GMS BI 793 Lecture 5

Affinity Mass Spectrometry: Strategies for Proteome Profiling

Joseph Zaia

Feb. 19, 2013

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

Discovery proteomics is less concerned with instrumental robustness

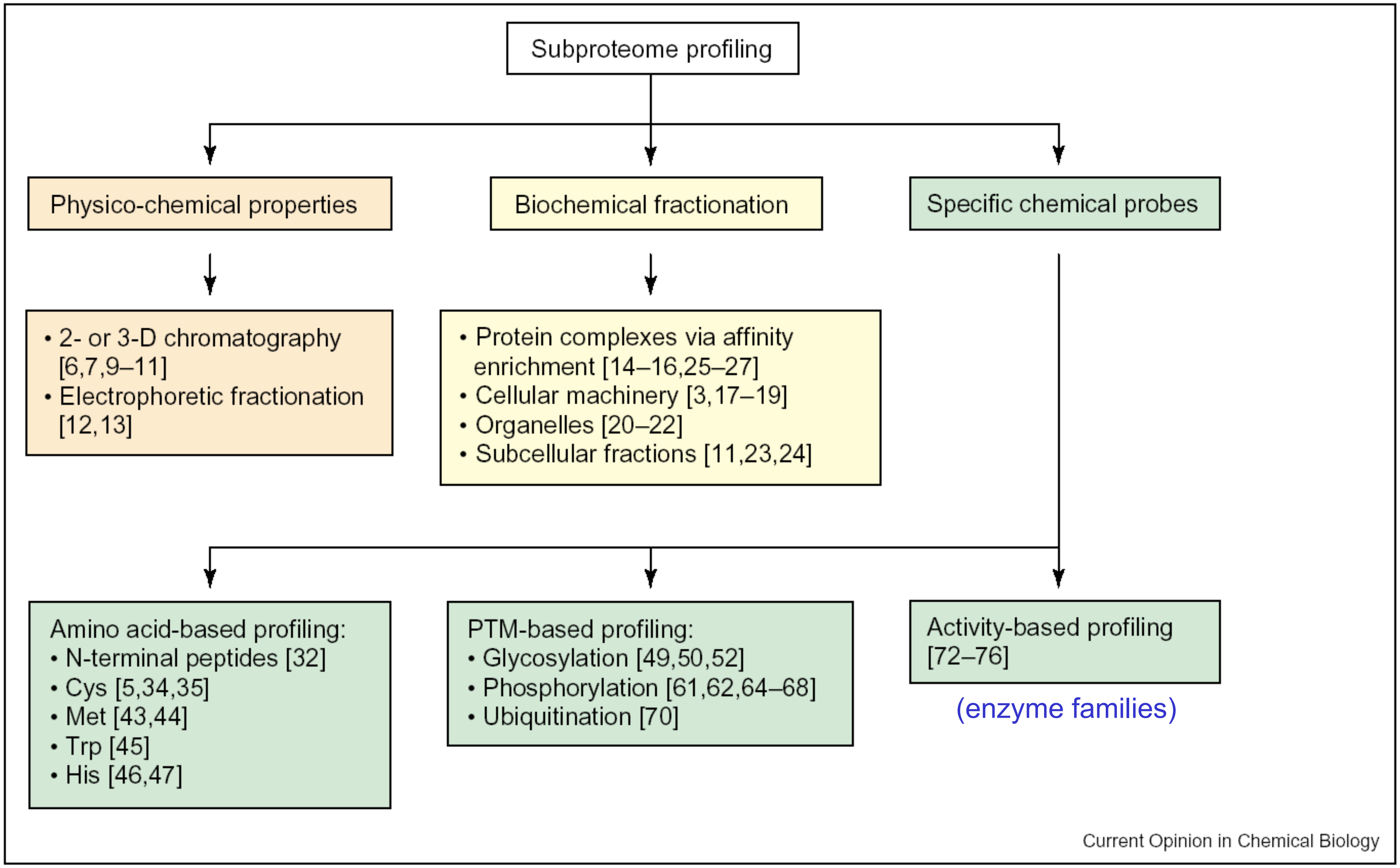
- When using stable isotope labels, can make pairwise comparisons
- Not appropriate for large clinical studies.
    - Instrumental variability
    - Effort, computation time

Quantitative 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 in mixture complexity
- Use of stable isotope labeled peptides (AQUA)
- Targeted SRM transitions: entails used of validated peptide data

Separation strategies for proteome profiling

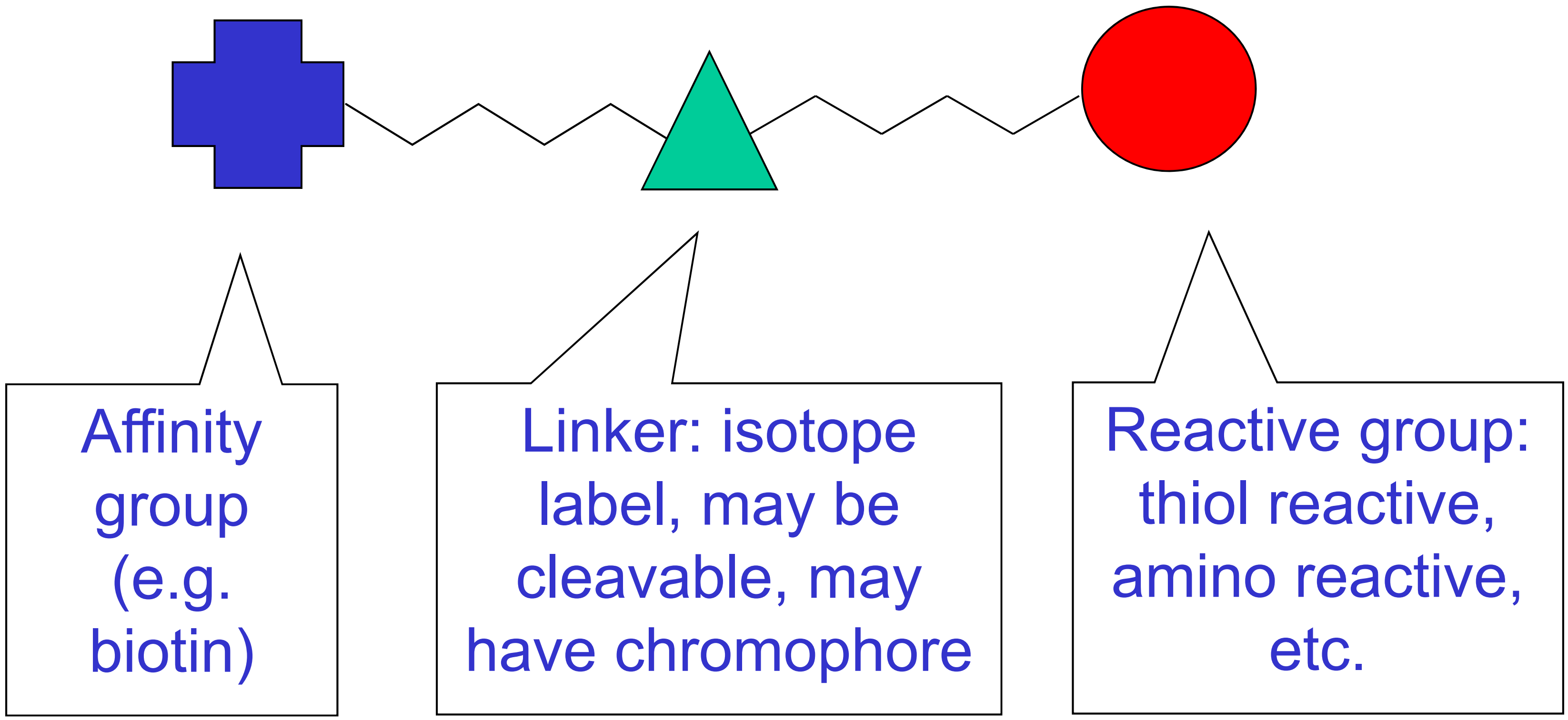
Figure 1



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



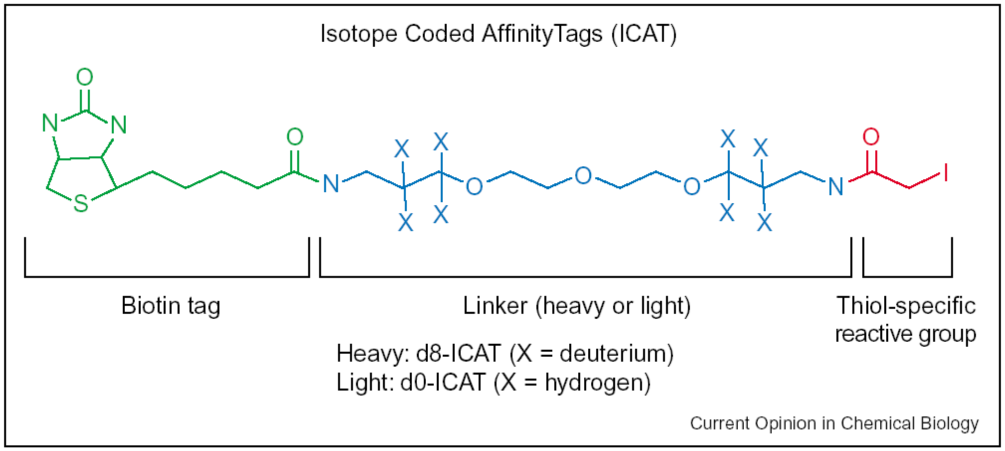
ICAT MS: biotin connected to a thiol-reactive group through an isotope-labeled linker

- **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.

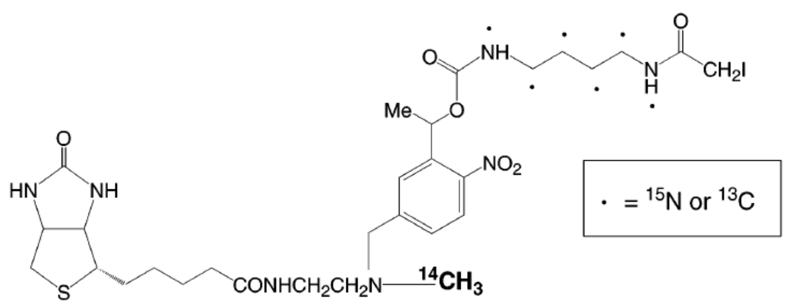
ICAT Reagents

Traditional

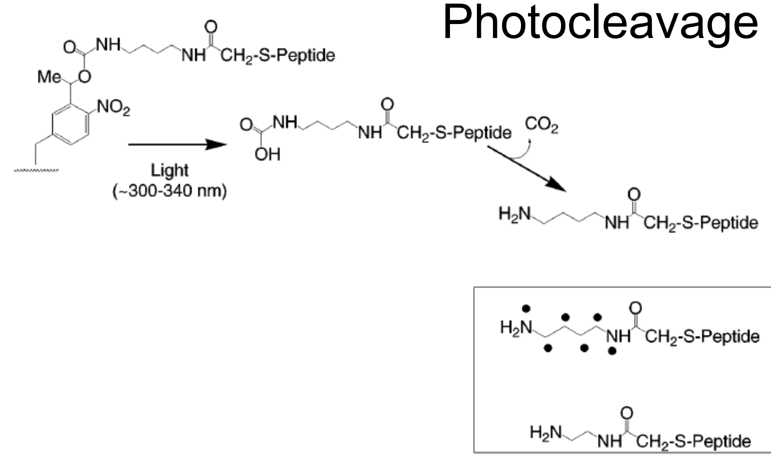


The structure of ICAT reagents, which comprise a cysteine-reactive group (red), a linker containing either heavy or light isotopes (blue) and a binding affinity tag (green).

VICAT (visible ICAT)



Photocleavage rxn



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

iTRAQ

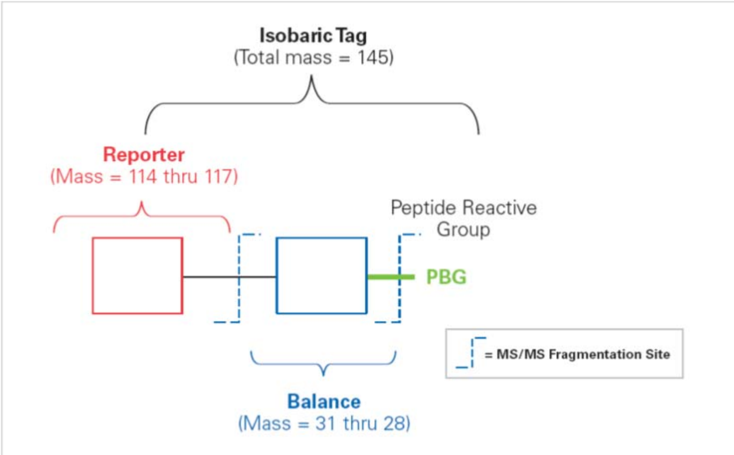


Figure 1. iTRAQ™ reagent structure

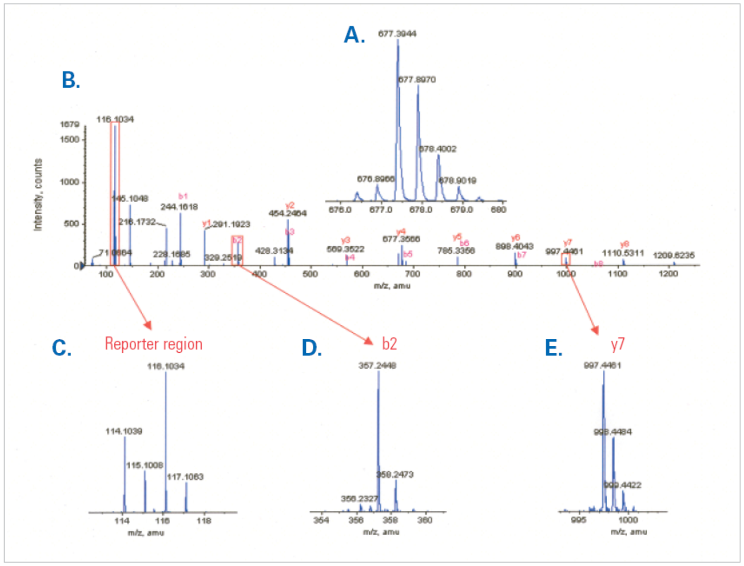


Figure 2. MS and MS/MS spectra from a multiplex sample labeled with 4 iTRAQ™ reagents showing **A.** doubly charged parent ion **B.** MS/MS spectrum corresponding to VLVDTDYK **C.** 4 diagnostic reporter ions and **D.** and **E.** peptide fragment ions

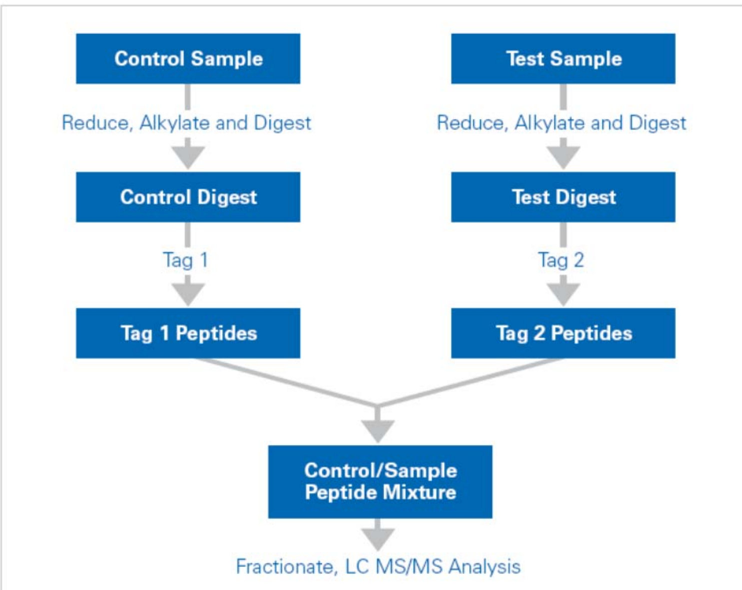
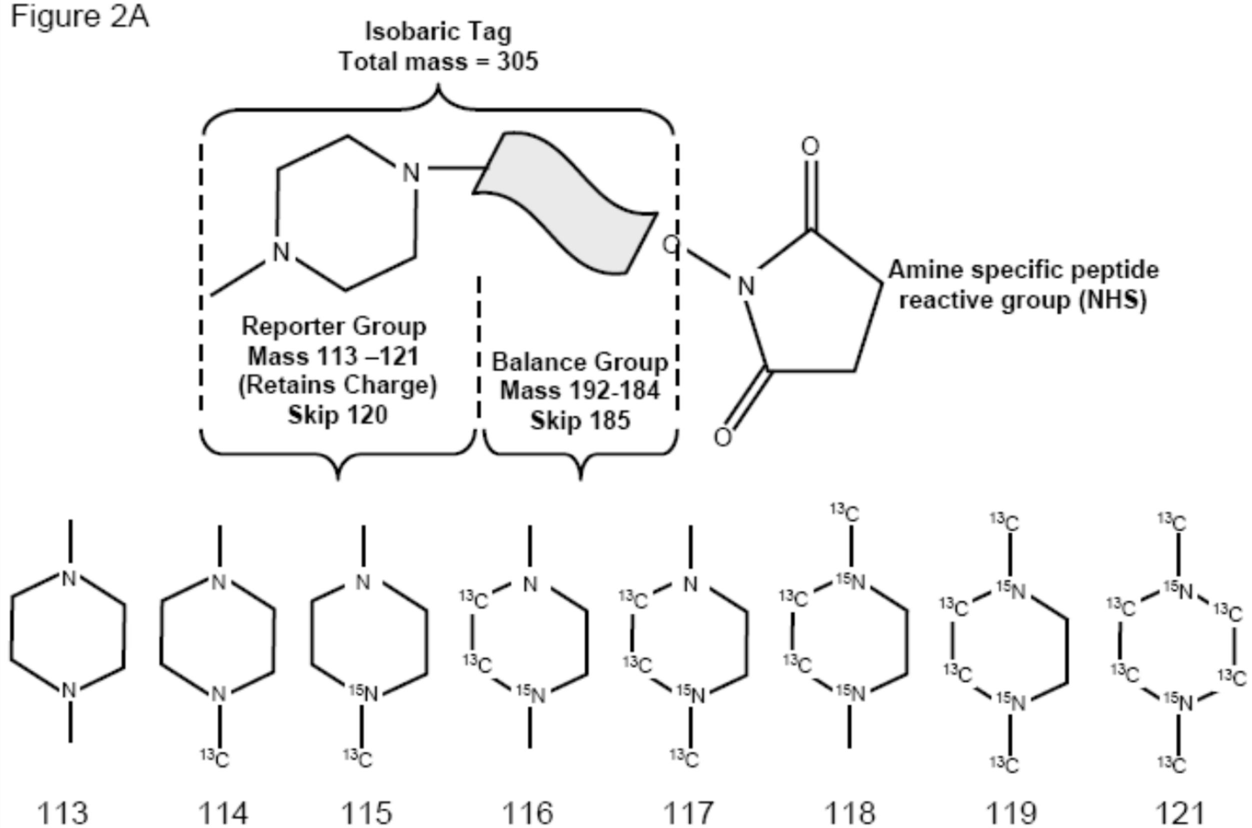


Figure 3. Duplex example of the iTRAQ™ reagent protocol.

Eight channel iTRAQ

Figure 2A



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.

Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry

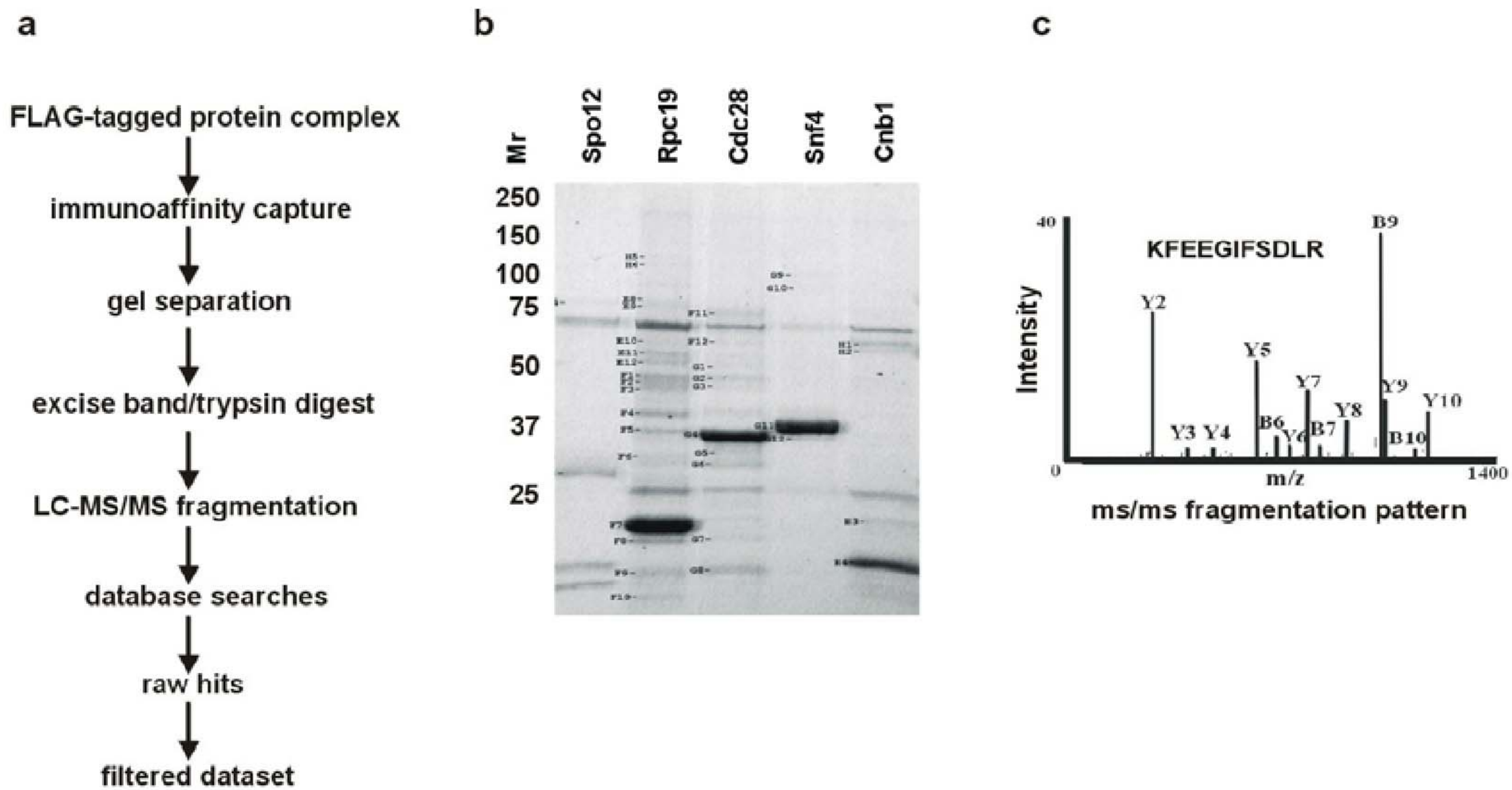
Nature **2002**, *415*, 180-183.

