GlycoScience™ & Nutrition

The scientific information in this journal is educational and is not to be used as a substitute for a doctor's care or for proven therapy.

PROVIDING SCIENTIFIC INFORMATION RELATED TO NUTRITIONAL SACCHARIDES AND OTHER DIETARY INGREDIENTS.

NOVEMBER 1, 2003 VOL 4, NO 7

Glycoproteins: Crucial Molecules for Health

Robert K. Murray, MD, PhD

INTRODUCTION

One cannot take glyconutritional supplements for long before encountering the term glycoprotein. The aim of this article is to help the reader obtain a general understanding of glycoproteins and their significance. Digesting and absorbing all of the contents of this article at one sitting may not prove easy, but if you have a genuine interest in learning about glycoproteins, this will hopefully prove a useful starting point. A subsequent article will be published that provides additional information: "Glycoproteins: Critical Molecules in Many Diseases."

The major organic (i.e. carbon-containing) components of human cells are proteins, lipids, various carbohydrates and carbohydrate-containing molecules, and the nucleic acids (DNA and RNA). The carbohydrate-containing molecules can be quite complex; those in which carbohydrates are attached to proteins or lipids are called glycoconjugates. There are three major classes of glycoconjugates: glycoproteins, proteoglycans, and glycolipids. Here we shall discuss only the glycoproteins, although it is certainly worth noting that the proteoglycans and glycolipids also have many implications for health and disease. Water and various ions (e.g., sodium, potassium, calcium and chloride) are the major inorganic components of human cells. Many of the terms used are defined in the Glossary, which is found at the end of this review.

One cannot take glyconutritional supplements for long before encountering the term glycoprotein. The aim of this article is to help the reader obtain a general understanding of glycoproteins and their significance.

ABOUT PROTEINS

We start our account of the glycoproteins by explaining the term protein. This is logical, because glycoproteins are proteins to which one or more sugars have been added. The proteins form the backbones to which the sugars are attached. Thus, one cannot understand glycoproteins without having some knowledge of proteins. Proteins are molecules generally containing all the 20 common amino acids linked together to form a chain by one particular type of bond (the pep-

tide bond). The sequence and number of amino acids can vary quite widely among the proteins in the human body, making each protein unique. Some proteins, such as insulin, contain as few as ~50 amino acid residues. Titin, a muscle protein, is the largest known protein and contains almost 26,000 residues. Each protein folds to form a specific three-dimensional shape (Figure C). A protein must have its correct shape in order to exercise its normal function. The shapes of different proteins vary widely, but all are suited for performing their biologic functions. The factors that determine how proteins fold are presently under intense investigation. If they are elucidated, it may become possible to produce proteins and glycoproteins of almost any desired shape and function. Altering the shape of a protein (e.g., by heating or by changes of pH), a phenomenon known as denaturation, causes it to lose its biologic activity.

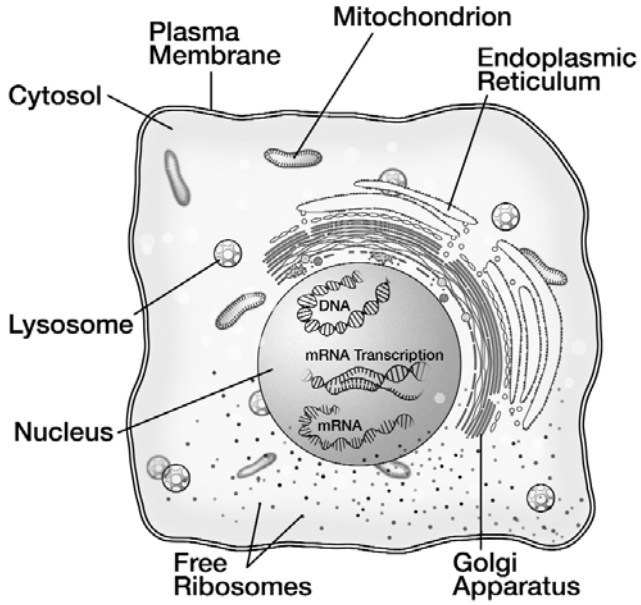
The sequence of amino acids in a protein can now be determined relatively easily, as can the sequence of the fundamental components (the bases) of DNA and RNA. As testimony to this latter statement, The Human Genome Project has practically completed sequencing the 3.2 billion bases present in the human genome. The sequence of amino acids in a protein is determined by the sequence of bases in its gene. It is estimated that there are approximately 30,000 human genes. The number of proteins will be larger, as sometimes one gene can give rise to multiple protein products. The gene (DNA in most cases, except in certain viruses) is copied (transcribed) to form a messenger RNA (mRNA). This acts as a mold (template) on which the amino acids are joined together to form the protein. This flow of information is summarized in the classic formulation (first proposed by Nobelist Francis Crick): DNA ---> mRNA---> protein. However, since the reports of Howard Temin and David Baltimore (also Nobelists) in 1970, it has been known that mRNA can be copied back into DNA by reverse transcriptase. This enzyme is a key constituent of HIV-1 and other retroviruses and an important target for drug therapy in the treatment of AIDS.

Each protein has its own messenger RNA. The mRNAs leave the nucleus and travel to small particles called ribosomes, on which the proteins are synthesized. Some ribosomes reside free in the cytoplasm, and others are bound to the membrane of the endoplasmic reticulum (ER). The proteins made on the free ribosomes leave them to enter the cytoplasm. Those made on the bound ribosomes pursue a longer journey. They first cross the membrane of the endoplasmic reticulum (ER) into the space in its interior. In this so-called intracisternal space, proteins travel to the Golgi apparatus (GA), which acts as a sorting center for newly made proteins. They then leave the GA for various destinations in the cell (Figure A). Alternatively, they can exit the cell to enter the matrix around it or the blood stream, as insulin does when it leaves the pancreatic islet cell in which it is made. There is considerable traffic of newly synthesized proteins inside cells. How do proteins know where to go? The destinations that proteins travel to are generally dictated by particular sequences of amino acids (zip codes, as they are sometimes referred to) in them. The sugar chains of glycoproteins also play a role in directing protein trafficking.

