Conformational sampling & Docking by a GRID-deployed genetic algorithm

Alexandru-Adrian Tantar[#], *Sébastien Conilleau*[♦], *Benjamin Parent*[♣], *Nouredine Melab*[#], *Lorraine Brillet*[◊], *Sylvaine Roy*[◊], *El-Ghazali Talbi*[#] *& Dragos Horvath*[♦*]

INRIA Lille, 40 Parc Scientifique de la Haute Borne, 59655 Villeneuve d'Ascq;[♣]ISEN Lille, 41 Bd. Vauban, 59046 Lille, France; [◊] CEA, DSV, iRTSV, CMBA, 17 rue des Martyrs, Grenoble, F-38054, France; [♦]Laboratoire d'Infochimie, Université Louis Pasteur, 4, rue Blaise Pascal, 6700 Strasbourg, France;* corresponding author: horvath@chimie.u-strasbg.fr

Abstract :

Conformational sampling and Docking are highly challenging problems, both due to the large number of degrees of freedom and the extreme ruggedness of the geometry-energy landscape that needs to be explored in search of low-energy conformations. Therefore, nature-inspired optimization and sampling techniques, such as Genetic algorithms (GA) are well suited to address such problems. Inspired from the Darwinian evolutionary theory, GAs are very well suited for parallelization on computer grids. State-of-the-art conformational sampling and docking programs typically perform a very limited exploration of the problem space and, unsurprisingly, need to rely on empirical correction schemes. Such empirical machine-learned correction functions cause docking procedures to behave like typical QSAR models, and fail beyond the intrinsic applicability ranges for which the validity of the docking score has been established. Proper, extensive conformational sampling accounting for both ligand and active site flexibility should, in principle, allow the obtention of ligand affinities without the need of additional empirical scoring.

The herein developed methodology as part of the ANR project "Docking@grid" aims to exploit the massively parallel computing resources in order to provide an in-depth exploration of the phase space of conformational sampling and flexible docking involving several hundreds of degrees of freedom – including, for example, the *ab initio* folding of small proteins or the specific analysis of key degrees of freedom in large structures – and to asses whether, and how, such in-depth knowledge of the potentially populated states may help improving the accuracy and robustness of molecular property prediction.