READER SERVICE

A NMR-spectroscopic database of complex carbohydrate structures

J. Albert van Kuik and
Johannes F.G. Vliegenthart.

NMR-spectroscopy is a non-destructive, powerful technique for determining the primary structure of complex carbohydrate chains. Even small amounts of material can be analyzed, provided that the carbohydrate structures under investigation do not have several repeating units (polysaccharides), and that reference ^{1}H-NMR or ^{13}C-NMR chemical shifts are available. These reference NMR data are usually derived from a small collection of in-house recorded spectra, and are augmented with reference data from literature. In particular, review articles containing carbohydrate-NMR data are in high demand for this type of analysis. Unfortunately, articles include only a part of all the available NMR data, and have to be inspected manually. A fast and easy way to analyze carbohydrate NMR spectra is to use the help of the database computer program SUGABASE. This database combines carbohydrate structures and bibliographic data with ^{1}H-NMR and ^{13}C-NMR chemical shifts.

The structures and bibliographic data are taken predominantly from the CCSD (Complex Carbohydrate Structural Database), and are linked to NMR data taken from literature. The ^{1}H- and ^{13}C-NMR database can be exported as an ASCII flat file. An example of a ^{1}H-NMR record in this format is presented in Figure 1. The NMR tables and structures in the database are grouped according to the type of carbohydrate chain, namely, N-linked, O-linked, lactose-type (i.e. carbohydrate chains containing a lactose sequence) and polysaccharides (all the rest). For each nucleus, all chemical shifts are calibrated to the same reference. The ^{1}H-NMR chemical shifts are relative to acetone, resonating at 2.225 ppm, and the ^{13}C-NMR chemical shifts are relative to dioxane at 67.40 ppm. The main effort in building the database has been focused on the accumulation of ^{1}H-NMR data from N-linked and O-linked structures. The database currently contains 888 structures that are connected to ^{1}H-NMR data and 380 structures that are linked to ^{13}C-NMR chemical shifts. The program runs on IBM-compatible personal computers under MS-DOS and requires 6 Mb disk space. Figure 2 shows an example of a SUGABASE session on a PC. Recently, the program has been transferred to Silicon Graphics workstations. An example of SUGABASE running in the X-Windows environment on a workstation is presented in Figure 3.

H#: L-0501-002618
CC: CCSD:2618
TI: Assignment of the 1H- and 13C-NMR Spectra of Eight Oligosaccharides of the Lacto-N-tetraose and Neotetraose Series
AU: Strecker G; Wieruszeski J-M; Michalski J-C; Montreuil J
CT: Glycoconjugate J. (1989) 6: 67-83
SC: 8

α-D-Neup5Ac-(2-6)-β-D-Galp-(1-4)-β-D-GlcpNAc-(1-3)-β-D-Galp-(1-4)-D-Glc

MHz 400
Temp 300
Solv D2O

Residue	Linkage	Proton	PPM	J	Hz	Note
D-Glc		H-1α	5.218			
		H-1β	4.661			
		H-2α	3.574			
		H-2β	3.280			
		H-3α	3.826			
		H-3β	3.638			
		H-4α	3.637			
		H-4β	3.634			
		H-5α	3.94			
		H-5β	3.60			
		H-6	3.92			
β-D-Galp	4	H-1(α)	4.442			a
		H-1(β)	4.440			a
		H-2	3.604			
		H-3	3.728			
α-D-Neup5Ac	6,4,3,4	H-3a	1.713			
		H-3e	2.668			
		H-4	3.666			
		H-5	3.810			
		H-6	3.67			
		H-7	3.56			
		H-8	3.87			
		H-9	3.67			
		H-9'	3.86			
		NAc	2.028			

Note a)Assignments may have to be interchanged.

Figure 1.

Essentially, the database program implements two procedures for searching for data. The first procedure is used to find carbohydrate structures that match the NMR spectrum of an unknown compound. The user selects characteristic signals from the NMR spectrum of this compound and constructs a search-profile. This search profile consists of a list of chemical shifts and some parameters to control the range and sensitivity. The program uses this search-profile to find matching structural elements (sub-structures of carbohydrate structures) in the database. After the search, the selected carbohydrate structures are displayed and the matching structural elements are highlighted. The related NMR data of these structures are also shown with the matching chemical shifts highlighted.