WHAT SHOULD WE KNOW ABOUT GLYCOPROTEINS?

The good news for the reader is that the above information on proteins also applies to glycoproteins. So, if you have followed the above, you already know quite a bit about glycoproteins. The key questions that we now have to consider about glycoproteins include the following: What sugars are present in them and how are they linked to each other? How and where in the cell are glycoproteins synthesized? How does addition of sugars affect the properties of proteins? What kinds of proteins contain sugars? What are the functions of glycoproteins in health? What are lectins, and what are their functions? Other questions may doubtless occur to interested readers, but the above are enough to get us started! In "Glycoproteins: Critical Molecules in Many Diseases" we shall ask what are the involvements of glycoproteins in disease states and how nutritional factors can affect glycoproteins.

Figure A

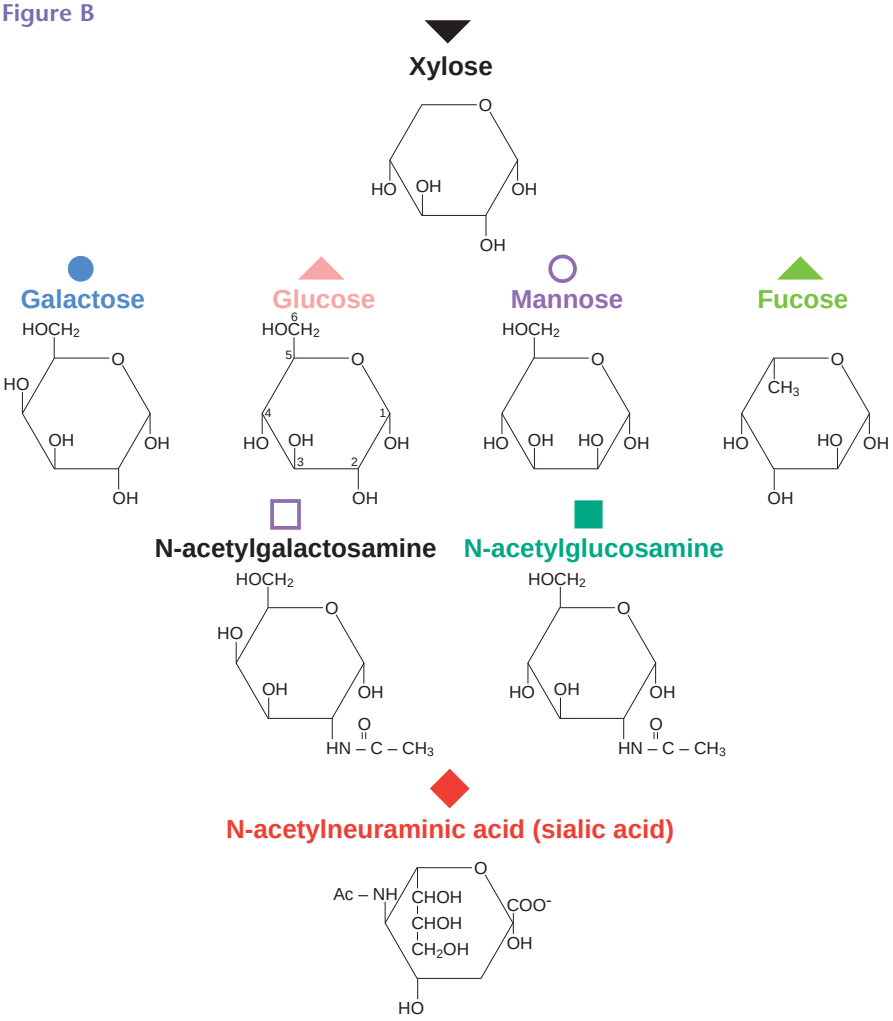


Diagrammatic representation of a human cell. mRNAs are transcribed from DNA in the nucleus. Proteins are then synthesized on free ribosomes and on ribosomes bound to the membranes of the endoplasmic reticulum using the mRNAs as templates. Glycosylation (i.e. sugar attachment) of the proteins occurs in the endoplasmic reticulum and Golgi apparatus.

WHAT SUGARS ARE PRESENT IN GLYCOPROTEINS AND HOW ARE THEY LINKED TO THE PROTEIN BACKBONE?

The principal sugars present in human glycoproteins are listed in Table 1 and illustrated in Figure B. Many readers will be familiar with their names. All are monosaccharides. They contain from 5 to 9 carbon atoms. Note that galactose (Gal), glucose (Glc) and mannose (Man) are very similar, except that they differ from each other by the positions of certain hydroxyl (OH) groups. Galactosamine (GalN), glucosamine (GlcN), and N-acetylneuraminic acid (NeuAc) all contain N groups, and are thus amino sugars. In glycoproteins, the N groups of GlcN and GalN are usually acetylated, so these sugars are referred to as GlcNAc and GalNAc. Glucosamine will be familiar to many readers, as it is widely

Figure B



Structures of xylose, galactose, glucose, mannose, fucose, N-acetylgalactosamine, N-acetylglucosamine and N-acetylneuraminic acid.
These are the most common sugars in glycoconjugates, but others are encountered. Numbering of carbon atoms is shown for glucose; the numbering for others is generally similar. Color coding and geometric shape symbols for sugars in this figure are used to symbolize sugars in selected subsequent figures. (Ac = CH_3 —— CO ——).

used in the treatment of osteoarthritis. Some sugars not listed are found in other species, such as some bacteria. The 8 sugars listed in Table 1 have sometimes been referred to as "The Big Eight." This is a useful device for thinking about them, but one should not be surprised to meet certain other sugars in some glycoproteins!

