Statistical study of membrane and soluble protein - protein and protein - nucleic acid interactions.

David Arnould, Nicolas Delsaux, Jacques Leroux, Olivier Bouffioux, Annick Thomas, Robert Brasseur.

Centre de Biophysique Moléculaire Numérique, Faculté Universitaire des Sciences Agronomiques, Passage des déportés 2, 5030 Gembloux, Belgium.

Protein-protein and protein-DNA interactions are involved in numerous essential biological processes and diseases. The prediction of the binding sites of these molecules is an interesting challenge to better understand biochemical pathways and to elaborate future pharmaceutical therapies. In this quest, bioinformatics is a fast and promising tool.

The aim of our study is the prediction of the interaction sites from a primary sequence. Our data set contains non-homologous proteins from the Protein Data Bank (PDB). The first subset is constituted of 78 soluble protein complexes from Chakrabarti P.[1], Heifetz[2] and Valdar[3]; the second subset contains 8 membrane proteins and the third subset contains 55 protein-DNA complexes from Mandel-Gutfreund and Margalit[4].

To detect the interacting amino acids, we have considered the variation of solvent Accessible Surface Area (ASA) using Zpex software[5,6]. Interacting amino acids have a maximal distance of interaction of 5Å. Using this selection, we plotted the frequencies of amino acid pairs and amino acid-nucleic acid pairs. We also considered the interacting secondary structures.

Preliminary results suggest that:

- For the membrane proteins, the major interactions occur between leucines and those belong mainly to alpha helix.
- For the globular proteins, most interactions are between leucine and leucine, leucine-phenylalanine and leucine-valine when the secondary structures implicated are mainly beta-sheets.
- For the amino acid-base interactions, the main amino acid is arginine, in interaction with guanine or thymine. The interacting arginines are in the alpha helix structure.

1. Chakrabarti P., Janin J., (2002) Proteins 47, 334-343.
2. Heifetz A., Katchalski-Katzir E., Eisenstein M., (2002) Protein Science 11, 571-587.
3. Valdar W.S.J., Thornton J. M., (2001) Proteins 42, 108-124.
4. Mandel-Gutfreund Y., Margalit H., (1998) Nucleic Acids Research 26 (10), 2306-2312.
5. Thomas A., Bouffioux O., Geeurickx D., Brasseur R., (2001) Proteins 43(1), 28-36.
6. Thomas A., Benhabiles N., Meurisse R., Ngwabije R., Brasseur R., (2001) Proteins 43(1), 37-44.

Poster n°8