communications in mathematical
match
and in computer chemistry

no. 43, March 2001

ISSN 0340-6253

MATCDY (43) 135-152 (2001)

On the Characterization of Molecular Stereostructure: 2. The Invariants of the Two-dimensional Graphs [†]

Giorgi Lekishvili
Laboratory of Modeling and Synthesis of Bioactive Compounds,
Faculty of Chemistry, Javakhishvili State University of Tbilisi, 3, Ilia Chavchavadze avenue, GE-380028 Tbilisi, Georgia

The present paper is an attempt of a systematic mathematical study of the extent of reflecting the 3D structural information by simple graph. The concept of two-dimensional graphs has been introduced. The usefulness in ascertaining the relative contributions of different structural phenomena was shown. The real and imaginary parts of the invariants of the two-dimensional graphs appeared to be mutually orthogonal.

The concepts of multi-dimensional graphs and universal algebraic models of molecules have been discussed.

[†] The present paper is dedicated to Prof. Randić, outstanding scientist and generous person, for his help in improving the presentation of the first paper of this series.

Introduction

As is well known, the precision of the QSAR/QSPR studies essentially depends on the accuracy of characterizing the molecular structure[1-5]. The perception of the latter occurs via at least the four steps: 1) The number and the types of the atoms composing the molecule; 2) The connectivity of the atoms and the bond types; 3) The localization of the atoms in space around a chiral atom, a double bond or a rigid ring, and 4) The localization of the atoms in space around a single bond. The constitution of molecules (composition and connectivity) can be well described by means of graph theory. The invariants of the weighted graphs are frequently applied in the QSAR/QSPR models despite the fact that they are unable to discriminate among the different configurations and conformations at identical constitution and connectivity. To overcome this inherent limitation of simple graph theory numerous molecular descriptors have been suggested[6-16]. It is noteworthy that even the application of highly discriminative indices does not always lead to a cardinal improvement of the model. To ascertain the reason of this is important not only from the theoretical, but also from the practical standpoint, as the calculation of most 3D descriptors requires much more efforts than of their simple analogues, at the conditions of relatively low extent of the model improvement.

Many publications were dedicated to the problem of the degree of conserving the information on 3D structure by applying simpler mathematical objects. As emphasized by Robinson et al.[17] "That exercise showed that most structural features were retained in the two-dimensional picture including, to a surprising level of accuracy, the three-dimensional distance information".

Our present task is a systematic mathematical study of this problem. We shall consider the approach enabling one to separate the information on the molecular structure at least into the two parts: the information on the molecular constitution and that on the spatial features of the molecular structure. In other words, one needs to apply such a set of indices, that some of them would characterize the molecular constitution only, being insensitive to the geometry variations of the molecules (so far many indices of this kind have been introduced). The second part, in contrary, would be zero for the molecules without chiral atoms and cis-trans isomerism, but be able to describe the latter. Then, we would ascertain the relative shares of the molecular constitution vs. the molecular configuration by employing the techniques of orthogonalization.

The present paper deals with an attempt to design such descriptors.

Orthogonal Descriptors

Multiple regression analysis is widely used in molecular modeling. Here a molecule is presented with a vector, the entries of which are independent topological indices. The latter are usually based on identical sources of invariants, such as the adjacency and the distance matrices. Therefore, they reveal high interrelation. Thus, the duplication of information occurs. Lukovits[18], Randić and Trinajstić have developed the concepts of the orthogonalized descriptors enabling one to overcome the above-mentioned drawback. Randić elaborated conceptually the approach[19-23] of obtaining qualitatively new information ("... give us insight, not numbers!"- Coulson). Trinajstić, Lučić and co-authors presented the unique possibilities[24-26] of optimizing the QSAR/QSPR models using orthogonalized descriptors.

The technique of orthogonalization leads to stable regressions, enabling one to ascertain the relative contribution of a given descriptor and to decrease the volume of calculations by removing the insignificant indices. All this, finally, results in the best possible, optimized models via applying relatively modern statistical techniques, see appendix.

One can find more details on the orthogonalized descriptors in refs 23, 25.

Topological Representation of Cis - Trans Isomerism

It is well known that the simple graphs are unable to reflect the spatial features of molecular structure, as different configuration/conformation molecules having the same composition and connectivity are presented with the isomorphic graphs. Therefore they have identical invariants - topological indices. In order to overcome this inherent limitation of the classic graph theory, which is devoid of utilizing the information on the 3D structure, we constructed a new set of the pairs of vertices denoted by Q. The Q set consists of the pairs characterized with the following peculiarities: they must be unbonded, connected with at least one path of the length three (d=3), basic subgraph (BSG), which may be not the shortest path, provided none of the vertices of the given pair is collinear with the central edge of this path. A pair of the vertices can be connected with more than one BSG[27].