A glycoprotein may contain as little as one sugar. Others contain many chains of sugars; some are branched and very complex (Figure C). For instance, glycophorin, a major constituent of the membrane of the human red blood cell, contains 16 sugar chains. The linkage between the amino acid of the protein chain that is bonded to the first sugar of a sugar chain is a very important consideration in glycoproteins. It is often used to classify them. Although many types of linkages have been reported, there are two major ones. The first is called an N-type linkage (N = nitrogen) and the second is an O-type linkage (O = oxygen). One reason for stressing the nature of this linkage is that the ways in which the sugar chains of these two types of glycoproteins are built up in a cell are quite different. Some glycoproteins contain both N- and O- linked sugar chains. A third type of linkage, found in a number of human glycoproteins, involves a complex structure containing a glycosylphosphatidylinositol (GPI) linkage; further details on this bond will not be presented here.

WHAT ARE THE BASIC PROPERTIES OF SUGAR CHAINS AND HOW ARE THE SUGARS LINKED TO EACH OTHER?

To understand glycoproteins, we must have some knowledge of their sugar chains. After all, the sugar chains distinguish glycoproteins from proteins and also confer on them many of their biological properties. In this section, we shall make a number of points about sugar chains, which if grasped (sooner or later!) should provide the reader with a reasonable foundation for future reading. They are summarized in Table 2.

1. The sugar chains are called glycans or oligosaccharide chains. We shall use the former term here, mainly because it is shorter. For a number of reasons, the same protein molecule may exhibit heterogeneity of its glycan composition. The variant forms are called glycoforms.
2. The precise natures of the bonds (linkages) between the different sugars of a glycan are critical for the structure and function of glycoproteins. We can illustrate this concept by discussing the sugar lactose, usually the first sugar to cross our lips after birth. Lactose is a disaccharide, made up of two monosaccharides, galactose (Gal), and glucose (Glc). As shown in Figure D, the Gal is linked to Glc by a bond between its carbon atom number 1 and carbon atom number 4 of Glc. The bond joining the two sugars is known to be of the beta (β) configuration, depicted in Figure D. The enzyme that synthesizes lactose makes the beta linkage. For comparison, the structure of maltose, another important disaccharide made up of two Glc units, is also shown. The linkage between carbon atom number one of the first Glc unit and carbon atom number 4 of the second is known to be in the alpha configuration. The enzyme that synthesizes maltose makes the alpha linkage. The beta configuration differs from the alpha configuration by 180 degrees.

Concerning lactose, if any other carbon atom of the Gal (e.g., numbers 2, 3 or 4) were joined to the Glc, the disaccharide would not be lactose. Similarly, if the bond were of the alpha configuration, and not beta, the sugar would not be lactose. The same considerations apply to

Table 1

The Principal Sugars* Found in Glycoproteins		
Name	**Abbreviation**	**No. of carbon (C) atoms**
Xylose	Xyl	5
Fucose	Fuc	6
Galactose	Gal	6
Glucose**	Glc	6
Mannose	Man	6
N-acetylgalactosamine	GalNAc	6****
N-acetylglucosamine	GlcNAc	6****
N-acetylneuraminic acid***	NeuAc	9

* All are monosaccharides.

** Glucose is present in N-linked glycoproteins during their synthesis, but is removed during the processing of glycans (see Figure F). It is not found in O-linked glycoproteins.

*** N-acetylneuraminic acid is a member of a family of 9-carbon sugars known as the sialic acids.

**** Six carbon atoms refer to glucosamine and galactosamine.

Table 2

Summary of Basic Points Concerning Sugar Chains
• They are called glycans or oligosaccharides
• The linkages between sugars in a glycan are critical in determining its structure and function. There is great diversity among glycans because of the many possible linkages, branching and the existence of glycoforms.
• Enzyme-catalyzed addition of sugars to protein or other glycoconjugates is called glycosylation
• The enzymes that add sugars to a glycoconjugate are called glycosyl transferases
• The enzymes that split off sugars from glycans are called glycosidases
• Mutations in genes encoding the above two classes of enzymes can result in various diseases
• Glycans contain biologic information and are recognized by other molecules, including protein and other glycans. The interaction of a glycan with an appropriate molecule can lead to a biologic response.

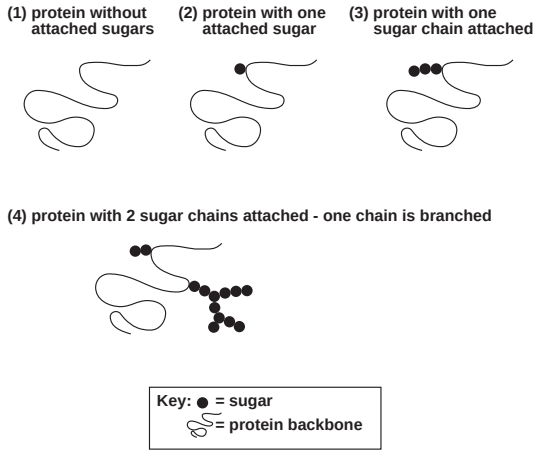
maltose. For instance, if the bond joining the two Glc units were of the beta configuration, the disaccharide would be a sugar other than maltose. The enzyme lactase, present in our intestines, can act on lactose to split the bond between Gal and Glc, thus converting lactose to Gal and Glc. Lactase would not split Gal and Glc bonded by an alpha linkage, nor would it act on maltose. This information illustrates enzyme specificity with regard to both forming and splitting bonds. A further example is given in Figure E, which shows the structure of a relatively small glycan. Each linkage shown is quite specific and each was formed by the action of an individual enzyme. Similarly, a different enzyme splits each linkage. Changes in the linkages would interfere with the structure and function of the glycan shown.

