High Resolution NMR in the Determination of Structure in Complex Carbohydrates

C. Allen Bush

Department of Chemistry,
Illinois Institute of Technology
Chicago, IL 60616 U.S.A.

Contents

I. Introduction: A Description of Complex Carbohydrates and their Biological Function 73

II. The Determination of Covalent Structure 75

III. The "Structural Reporter" Method in ^{1}H NMR 76

IV. Structure Determination by Complete Assignment of the ^{1}H NMR Spectrum 79

V. Prospects for Carbon NMR 88

VI. NMR Spectroscopy of Other Nuclei: ^{31}P, ^{17}O, ^{15}N 91

VII. High Resolution NMR of Intact Glycoproteins 93

VIII. References 94

I. Introduction: A Description of Complex Carbohydrates and their Biological Function

For the purposes of this review, we will define complex carbohydrates as polymers composed of such monosaccharides as mannose, galactose, N-Acetyl glucosamine, N-Acetyl galactosamine, rhamnose, fucose and sialic acid connected in highly varied linkages and anomeric configurations. These macromolecules, which contain glucose only rarely and are not used as energy sources by an organism, are found in glycoproteins, glycolipids and in the extracellular polysaccharides of bacteria. Proteoglycans containing such polysaccharides as heparin, chondroitin and dermatan covalently linked to protein could also be considered as glycoproteins but the glycosaminoglycans, including hyaluronic acid will not be discussed in detail in this review. This exclusion along with that of starch, glycogen, chitin and plant cell wall polysaccharides does not result from any lack of their biological importance but rather from the necessity to place some reasonable boundaries on this enormously varied subject.

The complex carbohydrate structures in glycoproteins, glycolipids and bacterial polysaccharides have generally distinct features and each class can be separately defined. On the other hand, they have many similarities and there is considerable biological evidence, such as immunological or lectin receptor activity which interrelates them. Typical oligosaccharide structures may have from two to 15 sugar residues in a non-repeating sequence. While the repetition of a common structural subunit is usually associated with the lipopolysaccharides and the

capsular polysaccharides of bacteria, this repeating structure motive has also been found in glycoproteins (cf. polylactosamines) as well as in lipids such as polyglycosyl ceramides.

These complex carbohydrates have highly varied structures and they occur in rather small proportions in the cells of higher organisms. It is thought that they do not have a storage function in energy metabolism as do the glucose polymers, starch and glycogen. The complex structures, whose function appears to be in control of cell metabolism, show remarkable differences among closely related animal species. These differences stand in contrast to such proteins as hemoglobins and cytochromes, extensive studies of whose amino acid sequences have revealed surprising homologies among distantly related organisms. The carbohydrate structures of glycoproteins show substantial differences among closely related species. For example, studies of the carbohydrate chains of the red blood cell membrane glycoprotein, glycophorin, in horse, sheep, cattle, mouse and man show quite different structures (1). In fact substantial genetically determined polymorphism in the carbohydrate structures of glycoproteins within a given species is well recognized, perhaps the best known example being case of blood group activity which is determined by specific complex oligosaccharide antigens.

Although the biosynthesis of complex carbohydrates is not understood in complete detail, it is clear that higher plants and animals expend substantial metabolic energy in both the anabolism and catabolism of these molecules. The biosynthesis of the carbohydrate chain of asparagine N-linked glycoproteins, which has been intensively studied in recent years has been found to be quite complicated with buildup of a tetradecasaccharide chain on a lipid donor followed by transfer to sites on the protein. This transfer is followed rapidly by cleavage of the tetradecasaccharide by a series of specific glycosidases and, in some glycoproteins, by subsequent resynthesis of a complex antenna or hybrid carbohydrate chain (2). A biosynthetic scheme of this complexity suggests that there must be some rather important function for the complex asparagine-linked oligosaccharides of the glycoproteins of eucaryotes.

Many studies have focused on the question of the function of complex carbohydrates yielding a diverse array of apparently unrelated functions. Specific oligosaccharide sequences in heparin are active in angiogenesis while other sequences are active in control of blood clotting. Complex carbohydrates have been shown to possess various cell surface receptor functions which include the activity of ganglioside GM1 as the receptor for cholera toxin. The nonreducing terminal galactose residue in serum glycoproteins controls the clearance of glycoproteins from circulation by receptor mediated endocytosis (3). Asparagine N-linked glycosylation of newly synthesized glycoproteins in the golgi apparatus is known to control intracellular migration to lysozomes (4). Antifreeze glycoprotein, a major serum glycoprotein of arctic fish results in the non-colligative lowering of the freezing point of the fish blood (5). The results from a new technique of blotting bacteria onto thin layer chromatograms of glycolipids isolated from the mucosal membranes implies that these glycolipids may serve as natural receptors for bacteria and thus play a role in their pathogenicity (6). Blood group antigens and a number of closely related tumor and differentiation antigens are known to be complex carbohydrate structures.

While it is possible to list numerous functions for complex carbohydrates, what is lacking in the catalog given above is some unifying hypothesis concerning function. Apparently we have overlooked some underlying principle which might contribute substantially to our understanding of such important biological problems as growth, differentiation, immunology, cancer and the organization of multicellular organisms in general. Perhaps the connection between cellular organization and the function of complex carbohydrates in higher animals and plants could be illuminated by an analogy to football uniforms. The brightly colored and distinctive uniform is detrimental to the function of an individual player; it would obviously be easier for a playerto take the ball to the goal without it. But the uniforms are essential to the function of the football team and the complex rules governing the game as a whole. Therefore hypotheses which assign to the carbohydrate of a glycoprotein a function in stabilizing a protein against denaturation or in protecting the cell surface against proteolysis might be compared with the hypothesis that the function of a football uniform with colorful insignia and numerals is just to keep the player warm. Biochemists have been taught to concentrate our attention on func-

tion at the cellular level and our failure to uncover the underlying principle which will unify our understanding the function of complex carbohydrates results from the difficulty in focusing on the organization of the whole organism.

In the absence of any clear synthetic idea about the biological function of complex carbohydrates, one might learn from history that the study of structure and conformation might be the source of valuable evidence concerning function. Thus there has been an increase in research activity in the area with a resulting growth in the precision of our knowledge about the structure and to a lesser extent about the three dimensional conformation of complex carbohydrates. High resolution NMR has made major contributions to this recent progress.

II. The Determination of Covalent Structure

The problem of structure determination in complex carbohydrates differs substantially from that in either proteins or nucleic acids. Although the structural subunit is generally limited to at most 15 residues and may be composed of no more than five or six different monosaccharides, the residues are interconnected by many different glycosidic linkages, eg. (1→2), (1→3), (1→4), (1→6) etc., with variable anomeric configuration. Since there are multiple substitution points, branching is not only possible but is in fact quite common. Determination of the structure from genetic information (DNA sequence) is impossible since the structure of a complex carbohydrate, unlike that of a protein or nucleic acid, is not under the control of a single gene; it is controlled by the interaction of a collection of interacting gene products (glycosyl transferases) whose activities depend not only the rate of their synthesis but also on other environmental factors.

The problem of structure determination of complex oligosaccharides has been attacked by many techniques in past years and the present state of the art continues to be confused by the lack of any single methodology which can be called upon for a new structural problem. Unlike proteins and nucleic acids for which well formulated strategies are available for sequencing, a plethora of competing techniques exist for complex oligosaccharides, none of which is truly adequate. Among the classical methods, methylation analysis involves tedious chemistry of limited reliability and its sensitivity (10 to 100 nanomoles) is unimpressive by modern biochemical standards. Furthermore, methylation analysis yields information only on positions of substitution and not the anomeric configuration or sequence of an oligosaccharide. Periodate oxidation of vicinal diols is still widely used in spite of difficulties in the interpretation of the data. Although enzymatic degradation would appear to be a valuable method, the battery of known exoglycosidases is quite small and the routine use of endoglycosidases, the "restriction enzymes" of complex carbohydrate chemistry, has been possible only in very recent years. The availability of endoglycosidases whose specificity is well documented remains very limited. Although oligosaccharide mass spectrometry, especially FAB mass spec, seems very attractive on account of its speed and sensitivity, no method now known can give more than limited partial structural information.

Of all the modern structural methods for complex carbohydrates, high field proton NMR yields the most complete and detailed structural information. It is the only method which can, in principle, give an *ab initio* structure without resort to any other method. In practice, a complete structure determination by NMR is rarely the best approach to a completely new complex carbohydrate structure and generally other methods are used in conjunction with NMR. The major weakness of NMR spectroscopy as a method for structure determination is its poor sensitivity (25 nanomoles to 10 micromoles). On the other hand, since the experiment is non-destructive, it should always be considered first after the isolation of a suitable sample. A simple proton NMR spectrum is a modest experimental effort and will give immediate information on the purity of the sample and perhaps some general information on the structure. In the most favorable cases, the structure can be completely determined by this simple experiment and in the worst case only time is lost since the sample can generally be recovered for subsequent analysis by methylation, enzymatic degradation, periodate oxidation, or mass spectrometry all of which are destructive. Perhaps the greatest utility of the NMR method will be in providing accurately known structures for calibration of other more sensitive methods such as mono-

clonal antibody recognition, enzymatic degradation, mass spectrometry or some as yet undiscovered technique.

III. The "Structural Reporter" Method in ^1H NMR

The proton NMR spectrum of a typical complex oligosaccharide in D_2O solution shows some isolated resonances which have been called "structural reporter resonances" in the pioneering work of Vliegenthart and coworkers who have assembled a large body of proton spectra relevant to glycoprotein structure determination (7). As an illustration of the method, the spectrum of the hexasaccharide, LND-1 from human milk (structure **1**) is shown in Figure 1.

fuc α-(1→2)↘
gal β-(1→3) glcNAc β-(1→3)gal β-(1→4)glc
fuc α-(1→4)↗

1

The resonances of the anomeric protons, those of methyl groups and of those of various protons with distinctive chemical shifts have been correlated with known structures to yield a powerful method for use in conjunction with classical structural methods which were outlined above. Since the method of Vliegenthart and coworkers is so simple and rapid, it should always be considered as the first step in the identification and structure determination of an unknown complex carbohydrate. A water soluble sample, such as a glycopeptide, oligosaccharide or oligosaccharide alditol, (100 nanomoles to 10 micromoles) is exchanged with D_2O and freeze dried in the 5 mm NMR tube. High purity D_2O is added along with a micromole of acetone as internal reference and a simple 1-Dimensional proton NMR spectrum is recorded at a field of 300 to 500 MHz. Chemical shifts are reported in ppm below DSS using 2.225 ppm for acetone as an internal reference (see Figure 1). While this chemical shift convention is not completely accurate, it has proved to be very reproducible among different laboratories facilitating exact comparison of chemical shifts. It is often possible to assign the structure of an oligosaccharide directly from the spectrum simply by comparison to published data on similar or identical oligosaccharides. Milk oligosaccharides similar to LND-1 have been studied by Dua and Bush (8) and by Dua *et al.* (9). In a typical spectrum (Figure 1) most of the resonances crowded into the 3.5 to 4.0 ppm region are those of methine protons with very similar chemical shifts. The "structural reporter resonances" in Figure 1 include those of the two overlapping methyl groups of the 6-deoxy sugars, fucose, in the 1.27 ppm region. Resonances assigned to the amide methyl groups of acetamido sugars are near 2.0 ppm and the resonance of H2 of β-glucose is at 3.28 ppm. The chemical shifts of some of these resonances may be distinctive of the position of the glycosidic linkage. Most of the structural reporter resonances appear in the low field region between 5.3 and 4.0 ppm as shown in the expanded plot of Figure 1 b. The resonances at lowest field are those of α-anomeric protons exemplified here by the two fucose anomeric proton signals at 5.15 and 5.02 ppm. The resonance of reducing terminal glucose α-anomer is at 5.23 ppm. The H5 resonance of fucose, which is extremely sensitive to linkage, is found at 4.88 ppm for the 4-linked residue and at 4.35 ppm for that which is 2-linked to galactose. The β-anomeric resonances are the 7 Hz doublets between 4.42 and 4.66 ppm. A characteristic sharp resonance at 4.14 ppm is equatorial H4 of gal which is substituted at C3 by glcNAc in the polylactosamine type of structure. This resonance overlaps with that of glcNAc H3 in this compound. Although the chemical shifts of a few of the protons are slightly temperature dependent, the effect is sufficiently small that the temperature can be adjusted to move the resonance of the residual HDO line so that it does not obscure any resonances. Since a large sample was used in recording the spectrum of Figure 1, the HDO resonanace at 4.8 ppm does not interfere with signals from the sample but residual HDO can pose a major problem when only a very small quantity of oligosaccharide sample is available for study.

