ModView, visualization of multiple protein sequences and structures

Valentin A. Ilyin, Ursula Pieper, Ashley C. Stuart, Marc A. Marti-Renom, Linda McMahan and Andrej Sali*

The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA

ABSTRACT

Summary: We describe ModView, a web application for visualization of multiple protein sequences and structures. ModView integrates a multiple structure viewer, a multiple sequence alignment editor, and a database querying engine. It is possible to interactively manipulate hundreds of proteins, to visualize conservative and variable residues, active and binding sites, fragments, and domains in protein families, as well as to display large macromolecular complexes such as ribosomes or viruses. As a Netscape plug-in, ModView can be included in HTML pages along with text and figures, which makes it useful for teaching and presentations. ModView is also suitable as a graphical interface to various databases because it can be controlled through JavaScript commands and called from CGI scripts.

Availability: ModView is available at http://guitar.rockefeller.edu/modview.

Contact: ilyin@neu.edu, sali@rockefeller.edu

Large numbers of gene sequences, protein 3D structures, and functional annotations are being determined individually as well as by large-scale efforts such as genome sequencing, structural genomics, functional genomics, and bioinformatics. The data are stored in an increasingly large variety of rapidly growing databases, and have to be organized, displayed, analyzed, understood, and exploited. Thus, there is a need for a publicly available program that allows web access to an integrated display and analysis of multiple protein sequences and structures, obtained efficiently from the sequence and structure databases. We outline here our initial attempt to create such a program, called ModView. A number of other tools for displaying protein sequences and structures are listed at http://guitar.rockefeller.edu/modview.

ModView provides an interactive display of single and multiple protein sequences and structures, and facilitates a visual analysis of relationships between them. Several ModView capabilities are illustrated by a sample snapshot of ModView windows displaying a large set of globin

*To whom correspondence should be addressed.

sequences and structures (Figure 1). ModView has four main windows.

The structure window of ModView displays multiple 3D structures of proteins, nucleic acids, and small ligands. Structure visualization is based on the program MIPA (Ilyin, 1994), with graphical rendering and scripting from RasMol (Sayle and Milner-White, 1995). There are a large number of options that control the display of structures, including atom type selections, presentation styles, and coloring schemes. It is also possible to superpose multiple structures at a click of a button, either by structure or according to an alignment in the sequence window. Each structure or a subset of atoms can be manipulated independently from the remaining structures. Some simple calculations, such as measuring atomic distances and angles, can also be performed.

The sequence window of ModView displays multiple protein sequences. Sequence visualization is based on the sequence editor of Jalview (Clamp, 1998). Pairwise and multiple sequence alignments can be calculated automatically by ModView's dynamic programming algorithm and ClustalW (Jeanmougin *et al.*, 1998), respectively. Alignments can also be edited manually, and the sequences in the alignment can be compared with the aid of dendrograms and principal components analysis.

A convenient feature of ModView is that the sequence and structure windows interact with each other. When a residue is selected in one window, it is highlighted in the other window. When the alignment is changed in the sequence window, structures can be re-superposed in the structure window. When ModView loads sequences, it also automatically attempts to download the corresponding structures from a database of structures into the structure window; when it reads a structure, it displays its sequence in the sequence window. The coloring scheme of proteins and residues in the sequence window corresponds to the coloring in the structure window. The sequence–structure link is especially convenient when considering both sequence and structure aspects at the same time.

