Carbohydrate Research, 205 (1990) 19–29
Elsevier Science Publishers B.V., Amsterdam

Computer-assisted interpretation of ^{1}H-n.m.r. spectra in the analysis of the structure of oligosaccharides*

Elizabeth F. Hounsell†
Section of Glycoconjugate Research, Clinical Research Centre, Watford Road, Harrow, Middlesex HA1 3UJ (Great Britain)

and Dennis J. Wright
Clinical Chemistry, Northwick Park Hospital, Watford Road, Harrow, Middlesex HA1 3UJ (Great Britain)

(Received December 9th, 1989; accepted for publication, February 3rd, 1990)

ABSTRACT

A computer program that can be implemented on IBM compatible microcomputers has been used to interpret the ^{1}H-n.m.r. spectra of a series of mucin-derived alditols. The program compares the chemical shifts of resonances in the spectra of unknown oligosaccharides with those in a standard library constructed from data in the literature. The program then compiles the identified sequences into possible structures, which, together with a second component of the program from which the individual assignments of chemical shifts can be made, facilitates ready access to the literature data for final confirmation of the structures.

INTRODUCTION

Several techniques in high-field ^{1}H-n.m.r. spectroscopy are now available for analysis of the structure of oligosaccharides, including ^{1}H–^{1}H correlated spectroscopy (COSY)[1–4], multistep relayed correlation spectroscopy (RECSY)[3–6], triple quantum filtered spectroscopy (TQCOSY)[7], ^{1}H-detected ^{13}C spectroscopy[8], and homonuclear Hartmann–Hahn spectroscopy (HOHAHA)[9], which have led to a rapid increase in chemical shift data that are now difficult to search manually. We have described[10] a computer program designed to search a database of the chemical shifts of the ^{1}H resonances of oligosaccharides and provide possible assignments of structure of unknown oligosaccharides by comparison with data derived from the literature. The use of this computer program was described with examples of oligosaccharides obtained by chemical or enzymic synthesis, from enzymic degradation of N-linked type chains, and from human milk[10]. We now report an updated version of the program which has been made more accessible by transfer to IBM PC compatible machines and improved to give automatically complete assignments of structures.

The program is applicable to the complete range of oligosaccharides, including those from O- and N-linked glycoproteins. The validation of this approach for analysing the extensive data obtained from multiple oligosaccharides of mucin glycoproteins

*Dedicated to Professor Leslie Hough in the year of his 65th birthday.
† Author for correspondence.

0008-6215/90/$ 03.50

is described. The program has been used to compare data from the spectra of oligosaccharides from human foetal gastrointestinal mucins[11,12] with those in the literature for oligosaccharides of gastrointestinal tract mucins of other species and bronchial and ovarian cyst mucins of humans[13–27].

EXPERIMENTAL

The computer program used in earlier work[10,11] has been rewritten in TURBO PASCAL V5.0 (Borland International, CA, U.S.A.) and runs on an IBM Personal Computer under MS DOS V4.00. Disc storage requirements are: program, 69 kbytes; standard file, 135 kbytes; and a single data file, 0.5 kbyte.

The data from ^{1}H-n.m.r. spectra for solutions in D_2O of unknown oligosaccharides are entered into a data file as a series of resonance frequencies. The program converts these data into p.p.m. relative to a standard signal from acetone at δ 2.225 (at 22°) and then makes a comparison with those in the standard library. A series of "standard sequences" has been defined in oligosaccharides from mammalian glycoproteins and those obtained from mammalian sources as reducing oligosaccharides. The linkage of each monosaccharide in the sequence is specified and the computer contains data on the resonances attributable to some or all of the protons for the specified monosaccharide in the series of standard sequences.

Data for the same oligosaccharide sequences in different oligosaccharides have been compared and a tolerance limit of ± 0.03 p.p.m. for the majority of protons and ± 0.005 p.p.m. for NAc resonances defined, within which it is possible to construct and search a manageable library of sequences. The computer compares the information on the standard sequences with the data in the unknown spectrum and produces a list of those standard sequences in which there is a resonance in the spectrum within ± 0.03 p.p.m. for protons in the sequence given to two decimal places in the library and within ± 0.005 p.p.m. for those given to three decimal places. The standard sequence nomenclature, described in the Results, has been retained from the earlier program[10], but the system has now been updated to compile possible structures from the list of standard sequences identified in the unknown.