In a pyranoside, it is found that the six-membered ring generally forms a chair of fixed conformation providing a classification of protons as axial or equatorial. Therefore the coupling patterns are characteristic of the stereochemistry of the carbohydrate. For example, if H2 is axial, as it is for the gluco and galacto stereochemistry, then a small coupling constant of 2-3 Hz is observed as a result

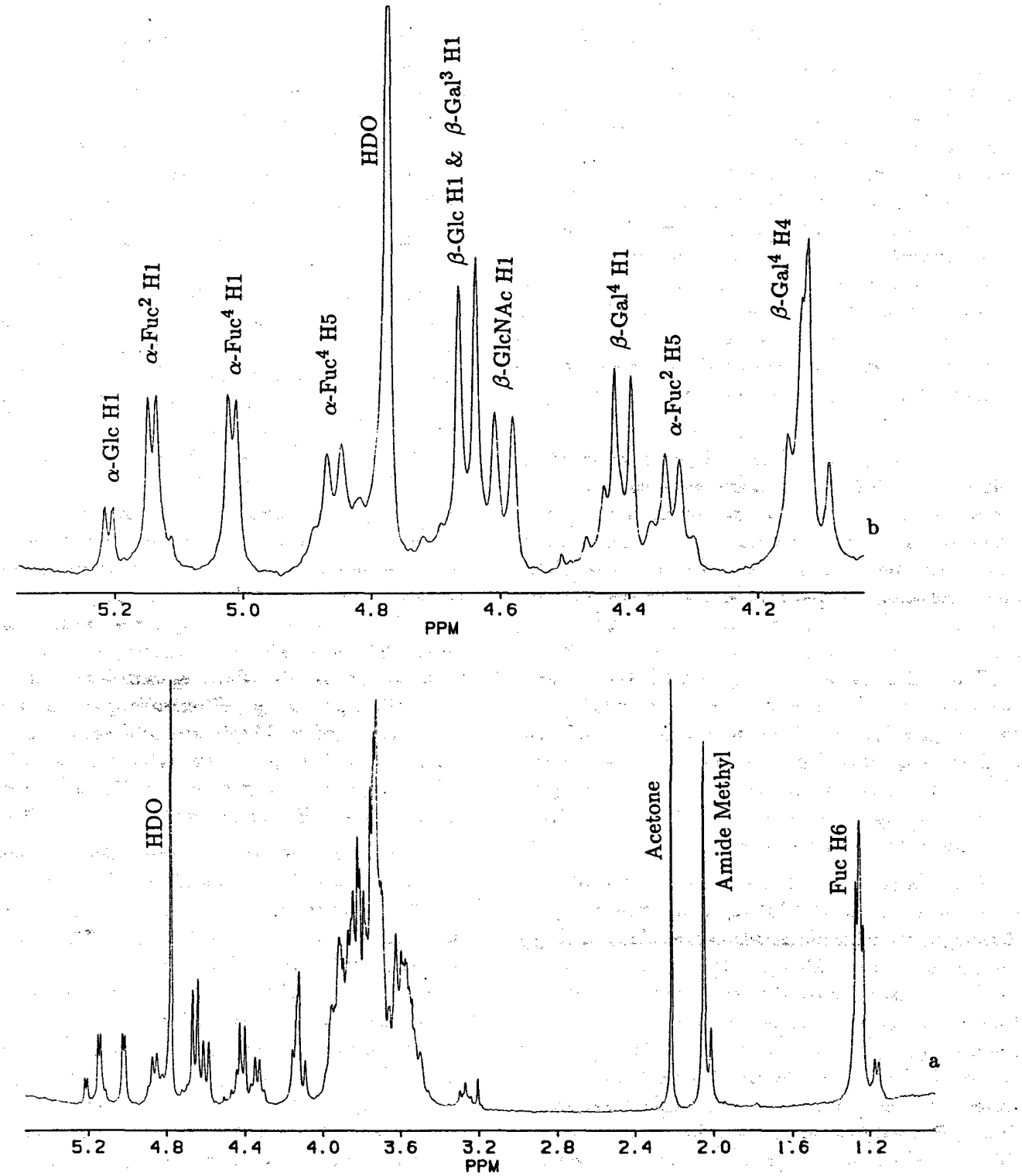


Figure 1. ^{1}H NMR spectrum of the human milk oligosaccharide, LND-I (structure **1**) recorded at 300 MHz and 24 °C. 1a is full spectrum and 1b is an expansion.

of the gauche conformation of H1 and H2 following the Karplus relation. The chemical shifts and coupling constants of the fucose α-anomeric proton resonances in Figure 1 are typical. The *trans* diaxial relationship of H1 and H2 in β-anomers of sugars with the gluco and galacto configuration leads to

larger coupling constants (7-9 Hz). Chemical shifts between 4.4 and 4.8 ppm are typical of the anomeric protons of these β-linked residues as shown in Figure 1. Assignment of the anomeric configuration in the case of mannose is made more difficult by the equatorial configuration of H2 for which the dihedral angle between H1 and H2 is small for both α- and β-anomers. Although interpretation of the chemical shift and coupling constant of the structural reporter resonances group is more subtle for mannose residues, Vliegenthart *et al.* (10) have shown that differences in structural reporter resonances can be used for structure assignment if analogies can be drawn to known structures and the interpretations are well controlled.

The resonances of equatorial protons are often shifted downfield into the position of structural reporter resonances and signals such as that of mannose H2 and galactose H4 are often distinctive and provide useful correlation with structure and linkage. Additional structural reporter resonances are observed in deoxy sugars which have rigid geminally coupled protons with distinctive chemical shifts as well as a distinctive coupling pattern. For example, the sialic acids N-acetyl neuraminic acid (NANA) and N-glycolyl neuraminic acid (NGNA) are found as 3-deoxy pyranosides and the characteristic chemical shifts of the resonances of the axial and equatorial H-3 signals have been correlated with linkage by Vliegenthart and coworkers (7,10).

A similar approach has also been applied to interpretation of the ^{1}H NMR spectra of glycolipids leading to valuable information on their carbohydrate structures. Although they are not generally soluble in water in a completely unaggregated form, Dabrowski and coworkers have shown that most glycolipids give well resolved proton NMR spectra in deuterated dimethyl sulfoxide (DMSO) (11). In the case of ionic glycolipids such as gangliosides, the addition of 2% D_2O improves solubility. The chemical shifts are reported relative to internal tetramethyl silane (TMS) and they differ somewhat from those of similar carbohydrates in D_2O (12). Although the structural reporter approach is quite useful for correlation of the carbohydrate structure of glycolipids with their ^{1}H NMR spectra, chemical shift analogies between spectra of glycolipids in DMSO and spectra of oligosaccharides and glycopeptides in D_2O cannot be effectively drawn due not only to the difference in the chemical shift reference but also to other perturbations in the chemical shifts (12). There is an extensive collection of reference spectra of glycolipids in DMSO solution to be found in the work of Dabrowski *et al.* (13) in gangliosides by Koerner and coworkers (14,15) and in blood group glycolipids by Hakomori and coworkers (16).

The method of structural reporter resonances has been applied not only to glycopeptides, oligosaccharides and glycolipids but also to bacterial polysaccharides. Unfortunately the method is less effective for this system because of the much greater structural diversity of bacterial polysaccharides. Therefore an enormous library of spectra of reference structures is necessary for drawing the close chemical shift analogies required by the structural reporter method. Some ^{1}H NMR data have been reported for pneumococcal polysaccharides (17) and for E. coli polysaccharides (18) but substantially more ^{13}C NMR data has been reported for high molecular weight polysaccharides. Gorin (19) has reviewed the general field and Jennings (20) has summarized the important work in bacterial polysaccharides. Expanded interest in the ^{1}H NMR spectroscopy of bacterial polysaccharides has been stimulated by the recent demonstration of complete proton assignments which greatly extends the capabilities of of proton NMR for the structure determination of bacterial polysaccharides (21,22).

The strength of structural reporter method is its simplicity. Since only a simple 1-Dimensional proton NMR spectrum is needed, it is the most sensitive method for very small samples. More sophisticated methods generally require more NMR observation preventing acquisition of large numbers of transients. The weakness of the method of structural reporters is that, in the absence of a sound fundamental theory of proton chemical shift, spectra of identical or closely related carbohydrate structures are required for reaching unambiguous conclusions. Therefore it is the method of choice when dealing with members of a class which has been previously studied in detail by proton NMR such as the complex N-asparagine-linked glycopeptides originally studied by Vliegenthart and coworkers (7,10). The high mannose N-linked glycopeptides have also been extensively studied not only by Vliegenthart and coworkers (10) but also by Carver (23) and by Atkinson (24). Glycopeptides of the mucin type

having the galNAc-O- glycosidic linkage to serine or threonine are usually isolated by alkaline borohydride degradation which produces the oligosaccharide alditols which have also been extensively studied by proton NMR (10,25). The oligosaccharides of human milk, which are structurally related to glycopeptides, have been studied by Dua *et al.* (8,9).

IV. Structure Determination by Complete Assignment of the ^{1}H NMR Spectrum

For interpretation of NMR spectra of oligosaccharides which are not identical to or closely related to known data, complete assignment of the methine resonances in the poorly resolved group of signals in the 3.5 to 4.0 ppm region adds greatly to the structural information over that provided by the method of structural reporter resonances. It has been shown that complete proton assignments in oligosaccharides in a size range up to 10 or 15 residues is in fact generally practical with sufficient observation time on 300-500 MHz NMR instruments (25,26). Composed mainly of linear chains of coupled spins, carbohydrates are especially suited to spin correlation methods such as difference decoupling or 2-dimensional spin correlation (COSY) and related techniques for identification of all the protons of a given sugar residue. The general approach is to assign an isolated structural reporter resonance, often an anomeric proton, then to correlate spins in a stepwise manner around the spin system of the ring. For the six-membered pyranoside ring, the stereochemistry of the sugar can be identified from the values of the vicinal coupling constants. Although spin correlation can be done by 1-Dimensional difference decoupling if only a few spin assignments are needed and sample quantity is very limited, difference decoupling is a difficult experiment requiring careful work with a very stable spectrometer (9). In most instances, 2-dimensional methods are preferred due to their more efficient use of the spectrometer resources for simultaneous determination of a large number of spin correlations.