- 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

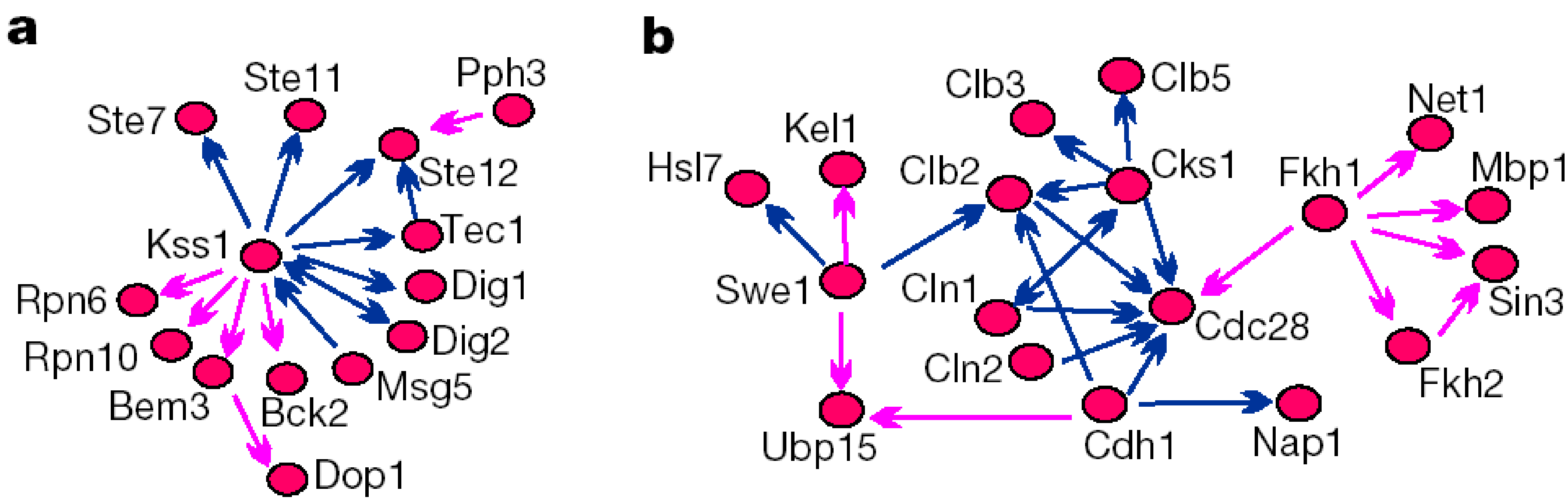
- Cited 1946 times as of 3/29/11

Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry



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

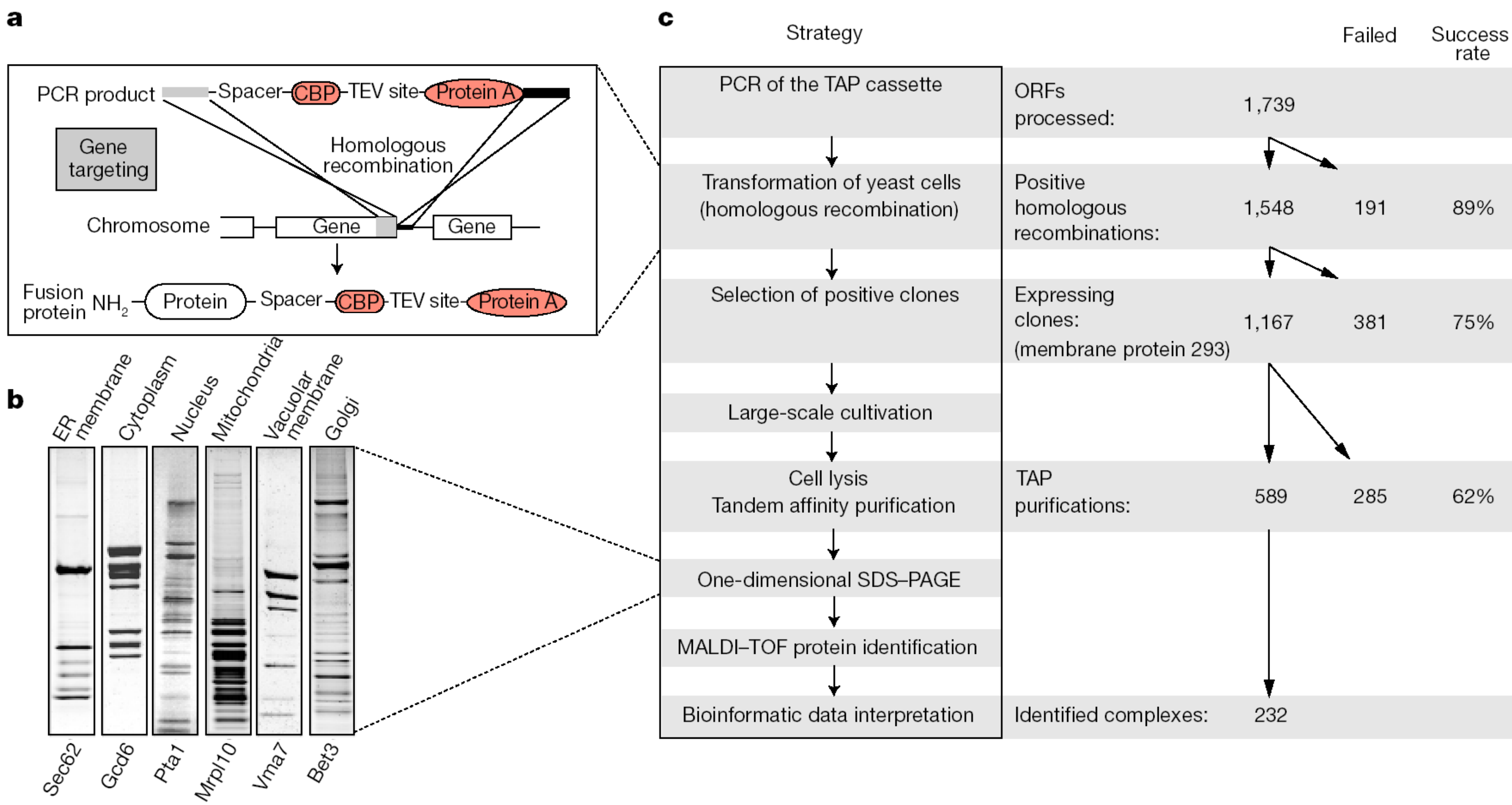
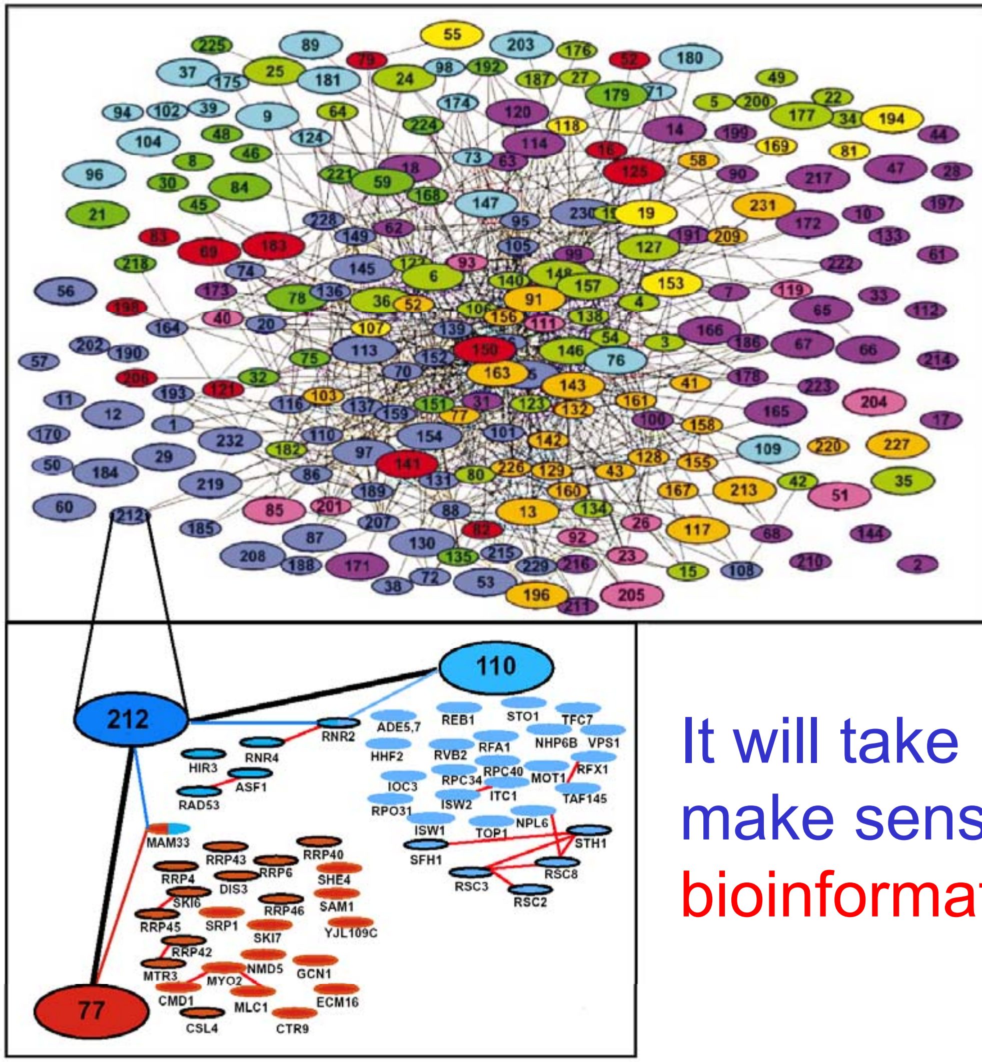


Figure 1 Synopsis of the screen. **a**, Schematic representation of the gene targeting procedure. The TAP cassette is inserted at the C terminus of a given yeast ORF by homologous recombination, generating the TAP-tagged fusion protein. **b**, Examples of TAP complexes purified from different subcellular 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



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

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].

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

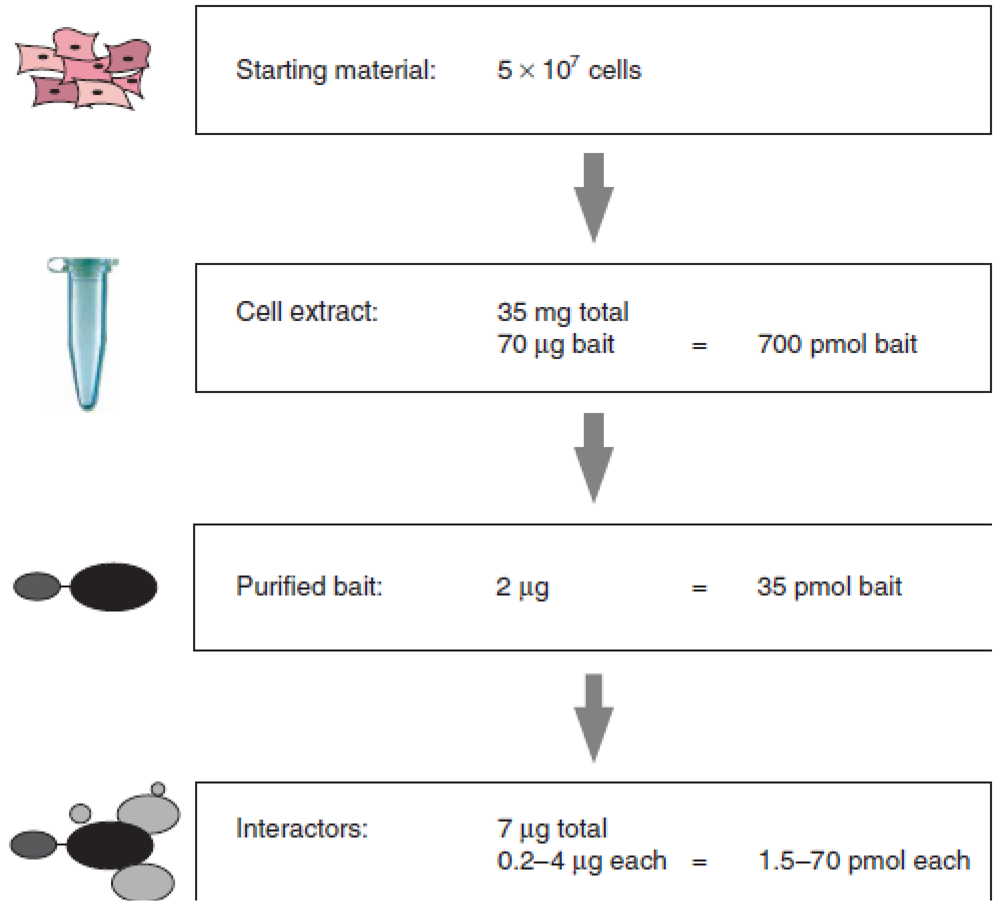


Figure 6 | Quantitative overview of the TAP procedure. A typical TAP purification is done from 5×10^7 cells, yielding 35 mg of total cell extract. Based on the quantification presented in **Figure 4**, the cell extract contains 70 µg (700 pmol) of bait protein. The final eluate contains 2 µg (35 pmol) 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.

10 recent articles referencing *Nature* **2002**, *415*, 141-147

1. Bonfiglio, J. J., Maccarrone, G., Rewerts, C., Holsboer, F., Arzt, E., Turck, C. W., and Silberstein, S. (2011) Characterization of the B-Raf interactome in mouse hippocampal neuronal cells, *Journal of Proteomics 74*, 186-198.
2. Feng, J. X., Jiang, R., and Jiang, T. (2011) A Max-Flow-Based Approach to the Identification of Protein Complexes Using Protein Interaction and Microarray Data, *Ieee-Acm Transactions on Computational Biology and Bioinformatics 8*, 621-634.
3. Haura, E. B., Muller, A., Breitwieser, F. P., Li, J. N., Grebien, F., Colinge, J., and Bennett, K. L. (2011) Using iTRAQ Combined with Tandem Affinity Purification to Enhance Low-Abundance Proteins Associated with Somatically Mutated EGFR Core Complexes in Lung Cancer, *J Proteome Res 10*, 182-190.
4. Hu, L., Huang, T., Liu, X. J., and Cai, Y. D. (2011) Predicting Protein Phenotypes Based on Protein-Protein Interaction Network, *PLoS ONE 6*.
5. Jager, S., Gulbahce, N., Cimermancic, P., Kane, J., He, N. H., Chou, S., D'Orso, I., Fernandes, J., Jang, G., Frankel, A. D., Alber, T., Zhou, Q. A., and Krogan, N. J. (2011) Purification and characterization of HIV-human protein complexes, *Methods 53*, 13-19.
6. Lin, M. Z., Shen, X. L., and Chen, X. (2011) PAIR: the predicted Arabidopsis interactome resource, *Nucleic Acids Res 39*, D1134-D1140.
7. Neilson, K. A., Ali, N. A., Muralidharan, S., Mirzaei, M., Mariani, M., Assadourian, G., Lee, A., van Sluyter, S. C., and Haynes, P. A. (2011) Less label, more free: Approaches in label-free quantitative mass spectrometry, *Proteomics 11*, 535-553.
8. Pflieger, D., Gonnet, F., van Bentem, S. D., Hirt, H., and de la Fuente, A. (2011) LINKING THE PROTEINS-ELUCIDATION OF PROTEOME-SCALE NETWORKS USING MASS SPECTROMETRY, *Mass Spectrom Rev 30*, 268-297.
9. Sendina-Nadal, I., Ofran, Y., Almendral, J. A., Buldu, J. M., Leyva, I., Li, D. Q., Havlin, S., and Boccaletti, S. (2011) Unveiling Protein Functions through the Dynamics of the Interaction Network, *PLoS ONE 6*.
10. Valentini, G. (2011) True Path Rule Hierarchical Ensembles for Genome-Wide Gene Function Prediction, *Ieee-Acm Transactions on Computational Biology and Bioinformatics 8*, 832-847.

The number of gene products is greater than the number of genes

ABTCDSEFGHTIJKLMNOP

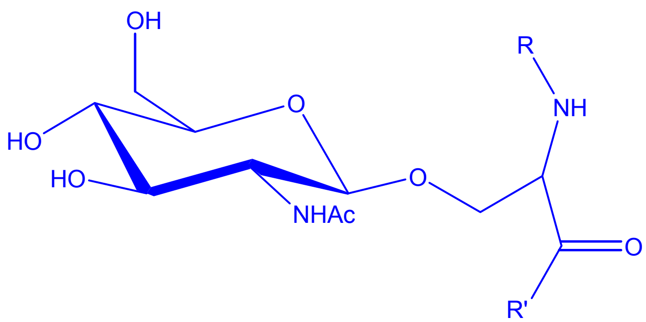
- A peptide sequence with three Ser/Thr residues, each of which may be unmodified, phosphorylated, or *O*-GlcNAcylated
- Number of possible structural variants = 3^3 = 27

Common PTM-amino acid linkages

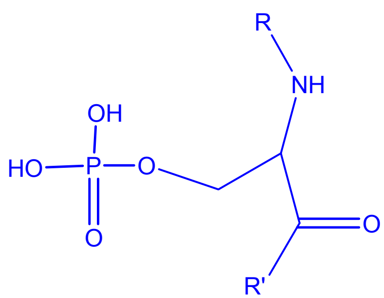
Intracellular

Cell surface, extracellular

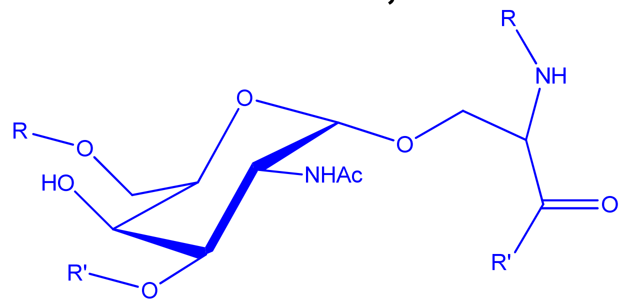
Eliminative release



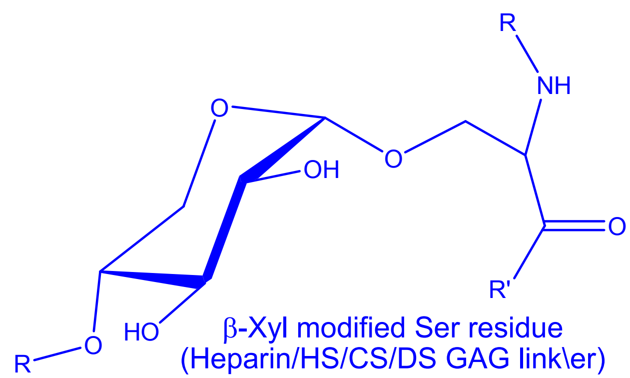
β-GlcNAc-modified Ser/Thr residue
(*O*-GlcNAc)



Phosphorylated Ser/Thr residue

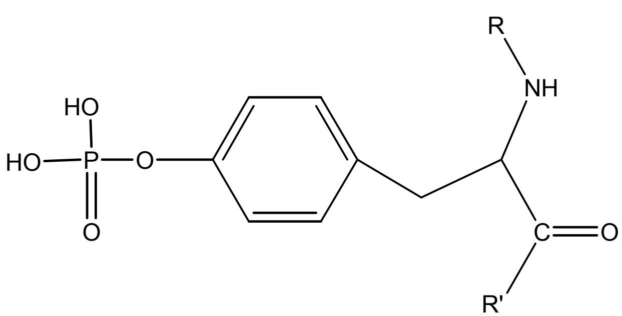


α-GalNAc-modified Ser/Thr residue
(mucin type *O*-linked glycans)

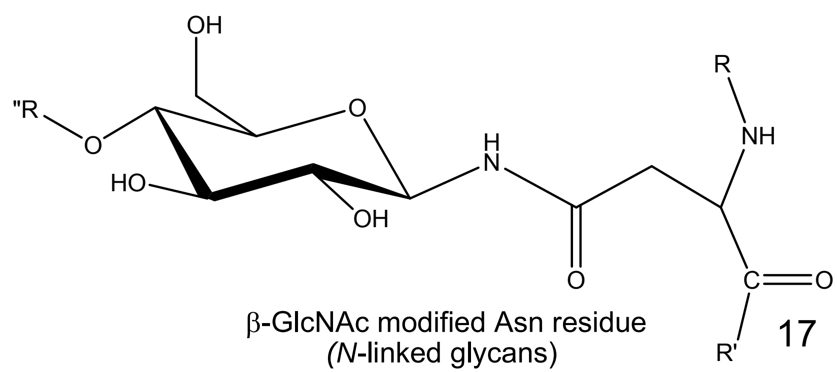


β-Xyl modified Ser residue
(Heparin/HS/CS/DS GAG link\er)

Eliminative release

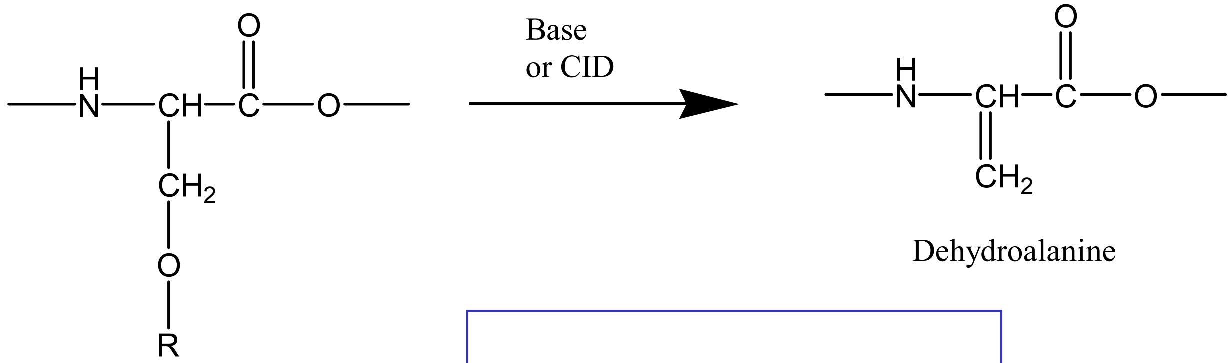


Phosphorylated Tyr residue
(Sulfated Tyrosine)

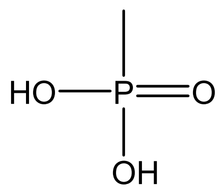


β-GlcNAc modified Asn residue
(*N*-linked glycans)

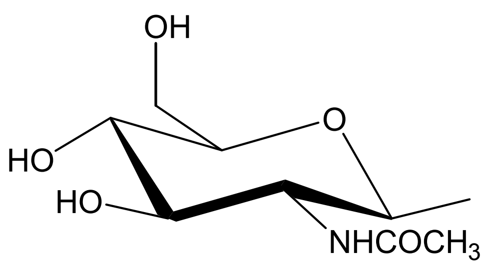
β-Elimination of PTM groups from Ser/Thr residues.



