



# GLYCAN DIVERSITY AND BIODIVERSITY EVOLUTION

**All cells are assembled from four building blocks: nucleic acids, proteins, lipids and carbohydrates or GLYCANS.**

**Glycans are biopolymers characterized by a significant diversity responsible for many biological functions.**

**Glycans stand for a significant part of the Earth biomass (e.g., *cellulose* from plants, and *chitin* from arthropods and fungi).**

**Glycans are the most abundant of all organic substances in living organisms, comprising about two-thirds of the dry weight of the total biomass.**

**The powerful forces of biological evolution, during more than 3 billion years, seem to have failed to generate any living cell that is devoid of a dense and complex array of cell surface glycans.**

**Thus, cell surface glycans seem to be as essential for life as having DNA, RNAs, proteins, lipids and different metabolites.**

Although we use the term *GLYCAN*, several other names for sugar polymers are found in the literature.

The former term “carbohydrates” or “hydrates of carbon” (coined more than 100 years ago) describes the naturally occurring substances with the general formula  $C_n(H_2O)_n$ .

The simplest of these polyhydroxylated carbonyl compounds (holding either an aldehyde or a ketone group) is called *monosaccharide*.

The *oligosaccharide* refers to any glycan that contains 2–20 monosaccharides connected by *glycosidic bonds*.

A *polysaccharide* (e.g., cellulose, starch, glycogen) is typically any linear or branched polymer consisting of monosaccharides.

The term *glycoconjugate* is often used to describe a macromolecule that contains monosaccharides covalently linked to proteins or lipids.

**The genetic code is virtually the same in all life forms, and gene transcription and energy generation tend to be conserved in various taxa.**

**On the contrary, glycans show a significant diversity of their structure and expression, both within and between evolutionary lineages.**

**For example, structurally, the glycans expressed by most prokaryotes have relatively little in common with those of eukaryotes (an exception occurs when microbial pathogens mimic eukaryotic host structures).**

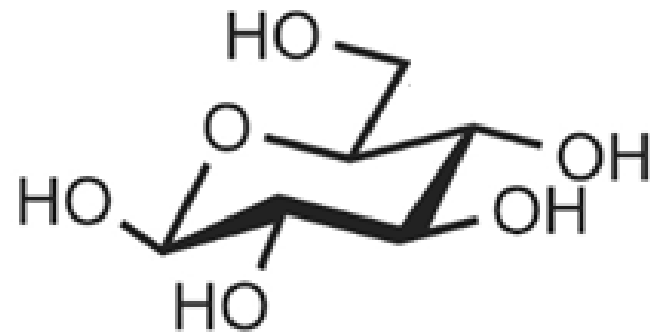
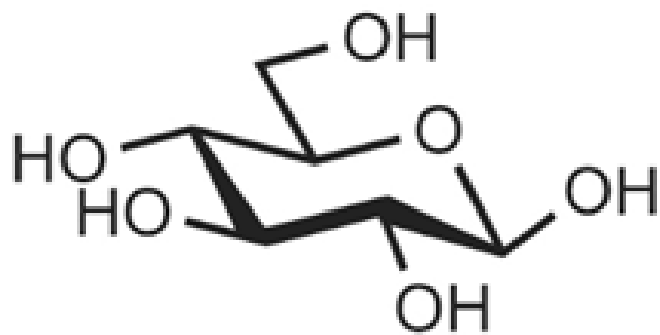
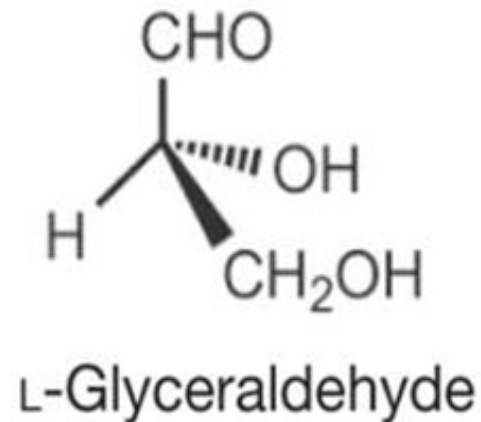
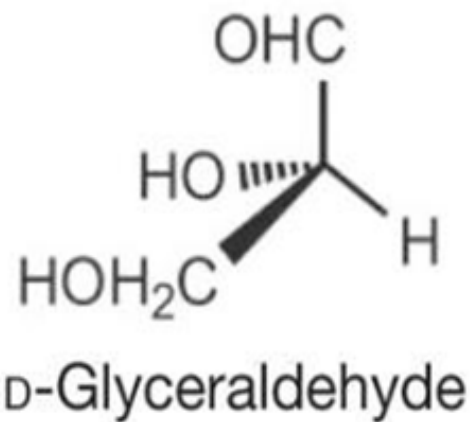
**Phosphodiester and peptide bonds lead to a linear structure defined completely by the sequence; its permutations are the sole source of coding capacity.**

**This situation is completely different for glycans.**

**But, for glycans there are 4 additional parameters underlying their superiority in coding capacity:**

- variability in linkage points (1–2, 1–3, 1–4 or 1–6);
- variability in anomeric position ( $\alpha$  or  $\beta$ );
- variability in ring size (pyranose or furanose);
- introduction of branches.

**In terms of coding capacity and given in absolute numbers,  $6.4 \times 10^7$  hexapeptides are possible with 20 different amino acids, a quantity effectively overshadowed by the  $1.44 \times 10^{15}$  theoretically possible hexasaccharides.**



D-Glucose

L-Glucose

**Why vertebrates use only such a small fraction of monosaccharide types for the assembly of their glycans remains a mystery.**

## **Principal building blocks of vertebrate glycans:**

**Sialic acids: *N*-acetylneuraminic acid (Neu5Ac), *N*-glycolylneuraminic acid (Neu5Gc);**

**Hexoses: Glucose (Glc), mannose (Man), galactose (Gal);**

**Hexosamines: *N*-acetylglucosamine (GlcNAc), *N*-acetylgalactosamine (GalNAc);**

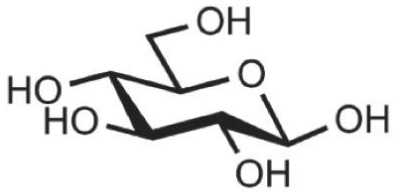
**Deoxyhexoses: Fucose (Fuc);**

**Pentoses: Xylose (Xyl);**

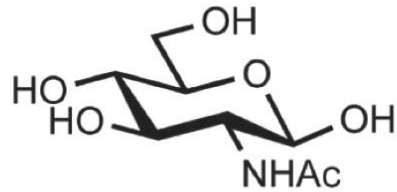
**Uronic acids: Iduronic acid (IduA), Glucuronic acid (GlcA).**



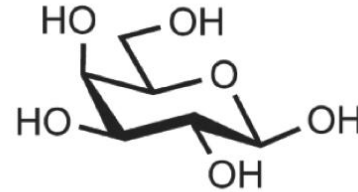
# Common monosaccharides found in vertebrates



