CMSP Proteomics Workshop Itinerary March 2023

- **Lecture Tuesday-Thursday:** Cargill Building 105, St Paul Campus, 1500 Gortner Avenue
- **In-person lab activities Tues-Thurs:** Gortner 46, St Paul Campus

### March 14, 2023

<table>
<thead>
<tr>
<th>Time</th>
<th>Topic</th>
<th>Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00 - 9:15 am</td>
<td>Welcome and Introductory Remarks</td>
<td>Tim Griffin</td>
</tr>
<tr>
<td>9:20 - 10:00 am</td>
<td>Introduction to Proteomics I</td>
<td>Tim Griffin</td>
</tr>
<tr>
<td>10:00 - 10:30 am</td>
<td>Introduction to Proteomics II</td>
<td>LeeAnn Higgins</td>
</tr>
<tr>
<td>10:35 - 10:45 am</td>
<td>Group presentation discussion and assignments</td>
<td>Tim Griffin</td>
</tr>
<tr>
<td>11:10 - 12:00 pm</td>
<td>Sample Preparation</td>
<td>Todd Markowski</td>
</tr>
<tr>
<td>1:00 - 4:00 pm</td>
<td>Workshops Ia &amp; Ib</td>
<td>Todd Markowski/ Cesar Velasquez</td>
</tr>
<tr>
<td></td>
<td>Ia: Stage-tip Clean-up for MALDI-TOF</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ib: MALDI-TOF Analysis</td>
<td>Tom Krick</td>
</tr>
<tr>
<td></td>
<td>Ic: Sample Preparation Technologies and Instrumentation</td>
<td>Andrew Rajczewski</td>
</tr>
</tbody>
</table>

### March 15, 2023

<table>
<thead>
<tr>
<th>Time</th>
<th>Topic</th>
<th>Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00 - 10:00 am</td>
<td>Introduction to Mass Spectrometry: Mass Spec 101</td>
<td>Tom Krick</td>
</tr>
<tr>
<td>10:00 - 11:05 am</td>
<td>Tandem MS and Intro to Liquid Chromatography-MS</td>
<td>LeeAnn Higgins</td>
</tr>
<tr>
<td>11:05 - 11:20 am</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>11:20 - 12:30 pm</td>
<td>Specialized Applications in Mass Spectrometry</td>
<td>Andrew Rajczewski</td>
</tr>
<tr>
<td>12:30 - 1:30 pm</td>
<td>Lunch</td>
<td></td>
</tr>
<tr>
<td>1:30 - 4:00 pm</td>
<td>Workshops IIa, IIb, &amp; IIc</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IIa: ESI-Ion Trap MS - Peptide MSMS and intact protein</td>
<td>LeeAnn Higgins</td>
</tr>
<tr>
<td></td>
<td>IIb: Manual Interpretation of MS/MS spectrum</td>
<td>Candace Guerrero</td>
</tr>
<tr>
<td></td>
<td>IIc: Proteomics vs. Metabolomics</td>
<td>Kevin Murray</td>
</tr>
</tbody>
</table>

### March 16, 2023

<table>
<thead>
<tr>
<th>Time</th>
<th>Topic</th>
<th>Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00 - 9:40 am</td>
<td>Database Searching</td>
<td>LeeAnn Higgins</td>
</tr>
<tr>
<td>9:40 -9:45 am</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>9:45 - 10:45 am</td>
<td>Protein Validation and Quantification</td>
<td>Candace Guerrero</td>
</tr>
<tr>
<td>10:50 - 12:00 pm</td>
<td>Untargeted Quantification</td>
<td>Kevin Murray</td>
</tr>
<tr>
<td>12:00 - 1:00pm</td>
<td>Lunch</td>
<td></td>
</tr>
<tr>
<td>1:00 - 3:30 pm</td>
<td>Proteoinformatics Workshops IIIa, IIb, &amp; IIc</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IIIa: Peptide Mass Fingerprinting (PMF) and MS/MS Analysis</td>
<td>LeeAnn Higgins/ Todd Markowski</td>
</tr>
<tr>
<td></td>
<td>IIIb: Viewing peptide and protein results: Scaffold</td>
<td>LeeAnn Higgins/ Kevin Murray</td>
</tr>
<tr>
<td></td>
<td>IIIc: Wrap-up and Q&amp;A</td>
<td>LeeAnn Higgins/ Candace Guerrero</td>
</tr>
</tbody>
</table>
Introduction to Proteomics 1.0

CMSP Workshop

Tim Griffin
Professor, BMBB
Faculty Director, CMSP
Why are we here?

Objectives

For participants:
• Learn basics of MS-based proteomics
• Learn what’s necessary for success using MS-based proteomics
  • Designing experiments; sample preparation; data analysis

For CMSP staff:
• Prepare users so they are equipped to have success working with CMSP
• Manage expectations – what can these technologies do and not do
Alien language made understandable

CMSP

Participants
Terminology made sensible

CMSP

Right on!

Participants

NanoLC, HCD, Precursor ion, MALDI, ES, monoisotopic, ES, quadrupole, TOF, MS/MS, iTRAQ, b-ion, Ion trap, Stage-tip, CMSP, Participants
Who we are

Center for Metabolomics and Proteomics

- Operated through the Department of Biochemistry, Molecular Biology and Biophysics
- Serving biological MS-related research needs across UofM campus and external institutions/private companies
- Fee-for-service Internal Service Organization (ISO)
- Supported by all Colleges at UofM using CMSP and Office of Vice President for Research (plus variety of granting sources)
- Extensive collaboration with Minnesota Supercomputing Institute/OIT/UMII

**Primary mission** to support research efforts at the University of Minnesota, but also train others in the use of advanced technologies and research approaches
Who we are

• 150+ collective years of experience in biological MS; hundreds of scientific publications
• Diverse expertise – design, sample preparation, instrumentation, data analysis
• Experience with MANY sample types and research studies
  • Fish…gophers…periodontal bacteria…snake venom
‘Omic technologies and the molecular biology paradigm


Image Source: http://fluorous.com/images/omics.JPG
Why proteomics and direct protein analysis?

(Genomic sequencing is cheaper, faster, more comprehensive…why proteomics?)

- DNA/RNA characterization cannot predict post-transcriptional events
Proteomics: A definition

“Proteomics includes not only the identification and quantification of proteins, but also the determination of their localization, modifications, interactions, activities, and, ultimately, their function.”


Alternatively: proteomics = high-throughput biochemistry
Proteomics as a complement to genomics

- measurement of protein response, which is not always indicated by mRNA response
- post-translational modifications
- macromolecular interactions
- sub-cellular location
- high-resolution structural and molecular characterization
- integration with genomic/transcriptomic data to comprehensively characterize biological systems
Proteomic technologies and approaches

• two-dimensional gel electrophoresis
• mass spectrometry
• protein chips
• yeast 2-hybrid
• phage display
• antibody engineering
• high-throughput protein expression
• high-throughput X-ray crystallography
• cell imaging
Enabling MS-based proteomics: “soft” ionization

• Making large, non-volatile biomolecules fly

Electrospray ionization (ESI)

Matrix-assisted laser desorption/ionization (MALDI)

Matrix

sample

5 mm

UV laser

200 µm
Nuts and bolts of mass spectrometry

- **ionization**
  - MALDI
  - Electrospray

- **separation by m/z**
  - quadrupole
  - ion trap
  - time-of-flight

- **detection**
  - mass analysis of proteins, peptides

*m/z = mass-to-charge*
Many instruments, same underlying process
Example of technology progress: more sensitive MS

---

Image from ASMS 2014 workshop (Speaker: Haas)
The information currency of MS
The “guts” of a mass spectrometer

m/z separation and detection

m/z separation

ionization

Figure 1. Schematic representation of the instrument.

Anal. Chem. 2013, 85, 11710–11714
Doing protein and proteomic analysis via MS

- Biological inquiry
  - Hypothesis
  - Experimental design

- Sample preparation

- MS analysis

- Data analysis

- Workshop structured to follow this ordering
- All aspects are important: each must be done well for success
- Challenge:
  - technologies within each component always changing
  - interdisciplinary
The importance of sample preparation

- Garbage in, garbage out

- Protein mixtures isolated from biological sources are complex (hundreds to thousands of components)

- Mass spectrometers have limited peak capacities requiring separation and fractionation of protein and peptide mixtures prior to analysis

- Separation methods include:
  - gels
  - liquid chromatography
  - affinity chromatography
  - immunochromatography
  - selective enrichment by covalent chemistry
Protein chemistry: a challenge

• Proteins offer unique challenges compared to other biomolecules (e.g. nucleic acids):
  
  – Solubility
  
  – Abundance (no PCR!)
  
  – Chemical heterogeneity

  Each protein is a unique character!
The workhorse: LC-MS

- Separating molecular mixtures prior to introduction into MS
Some example applications: from simple to complex

The “simple”: identifying a gel separated protein

2D gel electrophoresis: the original proteomics technology…but how to ID proteins?

Gygi, et. al. 1999, Molecular and Cellular Biology 19:1720
Even the simple still requires care.....

- Process of identifying a gel-separated protein

A bit more complicated: Identifying PTMs on a protein

• Phosphorylation
• Glycosylation
• Oxidations
• Acetylation
• Methylation
• Lipid anchors
• Ubiquitinylation/sumoylation

BUT….PTM analysis is not necessarily routine or easy!!

(abundance, enrichment, ionization, fragmentation….)

Center for Metabolomics and Proteomics
Still more complicated: identification of proteins in complex mixtures

- More complicated sample preparation (fractionation)

1. Separation of peptides by electrostatic charge
2. Separation by hydrophobicity
   - RP-uLC MS

Complex peptide mixture → UV absorbance → Cation exchange HPLC → MS
A bit more complicated: Quantitative proteomics

S. cerevisiae cell cycle
(compliments of J.A. Huberman)

- Protein abundance is **dynamic** in response to environmental, genetic, biochemical, pathological perturbations.
Quantitative proteomics: many methods available

- Labeled versus unlabeled

Sample:

<table>
<thead>
<tr>
<th>Sample</th>
<th>Digest to peptides</th>
<th>iTRAQ label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>+114</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>+115</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>+116</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>+117</td>
</tr>
</tbody>
</table>

Multidimensional separation

MS/MS spectrum

Intensity

m/z 114 115 116 117

Diagnostic ions used for quantitative analysis
Peptide fragments used for sequence identification
Systems biology: integrating

- Genome annotation
- Gene expression regulation
- Protein variants in disease
- Functional outcomes of genome mutation
Dealing with the data: the rate-limiting step?

Data acquisition

Raw data processing (Database searching)

Analysis of processed data (Statistical filtering, quantitative analysis)

Data organization and interpretation

Archiving and databasing
Workflow for protein identification

Raw MS/MS spectrum

Protein sequence and/or DNA sequence database search

Direct identification of 1000+ proteins from complex mixtures

Peptide sequence match

Protein identification
Bioinformatic interpretation and hypothesis generation

KEGG pathways

Proteomics 2012, 12, 992–1001
Thank you

Good luck

May all your ions fly well!
INTRODUCTION TO PROTEOMICS II

Experimental Design
<table>
<thead>
<tr>
<th>Outline</th>
<th>Terminology</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Goals</td>
<td>• Targeted vs. Discovery</td>
</tr>
<tr>
<td>• Experimental design</td>
<td>• Dynamic range</td>
</tr>
<tr>
<td>• Design choices</td>
<td>• Inference</td>
</tr>
<tr>
<td>• Tradeoffs</td>
<td>• Sampling error</td>
</tr>
<tr>
<td>• Workflows</td>
<td>• Bias</td>
</tr>
<tr>
<td>• Challenges</td>
<td>• Variability</td>
</tr>
<tr>
<td></td>
<td>• Blocking</td>
</tr>
<tr>
<td></td>
<td>• Randomization</td>
</tr>
</tbody>
</table>
Proteomics Research Areas

• Expression proteomics
  – Protein expression and regulation, including identification, quantification, differential expression, e.g., compare physiological or environmental conditions

• Functional proteomics
  – Protein function and protein interaction networks with other proteins, lipids, DNA, ligands

• Structural proteomics
  – Studying and modeling protein structure e.g., inferring function from protein structure

https://doi.org/10.1093/beheco/aru096
Common Goals

• Protein identification
• Protein-protein interactions via immunoprecipitation (IP) or co-IP
• Post translational modifications
• Protein quantification, relative or absolute
Typical Mass Spectrometry-Based Proteomics Workflow

Experimental Design

Sample Prep
- Organelle Fraction
- Purified Protein Complex
- Single Protein of Interest on a Gel
- Cell Lysate
- Sample for PTM analysis

LC* and Mass Spec
- Proteolytic Digest
- LC- ESI
- MS
- MS/MS

Data Analysis
- Cellular Composition
- Organellar Proteome
- Protein ID
- Interaction Protein
- PTM Analysis

* Liquid Chromatography

REF: Adapted from Walther T, Mann M. JCB 2010; 190: 491-500
Experimental Design

- Design your experiments with a particular biological question in mind
- Plan an exploratory survey or a hypothesis-driven question
- Plan your experimental design before you carry out an experiment
Design Choices

• Discovery vs Targeted
  – Trade-offs
  – Protein abundance

• Negative controls

• Number of Biological and/or Technical Replicates

• Statistical Rigor

• Workflow Technologies
Design Choice: Discovery vs Targeted

Trade-offs
Number of analytes versus number of samples

(Disclosure)  (Targeted)

Basic Research Biobanking New Technologies

Translational Research

Clinical Methods

Discovery  Qualification  Verification  Validation

Number of analytes

Number of samples
Design Choice: Discovery vs Targeted Inference challenges & mass spectrometry

“Schematic representation of the fraction of a proteome that can be identified or quantified by mass-spectrometry-based proteomics. Cellular proteins span a wide range of expression and current mass spectrometric technologies typically sample only a fraction of all the proteins present in a sample. Due to limited data quality, only a fraction of all identified proteins can also be reliably quantified” PMID: 17668192
The Mass Spec Inference Problem

• A molecule that is PRESENT in a sample could be UNDETECTABLE by mass spectrometry
  – Causes
    • High dynamic range of protein content
    • Sample preparation methods
    • Instrumentation choice
    • Data analysis methods
    • And more....
• Identification and quantification requires statistics, which requires experimental design
Design Choice: Discovery vs Targeted Protein Abundance Examples

High Abundant Protein Sequence Coverage

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Coverage</th>
</tr>
</thead>
</table>

Low Abundant Protein Sequence Coverage

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>F Y E S D R R R R R S S S L L L</td>
<td>R E L S A E L T R G</td>
</tr>
<tr>
<td>R R L G P O A V C Q</td>
<td>S D I G K Y I L V M</td>
</tr>
<tr>
<td>I D R F T I C G N R</td>
<td>I D</td>
</tr>
</tbody>
</table>
Design Choice: Negative Control
Example: Protein-Protein Interactions

1. Co-IP with fusion protein
2. Co-IP with FLAG peptide (negative control)
3. Elute protein complexes, run SDS-PAGE
4. In-gel trypsin digestion of protein bands of interest
5. Protein identification by peptide mass spectrometry
6. Repeat or use biological replicates

Co-IP with FLAG-tagged bait protein
NEGATIVE CONTROL

IP agarose: anti-FLAG anti-FLAG
HEK transfection: D1-FLAG FLAG-tag

PMID 19170023
Design Choice: Number of Biological Samples
Perform POWER ANALYSIS from pilot experiment to estimate minimum sample size required to find statistically significant differences between conditions

PILOT: $N = 3$

EXAMPLE: Quantify Protein Abundance Changes over Time in Embryonic Stem Cells
https://doi.org/10.1038/nrg3356

- On POWER: The higher the statistical power for a given experiment, the lower the probability of making a Type II (false negative) error. https://machinelearningmastery.com/statistical-power-and-power-analysis-in-python/
- Read more: DOI: 10.1080/19345747.2017.1342887
Design Choice: Statistical Rigor - BLOCKING

Experiments should be blocked properly to minimize confounding effects.

Fig. 2. Example of a good (A) and a bad (B) design. In design A, both the green and orange treatments are divided equally within each block. That way, the treatment effect can be estimated within a block. In design B, each block contains only one treatment, so the treatment effect is entirely confounded with the blocking effect. https://doi.org/10.1016/j.jprot.2017.04.004
Design Choice: Statistical Rigor
Numbers of Genotypes are BALANCED
Experiments are BLOCKED identically

Example: Compare relative protein abundances in cells with unique genotypes A, B, C, D

TMT* 10plex Experiment 1

<table>
<thead>
<tr>
<th>Isotope Label</th>
<th>126</th>
<th>127N</th>
<th>127C</th>
<th>128N</th>
<th>128C</th>
<th>129N</th>
<th>129C</th>
<th>130N</th>
<th>130C</th>
<th>131</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Category</td>
<td>Control / Ref.1</td>
<td>Control / Ref.2</td>
<td>GT.A.rep.1</td>
<td>GT.A.rep.2</td>
<td>GT.B.rep.1</td>
<td>GT.B.rep.2</td>
<td>GT.C.rep.1</td>
<td>GT.C.rep.2</td>
<td>GT.D.rep.1</td>
<td>GT.D.rep.2</td>
</tr>
</tbody>
</table>

TMT* 10plex Experiment 2

<table>
<thead>
<tr>
<th>Isotope Label</th>
<th>126</th>
<th>127N</th>
<th>127C</th>
<th>128N</th>
<th>128C</th>
<th>129N</th>
<th>129C</th>
<th>130N</th>
<th>130C</th>
<th>131</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Category</td>
<td>Control / Ref.1</td>
<td>Control / Ref.2</td>
<td>GT.A.rep.1</td>
<td>GT.A.rep.2</td>
<td>GT.B.rep.1</td>
<td>GT.B.rep.2</td>
<td>GT.C.rep.1</td>
<td>GT.C.rep.2</td>
<td>GT.D.rep.1</td>
<td>GT.D.rep.2</td>
</tr>
</tbody>
</table>

- GT = genotype
- Rep = biological replicate
- Control/Reference: Pooled Control for normalization across experiments

* TMT = Tandem Mass Tag (ThermoFisher Scientific relative protein quantification technology)
Design Choice: Statistical Rigor

VARIANCE and statistically significant differences

Example: Relative Protein Abundances in Two Conditions

https://doi.org/10.1038/s41598-018-25035-1
Sources of Bias and Variability

- gender
- age
- diet
- lifestyle
- individual
- batch effects:
  - column age
  - solvent variation
- ambient air
- preventative maintenance schedule
- technique
- storage
- personnel
- difficulty
SUMMARY: Control for Bias & Variability

• Control for Bias
  — Control group
  — Randomization
  — Normalization

• Reduce effects of sampling error
  — Replication
  — Balance (equal number of sample groups)
  — Blocking (identical replication of treatment groups/categories in different settings)
Workflow Technologies

Sample Prep
- Biopsy
- Biofluid
- Laser-capture microdissection
- Cell sorting (FACS)
- Primary cell culture
- Stable cell line culture
- Free-flow electrophoresis
- Gradient centrifugation

Mass Spec
- 1D and 2D gel electrophoresis
- Isoelectric focusing
- Capillary electrophoresis
- Column chromatography
- Immunoprecipitation
- Pulldowns with tagged proteins
- Cell surface labeling
- Active site labeling
- Affinity depletion
- Phosphoflow
- Glycocapture

Protein quantification
- Metabolic labeling (SILAC, 15N)
- Chemical protein labeling (CPL)
- Chemical peptide labeling (iCAT, cCAT, TMT, methylation, esterification)
- Enzymatic peptide labeling (18O)
- Absolute quantification (AQ, QconCAT)
- Label-free (spectrum counting, emPAI, APEX, XICs, expression)
- Single/multiple reaction monitoring (SRM, MRM)
- Express, Pepper, MSQuant, MaxQuant, iTracker, TPP, CPAAS, TOPP, ProteoWizard

Sample extraction
- 1D and 2D gel electrophoresis
- Isoelectric focusing
- Capillary electrophoresis
- Column chromatography
- Immunoprecipitation
- Pulldowns with tagged proteins
- Cell surface labeling
- Active site labeling
- Affinity depletion
- Phosphoflow
- Glycocapture

Protein identification
- Database searching
- De novo sequencing
- Peptide mass fingerprinting (PMF)
- Accurate mass and time tag (AMT)
- Mascot, Sequest, X!Tandem
- OMSSA, Phenix, Spectrum Mill
- PEAKS, PepNovo, InSpectr, PTM Score, A-Score, ModifiComb

Mass spectrometry
- Electrospray ionization (ESI)
- Matrix-assisted laser desorption/ionization (MALDI)
- Time-of-flight MS (TOF)
- Ion trap MS
- Quadrupole MS
- Orbitrap MS
- Fourier-transform ion cyclotron MS (FT-ICR)
- Liquid chromatography MS (LC-MS)
- Imaging MS
- Ion mobility MS
- Tandem mass spectrometry (MSn)
- Collision-induced dissociation (CID)
- Electron-transfer dissociation (ETD)
- Electron-capture dissociation (ECD)
- Post-source decay (PSD)

Peptide fractionation
- Ion-pairing reversed phase (RP-HPLC)
- Isoelectric focusing (IEF)
- Strong cation exchange (SCX)
- Weak anion exchange (WAX)
- Hydrophilic interaction (HILIC)
- Immobilized metal affinity (IMAC)
- Titanium dioxide, zirconium dioxide
- Lectin affinity chromatography
- Immunoprecipitation
- Biotinylation
- Fractional diagonal chromatography

Data Analysis
Before you start ...

PLAN

your experiments carefully

• Discuss the project with CMSP personnel
• Consult a statistician when necessary
Introduction to Proteomic Sample Preparation for Mass Spectrometry
Two Main Approaches to Proteomics

**Bottom Up Proteomics**
- Proteolytic Digestion
- HPLC or SPE
- Mass Spec.
- Analysis

**Top Down Proteomics**
- HPLC, Gelfree, CE
- FT-ICR Mass Spec.
- Analysis

Adapted from http://www.piercenet.com/method/sample-preparation-mass-spectrometry

© 2015 Regents of the University of Minnesota. All rights reserved.
Sample preparation is one of the most important aspects of any successful proteomics experiment!

Questions you should ask before proteomics sample prep:

1. How should I collect the samples?
   Consistency, consistency…
   ...and more consistency
   Snap freeze
   Sampling for biological variation
   Document all steps from sample collection thru prep – Good lab notes!

2. What do I hope to analyze?
   Simple protein ID
   Differentially expressed proteins (ITRAQ, SILAC, label free or targeted)
   Post-translational modifications (PTM)
   Cytoplasmic, membrane, nuclear, mitochondrial, etc.
   Immunoprecipitation/binding partners/TAP
Questions before starting prep (continued):

3. How do I best extract my proteins of interest?

- Extraction buffer
- In-gel: SDS-PAGE (1D or 2D)
- In-solution
- Enrichment/affinity strategies
- Depletion of high abundant proteins

4. Is the extraction buffer compatible with mass spectrometry? If not, how do I get rid of problematic buffer components?

- Solid Phase Extraction (SPE), detergent removal spin-column
- Cloud point extraction of non-ionic detergents
- Ethyl acetate extraction of detergents
- Protein precipitation
- Dialysis, molecular weight cut-off spin filter
Overview: Sample Prep to Mass Spectrometry Work Flow

**Lysate Preparation/Protein Extraction**

**Solid Tissue and cell culture samples**
- Disruption
  - sonication
  - bead beater/homogenizers
  - pressure cycling technology
  - freeze + mortar & pestle
  - Enzymatic
  - Heat, Freeze/thaw
- Extraction Buffer (denature proteins)
- Chaotropes (Disorder maker)
  - Urea, thiourea, guanidine HCl, salts(KCl, NaCl, etc.), MeOH, acetonitrile
- Detergents
  - Ionic (SDS), nonionic (NP40, triton x-100), zwitterionic (CHAPS)
- Buffer reagent – pH considerations
  - Tris, HEPES, ammonium bicarbonate, PBS, TEAB
- Reducing reagents
  - Dithiothreitol (DTT), TCEP (tris(2-carboxyethyl)phosphine)

**Biological Fluids, Media, other solution preps (i.e. IP’s)**
- MW Spin filter
- Dialysis
- Precipitation
- Denaturing Buffer
- Common Extraction Buffers: 4% SDS, 100mM Tris pH 7.6, 100mM DTT
  - 7M urea, 2M thiourea, 0.4M TEAB pH8.5, 20% acetonitrile, 4mM TCEP

Centrifuge insoluble material out & aspirate supernatant (protein extract)

Depletion of high abundant proteins typically done before digestion

Enrichment of post-translational modifications typically done after digestion

© 2015 Regents of the University of Minnesota. All rights reserved.
Disruption with the Barocycler® NEP2320
Maximizing Protein Extraction from Tissue

Pressure Cycling Technology Sample Preparation System (PCT SPS)
...the Power of PCT

- Pressure Cycling Technology (PCT) uses cycles of hydrostatic pressure between ambient and ultra high levels allowing for a high degree of speed, reproducibility, and convenience.

- 1 PSI input produces 440 PSI output resulting in a top pressure of 35 kPSI.

- Better extraction from various sample types and quicker enzymatic treatments (proteolytic cleavage, deglycosylation, etc.)
Pressure Compresses Lipids Beyond Equilibrium

Rapid De-pressurization Causes Membranes and Micelles to Disintegrate

Hydrostatic Pressure Rapidly Released

PCT = pressure cycling technology, 10 cycles, 20 sec. 35 kpsi, 20 sec ambient press.

GG = ground glass dounce homogenizer

500 ug loaded on each gel

<table>
<thead>
<tr>
<th>Method</th>
<th>Spots detected</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCT</td>
<td>2280 ± 173</td>
</tr>
<tr>
<td>GG</td>
<td>1620 ± 137</td>
</tr>
<tr>
<td>GG/sonication</td>
<td>1735 ± 144</td>
</tr>
<tr>
<td>GG/sonication/(\text{Vol} \times 2)</td>
<td>1682 ± 165</td>
</tr>
</tbody>
</table>

Values are mean ± SD (n = 5).
# +/- Barocycler Prep & Digestion

mouse liver tissue

<table>
<thead>
<tr>
<th>Treatment</th>
<th>ABI 4800 Protein ID’s</th>
<th>Thermo LTQ Protein ID’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std. Lysis/Std. Digest</td>
<td>115 (99% Protein Pilot) 203 (95% Protein Pilot)</td>
<td>118 (≥3 peptides 99%, Scaffold)</td>
</tr>
<tr>
<td>Std. Lysis/PCT Digest</td>
<td>NR</td>
<td>219</td>
</tr>
<tr>
<td>PCT Lysis/Std. Digest</td>
<td>275 471</td>
<td>264</td>
</tr>
</tbody>
</table>

PCT Lysis = 30 cycles @ 37C, 50 sec. 35kpsi, 10 sec. Ambient press.
PCT Digest = 60 cycles @ 37C, 50 sec. 20kpsi, 10 sec. Ambient press.

Lysis Buffer = 8M urea, 0.2% SDS, 0.5M TEAB, 5mM TCEP

Std. Digest = reduce, alkylate, dilute sample to 2M urea, add trypsin in 1:25 ratio, incubate overnight @ 37C

Std. Lysis = mortar & pestle + sonication
Std. Digest = overnight (16 hrs.)
Biological Fluid Prep Example - Urine

Urine Sample → MW Cut-off Spin Filter → Recover sample → Dialyze → Vacuum Centrifuge to Dryness → Resuspend with Solubilization buffer

Protein Concentration Assay (Bradford or BCA)

Proteolytic Digestion → HPLC or SPE → Mass Spec.
Two main sample prep paths to take...choose wisely!

