OCWLGI Descriptors: Theory and Praxis

Andrey A. Toropov[*,1], Alla P. Toropova[1], Emilio Benfenati[1] and Giuseppina Gini[2]

[1]*IRCCS - Istituto di Ricerche Farmacologiche Mario Negri, 20156, Via La Masa 19, Milano, Italy*

[2]*Department of Electronics and Information, Politecnico di Milano, piazza Leonardo da Vinci 32, 20133 Milano, Italy*

Abstract: The aim of this review is description of the logic and evolution of optimal descriptors OCWLGI calculated with the molecular graph and the demonstration of their ability as tools for the modeling of biological and physicochemical parameters of chemical compounds. The ability of optimal descriptors calculated with hydrogen suppressed graph (HSG), hydrogen filled graph (HFG) and graph of atomic orbitals (GAO) is demonstrated as a collection of quantitative structure-property relationships (QSPR) and quantitative structure-activity relationships (QSAR) for properties and endpoints available from the literature. The Monte Carlo method optimization of the correlation weights of local and global invariants (OCWLGI) of molecular graphs is used as the principle for building up descriptors which are discussed in this article. The statistical quality of the QSPR and QSAR models for physicochemical and biological properties which were obtained with the optimal descriptors are reasonably high.

Keywords: Global invariant, local invariant, Monte Carlo method, optimization of correlation weights of local and global invariants of graph (OCWLGI), quantitative structure-property relationships, QSPR, quantitative structure-activity relationships, QSAR.

INTRODUCTION

Many quantitative structure-property relationships (QSPR) and quantitative structure-activity relationships (QSAR) models are influenced by the work of Wiener on molecular graphs, topological indices and prediction of physicochemical properties [1-4]. The main idea of these studies is the use of molecular graphs, molecular matrices and topological indices that may be correlated with biological and physicochemical properties of organic compounds. Starting in the 1980's, the number of topological indices conceptually related to the Wiener number started to increase [5-14]. Most of these descriptors or indices were based on two matrices, namely the adjacency matrix and the distance matrix. Consider the molecular graph of 2-methyl butane, with the vertex labelling as shown in Fig. (**1**).

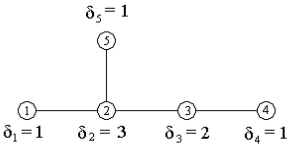

Fig. (1). The molecular graph of 2-methyl butane (CAS 78-78-4).

The adjacency matrix **A**(*G*) and the distance matrix **D**(*G*) of 2-methyl butane are:

A(*G*)

	1	2	3	4	5
1	0	1	0	0	0
2	1	0	1	0	1
3	0	1	0	1	0
4	0	0	1	0	0
5	0	1	0	0	0

D(*G*)

	1	2	3	4	5
1	0	1	2	3	2
2	1	0	1	2	1
3	2	1	0	1	2
4	3	2	1	0	3
5	2	1	2	3	0

*Address correspondence to this author at the Istituto di Ricerche Farmacologiche Mario Negri, 20156, Via La Masa 19, Milano, Italy;
Tel: +3902 3901 4595; Fax: +3902 3901 4735;
E-mail: andrey.toropov@marionegri.it

Although the topological indices that may be computed for chemical structures are diverse [15], the main principles in developing topological indices can be illustrated with the Wiener number (***W***) and connectivity indices of zero-order ($^0\chi$) and first-order ($^1\chi$), which is also known as the Randić index [8,12-15]. For example, these topological indices for 2-methyl butane (Fig. **1**) are calculated with the formulae:

$$W = (1/2)\Sigma\Sigma d_{ij} = (1/2)(0+1+2+3+2+ 1+0+1+2+1+2+1+0+1+2+ 3+2+1+0+3+2+1+2+3+0) = 18 \quad (1)$$

$$^0\chi = \Sigma\delta_i^{-1/2} = \delta_1^{-1/2} + \delta_2^{-1/2} + \delta_3^{-1/2} + \delta_4^{-1/2} + \delta_5^{-1/2} = (1)^{-1/2} + (3)^{-1/2} + (2)^{-1/2} + (1)^{-1/2} + (1)^{-1/2} = 1 + 0.577 + 0.707 + 1 + 1 = 4.284 \quad (2)$$

$$^1\chi = \Sigma\delta_i\delta_j^{-1/2} = (\delta_1\delta_2)^{-1/2} + (\delta_2\delta_3)^{-1/2} + (\delta_3\delta_4)^{-1/2} + (\delta_2\delta_5)^{-1/2} = (1\cdot 3)^{-1/2} + (3\cdot 2)^{-1/2} + (2\cdot 1)^{-1/2} + (3\cdot 1)^{-1/2} = 0.577 + 0.408 + 0.707 + 0.577 = 2.269 \quad (3)$$

where δ_i is the vertex degree equal to the sum of the elements of the adjacency matrix in a row, and d_{ij} is the element of the **D**(*G*) matrix. Detailed information on other structural descriptors applied in the last years to develop QSPR and QSAR models may be found in [15, 16]. Some topics related to the investigation of the topological indices are related to their definition. For instance, why the exponent for the vertex degree in Eq. (2) and Eq. (3) have been selected as –1/2, and not another value? Indeed, the QSAR analysis of several molecular properties [17] has shown that in most cases the optimal value of the degree, in models based on descriptors similar to the Eq.(2) or Eq. (3), is different from -1/2.

