Inverse Problem in QSAR/QSPR Studies for the Case of Topological Indices Characterizing Molecular Shape (Kier Indices)

Mariya I. Skvortsova, Igor I. Baskin, Olga L. Slovokhotova, Vladimir A. Palyulin, and Nikolai S. Zefirov*

Department of Chemistry, Moscow State University, Moscow 119899, Russia

Received August 7, 1992

An inverse problem of construction of chemical structures with given value of property is considered. An algorithm for the solution of the inverse problem for the case of correlating equation "property–topological "κ"-indices" is proposed, and examples of its application are given.

INTRODUCTION

The studies of quantitative structure–activity and structure–property relationships (QSAR/QSPR[1]) can be considered as one of important goals of modern theoretical/mathematical chemistry. The structure of molecules in QSAR/QSPR studies is usually described in terms of topological, electronic, and other characteristics, and the property (or activity) is correlated with these parameters for some training set of compounds. The main goal of construction of 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.

Earlier, we have pointed out[2-6] the fundamental problem which one can call the "inverse QSAR/QSPR approach". Indeed, if the structure–property correlating equation is found, the inverse problem of the generation of structures with a given value of property can be formulated. The topological indices (TI's) (invariants of weighted molecular graphs) are often used in QSAR/QSPR studies as descriptors of molecular structure.[7-10] The inverse problem is reduced in this case to the procedure of generation of structures with the given values of some TI's.

We have considered previously[2] the inverse problem for the cases of Randic and Wiener indices. That paper had caused the development of a new branch of mathematical chemistry dealing with the reconstruction of molecular graphs from their invariants. It should be emphasized that in graph theory the problem of graph reconstruction from a single topological index is not considered, as far as it can be always solved by looking through all possible variants and in the general case can have a great number of solutions. However, for some classes of molecular graphs, as it was shown,[2-6,11] this problem is of practical importance and can be solved using computer programs in reasonable time for Randic index[12] χ, Wiener index[13] W, and Balaban index[9] B avoiding the combinatorial explosion. The development of algorithms for solving the inverse problem for other indices and for multiple regression equations becomes an important problem of practical QSAR studies.

In the present paper an algorithm of this kind for regression equations containing "κ"-indices ${}^0\kappa$, ${}^1\kappa$, ${}^2\kappa$, ${}^3\kappa$, ${}^1\kappa_\alpha$, ${}^2\kappa_\alpha$, and ${}^3\kappa_\alpha$ characterizing molecular shape[14,15] and introduced by Kier for QSAR/QSPR studies is presented.

TOPOLOGICAL "κ"-INDICES OF MOLECULAR SHAPE

Indices ${}^i\kappa$ (i = 1, 2, 3) are defined in terms of the numbers of graph vertices n and the numbers of paths ${}^iP$ with length i (i = 1, 2, 3) according to the following formulas:[14,15]

$$ {}^1\kappa = \frac{n(n-1)^2}{({}^1P)^2}, \quad {}^2\kappa = \frac{(n-1)(n-2)^2}{({}^2P)^2} \tag{1} $$

$$ {}^3\kappa = \begin{cases} \dfrac{(n-3)(n-2)^2}{({}^3P)^2} & \text{for even } n\ (n>3) \\ \dfrac{(n-1)(n-3)^2}{({}^3P)^2} & \text{for odd } n\ (n>3) \end{cases} \tag{2} $$

Numerators of expressions 1 and 2 for ${}^i\kappa$ are equal with precision to constant multiplier to values ${}^iP_{max}{}^iP_{min}$ (i = 1, 2, 3), where ${}^iP_{min}$ is the number of i-bond paths in a chain (e.g. Figure 1a) with n vertices, ${}^iP_{max}$ is the number of i-bond paths in a star graph (i = 2) (e.g. Figure 1b), in a "double star" (i = 3) (e.g. Figure 1d,e), or in a complete graph (i = 1) (e.g. Figure 1c). The topological index ${}^0\kappa$ is an informational one, and it is defined by the following formula ${}^0\kappa = -n\sum_i(n_i/n)\log(n_i/n)$, where n_i is the number of topologically equivalent vertices in an ith equivalence class. The vertex partition into classes is carried out according to their topological characteristics, the most "detailed" partition resulting from the orbits of the graph automorphism group.

