Chem. Soc. Rev., 1987, **16**, 161—185

TATE AND LYLE LECTURE
Structural and Conformational Characterization of Carbohydrate Differentiation Antigens

By Elizabeth F. Hounsell
APPLIED IMMUNOCHEMISTRY RESEARCH GROUP, MRC CLINICAL RESEARCH CENTRE, WATFORD ROAD, HARROW, MIDDLESEX HA1 3UJ

1 Introduction

Recently there has been a resurgence of interest in the chemistry and biochemistry of carbohydrates which have been largely overshadowed in the past few years by the significant advances in nucleic acid technologies and then protein sequencing and synthetic methods. It is now becoming more widely realized that post-translational events, in particular protein glycosylation, also need to be well-characterized at the molecular level. In the carbohydrate field two main areas of research have contributed to this realization. First, improvements in methods of carbohydrate analysis have allowed the characterization of a large number of different structures and studies on their biosynthesis, tissue localization, and cellular distribution. Second, it has been shown that a number of naturally occurring and hybridoma-derived antibodies recognize carbohydrate structures as antigens which change during various stages of cell development and tumourigenesis.[1-3] The term differentiation antigen was coined[4] to describe cell surface antigens that are restricted to particular lineages or to specific stages within a cell lineage. Antibodies to these surface markers are extremely useful reagents in analysis of different cell types[5,6] and characterization of the determinants which they recognize can provide significant insight into cell surface biology.

The present review is a selective account of the characterization of antibodies which recognize carbohydrate differentiation antigens, highlighting the use of modern chromatographic, n.m.r., and f.a.b.-m.s. methods* in the structural and antigenic assignment of oligosaccharides with particular reference to the characterization of fucosylated and sulphated antigens based on the *N*-acetyllactosamine sequence, Galβ(1→4)GlcNAc.

[1] T. Feizi, *Immunol. Commun.*, 1981, **10**, 127.
[2] T. Feizi, *Nature*, 1985, **314**, 53.
[3] S. Hakomori, *Sci. Am.*, 1986, **254**, 32.
[4] E. A. Boyse and L. J. Old, *Ann. Rev. Genet.*, 1969, **3**, 269.
[5] 'Monoclonal Antibodies', ed. P. C. L. Beverley, Churcill Livingstone, 1986.
[6] 'Immunocytochemistry', ed. J. M. Polak and S. Van Noorden, 2nd Edn., John Wright, Bristol, 1986.

* Abbreviations used: m.s., mass spectrometry; f.a.b.-m.s., fast atom bombardment m.s.; e.i.-m.s., electron impact m.s.; ^{1}H n.m.r., proton nuclear magnetic resonance spectroscopy; h.p.t.l.c., high performance thin layer chromatography; h.p.l.c., high performance liquid chromatography; g.c., gas–liquid chromatography; Gal, D-galactopyranose; GlcNAc, 2-*N*-acetamido-D-glucopyranose; Fuc, L-fucopyranose; Man, D-mannopyranose; GalNAc, 2-*N*-acetamido-D-galactopyranose; Glc, D-glucopyranose; Asn, asparagine; Ser, serine; Thr, threonine; Cer, ceramide; 3-fucosyl-*N*-acetyllactosamine, (*continued over page*)

A. The Distribution of Carbohydrate Differentiation Antigens.—Carbohydrate antigens at the cell surface occur as oligosaccharide chains covalently linked to protein in glycoproteins and proteoglycans and to lipid in glycolipids. Proteoglycans[7] have long sulphated oligosaccharide chains which make up the bulk of the matrix surrounding cells. Glycoproteins[8] and glycolipids[9] can be found either anchored in the cell membrane by relatively hydrophobic, non-glycosylated regions or as extracellular components. The carbohydrate moieties of these glycoconjugates surround the cell surface where they are accessible to antibodies[1,10] and other carbohydrate-binding proteins (lectins[11]).

As described below for the antigen termed SSEA-1, the same carbohydrate antigen can exist on different types of glycoconjugate chain and also as part of free oligosaccharides such as those of milk and urine. In addition, carbohydrate antigens can be present in many different tissues and can, for example, show tumour-associated changes where an antigen absent from one type of tissue is expressed on tumour cells derived from it.

The antigen SSEA-1 recognized by a monoclonal antibody raised against mouse teratocarcinoma cells[12] is a stage-specific embryonic antigen (acronym SSEA) in mouse embryo development (appearing at the 8-cell stage of the embryo). However, in man this antigen is relatively widespread, the normal human distribution including the spleen, brain, and kidney.[13] The antigen is absent from other tissues, for example liver, pancreas, skeletal muscle, and non-lactating breast, but in this last organ the antigen appears during tumour development.[13]

As shown in Table 1, the carbohydrate structure characterized as the SSEA-1 antigen,[14,15] 3-fucosyl-*N*-acetyllactosamine, can form part of carbohydrate chains

---

3FLN, Galβ(1→4)GlcNAc
↑ 1,3
Fucα

LNFP-II, Galβ(1→3)GlcNAcβ(1→3)Galβ(1→4)Glc
↑ 1,4
Fucα

LNFP-III, Galβ(1→4)GlcNAcβ(1→3)Galβ(1→4)Glc
↑ 1,3
Fucα

FLNH, Galβ(1→4)GlcNAcβ 1 ↘ 6
↑ 1,3 Galβ(1→4)Glc
Fucα 3 ↗
Galβ(1→3)GlcNAcβ 1

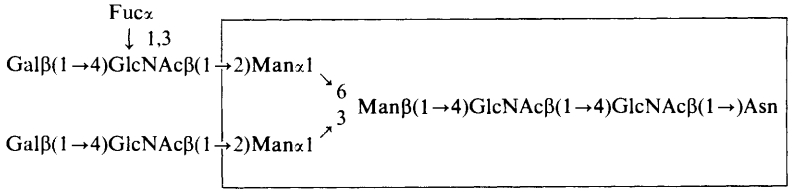
[7] M. Höök, L. Kjellán, S. Johansson, and J. Robinson, *Ann. Rev. Biochem.*, 1984, **53**, 847.
[8] R. Kornfeld and S. Kornfeld, *Ann. Rev. Biochem.*, 1985, **54**, 631.
[9] S. Hakomori, *Ann. Rev. Biochem.*, 1981, **50**, 733.
[10] E. F. Hounsell, H. C. Gooi, and T. Feizi, in 'Investigation and Exploitation of Antibody Combining Sites', ed. E. Reid, G. M. W. Cook, and D. J. Moore, Plenum Publishing Corporation, 1985, p. 317.
[11] H. Lis and N. Sharon, *Ann. Rev. Biochem.*, 1986, **55**, 35.
[12] D. Solter and B. B. Knowles, *Proc. Nat. Acad. Sci.*, 1978, **75**, 5565.
[13] N. Fox, I. Damjanov, B. B. Knowles, and D. Solter, *Cancer Res.*, 1983, **43**, 669.
[14] H. C. Gooi, T. Feizi, A. Kapadia, B. B. Knowles, D. Solter, and M. L. Evans, *Nature*, 1981, **292**, 156.
[15] E. F. Hounsell, H. C. Gooi, and T. Feizi, *FEBS Lett.*, 1981, **131**, 279.

Table 1 *Typical carbohydrate chains of glycoproteins, glycolipids, and secreted oligosaccharides bearing the SSEA-1 antigenic structure, 3-fucosyl-N-acetyllactosamine, on backbone sequences attached to different core regions (outlined in boxes)*

Glycoproteins[a]

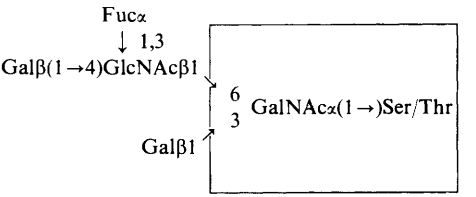
N-*linked chains*

e.g.



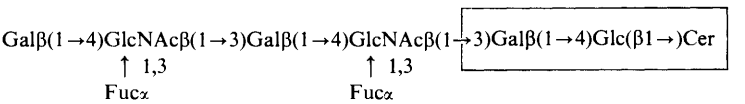
O-*linked chains*

e.g.



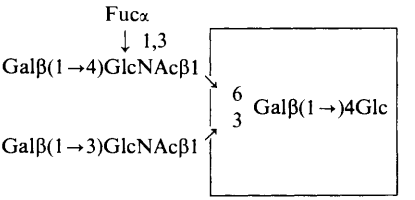
Glycolipids[b]

e.g.



Milk oligosaccharides[c]

e.g.



[a] For references see text (Section 1A). [b] Glycolipids having both of the Fuc residues shown accumulate in human adenocarcinoma (ref. 23). Those with only the left-hand Fuc are present on human erythrocytes (ref. 22) and the fucosylated pentasaccharide-Cer containing the right-hand side fucosylated pentasaccharide is found, for example, in dog small intenstine (ref. 24). [c] The oligosaccharide shown is fucosyllacto-*N*-hexaose, FLNH.[25,53]

either *N*-linked to glycoproteins[16] (*e.g.* in human secretory component[17] and promyelocytic HL60 leukaemia cells[18]), *O*-linked to glycoproteins[19] (*e.g.* in bronchial mucus[20] and human seminal plasma[21]), in glycolipids,[22–24] or as the oligosaccharides of human milk.[15,25]

B. Characterization of Carbohydrate Antigens.—High molecular weight secreted glycoproteins (mucins), which characteristically have *O*-linked chains,[26–28] and the oligosaccharides of milk[15,25,29] have been of particular value in the characterization of carbohydrate differentiation antigens because they are a relatively abundant source of oligosaccharides which express many of the antigens of cell membranes detected by monoclonal antibodies. They were used, for example, in the characterization of the Ii antigens of sheep gastric mucins[30,31] and the SSEA-1 antigen.[14,15] The strategy employed has been to inhibit with structurally defined glycoproteins and oligosaccharides the binding of antibodies to glycoproteins in inhibition radioimmunoassays.[14,30] For structural identification and correct antigenic assignment of oligosaccharides it is first necessary to carry out extensive purification (Section 2) and then to determine their complete structure.

