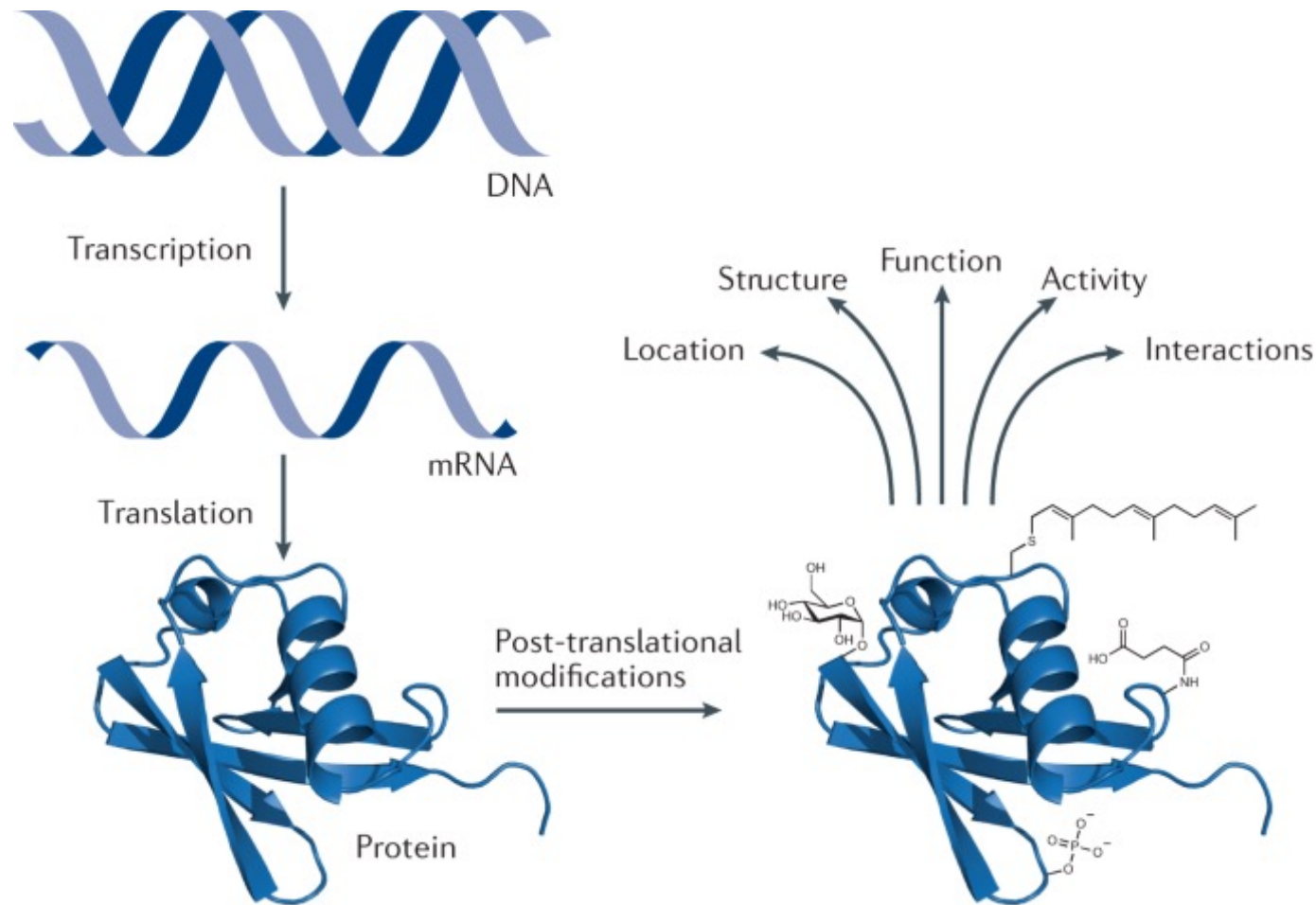


LECTURE 5: PROTEIN POST- TRANSLATIONAL MODIFICATIONS

Protein Post-translational Modifications (PTMs)



- PTM analysis is a discipline unique to proteomics.
 - Detect and/or quantify as many PTMs as possible from a complex mixture of proteins;
 - Find “all” PTMs in a single protein or small group of proteins.

Protein Post-translational Modifications (PTMs)

- What are they?

Addition of groups or deletion of parts/amino acids to make functionally different proteins.

Currently about 1,530 documented modifications

- Types of PTMs

Methylation

Acetylation

Glycosylation

Phosphorylation

Ubiquitination

etc. (how many??)

A Modification Database

UNIMOD protein modifications for mass spectrometry

Unimod

Logged as Guest

Log out

Change password

Advanced search

Help

Search for:

Any field

Contains

Search

Details found:

1530

Page 1 of 77

Records Per

Page::

20

Show all

	Accession #	PSI-MS Name	Interim name	Description	Monoisotopic mass	Average mass	Composition
View	1287		Arg-loss	Loss of arginine due to transpeptidation	-156.101111	-156.1857	H(-12) C(-6) N(-4) O(-1)
View	765		Met-loss	Removal of initiator methionine from protein N-terminus	-131.040485	-131.1961	H(-9) C(-5) N(-1) O(-1) S(-1)
View	676		Trp->Gly	Trp->Gly substitution	-129.057849	-129.1586	H(-7) C(-9) N(-1)
View	313	Lys-loss	-K	Loss of Lysine	-128.094963	-128.1723	H(-12) C(-6) N(-2) O(-1)
View	1224		Trp->Ala	Trp->Ala substitution	-115.042199	-115.1320	H(-5) C(-8) N(-1)
View	1239		Tyr->Gly	Tyr->Gly substitution	-106.041865	-106.1219	H(-6) C(-7) O(-1)
View	646		Arg->Gly	Arg->Gly substitution	-99.079647	-99.1344	H(-9) C(-4) N(-3)
View	673		Trp->Ser	Trp->Ser substitution	-99.047285	-99.1326	H(-5) C(-8) N(-1) O
View	400	Tyr->Dha	DehydroalaY	Dehydroalanine (from Tyrosine)	-94.041865	-94.1112	H(-6) C(-6) O(-1)
View	1237		Tyr->Ala	Tyr->Ala substitution	-92.026215	-92.0954	H(-4) C(-6) O(-1)
View	1093		Phe->Gly	Phe->Gly substitution	-90.046950	-90.1225	H(-6) C(-7)
View	766		Met-loss+Acetyl	Removal of initiator methionine from protein N-terminus, then acetylation of the new N-terminus	-89.029920	-89.1594	H(-7) C(-3) N(-1) S(-1)
View	1232		Trp->Pro	Trp->Pro substitution	-89.026549	-89.0947	H(-3) C(-6) N(-1)
View	1235		Trp->Val	Trp->Val substitution	-87.010899	-87.0788	H(-1) C(-6) N(-1)
View	1189		Arg->Ala	Arg->Ala substitution	-85.063997	-85.1078	H(-7) C(-3) N(-3)
View	1234		Trp->Thr	Trp->Thr substitution	-85.031634	-85.1060	H(-3) C(-7) N(-1) O
View	674		Trp->Cys	Trp->Cys substitution	-83.070128	-83.0670	H(-5) C(-8) N(-1) S
View	1117		His->Gly	His->Gly substitution	-80.037448	-80.0880	H(-4) C(-4) N(-2)
View	679		Tyr->Ser	Tyr->Ser substitution	-76.031300	-76.0960	H(-4) C(-6)
View	1090		Phe->Ala	Phe->Ala substitution	-76.031300	-76.0960	H(-4) C(-6)

Types of Protein Post-translational Modifications (PTMs)

Modification	Description
Acetylation	<i>N</i> -terminal of some residues and side chain of lysine, +42 Da
Amidation	Generally at the <i>C</i> -terminal of a mature active peptide, -1 Da
Formylation	Generally of the <i>N</i> -terminal methionine, +28 Da
Glycosylation	<i>N</i> -linked (Asn), <i>O</i> -linked (Ser or Thr), and <i>C</i> -linked (Trp), more than +162 Da, up to kDa
Phosphorylation	Of serine, threonine, tyrosine, +80 Da
Sulphation	Of tyrosine, serine or threonine, +80 Da
Methylation	Of lysine and arginine, +14 Da
Ubiquitination	Of lysine, +114 Da

For Δ masses, visit: <https://abrf.org/delta-mass>

PTMs

Location

Nucleus

Lysosome

Mitochondria

Golgi

ER

Cytosol

Ribosome

Plasma membrane

Extraceullar fluid

Extrallular matrix

Modification

acetylation, phosphorylation, methylation

mannose-6-phosphate labelled *N*-linked sugar

N-formyl acylation

N- and *O*-linked oligosaccharide, sulphation, palmitoylation

N-linked oligosaccharide, GPI-anchor

acetylation, methylation, phosphorylation,

myristoylation

N- and *O*-glycosylation, GPI-anchor

N- and *O*-glycosylation, acetylation, phosphorylation

N- and *O*-glycosylation, phosphorylation, hydroxylation

Why Are We Interested in Studying Protein Post-translational Modifications?

- Cells can rapidly respond to stimuli and perturbations
 - Protein expression is much too slow for quick adaptation
- PTMs are crucial regulators
 - Important cellular mechanisms are tightly controlled
 - Activation, folding, stabilization, interaction, localization, and secretion
- Often, diseases (e.g., cancer) are due to aberrantly activated proteins

Protein Methylation

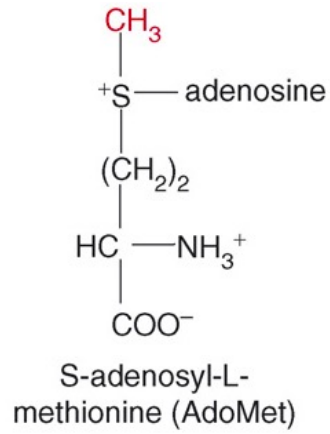
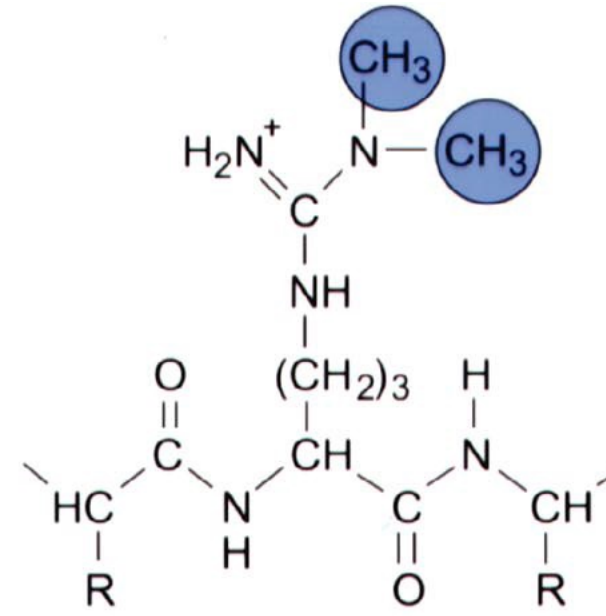
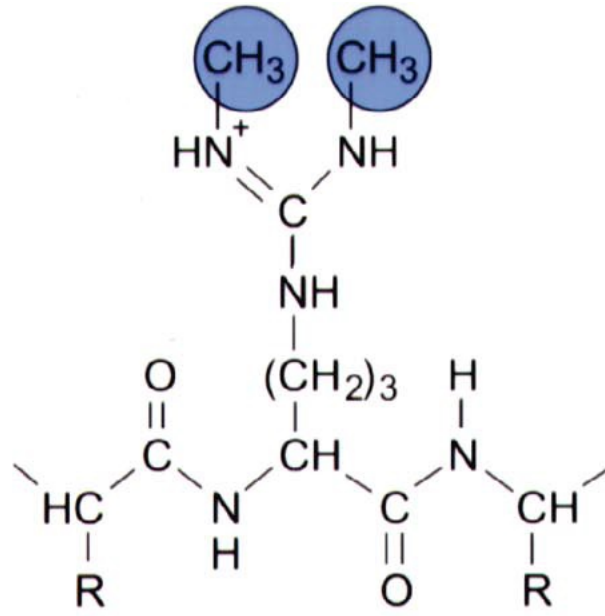
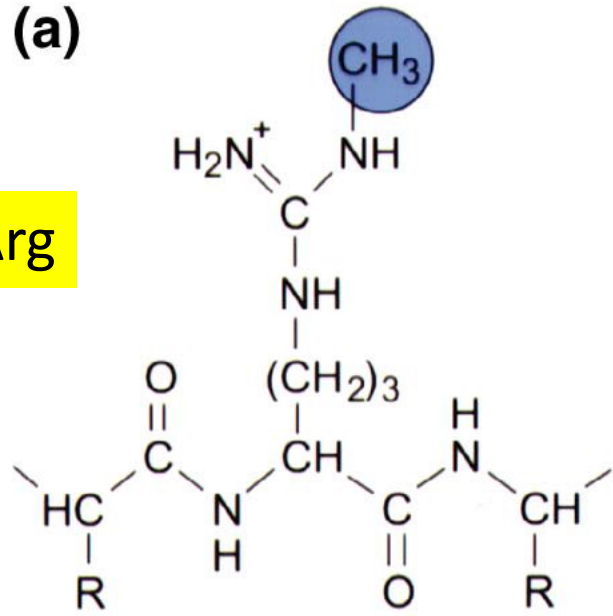
- Multiple sites and to different degrees (mono-, di-, trimethyl)
 - Lys, Arg
 - Histones, Transcription factors, etc.

