

Internet Electronic Journal of Molecular Design

February 2003, Volume 2, Number 2, Pages 70–85

Editor: Ovidiu Ivanciuc

Special issue dedicated to Professor Haruo Hosoya on the occasion of the 65th birthday
Part 6

Guest Editor: Jun–ichi Aihara

Molecular Design of Chemical Compounds with Prescribed Properties from QSAR Models Containing the Hosoya Index

Mariya I. Skvortsova, Konstantin S. Fedyaev, Vladimir A. Palyulin, and Nikolai S.
Zefirov

Department of Chemistry, Moscow State University, Moscow 119992 Russia

Received: October 4, 2002; Revised: December 6, 2002; Accepted: December 21, 2002; Published: February 28, 2003

Citation of the article:

M. I. Skvortsova, K. S. Fedyaev, V. A. Palyulin, and N. S. Zefirov, Molecular Design of Chemical Compounds with Prescribed Properties from QSAR Models Containing the Hosoya Index, *Internet Electron. J. Mol. Des.* **2003**, *2*, 70–85, <http://www.biochempress.com>.

Molecular Design of Chemical Compounds with Prescribed Properties from QSAR Models Containing the Hosoya Index[#]

Mariya I. Skvortsova, Konstantin S. Fedyaev, Vladimir A. Palyulin,* and Nikolai S. Zefirov*

Department of Chemistry, Moscow State University, Moscow 119992 Russia

Received: October 4, 2002; Revised: December 6, 2002; Accepted: December 21, 2002; Published: February 28, 2003

Internet Electron. J. Mol. Des. 2003, 2 (2), 70–85

Abstract

A problem of computer generation of chemical structures with given property values is considered. It is supposed that there exists a structure–property correlation equation containing the Hosoya index and the numbers of different atom types. Such an equation allows us to reduce the aforementioned problem to the task of generation of chemical graphs containing definite structural fragments. An algorithm for solving this problem and examples of its application are considered.

Keywords. Quantitative structure–property relationships; chemical structure reconstruction; inverse problem in QSPR/QSAR; Hosoya index; topological indices.

1 INTRODUCTION

The studies of quantitative structure–property relationships (QSPR) can be considered as one of the important goals of modern theoretical/mathematical chemistry. The structure of molecules in QSPR studies is usually described in terms of topological, electronic, steric and other characteristics, and the property is correlated with these parameters for some training set of compounds. The main goal of constructing the correlating equations is to estimate the properties of the compounds not belonging to the training set in order to select structures for further synthesis of compounds with required properties.

The so-called inverse problem (IP) consists in the exhaustive generation of chemical structures of some class with a prescribed value of some property. The generation process is based on the preliminary constructed QSPR equation, which usually can be written as $y = f(x_1, \dots, x_N)$, where y is the property value, and x_1, \dots, x_N are some topological indices. The IP in this case can be reduced to

[#] Dedicated to Professor Haruo Hosoya on the occasion of the 65th birthday.

* Correspondence author; phone: 7–095–9393557; fax: 7–095–9390290; E-mail: vap@org.chem.msu.su (VAP) and zefirov@org.chem.msu.su (NSZ).

the procedure of generation of structures with definite values of parameters x_1, \dots, x_N which are determined from the correlation equation at the condition $y_1 \leq y \leq y_2$.

Earlier the IP for correlation equations containing only one particular topological index (Randić, Wiener, centric Balaban index, etc.) or several indices of a special type (kappa–indices of molecular shape, information topological indices) were considered [1–11]. However, other topological descriptors are also widely used in QSPR [12]. Thus, a development of algorithms for solving the IP for other indices and for multiple linear regression is of great importance in practical QSPR studies [13].

One of the well known molecular descriptors used in these studies is the Hosoya index Z defined as:

$$Z = \sum_{k=0}^{\lfloor n/2 \rfloor} p_k, \quad (1)$$

where p_k is the number of subgraphs consisting of k non–adjacent edges in a molecular graph G representing a chemical structure, $p_0 = 1$, n is the number of vertices in G [14,15]. The index Z can be also defined by the formula:

$$Z = i^{-n} M(i),$$

where $M(x) = \sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^k p_k x^{n-2k}$ is the matching (or acyclic) polynomial of a graph G , and i is the imaginary unit [16]. There are some extensions of index Z to weighted graphs [17–21]. The Hosoya index has been used in correlations of thermochemical properties of alkanes, bond orders and π –electron energies of conjugated hydrocarbons [22,23]. Also some linear correlations between boiling points and parameters Z and $P = \prod_i v_i$ (v_i is the degree of the i –th vertex in graph G) were found [12]. Note that for acyclic graphs the Hosoya index can be expressed via the coefficients of a characteristic polynomial of G [14].

In the present paper an algorithm for solving the IP for the case of correlation equation

$$y = f(Z, n, n_1, n_2, n_3, n_4) \quad (2)$$

where n_i is the number of vertices of degree i ($i = 1, 2, 3, 4$) in a graph is considered. It is assumed that a molecular graph G is a simple one (without multiple edges and labeled vertices) and has no vertices of degree exceeding four. It is supposed that the Hosoya index can be expressed from Eq. (2), that is:

$$Z = f_1(y, n, n_1, n_2, n_3, n_4) \quad (3)$$

for some function f_1 . Evidently, the index Z does not take into account heteroatoms and multiple bonds in molecules. However, the suggested algorithm may be used for such compounds if there

exists a sufficiently good correlation (2) for a particular case.