Because of all the hydroxyl (OH) groups present in sugars (see Figure B), sugars are very reactive and can link up with others to form a great number of chemical bonds. For instance, it has been calculated that three different sugars can link with each other to form more than 1,000 different trisaccharides. In a human cell, only a small number (if any) of trisaccharides would be formed, because of the limited number of enzymes present that make trisaccharides that are not part of glycans. It is the great number of possible linkages present in glycans, their branching, and the existence of glycoforms that has slowed progress in determining their structures. However, new methods for sequencing and synthesizing glycans are now available, so progress is likely to be much more rapid in the future. In contrast, the linkages in proteins and nucleic acids are much more limited and branching does not occur. Thus, progress in sequencing them has been much more rapid.

Because we have just discussed lactose and lactase, it is worth noting that the condition known as lactase deficiency is due to a genetically determined deficiency of the enzyme lactase in the small intestine. In this condition, lactose is not split to Gal and Glc in significant amounts, and it accumulates and passes into the large intestine. Therein it is acted upon by bacterial enzymes, causing the production of various acids that lead to contractions of the colon, cramps, and diarrhea. Lactase deficiency can be treated with lactose-free milk or tablets containing lactase.

3. The process of adding sugars to a protein or another glycoconjugate is called glycosylation. It is directed and regulated by specific enzymes and occurs in several locations in the cell (see Figure F). It should be noted that under normal conditions, small amounts of sugars, mainly glucose, can be added to proteins by a reaction that does not require an enzyme. This is called glycation and is of importance in causing tissue damage in diabetes.
4. New glycans are continually being synthesized in cells.

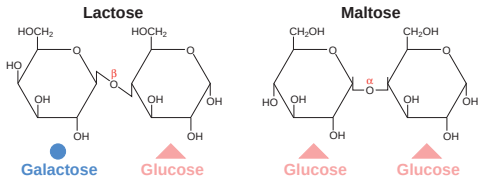
Figure C



Diagrammatic representations of (1) a protein (no glycan attached) and (2)-(4) several glycoproteins, illustrating the variable length and variable number of glycans that can be attached to these molecules. Note that one of the two glycans shown attached to glycoprotein (4) is branched. Many glycans are very large molecules, containing many more sugars than shown here. Others are also much more highly branched. Some glycoproteins contain numerous glycans.

Figure D

Structures of two disaccharides, lactose and maltose.



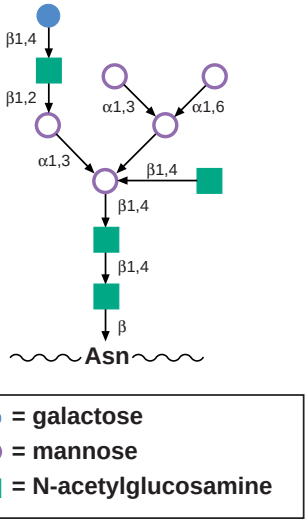
Lactose consists of galactose joined to glucose by a beta (β) 1 -> 4 linkage. Maltose consists of two molecules of glucose joined by an alpha (α) 1 -> 4 linkage. Lactase splits the bond between galactose and glucose, to release these two monosaccharides. Maltase splits the bond between the two glucose molecules, releasing them as free glucose.

The enzymes that add sugars to proteins, glycans, and other molecules are called glycosyl transferases. Many of them are located in the ER and the GA (see Figure F). Over 100 have been identified.

5. Glycans are also being continually degraded. The enzymes that split off sugars from glycans are called glycosidases. Some are located in the ER and GA, some in lysosomes, and yet others elsewhere.
6. Mutations in the genes encoding glycosyl transferases and glycosidases occur. These result in defective enzymes that are often incapable of performing their normal cellular functions, which can cause various genetic diseases. Some that are the result of mutations affecting enzymes involved in the synthesis of the glycans of glycoproteins are named congenital disorders of glycosylation, CDGs. They are discussed in "Glycoproteins: Critical Molecules in Many Diseases." Other conditions, due to abnormalities of glycosidases, lead to the accumulation in various organs of the molecule on which the normal enzyme acts, but which is no longer degraded. Many glycosidases are located in lysosomes (Figure A). For these reasons, many of these conditions are designated as lysosomal storage diseases. They affect not only glycoproteins, but also proteoglycans (e.g., Hurler syndrome) and glycolipids (e.g., Gaucher disease and Tay-Sachs disease).
7. Last but not least, some if not all glycans contain biologic information which may be integral to the normal functioning of the glycoproteins that contain them. The first clear indication that glycans contain information was derived from a series of experiments by Gilbert Ashwell and colleagues in the 1970s. They showed that removing NeuAc from them markedly shortened the life of certain plasma proteins in the blood stream. This was achieved by use of the enzyme neuraminidase. Removal of NeuAc exposed underlying molecules of Gal in the glycans of these proteins. These Gal molecules (unlike the NeuAc molecules) were recognized by a specific receptor (a protein) in the liver, leading to rapid removal of the plasma proteins from the blood. These and other experiments showed that the recognition of specific sugars in glycans by other molecules (e.g. proteins and other glycans) is of fundamental importance in helping to explain how glycoproteins perform their many functions. The points outlined above also apply to many aspects of the glycans of proteoglycans and glycolipids.