Most of the 2-D NMR methods used in the structural analysis of complex oligosaccharides differ only in a few details from those used in the complete proton assignments of proteins. Since the values of the coupling constants are related to the stereochemistry of the pyranoside ring, it is important to observe the individual components of the multiplets in the crosspeaks for carbohydrates. Thus, special attention should always be paid to the problem of adequate digital resolution in 2-dimensional spectra of oligosaccharides. The difficulty of meeting this requirement is alleviated by the somewhat reduced chemical shift dispersion of carbohydrates compared to proteins. Sugars have no aromatic protons and the resonances of exchangeable amide protons are not generally observed in D_2O solution. The highest field resonance is usually that of the methyl group of a 6-deoxy hexose near 1.2 ppm and lowest field anomeric resonance in D_2O is near 5.4 ppm. The rotational correlation times of most glycopeptides, oligosaccharides and glycolipids are such that 1 Hz splittings are resolved. For polysaccharides including those with high molecular weight, segmental motion contributes to well resolved lines especially if the spectra are measured at elevated temperature. As in the case of proteins, the first reported spin correlation experiments were magnitude COSY but the advantages of phase sensitive experiments with pure absorptive lineshapes have been amply demonstrated in protein chemistry. Since most current NMR spectrometers are capable of phased COSY with a double quantum filter (DQF), application of this technique is becoming more common in complex oligosaccharides. Since in the magnitude COSY experiment, one calculates the absolute value rather than attempting to phase the spectrum, the line shape is not pure absorptive and the dispersive contribution to the line shape causes excessive broadening and peak tailing. To improve this line shape, a sine bell apodization is generally used to form a pseudo echo (27). The radical resolution enhancement is achieved at the price of suppressing much of the original signal since sine apodization attenuates the early part of the FID as well as the end. If the data are encoded in such a way that the correct phases of a pure absorptive spectrum can be calculated, then good lineshape can be coupled with a data apodization which degrades the signal less severely. The two methods which are commonly used for encoding the phases, time proportional phase increment (TPPI) (28) and the method of States *et al.* (29) have been shown to be formally equivalent so the choice is dictated mainly by

instrumental effects (30). Although both methods require acquisition of a raw data set twice as large as that of a magnitude COSY, it has been shown that no data are lost in phasing and the less radical apodization leads to improved signal-to-noise ratios (30). Since there is no net magnetization transfer in COSY, the multiplet components have alternating signs in phased spectra and inadequate digital resolution can cause the overlapping multiplet components to cancel (31). The requirement for adequate digital resolution in phased COSY spectra places serious demands on the NMR computer for storage and processing of the large data matrices. Data sets of 2K × 2K points are in common use and the anticipated availability of more powerful computers capable of processing larger matrices will be a decided advantage. Since it is impossible to store and process such large data matrices on the primitive computers associated with NMR instruments more than a few years old, many workers have begun to utilize a separate general purpose computer for processing and storage after acquisition of the raw data using the computer on the NMR instrument. Although high speed data links capable of transfer of 5 to 10 megabytes of data in a reasonable time period are required, the advantage in computer speed, data storage convenience and flexibility of data processing gained by using a generic multiuser computer of the current generation is often worth this effort. Software for 2-dimensional NMR data processing on such general purpose machines as the VAX and Sun computers is available from several commercial sources.

In Figure 2 is shown an expanded section of the phased COSY spectrum of the milk hexasaccharide LND-1, (structure **1**) whose 1-D spectrum was discussed above (Figure 1). Both negative and positive contour levels are plotted and the resulting multiplet shapes are useful for the assignment of overlapping crosspeaks. In Figure 2, the crosspeaks between the α-anomeric resonances and their correlated H2 are in the rows between 5.0 and 5.3 ppm. The well resolved 3 Hz gauche couplings which appear as anti-phase multiplet components in Figure 2, cannot be generally discerned in magnitude COSY spectra. The larger coupling constants of the diaxial H1-H2 of the β-anomeric proton resonances are seen in the rows between 4.42 and 4.66 ppm with the corresponding H2 resonances on the columns between 3.2 and 3.85 ppm. Assignment of the two overlapping β-anomeric resonances at 4.66 ppm in Figure 1 is readily made in the COSY as the resonances of H2 of β-gal[3] and β-glc are well separated at 3.61 and 3.28 ppm respectively. The well resolved multiplet structure and the alternating signs of the antiphase crosspeaks are especially useful in resolving the crosspeak at row 4.14 ppm and column 3.72 ppm which contains superimposed crosspeaks between gal[4] H4-H3 and between glcNAc H3-H4. The multiplet shape is characteristic for pyranosides since the size of the coupling constants is determined by the stereochemistry (*trans* or *gauche*) of the protons which are coupled.

Since, when the phase of the crosspeaks in a COSY spectrum are chosen to be absorptive, the diagonal peaks are dispersive, the diagonal can interfere with crosspeaks between resonances with similar chemical shifts which are close to the diagonal. To better visualize these crosspeaks, a double quantum filter (DQF) is often incorporated in the pulse sequence as it was in the DQF-COSY spectrum of Figure 2. This sequence preferentially attenuates the single quantum resonances of the diagonal with respect to the crosspeaks. In order to better visualize peaks with small coupling constants (0.5 to 1 Hz) such as those between galactose H4 and H5, a delayed COSY will emphasize the small couplings at the expense of some signal loss during the delay.

In spectra such as those of Figure 1 and 2 in which many resonances are crowded into the 3.5 to 4.0 ppm region, one might anticipate that strong coupling could introduce some difficulties in the scheme outlined for the assignment of the spins of an individual ring by COSY even with 500 MHz spectrometers. Strong coupling, which arises when two coupled protons have similar chemical shifts, leads both to some distortion of the expected multiplet shape and to COSY cross peaks which lie close to the diagonal. The former effect may interfere with determination of the sugar stereochemistry and the latter interferes with tracing the chain of spins within the sugar residue. If multiplet distortion is not too severe it can be accurately interpreted by spin simulation but additional experimental methods are needed to complete the tracing of the spin connectivity. In these cases spin relay experiments or isotropic mixing techniques have been shown to be especially valuable in assignments in

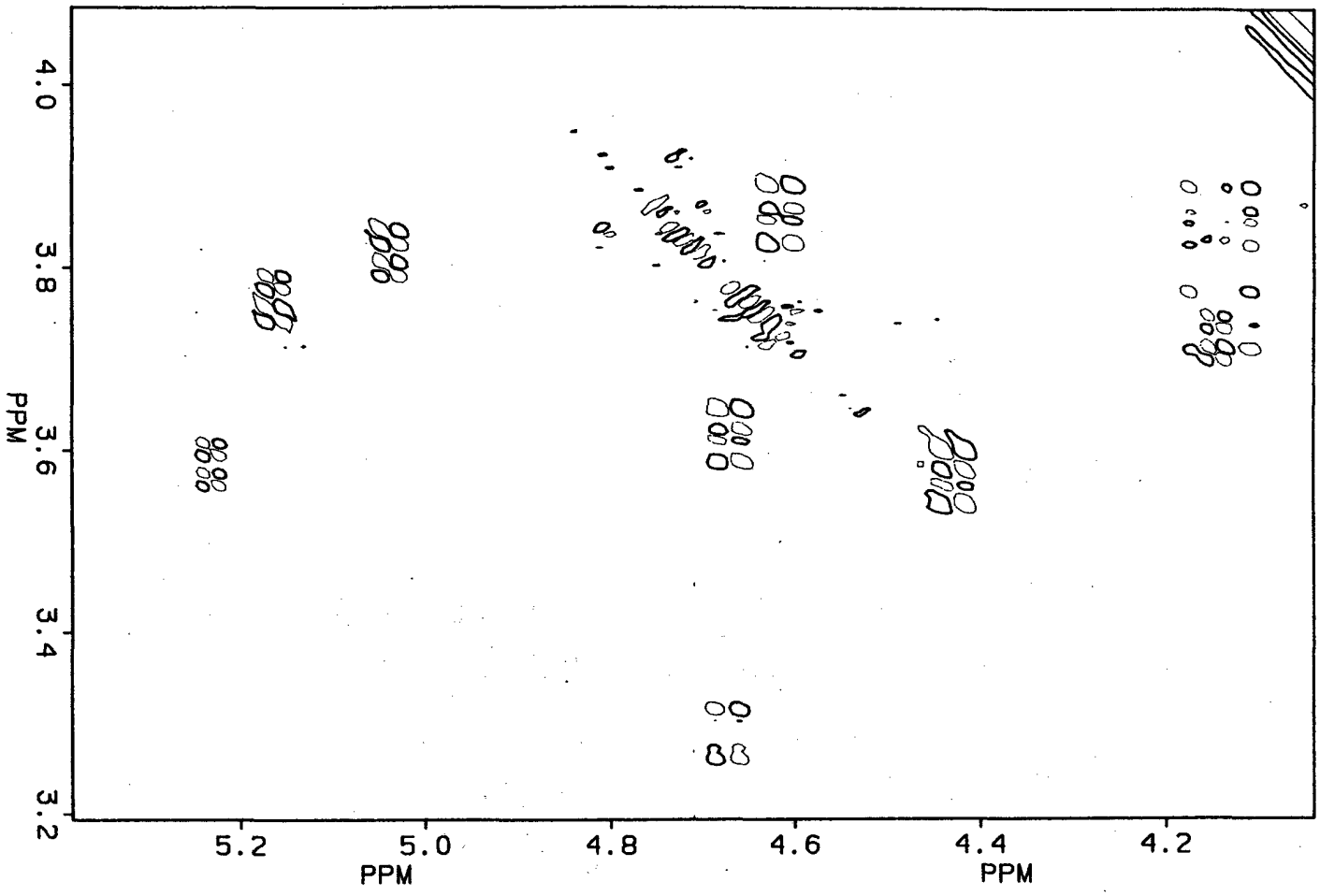
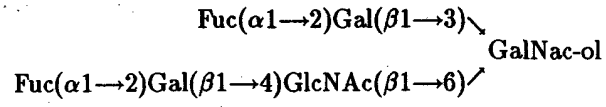


Figure 2. Expanded portion of the phase sensitive double quantum filtered COSY spectrum (300 MHz) of LND-I (structure **1**). Positive and negative contours are indicated by lines of different thickness.

oligosaccharides (12,32a). The most useful method for relay of coherence along the chain of spins is the isotropic mixing experiment in which net magnetization is transferred under spin locking. This experiment which is known as HOHAHA (homonuclear Hartmann-Hahn) or also as TOCSY, can be done either in 1-Dimensional difference mode or as a 2-dimensional experiment with phase encoding (33). In this technique, crosspeaks are observed between resonances within a single spin system (eg. pyranoside ring) which share common coupling partners. A complete spin system can thus be identified if there is at least one resonance in the spin system, such as the anomeric proton, which is well isolated and which has a reasonably large coupling to its neighboring spin. Figure 3 shows an example of the phased 2-D HOHAHA spectrum of a hexasaccharide alditol (12), structure **2**.



2

The spin systems of the glcNAc residue (H3 and H4) and of both fucosyl residues (H2 and H3) in the oligosaccharide show strong and intermediate coupled peaks which are difficult to interpret in the COSY spectrum because of their proximity to the diagonal peaks. But in Figure 3, the rows containing the anomeric proton resonances of fucose and GlcNAc show crosspeaks with the H2, H3 and H4 which are far from diagonal. Since fucose has the galacto configuration, the H4-H5 coupling is small and the spin propagates only as far as H4 in this ex-