The number of possible structures can be restricted by including only core initiator sequences that are likely to exist in the sample and by defining the probable molecular size, which can be estimated in several ways. For example, the position of elution in size-exclusion chromatography on Bio-Gel P-4 gives an approximate range of sizes for an oligosaccharide based on glucose units which can now be indicated in the program to provide assignments within that range. Additional information on size is usually readily obtained also from the ^{1}H-n.m.r. spectra, such as the number of signals for anomeric protons and groups, and/or from m.s. (liquid secondary-ion m.s. of native or methylated oligosaccharides[11,12]).

Once an oligosaccharide of appropriate size has been identified by the program, the chemical shifts can be assigned to the resonances of specific protons, using another section of the program, and the shifts then compared with the data for that oligosaccha-

ride in the literature to within ± 0.005 p.p.m., which is the normal laboratory-to-laboratory variation for all proton resonances from identical oligosaccharides.

The spectra used in the present analysis and the approximate molecular size based on elution from Bio-Gel P-4 are from the di- to hexa-saccharide-alditol fractions described in refs. 11 and 12. The standard library comprises the literature references noted in refs. 10 and 16–28.

TABLE I

List of standard sequences having all resonances in the library identified in the spectrum of the example pentasaccharide

α-D-Glc*p*NAc-(1→4)-β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→6)
GalNAcol
β-D-Gal*p*-(1→3)

Number of chemical shifts in the library	*Standard sequence*	*Sequence number*
6	HEXB3[HEXNB6]GALNOL	20
8	HEXB4GLCOL	90
5	HEXB4[DHEXA3]GLCOL	100
8	HEXNB3GALOL	160
3	SAA3GALB4[DHEXA3]HEX/HEXN	191
3	SAA3GALB3[SAA06]HEXN	193
3	HEXNA4GALB43	200
5	HEXNB3GALB43HEXNB/OL	210
3	HEXA3GALB4HEXN	211
2	HEXNB3GALB3[HEXNB6]HEXNOL	212
2	HEXNA4GALB3[HEXNB6]HEXNOL	213
6	DHEXA2GALB43HEXNB63[HEXB036/HEXNB036]HEXNOL	220
2	DHEXA2GALB3[HEXNB6]HEXNOL	243
3	HEXNB3GALB4HEX	250
4	HEXNB3GALB4HEXOL	260
4	HEXNB3[HEXNB6]GALB4HEXOL	261
2	HEXNB3GALB4[DHEXA3]HEXOL	262
2	HEXNB3[HEXNB6]GALB43HEXNB/OL	280
3	HEXB4/HEXNB4GLCNB4HEXN/OL	290
3	HEXB4GLCNB6[HEXB3]HEXNOL	302
3	HEXB4GLCNB6[HEXNB3]HEXNOL	304
7	DHEXA2[HEXNA03]HEXB4GLCNB6	321
2	SAA6/DHEXA2HEXB4GLCNB2HEX	351
2	GLCNB4[HEXA3][HEXA6]HEXB4	421
7	GALB643	430
3	GALB4[DHEXA3]HEXN	431
2	GALB3[DHEXA4]HEXN	432
3	GLCNB6[HEXB3]HEXNOL	440
3	GLCNA4HEX	500

TABLE II

Illustration of the use of rules *A–D* (see Results) by the program to combine the standard sequences identified in Table I to give possible structural assignments of structure for the pentasaccharide in Table I (the bold italics show monosaccharides added to the chain on each iteration)