QSPR and QSAR studies [18-21] have shown that the correction of the adjacency matrix (*e.g.* ethyl isopropyl

1875-6697/13 $58.00+.00

sulphide, CAS 5145-99-3, Fig. **2**) by means of replacing the zero values on the diagonal of the adjacency matrix with special selected coefficients *x* and *y* (Fig. **3**) may produce considerable improvement of correlation between the connectivity indices ($^0\chi$ and $^1\chi$) and various properties.

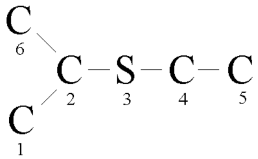

Fig. (2). The hydrogen-suppressed graph (HSG) of ethyl isopropyl sulfide (CAS 5145-99-3).

ADJACENCY MATRIX

	1	2	3	4	5	6	δ_i
1	0	1	0	0	0	0	1
2	1	0	1	0	0	1	3
3	0	1	0	1	0	0	2
4	0	0	1	0	1	0	2
5	0	0	0	1	0	0	1
6	0	1	0	0	0	0	1

MODIFIED ADJACENCY MATRIX

	1	2	3	4	5	6	δ_i
1	x	1	0	0	0	0	$1+x$
2	1	x	1	0	0	1	$3+x$
3	0	1	y	1	0	0	$2+y$
4	0	0	1	x	1	0	$2+x$
5	0	0	0	1	x	0	$1+x$
6	0	1	0	0	0	x	$1+x$

It has to be noted that by adding these parameters on the diagonal of the adjacency matrix, the calculated values for δ_i are also modified. Similar modifications have been carried out for the graph distance matrix [22], and the result was a significant improvement of the QSPR models. These descriptors have been named "variable" or "flexible", however, we will use the term "optimal descriptors", because descriptors which are discussed here are calculated by means of the Monte Carlo method optimization. All these optimal descriptors [18-22] have been computed from the hydrogen–suppressed molecular graph.

CONSTRUCTION OF OPTIMAL DESCRIPTORS

The first optimal descriptor, based on the hydrogen-suppressed graph, has been suggested by Randić [23, 24]. The main idea of the approach was to use diagonal entries of the adjacency matrix to consider the influence of heteroatoms, similarly to the well known generalization of the Hückel molecular orbitals method [23].

The optimal descriptors have been successfully used for the QSPR modeling of aliphatic alcohols [25], nitrogen-containing compounds [26], and sulfides [27]. Later on, optimal descriptors based on the hydrogen-filled graph have been suggested [28].

The optimization target may be the standard error of estimation (*s*) [23-27] (*i.e*. searching for minimum of *s*), or the correlation coefficient (*r*) [28, 29] (*i.e*. searching for maximum of *r*), and by comparing the results obtained with these target functions it has been shown that the maximization of correlation coefficient gave models with better statistical quality [29].

OPTIMAL DESCRIPTORS BASED ON HSG AND HFG

As an example of the general scheme based on the hydrogen-suppressed graphs we consider ethyl isopropyl sulfide (Fig. **2**). Accordingly to [30], the use of x=+0.25 and y=-0.95 in the calculation of the optimal connectivity index $^1\chi$ (x,y) for the correlation with normal boiling points of 21 sulfides gave the more accurate model in comparison with the standard formula with fixed coefficients $^1\chi$ (0,0).

An optimal descriptor is a modification of fixed coefficients (e.g., vertex degree, path of length 2, etc.), in the calculation of a "classic" topological index. This modification is made by using parameters that improve the statistical indices of the QSPR and QSAR models. In other words, each descriptor is a mathematical function of the representation of the molecular structure (MSR). Any MSR contains molecular invariants (MI), which define the molecular individuality. An MSR can be represented by

$$D = F(MSR) = F(MI_1, MI_2, ..., MI_m) \quad (4)$$

where MI_k is the *k*-th molecular invariant (k = 1, …, m, where m is the total number of molecular invariants in the molecule). The descriptor that is calculated with Eq. (4) is the fixed version which is similar to descriptors calculated with Eq.(1), Eq.(2), and Eq.(3).

Formula (4) can be modified by replacing the fixed components MI_k with flexible ones $CW(MI_k)$:

$$D = F(MSR) = F(CW(MI_1), CW(MI_2), ..., CW(MI_m) \quad (5)$$

where $CW(MI_k)$ is the correlation weight of the *k*-th molecular invariant. The descriptor *D* calculated with Eq. (5) is a flexible version of the descriptor calculated with Eq. (4). The correlation weights $CW(MI_k)$ are numerical coefficients used in the calculation with Eq.(5).

The correlation coefficient between a descriptor calculated with Eq. (5) and the property/activity (*PA*) of interest is also a mathematical function of the CWs,

$$R(PA, D) = R[PA, F(CW(MI_1), CW(MI_2), ..., CW(MI_m)) \quad (6)$$

where $R(PA,D)$ is the correlation coefficient between the *PA* and *D*, calculated with Eq. (5).

The optimization of the parameters $CW^*(MI_1)$, $CW^*(MI_2)$, … $CW^*(MI_m)$ is performed with the Monte Carlo method, which results in a maximum value for the $R(PA,D)$ for the training set of compounds. The predictive ability of the model can be tested with an external set of compounds.

As a possible extension of the scheme based on the Eq. (6), these correlation weights can be calculated not only for

numerical invariant of a molecular graph such as vertex degrees [28], extended connectivity of increasing orders [31], paths of length 2 and 3 [9], valence shells of increasing orders [9], but also for non-numerical features of a molecular structure, such as presence of different atoms, presence/absence of different cycles, and so on. The optimization of correlation weights of the local and global graph invariants (OCWLGI) is the basic principle of building up optimal descriptors considered in this article. Thus, these descriptors can be named OCWLGI descriptors.