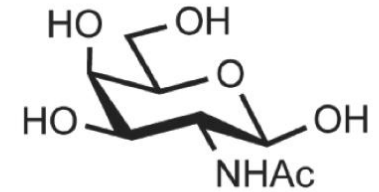
D-Glucose  
(Glc)



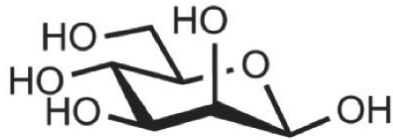
*N*-Acetyl-D-glucosamine  
(GlcNAc)



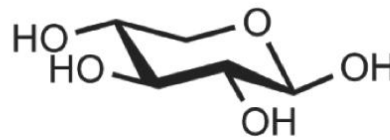
D-Galactose  
(Gal)



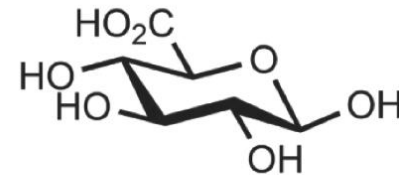
*N*-Acetyl-D-galactosamine  
(GalNAc)



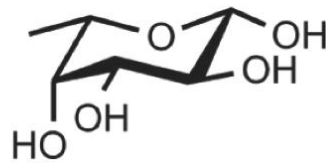
D-Mannose  
(Man)



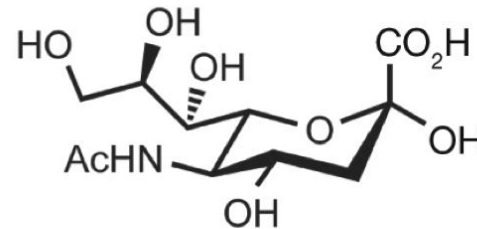
D-Xylose  
(Xyl)



D-Glucuronic acid  
(GlcA)

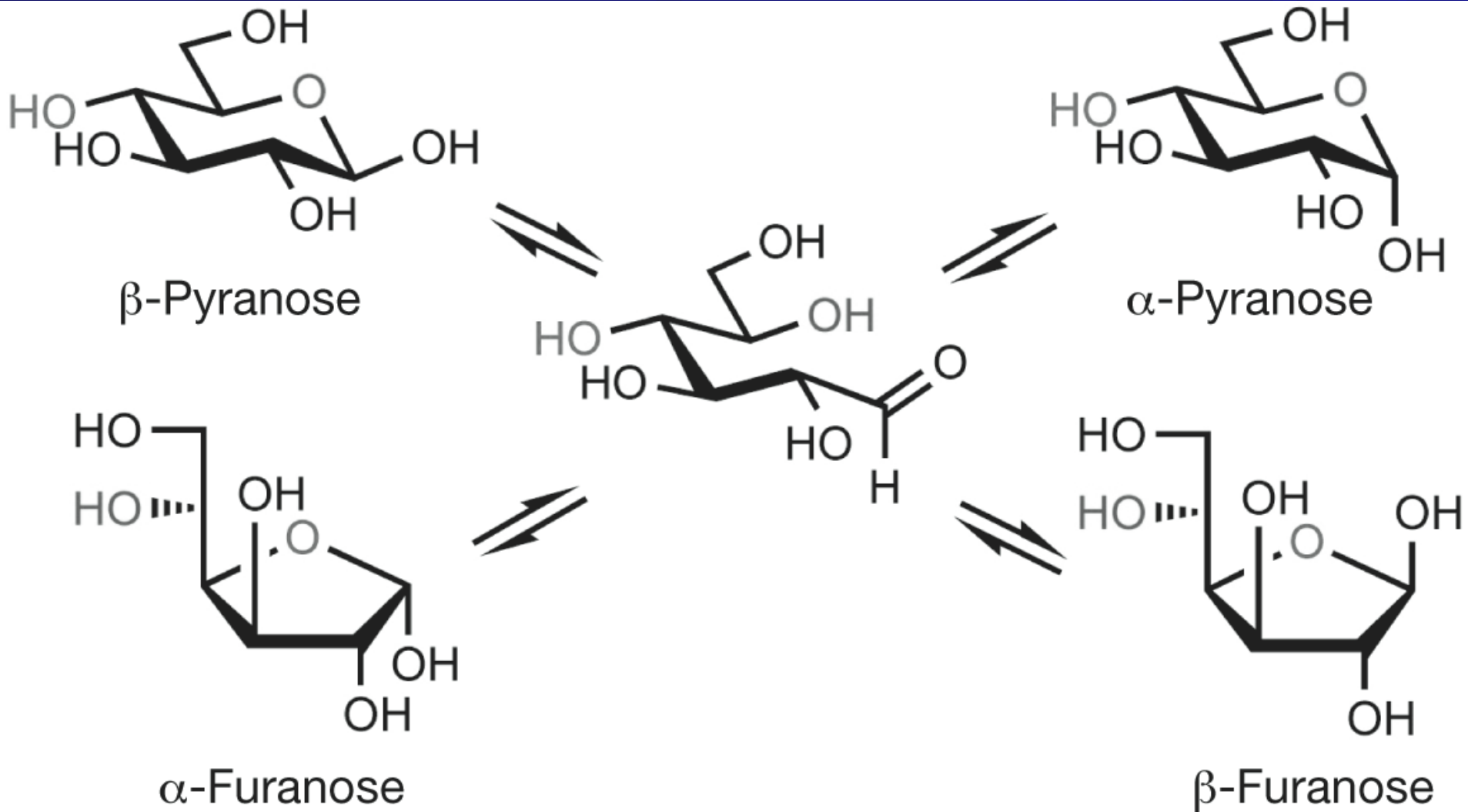


L-Fucose  
(Fuc)

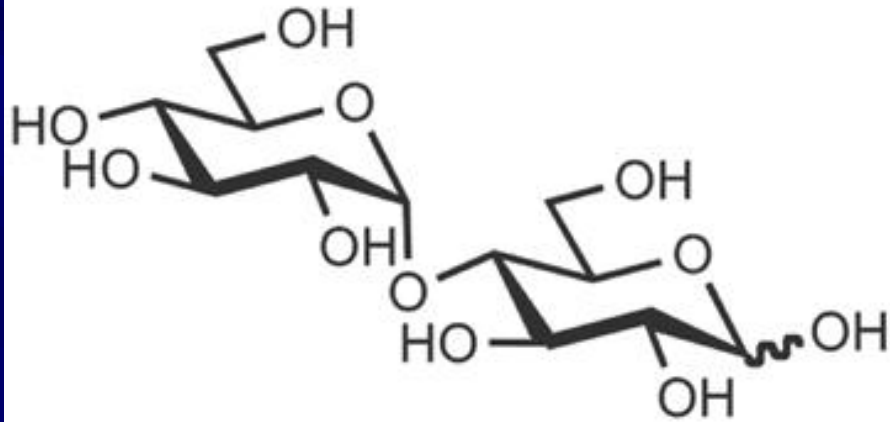


*N*-Acetylneuraminic acid  
(NeuAc)