		Rule
—B3[HEXNB6]GALNOL	20	
—A4GALB3[HEXNB6]GALNOL	200	*A*
GLCNA4GALB3[HEXNB6]GALNOL	500	*A*
—B4GLCNB6{GLCNA4GALB3}GALNOL	302	*A*
—A4GALB4GLCNB6{GLCNA4GALB3}GALNOL	200	*A*
GLCNA4GALB4GLCNB6{GLCNA4GALB3}GALNOL	500	*C*
TOO MANY SACCHARIDES IN CHAIN		
—B3GALB4GLCNB6{GLCNA4GALB3}GALNOL	210	*D*
—A3GALB4GLCNB6{GLCNA4GALB3}GALNOL	211	*D*
—A2GALB4GLCNB6{GLCNA4GALB3}GALNOL	220	*D*
—B3[HEXNB6]GALB4GLCNB6{GLCNA4GALB3}GALNOL	280	*D*
GALB4GLCNB6{GLCNA4GALB3}GALNOL	430	*B*
GALB4GLCNB6{GLCNA4GALB3}GALNOL		
—B4GLCNB6{GLCNA4GALB3}GALNOL	321	*A*
—A2GALB4GLCNB6{GLCNA4GALB3}GALNOL	220	*D*
—B4GLCNB6{GLCNA4GALB3}GALNOL	321	*D*
GLCNB6{GLCNA4GALB3}GALNOL	440	*B*
STRUCTURE NOT IN SIZE RANGE		
—B3GALB3[HEXNB6]GALNOL	212	*D*
—A4GALB3[HEXNB6]GALNOL	213	*A*
GLCNA4GALB3[HEXNB6]GALNOL	500	*A*
—B4GLCNB6{GLCNA4GALB3}GALNOL	300	*A*
—A4GALB4GLCNB6{GLCNA4GALB3}GALNOL	200	*A*
GLCNA4GALB4GLCNB6{GLCNA4GALB3}GALNOL	500	*C*
TOO MANY SACCHARIDES IN CHAIN		
—B3GALB4GLCNB6{GLCNA4GALB3}GALNOL	210	*D*
—A3GALB4GLCNB6{GLCNA4GALB3}GALNOL	211	*D*
—A2GALB4GLCNB6{GLCNA4GALB3}GALNOL	220	*D*
—B3[HEXNB6]GALB4GLCNB6{GLCNA4GALB3}GALNOL	280	*D*
GALB4GLCNB6{GLCNA4GALB3}GALNOL	430	*B*
STRUCTURE ALREADY IDENTIFIED		
—B4GLCNB6{GLCNA4GALB3}GALNOL	321	*A*
—A2GALB4GLCNB6{GLCNA4GALB3}GALNOL	220	*D*
—B4GLCNB6{GLCNA4GALB3}GALNOL	321	*D*
GLCNB6{GLCNA4GALB3}GALNOL	440	*B*
STRUCTURE NOT IN SIZE RANGE		
—A2GALB3[HEXNB6]GALNOL	243	*D*
GALB3[HEXNB6]GALNOL	430	*A*
—B4GLCNB6{GALB3}GALNOL	302	*A*
—A4GALB4GLCNB6{GALB3}GALNOL	200	*A*
GLCNA4GALB4GLCNB6{GALB3}GALNOL	500	*B*
GLCNA4GALB4GLCNB6{GALB3}GALNOL		
—B3GALB4GLCNB6{GALB3}GALNOL	210	*D*
—A3GALB4GLCNB6{GALB3}GALNOL	211	*D*
—A2GALB4GLCNB6{GALB3}GALNOL	220	*D*
—B3[HEXNB6]GALB4GLCNB6{GALB3}GALNOL	280	*D*
GALB4GLCNB6{GALB3}GALNOL	430	*B*

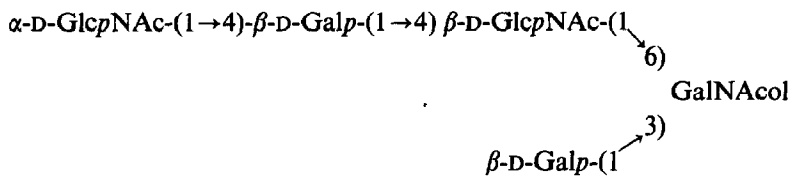
TABLE II

STRUCTURE NOT IN SIZE RANGE		
—*B4GLCN*B6{GALB3}GALNOL	321	*D*
—*A2GALB4GLCN*B6{GALB3}GALNOL	220	*D*
—*B4GLCN*B6{GALB3}GALNOL	321	*D*
*GLCN*B6{GALB3}GALNOL	440	*B*
STRUCTURE NOT IN SIZE RANGE		
—*B4GLCOL*	90	
—*A4GAL*B4GLCOL	200	*A*
*GLCN*A4GALB4GLCOL	500	*B*
STRUCTURE NOT IN SIZE RANGE		
—*B3GAL*B4GLCOL	260	*D*
—*B3[HEXNB6]GAL*B4GLCOL	261	*D*
*GAL*B4GLCOL	430	*B*
STRUCTURE NOT IN SIZE RANGE		
—*B4[DHEXA3]GLCOL*	100	
—*A4GAL*B4[DHEXA3]GLCOL	200	*A*
*GLCN*A4GALB4[DHEXA3]GLCOL	500	*D*
—*B3GAL*B4[DHEXA3]GLCOL	262	*D*
*GAL*B4[DHEXA3]GLCOL	430	*D*
—*B3GALOL*	160	

Good structures, 2; residue chain-length, 5; time taken, 20 s.

