
Mass Spectrometry in an “Omics” World ME.330.804

Hui Zhang
Department of Pathology, Clinical Chemistry Division
Johns Hopkins University
hzhang32@jhmi.edu
Topics

• Glycoprotein classes
• Glycoprotein enrichments
• Release of peptides and glycosite analysis using MS
• Glycomic analysis using MS
Glycoproteins

- Common protein modification: over 1/2 of the mammalian proteins

- Diverse biological processes such as immune response, cellular regulation, and cell signaling

- Alterations in glycosylation patterns are linked to diseases


Types of Glycoproteins

- **N-glycosylation**
  To the Asn side chain of proteins containing the sequon Asn-X-Ser/Thr (where X is any amino acid except Pro)

- **O-GalNAc glycosylation**
  Begins with the addition of a N-acetylgalactosamine to the OH of specific Ser or Thr side chains

- **O-GlcNAc modification**
  N-acetylglucosamine addition to the oxygen of specific Ser or Thr side chains

- **Glycosylphosphatidylinositol (GPI)**

- **Proteoglycans**: Glycosaminoglycans (GAG)
Structure and Names of Common Monosaccharides

- D-Galactose (Gal)
- D-Mannose (Man)
- L-Fucose (Fuc)
- N-acetyl-D-glucosamine (GlcNAc)
- D-Glucose (Glc)
- N-acetyl-D-galactosamine (GalNAc)
- D-Xylose (Xyl)
- N-acetylneuraminic acid (NeuAc)

Major Classes of N-Glycans

Oligomannose

Complex

Hybrid
O-GalNAc Glycans with Different Core Structures

Tn antigen  \( \text{GalNAc} \alpha \text{Ser/Thr} \)
Sialyl-Tn antigen  \( \text{Sia} \alpha 2-6\text{GalNAc} \alpha \text{Ser/Thr} \)

Complex O-GalNAc Glycans with Different Core Structures

Extended core 1

Extended core 2

Extended core 3

Extended core 4
O-linked Glycosylation

GalNAc  Man  Fuc  Glc  GlcNAc
Ser/Thr  Ser/Thr  Ser/Thr  Ser/Thr  Ser/Thr

Mucins  Notch Coagulation Factors  Fibrinolytic Factors

ALSO: Proteoglycans, Hydroxyproline/Hydroxylysine Glycosylation

Esko, J.
**O-GlcNAc Modifications**

- Modification at Ser/Thr residues;
- A ubiquitous and dynamic form of protein modification;
- Present in cytosolic proteins and nuclear proteins;
- Some modification sites overlap with phosphorylation sites;
- Protein interactions, signal transduction, glucose sensing;
  Implicated in insulin resistance, stress response, and regulation of proteosome’s functions.


Courtesy of Yingming Zhao
General Structure of GPI Anchors

\[ \text{H}_2\text{N} \quad \text{Protein} \quad \text{C} \quad \text{NH}_2 \]

\[ \text{H}_2\text{N} \quad \text{Protein} \quad \text{C} \quad \text{NH}_2 \]

Ethanolamine

Man\(\alpha 1-2\)

Man\(\alpha 1-6\)

Man\(\alpha 1-4\)

GlcN\(\alpha 1-6\)

R1 = fatty acid or OH
R2 = fatty acid or alkyl or alkenyl chain
(Note, in some cases, the lipid may also
be a ceramide rather than a glycerolipid)
R3 = fatty acid or OH
R4, R9 = ethanolamine phosphate or OH
R5,6,7,8,10 = carbohydrate substituents or OH

Phosphatidylinositol
Proteoglycans Consist of a Protein Core and One or More Covalently Attached Glycosaminoglycan

Diagram showing various types of proteoglycans:
- Glypican
- Syndecan
- Aggrecan
- Perlecan
- Decorin
- Biglycan

Extracted text from the image:
Proteoglycans Consist of a Protein Core and One or More Covalently Attached Glycosaminoglycan
Glycosaminoglycans Consist of Repeating Disaccharide Units