Table I. Functions of protein-bound arginine and lysine methylation

Type of methylation	Postulated or proven function	Proteins involved
Protein-arginine	mRNA splicing Signal transduction Cellular proliferation Chromatin remodeling Transcriptional coactivator Protein-protein interaction Translocation Myelogenesis	Motor neuron proteins Interferon receptor Transcription factor Histones Nuclear receptor, p-53 Inter- and intra-molecules hnRNP Myelin basic protein
Protein-lysine	Chromatin remodeling Increased protein stability <i>in vivo</i> Resistance to proteolytic digestion <i>in vitro</i> Biom mineralization Modulation of NAD kinase	Histones Cytochrome <i>c</i> Cytochrome <i>c</i> Silaffins in marine diatom Calmodulin

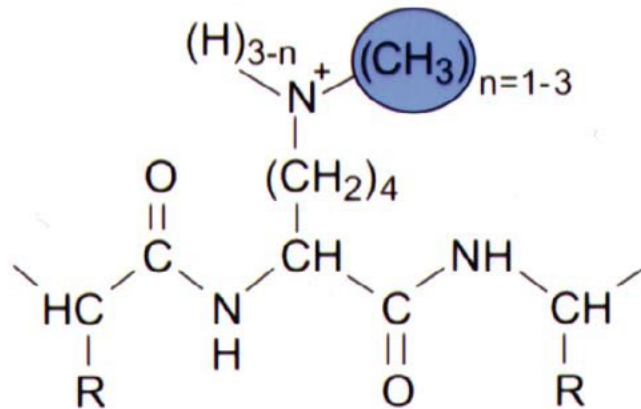
(a)

Arg



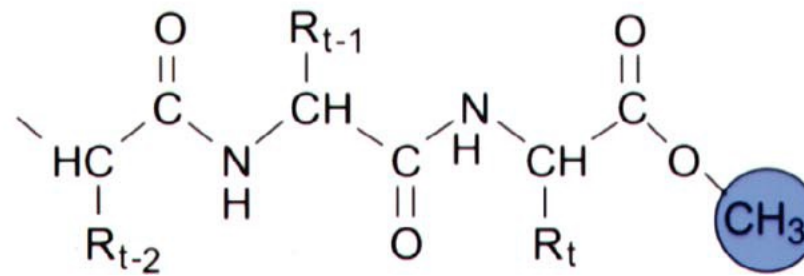
(b)

Lys



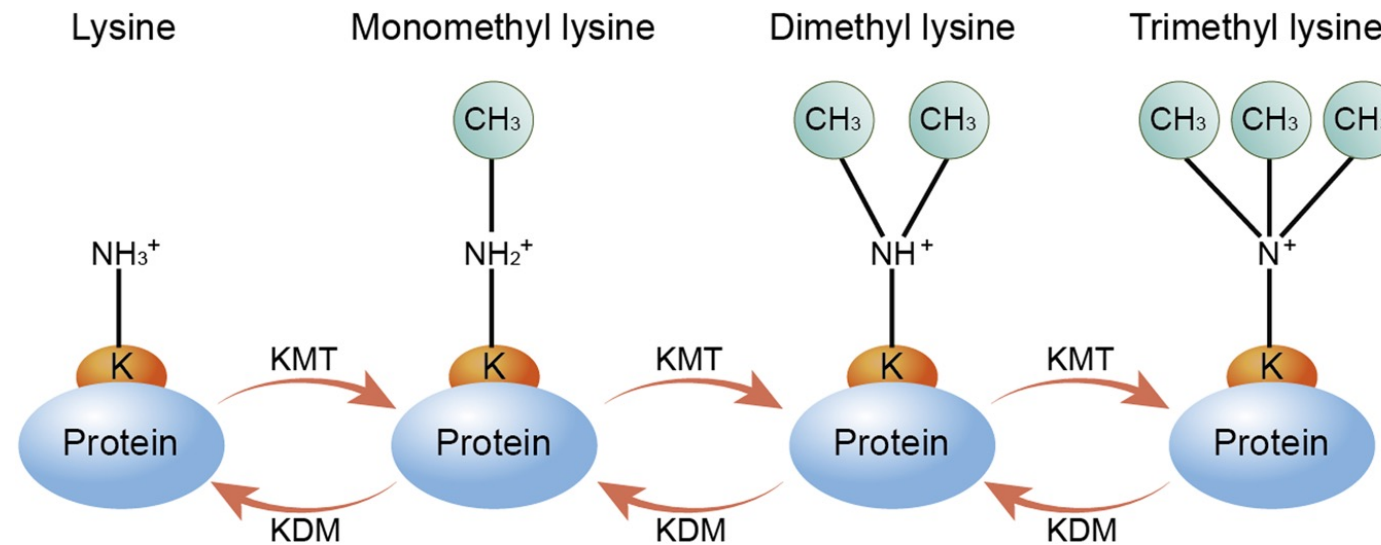
(c)

C-terminal carboxyl group



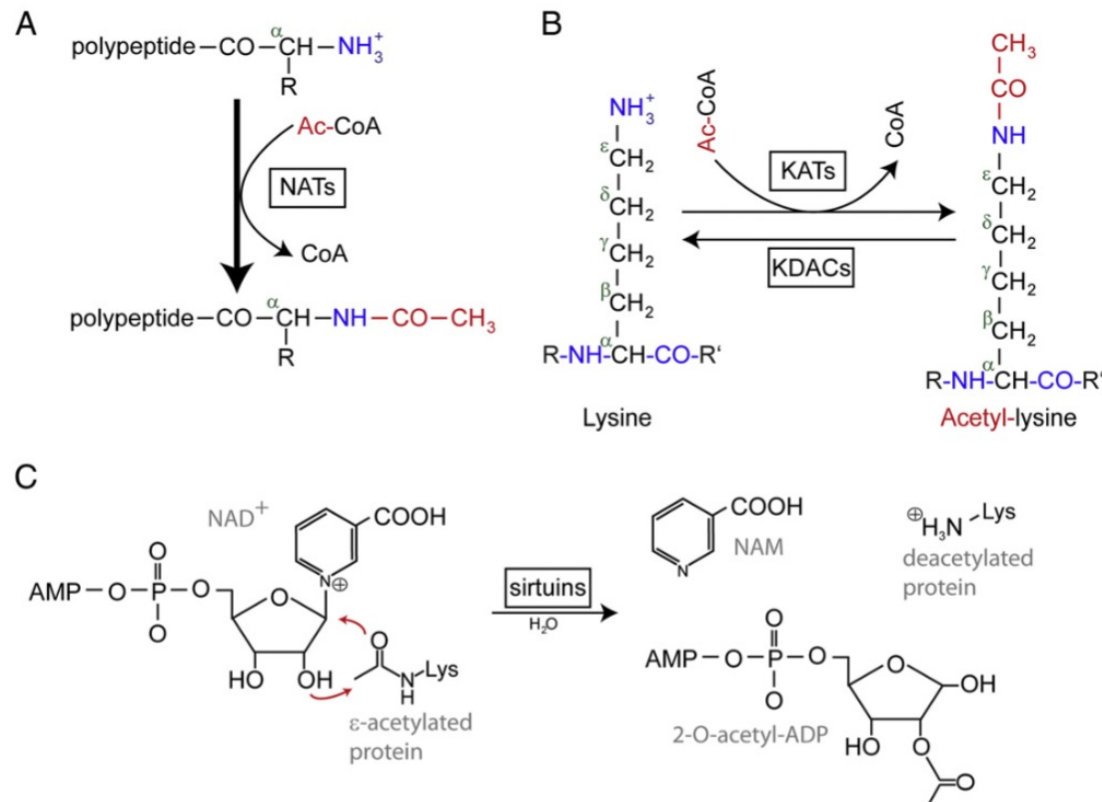
Lysine methylation of transcription factors in cancer

- Interfering with transcription factor lysine methylation can inhibit cancer cell proliferation, thereby reversing tumor progression.
- The inhibitors targeting lysine methyltransferases (KMTs) and lysine demethylases (KDMs) may be used as potential anticancer agents in the clinic.



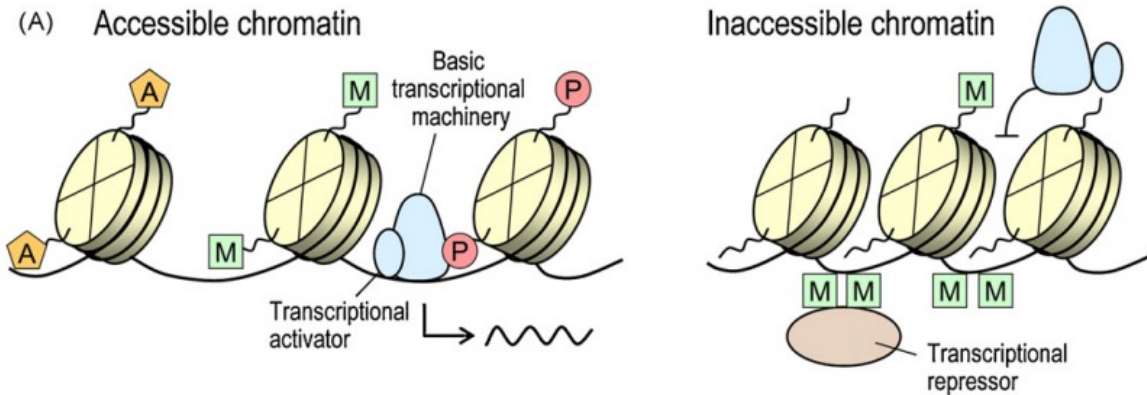
Acetylation

- Protein acetylation involves the process that acetyl group from acetyl coenzyme A (Ac-CoA) is transferred to a specific site on a polypeptide chain.

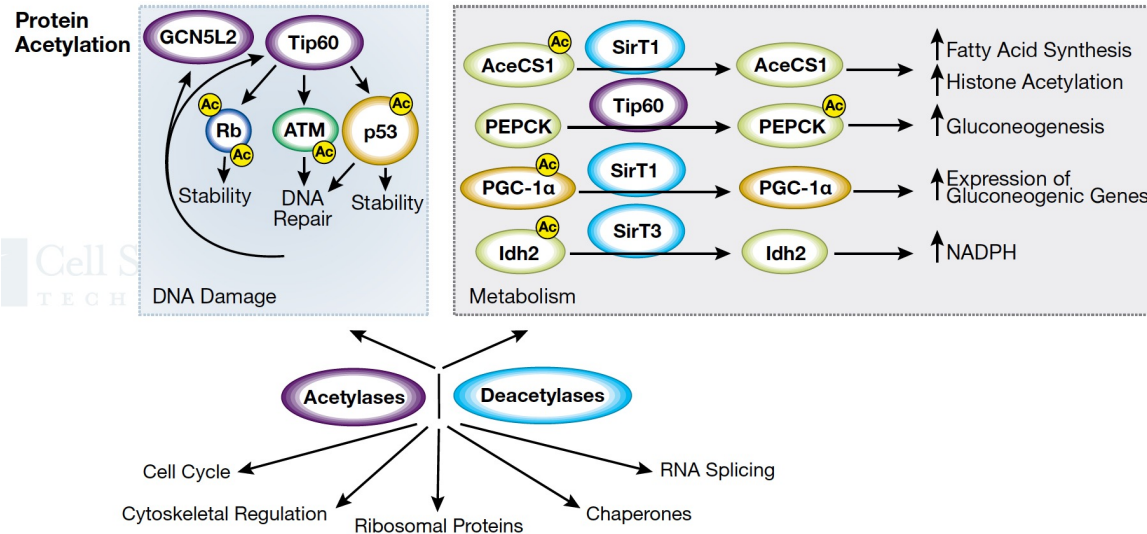


- In humans, a big majority (80-90%) of proteins are co-translationally acetylated at their N-termini by N-terminal acetyltransferases (NATs).
- Reversible acetylation on ϵ -amino group of a lysine residue is catalyzed by lysine acetyltransferases (KATs) and lysine deacetylases (KDACs).

Examples of Protein Acetylation

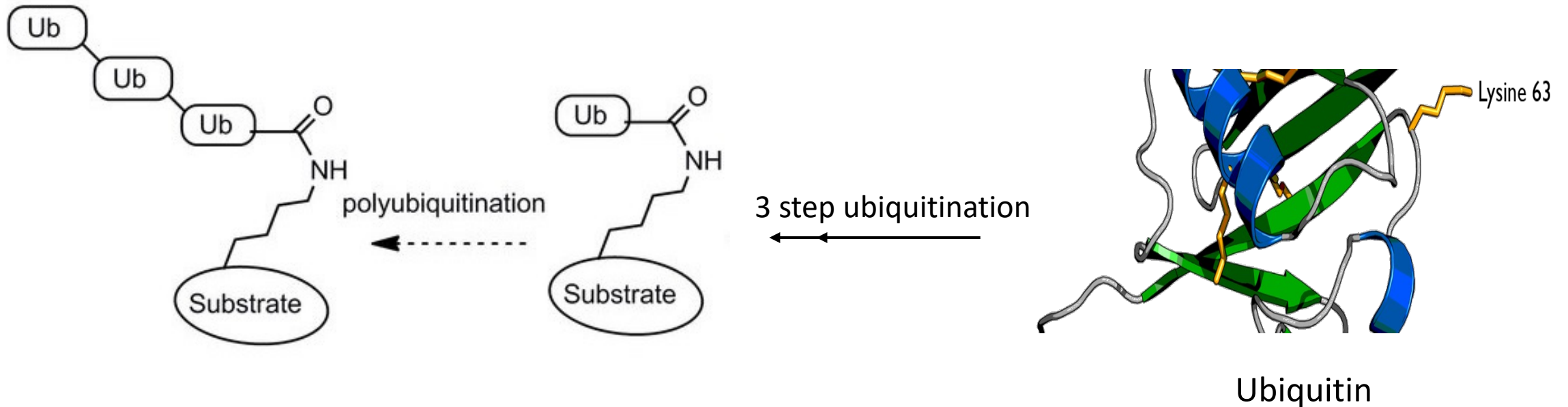


- Regulating chromatin structure and transcriptional activity
 - In histone acetylation, acetyl groups are attached to positively charged lysine in histone tails. Acetylation makes histones less positive, so less attraction to DNA. This generally loosens chromatin structure, promoting the initiation of transcription.
- Important role in immunity, circadian rhythmicity, and memory formation
- Favorable target in drug design for numerous disease conditions

