J. Chim. Phys. (1998) **95**, 216-220

A specialized database manager for interpretation of NMR spectra of synthetic glucides: JPD

J. Czaplicki[1] and C. Ponthus[2,*]

[1] *IPBS-CNRS, 118 route de Narbonne, 31062 Toulouse cedex, France*
[2] *SANOFI-Recherche, 195 route d'Espagne, 31036 Toulouse cedex, France*

* *Correspondence and reprints.*

RÉSUMÉ

Nous présentons ici un logiciel écrit spécifiquement pour créer et gérer une base de données spécialisées, contenant les motifs du couplage J des unités monosaccharidiques différentes. Le format de la base de données est compatible avec le format utilisé par le logiciel Aurelia/Amix de Bruker. Le logiciel facilite la recherche des motifs J inclus dans la base de données et leurs comparaisons avec un spectre expérimental, afin d'identifier les constituants de l'échantillon étudié, et ses éventuelles impuretés.

mots-clés : oligosaccharides, motifs-J, gestion de base de données

ABSTRACT

The current communication presents a program, written specifically to create and handle a specialized database, containing NMR spectral patterns of various monosaccharidic units. The program's database format is compatible with that of the Aurelia/Amix Bruker software package. The software facilitates the search for J patterns included in the database and their comparison with an experimental spectrum, in order to identify the components of the studied system, including the contaminants.

keywords : oligosaccharides, J pattern, database management

INTRODUCTION

A complete synthesis of oligosaccharides (up to 20-mer) is aimed at obtaining a substitute for heparin. Heparin is an oligosaccharide with a strong affinity for antithrombin III and is used as anticoagulant. The synthesis may involve up to 70

different stages. Under these circumstances a complete analytic control of products at each stage becomes crucial. NMR plays an important role in product characterization, allowing to obtain information on approximate sample purity, conformation of each monosaccharidic unit ($^4C_1$, $^1C_4$, $^2S_0$ ...), nature and position of protecting groups used and anomeric glycosidic bonds.

For each of the possible monosaccharides several hundred of structures exist because of combinatorial propagation of such factors as *(i)* different types of sugars (glucose, idose, mannose, galactose, etc.), *(ii)* different protecting groups used (several dozens) and *(iii)* position of protecting groups in a unit (1, 2, 3, 4 or 66').

The NMR parameters of interest are the chemical shifts and J coupling constants of the 5 or 7 glycosidic protons of each unit. The correlation pathways observed in magnitude mode COSY spectra, i.e. the J patterns of each unit, are unique for each structure. The complexity of spectra and the necessity to reduce the analysis time to a reasonable minimum (< 1 day) were the factors responsible for the development of a specialized database management program JPD (for **J P**attern **D**atabase). Its objective is to accelerate and facilitate the interpretation of NMR spectra of glucides.

Concept and functioning of JPD

In the absence of commercially available software exactly adapted to our needs, it was necessary to create the JPD program. It is designed to facilitate the recognition of spectral patterns. The J patterns from the database can be superposed on an experimental spectrum in order to identify the components of the studied system. This approach makes it possible to verify the presence or absence of specific substances, even on the level of contamination.

The JPD program is designed as an extension of the Aurelia/Amix Bruker software package, facilitating the use of the 'J Pattern' module of Amix. Its data format is compatible with the Aurelia/Amix format, hence files containing

descriptions of patterns can be freely exchanged between them. As a consequence it is possible to perform a semi-automatic search for the structures which match the J pattern directly on the COSY spectrum of the studied sample (Fig. 1).

The program was developed on a Silicon Graphics Indy workstation, running Irix 5.3. The graphic interface was coded in C, while the rest of the program is written in Fortran 77. The graphic interface is based on the XForms library [1], which facilitates creation of widgets (menus, buttons, browsers, etc.) and handling events under the X Windows system (Fig. 2). The program works largely in a standalone mode, relying on the operating system only in a few specific cases (data printing, memory configuration, etc.).

The general objective of the JPD program is to search the database according to an arbitrary combination of such criteria as :

(i) type of saccharidic unit,

(ii) type of subunit structure,

(iii) description (solvent, date, reference),

(iv) proton chemical shifts.

Specification of the search string may contain wildcards to facilitate database browsing. Results of a search according to one specific criterion form a subset of the data, which in turn can be searched on the basis of another criterion. Consecutive searches with different criteria lead to an arbitrarily narrow final data subset.

CONCLUSIONS

The currently used database, including more than 750 different monosaccharidic units and used on a daily basis in the NMR lab of Sanofi Toulouse, permits a significant gain of time. Further development of the software is planned, particularly in what concerns its graphic capabilities. Last but not least, we would like to profit

from the availability of such a vast database to establish empirical rules, permitting to predict chemical shifts of glycosidic protons of a given sugar unit.

