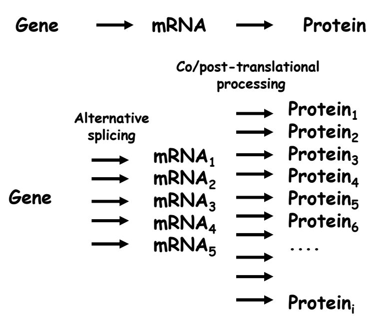
Applications of Mass Spectrometry to Proteomics

Raghothama Chaerkady, Ph.D.

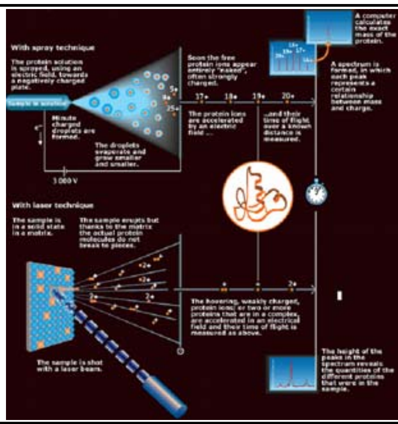
McKusick-Nathans Institute of Genetic Medicine and the Department of Biological Chemistry

Why Proteomics?

One Gene, Many Proteins



Why Mass Spectrometry?



http://nobelprize.org

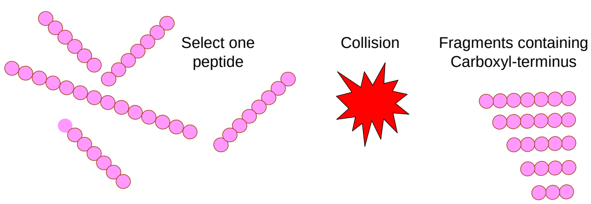
Protein Characterization by Mass Spectrometry

- Very sensitive: Less than 1 pmol of protein is required for identification (50 ng of a 50 kDa protein)
- Mixtures of proteins (hundreds to thousands of proteins) can be analyzed
- Proteins with a blocked N terminus can be identified
- Post-translational modifications such as phosphorylation can be identified and localized
- Quantitative mass spectrometry has numerous applications

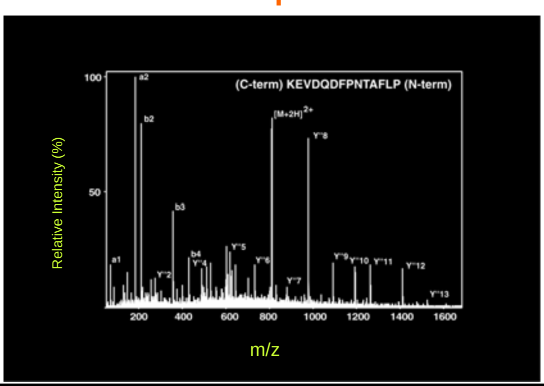
Electrospray Tandem Mass Spectrometry (MS/MS)

- MS1 Scan - Detection of all the peptides in a mixture
- MS/MS - When one of the peptides is isolated and subjected to collision induced dissociation for sequencing
- Tandem mass spectrometry provides the actual sequence of the peptide

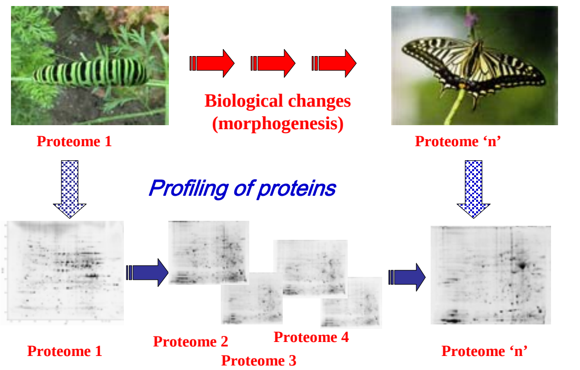
Peptide Sequencing by MS/MS



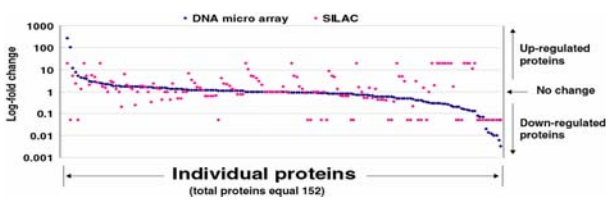
MS/MS Spectrum



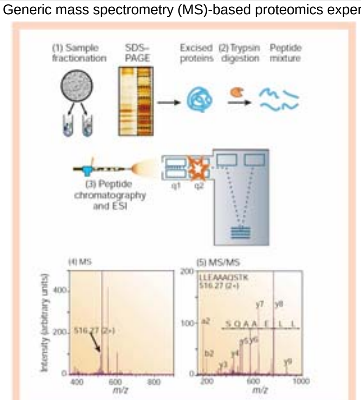
Genomics and Proteomics



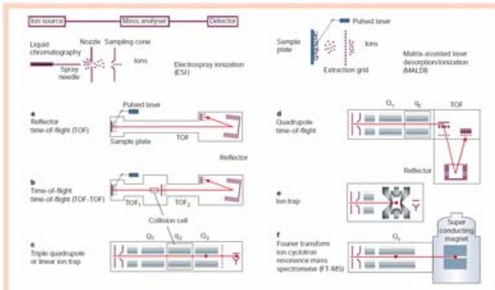
mRNA and Protein Correlation



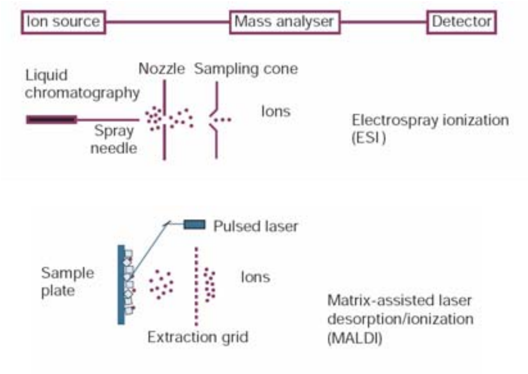
Generic mass spectrometry (MS)-based proteomics experiment.

