NCBI Bookshelf. A service of the National Library of Medicine, National Institutes of Health.

Berg JM, Tymoczko JL, Stryer L. Biochemistry. 5th edition. New York: W H Freeman; 2002.

Section 11.3 Carbohydrates Can Be Attached to Proteins to Form Glycoproteins

Carbohydrate groups are covalently attached to many different proteins to form *glycoproteins.* Carbohydrates are a much smaller percentage of the weight of glycoproteins than of proteoglycans. Many glycoproteins are components of cell membranes, where they play a variety of roles in processes such as cell adhesion and the binding of sperm to eggs.

11.3.1. Carbohydrates May Be Linked to Proteins Through Asparagine (*N*-Linked) or Through Serine or Threonine (*O*-Linked) Residues

In glycoproteins, sugars are attached either to the amide nitrogen atom in the side chain of asparagine (termed an N-*linkage*) or to the oxygen atom in the side chain of serine or threonine (termed an O-*linkage*), as shown in Figure 11.18. An asparagine residue can accept an oligosaccharide only if the residue is part of an Asn-X-Ser or Asn-X-Thr sequence, in which X can be any residue. Thus, *potential glycosylation sites can be detected within amino acid sequences.* However, which of these potential sites is actually glycosylated depends on other aspects of the protein structure and on the cell type in which the protein is expressed. All *N*-linked oligosaccharides have in common a pentasaccharide core consisting of three mannose and two *N*-acetylglucosamine residues. Additional sugars are attached to this core to form the great variety of oligosaccharide patterns found in glycoproteins (Figure 11.19).

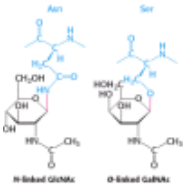


Figure 11.18

Glycosidic Bonds between Proteins and Carbohydrates. A glycosidic bond links a carbohydrate to the side chain of asparagine (*N*-linked) or to the side chain of serine or threonine (*O*-linked). The glycosidic bonds are shown in red.

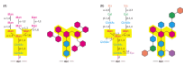


Figure 11.19

N -linked oligosaccharides. A pentasaccharide core (shaded yellow) is common to all *N*-linked oligosaccharides and serves as the foundation for a wide variety of *N*-linked oligosaccharides, two of which are illustrated: (A) high-mannose type; (B) complex (more...)

Abbreviations for sugars —

Fuc Fucose
Gal Galactose
GalNAc *N*-Acetylgalactosamine
Glc Glucose
GlcNAc *N*-Acetylglucosamine
Man Mannose

Sia Sialic acid
NeuNAc *N*-Acetylneuraminate (sialic acid)

Carbohydrates are linked to some soluble proteins as well as membrane proteins. In particular, many of the proteins secreted from cells are glycosylated. Most proteins present in the serum component of blood are glycoproteins (Figure 11.20). Furthermore, *N*-acetylglucosamine residues are *O*-linked to some intracellular proteins. The role of these carbohydrates, which are dynamically added and removed, is under active investigation.

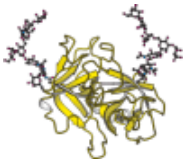


Figure 11.20

Elastase, a Secreted Glycoprotein, Showing Linked Carbohydrates on Its Surface.  Elastase is a protease found in serum. Note that the oligosaccharide chains have substantial size even for this protein, which has a relatively low level of glycosylation. (more...)

11.3.2. Protein Glycosylation Takes Place in the Lumen of the Endoplasmic Reticulum and in the Golgi Complex

Protein glycosylation takes place inside the lumen of the *endoplasmic reticulum* (ER) and the *Golgi complex*, organelles that play central roles in protein trafficking (Figure 11.21). One such glycoprotein (depicted in Figure 11.20) is the proteolytic enzyme elastase (Section 9.1.4), which is secreted by the pancreas as a zymogen (Section 10.5). This protein is synthesized by ribosomes attached to the cytoplasmic face of the ER membrane, and the peptide chain is inserted into the lumen of the ER as it grows, guided by a signal sequence of 29 amino acids at the amino terminus. This signal sequence, which directs the protein through a channel in the ER membrane, is cleaved from the protein in the transport process into the ER (Figure 11.22). After the protein has entered the ER, the glycosylation process begins. The *N*-linked glycosylation begins in the ER and continues in the Golgi complex, whereas the *O*-linked glycosylation takes place exclusively in the Golgi complex.