Some examples of solving the IP both for hydrocarbons and for molecules containing heteroatoms are given.

2 THEORETICAL RESULTS

Here we will present some graph-theoretical results concerning the numbers p_k of k -matching in a graph G ($k \geq 1$), obtained by us. Suppose that G is a simple connected graph with vertex degrees not exceeding four. It is known [24] that the number of edges p_1 can be found by the following formula:

$$p_1 = \frac{1}{2} \sum_{i=1}^4 i \cdot n_i. \quad (4)$$

Theorem 1.

$$p_2 = \frac{1}{2} p_1 + \frac{1}{2} p_1^2 - \frac{1}{2} \sum_{i=1}^4 i^2 \cdot n_i. \quad (5)$$

Proof. Let us select an arbitrary edge γ in graph G and consider a graph G_1 obtained from G by deleting the selected edge γ and all edges adjacent to it. Denote degrees of vertices incident to γ by $v_1(\gamma)$ and $v_2(\gamma)$. Then the number of edges in graph G_1 is equal to $p_1 - (v_1(\gamma) + v_2(\gamma) - 1)$. Denote by $p_{2|\gamma}$ the number of pairs of non-adjacent edges in G , one of which is γ . Then, obviously, $p_{2|\gamma}$ is equal to the number of edges in G_1 , that is $p_{2|\gamma} = p_1 - (v_1(\gamma) + v_2(\gamma) - 1)$. Summing these equalities over all possible edges γ , we get

$$\begin{aligned} \sum_{\gamma} p_{2|\gamma} &= \sum_{\gamma} (p_1 - (v_1(\gamma) + v_2(\gamma) - 1)) = \\ &= \sum_{\gamma} p_1 - \sum_{\gamma} (v_1(\gamma) + v_2(\gamma)) + \sum_{\gamma} 1 = p_1^2 - \sum_{\gamma} (v_1(\gamma) + v_2(\gamma)) + p_1. \end{aligned}$$

Each pair of non-adjacent edges is present in the sum $\sum_{\gamma} p_{2|\gamma}$ twice, since each edge of G may be considered both as γ and as the edge of G_1 . Hence $\sum_{\gamma} p_{2|\gamma} = 2p_2$ and

$$p_2 = \frac{1}{2} p_1^2 - \frac{1}{2} \sum_{\gamma} (v_1(\gamma) + v_2(\gamma)) + \frac{1}{2} p_1.$$

Transform the sum $\sum_{\gamma} (v_1(\gamma) + v_2(\gamma))$. Each vertex degree v_k is present in it as many times as many edges are incident to the k -th vertex, that is:

$$\sum_{\gamma} (v_1(\gamma) + v_2(\gamma)) = \sum_{k=1}^n v_k \cdot v_k = \sum_{i=1}^4 i^2 \cdot n_i.$$

Using the last formula, one can easily get the formula (5).

Theorem 2. For $2 \leq k < \lceil n/2 \rceil$ the following inequalities are true:

$$\frac{1}{k+1} p_k \left(p_1 - \sum_{i=1}^4 \alpha_i(k) \cdot i + k \right) \leq p_{k+1} \leq \frac{1}{k+1} p_k (p_1 - 2k), \quad (6)$$

where

$$\begin{aligned} \alpha_4(k) &= \min \{n_4, 2k\}, \\ \alpha_3(k) &= \min \{n_3, 2k - \alpha_4\}, \\ \alpha_2(k) &= \min \{n_2, 2k - \alpha_4 - \alpha_3\}, \\ \alpha_1(k) &= \min \{n_1, 2k - \alpha_4 - \alpha_3 - \alpha_2\}. \end{aligned}$$

Proof. Let us prove the left inequality in Eq. (6). Select and fix an arbitrary set $\gamma_1, \dots, \gamma_k$ of k non-adjacent edges of graph G . Denote by G_k a subgraph of G obtained by deleting from G the selected edges, all edges adjacent to it and vertices that became isolated. Denote by \bar{G}_k a subgraph of G consisting of deleted edges and vertices incident to it. Denote by q_k and \bar{q}_k the numbers of edges in G_k and \bar{G}_k , respectively. Then $q_k + \bar{q}_k = p_1$. Let us estimate \bar{q}_k . One can show that

$$\bar{q}_k \leq \sum_{i=1}^k (v_1(\gamma_i) + v_2(\gamma_i) - 1) = \sum_{i=1}^k (v_1(\gamma_i) + v_2(\gamma_i)) - k, \text{ where } v_1(\gamma_i) \text{ and } v_2(\gamma_i) \text{ are degrees of}$$

vertices incident to the edge γ_i . A maximum value of the sum $\sum_{i=1}^k (v_1(\gamma_i) + v_2(\gamma_i))$ is reached when

a maximum number of $2k$ vertices in it are of degree 4, a maximum number of remained vertices are of degree 3, etc. Note that the number of vertices of degree 4 in $\gamma_1, \dots, \gamma_k$ cannot be greater than n_4 nor $2k$, similarly, the number of vertices of degree 3 cannot be greater than n_3 nor the number of the remained vertices, etc. Introduce the numbers $\alpha_4(k) = \min \{n_4, 2k\}$, $\alpha_3(k) = \min \{n_3, 2k - \alpha_4\}$, $\alpha_2(k) = \min \{n_2, 2k - \alpha_4 - \alpha_3\}$, $\alpha_1(k) = \min \{n_1, 2k - \alpha_4 - \alpha_3 - \alpha_2\}$. Then:

$$\sum_{i=1}^k (v_1(\gamma_i) + v_2(\gamma_i)) \leq \sum_{i=1}^4 i \cdot \alpha_i(k).$$

