COMPUTATIONAL PROTEOMICS AND METABOLOMICS

Oliver Kohlbacher, Sven Nahnsen, Knut Reinert

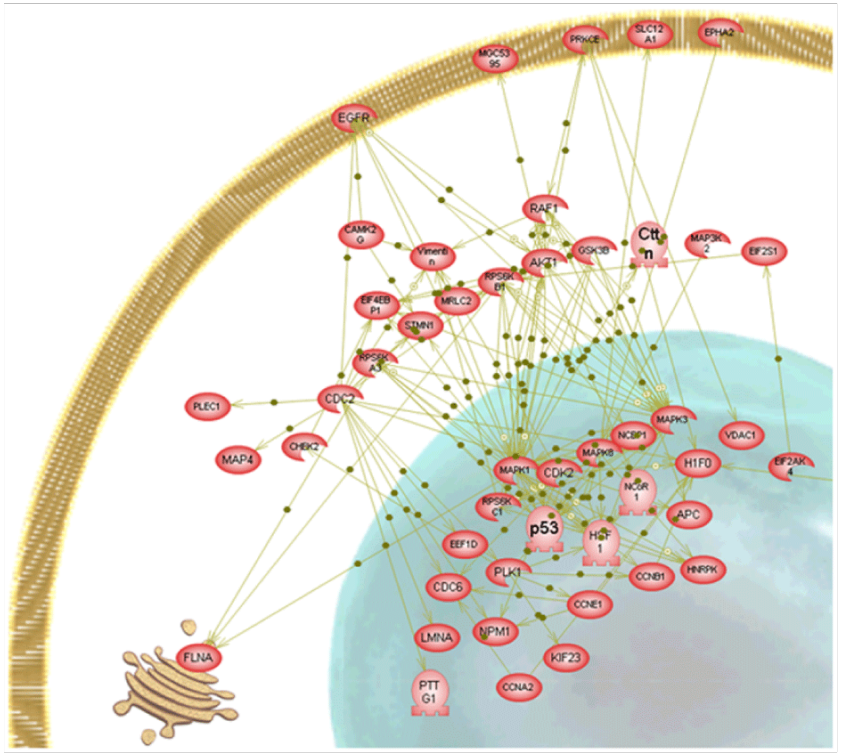
10. Post-translational modifications

LEARNING UNIT 10A
PROTEIN POST-TRANSLATIONAL MODIFICATIONS

- PTMs introduced by sample handling, in vivo PTMs
- Post-translational processing of proteins
- Phosphorylation, acetylation, glycosylation
- UniMOD
- Advantage of high mass accuracy
- PTMs increase size of search space

Motivation

- Cells can rapidly respond to stimuli and perturbations
- Important cellular mechanisms are tightly controlled
- Often, diseases (e.g. cancer) are due to aberrantly activated proteins
  - Protein expression is much too slow for quick adaption
  - PTMs are crucial regulator
  - MS-based proteomics allows to analyze complex networks of post-translationally modified proteins



http://www.omicsonline.org/ArchiveJPB/2010/April/03/JPB-03-113.php

Sample handling

Several Modification can be induced during the sample preparation

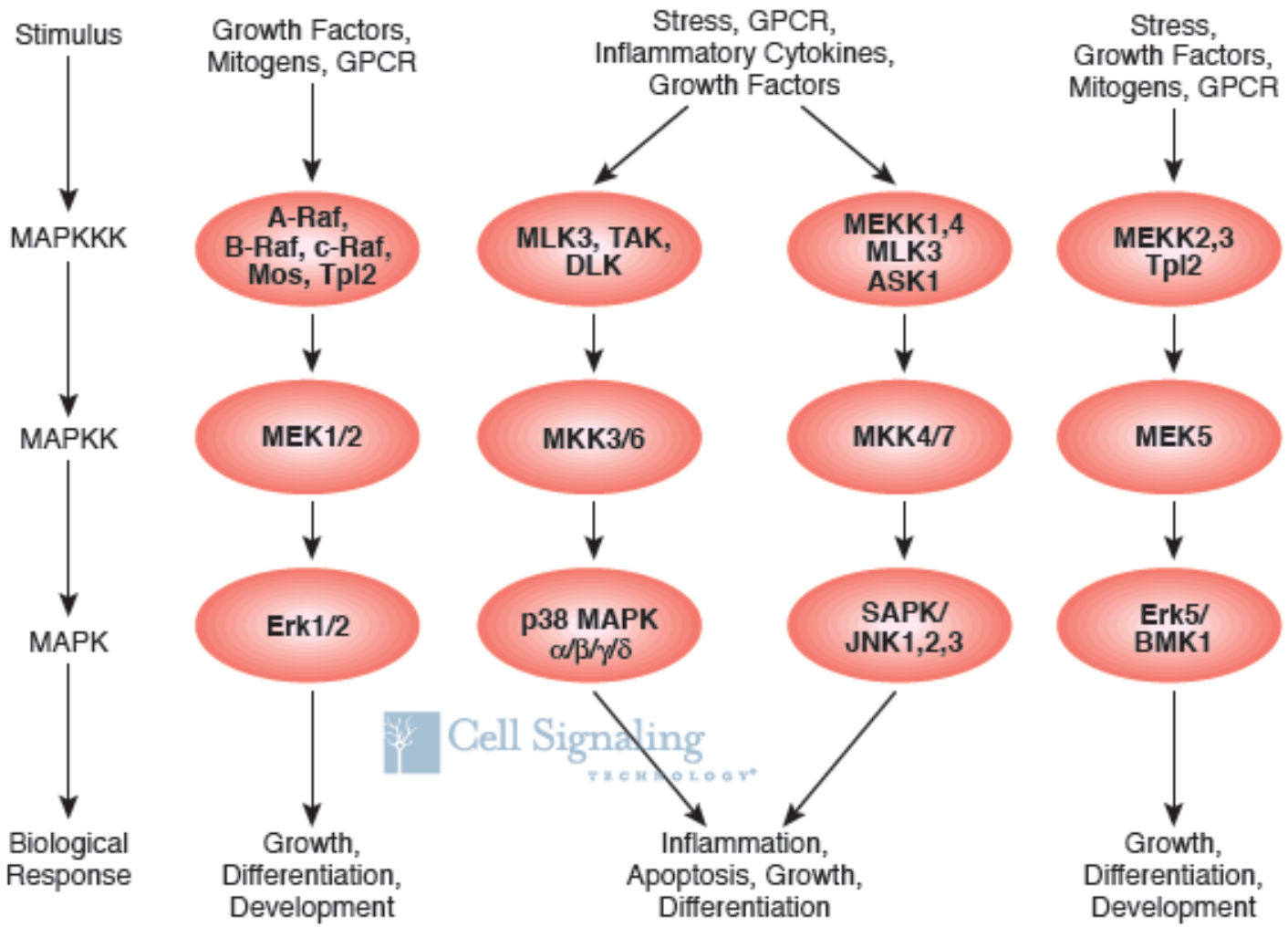
- Carbamidomethylation (Cys + 57 Da): protection of reduced sulfide groups with iodacetamide.
- Oxidation (Met + 16 Da): Exposure to air
- Pyro-Glu (N-terminal Glu – 17 Da): spontaneously
- Deamidation ([Asn – Gly] – 1 Da): spontaneously
- Sodium adducts (Asp, Glu + 22 Da) from salt
- Carbamylation (N-terminus and Lys + 43 Da): from metabolites of urea

- Note: the modification masses here are nominal masses

In vivo PTMs

- **Phosphorylation** (Ser, Thr, Tyr; +80 Da)
  - Phosphorylation is one of the most important PTMs
  - A key event in signaling
  - Catalyzed by kinases/phosphatases
- **Acetylation** (N-termini and Lys +42 Da)
  - often combined with removal of protein initial Met
- **Hydroxyprolination** (Pro; +16 Da)
  - stabilizes extra-cellular matrix on collagens
- **Ubiquitination** (Lys; +114 Da)
  - marks proteins for degradation

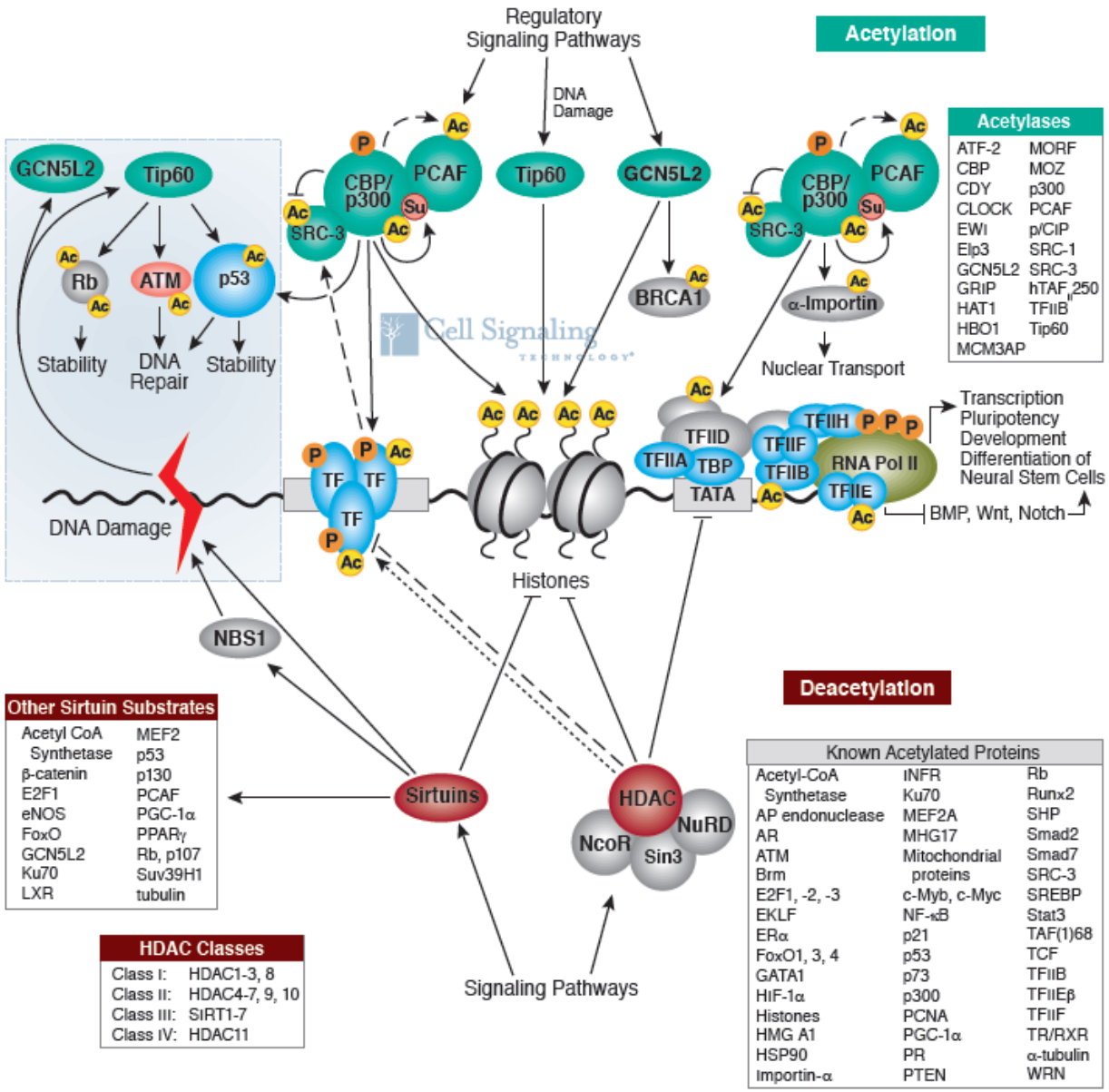
Phosphorylation – Example



http://www.cellsignal.com/reference/pathway/pdfs/MAPK_Cascades.pdf

Protein Acetylation

- Regulating chromatin structure and transcriptional activity
- Important role in immunity, circadian rhythmicity, and memory formation
- Favorable target in drug design for numerous disease conditions