We carried out the modification (reformulation) of the graph isomorphism and invariant definitions, as well as of the adjacency and of the distance matrices[27-29]. The elements of these matrices, complex numbers, can be calculated as follows:

$$a_{kl} = \alpha_{kl} + i\beta_{kl}$$

Where

$$\alpha_{kl}=\begin{cases}\mu_{kl},\ \text{if}\ \{k,l\}\in E,\text{i.e., if the } k^{th} \text{ and } l^{th} \text{ vertices are connected;}\\ 0,\ \text{otherwise.}\end{cases}$$

μ_{kl} is the edge weight, and

$$d_{kl}^{(i)}=d_{kl}+i\beta_{kl};$$

Here d is the path length, and β is a characteristic value of the pairs connected by a BSG. In case of double bonds as the central edges we can define it as follows:

$$\beta_{kl}=\begin{cases}-1, & \textit{if}\ d_{kl}=3 \text{ and } \quad k^{th} \text{ and } l^{th} \quad \text{vertices}\\ \text{are} & \text{on the different sides toward their central edge;}\\ 1, & \textit{if}\ d_{kl}=3 \text{ and } \quad k^{th} \text{ and } l^{th} \quad \text{vertices}\\ \text{are} & \text{on the same side toward their central edge;}\\ 0, & \text{otherwise.}\end{cases}$$

We demonstrated that one can obtain better correlations via applying the invariants of the complex adjacency matrix, e.g., a Randić type index[30], than the descriptors based on the simple one, e.g., the Kier-Hall index[31].

The model can function in the two following manners: first, one can apply only the BSGs with the double central bond, as we did in the first paper of the series for modeling standard entropies of alkenes. The second way of characterizing the geometry variations of molecular structures implies also the use of the BSGs with the single central bond. It enables one to consider also rotamers. We demonstrated the principal possibility of discrimination among different conformers by applying our approach in the preceding paper, however a genuine characterization of the molecular conformation requires some additional efforts, as the both terminal vertices of a basic subgraph are embedded in the same plane with the central edge in case when the central bond is of order two (see ref 32). The problem will be discussed in great detail in a next paper of the series [33,34].

The consideration of the BSGs only with the double central bond is convenient also because of another reason: mostly cis-trans isomers have different conformations. Therefore, one has to define which conformer should be taken as the representation of the given geometrical isomer. One has to be well aware of the fact that the longest conformations are not necessarily of the minimal energy, even in case of saturated hydrocarbons. We suggested a method to overcome this drawback in our dissertation [29], but the approach is computationally extremely intensive and not applicable in praxis. Thus, we shall still use only the BSGs with the double central bond formally considering the BSGs with the single central bond to be linear.

Below we shall employ the following complex topological indices:

$$W^{(i)} = \frac{1}{2}\sum_{k}^{n}\sum_{l}^{n} d_{kl}^{(i)}; \qquad d_{kl}^{(i)} = d_{kl} + i\beta_{kl},$$

And

$$\chi^{(i)} = \sum_{bonds}(\delta_k\delta_l)^{-1/2}; \qquad \delta_k = \sum_{l} a_{kl}\,.$$

Wiener[35] and Randić type invariants, respectively. The Kier-Hall index is a Randić-type invariant, which allows for consideration heteroatoms and bond multiplicity. The vertex degree is calculated[31] as follows:

$$\partial_k^V = Z_k^V - h_k$$

Here, $Z^V{}_k$ is the number of valence electrons of the kth atom and h_k is the number of the hydrogen atoms bonded with the kth atom.

Results and Discussion

A close inspection of the $\chi^{(i)}$ index reveals that its real part is insensitive (see Table, the dataset of alkenes is the same as in ref 27) to the geometry variations of molecular structure (cis-trans isomerism, in this case). It changes in the same way as the Kier-Hall χ^V index having similar numerical values each time when the molecule does not possess cis-trans isomerism. Moreover, $\mathrm{Re}\{\chi^{(i)}\}$ correlates well with χ^V:

$\mathrm{Re}\{\chi^{(i)}\}=0.9897\chi^V-0.0464;$

$r^2=0.9643, \quad S=0.1035, \quad F=620.9.$

On the contrary, the imaginary part of the index is always zero at the absence of the cis-trans isomerism describing only the spatial features of the molecules. Therefore, one can conclude that $\mathrm{Re}\{\chi^{(i)}\}$ presents the molecular constitution only, while $\mathrm{Im}\{\chi^{(i)}\}$ is solely responsible for the information on 3D structure.

