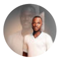
The Summation of Atomic Contributions is an Overly Simplified Characterization of the Holistic Molecular Behavior

Article *in* Letters in Drug Design & Discovery · August 2016

DOI: 10.2174/1570180813999160824160828



Yoan Martínez López
Universidad de Camagüey "Ignacio Agramonte Loynaz"

Yovani Marrero-Ponce
Universidad San Francisco de Quito (USFQ)

Stephen Jones Barigye
McGill University

Letters in Drug Design & Discovery, 2016, *13*, 00-00

The Summation of Atomic Contributions is an Overly Simplified Characterization of the Holistic Molecular Behavior

Yoan Martínez-López[a,b], Yovani Marrero-Ponce[a,c,d*], Gustavo Echeverri Jaramillo[d] and
Stephen J. Barigye[a,e]

[a]*Computer-Aided Molecular "Biosilico" Discovery and Bioinformatic Research International Network (CAMD-BIR International Network), Los Laureles L76MD, Nuevo Bosque, 130015, Cartagena de Indias, Bolivar, Colombia*

[b]*Department of Computer Sciences, Faculty of Informatics, Camaguey University, Camaguey City, 74650, Camaguey, Cuba*

[c]*Universidad San Francisco de Quito (USFQ), Escuela de Medicina, Colegio de Ciencias de la Salud (COCSA), Hospital de los Valles, Ed. Especialidades Médicas, PB, Oficina 109. Av. Interoceánica Km. 12 ½ y Av. Florencia, Casilla Postal 17-1200-841, Quito, Ecuador*

[d]*Grupo de Investigación Microbiología y Ambiente (GIMA). Programa de Bacteriología, Facultad Ciencias de la Salud, Universidad de San Buenaventura, Calle Real de Ternera, 130010, Cartagena (Bolivar), Colombia*

[e]*Departamento de Química, Universidade Federal de Lavras, CP 3037, 37200-000, Lavras, MG, Brazil*

Y. Marrero-Ponce

Abstract: The present report introduces a set of aggregation operators (AOs) to calculate total and local molecular descriptors (MDs) as a generalized approach for the sum of the components of an atomic weight vector. These AOs are classified in four groups, *i.e.* Norms, Means, Statistic Invariants and "Classical Algorithms". In order to evaluate the usefulness of the proposed MDs in correlation studies, QSAR models for the binding affinity to the corticosteroid-binding globulin of Cramer's steroid dataset were built using the multiple lineal regression technique. The statistical parameters of these models demonstrate that the indices obtained with AOs other than the summation yield better performance than those obtained using the summation exclusively. Additionally, a comparison between the statistics for the best model obtained using the AOs strategy and those reported in the literature reveals superior performance for the former, despite its simplicity. Therefore, it can be concluded that the proposed generalization scheme constitutes a useful tool to be taken into account in different chemo-informatics tasks.

Keywords: Vertex degree, Atomic property, Invariant, Molecular descriptor, Aggregation operator, MD-LOVIs Software, Multiple linear regression, QSPR.

1. INTRODUCTION

Molecular descriptors (MDs) have been defined by Todeschini and Consonni as, "the final result of a logical and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or result of some standardized experiment"[1]. There are different classifications of MDs reported in the literature although the most common is the one based on their dimensionality according to which MDs are stratified in 0D-3D, and most recently 4D-6D. Detailed discussions on the particularities of each class are beyond the scope of this report, see references [1-4]. For example, the 0 DMDs codify chemical information from the simplest chemical structure representation, i.e. **the chemical formula** [1] and these include examples such as: the atom number, molecular weight and relative atom-type counts. These MDs are also known as **constitutional descriptors** and are defined as functions of atomic characteristics/properties. A tool called Mind-Prot can calculate drug and protein Markov Mean Properties using atom properties such as Electronegativity (EM), Kang-Jhon Polarizability, Van der Waals area and Atom Contribution to P. Several QSAR models have been published using these descriptors [5, 6]. Generally, these descriptors are considered as linear combinations of atomic or group characters, which may be atom(or group) types or properties such as: atom numbers, bond type counts, cyclomatic numbers, atomic mass, atomic charge, Van der Waals volume, Pauling electronegativity, etc., to yield global molecular definitions.

The use of the linear combination (additive operator) in the global definition of molecular structure tendencies from their atomic contributions is a common practice in mathematical chemistry. In fact many global definitions of molecular structures reported in the literature are applications of the summation operator to contributions for atoms, functional groups or substructures for example: sum-like [7], reciprocal squared sum-like [8] and first Zagreb-like [9] indices.

---

*Address correspondence to this author at the Universidad San Francisco de Quito (USFQ), Escuela de Medicina, Colegio de Ciencias de la Salud (COCSA), Hospital de los Valles, Ed. Especialidades Médicas, PB, Oficina 109. Av. Interoceánica Km. 12 ½ y Av. Florencia, Casilla Postal 17-1200-841, Quito, Ecuador;
E-mails: ymarrero77@yahoo.es or ymponce@gmail.com

1570-1808/16 $58.00+.00

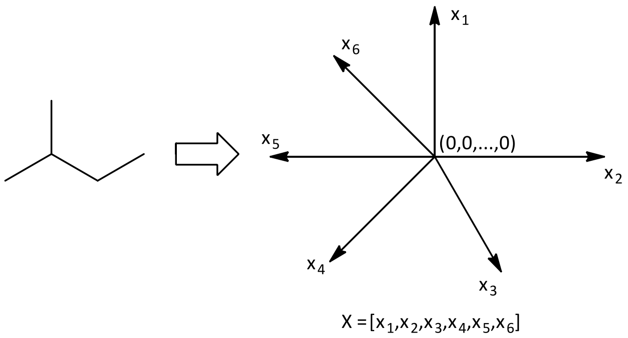


Fig. (1). Representation of the molecular structure of 2-methylpentane molecule as vector of atomic properties.

However, interesting trends have recently been obtained when other operators other than the summation have been applied on vectors of atomic indices, for example better correlations for determined bioactivities were yielded with global MDs based on other aggregation operators other than the summation exclusively [10-14]. The underlying reasoning for this approach is that the summation of atomic contributions is an overly simplified characterization of the holistic experience when the collective contribution of atoms is desired. Similar positions are regular in data structure analysis. For example, from a statistical perspective when one wishes to study the asymmetric tendency of data matrices; the variance offers no useful information to this respect as it is particularly related to the central tendency of data points while skewness and kurtosis are deeply instructive for such cases. Also, when one wishes to compare variables of different properties or ranges, the arithmetic mean may be misleading because of the disparity of the ranges while the geometric mean, "suppresses" this difference rendering them comparable. Therefore it is not implausible to suggest that not all global (molecular) chemical phenomena are ideally represented by the summation operator on atomic contributions. This report aims at extrapolating the generalization schemes proposed in the literature for the linear combination of atom-based indices to simple atomic weight vectors, bearing in mind their overly simplified structure, in order to assess the earnest contribution of these generalization schemes.

2. THEORETICAL FRAMEWORK

Before discussing the theoretical framework of the generalization scheme to be presented in this report, let us give a few definitions. A molecular graph (MG) is defined as a 2D representation of a chemical compound in terms of the constituting set of atoms (indicated by vertices) and their pairwise relations (represented by edges). Often, the vertices in a molecular graph are weighted with determined atomic properties, such as the atomic mass, electronegativity, polarity, Van der Waals radii, etc. to achieve greater discrimination of chemical compounds and these types of graphs are denoted as ***weighted molecular graphs*** [15].

