Ediciones pensando el futuro
Editions thinking the future
Informática 2013
XV CONVENCIÓN Y FERIA INTERNATIONAL
15TH INTERNATIONAL CONVENTION AND FAIR
www.informaticahabana.com
www.informaticahabana.cu
ISBN: 978-959-7213-02-4

SLD165 EMPLOY OF HYBRID REDUCED GRAPH FOR CLASSIFICATION STUDIES OF BIOACTIVE MOLECULES

SLD165 EMPLEO DE GRAFOS REDUCIDOS HÍBRIDOS PARA ESTUDIOS DE CLASIFICACIÓN DE MOLÉCULAS BIOACTIVAS

Ramón Carrasco-Velar[1], Julio Omar Prieto-Entenza [1], Aurelio Antelo-Collado[1], Juan Alexander Padrón-García[1,] Gonzalo Cerruela-García[2], Álvaro Luis Maceo-Pixa[1], Rubén Alcolea-Núñez[1] and Luis Guillermo Silva-Rojas[1]

1 *Universidad de las Ciencias Informáticas*, *Cuba*, *rcarrasco@uci.cu*,*Carretera a San Antonio Km 2 ½.Torrens.Boyeros. La Habana*

2 *University of Córdoba*, *Spain*, *gcerruela@uco.es*

ABSTRACT:*There were defined and evaluated new atomic and local hybrid indices in retrospective study. Inspired in the Refractotopological State Index for Atoms, the new indices are theoretically supported by graph theory principles. The local indices are obtained from the sum of the atomic values of the atoms in the selected group. The meta classifier Bagging-Reptreewas used for SAR studies, with more than 92% accurate prediction. These indices show a low mutual correlation coefficient.*

KeyWords:*Refractotopological State Index, Hybrid indices, meta-classifiers, LocalDescriptors, Reduced Graph,SAR*.

RESUMEN: Se definieron y evaluaron nuevos índices híbridos atómicos y locales en un estudio retrospectivo. Inspirado en el Índice del Estado Refractotopológico para Átomos, los nuevos indices están basados en la teoría del grafo químico. Los índices locales se obtienen a partir de la suma de los valores atómicos de los átomos del fragmento seleccionado. Los índices están poco correlacionados entre si. Se empleóBagging-REPTreecomo meta clasificador para los estudios de clasificacióncon una exactitud en la predicción del 92%.

Palabras Clave: Índice del Estado Refractotopológico, Índices híbridos, meta-clasificadores, Descriptores Locales, Grafo Reducido,SAR.

1. INTRODUCTION

In 2000, Todeschini and Consoni stated that the molecular descriptor is the result of a logical and mathematical procedure, which transforms chemical information encoded within a symbolic representation of a molecule into a useful number, or the result of some standardized experiment [1].

It is also known that molecular descriptors are closely connected to the concept of molecular structure, and as a consequence play a fundamental role in scientific research, being the theoretical core of a complex network of knowledge [2]. Molecular descriptors are based on several different theories, but those based in graph chemical theory may be the most popular ones, although to be fair, the molecular graphs are an invaluable interface for computational chemists and biologists. There are also several descriptors based on the complexity of molecular graphs. Bonchevhas published comprehensive overviews of these descriptors. [3, 4]

Since 2004 the first hybrid atomic index -The Refractotopological State Index for Atoms (ℜ, Rstate) [5] was published by two of the present authors, nothing else has appeared which focuses on this controversial approach. Nevertheless, others have published their own applications of it, generally included in the Electrotopological(*S*, Estate) ones [6-8]. Although reports of application of the index in SAR and QSAR studies, including in the design of new compounds, may be considered as sufficient to demonstrate the usefulness of this approach, it is necessary to take into account that it has been reported that in the definition of any topological index it is not possible to include chemical-physical properties[9]. Nevertheless, not always one or two atoms are responsible for the biological activity, as in the case, for example, of the estrogenic property of flavonoids as compared with 17-beta-estradiol [10]. In this case, there are several reports of the employ of molecular fragments in SAR and QSAR.

