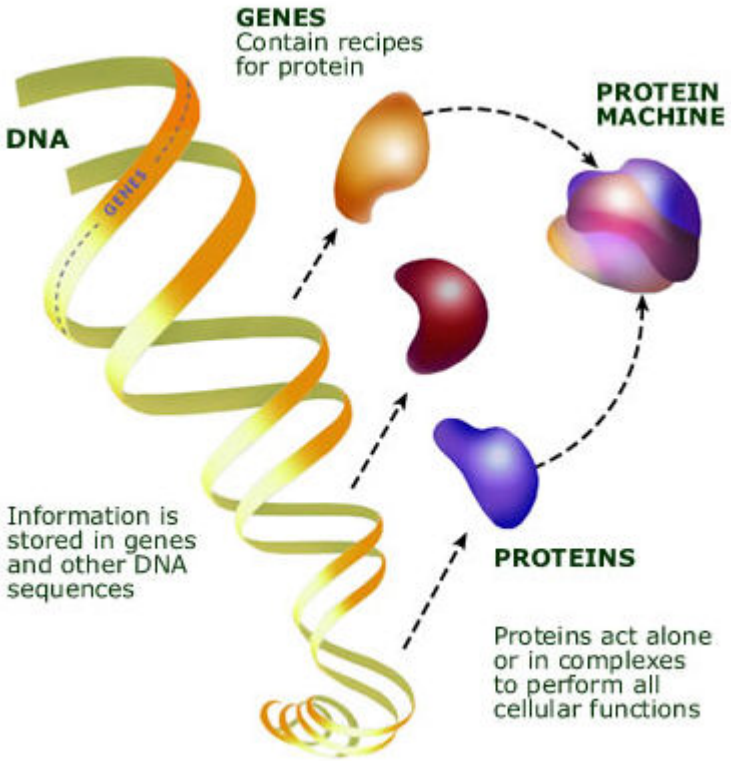
GMS BI 793 Lecture 8
Affinity Mass Spectrometry: Strategies for Proteome Profiling

Joseph Zaia

March 24, 2015

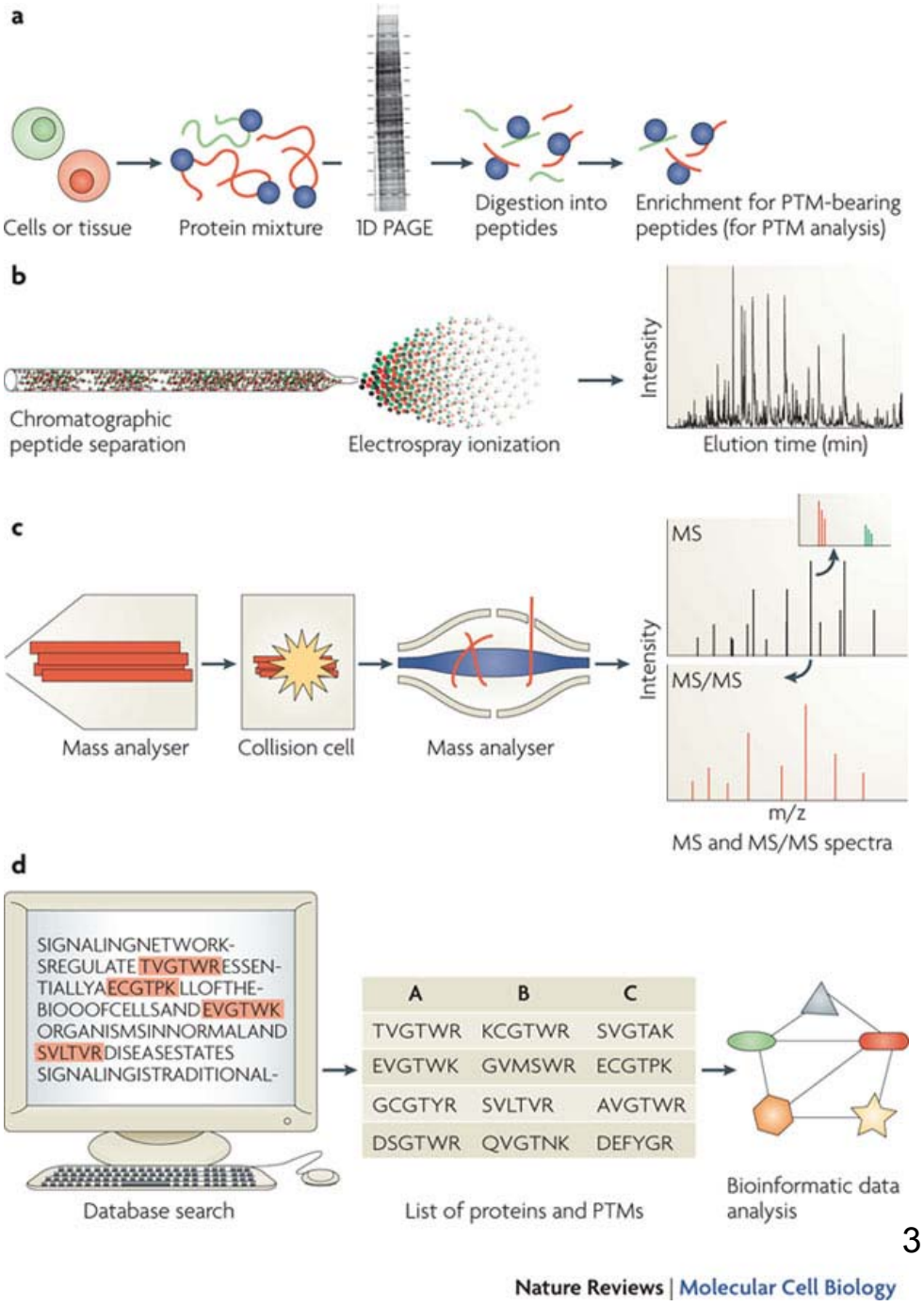
1. Discovery versus quantitative proteomics
2. Affinity proteomics
    - Tandem affinity purification
    - Isotope coding strategies
3. Affinity MS for identification of intracellular post-translational modifications
    - Phosphorylation
    - *O*-GlcNAc
    - *N*- and *O*-glycosylation

Proteomics: using genomic sequences to identify proteins



http://www.isaaa.org/siteimages/pocketkimages/clip_image002_0014.jpg

Proeomics: wide angle protein analysis



Choudhary, C. and M. Mann, Nat Rev Mol Cell Biol, 2010. **11**(6): p. 427-439.

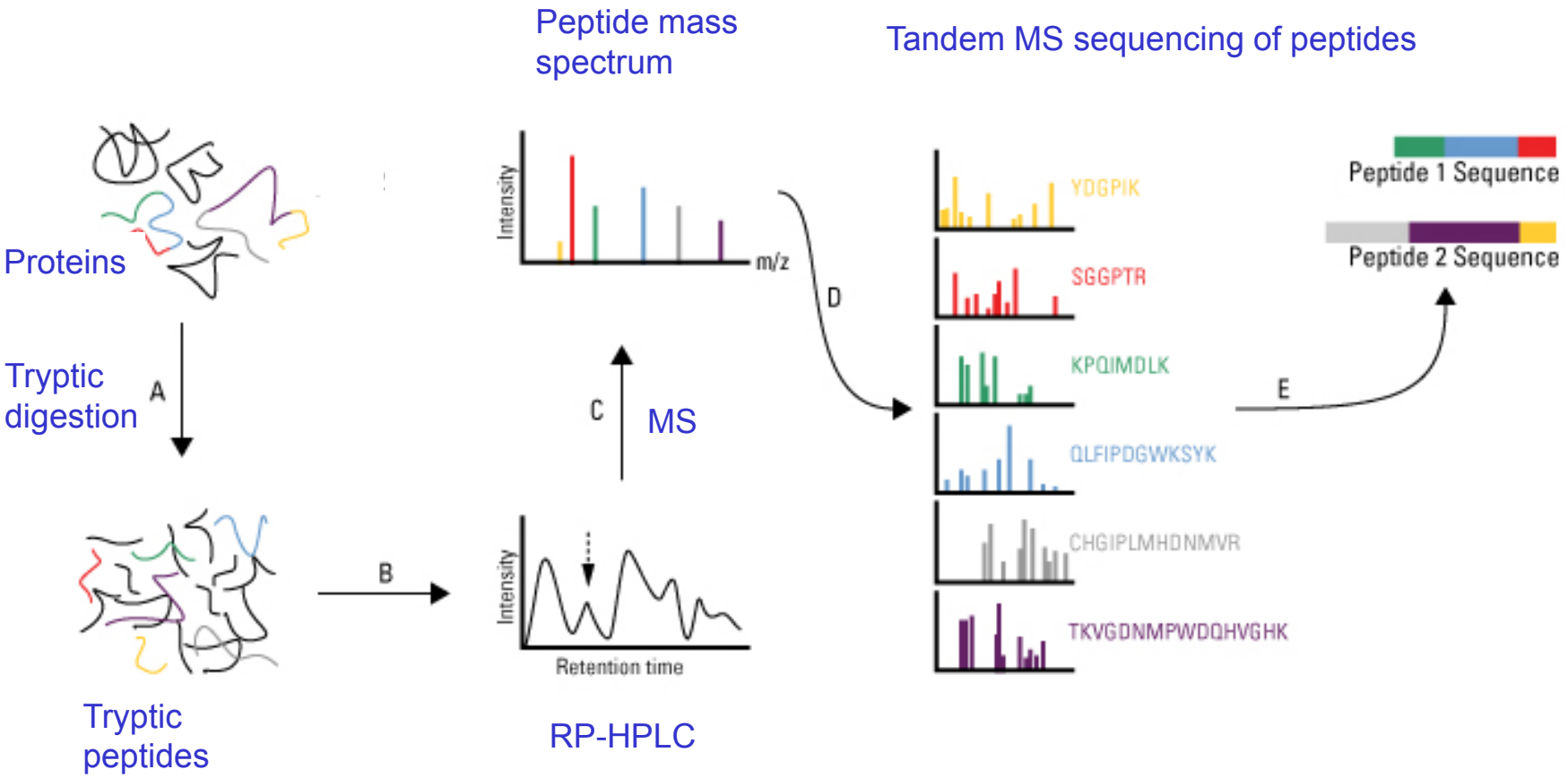
Discovery proteomics is less concerned with instrumental robustness

- When using stable isotope labels, can make pairwise comparisons
- Not appropriate for large clinical studies.
    - Instrumental variability
    - Effort, computation time

Targeted proteomics entails limiting scope of analysis so as to measure abundances of targeted peptides

- Reproducibility is essential
- Label free proteomics
- Accurate mass tags, often requires reducing mixture complexity
- Use of stable isotope labeled peptides (AQUA)
- Targeted SRM transitions: entails used of validated peptide data

Proteomics: discovery of proteins (qualitative)



http://www.piercenet.com/media/Proteomics%20Identification%20Workflow-700px.jpg

Tiers of Targeted MS measurements

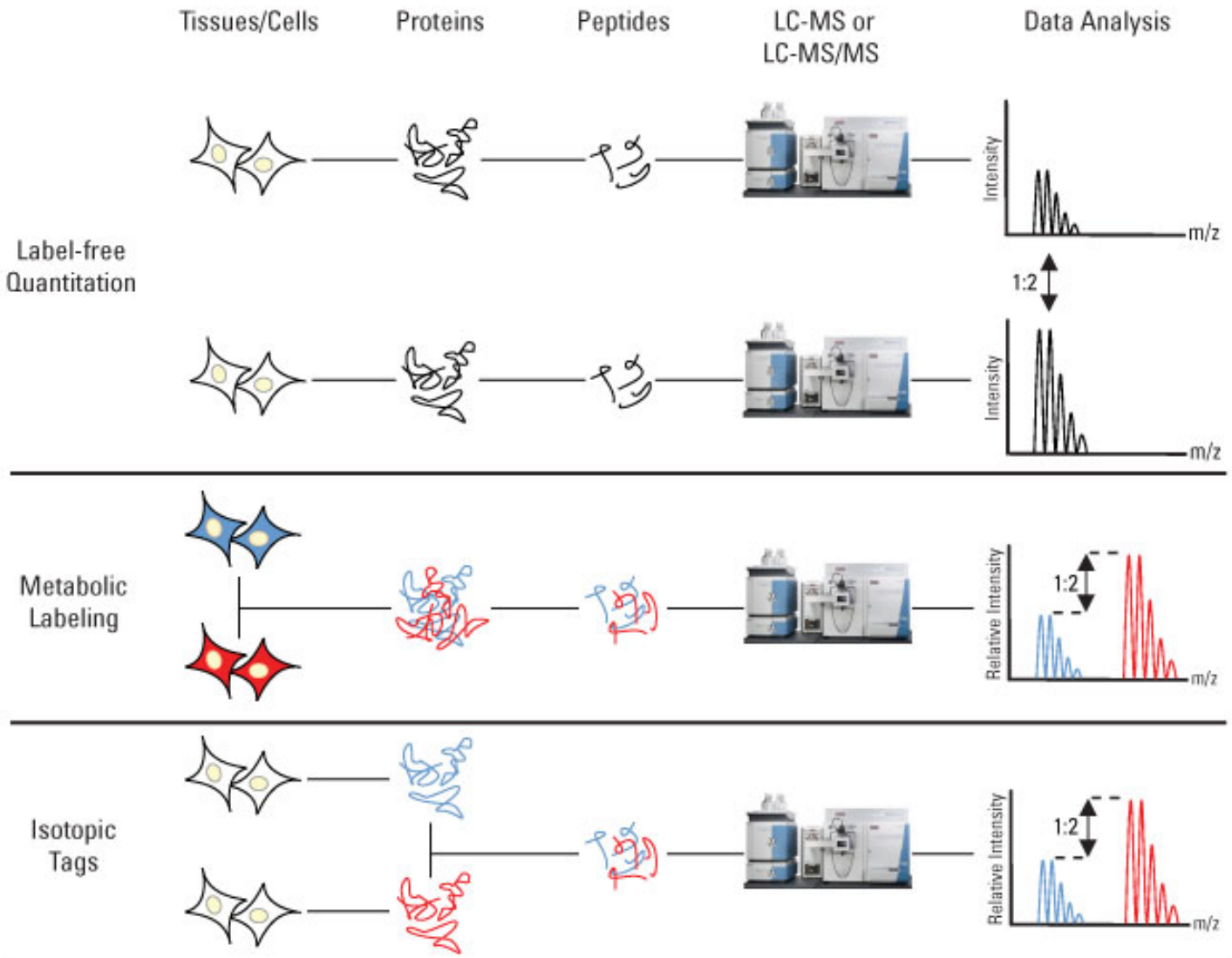
TABLE I

Three Tiers of Targeted MS Measurements; experimental design parameters and assay characteristics are listed for each tier

Tier and Areas of Application	Degree of Analytical Validation	Labeled Internal Standards	Reference Standards	Specificity	Precision	Quantitative Accuracy	Repeat-ability	Comments and Suggested References
Tier 1 **Clinical bioanalysis/ diagnostic laboratory test; single analyte or small numbers of analytes**	High, including batch-to-batch QC	Yes, for every analyte	Yes	High	High (typically <20-25% CV achieved)	Defining accuracy is a goal; true accuracy difficult to demonstrate.	High	Precise, quantitative assays; established, high performance; may need comply with FDA and CLIA guidance depending on use of assay Refs. 30, 41, 42, 53
Tier 2 **Research use assays for quantifying proteins, peptides, and post-translational modifications; 10's to 100's of analytes**	Moderate-to-high	Yes, for every analyte	Limited use	High	Moderate-to-high (typically <20-35% CV achieved)	Not applicable	High	Precise, relative quantitative assays; established performance; suitable for verification Refs. 30, 31, 36, 37, 40, 51, 70, 71
Tier 3 **Exploratory studies; 10's to 100's of analytes**	Low-to-moderate	None-to-limited	No	Moderate-to-high	Low-to-moderate: similar to label-free discovery	Not applicable	Moderate-to-high	Discovery in a targeted mode; performance not defined; results require further verification using quantitative techniques Refs. 36, 37, 86-89

Carr, S.A., et al., Molecular & cellular proteomics : MCP, 2014. **13**(3): p. 907-17.

Proteomics: targeted quantification of proteins



http://www.piercenet.com/media/Quantitative%20Protomics%20Workflows-700px.jpg

Selected reaction monitoring

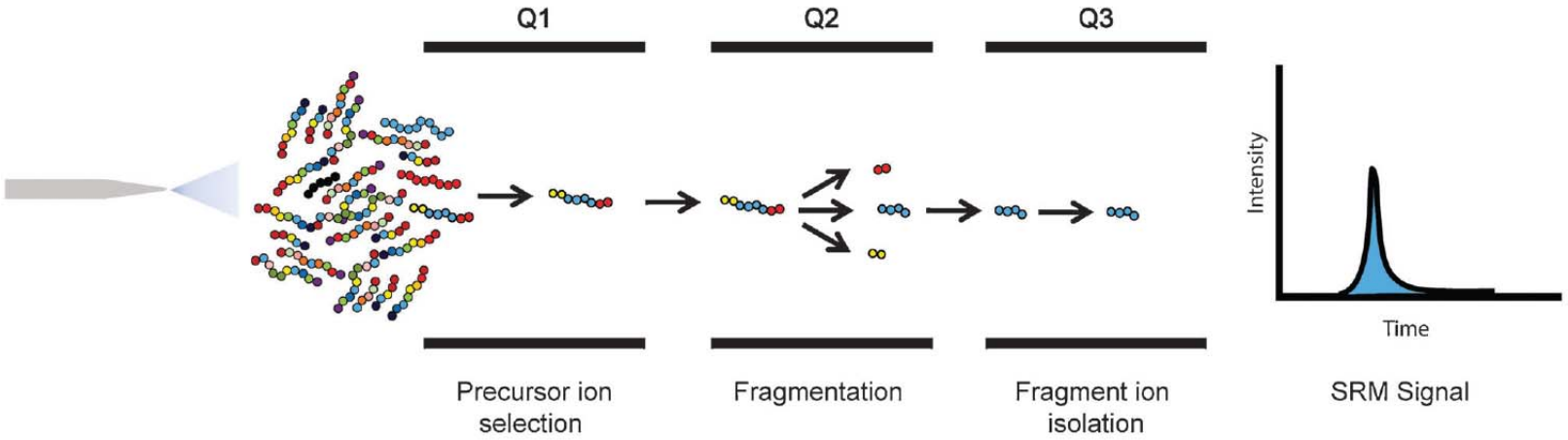
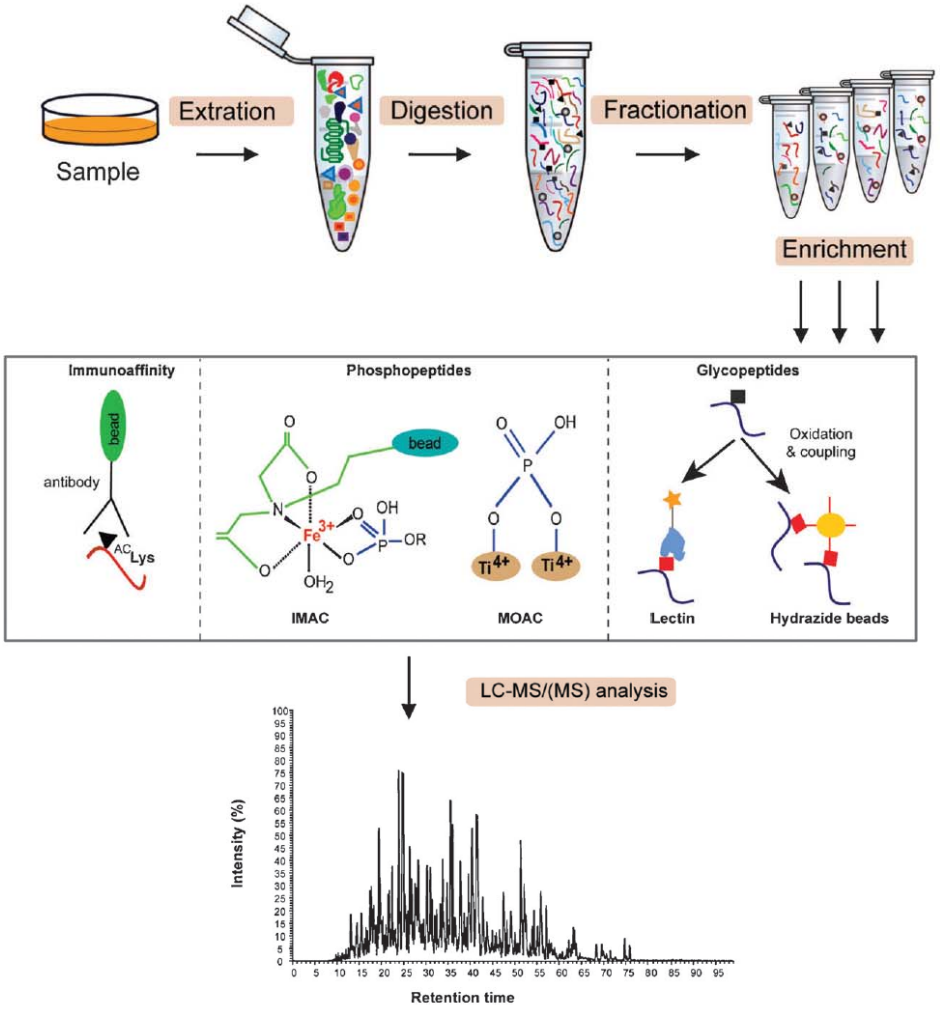


Fig. 3 Overview of selected reaction monitoring mass spectrometry (LC-SRM-MS). Proteotypic peptides are separated from complex biological samples by reversed phase liquid chromatography. The selected proteotypic peptides are isolated in Q1 reducing interfering background signal, and subsequently fragmented in Q2 and specific transition ions are isolated in Q3 prior to detection. Multiple rounds of isolation greatly reduce the background signal resulting in greatly improved signal to noise in typical SRM-MS quantification.

PTM proteomics



Angel, T.E., et al., Chemical Society reviews, 2012. **41**(10): p. 3912-28.

Fig. 4 Experimental strategies to study protein post-translational modification. Enrichment of lysine acetylated (▲), phosphorylated (●) and glycosylated peptides (■) using different affinity media is shown as examples. A typical workflow for PTM analysis involves extraction of proteins from cells or tissues followed by proteolysis of extracted proteins into peptides, reduction of sample complexity by fractionation (if required), enrichment of PTM peptides by using appropriate methods, LC-MS/(MS) and database searches for identification, quantification and PTM site matching. Different methods can be selected for protein extraction, fractionation, and enrichment of PTM peptides depending on sample type, complexity and targeted analysis.

