LVOS-DISTEVOL: a distributed, evolutionary computing system for ligand virtual optimization and screening

Piastra M[(1)], Spallarossa A[(2)], Baracchi S[(1)], Di Bernardo D[(1)], Rosano C[(3)]

[(1)] Dipartimento di Informatica e Sistemistica, via Ferrata, 1 – 27100 PAVIA
[(2)] Dipartimento di Scienze Farmaceutiche, viale Benedetto XV, 3 – 16132 GENOVA
[(3)] S.C. Nanobiotecnologie, Istituto Nazionale per la Ricerca sul Cancro (IST), Largo R. Benzi, 10 – 16132 Genova

Motivation

Automated docking simulations that calculate the bioactive conformation(s) of flexible ligands to macromolecular targets are becoming increasingly sophisticated and relevant for the design of new drugs. The vast majority of existing software tools generate, either systematically or via random probing, a number of docking poses that are ranked according to specific energy scoring functions. A typical virtual screening process involves the in-silico evaluation of a compound collection against a specific target; the compound set should include a large number of chemically diverse entities so that the probability of finding viable solutions to the specific drug-design problem being investigated is enhanced. In order to provide a valuable computational tool to support the design and the chemical modification of new molecules endowed with pharmacological properties, the LVOS-DISTEVOL system was developed. This system uses evolutionary computation techniques for the incremental generation of a vast collection of ligands which are evaluated by interacting, in a parallel and distributed fashion, with multiple instances of a third-party docking tool. These collected evaluations drive the overall stochastic optimization process.

Methods

The main software component of LVOS-DISTEVOL is a centralized server that runs a steady-state evolutionary process on a population of candidate ligands. The system is networked with a number of ‘helpers’, individual components each controlling an instance of the docking tool adopted. Helpers are only loosely coupled with the server: they first obtain their setup parameters, including the description of the target, then they iteratively obtain one candidate ligand at time and report back the results in a stateless protocol. As a result, helpers can be connected and disconnected at any time and the server is virtually impervious to their failure. Helper integration modules have been developed for AutoDock 4 and eHiTS 6. The genetic operators of LVOS-DISTEVOL are fundamental for shaping the ligand optimization space. Candidate molecules in LVOS-DISTEVOL have a two-level representation: an atomic-level microstructure and a macrostructure composed of rigid fragments tied together by flexible chains. Genetic operators are also divided into two categories: macro operators, that recombine and mutate molecules as a structure of macro-blocks, i.e. either rigid fragments or flexible chains, without introducing microscopic changes; micro operators, that act on individual rigid fragments or flexible chains by introducing small changes that – in the designers’ intention – are typical of a ‘manual’ optimization procedure. The micro operators implemented are divided into three further categories: 1) modifications of cycloalkyl rings; 2) modifications of aromatic systems; 3) modifications of carbon linkers. Micro operators are also associated to pre-conditions, which determine their applicability. Altogether, they also must obey a few systemic requirements in order to avoid biasing the process: a) reversibility – each micro operator must have a reverse; b) balance of population statistics – for instance the combined effect of the operators should not alter per se the percentage of heteroatoms in a heterocycles in the population. The chosen strategy for defining the initial population of ligands in LVOS-DISTEVOL is hybrid: one part of the initial molecules is defined explicitly, while another is defined only implicitly, as a library of selected rigid fragments and flexible chains, which are recombined at random to create further initial candidates. The fitness function for the evolutionary process is based on the inhibition constant, as estimated by the docking tool for the best docking pose. In addition, in order to lower the fitness of molecules with unwanted features, a selective pressure is also introduced. For instance, an extra weight may represent the amount by which the Lipinski’s rules are violated by the candidate molecule.

Results

Apart from the platform requirements of the third-party docking tools adopted, LVOS-DISTEVOL is a platformindependent Java software system. In its current implementation, the system is more a framework than a standalone product, as both the set of micro operators and the fitness function typically require some tuning that, depending on the experiment, may entail minor changes to the source code. The validation experiments performed, with several nodes working in parallel, have produced very interesting results. In all the comparative runs, with HIV-1 Reverse Transcriptase (RT) and an initial population created from known 33 inhibitors selected from the PDB, the system has detected plausible molecules with extremely low, estimated inhibition constants (in the order of 10E-11 M). The critical aspect emerged, which requires some further investigations, is the need to improve on the selective pressure in order to detect molecules with more sophisticated expected properties.

Contact : marco.piastra@unipv.it