They were used in first additive schemes developed in the 40s-50s to estimate physicochemical properties of organic compounds, although Hammett, Hansch and others developed the well-known extra thermodynamic approach for substituent constants [11, 12]. These molecular substructures not only may be regarded as connected graphs that are completely contained in the molecular graph, but also as portraying one or more of the three principal chemical-physical features: electronic, steric and solubility.

Mathematically speaking, many physicochemical properties can be related to the frequency of certain substructures in a molecule, as in the Free–Wilson approach [13].These molecular substructures or subgraphs are used for the screening process in most of the chemical databases. In the search of pharmacophoric groups, the 2D structure of the subgraph is the most commonly used approach. Nevertheless,research programs in the pharmaceutical industries seek to identify novel, synthetically feasible molecules with useful biological activity and minimal side effects. One such approach is the study of bioisosterism [14-16], but in order to do so it is necessary to include work with an adequate description of the molecular fragment or subgraph. On the other hand, there are many classifiers for SAR studies. Several of them are meta-classifiers, which are broadly applied when others fail. In Weka software, many of these are included, as Decorate [17-18] and Bagging[19]. From these two examples, the last one was selected to be employed in this work.

2. THEORETICALFEATURES

At the beginning of the QSAR era [12] the chemical-physical approach was addressed to the description of chemical structure with three essential properties related to electrostatic, bulky, and lipophilic characteristics of molecules. With the advent and growing of graph chemical theory, the employ of structural variables based on this approach grew dramatically. The great number of molecular descriptors bears to the appearance of prediction models with regression equations with five and more variables. The complexity of these models complicates the possibility of carrying out pathological and physiological analysis since the relationships between the biological activity and the chemical structure is not evident because it is related to a great number of independent variables that are difficult to analyze. With the approach here exposed, authors attempt to return to the early days when the physico-chemical approaches were the principal way to describe the performance of molecules in the biological media, without excluding the use of chemical graph theory. To do that, the first step is to define the different atomic descriptors weighted with electronic, bulky and lipophilic properties and later define the corresponding reduced graphs.

2.1. Definition of AtomicDescriptors

The atomic descriptors defined in this section, are the logical continuation of the Refractotopological State Index for Atoms [5, 20]. To extend the concept of essential structural features and chemical-physical properties, it was decided explicitly include tridimensional information and lipophilicity, as a new partitioned property for characterization purposes but also with explicit tridimensional information contents. The 3D structural information content is reached from the optimized structure with any quantum-chemical approach.

2.2. Definition of Lipotopologic State Index for Atoms.

The Lipotopologic State Index for Atoms (Λi, Lstate) like its analogous ℜi. is defined by equation I:

$$\Lambda i,j = ALi + \Delta ALi.j \qquad (I)$$

whereAL_i is the lipophilic intrinsic value of atom I and ΛALi,jis the perturbative term defined by equation II:

$$\Delta\Lambda i,j = \sum_{j=1}^{N} (ALi - ALj)/ri,j2 \qquad (II)$$

where the sum of all the j vertices adjacent in the

chemical graph, ALi and ALj are the intrinsic lipophilicity values of the atoms *I* and *j*, respectively, and $r_{i,j}^2$ is the atom number of the shorter path between atoms *i* and *j*, including both *i* and *j*. To the lipophilicity values of each one of the atoms different from hydrogen, the corresponding values of the hydrogen bonded are added.

2.3. Definition of the topographic indices for atoms.

The transformation of the topologic indices into the topographic ones is a very simple procedure. The substitution of the topologic distance in the equation II by the corresponding euclidean is sufficient. This last distance was obtained from the optimized geometry of ground state of molecules. Therefore, the Refractotopographic ($\Re_{3D}$, Rstate3D) and the Lipotopographic State Index for Atoms (Λ_{3D}, Lstate3D) that are proposed, are developed from the graph chemical theory and the partition of the atomic refractivity or the atomic lipophilicity as defined by Ghose and Crippen [21-23].