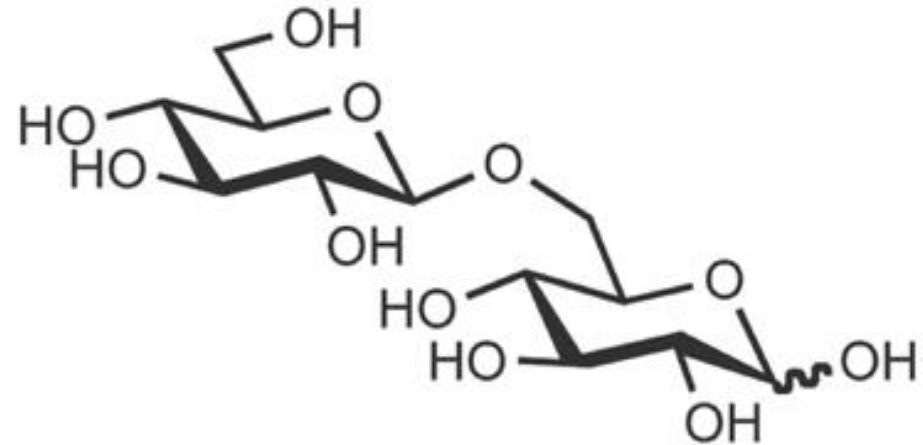
# Cyclization of the acyclic form of D-glucose to form pyranose and furanose structures



# Two isomeric disaccharides



Glc $\alpha$ 1-4Glc  
(maltose)



Glc $\beta$ 1-6Glc  
(gentiobiose)

**For many taxa, essentially no information is available on their glycan profiles.**

**There is no universal “glycan structure code” similar to the genetic code.**

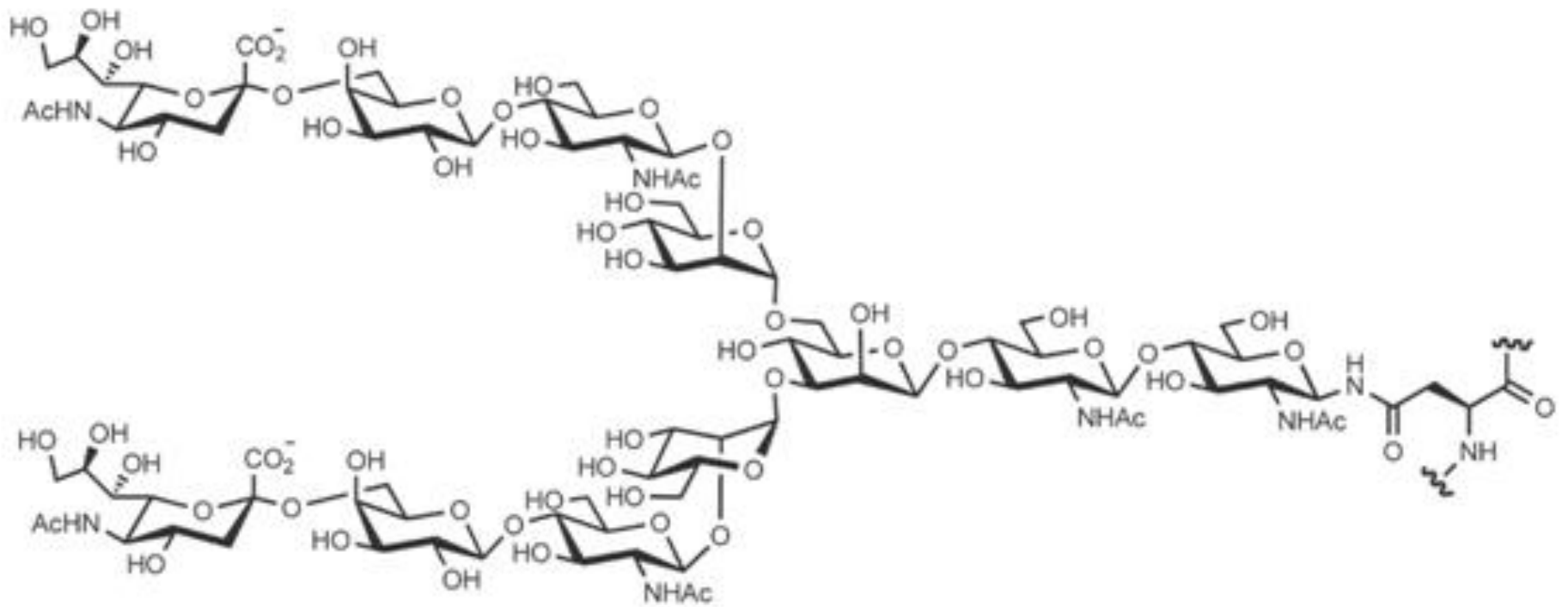
**For example, the glycans expressed by most free-living prokaryotes have relatively little in common structurally with those of eukaryotes.**

# Evolutionary variations in glycans: **N-GLYCANS**

The broadest base of evolutionary information concerns asparagine-*N*-linked glycans. All plants and animals seem to share the same early stages of the classic *N*-glycan processing pathway.

Such conservation is logical, given the critical role of this glycan structure in modulating the folding and maturation of newly synthesized glycoproteins in the endoplasmic reticulum.

Yeasts and vegetative slime molds are unable to generate typical “complex-type” *N*-glycans.



$\text{Neu5Ac}\alpha_6\text{Gal}\beta_4\text{GlcNAc}\beta_2\text{Man}\alpha_6$   
 $\text{Man}\beta_4\text{GlcNAc}\beta_4\text{GlcNAc}\beta\text{-Asn}$   
 $\text{Neu5Ac}\alpha_6\text{Gal}\beta_4\text{GlcNAc}\beta_2\text{Man}\alpha_3$