Dehydroalanine

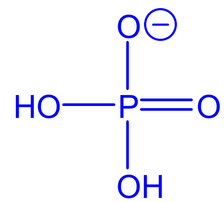


R = phosphate

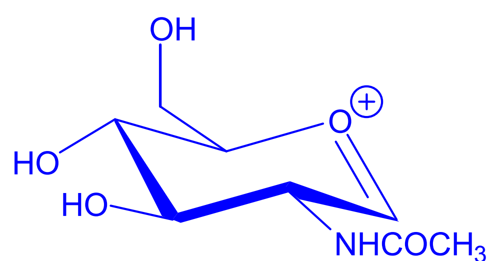


R = GlcNAc

Low *m/z* ions



$H_2O_4P^-$
Exact Mass: 96.97

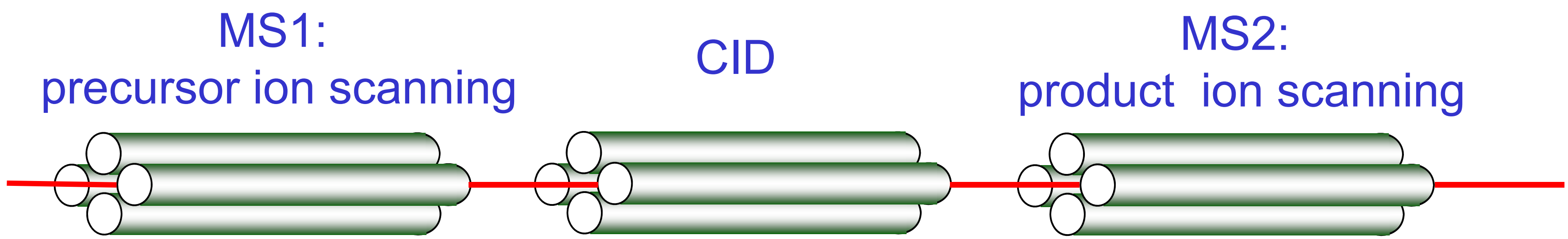


$C_8H_{14}NO_5^+$
Exact Mass: 204.09

Conditions:

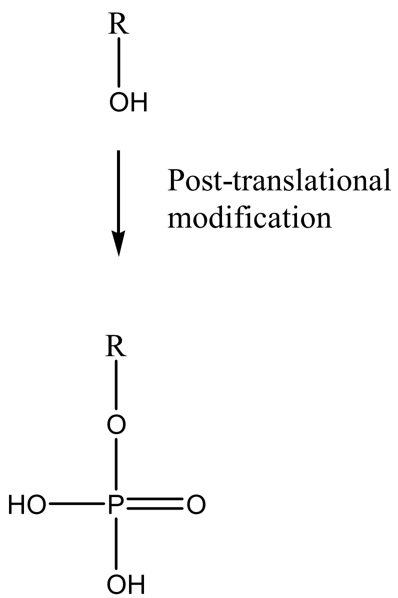
- *O*-GlcNAc 1% TEA, 0.1% NaOH, 2 hr. 50ºC
- *O*-GalNAc, 1M NaOH, 2M $NaBH_4$, o/n 20ºC
- *O*-phosphate, 4M LiOH, 1% ethanedithiol, 37ºC, 1 hr.

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



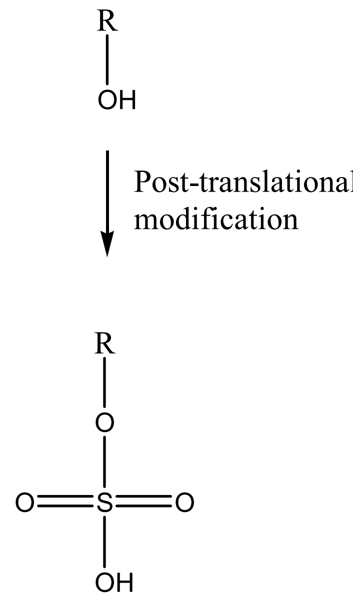
mass shift 79.96688

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

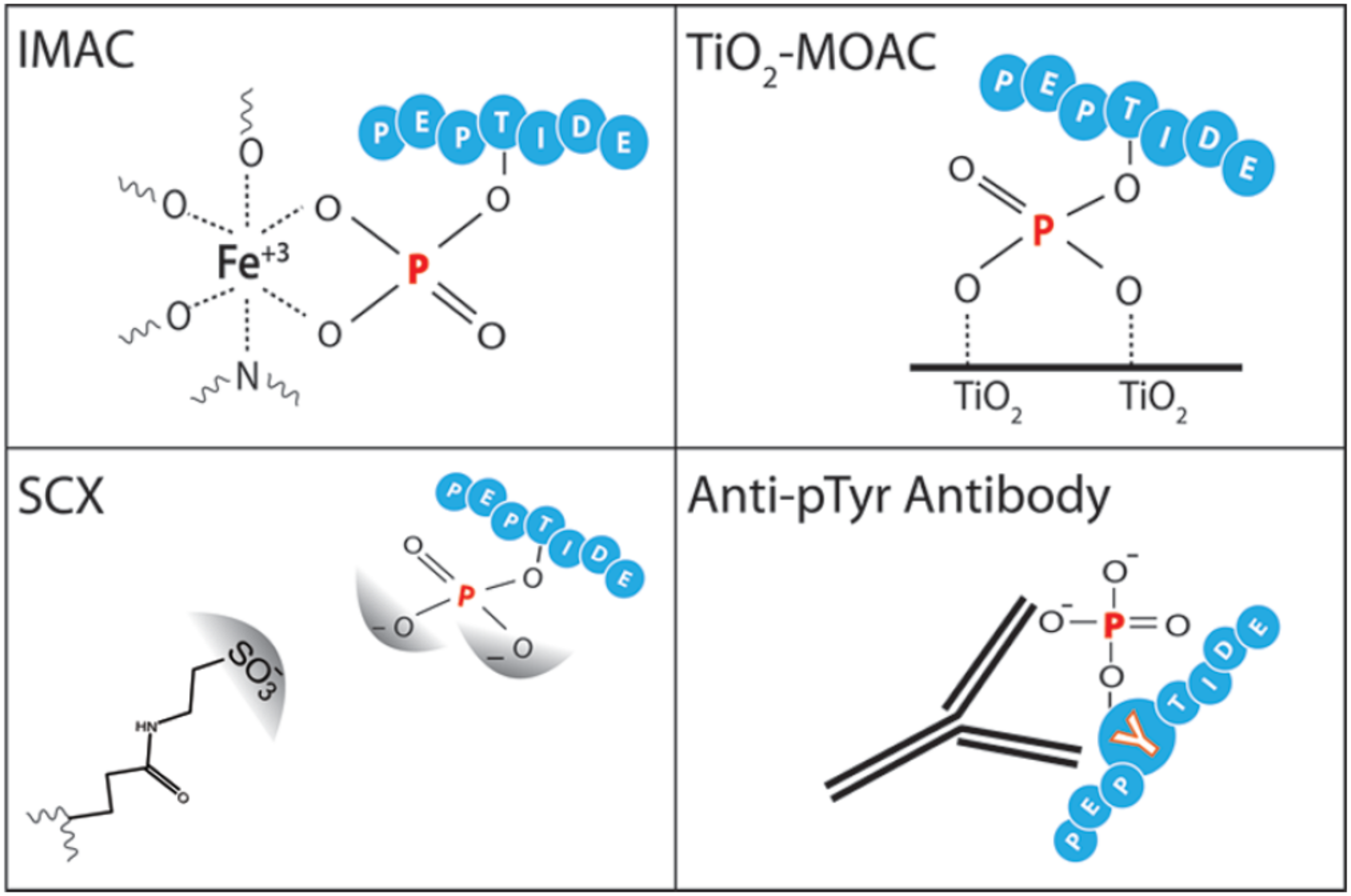
$[M-SO_3+H]^+$
loss of 79.95736

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, 2000 known
  - Phospho groups are somewhat labile during MS/MS, losses of 98 from precursor and product ions often observed

• Ficarro, S. B., McCleland, M. L., Stukenberg, P. T., Burke, D. J., Ross, M. M., Shabanowitz, J., Hunt, D. F., and White, F. M. (2002) *Nat Biotechnol* **20,** 301-5.
• Noble, M. E., Endicott, J. A., and Johnson, L. N. Protein Kinase Inhibitors: Insights into Drug Design from Structure. *Science* 2004,303**,** 1800-5.
• Kalume, D. E., Molina, H., and Pandey, A. Tackling the Phosphoproteome: Tools and Strategies. *Curr Opin Chem Biol* 2003,7**,** 64-9.

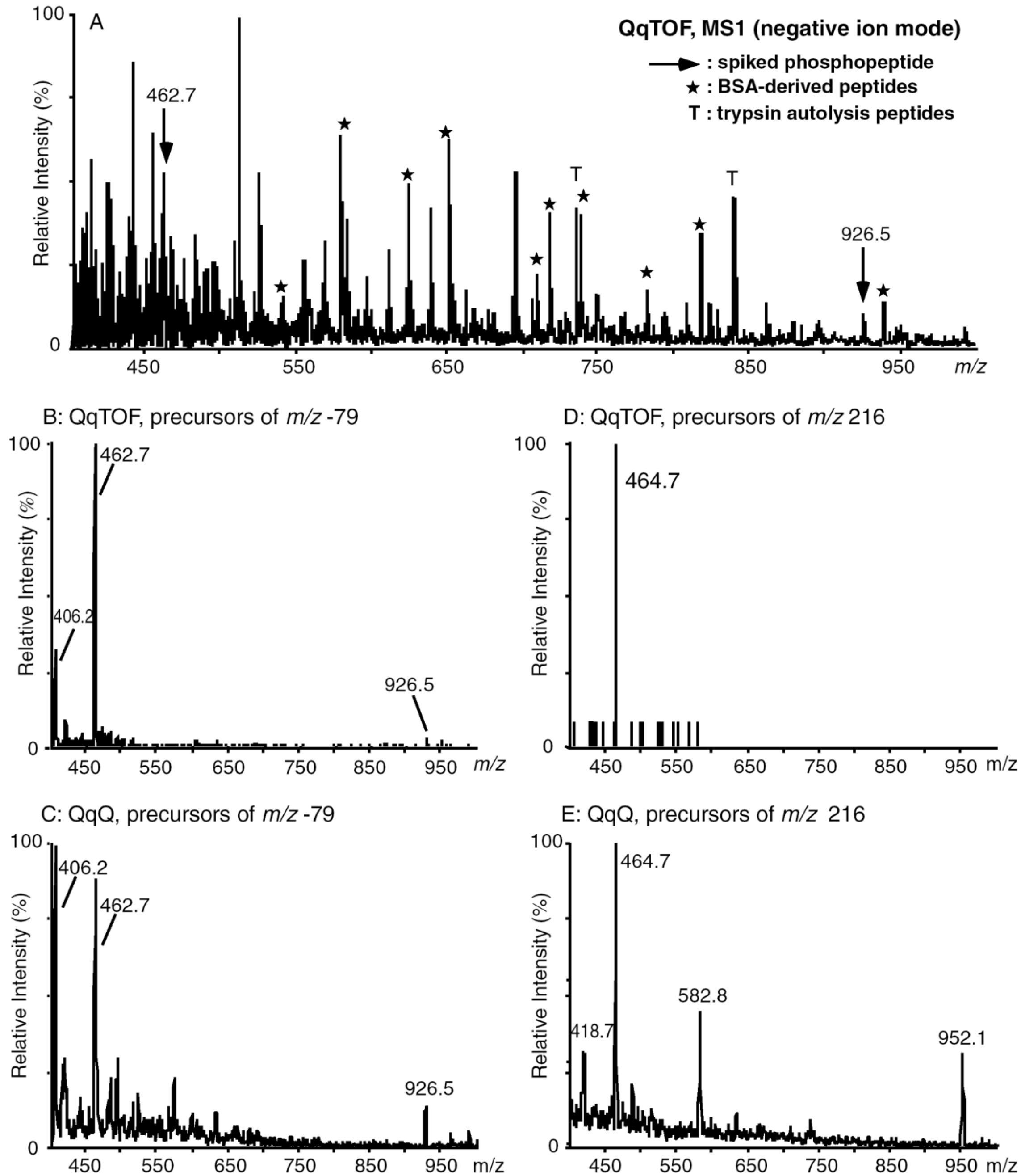
Affinity enrichment of phosphopeptides



[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.

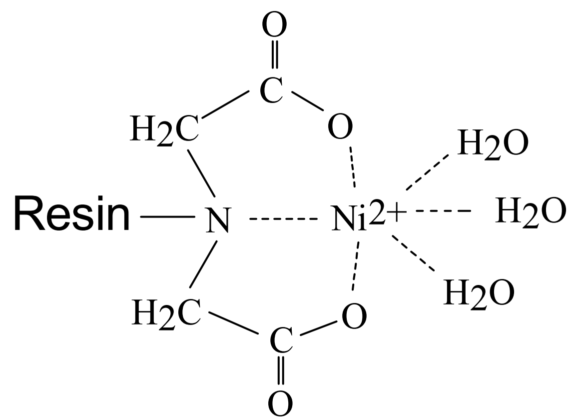
Precursor ion scanning for detection of phosphopeptides:

- Neg mode, detection of *m/z* 79 $(PO_3)^-$
- Pos mode detection of *m/z* 216, phosphotyrosine immonium ion
- Comparison of Q-oTOF vs QqQ.
- Q-oTOF has greater mass selectivity, more impt for *m/z* 216
- Q-oTOF is faster (LC/MS).

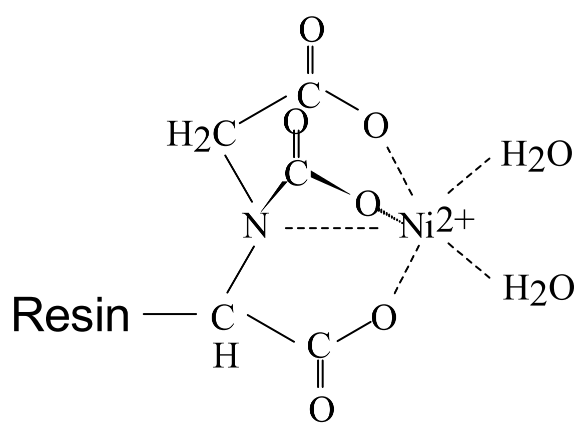


Steen, H.; Kuster, B.; Mann, M. *J Mass Spectrom* **2001**, *36*, 782-790.

Immobilized metal ion affinity chromatography: IMAC



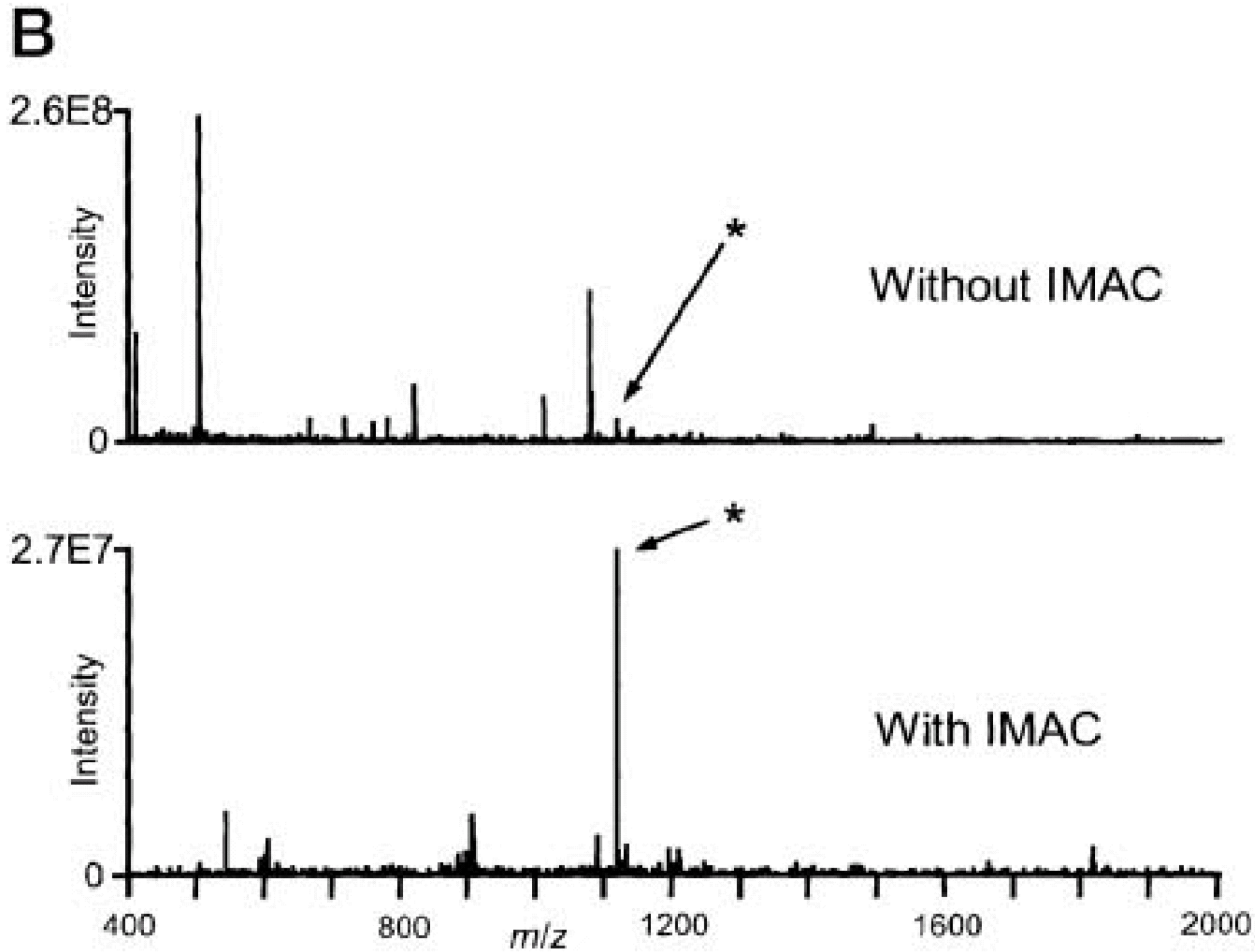
Nickel-IDA (Imino-DiAcetic acid) resin



Nickel-NTA (Nitrilo-TriAcetic acid) resin

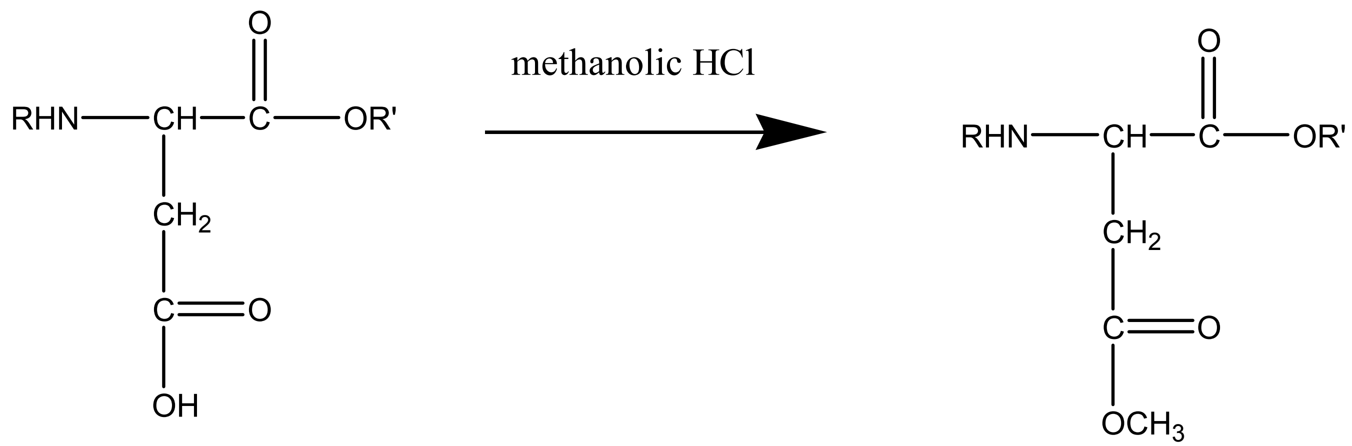
Common metals:
Ni^{2+}: His tags
Fe^{3+}: phosphopeptides

IMAC: immobilized metal affinity chromatography

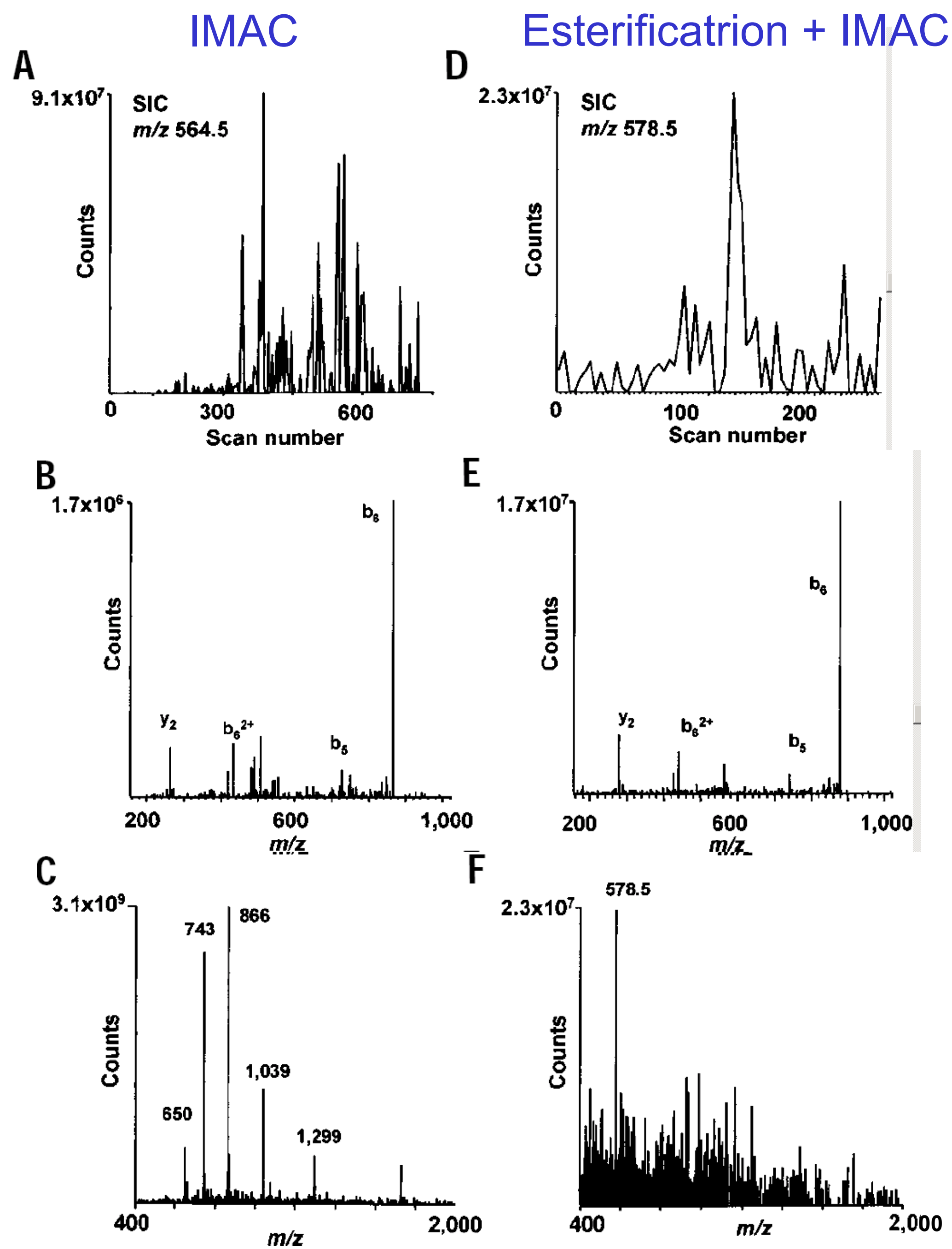


Gioeli, D.; Ficarro, S. B.; Kwiek, J. J.; Aaronson, D.; Hancock, M.; Catling, A. D.; White, F. M.; Christian, R. E.; Settlage, R. E.; Shabanowitz, J.; Hunt, D. F.; Weber, M. J. *J Biol Chem* **2002**, *277*, 29304-29314.

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.

