

CoPS

Application CoPS (**the Complex Comparison of Protein Structures**) is one of the BalticGrid-II project pilot applications. CoPS offers a method applied for exploration of potential evolutionary relationship between the CATH protein domains and their characteristics. The approach is based on the Evolutionary Secondary Structures Matching (ESSM) algorithm for detecting structural mutations, the output of which is used for construction of fold space graphs.

This approach is based on the assumption that protein structures, similarly to sequences, have evolved by a stepwise process, each step involving a small change in protein fold.

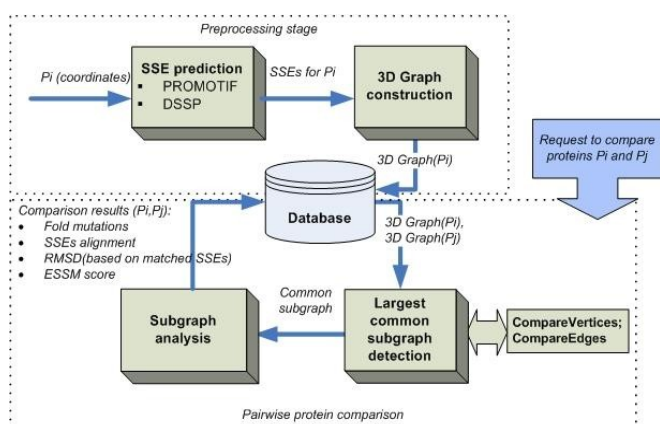


Fig. 1 Schema of ESSM tool.

The CoPS application is a convenient tool which allows easy detection of structural similarity and possible fold mutations between these proteins. This method consists of two stages:

- All-against-all comparison of CATH domains by the ESSM software;
- Construction of fold space graphs on the basis of the output from the 1st stage.

The CoPS is also implemented within the Migrating Desktop – an intuitive interface to Grid resources (see Fig.2).

Application needs

The CoPS is developed and used by the research communities in the Baltic States. Usage of the BalticGrid-II infrastructure significantly improves the outcome of the research as the number of pairwise protein comparisons is few billions. The application is also using the Protein DataBase, the size of which in uncompressed format is ~ 8GB.

Protein DataBase consists of:

- Protein Data Bank (with protein 3D coordinates);
- FASTA files (.txt files with protein sequences);
- Structural descriptions files with structural elements.

Total number of parallel tasks - 20 451 945, divided into 410 files.

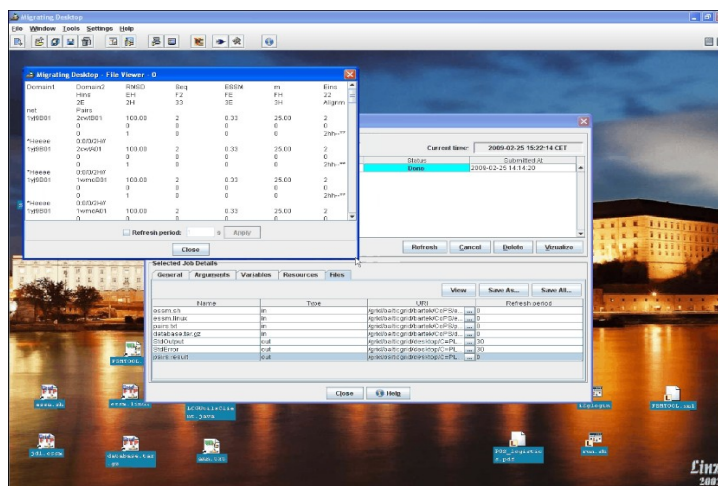


Fig. 2 Application running on Migrating Desktop.

Application results

The results obtained by the application were successfully used for the exploration of the CATH fold space by using fold space graphs for representation of comparison results and estimation of "evolution distance" on the basis of observed changes.