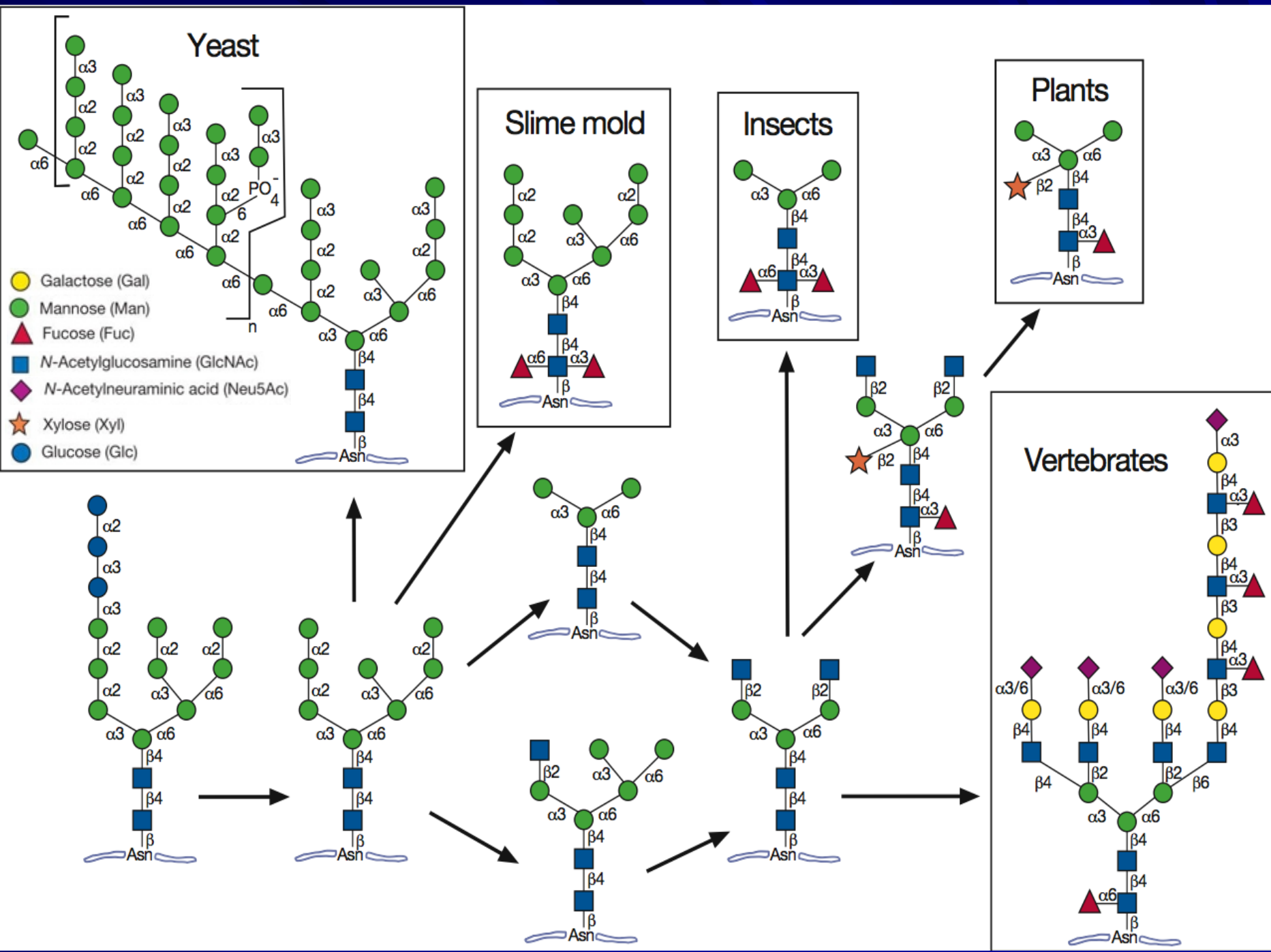
N-Glycan

# Evolutionary variations in glycans: N-GLYCANS

The *N*-glycan pathway is evolutionarily ancient and found throughout the eukaryotes.

But, limited information exists regarding the presence and distribution of most *N*-glycan enzymes and genes in most animals and plants.

Recent data show that Archaea and a few bacteria (*Campylobacter jejuni*) express *N*-glycans linked to asparagine.





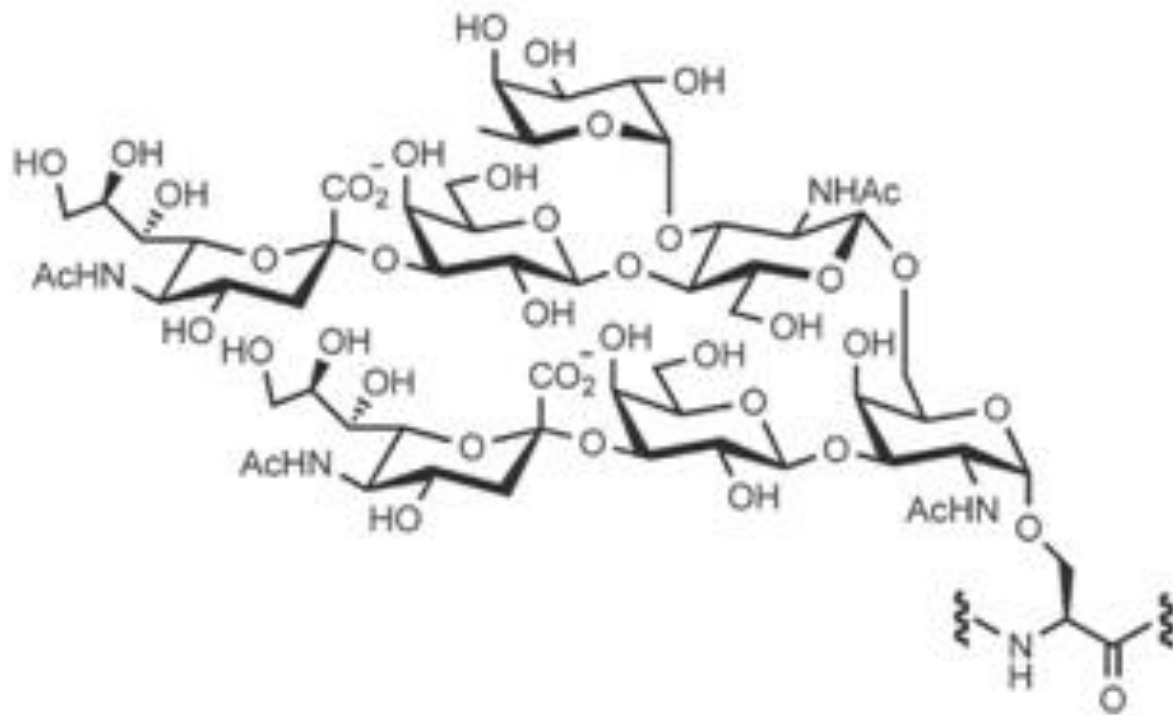
# Evolutionary variations in glycans:

## O-GLYCANS

*N*-acetylgalactosaminyl transferases that initiate synthesis of the most common O-glycans in vertebrates have been found in animal kingdom.

Plants do not appear to have O-linked GalNAc, but they express O-linked galactose to serine and threonine and arabinose to hydroxyproline.

There are few data on bacterial O-glycosylation, although it is clear that novel O-glycans can be found within bacterial “S” layers.



Fuc $\alpha_3$

Neu5Ac $\alpha_3$ Gal $\beta_4$ GlcNAc $\beta_6$

Neu5Ac $\alpha_3$ Gal $\beta_3$ GalNAc $\alpha$ -Ser

O-Glycan

# Evolutionary variations in glycans: GLYCOSPHINGOLIPIDS

*Glucosylceramide* is found in both plants and animals.

*Galactosylceramide* seems to be limited to the nervous system of the deuterostomes of “higher” animals, but all protostome nerves contain mainly glucocerebrosides.

A transition from gluco- to galactocerebrosides corresponds with changes in the nervous system from loosely structured to highly structured myelin.