Proteomics for PTMs

- Key issue in PTM analysis: modified proteins/peptides often have **low abundance**
- **Enrichment** can be used to increase their concentration in the sample
- The most established methodology is **phosphoproteomics**
- Development of phosphoproteomics has been driven by the deep interest in signalling research (pharma applications)
- Phosphopeptides are relatively easy to enrich and mass spectrometric methods have been well established
- The field is currently moving to high-throughput of profiling methods of other PTMs and the analysis of several PTMs in parallel

UniMOD – A Modification Database

Currently about 1,000 documented modifications

UNIMOD protein modifications for mass spectrometry

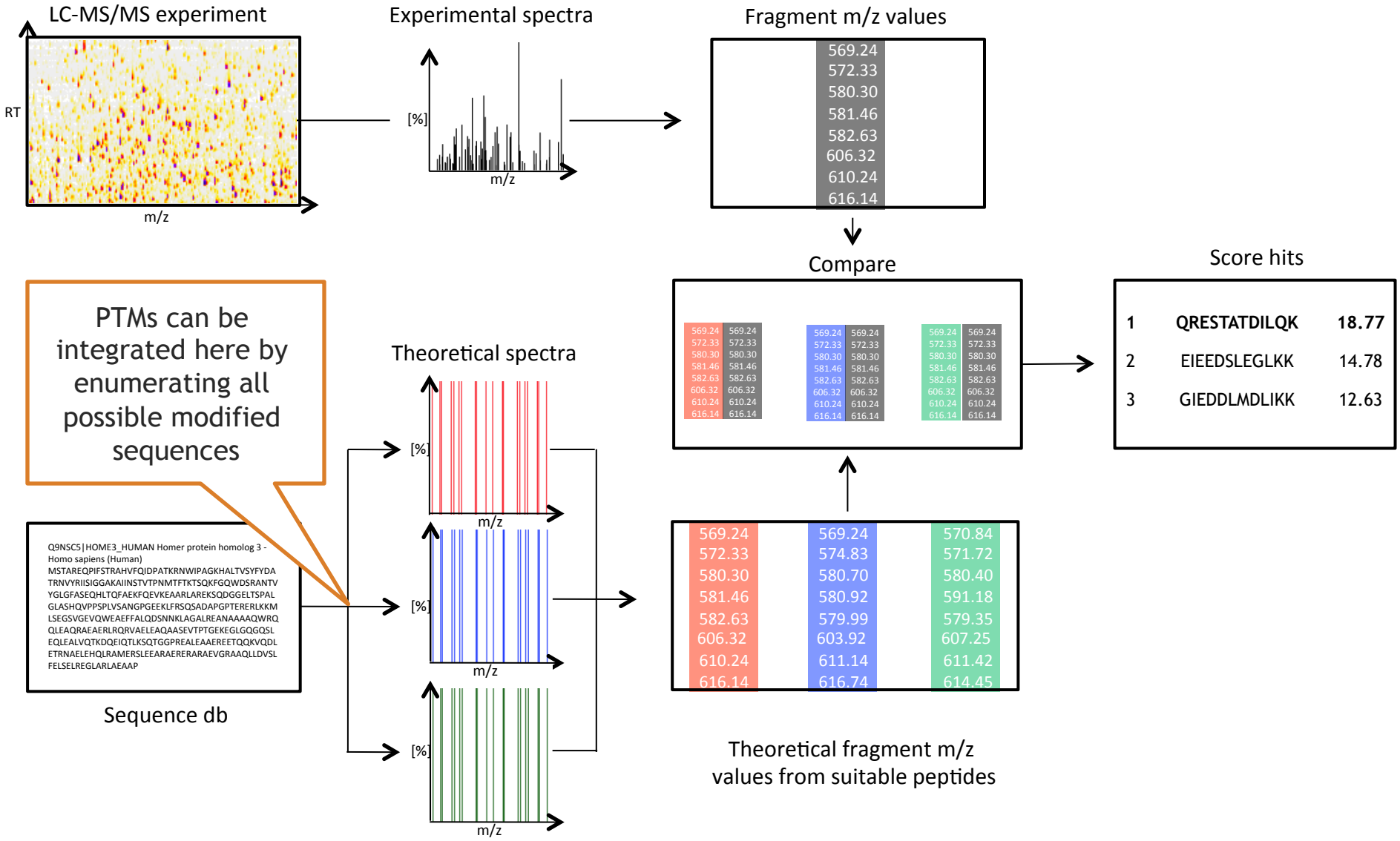
Unimod Logged as **Guest** Log out Change password Advanced search

Search for: Any field | Contains | Search | Show all | Details found: 1000 Page 1 of 50 | Records Per Page:: 20

	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 C-terminal K from Heavy Chain of MAb	-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)

www.unimod.org [accessed 01/19/2015]

Database Search for PTMs



Advantage of High Mass Accuracy

- Modifications often have similar masses
- Accurate precursor measurement improves discrimination of these possibilities

	Accession #	PSI-MS Name	Interim name	Description	Monoisotopic mass	Average mass	Composition
View	1	Acetyl	Acetyl	Acetylation	42.010565	42.0367	H(2) C(2) O
View	1197		Ser->Glu	Ser->Glu substitution	42.010565	42.0367	H(2) C(2) O
View	52	Guanidinyl	Guanidination	Guanidination	42.021798	42.0400	H(2) C N(2)
View	440	Amidino	amidino	amidino	42.021798	42.0400	H(2) C N(2)
View	37	Trimethyl	tri-Methylation	tri-Methylation	42.046950	42.0797	H(6) C(3)
View	575		Gly->Val	Gly->Val substitution	42.046950	42.0797	H(6) C(3)
View	1047		Ala->Xle	Ala->Leu/Ile substitution	42.046950	42.0797	H(6) C(3)
View	1305		Propyl	Propyl	42.046950	42.0797	H(6) C(3)
View	1163		Asn->Arg	Asn->Arg substitution	42.058184	42.0830	H(6) C(2) N(2) O(-1)

www.unimod.org [accessed 01/19/2015]

PTMs Increase Search Space

- PTMs are considered variable modifications for database search: they can be present at a certain amino acid, but do not have to be
- Each PTM position thus increases the search space: the original sequence and the modified sequence need to be generated
- Search space is further increased by nearby PTMs – multiple PTMs within one (tryptic) peptide lead to an exponential number of sequence

D I G S E S T E K

D I G S*E S T E K

D I G S E S*T E K

D I G S E S T*E K

D I G S*E S*T E K

D I G S*E S T*E K

D I G S E S*T*E K

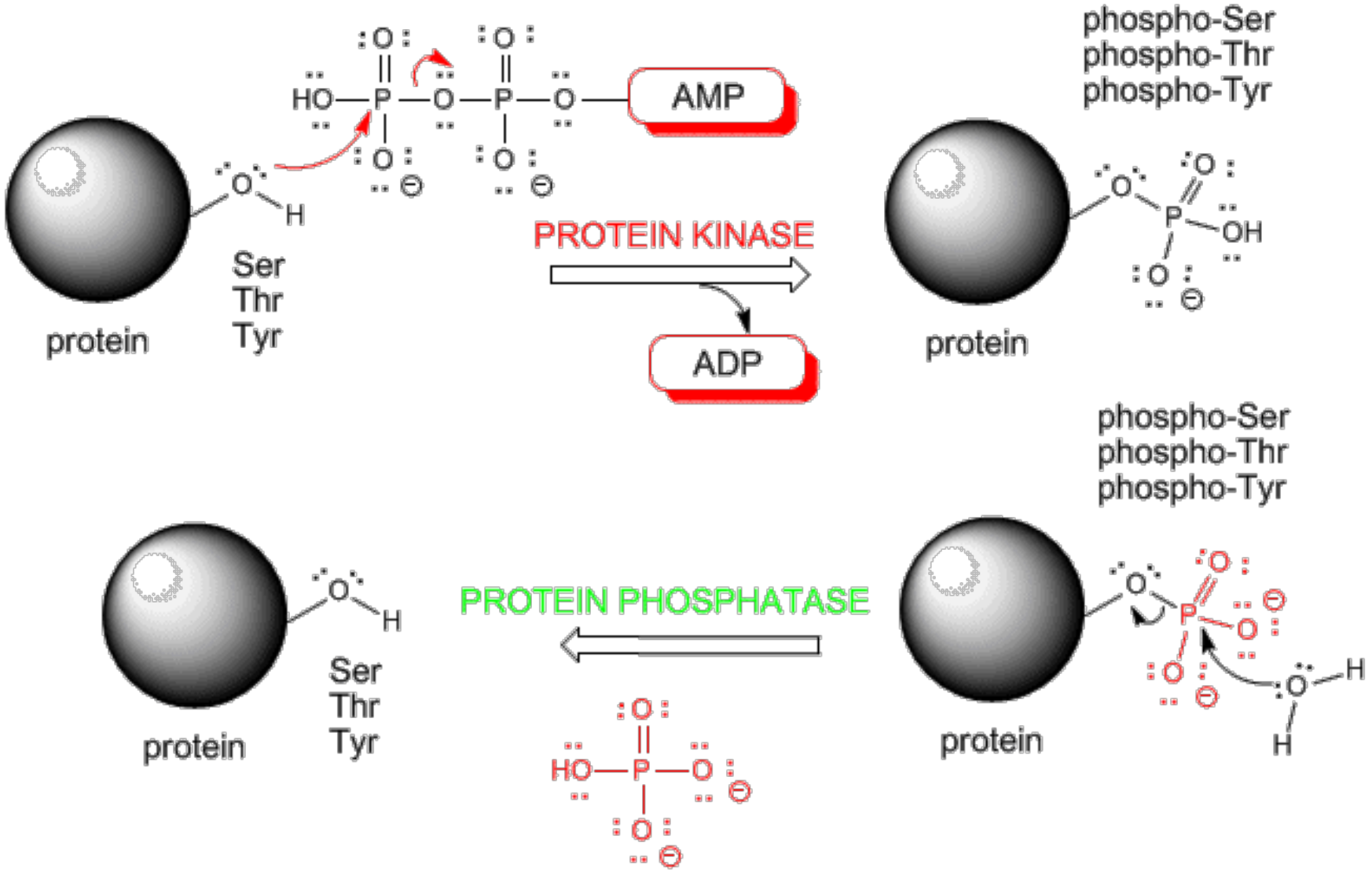
D I G S*E S*T*E K

Example:
This peptide contains three potential phosphorylation sites (marked by asterisks). We thus obtain $2^3 = 8$ potential peptide sequences.

