

Visualization and Analysis of Eukaryotic Gene Signals on Protein Structures

Vivek Gopalan¹, Shang Liang², Tin Wee Tan³, Shoba Ranganathan⁴

Keywords: evolution, protein, gene, structure

1 Introduction.

Any relation between eukaryotic gene signals on protein structures is essential for analyzing the origin and evolution of genes. Here we present an approach for analyzing the correspondence between boundaries of the protein structural elements such as go-modules, domains and secondary structures and the intron positions in homologous eukaryotic genes. As a first step, we have developed a program that visualizes the various structural elements and provides detailed statistics about the significance of the intron positions in a single or set of protein structures.

The results from the analysis can be used for testing the Exon Theory of Genes (ETG) or intron early theory, which advocates that genes in primordial cells contain introns and proposes exon shuffling as the main process for protein diversity.

2 Methods

The intron positions and phases of eukaryotic genes are extracted from Xpro [1], which is eukaryotic protein encoding database based on GenBank (version 137)[2]. The algorithm for prediction of go module boundaries is based on de Souza et al [3]. The intron positions and phases are mapped on the PDB protein chain sequences based on conserved region in the BLAST alignment against the subset of protein sequences that represent intron containing genes in Xpro database. The E-value cutoff for the BLAST search was kept as 10^{-4} [4]. The program is written in Java 2 and Java 3D API is required for visualization of protein structures. The cartoon model of the protein structure is obtained based on the MOLSCRIPT program [5]. The secondary structures of the protein chains are obtained from the definitions in the PDB file.

3 Results

Figure 1 shows the Go plot for the chain A of the PDB code 1TIM. The module regions and their boundaries are shown as rectangular blocks in the hypotenuse of the Go plot. Figure 2 shows the screen shot of the of the protein structure for the PDB code 1TIM with Go-module mapped on it.

¹ Department of Biochemistry, National University of Singapore, Singapore 119260. E-mail: vivek@bic.nus.edu.sg

² Department of Biochemistry, National University of Singapore, Singapore 119260. E-mail: sci11162@nus.edu.sg

³ Department of Biochemistry, National University of Singapore, Singapore 119260. E-mail: tinwee@bic.nus.edu.sg

⁴ Department of Biochemistry, National University of Singapore, Singapore 119260. E-mail: shoba@bic.nus.edu.sg

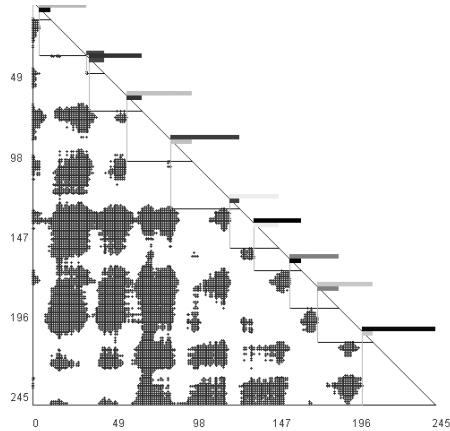


Figure 1: Screen shot showing Go plot for chain A for PDB code 1TIM at 28 Å diameter.

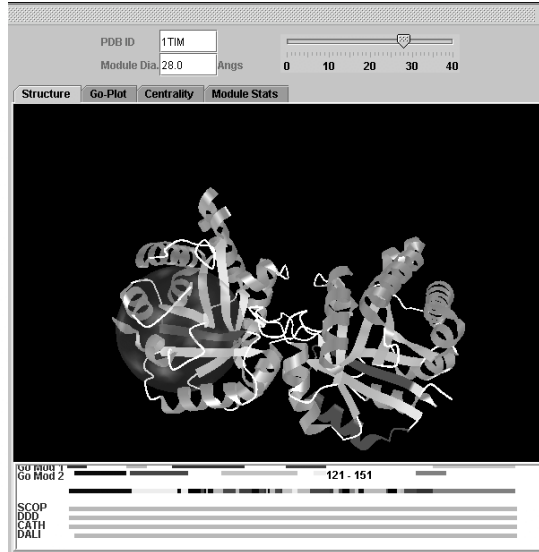


Figure 2: Cartoon model of a PDB structures with Go-modules mapped on protein 3D structure. The spherical region in the left chain indicates a Go-module (PDB code – 1TIM, chains – A,B).

The lower region of figure 2 shows the Go-modules mapped a PDB chain in addition to the mapped intron positions and domain definitions from various domain definition databases such as CATH, SCOP, DALI, and 3Dee. Provisions for changing the Go-module diameter and visualizing each of the domain regions, exon encoded regions or module boundaries are built-in within the program. In conclusion, we have developed an integrated platform for analyzing and visualizing features related to protein structures and eukaryotic gene signals.

4. References

- [1] Gopalan, V., Tan, T.W, Lee, B.T. and Ranganathan, S. 2004. Xpro: database of eukaryotic protein-encoding genes. *Nucleic Acids Research* 32:D59-63.
- [2] Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Wheeler, D.L. 2003. GenBank. *Nucleic Acids Research*. 31:23–27.
- [3] de Souza, S.Long, M., Schoenbach, L. and Gilbert, W. 1996. Intron positions correlate with module boundaries in ancient proteins. *Proceedings of the National Academy of Sciences USA* 93: 14632-14636.
- [4] Fedorov, A., Cao, X., Saxonov, S., de Souza, S.J., Roy, S.W., Gilbert, W. 2001. Intron distribution difference for 276 ancient and 131 modern genes suggests the existence of ancient introns. *Proceedings of the National Academy of Sciences USA* 93: 14632-14636 98:13177-82.
- [5] Kraulis, P.J. 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *Journal of Applied Crystallography* 24: 946-950.