The indices of molecular shape ${}^i\kappa_\alpha$ ($i = \overline{1,3}$) distinguishing atom differences were also introduced by Kier.[14,15] They are calculated similarly to ${}^i\kappa$ using $n+\alpha$ and $({}^iP+\alpha)$ instead of n and ${}^iP$ ($i = \overline{1,3}$) in formulas 1 and 2, α being some parameter. The value of parameter α_j for a non-$C_{(sp^3)}$ atom j is evaluated using covalent atomic radii r_j:

$$ \alpha_j = (r_j/r_{C(sp^3)}) - 1, \quad \alpha = \sum_{j=1}^{n}\alpha_j $$

The values of α_j used[14,15] are given in Table I.

ALGORITHM FOR THE CASE OF ${}^i\kappa(i = \overline{0,3})$ INDICES

We shall give the methodology of the solution of an inverse problem for the case when the vertex partition into equivalence classes is conducted according to their degrees, that is ${}^0\kappa = -n\sum_{i=1}^{4}(d_i/n)\log(d_i/n)$, and the regression equation connecting property "y" and parameters ${}^i\kappa$ (i = 0, 1, 2, 3) has the following form:

$$ y = \sum_{i=0}^{3} {}^ia\,{}^i\kappa + a_4 \qquad {}^ia = \text{constant}, \quad i = 0, 1, 2, 3 \tag{3} $$

The solution of the inverse problem can be divided into seven main steps:

0095-2338/93/1633-0630$04.00/0

Figure 1. Graphs: (a) chain ($n = 5$); (b) star ($n = 8$); (c) complete graph ($n = 5$); (d) "double star" for even n ($n = 8$); (e) "double star" for odd n ($n = 9$).

Table I. Values of Parameters α_i for Atoms of Different Types

type of atom		α_i
C(sp^3)	**1**, —C(—)—; **2**, —C(—)—; **3**, —C—; **4**, C—	0
C(sp^2)	**5**, =C(—)(—); **6**, =C—; **7**, C=	−0.13
C(sp)	**8**, ≡C—; **9**, =C=; **10**, C≡	−0.22
N(sp^3)	**11**, —N(—)—; **12**, —N—; **13**, N—	−0.04
N(sp^2)	**14**, =N—; **15**, N=	−0.20
N(sp)	**16**, N≡	−0.29
O(sp^3)	**17**, —O—; **18**, O—	−0.04
O(sp^2)	**19**, O=	−0.20
P(sp^3)	**20**, —P(—)—; **21**, —P—; **22**, P—	0.43
P(sp^2)	**23**, =P(—)(—)	0.30
S(sp^3)	**24**, —S—; **25**, S—	0.35
S(sp^2)	**26**, =S(—)(—); **27**, S=	0.22
F	**28**, F—	−0.07
Cl	**29**, Cl—	0.29
Br	**30**, Br—	0.48
I	**31**, I—	0.73

(1) Let n be a number of graph vertices. Find all partitions of n into four integers including zero: $n = d_1 + d_2 + d_3 + d_4$, where d_i is the number of vertices with degree i ($i = 1, 2, 3, 4$).

(2) For the set $\{d_i\}$ check necessary and sufficient conditions for the existence of molecular graphs with given distribution of vertex degrees and eliminate unsuitable variants. Let $v_1 \geq v_2 \geq \ldots \geq v_n > 0$ be an ordered set of n numbers. A molecular graph with vertex degrees $\{v_i\}$ exists if and only if[16]

(a) $\sum_{i=1}^{n} v_i$ is even and for any integer r, $1 \leq r \leq n-1$

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

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

(3) For the set $\{d_i\}$ calculate the index ${}^0\kappa = -n\sum_{i=1}^{4}(d_i/n) \times \log(d_i/n)$ and parameters ${}^1P = (d_1 + 2d_2 + 3d_3 + 4d_4)/2$, ${}^2P = d_2 + 3d_3 + 6d_4$.