RESULTS

Tables I and II demonstrate the working of the program to give possible assignments of structure based on comparison of the standard sequences in the library with the spectral data for the pentasaccharide[12]



In order to identify this oligosaccharide, the program has to find chemical shift matches for the following three groups of "standard sequences":

chain initiators, *e.g.*, HEXB3[HEXNB6]GALNOL;

chain extenders, *e.g.*, HEXB4GLCNB6[HEXB3]HEXNOL and HEXNA4GALB43;

chain terminators, *e.g.*, GALB643 and GLCNA4HEX.

Taking each of the chain initiator standard sequences, the program works systematically through all of the remaining standard sequences, as indicated in Table II, by applying the following rules: *A*, a matching standard sequence is found; or *B*, a valid structure is identified with a chain-terminating sequence; or *C*, the constructed chain

TABLE III

The structures identified by the program from the spectra of a series of mucin-derived oligosaccharide-alditols (the structures underlined have been confirmed as the major oligosaccharides present by reference to the original data used in the standard library and additional structural methods[11,12])

OLIGOSACCHARIDE 1	
GALB3GALNOL	385.4 Da
Good structures, 1; residue chain-length, 2–3; time taken, 3 s.	
OLIGOSACCHARIDE 2	
GLCNB3GALNOL	426.4 Da
Good structures, 1; residue chain-length, 2–4; time taken, 3 s.	
OLIGOSACCHARIDE 3	
GLCNB6{GALB3}GALNOL	588.6 Da
Good structures, 1; residue chain-length, 3–4; time taken, 3 s.	
OLIGOSACCHARIDE 4	
GLCNB3GALB4GLCNB3GALNOL	791.8 Da
GALB4GLCNB3GALNOL	588.6 Da
GLCNB3GALB3GLCNB3GALNOL	791.8 Da
GALB3GLCNB3GALNOL	588.6 Da
GLCNB3GALB4GLCOL	547.6 Da
GLCNB3GALB3GLCNOL	588.6 Da
Good structures, 6; residue chain-length, 3–4; time taken, 9 s.	
OLIGOSACCHARIDE 5	
GALB4GLCNB6{GALB3}GALNOL	750.8 Da
Good structures, 1; residue chain-length, 4–5; time taken, 4 s.	
OLIGOSACCHARIDE 6	
GLCNB3GALB3GLCNB3GALB3GALB3GALNOL	954.0 Da
GALB3GLCNB3GALB3GALNOL	750.8 Da
GLCNB3GALB4GLCNB3GALB3GALNOL	954.0 Da
GALB4GLCNB3GALB3GALNOL	750.8 Da
GLCNB3GALB3GLCNB3GALB4GLCOL	913.0 Da
GALB3GLCNB3GALB4GLCOL	709.8 Da
GLCNB3GALB4GLCNB3GALB4GLCOL	913.0 Da
GALB4GLCNB3GALB4GLCOL	709.8 Da
Good structures, 8; residue chain-length, 4–5; time taken, 56 s.	

TABLE III

OLIGOSACCHARIDE 7

GALB4GLCNB6{FUCA2GALB3}GALNOL	897.0 Da
FUCA2GALB4GLCNB6{GALB3}GALNOL	897.0 Da
GALB4GLCNB6{GALB3}GALNOL	750.8 Da
GALB4GLCNB6{GALB4GLCNB3}GALNOL	954.0 Da
GALB4GLCNB6{GLCNB3}GALB4GLCNOL	954.0 Da
GLCNB4GLCNB4GLCNB4GLCNB4GLCNOL	1036.0 Da
GALB4GLCNB4GLCNB4GLCNB4GLCNOL	995.0 Da
GLCNB4GLCNB4GLCNB4GLCNOL	832.8 Da
GALB4GLCNB4GLCNB4GLCNOL	791.8 Da

Good structures, 9; residue chain-length, 4–5; time taken, 113 s.

