Covalent carbohydrate-protein compounds are widely distributed natural products. They are found mainly in animals, vertebrates and invertebrates, but recently their occurrence in green plants has also been reported. All carbohydrate-protein compounds have the same fundamental structure. Their overall design is that of a polypeptide backbone to which are joined covalently one or more carbohydrate groups. According to the nature and the structure of the carbohydrate moiety two main classes of carbohydrate-protein compounds may be distinguished: \( \text{proteo-glycosaminoglycans} \), or shortly \( \text{proteoglycans} \) and \( \text{glycoproteins} \). In the first class the carbohydrate is characterized by a small repeating unit, frequently a disaccharide, composed of uronic acid and an N-substituted hexosamine, a high degree of polymerization and a linear structure. In glycoproteins the carbohydrate moiety consists as a rule of one or more heterooligosaccharides or heteropolysaccharides. The composition of the heterosaccharides and their number vary from glycoprotein to glycoprotein. In ovalbumin (mol. wt. 45,000) with about 3.5 % carbohydrate only one prosthetic group of mol. wt. 1420 is present. In the human blood-group specific substances A, B and H (see Fig. 1) of an average mol. wt. of 500,000, containing 85 % carbohydrate, about 185 carbohydrate chains are present. GalNAc, GlcNAc, D-galactose, L-fucose and sialic acid are components of these chains with an approximate mol. wt. of 2300. As is evident from these figures, the degree of polymerization is low. In both glycoproteins (ovalbumin and blood-group specific substances) the chains are heavily branched. There is no report of the presence of repeating units in the carbohydrate chains of glycoproteins. L-Fucose and sialic acid, if present, invariably are terminating the chain or a side-chain.

One of the most fascinating aspects of glycoprotein biochemistry is their biosynthesis. Even this selected topic is so voluminous that I will mainly discuss the general mechanism of glycoprotein biosynthesis, the specificity of the enzymes involved and their localization inside the cell. I will illustrate the principal features in some detail on a few well investigated glycoproteins rather than give a review of all relevant data.

As to the biosynthesis of the protein moiety of glycoproteins it has been shown for mouse immunoglobulins quite unambiguously that the polypeptide is synthes-
ized at membrane-bound polyribosomes of the rough endoplasmic reticulum (for references see Melchers 7). Most probably these polyribosomes are attached to the outside of the rough membrane. The protein is synthesized by the well studied mechanism in which the genetic message is used as a template for the nature, number and sequence of amino acids, all joined together by a single, most probably unspecific enzyme system. The synthesis of the polypeptide takes only a few minutes.

By contrast, the heterosaccharide chains of glycoproteins are synthesized by highly specific multiglycosyltransferase systems (Roseman 8). The synthesis of these enzymes is of course under genetic control and in this way the carbohydrate chains are also genetically determined, albeit indirectly. Their synthesis may take 90 min or more.

The carbohydrate residues are added to the polypeptide chain in a stepwise manner, one at a time, the first residue to the functional group of a side-chain of a constituent amino acid, the others to the growing carbohydrate chain. The mechanism is similar to that known to hold for the biosynthesis of oligo- or polysaccharides. The donor of the sugar residue is a sugar-nucleotide and the acceptor either a peptide with a unique sequence of amino acids and conformation or a carbohydrate chain with a unique structure. The catalyst, i.e. the glycosyltransferase, has to fit both the donor and the acceptor.