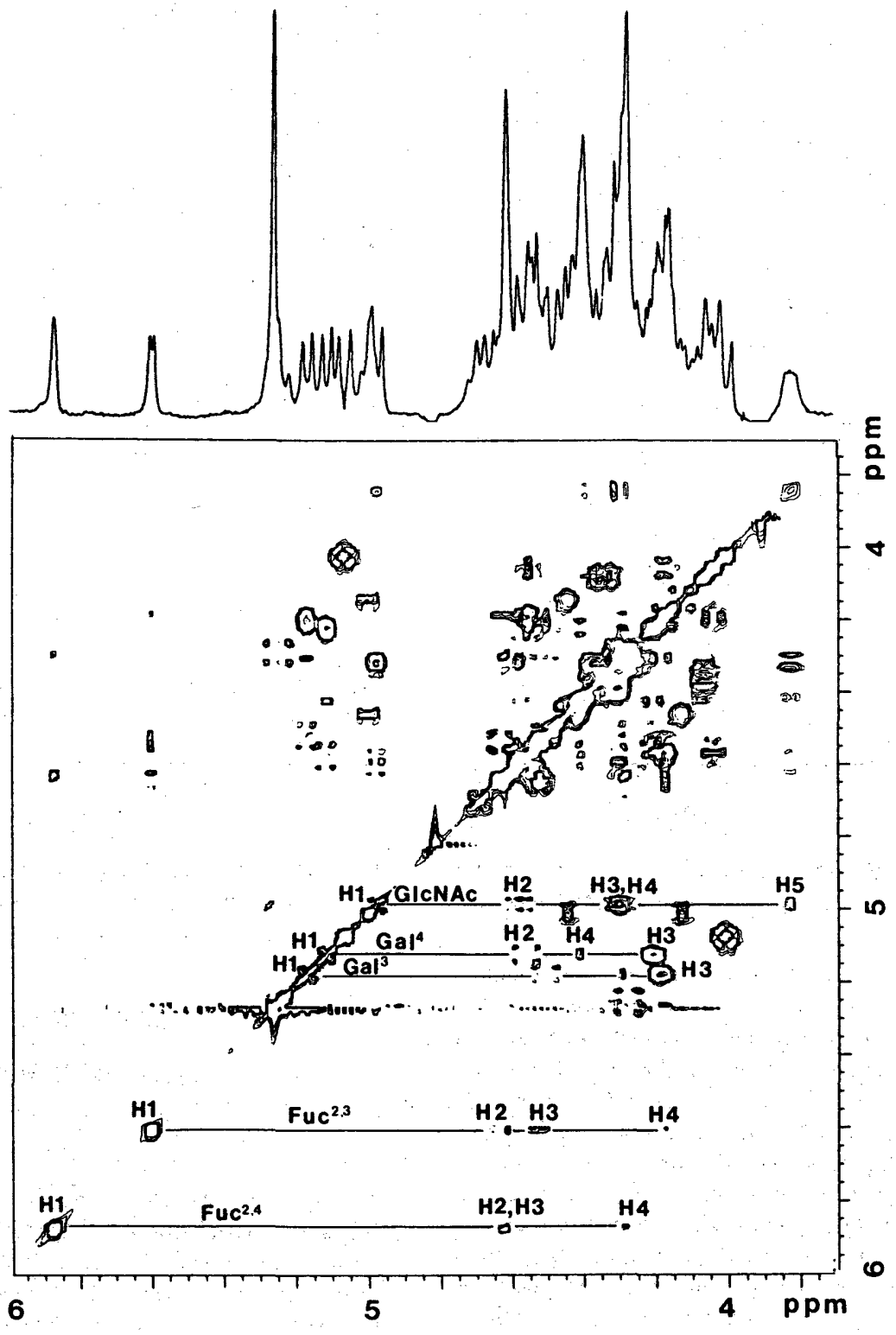


Figure 3. Phase sensitive 2-dimensional Homonuclear Hartmann-Hahn (HOHAHA) spectrum (300 MHz) of the H hexasaccharide, structure **3** in pyridine solution from ref. (12). Chemical shifts are referenced to internal TMS.

periment. Since there is net magnetization transfer in the HOHAHA experiment, the spectrum can be phased such that all peaks have a positive sign and the effects of limited digital resolution are less serious than in phased COSY. The technical implementation of the HOHAHA sequence requires spin locking at moderate power which is not always readily available on older spectrometers. Bax and coworkers (34) have recommended the use of the transmitter "low-power" mode augmented by an outboard linear amplifier while Rance (35) has demonstrated schemes in which the decoupler can be used for spin locking.

While a complete assignment of the resonances of each spin system to individual pyranoside rings can usually be achieved by the DQF COSY method augmented perhaps by HOHAHA, other methods have been found useful for the solution of specific problems in the assignment. The triple quantum filter COSY (TQF COSY) selects only those spin systems for which there are three mutually coupled spins. In the absence of strong coupling, the pyranosides commonly found in glycoproteins have only a few systems with three mutually coupled spins. One such system in hexapyranosides is H5,H6 and H6' which often presents some difficulty in assignment if the ring contains equatorial protons with small coupling constants which prevent transfer of coherence from the anomeric proton to H5 and H6. Dwek and coworkers have suggested this experiment for resolving the difficult and important problem of the assignment of H5,H6, H6' in oligomannosides in the N-linked glycopeptides (36). They have observed that if any two spins in the three mutually coupled proton system are chemically equivalent (i.e. they have the same chemical shift) no TQF crosspeak will be seen. Further difficulties in this experiment result from strong coupling among two protons of the ring which may cause an artifactual crosspeak. Although this is a powerful experiment, the results must be carefully interpreted.

The presence of acetamido sugars in a structure is indicated by the amide methyl resonance near 2.0 ppm but this signal does not generally indicate to which residue the amide group belongs. The residue containing the amide group can be uniquely identified by observation of the amide NH resonance which is coupled to the main ring spin system at H2 in the common 2-acetamido sugars galNAc and glcNAc. Since in aqueous solution the amide proton exchanges with solvent, it is generally necessary to identify the amide proton resonance with experiments either in H_2O solution or in an H_2O exchanged sample in non-protic solvents which do not promote exchange. For the case of experiments in H_2O solution, some water suppression method for improving the dynamic range is usually necessary in order to get reasonable sensitivity. Selective pulses in 1-D difference decoupling spectroscopy are simple and give a sensitivity equivalent to that of the D_2O sample (25). The usual practice in 2-D spectroscopy (COSY) involves water suppression with the decoupler channel which, because of dynamic range problems on most spectrometers, generally leads to slightly degraded sensitivity. The capabilities of more modern spectrometers in generating selective pulses may be exploited to improve the sensitivity of 2-D spectroscopy in H_2O (37).

It is generally thought that the method of spin correlation (eg. COSY) which is so crucial to the successful application of the methods outlined above, cannot be applied to globular proteins over about 20,000 Daltons as a result of the slow rotational correlation times which shorten T_2 beyond the point at which coupling correlations of 7 to 10 Hz can be detected. One should therefore ask whether application of these methods to polysaccharides might encounter a similar limitation. Especially if the polysaccharide adopts an extended rod-like conformation, quite low molecular weight polymers might exhibit long rotational correlation times preventing the observation of spin correlation about the pyranoside ring. Currently existing NMR data on polysaccharides suggest that most of them are sufficiently flexible that this is not the case. While the extent of segmental motion depends on the specific glycosidic linkages of the polymer, the effective rotational correlation times of most polysaccharides are such that couplings of 1 to 3 Hz are resolved. For more rigid polysaccharides, raising the probe temperature to 75°C may lengthen T_2 and improve the resolution.

Once each sugar residue has been identified and its anomeric configuration determined, completion of the structure determination requires that the sequence of the sugars and the linkage positions be determined. The vicinal proton coupling, which is so useful in the proton assignments of the individ-

ual pyranoside rings, is less valuable for relating the rings to each other. Although the four-bond proton couplings across the glycosidic linkage are large enough to be useful in some aromatic glycosides, they are only about 0.2 Hz in disaccharides (38). Unfortunately such small coupling correlations are difficult to detect in larger oligosaccharides and in polysaccharides because of the longer rotational correlation times and short T_2 associated with these structures. Proton nuclear Overhauser enhancement (NOE) which depends on proton proximity is a more effective method for determination of the sequence of the sugar residues and in some cases their linkage positions as well. Although NOE is most often observed between the anomeric proton and the proton connected to the carbon atom of the linkage (the aglycone proton), the effect depends on the conformation of the glycosidic linkage and some care must be used in deduction of the oligosaccharide linkage directly from proton NOE data. Several instances have been reported in blood group oligosaccharides and bacterial polysaccharides in which the largest NOE between the anomeric proton and that of the aglycone sugar is not to the aglycone proton itself, but to one adjacent to it. For the galNAc α-(1→3) gal linkage in the blood group A oligosaccharides, the NOE between galNAc H1 and gal H4 is much greater than to the aglycone proton, gal H3 (25). This effect has been explained on the basis of the proposed conformation of the non-reducing terminal trisaccharide fragment of the blood group A oligosaccharides (39). A similar effect observed in the gal α-(1→3) gal linkage of blood group B glycolipids presumably arises from a similar conformation (13). In a bacterial polysaccharide from *Streptococcus sanguis* 34 containing the sequence galNAc α-(1→3)rhamnose, the anomeric proton of galNAc shows the major NOE to the resonance of rha H2 rather than to that of H3 (22). Although it is an equatorial proton adjacent to the linkage position which shows the anomalously large NOE in each of these examples, no simple rule can be extrapolated from these observations since the man α-(1→3)man sequence in N-asparagine glycopeptides does not show this effect (40). The NOE results depend on the conformation of the glycosidic linkage, a topic which is not yet fully understood. Therefore, in any determination of the structure of a completely new and unknown oligosaccharide system, NOE data will reliably indicate the carbohydrate sequence but there may be some ambiguities in linkage position. Thus the NOE experiment may require some confirming data such as methylation analysis, enzymatic degradation, periodate oxidation or additional NMR techniques, such as long range ^{13}C-^{1}H coupling correlation which will be discussed below.

NOE depends not only on proximities of protons but also on the rotational correlation time of the molecule, being positive for small molecules (hundreds of Daltons) and negative for large ones (thousands of Daltons). Although this effect can present some technical problems in the measurement of NOE in oligosaccharides, the extent of segmental motion in high molecular weight polysaccharides is usually such that substantial negative NOE are observed. By regulation of the probe temperature, conditions can generally be found for most polysaccharides such that NOE can be measured either by 1-D difference methods or by NOESY using procedures similar to those used for small proteins. When displayed in the phase sensitive mode, all the 2-D NOESY cross peaks have the same sign. The phases can be encoded either by the TPPI method (28) or by the method of States *et al.* (29) as described above for COSY.

We have alluded to the potential problem in measuring NOE for molecules of intermediate size. The rotational correlation times of oligosaccharides having 4 to 6 residues are such that these compounds generally show no NOE when studied at field strengths of 300 to 500 MHz in D_2O solution. While substantial negative NOE is generally observed for glycolipids in DMSO (41), NOE measurements on oligosaccharides and small glycopeptides may require special measures. In our work, we have found that the rotational correlation time of oligosaccharides is a strong function of temperature. Figures 4 and 5 show the 1-D difference NOE spectra for a hexasaccharide alditol (structure **2**) at two different temperatures.

Since this oligosaccharide exhibits negative NOE at 5°C and positive NOE at 70°C , we conclude that measurable effects can be measured quite generally for any size oligosaccharide by the 1-D difference method (12,42). When ω is very close to τ_c, the effects are modest (about 5 to 10 %) but can be measured with sufficient attention to control of the probe temperature and to frequency stability of the