2.4. Definition of Local Descriptors

A common practice to describe the chemical structure in chemometry is splitting the molecule in fragments or subgraphs. The graph theoretical representation of molecules allows the manipulation of the chemical structure very easily. This approach also allows an approximation to the concept of pharmacophoric group as the part or region of molecule responsible of the biological activity. For these fragmentations, several topological graph-theoretical approaches have been employed [24, 25].

Typical subgraphs, like Structure-Connected Clusters, rings and selected vertices, can be used as molecular fragments and also to define the local descriptors. In thiswork, thefollowing subgraphs wereselected:

- Structure-Connected Clustersorderthree
- Structure-Connected Clustersorderfour
- Heteroatom
- Ringsorderthreeto ten

The simplicity and advantage of this approach is that, for example, Structure-Connected Clusters order three includes functional groups as carboxyl, amide, ester, amidine, keto, thiocarboxyl, thioester, N,N-disubstituted amine, tertbutil, and the substitutions in rings among other organic functions. The Structure-Connected Clusters order four take into account the double substitutions in rings, the spiro compounds, etc. The heteroatoms include the phenol, amine, ether, thiol, halogen, etc. functions. In the same way, the difference between features such as aromaticity, conjugation, saturation, size and the presence or not of endocyclic heteroatoms, among others, is achieved from the values acquired by cycles in the molecular environment. It may be take into account that all these descriptors are closely related to the environment around the atom or group of atoms considered.

Another feature evaluated by the new indices was the capability to offer different structural information related to the properties used for weighting. The local indices here presented will be called Descriptors Centers (CD) and are defined by equation III.

$$CD = \sum_{i}^{k} \phi Ai \qquad \text{III}$$

where CD represent the overall value of the property in the molecule or fragment given by the sum of all the corresponding values of atom indices in the structure and ϕ is the property of atom $A_i$represented in this case by the atomic indices, both topologic and hybrid, Estate, Rstate, Lstate and the corresponding topographic Estate3D, Rstate3D and Lstate3D. For Estate and Estate3D, the sum includes all atoms different from hydrogen included in the subgraph. In the case of hybrids, the hydrogen atoms are included. For cycles, this definition takes in to account all endocyclic atoms of cycles order *k,* where *k*, belongs to the interval $3 \leq k \leq 10$.

3. METHODSAND PROGRAMS

A statistical analysis to the CDs of each sample as a whole and later to each descriptor was made. To evaluate the different information contents, the mutual correlation matrices for the different variables, were calculated. Also applied was a multiple regression analysis among the Electrotopologic and Electrotopographic indices against the Refracto and Lipohomologous.The box and whisker's diagrams for the mean and the standard deviation were determined.

To evaluate the capability of the CDs to classify molecules, it was selected the confirmatory assays in the NCBI databank, AID:941. As meta-classifier was used Bagging-REPTree.A ten-fold cross validation was applied to test the models. Also were evaluated libSVM, and Multilayer Perceptron. To carry out these studies, the programs Statistics v. 7.0 [26] and Weka v. 3.7.0 [27] were used.

4. RESULTS AND DISCUSSION

4.1. Statistical analysis

The bioassay AID:941 was the choice to exemplify the statistical performance of the new indices. Table 1 shows the main calculated statistics, and Table 2, the correlation matrix between the topologic and the topographic ones. Another analysis for each fragment type is shown in Figure 1 for cycles order six where the mean and the standard deviation are showed in a box and whiskers graph.

We employ regression analysis to evaluate the relation between information content among variables. The Electrotopological State Index for Atomsand the corresponding topographic were the dependent variable in the regression analysis to identify the different information contents of its analogues, both for 2D and 3Dderivatives. These results are shown in Table 3.

