Glycobiology

The study of the biological functions, structures, recognition and biosynthesis of glycans (sugar chains, saccharides) in the context of the biological scaffolds to which they are attached (e.g. glycolipids & glycoproteins).
Nobel Laureates in Glycobiology

E. Fischer  K. Landsteiner  N. Haworth

C. Cori  G. Cori  L. Leloir  G. Palade

Also, James Rothman and Randy Schekman (2013)
Recommended Books on Glycobiology

*Introduction to Glycobiology*
Maureen E. Taylor and Kurt Drickamer
Oxford University Press, 2011

*Essentials of Glycobiology*
EDITED BY Varki, Cummings, Esko, Stanley, Hart, Aeby, Darvill, Kinoshita, Packer, Prestegard, Schnaar, Seiberg
CHS Press, 2017
Content freely available at: http://www.ncbi.nlm.nih/books/NBK310274
Basic Principles of Glycobiology

Occurrence

- All cells in nature are covered with a dense and complex array of sugar chains (glycans).
- The cell walls of bacteria and archaea are composed of several classes of glycans and glycoconjugates.
- Most secreted proteins of eukaryotes also carry large amounts of covalently attached glycans.
- In eukaryotes, these cell-surface and secreted glycans are mostly assembled via the ER-Golgi pathway.
- The extracellular matrix of eukaryotes is also rich in such secreted glycans.
- Cytosolic and nuclear glycans are common in eukaryotes.
- For topological, evolutionary, and biophysical reasons, there is little similarity between cell surface/secreted and nuclear/cytosolic glycans.
**Basic Principles of Glycobiology**

Chemistry and Structure

- Glycosidic linkages can be in $\alpha$- or $\beta$- linkage forms, which are biologically recognized as completely distinct
- Glycan chains can be linear or branched
- Glycans can be modified by a variety of different substituents, such as acetylation and sulfation
- Complete sequencing of glycans is feasible, but usually requires combinatorial or iterative methods
- Modern methods allow in vitro chemoenzymatic synthesis of both simple and complex glycans
Basic Principles of Glycobiology

Biosynthesis

✓ The final products of the genome are posttranslationally modified proteins, with glycosylation being the most common and versatile of these modifications.

✓ The primary units of glycans (monosaccharides) can be synthesized within a cell or salvaged from the environment.

✓ Monosaccharides must be activated into nucleotide sugars or lipid-linked sugars before they are used as donors for glycan synthesis.

✓ Whereas lipid-linked sugar donors can be flipped across membranes, nucleotide sugars must be transported into the lumen of the ER-Golgi pathway.

✓ Each linkage unit of a glycan or glycoconjugate is assembled by one or more unique glycosyltransferases.

✓ Many glycosyltransferases are members of multigene families with related functions.

✓ Most glycosyltransferases recognize the underlying glycan of their acceptor, but some are protein or lipid specific.

✓ Many biosynthetic enzymes (or glycosyltransferases, glycosidases, sulfotransferases, etc.) are expressed in a tissue-specific, temporally regulated manner.
Basic Principles of Glycobiology

Diversity

✓ Monosaccharides generate much greater combinatorial diversity than nucleotides or amino acids.
✓ Further diversity arises from covalent modifications of glycans.
✓ Glycosylation introduces a marked diversity in proteins.
✓ Only a limited subset of the potential diversity is found in a given organism or cell type.
✓ There is intrinsic diversity (microheterogeneity) within a cell type or even a single glycosylation site.
✓ The total expressed glycan repertoire (glycome) of a given cell type or organism is thus much more complex than the genome or proteome.
✓ The glycome of a given cell type or organism is also dynamic, changing in response to intrinsic and extrinsic signals.
✓ Glycome differences in cell type, space, and time generate biological diversity, and can help explain why only a limited number of genes are expressed from the typical genome.
Basic Principles of Glycobiology

Recognition

✓ Glycans are recognized by specific glycan-binding proteins that are intrinsic to an organism.

✓ Glycans are also recognized by many extrinsic glycan-binding proteins of pathogens and symbionts.

✓ Glycan-binding proteins fall in two general categories: those that can usually be grouped by shared evolutionary origins and/or similarity in structural folds (lectins) and those that emerged by convergent evolution from different ancestors (e.g, GAG-binding proteins).