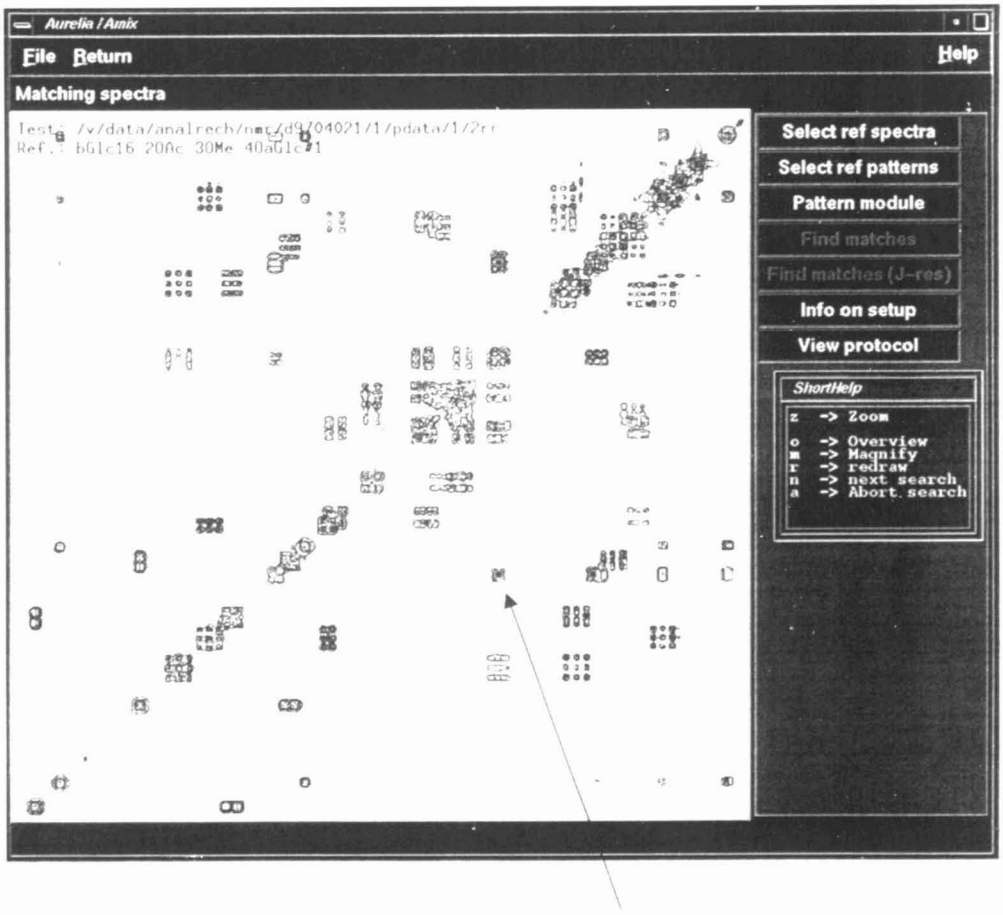

Figure 1: Example of pattern matching with Amix. Each square represents a match between an experimental peak and an entry from the JPD database.

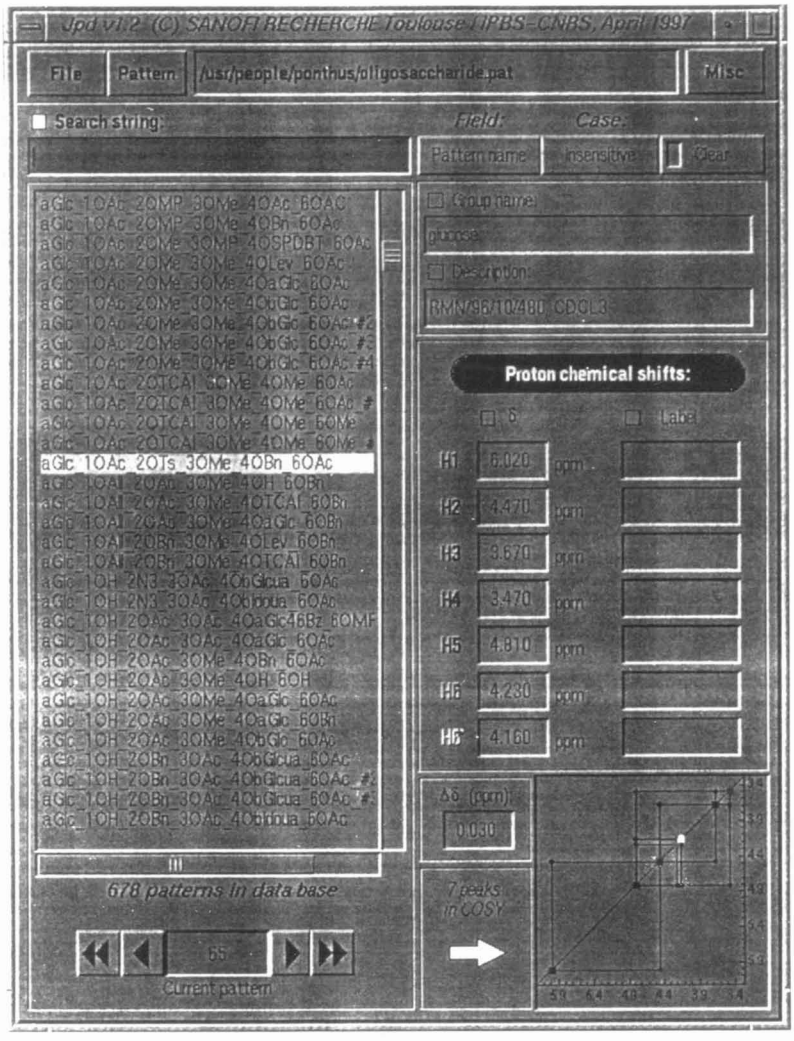

Figure 2: JPD graphic interface based on the XForms library.

REFERENCES

1 Zhao T C, Overmars M (1997) Forms Library; A Graphical User Interface Toolkit for X, v0.86. Available at *http://bragg.phys.uwm.edu/xforms*.