

PyETV: a PyMOL evolutionary trace viewer to analyze functional site predictions in protein complexes

Rhonald C. Lua and Olivier Lichtarge*

Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

Associate Editor: Anna Tramontano

ABSTRACT

Summary: PyETV is a PyMOL plugin for viewing, analyzing and manipulating predictions of evolutionarily important residues and sites in protein structures and their complexes. It seamlessly captures the output of the Evolutionary Trace server, namely ranked importance of residues, for multiple chains of a complex. It then yields a high resolution graphical interface showing their distribution and clustering throughout a quaternary structure, including at interfaces. Together with other tools in the popular PyMOL viewer, PyETV thus provides a novel tool to integrate evolutionary forces into the design of experiments targeting the most functionally relevant sites of a protein.

Availability: The PyETV module is written in Python. Installation instructions and video demonstrations may be found at the URL: <http://mammoth.bcm.tmc.edu/traceview/HelpDocs/PyETVHelp/pyInstructions.html>.

Contact: lichtarge@bcm.tmc.edu

Received on July 8, 2010; revised on September 14, 2010; accepted on October 3, 2010

1 INTRODUCTION

Since protein–protein interactions are ubiquitous and an emerging target for design and therapeutics (Mandell and Kortemme, 2009), it is critical to improve the tools that enable their characterization. Here, we present an ET Viewer that provides a high-quality interface to map and analyze evolutionary forces and the functional sites they define across multi-protein interfaces and assemblies. Importantly, it enables the integration of Evolutionary Trace (ET) analysis (Lichtarge and Wilkins, 2010; Lichtarge *et al.*, 1996) with any type of structural and biophysical information accessible to PyMOL (DeLano, 2002).

ET is a well-validated method to identify functional sites and their residue determinants in proteins. It drives experiments that efficiently elucidate the molecular basis of binding, catalysis and allostery, and that rationally perturb networks (Ribes-Zamora *et al.*, 2007; Rodriguez *et al.*, 2010). ET ranks the ‘evolutionary importance’ of sequence positions by tracking whether their variations during evolution correlate with large or small divergences among orthologs and paralogs (Lichtarge *et al.*, 1996). These ET ranks are robust, and the best-ranked sequence positions form continuous spatial clusters (Wilkins *et al.*, 2010) that reveal functional sites and their functional determinants. Varying the threshold of importance reveals these sites at varying levels of

detail. Recent studies show that small motifs of ET residues can be compared across all protein structures to predict functions in enzymes and non-enzymes alike, thereby validating ET predictions of functional determinants on a large scale (Erdis *et al.*, 2010; Kristensen *et al.*, 2008).

Yet, the study of the structural interactions of evolutionary important ET residues across structural interfaces has been limited. The ET report_maker (Mihalek *et al.*, 2006) only provided a PDF report of individual sequence or structure ET analysis, with no graphical user interface. A prior Evolutionary Trace Viewer (ETV) that did provide an interactive molecular viewer (Morgan *et al.*, 2006) could only display just a single protein chain at a time; and it could not show surfaces, secondary structure elements or any other type of information, such as crystallographic B-factors, electrostatics or bound ligands. By addressing these problems, PyETV offers a tool to study protein functional determinants and evolutionary forces in the more relevant structural context of quaternary interactions.

2 OVERVIEW OF PYETV

PyETV builds on the popular and extensible PyMOL (DeLano, 2002) platform. PyMOL is a molecular graphics package to view, select, label and perturb any number of structures or residues in many ways (e.g. cartoon, surface, stereo, etc.). Moreover, it is easily extended with plugins—scripts that overlay complementary information through custom interfaces, such as the APBS plugin (by M. G. Lerner) to generate electrostatics maps. Likewise, PyETV, which opens when selected among the items under the ‘Plugin’ menu, specifically maps ET ranks onto structures.

2.1 Loading evolutionary information

The primary source of precomputed ET rank data is the ET server (<http://mammoth.bcm.tmc.edu/ETserver.html>), which regularly updates to incorporate new PDB (Berman *et al.*, 2000) structures. If a user wishes to start with a new protein sequence or provide a custom alignment, new ET rank data can be generated via the ET Wizard feature of ETV (Morgan *et al.*, 2006).

ET rank data can be loaded into PyETV in several ways:

- (1) Most simply, PyETV has a ‘Load trace’ feature that takes a single PDB code and chain identifier and fetches ET ranks and the matching PDB chain directly from the ET server. Similarly, traces and structures of entire biological units can be loaded at once (see Section 2.4).
- (2) Alternatively, PyMOL scripts can be used to launch PyMOL and to simultaneously load multiple structures and their corresponding traces. After selecting PyETV in the ‘Plugin’

*To whom correspondence should be addressed.

menu, the ranks are mapped to their corresponding structures. Web links to such scripts are available in the ET server.

- (3) Finally, users can explicitly specify in PyETV a directory path to a rank data file they have produced themselves or obtained from sources mentioned above.

