

Structural bioinformatics

Friend, an integrated analytical front-end application for bioinformatics

Alexej Abyzov, Mounir Errami, Chesley M. Leslin and Valentin A. Ilyin*

Department of Biology, Northeastern University, Boston, MA, USA

Received on March 18, 2005; revised on June 13, 2005; accepted on July 27, 2005

Advance Access publication August 2, 2005

ABSTRACT

Friend is a bioinformatics application designed for simultaneous analysis and visualization of multiple structures and sequences of proteins and/or DNA/RNA. The application provides basic functionalities, such as structure visualization, with different rendering and coloring, sequence alignment and simple phylogeny analysis, along with a number of extended features to perform more complex analyses of sequence structure relationships, including structural alignment of proteins, investigation of specific interaction motifs, studies of protein–protein and protein–DNA interactions and protein super-families. It is also useful for functional annotation of proteins, protein modeling and protein folding studies. Friend provides three levels of usage: (1) an extensive GUI for a scientist with no programming experience, (2) a command line interface for scripting for a scientist with some programming experience and (3) the ability to extend Friend with user written libraries for an experienced programmer. The application is linked and communicates with local and remote sequence and structure databases.

Availability: <http://mozart.bio.neu.edu/friend>

Contact: ilyin@neu.edu

Many areas of research in biology, biomedicine and bioinformatics require tools to perform comparative analytical studies of DNA and protein sequence and structure families. There are numerous examples where sequence–structure function research is essential, including the analysis of conservative positions, active and binding site residues, variations in orthologous and paralogous sequences and structures, sequence alignments, structural alignments and classification, identification of the key residues for protein functionality and formation of macromolecular complexes, phylogenetic tree analysis and simple visual inspection. Therefore, extended applications that integrate data from different sources, facilitate in visualization and assist in analytical studies are an everyday need in biology, biochemistry, molecular biology and related research areas.

Integration of different types of data is challenging and not always straightforward. Despite the significant variety of applications in each particular research area, sequence analysis, structure studies and phylogeny, only a few such applications integrate the data, for example JalView (Clamp *et al.*, 2004), Cn3D (Wang *et al.*, 2000), Pfaat (Johnson *et al.*, 2003) and DeepView (Schwede *et al.*, 2003). One more example of such an attempt to integrate sequence and structure analysis, ModView, a Netscape plug-in, has been reported recently (Ilyin *et al.*, 2003). While the integrative analytical

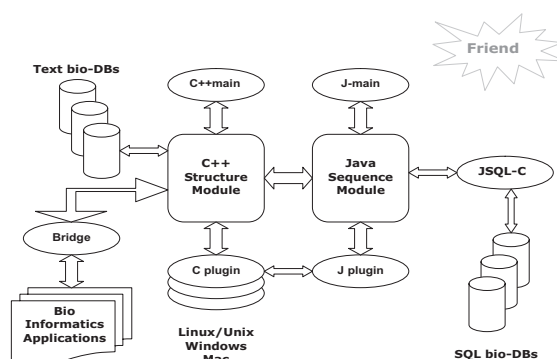


Fig. 1. General architecture of Friend.

ability of the application has been found useful for protein analysis and the plug-in is downloaded constantly by Linux/Netscape version 4 users, ModView has a clear downside because of the strong dependence on the evolution of the Netscape browser, in particular, Netscape does not provide backward compatibilities in its new versions and has restricted data handling abilities. To avoid dependence on any third party software a new stand-alone program, Friend, has been developed, which itself communicates with Internet resources, provides data integration between structure and sequence modules and local user files (Fig. 1). The majority of ModView's functionalities have been re-implemented in Friend, while many new options have been added, including one that allows a user to extend Friend with new functionality with dynamic libraries, i.e. its own plug-ins.

Friend is a stand-alone, multi-module, web-related application, designed to be a front-end user interface for visual and analytical studies on multiple sequences and multiple structures in real time on a personal computer (Fig. 2). Both structure and sequence data are linked to local and remote databases, providing researchers with a comprehensive picture about related proteins. Friend allows a user to visualize and manipulate hundreds of spatial protein or DNA/RNA structures and sequences, and provides an easy to use interface.

There are more than 200 basic commands, including different coloring and rendering of the protein structure, exhaustive selection of subsets of atoms, residues, stereoviewing of structures, interactive atom/residue identification and labeling, alignment editing, phylogeny clustering and tree viewing, pairwise sequence alignment, cross coloring of sequence and structures, changing residue representation in structure by simply clicking on it in the sequence, saving the results in a variety of text and graphical formats and producing publication quality pictures. The application supports several

*To whom correspondence should be addressed.