# **Evolutionary variations in glycans: GLYCOSPHINGOLIPIDS**

**Comparing reptiles to fish and to mammals, an increase in sialic acid content, a decrease in the complexity of ganglioside composition, and a decrease in “alkali-labile” molecules is found.**

**The poikilothermic animals tend to express many polysialylated (more polar) gangliosides in the brain.**

# Evolutionary variations in glycans: SIALIC ACIDS (SAs)

SAs are prominently expressed at the outer termini of *N*-glycans, *O*-glycans, and glycosphingolipids of the deuterostomes; this unique evolutionary innovation originated during the Cambrian Expansion.

*Caenorhabditis elegans*, the free-living nematode, does not contain genes for synthesizing or metabolizing SA.

SA genes derived from genes of the related pathway for 3-deoxy-octulosonic acid synthesis.

A variety of bacteria synthesize other SA-like molecules using a very similar pathway.

# Evolutionary variations in glycans: SIALIC ACIDS

The most SA diversity tends to be found in invertebrates, such as echinoderms (sea urchins and starfish), and the simplest profiles are found in humans.

Although there is a tendency for some types of SAs to be dominant in certain mammalian species, careful investigation reveals the presence of lower quantities of such SAs in most other species.

Humans are “knockout” primates for the enzyme CMP–Neu5Ac hydroxylase (CMAH), because they contain a mutated and inactive *CMAH* gene.

## Evolutionary variations in glycans: GLYCOSAMINOGLYCANS (GAGs)

There are heparan- and chondroitin sulfates in invertebrates, including insects and mollusks.

The chondroitins (not always sulfated), an evolutionarily ancient class, are widely distributed.

The more highly sulfated and epimerized forms of heparin and dermatan sulfate are found primarily in “higher” animals of the deuterostome lineage.

The sponges (the primordial multicellular organisms) have a novel class of glycans, *glyconectins*, completely distinct from GAGs.

Also, plants and bacteria do not have typical animal GAGs.

# **Evolutionary variations in glycans: LIPOPOLYSACCHARIDES (LPSs)**

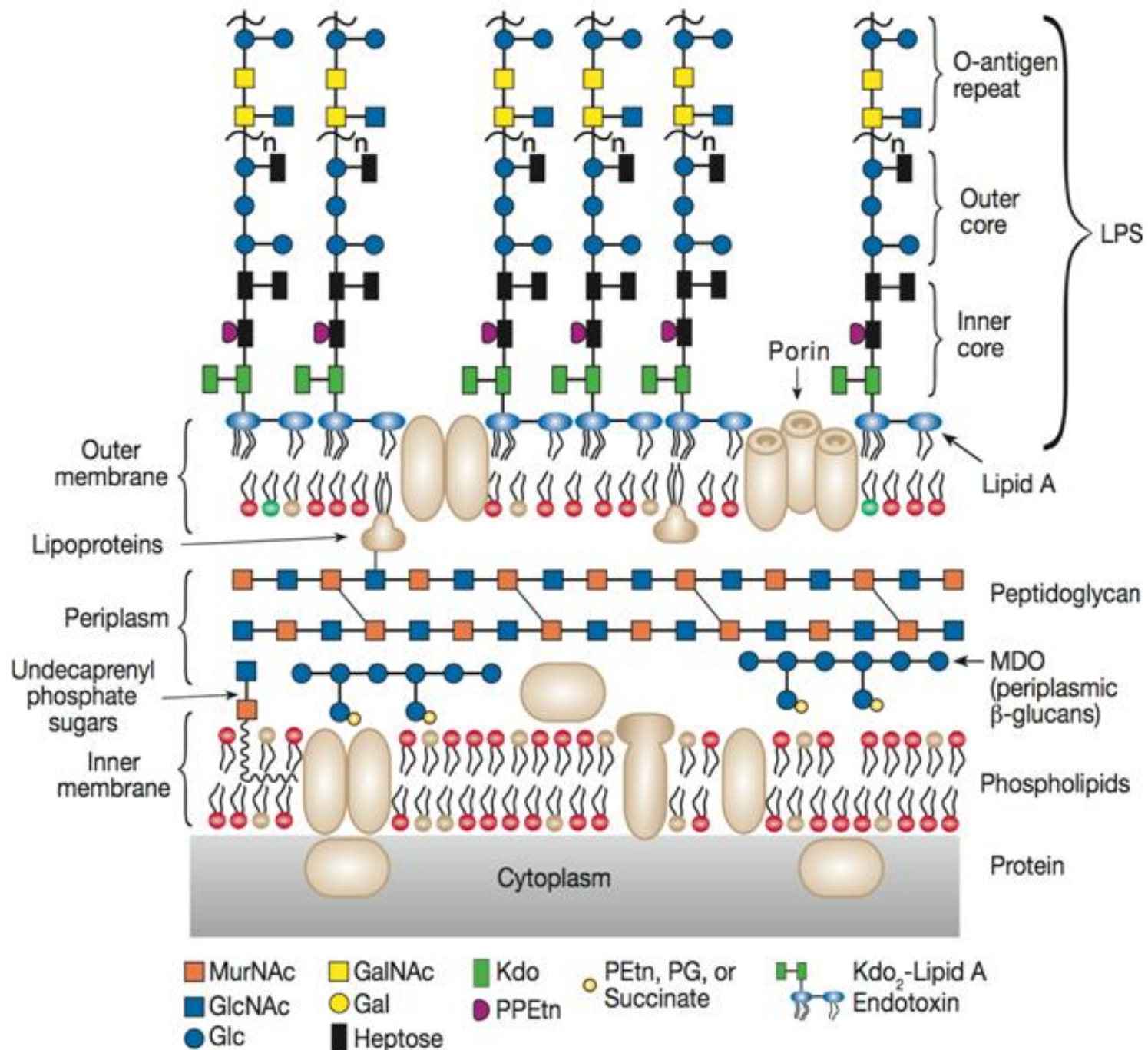
**LPSs are cell-wall components of Gram-negative bacteria.**

