

CASE STUDY

Company brings chemical database to the Internet

By Simson L. Garfinkel

Looking for methamphetamines on the Internet? How about a bit of lysergic acid? Or, for those more sublime encounters, how about a jolt

of ethanol? While individuals using the Internet to distribute these drugs are probably under investigation by the FBI, a small company

in Cambridge, Mass., founded by Stewart Rubenstein, is now making a database of these and other compounds of interest to organic chemists available on the Internet. The company is CambridgeSoft Inc., a firm that creates some of the most respected chemical drawing and modeling programs on the market. CambridgeSoft's programs run on Windows, Mac and Unix computers. But the company's new Internet chemical database is 100 percent Mac.

The Internet site plays two roles, said Jonathan Brecher, the company's Webmaster. He went to work for CambridgeSoft after graduating from Harvard two years ago with a degree in chemistry. On the one hand, the

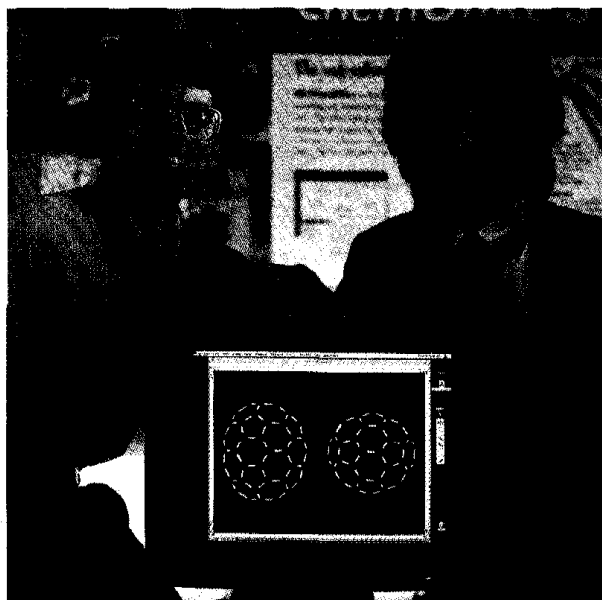
site is showing off CambridgeSoft's newest product, a World-Wide Web module for its chemical database product. But the real purpose is to encourage chemists to put their own Web servers with their own compounds on the Internet and in so doing create a global laboratory notebook for the benefit of chemists everywhere.

"Organic chemists generate a lot of intermediates," or chemicals that are slightly different than the particular compound that they are investigating, Brecher said. For example, a chemist trying to understand a single reaction might generate dozens or even a hundred new compounds, each with its own particular properties. These compounds can be interesting to other chemists who are working in different areas of research. "So you just make all of your compounds available on the Web, and people can search the universe of chemical compounds for structures that they are interested in, and they may find new uses of the compounds," he said.

Nowhere to go

Until this summer, chemists wanting to put their compounds on the Internet had a problem. Although they could simply publish their papers, there was no simple way to publish structural information about chemicals — the actual 2-D and 3-D representation of atoms and bonds. Just as important, there was no way to search a Web database by chemical structure or property. That's important, since a typical complex chemical can have many different names, all of which are equally accurate.

This is what is so revolutionary about CambridgeSoft's new site. The site is based on CambridgeSoft's ChemFinder Pro, a database



SETH RESNICK

CambridgeSoft founder Stewart Rubenstein (right) and Webmaster Jonathan Brecher (left) studied chemistry at Harvard and share the idea of making compounds information available on the Web.

that understands the laws of chemistry and can retrieve chemicals based on their chemical properties, structure or substructure.

To build the site, Brecher took an old Mac IIci that was in the office, put an Ethernet board in the back of the computer and downloaded a copy of WebStar from StarNine Technologies Inc.

In less than five minutes, the Web server was up and running. "Literally, I downloaded the software, and I uncompressed it, launched the program and that was it," Brecher said.

Next came the process of integrating WebStar with ChemFinder Pro. The task proved to be quite easy, Brecher said, because ChemFinder Pro supports AppleScript, and WebStar can use AppleScript to control remote programs and report the results back to users over the Internet.

One of the problems that Brecher had to deal with was developing a way for sending chemical structure information over the Web. Chemical structure, after all, isn't one of the primary data types that's built into the HTML (Hypertext Markup Language) standard.

All SMILES

What Brecher ended up doing was using a simple code called SMILES (Simplified Molecular Input Line Entry System), which allows chemists to represent structural information as text. To do a search by structure, a chemist would first draw the compounds on his own computer, have the computer determine the SMILES code and then copy the code into a search field on the Web page. Structures that are found during the course of a search are similarly displayed by SMILES

See Case study, Page 42

Case study

From Page 40

code. CambridgeSoft is working on free helper applications, which people will be able to download from the company's FTP site, to display the structures directly.

Not ready for prime time

By mid-September, the site was in beta testing. At first, the results were less than exciting.

The problem was that people kept searching for compounds that weren't there. It's the same problem that the company has had with a sample database that it distributes on CD-ROM. "What subset of 13 million compounds

.....
CambridgeSoft's chemical collection is getting bigger, but the goal of this project is to let research chemists publish their own notes and findings using the Internet.
.....

do you fit on a CD-ROM?" Brecher said.

Fortunately, it's much easier to update the Web site than to remaster a compact disc. Throughout the summer, Brecher monitored all searches performed over the Internet and added the compounds as they were requested. Eventually, he said, he hopes to have a large working subset that contains many of the most popular compounds.

Although CambridgeSoft's chemical collection is getting bigger, the eventual goal of this project is to allow research chemists all over the world to publish their own notes and findings using the Internet, much in the way that other professions have already started doing.

Reserving praise

Robert G. Salomon, a professor of chemistry at Case Western Reserve University in Cleveland and a devoted ChemDraw user, agreed with Brecher's vision but was somewhat reserved in his praise.

"The substructure-searchable ChemFinder Web server is a work in progress, he said. "Presently, it contains structures and their names, information of little interest to me. But that's not the point. The site demonstrates the feasibility of transmitting structure information through the Web ... and searching Web sites containing ChemFinder databases in various ways, including the powerful substructure method."

Salomon said he anticipates a future in which many laboratories and commercial sources are making chemical structure information on the Web. But in order for the final network to be usable, he said, it will be necessary for somebody to write a master index program, similar to Carnegie Mellon University's Lycos search engine. "If enough people who visit the page request certain features, CambridgeSoft will probably figure out how to provide them." □