TableI.DescriptiveStatistics. Full sample (n=2775)

	Mean	Min.	Max.	Std.Dev.
S	6.516	-18.041	20.958	6.155
$\mathcal{H}$	11.689	-19.057	50.603	12.793
Λ	0.456	-10.395	6.376	2.564
S_{3D}	6.735	-7.321	20.977	4.984
$\mathcal{H}_{3D}$	12.046	-9.081	40.797	11.575
Λ_{3D}	0.447	-7.387	4.309	1.930

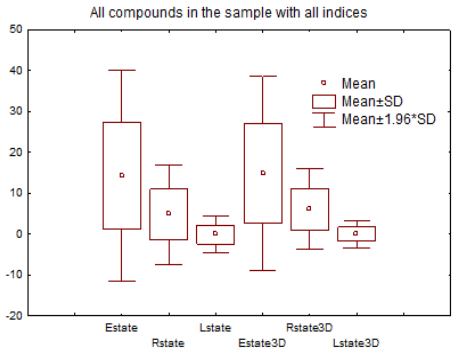


Figure 1.Comparison of the means of all indices

TableII. Correlation matrix between all indices in full sample (AID941)

	S	$\mathcal{H}$	Λ	S_{3D}	$\mathcal{H}_{3D}$	Λ_{3D}
S	1.00	0.32	0.37	0.98	0.45	0.41
$\mathcal{H}$		1.00	0.32	0.31	0.95	0.34
Λ			1.00	0.31	0.28	0.98
S_{3D}				1.00	0.47	0.36
$\mathcal{H}_{3D}$					1.00	0.31
Λ_{3D}						1.00

TableIII. Multiple regression. Cycles six member and Structure-Connected Clusters order 3.

Dependent Variable: ***S***, R= 0.57 R^2= 0.32 Adjusted R^2= 0.32 F(2,867)=207.02 p<0.00Std. Error of estimate: 4.39;***Cycles six member*** n=870

	B	Std.Err.	t(867)	p-level
Intercpt	6.36	0.51	12.47	0.00000
Λ	1.09	0.07	15.37	0.00000
$\mathcal{H}$	0.09	0.02	4.06	0.00005

Dependent Variable: ***S***, R= 0.64 R^2= 0.41 Adjusted R^2= 0.41 F(2,867)=298.49 p<0.00 Std.Error of estimate: 3.08; ***Cycles six member*** n=870

	B	Std.Err.	t(867)	p-level
Intercpt.	5.57	0.52	10.67	0.00000
Λ_{3D}	1.12	0.07	17.15	0.00000
$\mathcal{H}_{3D}$	0.14	0.02	6.48	0.00000

Dependent Variable: ***S*** R= 0.28 R^2= 0.08 Adjusted R^2= 0.07 F(2,223)=9.63 p<0.0001 Std.Error of estimate: 4.073; ***Cluster 3*** n=226

	B	Std.Err.	t(223)	p-level
Intercpt	5.68	0.55	10.28	0.00000
$\mathcal{H}_{3D}$	0.14	0.04	3.70	0.00026
Λ_{3D}	-0.59	0.20	-2.94	0.00362

Dependent Variable: ***S*** R= 0.33 R^2= 0.11 Adjusted R^2= 0.1 F(2,223)=13.801 p<0.00000 Std.Error of estimate: 3.203; ***Cluster 3*** n=226

	B	Std.Err.	t(223)	p-level
Intercpt	5.79	0.51	11.25	0.00000
$\mathcal{H}_{3D}$	0.16	0.04	4.36	0.00002
Λ_{3D}	-0.76	0.21	-3.67	0.00030

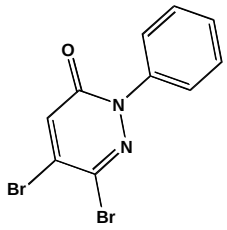
When analyzing the results showed in table 1, the different information content of each index when all fragments are considered together is evident. Besides this, the calculated cycles order six for example, reveals the difference in the space values among index type, although it is not so evident among the topologic and topographic ones. This can be seen in Figure 1, where the mean and the standard deviation are shown in a box and whiskers graphic. Even the correlation matrix (Table II) shows the high correlation between the topologic and the topographic indices.