(4) Substitute the values ${}^0\kappa$, ${}^1P$, ${}^2P$ in eq 3 and obtain an equation of the type $A = {}^3P$ (A = constant > 0).

(5) Let the type of edge (i,j) in the graph be designated by the pair of numbers (v_i, v_j), $v_i \leq v_j$, where v_i and v_j are degrees of vertices incident to the edge (i,j). Then find the values n_{kl} (number of edges of the type (k,l)) using the formulas

$$\begin{cases} {}^3P = n_{22} + 2n_{23} + 3n_{24} + 4n_{33} + 6n_{34} + 9n_{44} \\ {}^1P = n_{12} + n_{13} + n_{14} + n_{22} + n_{23} + n_{24} + n_{33} + n_{34} + n_{44} \end{cases}$$

For the rejection of unsuitable variants use the following criterion: if $0 < n_{kk} = d_k \leq 2$ ($k = 2, 3, 4$), than the set $\{n_{kl}\}$ should be rejected.

(6) Check that each set $\{n_{kl}\}$ corresponds to the sequence $\{d_i\}$; e.g. the following conditions are met:

$$d_k = \Big(\sum_{l,l \neq k} n_{kl} + 2n_{kk}\Big)/k \qquad n_{kl} \equiv n_{lk} \quad \text{for } k > l,\ n_{11} = 0$$

and reject all unsuitable variants.

(7) Use the sequence $\{d_i\}$ for generating all molecular graphs with this set of vertex degrees and with the set of types of edges $\{n_{kl}\}$ using, for example, algorithms from refs 17 and 18. These graphs will yield a solution of the inverse problem.

ALGORITHM FOR THE CASE OF ${}^i\kappa (i = \overline{0,3})$ AND ${}^j\kappa_\alpha (j = \overline{1,3})$ INDICES

The algorithm suggested in the present paper for the solution of inverse problem for the case of ${}^i\kappa$ ($i = \overline{0,3}$) can be generalized for the regression equation of the following type

$$y = \sum_{i=0}^{3} {}^i a\, {}^i\kappa + \sum_{i=1}^{3} {}^i a_\alpha\, {}^i\kappa_\alpha + b \qquad {}^i a, {}^i a_\alpha, b = \text{constant} \quad (4)$$

and consists of eight steps:

(1) Let n be the number of vertices in a hydrogen-depleted molecular graph having multiple edges. Find all sequences $\{k_j\}$, $n = \sum_{j=1}^{m} k_j$, k_j is the number of atoms of type j, $m = 31$ is the number of atom types (see the second column of Table I).

(2) Calculate d_i' (the number of vertices with degree i ($i = \overline{1,5}$) in a molecular graph with multiple edges):

$$\begin{cases} d_1' = k_4 + k_{13} + k_{18} + k_{22} + k_{25} + k_{28} + k_{29} + k_{30} + k_{31} \\ d_2' = k_3 + k_7 + k_{12} + k_{15} + k_{17} + k_{19} + k_{21} + k_{24} + k_{27} \\ d_3' = k_2 + k_6 + k_{10} + k_{11} + k_{14} + k_{16} + k_{20} \\ d_4' = k_1 + k_5 + k_8 + k_9 + k_{26} \\ d_5' = k_{23} \end{cases}$$

(3) For the set $\{d_i'\}$ check the following[16] necessary and sufficient conditions for existence of graphs with multiple edges and vertex degrees $v_1' \geq v_2' \geq \ldots \geq v_n'$: $\sum_{i=1}^{n} v_i'$ is even and ${}^1/_2\sum_{i=1}^{n} v_i' \geq v_1'$. Unsuitable sequences $\{k_j\}$ should be rejected.