Owing to this possibility (involving of local and global invariants in the modeling) one can estimate a measure of the influence for a given molecular attribute (invariant of molecular graph) on the property/activity of interest, that can be used as a hint on the study of mechanism related to the phenomena (property/activity) under consideration.

A comparison of the hydrogen-suppressed graph based and the hydrogen-filled graph (HFG) based optimal descriptors has been carried out [32]. It has been shown that the optimal descriptor based on the hydrogen-filled graph gave better models for the normal boiling points of alkyl alcohols.

OCWLGI DESCRIPTORS BASED ON GAO

The graph of atomic orbitals (GAO) is an attempt to take into account the structure of atoms in QSPR/QSAR analysis [31-37]. The conversion of the hydrogen filled graph into GAO can be carried out by the scheme:

1. Each vertex of the hydrogen filled graph is replaced by the group of atomic orbitals. Such groups of the atomic orbitals are listed in Table **1**.

2. Elements of the adjacency matrix of the graph of atomic orbitals are defined as

$$a_{ij} = \begin{cases} 1, & \text{if } i-th \text{ and } j-th \text{ GAO vertices fall in different atoms in HFG and these atoms have joint edge in HFG} \\ 0, & \text{otherwise} \end{cases}$$

The groups of atomic orbitals for some chemical elements are listed in Table **1**.

For a training set of graphs of atomic orbitals, one can carry out the same Monte Carlo optimization of correlation weights of the invariants using the same algorithms [31-38]. However, the models based on the HSG or HFG and GAO-based model are different. For the same list of compounds the number of different vertexes as well as vertex degree values for the GAO representation is larger. This approach improves the statistical quality of a model (calculated with the optimal descriptors) for the training set, but the statistical quality for the external test set can be poor [31]. Consequently, one should be careful with the GAO representation of chemical compounds, because it may lead to overtraining [31]. However, quite good GAO-based models are also possible [31-34].

Table 1. Groups of Atomic Orbitals, for Some Chemical Elements, Used in Constructing the Graph of Atomic Orbitals

Chemical Elements	Group of Atomic Orbitals
H	$1s^1$
C	$1s^2, 2s^2, 2p^2$
N	$1s^2, 2s^2, 2p^3$
O	$1s^2, 2s^2, 2p^4$
F	$1s^2, 2s^2, 2p^5$
P	$1s^2, 2s^2, 2p^6, 3s^2, 3p^3$
S	$1s^2, 2s^2, 2p^6, 3s^2, 3p^4$
Cl	$1s^2, 2s^2, 2p^6, 3s^2, 3p^5$
Br	$1s^2, 2s^2, 2p^6, 3s^2, 3p^6, 3d^{10}, 4s^2, 4p^5$

For the hydrogen-filled graph of nitrapyrin (CAS 1929-82-4) with vertex numbering shown in Fig. (**4**), this conversion gives the GAO shown in Fig. (**2**).

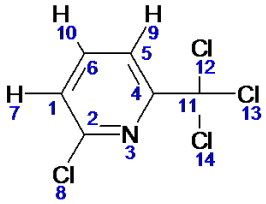

Fig. (3). The numbered HFG of nitrapyrin (CAS 1929-82-4).

COLLECTION OF QSPR/QSAR BASED ON OCWLGI DESCRIPTORS

The QSPR/QSAR models in this study are characterized by the number of compounds in a dataset (n); the square of correlation coefficient (r^2); standard error of estimation (s); and Fischer F-ratio. The generalized form of the optimal descriptor is the following:

$$DCW = CW(G) + \{ \sum_{k=1}^{N} CW(A_k) + \sum_{k=1}^{N} W(VI_k) \} \quad (7)$$

where N is the number of vertex in molecular graph, *i.e.* the number of atoms in the case of HSG (Fig. **2**) and HFG (Fig. **4**) and the number of AO in the case of GAO (Fig. **5**); $CW(A_k)$ is the correlation weight of chemical element (or AO in the case of the GAO) which is an image of the k-th vertex; $CW(VI_k)$ is the correlation weight of the invariant of k-th vertex such as vertex degree, Morgan extended connectivity [31, 38], valence shells [9, 47], the number of paths of length 2 or 3 [47]; $CW(G)$ is the correlation weight of a global invariant of the molecular graph, such as the number of cycles [7, 47], hydrogen bond indices [39], etc. As an alternative to the additive scheme for optimal descriptor (*i.e.* Eq. 7), the multiplicative scheme [31] can be used. Table **2** contains a collection of QSPR/QSAR models which were built up with OCWLGI descriptors.