Hyaluronan (HA)

4S

β4▶β3 ▶β4▶β3 ▶β4▶β3 ▶β4▶β3 ▶β4▶β3 ▶β4▶β3

Chondroitin sulfate (CS)

4S

β4▶β3 ▶β4▶β3 ▶β4▶α3 ▶β4▶α3 ▶β4▶α3 ▶β4▶β3

6S

β4▶β3 ▶β4▶β3 ▶β4▶β3 ▶β4▶β3

Dermatan sulfate (DS)

6S

α4▶α4 ▶α4▶β4 ▶α4▶β4 ▶α4▶β4 ▶α4▶β4

NS ▶2S ▶NS ▶NS ▶NS

Heparan sulfate/heparin (HS)

6S

β3 ▶β4 ▶β3 ▶β4 ▶β3 ▶β4 ▶β3 ▶β4 ▶β3

6S

Keratan sulfate (KS)
Keratan Sulfates Contain a Sulfated Poly-\(N\)-acetyllactosamine Chain
Topics

- Glycoprotein classes
- Glycoprotein enrichments
- Release of peptides and glycosite analysis using MS
- Glycomic analysis using MS
Glycoprotein Characterization

- Glycoprotein Identification
- Glycosylation type: glycan classes and their conjugation sites
- Glycan structures
- Glycans on each glycosylation site
- Glycosylation occupancy
- Quantitation
Characterization of Glycoproteins and Glycans Using MS

Sample preparation procedure → MALDI-MS-MSn

Glycoproteins → Proteolysis → Peptides + Glycopeptides → Enrichment → Glycopeptides → Release glycans → Formerly glycosylated peptides + Glycans → Formerly glycosylated peptides → Glycans

Analyzer → Separation
Detector → Fragmentation
Ion trap → Ionization
Sample plate →
Isolation of Glycopeptides

1. Detection with specific lectins or antibodies
2. Chemical reactions with constituent monosaccharides-A general method labeling glycans on proteins involves periodate oxidation followed by Schiff base formation with amine- or hydrazide-based probes.
3. Metabolic labeling with chemically reactive monosaccharides
4. Label in vitro using a purified glycosyltransferase
5. LC-based enrichments
1. Lectins

- Carbohydrate-binding proteins
- More than 2,000 lectins
- Many lectins became commercially available
- Multiple lectins with distinct binding specificities are used in combination or in series

- http://proline.physics.iisc.ernet.in/cgi-bin/cancerdb/input.cgi
- http://nscdb.bic.physics.iisc.ernet.in
# Lectin Affinity Chromatography

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Metal Ions</th>
<th>Specificity</th>
<th>Elution</th>
<th>Binding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Con A</td>
<td>Ca2+ Mn2+</td>
<td>a-Man&gt;a-Glc</td>
<td>0.1-0.5 M a-MeMan</td>
<td>N-linked</td>
</tr>
<tr>
<td>SNA</td>
<td>-</td>
<td>Siaα2-6Gal or GalNAc</td>
<td>0.1-0.5 Lactose</td>
<td>α2-6-linked Sia</td>
</tr>
<tr>
<td>UEA</td>
<td>-</td>
<td>A-L-Fuc</td>
<td>0.1-0.5 M L-Fuc</td>
<td>Fuc</td>
</tr>
</tbody>
</table>

Lectin Affinity Capture, Isotope-coded Tagging and Mass Spectrometry to Identify N-linked Glycoproteins

Other Affinity Reagents

- **Antibodies against glycans**: Monoclonal antibodies against Lewis X antigen. (Baeckström D et al. J Biol Chem. 1991; 266: 21537-21547)

- **Glycoprotein receptors**: Mannose-6-phosphate (M6P) receptors for M6P-motifs containing glycoproteins (Sleat DE et al Proteomics 2005; 5: 1520–1532)
2. Identification and Quantification of N-linked Glycoproteins Using Hydrazide Chemistry and Mass Spectrometry

3. Metabolic Incorporation of the N$_3$-GlcNAc into Proteins

Extracellular

Intracellular

Phospho-N-Acetylglucosamine Mutase

O-GlcNAcase

UDP-N$_3$-GlcNAc + UDP-N$_3$-GlcNAc transferase

O-GlcNAcase

Courtesy of Yingming Zhao
Chemoselective Conjugation Between $N_3$ and a Phosphine

$N_3$: small, inert, uncharged, non-polar, air-stable, and abiotic.