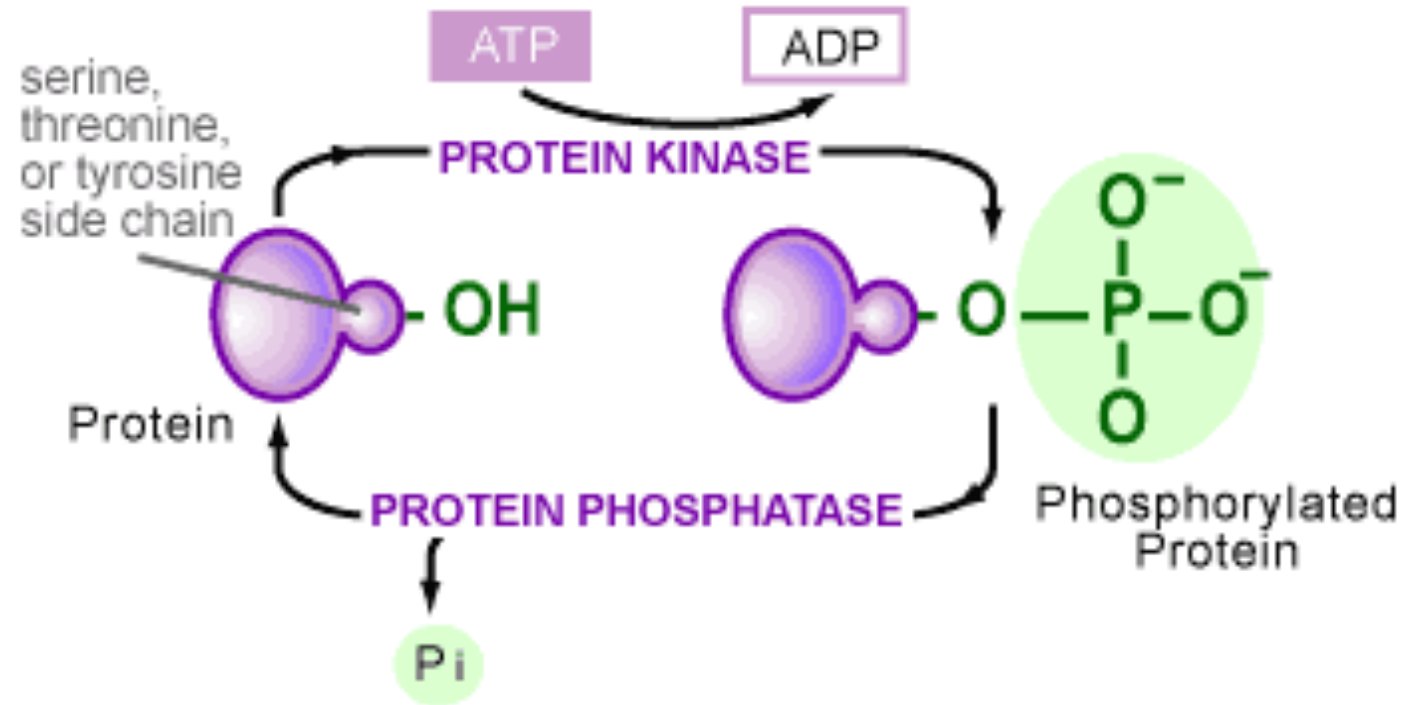


Ubiquitination

- Ubiquitination involves the attachment of the polypeptide **ubiquitin** to target proteins by a set of enzymes.
- Ubiquitin is a highly conserved 8 kD protein that tags proteins for destruction. Thereby denatured, damaged, or improperly translated proteins are removed from cells.



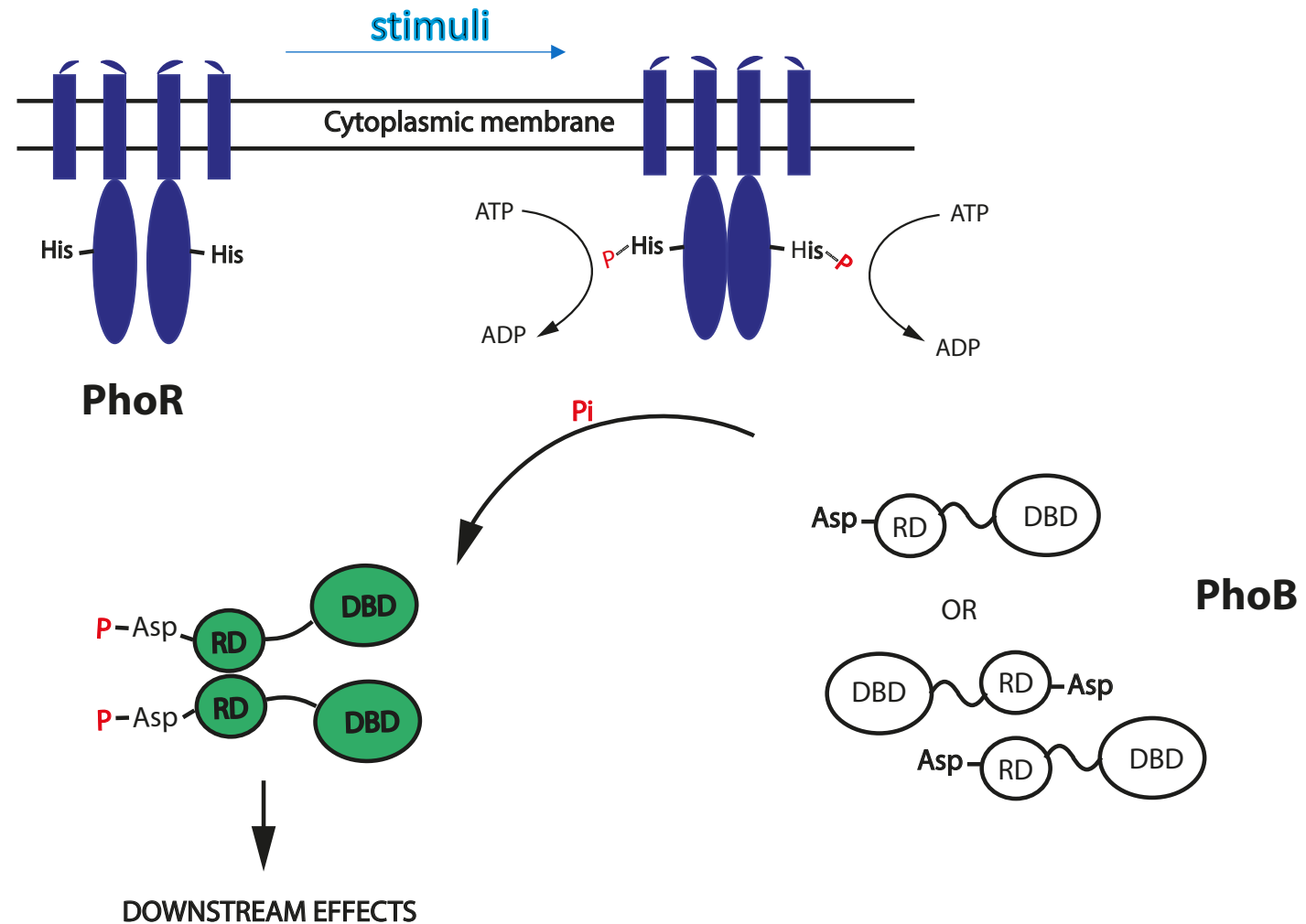
Phosphorylation



- Protein phosphorylation is a reversible modification catalyzed by **kinases** and **phosphatases**.

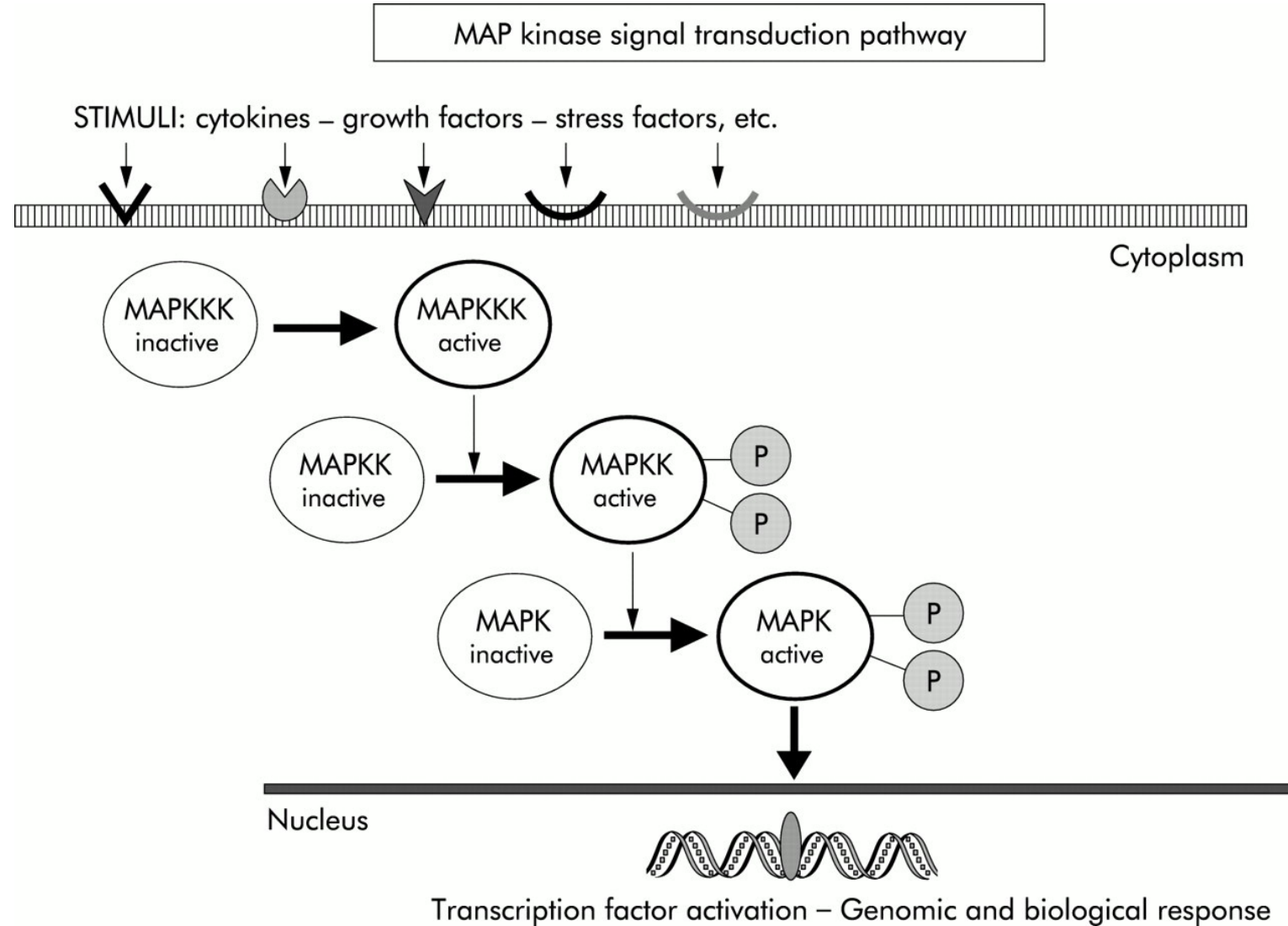
Examples of Protein Phosphorylation

Two component system in prokaryotes



Examples of Protein Phosphorylation

In eukaryotes



Glycosylation

- Glycosylation, the attachment of sugar moieties to proteins.
- **The very important and complex form of PTMs**

- Roles:

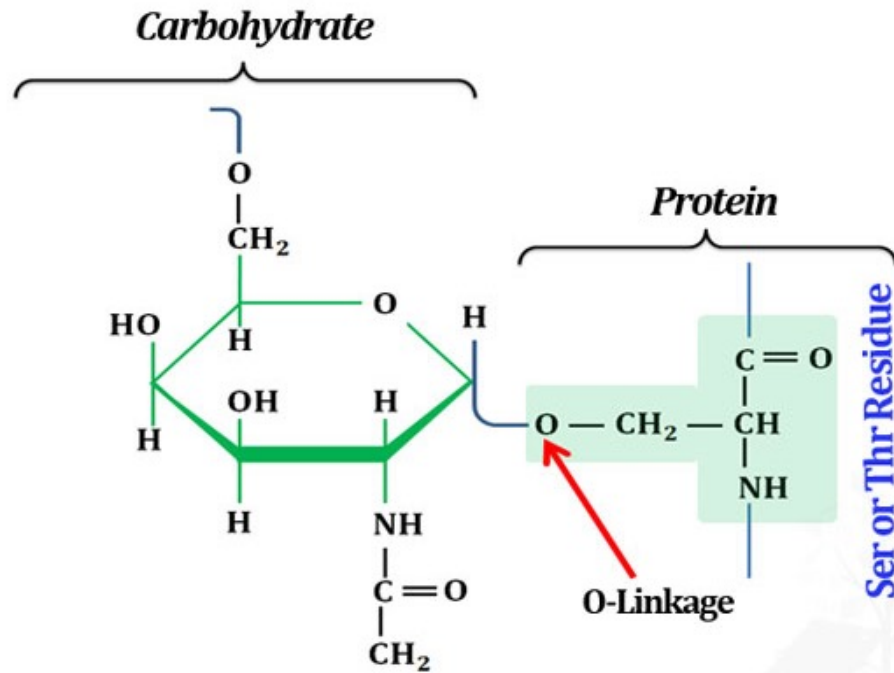
- Folding (activity)
- Localization
- Stability (half life)
- Interactions
- Linked to diseases

