BERICHTE

On the Biosynthesis of Glycoproteins *

ALFRED GOTTSCHALK

Max-Planck-Institut für Virusforschung, Tübingen

(Z. Naturforsch. 28c, 94—99 [1973]; received November 11, 1972)

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[1]. 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 (GOTTSCHALK [2], MATTHEWS and LOZAITYTE [3]). According to the nature and the structure of the carbohydrate moiety two main classes of carbohydrate-protein compounds may be distinguished: -proteo-glycosaminoglycans, or shortly proteoglycans, and 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 (GOTTSCHALK [4, 5]). In glycoproteins the carbohydrate moiety consists as a rule of one or more heterooligosaccharides or heteropolysaccharides (GOTTSCHALK [2]). 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

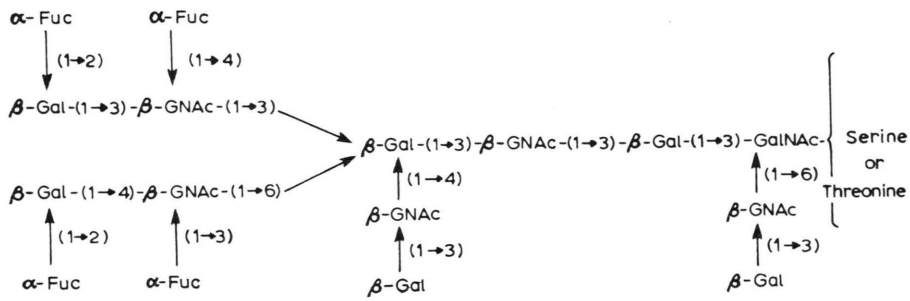

Fig. 1. Composite structure proposed for the carbohydrate chains in an HLe^b substance (from LLOYD and KABAT [6]).

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-

Requests for reprints should be sent to Prof. A. GOTTSCHALK, MPI für Virusforschung, *D-7400 Tübingen*, Spemannstr. 35.

* Plenary lecture given at the *Journées des Glucides*, organized by the *Société de Chimie Biologique et le Groupe d'étude des glucides de la Société Chimique de France* at Orsay on October 2nd and 3rd 1972.

Dieses Werk wurde im Jahr 2013 vom Verlag Zeitschrift für Naturforschung in Zusammenarbeit mit der Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V. digitalisiert und unter folgender Lizenz veröffentlicht: Creative Commons Namensnennung-Keine Bearbeitung 3.0 Deutschland Lizenz.

This work has been digitalized and published in 2013 by Verlag Zeitschrift für Naturforschung in cooperation with the Max Planck Society for the Advancement of Science under a Creative Commons Attribution-NoDerivs 3.0 Germany License.

Zum 01.01.2015 ist eine Anpassung der Lizenzbedingungen (Entfall der Creative Commons Lizenzbedingung „Keine Bearbeitung") beabsichtigt, um eine Nachnutzung auch im Rahmen zukünftiger wissenschaftlicher Nutzungsformen zu ermöglichen.

On 01.01.2015 it is planned to change the License Conditions (the removal of the Creative Commons License condition "no derivative works"). This is to allow reuse in the area of future scientific usage.

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, MOLNAR 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 ^{14}C-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 MOLNAR and SY [9] became difficult. ROBINSON [11] therefore re-investigated the question using *packed* polyribosome preparations from rats injected with ^{14}C-glucosamine. The radioactivity found in these preparations was not a consequence of contamination, but was due to the presence of ^{14}C-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 ^{14}C-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 8×10^5; 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-

SCHALK *et. al.* [14]). With a similar preparation as substrate ROSEMAN [15] has reconstituted the carbohydrate moiety of OSM. One or more *N*-acetylgalactosaminyltransferases 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*-acetylgalactosaminyltransferase (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 [15]). For the corresponding enzyme prepared from bovine submaxillary glands HAGOPIAN and EYLAR [16] 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 [17]. 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

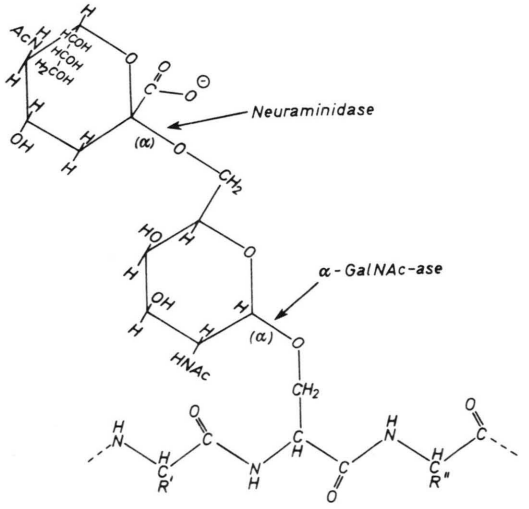

Fig. 2. *N*-Acetylneuraminosyl (α,2 → 6)-*N*-acetylgalactosaminyl (α,1 → 3)-seryl(threonyl) (see BUDDECKE *et al.*[13]).

