# METHODS IN PROTEOMICS

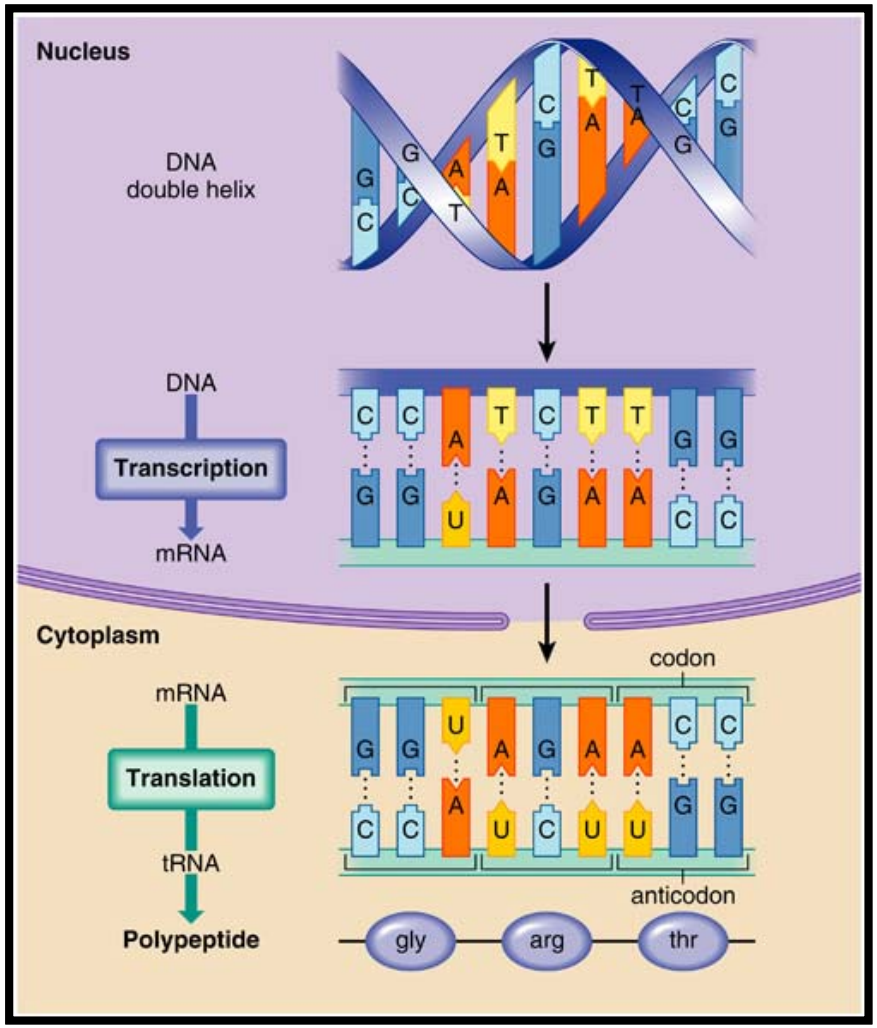
Lecturer: Ming-Hsien Tsai, PHD.

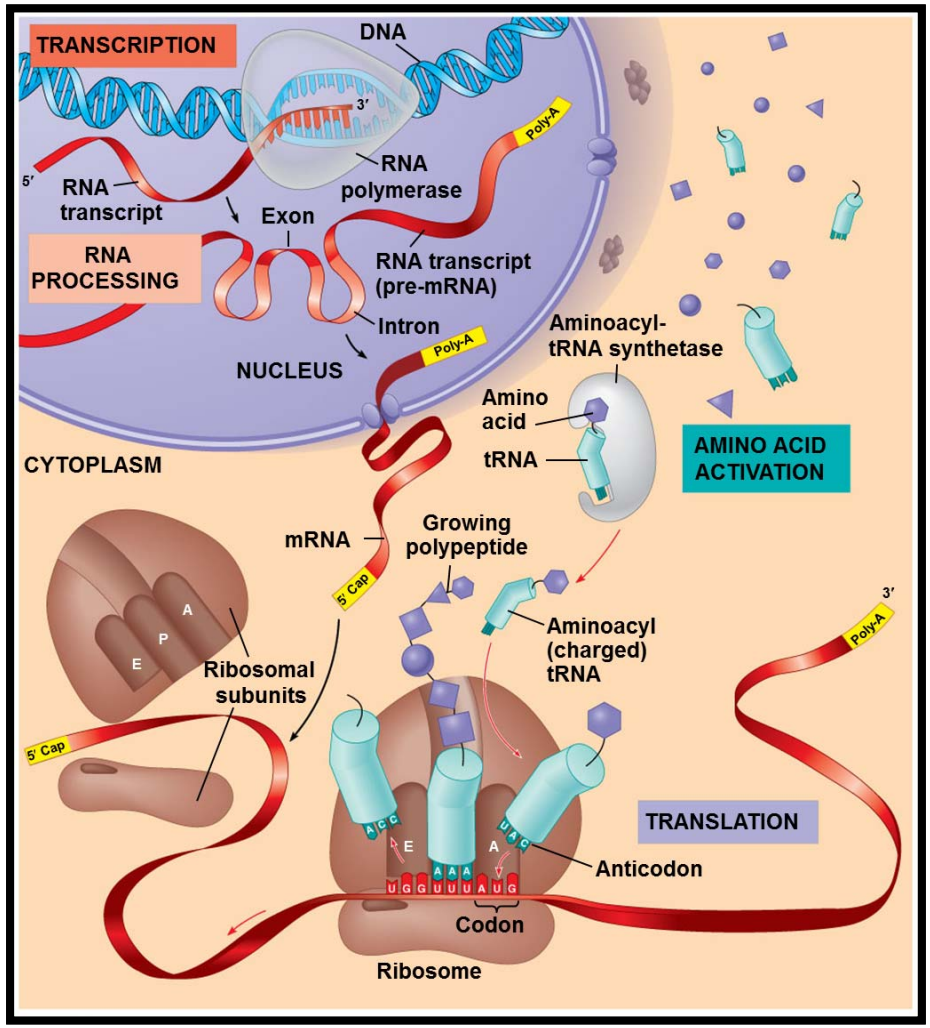
Assistant researcher

Center for Lipid bioscience, KMUH

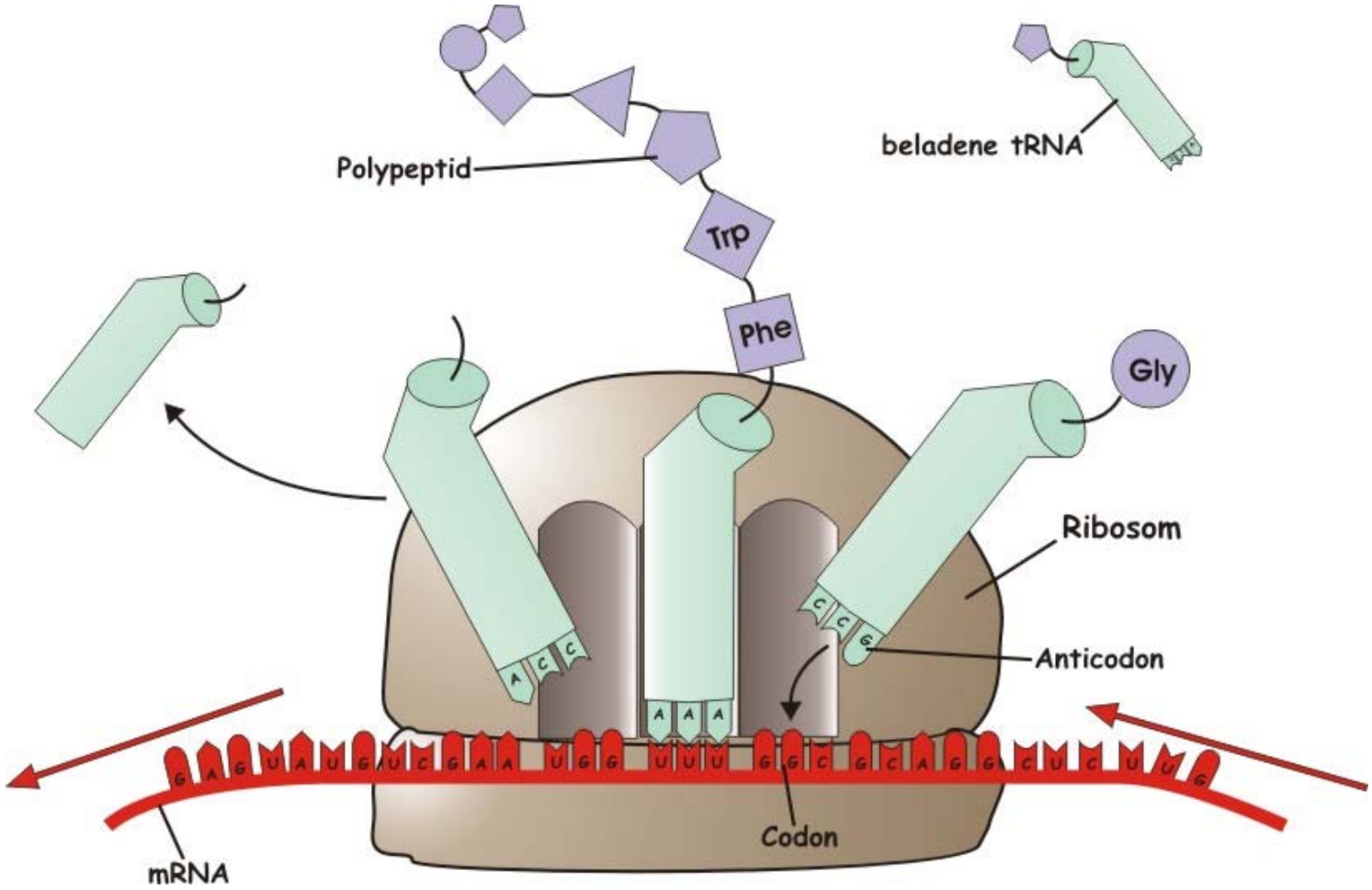
Lipid Science and Aging Research Center, KMU

# DNA → RNA → Protein



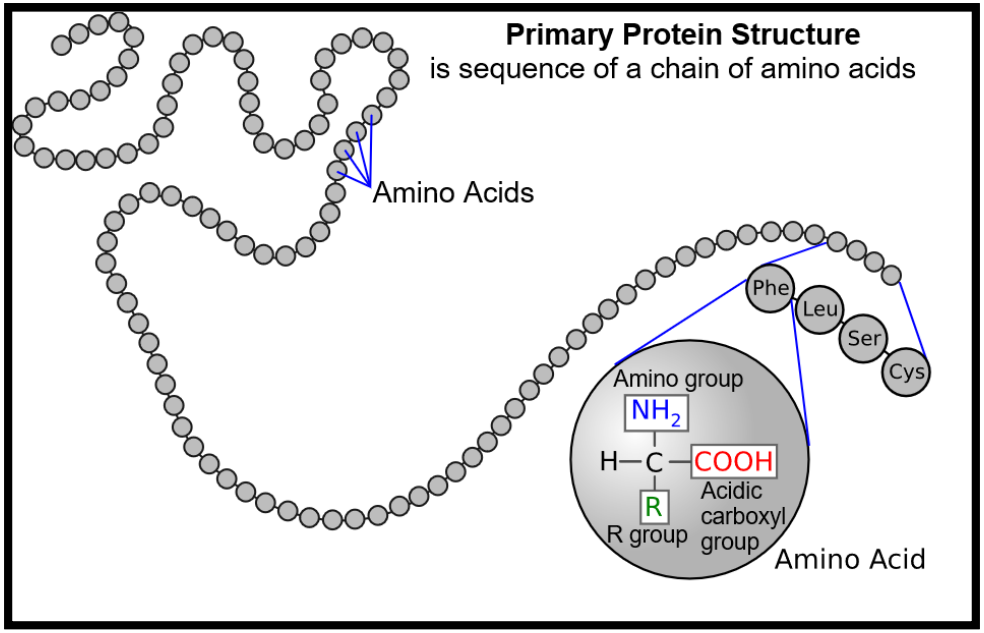



# Translation

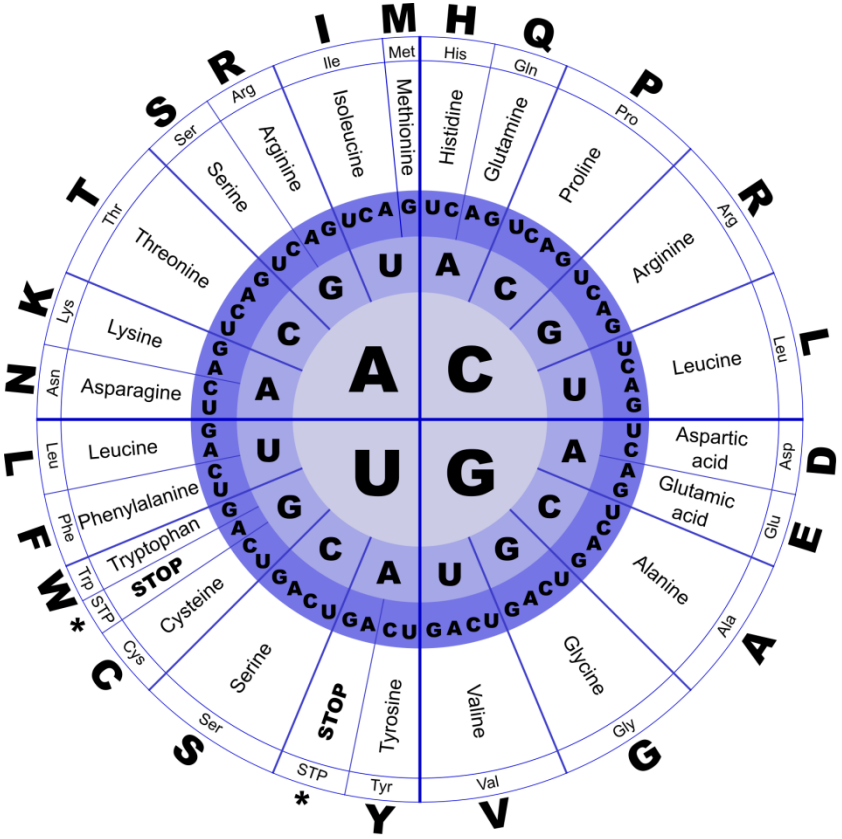


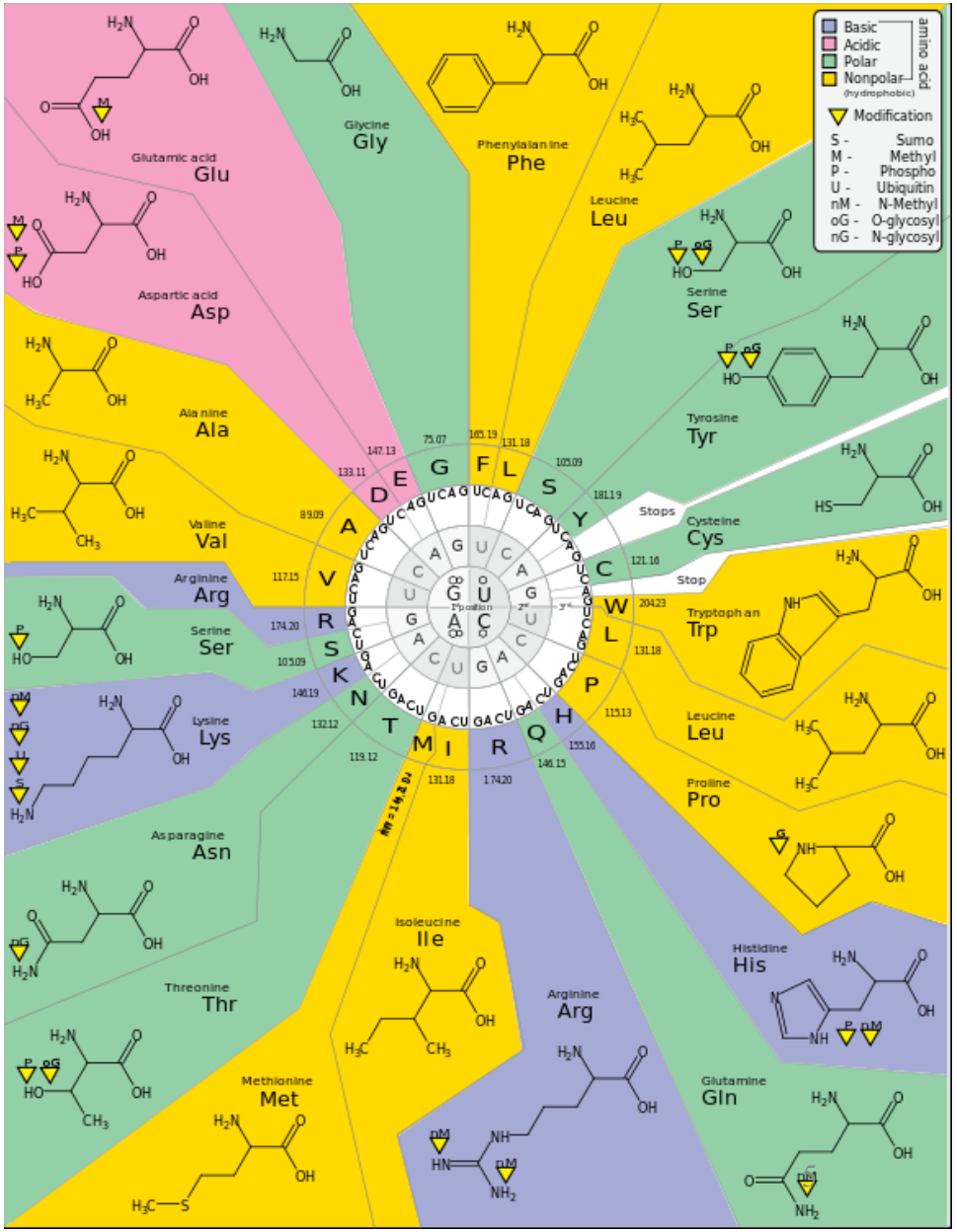
# How does DNA act as A code?