An algebraic operator (AO) is a function which assigns a real number y to an n-dimensional vector of real numbers $(x_1, x_2, x_3, \ldots, x_n)$ [16]:

$$y = AO(x_1, x_2, x_3, \ldots, x_n) \tag{1}$$

The information codified in a weighted MG may be projected in an $\mathbb{R}^n$ space, where $\mathbb{R}$ a set of real numbers and n is the number of atoms in a molecule. If the molecular vector **W** in the $\mathbb{R}^n$ space is considered to start from the origin, then the vector components represent given atom, atom-type or group properties. Consequently, the AOs may be applied to **W** to yield a set of total and local MDs. The AOs, as molecular structure characterizing parameters, have an important property of being invariant, in that they yield the same result irrespective of the arrangement (numbering) of the vector components. In this manuscript the term invariant and AO will be used interchangeably.

Let us take as an example the MG of 2-methylpentane. Firstly, in accordance to the corollaries above, this MG may be represented in the $\mathbb{R}^6$ space by a six dimensional vector, with components $[x_1, x_2, x_3, x_4, x_5, x_6]$ representing particular atomic traits (see Fig. **1**).

Therefore, if Pauling electronegativity (E) for each atom is taken as property and is divided by vertex degree [*i.e.* $W_i = E(S_i)/\delta(S_i)$], the following vector of atomic weights is obtained:

$$W = [E(C_1)/\delta(C1)\ ,\ E(C_2)/\delta(C2),\ E(C_3)/\delta(C3),\ E(C_4)/\delta(C4),\ E(C_5)/\delta(C5),\ E(C_6)/\delta(C6)]$$

$$W = [2.55/1,\ 2.55/1,\ 2.55/3,\ 2.55/2,\ 2.55/2,\ 2.55/1]$$

$$W = [2.55,\ 2.55,\ 0.85,\ 1.275,\ 1.275,\ 2.55]$$

2.1. Atomic Weight of the Molecular Structure from a Molecule

A group of atomic weights (AWs) [17] are used to generate the n-dimensional vector of atomic weights. These AWs are standard values or calculated from atoms of a molecule and are classified in three groups:

The *first group* constitutes **Chemical** AWs such as Atomic Number (Z), Atomic Mass (A), Van der Waals Volume (VW), Covalent Radius (R), Polarizability (P) and Pauling Electronegativity (E). The chemical properties are divided by the bond vertex degree (δ) of each vertex in the MG.

The *second group* is comprised of **Physical** AWs such as Topological Surface Area (T), Ghose-Crippen A LogP (L), Ghose-Crippen Molar Refractivity (M) and Charge(C), respectively.

The *third group* of AWs is comprised of **Vertex Degrees** computed on MGs and includes the following:

The Valence Degree (N), defined by Kier and Hall [18], is mathematically expressed as:

$$N_i = \frac{(Z_i^v - h_i)}{(Z_i - Z_i^v - 1)} \tag{2}$$

where, Z_i is the total number of electrons of the i^{th} atom, (*i.e.* atomic number), Z_i^v is the number of valence electrons

(σ electrons, π electrons and n lone pair electrons,) of the i^{th} atom and h_i is the number of hydrogen atoms bonded to it.

The Intrinsic State (I) is a modification of the Valence Degree (**N**), also proposed by Kier and Hall and is defined as [18]:

$$\mathrm{I_i} = \frac{(2/\mathrm{L_i})^2 . N_i + 1}{\delta_i} \tag{3}$$

where, L_i is the principal quantum number, N_i is the valence vertex degree and δ_i is the simple vertex degree of the i^{th} atom. The L_i is used to account for the increase in the screening effect of the inner electrons, δ_i is the sum of elements of the adjacency matrix of the MG [19].

The Electrotopological State (ES) or also called the E-state for atom, a_i is expressed as [20]:

$$\mathrm{ES_i} = \mathrm{I_i} + \sum_{j=1}^{A} \Delta \mathrm{I_{ij}} \tag{4}$$

where, I_i is the intrinsic state of atom a_i and ΔI_{ij} is the perturbation of I_i, which is in turn computed using the formula:

$$\Delta \mathrm{I_{ij}} = \frac{\mathrm{I_i} - \mathrm{I_j}}{\mathrm{d_{ij}^2}} \tag{5}$$

where d_{ij} the topological distance between atoms a_i and a_j of the MG [19].

The Kupchik's Vertex Degree (KU), proposed to take into account the covalent radius of atoms in the MG, is expressed by equation 6 as [21]:

$$\mathrm{KU_i} = \frac{\mathrm{R_c}}{\mathrm{R_i}} . \mathrm{N_i} \tag{6}$$

where, R_c and R_i are the covalent radius of the carbon atom and the i^{th} atom of the molecule, respectively, and N_i is the valence vertex degree.

The Bond's Vertex Degree (BD) accounts for connectedness as well as the bond multiplicity. It is calculated from the atom connectivity matrix **C** as sum of row entries,

$$\mathrm{BD_i} = \sum_{j=1}^{A} \mathrm{C_{ij}} \tag{7}$$

where, A is the number of graph vertices (atoms), C_{ij} are elements of **C**. The elements of this matrix are equal to one for simple bonds, two for double bonds, three for triple bonds (here, conjugated bonds are considered as double bonds) and zero if there exists no bond.

The Hu-Xu's Vertex Degree (HX), proposed by Hu-Xu *et al.* [22], is expressed as:

$$\mathrm{HX_i} = \delta_\mathrm{i} . \sqrt{\mathrm{Z_i}} \tag{8}$$

where, δ_i is the sum of the elements of adjacency matrix of MG and Z_i the atomic number of the considered atom, a_i.

The Li's Vertex Degree (LI), proposed by Li *et al.* [23], is defined as

$$\mathrm{LI_i} = \frac{\mathrm{Z_i^v} . (\mathrm{Z_i^v} - \mathrm{h_i})}{\mathrm{L_i^2}} \tag{9}$$

where, Z_i^v is the number of valence electrons (σ electrons, π electrons and n lone pair electrons) of the i^{th} atom, h^i is the number of hydrogen atoms bonded to it and L_i is the principal quantum number.

The Alikhanidi's Vertex Degree (Alk) is proposed as a modification of the Hu–Xu vertex degree and is expressed as [24]:

$$\mathrm{Alk_i} = \partial_\mathrm{i} \sqrt{\mathrm{Z'_i}} \tag{10}$$

where, Z'_I is a function of the atomic numbers Z_j of the atoms adjacent to the i^{th} atom (also known as consecutive AT number) and is defined as:

$$\mathrm{Z'_i} = \left[\sum_{j=1}^{A} \mathrm{a_{ij}} \sqrt{(\sqrt{2} + \mathrm{Z_j})}\right]^2 \tag{11}$$

where a_{ij} denotes the elements of the adjacency matrix.

