A NMR spectroscopic database of complex carbohydrate structures

by J. Albert van Kuik and Johannes F.G. Vliegenthart, Bijvoet Center

There is a growing interest in complex carbohydrates, a class of biologically important molecules that includes polysaccharides, glycoproteins, proteoglycans, and glycolipids. For example, complex carbohydrates play a role in the interaction of cells and proteins. From the biotechnological point of view, it is important to know the structure of the carbohydrate moiety of (recombinant) glycoproteins, because the type of carbohydrate chains are determined by the type of the cell in which the proteins are expressed. The characteristics of these glycoproteins are influenced by the different types of sugar chains.

The analysis of carbohydrate chains is a complicated task because oligosaccharides demonstrate an tremendous variety of (branched) structures, that are built from different monosaccharide residues. These residues can have different anomeric configurations and ring forms, and are connected via interglycosidic linkages at different positions. They also can contain noncarbohydrate substituents.

Nuclear magnetic resonance (NMR) spectroscopy is a nondestructive, powerful technique to determine the primary structure of complex carbohydrate chains. Often, small amounts of carbohydrate material can be analyzed, if 1H-NMR and 13C-NMR chemical shifts from the literature are used as references. However, this is a time consuming procedure, because these reference chemical shifts are not readily accessible and have to be inspected manually. Typically, review articles that contain carbohydrate- NMR data are used for this type of work, although they include only a part of all the NMR data available. A better tool for this kind of analysis is the database computer program SUGABASE. This database combines carbohydrate structures and bibliographic data with 1H-NMR and 13C-NMR chemical shifts.

Essentially, the database program has two functions. Its first function is 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. The program uses this 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 to get a first impression of the type of structure under investigation. Subsequently, the number of chemical shifts can be increased to get 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 result of the search narrows down 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 these 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 that are considered in a search.

The second function of the database is to provide all the NMR data that are 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 that contain 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 than embeds the structural element.

Currently, the database contains 815 structures that are connected to 1H-NMR data, and 372 structures that are linked to 13C-NMR chemical shifts. The program runs on IBM-compatible personal computers under MS-DOS and requires 5 MB of disk space, and 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