LC/MS differential quantitation of phosphopeptides

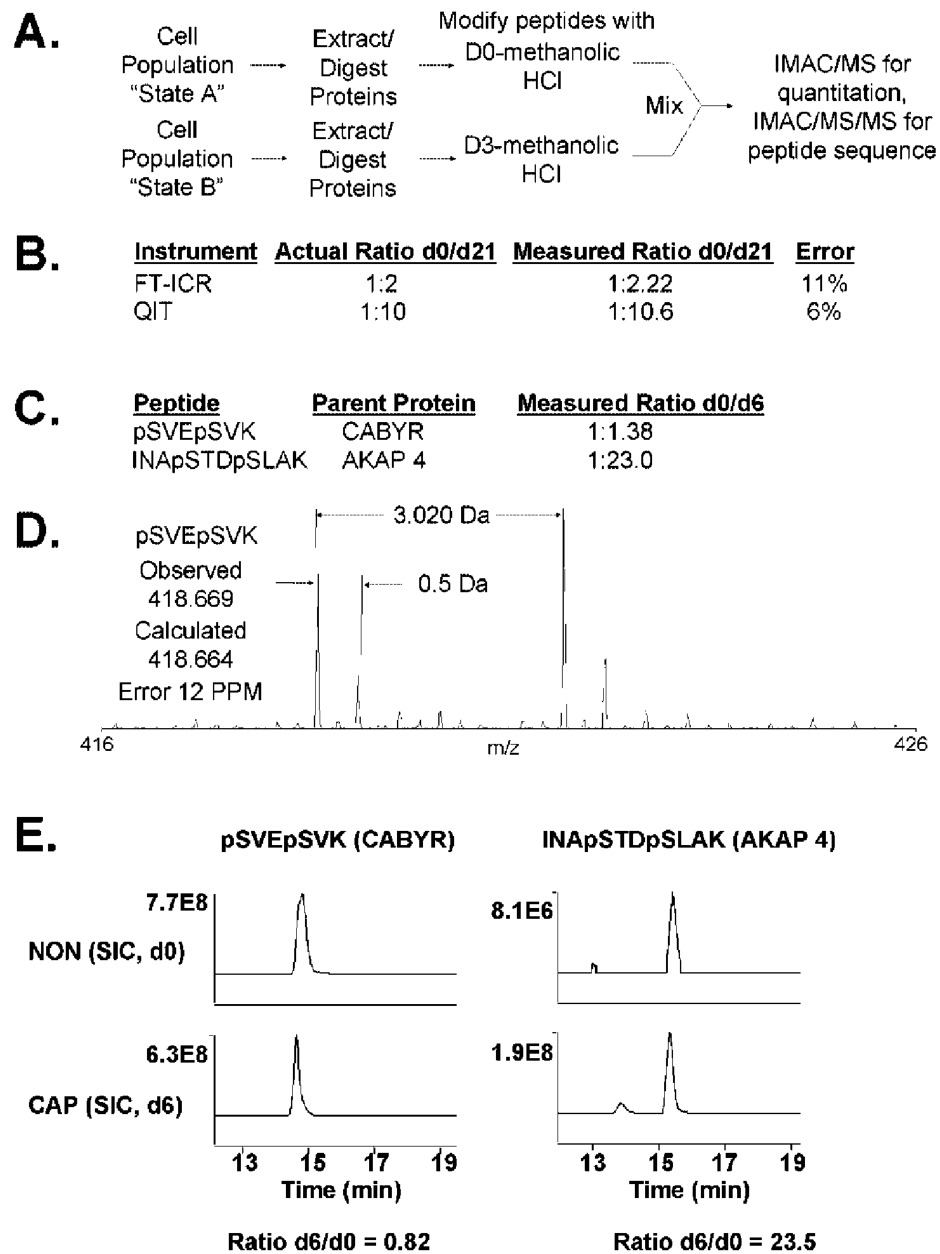
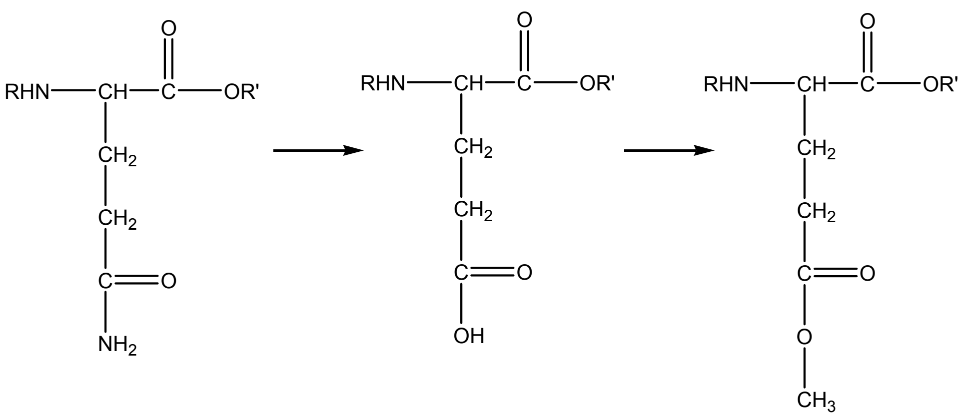


FIG. 3. **Quantitation of phosphorylated peptides in protein digests.** *A*, schematic representation of the procedure for phosphopeptide quantitation. *B*, ratios of deuterated and nondeuterated peptide FQpSEEQQQTEDELQDK present in trypsin digests of β-casein. *QIT*, quadrupole ion trap. *C*, capacitated and noncapacitated sperm total protein extracts measured using the approach described above. *D*, mass spectrum recorded during coelution of deuterated and nondeuterated pSVEpSVK peptides. *E*, single ion chromatograms (*SIC*; obtained by plotting the sum of the intensities of ions within a small mass window, *i.e.* 418.67 ± 0.1 Da, *versus* time) derived from MS analysis of differentially methyl ester-modified tryptic sperm peptides using an FT-ICR mass spectrometer. Double charged ions corresponding to deuterated (d_6) and nondeuterated (d_0) forms of peptides P1 (pSVEpSVK) and P2 (INApSTDpSLAK) were observed. Note that both peptides contain two carboxyl groups so that derivatized analogs differ in mass by 6 Da. Peptide sequences were derived from parallel MS/MS experiments performed on an ion trap mass spectrometer.

Ficarro, S.; Chertihin, O.; Westbrook, V. A.; White, F.; Jayes, F.; Kalab, P.; Marto, J. A.; Shabanowitz, J.; Herr, J. C.; Hunt, D. F.; Visconti, P. E. *J Biol Chem* **2003**, *278*, 11579-11589.

Side reactions in the methyl esterification procedure



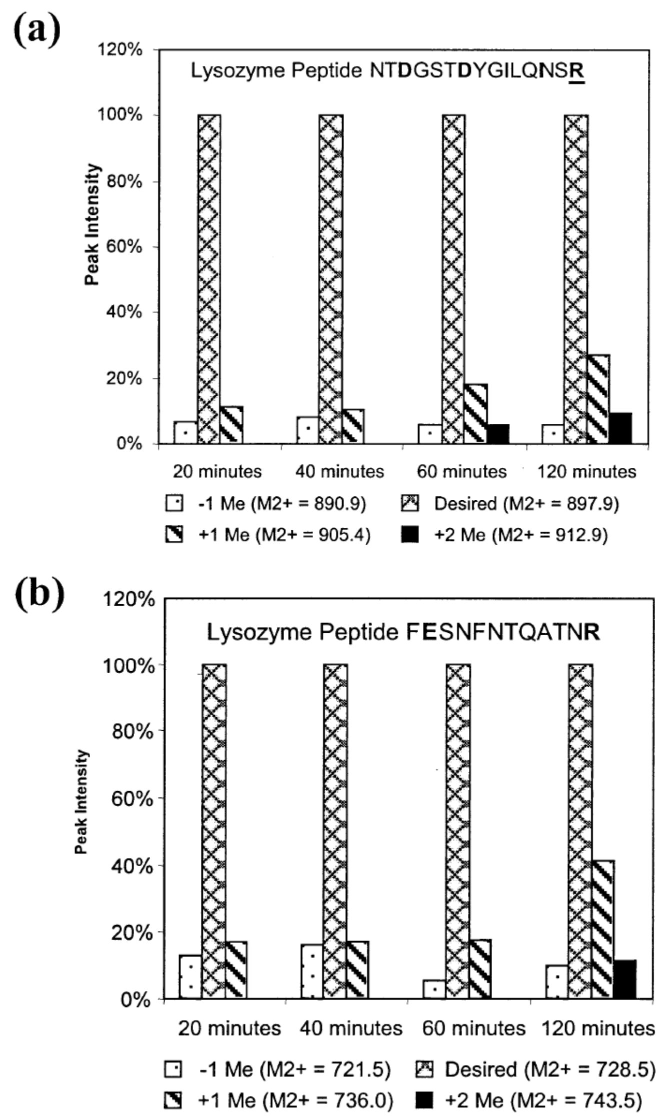


Figure 5. The relative intensity (by XIC of interest ions) of incomplete-reaction and side-reaction for the peptides (NT**D**GST**D**YGILQ<u>N</u>S**R**) (**a**) and (F**E**S<u>N</u>FNTQATN**R**) (**b**) with side-reaction on asparagine residues at different reaction times (20, 40, 60, and 120 min). 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

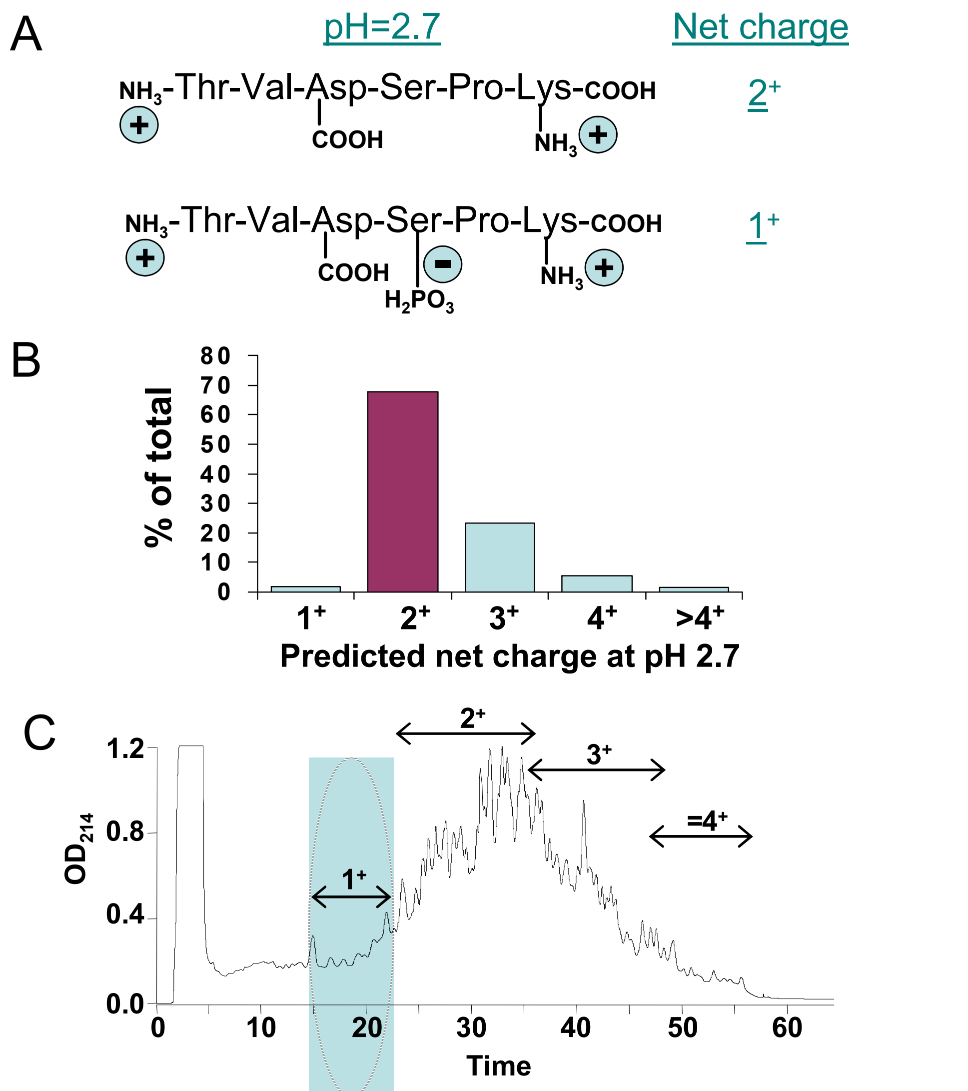
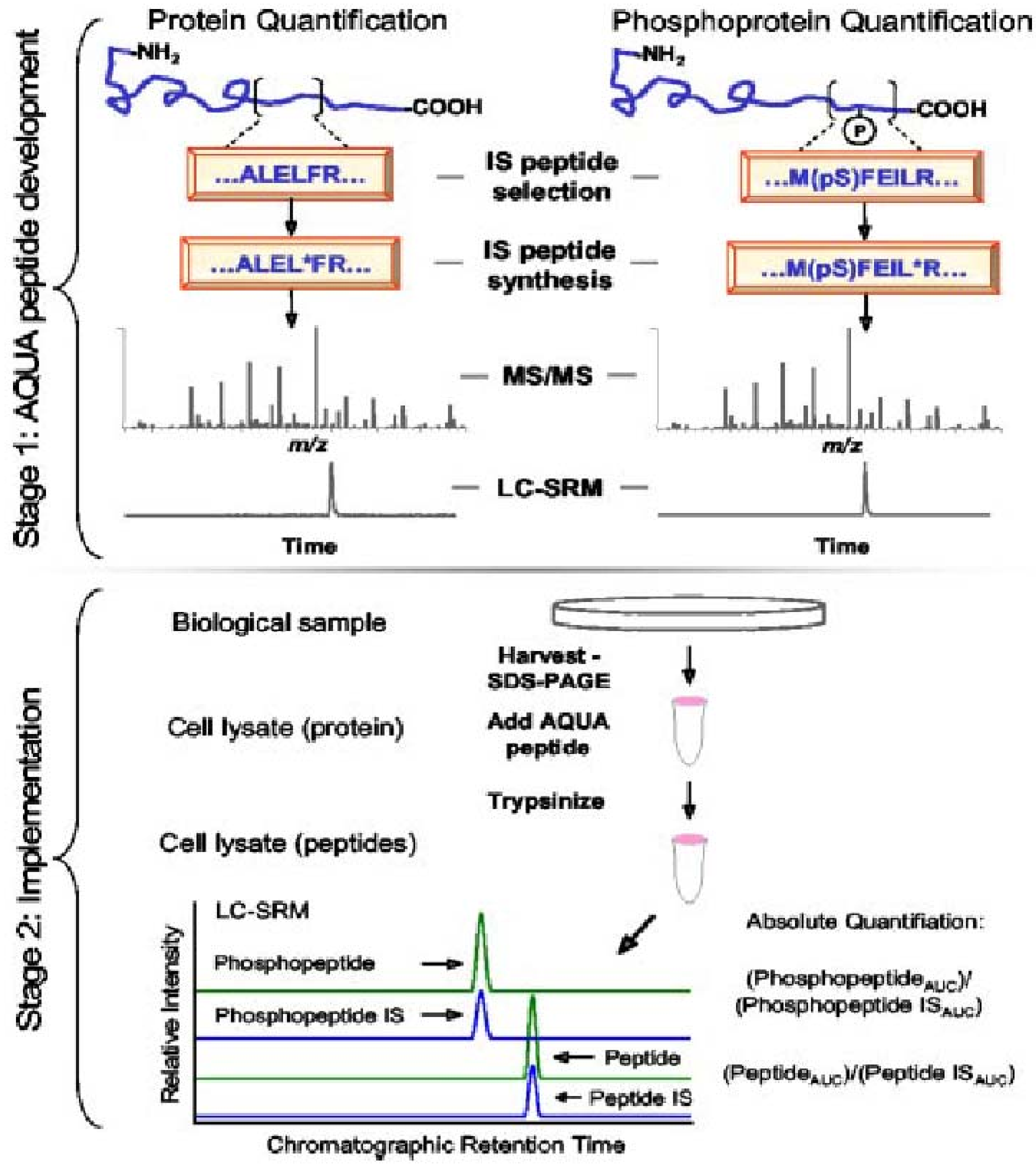


Figure 1. Scheme for phosphopeptide enrichment by strong cation exchange (SCX) chromatography. A) At pH 2.7, peptides produced by trypsin proteolysis generally have a solution charge state of 2^+ while phosphopeptides have a charge state of only 1^+. B) Solution charge state distribution of peptides (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†

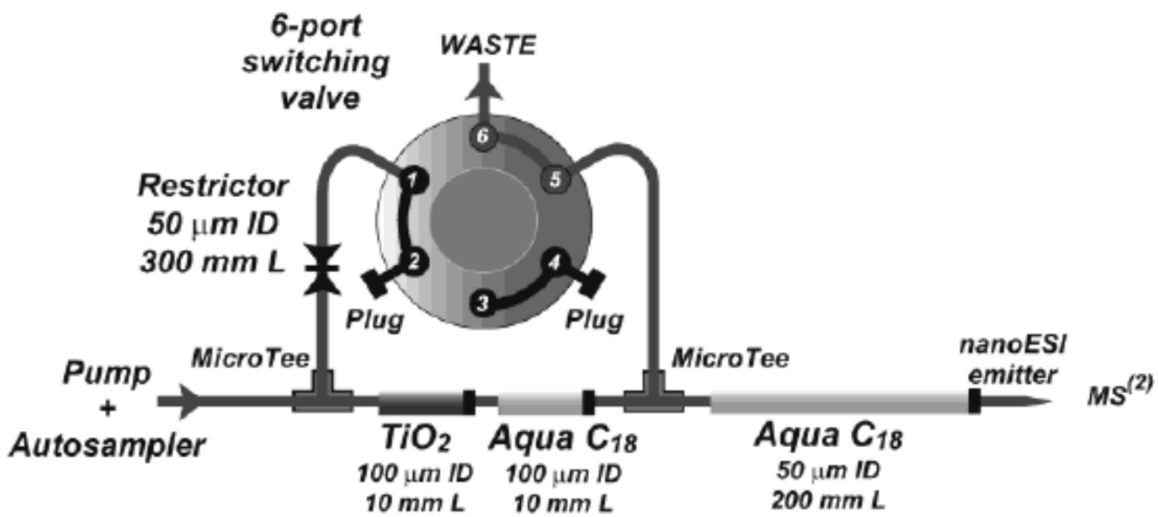


Figure 1. Schematic representation of the two-dimensional LC-MS setup. This setup comprises a six-port switching valve, a dual TiO_2/C_{18} precolumn, and a C_{18} analytical column. During sample loading from the injection loop of the autosampler the flow is 3 μL/min and the restrictor is closed. In 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†

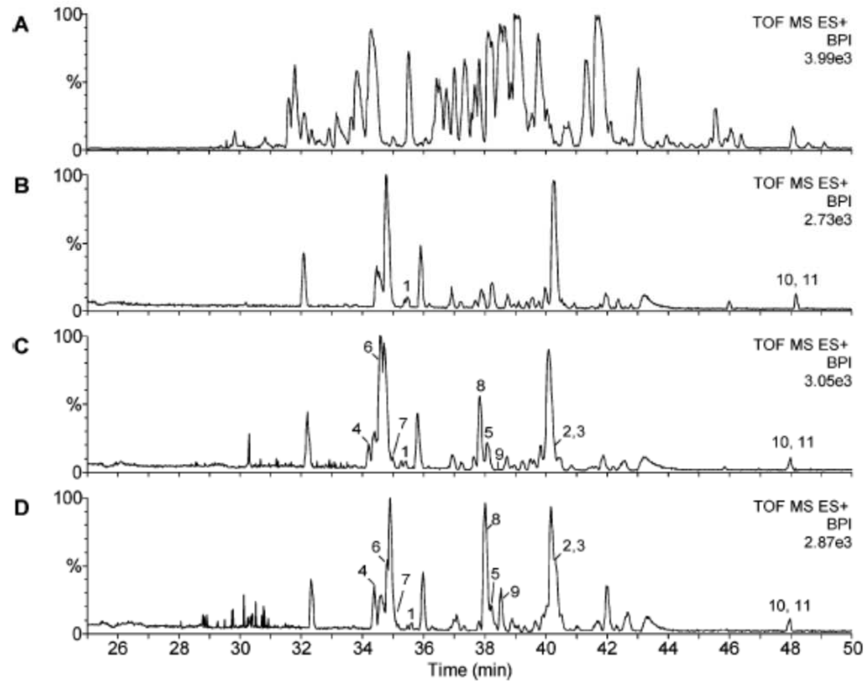


Figure 3. Nano-2D-LC-MS analysis of 250 fmol of tryptic PKG, phosphorylated to different degrees. Shown are the BPI elution profiles of (A) the first gradient of native (nonautophosphorylated) PKG and the second gradients of (B) native PKG, (C) partially autophosphorylated PKG, and (D) highly autophosphorylated 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

no.	residues	sequence [a]	mass [b] M_r found	mass [b] M_r calc	peptide observed in[c] non	partial	high
1	24−37	(K)-RLpSEKEEEIQELKR-(K)	1865.91	1865.92	+	+	+
2	42−56	(K)-C*QSVLPVPpSTHIGPR-(T)	1726.87	1726.82	−	++	++
3	42−56	(K)-C*QpSVLPVPpSTHIGPR-(T)	1806.86	1806.78	−	+	++
4	72−77	(R)-pSFHDLR-(Q)	853.33	853.35	−	+	++
5	60−77	(R)-AQGISASEPQTYRpSFHDLR-(Q)	2155.08	2154.99	−	+	−
6	57−71	(R)-TpTRAQGISAEPQTYR-(S)	1757.84	1757.81	−	++	+
7	57−71	(R)-TpTRAQGIpSAEPQTYR-(S)	1837.82	1837.78	−	+	+
8	57−77	(R)-TpTRAQGISAEPQTYRpSFHDLR-(Q)	2593.18	2593.15	−	+	++
9	57−77	(R)-TpTRAQGIpSAEPQTYRpSFHDLR-(Q)	2673.21	2673.11	−	+	++
10	514−532	(K)-TWpTFC*GTPEYVAPEIILNK-(G)	2318.21	2318.07	+	+	+
11	513−532	(K)-KTWpTFC*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.