C. Structural Analysis of Oligosaccharides.—The structural analysis of oligosaccharides requires determination of the monosaccharide composition, the sequence of monosaccharides, their position of linkage and anomeric configuration. As outlined in Table 2 this is usually achieved by the combined use of g.c., m.s., and n.m.r. Modern high field n.m.r. (*e.g.* 500 MHz ^{1}H n.m.r.) can achieve a complete structural analysis without recourse to other methods (Section 3), but *de novo* determination of previously undocumented oligosaccharides is expensive on instrument and operator time and requires relatively large amounts of material

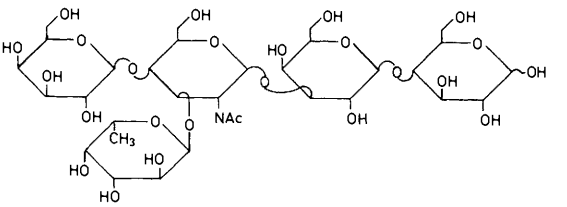
[16] M. D. Snider in 'Biology of Carbohydrates' vol. 2, ed. V. Ginsburg and P. W. Robbins, John Wiley and Sons, 1984, p. 163.
[17] A. Mizoguchi, T. Mizuochi, and A. Kobata, *J. Biol. Chem.*, 1982, **257**, 9612.
[18] A. Mizoguchi, S. Takasaki, S. Maeda, and A. Kobata, *J. Biol. Chem.*, 1984, **259**, 11949.
[19] J. E. Sadler in 'Biology of Carbohydrates' vol. 2, ed. V. Ginsburg and P. W. Robbins, John Wiley and Sons, 1984, p. 199.
[20] H. Van Halbeek, L. Dorland, J. F. G. Vliegenthart, W. E. Hull, G. Lamblin, M. Lhermitte, A. Boersma, and P. Roussel, *Eur. J. Biochem.*, 1982, **127**, 7.
[21] F.-G. Hanisch, H. Egge, J. Peter-Katalinic, G. Uhlenbruck, C. Dienst, and R. Fangmann, *Eur. J. Biochem.*, 1985, **152**, 343.
[22] R. Kannagi, E. Nudelman, S. B. Levery, and S. Hakomori, *J. Biol. Chem.*, 1982, **257**, 14865.
[23] S. Hakomori, E. Nudelman, R. Kannagi, and S. B. Levery, *Biochem. Biophys. Res. Commun.*, 1982, **109**, 36.
[24] K.-E. Falk, K.-A. Karlsson, and B. E. Samuelson, *Arch. Biochem. Biophys.*, 1979, **192**, 191.
[25] Y. Tachibana, K. Yamashita, and A. Kobata, *Arch. Biochem. Biophys.*, 1978, **188**, 83.
[26] E. A. Kabat in 'Carbohydrates in Solution', Advances in Chemistry Series No. 117, American Chemical Society, 1973, p. 334.
[27] E. F. Hounsell and T. Feizi, *Med. Biol.*, 1982, **60**, 227.
[28] E. F. Hounsell, A. M. Lawson, J. Feeney, H. C. Gooi, N. J. Pickering, M. S. Stoll, S. C. Lui, and T. Feizi, *Eur. J. Biochem.*, 1985, **148**, 367.
[29] L. C. Huang, C. I. Givin, J. L. Magnani, J. H. Shaper, and V. Ginsburg, *Blood*, 1983, **61**, 1020.
[30] E. Wood, E. F. Hounsell, J. Langhorne, and T. Feizi, *Biochem. J.*, 1980, **817**, 711.
[31] E. F. Hounsell, E. Wood, T. Feizi, M. Fukuda, M. E. Powell, and S. Hakomori, *Carbohyd. Res.*, 1981, **90**, 283.

Table 2 *Example of the structural analysis of oligosaccharides using g.c., m.s., n.m.r., and chemical or enzymatic methods for the oligosaccharide lacto-*N*-fucopentaose* III [*LNFP*-III) *of human milk*

Question asked	*Method*	*Results*
Composition	G.c. analysis of trimethylsilyl ethers of methyl glycosides of the constituent monosaccharides[a]	Molar ratio of 2:1:1:1 of Gal, Fuc, GlcNAc, Glc
Molecular weight	F.a.b.–m.s.[b] of native oligosaccharides	Mol. wt. 853 (elemental analysis for 3 hexoses, 1 deoxyhexose, and 1 *N*-acetylhexosamine
Monosaccharide sequence	F.a.b.–m.s. and e.i.–m.s. of permethylated oligosaccharides[c]	Hex–O–HexNAc–O–Hex–O–Hex deoxy-Hex 1, 2, 3, are fragment ions of mol. wt. 219, 638, and 842
Position of linkage	G.c.–m.s. of partially methylated alditol acetates[d] before and after removal of fucose by mild acid hydrolysis[e]	Gal(1→), Fuc(1→), $^{\rightarrow 4}_{\rightarrow 3}$GlcNAc(1→) [and (→4)GlcNAc(1→) after hydrolysis], (→3)Gal(1→), (→4)Glc
Anomeric configuration	Oxidation with chromic oxide,[f] degradation with glycosidases of known specificity, n.m.r.	α or β

Overal conclusion

Galβ(1→4)GlcNAcβ(1→3)Galβ(1→4)Glc
| 1,3
Fuc α



[a] Derivatization largely as described by T. Bhatti, R. E. Chambers, and J. R. Clamp, *Biochim. Biophys. Acta*, 1970, **222**, 339. [b] F.a.b.–m.s. is discussed in Section 5. [c] Refs. 94–98. [d] H. Björndal, C. G. Hellerqvist, B. Lindberg, and S. Svensson, *Agnew. Chem., Int. Ed. Engl.*, 1970, **9**, 610; K. Stellner, H. Saito, and S. Hakomori, *Arch. Biochem. Biophys.*, 1973, **155**, 464; A. M. Lawson, E. F. Hounsell, and T. Feizi, *Int. J. Mass. Spectrom. Ion Phys.*, 1983, **48**, 149. [e] 0.02M, H_2SO_4, 100 °C, 1 h. [f] For example, as described in ref. 21

(> 100 nM). N.m.r. analysis of 100 nM samples can identify oligosaccharides for which previous data are available with the added advantage of leaving the sample intact for biological or immunological assays. For oligosaccharides for which n.m.r. data have not previously been documented, n.m.r. analysis gives valuable information which supplements that obtained by g.c. and m.s. methods (< 10 nM required) or by enzymatic and chemical degradation of radioactively labelled material followed by chromatographic identification (< 1nM required).

D. Conformational Studies of Oligosaccharide Antigens.—Where 100 nM material is available, n.m.r. analysis can go one step further than providing sequence and linkage information, in giving data which can be interpreted to deduce the conformation in water solution of oligosaccharides (Section 3). Knowledge of the solution conformation is important for understanding the 3D shape recognized by anti-carbohydrate antibodies and lectins: for example, in understanding the differences in fine specificities of monoclonal antibodies which recognize different facets of the same molecule (Sections 4 and 5). This type of information is crucial for obtaining biological information from the use of carbohydrate-binding proteins in studies on the tissue distribution of oligosaccharide sequences which are differentiation antigens and tumour markers.

2 The Oligosaccharides of Mucins and Milk: Purification and Antigenic Assignment

The field of carbohydrate research presents a unique challenge in purification and analysis because oligosaccharides vary, not only in composition and sequence of their constituent monosaccharides, but also by the positions and anomeric configuration of monosaccharide linkage. An abundance of oligosaccharides with closely related composition and of isomeric molecules is particularly found with the oligosaccharide chains of mucins. The carbohydrate, which makes up as much as 85% by weight of these large molecular weight glycoproteins, is arranged as multiple chains consisting of from one to twenty or more monosaccharides.[27]

A. Oligosaccharide Chain Biosynthesis, Structure, and Antigenicity.—Oligosaccharide chains are built up by stepwise addition of monosaccharides by glycosyltransferases which are specific for substrate, donor, and linkage. In the biosynthesis of mucin oligosaccharide chains[19,32] the initial step is the attachment of a GalNAc residue to the hydroxyl group (*O*-linked) of Ser or Thr amino acids in the polypetide. The chains are then lengthened by the addition of Gal and GlcNAc residues in a variety of different linkages. The resulting types of core sequences found in mucins of human meconium,[28] the bronchus,[20,33] and colon[34] are shown in Table 3. Backbone sequences are built up of repeating Gal and GlcNAc residues as shown for sheep gastric mucins.[31] From these latter studies a composite picture of the carbohydrate chains of uncharged mucin glycoproteins can be constructed

[32] H. Schacter, *Can. J. Biochem. Cell. Biol.*, 1986, **64**, 163.
[33] G. Lamblin, A. Boersma, A. Klein, P. Roussel, H. van Halbeek, and J. F. G. Vliegenthart, *J. Biol. Chem.*, 1984, **259**, 9051.
[34] D. K. Podolsky, *J. Biol. Chem.*, 1985, **260**, 15510.