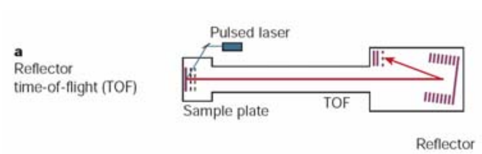


insight
Nature 422, 198-207

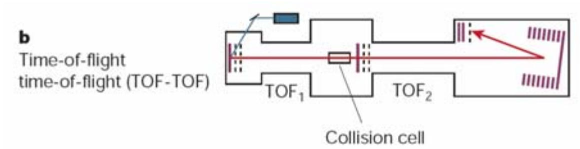


Mass spectrometers used in proteome research.

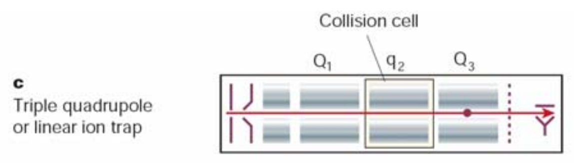




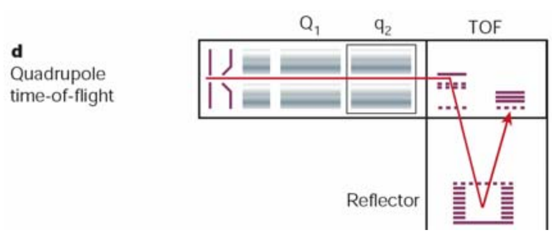
In reflector time-of-flight (TOF) instruments, the ions are accelerated to high kinetic energy and are separated along a flight tube as a result of their different velocities. The ions are turned around in a reflector, which compensates for slight differences in kinetic energy, and then impinge on a detector that amplifies and counts arriving ions.



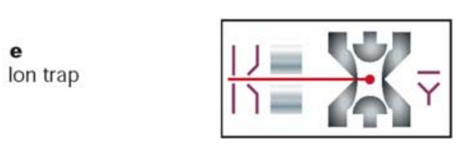
b, The TOF-TOF instrument incorporates a collision cell between two TOF sections. Ions of one mass-to-charge (m/ z) ratio are selected in the first TOF section, fragmented in the collision cell, and the masses of the fragments are separated in the second TOF section



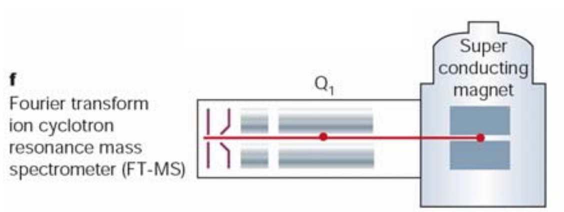
Quadrupole mass spectrometers select by time-varying electric fields between four rods, which permit a stable trajectory only for ions of a particular desired m/z. Again, ions of a particular m/z are selected in a first section (Q1), fragmented in a collision cell (q2), and the fragments separated in Q3. In the linear ion trap, ions are captured in a quadruple section, depicted by the red dot in Q3. They are then excited via resonant electric field and the fragments are scanned out, creating the tandem mass spectrum



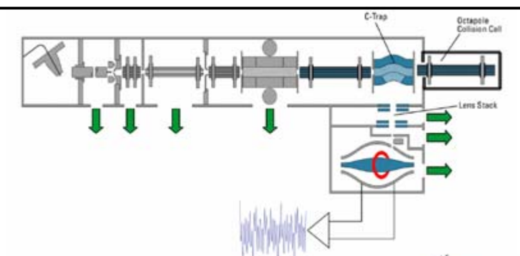
The quadrupole TOF instrument combines the front part of a triple quadruple instrument with a reflector TOF section for measuring the mass of the ions.



The (three-dimensional) ion trap captures the ions as in the case of the linear ion trap, fragments ions of a particular m/ z, and then scans out the fragments to generate the tandem mass spectrum.



The FT-MS instrument also traps the ions, but does so with the help of strong magnetic fields. The figure shows the combination of FT-MS with the linear ion trap for efficient isolation, fragmentation and fragment detection in the FT-MS section.



During ion detection, both the central electrode and deflector are maintained at very stable voltages so that no mass drift can take place. The outer electrode is split in half at z=0, allowing the ion image current in the axial direction to be collected. The image current on each of half of the outer electrode is differentially amplified and then undergoes analog-to-digital conversion before processing using the fast Fourier transform algorithm.