LEARNING UNIT 10B
EXPERIMENTAL METHODS FOR PHOSPHOPROTEOMICS

- Typical phosphoproteomics workflows
- Enrichments techniques
- MS techniques

Phosphoproteomics

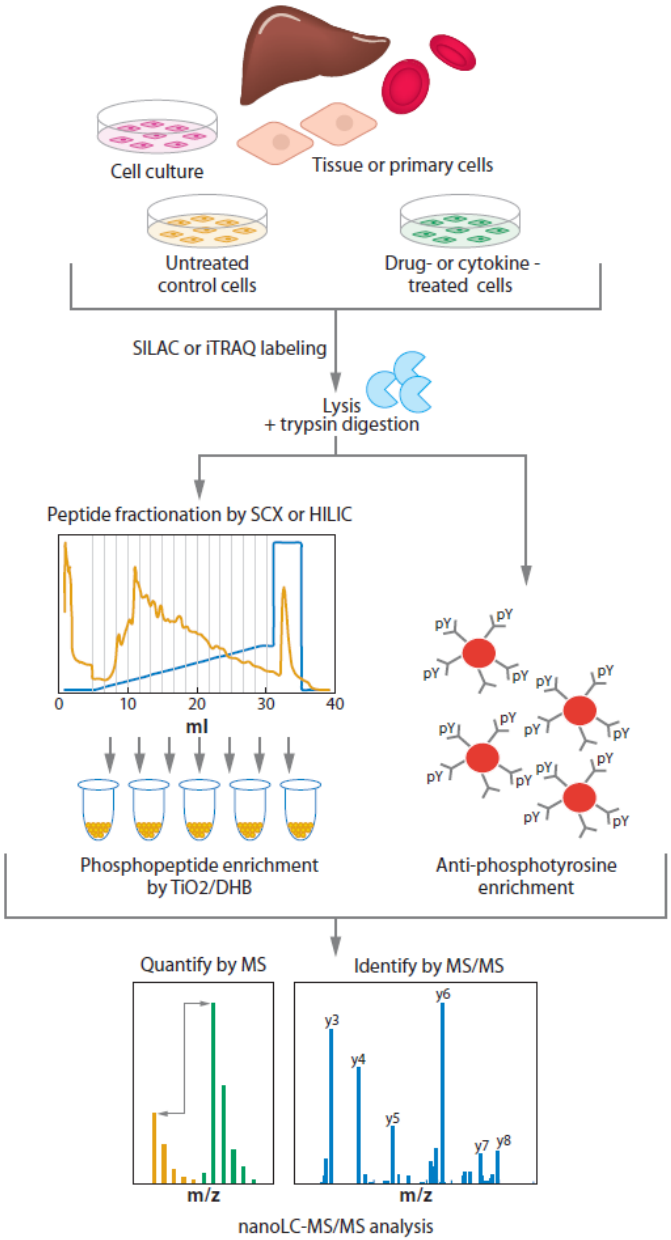


- http://employees.csbsju.edu/hjakubowski/classes/Chem%20and%20Society/Signal_Transduction/cellsdifferentsig.htm

Problems of Analyzing Phosphoproteins

- Site-specific phosphorylation is often substoichiometric, thus, phosphopeptides represent a small proportion of all peptides present in a total cell lysate
- Biochemical enrichment strategies have been developed. The most important methods include affinity- and antibody-based methods
- The mass spectrometric analysis of phosphopeptides is difficult since phospho groups might get lost during fragmentation or they suppress other ions

Phosphoproteomics Workflow



Sample preparation

↓

Quantification strategy

Digestion

Phosphopeptide enrichment

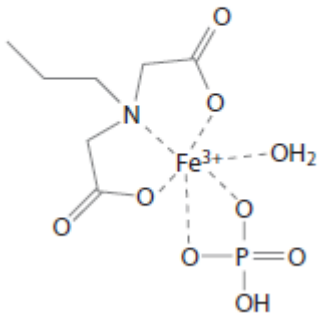
Mass spectrometric measurement

Macek et al, Annu Rev Pharmacol Toxicol. 2009;49:199-221. PMID: 18834307

Phosphopeptide Enrichment

- Immobilized metal-affinity chromatography (IMAC)
  - Phosphates have high affinity to trivalent metal ions
  - Metal ions are immobilized on columns
  - A variety of metals has been used, including Fe^{3+}, Ga^{3+}, Al^{3+} and Zr^{3+}
- Problem with IMAC
  - If the pH during the loading is out of [2,3.5], then non-phosphopeptides bind as well
  - Strongly acidic peptides (rich in E and D) are also affine to the metal complexes

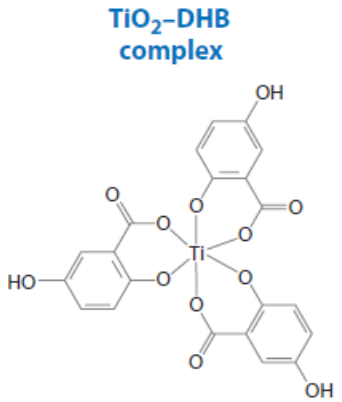
IMAC-Fe^{3+} phosphate complex



Macek et al, Annu Rev Pharmacol Toxicol. 2009;49:199-221. PMID: 18834307

Titanium Dioxide (TiO_2) enrichment

- 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



Macek et al, Annu Rev Pharmacol Toxicol. 2009;49:199-221. PMID: 18834307

Strong Cation Exchange

- 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 -> 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

Antibody-based Enrichment

- Immunopurification with immobilized anti-phospho-Y antibodies
- Antibody-based method is well established for Y, but not for other residues
- It is limited in throughput and hard to automate

MS for Phosphopeptides

Detection of phosphopeptides by MS is difficult, because:

- Phosphopeptides are very low abundant
- They have low MS response values
- They show inadequate fragmentation patterns

That is why alternative methods emerged:

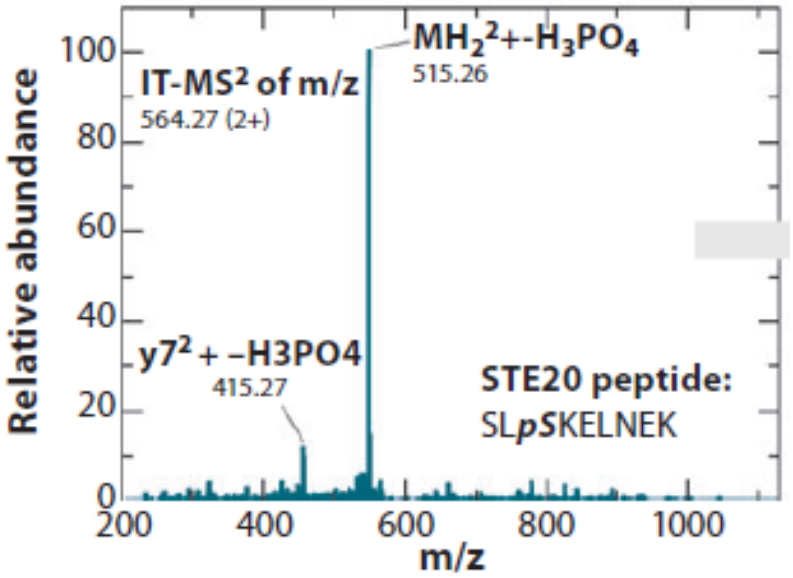
1. Precursor ion scanning and reporter ions
2. Neutral loss dependent MS^n
3. Alternative fragmentation methods (e.g ETD)

Precursor/ Reporter ion scanning

- QQQ instruments can be used to detect diagnostic fragment ions at m/z 79 (HPO_3-) using precursor ion scanning in negative mode
  - Very sensitive, however, fragment spectra recorded in negative mode are of poor quality
  - Switching between ionization modes takes time
- For the analysis of Y-phosphorylation, the reporter ion scanning method is used to detect the pY immonium ion (cleavage at either side of pY) at m/z 216.043 (very specific for pY!)
- The pY immonium ion is mass deficient due to the high content of O and P, thus high resolution instruments can easily distinguish between ions with the same nominal mass

Neutral-Loss-Dependent MS^n

- Fragmentation (using CID) occurs via the lowest energy dissociation pathway (e.g. the O-P bond in S and T phosphorylated peptides)
- Poor coverage of peptide backbone ions in the MS/MS spectrum



On modern mass spectrometers an additional MS event (MS3) can be triggered at m/z [m/z precursor – neutral loss (98 Da)/ z]

- Macek et al, Annu Rev Pharmacol Toxicol. 2009;49:199-221. PMID: 18834307

LEARNING UNIT 10C
ALGORITHMS FOR PHOSPHOPROTEOMICS

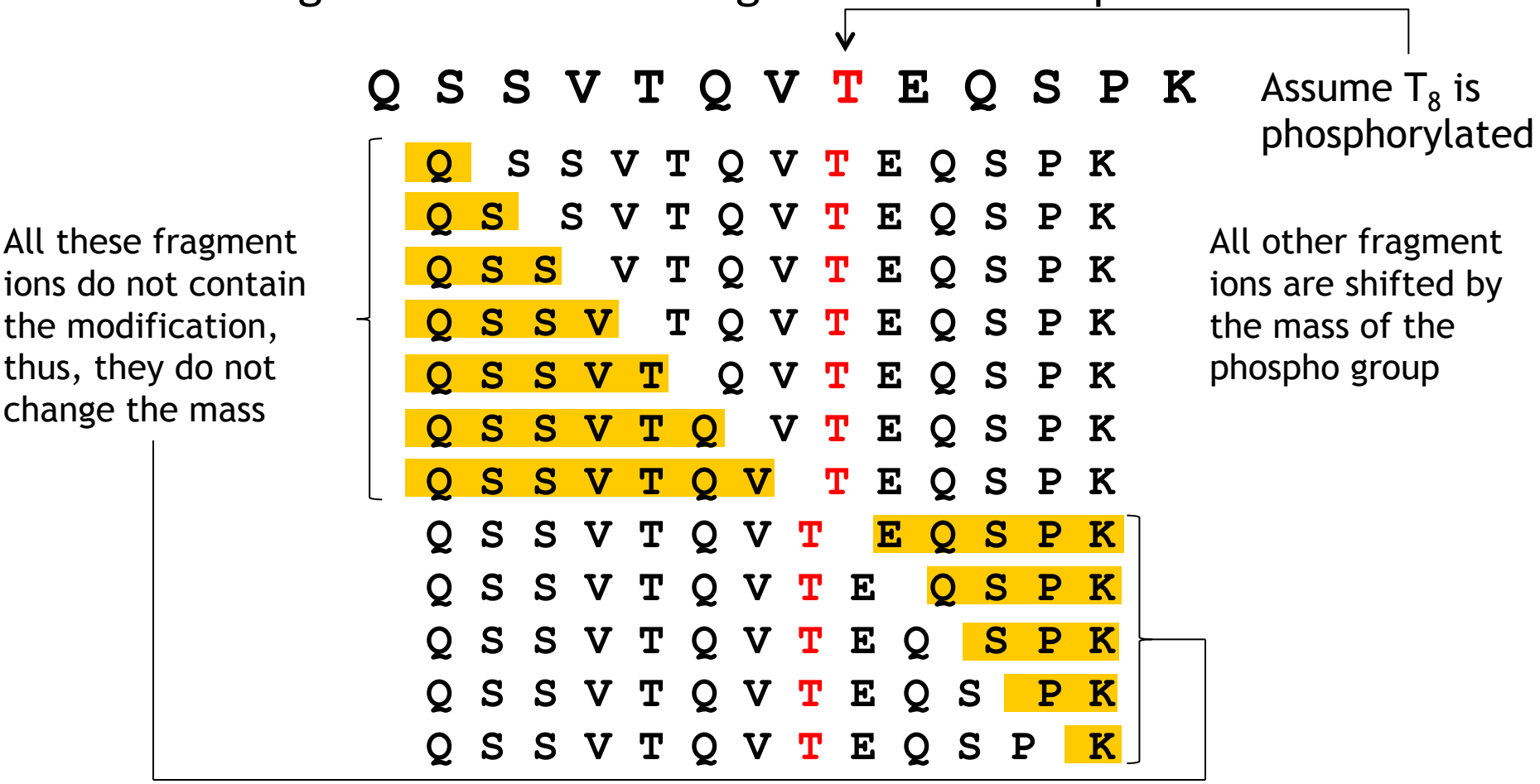
- Site assignment in phosphoproteomics
- Ascore algorithm
- Phosphosite stoichiometry