✓ Lectins often show a high degree of specificity for binding to specific glycan structures, but they typically have relatively low affinities for single-site binding.

✓ Thus, biologically relevant lectin recognition usually requires multivalency of both the glycan and glycan-binding protein, to generate high avidity of binding.
Basic Principles of Glycobiology

Genetics

✓ Naturally occurring genetic defects in glycans seem to be relatively rare in intact organisms. However, this apparent rarity may be due to a failure of detection, caused by unpredictable or pleiotropic phenotypes.

✓ Genetic defects in cell-surface/secreted glycans are easily obtained in cultured cells but have somewhat limited biological consequences.

✓ The same mutations typically have major phenotypic consequences in intact multicellular organisms.

✓ Thus, many of the major roles of glycans likely involve cell–cell or extracellular interactions.

✓ Nuclear/cytosolic glycans may have more cell-intrinsic roles, e.g., in signaling.

✓ Complete elimination of major glycan classes generally causes early developmental lethality.

✓ Organisms bearing tissue-specific alteration of glycans often survive, but they exhibit both cell-autonomous and distal biological effects.
Basic Principles of Glycobiology

Biological Roles

- Biological roles for glycans span the spectrum from nonessential activities to those that are crucial for the development, function, and survival of an organism.
- Many theories regarding the biological roles of glycans appear to be correct, but exceptions occur.
- Glycans can have different roles in different tissues or at different times in development.
- Terminal sequences, unusual glycans, and modifications are more likely to mediate specific biological roles.
- However, terminal sequences, unusual glycans, or modifications may also reflect prior evolutionary interactions with microorganisms and other noxious agents.
- Thus, a priori prediction of the functions of a specific glycan or its relative importance to the organism is difficult.
Basic Principles of Glycobiology

Evolution

- Relatively little is known about glycan evolution.
- Interspecies and intraspecies variations in glycan structure are relatively common, suggesting rapid evolution.
- The dominant mechanism for such evolution is likely the ongoing selection pressure by pathogens that recognize glycans.
- However, glycan evolution must also preserve and/or elaborate critical intrinsic functions.
- Interplay between pathogen selection pressure and preservation of intrinsic roles could result in the formation of "junk" glycans.
- Such "junk" glycans could be the substrate from which new intrinsic functions arise during evolution.
Glycobiology is ... NOT Carbohydrates as Food

Carbohydrates in food are important sources of energy. Starch found in plant-derived food such as pasta ....

Cellulose
Chitin
Glycobiology is NOT “a” Post-Translational Modification…. it is Thousands of Modifications

- Combinatorial linkage position, orientation and branching provide the potential for millions* of different glycan structures – functional diversity
- Glycans may be larger and are more diverse than their (protein) carriers
- A glycan’s function can supersede that of its (protein) carrier
- 1-2% of the human genome is devoted to glycosylation

Modified from http://www-che.syr.edu/ faculty/hougland.html

*Glycoproteins and glycolipids may contain ~3000 glycan determinants with an additional ~4000 theoretical pentasaccharide sequences in glycosaminoglycans

Glycobiology is ... NOT a “Decoration”

N-linked carbohydrate can form both an immunologically silent face—with carbohydrate masquerading as "self"—and also can protect neighboring epitopes through an "evolving glycan shield"

“Structure” of HIV gp120
Zolla-Pazner (2004)
*Nature Reviews Immunology* 4, 199

Proteins typically fold inward, whereas glycans spread out in space.

CD59, a complement defense glycoprotein


Basic Definitions

- **Monosaccharide**: A carbohydrate that cannot be hydrolyzed into a simpler carbohydrate. The building block of oligosaccharides and polysaccharides.

- **Oligosaccharide**: Linear or branched chain of monosaccharides attached to one another via glycosidic linkages. The number of monosaccharide units can vary.

- **Polysaccharide**: Glycan composed of repeating monosaccharides, generally greater than ten monosaccharide units in length.

