transferred to the growing glycoprotein or proteoglycan on its way through the channels of the endoplasmic reticulum. [Part of the hexosamine appears to be added to the protein while still attached to ribosomes (Lawford & Schachter, 1968; Stoolmiller & Dorfman, 1969).] The glycosyltransferases forming 'multienzyme systems' (Hagopian *et al.*, 1968; Roseman, 1970) are closely associated with the membranous particles, consisting of the rough and especially of the smooth portions of the endoplasmic reticulum and of the Golgi apparatus (Horwitz & Dorfman, 1968; Lawford & Schachter, 1968; Stoolmiller & Dorfman, 1969; Spiro, 1970). In contrast with a free solution, such a specific spatial arrangement of enzymes enables a co-ordinated series of metabolic reactions and a local concentration of the substrates sufficient for optimal enzyme activity. In the Golgi region the glycoprotein is concentrated and packaged for secretion. Only a few glycosyltransferases have been subisolated and purified.

Bovine submaxillary-gland glycoproteins have been found to be synthesized in subcellular membranes by the following sequence of reactions (Schauer & Wember, 1971; fundamentally the same was observed with the corresponding porcine glycoproteins). At first *N*-acetylgalactosamine residues are linked to serine or threonine residues of the protein core. *N*-Acetyl-, *N*-glycolyl- and *N*-acetyl-*O*-acetylneuraminic acid are then transferred from their CMP-glycosides by acylneuraminic acid-specific (?) transferases. However, as could be shown by experiments with isolated subcellular particles sedimenting at 10000 g, hydroxylation or *O*-acetylation of *N*-acetylneuraminic acid can also occur after incorporation into the growing glycoprotein molecules on the microsomal membranes. The transferases and the enzymes modifying free and membrane-bound *N*-acetylneuraminic acid are believed to form multienzyme complexes in the membranes of the endoplasmic reticulum of submaxillary glands.

- Davidson, E. A. (1966) *Amino Sugars* 2B, 1
 Gottschalk, A. (1969) *Nature (London)* 222, 452
 Hagopian, A., Bosmann, H. B. & Eylar, E. H. (1968) *Arch. Biochem. Biophys.* 128, 387
 Hassid, W. Z. (1970) in *The Carbohydrates—Chemistry and Biochemistry* (Pigman, W. & Horton, D., eds.), p. 301, Academic Press, New York and London
 Heath, E. C. (1971) *Annu. Rev. Biochem.* 40, 29
 Horwitz, A. L. & Dorfman, A. (1968) *J. Cell Biol.* 38, 258
 Kornfeld, S., Kornfeld, R., Neufeld, E. F. & O'Brien, P. J. (1964) *Proc. Nat. Acad. Sci. U.S.A.* 52, 371
 Lawford, G. R. & Schachter, H. (1968) *J. Biol. Chem.* 241, 5408
 Neufeld, E. F. & Hall, C. W. (1965) *Biochem. Biophys. Res. Commun.* 19, 456
 Rodén, L. (1970) in *Chemistry and Molecular Biology of the Intercellular Matrix* (Balasz, E. A., ed.), vol. 2, p. 797, Academic Press, New York and London

- Roseman, S. (1968) in *Biochemistry of Glycoproteins and Related Substances: Cystic Fibrosis* (Rossi, E. & Stoll, E., eds.), part 2, p. 244, S. Karger, Basel and New York
 Roseman, S. (1970) *Chem. Phys. Lipids* 5, 270
 Schauer, R. (1970a) *Hoppe-Seyler's Z. Physiol. Chem.* 351, 595
 Schauer, R. (1970b) *Hoppe-Seyler's Z. Physiol. Chem.* 351, 749
 Schauer, R. (1970c) *Hoppe-Seyler's Z. Physiol. Chem.* 351, 783
 Schauer, R. & Wember, M. (1971) *Hoppe-Seyler's Z. Physiol. Chem.* 352, 1282
 Schoop, H. J., Schauer, R. & Faillard, H. (1969) *Hoppe-Seyler's Z. Physiol. Chem.* 350, 155
 Spiro, R. G. (1970) *Annu. Rev. Biochem.* 39, 599
 Stoolmiller, A. C. & Dorfman, A. (1969) *Compr. Biochem.* 17, 241
 Warren, L. (1966) *Glycoproteins* (Gottschalk, A., ed.) p. 570, Elsevier Publishing Co., Amsterdam

Structure, Genetics and Biosynthesis of Blood-Group-Specific Glycoproteins

By WINIFRED M. WATKINS (*Lister Institute of Preventive Medicine, London SW1W 8RH, U.K.*)

Five blood-group specificities, namely A, B, H, Le^a and Le^b, are detectable in human secretions. The structures responsible for these specificities arise from the action of four gene systems that are inherited independently but are closely related in their phenotypic expression, namely *ABO*, *Hh*, *Lele* and *Sese* (see Race & Sanger, 1968).