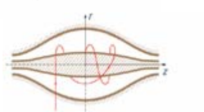
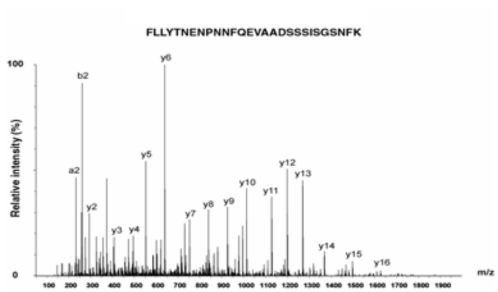
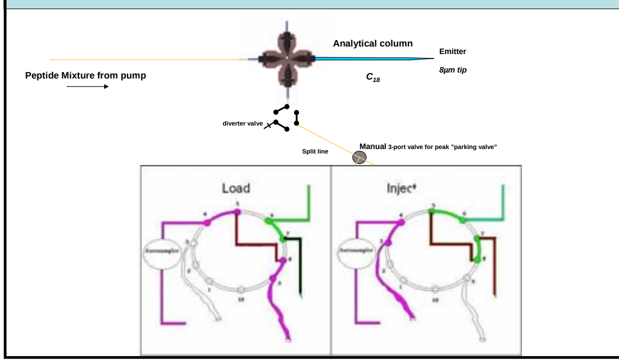


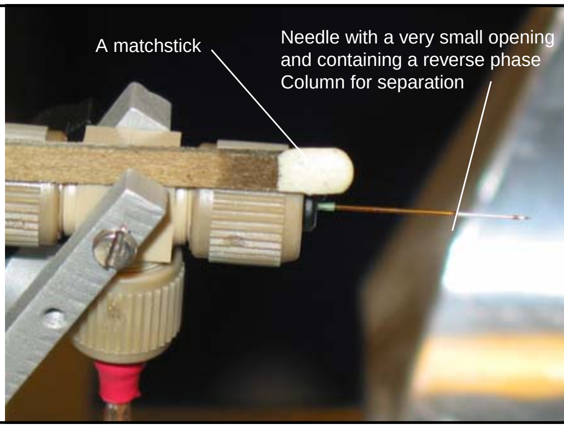
Figure 1-10. Schematic view of the orbitrap cell and example of a stable ion trajectory

MS/MS spectrum (sequencing)

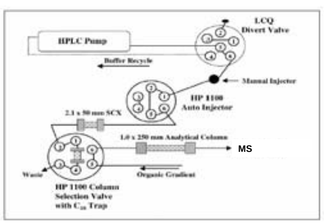


1D Liquid Chromatography Setup



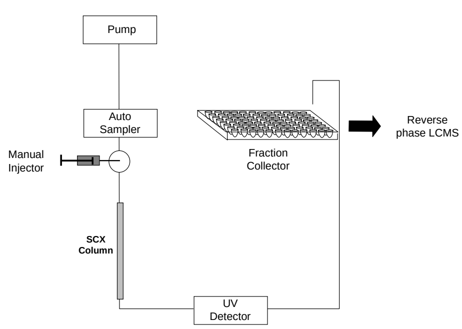


Multidimensional Liquid Chromatography (MudPIT) Setup

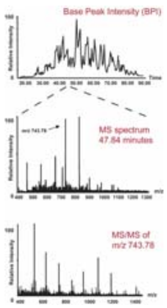


Liu, H. et al. (2002)

Offline Fractionation in the First Dimension

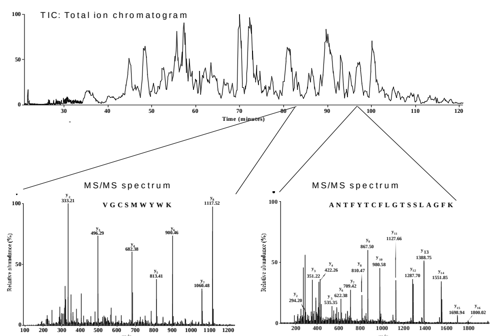


Automated nanoLC-MS/MS



- Purification of sample
- Very complex samples can be analyzed
- Identical peptides will elute in one peak (enhanced signal)
- Automated
- Many proteins can be identified in one run (100-2,000 proteins)

Coupling Liquid Chromatography to Tandem MS



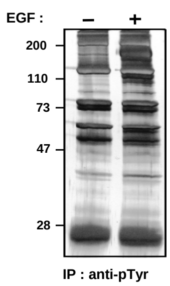
Quantitative Proteomics

- Based on difference in intensity
- By relative quantitation using MS

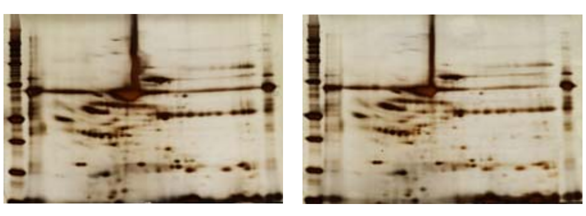
Quantitative Proteomics

- Based on difference in intensity
  - 1D gels
  - 2D gels
  - Use of fluorescence-based methods

1D Gel-based Comparison



2D Gel-based Comparison



2D Gels - Limitations

- Sample preparation - lot of optimization required
- Loading capacity limited
- Do not resolve very small (<10 kD) or large (>100 kD) proteins
- Do not resolve hydrophobic (e.g. membrane) proteins
- Issues with reproducibility

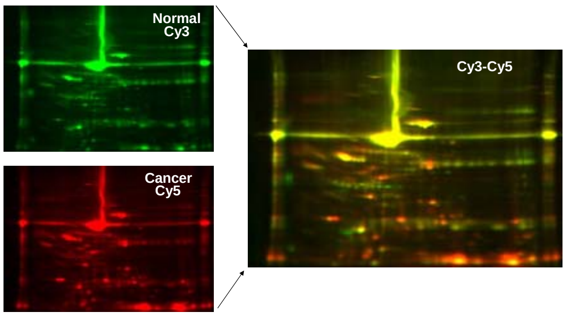
Quantitative Proteomics

- Fluorescence-based quantitation
  - DIGE (Difference in-gel electrophoresis)

DIGE

- Samples to be labeled are labeled with Cy3 (green) and Cy5 (red)
- Samples are ‘mixed’ and resolved by 2D gels
- Fluorescence measured and quantitated