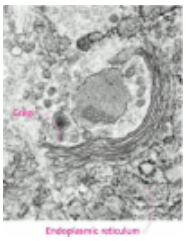


Figure 11.21

Golgi Complex and Endoplasmic Reticulum. The electron micrograph shows the Golgi complex and adjacent endoplasmic reticulum. The black dots on the cytoplasmic surface of the ER membrane are ribosomes. [Micrograph courtesy of Lynne Mercer.]

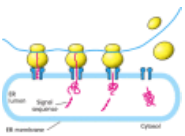


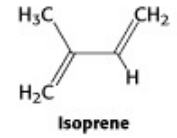
Figure 11.22

Transport Into the Endoplasmic Reticulum. As translation takes place, a signal sequence on membrane and secretory proteins directs the nascent protein through channels in the ER membrane and into the lumen. In most cases, the signal sequence is subsequently (more...)

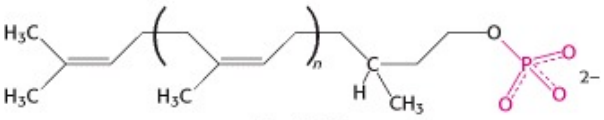
11.3.3. *N*-Linked Glycoproteins Acquire Their Initial Sugars from Dolichol

Donors in the Endoplasmic Reticulum

A large oligosaccharide destined for attachment to the asparagine residue of a protein is assembled attached to *dolichol phosphate*, a specialized lipid molecule containing as many as 20 isoprene (C_5) units (Section 26.4.8).



Isoprene



n = 15–19

Dolichol phosphate

The terminal phosphate group is the site of attachment of the activated oligosaccharide, which is subsequently transferred to the protein acceptor. Dolichol phosphate resides in the ER membrane with its phosphate terminus on the cytoplasmic face.

The assembly process proceeds in three stages. First, 2 *N*-acetylglucosamine residues and 5 mannose residues are added to the dolichol phosphate through the action of a number of cytoplasmic enzymes that catalyze monosaccharide transfer from sugar nucleotides. Then, in a remarkable (and, as yet, not well understood) process, this large structure is "flipped" through the ER membrane into the lumen of the ER. Finally, additional sugars are added by enzymes in the ER lumen, this time with the use of monosaccharides activated by attachment to dolichol phosphate. This process ends with the formation of a 14-residue oligosaccharide attached to dolichol phosphate (Figure 11.23).

Figure 11.23

Assembly of an *N* -linked oligosaccharide precursor on dolichol phosphate. The first stage of oligosaccharide synthesis takes place in the cytoplasm on the exposed phosphate of a membrane-embedded dolichol molecule. Synthesis of the precursor is completed (more...)

The 14-sugar-residue precursor attached to this dolichol phosphate intermediate is then transferred en bloc to a specific asparagine residue of the growing polypeptide chain. In regard to elastase, oligosaccharides are linked to the asparagine residues in the recognition sequences Asn 109-Gly-Ser and Asn 159-Val-Thr. Both the activated sugars and the complex enzyme that is responsible for transferring the oligosaccharide to the protein are located on the lumenal side of the ER, accounting for the fact that proteins in the cytosol are not glycosylated by this pathway. Before the glycoprotein leaves the lumen of the ER, 3 glucose molecules are removed from the 14-residue oligosaccharide. As we will see in Section 11.3.6, the sequential removal of these glucose molecules is a quality-control step that ensures that only properly folded glycoproteins are further processed.

Dolichol pyrophosphate released in the transfer of the oligosaccharide to the protein is recycled to dolichol phosphate by the action of a phosphatase. This hydrolysis is blocked by

bacitracin, an antibiotic. Another interesting antibiotic inhibitor of *N*-glycosylation is *tunicamycin*, a hydrophobic analog of the sugar nucleotide uridine diphosphate-*N*-acetylglucosamine (UDP-GlcNAc), the activated form of *N*-acetylglucosamine used as a substrate for the enzymes that synthesize the oligosaccharide unit on dolichol phosphate. Tunicamycin blocks the addition of *N*-acetylglucosamine to dolichol phosphate, the first step in the formation of the core oligosaccharide.

11.3.4. Transport Vesicles Carry Proteins from the Endoplasmic Reticulum to the Golgi Complex for Further Glycosylation and Sorting

Proteins in the lumen of the ER and in the ER membrane are transported to the Golgi complex, which is a stack of flattened membranous sacs. The Golgi complex has two principal roles. First, *carbohydrate units of glycoproteins are altered and elaborated in the Golgi complex.* The *O*-linked sugar units are fashioned there, and the *N*-linked sugars, arriving from the ER as a component of a glycoprotein, are modified in many different ways. Second, *the Golgi complex is the major sorting center of the cell.* Proteins proceed from the Golgi complex to lysosomes, secretory granules (as is the case for the elastase zymogen), or the plasma membrane, according to signals encoded within their amino acid sequences and three-dimensional structures (Figure 11.24).