Some common post-translational modifications of proteins

Table 1. Some common and important post-translational modifications

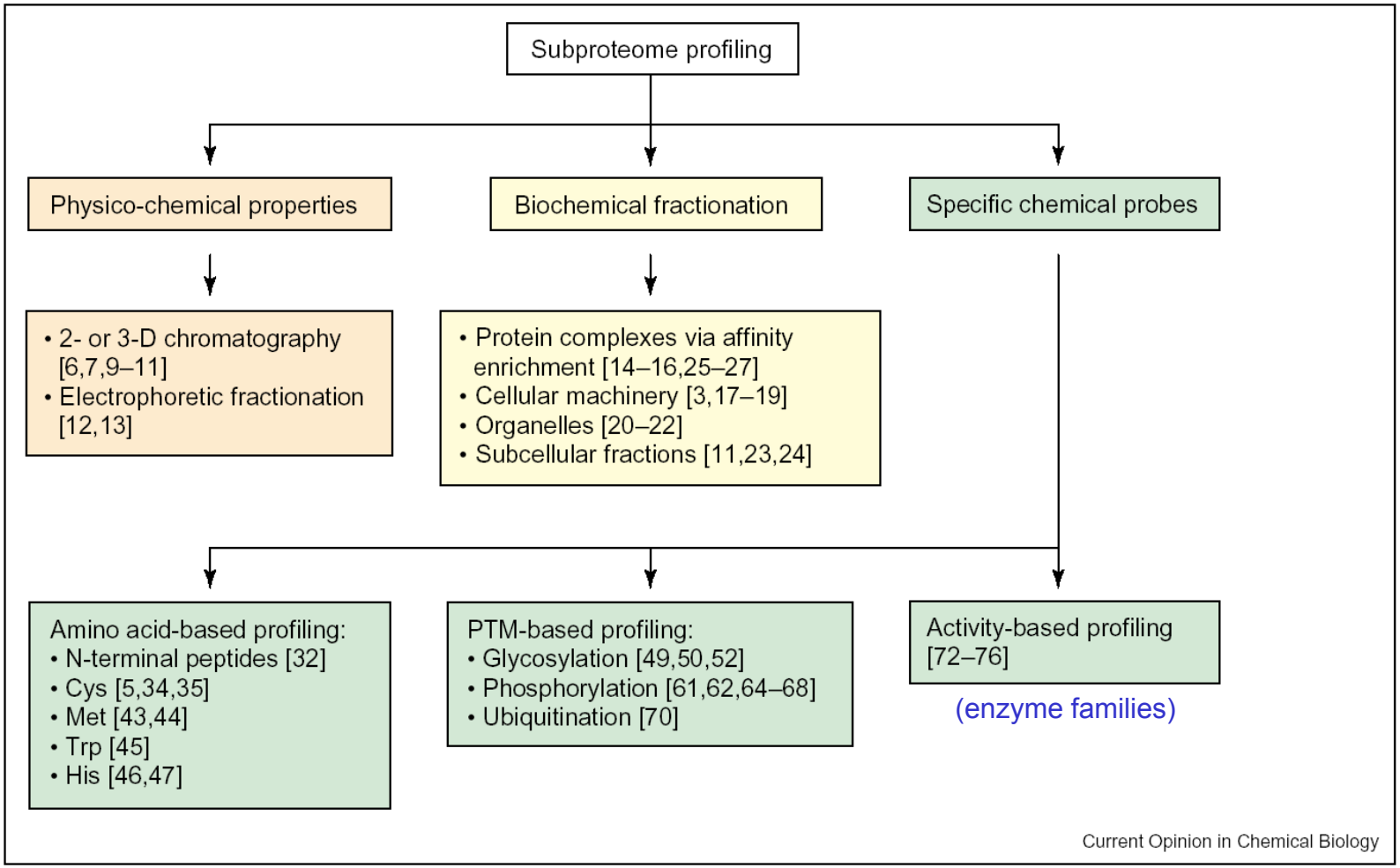
PTM type	ΔMass[a] (Da)	Stability[b]	Function and notes
Phosphorylation			Reversible, activation/inactivation of enzyme activity, modulation of molecular interactions, signaling
pTyr	+80	+++	
pSer, pThr	+80	+/++	
Acetylation	+42	+++	Protein stability, protection of N terminus. Regulation of protein–DNA interactions (histones)
Methylation	+14	+++	Regulation of gene expression
Acylation, fatty acid modification			Cellular localization and targeting signals, membrane tethering, mediator of protein–protein interactions
Farnesyl	+204	+++	
Myristoyl	+210	+++	
Palmitoyl	+238	+/++	
etc.			
Glycosylation			
N-linked	>800	+/++	Excreted proteins, cell–cell recognition/signaling
O-linked	203, >800	+/++	*O*-GlcNAc, reversible, regulatory functions
GPI anchor	>1,000	++	Glycosylphosphatidylinositol (GPI) anchor. Membrane tethering of enzymes and receptors, mainly to outer leaflet of plasma membrane
Hydroxyproline	+16	+++	Protein stability and protein–ligand interactions
Sulfation (sTyr)	+80	+	Modulator of protein–protein and receptor–ligand interactions
Disulfide bond formation	–2	++	Intra- and intermolecular crosslink, protein stability
Deamidation	+1	+++	Possible regulator of protein–ligand and protein–protein interactions, also a common chemical artifact
Pyroglutamic acid	–17	+++	Protein stability, blocked N terminus
Ubiquitination	>1,000	+/++	Destruction signal. After tryptic digestion, ubiquitination site is modified with the Gly-Gly dipeptide
Nitration of tyrosine	+45	+/++	Oxidative damage during inflammation

[a]A more comprehensive list of PTM Δmass values can be found at: http://www.abrf.org/index.cfm/dm.home
[b]Stability: + labile in tandem mass spectrometry, ++ moderately stable; +++ stable.

Mann, M., and Jensen, O. N. (2003). *Nat Biotechnol 21*, 255-61.

Separation strategies for proteome profiling

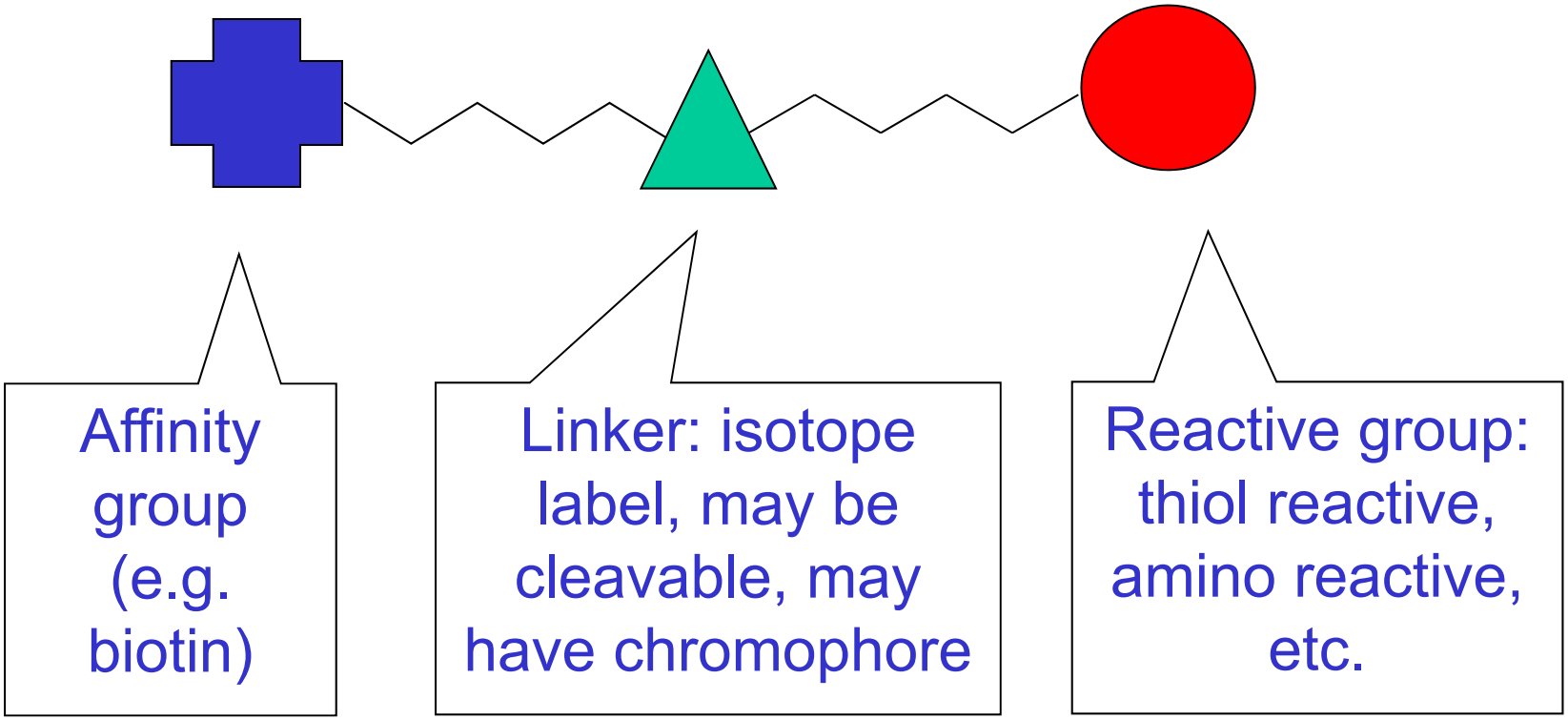
Figure 1



Strategies for subproteome profiling, with key references.

Zhang, H.; Yan, W.; Aebersold, R. *Curr Opin Chem Biol* **2004**, *8*, 66-75.

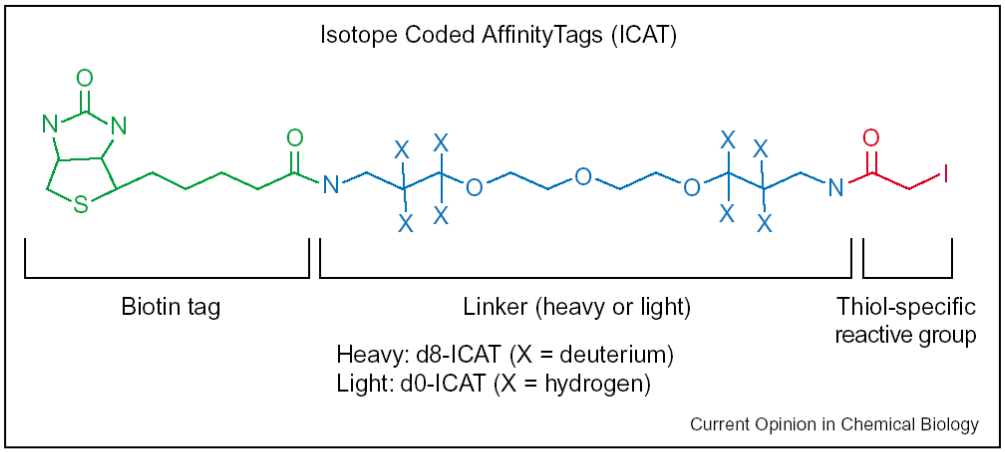
The Isotope Coding Reagents



- Non-isobaric labeling (ICAT, mTRAQ, SILAC)
- Isobaric labeling (iTRAQ, TMT)

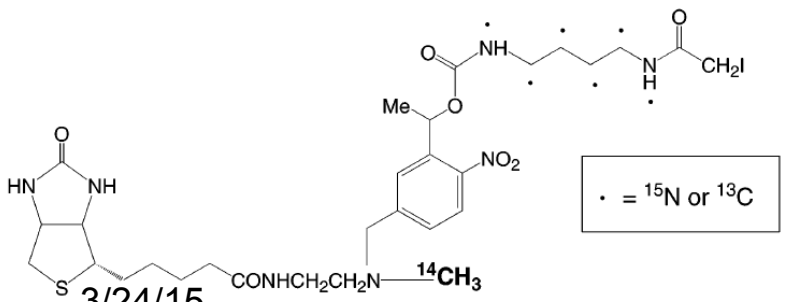
ICAT Reagents

Traditional

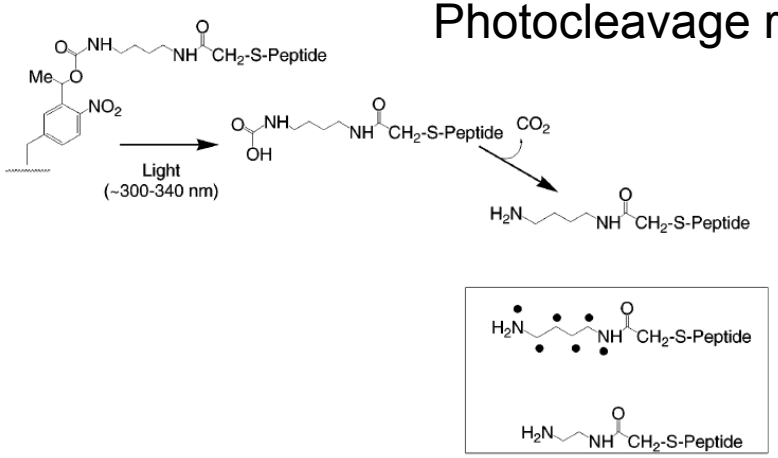


The structure of ICAT reagents, which comprise a cysteine-reactive group (red), a linker containing either heavy or light isotopes (blue) and a binding affinity tag (green).

VICAT (visible ICAT)



Photocleavage rxn



Bottari, P.; Aebersold, R.; Turecek, F.; Gelb, M. H. *Bioconjug Chem* **2004**, *15*, 380-388.

ICAT MS: biotin connected to a thiol-reactive group through an isotope-labeled linker

- **Traditional ICAT**
  - Deuterated
  - Biotinylated
  - Non-cleavable
  - No chromophore
  - $$$$$$
- **Improved ICAT**
  - ^{13}C and ^{15}N
  - Reducible or photocleavable linkers
  - Cleavable linker
  - Addition of chromophore
  - $$$$$$

Bottari, P.; Aebersold, R.; Turecek, F.; Gelb, M. H. *Bioconjug Chem* **2004**, *15*, 380-388.

Isobaric tag for relative and absolute quantitation (iTRAQ, Sciex)

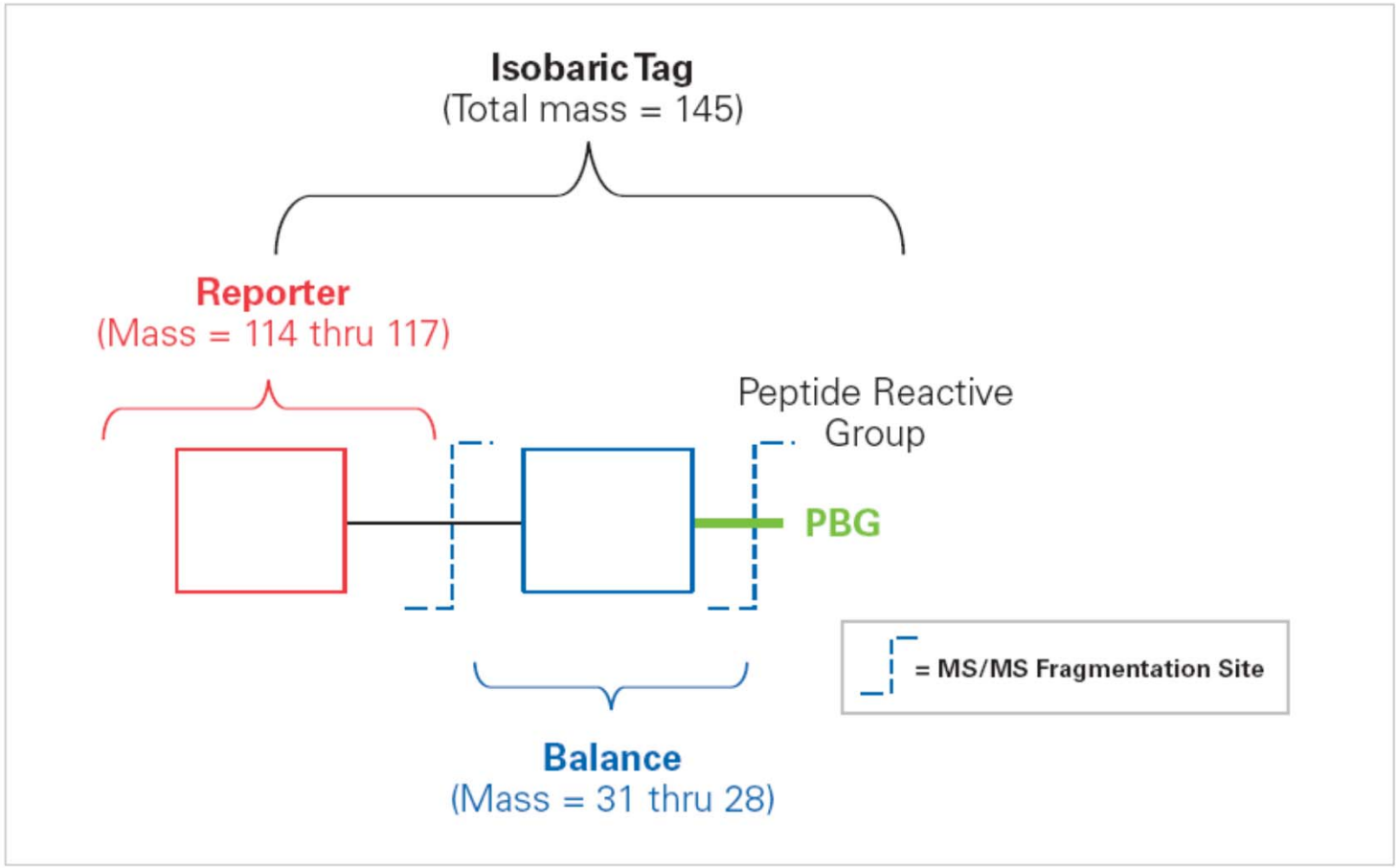


Figure 1. iTRAQ™ reagent structure

iTRAQ

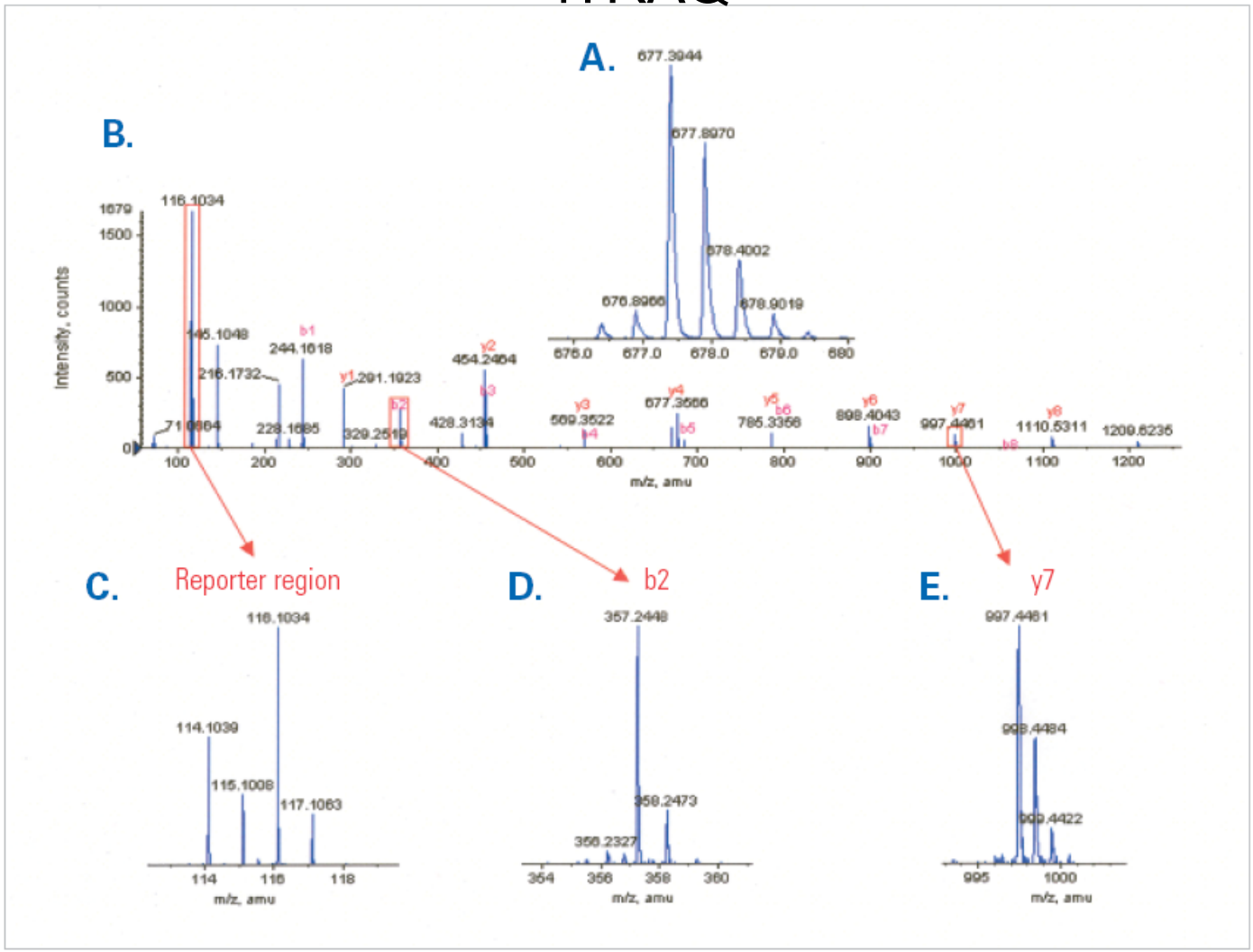
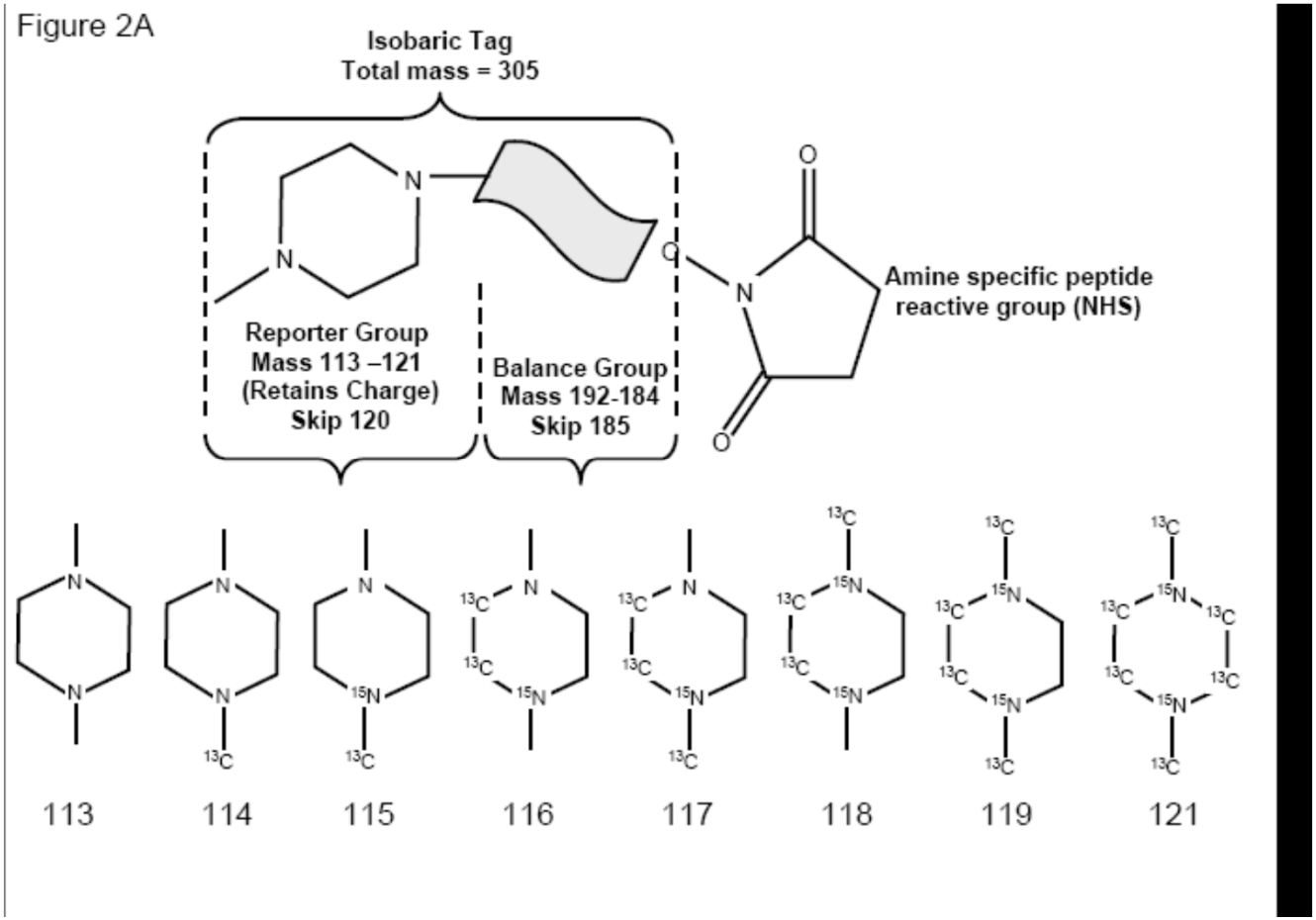