- **Carbohydrate, glycan, saccharide, sugar**: Generic terms used interchangeably. Includes monosaccharides, oligosaccharides, polysaccharides, and derivatives of these compounds. Carbohydrates consist of “hydrated carbon”, $[\text{CH}_2\text{O}]_n$

- Preferred generic term is “Glycan”
Glycoconjugates

Glycoconjugate: A compound in which one or more glycans (the glycone) are covalently linked to a non-carbohydrate moiety (the aglycone)

Glycoproteins: Protein with one or more covalently bound glycans

Glycolipids: A molecule containing a saccharide linked to a lipid

Proteoglycans: Any glycoprotein with one or more covalently attached glycosaminoglycan chains
Intra- & Extracellular Localization of Glycoconjugates
Types of Glycan-Protein Linkages in Nature

N-glycosyl
- Asn
- Arg

C-glycosyl
- Trp

P-glycosyl
- Ser

O-glycosyl
- Ser/Thr

C-term Glypiation
- Man-6-P-EthN
Thy-1 Glycoprotein – A Cell Surface Antigen

Involved in cell adhesion and communication, particularly in the immune and nervous systems
Major Glycan Classes in Vertebrate Cells
Independent Functions of Protein & Glycan

Different enzyme activities
Clearance through common receptor

Same enzyme
Targeting to different tissues
Encoding & Decoding Information in Glycans

Genes

Expression dependent on cell type and developmental stage

Glycosyl transferases

Information encoding into glycans

Glycans

Intrinsic effects

Lectins

Information decoding from glycans

Extrinsic effects

Biological functions
Summary of Glycan Functions

Providing structural components
  Cell walls
  Extracellular matrix

Modifying protein properties
  Solubility
  Stability

Intrinsic functions performed by glycans

Directing trafficking of glycoconjugates
  Intracellular
  Extracellular

Mediating and modulating cell adhesion
  Cell–cell interactions
  Cell–matrix interactions

Extrinsic functions resulting from glycan–lectin interactions

Mediating and modulating signalling
  Intracellular
  Extracellular
DNA-Centric View of Molecular & Cellular Biology
Holistic View of Molecular & Cellular Biology

DNA → RNA → PROTEINS → ENZYMES

REGULATORY FACTORS → SIGNALLING MOLECULES

LIPIDS → GLYCANS

GLYCOLIPIDS → GLYCOPROTEINS PROTEOGLYCANS

MICROBES PARASITES → CELL → TISSUES & ORGANS → ORGANISM

PHYSICAL ENVIRONMENT

CULTURAL ENVIRONMENT

DNA → DNA

DIET
Tiny sugar “decorations”... added as an afterthought

Eukaryotic Cell Surface circa 1995

Historical electron micrograph of endothelial cells from a blood capillary in diaphragm muscle of a rat, showing the lumenal cell membrane of the cells (facing the blood) decorated with particles of cationized ferritin (arrowheads).
**What if a Major Cell Component is Invisible by Light Microscopy?**

Light microscopy micrograph of *Cryptococcus neoformans* capsule delineated by India ink. The inner circle represents the fungal cell, with the wide outer circle being the capsule.


Scanning electron microscopy of *C. neoformans* yeast cells.

Unique to Yeast – NO!

Electron microscopic thin section of *Escherichia coli* K1

Electron microscopic thin section of *Klebsiella pneumoniae*
The “glycocalyx” surrounding a fibroblast. Cell surface carbohydrates are stained black.
All Cells Are Coated with Glycans

Electron micrograph of a human lymphocyte (Ruthenium Red staining)
“Evolution has failed to generate a living cell devoid of surface glycosylation” - A. Varki
Glycobiology as a Language (Semiotics)

Cohen & Varki (2010)  
OMICS 4:455
Carbohydrates – The Building Blocks of Glycobiology

\[(\text{CH}_2\text{O})_n = \text{“carbo” “hydrate”}\]

Generalized Aldose

Generalized Ketose
Enantiomers: Mirror images of each other that are not superimposable.

Diastereomers: Stereoisomers that are not enantiomers.

Monosaccharide identity is all about stereochemistry.

