A New Structural approach for Molecular Similarity and Diversity in Chemoinformatics

Ana G. Maldonado, Florent Barbault, Michel Petitjean, Jean-Pierre Doucet, Bo-Tao Fan*

ITODYS, Université Paris 7 - Denis Diderot, CNRS UMR-7086.
1 Rue Guy de la Brosse, 75005, Paris, France.
* Corresponding author email: fan@paris7.jussieu.fr

Abstract. The molecular *similarity* provides a powerful method for virtual screening and underlies the use of clustering methods on chemical databases. Furthermore, molecular *diversity* analysis explores the way of molecules to cover a determined structural space and is the basis of many approaches for compound selection and design of combinatorial libraries. The choice of an optimal metric space that represents bitterly the structural diversity, as well as, the descriptors which express the chemical reality, are determinant in the efficiency of the model.

Fragment approaches have been shown to be efficient in molecular analysis for virtual screening. Some key findings reported in recent investigations about the use of fragment approaches in molecular similarity and diversity for HTS are selectively highlighted and summarized in our recently published review [1]. A such careful study led us to propose a new fragment approach for molecular similarity and diversity analysis. This method is essentially based on vectorization of structural information combined with expected physicochemical properties. Although our approach is mainly centred in chemoinformatics, applications could cover other areas such as pharmaceutical and medical chemistry, combinatorial chemistry and chemical databases.

[1] Ana Maldonado, Jean-Pierre Doucet, Michel Petitjean, and Bo Tao Fan. *Molecular similarity and diversity in chemoinformatics: from theory to applications.* Molecular Diversity. 10: 39-79 (2006).