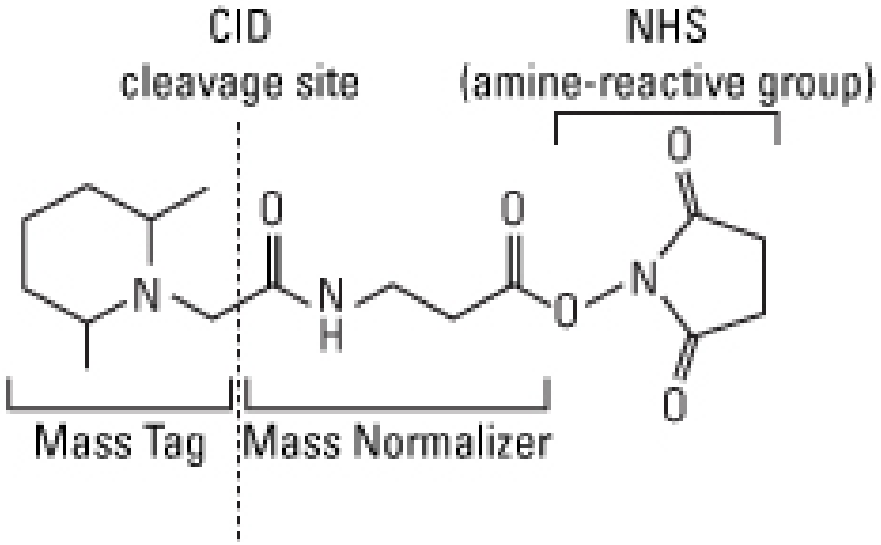
Figure 2. MS and MS/MS spectra from a multiplex sample labeled with 4 iTRAQ™ reagents showing **A.** doubly charged parent ion **B.** MS/MS spectrum corresponding to VLVDTDYK **C.** 4 diagnostic reporter ions and **D.** and **E.** peptide fragment ions

Eight channel iTRAQ

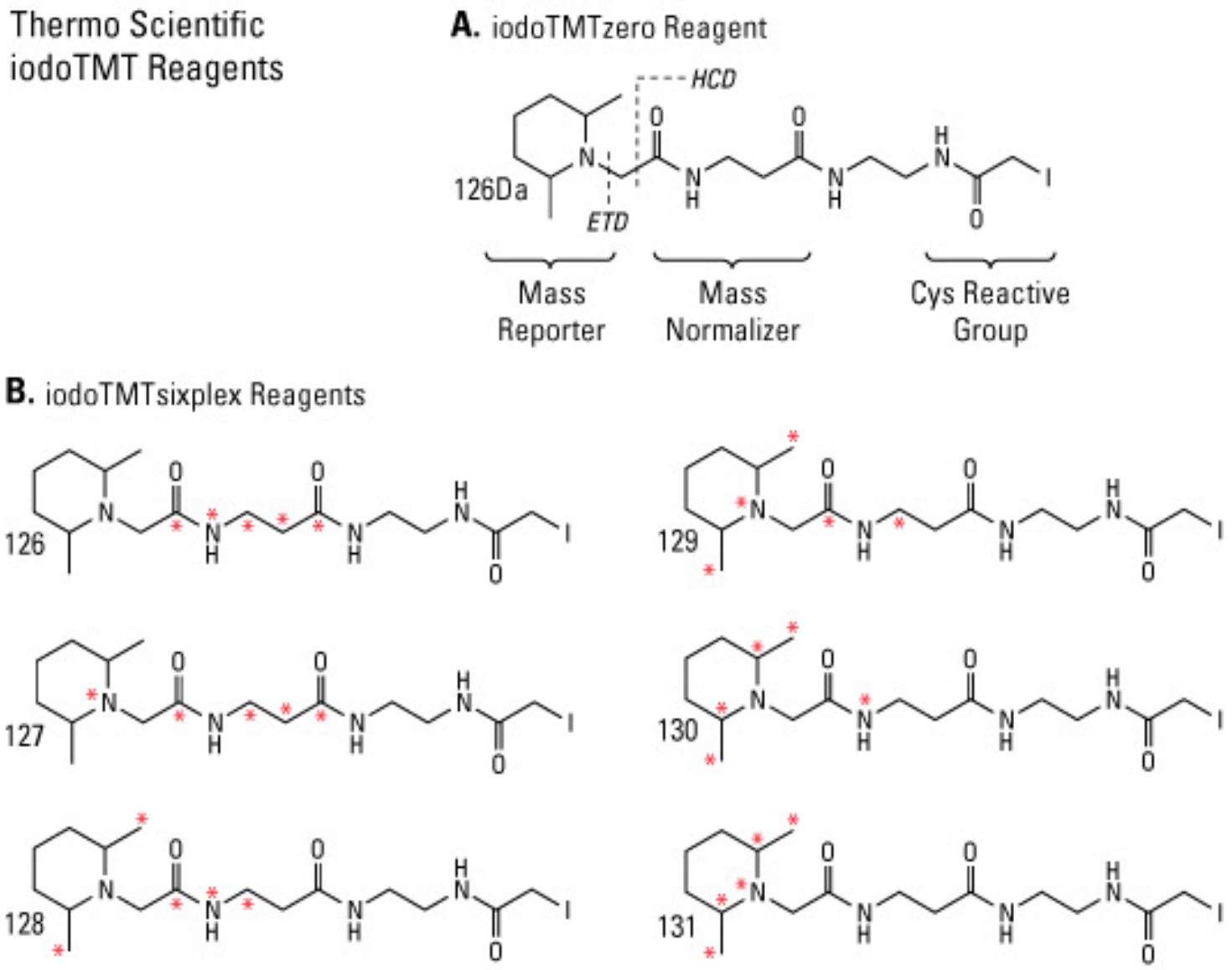


Pierce, A., Unwin, R.D., Evans, C.A., Griffiths, S., Carney, L., Zhang, L., Jaworska, E., Lee, C.F., Blinco, D., Okoniewski, M.J., Miller, C.J., Bitton, D.A., Spooncer, E., and Whetton, A.D. (2007). Eight-channel iTRAQ enables comparison of the activity of 6 leukaemogenic tyrosine kinases. Mol Cell Proteomics.

Tandem mass tags (TMT, Thermo-Fisher)

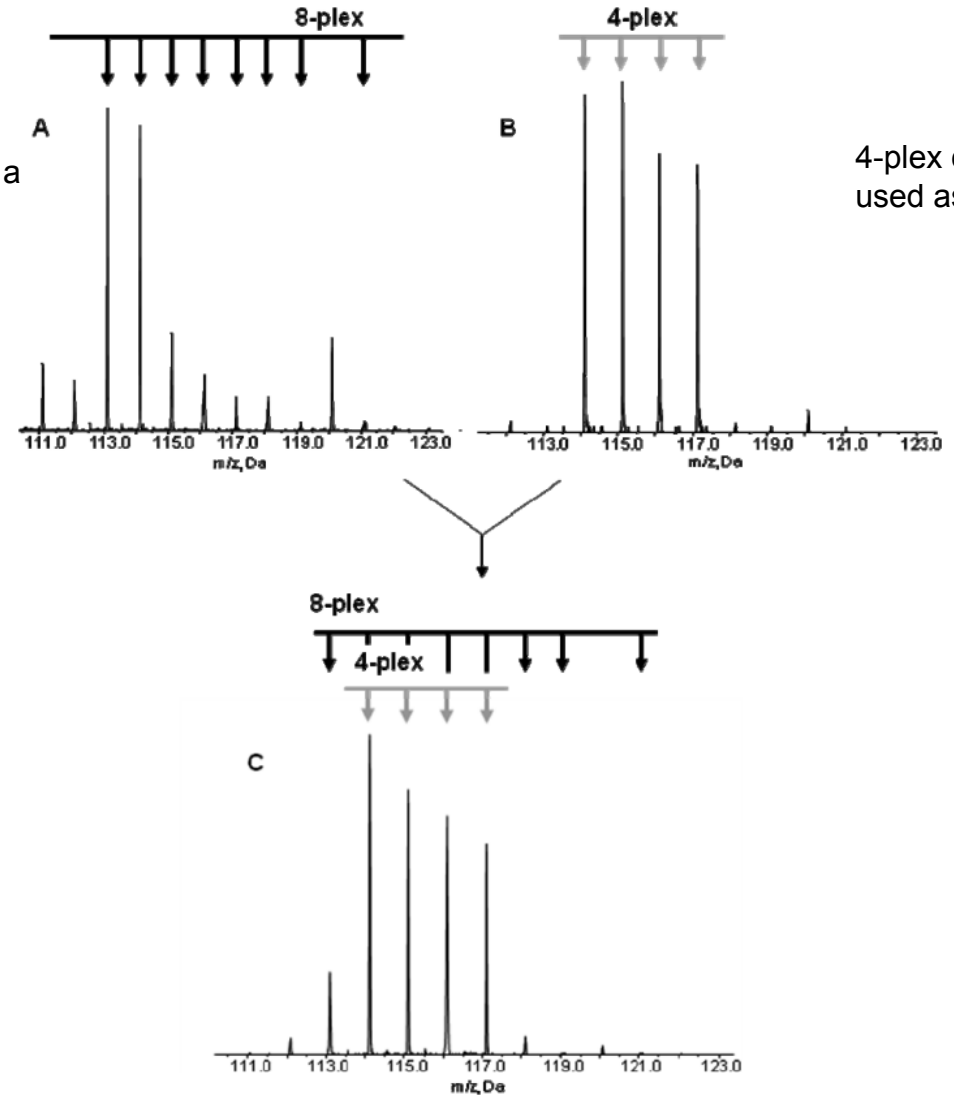


Iodo TMT reagents



iTRAQ and complex mixtures

8-plex samples in a dilution series (4 protein mix)

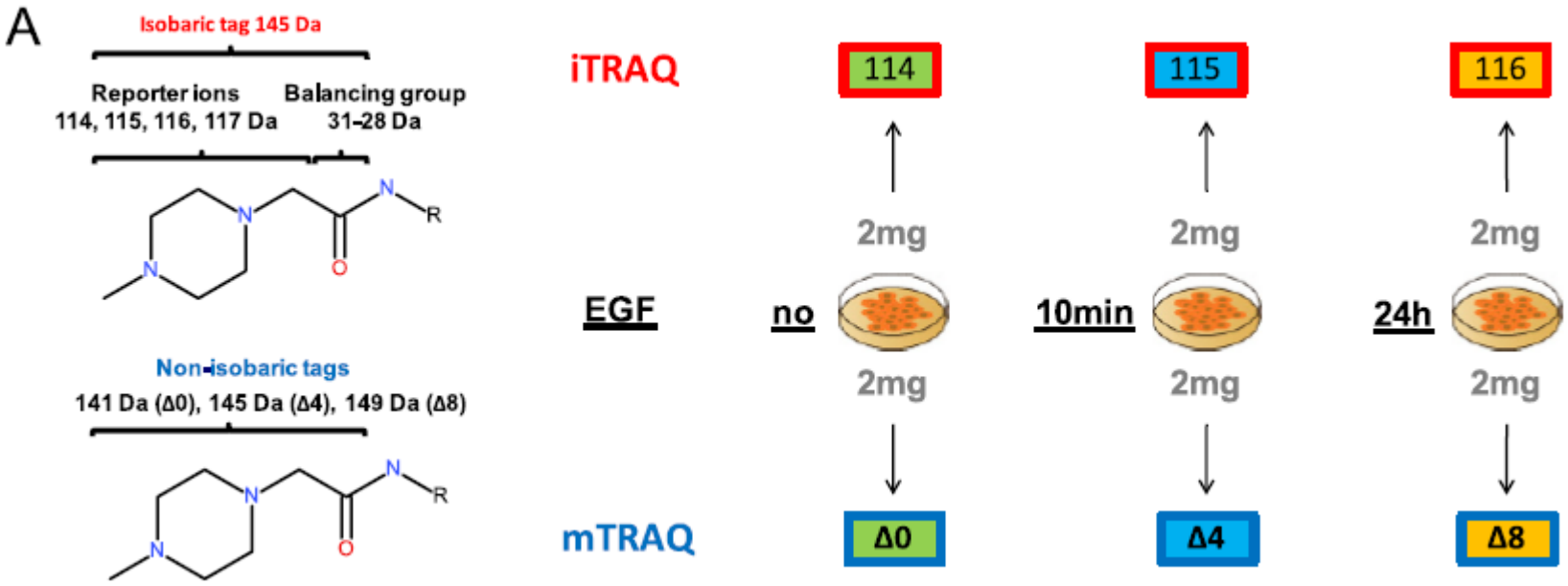


4-plex cyanobacterium proteome used as a complex background

Figure 6. Interference from MS/MS mixing between 8-plex master mix and 4-plex sample background. The spectrum describes a single scan of interference *via* the isolation of precursor ion 764.83 *m*/*z* (monoisotopic). This datum demonstrates a severe case of actual background mixing, the resultant quality for sequencing of fragment *y*- and *b*-ions are also adversely affected (copopulation of 4-plex and 8-plex peptide fragments). (A) The 8-plex master mix iTRAQ reporters; (B) 4-plex sample background iTRAQ reporters; (C) mixed MS/MS between the two sets of labels. Arrows denote the *m*/*z* position of the labels.

Ow, S.Y., et al., Journal of proteome research, 2009. **8**(11): p. 5347-55.

iTRAQ vs mTRAQ

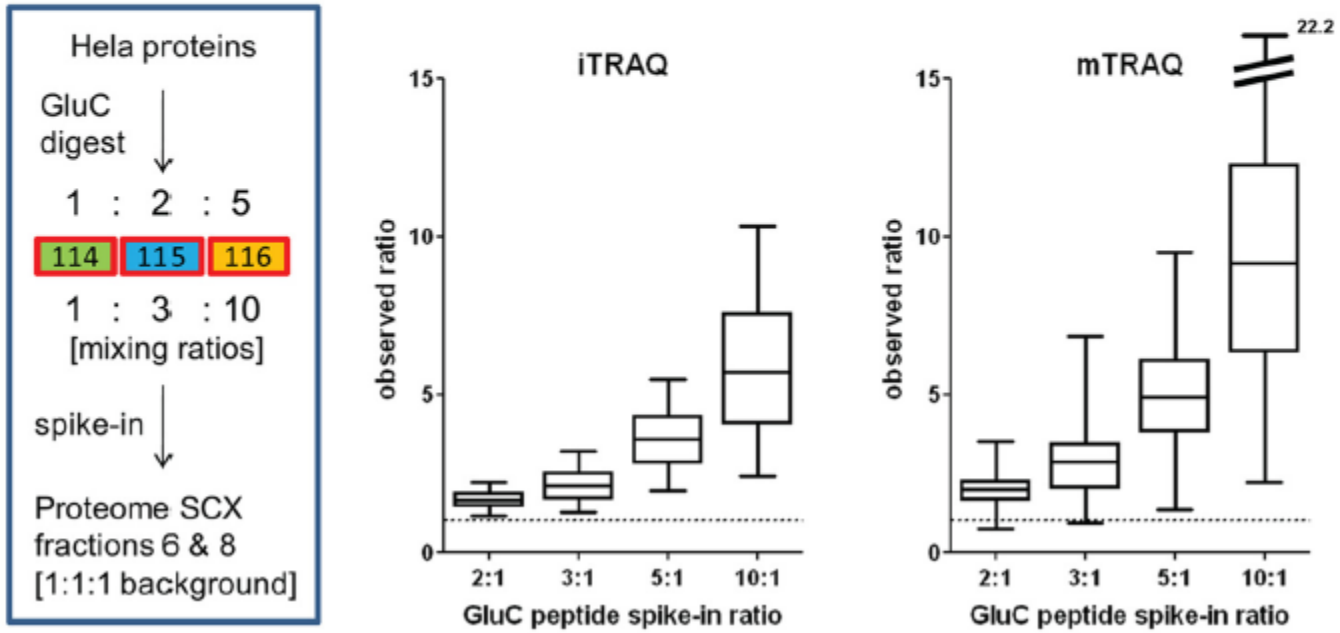


Mertins, P., et al., Molecular & cellular proteomics : MCP, 2012. 11(6): p. M111 014423.

iTRAQ vs mTRAQ: FIG. 6. iTRAQ quantification is more precise but less accurate than mTRAQ. A, GluC peptide spike-in experiments to test accuracy and variability of quantification.

A

	iTRAQ				mTRAQ			
spike ratio	**2:1**	**3:1**	**5:1**	**10:1**	**2:1**	**3:1**	**5:1**	**10:1**
GluC peptides	2,444	2,444	2,447	2,452	605	606	685	673
median ratios	**1.7**	**2.1**	**3.6**	**5.7**	**2.0**	**2.9**	**4.9**	**9.2**
CV [%]	22.4	33.4	30.8	42.2	67.7	73.6	116.7	188.0



Mertins, P., et al., Molecular & cellular proteomics : MCP, 2012. 11(6): p. M111 014423.

Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry

Nature **2002**, *415*, 180-183.

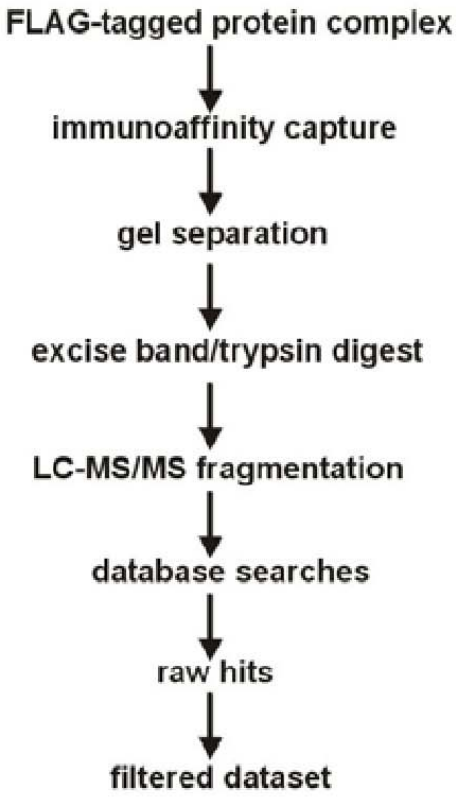
- Ho, Yuen*; Gruhler, Albrecht*; Heilbut, Adrian*; Bader, Gary D.†‡; Moore, Lynda*; Adams, Sally-Lin*; Millar, Anna*; Taylor, Paul*; Bennett, Keiryn*; Boutilier, Kelly*; Yang, Lingyun*; Wolting, Cheryl*; Donaldson, Ian*; Schandorff, Søren*; Shewnarane, Juanita*; Vo, Mai*†; Taggart, Joanne*†; Goudreault, Marilyn*†; Muskat, Brenda*; Alfarano, Cris*; Dewar, Danielle†; Lin, Zhen†; Michalickova, Katerina†‡; Willems, Andrew R.†§; Sassi, Holly†; Nielsen, Peter A.*; Rasmussen, Karina J.*; Andersen, Jens R.*; Johansen, Lene E.*; Hansen, Lykke H.*; Jespersen, Hans*; Podtelejnikov, Alexandre*; Nielsen, Eva*; Crawford, Janne*; Poulsen, Vibeke*; Sørensen, Birgitte D.*; Matthiesen, Jesper*; Hendrickson, Ronald C.*; Gleeson, Frank*; Pawson, Tony†§; Moran, Michael F.*; Durocher, Daniel†§; Mann, Matthias*; Hogue, Christopher W. V.*†‡; Figeys, Daniel*; Tyers, Mike†§

- 725 yeast proteins selected and expressed as Flag epitope fusions
  - 100 protein kinases, 36 phosphatases, 86 DNA repair proteins
- One step immunoaffinity purification, eluted with excess FLAG peptide
- Complexes separated by SDS-PAGE
- 15,683 gel slices processed by in-gel tryptic digestion, followed by tandem MS
- 940,000 tandem mass spectra that matched proteins in the yeast databases
- 35,000 protein identifications made