OLIGOSACCHARIDE 8

GALB4GLCNB6{FUCA2GALB3}GALNOL	897.0 Da
GALB4GLCNB4GLCNB4GLCNB4GLCNOL	995.0 Da
GALB4GLCNB4GLCNB4GLCNOL	791.8 Da

Good structures, 3: residue chain-length, 4–5; time taken, 10 s.

OLIGOSACCHARIDE 9

GLCNB3GALB3GLCNB3GALB4GLCNB3GALNOL	1157.2 Da
GALB3GLCNB3GALB4GLCNB3GALNOL	954.0 Da
GLCNB3GALB4GLCNB3GALB4GLCNB3GALNOL	1157.2 Da
GALB4GLCNB3GALB4GLCNB3GALNOL	954.0 Da

Good structures, 4: residue chain-length, 5–6; time taken, 5 s.

OLIGOSACCHARIDE 10

GALB4GLCNB6{GALB3GLCNB3}GALNOL	954.0 Da
GALB4GLCNB6{GALB4GLCNB3}GALNOL	954.0 Da

Good structures, 2; residue chain-length, 5–6; time taken, 32 s.

OLIGOSACCHARIDE 11

GLCNB6{GLCNB3GALB3GLCNB3GALB3}GALNOL	1157.2 Da
GALB4GLCNB6{GALB3GLCNB3GALB3}GALNOL	1116.2 Da
GLCNB6{GALB3GLCNB3GALB3}GALNOL	954.0 Da
GLCNB6{GLCNB3GALB4GLCNB3GALB3}GALNOL	1157.2 Da
GALB4GLCNB6{GALB4GLCNB3GALB3}GALNOL	1116.2 Da
GLCNB6{GALB4GLCNB3GALB3}GALNOL	954.0 Da
GLCNB3GALB4GLCNB6{GLCNB3GALB3}GALNOL	1157.2 Da
GALB4GLCNB6{GLCNB3GALB3}GALNOL	954.0 Da
GALB3GLCNB3GALB4GLCNB6{GALB3}GALNOL	1116.2 Da
GALB4GLCNB3GALB4GLCNB6{GALB3}GALNOL	1116.2 Da
GLCNB3GALB4GLCNB6{GALB3}GALNOL	954.0 Da
GALB3GLCNB3GALB3GLCNB3GALB4GLCOL	1075.2 Da
GALB4GLCNB3GALB3GLCNB3GALB4GLCOL	1075.2 Da
GLCNB3GALB3GLCNB3GALB4GLCOL	913.0 Da

TABLE III

GALB3GLCNB3GALB4GLCNB3GALB4GLCOL	1075.2 Da
GALB4GLCNB3GALB4GLCNB3GALB4GLCOL	1075.2 Da
GLCNB3GALB4GLCNB3GALB4GLCOL	913.0 Da
GALB3GLCNB3GALB3GLCNB3GALB3GLCNOL	1116.2 Da
GALB4GLCNB3GALB3GLCNB3GALB3GLCNOL	1116.2 Da
GLCNB3GALB3GLCNB3GALB3GLCNOL	954.0 Da
GALB3GLCNB3GALB4GLCNB3GALB3GLCNOL	1116.2 Da
GALB4GLCNB3GALB4GLCNB3GALB3GLCNOL	1116.2 Da
GLCNB3GALB4GLCNB3GALB3GLCNOL	954.0 Da

Good structures, 23; residue chain-length, 5–6; time taken, 263 s.

OLIGOSACCHARIDE 12

GLCNA4GALB4GLCNB6{GLCNA4GALB3}GALNOL	1157.2 Da

Good structures, 1; residue chain-length, 5–6; time taken, 13 s.

OLIGOSACCHARIDE 13

GALB4GLCNB6{GALB3GLCNB3GALB3}GALNOL	1116.2 Da
GLCNB6{GALB3GLCNB3GALB3}GALNOL	954.0 Da
GALB3GLCNB3GALB4GLCNB6{GALB3}GALNOL	1116.2 Da
GALB3GLCNB3GALB3GLCNB3GALB4GLCOL	1075.2 Da

Good structures, 4; residue chain-length, 5–6; time taken, 76 s.