Summary of phosphoproteomics enrichment methods

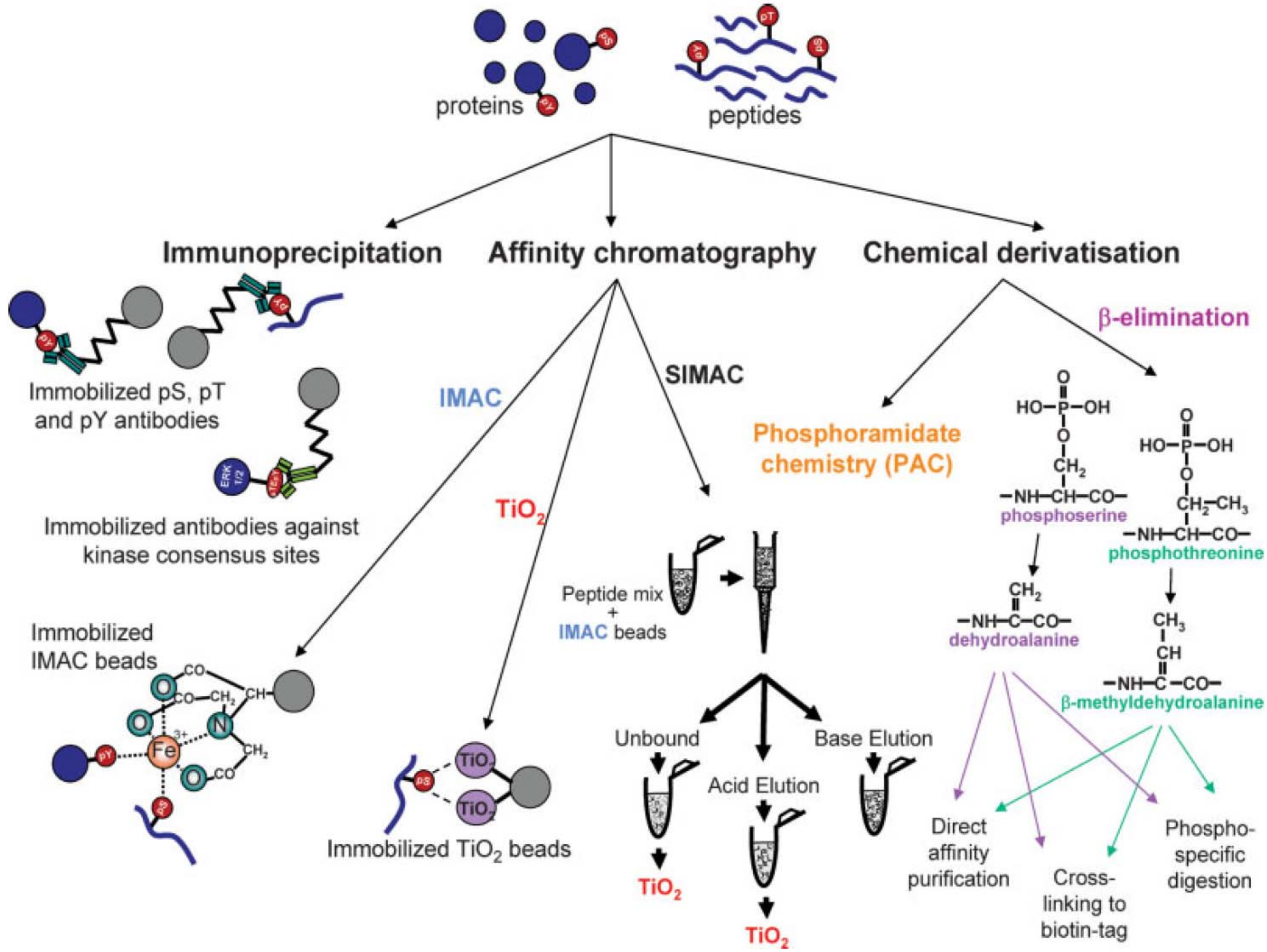


Figure 1. Strategies for phospho-specific enrichment. Most commonly used strategies for immunoprecipitation, affinity chromatography or chemical modification applied for enrichment of phosphoproteins and phosphopeptides are illustrated.

Thingholm, T. E., Jensen, O. N., and Larsen, M. R. (2009) Analytical strategies for phosphoproteomics, Proteomics 9, 1451-1468.

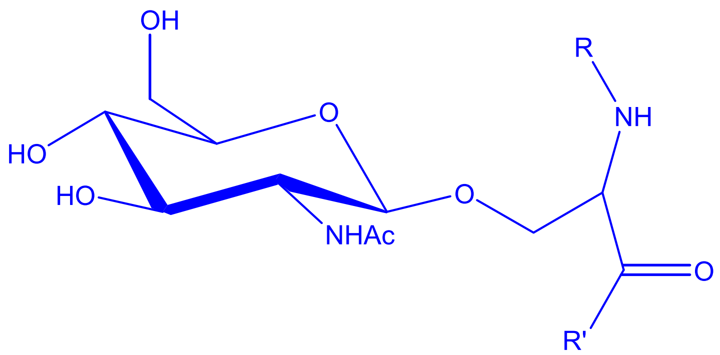
Ser/Thr *β-O*-GlcNAc Modification

- 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

BREAK

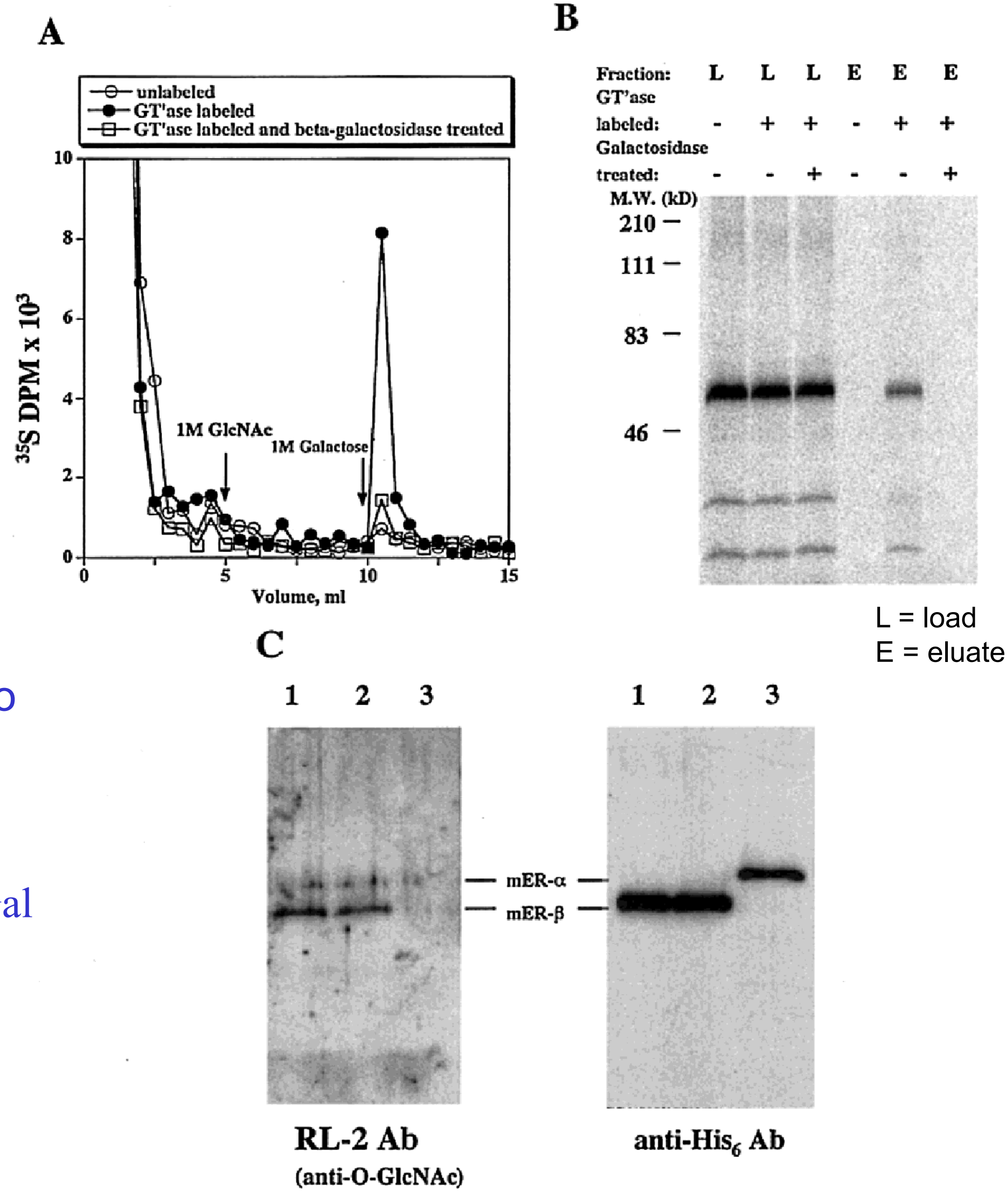
Wells, L.; Vosseller, K.; Cole, R. N.; Cronshaw, J. M.; Matunis, M. J.; Hart, G. W. *Mol Cell Proteomics* **2002**, *1*, 791-804.

Analysis of the β-O-GlcNAc-ome



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

Identification of *O*-GlcNAc sites in expressed murine estrogen receptor β

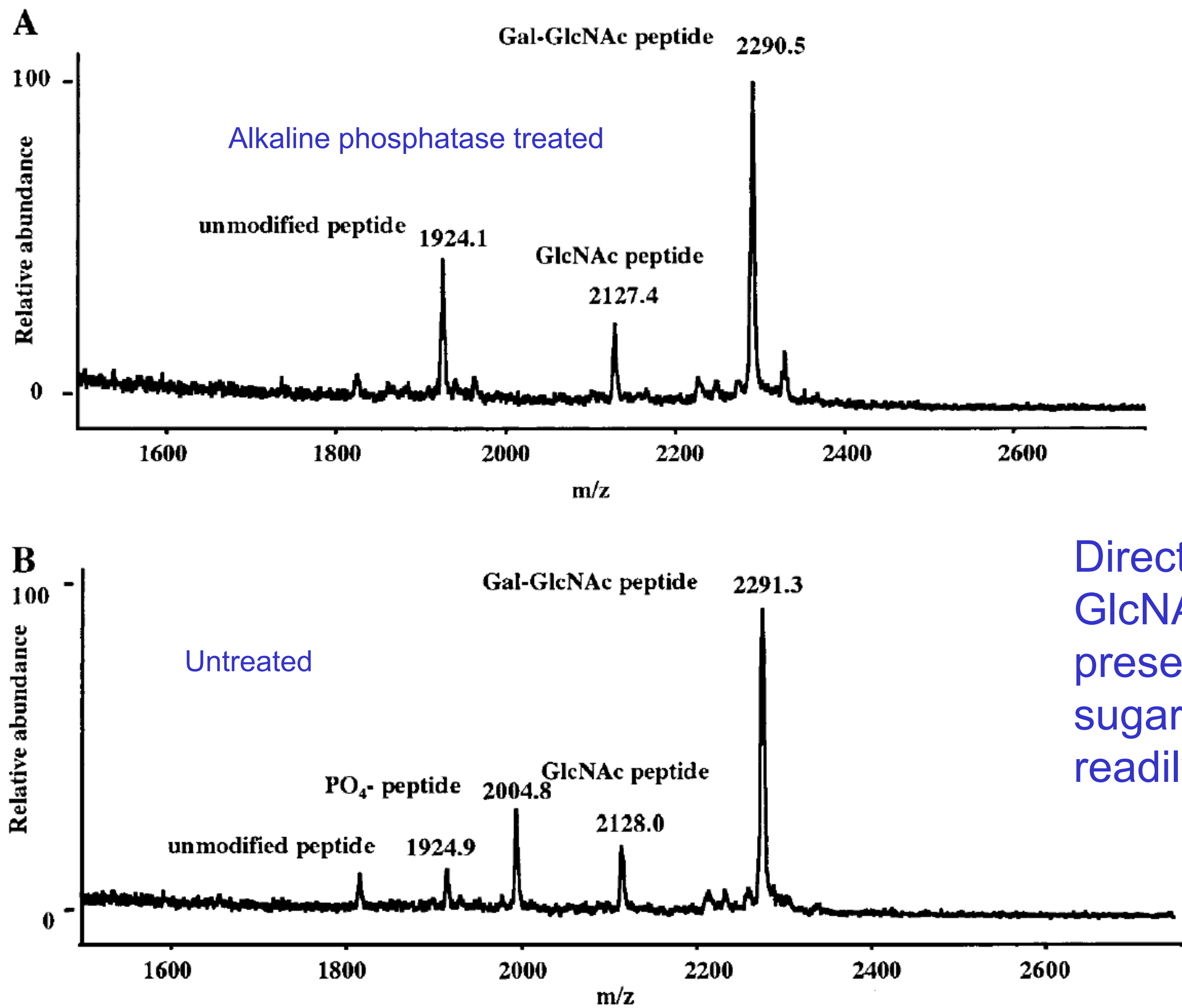


Gal-transferase adds Gal to GlcNAc-R to form Gal(β4)GlcNAc-R.

- RCA lectin can be used to specifically bind terminal Gal residues.
- ^{3}H-UDP-Gal can be used to radiolabel the GlcNAc-R peptides.

Cheng, X.; Cole, R. N.; Zaia, J.; Hart, G. W. *Biochemistry* **2000**, *39*, 11609-11620.

Profiling of reciprocal *O*-phosphorylation/*O*-GlcNAc modification of murine estrogen receptor β



Direct tandem MS of *O*-GlcNAc modified peptides presents a problem: the sugar group dissociates readily during CID.

Cheng, X.; Cole, R. N.; Zaia, J.; Hart, G. W. *Biochemistry* **2000**, *39*, 11609-11620.

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

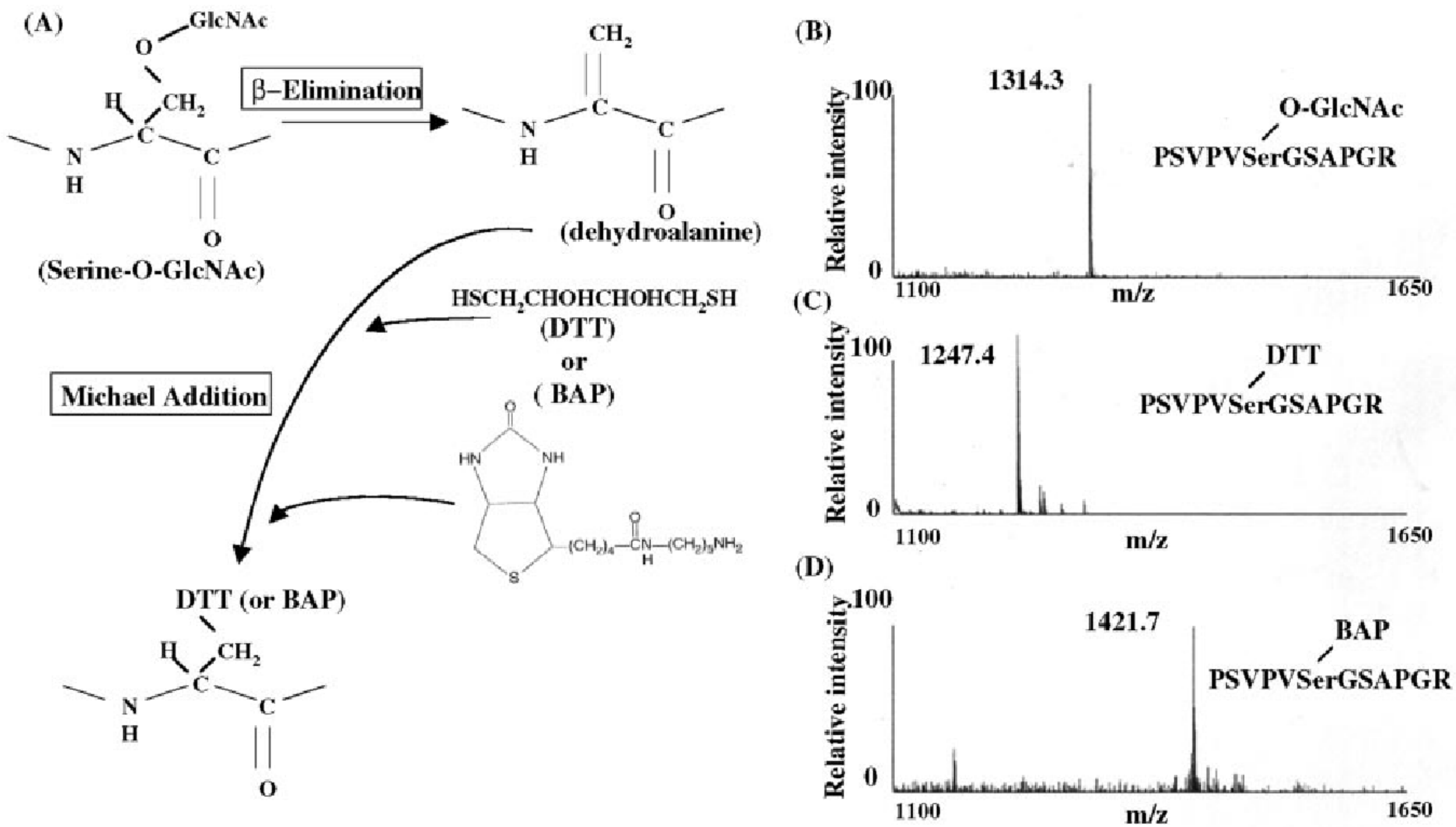
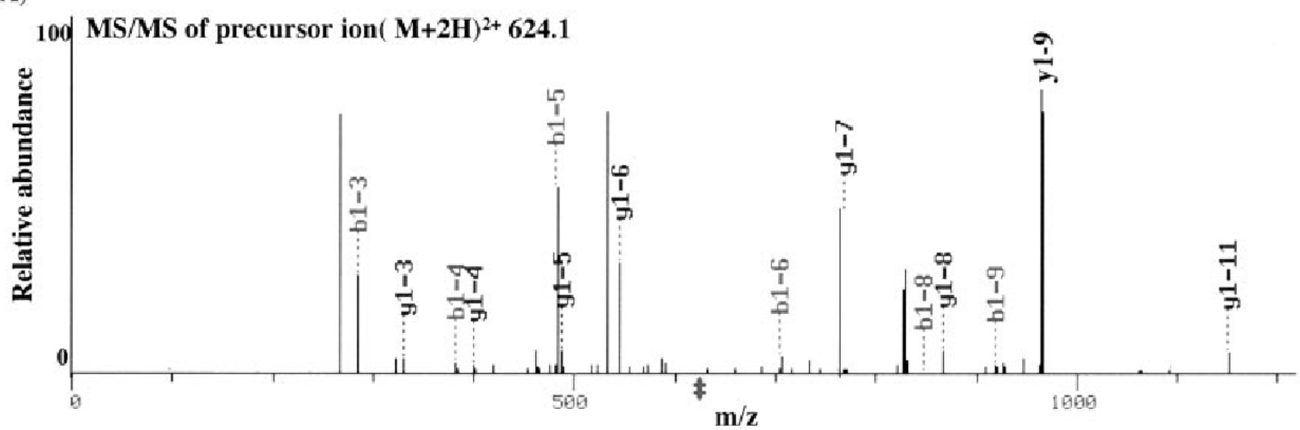


FIG. 2. **β-Elimination of *O*-GlcNAc and replacement with DTT (BEMAD) or BAP through Michael addition chemistry.** *A*, strategy for replacement of serine- or threonine-linked *O*-GlcNAc with the stable affinity tags DTT or BAP after β-elimination. *B*, *C*, and *D*, MALDI-TOF analysis of a synthetic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

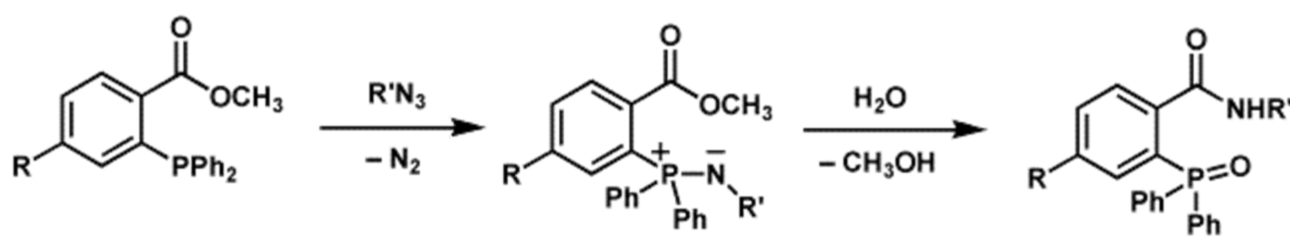
y ions	1247.1	**1150**	**1063**	**963.9**	**866.9**	**767.8**	**544.3**	**487.3**	**400.2**	**329.2**	232.1	175.1	- y ions
#	12	11	10	9	8	7	6	5	4	3	2	1	#
Peptide sequence	**P**	**S**	**V**	**P**	**V**	**S-DTT**	**G**	**S**	**A**	**P**	**G**	**R**	
#	1	2	3	4	5	6	7	8	9	10	11	12	
b ions	-98.1	**185.1**	**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 assignme of the site of modification unambiguous.

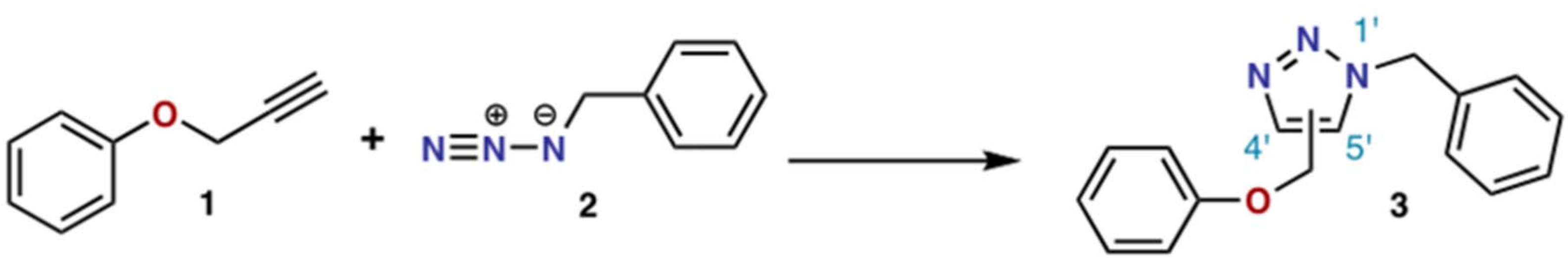
Wells, L.; Vosseller, K.; Cole, R. N.; Cronshaw, J. M.; Matunis, M. J.; Hart, G. W. *Mol Cell Proteomics* **2002**, *1*, 791-804.