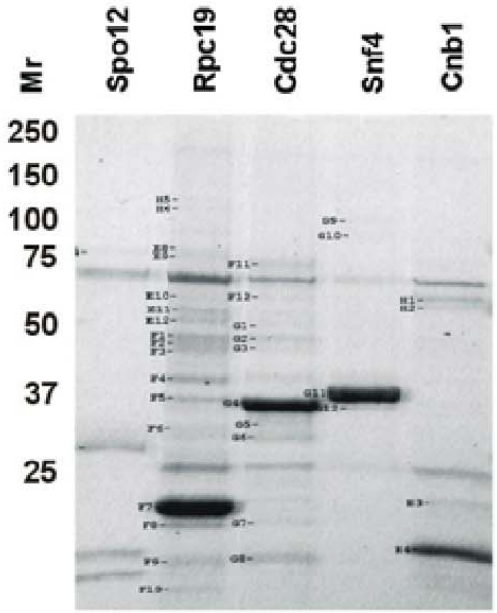
- Cited 2327 times as of 4/1/14

Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry

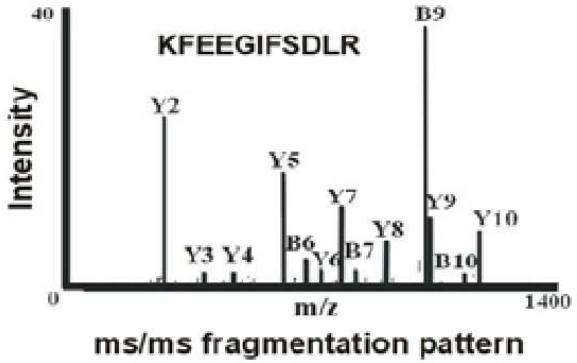
a



b



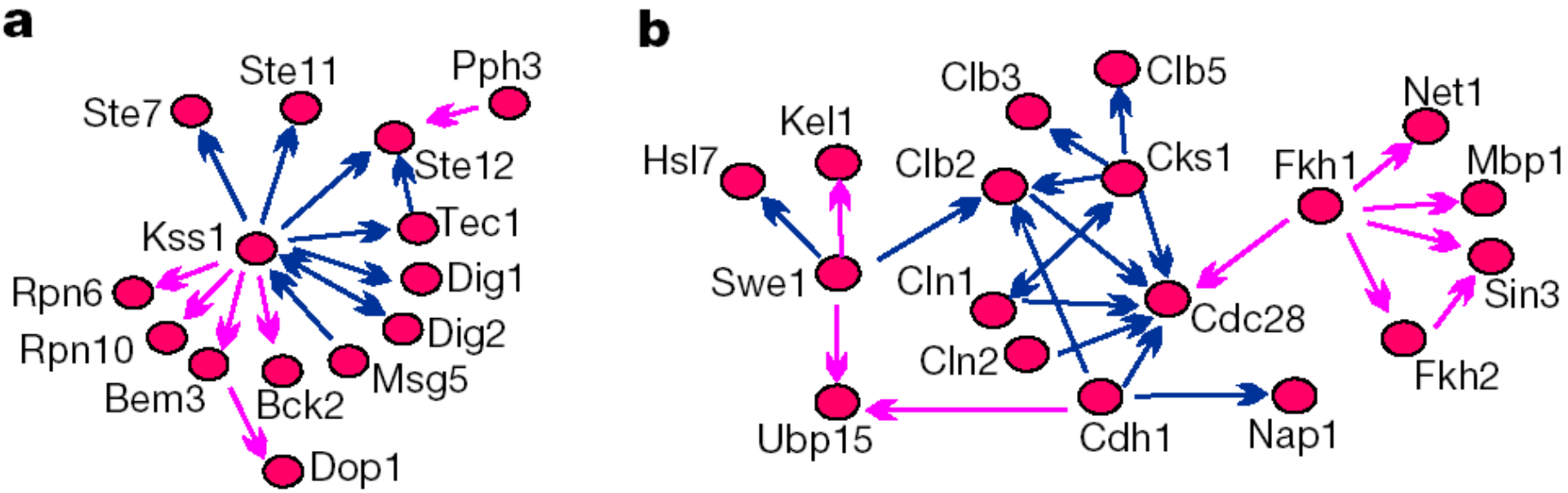
c



3/24/15

Nature **2002**, *415*, 180-183.

Example signaling diagrams determined from the affinity MS/MS data.



Nature **2002**, *415*, 180-183.

TAP: tandem affinity purification, applied to the yeast proteome

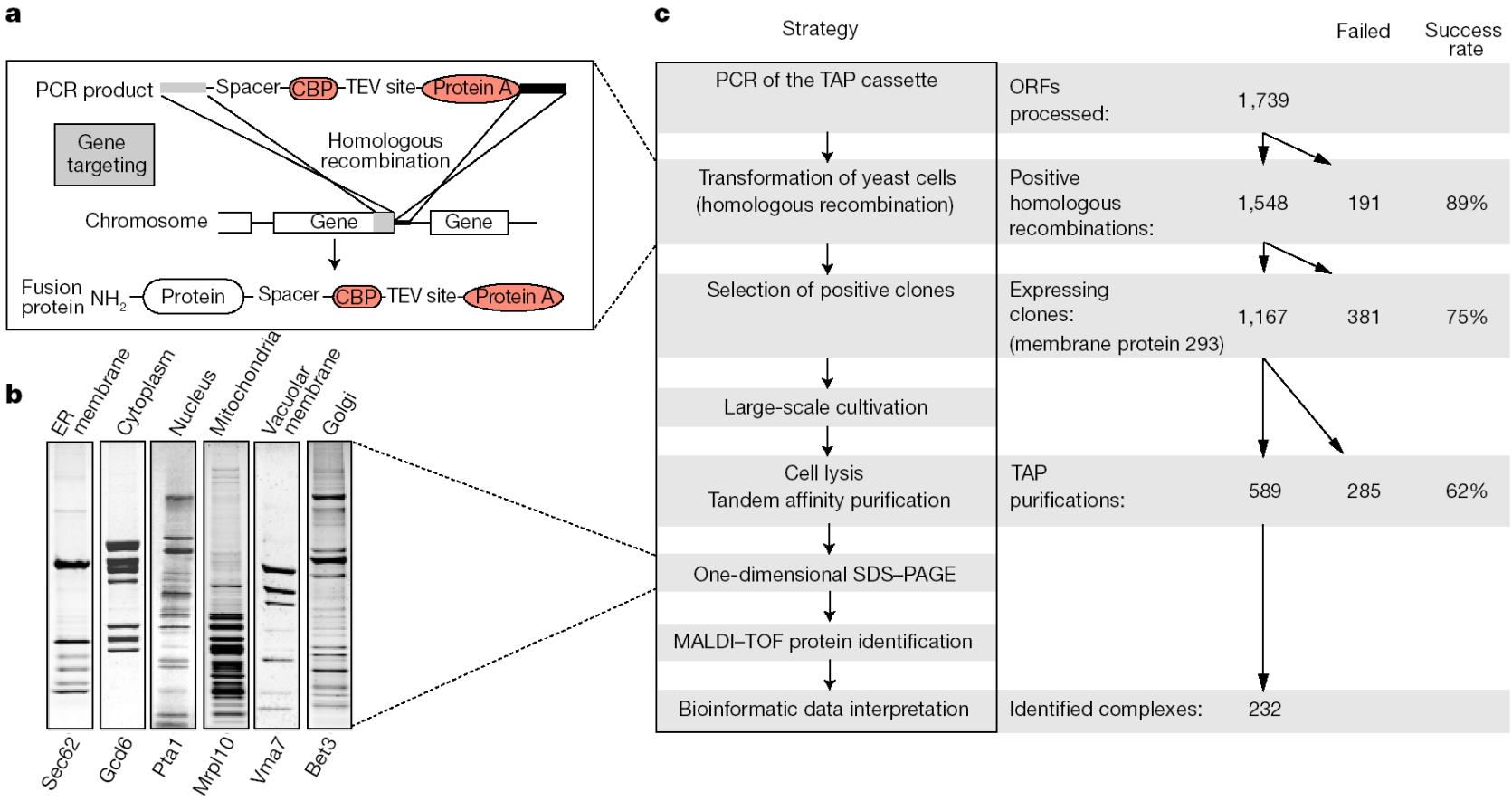


Figure 1 Synopsis of the screen. **a**, Schematic representation of the gene targeting procedure. The TAP cassette is inserted at the C terminus of a given yeast ORF by homologous recombination, generating the TAP-tagged fusion protein. **b**, Examples of TAP complexes purified from different subcellular compartments separated on denaturing protein gels and stained with Coomassie. Tagged proteins are indicated at the bottom. ER, endoplasmic reticulum. **c**, Schematic representation of the sequential steps used for the purification and identification of TAP complexes (left), and the number of experiments and success rate at each step of the procedure (right).

Nature **2002**, *415*, 141-147.

Connected complexes

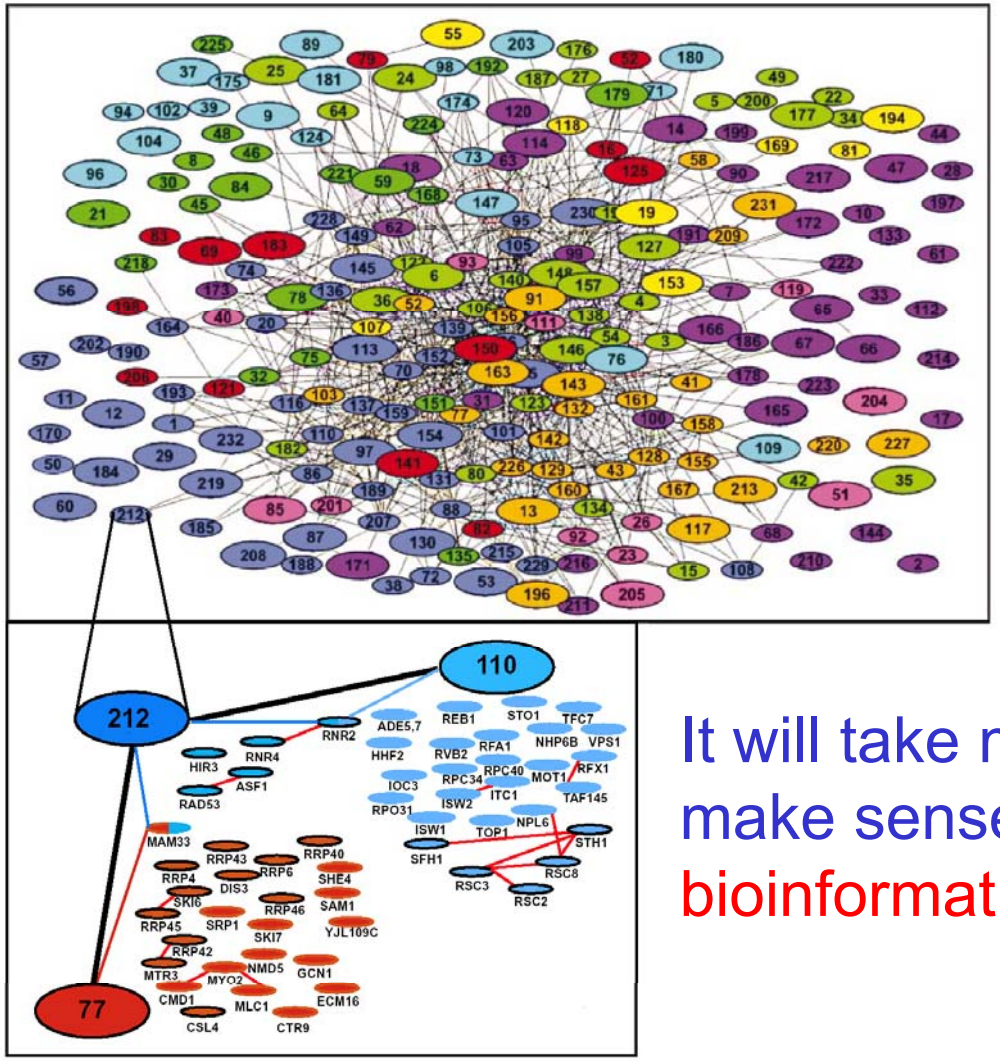


Figure 4 The protein complex network, and grouping of connected complexes. Links were established between complexes sharing at least one protein. For clarity, proteins found in more than nine complexes were omitted. The graphs were generated automatically by a relaxation algorithm that finds a local minimum in the distribution of nodes by minimizing the distance of connected nodes and maximizing distance of unconnected nodes. In the upper panel, cellular roles of the individual complexes (ascribed in Supplementary Information, Table S3) are colour coded: red, cell cycle; dark green, signalling; dark blue, transcription, DNA maintenance, chromatin structure; pink, protein and RNA transport; orange, RNA metabolism; light green, protein synthesis and turnover; brown, cell polarity and structure; violet, intermediate and energy metabolism; light blue, membrane biogenesis and traffic. The lower panel is an example of a complex (yeast TAP-C212) linked to two other complexes (yeast TAP-C77 and TAP-C110) by shared components. It illustrates the connection between the protein and complex levels of organization. Red lines indicate physical interactions as listed in YPD[22].

It will take many years to make sense of these data: bioinformatics←→techniques

Nature **2002**, *415*, 141-147.

TAP in mammalian Cells: need to improve yield, sensitivity

then eluted with EGTA. Although this technology has been applied successfully in mammalian cells[12,14], it also has limitations. First, and most importantly, the overall yield of the process is very low. As a consequence, TAP requires large initial quantities of cells (typically 5×10^8–1×10^9 cells). Second, the technology has not been applied to highly differentiated cells (for example, neuronal cells). Third, primary cells have not been amenable to the TAP procedure

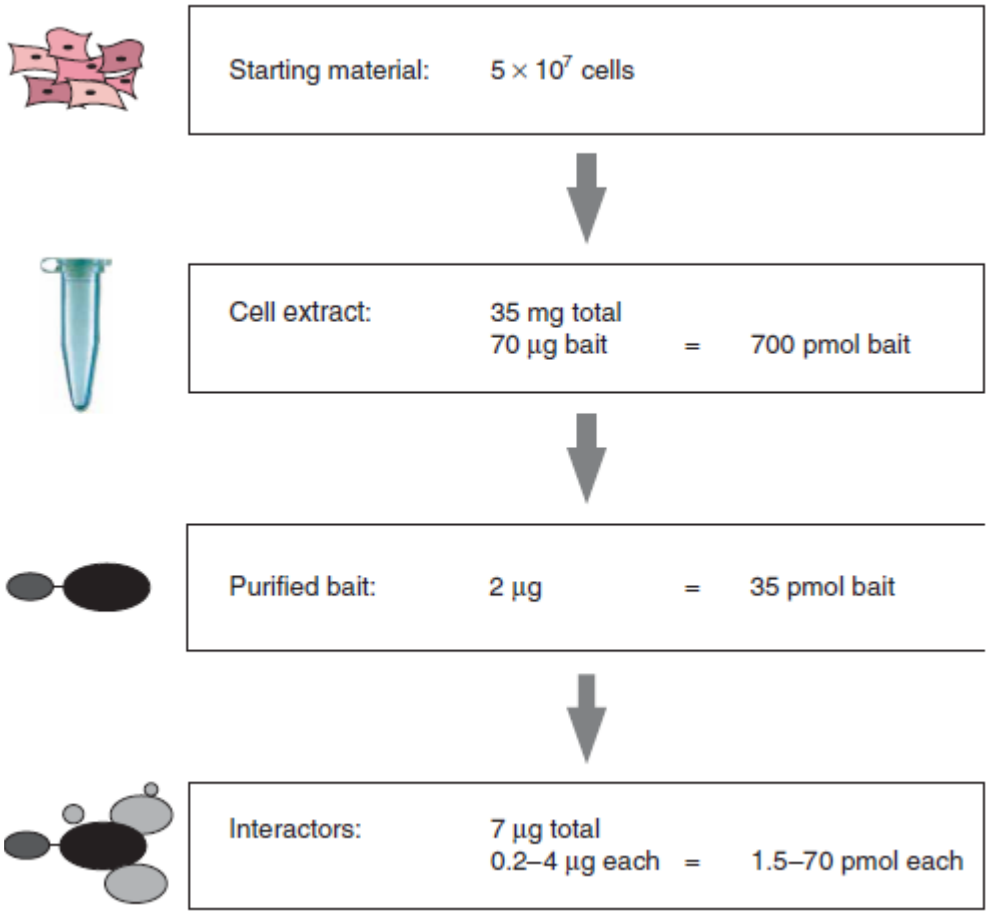


Figure 6 | Quantitative overview of the TAP procedure. A typical TAP purification is done from 5×10^7 cells, yielding 35 mg of total cell extract. Based on the quantification presented in **Figure 4**, the cell extract contains 70 µg (700 pmol) of bait protein. The final eluate contains 2 µg (35 pmol) of bait, which correspond to an overall yield of 5%. Depending on the stoichiometry of the interactions (2:1 to 1:20), 0.2–4 µg (1.5–70 pmol) of interactors will be copurified.

Burckstummer, T.; Bennett, K. L.; Preradovic, A.; Schutze, G.; Hantschel, O.; Superti-Furga, G.; Bauch, A. Nat Methods 2006, 3, 1013-1019.

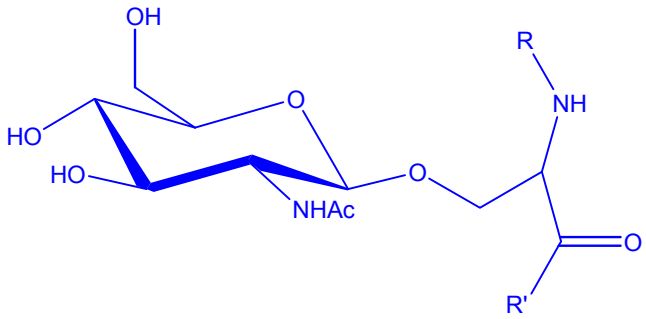
PTM Proteomics: The number of gene products is greater than the number of genes

ABTCDSEFGHTIJKLMNOP

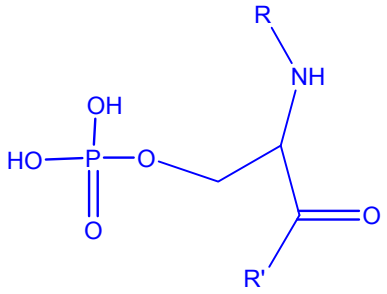
- A peptide sequence with three Ser/Thr residues, each of which may be unmodified, phosphorylated, or *O*-GlcNAcylated
- Number of possible structural variants = $3^3 = 27$

Common PTM-amino acid linkages

Intracellular



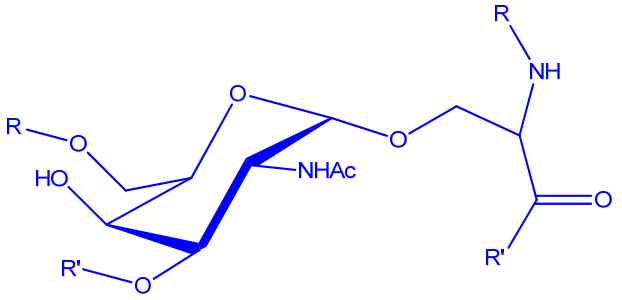
β-GlcNAc-modified Ser/Thr residue
(*O*-GlcNAc)



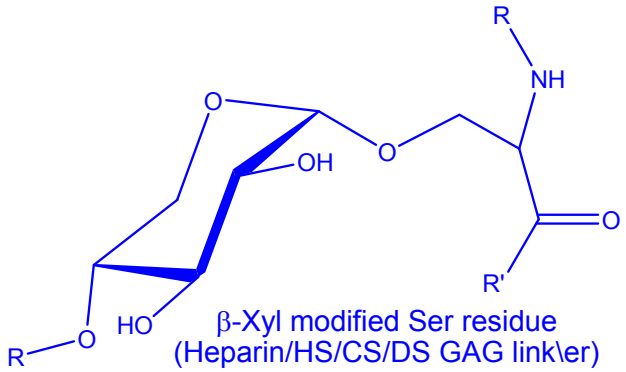
Phosphorylated Ser/Thr residue

Eliminative release

Cell surface, extracellular

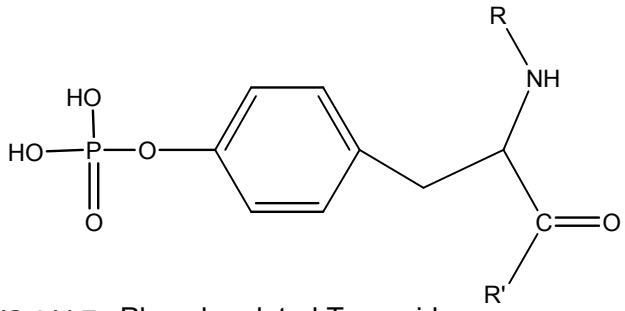


α-GalNAc-modified Ser/Thr residue
(mucin type *O*-linked glycans)

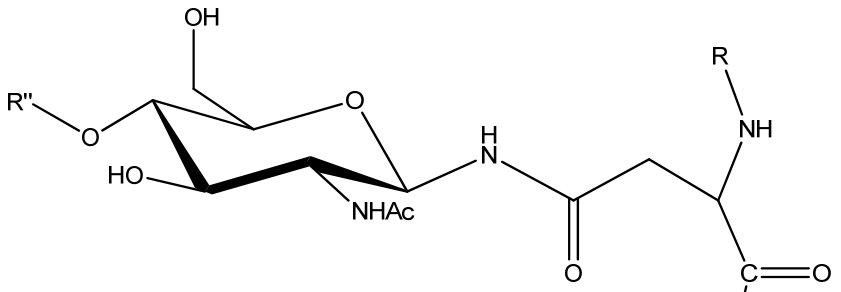


β-Xyl modified Ser residue
(Heparin/HS/CS/DS GAG link\er)

Eliminative release

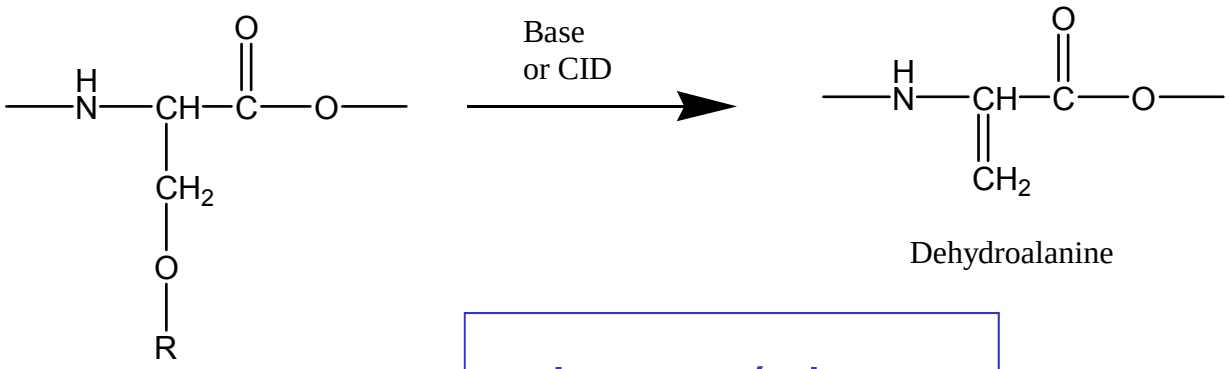


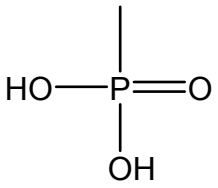
Phosphorylated Tyr residue
(Sulfated Tyrosine)

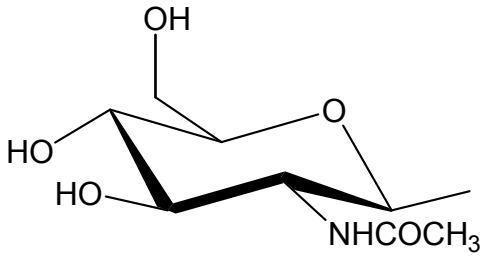


β-GlcNAc modified Asn residue
(*N*-linked glycans)

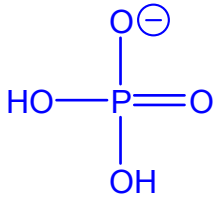
β-Elimination of PTM groups from Ser/Thr residues.