Generalized Aldose
Fischer Projections

D-glyceraldehyde
Enantiomers

D-glyceraldehyde

L-glyceraldehyde

* Highest numbered asymmetric carbon = reference carbon
Diasteriomers

D-Erythrose

D-Threose

*Highest numbered asymmetric carbon = reference carbon
Enantiomers & Diasteriomers

D-Erythrose

L-Erythrose

D-Threose

L-Threose

*Highest numbered asymmetric carbon = reference carbon

<table>
<thead>
<tr>
<th># carbon</th>
<th>asymmetric carbons</th>
<th>diasteriomers</th>
<th>diasteriomers &amp; enantiomers</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>8</td>
<td>16</td>
</tr>
</tbody>
</table>
Example of Aldose & Ketose

Glucose

Fructose
Stereochemistry of Glucose & Galactose

Glucose

1. HC=O
2. HCOH
3. HOCH
4. HCOH
5. HCOH
6. CH₂OH

Galactose

1. HC=O
2. HCOH
3. HOCH
4. HOCH
5. HCOH
6. CH₂OH

Epimerization at C4

Above page

Below page
Monosaccharides – The Basic Structural Unit

• Carbonyl group at the end of the carbon chain (aldoses) or at an inner carbon (ketoses) has potential reducing power. This end is called the reducing terminus, or reducing end.

• The ring form of a monosaccharide generates a chiral (anomeric) center (at C-1 for aldo sugars or at C-2 for keto sugars). Notice that other positions are chiral, which therefore imparts stereochemical information.

\[
\begin{align*}
\text{D-glucose} & : \quad \text{HO-} & \text{CH}_2\text{OH} \\
\text{ \quad 3} & \quad \text{H-} & \text{C-} \quad \text{H} \\
\text{ \quad 5} & \quad \text{H-} & \text{C-} \quad \text{OH} \\
\text{ \quad 6} & \quad \text{H-} & \text{C-} \quad \text{OH} \\
\text{ \quad 1} & \quad \text{CHO} \\
\end{align*}
\]
Structural Representations of Glucose: Fischer & Haworth Projections, Anomeric Center
epimers: differ in only one stereogenic center
Graphical Depiction of Monosaccharides
More Monosaccharide Symbols

6-Deoxy Sugars
NAc and Amines are in the 2 position

- 6-deoxy-D-Galactose (D-Fucose)
- 6-deoxy-D-Glucose (D-Quinvose)
- 6-deoxy-D-Mannose (D-Rhamnose)

- Fuc
- Qui
- Rha

FucNAc
QuiNAc
RhaNAc

- FucN
- QuiN
- RhaN

Sialic Acids
- Neu5Ac
- Neu5Gc
- KDN

Ketoses
- Fru
- Sor
- Psi
- Tag
### Essentials of Glycobiology, 3rd Edition — Symbol Nomenclature for Glycans (Appendix 1B)


**Table:** Monosaccharide Symbol Nomenclature

<table>
<thead>
<tr>
<th>SHAPE</th>
<th>White</th>
<th>Blue</th>
<th>Green</th>
<th>Yellow</th>
<th>Orange</th>
<th>Pink</th>
<th>Purple</th>
<th>Light Blue</th>
<th>Brown</th>
<th>Red</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filled Circle</td>
<td>Hexose</td>
<td>Glc</td>
<td>Man</td>
<td>Gal</td>
<td>Gul</td>
<td>Alt</td>
<td>All</td>
<td>Tal</td>
<td>Ido</td>
<td></td>
</tr>
<tr>
<td>Filled Square</td>
<td>HexNAc</td>
<td>GlcNAc</td>
<td>ManNAc</td>
<td>GalNAc</td>
<td>GulNAc</td>
<td>AltNAc</td>
<td>AllNAc</td>
<td>TalNAc</td>
<td>IdoNAc</td>
<td></td>
</tr>
<tr>
<td>Crossed Square</td>
<td>Hexosamine</td>
<td>GlcN</td>
<td>ManN</td>
<td>GalN</td>
<td>GuilN</td>
<td>AltN</td>
<td>AllN</td>
<td>TalN</td>
<td>IdoN</td>
<td></td>
</tr>
<tr>
<td>Divided Diamond</td>
<td>Hexuronate</td>
<td>GlcA</td>
<td>ManA</td>
<td>GalA</td>
<td>GulA</td>
<td>AltA</td>
<td>AllA</td>
<td>TalA</td>
<td>IdoA</td>
<td></td>
</tr>
<tr>
<td>Filled Triangle</td>
<td>Deoxyhexose</td>
<td>Qui</td>
<td>Rha</td>
<td>6dGul</td>
<td>6dAlt</td>
<td>6dTal</td>
<td>Fuc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Divided Triangle</td>
<td>DeoxyhexNAc</td>
<td>QuiNAc</td>
<td>RhaNAc</td>
<td>6dAltNAc</td>
<td>6dTalNAc</td>
<td>FucNAc</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flat Rectangle</td>
<td>Dideoxyhexose</td>
<td>Oli</td>
<td>Tyv</td>
<td>Abe</td>
<td>Par</td>
<td>Dig</td>
<td>Col</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Filled Star</td>
<td>Pentose</td>
<td>Ara</td>
<td>Lyx</td>
<td>Xyl</td>
<td>Rib</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Filled Diamond</td>
<td>Deoxynonulosonate</td>
<td>Kdn</td>
<td></td>
<td>Neu5Ac</td>
<td>Neu5Gc</td>
<td>Neu</td>
<td>Sia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flat Diamond</td>
<td>Deoxynonulosonate</td>
<td>Pse</td>
<td>Leg</td>
<td>Aci</td>
<td></td>
<td>4eLeg</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flat Hexagon</td>
<td>Unknown</td>
<td>Bac</td>
<td>LDmanHep</td>
<td>Kdo</td>
<td>Dha</td>
<td>DDmanHep</td>
<td>MurNAc</td>
<td>MurNGc</td>
<td>Mur</td>
<td></td>
</tr>
<tr>
<td>Pentagon</td>
<td>Assigned</td>
<td>Api</td>
<td>Fru</td>
<td>Tag</td>
<td>Sor</td>
<td>Psi</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Monosaccharide Conformation – Chair Configurations

