

METHODS IN PROTEOMICS



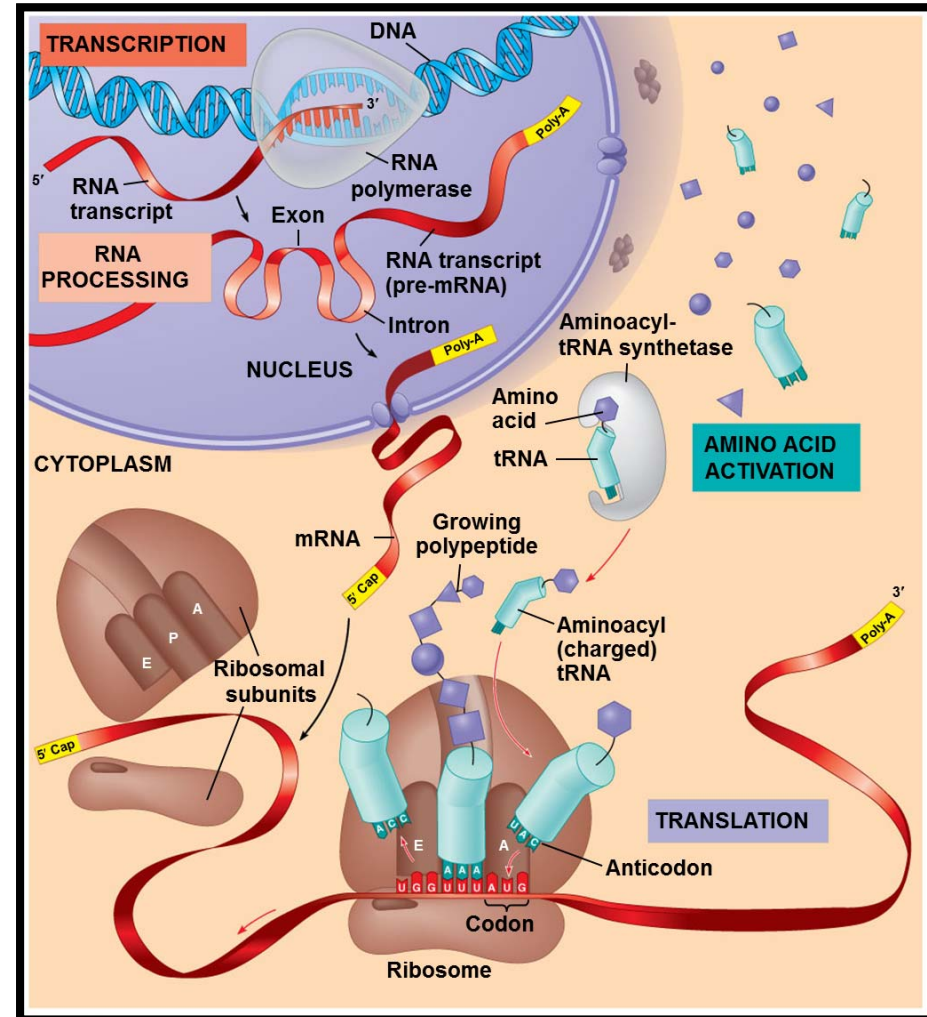
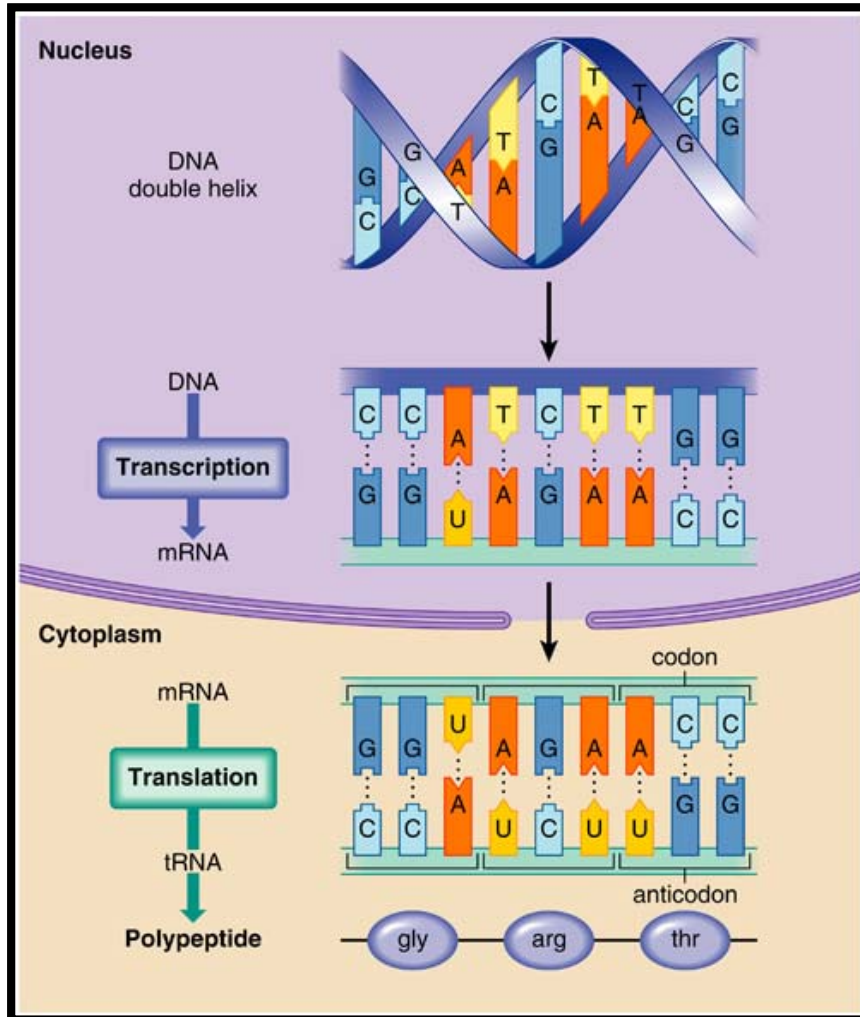
Lecturer: Ming-Hsien Tsai, PHD.

