neural networks, support vector machines, decision trees) is combined into a consensus predictor that surpasses each individual method in prediction accuracy. The framework is benchmarked with respect to robustness of the descriptor optimization scheme, enrichment of active compounds in virtual screening, and scaffold hopping capability.

Allosteric modulation of subtypes of the human metabotropic glutamate receptors (mGluRs) are attractive strategies for treatment of neurological disorders and diseases. Specifically we report our results in applying the cheminformatics framework to potentiators of mGluR5 (treatment strategy for schizophrenia), potentiators of mGluR4 (treatment strategy for Parkinson's disease), and negative allosteric modulators of mGluR5 (treatment strategy for fragile X syndrome). For all systems a high-throughput calcium mobilization assay screen of ~150,000 compounds serves as input for the machine learning approach. The final models were utilized to virtually screen external compound libraries and obtain a total of 2,682 compounds for testing based on the highest predicted potency values (EC_{50}). A robust descriptor set was optimized for each dataset, and the enrichment factors in these compound libraries ranged from 7.4 for negative allosteric modulators of mGluR5 to 30 for positive allosteric modulators of mGluR5. In addition, the enrichment factor for mGluR4 potentiators was found to be 21.

In a second experiment, the scaffold hopping capacity of the framework was determined: A diverse library of 3,147 compounds was ordered based on the predictions by the models. Analyzing the chemical structure of the predicted efficacious compounds has led to the discovery of four new scaffolds for mGluR4 potentiators and one novel scaffold for negative allosteric modulators of mGluR5. Two scaffolds in the mGluR4 potentiators result set were similar to pharmaceutical leads, which were not represented in the training data set.

Improving molecular docking through tunable scoring function: application

Orr Ravitz, Zsolt Zsoldos, and Dan Harris, SimBioSys

The molecular docking paradigm, has thus far failed to produce a generic approach that would deliver accurate pose prediction capabilities, and reliable rank-ordering of conformations and ligands consistently for any biological system of interest. This reality, which has been addressed by numerous methodology papers and comparative studies, has been largely attributed to the inability of scoring functions to capture different chemical interaction types at a uniform level of accuracy. Several studies attempted to develop guidelines for choosing the most suitable docking and scoring method for a specific problem based on protein family classification of the target, dominant interactions, and other properties of the studied system. Consensus techniques, on the other hand, try to synergistically integrate information from multiple sources assuming agreement between different methods is indicative of more accurate values. Both approaches, however, have shown only limited success in improving binding mode and activity prediction capabilities.

An alternative solution, and arguably a more rigorous one, would be to tailor the scoring function for the system of interest. eHiTS, with its knowledge-based approach, offers users the capability to fine-tune the scoring function using their data and thus incorporate their full body of knowledge in a systematic and automatic fashion. In many realistic drug discovery scenarios, structural and ligand -activity information is sufficient in a statistical sense to adjust a limited set of parameters representing the relative weights of the various terms in the eHiTS scoring function. In this talk, an overview of the eHiTS' tuning utility will be given, outlining the implementation of scoring function weight training techniques employing multiple crystal structures of the target, complexed with known active ligands, and associated binding affinity data. Acetylcholine binding protein (surrogates of nicotinic acetylcholine receptors), beta secretase and other systems of pharmaceutical interest will be used to demonstrate the improvement in docking performance in terms of score discrimination between low and high RMSD poses, of enrichment levels in screening runs, and of correlation between score and binding affinity. Guidelines for choosing the optimal data set for training will be discussed.

MOLA: a tool for automation of parallel virtual docking using AutoDock Vina in a heterogeneous set of computer clusters

Rui M.V. Abreu[a,b,], Hugo Froufe[a], Maria João R.P. Queiroz[c], Isabel C.F.R. Ferreira[a]*

[a]CIMO-ESA, Instituto Politécnico de Bragança,. [b]Instituto de Biotecnologia e Bioengenharia, Centro de Genética e Biotecnologia, Universidade de Trás-os-Montes e Alto Douro. [c]Centro de Química, Universidade do Minho

The use of molecular docking to search large databases of compounds for possible ligands of a protein receptor is usually termed virtual screening and has been successfully applied in several therapeutic programs at the lead discovery stage. However, large scale virtual screening is time demanding and usually requires dedicated High Performance Computing (HPC) systems. We present MOLA, an easy-to-use graphical user interface tool that automates parallel virtual screening using AutoDock Vina as docking engine. AutoDock Vina is a new software for molecular docking and virtual screening that uses a gradient optimization method in its local optimization procedure. MOLA automates several tasks including: ligand preparation, parallel run distribution, result analysis and ligand ranking. The HPC system (computer cluster) is set up using a customized version of Pelican HPC operating system that boots from the CD and clusters any kind of non-dedicated heterogeneous computers available. All input files and output results are automatically recorded on a USB-flash drive. With a cluster of 10 computer processors, the parallel algorithm of MOLA performed with a speed-up 8.60 using AutoDock Vina. MOLA converts these computers into a HPC cluster that can be used for parallel virtual screening. MOLA is not intended for large clusters rather his main strengths are: (1) ease-of-use for users with little knowledge on Linux systems, (2) the facility to integrate a very heterogeneous set of computers and (3) the ability to use non-dedicated desktop computers.