Database of C-Glycosylporphyrins in Web Fashion

Renato Borromei,[†] Pietro Cozzini,*[,†] Silvia Capacchi,[†] and Mara Cornia[‡]

Department of General and Inorganic Chemistry, Laboratory of Molecular Modeling, and Department of Organic Chemistry, University of Parma, Parco Area delle Scienze, 17A, I-43100 Parma, Italy

Received March 25, 2000

The aim of this work was to organize chemical data in a *client-server* environment using Database Management System and Web fashion for the client interface. To solve this *ancient* problem (for us) merging text data, reaction schemes, tridimensional structures, and NMR, CD, and UV spectra images, we have based our implementation on a few fundamental points: no cost for the user, availability of data via the Internet, standard and freeware software, and a Web browser for the database inquiry. These functions are delivered in a platform-independent manner via the Internet and are used by computational experts and nonexperts alike. C-Glycosylporphyrins is the class of compounds chosen to test our applications. These results can be exportable for many other classes of chemical compounds.

INTRODUCTION

After a few years of study, any chemical researcher has stored thousands of data on paper or in electronic files. Usually we hardly find time enough to organize all data in a way to retrieve them easily and quickly. In the past we tried to solve the problem, but the lack of time led us to develop small systems using DBMS on the PCs to collect and retrieve a small amount of information about our compounds.

Now we have decided to face the problem in a complete way building a system to manage all kinds of data collected about C-glycosylporphyrins, a class of compounds we have been studying for several years.[1] These products have no crystals suitable for the X-ray diffraction analysis, but, in the past, we have produced many structural files using molecular modeling packages, NMR analysis, and UV and CD characterization in addition to bibliographic data. In recent years porphyrinic macrocycles and particularly water-soluble porphyrins were found to be of great interest because of their wide biological and biomedical applications to cancer chemotherapy and diagnosis (NMR imaging), photodynamic therapy (PDT),[2] electron transfer, enantioselective epoxidation catalyzed by cytochrome P_{450} enzymes, and bacterial photosynthesis (Figure 1). Moreover, a few types of porphyrins appeared to exhibit antiviral activity. In addition, it is known that porphyrins assuming nonplanar macrocyclic distortions (*ruffling*)[3] would be most useful in most research about the effects of nonplanarity on light absorption, on redox, and on other properties related to selective interactions between porphyrins and biomolecules and also as a model for protein-induced nonplanar distortions of metal-tetrapyrrole cofactors.

Consequently, macrocycles showing in the same molecule either hydrophilic or amphiphilic properties and planarity deviation of the macrocycle could be of great interest in the fields of molecular recognition and medicinal chemistry.

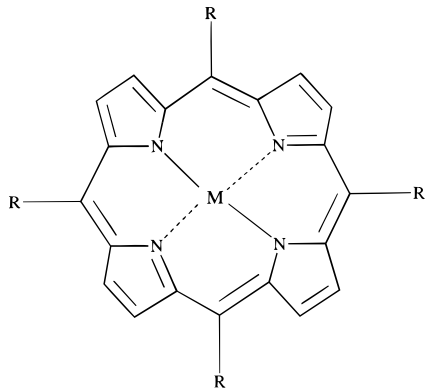

Figure 1. Schema of the porphyrins.

In our previous reports on C-glycosylated porphyrins,[1] we studied the selective synthesis and the characterization of ruffled lipophilic, amphiphilic, and water-soluble C-*glyco*-porphyrins (with and without metallic ions in the cavity), where the tetrapyrrolic macrocycle and the sugar moiety, deriving from glyceraldehyde, are connected by a strong carbon−carbon linkage. In lack of crystals, we have tried to understand the stereodisposition of the glycosidic *meso* substituents up and down with respect to the mean molecular plane (*atropisomers*) and the nonplanar distortion of the pyrrole rings (*ruffling*) also by means of NMR and molecular modeling (MM) techniques. Molecular modeling simulations agreed and confirmed our experimental data. Simulations were carried out in two different directions making, at first, a conformational analysis on the four atropoisomers synthesized, both protected and deprotected, and on the influence of the metal fitted in the macromolecular cavity and of its ionic radius on the conformation of the macrocycles.