Hence, $\bar{q}_k \leq \sum_{i=1}^4 i \cdot \alpha_i(k) - k$ and:

$$q_k \geq p_1 - \sum_{i=1}^4 i \cdot \alpha_i(k) + k.$$

The number of sets of $k+1$ non-adjacent edges at fixed edges $\gamma_1, \dots, \gamma_k$ is equal to the number of edges in G_k , that is, q_k . Summing these inequalities for q_k over all sets $\gamma = \{\gamma_1, \dots, \gamma_k\}$, which number is equal to p_k , one can find:

$$\sum_{\gamma} q_k \geq p_k \left(p_1 - \sum_{i=1}^4 i \cdot \alpha_i(k) + k \right).$$

Each set of $k + 1$ non-adjacent edges is present $k + 1$ times in the sum $\sum_{\gamma} q_k$: in each case k edges of these $k + 1$ ones are fixed as $\gamma_1, \dots, \gamma_k$ and the $(k + 1)$ -th belongs to G_k . Hence $\sum_{\gamma} q_k = (k + 1)p_{k+1}$. Using this relationship and the last inequality, we obtain the left inequality in Eq. (6).

Consider now the right inequality in Eq. (6). It is easy to show that $\bar{q}_k \geq 2k$. Indeed, \bar{q}_k is the number of edges in graph \bar{G}_k containing k edges $\gamma_1, \dots, \gamma_k$ and at least k edges adjacent to it. Hence, $q_k \leq p_1 - 2k$. Summing these inequalities over all sets $\gamma = \{\gamma_1, \dots, \gamma_k\}$ in the same manner as in the first part of this proof, we obtain the right inequality in Eq. (6).

3 AN ALGORITHM FOR SOLVING THE INVERSE PROBLEM

In this section we will give the general methodology of the IP solution for the case of equation (2) for simple molecular graphs with vertex degrees not exceeding four. The solution process can be subdivided into the following steps.

1. Let $[y_1, y_2]$ be an interval of possible values of property under consideration, n is a number of graph vertices. Find all partitions of n into 4 integers (including zero): $n = n_1 + n_2 + n_3 + n_4$, where n_i is the number of vertices with degree i ($i = 1, 2, 3, 4$).

2. For the sets $\{n_i\}$ check necessary and sufficient conditions for the existence of molecular graphs with given distribution of vertex degrees and eliminate unsuitable variants. Let $v_1 \geq v_2 \geq \dots \geq v_n > 0$ be an ordered set of numbers. A molecular graph with vertex degrees $\{v_i\}$ exists if and only if [24]:

a. $\sum_{i=1}^n v_i = \sum_{i=1}^4 i \cdot n_i$ is even;

b. for any integer r , $1 \leq r \leq n - 1$,

$$\sum_{i=1}^r v_i \leq r(r - 1) + \sum_{i=r+1}^n \min(r, v_i);$$

c. $\sum_{i=1}^n v_i \geq 2(n - 1)$, $n \geq 3$.

3. Calculate p_1 and p_2 using formulas (4) and (5). On this step some additional criteria related to the considered class of compounds may be also used for elimination of unsuitable sets $\{n_i\}$ (for example, $p_1 = n - 1$ for acyclic structures, $p_1 > n - 1$ for cycle-containing ones; $p_2 = n_2 + 3n_3$ for simple graphs corresponding to cata-condensed benzenoid hydrocarbons, etc.).

4. For the remaining sets $\{n_i\}$ using Eq. (3) and given interval of property values $[y_1, y_2]$ define an interval of possible values of index Z . Further the following elimination criteria are used:

$$1 + p_1 + p_2 \leq Z \leq 1 + p_1 + p_2 + \sum_{k=3}^{[n/2]} p_{k,\max} \quad (7)$$

where $p_{2,\max} = p_2$, $p_{k+1,\max} = \frac{1}{k+1} p_{k,\max} (p_1 - 2k)$, $k \geq 3$ – a priori found maximum values of p_k derived from inequalities (6).

5. Find all partitions of Z into $[n/2] + 1$ integers p_k , $k \geq 0$ (see (1)), where $p_0=1$, p_1 and p_2 are found above, and p_k satisfies the conditions (6).

6. For the sets $\{p_k\}$ check the following necessary conditions for the existence of graphs with given p_k , $k \geq 0$ [25]:

a. All roots of the acyclic polynomial $M(x) = \sum_{k=0}^{[n/2]} (-1)^k p_k x^{n-2k}$ are real numbers.

b. For the maximal root μ_1 of $M(x)$ the following inequalities are true:

$$\frac{1}{2} \sqrt{\Delta_{\max} + 1 + \sqrt{\Delta_{\max}^2 - 2\Delta_{\max} + 5}} \leq \mu_1 \leq 2\sqrt{\Delta_{\max} - 1} \quad ,$$

where $\Delta_{\max} = \max\{i : n_i > 0\}$.

7. Construct all molecular graphs with obtained sets $\{n_i\}$ and $\{p_k\}$, using, for example, the structure generator SMOG [26]. This program allows one to construct all chemical structures with given molecular formula and lists of required and forbidden fragments. The number of the required fragments can be also given. The generated graphs yield a solution of the IP.