The Ivanciuc's Vertex Degree (IN), proposed as a combination of topological distances and vertex degrees in the MG, is computed according to the following general expression [25]:

$$\mathrm{IN_i} = \sum_{j=1}^{A} \mathrm{d_{ij}}^{\alpha} \partial \mathrm{i}^{\beta} \partial \mathrm{j}^{\gamma} \tag{12}$$

where, d_{ij} is the topological distance between vertices v_i and v_j; α = 0, 1; β = 0, 1, -1; γ = 0, 1, -1; ∂_i and d ∂_j are the vertex degrees for v_i and v_j, respectively.

The Distance Count (DC) indicates the frequencies {$^1\mathrm{f_i}$, $^2\mathrm{f_i}$, $^3\mathrm{f_i}$, ..., $^D\mathrm{f_i}$} of distances equal to {1, 2, 3, ...D}, respectively, from vertex v_i to any other vertex of a MG; D is the eccentricity (maximum distance from v_i).

The Eccentric Connectivity (Y) is defined as the sum of products between eccentricity D_i and valence vertex degree δ_i of a MG and is expressed as:

$$\mathrm{Y_i} = \sum_{i=1}^{A} D_i \, \delta_i \tag{13}$$

Note that the atom weights may be computed from an H-suppressed or H-filled MG. Therefore, with the set of AWs computed for a given molecular structure, the corresponding atomic weights vector **W** is constructed and possesses the following structure:

$$\mathrm{W} = [\mathrm{W_1}, \mathrm{W_2}, \mathrm{W_3}, \ldots, \mathrm{W_n}] \tag{14}$$

where, $\mathrm{W_i}$ is the weight for atom a_i and n is the number of atoms of MG.

2.2. Invariants Applied to Atomic Weight's Vector

The Aggregation Operators (AO) as mathematical invariants applied to the atomic weight vector, **W**, constitute a generalization of the classical approach of computing the global (or local) MDs by summation of vector components. These AOs are classified into four main groups. The first group is the **Norms** (or **Metric**s) which include: Minkowski's norms [*i.e.* N1 (equivalent to the summation), N2, N3] and Penrose's size (PN). The second group is the **Mean Invariants** (**first statistical moment**) and comprises: Geometric Mean (G), Arithmetic Mean (M), Quadratic Mean (P2), Power Mean (P3) and Harmonic Mean (A). The third group is the **Statistical Invariants** (**highest statistical moments**) which entails: Variance (V), Skewness (S), Kurtosis (K), Standard Deviation (DE), Variation Coefficient (CV), Range (R), Percentile 25 (Q1), Percentile 50 (Q2), Percentile 75 (Q3), Inter-quartile Range (I50), X max (MX) and X min (MN). The fourth group is the **"Classical algorithms" Invariants** which comprises: Autocorrelation (AC), Gravita-

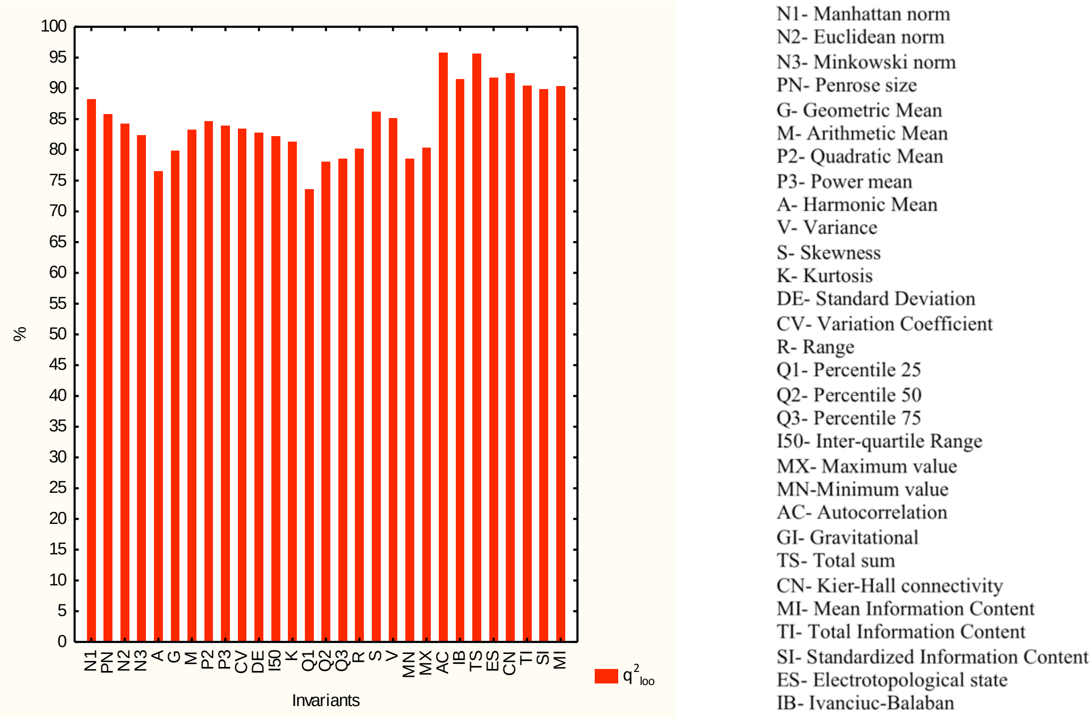


Fig. (2). Performance of NormsMeans and Statistics Invariants in modeling of the physicochemical property log 1/K for Cramer's 31 steroid molecules.

tional (GI), Total Information Content (TI), Mean Information Content (MI), Standardized Information Content (SI), Total Sum (TS), Ivanciuc – Balaban (IB), Electrotopological State (ES) and Kier-Hall Connectivity (CN). Note that most of the AOs in this group can be generalized using the first three groups, given that these algorithms in turn involve the summation operator, see Table **1**.

2.3. Local and Total Molecular Descriptors from Atomic Weights Vector

Motivated by the understanding that the chemical, physicochemical or biological activity of compounds does not always depend on the molecular structure as a whole but also on particular portions which may be functional groups, substructures or fingerprints, the present approach is further refined to permit the computation of total or local definitions for a molecular structure. In this sense, all or a subset of vector components may be considered, to obtain the total or local MDs, respectively. Ten local types (or groups) are analyzed in the present report, *i.e.*, hydrogen bond acceptors (HA), carbon atoms in aliphatic chains (LA), hydrogen bond donors (HD), halogens (HL), terminal methyl groups (MC), Carbon atoms (Cb), carbon atoms in aromatic systems (RA), heteroatoms (O, N and S in all valence states, denoted as HT), unsaturated bonds (IS) and group at lag *k* (GL).

3. QSPR MODELING OF THE CRAMER'S STEROIDS WITH LTMD-W

Cramer's Steroid dataset consists of 31 steroids analyzed for the binding affinity to corticosteroid binding globulin (CBG) [26]. This dataset has been used in several studies to evaluate the performance of novel procedures or methodologies, proposed by several authors, and to compare them with the existing ones in modeling tasks [26-31]. It can thus be considered as a "benchmark" for comparing molecular structure characterizing strategies.

In this study, the Cramer's Steroid dataset was used to validate the local and total MDs obtained from the **W** and to compare N1 invariant (sum of the **W** components) with the other AOs, taking into account the prediction of the binding affinity to CBG. All invariants were calculated with using an "*in house*" software denominated **MD-LOVIs** (Acronym of **M**olecular **D**escriptors from **Lo**cal **V**ertex **I**nvariants and Related Map**s**) [32, 33].

