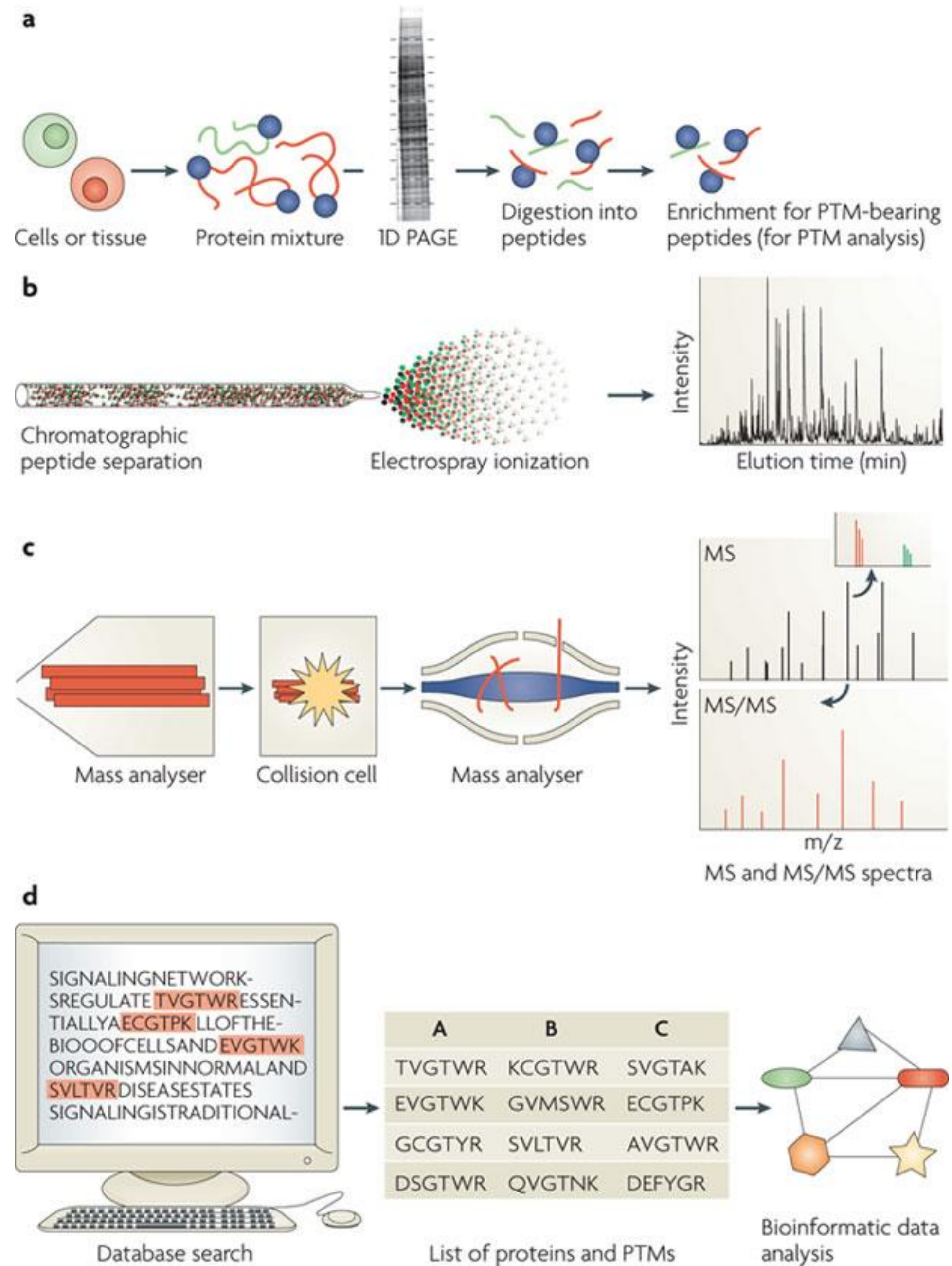
GMS BI 793 Lecture 9
Affinity Mass Spectrometry: Strategies for Proteome Profiling

Joseph Zaia

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: wide angle protein analysis



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

Discovery proteomics: data-dependent analysis (DDA) tandem MS

- When using stable isotope labels, can make pairwise comparisons
- Label free proteomics
  - Instrumental variability an issue
  - Effort, computation time
- Isobaric tagging: iTRAQ, TMT, multiplexing using isobaric tags with tandem MS reporter ions
  - Up to 16-plex

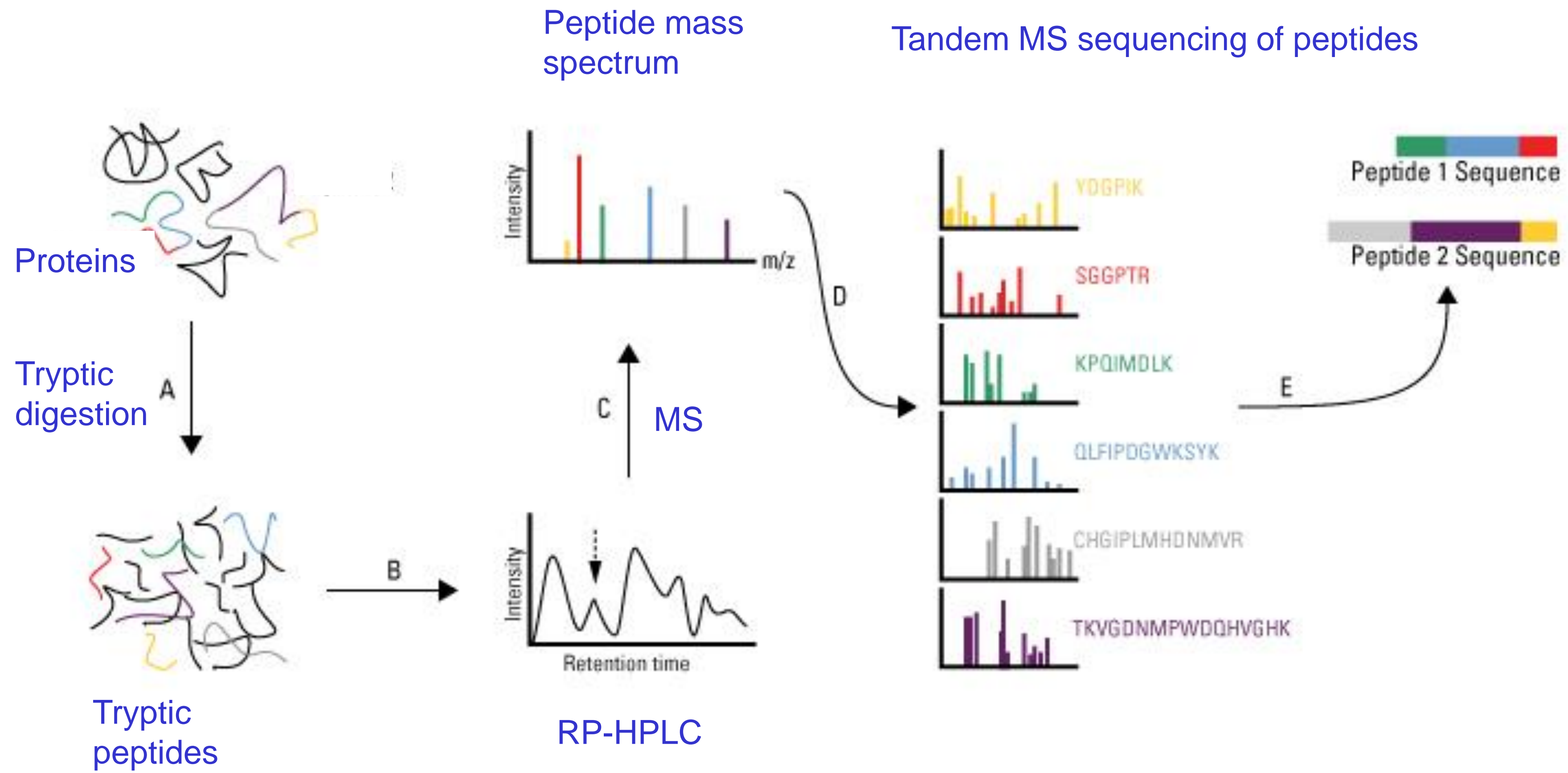
Targeted proteomics: entails limiting scope of analysis so as to measure abundances of targeted peptides

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

Data independent analysis: Acquire alternating MS and tandem MS scans on entire mass window.

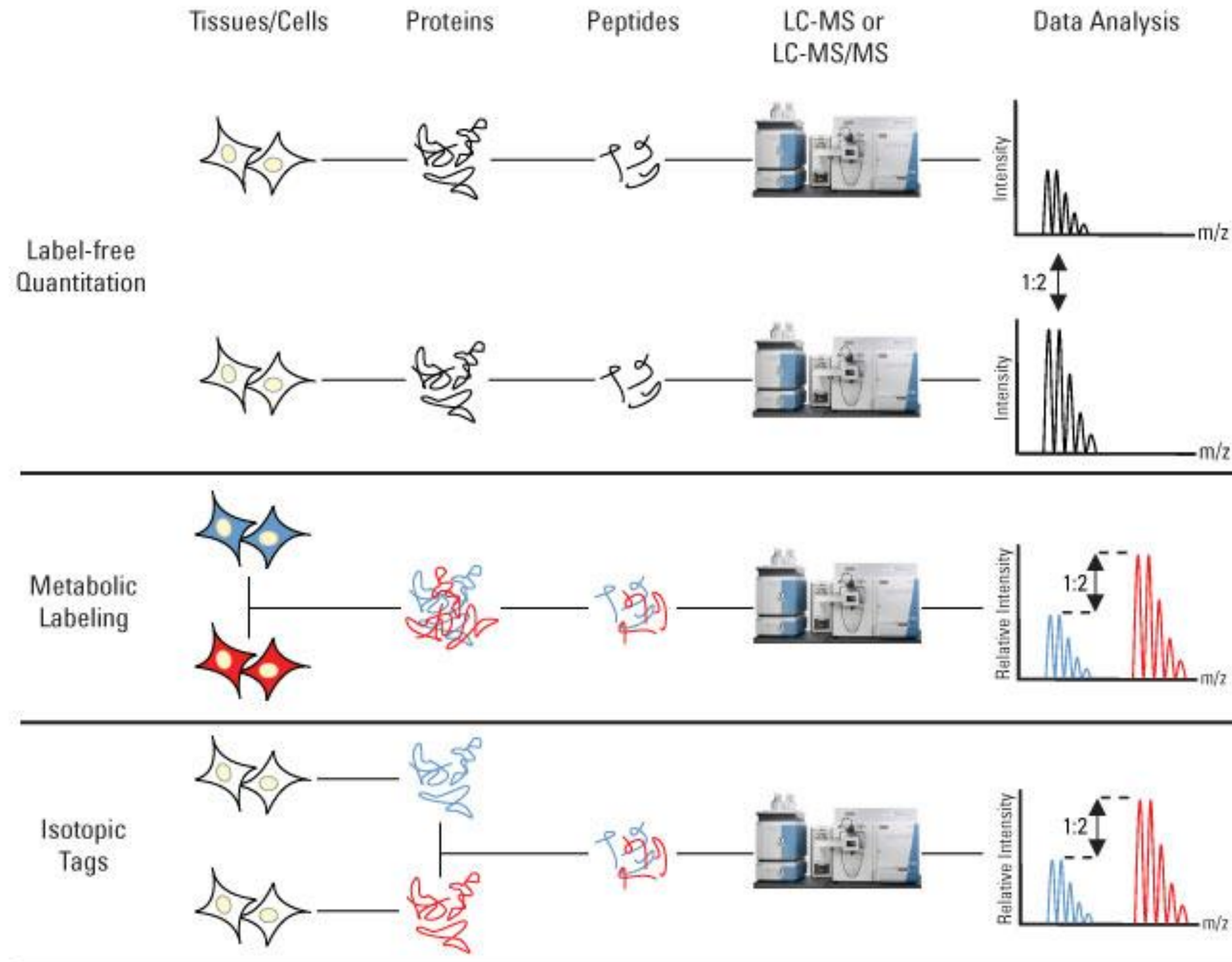
- Scanning tile DIA: set quadrupole to step through the desired mass range. SWATH. Requires spectral library for interpretation of tandem mass spectra.
- MS^E DIA using full quad mass range.

Proteomics: discovery of proteins



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

Proteomics: quantification of proteins



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

Selected reaction monitoring

Targeted proteomics

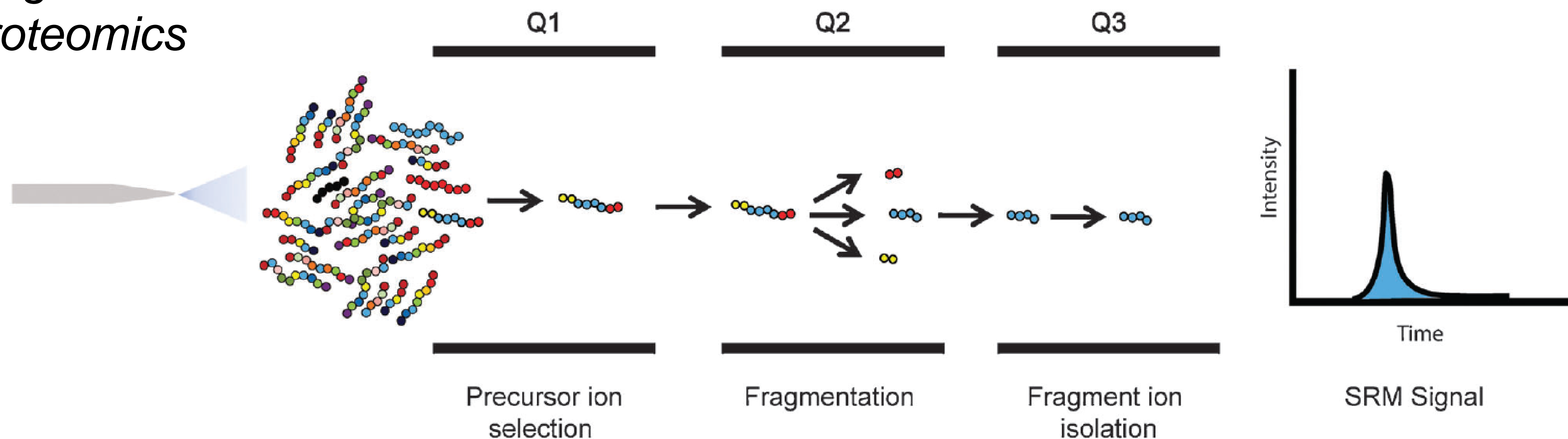


Fig. 3 Overview of selected reaction monitoring mass spectrometry (LC-SRM-MS). Proteotypic peptides are separated from complex biological samples by reversed phase liquid chromatography. The selected proteotypic peptides are isolated in Q1 reducing interfering background signal, and subsequently fragmented 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.

This journal is © The Royal Society of Chemistry 2012
Chem. Soc. Rev., 2012, **41**, 3912–3928 | 3917

Tiers of Targeted MS measurements

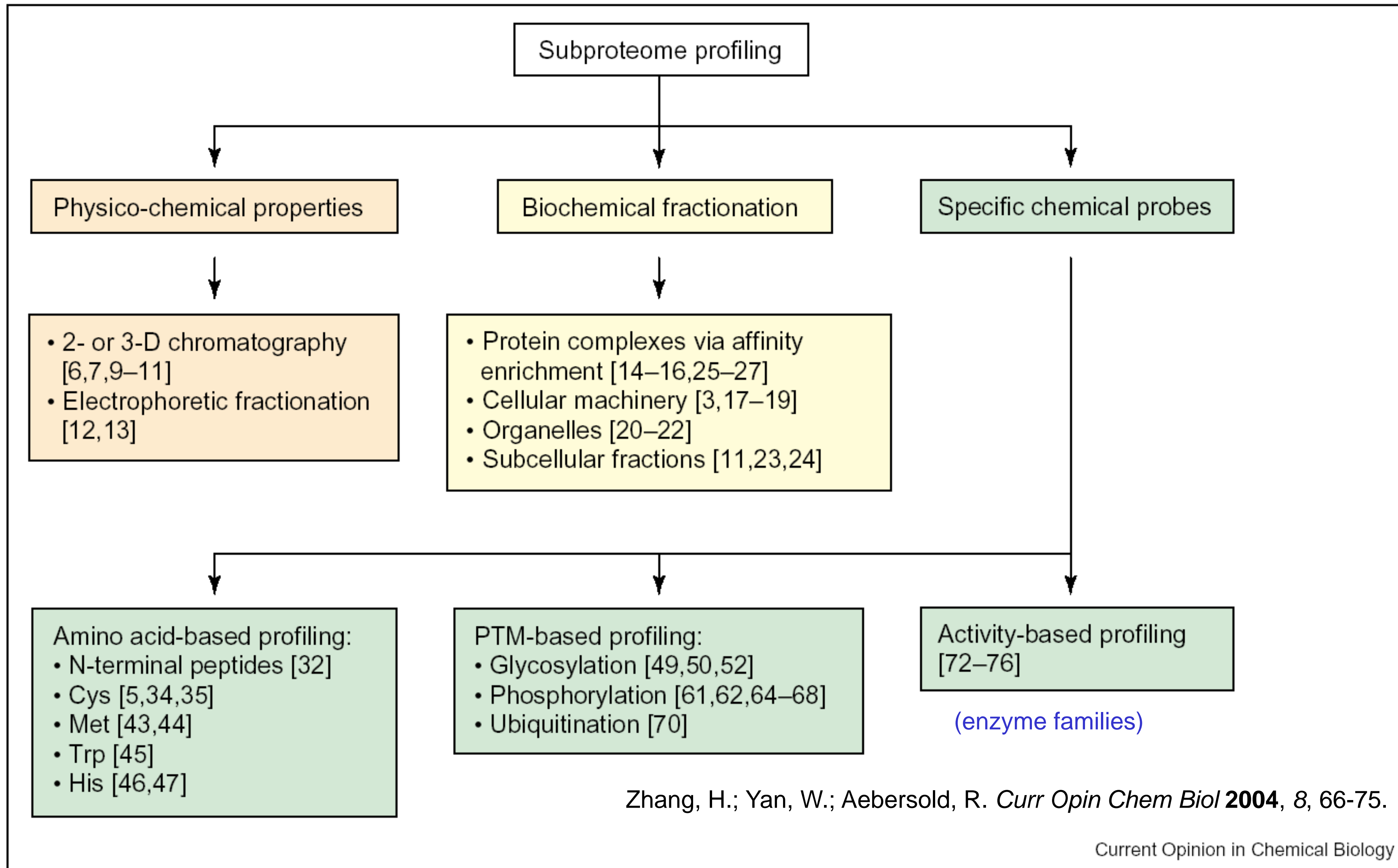
TABLE I

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.

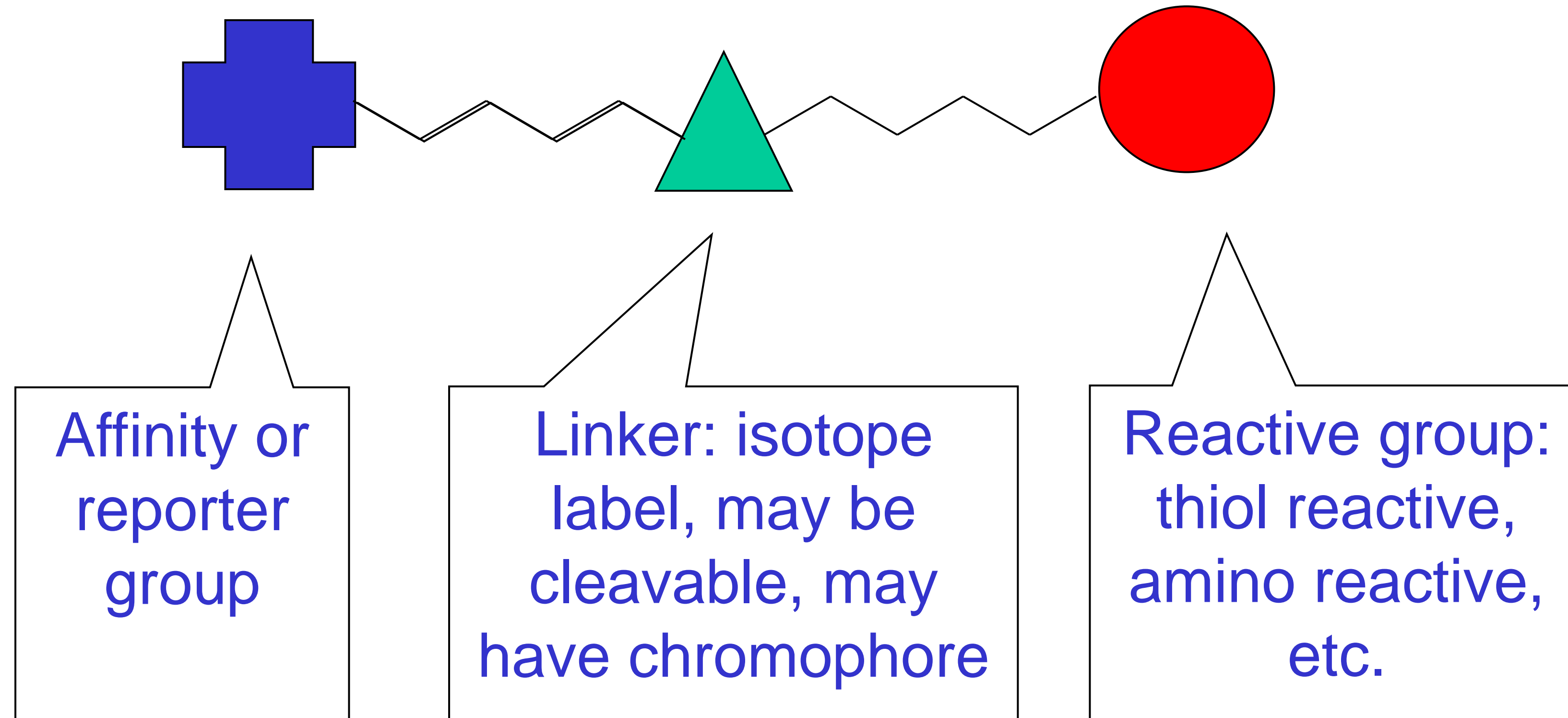
Separation strategies for proteome profiling



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

Strategies for subproteome profiling, with key references.