Assistant researcher

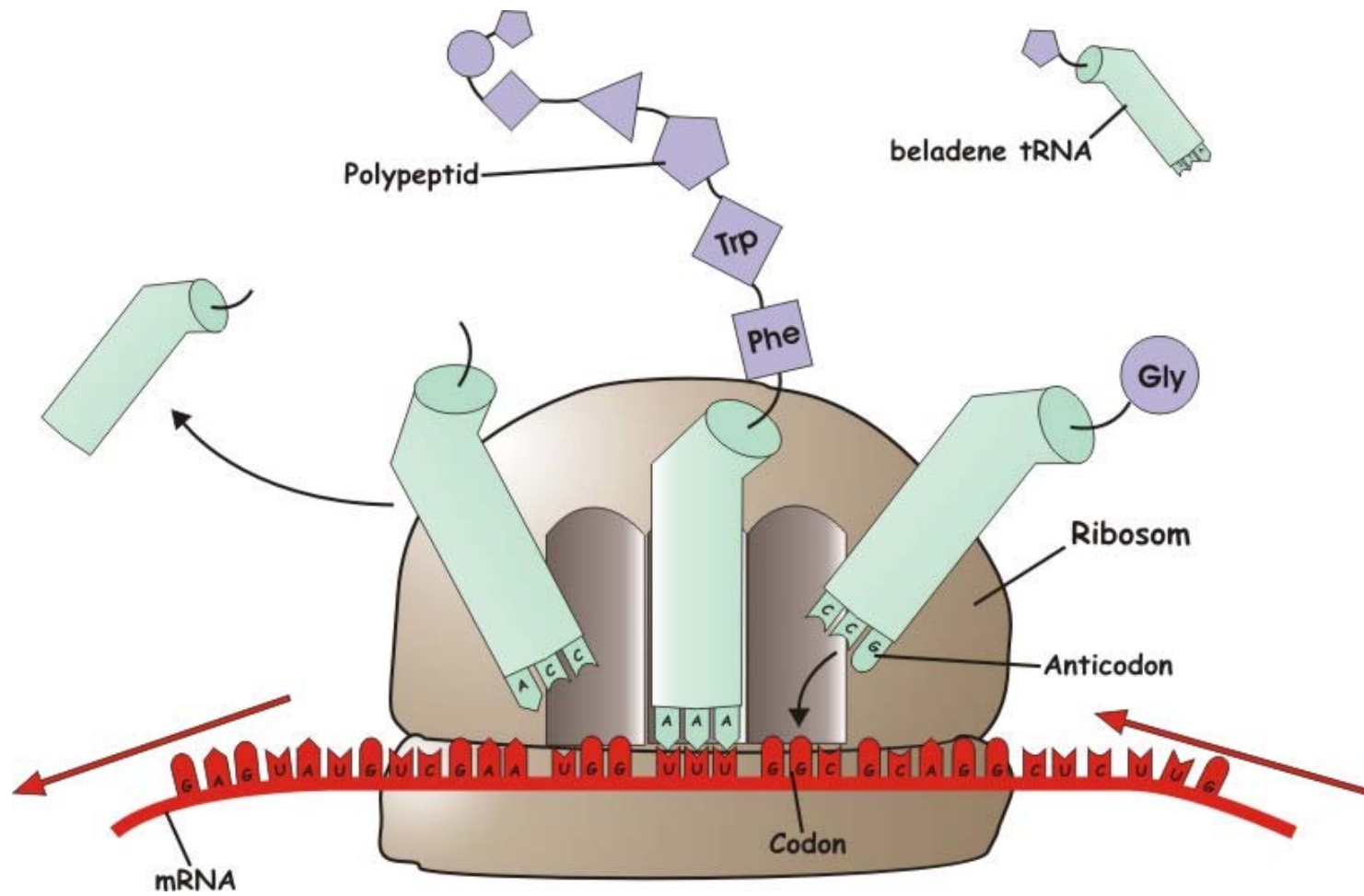
Center for Lipid bioscience, KMUH

Lipid Science and Aging Research Center, KMU

DNA → RNA → Protein

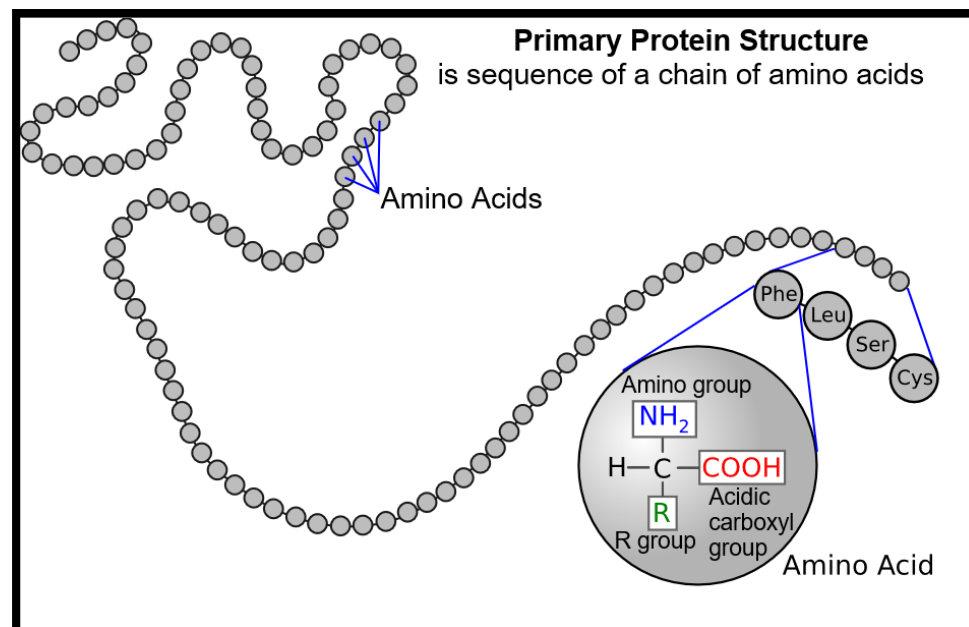


Translation

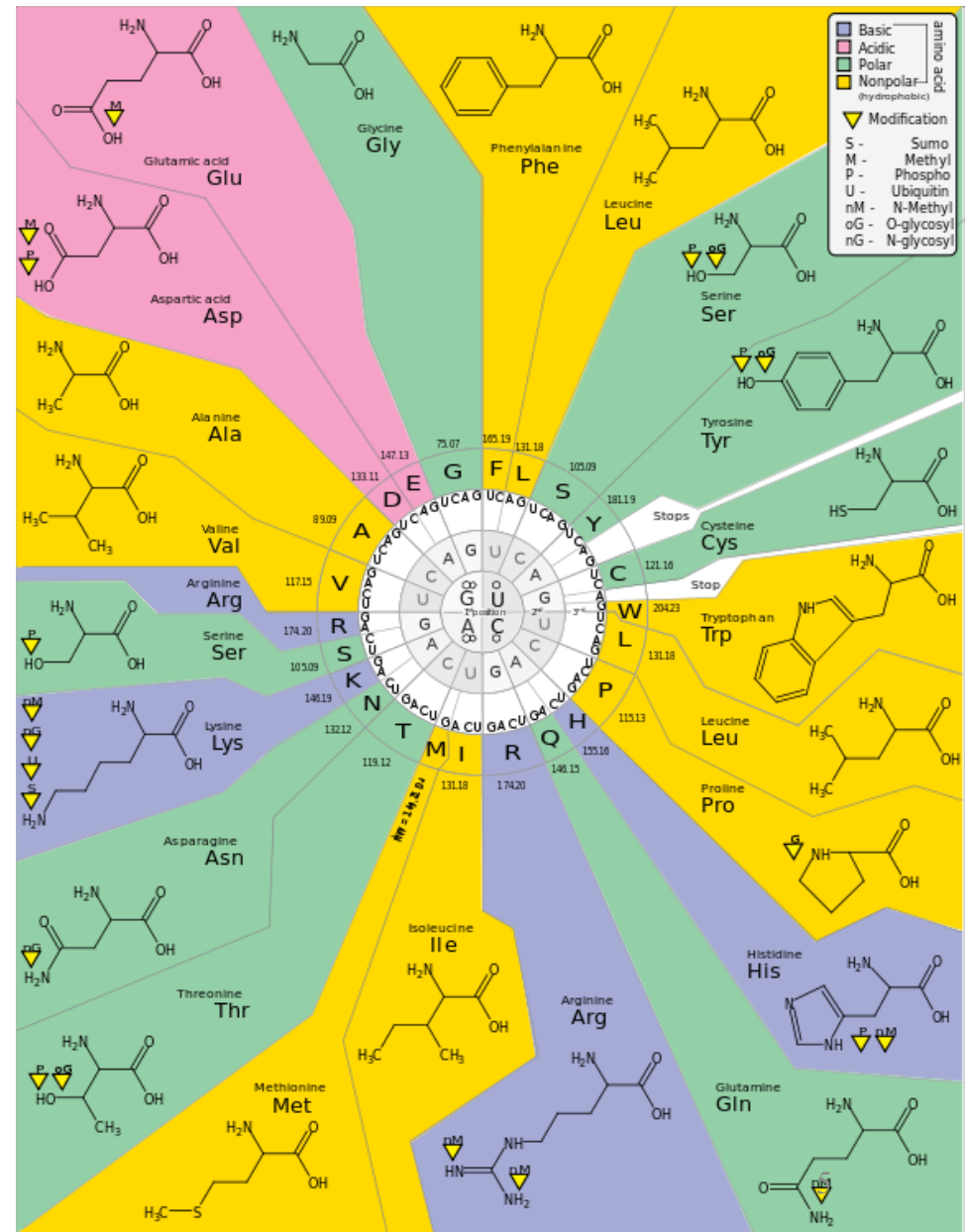
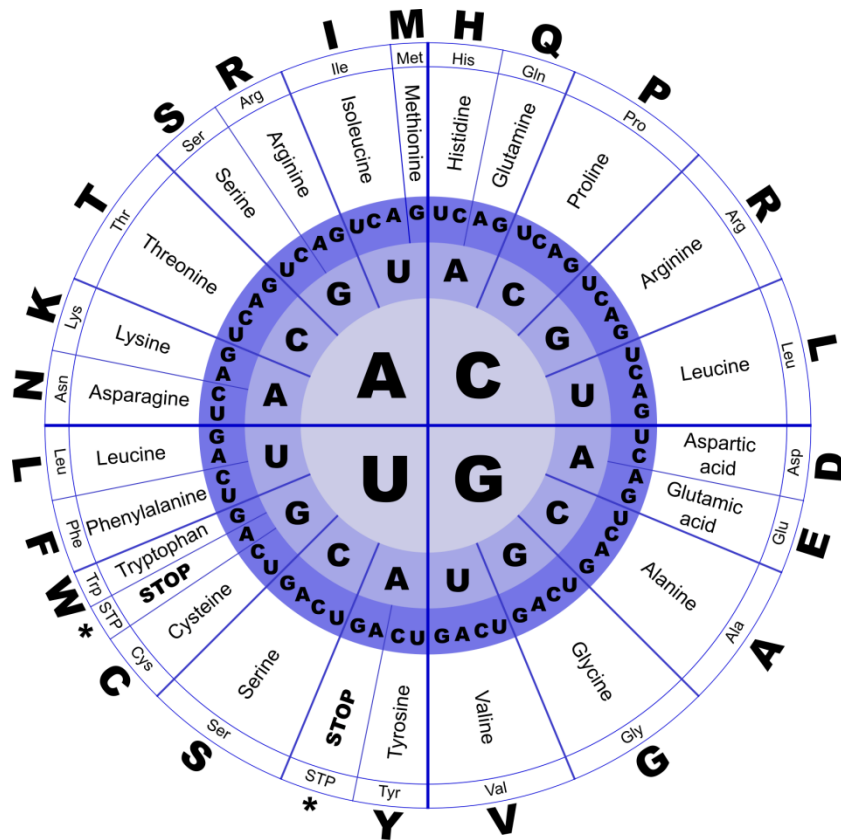


How does DNA act as A code?

- 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

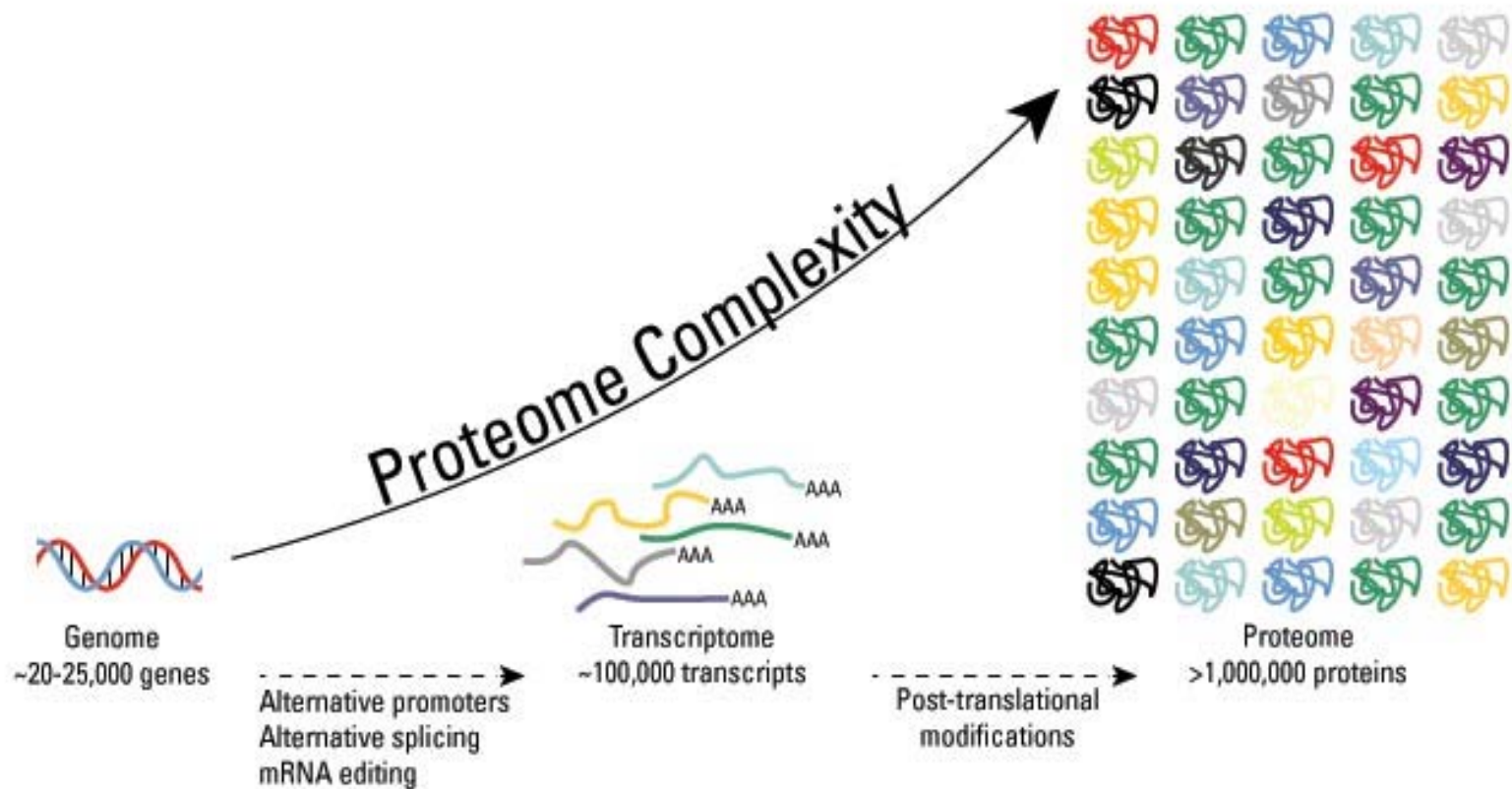


Amino Acid codon



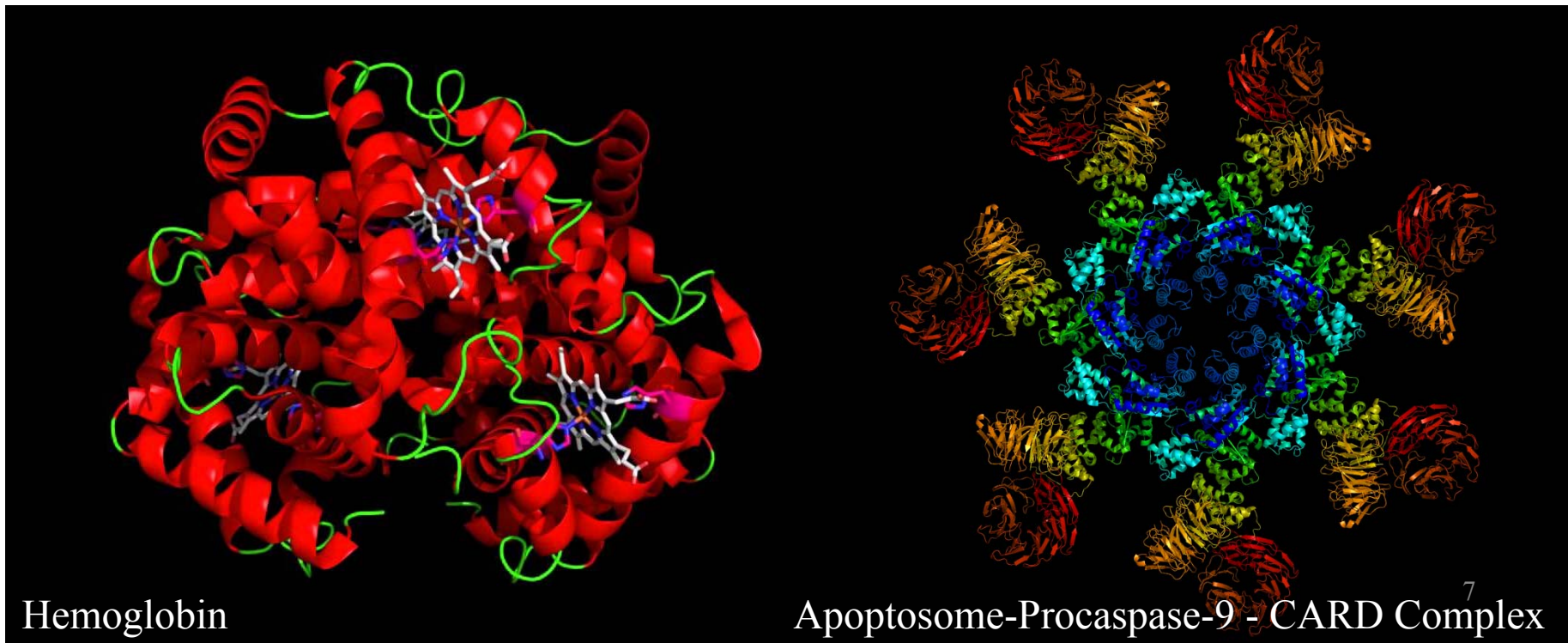
But in Fact...

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

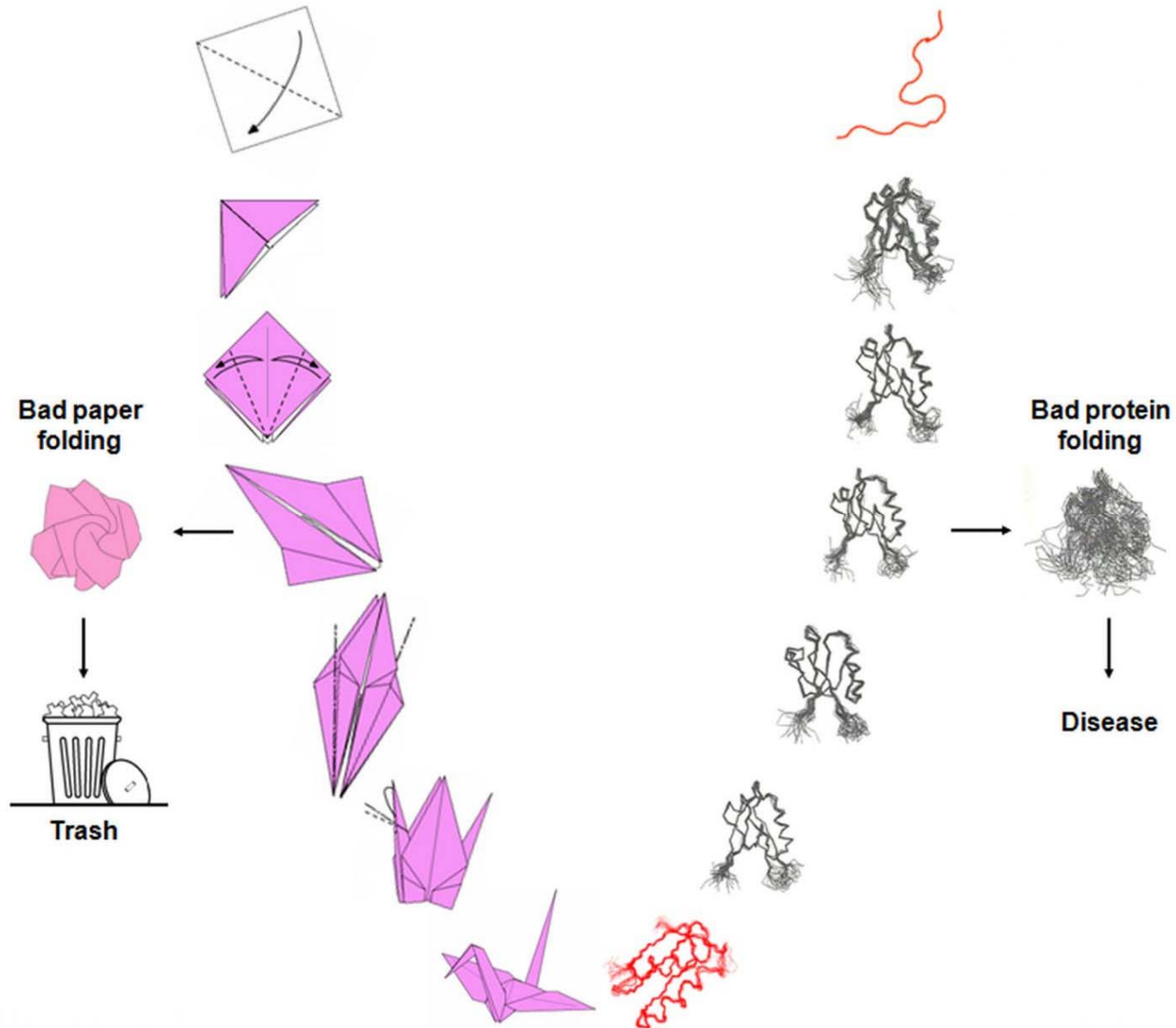


Proteomics and the proteome

- 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.