The question whether the first sugar residue is attached to the polypeptide when still at the ribosomes remains open for discussion. For immunoglobulin G, for instance, Melchers (personal communication) has shown that the growing polypeptide contains very little (about 0.1 %) or no sugar at all, due probably to the fact that there is a pool of polypeptides of different chain length. On the other hand, Molnár and Sy 9 reported experiments showing that puromycin promoted the release of glucosamine-labeled protein from rat liver polyribosomes. These authors concluded that their results demonstrated the presence of glucosamine-labeled nascent protein in the ribosomes. It was then established that rat liver polyribosome preparations can be contaminated by 14C-glucosamine-labeled protein that is not of ribosomal origin (Robinson 10). Since puromycin might effect the release of this contaminating protein, the interpretation of the results of Molnár and Sy 9 became difficult. Robinson 11 therefore re-investigated the question using packed polyribosome preparations from rats injected with 14C-glucosamine. The radioactivity found in these preparations was not a consequence of contamination, but was due to the presence of 14C-glucosamine bound to trichloracetic acid-insoluble material that could be partly released from the ribosomes on incubation with energy sources and supernatant enzyme; the release of the material was doubled when puromycin was added to the incubation medium. In control experiments with polyribosomes obtained from rat liver shortly after injection with 14C-leucine the nascent ribosomal protein was labeled. In this case glycosylation of the polypeptide must have occurred while the nascent polypeptide was still associated with the ribosome. Robinson 11 proposes that initiation of sugar attachment to the nascent protein results from the close vicinity of the transferases to the polyribosomes. If close vicinity is lacking, attachment of the first sugar will occur only after the completed polypeptide has dissociated from the ribosome.

The high specificity of the transferase initiating glycosylation of the polypeptide chain may be best demonstrated by the biosynthesis of the carbohydrate groups of ovine submaxillary glycoprotein, referred to as OSM. This compound has all characteristic properties of a glycoprotein. Its molecular structure, enzymatic degradation and biosynthesis have been studied in detail and are well understood. Since it is the simplest glycoprotein, it may be regarded as a prototype of this class. The compound has been prepared in my laboratory in a highly purified state, though it is well to remember that the usual homogeneity criteria have to be interpreted with reservation in case of mucins because of their polydispersity, high degree of molecular asymmetry and particle-particle interaction. OSM has a mol. wt. of about 8 X 104; it contains 30 % N-acetylneuraminic acid and about an equimolar amount of N-acetylgalactosamine, i.e. it contains about 50 % carbohydrate and 50 % protein. The carbohydrate moiety is composed of about 800 individual disaccharides of the structure N-acetylneuraminosyl-(α,2→6)-N-acetylgalactosamine (for references see Gottschalk, Bhargava and Murty 12). The N-acetylgalactosamine is attached in α-D-configuration to the OH-groups of serine and threonine (Fig. 2) (for references see Buddecke et. al. 13). The disaccharide can be stepwise enzymatically removed, the sialic acid by crystalline or non-crystalline neuraminidase from Vibrio cholerae, the hexosamine by highly purified α-N-acetylgalactosaminidase, both enzymes being free of any measurable peptidase activity (Gott-
With a similar preparation as substrate ROSEMAN has reconstituted the carbohydrate moiety of OSM. One or more N-acetylglactosaminyltransferases from ovine submaxillary glands transfer the hexosamine from UDP-N-acetylgalactosamine to serine and threonine residues of the polypeptide chain of OSM. The enzyme N-acetylglactosaminyltransferase (from sheep submaxillary glands) is highly selective in its acceptor requirements. Of some fifty potential acceptors tested in ROSEMAN'S laboratory only the polypeptide obtained from OSM by partial removal of the sugar residues was active. Inactive compounds included a number of proteins, glycoproteins, serine and threonine and peptides from the active polypeptide (ROSEMAN). For the corresponding enzyme prepared from bovine submaxillary glands HAGOPIAN and EYLA reported that treatment of BSM with trypsin reduced the acceptor activity by 66%, treatment with pronase resulted in complete loss of acceptor activity. A similar observation was made by SPIRO and SPIRO. They prepared an enzyme from rat kidney cortex which transferred a galactosyl residue to the OH-group of hydroxylysine in glomerular basement membrane and collagen (Fig. 3). Free hydroxylysine and hydroxylysine present in small peptides did not act as acceptors. The best acceptor was the intact collagen polypeptide free of carbohydrate.