* Corresponding author phone: +39 0521 905669; fax: +39 0521 905556; e-mail: cozzi@unipr.it.

[†] Department of General and Inorganic Chemistry, Laboratory of Molecular Modeling.

[‡] Department of Organic Chemistry.

10.1021/ci000028u CCC: $19.00
Published on Web 08/19/2000

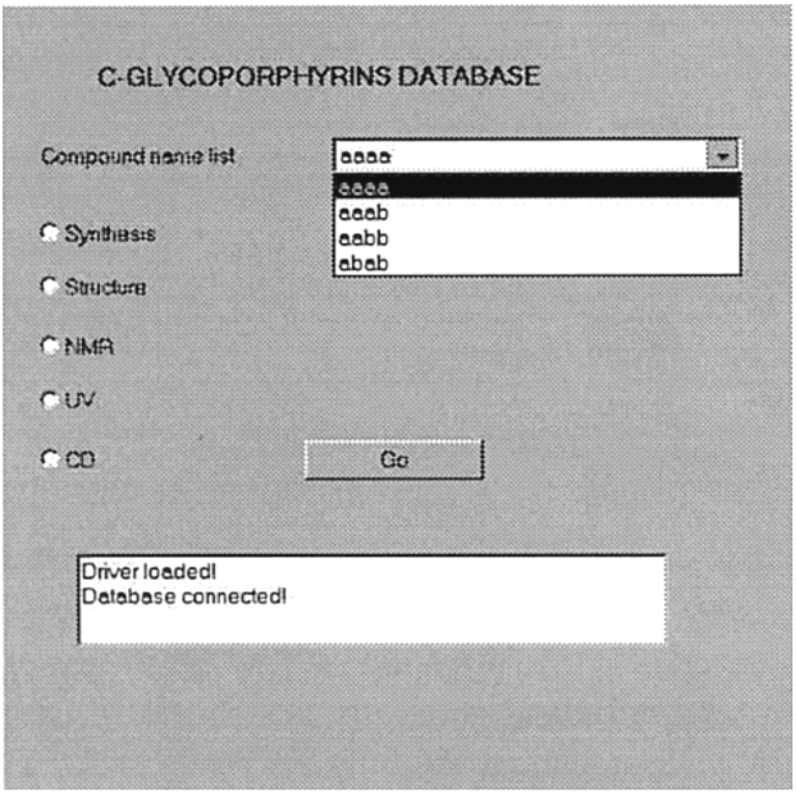

Figure 2. Graphic interface produced by applet.

COMPUTATIONAL METHODS

At the department of chemistry of the University of Parma there is a great deal of experience in the chemist community in developing software, database management systems, and Web programming. The great freedom offered by the Internet led us to use this approach to solve our problems allowing all chemists to consult our database without any investment. This choice was very important for us because the main users are organic chemists, not informatic experts, but now all our colleagues are provided with a PC on the net.

Our starting points are the following: availability of data via the Internet; search and interactions with the database made simpler; structure and implementation of programs for the query on DB made more flexible to be adapted to other chemical systems having different data structure; and no cost to manage the DB core.[4,5] To achieve the previous aims the only choice is to use HTML and Java programming that enable the development of interactive, platform-independent applications on the Web.

The simple way to share information to all researchers is to convert it in an HTML format stored on a computer running as an Internet server and to inquire about it via a Web browser running as a client on a simple PC without any other software.

On the server side we used LINUX as the operating system and APACHE as the Web server. This choice was made because this package is allowable for any operating system, freeware, and upgradable. On the side of the DBMS server we chose an SQL server, MYSQL,[6] because SQL[7] is a standard *de iure* and *de facto* for relational DBMS[8] and MYSQL is distributed for several operating systems.

On the client side we developed a Java program running as an applet within an HTML page.

