

CHARACTERIZATION AND STRUCTURE OF PLASMA GLYCOPROTEINS

KARL SCHMID

Department of Biochemistry, Boston University School of Medicine, Boston, Massachusetts 02118, USA

ABSTRACT

The present state of knowledge of glycoproteins is briefly reviewed. The recent discovery of hitherto unknown extremely important biological functions of this class of macromolecules has given a new impetus to the basic research of glycoproteins. In an effort to understand the relationship between structure and function of these macroglobulins, the amino acid and monosaccharide sequences of the polypeptide chains and heterosaccharide units of these conjugated proteins are under intense study. The difficulties and problems encountered are discussed. The three-dimensional structure of any glycoprotein is eagerly awaited.

Interest in the various areas of study of glycoproteins has increased greatly over the past decade†, particularly since the discovery that these conjugated proteins fulfil a large number of diverse biological roles‡. For the complete understanding of the significance of this class of proteins, the elucidation of their structures is of primary importance. To date, the three-dimensional structure of no one glycoprotein has as yet been elucidated. Further, the amino acid sequence of only the 7S γ -globulins^{25, 23, 6}, porcine ribonuclease¹⁵ and avidin⁴ has been established completely. With regards to the structure^{18, 38} and biosynthesis^{26, 28} of its carbohydrate moiety, it is α_1 -acid glycoprotein (orosomucoid), a normal human plasma protein which has a high carbohydrate content of approximately 40 per cent, that is probably the most extensively studied glycoprotein.

Glycoproteins may be defined as proteins that carry covalently bound sugars of which the prominent ones are D-galactose, D-mannose, D-glucose, L-fucose, D-

† A list of over thirty reviews on glycoproteins was compiled by Schmid²⁹. Recent important reviews were published by Schmid²⁹, Spiro^{34, 36}, Simmons *et al.*³², and Ginsberg and Neufeld⁹. The treatises by Jeanloz and Balazs¹⁶, Gottschalk¹⁰, Rossi and Stoll²⁷ contain further invaluable information on the investigations of glycoproteins.

‡ The following examples should illustrate the wide range of biological roles exhibited by glycoproteins: some of these proteins are known to be responsible for certain enzymatic, hormonal and antibody activities, for the specific binding of the vitamin B₁₂-intrinsic factor complex by the microvillous membranes of the intestinal mucosa¹⁹, the cellular recognition and cellular adhesion, antigenic sites of membranes, active transport of molecules across membranes, and the conduction of nervous impulses.

xylose, *N*-acetyl-D-glucosamine, *N*-acetyl-D-galactosamine, and several sialic acids. Uronic acids are not present in glycoproteins except in the glycosaminoglycan-protein complexes. There are several classes of glycoproteins: 1. Globular glycoproteins (plasma proteins, ovalbumin). The carbohydrate-protein linkages of these macromolecules are *N*-glycosidic and involve the β -amide nitrogen of asparagine and *N*-acetylglucosamine residues. 2. Submaxillary mucoproteins. The carbohydrate-protein linkages of this class of glycoproteins involve *O*-glycosidic bonds between *N*-acetylgalactosamine and serine or threonine residues. 3. The glycosaminoglycans (for example, dermatan). The linkages between the carbohydrate units and the polypeptide chain(s) of these compounds are provided by xylose and serine residues. Characteristic of this type of linkage is its relatively high sensitivity towards alkaline hydrolysis (β -elimination). 4. Structural proteins (for example, the collagens). The carbohydrate units of these proteins are linked *O*-glycosidically by galactose residues to the hydroxyl group of hydroxylysine residues. This linkage is unusually stable towards alkaline hydrolysis. 5. Glycoproteins of cell walls of certain plants. These have been shown to possess a carbohydrate-polypeptide linkage that involves arabinose and hydroxyproline residues²⁰. 6. Cell wall glycoproteins of certain bacteria, the carbohydrate is linked to the protein moiety through phosphodiester bonds¹⁴.

For the Symposium lecture the globular glycoproteins were selected as the topic, but the discussion was limited largely to plasma glycoproteins. Further, the biosynthesis, catabolism and the biological roles of these proteins were also excluded from this presentation.

ISOLATION AND CHARACTERIZATION

Considerable progress has been reported during the past decade on the techniques used for the isolation and characterization of proteins and glycoproteins. The methods employed for isolating these conjugated proteins have constantly been refined and are those utilized for the preparation of proteins in general. A newly introduced technique is the centrifugation in a density gradient system⁵ where advantage is taken of the difference in the average partial specific volume of the amino acid ($\bar{v} = 0.7$) and monosaccharide ($\bar{v} = 0.6$) residues. This procedure permits separation, for example, of certain mucopolysaccharide-protein complexes from each other based on differences in their carbohydrate contents. Further, the two criteria for homogeneity of protein preparations using disc electrophoresis and electrofocusing have become routine tests. These two procedures have now been adopted successfully for preparative techniques.

