Comparison of protein topology graphs using graphlet-based methods

Tatiana Bakirova[1,2], Tim Schäfer[1], Ina Koch[1]

[1] Molecular Bioinformatics Group, Institute of Computer Science, Cluster of Excellence Frankfurt "Macromolecular Complexes", Robert-Mayer-Str. 11-15, 60325 Frankfurt am Main, Germany
[2] Department of Informatics, Karlsruhe Institute of Technology, Germany

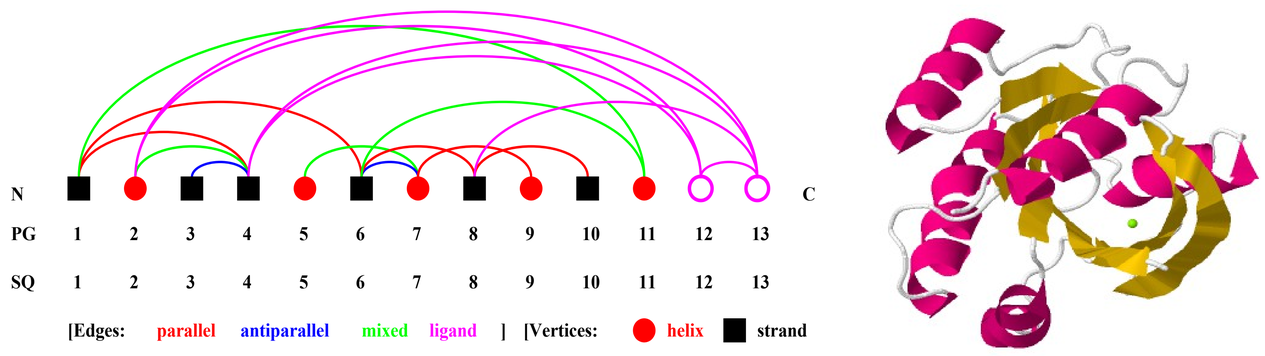
With the rapidly growing amount of protein structures available in public databases like the RCSB Protein Data Bank, there is a strong need for developing fast and accurate comparison methods for proteins. Detecting functional similarity, evolutionary relationships and/or structural motifs at different description levels of proteins is of major interest for many applications in biology and molecular medicine. Although there already exist different methods for protein structure alignment, the search of protein structure databases is still time consuming and similarity between proteins can be defined in many ways.

Project Objective

For a given query protein in the protein topology library (PTGL) [1] retrieve *k* similar proteins based on their protein graph similarity:

- compare the query graph to all graphs in PTGL,
- return *k* most similar candidates.

Protein topology graph for the Ras protein 121p_A



Methods

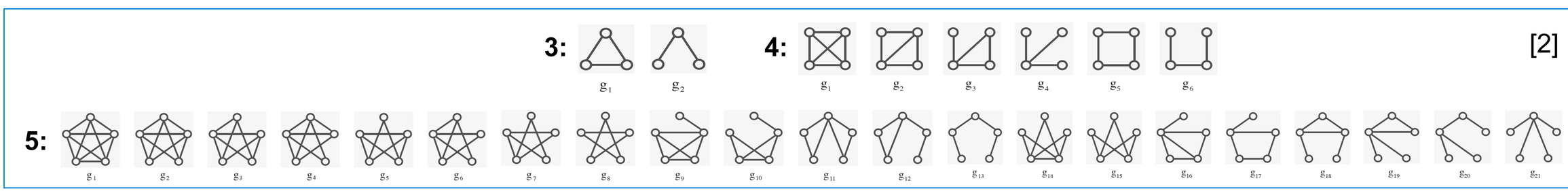
Graph comparison by feature-based approach using unlabeled graphlet counts.

Unlabeled graphlet counting algorithms used for experiments: “Enumerating all graphlets” by Shervashidze et al. [2]

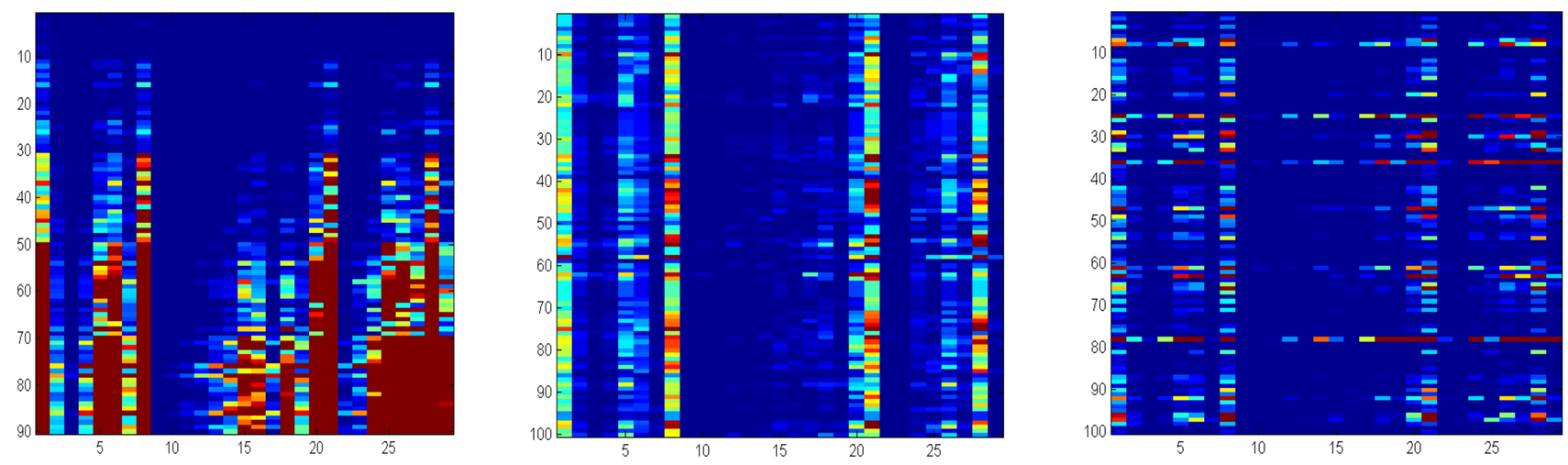
Future Work

- Bio-motivated labeled graphlets
- Similarity measure function

Graphlets



Graphlet Counts



a) random protein graphs with {5, 10, 15, 20, 25} vertices

b) proteins similar to 121p_A

c) proteins from different CATH superfamilies

References

[1] Schäfer, T., May, P., & Koch, I. (2012). Computation and Visualization of Protein Topology Graphs Including Ligand Information. In *GCB* (pp. 108-118).

[2] Shervashidze, N., Petri, T., Mehlhorn, K., Borgwardt, K. M., & Viswanathan, S. (2009). Efficient graphlet kernels for large graph comparison. In *International Conference on Artificial Intelligence and Statistics* (pp. 488-495).