The models were built with MobyDigs software [34] using Multiple Linear Regression coupled with the GA (MLR-GA) as the stochastic optimization tool and the following configurations: the statistical parameter q^2_{loo} as the optimization function, initial population size of 100 individuals and the reproduction/mutation ratio of 0.5. The validation techniques bootstrapping (q^2_{boot}) and y-scrambling [a (q^2)] were used to evaluate the models' robustness and check the possibility of fortuitous correlations of the obtained QSPR models, respectively.

3.1. Internal Comparison of the Local and Total Molecular Descriptors

Firstly, a comparison of all the invariants on the basis of their modeling capacity of the binding affinity to CBG of the Cramer's steroids was performed. For this study, all the 31 steroids are considered as training set (the 31st steroid previously reported as an outlier in the literature was not excluded [35-37]). Fig. (**2**) shows the performance of the six-variable

Table 1. Norms, Means and Statistical Invariants as Generalizations of the Linear Combination of LOVIs as Global (or Local) MDs Operator, as well as Classical algorithms which generalize the first three groups.

No.	Group[a]	Name	ID	Formula[b]
1	Norms (Metrics)	Minkowsky norm (p = 1) Manhattan norm	N1	$\mathrm{N1}=\sum_{a=1}^{n}\lvert L_a\rvert$
2		Minkowsky norm (p = 2) Euclidean norm	N2	$\mathrm{N2}=\sqrt{\sum_{a=1}^{n}\lvert L_a\rvert^2}$
3		Minkowsky norm (p = 3)	N3	$\mathrm{N3}=\sqrt[3]{\sum_{a=1}^{n}\lvert L_a\rvert^3}$
4		Penrosesize	PN	$\mathrm{PN}=\sqrt{\frac{1}{n^2}\left[\sum_{a=1}^{n}(L_a)\right]^2}$
5	Mean (firststatisticalmoment)	Geometric Mean	GM	$\mathrm{G}=\sqrt[n]{\prod_{a=1}^{n}L_a}$
6		Arithmetic Mean (Power mean of degree β = 1)	M	$\mathrm{M}_{\beta}=\left(\frac{L_1^{\beta}+L_2^{\beta}+\ldots+L_n^{\beta}}{n}\right)^{\frac{1}{\beta}}$
7		Quadratic Mean (Power mean of degree β = 2)	P2	
8		Power mean of degree β = 3	P3	
9		Harmonic Mean (Power mean of degree β = -1)	A	
10	Statistical (higheststatisti-calmoments):	Variance	V	$\mathrm{V}=\frac{\sum_{a=1}^{n}(L_a-M)^2}{n-1}$
11		Skewness	S	$S=\frac{n*(X_3)}{(n-1)(n-2)(DE)^3}$ $X_3=\sum_{a=1}^{n}(L_a-M)^3$ M, arithmetic mean DE, standard deviation
12		Kurtosis	K	$k=\frac{n(n+1)X_4-3(X_2)(X_2)(n-1)}{(n-1)(n-2)(n-3)(DE)^4}$ $X_j=\sum_{a=1}^{n}(L_a-M)^j$ M, arithmetic mean DE, standard deviation
13		Standard Deviation	DE	$\mathrm{DE}=\sqrt{\frac{\left(\sum L_a-M\right)^2}{n-1}}$
14		VariationCoefficient	CV	$\mathrm{CV}=DE/M$

Table **1**. contd…

No.	**Group**[a]	**Name**	**ID**	**Formula**[b]
15		Range	R	$R = L_{\max} - L_{\min}$
16		Percentile 25	Q1	$Q1 = \left[\frac{N}{4} + \frac{1}{2}\right]$ N, La number
17		Percentile 50	Q2	$Q2 = \left[\frac{N}{2} + \frac{1}{2}\right]$ N, La number
18		Percentile 75	Q3	$Q3 = \left[\frac{3N}{4} + \frac{1}{2}\right]$ N, La number
19		Inter-quartileRange	I50	$I50 = Q3 - Q1$
20		Maximumvalue	MX	MX = La max
21		Minimumvalue	MN	MN = La min
22	Classical	Autocorrelation	ACk	$AC_k = \sum_{i=1}^{n}\sum_{j\geq 1}^{n} L_i \times L_j \bullet (\delta(d_{ij},k))$ $k = 1,2,..7$
23		Gravitational	GIk	$GI_k = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{L_iL_j}{{}^{k}d_{ij}} \bullet \delta(d_{ij},k))$ $k = 1,2,..7$
24		Total sum at lag k	TSk	$TS_k = \sum_{i=1}^{n}\sum_{j=1}^{n} L_{ij} \bullet \delta(d_{ij},k))$ $k = 1,2,\ldots,7$
25		Kier-Hall connectivity	CNm	${}^{m}KH_t = \sum_{i=1}^{K}(\prod_{i=1}^{n_k} L_i, w)_k^{\lambda}$ where, K is the number of sub-graphs, nk is the number of atoms in a fragment, λ is equal to ½, m and t are the sub-graph order and type, respectively
26		Mean Information Content	MI	$MI = -\sum_{i=1}^{A}\frac{N_g}{N_o} \cdot \log_2 \frac{N_g}{N_o}$ where, Ng is the number of atoms with the same LOVI value. No is the number of atoms in a molecule
27		Total Information Content	TI	$TI = N_0 \cdot \log_2 N_0 - \sum_{g=1}^{G} N_g \cdot \log_2 N_g$
28		StandardizedInformation Content	SI	$SI = \frac{IT}{N_0 \cdot \log_2 N_0}$
29		Electrotopological state (E-state index)	ES	$S_i = I_i + \Delta I_i = I_i + \sum_{j=1}^{n}\frac{I_i - I_j}{(d_{ij}+1)^2}$ where, Ii is the intrinsic state of the ith atom and ΔIi is the field effect on the ith atom calculated as perturbation of the Ii of ith atom by all other atoms in the molecule, dij is the topological distance between the ith and the jth atoms, and n is the number of atoms. The exponent k is 2.

Table **1**. contd…

No.	Group[a]	Name	ID	Formula[b]
30		Ivanciuc-Balaban Type-Indices	IB	$$J_k = \frac{n^2 \cdot B}{n+C+1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_{ij}\left[L_i \times L_j\right]^{-\frac{1}{2}}$$ where, the summation goes over all pairs of atoms but only pairs of adjacent atoms are accounted for by means of the elements aij of the adjacency matrix. The n, B, and C are the number of atoms, bonds, and rings (cyclomatic number), respectively.

[a] The second group (invariants 5-9) could be re-named as "location statistics" if percentiles and maximum (minimum) are taken into consideration in this group. In this case, the third group (invariants 10-21) could be re-named as "spread and shape statistics", [b] LOVIs for "a" atoms in molecule.

