Molecular Descriptor to Predict Biological Activity of Analogues Cocaine

Luis Puerta[12], Carlos Gonzalez[2]

[1] Departamento de Química, FACYT, Universidad de Carabobo, Apartado 2005, Valencia, Estado Carabobo, República Bolivariana de Venezuela
[2] Chemical and Biochemical Reference Data Division, NIST, Gaithersburg, Maryland 20899, USA

Cocaine is an alkaloid ester extracted from leaves of plants including family Erythroxylaceae[1]. Analogs Cocaine are synthetic drugs with a cocaine-like structure. The biological activity for some analogs has been measured experimentally[2]. Meyers et al[3] used molecular descriptors to characterize three-dimensionality of molecular structure (DCDM): The Principal Moment of Inertia (PMI) [4] and Plane of Best Fit, have found molecular three-dimensionality should be considered in drug design. We used the following descriptors PMI, NPR1 and NPR2, which are obtained for normalized PMI, the Asphericity (Asphe), Eccentricity (Eccen), Radius of Gyration (Rgy), Spherocity Index (ShIn)[5], we considered that last four descriptors are also DCDM, at the same time we used descriptors from electronic structure calculations to obtain molecular descriptors (Quantum descriptor): dipole moments (μ), LUMO energies (E_{LUMO}), and the specific rotation (α) for each Analogues. These descriptors were divided into tree set. First set formed: DCDM, Asphericity (Asphe), Eccentricity (Eccen), Radius of Gyration (Rgy), Spherocity Index (ShIn)[5]. Second set formed Quantum descriptor, and at the end a third set formed by combination of descriptors of the two previous groups. We realized the use of these descriptors set as inputs to aArtificial Neural Network (ANN) for Biological Activity calculation implemented as a multilayer perceptron (MLP), based on Resilient back propagation (Rprop) learning algorithm[4]. Neurons in the MLP layers were distributed in the following way: First layer (inputs) are equal number descriptor depending the case, layer hidden are equal number input - 1, and one layer which output. To validate the effectiveness of different sets formed by previous descriptors, the technical K- Fold Cross Validation was used. The first sets were more efficient in the Biological Activity calculation. On the other hand, through a sample 65 compounds, the network produced poor results only in tree samples. In this case the cross validate score = 0.04, Training error = 0.02 and Validation error = 0.05. The other sets produced poor result. Therefore, the obtained results by applying the first set are promising. Those descriptors suggest that the use might be suitable to predict Biological Activity as a greater number of illicit drugs similar to cocaine.

Reference

1. https://en.wikipedia.org/wiki/Coca
2. Satendra Singh. Chem. Rev. 2000, 100, 925-1024
3. Joshua Meyers, Michael Carter, N. Yi Mok, and Nathan Brown. Future Med Chem. 2016 September; 8(14) : 1753-1767.
4. RDKIT v2018.09.1. Copyright (C) 2018 Manish Sud. All rights reserved.
5. Maya Chem Tools. Copyright (C) 2019 Manish Sud. All rights reserved.
6. Encog Java 3.4 © 2019 by Heaton Research, Inc.