The enzymes called sialyltransferases are a family of transferases transferring N-acetyleneuraminic acid (NANA) and N-glycolylneuraminic acid from cytidine monophosphate (CMP) sialic acid to acceptors containing terminal galactose, N-acetylgalactosamine or NANA residues. The transferases have been found in particulate preparations of various rat tissues including mammary gland, in goat colostrum, ovine submaxillary gland and were characterized mainly by their acceptor requirements. Thus the sialyltransferases of rat mammary gland and goat colostrum use β-D-galactopyranosides as acceptors, whereas the corresponding transferase of ovine submaxillary glands requires N-acetylgalactosamine residues as they are present in OSM pretreated with neuraminidase. The mammary gland enzyme transfers NANA from CMP-NANA to C3 of the galactose moiety of lactose, the colostrum enzyme joins NANA to C6 of the galactose residue. In OSM pretreated with neuraminidase the activity of the submaxillary gland sialyltransferase results in the formation of the structure N-acetylneuraminosyl (α, 2→6)N-acetylgalactosaminyl (α, 1→3) seryl peptide (see Fig. 2). The goat colostrum transferase is sensitive not only to the ultimate sugar residue of the acceptor molecule but also to the nature of the penultimate sugar and to the position of the O-glycosidic linkage between the ultimate and the penultimate sugar. N-Acetyllactosamine proved to be nearly 8 times more efficient as acceptor than was lactose. Of the three possible position isomers of β-D-galactopyranosyl-GlcNAc (1→4, 1→3, 1→6), N-acetyllactalactos-
mine (1→4) was 5- and 28-fold more active as acceptor than were the 1→3 and 1→6 isomers respectively (Roseman 18).

It would thus appear that the transferases joining a sugar residue to serine, threonine or hydroxylysine residues are highly specific for the surrounding amino acids and for the conformation of the polypeptide chain. The surrounding amino acids contain one or more marker-amino acids, guiding the transferase to the correct amino acid residue. Thus, it has been shown that the sequence

Gly-Met-Hyl-(Gal-Glc)-Gly-His-Arg

is present not only in digests of soluble guinea pig skin collagen, but also in glycopeptides prepared from carp swim bladder and human skin collagens (Morgan et al. 18). An optimal conformation of the acceptor adds probably to the selectivity of the transferases. In the serum type of glycoproteins nearly almost N-acetylgalactosamine in β-D-configuration is bound to the amide group of asparagine or by an N-glycosidic linkage to aspartic acid. As to the sequence of amino acids in the vicinity of the glycosylated L-asparagine, the structure X-Asn-Y-Ser (or Thr), where Asn represents the glycosylated asparagine or aspartic acid residue and where X and Y are amino acid residues, is frequently, but by no means regularly, found. Conformational factors may play a more important role, as is suggested by the presence of this structure in nonglycosylated proteins (Neuberger et al. 19). My laboratory is just investigating the sequence of amino acids surrounding glycosylated serine and threonine residues as in OSM.

The direct transfer of a single sugar residue from its activated form, the sugar nucleotide, to a specific acceptor, catalysed by a specific transferase, is only one type of biosynthesis of the carbohydrate group of glycoproteins. Thus it is well known that the cell wall of Staphylococcus aureus (Gram-positive) is composed of at least 25 polysaccharide chains consisting of two alternating amino sugars, N-acetylgalactosamine and the 3-O-D-lactic acid ether of N-acetylgalactosamine, known as N-acetylmuramic acid. The N-acetylmuramic acid residues are substituted on their carboxyl groups by a pentapeptide of the sequence L-alanyl-D-isoglutaminyl-L-lysyl-D-alanyl-D-alanine. The strands of polysaccharide substituted by the pentapeptide are in turn cross-linked through an interpeptide bridge (pentaglycine). The pentaglycine extends from the ε-amino group of L-lysine of one peptidoglycan chain and then reacts with the terminal peptide bond in the pentapeptide chain of a vicinal peptidoglycan chain eliminating the terminal D-alanine in a transpeptidation reaction, thus attaching the other pentaglycine end to the carboxyl group of the previously penultimate D-alanine residue.

The interest in the biosynthesis of this peptidoglycan derives from the facts that not a monosaccharide but a disaccharide is built into the growing polysaccharide chain and that this disaccharide is synthesized on a phospholipid as intermediate carrier. The phospholipid cycle in peptidoglycan synthesis is shown in Fig. 4. The structure of the lipid was ascertained by mass spectrometry. It is a C-55-isoprenoid alcohol containing 11 isoprene units CH₃·C(CH₃)·CH(CH₃) in a chain ending in an alcoholic function to which the disaccharide-pentapeptide fragment is linked by a pyrophosphate bridge (Higashi et al. 20, 21).