Tagging-via-substrate strategy for probing O-GlcNAc modified proteins. J
Identification of O-N₃-GlcNAc-modified Proteins

O-N₃-GlcNAc Conjugation Affinity-purified by streptavidin-beads
2% SDS 8 M Urea

Digestion

HPLC/MS/MS for protein identification

Courtesy of Yingming Zhao
4. Carbohydrate-tags Via Chemo-enzymatic Labeling

- Utilizes a genetically engineered galactosyltransferase
- Incorporate ketone analogs of galactose to cellular O-glycosylated proteins
- Incorporate a biotin label through coupling with aminooxy-biotin

Chemoenzymatic Strategy for Identifying O-GlcNAc-glycosylated Proteins from Cellular Lysates
Application of the Strategy Toward Crystallin

(a) MS analysis revealed the tagged $O$-GlcNAc peptide.
(b) MS2 spectrum of the precursor ion revealed the signature loss of the ketone-biotin moiety.
(c) MS3 analysis revealed the loss of the GlcNAc moiety.
(d) MS4 analysis generated additional y and b ions that were used to sequence the peptide.
(e) Summary of the y and b fragment ions.
5. LC-based Enrichment

- **Hydrophilic interaction LC (HILIC):** the hydrophilic nature of glycopeptides (Wada Y, Tajiri M, Yoshida S. Anal Chem 2004; 76: 6560-6565)

- **Size exclusion chromatography:** masses of \(N\)-glycans are larger than 1200 Da; thus, most \(N\)-glycopeptides could be enriched by size-exclusion chromatography (Alvarez-Manilla G et al. J Proteome Res 2006; 5: 701-708)

- **Boronic acid:** Boronic acid forms boronic diesters through reaction of geminal diols (Sparbier K et al. J Chromatogr B 2006; 840: 29–36)

- **Strong cation exchanger (SCX):** Glycopeptides with a terminal sialic acid can be enriched by LC on an SCX column (Lewandrowski U et al. Mol Cell Proteomics 2007; 6: 1933–1941)

- **Titanium dioxide:** Sialic acid binds TiO2 (Larsen MR et al. Mol Cell Proteomics 2007; 6: 1778–1787)
Topics

- Glycoprotein class
- Glycoprotein enrichments
- Release of peptides and glycosite analysis using MS
- Glycomic analysis using MS
Release of N-Linked Glycans Using PNGase F

- PNGase conversion of Asn to Asp
- Mass shift confirms that the peptide was glycosylated and localizes site of N-linked glycosylation
- Labeling peptides and glycans with 18O water on the glycosylation site

Cleavage converts Asn to Asp

Reducing glycan and 18O labeling
Mass Spectra of Glyco and Non-glycopeptides After Releasing N-glycans
Enzymatic Release N-Glycans

PNGase F

R₁ = N- and C- substitution by groups other than H
R₂ = H or the rest of an oligosaccharide structure
R₃ = H or α(1-6) fucose

N-Glycosidase A

R₁ = N- and C- substitution by groups other than H
R₂ = H or the rest of an oligosaccharide
R₃ = H or α 1,6 fucose

Endoglycosidase F₁

R₁ = Asn or H
R₂ = Oligomannose or hybrid configurations
R₃ = H or α 1,6 fucose

Endoglycosidase F₂

R₁ = H or Asn
R₂ = Biantennary and oligomannose configurations
R₃ = H or α 1,6 fucose

