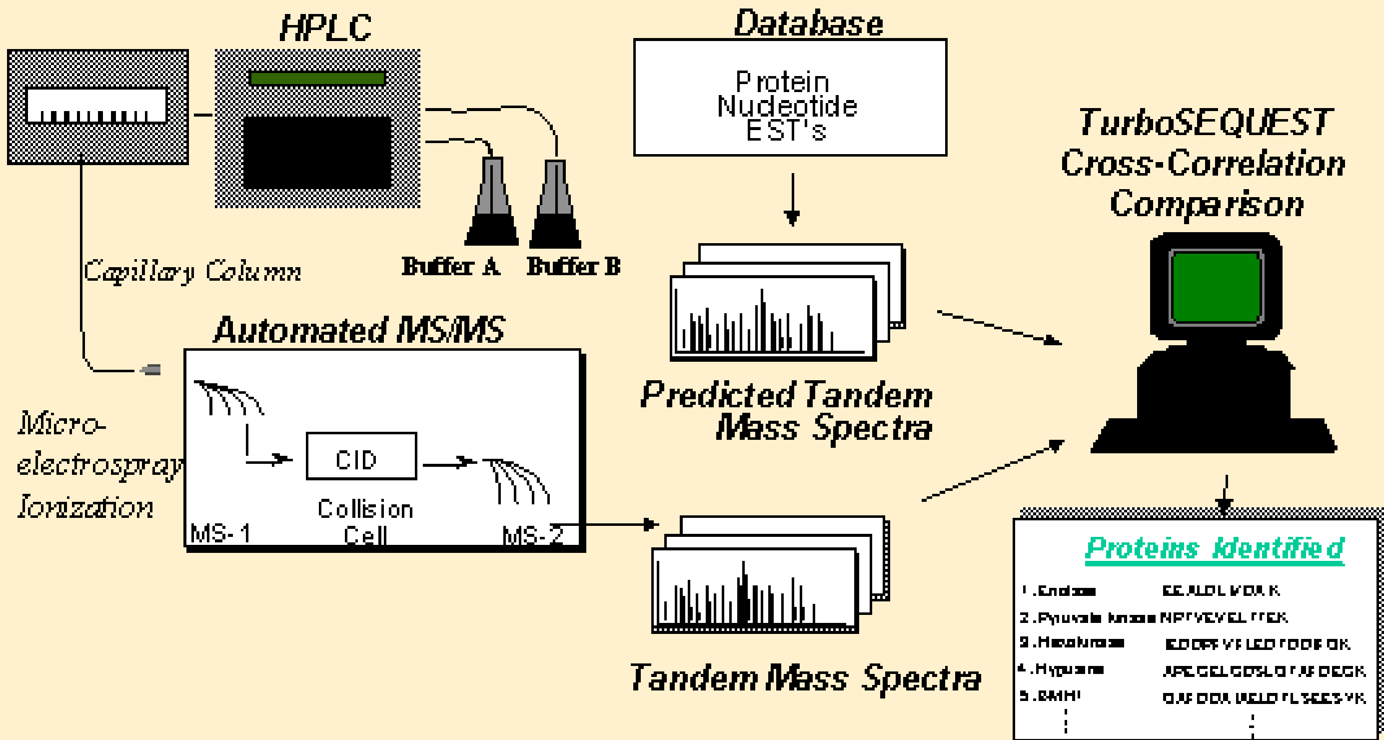
BIOC4004 - Industrial Biochemistry
Lecture 20 - Wed Mar 17, 04

Topics for the Day:

- Proteomics
  - Mass Spectrometry
  - Sequencing using Mass Spec
  - FTICR-MS

MS analysis of samples can be automated



Major bottlenecks in proteomics:

- 2-D gel analysis & spot excision
- analysis of MS spectra

Solutions:

- automating gel analysis & spot picking robots
- move away from 2-D gel electrophoresis
    - use liquid chromatography (LC)
    - use capillary electrophoresis (CE)
- integrated separation / identification systems (ie. couple separation with MS)
- more sophisticated bioinformatics
    - more complete databases
    - better fingerprint matching

MS for dummies (or "how I learned to interpret MS data")

Information Obtained From a Mass Spectrum

1) Ion Identification: The spectrometer can identify ions based upon their ratio of mass to charge (or *m/z* ratio). Molecules can have inherent charge or can be charged as a result of molecule fragmentation.

