Selecting structure-based virtual screening hits using chemoinformatics tools: a case study with HIV-1 reverse transcriptase

LH Santos[1], RS Ferreira[2], ER Caffarena[1]

[1]*Fundação Oswaldo Cruz, Programa de Computação Científica, Rio de Janeiro, RJ, Brasil*
[2]*Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas, Belo Horizonte, MG, Brasil*

Reverse transcriptase (RT) is a major drug target for the treatment of HIV infection. RT is inhibited by two inhibitors classes: the nucleoside RT inhibitors (NRTI) and the non-nucleoside RT inhibitors (NNRTI). The NRTIs, when bound to RT active site, obstruct the conversion of single-stranded RNA to double-stranded DNA provirus, which is then integrated into the human genome. Whereas, NNRTIs, bind to an allosteric site, causing conformation changes that impair DNA synthesis. Despite the large number of drugs targeting HIV-1 RT, problems like resistance, toxicity, and especially mutations turn the development of more effective and less toxic inhibitors an urgent matter. Due to the availability of several crystal structures, structure-based virtual screening might be a preliminary effort to rational drug design of RT inhibitors. However, after the screening of a compound library through molecular docking, there is still the task of analyzing and interpreting hundreds to thousands of ranked compounds. To this task, chemoinformatics tools use unique representations of chemical structures in the form of descriptors, utilizes metrics of similarity, and apply statistical and other techniques to establish relationships between chemical structures and their properties. Therefore, this study aimed to identify a subset of hit compounds after applying chemoinformatics tools to the virtual screening outcomes of an extensive compound library using the non-nucleoside binding pocket (NNBP). The lead-like now ligand subset from the ZINC database was chosen as the screening library, at the time composed of 2,797,315 compounds. All compounds were submitted to molecular docking using DOCK6.6. The use of molecular docking method ensured that the subsequently chosen compounds, would fit, and possible have interactions with the NNBP. The screening was performed using an RT structure (PDB: 4G1Q) bound to the known NNRTI, rilpivirine. This structure achieved the best performance amongst ten others in the prior assessment of DOCK6.6. Among the 50% top-scoring compounds (GridScore of -40.52 kcal/mol or lower), we chose two samples of 5,000 compounds each. One sample containing compounds ranked by the lowest (more favorable) GridScore, and another containing randomly selected compounds to explore the possibility of finding out suitable candidates in the remaining database. With the help of the R package ChemmineR, clustering of the sets was done to identify discrete similarity groups using the binning clustering function with Tanimoto coefficient at a cut-off of 0.55. In the end, 68 and 86 chemical diverse compounds were obtained from each subset, respectively. To further filter the remaining compounds, we selected only compounds with one or more specific interactions with important NNBP amino acids, known to interact with known inhibitors. Therefore, we were able to propose a more manageable number of hit candidates for further testing from the large initial screening library, making use of valuable metrics to ensure the chemical diversity of ranked compounds by docking.

Supported by: Fiocruz, Capes.