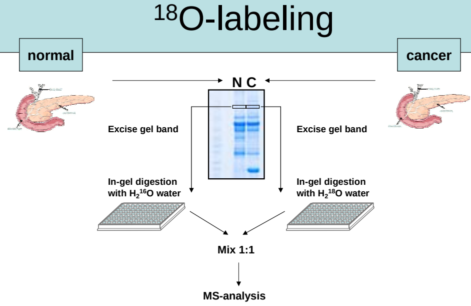
2D-DIGE of Pancreatic Juice



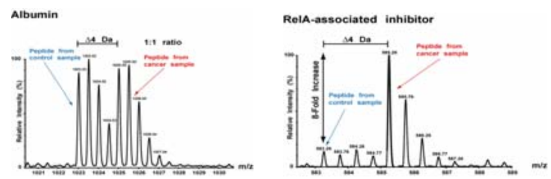
Quantitative Proteomics

- By relative quantitation using MS
  - *in vitro* labeling
    - ^{18}O-labeling
    - Peptide mass tagging (ICAT)
  - *in vivo* labeling
    - Labeling with stable isotope containing amino acids (SILAC)

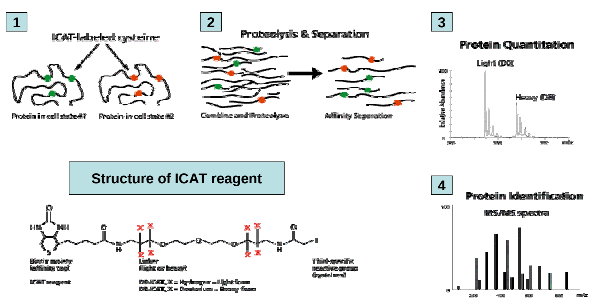
^{18}O-labeling



^{18}O-labeling



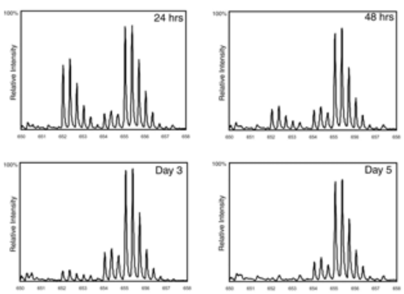
Isotope-Coded Affinity Tag (ICAT)



Stable Isotope Labeling in Cell Culture (SILAC) for Protein Quantitation

- Mammalian cell culture models are used for studying a number of biological processes
- In the SILAC approach, cells are grown continuously in media containing one or more stable isotopes (e.g. ^{13}C). All the proteins in the cells are heavier and can be used to 'mark' a given state in mass spectrometric analysis

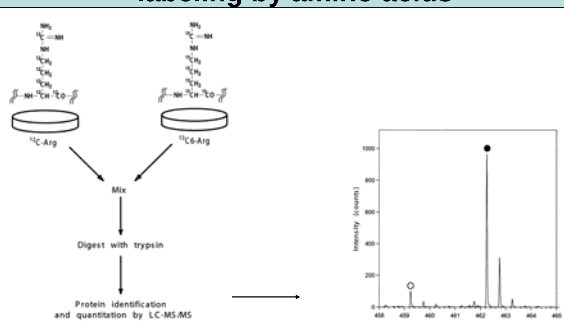
Time Course of Heavy Amino Acid Incorporation



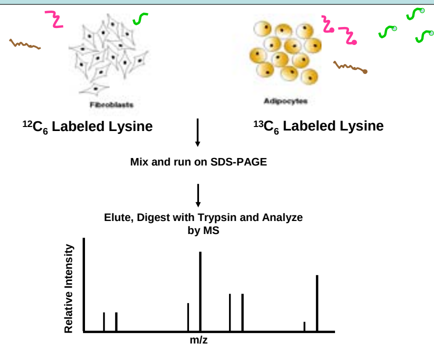
Advantages of the SILAC Method

- Simple
- *In vivo*
- Does not require any extra processing steps
- All proteins are uniformly labeled
- Complete and predictable incorporation
- Choice of labeled amino acids
- De novo sequencing can be performed

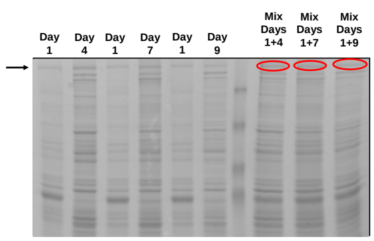
General strategy for stable isotope labeling by amino acids



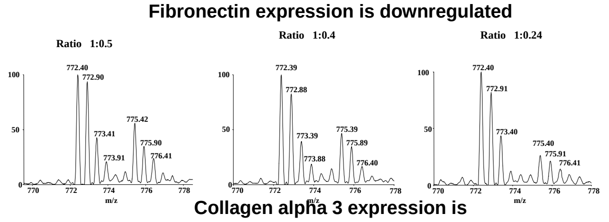
SILAC for Quantitation of Secreted Proteins



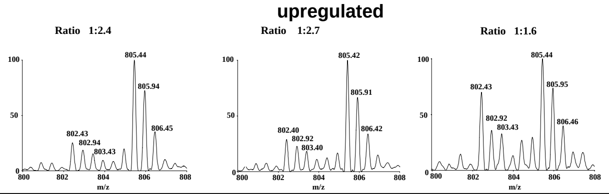
Profile of Proteins Secreted by Adipocytes