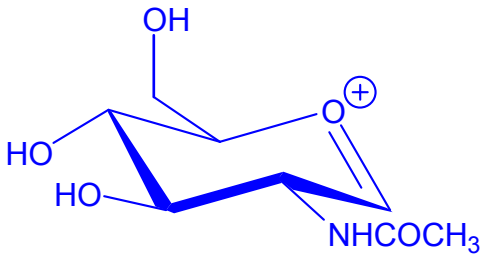

R = phosphate


R = GlcNAc

Low *m/z* ions


$H_2O_4P^-$

Exact Mass: 96.97

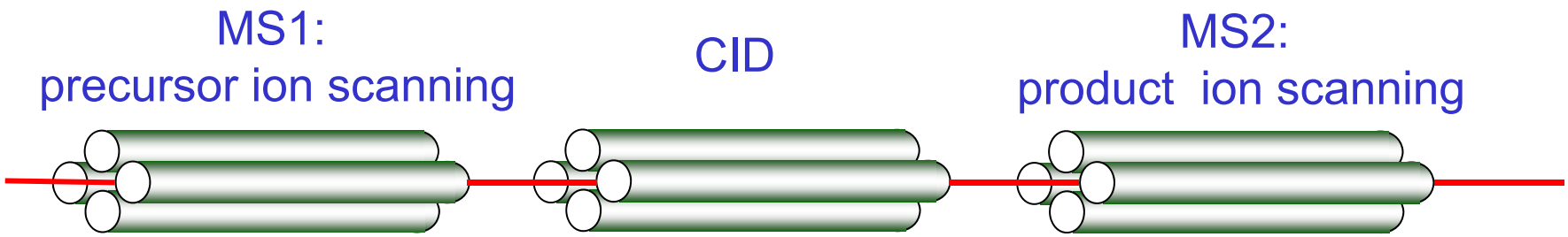

$C_8H_{14}NO_5^+$

Exact Mass: 204.09

Conditions:

- *O*-GlcNAc 1% TEA, 0.1% NaOH, 2 hr. 50ºC
- *O*-GalNAc, 1M NaOH, 2M $NaBH_4$, o/n 20ºC
- *O*-phosphate, 4M LiOH, 1% ethanedithiol, 37ºC, 1 hr.

Precursor ion scans and neutral loss scans



Product ion scan

Fixed *m/z*	fragmentation	full scan

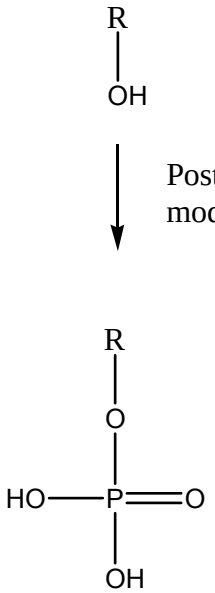
Precursor ion scan

Scanning	fragmentation	Fixed *m/z*

Constant neutral loss scan

Scanning	fragmentation	Scaning a set *m/z* loss

Mass shifts for phosphorylation and sulfation



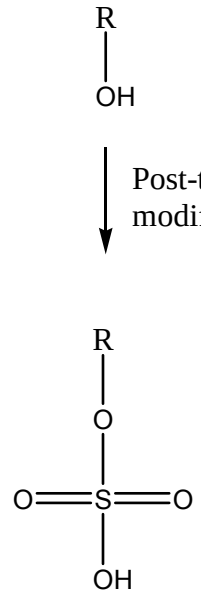
mass shift 79.96688

Neg CID

PO_2^-
m/z 62.96414
PO_3^-
m/z 78.95905
$H_2PO_4^-$
m/z 96.96962

Pos CID

$[M-H_3PO_4+H]^+$
loss of 97.97744
$[M-HPO_3+H]^+$
Loss of 79.96688



mass shift 79.95736

Neg CID

HSO_4^-
m/z 96.96010

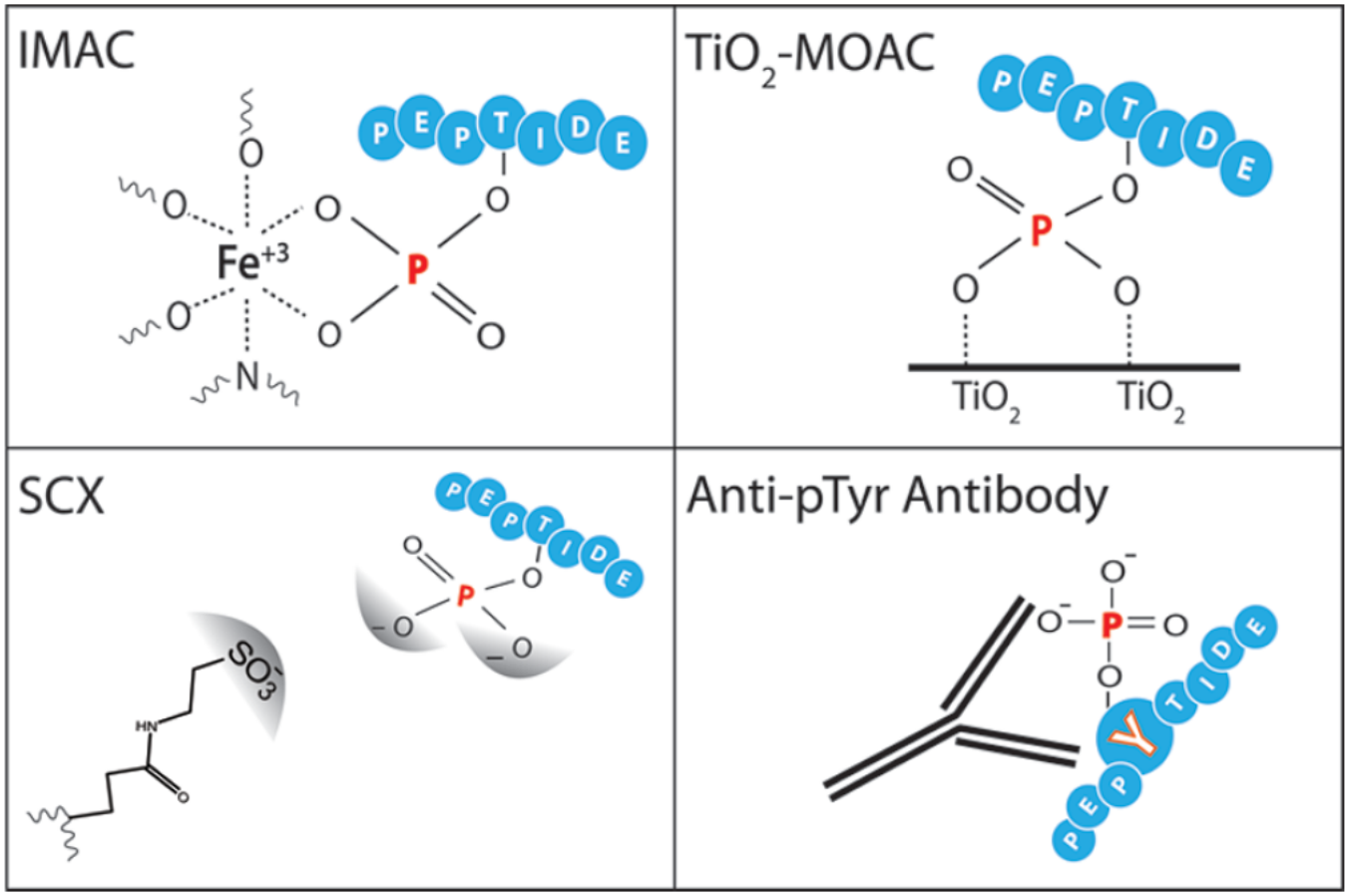
Pos CID

$[M-SO_3+H]^+$
loss of 79.95736

The Phosphoproteome

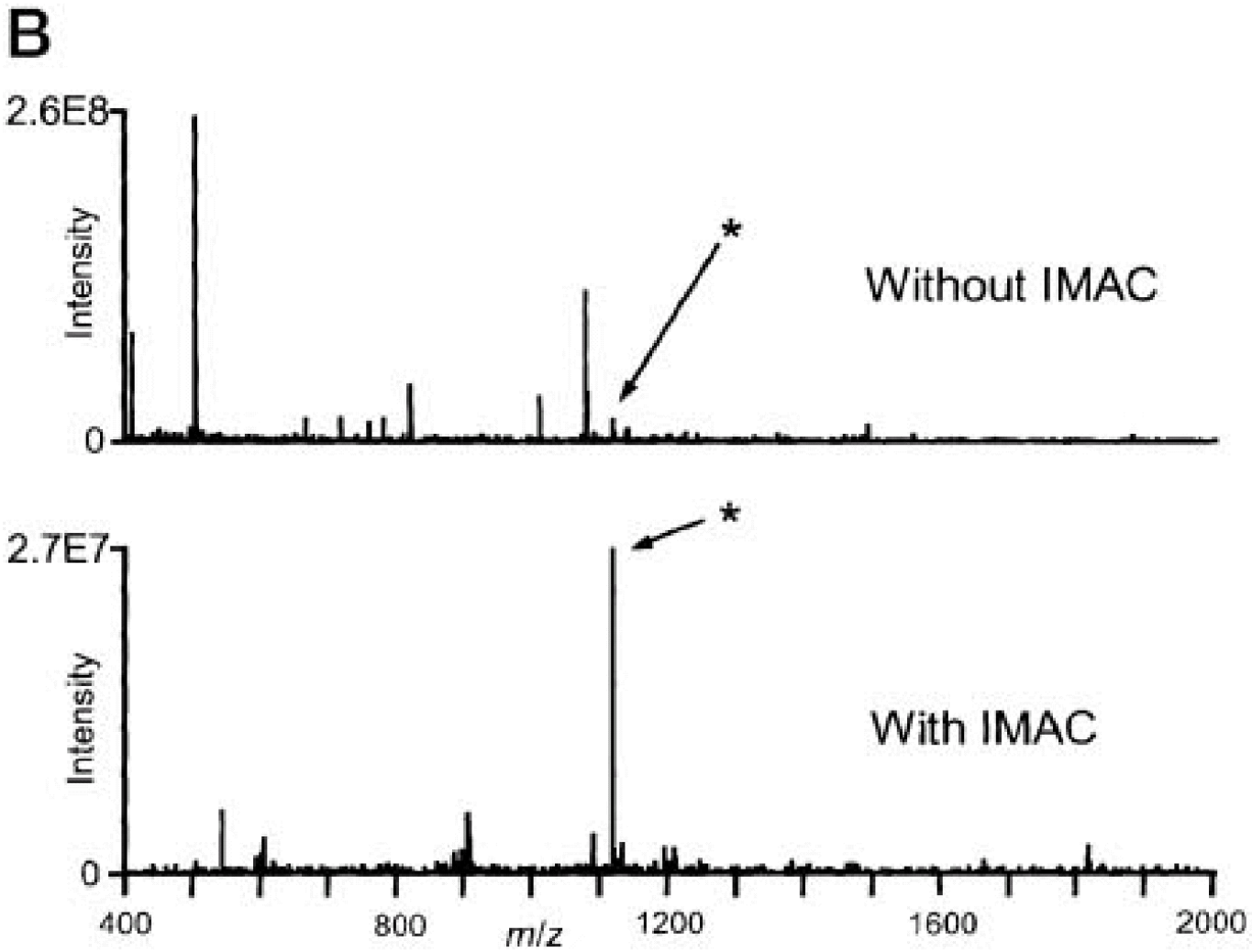
- ~30% of all proteins are thought to be phosphorylated
- Protein kinases are coded by >2000 genes (518 human protein kinases with a conserved catalytic domain)
  - Receptor tyrosine kinases (growth factor receptors: EGFR, FGFR, VEGFR)
  - Non-receptor tyrosine kinases
  - Signal transducing serine/threonine kinases (mitogen activated protein kinases, MAPK)
  - Cyclin dependent kinases (CDKs, pRb phosphorylation required for progression through G1 phase)n
- Identification of phosphorylation sites is a challenge
  - Phosphorylation is dynamic, it is possible that only a few percent of the sites of a gene product are phosphorylated at a given time
  - 100K potential human phosphoylatioin sites
  - Phospho groups are somewhat labile during MS/MS, losses of 98 from precursor and product ions often observed

Affinity enrichment of phosphopeptides



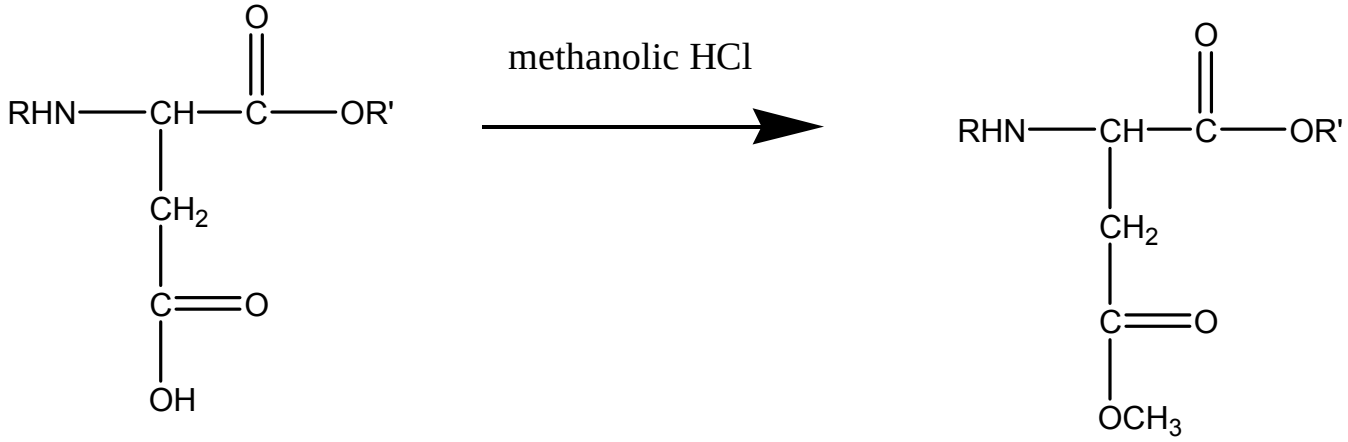
[1]P.A. Grimsrud, D.L. Swaney, C.D. Wenger, N.A. Beauchene, and J.J. Coon, Phosphoproteomics for the masses. ACS Chem Biol 5 (2010) 105-19.

IMAC: immobilized metal affinity chromatography

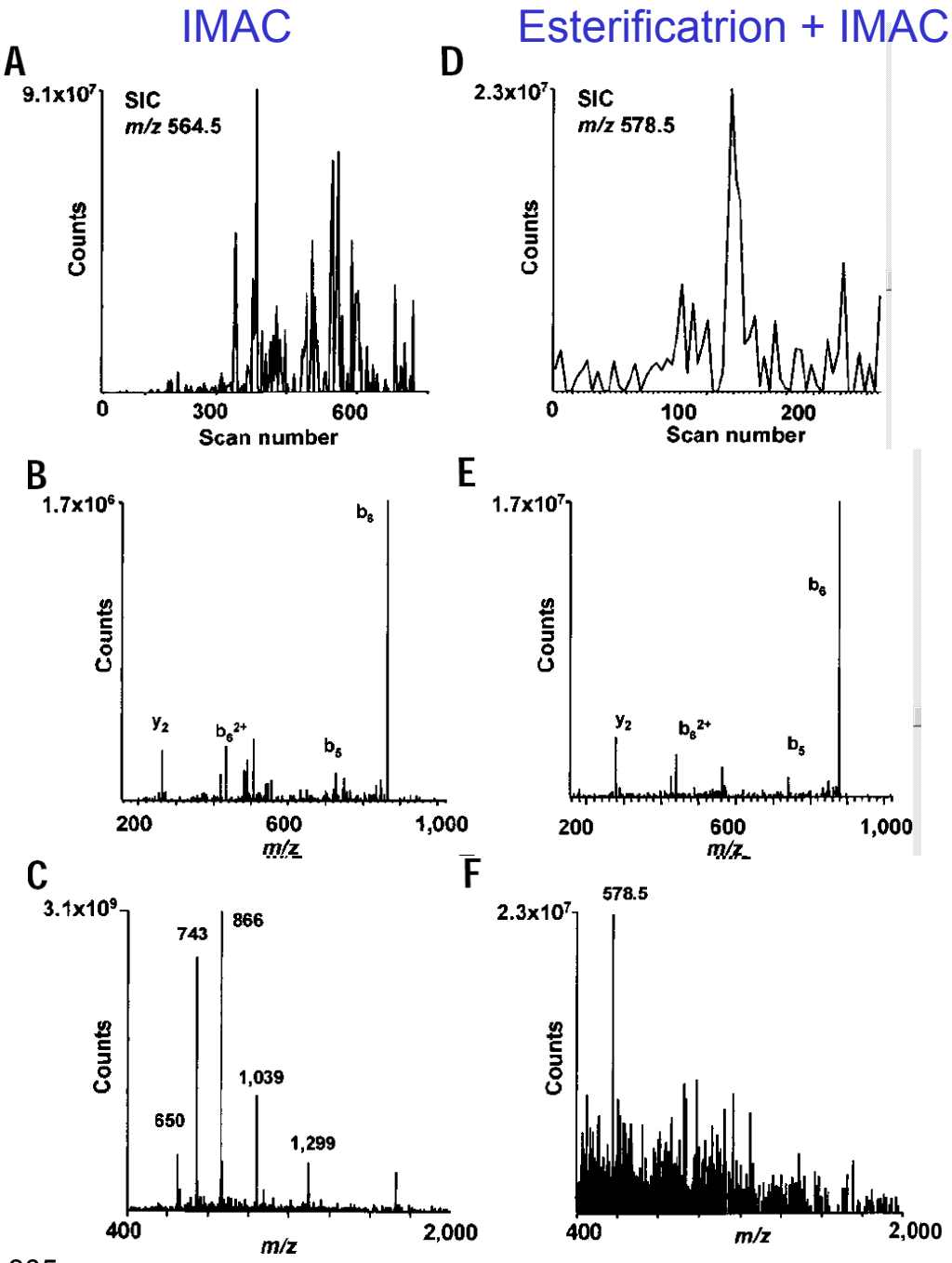


Gioeli, D.; Ficarro, S. B.; Kwiek, J. J.; Aaronson, D.; Hancock, M.; Catling, A. D.; White, F. M.; Christian, R. E.; Settlage, R. E.; Shabanowitz, J.; Hunt, D. F.; Weber, M. J. *J Biol Chem* **2002**, *277*, 29304-29314.

Methyl esterification of acidic amino acid residues

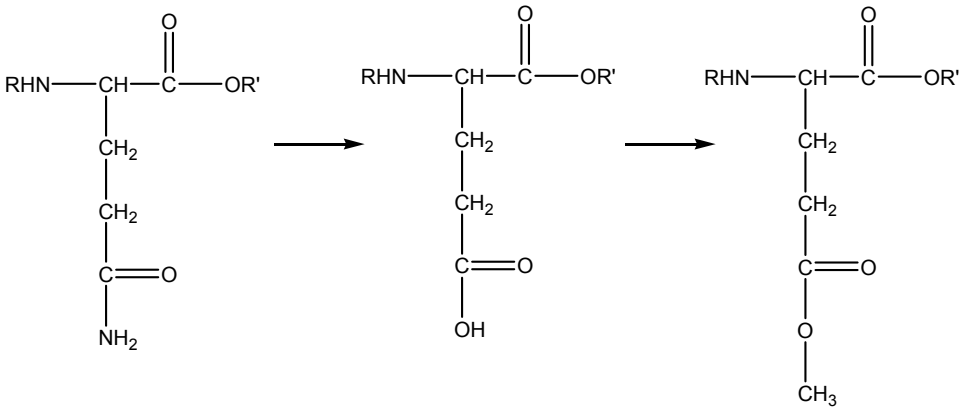


The effect of methyl esterification on IMAC phosphoproteome analysis



Ficarro, S. B.; et al. *Nat Biotechnol* **2002**, *20*, 301-305.

Side reactions in the methyl esterification procedure



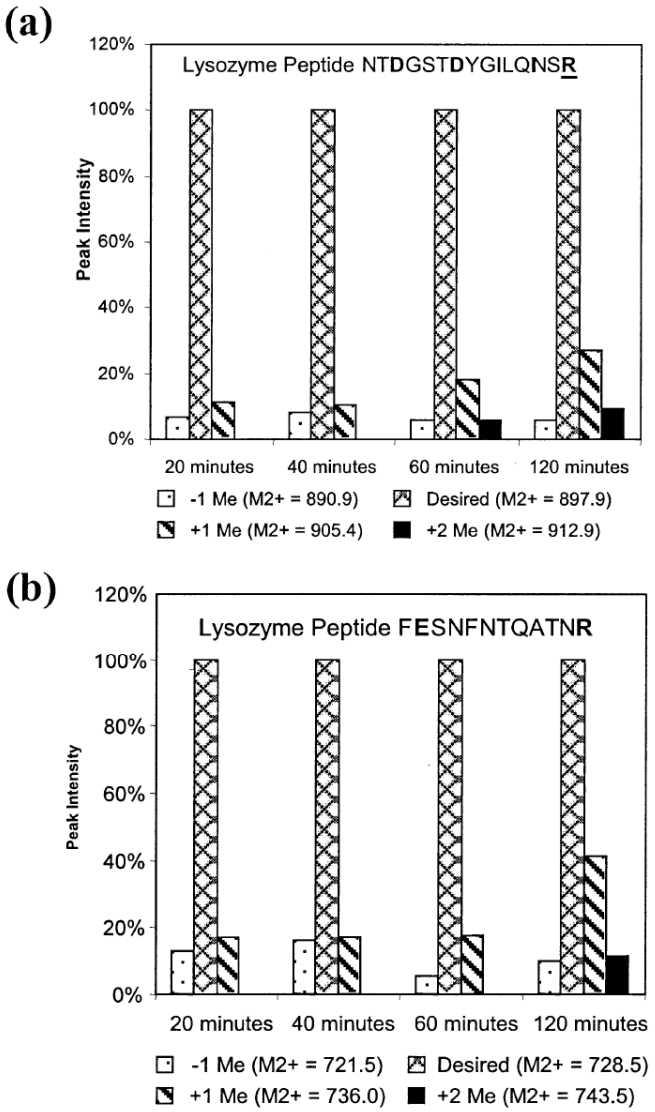


Figure 5. The relative intensity (by XIC of interest ions) of incomplete-reaction and side-reaction for the peptides (NT**D**GST-**D**YGILQINS**R**) (**a**) and (F**E**S**N**FNTQATN**R**) (**b**) with side-reaction on asparagine residues at different reaction times (20, 40, 60, and 120 min). The MS/MS spectra obtained from the peptides modified on asparagine residues after the 120 min reaction are shown in Figure 4.