Organic chemical ligation reactions

Staudinger ligation:



Click chemistry:



Chemoenzymatic enrichment of *O*-GlcNAc-modified peptides

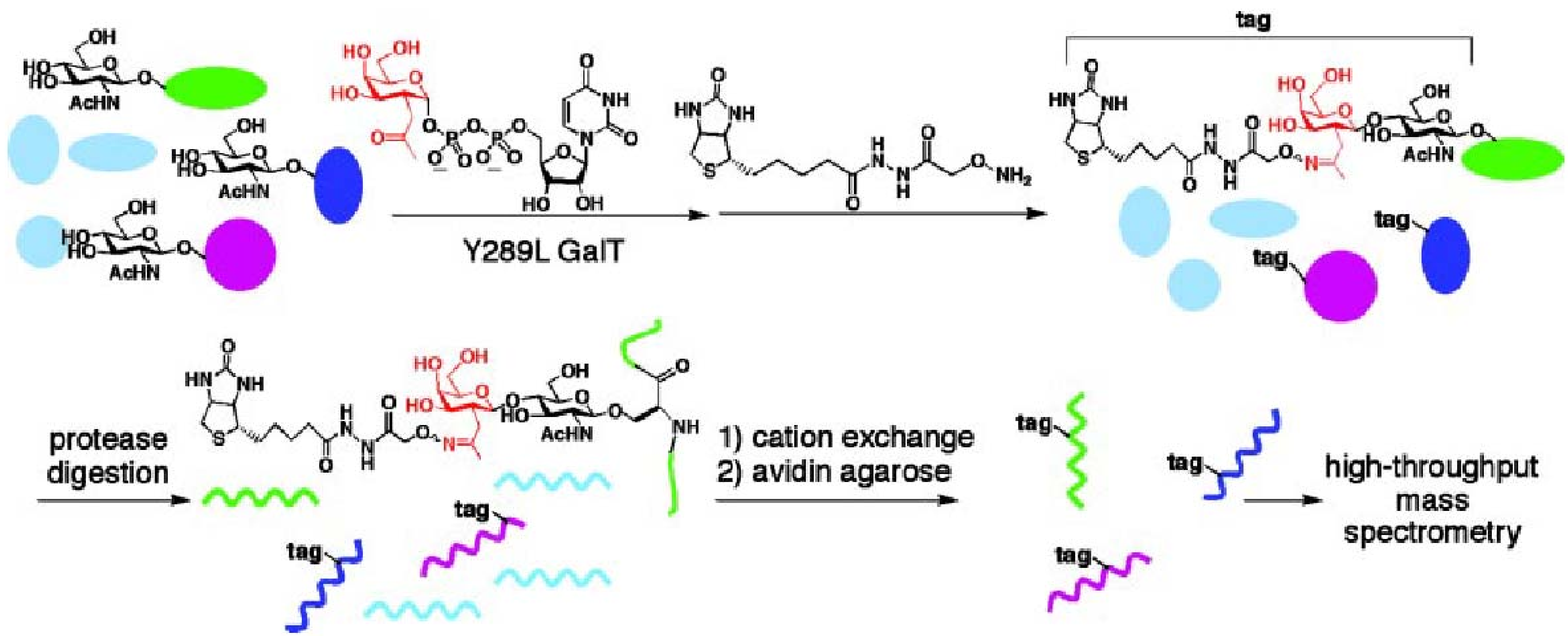
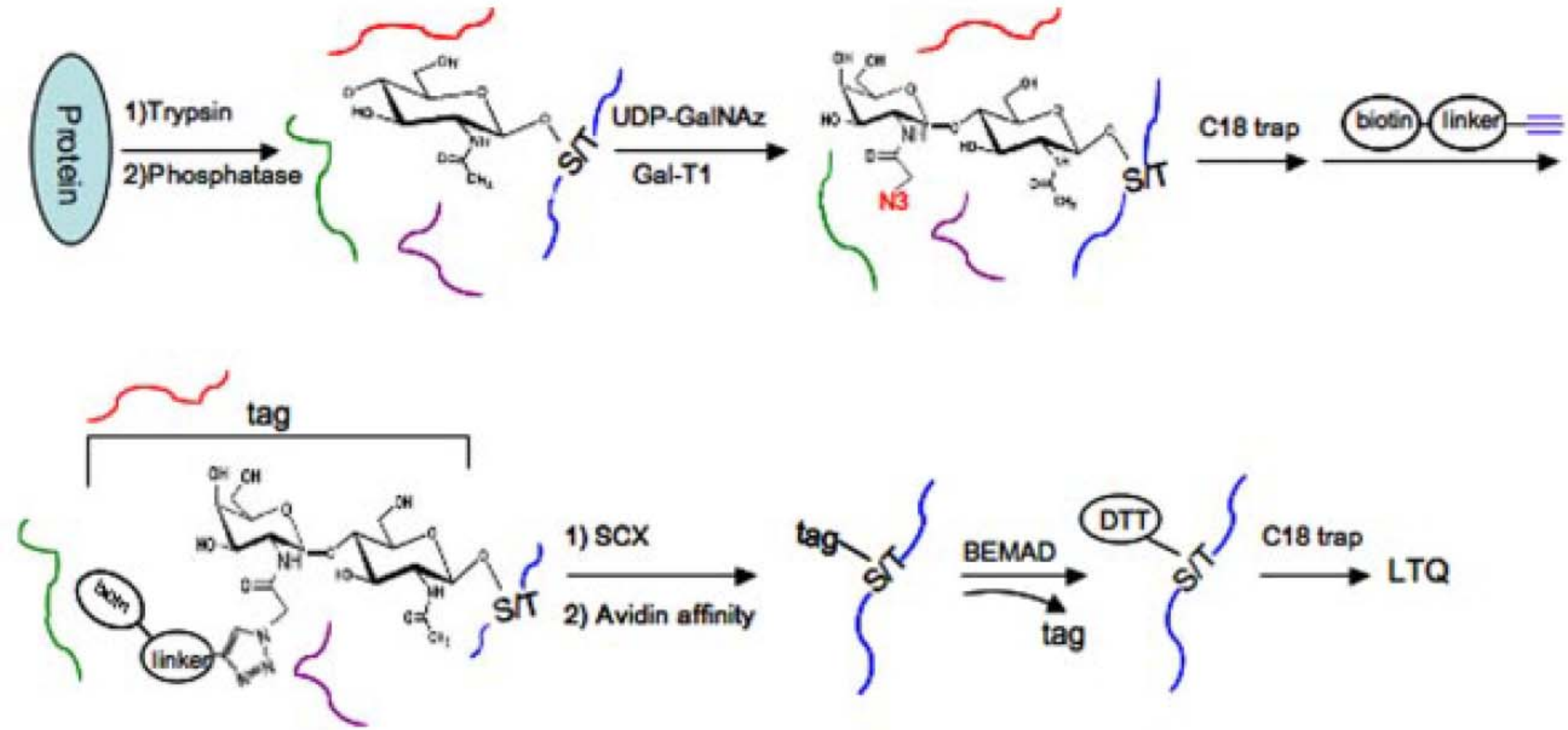


Fig. 1. Chemoenzymatic strategy for identifying *O*-GlcNAc-glycosylated proteins from cellular lysates.

Khidekel, N., et al. (2004) *PNAS* **101,** 13132-13137.

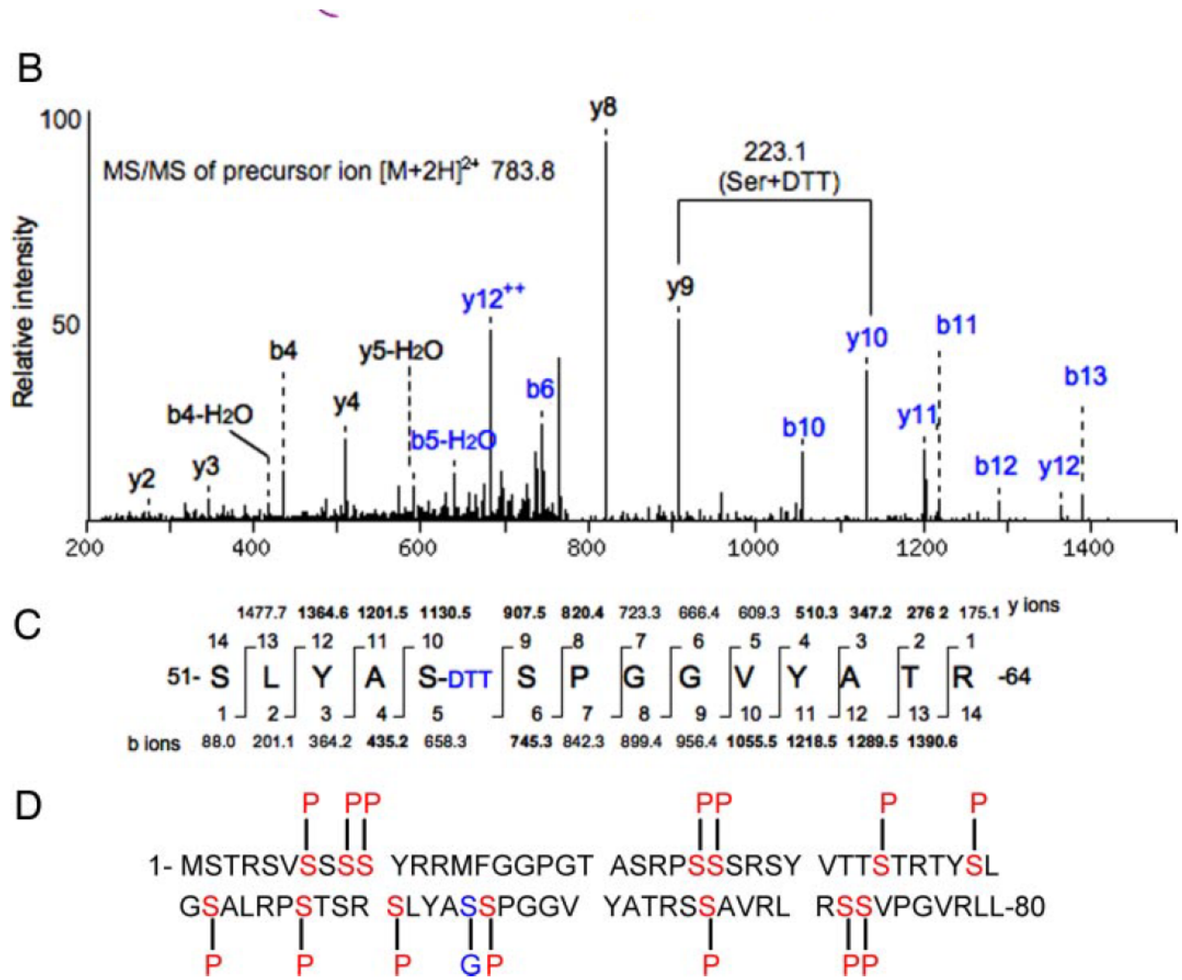
A combination of immunoaffinity enrichment, chemo-enzymatic labeling, affinity enrichment, BEMAD for site-specific determination of O-GlcNAc site in the midst of a phosphorylation domain.



Wang, Z., Pandey, A., and Hart, G.W. (2007). Mol Cell Proteomics *6*, 1365-1379.

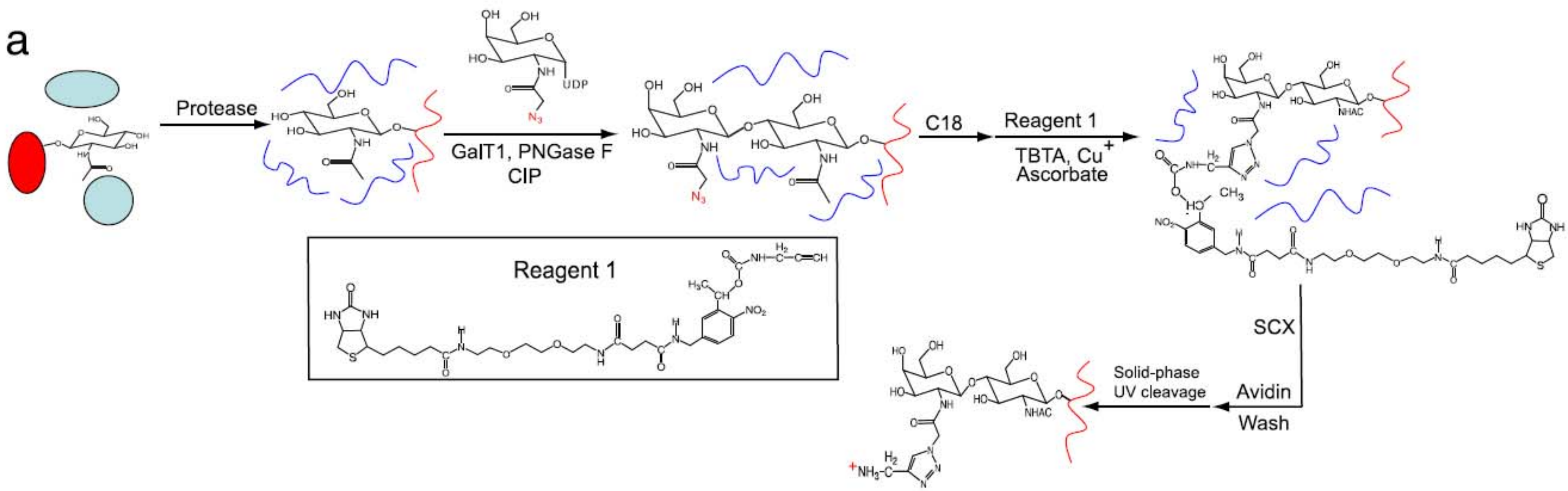
A combination of immunoaffinity enrichment, chemo-enzymatic labeling, affinity enrichment, BEMAD for site-specific determination of O-GlcNAc site in the midst of a phosphorylation domain.

FIG. 5. **Mapping *O*-GlcNAc site(s) on vimentin.** *A*, schematic for purification of *O*-GlcNAc-modified peptides and site mapping. *B*, MS/MS fragmentation of an *O*-GlcNAc-modified vimentin peptide (*m/z* 783.8 $[M + 2H]^{2+}$). *b* and *y* ions carrying DTT are shown in *blue*. *C*, summary of fragmentation patterns. Detected *b* and *y* ions are shown in *bold*. The *O*-GlcNAc site is unambiguously assigned to Ser-55. *D*, sequence of the human vimentin head domain (amino acids 1–80) showing known phosphorylation (43) and *O*-GlcNAc sites. *P*, known phosphorylation sites; *G*, *O*-GlcNAc site; *SCX*, strong cation exchange.



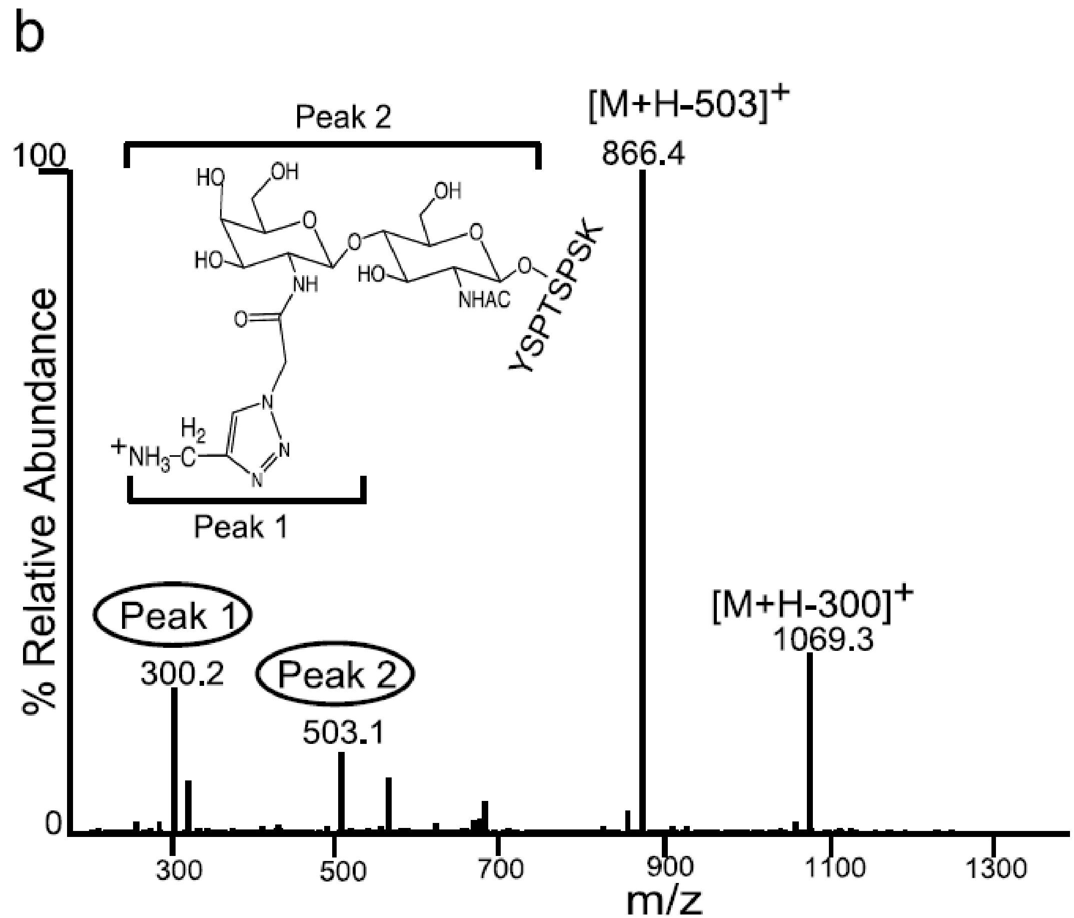
Wang, Z., Pandey, A., and Hart, G.W. (2007). Mol Cell Proteomics *6*, 1365-1379.

Metabolic labeling and tagging with photocleavable biotin



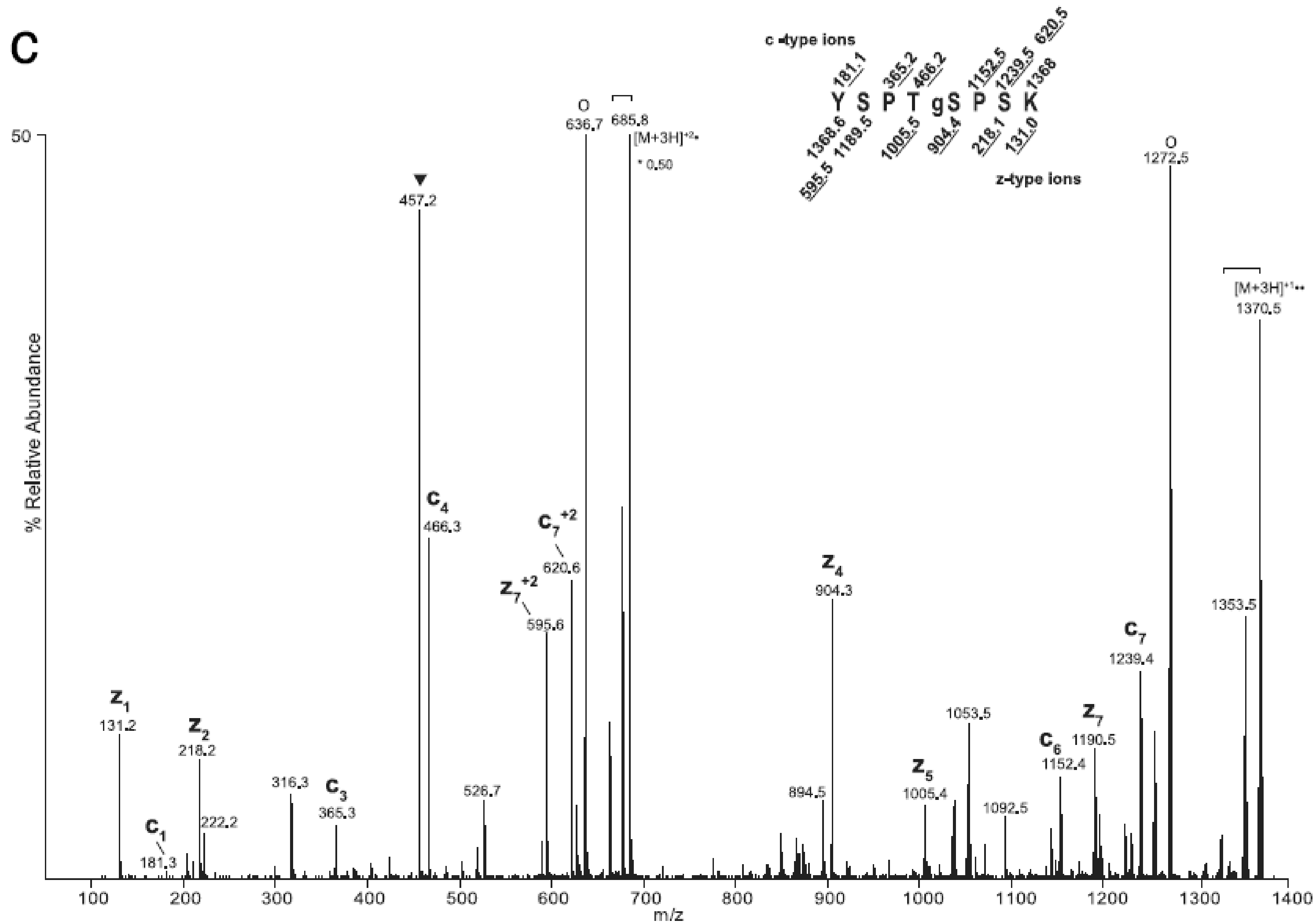
Wang, Z.; Udeshi, N. D.; O'Malley, M.; Shabanowitz, J.; Hunt, D. F.; Hart, G. W. Mol Cell Proteomics 2010, 9, 153-160.

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

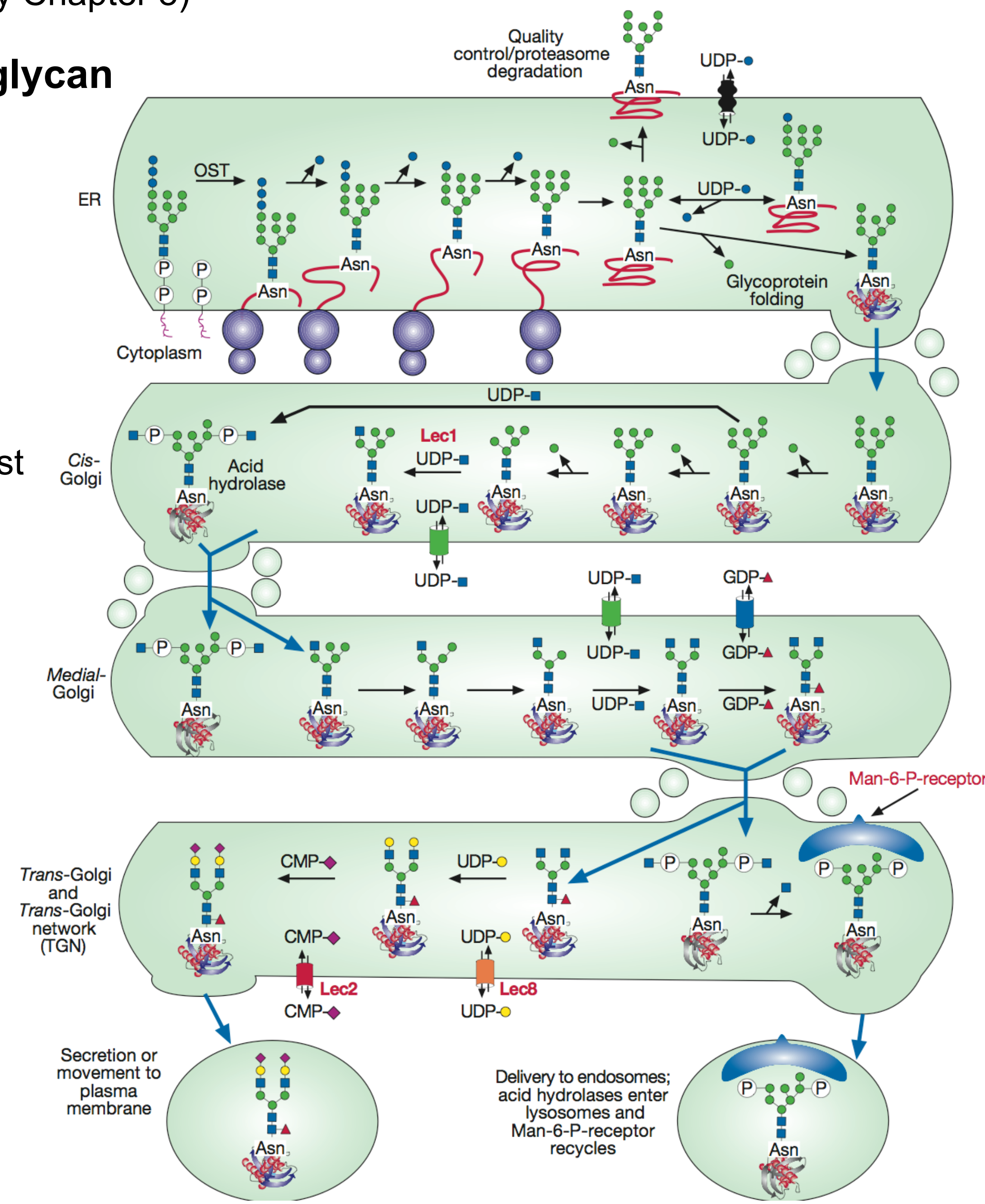


Wang, Z.; Udeshi, N. D.; O'Malley, M.; Shabanowitz, J.; Hunt, D. F.; Hart, G. W. Mol Cell Proteomics 2010, 9, 153-160.