For the characterization of glycoproteins the following new technique has recently become common use. Gas-liquid chromatography of appropriate derivatives of the monosaccharides, obtained after methanolysis of such proteins, seems to afford the most accurate qualitative and quantitative values². Under the experimental conditions employed, only a minute destruction of the sugars takes place. Significant progress has also been made in the gas-liquid chromatographic analysis of amino acids⁸. A further, very promising method is mass spectrometry. Mass spectrometric techniques originally introduced by Lederer have recently been applied successfully to the determination of the

CHARACTERIZATION AND STRUCTURE OF PLASMA GLYCOPROTEINS

structures of certain 'difficult to elucidate' peptides¹. Efforts are being made to adopt this method for the use in carbohydrate chemistry.

STRUCTURAL STUDIES

The structural studies of glycoproteins may be divided into three parts, one pertaining to the primary, the other to the secondary, and the third to the tertiary structure. It should be noted that the conformations of glycoproteins have been investigated to a relatively small extent and the tertiary structures very little. Recent studies by Greenfield and Fasman¹¹ demonstrated that the three classes of conformations of, at least, certain proteins can now be determined accurately using circular dichroism measurements.

The primary structure of a glycoprotein involves the elucidation of the amino acid sequence, the sequence of the monosaccharide residues and the linkages between the carbohydrate units and the polypeptide chain(s). The methods for the determination of the amino acid sequences of 'simple' proteins¹³ are applicable for the largest part to glycoproteins, as demonstrated by the investigations, for example, on the 7S γ_2 -globulins²³ and porcine ribonuclease¹⁵. The cleavage by cyanogen bromide and the degradation by the Edman technique of the polypeptide chain yielded the expected results when applied to glycoproteins. Although sequenators will be used for the elucidation of the amino acid sequences during the coming years, it is hoped that later mass spectrometric procedures can be developed for this purpose.

With the availability of glycopeptides, earlier isolated primarily from pronase digests of glycoproteins, it was possible to establish the amino acid sequences adjacent to the carbohydrate-polypeptide linkages of many of these conjugated proteins^{29, 15, 4}. These studies have led to speculations as to a recognition signal for an enzyme attaching *N*-acetylglucosamine to certain asparagine residues of the completely synthesized polypeptide chain of a glycoprotein^{7, 34, 35, 15, 28}. Such apparently predetermined asparaginyl residues become substituted if they are within the following amino acid sequence: -Asn-X-($\frac{\text{Thr}}{\text{Ser}}$)^{7, 34, 35, 22, 15}. However, not enough investigations have been carried out to demonstrate whether all such tripeptides will indeed be substituted by this enzymatic reaction.^{28, 4} It should be kept in mind that such peptide sections must be located at the surface of the polypeptide molecule because the polypeptide chain has probably assumed its final three-dimensional structure before the first monosaccharide residue of the carbohydrate unit is attached. A critical analysis, permitting evaluation of the above mentioned contentions, would require the study of the surface of a protein devoid of carbohydrate (*e.g.* serum albumin) for tripeptides with the mentioned sequence. Indeed, as Hunt and Dayhoff¹² pointed out, a large number of such tripeptides that are free of sugars, is already known. In the case of albumin it might be that the cells responsible for the production of this protein do not synthesize the required sugar transferases. Moreover, the finding that ribonuclease B possesses a carbohydrate unit, although ribonucleases A and B have the same amino acid sequence, suggests that probably additional factors play a role in influencing the mechanism of the attachment of the initial *N*-acetyl-glucosamine residues to nascent proteins.

A further problem fundamental to the investigation of the primary structure

of glycoproteins involves the determination of the distances between the various carbohydrate units in terms of the exact number of amino acid residues and not, as it has been done in the past, in terms of an average number of such residues. Moreover, the study of the amino acid sequences of glycoproteins will also yield the genetically determined amino acid substitutions accounting for the so-called variants. Some of these variants, whose amino acid difference(s) lead to a difference in the electrostatic net charge, are thus demonstrable by electrophoresis. Genetically transmitted amino acid substitutions have already been reported for a large number of 'simple' proteins³.

The various types of linkages between carbohydrate and polypeptide moieties of glycoproteins have been enumerated above. Of particular interest is the observation by Spik *et al.*³³ and others³⁶ who reported that certain glycoproteins possess two distinct types of such linkages.