Suppose that molecular structures include heteroatoms and multiple bonds. In this case simple molecular graphs are generated by the algorithm described above and then labeled graphs are constructed by arranging labels on edges and vertices in all non-equivalent ways.

4 EXAMPLES

Let us consider the examples of solving the IP for Eq. (2). Below r denotes a correlation coefficient, r_{LOO} – a leave-one-out cross-validation correlation coefficient, s – a standard deviation, F – a Fisher test, N – the number of structures in the training set.

Example 1. For the boiling points of alkylsulfides [27] we have obtained the following equation:

$$\begin{aligned} \text{B.p.} &= -17.61 + 52.744 \ln Z \quad , \\ r &= 0.995, r_{\text{LOO}} = 0.994, s = 2.87, F = 1886, N = 21, \end{aligned}$$

where B.p. is the boiling point ($^{\circ}\text{C}$), the training set consists of structures 1–21 (see Table 1).

Consider the problem of finding all sulfides with the number of carbon atoms from 2 to 8 and the boiling point from 113°C to 126°C. As a result of using the described algorithm structures 7, 9, 19, 20, 22, 23 from Table 1 were obtained. These structures are shown in Figure 1. The calculated values of the boiling point are in a good agreement with the experimental values. The experimental value for compound 23 is absent.

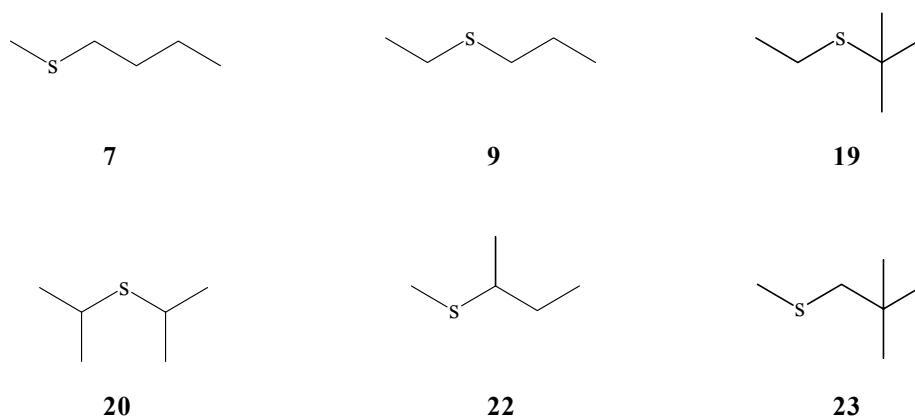


Figure 1. Structures generated in the example 1.

Table 1. Experimental and predicted values of boiling point for alkyl sulfides used in the Example 1.

No	Sulfide	B.p. exp., °C	B.p. calc., °C	Res.
1	dimethyl sulfide	37.3	40.3	-3.0
2	methyl ethyl sulfide	66.6	67.2	-0.6
3	methyl propyl sulfide	95.5	92.1	3.4
4	diethyl sulfide	92.0	92.1	-0.1
5	methyl isopropyl sulfide	84.4	85.0	-0.6
6	ethyl isopropyl sulfide	107.4	108.9	-1.5
7	methyl butyl sulfide	123.2	117.7	5.5
8	methyl isobutyl sulfide	112.5	108.9	3.6
9	ethyl propyl sulfide	118.5	117.7	0.8
10	methyl t-butyl sulfide	101.5	98.3	3.2
11	methyl amyl sulfide	145.0	143.0	2.0
12	ethyl butyl sulfide	144.2	143.0	1.2
13	dipropyl sulfide	142.8	143.0	-0.2
14	propyl isopropyl sulfide	132.0	134.9	-2.9
15	ethyl isobutyl sulfide	134.2	134.8	-0.6
16	methyl isoamyl sulfide	137.0	134.8	2.2
17	methyl 2-methylbutyl sulfide	139.0	137.7	1.3
18	ethyl s-butyl sulfide	133.6	137.7	-4.1
19	ethyl t-butyl sulfide	120.4	121.6	-1.2
20	diisopropyl sulfide	120.0	125.2	-5.2
21	methyl 1-ethylpropyl sulfide	137.0	140.4	-3.4
22	methyl s-butyl sulfide	114.5	113.5	1.0
23	methyl 2,2-dimethylpropyl sulfide	—	121.59	—

Example 2. For the solubility of saturated alcohols in water [28] we have obtained the following correlation:

$$-\log X = -0.8 + 1.186 \ln Z,$$
$$r = 0.976, r_{\text{LOO}} = 0.975, s = 0.21, F = 966, N = 50,$$

where X denotes the mole fraction solubility of alcohols in water. Let us find all such compounds with $2.6 \leq -\log X \leq 3.0$. Structural formulas of 20 generated alcohols are shown in Figure 2. Experimental and calculated values of $-\log X$ for the training set (structures 1–50) and calculated values for the generated structures (51–59) not belonging to the training set are given in Table 2.