Enrichment of phosphopeptides using strong cation exchange chromatography

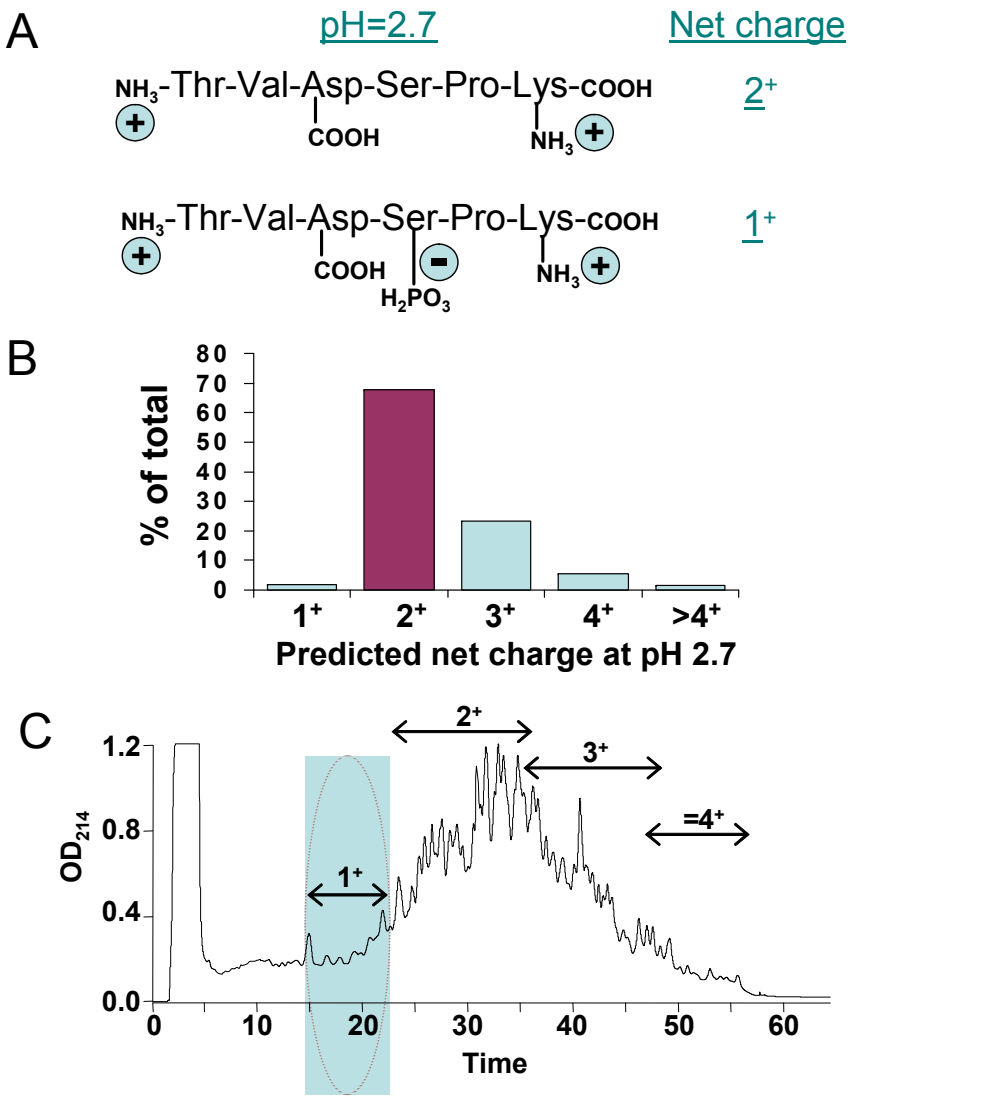
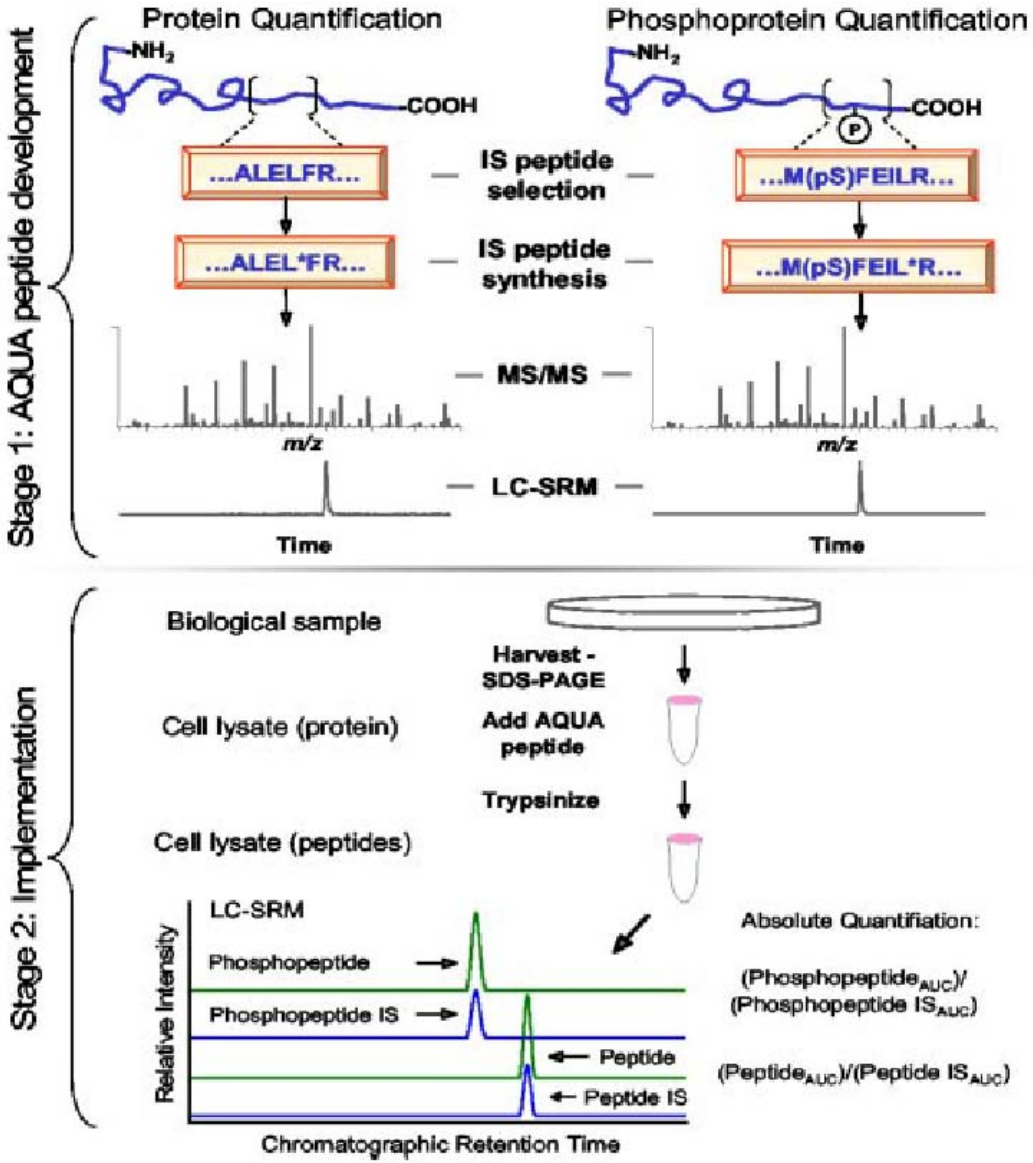


Figure 1. Scheme for phosphopeptide enrichment by strong cation exchange (SCX) chromatography. A) At pH 2.7, peptides produced by trypsin proteolysis generally have a solution charge state of 2^+ while phosphopeptides have a charge state of only 1^+. B) Solution charge state distribution of peptides (5-40 amino acids in length) produced by a theoretical digestion of the human protein database with trypsin (n=6.8 x 10^8 peptides). Sixty-eight percent of the predicted peptides have net charge of 2^+. Any peptide in this category would shift to a 1^+ charge state upon phosphorylation. C) SCX chromatography separation at pH 2.7 for a highly complex peptide mixture of human proteins after trypsin digestion. The circled region is highly enriched for phosphopeptides.

Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villen, J.; Li, J.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P. *PNAS* **2004**, *101*, 12130-12135.

Absolute quantification (AQUA) of proteins and phosphoproteins from cell lysates by tandem MS



Gerber, S. A.; Rush, J.; Stemman, O.; Kirschner, M. W.; Gygi, S. P. *Proc Natl Acad Sci U S A* **2003**, *100*, 6940-6945.

Anal. Chem. **2004**, *76*, 3935−3943

Selective Isolation at the Femtomole Level of Phosphopeptides from Proteolytic Digests Using 2D-NanoLC-ESI-MS/MS and Titanium Oxide Precolumns

Martijn W. H. Pinkse,*,† Pauliina M. Uitto,† Martijn J. Hilhorst,‡ Bert Ooms,‡ and Albert J. R. Heck†

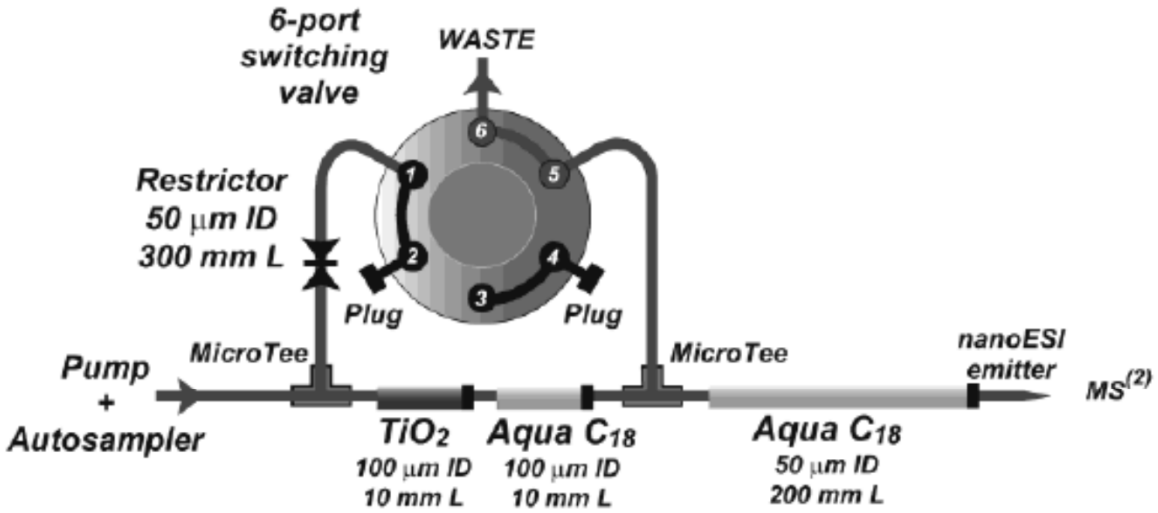


Figure 1. Schematic representation of the two-dimensional LC-MS setup. This setup comprises a six-port switching valve, a dual TiO_2/C_{18} precolumn, and a C_{18} analytical column. During sample loading from the injection loop of the autosampler the flow is 3 μL/min and the restrictor is closed. In this situation, the content of the sample loop is transported directly over the double precolumn, where phosphopeptides are trapped on the TiO_2 particles and non-phosphopeptides are trapped on the C_{18} particles. After sample loading, the six-port valve switches the restrictor to the waste line and simultaneously the flow is increased to 400 μL/min. This results in a pressure increase to ~150 bar and a column flow over the analytical column of ~100 nL/min is obtained. The gradient pump delivers a linear H_2O/acetonitrile gradient, which analytically separates the content of the C_{18} precolumn. After this first analysis, the content of the TiO_2 precolumn is loaded onto the C_{18} precolumn using a strong base. A second H_2O/acetonitrile gradient is used to analytically separate the trapped phosphopeptides.

Anal. Chem. **2004**, *76*, 3935−3943

Selective Isolation at the Femtomole Level of Phosphopeptides from Proteolytic Digests Using 2D-NanoLC-ESI-MS/MS and Titanium Oxide Precolumns

Martijn W. H. Pinkse,*,† Pauliina M. Uitto,† Martijn J. Hilhorst,‡ Bert Ooms,‡ and Albert J. R. Heck†

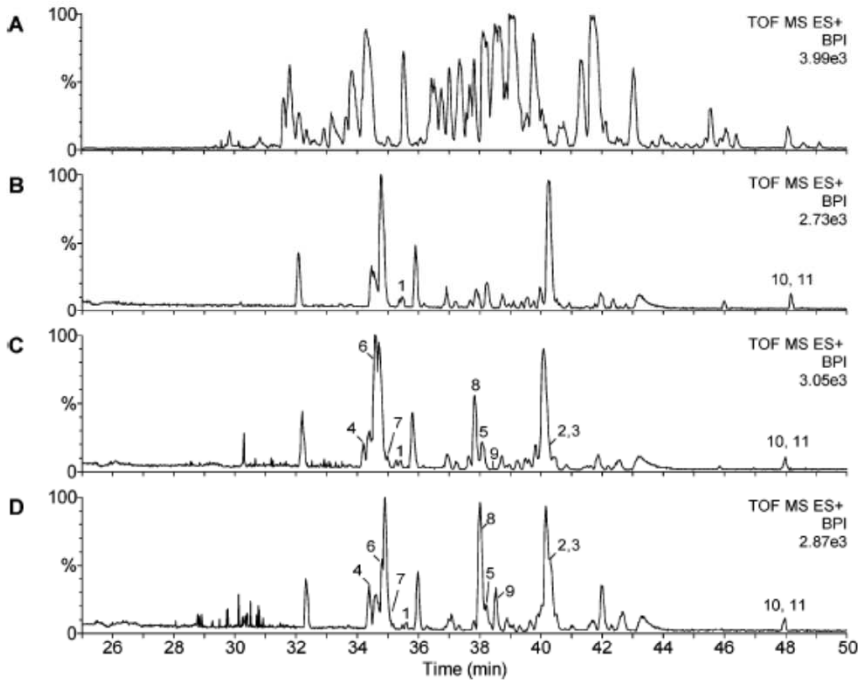


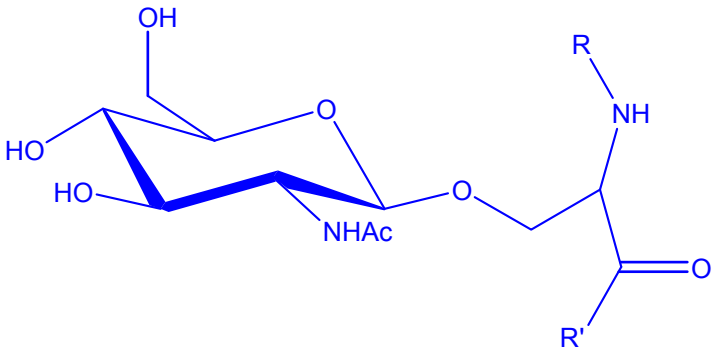
Figure 3. Nano-2D-LC-MS analysis of 250 fmol of tryptic PKG, phosphorylated to different degrees. Shown are the BPI elution profiles of (A) the first gradient of native (nonautophosphorylated) PKG and the second gradients of (B) native PKG, (C) partially autophosphorylated PKG, and (D) highly autophosphorylated PKG. The phosphopeptides, which were identified from low-energy CID spectra in a separate 2D-LC-MS/MS analysis, are labeled and listed in Table 1. Values given for the base peak intensity, listed in the upper right corner of each chromatogram, are in arbitrary units.

Table 1. Identified Tryptic Phosphopeptides from PKG

			mass [b]		peptide observed in[c]		
no.	residues	sequence [a]	$M_{r\ found}$	$M_{r\ calc}$	non	partial	high
1	24−37	(K)-RL**pS**EKEEEIQELKR-(K)	1865.91	1865.92	+	+	+
2	42−56	(K)-C*QSVLPVP**pS**THIGPR-(T)	1726.87	1726.82	−	++	++
3	42−56	(K)-C*Q**pS**VLPVP**pS**THIGPR-(T)	1806.86	1806.78	−	+	++
4	72−77	(R)-**pS**FHDLR-(Q)	853.33	853.35	−	+	++
5	60−77	(R)-AQGISASEPQTYR**pS**FHDLR-(Q)	2155.08	2154.99	−	+	−
6	57−71	(R)-T**pT**RAQGISAEPQTYR-(S)	1757.84	1757.81	−	++	+
7	57−71	(R)-T**pT**RAQGI**pS**AEPQTYR-(S)	1837.82	1837.78	−	+	+
8	57−77	(R)-T**pT**RAQGISAEPQTYR**pS**FHDLR-(Q)	2593.18	2593.15	−	+	++
9	57−77	(R)-T**pT**RAQGI**pS**AEPQTYR**pS**FHDLR-(Q)	2673.21	2673.11	−	+	++
10	514−532	(K)-TW**pT**FC*GTPEYVAPEIILNK-(G)	2318.21	2318.07	+	+	+
11	513−532	(K)-KTW**pT**FC*GTPEYVAPEIILNK-(G)	2446.31	2446.16	+	+	+

[a] Amino acid sequence of phosphorylated peptides identified from tryptic digests on the basis of their low-energy CID spectrum. C* denotes carbamidomethyl cysteine, pS denotes phosphoserine, and pT denotes phosphothreonine. [b] All mass values are listed as monoisotopic mass. [c] Semiquantitative information about the presence of the identified phosphopeptide derived from its elution profiles (extracted ion chromatogram): (−) not present; (+) low abundant; (++) highly abundant.

Analysis of the β-O-GlcNAc-ome



β-GlcNAc-modified Ser/Thr residue
(*O*-GlcNAc)

Ser/Thr *β-O*-GlcNAc Modification

- Many nuclear and clytoplasmic proteins are transiently modified with Ser/Thr-O-GlcNAc.
- All *O*-GlcNAc modified proteins are potential phosphoproteins (reciprocal switches)
- Added by UDP-GlcNAc-peptide-β-GlcNAc transferase (OGT) using UDP-GlcNAc
- Removed by *N*-acetyl-β-D-glucosaminidase (*O*GlcNAcase)
- Anti-*O*-GlcNAc monoclonal antibodies are now available
- Metabolic labeling using Gal transferase and Gal
- *β-O*-GlcNAc is very labile
- Need methods to both enrich *O*-GlcNAc and determine sites of occupancy

Wells, L.; Vosseller, K.; Cole, R. N.; Cronshaw, J. M.; Matunis, M. J.; Hart, G. W. *Mol Cell Proteomics* **2002**, *1*, 791-804.

Affinitry enrichment of *O*-GlcNAc using β-elimination and Michael addition (BEMAD)

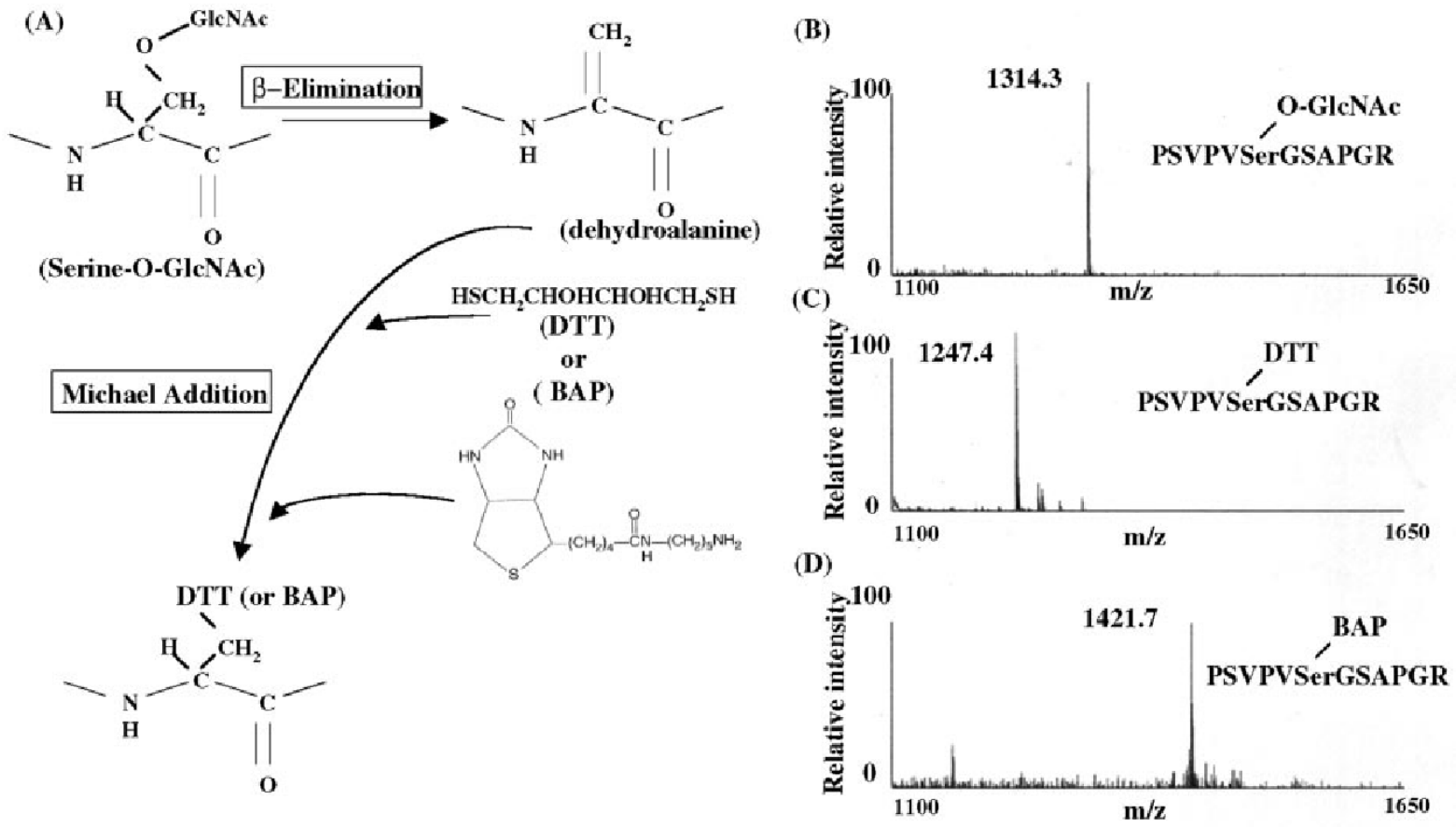
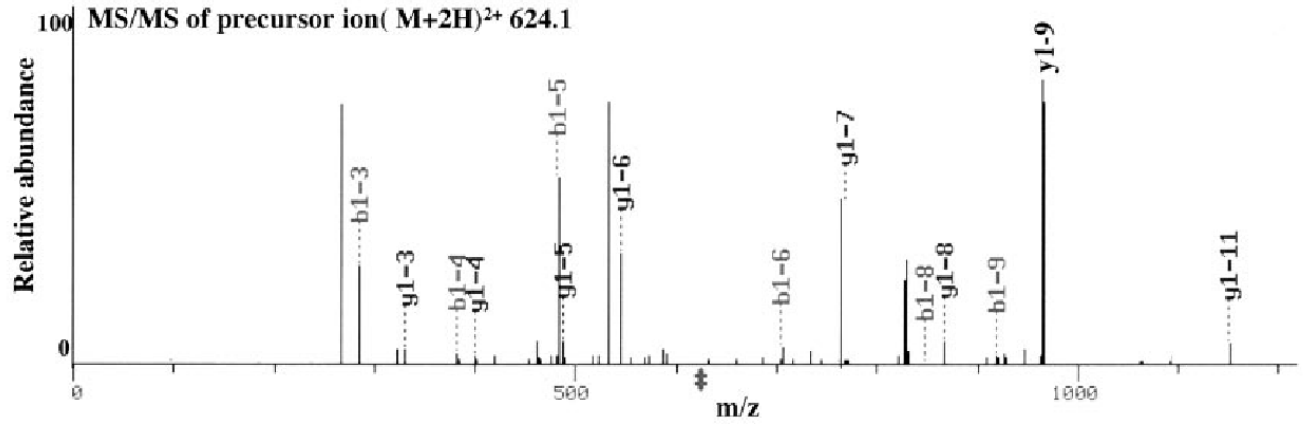


FIG. 2. **β-Elimination of *O*-GlcNAc and replacement with DTT (BEMAD) or BAP through Michael addition chemistry.** *A*, strategy for replacement of serine- or threonine-linked *O*-GlcNAc with the stable affinity tags DTT or BAP after β-elimination. *B*, *C*, and *D*, MALDI-TOF analysis of a synthetic *O*-GlcNAc-modified peptide that was untreated (*B*) or incubated at 50 °C for 2 h in 1% triethylamine, 0.1% NaOH in the presence of 10 mM DTT (*C*) or 20 mM BAP (*D*). Mass shifts in *C* and *D* correspond to loss of *O*-GlcNAc (203 daltons) and addition of DTT (136.2) and BAP (310.5), respectively.

