Applying Model Trees on Flexible-Receptor Docking Experiments to select promising protein receptor snapshots

Karina S. Machado1*, Ana T. Winck2*, Duncan D. A. Ruiz2, Osmar Norberto de Souza1§
1 LABIO - PPGCC, Faculdade de Informática, PUCRS, Av. Ipiranga, 6681 – Prédio 32, sala 602, 90619-900, Porto Alegre, RS, Brazil
2 GPIN - PPGCC, Faculdade de Informática, PUCRS, Av. Ipiranga, 6681 – Prédio 32, sala 628, 90619-900, Porto Alegre, RS, Brazil

Background
In-silico based Rational Drug Design (RDD) [1] is a four-step cycle that combines structural information and computational efforts based on a detailed understanding of the target protein (or receptor) and ligand interactions [2]. The most important step in RDD is the molecular docking simulation that investigates the affinity between protein receptors and ligands. Typically, molecular docking algorithms consider receptors as rigid bodies. Receptors are, however, intrinsically flexible in the cellular environment. There are several ways to address the receptor flexibility [3]. In this work we model the full receptor explicit flexibility in molecular docking simulations by using different conformations of the receptor generated by a molecular dynamics (MD) simulation.

The problem is that this approach is extensively time-consuming. Hence, selection of the most promising conformations can accelerate docking experiments and, consequently, the RDD efforts.

Results
We previously docked two ligands (NADH and TCL) to 3,100 snapshots of the InhA receptor from M. tuberculosis [4] using a scientific workflow [5]. Based on the receptor residues-ligands distances we preprocessed all docking results stored on a proper database [6] to generate appropriate input for data mining using the M5P model tree algorithm [7]. The predictive attributes were the shortest inter-atomic distance between the ligand and the receptor’s residues for each docking result and the target attribute was the estimated free-energy of binding (FEB) value. The mining inputs were submitted to M5P algorithm. These models produced short and understandable trees. On the basis of the correlation values of the produced model trees, NADH and TCL obtained more than 95% correlation. Post processing the generated model trees, for each linear modelo (LM), we calculate the average FEB for their associated instances. From these values we considered a LM as representative if its average FEB is smaller or equal than the average FEB of the test set. The instances in the selected LMs were considered the most promising snapshots. It totalized 1,521 and 1,780 snapshots for NADH and TCL respectively.

Conclusions
By post processing the generated model trees we were able to propose a criterion of selection of LMs which, in turn, is capable of selecting a set of the most promising receptor conformations. As future work we intend to select another compounds, among the million catalogued, that may be promising as new drugs to be tested considering the selected receptor conformations.

References