(4) Calculate d_i (the number of vertices with degree i ($i = \overline{1,4}$) in a "simple" molecular graph (e.g. with ordinary edges) corresponding to the molecular graph with multiple edges).

$$\begin{cases} d_1 = k_4 + k_7 + k_{10} + k_{13} + k_{15} + k_{16} + k_{18} + k_{19} + \\ \qquad k_{22} + k_{25} + k_{27} + k_{28} + k_{29} + k_{30} + k_{31} \\ d_2 = k_3 + k_6 + k_8 + k_9 + k_{12} + k_{14} + k_{17} + k_{21} + k_{24} \\ d_3 = k_2 + k_5 + k_{11} + k_{20} + k_{26} \\ d_4 = k_1 + k_{23} \end{cases}$$

(5) For the set $\{d_i\}$ check the necessary and sufficient conditions for the existence of connected simple graphs with vertex degrees $v_1 \geq v_2 \geq \ldots \geq v_n > 0$ (see step 2 of the algorithm for the case of ${}^i\kappa$, $i = \overline{0,3}$).

(6) Find $\{n_{kl}\}$ (the number of edges of type (k,l) of a simple molecular graph from conditions):

$$^1/_2(2d_2 + 3d_3 + 4d_4 - d_1) = n_{23} + n_{24} + n_{34} + n_{22} + n_{33} + n_{44}$$

$$\begin{cases} n_{23} + n_{24} + 2n_{22} \le 2d_2 \\ n_{23} + n_{34} + 2n_{33} \le 3d_3 \\ n_{24} + n_{34} + 2n_{44} \le 4d_4 \end{cases}$$

If $0 < n_{kk} = d_k \le 2$ $(k = \overline{2, 3, 4})$, then $\{n_{kl}\}$ is unsuitable.

(7) Calculate $^0\kappa$, ^iP $(i = \overline{1,3})$ using the values of $\{d_i\}$:

$$^0\kappa = -n\sum_{i=1}^{4}(d_i/n)\log(d_i/n)$$

$$^1P = {}^1/_2(d_1 + 2d_2 + 3d_3 + 4d_4)$$

$$^2P = d_2 + 3d_3 + 6d_4$$

$$^3P = n_{22} + 2n_{23} + 3n_{24} + 4n_{33} + 6n_{34} + 9n_{44}$$

and $\alpha = \sum_{j=1}^{m}\alpha_j k_j$. Check that these parameters satisfy eq 4 with a given value of y and reject unsuitable variants.

(8) Generate all molecular graphs by using parameters n, $\{k_j\}$, and Table I (see step 7 in algorithm for the case of $^i\kappa$ $(i = \overline{0,3})$).

REMARKS

(1) The suggested algorithm can be easily expanded for the case when the correlating equation is not linear with respect to indices $^i\kappa$ $(i = 0, 1, 2, 3)$, $^i\kappa_\alpha$ $(i = 1, 2, 3)$ (for example, it can contain functions $(^i\kappa)^2$, $(^i\kappa)^{-1}$, etc.). Apparently, the algorithm becomes much simpler if some indices $^i\kappa$ are absent in the equation.

(2) From the regression equation and given value of property y it is possible to estimate the parameter n and to take proper values of n on the first step of the algorithms. This estimation is based on the following inequalities:

$$n/2 \le {}^1P \le 2n, \quad n \le {}^2P \le 6n, \quad n/2 \le {}^3P \le 18n$$

From Table I it follows that $\max_i \alpha_i = 0.73$, $\min_i \alpha_i = -0.29$. Hence $-0.29n \le \alpha \le 0.73n$.

$$\begin{cases} 0.21n \le {}^1P + \alpha \le 2.73n \\ 0.71n \le {}^2P + \alpha \le 6.73n \\ 0.21n \le {}^3P + \alpha \le 18.73n \end{cases}$$

$$\begin{cases} 0.71n \le n + \alpha \le 1.73n \\ 0.71n - 1 \le n + \alpha - 1 \le 1.73n - 1 \\ 0.71n - 2 \le n + \alpha - 2 \le 1.73n - 2 \\ 0.71n - 3 \le n + \alpha - 3 \le 1.73n - 3 \end{cases}$$