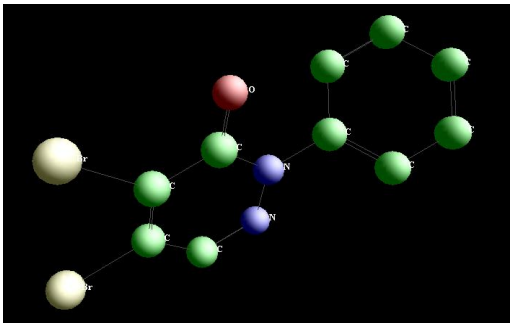
This behavior is understandable because the standard Ghose and Crippen parameters were calculated from the structure fully optimized by semi empirical methodsand the resultsare averaged. This approach include tridimensional information content, but it is increased with the algorithm for the calculation of hybrid indices because it is done for

specific molecules included in thestudied sample, and it is not a simple statistical approximation. Moreover, Randic [28] establish that the information content among indices is the same, only when mutual correlation is 1. To demonstrate it, he developed the Dominant Component Analysis. These results together with the low multiple regression coefficients found in both cases (Table III), strongly suggest that the indices possess different information contents.

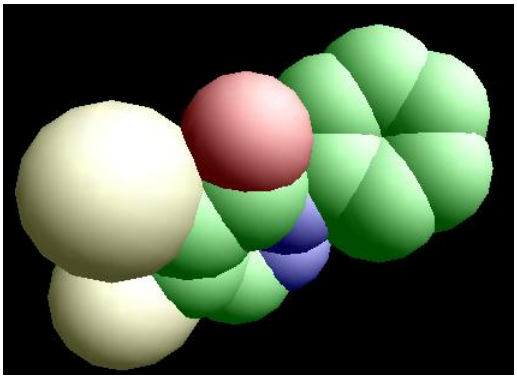
4.2. Graphical representation

The distribution of values of each index suggests that the molecule looks different when the corresponding property is visualized. It is corroborated with the graphical representation of values of compound I (CID: 203396, PubChemBioasssay AID:941) in a simple computational graphic as shown in Figure2[29].