Table

The Topological Indices and Some Physical Properties of the Set of 25 Alkenes

#	Compound	bp, *)	S^0_{298} **)	$W^{(i)}$	$\chi^{(i)}$	χ^V
1	ethene	<0	55.45	1	0.5000	0.5000
2	propene	<0	63.80	4	0.9856	0.9856
3	butene	<0	73.04	10	1.5236	1.5236
4	*cis*-butene	0.88	71.90	10+i	1.2304-0.3718i	1.4881
5	*trans*-butene	3.72	70.86	10-i	1.2304+0.3718i	1.4881
6	2-me-propene	<0	70.17	9	1.3536	1.3536
7	pentene	29.9	82.65	20	2.0235	2.0235
8	*cis*-pent-2-ene	36.9	82.76	20+i	1.8085-0.4282i	2.026
9	*trans*-pent-2-ene	36.4	81.36	20-i	1.8085+0.4282i	2.026
10	2-Me-butene	31.2	81.73	18	1.9143	1.9143
11	2-Me-but-2-ene***)	38.5	80.92	18	1.8661	1.8661
12	3-Me-butene	20.1	79.70	18	1.8963	1.8963
13	hexene	63.3	91.93	35	2.5235	2.5235
14	*cis*-hex-3-ene	66.4	90.73	35+i	2.3867-0.4848i	2.5639
15	*trans*-hex-3-ene	67.1	89.59	35-i	2.3867+0.4848i	2.5639
16	2-Me-pentene	60.7	91.34	32	2.4143	2.4143
17	3-Me-pentene	54.1	90.06	31	2.4342	2.4342
18	4-Me-pentene	53.9	87.89	32	2.3794	2.3794
19	2-Me-pent-2-ene***)	67.3	90.45	32	2.4040	2.4040
20	*cis*-4-Me-pent-2-ene	56.3	89.23	32+i	2.2125-0.3140i	2.3988
21	*trans*-4-Me-pent-2-ene	58.6	88.02	32-i	2.2125+0.3140i	2.3988
22	2-Et-butene	64.7	90.01	31	2.4750	2.4750
23	2,3-diMe-butene	55.7	87.39	29	2.2971	2.2971
24	3,3-diMe-butene	41.2	82.16	28	2.1969	2.1969
25	2,3-diMe-but-2-ene***)	73.2	86.67	29	2.2500	2.2500

*) Taken from ref 11; we used only 21 compounds (with positive values of bp-s) in modeling bp-s.

**) Taken from ref 35a.

***) See ref 32.

In order to ascertain the relative shares of each structural phenomena (constitution vs. configuration) in the selected physical property (here: S^0_{298} [36a]) we constructed the following model:

S^0_{298}=aA+bB+c,

Where A=Re$\{\chi^{(i)}\}$ and B=Im$\{\chi^{(i)}\}$,

a=17.5246; b=-1.4642; c=47.8449;

r^2=0.9710; S=1.6979; F= 367.9

Next we started to orthogonalize B against A to find out the extent of content of the information on 3D structure in A and that on connectivity in B. The result was quite unexpected: B has turned to be already orthogonal to A, i.e., r(B,A)=0.0000! Therefore, we obtained the stable models without additional efforts:

S^0_{298}=aA+ c,

a=17.5246; c=47.8449;

r^2=0.9697; S=1.6969; F=735.69.

and

S^0_{298}= bB+c_1,

b=-1.4642; c_1=81.9924;

r^2=0.0013; S=9.7399; F= 0.030.

As one can see, the introduction of the information on the 3D geometry variations of the structures has not led us to a sufficient improvement of the model, in contrary, the S value along with the F ratio decreased.

The similar results were observed when using the correlation of A and B with the boiling points of alkenes[36b].

bp=aA+bB+c,

a=51.3617; b=1.4525; c=-61.8016;

r^2=0.8499; S=8.4817; F=50.94.

bp=aA+c,

a=51.3617; c=-61.8016;

r^2=0.8495; S=8.2649; F=107.27.

bp= bB+c_1,

b=1.4525; c_1=46.6715;

r^2=0.00032; S=21.3016; F=0.0061.