The size, number and chemical composition of the carbohydrate units have been established for many glycoproteins³⁶. It is now well known that the carbohydrate moiety of all glycoproteins is present as several units, an exception being ovalbumin that has only one such unit. As to the composition of the carbohydrate moiety of these macromolecules, it should be noted that so far no integral ratios of the monosaccharide residues have been obtained. The reason for this observation is severalfold, primarily the microheterogeneity of the carbohydrate units and the analysis of heterogenous preparations (see below). It should be noted that appropriate criteria to assess the degree of heterogeneity of preparations of carbohydrate units are lacking at present.

The partial sequence of certain carbohydrate units of several glycoproteins has been determined^{10, 16, 17, 34, 36}. Sialic acid and fucose have always been found to occupy the terminal positions of the polysaccharide chains of such units. The sialyl residues are glycosidically linked to galactose, the penultimate monosaccharide. The linkages of the galactosyl to the third monosaccharide residue, *N*-acetyl-glucosaminyl, and that of the *N*-acetyl-glucosaminyl to the mannosyl residues have been well established^{16, 17}. The core of the carbohydrate units consisting of mannose and *N*-acetylglucosamine has not yet been elucidated for any carbohydrate unit with absolute certainty. Different branching (microheterogeneity) of the carbohydrate chains within this core has been found to be the cause of an almost insurmountable problem regarding the purification of glycopeptides. As indicated above, another major, very closely related, problem involves the assessment of the degree of heterogeneity of the carbohydrate units of a glycopeptide preparation. Thus, it would appear that the results reported so far on the composition and structure of carbohydrate units probably represent, for the largest part, average values. Great efforts are being made to solve these important questions (see *Symposium on Glycoproteins*, Spiro³⁵). The proposed structures of certain carbohydrate units of macro-immunoglobulins³⁷, α_1 -acid glycoprotein^{16, 17, 39}, fetuin³⁶, Taka-amylase A⁴¹ and stem bromelian⁴³ reflect these problems.

The three known linkages between the sialyl and galactosyl residues^{16, 17} of certain glycoproteins, referred to as positional isomorphism³⁴, have been suggested to be the cause of the polymorphism of α_1 -acid glycoprotein^{29, 42} and other glycoproteins^{29, 31, 24}. The chemical composition of the polymorphic forms of the former glycoprotein differs very little from each other^{30, 42}. While there is only a small difference in their sialic acid contents²¹, there is a considerable

difference in their apparent electrophoretic mobilities and their isoionic points³⁰.

While two decades ago very little was known about glycoproteins, an astounding wealth of knowledge has been accumulated in the meantime on the various aspects of fundamental investigations of this class of macromolecules.

ACKNOWLEDGEMENT

The studies reported in this paper were supported by grants from the National Science Foundation (GB-4531) and United States Public Health Service, National Institute of Health (GM-10374 and 1-K3-GM-32,160).

