

# Northeast Structural Genomics Consortium

## The MarkUs Function Annotation Server

<http://luna.bioc.columbia.edu/honiglab/mark-us>

It is now well established that structural similarity of overall chain fold can often provide clues to evolutionary relationships that are not evident from sequence information alone, and thus provide clues to molecular function. However, functionally relevant structural similarities often exist between regions of two proteins that have quite different global topologies or chain folds.

The MarkUs server, <http://luna.bioc.columbia.edu/honiglab/mark-us/cgi-bin/submit.pl>, developed by the Honig group as part of NESG PSI-2, uses structural similarity to guide molecular function annotation (Petrey *et al.*, 2009). MarkUs identifies structurally similar proteins with cutoff criteria that are loose enough to accept alignments where two proteins have as few as three secondary structure elements in common. Information about proteins with related structural fragments are often obtained, while avoiding classifications based on fold and topology. This strategy significantly expands the number of relationships that can be detected.

The MarkUs user is offered a variety of options for filtering “hits” and for displaying functional information. Fig. 1 provides an example of a MarkUs application to the identification of sugar binding sites in a protein of unknown function. The idea is to find multiple possible relationships, based on structural similarity, followed by user-directed filtering based on other biophysical and biochemical properties. Rather than focusing entirely on automatic tools, MarkUs is more in tune with the way research is traditionally carried out by biologists, and is designed to allow a user, with interest in a particular problem and specific hypotheses in mind, to manipulate and visualize a set of from different sequence and structural databases. Precalculated results have also been generated for all PSI proteins; Markus functional annotations for nearly all of the structured NESG proteins are available at <http://luna.bioc.columbia.edu/honiglab/nescg/cgi-bin/browse.pl?>.



**Fig. 1.** The MarkUs “annotation map”. The grey lines (A) represent the sequences of a query protein (first grey line) and its structural neighbors. The magenta rectangles indicate residues associated with a given function. Hovering the mouse over different areas of the annotation map will display “tool tips” (B) that provide additional functional details about each protein. By clicking on the GO annotation in the tool tip, the set of proteins displayed can be filtered to include those that share that particular annotation. The information displayed on the annotation map can be changed using the controls to the left (C,D). In this case, all ligand contacts are displayed but this can be restricted to ligands of a certain type based on the ChEBI annotation system (D). A wide array of other structural and functional information, including UniProt sequence features, sequence conservation, protein-protein interactions, SNPS, and secondary structure can be displayed using the menu (C) at the top left. E) Colored boxes in this region indicate residues lining predicted functional cavities identified by the program SCREEN.

Petrey, D.; Fischer, M.; Honig, B. *Proc. Natl. Acad. Sci. USA* 2009, 106, 17377 Structural relationships among proteins with different global topologies and their implications for function annotation strategies.