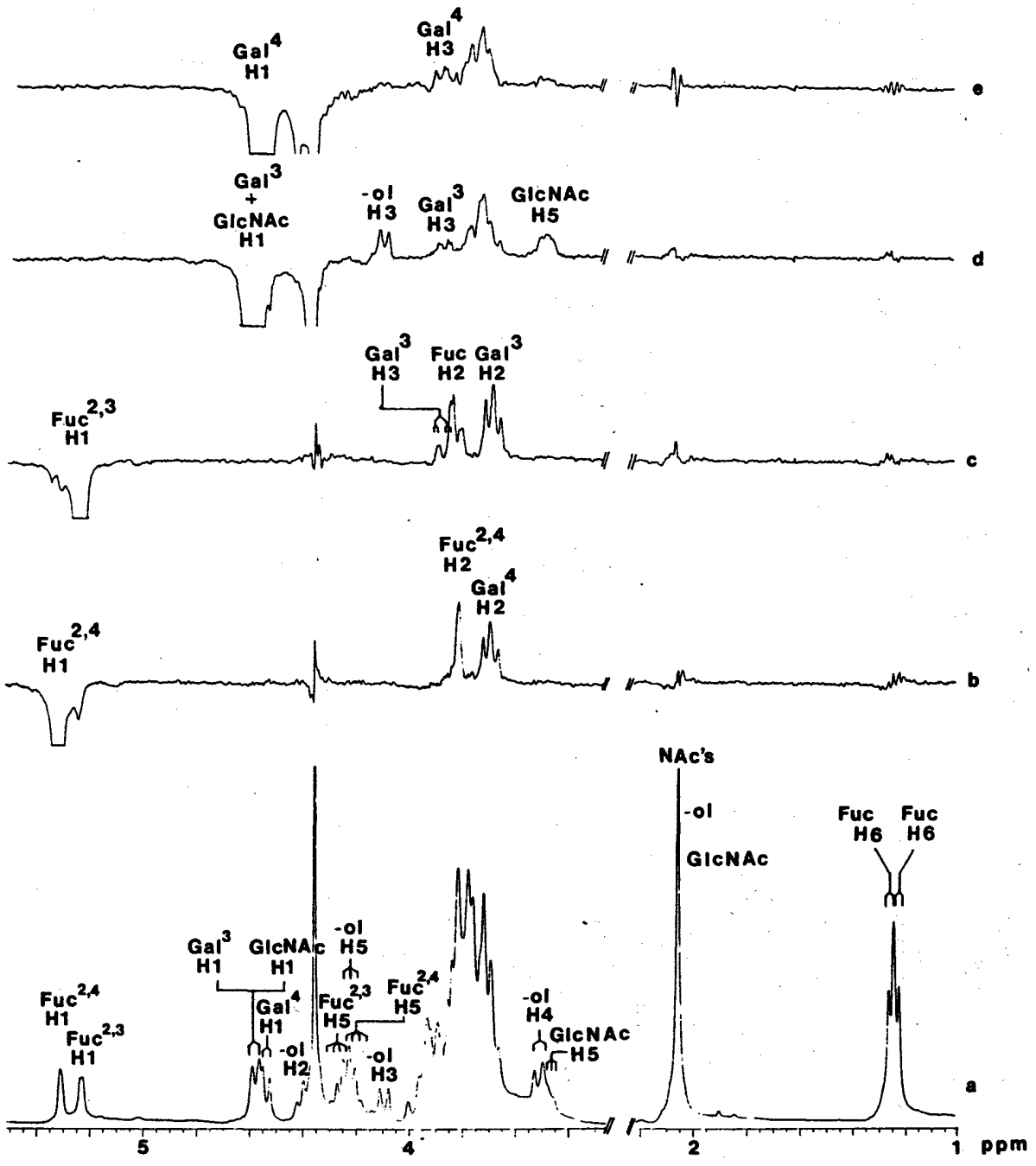


Figure 4. Difference NOE spectra (300 MHz) of the H hexasaccharide (structure 2) in D_2O at 70°C shows positive enhancements (42).

spectrometer. Two-dimensional methods (NOESY) place lesser demands on spectrometer stability and suffer less from spectral overlap and cross saturation effects for resonances with similar chemical shifts. The rotational correlation times of most polysaccharides and of decasaccharides at 5 °C are suffi-

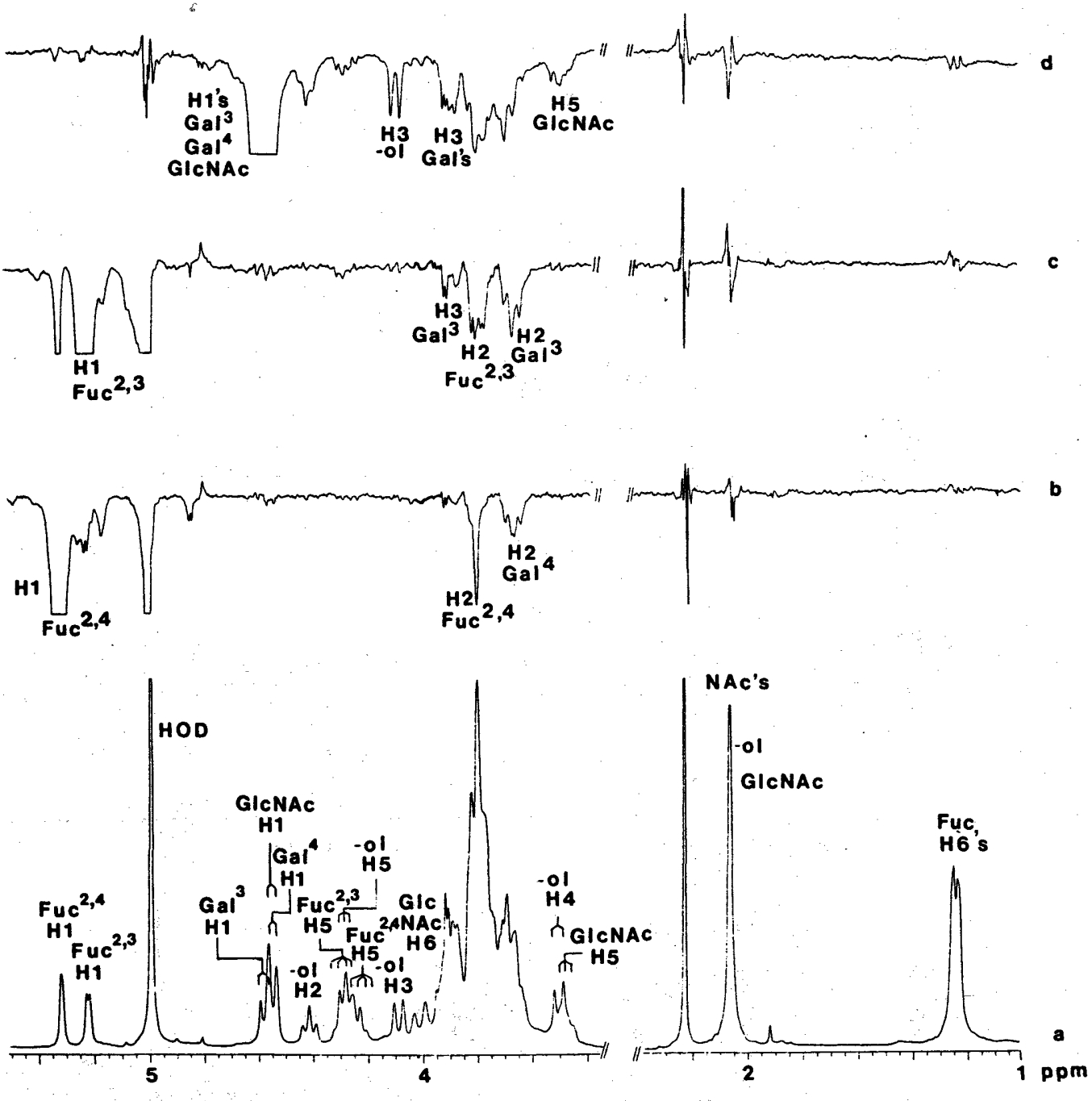


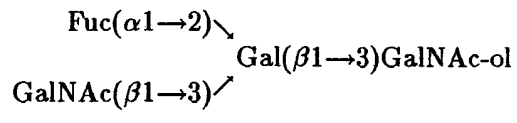
Figure 5. Difference NOE spectra (300 MHz) of the H hexasaccharide (structure 2) in D_2O at 5°C shows negative enhancements (42).

ciently long that the NOESY experiment works well but for smaller oligomers, good NOESY data require mixed DMSO-water solvents and reduced temperatures. Although it might appear that NOE measurements in the rotating frame (ROESY) might offer a solution to the experimental problem of 2-D NOE

measurement in oligosaccharides of medium size, it has been pointed out that artifactual crosspeaks can arise when rotating frame NOE occurs to protons which are coupled and whose chemical shifts differ by less than a few hundred Hz (43). This is precisely the case for most oligosaccharides in which there are invariably several coupled protons on the aglycone residue whose chemical shifts are quite similar.

In addition to the value of the NOE for structure determination, the technique has been a major source of experimental information for determining the conformation of complex carbohydrates. Since NOE depend on distances between protons, it is possible in principle to determine inter-proton distances directly from NOE data. In practice it is generally impossible to measure a sufficient number of independent NOE data to rigorously determine an oligosaccharide conformation. However, when the NOE data are combined with the geometric constraints dictated by the known covalent structure and the conformation of the pyranoside rings, a detailed conformation can often be deduced. There is a growing body of literature in which various computational methods of molecular modeling are combined with proton NOE data to deduce conformations of such complex carbohydrates as the N-asparagine glycopeptides (40,44) and blood group oligosaccharides (39,45).

Essentially all of the structural studies of complex oligosaccharides, glycopeptides and of bacterial polysaccharides by proton NMR spectroscopy have utilized D_2O as the solvent. For glycolipids, which aggregate into micelles in D_2O solution, deuterated dimethyl sulfoxide has invariably been the solvent of choice for structural studies and a substantial literature exists both for gangliosides and for neutral glycolipids. It is clear from comparison of the ^1H spectra of similar oligosaccharides in DMSO and D_2O that there are real differences in the relative proton chemical shifts in these two solvent systems which does not result simply from the difference in chemical shift referencing. In recent studies on blood group oligosaccharides which are soluble in DMSO, we have completed proton assignments of the spectra by the same methods as those used in D_2O solutions and we have extended this to the use of deuterated pyridine as the solvent (12). In Figure 6 are compared the spectra of a blood group A tetrasaccharide (structure 3) in D_2O and in pyridine showing even greater differences in the chemical shifts in this solvent.



3

While the anomeric proton of fucose resonates downfield from that of galNAc in the aqueous system, the galNAc anomeric is the lowest field resonance in the spectrum in pyridine solution. Also the common rule that the anomeric resonances are the most downfield is not followed for the spectrum in pyridine solution, the resonance assigned to galNAc H2 being downfield of that of the anomeric proton of β-gal. The substantial influence of the solvent on the chemical shifts can be used to advantage in structural studies on complex carbohydrates for which unfavorable signal overlapping and strong coupling of the spectra in D_2O solution causes difficulties in the tracing of the spin systems by coupling correlation and in identification of the sequence and carbohydrate linkage by NOE (12). If a ^1H NMR study in D_2O fails to yield data adequate to fully identify a new oligosaccharide structure, the experiment can be repeated in pyridine solution without compromising the sample. This solvent can be evaporated for subsequent studies in another solvent.

It might be suspected that the remarkable dependence of relative chemical shifts on solvent could arise from a difference in the oligosaccharide conformation. But at least for the blood group A and H oligosaccharides, we have evidence based on measurements of the NOE in different solvents that the conformations of these oligosaccharides are similar in water, DMSO and in pyridine (45). We propose instead that the chemical shift differences depend on different local magnetic susceptibility. This hypothesis is consistent with the observation that pyridine, with its highly anisotropic aromatic ring shows the greatest effect as a result of the "ring current" effects. Although a more detailed explanation of the solvent perturbations of the chemical shifts might provide a valuable correlation with oligosaccharide structure, we cannot yet provide any simple rationalization of these effects.