Diseases	Glycoprotein	Alteration
Hepatic cancer	α -fetoprotein	Different N-glycan structures
Immune disorder	CD43	Different O-glycan structures
Rheumatoid arthritis	IgG	N-glycans, reduction in terminal galactosylation
Choriocarcinoma	hCG	N- and O-linked glycans, hyperbranching
Alcohol abuse	Transferrin	N-glycans, desialylation

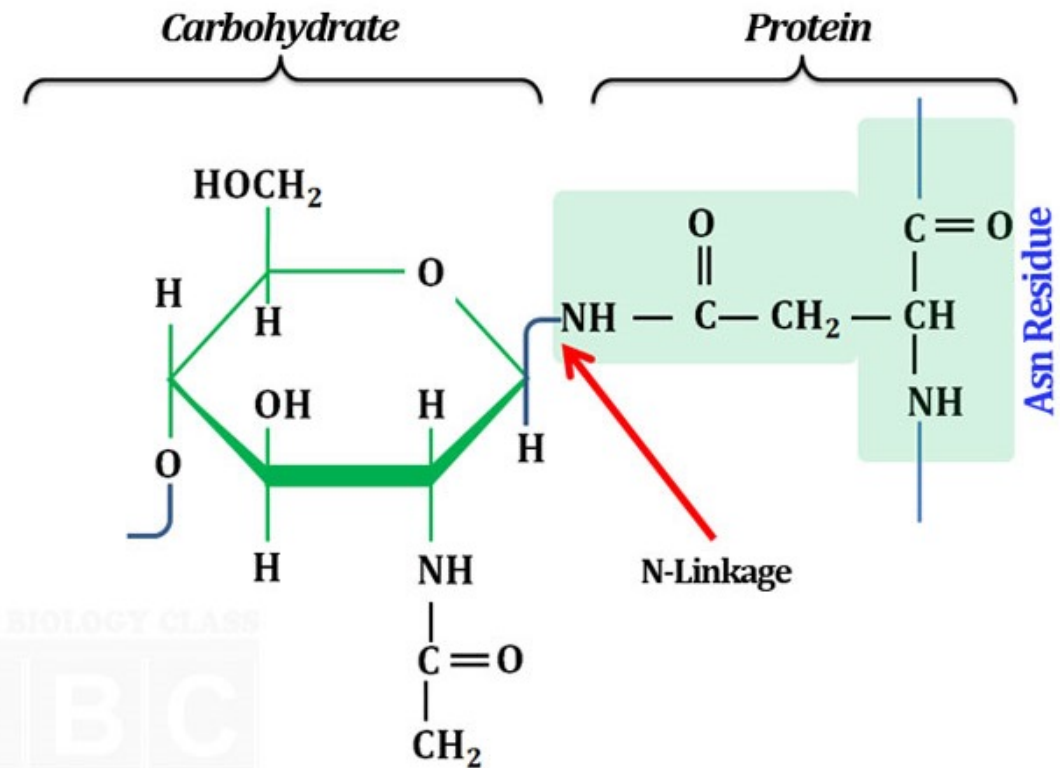
Types of Glycosylation

- Based on the nature of the sugar-peptide bond and the oligosaccharide attached, glycosylation includes:
 - N-linked: Glycan binds to amino group of Asn (Asn-X-Ser/Thr/Cys) in the ER
 - O-linked: Monosaccharides bind to OH group of Ser/Thr in the Golgi stack
 - Glypiation: Glycan core links a phospholipid and a protein
 - C-linked: Mannose binds to the indole ring of Trp
 - Phosphoglycosylation: Glycan binds to Ser via phosphodiester bond

Chemistry of Glycosylation



O-Linked Glycosylation



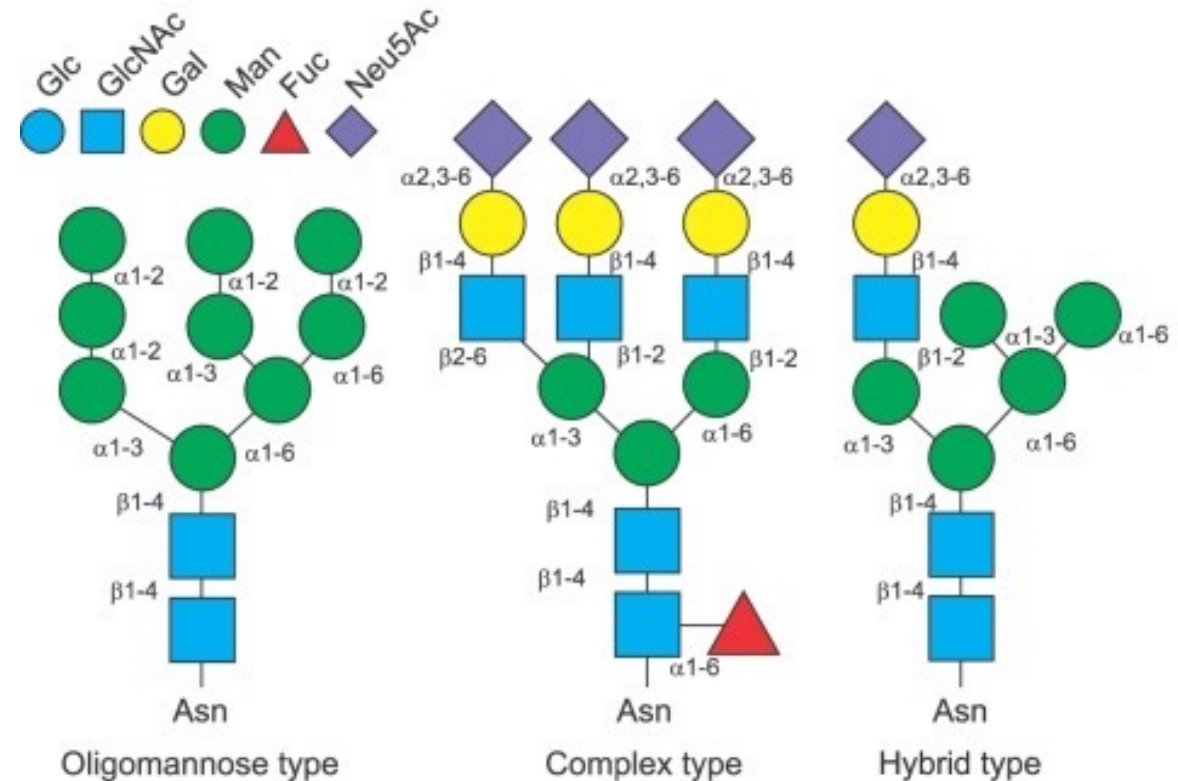
N-Linked Glycosylation

Sugar Abbreviations

Glucose	Glc
Galactose	Gal
Mannose	Man
Rhamnose	Rha
Fucose	Fuc
Glucosamine	GlcN
Galactosamine	GalN
Mannosamine	ManN
N-acetylglucosamine	GlcNAc
N-acetylgalactosamine	GalNAc
N-acetylmannosamine	ManNAc
Glucuronic acid	GlcA
Galacturonic acid	GalA
Mannuronic acid	ManA
N-acetylneuraminic acid	NeuNAc
3-deoxy-D-manno-2-octulosonic acid	Kdo

N-Linked Glycans

- N-linked glycans are covalently attached to Asn residues within a consensus sequence (Asn-Xaa-Ser/Thr/Cys), enabling prediction of the modification sites by protein sequence analysis
- All N-linked glycans share a common pentasaccharide core (**GlcNAc2Man3**) recognized by **lectins** and **N-glycanase enzymes (PNGase F)**
- These reagents have been used to visualize proteins bearing N-linked glycans from cell or tissue lysates and to enrich them for mass spectrometry analysis



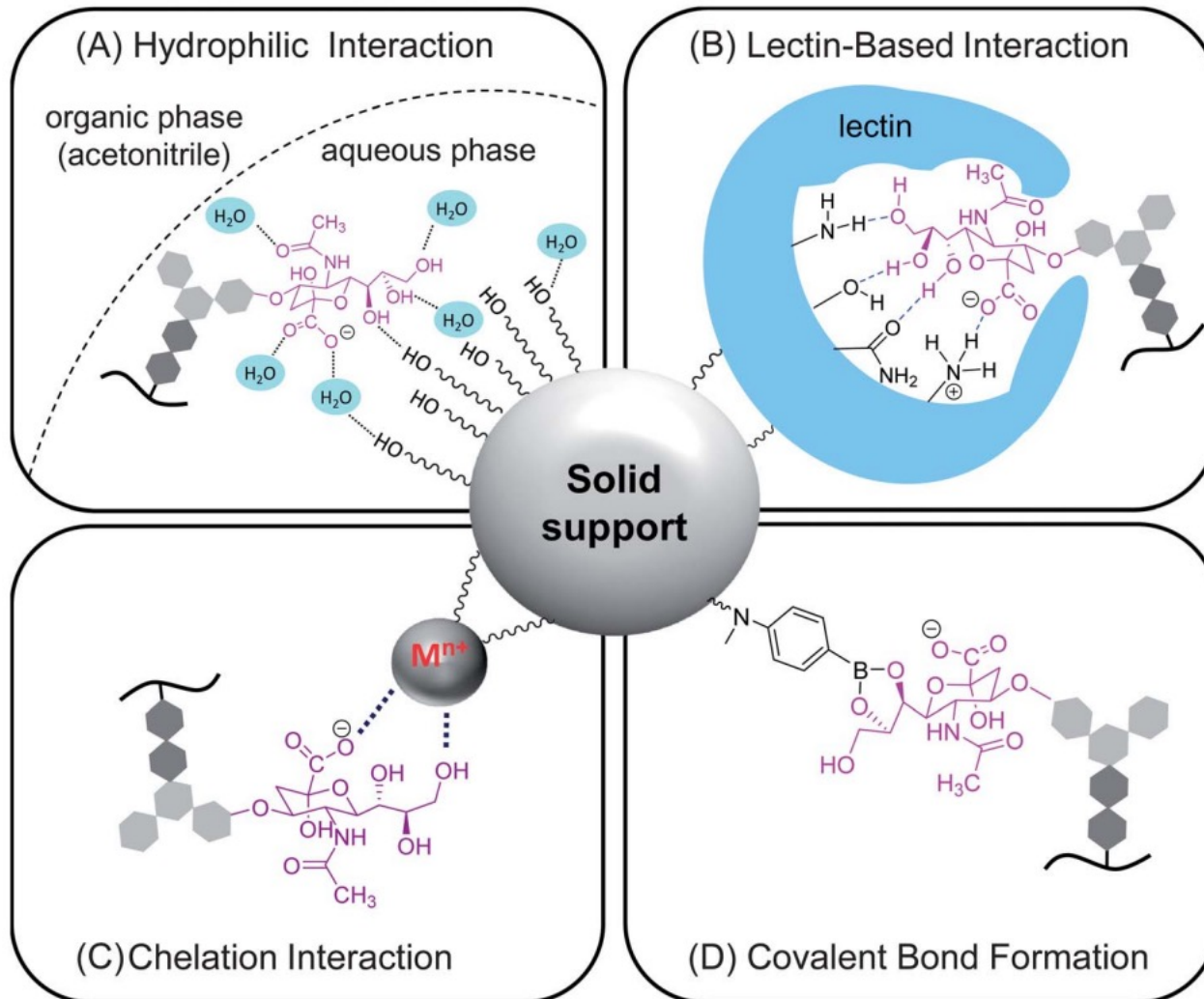
O-Linked Glycans

- Comparable tools are lacking for the study of proteins bearing O-linked glycans.
- Mucin-type, the most prevalent O-linked glycosylation is characterized by an *N*-acetylgalactosamine (GalNAc) residue -linked to the hydroxyl group of Ser or Thr. GalNAc residue is installed by a family of 24 *N*-acetyl-galactosaminyltransferases, then further elaborated by a series of glycosyltransferases to generate higher-order O-linked structures.
- Because of the complex biosynthetic origin, O-linked glycans are not installed at a defined consensus motif and their presence cannot be accurately predicted based on the protein's primary sequence.