Phosphopeptide Identification

- Precise phosphorylation site assignment can be difficult
- Search engines often suggest the correct phosphopeptide, but fail to correctly localize the phosphorylation site, if more than one residue can potentially carry the phospho group
  - Note that in eukaryotes phospho groups are predominately attached the S, T and Y residues (other phosphorylated amino acids exists (H,L or R), but are very rare)

Phosphorylated Example

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



Phosphosite Localization

- The **Ascore** algorithm has been published in 2006 by the Gygi lab
- Two step algorithm:
    i. Determine the most likely site locations
    ii. Use *site-determining ions* to calculate the probability for correct assignment
- One of the first algorithms addressing the problem of modification site localization
- Similar version is also implemented in MaxQuant
- OpenMS implementation (OpenMS::TOPP::PhosphoScoring)

Beausoleil et al., Nat Biotechnol. 2006 Oct;24(10):1285-92. PMID: 16964243

Ambiguous Localization

- Example: 13 residue peptide from *Zinc finger protein 638*
- Multiple possibilities for phosphosite assigment
- Phosphorylation *is allowed* (by search settings) on STY

NH_2 - QSSVTQVTEQpSPK-COOH

or ?

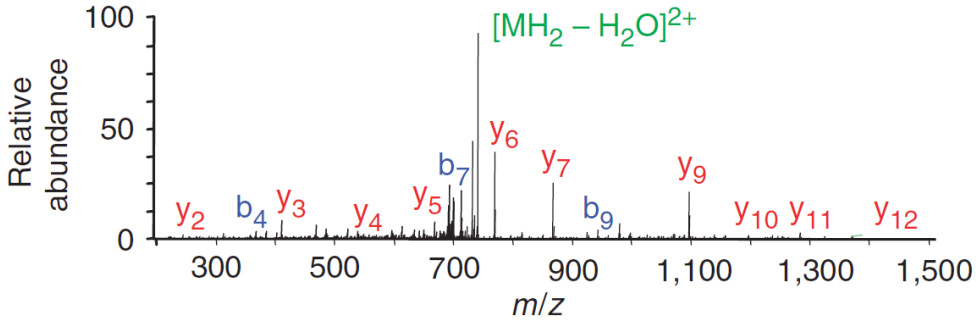
NH_2 - QSSVTQVpTEQSPK-COOH

- Which one is the correct phospho site?

Beausoleil et al., Nat Biotechnol. 2006 Oct;24(10):1285-92. PMID: 16964243

Ascore Algorithm

(i) Determine the most likely phosphorylation site



- MS/MS spectrum is separated into windows of 100 m/z units
- Retain *i*, with $i \in [1, .., N]$, of the most intense peaks per spectrum
- Predicted *b* and *y* ions are then overlaid with the processed spectrum
- The cumulative binomial probability *P* is then calculated using the number of trials *N*, the number of successes n and the probability of success *p*:

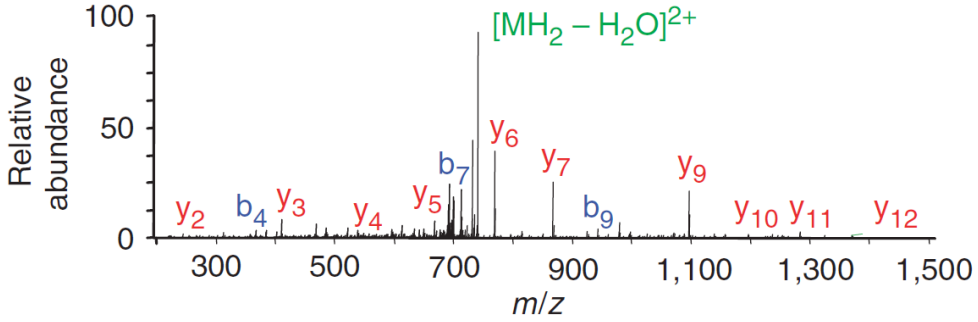
$$P(x) = \sum_{k=n}^{N} \binom{N}{k} p^k (1-p)^{N-k}$$

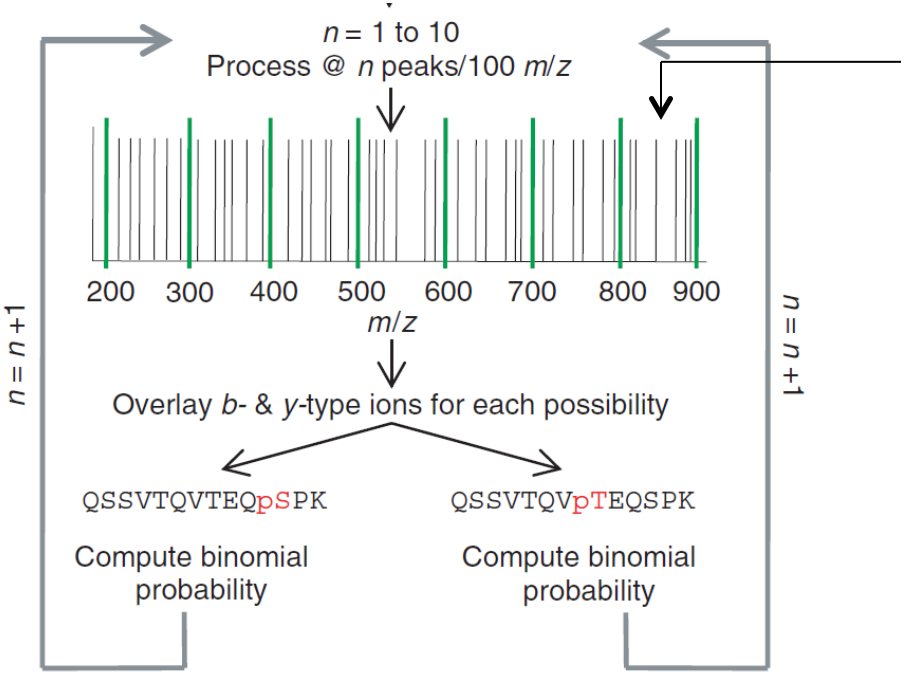
P is the probability for random matchings of the given number of fragment ions; the total number of trials (*N*) equals the number of fragment ions; the number of successes (*n*) is the number of matches

Beausoleil et al., Nat Biotechnol. 2006 Oct;24(10):1285-92. PMID: 16964243

Ascore Algorithm

(i) Determine the most likely phosphorylation site





(i) Retain top *N* peaks per window

(ii) This is repeated for 1 to N (here 10) ions per window

(iii) The cumulative binomial probability uses the number of trials (all *b* and *y* ions) and the number of successes (matched ions)

Beausoleil et al., Nat Biotechnol. 2006 Oct;24(10):1285-92. PMID: 16964243

Ascore Algorithm

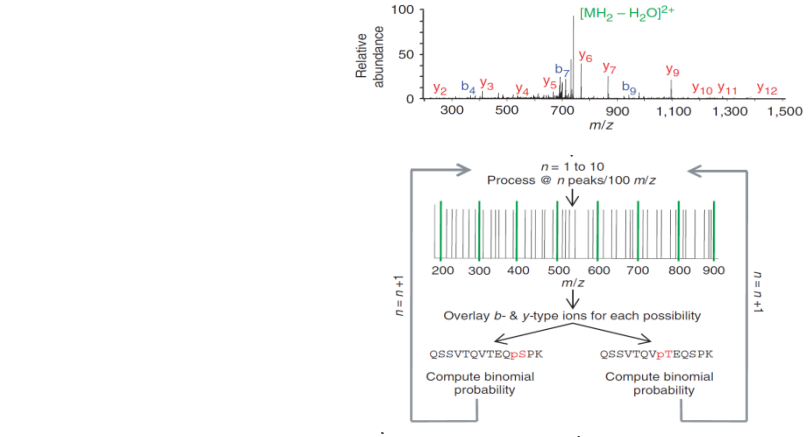
- Within a given window, the probability of matching a peak is $\frac{i}{100}$ (this is *p*)
- The final peptide score is calculated with:

$$Score = -10\ log(P)$$

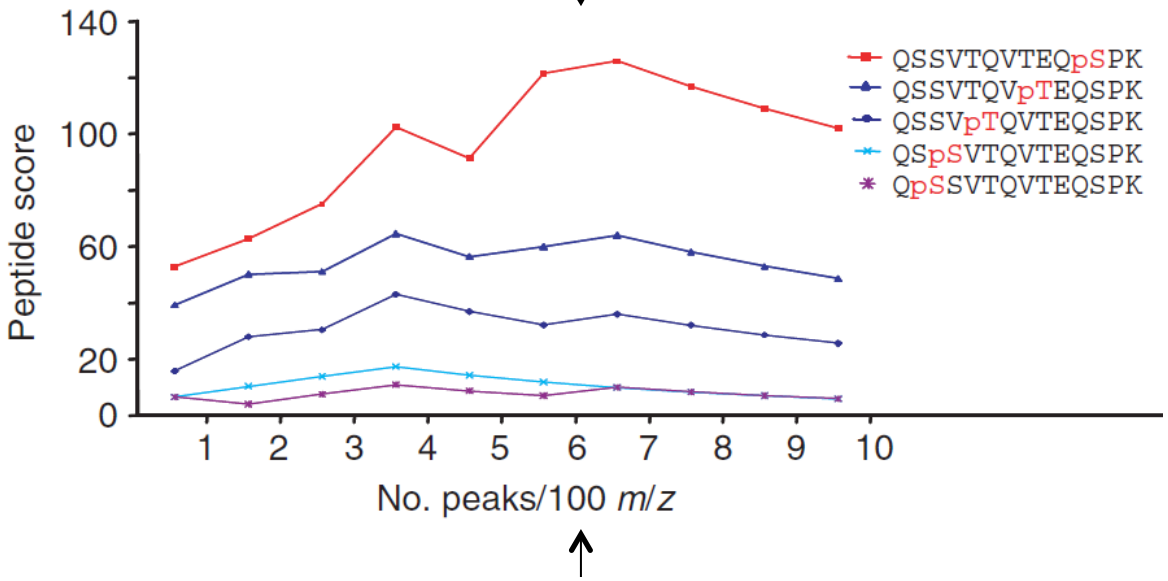
and the ambiguity score (Ascore) is

Ascore = Score(top hit) - Score(2nd hit)