The control window of ModView has two panels. The top panel contains menus, buttons, and check boxes for
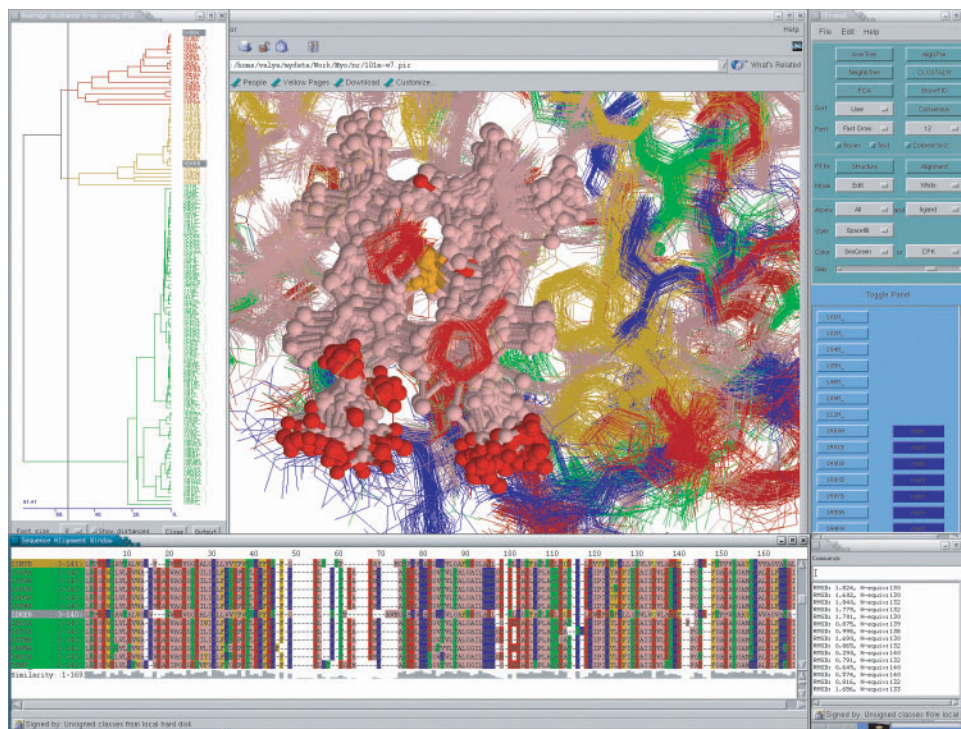
Fig. 1. Sample ModView display.

executing various commands and selecting their options, including the selection of displayed atom types, their presentation styles, and coloring. The toggle panel allows for quick selection of the displayed molecules and chains.

The command window is for typing of RasMol commands, and reports results from ModView calculations in its output panel.

As a web application, ModView can be used locally to retrieve, visualize and analyze sequence/structure information from a remote server. ModView as a visual interface to a database is illustrated by the ModBase database of comparative protein structure models (Pieper *et al.*, 2002), the LigBase database of families of ligand binding sites of known structure (Stuart *et al.*, 2002), and the DBAli database of many protein structure-based alignments (Marti-Renom *et al.*, 2001).

ModView is written in C++ and Java. As a Netscape plug-in, it can communicate with local and remote sequence and structure databases over the Internet using the CGI/HTML, JavaScript and RasMol scripting languages. It can also interact with other programs, and saves sequences, structures, and alignments in several common formats. In addition, the contents of the sequence, structure, and dendrogram windows can be printed to a PostScript file. ModView provides a relatively intuitive interface for novice users, while the CGI/HTML, JavaScript and RasMol programming interfaces allow

experts considerable flexibility. ModView is currently supported under the PC Linux operating system, although it also runs in the X11 environment on Microsoft Windows computers.

ACKNOWLEDGEMENTS

REFERENCES

Clamp,M. (1998) Jalview—a java multiple alignment editor. http://circinus.ebi.ac.uk:6543/jalview/.

Ilyin,V.A. (1994) Non-polar nuclei in fungal microbial RNases. *Protein Eng.*, **7**, 1189–1195.

Jeanmougin,F. *et al.* (1998) Multiple sequence alignment with Clustal X. *Trends Biochem.Sci.*, **23**, 403–405.

Marti-Renom,M.A., Ilyin,V.A. and Sali,A. (2001) DBAli: a database of protein structure alignments. *Bioinformatics*, **17**, 746–747.

Pieper,U. *et al.* (2002) ModBase, a database of annotated comparative protein structure models. *Nucleic Acids Res.*, **30**, 255–259.

Sayle,R.A. and Milner-White,E.J. (1995) RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.*, **20**, 374.

Stuart,A.C., Ilyin,V.A. and Sali,A. (2002) LigBase: a database of families of aligned ligand binding sites in known protein sequences and structures. *Bioinformatics*, **18**, 200–201.