• ESI is used for smaller molecules. At low voltages ESI can be a soft ionization method. If voltage is increased, you can get molecule fragmentation

• MALDI is used for large, thermo-labile, organic samples (such as protein, peptides, sugars, nucleic acid oligos) and generally leads to fragmentation.

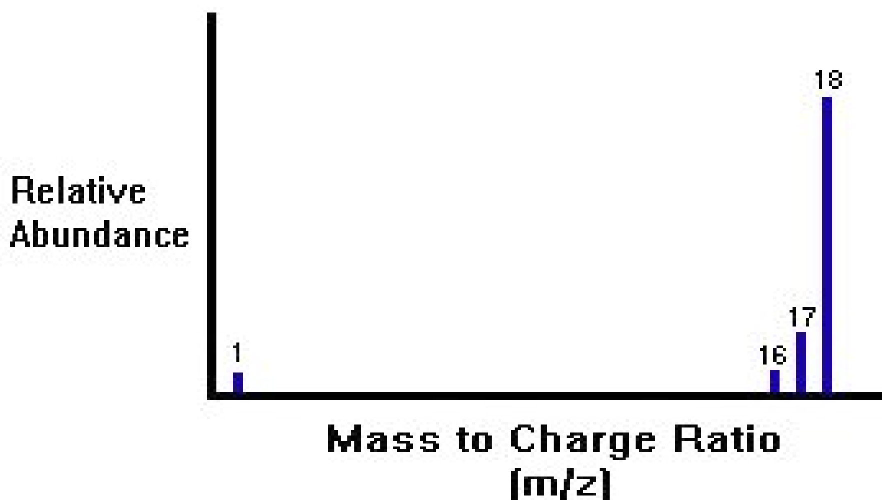
For proteins:

- trypsin digestion is used to digest the protein into smaller peptides which are small enough to “fly” (ie. intact proteins too heavy to move)
- molecules can be “charged”
    - positive mode: add formic acid, helps with protonation
    - negative mode: add ammonia, helps with deprotonation
    - proteins are run generally in positive mode

2) Fragmentation of Ions: In the process of creating gas phase ions, the whole, intact molecule (parent) may fragment into a variety of smaller pieces. These fragments can also be charged and thus be detected at a different mass to charge ratio than their parental precursor.

3) Structure of Parent: The total pattern of ions detected in a mass spectrum can reveal information about the composition and structure of the parent molecule.

MS for dummies (pt I)



The mass spectrum of water would be similar to the spectrum shown here. Each peak of the spectrum represents the m/z of a unique ionic species. To illustrate this, compare the spectrum with the following information:

m/z =18 $(H_20)^+$ since O + H + H = 16 + 1 + 1 = 18 amu

m/z=17 $(HO)^+$ since O + H = 16 + 1 = 17 amu

m/z=16 $(O)^+$ since O = 16 amu

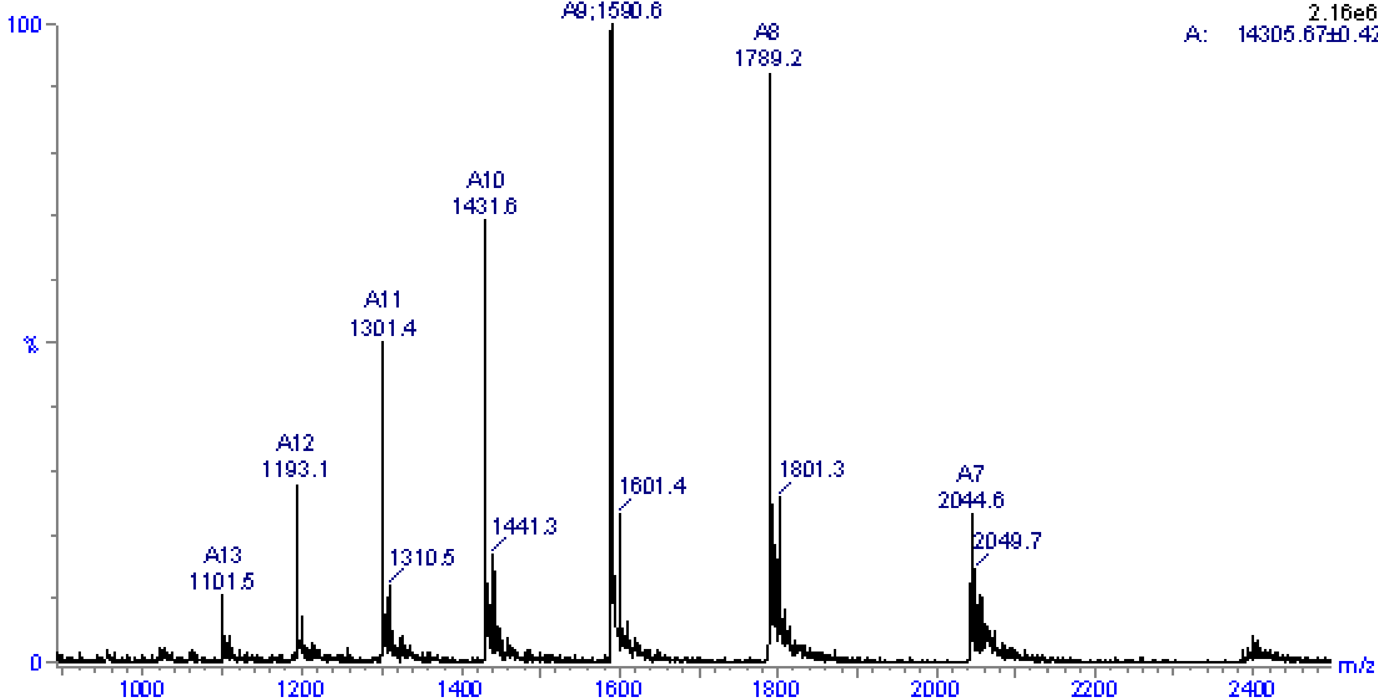
m/z=1 $(H)^+$ since H = 1 amu

- can't forget about isotope replacements in the molecule (would be trace peaks because of lower abundance)

Steps for Revealing Mass Spectral Information
Some general ways to reveal information in a mass spectrum include:

(1) Determine the exact mass of a peak to reveal its elemental composition.
(2) Assign the heaviest m/z as the parent molecule.
(3) Check for isotopic abundances in a greater magnification of the peak to to reveal the presence of isotopic elements.
(4) Record the fragmentation pattern to "fingerprint" the molecule and to reveal details of its structure.

MS for dummies (pt II)



Why do we see many different M/Z for a given peptide ?

- Proteins have many suitable sites for protonation
  - backbone amide nitrogen atoms could be protonated
  - certain amino acid side chains such as lysine and arginine which contain primary amine functionalities

