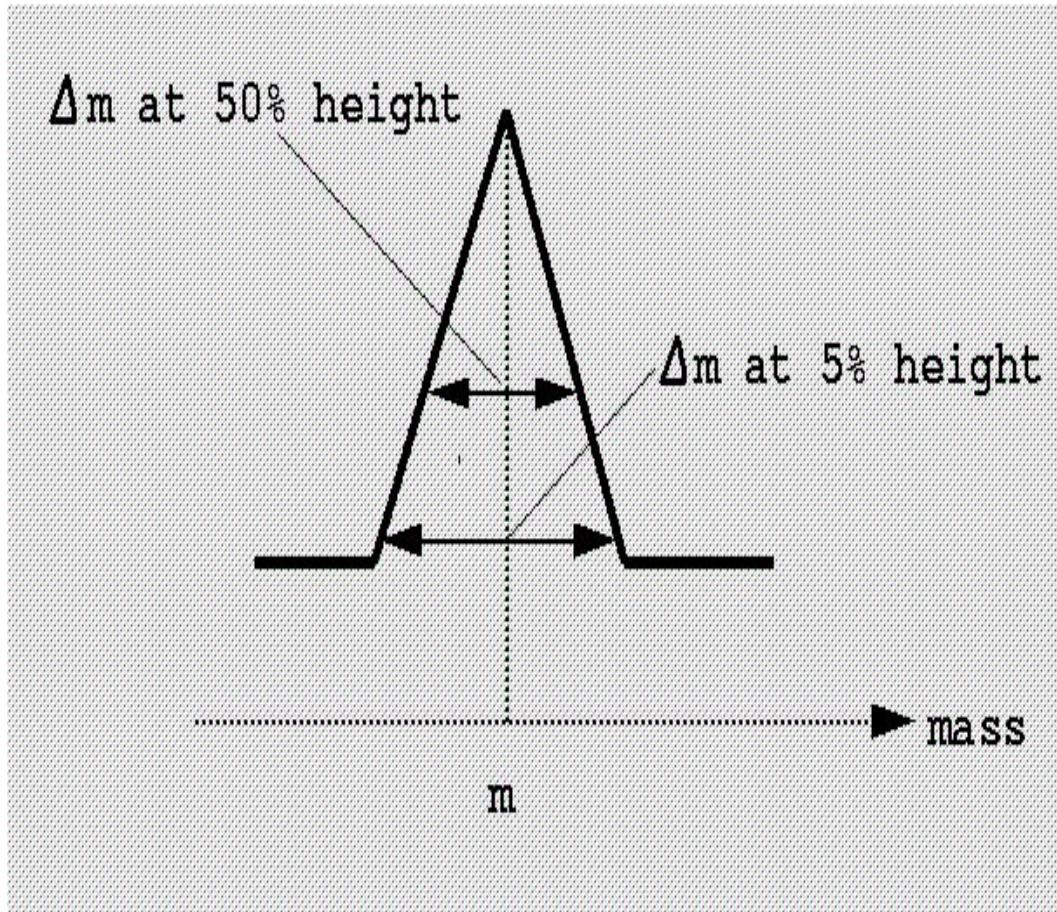
Typical Mass Spectrum

- Characterized by sharp, narrow peaks
- X-axis position indicates the m/z ratio of a given ion (for singly charged ions this corresponds to the mass of the ion)
- Height of peak indicates the relative abundance of a given ion (not reliable for quantitation)
- Peak intensity indicates the ion's ability to desorb or “fly” (some fly better than others) – this is called ionization efficiency

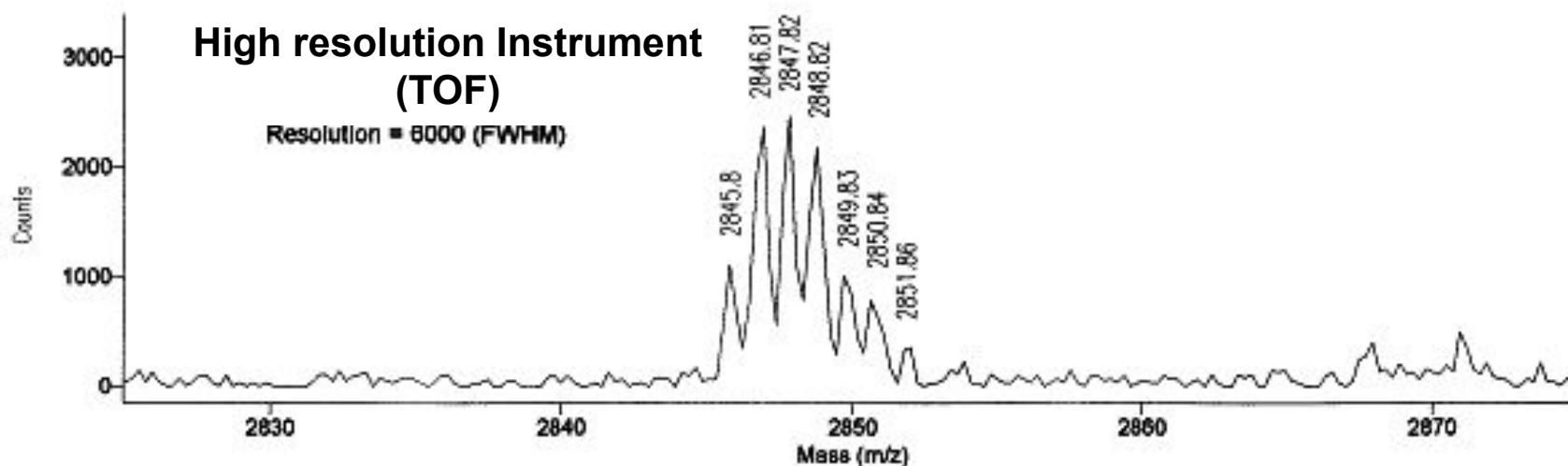
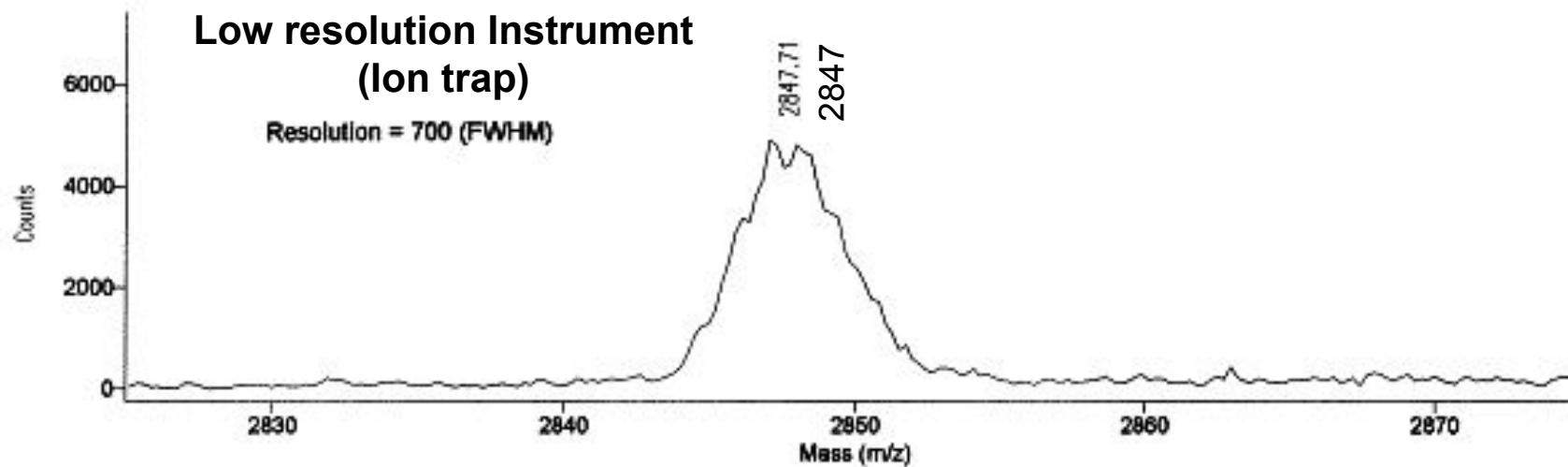
Resolution & Resolving Power

- Width of peak indicates the resolution of the MS instrument
- The better the resolution or resolving power, the better the instrument and the better the mass accuracy
- Resolving power is defined as: $\frac{M}{\Delta M}$
- M is the mass number of the observed mass (ΔM) is the difference between two masses that can be separated

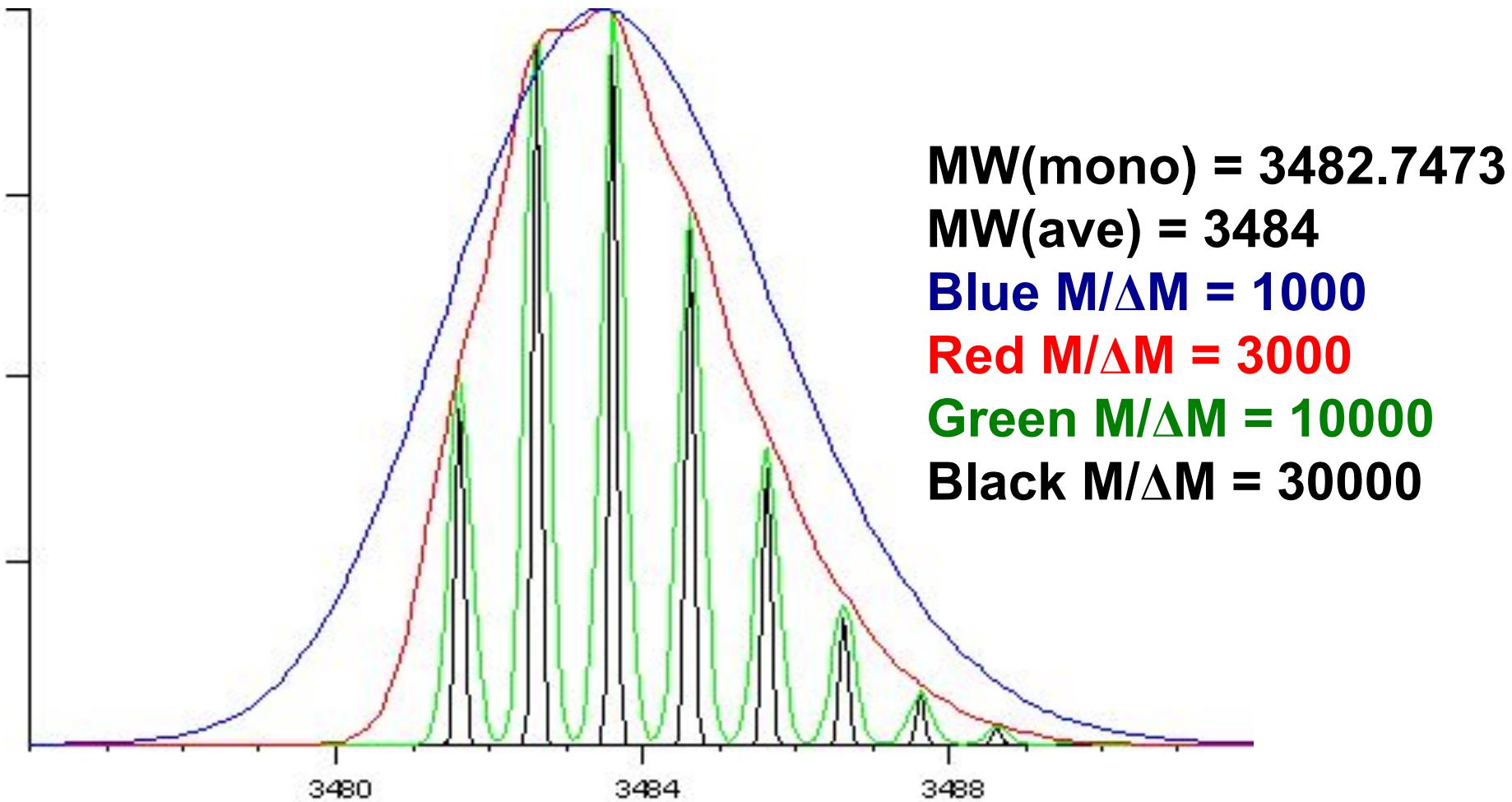
Resolution in MS



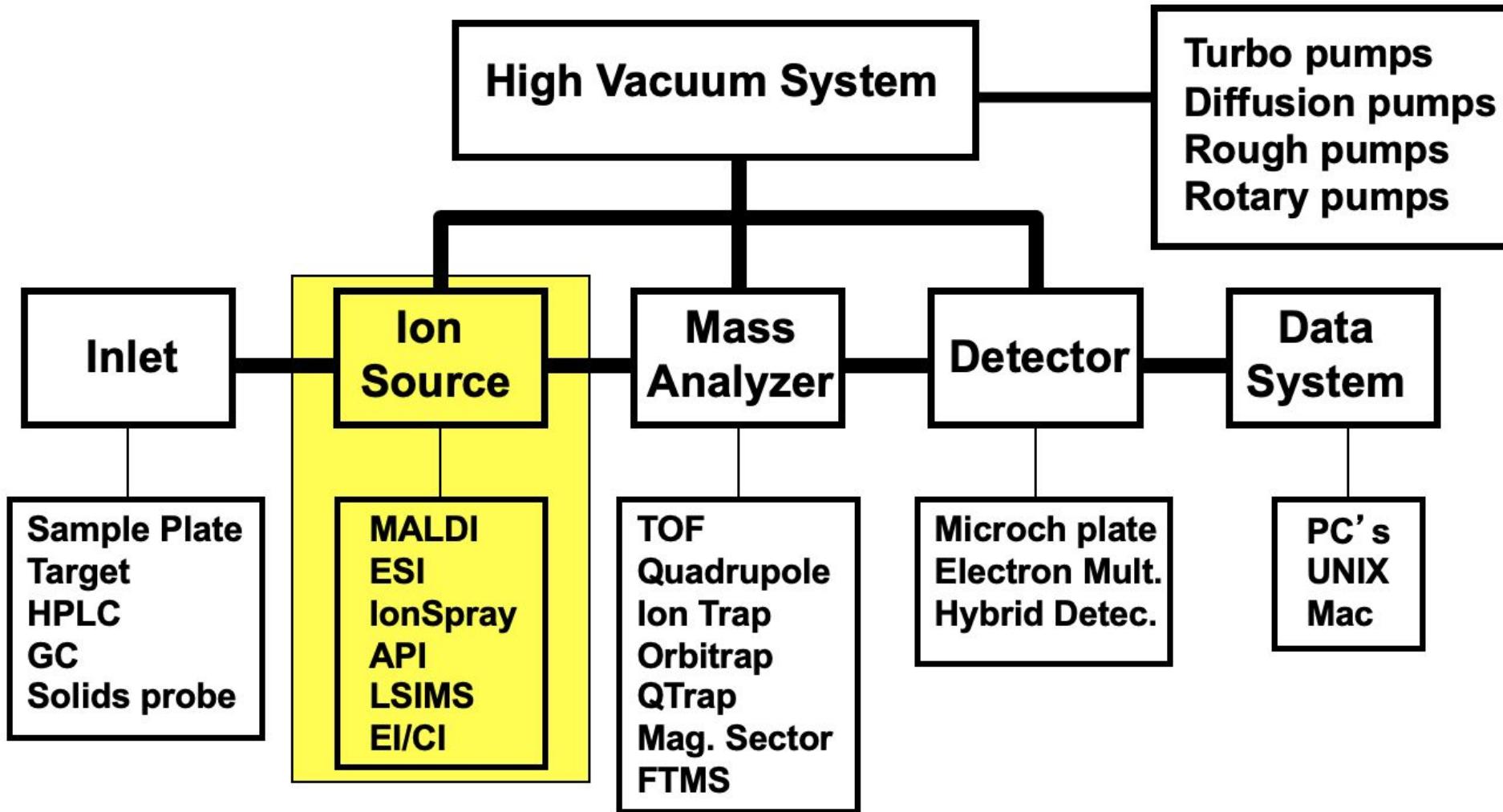
Resolution in MS



Resolution/Resolving Power



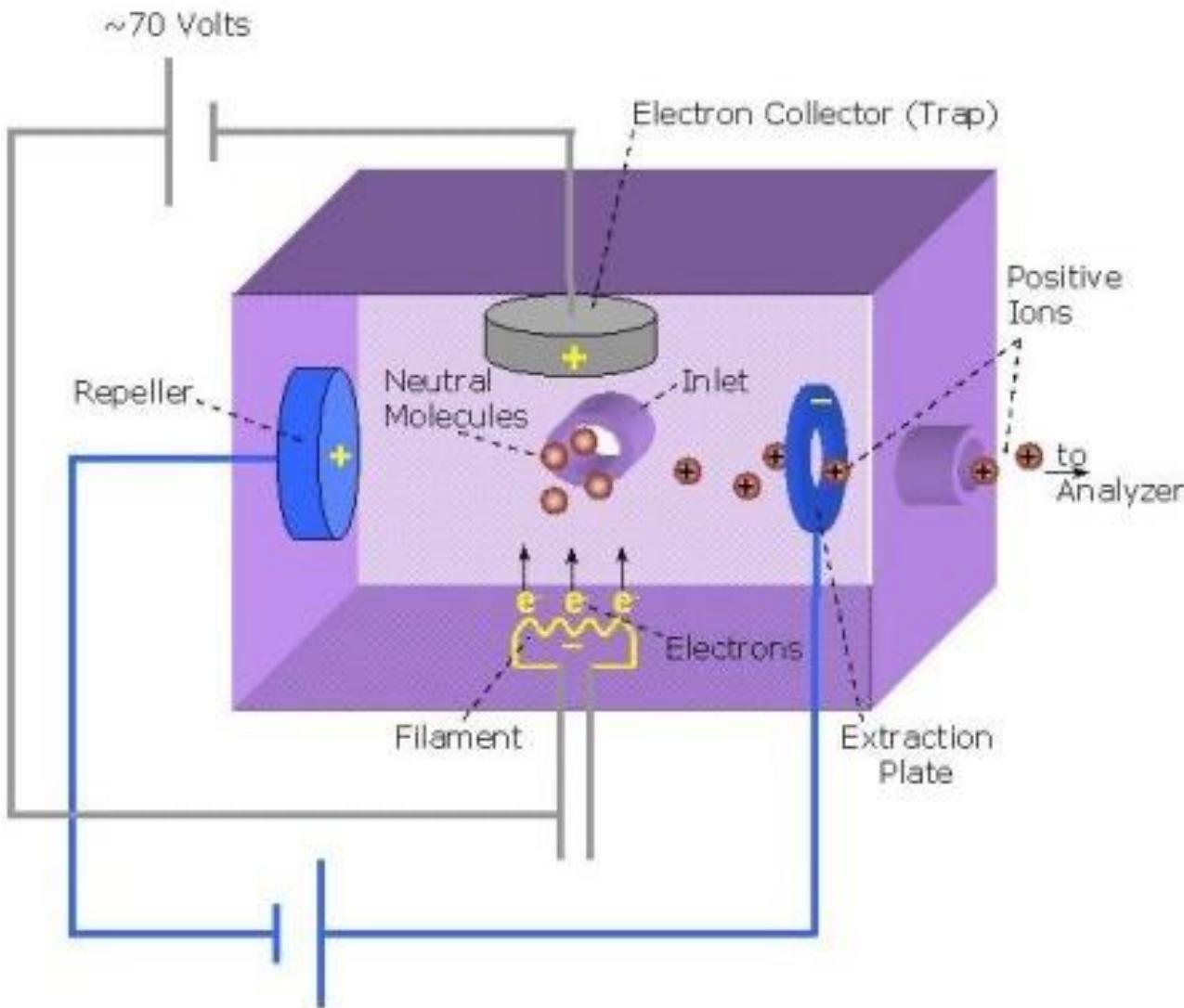
Mass Spectrometer Schematic



Different Ionization Methods

- Electron Ionization (EI - Hard method)
 - Small molecules, 1-1000 Daltons, structure
- Chemical Ionization (CI – Semi-hard)
 - Small molecules, 1-1000 Daltons, simple spectra
- Electrospray Ionization (ESI - Soft)
 - Small molecules, peptides, proteins, up to 200,000 Daltons
- Matrix Assisted Laser Desorption (MALDI-Soft)
 - Smallish molecules, peptides, proteins, DNA, up to 500 kD

Electron Impact Ionization Source

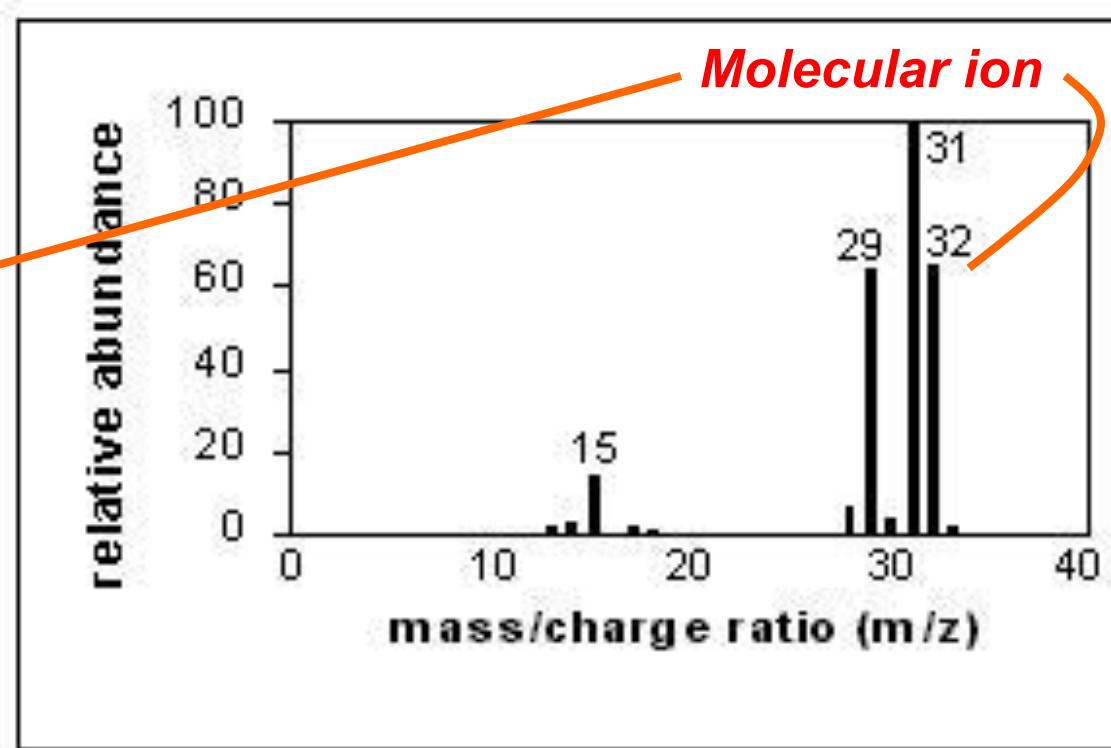


Electron Impact Ionization

- Sample introduced into the instrument by heating it until it evaporates
- Gas phase sample is bombarded with electrons coming from a rhenium or tungsten filament (electron energy = 70 eV)
- Molecule is “shattered” into fragments ($70\text{ eV} \gg 5\text{ eV}$ bonds)
- Fragments are sent to mass analyzer for detection
- EI is most commonly used in GC-MS

Electron Impact MS of CH₃OH

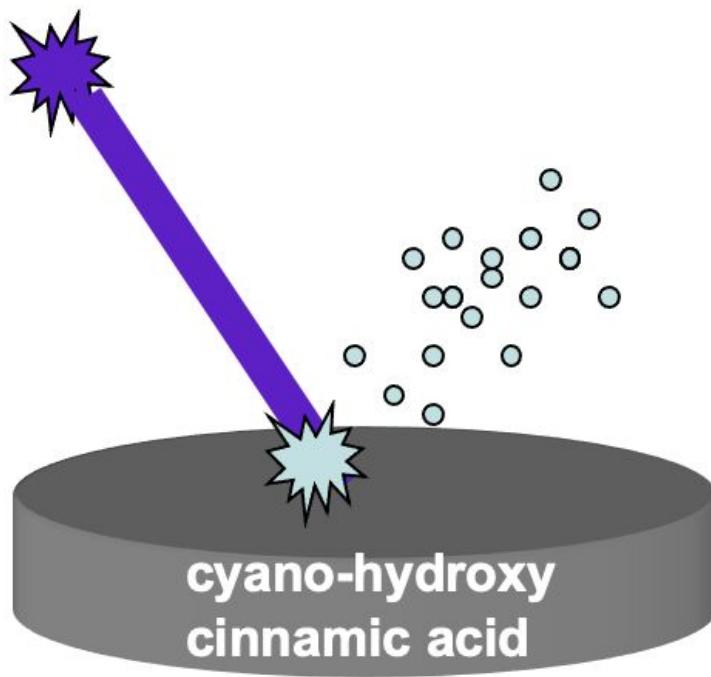
ions	m/z
CH ₃ OH ⁺	32
H ₂ C=OH ⁺	31
HC≡O ⁺	29
H ₃ C ⁺	15



