Development of Empirical Scoring Functions for Predicting Protein-Ligand Binding Affinity

Guedes, I.A.[1]; Barreto, A.M.S.[1]; Miteva, M.A.[2], Dardenne, L.E.[1]

[1] *GMMSB, LNCC/MCTI, Petrópolis, Brazil*, [2] *Molécules Thérapeutiques in silico, Paris Diderot – Paris VII, Paris, France*

INTRODUCTION. DockThor has obtained promising results in comparative studies with other well established docking programs for predicting experimental binding mode. Despite useful for pose prediction, the current scoring function implemented in DockThor is not suitable for predicting binding affinities. **OBJECTIVES:** In this work, we develop several scoring functions from first-principles with interaction-driven features for predicting binding affinities of protein-ligand complexes. **MATERIALS AND METHODS:** We developed general and specific scoring functions for target-classes, the last to account for binding characteristics associated with a target class of interest, focusing on proteases, kinases and protein-protein interactions complexes (PPIs). The scoring functions were derived using linear regression (MLR) and seven machine learning techniques for nonlinear problems using the PDBbind refined set 2013 (N = 2959) for training and testing. We also trained and evaluated general scoring functions using docking results obtained with DockThor. **DISCUSSION AND RESULTS:** The linear and the best nonlinear (random forest) general scoring functions trained with experimental structures obtained great performances when evaluated on the benchmark PDBbind core set 2013 (N = 195) (R_{MLR} = 0.602 and R_{RF} = 0.704). These results demonstrated that our scoring functions are competitive with the best scoring functions evaluated in such benchmarking studies, i.e. X-ScoreHM (linear) with R = 0.644 and RF::VinaElem (nonlinear) with R = 0.752. The scoring functions specific for target classes also obtained good performances on independent test sets: R_{SVM} = 0.723 for proteases, R_{LWL} = 0.702 for kinases and R_{SVM} = 0.613 for iPPIs. Furthermore, the scoring functions trained with docking results obtained promising performances when evaluated in both experimental and docking structures, indicating that they are reliable to be applied in both cases. **CONCLUSIONS:** The development of the scoring functions implemented in this work is a crucial step to make the DockThor a even more competitive program, encouraging the development of the virtual screening program and portal DockThor-VS.

Keyworks: binding affinity prediction, molecular docking, scoring functions.
Supported by: CNPq and CAPES.