$$m/z = \frac{(MW + nH^{+})}{n}$$

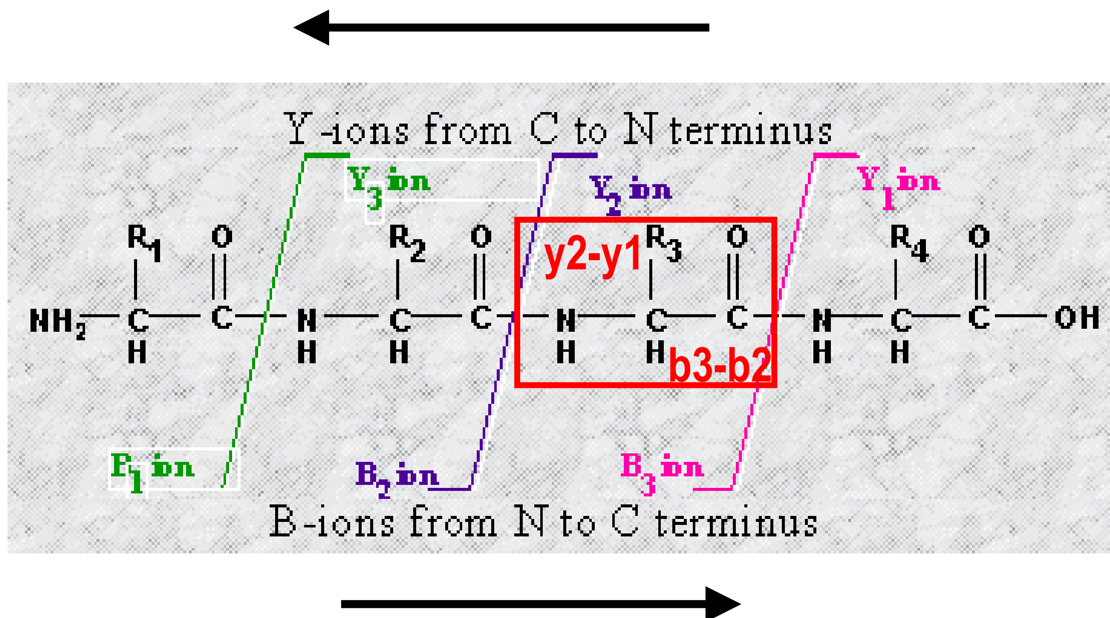
MW = the molecular weight of the sample
n = the integer number of charges on the ions
H = the mass of a proton = 1.008 Da.

- Assume each peak has n+1 charge relative to the previous peak
- can solve the equation using this assumption
- The MW estimated from each peak will converge to same answer

Protein Sequencing Using MS-MS (pt.I)

• MS relies on detection of charged molecules
• molecules can have inherent charge or can result from molecule fragmentation
• **Ion traps** (Quadropole ion traps) can be used to keep certain ions for a second round of fragmentation and subsequent MS analysis of the resulting ions (tandem MS).
• Carry out **precursor ion scan**: look for ions that generate signal
• Isolate an individual ion and break it by collision with small gas molecules
  • **CID**: collision induced dissociation
  • peptides always break at the peptide bonds
  • generate ladder of fragments

• the protein sequence data is generated like DNA sequence
  • separation of ions based on mass
  • subtracting the mass of a ion X from the next largest ion (X + 1) tells us the mass of the extra amino acid on (X + 1)