Table 3 *Core region oligosaccharides isolated from human gastroinstestinal mucin glycoproteins*[a]

GalNAc→

GalNAcα(1→3)GalNAc→

Galβ(1→3)GalNAc→

GlcNAcβ(1→3)GalNAc→

GlcNAcβ(1→6)GalNAc→

GlcNAcβ1 ↘ 6 GalNAc→
3
Galβ1 ↗

Galβ(1→4)GlcNAcβ1 ↘ 6 GalNAc→
3
GlcNAcβ1 ↗

[a] Data taken from refs. 20, 28, 33, and 34 and substantiated by studies on ovarian cyst mucins (*e.g.* refs. 35, 36). GalNAcα(1→3)GalNAc has also been found in sialylated form in humans in mucin-type glycoprotein of rectal adenocarcinoma (A. Kurosaka, H. Nakajima, T. Funakoshi, M. Matsuyama, T. Nagayo, and T. Yamashina, *J. Biol. Chem.*, 1983, **258**, 11594). Sialic acid is a common component linked to all these cores (*e.g.* see also refs. 21, 33, 34). Terminal GlcNAc residues are more often found with either Galβ(1→3) or Galβ(1→4) substituents. Oligosaccharide chains are built up from these cores as shown in Figure 1

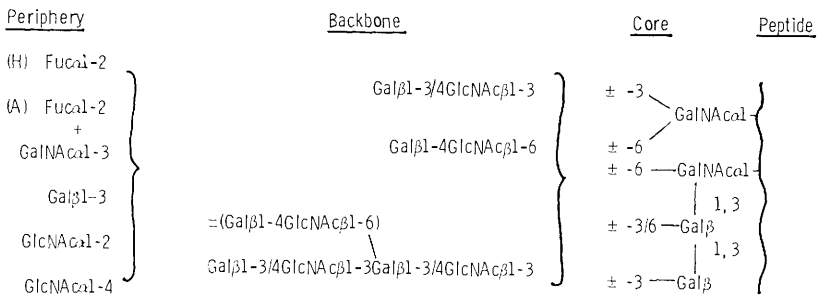


Figure 1 *Composite scheme of the carbohydrate chains of mucins. (H) and (A) are examples of blood group active substitutions. Backbone regions made up of Galβ(1→4)GlcNAc sequences express* I *(branched) and* i *(linear) antigens. Fucose residues (Fuc) also occur linked to backbone GlcNAc to form the Le*[a], and *SSEA-1 antigens. The blood group related T antigen Galβ(1→3)GalNAc and the Tn antigen GalNAcα(1→)Ser/Thr are expressed at the core regions in the absence of additional glycosylation. Peripheral sialic acid residues linked to backbone or core region sequences and sulphate ester groups can occur which mask antigens associated with internal sequences*

(Figure 1). This picture is consistent with the data obtained by several groups on the gastrointestinal mucins of man, horse, hog, and rat (reviwed in ref. 27) and human ovarian cyst mucins.[26,35,36] In addition to the core and backbone regions, a third 'peripheral' domain can be identified and, as shown in Figure 1 and discussed in ref. 27, each of these domains is associated with a particular set of antigens.

The oligosaccharides of human milk (and for that matter the *N*-linked chains of glycoproteins and oligosaccharides of glycolipids *e.g.* see Table 2) express many of the antigenic sequences associated with the backbone and peripheral regions of mucin carbohydrate chains, the difference in milk being that the chains are built up on a lactose core, Galβ(1→4)Glc (Table 1), and only relatively short back bones have so far been characterized.

B. Characterization of Oligosaccharide Antigens.—Except for SSEA-1, the antigens discussed in the legend to Figure 1 were originally characterized as the structures recognized by naturally occurring antibodies.[26,27,30,31,37] SSEA-1 was the first hybridoma-defined antigen to be characterized at the molecular level.[14,15] The anti-SSEA-1 antibody was found to bind to mucin glycoproteins of non-secretor type (lacking in the peripheral ABH antigens which mask antigens associated with backbone sequences). This binding could be inhibited by mucin oligosaccharides containing the 3-fucosyl-*N*-acetyllactosamine sequence.[14] However, a very good inhibitor of binding was a preparation of the milk oligosaccharide lacto-*N*-fucopentose II (LNFP-II), having an isomeric non-reducing end trisaccharide sequence

$$\begin{array}{l} \text{Gal}\beta(1\rightarrow3)\text{GlcNAc} \\ \quad\uparrow 1,4 \\ \quad\text{Fuc}\alpha \end{array} \quad \text{rather than} \quad \begin{array}{l} \text{Gal}\beta(1\rightarrow4)\text{GlcNAc} \\ \quad\uparrow 1,3 \\ \quad\text{Fuc}\alpha \end{array}$$

It therefore appeared at this stage in its characterization that the anti-SSEA-1 antibody might be relatively unspecific in its recognition of antigen, but it turned out that LNFP-II purified by classical chromatographic methods contained a significant proportion of contaminant LNFP-III having the 3-fucosyl-*N*-acetyllactosamine non-reducing end sequence. The two isomers were purified by resorting to acetylation and h.p.t.l.c. and in this way it could be proved that LNFP-III was the active component.[15]

C. Purification of Oligosaccharides.—The characterization of the SSEA-1 antigen demonstrated the need for efficient purification methods for oligosaccharides. Several studies have now shown that oligosaccharides purified by classical chromatographic techniques, and more recently by gel filtration chromatography on BioGel P4, contain more than one structural isomer or indeed different oligosaccharides of closely related size and composition. BioGel P4 (-400 mesh, BioRad) remains as the material of choice for initial purification because the

[35] A. M. Wu, E. A. Kabat, B. Nilsson, D. A. Zopf, F. G. Gruezo, and J. Liao, *J. Biol. Chem.*, 1984, **259**, 7178.
[36] V. K. Dua, B. N. N. Rao, S.-S. Wu, V. E. Dube, and C. A. Bush, *J. Biol. Chem.*, 1986, **261**, 1599.
[37] W. M. Watkins, *Adv. Hum. Genet.*, 1980, **10**, 1 and 379.

chromatography is based on size exclusion where an acetamido sugar is consistently eluted in twice the volume of a neutral sugar, thus giving useful compositional information.[28,38]

H.p.l.c. has now largely superseded h.p.t.l.c. as the preferred method for further purification because of the efficiency in time and yield of each chromatographic step. This is important as often more than one type of chromatographic separation is required for oligosaccharide purification.[39,40] The h.p.l.c. systems widely used for oligosaccharide purification and analysis are reverse-phase octadecylsilyl column packings with water or water–acetonitrile elution, normal-phase chromatography with amine-bonded packings such as aminopropyl silica or anion exchange and water–acetonitrile elution, anion exchange chromatography on amine-bonded and anion exchange column packings eluted with buffers, and cation exchange chromatography.[40]

A combination of reverse- and normal-phase h.p.l.c. was used to purify the core region oligosaccharides of mucins[20,28,34–36] such as those shown in Table 3. Sialylated oligosaccharides of mucins[33,34] and milk[40,41] and sialylated glycolipids (gangliosides)[42] have been separated on amino-bonded column packings with phosphate buffer–acetonitrile elution. H.p.l.c. anion exchange column packings eluted with buffer gradients have been used to separate sialylated and phosphorylated *N*-linked oligosaccharides[43] and sulphated oligosaccharides (Section 5).[44] The h.p.l.c. anion exchange column packings eluted with water–acetonitrile[34,40,45] together with cation exchange chromatography,[46] offer additional systems for separation of non-anionic oligosaccharides.

The use of more than one column and solvent system for oligosaccharide analysis gives greater assurance that purification has been achieved and knowledge of the different chromatographic behaviour of each oligosaccharide can be used in their structural identification. For previously characterized oligosaccharides, this and n.m.r. analysis serve to identify the structure and purity of an oligosaccharide without recourse to degradative structural methods such as those involving g.c. and m.s.

3 The Structural and Conformational Analysis of Oligosaccharides using 500 MHz ^{1}H N.m.r.

High resolution ^{1}H n.m.r. has emerged in the past few years as a powerful technique in the structural analysis of oligosaccharides. Several groups have now published data obtained at 270—500 MHz for various types of oligosaccharide structures

[38] A. Kobata in 'Biology of Carbohydrates' vol. 2, ed. V. Ginsburg, and P. W. Robbins, John Wiley and Sons, 1984, p. 87.
[39] E. F. Hounsell, N. J. Jones, and M. S. Stoll, *Biochem. Soc. Trans.*, 1985, **13**, 1061.
[40] E. F. Hounsell in 'HPLC of Small Molecules', ed. C. K. Lim, 1986, p. 49.
[41] M. L. E. Bergh, P. Koppen, and D. H. van den Eijnden, *Carbohyd. Res.*, 1981, **94**, 225.
[42] G. Gazzotti, S. Sonnino, and R. Ghidoni, *J. Chromatogr.*, 1985, **348**, 371.
[43] J. U. Baenziger and M. Natowicz, *Anal. Biochem.*, 1981, **112**, 357.
[44] P. Scudder, P. W. Tang, E. F. Hounsell, A. M. Lawson, H. Mehmet, and T. Feizi, *Eur. J. Biochem.*, 1986, **157**, 365.
[45] S. J. Mellis and J. U. Baenziger, *Anal. Biochem.*, 1981, **114**, 276.
[46] A. S. R. Donald and J. Feeney, *Carbohydr. Res.*, in press.

including the *N*-linked chains of glycoproteins,[47-50] *O*-linked chains released from protein by base/borohydride degradation,[20,28,36,51,52] naturally occurring free oligosaccharides,[46,53-54] chemically synthesized oligosaccharides and glycosides[55-57] and glycolipids.[58,59]

A. An Introduction to N.m.r. Analysis of Oligosaccharides.—The n.m.r. analyses of oligosaccharides described above were carried out in solution in deuterium oxide (D_2O). The oligosaccharides were first lyophilized several times in D_2O to exchange the OH and NH protons with deuterium. Each of the remaining protons of the glycosidic rings resonates in the applied magnetic field at a specific radiofrequency to give a characteristic signal along the radiofrequency scale of the spectrum (chemical shift, given as p.p.m. of the operating frequency of the instrument). The signal is split into a doublet or multiplet pattern depending on the number of interactions (couplings) with protons on adjacent carbons in the ring and from these splittings coupling constants can be calculated which are characteristic of each monosaccharide type. The chemical shifts of the protons of each monosaccharide are dependent on their local chemical environment and are thus characteristic of monosaccharide sequence and linkage. In general, chemical shifts are comparable for the same oligosaccharide sequences of four or five monosaccharides in different oligosaccharides, enabling a data base for ^1H n.m.r. data to be set up.[60] This is important because n.m.r. analysis relies strongly on the comparison of data with those from standard compounds.