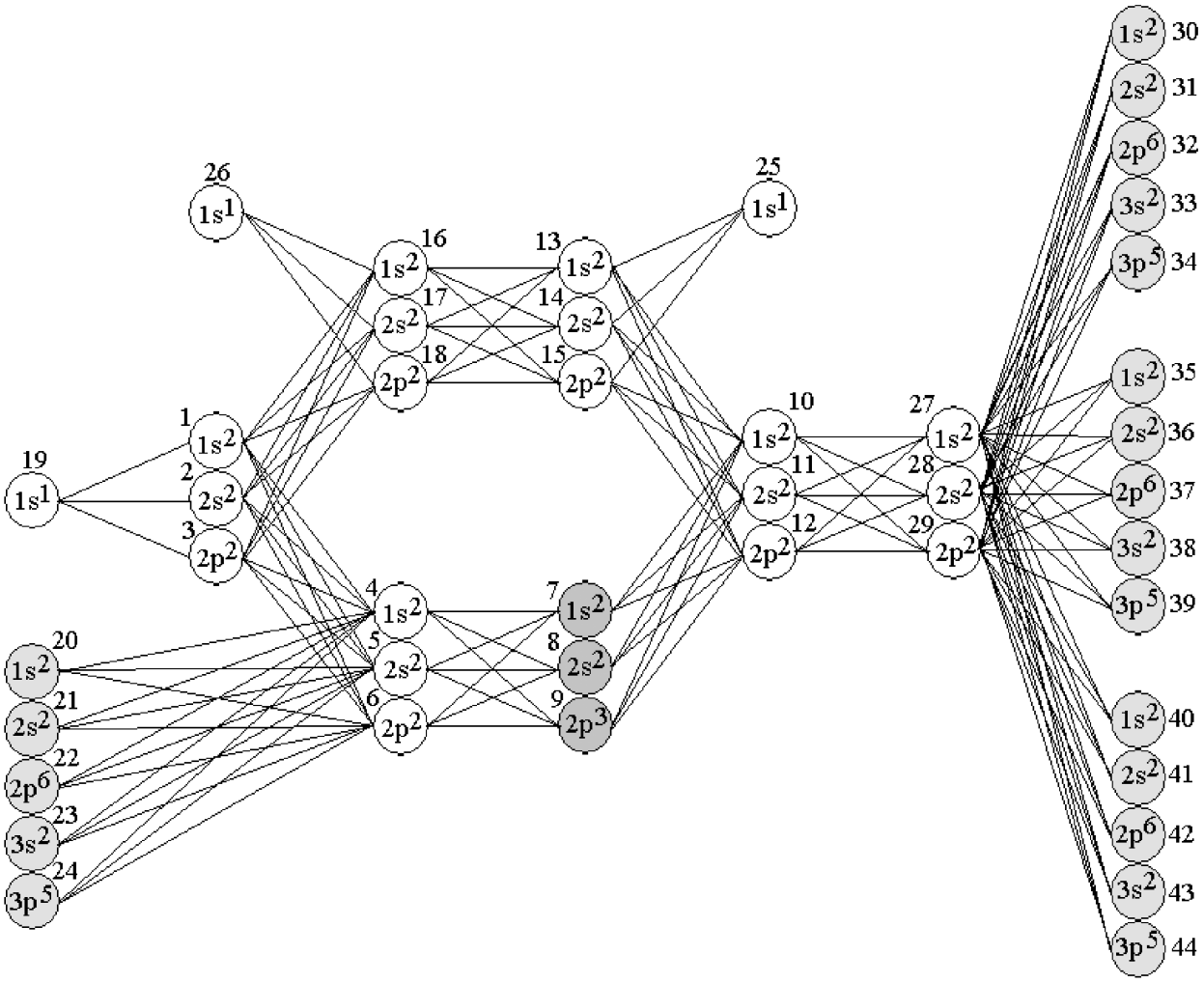

Fig. (4). The numbered, accordingly HFG from Fig. (**3**), GAO of nitrapyrin.

DISCUSSION

The model for normal boiling points [48] (developed with the same chemical compounds as for *Model 7*, Table **2**) is characterized by n = 134, r^2 = 0.9886, s = 2.86 ^{o}C, F = 3770. Thus, the statistical characteristics for the *Model 7* are better. The statistical indices of the best QSPR for molar heat capacity (at 300^{o}K) [48] are n = 134, r^2 = 0.9771, s = 3.89 J·K^{-1} mol^{-1}, F = 1848. The statistical characteristics of the *Model 8* are better (Table **2**). The best QSPR for standard Gibbs energy of formation in the gas phase at 300^{o}K has n = 134, r^2 = 0.9172, s = 4.29 kJ/mol, F = 480 [48]. The statistical quality of the *Model 9* is better. The best model for alkane vaporization enthalpy at 300^{o}K is characterized by n = 134, r^2 = 0.9801, s = 0.61 kJ/mol, F = 2138 [48]. The statistical quality of the *Model 10* is better. The best QSPR for refractive index n_D^{25} [48] is characterized by n = 134, r^2 = 0.9683, s = 0.0025, F = 1309. The statistical quality of the *Model 11* is better. Finally, the statistical quality of the best model for alkane density is n = 134, r^2 = 0.9805, s = 3.73 kg/m^3, F = 2156 [48]. The statistical quality of the *Model 12* is better. It is to be noted, that in ref. [48] the external test set is absent.

A QSAR analysis based on the quantum chemical descriptors for 57 anti-HIV-1 agents of tetrahydro-imidazo[4,5,l-*jk*][1,4]-benzodiazepin-2-(1*H*)-one (TIBO) derivatives together with 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine (HEPT) derivatives is presented. The QSAR model for TIBO derivatives has r^2 = 0.8649, s = 0.597 [49], and the statistical quality of the QSAR model for HEPT derivatives is r^2 = 0.9063, s = 0.371. The *Model 15* is related to both TIBO and HEPT derivatives. The model is checked with the external test dataset. Thus, the statistical quality of *Model 15* and statistical quality of models from ref. [49] should be considered as similar.