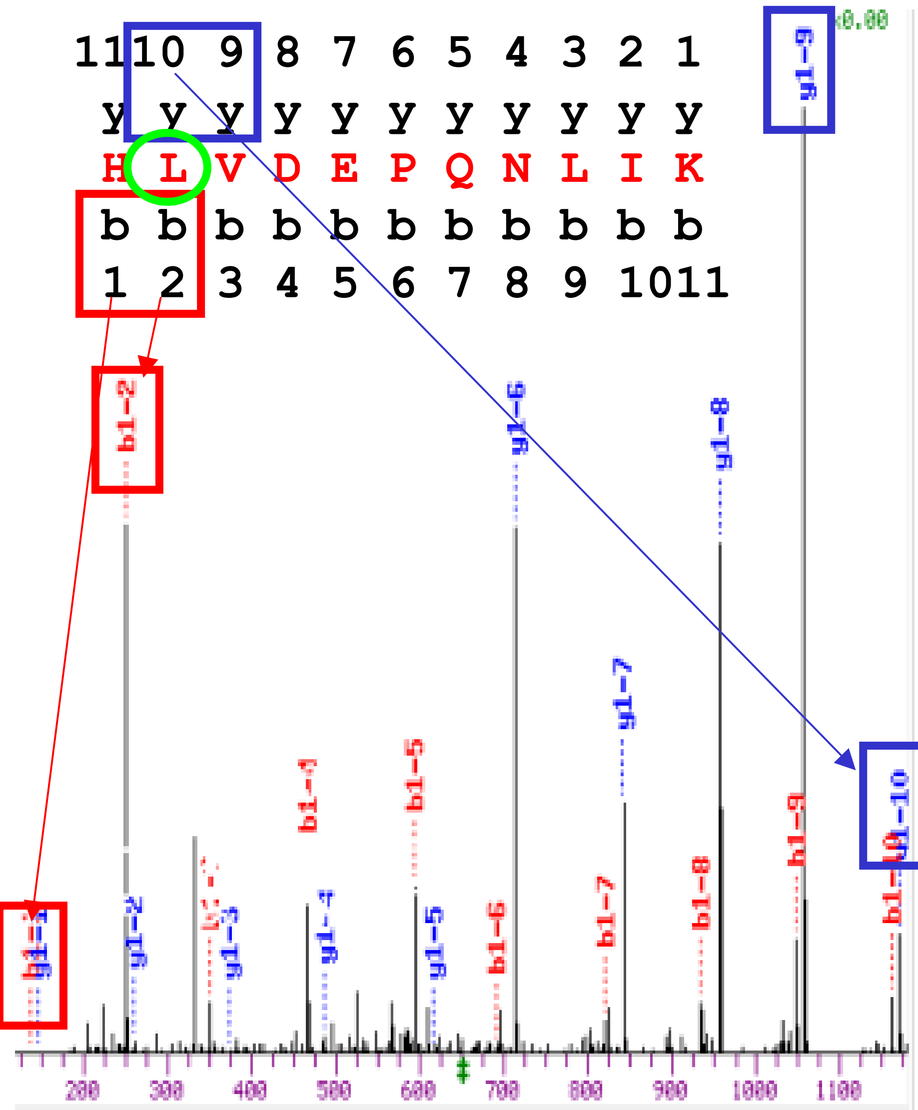


• We can use this information to derive peptide sequence information
  • peptide sequence data is better than peptide fingerprint data

Protein Sequencing Using MS-MS (pt.II)

Seq	#	b	y	(+1)
H	1	138.1	1305.7	11
L	2	251.1	1168.6	10
V	3	350.2	1055.6	9
D	4	465.2	956.5	8
E	5	594.3	841.5	7
P	6	691.3	712.4	6
Q	7	819.4	615.4	5
N	8	933.4	487.3	4
L	9	1046.5	373.3	3
I	10	1159.6	260.2	2
K	11	1287.7	147.1	1

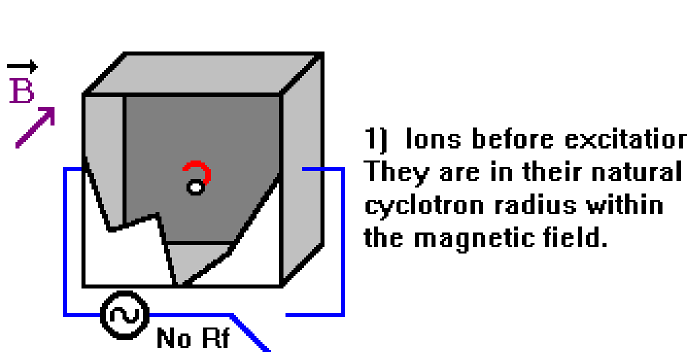
Seq	#	b	y	(+2)
H	1	69.5	653.4	11
L	2	126.1	584.8	10
V	3	175.6	528.3	9
D	4	233.1	478.7	8
E	5	297.6	421.2	7
P	6	346.2	356.7	6
Q	7	410.2	308.2	5
N	8	467.2	244.2	4
L	9	523.8	187.1	3
I	10	580.3	130.6	2
K	11	644.4	74.1	1