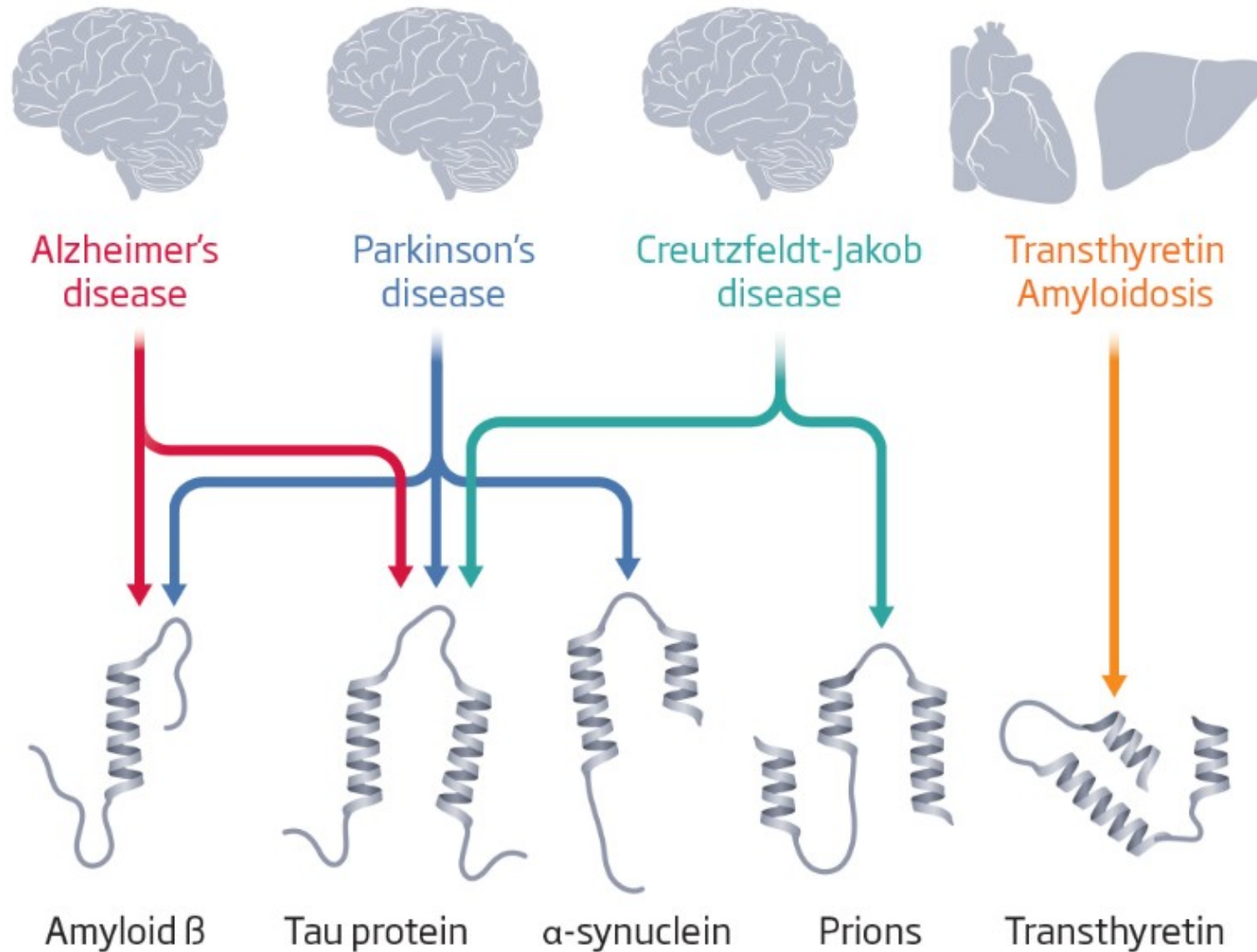


Bad protein folding and Disease



Proteins gone rogue

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



SOURCE: NEUROPHAGE PHARMACEUTICALS

Challenges

Avoiding contamination

Abundant proteins

Reliable quantitation

Patients plasma (comorbidity)

Maximising number of
confidently assigned
proteins

Experimental design

Throughput

Normalisation

Large data files

Protein degradation

What to do with
low confidence
proteins

Maintaining system
performance over a long
period of analyses

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 (PO_3^-) to a molecule. In biology, phosphorylation and its counterpart, dephosphorylation, are critical for many cellular processes.

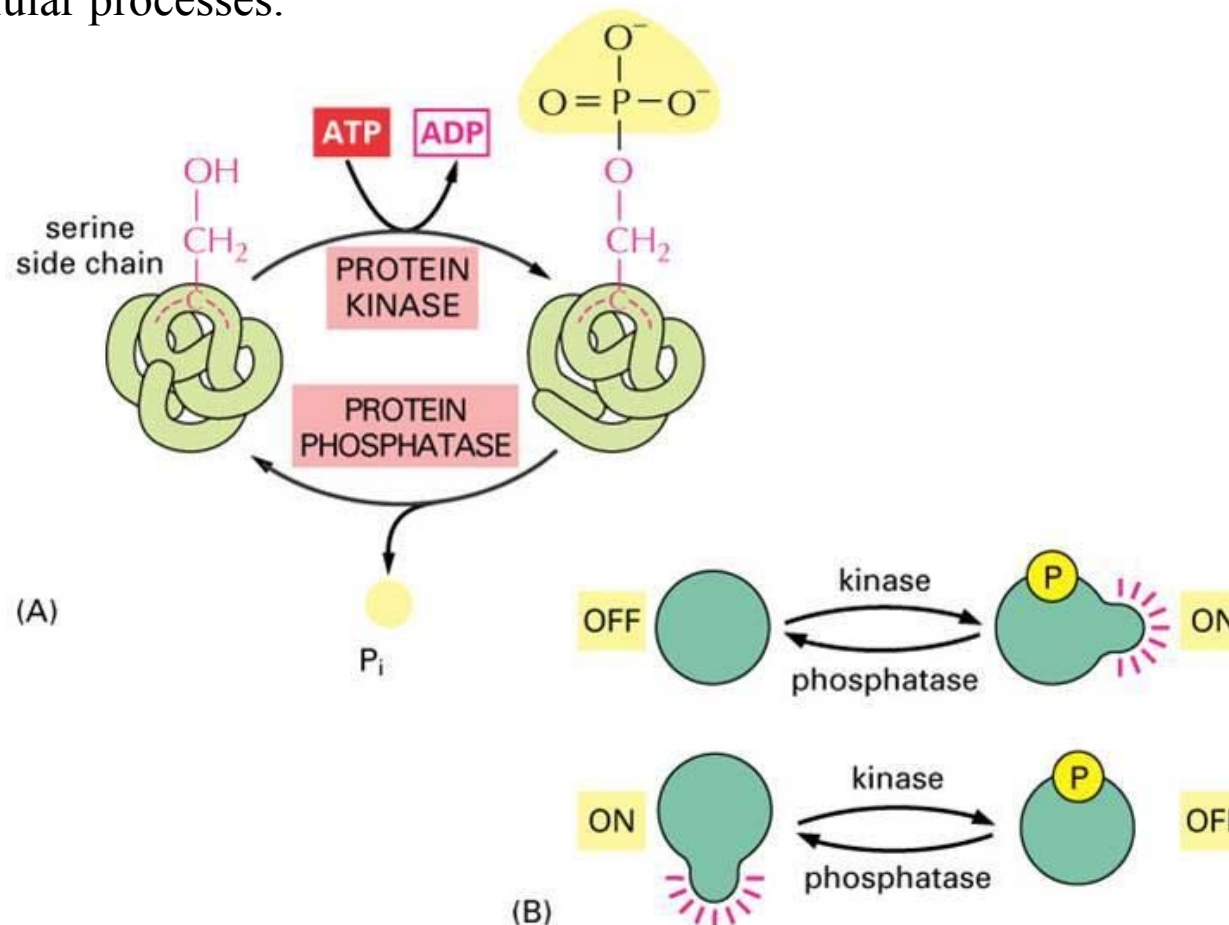
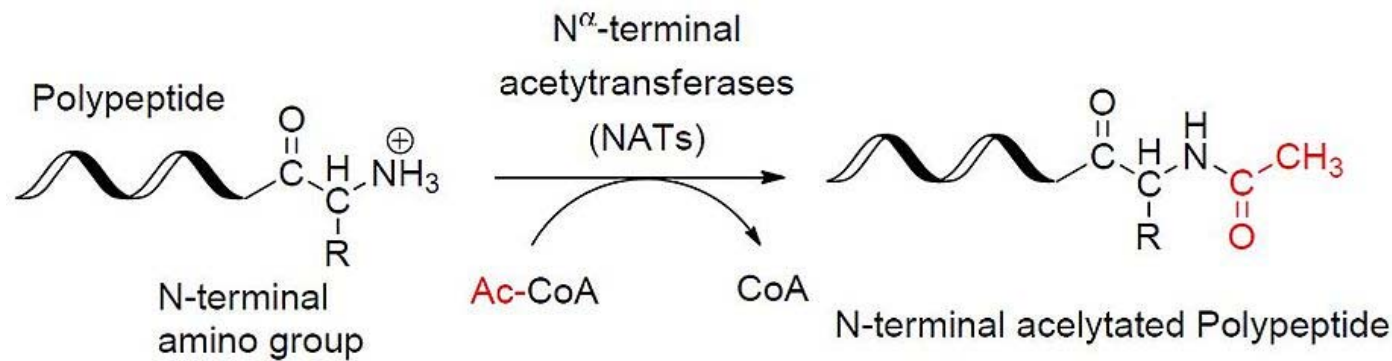
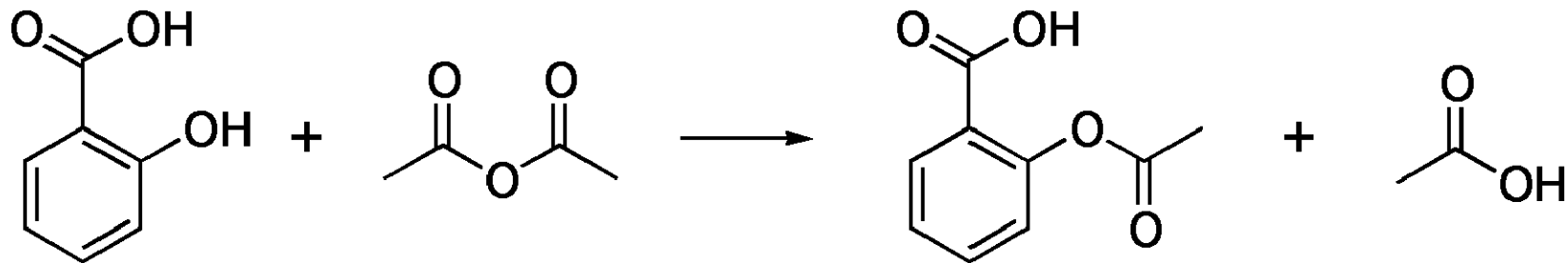


Figure 4-41 Essential Cell Biology, 2/e. (© 2004 Garland Science)

Acetylation

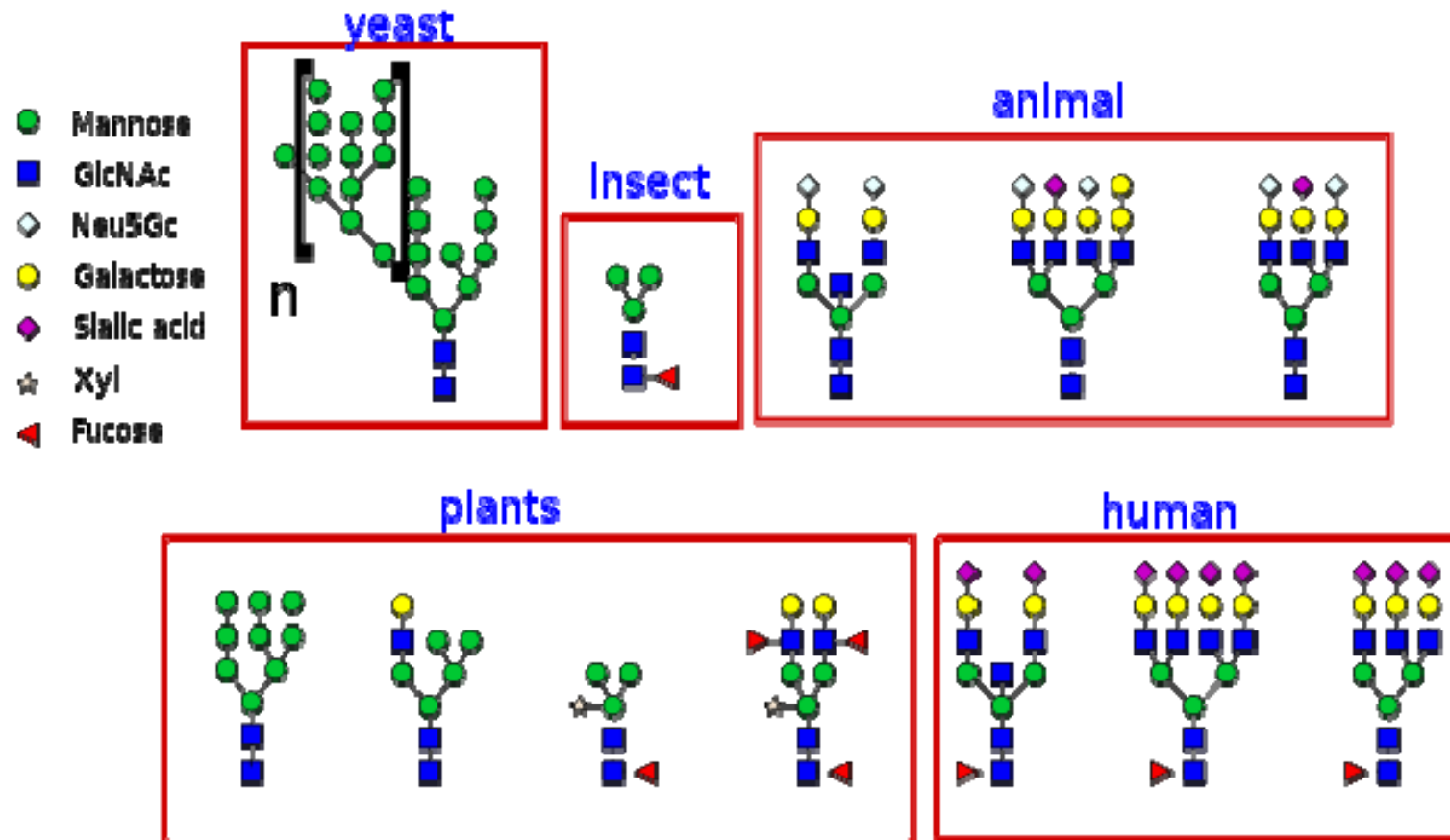
Acetylation (or in IUPAC nomenclature ethanoylation) describes a reaction that introduces an acetyl functional group into a chemical compound. Deacetylation is the removal of an acetyl group.



