Pacific Symposium on Biocomputing 4:338-340 (1999)

MOLECULES TO MAPS: TOOLS FOR VISUALIZATION AND INTERACTION IN COMPUTATIONAL BIOLOGY

THOMAS E. FERRIN
Computer Graphics Laboratory
University of California
San Francisco, CA 94143-0446

EILEEN T. KRAEMER
Department of Computer Science
The University of Georgia
Athens, GA 30602

Data continues to accumulate rapidly from the various genome projects and from experimental methods such as X-ray crystallography, NMR spectroscopy, and electron and confocal microscopy. The vast volume of sequence, structural, and functional data, the wide variety of analyses and annotations to be performed, and the variety of organisms and projects represented in this deluge of information combine to present the bioinformatics community with unprecedented challenges. This conference track explores challenges of enormous importance to the molecular biology and structural biology communities: tools and techniques to assist scientists in evaluating, absorbing, navigating, and correlating this sea of information, especially through visualization and user interaction.

The papers in this track fall into roughly three areas: tools for performing specific tasks, techniques for improving the productivity of tool developers, and multimedia approaches for presenting results. All represent powerful visualization techniques that can help the scientist user gain an understanding of an otherwise large volume of nearly incomprehensible data. The human mind can often quickly grasp spatial relationships when presented in visual, and now also auditory, form through models, but if left in their original form, columns upon columns of numerical data will have little meaning to even the most astute scientist.

Meads, Hansen, and Pang describe a visualization tool, ProtAlign, for the display of low resolution protein structures and the editing of protein sequence alignment directly via the three dimensional structure. Low resolution protein structures, that is, structures which do not take amino acid side chain position into account, are frequently found in protein threading work. Here an alignment is made between a known structure and a target sequence, with the aligned portions of the target sequence backbone placed in the same orientations as the corresponding backbone segments of the known structure without regard to side chain position.

Existing molecular visualization software typically displays models using an all-atom approach, while ProtAlign utilizes a stylized approach to displaying side chains based on depictions of multishaped building blocks, with similar shapes representing compatible residues. Furthermore, ProtAlign allows the user to conveniently modify the alignment of the two protein sequences by direct manipulation of the three dimensional model instead of forcing the user to move back and forth between a two-dimensional representation of the sequence alignment and a three-dimensional representation of the structure. These and other features of ProtAlign make it an innovative and useful application to the protein threading community.

Lawton *et al.* also use a low resolution approach in their program Proteinmorphosis, a physically based modeling system for simulating large or small conformational changes of proteins and protein complexes. By using a macroscopic approach to the problem, coupled with a novel parameterization of the degrees of freedom in terms of global translations and rotations of nodal atoms, they achieve results which closely match the conformational changes observed in two significant experiments: that of a calmodulin-peptide complex which undergoes significant tertiary conformational change upon binding, and the rearrangement of the quaternary structure of hemoglobin during cooperative oxygen binding. Their findings support the notion that such simplified approaches are reasonable to an otherwise very computationally intensive problem.

Metabolic pathways are of critical importance to molecular biologists, since biological function is the ultimate role of molecules in life. The creation and visualization of metabolic pathway diagrams has always posed a challenge to researchers, however, since such diagrams are often complex and subject to change as new knowledge is gained. Salamonsen and his collaborators describe a program, BioJAKE, that simplifies the process of creation, manipulation, and visualization of metabolic pathways. The program provides an easy mechanism for constructing pathways from scratch, or automatically from information stored within databases. It also has the ability to compute the reactions in which any given molecule is involved in, or the sequence of reactions required for a specific pathway to take place. BioJAKE can also perform remote database queries on particular molecules within a pathway based on a list of databases and query parameters coded into the program, with results displayed in a web browser window. BioJAKE is written in Java, and hence portable to many systems.

The dedicated tools described above represent a tremendous software development effort. Sanner *et al.* describe timely work on improving the productivity of their programming environment through use of the Python programming language and the AVS computation and visualization system. Python, an interpreted, object-oriented programming language available for a large variety of

platforms, was designed from the beginning for extending and embedding in other applications. Sanner *et al.* found that using Python ''wrappers'' around other existing computational tools increased the flexibility and interoperability of these tools and significantly increased the ease with which these applications could be incorporated into AVS dataflow networks. Moreover, the process of writing customized AVS modules, previously restricted to highly skilled programmers, was considerably simplified by the use of Python. Several examples are provided in their paper as evidence of the benefits of using Python.

Two papers in this track have the word 'multimedia' in their title, illustrating how important this approach has become for perception and learning. Hansen *et al.* describe an innovative application called PROMUSE that makes use of both visualization and sonification (sound) to perceive protein structure alignment. Sonification presents several opportunities to researchers. For those with visual impairment, data sonification can be a useful alternative to visualization. Sonification can also be used to increase the amount of information being conveyed to the user or to disambiguate otherwise visually cluttered displays. Hansen and colleagues describe a series of experiments designed to evaluate the use of sonification in discriminating four parameters found in protein structural alignment comparisons: secondary structure, polarity, exposure, and goodness-of-fit. Their experimental results found that overall accuracy scores were much higher when sound, or sound and visualization were used to depict these parameters compared to visualization alone. The sounds files used in this work are downloadable via the web.

If you have ever wondered how to go about creating a web-based HTML, sound, and video multimedia presentation, then the Quinn *et al.* paper entitled ''Developing Protein Documentaries and Other Multimedia Presentations for Molecular Biology'' is for you. The authors describe a variety of tools and techniques that represent the cutting edge of multimedia content-authoring technology, and provide seven examples that illustrate various approaches. URLs are provided so that you can download each example (you'll also need to download a number of free plug-in applications for either Netscape Navigator or Internet Explorer), and you'll have lots of fun experiencing the presentations.

Together this collection of papers provides an excellent sample of ongoing research activities in software tool development for visualization and interaction in computational biology.