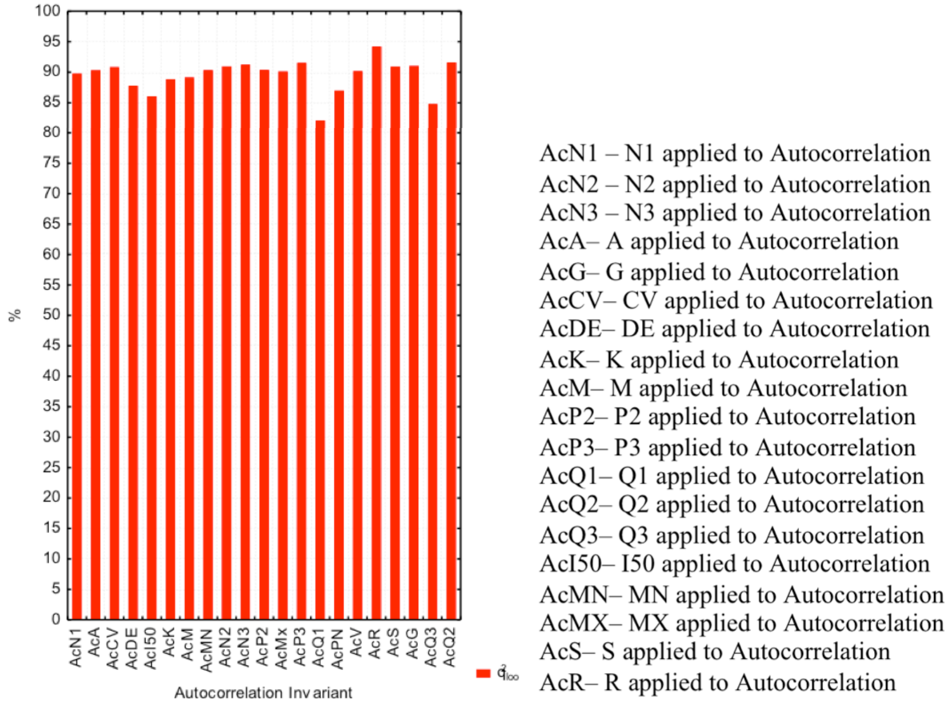


Fig. (3). Performance of norms, mean and statistical invariants applied to the Autocorrelations algorithmin modeling of the physicochemical property log 1/K for Cramer's 31 steroid molecules.

QSAR models, in terms of the cross-validation parameter q^2_{loo}, built with the different AOs.

As can be observed, many mathematical invariants exhibited better performance than N1, for example, Autocorrelation (AC), Ivanciuc-Balaban Type-Indices (IB), Total Sum at lag *k* (TS), Electro-topological State (ES), Kier-Hall Connectivity (CN), Total Information Content (TI), Standardized Information Content (SI), Mean Information Content (MI) and Gravitational (GI). Moreover, the following observations were made for each group of AOs: for the norms, the sum of parts (N1) had the best performance while the Quadratic Mean (P2) had the best performance in the means. As for the statistical invariants and "classical algorithms", the best operators were Skewness (S) and Gravitational (GI) AOs, respectively. Generally, the "classical algorithms" invariants had better performance than the norms, means and statistical AOs. The best overall AO was the Gravitational (GI), with q^2_{loo} = 0.963. For more information see Support Information (SI3).

Additionally, bearing in mind that the "classical algorithms" are in turn generalizable with the first three groups of invariants (*i.e.* norms, means and statistical invariants), it is of interest to evaluate the performance of these AOs within the "classical algorithms". For this study, the Autocorrelation (AC) "classical algorithm" was selected and a comparison of the statistical parameters for the QSAR models obtained with the summation operator-based MDs (equivalent to the original AC definition [38]) with models for the rest of the norms, means and statistical invariants was performed. The following results were obtained (see Fig. **3**): the Range (R), Geometric Mean (G), Percentile 50(I50) and Minkowski norm (N3) had the best performance, with q^2_{loo}= 0.942, q^2_{loo}=0.910, q^2loo=0.916, q^2_{loo}=0.912, respectively, superior to the summation operator's (ACN1) q^2_{loo}=0.89.

Table 2. Statistical parameters for the best models for 1-6 variables for the physicochemical property log 1/K, considering the 31compounds of steroids.

Molecules with Hydrogen Filled (WH)							
Size	**r^2**	**q^2_{loo}**	**$q2_{boot}$**	**a(q$^{2)}$**	**SDEC**	**Models**	**Eq.**
1	0.766	0.733	0.74	-0.171	0.515	Log 1/K = -5.1735(±0.15673) -0.00013(±0.00001) AC(R)_IN_IS	(17)
2	0.876	0.853	0.851	-0.212	0.374	Log 1/K = -4.61042(±0.19336) -0.02738(±0.00237) R_IN_IS + 0.76142(±0.08656)TS(S)_A_IS	(18)
3	0.932	0.911	0.908	-0.279	0.276	Log 1/K = -4.69857(±0.14677) -0.02185(±0.00214)R_IN_IS + 0.76806(±0.06519)TS(S)_A_IS - 0.0578(±0.01221)ES(R)_DC_HT	(19)
4	0.962	0.952	0.949	-0.288	0.205	Log1/ K = -3.96824(±0.19328) -0.01526(±0.00331)AC(K)_C_T - 0.02133(±0.00162)R_IN_IS + 0.75706(±0.04932)TS(S)_A_IS -0.06245(±0.00928)ES(R)_DC_HT	(20)
5	0.968	0.958	0.952	-0.404	0.189	Log 1/K = -4.09192(±0.1916) -0.01319(±0.00327) AC(K)_C_T -0.02147(±0.00152)R_IN_IS + 0.71758(±0.05021)TS(S)_A_IS -0.06751(±0.00908)ES(R)_DC_HT + 0.04124(±0.01989)ES(V)_KU_HA	(21)
6	0.973	0.964	0.954	-0.514	0.173	Log 1/K = -4.15233(±0.18077) -0.01272(±0.00306) AC (K)_C_ T + 0.04838(±0.02218)GI(A)_DC_ IS -0.02188(±0.00143)R_IN_IS + 0.65969(±0.05381) TS (S)_A_ IS -0.06211(±0.00882)ES (R)_DC_HT + 0.05098(±0.01907)ES (V)_KU_ HA	(22)
Molecules with Hydrogen Suppressed (RH)							
Size	r^2	q^2_{loo}	$q2_{boot}$	a(q^2)	SDEC	Models*	Eq.
1	0.690	0.653	0.653	-0.159	0.593	Log 1/K = -4.55546(±0.25264) +0.24844(±0.03089) TS (N1)_L_IS	(23)
2	0.899	0.881	0.879	-0.236	0.338	Log 1/K = -6.96654(±0.23444)-0.00031(±0.00002) AC(R)_IN_IS + 0.11241(±0.0141) TS (V)_KU_LA	(24)
3	0.930	0.915	0.912	-0.292	0.28	Log 1/K = -6.61638(±0.22111) -0.0003(±0.00002) AC(R)_IN_IS + 17.32924 (±4.79888) GI (P3)_C_IS + 0.11566(±0.01223) TS (V)_KU_T	(25)
4	0.953	0.938	0.933	-0.35	0.229	Log1/ K = -6.93139 (±0.1733)-0.00032 (±0.00001) AC(R)_IN_IS + 0.09997 (±0.02135) TS (G)_L_HD + 0.12913(±0.01174) TS (V)_KU_T -0.00144 (±0.0004) TS (V)_IN_HT	(26)
5	0.972	0.961	0.956	-0.38	0.177	Log 1/K = -4.01577(±0.10933) + 0.23161(±0.02385) K_HX_HT + 0.19661(±0.03017) AC (CV)_IN_IS -0.0001(±0.00001) AC (I50)_IN_LA -0.01277(±0.00107) R_IN_IS + 0.35946(±0.05223) TS (S)_Alk_IS	(27)
6	0.977	0.968	0.958	-0.468	0.161	Log 1/K = -4.846(±0.1302) + 0.23243(±0.02523) K_ HX_ HA + 0.00003 (±0.00001) AC_NE(MN)_IN_IS -0.00009(±0.00002)GI (Q2)_IN_LA -0.12313 (±0.01487) DE_ IN _IS +0.39984(±0.04408)TS (S)_Alk_ IS + 0.31411 (±0.0412) TS (S)_KU_T	(28)

In a general sense, this internal comparison showed that the global and local molecular descriptors from norms, mean, statistic and “classical algorithms” invariants, were better predictors of the Log*1/k (CGB) property* than just the sum of its parts (N1).