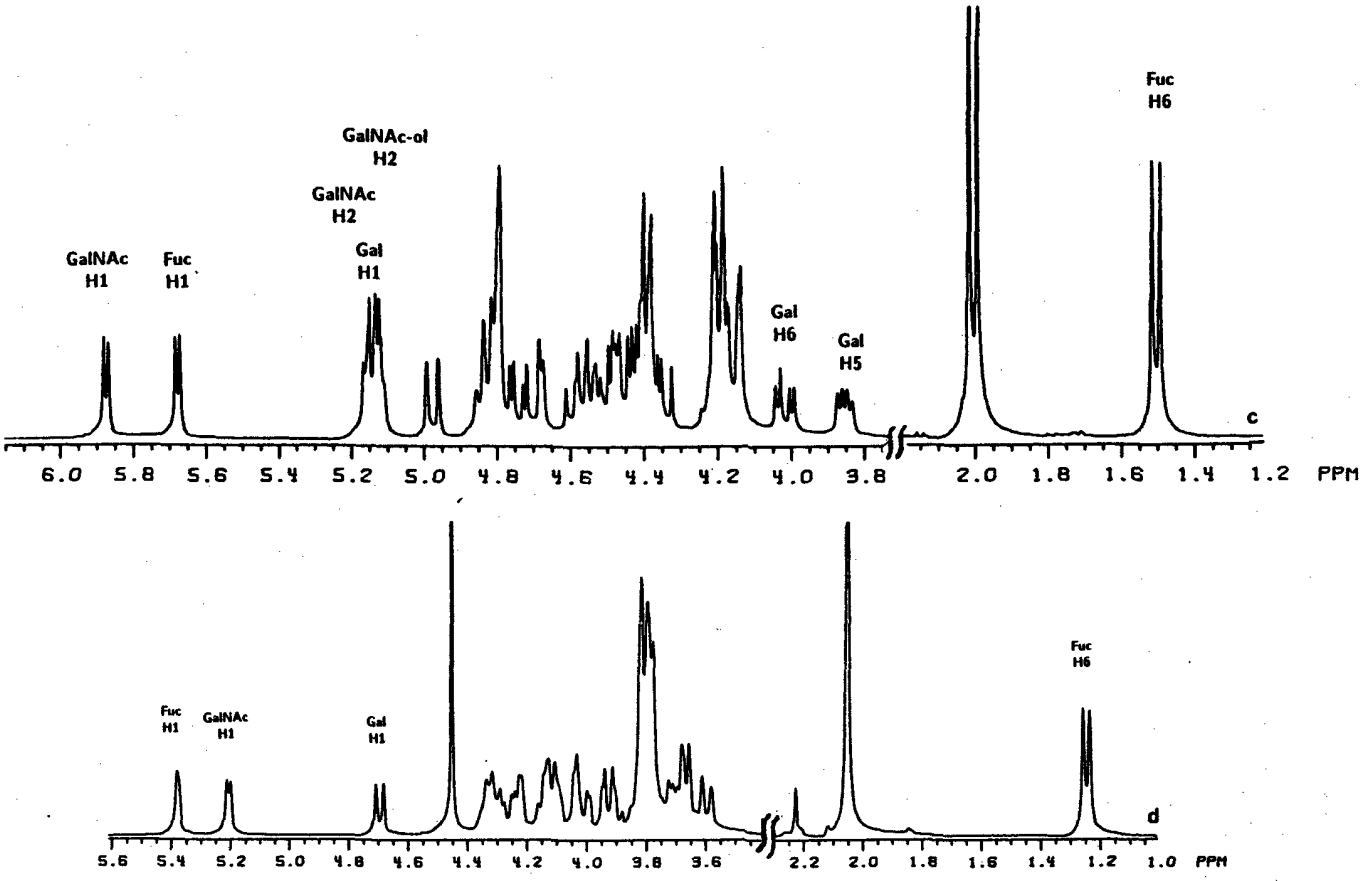


Figure 6. Spectrum of the blood group A tetrasaccharide (structure 3) at 300 MHz and 60°C in D_2O solution referenced to DSS (lower trace) and in pyridine solution referenced to TMS (upper trace) showing changes in relative chemical shifts (12).

V. Prospects for Carbon NMR

While ^1H NMR spectroscopy has been the most important source of structural information for complex carbohydrates, ^{13}C NMR spectroscopy has enormous potential for carbohydrates as a result of its chemical shift dispersion which is much greater than that of proton NMR. In contrast to the rather crowded proton spectrum, the ^{13}C NMR spectrum of a pentasaccharide has few overlapping lines even when recorded at low field in an iron magnet spectrometer at 25 MHz. In higher field superconducting spectrometers individual lines are observed for essentially all the carbons of an oligosaccharide of modest size (see Figure 7). In the proton-decoupled carbon spectrum of a milk pentasaccharide LNF-II (structure 4) measured at 75 MHz, the anomeric region between 95 and 105 ppm contains resolved signals for the α-and β-anomers of reducing terminal glucose and one line for each of the other four residues (46).

gal β-(1→3)glcNAc β-(1→3)gal β-(1→4)glc
fuc α-(1→4)↗

4

The resonances of the two methyl carbons, one from C6 of fucose and one from the amide methyl of GlcNAc are at 16.5 and 23 ppm respectively. The resonances at 80 and 83 ppm are assigned to glc C4 and to gal[4] C3, the positions of glycosylation.

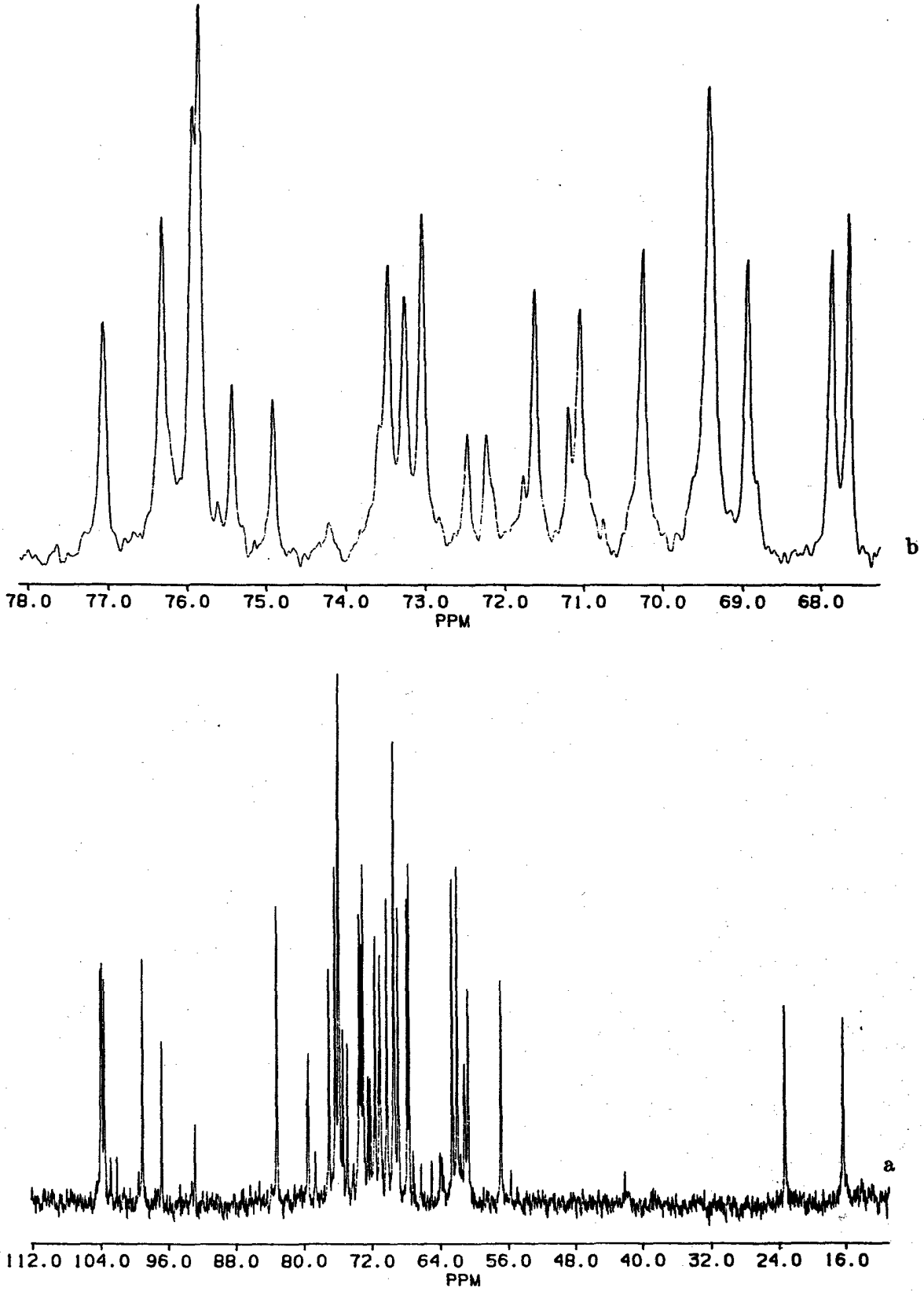


Figure 7. ^{13}C NMR spectrum (75 MHz) of the milk tetrasaccharide LNF-II (structure 4). 7a is full spectrum and 7b is an expansion.

These downfield shifts are characteristic of glycosidic substitution. The resonance of glcNAc C2 is at 57 ppm, an upfield position which is characteristic of the acetamido substituted carbons. The methylene carbons of the C6 are all resolved and located between 61 and 63 ppm. In addition to these relatively isolated and well resolved resonances, the expanded plot of Figure 7b between 66 and 78 ppm contains 19 resolved lines in a region in which 21 resonances are expected (46). In spite of the improved chemical shift dispersion of the ^{13}C spectrum of Figure 7 over that of the ^{1}H spectrum of a similar oligosaccharide (Figure 1), carbon spectroscopy has not found wide use in structure determination of glycopeptides and related oligosaccharides. One of the major barriers to wider exploitation of ^{13}C NMR spectroscopy in complex carbohydrates has been its rather poor sensitivity. Therefore, published data on ^{13}C spectroscopy of glycopeptides from higher animals, which have been available only in very limited sample sizes, have been modest. In contrast, polysaccharides such as those from yeast and bacterial cell surfaces can often be isolated in 100 mg quantities. Therefore, the availability of Fourier transform ^{13}C NMR spectrometers soon led to an extensive literature on the ^{13}C spectra of polysaccharides. In a review of this field, Gorin (19) has described the correlation of the ^{13}C spectra with structure of the polysaccharide repeating unit. For most polysaccharides, line widths of only a few Hz were observed and the ^{13}C NMR data have made significant contributions to the overall understanding of the structure of bacterial polysaccharides (20).

Although it might appear that complete assignment of all the well resolved lines of the ^{13}C spectrum of a complex carbohydrate could be easier than the assignment of a ^{1}H spectrum, with its array of overlapping multiplets, a truly reliable assignment of the carbon spectrum of an oligosaccharide presents some special problems. As a result of the low natural abundance of ^{13}C, there is no simple analog of the vicinal coupling of protons which provides a rigorous assignment of the resonances of the sugar ring. Therefore other approaches to the assignment of ^{13}C resonances have been employed in the studies on bacterial polysaccharides as well as in the limited number of studies on glycopeptides and glycolipids. Many of the assignments were made by analogies of the chemical shifts to the assigned resonances of the constituent monosaccharides and to those of simple oligosaccharides following a build-up scheme which incorporates the effect of α-and β-substituents in an empirical manner. This approach, which has been called the glycosylation shift method, has been applied in the most systematic way to the mannose oligosaccharides of the type seen in N-linked glycopeptides (47,48). This empirical method suffers from the lack of a fundamental theory of chemical shift and has no scheme to account for conformational perturbations which result in long range effects. Therefore it is not very reliable and has led to many errors in the assignments of the ^{13}C NMR spectra of carbohydrates (46). Among other techniques which have been employed to assist in assignment of the carbon resonances, specific labeling with ^{13}C is most reliable but its utility for assignments in complex carbohydrates isolated from natural sources is extremely limited (49). The "attached proton test" (APT) and related DEPT pulse sequences involving transfer of polarization from ^{1}H to ^{13}C are very useful in organic chemistry for distinguishing among methine, methylene and methyl carbons, each of which is directly coupled to a different number of protons (50). For carbohydrates, most of whose carbons are of the methine type, this experiment is less useful serving mainly for rigorous identification of C6 in hexapyranosides (46). A third method which has been used for carbon assignments in complex carbohydrates is the deuterium isotope shift (DIS) which depends on the difference between the chemical shift of carbon atoms connected to OH and those connected to OD. The differential of a few hertz in ^{13}C chemical shift between carbohydrates in D_2O and in H_2O solution can be measured either in a coaxial tube scheme (51) or with two separate carefully referenced spectra, one measured in H_2O and one measured in D_2O (47). Since the deuterium isotope shift depends on whether a ^{13}C is α-or β-to a hydroxyl group, it should be possible to distinguish not only the signals assigned to glycosidically linked carbons, but also those adjacent. Unfortunately differences in magnetic susceptibility and concentration dependence of the chemical shift result in some technical difficulties in making the highly accurate chemical shift measurements required by this experiment and the empirical interpretation requires some caution (52). The above methods have been used to derive com-

plete carbon assignments for complex carbohydrate systems including milk oligosaccharides (46), mannosidosis oligosaccharides (53) blood group oligosaccharides (54) gangliosides (55) and for glycopeptides (56). In a review Bock *et al.* (57) have compiled tables of the chemical shifts of many oligosaccharides and Dill *et al.*, (58) have reviewed the literature in glycopeptides. Although most of these resonance assignments are likely to be correct, a few of them may have been interchanged in these studies.

