Mass Spectrometry Based Proteomics



Mass Spectrometry Based Proteomics

This course is concerned with.....

Protein primary structure sequencing

Post translational modifications

Bottom up and Top down proteomics



<u>Individual Protein</u> or <u>Complex Proteome</u> analysis for Identification and Characterization of <u>Overexpressed Proteins</u> or <u>Disease States</u>







Other Systems Biology Mass Spectrometry The 'omics buzz'

Transcriptomics

Proteomics

Lipidomics

Glycomics

Metabolomics

Interactomics

Fluxomics (dynamic changes over time)

Mass Spectrometry Based Proteomics

- I. Mass Spectrometers The Basics
- **II. Biological Mass Spectrometry**
- **III. Multidimensional Separations**
- **IV.** Post-translational Modification Analysis
- **V. Protein Quantification**
- **VI.** Clinical Proteomics



Ion Traps and RF Mass Filters

Quadrupole Mass Filter



Adapted From -> http://www.abrf.org/ABRFNews/1996/September1996/sep96iontrap.html

Linear Quadrupole Ion Trap



http://www.freepatentsonline.com/7180057.html

lons of a certain *m/z range* may be trapped with certain applied voltages with radio frequency (RF) devices, i.e. *electric fields*.

Time-of-Flight (TOF) Mass Spectrometers



Note: Quadrupole Mass Filters are added in-line to enhance the performance of modern MS instruments. These are called "hybrid instruments".

TOF instruments separate different *m/z ions* based on the Flight Time though a field free region after normalized acceleration (i.e. kinetic energy)

Hybrid Quadrupole/Time-of-Flight Mass Spectrometers for Analysis of Biomolecules Werner Ens; Kenneth G. Standing, Methods in Enzymology, Volume 402, 2005, Pages 49–78 Biological Mass Spectrometry

Fourier Transform (FT) Mass Spectrometers



Ions go into "cyclotron frequency" once trapped in a fixed magnetic field. Each ion *m/z* has it's own "cyclotron frequency".

The lons are "excited" to a higher orbit in the "ICR" cell. A second "detector plate" measures the "frequency" to which ions of a given m/z pass.



National High Magnetic Field Laboratory ICR Program Tallahassee, Florida

> 14.5 T Actively Shielded Solenoid Magnet, 104 mm Bore

LTQ

Fourier Transform (FT) Mass Spectrometers





Annu. Rev. Biomed. Eng. 11:49–79

Operates with similar principles as FTICR MS; however without the expensive \$ superconducting magnet \$.

The ions are trapped in the "The Kingdon trap" (aka. OrbiTrap) after transfer from the C-trap. The ions then oscillate back and forth along the central spindle (aka. central electrode). Ions of different *m/z* have different frequencies of oscillation.

I. Mass Spectrometers – Types Summary

Ion Traps (IT); Linear Ion Trap Quadruple (LTQ) and Quadruple [Mass Filters] (Q) = measure stability of ions in an electric field

TOF = measuring the flight time or "kinetic energy" of ions

FT instruments (FT ICR and OrbiTrap) = measure "frequency" of ions moving past a detector within a magnetic or electric field.

(There are other types of mass spectrometers; however, not mentioned)

Information Content – Resolving Power



Lower Resolution

Improved Resolution

I. Resolving Power and Mass Accuracy

The <u>mass resolving power</u> is the measure of the ability to distinguish two peaks of slightly different m/z. The <u>mass</u> <u>accuracy</u> is the ratio of the m/z measurement error to the true m/z. Mass accuracy is usually measured in <u>ppm</u> or <u>milli mass units</u>. The mass range is the range of m/z amenable to analysis by a given analyzer. The <u>linear dynamic</u> <u>range</u> is the range over which ion signal is linear with analyte concentration. Speed refers to the time frame of the experiment and ultimately is used to determine the number of spectra per unit time that can be generated.

