

Predicting Proteasome Inhibition using Atomic Weighted Vector and Machine Learning

Yoan Martínez-López^{a,b}, Efrain Edgar Chaluisa Quishpe^a, Yaile Caballero^{a,b}, Stephen J. Barigye,^c Francisco Torrens^d and Juan Alberto Castillo-Garit^e

^aDepartment of Computer Sciences, Faculty of Computer Sciences, Camagüey University, Camagüey city, 74650, Camagüey, Cuba

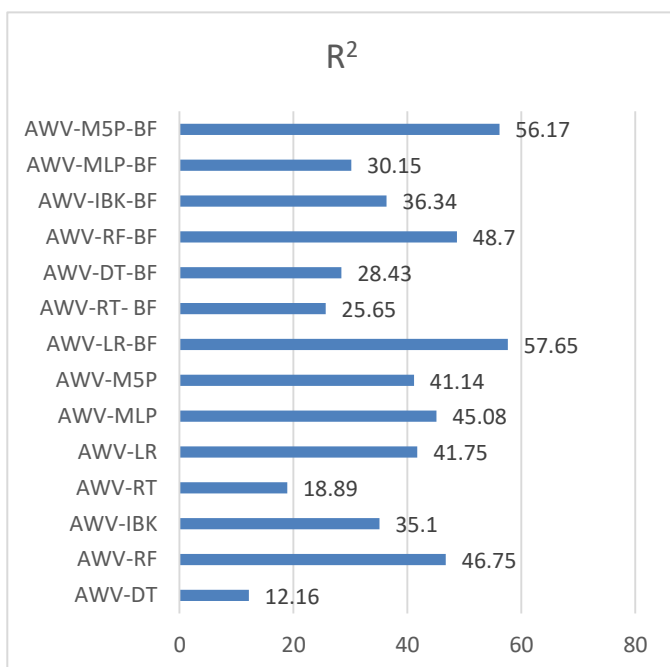
^bArtificial Intelligence Researches Group (AIRES), Faculty of Computer Sciences, Camagüey University, Camagüey, Cuba

^cDepartment of Chemistry, McGill University, 801 Sherbrooke Street West, Montreal, QC H3A 0B8, Canada

^dInstitut Universitari de Ciència Molecular, Universitat de València, Edifici d'Instituts de Paterna, P.O. Box 22085, E-46071, València, Spain;

^eUnidad de Toxicología Experimental, Universidad de Ciencias Médicas de Villa Clara, Carretera a acueducto y circunvalación, Santa Clara, Villa Clara, Cuba. CP: 50200, Cuba.

Graphical Abstract



Abstract.

Ubiquitin/Proteasome System (UPS) is a highly regulated mechanism of intracellular protein degradation and turnover. Through the concerted actions of a series of enzymes, proteins are marked for proteasomal degradation by being linked to the polypeptide co-factor, ubiquitin. The UPS participates in a wide array of biological functions such as antigen presentation, regulation of gene transcription and the cell cycle, and activation of NF- κ B. Some researchers have applied QSAR method and machine learning in the study of proteasome inhibition ($EC_{50}(\mu\text{mol/L})$), such as: the analysis of proteasome inhibition prediction, in the prediction of multi-target inhibitors of UPP and in the prediction of protein contact map. Following this idea, we applied the new tool for obtaining molecular descriptor for modeling of proteasome Inhibition $EC_{50}(\mu\text{mol/L})$, in which has used this novel molecular descriptors (MDs) and different classification algorithms for these quantitative structure-activity relationship (QSAR) studies. In the

present research, we use the Atomic Weighted Vector (AWV) as attributes with the objective to develop the QSAR modeling of this datasets and also compare a set of different machine learning (ML) techniques to solve this problem, such as: Linear Regression (LR), Multiple linear regression (MLR), Decision tree(DT), Regression Tree(RT), Random Forest(RF), M5P, K-nearest neighbors (IBK or kNN), Multi-Layer perceptron (MLP), Best-first search (BF) and Genetic Algorithm (GA). The figure shows the results of R^2 of the ML-QSAR using ten- folds cross validation for 258 compounds. The results indicate that AWVs are very important tool for modeling the proteasome inhibitory regardless of the ML algorithm used. It can be suggested that the MD-AWV are suitable for codifying important structural information of the molecules and, thus, constitute an interesting alternative to building effective models for the prediction of the values of EC_{50} ($\mu\text{mol/L}$).

References (mandatory)

- Crawford, L.J., B. Walker, and A.E. Irvine, Proteasome inhibitors in cancer therapy. J Cell Commun Signal 2011. 5(2): p. 101–110.
- Martínez-López, Y., *et al.*, The Summation of Atomic Contributions is an Overly Simplified Characterization of the Holistic Molecular Behavior. Letters in Drug Design & Discovery, 2016. 13(7): p. 12.
- Martínez-López, Y., *et al.*, Prediction of aquatic toxicity of benzene derivatives using molecular descriptor from atomic weighted vectors. Environmental Toxicology and Pharmacology, 2017. 56: p. 314-321.
- Castillo-Garit, J.A., *et al.*, Machine learning-based models to predict modes of toxic action of phenols to Tetrahymena pyriformis. SAR QSAR Environ. Res. 2017, 28 (9), 735–747.
- Martínez-López, Y., *et al.*, State of the Art Review and Report of New Tool for Drug Discovery. Current Topics in Medicinal Chemistry, 2017. 17: p. 1-20.
- Svetnik, V., *et al.*, Random forest: a classification and regression tool for compound classification and QSAR modeling. Journal of chemical information and computer sciences, 2003. 43(6): p. 1947-1958.
- Liaw, A. and M. Wiener, Classification and regression by randomForest. R news, 2002. 2(3): p. 18-22.
- Zhan, C., A. Gan, and M. Hadi, Prediction of lane clearance time of freeway incidents using the M5P tree algorithm. IEEE Transactions on Intelligent Transportation Systems, 2011. 12(4): p. 1549-1557.

- Quinlan, R.J., Learning with Continuous Classes., in 5th Australian Joint Conference on Artificial Intelligence. 1992: Singapore. p. 343-348.
- Cover, T. and P. Hart, Nearest neighbor pattern classification. IEEE transactions on information theory, 1967. 13(1): p. 21-27.
- Prasad, A.M., L.R. Iverson, and A. Liaw, Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems, 2006. 9(2): p. 181-199.
- Hess, K.R., et al., Classification and regression tree analysis of 1000 consecutive patients with unknown primary carcinoma. Clinical Cancer Research, 1999. 5(11): p. 3403-3410.