Figure E

Structure of a branched N-glycan.



The glycan is attached to the protein via the N-linkage between GlcNAc and asparagine (Asn). Enzymes (glycosyl transferases) build up each linkage and other enzymes (glycosidases) split them. Note that some bonds are α and others β and the positions on the sugars involved in the linkages vary.

HOW AND WHERE IN THE CELL ARE GLYCOPROTEINS SYNTHESIZED?

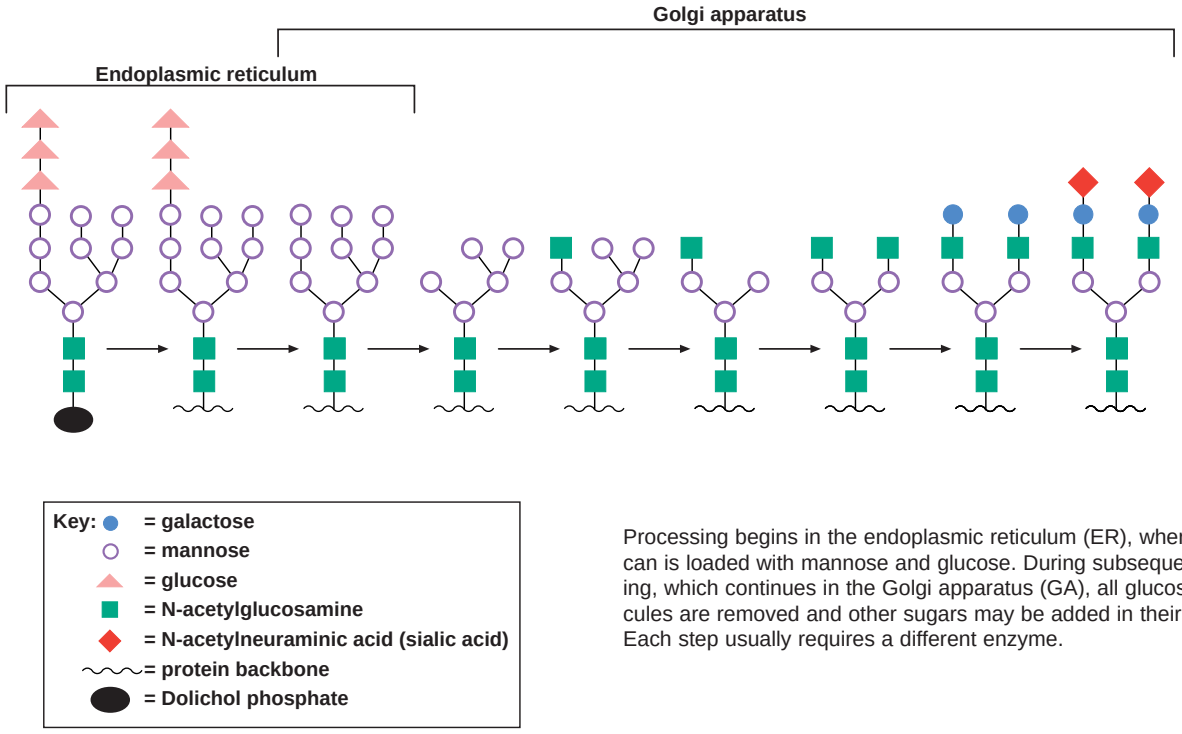
We have already addressed certain aspects of these two questions. As mentioned earlier, there are two major classes of glycoproteins, N- and O- linked. Their protein backbones are synthesized on ribosomes and are then glycosylated. The mechanisms of glycosylation of the two classes differ. In the N-linked class, a glycan consisting of 14 sugars is first built attached to a particular lipid, and the glycan is then transferred in one step to the protein backbone. A GlcNAc present at one end of the glycan is joined to a nitrogen atom of the amino acid asparagine (Asn), forming the N-type linkage. Then, surprisingly, a number of sugars are removed and, in certain glycoproteins, additional new sugars are added back on to the glycan (so-called glycan processing). Each reaction, including the build-up of the glycan containing 14 sugars, its transfer to protein, and subsequent processing, is catalyzed by an enzyme (Figure F). The initial transfer of the glycan occurs as the protein is moving through the membrane of the ER into the intracisternal space (Figure A). Subsequent processing reactions occur in the ER and the several sub-compartments of the GA.

The mode of synthesis of O-linked glycans is simpler. Individual sugars are added one at a time elongating the O-linked chains. These reactions are catalyzed by various glycosyl transferases (one enzyme:one linkage), located in the ER and GA. Processing of O-linked chains does not occur.

Variations in the synthesis of glycans are important. For instance, certain enzymes destined for lysosomes have a specific identity tag (mannose-6-phosphate) attached to them. If addition of this tag is affected by a genetic problem, a condition known as I-cell disease results, because a number of enzymes do not reach the lysosomes. Overall, glyco-

Figure F

Processing of an N-linked glycan.



Processing begins in the endoplasmic reticulum (ER), where the glycan is loaded with mannose and glucose. During subsequent processing, which continues in the Golgi apparatus (GA), all glucose molecules are removed and other sugars may be added in their place. Each step usually requires a different enzyme.

proteins can find their way through the cell to many sites, both inside and outside the cell (Figure G). The surface membrane (plasma membrane), in particular, contains many glycoproteins (Figure H) and is surrounded by a fuzzy layer known as the glycocalyx, which is also rich in these molecules and other glycoconjugates. Others that are secreted by the cell enter the extracellular matrix or the blood stream.

HOW DOES GLYCOSYLATION AFFECT THE PROPERTIES OF PROTEINS?