Glycoproteomics

- **Glycoproteomics** analyzes the totality of **glycoproteins** (e.g., molecules consisting of a protein part and a sugar part [**glycan**]).
- Glycosylation of proteins mainly takes place in the endoplasmic reticulum (ER) and the Golgi
- Enzymes, such as glycosyltransferases and glycosidases, attach or remove the sugar groups to/from the proteins
- These enzymes mainly target S, T and N
- It impacts on charge, conformation and protein stability
- Glycosylation is the most complex of all PTMs
 - Glycans (polysaccharides) represent a huge variety in their composition

Enrichment of Glycopeptides



- Due to low abundance glycoproteins/glycopeptides usually need to be enriched
- **Lectin affinity chromatography**
 - Lectins are sugar-binding proteins
 - Lectins play a role in many cellular processes, for example in virus attachment to host cells, where the lectin's affinity to glycosylated membrane proteins is used by the virus
 - Many lectins are available that have very specific ligands

Lectins- carbohydrate-binding proteins

Lectins **read sugar code** and mediate many biological processes :

- Cell-cell recognition
- Signaling
- Adhesion
- Intracellular targeting of newly synthesized proteins

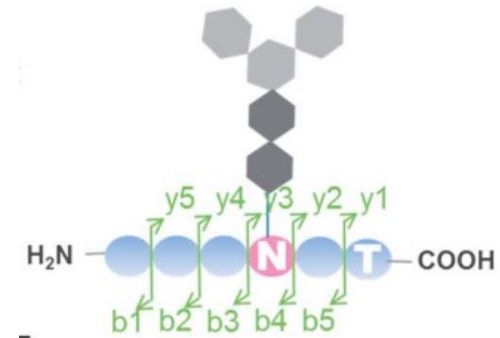
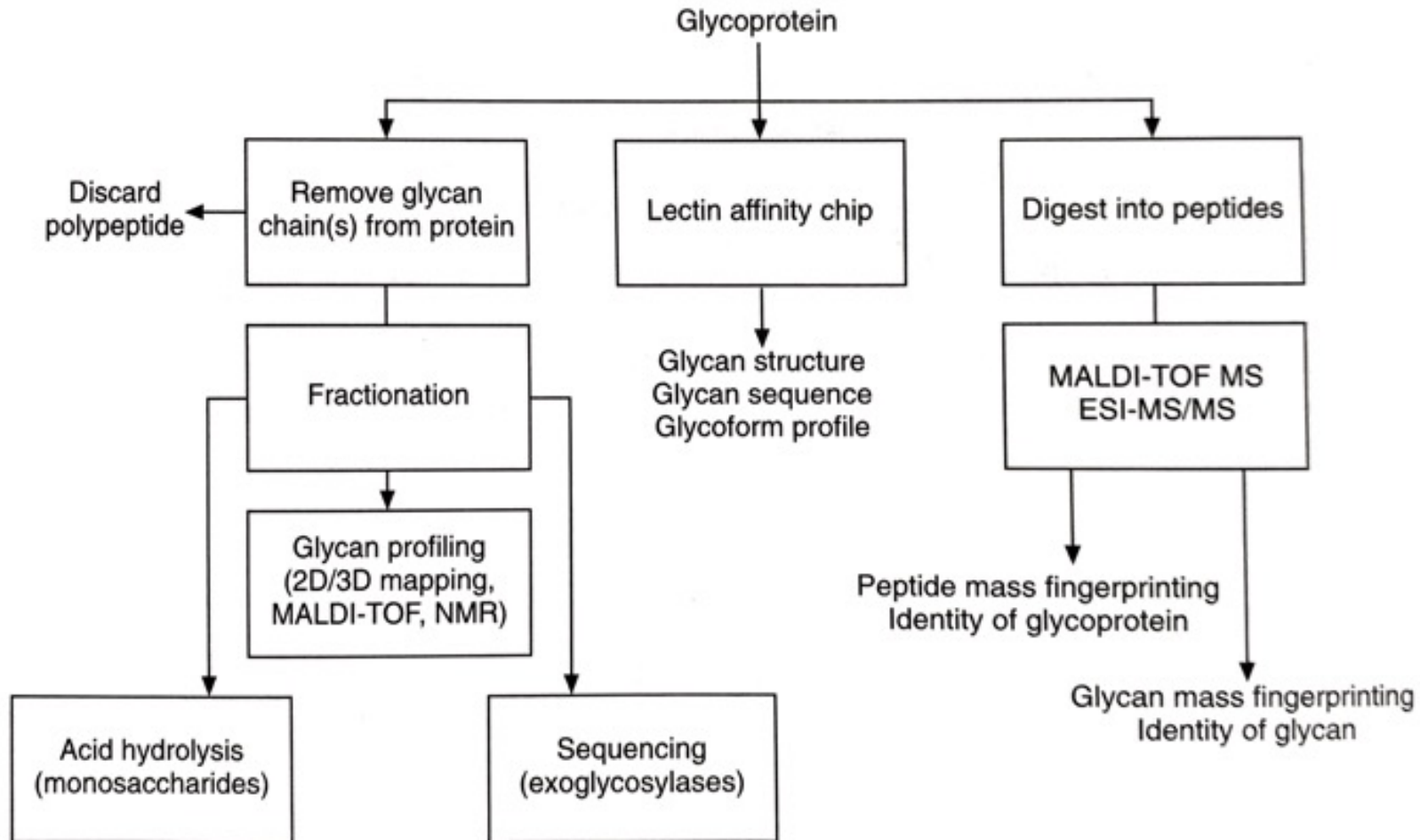
Commonly used lectins for enrichment of glycoproteins or glycopeptides

Lectin	Specific Affinity
<i>Aleuria aurantia</i> lectin (AAL)	Fuc α 1-6 GlcNAc, Fuc α 1-3(Gal β 1-4)GlcNAc
Concanavalin A (Con A)	High-Mannose, Man α 1-6(Man α 1-3)Man
Jacalin	Gal β 1-3GalNAc, GalNAc
<i>Phaseolus vulgaris</i> leucoagglutinin (PHA-L)	Tri/tetra-antennary complex-type N-glycan
<i>Ricinus communis</i> agglutinin-I (RCA-I)	Gal β 1-4GlcNAc β 1-R
<i>Sambucus nigra</i> lectin (SNA, EBL)	Sia α 2-6Gal/GalNAc
<i>Ulex europaeus</i> agglutinin-I (UEA-I)	Fuc α 1-2Gal β 1-4GlcNAc
Wheat Germ Agglutinin (WGA)	Chitin oligomers, Sia

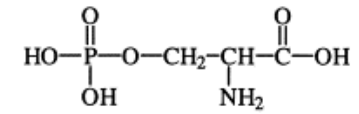
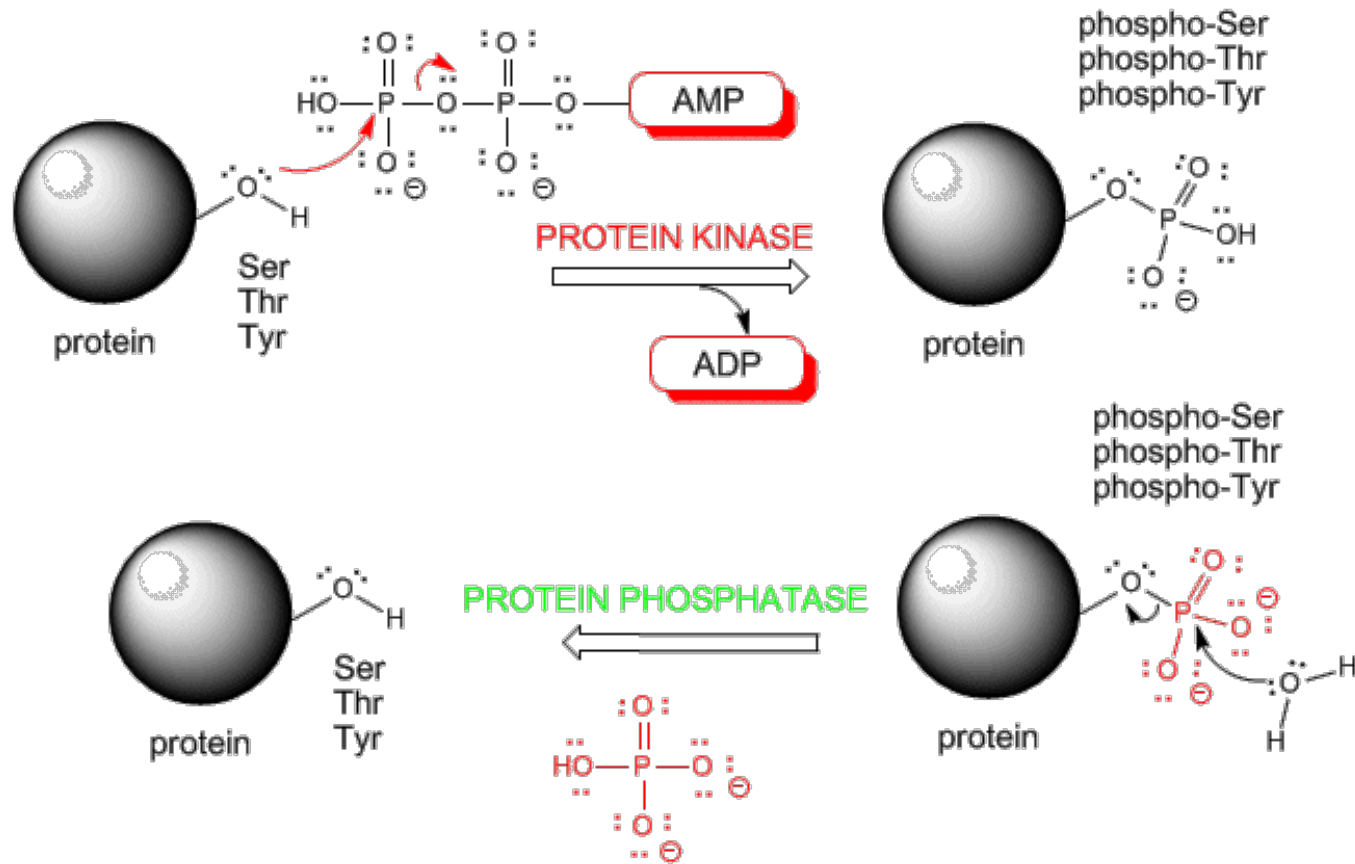
Mass Spec Analysis of Glycopeptides

- The hydrophilic nature of glycans limits the surface activity and the ionization efficiency
 - Natural and basic glycoconjugates can be protonated
 - Acidic glycoconjugates can only be deprotonated (negative ESI mode!)
- Often, derivatization is used to increase hydrophobicity and volatility (and thereby ionization efficiency)
- As for unmodified peptides, MS/MS can be performed to sequence the glycan as well as the underlying peptides
- MS/MS of glycopeptides is more complicated than that of unmodified peptides, since the chemical properties of the peptide and glycan are not similar
 - Heterogeneity in glycan size and charge
- Using CID, the collision energy is highly important to the content of the MS/MS spectrum

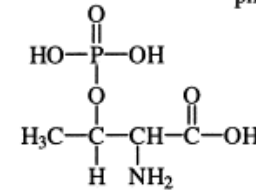
General Work-flow For The Full Analysis Of Glycoproteins



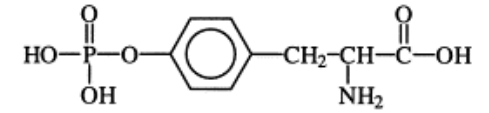
PhosphoProteomics



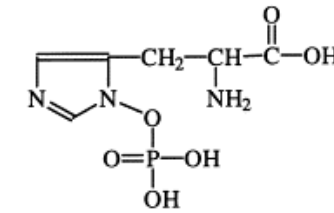
phospho-serine



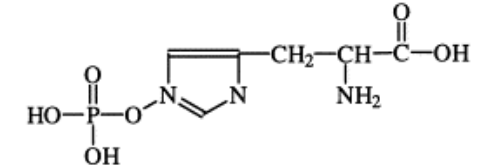
phospho-threonine



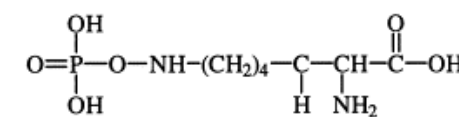
phospho-tyrosine



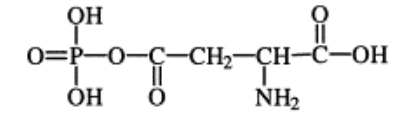
1-phospho-histidine



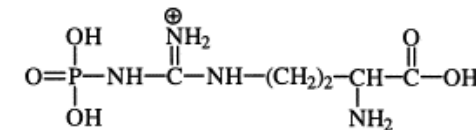
3-phospho-histidine



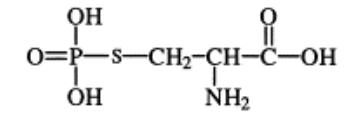
N^ε-phospho-lysine



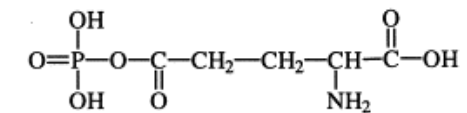
phospho-aspartic acid



N^ω-phospho-arginine



phospho-cysteine



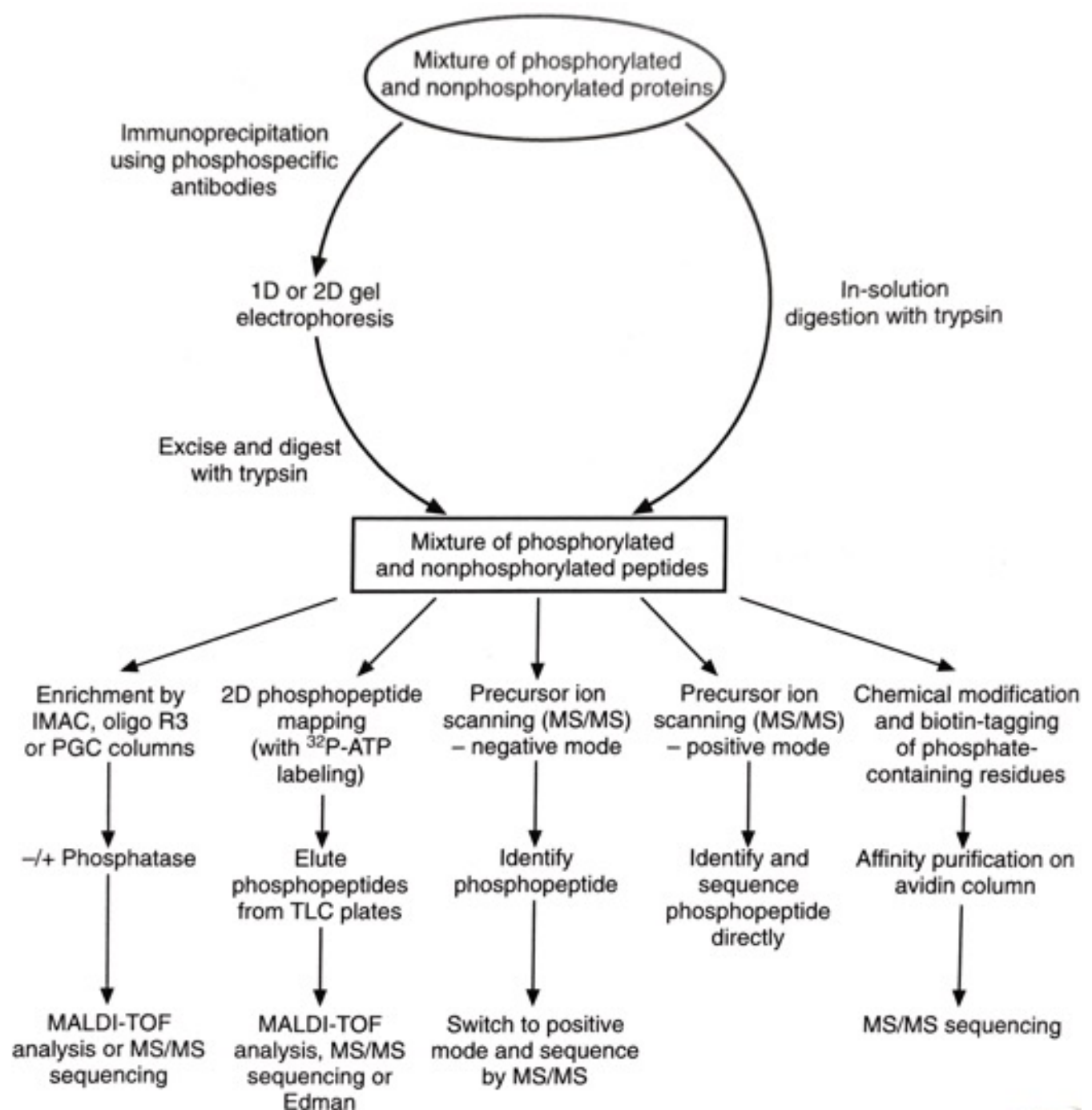
phospho-glutamic acid

PhosphoProteomics

- Phosphorylation lies at the heart of many many biological processes, including signal transduction, gene expression, and the regulation of cell division.
- Most phosphoproteins have more than one phosphorylation site and exist as a mixture of alternative phosphoforms.
- The phosphoproteome is not only complex but also extremely dynamic.
 - A catalog of all phosphoproteins
 - Distribution of phosphorylation site
 - Abundance of alternative phosphoforms under different conditions