Our next experiment[37] was aimed at obtaining an empirical equation describing the statistical dependence of S^0_{298} on $W^{(i)}$. Below these equations are given:

S^0_{298}=aW+bB+c,

a=0.8634; b=-0.5988; c=62.2712;

r^2=0.9266; S=2.7001; F=138.83;

S^0_{298}=aW+ c,

a=0.8634; c=62,2712;

r^2=0.9253; S=2.6643; F=284.8;

and

S^0_{298}= bB+c_1,

b=0.5988; c_1=81.9924;

r^2=0.0013; S=9.7399; F= 0.030.

Also here W index is orthogonal to B.

One has to emphasize that the share of molecular constitution did overweight that of configuration in each of these cases. Thus, this formalism could be useful in searching for leading substructures. This corollary also demonstrates that the role of the 2D indices in the QSPR researches did not vanish.

As is evident, we have solved the task of our paper: a systematic mathematical study of the extent of reflecting the 3D structural information by simple graph. Nevertheless, the obtained results enable us to explore some other benefits from our efforts.

A careful consideration of the results reveals that the real part of the $W^{(i)}$ index is the classical Wiener number[35], an invariant of the simple graph {V, E}, where the set Q does not take part. Thus, as was expected, the invariant of the simple graph characterized the molecular connectivity. Analogically, the imaginary part of the $W^{(i)}$ index is an invariant of the {V, Q} object. It stores the information on the spatial features of molecular structure ('localization'). As one can see, something akin of 'the information decomposition' effect occurs, i.e., we can evaluate the contributions of connectivity and of localization separately. It is important either from practical and theoretical standpoint. In order to get acquainted with the nature of the {V, Q} object, we need to recall that the elements of the set Q present the pairs of the elements of the set V, which (the vertices making the pairs) are in the binary relation to each another. Therefore, the {V, Q} object satisfies the definition of graph. Indeed, a genuine analogy is observed. The set E contains the pairs of the vertices being in the binary relation[38] of binding. To discriminate among the different bond types and heterobonds, the weights may be employed:

$$\varphi:E\to W(E)$$

Here E is the set of edges, φ is the surjective mapping of the elements of E onto weight set W(E).

In the same manner, one may consider the set Q as the set of edges, with the difference that the elements of Q are in the binary relation[38] of localization to each another. The pair of vertices v_i and v_j is said to belong to the set Q if at least one path d(i, j) exists such that d(i, j)=3, and none of the vertices v_i and v_j is bonded and collinear to the central edge of this path[27]. Obviously, we have to employ a set of weights also here to discriminate among different alignment of the terminal vertices relatively to the central edge:

$$\psi:Q\rightarrow W(Q)$$

In our experiments, the set W(Q) contains two elements –1 and 1, but in case of the single bond as the central edge one needs to use a more sophisticated system of weights[33]. Also here, ψ is the surjective mapping of the elements of Q onto the weight set W(Q).

Of course, we can construct a simple graph {V, E+Q} and handle the objects {V, E} and {V, Q} as its subgraphs (note that $E\cap Q=\varnothing$), but it looks like 'a violation', while the sets E and Q model the molecular-structural phenomena of completely different physical-chemical nature. We shall suggest a different approach. As it will be shown below as well as in other publications,[33, 34, 39, 41] this approach is a convenient starting point to create a universal algebraic model of molecule and retains the information decomposition effect.

Let us create the two-dimensional graph, as an abstract algebraic object $G=\{V, E_1, E_2\}$, where V is the set of vertices and E_1 and E_2 are the sets of edges constructed via two different binary relations[38,42] (the latter are defined on the set V). At molecular modeling, the set E_1 is given as the relation of binding, and the set $E_2=Q$ - via the spatial location of atoms ($q=\{v_iv_j\}\in E_2$, if $d_{ij}=3$ and v_i, v_j are neither bonded nor collinear to the central edge of this path). Therefore, an invariant of $\{V, E_1\}$ becomes a characteristic of the constitution, while that of $\{V, E_2\}$ presents geometrical variations of molecular structure.