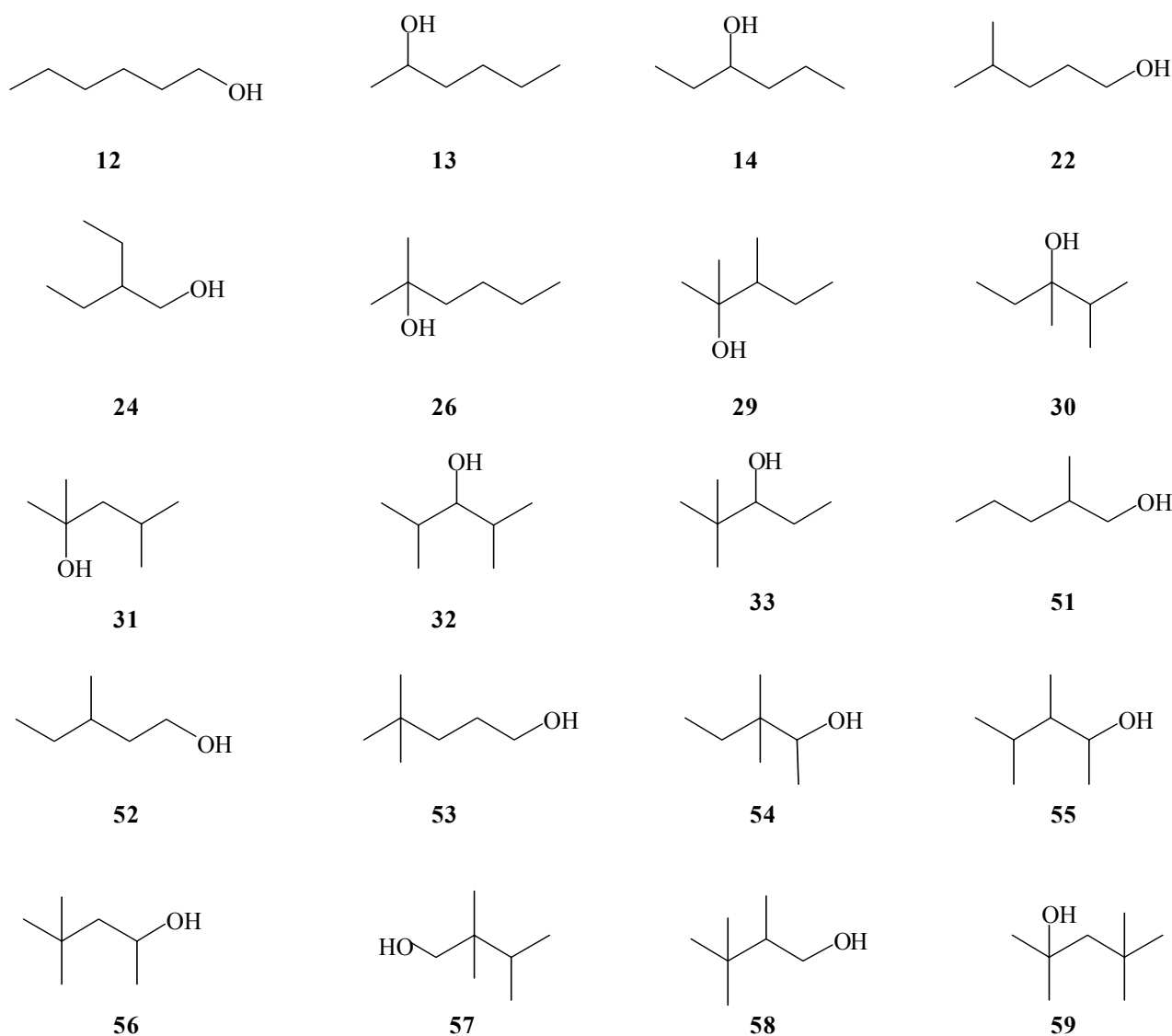


Figure 2. Structures of alcohols generated in the example 2.

Table 2. Experimental and predicted values of the mole fraction solubility in water for alcohols (example 2).