EI Breaks up Molecules in Reasonably Predictable Ways

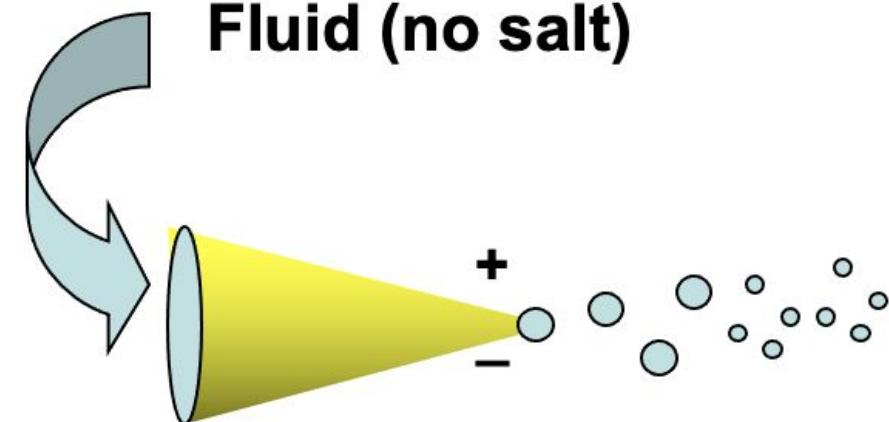
Soft Ionization Methods

337 nm UV laser



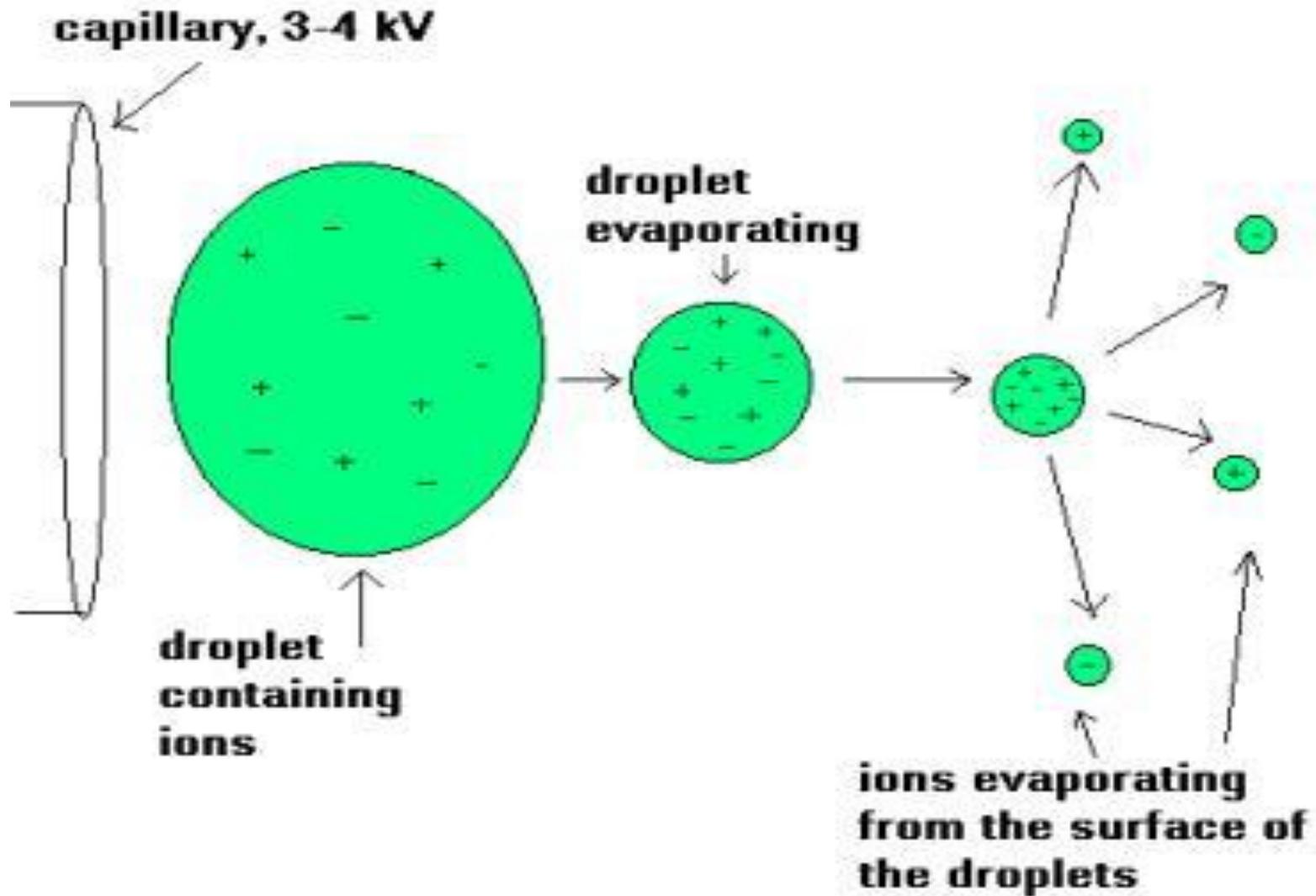
MALDI

Fluid (no salt)



ESI

Electrospray (Detail)



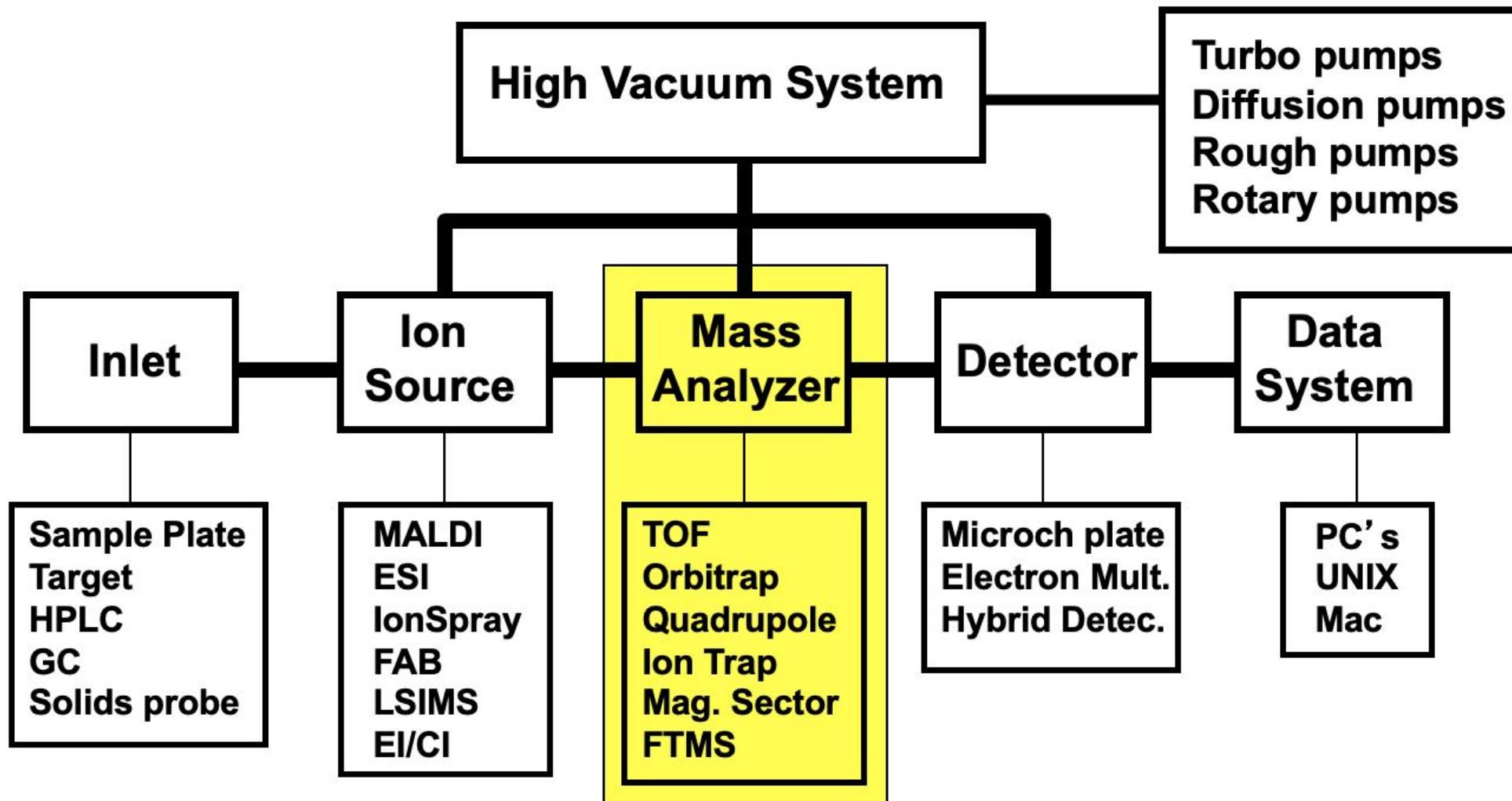
Electrospray ionization

- Sample dissolved in polar, volatile buffer (no salts) and pumped through a stainless steel capillary (70 - 150 μm) at a rate of 10-100 $\mu\text{L}/\text{min}$
- Strong voltage (3-4 kV) applied at tip along with flow of nebulizing gas causes the sample to “nebulize” or aerosolize
- Aerosol is directed through regions of higher vacuum until droplets evaporate to near atomic size (still carrying charges)

Electrospray ionization

- Can be modified to “nanospray” system with flow < 1 $\mu\text{L}/\text{min}$
- Very sensitive technique, requires less than a picomole of material
- Strongly affected by salts & detergents
- Positive ion mode measures $(M + H)^+$ (add formic acid to solvent)
- Negative ion mode measures $(M - H)^-$ (add ammonia to solvent)

Mass Spectrometer Schematic



Different Types of Mass Analyzers

- Magnetic Sector Analyzer (MSA)
 - High resolution, exact mass, original MA
- Quadrupole Analyzer (Q or Q*)
 - Low (1 amu) resolution, fast, cheap
- Time-of-Flight Analyzer (TOF)
 - No upper m/z limit, high throughput
- OrbiTrap Analyzer (Orbitrap)
 - Very high m/z limit, highest resolution
- Ion Cyclotron Resonance (FT-ICR)
 - Highest resolution, exact mass, costly

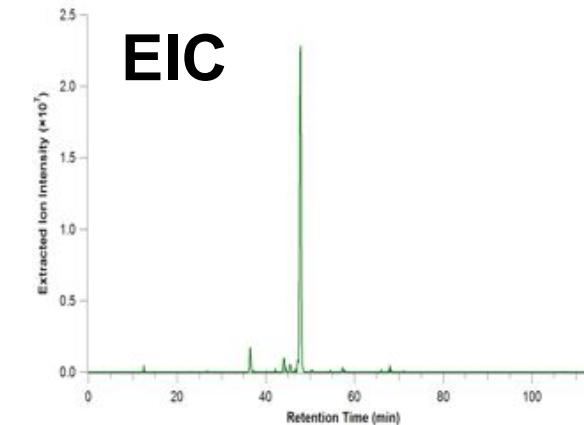
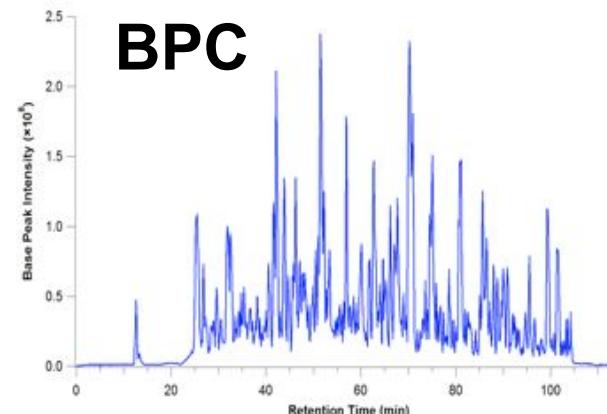
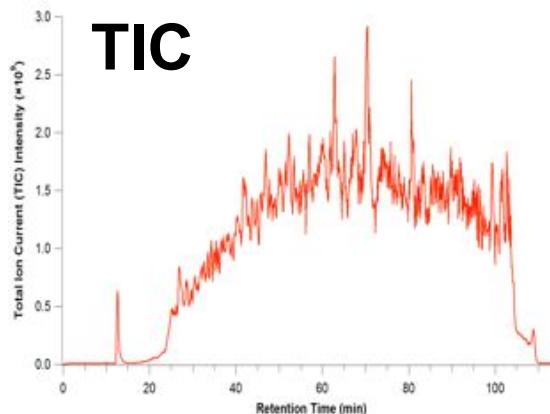
MS Mass Accuracy

<u>Type</u>	<u>Mass Accuracy</u>
FT-ICR-MS	0.1 - 1 ppm
Orbitrap	0.5 - 1 ppm
Magnetic Sector	1 - 2 ppm
TOF-MS	3 - 5 ppm
Q-TOF	3 - 5 ppm
Triple Quad	3 - 5 ppm
Linear IonTrap	50-200 ppm (10 ppm in Ultra-Zoom)

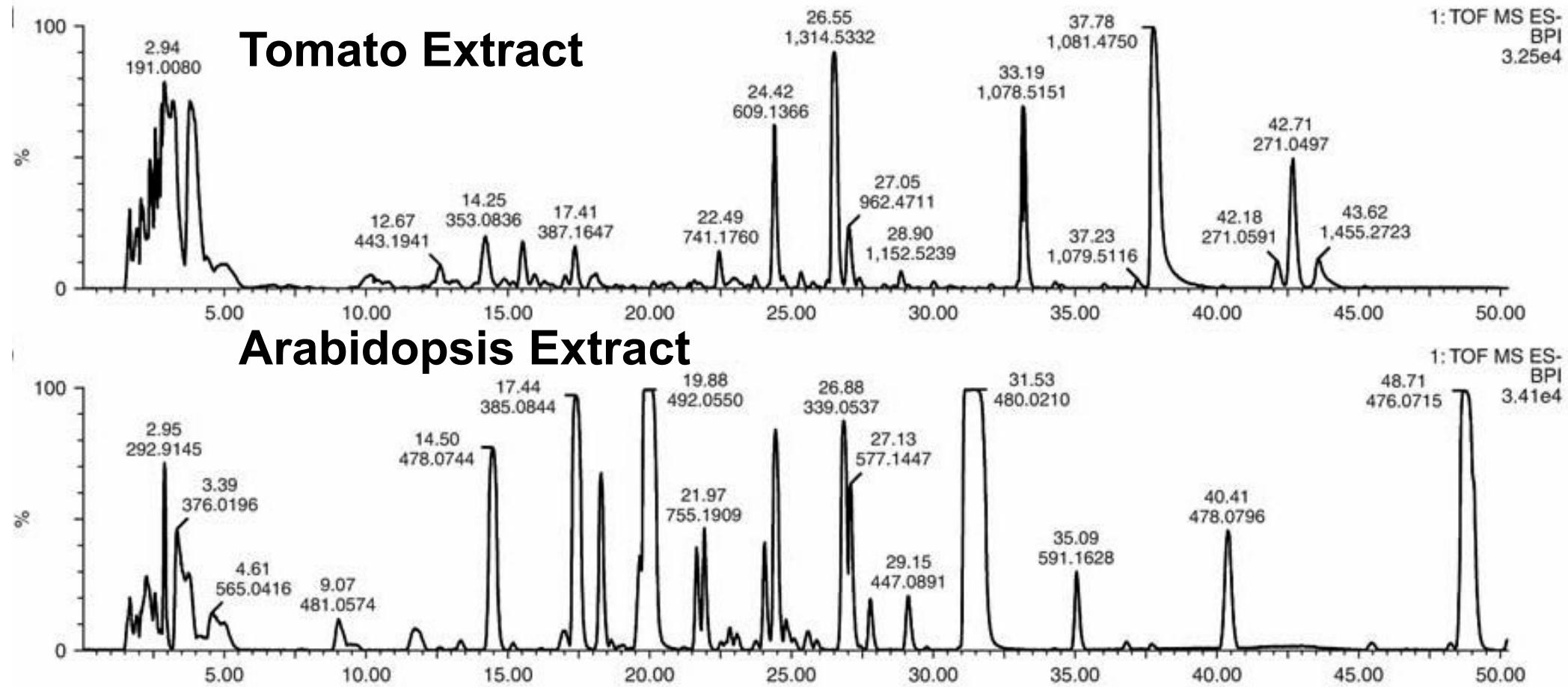
$$\text{ppm} = \left(\frac{m_{\text{exp}} - m_{\text{calc}}}{m_{\text{exp}}} \right) * 1E+6$$

Mass Chromatograms

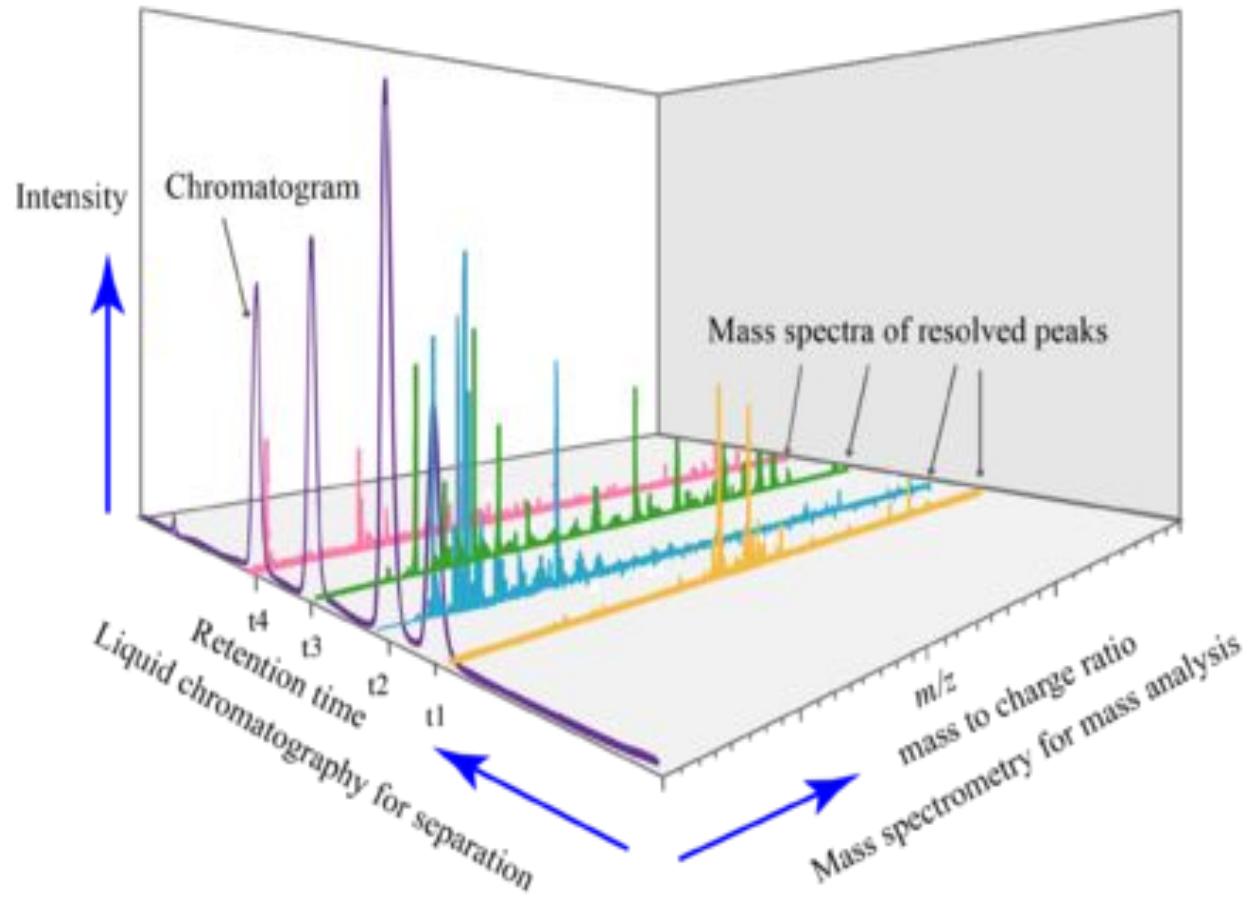
- Standard “output” from an LC-MS or GC-MS experiment
- X-axis is retention time, Y-axis is signal intensity
- **Total Ion Current (TIC)** chromatogram is summed intensity across the entire range of masses being detected at every point in the analysis
- **Base Peak chromatogram (BPC)** is like a TIC but displays only the most intense peak in each spectrum
- **Extracted Ion chromatogram (EIC)** contains one or more analytes extracted from the TIC or BPC



Base Peak Chromatograms of Biological Mixtures



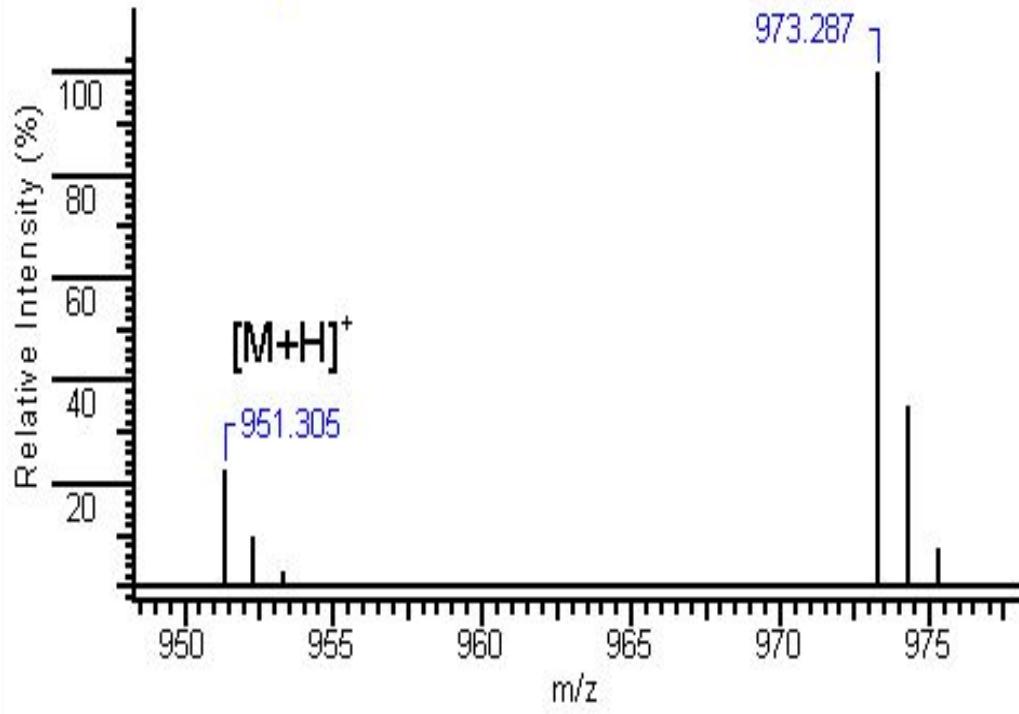
Metabolite ID by LC-MS



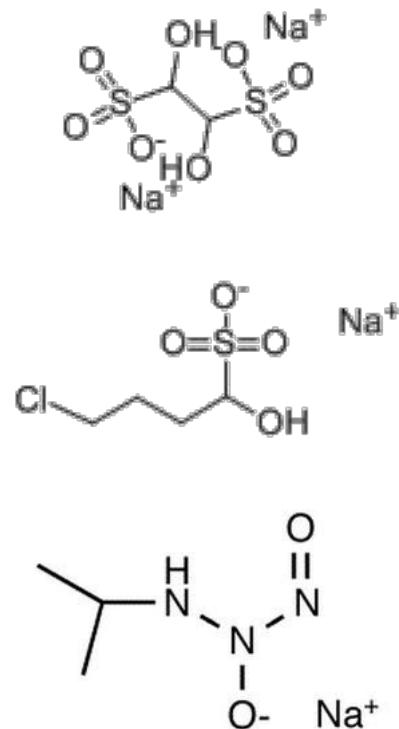
- **Features: 3D LC-MS signals (m/z , RT, intensity) produced by the same molecular ion**
- **Single compound can give rise to # of mass signals (salt adducts, neutral loss species and multiply charged species)**
- **Up to 80% of LC-MS signals arise from these “noise” sources**
- **Key challenge is to distinguish adducts or multiply charged species from parent ions**

Adduct Formation

ESI+ mass spectrum



Effect on ESI Mass Spectrum



Sample Na Adducts

Table 1. Monoisotopic exact masses of molecular ion adducts often observed in ESI mass spectra