Then,

$$0 \le {}^0\kappa \le n, \qquad \frac{n-2}{4} \le {}^1\kappa \le 4n - 4,$$

$$\frac{n-5}{36} \le {}^2\kappa \le n - 2, \quad \frac{n-7}{324} \le {}^3\kappa \le 4n - 12$$

$$\begin{cases} \dfrac{(0.71)^3 n - 2\cdot(0.71)^2}{(2.73)^2} \le {}^1\kappa_\alpha \le \dfrac{(1.73)^3 n}{(0.21)^2} \\ \dfrac{(0.71)^3 n - 5\cdot(0.71)^2}{(6.73)^2} \le {}^2\kappa_\alpha \le \dfrac{(1.73)^3 n - 2\cdot(1.73)^2}{(0.71)^2} \\ \dfrac{(0.71)^3 n - 7\cdot(0.71)^2}{(18.73)^2} \le {}^3\kappa_\alpha \le \dfrac{(1.73)^3 n - 3\cdot(1.73)^2}{(0.21)^2} \end{cases}$$

If for some value n (for example, $n < 2$ for $^1\kappa$, $n < 5$ for $^2\kappa$, and $n < 7$ for $^3\kappa$) the lowest bound in inequality for $^i\kappa$ or $^i\kappa_\alpha$ is negative, then one should replace it by zero.

As a result, the inequalities of the following type are established:

$$^iB \le {}^i\kappa \le {}^iA \quad i = \overline{0,3}; \qquad {}^iB_\alpha \le {}^i\kappa_\alpha \le {}^iA_\alpha \quad i = \overline{1,3}$$

where $^iA, {}^iB, {}^iA_\alpha, {}^iB_\alpha \ge 0$ are expressions of the following kind: $dn + c$ (d, c = constant). It is supposed that in regression eq 4 $^ia \ge 0$ $(i = \overline{0,3})$, $^ia_\alpha \ge 0$ $(i = \overline{1,3})$. Then one can write

$$\sum_{i=0}^{3} {}^iB^ia + \sum_{i=1}^{3} {}^iB_\alpha {}^ia_\alpha \le y - b \le \sum_{i=0}^{3} {}^iA^ia + \sum_{i=1}^{3} {}^iA_\alpha {}^ia_\alpha$$

Suppose that for some value j we have $^ja < 0$ (or $^ja_\alpha < 0$) and other coefficients are nonnegative. Then the following estimation is true:

$$\sum_{i\ne j} {}^iB^ia + \sum_{i} {}^iB_\alpha {}^ia_\alpha + {}^ja^jA \le y - b \le \sum_{i\ne j} {}^iA^ia + \sum_{i} {}^iA_\alpha {}^ia_\alpha + {}^ja^jB$$

Evidently, the left and right parts of these inequalities are linear functions of n with known coefficients. Hence it is possible to estimate the suitable values of n: $n_1 \le n \le n_2$.

EXAMPLES

Now we shall give some examples of solving the inverse problem for κ-indices. We have obtained the following correlations:

$$y_1 = 3.971(^1\kappa) + 1.285(^2\kappa) - 0.253(^3\kappa) + 2.683$$
$$r = 0.996, \quad s = 0.48, \quad n = 69$$

$$y_2 = 0.41(^1\kappa) + 0.16(^2\kappa) + 0.021(^3\kappa) - 0.96$$
$$r = 0.985, \quad s = 0.17, \quad n = 50$$

$$y_3 = 0.362(^1\kappa_\alpha) + 0.105(^2\kappa_\alpha) + 0.0318(^3\kappa_\alpha) - 1.570$$
$$r = 0.986, \quad s = 0.14, \quad n = 64$$

where y_1 is the heat of evaporation of alkanes (kJ/mol); y_2 is the mole fraction solubility of alcohols in water, $-\log x$; y_3 is the partition coefficient in an octanol–water system ($\log P$) for oxygen-containing compounds (15 ketones, 16 unsaturated alcohols, 38 saturated alcohols, 5 carboxylic acids). Experimental data were taken from refs[19-21].