- The order of bases on the DNA strand instructs the ribosomes how to synthesize proteins
- Gene: portion of DNA that codes for the production of a specific polypeptide
- Polypeptide: building block of a protein



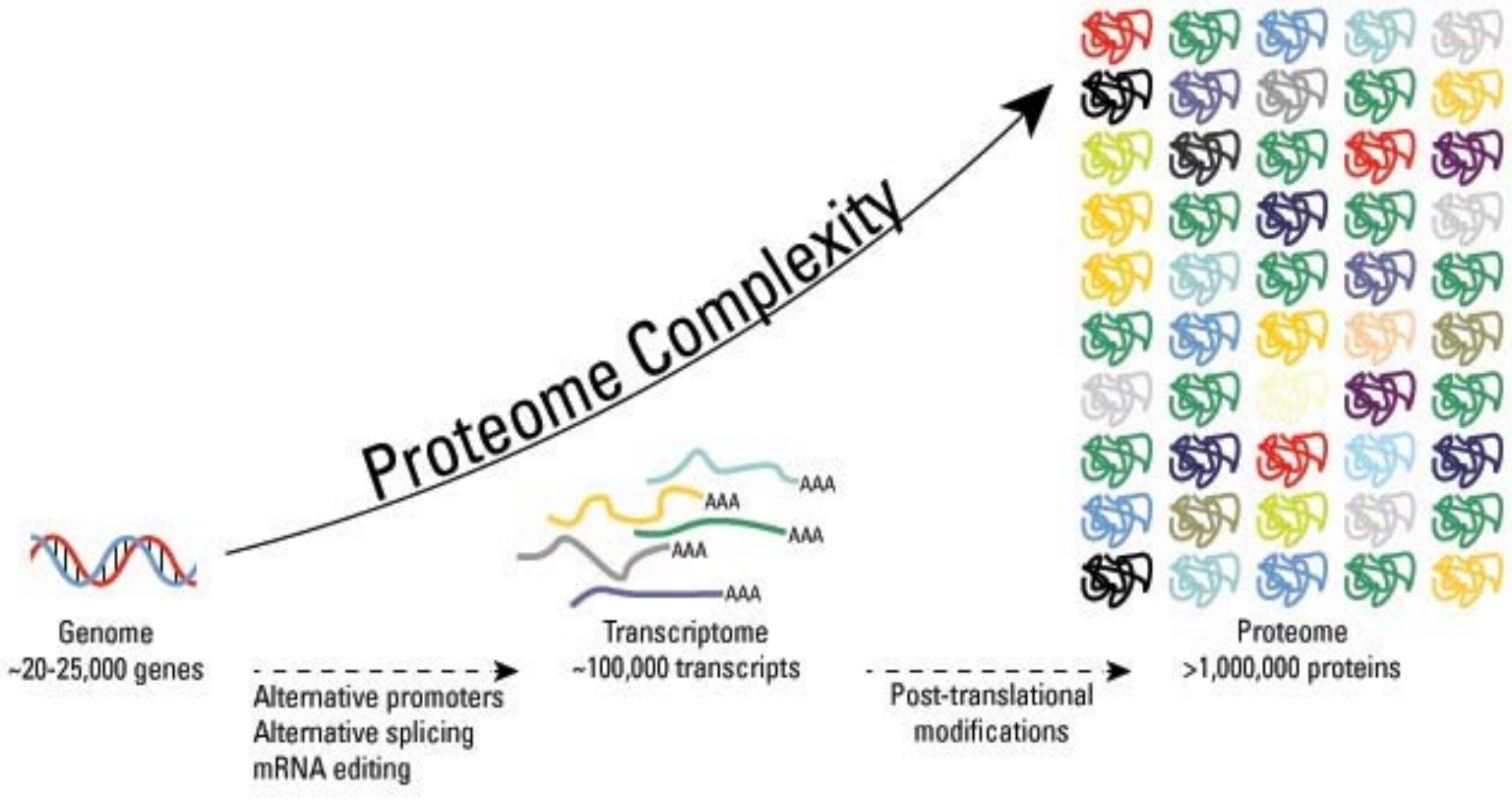
# Amino Acid codon



# But in Fact…

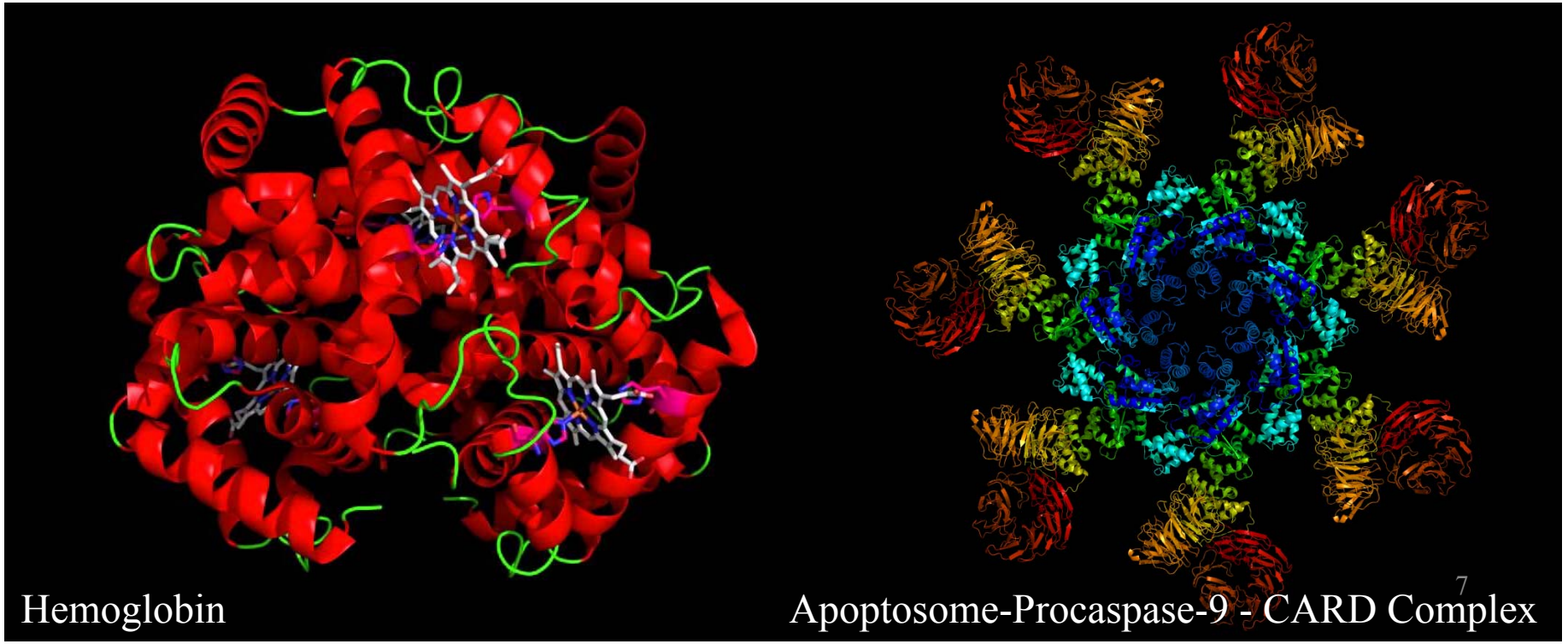
- Genome ~ 20,000-25,000 protein encoding genes
- Human proteins ≥ 1 million




Zimmermann J and Brown LR. (2001)

# Proteomics and the proteome

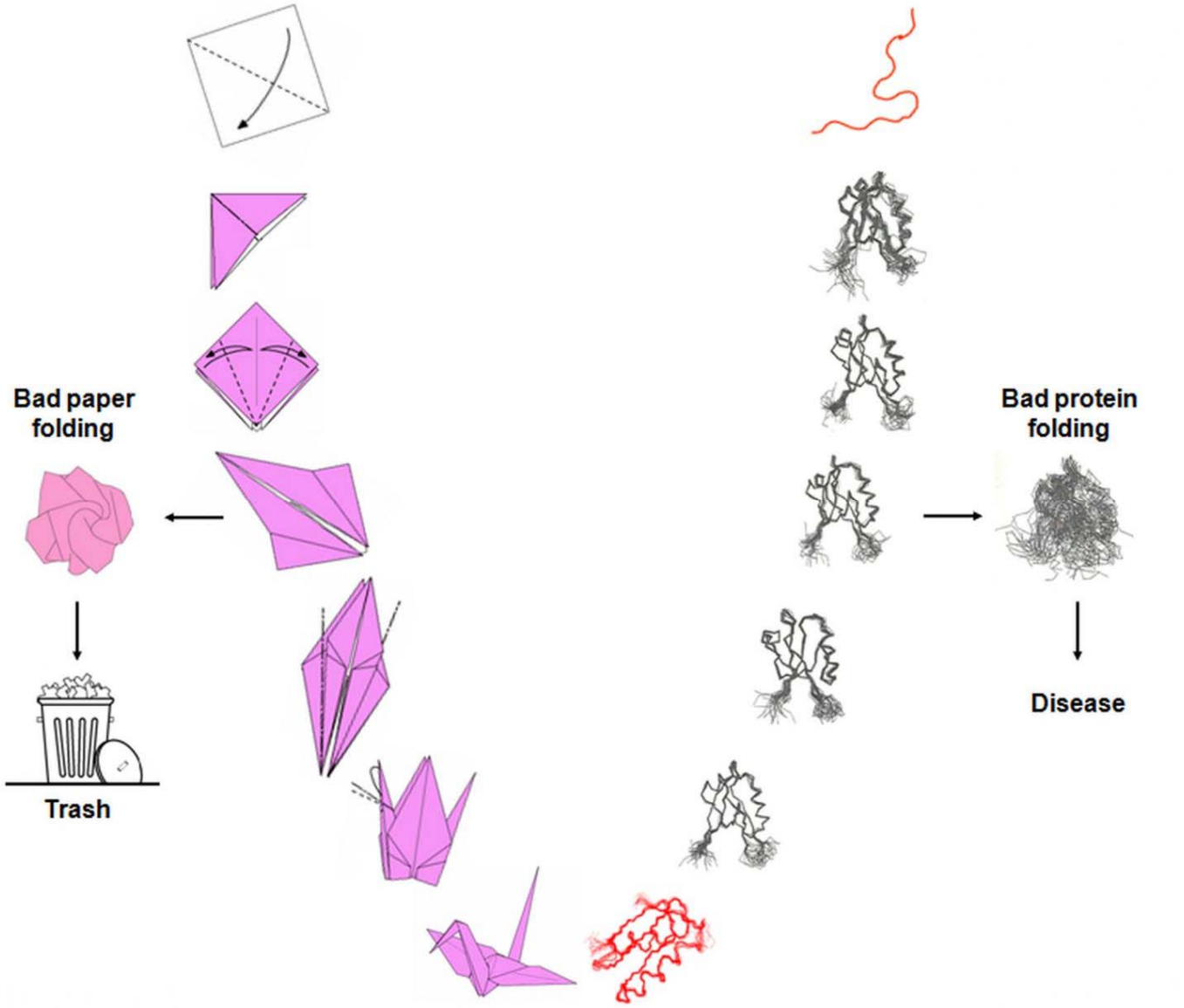
- Proteomics is the study of the proteome, the full protein complement of organisms e.g. plasma, cells and tissue.
- Understanding the proteome allows to characterisation of proteins, understanding protein interactions and identification of disease biomarkers.
- Unlike related fields like genomics, proteomics allows for the study of post-translational modifications and interactions.