Let us consider another experiment. The first descriptor is the Kier-Hall index, the best Randić-type invariant for the weighted graphs. The second one is the same halfsum of the adjacency matrix elements of the $\{V, E_2=Q\}$ graph, hereafter denoted as T. Below the statistical equations are listed:

$\mathbf{S^0_{298}}=a\chi^V+bT+c$,

a=17.7917; b=-0.5988; c=46.1308;

r^2=0.9852; S=1.21; F=730.81;

$S^0_{298}=a\chi^V+c$,

a=17.7917; c=46.1307;

r^2=0.9839; S=1.24; F=1401.9;

and

$S^0_{298}=bT+c_1$,

b=0.5988; c_1=81.9924;

r^2=0.0013; S=9.7399; F= 0.030;

$r(\chi^V,B)=0.0000$.

We see that the information decomposition still has place.

A question arises: may we consider the invariants of the (sub) graphs {V, E_1} and {V, E_2} strictly orthogonal?

We would like to answer in the following manner: firstly, the orthogonality of these descriptors is a statistical fact, as well as actually all of the QSPR/QSAR relationships are statistical, and as such, (the orthogonality) depends on the quality of the training dataset. However, it is not a disadvantage, as the quality of the datasets is always crucial in the QSPR/QSAR studies. We shall once again emphasize, that the task of this study is to design not simply 3D indices, but the descriptors, which reflect the information on the molecular connectivity and on configuration separately. Nevertheless, the orthogonality of these descriptors, even statistical (once, no other-type-relationships are available in QSPR/QSAR), has obvious theoretical importance. Note, that we have applied the same dataset, as three years ago, therefore, it was not prepared specially. Next, the T index is designed in the way that it differs from zero only at the presence of cis-trans-isomerism. More important is that the traditional adjacency matrix does not take part in constructing the T number, if we may neglect the fact that the distance matrix is required to create the set E_2=Q. The latter matrix can be reduced to the adjacency matrix. Much more significant is that none of the pairs of vertices can belong to the sets E_1 and E_2 at the same time as it follows from our method of constructing the set E_2.

In order to test the orthogonality of these descriptors in a more profound manner, we carried out another experiment. The 3-, 4- and 5-vertex chain configurations were studied (the same as on Fig. 7 in ref 5). The structures and the corresponding values of the W and T numbers are given on Figure. As was expected, these descriptors are orthogonal, r(W, T)=0!

That means that one has to consider our task to create a 3D model of molecule with the separated information on the constitution and the configuration to be solved.

Concluding Remarks

Thus, the two-dimensional graphs present the sources of the molecular descriptors of the principally new type. The entries of these descriptors, i.e., their real and imaginary parts reflect separately the molecular connectivity and the features of stereostructure, respectively. Such graphs enable one to design sets of linearly independent descriptors. Apart from the theoretical importance, our method is practically valuable. As mentioned above, many highly discriminative 3D descriptors fail to achieve any significant improvement of the QSPR quality in comparison with their simple analogues. Sometimes the former are even inferior to the latter[22]. That is why one may need a method to foresee the extent of improving the QSPR quality by substituting simple (basic 2D) indices with their 3D analogues. Obviously, an improvement of the correlations (if any) is conditioned by the contribution of the structural spatial features relatively to the molecular constitution. We suggest a method that works as follows: subtract a smaller training dataset from the real training one, keep it balanced (the methods of making high quality datasets are considered in any textbook of chemometrics), calculate the basic 2D index and the T number. If they are already orthogonal, one has the desired evaluation of the contribution of the structural spatial features to the given physical property/biological activity. If these indices are not orthogonal, one needs to find the dominant descriptor and carry out the orthogonalization by the standard techniques[23,25]. This will lead one to the desired evaluation. E.g., considering the results of our study, one may conclude that a substitution of the Kier-Hall index with its 3D analog cannot lead us to a sufficient improvement of the QSPR correlations in the alkenes series. Indeed, Estrada[11] gives a correlation **bp**~Ω, Ω is an excellent 3D analog of χ^V, here $\Delta r=0.0018$, $\Delta S=-0.31$, $\Delta F=165$, n=53.

Taking into account the above-mentioned one can make another important conclusion. We can consider (di)graphs as particular cases of universal algebras (V, Ω) [42], when the set (system) Ω of the latter consists of one binary relation only. This is of certain importance for computational and for mathematical chemistry indicating a new way to modernize and to generalize chemical graph theory. Indeed, the newly introduced two-dimensional graph should be considered as another particular case of universal algebras with the set Ω containing the two binary relations. The first serves for presenting connectivity while the second for configuration/conformation.

