MOL2NET, International Conference Series on Multidisciplinary Sciences
Insert the title of the workshop here

PTML model of CHEMBL neurological diseases assays vs. protein sequence, and protein interaction networks in different brain regions

Brenda de la Caridad Fundora Ortiz[a], *Sonia Arrasate*[b], *Humbert González-Díaz*[a,b]

[a]*Department of Organic Chemistry II, University of Basque Country UPV/EHU, 48940, Leioa, Spain*
[b]*IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Spain*

.
.
.

Abstract.

Degenerative neurological diseases have become serious risks to human health. These diseases depend on age and are becoming more common today, as the number of older people in society increases. The discovery of new drugs for the treatment of neurodegenerative diseases such as Alzheimer's, Parkison's, and Huntington's diseases, Friedreich ataxia and others is an important goal of medicinal chemistry. For this reason, it is very useful to use the existing public information on preclinical assays with a high number of combinations of experimental conditions to create models that allow predicting new compounds useful for the treatment of these diseases. ChEMBL is a chemical database of bioactive molecules with drug-like properties. This database manages Big Data feature with a complex data set, which is hard to organize. This makes information difficult to analyze due to a big number of characteristics described in order to predict new drug candidates for neurodegenerative diseases. In this context, we propose to combine perturbation theory (PT) ideas and machine learning (ML) modeling to solve this combinatorial-like problem. The PT operators used are founded on multi-condition moving averages, combining different features and simplifying the difficulty to manage all data. For the construction of this model, the structure of the drug, the sequence of the proteins with which these drugs interact, the protein interaction network and the brain region in which these proteins are expressed were considered. The bondaring conditions that were taken into account were: the activity of the drug, the cell line in which the drug was tested, the brain region and the test organism. The developed PTML model reached considerable values in sensibility (80.89% for training and 80.94% for validation), specificity (80.18% for training and 80.33% for validation), and accuracy (80.25% for training and 80.39% for validation). We can conclude that this PTML model is the first one that can predict the activity of drug candidate compounds against degenerative neurological diseases taking into account the structure of the drug,

the sequence of the proteins with which these drugs interact, the protein interaction network and the brain region in which these proteins are expressed.

References

.Fundora-Ortiz, B.; Arrasate, S.; González-Diaz, H. Master's Thesis: PTML model of CHEMBL neurological diseases assays vs. protein sequence, and protein interaction networks in different brain regions. **2020**

.
.