BioEditor—simplifying macromolecular structure annotation

Peng Yang¹, Paul A. Craig², David Goodsell³ and Philip E. Bourne¹, 4, *

¹San Diego Supercomputer Center, University of California, San Diego, 9500 Gilman Drive, MC 0505, La Jolla, CA 92037-0505, ²Rochester Institute of Technology, Department of Chemistry, 85 Lomb Memorial Drive, Rochester, NY 14623, ³The Scripps Research Institute, Department of Molecular Biology, 10550 N. Torrey Pines Road, La Jolla, CA 92037 and ⁴Department of Pharmacology, University of California, San Diego, 9500 Gilman Drive, MC 0505, La Jolla, CA 92093

Received on September 9, 2002; revised on December 13, 2002; accepted on December 26, 2002

ABSTRACT
Summary: BioEditor is an application to enable scientists and educators to prepare and present structure annotations containing formatted text, graphics, sequence data, and interactive molecular views. It is intended to bridge the gap between printed journal articles and Internet presentation formats. BioEditor is relevant in the era of structural genomics, where annotation and publication could become the rate determining step in structure determination.

Availability: BioEditor is available at http://bioeditor.sdsc.edu. The Web site includes the latest version of the software for Microsoft Windows, including documentation, the opportunity to submit bug reports and suggestions, example documentaries prepared with BioEditor and a repository where users can submit documentaries for posting to the site.

Contact: bourne@sdsc.edu

INTRODUCTION
When translating biomolecular data into text and static figures for print formats, much of the information that is available must be discarded or radically simplified. Electronic publication allows a much tighter integration of the data than is available with the information that is disseminated. Electronic documents allow readers to explore the information on their own, in ways that the authors may not have predicted.

In regard to macromolecular structures, kinemages, developed by David and Jane Richardson, were an early development to utilize this concept in the study of structure-function relationships by allowing authors to present and annotate atomic structures (Richardson and Richardson, 1994). Other popular molecular visualization tools including Chime (http://www.mdli.com), Rasmol (Sayle and Milner-White, 1995), DeepView (Guex and Peitsch, 1997), WebMol (Walther, 1997) and Protein Explorer (Martz, 2002) have been used for the same purpose. Hence skilled users have created excellent resources on the Internet. The IMB JENA Image Library of Macromolecules (http://www.imb-jena.de/IMAGE.html, Reichert et al., 2000) and the Online Macromolecular Museum (OMM; http://www.clunet.edu/BioDev/omm/gallery.htm) are examples. Moreover, detailed instructions exist on preparing online resources (Cummer, 2000; Horton, 1999). However, doing so usually requires the ability to write HTML and perhaps JavaScript (Horton, 1999). In short, while very useful, these tools represent a barrier to the content provider intent on a detailed explanation of structure-function relationships.

As a potential means to overcome this barrier we describe BioEditor, an XML-based online-publication method for presenting general biomolecular information. To simplify the preparation of ‘structure documentaries’, BioEditor’s user interface is designed to look and feel like Microsoft Word with appropriate extensions. To fully utilize BioEditor’s interactive structure display tools users must become familiar with the scripting language associated with Chime, but the remaining functionality will be familiar and comfortable for people who consider themselves computer users rather than programmers. Hence, BioEditor is designed for a wide target audience. It allows crystallographers and structural biologists to create technical electronic documents describing structures they have determined. In addition to information normally found in journal articles, additional views and structural animations can be included. BioEditor is also a convenient tool for collecting, organizing and disseminating information compiled from a bioinformatics perspective, for example, relating sequence, structure and function.
within a protein family. Those engaged in structural genomics may find it a useful way to annotate structures and perhaps use it as a form of electronic notebook of each structure under study. At the college level, both students and faculty may prepare structure documentaries as part of the teaching and learning process.

SYSTEMS AND METHODS
Collecting Information
Users collect a variety of data types in BioEditor and organize them into folders, which are patterned after standard graphical user interfaces. Each BioEditor document contains six folder types: authors, topics, structures, sequences, graphics and references. These data are stored in two locations by BioEditor—an XML document and its associated data folder. Imported graphics and PDB structure files are stored in the data folder. The XML document contains the remaining information (author, topic, sequence and reference) as well as information describing any molecular views that have been created. One of the major advantages of BioEditor is that all the data are accumulated in an organized fashion in one location.

Interface
BioEditor is written using Visual Basic for the Windows environment, with standard drop-down menus and toolbars. In addition to standard menu functionality (file, edit, window, help), it has menus (view, format, insert) which have common names, but contain unique functions. For example, the insert menu enables users to insert sequence information or 3D molecular views into a topic. Selections in the structure menu will open well-known structural biology databases in the browser window in BioEditor and provide tools for molecular animation. The sequence menu provides access to pertinent protein and DNA sequence databases.

Creating Molecular Animations
BioEditor incorporates the Chime plug-in (http://www.mdl.com) for molecular visualization, using the commands associated with the Chime scripting language. Once a structure (in the form of a PDB file) has been loaded into BioEditor, available menus and toolbars enable the author to create a molecular animation of the structure, using all the tools and commands that are part of Chime.

Integrating Data with Text
Text, graphics, sequence and molecular animations are all integrated into topics within BioEditor. Once a topic has been created, any of these other data types can be inserted anywhere in the text. For example, selecting 3D view from the insert menu (or the related icon in the insert toolbar) opens a dialog box that contains a listing of all the 3D views that have been generated by the user. Similar dialog boxes are available for inserting graphics and sequences into a topic.

The BioEditor help file contains detailed descriptions of the software and a step-by-step tutorial on preparing structure documentaries.

DISCUSSION
Future goals include, first, a version written in Java to support multiple operating systems, including the Macintosh operating system. Second, removal of the dependency on Chime as a molecular viewer by using a Java3D toolkit that is currently in the alpha stage of development (see http://mbt.sdsc.edu). Finally while much effort is devoted to the analysis of free text to extract useful knowledge, authoring in tools like BioEditor provides a level of organization and data typing not present in conventional word processing documents which are just large blobs of free text. In the long term we are seeking to exploit this relationship for a new generation of structure-function databases, this will involve providing an XML specification which is compatible with other overlapping efforts worldwide.

ACKNOWLEDGEMENTS
We would like to thank the National Science Foundation for their support (Craig—NSF DBI-0132856; Goodsell, Yang, Bourne – NSF DBI 9904796).

REFERENCES