Hemoglobin

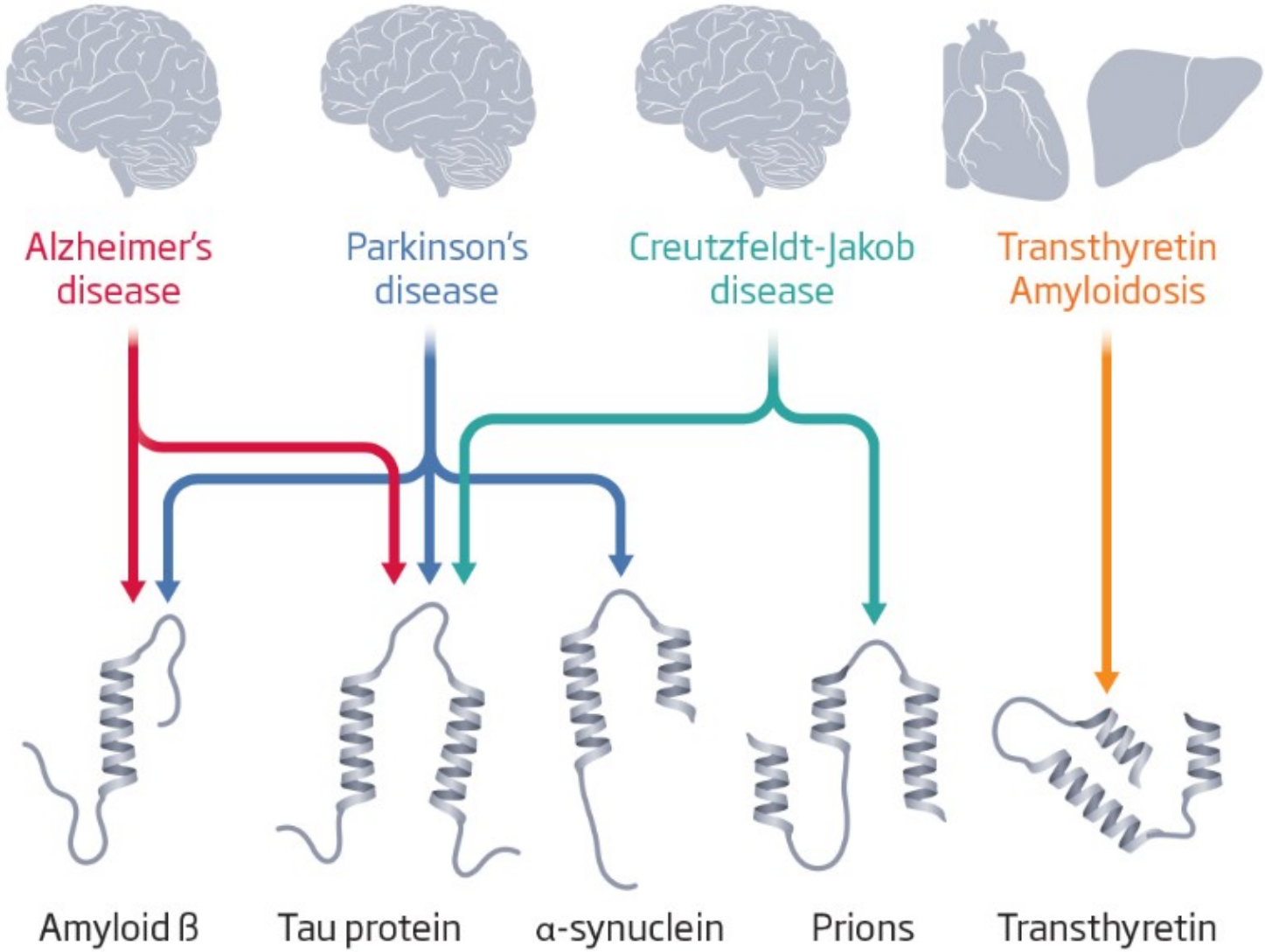
Apoptosome-Procaspase-9 - CARD Complex

# Bad protein folding and Disease

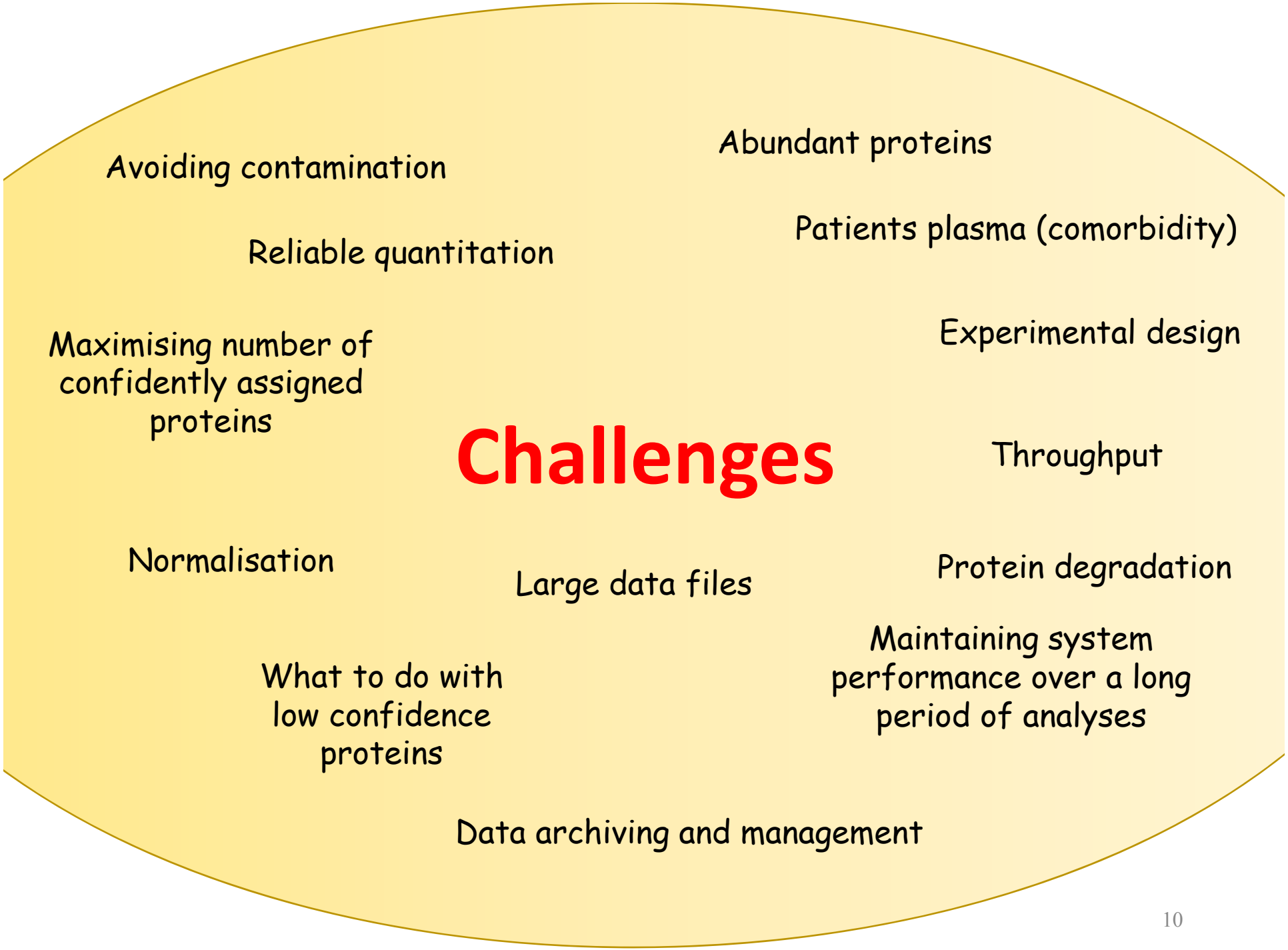


# Proteins gone rogue

Multiple kinds of misfolded plaques are implicated in different diseases, but they share a common feature




SOURCE: NEUROPHAGE PHARMACEUTICALS


Avoiding contamination
Abundant proteins
Patients plasma (comorbidity)
Reliable quantitation
Maximising number of confidently assigned proteins
Experimental design
Challenges
Throughput
Normalisation
Large data files
Protein degradation
Maintaining system performance over a long period of analyses
What to do with low confidence proteins
Data archiving and management

# Post-Translational Modifications

- Phosphorylation
- Glycosylation
- Ubiquitination
- Methylation
- SUMOylation
- ⋮

- smaller chemical groups
  - ➢ Acylation
  - ➢ Alkylation
  - ➢ Amidation
  - ➢ Hydroxylation
  - ➢ N-Acetylation
  - ➢ S-Nitrosylation
  - ➢ S-glutathionylation
  - ⋮

# Common PTMs by frequency

- In 2011, statistics of each post-translational modification experimentally and putatively detected have been compiled using proteome-wide information from the Swiss-Prot database

| Frequency | Modification |
|---|---|
| 58383 | Phosphorylation |
| 6751 | Acetylation |
| 5526 | N-linked glycosylation |
| 2844 | Amidation |
| 1619 | Hydroxylation |
| 1523 | Methylation |
| 1133 | O-linked glycosylation |
| 878 | Ubiquitylation |
| 826 | Pyrrolidone Carboxylic Acid |
| 504 | Sulfation |

# Protein Phosphorylation

Phosphorylation is the addition of a phosphoryl group (PO3)− to a molecule. In biology, phosphorylation and its counterpart, dephosphorylation, are critical for many cellular processes.

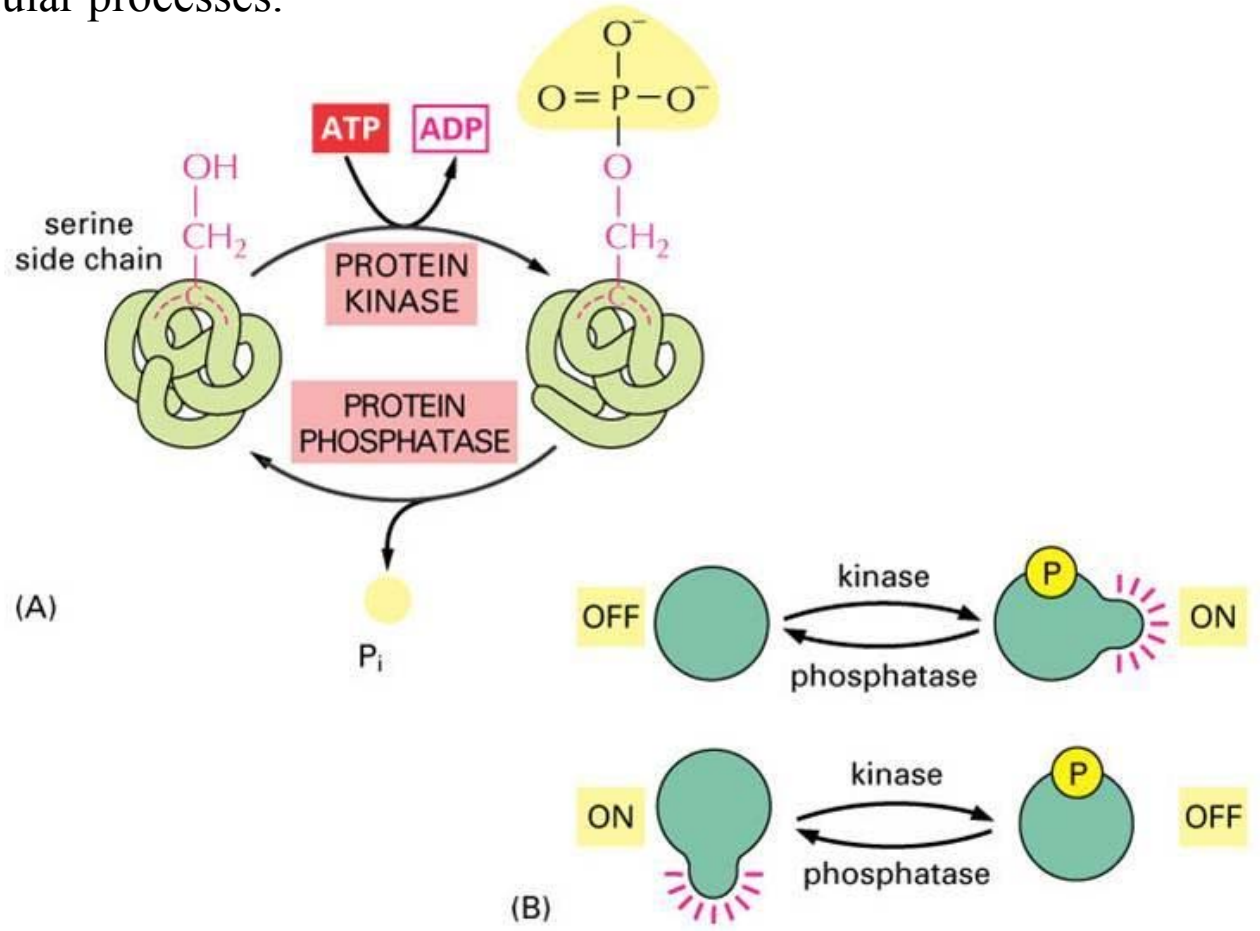
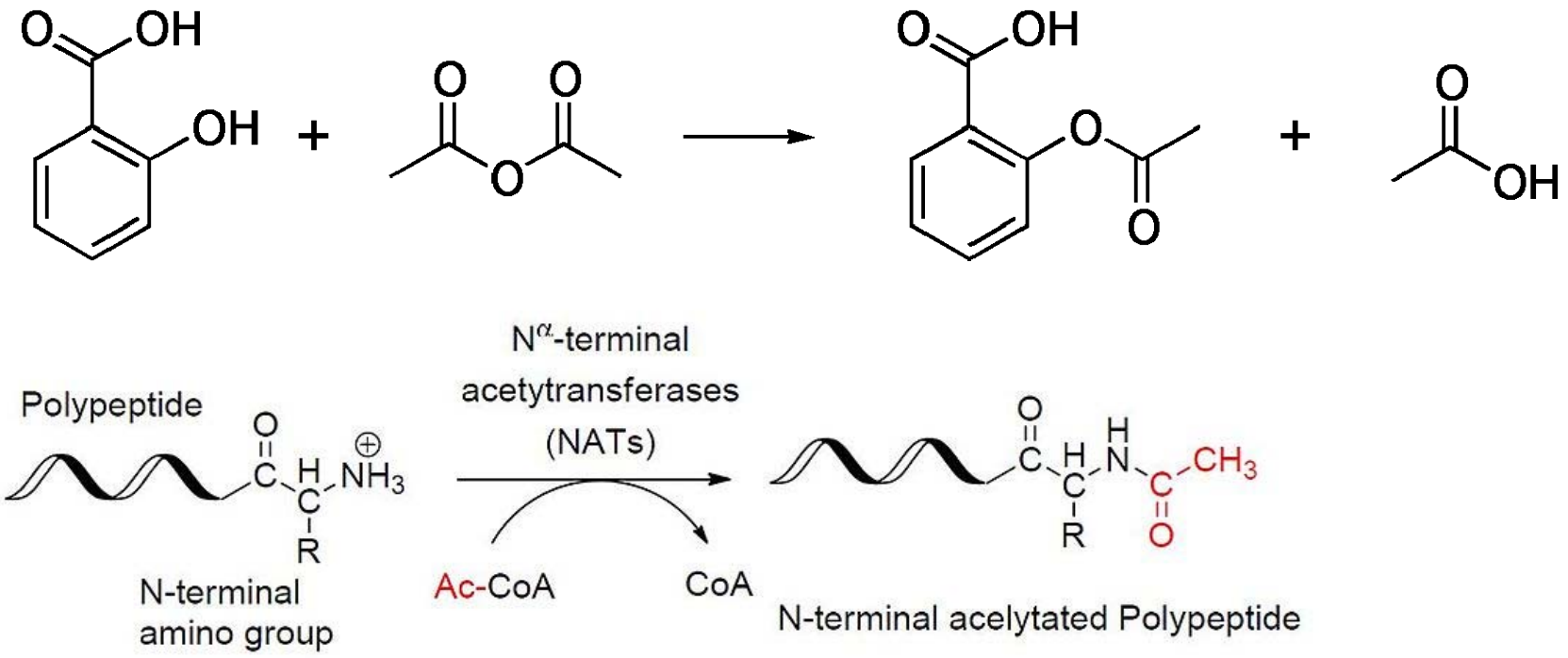



Figure 4-41 Essential Cell Biology, 2/e. (© 2004 Garland Science)

# Acetylation

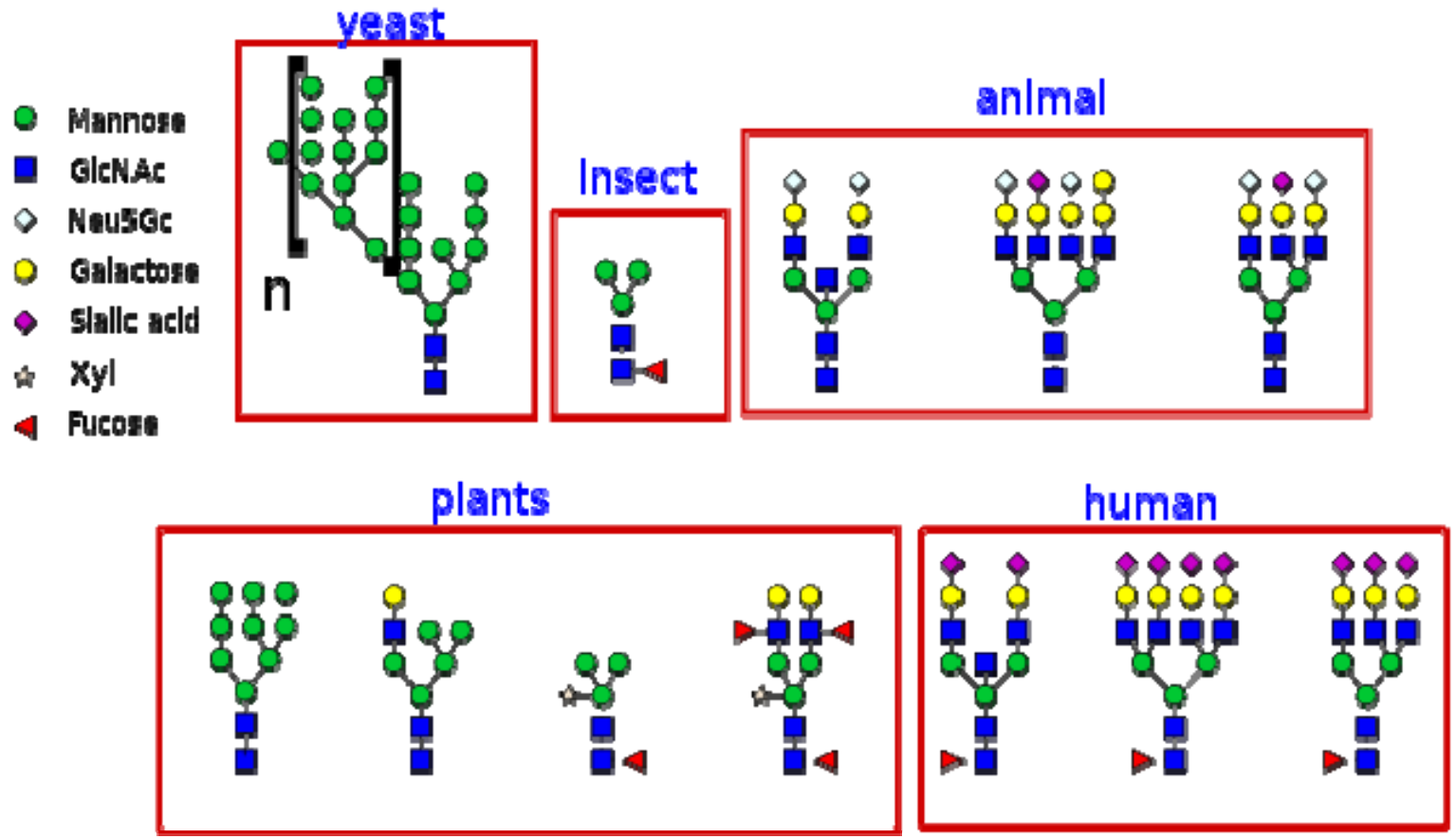
Acetylation (or in IUPAC nomenclature ethanoylation) describes a reaction that introduces an acetyl functional group into a chemical compound. Deacetylation is the removal of an acetyl group.




N-terminal acetylation

# N-linked glycosylation

N-linked glycosylation, is the attachment of the sugar molecule oligosaccharide known as glycan to a nitrogen atom (amide nitrogen of asparagine (Asn) residue of a protein), in a process called N-glycosylation, studied in biochemistry.