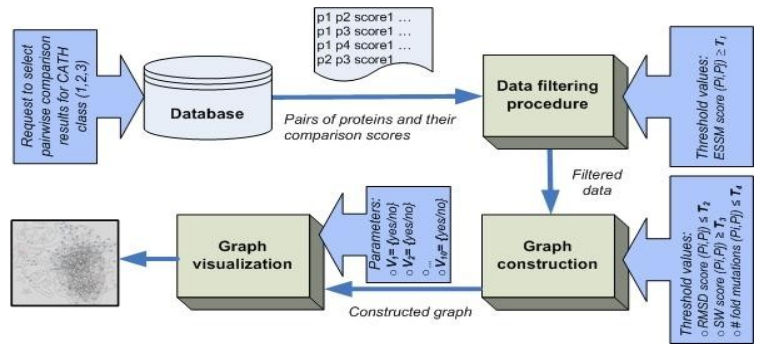


Fig. 3 Schema of fold space graph construction.

The creation of fold space graphs is done using a self-made tool that has been developed on the basis of some GRADE (advanced business and system modeling tool, developed by Institute of Mathematics and Computer Science University of Latvia (IMCS UL) - one of the BalticGrid-II partners) components. Some of the generated graphs can be seen in Fig.4.

There are two main aspects in the graph creation procedure: data filtering and selection of thresholds for the construction and visualization procedures.

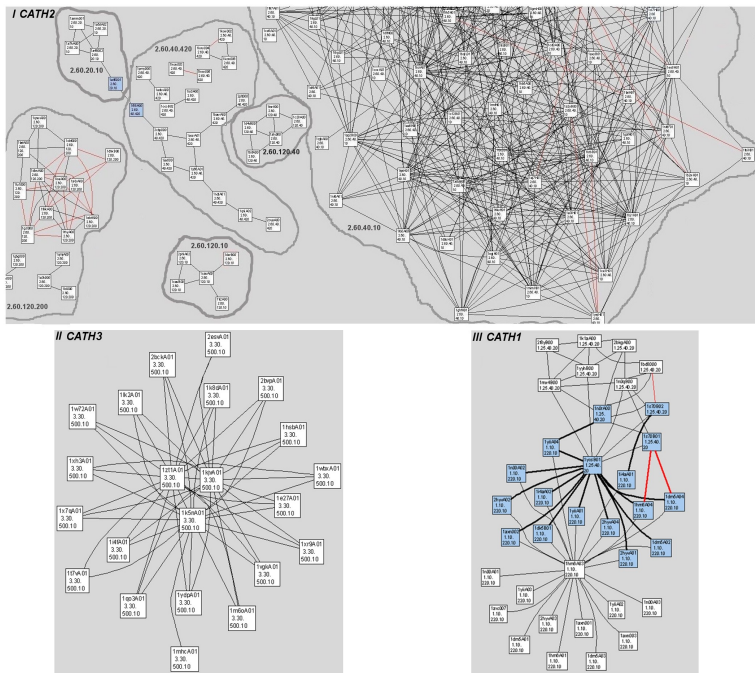


Fig. 4 Fold space graphs for CATH.

Figure 5 demonstrates chains of proteins extracted from fold space graphs, where neighbouring proteins are evolutionary related. Using these chains biologists are able to detect the most probable/interesting types of mutations of the affected protein structures.

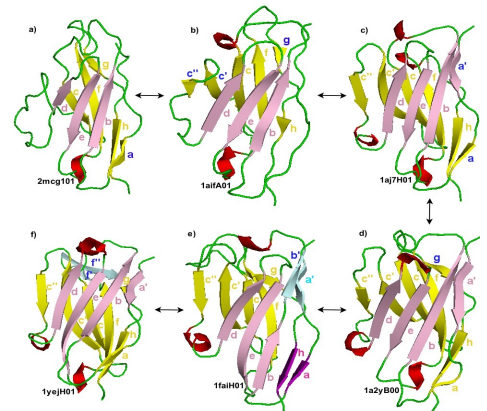


Fig. 5 Chains of proteins.

Project Director

Åke Edlund
KTH, Stockholm, Sweden
edlund@nada.kth.se

Technical Contact



IMCS UL; Dana Ludviga
dana.ludviga@sigmanet.lv