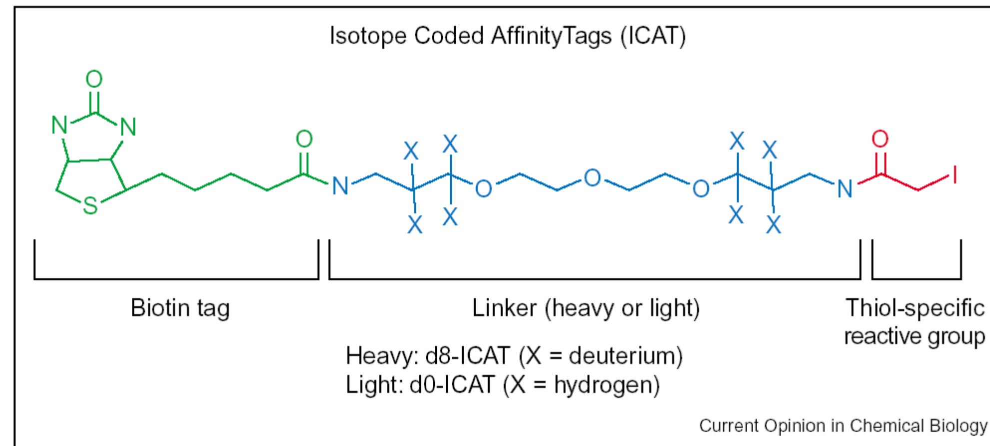
The Isotope Coding Reagents



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

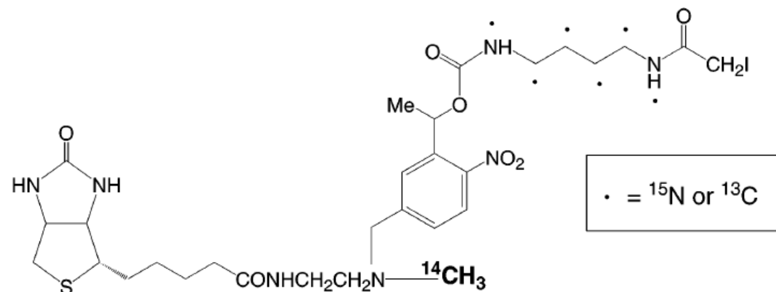
ICAT Reagents

Traditional

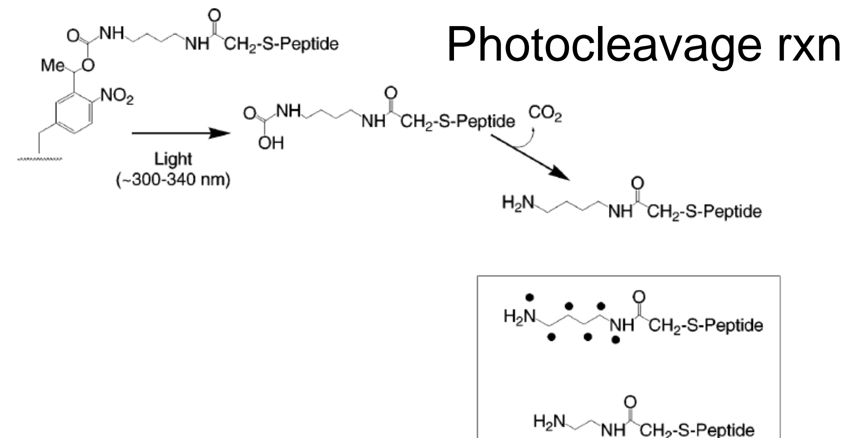


The structure of ICAT reagents, which comprise a cysteine-reactive group (red), a linker containing either heavy or light isotopes (blue) and a binding affinity tag (green).

VICAT (visible ICAT)



Photocleavage rxn



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

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

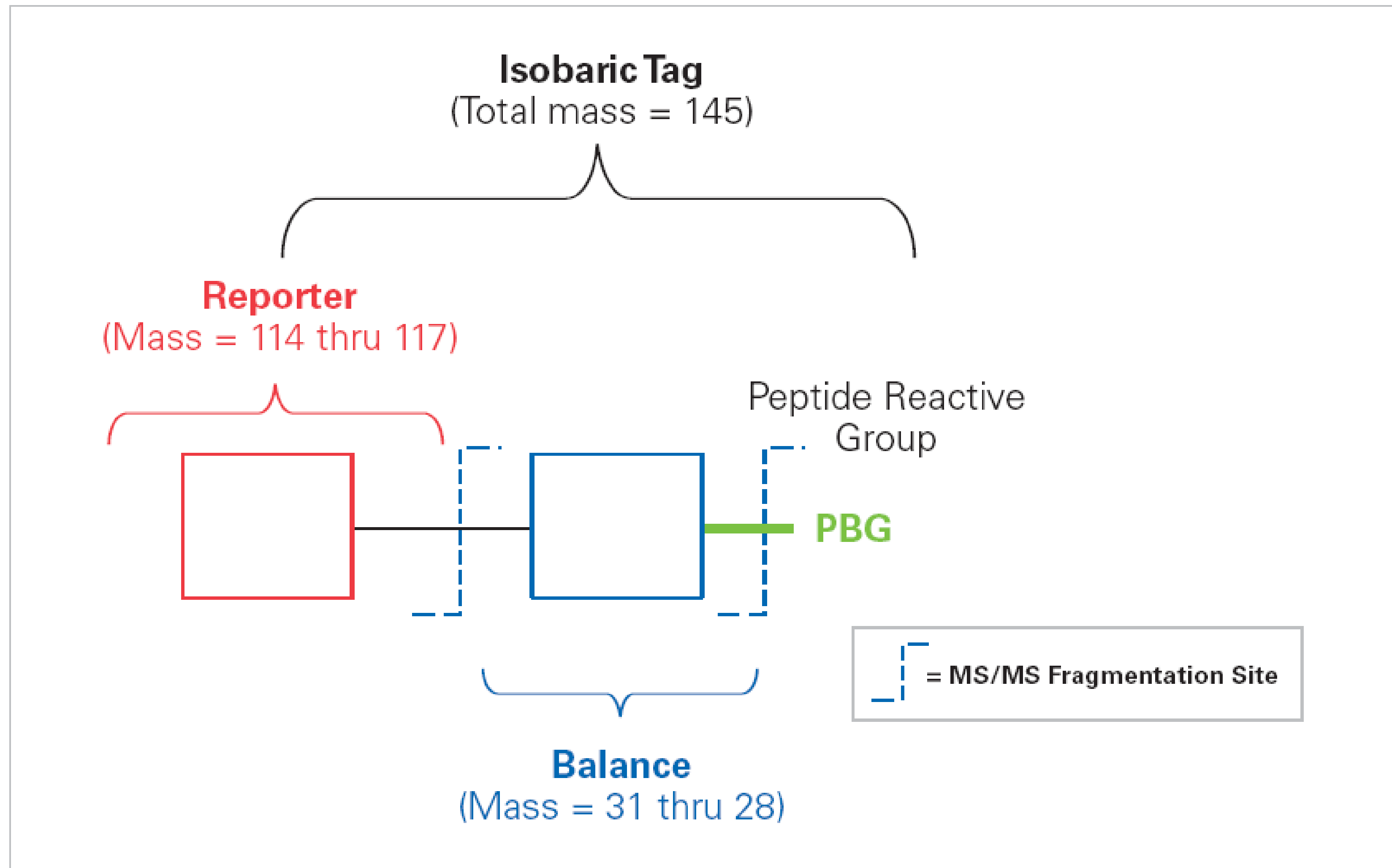
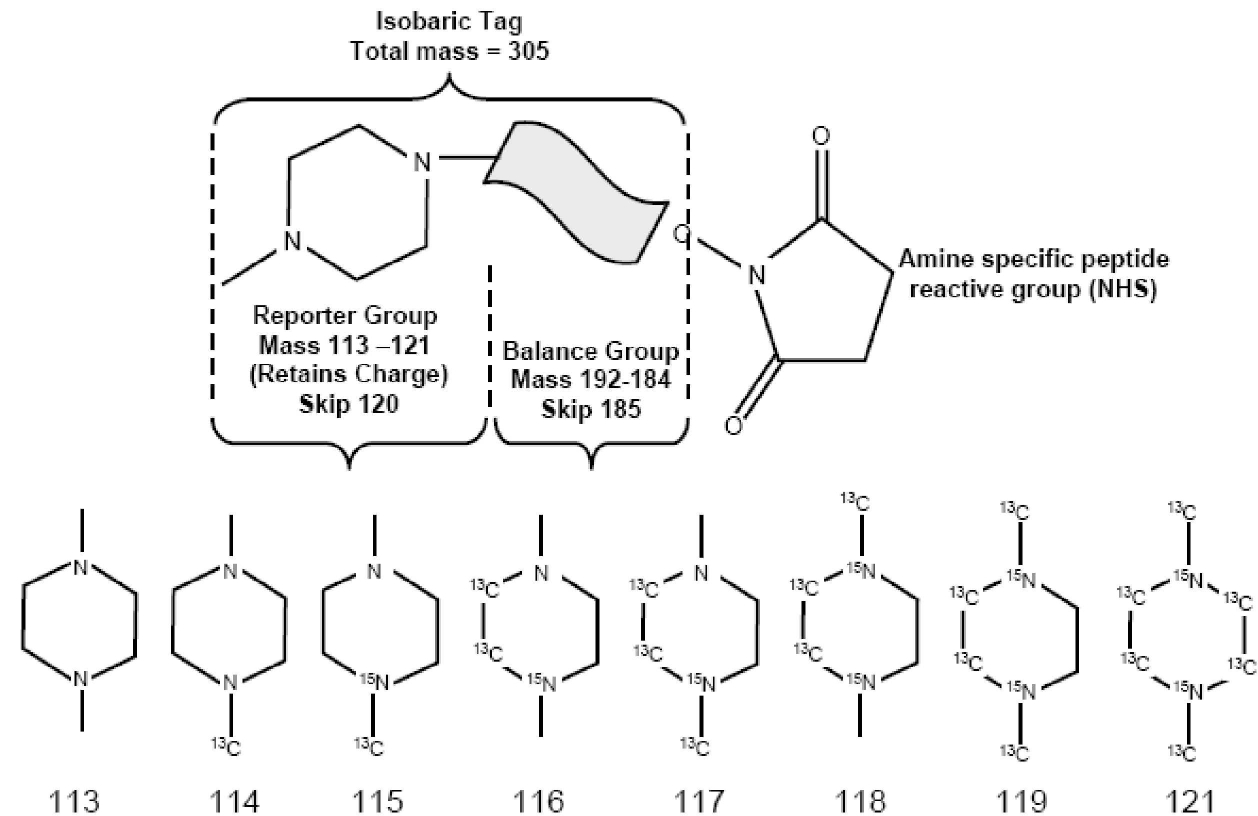


Figure 1. iTRAQ™ reagent structure

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.

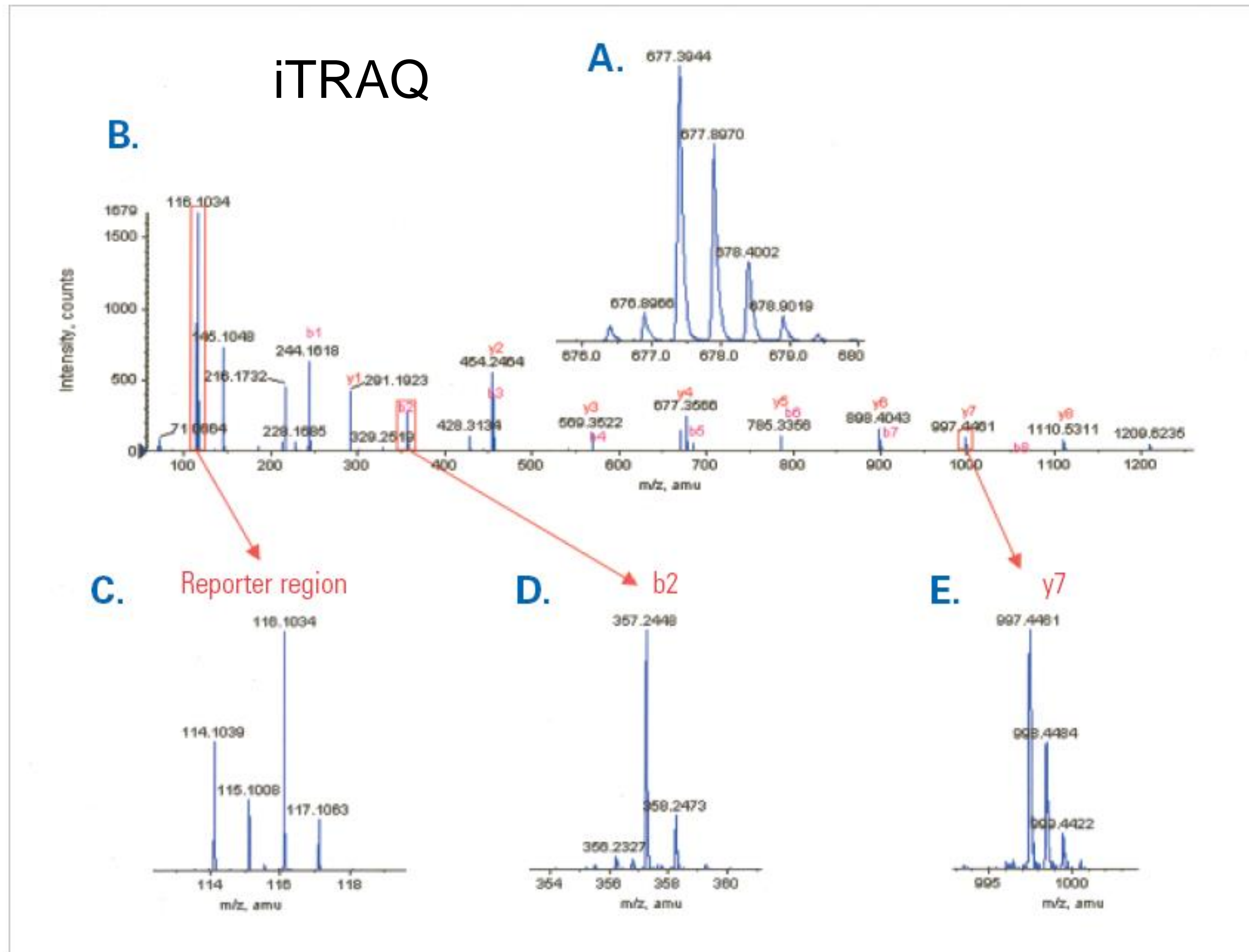
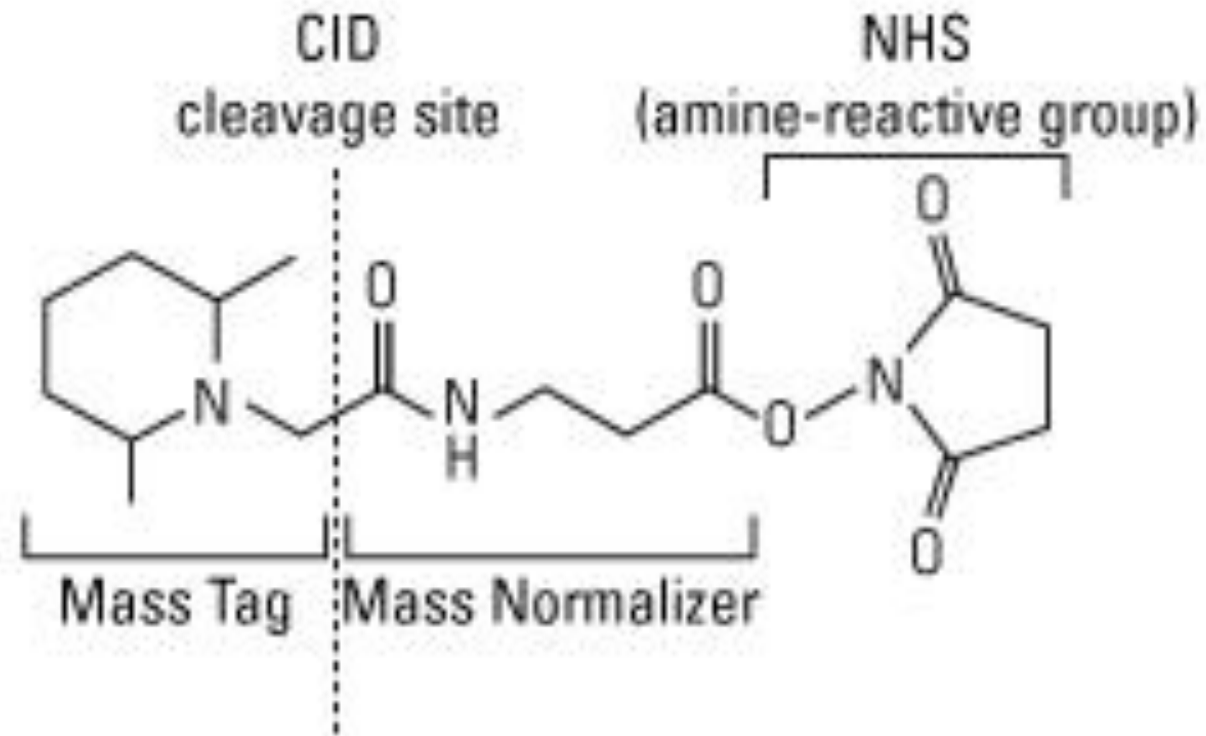