(1) Let us find all alkanes with six carbon atoms for which $28 \le y_1 \le 30$. There are 35 partitions of the number $n = 6$ into the sum of four natural numbers (step 1). The conditions for existence of the graph are met only for the following sets

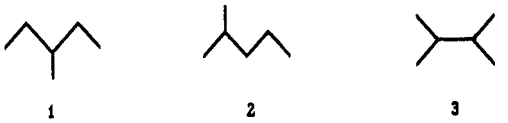

Figure 2. Structural formulas of generated alkanes for which the estimated value of the heat of evaporation is $28 \le y_1^{est} \le 30$ kJ/mol. The experimental values of the heat of evaporation y_1^{exp} for these compounds are (1) 30.27, (2) 29.86, and (3) 29.12 kJ/mol.

Table II. Minimal $^{min}d_i$ and Maximal $^{max}d_i$ Numbers of Atoms of Different Types in Compounds of Class Considered in Example 3

i	atom type	$^{min}d_i$	$^{max}d_i$
1		0	2
2		0	2
3	—C—	0	8
4	C—	1	6
5		0	1
6		0	1
7	=C	0	1
8	—O	0	1
9	=O	0	1

of $\{d_1, d_2, d_3, d_4\}$: (1) {2, 4, 0, 0}, (2) {3, 2, 1, 0}, (3) {4, 0, 2, 0} (step 2). For these sets we calculate $^0\kappa$, ^1P, ^2P (step 3); then by using the proper correlation and condition $28 \le y_1 \le 30$, we find ^3P: (1) $^3P = 2$; (2) $^3P = 3, 4$; (3) $^3P = 3, 4$ (step 4). After this we obtain sets $\{n_{kl}\}$ for each case and check all criteria from the algorithm (steps 5, 6). The vertex degree distribution (2, 4, 0, 0) must be rejected since the system of equations (step 5)

$$\begin{cases} {}^3P = n_{22} + 2n_{23} + 3n_{24} + 4n_{33} + 6n_{34} + 9n_{44} \\ {}^1P = n_{12} + n_{13} + n_{14} + n_{22} + n_{23} + n_{24} + n_{33} + n_{34} + n_{44} \end{cases}$$

does not have solutions $\{n_{ij}\}$ (for $^3P = 2, {}^1P = 5$). The remaining

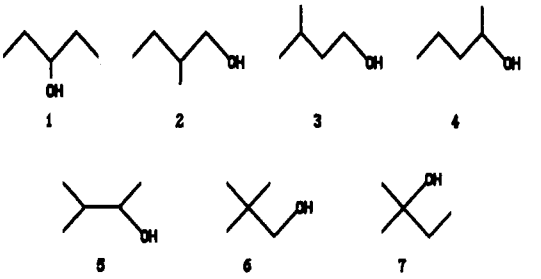

Figure 3. Structural formulas of generated alcohols for which the estimated value of solubility in water is $1.6 \le y_2^{est} \le 2.4$. The experimental values of solubility in water y_2^{exp} for these compounds are (1) 1.961, (2) 2.207, (3) 2.254, (4) 2.025, (5) 1.926, (6) 2.03, and (7) 1.608.

distributions of the vertex degrees 2 and 3 and of types of edges are used as input data for structure generator. As a result we come to structures corresponding to carbon skeletons of 3-methylpentane, 2-methylpentane, and 2,3-dimethylbutane for which $28 \le y_1 \le 30$ (Figure 2). These results are in good agreement with the experimental data: the experimental values of the heat of evaporation y_1^{exp} for these compounds are 30.27, 29.86, and 29.12 kJ/mol, respectively.

(2) Let us consider the second example. Let us find all saturated alcohols with one hydroxy group and 5 carbon atoms for which $1.6 \le y_2 \le 2.4$. The structural formulas generated by computer are given in Figure 3. The experimental values of solubility y_2^{exp} of these compounds are, respectively, (1) 1.961, (2) 2.207, (3) 2.254, (4) 2.025, (5) 1.926, (6) 2.03, and (7) 1.608. It should be mentioned that there are no other structures of this class satisfying all restrictions on the number of atoms and on the value of property y_2; therefore, all structures to be found are given in Figure 3.