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**

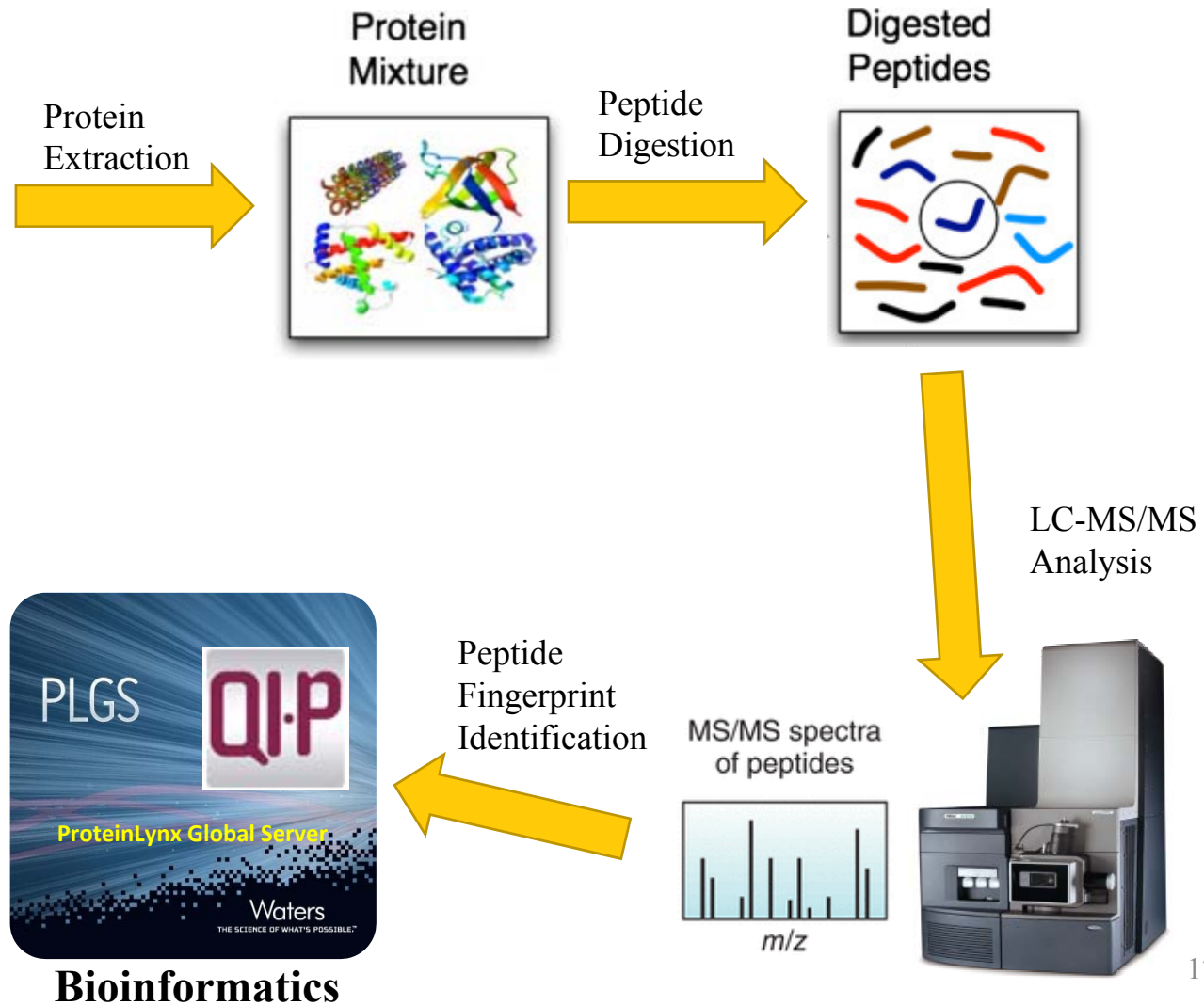
DNA methylation in vertebrates typically occurs at CpG sites (cytosine-phosphate-guanine sites—that is, where a cytosine is directly followed by a guanine in the DNA sequence).
- **Protein methylation**

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

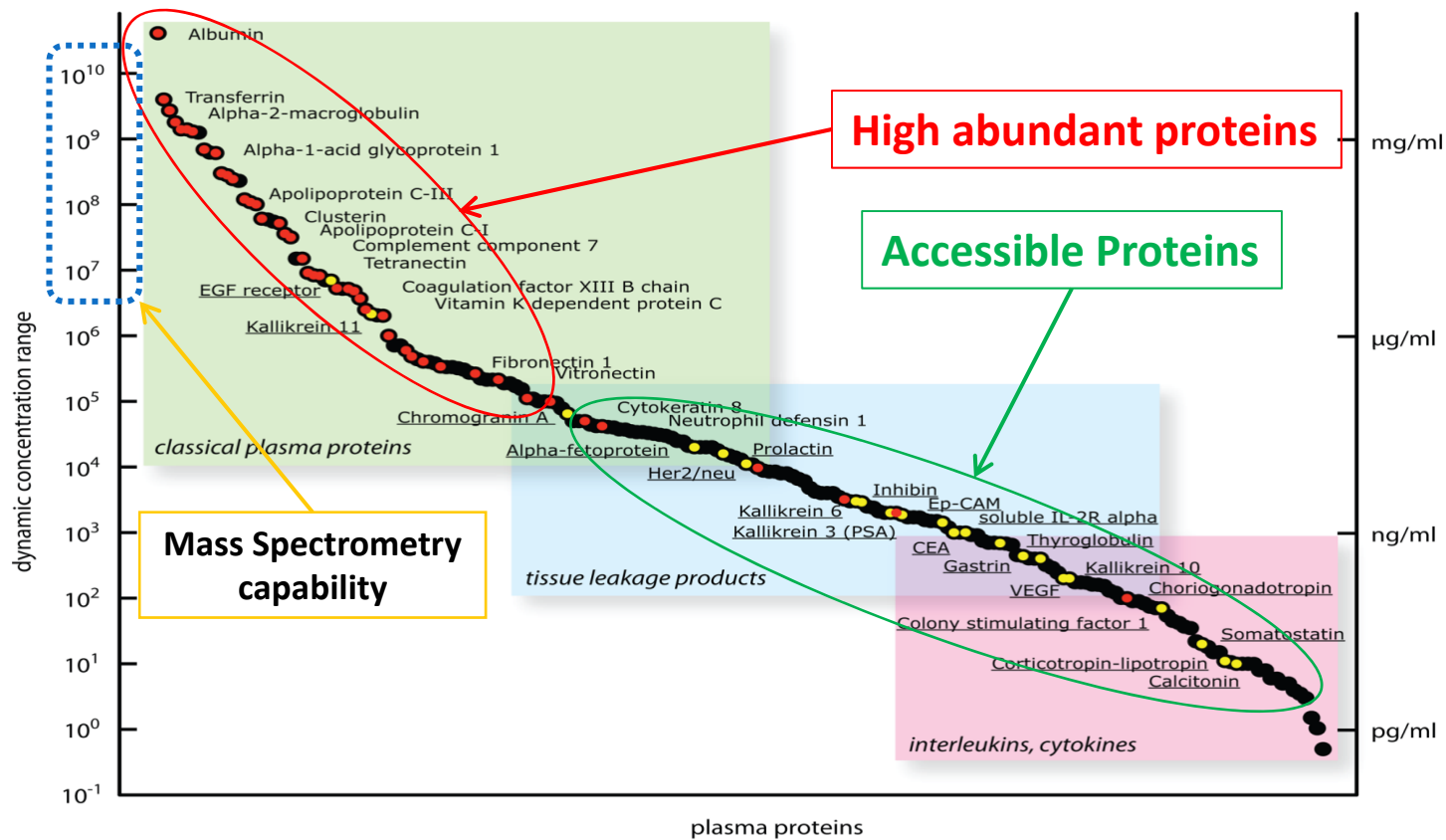
Workflow of Proteomics



Cell Culture
Clinical Samples
Animal Tissue

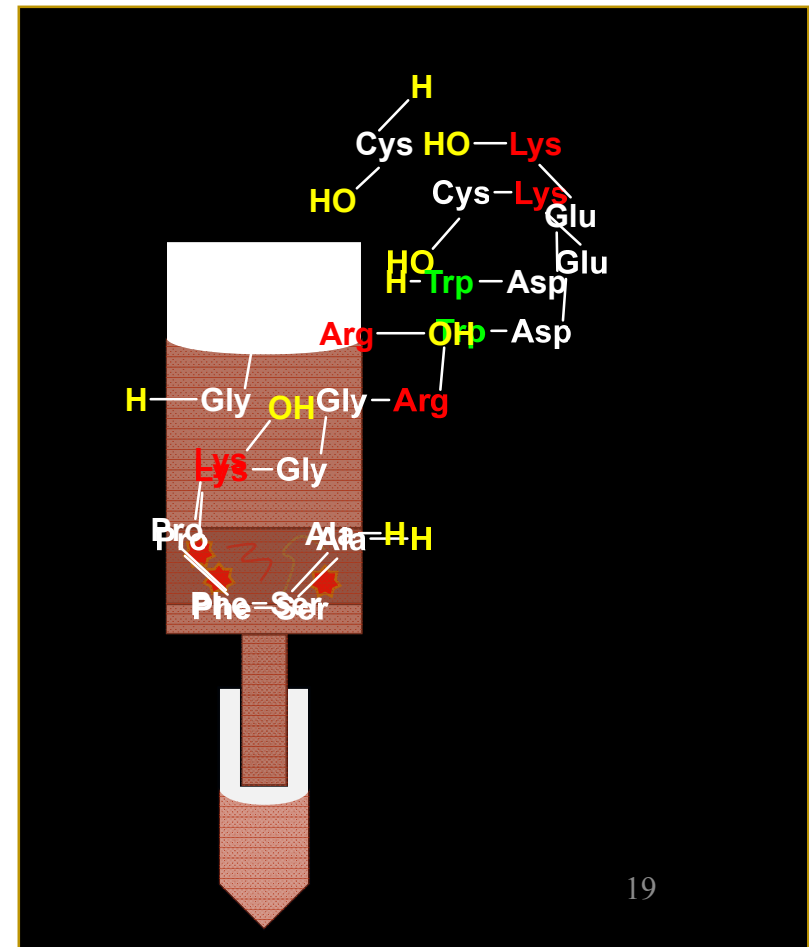


Sample Preparation (Cont.)



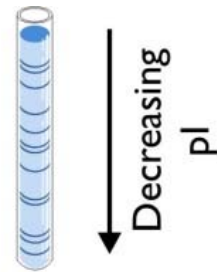
Techniques

- 2D-Gel electrophoresis, Sample enrichment (Beads, Affinity Matrix)
- BCA Assay
- Shot gun proteomics (Tryptic digestion)
- Solid-phase extraction

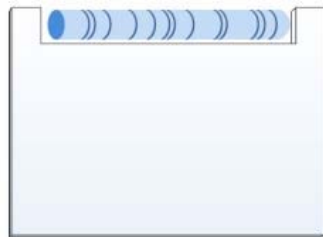


2D-Gel electrophoresis

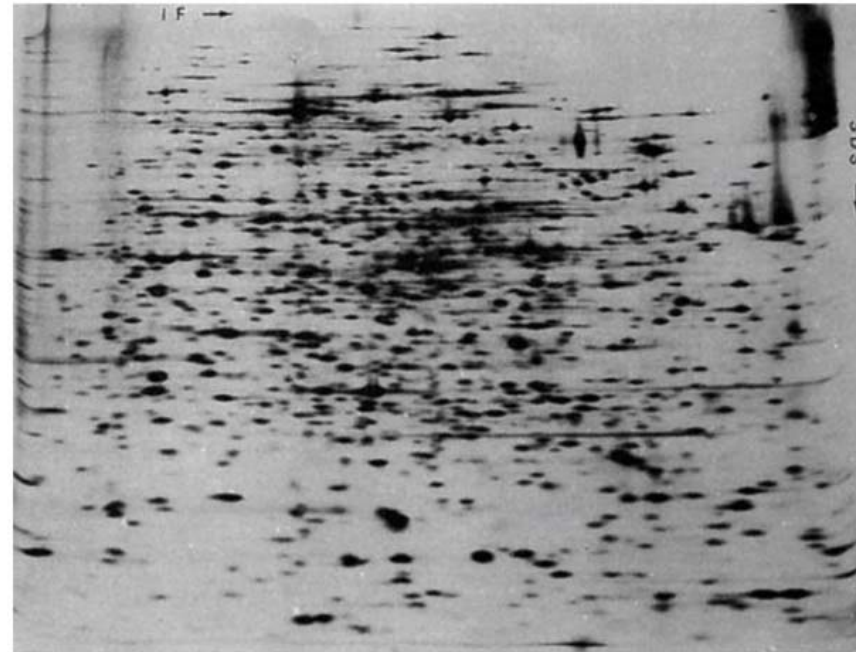
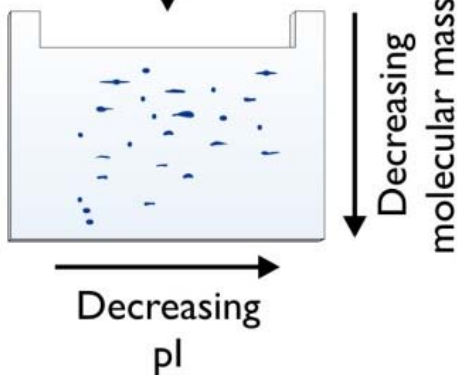
1. Separation of proteins by pI value