An example of the use of the data base is given in ref. 28 for the structural assignment of the core region sequences of human meconium glycoproteins (Table 3). Two of the oligosaccharide sequences could not be assigned in this way as their n.m.r. data had not previously been documented. Further n.m.r. experiments were therefore carried out to give complete proton assignments for these two oligosaccharides[61] using 2D-correlated spectroscopy (COSY)[62] and spin-decoupling (irradiation at the frequency of a proton signal causing 'decoupling' of signals from protons on adjacent carbon atoms which can then be traced in the spectrum by observing the collapse of their multiplets).

[47] J. F. G. Vliengenthart, L. Dorland, and H. van Halbeek, *Adv. Carbohydr. Chem. Biochem.*, 1983, **41**, 209.
[48] K. Bock, J. Arnarp, and J. Lönngren, *Eur. J. Biochem.*, 1982, **129**, 171.
[49] J.-P. Brisson and J. P. Carver, *Biochemistry*, 1983, **22**, 3671.
[50] S. W. Homans, R. A. Dwek, D. L. Fernandes, and T. W. Rademacher, *Biochim. Biophys. Acta*, 1983, **760**, 256.
[51] H. van Halbeek, *Biochem. Soc. Trans.*, 1984, **12**, 601.
[52] B. N. N. Rao, V. K. Dua, and C. A. Bush, *Biopolymers*, 1985, **24**, 2207.
[53] V. K. Dua, K. Goso, V. E. Dube, and C. A. Bush, *J. Chromatogr.*, 1985, **328**, 259.
[54] E. F. Hounsell, N. J. Jones, H. C. Gooi, T. Feizi, A. S. R. Donald, and J. Feeney, *Carbohydr. Res.*, in press.
[55] C. Augé, S. David, and A. Veyrières, *Nouv. J. Chim.*, 1979, **3**, 491.
[56] R. U. Lemieux, K. Bock, L. T. J. Delbaere, S. Koto, and V. S. Rao, *Can. J. Chem.*, 1980, **58**, 631.
[57] O. Hindsgaul, T. Norberg, J. LePerdu, and R. U. Lemieux, *Carbohydr. Res.*, 1982, **109**, 109.
[58] U. Dabrowski, H. Egge, and J. Dabrowski, *Arch. Biochem. Biophys.*, 1983, **224**, 254.
[59] T. A. W. Koerner, J. H. Prestegard, P. C. Demou, and R. K. Yu, *Biochemistry*, 1983, **22**, 2676.
[60] E. F. Hounsell, D. J. Wright, A. S. R. Donald, and J. Feeney, *Biochem. J.*, 1984, **223**, 129.
[61] J. Feeney, T. A. Frenkiel, and E. F. Hounsell, *Carbohydr. Res.*, 1986, **152**, 63.
[62] A. Bax and L. Lerner, *Science*, 1986, **232**, 960.

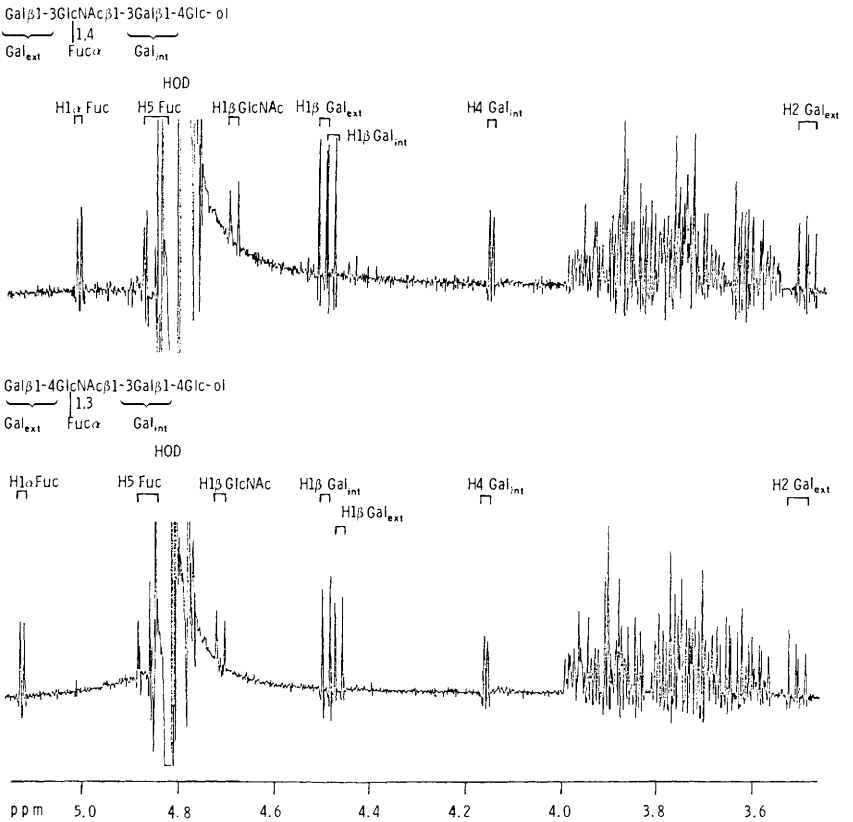


Figure 2 *The* 500 MHz ^{1}H *n.m.r. spectra of LNFP-*II*-ol and LNFP-*III*-ol*

Besides giving sequence and linkage information, comparison of the data from ^{1}H n.m.r. analysis of a series of structurally related oligosaccharides can lead to information on their solution conformations, as described in Sections 3B—D below for the oligosaccharides of milk used in the characterization of the SSEA-1 antigen.

B. The Spectra of the Alditols of LNFP-II and LNFP-III.—The 500 MHz ^{1}H n.m.r. spectra of LNFP-II-ol and LNFP-III-ol are shown in Figure 2. Approximately 120μg (140nM) were used for the analysis. The signals that can be readily recognized have been assigned in the figure. These are the structural reporter groups[47] which resonate at frequencies away from the main bulk of the proton signals. For example, the ^{1}H signals for the H-1α and H-5 of Fuc, the H-1β of GlcNAc, and the H-1β of the two galactose residues are to low-field (high

radiofrequency) of the bulk of the proton signals. The H-2 signal for the non-reducing end Gal (Gal_{ext}) is also readily distinguished (by a decoupling experiment at the H-1 frequency) and is located at a slightly higher field than the bulk region proton signals. The methyl protons of the acetamido group and C-6 of Fuc are also at high-field near the internal standard acetone, which has chemical shift 2.225 p.p.m. with respect to sodium 4,4-dimethyl-4-silapentane-1-sulphonate (DSS) at 295 K (the temperature of the experiments). The H-1 protons marked in the spectra of Figure 2 are doublets (coupling to H-2 only), the H-2 is a quadruplet (coupling to H-1 and H-3) and the H-5's are octuplets (coupling to H-6, H-6′, and H-4). The H-1β signals can be distinguished because that for GlcNAc characteristically has a larger coupling constant than those of Gal residues. The signals arising from the two Gal residues of each molecule can be distinguished because the chemical shifts for the Gal attached to Glc-ol and Glc are different in the reduced and unreduced oligosaccharides (Table 4).

C. Interpretation of ^{1}H N.m.r. Spectra to Give Structural and Conformational Information.—As can be seen from the chemical shift data given in Table 4, in general comparable residues in the related oligosaccharides have very similar chemical shifts *e.g.* the internal Gal residue (Gal_{int}), Glcα, and Glcβ in LNT, LNNT, LNFP-II, and LNFP-III. The addition of Fuc to LNT and LNNT to give, respectively, the non-reducing end trisaccharide sequence of LNFP-II and LNFP-III, leads to differences in the chemical shifts of residues in these sequences as does change of linkage from Galβ(1→3)GlcNAc to Galβ(1→4)-GlcNAc.

As illustrated in Figure 3, the differences in chemical shift for the non-reducing end trisaccharide sequences of LNFP-II and LNFP-III have been attributed [56,63,64] to the close proximity of the H-5 of Fuc to the ring oxygen of the external Gal residue (Gal_{ext}) and the oxygen of the Galβ(1→3) or Galβ(1→4)GlcNAc glycosidic bonds. The H-1 of Galβ(1→3) is then adjacent to the acetamido group whereas that of Galβ(1→4) is adjacent to GlcNAc C-6. The H-5 signals of Fuc are significantly deshielded [63,64] by close proximity to the two oxygen atoms giving low-field signals of 4.882 and 4.830 p.p.m. (LNFP-II and LNFP-III, respectively) as compared for example, with the H-5 of Fucα (1→2) of another isomer, LNFP-I

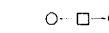
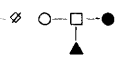
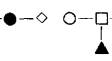
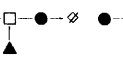
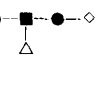
$$Fuc\alpha(1\rightarrow2)Gal\beta(1\rightarrow3)GlcNAc\beta(1\rightarrow3)Gal\beta(1\rightarrow3)Glc$$

which has a Fuc H-5 chemical shift of 4.293 p.p.m.[54]

The similarity in the chemical shifts of the non-reducing/non-reduced end portions of each oligosaccharide/alditol pair can be interpreted as the reducing/reduced end being orientated away from the rest of the molecule as predicted by empirical energy calculations [65] and shown in Figure 4 for LNFP-III-ol.