(3) Let us consider the next example. The class of compounds is described by the following way: it consists of acyclic structures for which the number of oxygen atoms is equal to 1 or 2 and the numbers d_i of different atom types i ($1 \le i \le 9$) lie within limits $^{min}d_i \le d_i \le {}^{max}d_i$, shown in Table II. Besides, there are restrictions on the bond types: the

Table III. Maximal Numbers $^{max}c_j$ ($1 \le j \le 45$) of Bonds of Different Types in Compounds of Class Considered in Example 3

			—C—	—C		=C—	=C	—O	=O
	1	1	3	3	1	0	0	1	0
		2	2	2	1	1	0	1	0
—C—			7	3	2	1	0	1	0
—C				0	2	1	0	1	0
					0	0	0	1	1
=C—						0	1	0	0
=C							0	0	0
—O								0	0
=O									0

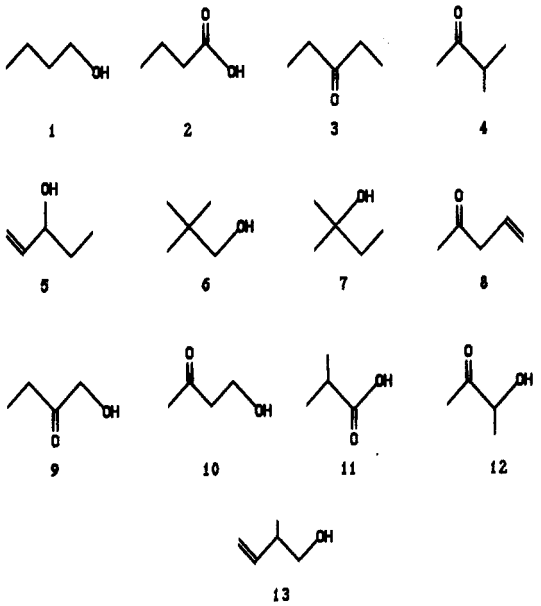

Figure 4. Structural formulas of generated oxygen-containing compounds for which the estimated value of log $P = y_3$ is $0.74 \leq y_3^{est} \leq 0.94$. The experimental values of log P for compounds **1–7** are (1) 0.88, (2) 0.79, (3) 0.79, (4) 0.59, (5) 0.81, (6) 1.36, (7) 0.89; for compounds **8–13** experimental values were not available.

numbers c_j ($1 \leq j \leq 45$) of bond types j lie within the limits $0 \leq c_j \leq {}^{max}c_j$; ${}^{max}c_j$ are given in Table III. The bond type in Table III is defined by the atom types from the corresponding column and line.

These conditions are deduced from the analysis of the training set of compounds. They provide the similarity of the training set and the class of compounds, in which the inverse problem is considered. Suppose that it is necessary to find all structures of the described class for which $0.74 \leq y_3 \leq 0.94$. The computer has generated 13 structures given in Figure 4, which yield the solution of inverse problem. The experimental values of log P for compounds **1–7**, included in the training set, are equal to 0.88, 0.79, 0.79, 0.59, 0.81, 1.36, and 0.89, respectively. For other compounds we do not have experimental data. It is seen that two experimental values do not belong to interval (0.74, 0.94). This fact is connected with the quality of the regression model including κ-indices.

CONCLUSION

In the present paper further important steps in the development of the inverse problem methodology have been made. The methodology proposed allows to reconstruct molecular graphs from the values of one more class of indices, the Kier κ-indices of molecular shape. This methodology has been also advanced for the application to the structures containing heteroatoms and multiple bonds. Moreover, the solution of the inverse problem for the case of multiple regression equations has been considered for the first time. The group of criteria providing the similarity of the generated structures to the structures from the training set has been elaborated.