Wells, L.; Vosseller, K.; Cole, R. N.; Cronshaw, J. M.; Matunis, M. J.; Hart, G. W. *Mol Cell Proteomics* **2002**, *1*, 791-804.

Tandem MS of a BEMAD S-DTT peptide

(A)



(B)

Sequest summary:

z	observed MH+	Xcorr	matched ions	protein	peptide sequence
2	1247.3	3.27	15/22	BPP	PSVPVS(*)GSAPGR

(C)

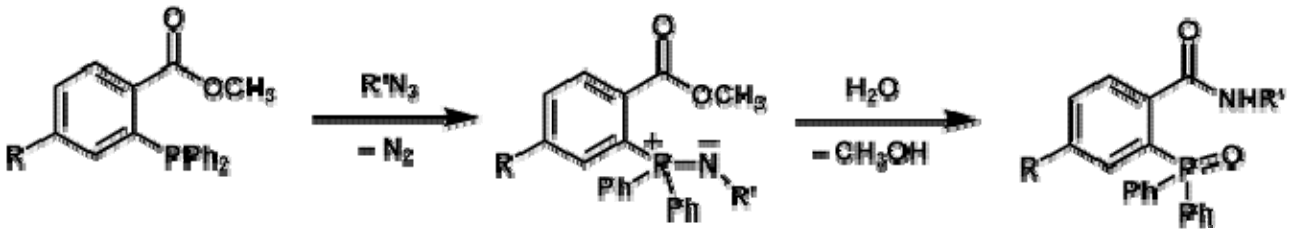
	1247.1	1150	1063	963.9	866.9	767.8	544.3	487.3	400.2	329.2	232.1	175.1- y ions	
	12	11	10	9	8	7	6	5	4	3	2	1	#
Peptide sequence	P	S	V	P	V	S-DTT	G	S	A	P	G	R	
#	1	2	3	4	5	6	7	8	9	10	11	12	
b ions -98.1	185.1	284.2	381.2	480.3	703.8	760.8	847.9	918.9	1016	1073	1229.1		

FIG. 5. **DTT replacement of *O*-GlcNAc through BEMAD is stable during tandem mass spectrometry, allowing for identification of th peptide and the DTT-modified residue.** BEMAD was performed on the peptide PSVPVS(*O*-GlcNAc)GSAPGR, and the sample was analyze by nanospray LC-MS/MS. *A*, MS/MS spectrum from collision-induced dissociation of a precursor ion selected at 624.1 $[M + 2H]^{2+}$. A theoretical b and y ions are indicated by *dashed lines*. *B*, interpretation of MS/MS data in *A* by Turbosequest search against the Owl data bas allowing for addition of 136.2 daltons to serine (*) or threonine (#) correctly identifies the peptide PSVPVS(DTT)GSAPGR. *C*, b and y io fragments correctly interpreted are shown in *bold*. Both the b and y ions ending at the DTT-modified serine are present, making assignmer of the site of modification unambiguous.

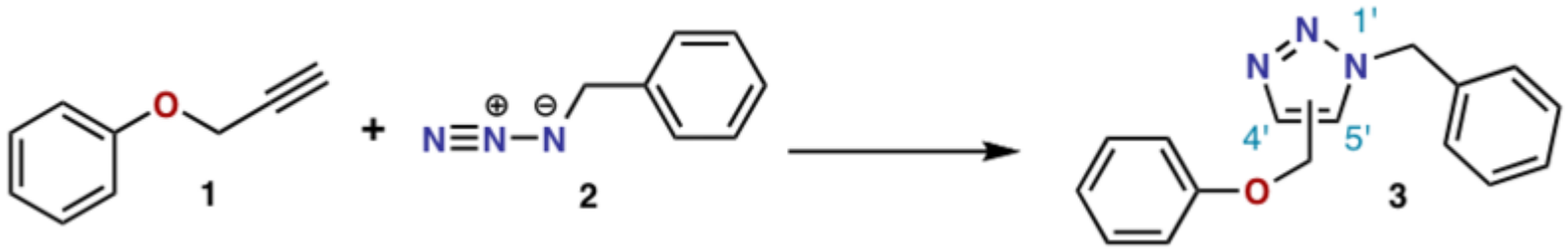
Wells, L.; Vosseller, K.; Cole, R. N.; Cronshaw, J. M.; Matunis, M. J.; Hart, G. W. *Mol Cell Proteomics* **2002**, *1*, 791-804.

Organic chemical ligation reactions

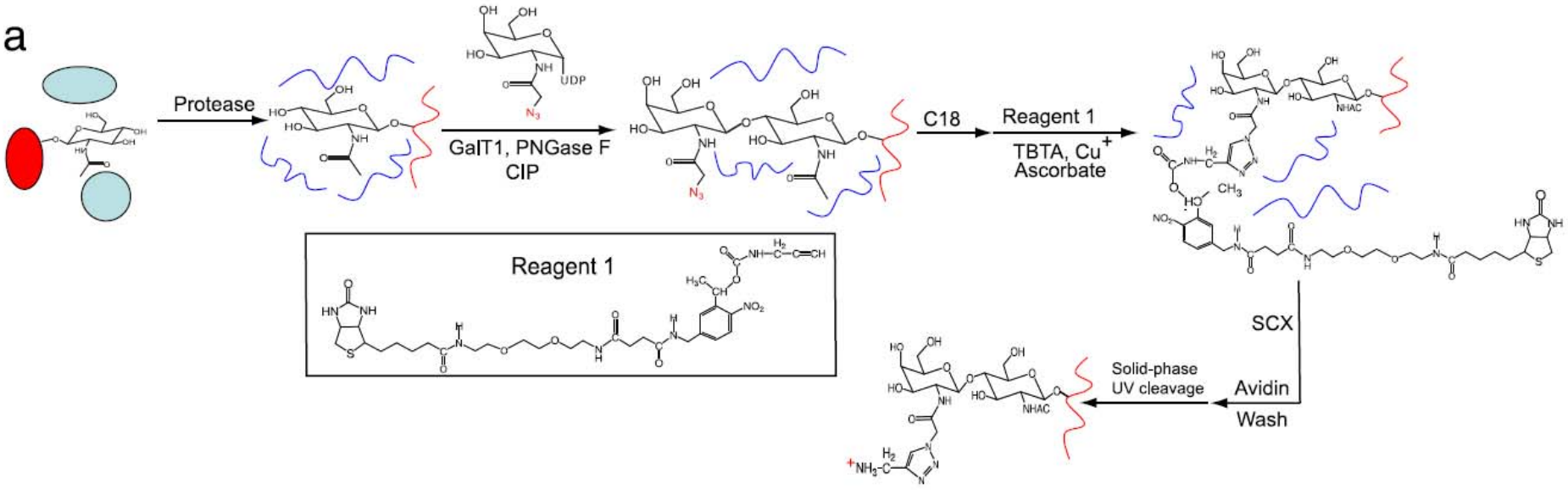
Staudinger ligation:



Click chemistry:

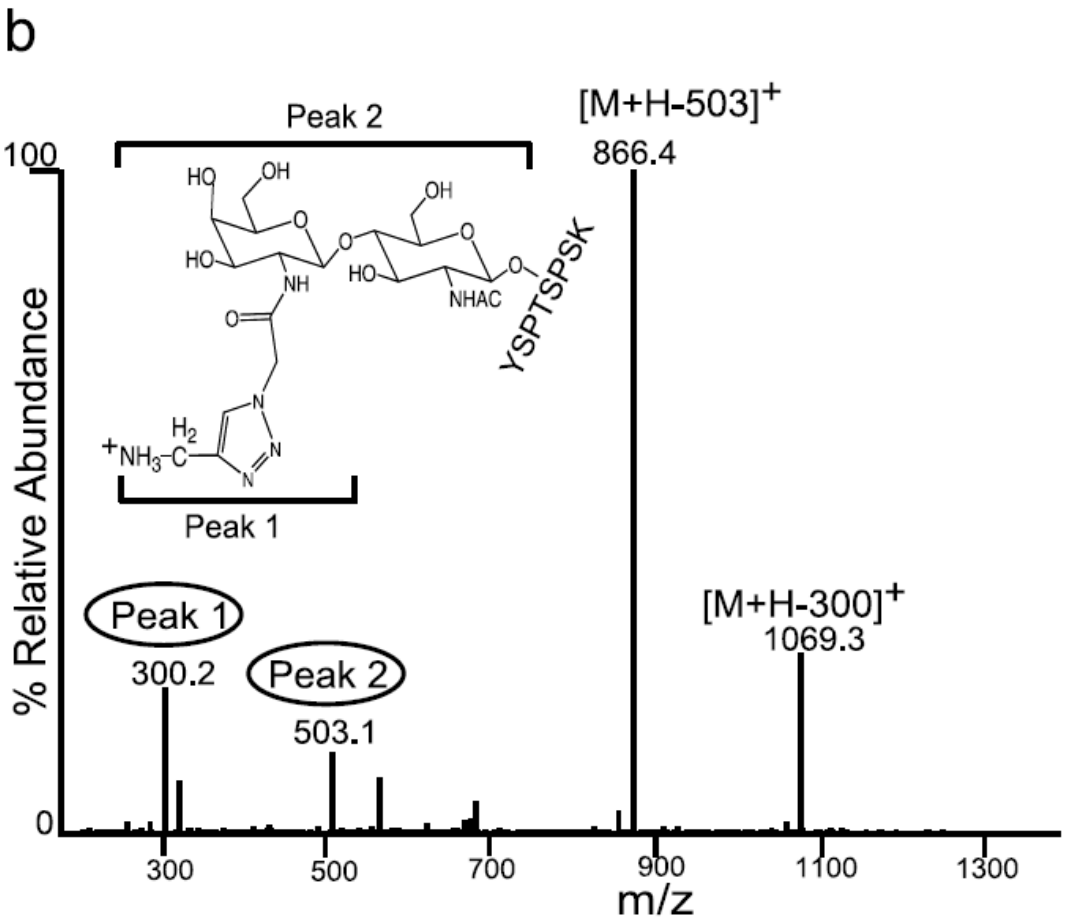


GalT1 labeling and tagging with photocleavable biotin



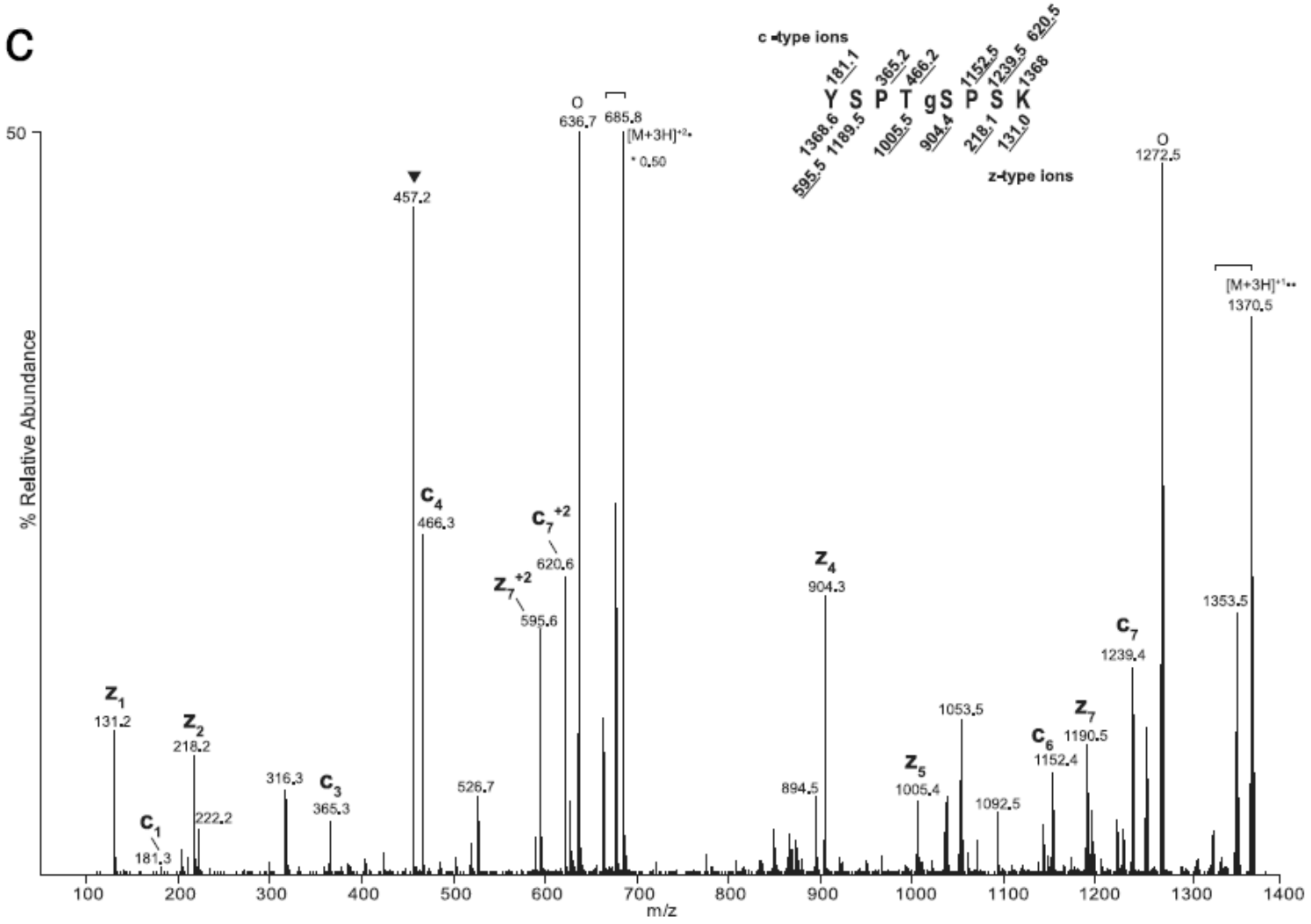
Wang, Z.; Udeshi, N. D.; O'Malley, M.; Shabanowitz, J.; Hunt, D. F.; Hart, G. W. Mol Cell Proteomics 2010, 9, 153-160.

CAD of cleaved O-GlcNAc-peptide



Wang, Z.; Udeshi, N. D.; O'Malley, M.; Shabanowitz, J.; Hunt, D. F.; Hart, G. W. Mol Cell Proteomics 2010, 9, 153-160.

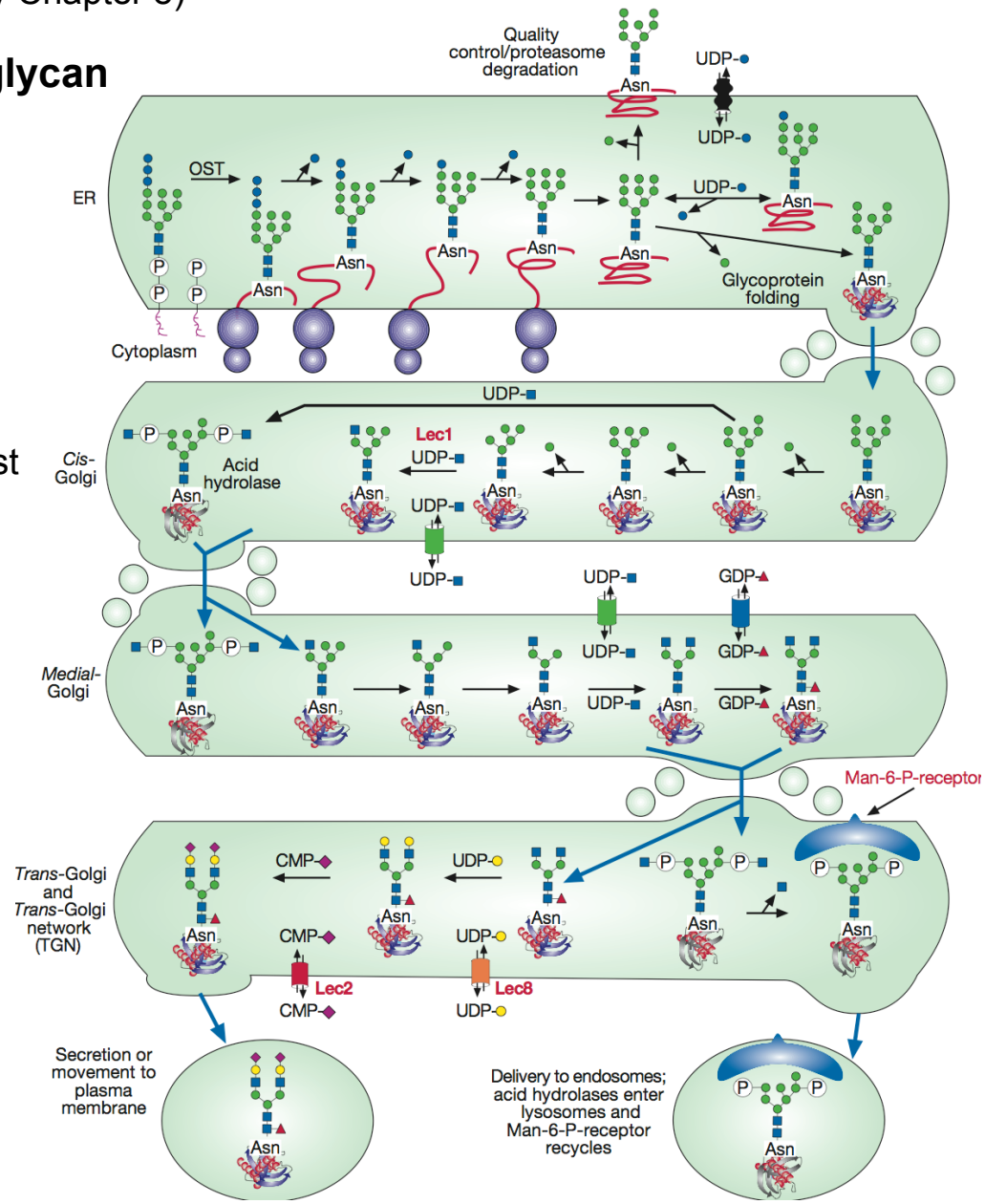
ETD of the same peptide



Wang, Z.; Udeshi, N. D.; O'Malley, M.; Shabanowitz, J.; Hunt, D. F.; Hart, G. W.
Mol Cell Proteomics 2010, 9, 153-160.