In the future, the two-dimensional graphs will be generalized further to N-dimensional (N>2) graphs via describing new binary relations, e.g., for hydrogen bonds, molecule-molecule alignment, etc. Moreover, we are able to define ternary (N=3), quaternary (N=4) relations presenting molecular topological forms[4, 5, 43], 5-ary relation for modeling tetrahedral chirality[39, 40], 11-ary relation for describing metalocenes[41] and so on. We can quantify N-ary (N>2) relations applying the techniques of hypergraphs[44] being another particular case of universal algebras. The use of universal algebras as the universal algebraic models of molecules has two important advantages: first, we can employ the universal algebras in the form required by the given task, i.e., constructing the set Ω in the convenient form to avoid the use of insignificant structural phenomena. In other words, we can optimize molecular models. The second positive feature is that universal algebras seem to be able to produce linearly orthogonal sets of descriptors.

Therefore, a further work on multidimensional (N-dimensional) graphs and universal algebraic models is of practical and theoretical importance.

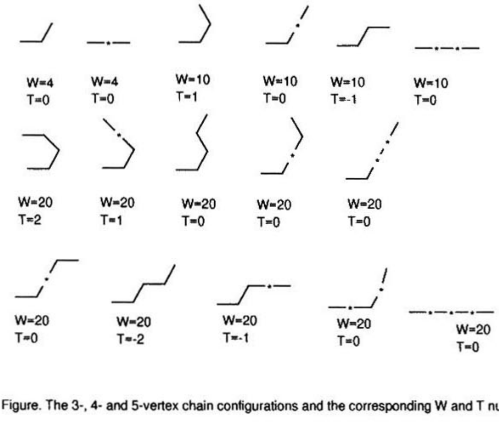

Figure. The 3-, 4- and 5-vertex chain configurations and the corresponding W and T numbers

Acknowledgement

The author would like to thank Prof. Trinajstić, Zagreb, Croatia, for reading the manuscript, for the helpful comments and for the constant support. The amendments and suggestions of the respected referee of the paper, Prof. Gutman, Kraguevac, Yugoslavia, is gratefully acknowledged.

Appendix

We have not used the following equation in our study:

$$S = \sqrt{\frac{(Y-\hat{Y})^2}{M}} \qquad (1)$$

Here Y is experimental physical property (biological activity), $\hat{Y}$ is the one estimated with the model, M is the number of objects (observations) in the dataset.

It is evident, that to add a new descriptor causes the S to decrease, independent on the goodness of the descriptor. Therefore, S of the equation (1) cannot preserve the model from overfitting.

Instead of eq 1, we applied the following expression:

$$S = \sqrt{\frac{(Y-\hat{Y})^2}{M-I-1}} \qquad (2)$$

Here Y, $\hat{Y}$ and M are the same as in eq 1, and 'I' is the number of descriptors (predictors) in the model. Now, if we add a new descriptor to the model, there's no guarantee that S decreases, as both the denominator and numerator are diminished. Thus, adding only a good descriptor can improve the model's quality (evaluated via the standard error of estimate). It is clear, that S of eq 2 is much more useful in QSAR/QSPR studies. Furthermore, many standard statistical packages use eq 2 to calculate S, e.g., 'Datafit Engineering' Shareware (http://www.oakdate.com), which was applied in our experiments.

Note, that eq 2 is useless, when I-1>=M.

Lučić and Trinajstić often use another equation to get the S value:

$$S = Std(Y)\sqrt{\frac{M(1-R^2)}{M-I-1}} \qquad (3)$$

Here Std is standard deviation. R is correlation coefficient. The efficiency of eq 3 increases if it is used with mean-centered and scaled data, as Std(Y) becomes unit and thus, vanishes. If the predictors (descriptors) are orthogonal,

$$R^2 = \sum_i R_i^2 \qquad (4)$$

(Check eq 4 for our studies!) Now eq 3 can be rewritten in a very convenient form:

$$S^2 = \frac{M(1-\sum_i R_i^2)}{M-I-1} \qquad (5)$$

Only introduction of a good (relevant) descriptor can decrease S of eq5. The calculations using eq 2 and eq 3 produce similar results for our experiments.