“When You Come to a Fork in the Road, Take It!” –Yogi Berra
In-gel Proteolytic Digestion

Washing: A series of water, ammonium bicarbonate and acetonitrile additions and aspirations to remove impurities from the separation

Reduction: DTT or TCEP + heat

Alkylation: iodoacetamide (+57 mass shift on Cys)
           methyl methanethiosulfonate (+46 on Cys)

Washing and Dry: ammonium bicarb./acetonitrile washes
dry with 100% acetonitile – no speedvac

Digestion: Trypsin 5ng/ul 25-100mM ammonium bicarbonate
           other enzymes, AspN, GluC, Chymotrypsin
           Incubate trypsin digest overnight @ 37°C

Extraction of peptides: 50/50 acetonitrile/water, 0.1% TFA or FA
                       75/25 acetonitrile/water, 0.1% TFA or FA
                       pool digest and extracts for same sample
dry down in vacuum centrifuge for SPE or mass spec.

gel cutting pictures: http://sites.psu.edu/msproteomics/category/sample-prep/in-gel-digestion-tutorial/
workflow picture: www.piercenet.com/method/sample-preparation-mass-spectrometry

Adapted from: www.cshprotocols.cshlp.org
In-Solution Proteolytic Digestion

Reduction: DTT or TCEP + heat
Alkylation: iodoacetamide (+57 mass shift on Cys)
methyl methanethiosulfonate (+46 on Cys)

Dilute buffer components that interfere with proteolytic enzyme: urea < 2M, thiourea < 1M SDS < 0.1%

Digest: Add trypsin (proteolytic enzyme) in 1:35 to 1:100 enzyme to total protein ratio

Incubate overnight (12-16hrs) @ 37°C

Freeze digest at -80°C and dry down for SPE of peptides

*Consider if buffer components can be cleaned up with SPE or HPLC step before mass spec., otherwise need to rethink buffer

Adapted from: www.cshprotocols.cshlp.org
Filter Aided Sample Prep - FASP

a. Lysate, SDS and DTT
   Add 8 M urea
   Centrifuge (remove SDS, DTT and low-molecular-weight material)

b. Retenate
   Add 8 M urea and IAA
   Incubate
   Centrifuge (remove excess IAA)

c. Retenate
   Add 8 M urea
   Centrifuge (remove any remaining reagent)

d. Retenate
   Add endoproteinase
   Incubate
   Centrifuge to collect peptides (eluate)

### Many different proteolytic enzymes – choose based on your application

<table>
<thead>
<tr>
<th>Protease</th>
<th>Cleavage site</th>
<th>Example of use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lys-C Specific protease</td>
<td>NNNNK$_{N}$NNN (K is lysine)</td>
<td>Digests membrane and other proteolytically resistant proteins; generates larger peptides than trypsin peptides—advantage for certain mass spec methods (for example, ETD)</td>
</tr>
<tr>
<td>Arg-C Specific protease</td>
<td>NNNR$_{R}$NNN (R is arginine)</td>
<td>Arg-C can, at a lesser degree, cleave at lysine also Facilitates analysis of histone posttranslational modifications; used in proteome-wide analysis</td>
</tr>
<tr>
<td>Glu-C Specific protease</td>
<td>NNNNE$_{E}$NNN (E is glutamate)</td>
<td>Glu-C can, at a lesser degree, cleave at aspartate residue also Used as an alternative to trypsin if trypsin produces too short or too long peptides or if trypsin cleavage sites are not accessible</td>
</tr>
<tr>
<td>Asp-N Specific protease</td>
<td>NNNND$_{D}$NNN (D is aspartate)</td>
<td>Similar to Glu-C</td>
</tr>
<tr>
<td>Chymotrypsin Low Specific protease</td>
<td>NNNN(F/Y/W)$_{F/Y/W}$NNN (F, Y and W are aromatic residues phenylalanine, tyrosine and tryptophan)</td>
<td>Digests hydrophobic proteins (for example, membrane proteins)</td>
</tr>
<tr>
<td>Pepsin Nonspecific protease</td>
<td>Nonspecific protease (advantage—most active at low pH)</td>
<td>Used in structural protein studies and antibody analysis; digests proteolytically resistant, tightly folded proteins</td>
</tr>
<tr>
<td>Thermolysin Nonspecific protease</td>
<td>Nonspecific protease (advantage—remains active at high temperature)</td>
<td>Digests proteolytically difficult, tightly folded proteins; used in structural protein studies</td>
</tr>
<tr>
<td>Elastase Nonspecific protease</td>
<td>Nonspecific protease</td>
<td>Used to increase protein coverage</td>
</tr>
</tbody>
</table>
Solid Phase Extraction (SPE)/Clean-up of Peptides

Peptides + Digest Buffers

- Only Buffer & Salts
- C18 material
- Peptide Enrichment/Cleanup

Other anionic hydrophobic contaminants

Sodium Dodecyl Sulfate (SDS)

Nonionic/Zwitterionic/ionic Detergents (triton-x100, CHAPS, NP40, SDS, etc.)

MCX - Mixed Mode Cation Exchange

- 2 punches
- Ethyl acetate precipitation of detergents

Cloud Point Extraction – only for nonionic detergents
(J. Proteome Res., 2010, 9 (8), pp 3903–3911)

Detergent removal spin columns

Pierce detergent removal: http://www.fishersci.com

Cartridge images: www.perkinelmer.com
tips & Empore image: www.sigmaaldrich.com
C18 resin image: www.lamondlab.com

Image: biotage.phosdev.se
**Lysate Preparation/Protein Extraction**

**Solid Tissue and cell culture samples**
- Disruption
  - sonication
  - bead beater/homogenizers
  - pressure cycling technology
  - freeze + mortar & pestle
  - Enzymatic
  - Heat, Freeze/thaw

**Extraction Buffer** (denature proteins)

**Biological Fluids, Media, other solution preps (i.e. IP’s)**
- MW Spin filter
- Dialysis
- Precipitation

**Disruption**
- Chaotropes (Disorder maker)
  - Urea, thiourea, guanidine HCl, salts (KCl, NaCl, etc.), MeOH, acetonitrile

**Detergents**
- Anionic (SDS), nonionic (NP40, triton x-100), zwitterionic (CHAPS)

**Buffer reagent – pH considerations**
- Tris, HEPES, ammonium bicarbonate, PBS, TEAB

**Reducing reagents**
- Dithiothreitol (DTT), TCEP (tris(2-carboxyethyl)phosphine)

**Common Extraction Buffers:**
- 4% SDS, 100mM Tris pH 7.6, 100mM DTT
- 7M urea, 2M thiourea, 0.4M TEAB pH 8.5, 20% acetonitrile, 4mM TCEP

Centrifuge insoluble material out & aspirate supernatant (protein extract)

**Depletion** of high abundant proteins typically done before digestion

**Enrichment** of post-translational modifications typically done after digestion

© 2015 Regents of the University of Minnesota. All rights reserved.
Dynamic Range Problem In Samples – A Case for Depletion


Beckman Coulter IgY-12 Protein Partitioning Column

12 Proteins comprise up to 96% of the protein mass in plasma.

Pool containing enriched Medium- and Low-Abundance Proteins

Protein Partitioning – Next Generation Protein Depletion

Protein depletion has been used for some years to remove most of the albumin and/or IgG from biofluids such as plasma and serum prior to analysis, but it is clear that this alone is insufficient to enable progress to be made in biomarker discovery. The presence of highly abundant proteins significantly complicates the discovery process by masking the presence and limiting the detection of low abundance species. ProteomeLab IgY partitioning addresses this issue by reversibly capturing the more abundant proteins from human biofluids such as plasma and serum, yielding an enriched pool of low abundance proteins for further study. The captured proteins can also be easily recovered for investigation if required - hence the term partitioning rather than depletion.

Current Agilent Multiple Affinity Removal System (MARS)

- Anti-Albumin-resin
- Anti-Transferrin-resin
- Anti-Haptoglobin-resin
- Anti-α-1-antitrypsin-resin
- Anti-IgA-resin
- Anti-IgG-resin

Apply Crude Human Serum

Buffer A  Buffer B

Low-Abundant Proteins Free from Interferences
High-Abundant Proteins

Individual Ab materials are mixed in selected percentages and packed into a column format. Cost of the product is driven by cost of the antigen.

Reference: Agilent web seminar slides
Sigma-Aldrich Human IgY14 and SuperMix Columns

1DE of IgY 14/SuperMix Column Samples

M: MW Marker
P: Plasma
F1: Flow-Through of IgY 14
E: Bound/Eluted Fraction from SuperMix
F2: Flow-Through of IgY 14 + SuperMix

4-20% SDS-PAGE under reducing conditions

4μg Protein loaded per lane


Enrichment - Post Translational Modifications (PTM’s)

Why Enrich for PTM of interest?
In most cases, stoichiometric ratio of PTM species to unmodified species is extremely low.

Jensen Nature Reviews Molecular Cell Biology 7, 391-403 (June 2006) | doi:10.1038/nrm1939
Phosphoproteomic Enrichment

PhosphoSerine
\[ \text{C}_2\text{H}_6\text{NPO}_5 \]

PhosphoThreonine
\[ \text{C}_4\text{H}_8\text{NPO}_5 \]

PhosphoTyrosine
\[ \text{C}_2\text{H}_10\text{NPO}_5 \]

Reference: TE Thingholm, Proteomics 2009, 9, 1451–1468

http://www.ionsource.com/Card/phos/phos.htm
Acetylation Enrichment with Anti-Acetyl Lysine Antibody

IP Method Considerations

Reference: piercenet.com
Proximity Labeling – Biotin Ligase Fusion Protein

BioID/TurboID

Comparison of BioID and APEX proximity labeling methods. Both methods utilize biotin affinity capture with streptavidin and purification strategies. (BioID image adapted from Roux et al., 2012 Figure 1A and APEX image adopted from Rhoe et al., 2013 Figure 1A)

APEX

## In-Solution vs. In-Gel Strategies – Dynamic Range

### In-Solution Prep, 11 proteins

<table>
<thead>
<tr>
<th>#</th>
<th>Value</th>
<th>Accession Number</th>
<th>Molecular Weight</th>
<th>Protein Grouping Abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Serum albumin precursor (Allergen Bos d 6) (BSA) cRAP</td>
<td>gi</td>
<td>1351907...</td>
<td>69 kDa</td>
</tr>
<tr>
<td>2</td>
<td>keratin, type II cytoskeletal 1 [Homo sapiens]</td>
<td>gi</td>
<td>1193957...</td>
<td>66 kDa</td>
</tr>
<tr>
<td>3</td>
<td>keratin, type I cytoskeletal 9 [Homo sapiens]</td>
<td>gi</td>
<td>5595689...</td>
<td>62 kDa</td>
</tr>
<tr>
<td>4</td>
<td>Trypsin precursor cRAP</td>
<td>gi</td>
<td>126429...</td>
<td>24 kDa</td>
</tr>
<tr>
<td>5</td>
<td>alpha-2-HS-glycoprotein proproteins [Homo sapiens]</td>
<td>gi</td>
<td>1565339...</td>
<td>39 kDa</td>
</tr>
<tr>
<td>6</td>
<td>keratin, type I cytoskeletal 10 [Homo sapiens]</td>
<td>gi</td>
<td>1959728...</td>
<td>59 kDa</td>
</tr>
<tr>
<td>7</td>
<td>hemoglobin subunit alpha [Homo sapiens]</td>
<td>gi</td>
<td>4504345...</td>
<td>15 kDa</td>
</tr>
<tr>
<td>8</td>
<td>keratin, type I cytoskeletal 14 [Homo sapiens]</td>
<td>gi</td>
<td>1543131...</td>
<td>52 kDa</td>
</tr>
<tr>
<td>9</td>
<td>alpha-2-macroglobulin precursor [Homo sapiens]</td>
<td>gi</td>
<td>6692929...</td>
<td>163 kDa</td>
</tr>
<tr>
<td>10</td>
<td>inter-alpha-trypsin inhibitor heavy chain HZ precursor [Homo sapiens]</td>
<td>gi</td>
<td>7077891...</td>
<td>106 kDa</td>
</tr>
<tr>
<td>11</td>
<td>alpha-fetoprotein precursor [Homo sapiens]</td>
<td>gi</td>
<td>4501989...</td>
<td>69 kDa</td>
</tr>
</tbody>
</table>

### In-Gel Prep, 401 proteins

<table>
<thead>
<tr>
<th>#</th>
<th>Value</th>
<th>Accession Number</th>
<th>Molecular Weight</th>
<th>Protein Grouping Abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Serum albumin precursor (Allergen Bos d 6) (BSA) cRAP</td>
<td>gi</td>
<td>1351067...</td>
<td>58 kDa</td>
</tr>
<tr>
<td>2</td>
<td>Alpha-2-macroglobulin OS=Bos taurus sp</td>
<td>QYS1H...</td>
<td>168 kDa</td>
<td>151</td>
</tr>
<tr>
<td>3</td>
<td>Alpha-2-HS-glycoprotein OS=Bos taurus sp</td>
<td>P12763...</td>
<td>151 kDa</td>
<td>431</td>
</tr>
<tr>
<td>4</td>
<td>Serotransferrin OS=Bos taurus G. sp</td>
<td>SP9443...</td>
<td>70 kDa</td>
<td>629</td>
</tr>
<tr>
<td>5</td>
<td>Complement C3 OS=Bos taurus G. sp</td>
<td>Q2YUK...</td>
<td>187 kDa</td>
<td>473</td>
</tr>
<tr>
<td>6</td>
<td>Alpha-1-antitrypsin OS=Bos taurus sp</td>
<td>P34955...</td>
<td>46 kDa</td>
<td>416</td>
</tr>
<tr>
<td>7</td>
<td>Uncharacterized protein OS=Bos taurus sp</td>
<td>b</td>
<td>E1BH9R...</td>
<td>51 kDa</td>
</tr>
<tr>
<td>8</td>
<td>Inter-alpha-trypsin inhibitor heavy chain HZ precursor [Homo sapiens]</td>
<td>gi</td>
<td>7077891...</td>
<td>106 kDa</td>
</tr>
<tr>
<td>9</td>
<td>Apolipoprotein A-I OS=Bos taurus sp</td>
<td>P15749...</td>
<td>30 kDa</td>
<td>285</td>
</tr>
<tr>
<td>10</td>
<td>Hemoglobin fetal subunit beta OS=Bos taurus sp</td>
<td>P02681...</td>
<td>15 kDa</td>
<td>285</td>
</tr>
<tr>
<td>11</td>
<td>Inter-alpha-trypsin inhibitor heavy chain HZ sp</td>
<td>P56652...</td>
<td>106 kDa</td>
<td>263</td>
</tr>
<tr>
<td>12</td>
<td>Inter-alpha-trypsin inhibitor heavy chain HZ sp</td>
<td>Q3T052...</td>
<td>102 kDa</td>
<td>230</td>
</tr>
<tr>
<td>13</td>
<td>Uncharacterized protein OS=Bos taurus sp</td>
<td>b</td>
<td>E1BH96...</td>
<td>192 kDa</td>
</tr>
<tr>
<td>14</td>
<td>Hemoglobin subunit alpha OS=Bos taurus sp</td>
<td>P03663...</td>
<td>15 kDa</td>
<td>285</td>
</tr>
<tr>
<td>15</td>
<td>Uncharacterized protein OS=Bos taurus sp</td>
<td>b</td>
<td>F1M118...</td>
<td>192 kDa</td>
</tr>
<tr>
<td>16</td>
<td>Transthyretin OS=Bos taurus G. sp</td>
<td>SP04637...</td>
<td>15 kDa</td>
<td>198</td>
</tr>
<tr>
<td>17</td>
<td>Alpha-fetoprotein OS=Bos taurus sp</td>
<td>Q35757...</td>
<td>69 kDa</td>
<td>183</td>
</tr>
<tr>
<td>18</td>
<td>Plasminogen OS=Bos taurus G. sp</td>
<td>tr</td>
<td>E1B726...</td>
<td>91 kDa</td>
</tr>
<tr>
<td>19</td>
<td>Alpha-1-microglobulin OS=Bos taurus sp</td>
<td>Q1HMK...</td>
<td>39 kDa</td>
<td>163</td>
</tr>
<tr>
<td>20</td>
<td>Fetuin B OS=Bos taurus G. sp</td>
<td>FETU...</td>
<td>43 kDa</td>
<td>138</td>
</tr>
</tbody>
</table>

---

**Probability Legend:**
- **over 95%**
- 80% to 94%
- 60% to 79%
- 50% to 79%
- 20% to 49%
- 0% to 19%

---

*University of Minnesota Center for Mass Spectrometry and Proteomics | Phone: (612) 625-2280 | Fax: (612) 625-2279*
INTRO: Protein Quantitation Strategies at the Peptide Level by Mass Spectrometry†

• **Discovery-based** (complex mixture)
  – Stable isotope incorporation
    • iTRAQ® † *
    • TMT † – tandem mass tag
    • SILAC **
  – ‘Label free’
    • Spectral counting
    • Peptide Ion Peak Intensity: XIC-based (extracted ion chromatogram)

• **Targeted analyses**
  – Select peptides of interest
  – Internal standard for absolute quantitation

† Instrument-specific capabilities required
* iTRAQ® Isobaric Tag for Relative and Absolute Quantitation
** SILAC Stable isotope labeling with amino acids in cell culture
iTRAQ® 8-Plex Reagent Chemical Structure

Isobaric Tag
Total mass = 305

Reporter Group
113 – 119, 121 m/z

Balance Group (?)
Mass 184, 186 – 192 m/z

Amine specific peptide reactive group (NHS)
N-hydroxysuccinimide

This will react with primary amines...need to ensure buffer components are compatible or cleaned-up!

Applied Biosystems has granted permission to use this slide.
Isobaric Tags for Relative & Absolute Quantitation (iTRAQ®)

Example: Compare Relative Protein Expression Levels in **Healthy vs Disease** Tissues

**Experimental Design for 4-plex iTRAQ® Experiment**

1. **Obtain protein-containing sample, extract protein**
   - Obtain protein-containing sample, extract protein

2. **Proteolytic Digestion**
   - Reduce, alkylate Cysteines
   - Trypsin Digest

3. **Label peptides with iTRAQ® Reagents**
   - iTRAQ TAG 114
   - iTRAQ TAG 115
   - iTRAQ TAG 116
   - iTRAQ TAG 117

4. **MIX**

5. **2D LC-MS/MS**

Try to break down sample complexity

1st Dimension offline HPLC
High pH C18 RP

2nd Dimension HPLC
Inline w/Mass Spec.
Low pH C18 RP

SILAC Metabolic Labeling Experimental Workflow

http://www.piercenet.com/method/quantitative-proteomics
SILAC Experiment: Proteome Dynamics of *B. subtilis* in Response to Two Nutritional Challenges

**growth on succinate**

"normal AA"  
Lys-$^{12}$C$_6$$^{14}$N$_2$

"heavy AA" (+8Da)  
Lys-$^{13}$C$_6$$^{15}$N$_2$

Treated cells  
(succinate or low P)

Control cells

Lyse and Combine 1:1

SILAC incorporation check

adapted from Soufi B et al; J. Proteome Res. 2010, 9, 3638-3646. Copyright 2010 ACS

© 2015 Regents of the University of Minnesota. All rights reserved.
Label Free and Targeted Quantification Prep

Label-free Quantitation

Targeted

Metabolic Labeling

Spiked Heavy Peptides

adapted from http://www.piercenet.com/method/quantitative-proteomics
WHAT is MASS SPECTROMETRY?

Mass spectrometry is a powerful analytical technique that is used to identify unknown compounds, to quantify known compounds, and to elucidate the structure and chemical properties of molecules.

- Identify structures of biomolecules, such as carbohydrates, nucleic acids and steroids
- Sequence biopolymers such as proteins and oligosaccharides
- Determine how drugs are used by the body
- Perform forensic analyses such as conformation and quantitation of drugs of abuse
- Analyze for environmental pollutants
- Determine the age and origins of specimens in geochemistry and archaeology
- Identify and quantitate compounds of complex organic mixtures
- Perform ultrasensitive multielement inorganic analyses

(Ref: http://www.asms.org/whatisms)
Goals Attainable by MS

• Small molecule MW & structural composition
• Intact protein molecular weight
• Peptide mass
• Peptide sequence $\rightarrow$ protein ID
• Identification and location of post-translational modifications of amino acids
• de novo Sequencing of unknown proteins
• Relative Quantification of Proteins
History

1803  John Dalton proposes theory that each element has distinct, measurable atomic weight*

1905  J. J. Thomson produces the first mass spectrum (e/m) (Cambridge University)*

1919  Francis Aston discovers elements have stable isotopes (Cambridge University)*

2002  John Fenn (Virginia Commonwealth) and Koichi Tanaka (Shimadzu Corp., Japan) joint Chemistry Nobel Prize winners for ionization techniques ES and MALD

* Source: Measuring Mass: from positive rays to proteins; Grayson MS, ed. Chemical Heritage Press 2002
How does a mass spectrometer work?

- Ion Source: Makes ions
- Mass Analyzer: Separates Ions
- Detector: Presents information as a Mass Spectrum

So for a mass spectrometer to work, the molecules must be in the gas phase and ionized so they can be mass separated
Mass Spec Principles

Figure: Dr. David Wishart, University of Alberta
Electron Impact (EI) and Chemical Ionization (CI) Spectra of Ephedrine

Reference: asms.org
Mass Spectrometry Data: $m/z$

- **Mass ($m$):**
  - calibrate instrument with standard compound (molecule with a known MW)

- **Charge ($z$):**
  - Calculate CHARGE STATE OF AN ION from peak spacing in RESOLVED ISOTOPE SERIES/ENVELOPE
  - MALDI mainly produces singly charged ions - easy to interpret

**SINGLY CHARGED**

![Singly Charged Spectrum](image)

**DOUBLY CHARGED**

![Doubly Charged Spectrum](image)
# Atomic Masses and Abundances for a Subset of Naturally Occurring Biologically Relevant Isotopes

<table>
<thead>
<tr>
<th>Iso</th>
<th>A</th>
<th>%</th>
<th>Iso</th>
<th>A+1</th>
<th>%</th>
<th>Iso</th>
<th>A+2</th>
<th>%</th>
<th>Iso</th>
<th>A+3</th>
<th>%</th>
<th>Iso</th>
<th>A+4</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>^12C</td>
<td>12</td>
<td>98.93(8)</td>
<td>^13C</td>
<td>13.0033458378(10)</td>
<td>1.07(8)</td>
<td>^14C</td>
<td>14.003241988(4)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>^1H</td>
<td>1.0078250321(4)</td>
<td>99.9885(70)</td>
<td>^2H</td>
<td>2.0141017780(4)</td>
<td>0.0115(70)</td>
<td>^3H</td>
<td>3.0160492675(11)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>^14N</td>
<td>14.0030740052(9)</td>
<td>99.6327(7)</td>
<td>^15N</td>
<td>15.0001088984(9)</td>
<td>0.368(7)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>^16O</td>
<td>15.9949146221(15)</td>
<td>99.757(16)</td>
<td>^17O</td>
<td>16.99913150(22)</td>
<td>0.038(1)</td>
<td>^18O</td>
<td>17.9991604(9)</td>
<td>0.205(14)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>^32S</td>
<td>31.97207069(12)</td>
<td>94.93(31)</td>
<td>^33S</td>
<td>32.97148580(12)</td>
<td>0.76(2)</td>
<td>^34S</td>
<td>33.96786683(11)</td>
<td>4.29(28)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>^36S</td>
<td>35.96708088(25)</td>
<td>0.02(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>^19F</td>
<td>18.99840320(7)</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>^23Na</td>
<td>22.98979667(23)</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>^39K</td>
<td>38.9637069(3)</td>
<td>93.2581(44)</td>
<td>^40K</td>
<td>39.96399867(29)</td>
<td>0.0117(1)</td>
<td>^41K</td>
<td>40.96182597(28)</td>
<td>6.730(244)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>^31P</td>
<td>30.97366151(20)</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>^35Cl</td>
<td>34.96885271(4)</td>
<td>75.781(4)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>^37Cl</td>
<td>36.96590260(5)</td>
<td>24.22(4)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>^55Mn</td>
<td>54.93834206(14)</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>^54Fe</td>
<td>53.9396148(14)</td>
<td>5.845(35)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>^56Fe</td>
<td>55.9349421(15)</td>
<td>91.75(36)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>^63Cu</td>
<td>62.9296011(15)</td>
<td>69.17(3)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>^65Cu</td>
<td>64.9277937(19)</td>
<td>30.83(3)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>^78Br</td>
<td>78.91833762(70)</td>
<td>50.69(7)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>^81Br</td>
<td>80.916291(3)</td>
<td>49.31(7)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>^127I</td>
<td>126.904468(4)</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>


**Original Literature Source:**

The above table has been double checked against the reference material. If you wish to validate this table against the original IUPAC values please send your findings to contributions@ionsource.com
Peaks in an Isotope Series have Identical Structural Formulas but Different Isotope Components

$^{12}\text{C, }^{14}\text{N, }^{16}\text{O, etc – containing ions}$

$^{13}\text{C, }^{15}\text{N, }^{18}\text{O, etc – containing ions}$
Mass Spectrometry: Resolution
Monoisotopic vs. Average $m/z$

“Mass resolution is the dimensionless ratio of the mass of the peak divided by its width. Usually, the peak width is taken as the full width at half maximum intensity, (FWHM).”