Ascore Algorithm



(iv) Each possibility is plotted as -10 log(P) vs. peak depth (peaks per m/z window)
(v) Now the *site-determining* ions can be identified. Here only six ions can potentially differentiate between the 2 sites. The corresponds to the earliest maximal difference

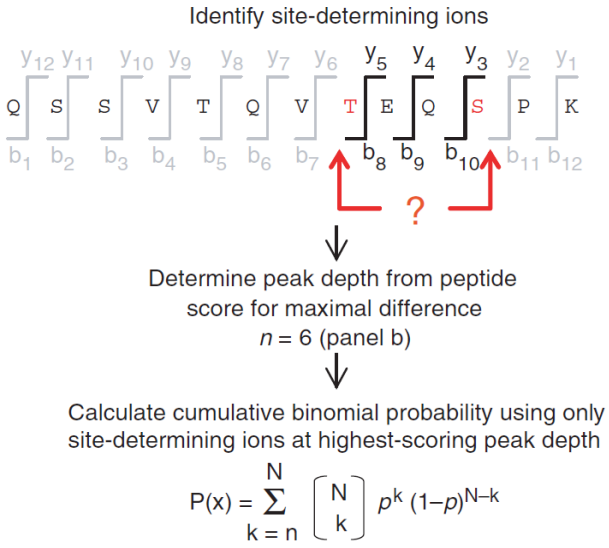


Earliest maximal difference

Beausoleil et al., Nat Biotechnol. 2006 Oct;24(10):1285-92. PMID: 16964243

Ascore Algorithm

- The calculate the final Ascore only the two top scoring sequences (based on the peptide score) are used



Phosphopeptide	QSSVTQVTE QpSPK	QSSVTQ VpTEQSP K
Trials (*N*)	6 (y_3, y_4, y_5, b_8, b_9, b_{10})	6 (y_3, y_4, y_5, b_8, b_9, b_{10})
Successes (*n*)	5 (y_3, y_4, y_5, b_9, b_{10})	0
p (6 peaks / 100 *m/z*)	0.06	0.06
P	0.0000044	1.0
Score [$-10 \times \log(P)$]	53.57	0
Ascore = ambiguity score (difference of the top two candidates)	53.57 – 0 = 53.57	

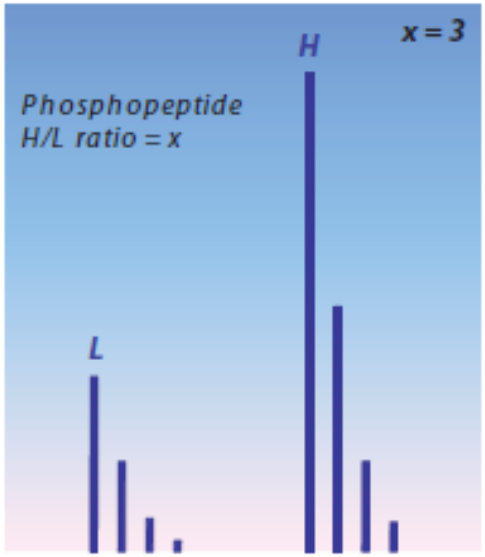
Beausoleil et al., Nat Biotechnol. 2006 Oct;24(10):1285-92. PMID: 16964243

Phosphosite Stoichiometry

- In quantitative (phospho)proteomics studies it is often important to differentiate between differential protein expression and differential protein phosphorylation
- Note that phosphorylation can be a very rapid event (in the range of seconds and minutes), whereas protein expression takes more time (hours). However, differential phosphorylation can still be present after longer periods.

Differential Analysis

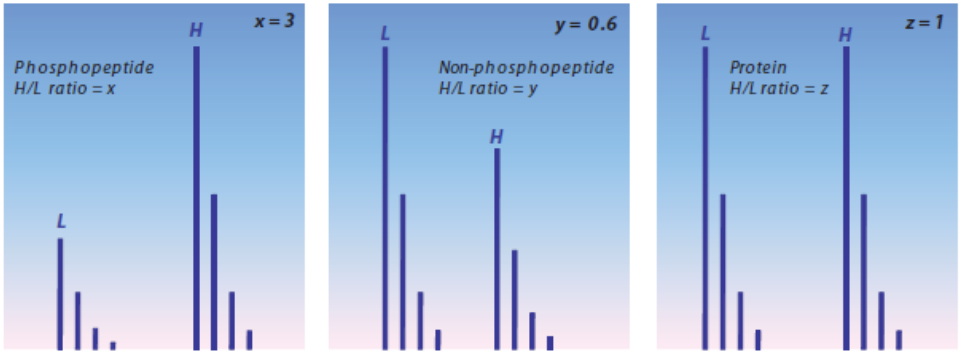
- This is a SILAC pair of a phosphorylated peptide



- Is the protein of origin more abundantly expressed in the state with the heavy (H) label?
- Is the phosphosite more abundantly occupied in the H state?

Phosphosite Stoichiometry

- Olsen et al. published a strategy for the calculation of phosphosite occupancy (stoichiometry)
- This information can only be inferred if non-phosphorylated form of the peptide has also been identified and if the protein ratio has been calculated



- In the initial paper SILAC was used as a quantification approach

Olsen et al., 2010. *Science Signaling* 3 (104), ra3. [DOI: 10.1126/scisignal.2000475]

Phosphosite Stoichiometry

- *Known:* The ratio (heavy SILAC/light SILAC) of the phosphorylated peptide as *x*, the non-phosphorylated peptide as *y* and the ratio of the protein (median of all ratios of peptides assigned to the same protein) as *z*
- *Unknown:* Absolute phosphorylation site stoichiometry in the L and H state
  - $N_L(Phos)$ and $N_L(Pep)$ are the amounts (copy numbers) of the phosphorylated and non-phosphorylated version of the same peptide. The proportion of phosphorylated peptide in the L SILAC state is given by a,

$$a = \frac{N_L(Phos)}{N_L(Pep)}$$

  - In the H state as b, respectively,

$$b = \frac{N_H(Phos)}{N_H(Pep)}$$

Phosphosite Stoichiometry

- The protein copy number *N(Prot)* is modeled by the copy number of the peptides, we can thus assume

$$\frac{N_H(Phos)+N_H(Pep)}{N_H(Prot)} = \frac{N_L(Phos)+N_L(Pep)}{N_L(Prot)}$$

- So far these are theoretical considerations, but the MS data delivers:

$$x = \frac{N_H(Phos)}{N_L(Phos)}$$

$$y = \frac{N_H(Pep)}{N_L(Pep)}$$

$$z = \frac{N_H(Prot)}{N_L(Prot)}$$

Phosphosite Stoichiometry

- This information allows to calculate the phosphopeptide proportions as follows:

$$a = \frac{N_L(Phos)}{N_L(Pep)} = \frac{z-y}{x-z}$$

$$b = \frac{N_H(Phos)}{N_H(Pep)} = \frac{x(z-y)}{y(x-z)}$$

- Furthermore, with the assumption,

$$N_H(Phos) + N_H(Pep) = N_L(Phos) + N_L(Pep) = 1$$

the exact occupancies for light and heavy are calculated as,

$$\frac{a}{1+a} \text{ and } \frac{b}{1+b}$$

LEARNING UNIT 10D
GLYCOPROTEOMICS

- Function and relevance of glycosylation
- Glycan structure
- Glycan enrichment
- MS methods
- Glycan identification

Glycoproteomics

glyco- or before a vowel ***glyc-***

combining form -- indicating sugar: glycogen. [from Greek *glukus* sweet]

Glycoproteomics thus analyzes the totality of **glycoproteins** (i.e., molecules consisting of a protein part and a sugar part [**glycan**]).

http://graphics8.nytimes.com/images/2012/11/18/magazine/18WMT-t_CA0/mag-18WMT-t_CA0-superJumbo.jpg

Glycoproteomics

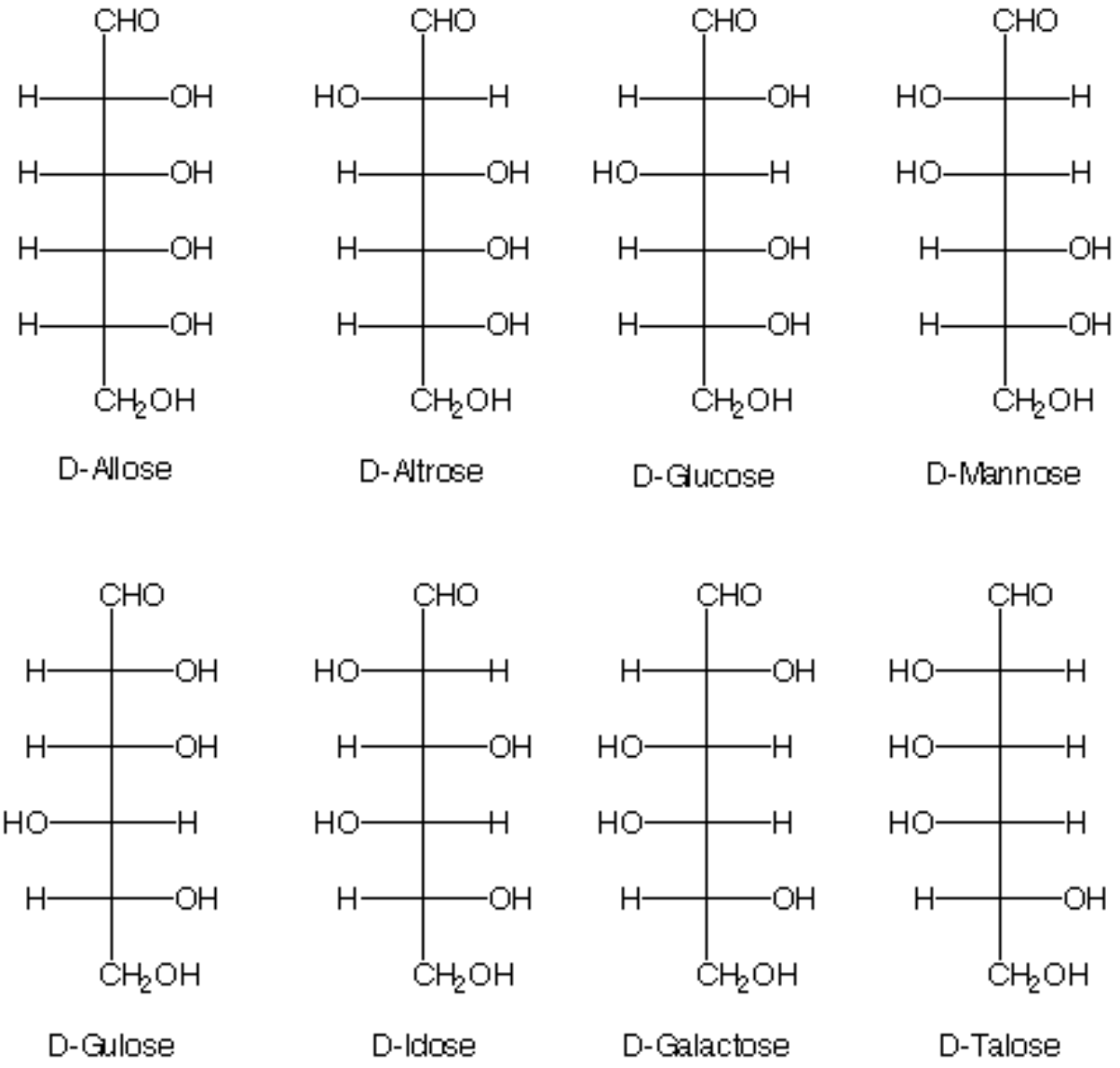
- Glycosylation is the most complex of all PTMs
- Glycans (polysaccharides) represent a huge variety in their composition
- Glycosylation of proteins takes mainly place in the endoplasmic reticulum (ER) and the Golgi
- There, enzymes, such as glycosyltransferases and glycosidases attach the sugar groups to the proteins
- These enzymes mainly target S, T and N
- It impacts on charge, conformation and protein stability