References and Notes

1. Trinajstić, N. Chemical Graph Theory, 2nd revised ed.; CRC; Boca-Raton, FL, 1992.
2. Stankevich, M. I.; Stankevich, I. V.; Zefirov, N. S. Topological Indexes in Organic Chemistry, Russ. Chem. Rev. 1988, 57, 191-208.
3. Balaban, A. T.; From Chemical Topology to 3D Molecular Modeling. In: From Chemical Topology to Three Dimensional Geometry; Balaban, A. T.; Ed.; Plenum Press: New York and London, 1996.
4. Tratch, S. S.; Zefirov, N. S. Combinatorial Models and Algorithms in Chemistry. Ladder of Combinatorial Objects and its Application to Formalization of Structural Problems of

Organic Chemistry. In: Principles of Symmetry and Systemology in Chemistry; Stepanov, N. F., Ed.; Moscow State University Press: Moscow, 1987, pp. 54-86 (Russian).

5. Tratch. S. S.; Devdariani, R. O.; Zefirov, N. S. Combinatorial Models and Algorithms in Chemistry. Configuration-Topological Analogs of Wiener Index. Zh. Organ. Khim. 1990, 26, 921-932. (Russian).
6. Whitley, D. C. Van der Waals Surface Graphs and Molecular Shape. J. Math. Chem. 1998, 23, 377-397.
7. Randić, M.; Jerman-Blazić, B.; Trinajstić, N. Development of 3 Dimensional Molecular Descriptors. Comput. Chem. 1990, 14, 237-246.
8. Bogdanov, B. On the Three Dimensional Wiener Number. J. Math. Chem. 1989, 3, 299-302.
9. Pogliani, L. On a Graph Theoretical Characterization of Cis/Trans Isomers. J. Chem. Inf. Comput. Sci. 1994, 34, 801-804.
10. Diudea, M. V.; Horvath, D.; Graovac, A. Molecular Topology. 15. 3D Distance Matrices and Related Topological Indices. J. Chem. Inf. Comput. Sci. 1995, 35, 708-713.
11. Estrada, E. Three-Dimensional Molecular Descriptors Based on Charge Density Weighted Graphs. J. Chem. Inf. Comput. Sci. 1995,35, 708-713.
12. Nikolić, S.; Trinajstić, N.; Mihalić, Z.; Carter, S. On the Geometric Distance Matrix and the Corresponding Structural Invariants of Molecular Systems. Chem. Phys. Lett. 1991, 176, 21-28.
13. Schultz, H. P.; Schultz, E. B.; Schultz, T. P. Topological Organic Chemistry. 9. Graph Theory and Molecular Topological Indices of Streoisomeric Organic Compounds. J. Chem. Inf. Comput. Sci. 1995, 35, 864-870.
14. Schuur, J. H.; Seltzer, P.; Gasteiger, J. The Coding of the Three-Dimensional Structure of Molecules by Molecular Transform and its Application to Structure-Spectra Correlations and Studies of Biological Activity. J. Chem. Inf. Comput. Sci. 1996, 36, 334-344.
15. Wagener, M.; Sadowski, J.; Gasteiger, J. Autocorrelation of Molecular Surface Properties for Modeling Corticosteroid Binding Globulin and Cytosolic Ah Receptors Activity by Neural Networks. J. Amer. Chem. Soc. 1995, 117, 7769-7775.
16. Bogdanov, B.; Nikolić, S.; Trinajstić, N. On the Three-Dimensional Wiener Number. A Comment. J. Math. Chem. 1990, 5, 305-306.

17. Robinson, D. D.; Barlow, T. M.; Richards, W. G. Reduced Dimensional Representation of Molecular Structure. J. Chem. Inf. Comput. Sci. 1997,37, 939-942.

18. Lukovits, I. Quantitative Structure-Activity Relationships Employing Independent Quantum Chemical Indexes. J. Med. Chem. 1983, 26, 1104-1109.

19. Randić, M. Orthogonal Molecular Descriptors. New J. Chem. 1991, 15, 517-525.

20. Randić, M. Fitting of Non Linear Regressions by Orthogonalized Power Series. J. Comput. Chem. 1993, 14, 363-370.

21. Randić, M. Curve-Fitting Paradox. Int. J. Quant. Chem.: Quant. Biol. Symp. 1994, 21, 215-225.

22. Randić, M. Quantitative Structure-Property Relationships. Boiling Points of Planar Benzenoids. New J. Chem. 1996, 20, 1001-1009.

23. Randić, M. On Characterization of Chemical Structure. J. Chem. Inf. Comput. Sci. 1997, 37, 670-687.

24. Amić, D.; Davidović-Amić, D.; Trinajstić, N. Calculations of Retention Times of Anthocyanins with Orthogonalized Molecular Descriptors. J. Chem. Inf. Comput. Sci. 1995, 35, 131-139.