Purified blood-group substances isolated from secretions are glycoproteins composed of about 85% carbohydrate and 15% amino acids. The substances, irrespective of their blood-group-specific properties, each contain the sugars L-fucose, D-galactose, *N*-acetylglucosamine and *N*-acetylgalactosamine and 15 amino acids; the composition of the peptide moiety is unusual in that serine, threonine, proline and alanine together account for about two-thirds by weight of the total amino acids present (see Watkins, 1966; Morgan, 1970). The average molecular weights of the glycoproteins from different individuals range from 3×10^5 to 1×10^6 . The detailed macromolecular organization of these glycoproteins is not completely established, but their properties are consistent with the interpretation that a large number of oligosaccharide chains are covalently attached at intervals to a peptide backbone. The carbohydrate-peptide linkage is through α -*O*-glycosidic bonds involving *N*-acetylgalactosamine and the hydroxyl groups of serine and threonine (Anderson *et al.*, 1964; Adams, 1965; Kabat *et al.*, 1965; Donald *et al.*, 1969). About 1% of individuals do not secrete A, B, H, Le^a or Le^b substances, but they have in their secretions a glycoprotein very similar in general composition and properties, apart from a much lower fucose content, to the blood-group-active

glycoproteins (Watkins & Morgan, 1959; Vicari & Kabat, 1969).

The assignment of a defined structure to each of the A, B, H, Le^a and Le^b determinants was finally achieved by the isolation of serologically active oligosaccharides from the acid hydrolysis or alkaline degradation products of the group-specific glycoproteins. The A determinant is GalNAc α 1 $\rightarrow$ 3-[Fuc α 1 $\rightarrow$ 2]Gal β 1 $\rightarrow$ 3 or 4GlcNAc (Painter *et al.*, 1965; Lloyd *et al.*, 1966), the B determinant Gal α 1 $\rightarrow$ 3-[Fuc α 1 $\rightarrow$ 2]Gal β 1 $\rightarrow$ 3 or 4GlcNAc (Painter *et al.*, 1965; Lloyd *et al.*, 1966) and the H determinant Fuc α 1 $\rightarrow$ 2Gal1 $\rightarrow$ 3 or 4GlcNAc (Rege *et al.*, 1964a). The isolation of two types of active structures from A, B and H glycoproteins differing only in the nature of the positional linkage of the subterminal galactose to N-acetylglucosamine, 1 $\rightarrow$ 3 in one (Type 1 chain) and 1 $\rightarrow$ 4 in the other (Type 2 chain), demonstrates that two types of carbohydrate chain endings form the basis of the blood-group-active structures. The Le^a determinant is Gal β 1 $\rightarrow$ 3[Fuc α 1 $\rightarrow$ 4]GlcNAc (Rege *et al.*, 1964b) and the Le^b determinant is Fuc α 1 $\rightarrow$ 2Gal β 1 $\rightarrow$ 3[Fuc α 1 $\rightarrow$ 4]GlcNAc (Marr *et al.*, 1967). Many other oligosaccharides have been isolated from the degradation products of the blood-group-active glycoproteins, and these are identical irrespective of the blood group of the starting material, indicating the general structural similarity of the carbohydrate moieties (see Watkins, 1966; Morgan, 1970). The isolation of a pentasaccharide in which the disaccharides Gal β 1 $\rightarrow$ 3GlcNAc (Type 1) and Gal β 1 $\rightarrow$ 4GlcNAc (Type 2) occurred as branches on the same galactose residue led Lloyd *et al.* (1968) to propose a structure incorporating all the known determinants on one carbohydrate chain.

