Russian Chemical Bulletin, International Edition, Vol. 56, No. 9, pp. 1681—1693, September, 2007

Full Articles

Octane numbers (ONs) of hydrocarbons: a QSPR study using optimal topological indices for the topological equivalents of the ONs*

E. A. Smolenskii,★ A. N. Ryzhov, V. M. Bavykin, T. N. Myshenkova, and A. L. Lapidus

N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences,
47 Leninskiy prosp. 119991 Moscow, Russian Federation.
E-mail: smolensk@ioc.ac.ru

A new method proposed for solving QSPR tasks is based on transition from numerical values to topological equivalents (TEs) of physicochemical properties of chemical compounds. The TEs are unambiguously related to corresponding properties; for *n*-alkanes, they are linear functions of the number, *n*, of carbon atoms. Since the TE depends only on the corresponding physicochemical parameter, it can be calculated for any hydrocarbon using the same relationships as those known for *n*-alkanes. The optimal topological index (OTI) constructed using the chemical structure matrix for TEs usually has a much smaller basis compared to the topological index obtained by analogous procedure for the physicochemical property. An algorithm for modeling of physicochemical properties using the TEs was developed and evaluated taking the octane numbers of alkanes and cycloalkanes as examples.

Key words: hydrocarbons, alkanes, cycloalkanes, descriptors, indices topological, chemical structure matrix, optimal topological indices, structure—property relations, topological equivalents, physicochemical properties, octane numbers.

Modelling of physicochemical properties of hydrocarbons is an important problem in theoretical chemistry. Usually, in constructing such models each molecule is assigned to a certain numerical parameter called descriptor. Various topological indices, which are as a rule used as descriptors,[1] can be treated as molecular graph invariants constructed ignoring hydrogen atoms.[2] Any topological index and any physicochemical property can be represented by a vector in an N-dimensional linear space (N is the number of compounds in the set).[3] An example of topological indices is provided by the quantities

$$\mathbf{g}_j = [g_j]_i, \qquad (1)$$

where $[g_j]_i$ is the number of different subgraphs g_j in the graph g_i corresponding to the ith alkane ($1 \le i, j \le N$). The

---

* Dedicated to Academician G. A. Abakumov on the occasion of his 70th birthday.

Published in Russian in *Izvestiya Akademii Nauk. Seriya Khimicheskaya,* No. 9, pp. 1619—1631, September, 2007.

numbers $[g_j]_i$ are components of the chemical structure matrix (CSM)

$$S = \|s_{ij}\|, \qquad (2)$$

whose columns represent the vectors $\mathbf{g}_j$: $[g_j]_i = s_{ij}$. The set under study, which includes N compounds, can be ordered in such a manner that the graph of the jth compound can not be a subgraph of the ith compound at $i > j$.[3] In this case the matrix S becomes a triangular matrix with the principal diagonal filled with unities, because any graph is the sole subgraph of itself. Therefore, $\det(S) = 1$ and, hence, the vectors $\mathbf{g}_j$ form a basis of an N-dimensional linear space; any topological index and property **P** can be expanded over this basis:

$$\mathbf{P} = \sum_{j=1}^{N} a_j \mathbf{g}_j, \qquad (3)$$

where a_j are the expansion coefficients.

It follows, in particular, that the number of the known TIs much exceeds that of linearly independent TIs. By analogy with **P** the coefficients a_j can be treated as a vector **A**. Then, one can write the following matrix equation for these vectors:

$$\mathbf{P} = S\mathbf{A}. \qquad (4)$$

When considering topological indices, the following is valid for coefficients a_j in expression (3). As N increases, the number of nonzero coefficients a_j either increases to a finite limiting value (the case for the Randič index) or tends to infinity but always remains much smaller than N (Wiener and Hosoya indices); the coefficients a_j depend on j or n in an *a priori* known, simple manner (complexity index: $a_j \equiv 1$; Wiener index: $a_j \equiv n$ for non-branched subgraphs and $a_j \equiv 0$ otherwise).[3] This is not true for the physicochemical properties, *e.g.*, for the boiling points (T_b) of alkanes almost all coefficients a_j differ from zero. In this case one can neglect "sufficiently small" coefficients a_j, which makes it possible to construct models for prediction of physicochemical properties with preset accuracy.[4] The index

$$\mathbf{I}_P(\varepsilon) = \sum_{j=1}^{N} b_j(\varepsilon)\mathbf{g} \qquad (5)$$

is called the optimal topological index (OTI) of the property P to the accuracy of ε;

$$b_j = \begin{cases} a_j, & \text{if } |a_j| \ge \varepsilon \\ 0, & \text{if } |a_j| < \varepsilon. \end{cases}$$