2.2 Basic viewing

Once the ET data are loaded for any number of chains, the top-ranked residues of each one can be highlighted at any desired threshold of evolutionary importance by dragging a slider (horizontal scrollbar) specific to that structure. The color scheme is flexible: it can be made uniform (all red for the top n -th percentile rank of importance), or it can follow a rainbow spectrum in which red is the most important and purple the least so. For added clarity, top-ranked residues may also be distinguished from the rest of the structure by their representation (e.g. spheres, C-alpha atoms, sticks). (For a video demonstration, see <http://www.youtube.com/watch?v=Wt5Q0Nwvu24>.) These operations can be repeated individually for each chain, or they can be coordinated among all chains to display the same threshold of importance throughout a complex.

2.3 Statistics of trace clusters

After selecting top-ranked residues for some threshold, PyETV provides z -scores (Mihalek *et al.*, 2003; Wilkins *et al.*, 2010) that quantitatively assess whether these residues cluster within a single protein structure or across a protein–protein interface more so than expected by chance (i.e. if z -score > 2). High z -scores indicate that surface groupings of ET residues may well reveal functional sites. Two types of clustering z -scores are available for single structures: one that accounts for close sequence neighbors (3D biased z -score) and another that does not (3D no-bias z -score). For protein–protein complexes, a similar z -score, called Z_{coupling} , is also available.

2.4 PISA assembly tool

It is often non-trivial to extract the biologically relevant complex from a PDB file. As an aid to general users, PyETV can load a macromolecular assembly directly from Protein Interfaces, Surfaces and Assemblies (PISA, http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) (Krissinel and Henrick, 2007), and match precomputed traces to each chain. The user need only to supply a PDB code. Mappings of ET ranks can be simultaneously controlled over all chains at once to examine the joint distribution of importance over the entire complex. To assist with interface analysis, one may select residues in a structure that is in contact with any ligand, such as a protein complex partner.

For demonstrations of these features, see the following: (i) to load PISA structures, map ET ranks and select a protein–protein interface, view <http://www.youtube.com/watch?v=1VCdKPKqLdg>. (ii) for another video example of selecting and isolating interface

residues with ET ranks mapped to these residues, visit <http://www.youtube.com/watch?v=OUyaJCSYQQA>.

3 CONCLUSION

PyETV integrates data from several sources (ET, PDB, PISA) and extends the trace-to-structure mapping originally implemented in the Java-based ETV to any number of structures and traces. PyETV relies on the power, flexibility and wide availability of the PyMOL molecular visualization system and its extensibility through Python. Future updates may include new tools for analyzing protein–protein interfaces, mapping new attributes such as correlations between two residues or ET ranks for residue pairs, and incorporating quality measures that supplement the clustering z -score.

ACKNOWLEDGEMENTS

We thank P. Katsonis, D. H. Morgan, J. Quiros, E. Venner, A. D. Wilkins and R. Yao.

Funding: National Science Foundation (DBI-0547695 and CCF-0905536 to O.L.); National Institutes of Health (GM079656 and GM066099 to O.L.).

Conflict of Interest: none declared.

REFERENCES

- Berman, H.M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
- DeLano, W.L. (2002) *The PyMOL Molecular Graphics System*. DeLano Scientific, San Carlos, CA.
- Erdin, S. *et al.* (2010) Evolutionary trace annotation of protein function in the structural proteome. *J. Mol. Biol.*, **396**, 1451–1473.
- Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372**, 774–797.
- Kristensen, D.M. *et al.* (2008) Prediction of enzyme function based on 3D templates of evolutionarily important amino acids. *BMC Bioinformatics*, **9**, 17.
- Lichtarge, O. *et al.* (1996) An evolutionary trace method defines binding surfaces common to protein families. *J. Mol. Biol.*, **257**, 342–358.
- Lichtarge, O. and Wilkins, A. (2010) Evolution: a guide to perturb protein function and networks. *Curr. Opin. Struct. Biol.*, **20**, 351–359.
- Mandell, D.J. and Kortemme, T. (2009) Computer-aided design of functional protein interactions. *Nat. Chem. Biol.*, **5**, 797–807.
- Mihalek, I. *et al.* (2003) Combining inference from evolution and geometric probability in protein structure evaluation. *J. Mol. Biol.*, **331**, 263–279.
- Mihalek, I. *et al.* (2006) Evolutionary trace report_maker: a new type of service for comparative analysis of proteins. *Bioinformatics*, **22**, 1656–1657.
- Morgan, D.H. *et al.* (2006) ET viewer: an application for predicting and visualizing functional sites in protein structures. *Bioinformatics*, **22**, 2049–2050.
- Ribes-Zamora, A. *et al.* (2007) Distinct faces of the Ku heterodimer mediate DNA repair versus telomeric functions. *Nat. Struct. Mol. Biol.*, **14**, 301–307.
- Rodriguez, G.J. *et al.* (2010) Evolution-guided discovery and recoding of allosteric pathway specificity determinants in psychoactive bioamine receptors. *Proc. Natl Acad. Sci. USA*, **107**, 7787–7792.
- Wilkins, A.D. *et al.* (2010) Sequence and structure continuity of evolutionary importance improves protein functional site discovery and annotation. *Protein Sci.*, **19**, 1296–1311.