No	Alcohols	$-\log X$, exp.	$-\log X$, calc.	Res.
1	1-Butanol	1.750	1.666	0.084
2	2-Methylpropanol	1.743	1.508	0.235
3	2-Butanol	1.724	1.508	0.216
4	1-Pentanol	2.332	2.242	0.090
5	3-Methylbutanol	2.254	2.044	0.210
6	2-Methylbutanol	2.207	2.147	0.060
7	2-Pentanol	2.025	2.044	-0.019
8	3-Pentanol	1.961	2.147	-0.186
9	3-Methyl-2-butanol	1.926	1.931	-0.005
10	2-Methyl-2-butanol	1.608	1.806	-0.198
11	2,2-Dimethyl-1-propanol	2.030	1.806	0.224
12	1-Hexanol	2.957	2.811	0.146
13	2-Hexanol	2.612	2.628	-0.016
14	3-Hexanol	2.542	2.692	-0.150
15	3-Methyl-3-pentanol	2.109	2.488	-0.379
16	2-Methyl-2-pentanol	2.233	2.330	-0.097
17	2-Methyl-3-pentanol	2.445	2.560	-0.115
18	3-Methyl-2-pentanol	2.458	2.560	-0.102
19	2,3-Dimethyl-2-butanol	2.118	2.242	-0.124
20	3,3-Dimethyl-1-butanol	2.870	2.330	0.540
21	3,3-Dimethyl-2-butanol	2.359	2.242	0.117
22	4-Methylpentanol	2.737	2.628	0.109
23	4-Methyl-2-pentanol	2.534	2.412	0.122
24	2-Ethyl-1-butanol	2.956	2.753	0.203
25	1-Heptanol	3.554	3.383	0.171
26	2-Methyl-2-hexanol	2.820	2.919	-0.099
27	3-Methyl-3-hexanol	2.729	3.018	-0.289
28	3-Ethyl-3-pentanol	2.579	3.152	-0.573
29	2,3-Dimethyl-2-pentanol	2.615	2.866	-0.251
30	2,3-Dimethyl-3-pentanol	2.588	2.919	-0.331
31	2,4-Dimethyl-2-pentanol	2.678	2.692	-0.014
32	2,4-Dimethyl-3-pentanol	2.962	2.969	-0.007
33	2,2-Dimethyl-3-pentanol	2.893	2.866	0.027
34	3-Heptanol	3.132	3.273	-0.141
35	4-Heptanol	3.133	3.234	-0.101
36	1-Octanol	4.091	3.953	0.138
37	2,2,3-Trimethyl-3-pentanol	3.018	3.234	-0.216
38	2-Octanol	3.811	3.767	0.044
39	2-Ethylhexanol	3.915	3.886	0.029
40	1-Nonanol	4.745	4.524	0.221
41	2-Nonanol	4.490	4.337	0.153
42	3-Nonanol	4.402	4.412	-0.010
43	4-Nonanol	4.330	4.383	-0.053
44	5-Nonanol	4.240	4.397	-0.157
45	2,6-Dimethyl-3-heptanol	4.253	4.076	0.177
46	3,5-Dimethyl-4-heptanol	4.046	4.239	-0.193
47	1,1-Diethylpentanol	4.165	4.273	-0.108
48	7-Methyloctanol	4.240	4.337	-0.097
49	3,5,5-Trimethylhexanol	4.251	3.886	0.365
50	1-Decanol	5.444	5.095	0.349
51	2-Methyl-1-pentanol	—	2.692	—
52	3-Methyl-1-pentanol	—	2.692	—
53	4,4-Dimethyl-1-pentanol	—	2.919	—
54	3,3-Dimethyl-2-pentanol	—	2.919	—
55	3,4-Dimethyl-2-pentanol	—	2.969	—
56	4,4-Dimethyl-2-pentanol	—	2.692	—
57	2,2,3-Trimethyl-1-butanol	—	2.919	—
58	2,3,3-Trimethyl-1-butanol	—	2.866	—
59	2,4,4-Trimethyl-2-pentanol	—	2.969	—

Example 3. For a training set of hydrocarbons with the number of carbon atoms n from 6 to 8 and known values of the boiling point t_b [29], we obtained an equation

$$\text{B.p.} = -98.2 + 41.81 \ln Z + 11.1n_1 + 9.22n_2 + 5.1n_3, \\ r = 0.983, r_{\text{LOO}} = 0.981, s = 4.28, F = 525, N = 78$$

Experimental data and calculated values of the boiling point for the training set are given in Table 3 (structures 1–78). Further a problem of generating hydrocarbons with $6 \leq n \leq 8$ and the boiling point $100 \pm 1.5^\circ\text{C}$ was considered. Using the described algorithm, 77 structures were generated. Then 57 tetrahedranes, pyramidananes, propellanes, and other extremely strained and unrealistic structures were removed manually. The remained 20 structures (no. 18, 33, 34, 55, 74, 79–93) are shown in Table 3. The structures 18, 33, 34, 55, 74 were present in the training set, structures 79 and 80 were present in the database [29] while the other generated structures were not present in that database.

Table 3. Experimental and predicted values of boiling point for hydrocarbons considered in example 3.

No	Structure	Name	B.p. exp., °C	B.p. calc., °C	Res.
1		n6	68.7	68.2	0.5
2		2mn5	60.3	58.9	1.4
3		22mn4	49.7	47.3	2.4
4		1ipc3	58.3	63.0	-4.7
5		1e2mc3	63.0	65.9	-2.9
6		112mc3	52.6	53.3	-0.7
7		1ec4	70.7	73.3	-2.6
8		11mc4	53.6	58.9	-5.3
9		1mc5	71.8	70.8	1.0
10		bcpr	76.0	74.1	1.9
11		bc220hx	83.0	78.1	4.9
12		mbc210p	60.5	66.5	-6.0
13		n7	98.5	97.4	1.1

Table 3. (Continued).

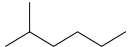
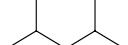
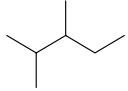
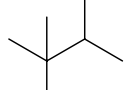
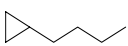

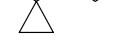


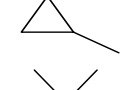
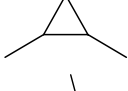
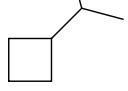
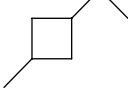
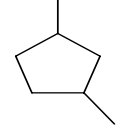
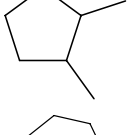
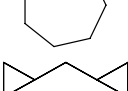

No	Structure	Name	B.p. exp., °C	B.p. calc., °C	Res.
14		2mn6	90.0	88.7	1.3
15		24mn5	80.5	78.9	1.6
16		23mn5	89.8	84.1	5.7
17		223mn4	80.9	70.0	11.2
18		1bc3	98.0	100.3	-2.3
19		12ec3	90.0	96.4	-6.4
20		1m1pc3	84.9	90.1	-5.2
21		11ec3	88.6	93.8	-5.2
22		1m1ipc3	81.5	83.9	-2.4
23		12m1ec3	85.2	85.9	-0.7
24		1123mc3	78.0	79.5	-1.5
25		1ipc4	92.7	94.7	-2.0
26		1e3mc4	89.5	94.7	-5.2
27		13mc5	91.3	93.0	-1.7
28		12mc5	95.6	94.7	0.9
29		c7	118.4	107.1	11.3
30		dcprm	102.0	103.0	-1.0

Table 3. (Continued).