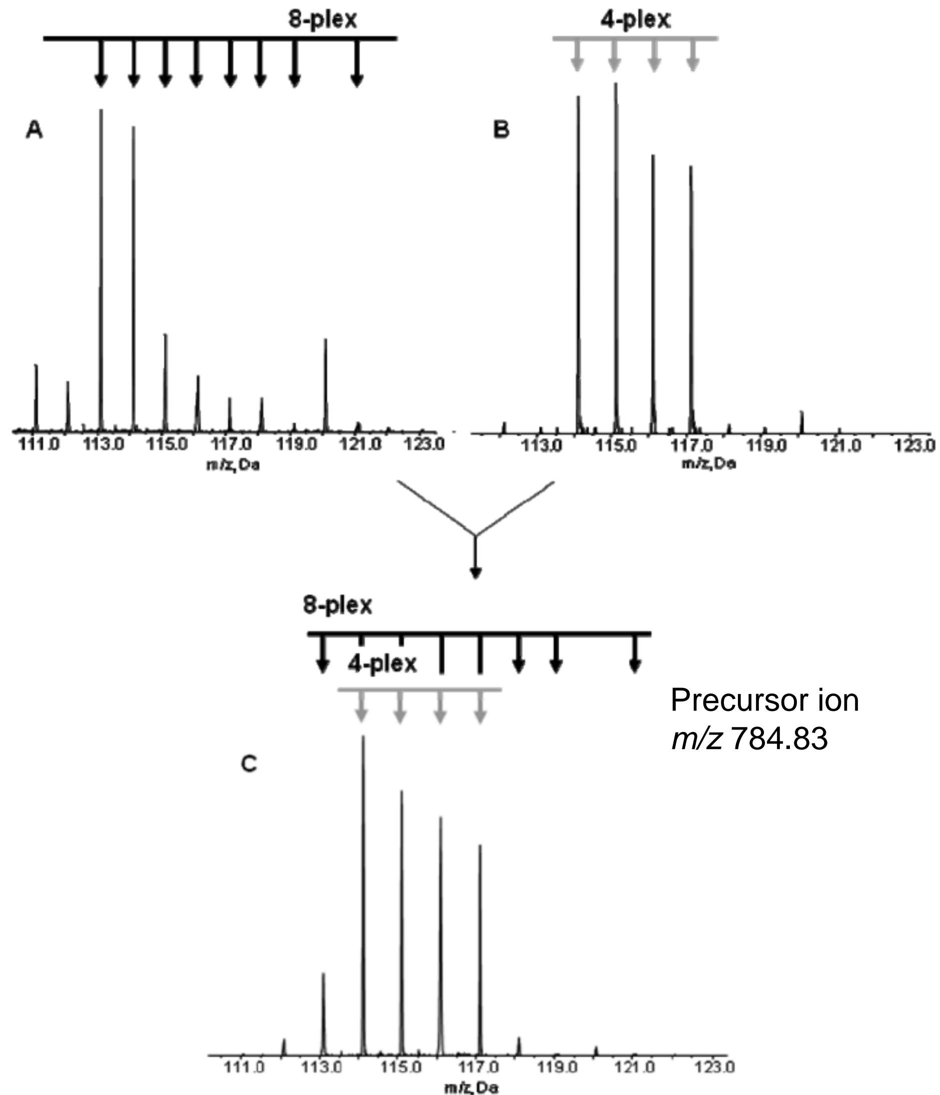
Figure 2. MS and MS/MS spectra from a multiplex sample labeled with 4 iTRAQ™ reagents showing **A.** doubly charged parent ion **B.** MS/MS spectrum corresponding to VLVDTDYK **C.** 4 diagnostic reporter ions and **D.** and **E.** peptide fragment ions

Tandem mass tags (TMT, Thermo-Fisher)



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

4-plex cyanobacterium proteome used as a complex background

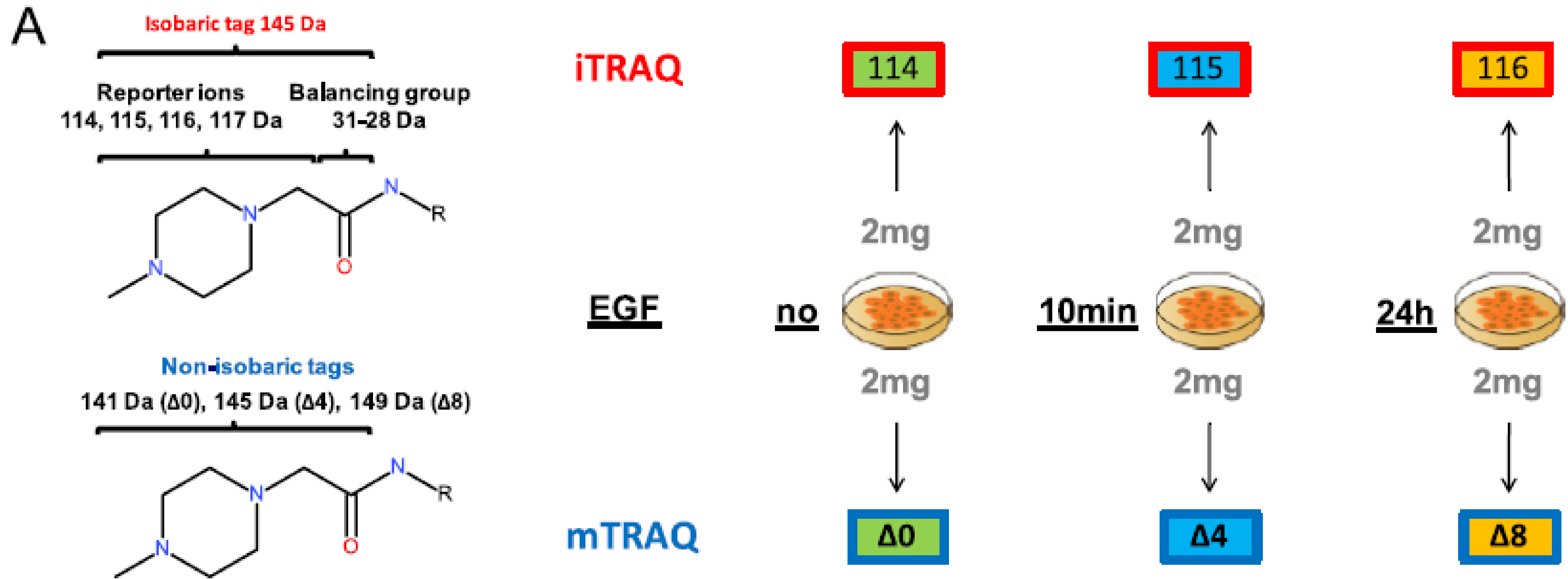


Precursor ion
m/z 784.83

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.

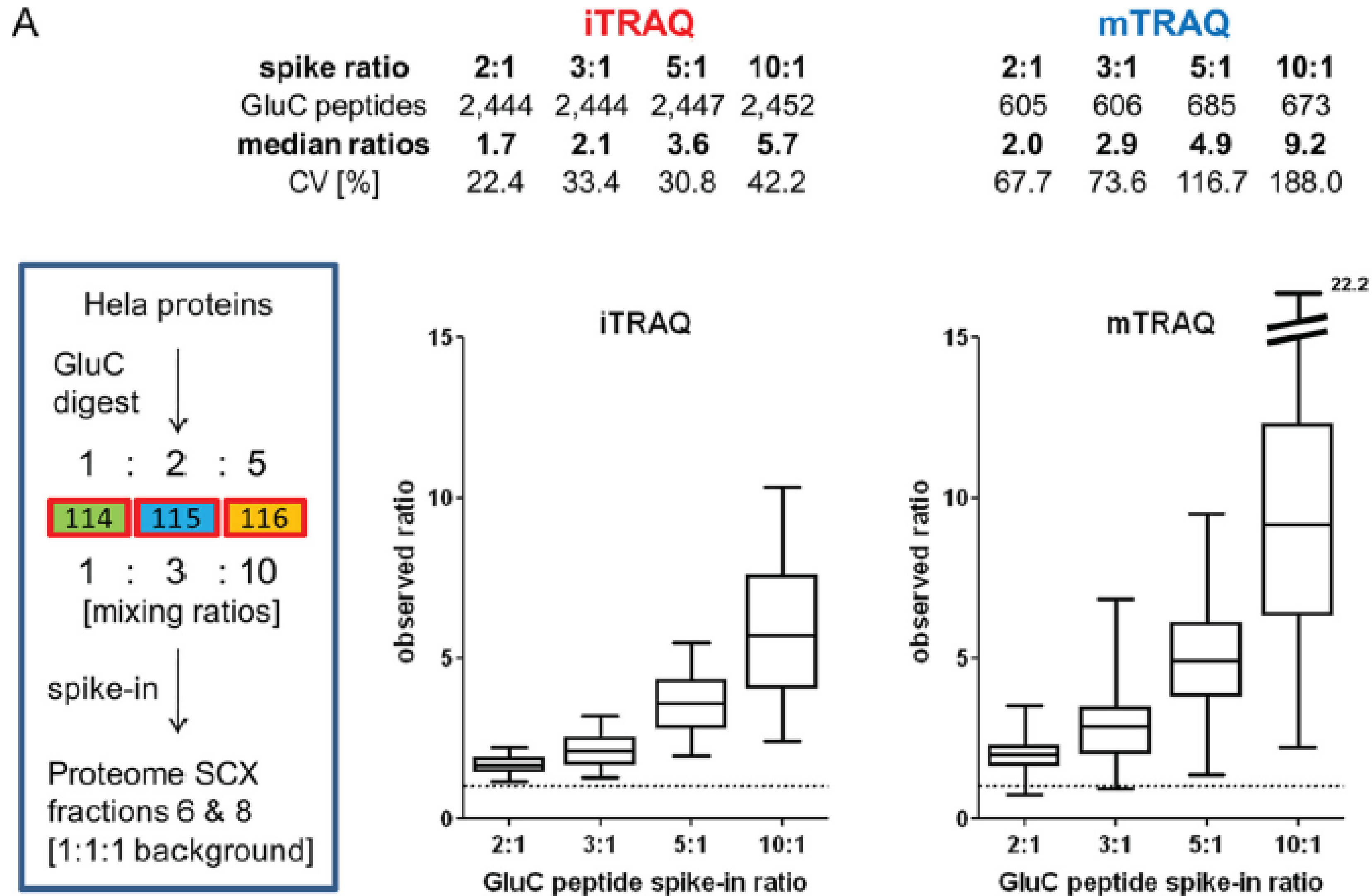
Ow, S.Y., et al., Journal of proteome research, 2009. **8**(11): p. 5347-55.

Comparison of quantitative performance of iTRAQ vs mTRAQ



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

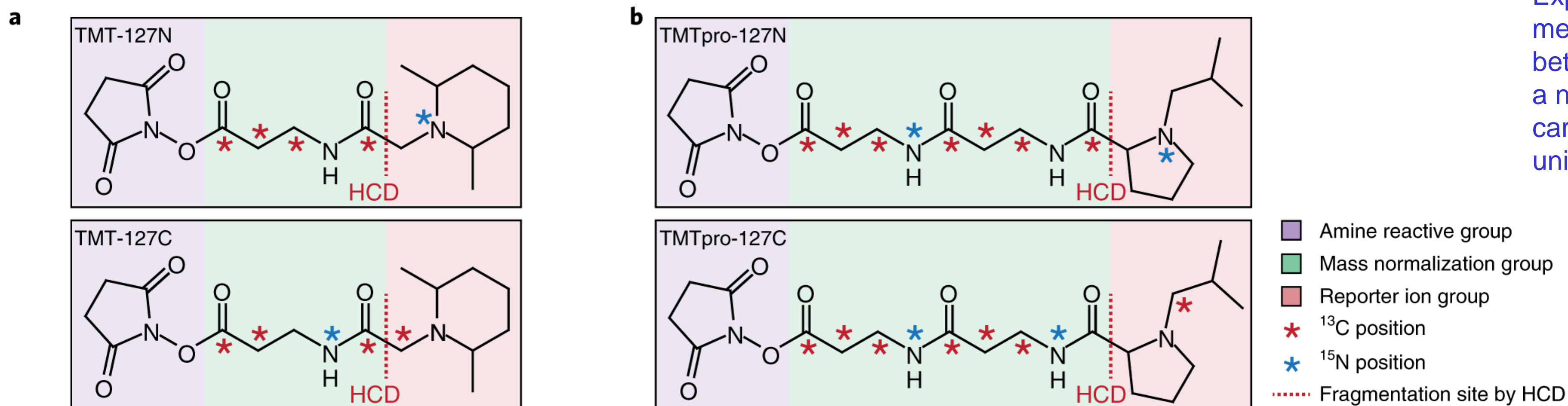
iTRAQ vs mTRAQ: FIG. 6. iTRAQ quantification is more precise but less accurate than mTRAQ. A, GluC peptide spike-in experiments to test accuracy and variability of quantification.



	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.

TMTpro reagents: a set of isobaric labeling mass tags enables simultaneous proteome-wide measurements across 16 samples

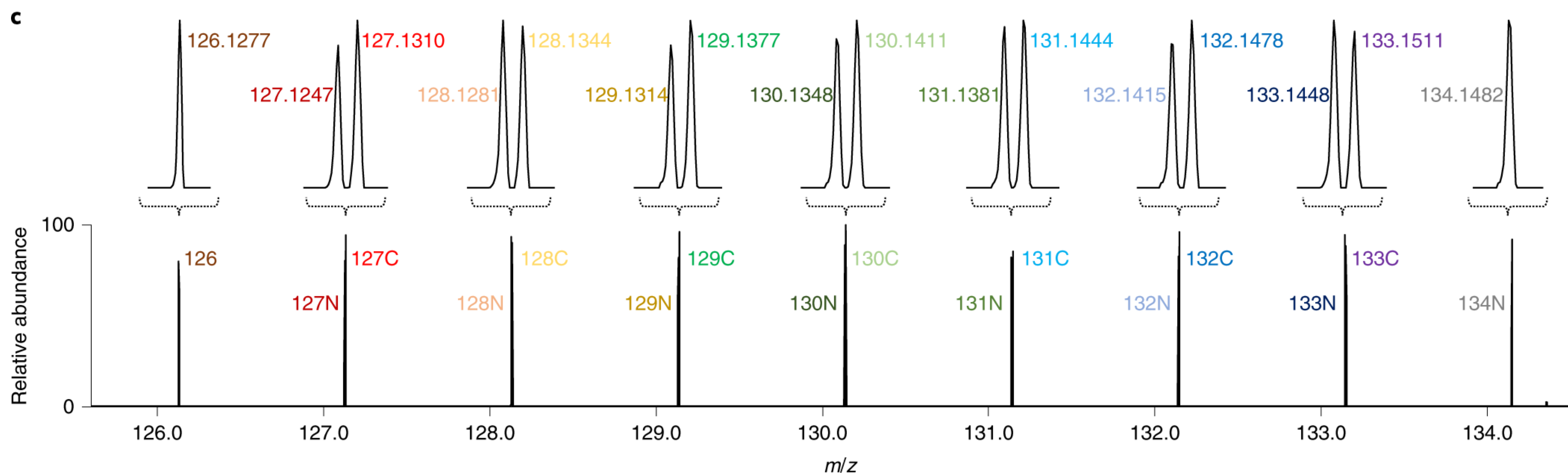


Exploiting the small, but measurable, mass difference between an extra neutron in a nitrogen-15 versus a carbon-13 atom (6 millimass units)

TMT11plex

TMTpro16plex

9 tagged positions ≤ 18 reagents



Li, J., et al., Nat Methods, 2020. **17**(4): p. 399-404.

Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry

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

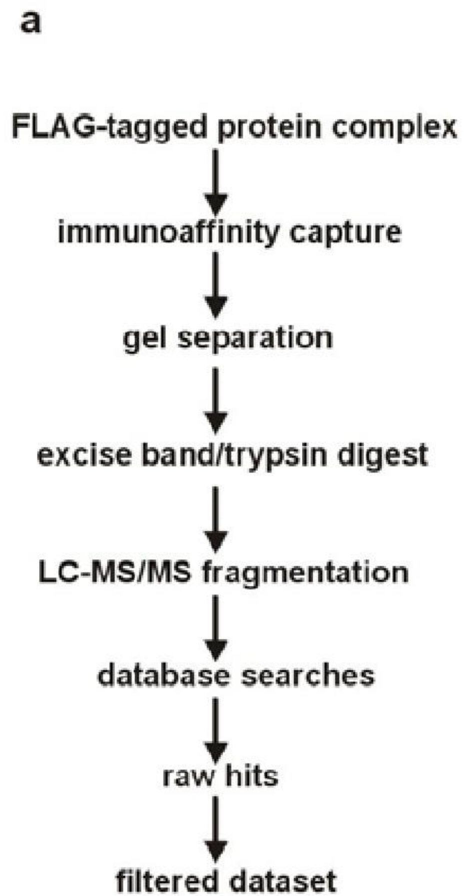
- Ho, Yuen*; Gruhler, Albrecht*; Heilbut, Adrian*; Bader, Gary D.†‡; Moore, Lynda*; Adams, Sally-Lin*; Millar, Anna*; Taylor, Paul*; Bennett, Keiryn*; Boutilier, Kelly*; Yang, Lingyun*; Wolting, Cheryl*; Donaldson, Ian*; Schandorff, Søren*; Shewnarane, Juanita*; Vo, Mai*†; Taggart, Joanne*†; Goudreault, 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

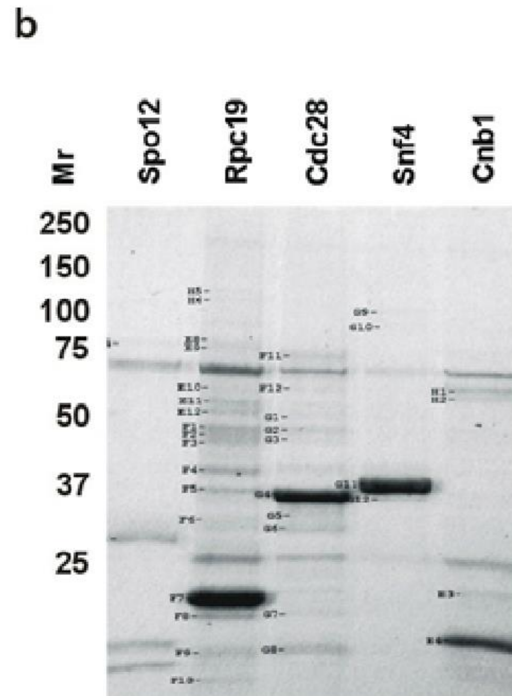
•Cited 4234 times as of 3/30/21, Google Scholar

Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry

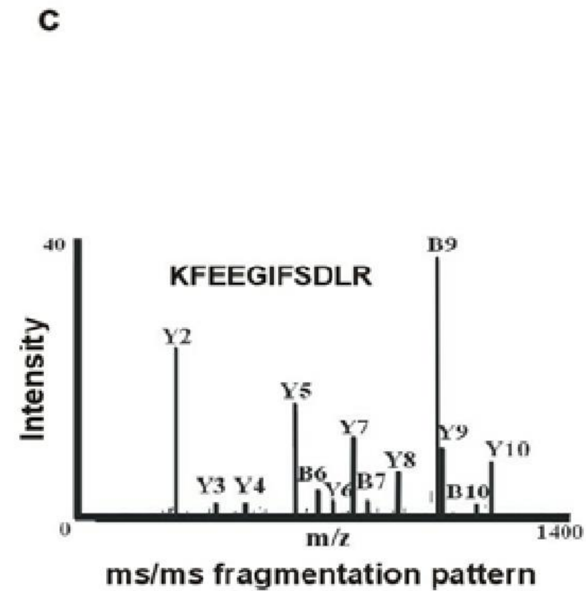
a



b



c



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

Connected complexes

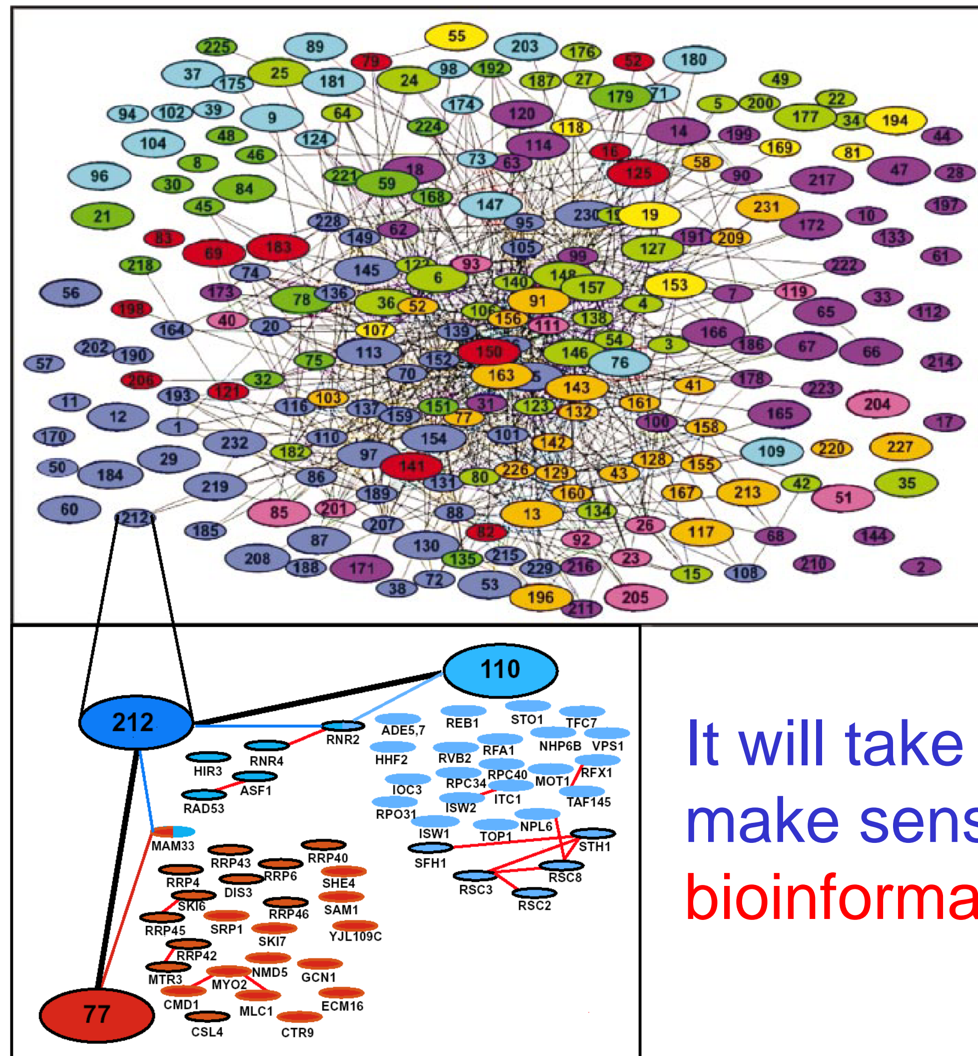
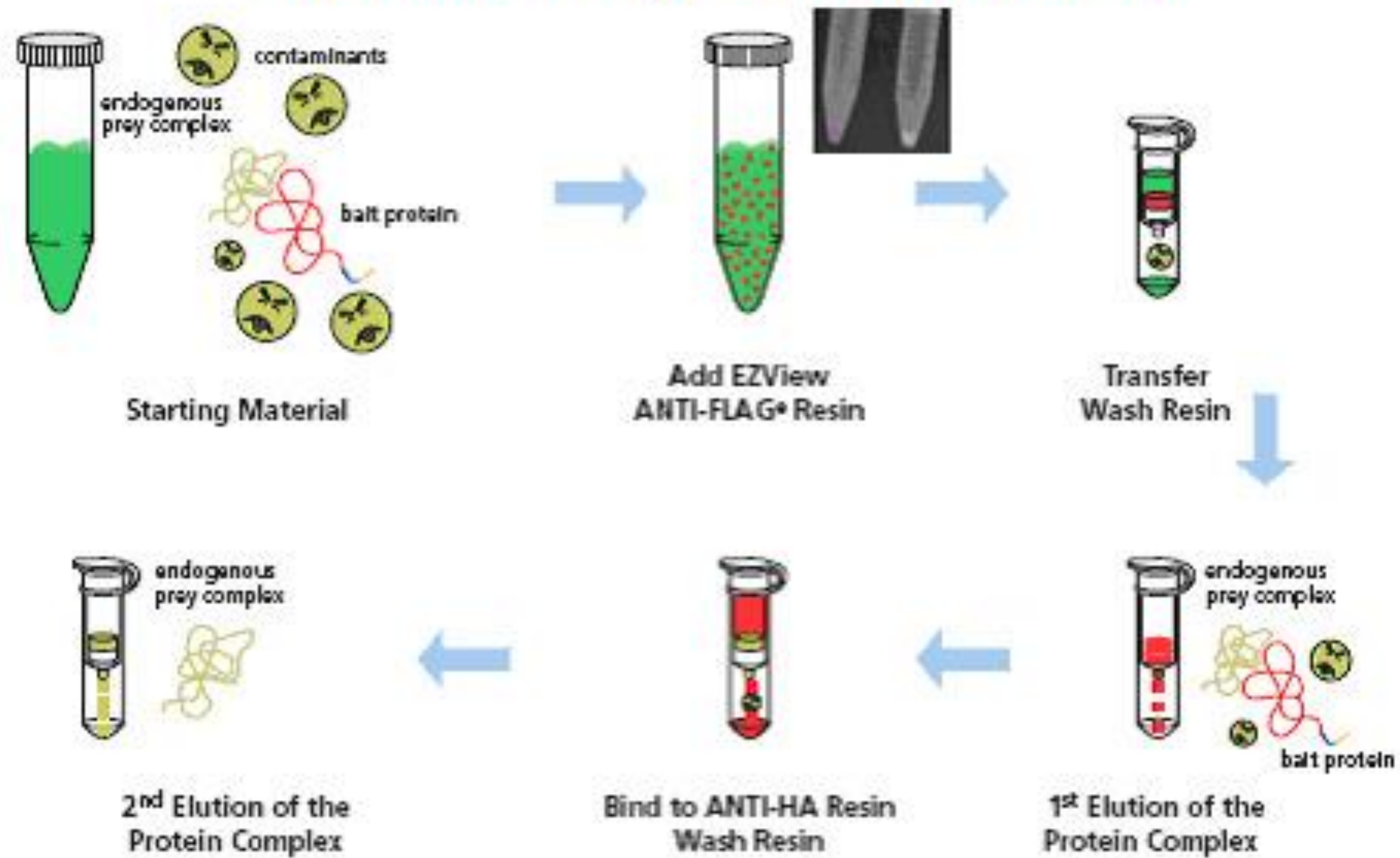


Figure 4 The protein complex network, and grouping of connected complexes. Links were established between complexes sharing at least one protein. For clarity, proteins found in more than nine complexes were omitted. The graphs were generated automatically by a relaxation algorithm that finds a local 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.

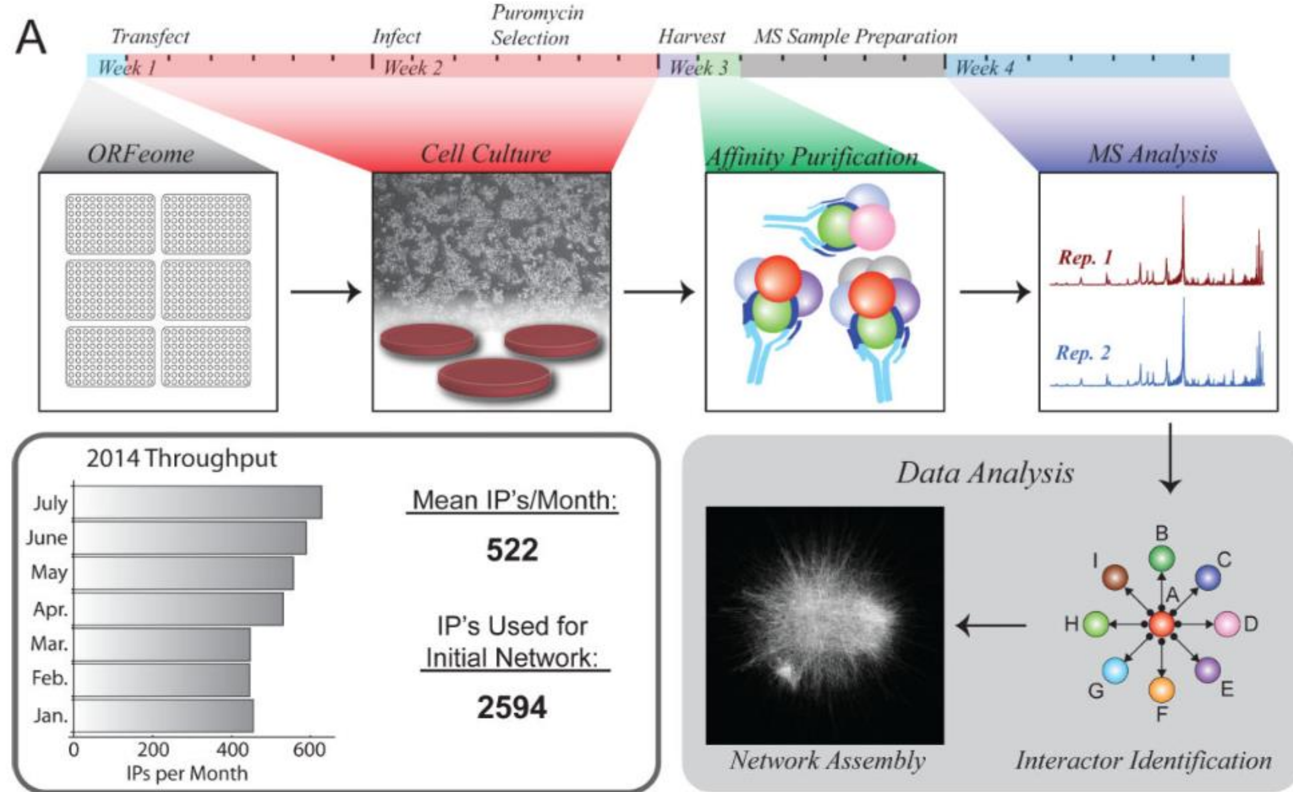
FLAG HA
Tandem Affinity Purification (TAP)



https://www.sigmaaldrich.com/life-science/proteomics/recombinant-protein-expression/purification-detection/flag-system/flag-reg-ha-system.html

Cell

The BioPlex Network: A Systematic Exploration of the Human Interactome

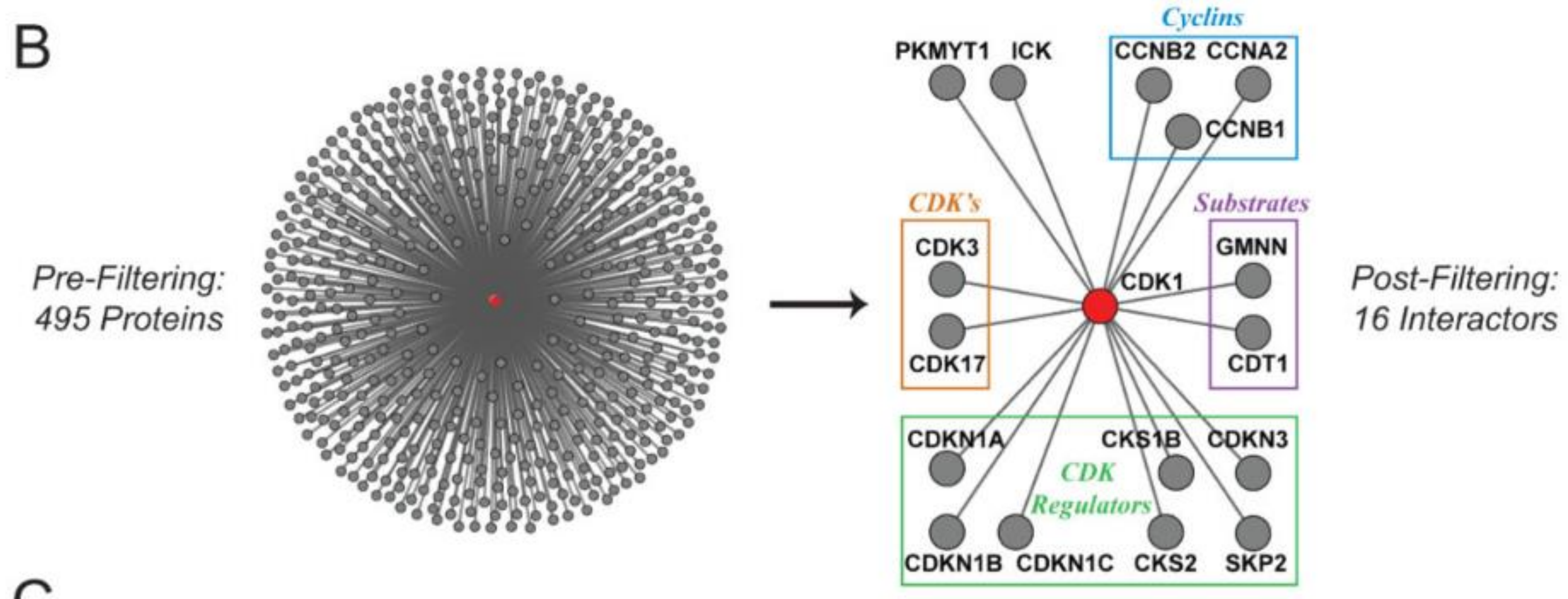


(A) AP-MS platform: 1) A lentiviral library of 13,000 FLAG-HA-tagged ORFs was constructed from the Human ORFEOME; 2) 293T cells were infected and expanded under puromycin selection; 3) Baits and preys were immuno-purified; 4) tryptic digests were analyzed in technical duplicate by LC-MS; 5) Proteins were identified and specific interactors found; 6) Interactions were assembled to model the human interactome. Up to 600 AP-MS experiments may be completed per month.

Identify interacting partners for 2,594 human proteins in HEK293T cells. The resulting network (BioPlex) contains 23,744 interactions among 7,668 proteins with 86% previously undocumented.

Huttlin, E.L., et al., Cell, 2015. **162**(2): p. 425-440.

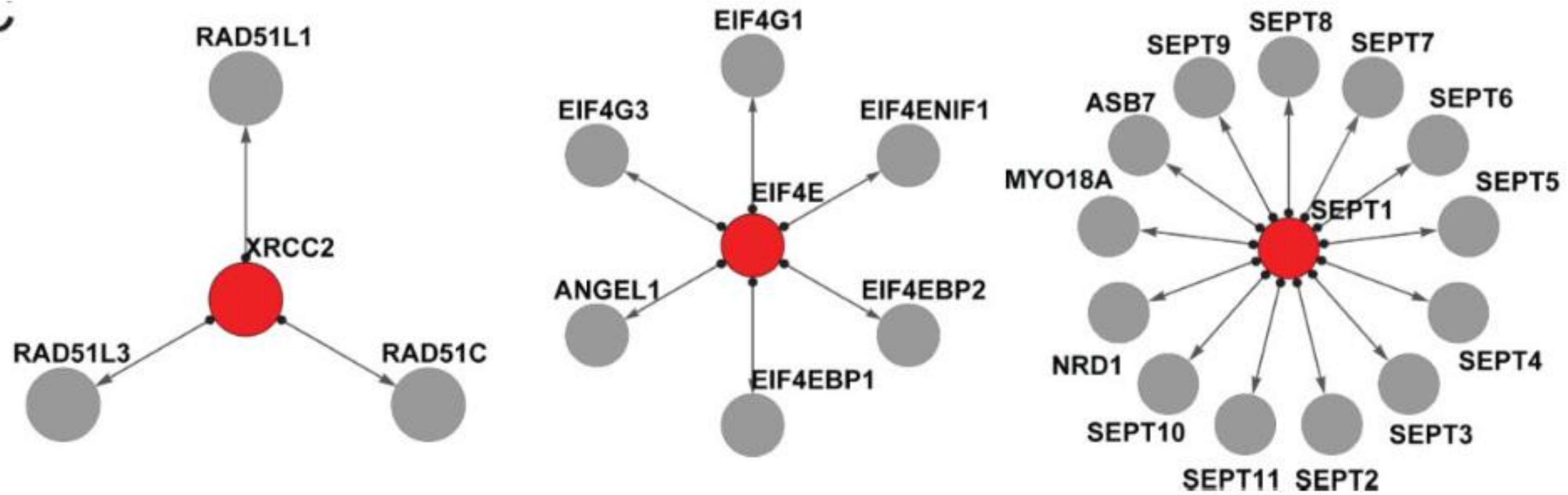
"The proteome can be viewed as constellations of interacting protein modules organized into signal transduction networks, molecular machines, and organelles."



(B) *CompPASS-Plus* extracts 16 interactors for bait CDK1 from a background of nearly 500 proteins.

Our work employs *CompPASS, which has identified high-confidence interacting proteins* (HCIP's) for up to ~100 baits (Sowa et al., 2009). *CompPASS quantifies enrichment of each* protein in each IP, relative to other unrelated AP-MS datasets, based on abundance, detection frequency, and reproducibility.

Huttlin, E.L., et al., Cell, 2015. **162**(2): p. 425-440.



(C) Interaction maps for baits XRCC2, EIF4E, and SEPT1 (red). Nearly all interactions have been previously described. Interactors were identified from backgrounds of 487, 778, and 749 proteins, respectively.

Huttlin, E.L., et al., Cell, 2015. **162**(2): p. 425-440.

Published online 1 November 2009 Nucleic Acids Research, 2010, Vol. 38, Database issue D497–D501
doi:10.1093/nar/gkp914

CORUM: the comprehensive resource of mammalian protein complexes—2009

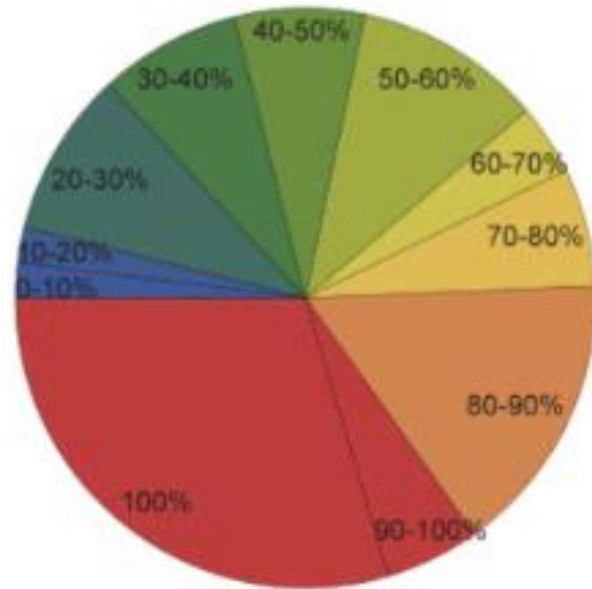
Andreas Ruepp[1,*], Brigitte Waegele[1,2], Martin Lechner[1], Barbara Brauner[1], Irmtraud Dunger-Kaltenbach[1], Gisela Fobo[1], Goar Frishman[1], Corinna Montrone[1] and H.-Werner Mewes[1,2]

[1]Institute for Bioinformatics and Systems Biology (IBIS), Helmholtz Zentrum München—German Research Center for Environmental Health (GmbH), Ingolstädter Landstraße 1, D-85764 Neuherberg and [2]Technische Universität München, Chair of Genome Oriented Bioinformatics, Center of Life and Food Science, D-85350 Freising-Weihenstephan, Germany

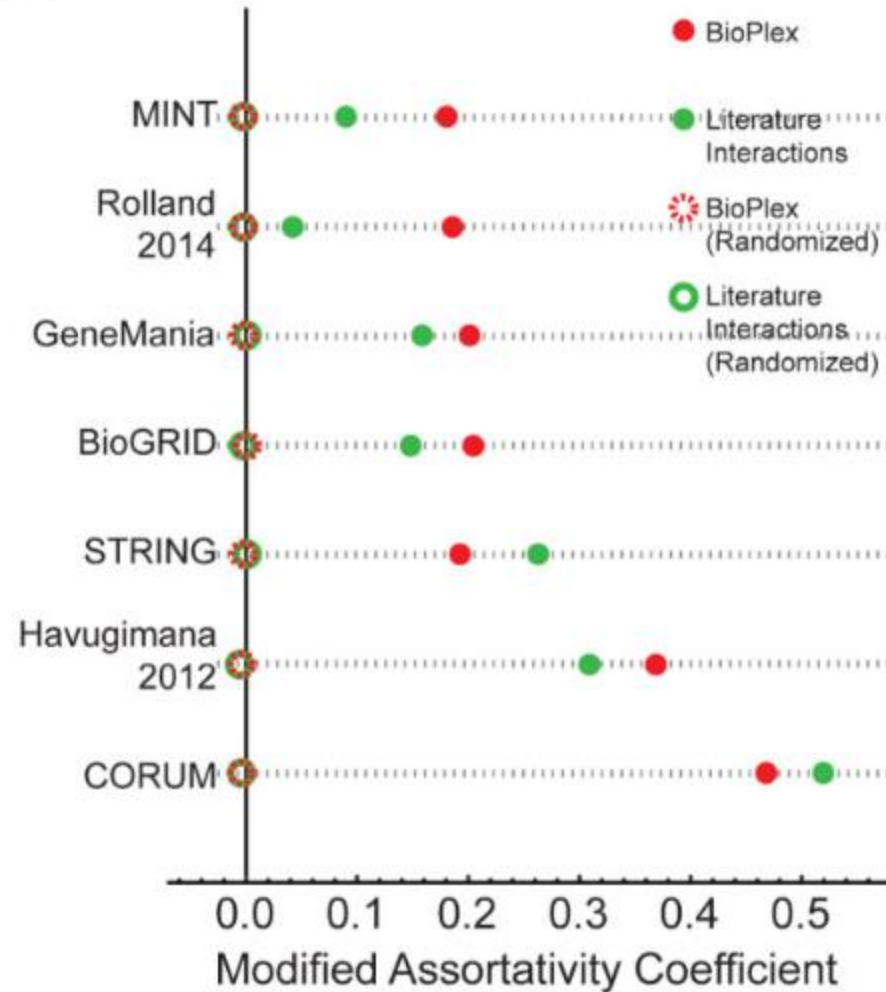
ABSTRACT

CORUM is a database that provides a manually curated repository of experimentally characterized protein complexes from mammalian organisms, mainly human (64%), mouse (16%) and rat (12%). Protein complexes are key molecular entities that integrate multiple gene products to perform cellular functions. The new CORUM 2.0 release encompasses 2837 protein complexes offering the largest and most comprehensive publicly available dataset of mammalian protein complexes. The CORUM dataset is built from 3198 different genes, representing ~16% of the protein coding genes in humans. Each protein complex is described by a protein complex name, subunit composition, function as well as the literature reference that characterizes the respective protein complex. Recent developments include mapping of functional annotation to Gene Ontology terms as well as cross-references to Entrez Gene identifiers. In addition, a 'Phylogenetic Conservation' analysis tool was implemented that analyses the potential occurrence of orthologous protein complex subunits in mammals and other selected groups of organisms. This allows one to predict the occurrence of protein complexes in different phylogenetic groups. CORUM is freely accessible at (http://mips.helmholtz-muenchen.de/genre/proj/corum/index.html).