# Methylation

- In the chemical sciences, methylation denotes the addition of a methyl group on a substrate, or the substitution of an atom (or group) by a methyl group. Methylation is a form of alkylation, with a methyl group, rather than a larger carbon chain, replacing a hydrogen atom. These terms are commonly used in chemistry, biochemistry, soil science, and the biological sciences.
- DNA/RNA methylation

  DNA methylation in vertebrates typically occurs at CpG sites (cytosine-phosphate-guanine sites–that is, where a cytosine is directly followed by a guanine in the DNA sequence).

- Protein methylation

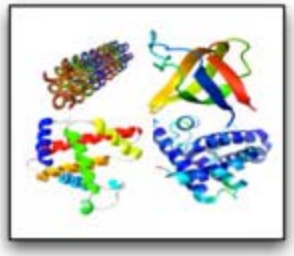
  Protein methylation typically takes place on arginine or lysine amino acid residues in the protein sequence.

# Workflow of Proteomics

Cell Culture
Clinical Samples
Animal Tissue

Protein Extraction

Protein Mixture

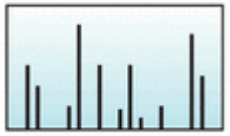


Peptide Digestion

Digested Peptides

LC-MS/MS Analysis

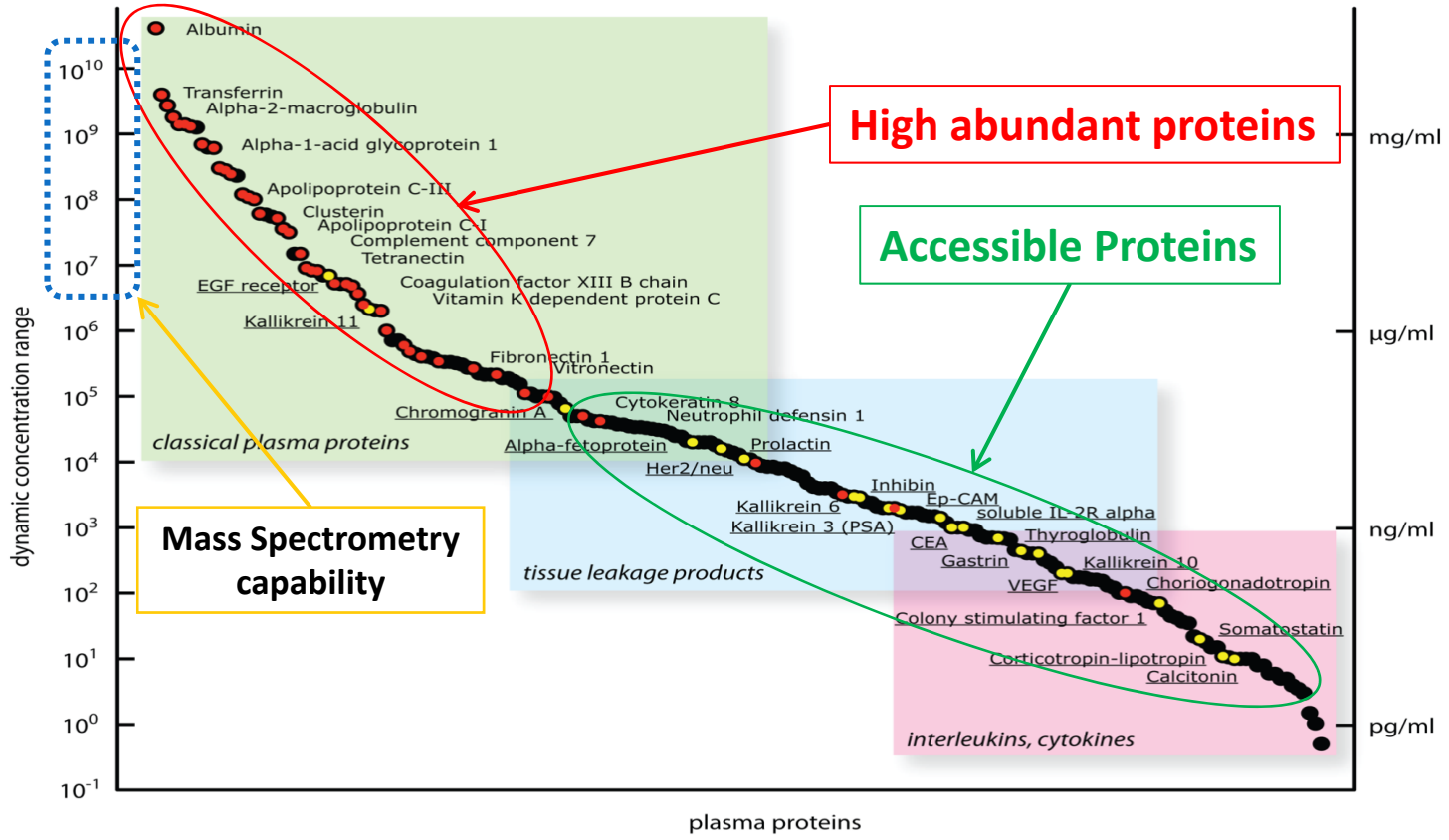
MS/MS spectra of peptides




Peptide Fingerprint Identification


**Bioinformatics**

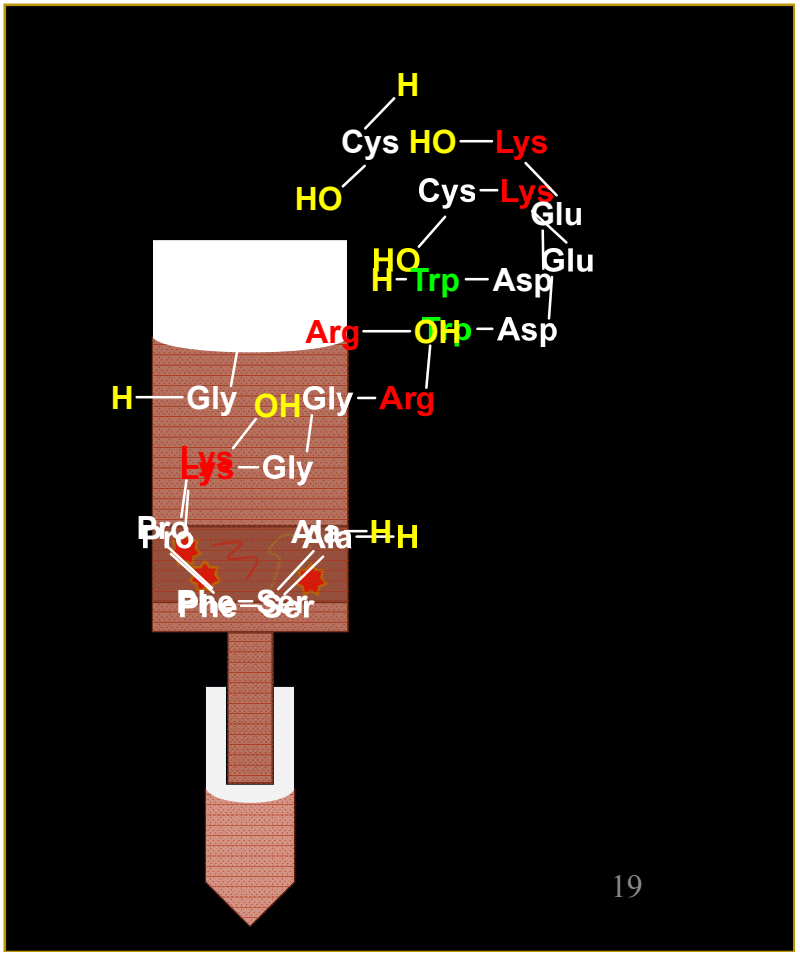
# Sample Preparation (Cont.)



# Techniques

- 2D-Gel electrophoresis, Sample enrichment (Beads, Affinity Matrix)
- BCA Assay
- Shot gun proteomics

  (Tryptic digestion)
- Solid-phase extraction

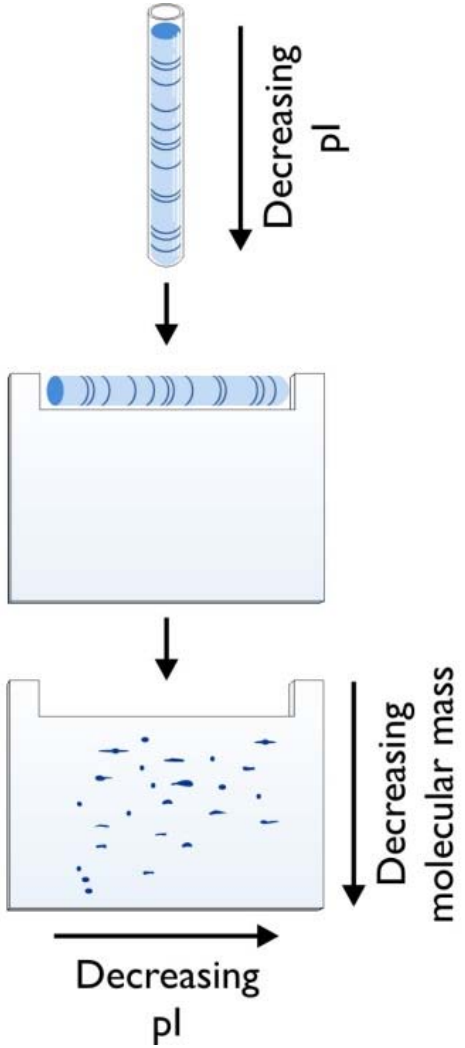


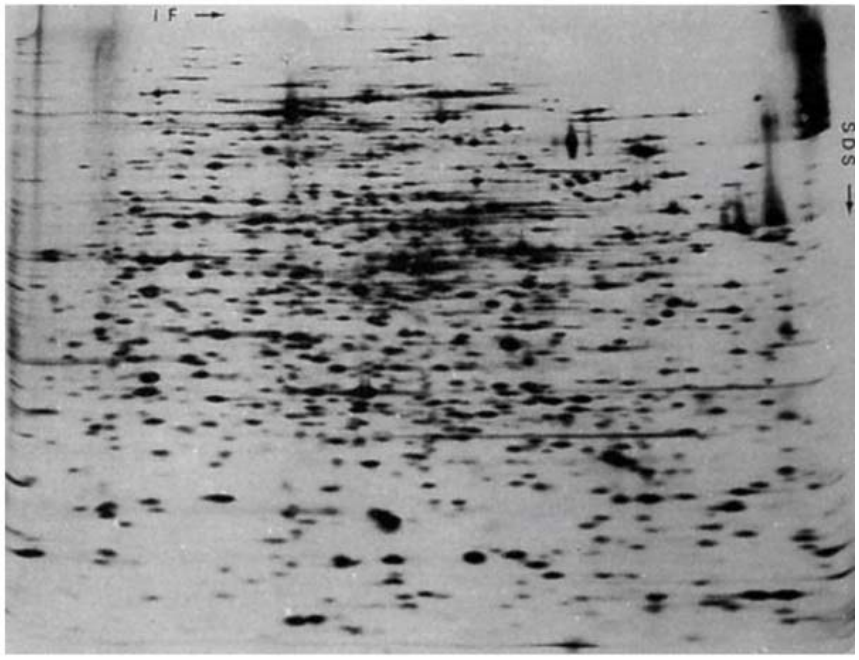
# 2D-Gel electrophoresis

1. Separation of proteins by pI value

2. Soaking the gel in SDS solution and fitting it on an SDS PA gel

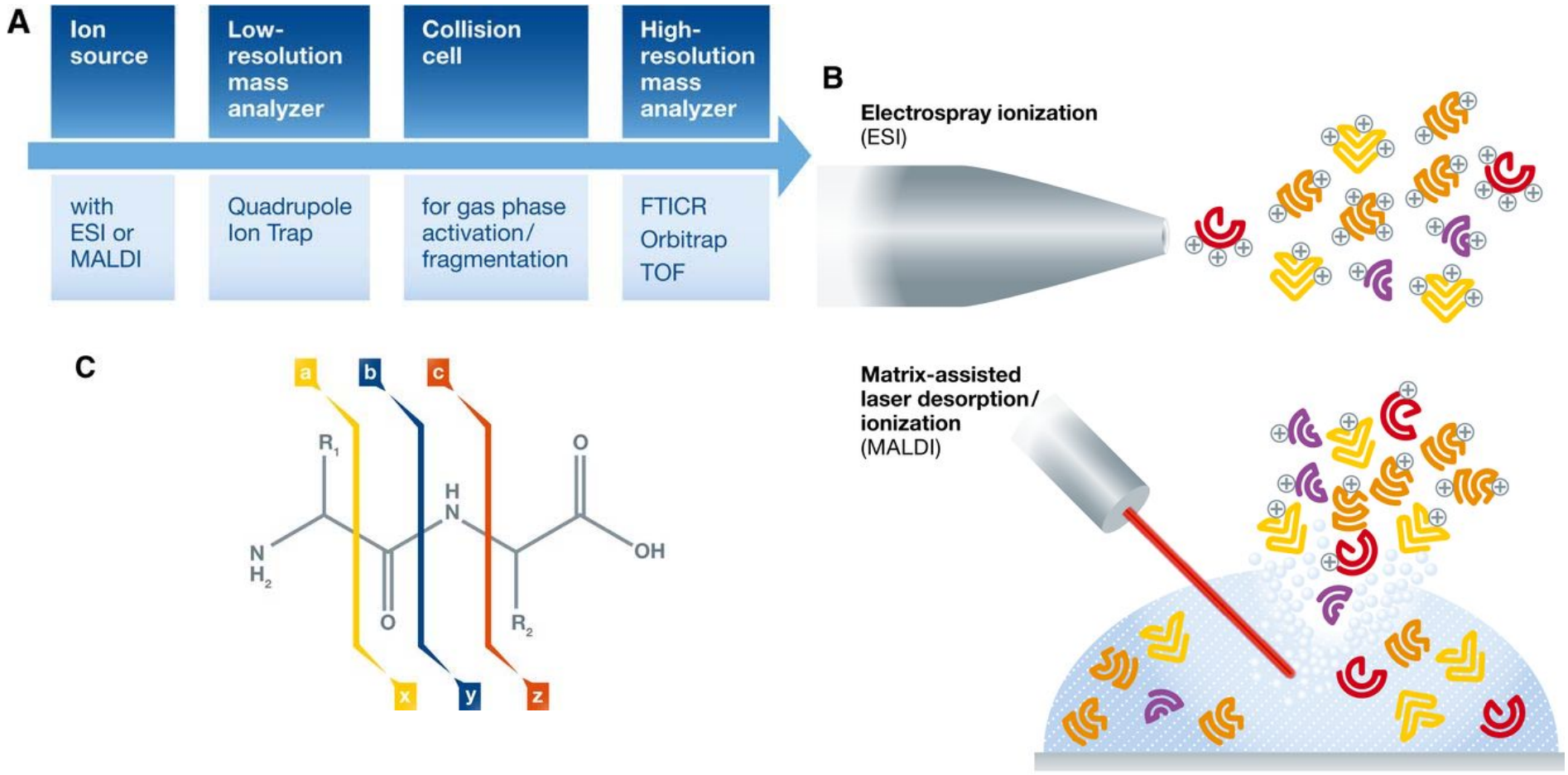
3. Separating the proteins by molecular mass with SDS PAGE