OLIGOSACCHARIDE 14

GLCNA4GALB4GLCNB6{GLCNA4GALB3}GALNOL	1157.2 Da
GLCNA4GALB4GLCNB6{FUCA2GALB3}GALNOL	1100.2 Da

Good structures, 2; residue chain-length, 5–6; time taken, 20 s.

becomes longer than the definable number of saccharide residues; or *D*, no matching standard sequence is found.

The presence of oligosaccharides suggested by the program can be substantiated by reference to the literature that gives the original spectral data for the suggested oligosaccharide(s). In addition to identifying structures for which data have been published, the program can construct likely structures from compatible sequences. This capability is illustrated by the results of the analysis of the pentasaccharide noted above, in that data for this oligosaccharide itself are not in the literature but only those for related oligosaccharides (Tables I–III).

Table III shows the structures derived by the computerised interpretation from the spectra of a series of mucin-derived di- to hexa-saccharide-alditol fractions. The structure(s) underlined are those which were proved subsequently to be the major oligosaccharides present[11,12]. This series illustrates the wide variety of structures of

mucin-type oligosaccharides that can be assigned by the program. Those identified in Tables II and III are a particular subset based on four different core regions, namely, β-D-Gal*p*-(1→3)-GalNAcol, β-D-Glc*p*NAc-(1→3)-GalNAcol, β-D-Gal*p*-(1→3)[β-D-Glc*p*NAc-(1→6)]-GalNAcol, and β-D-Glc*p*NAc-(1→3)[β-D-Glc*p*NAc-(1→6)]-GalNAcol, that have β-D-Gal*p*-(1→3/4) or β-D-Gal*p*-(1→3/4)-β-D-Glc*p*NAc-(1→3) backbones, with or without α-D-Glc*p*NAc-(1→4) or α-L-Fuc*p*-(1→2) peripheral regions linked to β-D-Gal*p*-(1→3)-GalNAcol or β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→6)-GalNAcol.

In this mucin series, there were two additional spectra of oligosaccharides in the di- and tri-saccharide size range for which the program gave no possible structures and one further spectrum of a hexasaccharide which gave no structures that had a GalNAcol core. The major oligosaccharides in these fractions were shown subsequently, by additional n.m.r. spectral assignment[3,29] and structural analysis methods involving m.s.[11,12], to be structures based on GalNAcol-containing cores but with novel sequences for which previous literature data had not been published. Two of these sequences were the di- and tri-saccharides[3] α-D-Gal*p*NAc-(1→3)-GalNAcol and β-D-Gal*p*-(1→4)-Glc*p*NAc-(1→6)-GalNAcol that have a novel monosaccharide or linkage at GalNAcol. The third oligosaccharide, shown to be the hexasaccharide[29] β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→6)-β-D-Gal*p*-(1→3)[β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→6)]-GalNAcol, has a novel backbone linkage, namely, β-D-Glc*p*NAc-(1→6)-β-D-Gal*p*.

In addition to the specificity of the program for distinguishing novel oligosaccharides, it can be seen that, although the standard database includes other monosaccharides besides those commonly found in O-linked-type chains of mucins, these are excluded by the search of the data from the oligosaccharides in the present study. On the other hand, the computer analysis of the spectra of oligosaccharides from N-linked glycoprotein chains[30] gives assignments based on the typical α-D-Man*p*-(1→6)[α-D-Man*p*-(1→3)]-β-D-Man*p*-(1→4)-β-D-Glc*p*NAc sequence, and spectra from sialylated compounds, for example, are given correct structural assignments[10,30].

DISCUSSION

The identification of individual oligosaccharides from the large number of possible structures in mammalian glycoconjugates is an increasingly important task as more structures are identified and the roles of their specific sequences in biological systems are recognised. When enough material is available (~ 50 μg or more), ^{1}H-n.m.r. analysis is an important first-line procedure which has the advantage of providing complete information on structure, identifying novel sequences, and giving insights into conformation[4,10,28–32] without degradation of the material. Where practicable, a combined approach to the assignment of structure, including chemical, enzymic, chromatographic, mass-spectrometric, and n.m.r. analysis, is essential[31]. Newer methods of analysis involving specific degradation and direct t.l.c.–m.s. are also available which complement the existing techniques[33,34]. The present study demonstrates the extent of

information to be obtained from ^{1}H-n.m.r. spectroscopy alone and the use of computer-assisted analysis of the spectra.