The best approach for this type of search is to start with a small number of well-defined chemical shifts in order to gain initial insight into the type of structure under investigation. Subsequently, the number of chemical shifts can be increased to arrive at a smaller number of matching structures, which, in turn, contain larger matching structural elements. If the database contains structures that are present in the unknown compound, the results of the search narrows down

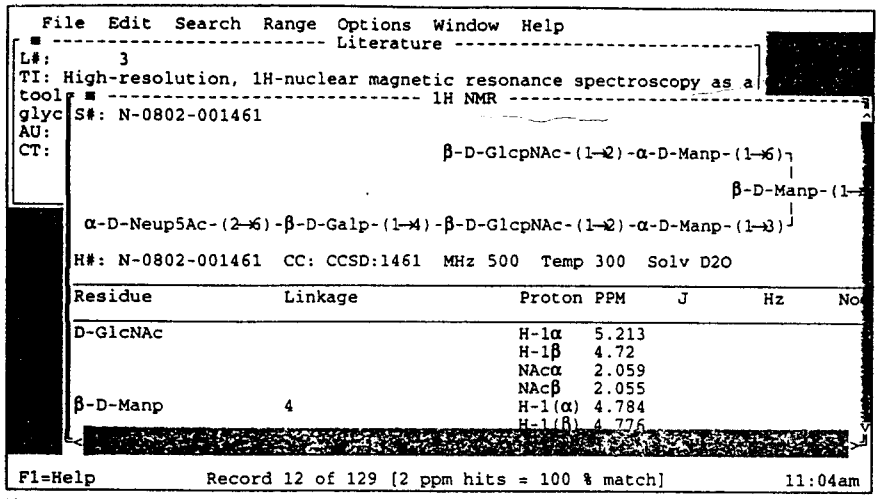

Figure 2.

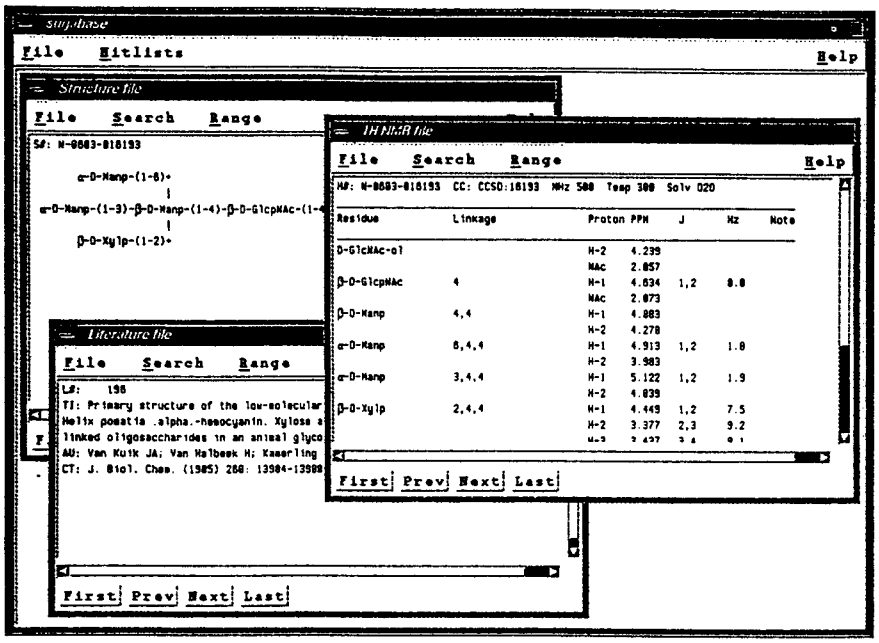

Figure 3.

to those structures. If there are no completely-matching structures in the database, the program presents a number of structures with highlighted structural-elements. These elements match parts of the unknown compound. The highlighting of these structural elements facilitates the recognition of the elements in the structure. To tune the search, the required minimum percentage of matching chemical shifts per structure for a hit and the tolerated variation of the matching chemical shifts can be defined. Moreover, an optional list of monosaccharide residues can be used to restrict the type of carbohydrate chains which are considered in a search.

The second procedure to search the database can be used to select all NMR data available for a known structural element. The user constructs a search profile that contains a structural element, and the program searches the database for all structures containing this sub-structure. These selected structures are then presented in combination with the related NMR data. Because the NMR data of a structural element are influenced by its surroundings, NMR parameters often give some information on that part of the structure that embeds the structural element.

The program can be acquired by anonymous ftp. The internet address of the ftp server is 'ruucj1.chem.ruu.nl' and the program can be found in the directory '/pub/sugabase/msdos'. The program will ask for a licence number the first time it is activated. This licence number can be obtained free of charge for non-profit organizations. Commercial organizations pay NLG 300. For organizations that can not use ftp, a copy of the program is available on diskette.

For more information contact:
Dr. Albert van Kuik
Department of Bio-Organic Chemistry,
Bijvoet Center for Biomolecular Research,
Utrecht University,
P.O. Box 80.075,
NL-3508 TB Utrecht,
The Netherlands

telephone: 31-30-533498
telefax: 31-30-540980
Email: kuik@boc.chem.ruu.nl