No	Structure	Name	B.p. exp., °C	B.p. calc., °C	Res.
31		bc320h	110.5	106.7	3.8
32		bc410h	116.0	105.5	10.5
33		2mbc310hx	100.0	99.4	0.6
34		6mbc310hx	103.0	100.8	2.2
35		mbc310hx	92.0	95.7	-3.7
36		11ms22p	78.0	84.5	-6.5
37		tc410024h	105.0	104.1	0.9
38		tc3110024h	107.0	105.1	1.9
39		tc410013h	107.5	102.0	5.5
40		tec320h	108.5	106.0	2.5
41		2mn7	117.6	117.9	0.3
42		3mn7	118.9	120.7	-1.8
43		3en6	118.5	122.0	-3.5
44		24mn6	109.4	111.1	-1.7
45		22mn6	106.8	105.0	1.8
46		3e2mn5	115.6	114.2	1.4
47		224mn5	99.2	94.8	4.4
48		3e3mn5	118.2	113.2	5.0
49		1pec3	128.0	129.6	-1.6
50		1spec3	117.7	123.2	-5.5

Table 3. (Continued).

No	Structure	Name	B.p. exp., °C	B.p. calc., °C	Res.
51		b2mc3	124.0	124.3	-0.3
52		5msbc3	115.5	116.3	-0.8
53		ib2mc3	110.0	115.0	-5.0
54		112m2ec3	104.5	104.0	0.5
55		11223mc3	100.5	98.7	1.8
56		1sbc4	123.0	126.3	-3.3
57		12ec4	119.0	127.4	-8.4
58		1pc5	131.0	130.6	0.4
59		1ipc5	126.4	123.2	3.2
60		124mc5	115.0	116.3	-1.3
61		1e1mc5	121.5	121.1	0.4
62		112mc5	114.0	112.8	1.2
63		1ec6	131.8	131.5	0.3
64		12mc6	126.6	124.3	2.3
65		11mc6	119.5	117.5	2.0
66		bcprm	129.0	132.5	-3.5
67		bc330o	137.0	135.6	1.4
68		bc510o	141.0	134.8	6.2

Table 3. (Continued).

No	Structure	Name	B.p. exp., °C	B.p. calc., °C	Res.
69		2mhc221h	125.0	128.6	-3.6
70		2mhc320h	130.5	130.3	0.3
71		1mhc221h	117.0	124.2	-7.2
72		1mhc410h	125.0	125.1	-0.1
73		33mhc310hx	115.0	115.0	0.0
74		2244mhc b	104.0	101.3	2.7
75		1223mhc b	105.0	101.6	3.3
76		tc510035o	142.0	132.9	9.1
77		tc3210o	136.0	134.2	1.7
78		3mtc2210h	120.5	129.4	-8.9
79		1mc6	101.0	100.3	0.7
80		223mn5	109.8	100.9	8.9
81		1cpbc110b		101.0	
82		2scpbc210p		101.0	
83		3mtc310026hx		99.7	
84		3mtc220026hx		100.8	

Table 3. (Continued).

No	Structure	Name	B.p. exp., °C	B.p. calc., °C	Res.
85		3mtc310024hx		99.7	
86		1etc210025p		99.8	
87		2etc210025p		98.7	
88		2pbc110b		100.8	
89		2ebc210p		100.8	
90		2mbc220hx		100.8	
91		5mbc112hx		100.8	
92		2mbc112hx		99.4	
93		1m2tbc3		100.9	

5 CONCLUSIONS

An algorithm has been developed for solving the inverse problem for correlation equation containing the Hosoya index and some other simple topological descriptors. A method has been tested for a number of correlation equations of described type. The results obtained show a sufficiently high efficiency of the suggested approach for searching the structures of some simple classes of organic compounds with prescribed property values.

Acknowledgment

This work was partially supported by the Virtual Computational Chemistry Laboratory INTAS 00–0363 grant.

6 REFERENCES

- [1] M. I. Skvortsova, I. I. Baskin, V. A. Palyulin, O. L. Slovokhotova and N. S. Zefirov, Structural Design. Inverse Problems for Topological Indices in QSAR/QSPR Studies; in: *AIP Conference Proceedings 330, E.C.C.C.1, Computational Chemistry, F.E.C.S. Conference, Nancy, France, May 1994*, Eds. F. Bernardi, and J.–L. Rivail, AIP Press, Woodbury, N.Y., 1995, 486–499.
- [2] V. Kvasnička and J. Pospichal, Canonical Indexing and Constructive Enumeration of Molecular Graphs, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 99–105.
- [3] L. B. Kier, L. H. Hall, and J. W. Frazer, Design of Molecules from Quantitative Structure–Activity Relationship Models. 1. Information Transfer between Path and Vertex Degree Counts, *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 143–147.
- [4] L. B. Kier, L. H. Hall, and J. W. Frazer, Design of Molecules from Quantitative Structure–Activity Relationship Models. 2. Derivation and Proof of Information Transfer Relating Equations, *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 148–152.
- [5] L. H. Hall, R. S. Dailey, and L. B. Kier, Design of Molecules from Quantitative Structure–Activity Relationship