Importance of Glycoproteomics

- Abnormal protein glycosylation has been correlated with numerous diseases
  - Cancer
    - Aberrant glycosylation can serve as biomarker for certain cancers
  - Inflammatory diseases
  - Neurodegenerative diseases
  - …

Kim et al.. Protein Pept Lett. 2009;16(5):499-507. PMID: 19442229
Brooks. s. Mol Biotechnol. 2009 Sep;43(1):76-88. Epub 2009 Jun 9. PMID: 19507069

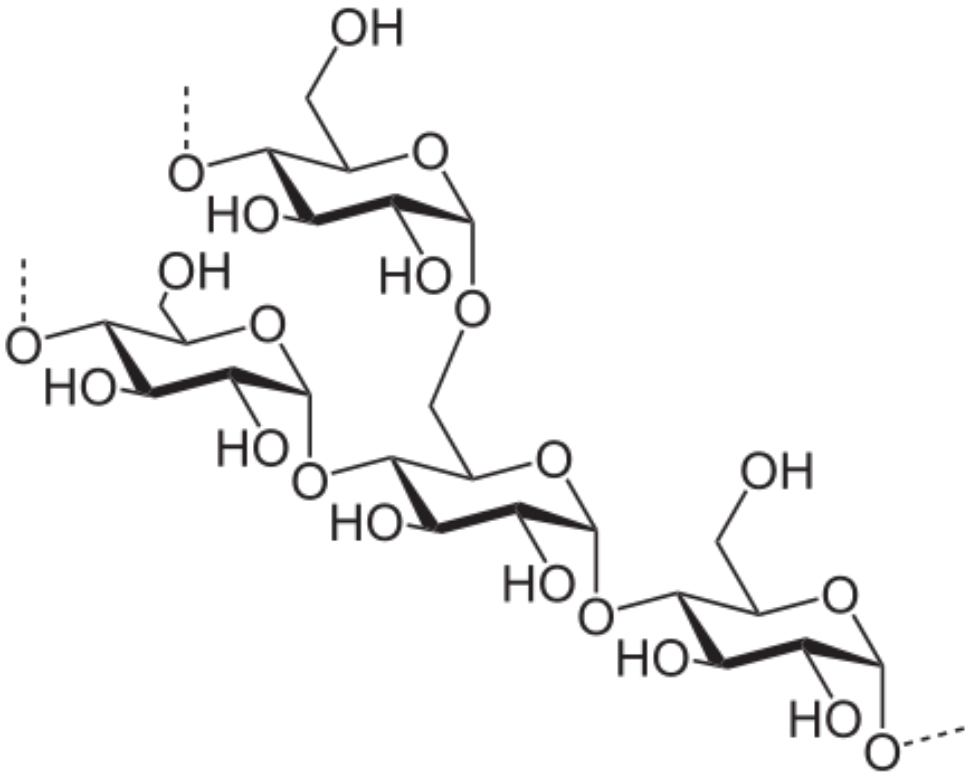
Monosaccharides – Hexoses



http://tonga.usp.edu/gmoyna/biochem341/hexoses.gif

Monosaccharides of Glycoproteins

- In contrast to other biomacromolecules (proteins, NAs), sugars can form **branched structures**
- This



http://www.genomebc.ca/files/9612/7446/9353/6.6.1.3%20branched%20sugar%20image.png

Glycan Monomers

- Human glycans are composed of eight different monosaccharides:
  - Mannose (Man)
  - Glucose (Glc)
  - Galactose (Gal)
  - Fucose (Fuc)
  - N-acetylgalactosamine(GalNAc)
  - N-Acetylglucosamine (GlcNAc)
  - N-Acetylneuraminic acid (NeuNAc)
  - Xylose (Xyl)

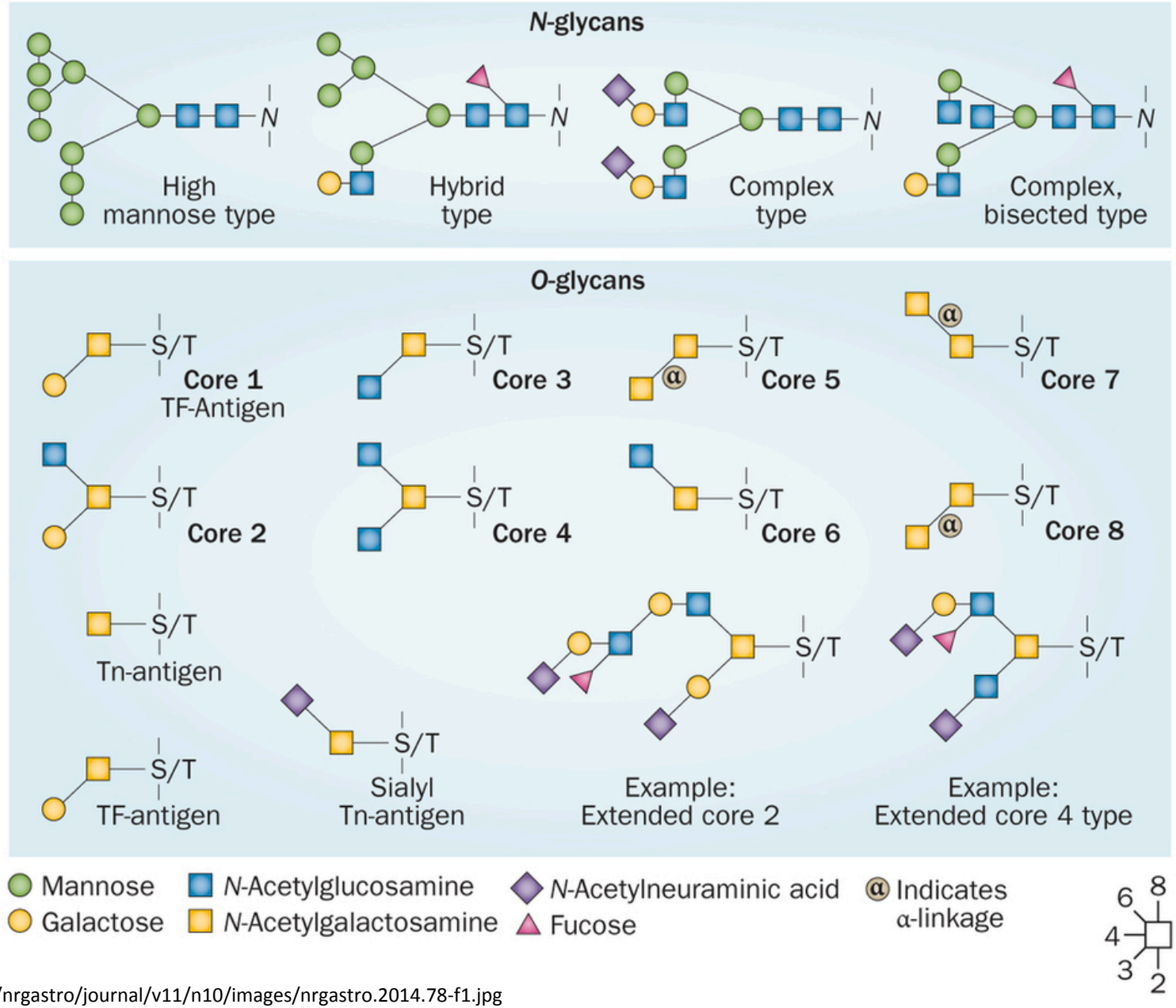
Glycoproteome Analysis

- For the complete glycoprotein picture one needs: the peptide sequence, glycosylation site and the glycan structure
- Glycosylation does not conform to a single structure and there is no underlying template
- These sugars can be linked in linear or branching chains of various sequences and lengths
- Glc, Gal and Man have identical masses and charges
- These are different stereoisomers and their permutations in complex glycans can result in a broad range of different glycoforms

Protein-linked Glycans

- N-linked glycans, with the glycan attached to the amide group of N via a GlcNAc in a consensus sequence N-X-S/T (X any amino acid except P)
- O-linked glycans, with glycan attached to S or T
- C-glycans, with the glycan (Man) attached to W with a C-C bond in a consensus sequences W-X-X-W or W-S-X-C
- Glycosylphosphatidylinositol anchors, with the glycan attached to the protein C-terminus by a phosphoethanolamine bridge with Man (this occurs only on membrane associated proteins)
- **The two most common forms are O- and N- linked glycosylation**

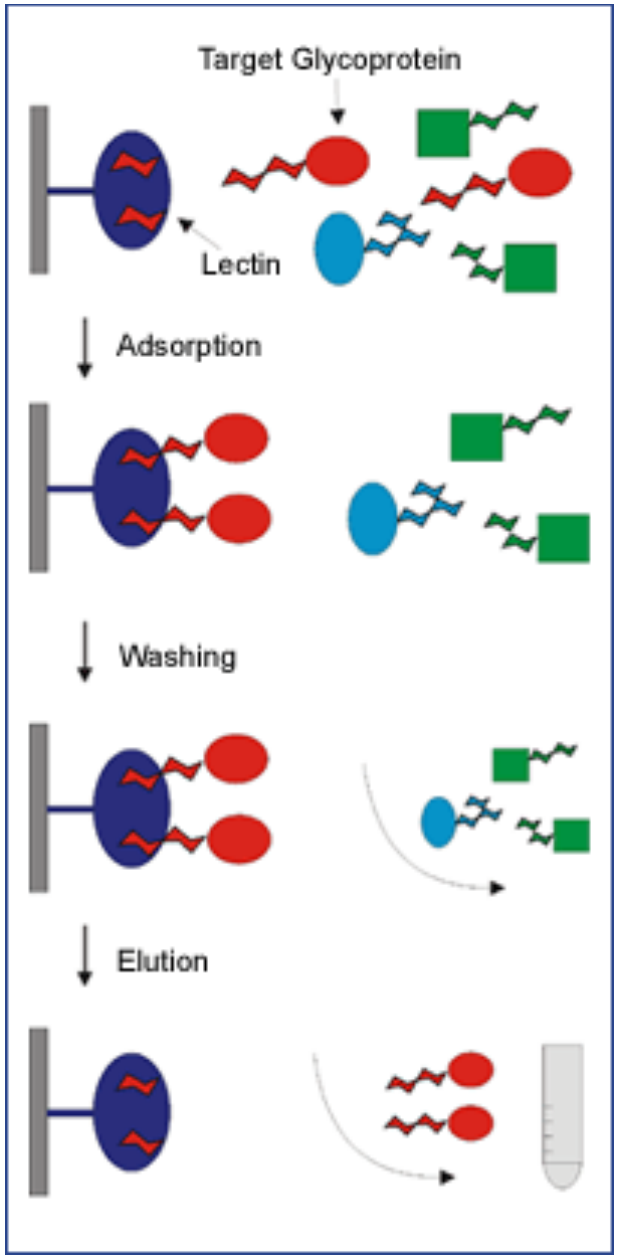
Protein-linked Glycans



http://www.nature.com/nrgastro/journal/v11/n10/images/nrgastro.2014.78-f1.jpg

Enrichment of Glycopeptides

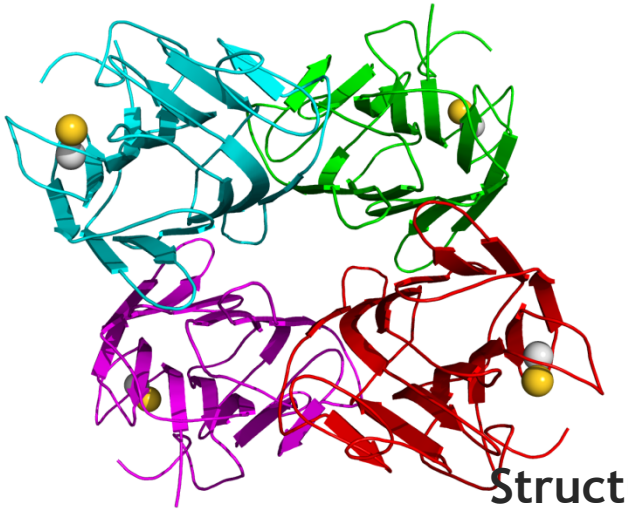
- Due to low abundance glycoproteins/glycopeptides usually need to be enriched
- **Lectin-based enrichment of glycoproteins**
  - Lectins are sugar-binding proteins
  - Lectins play a role in many cellular processes, for example in virus attachment to host cells, where the lectin's affinity to glycosylated membrane proteins is used by the virus



