

UC Irvine

UC Irvine Previously Published Works

Title

Electronic publishing: Protein science at the edge of a revolution

Permalink

<https://escholarship.org/uc/item/4m30v7hn>

Journal

Protein Science, 3(11)

ISSN

0961-8368

Author

White, Stephen H

Publication Date

1994-11-01

DOI

10.1002/pro.5560031101

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

EDITORIAL

Electronic Publishing: *Protein Science* at the Edge of a Revolution

Many years from now scholars will look back at our century in search of its pivotal events. Two of those events will almost certainly be the emergence of the World Wide Web (WWW) conceived of at CERN in 1990 (Berners-Lee et al., 1994) and the creation of the computer software Mosaic introduced in 1993 by the National Center for Supercomputing Applications (NCSA). These events are merging into a single phenomenon: WWW/Mosaic. What is this phenomenon, why is it so special, and why should protein scientists care about it? A thorough and readable account of WWW/Mosaic has been presented in a recent issue of *Science* by Schatz and Hardin (1994). Briefly, the "Web" is a collection of computer servers (data repositories) and computer clients (users) connected together by the Internet. Mosaic is a hypermedia "browser" for retrieving and viewing data from the global network of servers. Invented by Marc Andreessen at NCSA, Mosaic presents to the user a document with nicely formatted text containing in-line images and hyperlinks to other images, sounds, and movies as well as text documents. The result of the WWW/Mosaic innovation is a rapidly expanding global hypermedia library of immense size. By means of a few clicks of the mouse on a computer, one can view the latest satellite weather maps, listen to recorded interviews, tour the Library of Congress "Scrolls from the Dead Sea" exhibit, or examine molecular graphic images of proteins. The quality, quantity, and variety of information available and the phenomenal growth in WWW/Mosaic usage suggest that we are at the dawn of a revolution in scientific communication comparable to the printing revolution of the fifteenth century.

Protein Science is an active participant in this revolution. With the help of Louise Falevsky, the electronic publishing systems specialist for the *Protein Science* Electronic Publishing Initiative, and Stephen Franklin and Lyle Wiedeman of the Office of Academic Computing, a Web server has been established at the University of California–Irvine for the eventual publication of *Protein Science* over the Internet. The simultaneous publication of the electronic and printed *Protein Science* is an important experiment that will allow the technical, administrative, and economic principles of electronic publishing to be determined and compared with those of print publishing. While that experiment goes forward, we have endeavored to make the *Protein Science* Web server useful to the protein science community as a source of information about protein science and about WWW/Mosaic. Included in this issue of *Protein Science* are detailed instructions for contacting the Web server and an overview of WWW/Mosaic prepared by Dan Jacobson of the Johns Hopkins BioInformatics Project.

The principle of WWW/Mosaic is simple. Servers and clients communicate by WWW protocols that tell how to specify the location of documents on the Web, how to form hyperlinks among documents, and how to request and send documents. The location of documents and the protocol the client

should use to retrieve them are specified in a Uniform Resource Locator (URL). For example, to reach the default document (home page) on the newly established *Protein Science* Web server, the URL is <http://www.prosci.uci.edu/> where http indicates that the Hypertext Transfer Protocol is to be used to retrieve the document located at the Internet address www.prosci.uci.edu in the default directory of the server. Because no particular document is specified, the home page will be delivered. The URL <http://www.prosci.uci.edu/kinemages/kinpage.html> will take the client to the subdirectory kinemages and the document kinpage.html which, in this case, will provide access to all kinemages published in *Protein Science* to date. The Hypertext Transfer Protocol is used when the document to be delivered has been prepared using the Hypertext Markup Language (HTML). This type of document is written in plain text and contains "marks" such as and (which mark the beginning and end of bold text). Most important, HTML documents contain hyperlink marks containing URLs of documents on local or remote computers. It is these hyperlinks that allow a particular computer to access vast amounts of information without storing it locally. Furthermore, the technology exists for indexing the information so that full-text searches over the Web are possible. One of the important innovations of WWW/Mosaic is its support of all of the earlier Internet protocols, including telnet, ftp, and gopher so that if the URL specifies, say, the gopher protocol (gopher://) instead of http, then Mosaic will retrieve and display data from a gopher server.