The discovery of the carrier principle signified a second very interesting type of biosynthesis of the carbohydrate chain of carbohydrate-peptide compounds. The carrier mechanism proved to be widely spread in microbes (for references see Osborn 23).

A similar isoprenoid alcohol was found in many animal tissues. Its structure was established by Burgos et al. 24. It has 20 isoprene residues, the one carrying the alcohol group being saturated:
Fig. 5. Structures of goat colostrum nucleotide trisaccharides. Four compounds are known, two containing N-acetylneuraminic acid (R = acetyl), and two containing N-glycolylneuraminic acid (R = glycolyl) (from Jouzdian and Rosemann 26).

Because of the length of the carbon chain the compound was called dolichol (greek dolichos, long). Over the last years Leloir's laboratory has studied its function as intermediate carrier in the transfer of sugars from donor to acceptor. It was found that liver microsomal enzymes catalyse the transfer of glucose, mannose, N-acetylglucosamine and N-acetylgalactosamine from their UDP, GDP, UDP and UDP activated forms to an endogenous acceptor. In this way oligosaccharides were formed bound to dolichol through a phosphate or pyrophosphate bridge. From this intermediate carrier the oligosaccharide is transferred to its final acceptor, not yet obtained in a purified state, but apparently a glycoprotein (Behrens et al. 24, Leloir 25). In view of the presence in goat colostrum of UDP-trisaccharides (Fig. 5), the trisaccharides having composition and structures not unfrequently met in glycoproteins (Jouzdian and Roseman 26), the possibility of a trisaccharide transport to a growing carbohydrate chain cannot be excluded, though no transferase has yet been found catalysing such a transfer.

Now back to the immunoglobulins with which we started our discussion on the biosynthesis. Their protein moiety is invariably assembled at polyribosomes situated at the rough membranes. Immunoglobulins are composed of two heavy chains (mol. wt. 55,000) and two light chains (mol. wt. 23,000). The carbohydrate chain is in general attached to an asparagine residue in the carboxy-terminal half of the heavy chain, whereas the light chains are usually free of carbohydrate. A few years ago Melchers et al. 27 reported that the light chain produced and secreted by the mouse plasma cell tumor MOPC 46 is of the x-type and does contain one carbohydrate chain bound N-glycosidically to asparagine. The monomeric form of the light chain appears in three forms containing either 2, 1 or 0 molecules of sialic acid. Recently Choi et al. 28 investigated the kinetics of secretion of this light chain. They dissected the characteristic transit time between synthesis of the protein and release of the light chain from the cell by examining the distribution of leucine-labeled light chains among subcellular fractions and the pattern of flow between the fractions. They concluded that light chains pass from rough membrane-containing fractions to smooth membrane-containing fractions before being secreted. The carbohydrate composition of light chains isolated from different subcellular fractions may be seen from Fig. 6. It is obvious that the single polysaccharide attached to the light chain is formed by the acquisition of different sugars in the rough and smooth membrane-containing fractions. It seems that transferases catalysing the transport of sugars from donors to acceptors which are eventually secreted are membrane-bound. However, the localization of these transferases may vary from tissue to tissue even within the same animal.

It is significant that rat serum albumin, which is carbohydrate-free, is synthesized and secreted by liver cells by the same pathway as that described for glycoproteins (Peters et al. 29).

The formation of a complete chain would require optimal concentrations of donors, transferases, acceptors and pH, conditions which will not often be
fulfilled. There may be also sometimes a competition between two different transferases for the same acceptor. These circumstances result in what is known as microheterogeneity of the carbohydrate chains; some of them are incomplete, others have a sugar residue not present in the bulk of the chains (Gottschalk, Roseman, Gottschalk). The microheterogeneity of the carbohydrate chains is met in nearly all glycoproteins which have been investigated closely. The enzymatic, non-ribosomal mechanism for sequence determination does not work as faultlessly as does the template mechanism.

The advancement made in our knowledge of the biosynthesis of glycoproteins in the last decade is tremendous. But as always in science, the widening of the horizon poses more problems than it answers questions.

28. Y. S. Choi, P. M. Knopp, and E. S. Lennox, Biochemistry 10, 668–679 [1971].