Extension of protein topologies by ligand interactions: computation and visualization

Tim Schäfer, Patrick May, Ina Koch

August 14, 2011

Our knowledge on proteins expands quickly and with more and more 3D structures available in databases like the RCSB Protein Data Bank (PDB), computational tools to automatically search, compare and classify protein structures are needed. Not surprisingly, many secondary databases and software tools to perform these tasks on all levels of protein structure exist today.

The databases CATH [4], SCOP [3] and TOPS [6] all operate on the super-secondary structure level, i.e., they compare secondary structure elements (SSEs) and their spatial relation between different proteins. Protein structure databases usually compare protein topologies based on atom coordinates from PDB files, but all methods listed above ignore the ligand atoms stored in those files. TOPS+ [5] is the first method which includes ligand information in the comparison of proteins on the secondary structure level. It works on the domain level, uses a string-based description of protein secondary structure and includes additional biochemical and structural features like length of SSEs. String alignment methods are used to find similarities between domains.

The Protein Topology Graph Library (PTGL, [2]) is a database of protein topologies that provides a web interface to compare and visualize protein folds based on SSEs. It uses a graph-based protein model related to earlier work by Koch et al. [1]: the vertices of the protein graph represent the secondary structure elements (SSEs) α-helix and β-sheet while the edges model contacts and spatial relations between these SSEs. Similarity between proteins is defined via common substructures in their protein graphs. Protein-ligand contacts are computed from PDB files and a new vertex type for ligands is added to the protein model in this work. A new desktop application named Visualization of Protein-Ligand Graphs (VPLG), which computes protein graphs from PDB files and the SSE assignments of the DSSP algorithm, is presented. The protein graphs can also be visualized and written to image files in bitmap or vector formats. The program supports the computation of ligand contacts and allows the results to be saved to a database or to text files in a custom format for protein graphs. Optionally, coiled regions in proteins can also be included in the graph representations. An example output image is shown in Figure 1. Both the new version of the PTGL and the VPG software package enable comparison of proteins on the super-secondary structure level including ligands and are thus of interest in the fields of proteomics, drug design and molecular medicine.

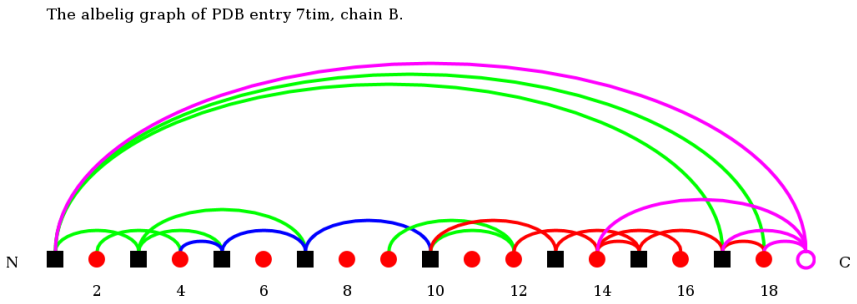

Figure 1: The albelig-graph of the β-chain of triosephosphate isomerase. The α-helices are shown as red circles and the β-strands as black square. Ligands are represented by magenta circles. From left to right, the vertices are ordered by their position in the amino acid sequence. The arcs mark spatial contacts (red, parallel; blue, anti-parallel; green, mixed; magenta, ligand contact).

References

[1] I. Koch, T. Lengauer, and E. Wanke. An algorithm for finding maximal common subtopologies in a set of protein structures. *J.Comp.Biol.*, 3:289–306, 1996.

[2] P. May, A. Kreuchwig, T. Steinke, and I. Koch. PTGL: a database for secondary structure-based protein topologies. *Nucl. Acid Res.*, 38:D326–D330, 2010.

[3] A.G. Murzin, S. E. Brenner, T. Hubbard, and C. Chothia. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, 247:536–540, 1995.

[4] C.A. Orengo, A.D. Michie, S. Jones, D.T. Jones, M.B. Swindells, and J.M. Thornton. CATH – a hierarchical classification of protein domain structures. *Structure*, 5:1093–1108, 1997.

[5] M. Veeramalai and D. Gilbert. A novel method for comparing topological models of protein structures enhanced with ligand information. *Bioinformatics*, 24:2698–2705, 2008.

[6] J. Viksna and D. Gilbert. Pattern matching and pattern discovery algorithms for protein topologies. In *Algorithms in Bioinformatics: First International Workshop*, 2001.