Different Techniques for Phosphoproteins Analysis

- Because low abundance of many phosphoproteins, enrichment is usually necessary.



Enrichment of Phosphoproteins: Antibodies

- Immunopurification with immobilized anti-phospho-Tyr antibodies
- Antibody-based method is well established for Tyr, but not for other residues (Ser & Thr)
 - Antiphosphoserine or Antiphosphothreonine antibodies don't have satisfactory immunoprecipitation result
- The antibodies don't bond phosphopeptide strongly, so antibodies are usually employed before digestion
- It is limited in throughput and hard to automate

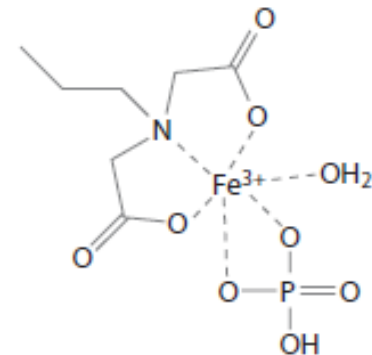
Enrichment of Phosphoproteins: IMAC₁

- Immobilized Metal Affinity Chromatography
 - Phosphates have high affinity to trivalent metal ions
 - Metal ions are immobilized on columns
 - A variety of metals has been used, including Fe³⁺, Ga³⁺, Al³⁺ and Zr³⁺

- Problems with IMAC

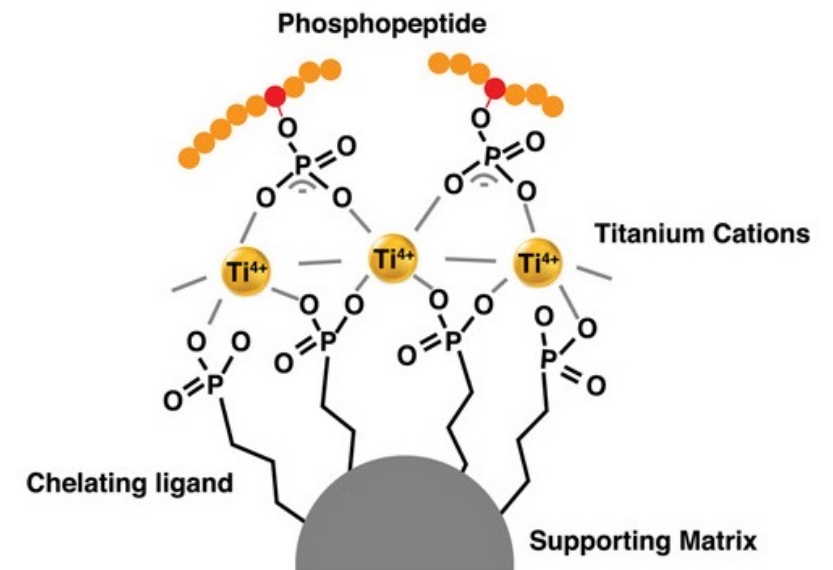
- If the pH during the loading is very basic, then non-phosphopeptides bind as well
- Strongly acidic peptides (rich in E and D) are also affine to the metal complexes

IMAC-Fe³⁺
phosphate complex



Enrichment of Phosphoproteins: IMAC₂

- High chemical stability
- Unique amphoteric ion-exchange properties
- Organic phosphates are effectively adsorbed to TiO_2 in acidic conditions and desorbed in alkaline conditions
 - Selective enrichment for phosphopeptides
- Dihydroxy benzoic acid (DHB) as a competitive binder in TiO_2 enrichment to avoid unspecific binding of D and E

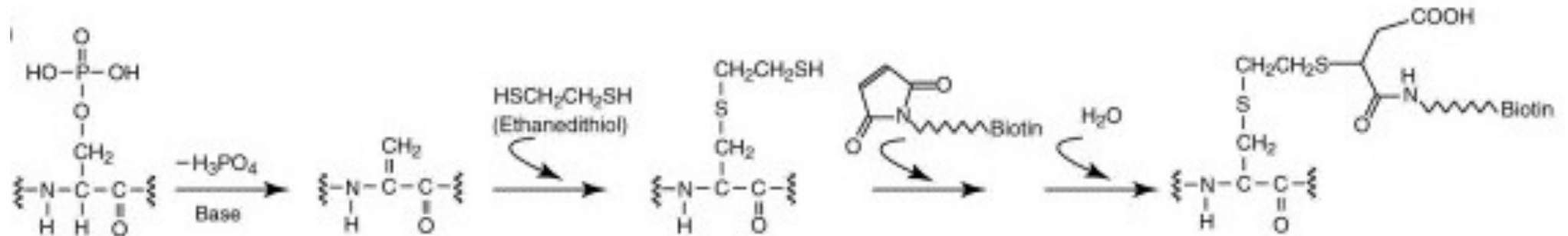


Enrichment of Phosphoproteins: SCX

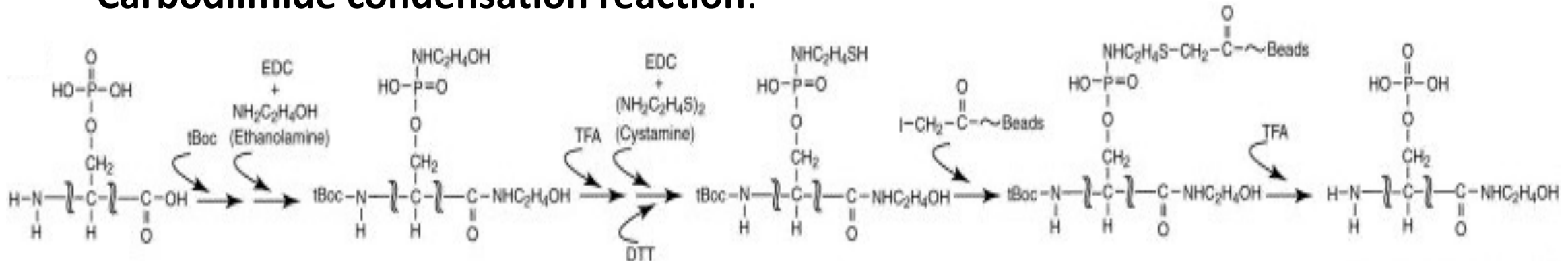
- Strong Cation Exchange (SCX) strategy is based on the difference in the solution **charge state** of phosphorylated and non-phosphorylated peptides
 - At pH 2.7 a (typical) tryptic peptide has charge $z = +2$ (N-terminal amine group + C-terminal K or R)
 - If this peptide is phosphorylated $\rightarrow z = +1$, since the phosphate group is negatively charged
 - Using a linear salt gradient the phosphopeptides can be enriched in early SCX fraction.
 - Note that multiply phosphorylated peptides will be in the flow-through

Enrichment of Phosphoproteins: Chemical Modification

- For larger amounts of starting samples, affinity purification with chemical modifications of phosphates can be used.
 - β -elimination reaction:



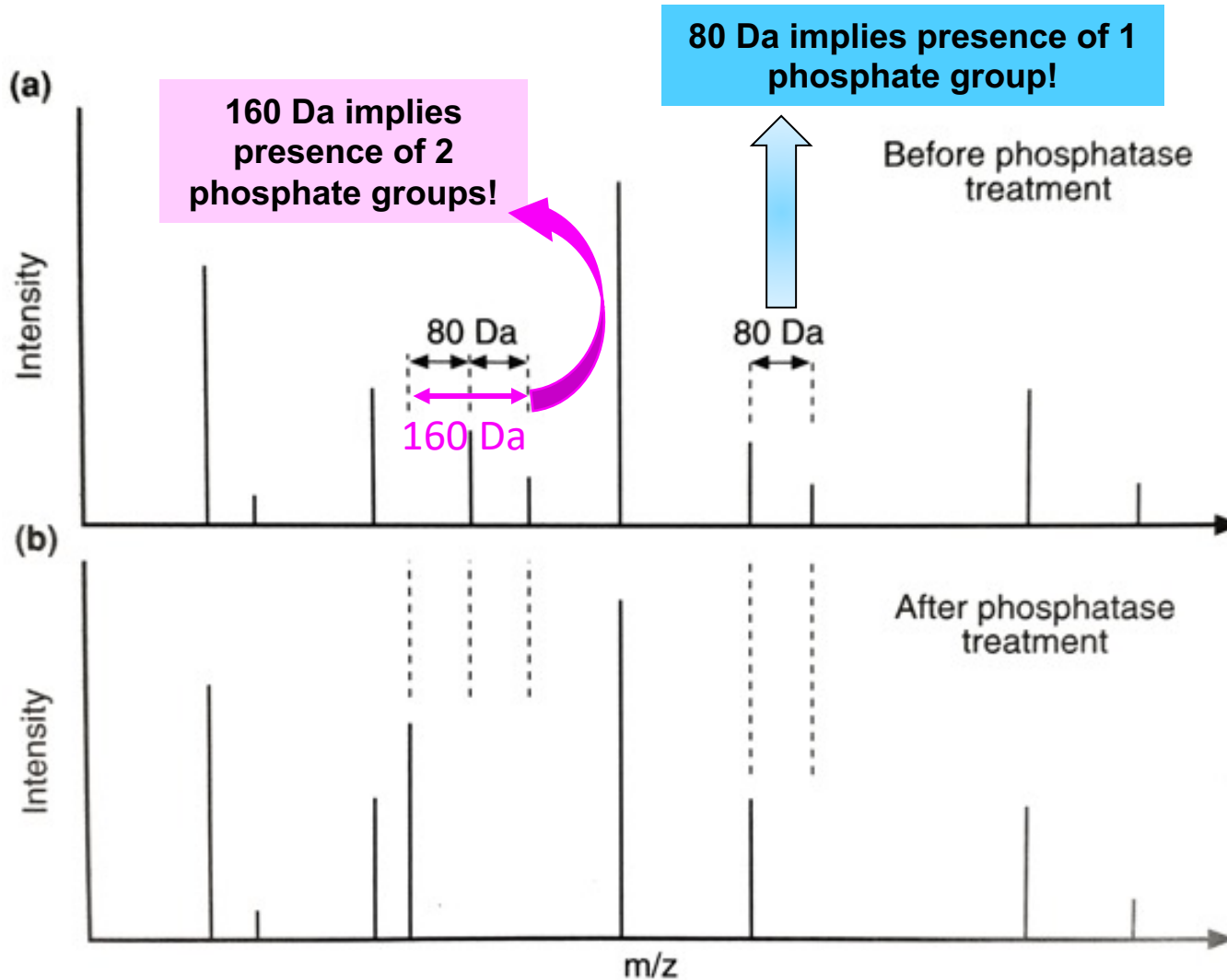
- Carbodiimide condensation reaction:



MS for Phosphopeptides: **MALDI-TOF MS**

- **MALDI-TOF MS** is most often used to analyze intact peptides, and correlative database searching (peptide mass fingerprinting) allows the derived masses to be matched against the theoretical peptides of known proteins.
- If the identity of the protein is known or can be deduced, phosphopeptides can be identified simply by examining the mass spectrum for masses shift of 79.983 ($\text{HPO}_3 = 80 \text{ Da}$) compared to predicted masses.