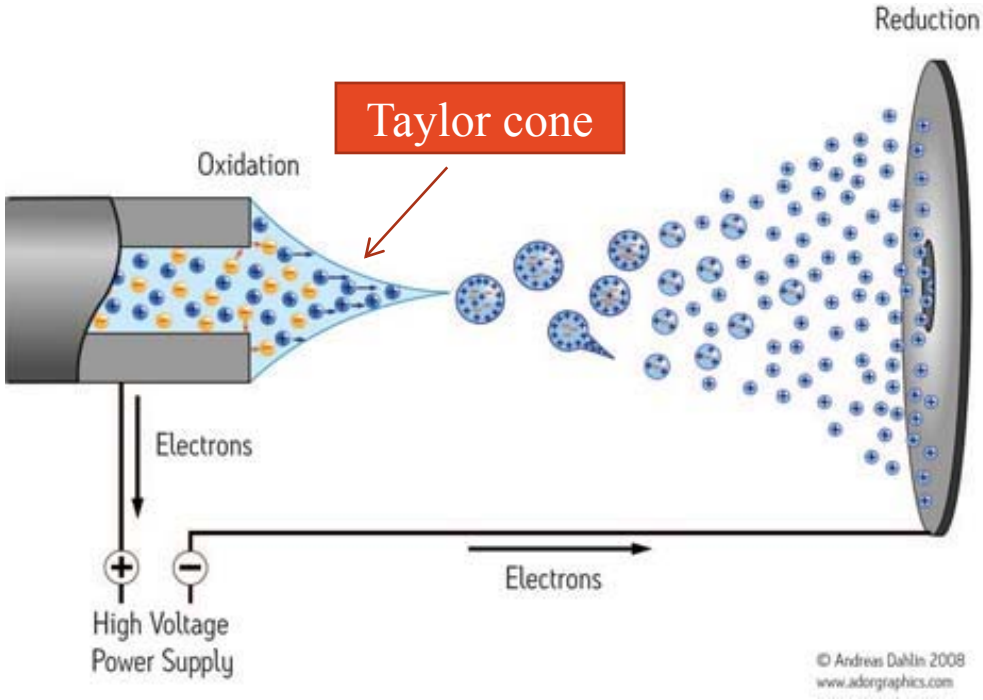

Approximately 1000 *E. coli* proteins on a single 2D gel

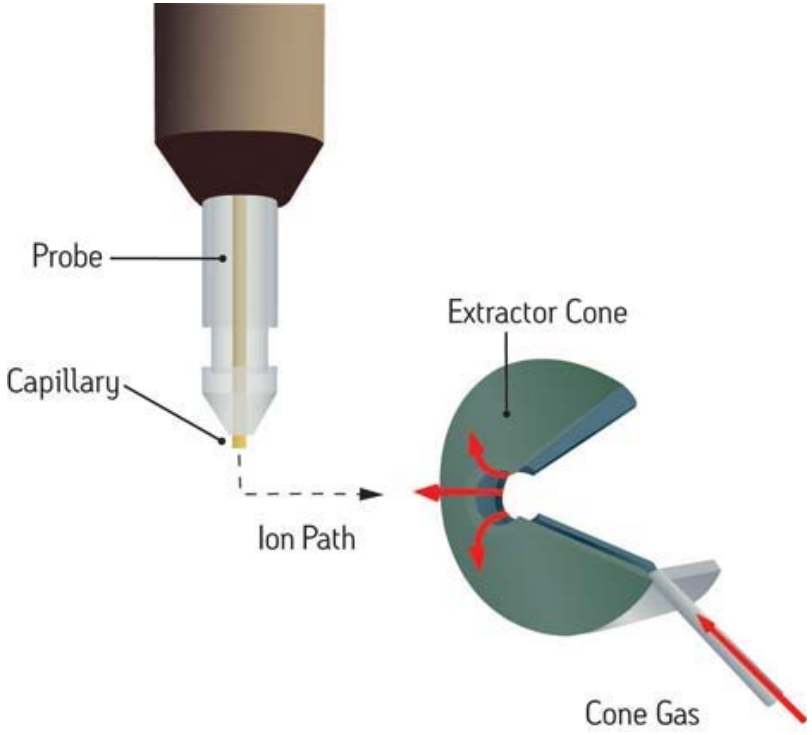
# Basic principles of biomolecular MS



# Principle of Electrospray Ionization (ESI)

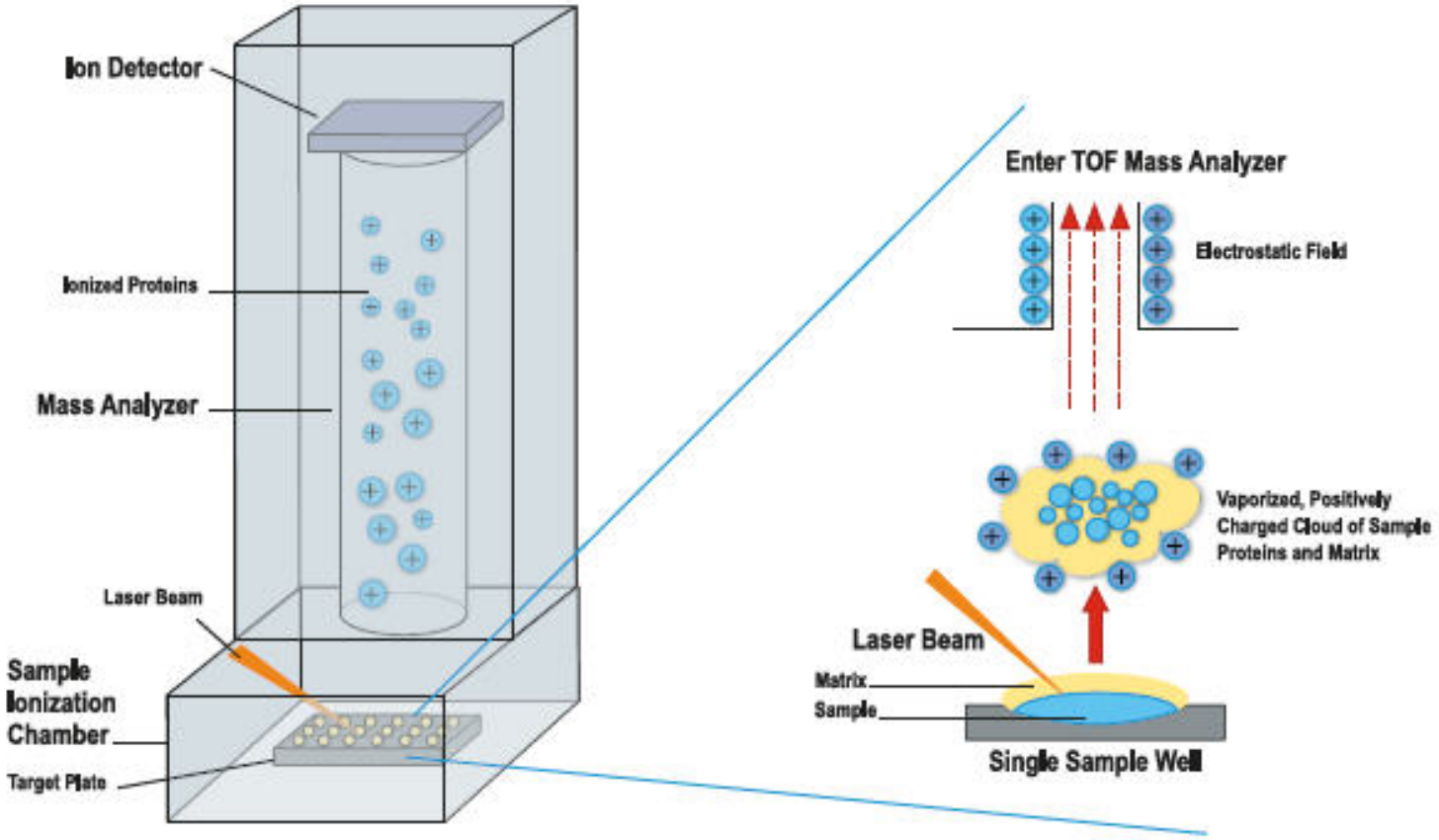
Electrospray ionization (ESI) is a technique used in mass spectrometry to produce ions using an electrospray in which a high voltage is applied to a liquid to create an aerosol.



# Principle of Matrix-assisted laser desorption/ionization (MALDI)

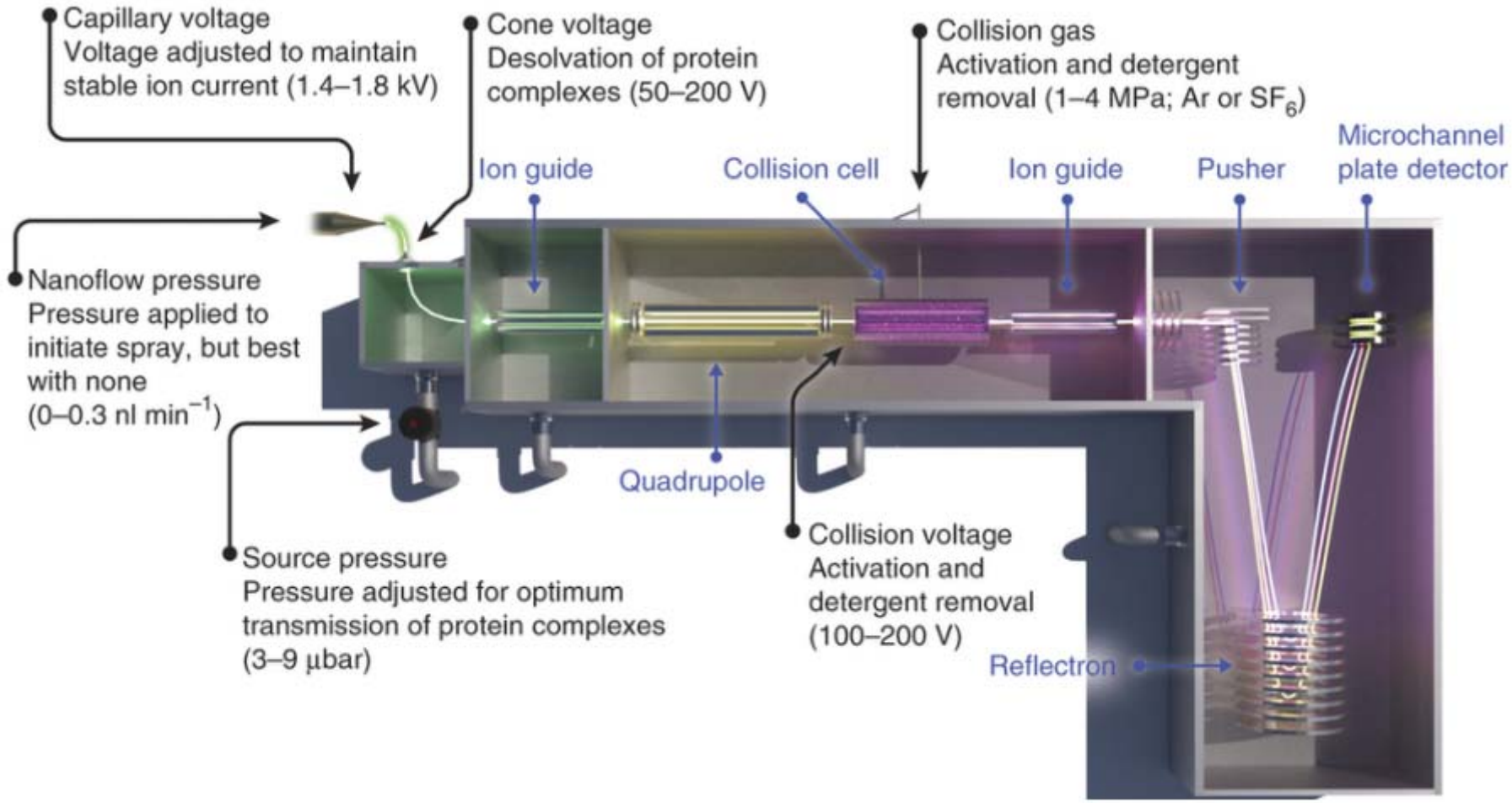
MALDI is the abbreviation for "Matrix Assisted Laser Desorption/Ionization." The sample for MALDI is uniformly mixed in a large quantity of matrix.



# Mass Analyzers (MS)

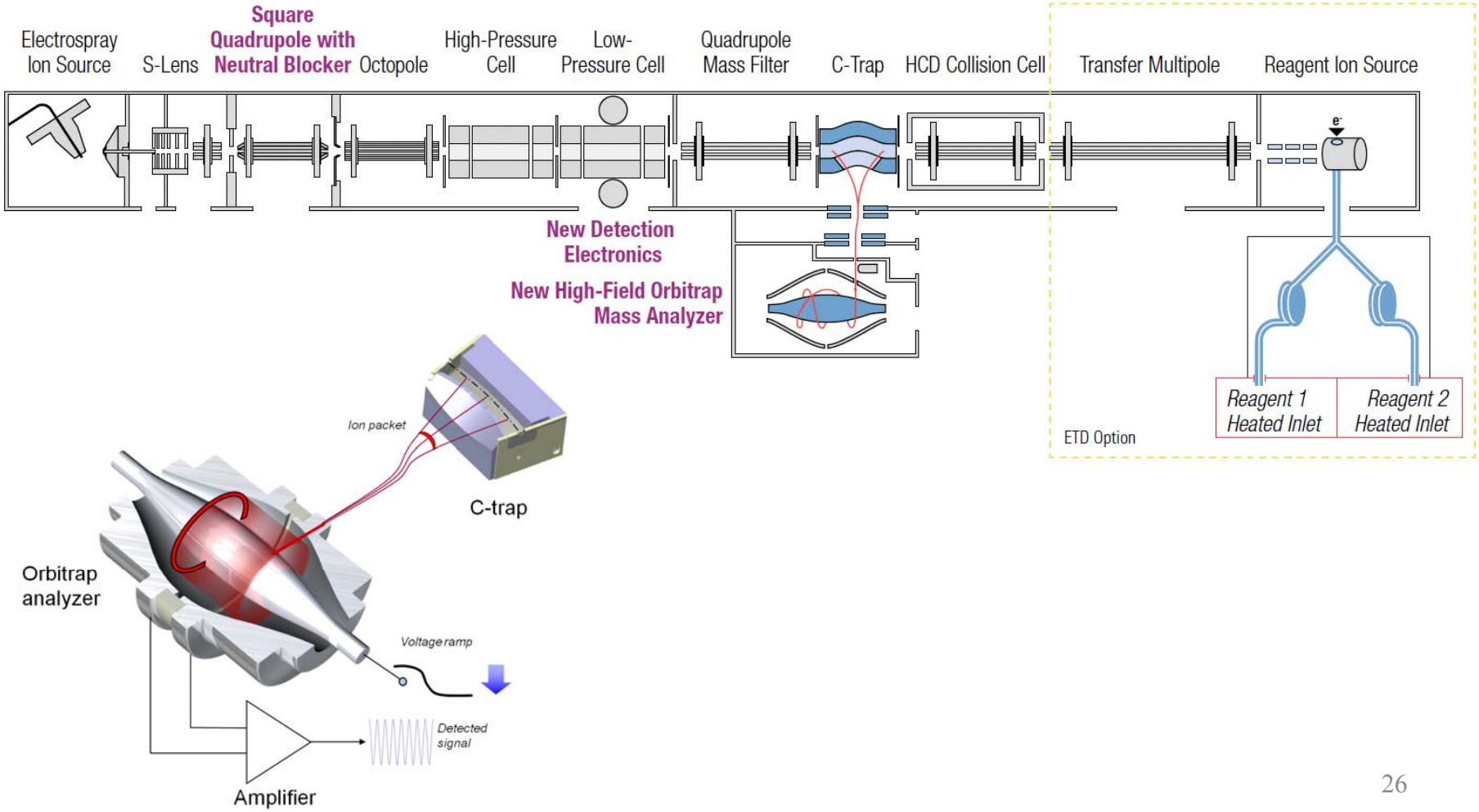
- **Quadrapole**
  - High Sensitivity, acceptable mass accuracy and resolution
  - Easily coupled to chromatography
- **Time of Flight**
  - High Sensitivity, high mass accuracy, high resolution
  - Limited to small m/z ratios
  - Not easily coupled to chromatography
  - Easily coupled to MALDI
- **Ion Trap**
  - High Sensitivity
  - Low mass accuracy and resolution