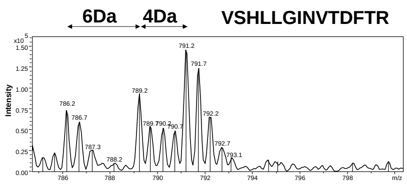
Fibronectin expression is downregulated



Collagen alpha 3 expression is upregulated



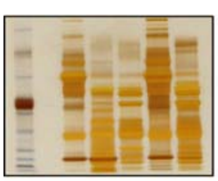
Studying Dynamics Using SILAC



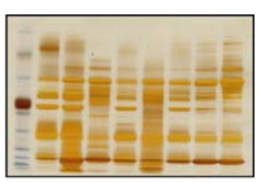
Biomarker Discovery Using Proteomics

- Ideal targets for biomarker
  - Protein (differential expressed proteins)
  - DNA (mutations, methylation)
  - RNA (differential expressed genes)
- Biological specimen
  - Tissue (whole tissue or isolated tumor cells)
  - Pancreatic juice
  - Serum
  - Plasma

Pancreatic juice

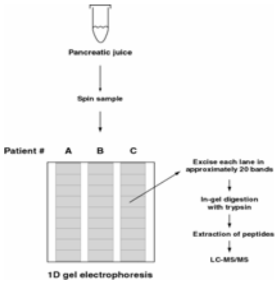


Chronic Pancreatitis



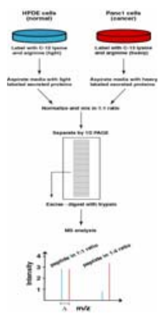
Pancreatic Cancer

Physiological proteome of human pancreatic juice



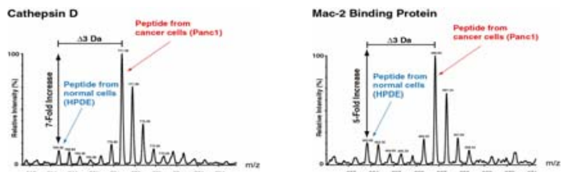
- Collection of pancreatic juice (cancer) during surgery
- Run on 1D gel
- LC-MS/MS analysis
- Bioinformatic analysis of identified proteins
- Compare data with known microarray data

Differential Proteomic Analysis of Pancreatic Cancer Secretome



- *In-vivo* labeling with both arginine and lysine
- Collect conditioned media and concentrate with centricon 3,000 Da MWCO
- Normalize and mix in 1:1 ratio
- Resolve proteins by 1D gel electrophoresis
- Excise bands and digest proteins by trypsin
- Identify proteins by nanoLC-MS/MS (2x30 bands)
- Verify identified proteins (manually)
- Relative quantitation of ID proteins (manually)

Quantitation of Secreted Proteins



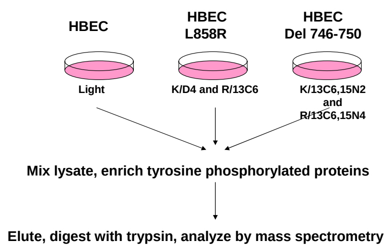
Post-translational Modifications

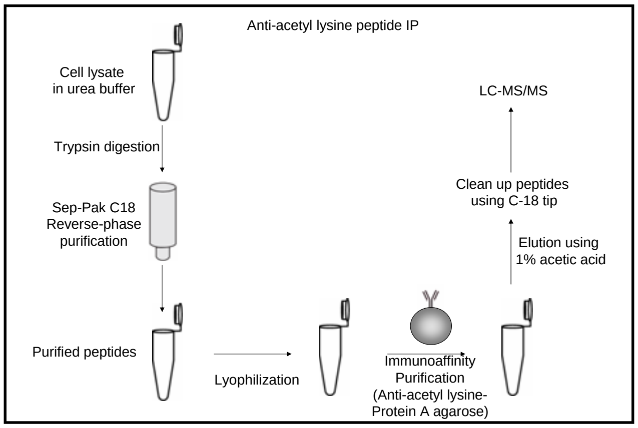
- Peptides can have a number of modifications
- During database searching, a variable modification has to be specified – otherwise, no ‘hit’
- Common PTMs are phosphorylation, acetylation, ubiquitination, glycosylation etc.

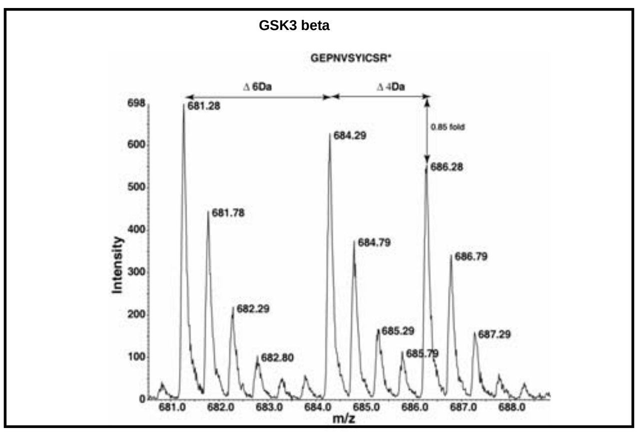
Protein Phosphorylation

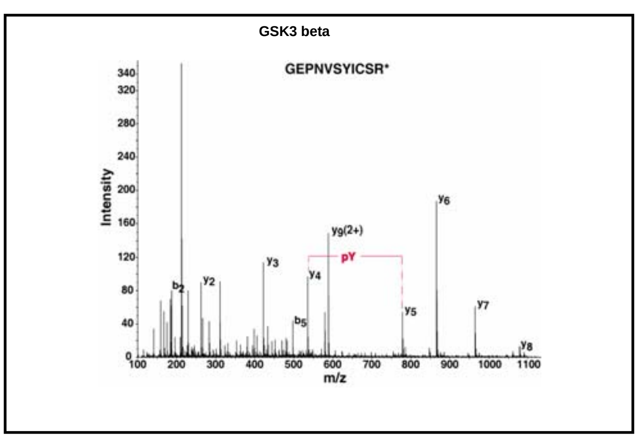
- One-third of all cellular proteins are phosphorylated at one time or another
- Phosphoamino acid content of a vertebrate cell: Serine - 90%; Threonine - 10%; Tyrosine - 0.05%
- Ser:Thr:Tyr - 1800:200:1
- Tyrosine phosphorylation is tightly regulated