The statistical indices of the best model for the acute aquatic toxicity are n = 69, r^2 = 0.863, s = 0.30 [50]. The statistical characteristics for *Model 19* are better. In conclusion, these comparisons with models for the same endpoints from the literature indicate that the optimal OCWLGI-descriptors can be useful for the QSPR/QSAR analyses. Finally, we note that there are several topological indices which can be translated into their OCWLGI type of molecular descriptors [54-61].

Table 2. A Collection of QSPR/QSAR Based on the OCWLGI-Descriptors

No.	Description	Statistical Characteristics	Refs.
Physicochemical Parameters			
1	QSPR models for normal boiling points of alkanes, alkylbenzenes, and polyaromatic hydrocarbons	n = 70, r^2 = 0.9988, s = 5.8° C, F = 57437 (training set) n = 70, r^2 = 0.9985, s = 6.7° C, F = 45154 (test set)	[38]
2	QSPR modelling of the constants of stability of 110 biometal M^{2+} complexes with α-amino acids and phosphate derivatives of adenosine	n = 55, r^2 = 0.9843, s = 0.279, F = 3328 (Training set) n = 55, r^2= 0.9935, s = 0.248, F = 4027 (Test set)	[39]
3	QSPR model of normal boiling points of alcohols	n=29, r^2=0.9906, s=2.9°C F=5733 (training set) n=29, r^2=0.9896, s=3.0°C F=2595 (test set)	[32]
4	QSPR model of normal boiling points of acyclic carbonyl compounds	n = 100, r^2 = 0.972, s = 6.12°C, F = 3464 (training set) n = 100, r^2 = 0.975, s = 6.00°C, F = 3905 (test set)	[40]
5	QSPR model of normal boiling points of normal boiling points of haloalkanes (GAO)	n = 138, r^2 = 0.9841, s = 9.80°C, F = 3464 (training set) n = 138, r^2 = 0.9854, s = 7.39°C, F = 3905 (test set)	[35]
6	The challenged study involved predictions of normal boiling points for organic molecules of varied composition. These molecules included species with both linear and cyclic structures, comprise ketones, esters, aldehydes, nitriles, amines, alcohols, and hydrocarbons and a wide variety of atoms, such as C, H, O, N, Si, Cl, Br, F, P, and S	n = 126, r^2 = 0.9279, s = 33.3°C, F = 1599 (training set); n = 32, r^2= 0.8819, s = 39.1°C, F = 224 (test set)	[41]
7	The short list of additional examples relevant to this review includes: Normal boiling points of alkanes	n = 67, r^2 = 0.9984, s = 1.126°C, F = 39180 (training set); n = 67, r^2 = 0.9910, s = 2.553°C, F = 7118 (test set)	[31]
8	Molar heat capacity at 300 K (J/(K mol)) of alkanes	n = 67, r^2 = 0.9892, s = 2.790, F = 5869 (training set); n = 67, r^2 = 0.9902, s = 2.442, F = 6607 (test set)	[31]
9	Standard Gibbs energy of formation in the gas state at 300 K (kJ/mol) of alkanes	n = 67, r^2 = 0.9884, s = 1.804, F = 5465 (training set); n = 67, r^2 = 0.9803, s = 1.791, F = 3224 (test set)	[31]
10	Vaporization enthalpy at 300 K (kJ/mol) of alkanes	n = 67, r^2 = 0.9884, s = 0.471, F = 5428 (training set); n = 67, r^2 = 0.9858, s = 0.537, F = 4529 (test set)	[31]
11	Refractive index at 25°C of alkanes	n = 67, r^2 = 0.9735, s = 0.0024, F = 2385 (training set); n = 67, r^2 = 0.9627, s = 0.0024, F = 1684 (test set)	[31]
12	Density (kg/m^3) of alkanes	n = 67, r^2 = 0.9844, s = 3.602, F = 4036 (training set) n = 67, r^2 = 0.9763, s = 3.790, F = 2680 (test set)	[31]
13	Flory-Huggins parameter for binary polymer-solvent mixtures	n = 30, r^2 = 0.9990, s = 0.028, F = 27537 (training set); n = 30, r^2 = 0.9972, s = 0.053, F = 10294 (test set);	[42]
14	The intrinsic viscosity of polymers	n = 17, r^2 = 0.9130, s = 0.126 cm^3/g, F = 157 (training set); n = 9, r^2 = 0.9231, s = 0.143 cm^3/g, F = 84 (test set).	[42]
Biological Activity			
15	Anti-HIV-1 activity TIBO and HEPT derivatives	n = 37, r^2 = 0.8688, s = 0.557, F = 232 (training set) n = 20, r^2 = 0.8759, s = 0.588, F = 127 (test set)	[43]
16	Toxicity, *V. fischeri*, log(1/IGC_{50}), valence shells has been used as local graph invariant	n = 45, r^2 = 0.8299, s = 0.402, F = 210 (training set) n = 21, r^2 = 0.8902, s = 0.339, F = 154 (test set)	[44]
17	Toxicity to *Tetrahymena pyriformis* of heterogeneous set of benzene derivatives	n = 157, r^2 = 0.883, s = 0.27, F = 1170 (training set); n = 60, r^2 = 0.863, s = 0.28, F = 372 (test set);	[45]
18	The mutagenic activities of 95 heteroaromatic compounds in *S. typhimurium* TA98 S9, graph of atomic orbitals	n = 47, r^2 = 0.7637, s = 1.05, F = 145 (training set); n = 48, r^2 = 0.7569, s = 0.86, F = 144 (test set)	[34]
19	Aquatic toxicity, *Pimephales promelas*, log(1/LC50), Morgan extended connectivity is used as local graph invariant	n = 44, r^2 = 0.8982, s = 0.251, F = 371 (training set) n = 25, r^2 = 0.9181, s = 0.234, F = 258 (test set)	[46]
20	Acute toxicity LC_{50}-96h to rainbow trout (*Oncorhynchus mykiss*) of 274 organic pesticides	n = 233, r^2 = 0.7689, s = 0.75, F = 769 (training set) n = 41, r^2 = 0.6421, s = 1.14, F = 70 (test set)	[47]
21	Carcinogenic activity of methylated polycyclic aromatic hydrocarbons	n=30, r^2=0.8909, s=0.689 (training set) n=16, r^2=0.9247, s=0.594 (test set)	[51]
22	Lipophilicity (logP) of 76 industrial chemicals	n=36, r^2=0.8857, s=0.500, F=279 (training set) n=36, r^2=0.9251, s=0.382, F=414 (test set)	[52]
23	The mutagenic activities of these compounds in *S. typhimurium* TA100 + S9 microsomal preparation are expressed in log of revertant per nonamole, ln *R*.	n=36, r^2=0.6446, s=0.861, F=62 (training set) n=37, r^2=0.7843, s=0.616, F=142 (test set)	[53]