[63] K. Bock, *Pure Appl. Chem.*, 1983, **55**, 605.
[64] U. Spohr, N. Morishima, O. Hindsgaul, and R. U. Lemieux, *Can. J. Chem.*, 1985, **63**, 2659.
[65] M. Biswas and V. S. R. Rao, *Int. J. Quantum Chem.*, 1981, **20**, 99.

Table 4 *The ^1H chemical shifts (p.p.m. from* DSS*) of the structural reporter groups of human milk oligosaccharides and their alditols*[a]

		LNT	LNT-ol	LNNT	LNNT-ol	LNFP-II	LNFP-II-ol	LNFP-III	LNFP-III-ol	FLNH
										
Fuc ▲, △	H-1					5.028	5.029	5.127	5.131	5.105
	H-5					4.882	4.870	4.830	4.842	4.864
	H-6					1.180	1.179	1.174	1.175	1.172
Gal_{ext} ○	H-1	4.441	4.442			4.505	4.507			4.440
	H-2	3.524	3.529			3.485	3.486			3.524
Gal_{ext} ●	H-1			4.480	4.479			4.462	4.463	4.452
	H-2			3.539	3.536			3.497	3.497	3.496
GlcNAc □	H-1α	4.731[b]	4.731			4.706	4.697			4.730
	H-1β	4.726				4.695				4.726
	NAc	2.027	2.025			2.032	2.032			2.025
GlcNAc ■	N-1α			4.707	4.708			4.715	4.713	4.638
	H-1β			4.703				4.710		4.638
	NAc			2.034	2.034			2.022	2.023	2.052
Gal_{int} ●	H-1	4.441	4.492	4.437	4.495	4.433	4.489	4.432	4.488	4.431
	H-4	4.154	4.144	4.155	4.150	4.156	4.152	4.156	4.152	4.140
Glcα ◇	H-1	5.220	—[c]	5.220	—	5.220	—	5.219	—	5.219
Glcβ	H-1	4.662	—	4.663	—	4.663	—	4.662	—	4.666
	H-2	3.279	—	3.278	—	3.279	—	3.277	—	3.290

[a] Symbols used: △, Fucα(1→3); ▲, Fucα(1→4); ○, Galβ(1→3); ●, Galβ(1→4); □, GlcNAc linked at C-3 by Gal; ■, GlcNAc linked at C-4 by Gal; ◇, Glc; ⊘, Glc-ol. [b] Signals arising from the α and β anomers of reducing oligosaccharide (anomerization effect *e.g.* see ref. 90). [c] Not determined.

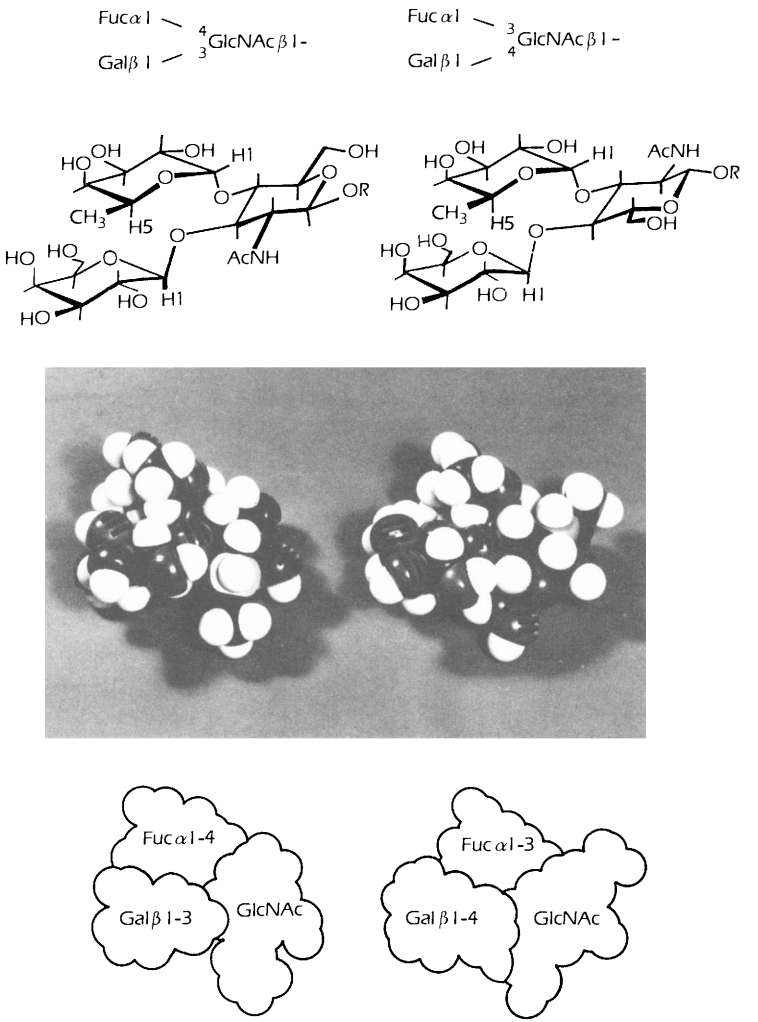
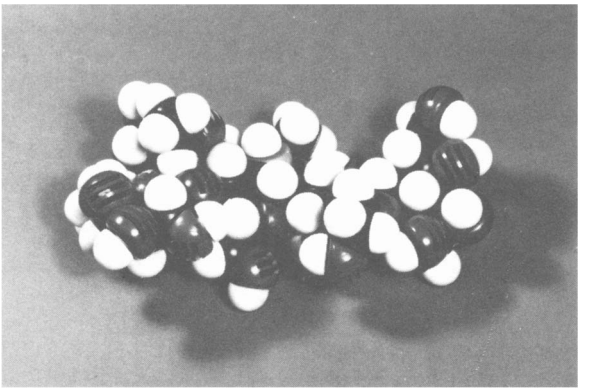


Figure 3 *The structural formulae, molecular models and outline of the molecular models proposed for the non-reducing end trisaccharide sequences of LNFP-*II *and LNFP-*III. *The atoms making up the molecular models are white,* H; *light grey,* N; *dark grey,* O; *and black,* C

Several predictions can also be made about the conformation of FLNH from a comparison of the ^{1}H n.m.r. chemical shift data in Table 4 (similar data have been obtained by Dua *et al.*[53]). The chemical shifts of GlcNAc and non-reducing end Gal, (Gal_{ext}), in the sequence Galβ(1→3)GlcNAcβ(1→3)Gal are essentially identical in LNT and FLNH. The chemical shifts for the Fuc and GlcNAc residues in the sequence Galβ(1→4)[Fucα(1→3)]GlcNAc of FLNH are close to those of the

LNFP-III-ol

Galβ1-4 GlcNAcβ1-3 Galβ1-4 Glc-ol
|1,3
Fucα



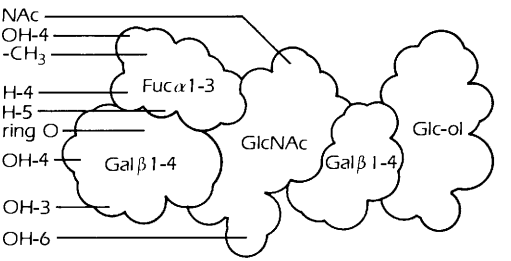
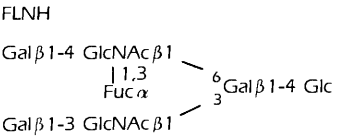


Figure 4 *The structural formula, molecular model and outline of the molecular model proposed for LNFP-*III*-ol. The atoms appear coloured as described in Figure* 3

corresponding shifts of LNFP-III, thus confirming that the single fucose residue is in this 3-fucosyl-*N*-acetyllactosamine sequence. The differences in chemical shift between the fucosylated terminal trisaccharide unit of LNFP-III and FLNH can therefore be accounted for by the 3-fucosyl-*N*-acetyllactosamine sequence being followed by a 1→3 linkage to Gal_{int} in LNFP-III and a 1→6 linkage to Gal_{int} in FLNH. In addition, the chemical shift of the anomeric proton of the fucosylated GlcNAc is smaller in FLNH, whereas that of the *N*-acetamido methyl group is larger (by 0.074 and 0.029 p.p.m., respectively). An explanation for this in terms of relative deshielding is given by the molecular model of FLNH depicted in Figure 5. This shows the anomeric proton of GlcNAc adjacent to the hydrogens of C-6 of galactose and the methyl protons of the *N*-acetamido group in close proximity to



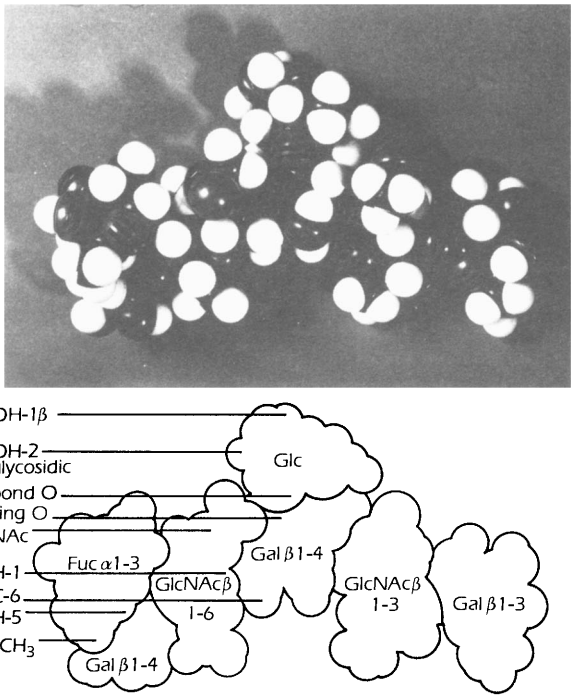


Figure 5 *The structural formula, molecular model and outline of the molecular model proposed for FLNH. The atoms appear coloured as described in Figure 3*

the ring oxygen of the galactose linked to glucose and the glycosidic oxygen of the Galβ(1→4)Glc bond.