Ion name	Ion mass	Charge	Mult.	Mass	Your M here:	Your M+X or M-X
1. Positive ion mode						
M+3H	M/3 + 1.007276	3+	0.33	1.007276	285.450906	291.099391
M+2H+Na	M/3 + 8.334590	3+	0.33	8.334590	292.778220	283.772077
M+H+2Na	M/3 + 15.7661904	3+	0.33	15.766190	300.209820	276.340476
M+3Na	M/3 + 22.989218	3+	0.33	22.989218	307.432848	269.117449
M+2H	M/2 + 1.007276	2+	0.50	1.007276	427.672721	437.152724
M+H+NH4	M/2 + 9.520550	2+	0.50	9.520550	436.185995	428.639450
M+H+Na	M/2 + 11.998247	2+	0.50	11.998247	438.663692	426.161753
M+H+K	M/2 + 19.985217	2+	0.50	19.985217	446.650662	418.174783
M+ACN+2H	M/2 + 21.520550	2+	0.50	21.520550	448.185995	416.639450
M+2Na	M/2 + 22.989218	2+	0.50	22.989218	449.654663	415.170782
M+2ACN+2H	M/2 + 42.033823	2+	0.50	42.033823	468.699268	396.126177
M+3ACN+2H	M/2 + 62.547097	2+	0.50	62.547097	489.212542	375.612903
M+H	M + 1.007276	1+	1.00	1.007276	854.338166	875.312724
M+NH4	M + 18.033823	1+	1.00	18.033823	871.364713	858.286177
M+Na	M + 22.989218	1+	1.00	22.989218	876.320108	853.330782
M+CH3OH+H	M + 33.033489	1+	1.00	33.033489	886.364379	843.286511
M+K	M + 38.963158	1+	1.00	38.963158	892.294048	837.356842
M+ACN+H	M + 42.033823	1+	1.00	42.033823	895.364713	834.286177
M+2Na-H	M + 44.971160	1+	1.00	44.971160	898.302050	831.348840
M+IsoProp+H	M + 61.06534	1+	1.00	61.065340	914.396230	815.254660
M+ACN+Na	M + 64.015765	1+	1.00	64.015765	917.346655	812.304235
M+2K-H	M + 76.919040	1+	1.00	76.919040	930.249930	799.400960
M+DMSO+H	M + 79.02122	1+	1.00	79.021220	932.352110	797.298780
M+2ACN+H	M + 83.060370	1+	1.00	83.060370	936.391260	793.259630
M+IsoProp+Na+H	M + 84.05511	1+	1.00	84.055110	937.386000	792.264890
2M+H	2M + 1.007276	1+	2.00	1.007276	1707.669056	1751.632724
2M+NH4	2M + 18.033823	1+	2.00	18.033823	1724.695603	1734.606177
2M+Na	2M + 22.989218	1+	2.00	22.989218	1729.650998	1729.650782
2M+3H2O+2H	2M + 28.02312	2+	2.00	28.023120	1734.684900	1724.616880
2M+K	2M + 38.963158	1+	2.00	38.963158	1745.624938	1713.676842

Resolving LC-MS Complications

- Identify, remove (or consolidate) adducts and multiply charged species
- Identify, remove (or consolidate) fragments (neutral losses, breakdown products, rearrangements)
- Identify, remove (or consolidate) isotope peaks
- Remove noise peaks (from sample blanks or peaks that do not appear in >2/3 technical replicates or peaks that do not show dilution trends in 4 dilution replicates)

Feature Simplification in LC-MS

- Raw +ve mode spectrum • 15,000 features
- Remove adducts • 12,000 features
- Remove multiple charges • 10,000 features
- Remove neutral losses • 8,000 features
- Remove isotope peaks • 3,000 features
- Remove noise peaks • 2,500 features
- Final spectrum • 2,500 M+H peaks
- Repeat for –ve mode • 1,500 M-H peaks

These identified peaks can then be used to determine molecular formulas

Molecular Formula Generators

- Formula generators are used to create molecular formulae from very accurate m/z obtained by QTOF, FT-MS or OrbiTrap
- Assist in compound ID by LC-MS (formula is more restrictive than MW)
- Input typically requires:
 - Accurate isotopic mass (with or without adduct)
 - Error in ppm or mDa (milliDaltons)

The screenshot shows the SmartFormula Manually software interface. The input fields at the top are set to "Lower formula: CH", "Upper formula: C₄₀H₇₀O₅₀N₁₀S₁₀P₅Cl₁₀F₈I₅", and "Measured m/z: 525.0808". The tolerance is set to 3 ppm and the charge to 1. The results table below lists several molecular formulas along with their scores, m/z values, and other parameters like rdb and e⁻ Conf.

Meas. m/z	#	Ion Formula	Score	m/z	err [ppm]	mSigma	rdb	e⁻ Conf	N-Rule	Adduct
525.0808	1	C ₂₄ H ₁₅ F ₃ N ₅ O ₄ P	100.00	525.0808	0.0	11.8	19.5	odd	ok	M+H
525.0808	2	C ₂₁ H ₁₈ FN ₉ OP ₂ S	79.55	525.0809	0.1	21.9	18.5	odd	ok	M+H
525.0808	3	C ₂₃ H ₂₇ F ₂ N ₂ P ₄ S	80.63	525.0808	-0.0	22.1	13.0	even	ok	M+H
525.0808	4	C ₂₇ H ₁₈ F ₄ N ₂ OPS	80.55	525.0808	-0.0	22.3	19.0	even	ok	M+H
525.0808	5	C ₂₂ H ₂₄ FN ₂ O ₆ P ₂ S	67.94	525.0809	0.1	28.7	13.0	even	ok	M+H
525.0808	6	C ₁₉ H ₂₁ N ₅ O ₉ P ₂	56.59	525.0809	0.2	35.8	13.5	odd	ok	M+H
525.0808	7	C ₂₁ H ₁₇ F ₇ N ₂ O ₄ P	54.84	525.0809	0.1	37.5	12.0	even	ok	M+H

Below the table are several checkboxes and settings:
- Automatically locate monoisotopic peak (unchecked)
- Maximum number of formulae: 20
- Check rings plus double bonds (checked)
- Minimum: -0.5, Maximum: 20
- Electron configuration: both
- Filter H/C element ratio (checked)
- Minimum H/C: 0, Maximum H/C: 3
- Estimate carbon number (checked)
- Generate immediately (checked)

Metabolite ID by LC-MS

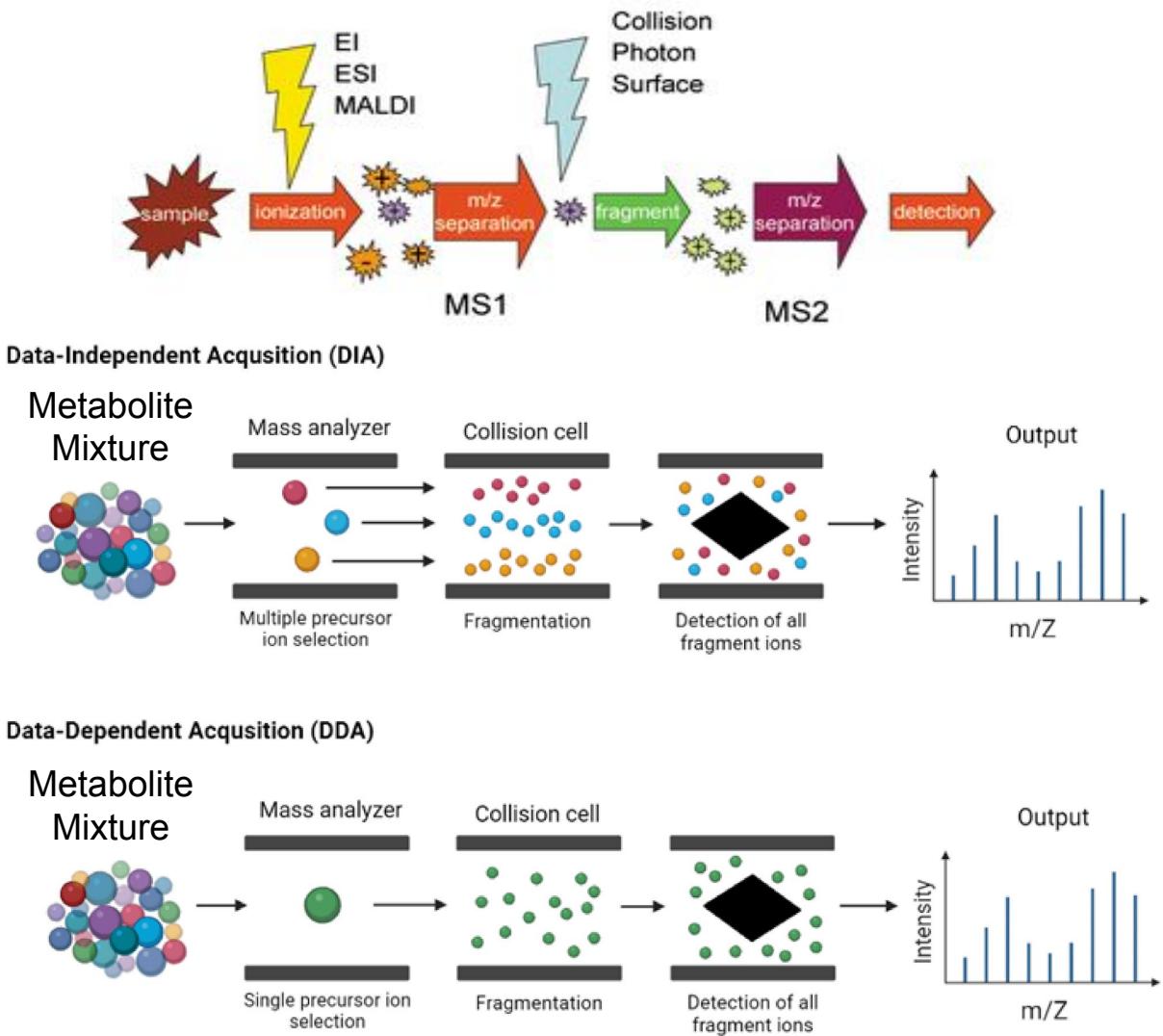
- Searching a metabolomic database (HMDB) for a m/z match or formula match can often give a tentative metabolite ID or a match to a small number of other possible metabolites
- Mass matching alone is not considered ideal for metabolite ID
- Use of additional information (RT data, CCS data, **MS/MS data** or authentic standards) is often required to make confident metabolite identifications

The screenshot displays two views of the HMDB website. The top view shows the homepage with the HMDB logo and navigation links. The bottom view shows a detailed search interface for a specific mass (175.01). The search form includes fields for Query Masses (Da), Ionization, Ion Mode (Positive), Adduct Type (Unknown), and Molecular Weight Tolerance. The results table lists 10 entries, each with the compound name, adduct type, adduct MW, compound MW, and delta value. The results are as follows:

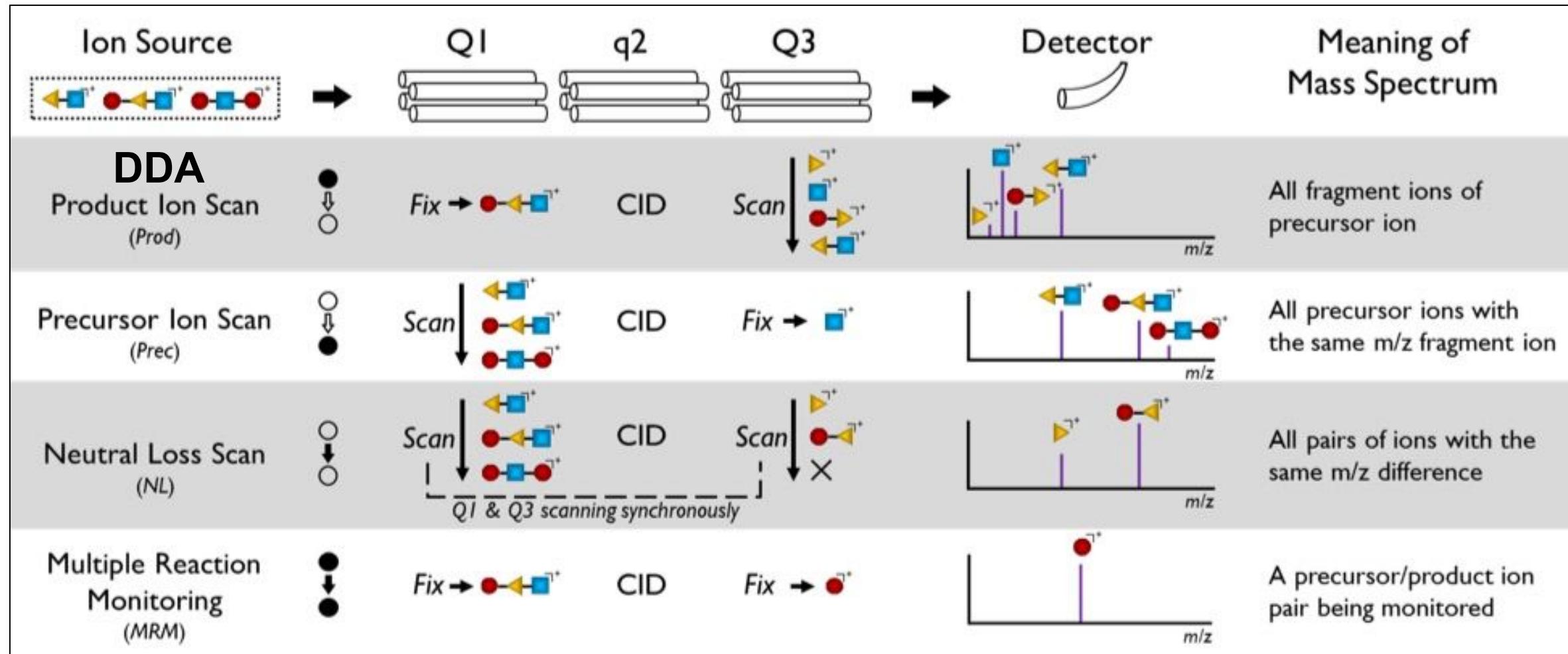
Compound	Name	Adduct	Adduct MW (Da)	Compound MW (Da)	Delta
HMDB00293	Hydroxidiodoxidosulfosulfate	M+IsoProp+H	175.009675	113.944535	0.000125
HMDB01436	Silicic acid	M+DMSO+H	175.009105	95.987885	0.000895
HMDB03657	De-O-methylsterigmatocystin	M+H+K	175.009086	310.047738	0.000914
HMDB03520	Aurantrichlide B	M+H+K	175.009086	310.047738	0.000914
HMDB34155	Thiourea	2M+Na	175.008256	76.009519	0.001744
HMDB01570	Thymidine 3',5'-cyclic monophosphate	M+2Na	175.012237	304.046037	0.002237
HMDB01270	Glyceric acid 1,3-biphosphate	M+2ACN+2H	175.013458	265.95927	0.003458
HMDB01294	2,3-Diphosphoglyceric acid	M+2ACN+2H	175.013458	265.95927	0.003458
HMDB00394	5-Fluorodeoxyuridine monophosphate	M+H+Na	175.014012	326.03153	0.004012
HMDB00015	Phenol sulphate	M+H	175.005955	173.998679	0.004045

Tandem Mass Spectrometry (MS/MS)

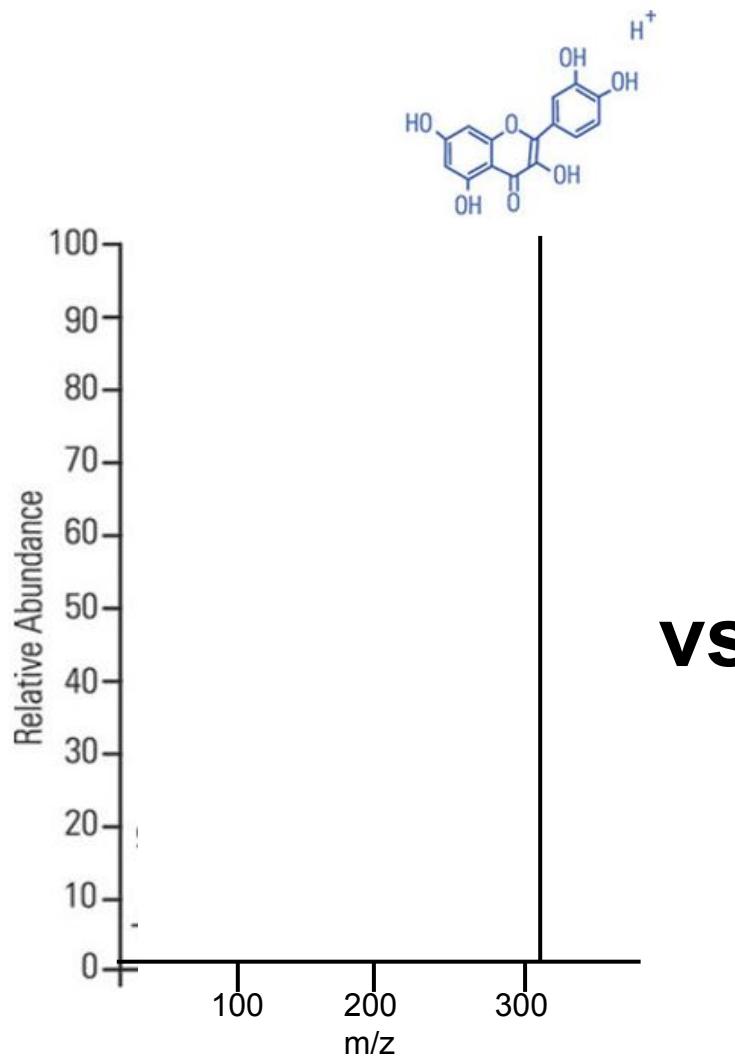
- A technique to fragment selected ions via collision induced dissociation (CID) and separating these via a second MS instrument
- Further fragmentation of the parent ion to produce product ions gives additional structural information over just an m/z value
- Can be done with QqQ, QToF or OrbiTrap MS instruments
- Can be done in several modes – DIA, DDA, product ion scan, etc.



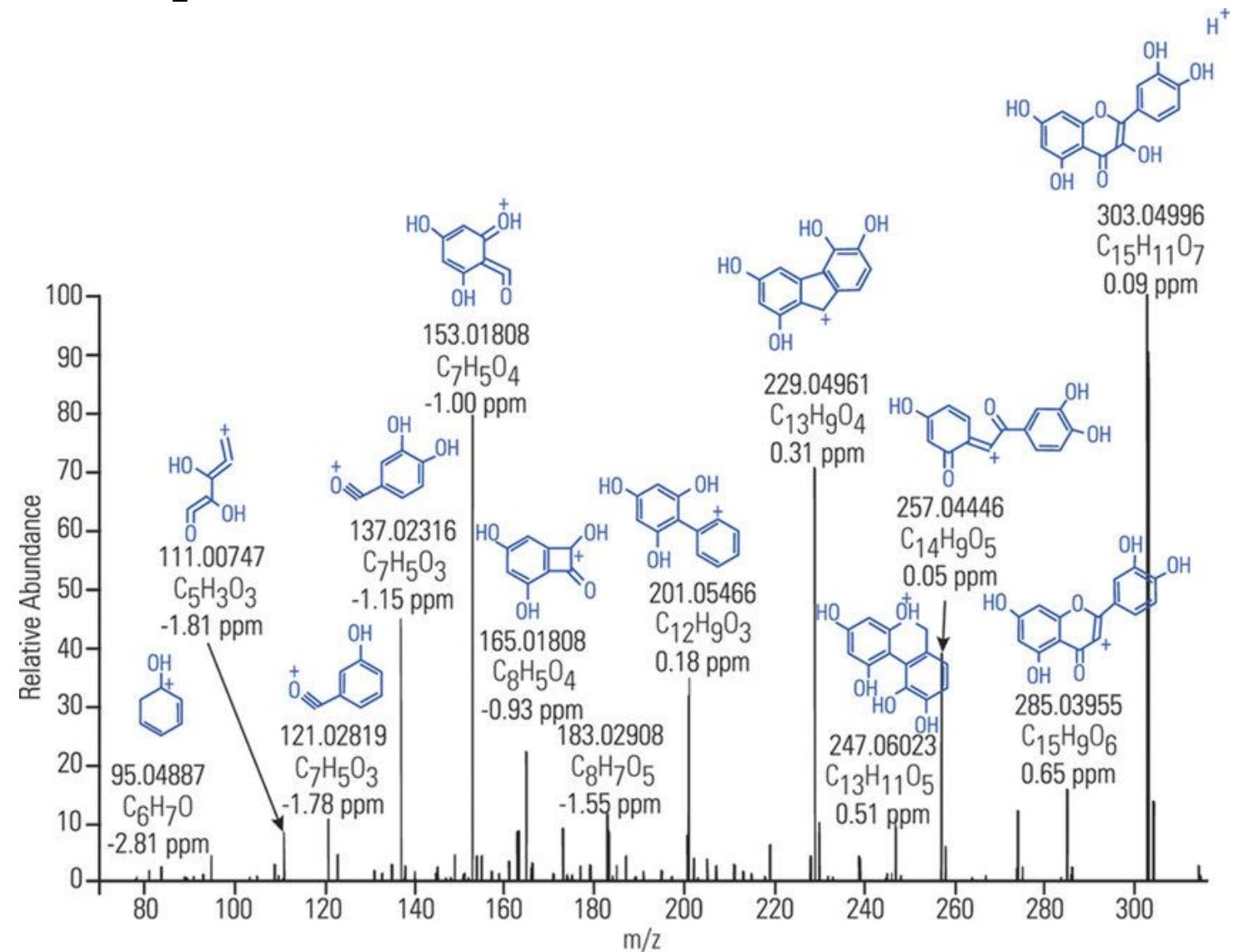
Tandem Mass Spectrometry



MS vs. MS/MS Spectrum of Quercetin

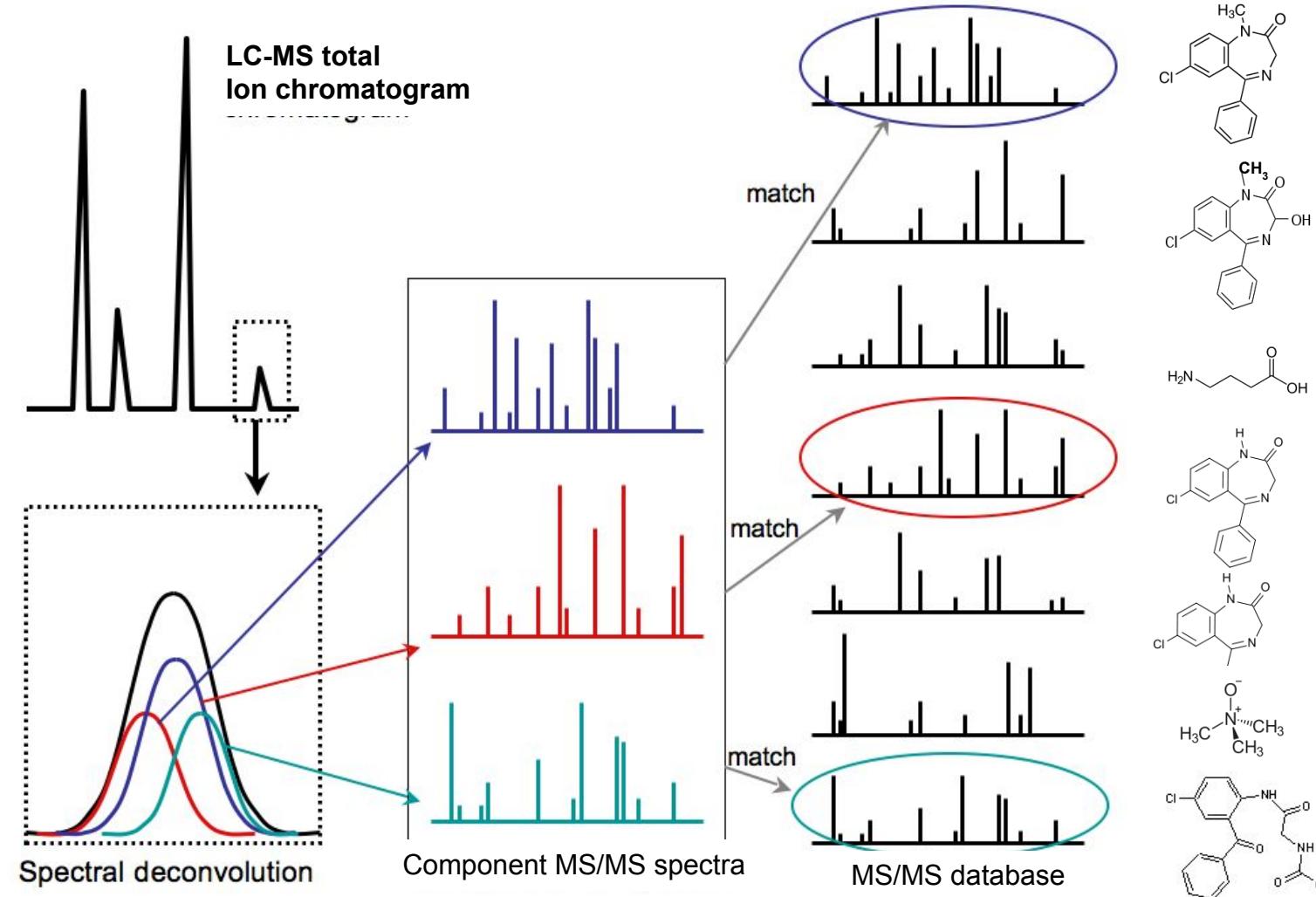


VS



Metabolite ID by LC-MS/MS

- 4 levels of metabolite ID
- Positively identified compounds
 - Confirmed by match to known standard
- Putatively identified compounds
 - Match to MS+RT or MS/MS+RT
- Compounds putatively identified in a specific compound class
- Unknown compounds