Several recent technical developments in ^{13}C NMR spectroscopy may alleviate the difficulties in deriving reliable ^{13}C assignments as well as contribute powerful new tools for structure determination of complex carbohydrates. First the recent progress in obtaining complete proton assignments for complex carbohydrates can be exploited in assignment of the carbon resonances by means of ^{1}H-^{13}C chemical shift correlation. Although this correlation can be done in principle by established ^{13}C detected heteronuclear COSY techniques which have been widely used in organic chemistry, the low sensitivity of this method presents a major obstacle to its application for higher molecular weight complex carbohydrates. Recent hardware developments make possible proton detected 2-D heteronuclear multiple quantum coherence spectra which gives a gain in sensitivity sufficient to make ^{13}C-^{1}H correlation spectroscopy of complex oligosaccharides a real possibility. The experiment requires a decoupler operating at the ^{13}C frequency and a dual tuned probe, preferably one with the proton detection coil inside the carbon coil for improved sensitivity (34). Through special pulse sequences one can suppress the ^{1}H signals from protons connected to ^{12}C and the experiment requires as little as a micromole of sample for a modern 500 MHz spectrometer. An extended version of this reverse detected COSY, in which the spin is relayed from ^{13}C to ^{1}H to vicinally coupled ^{1}H further extends its power in deriving complete assignment of the carbon spectrum from the assigned proton spectrum (59). Although there are few reports in which this approach has been used for assignment of a large unknown complex carbohydrate, it appears that the added resolution gained by spreading the congested proton spectrum in the carbon dimension may extend considerably the high resolution NMR methods for complicated structures (60a).

Once the carbon spectrum has been completely assigned, an unambiguous determination of the glycosidic linkage position can be obtained from the 3-bond C-H coupling correlation. Detection of long range coupling between either the anomeric carbon and the aglycone proton or between the anomeric proton and the aglycone carbon would serve to identify the linkage much more reliably than can be done by proton NOE. Although these coupling constants are known to depend on geometry, unambiguous determination of the linkage requires only that either one of these couplings be greater than about 5 Hz permitting detection in long-range ^{13}C-^{1}H correlation spectra. It has been proposed that this correlation can be observed either by the selective INEPT method with ^{13}C detection (61) or by the 2-dimensional heteronuclear multiple quantum coherence method with proton detection of the 3-bond correlation (59,62). It should be noted that the values of these long range carbon-proton coupling constants across the glycosidic linkage could be very useful in determining the conformation of the glycosidic linkage. The long range $^{3}J_{CH}$ between the anomeric proton and the aglycone carbon is related to the glycosidic dihedral angle ϕ and that between the anomeric carbon and the proton attached to the aglycone carbon is related to ψ by a Karplus relation. While the correct parameterization of this Karplus relation is not yet understood, detailed studies of these coupling constants for model systems could contribute a valuable tool for investigation of the conformation of complex oligosaccharides.

VI. NMR Spectroscopy of Other Nuclei: ^{31}P, ^{17}O, ^{15}N

Few nuclei other than ^{1}H and ^{13}C have been exploited in structural studies of complex carbohydrates. ^{17}O has very low natural abundance and since it is quadrupolar it does not give high resolution spectra. The nitrogen nucleus of acetamido sugars is a candidate for study but the natural abundance of ^{15}N is low and since no interesting structural information is expected few data have been reported. Although study by NMR of other carbohydrate substituents such as sulfate would be of great interest, detection of the sulfur nucleus by NMR is impractical and only the effect of sulfation on ^{1}H

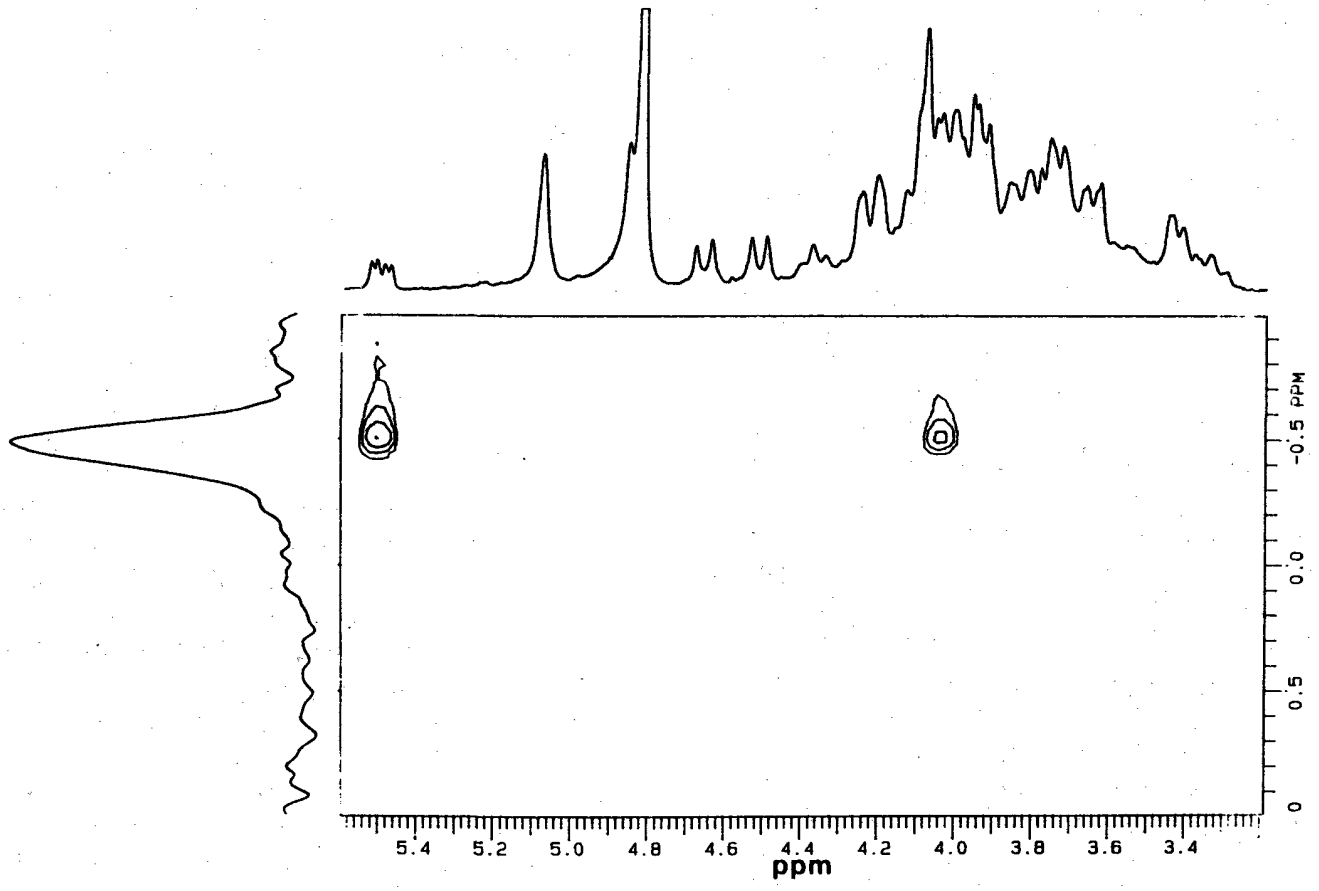


Figure 8. Two dimensional ^{31}P-^{1}H long range coupling correlated spectrum of the capsular polysaccharide from *Streptococcus sangus* 34, structure 5.

or ^{13}C chemical shift has been used. Phosphorylation of complex oligosaccharides is relatively common and the sensitivity of ^{31}P makes this nucleus particularly suitable for high resolution NMR. The chemical shift of ^{31}P in phosphates is quite sensitive to the state of ionization of the phosphate so phosphomonoesters and diesters can generally be distinguished from inorganic phosphate as a result of their different pK_a (63). Since ^{31}P in phosphate esters shows reasonably large vicinal and geminal coupling constants to protons and to carbon atoms (5-7 Hz), long range coupling correlation can be used to assign proton or carbon resonances which are coupled to ^{31}P (64). This approach can be used in the assignment of the position of phosphate ester substitution in phosphorylated carbohydrates and in bacterial polysaccharides containing internal phosphodiester linkages (65,66). Figure 8 shows the long range ^{31}P-^{1}H 2-D correlated spectrum and the ^{1}H NMR spectrum of the capsular polysaccharide of the bacterium, *Streptococcus Sanguis*

This bacterial polysaccharide is composed of a repeating linear hexasaccharide linked with phosphodiester bonds as shown in structure 5.

$$[\rightarrow PO_4^- \rightarrow 6)\ \alpha\text{-D-Gal}p\text{NAc}\ (1\rightarrow 3)\beta\text{-L-Rha}p$$
$$(1\rightarrow 4)\beta\text{-D-Glc}p\ (1\rightarrow 6)\beta\text{-D-Gal}f$$
$$(1\rightarrow 6)\beta\text{-D-Gal}p\text{NAc}\ (1\rightarrow 3)\alpha\text{-D-Gal}p\text{-(-1-}]_n$$

5

The ^{1}H spectrum was fully assigned using methods such as COSY, HOHAHA and NOE which have been described above (22). The position of the phosphodiester linkage was then determined by the vicinal ^{31}P-^{1}H coupling correlation between phosphorus nucleus and the protons of the carbohydrate at the linkage positions, α-gal H1 and α-galNAc H6.

VII. High Resolution NMR of Intact Glycoproteins

One of the major advantages of structural analysis by NMR spectroscopy, that no chemical treatment of an oligosaccharide or glycopeptide is required, not only simplifies the experimental procedure but also allows easy recovery of the sample for further study by other methods. A valuable extension of this idea would be the extraction of information about the structure of the carbohydrate sidechains directly from the NMR spectrum of an intact glycoprotein. While ^1H NMR experiments on intact polysaccharides and glycolipids are routine, ^1H NMR spectroscopy of glycoproteins has been somewhat limited for two reasons. First, the long rotational correlation times of globular proteins larger than about 20,000 Daltons shortens T_2 beyond the point at which spin correlation can be used in the proton assignment. Second, the resonances of many of the protons of the peptide overlap those of the sugars. For example the resonances of the α-and β-protons of the amino acids fall in the 4.8 to 4.0 ppm region and cover many anomeric proton resonances. For proteins in which amino acids are much more abundant than carbohydrate residues, it is impossible to distinguish the sugars so complete proton assignments for glycoproteins have been limited to some special types of glycoproteins that are internally flexible and have high sugar content and some repetitive structure. Our laboratory has carried out complete proton assignments of several antifreeze glycoproteins, important components of the blood of polar fish. This class of glycoproteins has a simple repeating sequence and the completely assigned proton NMR spectra have been used in conformational modeling studies (67,68). A somewhat similar situation occurs in mucin glycoproteins which are structurally related to antifreeze glycoproteins and have been studied by proton NMR spectroscopy (69).

The improved chemical shift dispersion of ^{13}C NMR spectroscopy provides greater possibilities for experiments on intact glycoproteins. Dill and Allerhand (70) have pointed out that there is a window in the carbon resonances of the peptide residues above the aromatic resonances (160-110 ppm) and below the resonances of the α-carbons at about 70 ppm. Since the resonances of all the anomeric carbons of the sugar fall in between 110 and 90 ppm they can be detected even in the presence of a substantial excess of peptide signal. Some of the remaining carbon resonances fall into this window region permitting the extraction of information on the carbohydrate structure. In spite of the poor sensitivity of direct ^{13}C detection, valuable studies on intact glycoproteins have been reported including some on globular proteins (71) and the ^{13}C spectrum of antifreeze glycoprotein has been completely assigned (72). Several extensive studies of mucins, including those with blood group active carbohydrate sidechains, have explored the dynamics of the carbohydrate showing that it is generally more flexible than the peptide core (73,74).