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

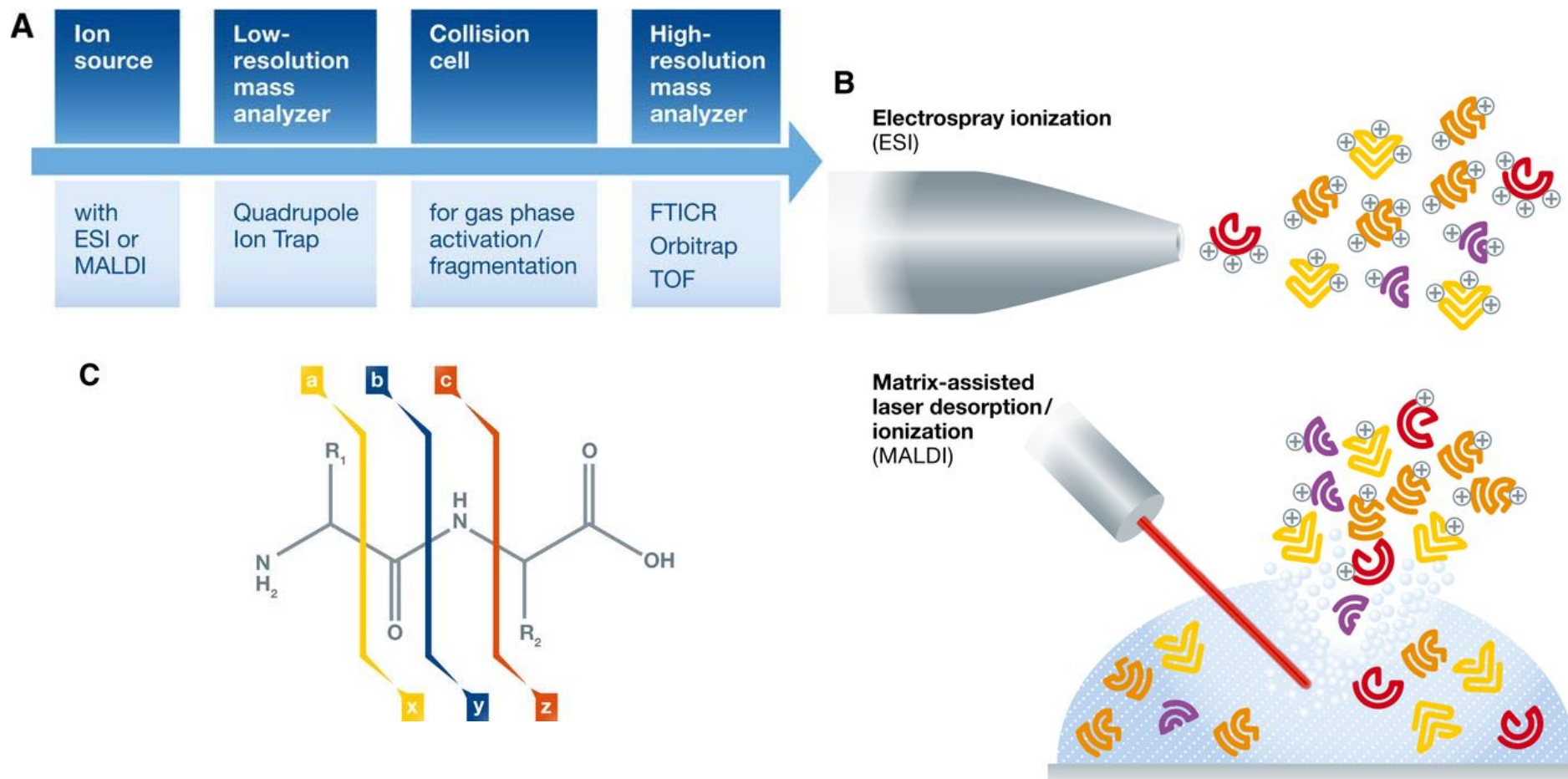


3. Separating the proteins by molecular mass with SDS PAGE



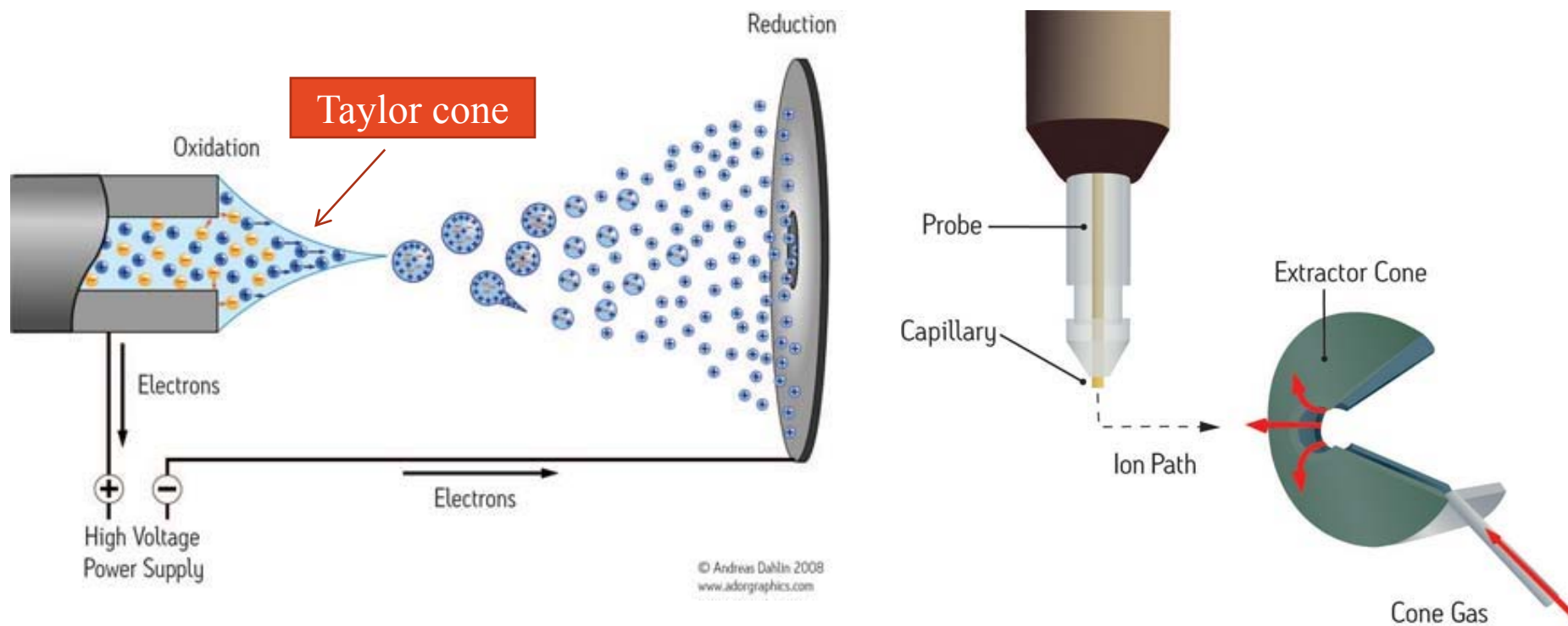
Approximately 1000 *E. coli* proteins on a single 2D gel

Basic principles of biomolecular MS



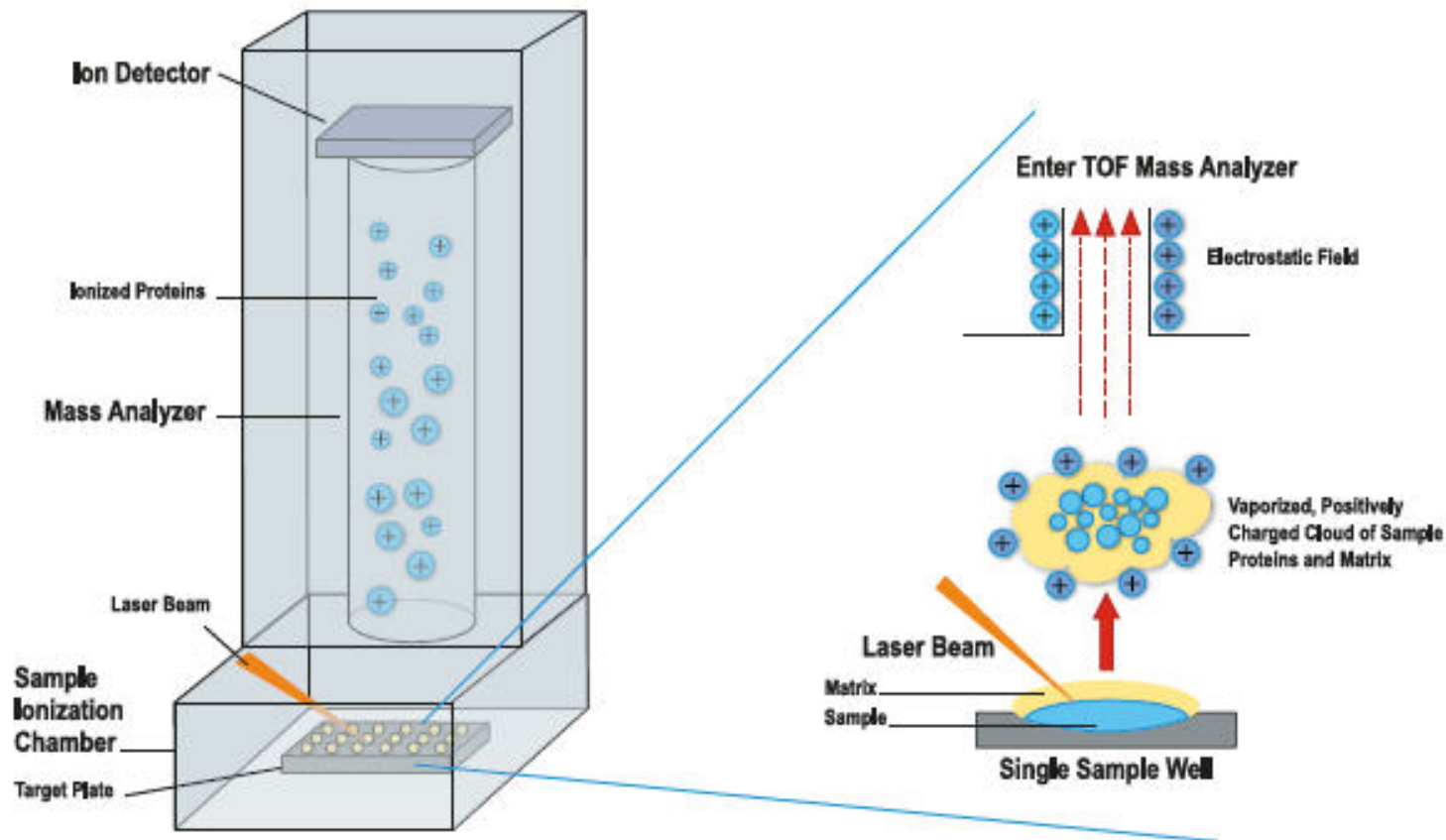
Principle of Electrospray Ionization (ESI)

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)

- **Quadrupole**

- High Sensitivity, acceptable mass accuracy and resolution
- Easily coupled to chromatography

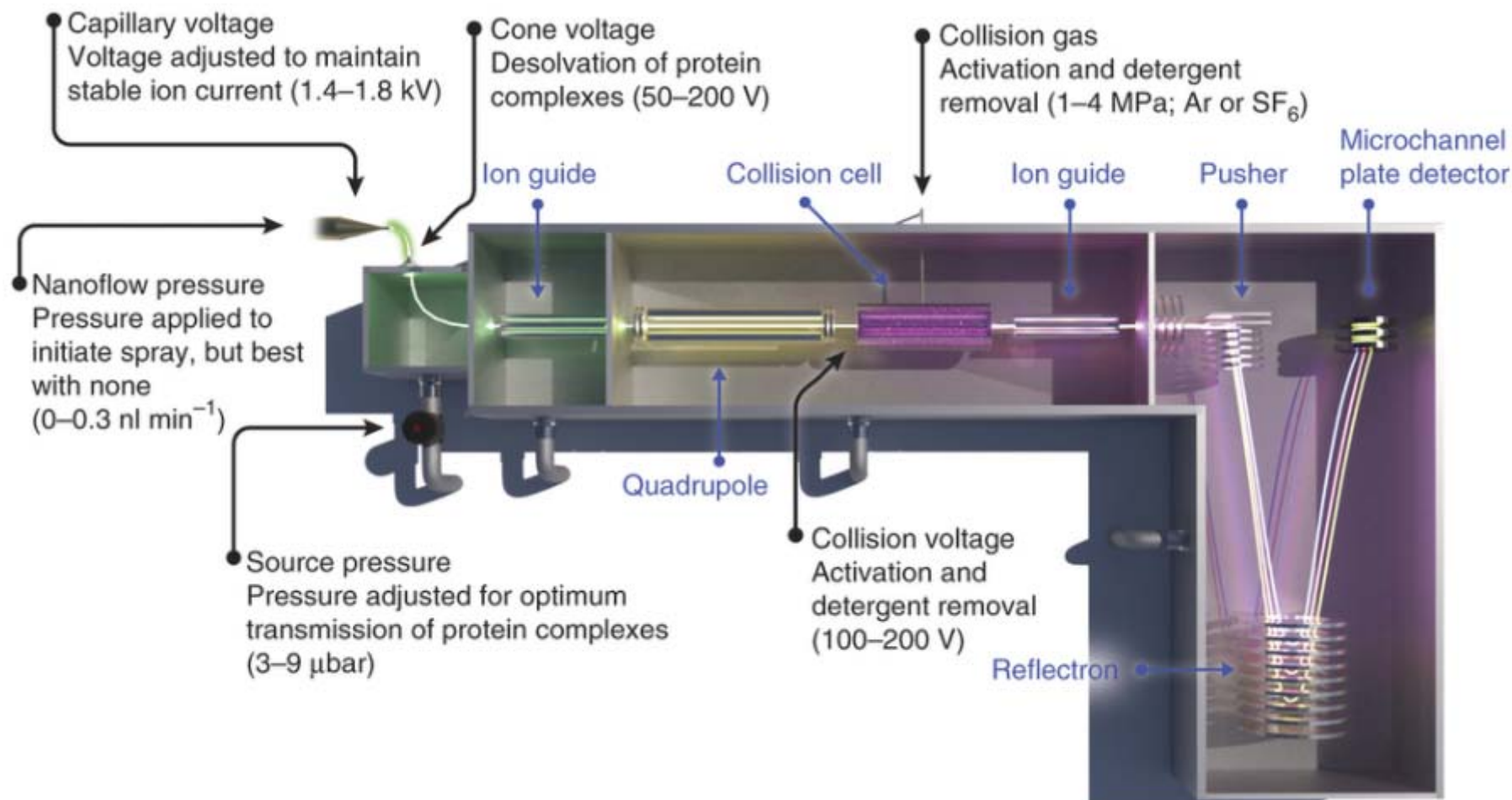
- **Time of Flight**

- High Sensitivity, high mass accuracy, high resolution
- Limited to small m/z ratios
- Not easily coupled to chromatography
- Easily coupled to MALDI