BioPlex Recapitulates Known Complexes and Reveals Thousands of New Interactions



(A) AP-MS interactions superimposed onto CORUM complexes. The pie chart depicts the fraction of complexes achieving the indicated coverage in BioPlex



(E) Pairwise comparisons of BioPlex with published interaction networks were performed, using graph assortativity to quantify preferential interaction in cases of shared localization among proteins detected in both networks. Literature datasets included BioGRID, CORUM,

Huttlin, E.L., et al., Cell, 2015. **162**(2): p. 425-440.

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.

PTM proteomics

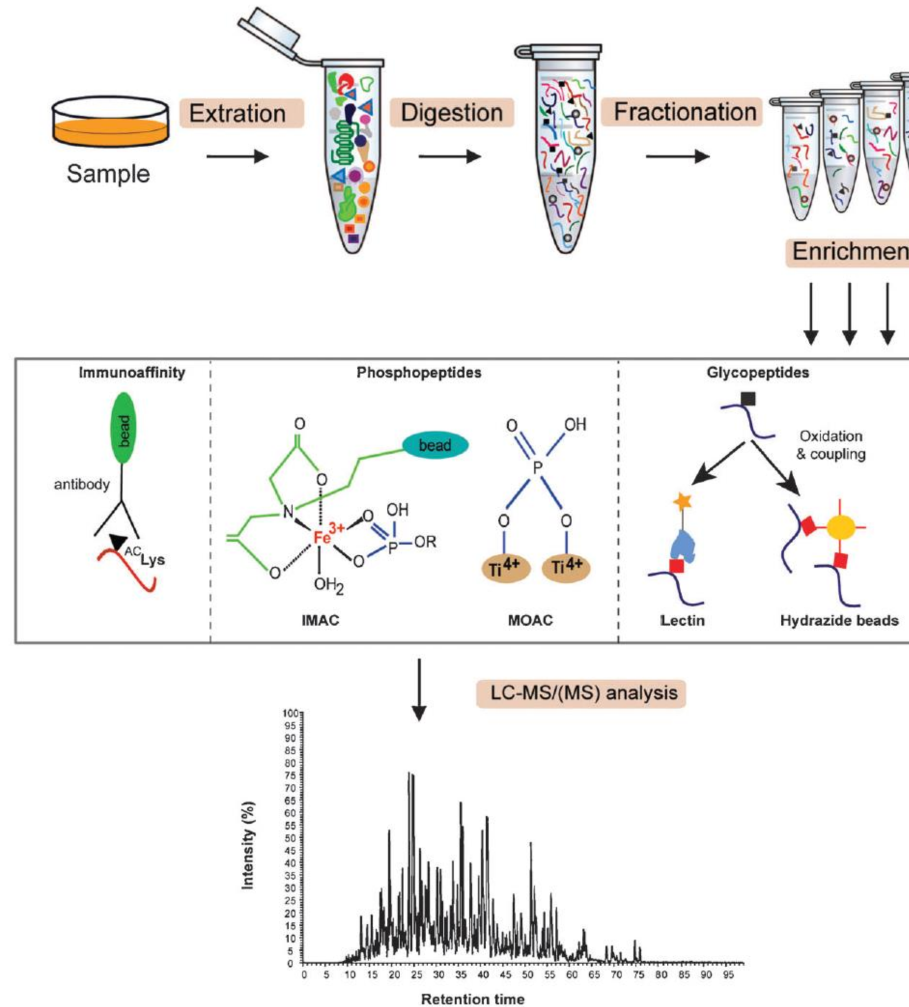


Fig. 4 Experimental strategies to study protein post-translational modification. Enrichment of lysine acetylated (▲), phosphorylated (●) and glycosylated peptides (■) using different affinity media is shown as examples. A typical workflow for PTM analysis involves extraction of proteins from cells or 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.

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

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

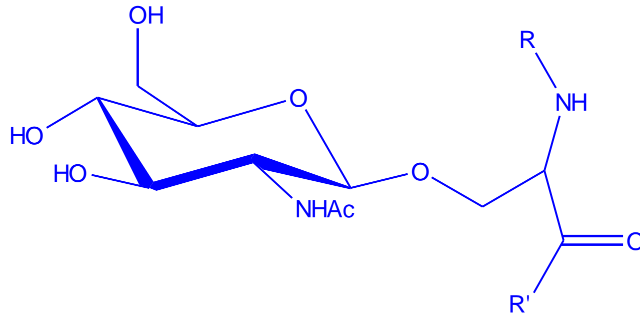
ABTCDSEFGHTIJKLMNOP

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

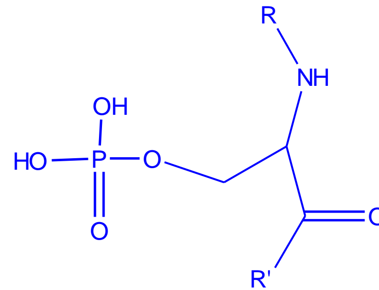
Common PTM-amino acid linkages

Intracellular | Cell surface, extracellular

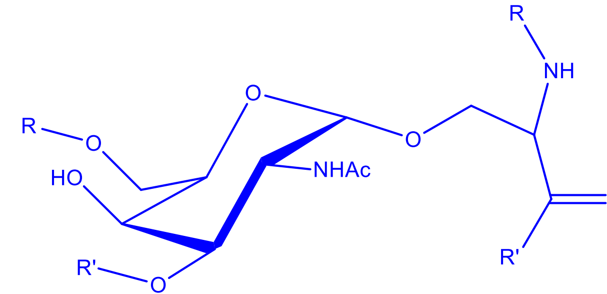
Eliminative release



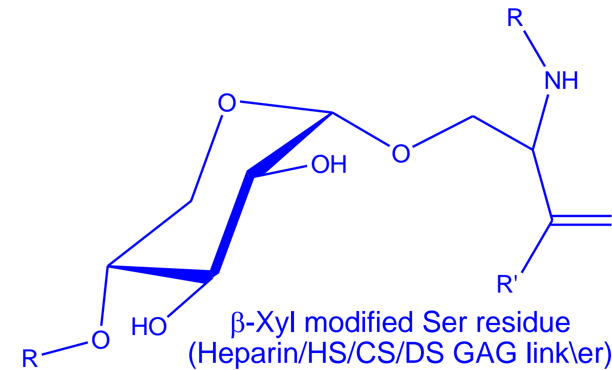
β-GlcNAc-modified Ser/Thr residue
(*O*-GlcNAc)



Phosphorylated Ser/Thr residue

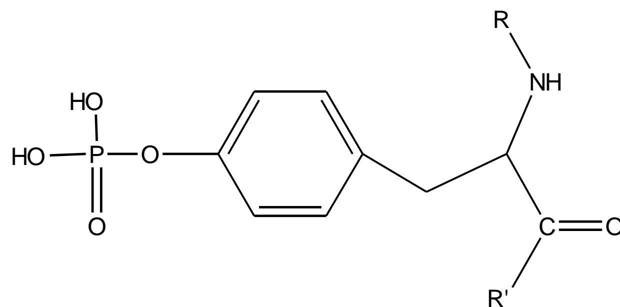


α-GalNAc-modified Ser/Thr residue
(mucin type *O*-linked glycans)

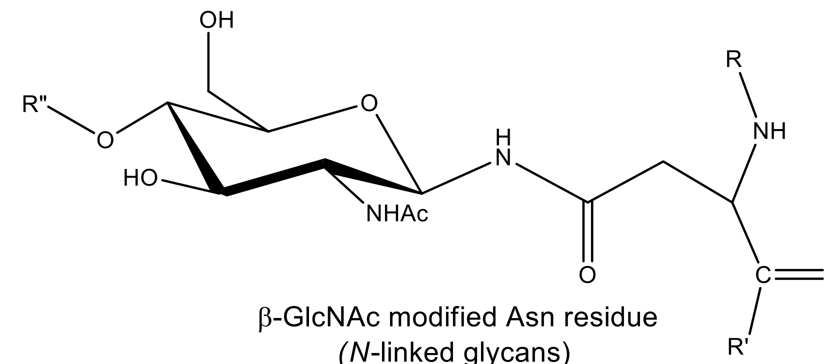


β-Xyl modified Ser residue
(Heparin/HS/CS/DS GAG link\er)

Eliminative release

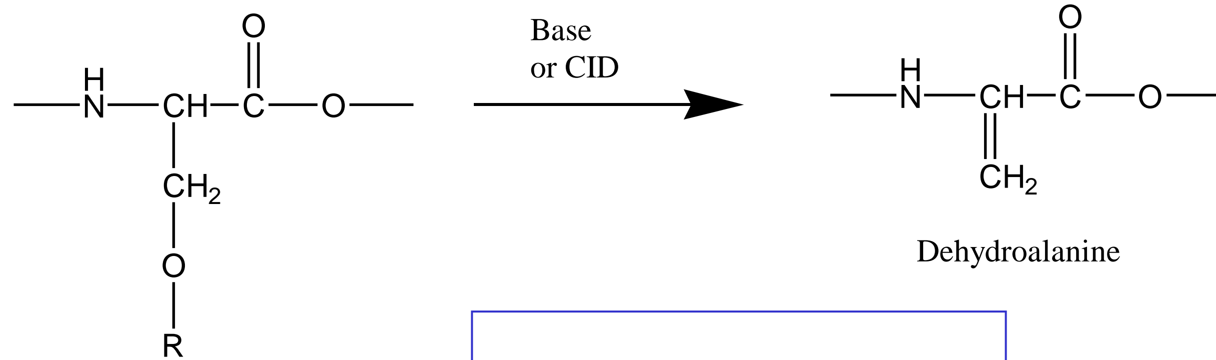


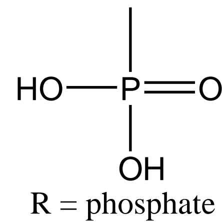
Phosphorylated Tyr residue
(Sulfated Tyrosine)



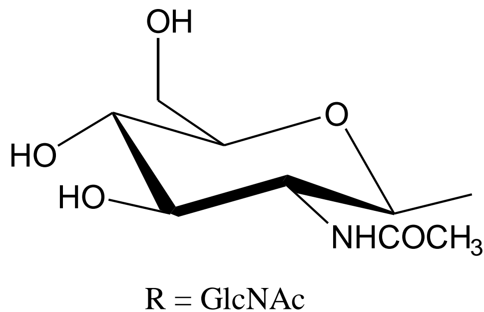
β-GlcNAc modified Asn residue
(*N*-linked glycans)

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



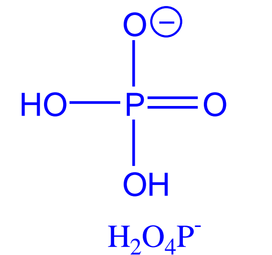


R = phosphate



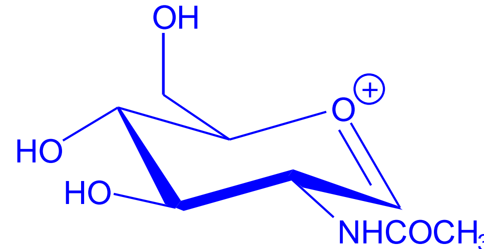
R = GlcNAc

Low *m/z* ions



$H_2O_4P^-$

Exact Mass: 96.97



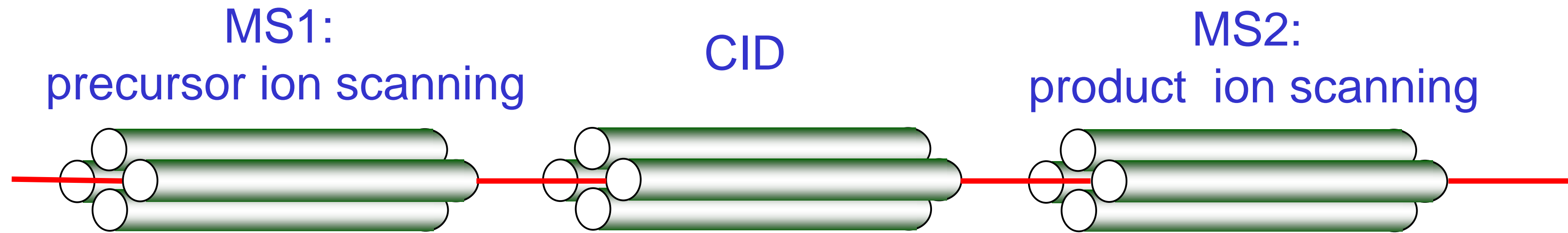
$C_8H_{14}NO_5^+$

Exact Mass: 204.09

Conditions:

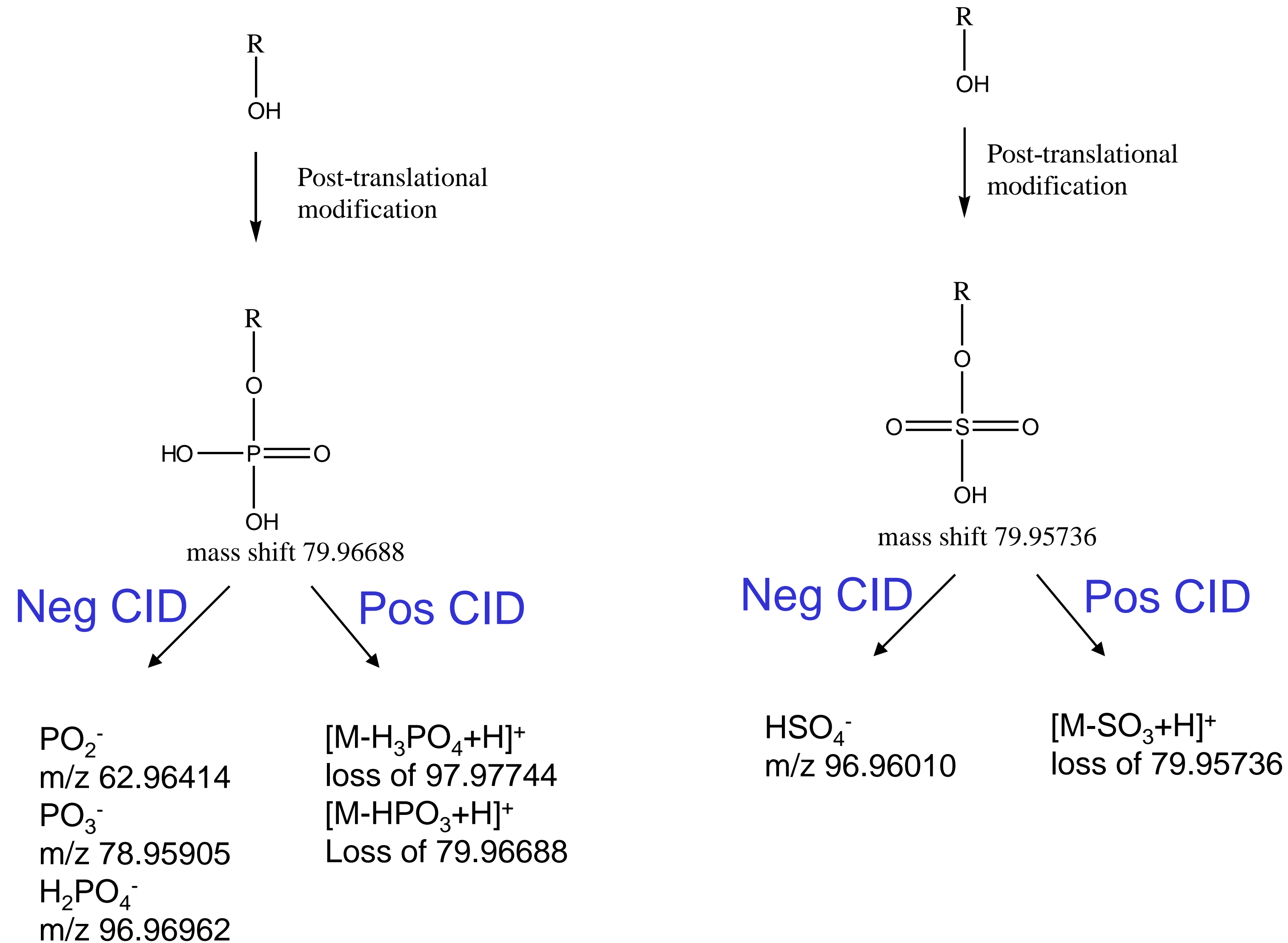
- *O*-GlcNAc 1% TEA, 0.1% NaOH, 2 hr. 50ºC
- *O*-GalNAc, 1M NaOH, 2M $NaBH_4$, o/n 20ºC
- *O*-phosphate, 4M LiOH, 1% ethanedithiol, 37ºC, 1 hr.

Precursor ion scans and neutral loss scans



	MS1	CID	MS2
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

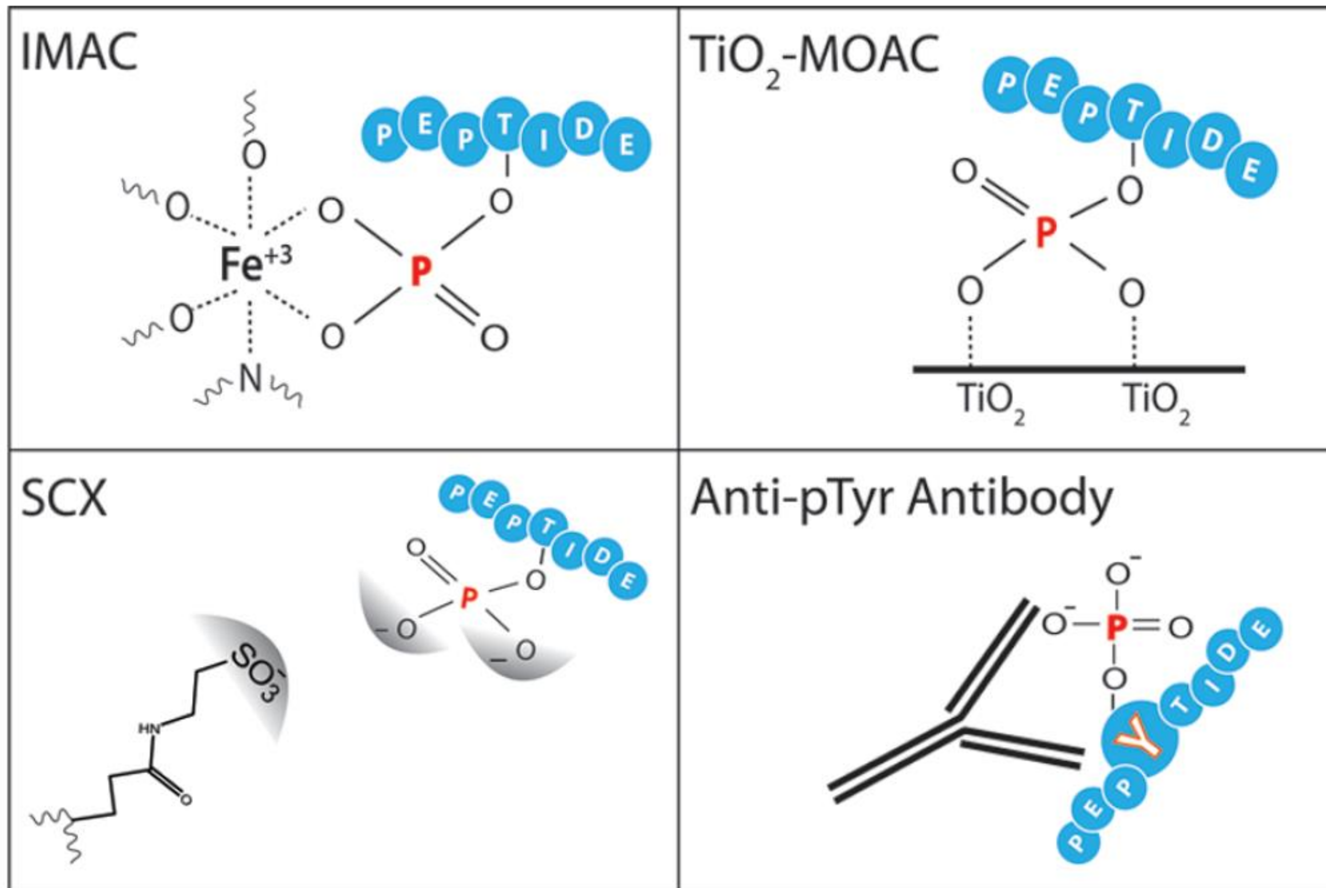
Mass shifts for phosphorylation and sulfation