The Resolving Power of a Mass Spectrometer Dictates the Accuracy of the \( m/z \) Values Produced

<table>
<thead>
<tr>
<th>( m/z ) Value</th>
<th>tolerance</th>
<th>accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>± 1</td>
<td>1000 ppm</td>
</tr>
<tr>
<td>1000.3</td>
<td>± 0.1</td>
<td>100 ppm</td>
</tr>
<tr>
<td>1000.345</td>
<td>± 0.001</td>
<td>1 ppm</td>
</tr>
</tbody>
</table>

RESOLUTION = 8000

![Mass Spectrum Graph](image_url)
Mass Accuracy Calculations

Accuracy in ppm (parts per million):

\[
\left( \frac{\Delta}{\text{theoretical value}} \right) \times 10^6 = \text{ppm error}
\]

Accuracy in Percent

\[
\left( \frac{\Delta}{\text{theoretical value}} \right) \times 100 = \% \text{ error}
\]

\( \Delta = \) experimental/observed value (mass spec data) – theoretical value

300 ppm is equivalent to 0.03 %
Comparison of Resolving Power: 645.7 m/z

FT-ICR MS
(Fourier Transform-Ion Cyclotron Resonance)

Quadrupole-TOF MS
(Time of Flight)
MALDI-TOF: Peptide Mass Fingerprint

In-Gel Tryptic Digest: GOAL = Protein Identification

![MALDI-TOF Mass Spectrum Graph]
ESI-Ion Trap MS: Product Ion Spectrum

MS/MS Spectrum from [M + 2H^2], Precursor m/z = 467.8 (MW = 933)
MALDI-TOF MS of Intact Protein:
Unoxidized Calmodulin
MALDI-TOF MS of Intact Protein:
Mixture of Calmodulin with Multiply Oxidized Methionines

16817, 6 ox-Met
16832, 7 ox-Met
16849, 8 ox-Met

E. Balog, Univ MN, BMBB
Intact Protein ESI-MS: Cytochrome C

1) Raw MS data

2) Protein MW (after mathematical deconvolution)

Mathematical deconvolution
Mass Spectrometry Principles

Sample

Ionizer + Mass Analyzer

Detector

Figure: Dr. David Wishart, University of Alberta
IONIZATION

• Mass spectrometers can detect the presence of *IONS* (+ or −)
• Ions must be in the *GAS PHASE*
  • Ions must be free from salts, solvents...
• Ionization from *SOLID CRYSTALS*
  • Destroys non-covalent interactions
  • **MALDI** *(Matrix Assisted Laser Desorption Ionization)*; typically produces singly charged ions
• Ionization from *LIQUID STATE*:
  • **ESI** *(Electrospray Ionization)*; typically produces multiply charged ions
  • Non-covalent interactions are retained
  • Compatible with in-line HPLC
    • solvents must be volatile- e.g. ammonium acetate, ACN, methanol
• Not all ionization modes are applicable to a given molecule
Electron Impact Ionization

A schematic diagram of a “Nier-type” ion source is shown in Figure 3. Electrons are produced by heating a metal filament, usually formed from a fine wire or ribbon of tungsten or rhenium. These electrons are accelerated by a potential difference between the filament and the ions source box, pass through entrance and exit apertures in the box, and are collected by a trap electrode. Voltage is applied to a repeller electrode within the box that accelerates ions toward the ion exit aperture. A collimating magnetic field is applied parallel to the electron beam axis, and the field strength is chosen to provide high transmission of the electrons with minimal perturbation of the ion beam. A field on the order of 100 G is typically used. The electron current is stabilized by monitoring either the total emission current or the current to the trap with feedback to the filament heater power supply to maintain a constant current. Samples are introduced directly into the ion source box, and for most analytical applications, the vapor input rate and the orifices in the source box must be chosen so that the probability of ion–molecule reactions is minimal. In the “chemical ionization” (CI) source, discussed

Odd-electron ions (radical cations) are formed during Electron Impact
Ionization Techniques for MS

- Electron Impact (EI)
- Chemical Ionization (CI)
- Inductively Coupled Plasma (ICP)
- Field Desorption (FD)
- Fast Atom Bombardment (FAB)
- Thermospray
- Desorption Electrospray Ionization (DESI)
- Direct Analysis in Real Time (DART)
- Atmospheric Pressure Chemical Ionization (APCI)
- Secondary Ion Mass Spectrometry (SIMS)
- Matrix-assisted Laser Desorption (MALDI)
- Electrospray (ESI)
Technology Breakthroughs (1990’s) in Mass Spectrometry Accelerated Proteomics Studies

Ionization Methods suitable for Peptides/Proteins
(i.e. methods for producing gas phase ions):

• **Electrospray Ionization:** John Fenn (Virginia Commonwealth) 2002 Nobel prize Chemistry

• **MALDI** (matrix-assisted laser desorption ionization): Koichi Tanaka (Shimadzu Corp) 2002 Nobel prize Chemistry
Soft Ionization Methods

MALDI
- Unlimited analyte mass
- Insensitive to moderate salts
- Excellent for mixtures
- Off-line sample prep

ESI
- Unlimited analyte mass
- Some sample cleanup
- Mixtures problematic
- On-line sample injection

Adapted from Dr. David Wishart, University of Alberta
API-Electrospray Mechanism

Mechanisms of Ionization

- Ion Evaporation: Electric field charge liquid → charged droplets → Desolvation to reach high field strengths (10^8 V/cm^2) → Ionization
**APCI Mechanisms**

Mechanisms of Ionization

- Vaporation $\rightarrow$ Solvent ionized $\rightarrow$ Charge transfer to analyte

\[
\begin{align*}
N_2 + e^- & \rightarrow N_2^{+\cdot} + 2e^- \\
N_2^{+\cdot} + 2N_2 & \rightarrow N_4^{+\cdot} + N_2 \\
N_4^{+\cdot} + H_2O & \rightarrow H_2O^{+\cdot} + 2N_2 \\
H_2O^{+\cdot} + H_2O & \rightarrow H_3O^+ + OH^- \\
H_3O^+ + M & \rightleftharpoons [M+H]^+ + H_2O
\end{align*}
\]
DESI
Desorption Electrospray Ionization

Reference: Prosilia Omni Spray™ Ion Source manual
Instrumentation

Multiple types of Mass Analyzers are used for a variety of purposes
MALDI with Time-of-Flight (TOF) MS

velocity of the ion in the flight tube depends on the mass-to-charge ratio

Target: **Analyte Ions** + “**matrix**” crystals
Apply Accelerating Potential (19kV)
MALDI-TOF MS: basic principles

Detection

Separation

Acceleration

Ionization

Desorption

\[ \frac{m}{z} = \frac{2eU}{L^2} t^2 \]

- \( m \): mass
- \( z \): charge
- \( U \): acceleration voltage
- \( L \): path length
- \( t \): time
- \( e \): elementary charge

Image source: http://www.anagnostec.de/index/modul/portal/kernwert/technology/
The Time an Ion Spends in Flight

Summary: Time of flight of the ion varies with the square root of its mass-to-charge ratio.

$$t = \Delta x \sqrt{\frac{m}{2zeU_{ex}}}$$

$$RP = \frac{m}{\Delta m} = \frac{t}{2\Delta t}$$
QUADRUPOLE Mass Filter

Reference: http://www.chm.bris.ac.uk/ms/theory/quad-massspec.html
Quadrupole Mass Analyzer

• Uses a combination of RF and DC voltages to operate as a mass filter
• Has four parallel metal rods with alternating positive and negative voltages
• Can scan through all masses or sit at one fixed mass
Comparison of Linear Quadrupole to 3-D Quadrupole (Ion Trap)

Quadrupole Mass Filter

Quadrupole Ion Trap

3D Ion Trap

The ion trap is an energy well. Ions with sufficient energy to enter the trap are retained by an energy barrier on the exit side of the trap. The advantage of the ion trap is that it accumulates selected ions prior to their analysis giving it high initial sensitivity. Ions can be fragmented by collision with helium gas and their product ions analyzed within the trap. Selected product ions can undergo further fragmentation, thus allowing MSn. The ion trap has a high efficiency of transfer of fragment ions to the next stage of fragmentation (unlike the triple quadrupole instrument).

The 3D ion trap consists of two end cap electrodes and a center ring electrode.
ION TRAP MASS ANALYZER SCHEMATIC

Figure from: http://www.matrixscience.com/help/ion_trap_main_help.html
HYBRID QUADRUPOLE TIME-OF-FLIGHT MASS SPECTROMETER

Figure 1. Schematic diagram of the tandem QqTOF mass spectrometer.

Reference: Chernushevich IV et al. J. Mass Spectrom. 2001; 36:
LTQ Orbitrap Velos Schematic

Figure 1-2. Schematic view of the LTQ Orbitrap Velos

\[ \omega = \sqrt{\frac{k}{m \cdot z}} \]

http://en.wikipedia.org/wiki/Orbitrap
### Outline

- Liquid Chromatography
- Electrospray
- MS Acquisition
- Peptide Fragmentation
- PTM Detection
- Protein ESI-MS

### Terminology

- ESI
- Liquid chromatography
- Acquisition
- DDA (data dependent acquisition)
- MS1 vs MS2
- Precursor ion
- MS/MS
- Product Ions
- b-ion, y-ion
EXAMPLE Workflow:
MIXTURE OF UNKNOWN PROTEINS

A. Obtain protein-complex
B. Separate proteins on 1D gel
C. Excise protein bands > trypsin digestion
D. Liquid Chromatography-ESI-Mass Spectrometry

Gingras A et al, 2007 Nat Rev Mol Cell Biol 8, 645
Example Sample Analysis Workflow

Liquid Chromatography (LC): Peptide Separation

Electrospray Ionization (ESI)

Mass Spectrometry (MS)
Liquid Chromatography: Separation Technique

Mixtures are separated or partially separated before MS analysis with Chromatographic Technologies
Liquid Chromatography Basic Overview

“... modern Liquid Chromatography (LC), uses a liquid mobile phase to transport the sample components through a column packed with a solid material - the stationary phase.” Reference: http://www.earl2learn.com

http://www.chemistry.adelaide.edu.au/external/soc-rel/content/lc-col.htm

Mikhail Tswett (1872 – 1919)

From Tswett’s notebook (1910) on the early chromatographic experiments: plant pigments were passed through calcium carbonate using petroleum ether
• Peaks represent analyte elution profiles

• Increased Retention Time = increased peptide hydrophobicity on a C18 column
• **Column 1** Elution Profile
• Use fraction collector – collect peptides in separate tubes

Liquid Chromatography, Peptide Elution Profiles
2D LC

**Column 2**: LC-MS Peptide Elution Profiles
EXAMPLE Workflow:
MIXTURE OF UNKNOWN PROTEINS

Typical Numbers of Proteins Identified:

- 1D LC-MS: up to 800-1000
- 2D LCMS: up to 6000

(sample and dynamic range dependent)
Example Sample Analysis Workflow

Liquid Chromatography (LC): Peptide Separation

Electrospray Ionization (ESI)

Mass Spectrometry (MS)
Electrospray Ionization (ESI)
Produce Analyte Ions

GOAL: eliminate solvent, get analyte into the gas phase; apply high voltage to a liquid to create aerosol

Electrospray Ionization

Ref: IJAC Vol 2012 ID 282574
Example Sample Analysis Workflow

Liquid Chromatography (LC): Peptide Separation

Electrospray Ionization (ESI)

Mass Spectrometry (MS)
EXAMPLE:
MIXTURE OF UNKNOWN PROTEINS

Data-Dependent Acquisition Scan Mode (DDA)
on the Mass Spectrometer
MS1 Data Acquisition

- MS1 spectrum (below)
- Peaks below represent unfragmented peptide m/z values

Mass Spectrum at 39.22 minutes shows co-eluting peptides
Data Dependent Acquisition:
• Top 6 (most abundant) Peaks Identified
• Next 6 Scan Events are MS/MS

Sequential MS/MS scan events are triggered for the 6 most abundant peaks
MS2 Data Acquisition (1\textsuperscript{st} peak triggered in DDA)

MS/MS spectrum

Precursor 851.95 \( m/z \)

good_hamel_011713_12503_chymo_tubA_025dda #4329

RT: 39.20  AV: 1  NL: 7.09E6

T: FTMS + c NSI d Full ms2 851.95@hcd40.00 [111.00-1715.00]

peptide fragment ions peak
Tandem Mass Spectrometry (MS/MS)

- Select Precursor Peptide
- Isolate Precursor ions in Collision Cell
- Fragment precursor ions with Collision Induced Dissociation (CID)
- Measure \( m/z \) of Product ions
Collision Induced Dissociation (CID or HCD)

Q1
Quadrupole Mass Filter

+LLLYSSQICK+

q2
Collision Cell

Collision Induced Dissociation
1) Electric potential applied to collision cell
2) Ions accelerated to high kinetic energy
3) Ions collide with neutral gas molecule
4) Kinetic energy is converted to internal energy
5) Chemical bond breakage occurs

Tandem Mass Spectrometry
1st measurement (MS1) = Intact Peptide m/z
2nd measurement (MS2) = Peptide fragment ion m/z values

N2 collision gas

Measure m/z values of product ions

Precursor ion 612.8 m/z (LLLYSSQICK) filtered in Q1
Predictable Fragment Ions Types from Peptide Dissociation

- Peptide Backbone has 3 Bond Types (peptide bond is the weakest of the 3 bonds)
- Bond Breakage Yields Complimentary Ion Types, e.g., b- and y-type
Peptide Fragment Ions (or product ions)

Peptide MQIFVKTLTK
- 604.8572 \( m/z \) precursor ion, monoisotopic
- 1208.7077 Intact Peptide Mass, monoisotopic

<table>
<thead>
<tr>
<th>( m/z )</th>
<th>b series</th>
<th>y series</th>
<th>m/z</th>
</tr>
</thead>
<tbody>
<tr>
<td>( b_9 )</td>
<td>1062.6</td>
<td>MQIFVKTLT</td>
<td>K 147.1</td>
</tr>
<tr>
<td>( b_8 )</td>
<td>961.6</td>
<td>MQIFVKTL</td>
<td>TK 248.1</td>
</tr>
<tr>
<td>( b_7 )</td>
<td>848.5</td>
<td>MQIFVKT</td>
<td>LTK 361.3</td>
</tr>
<tr>
<td>( b_6 )</td>
<td>747.4</td>
<td>MQIFVK</td>
<td>TLTK 462.3</td>
</tr>
<tr>
<td>( b_5 )</td>
<td>619.3</td>
<td>MQIFV</td>
<td>KTLTK 590.4</td>
</tr>
<tr>
<td>( b_4 )</td>
<td>520.3</td>
<td>MQIF</td>
<td>VKTLTK 689.5</td>
</tr>
<tr>
<td>( b_3 )</td>
<td>373.2</td>
<td>MQI</td>
<td>FVKTLTK 836.5</td>
</tr>
<tr>
<td>( b_2 )</td>
<td>260.1</td>
<td>MQ</td>
<td>IFVKTLTK 949.6</td>
</tr>
<tr>
<td>( b_1 )</td>
<td>132.0</td>
<td>M</td>
<td>QIFVKTLTK 1077.7</td>
</tr>
</tbody>
</table>
Peptide Tandem MS can provide Unambiguous Evidence for Location of Amino Acid Modifications

Rule: Experimental data must contain fragment ions that provide site localization evidence
Tandem MS (or MS/MS) for Identification of Amino Acid Post-translational Modification (PTM) Site

MQIFVKTLTK  $\text{MW}_{\text{mono}} = 1208.7077$
MQIFVKTpLTK  $\text{MW}_{\text{mono}} = 1288.6740$ (phosphorylation $+80 \text{ Da} = \text{HPO}_3$)

<table>
<thead>
<tr>
<th>$m/z$</th>
<th>b series</th>
<th>y series</th>
<th>$m/z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1062.6</td>
<td>MQIFVKT</td>
<td>K</td>
<td>147.1</td>
</tr>
<tr>
<td>961.6</td>
<td>MQIFVK</td>
<td>TK</td>
<td>248.1</td>
</tr>
<tr>
<td>848.5</td>
<td>MQIFV</td>
<td>LTK</td>
<td>361.3</td>
</tr>
<tr>
<td>747.4</td>
<td>MQIFV</td>
<td>TLTK</td>
<td>462.3</td>
</tr>
<tr>
<td>619.3</td>
<td>MQIFV</td>
<td>KTLTK</td>
<td>590.4</td>
</tr>
<tr>
<td>520.3</td>
<td>MQIFV</td>
<td>VKTLTK</td>
<td>689.5</td>
</tr>
<tr>
<td>373.2</td>
<td>MQIFV</td>
<td>FVKTLTK</td>
<td>836.5</td>
</tr>
<tr>
<td>260.1</td>
<td>MQIFV</td>
<td>IFVKTTLTK</td>
<td>949.6</td>
</tr>
<tr>
<td>132.0</td>
<td>MQIFV</td>
<td>QIFVKTTLTK</td>
<td>1077.7</td>
</tr>
</tbody>
</table>

Mass shifts will occur only in fragments containing the phos-Thr, therefore location of MODIFICATION can be pinpointed
### Predictable Mass Shifts for Phosphorylated Fragments

**MQIFVKTpLTK**

<table>
<thead>
<tr>
<th>Mass Shifts (+80)</th>
<th>Fragment</th>
<th>b series</th>
<th>y series</th>
<th>Mass Shifts (+80)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1062.6</td>
<td>MQIFVKTLT</td>
<td></td>
<td>K 147.1</td>
<td></td>
</tr>
<tr>
<td>961.6</td>
<td>MQIFVKTL</td>
<td>TK 248.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>848.5</td>
<td>MQIFVKT</td>
<td>LTK 361.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>747.4</td>
<td>MQIFVK</td>
<td>TLTK 462.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>619.3</td>
<td>MQIFV</td>
<td>KTLTK 590.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>520.3</td>
<td>MQIF</td>
<td>VKTTLTK 689.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>373.2</td>
<td>MQI</td>
<td>FVKTTLTK 836.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>260.1</td>
<td>MQ</td>
<td>IFVKTTLTK 949.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>132.0</td>
<td>M</td>
<td>QIFVKTTLTK 1077.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mass shifts will occur only in fragments containing the **phos-Thr**, therefore location of MODIFICATION can be pinpointed.
Predictable Mass Shifts for Phosphorylated Fragments

MQIFVKTpLTK (+80 = HPO$_3$); MW$_{mono}$ = 1288.67

<table>
<thead>
<tr>
<th>m/z</th>
<th>b series</th>
<th>y series</th>
<th>m/z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1142.6</td>
<td>MQIFVKTpLTK</td>
<td>K 147.1</td>
<td></td>
</tr>
<tr>
<td>1041.5</td>
<td>MQIFVKTpL</td>
<td>TK 248.1</td>
<td></td>
</tr>
<tr>
<td>928.4</td>
<td>MQIFVKTP</td>
<td>LTK 361.3</td>
<td></td>
</tr>
<tr>
<td>747.4</td>
<td>MQIFVKTp</td>
<td>TpLTK 542.3</td>
<td></td>
</tr>
<tr>
<td>619.3</td>
<td>MQIFVKTp</td>
<td>KTpLTK 670.4</td>
<td></td>
</tr>
<tr>
<td>520.3</td>
<td>MQIFVKTp</td>
<td>VKTpLTK 769.4</td>
<td></td>
</tr>
<tr>
<td>373.2</td>
<td>MQIFVKTp</td>
<td>FVKTpLTK 919.5</td>
<td></td>
</tr>
<tr>
<td>260.1</td>
<td>MQIFVKTp</td>
<td>IFVKTpLTK 1029.6</td>
<td></td>
</tr>
<tr>
<td>132.0</td>
<td>MQIFVKTp</td>
<td>QIFVKTpLTK 1157.6</td>
<td></td>
</tr>
</tbody>
</table>

Mass shifts will occur only in fragments containing the phos-Thr (bold), therefore location of MODIFICATION can be pinpointed
EXAMPLE: MS2 Spectrum and Confirmed Peptide Match

Protein ID: Tyrosine-protein kinase JAK3 (human)
Intact Protein Electrospray Mass Spectrometry
Intact Protein Electrospray - Mass Spectrum

Mathematically deconvoluted Data → Protein MW

12359 +/- 2
Intact Protein ESI-MS
Cytochrome C, 12361 Da

\[
\frac{m}{z} = \frac{12361 + 12}{12} = 1031
\]

Charge = +8

\[
\frac{m}{z} = \frac{12361 + 8}{8} = 1546
\]
Intact Protein Electrospray-MS
Infusion of Solubilized Relatively Pure Protein

- Relatively pure sample
- 20 – 50 μM protein concentration
- Detergent and salt free solution
- Typical solvent: 50:50, acetonitrile:water, 0.1% formic acid
- Difficult (but possible) to achieve high quality data
- Non-covalent interactions are retained with ESI (not MALDI), usually with neutral/basic buffer system and a lot of trial and error
Glycosylated Intact Protein ESI Mass Spec

- Intermediate 1 Protein

m/z, amu
0 2000 4000 6000 8000
Intensity, counts

m/z
1568 1618 1672 1730 1792 1815 1858 1882 1930 2007 2090 2200
Intermediate 1: Deconvoluted MS Data

Using Bayesian Reconstruct Tool; Theoretical mass = 50140.01 Da; error=4 ppm

- Mass reconstruction of +TOF MS: ...
- a=3.57017668630689880e-004, t0...
- Max. 1.1e5 cps.

Mass, amu

- 0.00
- 2.00e4
- 4.00e4
- 6.00e4
- 8.00e4
- 1.00e5
- 1.13e5

Intensity, cps

- 50140
- 50180
- 50797
- 50229
- 50497
- 49849

- Sialic acid
- Δ 291
- +K

- NAcNeu
- Gal
- GlcNAc

- Fucose
- Δ 656
- Δ 146

- Mass reconstruction of +TOF MS: ...
- a=3.57017668630689880e-004, t0...
- Max. 1.1e5 cps.
Special Applications in Mass Spectrometry

Andrew T. Rajczewski
CMSP Laboratory in Mass Spectrometry
January 10th, 2023
Quantification of small molecules, peptides and proteins

- Survey (MS – Relative between samples)
- Targeted (MS/MS – Absolute)
- Focus - Selected Reaction Monitoring (SRM)/Multiple Reaction Monitoring (MRM) scan mode as method of quantitation
- External or internal standards
Survey Scan (MS of intact molecule) and Identification (MS/MS)

- Accurate mass
- Retention time
- Product Ions
- Ultimately Relies on matching these to the known compound
Total Ion Chromatograph – Plasma in negative mode

Predominant analyte of peak at 9.67 minutes

255.2326 m/z (C16:0 fatty acid)
Extracted Ion Chromatograph – for m/z 255.2326

Integrated area under the curve for C16:0 fatty acid
Targeted: Selective/Multiple Reaction Monitoring

• Relies on intact m/z
• Relies on product ions produced under optimized conditions
• Relies on retention time
Q-Trap Mass Spectrometer
Single Reaction Monitoring (SRM) Theory

Quadrupole 1 (Q1)  
Collision Cell (Q2)  
Quadrupole 3 (Q3)  
Electron Multiplier (EM)
Selected Reaction Monitoring (SRM) Theory

Quadrupole 1 (Q1)  
Collision Cell (Q2)  
Quadrupole 3 (Q3)  
Electron Multiplier (EM)
Multiple Reaction Monitoring (MRM) Theory

Quadrupole 1 (Q1)  
Collision Cell (Q2)  
Quadrupole 3 (Q3)
MRM Case Studies

Example 1: MRM mode, Quantification w/o Internal Standard
Example 2: MRM mode, Quantitation WITH Internal Standard
Calibration - External Standard Method

**Standards** = analyte \( n_{A,i} \) \( i = 1 \ldots 5 \) in \( V_A \)

\[ c_1 = \frac{n_1}{V_A} \quad c_2 = \frac{n_2}{V_A} \ldots \]

**Unknown Sample** = analyte \( n_U \) in \( V_A \)

Analysis

\[ R_1, R_2, R_3 \ldots R_5 \]

Plot

\[ R_A = k \cdot n_A + S_{bl} \]

\[ y = mx + b \]

Measure \( R_U \)

\( n_U \)

amount, \( n_{A,i} \)

5/11/2008
FIGURE 1. The secondary structure of *E. coli* tRNALys(UUU) (a) and the chemical structure of t6A (b).

© 2012 by The American Society for Biochemistry and Molecular Biology, Inc. Published in the U.S.A.
Experimental Design

• Bulk purification of tRNA
• Enzymatically hydrolysed to nucleotides
• Purified by running zip-tip c18
• LC on C18 RP in presence of ammonium acetate
• MRM analysis
  • T6A  413 → 280.9 and 413 → 136
  • Adenosine 268 → 136
• Ratio of T6A to Adenosine was determined
Selected Reaction Monitoring in a Triple Quadrupole

Q1 (413)  Q2 collision cell  Q3 (280.9)

Fragment precursor

Set on mass of precursor ion  Transmit diagnostic product ion

6-Threonylcarbamoyl-adenosine (t6A)  413 m/z

6-Threonylcarbamoyl-adenine  280.9 m/z
Calibration for T6A_281: $y = 0.51544 \times + 186.90287 \ (r = 0.99996) \ (weighting: \ None)$
Wild-type
Y - uninduced
Y+IPTG
• Determined that the T6A had increased when a target gene was induced with an IPTG inducible promoter.
• Used ratio of an unlabeled internal standard housekeeping small molecule with similar characteristics as the target molecule.
Internal Standard (IS) Method

Standards = analyte \( n_{A,i} \) \( i = 1 \ldots 3 \) in \( V_A \)

\[ c_1 = \frac{n_1}{V_A} \]
\[ c_2 = \frac{n_2}{V_A} \]

\[ c_{1,f} = \frac{n_1}{V_A + v_{IS}} \]
\[ c_{IS,f} = \frac{n_{IS}}{V_A + v_{IS}} = \text{constant} \]

Internal Standard = IS \( n_{IS} \) in solvent \( V_{IS} \)

\[ v_{IS} \]

\[ c_{IS} = \frac{n_{IS}}{V_{IS}} \]

\[ c_U = \frac{n_U}{V_A} \]

Sample Work-up

Analysis

- Standard
- IS
- Analyte
- IS

\[ c_{U,f} = \frac{n_U}{V_A + v_{IS}} \]
\[ c_{IS,f} = \frac{n_{IS}}{V_A + v_{IS}} \]

= constant

5/11/2008
Acrylamide Experimental design

- Extract acrylamide from French fries
- Spike with stable isotope
- Process samples with SPE column
- LC-MS/MS on product ions
Acrylamide Multiple Reaction Monitoring

Q1 Precursor mass | Q2 collision cell | Q3 product ions

Capillary LC column

Precursor m/z values transmitted to Q2

CE optimized to produce diagnostic ions

Detection

Diagnostic product ions of acrylamide

75 m/z
72 m/z

58 m/z
55 m/z
44 m/z
Acrylamide
Standard Curve ratio of H/L

Calibration for acryl72/55: $y = 1.04935x + -0.00474$ (weighting: None)
## Acrylamide Spread Sheet

<table>
<thead>
<tr>
<th>Sample Name</th>
<th>Conc</th>
<th>Area</th>
<th>Area Ratio</th>
<th>Calc Conc</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>575</td>
<td>-</td>
<td>438710.8</td>
<td>0.696485206</td>
<td><strong>668.2531505</strong></td>
<td>-</td>
</tr>
<tr>
<td>575dup</td>
<td>-</td>
<td>405065.5</td>
<td>0.597388917</td>
<td><strong>573.8170576</strong></td>
<td>-</td>
</tr>
<tr>
<td>583</td>
<td>-</td>
<td>1258691</td>
<td>1.86163436</td>
<td><strong>1778.608893</strong></td>
<td>-</td>
</tr>
<tr>
<td>583dup</td>
<td>-</td>
<td>1181994</td>
<td>1.65007301</td>
<td><strong>1576.996628</strong></td>
<td>-</td>
</tr>
<tr>
<td>619</td>
<td>-</td>
<td>360802.5</td>
<td>0.482302645</td>
<td><strong>464.142942</strong></td>
<td>-</td>
</tr>
<tr>
<td>619dup</td>
<td>-</td>
<td>492556.5</td>
<td>0.705858204</td>
<td><strong>677.1853648</strong></td>
<td>-</td>
</tr>
<tr>
<td>standard 100pg</td>
<td>100</td>
<td>171811.6</td>
<td>0.099074211</td>
<td>98.93657198</td>
<td>98.93657198</td>
</tr>
<tr>
<td>standard 200pg</td>
<td>200</td>
<td>537783.4</td>
<td>0.207571201</td>
<td>202.3312791</td>
<td>101.1656395</td>
</tr>
<tr>
<td>standard 400pg</td>
<td>400</td>
<td>907843.4</td>
<td>0.397525022</td>
<td>383.352152</td>
<td>95.838038</td>
</tr>
<tr>
<td>standard 600pg</td>
<td>600</td>
<td>1191880</td>
<td>0.651920405</td>
<td>625.7840957</td>
<td>104.2973493</td>
</tr>
<tr>
<td>standard 800pg</td>
<td>800</td>
<td>1782703</td>
<td>0.82604448</td>
<td>791.719647</td>
<td>98.9649558</td>
</tr>
<tr>
<td>standard 1000pg</td>
<td>1000</td>
<td>3218397</td>
<td>1.033375724</td>
<td>989.3007344</td>
<td>98.93007344</td>
</tr>
<tr>
<td>standard 100pg</td>
<td>100</td>
<td>199147.7</td>
<td>0.097779106</td>
<td>97.70237165</td>
<td>97.70237165</td>
</tr>
<tr>
<td>standard 200pg</td>
<td>200</td>
<td>577656.6</td>
<td>0.206046147</td>
<td>200.8779445</td>
<td>100.4389722</td>
</tr>
<tr>
<td>standard 400pg</td>
<td>400</td>
<td>1029899</td>
<td>0.40132209</td>
<td>386.9706552</td>
<td>96.7426638</td>
</tr>
<tr>
<td>standard 600pg</td>
<td>600</td>
<td>1318852</td>
<td>0.658099536</td>
<td>631.6726406</td>
<td>105.2787734</td>
</tr>
<tr>
<td>standard 800pg</td>
<td>800</td>
<td>1980644</td>
<td>0.83004524</td>
<td>795.5322633</td>
<td>99.44153292</td>
</tr>
<tr>
<td>standard 1000pg</td>
<td>1000</td>
<td>3399542</td>
<td>1.037169871</td>
<td>992.9164546</td>
<td>99.29164546</td>
</tr>
</tbody>
</table>
• Multiple samples found to have acrylamide contamination
• Internal standards allowed for confident quantitation of acrylamide content
Mass Spectrometry Imaging (MSI)

- Often it is useful to know the spatial distribution of an analyte within a sample.
- Mass Spectrometry Imaging (MSI) selects specific analytes of interest and detects local concentration values.
- Readings can be added together to create a gradient image of the sample.
• Tissue sectioning exposes internal components of organs/organisms for MSI analysis
• Tissues are harvested, fixed, and embedded in paraffin
• Embedded tissues are thinly sliced, mounted on slides for analysis

MALDI Imaging

MALDI Imaging

DESI Imaging

DESI Imaging

MS/MS Provides Information for MRM Transitions

ProteinPilot interpretation of MS/MS data

Fragmentation Evidence

<table>
<thead>
<tr>
<th>Residue</th>
<th>b</th>
<th>b+2</th>
<th>y</th>
<th>y+2</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>115.0502</td>
<td>58.0287</td>
<td>1447.6852</td>
<td>724.3462</td>
</tr>
<tr>
<td>F</td>
<td>262.1186</td>
<td>131.5629</td>
<td>1333.6423</td>
<td>667.3248</td>
</tr>
<tr>
<td>Y</td>
<td>425.1819</td>
<td>213.0946</td>
<td>1186.5739</td>
<td>593.7906</td>
</tr>
<tr>
<td>L</td>
<td>538.2660</td>
<td>269.6366</td>
<td>1023.5106</td>
<td>512.2589</td>
</tr>
<tr>
<td>V</td>
<td>637.3344</td>
<td>319.1709</td>
<td>910.4265</td>
<td>455.7169</td>
</tr>
<tr>
<td>D</td>
<td>752.3614</td>
<td>376.6843</td>
<td>811.3581</td>
<td>406.1827</td>
</tr>
<tr>
<td>A</td>
<td>823.3985</td>
<td>412.2029</td>
<td>696.3311</td>
<td>348.6692</td>
</tr>
<tr>
<td>D</td>
<td>938.4254</td>
<td>469.7164</td>
<td>625.2940</td>
<td>313.1506</td>
</tr>
<tr>
<td>S</td>
<td>1025.4575</td>
<td>513.2324</td>
<td>510.2671</td>
<td>255.6372</td>
</tr>
<tr>
<td>F</td>
<td>1172.5259</td>
<td>586.7666</td>
<td>423.2350</td>
<td>212.1212</td>
</tr>
<tr>
<td>T</td>
<td>1273.5735</td>
<td>637.2904</td>
<td>276.1666</td>
<td>138.5870</td>
</tr>
<tr>
<td>R</td>
<td>1429.6747</td>
<td>715.3410</td>
<td>175.1190</td>
<td>88.0631</td>
</tr>
</tbody>
</table>

Precursor m/z 724.3

MS/MS data from captured peptide
BSA MRM/SRM spectrum

![Graph showing BSA MRM/SRM spectrum with peak IDs and masses.]