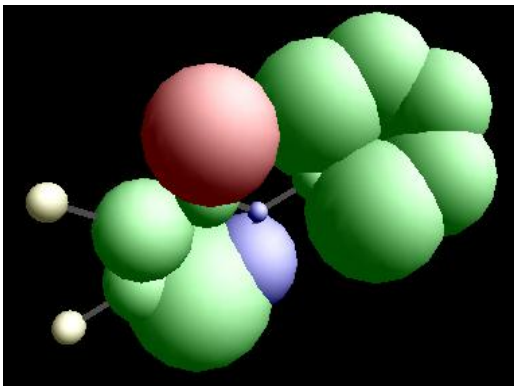




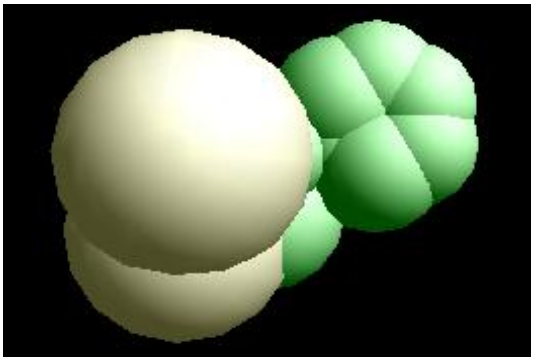
Wire and balls



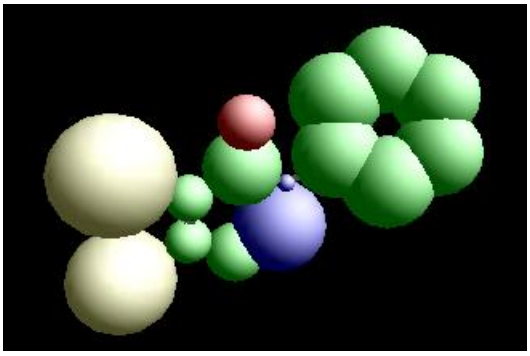
Van der Waals Radii



Electrotopographic



Refractotopographic



Lipotopographic

Figure2.Representation of the hydrogen depleted compound I. Wire and balls, van der Waals and the schematic representation of indices values.

With this representation is possible to distinguish the different performance of each atom in the molecule. For example, the relevance of the oxygen atom is less when the lipophilicity is considered, or in another case, the molar refractivity of bromine atoms is more important than electrostatic or lipophilic property when they are analyzed, and so on. It can also be observed the great difference between the two nitrogen atoms.

4.3. Activity prediction study

In data mining, the problem P in the main learning process Lp is decomposed into sub-problems:

$$P = [P_1, \dots P_2]$$

where $P = \langle Di, Mi \rangle$ and D ⊆ M is a learning dataset and M is a model space. In computational intelligence attractive models m ∈ M are determined with learning process: *Lp:D→ M*, where *p* defines the parameters of the learning machine and is decomposed to the vector *[Lpii, . . . ,Lpnn]*. Meta-learning algorithms are also learning machines; however, the goal of meta-learning is to find the best decomposition in an automated way.

An ensemble is itself a supervised learning algorithm, because it is trained, and then used to make predictions. The trained ensemble, therefore, represents a single hypothesis. This hypothesis, however, is not necessarily contained within the hypothesis space of the models from which it is built. It can be said that ensembles have more flexibility in the functions they can represent. This flexibility can, in theory, enable them to over-fit the training data more than a single model would, but in practice, some ensemble techniques like for example, bagging (Bootstrap aggregating), tend to reduce problems related to over-fitting of the training data. It trains each model in the ensemble using a randomly drawn subset of the training set, and then achieves very high classification accuracy [19]. Fast algorithms such as decision trees are commonly used with ensembles (for example *Random Forest*), although slower algorithms can benefit from ensemble techniques as well.

The results were attained through a simple process that can be summarized in Figure 3,where the first step was the calculation of all indices both news and precursors, followed by the calculation of CDs. The third step consisted in the classification study using typical classifiers as SVM and MLP among others, and meta classifiers. The non-conclusive results required to include the distance between the descriptors centers as a new structural element and the classification study repeated, including the use of metaclassifiers.

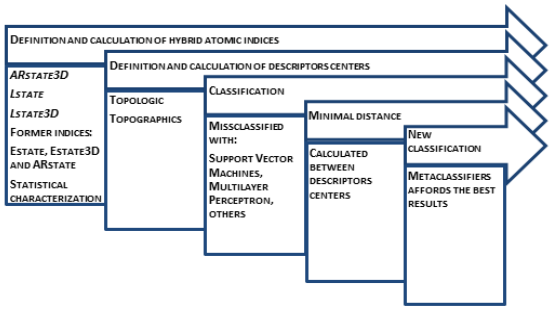


Figure 3.Workflow chart for the classification process.

For the classification study we used only the local topographic indices before defined, taking into account the tridimensional information content against the corresponding topologic. The employ of SVM or neural networks did not afford classification results over 60%. To solve that, meta-classifiers were applied but also without satisfactory results. Therefore, the descriptorscenterswerecountedinpairs without repeatinganypair, the minimal distance between CDs calculated and then, the computational experimentwas restarted. The results for the studied sample are shown in table 4, including both ROC area and the respective confusion matrices. The ROC area analysis [30,31] is useful to select possible optimal models and to discard suboptimal ones independently from the cost context or the class distribution. Given the high values of this parameter and the percentage of correct classification for each case, it can be assessed that the selected classification techniques and descriptors are suitable to describe the samples. The good quality of the obtained models confirms the usefulness to predict the molecular activity.