Instrument FT-ICR-MS	<u>Can be Accurate To</u> 0.0001 Da	Typical Mass Error 0.2 to 1.0 ppm	Error calculation (500.0001) – (500.0000)
FT-Orbitrap	0.0010 Da	0.2 to 2.0 ppm	(500.0000)
High-Res-TOF	0.0025 Da	0.5 to 10 ppm	multiply by x10 ⁶ = 0.2 ppm error
Quadrupole/Ion Tra	p 0.5 to 1.0 Da	> 1000 ppm	

With *decreasing* mass error, there is a *increased* likelihood of identifying components correctly.

Chemical Formulas of Small Molecules can be predicted by accurate mass measurement (< 2 ppm mass error eliminates 90 - 95% of possible elemental formulas [and peptides])

I. Resolving Power and Mass Accuracy



"Routine Experiments"

Туре	Resolving Power			
FT-ICR-MS	1,000,000+			
FT-Orbitrap	100,000			
High-Res-TOF	60,000			
Quadupole/Ion Trap	1,000			

Example Proteomics Work Flow

Proteomics is a Multistep Process







Tipton, J.D.; Tran, J.C.; Catherman, A.D.; Ahlf, D.R.; Durbin, K.R.; Kelleher, N.L. "Analysis of Intact Protein Isoforms by Mass Spectrometry" Journal of Biological Chemistry, 286 (29), 25451-25458, (2011)



Bottom up and Top down Protein Analysis



Reduce Complexity of Sample

Enrichment of Cell Type, Organelle, Protein Complex Enrichment of PTM (Protein or Peptide Level) Orthogonal Separations, i.e. SCX – RP LC

Concentrations of Proteins in Yeast – The Dynamic Range Problem



Cellular Concentrations of the Set of Measured Proteins Protein abundances are derived from <u>Ghaemmaghami et al. (2003)</u>. Proteins detected by *SRM assays* are sorted by abundance to show the even distribution across the whole range of concentration (blue circles). Proteins for which the absolute abundance was measured *using isotopically-labeled standards* are indicated on top of the graph (open circles).

http://www.sciencedirect.com/science/article/pii/S0092867409007156



http://pubs.acs.org/cgi-bin/sample.cgi/ancham/2008/80/i19/html/ac8013669.html

(1) Strong Cation Exchange (SCX) – Reversed Phase Liquid Chromatography (RP LC)

(2) Weak Anion Exchange (WAX) – RP LC

(3) hydrophilic interaction liquid chromatography (HILIC) – RP LC

(4)



Trypsin Cleaves on the C-terminal Side of Lysine and Arginine

(Leaving a positively charged amino acid side group)

Different peptide combinations have different physiological properties.

(i.e. Tyrosine, tryptophan, etc. have more *hydrophobicity*)

TRYRPYTR

TEQSFTQR

SGILATQR

ALANMARSHALNHMFLMS

(See Previous Page For Structures of Amino Acids)

(Please Draw the Above Peptides For Homework)



The first dimension is often 'ion exchange'

S-cation exchanger

CM-cation exchanger



Peptides with different numbers of 'charges' will be retained under different buffer conditions. (often, increase a 'salt' concentration to obtain separations)





Sample amounts vary from 5 μ g to 50 mg +

Ion Exchange



Pipette tips can be used, however pipette tip sample clean up is associated with *Reversed phase* Clean up

Fractionation by Column

SCX – Collecting Fractions to Reduce Complexity



Collect each fraction \rightarrow Analyze Fractions with nanoRP-LC



C-18 is most common reverse phase chemistry

Common reversed phase structures



http://mach7.bluehill.com/proteinc/tutorial/rpc.html

н ЦС

нцС

н ₂(

Nano Liquid Chromatography and ESI Tips



Capillary is Packed with <u>Reversed Phase</u> Material – nRP-LC

Slide: John Quinn

SDS-PAGE – In-Gel Trypsin Digestion

First Dimension of Separation is SDS-Page (Separation by protein molecular weight)



3D Separations

- (1) Solution Isoelectric Focusing
- (2) Gel-eluted Liquid Fraction Entrapment Electrophoresis (GELFrEE)
- (3) nano RP LC MS



Quick Summary

Peptides with different amino acid combinations have different properties (Homework. Draw Peptide Structures)

The 'charge' and the 'hydrophobic' properties of the peptides allow for multidimensional separations – along with other chemical properties.

