

Friend, an Integrated Analytical Front-end Application for Bioinformatics

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

Department of Biology, Northeastern University, Boston, MA

Email: ilyin@neu.edu

Almost every area of research in biology, biomedicine and bioinformatics requires tools to provide comparative analytical studies of DNA and protein sequences, along with their encoded structures, and their families. There are numerous examples where sequence-structure function research is essential, including analysis of conservative positions, active and binding site residues, variations in orthologs and paralog sequences and structures, pairwise sequence alignments, structural alignments and classification, identification of the key residues for protein functionality and complex formation, phylogenetic tree analysis, and simple visual inspection. Therefore, applications which integrate the data from different sources, facilitate in the visualization and assist in the analytical analysis become an everyday need in biological, biochemical, molecular biology and other 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 visualization and clustering (phylogeny) analysis only a few such applications integrate the data together. One example of such an attempt to integrate sequence and structure analysis, ModView, a Netscape plug-in, has been recently reported (Ilyin *et al.*, 2003). Here we present the next step in the development of the integrated environment for bioinformatics, a more powerful, integrated, stand-alone application Friend, which inherits ModView's functionalities and substantially extends them.

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

There are more than 200 basic options, including various coloring and rendering of the protein structure, exhaustive selection of subsets of atoms, residues, stereo view of structure, 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, and saving the results in a variety of text and graphical formats. Several popular alignment and structure representation formats are supported: PDB, FASTA, PIR, CLUSTAL, and SKY (introduced as a part of the Friend development) formats. SKY-format was created to fulfill the demand for saving and linking sequence and structure data cooperatively.

There are three levels of usage in Friend: 1) for a scientist with no programming experience, Friend provides users with an extensive GUI (its 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 a XML-file 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 code and compile it into the 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; it 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 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 SEDB (Leslin *et al.*, 2004) and StSNP (Uzun *et al.*, 2004) databases. 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, 1996) library for fast Delaunay Tessellation (Delaunay, 1934) is integrated in Friend to analyze protein and DNA/RNA atom-atom, residues-residue, residues-base, base-base interactions along with the visualization of the tessellation in different views.

Friend is an ongoing, developing project. Future directions in development include adding several modules to: broaden the number of the databases it can provide interfaces to (NCBI databases, SWISS-PROT, UniProt, etc), porting it to the Mac OS platform and supplying the application in the form of a plug-in for a number of popular browsers (Internet Explorer, Netscape, Mozilla, and FireFox).

Friend is extensively used in Bioinformatics courses and also has a constant rate of outside installations ~25-30 per month since 2003.

The application runs under Windows and Linux platforms, requires 128 Mb of memory and Java 1.4+ installed. It is free and publicly available from <http://mozart.bio.neu.edu/friend>