TableIV.Prediction (in %, Confusion Matrix and average ROC Area).

Bagging-REPTree: 91.7 %			
ConfusionMatrix			AverageROCArea
a	b	classified as	
5691	352	a = Active	0.975
550	4336	b = Inactive	

The high quality of the model can also be assessed taking into account other statistics from the confusion matrix. Good values are obtained both for active and inactive compounds observing that for example, for the employed sample are reached values of 0.935, 0.926 and 0.893 for sensitivity, accuracy and specificity respectively. These results suggest the high quality of the models in the studied sample.

5. CONCLUSION

These results strongly suggest that is possible the developing of good classification models for qualitative structure activity relationships by using hybrid descriptors, and that the reduced graphs have, physical chemical properties weighted, be useful for SAR studies with the use of meta-classifiers More work is in progress to make precisions on structural description with hybrid indices and the select the best classifiers.

6. ACKNOWLEDGES

Authors wish to acknowledge the University of Informatics Sciences for financial support.

6. BIBLIOGRAPHIC REFERENCES

1.R. Todeschini, and V. Consonni, Handbook of Molecular Descriptors, Wiley-VCHVerlag GmbH, Weinheim, Germany, 2009, p. 668.
2.R. Todeschini, and V. Consonni, Molecular Descriptors for Chemoinformatics Second, Revised and Enlarged Edition WILEY-VCHVerlag GmbH & Co. KGaA, Weinheim ISBN 978-3-527-31852-0; 2009.
3.D. Bonchev and D. Rouvray, Complexity in Chemistry: Introduction and Fundamentals. London and NewYork : Taylor & Francis, 2003.
4.D. Bonchev.Overall connectivities/topological complexities: A new powerful tool for QSPR/QSAR. J. Chem. Inf. Comput. Sci., 40(2000) 934–941.
5. R. Carrasco, J. A. Padrón and J. Gálvez.Definition of a novel atomic index for QSAR: the refractotopological state. J. Pharm. Pharmaceut. Sci. 7(2004) 19-26.
6.S. Samanta, Sk. M. Alam, P. Panda and T. Jha. Pharmacophore Mapping of Tricyclic Isoxazoles for Their Affinity Towards Alpha–2 Adrenoreceptors. Internet Electron. J. Mol. Des. 5(2006) 503–514.
7. T. Jha, S. Samanta, S. Basu, A. K. Halder, N. Adhikari, and M. K. Mait. QSAR Study On Some Orally Active Uracil Derivatives as Human Gonadotropin–Releasing–Hormone Receptor Antagonists. Internet Electron. J. Mol. Des.7(2008) 234–250.
8.N. Adhikaria, M. K. Maitia and T. Jha.Predictive comparative QSAR modelling of (phenylpiperazinyl-alkyl) oxindoles as selective 5-HT1Aantagonists by stepwise regression, PCRA, FA-MLR and PLS techniques. Eur. J. Med. Chem. 45(2010) 1119-1127.
9. M.Randic. On Characterization of Chemical Structure. J. Chem. Inf. Comput. Sci., 37(4) 672-687(1997).
10. J.C. Escalona, R. Carrasco, K. Guerra and J. Centeno. Estudio QSAR de flavonoides con actividad estrogénica. Rev. Cub. Quím. XI (1999) 21-31.
11.L.P. Hammett. Linear free energy relationships in rate and equilibrium phenomena. Trans. Faraday Soc. 34(1938) 156–165.
12.T. Fujita and C. Hansch, ρ-σ-π-analysis, a method for the correlation of biological activity and chemical structure. J. Am. Che. Soc. 86(1964) 1616–1626.