Endoglycosidase F₃

R₁ = N- and C- substitution by groups other than H
R₂ = Biantennary and biantennary complex oligosaccharides or trimannosylchitobiose core
R₃ = H or α 1,6 fucose
R₄ = Asn (H or Asn if core fucosylated)

Endoglycosidase H

R₁ = Oligomannose (2-150)
R₂ = H or mono or oligo saccharide at the C₂ or C₄ position
R₃ = H or α 1,6 fucose
R₄ = Asn or Dolichol pyrophosphate

Sigma
Enzymatic Release of O-Glycans

Sigma
Chemical Release of O-Glycans

A. Beta-elimination of Ser/Thr PTMs

PTM-Ser

PTM-Thr

Conditions:
- O-GlcNAc 1% TEA, 0.1% NaOH, 2 hr. 50°C
- O-GalNAc, 1M NaOH, 2M NaBH₄, o/n 20°C
Release of GPI

- Successful cleavage by GPI-specific phospholipases can be assessed by subsequently analyzing samples by MS, because removal of the GPI anchor causes a shift in molecular mass. This is a common diagnostic method for identifying the presence of a GPI anchor on a protein of interest.
- Another method is to treat the GPI-anchored protein with nitrous acid, which cleaves the unsubstituted glucosamine residue that links the glycan to the phosphatidylinositol (PI).
Enzymatic Release of GPI Anchors

Cleavage sites of phospholipases. Phospholipase C cut just before the phosphate attached to the $R_3$ moiety.
Release of Proteoglycans

- Proteoglycans typically contain more glycan than protein. They may be separated by agarose gel electrophoresis and by ion-exchange chromatography, which separates on the basis of the charge conferred by sulfate groups.
- Treatment of proteoglycans with GAG lyases will produce a shift in mass to remove most of the glycan portion.
- Antibodies that recognize the remaining structures (“stubs”) may be used in western analysis. The lyases cleave a 4,5 unsaturated uronic acid at the no reducing end. Anti-“stub” antibodies recognize the sulfation of the penultimate N-acetylglicosamine or N-acetylgalactosamine residue.
Treating Glycoproteins with Proteases

N-Glycans and O-glycans can be obtained nonselectively by degradation of the protein by proteases to generate glycopeptides.
Chemical Methods to Release Glycans

• Hydrazinolysis: A chemical method that uses hydrazine to cleave amide bonds (e.g., the glycosylamine linkage between a sugar residue and asparagine or the acetamide bond in N-acetylhexosamines) to release both N-glycans and O-glycans or, under controlled conditions, cleaves only the N-glycans.

• Anhydrous hydrogen fluoride treatment: cleaves all the linkages of glycans while leaving peptide bonds and glycopeptide linkage linkages of amino sugars intact.
Topics

• Glycoprotein classes
• Glycoprotein enrichments
• Release of peptides and glycosite analysis using MS
• Glycomic analysis using MS
Control of Glycoconjugate Biosynthesis

DNA → RNA → Proteins → Enzymes → Carbohydrates → Lipids

Glycocalyx
Collagens
NCPs
Adhesion molecules
Proteoglycans

ECM
Cell
Organism

Glycoconjugates
Control of Glycan Structures

- Expression and activities of enzyme
- Nucleotide sugar availability
- Kinetics of transports
- Glycoprotein expression
- Availability of glycosylation sites
- Glycans are mixtures of variants (glycoforms) on a core structure
Mass Spectrometric Analysis of Glycans

- Sample preparation/purification
- Separation
- Ionization
Sample Preparation Strategies for Glycans

Joseph Zaia  
Separation of Glycans for MS Analysis

- Reversed phase LC/MS: reductive amination is applied to increase the hydrophobicity
- Graphitized carbon chromatography-MS: separate structural isomers
- Hydrophilic interaction chromatography (HILIC)
- Lectin affinity
Glycomics Using Mass Spectrometry

- Putative structures are assigned to each molecular ion based on the usually unique glycan composition for a given mass.
- Prior knowledge of N- and O-glycan biosynthesis.
- Assignments can be confirmed in a second experiment employing ESI-MS/MS instrumentation by selecting each molecular ion for collisional activation and recording its fragment ion spectrum.
- Additional information can be provided by MS experiments on chemical and enzymatic digests, the choice of which is guided by the sequence information provided by mass mapping and MS/MS experiments.
Glycan Fragmentation Ions

Domon and Costello, 1988


November 28, 2012
Data from a Glycomics Study of N-glycans from Mouse Kidney
Linkage Analysis

• The principle of this method is to introduce a stable substituent (an ether-linked methyl group) onto each free hydroxyl group of the native glycan.