www.galab.de/technologies/img/trennung.gif

Enrichment of Glycoproteins

Lectins for the enrichment of N-glycans	Lectins for the enrichment of O-glycans
Concanavaline A (ConA)	*Helix pomatia* agglutinin
Lens culinaris agglutinin	AIL (*Artocarpus integrifolia* lectin)
Aleuria aurantia lectin	Peanut agglutinin

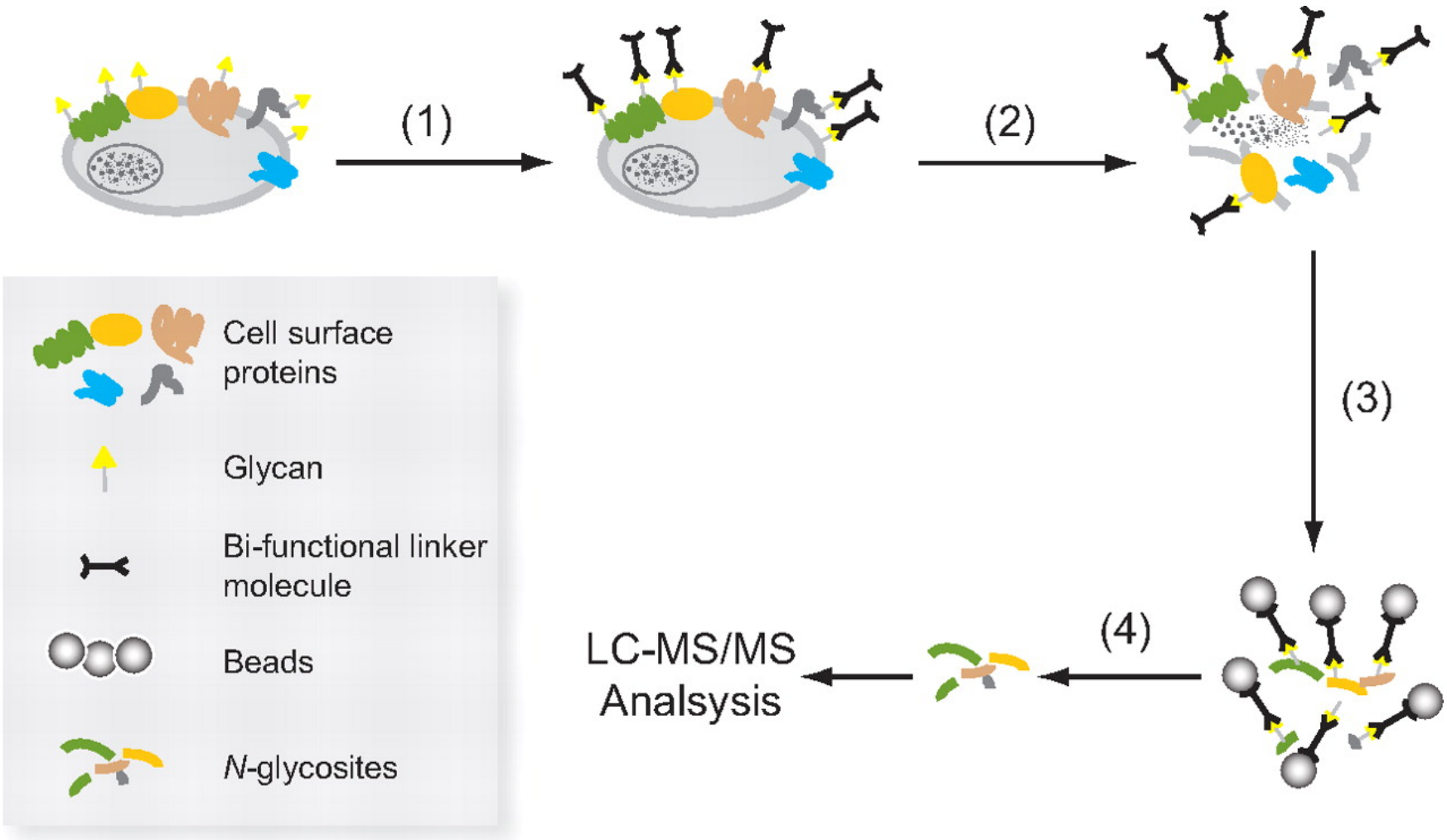


Structure of concanavaline A

http://en.wikipedia.org/wiki/Concanavalin_A. 02/01/012 2 PM CET

Enrichment of Glycoproteins

2. Linker-based enrichment of glycoproteins



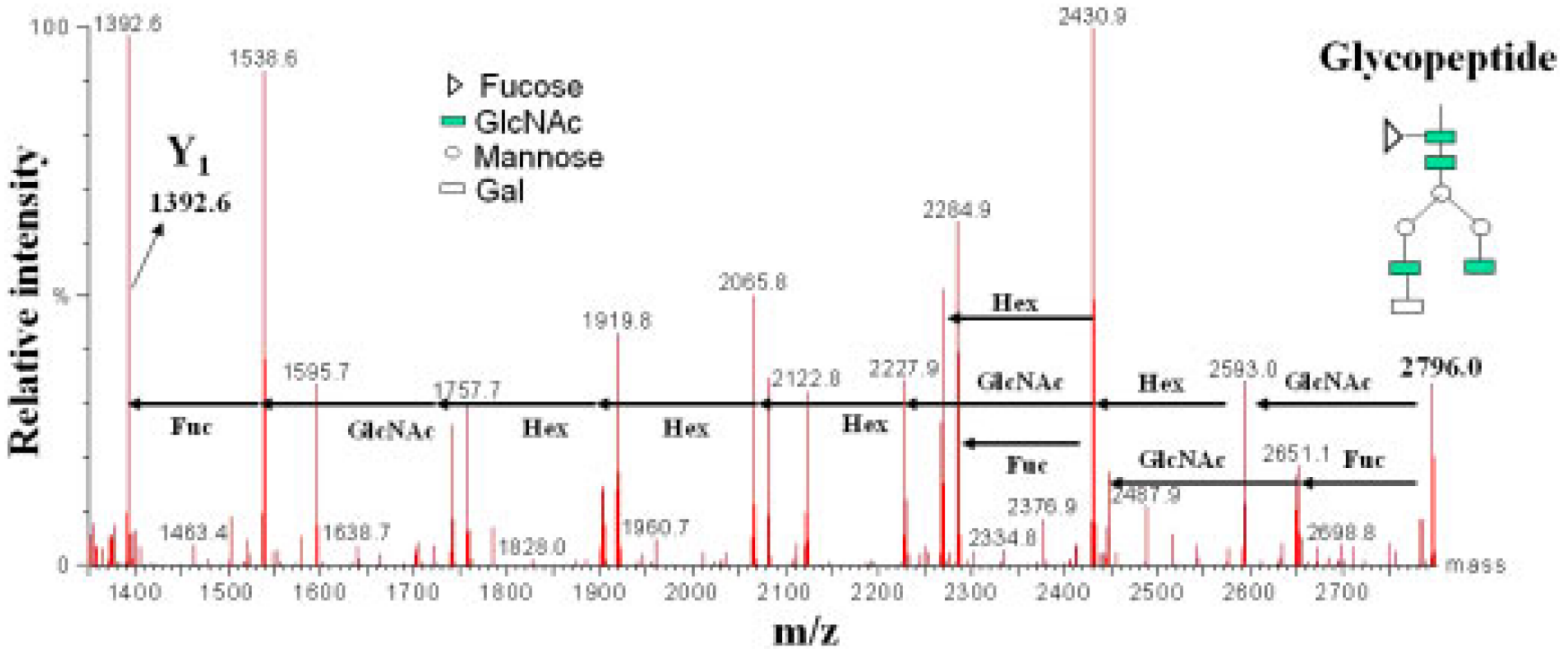
Schiess et al.,. Molecular and Cellular Proteomics 2009 Apr;8(4):624-38. doi: 10.1074/mcp.M800172-MCP200

MS 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 with unmodified peptides, since the chemical properties of the peptide and glycan are dissimilar
- Using CID, the collision energy is highly important to the content of the MS/MS spectrum