**Many bacteria also produce polysaccharides that are excreted outside the cell wall.**

**They may be components of a capsule (capsular polysaccharides) or soluble in the extracellular medium.**

**In either case, they are called exopolysaccharides.**





## Evolutionary variations in glycans: PEPTIDOGLYCAN or MUREIN

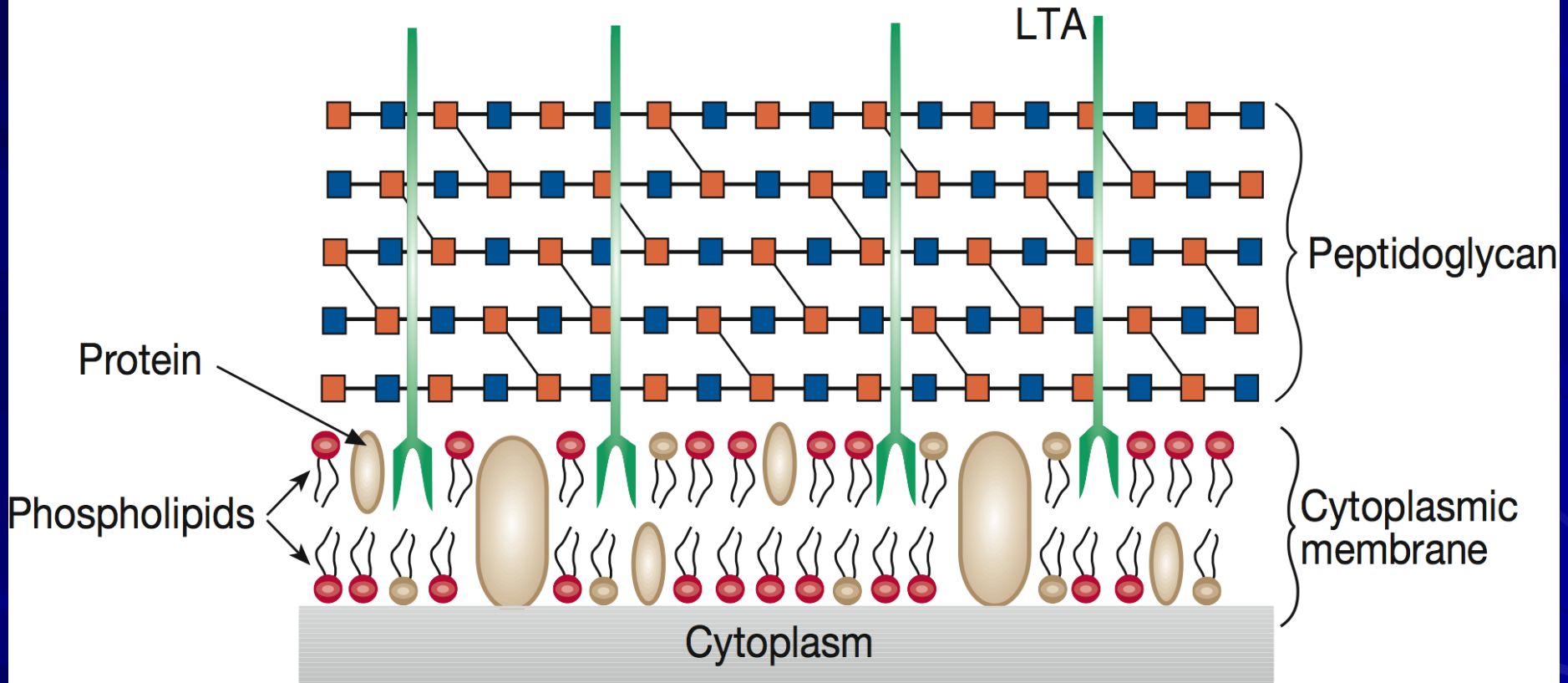
A polymer called peptidoglycan is present in most, perhaps all, bacterial cell walls as the major structural component.

Because it is unique to bacteria, peptidoglycan is one of the most valuable targets to which antibiotics may be directed.

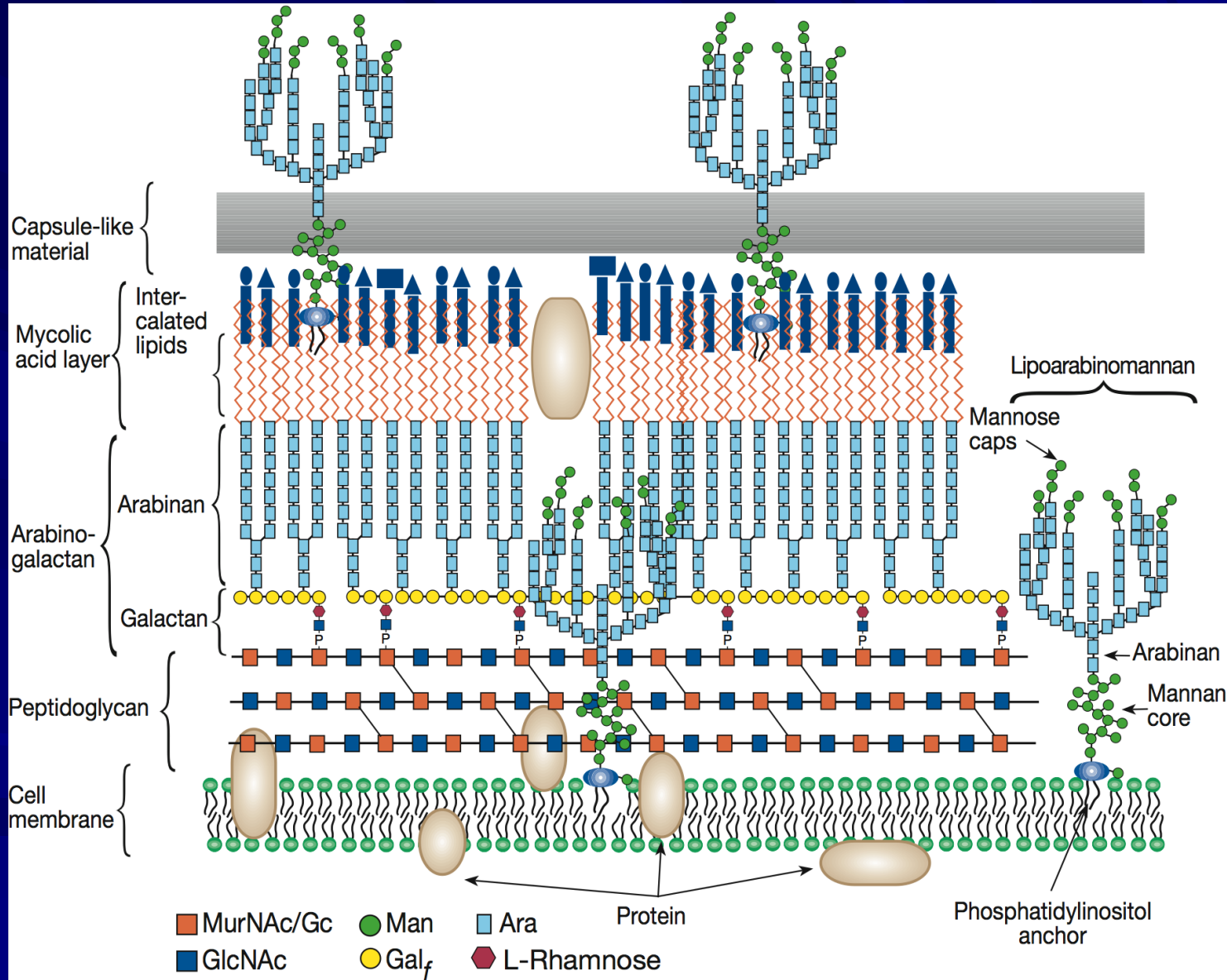
There are over 100 peptidoglycan types, distinguished by different amino acids in the side-chains and different types of cross-linking.

Most *N*-acetylmuramic acid-linked peptides contain a diamino acid in the 3<sup>rd</sup> position of the chain.