N-glycosylation (Essentials of Glycobiology Chapter 8)

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

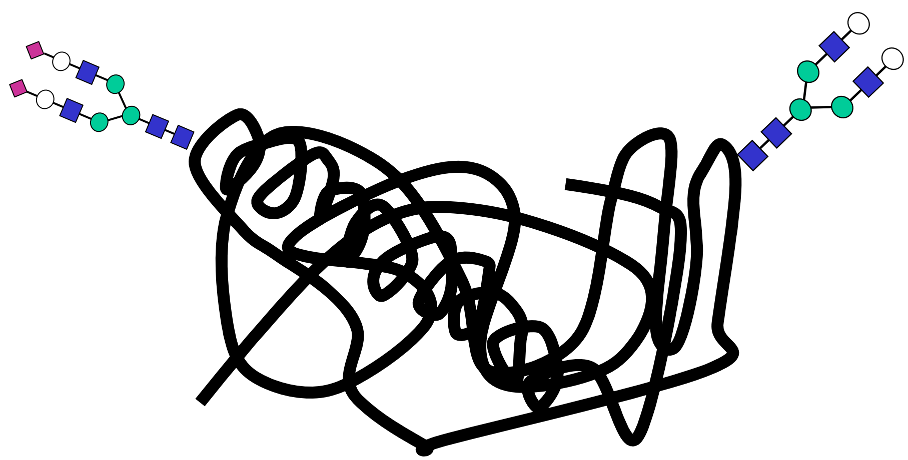


Essentials of Glycobiology
Second Edition

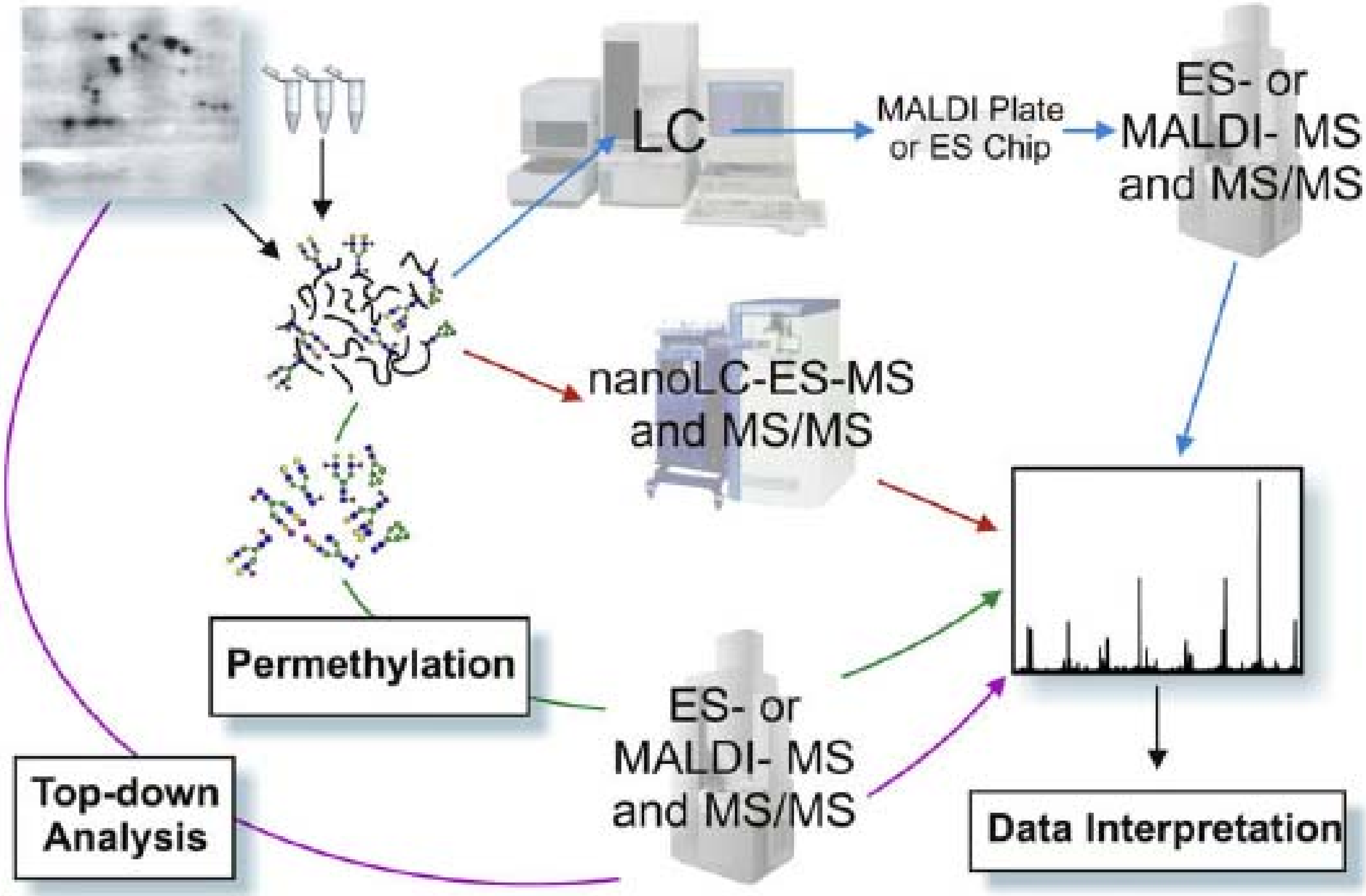
Chapter 8, Figure 4

Glycoproteomics:

A glycoprotein

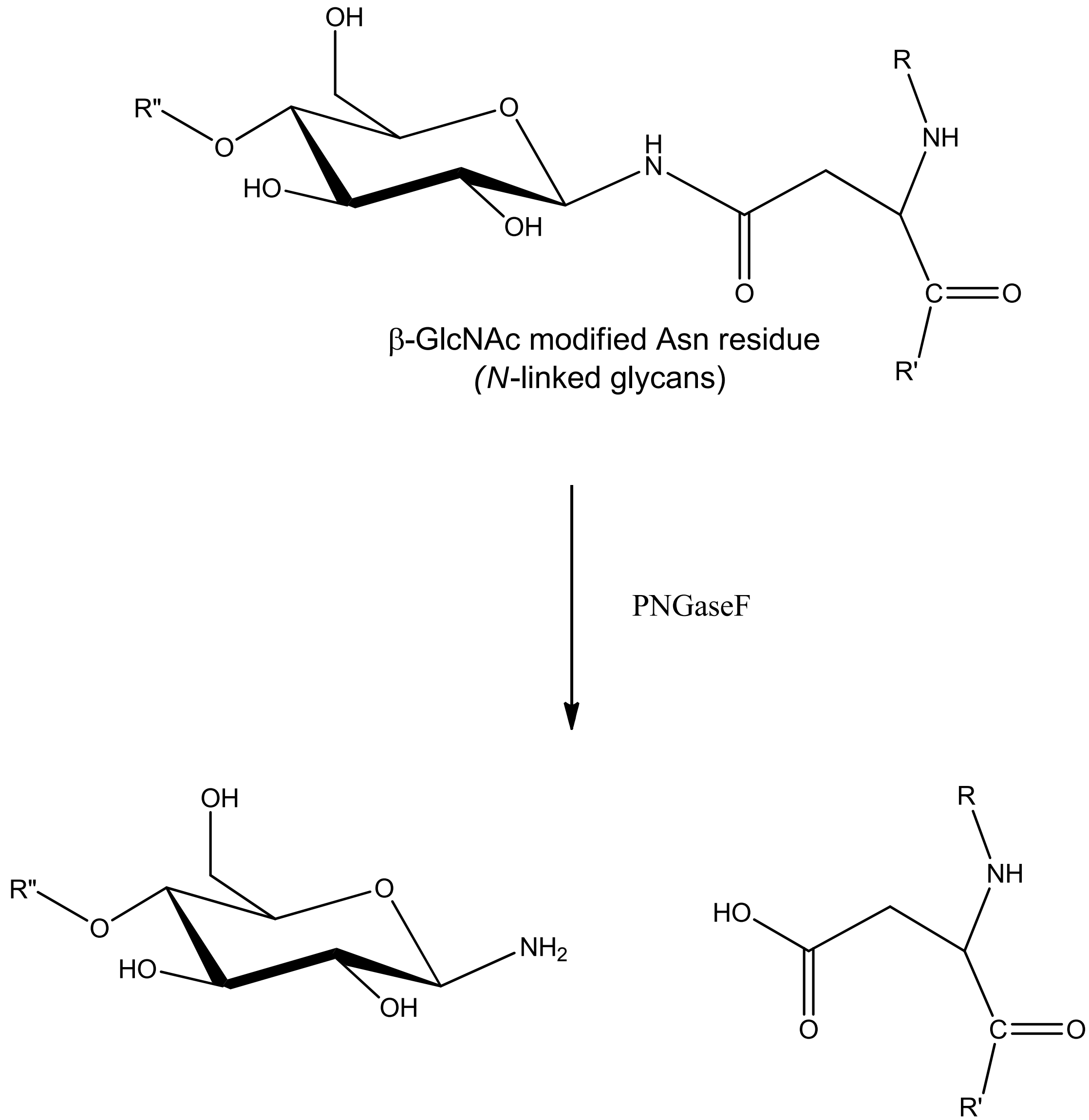


Glycoproteomics overview

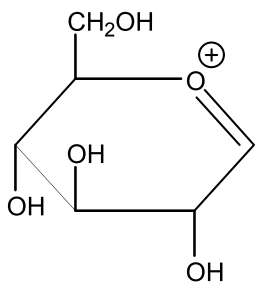


Tissot, B.; North, S. J.; Ceroni, A.; Pang, P. C.; Panico, M.; Rosati, F.; Capone, A.; Haslam, S. M.; Dell, A.; Morris, H. R. FEBS Lett 2009, 583, 1728-1735.

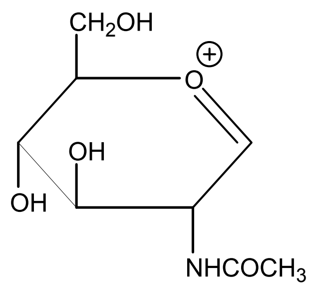
Peptide N-glycosidase F (PNGase F)



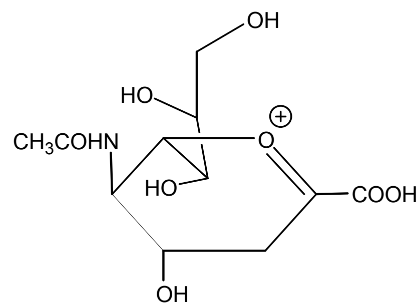
Common glycan oxonium ions



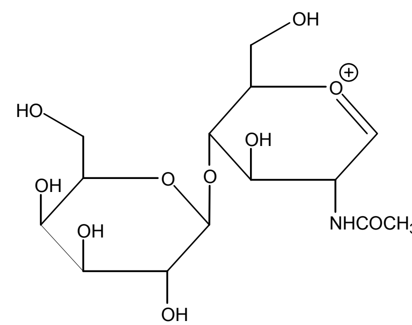
Hex
$C_6H_{11}O_5^+$
m/z 163.06



HexNAc
$C_8H_{14}NO_5^+$
m/z: 204.09



Neu5Ac
$C_{11}H_{18}NO_8^+$
m/z 292.10

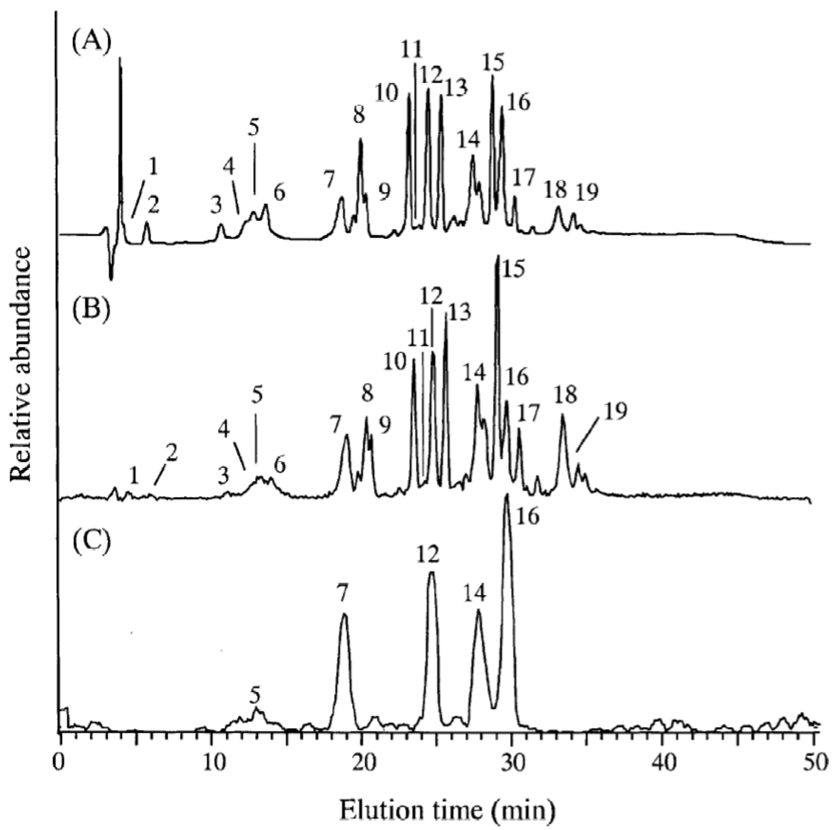


HexHexNAc
$C_{14}H_{24}NO_{10}^+$
m/z 366.14

- 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

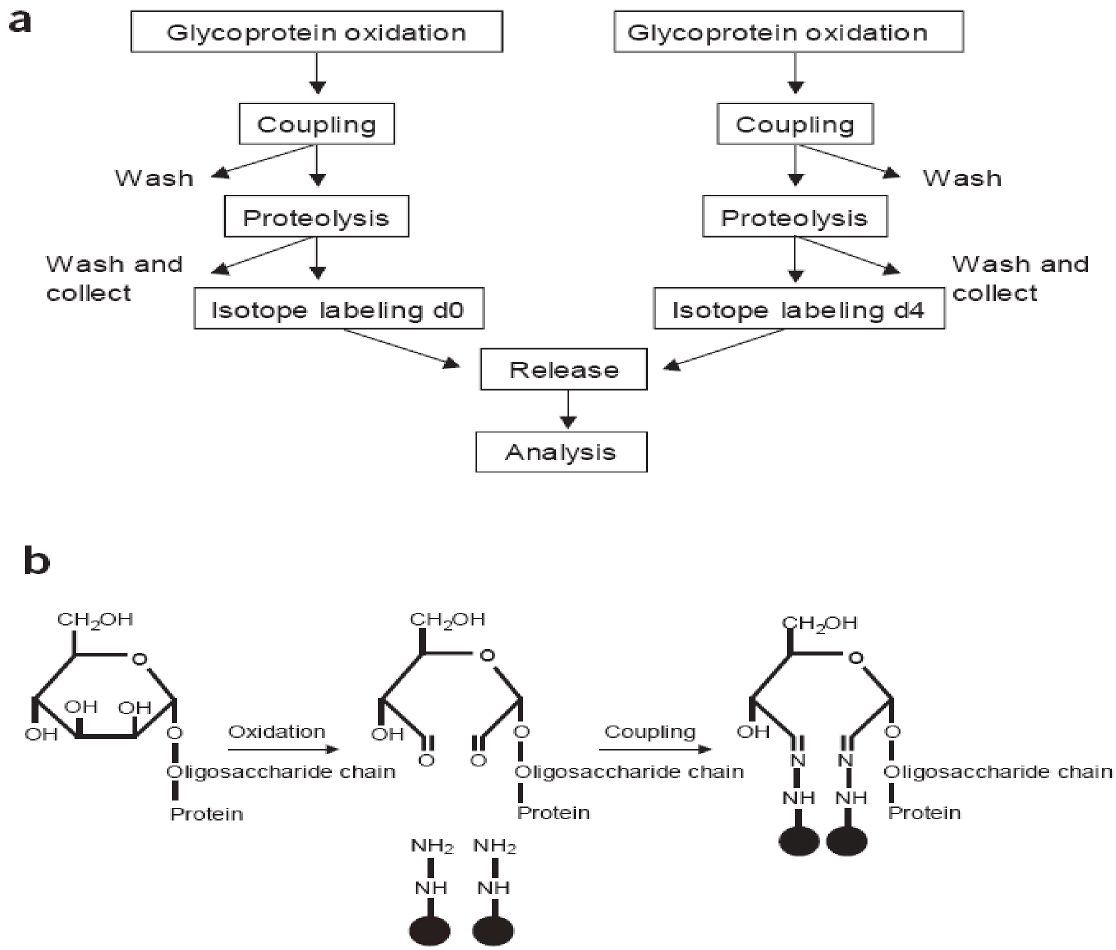


Figure 1 Schematic diagram of quantitative analysis of N-linked glycopeptides. (**a**) Strategy for quantitative analysis of glycopeptides. Proteins from two biological samples are oxidized and coupled to hydrazide resin. Nonglycosylated peptides are removed by proteolysis and extensive washes. The nonglycosylated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

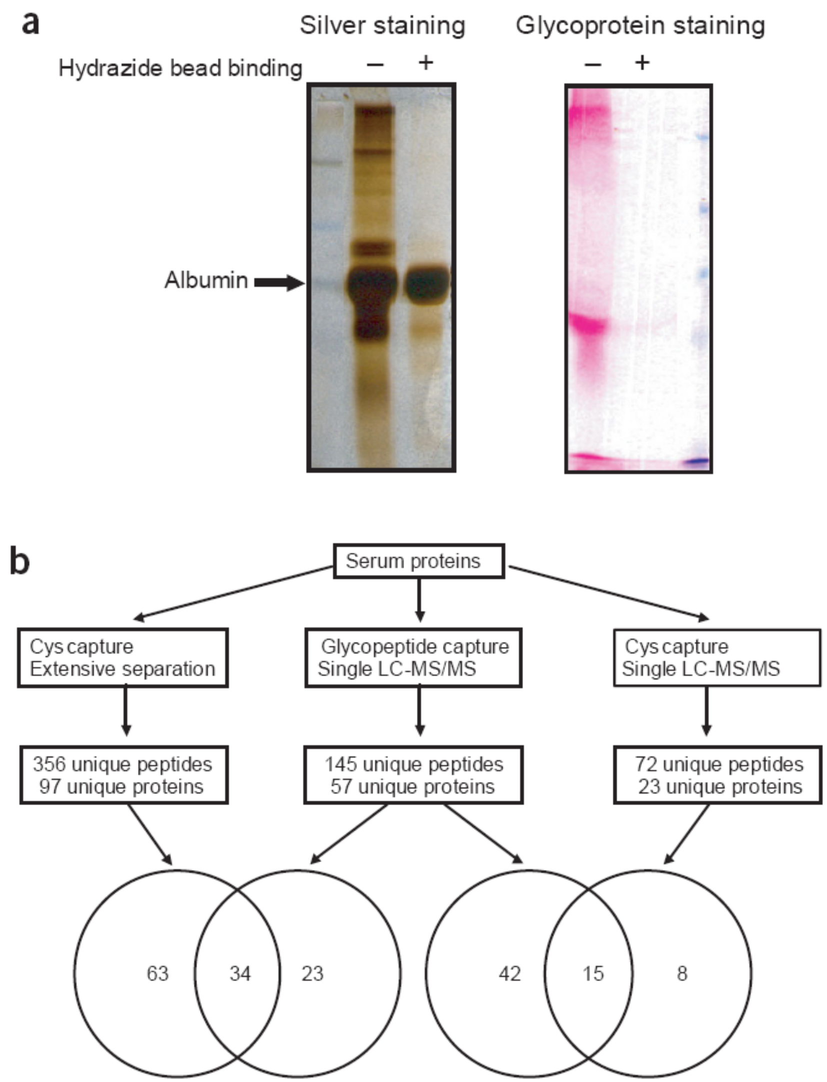


Figure 2 Isolation of glycoproteins from serum. (**a**) Specific capture of glycoproteins by hydrazide resin. Total protein staining or glycoprotein staining of serum before (–) and after (+) capture of glycoproteins by hydrazide resin. Proteins were separated by SDS-PAGE and stained with silver (left) or Gel Code Blue glycoprotein staining reagent (right). (**b**) Comparison of total number of proteins or peptides identified from serum samples isolated by either cysteine-reactive tag or glycopeptide capture method.

Glycopeptide enrichment increases efficiency or serum analysis by MS

Zhang Nat Biotech 2003 21:660-666

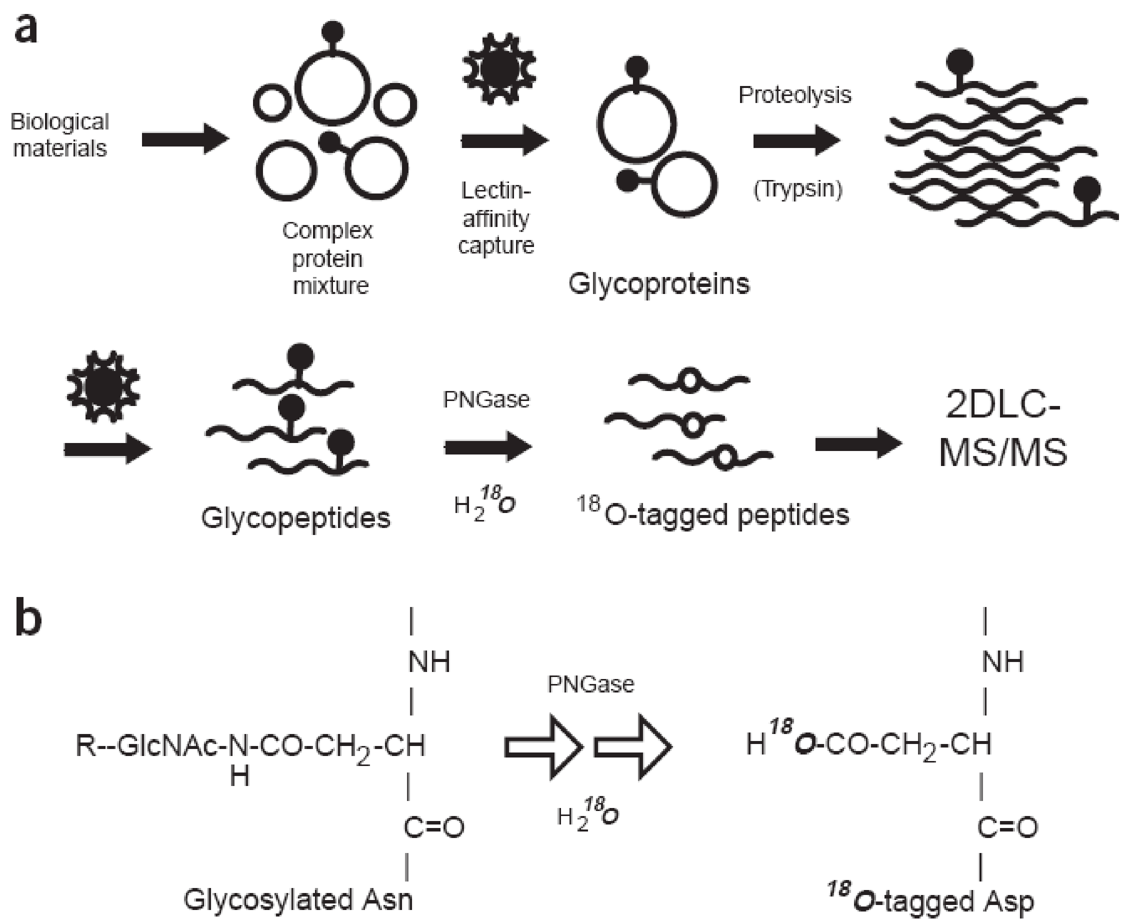
Lectins for affinity enrichments of glycoproteins/glycopeptides

- ~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.