Various experiments have shown that glycosylation of proteins can, in many cases, significantly alter the properties of these proteins. They include chemical and physical properties such as solubility, shape, viscosity, and resistance to digestion by enzymes (proteinases) that digest proteins. Biologic properties are also affected, because glycosylation or changes in glycosylation can confer new biologic information on a protein, as discussed above. A specific type of glycan has been shown to play an important role in helping to ensure the correct folding and normal trafficking inside cells of a number of glycoproteins. Certain hormones are glycoproteins (e.g., luteotrophin and follitropin). Their glycans are important in the interactions of these hormones with receptors present on the surface membranes of the cells on which they act. The bottom line is that many chemical, physical, and biologic properties of proteins are affected by glycosylation. In particular, the addition of sugars greatly expands the ability of proteins to be recognized specifically by other molecules. It should be noted, however, that many human proteins are never glycosylated, so that addition of sugars is not essential for the normal structure and function of all proteins.

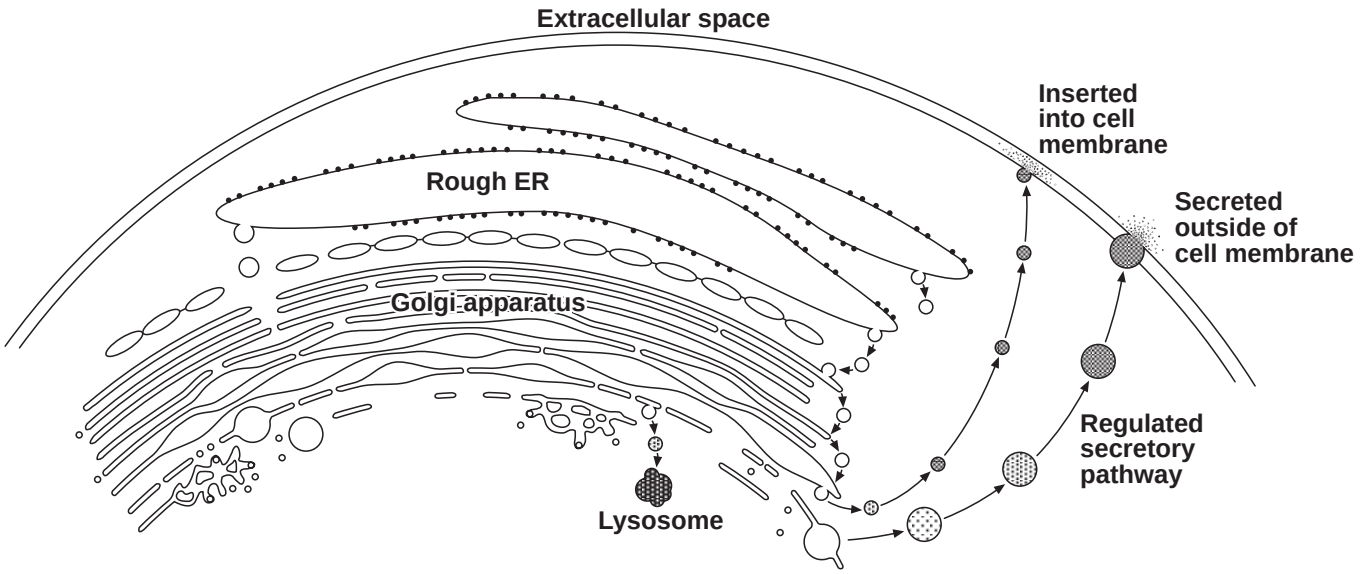
WHAT ARE THE FUNCTIONS OF GLYCOPROTEINS?

Proteins are the molecules that perform most of the work done in organisms, including humans. They have an enormous range of functions. Similarly, glycoproteins also have a wide range of functions; examples are listed in Table 3. Some structural proteins, such as certain types of collagen and also other proteins in the extracellular matrix (ECM), contain glycans. Many proteins involved in cell-cell interactions (i.e. how cells stick together to form tissues and communicate with each other) are glycosylated. Glycoproteins are thought to play an important role in the sperm-egg interactions that occur during fertilization. Cell signaling is the related study of how healthy cells respond appropriately in health to various stimuli (e.g., hormones, growth factors, drugs, cell-cell contact, etc). Conversely, the responses may be inappropriate in certain diseases, such as cancer. Many molecules involved in these responses are glycoproteins (e.g., receptors).

Recent findings have shown that certain of the major signaling pathways that govern normal development involve glycoconjugates. This looks likely to be a very “hot” area of future research. The probable importance of glycoproteins in development of the nervous system has been highlighted by the fact that mental retardation is a common finding in diseases due to abnormalities of glycosylation. Most of our plasma proteins (plasma is the fluid component of our blood) are glycosylated, including the clotting factors and lipoproteins. Many key molecules of the immune system are glycosylated (e.g., the immunoglobulins [antibodies]), and carbohydrate-containing molecules play major roles in its normal function. Ion channels regulate the passage of ions such as sodium, potassium, calcium, and chloride into and

Figure G

Overview of possible destinations of glycoproteins.



Glycoproteins released from the Golgi apparatus (GA) can be used for many purposes. For example, they may become components of lysosomes, be secreted by the cell, or inserted into the plasma membrane.