Carbon-13 detection, which was used in all the studies described above, requires 10-30 micromoles of sample in the NMR probe which translates into 100 mg of a 30,000 Dalton protein and presents a major obstacle to these studies. Even if a glycoprotein can be obtained in such prodigious quantities, it may not form well behaved aqueous solutions at such high concentration. The prospect of overcoming the sensitivity problem is offered by the method of ^1H detected ^{13}C spectroscopy which can provide 2-dimensional C-H correlated spectra on a few micromoles of sample. This technique, which was described above in connection with the problem of assignment of the ^{13}C spectra of oligosaccharides, has the effect of spreading out the proton spectrum in the carbon dimension. Thus the anomeric proton resonances which occur in an intact glycoprotein between 4.4 and 5.4 ppm, would appear between 95 and 105 ppm in the ^{13}C axis, a region in which peptide resonances are absent. It is possible that inverse detected 2-dimensional ^{13}C-^1H correlation spectroscopy could provide detailed structural data on a few milligrams of an intact glycoprotein.

Acknowledgments

I wish to thank Perseveranda Cagas for the ^1H data on LND-I and C. Abeygunawardana for the ^1H-^{31}P correlation data on the *S.s.* 34 polysaccharide. Research supported by NSF grant DMB 8517421 and by NIH grant GM-31449.

VIII. References

[1]Murayama, J., Tomita, M., and Hamada, A. *J. Biochem. (Tokyo)*, **91**, 1829-1836 (1982).

[2]Kornfeld, R. and Kornfeld, S. *Ann. Rev. Biochem.*, **54**, 631-664 (1985).

[3]Harford, J. and Ashwell, G. *Methods in Enzymol.* **109**, 232-246 (1985).

[4]Kornfeld, S. *Federation J.* **1**, 462-468 (1987).

[5]Feeney, R. E., Burcham, T. S., and Yeh, Y. *Annu. Rev. Biophys. Biophys. Chem.* **15**, 59-78 (1986).

[6]Karlsson, K. A. *Chem Phys Lipids* **42**, 153-172 (1986).

[7]Vliegenthart, J. F. G., Dorland, L., and van-Halbeek, H. *Advan. Carbohyd. Chem. Biochem.*, **41**, 209-374 (1983).

[8]Dua, V. K. and Bush, C. A. *Analyt. Biochem.* **133**, 1-10. (4), 386-397 (1983).

[9]Dua, V. K., Goso, K., Dube, V. E., and Bush, C. A. *J. Chromatogr.*

328, 259-269 (1985).

[10]Vliegenthart, J. F. G., vanHalbeek, H., and Dorland, L. *Pure Appl. Chem.* **53**, 45-77 (1981).

[11]Hanfland, P., Egge, H., Dabrowski, U., Kuhn, S., Roelke, D., and Dabrowski, J. *Biochemistry* **20**, 5310-5319 (1981).

[12]Rao, B. N. N. and Bush, C. A. *Carbohyd. Res.* **180**, 111-128 (1988).

[13]Dabrowski, J., Egge, H., and Hanfland, P. *Pure Appl. Chem.* **56**, 807-819 (1984).

[14]Yu, R. K., Koerner, T. A., Scarsdale, JN, and Prestegard, J. H. *Chem. Phys. Lipids* **42**, 27-48 (1986).

[15] Koerner, T. A. W., Prestegard, J. H., and Yu, R. K. *Methods Enzymol.*, **138**, 38-59 (1987).

[16] Clausen, H., Levery, S. B., Nudelman, E. D., Stroud, M., Salyan, M. E., and Hakomori, S. *J. Biol. Chem.*, **262**, 14228-34 (1987).

[17]Jones, C. *Carbohyd. Res.*, **139**, 75-83 (1985).

[18]Kenne, L. *Stud. Org. Chem.*, **26**, 199-210 (1986).

[19]Gorin, P. A. J. *Advan. Carbohyd. Chem. Biochem.*, **38**, 13-104 (1981).

[20]Jennings, H. J. *Advan. Carbohyd. Chem Biochem.*, **41**, 155-208 (1983).

[21]Romanowska, E., Gamian, A., and Dabrowski, J. *Eur. J. Biochem.*, **161**, 557-564 (1986).

[22]McIntire, F. C., Bush, C. A., Wu, S. S., Li, S. C., Li, Y. T., McNeil, M., Tjoa, S. S. and Fennessey, P. V. *Carbohyd. Res.*, **166**, 133-143 (1987).

[23]Carver, J. P. and Grey, A. A. *Biochemistry* **20**, 6607-6616 (1981).

[24]Trimble, R. B. and Atkinson, P. H. *J. Biol. Chem.* **261**, 9815-24 (1986).

[25]Dua, V. K., Rao, B. N., Wu, S. S., Dube, V. E., and Bush, C. A. *J. Biol. Chem.* **261**, 1599-608 (1986).

[26]Homans, S. W., Dwek, R., A, and Rademacher, T. W. *Biochemistry* **26**, 6571-6578 (1987).

[27]Bax, A. "Two-dimensional Nuclear Magnetic Resonance in Liquids" D. Reidel, Dordrect, Holland, 1982.

[28]Marion, D. and Wuthrich, K. *Biochem. Biophys. Res. Comm.* **113**, 967-974 (1983).

[29]States, D. J., Haberkorn, R. A., and Ruben, D. J. (1982) *J. Magn. Reson.* **48**, 286-292.

[30]Keeler, J. and Neuhaus, D. *J. Magn. Reson.* **63**, 454-72 (1985).

[31]Widmer, H. and Wuthrich, K. *J. Magn. Res.* **74**, 316-336 (1987).

[32]Homans, S. W., Dwek, R. A., Boyd, J., Soffe, N., and Rademacher, T. W. *Proc. Natl. Acad. Sci., USA* **84**, 1202-1205 (1987).

[32a]Dabrowski, J., Dabrowski, U., Bermel, W., Kordowicz, W. and Hanfland, P. *Biochemistry* **27**, 5149-5155 (1988).

[33] Bax, A. and Lerner, L. *Science (Washington)* **232**, 960-967 (1986).

[34]Summers, M. F., Marzilli, L. G., and Bax, A. *J. Am. Chem. Soc.* **108**, 4285-4894 (1986) .

[35]Rance, M. *J. Magn. Res.* **74**, 557-564 (1987).

[36]Homans, S. W., Dwek, R. A., Boyd, J., Mahmoudian, M., Richards, W. G. and Rademacher, T. W. *Biochemistry* **25**, 6342-6350 (1986).

[37]Skelnar, V., Tschudin, R., and Bax, A. *J. Magn. Res.* **75**, 352-357 (1987).

[38]Batta, G. and Liptak, A. *J. Am. Chem. Soc.* **106**, 248-250 (1984).

[39]Bush, C. A., Yan, Z. -Y., and Rao, B. N. N., *J. Am. Chem. Soc.* **108**, 6168-6173 (1986).

[40]Homans, S. W., Dwek, R. A. and Rademacher, T. W., *Biochemistry* **26**, 6553-6560 (1987).

[41]Koerner, T. A. W., Prestegard, J. C., Demou, P. C., and Yu, R. K. *Biochemistry* **22**, 2687-2690 (1983).

[42]Rao, B. N. N., Dua, V. K. and Bush, C. A. *Biopolymers* **24**, 2207-2229 (1985).

[43]Bax, A. *J. Magn. Resonance* **77**, 134-147 (1988).

[44]Cumming, D. A. and Carver, J. P. *Biochemistry* **26**, 6664-6676 (1987).

[45]Yan, Z. -Y., Rao, B. N. N. and Bush, C. A. *J. Am. Chem. Soc.* **109**, 7663-7669 (1987).

[46]Bush, C. A., Panitch, M. M., Dua, V. K., and Rohr, T. E. *Anal. Biochem.* **145**, 124-136 (1985).

[47]Allerhand, A. and Berman, E. *J. Am. Chem. Soc.* **106**, 2400-2412 (1984).

[48]Allerhand, A. and Berman, E. *J. Am. Chem. Soc.* **106**, 2412-2420 (1984).

[49]Barker, R. and Serianni, A. S. *Acc. Chem. Res.* **19**, 307-313 (1986).

[50]Doddrell, D. M. and Pegg, D. T. *J. Am. Chem. Soc.* **102**, 6388-6390 (1980).

[51]Pfeffer, P., Valentine, K., and Parrish, F. *J. Am. Chem. Soc.* **101**, 1265-1274 (1979).

[52]Blumberg, K. and Bush, C. A. *Analyt. Biochem.* **119**, 397-406 (1982).

[53]Matsuura, F., Nunez, H. A., Grabowski, G. A., and Sweeley, C. C. Archives Biochem. and Biophys. **207**, 337-352 (1981).

[54]Shashkov, A. S., Arbatsky, N. P., Derevitskaya, V. A., and Kochetkov, N. K. *Carbohyd. Res.* **72**, 218-221 (1979).

[55]Sillerud, L. O., Prestegard, J. H., Yu, R. K., Schafer, D. E., and Konigs berg, W. H. *Biochemistry* **17**, 2619-2628 (1978).

[56]Prohaska, R., Koerner, T. A. W., Armitage, I. M., and Furthmayr, H. *J. Biol. Chem.* **256**, 5781-5791 (1981).

[57]Bock, K., Pedersen, C. and Pedersen, H. *Adv. Carbohyd. Chem. Biochem.* **42**, 193-225 (1984).

[58]Dill, K., Berman, E., and Pavia, A. A. *Advan. Carbohyd. Chem.*

and Biochem. **43**, 1-49 (1985).

[59]Lerner, L. and Bax, A. *Carbohyd. Res.* **166**, 35-46 (1987).

[60]Dabrowski, J., Hauck, M., Romanowska, E. and Gamian, A. *Biochem. Biophys. Res. Commun.* **146**, 1283-1285 (1987).

[60a]Tsui, F. P., Egan, W., Summers, M. F., Byrd, R. A., Schneerson, R. and Robbins, J. B. *Carbohyd. Res.* **173**, 65-74 (1988).

[61]Bax, A., Egan, W., and Kovac, P. *J. Carbohyd. Chem.* **3**, 593-611 (1984).

[62] Byrd, R. A., Egan, W., Summers, M. F., and Bax, A. *Carbohyd. Res.* **166**, 47-58 (1987).

[63]Gorenstein, D. G. *Ann. Rev. Biophys. Bioengin.* **10**, 355-386 (1981).

[64]Jennings, H. J., Rosell, K. . G., and Carlo, D. J. *Canad. J. Chem.* **58**, 1069-1074 (1980).

[65]Hellerqvist, C. G., Sundell, H., and Gettins, P. *Proc. Natl. Acad. Sci., USA* **84**, 51-55 (1987).

[66]McIntire, F. C., Crosby, L. K., Vatter, A. E., Cisar, J. O., McNeil, M. R., Bush, C. A., Tjoa, S. S., and Fennessey, P. V. *J. Bacteriol.* **170**, 2229-2235 (1988).

[67]Bush, C. A. and Feeney, R. E. *Int. J. Pept. Protein Res.* **28**, 386-397 (1986).

[68] Rao, B. N. N. and Bush, C. A. *Biopolymers* **26**, 1227-1244 (1987).

[69]Gerken, T. A. *Arch. Biochem. Biophys.* **247**, 239-253 (1986).

[70]Dill, K. and Allerhand, A. *J. Biol. Chem.* **254**, 4524-4531 (1979).

[71]Berman, E., Walters, D. E., and Allerhand, A. *J. Biol. Chem.* **256**, 3852-3857 (1981).

[72]Berman, E., Allerhand, A., and DeVries, A. L. *J. Biol. Chem.* **255**, 4407-4410 (1980).

[73]Gerken, T. A. and Dearborn, D. G. *Biochemistry* **23**, 1485-1497 (1984).

[74]Gerken, T. A. and Jentoft, N. *Biochemistry* **26**, 4689-4699 (1987).