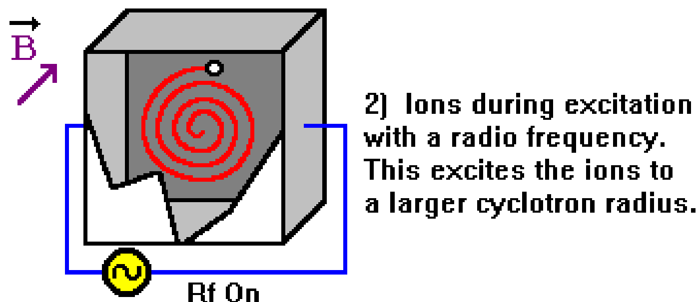


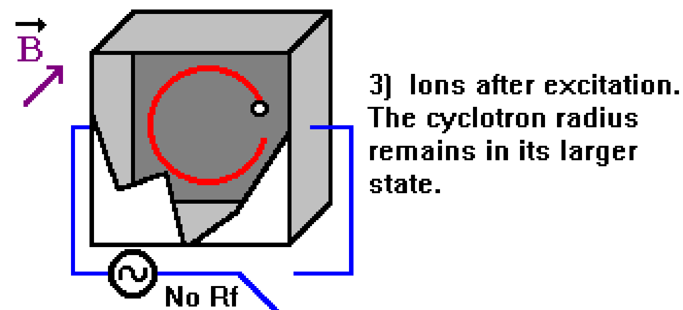
b2-b1	=	y10-y9	= 113 =	MW Leu !!!
251.1-138.1	=	1168.6-1055.6	= 113 =	MW Leu !!!

Down the pipeline: FTICR-MS (Part I)

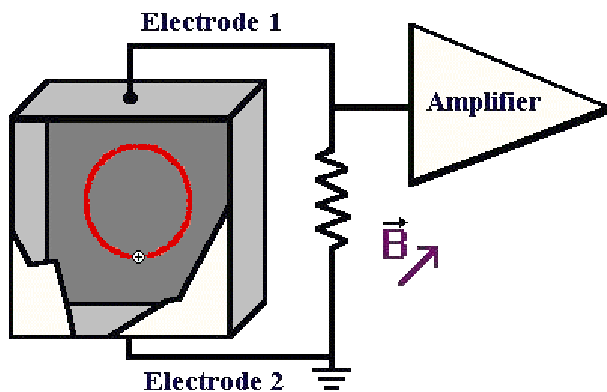
- FTICR (Fourier Transform Ion Cyclotron Resonance)
  - ions in static magnetic field move in circles of given radius at a given rate
    - **cyclotron frequency**
  - frequency of the cyclotron gyration of an ion is inversely proportional to its mass-to-charge ratio (*m/z*)
  - Ions can be excited by radio frequency (Rf) of the same wavelength as cyclotron frequency --> **acceleration, larger radius but same cyclotron frequency**







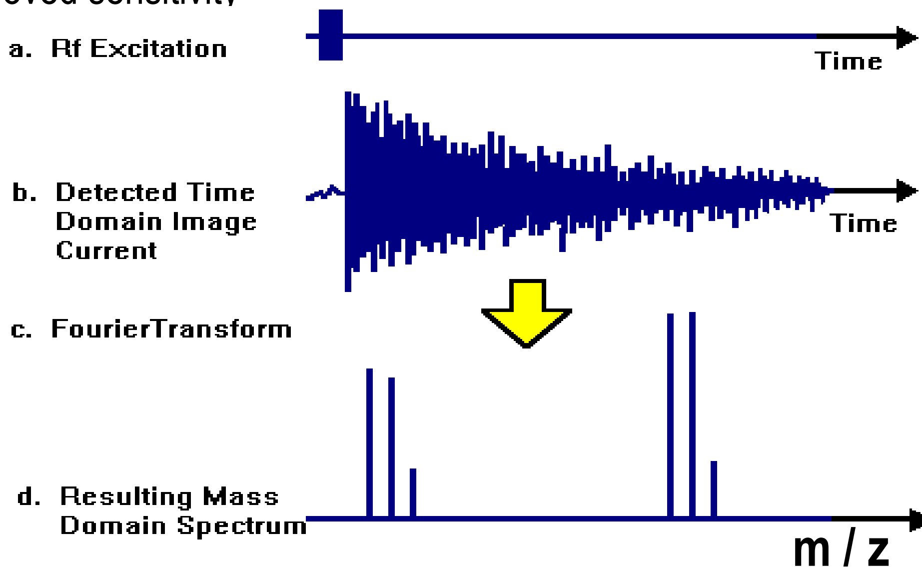
When the radio frequency (Rf) signal is terminated, the accelerated ions continue to gyrate at a constant radius. This phenomenon provides the basis for **ion cyclotron resonance** mass spectrometry because ions having a different cyclotron frequency (ie. ions with different m/z) are not accelerated.