Three Basic Multi dimensional Separation Schemes: Ion Exchange + RPLC + MS SDS-PAGE + RPLC + MS Solution IEF + GELFrEE + RPLC + MS

What are we actually measuring with mass spectrometry?



What are we actually measuring with mass spectrometry?

Amino acid	3-lette code	1-letter code	MW (Da)		Structure
Alanine	Ala	A	89.1	\square	CH3-CH (NH2)-COOH
Arginine	Arg	R	174.2	H	I=C (NH2) -NH- (CH2) 3-CH (NH2) -COOH
Asparagine	Asn	N	132.1	Γ	H2N-CO-CH2-CH (NH2)-COOH
Aspartic Acid	Asp	D	133.1		HOOC-CH2-CH (NH2)-COOH
Cysteine	Суз	С	121.2	Γ	HS-CH2-CH (NH2)-COOH
Glutamic Acid	Glu	E	147.1	Γ	HOOC- (CH2) 2-CH (NH2) -COOH
Glutamine	Gln	Q	146.1	Γ	H2N-CO-(CH2)2-CH(NH2)-COOH
Glycine	Gly	G	75.1	Γ	NH2-CH2-COOH
Histidine	His	н	155.2		NH-CH=N-CH=C-CH2-CH (NH2)-COOH
Isoleucine	Ile	I	131.2		CH3-CH2-CH (CH3)-CH (NH2)-COOH
Leucine	Leu	L	131.2		(CH3) 2-CH-CH2-CH (NH2)-COOH
Lysine	Lys	K	146.2		H2N-(CH2)4-CH(NH2)-COOH
Methionine	Met	М	149.2		CH3-S-(CH2)2-CH(NH2)-COOH
Phenylalanine	Phe	F	165.2		Ph-CH2-CH (NH2)-COOH
Proline	Pro	₽	115.1		NH- (CH2) 3-CH-COOH
Serine	Ser	S	105.1		HO-CH2-CH (NH2)-COOH
Threonine	Thr	Т	119.1		CH3-CH (OH) -CH (NH2) -COOH
Tryptophan	Trp	W	204.2		Ph-NH-CH=C-CH2-CH (NH2)-COOH
Tyrosine	Tyr	Y	181.2		HO-p-Ph-CH2-CH (NH2)-COOH
Valine	Val	v	117.1		(CH3) 2-CH-CH (NH2) -COOH

List of amino acids and masses

Individual masses used for Tandem MS Sequencing

Amino acid sequences of proteins may be predicted from Genome Database.

Individual masses used for Tandem MS Sequencing

http://www.geneinfinity.org/sp/sp_aaprops.html

What are we actually measuring with mass spectrometry?

Peptide Sequencing with Mass Spectrometry

- 1) Measure the "Precursor" ion molecular weight See Slide 34
- 2) Select the "Precursor" ion for fragmentation (Collisional Induced/Activated Dissociation – CAD or CID) (Electron Capture/Transfer Dissociation ECD or ETD)

CAD/CID is the most common fragmentation method for proteomics and peptide sequencing.

- 3) Measure the fragment ion mass to charge
- 4) Search measured precursor *m/z* and fragmentation pattern against protein sequence database.

What are we actually measuring with mass spectrometry?

Gas Phase Fragmentation of Peptides/Proteins





http://www.currentprotocols.com/WileyCDA/CPUnit/refId-ps1601.html

What are we actually measuring with mass spectrometry?



	b ion series		y ion series	
G		1		6
F	205.0972	2	665.3253	5
D	320.1241	3	518.2569	4
V	419.1925	4	403.23	3
E	548.2351	5	304.1615	2
R		6	175.119	1

Peptides Have a "Finger Print" Which can be Matched to a Database





Protein identifications are inferred based on the likelihood of a peptide match to a database. (Informatics and probabilities..... another 2 week lecture)

What are we actually measuring with mass spectrometry?

