

## WWW GUIDE

# Dissecting complex carbohydrates with “virtual” tools

Edward McSweegan

Most medically important proteins are actually glycoproteins—proteins coated with varying amounts of sugar molecules. Sugar-coated proteins include antibodies, growth and clotting factors, blood group antigens, hormones, and cell receptors. Carbohydrates are important in cell-to-cell recognition and communication, microbial virulence, and host immune responses.

Despite their obvious importance in health and disease, pharmaceutical research on carbohydrates has lagged behind that on proteins and nucleic acids. Part of the reason

for this is the diversity and complexity of many large carbohydrate molecules. Proteins and nucleic acids are essentially linear molecules and their individual building blocks are strung together like beads on a string. In contrast, carbohydrates are complex in structure and highly branched.

Over the past few years, several companies have introduced products for automating the analysis and synthesis of complex carbohydrate structures. These new tools have opened the previously esoteric world of carbohydrate chemistry to greater numbers of researchers.

At the same time, a growing number of on-line databases and analytical software are giving researchers high-speed access to the

kinds of information and tools that users of GenBank and other DNA/protein databases have relied on for years. Some of these World Wide Web resources are also providing “how to” information about hundreds of previously synthesized sugar bonds, neural network estimations of glycosylation sites on proteins, and structural analyses of carbohydrates.

Internet-based resources for carbohydrate chemists do not yet constitute the kind of “virtual laboratory” that is available on the Web for protein and DNA researchers. Yet the Web’s expanding capacity for research and communication, and the belated recognition of the commercial and medical importance of carbohydrates, suggests that such a capacity is not too far in the future.

---

Edward McSweegan is a science writer based in Maryland ([edwardmc@qis.net](mailto:edwardmc@qis.net)).

## A sampler of sites for working with complex carbohydrate structures

### Databases

#### *The Complex Carbohydrate Structure Database*

<http://bssv01.lanccs.ac.uk/gig/pages/gag/carbbank.htm>

CCSD is the carbohydrate equivalent of GenBank. Also known as CarbBank, it is a database of carbohydrate structures and related information that can be searched by accession number, author, date, structure, and other criteria. Software for accessing CCSD (CarbBank 3.1 for Windows) also allows researchers to create their own databases and records that can be submitted to the CCSD. CarbBank software runs under Windows 95 and is available from the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>) and at the ftp site provided in the URL shown above.

#### *Sugabase*

<http://www.boc.chem.ruu.nl/sugabase/databases.html>

An experimental Web interface to the CCSD. The interface allows users to run simple searches for carbohydrate structures and nuclear magnetic resonance (NMR) data using Java or HTML forms.

#### *University of Alberta Glycobiology Group*

[http://glyco2.chem.ualberta.ca/MISC/storage/glycosyl\\_top.html](http://glyco2.chem.ualberta.ca/MISC/storage/glycosyl_top.html)

This searchable index of glycosidic bonds and linkage reactions is a useful tool for answering the question, “Has anyone synthesized this yet?” Over 700 synthesized glycosidic linkages have been culled from the 1994 literature and tabulated at this site. Information on reactants, product size, anomeric configuration, and yield are provided. Only glycosidic linkages formed between individual sugar molecules are reported.

### Tools

#### *The Complex Carbohydrate Research Center*

<http://www.ccrcc.uga.edu>

CCRC at the University of Georgia supports an on-line neural net recognition system (CCRD-Net) for the research community. The automated recognition system identifies gas chromatography (GC) and NMR spectra from submitted oligosaccharide samples. Currently, on-line analysis is limited to GC spectra of partially methylated alditol acetate derivatives and NMR spectra of xyloglucan oligosaccharides.

#### *Carbhyd*

<http://www.public.iastate.edu/~pedro/carbhyd/carbhyd.html>

Provides a variety of WWW links to information on carbohydrate structures, biochemistry, and related graphics tools for modeling sugar molecules. Graphic tools include the Monosaccharide Browser, which displays space-filling Fischer projections of sugar molecules, and Carblocks, which display a set of oligosaccharides in the Haworth Projection. Links to CarbBank and Sugabase are also included.

#### *NetOGlyc*

<http://www.cbs.dtu.dk/services/NetOGlyc/>

The NetOGlyc WWW server returns probable O-glycosylation sites of mammalian mucin glycoproteins based on the predictions of a neural network. Amino acid sequences are pasted into the interface and data is returned as a gif or postscript graph. Submitted sequences are automatically deleted after site prediction in order to ensure confidentiality.

### Network links

#### *The Glycoscience Network*

<http://www.vei.co.uk/TGN/>

The Glycoscience Network provides hotlinks to various glyco-related directories and databases, including Sugabase and CarbBank. The Network Web page also links to academic centers, meetings, courses, and a hypermail service for researchers interested in carbohydrate chemistry.