REFERENCES

- ¹ K. L. Agarwal, G. W. Kenner and R. C. Sheppard. *J. Am. Chem. Soc.* **91**, 3096 (1969).
- ² J. R. Clamp, G. Dawson and L. Hough. *Biochim. Biophys. Acta* **148**, 342 (1967).
- ³ M. O. Dayhoff. *Atlas of Protein Sequence and Structure*. National Biomedical Research Foundation, Silver Spring, Maryland, 1969.
- ⁴ R. J. DeLange. *J. Biol. Chem.* **245**, 911 (1970).
- ⁵ J. R. Dunston. *Separ. Sci.* **4**, 267 (1969).
- ⁶ G. Edlman and coworkers. *Biochemistry* **9**, 3155 (1970).
- ⁷ E. H. Eylar. *J. Theoret. Biol.* **10**, 89 (1966).
- ⁸ C. W. Gehrke, H. Nakamoto and R. Zumwalt. *J. Chromatog.* **45**, 24 (1969).
- ⁹ V. Ginsberg and E. F. Neufeld. *Ann. Rev. Biochem.* **38**, 371 (1969).
- ¹⁰ A. Gottschalk (Ed.). *Glycoproteins*, Vol. 5. In BBA Library, Elsevier Publishing Company, 1966.
- ¹¹ N. Greenfield and G. D. Fasman. *Biochemistry* **8**, 4108 (1969).
- ¹² L. T. Hunt and M. O. Dayhoff. *Biochem. Biophys. Res. Commun.* **39**, 757 (1970).
- ¹³ C. H. W. Hirs (Ed.) in *Methods in Enzymology*, Vol. 11, Academic Press, New York, 1967.
- ¹⁴ R. C. Hughes and P. F. Thurman. *Biochem. J.* **117**, 431 (1970).
- ¹⁵ R. L. Jackson and C. H. W. Hirs. *J. Biol. Chem.* **245**, 624 (1970).
- ¹⁶ R. W. Jeanloz and E. A. Balazs (Eds.). *The Amino Sugars*, especially volumes IIA and IIB. Academic Press, New York, 1966.
- ¹⁷ R. W. Jeanloz in *Glycoproteins* Vol. 5, Ed. A. Gottschalk, In BBA Library, Elsevier Publishing Company, Amsterdam, 1966, p. 362.
- ¹⁸ R. W. Jeanloz and T. Kojima. *The Chemical Structure of the Carbohydrate Components of α_1 -Acid Glycoprotein* 158th Am. Chem. Soc. Meeting, Division of Carbohydrate Chemistry, September 8-12, 1969.
- ¹⁹ W. L. Kopp, J. S. Trier, I. L. Mackenzie and R. M. Donaldson. *J. Exp. Med.* **128**, 357 (1968).
- ²⁰ D. T. A. Lamport. *Biochemistry* **8**, 1155 (1969).
- ²¹ J. Montreuil. Private communication.
- ²² A. Neuberger and R. D. Marshall. *Symposium on Foods-Carbohydrates* Avi, Westport, Conn. (1970).
- ²³ J. O'Donnell, B. Frangione and R. R. Porter. *Biochem. J.* **116**, 261 (1970).
- ²⁴ Y. Oshiro and E. H. Eylar. *Arch. Biochem. Biophys.* **127**, 476 (1968).
- ²⁵ F. W. Putnam. *Science* **163**, 633 (1969).
- ²⁶ S. Roseman in *4th International Conference on Cystic Fibrosis of the Pancreas [Mucoviscidosis]*, Part II, Ed. E. Rossi and E. Stoll, S. Karger, Basel, 1968, p. 244.
- ²⁷ E. Rossi and E. Stoll (Eds.). *4th International Conference on Cystic Fibrosis of the Pancreas [Mucoviscidosis]*, Part II, S. Karger, Basel, 1968.
- ²⁸ H. Schachter, I. Jabbal, R. L. Hudgin and L. Pinteric. *J. Biol. Chem.* **245**, 1090 (1970).
- ²⁹ K. Schmid in *4th International Conference on Cystic Fibrosis of the Pancreas [Mucoviscidosis]*, Part II, Eds. E. Rossi and E. Stoll, S. Karger, Basel, 1968, p. 4.
- ³⁰ K. Schmid, J. P. Binette, S. Kamiyama, V. Pfister and S. Takahashi. *Biochemistry* **1**, 959 (1962).
- ³¹ K. Schmid and S. Takahashi. *Nature* **203**, 407 (1964).

KARL SCHMID

- ³² P. Simmons, R. Penny and I. Goller. *Med. J. Australia* **2**, 494 (1969).
- ³³ G. Spik, M. Monsigny and J. Montreuil. *C.R. Acad. Sci. [Paris]* **263**, 893 (1966).
- ³⁴ R. G. Spiro. *New Engl. J. Med.* **281**, 911, 1043 (1969a).
- ³⁵ R. G. Spiro. *Symposium on Glycoproteins: Structures and Metabolism* 158th Meeting of the Am. Chem. Soc., Division of Carbohydrate Chemistry, September 8-12, 1969, New York.
- ³⁶ R. G. Spiro. *Ann. Rev. Biochem.* **39**, 599 (1970).
- ³⁷ B. P. Spragg and J. R. Clamp. *Biochem. J.* **114**, 57 (1969).
- ³⁸ D. B. Thomas and R. J. Winzler. *J. Biol. Chem.* **244**, 5943 (1969).
- ³⁹ P. V. Wagh, I. Bornstein and R. J. Winzler. *J. Biol. Chem.* **244**, 663 (1969).
- ⁴⁰ R. J. Winzler. *Structural Studies on a Glycoprotein from Human Erythrocyte Membranes*. Division of Carbohydrate Chemistry, 158th Meeting of the Am. Chem. Soc., September 8-12, 1969, New York.
- ⁴¹ H. Yamaguchi, T. Ikenaka and Y. Matsushima. *J. Biochem. [Tokyo]* **65**, 793 (1969).
- ⁴² T. Yamauchi and I. Yamashina. *J. Biochem. [Tokyo]* **66**, 213 (1969).
- ⁴³ Y. Yasuda, N. Takahashi and T. Murichi. *Biochemistry* **9**, 25 (1970).