Genetic schemes for the biosynthesis of the blood-group-active glycoproteins (Ceppellini, 1959; Watkins & Morgan, 1959) proposed that the blood-group genes modify a precursor glycoprotein that occurs in all individuals. This precursor is thought to be the 'inactive' glycoprotein found in the secretions of those who do not secrete A-, B-, H-, Le^a- or Le^b-active substances. It was suggested that the primary products of the A and B alleles in the ABO system, the H gene in the Hh system and the Le gene in the Lele gene system are enzymes catalysing the transfer of sugars from an activated donor substrate to carbohydrate chains in the precursor. The alleles O, h and le are inactive. The secretor gene Se controls the expression of the H gene in cells synthesizing the secreted glycoproteins. The protein specified by the H gene is an α -L-fucosyltransferase that conveys fucose to the terminal galactosyl residues in either Type 1 or Type 2 chains to give H-active structures. The enzymes specified by the allelic A and B genes are an α -N-acetylgalactosaminyltransferase and an α -galactosyltransferase respectively that add N-acetylgalactosamine and D-galactose to the C-3 position of

the terminal β -galactosyl residue in the H determinant. In the absence of H-active structures, A and B determinants cannot be completed. The Le gene product is an α -L-fucosyltransferase that catalyses the addition of L-fucose to the C-4 position of the subterminal N-acetylglucosamine residue in the Type 1 chain endings to give Le^a-active structures. The Le^b determinant is an interaction product formed when the H- and Le-gene-specified α -L-fucosyltransferases convey fucose to both the galactose and N-acetylglucosamine residues in the Type 1 structures.

Experimental proof of these pathways has been obtained by examination of tissues from donors of known blood group and secretor status for glycosyltransferases of the required specificity. An α -N-acetylgalactosaminyltransferase found only in tissues from group A and AB subjects occurs in human submaxillary glands and stomach linings (Hearn *et al.*, 1968; Tuppy & Schenkel-Brunner, 1969), milk (Kobata *et al.*, 1968a) and plasma (Sawicka, 1971). An α -galactosyltransferase occurring only in group B and AB subjects is demonstrable in the same tissues (Race *et al.*, 1968; Kobata *et al.*, 1968b; Sawicka, 1971). These enzymes have a requirement for H-active structures in either low-molecular-weight or macromolecular acceptors. The enzymes are found in tissues of both secretors and non-secretors of the appropriate blood groups, thus supporting the concept that failure to secrete A and B substances arises from the absence of the requisite H structure and not from lack of expression of the A and B genes.

Direct evidence that the secretor status is dependent on the expression of the H gene has also been obtained. α -2-L-Fucosyltransferase activity is detectable in milk (Shen *et al.*, 1968) and submaxillary glands (Chester & Watkins, 1969) from secretors but is not found in these same tissues from non-secretors. In contrast α -4-L-fucosyltransferase activity occurs in tissues from both secretors and non-secretors; with Gal β 1 $\rightarrow$ 3GlcNAc the enzyme gives a trisaccharide that co-chromatographs with the Le^a-active trisaccharide (Chester & Watkins, 1969), and with a glycoprotein devoid of blood-group activity yields a product with Le^a activity (Jarkovsky *et al.*, 1970).

Adams, J. B. (1965) *Biochem. J.* **97**, 345

Anderson, B., Seno, N., Sampson, P. Riley, R. G., Hoffman, P. & Meyer, K. (1964) *J. Biol. Chem.* **239**, 2716

Ceppellini, R. (1959) *Biochem. Hum. Genet., Ciba Found. Symp.* **1959**, 342

Chester, M. A. & Watkins, W. M. (1969) *Biochem. Biophys. Res. Commun.* **34**, 835

Donald, A. S. R., Creeth, J. M., Morgan, W. T. J. & Watkins, W. M. (1969) *Biochem. J.* **115**, 125