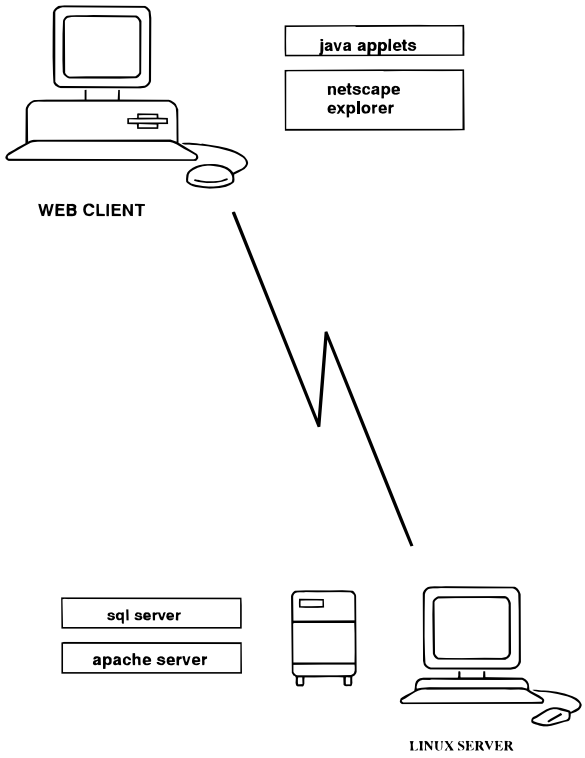

Figure 3. Flowchart of the system.

We chose Java because the code is running in any client operating system, it is simple and easy to modify on any server operating system, and it is not resident on the client but dynamically loaded at the connection allowing the query via a simple graphic interface with a SQL server.[9] The interaction between users and DB is generated by a specific Java applet which produces an interface as shown in Figure

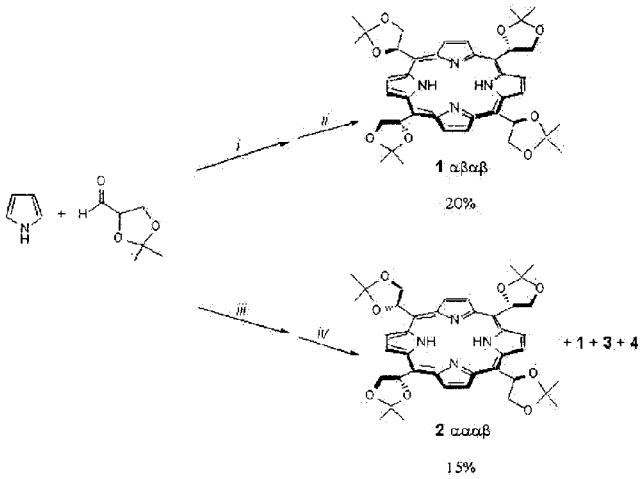

Figure 4. Bibliographic data and reaction schema: an example of numeric ingormation.

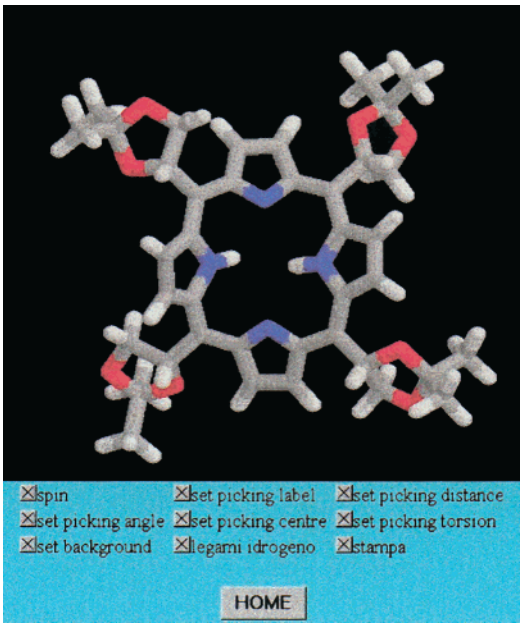

Figure 5. 3D structure of the molecule.

2. JDBC drivers are used to query the database with SQL commands; the drivers used are *type IV*, a native protocol Java driver which converts JDBC calls into the network protocol used by DBMS directly. Figure 3 shows the scheme of the links between modules needed to be installed on a client and a server.

A client base, as illustrated before, allows one to have a program divided into several classes, with a code that is easy to adapt to any chemical class of compounds and to any database structure and with a simple query on any database just modifying a little part of the code, JDBC driver. (The applet is available from the authors.)

