Structural bioinformatics

ET viewer: an application for predicting and visualizing functional sites in protein structures

Daniel H. Morgan\textsuperscript{1,3}, David M. Kristensen\textsuperscript{1,3}, David Mittelman\textsuperscript{2,3} and Olivier Lichtarge\textsuperscript{1,3,*}

\textsuperscript{1}Department of Molecular and Human Genetics, \textsuperscript{2}Department of Biochemistry and \textsuperscript{3}Program in Structural and Computational Biology and Molecular Biophysics, One Baylor Plaza, Houston, TX 77030, USA

Received on April 25, 2006; accepted on May 31, 2006
Advance Access publication June 29, 2006

ABSTRACT

Summary: The Evolutionary Trace Viewer (ETV) provides a one-stop environment in which to run, visualize and interpret Evolutionary Trace (ET) predictions of functional sites in protein structures. ETV is implemented using Java to run across different operating systems using Java Web Start technology.

Availability: The ETV is available for download from our website at http://mammoth.bcm.tmc.edu/traceview/index.html. This webpage also links to sample trace results and a user manual that describes ET Viewer functions in detail.

Contact: lichtarge@bcm.edu

1 INTRODUCTION

Proteins run the cellular machinery and most drugs work by altering their function. Thus, a key aim of therapeutics design is to identify and then interfere with the molecular determinants of protein action. While the discovery of novel protein structures is essential to this goal, and becoming an increasingly robotic task through structural proteomics, the high-resolution characterization of protein functional sites and molecular determinants remains rooted in painstaking experimental mutational analysis.

In order to relieve this bottleneck, the Evolutionary Trace (ET) (Lichtarge \textit{et al}., 1996; Mihalek \textit{et al}., 2004) was developed as a scalable alternative that identifies functional residues computationally. ET analyzes evolutionary patterns of sequence variations and then ranks every residue in a protein’s sequence by relative importance. Top-ranked residues consistently exhibit useful structural and functional features: they cluster spatially in protein structures (Madabushi \textit{et al}., 2002); they overlap known functional sites (Madabushi \textit{et al}., 2004; Yao \textit{et al}., 2003); and clustering quality correlates with functional site overlaps (Mihalek \textit{et al}., 2006). Thus mutations targeted to top-ranked clustering residues predictably alter function. For example, ET studies have led to \textit{bona fide} predictions of new functional sites, to the rational re-design of protein functional specificity, to uncoupling signaling pathways arising from the same receptor, and to peptides that inhibit protein–protein interactions (Gu \textit{et al}., 2005; Lichtarge and Sowa, 2002; Madabushi \textit{et al}., 2004; Shenoy \textit{et al}., 2006).

The ET Viewer (ETV) is a set of integrated modules that make such predictions of functional sites and specificity determinants widely available. ETV allows a user to launch traces, and then interactively view the alignment, the related phylogenetic tree, and a molecular graphics display of top-ranked residues mapped on the structure for any user-specified adjustable rank threshold (Fig. 1). Unlike other ET servers (Joachimiak and Cohen, 2002; Shi, 2004, \textit{http://www-cryst.bioc.cam.ac.uk/~jiye/evoltrace/evoltrace.html}; Toh and Nohara, 2005, \textit{http://timpani.genome.ad.jp/~ash/calcevolutionary.html}), the ETV requires only a PDB as input, not a full alignment, and it provides two unique features: an objective assessment of the statistical significance of trace clusters (Z-score), and an rvET option (Mihalek \textit{et al}., 2004), which is a version of ET that is more robust to sequence errors and fragments.

2 DESIGN

ETV is implemented in Java and launched via Java Web Start, which uploads the initial application to the client machine and automatically checks for updates. There are four major modules: a structure viewer, a tree viewer, a MSF viewer (Fig. 1) and the ET wizard. These modules are synchronized so that ET results can be viewed at any chosen rank, and they are coupled so that branches and sequences can be selected and then used as input to run successively more refined ET analyses.

2.1 Structure viewer

The structure viewer is the main window from which other modules are launched. Its primary function is to display ET rank information on the protein structure. The core file used by ETV contains structural information, the ET ranks file and tree file. The protein can be rotated, translated, zoomed and shown in either space-fill (with and without side chains) or stick representations. Individual residues can be clicked and identified. ETV also provides PyMol (DeLano, 2002) and MolMol (Koradi \textit{et al}., 1996) scripts, for more advanced visualizations.

A key feature is that users can quickly scan through ET ranks by dragging a slider. This dynamically updates all of the modules to the current rank, percent coverage and Z-score statistics. The entire list of trace ranks is available and can be visualized on the structure as a red to blue spectrum. Alternatively, functional sites can be most easily examined in cluster coloring mode and viewing a space-fill model while sliding across ranks. Trace residues in the protein core can be seen using bond or backbone views.

2.2 ET tree viewer

The ET tree viewer is adapted from ATV (Zmasek and Eddy, 2001) and displays the family tree used to compute ET ranks. As the rank slider is moved in the structure viewer, the ET tree is continuously...
updated to show the node at which the current rank diverges [this feature is for integer ET only since the rank-to-node correspondence is less well defined in rET (Mihalek et al., 2004)]. Nodes may be clicked to highlight one or many branches of the tree. These sequences are also highlighted in the MSF viewer where they may be saved and, in turn, used as input to the ET wizard so that they can be traced by themselves for difference ET analysis (Madabushi et al., 2004).

2.3 MSF viewer

The MSF viewer displays the multiple sequence alignment (MSA) associated with the current trace. If a user wishes to use their own MSA, it must be in GCG or FASTA format. File loading and sequence data structures are derived from PFAAT (Johnson et al., 2003). All or some of the sequences can be selected and saved for input to the ET wizard. Selected sequences are also shown on the ET tree if the ETV file is loaded. It implements ClustalX (Thompson et al., 1997) coloring schemes to aid the user in alignment analysis. The top of the window displays an ET rank profile with the current rank threshold and currently ranked alignment columns highlighted in red.

2.4 ET wizard

The ET wizard is an interface, called from the Utility dropdown menu in the structure viewer, that lets a user launch a customized ET. The minimum input is a PDB ID code, but users may also upload their own alignment files and tree files, and they can modify ET default settings. The wizard then sends these queries to an ET server and the trace is run. While it runs, the server sends information to the client about trace status, including failures. When completed, all data files (including the entire list of trace ranks in both rank order and residue order) are zipped and downloaded back to the user’s computer. The user is then able to visualize the resultant trace and make any modifications that are desired, independent of the ET server.

3 SYSTEM REQUIREMENTS

The ET Viewer requires Java JRE 1.4.2 to run. Efficient protein structure viewing requires a 3D hardware accelerated video card with a minimum of 32 MB of video memory. The ET wizard requires an internet connection for job submission and retrieval. Future updates will add ribbon representations and Difference ET. Feedback and suggestions can be sent to etviewer@bcm.edu.

ACKNOWLEDGEMENTS

The authors would like to thank Chris Agerton, from Bellaire High School, and NSF DGE-0086397 and DGE-0440525 ( GK12), DBI-0318415, MRI-0420984; the March of Dimes FY03-93, and NIH GM066099 for grant support.

Conflict of Interest: none declared.

REFERENCES