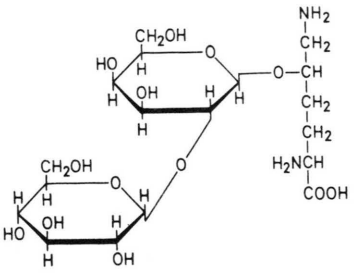

Fig. 3. Structure and peptide attachment of the disaccharide unit of basement membranes and collagens. (2-O-α-D-glucopyranosyl-O-β-D-galactopyranosylhydroxylysine) (SPIRO and SPIRO [17]).

was the intact collagen polypeptide free of carbohydrate.

The enzymes called sialyltransferases are a family of transferases transferring *N*-acetylneuraminic 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 C_3 of the galactose moiety of lactose, the colostrum enzyme joins NANA to C_6 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*-acetyllactosa-

mine (1→4) was 5- and 28- fold more active as acceptor than were the 1→3 and 1→6 isomers respectively (ROSEMAN [15]).

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*-acetylglucosamine 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 non-glycosylated 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*-acetylglucosamine and the 3-O-D-lactic acid ether of *N*-acetylglucosamine, 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-lylsyl-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 peptoglycan 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.

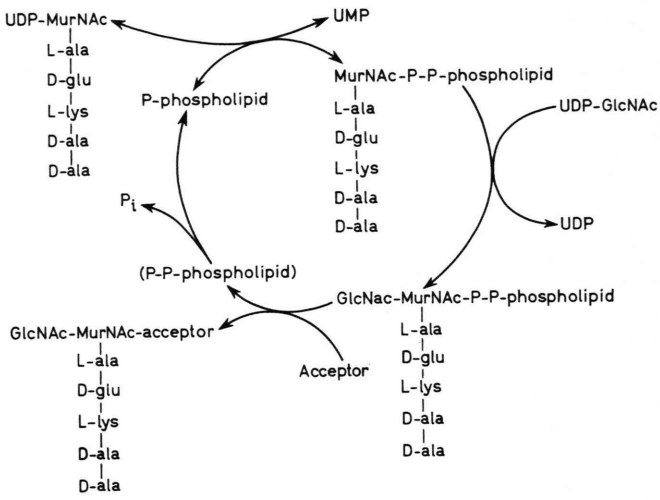

Fig. 4. Phospholipid cycle in peptidoglycan synthesis of *Staphylococcus aureus* (HIGASHI *et al.*[20]).

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 $CH_3 \cdot C(CH_3) : CH(CH_3)$ 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 [22]).

A similar isoprenoid alcohol was found in many animal tissues. Its structure was established by BURGOS *et al.*[23]. It has 20 isoprene residues, the one carrying the alcohol group being saturated:

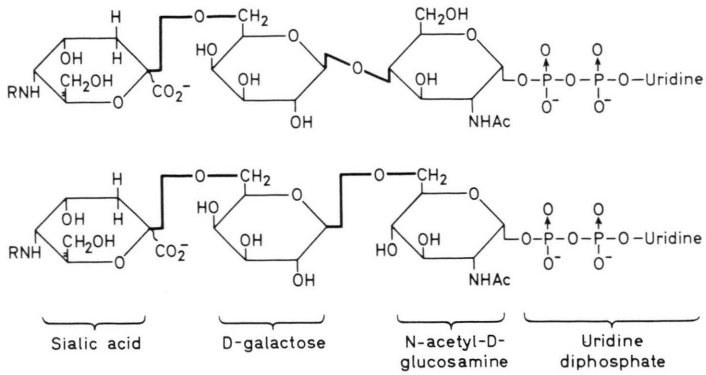

Fig. 5. Structures of goat colostrum nucleotide trisaccharides. Four compounds are known, two containing *N*-acetylneuraminic acid (R = acetyl), and two containing *N*-glycolylneuraminic acid (R = glycolyl) (from JOURDIAN and ROSEMANN [26]).

$$H \cdot (CH_2 \cdot C(CH_3) : CH \cdot CH_2)_{19} \cdot CH_2 \cdot CH(CH_3) \cdot CH_2 \cdot CH_2 \cdot OH$$

Because of the length of the carbon chain the compound was called dolichol (greek dolichos, long). Over the last years L e l o i r ' 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 (JOURDIAN 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 $\varkappa$-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

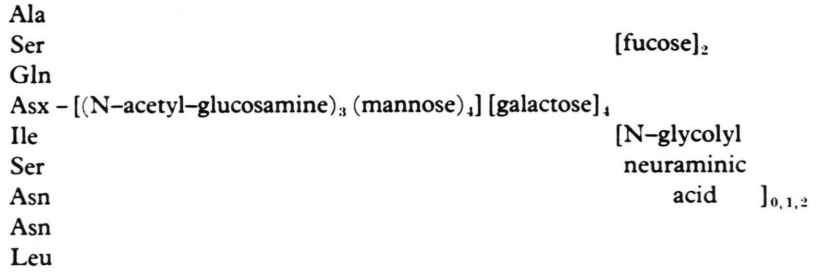

Fig. 6. Possible structure of the carbohydrate component of urinary MOPC46 κ-type light chain (from Melchers[7]).