MS/MS of glycopeptides

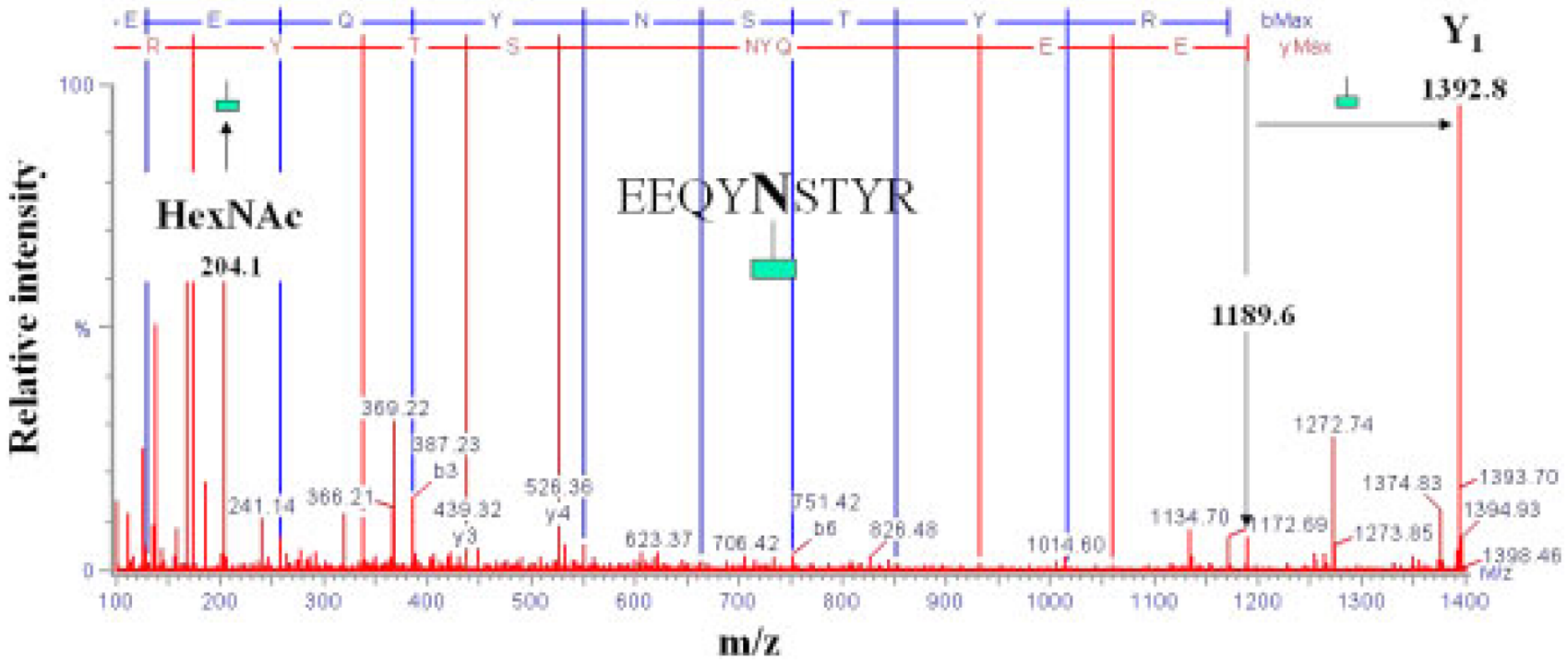
- Using low-energy CID, the tandem spectrum is dominated by ions from the sequential loss of sugars and occasionally the precursor ion, but there is no fragmentation of the peptide backbone



Lazar et al., Recent advances in the MS analysis of glycoproteins: Theoretical considerations. Electrophoresis 2011, 32, 2-13

MS/MS of glycopeptides

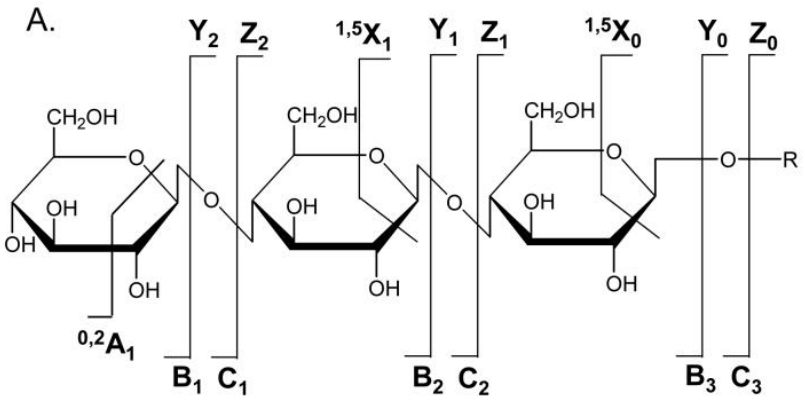
- Increasing the collision energy during CID fragmentation, signals for the sugar residues diminish, but the peptide backbone ions become visible



Lazar et al., Recent advances in the MS analysis of glycoproteins: Theoretical considerations. Electrophoresis 2011, 32, 2-13

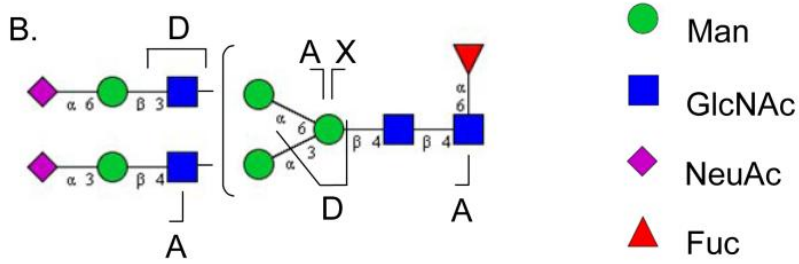
MS/MS of glycopeptides

- Information in the sugar stereochemistry (which Glc, Gal or Man ?), the linkage (1->4 or 1->6 ?) or branching pattern cannot be obtained using conventional CID fragmentation
- This can be achieved by “cross-ring” fragmentation on MALDI-TOF-TOF instruments. Very high collision energy is needed (orders of KeV)

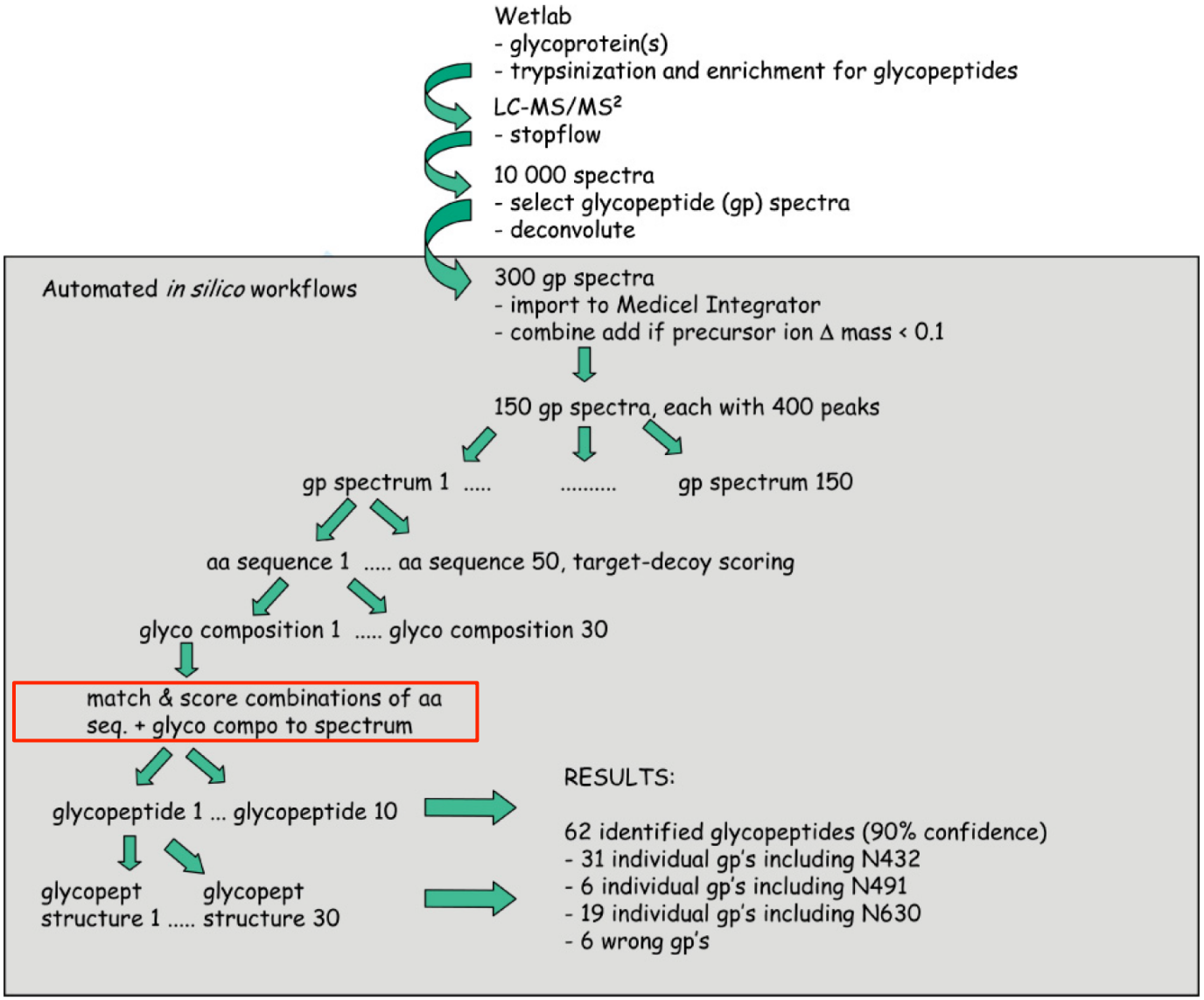


Nomenclature for tandem mass spectrometric product ions of glycans and glycoconjugates

D-ion (not part of the original nomenclature) consists of combined C- and Z-type fragmentation. They differentiate *N*-acetyllactosamine linkages and mannose branching structure



An N-Glycoproteomics Workflow



Joenväärä et al., N-Glycoproteomics – An automated workflow approach. Glycobiology. 2008 Vol. 18 no. 4 pp. 339-349.

Scoring function *Joenväärä et al.*

- Score *S* for glycopeptide *G,*

$$S(G) = -log(P(SPC(R,s) \geq SPC(G,s)))$$

  SPC is the shared peak count, *R* a random spectrum, *s the measured* spectrum.

- Furthermore, *s* has *M* mass values $\{m_1, ..., m_M\}$ and M_p is the number of possible mass values with the given mass range and tolerance *T*
- *G*, the theoretical glycopeptide spectrum has *N*, $\{g_1, ..., g_N\}$ mass values. R has also *N* mass values
- *SPC* counts the number of peak pairs with

$$|g_i - m_j| < T \; \forall \; i \in [1, N] \quad \wedge \quad j \in [1, M]$$

Scoring function *Joenväärä et al.*

- The probaility $P_R = P(SPC(R,s) \geq SPC(G,s))$ is calculated using the binomial distribution

$$P_R = 1 - \sum_{k=0}^{N_h - 1} \binom{N}{k} p^k (1-p)^{N-k}$$

where $N_h = SPC(G,s)$ and $p = \frac{M}{M_p}$

- The score, designed for glycopeptide scoring by *Joenväärä et al.* has similarity to the Ascore designed for phosphosite assignment by *Beausoleil et al.*

References

- **Phosphoproteomics**
  - Macek et al, Annu Rev Pharmacol Toxicol. 2009;49:199-221. PMID: 18834307
  - Beausoleil et al., A probability-based approach for high-throughput protein phosphorylation analysis and site localization. Nat Biotechnol. 2006 Oct;24(10):1285-92. PMID: 16964243
  - Olsen et al., Quantitative Phosphoproteomics Reveals Widespread Full Phosphorylation Site Occupancy During Mitosis. 2010. *Science Signaling* **3** (104), ra3. [DOI: 10.1126/scisignal. 2000475]
- **Acetylation proteomics**
  - Choudhary*, et al.* Lysine Acetylation Targets Protein Complexes and Co-Regulates. *Science* 325, 834 (2009); DOI: 10.1.126/science.1175371
- **Gycoproteomics**
  - Joenväärä et al., N-Glycoproteomics – An automated workflow approach. Glycobiology. 2008 Vol. 18 no. 4 pp. 339-349.
  - Zaia, Mass Spectrometry and the Emerging Field of Glycomics. Chem Biol. 2008 September 22; 15(9): 881–892.
  - Lazar et al., Recent advances in the MS analysis of glycoproteins: Theoretical considerations. Electrophoresis 2011, 32, 2-13
  - Schiess et al., Analysis of Cell Surface Proteome Changes viaLabel-free, Quantitative Mass Spectrometry. Molecular and Cellular Proteomics 2009,.

Materials

- Learning Units 10A-D