Identifying Activated Tyrosine Kinases in Lung cancer




Anti-acetyl lysine peptide IP
Cell lysate in urea buffer
Trypsin digestion
Sep-Pak C18 Reverse-phase purification
Purified peptides
Lyophilization
Immunoaffinity Purification (Anti-acetyl lysine-Protein A agarose)
Elution using 1% acetic acid
Clean up peptides using C-18 tip
LC-MS/MS

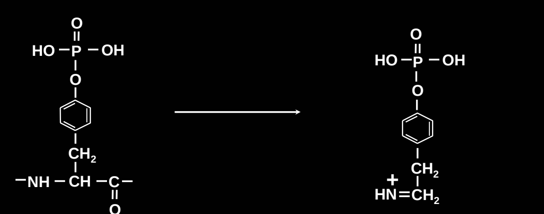

GSK3 beta
GEPNVSYICSR*
Δ 6Da
Δ 4Da
0.85 fold
681.28
681.78
682.29
682.80
684.29
684.79
685.29
685.79
686.28
686.79
687.29
Intensity
m/z


GSK3 beta
GEPNVSYICSR*
Intensity
m/z
pY

Why is Phosphorylation Analysis Difficult?

- Stoichiometry of phosphorylation is low
- Complete coverage of proteins is difficult to obtain
- Phosphorylated serine and threonine residues are labile whereas phosphotyrosines are more stable
- Phosphoserines and phosphothreonine residues can be subjected to a beta-elimination reaction but not phosphotyrosine residues
- Antibodies to enrich for serine and threonine phosphorylated proteins are not available
- Phosphopeptides are ‘suppressed’ in a mass spectrum

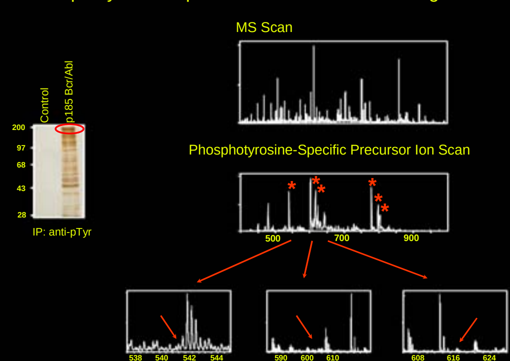
Immonium Ion of Phosphotyrosine as a Reporter Ion



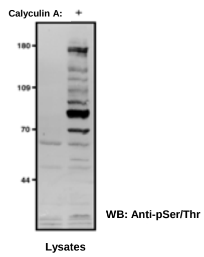
Phosphotyrosine

Immonium Ion (216.043 Da)

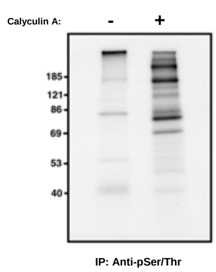
Phosphotyrosine-Specific Precursor Ion Scanning – Bcr/Abl



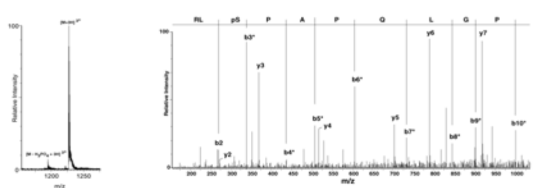
Large scale IP with an anti-pSer/Thr antibody



In Vivo Labeling with ^{32}P

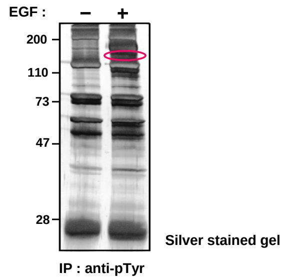


Identification of Phosphorylated Ser/Thr residues



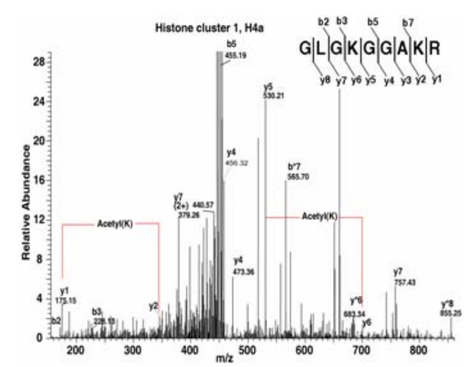
^{10}RLpSPAPQLGP^{19}

MS-Based Identification of a 130 kDa Protein in the EGF Receptor Signaling Pathway



Partial list of acetylated lysine containing peptides identified in an IP experiment

Protein name	Peptide	Score
Histone cluster 1, H4a	GLGKGGAKR	40
Immunoglobulin superfamily, member 9	GKPEVVSVVGR	30
Hypothetical protein [gi|89059164]	KGKPQVQDKVVK	54
RBP-binding zinc finger protein, c	KNQLVQK	43
Putative homeodomain transcription factor 1	LKKVENIK	41
H2A histone family, member Z	TKAVSR	40
Block of proliferation 1	QELTKKLMPNCK	36



Assignment of the initiator methionine in a cDNA 'fragment' based on an N-terminal peptide