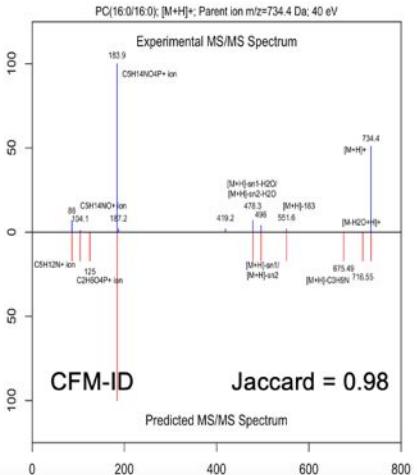
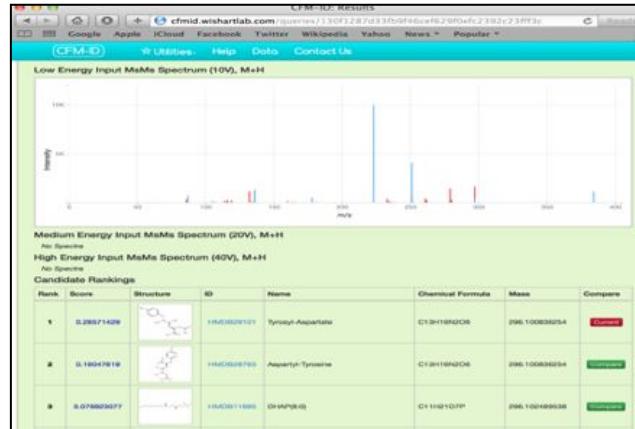


MS/MS Spectral DBs

- **MoNA** – 658,790 spectra, 223,614 cmpds* (12,000)
- **METLIN** – 2,500,000 spectra, 500,000 cmpds (80,038)
- **mzCloud** – 2,238,933 recal. spectra, 8779 cmpds
- **NIST17 MS/MS** – 574,826 spectra, 13,808 cmpds
- **MassBankEU** – 80,661 spectra, 14,382 cmpds
- **ReSpect** – 9017 spectra, 3595 cmpds
- **GNPS** – 154,820 spectra, 13,717 natural products

Total #compounds with exp. MS/MS spectra ~85,000

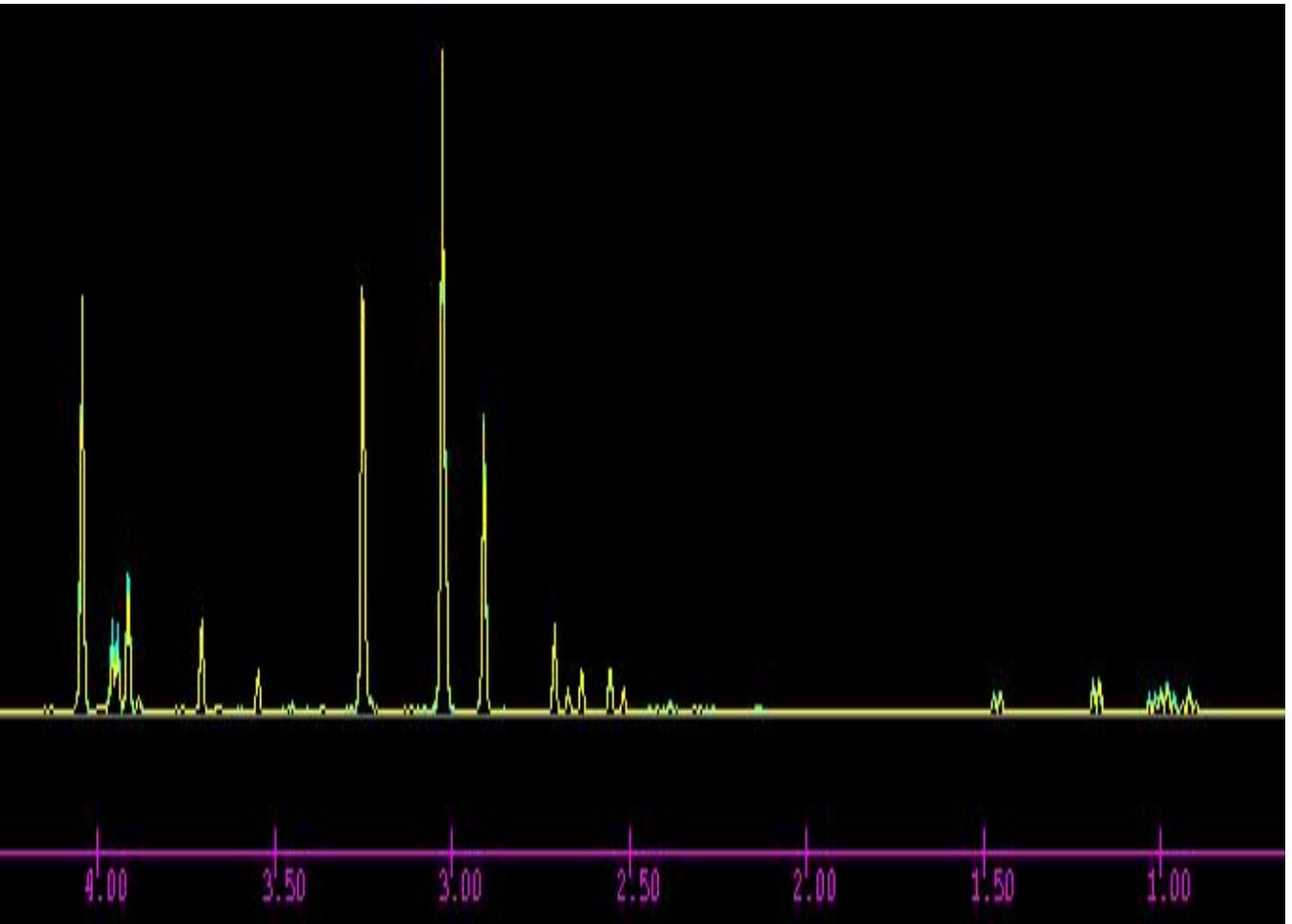
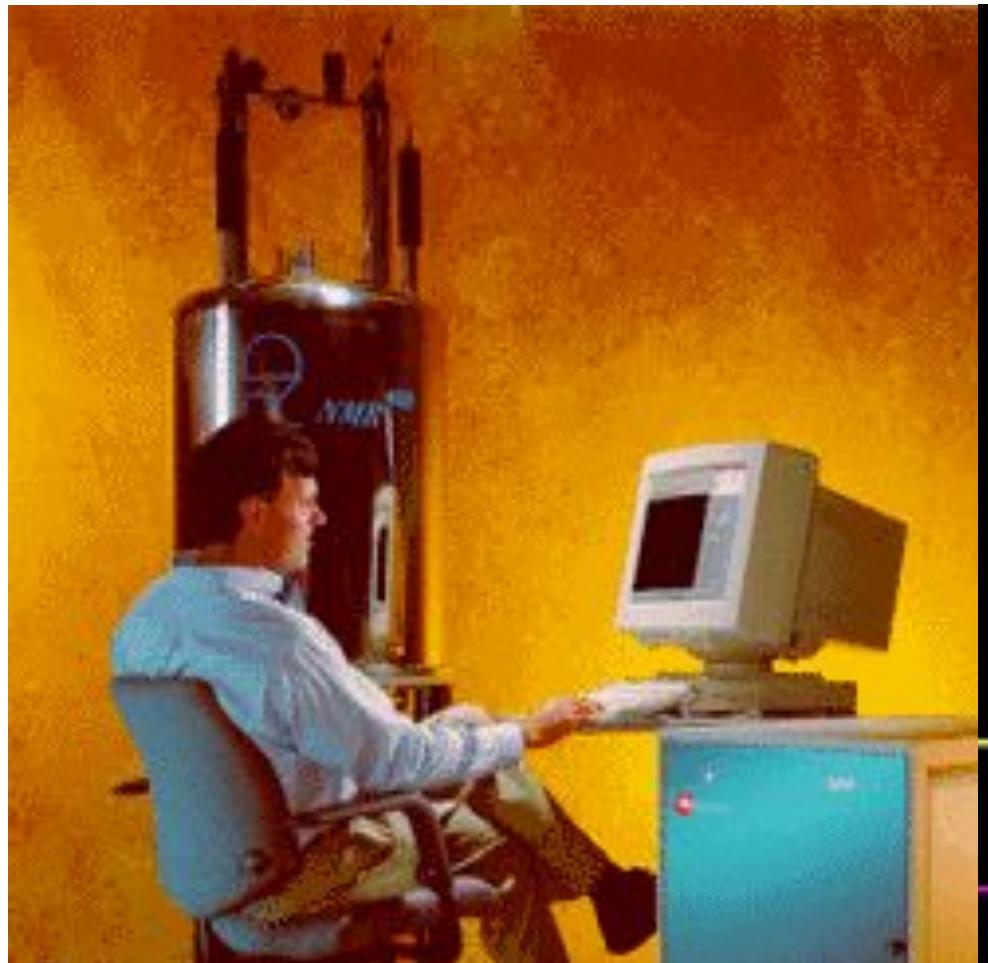
MS/MS Spectra Are Predictable



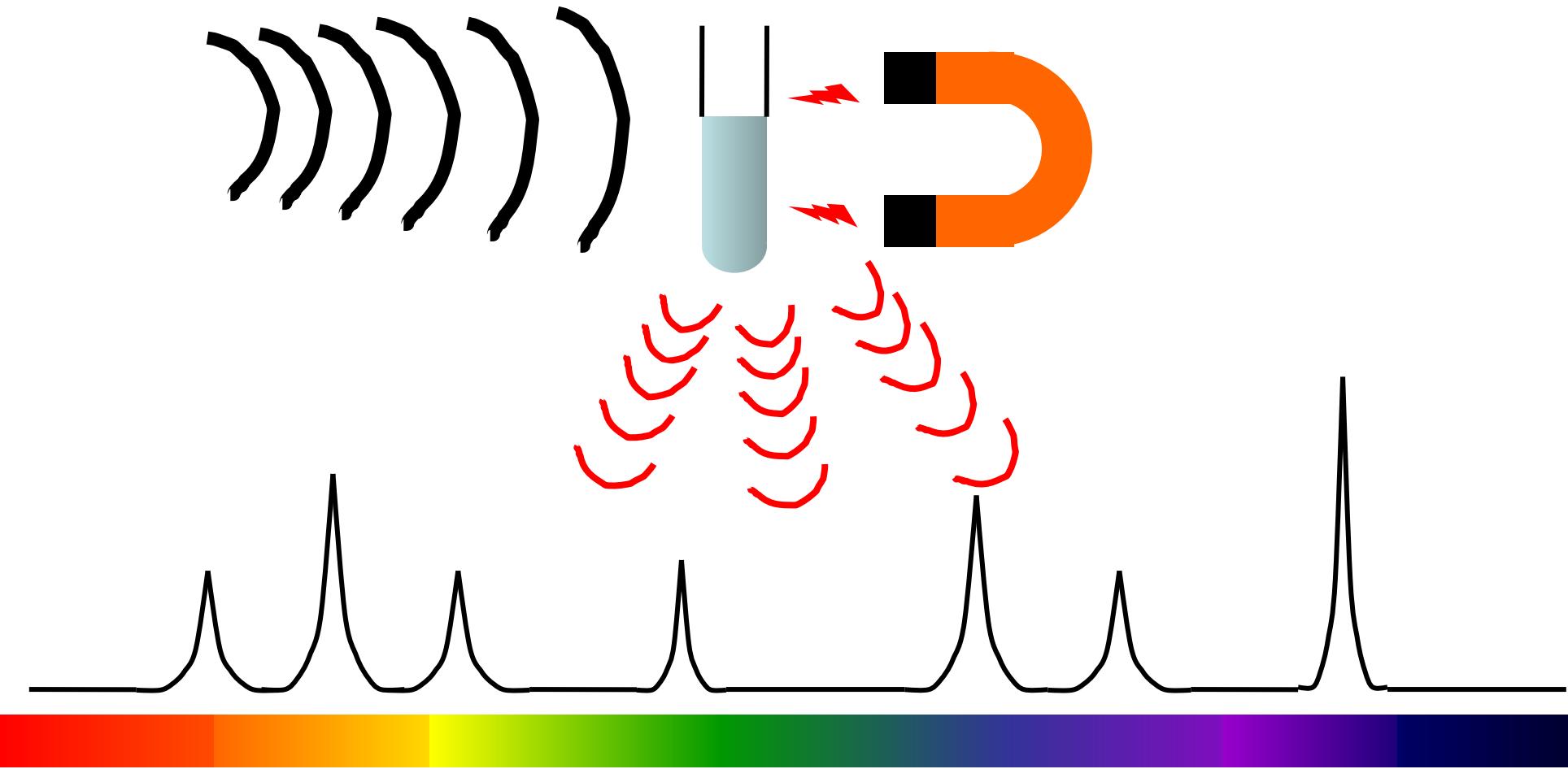
	CFM-ID 2.0 (2016)	CFM-ID 3.0 (2018)	CFM-ID 4.0 (2020)
10 eV collision energy Dice Score (non-lipid)	35.1	35.1	44.8
20 eV collision energy Dice Score (non-lipid)	29.2	29.2	34.9
40 eV collision energy Dice Score (non-lipid)	24.5	24.5	31.9
Lipid (ave. over 14 lipid types) Dice Score	8.5	90.7	90.7
CASMI 2016 performance (208 test set)	120/208	149/208	165/208

- **CFM-ID - Competitive Fragment Modeling for Identification**
- **Now in its 4th Edition**
- **Uses a large training set of high-resolution MS/MS data of known compounds at 10, 20, 40 eV CID collision energies to train an MS/MS “fragmenter”**
- **The fragmenter slowly learns from its training data (HMM)**
- ***Supports novel compound ID, MS/MS spectral prediction and MS/MS spectral annotation and molecular formula prediction***
- **Version 4.0 is now available with expt. and predicted MS/MS spectra for >300,000 different compounds**

NMR Spectroscopy



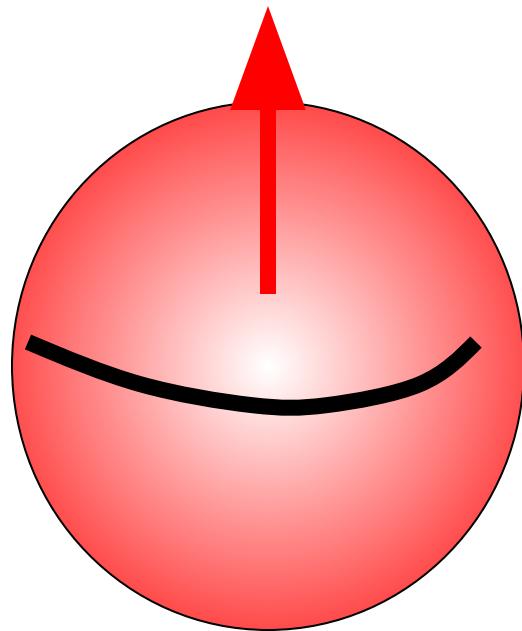
Explaining NMR



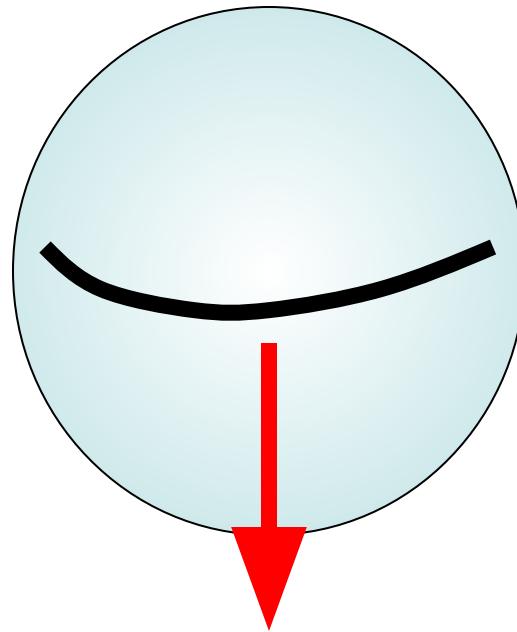
Principles of NMR

- Measures **nuclear magnetism or changes in nuclear magnetism in a molecule**
- **NMR spectroscopy measures the absorption of light (radio waves) due to changes in nuclear spin orientation**
- **NMR only occurs when a sample is placed in a strong magnetic field**
- **Different nuclei absorb at different energies (frequencies) leading to different absorption peaks**

Protons (and other nucleons) Have A Property Called Spin

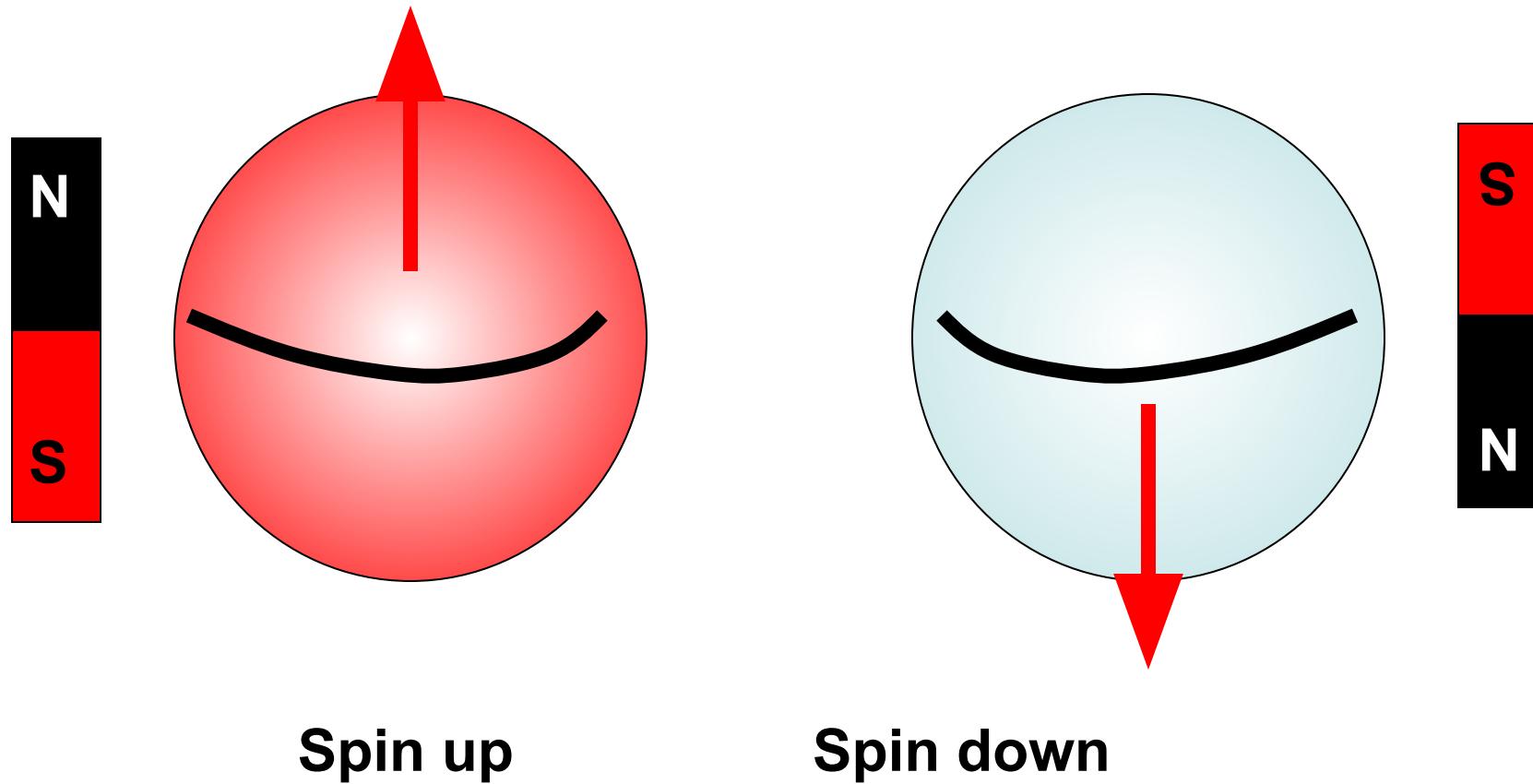


Spin up

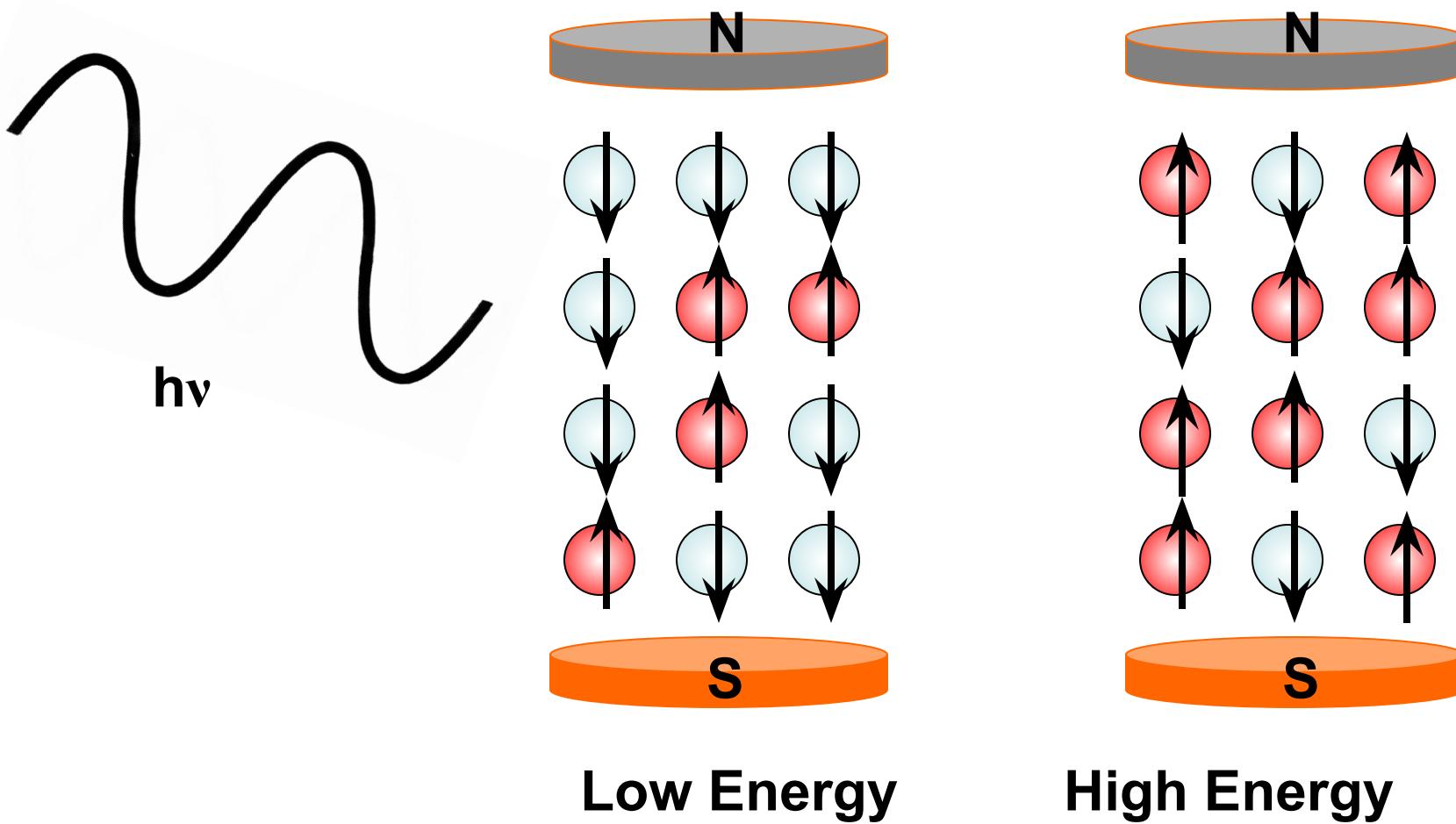


Spin down

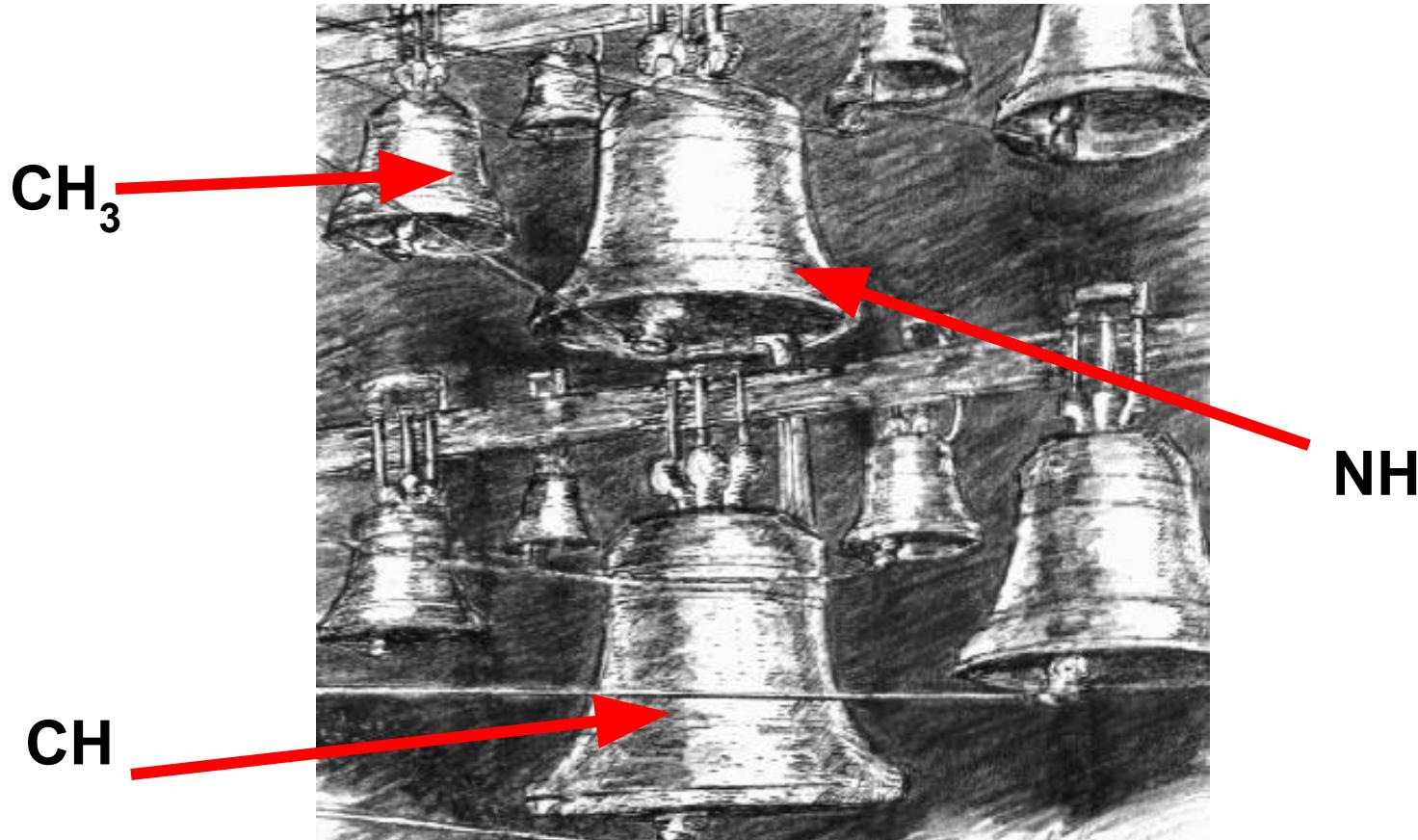
Each Spinning Proton is Like a “Mini-Magnet”