CONCLUSIONS

The OCWLGI descriptors calculated from the molecular graph are a robust approach for the predictive modeling of physicochemical and biomedical properties. Our experience with the OCWLGI descriptors may be summarized in the following rules to obtained high quality QSPR and QSAR models: (1) complex molecular descriptors, such as the Morgan extended connectivity, or valence shells, may result in overfitting and poor prediction statistics; (2) the QSPR and QSAR models based on the OCWLGI descriptors should be validated by splitting the dataset into a training dataset and a test dataset, and then repeating several times this procedure; (3) infrequent structural descriptors computed from the molecular graph should be removed from the process of training a QSPR or QSAR model. The CORAL software (http://www.insilico.eu/coral/) is available for building up described models.

CONFLICT OF INTEREST

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

ACKNOWLEDGEMENTS

The authors express their gratitude to CALEIDOS (the project number LIFE11-INV/IT 00295) for financial support. Also the authors express their gratitude to Dr. L. Cappellini, Dr. R. Bagnati, and Dr. G. Bianchi for technical assistance. This paper is dedicated to Professor Milan Randi on the occasion of his 80 birthday.

ABBREVIATIONS

OCWLGI = Optimization of correlation weights of local and global invariants of graph

QSPR/QSAR = Quantitative structure-property/activity relationships

HFG = Hydrogen-filled graph

HSG = Hydrogen-suppressed graph

GAO = Graph of atomic orbitals

REFERENCES

[1] Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.*, **1947**, *69*, 17-20.

[2] Wiener, H. Correlation of heats of isomerization, and differences in heats of vaporization of isomers, among the paraffin hydrocarbons. *J. Am. Chem. Soc.*, **1947**, *69*, 2636-2638.

[3] Wiener, H. Relation of the physical properties of the isomeric alkanes to molecular structure. surface tension, specific dispersion, and critical solution temperature in aniline *J. Phys. Chem.*, **1948**, *52*, 1082-1089.

[4] Wiener, H. Vapor pressure-temperature relationships among the branched paraffin hydrocarbons *J. Phys. Chem.*, **1948**, *52*, 425-430.

[5] Hosoya, H. Topological index as a sorting device for coding chemical structures. *J. Chem. Doc.*, **1972**, *12*, 181-183.

[6] Amidon, G.L.; Anik, S.T. Comparison of several molecular topological indexes with molecular surface area in aqueous solubility estimation. *J. Pharm. Sci.*, **1976**, *65*, 801-808.

[7] Bonchev, D.; Balaban, A.T.; Mekenyan, O. Generalization of the graph center concept, and derived topological centric indexes. *J. Chem. Inf. Comput. Sci.*, **1980**, *20*, 106-113.

[8] King, R.B. In: Chemical applications of topology and graph theory. Proceedings of a symposium held at the University of Georgia, Athens, Georgia, USA, **1983**

[9] Randić, M. Graph Valence Shells as Molecular Descriptors. *J. Chem. Inf. Comput. Sci.*, **2001**, *41*, 627-630.

[10] Randić, M.; Basak, S.C. On Use of the variable connectivity index $^{1}\chi^{f}$ in QSAR: Toxicity of aliphatic ethers. *J. Chem. Inf. Comput. Sci.*, **2001**, *41*, 614-618.

[11] Randić, M.; Plavšić, D.; Lers, N. Variable connectivity index for cycle-containing structures. *J. Chem. Inf. Comput. Sci.*, **2001**, *41*, 657-662.

[12] Roy, K.; Leonard, T.J. QSAR modelling of HIV-1 reverse transcriptase inhibitor 2-amino-6-arylsulfonylbenzonitriles and congeners using molecular connectivity and E-state parameters. *Bioorg. Med. Chem.*, **2004**, *12*, 745-754.

[13] Jalbout, A.F.; Li, X. Anti-HIV-1 inhibitors of various molecules using principles of connectivity. *J. Mol. Struct. (Theochem)*, **2003**, *663*, 19-23.

[14] Visco, Jr., D.P.; Poppale, R.S.; Rintoul, M.D.; Faulon, J.-L. Developing a methodology for an inverse quantitative structure-activity relationship using the signature molecular descriptor. *J. Mol. Graph. Mod.*, **2002**, *20*, 429-438.