IGOT: Isotope-coded glycosylation-site-specific tagging



ConA lectin

Figure 1 Schematic representation of the IGOT strategy. (**a**) A subset of glycoproteins is collected by lectin-affinity chromatography from a biological mixture. After tryptic digestion of the glycoproteins, the glycopeptides are recovered by the same lectin and then are treated with peptide-N-glycosidase in $H_2{}^{18}O$ to remove the sugar moieties and to incorporate the 'glycosylation-site-specific' tag. The peptides are finally analyzed by the multi-dimensional LC-MS-based protein identification technology. (**b**) PNGase-mediated conversion of the glycosylated asparagine to aspartic acid and concomitant incorporation of ^{18}O from the solvent $H_2{}^{18}O$.

Kaji H, et al. (2003) Nature Biotechnology 21:667-672.

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 ($$$)

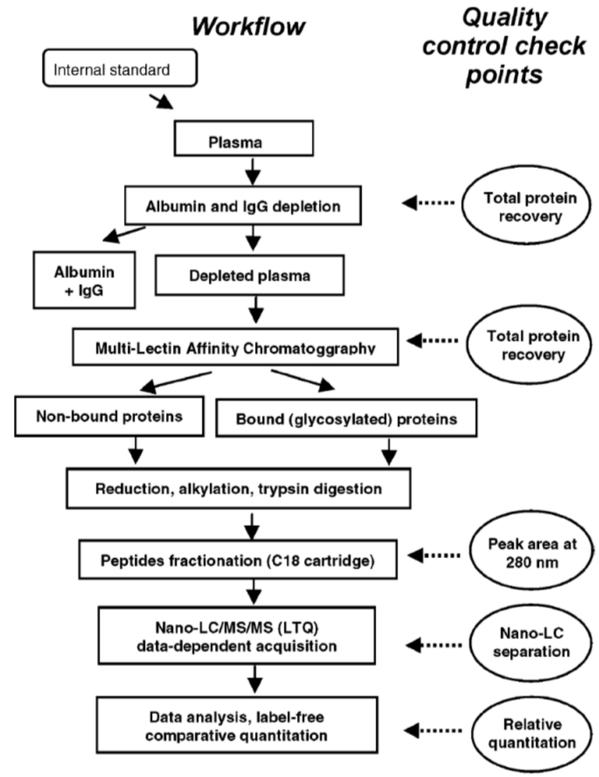


Figure 1. Flowchart of plasma analysis. To maintain the quality of the analysis, standard operating procedures were devised for each step of the method. The following quality control check-points were monitored: total protein recovery during the abundant protein depletion and M-LAC fractionation, as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

MLAC results: comparison of glycoprotein abundances in plasmas in normal versus psoriasis (LTQ linear ion trap)

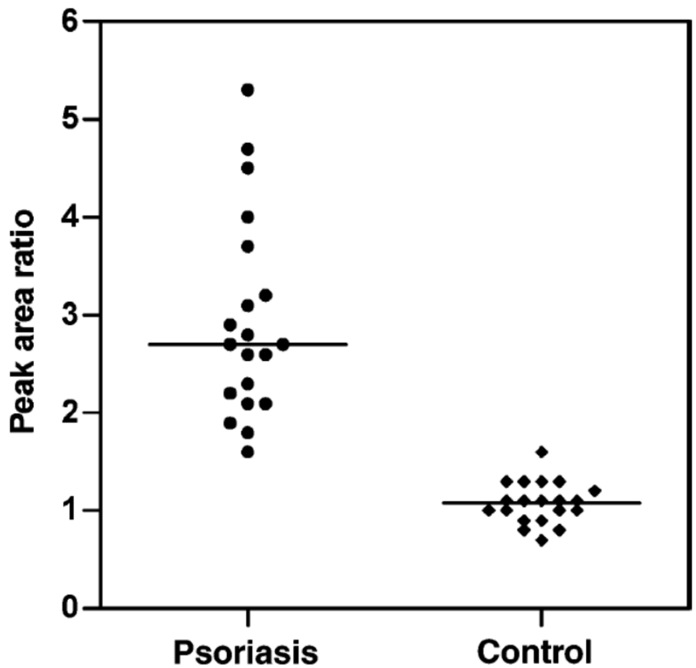


Figure 2. Levels of galectin 3 binding protein (G3BP) assess by peak area ratios of EIC (● - psoriasis samples, ◆ - contı samples). G3BP levels were significantly elevated (p < 0.00(in psoriasis plasma samples, compared to control plasr samples. Horizontal lines show the medians.

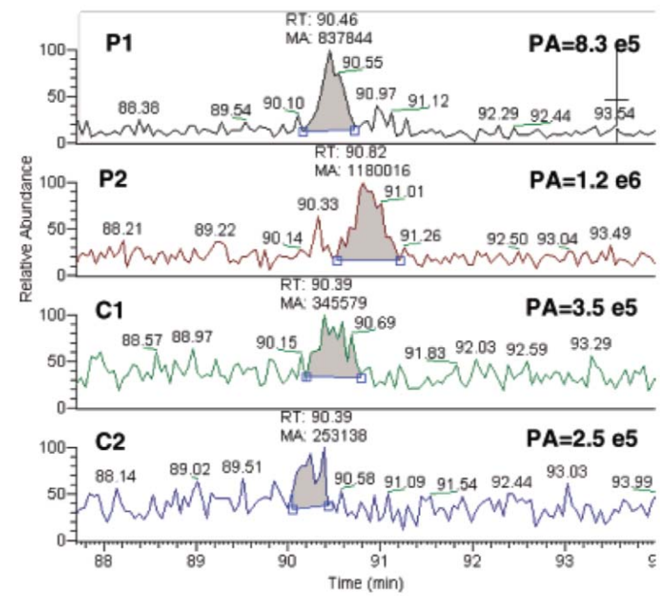


Figure 3. Example of comparative quantitation for galectin 3 binding protein (G3BP). Extracted ion chromatograms (EIC) of a tryptic peptide (R.GQWGTVCDNLWDLTDASVVCR.A) derived from G3BP: P1 - psoriasis sample 1, P2 - psoriasis sample 2, C1 - control sample 1, C2 - control plasma sample 2. PA represents peak area of extracted ion chromatograms. The limit of detection for peak area of EIC is 2.5×10^5.

Results verified using ELISA

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

Another multiple lectin affinity chromatography approach:

- Madera M, et al. 2007. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences 845:121-137.
- four silica-bound lectin microcolumns prepared:
  - ConA, broad specificity, Man, Glc, Gal, NLs
  - SNA-1 α6NeuAc
  - UEA-1 α2Fuc > α3Fuc, α6Fuc
  - PHA-L Galβ4GlcNAcβ2Man trisacc (cancer interest)
- MARS-depletion of abundant blood proteins, sequential silica-based lectin affinity chromatography followed by high temperature chromatography of the enriched proteins

Table 1. Elution Buffers Used for the Displacement of Glycoproteins from the Different Lectin Microcolumns

used lectin (source)	binding buffer	elution buffer
Con A (*Canavalia ensiformis*)	10 mM TRIS-HCl (pH 7.4), 150 mM NaCl, 1 mM $MnCl_2$, 1 mM $CaCl_2$, 1 mM $MgCl_2$, 0.02% NaN_3	0.1 M methyl-α-D-mannoside in binding buffer
SNA (*Sambucis nigra*)	10 mM sodium phosphate buffer (pH 7.4), 150 mM NaCl, 0.02% NaN_3	0.1 M D-lactose in binding buffer
UEA-I (*Ulex europaeus*)	10 mM sodium phosphate buffer (pH 7.4), 150 mM NaCl, 0.02% NaN_3	0.1 M α-L-fucose in binding buffer
PHA-L (*Phaseolus vulgaris*)	10 mM sodium phosphate buffer (pH 7.4), 150 mM NaCl, 0.02% NaN_3	0.4 M *N*-acetyl-D-glucosamine in binding buffer

2006. Journal of Proteome Research 5:2348-2363.

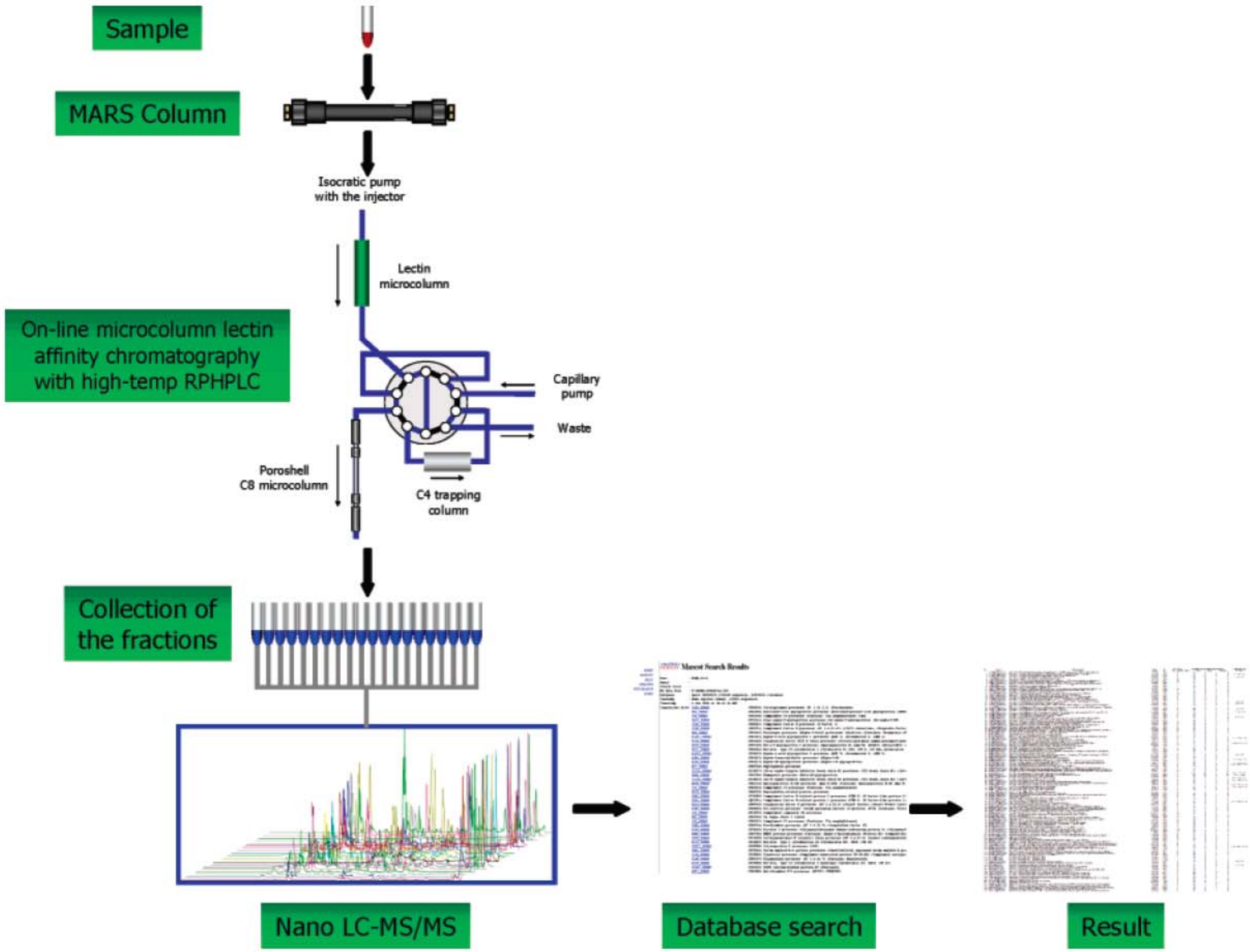


Figure 1. Experimental workflow chart.

2006. Journal of Proteome Research 5:2348-2363.

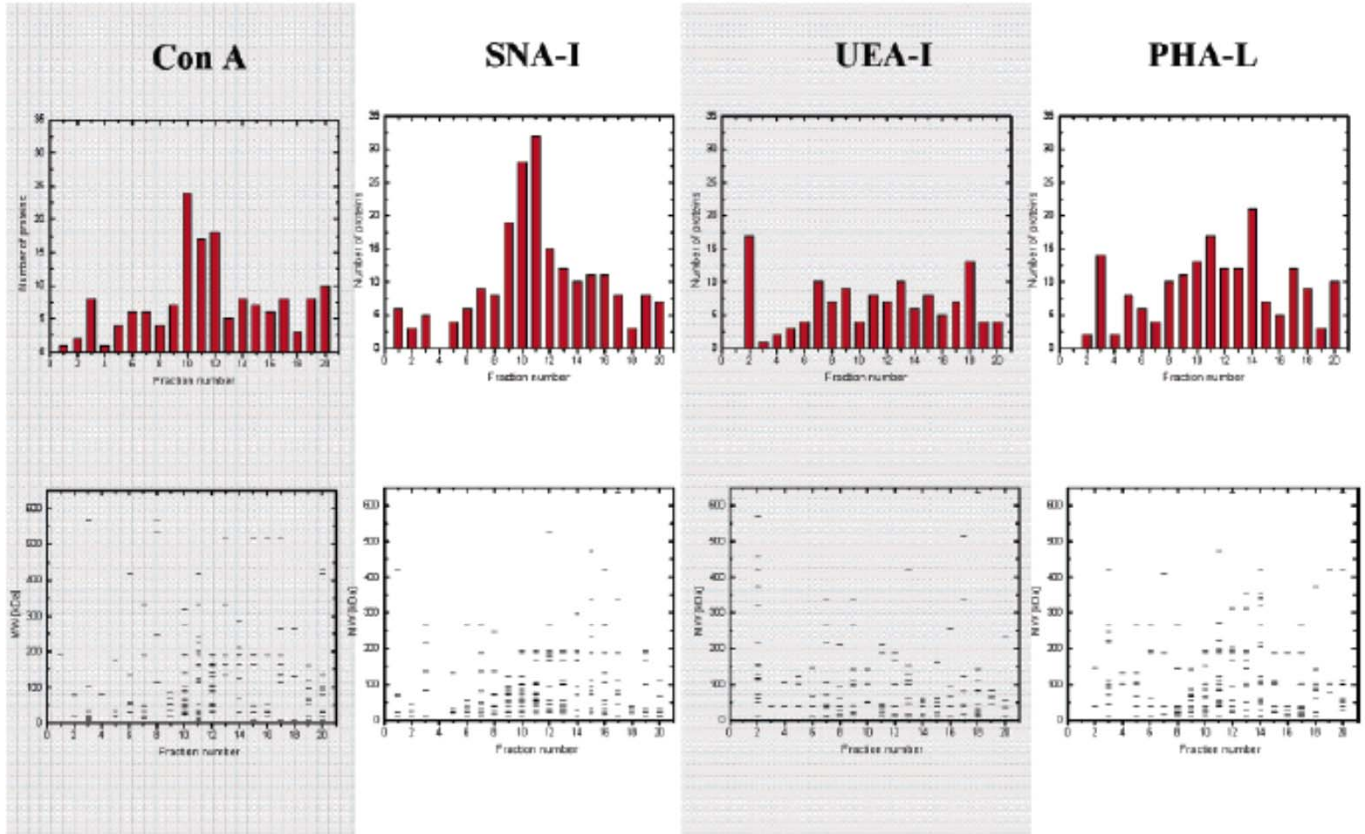
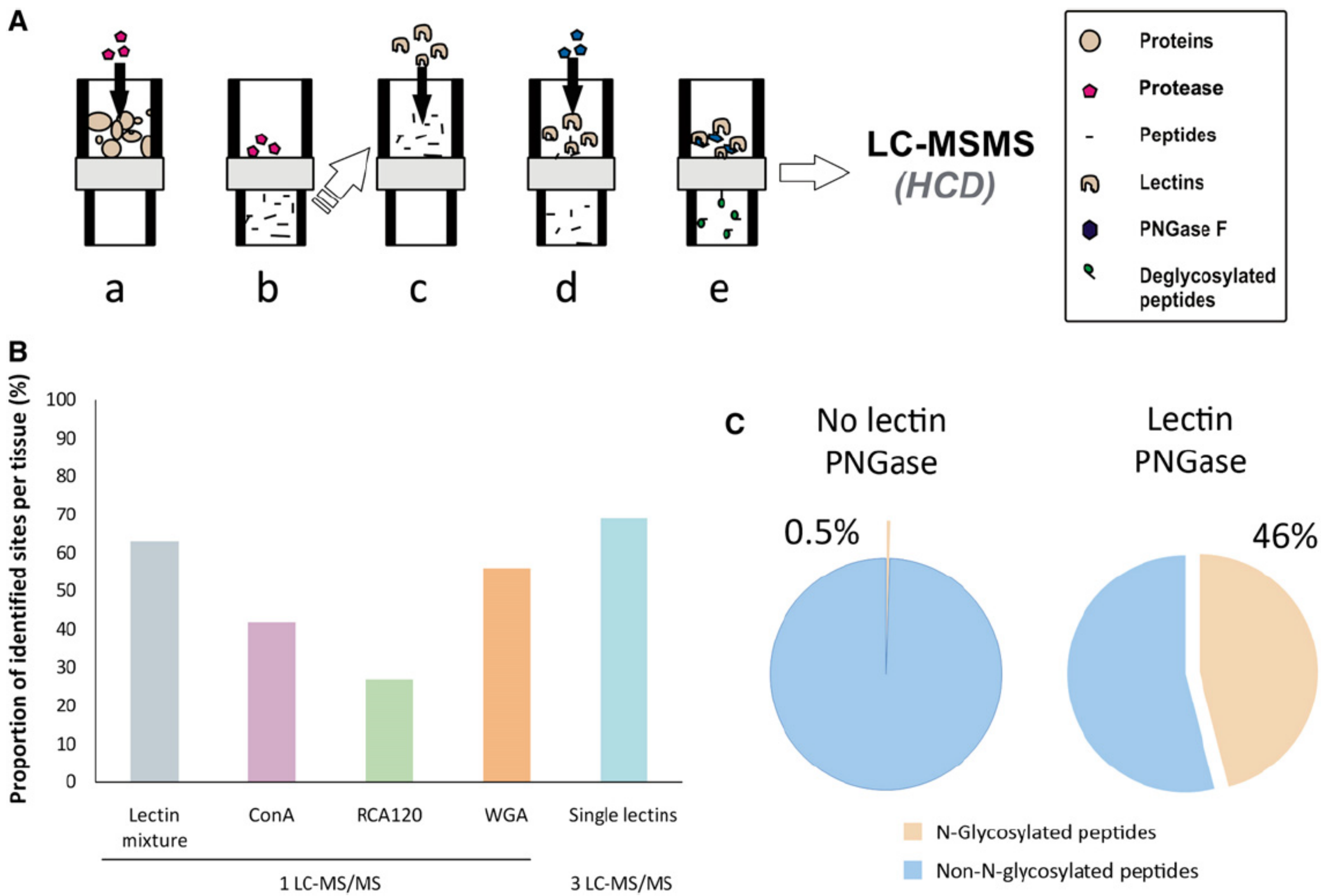


Figure 4. Distribution of the numbers (upper plots) and molecular weights (lower plots) of the different glycoproteins identified by LC–MS/MS of their tryptic digests after lectin-enriched and Poroshell separation.

2006. Journal of Proteome Research 5:2348-2363.

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

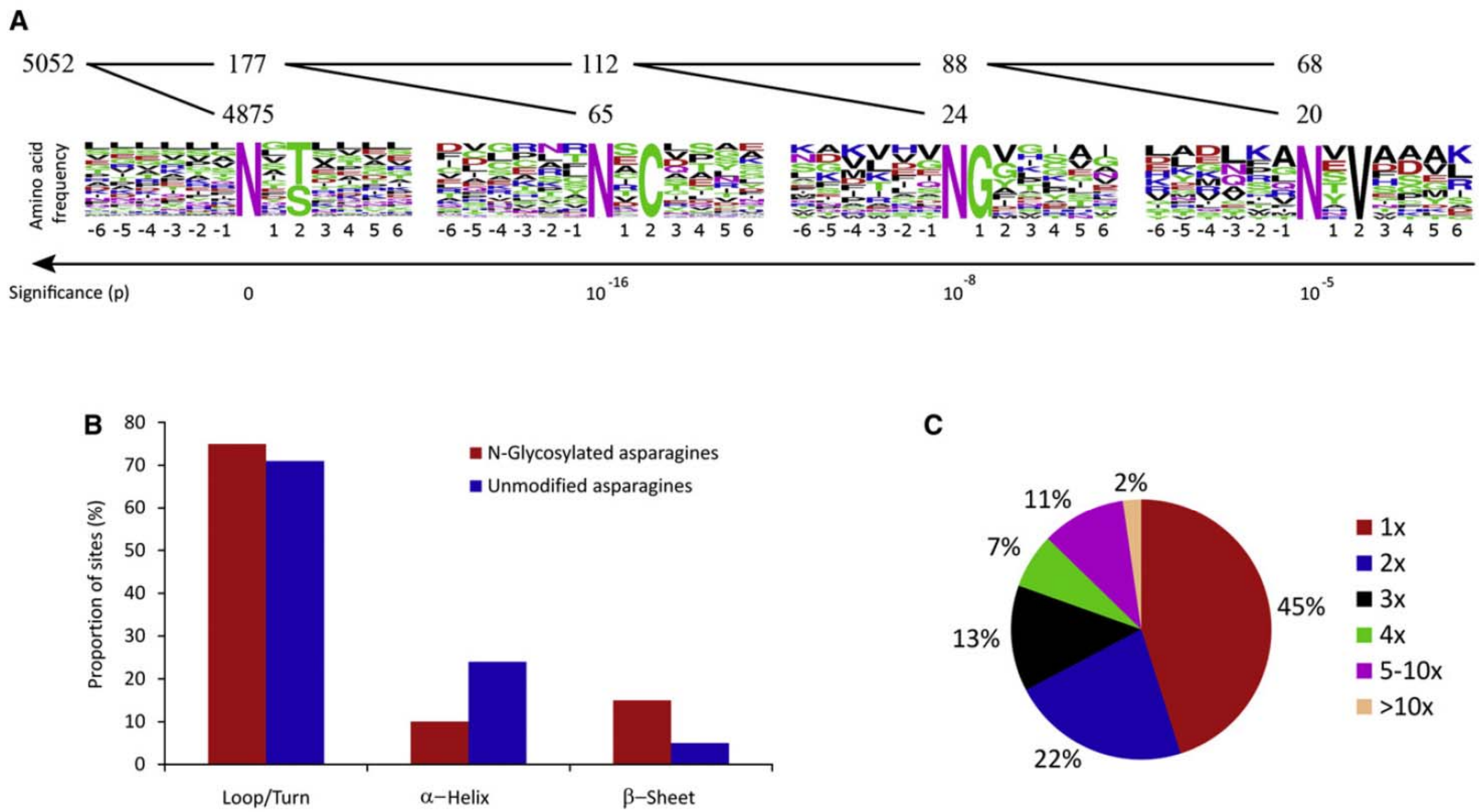


Figure 4. Sequence Recognition Motifs, Structure Preference, and Multiple Glycosylation

(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.