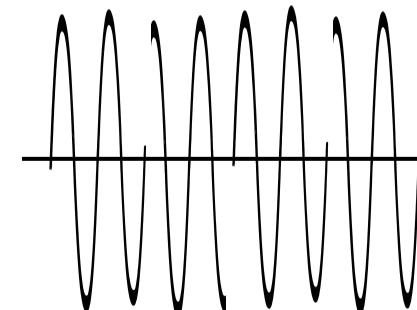
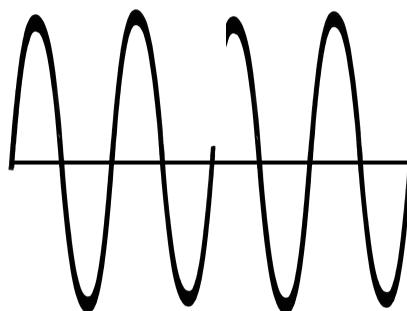
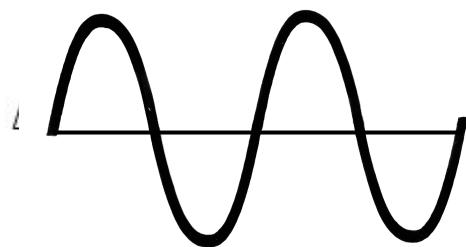
Principles of NMR



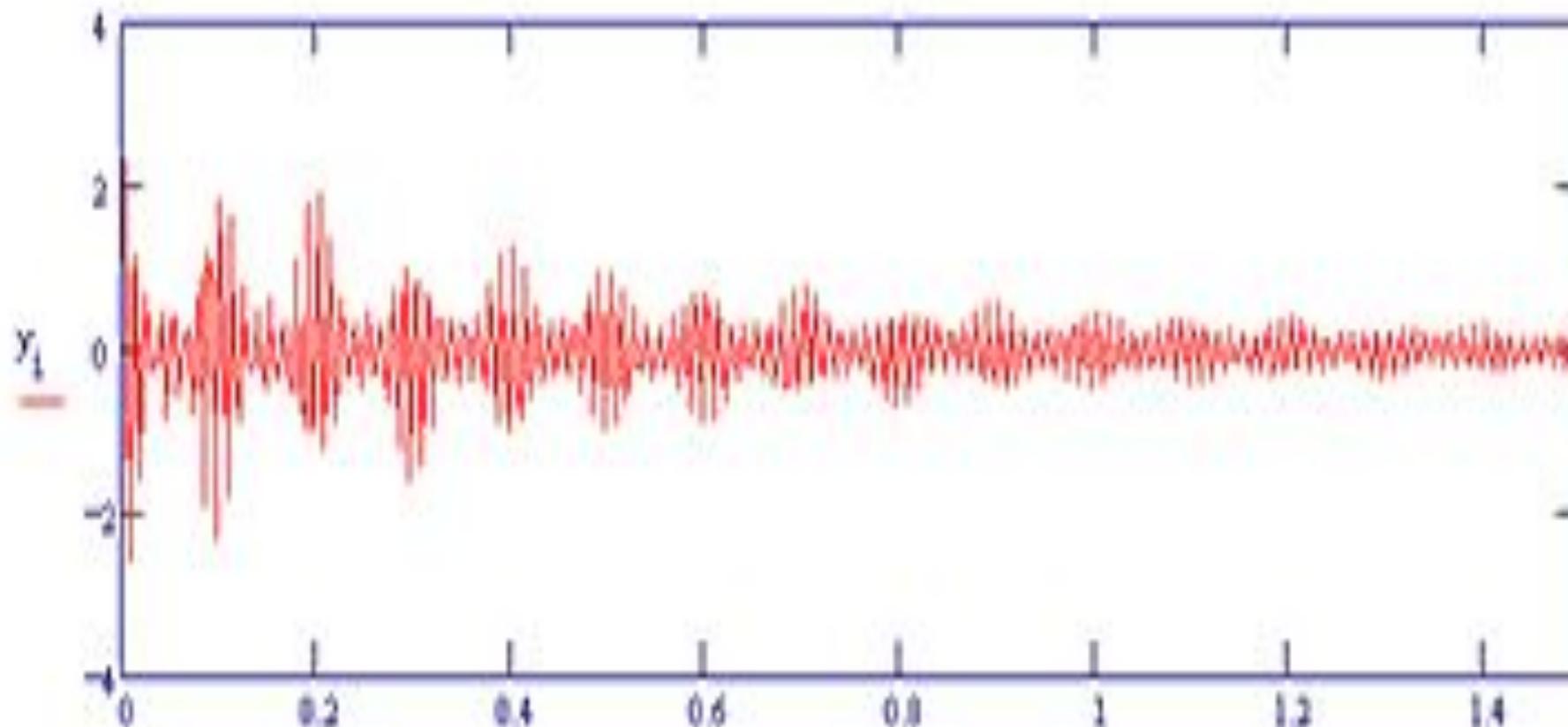
Different Chemical Groups Behave Like Different Bells



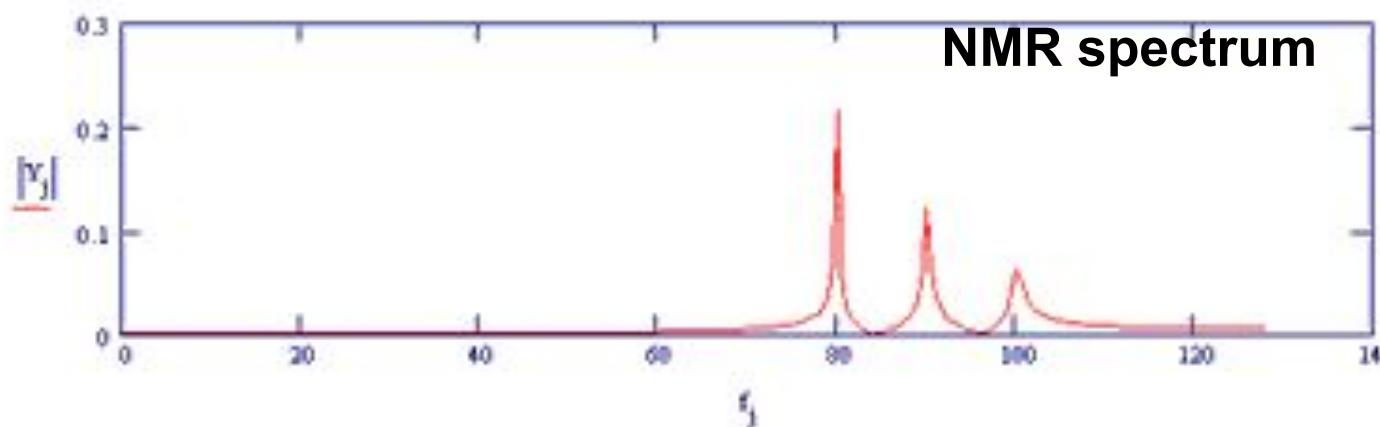
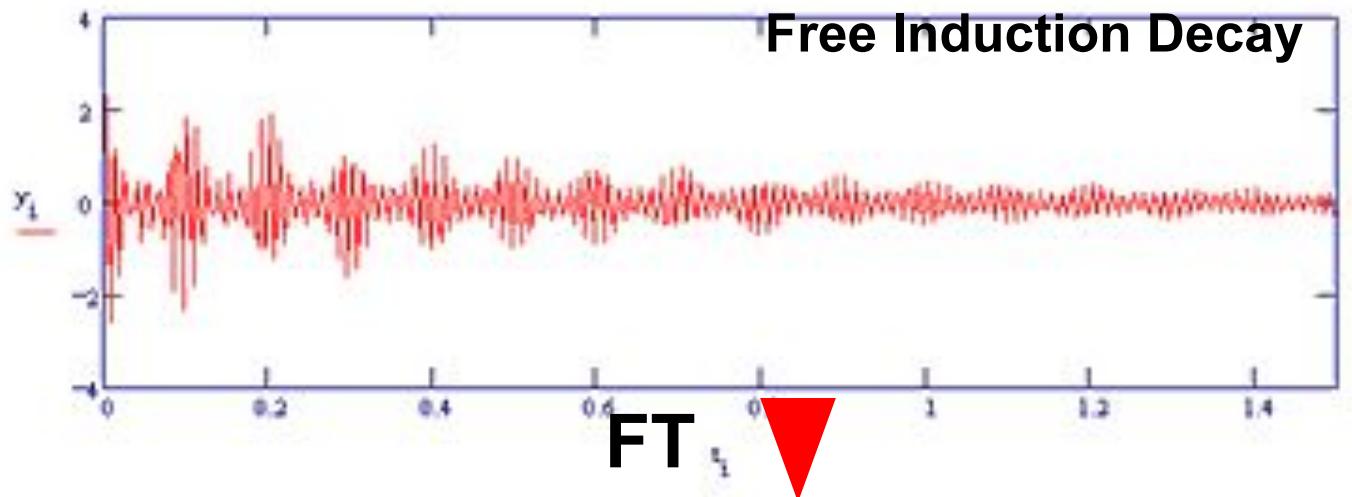
Different Bells (chemical groups) Ring At Different Frequencies



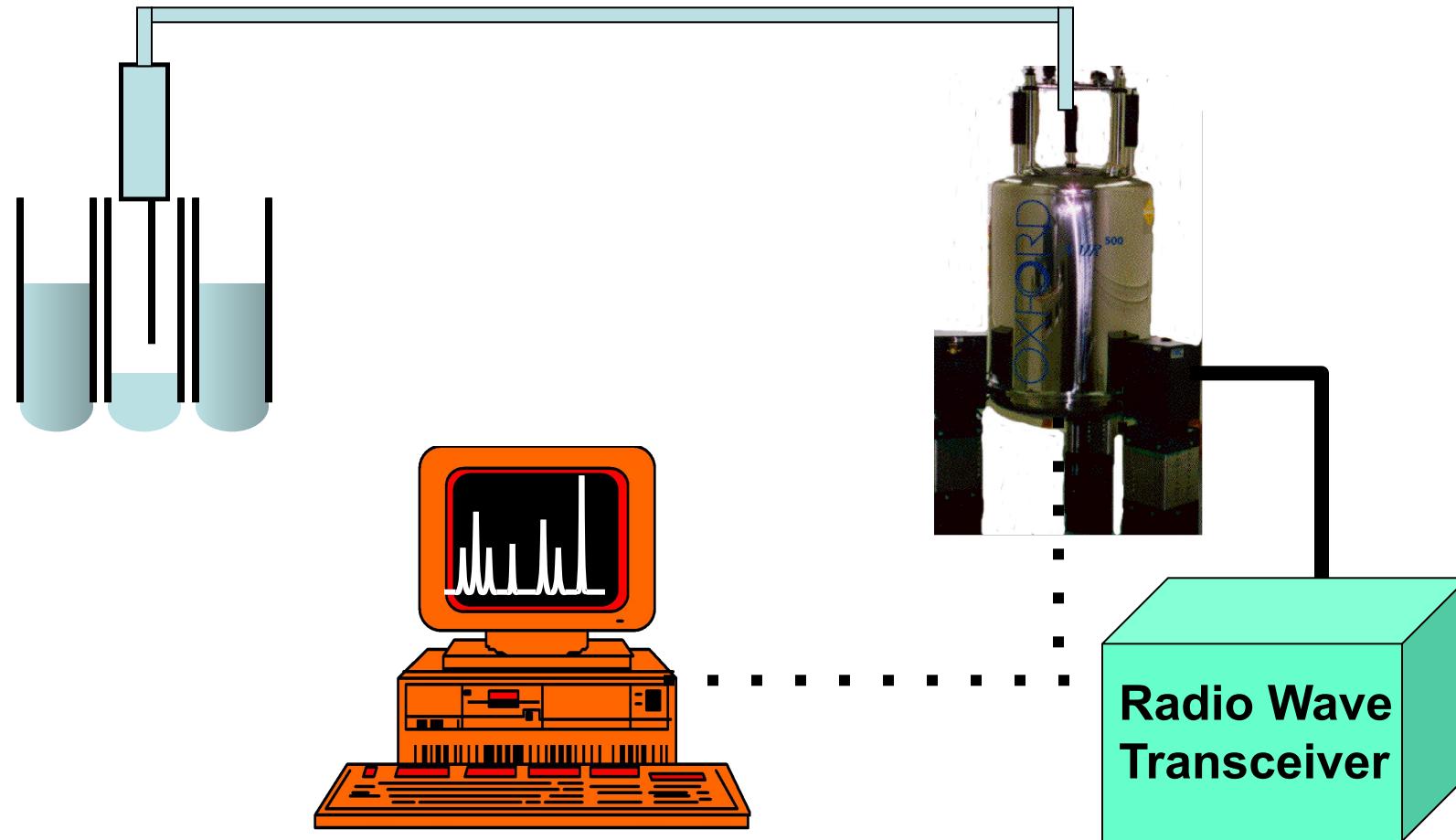
**Ring All The Bells Together
And You Get...**



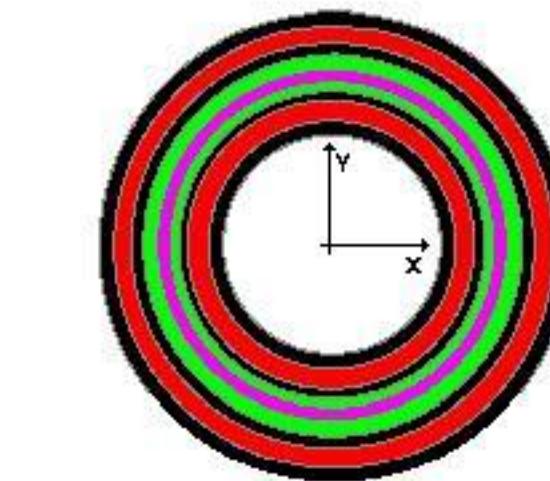
Changing Time Signals to Frequency Signals



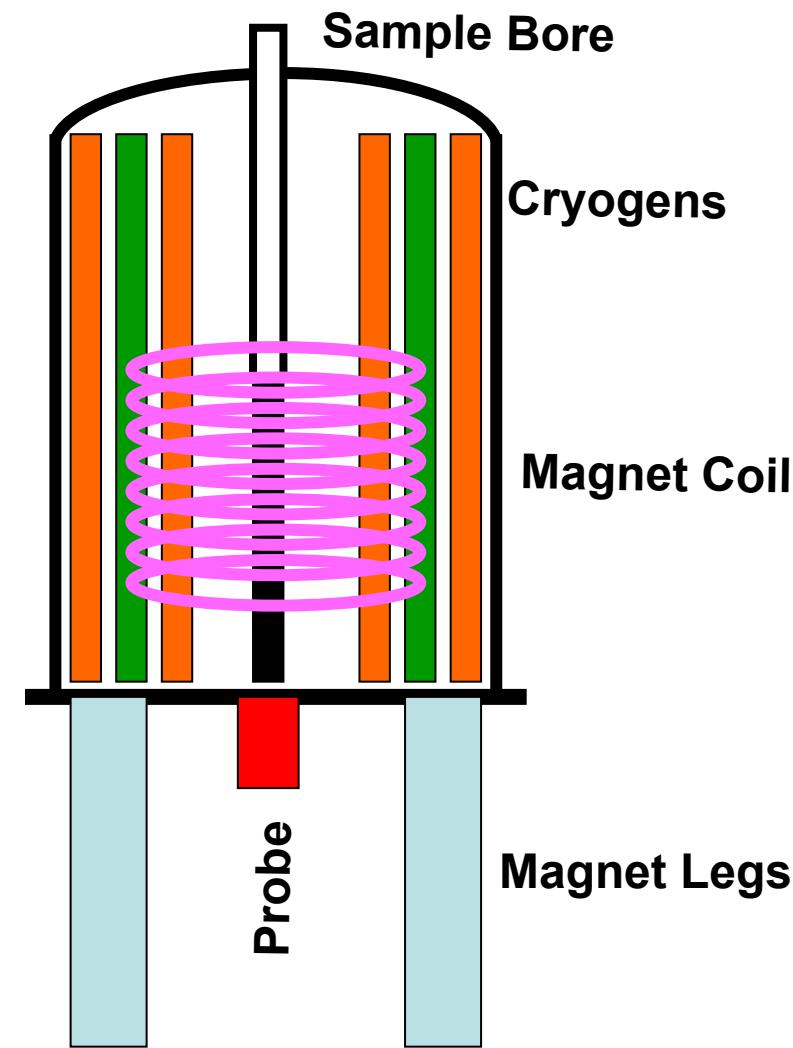
A Modern NMR Instrument



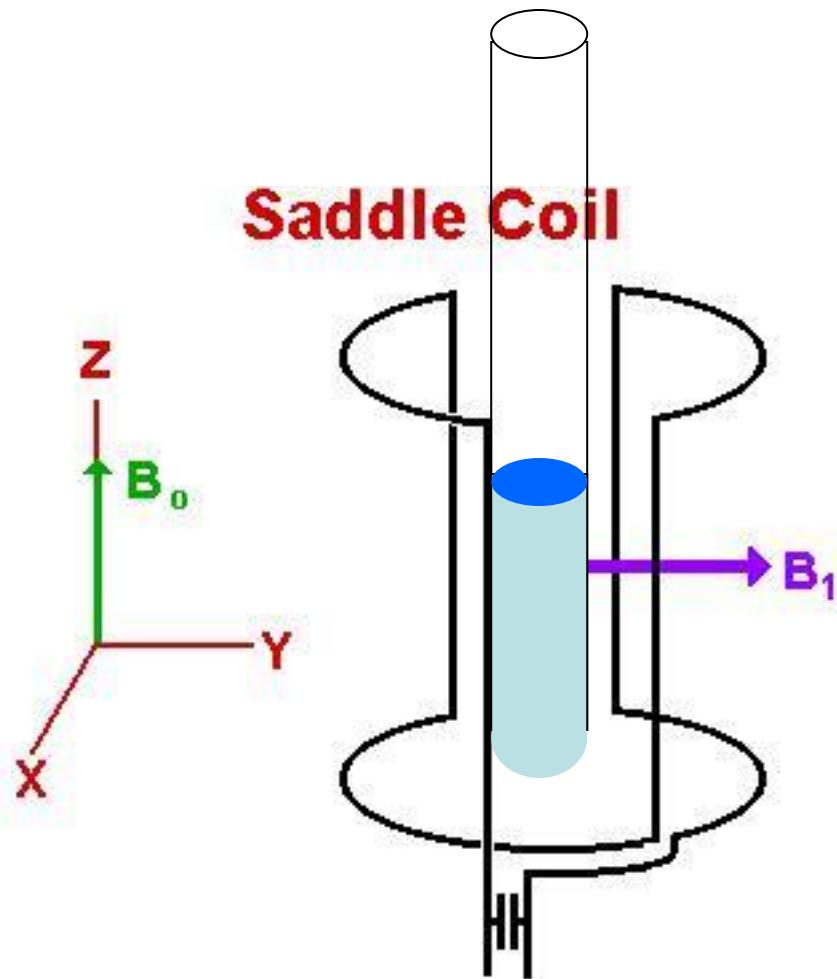
NMR Magnet Cross-Section



- Vacuum
- Liquid Helium
- Liquid Nitrogen
- Container & Support
- Superconducting Coil

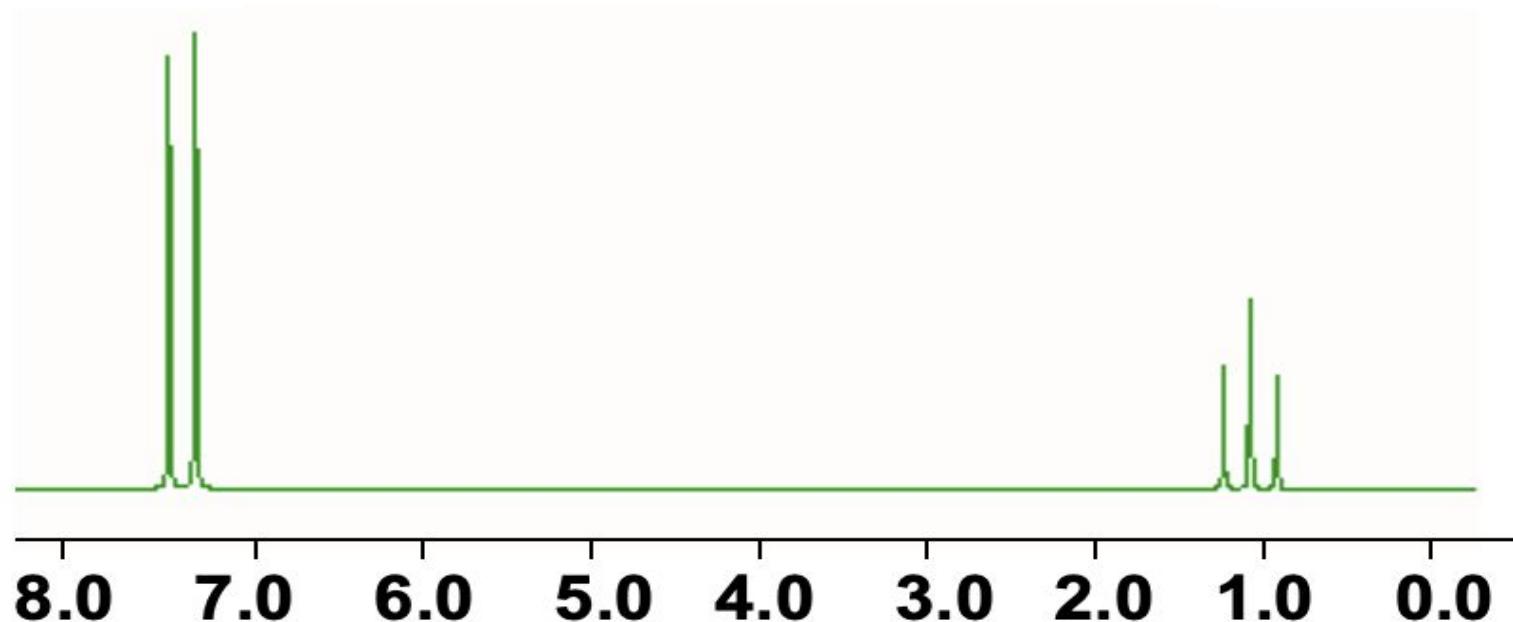


NMR Sample & Probe Coil



^1H NMR Spectra Exhibit...