The Phosphoproteome

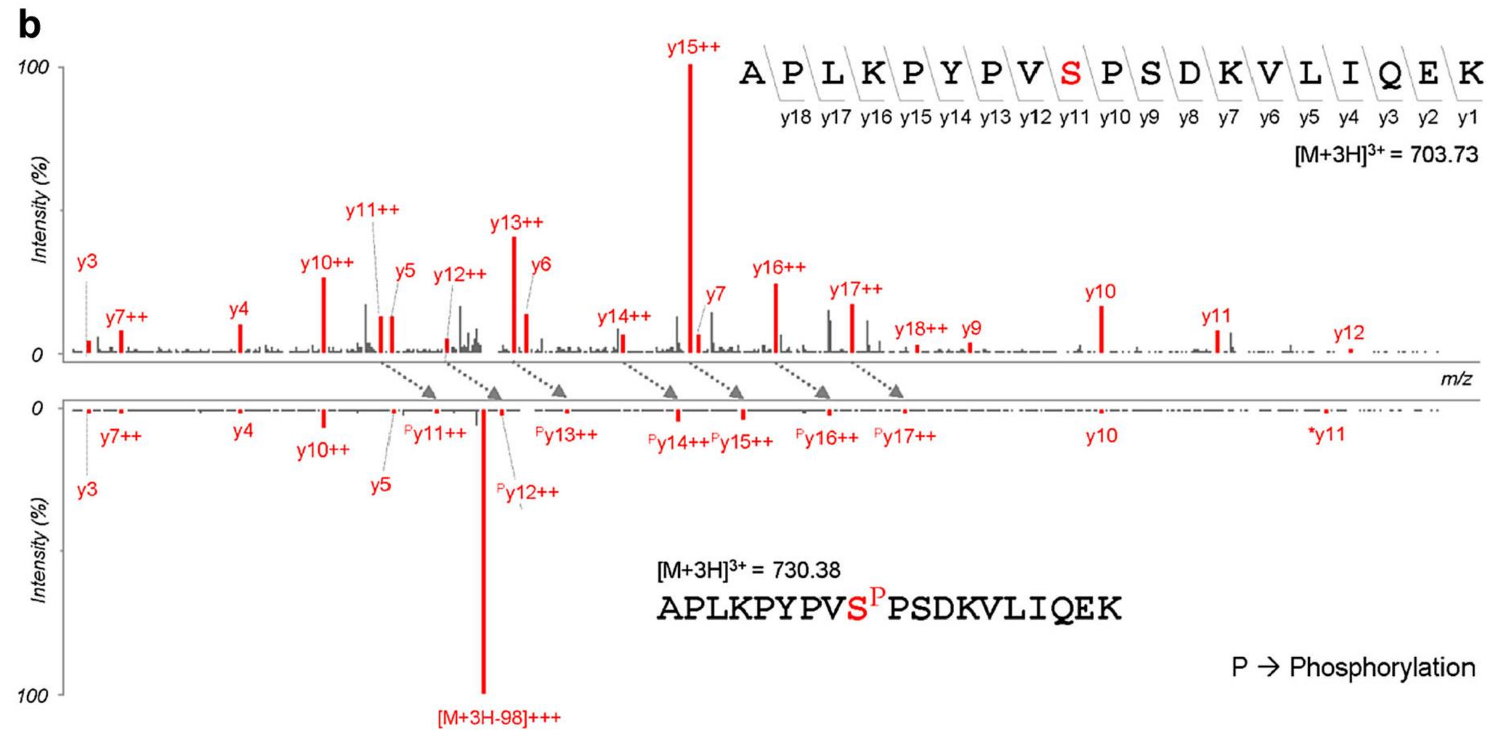
- ~30% of all proteins are thought to be phosphorylated
- Protein kinases are coded by >2000 genes (518 human protein kinases with a conserved catalytic domain)
  - Receptor tyrosine kinases (growth factor receptors: EGFR, FGFR, VEGFR)
  - Non-receptor tyrosine kinases
  - Signal transducing serine/threonine 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 phosphoylation sites
  - Phospho groups are somewhat labile during MS/MS, losses of 98 from precursor and product ions often observed

Affinity enrichment of phosphopeptides



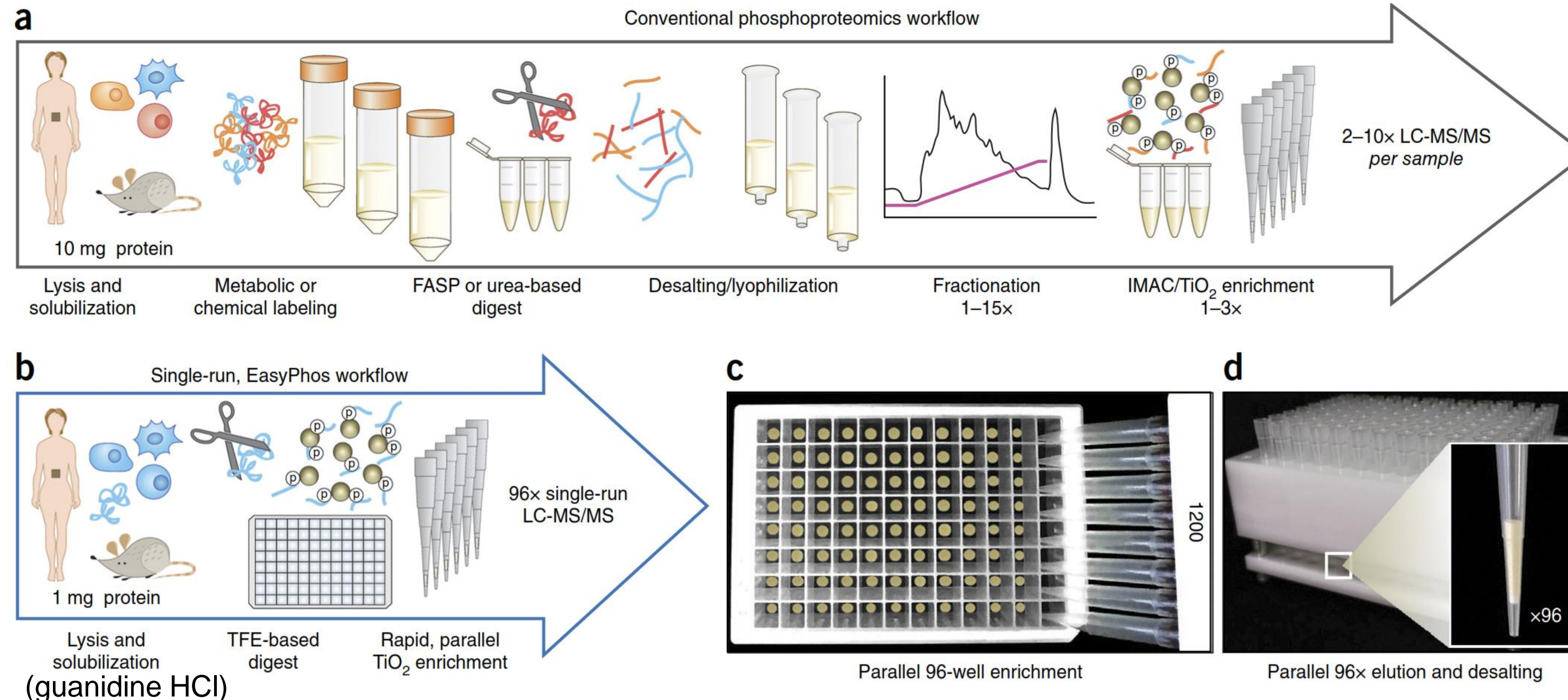
[1]P.A. Grimsrud, D.L. Swaney, C.D. Wenger, N.A. Beauchene, and J.J. Coon, Phosphoproteomics for the masses. ACS Chem Biol 5 (2010) 105-19.

Influence of phosphorylation on CAD tandem MS



Na, S. and E. Paek, Mass Spectrometry Reviews, 2015. **34**(2): p. 133-147.

EasyPhos phosphoproteomics metho

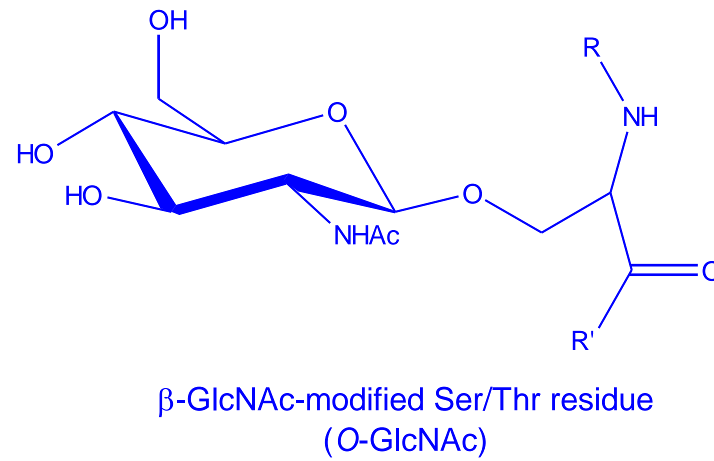


(guanidine HCl)

Figure 1 Scalable, EasyPhos phosphoproteomics platform for single-run analysis of phosphoproteomes compared with conventional workflows. (**a**) Conventional phosphoproteomics workflows generally require around 10 mg of sample lysate, use FASP (filter-assisted sample preparation) or urea-based protein digestion, followed by peptide desalting and lyophilization, peptide fractionation using strong cation exchange or high-pH reversed-phase chromatography. Phosphopeptides are enriched by IMAC or TiO_2, sometimes multiple times, resulting in numerous LC-MS/MS measurements per biological sample to be analyzed. (**b**) The phosphoproteomics workflow described here requires minimal starting materials, no fractionation, and uses TFE-based digestion, eliminating the need for peptide desalting before streamlined phosphopeptide enrichment in tubes or 96-well format. Phosphopeptides are subsequently analyzed by single-run LC-MS/MS measurements. (**c**,**d**) Rapid, parallel enrichment in 96-well plate format (**c**) and StageTip-based elution (**d**).

Humphrey, S.J., S.B. Azimifar, and M. Mann, Nat Biotechnol, 2015. **33**(9): p. 990-5.

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)

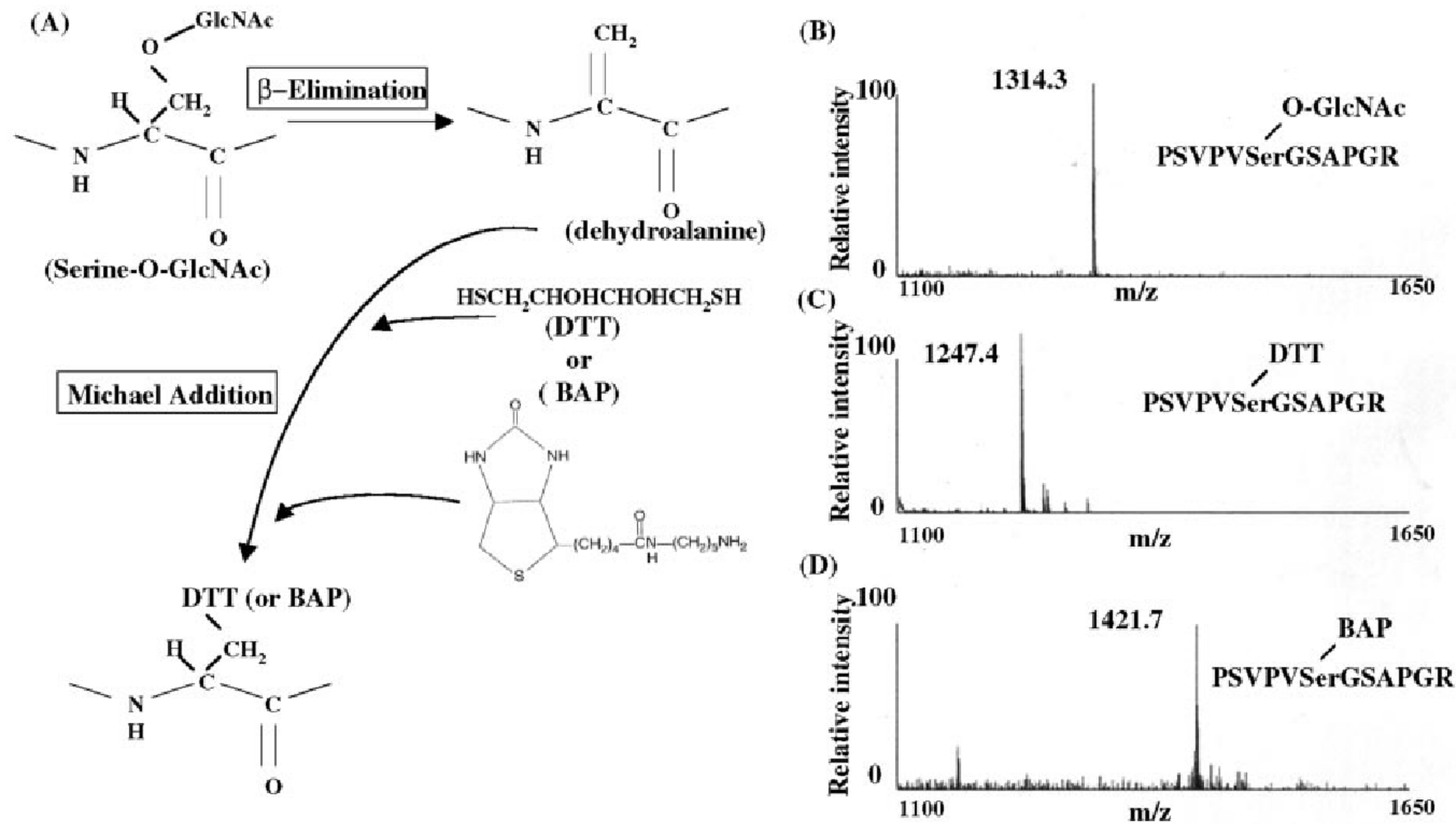


FIG. 2. ***β*-Elimination of *O*-GlcNAc and replacement with DTT (BEMAD) or BAP through Michael addition chemistry.** *A*, strategy for replacement of serine- or threonine-linked *O*-GlcNAc with the stable affinity tags DTT or BAP after *β*-elimination. *B*, *C*, and *D*, MALDI-TOF analysis of a synthetic *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

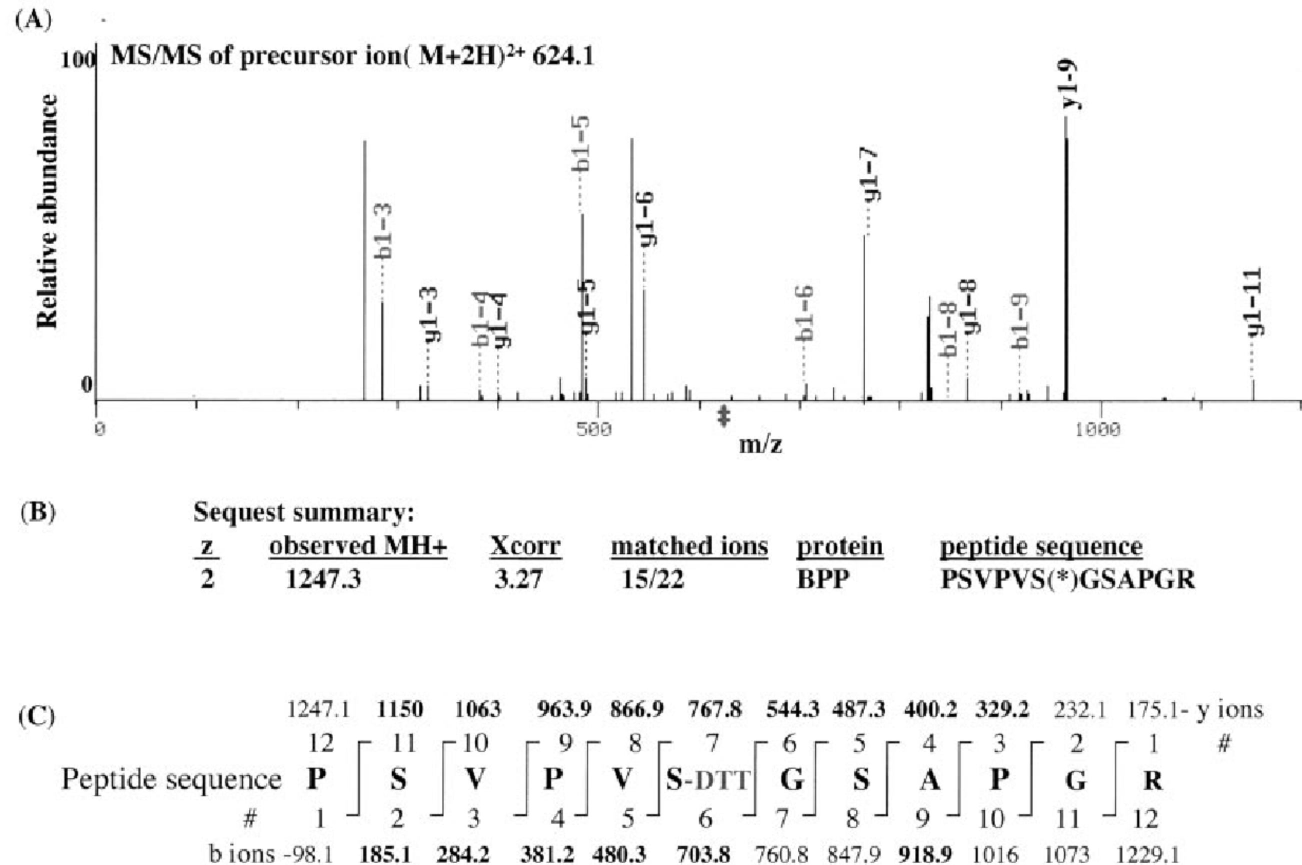
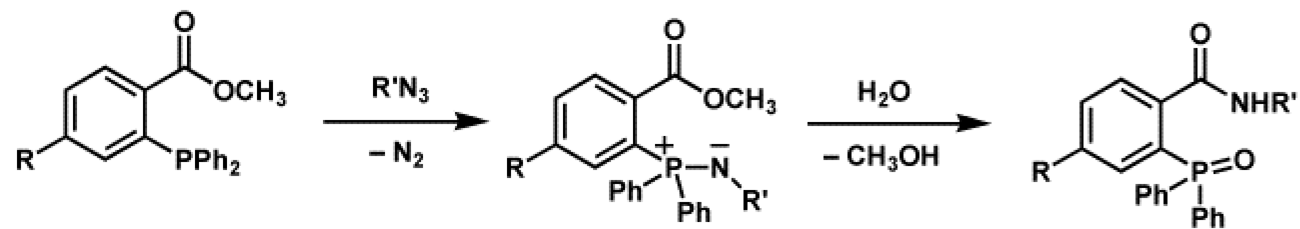


Fig. 5. **DTT replacement of *O*-GlcNAc through BEMAD is stable during tandem mass spectrometry, allowing for identification of th peptide and the DTT-modified residue.** BEMAD was performed on the peptide PSVPVS(*O*-GlcNAc)GSAPGR, and the sample was analyze by nanospray LC-MS/MS. *A*, MS/MS spectrum 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.

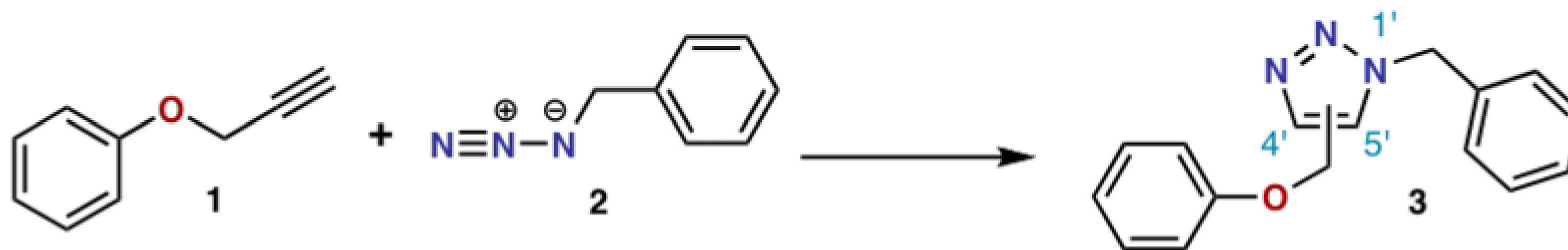
Wells, L.; Vosseller, K.; Cole, R. N.; Cronshaw, J. M.; Matunis, M. J.; Hart, G. W. *Mol Cell Proteomics* **2002**, *1*, 791-804.