[15] Accelrys, http://www.esi.umontreal.ca/accelrys/life/cerius46/qsar/theory_descriptors.html (Accessed March 1, **2010**)

[16] Hu, Q.; Liang, Y.; Fang, K. The Matrix Expression, Topological Index and Atomic Attribute of Molecular Topological Structure. *J. Data Sci.*, **2003**, *1*, 361-389.

[17] Amić, D.; Beslo, D.; Lucić, B.; Nikolić, S.; Trinajstić, N.; The Vertex-Connectivity Index Revisited. *J. Chem. Inf. Comput. Sci.*, **1998**, *38*, 819-822.

[18] Randić, M.; Basak, S.C.; Optimal Molecular Descriptors Based on Weighted Path Numbers. *J. Chem. Inf. Comput. Sci.*, **1999**, *39*, 261-266.

[19] Randić, M.; Pompe, M. The variable connectivity index $^{1}\chi^{f}$ versus the traditional molecular descriptors: A comparative study of $^{1}\chi^{f}$ against descriptors of CODESSA. *J. Chem. Inf. Comput. Sci.*, **2001**, *41*, 631-638

[20] Benfenati, E.; Carbó-Dorca, R.; Gini, G.; Grauel, A. Multiple toxicity prediction using different molecular descriptors, Proceeding of the V Girona Seminar on Molecular Similarity, Girona, Spain, July 12-20, **2001**.

[21] Randić, M.; Basak, S.C.; New Descriptor for structure-property and structure-activity correlations. *J. Chem. Inf. Comput. Sci.*, **2001**, *41*, 650-656.

[22] Randić, M.; Pompe, M. The variable molecular descriptors based on distance related matrices. *J. Chem. Inf. Comput. Sci.*, **2001**, *41*, 575-585.

[23] Randić, M. Novel graph theoretical approach to heteroatoms in quantitative structure—activity relationships. *Chemometr. Intel. Lab.*, **1991**, *10*, 213-227.

[24] Randić, M. On computation of optimal parameters for multivariate analysis of structure-property relationships. *J. Chem. Inf. Comput. Sci.*, **1991**, *12*, 970-980.

[25] Amić, D.; Basak, S.C.; Lucić, B.; Nikolić, S.; Trinajstić, N. Structure-Water solubility modelling of aliphatic alcohols using the weighted path number. *SAR and QSAR in Environ. Res.*, **2002**, *13*, 281-295.

[26] Randić, M.; Dobrowolski, J.Cz. Optimal molecular connectivity descriptors for nitrogen-containing molecules. *Int. J. Quant. Chem.*, **1998**, *70*, 1209-1215.

[27] Randić, M.; Basak, S.C. Construction of high-quality Structure-Property-Activity Regerssions: The boiling points of Sulfides. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 899-905.

[28] Toropov, A.A.; Toropova, A.P. Optimization of correlation weights of the local graph invariants: Use of the enthalpies of formation of complex compounds for the qspr modelling. *Russ. J. Coord. Chem.*, **1998**, *24*, 81-85.

[29] Toropov, A.A.; Toropova, A.P.; Voropaeva, N.L.; Ruban, I.N.; Rashidova, S.Sh. Two concepts of weighing molecular graph local invariants in QSPR modelling of the enthalpies of complexes: sampling of increments and optimization of correlation weights. *Russ. J. Coord. Chem.* **1999**, *25*, 618-623.

[30] Gutman, I.; Toropov, A.A.; Toropova, A.P. The graph of atomic orbitals and it's basic properties. 1. Wiener index. *MATCH Commun. Math. Comput. Chem.*, **2005**, *53*, 215-224.

[31] Toropov, A.A.; Toropova A.P. QSPR modelling of alkane properties based on graph of atomic orbitals. *J. Mol. Struct. (Theochem)*, **2003**, *637*, 1-10.

[32] Krenkel, G.; Castro, E.A.; Toropov, A.A. Improved molecular descriptors to calculate boiling points based on the optimization of correlation weights of local graph invariants. *J. Mol. Struct. (Theochem)*, **2001**, *542*, 107-113.

[33] Toropov, A.A.; Toropova, A.P. QSPR Modelling of the Formation Constants for Complexes Using Atomic Orbital Graphs. *Russ. J. Coord. Chem.* **2000**, *26*, 398-405.

[34] Toropov, A.A.; Toropova A.P. Prediction of heteroaromatic amine mutagenicity by means of correlation weighting of atomic orbital graphs of local invariants. *J. Mol. Struct. (Theochem)*, **2001**, *538*, 287-293.

[35] Toropov, A.A; Toropova A.P. Nearest neighboring code and hydrogen bond index in labeled hydrogen-filled graph and graph of atomic orbitals: application to model of normal boiling points of haloalkanes. *J.Mol. Struct. (Theochem)*, **2004**, *711*, 173-183.

[36] Toropov, A.A.; Gutman, I.; Furtula B. Graph of atomic orbitals and the molecular structure descriptors based on it. *J. Serb. Chem. Soc.*, **2005**, *70*, 669-674.

[37] Gutman, I.; Furtula, B.; Toropov, A.A.; Toropova A.P. The graph of atomic orbitals and it's basic properties. 2. Zagreb index. *MATCH Commun. Math. Comput. Chem.*, **2005**, 53, 225-230.

[38] Toropov, A.A.; Toropova, A.P.; Gutman, I. Comparison of QSPR models based on hydrogen-filled graphs and on graphs of atomic orbitals. *Croat. Chem. Act.*, **2005**, *78*, 503-509.