• The glycosidic linkages, which are much more labile than the ether-linked methyl groups, are then cleaved with free hydroxyl groups at the positions that were previously involved in a linkage.
Differentiation of Glycan Isomers Using Tandem MS Analysis of Permethylated Glycans

(A) $m/z$ 445

(B)

November 28, 2012
Glycosidases Used for Structural Analysis

Glycosidases remove sugars

- Exoglycosidases remove terminal sugars
- Endoglycosidases remove glycans

High mannose- or hybrid-type N-glycan

Must have this residue to be cleaved by Endo H

Endoglycosidase H (Endo H)
Quantitative Glycomics

• Label-free: Permethylation and MS
• Stable isotope labels for glycomics: based on differential stable isotope labeling (CH3I/CD3I) and permethylation
• Reductive Amination Labeling: d0/d4 pyridyl amine (PA)
Determination of Anomericy

• Sequential exoglycosidase digestions: Cleavage by α- or β-exoglycosidases indicates the anomericy of specific terminal sugar residues.

• Many glycosidases are specific for both monosaccharide residue and linkage type, allowing detailed structural conclusions, although the number of such enzymes available is limited.
Glycoproteomics: Analysis of Glycopeptides with Glycan Attached

• Mass spectrometric analysis of glycopeptides is made challenging by the differing chemical properties of glycans and peptides.
• The ultimate goal of glycoproteomics is to quantify the site occupancy of glycosylation in the proteome and the structures of glycoforms at each site.
Enrichment of Glycopeptides and MS Analysis

- Lectins for affinity capture
- Graphitized carbon solid phase extraction
- Abundant peptide backbone dissociation is observed for glycopeptides using ETD. CAD results in preferential fragmentation of the glycan moiety of glycopeptides.
The Informatics Challenges of Diverse Glycomic Data

- Efforts to correlate large data sets obtained from glycomic, transcriptomic, genomic, and proteomic studies have met with several challenges.
- Representation of glycan chemical structures is difficult because of their complexity and branching patterns. The use of single alphabet codes, as employed to describe nucleic acid and amino acid sequences, is not applicable to glycans.
- The field is in need of a comprehensive bioinformatics platform that stores, integrates, and processes data from glycomic and other “omic” studies and disseminates them in a meaningful fashion via the Internet to the scientific community.
Databases and Bioinformatics Platforms

- GlycoSuiteDB, Sweet, KEGG GLYCAN
- The Consortium for Functional Glycomics (CFG)
- EuroCarbDB
- National Center for Glycomics and glycoproteomics
- Glycomod: all possible compositions of a glycan structure
- GlycoPep DB: N-glycopeptide compositional assignment
- Cartoonist: automated annotation of \textit{N-glycan} MALDI TOF mass spectra with cartoons representing the most plausible glycan assemblies synthesized by mammals using 300 manually determined archetypes.
Databases and Bioinformatics Platforms

- Peptoonist: automated identification of \( N\)-glycopeptides using a combination of MS and MS/MS data
- Glyco-Peakfinder: rapid assignment of glycan compositions, is intended to be entirely a \textit{de novo} platform for compositional analysis
- SysBioWare: carbohydrate assignment
- NCRR GlycomicsPortal
- SimGlycan
- Accurate Glycan Analyzer
- GlycoWorkbench
Summary

- Glycoprotein classes
- Glycoprotein enrichments
- Release of peptides and glycosite analysis by MS
- Glycomics analysis