Hearn, V. M., Smith, Z. G. & Watkins, W. M. (1968) *Biochem. J.* **109**, 315

Jarkovsky, Z., Marcus, D. M. & Grollman, A. P. (1970) *Biochemistry* **9**, 1123

- Kabat, E. A., Basset, E. W., Pryzwansky, K., Lloyd, K. O., Kaplan, M. E. & Layug, E. J. (1965) *Biochemistry* **4**, 1632
- Kobata, A., Grollman, E. F. & Ginsburg, V. (1968a) *Arch. Biochem. Biophys.* **124**, 609
- Kobata, A., Grollman, E. F. & Ginsburg, V. (1968b) *Biochem. Biophys. Res. Commun.* **32**, 272
- Lloyd, K. O., Kabat, E. A., Layug, E. J. & Gruezo, F. (1966) *Biochemistry* **5**, 1489
- Lloyd, K. O., Kabat, E. A. & Licerio, E. (1968) *Biochemistry* **7**, 2976
- Marr, A. M. S., Donald, A. S. R., Watkins, W. M. & Morgan, W. T. J. (1967) *Nature (London)* **215**, 1345
- Morgan, W. T. J. (1970) in *British Biochemistry Past and Present* (Goodwin, T. W., ed.), p. 99, Academic Press, London
- Painter, T. J., Watkins, W. M. & Morgan, W. T. J. (1965) *Nature (London)* **206**, 594
- Race, C., Zideman, D. & Watkins, W. M. (1968) *Biochem. J.* **107**, 733
- Race, R. R. & Sanger, R. (1968) *Blood Groups in Man*, 5th. edn., Blackwells, Oxford
- Rege, V. P., Painter, T. J., Watkins, W. M. & Morgan, W. T. J. (1964a) *Nature (London)* **203**, 360
- Rege, V. P., Painter, T. J., Watkins, W. M. & Morgan, W. T. J. (1964b) *Nature (London)* **204**, 740
- Sawicka, T. (1971) *FEBS Lett.* **16**, 364
- Shen, L., Grollman, E. F. & Ginsburg, V. (1968) *Proc. Nat. Acad. Sci. U.S.A.* **59**, 224
- Tuppy, H. & Schenkel-Brunner, H. (1969) *Eur. J. Biochem.* **10**, 152
- Vicari, G. & Kabat, E. A. (1969) *J. Immunol.* **102**, 821
- Watkins, W. M. (1966) *Glycoproteins* 462
- Watkins, W. M. & Morgan, W. T. J. (1959) *Vox Sang.* **4**, 97

Structural Analysis of Glycoproteins

By J. R. CLAMP (*Department of Medicine, University of Bristol, Bristol BS8 1TD, U.K.*)

Although there has been little fundamental advance in recent years in methods for investigating the structure of the oligosaccharide units of glycoproteins, there has been some improvement in analytical techniques. Automated column procedures and g.l.c. have both been used for this purpose, and, as our experience has been with the latter technique, this will be used to illustrate the application of structural methods.

Two main procedures are in use for the analysis of the carbohydrate content of glycoproteins by g.l.c., differing in the method of cleavage of the polymeric material. The first of these is based on hydrolysis with aqueous acid followed by reduction of the liberated aldoses to the glycitols. The second depends on methanolysis, yielding methyl glycosides together with the methyl esters of carboxyl groups as in hexuronic acids and sialic acids. The final products in each procedure, namely glycitols or methyl glycosides, are then chromatographed as the *O*-trimethyl-

silyl ethers. Both systems have certain advantages, although our experience favours methanolysis. Thus it appears to be more efficient than hydrolysis in cleaving glycosidic linkages and the products are probably more stable (Chambers & Clamp, 1971). In addition, each monosaccharide unit after methanolysis gives a series of methyl glycosides, thereby producing a characteristic pattern of peaks on g.l.c., and this makes the identification of each monosaccharide more certain. This multiple-peak pattern has an additional advantage when, as is common with biological samples, there is some contamination and a background peak coincides with one of the isomer peaks. In this case the amount of monosaccharide can still be calculated from the unobserved peaks.

By using this technique, all the monosaccharide units commonly occurring in glycoproteins can be separated, identified and determined in a single procedure. Such monosaccharide units include fucose, mannose, galactose, glucose, *N*-acetylglucosamine, *N*-acetylgalactosamine and sialic acid. In addition, the system can cope with pentoses, hexuronic acids, glycitols and most other monosaccharides and derivatives, and even with those amino acids likely to occur in glycopeptides (Bhatti & Clamp, 1971).

G.l.c. can also be used to considerable advantage as an analytical procedure in conjunction with classical methods such as methylation, periodate oxidation, partial acid hydrolysis and so on. These methods are usually carried out on glycopeptides isolated from the glycoproteins. Methylation studies give information on positions of substitution of monosaccharides. A number of systems have been devised to deal with the *O*-methyl derivatives of those carbohydrate residues usually found in glycoproteins, in which the partially methylated residues are usually separated as mixed derivatives (Clamp *et al.*, 1971).

The technique of periodate oxidation also provides information on positions of substitution, although the information is less precise than that given by methylation. Despite this, the technique is used more often, probably because it is simpler and uses less material, and because the extent and completeness of the reaction can be assessed by determining the amount of unchanged periodate. In this way the carbohydrate content of the sample can be determined after periodate consumption has levelled off, with particular interest in hexose residues resistant to, and *N*-acetylhexosamine residues susceptible to, oxidation. Three unsubstituted vicinal hydroxyl groups yield formic acid, and in mammalian glycoproteins this would be likely to arise from terminal or 6-*O*-substituted hexoses or from *N*-acetylneuraminic acid. Periodate oxidation followed by reduction yields a number of short-chain aldehydes and alcohols. For example, threitol arises from 4-*O*-substituted galactose, whereas the corresponding mannose deri-