- Models. 3. Role of Higher Order Path Counts: Path 3, *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 598–603.
- [6] I. I. Baskin, E. V. Gordeeva, R. O. Devdariani, N. S. Zefirov, V. A. Palyulin, and M. I. Stankevich, Methodology of Solution of the Inverse Problem in “Structure – Property” Relationship Studies for the Case of Topological Indices, *Dokl. Akad. Nauk SSSR (Russ.)* **1989**, *307*, 613–617.
- [7] M. I. Skvortsova, I. V. Stankevich, and N. S. Zefirov, Generation of Molecular Structures of Polycondensed Benzenoid Hydrocarbons by the Randić Index, *Zh. Strukt. Khim. (Russ.)* **1992**, *33*, 99–104.
- [8] N. S. Zefirov, V. A. Palyulin, and E. V. Radchenko, The Problem of Generation of Structures with Given Properties. The Solution of Inverse Problem for Centric Balaban Index, *Dokl. Akad. Nauk SSSR (Russ.)* **1991**, *316*, 921–924.
- [9] M. I. Skvortsova, I. I. Baskin, O. L. Slovokhotova, V. A. Palyulin, and N. S. Zefirov, Inverse Problem in QSAR/QSPR Studies for the Case of Topological Indices Characterizing Molecular Shape, *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 630–634.
- [10] E. V. Gordeeva, M. S. Molchanova, and N. S. Zefirov, General Methodology and Computer Program for the Exhaustive Restoring of Chemical Structures by Molecular Connectivity Indexes. Solution of the Inverse Problem in QSAR/QSPR, *Tetrahedron Computer Methodology* **1990**, *3*, 389–415.
- [11] M. I. Skvortsova, O. L. Slovokhotova, I. I. Baskin, V. A. Palyulin, and N. S. Zefirov, Inverse Problem in the Structure–Property Relationship Analysis for the Case of Information Topological Indices, *Doklady Chemistry*, **1997**, *357*, 252–254.
- [12] M. I. Stankevich, I. V. Stankevich, and N. S. Zefirov, Topological Indexes in Organic Chemistry, *Usp. Khim. (Russ.)* **1988**, *57*, 337–366.
- [13] G. W. A. Milne, Mathematics as a Basis for Chemistry, *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 639–644.
- [14] H. Hosoya, Topological Index. A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons, *Bull. Chem. Soc. Japan*. **1971**, *44*, 2332–2339.
- [15] H. Hosoya, The Topological Index Z Before and After 1971, *Internet Electron. J. Mol. Des.* **2002**, *1*, 428–442, <http://www.biochempress.com>.
- [16] I. Gutman and A. Shalabi, Topological Properties of Benzenoid Systems. XXIX. On Hosoya’s Topological Index, *Z. Naturforsch.* **1984**, *39a*, 797–799.
- [17] O. Ivanciuc, Design of Topological Indices. Part 18. Modeling the Physical Properties of Alkanes with Molecular Graph Descriptors Derived from the Hosoya Operator, *Rev. Roum. Chim.* **2001**, *46*, 129–141.
- [18] R. Todeschini and V. Consonni. *Handbook of Molecular Descriptors*, Wiley–VCH, Weinheim, 2000.
- [19] J. Devillers and A. T. Balaban (Eds.), *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon and Breach, Amsterdam, 1999.
- [20] S. Nikolić, D. Plavšić, and N. Trinajstić, On the Z–counting Polynomial for Edge–weighted Graphs, *J. Math. Chem.* **1992**, *9*, 381–387.
- [21] D. Plavšić, M. Šoškić, Z. Daković, I. Gutman, and A. Graovac, Extension of the Z–matrix to Cycle–Counting and Edge–Weighted Molecular Graphs, *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 529–534.
- [22] H. Hosoya, K. Hosoi, and I. Gutman, A Topological Index for the Total π –Electron Energy. Proof of a Generalized Hückel Rule for an Arbitrary Network, *Theor. Chim. Acta (Berlin)* **1975**, *38*, 37–47.
- [23] H. Hosoya and M. Murakami, Topological Index as Applied to π –Electronic Systems. II. Topological Bond Order, *Bull. Chem. Soc. Japan* **1975**, *48*, 3512–3517.
- [24] F. Harary, *Graph Theory*, Addison–Wesley, Reading, MA, 1969.
- [25] O. J. Heilmann and E. H. Lieb, Theory of Monomer–Dimer Systems, *Commun. Math. Phys.* **1972**, *25*, 190–232.
- [26] M. S. Molchanova, V. V. Shcherbukhin and N. S. Zefirov, Computer Generation of Molecular Structures by SMOG program, *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 888–899.
- [27] N. S. Zefirov and V. A. Palyulin, QSAR for Boiling Points of “Small” Sulfides. Are the “High–Quality Structure–Property–Activity Regressions” the Real High Quality QSAR Models? *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1022–1027.
- [28] A. Cammarata, Molecular Topology and Aqueous Solubility of Aliphatic Alcohols, *J. Pharm. Sci.* **1979**, *68*, 839–842.
- [29] Rücker G. and Rücker C. On Topological Indices, Boiling Points, and Cycloalkanes, *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 788–802.