AlgoGen: A tool coupling a linear-scaling quantum method with a genetic algorithm for exploring non-covalent interactions

C. Barberot, J.C. Boisson, S. Gérard, H. Khartabil, E. Thiriot, G. Monard, E. Hénon

Abstract

AlgoGen-DivCon is a program that greatly benefits from algorithmic advances in quantum chemistry. It was initially designed to perform rigid molecular docking in order to ultimately pose a ligand in the receptor site by combining the Divide and Conquer linear-scaling quantum-chemistry method with a genetic algorithm (GA). A new version of this program with several enhancements is presented, interfaced with MOPAC/MOZYMEx. A biological application on seven docking structures leads to a pose in good agreement with known crystallographic structures. But, more generally, AlgoGen can explore intermolecular potential energy surfaces without preconceived idea, what yields an alternative use of this program. This feature was employed to investigate the possible presence of computational artefacts on the semi-empirical PM6-DH+ potential energy surfaces (PES) of 22 relative small complexes. For all dimers, the PM6-DH+ PES features a minimum geometry almost identical to the high-level reference equilibrium geometry. This method is found to perform remarkably well in predicting properties of hydrogen bonded complexes. However, in addition to the expected minima, false positive structures associated with well-characterized minima on the PES were identified for the ammonia and water dimers. Detection of these artefact makes AlgoGen PES scans an interesting tool for semi-empirical method developments aiming at reproducing non-covalent interactions and their evaluation. Additionally, a complementary post-treatment using NCI analysis turns out to give significant insight into chemical weak interactions found by AlgoGen.

1. Introduction

In this paper, we address the potential of the AlgoGen software [1] in the context of computational studies involving intermolecular interactions. Since initially designed to perform rigid molecular docking using quantum chemistry evaluation, a first focus is on intermolecular interactions in the framework of biological molecular recognition. Several enhancements to the AlgoGen algorithm are described. Additionally, the potential of the AlgoGen algorithm to explore intermolecular situations without preconceived idea is employed to investigate the possible presence of computational artefacts on the semi-empirical PM6-DH+ potential energy surface (PES) of small complexes.

Molecular modelling tools are widely used to search for new drugs in medicinal and pharmaceutical fields. Drugs are usually (relatively) small organic molecules called «ligand» and mostly produced their effect by interacting with a biological macromolecule such as protein. Among several molecular modelling techniques used to examine the molecular recognition, molecular docking is employed to describe the best orientation of a ligand that binds to a particular protein of interest [3]. Such programs sample the orientation space of a small molecule into the receptor site by evaluating the favourable docking structures. The simplest algorithms consider the two molecules as rigid bodies and then six degrees of freedom: three translational and three rotational, are explored leading to an optimization problem over a six dimensional space. The receptor is practically always assumed to be rigid but attempts to incorporate (partly) the internal conformational space of the ligand to the optimization problem lead to what is called the «flexible docking» technique, increasing the solution space.

To identify the best orientation of a small molecule into a biological receptor by molecular docking a scoring function is needed. It reflects the strength of the protein–ligand association. Also, the scoring function has to be able to rank a ligand relative to another. Most of the scoring functions attempt to approximate the binding free energy of the complex: the better score, the lower free energy of interaction. The prediction of the scoring function is a major challenge. Because the search space is usually large, most of these functions are based on very simple descriptions, either using an