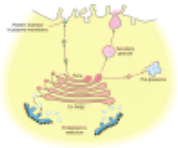


Figure 11.24

Golgi Complex as Sorting Center. The Golgi complex is the sorting center in the targeting of proteins to lysosomes, secretory vesicles, and the plasma membrane. The cis face of the Golgi complex receives vesicles from the ER, and the trans face sends (more...)

The Golgi complex of a typical mammalian cell has 3 or 4 membranous sacs (cisternae), and those of many plant cells have about 20. The Golgi complex is differentiated into (1) a *cis* compartment, the receiving end, which is closest to the ER; (2) *medial* compartments; and (3) a *trans* compartment, which exports proteins to a variety of destinations. These compartments contain different enzymes and mediate distinctive functions.

The *N*-linked carbohydrate units of glycoproteins are further modified in each of the compartments of the Golgi complex. In the cis Golgi compartment, three mannose residues are removed from the oligosaccharide chains of proteins destined for secretion or for insertion in the plasma membrane. The carbohydrate units of glycoproteins targeted to the lumen of lysosomes are further modified. In the medial Golgi compartments of some cells, two more mannose residues are removed, and two *N*- acetylglucosamine residues and a fucose residue are added. Finally, in the trans Golgi, another *N*-acetylglucosamine residue can be added, followed by galactose and sialic acid, to form a complex oligosaccharide unit. The sequence of *N*-linked oligosaccharide units of a glycoprotein is determined both by (1) the sequence and conformation of the protein undergoing glycosylation and by (2) the glycosyltransferases present in the Golgi compartment in which they are processed. Note that, despite all of this processing, *N*-glycosylated proteins have in common a pentasaccharide core (see Figure 11.19). Carbohydrate processing in the Golgi complex is called *terminal glycosylation* to distinguish it from *core glycosylation*, which takes place in the ER. Tremendous structural diversification can occur as a result of the terminal glycosylation process.

11.3.5. Mannose 6-phosphate Targets Lysosomal Enzymes to Their Destinations

A carbohydrate marker directs certain proteins from the Golgi complex to lysosomes. A clue to the identity of this marker came from analyses of *I-cell disease* (also called *mucolipidosis II*), a lysosomal storage disease. *Lysosomes* are organelles that degrade and recycle damaged cellular components or material brought into the cell by endocytosis. Patients with I-cell disease suffer severe psychomotor retardation and skeletal deformities. Their lysosomes contain large *inclusions* of undigested glycosaminoglycans (Section 11.2.4) and glycolipids (Section 12.2.3)—hence the "I" in the name of the disease. These inclusions are present because at least eight acid hydrolases required for their degradation are missing from affected lysosomes. In contrast, very high levels of the enzymes are present in the blood and urine. Thus, active enzymes are synthesized, but they are exported instead of being sequestered in lysosomes. In other words, *a whole series of enzymes is mislocated in I-cell disease.* Normally, these enzymes contain a mannose 6-phosphate residue, but, in I-cell disease, the attached mannose is unmodified (Figure 11.25). *Mannose 6-phosphate is in fact the marker that normally directs many hydrolytic enzymes from the Golgi complex to lysosomes. I-cell patients are deficient in the phosphotransferase catalyzing the first step in the addition of the phosphoryl group; the consequence is the mistargeting of eight essential enzymes.*

Figure 11.25

Formation of a Mannose 6-Phosphate Marker. A glycoprotein destined for delivery to lysosomes acquires a phosphate marker in the cis Golgi compartment in a two-step process. First, a phosphotransferase adds a phospho-*N*-acetylglucosamine unit to the 6-OH (more...)