# Schematic of a Q-TOF mass spectrometer

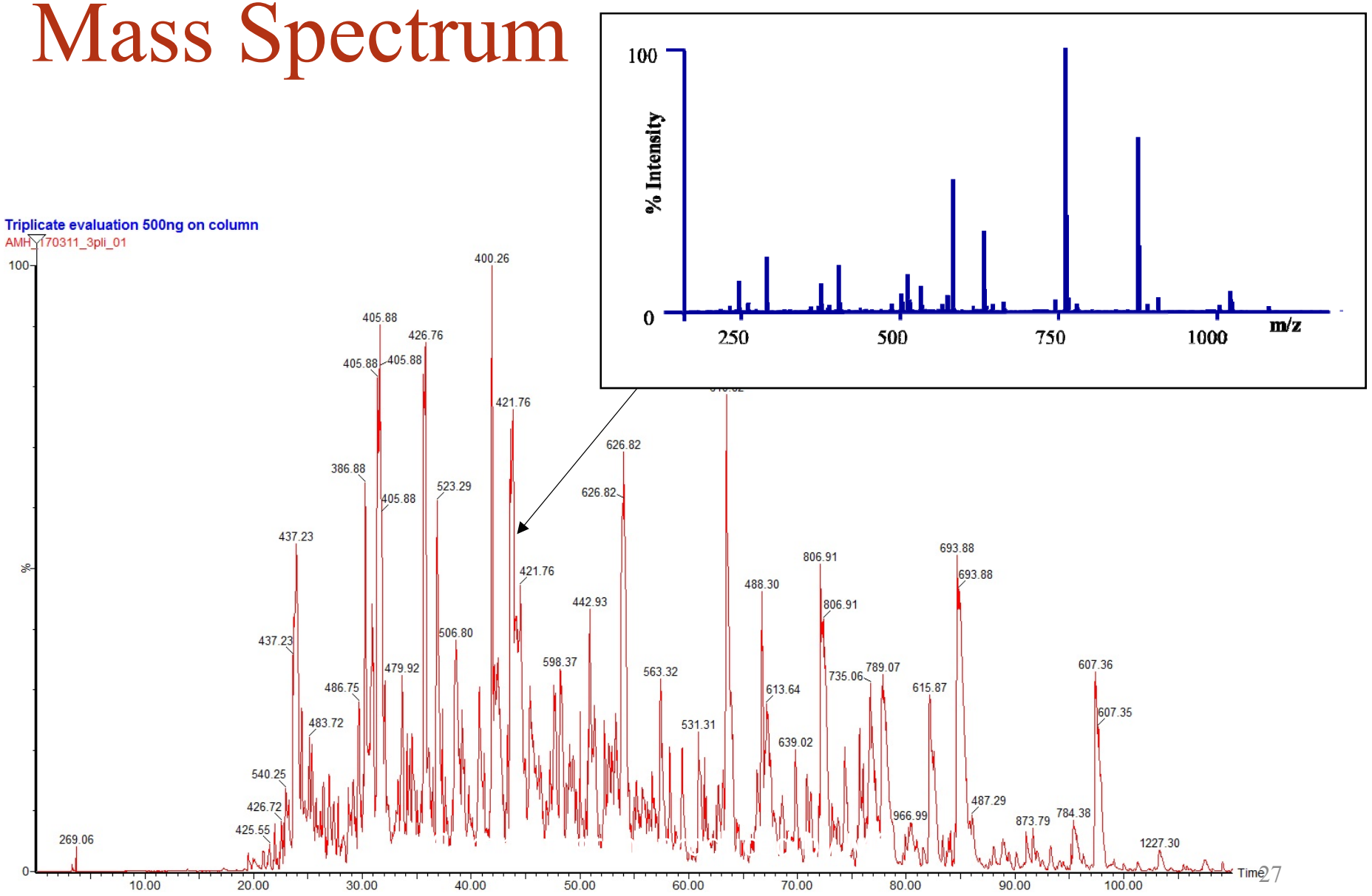



Nature Protocols 8, 639–651 (2013) doi:10.1038/nprot.2013.024

# Schematic of a orbitrap mass spectrometer



# Chromatography and Mass Spectrum



# Equipment in CLB, KMUH

Waters UPLC

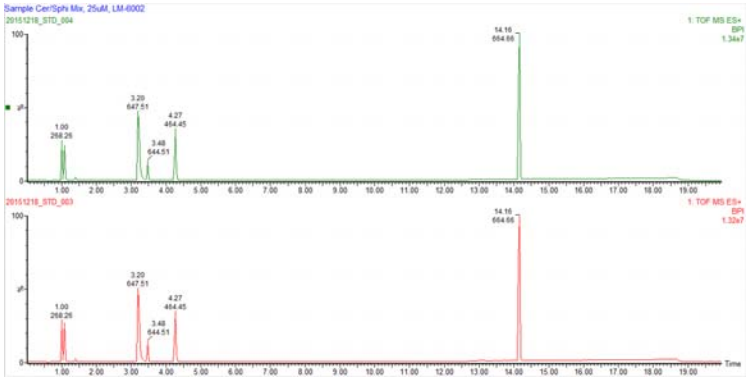
Low Flow ESI

Waters G2 qTof

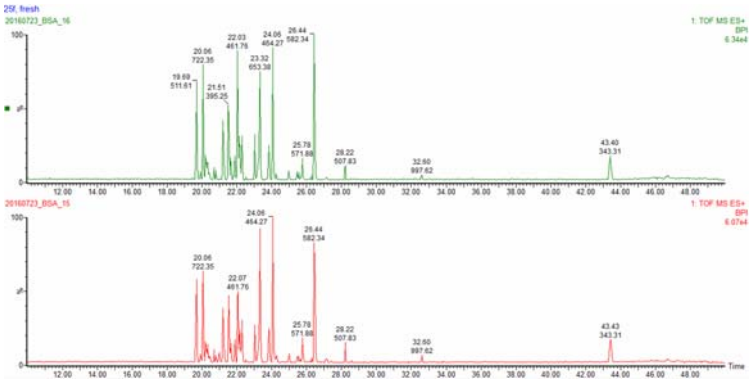
Waters M-class UPLC

NanoFlow ESI

Lipid Identification



Peptide Identification



# Bioinformatics server in CLB, KMUH

Statistic software service

Microarray data analysis service

NGS analysis service

Proteomics analysis service

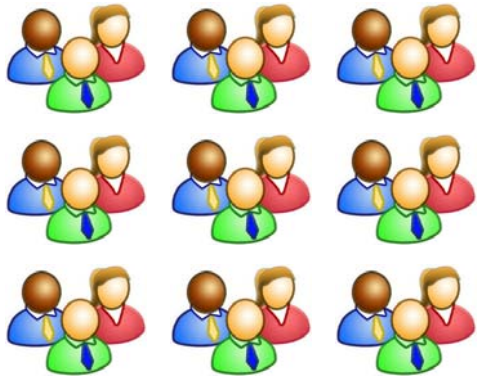
Meta-analysis service

Lipidomics analysis service

Image process service

Glycosylation analysis service

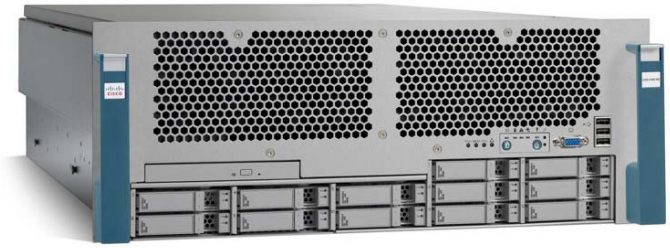
Remote control software



Users

Windows Server Hyper-V

Stimulated virtual machine

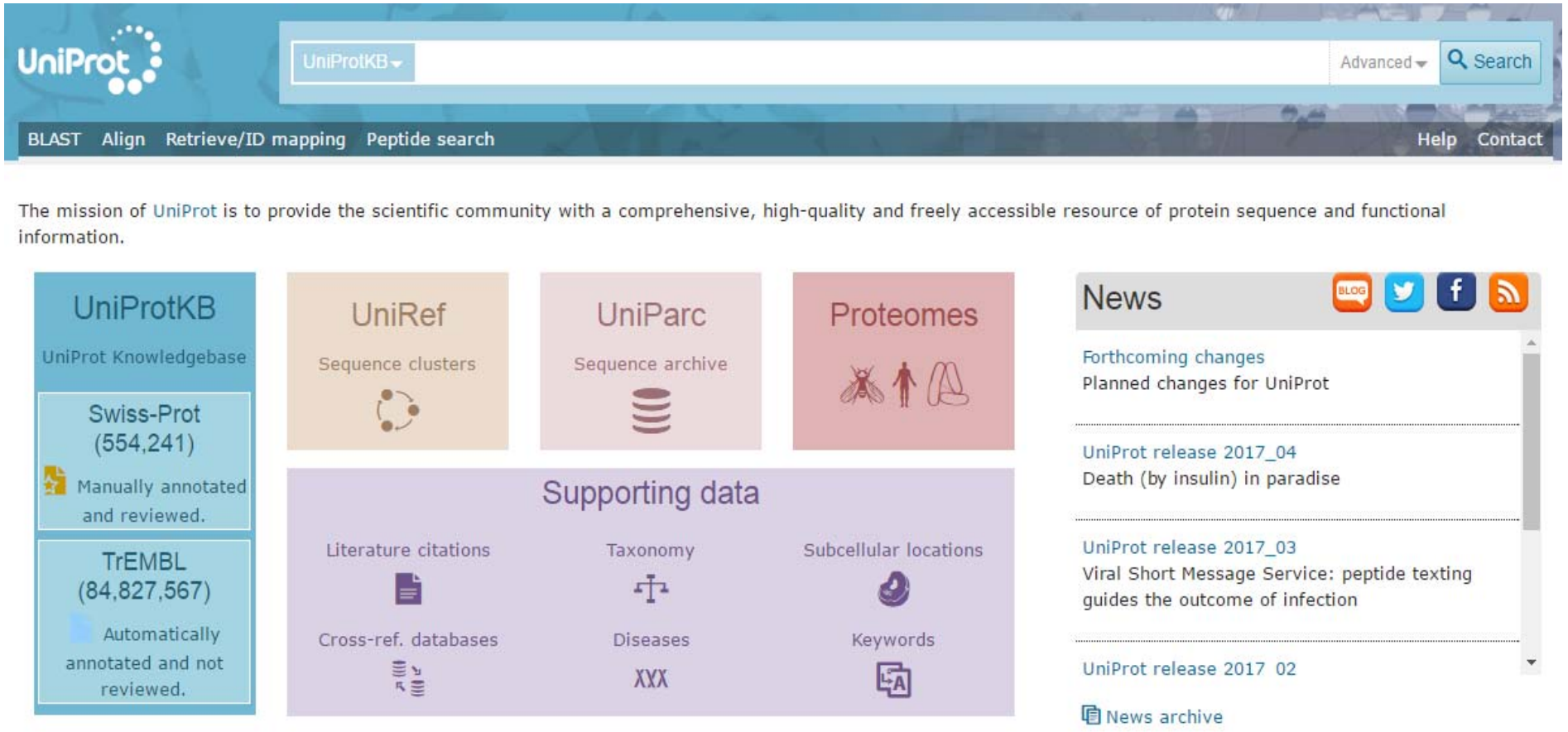


Cisco UCS C460 M2 High Performance Rack Server Equip with 4 Intel Xeon E4860 10 core processors