An important finding from this evaluation is the fact that AOs permit obtaining MDs with varied performance in regression models, and thus combinations of these should yield greater performance.

3.2. LTMD-W Models *vs.* Other Approaches Reported in the Literature

In this section we wish to compare the performance of the LTMD-W approach as a whole with other approaches reported in the literature in the modeling of the Log 1/k (CGB) of Cramer‘s steroids database. Table **2** shows the best regression models for LTMD-W using Multiple Linear Regression coupled with the GA (MLR-GA) method, for 1-6 variable models.

Generally, the obtained models (Equation 23-28) are robust and, are not influenced by the chance correlation phenomenon, evidenced by the favorable cross validation (q^2loo and q^2boot) and y-randomization [a (q^2)] parameters, respectively.

In the comparison with other approaches reported in the literature (see Table **3**), it is observed that the LTMD-W models yield better performance than the best model reported in the literature, based on the Combined Electrostatic and

Table 3. Comparison of q^2loo values for models obtained with LTMD-W and other QSAR methods for 31 compounds of Cramer's steroids.

Method	PCs/Var.	Statistical Method	q^2_{loo}	Eq./Ref.
LTMD-W(H-suppressed)	6	GA and MLR	0.968	Eq. 28
LTMD-W(H-filled)	6	GA and MLR	0.964	Eq. 22
LTMD-W(H-suppressed)	5	GA and MLR	0.961	Eq. 27
LTMD-W(H-filled)	5	GA and MLR	0.958	Eq. 21
LTMD-W(H-filled)	4	GA and MLR	0.952	Eq. 20
Combined electrostatic and shape similarity matrix (CESSM)	6	Genetic NN	0.941	[39, 40]
LTMD-W(H-suppressed)	4	GA and MLR	0.938	Eq. 26
LTMD-W(H-suppressed)	3	GA and MLR	0.915	Eq. 25
LTMD-W(H-filled)	3	GA and MLR	0.911	Eq. 19
Hodking SM	6	genetic NN	0.903	[40]
Fragment QS-SM	4	PLS	0.886	[41]
MEDV-13	5	GA-RLM	0.882	[42]
MiDSASA – "template"	2"compounds"		0.88	[43]
LTMD-W(H-suppressed)	2	GA and MLR	0.881	Eq. 24
LTMD-W(H-filled)	2	GA and MLR	0.853	Eq. 18
TQSI	5	MLR	0.848	[27]
Tuned-QSAR	6	MLR after PCA	0.842	[28]
CoMMA	3	PLS	0.828	[29]
SOMFA/esp + ALPHA	-	SOR	0.82	[30]
CESSM	6	GA and RLM	0.819	[40]
EEVA	4	PLS	0.81	[30]
Charges and Properties from MEPS-AM1	5	RLM	0.80	[31]
CoSA	3 "Bins"	PLS	0.78	[44]
TQSI	4	MLR	0.775	[41]
EVA	5	PLS	0.77	[30]
CoMSA	1	PLS	0.76	[45]
MQSM	5	MLR after PLS	0.759	[41]
EVA + ALPHA	-	SOR	0.75	[30]
GRIND		PLS	0.75	[46]
SEAL	3	PLS	0.748	[47]
SOMFA/esp	6	PLS	0.74	[30]
CoSCoSA	3		0.74	[48]
Similarity Indices (ESP MC matrix)	1	PLS	0.734	[49]

Table **3**. contd…

Method	PCs/Var.	Statistical Method	q^2_{loo}	Eq./Ref.
LTMD-W(H-filled)	1	GA and MLR	0.733	Eq. 17
CoSASA	3 “atoms”	PLS	0.73	[44]
MQSM	4	MLR after PLS	0.727	[28]
E-State and kappa shape index	4	MLR	0.72	[50]
TARIS	2		0.71	[51]
MQMS	3	MLR and PCA	0.705	[41]
CESSM	5	PLS	0.70	[40]
SAMFA-RF	-	RF	0.69	[52]
SAMFA-PLS	4-5	PLS	0.69	[52]
4D-QSAR	2	PLS	0.69	[53]
CoMMA (ab initio)	6	PLS	0.689	[54]
SOM-4D-QSAR	4	SOM Neural Network	0.68	[53]
LTMD-W(H-suppressed)	1	GA and MLR	0.653	Eq. 23
MEDV-13	4	GA and RLM	0.648	[27]
Wagener’s (AMSP Method)	-	k-NN and FNN	0.630	[55]
SAMFA-SVM	-	SVM	0.60	[52]
ALPHA	2	PLS	0.57	[30]_ENREF_9
SOMFA		MLR	R^2 0.577	[56]

[a]When it is applicable, specifies the number of components (PCs) or variables, [b]steric field, [c]electrostatic field.

Shape Similarity Matrix (CESSM) method, even with a much lower degree of freedom [*i.e.* q^2_{loo} (LTMD-W) = 0.952, q^2_{loo} (LTMD-W)= 0.958, q^2_{loo} (LTMD-W)= 0.964 for 4, 5, and 6 variables, respectively, compared to q^2_{loo}(CESSM) = 0.941 obtained with 6 variables]. The CESSM method employs Neural Networks as the fitting method, which is known to yield better optimized models compared with other traditional regression procedures such as Multiple Linear Regression, Principal Component Regression or Multiple Logistic Regression. Nonetheless, better statistical parameters are obtained with the LTMD-W approach using a much simpler technique, demonstrating the theoretical robust of the proposed strategy, notwithstanding its simplicity.

In addition, the 2 variable LTMD-W models computed on H-filled and H-suppressed MGs, respectively, yielded superior performance (q^2_{loo} = 0.881 and q^2_{loo} = 0.853, respectively) than many other methods, whose models were constructed with a higher degree of freedom, for example TQSI (q^2_{loo} = 0.848), CoMMA (q^2_{loo} =0.828), SOMFA (q^2_{loo} = 0.74) are evidently inferior to the 2 variable LTMD-W model.

The good performance of the LTMD-W strategy in QSAR modeling supports the hypothesis that global definitions of chemical behavior from atomic characteristics may after all not imply their linear combinations and therefore other relations should also be considered. Additionally, the LTMD-W strategy offers a remarkable advantage over most 3D techniques due to its inherent simplicity, in that procedures such as molecular alignment and structural optimization, as well as the limitation of their inapplicability to structurally diverse datasets are avoided.