D. The Construction of Space-filling Molecular Models Incorporating ^{1}H N.m.r. Data.—The molecular models shown in Figures 3—5 were constructed using the information obtained from the n.m.r. data as discussed above, together with information from other sources[52,56,57,63–70] obtained by *X*-ray crystallography,

[66] J. F. Stoddart, 'Stereochemistry of Carbohydrates', John Wiley and Sons Inc., 1971.
[67] D. A. Rees, *Adv. Carbohydr. Chem. Biochem.*, 1969, **24**, 267.
[68] R. U. Lemieux and S. Koto, *Tetrahedron*, 1974, **30**, 1933.
[69] H. Thøgersen, R. U. Lemieux, K. Bock, and B. Meyer, *Can. J. Chem.*, 1982, **60**, 44.
[70] R. U. Lemieux and K. Bock, *Arch. Biochem. Biophys.*, 1983, **221**, 125.

empirical energy calculations, and n.m.r. nuclear Overhauser experiments (measurement of the change in intensity of a signal on irradiation of another resonance caused by a through-space interaction of close protons).[62] These studies have shown that in solution oligosaccharides tend to adopt conformations in which D-sugars have the $^4C_1$ conformation and L-sugars the $^1C_4$ conformation (as shown in Figure 3) and in the most energetically favoured conformers the angles which define the orientation of residues around the glycosidic bonds (the φ and ψ angles) are such that the anomeric and aglyconic protons of each linkage are on the same face of the molecules (φH approximately 60° for β-D-glycosides and approximately −60° for α-D-glycosides[63]). Although this is only an approximation to the likely solution conformation of an oligosaccharide at any particular time, the molecular models of structurally similar oligosaccharides are useful in interpreting major differences in their chemical shifts and in envisaging the molecular features recognized by carbohydrate binding proteins as has been discussed previously[54] and is explained further in Section 4.

4 Studies of Carbohydrate Antigen Recognition using Space-filling Molecular Models

As discussed in Section 2 the anti-SSEA-1 antibody specifically recognizes oligosaccharides having the 3-fucosyl-*N*-acetyllactosamine sequence,

$$\begin{array}{l} \text{Gal}\beta(1\rightarrow4)\text{GlcNAc} \\ \qquad\qquad\uparrow 1,3 \\ \qquad\qquad\text{Fuc}\alpha \end{array}$$

Oligosaccharides having the isomeric Le^a active structure,[37]

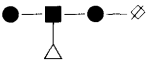
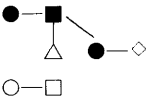
$$\begin{array}{l} \text{Gal}\beta(1\rightarrow3)\text{GlcNAc} \\ \qquad\qquad\uparrow 1,4 \\ \qquad\qquad\text{Fuc}\alpha \end{array}$$

are not recognized by anti-SSEA-1. As shown by the molecular models of these two trisaccharides (Figure 3), the significant features which discriminate between the two models are that in the LNFP-II-type structure the Fuc-methyl and GlcNAc *N*-acetamido group are on different surfaces of the model (top and bottom, respectively) whereas in the LNFP-III-type structure they are in close proximity on the top surface of the molecule. As it is known that the fucose residue and acetamido group are necessary for binding of anti-SSEA-1 (*N*-acetyllactosamine and 3-fucosyllactose are relatively inactive as inhibitors of binding) it is thought that this part of the molecule is specifically recognized and that it is the orientation of the Fuc-methyl and acetamido groups to each other and to those of other atoms in the molecule, *e.g.* hydroxy groups of the Fuc and non-reduced end Gal residue, that are recognized by the antibody combining site (Figures 4 and 6).

From other studies of oligosaccharide binding by lectins,[57,71–74] anti-

[71] R. U. Lemieux in 'Frontiers of Chemistry', ed. K. J. Laidler, Pergamon Press, 1982, p. 3.
[72] U. Spohr, O. Hindsgaul, and R. U. Lemieux, *Can. J. Chem.*, 1985, **63**, 2644.
[73] O. Hindsgaul, D. P. Khare, M. Bach, and R. U. Lemieux, *Can. J. Chem.*, 1985, **63**, 2653.
[74] K. A. Kronis and J. P. Carver, *Biochemistry*, 1985, **24**, 834.

Table 5 *Differing reaction patterns of monoclonal antibodies recognizing 3-fucosyl-N-acetyllactosamine*

Antibody	*Immunogen*[a]	*Activity with different oligosaccharides*[b]		
		3FLN[c] 	LNFP-III-ol 	FLNH 
Anti-SSEA-1	Mouse teratocarcinoma cells	+ + + +	+ + +	+
Vim/Vep	Human myeloid cells	+	+ +	NT
3C1B12	EGF receptor	−	+ +	+

[a] References as in text (Section 4). [b] Activity measured as nM oligosaccharide giving 50% inhibition of binding; + + + +, < 1 nM; + + +, < 2 nM; + +, < 4 nM; +, < 10 nM; NT, not tested. [c] Symbols as in legend to Table 4

bodies,[52,54,56,64,75–77] and other carbohydrate-binding proteins[78] it is thought that areas with a predominance of protons interact with hydrophobic amino acids in the combining site and that surrounding hydroxyl groups form specific hydrogen bonds with polar amino acids.

Prediction of the molecular features recognized by carbohydrate-binding proteins can be made by analysis of the inhibition data of a series of proteins with related specificities and a series of oligosaccharides with different inhibitory activities isolated from biological sources or synthesized by site-specific chemical reactions.[64,69,71–73,75–77] For studies on antibody recognition of 3-fucosyl-*N*-acetyllactosamine there were available three oligosaccharides having this trisaccharide sequence; the trisaccharide itself (3FLN), LNFP-III-ol, and FLNH. These had different inhibitory activities towards several antibodies with specificities related to SSEA-1. In addition to the SSEA-1 antibody, other antibodies recognizing the 3-fucosyl-*N*-acetyllactosamine sequence which have now been characterized include a series of related antibodies belonging to the Vim and Vep series which were raised against human myeloid cells,[79,80] and the antibody 3C1B12 raised against the receptor for epidermal growth factor (EGF).[81] An

[75] R. U. Lemieux, A. P. Venot, U. Spohr, P. Bird, G. Mandal, N. Morishima, O. Hindsgaul, and D. R. Bundle, *Can. J. Chem.*, 1985, **63**, 2664.
[76] E. A. Kabat, J. Liao, M. H. Burzynska, T. C. Wong, H. Thøgersen, and R. U. Lemieux, *Mol. Immunol.*, 1981, **18**, 873.
[77] R. U. Lemieux, T. C. Wong, J. Liao, and E. A. Kabat, *Mol. Immunol.*, 1984, **21**, 751.
[78] F. A. Quiocho, *Ann. Rev. Biochem.*, 1986, **55**, 287.
[79] H. C. Gooi, S. J. Thorpe, E. F. Hounsell, H. Rumpold, D. Kraft, O. Förster, and T. Feizi, *Eur. J. Immunol.*, 1983, **13**, 306.
[80] H. C. Gooi, E. F. Hounsell, A. Edwards, O. Majdic, W. Knapp, and T. Feizi, *Clin. Exp. Immunol.*, 1985, **60**, 151.
[81] H. C. Gooi, E. F. Hounsell, I. Lax, R. M. Kris, T. A. Libermann, J. Schlessinger, J. D. Sato, T. Kawamoto, J. Mendelsohn, and T. Feizi, *Biosci. Rep.*, 1985, **5**, 83.

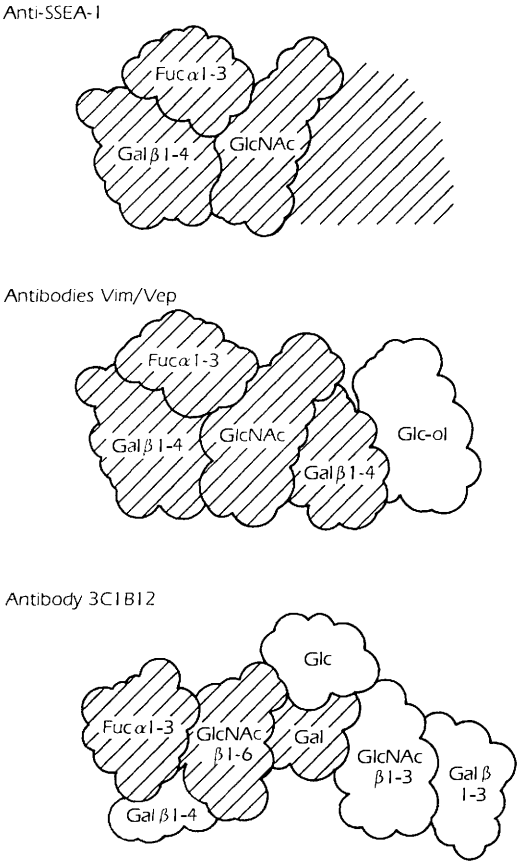


Figure 6 *The outlines for the molecular models of FLN, LNFP-*III*-ol, and FLNH taken from the molecular models shown in Figures* 3, 4, *and* 5, *respectively, with shading signifying regions containing features recognized by antibody anti-SSEA-*1, *antibodies of the Vim and Vep series, and antibody* 3C1*B*12 *raised against the receptor for epidermal growth factor (for references see text). It is proposed that antibody SSEA-*1 *recognizes molecular features in addition to those represented in* 3*-fucosyl-*N*-acetyllacetosamine, but which are not present in LNFP-*III*-ol and FLNH,* e.g. β(1→3)*Gal*β(1→4)*GlcNAc rather than* β(1→3)*Gal*β(1→4)*Glc-ol. On the other hand, antibodies of the Vim and Vep series recognize mostly features of* 3*-fucosyl-*N*-acetyllactosamine itself and antibody* 3C1*B*12 *may recognize the branched sequence*