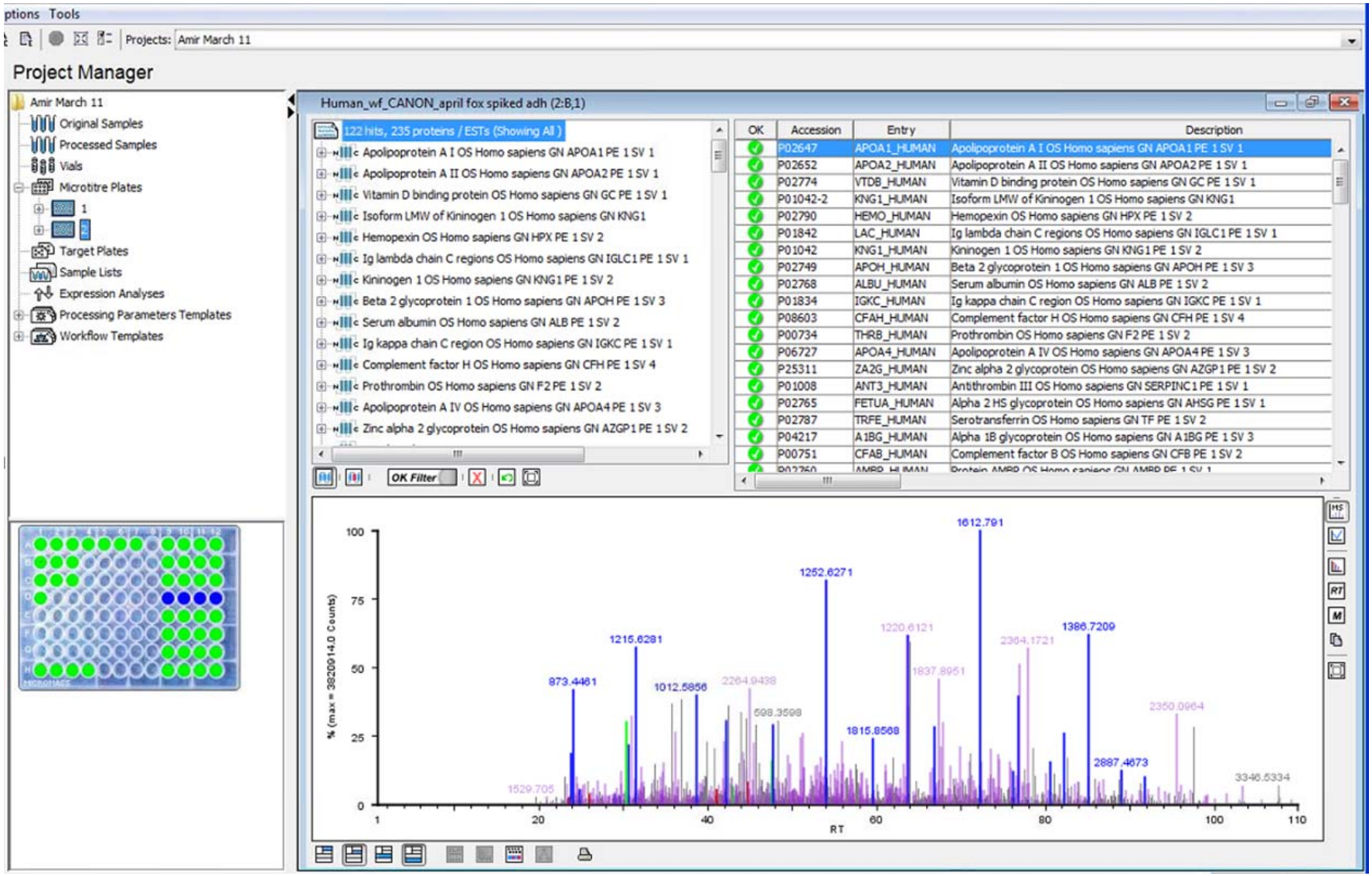
Data storage

Synology RackStation NAS RS18016xs+

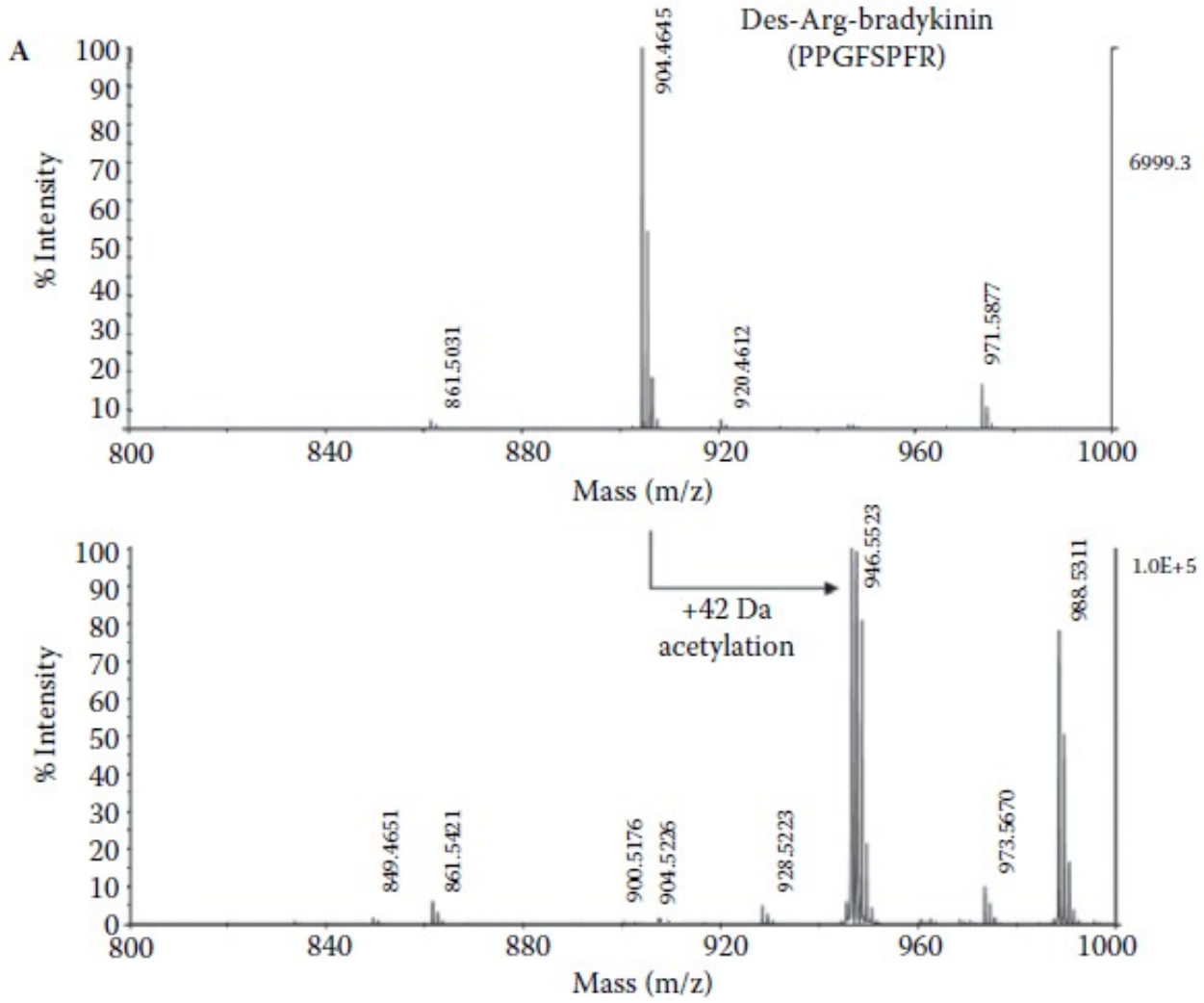
# Protein Database



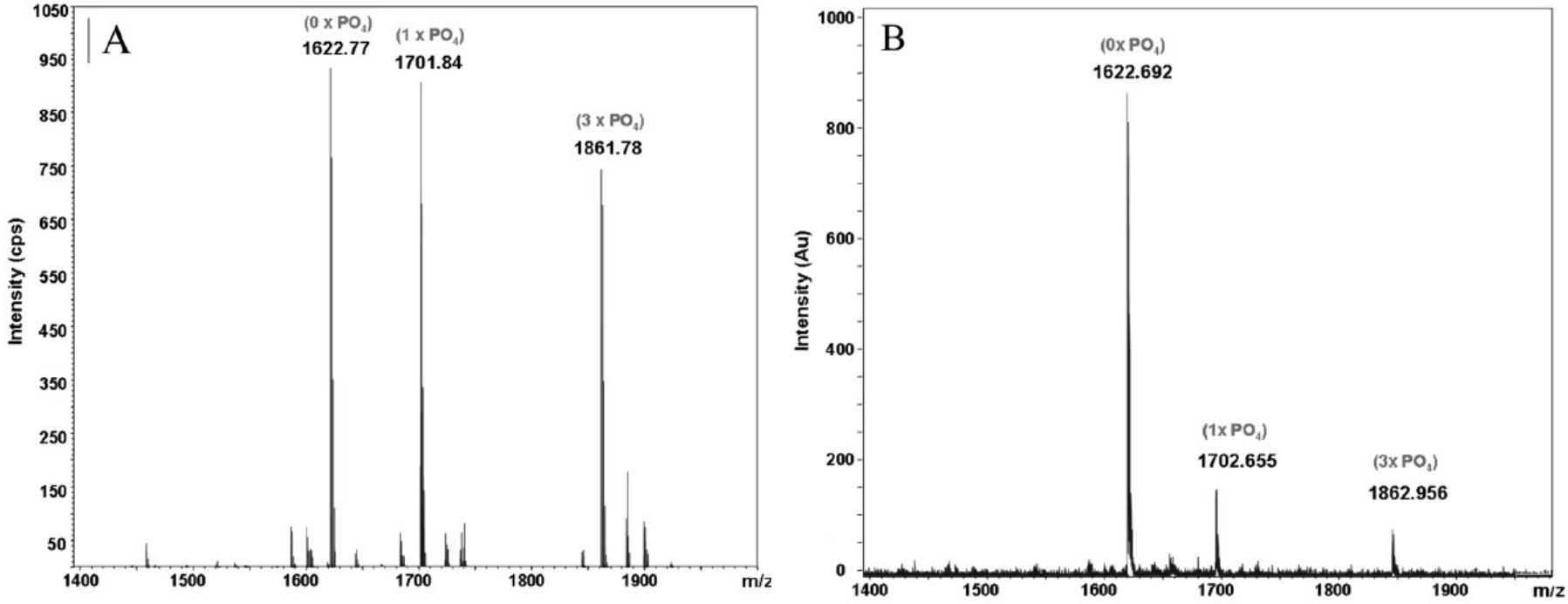
# Identity $MS^E$ Results from PLGS



# Acetylation in MSMS

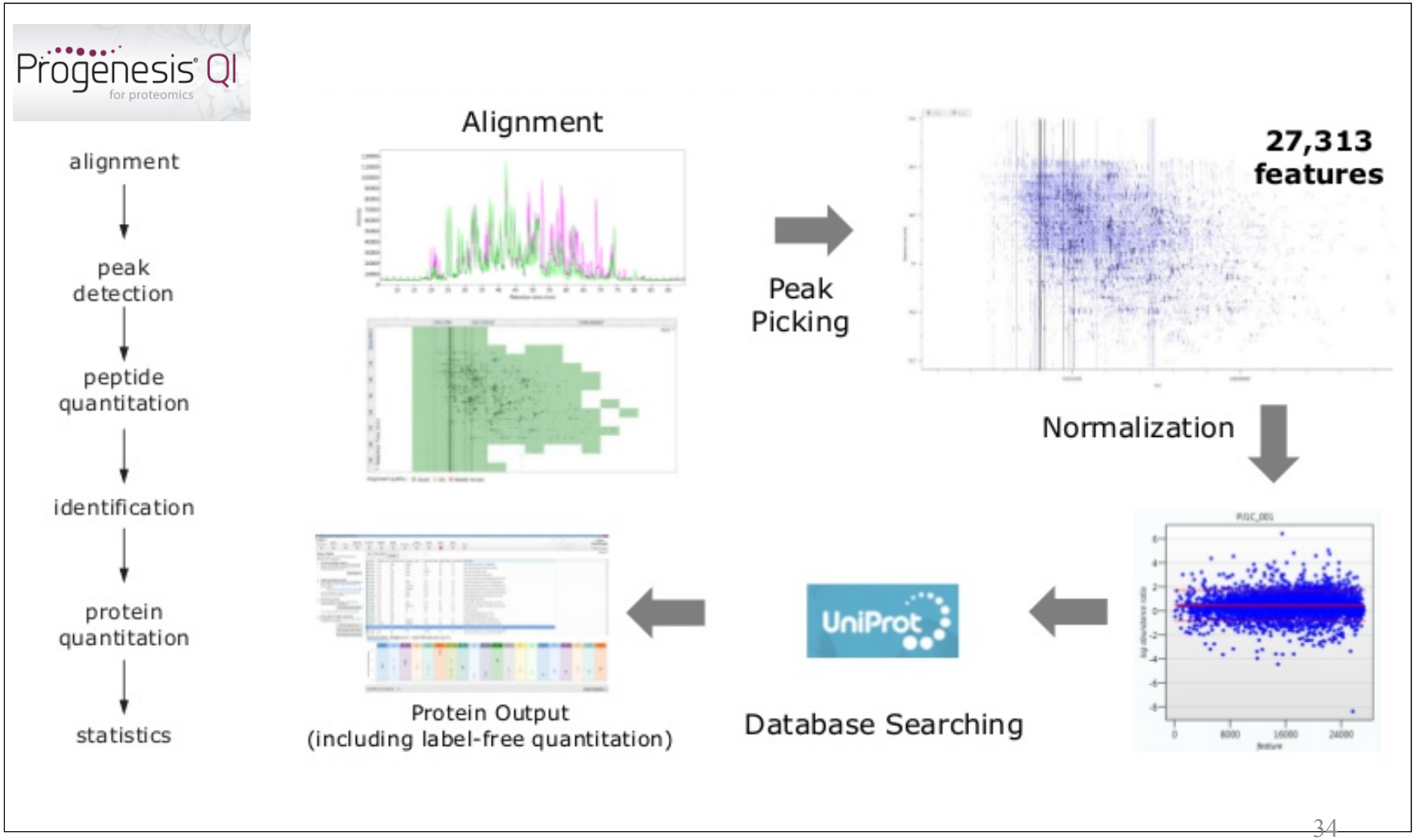


# Phosphorylation in MSMS

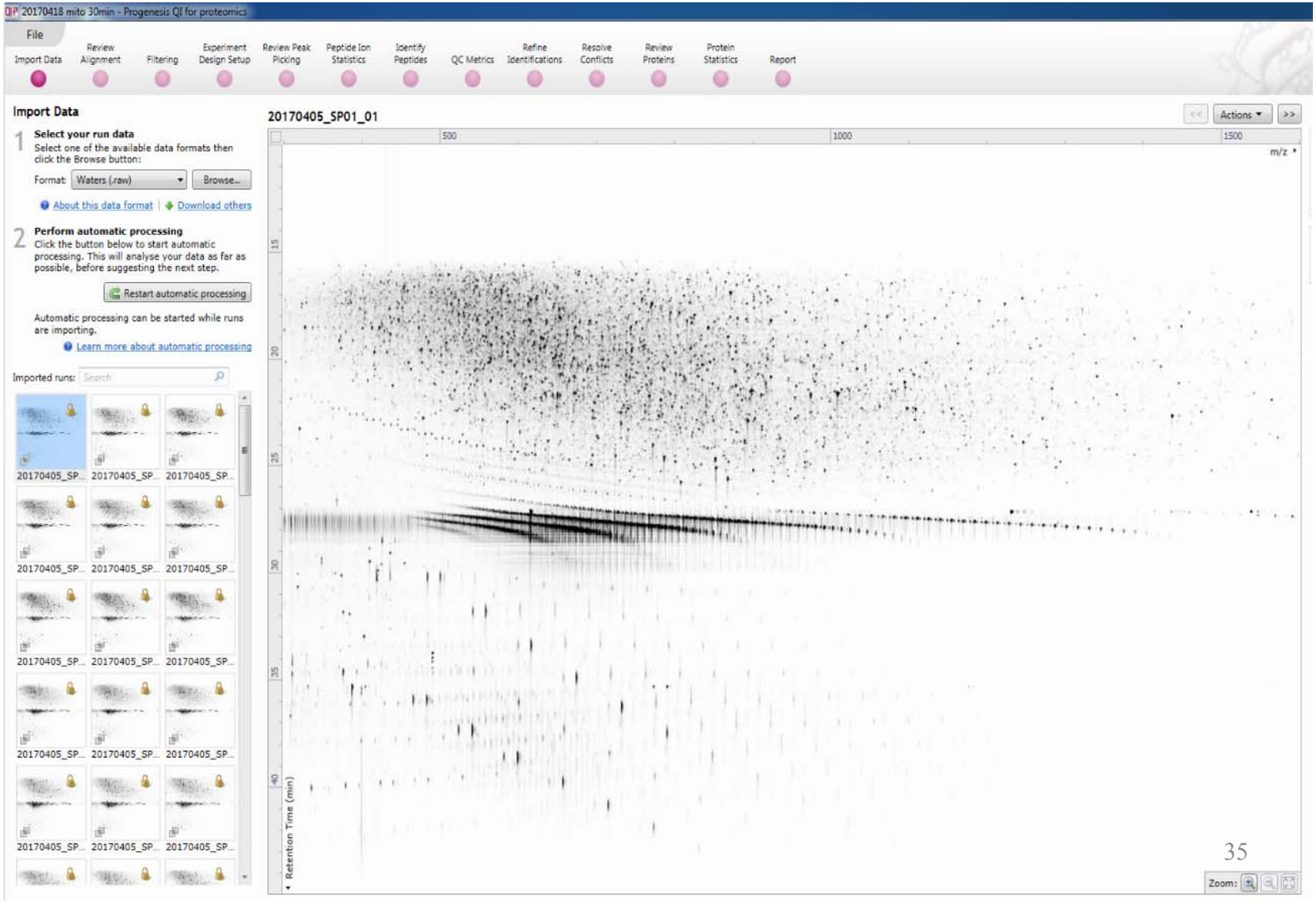



Sensitivity as a function of the number of phosphoryl groups. A: Positive ion electrospray (deconvoluted spectrum). B: Positive ion MALDI-MS (DHB matrix).

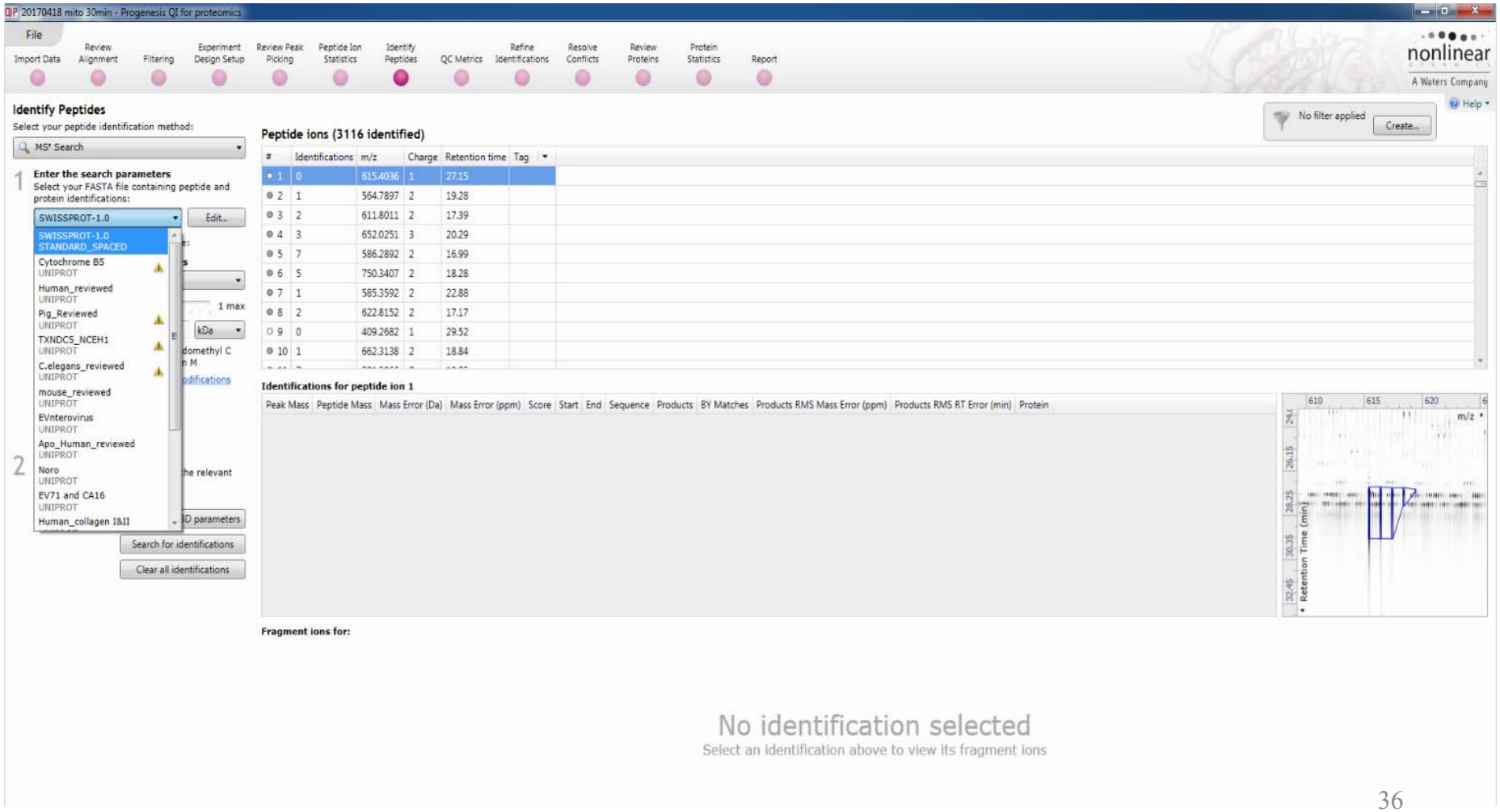
# Progenesis QI for proteomics workflow



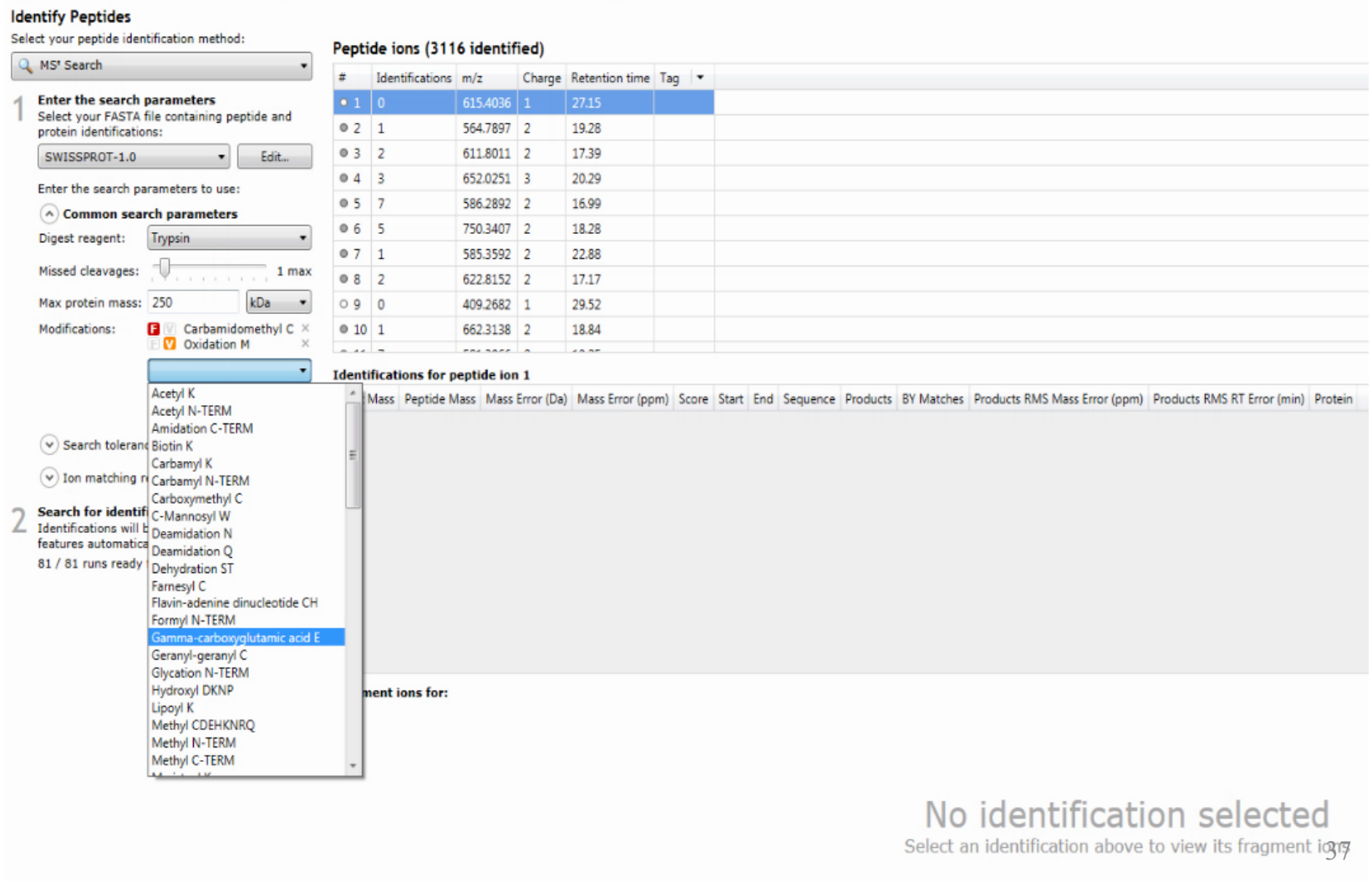
# Relative quantification by Progenesis QI for proteomics