RESULTS AND DISCUSSION

In this phase of the work we do not need sophisticated research procedures: we need access to data only in a sequential way on a list of compounds. The applet automatically sends a "SELECT * FROM" command to a SQL server producing the list shown on the screen. Obviously we are fully aware that the sequential search is the poorest query technique, but, at the moment, the most important goals to us are definition of data structures, data loading, and data consistency checking. Moreover we have decided to spend more efforts to make the data interactive. Another reason to limit the search capability to a sequential search is the low number of compounds considered in the first step of the work.

On the contrary the most important effort we have done has been to make data interactive.

Interactive data are separated in the following way:

•Bibliographic data and reaction schemes are stored in an alphanumeric way and converted in HTML format starting from .DOC format, in particular reaction schemes, found in several DOC files, needed to be converted to JPEG format (Figure 4).

•Structural data, moving from computational results, stored in PDB[9] or MOL2[10] format can be visualized using Chime MDL[11] software available for Netscape and Microsoft Internet Explorer. The strucures can be represented with several styles (stick, stick and bonds, CPK, etc.) or moved, rotated, and zoomed in any direction; geometries can be calculated as well (angles, torsions, distances, and hydrogens bonds), and moreover they can be extracted as .PDB files to be submitted to any molecular modeling package or other specific computational tools (Figure 5).

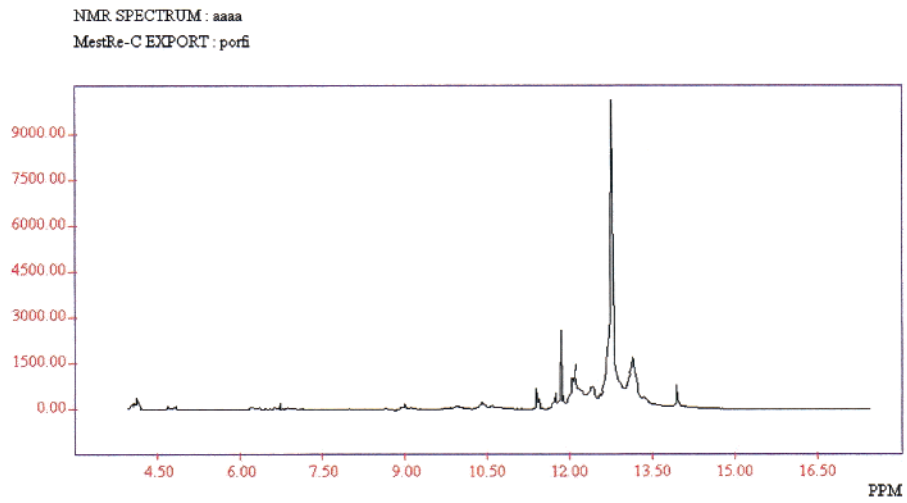

Figure 6. NMR spectrum.

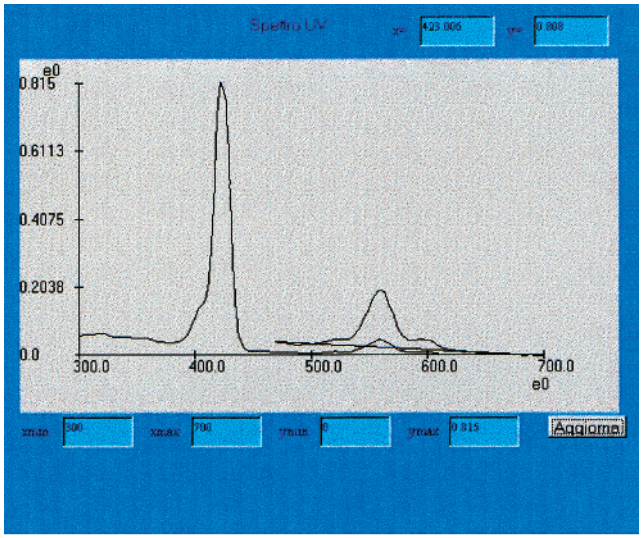

Figure 7. UV spectrum.