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 [30], ROSEMAN [8], GOTTSCHALK [31]). 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.

[1] E. BUDDECKE, Glycoproteins. Their Composition, Structure and Function (A. GOTTSCHALK, ed.), vol. **A**, p. 535—564, Elsevier, Amsterdam 1972.
[2] A. GOTTSCHALK, Nature [London] **170**, 662—663 [1952]
[3] M. B. MATTHEWS and I. LOZAITYTE, Arch. Biochem. Biophysics **74**, 158—174 [1958].
[4] A. GOTTSCHALK, Perspectives in Biology and Medicine **5**, 327—337 [1962].
[5] A. GOTTSCHALK, Glycoproteins. Their Composition, Structure and Function (A. GOTTSCHALK, ed.), p. 20—28, Elsevier, Amsterdam 1966.
[6] K. O. LLOYD and E. A. KABAT, Proc. nat. Acad. Sci. USA **61**, 1470—1477 [1968].
[7] F. MELCHERS, Histochem. J. **3**, 389—397 [1971].
[8] S. ROSEMAN, Chem. and Phys. of Lipids **5**, 270—297 [1970].
[9] J. MOLNAR and D. SY, Biochemistry **6**, 1941—1947 [1967].
[10] G. B. ROBINSON, FEBS Letters **4**, 190—192 [1969a].
[11] G. B. ROBINSON, Biochem. J. **115**, 1077—1078 [1969b].
[12] A. GOTTSCHALK, A. S. BHARGAVA, and V. L. N. MURTY, Glycoproteins. Their Composition, Structure and Function (A. GOTTSCHALK, ed.) vol. **B**. p. 810—829, Elsevier, Amsterdam [1972].
[13] E. BUDDECKE, H. SCHAUER, E. WERRIES, and A. GOTTSCHALK, Biochem. biophysic. Res. Commun. **34**, 517—521 [1969].
[14] A. GOTTSCHALK, H. SCHAUER, and G. UHLENBRUCK, Hoppe-Seyler's Z. physiol. Chem. **352**, 117—124 [1971].
[15] S. ROSEMAN, Biochemistry of Glycoproteins and Related Substances(E. ROSSI and E. STOLL, eds.) p. 244—269, S. Karger, Basel 1968.
[16] A. HAGOPIAN and E. H. EYLAR, Arch. Biochem. Biophysics **128**, 422—433 [1968].
[17] R. G. SPIRO and M. J. SPIRO, Federat. Proc. **27**, 345 [1968].
[18] P. H. MORGAN, H. G. JACOBS, J. P. SEGREST, and L. W. CUNNINGHAM, J. biol. Chemistry **245**, 5042—5048 [1970].
[19] A. Neuberger, A. GOTTSCHALK, R. D. MARSHALL, and R. G. SPIRO, Glycoproteins. Their Composition, Structure and Function (A. GOTTSCHALK, ed.) 2nd edn., p. 450—485 Elsevier, Amsterdam [1972].
[20] Y. HIGASHI, J. L. STROMINGER, and C. C. SWEELEY, Proc. nat. Acad. Sci. USA **57**, 1878—1884 [1967].
[21] Y. HIGASHI, G. SIEWERT, and J. L. STROMINGER, J. biol. Chemistry **245**, 3683—3690 [1970].
[22] M. J. OSBORN, Annu. Rev. Biochem. **38**, 501—533 [1969].
[23] J. BURGOS, F. W. HEMMING, J. F. PENNOCK, and R. A. MORTON, Biochem. J. **88**, 470—482 [1963].
[24] N. H. BEHRENS, A. J. PARODI, and L. F. LELOIR, Proc. nat. Acad. Sci. USA **68**, 2857—2860 [1971].
[25] L. F. LELOIR, Journées des Glucides-Orsay, organized by the Société de Chimie Biologique et la Groupe de étude des glucides de la Société Chimique de France. Bulletin de la Société Chimique de France, Nr. **14b**, p. 3 [1972].
[26] G. W. JOURDIAN and S. ROSEMAN, Ann. New York Acad. Sci. **106**, 202—217 [1963].
[27] F. MELCHERS, E. S. LENNOX, and M. FACON, Biochem. biophysic. Res. Commun. **24**, 244—251 [1966].
[28] Y. S. CHOI, P. M. KNOPF, and E. S. LENNOX, Biochemistry **10**, 668—679 [1971].
[29] T. PETERS, B. FLEISCHER, and S. FLEISCHER, J. biol. Chemistry **246**, 240—244 [1971].
[30] A. GOTTSCHALK, Nature [London] **222**, 452—454 [1969].
[31] A. GOTTSCHALK, Chimia [Zürich] **25**, 77—80 [1971].