Organic chemical ligation reactions

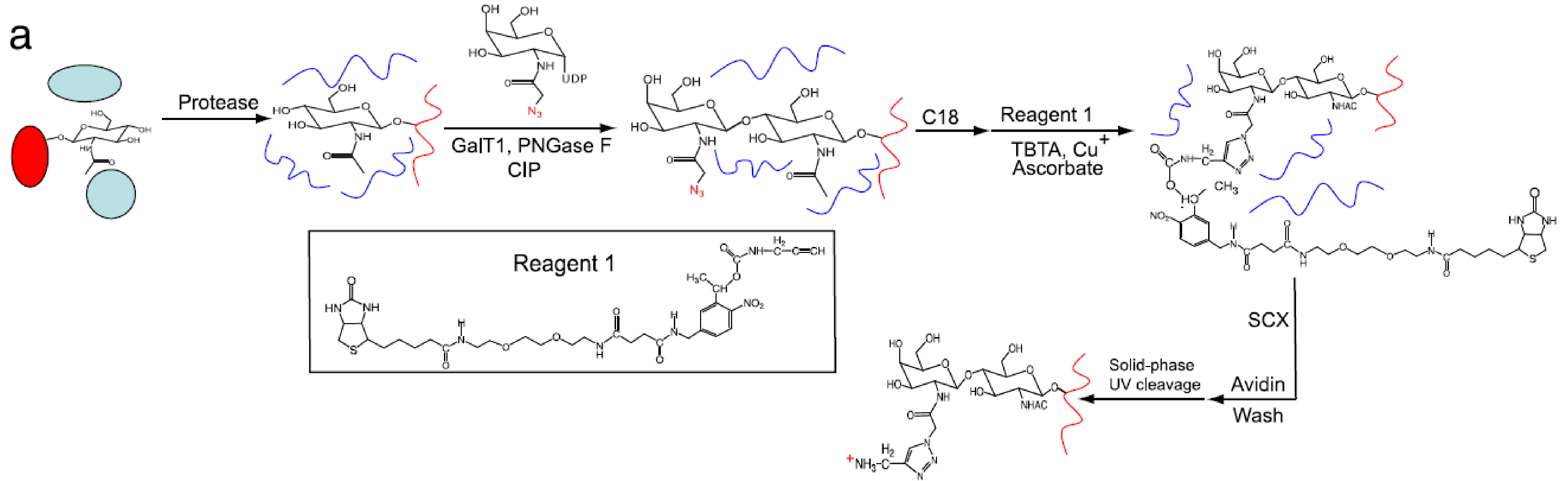
Staudinger ligation:



Click chemistry:

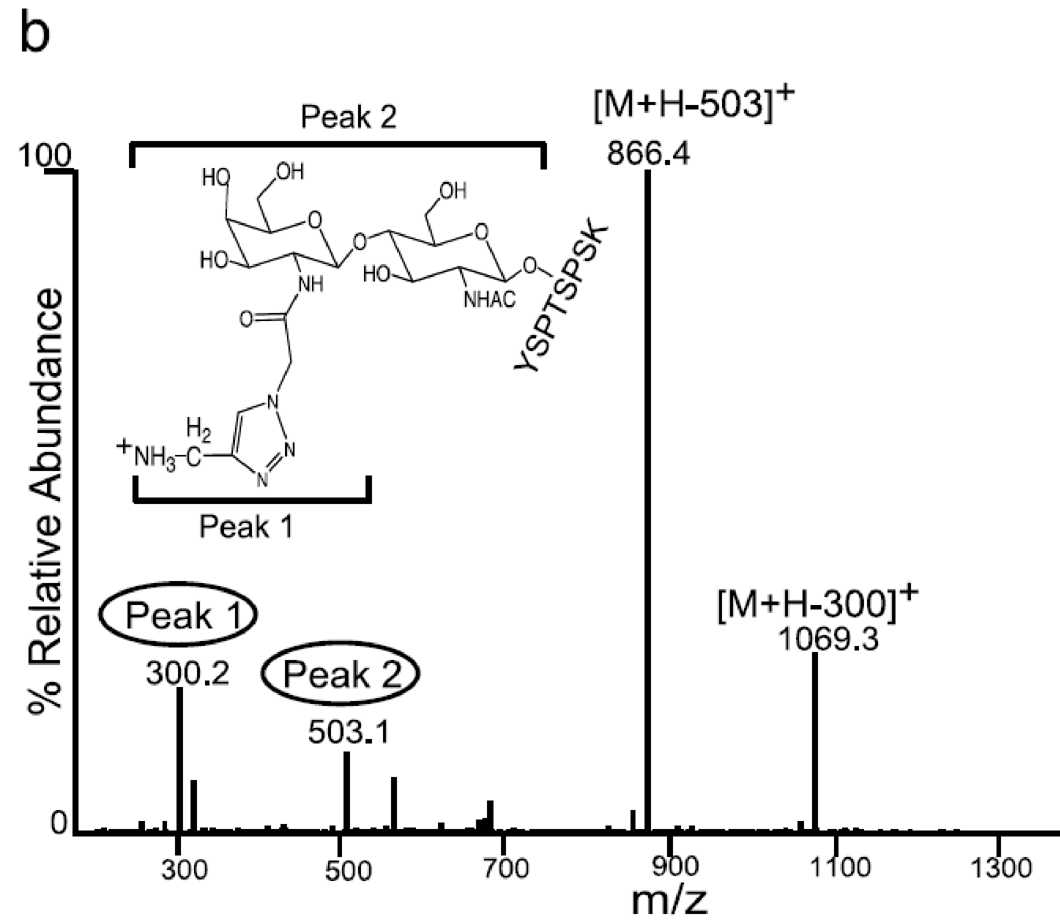


Enrichment of O-GlcNAc-peptides
GalT1 labeling and tagging with photocleavable biotin



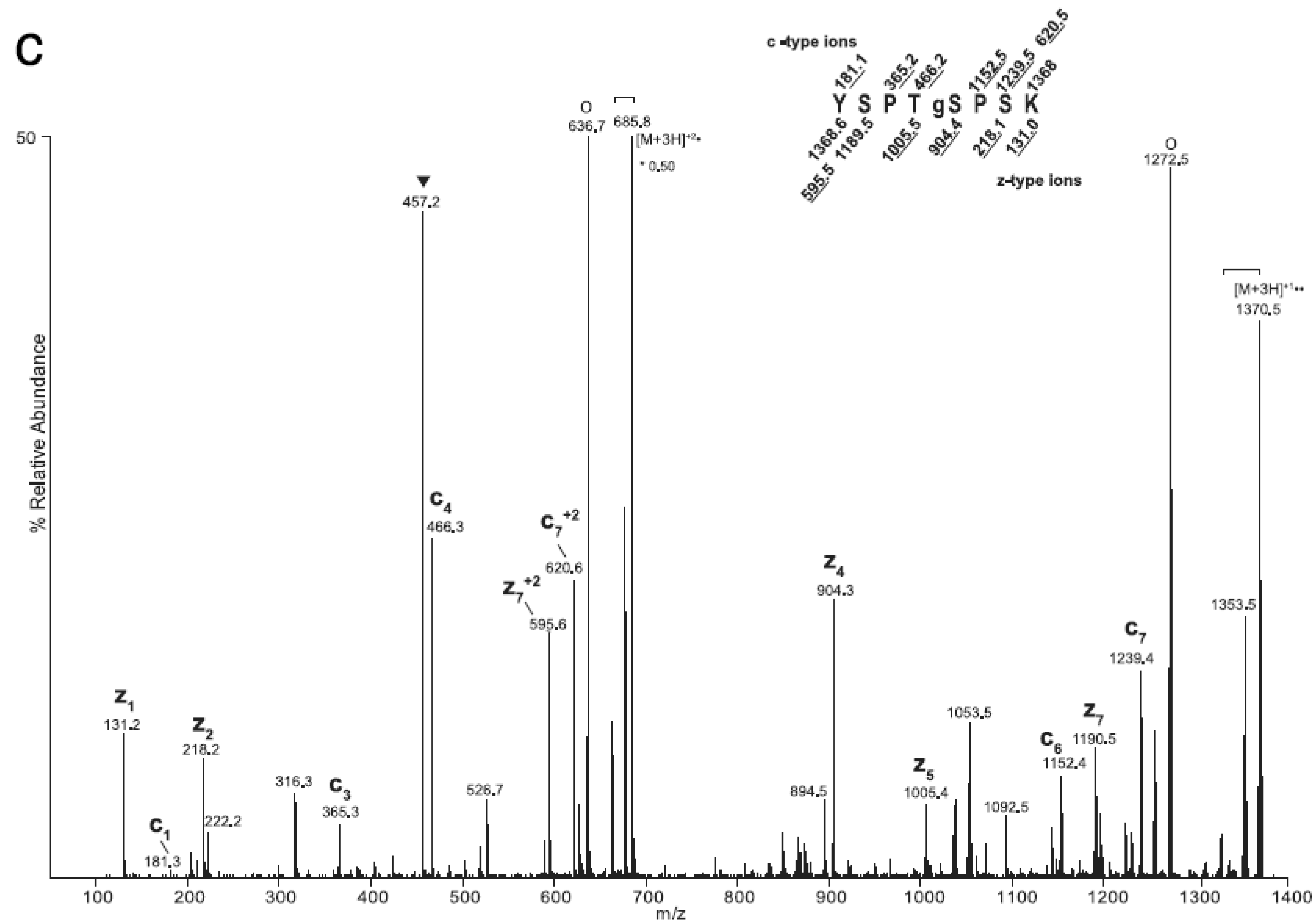
Wang, Z.; Udeshi, N. D.; O'Malley, M.; Shabanowitz, J.; Hunt, D. F.; Hart, G. W. Mol Cell Proteomics 2010, 9, 153-160.

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.

J. Ma, and G.W. Hart, O-GlcNAc profiling: from proteins to proteomes; Clin Proteomics; 11; (2014) 8.

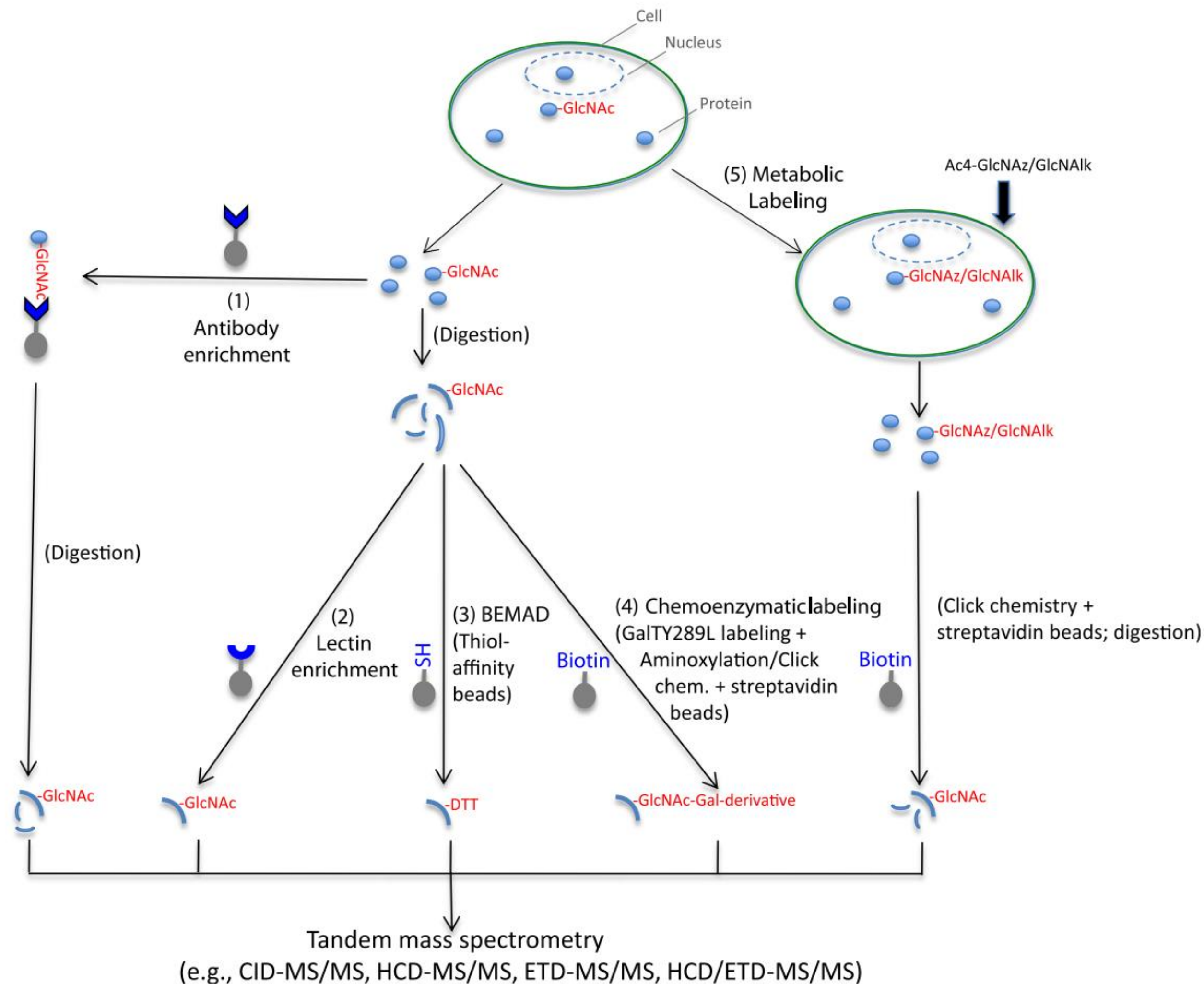
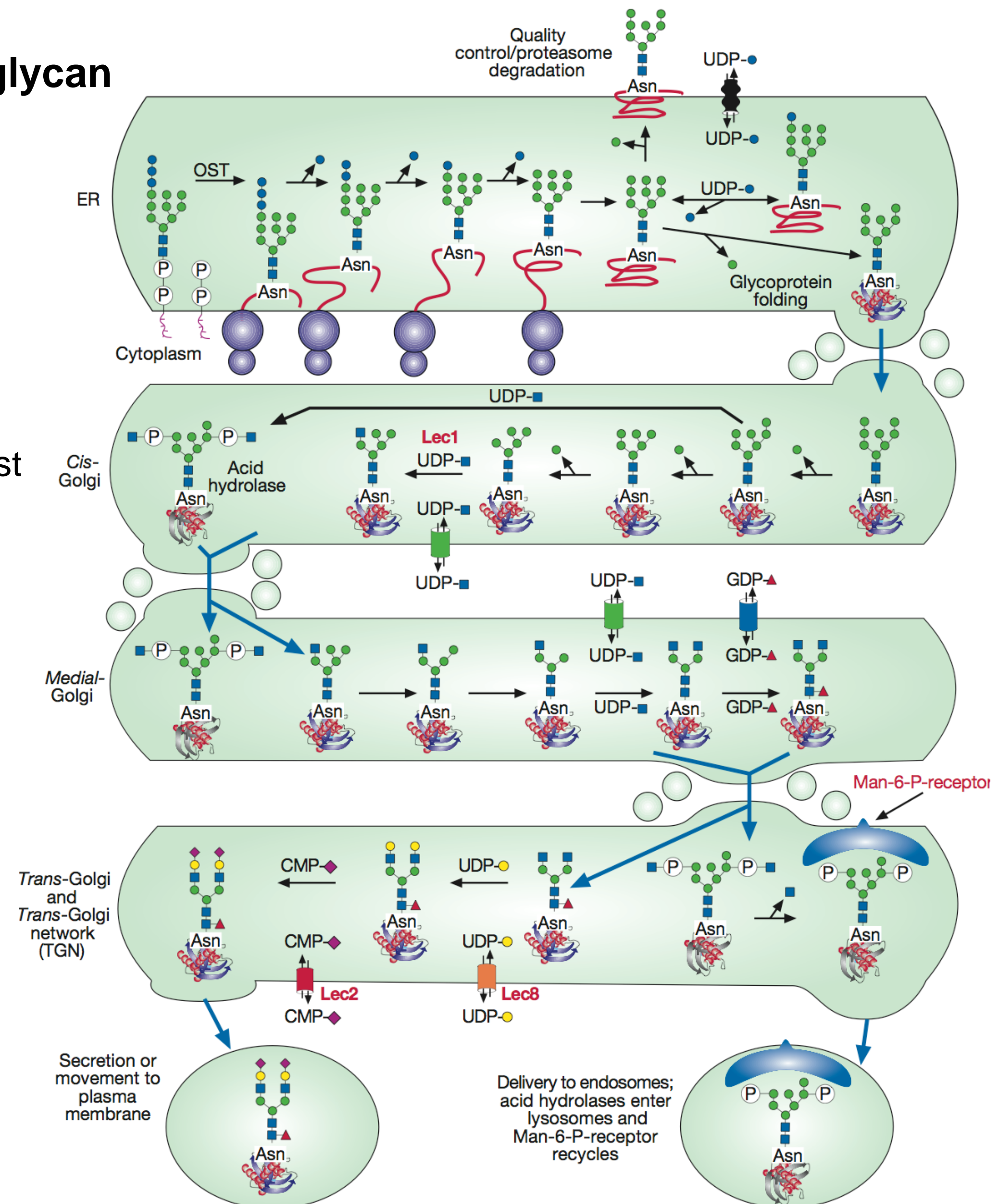


Figure 2 Scheme for the enrichment of *O*-GlcNAcylated proteins/peptides. Most commonly used strategies with antibody enrichment **(1)**, lectin enrichment **(2)**, BEMAD **(3)**, chemoenzymatic labeling **(4)** and metabolic labeling **(5)** are illustrated. In (1), proteins are captured onto antibody/antibodies-conjugated beads, and the enriched ones are digested and identified by tandem mass spectrometry. In (2), (3), and (4), proteins are digested into peptides, which are captured with lectin-conjugated resin (2), thio-capture column after BEMAD (3), and streptavidin-conjugated beads after chemoenzymatic labeling (4), with the enriched peptides identified by tandem mass spectrometry. In (5), cells are fed with GlcNAc analogs GlcNAz and GlcNAlk, and the GlcNAz- and GlcNAlk-containing proteins are subjected to click chemistry, streptavidin-conjugated beads enrichment and digestion, with the digests analyzed by tandem mass spectrometry. Note: The cocktail usage of several methods (e.g., chemo-enzymatic/metabolic labeling and BEMAD) has also been applied in some cases.

