The Combination between Artificial Intelligence and Molecular Docking on the Prediction of Natural Substrates Proteins

Dias, M. F. R.[1], Chagas, M. S.[1,2], da Silva, M. L. [1], Pascutti, P. G.[1,2]

[1]Grupo de Biologia Computacional, Diretoria de Metrologia Aplicada às Ciências da Vida – Dimav. Instituto Nacional de Metrologia, Qualidade e Tecnologia - Inmetro, Duque de Caxias/RJ, Brazil.
[2] Universidade Federal do Rio de Janeiro, Brazil.

Viral diseases affect populations over the globe. The use of computational methods linked to protein information for patterns' discovery in biological systems allows the screening and detection of new drug prototypes from compound data banks. The present work uses natural substrate of HIV-1 protease and computational methods to find compounds which show the interaction probability of HIV-protease and the potential to inhibit it. The linear sequences of the natural substrates were numerically represented by physicochemical properties (AAINDEX) and classified by the SVM algorithm in WEKA tool such as the potential to inhibit the protease of HIV-1. The three-dimensional structures of these substrates were obtained by Modeller software. The molecular docking with the Lamarckian algorithm simulated in AutoDockTools package. The peptides of the gag-pol polyproteins were represented by 31 physicochemical properties which were fractionated on 8, 6, 5 and 4 amino acids and used as positive and negative sets corresponding to the cleavage sites that do and do not interact with HIV-protease. The results of the cross-validation of the type 10-k-fold showed that 66% of the correct classification, had precision and that the recall ranging were from 0.63%-0.67% for peptides of different sizes. Subsequently, a subset of peptides, classified as positive and negative underwent a molecular docking simulations and analyzed using docking energy values. The machine learning results showed no significant differences in the classification of different size peptides, but the literature assumes that smaller substrates have great potential to inhibit the enzyme HIV-1 protease. The docking results showed that the peptide size, structure, position and amino acids that compose it are essential to find better interaction with the enzyme. The use of machine learning tools' coupled with molecular docking techniques allowed the data analysis in their linear and three-dimensional shapes and has provided a greater reliability in the prediction method.

Key Words: HIV protease, physicochemical properties, artificial Intelligence, molecular docking

Acknowledgements: CAPES-Biocomputacional

*email: marfedias@gmail.com