The results show that the program can give readily a first line of evidence for the structures of oligosaccharides which can be checked by reference to the primary source of information used to construct the database, *i.e.*, as far as possible all of the ^{1}H-n.m.r. data for oligosaccharides from mammalian glycoproteins quoted in the literature. The method chosen for cross-referencing the literature data involves the use of the CarbBank CCSD program[35]. Thus, the possible structures given by the computer search of the ^{1}H-n.m.r. data can be identified by CarbBank and the literature reference for the ^{1}H-n.m.r. data of the oligosaccharides obtained in this way. Because the program for the interpretation of ^{1}H-n.m.r. data includes a parameter to assign each peak, the chemical shift data for each monosaccharide in the sequence can be obtained and checked against the literature data at ± 0.005 p.p.m. (the laboratory-to-laboratory variation) in order to provide a unique assignment.

From the computer-assisted interpretation of ^{1}H-n.m.r. spectra so far carried out, it has been found that, using a ± 0.03–0.005 p.p.m. tolerance limit, the program can identify many different oligosaccharides and distinguish between closely related structures. This facility is a useful first step in the determination of structure by ^{1}H-n.m.r. spectroscopy for workers familiar or unfamiliar with spectra interpretation. Other computerised approaches to ^{1}H- and ^{13}C-n.m.r. analysis[36,37] and large databases of ^{1}H- and ^{13}C-n.m.r. data[28,38] offer different approaches to the interpretation of spectra. The present approach has the advantage of readily giving comparisons of the chemical shift data of related oligosaccharides that are important in considerations of both structure and conformation.

ACKNOWLEDGMENTS

We thank Dr. J. Feeney for collaboration on the ^{1}H-n.m.r. spectroscopy, and Dr. T. Feizi for helpful discussions.

REFERENCES

1 K. Nagayama, A. Kumar, K. Wuthrich, and R. R. Ernst, *J. Magn. Reson.*, 40 (1980) 321–334.
2 S. W. Homans, R. A. Dwek, D. L. Fernandes, and T. W. Rademacher, *Biochim. Biophys. Acta*, 760 (1983) 256–261.
3 J. Feeney, T. A. Frenkiel, and E. F. Hounsell, *Carbohydr. Res.*, 152 (1986) 63–72.
4 E. Berman, U. Dabrowski, and J. Dabrowski, *Carbohydr. Res.*, 176 (1985) 1–15.
5 G. Eich, G. Bodenhausen, and R. R. Ernst, *J. Am. Chem. Soc.*, 104 (1982) 3731–3732.
6 S. W. Homans, R. A. Dwek, D. L. Fernandes, and T. W. Rademacher, *Proc. Natl. Acad. Sci. U.S.A.*, 81 (1984) 6286–6289.
7 S. W. Homans, R. A. Dwek, J. Boyd, M. Mahmoudian, W. G. Richards, and T. W. Rademacher, *Biochemistry*, 25 (1986) 6342–6350.
8 L. Lerner and A. Bax, *Carbohydr. Res.*, 166 (1987) 35–46.
9 S. W. Homans, R. A. Dwek, J. Boyd, N. Soffe, and T. W. Rademacher, *Proc. Natl. Acad. Sci. U.S.A.*, 84 (1987) 1202–1205.
10 E. F. Hounsell, D. J. Wright, A. S. R. Donald, and J. Feeney, *Biochem. J.*, 223 (1984) 129–143.
11 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.*, 148 (1985) 367–377.