CONCLUSION

In this article, a set of AOs was introduced as generalization of sum of the components of the atomic weight vector, **W**. These AOs are used as ***invariants*** and are classified in Norms, Means, Statistical invariants and “classical algorithms”. The QSPR studies demonstrate that better predictors of the Log1/k (CGB) property may be obtained with other AOs other than the summation operator (N1 in our scheme). Additionally, it is demonstrated that the LTMD-W approach yields models with superior statistical parameters than other approaches reported in the literature, despite its simplicity. Altogether, it may be suggested that the LTMD-W method codifies relevant chemical information and thus constitutes an important tool that should be taken into consideration in chemo informatics tasks such as: diversity analysis of molecular structural datasets, QSPR/QSAR modeling and the high-throughput screening of chemical compound libraries.

Finally, it is important to note that the proposed generalization scheme may be applied to any family of MDs defined at the atomic level.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

Barigye, S. J. acknowledges financial support from CNPq.

PROGRAM AVAILABILITY

The MD-LOVIs software (portable standalone) and the respective user manual are freely available online at www.tomocomd.com.

SUPPLEMENTARY INFORMATION AVAILABLE

The molecular descriptor values for the steroids data set in excel file (SI1) are available via the Internet at http://

REFERENCES

[1] Todeschini, R.; Consonni, V., *Molecular Descriptors for Chemoinformatics*. 1st ed.; WILEY-VCH: Weinheim, 2009; Vol. 1, p. 667.

[2] Vedani, A.; Dobler, M., 5D-QSAR: The Key for Simulating Induced Fit? *Journal of Medicinal Chemistry* **2002,** *45* (11), 2139-2149.

[3] Vedani, A.; Dobler, M.; Lill, M. A., Combining Protein Modeling and 6D-QSAR. Simulating the Binding of Structurally Diverse Ligands to the Estrogen Receptor†. *Journal of Medicinal Chemistry* **2005,** *48* (11), 3700-3703.

[4] Barigye, S. J.; Marrero-Ponce, Y.; Pérez-Giménez, F.; Bonchev, D., Trends in information theory-based chemical structure codification. *Molecular diversity* **2014**, 1-14.

[5] Fernandez-Lozano, C.; Gestal, M.; Gonzalez-Diaz, H.; Dorado, J.; Pazos, A.; Munteanu, C. R., Markov mean properties for cell death-related protein classification. *Journal of theoretical biology* **2014,** *349*, 12-21.

[6] Fernandez-Lozano, C.; Fernandez-Blanco, E.; Dave, K.; Pedreira, N.; Gestal, M.; Dorado, J.; Munteanu, C. R., Improving enzyme regulatory protein classification by means of SVM-RFE feature selection. *Molecular bioSystems* **2014,***10* (5), 1063-71.

[7] Lučić, B.; Trinajstić, N.; Zhou, B., Comparison between the sum-connectivity index and product-connectivity index for benzenoid hydrocarbons. *Chem.l Phys. Lett.* **2009,***475* (1), 146-148.

[8] Zhou, B.; Trinajstić, N., On general sum-connectivity index. *JOMC* **2010,** *47* (1), 210-218.

[9] Feng, L.; Ilić, A., Zagreb, Harary and hyper-Wiener indices of graphs with a given matching number. *Applied Mathematics Letters* **2010,** *23* (8), 943-948.

[10] Marrero-Ponce, Y.; Santiago Martínez, O.; López Martínez, Y.; Barigye, S. J.; Torrens, F., Derivatives in discrete mathematics: a novel graph-theoretical invariant for generating new 2/3D molecular descriptors. I. Theory and QSPR application. *JCAMD* **2012,** *26* (11), 1229-1246.

[11] Barigye, S. J.; Marrero-Ponce, Y.; Martínez López, Y.; Martínez Santiago, O.; Torrens, F.; García Domenech, R.; Galvez, J., Event-Based Criteria in GT-STAF Information Indices: Theory, Exploratory Diversity Analysis and QSPR Applications. *SAR & QSAR Environ. Res.* **2013,** *24*, 3-34.

[12] Barigye, S. J.; Marrero-Ponce, Y.; Martínez López, Y.; Artiles Martínez, L. M.; Pino-Urias, R. W.; Martínez Santiago, O.; Torrens, F., Relations Frequency Hypermatrices in Mutual, Conditional and Joint Entropy-Based Information Indices. *J. Comput. Chem.* **2013,** *34*, 259-274.

[13] Barigye, S. J.; Marrero-Ponce, Y.; Martínez-López, Y.; Martínez-Santiago, O.; Perez-Giménez, F., N-tuple Relations Frequency Matrix Derivations from Event-based Criteria in GT-STAF Information Indices. *J Comp. Aided Mol. Des.* **2013**, En revisión.

[14] García-Jacas, C. R.; Marrero-Ponce, Y.; Acevedo-Martínez, L.; Barigye, S. J.; Valdés-Martiní, J. R.; Contreras-Torres, E., QuBiLS-MIDAS: A parallel free-software for molecular descriptors computation based on multilinear algebraic maps. *Journal of Computational Chemistry* **2014**.

[15] Basak, S.; Gute, B., Characterization of molecular structures using topological indices. *SAR and QSAR in Environmental Research* **1997,** *7* (1-4), 1-21.

[16] Detyniecki, M., Fundamentals on aggregation operators. *This manuscript is based on Detyniecki's doctoral thesis and can be downloaded from* **2001**.

[17] Todeschini, R., Consonni V., New Local Vertex Invariants and Molecular Descriptors Based on Functions of the Vertex Degrees. *MATCH Commun. Math. Comput. Chem* **2010,***64*, 359-372.

[18] Kier, L. B.; Hall, L. H., *Molecular Structure Description. The Electrotopological State*. Academic Press: New York, 1999.

[19] Harary, F.; Palmer, E.; Robinson, R.; Read, R.; Balaban, A., In Chemical Applications of Graph Theory. *In Chemical Applications of Graph Theory* **1976**.

[20] Kier, L. B.; Hall, L. H., An electrotopological-state index for atoms in molecules. *Pharm Res* **1990,***7* (8), 801-7.

[21] Kupchik, E. J., Structure—Molar Refraction Relationships of Alkylgermanes Using Molecular Connectivity. *Quantitative Structure-Activity Relationships* **1988,***7* (2), 57-59.

[22] Hu, Q.-N.; Liang, Y.-Z.; Yin, H.; Peng, X.-L.; Fang, K.-T., Structural Interpretation of the Topological Index. 2. The Molecular Connectivity Index, the Kappa Index, and the Atom-type E-State Index. *J. Chem. Inf. Comput. Sci.* **2004,***44*, 1193-1201.

[23] Li, X.; Jalbout, A.; Solimannejad, M., Definition and application of a novel valence molecular connectivity index. *J. Mol. Struc-THEOCHEM* **2003,** *663* (1), 81-85.

[24] Alikhanidi, S.; Takahash, Y., New Molecular Fragmental Descriptors and Their Application to the Prediction of Fish Toxicity. *MATCH Commun.Math.Comput.Chem* **2006,***55*, 205-232.

[25] Ivanciuc, O., Design on topological indices. 1. Definition of a vertex topological index in the case of 4-trees. *Revue Roumaine de Chimie* **1989,***34* (6), 1361-1368.

[26] Cramer, R. D.; Patterson, D. E.; Bunce, J. D., Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *JACS* **1988,***110* (18), 5959-5967.

[27] Atkinson, A. C., *Plots, transformations, and regression: an introduction to graphical methods of diagnostic regression analysis*. Clarendon Press Oxford: 1985.