# Gram-positive Cell Wall



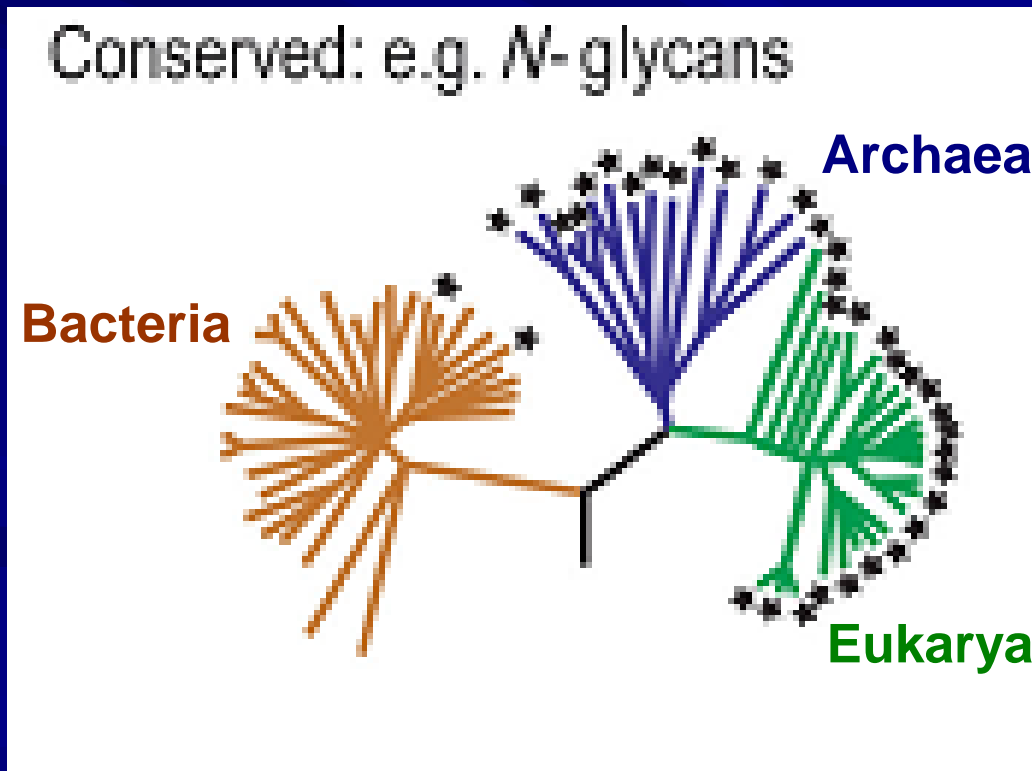
# Structure of the cell wall of mycobacteria



# There are four general glycan patterns:

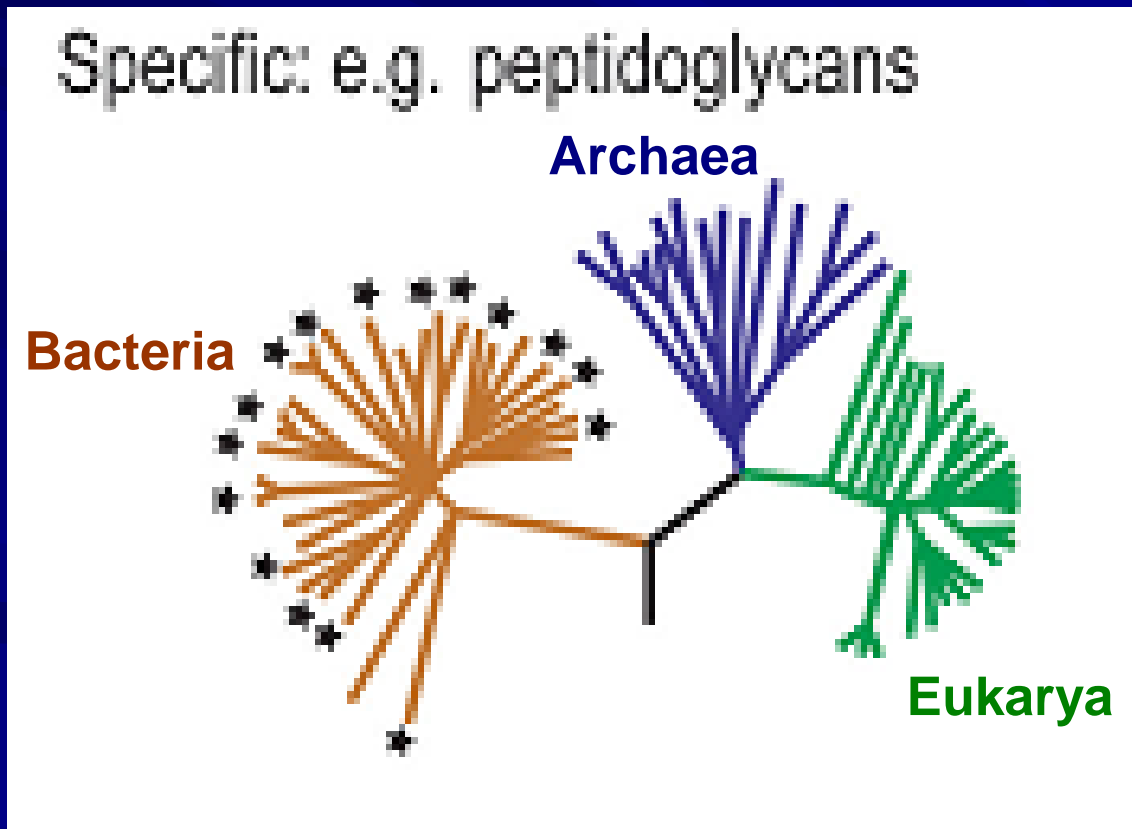
## 1. Glycans conserved across many taxa.

A relatively conserved class of glycan would be *N*-glycans found in organisms of all the three primary lineages of life, excepting many bacteria.



# There are four general glycan patterns:

2. Glycans specific to a particular lineage, such as capsule murein peptidoglycans in bacteria or gangliosides in vertebrates.