Peptide = DYPVVSIEDPFDQDDWGAWQK



Back To Post-translational Modifications

"The increase in complexity from the level of the genome to the proteome is further facilitated by protein post-translational modifications (PTMs). <u>PTMs are chemical modifications that play a key role in functional proteomics, because they regulate activity, localization and interaction with other cellular molecules such as proteins, nucleic acids, lipids, and cofactors."</u>

Phosphorylation is one of the most studied PTMs (by Mass Spectrometry)



Adapted From Pierce Web Page

PTM type	Average MH+	Modified amino acid residue	Postion	Remarks	Reported in Pub Med (case)
Acetylation	42.04	S K	N-term anywhere	Reversible, protein stability, regulation of protein function	11,069
Phosphorylation	79.98	Y, S, T, H, D	anywhere	Reversible, regulation of protein activity, signaling	103,235
Cys oxidation disulfide bond glutathionylation sulfenic acid sulfinic acid	-2.0 305.31 16.00 32.00	C C C C	anywhere	Reversible, oxidative regulation of proteins	23,538 63 228 642
Acylation famesylation myristoylation palmitoylation	204.36 210.36 238.41	C G K C (S, T, K)	anywhere N-term anywhere anywhere	Reversible, cellular localization to membrane	1,349 644 681
Glycosylation O-linked (O-Glc-NAc) N-linked	>800 203.20, >800	S, T N	anywhere	Reversible, cell-cell interaction and regulation of proteins	24,115
Deamidation	0. 98	N, Q	anywhere	N to D, Q to E	711
Methylation monomethylation dimethylation trimethylation	14.03 28.05 42.08	K K K	anywhere	Regulation of gene expression, protein stability	29,889
Nitration S-Nitrosylation	45.0 29.00	Y C		Oxidative damage	62 399
Ubiquitination Sumoylation		K K	anywhere [ILFV]K.D	Reversible/irreversible	1951 104
Hydroxyproline Pyroglutamic acid	16.00 -17	P Q	N-term	Proein stability	11,424 710

http://bmbreports.org/jbmb/jbmb_files/%5B37-1%5D0401271834_035-044.pdf

Post-translational Modifications: Phosphorylation Example

Increase peptide molecular weight by **80 Da**. NH_2 PO₃⁻ NH_2 H N _{///////} Н OH. Η N H 0 \cap SH Ν L Κ С Н F

The addition of post-translational modifications increases the molecular weight of the peptide. Tandem Mass Spectrometry (MS/MS) fragmentation allows for site characterization. The amount of proteins (peptides) that contain the phosphorylation can be quantified.

Phosphorylation: Reversible protein phosphorylation, principally on serine, threonine or tyrosine residues, is one of the most important and well-studied post-translational modifications. Phosphorylation plays critical roles in the regulation of many cellular processes including cell cycle, growth, apoptosis and signal transduction pathways.



Adapted From: Quantitative Phosphoproteomics Reveals Widespread Full Phosphorylation Site Occupancy During Mitosis, Mann, et.al.

Measure the differences between two (or more) proteomes.

Some Examples:

Change in the Regulation of <u>Protein Classes</u> (i.e. change in Phosphatase and Kinase Abundance)

Change in the Regulation of <u>Post translational modifications</u> (i.e. change in Phosphorylation)

Change in <u>Phenotype</u> due to Mutations (i.e. mutation in p53 as related to different types of cancers)

..... etc. – leads into <u>Clinical Proteomics</u>

Experiments for Quantitation in Proteomics

Stable isotope labeling by amino acids in cell culture (SILAC)