insight into the topographical array of atoms recognized by these antibodies is given by comparison of their relative reactivities with the three different oligosaccharides shown in Table 5, as follows. Anti-SSEA-1 reacts slightly better with 3FLN than LNFP-III-ol and is relatively unreactive with FLNH. This suggests that the glucitol of LNFP-III-ol, and even more the Galβ(1→3)GlcNAc sequence of FLNH interfere with antibody binding. Thus, it is proposed that the combining site for SSEA-1 recognizes the complete 'top' surface of 3FLN and

probably extends recognition beyond this to include for example a GlcNAc residue which is usually present in longer oligosaccharide chains of glycoconjugates in place of Glc-ol of LNFP-III-ol (Figure 6). On the other hand, antibodies of the Vim/Vep series and 3C1B12 react more strongly with LNFP-III-ol than 3FLN, and 3C1B12 shows two times less reactivity with FLNH than LNFP-III-ol. These antibodies therefore seem to recognize the internal Gal residues of LNFP-III-ol and FLNH which extend the hydrophobic face of the molecule of Figures 3—6. Some further aspect of the reducing end of LNFP-III-ol and the Galβ(1→3)GlcNAc sequence of FLNH may also fit into the combining site which indicates that the antibodies may be directed towards branched structures. From additional molecular models[54] and the knowledge that antibody 3C1B12 has some reactivity towards oligosaccharides with non-reducing end GalNAcα(1→3) and Fucα(1→2) linked to the Galβ(1→4)GlcNAc sequence, it is thought that the external Gal residue of LNFP-III-ol and FLNH is not involved in this antibody recognition.

5 Characterization of Sulphated Oligosaccharide Antigens Based on the *N*-Acetyllactosamine Sequence

Alterations in sulphated oligosaccharide sequences of proteoglycans and glycoproteins are being implicated increasingly in cell development and differentiation. For example: specifically sulphated sequences of heparin are thought to be involved in regulation of cellular proliferation;[82-84] several changes in proteoglycan composition are associated with differentiation[85] and age- and disease-related changes in cornea and cartilage;[86,87] sulphated glycoproteins show developmentally regulated changes in the liver and lung of chick embryos;[88] and the sulphation of *N*-acetyllactosamine sequences of the LFA-1 antigen of lymphocytes is restricted to the T-cell lineage.[89] For the study of these and related changes it is becoming increasingly important to have specific antibodies to sulphated sequences and methods available for their characterization.

A. Sulphated Antigens of the *N*-Acetyllactosamine Series.—Among the first sulphated oligosaccharide antigens to be characterized at the molecular level were sulphated poly-*N*-acetyllactosamine sequences which are the major component of the proteoglycan keratan sulphate. Three hybridoma antibodies raised against keratan sulphate were shown to recognize polysulphated hexa- and larger oligosaccharides isolated from bovine corneal keratan sulphate by endo-β-

[82] R. Crum, S. Szabo, and J. Folkman, *Science*, 1985, **230**, 1375.
[83] J. J. Castellot, J. Choay, J.-C. Lormeau, M. Petitou, E. Sache, and M. J. Karnovsky, *J. Cell Biol.*, 1986, **102**, 1979.
[84] J. T. Gallagher, M. Lyon, and W. P. Steward, *Biochem. J.*, 1986, **236**, 313.
[85] R. Kapoor and P. Prehm, *Eur. J. Biochem.*, 1983, **137**, 589.
[86] N. SunderRaj, J. Willson, J. D. Gregory, and S. P. Damle, *Curr. Eye Res.*, 1985, **4**, 49.
[87] A. R. Poole, *Biochem. J.*, 1986, **236**, 1.
[88] A. Heifetz, W. H. Kinsey, and W. J. Lennarz, *J. Biol. Chem.*, 1980, **255**, 4528.
[89] N. M. Dahms and G. W. Hart, *J. Immunol.*, 1985, **134**, 3978.

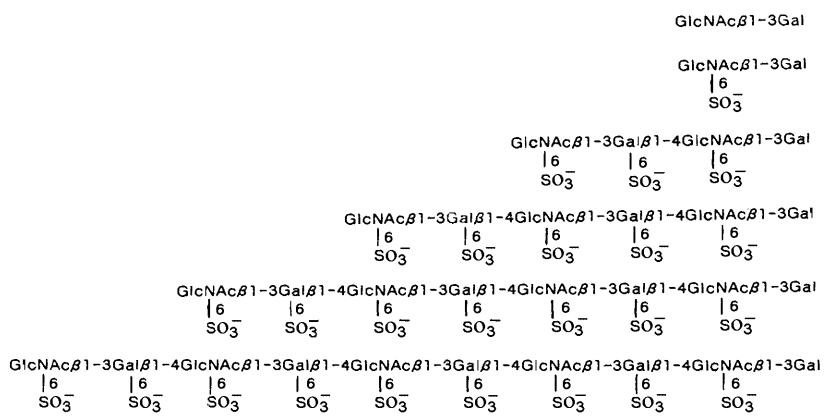


Figure 7 *The sequences of the major oligosaccharides isolated from bovine corneal keratan sulphate peptidoglycans by endo-β-galactosidase digestion, Bio-Gel P4 chromatography and h.p.l.c.*

galactosidase digestion of the peptidoglycan.[44,90,91] The major oligosaccharides characterized were as shown in Figure 7; a non-sulphated disaccharide, a monosulphated disaccharide, a trisulphated tetrasaccharide, a pentasulphated hexasaccharide, a heptasulphated octasaccharide, and a nonasulphated decasaccharide. The pentasulphated hexasaccharide was the smallest oligosaccharide with antigenic activity with the three antibodies. Activity was lost on desulphation indicating the importance of one or more sulphate groups for antibody recognition. The number and linkage position of the sulphate groups and the backbone sequence were deduced as described below and in references 44, 90, and 91.

B. Structural Characterization of Sulphated Oligosaccharides.—(i) *Chromatography.* The oligosaccharides released from keratan sulphate by the enzyme endo-β-galactosidase were purified by gel filtration chromatography on BioGel P4 (-400 mesh) eluted in ammonium bicarbonate buffer and anion exchange h.p.l.c. (Varian AX-4 column eluted with a phosphate buffer gradient from 10—400mM in 45 min at flow rate 1ml min^{-1}). From the known specificity of the enzyme[92] it could be deduced that GlcNAc–Gal sequences were present terminating in a non-sulphated Gal residue. From the chromatographic elution pattern of the oligosaccharides it could be deduced that these were di- to dodeca-saccharides increasing in size by mono- or di-sulphated *N*-acetylhexosamine-hexose units. The oligosaccharide composition was confirmed by GC analysis.

(ii) *Fast Atom Bombardment Mass Spectrometry.* The number of sulphates, *N*-acetylhexosamines, and hexoses present on the major oligosaccharides were

[90] E. F. Hounsell, J. Feeney, P. Scudder, P. W. Tang, and T. Feizi, *Eur. J. Biochem.*, 1986, **157**, 375.
[91] H. Mehmet, P. Scudder, P. W. Tang, E. F. Hounsell, B. Caterson, and T. Feizi, *Eur. J. Biochem.*, 1986, **157**, 385.
[92] P. Scudder, P. Hanfland, K. Uemura, and T. Feizi, *J. Biol. Chem.*, 1984, **259**, 6586.

calculated from the molecular weight given by their f.a.b.–m.s. spectra. This is a relatively new technique applied to the structural analysis of oligosaccharides as a method for determining the sequence and composition of constituent monosaccharides.[93] Previously, the physicochemical analysis of oligosaccharide sequence had been achieved by e.i. direct probe mass spectrometry of permethylated derivatives.[31,94–97] An advantage of the f.a.b. ionization method is that it can be used for the analysis of non-derivatized oligosaccharides. However, further studies *e.g.* refs.98, 99 have suggested that derivatized oligosaccharides give more useful f.a.b.–m.s. spectra containing fragment ions which provide sequence information. Methylated or acetylated derivatives are the method of choice analysed in positive ion f.a.b. in a 1 : 1 glycerol/thioglycerol matrix or 1 : 1 glycerol/triethylamine matrix.

For the series of sulphated oligosaccharides shown in Figure 7, negative ion f.a.b.–m.s. of the underivatized oligosaccharides using a thioglycerol matrix gave fragment ions resulting from the loss of sulphate residues (as their sodium or potassium adducts). No fragmentation due to glycosidic bond cleavage could be detected. However, the pattern of fragment ions given by loss of sulphate were most instructive in assigning the correct molecular weight as the molecular ion was relatively small. The sulphate groups were found not to be stable to methylation or acetylation and therefore additional sequence information could not be obtained by this method.

(iii) *N.m.r. Analysis.* For the non-sulphated and mono-sulphated disaccharides of keratan sulphate (Figure 7), complete assignment of all the signals in the spectra could be made using spin decoupling methods. From this analysis it was shown that the mono-sulphate residue was at the C-6 position of GlcNAc linked 1→3 to Gal. The protons on the sulphated C-6 atom had a characteristically large chemical shift, compared to those of the non-sulphated oligosaccharide, caused by deshielding by the sulphate group. The effect on the chemical shift of the other protons around the sulphated glycosidic ring was roughly proportional to their distance from C-6, suggesting that the sulphate group extended away from the oligosaccharide backbone.