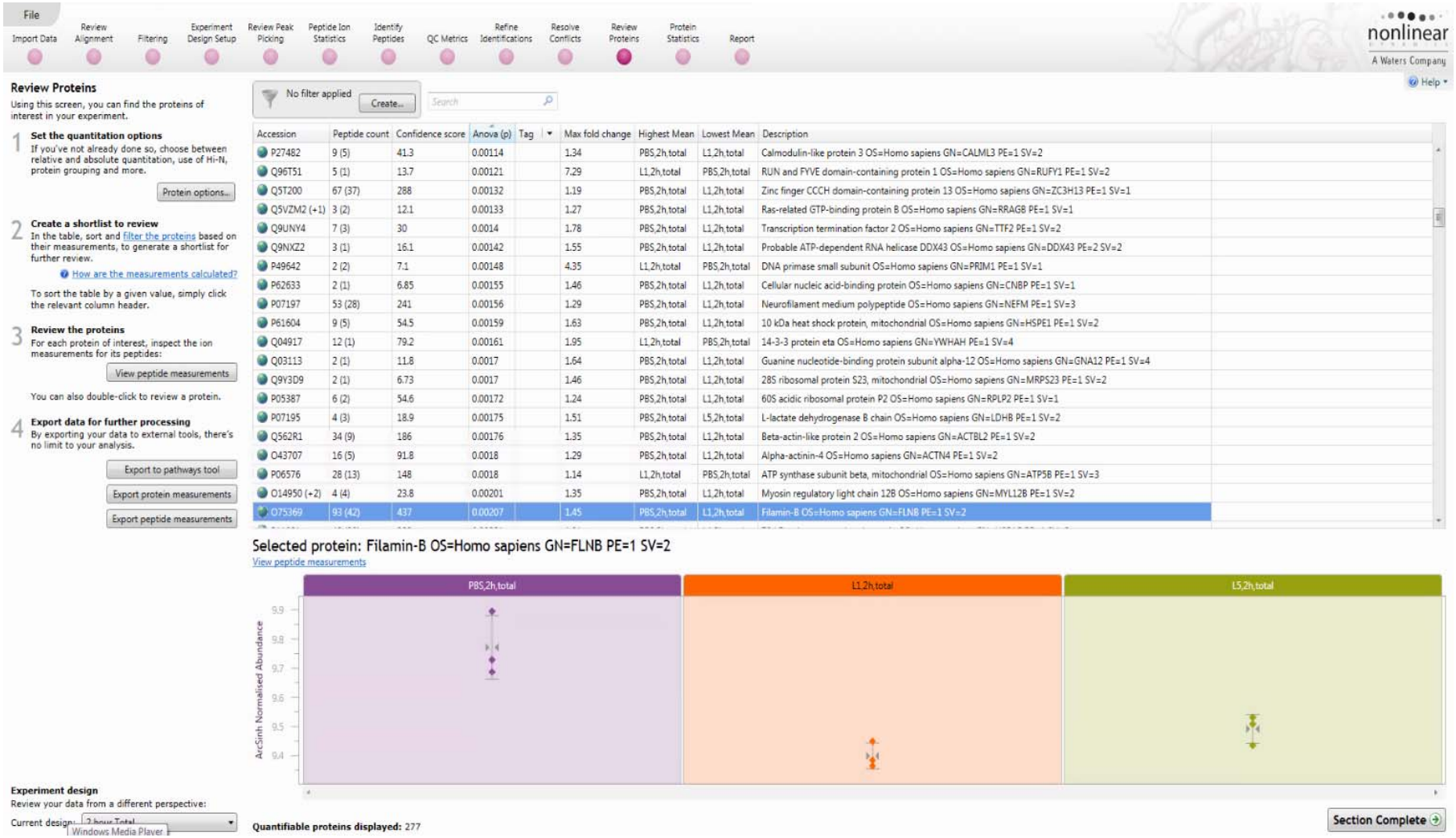
# Identify peptides fingerprint



# PTMs Search


Identify Peptides
Select your peptide identification method:
MS² Search
1 Enter the search parameters
Select your FASTA file containing peptide and protein identifications:
SWISSPROT-1.0
Edit...
Enter the search parameters to use:
Common search parameters
Digest reagent: Trypsin
Missed cleavages: 1 max
Max protein mass: 250 kDa
Modifications: Carbamidomethyl C
Oxidation M
Acetyl K
Acetyl N-TERM
Amidation C-TERM
Biotin K
Carbamyl K
Carbamyl N-TERM
Carboxymethyl C
C-Mannosyl W
Deamidation N
Deamidation Q
Dehydration ST
Farnesyl C
Flavin-adenine dinucleotide CH
Formyl N-TERM
Gamma-carboxyglutamic acid E
Geranyl-geranyl C
Glycation N-TERM
Hydroxyl DKNP
Lipoyl K
Methyl CDEHKNRQ
Methyl N-TERM
Methyl C-TERM
Search toleranc
Ion matching r
2 Search for identifi
Identifications will b
features automatica
81 / 81 runs ready
Peptide ions (3116 identified)
# Identifications m/z Charge Retention time Tag
1 0 615.4036 1 27.15
2 1 564.7897 2 19.28
3 2 611.8011 2 17.39
4 3 652.0251 3 20.29
5 7 586.2892 2 16.99
6 5 750.3407 2 18.28
7 1 585.3592 2 22.88
8 2 622.8152 2 17.17
9 0 409.2682 1 29.52
10 1 662.3138 2 18.84
Identifications for peptide ion 1
Mass Peptide Mass Mass Error (Da) Mass Error (ppm) Score Start End Sequence Products BY Matches Products RMS Mass Error (ppm) Products RMS RT Error (min) Protein
ment ions for:
No identification selected
Select an identification above to view its fragment ions

# Review proteins

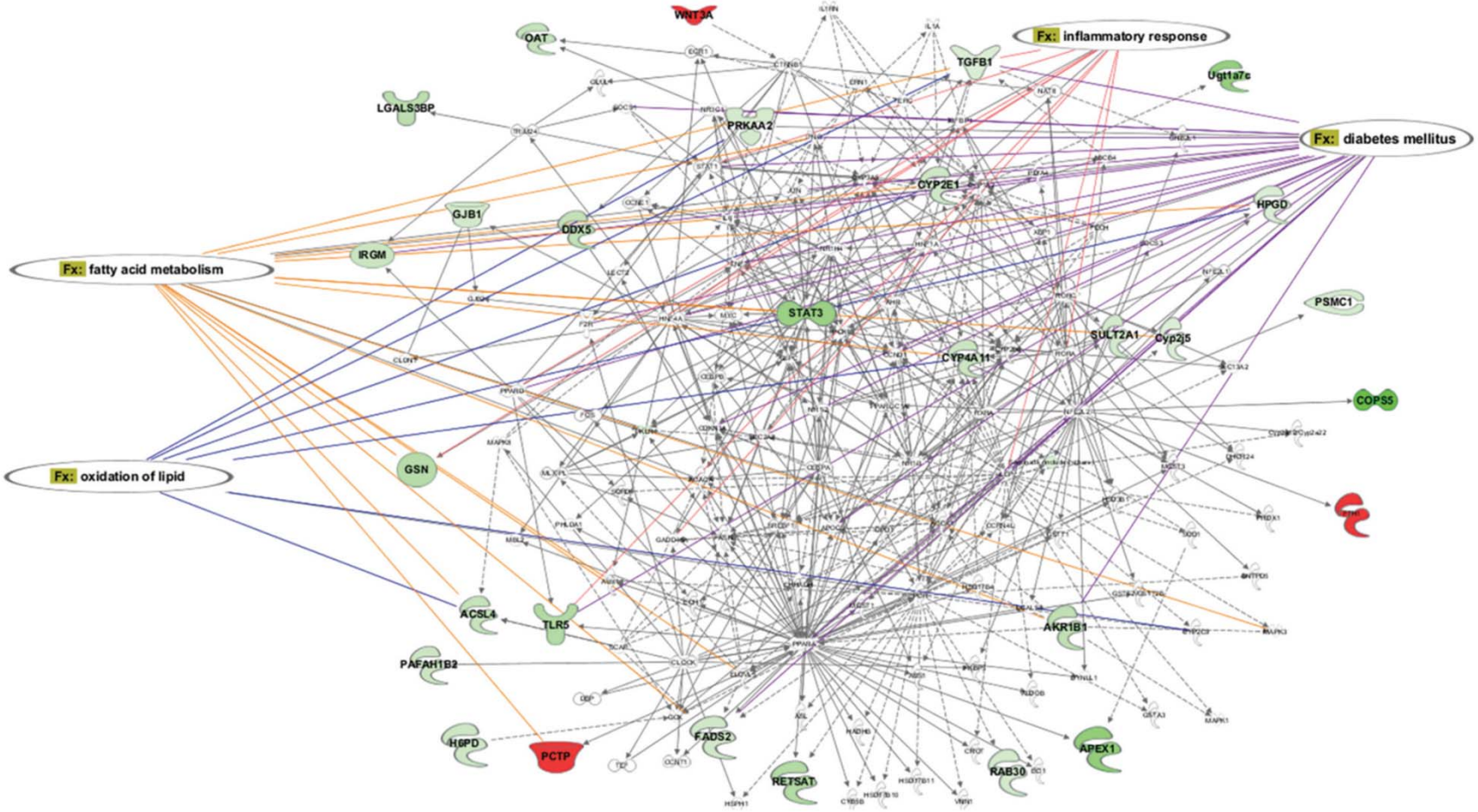


# Reports

| Accession | Peptides | Score | Anova (p)* | Fold | Tags | Description | Average Normalised Abundances | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | PBS | L1 | L5 |
| P08670 | 39 (34) | 312.55 | 8.47e-004 | 1.13 | | Vimentin OS=Homo sapiens GN=VIM PE=1 SV=4 | 5.58e+004 | 5.56e+004 | 6.30e+004 |
| P07355 | 32 (29) | 293.85 | 0.01 | 1.16 | | Annexin A2 OS=Homo sapiens GN=ANXA2 PE=1 SV=2 | 7.43e+004 | 7.12e+004 | 6.42e+004 |
| P08238 | 30 (15) | 241.90 | 0.03 | 1.14 | | Heat shock protein HSP 90-beta OS=Homo sapiens GN=HSP90AB1 PE=1 SV=4 | 5.96e+004 | 5.84e+004 | 6.67e+004 |
| P11021 | 33 (28) | 234.40 | 2.57e-003 | 1.20 | | 78 kDa glucose-regulated protein OS=Homo sapiens GN=HSPA5 PE=1 SV=2 | 2.75e+004 | 2.59e+004 | 3.12e+004 |
| P63261 | 22 (12) | 190.16 | 0.72 | 1.03 | | Actin, cytoplasmic 2 OS=Homo sapiens GN=ACTG1 PE=1 SV=1 | 2.56e+005 | 2.65e+005 | 2.56e+005 |
| A5A3E0 | 33 (4) | 188.64 | 0.07 | 1.07 | | POTE ankyrin domain family member F OS=Homo sapiens GN=POTEF PE=1 SV=2 | 6370.96 | 6568.73 | 6831.77 |
| P07900 | 26 (10) | 179.56 | 2.94e-004 | 1.25 | | Heat shock protein HSP 90-alpha OS=Homo sapiens GN=HSP90AA1 PE=1 SV=5 | 1.11e+004 | 1.08e+004 | 1.35e+004 |
| P07437 | 23 (4) | 177.81 | 3.71e-005 | 1.32 | | Tubulin beta chain OS=Homo sapiens GN=TUBB PE=1 SV=2 | 1.02e+004 | 7704.69 | 7805.81 |
| Q6S8J3 | 30 (0) | 176.93 | --- | --- | | POTE ankyrin domain family member E OS=Homo sapiens GN=POTEE PE=1 SV=3 | --- | --- | --- |
| P08758 | 23 (22) | 170.67 | 0.02 | 1.12 | | Annexin A5 OS=Homo sapiens GN=ANXA5 PE=1 SV=2 | 3.07e+004 | 2.96e+004 | 2.73e+004 |
| P11142 | 30 (23) | 150.82 | 9.76e-004 | 1.10 | | Heat shock cognate 71 kDa protein OS=Homo sapiens GN=HSPA8 PE=1 SV=1 | 2.07e+004 | 1.99e+004 | 2.18e+004 |
| P68104 | 22 (21) | 147.64 | 0.37 | 1.04 | | Elongation factor 1-alpha 1 OS=Homo sapiens GN=EEF1A1 PE=1 SV=1 | 3.47e+004 | 3.60e+004 | 3.58e+004 |
| P04406 | 18 (18) | 146.70 | 1.38e-005 | 1.32 | | Glyceraldehyde-3-phosphate dehydrogenase OS=Homo sapiens GN=GAPDH PE=1 SV=3 | 4.59e+004 | 3.48e+004 | 3.91e+004 |
| P68371 | 18 (1) | 143.13 | 6.88e-003 | 1.24 | | Tubulin beta-4B chain OS=Homo sapiens GN=TUBB4B PE=1 SV=1 | 2777.83 | 2234.63 | 2567.48 |
| P0CG38 | 25 (4) | 139.92 | 6.56e-006 | 1.32 | | POTE ankyrin domain family member I OS=Homo sapiens GN=POTEI PE=3 SV=1 | 1.30e+004 | 9886.78 | 1.06e+004 |
| P04083 | 18 (16) | 135.36 | 0.42 | 1.06 | | Annexin A1 OS=Homo sapiens GN=ANXA1 PE=1 SV=2 | 2.63e+004 | 2.49e+004 | 2.51e+004 |
| P30101 | 20 (16) | 129.19 | 2.30e-003 | 1.10 | | Protein disulfide-isomerase A3 OS=Homo sapiens GN=PDIA3 PE=1 SV=4 | 2.03e+004 | 1.99e+004 | 2.19e+004 |
| Q9BQE3 | 18 (4) | 125.00 | 5.95e-005 | 1.14 | | Tubulin alpha-1C chain OS=Homo sapiens GN=TUBA1C PE=1 SV=1 | 4591.27 | 4014.31 | 4589.83 |
| P68363 | 17 (0) | 122.99 | --- | --- | | Tubulin alpha-1B chain OS=Homo sapiens GN=TUBA1B PE=1 SV=1 | --- | --- | --- |
| Q13885 | 15 (0) | 118.95 | --- | --- | | Tubulin beta-2A chain OS=Homo sapiens GN=TUBB2A PE=1 SV=1 | --- | --- | --- |
| P27797 | 15 (15) | 116.58 | 3.04e-004 | 1.22 | | Calreticulin OS=Homo sapiens GN=CALR PE=1 SV=1 | 2.71e+004 | 2.58e+004 | 3.16e+004 |

http://163.15.167.51/QIP/20161229%20L1L5mito.htm

# Combine data with bioinformatics analysis software (IPA)



# Thank you for your attention!!!