<table>
<thead>
<tr>
<th>Q1 Mass</th>
<th>Q3 Mass</th>
<th>Dwell Time</th>
<th>ID</th>
<th>CE</th>
<th>Q1 Mass</th>
<th>Q3 Mass</th>
<th>Dwell Time</th>
<th>ID</th>
<th>CE</th>
</tr>
</thead>
<tbody>
<tr>
<td>450.2</td>
<td>786.4</td>
<td>100</td>
<td>LCVLHEK</td>
<td>31.7</td>
<td>554.8</td>
<td>600.3</td>
<td>100</td>
<td>EACFAVEGPK</td>
<td>29.8</td>
</tr>
<tr>
<td>461.7</td>
<td>Y6/722.4</td>
<td>100</td>
<td>AEFVEVTK</td>
<td>31.4</td>
<td>570.7</td>
<td>818.4</td>
<td>100</td>
<td>LVNELTEFAK</td>
<td>27.5</td>
</tr>
<tr>
<td>464.2</td>
<td>Y5/651.3</td>
<td>100</td>
<td>YLYEIAR</td>
<td>41.5</td>
<td>571.9</td>
<td>886.6</td>
<td>100</td>
<td>KQTALVELLK</td>
<td>28.1</td>
</tr>
<tr>
<td>474.2</td>
<td>552.3</td>
<td>100</td>
<td>SLHTLFGDELCK</td>
<td>40.7</td>
<td>582.3</td>
<td>951.5</td>
<td>100</td>
<td>LVNELTEFAK</td>
<td>28.8</td>
</tr>
<tr>
<td>480.6</td>
<td>587.4</td>
<td>100</td>
<td>RHPEYAVSVLLR</td>
<td>33.4</td>
<td>628</td>
<td>948.5</td>
<td>100</td>
<td>RPCFSALTPDETYVPK</td>
<td>28.7</td>
</tr>
<tr>
<td>487.7</td>
<td>Y6/746.4</td>
<td>100</td>
<td>DLGEEHFK</td>
<td>49</td>
<td>653.4</td>
<td>1055.6</td>
<td>100</td>
<td>HLVDEPQNLIK</td>
<td>29</td>
</tr>
<tr>
<td>547.3</td>
<td>589.3</td>
<td>100</td>
<td>KVPQVSTPTLVEVSR</td>
<td>32.1</td>
<td>722.8</td>
<td>1168.5</td>
<td>100</td>
<td>YICDNQDTISSLK</td>
<td>33.6</td>
</tr>
</tbody>
</table>
Phosphorylated Protein Detection by Mass Spectrometry

Challenges

• Low abundance of modified protein (relative to unmodified or other proteins)
• MS signal suppression
• Ionization efficiency
• Data analysis inefficiency (software limitations)
• Phosphate group is labile under mass spectrometric conditions

Methods for success

• **Enrichment**
• Variation of proteolytic enzymes
• Both ESI and MALDI MS
• Multiple software packages and manual interpretation
ENRICHMENT for PHOSPHOPEPTIDES is Crucial to Success

**Dual enrichment method**

1. SCX separation: at pH 2.7, phosphopeptides elute early
2. Immobilized Metal Affinity Chromatography (IMAC)

**IDENTIFICATION of 1000’s of phosphopeptides/proteins**

Villén J & Gygi SP 2008, Nature Protocols 3(10), 1630
Peptides Zip Tipped C18

Phosphopeptide

Sudha Marimanikkuppam (University of Minnesota)

© 2015 Regents of the University of Minnesota. All rights reserved.
Workshop IIb
Manual Interpretation of MS/MS Spectra
By
Candace Guerrero
Why MS/MS spectra?

• The information contained in an MS spectrum (m/z, isotope spacing and therefore “z”) is not enough to tell us the amino acid sequence

• Luckily, peptides fragment in a predictable way and a spectrum of these fragments can help us figure out the sequence
OK... how?

• 100,000’s of molecules of a particular peptide will be isolated and fragmented
• Each of these molecules will create a $b/y$ ion pair in an MS/MS spectrum
• The **spacing** (mass difference) between peaks in an MS/MS spectrum is characteristic of the amino acid residue(s) differing between the two sequences
Easiest Examples

• **Glycyl-glycine (GG)**

The spacing of these peaks (57.021 m/z) tells us that the heavier ion has an extra glycine residue compared to the lighter ion.
Residue masses

Table 4.1. Residue masses of the amino acids. The residue masses of the 20 genetically encoded amino acids and selected modified amino acids.*

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>One-letter code</th>
<th>Residue mass (Da)</th>
<th>Immonium ion (m/z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycine</td>
<td>G</td>
<td>57.02</td>
<td>30</td>
</tr>
<tr>
<td>Alanine</td>
<td>A</td>
<td>71.04</td>
<td>44</td>
</tr>
<tr>
<td>Serine</td>
<td>S</td>
<td>87.03</td>
<td>60</td>
</tr>
<tr>
<td>Proline</td>
<td>P</td>
<td>97.05</td>
<td>70</td>
</tr>
<tr>
<td>Valine</td>
<td>V</td>
<td>99.07</td>
<td>72</td>
</tr>
<tr>
<td>Threonine</td>
<td>T</td>
<td>101.05</td>
<td>74</td>
</tr>
<tr>
<td>Cystine</td>
<td>C</td>
<td>103.01</td>
<td>76</td>
</tr>
<tr>
<td>Leucine</td>
<td>L</td>
<td>113.08</td>
<td>86</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>I</td>
<td>113.08</td>
<td>86</td>
</tr>
<tr>
<td>Asparagine</td>
<td>N</td>
<td>114.04</td>
<td>87</td>
</tr>
<tr>
<td>Aspartate</td>
<td>D</td>
<td>115.03</td>
<td>88</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Q</td>
<td>128.06</td>
<td>101</td>
</tr>
<tr>
<td>Lysine</td>
<td>K</td>
<td>128.09</td>
<td>101</td>
</tr>
<tr>
<td>Glutamate</td>
<td>E</td>
<td>129.04</td>
<td>102</td>
</tr>
<tr>
<td>Methionine</td>
<td>M</td>
<td>131.04</td>
<td>104</td>
</tr>
<tr>
<td>Histidine</td>
<td>H</td>
<td>137.06</td>
<td>110</td>
</tr>
<tr>
<td>Oxidized Methionine</td>
<td>Mo</td>
<td>147.04</td>
<td>120</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>F</td>
<td>147.07</td>
<td>120</td>
</tr>
<tr>
<td>Arginine</td>
<td>R</td>
<td>156.10</td>
<td>129</td>
</tr>
<tr>
<td>Carboxymethylcysteine</td>
<td>C*</td>
<td>160.03</td>
<td>133</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Y</td>
<td>163.06</td>
<td>136</td>
</tr>
<tr>
<td>Acryloycysteine</td>
<td>C^</td>
<td>174.04</td>
<td>147</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>W</td>
<td>186.08</td>
<td>159</td>
</tr>
</tbody>
</table>

* This table also includes the one-letter abbreviations commonly used when writing peptide sequences and the m/z of the immonium ions with the form NH$_2$=$\text{CHR}^+$.

Lys $y_1$ = 147 m/z

Arg $y_1$ = 175 m/z

Look at the low end of the spectrum for these $y_1$ ions. If you have performed a trypsin digest then one of these ions should be present in the MS/MS fragmentation spectrum and this will reveal the C-terminal residue. Hint: The N-terminal b1 ion is never observed but often the b1 ion can be observed after acetylation of the peptide.
**b$_2$ ion is usually present**
Table 4.3. Amino acids combinations that are equal to a single amino acid residue mass.*

<table>
<thead>
<tr>
<th>Amino acid combination</th>
<th>Residue mass (Da)</th>
<th>Equivalent amino acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>GG</td>
<td>114</td>
<td>N</td>
</tr>
<tr>
<td>GA</td>
<td>128</td>
<td>Q, K</td>
</tr>
<tr>
<td>GV</td>
<td>156</td>
<td>R</td>
</tr>
<tr>
<td>GE</td>
<td>186</td>
<td>W</td>
</tr>
<tr>
<td>AD</td>
<td>186</td>
<td>W</td>
</tr>
<tr>
<td>SV</td>
<td>186</td>
<td>W</td>
</tr>
<tr>
<td>SS</td>
<td>174</td>
<td>C^a</td>
</tr>
</tbody>
</table>

* The single-letter amino acid codes shown in Table 4.1 are used in this table.

Table 4.4. Neutral losses observed from ions with different amino acid compositions.*

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Neutral loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>–</td>
</tr>
<tr>
<td>G</td>
<td>–</td>
</tr>
<tr>
<td>S</td>
<td>18</td>
</tr>
<tr>
<td>P</td>
<td>–</td>
</tr>
<tr>
<td>V</td>
<td>–</td>
</tr>
<tr>
<td>T</td>
<td>18</td>
</tr>
<tr>
<td>C</td>
<td>34</td>
</tr>
<tr>
<td>L/I</td>
<td>–</td>
</tr>
<tr>
<td>N</td>
<td>17</td>
</tr>
<tr>
<td>D</td>
<td>18</td>
</tr>
<tr>
<td>Q</td>
<td>17</td>
</tr>
<tr>
<td>K</td>
<td>17</td>
</tr>
<tr>
<td>E</td>
<td>18</td>
</tr>
<tr>
<td>M</td>
<td>48</td>
</tr>
<tr>
<td>H</td>
<td>–</td>
</tr>
<tr>
<td>Mo</td>
<td>64</td>
</tr>
<tr>
<td>F</td>
<td>–</td>
</tr>
<tr>
<td>R</td>
<td>17</td>
</tr>
<tr>
<td>C^a</td>
<td>92</td>
</tr>
<tr>
<td>Y</td>
<td>–</td>
</tr>
<tr>
<td>C^a</td>
<td>106</td>
</tr>
<tr>
<td>W</td>
<td>–</td>
</tr>
</tbody>
</table>

* The consecutive loss of small neutral molecules is an energetically favored process in collisionally induced dissociation. The nature of the neutral that is lost is dependent on the amino acid composition of the product ion. In this table, the one-letter amino acid codes given in Table 4.1 are used. The – designates that no neutral losses occur for that amino acid.
The Nine-Step Strategy for Interpretation (from Reference 1)

1. Inspect the low-mass region for immonium ions. The first step in the interpretation is to inspect the low-mass region of the spectrum, noting the presence of any immonium ions and the amino acid composition that they indicate.

2. Inspect the low-mass region for the b\(_2\)-ion. In the second step of the interpretation, the low-mass region of the spectrum is inspected to identify the b\(_2\)-ion, generally recognizable by the b\(_2\)-ion / a\(_2\)-ion pair separated by 28 Da. By using Table 4.2, the possible two-amino acid combinations indicated by the b\(_2\)-ion are noted. The m/z of the b\(_2\)-ion is then used to calculate the m/z of the corresponding y\(_{n-2}\)-ion, and the high-mass region of the product ion spectrum is inspected to identify this ion.

3. Inspect the low-mass region for the y\(_1\)-ion. The third step of the interpretation is to assign the C-terminal amino acid. The low-mass region of the spectrum is inspected to identify the y\(_1\)-ion at either m/z 147, for C-terminal lysine peptides, or m/z 175, for C-terminal arginine peptides. The m/z of the y\(_1\)-ion is then used to calculate the m/z of the b\(_{n-1}\)-ion, and the high-mass region of the product ion spectrum is inspected to identify that ion, if present.

4. Inspect the high-mass region to identify the y\(_{n-1}\)-ion. The fourth step of the interpretation is to attempt to assign the N-terminal amino acids from combinations indicated by the b\(_2\)-ion. The high-mass region of the spectrum is scrutinized to identify the y\(_{n-1}\)-ion, if present. The list of possible amino acid combinations derived from the b\(_2\)-ion limits the possible residue masses to consider. If an ion is identified the m/z of that ion is used to calculate the residue masses of the first two amino acids and to assign those peptides.

5. Extend the y-ion series toward lower m/z. Working with the residue masses listed in Table 4.1, begin to extend the y-ion series backwards (toward lower m/z) from the y\(_{n-2}\)-ion. As a y-ion is identified calculate the m/z of the corresponding b-ion and identify that ion in the spectrum. Work towards extending the y-ion series from the y\(_{n-2}\) ion to the y\(_1\)-ion.

6. Extend the b-ion series toward higher m/z. If progress extending the y-ion series falters, use the residue masses listed in Table 4.1 to extend the b-ion series from the last identified b-ion. As any b-ions are identified, use the m/z of that ion to calculate the m/z of the corresponding y-ion and identify that ion in the spectrum.

7. Calculate the mass of the peptide. When the interpretation of the spectrum is complete, calculate the mass of the proposed peptide sequence and check its agreement with the measured mass.

8. Reconcile the amino acid content with spectrum data. Check that the amino acid content agrees with the immonium ions observed. Also consider the charge state of the peptide in terms of the presence of histidine, and internal lysine or arginine residues.

9. Attempt to identify all ions in the spectrum. Work to identify the other ions in the spectrum based on the proposed peptide sequence and pay particular attention to the ions from the loss of H\(_2\)O, NH\(_3\), and HSOCH\(_3\); any doubly charged ions; and any ions due to internal cleavages.
Peptide MS/MS Spectrum

Precursor \([M + 2H]^+\) 698.86 \(m/z\)

Charge state information is calculated from full scan.

Full scan
698.86 \(m/z\)
precursor isotope series: \(z = 2\)

\[ \frac{1}{\Delta m/z} = \text{charge state} \]

\(\Delta m/z = .5\)
cguerrer_2016115_newcolumn_HeLa_QC5 #13190 RT: 27.40 AV: 1 NL: 6.65E3
T: ITMS + c NSI r d Full ms2 821.9176@cid35.00 [221.0000-1654.0000]
What possible b2 ion can this be?
What possible \( b_2 \) ion can this be? Q, L/I

\[ \Delta m = ? \]
$b_3 = V$

$b_2 = Q, \frac{L}{I}$

$\Delta m = 99.07$
\[ b_3 = V \]

\[ b_2 = Q, \frac{L}{I} \]

\[ \Delta m = 99.07 \]

\[ \Delta m = ? \]
\[ b_2 = Q, L/I \]
\[ \Delta m = 99.07 \]
\[ b_3 = V \]
\[ \Delta X = ? \]
b₁ = Q, L/I

Δm=99.07

b₃ = V

Δm=?
b₂ = Q, L/I

Δm = 99.07

b₃ = V

341.16

b₄ = L/I

Δm = 113.08
$b_2 = Q, L/I$

$\Delta m = 99.07$

$Q, L/I$

$V$

$b_3 = V$

$341.16$

$\Delta m = 113.08$

$L/I$

$b_4 = L/I$

$433.14$

$454.24$

$504.16$

$582.28$

$651.77$

$Q$

$b_5 = Q$

$452.17$

$469.24$

$565.26$

$597.28$

$682.38$

$\Delta m = 124.04$

$619.20$

$\Delta m = ?$
<table>
<thead>
<tr>
<th>Relative Abundance</th>
<th>m/z</th>
</tr>
</thead>
<tbody>
<tr>
<td>701.42</td>
<td></td>
</tr>
<tr>
<td>341.16</td>
<td></td>
</tr>
<tr>
<td>619.20</td>
<td></td>
</tr>
<tr>
<td>710.36</td>
<td></td>
</tr>
<tr>
<td>504.16</td>
<td></td>
</tr>
<tr>
<td>582.28</td>
<td></td>
</tr>
<tr>
<td>454.24</td>
<td></td>
</tr>
<tr>
<td>433.14</td>
<td></td>
</tr>
<tr>
<td>242.04</td>
<td></td>
</tr>
<tr>
<td>304.11</td>
<td></td>
</tr>
<tr>
<td>452.17</td>
<td></td>
</tr>
<tr>
<td>469.24</td>
<td></td>
</tr>
<tr>
<td>504.16</td>
<td></td>
</tr>
<tr>
<td>582.28</td>
<td></td>
</tr>
<tr>
<td>547.25</td>
<td></td>
</tr>
<tr>
<td>570.17</td>
<td></td>
</tr>
<tr>
<td>597.28</td>
<td></td>
</tr>
<tr>
<td>619.20</td>
<td></td>
</tr>
<tr>
<td>651.77</td>
<td></td>
</tr>
<tr>
<td>668.25</td>
<td></td>
</tr>
<tr>
<td>682.38</td>
<td></td>
</tr>
<tr>
<td>682.38</td>
<td></td>
</tr>
<tr>
<td>693.34</td>
<td></td>
</tr>
<tr>
<td>701.42</td>
<td></td>
</tr>
<tr>
<td>341.16</td>
<td></td>
</tr>
<tr>
<td>619.20</td>
<td></td>
</tr>
<tr>
<td>710.36</td>
<td></td>
</tr>
<tr>
<td>504.16</td>
<td></td>
</tr>
<tr>
<td>582.28</td>
<td></td>
</tr>
<tr>
<td>454.24</td>
<td></td>
</tr>
<tr>
<td>433.14</td>
<td></td>
</tr>
<tr>
<td>242.04</td>
<td></td>
</tr>
<tr>
<td>304.11</td>
<td></td>
</tr>
<tr>
<td>452.17</td>
<td></td>
</tr>
<tr>
<td>469.24</td>
<td></td>
</tr>
<tr>
<td>504.16</td>
<td></td>
</tr>
<tr>
<td>582.28</td>
<td></td>
</tr>
<tr>
<td>547.25</td>
<td></td>
</tr>
<tr>
<td>570.17</td>
<td></td>
</tr>
<tr>
<td>597.28</td>
<td></td>
</tr>
<tr>
<td>619.20</td>
<td></td>
</tr>
<tr>
<td>651.77</td>
<td></td>
</tr>
<tr>
<td>668.25</td>
<td></td>
</tr>
<tr>
<td>682.38</td>
<td></td>
</tr>
<tr>
<td>682.38</td>
<td></td>
</tr>
<tr>
<td>693.34</td>
<td></td>
</tr>
</tbody>
</table>

**Formula Diagram**

- $b_3 = V$  
  \[m/z = 341.16\]

- $b_2 = Q, L/I$  
  \[\Delta m = 99.07\]

- $b_4 = L/I$  
  \[m/z = 433.14, 454.24, 452.17, \ldots\]

- $b_5 = Q$  
  \[m/z = 504.16, 582.28, 547.25, \ldots\]

- $\Delta m = 113.08$

- $\Delta m = 124.04$

- $\Delta m = \Box$
$b_2 = Q, L/I$

$\Delta m = 99.07$

$Q, L/I$

$\Delta m = 113.08$

$Q$

$\Delta m = 124.04$

$Q$

$\Delta m = 128.08$

$b_5 = Q$

$b_3 = V$

$341.16$

$b_4 = L/I$

$433.14$

$454.24$

$b_6 = Q$

$504.16$

$582.28$

$b_6 = Q$

$619.20$

$597.28$

$693.34$

$701.42$

$710.36$
Q, L/I        V, L/I    Q    Q    Q    A    D

b_2          b_3          b_4          b_5          b_6          b_7          b_8          b_9

Δm=128.07    Δm=71.02    Δm=115.04

b_7 = Q       b_6 = A       b_5 = D
cguerrer_2016115_newcolumn_HeLa_QC5 #13190
RT: 27.40  AV:  1  NL: 9.35E3
T: ITMS + c NSI r d Full ms2 821.9176@cid35.00 [221.0000-1654.0000]

Q, L/I  V  L/I  Q  Q  Q  A  D  D

b₂ b₄ b₆ b₈ b₁₀

Δm=128.07  b₇ = Q
Δm=71.02   b₈ = A
Δm=115.04  b₉ = D
Δm=115.04  b₁₀ = D

m/z

100 90 80 70 60 50 40 30 20 10 0

Δm=128.07  b₇ = Q
Δm=71.02   b₈ = A
Δm=115.04  b₉ = D
Δm=115.04  b₁₀ = D

734.28  813.15  838.42  909.44  1006.44  1080.70  1101.40  1189.45  1210.57  1265.50  1302.50  1339.51  1359.44  1384.54  1401.61  1451.67  1504.00  1513.74  1540.78  1582.50  1600.54
Δm=128.07
b7 = Q

Δm=71.02
b8 = A

Δm=115.04
b9 = D

Δm=115.04
b10 = D

Δm=115.04
b11 = A

Q, L/I  V, L/I  Q  Q  Q  A  D  D  A
Δm=128.07

Δm=71.02

Δm=115.04

Δm=115.04

Δm=115.04

Δm=128.94

Δm=128.94
Δm=128.07

Δm=71.02

Δm=115.04

Δm=115.04

Δm=115.04

Δm=115.04

Δm=115.04

Δm=128.94

Δm=129.11

Δm=128.07

Δm=71.02

Δm=115.04

Δm=115.04

Δm=115.04

Δm=115.04

Δm=128.94

Δm=129.11

Δm=128.07

Δm=71.02

Δm=115.04

Δm=115.04

Δm=115.04

Δm=115.04

Δm=128.94

Δm=129.11

Δm=128.07

Δm=71.02

Δm=115.04

Δm=115.04

Δm=115.04

Δm=115.04

Δm=128.94

Δm=129.11

Δm=128.07

Δm=71.02

Δm=115.04

Δm=115.04

Δm=115.04

Δm=115.04

Δm=128.94

Δm=129.11

Δm=128.07

Δm=71.02

Δm=115.04

Δm=115.04

Δm=115.04

Δm=115.04

Δm=128.94

Δm=129.11

Δm=128.07

Δm=71.02

Δm=115.04

Δm=115.04

Δm=115.04

Δm=115.04

Δm=128.94

Δm=129.11

Δm=128.07

Δm=71.02

Δm=115.04

Δm=115.04

Δm=115.04

Δm=115.04

Δm=128.94

Δm=129.11

Δm=128.07

Δm=71.02

Δm=115.04

Δm=115.04

Δm=115.04

Δm=115.04

Δm=128.94

Δm=129.11
How do we know it’s the end of the peptide sequence?
How do we know it’s the end of the peptide sequence?
We use Trypsin to digest so it has to end with L or R
Q, L/I V L/I Q Q Q A D D A E E E R

How do we know it’s the end of the peptide sequence?

We use Trypsin to digest so it has to end with L or R

The whole peptide singly charged is 1642.62 m/z. 1642.62-1468.62=174

Lys $y_1 = 147 \text{ m/z}$

Arg $y_1 = 175 \text{ m/z}$
How do we work our way back to the beginning through the y-series ions?
Figuring out the y-series

• y-series typically have higher intensity ions
• The last ion in the y-series has to be either
  • Q or I/L
• Our single charge mass to charge of the peptide is 1642.64 m/z
• Let’s figure out what the m/z value of the ion we need to look for is.

<table>
<thead>
<tr>
<th>Isoleucine or Leucine</th>
<th>Δm of 113.08</th>
</tr>
</thead>
<tbody>
<tr>
<td>1642.64 - 113.08 = 1529.56 m/z</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Glutamine</th>
<th>Δm of 128.06</th>
</tr>
</thead>
<tbody>
<tr>
<td>1642.64 – 128.06 = 1514.58 m/z</td>
<td></td>
</tr>
</tbody>
</table>
Neither 1529.56 m/z or 1514.58 m/z are to be found.
Neither 1529.56 m/z or 1514.58 m/z are to be found.

What now....
Q, L/I V L/I Q Q Q A D D A E E R

Δm=128.07

Δm=71.02

Δm=115.04

Δm=115.04

Δm=115.04

Δm=115.04

Δm=128.94

Δm=129.11

Neither 1529.56 m/z or 1514.58 m/z are to be found.

What now….lets see if those ions are possibly doubly charged
Calculation to determine what m/z value we should look for.

\[
\frac{\text{mass} + n\text{H}}{\text{charge}}
\]

We are looking for a double charged ion for \(y_{13}\). Typically we would insert 2 for \(n\) into the equation. However, our ion is already singly charged. Therefore we only do the following:

Glutamine

\[
\frac{1514.58+1}{2} = 757.79 \text{ m/z}
\]

Isoleucine or Leucine

\[
\frac{1529.56+1}{2} = 765.28 \text{ m/z}
\]
$y_{13} = \text{Isoleucine or Leucine}$
cguerrer_2016115_newcolumn_HeLa_QC5 #13190  RT: 27.40  AV: 1  NL: 9.35E3
T: ITMS + c NSI r d Full ms2 821.9176@cid35.00 [221.0000-1654.0000]
Proteomics
vs
Metabolomics

Introduction to Discovery Small Molecule Analytics

Objectives

• What is Metabolomics?