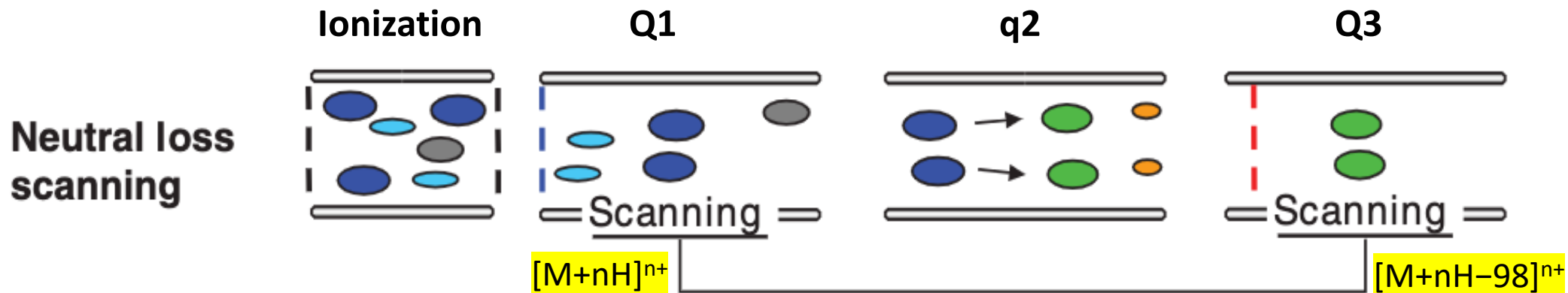
Phosphopeptide identification by MALDI-TOF MS combining with alkaline phosphatase treatment



- This method does not identify the phosphorylated residues directly.
- If there is only one possible phosphorylation site or if a consensus kinase target site is present, then the site can be identified accurately.

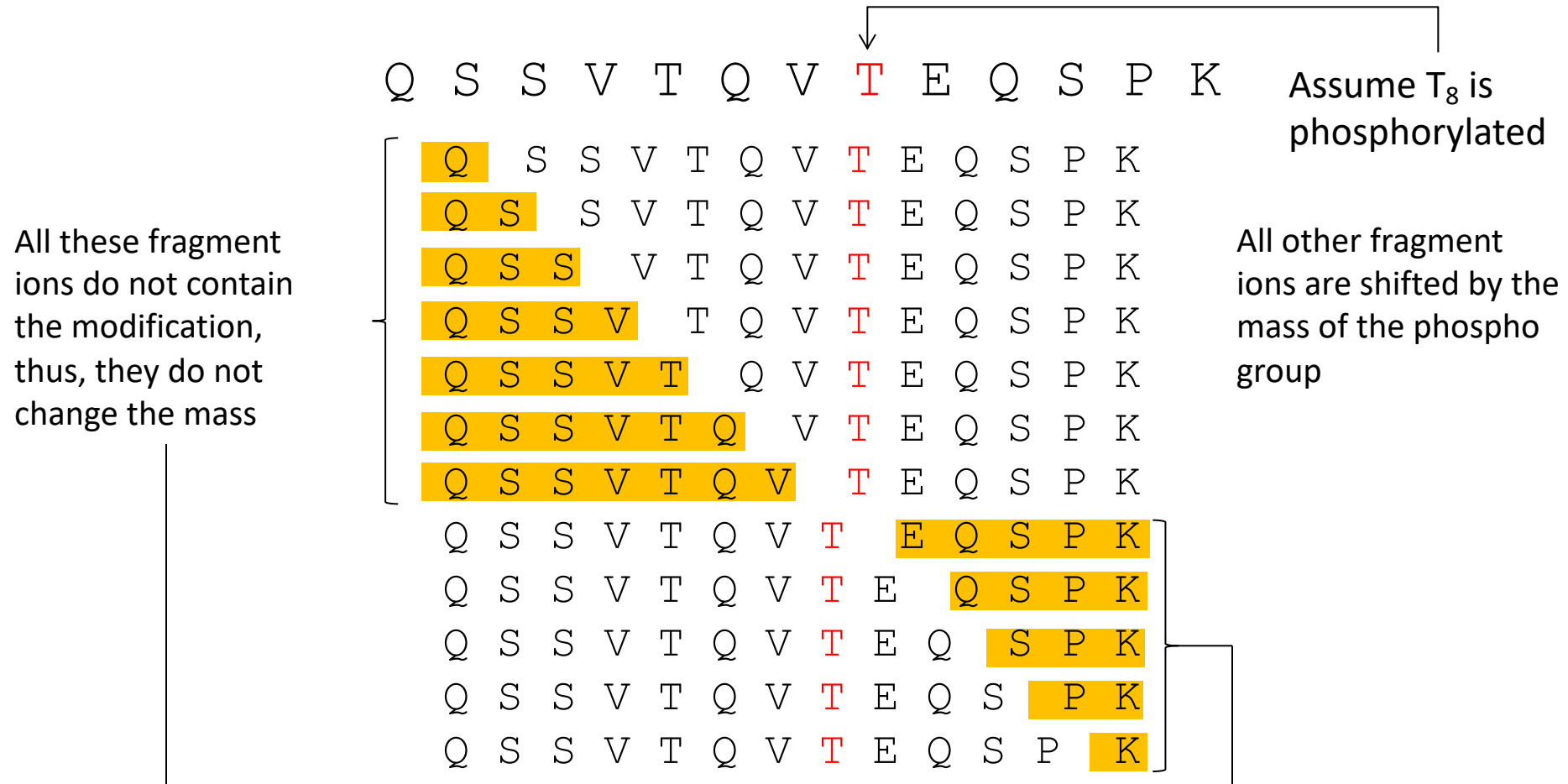
MS for Phosphopeptides: Neutral-Loss-Dependent MSⁿ

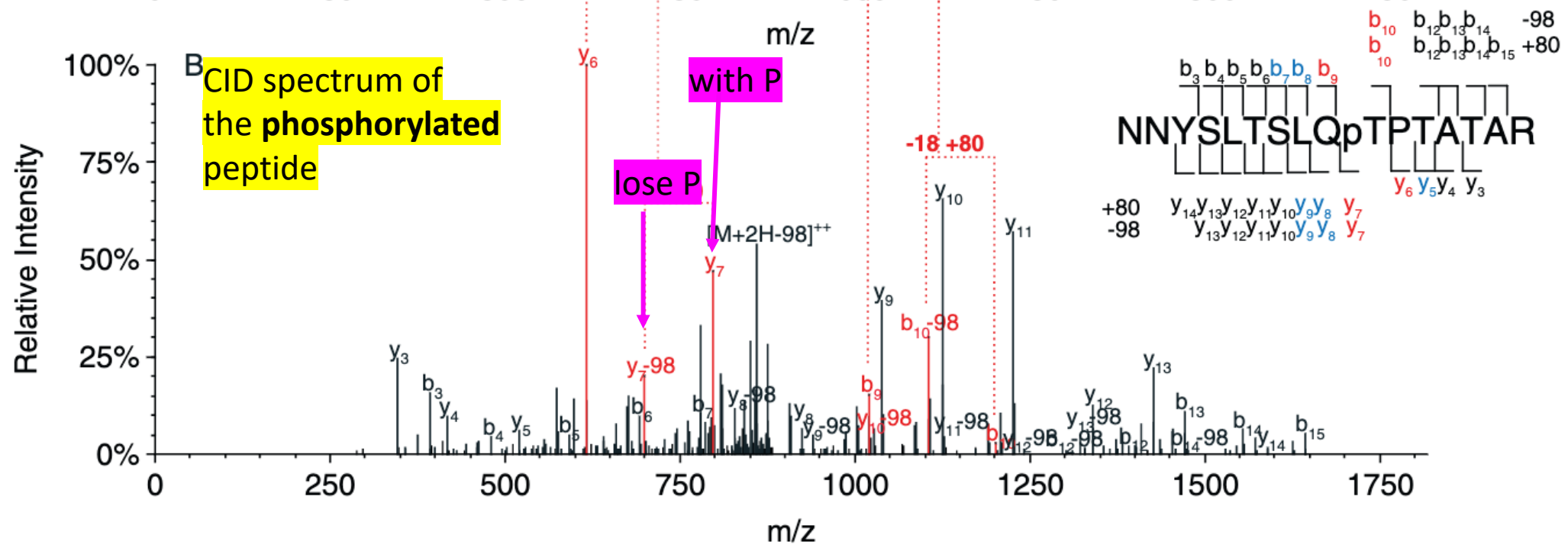
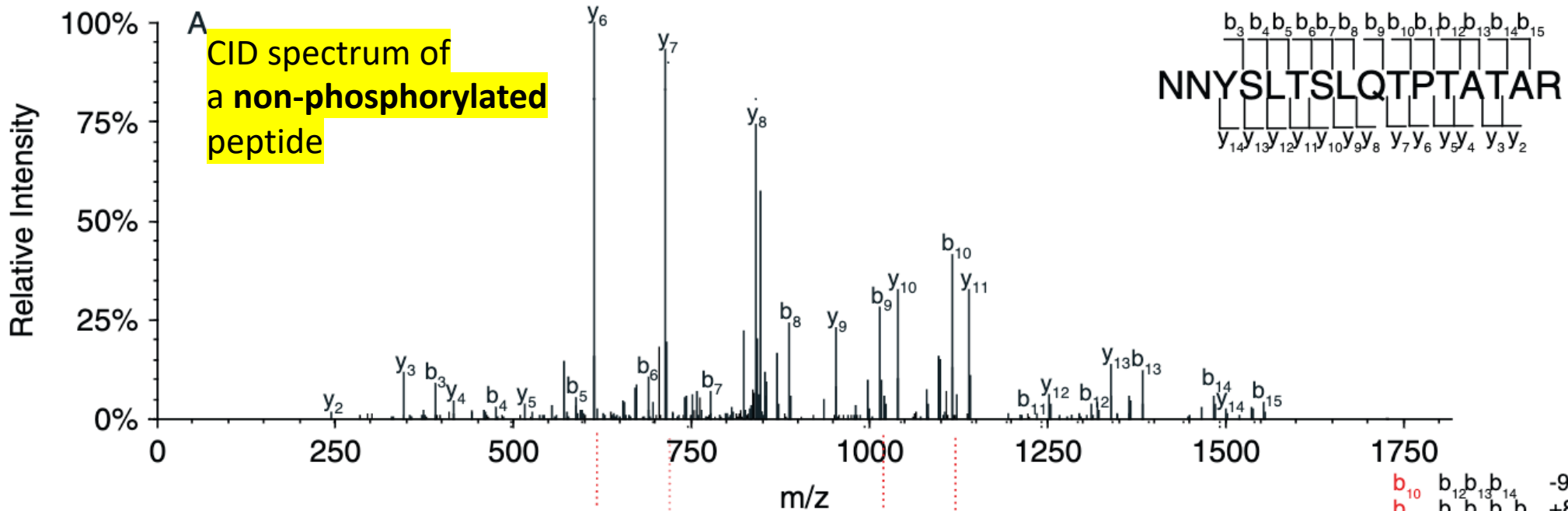
- Fragmentation of phosphopeptides by CID is often dominated by neutral losses
 - pS: predominantly H₂PO₄ (-98 Da)
 - pT: H₂PO₄ (-98 Da) and HPO₃ (-80 Da)
 - pY: HPO₃ (-80 Da) or (HPO₃ + H₂O) (-98 Da)
- **Neutral loss ion scanning** on a triple quadrupole mass spectrometer can be used to detect the loss of the phosphate group from phosphopeptides.
 - Q1 is used to scan the whole mass range
 - q2 is used as a collision cell
 - Q3 is set to scan at [M+nH-98]ⁿ⁺ to detect phosphopeptides that lose the phosphate group.



Phosphorylated Example

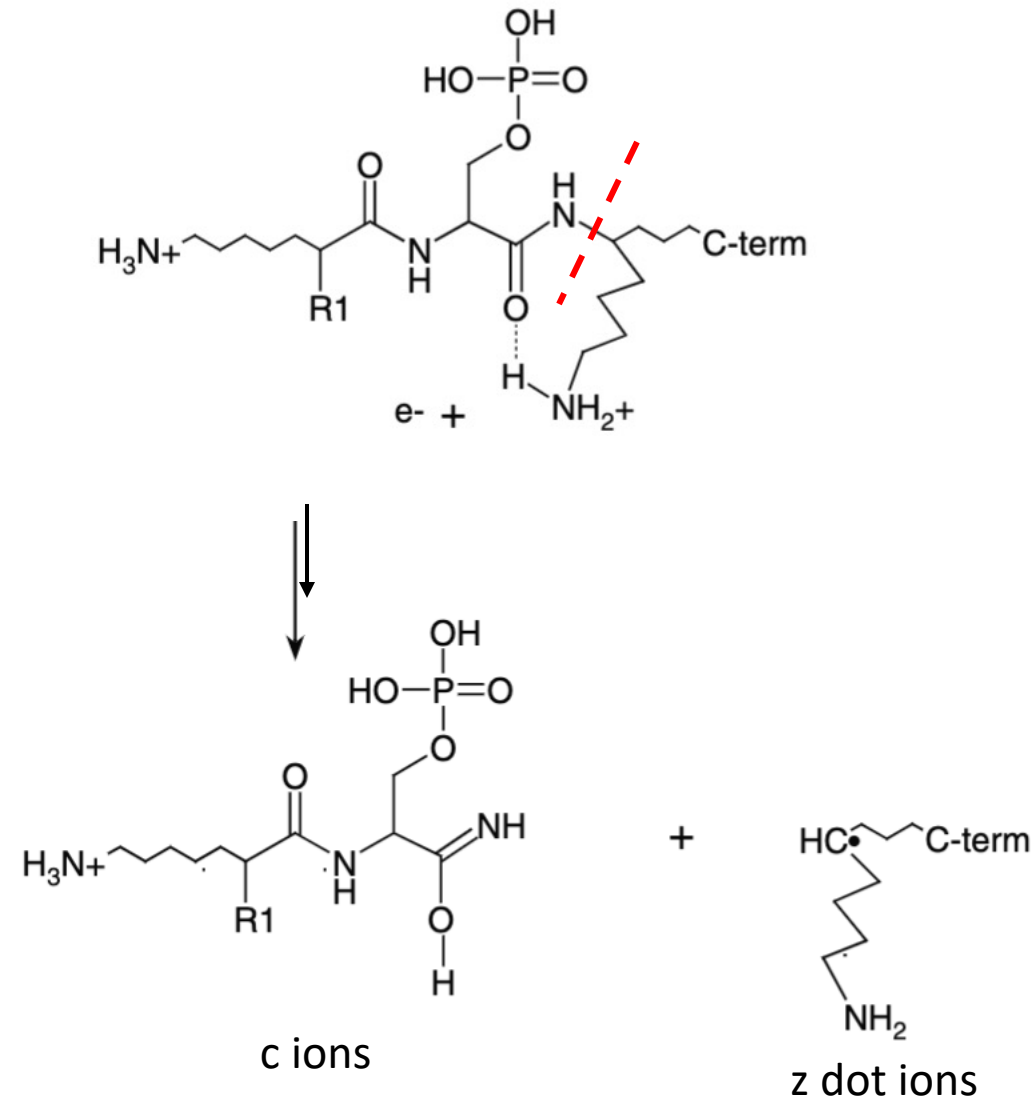
- QSSVTQVTEQSPK is phosphorylated at one position
 - precursor weight of unphosphorylated version + weight of phospho group
- Which fragment ions masses are changed in the MS² spectrum?