[39] Toropov, A.A.; Toropova, A.P. QSPR modelling of complex stability by optimization of correlation weights of the hydrogen bond index and the local graph invariants. *Russ. J. Coord. Chem.*, **2002**, *28*, 877-880.

[40] Toropov, A.A.; Toropova, A.P. Modelling of acyclic carbonyl compounds normal boiling points by correlation weighting of nearest neighboring codes. *J. Mol. Struct. (Theochem)*, **2002**, *581*, 11-15.

[41] González, M.P.; Toropov, A.A.; Duchowicz, P.R.; Castro, E.A. QSPR calculation of normal boiling points of organic molecules based on the use of correlation weighting of atomic orbitals with extended connectivity of zero- and first-order graphs of atomic orbitals. *Molecules*, **2004**, *9*, 1019-1033.

[42] Toropov, A.A.; Voropaeva, N.L.; Ruban, I.N.; Rashidova, S.Sh. Quantitative structure-property relationships for binary polymer-solvent systems: Correlation weighing of the local invariants of molecular graphs. *Polymer Science - Series A*, **1999**, *41*, 975-985.

[43] Marino, D.J.G.; Castro, E.A.; Toropov, A. Improved QSAR modelling of anti-HIV-1 acivities by means of the optimized correlation weights of local graph invariants. *Cent. Eur. J. Chem.*, **2006**, *4*, 135-148.

[44] Toropov, A.A.; Duchowicz, P.; Castro, E.A. Structure-toxicity relationships for aliphatic compounds based on correlation weighting of local graph invariants. *Inter. J. Mol. Sci.*, **2003**, *4*, 272-283.

[45] Toropov, A.A.; Schultz, T.W. Prediction of aquatic toxicity: Use of optimization of correlation weights of local graph invariants. *J. Chem. Inf. Comput. Sci*., **2003**, *43*, 560-567.

[46] Toropov, A.A.; Toropova, A.P. QSAR modelling of toxicity on optimization of correlation weights of Morgan extended connectivity. *J. Mol. Struct. (Theochem)*, **2002**, *578*, 129-134.

[47] Toropov, A.A.; Benfenati, E. Correlation weighting of valence shells in QSAR analysis of toxicity. Bioorg. Med. Chem., **2006**, *14*, 3923-3928.

[48] Ivanciuc, O.; Ivanciuc, T.; Balaban, A.T.; The complementary distance matrix, a new molecular graph metric. *ACH-Model. Chem*., **2000**, *137*, 57-82.

[49] González, O.G.; Murray, J.S.; Peralta-Inga, Z.; Politzer, P. Computed molecular surface electrostatic potentials of two groups of riverse transcriptase inhibitors: Relationships to anti- HIV-1 activities. *Int. J. Quantum Chem.*, **2001**, *83*, 115-124.

[50] Gute, B.D.; Basak, S.C. Predicting acute toxicity (LC50) of benzene derivatives using theoretical molecular descriptors: A hierarchical QSAR approach. *SAR QSAR Environ. Res.* **1997**, *7*, 117-131.

[51] Marino, D.J.G.; Peruzzo, P.J.; Castro, E.A.; Toropov, A.A. QSAR carcinogenic study of methylated polycyclic aromatic hydrocarbons based on topological descriptors derived from distance matrices and correlation weights of local graph invariants. *Internet Electron. J. Mol. Des*., **2002**, *1*, 115-133.

[52] Peruzzo, P.J.; Marino, D.J.G.; Castro, E.A.; Toropov, A.A. QSPR modelling of lipophilicity by means of correlation weights of local graph invariants. *Internet Electron. J. Mol. Des*., **2003**, *2*, 334-347.

[53] Toropov, A.A.; Toropova, A.P. QSAR modelling of mutagenicity based on graphs of atomic orbitals. *Internet Electron. J. Mol. Des.*, **2002**, *1*, 108-114.

[54] Bonchev, D. My life-long journey in mathematical chemistry. *Internet Electron. J. Mol. Des.*, **2005**, *4*, 434-490.

[55] Costescu, A.; Diudea, M.V. QSTR study on aquatic toxicity against *Poecilia reticulata* and *Tetrahymena pyriformis* using topological indices. *Internet Electron. J. Mol. Des.*, **2006**, *5*, 116-134.

[56] González, M.P.; Helguera, A.M.; Rodríguez, Y.M. TOPS-MODE and DRAGON descriptors in QSAR. 1. Skin permeation. *Internet Electron. J. Mol. Des.*, **2004**, *3*, 750-758.

[57] Hosoya, H. The topological index Z before and after 1971. *Internet Electron. J. Mol. Des.*, **2002**, *1*, 428-442.

[58] Kier, L.B. My journey through structure: The structure of my journey. *Internet Electron. J. Mol. Des.*, **2006**, *5*, 181-191.

[59] Lukovits, I.; Miličević, A.; Nikolić, S.; Trinajstić, N. On walk counts and complexity of general graphs. *Internet Electron. J. Mol. Des.*, **2002**, *1*, 388-400.

[60] Trinajstić, N.A life in science. *Internet Electron. J. Mol. Des.*, **2003**, *2*, 413-434.

[61] García-Domenech, R.; Catalá-Gregori, A.; Calabuig, C.; Antón-Fos, G.; del Castillo, L.; Gálvez, J. Predicting antifungal activity: A computational screening using topological descriptors. *Internet Electron. J. Mol. Des.*, **2002**, *1*, 339-350.

Received: March 10, 2011 **Revised: October 26, 2012** **Accepted: April 27, 2013**