Design and use of a non-redundant database of three-dimensional protein structures

Claire Schalon, Pascal Muller, Jean-Sébastien Surgand, Esther Kellenberger, Didier Rognan
CNRS UMR 7175 – LC1 Pharmacochimie de la communication cellulaire
Faculté de Pharmacie – Université Louis Pasteur

Using relationship between families of protein targets and sets of similar ligands is a recent approach for the rational conception of drugs. For this purpose, databases need to be available: it includes many databases of ligands but also specialized databases of target protein structures. There is only few of the last ones adapted to the topic.

The screening-PDB (sc-PDB) developed in the lab[1] contains 6415 3D structures of active sites (version may 2005) extracted from the Protein Data Bank (PDB)[2]. The sc-PDB is focused on binding sites of complexes that accommodate drug-like molecules. The entries have already been used for docking and virtual screening. Their detailed annotation facilitates the analysis of results generated.

A supplementary subset of the sc-PDB has been extracted in order to better fit the purpose of studying the relationships between proteins families and ligands chemotype. This collection of Non-Redundant Sites (cNRS) have been created to regroup a maximum of diversity for a minimum number of entries. This set contains 1437 entries selected after two steps: 1) some filters have removed the less appropriate entries; 2) two representatives of each functional class have been selected.

A first analysis of the cNRS has shown that the entries do not present sequences redundancy but the non-redundant subset will mostly be used to validate a program of structural alignment developed in the lab. The retrieval of proteins with similar binding sites and weak sequence identity will enable to propose new ligands for target proteins.

1 Kellenberger E., Muller P., Schalon C., Bret G., Foata N. and Rognan D. (2006) J. Chem. Inf. Model. 46, 717-727

2 Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) Nucleic Acids Res 28, 235-242