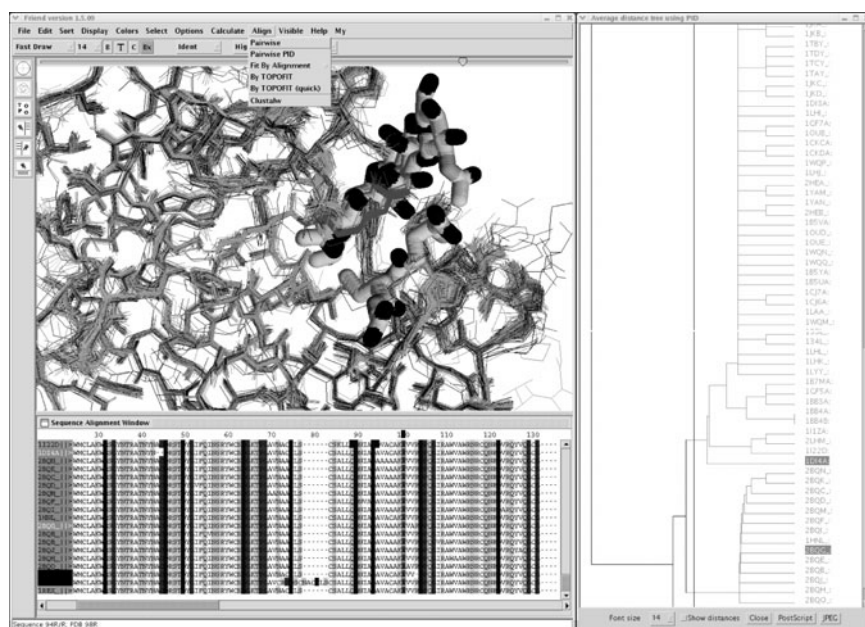


Fig. 2. Friend displays structures, sequences and phylogeny tree for 137 lysozymes.

popular alignment and structure representation formats, including PDB, FASTA, PIR, CLUSTAL and SKY. The SKY-format was created to fulfill the demand for saving and linking sequence and structure data cooperatively.

The advantage of Friend is the three levels of its usage: (1) for a scientist or users with no programming experience Friend provides an extensive GUI (the description can be found in the program manual); (2) for a scientist with some programming experience, it provides the possibility to create user-defined menus in an XML-file in order to use and combine any of the more than 200 commands; and (3) for a programmer, the ability to extend Friend with user written libraries. This is accomplished by use of abstract classes, providing functions to access and manipulate internal application data. Using the abstract interfaces one can write the code and compile it into a dynamically loaded library, which is loaded and executed during run time. The absence or presence of the library does not affect the functionality of the Friend core. The opportunity to add user written libraries allows users to perform more complex and specific analysis of the studied object in a time effective manner, since there is no need to develop basic routine functions. This also ensures that the code developed by different people does not interfere with each other. As an example, the TOPOFIT method for the structural alignment of proteins (Ilyin *et al.*, 2004) has been implemented as a separate dynamically loaded library.

Another powerful feature of the Friend application is the ability to provide an interface to various sequence and structure databases and other bioinformatics applications. Friend is used as a visual front-end interface to the Structural Exon Database, SEDB (Leslin *et al.*, 2004), and to a database for mapping non-synonymous SNPs, StSNP (Uzun *et al.*, 2005). Internal integrated client modules allow a user to perform similarity searches using BLAST and to load protein structures from the PDB on 'the fly'. Friend also provides an interface to the homology modeling software MODELLER (Sali and Blundell, 1993) and to the multiple sequence alignment program ClustalW (Jeanmougin *et al.*, 1998). The QHULL (Barber *et al.*, 1996,

<http://www.qhull.org>) library for fast Delaunay Tessellation (Delaunay, 1934) is integrated in Friend to analyze protein and DNA/RNA atom-atom, residues-residue, residues-base and base-base interactions along with the visualization of the tessellation in different views.

Friend is an ongoing project and future directions include development of several additional modules to broaden the number of database interfaces (to NCBI databases, SWISS-PROT, UniProt, etc.), porting it to the Mac OS and supplying the application as a plug-in for a number of popular browsers (Internet Explorer, Netscape, Mozilla and FireFox). Friend has been extensively used in a bioinformatics course and has a constant rate of outside installations, ~25–30 per month since January 2003.

REFERENCES

- Barber, C.B. *et al.* (1996) The Quickhull algorithm for convex hulls. *ACM Trans. Math. Softw.*, **22**, 469–483.
- Clamp, M. *et al.* (2004) The Jalview Java Alignment Editor. *Bioinformatics*, **20**, 426–427.
- Delaunay, B. (1934) Sur La Sphere Vide. *Bull. Acad. Science USSR VII: Class Sci. Mat. Nat.*, 793–800.
- Ilyin, V.A. *et al.* (2003) ModView, visualisation of multiple protein sequences and structures. *Bioinformatics*, **19**, 165–166.
- Ilyin, V.A. *et al.* (2004) Structural alignment of proteins by a novel TOPOFIT method, as a superimposition of common volumes at a topomax point. *Protein Sci.*, **13**, 1865–1874.
- Jeanmougin, F. *et al.* (1998). Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.*, **23**, 403–405.
- Johnson, J.M. *et al.* (2003). Protein family annotation in a multiple alignment viewer. *Bioinformatics*, **19**, 544–545.
- Leslin, C.M. *et al.* (2004) Structural Exon Database, SEDB, mapping exon boundaries on multiple protein structures. *Bioinformatics*, **20**, 1801–1803.
- Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.
- Schwede, T. *et al.* (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.*, **31**, 3381–3385.
- Uzun, A. *et al.* (2005) Structure SNP (StSNP) Database: A tool for modeling nsSNPs locations on proteins and presenting their metabolic pathways. RECOMB 2005, p. 179.
- Wang, Y. *et al.* (2000) Cn3D: sequence and structure views for entrez. *Trends Biochem. Sci.*, **25**, 300–302.