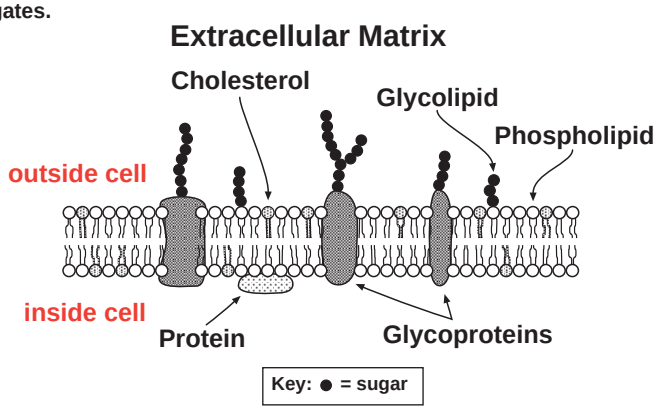
out of cells. It is absolutely essential for maintenance of health that the correct balance of these ions be maintained inside and outside cells. Again, some of these proteins contain carbohydrates. Mucins are proteins involved in lubrication and other functions in various parts of the body, such as the gastrointestinal and genitourinary tracts. Mucins are usually heavily glycosylated.

WHAT ARE LECTINS AND WHAT ARE THEIR FUNCTIONS?

Lectins are proteins that selectively bind glycans. For example, one lectin will recognize and bind mannose at the end of a glycan, whereas others will recognize and bind to glycans containing fucose, galactose, or other sugars. The sugars recognized by lectins bind to specific amino acids in the lectins, forming glycan-lectin complexes. Lectins have numerous functions. Some (e.g., calnexin) participate in the folding of glycoproteins into their correct shapes inside the ER. Others are thought to play a role in directing glycoproteins to specific sites in the cell. Yet others are involved in cell adhesion. In this regard, a class called selectins has been much studied. For instance, certain selectins help direct leukocytes (white blood cells) to sites of inflammation. They do this by being involved in adhesion between the surfaces of white cells and the lining cells of small blood vessels. Because of their important role in inflammation, a major factor in many diseases, selectins will be discussed further in "Glycoproteins: Critical Molecules in Many Diseases." The protein that is responsible for clearance of some glycoproteins from plasma (see the above discussion of Ashwell's experiments) is a lectin, recognizing and interacting with terminal galactose molecules in glycans. Many receptors interact with the glycan chains of glycoproteins, and are thus lectins. Overall, lectins are key players in glycobiology, because they recognize glycans specifically and interact with them, often initiating important cellular responses.

Figure H

Diagrammatic representation of the plasma membrane (PM) of a human cell, showing the locations of a number of glycoconjugates.



Several glycoproteins and glycolipids are shown; their glycans point outward, facing the extracellular matrix (ECM). A non-glycosylated protein is shown on the inner surface. The major lipids of the PM are phospholipids and cholesterol.

FUTURE DIRECTIONS

The Human Genome Project is having a major impact on biology and medicine. In particular, it is accelerating work on human glycomics. This is the characterization of the entire number of glycans present in humans (the glycome). The exact number of glycans is not known, but one thousand (many of which remain to be studied) is likely a conservative estimate. Parallel work will proceed on completely characterizing all the genes and enzymes involved in glycan synthesis and breakdown and on the proteins and other

Table 3

Some Functions Of Glycoproteins.	
Function	**Examples**
Structural	Some types of collagen; various proteins in the ECM*
Cell-cell interactions	Selectins; various cell adhesion molecules; proteins involved in sperm-egg interactions during fertilization
Cell signaling	Many receptors
Plasma proteins	Clotting factors; lipoproteins
Immunologic	Immunoglobulins (they are also plasma proteins); various other proteins
Ion channels	Cystic fibrosis transmembrane regulator (CFTR)
Mucins	Found in various membranes and in many secretions, such as those in the gastrointestinal, respiratory, and genitourinary tracts.
Protein folding and trafficking	Calnexin; calreticulin

* extracellular matrix

molecules to which the glycans are attached. The outcomes of these studies will greatly illuminate how glycoconjugates, and glycoproteins in particular, are involved in health and disease. The information provided should also serve as a basis for the development of new glyconutrients and new glycotherapies.

SUMMARY

- Glycoproteins are proteins that contain one or more sugars.
- Eight sugars predominate in glycoproteins, but others are also found.
- The nature of the linkage between the protein and carbohydrate is important in classifying glycoproteins. The two major types are referred to as N- and O-linkages.
- Sugar chains attached to glycoproteins or other glycoconjugates, such as proteoglycans and glycolipids, are referred to as glycans. The precise natures of the linkages between sugars in glycans are critical in determining the structures and functions of glycoproteins.
- Glycosylation is the process of adding sugars to a protein, and is carried out by a battery of enzymes (glycosyl transferases). A battery of glycosidases catalyzes removal of sugars from glycans. Mutations in the genes that encode these two classes of enzymes can result in a number of diseases.
- Glycans contain biologic information. That is, the specific sequence of sugars in a glycoprotein can determine its recognition and interaction with other molecules in the cell, often resulting in a biologic effect or response.
- The protein backbones of glycoproteins are synthesized on ribosomes and the glycans are assembled in the endoplasmic reticulum and the Golgi apparatus. The modes of synthesis of N-linked and O-linked glycoproteins are quite distinct.
- Glycosylation affects many chemical and biologic properties of proteins.
- Glycoproteins are especially located on the surface membranes of cells, in body fluids (such as plasma) and in mucins. They are also found in many other sites.
- Because of their wide distribution and numerous functions, glycoproteins are essential molecules in the maintenance of health.
- The new field of glycomics will yield much new information on the roles of glycoproteins and other glycoconjugates in health and disease. This will lead to the development of new glyconutrients and glycotherapies.

FURTHER READING

Maeder T. Sweet Medicines. *Sci Am.* 2002; 287(1):40-47.

McAnalley BH, Vennum E. Glycoscience: State of the science review. *GlycoScience & Nutrition (Official Publication of GlycoScience.com The Nutrition Science Site)*. 2001;2(14):1-8.

