GaudiMM: A Modular Multi-Objective Platform for Molecular Modeling

Jaime Rodríguez-Guerra Pedregal,* Giuseppe Sciortino, Jordi Guasp, Martí Municoy, and Jean-Didier Maréchal*

GaudiMM (for Genetic Algorithms with Unrestricted Descriptors for Intuitive Molecular Modeling) is here presented as a modular platform for rapid 3D sketching of molecular systems. It combines a Multi-Objective Genetic Algorithm with diverse molecular descriptors to overcome the difficulty of generating candidate models for systems with scarce structural data. Its grounds consist in transforming any molecular descriptor (i.e. those generally used for analysis of data) as a guiding objective for PES explorations. The platform is written in Python with flexibility in mind: the user can choose which descriptors to use for each problem and is even encouraged to code custom ones. Illustrative cases of its potential applications are included to demonstrate the flexibility of this approach, including metal coordination of multidentate ligands, peptide folding, and protein-ligand docking. GaudiMM is available free of charge from https://github.com/insilichem/gaudi. © 2017 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.24847

Introduction

Molecular modeling has become a major tool in many fields of chemistry and its interfaces. Its final objective is to accurately describe the structural and energetic properties of molecular systems; ultimately from scratch. Most of nowadays computational chemistry exercises need to start with clear structural data on the system of interest (i.e., X-ray or NMR structures). Unfortunately, this first piece of information is frequently missing or highly incomplete and in fact many projects strive to start because of the complexity to find convenient starting points in a reasonably short amount of time. When modelers try to overcome this "blank sheet" syndrome, they generally go through an iterative process of trial and error and manual adjustments, with the only guidance of his or her chemical intuition and/or experimental observations.

Most of the procedures used to rapidly identify physically sound initial models of a molecular system stand on finding the way to ally the exploration of wide conformational spaces and the adequate guiding variables. Among the most frequent tools for this task are Monte Carlo,[1,2] Random Walk,[3] Simulated Annealing,[4–6] and evolutionary algorithms (EA).[7,8] In those approaches, the way to bias the exploration generally stands on (1) using energetic evaluation of the molecular geometry (i.e., force field) and (2) impose additional simple Euclidian restraints like distances, angles, or dihedrals that could account on structural aspects hypothesis. The construction and assessment of an initial molecular candidate is therefore generally limited to potential energy surface (PES) considerations, eventually accounting for complex force field parametrization, and few guiding elements. However, there is much more structural information that the researchers could account on that are generally neglected at the moment.

Genetic Algorithms (GA)[9] are a kind of EA that have been increasingly applied for complex molecular problems such as molecular matching,[10,11] protein-ligand docking,[12,13] or conformational searches.[14,15] The most popular implementation in these applications are GAs that use a single objective strategy with a unique fitness function targeted. However, for years now, multi-objective genetic algorithms (MOGA) like NSGA[16–18] and SPEA[19–21] families are readily available and could bring novelties in the way we deal with molecular modeling. MOGAs are particularly helpful when different variables of the system fight ones against others to reach a tradeoff, especially if their importance is not known beforehand—a prototypical situation in initial model building. MOGAs tend to be applied when substantially different solutions could exist for the same problem. In principle, this kind of approaches could be useful on complex molecular modeling exercises when only partial information is available at the starting point of the study. The potential of multi-objective optimization in molecular modeling has been demonstrated by some recent developments,[22,23] but there is plenty of room for advances of MOGA applications in Molecular computational chemistry and more particularly in providing a modular platform able to deal with relevant molecular descriptors.

Here, we present GaudiMM (for Genetic Algorithms with Unrestricted Descriptors for Intuitive Molecular Modeling), a GA-based platform for 3D sketching of molecular systems that expands the idea of PES guiding using molecular objectives.

J. R.-G. Pedregal, G. Sciortino, J. Guasp, Martí Municoy, J.-D. Maréchal Departament de Química, Universitat Autónoma de Barcelona, Cerdanyola del Vallès, Barcelona 08193, Spain E-mail: jaime.rodriguezguerra@uab.cat E-mail: jeandidier.marechal@uab.cat

Contract grant sponsors: Spanish MINECO (project CTQ2014-54071-P), Generalitat de Catalunya (project 2014SGR989; to J.R.G.P.), and UAB (to G.S.)