• Differences and Similarities to Proteomics

• Considerations for Metabolomics Experiments
Metabolomics

Metabolites are precursors, intermediaries, or end-products of metabolism. (MW < 1500 Da)

Image Source: https://www.vanderbilt.edu/cit/introduction-metabolomics-research//
The Metabolome

• 1000’s of unique primary metabolites.

• Unknown number of secondary metabolites and exogenous chemicals.

• Most dynamic system in an organism.

Image Source: http://education.med.nyu.edu/mbm/carbohydrates/tca.shtml
Chemical Diversity

• Huge variety of compound classes, differing physical and biochemical properties.

• Greatest challenge is adequately sampling, characterizing, and identifying these molecules.
Sub classifications

Metabolomics

Lipidomics

Adductomics

Exposomics

Fluxomics*

Volatomics

Image Source: vectorstock/24005281
Metabolomics Technologies

- Liquid Chromatography-Mass Spectrometry (LC-MS)
- Gas Chromatography-Flame Ionization Detector (GC-FID)
- Infrared Ion Spectrometry (IR)
- Direct Infusion-Mass Spectrometry (DI-MS)
- Nuclear Magnetic Resonance Spectroscopy (NMR)
- Raman Spectroscopy
- Gas Chromatography-Mass Spectrometry (GC-MS)

NMR Review: Larive et al. 2015. NMR Spectroscopy for Metabolomics Metabolic Profiling. Analytical Chemistry. 87, 1, 133-146.
**LC-MS (LC-MS/MS)**  
*Thermo Scientific QExactive*

<table>
<thead>
<tr>
<th></th>
<th>LC-MS</th>
<th>GC-MS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Matrix</strong></td>
<td>Liquid</td>
<td>Solid/Liquid/Gas</td>
</tr>
<tr>
<td><strong>Ionization</strong></td>
<td>ESI (soft) - protonated</td>
<td>EI (hard) - radicals</td>
</tr>
<tr>
<td><strong>Selectivity</strong></td>
<td>ESI compatible and stable in aqueous</td>
<td>Thermally stable and volatile</td>
</tr>
<tr>
<td><strong>Avg # of Metabolites Detected</strong></td>
<td>1000’s</td>
<td>100’s</td>
</tr>
<tr>
<td><strong>Spectral Library Size</strong></td>
<td>Small/Moderate</td>
<td>Large</td>
</tr>
<tr>
<td><strong>Price</strong></td>
<td>$$$</td>
<td>$$</td>
</tr>
</tbody>
</table>
Proteomics vs Metabolomics

Preparation

MS Analysis

Data Processing

Cell Lysate

Organellar Fraction

Purified Protein Complex

Sample for PTM analysis

Single Protein of interest on a Gel

Proteolytic Digest

LC-ESI

MS

MS/MS

Data Analysis

Cellular Composition

Organellar Proteome

Interaction Protein

PTM Analysis

Protein ID
# Sample Preparation

## Proteomics

- **Matrix:**
  - **Cells**, Tissues, Biofluids...

- **Extract proteins of interest:**
  - Cell/Tissue Lysis
  - In-solution vs in-gel digestion
  - Peptide enrichment
  - Clean-up

## Metabolomics

- **Matrix:**
  - Cells, Tissues, **Biofluids**...

- **Extract metabolites of interest:**
  - Cell/Tissue Lysis
  - Protein Precipitation
  - Derivatization or fractionation
  - Clean-up

## Shared

- Consistency, consistency, consistency...
- Quenching – Cease activity as quickly as possible and be mindful how long you store samples at room temp.
- Document preparation careful and keep detailed notes anytime you change solvents, vial types, etc...
Protein Precipitation

- Utilize cold organic solvent to precipitate proteins and remove other large macromolecules.
  - Methanol/Acetone (90:10) stored at -80C.

- Vortex 1 min and chill sample at -20C for 15 min.
  - *Most variability in sample prep comes from this step.*

- Centrifuge at 13,000 rpm for 15 mins and transfer supernatant to new vial (repeat for dirty samples).

- Dry under nitrogen or with vacuum centrifuge.
Derivatization

- Process of making more suitable for analysis.
  - LC-MS: Improve ionization efficiency
  - GC-MS: Stabilize thermal labile compounds and improve volatility for heated injection.

```
R^\text{\text{-}}\text{NH}_2
\text{O}
R^\text{\text{-}}\text{COOH} + \text{CF}_3\text{N}^\text{-}\text{Si}^\text{-}
\text{R^\text{\text{-}}\text{O}}\text{NH}^\text{-}\text{Si}^\text{-}
\text{R^\text{\text{-}}\text{O}}\text{Si}^\text{-}
\text{R^\text{\text{-}}\text{O}}\text{O}^\text{-}\text{Si}^\text{-}
```

MSTFA: N-Methyl-N-(trimethylsilyl) trifluoroacetamide

Image Source: https://doi.org/10.3390/metabo1010003
Liquid Chromatography

**Proteomics LC Parameters**
- Column Dimensions: 100um x 50 cm
- Flow Rate: 325 nL/min
- Avg Run Length: 1-2 hours

**Metabolomics LC Parameters**
- Column Dimensions: 2.1mm x 50mm
- Flow Rate: 400 uL/min
- Avg Run Length: 20 min
Mass Spectrometry

• Both utilize high-resolution mass spectrometry

• Metabolomics instruments tuned for lower mass range than proteomics.

<table>
<thead>
<tr>
<th></th>
<th>Proteomics</th>
<th>Metabolomics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scan Range</strong></td>
<td>380-1580 m/z</td>
<td>70-1050 m/z</td>
</tr>
<tr>
<td><strong>Expected Charge</strong></td>
<td>2-5</td>
<td>1-3</td>
</tr>
<tr>
<td><strong>Ionization Type</strong></td>
<td>Positive</td>
<td>Positive or Negative</td>
</tr>
</tbody>
</table>
- RT on y-axis
- m/z on x-axis
- Intensity is color depth

Digger Deeper
More than meets the eye.
Retention Time

0.00  |  3.00  |  6.00  |  9.00  |  12.00  |  15.00  |  18.00
---|---|---|---|---|---|---
Amino Acids | Sugars | Carnitines-Conjugates | Bile Acids | Hormones | Cholesterol | Phospholipids | DGs and TGs

Total ion intensity

Retention Time
Negative Ion Polarity

• Proteomics utilizes positive ion polarity nearly always.
  • Peptides exhibit greater ionization efficiency in positive mode

• Numerous small molecules that preferentially gain negative charge during ESI ionization.
  • Some species make viewed with both polarity.

L-Isoleucine
Neutral mass: 131.094628665

\[
\begin{align*}
\text{[M+H]}^+ & : 132.1019 \text{ m/z} \\
\text{[M-H]}^- & : 130.0873 \text{ m/z}
\end{align*}
\]
Limitations of Negative Ion Mode

• On average, detect fewer compounds.
  • Considerable overlap with positive ion mode.

• Fewer MS/MS spectra in spectral libraries.

• More difficult to optimize.
Data Processing

• Visualization and empirical evaluation of proteomics and metabolomics data is similar.

• Data processing pipelines are completely independent and utilize different software.

• In general, processing metabolomics data is easier but obtaining identification are harder.
Basic Data Processing Workflow

- Data Import
- Mass Detection
- Chromatogram Building
- Deconvolution
- Alignment
- Export
Basic Data Processing Workflow

Data Import → Mass Detection → Chromatogram Building → Deconvolution → Alignment → Export

![Mass Detection Diagram](image)
Basic Data Processing Workflow

Data Import → Mass Detection → Chromatogram Building → Deconvolution → Alignment → Export

Retention time

Total ion intensity

0.00 5.00 10.00 15.00 20.00 25.00 30.00 35.00 40.00 45.00

0.00E0 2.00E6 4.00E6 6.00E6 8.00E6 1.00E7 1.20E7 1.40E7 1.60E7 1.80E7 2.00E7 2.20E7

Center for Mass Spectrometry and Proteomics | Phone | (612)625-2280 | (612)625-2279
Basic Data Processing Workflow

- Data Import
- Mass Detection
- Chromatogram Building
- Deconvolution
- Alignment
- Export
Basic Data Processing Workflow

1. Data Import
2. Mass Detection
3. Chromatogram Building
4. Deconvolution
5. Alignment
6. Export

Graph showing correlation between Sample A RT and Sample B RT.
Discovery with Quantitation

Proteomics

• Labeled:
  • TMT (6, 10, 11, 16 plex)
  • iTRAQ
  • SILAC

• Label-free Quant

Metabolomics

• Labeled:
  • TMT (6, 10, 11, 16 plex)
  • iTRAQ
  • SILAC

• Label-free Quant
Identification

• Peptides and metabolites are identified using very different methods.

• As polymers, peptides exhibit characteristic fragmentation between amino acid residues.

• Metabolite fragmentation is more unique, cannot reliably predict the fragmentation of compounds.
Caffeine
C8H10N4O2
Mono Mass: 194.0803

Creatine
C4H9N3O2
Mono Mass: 131.0694
L-Glutamine
C_5H_10N_2O_3
Mono Mass: 146.0691
[M+H]^+

Match Factor: 949

Match Factor = Weighted Dot-product
Ranges: (0...999)
**L-Tryptophan**
C11H12N2O2
Mono Mass: 204.0898
[M+H]+

**3-(4-Amino-phenyl)-3-methylpyrrolidine-2,5-dione**
C11H12N2O2
Mono Mass: 204.0898
[M+H]+

**Match Factor: 100**

**Match Factors:**
- 700 (fair)
- 800 (good)
- 900 (great)
LysoPC(14:0/0:0)
C22H46NO7P
Mono Mass: 467.3011

LysoPC(16:0/0:0)
C24H50NO7P
Mono Mass: 495.3324
Improving Identification Confidence

• Include precursor mass filters (10 ppm / 0.005 Da)

• Utilize retention time libraries.

• Evaluate authentic standards and/or labeled standards.

• **METABOLOMICS ID’S HAVE NO FDR SCORE!!!**
Identification Difficulties

- You will not identify everything in your samples...
- ID rate (5% - 25%) in common biofluids.
- Library bias – is compound of interest in library?
- Ambiguous IDs – too many similar compounds
Considerations for Future Metabolomics Studies
Beware Blind Experiments

• Global metabolomics is a myth...

• Hypothesis-generating ≠ Hypothesis-free

• Secondary hypotheses are valuable tools.
  • Any pathways or metabolites of interest? Let Us Know!
Untargeted vs Targeted

• Trade-offs for all experimental designs.

• Untargeted metabolomics
  • ↑ coverage of metabolites and pathways.
  • ↓ precision and confidence to infer differences.

• Targeted metabolomics
  • ↑ precision and estimate closer to true concentration
  • ↓ requires defined hypothesis
Requires More Attention to Experimental Factors

• Small molecules contaminants are everywhere!
  • Oils on skin, micro-plastics and polymers, and more!

• Some metabolites are very reactive, will create unexpected conjugates in solution.

• Metabolomics should not be an after thought or last resort! Plan your experiments appropriately!!
Questions?
Protein Identification by Mass Spectrometry

LeeAnn Higgins
<table>
<thead>
<tr>
<th>Outline</th>
<th>Terminology</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Protein sequence database</td>
<td>• Mass tolerance</td>
</tr>
<tr>
<td>• Database search parameters</td>
<td>• MS/MS search</td>
</tr>
<tr>
<td>• Database search types</td>
<td>• FASTA</td>
</tr>
<tr>
<td>• Peptide Mass Fingerprint (PMF) search</td>
<td>• Precursor mass</td>
</tr>
<tr>
<td>• Precursor mass-based search</td>
<td>• Fragment ion</td>
</tr>
<tr>
<td>• Sequence tag/de novo search</td>
<td>• Theoretical mass</td>
</tr>
<tr>
<td>• Spectral Libraries</td>
<td>• Experimental mass</td>
</tr>
<tr>
<td></td>
<td>• Sequence tag</td>
</tr>
<tr>
<td></td>
<td>• <em>de novo</em> sequencing</td>
</tr>
</tbody>
</table>
Three Strategies for Protein Inference from Peptide MS or MS/MS Data

DATA INPUT:
- Search TYPE:
  - Peptide Mass List (MS1)
  - Peptide Mass Fingerprint (PMF) (historical)
  - Peptide MS/MS (MS2)

1. Peptide Mass List (MS1)
2. Precursor Mass-based
3. Sequence-based (de novo or mass tag)
Selecting a Protein Sequence Database

• Public repositories (choose one)
  • NCBI Reference Sequence
  • UniProtKB Universal Proteome
    https://www.uniprot.org/proteomes/

• Custom protein sequences
  • not in public repository, format a text file with protein sequences (see next slide)

• Species specific
  • Example 1: HEK293 human kidney cell line
  • Example 2: Mouse protein expressed in E. coli

• Ideal size range ~ 2000 to < 1 million entries
FASTA-formatted Protein Sequence from NCBI example

<table>
<thead>
<tr>
<th>Accession</th>
<th>Protein name</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;AAD44166.1</td>
<td>cytochrome b, partial (mitochondrion) [Elephas maximus maximus]</td>
<td></td>
</tr>
<tr>
<td>LCLYTHIGRNIYYGSYLWSETWNTGIMLLLLITMATAFMGYVLPWQMSFWGATVITNLFSAIPTYIGTNLV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EWIWGGFSVDKATLNRRFAHFILPFTMVAGVHLTFHETGSNNPLGLTSDSDKIPFHPYTYTIKDFLG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LLILILLLLALASPDMLGDPDNMPADPLNTPLHIKPEWYFLFAYAILRSVPNKLGGLVALLFLSIVIL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLMPFLHTSKHRSMMLRPLSQALFWTLTMDLLLTLTWIGSQPVEPYTIIGQMASILYFSIIAFFPIAGX</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IENY</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Peptide Mass Fingerprint Search

Example analysis:
Peptide Masses from MALDI-TOF MS Experiment
### Peptide Mass Fingerprint (PMF)

- **Historical**: after 1D or 2D in-gel Digestion of single bands or spots
- **Current applications**: limited; decreasing utility

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input (Data)</strong></td>
<td>Peptide Mass List</td>
</tr>
<tr>
<td></td>
<td>• Mass type (Mr or MH+)</td>
</tr>
<tr>
<td></td>
<td>• Average or Monoisotopic</td>
</tr>
<tr>
<td><strong>Target</strong></td>
<td>Protein FASTA Sequence Database</td>
</tr>
<tr>
<td><strong>Search Parameters</strong></td>
<td>Proteolytic enzyme</td>
</tr>
<tr>
<td></td>
<td># Missed enzyme cleave sites</td>
</tr>
<tr>
<td></td>
<td>Amino acid modifications</td>
</tr>
<tr>
<td></td>
<td>Mass tolerance</td>
</tr>
</tbody>
</table>
Peptide Mass Fingerprint: INPUT

**UNKNOWN PROTEIN** *(amino acid sequence)*:

- GLSDGEWQQVLNVWGKVEADIAGHGQEVLR
- LFTGHPETLEK
- TEAEMK
- ASEDLK
- HGTVVLTALGGILK
- GHHEAELKPLAQSHATK
- YLEFISDAIIHVLSHPGNFGADAQGAMTK
- ALELFR
- NDIAAK
- ELGFAQG

**A) In gel Trypsin digestion**

**B) Acquire MALDI-TOF Mass Spectrum of peptide mixture**

**C) Experimental Peptide Mass (or m/z) List**

<table>
<thead>
<tr>
<th>m/z</th>
<th>Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>631.71</td>
<td>NDIAAK</td>
</tr>
<tr>
<td>650.71</td>
<td>ELGFAQG</td>
</tr>
<tr>
<td>662.72</td>
<td>ASEDLK</td>
</tr>
<tr>
<td>708.81</td>
<td>TEAEMK</td>
</tr>
<tr>
<td>748.90</td>
<td>ALELFR</td>
</tr>
<tr>
<td>1272.44</td>
<td>LFTGHPETLEK</td>
</tr>
<tr>
<td>1379.69</td>
<td>HGTVVLTALGGILK</td>
</tr>
<tr>
<td>1607.81</td>
<td>VEADIAGHGQEVLR</td>
</tr>
<tr>
<td>1817.01</td>
<td>GLSDGEWQQVLNVWGK</td>
</tr>
<tr>
<td>1855.06</td>
<td>GHHEAELKPLAQSHATK</td>
</tr>
<tr>
<td>3241.65</td>
<td>YLEFISDAIIHVLSHPGNFGADAQGAMTK</td>
</tr>
</tbody>
</table>
PMF TARGET: Protein FASTA Sequence Database

UniProt (example)

The Universal Protein Resource (UniProt) is a comprehensive resource for protein sequence and annotation data. UniProt is a collaboration between the European Bioinformatics Institute (EBI), the Swiss Institute of Bioinformatics (SIB) and the Protein Information Resource (PIR). Across the three institutes close to 150 people are involved through different tasks such as database curation, software development and support.  

http://www.uniprot.org/help/about

Example:

20,292 entries for taxonomy: "Homo sapiens (Human) [9606]" AND reviewed:yes
PMF Search Parameters Review:

- Trypsin Proteolytic Enzyme and
- 0 Missed Cleave Site (#MC)

Example:
Theoretical Trypsin Digest Peptide Mass List (truncated) for a Single Protein:
Trypsin = Enzyme Missed Cleave Sites = 0

PMF Search Parameters:

- Trypsin Proteolytic Enzyme and
- 1 Missed Cleave Site
- MSO = Methionine Oxidation (amino acid modification)

PMF Parameter: Peptide Mass Deviation (+/- m/z)

Low resolution measurement:
631 +/- 1

High resolution measurement:
631.345 +/- 0.006


3 Mouse proteins with Trypsin peptide in this range

197 Mouse proteins with Trypsin peptide in this range

Range: 630 – 632
## Mass Measurement Error Calculation

<table>
<thead>
<tr>
<th>Error Expression</th>
<th>Fractional Error Expression</th>
<th>Multiply by:</th>
</tr>
</thead>
<tbody>
<tr>
<td>ppm*</td>
<td>$\Delta / \text{Theoretical value}$</td>
<td>1,000,000</td>
</tr>
<tr>
<td>%</td>
<td>$\Delta / \text{Theoretical value}$</td>
<td>100</td>
</tr>
</tbody>
</table>

where $\Delta = \text{Experimental (or observed)} - \text{Theoretical } m/z \text{ value}$

### Example:

| | 
|---|---|
| **Experimental/observed value** (i.e., the data acquired by the mass spectrometer) | $= 1480.107 \text{ m/z}$ |
| **Theoretical value** (calculated from periodic table, after the peak is identified) | $= 1480.028 \text{ m/z}$ |
| **Delta ($\Delta$)** | $= 0.079$ |

### Error Expression

<table>
<thead>
<tr>
<th>Error Expression</th>
<th>Error Equation</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>ppm*</td>
<td>$(0.079/1480.028)*1,000,000$</td>
<td>53 ppm</td>
</tr>
<tr>
<td>%</td>
<td>$(0.079/1480.028)*100$</td>
<td>0.0053%</td>
</tr>
</tbody>
</table>

* parts per million

### A USEFUL REFERENCE:
MASCOT Peptide Mass Fingerprint Search Tool

http://www.matrixscience.com/

Access Mascot Server

You are welcome to submit searches to this free Mascot Server. Searches of MS/MS data are limited to 1200 spectra and some functions, such as no enzyme searches, are unavailable. Automated searching of batches of files is not permitted. If you want to automate search submission, perform large searches, search additional sequence databases, or customise the modifications, quantitation methods, etc., you'll need to license your own, in-house copy of Mascot Server.

Peptide Mass Fingerprint

The experimental data are a list of peptide mass values from the digestion of a protein by a specific enzyme such as trypsin.

Perform search | Example of results report | Tutorial
MASCOT Peptide Mass Fingerprint

Your name: LeeAnn
Email: higgi022@umn.edu
Search title: Workshop Example
Database(s): SwissProt, NCBInr
Enzyme: Trypsin
Allow up to: 1 missed cleavages
Taxonomy: Mammalia (mammals)
Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Protein mass: kDa
Peptide tol.: 300 ppm
Mass values: M+, M+H, M-H-
Monoisotopic
Data file: No file selected.
Query:
940.3
1027.4
1044.4
1324.4
1342.4
Decoy: off
Report top: AUTO
Start Search... Reset Form
### PMF Search

<table>
<thead>
<tr>
<th>m/z</th>
<th>Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>631.71</td>
<td>NDIAAK</td>
</tr>
<tr>
<td>650.71</td>
<td>ELGFQG</td>
</tr>
<tr>
<td>662.72</td>
<td>ASEDLK</td>
</tr>
<tr>
<td>708.81</td>
<td>TEAEMK</td>
</tr>
<tr>
<td>748.90</td>
<td>ALELFIR</td>
</tr>
<tr>
<td>1272.44</td>
<td>LFTGHPETLEK</td>
</tr>
<tr>
<td>1379.69</td>
<td>HGTVVTLALGGILK</td>
</tr>
<tr>
<td>1607.81</td>
<td>VEAADIAGHGQEVLIR</td>
</tr>
<tr>
<td>1817.01</td>
<td>GLSDGEWQQVLNWGK</td>
</tr>
<tr>
<td>1855.06</td>
<td>GHHEAELKPLAQSHATK</td>
</tr>
<tr>
<td>3241.65</td>
<td>YLEFISDAILHLHSHPNGFADAQGAMTK</td>
</tr>
</tbody>
</table>

### QUERY (Experimental Data)

**Search**

Compare QUERY to THEOREUTICAL PEPTIDE MASS LIST for each protein in the database.

Parameters:
- Enzyme
- Missed cleave site
- Amino acid mods
- Mass tolerance

### Score

Probability-based MOWSE Score (often, the protein with highest number of peptide matches has the highest score)
Mascot Search Results

User: LeeAnn
Email: biggi022@umn.edu
Search title:
Database: NCBI nr 20140305 (37425594 sequences: 13257553858 residues)
Taxonomy: Metazoa (Animals) (5200995 sequences)
Timestamp: 13 Mar 2014 at 00:03:18 GMT
Top Score: 121 for gi|30749793, Chain A, K46e Variant Of Horse Heart Myoglobin

Mascot Score Histogram

Protein score is -10*Log(P), where P is the probability that the observed match is a random event. Protein scores greater than 80 are significant (p<0.05).

Concise Protein Summary Report

Format As: Concise Protein Summary
Significance threshold p< 0.05
Max. number of hits: AUTO
Preferred taxonomy: All entries

1. gi|30749793  Mass: 16942  Score: 121  Expect: 4.1e-06  Matches: 9
   Chain A, K46e Variant Of Horse Heart Myoglobin
   gi|17846624  Mass: 16941  Score: 119  Expect: 6.5e-06  Matches: 9
   Chain A, Myoglobin (Horse Heart) Wild-Type Complexed With Co
   gi|1255653511 Mass: 17072  Score: 116  Expect: 1.3e-05  Matches: 9
   myoglobin [Equus caballus]
   gi|1396769172  Mass: 16967  Score: 116  Expect: 1.3e-05  Matches: 9
   Chain B, Horse Heart Myoglobin: D44K60K MUTANT WITH ZINC (II) - Deuteroporphyrin Dimethyl Ester
**MASCOT Search Results**

**Protein View: gi|30749793**

**Chain A, K45c Variant Of Horse Heart Myoglobin**

**Database:** NCBI nr  
**Score:** 121  
**Expect:** 4.1e-06  
**Nominal mass (M):** 16042  
**Calculated pI:** 6.63  
**Taxonomy:** Equus caballus

Sequence similarity is available as an NCBI BLAST search of gi|30749793 against nr.

**Search parameters**

- **Enzyme:** Trypsin: cuts C-term side of KR unless next residue is P.
- **Fixed modifications:** Carboxymethyl (C)
- **Variable modifications:** Oxidation (M)
- **Mass values searched:** 19
- **Mass values matched:** 9

**Protein sequence coverage:** 67%

Matched peptides shown in **bold red.**

1. **GDSGMQVVRNWRPEG** TADVPQPLKL LPTDRFSLK HSDESSLK
2. **RTQELVPEN** EHSGLPVIK AEGFAEFLK MAHKKFLK DIYKHTLC
3. **POQ**

Unformatted sequence string: **153 residues** (for pasting into other applications).

[Options to sort peptides by residue number, increasing mass, or decreasing mass]

**Show predicted peptides also**

<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
<th>Observed</th>
<th>Mr(expt)</th>
<th>Mr(calc)</th>
<th>ppm</th>
<th>M</th>
<th>Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>1016.0000</td>
<td>1141.9327</td>
<td>1814.8552</td>
<td>52.8</td>
<td>0</td>
<td>-GLSDESMQVVRNWRPEG.V</td>
</tr>
<tr>
<td>17</td>
<td>31</td>
<td>1607.0000</td>
<td>1607.9292</td>
<td>1605.8470</td>
<td>90.5</td>
<td>0</td>
<td>R.YKLADTVQPLKLR.L</td>
</tr>
<tr>
<td>32</td>
<td>42</td>
<td>1277.7000</td>
<td>1377.6297</td>
<td>1277.6557</td>
<td>29.1</td>
<td>0</td>
<td>R.LPTDRFSLK.F</td>
</tr>
<tr>
<td>64</td>
<td>77</td>
<td>1379.0000</td>
<td>1377.5957</td>
<td>1377.8344</td>
<td>110</td>
<td>0</td>
<td>R.HAVVTLAOGIQ.L</td>
</tr>
<tr>
<td>79</td>
<td>96</td>
<td>1982.0000</td>
<td>1981.1927</td>
<td>1981.0453</td>
<td>72.4</td>
<td>1</td>
<td>R.KEHAEILFLAQSHKTVH</td>
</tr>
<tr>
<td>80</td>
<td>96</td>
<td>1854.1000</td>
<td>1853.0937</td>
<td>1852.9564</td>
<td>74.7</td>
<td>0</td>
<td>R.KEHAEIKFLQAFMTH.E</td>
</tr>
<tr>
<td>103</td>
<td>112</td>
<td>1883.2000</td>
<td>1884.1927</td>
<td>1884.0160</td>
<td>94.6</td>
<td>0</td>
<td>R.YLEFLNADTVQPLKLR.M</td>
</tr>
<tr>
<td>134</td>
<td>139</td>
<td>748.6000</td>
<td>747.5927</td>
<td>747.4279</td>
<td>221</td>
<td>0</td>
<td>R.AKQFL.R</td>
</tr>
<tr>
<td>146</td>
<td>153</td>
<td>941.6000</td>
<td>940.5927</td>
<td>940.4854</td>
<td>123</td>
<td>1</td>
<td>R.YKLADTVQPLKLR.</td>
</tr>
</tbody>
</table>

• Is the protein ID experimentally rational?
• Does the MW of protein in search results match MW determined by SDS-PAGE?
• Does the pI of protein in search results match pI determined by 2D-PAGE?