A particularly important feature of Mosaic is that it utilizes Multipurpose Internet Mail Extensions (MIME), which is a protocol for attaching special files to ordinary electronic mail messages. If you examine the header on your e-mail messages, you will probably find the statement "Content-Type: TEXT/PLAIN". This is a MIME specification: file type = text, subtype = plain. Another content type is IMAGE/TIFF, meaning a binary image file of the type "tagged image format file". The MIME type tells the receiving computer how to treat the file. If Mosaic retrieves a file with the extension *.tiff, for example, it will launch a viewer program and load the tiff file into it. Virtually any computer program that uses a specific type of file with a MIME specification can serve as a viewer. It is this capability that makes WWW/Mosaic especially valuable to protein scientists. For example, the MAGE program that is now well known to readers of *Protein Science* can be used as a viewer for viewing kinemages on-line over the Internet. Full instructions for configuring MAGE to work in concert with Mosaic and other Web browsers such as Cello, WinWeb, MacWeb, and Netscape are provided on the *Protein Science* Web server. But MAGE is only one of many possible programs that will be of value to protein scientists using molecular simulation and visualization software. Henry Rzepa of Imperial College, London, is leading an effort to establish a chemical MIME type to stan-

dardize the handling of molecular coordinates and other chemical information over the Web. The kinemage is one of the proposed chemical subtypes. The *Protein Science* Web server has been configured to recognize the content type "chemical/x-kinemage" with file extension *.kin.

Johann Gutenberg's invention of printing by movable type revolutionized the transmission of information among scholars. The shift in the production of technical literature from scriptoria to printing shops made possible the precise reproduction not only of letters and numbers but of images as well. Elizabeth Eisenstein (1983) has suggested that the new forms of interplay among letters, numbers, and images made possible by the printed book were perhaps the most important aspect of the printing revolution. WWW/Mosaic has precisely this same quality. We are no longer constrained to view computer documents as text and simple graphs. We can now combine sound, images, text, and a host of other objects into a single hypermedia "document". The document, however, is no longer a printed page or a floppy disk. Rather, it is collection of hyperlinks to text and binary objects located on computers all over the world. This concept will alter forever our ideas about libraries and the manner in which we organize information. It is the beginning of the fulfillment of Vannevar Bush's (1945) dream of "a sort of mechanized private file and library" for individual use. In his prescient

article "As We May Think", he wrote: "If the user wishes to consult a certain book, he taps its code on the keyboard, and the title page of the book promptly appears before him, projected onto one of his viewing positions. . . . A special button transfers him immediately to the first page of the index. Any given book of his library can thus be called up and consulted with far greater facility than if it were taken from a shelf." With the advent of WWW/Mosaic, it appears that Vannevar Bush's future world has arrived.

References

- Berners-Lee T, Cailliau R, Luotonen A, Nielsen HF, Secret A. 1994. The World-Wide Web. *Commun ACM* 37:76-82.
Bush V. 1945. As we may think. *The Atlantic* 176(1):101.
Eisenstein EL. 1983. *The printing revolution in early modern Europe*. New York: Cambridge University Press.
Schatz BR, Hardin JB. 1994. NCSA Mosaic and the World Wide Web: Global hypermedia protocols for the Internet. *Science* 265:895-901.

STEPHEN H. WHITE

*Department of Physiology and Biophysics
University of California
Irvine, California 92717-4560*

How to Contact the Electronic *Protein Science*

Protein Science maintains WWW, Gopher, and FTP servers. The WWW server provides access to all *Protein Science* Internet resources.

To contact the WWW server: From the Mosaic menu bar, select "Navigate" and then "Files . . .". You will then be asked for the so-called WWW URL. Enter the line "http://www.prosci.uci.edu/". After you have contacted the *Protein Science* Web server, add it to your "Hot List" so that you can contact it later by a few clicks of your mouse.

To contact the Gopher server: Send the command "gopher ftp.uci.edu 1071". This will place you at the top of the University of California at Irvine's anonymous FTP directory. Select the item "Protein Science Published by the Protein Society".

To contact the FTP server: Send the command "ftp ftp.uci.edu" followed by a carriage return. Log on as "anonymous" using your e-mail address as password. Send the command "cd protein" followed by a carriage return to change to the *Protein Science* FTP directory.