11.3.6. Glucose Residues Are Added and Trimmed to Aid in Protein Folding

The oligosaccharide precursors added to proteins may play a role in protein folding as well as in protein targeting. As we have seen, before a glycoprotein leaves the ER, two glucosidases cleave the three glucose residues of the oligosaccharide in a step-by-step fashion. If the protein is properly folded, it moves to the Golgi complex for further processing (Section 11.3.3). However, if the protein is sufficiently unfolded that the oligosaccharide can act as a substrate for glucosyltransferase, another enzyme residing in the lumen of the ER, a glucose residue will be reattached (Figure 11.26). This residue, in turn, is bound by one of two chaperone proteins called *calnexin* and *calretic-ulin.* Calnexin, the more fully understood of the two proteins, is membrane bound, whereas calreticulin is a soluble component of the ER lumen. Unfolded proteins held by these carbohydrate-binding proteins (lectins, Section 11.4) cannot leave the ER, giving the unfolded proteins time to fold properly. When a chaperone releases the bound protein, the glucose residue will be cleaved by a glucosidase. If the folding is correct, the protein moves to the Golgi complex. Otherwise, the protein will repeat another cycle of glucose addition and binding until the glucose-free (and, hence, properly folded) protein can be translocated to the Golgi complex. This qualitycontrol system reveals an important principle: *carbohydrates carry information.* Here, the availability of carbohydrates to specific glycosyltransferases conveys information about

the folding state of the protein. Moreover, we see the reiteration of a theme in the control of protein folding: other chaperone proteins rely on the same essential mechanism of allowing misfolded proteins multiple attempts to reach a folded state (Section 3.6), even though carbohydrate modification is not a part of their reaction cycles.

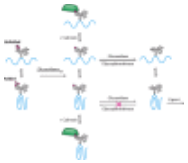


Figure 11.26

Quality-Control System for Protein Folding in the ER. A properly folded glycoprotein will move to the Golgi complex after the removal of glucose moieties (shown in red). An unfolded or misfolded protein will receive a glucose residue, through the action (more...)

11.3.7. Oligosaccharides Can Be “Sequenced”

Given the large diversity of oligosaccharide structures and the many possible points of attachment to most proteins, how is it possible to determine the structure of a glycoprotein? Most approaches are based on the use of enzymes that cleave oligosaccharides at specific types of linkages. For example, *N*-linked oligosaccharides can be released from proteins by an enzyme such as Peptide *N*-glycosidase F, which cleaves the *N*-glycosidic bonds linking the oligosaccharide to the protein. The oligosaccharides can then be isolated and analyzed. Through the use of MALDI-TOF or other mass spectrometric techniques (Section 4.1.7), the mass of an oligosaccharide fragment can be determined. However, given the large number of potential monosaccharide combinations, many possible oligosaccharide structures are consistent with a given mass. More complete information can be obtained by cleaving the oligosaccharide with enzymes of varying specificities. For example, β-1,4-galactosidase cleaves β-glycosidic bonds exclusively at galactose residues. The products can again be analyzed by mass spectrometry (Figure 11.27). The repetition of this process with the use of an array of enzymes of different specificity will eventually reveal the structure of the oligosaccharide.

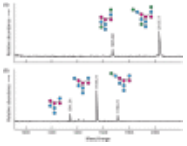


Figure 11.27

Mass Spectrometric “Sequencing” of Oligosaccharides. Carbohydrate-cleaving enzymes were used to release and specifically cleave the oligosaccharide component of the glycoprotein fetuin from bovine serum. Parts A and B show the masses obtained (more...)

The points of oligosaccharide attachment can be determined through the use of proteases. Cleavage of a protein by applying specific proteases yields a characteristic pattern of peptide fragments that can be analyzed chromatographically (Section 4.2.1). The chromatographic properties of peptides attached to oligosaccharides will change on glycosidase treatment. Mass spectrometric analysis or direct peptide sequencing can reveal the identity of the peptide in question and, with additional effort, the exact site of oligosaccharide attachment.

Posttranslational modifications such as glycosylation greatly increase the complexity of the proteome. A given protein with several potential glycosidation sites can have many different glycosylated forms (sometimes called *glycoforms*), each of which may be generated only in a specific cell type or developmental stage. Now that the sequencing of the human genome is essentially complete, the characterization of the much more complex proteome, including the

biological roles of specifically modified proteins, can begin in earnest.

Bookshelf ID: NBK22521

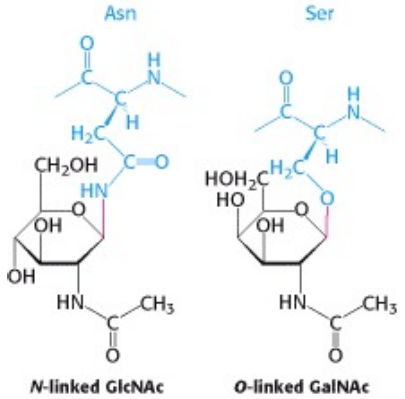


Figure 11.18

Glycosidic Bonds between Proteins and Carbohydrates. A glycosidic bond links a carbohydrate to the side chain of asparagine (*N*-linked) or to the side chain of serine or threonine (*O*-linked). The glycosidic bonds are shown in red.