>KIAA0229 (1180 residues) FRAGMENT

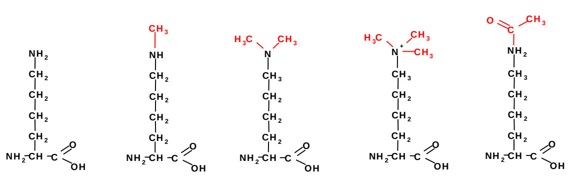
SWGKGREGVVSPAGLGGALPGDGKFGSPSRLGCSLGEGVQRVAALGMGKEQ
ELLRAARTGHLPAVEKLLSGKRLSSGFGGGGGGGSGGGGGGSGGGGGGLGS
SSHPLSSLLSMWRGPNVNCVDSTGYTPLHHAALNGHHRRSSSSRSQDSAEGQ
DGQVPEQFSGLLHGSSPVCEVGQDPFQLLCTAGQSHPDGSPQQGACHKASM
QLEETGVHAPGASQPSALDQSKRVGYLTGLPTTNSRSHPETLTHTASPHPGGA
EEGDRSGAR

Assignment of the initiator methionine in a cDNA 'fragment' based on an N-terminal peptide

>KIAA0229 (1180 residues) FRAGMENT

SWGKGREGVVSPAGLGGALPGDGKFGSPSRLGCSLGEGVQRVAALG**MGKEQ**
LLRAARTGHLPAVEKLLSGKRLSSGFGGGGGGGSGGGGGGSGGGGGGLGSS
SHPLSSLLSMWRGPNVNCVDSTGYTPLHHAALNGHHRRSSSSRSQDSAEGQD
GQVPEQFSGLLHGSSPVCEVGQDPFQLLCTAGQSHPDGSPQQGACHKASMQL
EETGVHAPGASQPSALDQSKRVGYLTGLPTTNSRSHPETLTHTASPHPGGAEE
GDRSGAR

Lysine Modifications



	Lysine	Mono-methylated lysine	Di-methylated lysine	Tri-methylated lysine	Acetylated lysine
Mass (Da)	128.095	142.111	156.127	170.143	170.105
Mass gain	-	14.016 Da	28.032 Da	42.048 Da	42.010 Da

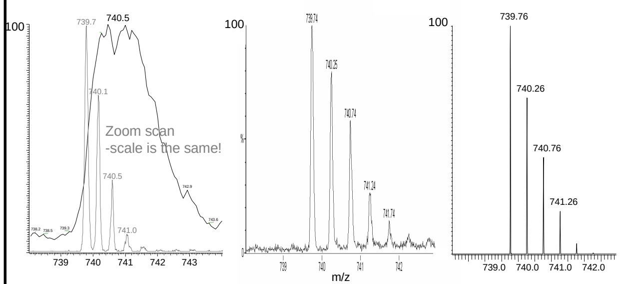
Acetylation vs Tri-methylation: a numbers game…

Tri-methylation: 42.046950 Da [H(6) C(3)]
Acetylation: 42.010565 Da [H(2) C(2) O]

Mass difference: 0.036385 Da

Resolution of various mass spectrometers

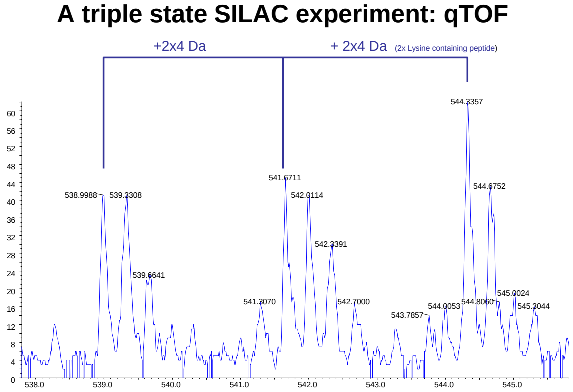
Normal ion trap | Quadrupole-Time-of-Flight | Orbitrap



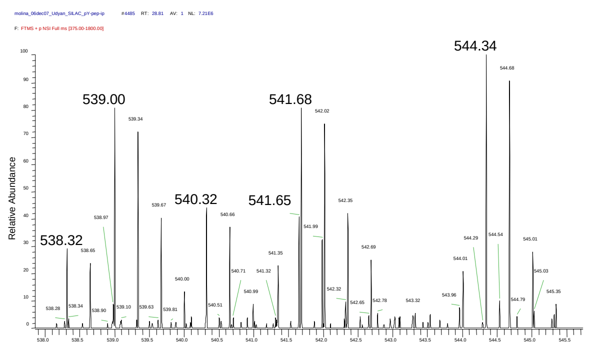
<3,000 | 8,000-14,000 | 30,000-60,000

Approx. resolution in ppm

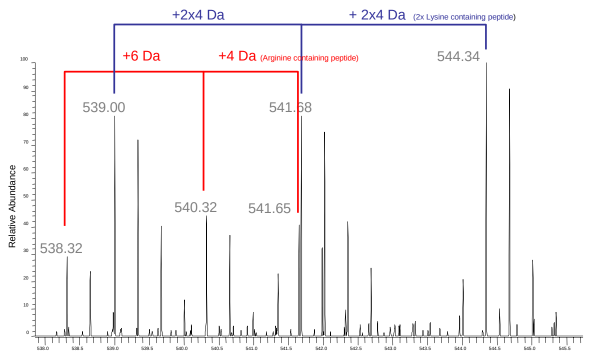
A triple state SILAC experiment: qTOF



A triple state SILAC experiment: Orbitrap



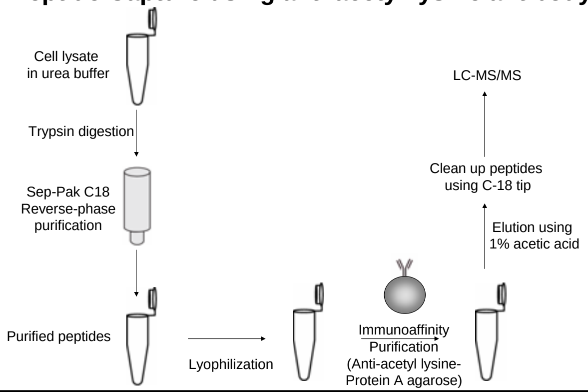
A triple state SILAC experiment: Orbitrap