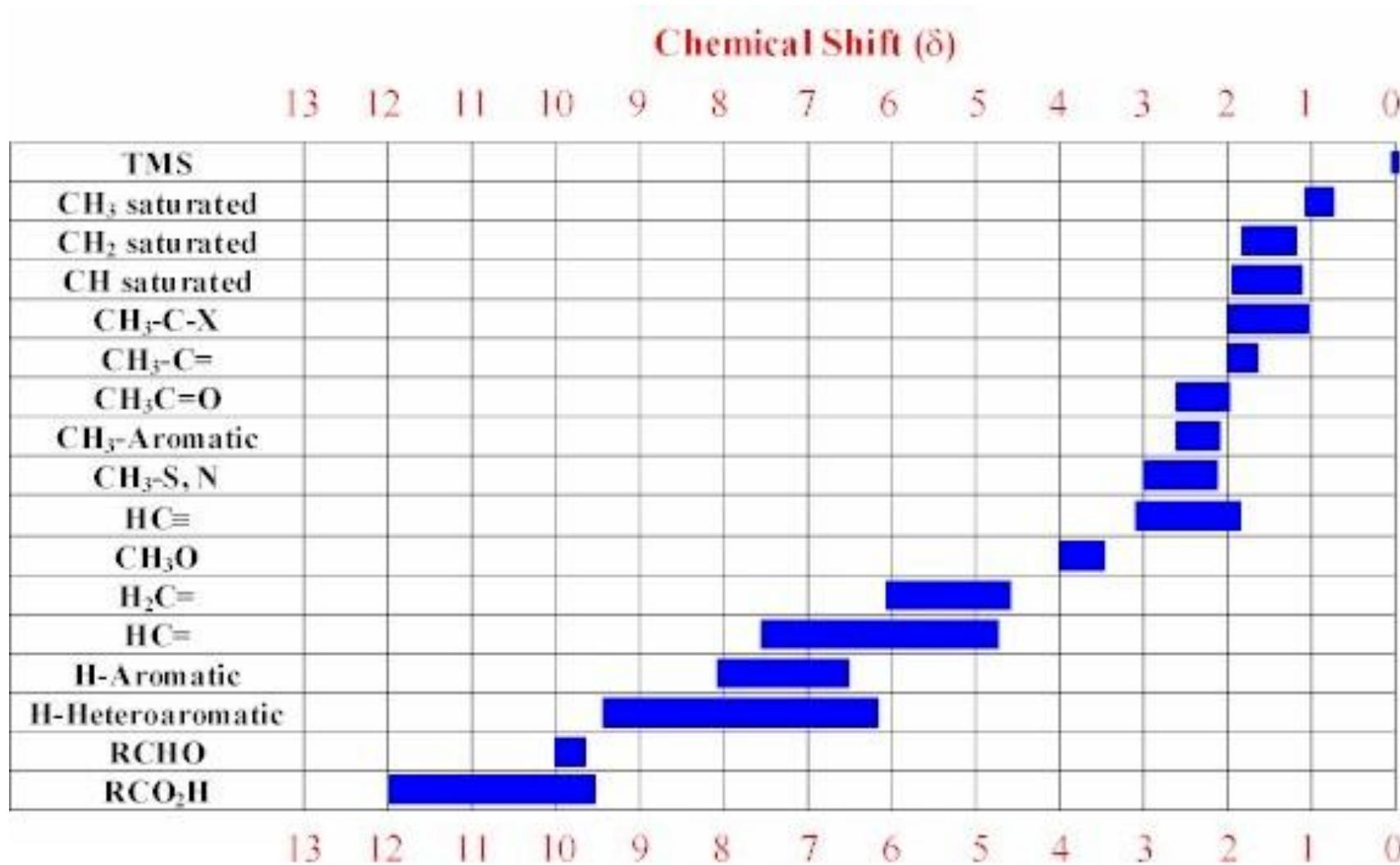
- Chemical Shifts (peaks at different frequencies or ppm values)
- Splitting Patterns (from spin coupling)
- Different Peak Intensities (# ^1H atoms)



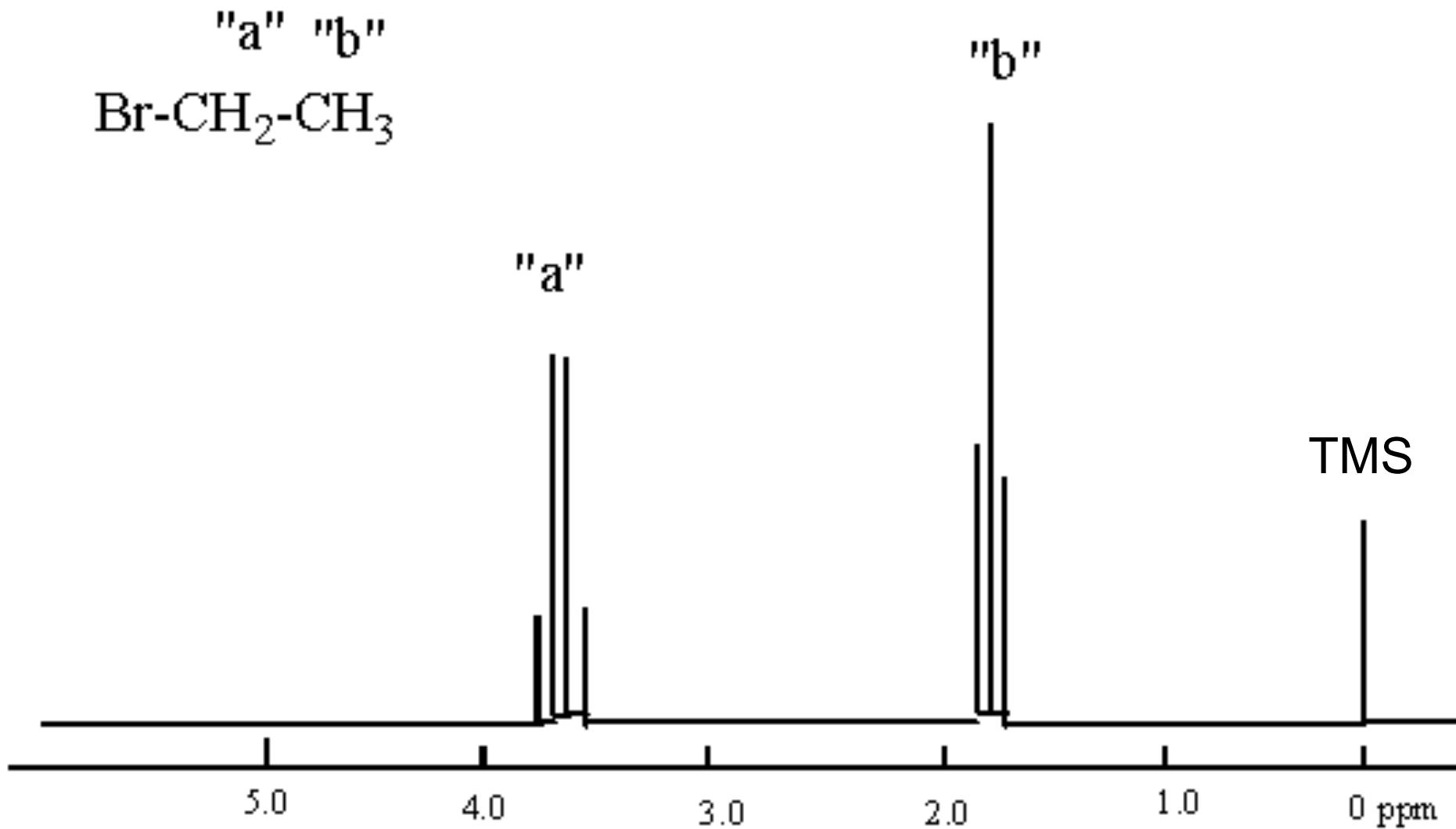
NMR Chemical Shifts

- Key to the utility of NMR in chemistry
- Chemical shifts reflect atom-specific NMR absorption frequencies
- Different ^1H atoms in different molecules exhibit different absorption frequencies or chemical shifts
- Each compound can be defined by a unique pattern of chemical shifts (a fingerprint)
- Chemical shifts are mostly affected by electronegativity of neighbouring atoms, bonds or groups

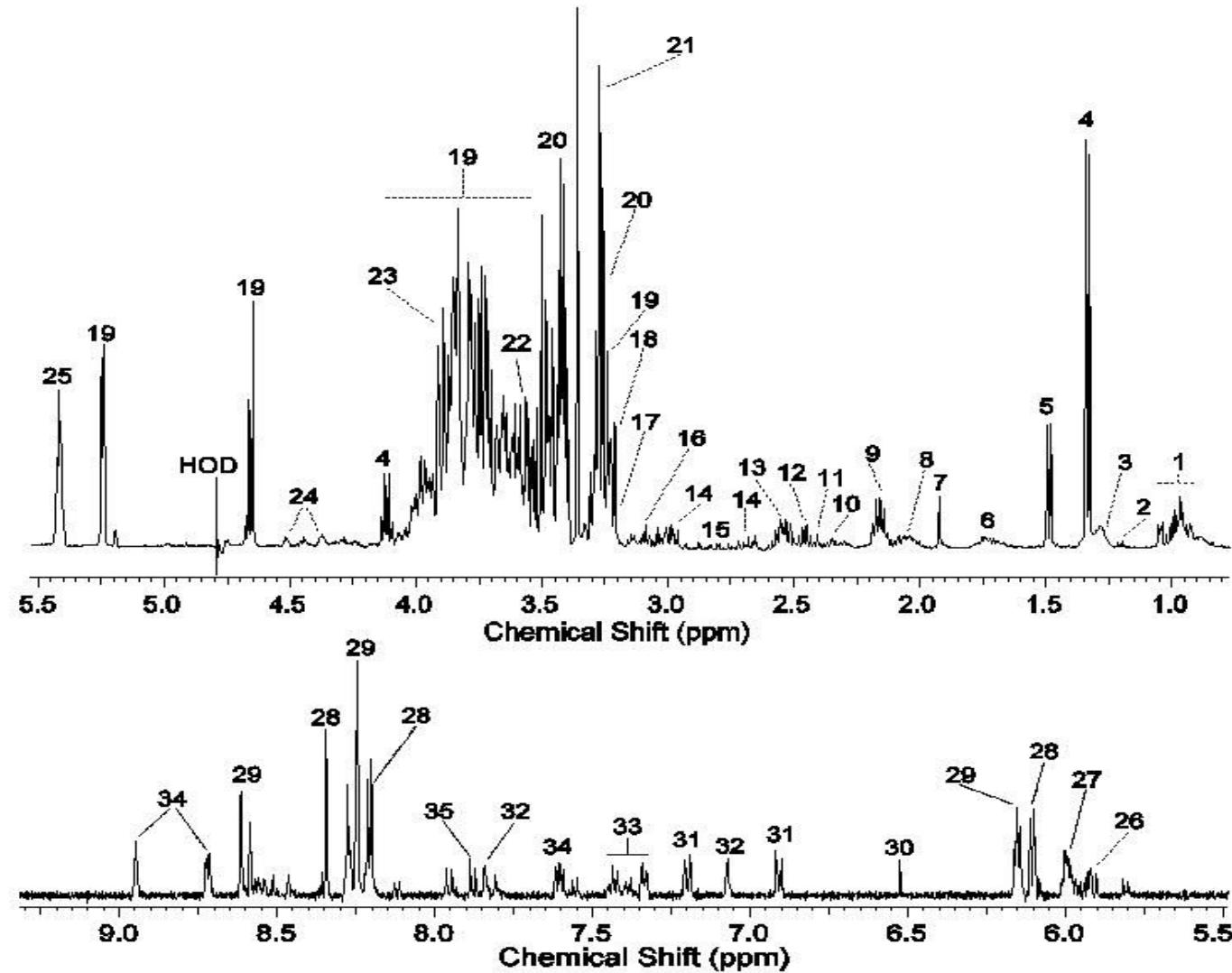
Characteristic Chemical Shifts



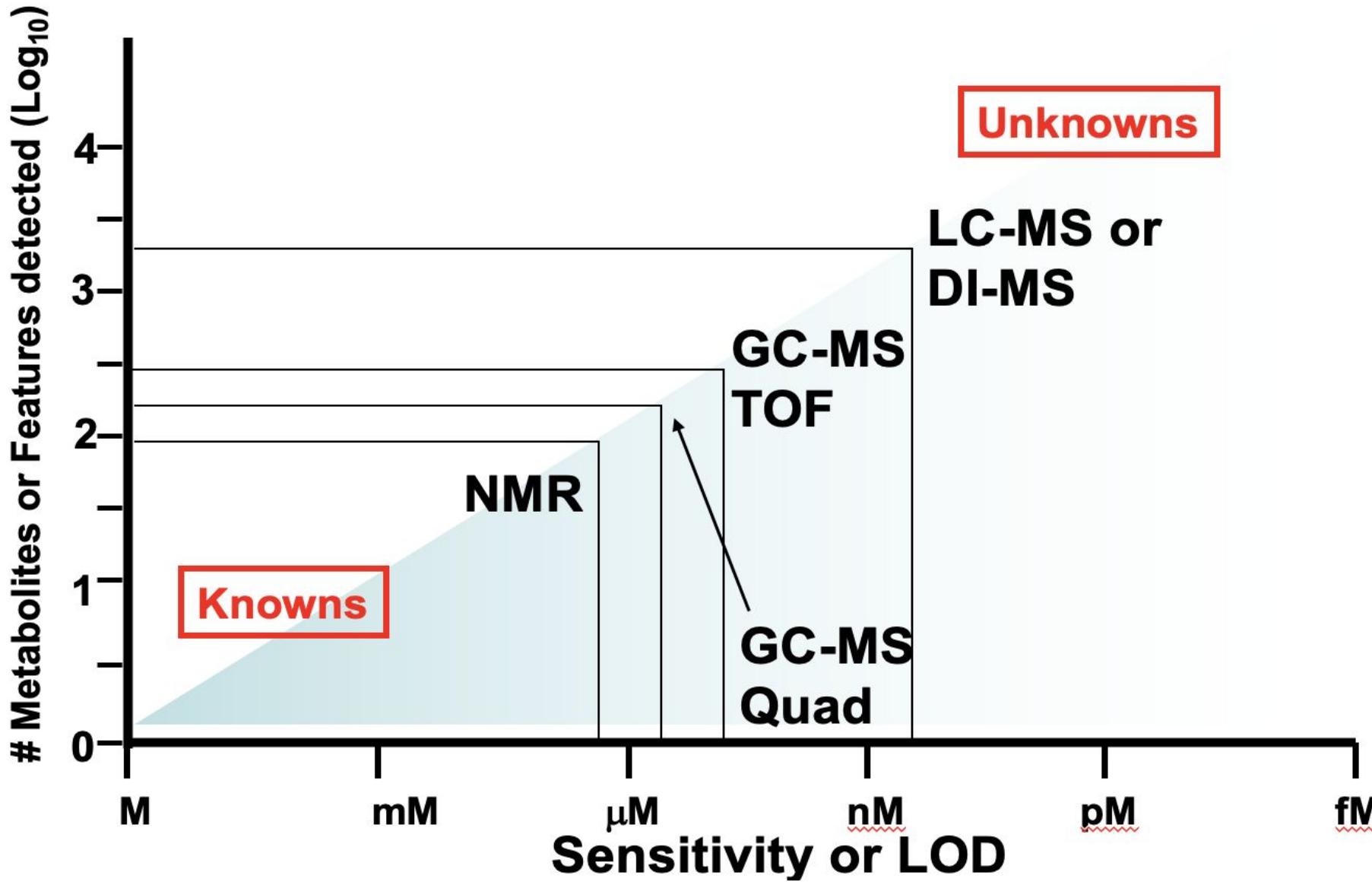
Assigning Simple NMR Spectra



NMR Spectrum of a Biological Mixture

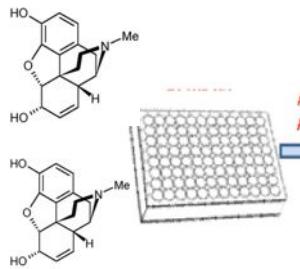


Technology & Sensitivity



Two Routes to Metabolomics

Targeted (Quantitative)



Add Internal Stds

Add Samples

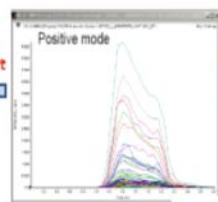
Protein Precipitation

Metabolite Extraction

MS Analysis

MS Analysis

QqQ or Qtrap MS
NMR
¹³C-MS



MS-Autofit

Asymmetric Dimethylarginine	0.36
L-arginine	2.14
Delta-1-hydroxybutyrate	50
Beta-alanine	125
All in hair	15
Carnosine	0.85
Choline	1.2
Creatinine	0.5
Folate	0.025
Ketone	3450
Glucose	3450



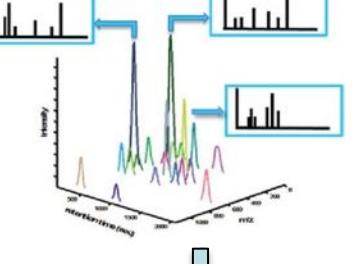
Untargeted (Non-Quantitative)



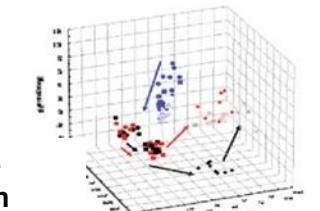
Liquid Chromatography
High Res Mass Spectrometry

Clustering, Peak detection

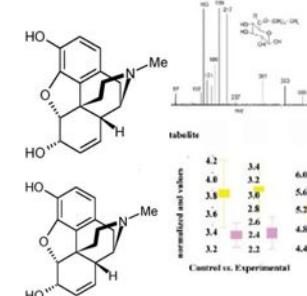
Data acquisition



Feature Selection



Metabolite Annotation

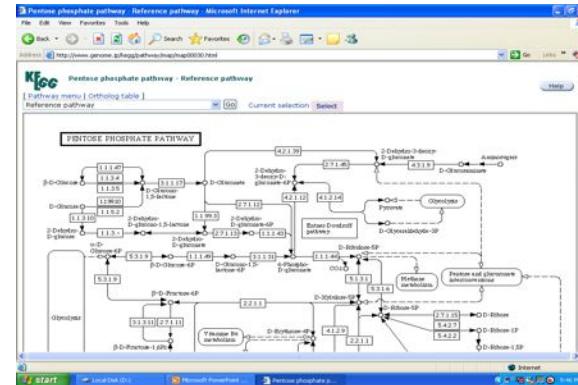
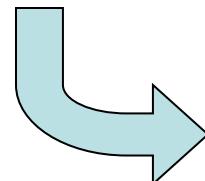


Targeted vs. Untargeted

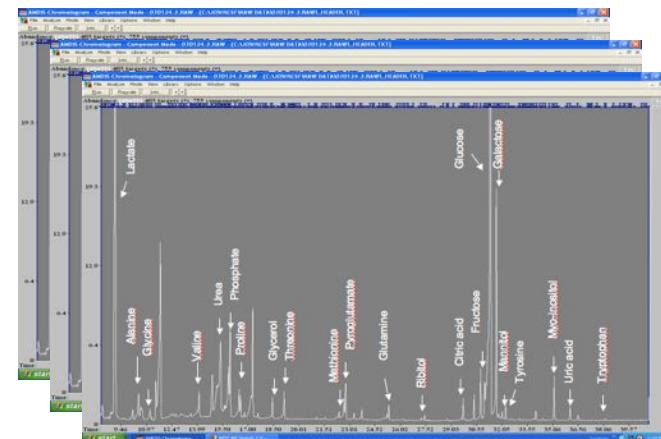
- Limited coverage (10-1500 pre-selected molecules)
- Limited in potential for discovery (hypothesis testing)
- Focus on absolute quantification
- Amenable to automation & kits
- Highly standardized or standardizable
- Almost unlimited coverage (>10,000 features)
- Good potential for discovery (hypothesis generating)
- Limited to relative quantification
- Not very fast or automated
- Needing much more standardization

Targeted Metabolomics

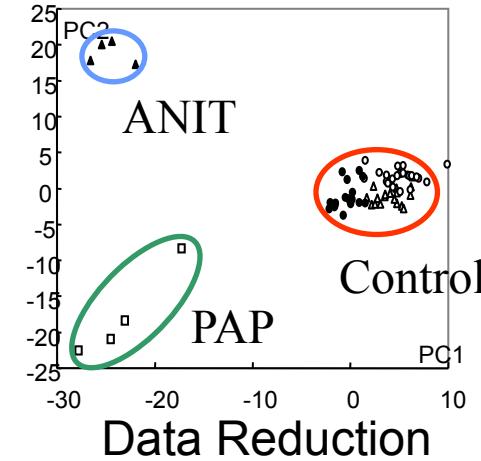
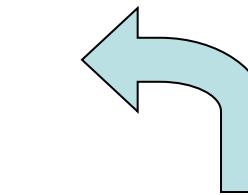
Sample Prep



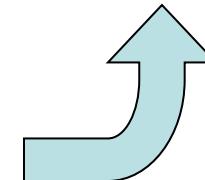
Biological Interpretation



Metabolite Identification & Quantification



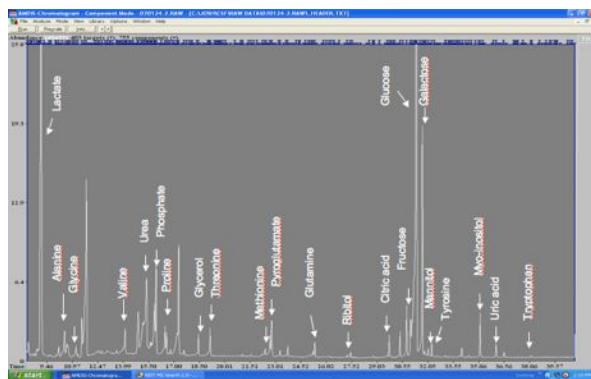
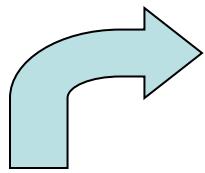
Data Reduction



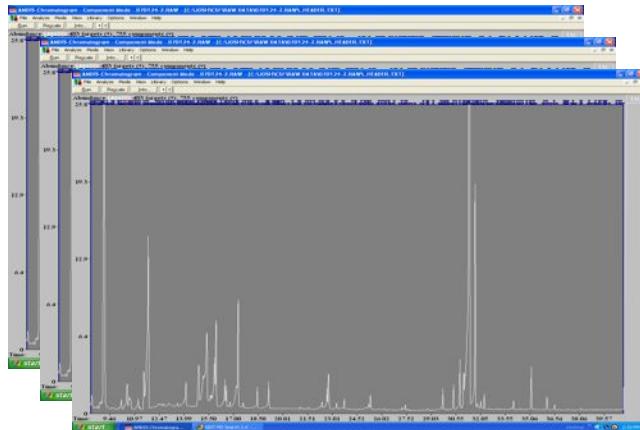
Untargeted Metabolomics



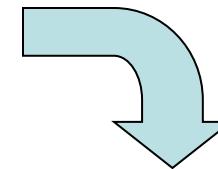
Sample Prep



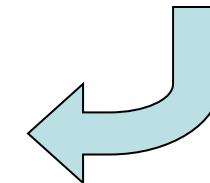
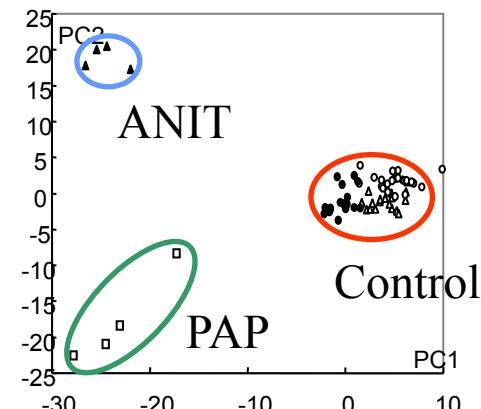
Metabolite Identification



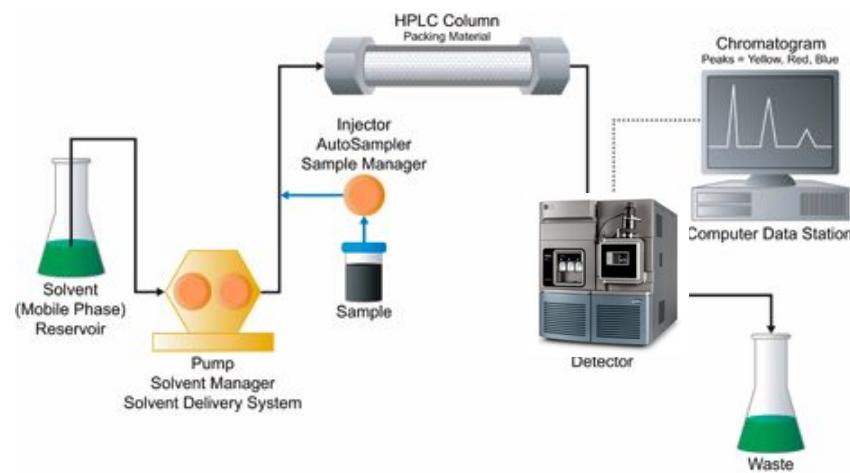
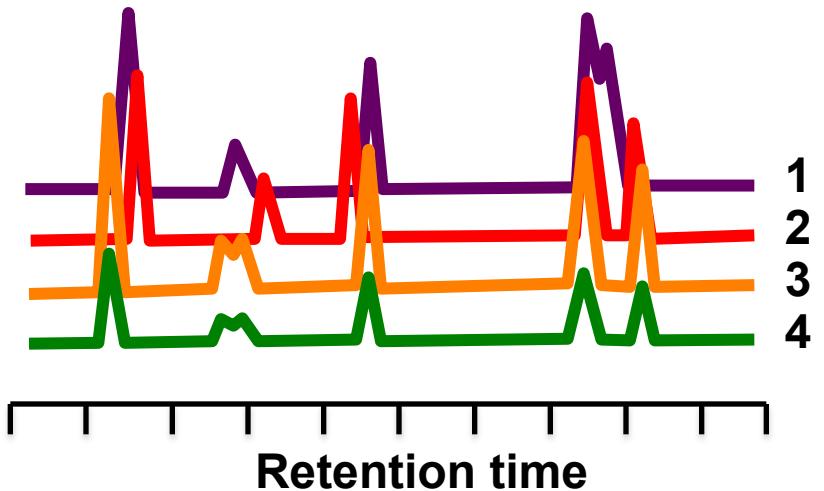
Data Collection