25. Lučić, B.; Nikolić, S.; Trinajstic, N. The Structure-Property Models Can Be Improved Using the Orthogonalized Descriptors. J. Chem. Inf. Comput. Sci. 1995, 35,532-538.

26. Lučić, B.; Trinajstić, N. Multivariate Regression Outperforms Several Robust Architectures of Neural Networks in QSAR Modeling. J. Chem. Inf. Comput. Sci. 1999, 39, 121-132.

27. Lekishvili, G. On the Characterization of Molecular Stereostructure: 1. Cis-Trans Isomerism. J. Chem. Inf. Comput. Sci. 1997 37, 5, 924-928.

28. Lekishvili, G. Mathematical Investigation of Cis-Trans Isomeric Transformations of Alkenes. Proceed. Acad, Sci. Georg. 1994, 20, 1-4, 152-153. (Georgian).

29. Lekishvili, G. Optimization Based Non Linear Models of Organic and Elementorganic Molecules. Dissertation, Tbilisi, 1998 (Georgian).

30. Randić, M. On Characterization of Molecular Branching. J. Amer. Chem. Soc. 1975, 69, 6609-6615.

31. Kier, L. B.; Hall, L. H. Molecular Connectivity in Chemistry and Drug Design Research. Academic Press, NY, 1976.

32. The compounds #11 (2-Me-but-2-ene), #19 (2-Me-pent-2-ene) and #25 (2,3-diMe-but-2-ene) are visible outliers within the framework of this dataset. That's why we considered neither the BSGs of this type in our studies, including the first paper of this series, as it is evident from the Table 2 in ref 27. It is interesting to note that even if one considers such BSGs in the model, the resultant indices $\chi^{(i)}$ and $W^{(i)}$ will not have the imaginary parts. Readers can check it easily, or visit the web-site http://lekishvili.science.directnic.com/fig2.htm , for an example for the compound #11, in case if one considers such BSGs.

Note, that such flexibility is an important advantage, as we are searching for statistical relationships, not for analytical functions. As one can see, our method enables for handling the outliers without their removal from the dataset. It increases the robustness of the regressions. We believe that the optimization-based descriptors belong to a new generation of topological indices.

The Figure, which traverses the algorithm of computing the $\chi^{(i)}$ and $W^{(i)}$ indices can be found at http://lekishvili.science.directnic.com/fig1.htm , or requested by mail: gleki@gol.ge .

33. Lekishvili, G. On the Characterization of Molecular Stereostructure: 3. Rotamers (submitted).

34. Lekishvili, G. On the Characterization of Molecular Stereostructure: 7. Coding 3D Structures (work in progress).

35. Wiener, H. J. Structural determination of Paraffin Boiling Points. J. Amer. Chem. Soc. 1947, 69, 17-20.

36. a) The experimental values of $S°_{298}$ are taken from Kireev, V. A. Methods of Practical Calculations in Thermodynamics of Chemical Reactions. Khimia, Moscow, 1975, pp. 462-463 (Russian); b) the experimental values of bp are taken from ref 11.

37. We would like to emphasize here, that getting the best correlations was outside the scope of the given study.

38. Berge, C. Théorie des Graphes et ses Applications. Dunod, Paris, 1958 (French).

39. Lekishvili, G. On the Characterization of Molecular Stereostructure: 4. An Algebraic Criterion and a Quantitative Measure of the Chirality of the Environment of Tetrahedral Atoms (to be submitted).

40. Ihlenfeldt, W. -D., Gasteiger, J. Augmenting Connectivity Information by Compound Name Parsing: Automatic Assignment of Stereochemistry and Isotope Labeling. J. Chem. Inf. Comput. Sci. 1995, 35, 663-674.

41. Lekishvili, G. On the Characterization of Molecular Stereostructure: 5. Metalocenes (work in progress).

42. Kurosh, A. G. Lectures in General Algebra. PhysMatGlz, Moscow, 1962 (Russian).

43. Zefirov, N. S.; Tratch, S. S. Some Notes on Randić-Razingers's Approach to Characterization of Molecular Shapes. J. Chem. Inf. Comput. Sci. 1997, 37, 900-912.

44. Konstantinova, E. V.; Skorobogatov, V. A. Molecular Hypergraphs: The New Representation of Nonclassical Molecular Structures with Polycentric and Delocalized Bonds. J. Chem. Inf. Comput. Sci. 1995, 35, 472-478.