➢ Perform MS/MS if no protein match, then perform tandem MS database search
Precursor Mass-based Search

Example analysis:
Peptide Tandem Mass Spectrometry from LC-MS/MS Experiment
Three Strategies for Protein Inference from Peptide MS or MS/MS Data

DATA INPUT:
- Peptide Mass List (MS1)
- Peptide Mass Fingerprint (PMF) (historical)

Search TYPE:
- Peptide MS/MS (MS2)

Precursor Mass-based

Sequence-based (de novo or mass tag)
## Tandem MS: Precursor Mass-based Search

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input (Data)</strong></td>
<td>Precursor Mass &amp; Product Ion Masses</td>
</tr>
<tr>
<td></td>
<td>• Charge state</td>
</tr>
<tr>
<td></td>
<td>• Average or Monoisotopic</td>
</tr>
<tr>
<td><strong>Target</strong></td>
<td>Protein FASTA Sequence Database</td>
</tr>
<tr>
<td><strong>Search Parameters</strong></td>
<td>Proteolytic enzyme</td>
</tr>
<tr>
<td></td>
<td># Missed enzyme cleave sites</td>
</tr>
<tr>
<td></td>
<td>Amino acid modifications</td>
</tr>
<tr>
<td></td>
<td>Mass tolerances:</td>
</tr>
<tr>
<td></td>
<td>• Peptide/Precursor Mass</td>
</tr>
<tr>
<td></td>
<td>• Product Ions Masses</td>
</tr>
</tbody>
</table>
Tandem MS: Precursor Mass-based Search

1. Each MS/MS spectra has 2 data-rich Components

2. Database search for each MS/MS spectra has 2 STEPS
Tandem MS: Precursor Mass-based Search

Calculate Intact Peptide Mass from precursor \( m/z \) signal

- Precursor \( m/z \) (monoisotopic) = 894.0941
- Precursor Charge State = 3
- Precursor Mass (Da) = 2679.2559
2. Search the protein database for peptides within a specified mass tolerance*

- Target peptide mass
  - 2679.2559

- Target peptide mass range
  
  example = 50 ppm (0.0005%) mass tolerance
  
  - 2679.2559 +/- 0.1340
  - 2679.1219 – 2679.3899

* Tolerance is chosen based on the expected error in mass values
Find Peptide Candidate Matches in the range 2679.1219 – 2679.3899 m/z

<table>
<thead>
<tr>
<th>Mr(expt)</th>
<th>Mr(calc)</th>
<th>ppm</th>
<th>Peptide Candidate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2679.2559</td>
<td>2679.3798</td>
<td>-46.2</td>
<td>GAVQVFIMLLLATVSDCAVITGACER</td>
</tr>
<tr>
<td>2679.2559</td>
<td>2679.1846</td>
<td>26.6</td>
<td>SYDPPCPGHWTPEAPGSHTCPCGLPR</td>
</tr>
<tr>
<td>2679.2559</td>
<td>2679.2889</td>
<td>-12.3</td>
<td>DGETAEQQGVPPVFAPGPGPGGLGGAPGGR</td>
</tr>
<tr>
<td>2679.2559</td>
<td>2679.3286</td>
<td>-27.1</td>
<td>GAGALALGASEPCNLSSAAPLPDGAATAAR</td>
</tr>
<tr>
<td>2679.2559</td>
<td>2679.2010</td>
<td>20.5</td>
<td>AVDFQEAQSYADDNSLLFMETSAK</td>
</tr>
<tr>
<td>2679.2559</td>
<td>2679.3102</td>
<td>-20.3</td>
<td>GTYLPQTYIQEQEMVVTEHVSDK</td>
</tr>
<tr>
<td>2679.2559</td>
<td>2679.25575</td>
<td>0.1</td>
<td>ERPTPSLNNNC(carbamidomethyl)TTSEDSLVLYNR</td>
</tr>
</tbody>
</table>

... and many more...
Tandem MS: Precursor Mass-based Search

For each peptide candidate....

Compare theoretical fragment ion $m/z$ values to experimental fragment ion $m/z$ values from MS/MS spectrum
2. **COMPARE**

theoretical product ions (table below) to experimental product Ions from MS2 spectrum (matched ions are colored blue and red)

<table>
<thead>
<tr>
<th>#</th>
<th>Immonium</th>
<th>b</th>
<th>b-H2O</th>
<th>a</th>
<th>a-H2O</th>
<th>b (2+)</th>
<th>Seq</th>
<th>y</th>
<th>y-H2O</th>
<th>y-NH3</th>
<th>y (2+)</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>102.055</td>
<td>130.087</td>
<td>112.087</td>
<td>102.055</td>
<td>84.045</td>
<td>65.525</td>
<td>E</td>
<td>2551.220</td>
<td>2533.210</td>
<td>2534.193</td>
<td>1276.110</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>129.103</td>
<td>286.152</td>
<td>268.142</td>
<td>258.156</td>
<td>240.146</td>
<td>143.576</td>
<td>R</td>
<td>2378.092</td>
<td>1198.060</td>
<td>1050.483</td>
<td>1799.992</td>
<td>22</td>
</tr>
<tr>
<td>3</td>
<td>70.066</td>
<td>383.189</td>
<td>365.189</td>
<td>355.214</td>
<td>337.199</td>
<td>192.102</td>
<td>P</td>
<td>2395.119</td>
<td>2377.109</td>
<td>2378.092</td>
<td>1198.060</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>74.061</td>
<td>484.253</td>
<td>466.241</td>
<td>456.258</td>
<td>438.243</td>
<td>242.626</td>
<td>T</td>
<td>2280.056</td>
<td>2281.040</td>
<td>1149.533</td>
<td>1439.972</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>70.066</td>
<td>581.305</td>
<td>563.294</td>
<td>553.310</td>
<td>535.299</td>
<td>291.152</td>
<td>P</td>
<td>2197.019</td>
<td>2179.009</td>
<td>1099.010</td>
<td>950.425</td>
<td>19</td>
</tr>
<tr>
<td>6</td>
<td>60.045</td>
<td>668.338</td>
<td>650.329</td>
<td>640.354</td>
<td>622.331</td>
<td>334.668</td>
<td>S</td>
<td>2099.966</td>
<td>2081.956</td>
<td>2082.939</td>
<td>1050.483</td>
<td>18</td>
</tr>
<tr>
<td>7</td>
<td>86.097</td>
<td>781.423</td>
<td>763.413</td>
<td>753.428</td>
<td>735.428</td>
<td>391.194</td>
<td>L</td>
<td>2012.934</td>
<td>1994.924</td>
<td>1995.907</td>
<td>1006.967</td>
<td>17</td>
</tr>
<tr>
<td>9</td>
<td>87.056</td>
<td>1009.513</td>
<td>991.520</td>
<td>981.519</td>
<td>963.501</td>
<td>505.253</td>
<td>N</td>
<td>1785.807</td>
<td>1767.797</td>
<td>1768.780</td>
<td>893.379</td>
<td>15</td>
</tr>
<tr>
<td>10</td>
<td>87.056</td>
<td>1123.553</td>
<td>1105.556</td>
<td>1095.559</td>
<td>1077.544</td>
<td>562.275</td>
<td>N</td>
<td>1671.764</td>
<td>1653.754</td>
<td>1654.737</td>
<td>836.382</td>
<td>14</td>
</tr>
<tr>
<td>11</td>
<td>133.043</td>
<td>1283.580</td>
<td>1265.570</td>
<td>1255.599</td>
<td>1237.575</td>
<td>642.290</td>
<td>C(+57.02)</td>
<td>1557.699</td>
<td>1539.711</td>
<td>1540.694</td>
<td>779.361</td>
<td>13</td>
</tr>
<tr>
<td>12</td>
<td>74.061</td>
<td>1384.628</td>
<td>1366.617</td>
<td>1356.616</td>
<td>1338.622</td>
<td>692.814</td>
<td>T</td>
<td>1397.692</td>
<td>1379.680</td>
<td>1380.664</td>
<td>699.345</td>
<td>12</td>
</tr>
<tr>
<td>13</td>
<td>74.061</td>
<td>1485.676</td>
<td>1467.666</td>
<td>1457.681</td>
<td>1439.706</td>
<td>743.338</td>
<td>T</td>
<td>1296.653</td>
<td>1278.632</td>
<td>1279.616</td>
<td>648.821</td>
<td>11</td>
</tr>
<tr>
<td>14</td>
<td>60.045</td>
<td>1572.708</td>
<td>1554.697</td>
<td>1544.713</td>
<td>1526.702</td>
<td>786.854</td>
<td>N</td>
<td>1195.599</td>
<td>1177.585</td>
<td>1178.581</td>
<td>598.298</td>
<td>10</td>
</tr>
<tr>
<td>15</td>
<td>102.055</td>
<td>1701.750</td>
<td>1683.740</td>
<td>1673.755</td>
<td>1655.745</td>
<td>851.375</td>
<td>E</td>
<td>1108.568</td>
<td>1090.577</td>
<td>1091.536</td>
<td>554.782</td>
<td>9</td>
</tr>
<tr>
<td>16</td>
<td>88.040</td>
<td>1816.777</td>
<td>1798.767</td>
<td>1788.782</td>
<td>1770.772</td>
<td>908.889</td>
<td>D</td>
<td>979.526</td>
<td>961.510</td>
<td>962.491</td>
<td>490.260</td>
<td>8</td>
</tr>
<tr>
<td>17</td>
<td>60.045</td>
<td>1903.809</td>
<td>1885.799</td>
<td>1875.814</td>
<td>1857.804</td>
<td>952.405</td>
<td>S</td>
<td>864.497</td>
<td>846.483</td>
<td>847.478</td>
<td>432.747</td>
<td>7</td>
</tr>
<tr>
<td>19</td>
<td>72.081</td>
<td>2115.962</td>
<td>2097.951</td>
<td>2087.967</td>
<td>2069.956</td>
<td>1058.481</td>
<td>V</td>
<td>664.380</td>
<td>646.367</td>
<td>647.351</td>
<td>332.689</td>
<td>5</td>
</tr>
<tr>
<td>20</td>
<td>86.097</td>
<td>2229.046</td>
<td>2211.035</td>
<td>2201.051</td>
<td>2183.040</td>
<td>1115.023</td>
<td>L</td>
<td>565.311</td>
<td>547.299</td>
<td>548.286</td>
<td>283.155</td>
<td>4</td>
</tr>
<tr>
<td>22</td>
<td>87.056</td>
<td>2506.152</td>
<td>2488.141</td>
<td>2478.157</td>
<td>2460.146</td>
<td>1253.576</td>
<td>N</td>
<td>289.163</td>
<td>271.151</td>
<td>272.137</td>
<td>145.081</td>
<td>2</td>
</tr>
<tr>
<td>23</td>
<td>129.114</td>
<td>2718.198</td>
<td>2690.187</td>
<td>2670.192</td>
<td>2652.181</td>
<td>1327.606</td>
<td>R</td>
<td>175.119</td>
<td>157.099</td>
<td>158.092</td>
<td>88.059</td>
<td>1</td>
</tr>
</tbody>
</table>
Evaluate Spectrum for Peptide Candidate Match Quality
Tandem MS: Precursor Mass-based Search

SUMMARY- for Each MS/MS Spectrum:
• Generate **theoretical** fragment ion tables for all peptide candidates
• Compare **theoretical** fragment ions to **experimental** fragment ions from MS2 spectrum
• **Score** all candidate peptides
• **Rank** peptides by Score
Tandem MS: Precursor Mass-based Search

STEP 1
Find *Theoretical* Peptide Matches from *in silico* Protein Digest in the range: “*Experimental* Precursor Mass +/- Mass Tolerance”

STEP 2
For each candidate peptide from Step 1: Compare *Theoretical* Product Ions (b & y, etc) to *Experimental* Product Ions (data)

STEP 3
*SCORE*: many software programs/ algorithms
Peptide rank
PROTEIN Report and Grouping: many variations
Tandem MS: Precursor Mass-based search software at UM

• Sequest
• X!Tandem
• MASCOT
Sequence-based Search

Example analysis:
Peptide Tandem Mass Spectrometry from LC-MS/MS Experiment
Three Strategies for Protein Inference from Peptide MS or MS/MS Data

DATA INPUT:

Search TYPE:

1. Peptide Mass List (MS1)
2. Peptide Mass Fingerprint (PMF) (historical)
3. Peptide MS/MS (MS2)

Precursor Mass-based
(de novo or mass tag)
MS/MS Sequence-based Software at UM

- PEAKS (BSI)
- ProteinPilot® (AB Sciex)
3 Sequence-based Searches: WHEN?

- Unsequenced genome (protein(s) of interest are not in the database) → HOMOLOGY-based search
- Search for amino acid MUTATIONS
- Search for LARGE NUMBERS of Post Translational Modifications (PTM’s)
STEP 1:
Obtain [full or partial] amino acid sequence string directly from the spectrum by de novo Sequencing

de novo (Latin) "from the beginning"
STEP 1: *de novo sequencing*: amino acid sequence is determined from delta mass values for a series of successive peptide b- or y-type product ions from a Product Ion (MS/MS) spectrum.

**EXAMPLE de novo sequencing:**

Mass differences between peaks correspond to amino acid residue masses.

See amino acid residue masses: http://www.ionsource.com/Card/aatable/aatable.htm
STEP 3: Candidate peptide sequences within protein sequences are identified with Precursor mass AND sequence tag information

Include:
- Enzyme specificity
- Mass tolerance
- Amino acid modifications
- Amino Acid mutations

STEP 4: Candidates are Ranked and SCORED
Peptide Library Search

• Database search performed on TANDEM MASS SPECTRA (instead of protein sequence)
• Spectral Libraries originate from previously acquired peptide tandem MS
• Example public spectral library repositories
  • http://www.peptideatlas.org/
  • https://www.nist.gov/programs-projects/peptide-mass-spectral-libraries
  • https://massive.ucsd.edu/ProteoSAFe/static/massive-kb-libraries.jsp
Galaxy-p.org
Customizable multi-omics informatics platform

Galaxy-p is funded by the NSF and NCI's Informatics Technology for Cancer Research program
Project PI: Timothy Griffin (tgriffin@umn.edu)
Summary

• Database search programs match peptide mass spec data to amino acid sequences from proteins
• Search programs provide protein—level reports that contain peptide spectral match results
• Algorithms assign scores for Peptide Spectral Matches (PSM’s) and rank protein and peptide level results by scores
• Interactive database search RESULTS inspection TODAY in the afternoon using Scaffold
Protein Validation (Statistical Inference)
<table>
<thead>
<tr>
<th>Outline</th>
<th>Terminology</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Statistics</td>
<td>• P-value</td>
</tr>
<tr>
<td>• Peptide scoring</td>
<td>• Multiple hypothesis testing</td>
</tr>
<tr>
<td>• Multiple hypotheses</td>
<td>• False discovery rate</td>
</tr>
<tr>
<td>• False discovery rate</td>
<td>• Target / Decoy</td>
</tr>
<tr>
<td>• Protein inference</td>
<td>• One hit wonders</td>
</tr>
<tr>
<td></td>
<td>• Shared peptide</td>
</tr>
<tr>
<td></td>
<td>• Parsimony</td>
</tr>
</tbody>
</table>
Statistics
Classical Example

PROBABLE CAUSE
A *P* value measures whether an observed result can be attributed to chance. But it cannot answer a researcher’s real question: what are the odds that a hypothesis is correct? Those odds depend on how strong the result was and, most importantly, on how plausible the hypothesis is in the first place.

Before the experiment
The plausibility of the hypothesis — the odds of it being true — can be estimated from previous experiments, conjectured mechanisms and other expert knowledge. Three examples are shown here.

The measured *P* value
A value of 0.05 is conventionally deemed ‘statistically significant’; a value of 0.01 is considered ‘very significant’.

After the experiment
A small *P* value can make a hypothesis more plausible, but the difference may not be dramatic.
Significance Testing

• Assumptions
  – Normal distribution
  – Independent
  – Scoring model
• Null Hypothesis
• Test statistic
• P-value
• Conclusion

http://en.wikipedia.org/wiki/P-value
Peptide Spectrum Match Scoring

• Visual inspection
• Scoring function

\[ f(S,P) \] measures the quality of the match between spectrum \( S \) and peptide sequence \( P \)
Peptide Spectrum Match Scoring

• Look at fragment ions
  – predicted mass
  – expected intensity
  – compare to next best
  – presence of immonium ions
  – etc.
MULTIPLE HYPOTHESIS TESTING & FALSE DISCOVERY RATE
Multiple Hypothesis Testing Errors

The problem of multiplicity arises from the fact that as we increase the number of hypotheses in a test, we also increase the likelihood of witnessing a rare event, and therefore, the chance to reject the null hypotheses when it's true (Type I error).
# Common Statistical Measures

### Truth

<table>
<thead>
<tr>
<th></th>
<th>$H_0$ True</th>
<th>$H_0$ False</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ True</td>
<td>Correct (TN)</td>
<td>Type II Error ($\beta$) (FN)</td>
</tr>
<tr>
<td>$H_0$ False</td>
<td>Type I Error ($\alpha$) (FP)</td>
<td>Correct (TP)</td>
</tr>
</tbody>
</table>

$H_0 = \text{null hypothesis}$

- $= \text{no quantitative difference}$
- $= \text{no match to peptide}$

### Your Findings

- **True Positive Rate:** $\frac{TP}{(TP + FN)}$ Sensitivity
- **False Positive Rate:** $\frac{FP}{(FP + TN)}$ Specificity
Multiple Hypothesis Testing Correction

• Bonferroni
  – Simplest correction: Bonferroni correction – divide by number of tests.
    • Example – If you want final error rate of 5% and you have 1000 proteins (1000 hypothesis tests), then divide the single test P-value by 1000, giving you $P = 0.00005$ as the threshold you would apply to have a 5% error rate in the proteins that pass this threshold
  – Too stringent – want to accept more errors
Multiple Hypothesis Testing Correction

- **False Discovery Rate (FDR)**
  - Benjamini & Hochberg, 1994
  - Controls for the expected proportion of falsely rejected hypotheses
- Requires we know false identifications
- **Decoy database**
  - Reversed or random sequences
Common Statistical Measures

<table>
<thead>
<tr>
<th>Truth</th>
<th>H₀ True</th>
<th>H₀ False</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₀ True</td>
<td>Correct (TN)</td>
<td>Type II Error (β) (FN)</td>
</tr>
<tr>
<td>H₀ False</td>
<td>Type I Error (α) (FP)</td>
<td>Correct (TP)</td>
</tr>
</tbody>
</table>

**H₀ = null hypothesis**
- = no quantitative difference
- = no match to peptide

**False Discovery Rate:**  \( \frac{FP}{(FP + TP)} \)

**True Positive Rate:**  \( \frac{TP}{(TP + FN)} \) Sensitivity

**False Positive Rate:**  \( \frac{FP}{(FP + TN)} \) Specificity

**Positive Predictive Value:**  \( \frac{TP}{(TP + FP)} \)
Elias JE, Gygi SP.

---

Target (Forward) / Decoy (Reverse)

---

>gi|6319436|ref|NP_009518.1| Pol12p [Saccharomyces cerevisiae S288c]
MSGSIDVITHFGPDAKDPEIITAENLTKHALSVEVLDYIKWEQFSNQRRTHTDLTSKNIDEFKQFLQMEK
RANQISSSSKVNTSTKPVIKKSNLSSPLGALSIPKPTPLKRRKLGPFSLSDSKQTYNGSEATNEKGNSSL
KLEFTPGMAEADVGPSAPLHAKSSDAKTPGSSSTFQTPTTTNTRSNQVPAEGILDSLPENIEISSGPNV
GLLSTEEPSYNYQVKEPYADKYYKFRTRMRQNLQEASDVLDQQIESFTKIQQNYHKLSPNDFAADPTIQSSEIY
AVGRIVPDSPTYDKFLNPESSLESLETSMGGVRVRDLQSVNELSFLGQVAFKGKNANGDYFTVNSILPL
YPYPNSPVSTSQELQEFQANLEGSSLKVIVTCGPYFANDNFSLELLQEFIDSINNEVKPHLIMFGPFIIDT
PSGKLNPFPQFQTKPLDELFLSKFTPIKIPHTQTQIPSTKDAISNHAYPASLIRKAKLQPLKRNFKCMAN
PSSFQINEIYFVICGNSNVTDKFLKKEVIGGTSSSSRLDRVRSEHILQFRRYPPFFGSIIVTIKPDVST
KETNDMSKEEKEVYIEISGADLDVSYLGLTEFVGGFSPDUMIIPSEQLHQFARVQNVVINPGRFIRATGN
RGSYAQiTVQCPDLEDGKTLVEEGEPVYLHNWKRARVDLIS

###REV###>gi|6319436|ref|NP_009518.1| Pol12p [Saccharomyces cerevisiae S288c]
SAILDVRARKWNHLVPEEVEGLTKGDDELDCVQITQAYSegratedTARIFRGPNIVVQVVRAFHQLLESPI
MIDPSFGVETGLYVLDAGSIHEYVEEKEMDNTETKTSVDPKIRISGFIPYYRROQLIESVRRD
LRLYRSSTTTGKIEVCLKDFTVNSCGFYIENIQFSSPAMCKFNRKQPLAKRISAOHPHAIISATKPSILV
TQIHPSTKLIPTFLKFLFLEDLTKPTKFQFPNPLKGSAILPHTIDFFPGFMILVHPKVENNISDIEQLELSFN
DAFYPGCTIVKLSGELNAQFEQLERQSTSVPNYPLPISNTVFYDQGNNAGKFQAVIQLFFSLENQSLDR
VRRVGGMSTESSEPNFLKDYTPSVPVRGAVIESQSIQITPAFNDNPSLYHNYIQKTFSEIQDDLVDSDA
EQLNQRMTRFKYKADYFPEVKYPVNYIPNOTSSLGVNPNGSSIEINEPNLSLDIEGAPVQNRSTTPTNTPTQ
FTSSGPTKAHSPLASDGVDAEMGPTFELKLSSNKENTAEESGVNTQKSSDLFSFGLKRLKSTTP
TKPISLGFLPSSNLKIKVPKKTSTNVKSSSSIQNARKEMQQLFLQKFREDINKSTLDTHTQRQQNSFQEWKIL
DEVSLAHLKTLENAtiEIPKDAPFHTIVDISGSM
Target / Decoy

Elias JE, Gygi SP.
FDR in Practice

1. Sort all peptide hits by score, descending.
2. Count how many target hits are greater than or equal to a given score.
3. Count how many decoy hits are greater than or equal to a given score.
4. Estimate the total number of incorrect hits (false positive, FP) from observed decoy hits \(d\) greater than or equal to a given score: \(FP=dfs\)
5. Calculate statistics related to FP for each given score threshold.
6. Select score threshold based on a desired statistic threshold.

