MOL2NET, International Conference Series on Multidisciplinary Sciences
http://sciforum.net/conference/mol2net-03

Multivariate classification of a series of organic compounds of pharmaceutical interest using MODESLAB methodology

<Luis SA Torres Gómez> (luistg@ifal.uh.cu:) [a], <Juan Carlos Polo Vega> (polo@ifal.uh.cu)[a], < Tapiwa Brine Mutsauri >(mutsaurit@gmail.com)[a]<Yaima Garcia Guevara> (yaimag@aica.cu.cu:)[b].
[a] < Department of Pharmacy. Institute of Pharmacy and Foods. University of the Havana >
[b] < Company Laboratories AICA>

.
.
.

Due to the high cost of the development of new active pharmaceutical ingredients and excipients for the pharmaceutical industry, molecular modeling methods have been included in this process more and more frequently in recent years. In this work the calculation of the spectral moments of the matrix of adjacency between edges of the molecular graph with suppressed hydrogens was made, weighted in the main diagonal with different parameters that characterize both the bonds and the atoms in the molecules of compounds of pharmaceutical interest, using the MODESLAB methodology. 91 descriptors related to solubility were calculated, which were used in a training series divided into five groups, according to the priority rules of the IUPAC. With the aim of identifying the descriptors that best discriminate between the compounds of each group and defining the set of functions of these descriptors able to distinguish with the greatest possible precision the members of one or the other group, a discriminant analysis was developed using the stepwise inclusion method using the statistical software IBM SPSS version 22. Four functions were generated that constitute combinations linear of 16 molecular descriptors, which encode both steric and electronic information of the molecules of each group. The functions obtained have a very low minimum Wilks Lambda (0.067) and a high canonical correlation (0.89), which demonstrates their discriminant power and allows the use of the descriptors included in them in future studies of structure-property or structure-activity relationship.

Abstract. Due to the high cost of the development of new active pharmaceutical ingredients and excipients for the pharmaceutical industry, molecular modeling methods have been included in this process more and more frequently in recent years. In this work the calculation of the spectral moments of the matrix of adjacency between edges of the molecular graph with suppressed hydrogens was made, weighted in the main diagonal with different parameters that characterize both the bonds and the atoms in the molecules of compounds of pharmaceutical interest, using the MODESLAB methodology. 91 descriptors related to solubility were calculated, which were used in a training series divided into five groups, according to the priority rules of the IUPAC. With the aim of identifying the descriptors that best discriminate between the compounds of each group and defining the set of functions of these descriptors able to distinguish with the greatest possible precision the members of one or the other group, a discriminant analysis was developed using the stepwise inclusion method using the statistical software IBM SPSS version 22. Four functions were generated that constitute combinations linear of 16 molecular descriptors, which encode both steric and electronic information of the molecules of each group. The functions obtained have a very low minimum

	Wilks Lambda (0.067) and a high canonical correlation (0.89), which demonstrates their discriminant power and allows the use of the descriptors included in them in future studies of structure-property or structure-relationship. activity.

Introduction Due to the high cost of developing new active pharmaceutical ingredients (IFA) as well as excipients for the Pharmaceutical Industry, molecular modeling methods have been applied in recent years. These methods are based on the study of the relationship between the molecular structure of substances and their properties, which include the physical, biological and toxicological chemistries, among others. Solubility, especially in water, is one of the most important properties of the compounds used in the pharmaceutical industry, on the one hand because many times their biological activity is conditioned by this behavior and on the other hand because the success of a formulation also depends largely on measure of said property. The approach called Modeslab (8) has been used to calculate a series of molecular descriptors in order to obtain mathematical models that allow to classify the compounds in different groups according to their chemical structure..

.

.

Materials and Methods The SMILES generated in ChemOffice 2010 were imported using the MODESLAB computer program, to proceed with the calculation of the molecular descriptors (spectral moments) of each compound. The spectral moments of each compound were calculated by weighting the following molecular graphs: bond distance (Std), dipole moments 1 (Dip), hydrophobicity (Hyd), polarization (Pol), atomic radius of Van der Waals (Van) and atomic weight (Ato) With this data, a matrix containing the spectral moments from μ0 to μ15 was obtained for each of the compounds included in the series. MODESLAB generates data in .txt extension files compatible with the Microsoft Office Excel electronic tabulator, from which a database was built using the statistical software IBM SPSS version 22 for Windows. To obtain the function that allows the classification of the series of organic compounds included in the work according to structural characteristics related to the solubility, it was developed, by means of the statistical software IBM SPSS, the Linear Discriminant Analysis

.

.

.

Results and Discussion Once the molecules were represented by the MODESLAB software, the atomic or binding parameters that were used to calculate the molecular descriptors were selected.
The selection criteria of the weighting parameters for the calculation of the molecular descriptors consisted in the implications that they have for the explanation of the behavior of different physical, chemical, chemical-physical or biological properties of the organic compounds of pharmaceutical interest.
The database has been made with the greatest possible heterogeneity from the structural point of view, which on the one hand guarantees a broad representation of the different families of organic

compounds (a high percentage of the compounds used in the biopharmaceutical industry are polyfunctional) and on the other, it will increase the domain of application of the mathematical models obtained.

In order to obtain the mathematical models, the statistical processing of the data was carried out using the SPSS software. Following the stepwise linear discriminant analysis technique, the referred database of 536 compounds and 91 independent variables (molecular descriptors) and a dependent nominal variable were analyzed) of Wilks.λThe selection of the models was made based on the statistical quality of the same, the multivariate comparison statisticians taken into account for this purpose were, first of all, the Lambda

.

.

.

Conclusions A wide series of training was conformed by 536 organic compounds of pharmaceutical interest, 91 molecular descriptors were calculated to the compounds of the training series and discriminant functions are obtained that allow to classify the polyfunctional organic compounds according to their chemical nature

References *()*

.

1. E, E., *Aplications of Aproximations in adyacence Matrix of edge.* J. Chem. Inf. Comput,2012.
2. Lajiness, M.S., *Molecular similarity-Based Methods for Selecting Compounds for Screenig. In Computacional Chemical Graph Theory.* 2010, new york.
3. Adler, M., *A detailed discussion of the crystal structure of compound 31 bound to for Xais described elsewhere.* M.Biochemistry, 2012.
4. Estrada, E., *Aplication of aproximations Toss Mode.* J. Chem. Inf. Comput, 1995. **35**.
5. H, Y., *QSAR studies of HIV-1 integrase inhibition.* Bioorg Med Chem., 2012. **12**.
6. Helmut Mack, *Orally active thrombin inhibitors .Par t1: Optimization of the P1-moiety.* Bioorg Med Chem., 2006.
7. MG, F., *QSAR studies of the pyrethroid insecticides. Part 3. A putative pharmacophore derived using methodology based on molecular dynamics and hierarchical cluster analysis.* J Mol Graph Model., 2003.
8. Rodríguez, L., *Topological Substructure Molecular Design.* 1997: Cuba.
9. T, S., *Classification of environmental estrogens by physicochemical properties using principal component analysis and hierarchical cluster analysis.* J Chem Inf Comput, 2013.
10. T, N., *Structural classification of protein kinases using 3D molecular interaction field analysis of their ligand binding sites:* J Med Chem, 2013.

Deng, H., *Synthesis, SAR exploration ,and X-ray crystal structures of factor XIa inhibitors*

.