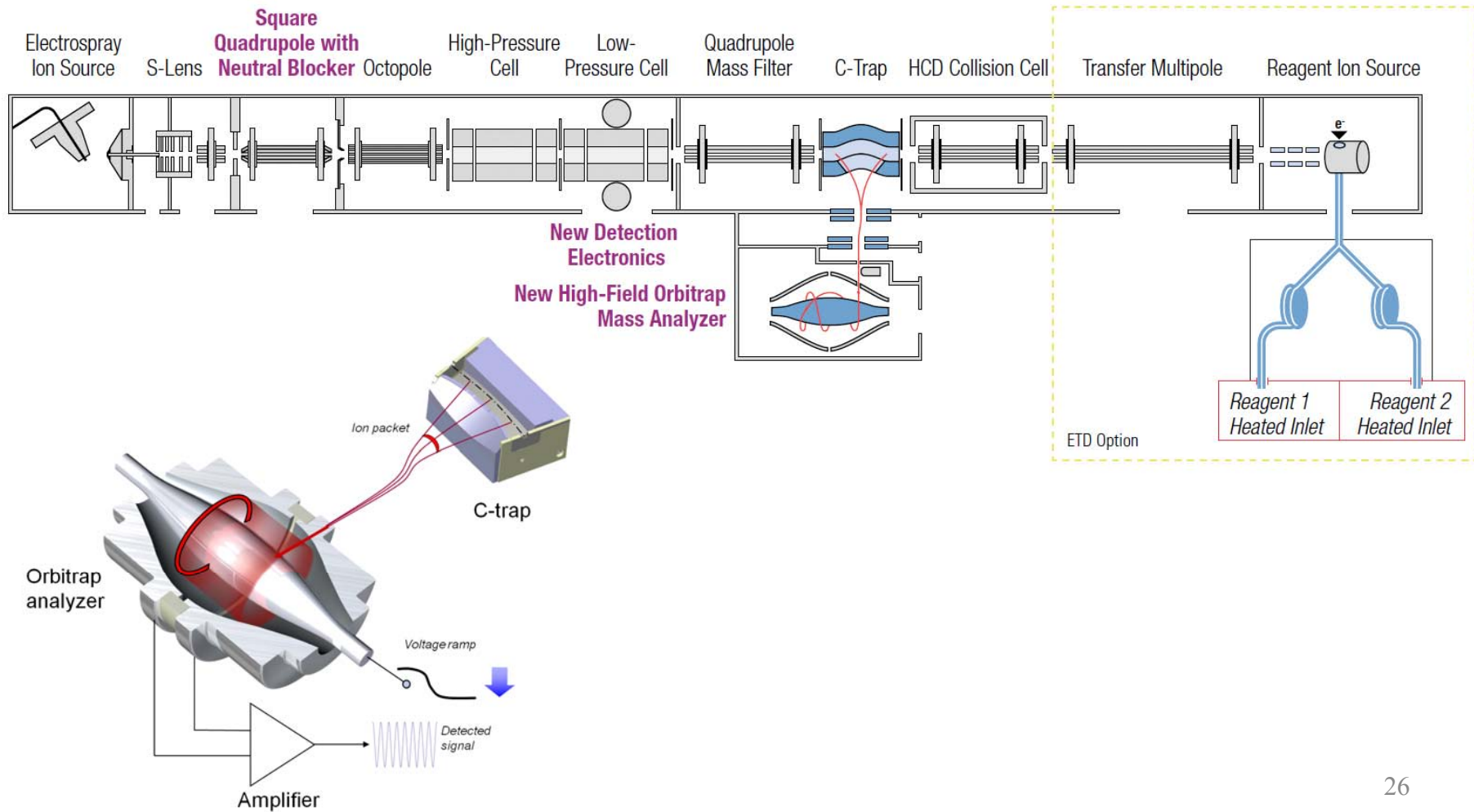
- **Ion Trap**

- High Sensitivity
- Low mass accuracy and resolution

Schematic of a Q-TOF mass spectrometer

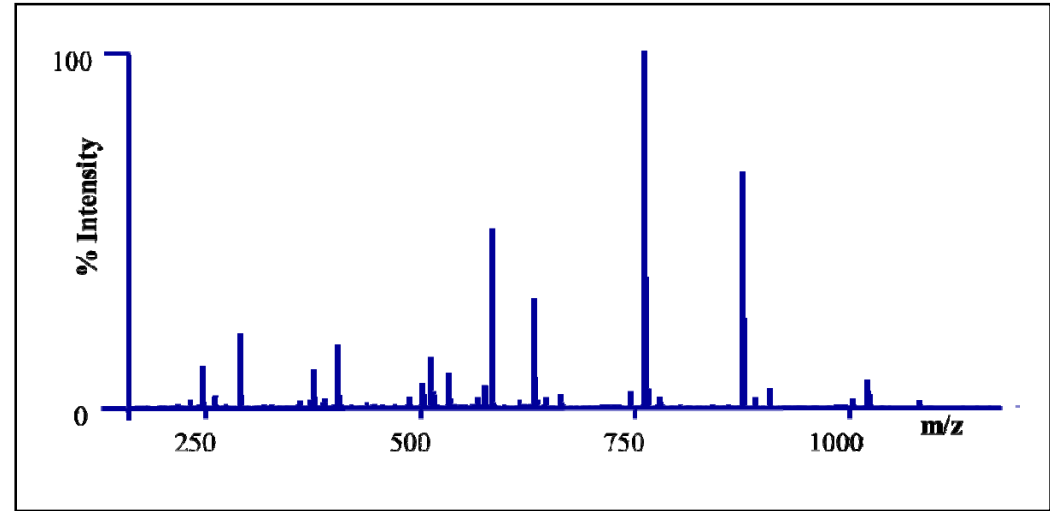
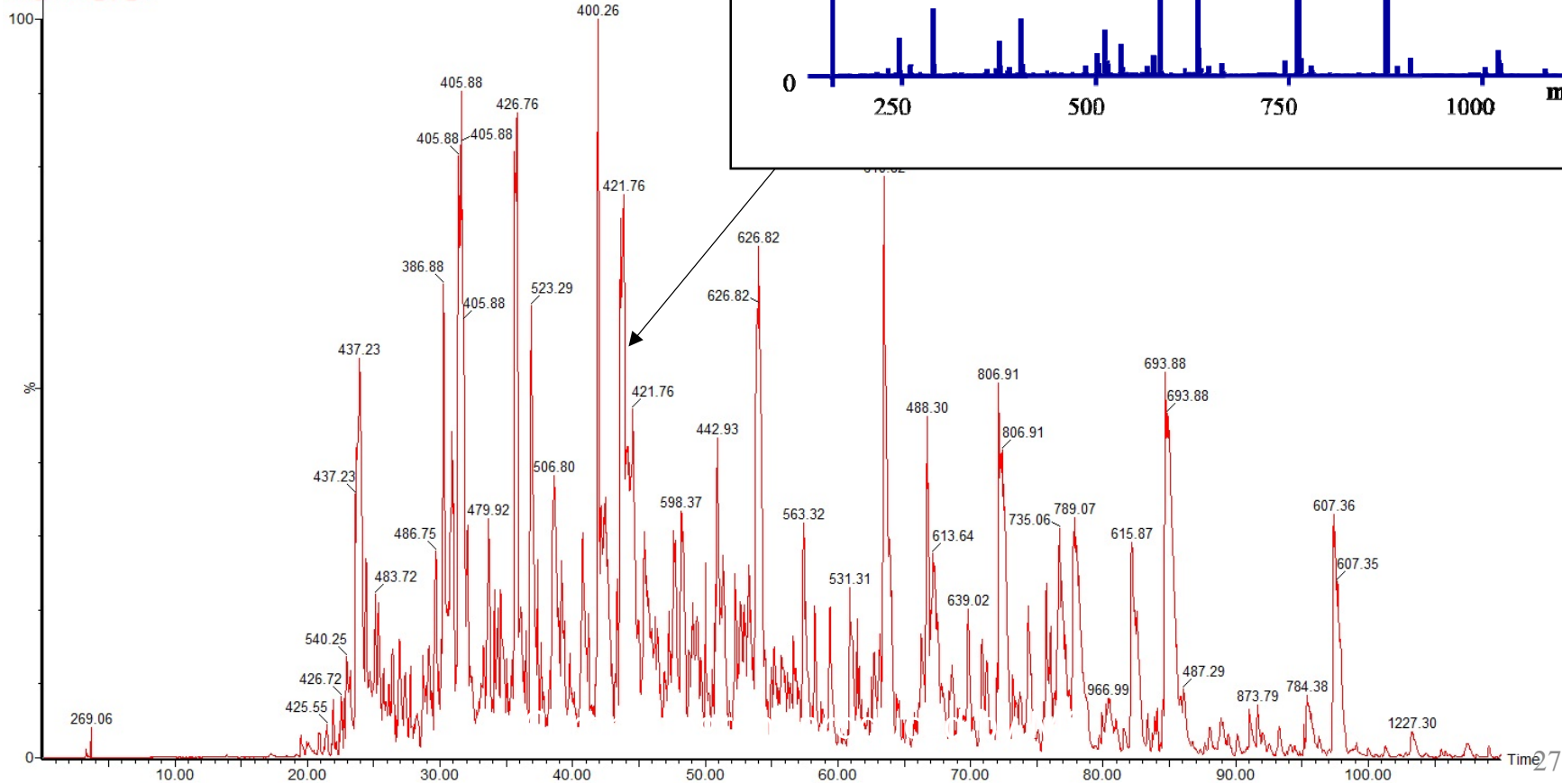


Schematic of a orbitrap mass spectrometer



Chromatography and Mass Spectrum

Triplicate evaluation 500ng on column
AMH170311_3pl_01



Equipment in CLB, KMUH



Waters UPLC



Low Flow ESI



Waters G2 qTOF



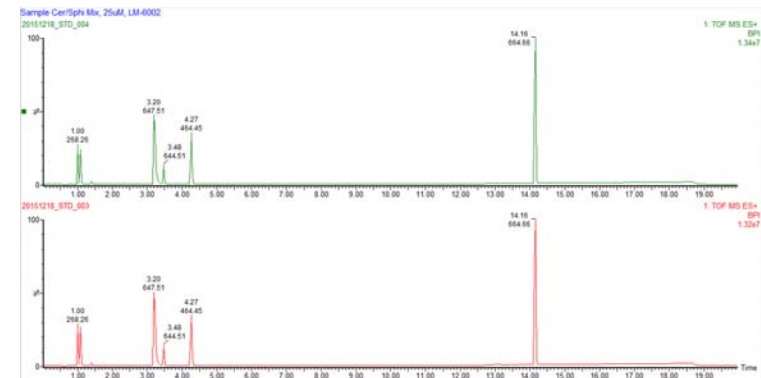
Waters M-class UPLC



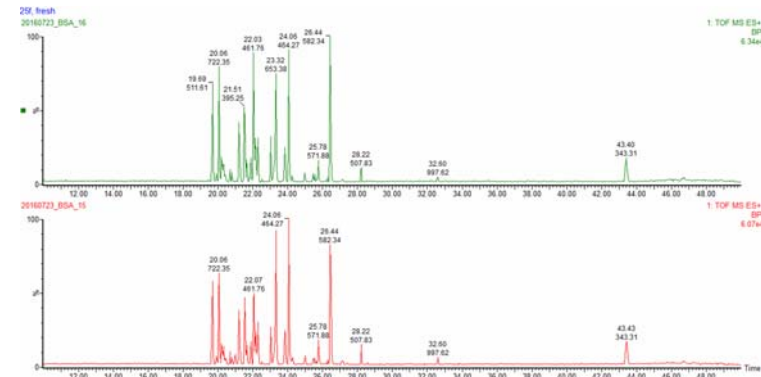
NanoFlow ESI



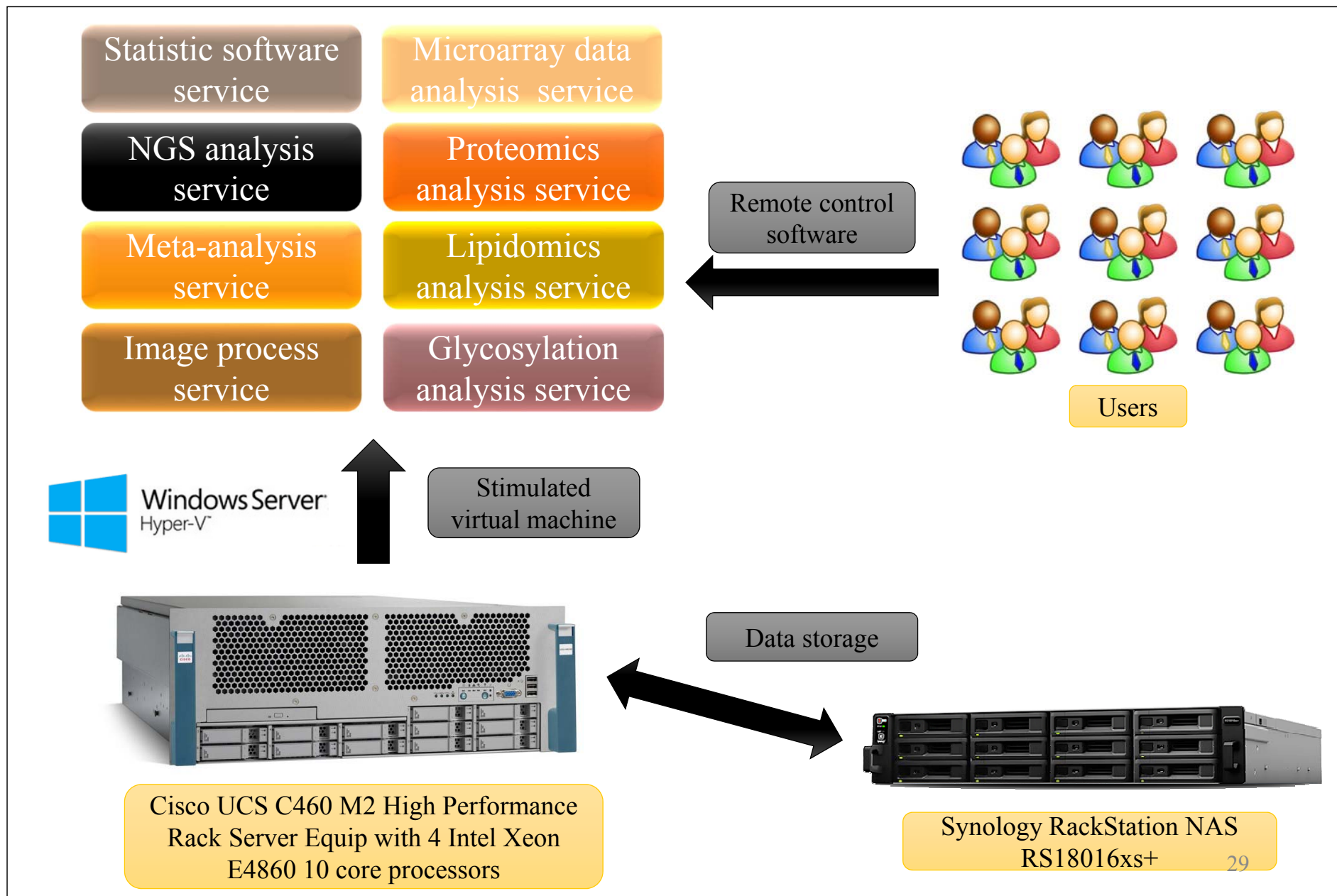
Lipid Identification



Peptide Identification



Bioinformatics server in CLB, KMHU



Protein Database



The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

UniProtKB
UniProt Knowledgebase

Swiss-Prot (554,241)
Manually annotated and reviewed.

