Application of Molecular-Similarity Networks for Chemoinformatics Research

Zsolt Lepp*, Chunfei Huang, Takashi Okada
Kwansei Gakuin University, Japan

Chemoinformatics projects are required to deal with large datasets of molecules. It is a crucial task to organize these compounds into meaningful groups, and also to integrate additional, non-stuctural information in a systematic way. A potential method is to create networks of the molecules on the basis of structural similarity and/or other properties. Although establishing such molecular networks is not a novel idea [1], there was relatively little activity on this area, so far. However, in the recent years, the explosion of network analyses techniques, led by various social- and system-biological researches gave us additional tools and opportunities to further advance this field. In the current research various topics on the assembly of molecular networks and use of those for chemoinformatics research were considered.

1. Analyzing the networks of molecules linked together by molecular similarity, and studying how these analyses could be related to common chemo and biological tasks. In this research we examined how four different correlation methods, which were used to calculate the similarity of molecular fingerprints, influenced the topology of networks. It was shown, that the structure of threshold networks was highly different in each cases. We found, that the most sensitive parameter to examine the influence of the similarity threshold on the topology of a network, was the number of strongly interconnected cliques. This parameter responds to the change of threshold by a high order function. It was also examined how the network-based approaches could be applied for clustering molecules. Minimum spanning trees as well as threshold networks were used, and it was shown, that although the results were different, all were meaningful and each gave answer to different chemoinformatics problems. In order to be able to visualize the network of molecules, two computer programs were integrated. One of those was Cytoscape, a software used in bioinformatics research to visualize system biology networks, and the other was Jchem, a system for chemoinformatics and molecular visualization.

2. Integrating chemical and biological knowledge. A fundamental task of pharmacological research is to categorize proteins based on activity. A chemo-centric approach to this problem is to compare not the biological targets themselves but rather the chemistry of their ligands. There are some pioneering works considering the practical approaches for this problem. [2-3].

In the frame of this research, a similar way of thinking was followed, but the focus was on organizing significant molecular fragments. The maximum common substructures (MCS) of the ligands of 240 proteins were determined. All of the generated substructures were unified into a central database. Although the database can be used to perform standard chemionformatics tasks, such as virtual screening, the main emphasis was on assembling a network, consisting of proteins and the most characteristic MCSes, which would graphically depict the map of pharmacologically significant molecular substructures.

[1] Mount, J.; Ruppert, J.; Welch, W.; Jain, A. N.J. Med. Chem.; **1999**; 42(1); 60-66.
[2] Paolini, G.V., Shapland, R.H.B. & v Hoorn, W.P. Mason, J.S. & Hopkins, A.L. Nat. Biotechnol. **2006**; 24; 805-815.
[3] M.J. Keiser, B. L. Roth, B. N. Armbruster, P Ernsberger, J. J. Irwin, K, Shoichet Relating protein pharmacology by ligand chemistry. Nat. Biotechnol. 2**007**; 25;197-206.