Nearly all Vertebrate Glycans are Built from Only 9 Sugars

3 hexoses*
- Glucose
- Mannose
- Galactose

2 N-acetylhexosamines
- N-acetylglucosamine
- N-acetylgalactosamine
- xylose
- glucuronic acid
- sialic acid
- L-fucose

*all D configuration except fucose
## Mammalian Monosaccharide Abundance

<table>
<thead>
<tr>
<th>Monosaccharides</th>
<th>Abundance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-GlcNAc</td>
<td>31.8</td>
</tr>
<tr>
<td>D-Gal</td>
<td>24.8</td>
</tr>
<tr>
<td>D-Man</td>
<td>18.9</td>
</tr>
<tr>
<td>Neu5Ac</td>
<td>8.3</td>
</tr>
<tr>
<td>L-Fuc</td>
<td>7.2</td>
</tr>
<tr>
<td>D-GalNAc</td>
<td>4.8</td>
</tr>
<tr>
<td>D-Glc</td>
<td>2.5</td>
</tr>
<tr>
<td>D-GlcA</td>
<td>0.3</td>
</tr>
<tr>
<td>D-Xyl</td>
<td>0.1</td>
</tr>
<tr>
<td>L-IdoA</td>
<td>0.1</td>
</tr>
<tr>
<td>Others</td>
<td>1.2</td>
</tr>
</tbody>
</table>
Common Derivatives of the Hexoses

D-Glucose (Glc)
D-Galactose (Gal)
D-Xylose (Xyl)
D-Glucuronic acid (GlcA)

6-Carboxyl group
Oxidation of C6
Removal of C6
Removal of 6-OH
Mirror image
Relationships Between Common Hexoses & N-Acetylhexosaminates
N-Acetylneuraminic Acid: The Most Common Sialic Acid

Carbons C1-C3 are derived from Pyruvate, and C4-C9 are from N-Acetylmannosamine
Glycan Properties for Molecular Recognition & Binding Energy

Glucose

Galactose

Mannose

Sialic Acid
Seven Eukaryotic Sugars & Their Relationship to Glucose

N-acetyl sugars

N-acetyl-d-galactosamine (GalNAc)

N-acetyl-d-glucosamine (GlcNAc)

hexoses

D-Galactose (Gal)

C4 Epimer

C2 N-acetyl

D-Glucose (Glc)

C2 Epimer

C6 acid

D-Mannose (Man)

C2 Epimer

C6 removed

D-Gluconic acid (GlcA)

d-Xylose (Xyl)

2 - axial group

2 - acetamido group

6-carboxyl group

glucose 6-modifications