TrEMBL (84,827,567)
Automatically annotated and not reviewed.

UniRef
Sequence clusters

UniParc
Sequence archive

Proteomes

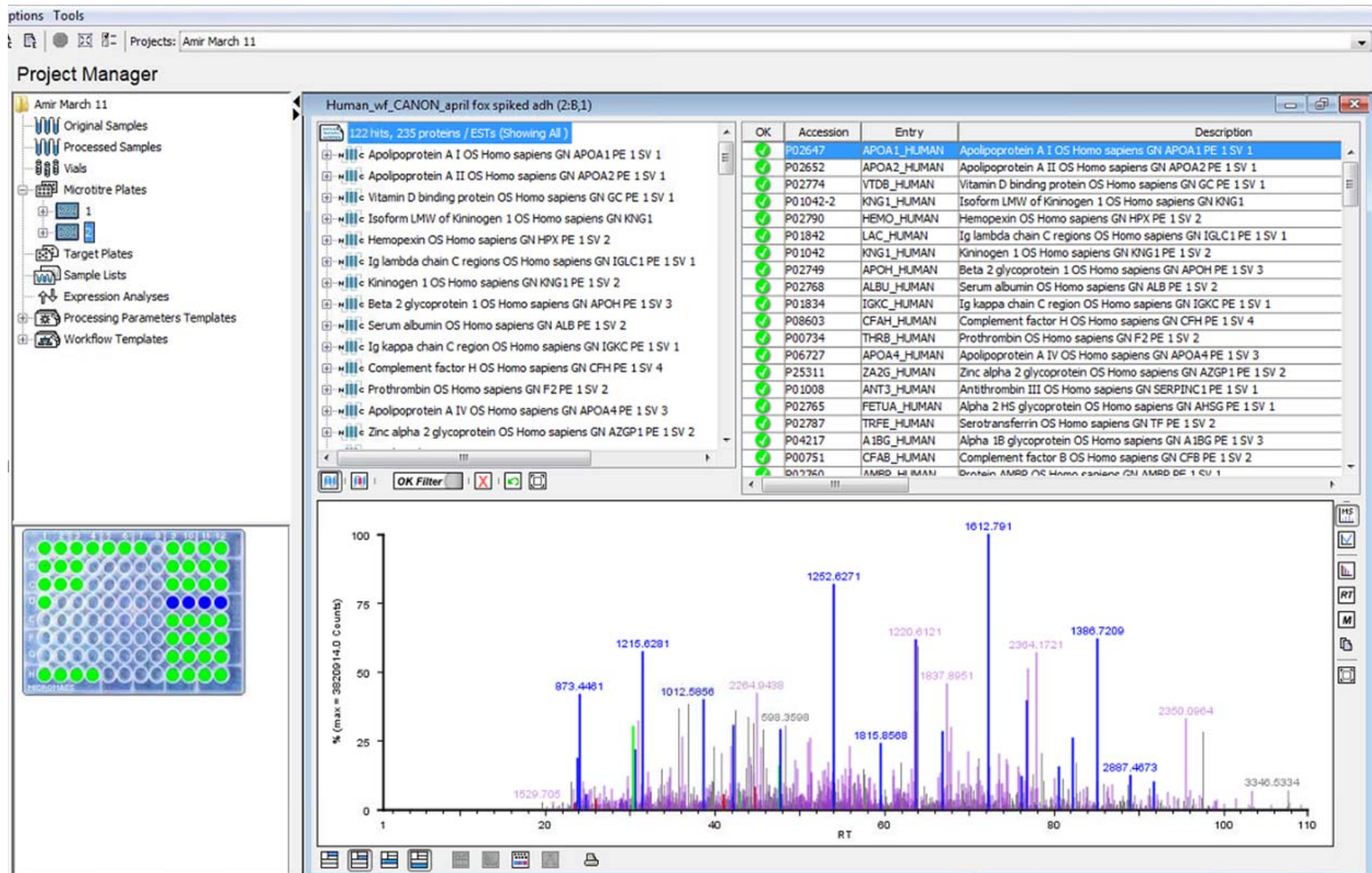
Supporting data

- Literature citations
- Taxonomy
- Subcellular locations
- Cross-ref. databases
- Diseases
- Keywords

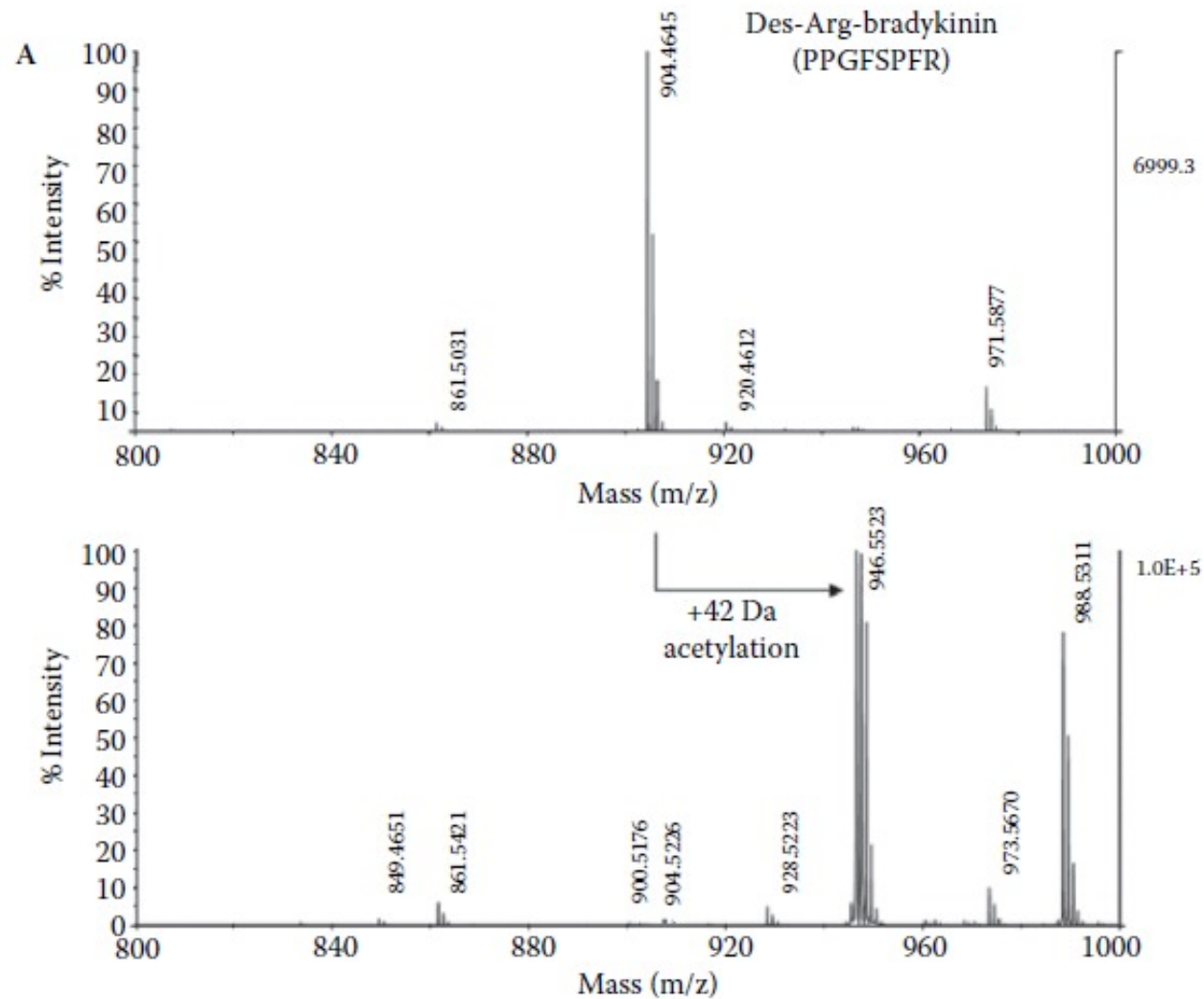
News

- Forthcoming changes
Planned changes for UniProt
- UniProt release 2017_04
Death (by insulin) in paradise
- UniProt release 2017_03
Viral Short Message Service: peptide texting guides the outcome of infection
- UniProt release 2017_02
- News archive