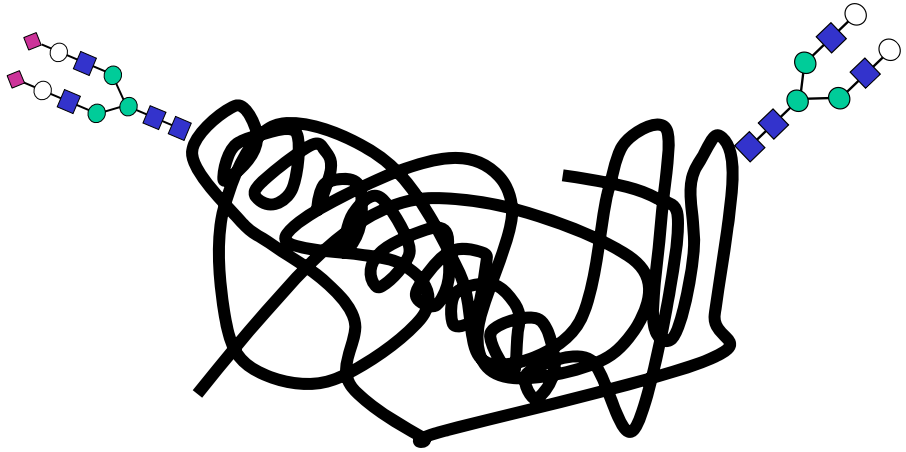
Processing and maturation of an N-glycan

Only Man_5 glycans acted upon by Lec1; most mature glycoproteins contain Man_{5-9} that escape Golgi processing and extension

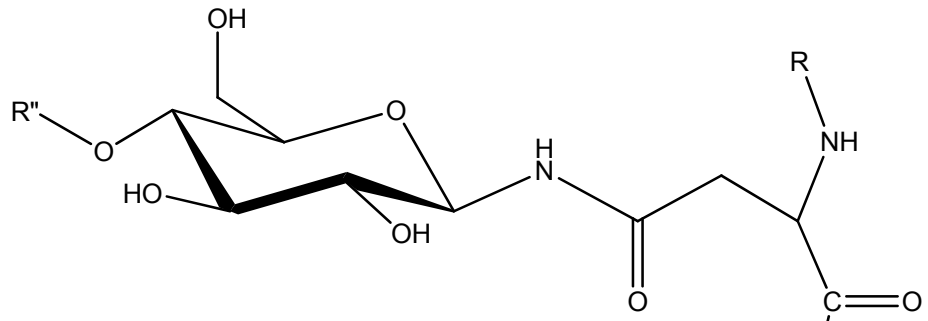


Glycoproteomics:

A glycoprotein

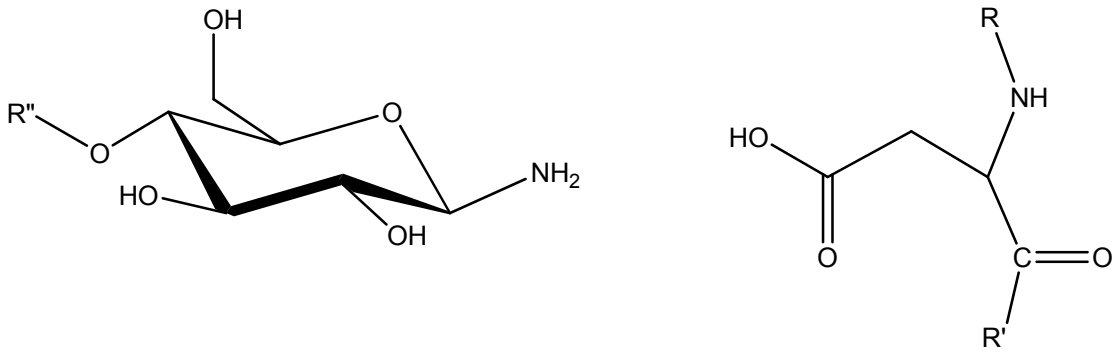


Peptide N-glycosidase F (PNGase F)

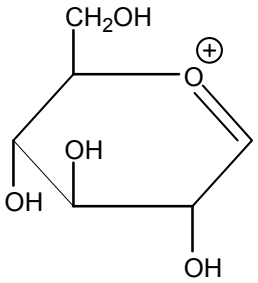


β-GlcNAc modified Asn residue
(*N*-linked glycans)

PNGaseF



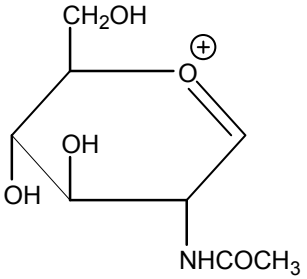
Common glycan oxonium ions



Hex

$C_6H_{11}O_5^+$

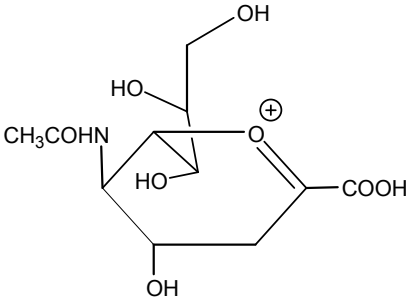
m/z 163.06



HexNAc

$C_8H_{14}NO_5^+$

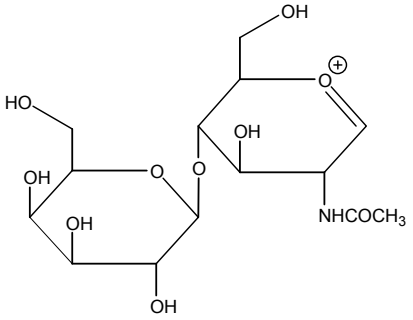
m/z: 204.09



Neu5Ac

$C_{11}H_{18}NO_8^+$

m/z 292.10



HexHexNAc

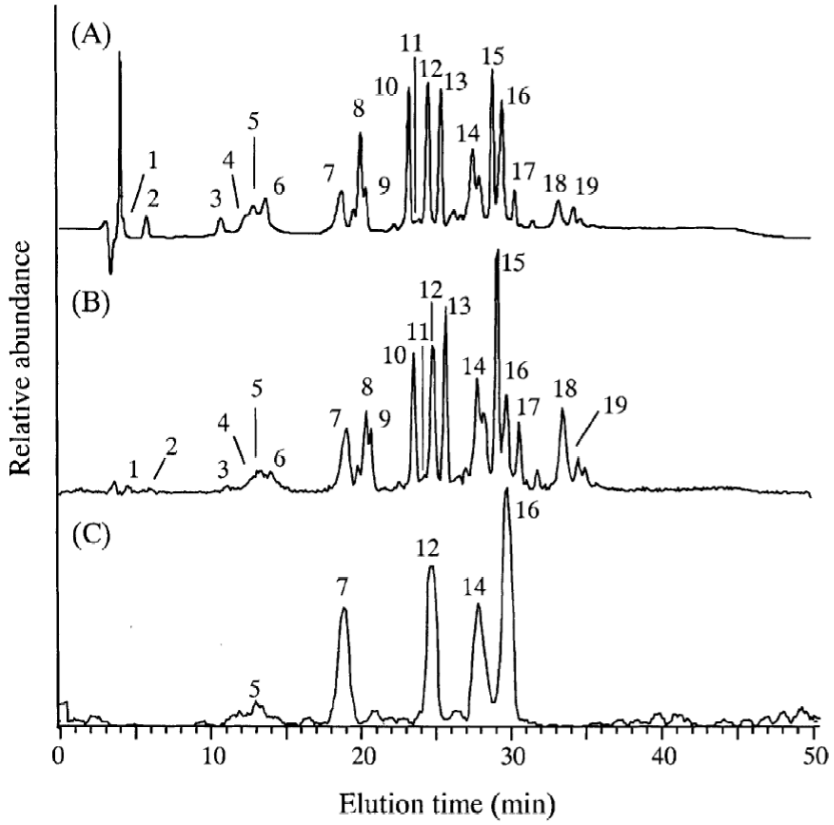
$C_{14}H_{24}NO_{10}^+$

m/z 366.14

- Pos ion CID of glycoconjugates forms oxonium ions, (analogous to peptide immonium ions)
- Precursor ion scans for these ions serve to identify glycopeptides in a glycoprotein digest

Precursor ion scans for glycopeptide detection

Huddleston, M. J., Bean, M. F., and Carr, S. A. *Anal Chem* 1993,65**,** 877-884.



Recombinant human thrombomodulin tryptic digest

(A) UV 206 nm

(B) MS TIC

(C) Precursor ion scan *m/z* 204 ($HexNAc^+$)

Itoh, S.; Kawasaki, N.; Ohta, M.; Hayakawa, T. *Journal of Chromatography A* **2002**, *978*, 141-152.

Enrichment of glycopeptides by periodate labeling followed by amine-affinity chromatography

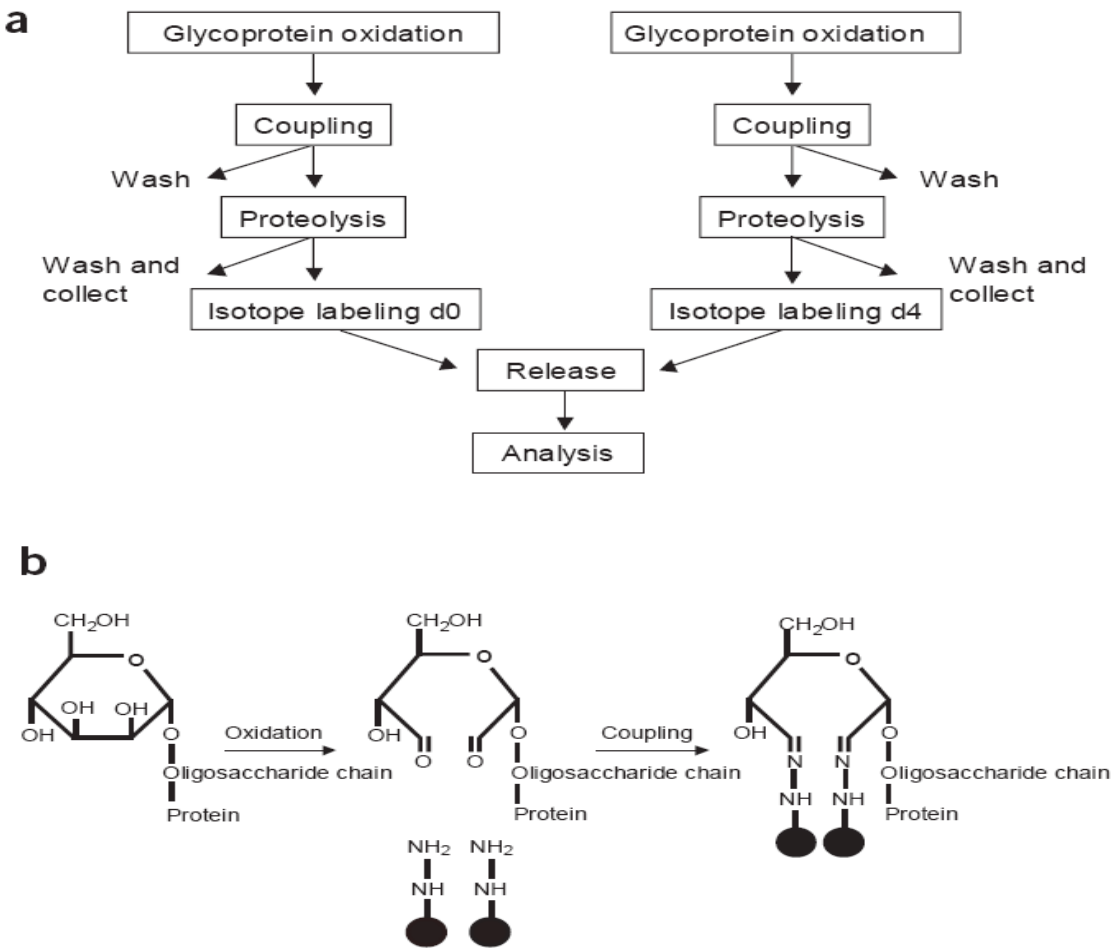


Figure 1 Schematic diagram of quantitative analysis of N-linked glycopeptides. (**a**) Strategy for quantitative analysis of glycopeptides. Proteins from two biological samples are oxidized and coupled to hydrazide resin. Nonglycosylated peptides are removed by proteolysis and extensive washes. The nonglycosylated peptides are optionally collected and analyzed. The N-terminus of glycopeptides are isotope labeled by succinic anhydride carrying either d0 or d4. The beads are then combined and the isotopically tagged peptides are released by PNGase F and analyzed by MS. (**b**) Oxidation of a carbohydrate to an aldehyde followed by covalent coupling to hydrazide resin.

Zhang Nat Biotech 2003 21:660-666

Lectins for affinity enrichments of glycoproteins/glycopeptides

- ~50% of serum/plasma proteins glycosylated
- Lectins are carbohydrate-binding proteins. Many sources and affinities. Used for decades for affinity purification of glycoconjugates and carbohydrates
- For glycoproteomics, intact glycoproteins may be captured, or may be digested with proteases before lectin affinity step.
- Multiple glycos sites on intact glycoproteins may facilitate higher affinity interactions.
- Captured glycopeptides are often deglycosylated prior to tandem MS.
- Lectin targets: sialylated, fucosylated glycans

> The binding of an oligosaccharide to an immobilized lectin may be of high, intermediate, or low affinity. We usually think of binding as being high affinity in nature, requiring haptenic sugars for elution; however, low-affinity binding can be just as useful. In that case the interactive oligosaccharides may be retarded in their elution from the column and well separated from the unbound material, and may not even require haptenic sugars for elution. Such methods have worked extremely well for several types of lectins.[9,14,16–18]

Cummings R. 1994. Use of lectins in analysis of glycoconjugates. Methods in Enzymology Volume 230:66-86.

MLAC: multiple lectin affinity chromatography

- MLAC is a single column with three lectins: ConA; WGA and jacalin
- ConA: specific for glycans with Man and Glc (biantennary NL low affinity, high Man NL, high affinity)
- WGA: GlcNAc, (NeuAc)
- Jacalin: Galβ3GlcNAc of *O*-glycans, αGal of other glycans
- Advantages over single lectin: better overall binding coverage
- Initial work showed that high abundance proteins (albumin, etc) interfere, and better results are obtained when they are depleted (2005. Proteomics 5:3353-3366).
- Poros protein G and Poros anti HSA cartridges used for depletion ($$$)

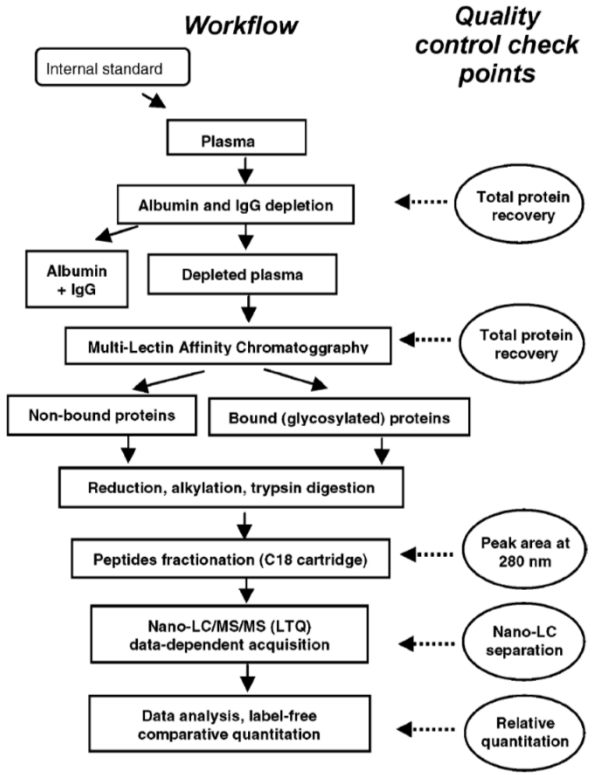
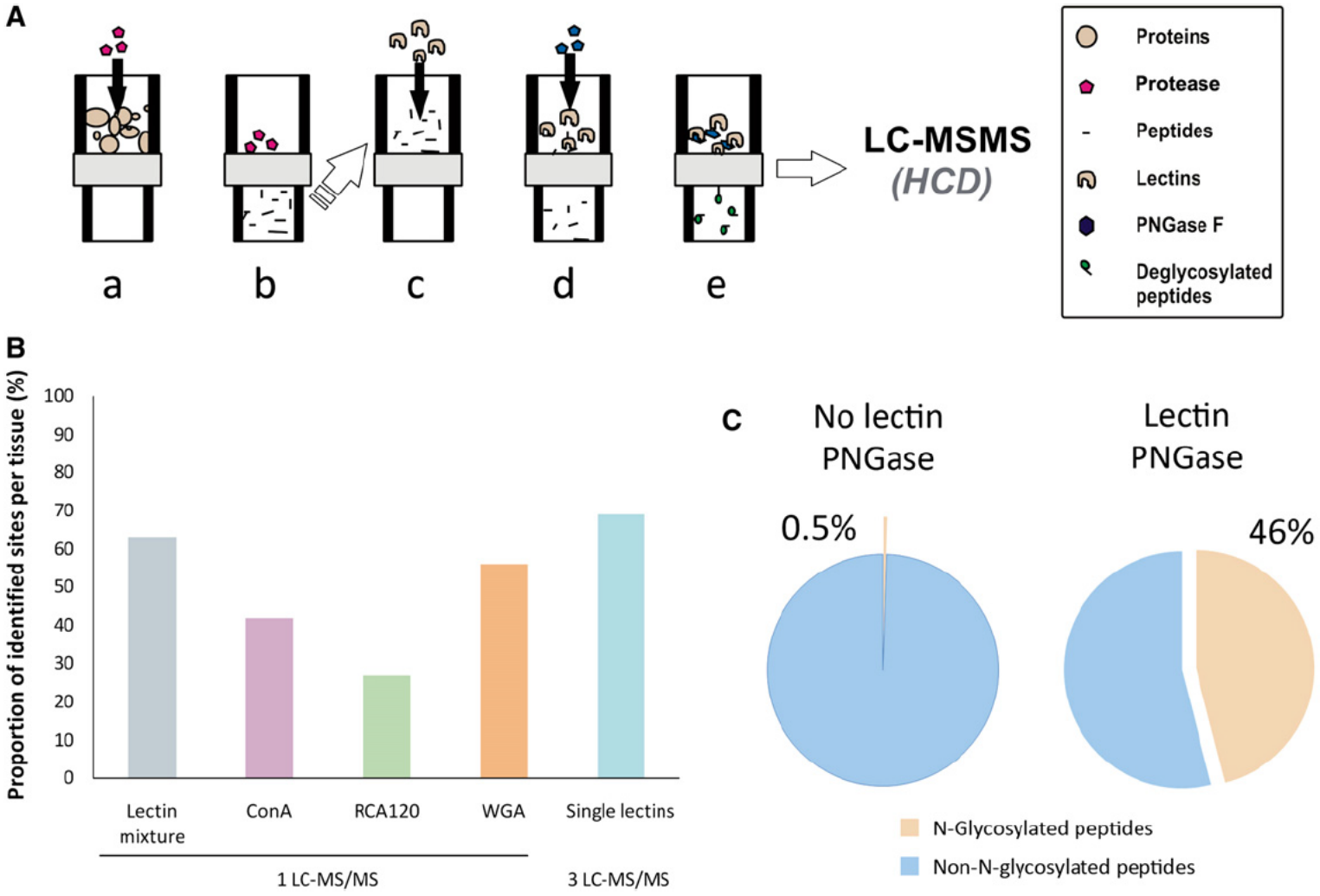


Figure 1. Flowchart of plasma analysis. To maintain the quality of the analysis, standard operating procedures were devised for each step of the method. The following quality control checkpoints were monitored: total protein recovery during the abundant protein depletion and M-LAC fractionation, as calculated from measured total protein concentrations; peak areas of chromatographic traces at 280 nm during peptide fractionation step, to monitor reproducibility of trypsin digestion; retention times and peak areas of selected peptides, as measures of reproducibility of nanoLC separation and comparative quantitation, respectively. To ensure accurate comparative quantitation of proteins in each sample, an internal standard (bovine fetuin) was spiked into plasma samples prior to analysis and used for data normalization.

Plavina T, (2007). Journal of Proteome Research 6:662-671.

Filter-aided sample preparation (FASP)



D.F. Zielinska, F. Gnad, J.R. Wisniewski, and M. Mann, Cell 141 (2010) 897-907.

Proteomic data on FASP lectin enriched glycopeptides

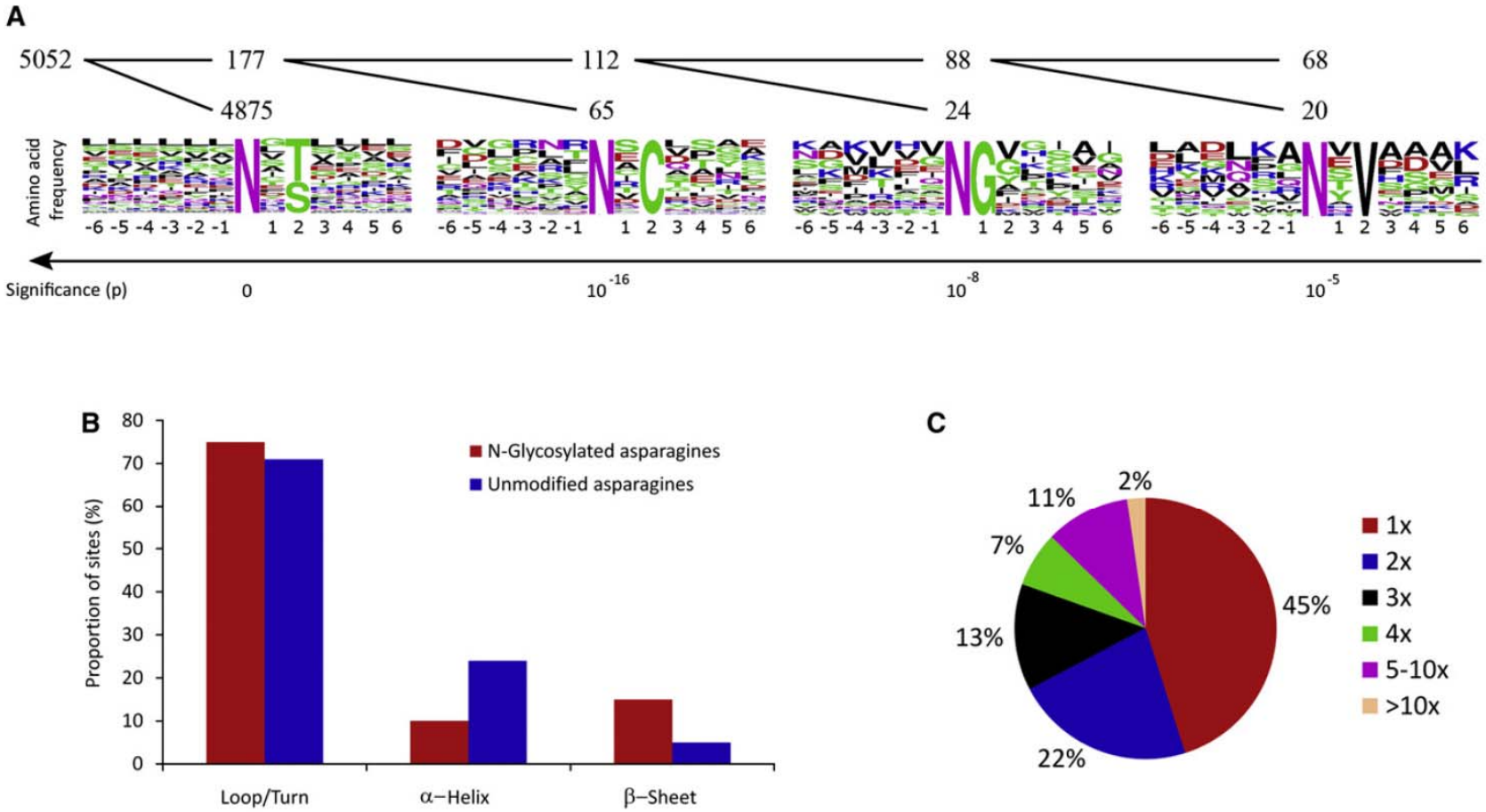


Figure 4. Sequence Recognition Motifs, Structure Preference, and Multiple Glycosylation

(A) N-glycosylation consensus sequence as derived using MotifX. WebLogo (Schneider and Stephens, 1990) was used to create relative frequency plots. The most significant sequence motif is the canonical one, with serine and threonine on position 2. In following iterative steps the consensus sequences N-X-C, N-G, and N-X-V were statistically identified.

(B) Proportion of N-glycosylated and non-N-glycosylated asparagines localized in loops, α helices, and β sheets.

(C) Distribution of singly and multiply glycosylated proteins.

D.F. Zielinska, F. Gnad, J.R. Wisniewski, and M. Mann, Cell 141 (2010) 897-907.