"Stable Isotope Labeling by Amino acids in Cell culture (SILAC) is a simple, robust, yet powerful approach in mass spectrometry (MS)-based quantitative proteomics. SILAC labels cellular proteomes through normal metabolic processes, incorporating non-radioactive, stable isotope containing amino acids in newly synthesized proteins. Growth medium is prepared where natural ('light') amino acids are replaced by 'heavy' SILAC amino acids. Cells grown in this medium incorporate the heavy amino acids after five cell doublings and SILAC amino acids have no effect on cell morphology or growth rates. When light and heavy cell populations are mixed, they remain distinguishable by MS and protein abundances are determined from the relative MS signal intensities. A variant of the approach, Triple Encoding SILAC, allows three protein populations to be compared simultaneously. SILAC provides accurate relative quantification without any chemical derivatization or manipulation and enables steady-state proteome quantification or pulse-chase experiments."

Experiments for Quantitation in Proteomics

Isobaric tags for relative and absolute quantitation (iTRAQ)







The method is based on the covalent labeling of the N-terminus and side chain amines of peptides from protein digestions with tags of varying mass. There are currently two mainly used reagents: 4-plex and 8-plex, which can be used to label all peptides from different samples/treatments. These samples are then pooled and usually fractionated by SCX followed by nano liquid chromatography and analyzed by tandem mass spectrometry (MS/MS). A database search is then performed using the fragmentation data to identify the labeled peptides and hence the corresponding proteins. The fragmentation of the attached tag generates a low molecular mass reporter ion that can be used to relatively quantify the peptides and the proteins from which they originated.

The signals of the reporter ions of each MS/MS spectrum allow for calculating the relative abundance (ratio) of the peptide(s) identified by this spectrum. The abundance of the reporter ions may consist of more than one single signal in the MS/MS data and the signals have to be integrated in some way from the histogram spectrum.

Experiments for Quantitation in Proteomics

Isotopic Dilution – Addition of "heavy" Labeled Synthetic Peptide

Protein Standard Addition Curve



A synthetic peptide with a "heavy" label is introduced to the mixture as an internal standard.

Absolute quantitation is achieved for single biomarker peptides

VI. Clinical Proteomics

Basic Science \rightarrow Translational Science \rightarrow <u>Clinical Proteomics</u> ('Omics') \rightarrow New Treatments for Human Disease

"<u>Clinical Proteomics</u> is an emerging <u>interdisciplinary research field</u> that coalesces researchers from many different areas of biomedical research into one of the most likely disciplines to successfully foster the <u>translation of basic scientific knowledge into clinical</u> <u>applications for the benefit of the patient</u>. Major research interests that are expected to profit from the application of proteomic technologies to clinical issues include:

(i) early detection/diagnosis of disease

(ii) prediction of how a disease will behave over time and how a specific patient will respond to a given treatment

(iii) identification of novel targets for therapeutic intervention."

VI. Clinical Proteomics

"Today, clinical proteomics studies are increasingly shifting toward the analysis of tissues and bio-fluids and as a result, concerns such as tissue heterogeneity, abundance of the cell type of interest, marker quantitation, availability of samples with long-term clinical follow-up, as well as experimental design and optimal method selection have become some of the main challenges that researchers face. These and other inherent problems associated with using clinical samples make interdisciplinary joint efforts a necessity for any high-quality clinical proteomics study. Furthermore, the translation of basic proteomic discoveries to a clinical setting is a very long and expensive process that requires broad multidisciplinary collaborations between basic researchers, clinicians, surgeons, pathologists, and epidemiologists as well as infrastructural support and industry participation."

VI. Clinical Proteomics

Apply the basic methods described in previous slides to analyze real human samples.

Often need special certifications to work on clinical samples: Institutional review board (IRB) a.k.a. independent ethics committee (IEC) or Ethical Review Board (ERB)

Clinical Proteomics is often combined with other 'Omics' (*slide 3*) results to describe the full 'systems biology' of a 'phenotype or organism'

With systems biology understanding of a human, we can start to prescribe "Personalized Medicine"