13.S. M. Free and J.W. Wilson. A mathematical contribution to structure–activity studies., J. Med. Chem. 7(1964) 395–399.
14.H. L Friedman. Influence of Isosteric Replacements Upon Biological Activity; National Academy of Sciences-USA: Washington, DC, 206(1951) 295-300.
15.C. W. Thornber. Isosterism and Molecular Modification in Drug Design.Quart.Rev. Chem., 8(1979) 563–579.
16.K. Birchall, V. J. Gillet and P. Willett.Use of Reduced Graphs to Encode Bioisosterism for Similarity-Based Virtual Screening. J. Chem. Inf. Model., 49(2009) 1330–1346.
17.P.Mooney and R. J. Melville.Constructing Diverse Classifier Ensembles Using Artificial Training Examples. Eighteenth International Joint Conference on Artificial Intelligence.(2003) 505-510.
18.P.Mooney and R. J. Melville. Creating Diversity in Ensembles Using Artificial Data.InformationFusion.Special Issue on Diversity in Multiclassifier Systems.2004.
19.L. Breiman, BaggingPredictors. Machine Learning, 24(1996) 123-140.
20.R. Carrasco-Velar, Nuevos Descriptores Atómicos y Moleculares. Aplicaciones. [ed.] Ramon Carrasco-Velar. Doctoral Thesis. (2003) La Habana, Cuba : Editorial Universitaria. ISBN 978-959-16-0646-4.
21.A.K. Ghose and G.M. Crippen. J. Comput. Chem., 7(1986), 565-577.
22.A.K. Ghose and G.M. Crippen. J. Chem. Inf. Comput. Sci., 27(1987), 21-35.
23.A.K. Ghose, A. Pritchett. and G.M. Crippen. J. Comput. Chem., 9(1988) 80.
24.A.Varnek and A.TropshaChemoinformatics Approaches to Virtual Screening., Eds. ISBN: 978-0-85404-144-2, Royal Society of Chemistry,.Chapter 1, (2008) pp.2-20.
25.F. Ruggiu, et al. ISIDA Property-Labelled Fragment Descriptors. Molecular Informatics, 29:855–868, 2010, doi:10.1002/minf.201000099.
26.StatSoft, Inc. (2004). STATISTICA (data analysis software system), version 7.www.statsoft.com.
27.M.Hall, E.Frank, G.Holmes, B.Pfahringer, P.ReutemannandI. H. Witten; The WEKA Data Mining Software: An Update; SIGKDD Explorations, 11(2009), Issue 1.
28.M. Randic, Orthogonal Molecular Descriptors. J. Chem. Inf. Comput. Sci., 31(1991) 311-320.
29. A.L. Maceo-Pixa, R. Carrasco-Velar; R. Alcolea-Nuñez, andL.G. Silva-Rojas. *MolVis*(Software for visualization of topographic atomic indices, Prototype). University of Informatic Science, Visualization and Virtual Reality Department, La Habana, Cuba, 2012.
30. T. Fawcett. An introduction to ROC analysis.Pattern Recognition Letters - Special issue.27 (8) 861-874 (2006).
31. Wen Zhu, Nancy Zeng and Ning Wang.

Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS Implementations.NESUG 2010, Health Care and Life Sciences. http://www.cpdm.ufpr.br/documentos/ROC.pdf, Last accessed 9/17/12.

7. CURRICULUM SUMMARY OF THE FIRST AUTHOR

Ramón Carrasco-Velar. Lic. in Chemistry at Havana University (1975). Master in Organic Chemistry (1993) and Dr. in Chemistry (2003) at the same university. Research interest areas: Cheminformatics, Computational Drug Design and Graph Mining. Senior Molecular Modeling and Drug Design Laboratory of the Center of Pharmaceutical Chemistry, from 1991 to 2001.Secretary of the QSAR Section of the Cuban Chemical Society from 1994 to 2003.