REFERENCES AND NOTES

(1) Seybold, P. G.; May, M.; Bagal, U. A. Molecular Structure-Property Relationships. *J. Chem. Educ.* **1987**, *64*, 575–581.
(2) Baskin, I. I.; Gordeeva, E. V.; Devdariani, R. O.; Zefirov, N. S.; Palyulin, V. A.; Stankevich, M. I. Methodology for Solving the Inverse Problem of Structure-Property Relationships for the Case of Topological Indexes. *Dokl. Akad. Nauk SSSR* **1989**, *307*, 613–617.
(3) Skvortsova, M. I.; Stankevich, I. V.; Zefirov, N. S. Generation of Molecular Structures of Polycondensed Benzenoid Hydrocarbons from Randic Index. *Zh. Strukt. Khim.* **1992**, *33* (3), 99–104.
(4) Zefirov, N. S.; Palyulin, V. A.; Radchenko, E. V. Problem of Generation of Structures with Definite Properties. Solution of Inverse Problem for Balaban Centric Index. *Dokl. Akad. Nauk SSSR* **1991**, *316*, 921–924.
(5) Skvortsova, M. I.; Stankevich, I. V.; Zefirov, N. S. Topological Properties of Katacondensed Benzenoid Hydrocarbons: Randic Index and Its Relation to Chemical Structure. In *Proceedings of the Conference "Molecular Graphs in Chemical Studies"*; Kalinin State University: Kalinin, USSR, 1990; p 84 (in Russian).
(6) Gordeeva, E. V.; Molchanova, M. S.; Zefirov, N. S. 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 Comput. Methodol.* **1990**, *3*, 389–415.
(7) Stankevich, M. I.; Stankevich, I. V.; Zefirov, N. S. Topological Indexes in Organic Chemistry. *Usp. Khim.* **1988**, *57*, 337–366.
(8) Rouvray, D. H. Should We Have Designs on Topological Indexes? *Stud. Phys. Theor. Chem.* **1983**, *28* (*Chem. Appl. Topol. Graph Theory*), 159–177.
(9) Balaban, A. Chemical Graphs. XXXIV. Five New Topological Indices for the Branching of Tree-like Graphs. *Theor. Chim. Acta* **1979**, *53*, 355–375.
(10) Randic, M. Resolution of Ambiguities in Structure-Property Studies by Use of Orthogonal Descriptors. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 311–320.
(11) Kvasnicka, V.; Pospichal, J. Canonical Indexing and Constructive Enumeration of Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 99–105.
(12) Randic, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.
(13) 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.
(14) (a) Kier, L. B. A Shape Index from Molecular Graphs. *Quant. Struct.-Act. Relat.* **1985**, *4*, 109–116. (b) Kier, L. B. Shape Indexes of Orders One and Three from Molecular Graphs. *Quant. Struct.-Act. Relat.* **1986**, *5*, 1–7. (c) Kier, L. B. Inclusion of Symmetry as a Shape Attribute in Kappa Index Analysis. *Quant. Struct.-Act. Relat.* **1987**, *6*, 8–12.
(15) Kier, L. B. Indexes of Molecular Shape from Chemical Graphs. *Med. Res. Rev.* **1987**, *7*, 417–440.
(16) Harary, F. *Graph Theory*; Addison-Wesley: Reading, MA, 1969.
(17) Faradzhev, I. A. Generation of Nonisomorphic Graphs with Given Partition of Vertex Degrees. In *Algorithmic Studies in Combinatorics*; Faradzhev, I. A., Ed.; Nauka: Moscow, 1978; pp 11–19 (in Russian).
(18) Molodtsov, S. G.; Piottukh-Peletskii, V. N. Generation of All Nonisomorphic Chemical Graphs from a Given Set of Structural Fragments. *Vychisl. Sist.* **1984**, *103*, 51–58 (in Russian).
(19) Needham, D. E.; Wei, I. C.; Seybold, P. G. Molecular Modeling of the Physical Properties of the Alkanes. *J. Am. Chem. Soc.* **1988**, *110*, 4186–4194.
(20) Cammarata, A. Molecular Topology and Aqueous Solubility of Aliphatic Alcohols. *J. Pharm. Sci.* **1979**, *68*, 839–842.
(21) Murray, W. J.; Hall, L. H.; Kier, L. B. Molecular Connectivity III: Relation to Partition Coefficients. *J. Pharm. Sci.* **1975**, *64*, 1978–1981.