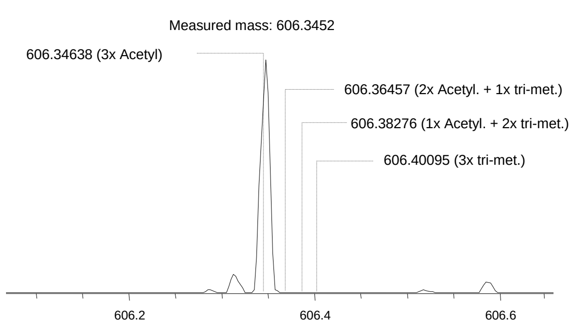
Identifying acetylation sites using Orbitrap

- Digest protein mixture into peptides
- Capture acetylated lysine containing peptides using anti-acetyl lysine antibodies
- Analyze on Orbitrap
- If the samples are labeled for quantitative proteomics, one can also obtain data on differential regulation of acetylation at individual lysine residues in this strategy

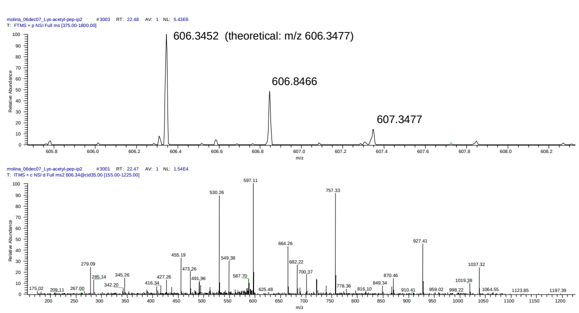
Peptide Capture using anti-acetyl lysine antibody



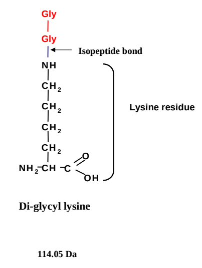
GGKGLGKGGAKR peptide from Histone Cluster 1, H4a: What are the PTMs?



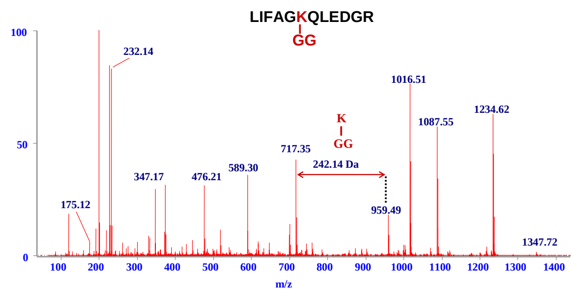
Answer: Acetylation at each of 3 lysine residues: GGK*GLGK*GGAK*R



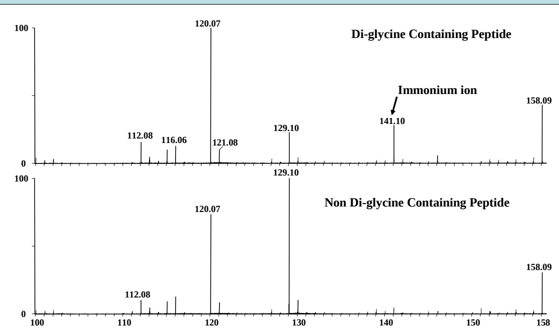
Lysine Modifications



Tryptic Digest of Ubiquitylated Peptide



Signature of a Ubiquitylated Peptide



Genome Annotation

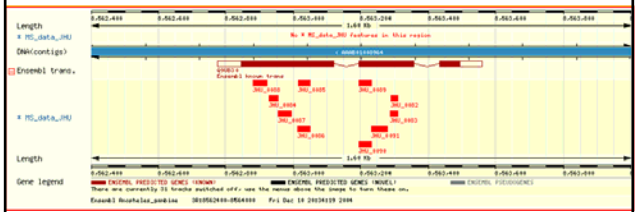
Genome Annotation by Mass Spectrometry: What Can We Gain?

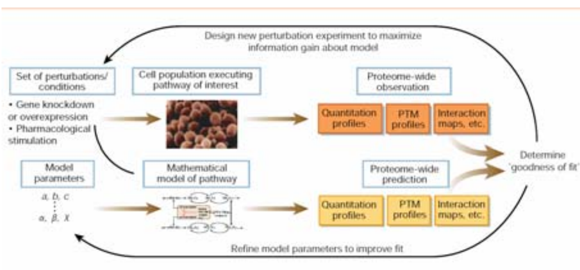
- Assigning start codons
- Proteins isoforms (alternative splicing, novel exons)
- Novel genes (proteins less than 100 amino acids not predicted by programs)
- cSNPs
- Correction of incorrect gene predictions (~50% of the genes in human are predicted)
- Validation of gene predictions

Transcripts and Proteins in *Anopheles gambiae*

- Proteins ~ 700 known
- Annotated genes: 15,189 (11,757 predicted)
  7,840 unique to predictions by Celera
  1,375 unique to predictions by Ensembl
  5,974 common to both predictions
- Gomez et al. (Genome Biology: 6:R39, 2005) 35,000 full-length enriched cDNAs
  3,700 genes of which only 650 are novel
- Krisventseva et al. (Genome Research, 2005)
  215,634 ESTs/cDNAs
  7,961 clusters of which 3,100 are novel

Use of Ensembl Distributed Annotation System to Validate a Known Transcript





Schematic representation of the systems biology paradigm involving proteomics.