Processing and maturation of an N-glycan

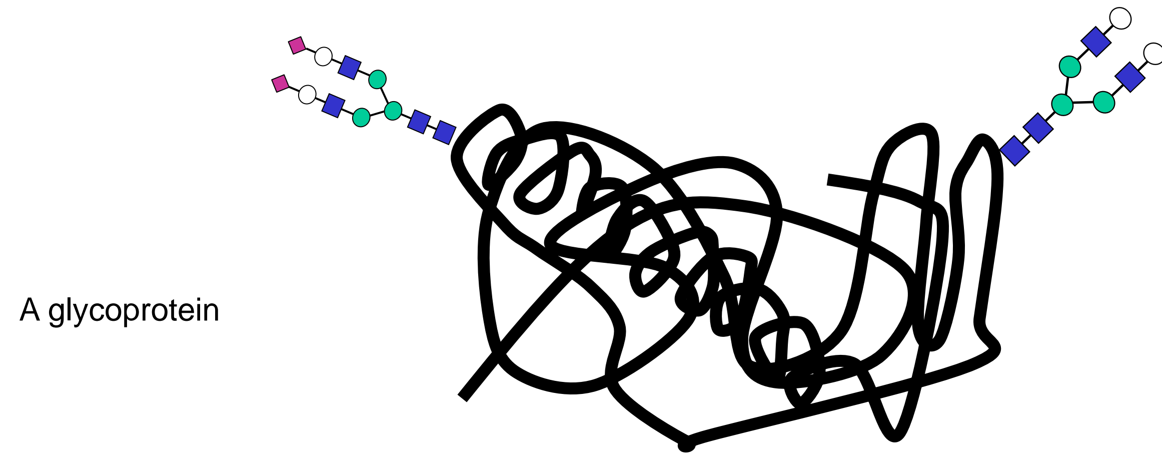
Only Man_5 glycans acted upon by Lec1; most mature glycoproteins contain $Man_{5\text{-}9}$ that escape Golgi processing and extension



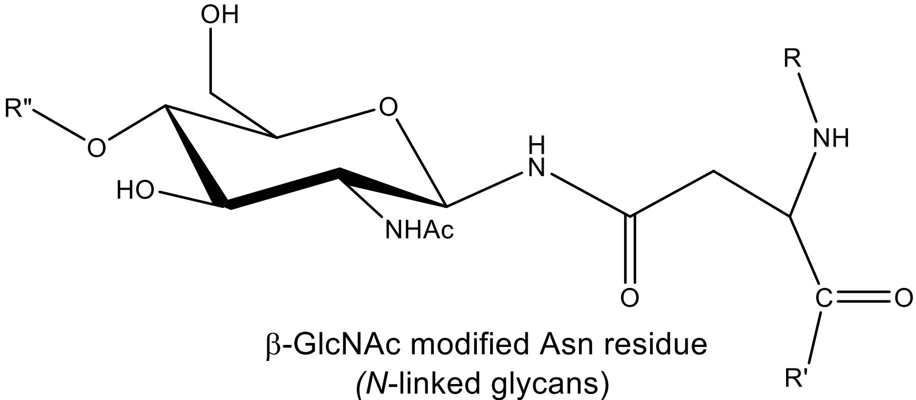
Essentials of Glycobiology
Second Edition

Chapter 8, Figure 4

Glycoproteomics:

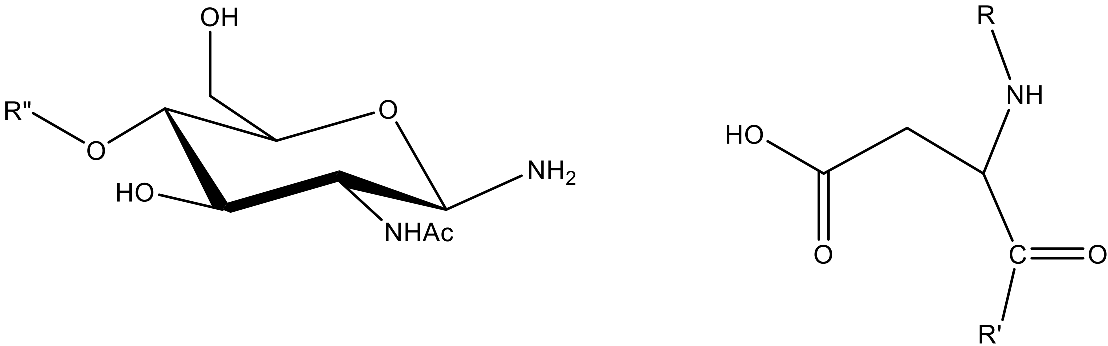


Peptide N-glycosidase F (PNGase F)

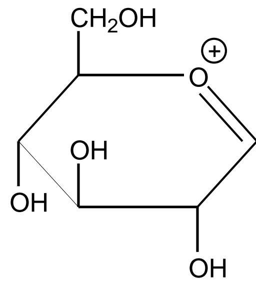


β-GlcNAc modified Asn residue
(*N*-linked glycans)

PNGaseF



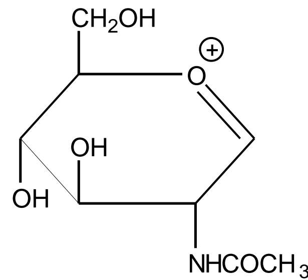
Common glycan oxonium ions



Hex

$C_6H_{11}O_5^+$

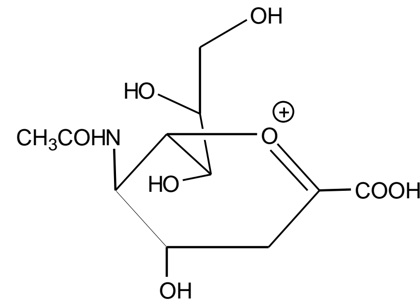
m/z 163.06



HexNAc

$C_8H_{14}NO_5^+$

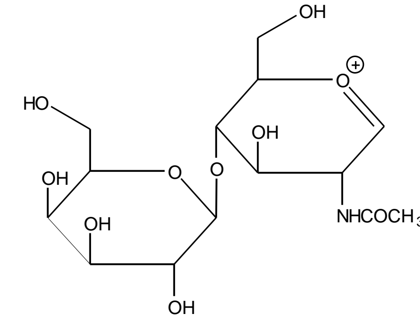
m/z: 204.09



Neu5Ac

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

m/z 292.10



HexHexNAc

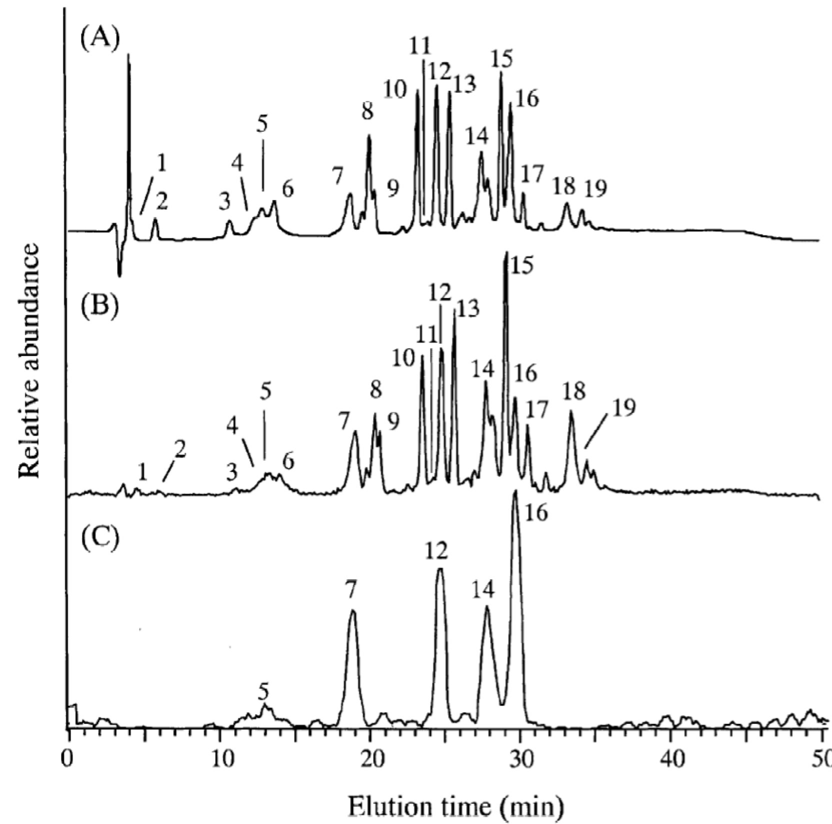
$C_{14}H_{24}NO_{10}^+$

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

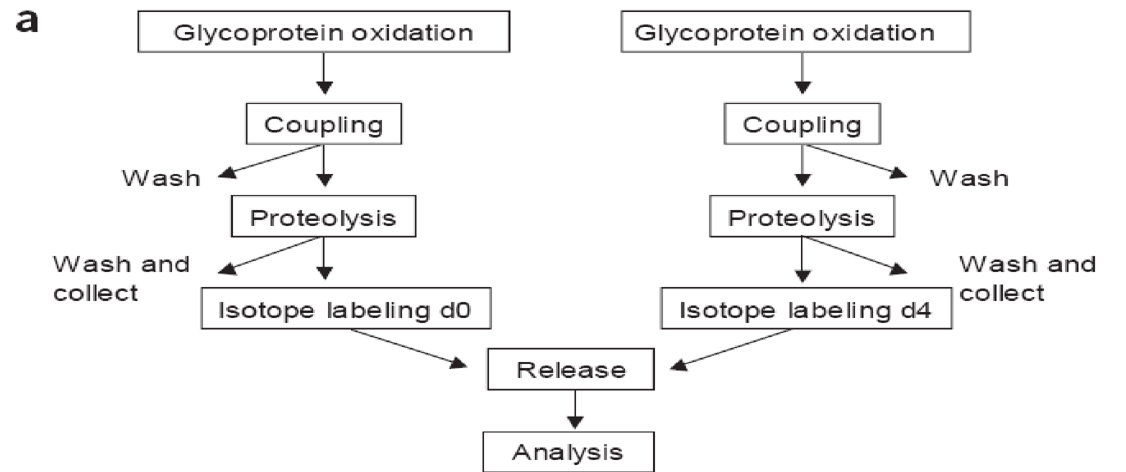
(A) UV 206 nm

(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



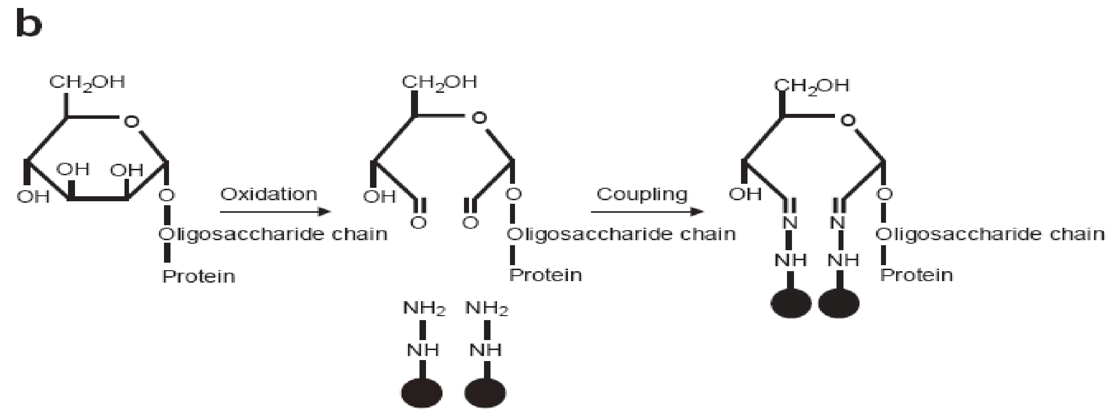


Figure 1 Schematic diagram of quantitative analysis of N-linked glycopeptides. (**a**) Strategy for quantitative analysis of glycopeptides. Proteins from two biological samples are oxidized and coupled to hydrazide resin. Nonglycosylated peptides are removed by proteolysis and extensive washes. The nonglycosylated 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 ($$$)

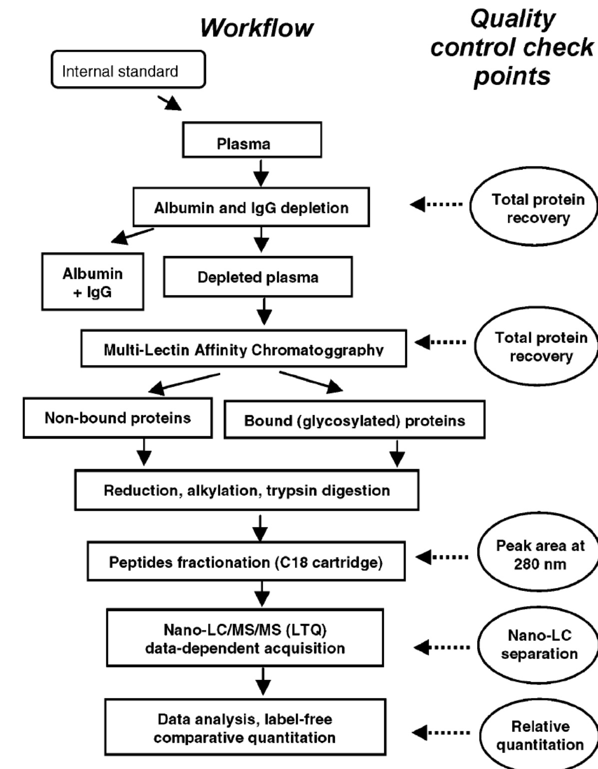
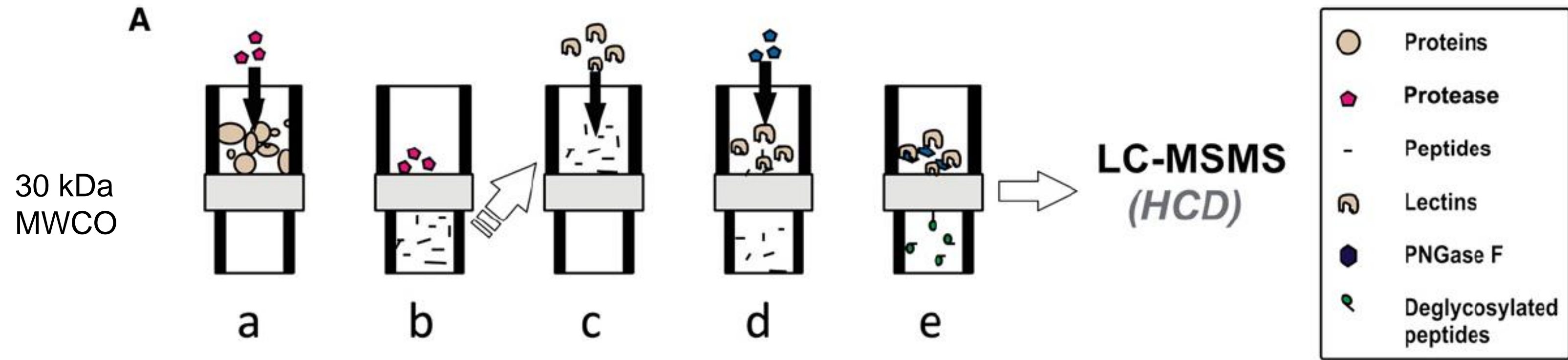


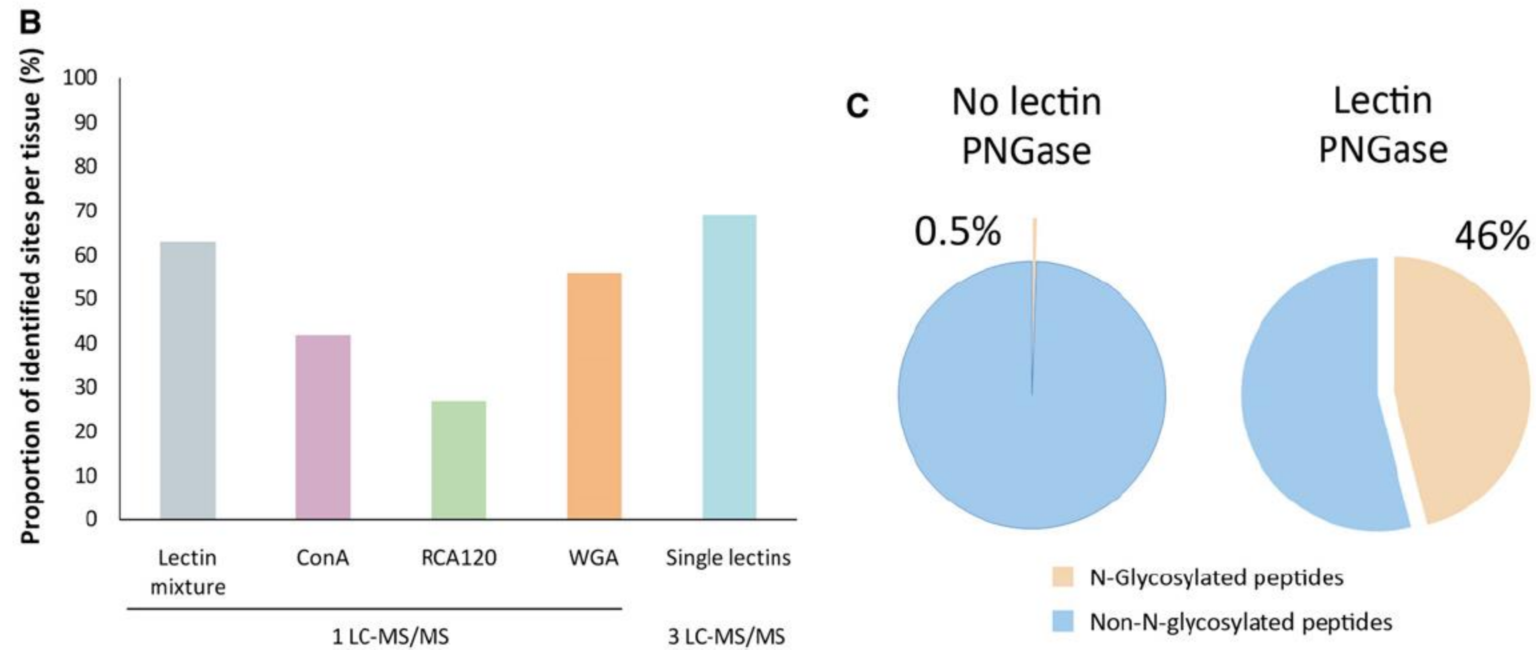
Figure 1. Flowchart of plasma analysis. To maintain the quality of the analysis, standard operating procedures were devised for each step of the method. The following quality control check-points were monitored: total protein recovery during the abundant protein depletion and M-LAC fractionation, as 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)

30 kDa
MWCO





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

Proteomic data on FASP lectin enriched glycopeptides

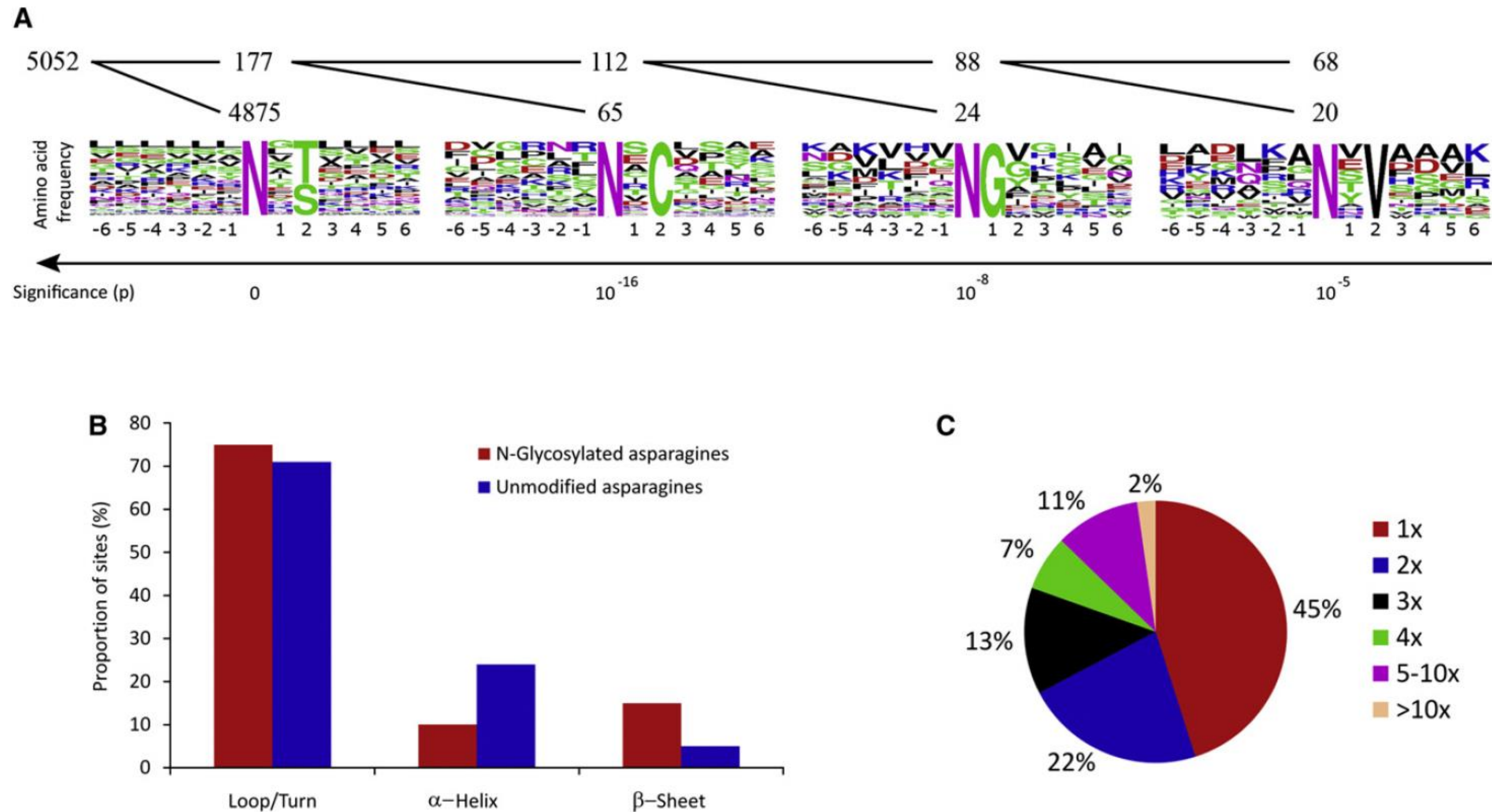


Figure 4. Sequence Recognition Motifs, Structure Preference, and Multiple Glycosylation

(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.