Detection: As the ion(s) in a circular orbit approach the top plate, electrons are attracted to this plate from ground. Then as the ion(s) circulate towards the bottom plate, the electrons travel back down to the bottom plate. This motion of electrons moving back and forth between the two plates produces a detectable current.

Down the pipeline: FTICR-MS (Part II)

- Conventional MS: destructive detection of ions in a "race towards the detector"
  - detect collision between ion and electrode
- FTICR detects the current generated by ion (ie.non destructive), signal can be amplified
  - improved sensitivity



- As the electrical signal decays, a spectrum as a function of *m* or *m/z* can be obtained using "math". Cyclotron frequencies can be measured with very high precision
  - high accuracy mass measurements (ie. ultra-high resolving power)

FTICR (Fourier Transform Ion Cyclotron Resonance) MS has:

- insane sensitivity (100 X more than current tandem MS)
- insane resolution (can get down to less than 1 ppm error)
- insane speed (10-100 times faster analysis)

- We can start using MS to analyze real proteomes, not single proteins spots from gels or single LC peaks
- Near single-cell sensitivity
- Aug 2002: identified 1,900 proteins from *Deinococcus radiodurans*
  - broadest proteomics analysis to date

The catch:

- $1,5 Million USD
- no one makes them commercially
- huge amounts of data generated…require huge informatics infrastructure

Metabolomics

• same as proteomics, except you are looking at small molecules, not proteins:

The term '**metabolome**' refers to the entire complement of all the small molecular weight metabolites inside a cell suspension (or other sample) of interest.

Why ???

• It is likely that the measurement of the metabolome in different physiological states will be much more discriminating for the purposes of functional genomics.

- Changes in individual enzyme concentrations can have substantial effects on metabolite concentrations
- These changes reflect what's going on inside the cell

The Problems

• using conventional mass spec, mass accuracies are a big problem because the margin of error can give ambiguous identifications:

eg. $M/z = 16 \pm 1$ is it heavy water or just regular water ???

• also:

- need **lots** of starting material
- need to restrict yourself to a few metabolites
- for comprehensive metabolomics, need much more power than conventional MS can deliver……..

FTICR (Fourier Transform Ion Cyclotron Resonance) MS has:

- insane sensitivity (100 X more than current tandem MS)
- insane resolution (can get down to less than 1 ppm error)
- insane speed (10-100 times faster analysis)