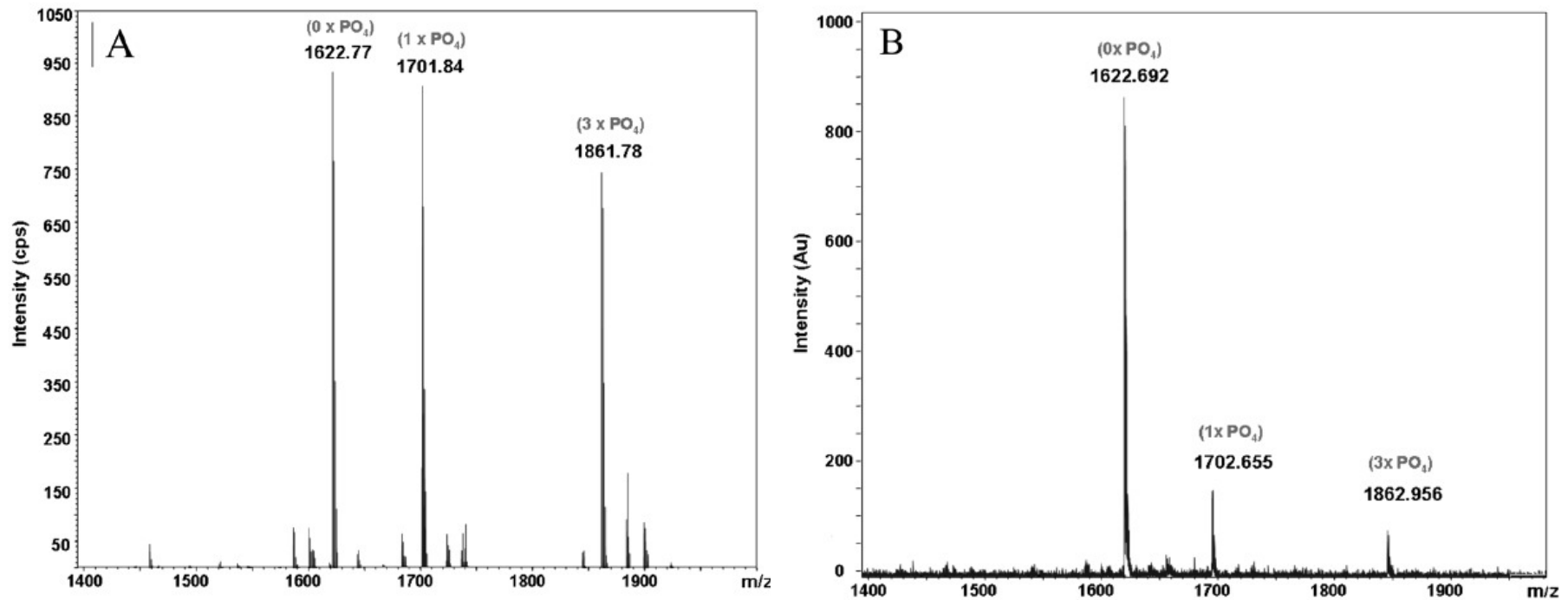
Identity MS^E Results from PLGS



Acetylation in MSMS

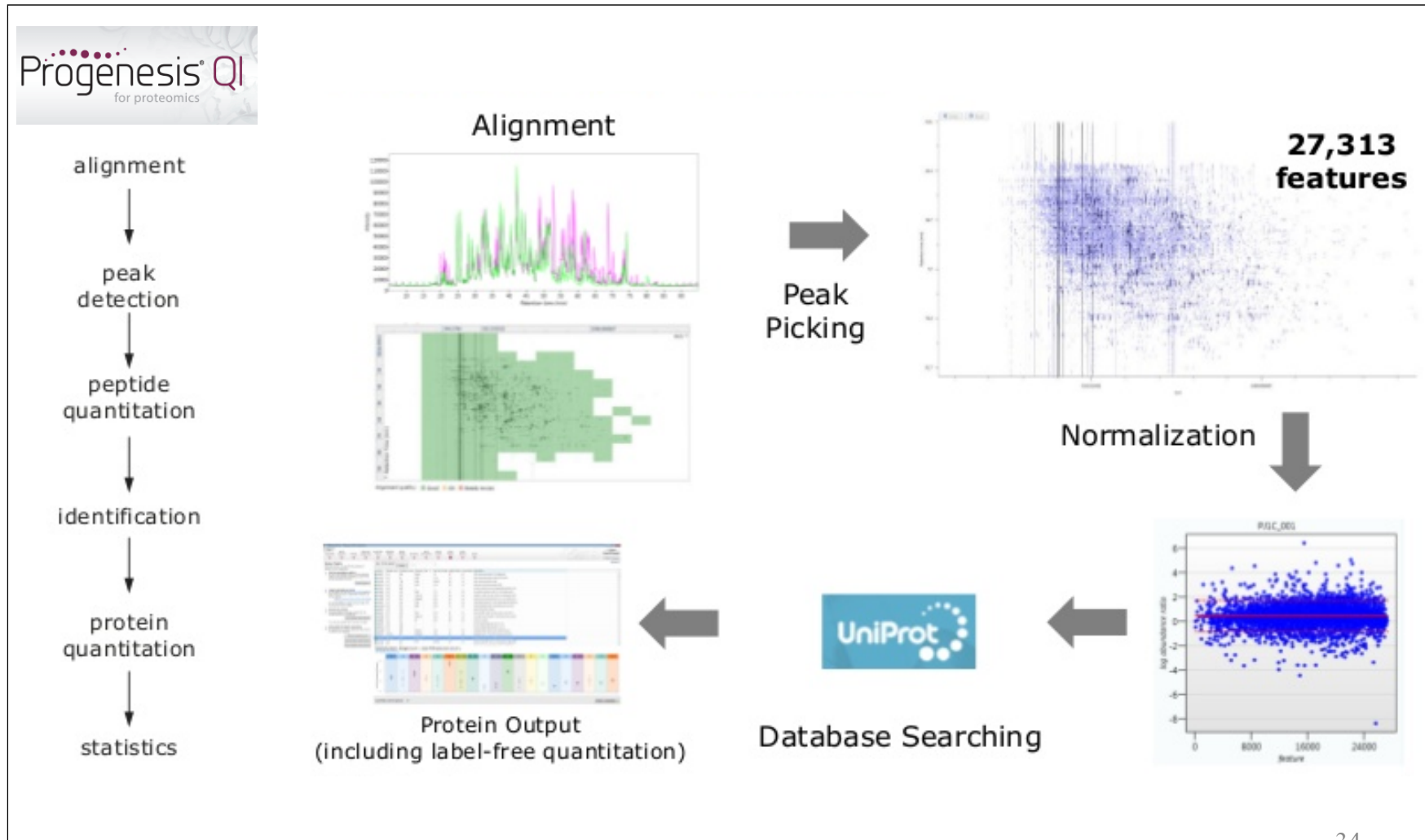


Phosphorylation in MSMS



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

The screenshot displays the Progenesis QI software interface. The top menu bar includes 'File', 'Review Alignment', 'Filtering', 'Experiment Design Setup', 'Review Peak Picking', 'Peptide Ion Statistics', 'Identify Peptides', 'QC Metrics', 'Refine Identifications', 'Resolve Conflicts', 'Review Proteins', 'Protein Statistics', and 'Report'. The 'Import Data' section on the left contains two main steps: 1. 'Select your run data' with a dropdown menu set to 'Waters (.raw)' and a 'Browse...' button; 2. 'Perform automatic processing' with a 'Restart automatic processing' button. Below these steps is a search bar for 'Imported runs' and a grid of 24 small thumbnail images representing individual runs. The main window displays a large data plot titled '20170405_SP01_01' with 'Retention Time (min)' on the y-axis (ranging from 15 to 40) and 'm/z' on the x-axis (ranging from 500 to 1500). The plot shows a dense cloud of points with a prominent horizontal band of points around 25-30 minutes retention time. The number '35' is visible in the bottom right corner of the plot area, and there are zoom controls at the bottom right of the interface.

Identify peptides fingerprint

The screenshot displays the Progenesis Q1 software interface for peptide identification. The main window is titled "20170418_mito_30min - Progenesis Q1 for proteomics". The "Identify Peptides" section is active, showing a list of peptide ions (3116 identified) and a fragment ion plot for the first ion.

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

Peak 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
-----------	--------------	-----------------	------------------	-------	-------	-----	----------	----------	------------	-------------------------------	-----------------------------	---------

Fragment ions for:

The fragment ion plot shows the m/z of the fragment ions versus their retention time (min). The x-axis ranges from 24 to 28 minutes, and the y-axis ranges from 240 to 280 m/z. A prominent peak is observed at m/z 615.4036, which corresponds to the parent ion. Other peaks are visible at lower m/z values, indicating fragmentation of the peptide.

No identification selected
Select an identification above to view its fragment ions

PTMs Search

Identify Peptides

Select your peptide identification method:
MS¹ 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

2 Search for identifications
Identifications will be displayed as they appear in the database. Features automatically identified in the search are highlighted in yellow.
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
No identification selected												

Select an identification above to view its fragment ions for:

- 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

No identification selected
Select an identification above to view its fragment ions for:

Review proteins

File Import Data Alignment Filtering Experiment Design Setup Review Peak Picking Peptide Ion Statistics Identify Peptides QC Metrics Refine Identifications Resolve Conflicts Review Proteins Protein Statistics Report

nonlinear
A Waters Company

Review Proteins

Using this screen, you can find the proteins of interest in your experiment.

1 Set the quantitation options

If you've not already done so, choose between relative and absolute quantitation, use of Hi-N, protein grouping and more.

[Protein options...](#)

2 Create a shortlist to review

In the table, sort and [filter the proteins](#) based on their measurements, to generate a shortlist for further review.

[How are the measurements calculated?](#)

To sort the table by a given value, simply click the relevant column header.

3 Review the proteins

For each protein of interest, inspect the ion measurements for its peptides:

[View peptide measurements](#)

You can also double-click to review a protein.

4 Export data for further processing

By exporting your data to external tools, there's no limit to your analysis.

[Export to pathways tool](#)

[Export protein measurements](#)

[Export peptide measurements](#)

No filter applied [Create...](#)

Search

Accession	Peptide count	Confidence score	Anova (p)	Tag	Max fold change	Highest Mean	Lowest Mean	Description
P27482	9 (5)	41.3	0.00114		1.34	PBS,2h,total	L1,2h,total	Calmodulin-like protein 3 OS=Homo sapiens GN=CALML3 PE=1 SV=2
Q96751	5 (1)	13.7	0.00121		7.29	L1,2h,total	PBS,2h,total	RUN and FYVE domain-containing protein 1 OS=Homo sapiens GN=RUFY1 PE=1 SV=2
Q5T200	67 (37)	288	0.00132		1.19	PBS,2h,total	L1,2h,total	Zinc finger CCCH domain-containing protein 13 OS=Homo sapiens GN=ZC3H13 PE=1 SV=1
Q5VZM2 (+1)	3 (2)	12.1	0.00133		1.27	PBS,2h,total	L1,2h,total	Ras-related GTP-binding protein 8 OS=Homo sapiens GN=RRAG8 PE=1 SV=1
Q9UNY4	7 (3)	30	0.0014		1.78	PBS,2h,total	L1,2h,total	Transcription termination factor 2 OS=Homo sapiens GN=TTF2 PE=1 SV=2
Q9NX22	3 (1)	16.1	0.00142		1.55	PBS,2h,total	L1,2h,total	Probable ATP-dependent RNA helicase DDX43 OS=Homo sapiens GN=DDX43 PE=2 SV=2
P49642	2 (2)	7.1	0.00148		4.35	L1,2h,total	PBS,2h,total	DNA primase small subunit OS=Homo sapiens GN=PRIM1 PE=1 SV=1
P62633	2 (1)	6.85	0.00155		1.46	PBS,2h,total	L1,2h,total	Cellular nucleic acid-binding protein OS=Homo sapiens GN=CNBP PE=1 SV=1
P07197	53 (28)	241	0.00156		1.29	PBS,2h,total	L1,2h,total	Neurofilament medium polypeptide OS=Homo sapiens GN=NEFM PE=1 SV=3
P61604	9 (5)	54.5	0.00159		1.63	PBS,2h,total	L1,2h,total	10 kDa heat shock protein, mitochondrial OS=Homo sapiens GN=HSP10 PE=1 SV=2
Q04917	12 (1)	79.2	0.00161		1.95	L1,2h,total	PBS,2h,total	14-3-3 protein eta OS=Homo sapiens GN=YWHAE PE=1 SV=4
Q03113	2 (1)	11.8	0.0017		1.64	PBS,2h,total	L1,2h,total	Guanine nucleotide-binding protein subunit alpha-12 OS=Homo sapiens GN=GNA12 PE=1 SV=4
Q9Y3D9	2 (1)	6.73	0.0017		1.46	PBS,2h,total	L1,2h,total	28S ribosomal protein S23, mitochondrial OS=Homo sapiens GN=MRPS23 PE=1 SV=2
P05387	6 (2)	54.6	0.00172		1.24	PBS,2h,total	L1,2h,total	60S acidic ribosomal protein P2 OS=Homo sapiens GN=RPLP2 PE=1 SV=1
P07195	4 (3)	18.9	0.00175		1.51	PBS,2h,total	L5,2h,total	L-lactate dehydrogenase B chain OS=Homo sapiens GN=LDB PE=1 SV=2
Q562R1	34 (9)	186	0.00176		1.35	PBS,2h,total	L1,2h,total	Beta-actin-like protein 2 OS=Homo sapiens GN=ACTBL2 PE=1 SV=2
O43707	16 (5)	91.8	0.0018		1.29	PBS,2h,total	L1,2h,total	Alpha-actinin-4 OS=Homo sapiens GN=ACTN4 PE=1 SV=2
P06576	28 (13)	148	0.0018		1.14	L1,2h,total	PBS,2h,total	ATP synthase subunit beta, mitochondrial OS=Homo sapiens GN=ATP5B PE=1 SV=3
O14950 (+2)	4 (4)	23.8	0.00201		1.35	PBS,2h,total	L1,2h,total	Myosin regulatory light chain 12B OS=Homo sapiens GN=MYL12B PE=1 SV=2
Q75369	93 (42)	437	0.00207		1.45	PBS,2h,total	L1,2h,total	Filamin-B OS=Homo sapiens GN=FLNB PE=1 SV=2

Selected protein: Filamin-B OS=Homo sapiens GN=FLNB PE=1 SV=2

[View peptide measurements](#)



Experiment design

Review your data from a different perspective:

Current design: L1,2h>Total

Quantifiable proteins displayed: 277

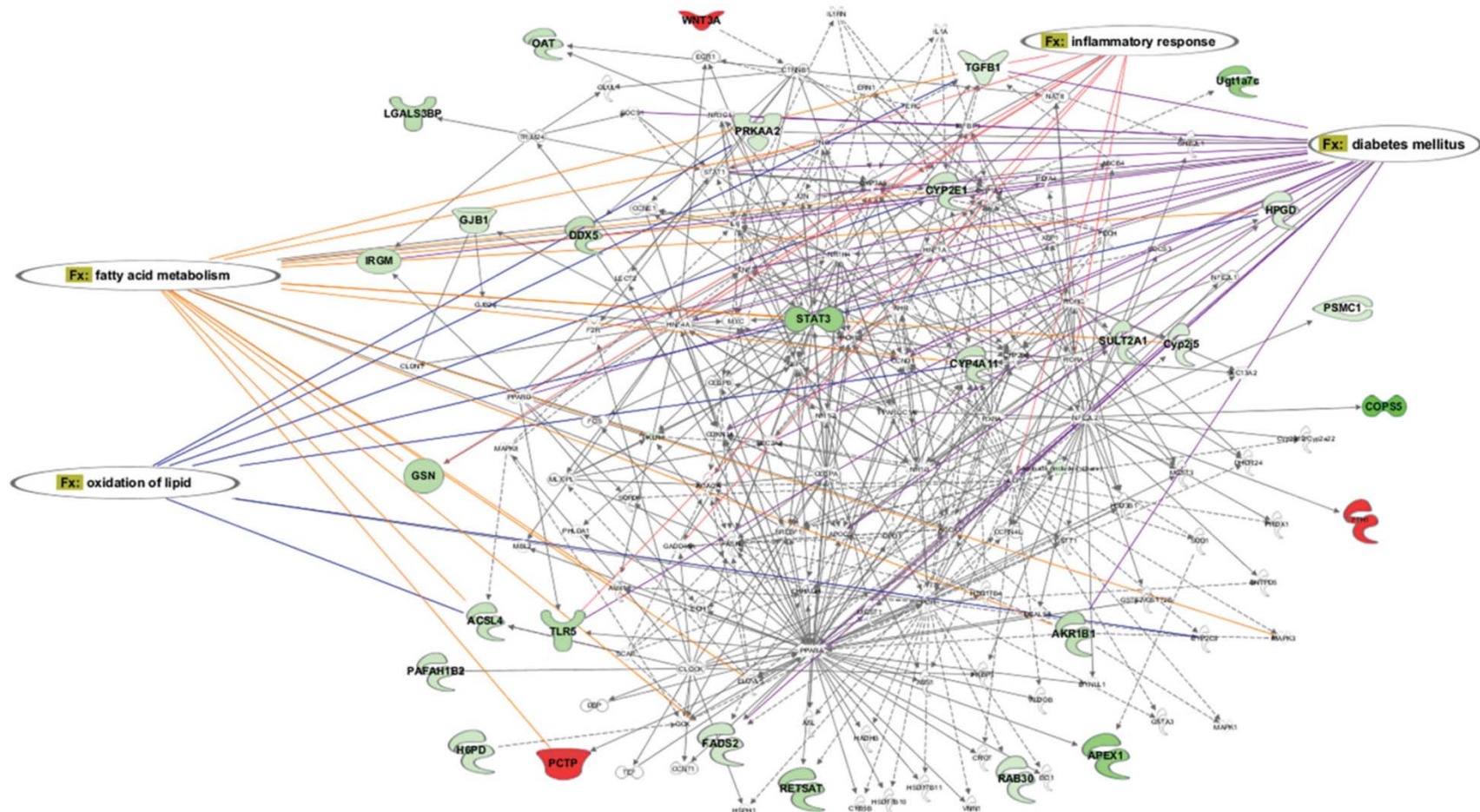
Section Complete

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
Q658J3	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!!!**