12 E. F. Hounsell, A. M. Lawson, M. S. Stoll, D. P. Kane, G. C. Cashmore, R. A. Carruthers, J. Feeney, and T. Feizi. *Eur. J. Biochem.*, 186 (1989) 597–610.
13 H. van Halbeek, L. Dorland, J. Haverkamp, G. A. Veldink, J. F. G. Vliegenthart, B. Fournet, G. Ricart, J. Montreuil, W. D. Gathmann, and D. Aminoff, *Eur. J. Biochem.*, 118 (1981) 487–495.
14 H. van Halbeek, L. Dorland, J. F. G. Vliegenthart, A.-M. Fiat, and P. Jolles, *FEBS Lett.*, 133 (1981) 45–50.
15 H. van Halbeek, L. Dorland, J. F. G. Vliegenthart, N. K. Kochetkov, N. P. Arbatsky, and V. A. Derevitskaya, *Eur. J. Biochem.*, 127 (1982) 21–29.
16 H. van Halbeek, L. Dorland, J. F. G. Vliegenthart, W. E. Hull, G. Lamblin, M. Lhermitte, A. Boersma, and P. Roussel, *Eur. J. Biochem.*, 127 (1982) 7–20.
17 H. van Halbeek, G. J. Gerwig, J. F. G. Vliegenthart, H. L. Smits, P. J. M. van Kerkhof, and M. F. Kramer, *Biochim. Biophys. Acta*, 747 (1983) 107–116.
18 G. Lamblin, A. Boersma, M. Lhermitte, P. Roussel, J. H. G. M. Mutsaers, H. van Halbeek, and J. F. G. Vliegenthart, *Eur. J. Biochem.*, 143 (1984) 227–236.
19 G. Lamblin, A. Boersma, A. Klein, P. Roussel, H. van Halbeek, and J. F. G. Vliegenthart, *J. Biol. Chem.*, 259 (1984) 9051–9058.
20 B. N. N. Rao, V. K. Dua, and C. A. Bush, *Biopolymers*, 24 (1985) 2207–2229.
21 V. K. Dua, V. E. Dube, Y-T. Li, and C. A. Bush, *Glycoconjugate J.*, 2 (1985) 17–30.
22 V. K. Dua, B. N. Narasinga Rao, S-S. Wu, V. E. Dube, and C. A. Bush, *J. Biol. Chem.*, 261 (1986) 1599–1608.
23 N. Ud-Din, R. W. Jeanloz, G. Lamblin, P. Roussel, H. van Halbeek, J. H. G. M. Mutsaers, and J. F. G. Vliegenthart, *J. Biol. Chem.*, 261 (1986) 1992–1997.
24 J. H. G. M. Mutsaers, H. van Halbeek, J. F. G. Vliegenthart, A. M. Wu, and E. A. Kabat, *FEBS Lett.*, (1986) 139–146.
25 J. Breg, H. van Halbeek, J. F. G. Vliegenthart, G. Lamblin, M.-C. Houvenaghel, and P. Roussel, *Eur. J. Biochem.*, 168 (1987) 57–68.
26 A. Klein, G. Lamblin, M. Lhermitte, P. Roussel, J. Breg, H. van Halbeek, and J. F. G. Vliegenthart, *Eur. J. Biochem.*, 171 (1988) 631–642.
27 J. Breg, H. van Halbeek, J. F. G. Vliegenthart, A. Klein, G. Lamblin, and P. Roussel, *Eur. J. Biochem.*, 171 (1988) 643–654.
28 J. F. G. Vliegenthart, L. Dorland, and H. van Halbeek, *Adv. Carbohydr. Chem. Biochem.*, 41 (1983) 209–374.
29 E. F. Hounsell, A. M. Lawson, J. Feeney, G. C. Cashmore, D. P. Kane, M. Stoll, and T. Feizi, *Biochem. J.*, 256 (1988) 397–401.
30 D. J. Wright and E. F. Hounsell, in N. Sharon, H. Lis, D. Duksin, and I. Kahane (Eds.), *Proc. Int. Symp. Glycoconjugates, Xth, Jerusalem, 1989*, pp. 135–136.
31 E. F. Hounsell, *Symposia Soc. Exp. Biol.* XLIII (1989) 149–154.
32 E. F. Hounsell, *Chem. Soc. Rev.*, 16 (1987) 161–185.
33 A. M. Lawson, W. Chai, G. Cashmore, M. S. Stoll, E. F. Hounsell, and T. Feizi, *Carbohydr. Res.*, 200 (1990) 47–58.
34 M. S. Stoll, E. F. Hounsell, A. M. Lawson, W. Chai, and T. Feizi, *Eur. J. Biochem.*, 189 (1990) 499–507.
35 CarbBank Complex Carbohydrate Structure Database and CCSD, 1989, University of Georgia.
36 D. R. Anderson and W. J. Grimes, *Anal. Biochem.*, 146 (1985) 13–22.
37 P.-E. Jansson, L. Kenne, and G. Widmalm, *Carbohydr. Res.*, 168 (1987) 67–77.
38 K. Bock, C. Pederson, and H. Pederson, *Adv. Carbohydr. Chem. Biochem.*, 42 (1984) 193–225.