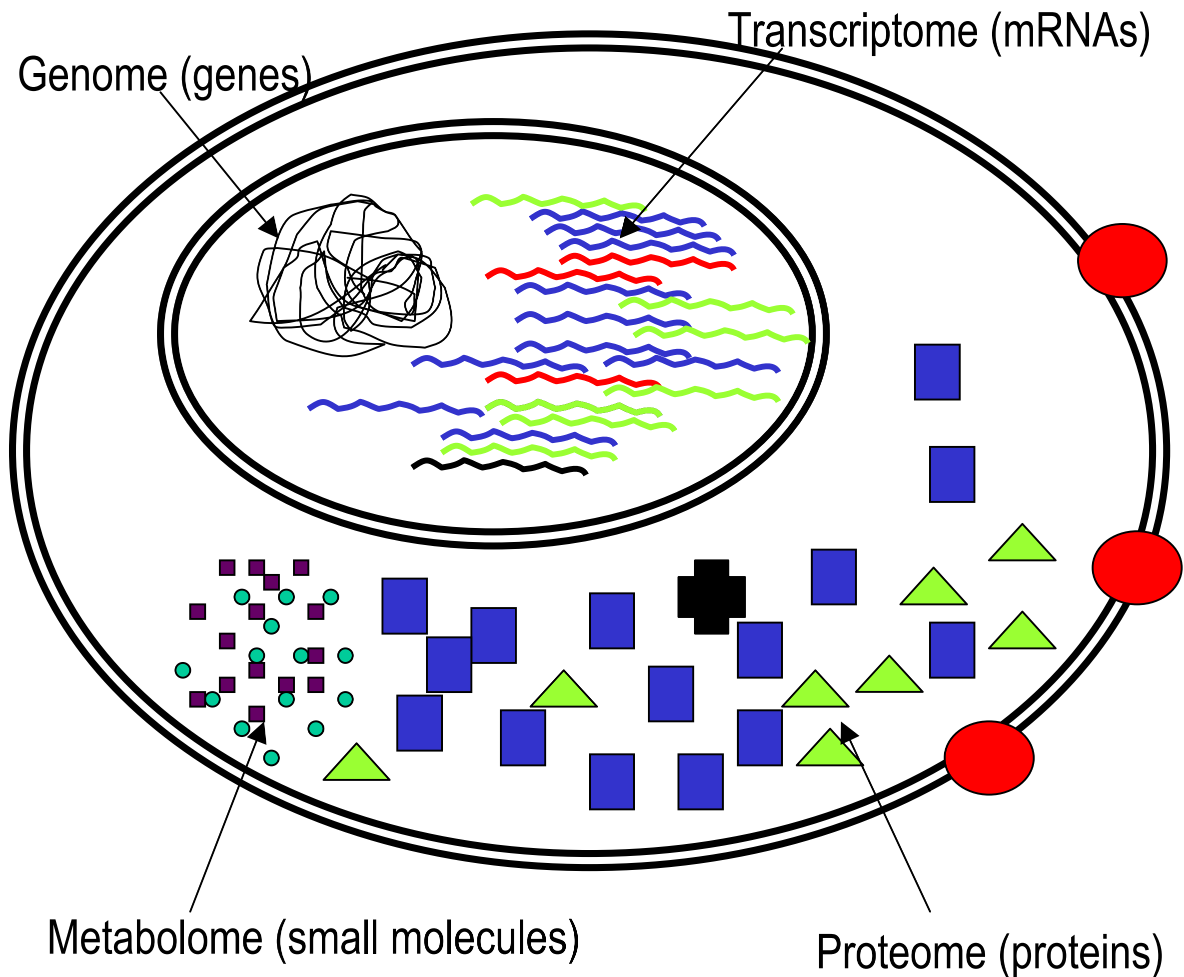
• We can start using MS to analyze **METABOLOMES**

• Near single-cell sensitivity

The catch:

- $1,5 Million USD
- no one makes them commercially
- huge amounts of data generated…require huge informatics infrastructure

Transcriptome vs Proteome vs Metabolome (Pt I)



- We're always interested in comparing a test sample to a control sample
- We're interested in determining if there are differences in the repertoire and relative abundance of different analytes in different types of samples

DNA, RNA, Protein and Small Molecules are all a part of life processes, but which level of study is more relevant ?

- DNA is the most indirect because it measures "potential" for gene expression
- RNA is a direct measurement of gene expression but an indirect measure of life processes
- Protein and metabolite profiling are more direct measures of life processes because they measure the "effectors"

DNA < RNA < proteins < metabolites ???

Transcriptome vs Proteome vs Metabolome (Pt II)

"Biological Relevance"

DNA < RNA < Proteins, Metabolites

"Accessibility"

DNA > RNA > proteins > metabolites

	The GOOD	The BAD
Genome	• determines total potential of organism • "easy" to obtain (bioinformatics, automation • a finite problem • "stable"	• only determines "potential" • unknown genes • nothing about what's going on "right now"
Transcriptome	• determines gene activity • gene activity ~ phenotype • data is "ordered" (ie. arrays) • every data point tested is known • "all RNA behaves similarly"	• gene activity ~ phenotype • RNA vs protein levels do not always correlate • post-transcriptional events ? • Limit of detection ?
Proteome	• determines what proteins are present • good proteome/phenotype correlation • can see post-transcriptional events • can see post-translational events • protein localization	• proteome is very variable • many proteins are never detected • bioinformatics complexity (modifications, splicing…) • expensive equipment • proteins behave differently
Metabolome	• determines metabolic processes going on • metabolites "drive life" • small changes in RNA/protein can give large differences in metabolites • every phenotypic difference due to metabolic changes	• requires super expensive equipment and infrastructure • exceedingly complex bioinformatics • interconnected pathways