Methods Mol Biol. 2010; 604: 55–71
<table>
<thead>
<tr>
<th>#</th>
<th>Spectrum</th>
<th>Accession</th>
<th>Peptide</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>P35908</td>
<td>scan 3632</td>
<td>GFSSGSAVVSGGSR</td>
<td>4.6</td>
</tr>
<tr>
<td>2</td>
<td>P0AFY8</td>
<td>scan 3609</td>
<td>FSAASQPAAPVTK</td>
<td>3.7</td>
</tr>
<tr>
<td>3</td>
<td>P0A940</td>
<td>scan 3629</td>
<td>GFQSNTIGPK</td>
<td>3.0</td>
</tr>
<tr>
<td>4</td>
<td>P0A6F9</td>
<td>scan 3635</td>
<td>STRGEVLAVGNGR</td>
<td>2.2</td>
</tr>
<tr>
<td>5</td>
<td>P0A870</td>
<td>scan 3636</td>
<td>ELAESEGAIER</td>
<td>2.1</td>
</tr>
<tr>
<td>6</td>
<td>P0A799</td>
<td>scan 3607</td>
<td>ADLNVPVKDGK</td>
<td>1.9</td>
</tr>
<tr>
<td>7</td>
<td>P0ABC7</td>
<td>scan 3626</td>
<td>EAEAYTNEVQPR</td>
<td>1.6</td>
</tr>
<tr>
<td>8</td>
<td>P0A853</td>
<td>scan 3602</td>
<td>IRVIEPVKR</td>
<td>1.4</td>
</tr>
<tr>
<td>9</td>
<td>P38489</td>
<td>scan 3623</td>
<td>KLTPEOAEQIK</td>
<td>0.9</td>
</tr>
<tr>
<td>10</td>
<td>P00448</td>
<td>scan 3616</td>
<td>GTTLQGDLK</td>
<td>0.8</td>
</tr>
<tr>
<td>11</td>
<td>P09546</td>
<td>scan 3621</td>
<td>LLPGPTGER</td>
<td>0.4</td>
</tr>
<tr>
<td>12</td>
<td>P0AFG8</td>
<td>scan 3615</td>
<td>AFLEGR</td>
<td>0.2</td>
</tr>
<tr>
<td>13</td>
<td>P14565</td>
<td>scan 3624</td>
<td>SAADVAIMK</td>
<td>0.0</td>
</tr>
<tr>
<td>14</td>
<td>rev_P06864</td>
<td>scan 3613</td>
<td>EGSLAVNVQGDAAIR</td>
<td>-0.4</td>
</tr>
<tr>
<td>15</td>
<td>P36562</td>
<td>scan 3604</td>
<td>DEEEVVGIGANLPTDK</td>
<td>-0.7</td>
</tr>
<tr>
<td>16</td>
<td>P0A9C5</td>
<td>scan 3606</td>
<td>IPVVSPK</td>
<td>-0.7</td>
</tr>
<tr>
<td>17</td>
<td>P0ABB0</td>
<td>scan 3611</td>
<td>ASTISNVVR</td>
<td>-0.7</td>
</tr>
<tr>
<td>18</td>
<td>rev_Q2EEU2</td>
<td>scan 3614</td>
<td>KFVALTCDDLLGER</td>
<td>-0.8</td>
</tr>
<tr>
<td>19</td>
<td>rev_P0ACL5</td>
<td>scan 3620</td>
<td>NNESAALMKEYCR</td>
<td>-0.9</td>
</tr>
<tr>
<td>20</td>
<td>rev_P37309</td>
<td>scan 3633</td>
<td>SDGSCNQRALNR</td>
<td>-0.9</td>
</tr>
<tr>
<td>21</td>
<td>P32132</td>
<td>scan 3627</td>
<td>VEETEDADAFRVSGR</td>
<td>-1.0</td>
</tr>
<tr>
<td>22</td>
<td>P37342</td>
<td>scan 3618</td>
<td>ILTQDEIDV</td>
<td>-1.0</td>
</tr>
<tr>
<td>23</td>
<td>rev_P0ADK0</td>
<td>scan 3610</td>
<td>IANVSDVVR</td>
<td>-1.2</td>
</tr>
<tr>
<td>24</td>
<td>P0AG93</td>
<td>scan 3601</td>
<td>LGMKREHMLQQK</td>
<td>-1.3</td>
</tr>
<tr>
<td>#</td>
<td>Spectrum</td>
<td>Accession</td>
<td>Peptide</td>
<td>Score</td>
</tr>
<tr>
<td>----</td>
<td>----------</td>
<td>-----------</td>
<td>------------------</td>
<td>-------</td>
</tr>
<tr>
<td>1</td>
<td>scan 3632</td>
<td>P35908</td>
<td>GFSSGSAVVSGGSR</td>
<td>4.6</td>
</tr>
<tr>
<td>2</td>
<td>scan 3609</td>
<td>P0AFY8</td>
<td>FSAASQPAAPVTK</td>
<td>3.7</td>
</tr>
<tr>
<td>3</td>
<td>scan 3629</td>
<td>P0A940</td>
<td>GQOSNTIGPK</td>
<td>3.0</td>
</tr>
<tr>
<td>4</td>
<td>scan 3635</td>
<td>P0A6F9</td>
<td>STRGEVLAVNGR</td>
<td>2.2</td>
</tr>
<tr>
<td>5</td>
<td>scan 3636</td>
<td>P0A870</td>
<td>ELAEEGAIER</td>
<td>2.1</td>
</tr>
<tr>
<td>6</td>
<td>scan 3607</td>
<td>P0A799</td>
<td>ADLNVPVKGDK</td>
<td>1.9</td>
</tr>
<tr>
<td>7</td>
<td>scan 3626</td>
<td>P0ABC7</td>
<td>EAEAYTEVQPR</td>
<td>1.6</td>
</tr>
<tr>
<td>8</td>
<td>scan 3602</td>
<td>P0A853</td>
<td>IRVIEPVKR</td>
<td>1.4</td>
</tr>
<tr>
<td>9</td>
<td>scan 3623</td>
<td>P38489</td>
<td>KLTPEOAEQIK</td>
<td>0.9</td>
</tr>
<tr>
<td>10</td>
<td>scan 3616</td>
<td>P00448</td>
<td>GTTLQGDLK</td>
<td>0.8</td>
</tr>
<tr>
<td>11</td>
<td>scan 3621</td>
<td>P09546</td>
<td>LLPGPTGER</td>
<td>0.4</td>
</tr>
<tr>
<td>12</td>
<td>scan 3615</td>
<td>P0AFG8</td>
<td>AFLEGR</td>
<td>0.2</td>
</tr>
<tr>
<td>13</td>
<td>scan 3624</td>
<td>P14565</td>
<td>SAADVAMK</td>
<td>0.0</td>
</tr>
<tr>
<td>14</td>
<td>scan 3613</td>
<td>rev_P06864</td>
<td>EGSLAVNVQGDAAIR</td>
<td>-0.4</td>
</tr>
<tr>
<td>15</td>
<td>scan 3604</td>
<td>P36562</td>
<td>DPEEVVGIANLPTDK</td>
<td>-0.7</td>
</tr>
<tr>
<td>16</td>
<td>scan 3606</td>
<td>P0A9C5</td>
<td>IPVVSPK</td>
<td>-0.7</td>
</tr>
<tr>
<td>17</td>
<td>scan 3611</td>
<td>P0ABB0</td>
<td>ASTISNVVR</td>
<td>-0.7</td>
</tr>
<tr>
<td>18</td>
<td>scan 3614</td>
<td>rev_Q2EEU2</td>
<td>KFVALTCDTLLGGER</td>
<td>-0.8</td>
</tr>
<tr>
<td>19</td>
<td>scan 3620</td>
<td>rev_P0ACL5</td>
<td>NNESAALMKEYCR</td>
<td>0.9</td>
</tr>
<tr>
<td>20</td>
<td>scan 3633</td>
<td>rev_P37309</td>
<td>SDGSCNQRALNP</td>
<td>-0.9</td>
</tr>
<tr>
<td>21</td>
<td>scan 3627</td>
<td>P32132</td>
<td>KEFTGADAFAFRVSGR</td>
<td>-1.0</td>
</tr>
<tr>
<td>22</td>
<td>scan 3618</td>
<td>P37342</td>
<td>ILTQDEIDVPR</td>
<td>-1.0</td>
</tr>
<tr>
<td>23</td>
<td>scan 3616</td>
<td>rev_P0ADK0</td>
<td>IANVSDVVPR</td>
<td>-1.2</td>
</tr>
<tr>
<td>24</td>
<td>scan 3601</td>
<td>P0AG93</td>
<td>LGMKREHMLQQK</td>
<td>-1.3</td>
</tr>
<tr>
<td>#</td>
<td>Spectrum</td>
<td>Accession</td>
<td>Peptide</td>
<td>Score</td>
</tr>
<tr>
<td>----</td>
<td>----------</td>
<td>-----------</td>
<td>---------------------</td>
<td>-------</td>
</tr>
<tr>
<td>1</td>
<td>scan 3632</td>
<td>P35908</td>
<td>GFSSGSAVVSGGSR</td>
<td>4.6</td>
</tr>
<tr>
<td>2</td>
<td>scan 3609</td>
<td>P0AFY8</td>
<td>FSAASQPAAPVTK</td>
<td>3.7</td>
</tr>
<tr>
<td>3</td>
<td>scan 3629</td>
<td>P0A940</td>
<td>GFQSNTIGPK</td>
<td>3.0</td>
</tr>
<tr>
<td>4</td>
<td>scan 3635</td>
<td>P0A6F9</td>
<td>STRGEVLAVGNGR</td>
<td>2.2</td>
</tr>
<tr>
<td>5</td>
<td>scan 3636</td>
<td>P0A870</td>
<td>ELAESEGAIER</td>
<td>2.1</td>
</tr>
<tr>
<td>6</td>
<td>scan 3607</td>
<td>P0A799</td>
<td>ADLNVPVKDGK</td>
<td>1.9</td>
</tr>
<tr>
<td>7</td>
<td>scan 3626</td>
<td>P0ABC7</td>
<td>EAEAYTNEVQPR</td>
<td>1.6</td>
</tr>
<tr>
<td>8</td>
<td>scan 3602</td>
<td>P0A853</td>
<td>IRVIEPVKR</td>
<td>1.4</td>
</tr>
<tr>
<td>9</td>
<td>scan 3623</td>
<td>P38489</td>
<td>KLTPEOAEQIK</td>
<td>0.9</td>
</tr>
<tr>
<td>10</td>
<td>scan 3616</td>
<td>P00448</td>
<td>GTTLQGDLK</td>
<td>0.8</td>
</tr>
<tr>
<td>11</td>
<td>scan 3621</td>
<td>P09546</td>
<td>LLPGPTGER</td>
<td>0.4</td>
</tr>
<tr>
<td>12</td>
<td>scan 3615</td>
<td>P0AFG8</td>
<td>AFLEGR</td>
<td>0.2</td>
</tr>
<tr>
<td>13</td>
<td>scan 3624</td>
<td>P14565</td>
<td>SAADVAIMK</td>
<td>0.0</td>
</tr>
<tr>
<td>14</td>
<td>scan 3613</td>
<td>rev_P06864</td>
<td>EGSLAVNVQGDAAIR</td>
<td>-0.4</td>
</tr>
<tr>
<td>15</td>
<td>scan 3604</td>
<td>P36562</td>
<td>DPEEVVIGANLPTDK</td>
<td>-0.7</td>
</tr>
<tr>
<td>16</td>
<td>scan 3606</td>
<td>P0A9C5</td>
<td>IPVVSSPK</td>
<td>-0.7</td>
</tr>
<tr>
<td>17</td>
<td>scan 3611</td>
<td>P0ABB0</td>
<td>ASTISNVVR</td>
<td>-0.7</td>
</tr>
<tr>
<td>18</td>
<td>scan 3614</td>
<td>rev_Q2EEU2</td>
<td>KFVALTCDTLLLGER</td>
<td>-0.8</td>
</tr>
<tr>
<td>19</td>
<td>scan 3620</td>
<td>rev_P0ACL5</td>
<td>NNESAALMKEYCR</td>
<td>-0.9</td>
</tr>
<tr>
<td>20</td>
<td>scan 3633</td>
<td>rev_P37309</td>
<td>SGSCNQRALNP</td>
<td>-0.9</td>
</tr>
<tr>
<td>21</td>
<td>scan 3627</td>
<td>P32132</td>
<td>LFEETSADAFAFRVSGR</td>
<td>-1.0</td>
</tr>
<tr>
<td>22</td>
<td>scan 3618</td>
<td>P37342</td>
<td>ILTQDEIDVR</td>
<td>-1.0</td>
</tr>
<tr>
<td>23</td>
<td>scan 3616</td>
<td>rev_P0ADK0</td>
<td>IANVSDVVPR</td>
<td>-1.2</td>
</tr>
<tr>
<td>24</td>
<td>scan 3601</td>
<td>P0AG93</td>
<td>LGMKREHMLQQK</td>
<td>-1.3</td>
</tr>
</tbody>
</table>
From peptides to proteins...

PROTEIN INFERENCE
From Peptides to Proteins

AEPTIR 85%
IDVCIVLLQHK 65%
NTGDR 25%

0.15 * 0.35 * 0.75 = 0.04

Peptide 1
Peptide 2
Peptide 3
Peptide 4
Peptide 5
Peptide 6
Peptide 7
Peptide 8
Peptide 9
Peptide 10

Correct Protein A
Correct Protein B
Incorrect Protein C
Incorrect Protein D

80% Peptides  →  50% Proteins
One Hit Wonders

95%  IDVCIVLLQHK  →  Protein (5%)
One Hit Wonders

• Quantify as a score:
  If different peptides agree: Good!
  If peptides are one-hit-wonders: Bad!
• Protein Prophet, etc.
• Then can use FDR at protein level
<table>
<thead>
<tr>
<th>#</th>
<th>Accession</th>
<th>Protein Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>P0ABH7</td>
<td>4258.08</td>
</tr>
<tr>
<td>2</td>
<td>P0ABJ9</td>
<td>2423.84</td>
</tr>
<tr>
<td>3</td>
<td>P0A7S3</td>
<td>1670.86</td>
</tr>
<tr>
<td>4</td>
<td>P0ACF0</td>
<td>1230.35</td>
</tr>
<tr>
<td>5</td>
<td>P0AES0</td>
<td>896.12</td>
</tr>
<tr>
<td>6</td>
<td>P21165</td>
<td>702.89</td>
</tr>
<tr>
<td>7</td>
<td>P0AG59</td>
<td>524.04</td>
</tr>
<tr>
<td>8</td>
<td>P17952</td>
<td>409.74</td>
</tr>
<tr>
<td>9</td>
<td>P08997</td>
<td>327.85</td>
</tr>
<tr>
<td>10</td>
<td>rev_P76577</td>
<td>276.03</td>
</tr>
<tr>
<td>11</td>
<td>P41407</td>
<td>246.88</td>
</tr>
<tr>
<td>12</td>
<td>P39177</td>
<td>219.44</td>
</tr>
<tr>
<td>13</td>
<td>P37689</td>
<td>195.37</td>
</tr>
<tr>
<td>14</td>
<td>P0A951</td>
<td>177.02</td>
</tr>
<tr>
<td>15</td>
<td>P0AGG4</td>
<td>164.52</td>
</tr>
<tr>
<td>16</td>
<td>P29131</td>
<td>153.92</td>
</tr>
<tr>
<td>17</td>
<td>rev_P0AEQ1</td>
<td>146.86</td>
</tr>
<tr>
<td>18</td>
<td>rev_P09155</td>
<td>140.07</td>
</tr>
<tr>
<td>19</td>
<td>P0A9S5</td>
<td>132.29</td>
</tr>
<tr>
<td>20</td>
<td>P0AE45</td>
<td>125.41</td>
</tr>
<tr>
<td>21</td>
<td>P77718</td>
<td>120.12</td>
</tr>
<tr>
<td>22</td>
<td>P76115</td>
<td>116.15</td>
</tr>
<tr>
<td>23</td>
<td>rev_P76463</td>
<td>111.37</td>
</tr>
<tr>
<td>24</td>
<td>rev_P0A6E4</td>
<td>107.58</td>
</tr>
<tr>
<td>#</td>
<td>Accession</td>
<td>Protein Score</td>
</tr>
<tr>
<td>---</td>
<td>--------------</td>
<td>---------------</td>
</tr>
<tr>
<td>1</td>
<td>P0ABH7</td>
<td>4258.08</td>
</tr>
<tr>
<td>2</td>
<td>P0ABJ9</td>
<td>2423.84</td>
</tr>
<tr>
<td>3</td>
<td>P0A7S3</td>
<td>1670.86</td>
</tr>
<tr>
<td>4</td>
<td>P0ACF0</td>
<td>1230.35</td>
</tr>
<tr>
<td>5</td>
<td>P0AES0</td>
<td>896.12</td>
</tr>
<tr>
<td>6</td>
<td>P21165</td>
<td>702.89</td>
</tr>
<tr>
<td>7</td>
<td>P0AG59</td>
<td>524.04</td>
</tr>
<tr>
<td>8</td>
<td>P17952</td>
<td>409.74</td>
</tr>
<tr>
<td>9</td>
<td>P08997</td>
<td>327.85</td>
</tr>
<tr>
<td>10</td>
<td>rev_P76577</td>
<td>276.03</td>
</tr>
<tr>
<td>11</td>
<td>P41407</td>
<td>246.88</td>
</tr>
<tr>
<td>12</td>
<td>P39177</td>
<td>219.44</td>
</tr>
<tr>
<td>13</td>
<td>P37689</td>
<td>195.37</td>
</tr>
<tr>
<td>14</td>
<td>P0A951</td>
<td>177.02</td>
</tr>
<tr>
<td>15</td>
<td>P0AGG4</td>
<td>164.52</td>
</tr>
<tr>
<td>16</td>
<td>P29131</td>
<td>153.92</td>
</tr>
<tr>
<td>17</td>
<td>rev_P0AEQ1</td>
<td>146.86</td>
</tr>
<tr>
<td>18</td>
<td>rev_P09155</td>
<td>140.87</td>
</tr>
<tr>
<td>19</td>
<td>P0RHS5</td>
<td>132.29</td>
</tr>
<tr>
<td>20</td>
<td>P0A6E4</td>
<td>125.41</td>
</tr>
<tr>
<td>21</td>
<td>P77718</td>
<td>120.12</td>
</tr>
<tr>
<td>22</td>
<td>P76115</td>
<td>116.15</td>
</tr>
<tr>
<td>23</td>
<td>rev_P76463</td>
<td>111.37</td>
</tr>
<tr>
<td>24</td>
<td>rev_P0A6E4</td>
<td>107.56</td>
</tr>
</tbody>
</table>
Protein FDRs only accurate with >100 Proteins

Uncertainty in Protein FDR vs. Number of Confidently IDed Proteins

1% Error In FDR Estimation
<table>
<thead>
<tr>
<th>#</th>
<th>Accession Number</th>
<th>Protein Description</th>
<th>Retention Time (min)</th>
<th>Molecular Weight (kDa)</th>
<th>Protein Group Abundance</th>
<th>Protein Source Abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>75</td>
<td>gi</td>
<td>6322505</td>
<td>Hsp70 family ATPase Ssa1 [Saccharomyces cerevisiae S288c]</td>
<td>71.0</td>
<td>77%</td>
<td>99%</td>
</tr>
<tr>
<td>76</td>
<td>gi</td>
<td>6324768</td>
<td>Toa1p [Saccharomyces cerevisiae S288c]</td>
<td>32.0</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>77</td>
<td>gi</td>
<td>6322697</td>
<td>phosphoglycerate mutase Gpm1 [Saccharomyces cerevisiae S288c]</td>
<td>28.0</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>78</td>
<td>gi</td>
<td>3983653</td>
<td>Rpi43bp [Saccharomyces cerevisiae S288c]</td>
<td>10.0</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>79</td>
<td>gi</td>
<td>6323107</td>
<td>Bos1p [Saccharomyces cerevisiae S288c]</td>
<td>28.0</td>
<td>95%</td>
<td>27%</td>
</tr>
<tr>
<td>80</td>
<td>gi</td>
<td>6320433</td>
<td>Sir4p [Saccharomyces cerevisiae S288c]</td>
<td>15.2</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>81</td>
<td>gi</td>
<td>3304434</td>
<td>Ecm33p [Saccharomyces cerevisiae S288c]</td>
<td>44.0</td>
<td>100%</td>
<td>83%</td>
</tr>
<tr>
<td>82</td>
<td>gi</td>
<td>6321599</td>
<td>Hsp1p [Saccharomyces cerevisiae S288c]</td>
<td>45.0</td>
<td>100%</td>
<td>83%</td>
</tr>
<tr>
<td>83</td>
<td>gi</td>
<td>6324162</td>
<td>Sks1p [Saccharomyces cerevisiae S288c]</td>
<td>70.0</td>
<td>100%</td>
<td>83%</td>
</tr>
<tr>
<td>84</td>
<td>gi</td>
<td>6323004</td>
<td>Hsp70 family chaperone SSA2 [Saccharomyces cerevisiae S288c]</td>
<td>59.0</td>
<td>100%</td>
<td>82%</td>
</tr>
<tr>
<td>85</td>
<td>gi</td>
<td>6323758</td>
<td>hypothetical protein YMIR11c [Saccharomyces cerevisiae S288c]</td>
<td>52.0</td>
<td>100%</td>
<td>82%</td>
</tr>
<tr>
<td>86</td>
<td>gi</td>
<td>8279524</td>
<td>Asp3p [Saccharomyces cerevisiae S288c]</td>
<td>50.0</td>
<td>95%</td>
<td>82%</td>
</tr>
<tr>
<td>87</td>
<td>gi</td>
<td>6322468</td>
<td>Tdh2p [Saccharomyces cerevisiae S288c]</td>
<td>36.0</td>
<td>77%</td>
<td>99%</td>
</tr>
<tr>
<td>88</td>
<td>gi</td>
<td>6323474</td>
<td>ribosomal 40S subunit protein S1A [Saccharomyces cerevisiae S288c]</td>
<td>29.0</td>
<td>77%</td>
<td>98%</td>
</tr>
<tr>
<td>89</td>
<td>gi</td>
<td>3983649</td>
<td>ribosomal 60S subunit protein L148 [Saccharomyces cerevisiae S288c]</td>
<td>39.0</td>
<td>100%</td>
<td>71%</td>
</tr>
<tr>
<td>90</td>
<td>gi</td>
<td>6322153</td>
<td>Csf6p [Saccharomyces cerevisiae S288c]</td>
<td>65.0</td>
<td>100%</td>
<td>67%</td>
</tr>
<tr>
<td>91</td>
<td>gi</td>
<td>6323763</td>
<td>Asc1p [Saccharomyces cerevisiae S288c]</td>
<td>35.0</td>
<td>98%</td>
<td>38%</td>
</tr>
<tr>
<td>92</td>
<td>gi</td>
<td>3983642</td>
<td>Gcn4p [Saccharomyces cerevisiae S288c]</td>
<td>31.0</td>
<td>100%</td>
<td>32%</td>
</tr>
<tr>
<td>93</td>
<td>gi</td>
<td>1251717</td>
<td>Keratin, type II microfibrillar, component 5 [RAP]</td>
<td>55.0</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>94</td>
<td>gi</td>
<td>6321362</td>
<td>ribosomal 60S subunit protein L7A [Saccharomyces cerevisiae S288c]</td>
<td>28.0</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>95</td>
<td>gi</td>
<td>6323888</td>
<td>chaperone ATPase Hsp60 [Saccharomyces cerevisiae S288c]</td>
<td>81.0</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>96</td>
<td>gi</td>
<td>3983645</td>
<td>Hsp1p [Saccharomyces cerevisiae S288c]</td>
<td>87.0</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>97</td>
<td>gi</td>
<td>3983660</td>
<td>Pbp1p [Saccharomyces cerevisiae S288c]</td>
<td>79.0</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>98</td>
<td>gi</td>
<td>6320907</td>
<td>Vh2p [Saccharomyces cerevisiae S288c]</td>
<td>37.0</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>99</td>
<td>gi</td>
<td>6324268</td>
<td>Hsp3p [Saccharomyces cerevisiae S288c]</td>
<td>44.0</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>100</td>
<td>gi</td>
<td>6319673</td>
<td>glucose-6-phosphate isomerase [Saccharomyces cerevisiae S288c]</td>
<td>61.0</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>101</td>
<td>gi</td>
<td>1250911</td>
<td>Keratin, type I microfibrillar, 47.6 kDa [Low-sulfur keratin] [RAP]</td>
<td>46.0</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>102</td>
<td>gi</td>
<td>6322668</td>
<td>ribosomal 60S subunit protein L14A [Saccharomyces cerevisiae S288c]</td>
<td>21.0</td>
<td>99%</td>
<td>99%</td>
</tr>
<tr>
<td>103</td>
<td>gi</td>
<td>547810</td>
<td>Keratin, glycine-tyrosine-rich of hair [RAP]</td>
<td>8.0</td>
<td>98%</td>
<td>98%</td>
</tr>
<tr>
<td>104</td>
<td>gi</td>
<td>3304433</td>
<td>Dcp1p [Saccharomyces cerevisiae S288c]</td>
<td>47.0</td>
<td>90%</td>
<td>90%</td>
</tr>
<tr>
<td>105</td>
<td>gi</td>
<td>6319972</td>
<td>Reverse sequence, was Pol12p [Saccharomyces cerevisiae S288c]</td>
<td>57.0</td>
<td>98%</td>
<td>98%</td>
</tr>
<tr>
<td>106</td>
<td>gi</td>
<td>6323514</td>
<td>Hsp70 family ATPase Ssb1 [Saccharomyces cerevisiae S288c]</td>
<td>36.0</td>
<td>97%</td>
<td>97%</td>
</tr>
</tbody>
</table>
Shared Peptides

Peptides identified:

1. TIGGGDDSFTFFSETGAGK
2. AVFVDLEPTVIDEVR
3. QLFHPEQILITGKEDANNNAR
4. NLDIERPTYTNLNR
5. IHFPLATYAPVISAEK
6. AYHEQLSVAEITNAEFPANQMVK
7. YMACCLLYR
8. SIQFVDWCPFGK
9. VGINYQPPTVVPGDLAK
10. AVCMLSTNAIAMEWAEAR
11. LDHKFDMYAK

Assignment of peptides to proteins:

<table>
<thead>
<tr>
<th>Proteins</th>
<th>Peptides</th>
</tr>
</thead>
<tbody>
<tr>
<td>P05209</td>
<td>alpha-1</td>
</tr>
<tr>
<td>Q13748-1</td>
<td>alpha-2</td>
</tr>
<tr>
<td>Q13748-2</td>
<td>alpha-2</td>
</tr>
<tr>
<td>NP_006000</td>
<td>alpha-3</td>
</tr>
<tr>
<td>P05215</td>
<td>alpha-4</td>
</tr>
<tr>
<td>Q9BQE3</td>
<td>alpha-6</td>
</tr>
<tr>
<td>Q9NY65</td>
<td>alpha-8</td>
</tr>
</tbody>
</table>

Nesvizhskii, A. I.; Aebersold, R.
Shared Peptides

Parsimony

- YMACCLLYR
- SIQFVDWCPTGFK
- Tubulin alpha 6
- Tubulin alpha 3
- Tubulin alpha 4

85% similarity
Peptide / Protein Quantification
<table>
<thead>
<tr>
<th>Outline</th>
<th>Terminology</th>
</tr>
</thead>
<tbody>
<tr>
<td>• MS Quantification</td>
<td>• Absolute Quantification</td>
</tr>
<tr>
<td>• Why variance matters</td>
<td>• Relative Quantification</td>
</tr>
<tr>
<td>• Label-free / normalization</td>
<td>• Bias</td>
</tr>
<tr>
<td>• Labelled</td>
<td>• Variability / Variance</td>
</tr>
<tr>
<td>- iTRAQ</td>
<td>• Label-free</td>
</tr>
<tr>
<td>• Peptides to proteins</td>
<td>• Normalization</td>
</tr>
<tr>
<td></td>
<td>• Labelled</td>
</tr>
<tr>
<td></td>
<td>• iTRAQ</td>
</tr>
<tr>
<td></td>
<td>• Isobaric</td>
</tr>
<tr>
<td></td>
<td>• Reporter ions</td>
</tr>
</tbody>
</table>
Peptide / Protein Quantification

• Absolute – Estimate the molar amount of protein / peptide in the biological sample
  – PTMs
  – Validation

• Relative – Fold change / statistically significant difference between 2 biological states
  – Biological variation
  – Biomarker studies
Label-free

• Area Under Curve
  – MS1
  – Integrate XIC

• Spectral Counting
  – MS2
  – High abundant proteins

MS1 AUC

XIC for m/z = 421.76

\[ y = f(x) \]

\[ \text{Area} = \int_{a}^{b} f(x) \]
Typical Discovery Experiment

Patient Group A  Patient Group B
Prepare Samples

HPLC-MS/MS

Peptides

MS$^1$ & MS$^2$ Spectra

Preprocess

Quantify  Identify

Analyze Statistically

Sources of Bias and Variability

Population

Biological

Instrument

Sample Handling
MS Quantification

• MS not inherently quantitative
• Physiochemical properties invoke different MS responses
• MS only samples a small percentage of total peptides
• Bias and variability
Why Variance Matters

3 Replicates of Analyte 3
High Variance

Mean Fold Change 2.09  
t-test p-value 0.07

3 Replicates of Analyte 3
Low Variance

Mean Fold Change 1.89  
t-test p-value 0.0001
Normalization

• Remove bias and variability *between* runs
• Global – commonly used
  – Median scale
  – Total ion current (TIC)
• Local – very recent development
  – Proximity-based intensity normalization (PIN)
Labeled Quantification

• Run samples simultaneously on in a single run
• Add label to samples
• Mix samples together
• Compute ratios / statistically significant diffs.
Labeled

- **Isobaric**
  - MS2, iTRAQ
  - Number of samples

- **Synthetic Peptides**
  - MS1
  - Absolute (AQUA)

- **Metabolic**
  - MS1, SILAC
  - Not higher life forms

Isobaric Example - iTRAQ

Isobaric tag (Total mass = 145)
Reporter group (Mass = 114-117)

Peptide reactive group
NHS-

Balance group-neutral loss (Mass = 31-28)

Labeled peptide

www.moffitt.org
iTRAQ® 8-Plex Reagent Chemical Structure

Isobaric Tag
Total mass = 305

Reporter Group
113 – 119, 121 m/z

Amine specific peptide reactive group (NHS)
N-hydroxysuccinimide

Balance Group (?)
Mass 184, 186 – 192 m/z

Isobaric Reporter Groups
113 – 119, 121 m/z

Applied Biosystems has granted permission to use this slide.
iTRAQ Experiment

Obtain protein-containing sample, extract protein

Proteolytic Digestion

Label peptides with iTRAQ® Reagents

Reduce, alkylate Cysteines

Reduce, alkylate Cysteines

Reduce, alkylate Cysteines

Reduce, alkylate Cysteines

Trypsin Digest

Trypsin Digest

Trypsin Digest

Trypsin Digest

iTRAQ TAG 114

iTRAQ TAG 115

iTRAQ TAG 116

iTRAQ TAG 117

MIX

2D LC-MS/MS

iTRAQ Experiment MS2 Spectrum

Peptide: VAIVVGAPR

Protein Match: Platelet membrane glycoprotein 11b

"Reporter Ion Mass Tags" from which quantitation is calculated

Peptide match is made from product ions, e.g., b- and y-ion series

MW_{mono} = 1024.62

Peptide match is made from product ions, e.g., b- and y-ion series

Center for Mass Spectrometry and Proteomics | Phone | (612)625-2280 | (612)625-2279

© 2014 Regents of the University of Minnesota. All rights reserved.
iTRAQ Results

• Reporter ion intensities reflect relative peptide amounts

No change disease:control
iTRAQ Results

• What fold changes are significant?

• Do they represent biological relevance as opposed to experimental variability?

Increase disease:control

Decrease disease:control
<table>
<thead>
<tr>
<th>Label-free quantification</th>
<th>Label-based quantification</th>
</tr>
</thead>
<tbody>
<tr>
<td>counting MS/MS spectra</td>
<td>isobaric peptide labeling</td>
</tr>
<tr>
<td>peak integration</td>
<td>(ITRAQ)</td>
</tr>
<tr>
<td>lysis and fractionation</td>
<td>labeled synthetic peptides</td>
</tr>
<tr>
<td>digestion</td>
<td></td>
</tr>
</tbody>
</table>

From Peptides to Proteins

From peptides to proteins involves the following steps:

1. **Protein mix, with \( c_i \) copies of protein \( i \)**
2. **Injected tryptic peptides**
3. **Fractionate by 2D HPLC**
4. **Electrospray ionization**
5. **MS spectrum of peptide mixture**
6. **Repeat for other peaks**
7. **Identify spectrum from database**

**Affect probability** \( p_{ij} \) of observing each peptide \( j \) from protein \( i \)

Proportional after correction by

\[
\text{Fraction contributed by protein } i = \left( c_i \text{ copies } \times O_i \right) / \left( \sum_{k=1}^{\text{# injected proteins}} (c_k \times O_k) \right)
\]

\[
\text{Fraction contributed by protein } i = \left( n_i \text{ MS/MS scans } \times \text{confidence } p_i \right) / \left( \sum_{k=1}^{\text{# observed proteins}} (n_k \times p_k) \right)
\]

Label-free Quantification

Relative Quant in Discovery Proteomics and Metabolomics
Objectives

• Understand how relative quant is estimated.

• Limitations of label-free quant and sources of error.

• Exploratory analyses and Biomarker discovery
Label-Free Quant (LFQ)

- Abundance measured directly from LC-MS peak characteristics.