[28] Robert, D.; Amat, L.; Carbo-Dorca, R., Three-Dimensional Quantitative-Activity Relationships from Tuned Molecular Quantum Similarity Measures: Prediction of the Corticosteroid-Binding Globulin Binding Affinity for a Steroid Family *J. Chem. Inf. Comput. Sci.* **1999,***39*, 333-344.

[29] Silverman, B. D.; Platt, D. E., Comparative molecular moment analysis (CoMMA): 3D-QSAR without molecular superposition *J. Med. Chem.* **1996,***39*, 2129-2140.

[30] Golbraikh, A.; Bonchev, D.; Tropsha, A., Novel chirality descriptors derived from molecular topology. *J. Chem. Inform. Comput. Sci.* **2001,***41* (1), 147-158.

[31] De, K.; Sengupta, C.; Roy, K., QSAR modeling of globulin binding affinity of corticosteroids using AM1 calculations. *Bioorg. Med. Chem.* **2004,***12*, 3323-3332.

[32] Martínez-López , Y.; Marrero-Ponce, Y.; SJ., B. MD-LOVIs for windows: Software for molecular descriptors calculator from LOVIs. Version 1.0. . 2010.

[33] Galvez, J.; García-Domenech, R., Improving the local vertex invariants in alkane graphs through a standard molecular orbital approach. *Chem. Phys. Lett.* **2007,***449* (1), 249-254.

[34] Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M.; Leardi, R., MobyDigs: software for regression and classification models by genetic algorithms. In *Data Handling in Science and Technology*, Elsevier: 2003; Vol. Volume 23, pp 141-167.

[35] Tuppurainen, K.; Viisas, M.; Peräkylä, M.; Laatikainen, R., Ligand intramolecular motions in ligand-protein interaction: ALPHA, a novel dynamic descriptor and a QSAR study with extended steroid benchmark dataset. *JCAMD* **2004,***18*, 175–187.

[36] Coats, E. A., The CoMFA steroids as a benchmark dataset for development of 3D QSAR methods. *Perspect. Drug DiscoVery Des.* **1998,***12-14*, 199-213.

[37] Hodge, V. J.; Austin, J., A Survey of outlier detection methodologies. *Artif. Intell. Rev.* **2004,***22*, 85–126.

[38] Broto, P.; Moreau, G.; Vandycke, C., Molecular structures: perception, autocorrelation descriptor and sar studies: system of atomic contributions for the calculation of the n-octanol/water partition coefficients. *Eur. J. Med. Chem.* **1984,***19* (1), 71-78.

[39] Aires-de-Sousa, J.; Gasteiger, J.; Gutman, I.; Vidovic, D., Chirality codes and molecular structure. *J. Chem. Inform. Comput. Sci.* **2004,***44* (3), 831-836.

[40] So, S. S.; Karplus, M., Three-dimensional quantitative structure-activity relationships from molecular similarity matrices and genetic neural networks. 1. Method and validations. *J Med Chem* **1997,***40*, 4347-4359.

[41] Alvarez-Ginarte, Y. M.; Crespo-Otero, R.; Marrero-Ponce, Y.; Noheda-Marin, P.; Garcia de la Vega, J. M.; Montero-Cabrera, L. A.; Ruiz García, J. A.; Caldera-Luzardo, J. A.; Alvarado, Y. J., Chemometric and chemoinformatic analyses of anabolic and androgenic activities of testosterone and dihydrotestosterone analogues. *Bioorg. Med. Chem.* **2008,***16* (12), 6448-6459.

[42] Bajzer, Z.; Randic, M.; Plavšic, D.; Basak, S. C., Novel map descriptors for characterization of toxic effects in proteomics maps. *J. Mol. Graph. Model* **2003,***22*, 1-9.

[43] Ajay, N. J.; Kimberle, K.; Chapman, D., Compass: Predicting Biological Activities from Molecular Surface Properties. Performance Comparisons on a Steroid Benchmark. *J. Med. Chem.* **1994,***37*, 2315-2327.

[44] *ADRIANA.Code*, Molecular Networks: Erlangen, Germany, 2011.

[45] Alcalá-Fdez, J.; Sánchez, L.; García, S.; Jesus, M. J. d.; Ventura, S.; Garrell, J. M.; Otero, J.; Romero, C.; Bacardit, J.; Rivas, V. M.; Fernández, J. C.; Herrera, F., KEEL: a software tool to assess evolutionary algorithms for data mining problems. *Soft Comput.* **2009,***13*, 307–318.

[46] Pastor, M.; Cruciani, G.; McLay, I.; Pickett, P.; Clementi, S., GRid-INdependent Descriptors (GRIND): A Novel Class of Alignment-Independent Three-Dimensional Molecular Descriptors. *J. Med. Chem.* **2000,** *43*, 3233-3243.

[47] Kubinyi, H.; Hamprecht, F. A.; Mietzner, T., Three-Dimensional Quantitative Similarity-Activity Relationships (3D QSiAR) from SEAL Similarity Matrices. *J. Med. Chem.* **1998,** *41*, 2553-2564.

[48] Ajmani, S.; Jadhav, K.; Kulkarni, S. A., Three-dimensional QSAR using the k-nearest neighbor method and its interpretation. *J. Chem. Inf. Model.* **2006,** *46* (1), 24-31.

[49] Parretti, M. F.; Kroemer, R. T.; Rothman, J. H.; R., G. W., Alignment of Molecules by the Monte Carlo Optimization of Molecular Similarity Indices. *J. Comput. Chem.* **1997,** *18*, 1344-1353.

[50] Aires-de-Sousa, J.; Gasteiger, J.; Gutman, I.; Vidovic, D., Chirality Codes and Molecular Structure. *J. Chem. Inf. Comput. Sci.* **2004,** *44*, 831-836.

[51] Marín, R. M.; Aguirre, N. F.; Daza, E. E., Graph Theoretical Similarity Approach To Compare Molecular Electrostatic Potentials. *J. Chem. Inf. Model.* **2008,** *48*, 109-118.

[52] Basak, S. C.; Magnuson, V. R., Molecular topology and narcosis. A quantitative structure-activity relationship (QSAR)study of alcohols using complementary information content (CIC). *Arzneim.-Forsch./Drug Res.* **1983,** *33*, 501-503.

[53] Polanski, J.; Bak, A., Modeling Steric and Electronic Effects in 3D- and 4D-QSAR Schemes: Predicting Benzoic pKa Values and Steroid CBG Binding Affinities. *J. Chem. Inf. Comput. Sci.* **2003,** *43*, 2081-2092.

[54] Silverman, B. D.; Platt, D. E.; Pitman, M.; Rigoutsos, I., *Comparative molecular moment analysis (COMMA)*. Kluwer Academic Publishers: Dordrecht, The Netherlands, 1998; Vol. 3, p 183–196.

[55] Wagener, M.; Sadowski, J.; Gasteiger, J., Autocorrelation of Molecular Surface Properties for Modeling Corticosteroid Binding Globulin and Cytosolic Ah receptor. *J. Am. Chem. Soc.* **1995,** *117*, 7769-7775.

[56] Robinson, D. D.; Winn, P. J.; Lyne, P. D.; Richards, W. G., Self-organizing molecular field analysis: A tool for structure-activity studies. *J. Med. Chem.* **1999,***42*, 573–583.

Received: April 03, 2015 Revised: September 01, 2015 Accepted: September 30, 2015