Murray RK, Granner DK, Mayes PA, Rodwell VR. *Harper's Illustrated Biochemistry.* 26th Edition. New York, NY: Lange Medical Books/McGraw-Hill, 2003.

Perkel JM. Glycobiology Goes to the Ball. *The Scientist.* 2002;16(9):32.

Schmidt K. Sugar rush. *NewScientist* 2003;176(2366):34.

Hurtley S, Service R, Szuromi P. Cinderella's Coach Is Ready. *Science* 2001;291(5512):2337.

Taylor ME, Drickamer K. *Introduction to Glycobiology*. Oxford, England: Oxford University Press, 2003.

Roush W, Goho AM, Scigliano E, et al. 10 Emerging Technologies That Will Change the World. *Technology Review;* 2003. February:33-49.

Varki A, Cummings R, Esko J, et al. *Essentials of Glycobiology.* Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press, 1999.

GLOSSARY

Amino acid: A specific type of acid; twenty different amino acids comprise the building blocks of proteins. When an amino acid is part of a protein, it is referred to as an amino acid residue.

Amino sugar: A monosaccharide in which a hydroxyl (OH) group is replaced by an amino (NH_2) group. Glucosamine is an example.

Carbohydrate: Originally used to refer to a compound with the general formula $C_n(H_2O)_n$ where n = 3 or more. It now includes molecules that have additional groups (e.g., amino groups) and is used more or less interchangeably with the terms sugar, saccharide and glycan. It includes monosaccharides, disaccharides, oligosaccharides and polysaccharides.

Enzyme: A biologic catalyst, usually a protein. Enzymes greatly speed up reactions, a phenomenon known as catalysis.

Enzyme specificity: The ability of enzymes to recognize and act on only certain compounds. It also includes their ability to form only certain linkages (e.g., alpha or beta).

Extracellular matrix (ECM): The material between and surrounding cells. Contains structural proteins (e.g., collagen), glycoproteins, proteoglycans and other molecules.

Glycan: A general term for a sugar, or assembly of sugars, in free form or attached to another molecule such as a protein or lipid. Used interchangeably with carbohydrate and saccharide.

Glycan processing: The removal of certain sugars from a glycan during biosynthesis of N-linked glycoproteins and often their subsequent replacement by other sugars.

Glycation: Non-enzymatic addition of a sugar (e.g., glucose) to a protein, such as occurs excessively in diabetes mellitus.

Glycobiology: The study of the chemistry and biology of carbohydrates.

Glycocalyx: A thin layer of fuzzy material around the plasma membrane of cells; it contains various glycoconjugates.

Glycoconjugate: A molecule with one or more sugars attached to protein or lipid. The three major classes are glycoproteins, proteoglycans and glycolipids.

Glycoform: Different forms of a glycoprotein, due to variability in the composition and structure of its glycan(s).

Glycolipid: A lipid with one or more sugars attached.

Glycomics: The study of all the glycans (the glycome) in humans and other organisms.

Glycoprotein: A protein with one or more sugars attached.

Glycosaminoglycan (GAG): A biopolymer made up of repeating units of disaccharides, one of which is usually an amino sugar and the other a sugar acid. Heparin and hyaluronate are examples. GAGs usually occur complexed with proteins and are then designated proteoglycans.

Glycosidase: An enzyme that releases one or more sugars from a glycan. For example, neuraminidase is a glycosidase that releases N-acetylneuraminic acid (NeuAc) from glycans.

Glycosylation: Enzyme-catalyzed attachment of a sugar to another molecule, such as a protein or lipid.

Glycosyl transferase: An enzyme involved in glycosylation reactions. Well over one hundred have been identified.

Hexose: A sugar containing 6 carbon atoms (e.g., galactose, glucose or mannose). Xylose (5 C atoms) is a pentose and N-acetylneuraminic acid (9 C atoms) is a nonose.

Ion: A small, charged molecule such as sodium, potassium, calcium or chloride.

Lectin: A protein that binds one or more specific glycans.

Ligand: A molecule that is recognized by and interacts with a specific protein or receptor. For example, specific glycans are ligands for certain lectins.

Lipids: Members of a major class of biomolecules that are relatively insoluble in water but soluble in various organic solvents.

Monosaccharide: A carbohydrate consisting of one simple sugar (e.g., mannose).

Mucin: A type of glycoprotein containing many O-linked glycans.

N-linked glycan: A glycoprotein glycan that is joined to the protein backbone via a nitrogen (N) atom of the amino acid asparagine (Asn).

Oligosaccharide: A chain of up to approximately 10 sugars joined to each other. Glycan is an equivalent term.

O-linked glycan: A glycoprotein glycan that is joined to the protein backbone via an oxygen (O) atom of the amino acids serine (Ser) or threonine (Thr).

Organelle: A structure within a cell (e.g. endoplasmic reticulum, Golgi apparatus, lysosome, mitochondrion, nucleus, etc) that performs specific functions.

Polysaccharide: A large glycan composed of one or more repeating monosaccharides. Starch and glycogen are two important examples, in which the repeating monosaccharide is glucose.

Proteoglycan: A large aggregate of one or more proteins with one or more types of GAG (see above). They are widely distributed in the body, but are often located in the ECM (see above).

Receptor: A protein that binds a ligand (see above), initiating a biologic response.

Saccharide: An equivalent term to carbohydrate (see above). Includes mono-, di-, tri-, oligo- and polysaccharides.

Selectin: A specific type of lectin found in cells lining blood vessels and in circulating white blood cells.

Signaling: Usually referred to as signal transduction or transmembrane signaling. Refers to the processes whereby molecules external to the cell (e.g., hormones or drugs) interact with the cell and lead to a cellular response.

Sugar: A general term more or less equivalent to carbohydrate (see above).