![Graph showing total ion intensity vs. retention time]

- Peak Area (AUC)
- Peak Height
- Full Width at Half Maximum (FWHM)
- Peak Width
Label-Free Quant (LFQ)

• Estimate of **relative abundance** between samples.

• Compare peak characteristics between samples, but cannot compare between peaks.
  • Difference in ionization efficiency.

• Peaks characteristics must be estimated *per* sample.
Basic Data Processing Workflow

Data Import → Mass Detection → Chromatogram Building → Deconvolution → Alignment → Export
Peak Detection (Deconvolution)

- Extremely complicated task. Very active area of research.

Image Source: https://doi.org/10.1021/acs.analchem.9b02983
Signal-to-Noise Ratio

- Ratio of Signal (S) strength to Noise (N) strength

Signal-to-Noise Ratio

- \( \frac{S}{N} \) ratio = \( \frac{S \text{ (peak height)}}{N \text{ (noise frequency)}} \)
Peak Detection in Discovery Experiments

Image Source: https://doi.org/10.1016/j.aca.2017.09.039
Retention Time Alignment

• Variability in chromatography – RT differences

• Simplified if MS/MS show high similarity.

Image Source: https://doi.org/10.1016/j.aca.2017.09.039
Retention Time Alignment

Linear Retention Time Alignment of File A and File B

- Very active area of research. Many different methods.
Workhorse Algorithms and Software

• Luckily, we don’t need to do this manually!

Proteomics

- ProteomeDiscoverer
- MaxQuant
- Scaffold
- GalaxyP

Metabolomics

- Progenesis QI
- MZmine
- XCMS
- MS-Dial

University of Minnesota
Center for Mass Spectrometry and Proteomics | Phone | (612)625-2280 | (612)625-2279
What Parameters Do You Need to Pay Attention To?

• Dozens of potential parameters to optimize!
  • Use default when parameters when you can...

• The most important parameters to optimize:
  • Mass accuracy
  • Signal-to-noise (S/N)
  • Retention time tolerance
  • Expected peak width

• Evaluate quantitative results *(if possible)*...
The Results

• Obtain a feature matrix of (metabolite or protein / peptide / psm) abundances per file.

• Each unique entry will have a m/z, charge, RT, and identity characteristics (if applicable).

• These signals are utilized for hypothesis testing and inference between experimental groups.
CMSP Statistics

• **CMSP is not a statistics core facility...!**

• Software we utilize will conduct hypothesis testing but it your responsibility to evaluate the results.

• We provide all raw results per experiment, feel free to reprocess and model as you please...
Statistics Primer

• Statistics is the characterization of variance.
  • Without variance we cannot do statistics...

• Research studies sample variance of subset population and we utilize statistics to infer distribution in larger population.
Hypothesis Testing

• Difference in means between population.
  • 2 groups – Student’s T Test
  • >2 groups – Analysis of Variance (ANOVA)

• Null hypothesis \((H_0)\) – No difference in means
• Alternative hypothesis \((H_A)\) – Difference of means

• We can’t just calculate average – because there is variation around the mean.
• Mean alone doesn’t describe the difference between these two values.

• Describe the variation around the mean.
Test Statistics

- Depending on experimental design, characterize variation around mean between pops differently.
  - $t$ test statistic (Student’s T Test)
  - F test statistic (ANOVA)

- Test statistics measure the signal-to-noise of differences between group distributions.
  - $\uparrow$ between group variance, $\downarrow$ within group variance
  - $\downarrow$ between group variance, $\uparrow$ within group variance
• Statistics estimate probability that these differences could happen by chance ($H_0 = \text{True}$)

• We report that confidence as a percentage with a significance threshold.

$t = 32.944 \quad p < 2.2 \times 10^{-16}$
95% true diff (1.755, 1.982)

$t = 3.066 \quad p < 0.0034$
95% true diff (0.818, 3.911)
Confidence and P-values

• P-value is probability that you would receive a similar or more extreme sample by chance (assuming H₀ = True). **P value ~ confidence!!**

• Confidence threshold (95% or 99%) inherently indicate some false discoveries.

• Multiple tests increase chance of false discovery
  • Multiple testing corrections (**Q value**)
Non-significant results

• For false negatives, either the effect size is very small and/or variation is too great.

Option 1: Increase sample size (N)

Option 2: Decrease non-relevant sources of variation
Limitations of LFQ

• Peak detection and integration is not perfect.

• High degree of variance, more missing values.

• Retention time drift may be difficult to correct
  • Misalignments and cross-alignments.

• Greater influence of instrument performance.
Difficulties in Peak Detection

- Labeled quant is estimated directly from MS/MS spectra. LFQ from peak characteristics – fit to idealized model.

- Require robust algorithms to discriminate closely eluting peaks, non-Gaussian peaks, and other non-ideal peaks from background noise.
Difficulties in Peak Detection

![Graph showing retention time vs. base peak intensity]
Difficulties in Peak Detection
Chromatographic Misalignment

• Incorrectly aligning peaks outside retention time window or cross-aligning to non-target peak.

• No chance of misalignment in labeled experiment!

• Must carefully assess retention time reproducibility to optimize alignment per experiment.
Condition A

Condition B
Bad Detection or Alignment, Increased Variance

• Inconsistent Peak Detection results in increased variability.

<table>
<thead>
<tr>
<th>Method</th>
<th>Group A Mean</th>
<th>Group B Mean</th>
<th>Difference</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMT-10plex</td>
<td>15.22</td>
<td>12.87</td>
<td>2.35</td>
<td>(1.69, 3.00)</td>
<td>2.34e-05</td>
</tr>
<tr>
<td>LFQ</td>
<td>15.07</td>
<td>13.77</td>
<td>1.29</td>
<td>(-1.51, 4.10)</td>
<td>0.32</td>
</tr>
</tbody>
</table>
Missing Values and Imputation

• Missing peak detection / quant plague LFQ.

• Imputation – replacement of missing through inference of underlying distribution or properties.
  • Replace 0’s with min or median value.
  • Machine learning – $k$-nearest neighbors, random forest...

• Too many missing values impedes inferential statistics and results quality.
Avoid LFQ? Not necessarily...

- Good experimentation, instrumentation, and data processing parameters avoid most issues...
  - Sacrificing coverage for increased reproducibility

- Quality control samples and/or spike-in to monitor quality and performance.

- More accurate than labeled...
  - Avg estimated difference closer to true difference.
### Quant in your experiment: LFQ vs Labeled

<table>
<thead>
<tr>
<th>Factor</th>
<th>LFQ</th>
<th>Labeled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complexity</td>
<td>Simple</td>
<td>Complex</td>
</tr>
<tr>
<td>Comparability</td>
<td>High similarity</td>
<td>Flexible</td>
</tr>
<tr>
<td>Study Design</td>
<td>Flexible</td>
<td>Limited</td>
</tr>
<tr>
<td>Price</td>
<td>Linear Increase</td>
<td>Fixed</td>
</tr>
</tbody>
</table>
Questions?
Useful Links

Introduction to Mass Spectrometry
University of Minnesota – Biodale Resources: https://cbs.umn.edu/research/resources/biodale
About Mass Spec: https://www.asms.org/about-mass-spectrometry
What is Electrospray? http://newobjective.com/electrospray/

Associations
American Society for Mass Spectrometry https://www.asms.org/
Association of Biomolecular Resource Facilities (ABRF) https://abrf.org/
Minnesota Chromatography Forum http://www.minnchrom.com/
Minnesota Mass Spec Discussion Group (MinnMass) http://minnmass.org/
HUPO https://hupo.org Human Proteome Organization
US HUPO www.ushupo.org

Companies
SCIEX https://sciex.com/
Bruker Daltonics https://www.bruker.com/
Waters http://www.waters.com/
Thermo https://www.thermofisher.com
Cambridge Isotope Laboratories http://www.isotope.com/

Software & Search Tools
Bioinformatics Solutions (PEAKS software) http://www.bioinfor.com/
Proteome Software (Intro. to Proteomics and Scaffold software) https://proteomesoftware.zendesk.com/hc/en-us
Matrix Science (Mascot software) http://www.matrixscience.com/
mMass (free, open source software tool for MS) http://www.mmass.org/
MoIE – Molecular Mass Calculator v2.02  http://mods.rna.albany.edu/massspec/MoIE

UCSF and Protein Prospector  http://prospector.ucsf.edu/prospector/mshome.htm

Sequest home page at Scripps  http://fields.scripps.edu/yates/wp/?page_id=17

Sequest Tips from Novatia  https://www.enovatia.com/ms/ms-resources/sequest-tips/

XCMS Online at Scripps  https://xcmsonline.scripps.edu/landing_page.php?pgcontent=mainPage

**Protein/Mass Spec Information**


DeltaMass (database of protein post translational modifications at ABRF)  https://abrf.org/delta-mass

IonSource (mass spectrometry educational resource)  www.ionsource.com


UniMod (protein modifications for MS – login as guest)  http://www.unimod.org/login.php?message=expired

UniProt Protein Database  https://www.uniprot.org/

Common MS Contaminants (from NC State MSF)  https://chemistry.sciences.ncsu.edu/msf/usefulMSFinfo.html

Protein Pilot Web Tutorial Modules  https://netfiles.umn.edu/users/pjagtap/Proteomics_Webinars_at_MSI/ProteinPilot_Web_Modules.pdf
### IonSource Amino Acid Table

#### Return to IonSource.Com Home

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>AA Codes</th>
<th>Mono.</th>
<th>Amino Acid</th>
<th>AA Codes</th>
<th>Mono.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly</td>
<td>G</td>
<td>57.021464</td>
<td>Asp</td>
<td>D</td>
<td>115.02694</td>
</tr>
<tr>
<td>Ala</td>
<td>A</td>
<td>71.037114</td>
<td>Gln</td>
<td>Q</td>
<td>128.05858</td>
</tr>
<tr>
<td>Ser</td>
<td>S</td>
<td>87.032029</td>
<td>Lys</td>
<td>K</td>
<td>128.09496</td>
</tr>
<tr>
<td>Pro</td>
<td>P</td>
<td>97.052764</td>
<td>Glu</td>
<td>E</td>
<td>129.04259</td>
</tr>
<tr>
<td>Val</td>
<td>V</td>
<td>99.068414</td>
<td>Met</td>
<td>M</td>
<td>131.04048</td>
</tr>
<tr>
<td>Thr</td>
<td>T</td>
<td>101.04768</td>
<td>His</td>
<td>H</td>
<td>137.05891</td>
</tr>
<tr>
<td>Cys</td>
<td>C</td>
<td>103.00919</td>
<td>Phe</td>
<td>F</td>
<td>147.06841</td>
</tr>
<tr>
<td>Leu</td>
<td>L</td>
<td>113.08406</td>
<td>Arg</td>
<td>R</td>
<td>156.10111</td>
</tr>
<tr>
<td>Ile</td>
<td>I</td>
<td>113.08406</td>
<td>CMC</td>
<td></td>
<td>161.01467</td>
</tr>
<tr>
<td>Asn</td>
<td>N</td>
<td>114.04293</td>
<td>Tyr</td>
<td>Y</td>
<td>163.06333</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Trp</td>
<td>W</td>
<td>186.07931</td>
</tr>
</tbody>
</table>

*home* | *disclaimer*

Copyright © 2000-2016 IonSource, LLC All rights reserved.

Last updated: Tuesday, January 19, 2016 02:48:48 PM
On the use of SDS-PAGE gels

For in-gel digestion of proteins, basically any type of SDS-PAGE gel will work (gradient or non-gradient, self-cast or precast, reducing or non-reducing), as long as you use high-grade reagents to prepare them, and fresh buffers to run them. If you use home-made gels, don’t cast and run them on the same day. Leaving a gel overnight will allow polymerization to complete, thus preventing crosslinking of proteins in the gel by residual radicals. This is best done at room temperature (not in the fridge/cold room) by covering the gel in plastic, adding some wet tissue to prevent drying of the gel. Gels can be kept at least a week before running. Please observe the shelf-life of precast gels.

Coomassie-stained gels

Not all coomassie stains are compatible with mass spectrometry. We strongly recommend that you mix your own coomassie stains (see protocol below for the recipe). Always try to only stain your gel for the minimum time to just detect your protein(s) of interest. Remember that staining is only meant to visualize the band. Extended staining will increase the background in subsequent MS analysis, it will NOT increase the amount of protein. Destain the gel thoroughly to clear the background and to enhance visibility of the band. Do NOT heat your gel in the microwave to speed up the staining! This will bake your protein in the gel, we will never get it out again.

We DO NOT accept samples that have been stained with any commercial 'instant' or 'ready-to-use' staining kits. Please NEVER use overhead projector foils for gel scanning as this prevents the gel from being used for MS experiments.

A protocol for colloidal coomassie staining can be downloaded here.

Scanning of gel bands

It is very helpful to have a scan of the gel before processing it for MS analysis. To prevent contamination (see below), first seal the gel in plastic (foil and sealer and scanner are available in the Facility, which you are welcome to use). Only use a dedicated lab-scanner that is not used for other purposes (e.g. in the office). A gel scanner is available in the Facility – feel free to use them. Do NOT use overhead transparencies for scanning – this WILL introduce contaminating polymers that inhibit protease activity.
Cutting of gel bands

We prefer to receive intact gels and do the cutting ourselves (if possible, include a scan appropriately indicating your bands of interest). Alternatively, you can cut bands yourself, but please observe these guidelines (also read the section below on ‘contamination’). After cutting, gel bands can be kept for months at -20°C before MS analysis without adverse effect. For uniformity in sample preparation, it may be a good idea to prepare other bands from the same gel at the same time, even if you are not planning to send them now.

Contamination

Contaminants may be introduced at several steps during sample preparation. Some will go unnoticed, others will totally obscure all proteins. Therefore, minimizing contamination is essential.

Keratins

Keratins are omni-present and notorious contaminating proteins originating from skin, hair, dust, clothes, chemicals, etc. Total elimination is virtually impossible (and not necessary), but over-abundance will result in the repeated identification of keratins, and not your protein of interest. Most contaminating keratins are believed to originate from dust collected on the gel after running it. Therefore, a clean lab environment will certainly benefit the outcome of the MS experiment. If you don’t trust your lab is sufficiently dust-free, you are most welcome to run/cut your gel in our Facility. Alternatively, there are a couple of measures you can take. For starters, wear powder-free nitrile gloves at all times. Latex gloves often contain starch, and thus a lot of protein, so do not use these. Never touch the gel (or the liquid it is floating in) with bare hands. Use clean trays/containers that are only used for the purpose of gel staining (and not for e.g. blocking solutions for Western blots), and leave them closed with a lid during staining/storage. Don’t lean over gels, scratching your head how to interpret the band pattern. If you take the gel out of the container (e.g. for scanning, cutting), thoroughly clean the area with water (no soap!) or 70% ethanol. Always use fresh solutions to prepare and run gels, including sample loading buffer. Old solvents tend to collect dust – literally.

Other contaminating proteins

Other frequently observed proteins are BSA, immunoglobulins, and other serum proteins. These usually originate from cell culture media, and can be reduced by extensive washing of cell pellets.

Polymers

In a way, polymers are more serious contaminants than keratins, since they tend to stick to HPLC columns either ruining them, or at least causing a severe memory effect in subsequent LC-MS runs. On top of this, most polymers ionize more easily than peptides - so that even minute amounts can be deleterious for the experiment. The most frequently observed polymers are various forms of PEG, which are present in some plastics but also in soaps, hand-creams (wear gloves!) and detergents. It is not always easy to trace back the source of a polymer contamination, but we have had issues with some brands of pipet tips (TopLab), and pipet tips that have been autoclaved in-house. Do not use coated (‘low-binding’) eppendorf tubes, and avoid using soft plastics (Parafilm!). Furthermore, NO detergents should be present in samples to be submitted for MS, especially NP-40 and Triton X-100 have a bad reputation.
Reversed Phase Chromatography

In reversed phase chromatography peptides or proteins (or other molecules) partition between a hydrophobic stationary phase (e.g., C18 beads) and a less hydrophobic mobile phase (e.g., water or aqueous buffers). The partitioning of the sample components between the two phases is dependent on their respective solubility characteristics. Less hydrophobic components will spend more time in the mobile phase while more hydrophobic components will spend more time in the stationary phase.

Silica particles covered with chemically-bonded hydrocarbon chains (C2 to C18) form the hydrophobic stationary phase, while the less hydrophobic mobile phase is composed of an aqueous mixture of an organic solvent that surrounds the silica particles.

When a peptide or protein passes through a reversed phase column the partitioning mechanism operates continuously as the hydrophobicity of the aqueous mobile phase is increased by addition of an organic solvent such as acetonitrile.

Less hydrophobic peptides or proteins will always move faster than more hydrophobic ones, since the mobile phase is always less hydrophobic than the stationary phase.

Example: ZipTip® pipette tip

The ZipTip® pipette tip is a 10 uL pipette tip with a 0.6 uL bed typically packed with reversed phase (C18 or C4) chromatography media. It is intended for concentrating and purifying peptide, protein or oligonucleotide samples prior to MALDI or electrospray mass spectrometry.
Mass Definitions

Molecular masses are measured in Daltons (Da), or mass units (u).

One Dalton = 1/12 of the mass of a $^{12}$C atom.

**Monoisotopic mass** = sum of the exact masses of the most abundant isotope of each element present, i.e., $^1$H=1.007825, $^{12}$C=12.000000, $^{16}$O=15.994915, etc. The monoisotopic mass is the most accurately defined molecular mass and is preferred if a measurement of it can be determined.

**Average mass** = sum of the averaged masses (“isotope abundance weighted”) of the constituent elements of a given molecule.

The result is a weighted average over all of the naturally occurring isotopes present in the compound. This is the common chemical molecular weight that is used for stoichiometric calculations (H=1.0080, C=12.011, O=15.994, etc.). The average mass cannot be determined as accurately as the monoisotopic mass because of variations in natural isotopic abundances.

**The mass to charge ratio (m/z)**. A quantity formed by dividing the mass (in u) of an ion by its charge number; unit: Thomson or Th
## Isotopic Abundances of Common Elements*

<table>
<thead>
<tr>
<th>Element</th>
<th>Isotope</th>
<th>Mass</th>
<th>Natural Abundance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>$^1$H</td>
<td>1.0078</td>
<td>99.99%</td>
</tr>
<tr>
<td></td>
<td>$^2$H</td>
<td>2.0141</td>
<td>0.015</td>
</tr>
<tr>
<td>C</td>
<td>$^{12}$C</td>
<td>12</td>
<td>98.89</td>
</tr>
<tr>
<td></td>
<td>$^{13}$C</td>
<td>13.0034</td>
<td>1.11</td>
</tr>
<tr>
<td>N</td>
<td>$^{14}$N</td>
<td>14.0031</td>
<td>99.64</td>
</tr>
<tr>
<td></td>
<td>$^{15}$N</td>
<td>15.0001</td>
<td>0.36</td>
</tr>
<tr>
<td>O</td>
<td>$^{16}$O</td>
<td>15.9949</td>
<td>99.76</td>
</tr>
<tr>
<td></td>
<td>$^{17}$O</td>
<td>16.9991</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>$^{18}$O</td>
<td>17.9992</td>
<td>0.2</td>
</tr>
<tr>
<td>P</td>
<td>$^{31}$P</td>
<td>30.9737</td>
<td>100</td>
</tr>
<tr>
<td>S</td>
<td>$^{32}$S</td>
<td>31.9721</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>$^{33}$S</td>
<td>32.9715</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>$^{34}$S</td>
<td>33.9679</td>
<td>4.22</td>
</tr>
<tr>
<td></td>
<td>$^{36}$S</td>
<td>35.9671</td>
<td>0.02</td>
</tr>
</tbody>
</table>

By coincidence, the most abundant isotope of these common elements has the lowest mass.

*A. Burlingame, UCSF*
In-Gel Trypsin Digest

Center for Mass Spectrometry & Proteomics, University of Minnesota

(Adapted from EMBL Method)

- Wear gloves to minimize contamination from keratins. Do not lean, talk or breathe over your gel when cutting bands. Do not have anyone walk by you when cutting out bands. Keratin is everywhere, so please take these precautions to minimize any exogenous keratin.
- Siliconized tubes are highly recommended for proteolytic digestions. We prefer 1.5mL snap-cap Eppendorf Protein Lo-Bind microfuge tubes.
- A typical gel band is ~2x10 mm (1mm thick). The volumes in this protocol correspond to gel volumes of this size. Please scale accordingly if your gel band/region is larger. If the gel band is smaller, the stated volumes are suitable.

1. Excision of protein bands from polyacrylamide gels.
   a) Excise the band of interest using a sterile scalpel, razor blade or sharpened microspatula. Cut as close to the protein band as possible to reduce the amount of "background" gel. Cut into ~2 x 2 mm cubes and place in 1.5 ml snap-cap microfuge tube. If cutting multiple bands, wash cutting utensil with 1:1 water/ethanol in between each gel band.
   b) If silver stain was used, destain bands with the protocol below; otherwise continue to step 2:
      Silver Destaining
      i. Excise the spot/band and place in very clean water (such as Milli-Q or nanopure water) in a siliconized microcentrifuge tube. Discard water.
      ii. De-stain (remove silver stain) gel piece with 1:1 solution of 30 mM potassium ferricyanide and 100 mM sodium thiosulphate made FRESHLY before use. Use 300 – 500 uL, incubate for 8 minutes, and discard.
      iii. Add 100 uL of the destaining solution if necessary, incubate for 1 minute and discard.
      iv. Wash 4 x with 1 mL of Mill-Q water, 8 minutes per wash

2. Washing of gel pieces
   a) Transfer 75 ul 1:1 100mM ammonium bicarbonate:acetonitrile to gel pieces, vortex briefly and incubate 15 min at room temperature. Remove solution, discard to waste and repeat this step for a total of two washes. If the band is heavily Coomassie stained, an extra wash is recommended.
   b) Remove previous wash and add 75 ul of 100% acetonitrile. After pieces shrink and turn white and semi-opaque, remove acetonitrile (~30 sec – 1 min).
   c) Remove acetonitrile and proceed to step 3.

3. Reduction & Alkylation - *note that DTT and iodoacetamide solutions should be made fresh for each digest.*
   a) Rehydrate gel pieces with 75 ul 10 mM DTT in 50 mM ammonium bicarbonate. Incubate 1 hr at 56 °C in water bath. After incubation, briefly spin down tubes then remove DTT solution.
   b) Add 75 ul of 55 mM iodoacetamide in 50 mM NH4HCO3 and incubate 30 min at room temperature in the dark (iodoacetamide is light sensitive). Remove iodoacetamide solution.
   c) Wash gel plugs with 75 ul 1:1 acetonitrile:100 mM ammonium bicarbonate as in step 2. Repeat for a total of 2 washes. All the Coomassie should be removed at this time. If residual Coomassie remains, repeat wash with the acetonitrile:100 ammonium bicarbonate solution.
   d) Remove previous wash and add 75 ul of 100% acetonitrile. After pieces shrink and turn opaque white, remove acetonitrile (~30 sec-1 min).
4. In-Gel Digest
   a) Gel pieces should have the opaque white look from the last wash with 100% acetonitrile.
   b) Rehydrate gel pieces in digestion buffer at 4 °C (50 mM NH₄HCO₃, 5 mM CaCl₂, 5 ng/ul trypsin). Add a sufficient volume of buffer to cover gel pieces, ~20 ul, inspect visually. Add more if necessary.
   c) Set on ice for 15 min. Remove supernatant and replace with 70 ul 50 mM NH₄HCO₃, 5 mM CaCl₂. Incubate at 37 °C overnight in a warm-air incubator.

5. Extraction of Peptides
   a) Briefly spin down tubes in a centrifuge at low speed and recover supernatant from overnight incubation. Place supernatant in new 1.5 ml tube labeled accordingly.
   b) Add sufficient volume of 50% acetonitrile, 0.3% formic acid to cover gel pieces (approximately 60 ul), incubate 15 min.
   c) Recover supernatant and place in corresponding tube.
   d) Add the same volume used in step (b) of 80% acetonitrile, 0.3% formic acid, incubate 15 min. Recover supernatant and place in corresponding tube.
   e) Freeze pooled extracts in -80 °C freezer 30 minutes, then dry in speed vac.
   f) Proceed to zip tip/stage tip protocol (for desalting) or store at -80 °C prior to submission for mass spec analysis.
Peptide Mass Fingerprint Database Search Parameters

- 1 or 0 Enzyme missed cleave sites (try each setting)
- Database NCBI
- Taxonomy (select appropriate, based on protein/sample origin)
- Mass values MH+
- Peptide mass error tolerance 150 parts per million (ppm) (or, try 200 ppm alternatively)
- Monoisotopic
- Fixed modifications: carboxymethyl cysteine or carbamidomethyl cysteine (based on protein source)
- Variable modification: oxidized methionine

<table>
<thead>
<tr>
<th>Protein Digest Sample</th>
<th>Species</th>
<th>Taxonomy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acid glycoprotein</td>
<td>Human</td>
<td>Mammalian</td>
</tr>
<tr>
<td>Alcohol dehydrogenase</td>
<td>Yeast</td>
<td>Yeast or Saccharomyces cerevisiae</td>
</tr>
<tr>
<td>Aldolase</td>
<td>Rabbit</td>
<td>Mammalian</td>
</tr>
<tr>
<td>Apotransferrin</td>
<td>Rat</td>
<td>Mammalian</td>
</tr>
<tr>
<td>Bovine serum albumin</td>
<td>Bovine</td>
<td>Mammalian</td>
</tr>
<tr>
<td>Cytochrome C</td>
<td>Horse</td>
<td>Mammalian</td>
</tr>
<tr>
<td>Enolase</td>
<td>Yeast</td>
<td>Yeast or Saccharomyces cerevisiae</td>
</tr>
<tr>
<td>Myoglobin</td>
<td>Horse</td>
<td>Mammalian</td>
</tr>
</tbody>
</table>
# MASCOT Peptide Mass Fingerprint

<table>
<thead>
<tr>
<th><strong>Your name</strong></th>
<th>LeeAnn</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Email</strong></td>
<td><a href="mailto:higgi022@umn.edu">higgi022@umn.edu</a></td>
</tr>
<tr>
<td><strong>Search title</strong></td>
<td>Peptide Mass Fingerprint</td>
</tr>
<tr>
<td><strong>Database(s)</strong></td>
<td>Vertebrates_EST contaminants crAP <strong>NCBIprot</strong> SwissProt</td>
</tr>
<tr>
<td><strong>Enzyme</strong></td>
<td>Trypsin</td>
</tr>
<tr>
<td><strong>Allow up to</strong></td>
<td>1 missed cleavages</td>
</tr>
<tr>
<td><strong>Taxonomy</strong></td>
<td>Mammalia (mammals)</td>
</tr>
<tr>
<td><strong>Fixed modifications</strong></td>
<td>Carbamidomethyl (C)</td>
</tr>
<tr>
<td><strong>Variable modifications</strong></td>
<td>Oxidation (M)</td>
</tr>
<tr>
<td><strong>Protein mass</strong></td>
<td>kDa</td>
</tr>
<tr>
<td><strong>Peptide tol. ±</strong></td>
<td>150 ppm</td>
</tr>
<tr>
<td><strong>Mass values</strong></td>
<td>MH⁺, Mr, M-H⁻</td>
</tr>
<tr>
<td><strong>Monoisotopic</strong></td>
<td>Average</td>
</tr>
<tr>
<td><strong>Data file</strong></td>
<td>No file selected.</td>
</tr>
<tr>
<td><strong>Query</strong></td>
<td>807.442, 821.450, 1286.720, 1287.718</td>
</tr>
<tr>
<td><strong>Data input</strong></td>
<td>Browse...</td>
</tr>
</tbody>
</table>