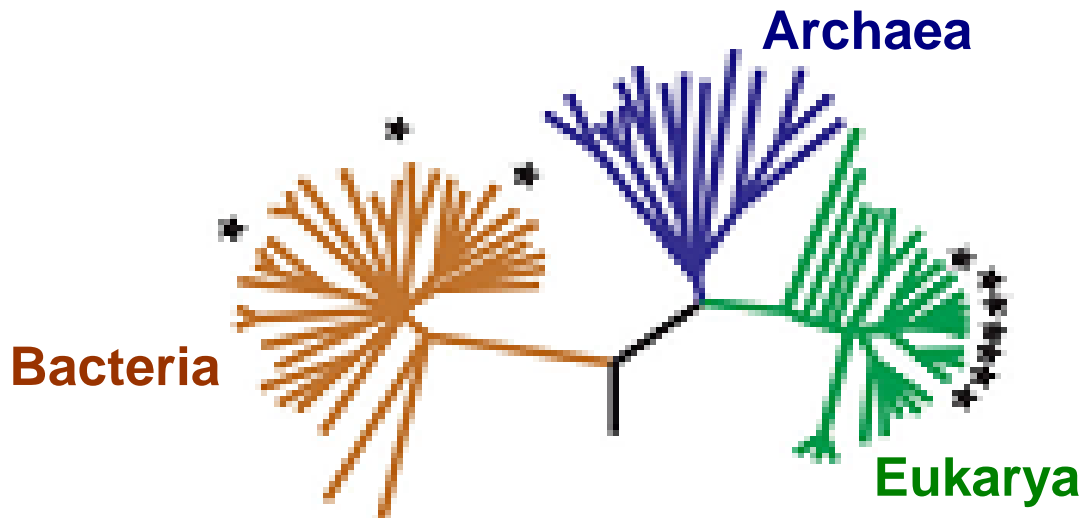


# There are four general glycan patterns:

## 3. Glycans similar across distant taxa:

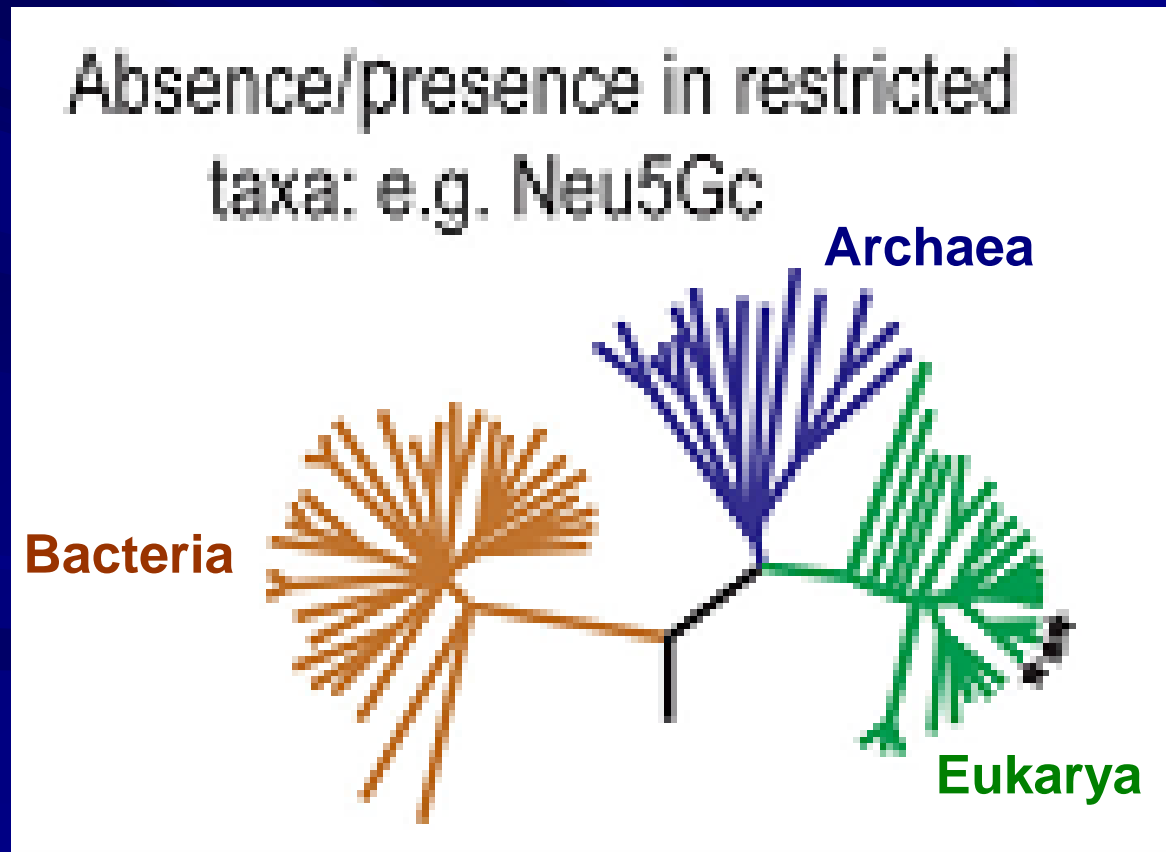
- cellulose in plants, bacteria and tunicates;
- glycosaminoglycans in metazoans and bacteria;
- sialic acids in deuterostomes, many bacteria and cephalopod mollusks (squid and octopus).

Similar across certain distant taxa:  
e.g. glycosaminoglycans



# There are four general glycan patterns:

4. Glycans evidently absent from very restricted taxa only (species, families, or higher units) within lineages that otherwise possess such glycans.





# Why do glycans evolve?

**Divergence:** Like that of any biological molecule, glycan evolution is likely to occur simply due to the divergence of evolutionary lineages.

**Natural selection:** Selective pressures resulting from recognition processes disproportionately affect the glycans covering cell surfaces.

Natural selection acts on glycans, either by favoring the maintenance of a particular glycan (stabilizing or purifying selection) or by diminishing survival and/or reproductive success of organisms carrying a certain glycan (negative selection).

Maintenance of the *N*-glycan synthesis pathway in all eukaryotes is an example of stabilizing selection, since disruptions often lead to lethal consequences.

# Why do glycans evolve?

**Negative selection on glycans could occur whenever an important pathogen exploits a particular glycan as a receptor for infection.**

**Positive selection would entail selection for rapid change in glycans e.g., to accommodate novel endogenous functions.**

**Convergence: Organisms belonging to distantly related lineages recruit or “reinvent” similar subsets of glycan repertoires.**

# Why do glycans evolve?

**Coevolution:** When organisms belonging to different lineages repeatedly interact, as is the case in most natural ecological communities, then their glycomes can become involved in coevolutionary processes.

Thus, the interactions of two distinct glycomes of the interacting lineages directly influence their mutual evolution.

# Variation of animals glycan antigens in time and space

Transient glycan variation in animals has been shown during key processes such as pregnancy, lactation, infection, or acute phase response.

The ontogenetic glycan variation plays key roles in the regulation of metazoan development.

Glycans can also vary in space, as different compartments and adjacent tissues in many animal species carry different glycan repertoires.

## **Genes coding for glycan biosynthetic enzymes contributing to glycan diversity in metazoans**

**A substantial fraction (1–2%) of animal genes operate in glycan biosynthesis and modification.**

**In mammals, there are 9 fucosyl transferases, compared to 4 in *Drosophila* and up to 18 in *Caenorhabditis elegans*.**

**Genetic studies in model organisms with null mutations in biosynthesis genes have proved that many glycans are required for proper metazoan development.**

# Evolutionary forces driving glycan diversification in nature

The glycan diversification in complex multicellular organisms has been driven by intrinsic and extrinsic evolutionary selection pressures.

The glycans are particularly susceptible to the “Red Queen” effect; glycans must keep on changing in order to stay ahead of the pathogens.

The overall diversity in vertebrate cell-surface glycan structure reflects the extrinsic pathogen-mediated selection processes.

# Evolutionary forces driving glycan diversification in nature

More systematic comparative glycobiology could also contribute, by making predictions about intrinsic glycan function.

Such work might also help to a better understanding of the functional significance of glycan diversification during evolution of biodiversity.

The possibility that glycan diversification might even drive the process of speciation (via reproductive isolation) also needs to be considered.