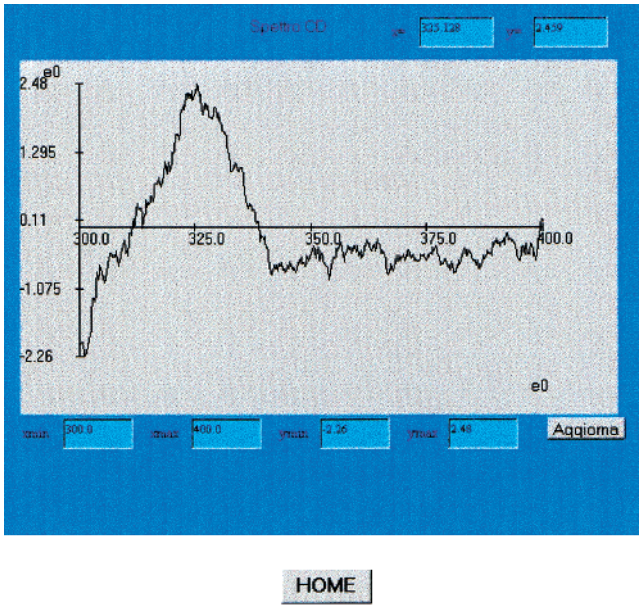

Figure 8. CD spectrum.

•NMR data, coming from Brucker spectrophotometer and stored in JDX format,[12] are visualized using Chime software as shown in Figure 6; the user can read interesting peaks or enlarging spectrum clicking with the mouse.

•UV and CD data are stored as ASCII files and visualized using applets within a Web page as shown in Figures 7 and 8; a portion of the spectra can be enlarged, and coordinates of the peaks can be read simply by clicking at the top of the peaks. In this case we use a specific applet because UV data are in ASCII format from the instrument; moreover they need to be analyzed with other our personal programs. Table 1 contains a list of file types used in Web porphyrins.

Table 1. File Types Used by Web Porphyrins

text, synthesis, bibliographic information	TXT, DOC (WORD), or HTML
NMR	JDX
CD, UV	ASCII
3D computational structures	.MOL2 or PDB

Future developments will be planned in two directions: to have more flexibility in the search using a combination of several keywords and to upgrade the system from applets to servlet approach. The peculiar phase of the next step of the work will be to implement sophisticated research techniques: the user will find bibliographic data and synthesis' schemes via string search and UV and NMR data via wavelength intervals or ppm, and the results will be related to structural data.

CONCLUSIONS

The experience of a few months with the use of this DB by organic chemists led us to continue the development of the software. Furthermore we will apply the software to another class of compounds, such as calixarenes, because many studies on crystals and in solution have been carried out on many calixarenes in our department during the last several years.

REFERENCES AND NOTES

(1) Cornia, M.; Capacchi, S.; Valenti, C.; Cozzini, P. *Tetrahedron* **1998**, *54*, 8091.
(2) Marzilli, C.; Petho, G.; Lin, M.; Kim, M. S.; Oixon, D. W. *J. Am. Chem. Soc.* **1992**, *114*, 7575.
(3) Jentzen, W.; Simpson, M. C.; Hobbs, J. D.; Song, X.; Ema, T.; Nelson, N. Y.; Medforth, C. J.; Smith, K. M.; Veyrat, M.; Mazzanti, M.; Ramasseul, R.; Marchon, J. C.; Takeuchi, T.; Goddard, W. A., III.; Shelnutt, J. A. *J. Am. Chem. Soc.* **1995**, *117*, 11085.
(4) Rzepa, H. S.; Tonge, A. P. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 1048.
(5) Rzepa H. S. *J. Mol. Struct. (THEOCHEM)* **1999**, *463*, 217.
(6) Yarser, R. J.; Reese, G.; King, T. In *MYSQL and SQL*; O'Reilly, Ed.; 1999.
(7) *SQL*; IBM DB2, User's Manual Version 2.1.1.
(8) Reese, G. In *Database Programming with JDBC and JAVA*; O'Reilly, Ed.; 1998.
(9) *PDB*; Protein Data Bank, Brookhaven National Laboratories. http: www.rcsb.org/pdb.
(10) *SYBYL*; User's Manual, Version 6.5, Tripos Inc.: 1998.
(11) *Chime 2.0.3*; Win95, MDL Information Systems Inc.: 1999; http: www.mdli.com/support/chime.
(12) *Brucker*; User's Manual, Version 97.01.01.3.

CI000028U