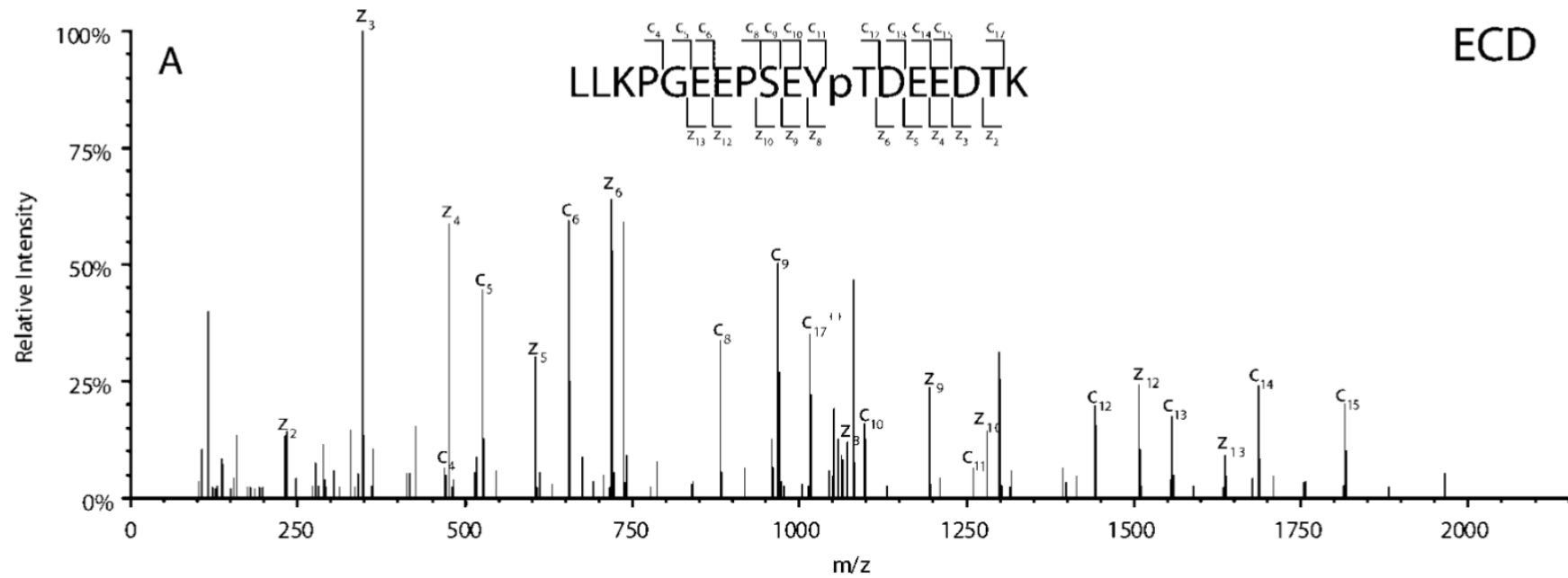




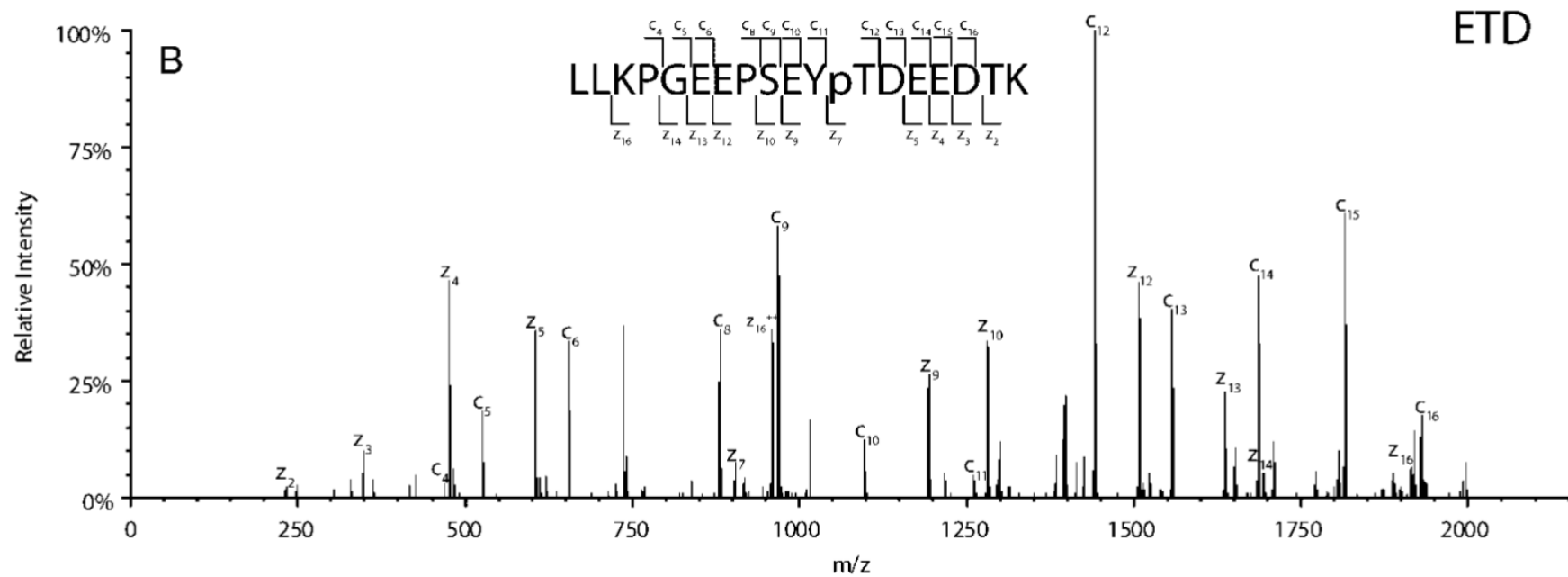
MS for Phosphopeptides: Electron-Driven Dissociation Methods

- As discussed above, fragmentation of phosphopeptides by CID is often dominated by neutral losses which can seriously hamper identification and phosphosite determination.
- **Electron-based dissociation methods (ECD/ETD)** have been introduced that leave posttranslational modifications such as phosphorylation and glycosylation intact during fragmentation.



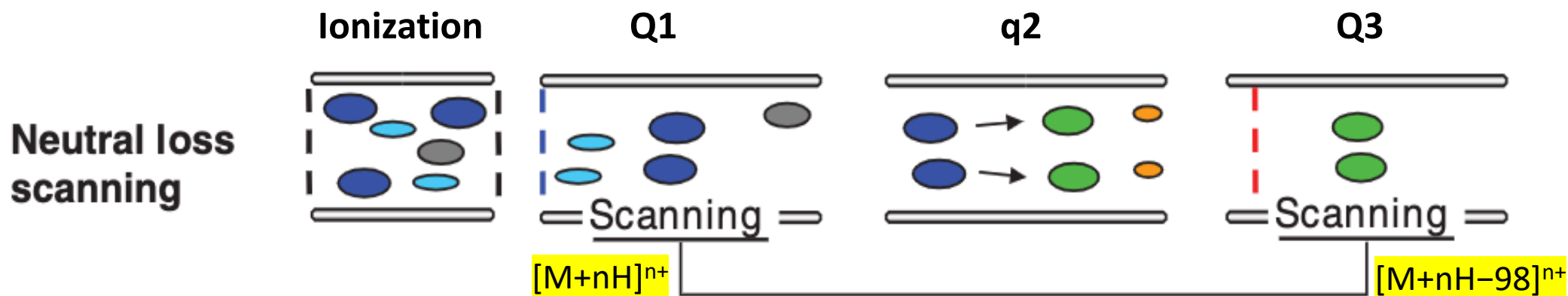


No neutral loss

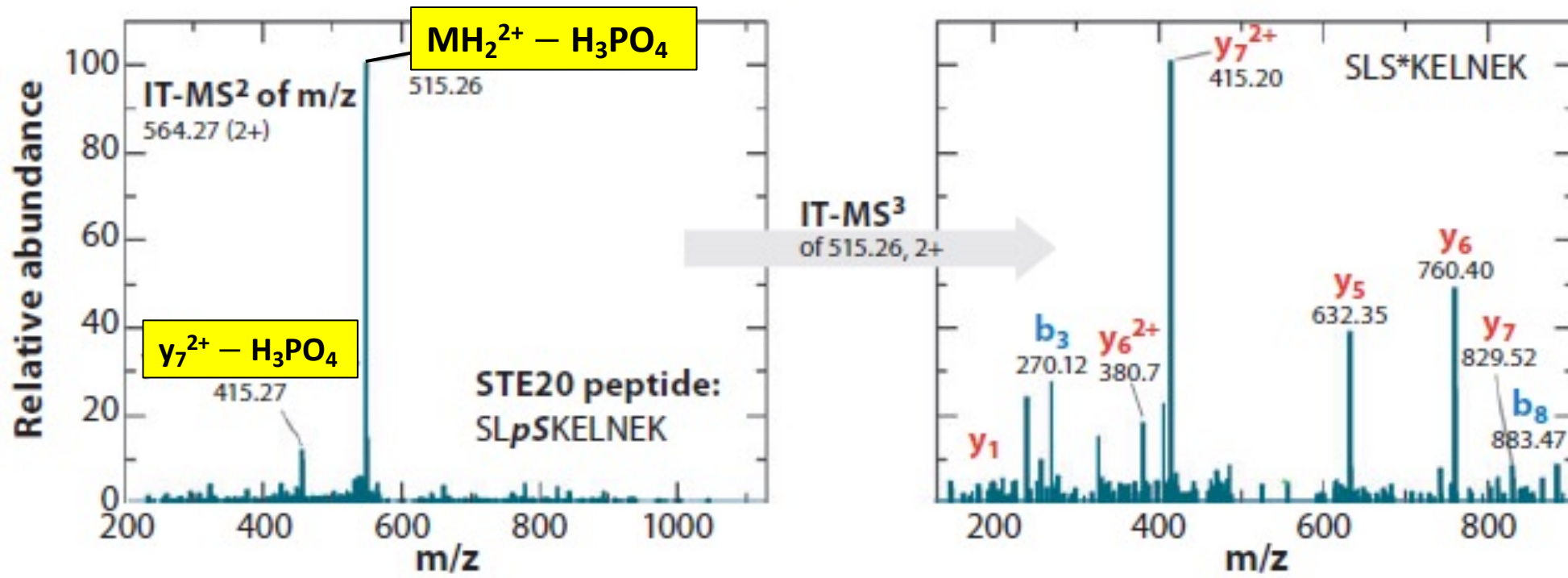


MS for Phosphopeptides: Neutral-Loss-Dependent MSⁿ

- Fragmentation of phosphopeptides by CID is often dominated by **neutral losses**
 - pS: predominantly H₂PO₄ (-98 Da)
 - pT: H₂PO₄ (-98 Da) and HPO₃ (-80 Da)
 - pY: HPO₃ (-80 Da) or (HPO₃ + H₂O) (-98 Da)
- **Neutral loss ion scanning** on a **triple quadrupole** mass spectrometer can be used to detect the loss of the phosphate group from phosphopeptides.
 - Q1 is used to scan the whole mass range
 - q2 is used as a collision cell
 - Q3 is set to scan at [M+nH-98]ⁿ⁺ to detect phosphopeptides that lose the phosphate group.



- When spectra are noisy or when backbone fragments have been suppressed by phosphate groups, an additional MS event (MS³) can be triggered at m/z [m/z precursor – neutral loss ($H_3PO_4 = 98$ Da)/ z]



MS for Phosphopeptides: Precursor/Reporter Ion Scanning

- Phosphopeptides preferentially yield diagnostic, **phosphate-specific fragment ions**, which indicates the presence of phosphopeptides.
 - H_2PO_4^- (97 Da)
 - PO_3^- (79 Da)
 - PO_2^- (63 Da)
- Since the phosphate group (above) provides a negative charge, phosphate-specific fragment ions are usually obtained in **negative ion mode**.
- **Higher collision energies** are used to generate these phosphate fragments
- The Precursor/Reporter Ion Scanning is usually carried out in triple quadrupole or more sensitive hybrid Q-q-TOF/Ion trap.
- The mass spectrometer can then be switched to **positive ion mode** to perform CID on only the peptide precursors that showed the loss that was scanned for.

MS for Phosphopeptides: Precursor/Reporter Ion Scanning

- Triple quadrupole (QqQ) instruments can be used to detect **diagnostic fragment ions** (e.g., at m/z of 79 for PO_3^-) using precursor ion scanning in negative mode
 - For the analysis of Tyr phosphorylation, the reporter ion scanning method is used to detect the **p-Tyr immonium ion** (cleavage at both sides of p-Tyr) at m/z **216.043** at positive mode. (very specific for p-Tyr!)