Data Reduction

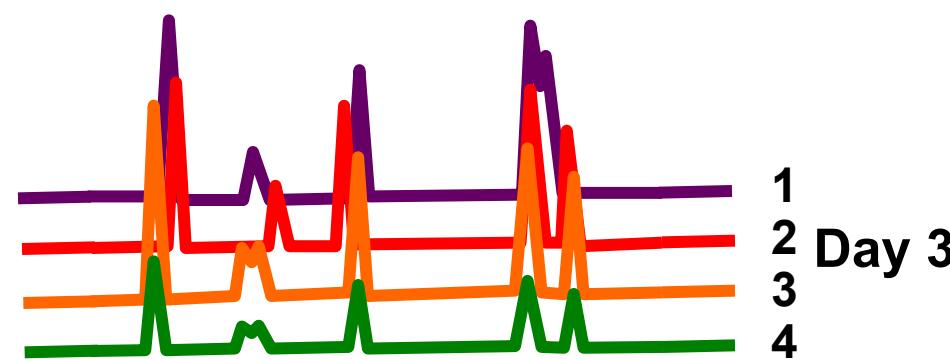
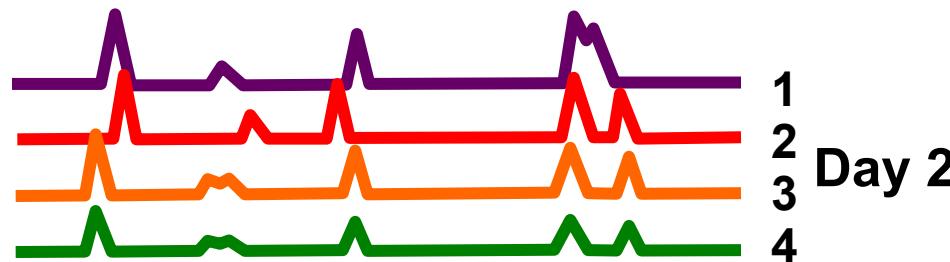
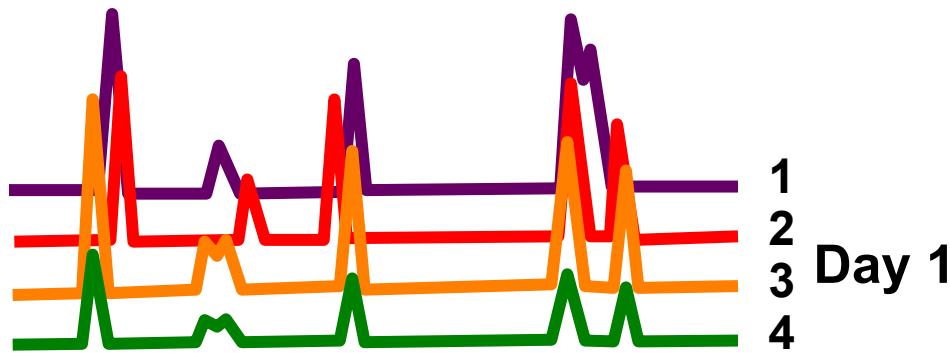


Untargeted Metabolomics

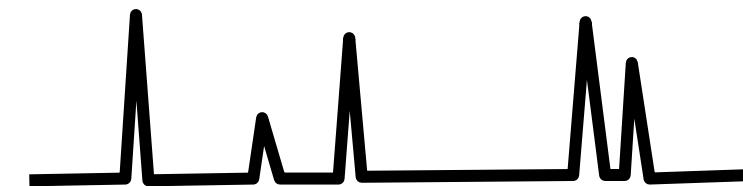
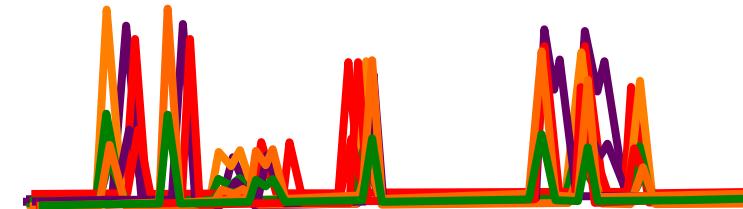


- 1000s of metabolites need to be separated by LC
- 100s of samples to be run over many days or weeks
- LC & MS are not very reproducible (varies with runs, varies with days – ***called batch variation***)
- Changes in retention time, intensity, mass often seen
- How to compare? how to find the “right” features? how to scale?

Liquid Chromatography & Batch Effects

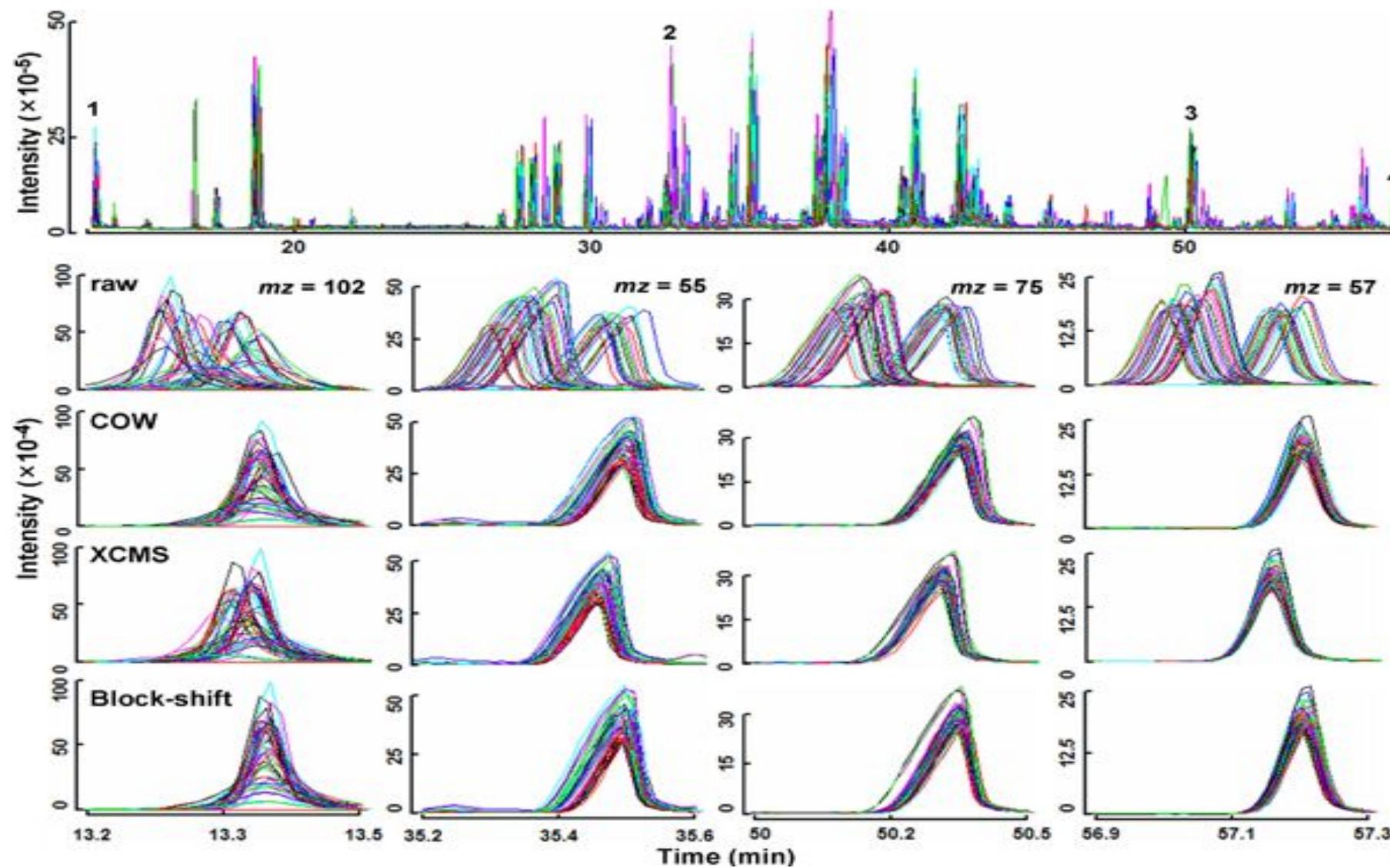


What you get

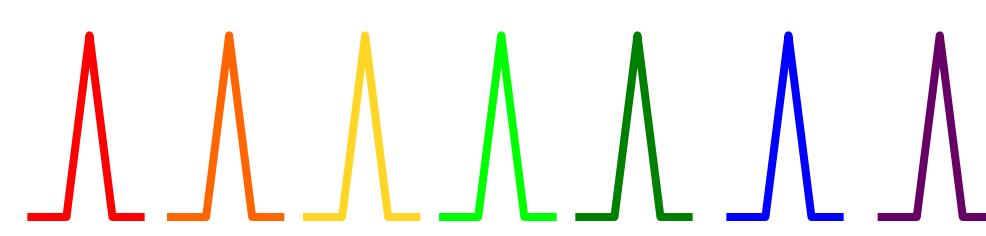
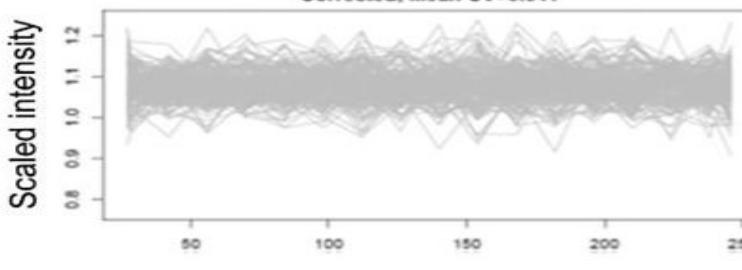
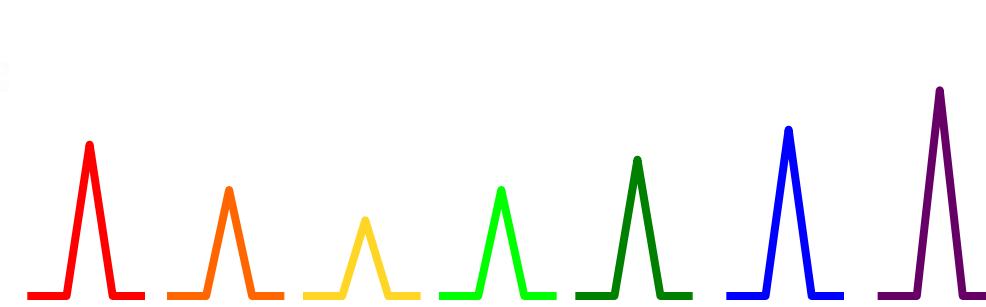
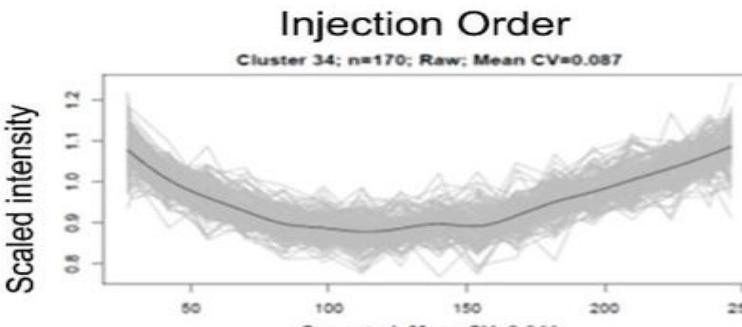
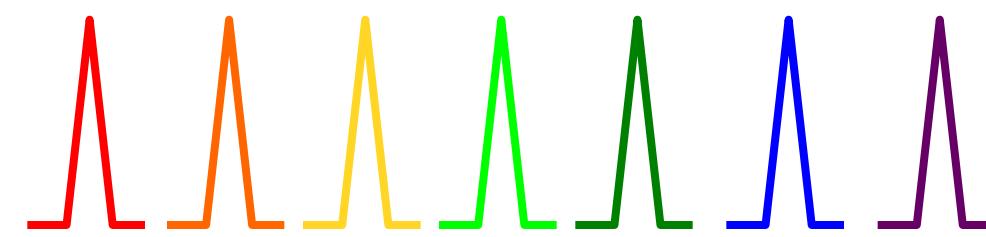
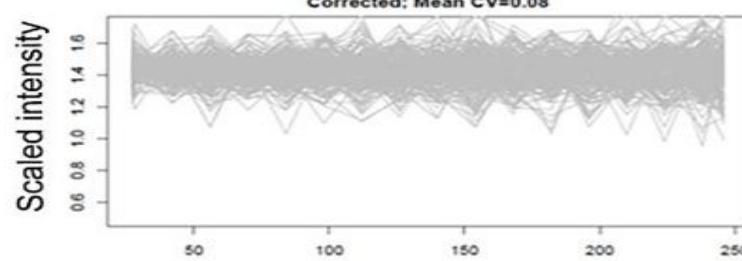
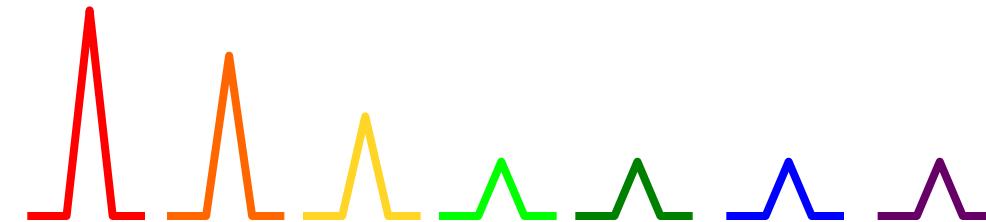
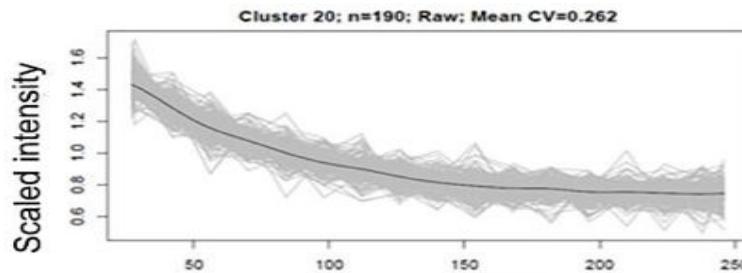


What you want

A Real Example (Alignment)



A Real Example (Scaling)



Injection Order

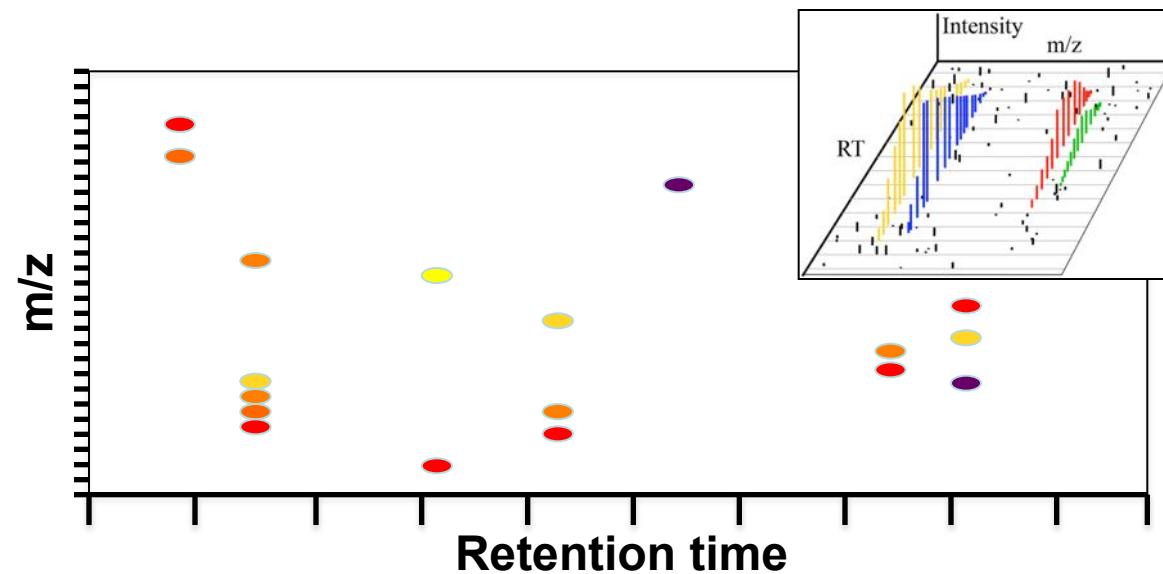
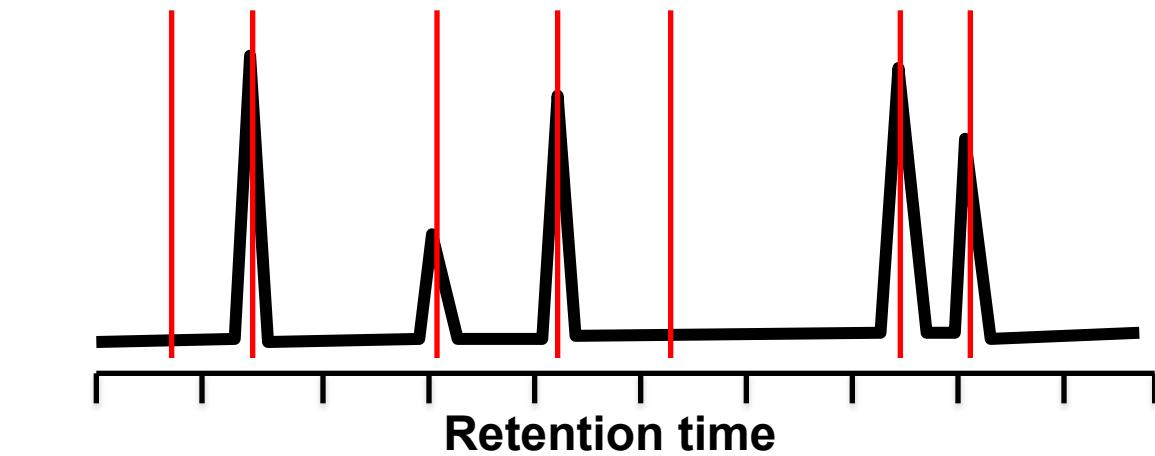
HPLC-MS (Ideal)



HPLC



MS



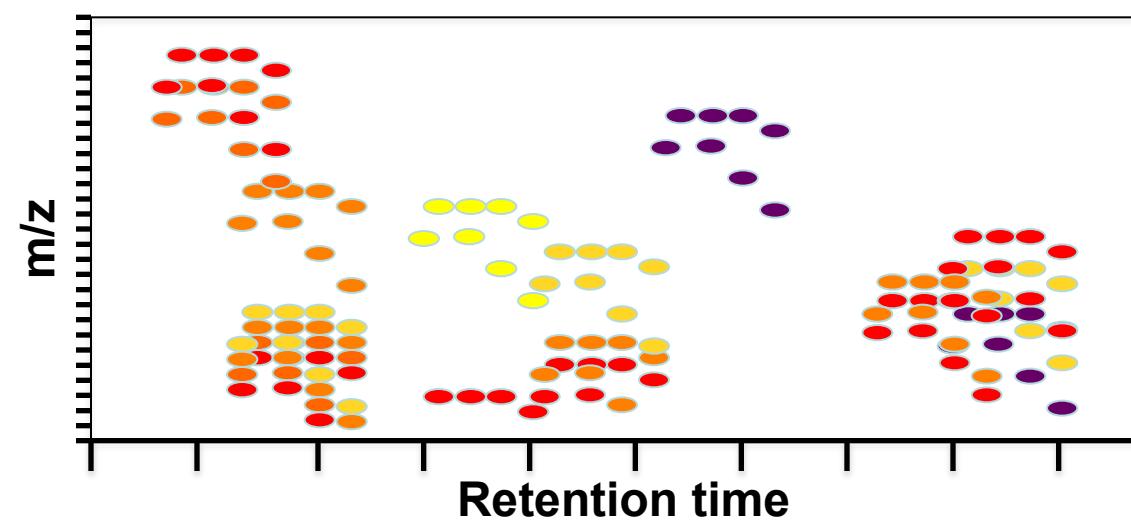
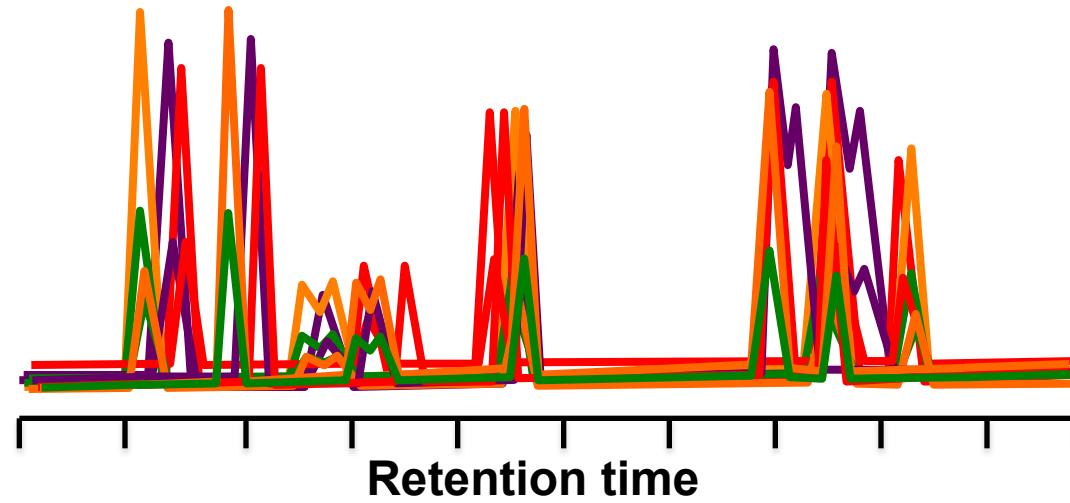
HPLC-MS (Batch Effects)



