

Title:

Implementation of the PTML-LDA Model in the discovery of drug candidates for inhibitors of viral diseases of the flaviviridae family.

Author:

Yendrek Andres Velsquez López; Sonia Arrasate Gil

Abstract:

Currently, diseases transmitted by the aedes aegypti mosquito in tropical areas have become a great risk to public health in these areas, within these diseases we find members of the Flaviviridae family, the Dengue Virus is the most known among them. Currently, there are other members such as: Chikungunya (CHIKV), Yellow Fever Virus (YFV), ZIKA Virus (ZIKV), etc. Although these diseases are not new in the present, there is no specific treatment to deal with them. Several of these diseases are classified as serious by the World Health Organization - WHO, taking hundreds of lives today. From phylogenetic studies it is known that members of this family possess highly conserved sequences making them an optimal target for drug development. The drug discovery model is not something new, it has traditionally been carried out under the exercise of "trial and error" and, given the arrival of new and increasingly resistant diseases, it was necessary develop new methodologies to accelerate this process. Computational chemistry was born from the need to reduce costs, reduce time in drug development and improve the discovery of new compounds, under these three requirements over the years several techniques such as QSAR methods were the fundamental axis of chemoinformatics, With the arrival of the Big Data era, a range of possibilities opens up for the study and development of drugs, such as the implementation of the perturbation theory (PT) and machine learning (ML) models - PTML. Through the use of databases such as ChEMBL it is possible to generate a sufficient data set for the development of a prediction model using PT operators, which are based on moving averages of multiple conditions (Moving Average), which combine the characteristics and simplify data management. In this study, several PTMLLDA prediction models were evaluated based on 47815 tests obtained from ChEMBL, taking as input variables: the reference function, three molecular descriptors: AlogP, MW, TPSA, six test conditions and the interaction between the disturbance conditions and operators. Three different PTML-LDA models were evaluated under different treatments and data processing, the proposed model presents precision values of 77.25%, on the contrary, models N°2 and N°3 did not exceed the 77% range in their training stages. The model was validated by ROC curve obtaining a value of 86.2% indicating that the discrimination is exact and not a random pattern. The proposed PTML-LDA model was selected with a specificity of 75.95% and a sensitivity of 78.88% (see table 9). With these values, the evaluation of the model obtained by using its resulting equation (See Eq. 12) was carried out, compared to 45 new compounds synthesized by our research group, obtaining 8 compounds with high probabilities of presenting activity against this type of diseases.