Spectra of the tetra- and hexa-saccharides could be interpreted, by the use of 2D-spectroscopy methods and comparison of the data with those for the disaccharides, to give the structures as shown in Figure 7. The larger oligosaccharides, which were initially analysed as mixtures, gave spectra which were consistent with them being of the same 6-sulphated poly-*N*-acetyllactosamine sequence, but little detailed information could be obtained. For both the f.a.b.–m.s. and ^{1}H-n.m.r. studies, analysis of the octa-, deca-, and dodeca-saccharides was greatly aided by having an homologous series of oligosaccharides of increasing size. The information obtained

[93] A. Dell and C. Ballou, *Biomed. Mass Spectrom.*, 1983, **10**, 50.
[94] K.-A. Karlsson, I. Pascher, W. Pimlott, and B. E. Samuelsson, *Biomed. Mass Spectrom.*, 1974, **1**, 49.
[95] K. Watanabe, M. E. Powell, and S. Hakomori, *J. Biol. Chem.*, 1978, **253**, 8962.
[96] N. K. Kochetkov, V. A. Derevitskaya, and N. P. Arbatsky, *Eur. J. Biochem.*, 1976, **67**, 129.
[97] M. McNeil, A. G. Darvill, P. Amån, L.-E. Franzén, and P. Albersheim, *Methods Enzymol.*, 1982, **83**, 3.
[98] E. F. Hounsell, M. J. Madigan, and A. M. Lawson, *Biochem. J.*, 1984, **219**, 947.
[99] S. Naik, J. E. Oates, A. Dell, G. W. Taylor, P. M. Dey, and J. B. Pridham, *Biochem. Biophys. Res. Commun.*, 1985, **132**, 1.

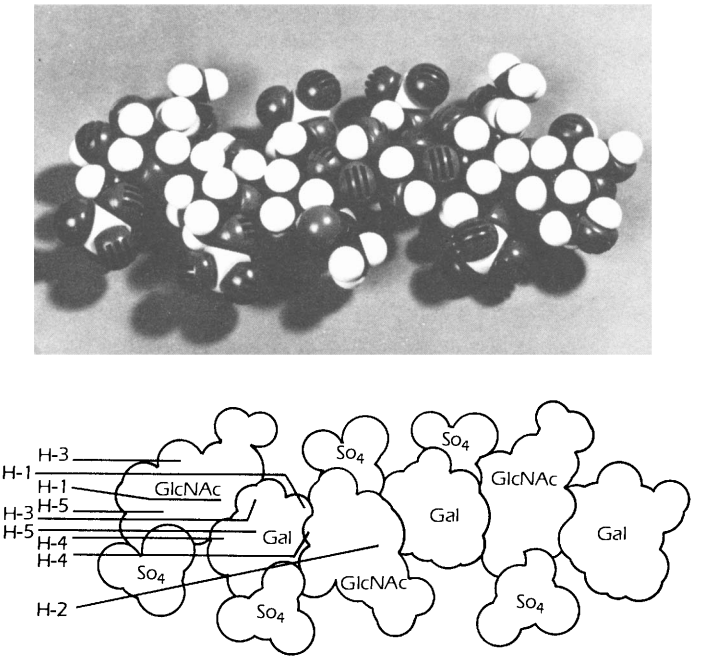


Figure 8 *A molecular model for the penta-sulphated hexasaccharide isolated from bovine corneal keratan sulphate based on n.m.r. studies and* X-*ray crystallographic data* (*refs.* 90, 100, 101)

from these studies serves as a model for the characterization of other sulphated sequences now increasingly being found in glycoproteins and proteoglycans.

The n.m.r. data were consistent with the molecular model for the pentasulphated hexasaccharide based on *X*-ray crystallographic data[100,101] shown in Figure 8. The anionic sulphate groups are arrayed in pairs along both the top and bottom surfaces of the molecule to give a distinct topography of hydrophilic and hydrophobic areas. The two non-reducing end pairs of sulphates surround a relatively hydrophobic region on the face of the molecule shown in the Figure, including the H-1, -3, and -5 of the non-reducing end GlcNAc, the H-1, -3, -4, and -5 of the Gal to which it is linked, and the H-2, and -4 of the next GlcNAc residue. The fully sulphated GlcNAc–Gal–GlcNAc–Gal sequence is repeated three times on the two faces of the decasaccharide (Figure 9) which was the best inhibitor of antibody binding with two of the antibodies,[91] suggesting the importance of this sequence in antibody binding and cooperativity effects.

[100] S. Arnott, J. M. Guss, D. W. L. Hukins, I. C. M. Dea, and D. A. Rees, *J. Mol. Biol.*, 1974, **88**, 175.
[101] D. A. Rees, *MTP Int. Rev. Sci. Biochem. Ser. 1*, 1975, **5**, 1.

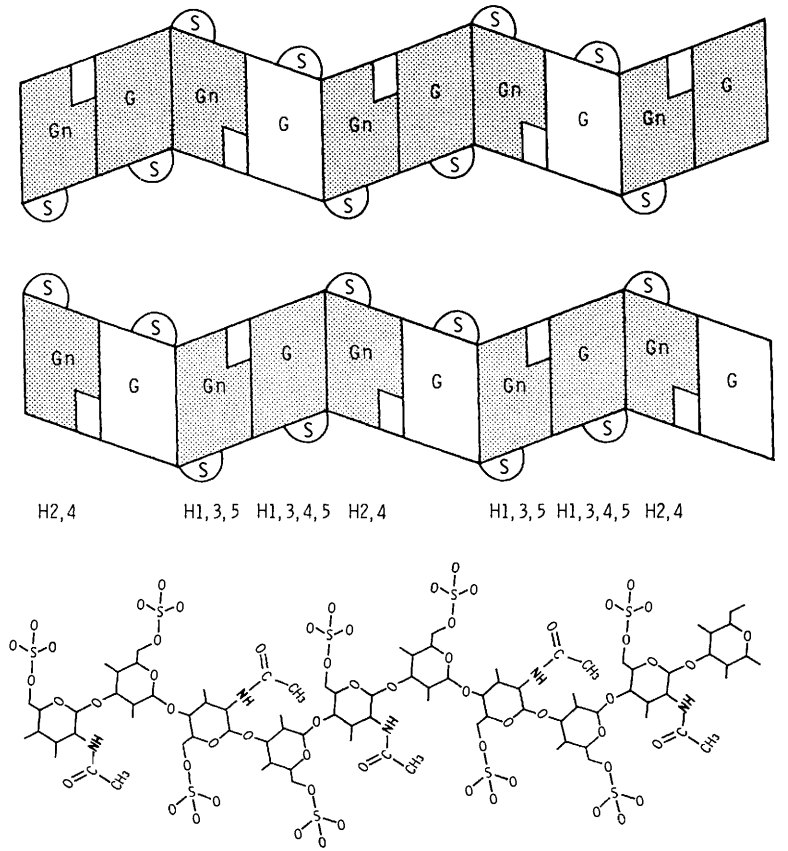


Figure 9 *A schematic diagram of the nona-sulphated decasaccharide isolated from keratan sulphate. The face of the hexasaccharide molecule shown in Figure* 8 *is represented in the upper diagram. The middle diagram shows the reverse side of the molecule and the lower, the conventional structure for this face. (Adopted from refs.* 90, 100, 101). *The stippled areas are those having the* H-1, -3, -5 *of GlcNAc (Gn), the* H-1, -3, -4, -5 *of Gal(G), and* H-2 *and* H-4 *of Gn. The small square is the* N-*acetamido group and* S *the sulphate. The tetra-sulphated tetra-saccharide sequence having this relatively hydrophobic Gn–G–Gn sequence linked to sulphated Gal is absent from the tetrasaccharide (Figure 7) which had negligible antigenic activity and is present one, two, and three times, respectively, in the hexa-, octa-, and deca-saccharides which had increasing antigenic activities, suggesting that a fully sulphated tetra-saccharide is the predominant feature recognized by the antibody-combining site*

6 Conclusions

There is a considerable diversity of carbohydrate structures possible from a combination of a few monosaccharides—taking into account different monosaccharide sequence, position of linkage, and anomeric configuration. They have, therefore, great potential as biological information molecules. The studies

described in this review show how changes in glycosylation pattern *i.e.* 6-sulphation and 3-fucosylation of *N*-acetyllactosamine sequences, result in different antigens defined by monoclonal antibodies and, further, that the features recognized are topographical arrays of atoms giving both hydrophilic and hydrophobic areas which are thought to interact, respectively, with polar and non-polar amino acids in the antibody-combining site. These studies provide a basic understanding of oligosaccharide antigen conformation and molecular recognition. More detailed characterization using empirical energy calculations, *X*-ray crystallography, and nuclear Overhauser experiments can be used to give additional insight into carbohydrate–protein interactions (*e.g.* refs. 52, 56, 57, 63—65, 69—78). The principles established in elucidating the specificities of monoclonal antibodies to cell surface markers are also relevant to future studies on the recognition of carbohydrate receptors by microorganisms,[102–104] to studies on carbohydrates as targets for autoantibodies,[30,31,103] and to work on cell-to-cell interactions mediated by endogeneous lectins.[104–106]

Acknowledgements. The personal studies described in this review could not have been carried out without valuable collaboration with Dr. T. Feizi, Dr. A. Lawson, and Dr. J. Feeney and colleagues. The author is also grateful to the Royal Society of Chemistry and Tate and Lyle Ltd. for the 1984 Carbohydrate Chemistry Prize, awarded for the work on which this review is largely based, and to Mrs. M. Runnicles for patient typing of the manuscript.

[102] 'Microbial Lectins and Agglutinins', ed. D. Mirelman, John Wiley and Son, 1986.
[103] L. M. Loomes, K. Uemura, R. A. Childs, J. C. Paulson, G. N. Rogers, P. R. Scudder, J.-C. Michalski, E. F. Hounsell, D. Taylor-Robinson, and T. Feizi, *Nature*, 1984, **307**, 560.
[104] S. H. Barondes, *Science*, 1984, **223**, 1259.
[105] 'The Biology of Glycoproteins', ed. R. J. Ivatt, Plenum Press, 1985.
[106] 'Receptors in Cellular Recognition and Developmental Processes', ed. R. M. Gorczynski, Academic Press, 1986.