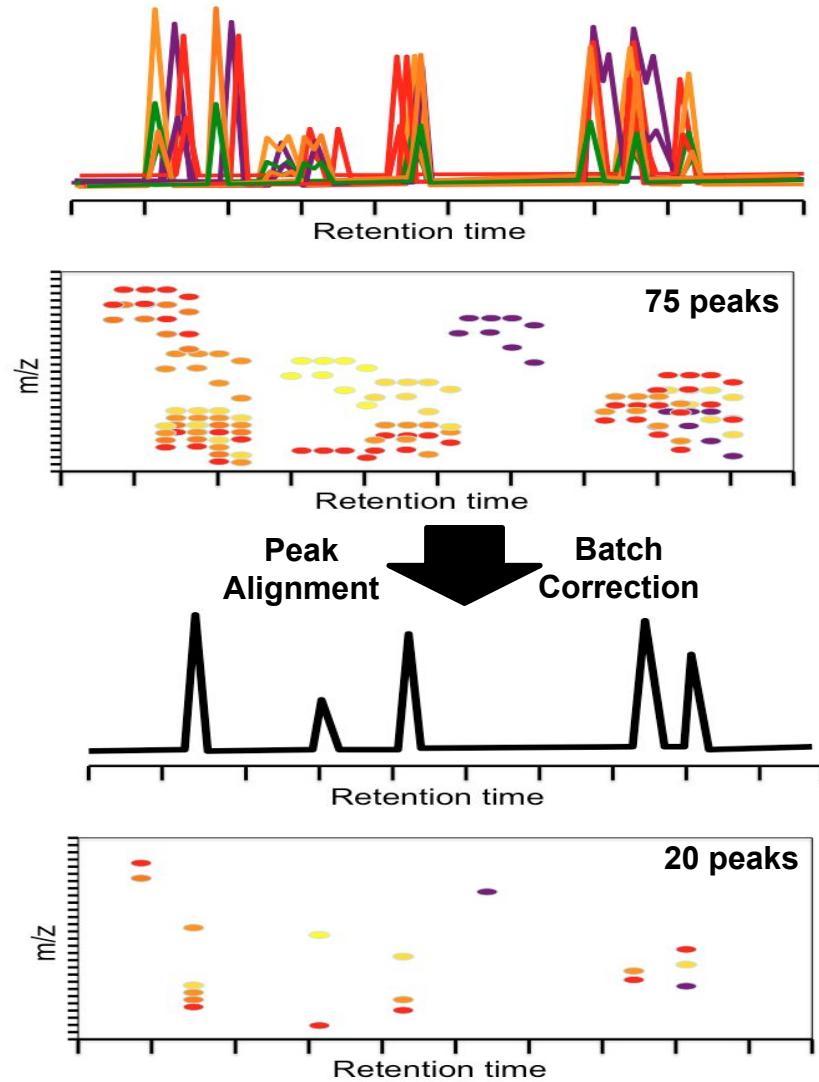
HPLC



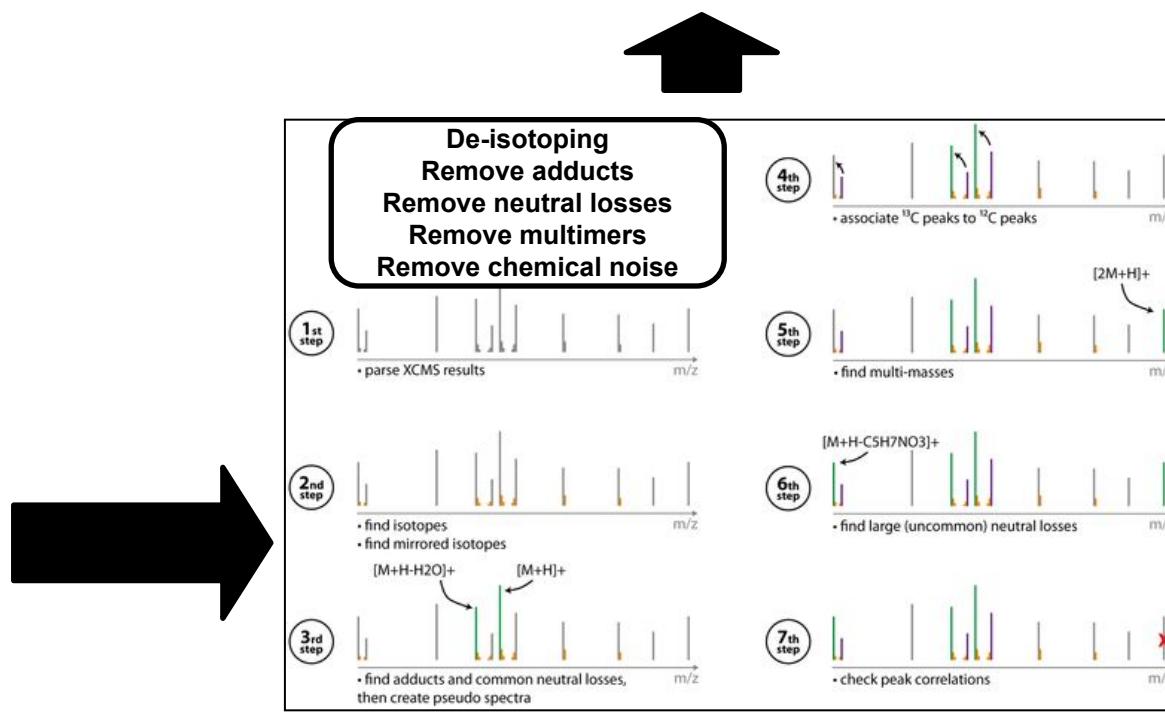
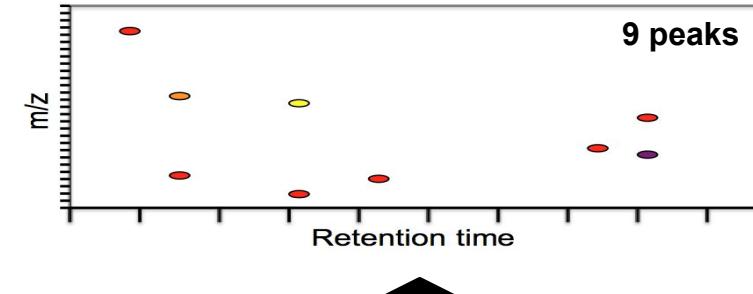
MS



Untargeted LC-MS Data Processing



Final RT vs m/z Feature List



Peak Matching & Retention Time Correction

- Groups the lists of ions into a single matrix of ion intensities across all samples
- Steps:
 1. Match peaks across samples based on RT
 2. Correct RT drift using peak groups
 3. Re-do alignment and update groups
 4. Steps 1-3 iteratively until no more changes are evident

Peak Matching & Retention Time Correction

Sample 1

m/z	rt	int
389.1235	31.49	562472
126.7911	51.98	451145
427.9103	78.12	3851039

Sample 2

m/z	rt	int
389.1235	31.49	562479
102.1157	51.99	460109
126.7905	20.18	572901

Sample 3

m/z	rt	int
389.1235	31.49	562482
126.7912	51.97	240510
102.1159	52.00	1049100

Group independent ion lists by m/z

m/z	rt	int
389.1235	31.49	562472
126.7911	51.98	451145
427.9103	78.12	3851039

m/z	rt	int
389.1235	31.49	562479
102.1157	51.99	460109
126.7905	20.18	572901

m/z	rt	int
389.1235	31.49	562482
126.7912	51.97	240510
102.1159	52.00	1049100

Peak Matching & Retention Time Correction

Sample 1			Sample 2			Sample 3		
m/z	rt	int	m/z	rt	int	m/z	rt	int
389.1235	31.49	562472	389.1235	31.49	562479	389.1235	31.49	562482
126.7911	51.98	451145	102.1157	51.99	460109	126.7912	51.97	240510
427.9103	78.12	3851039	126.7905	20.18	572901	102.1159	52.00	1049100

↓

Group ions by RT

↓

m/z	rt	int
389.1235	31.49	562472
126.7911	51.98	451145
427.9103	78.12	3851039

m/z	rt	int
389.1235	31.49	562479
102.1157	51.99	460109
126.7905	20.18	572901

m/z	rt	int
389.1235	31.49	562482
126.7912	51.97	240510
102.1159	52.00	1049100

Peak Matching & Retention Time Correction

m/z	rt	int
389.1235	31.49	562472
126.7911	51.98	451145
427.9103	78.12	3851039

m/z	rt	int
389.1235	31.49	562479
102.1157	51.99	460109
126.7905	20.18	572901

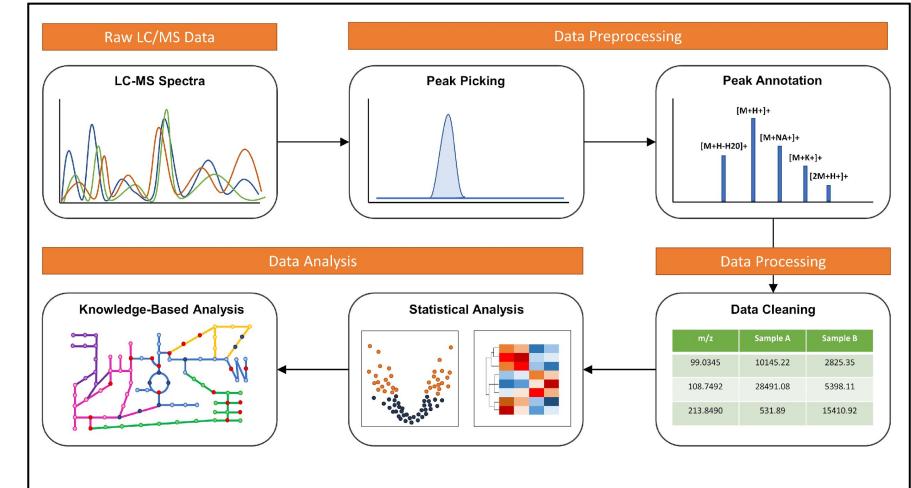
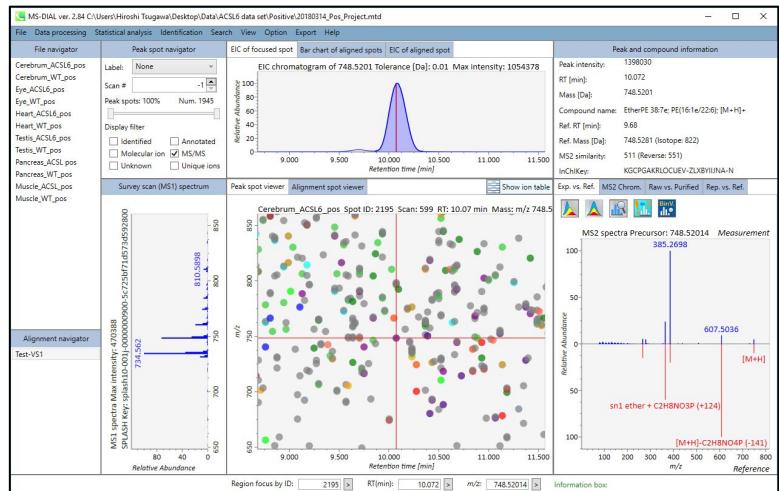
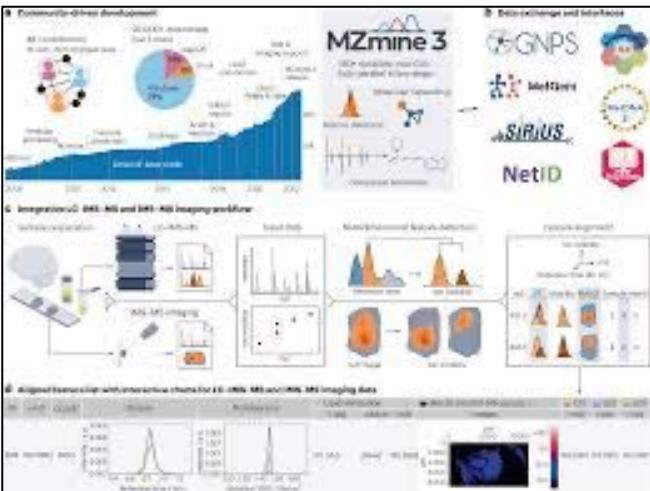
m/z	rt	int
389.1235	31.49	562482
126.7912	51.97	240510
102.1159	52.00	1049100

Final matrix

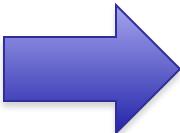
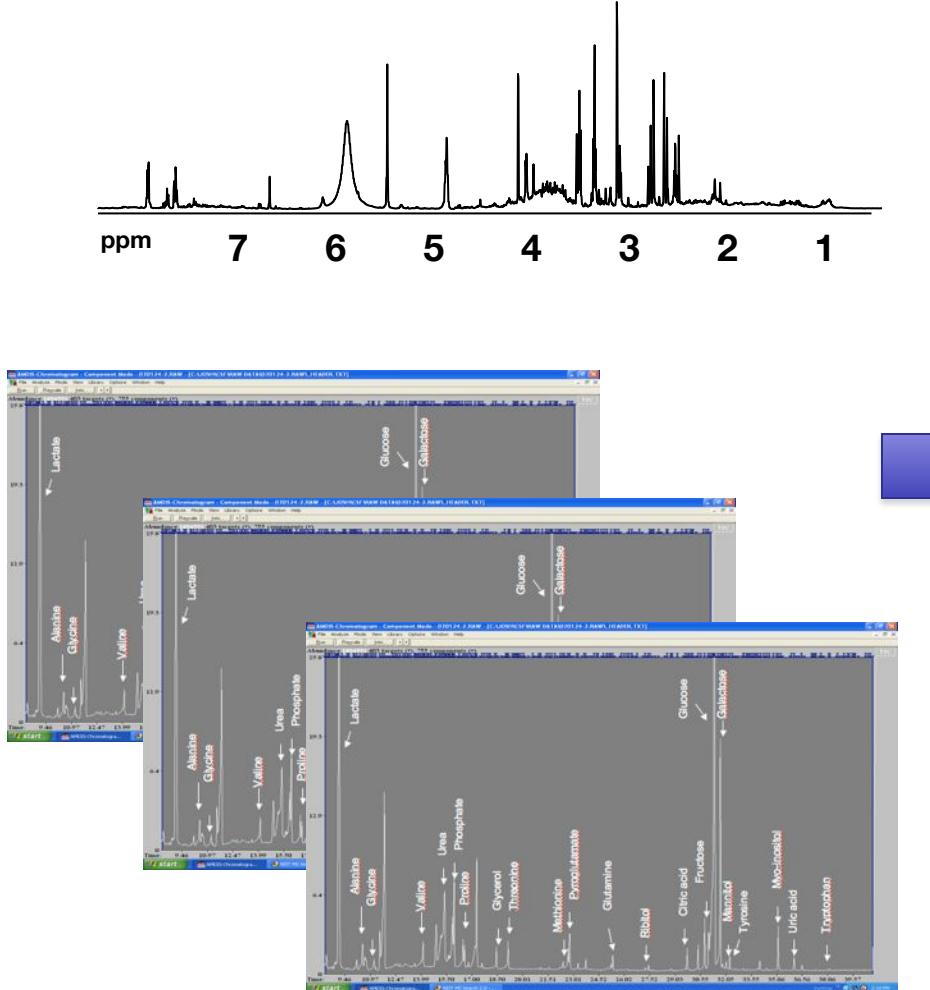
m/z	rt	Sample1	Sample2	Sample3
389.1235	31.49	562472	562479	562482
102.1157	51.99		460109	1049100
126.7911	51.98	451145		240510
427.9103	78.12	3851039		
126.7905	20.18		572901	

Feature Simplification

- Raw +ve mode spectrum
- Remove adducts
- Remove multiple charges
- Remove neutral losses
- Remove isotope peaks
- Remove noise peaks
- Final spectrum
- Repeat for -ve mode
- 15,000 features
- 12,000 features
- 10,000 features
- 8,000 features
- 3,000 features
- 2,500 features
- 2,500 M+H peaks
- 1,500 M-H peaks



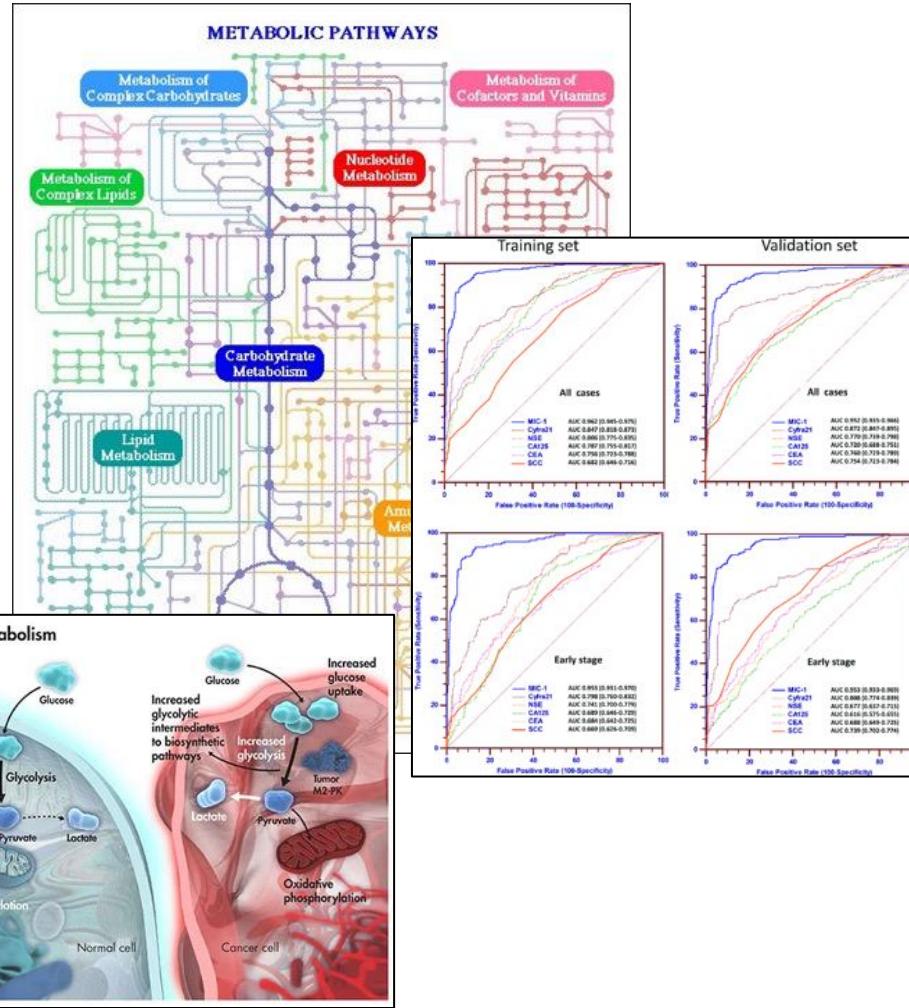
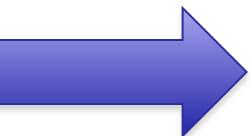
The Goal for Both Targeted & Untargeted: From Spectra to Lists



Compound	Retention Time (min)	Conc. in Urine (μM)	Compound	Retention Time (min)	Conc. in Urine (μM)
Dns-o-phospho-L-serine	0.92	<DL*	Dns-Ile	6.35	25
Dns-o-phospho-L-tyrosine	0.95	<DL	Dns-3-aminosalicylic acid	6.44	0.5
Dns-adenosine monophosphate	0.99	<DL	Dns-pipeolic acid	6.50	0.5
Dns-o-phosphoethanolamine	1.06	16	Dns-Leu	6.54	54
Dns-glucosamine	1.06	22	Dns-cystathione	6.64	0.3
Dns-o-phospho-L-threonine	1.09	<DL	Dns-Leu-Pro	6.60	0.4
Dns-6-dimethylamino purine	1.20	<DL	Dns-5-hydroxylysine	6.65	1.6
Dns-3-methyl-histidine	1.22	80	Dns-Cysteine	6.73	160
Dns-taurine	1.25	834	Dns-N-norleucine	6.81	0.1
Dns-carnosine	1.34	28	Dns-5-hydroxydopamine	7.17	<DL
Dns-Arg	1.53	36	Dns-dimethylamine	7.33	293
Dns-Asn	1.55	133	Dns-5-HIAA	7.46	18
Dns-hypotaurine	1.58	10	Dns-umbelliferone	7.47	1.9
Dns-homocarnosine	1.61	3.9	Dns-2,3-diaminopropionic acid	7.63	<DL
Dns-guanidine	1.62	<DL	Dns-L-ornithine	7.70	15
Dns-Gln	1.72	633	Dns-4-acetylamidophenol	7.73	51
Dns-allantoin	1.83	3.8	Dns-procaine	7.73	8.9
Dns-L-citrulline	1.87	2.9	Dns-homocystine	7.76	3.3
Dns-1-(or 3-)methylhistidine	1.94	1.9	Dns-acetaminophen	7.97	82
Dns-adenosine	2.06	2.6	Dns-Phe-Phe	8.03	0.4
Dns-methylguanidine	2.20	<DL	Dns-5-methoxy salicylic acid	8.04	2.1
Dns-Ser	2.24	511	Dns-Lys	8.16	184
Dns-aspartic acid amide	2.44	26	Dns-aniline	8.17	<DL
Dns-4-hydroxy-proline	2.56	2.3	Dns-leu-Phe	8.22	0.3
Dns-Glu	2.57	21	Dns-His	8.35	1560
Dns-Asp	2.60	90	Dns-4-thiolsine	8.37	<DL
Dns-Thr	3.03	157	Dns-benzylamine	8.38	<DL
Dns-epinephrine	3.05	<DL	Dns-1-ephedrine	8.50	0.6
Dns-ethanolamine	3.11	471	Dns-tryptamine	8.63	0.4
Dns-aminoadipic acid	3.17	70	Dns-pyridoxamine	8.94	<DL
Dns-Gly	3.43	2510	Dns-2-methyl-benzylamine	9.24	<DL
Dns-Ala	3.68	593	Dns-5-hydroxytryptophan	9.25	0.12
Dns-aminolevulinic acid	3.97	30	Dns-13-diaminopropane	9.44	0.23
Dns- <i>l</i> -amino-butyric acid	3.98	4.6	Dns-putrescine	9.60	0.5
Dns- <i>d</i> -amino-hippuric acid	3.98	2.9	Dns-12-diaminopropane	9.66	0.1
Dns-5-hydroxymethyluric acid	4.58	1.9	Dns-tyrosinamide	9.79	29
Dns-tryptophanide	4.70	5.5	Dns-dopamine	10.08	140
Dns-isoguanine	4.75	<DL	Dns-cadaverine	10.08	0.08
Dns-5-aminoenthanoic acid	4.79	1.6	Dns-histamine	10.19	0.4
Dns-sarcosine	4.81	7.2	Dns-3-methoxy-tyramine	10.19	9.2
Dns-3-amino-isobutyrate	4.81	85	Dns-Tyr	10.28	321
Dns-2-aminobutyric acid	4.91	17	Dns-cysteamine	10.44	<DL

From Lists to Biology

Compound	Retention Time (min)	Conc. in Urine (μM)	Compound	Retention Time (min)	Conc. in Urine (μM)
Dns-o-phospho -L-serine	0.92	<DL*	Dns-Lle	6.35	25
Dns-o-phospho -L-tyrosine	0.95	<DL	Dns-3-aminosalicylic acid	6.44	0.5
Dns-adenosine monophosphate	0.99	<DL	Dns-pipeolic acid	6.50	0.5
Dns-o-phosphoethanolamine	1.06	16	Dns-Leu	6.54	54
Dns-glucosamine	1.06	22	Dns-cystathione	6.54	0.3
Dns-o-phospho -L-threonine	1.09	<DL	Dns-Leu-Pro	6.60	0.4
Dns-6-dimethyllysine	1.20	<DL	Dns-5-hydroxylysine	6.65	1.6
Dns-3-methyl -histidine	1.22	80	Dns-Cysteine	6.73	160
Dns-taurine	1.25	834	Dns-N-norleucine	6.81	0.1
Dns-carnosine	1.34	28	Dns-5-hydroxydopamine	7.17	<DL
Dns-Arg	1.53	36	Dns-dimethylamine	7.33	293
Dns-Asn	1.55	133	Dns-5-HIAA	7.46	18
Dns-hypotaurine	1.58	10	Dns-umbelliferone	7.47	1.9
Dns-homocarnosine	1.61	3.9	Dns-2,3-diaminopropionic acid	7.63	<DL
Dns-guanidine	1.62	<DL	Dns-L-ornithine	7.70	15
Dns-Gln	1.72	633	Dns-4-acetylamidophenol	7.73	51
Dns-allantoin	1.83	3.8	Dns-procaine	7.73	8.9
Dns-L-citrulline	1.87	2.9	Dns-homocystine	7.76	3.3
Dns-1 (or 3-)methylhistamine	1.94	1.9	Dns-acetaminophen	7.97	82
Dns-adenosine	2.06	2.6	Dns-Phe-Phe	8.03	0.4
Dns-methylguanidine	2.20	<DL	Dns-5-methoxy salicylic acid	8.04	21
Dns-Ser	2.24	511	Dns-Lys	8.16	184
Dns-aspartic acid amide	2.44	26	Dns-aniline	8.17	<DL
Dns-4-hydroxy -proline	2.56	2.3	Dns-Ieu-Phe	8.22	0.3
Dns-Glu	2.57	21	Dns-His	8.35	1550
Dns-Asp	2.60	90	Dns-4-thialysine	8.37	<DL
Dns-Thr	3.03	157	Dns-benzylamine	8.38	<DL
Dns-epinephrine	3.05	<DL	Dns-1-ephedrine	8.50	0.6
Dns-ethanolamine	3.11	471	Dns-tryptamine	8.63	0.4
Dns-aminoacidic acid	3.17	70	Dns-pyridoxamine	8.94	<DL
Dns-Gly	3.43	2510	Dns-2-methyl -benzylamine	9.24	<DL
Dns-Ala	3.88	593	Dns-5-hydroxytryptophan	9.25	0.12
Dns-aminolevulinic acid	3.97	30	Dns-1,3-diaminopropane	9.44	0.23
Dns- <i>l</i> -amino-butyric acid	3.98	4.6	Dns-putrescine	9.60	0.5
Dns- <i>d</i> -amino-hippuric acid	3.98	2.9	Dns-1,2-diaminopropane	9.66	0.1
Dns-5-hydroxymethyluric acid	4.58	1.9	Dns-tyrosinamide	9.79	29
Dns-tryptophanamide	4.70	5.5	Dns-dopamine	10.08	140
Dns-isoguanine	4.75	<DL	Dns-cadaverine	10.08	0.08
Dns-5-aminopentanoic acid	4.79	1.6	Dns-histamine	10.19	0.4
Dns-sarcosine	4.81	7.2	Dns-3-methoxy -tyramine	10.19	9.2
Dns-3-amino -isobutyrate	4.81	85	Dns-Tyr	10.28	321
Dns-2-aminobutyric acid	4.91	17	Dns-cysteamine	10.44	<DL



We are on a Coffee/Lunch Break & Networking Session

Workshop Sponsors:



Canadian Centre for
Computational
Genomics



HPC4Health