The OTIs were used for successful prediction of T_b values of alkane isomers having $n = 7$ and 8 carbon atoms using the training set including alkanes with $n < 7$.[4]

Linear prediction theorem

It was at least equally important to study how the "sufficiently large" coefficients a_j behave as N increases. Indeed, if for a certain model the number of such coefficients increases without limit and their values vary randomly, this model seems to be unsuitable for prediction. In this connection we will prove a theorem that makes it possible to formulate a necessary (but, probably, not sufficient) condition for applicability of the optimal topological indices in mathematical modeling of the properties of hydrocarbons.

Linear prediction theorem. If the training set includes the values of a parameter (characteristic) of n-alkanes with $n \le n_0$, and the test set includes the values of the same parameter (characteristic) of the n-alkanes with $n > n_0$, the values of this parameter of the alkanes from the test set, calculated using the optimal topological indices are linear functions of n.

Proof. Let g_n be a molecular graph of an n-alkane C_nH_{2n+2}. Since

$$[g_j]_n = n - j + 1 \text{ at } j \le n \qquad (6)$$

and

$$[g_j]_n = 0 \text{ at } j > n$$

(this follows from the definition of $[g_j]_n$), then if the alkane C_nH_{2n+2} is included in the test set, the corresponding optimal topological index $I_P(n, n_0, \varepsilon)$ can be calculated using the relation

$$I_P(n, n_0, \varepsilon) = \sum_{j=1}^{n_0} b_j(\varepsilon)[g_j]_n = \sum_{j=1}^{n_0}(n - j + 1)b_j(\varepsilon) = \left[n_0 \sum_{j=1}^{n_0} b_j(\varepsilon)\right]n + \left[\sum_{j=1}^{n_0}(1 - j)b_j(\varepsilon)\right]. \qquad (7)$$

The terms in square brackets on the right side of expression (7) depend only on the parameters n_0 and ε.

Let us denote

$$a = \left[n_0 \sum_{j=1}^{n_0} b_j(\varepsilon)\right], \quad b = \left[\sum_{j=1}^{n_0}(1 - j)b_j(\varepsilon)\right];$$

then one gets

$$I_P(n, n_0, \varepsilon) = an + b.$$

The theorem is proved.

Consequence. The optimal topological indices can be used for mathematical modeling of physicochemical properties of normal and branched alkanes only if for n-alkanes C_nH_{2n+2} there exists $n = n_0$ such that at $n > n_0$ one has $P(n + 1) - P(n) = \text{const}$ ($P(i)$ is the value of the physicochemical property of n-alkane C_iH_{2i+2}). Otherwise, the error in prediction (extrapolation) will not decrease with increasing n.

For instance, prediction of T_b for n-$C_{40}H_{82}$ at 760 Torr (extrapolation of the dependence of corresponding optimal topological index at $n_0 = 8$ and $\varepsilon = 0$ to $n = 40$) using the training set including eight most light n-alkanes gives a value of 997 °C (*cf.* experimental value of 540 °C). The error is 457 °C.

Nevertheless, the optimal topological indices can be used in this case in the framework of an approach described below.

Method of topological equivalents

Consider a structure—property relation for the octane numbers (ONs) of n-alkanes[5,6] (set S_1; hereafter denoted as $P_{n,\mathrm{exp}}$) the number of C—C bonds in the alkane molecule using as descriptor, $n_1 = n - 1$. The ON(n_1) dependence can be correctly approximated by a branch of a hyperbola. The hyperbola is unambiguously defined by

1) coordinates of the intersection point of asymptotes (n_0, P_0);

2) angle α between the abscissa axis and the real axis of hyperbola (hereafter the angle is measured with respect to the real axis of hyperbola);

3) lengths of the real and imaginary axes ($2a$ and $2b$, respectively).

Parameters of the equation of a second-order curve equation

$$An_1^2 + BnP_{n,\mathrm{calc}} + CP_{n,\mathrm{calc}}^2 + Dn + EP_{n,\mathrm{calc}} + F = 0 \quad (8)$$

($P_{n,\mathrm{calc}}$ is the calculated value of the ON of alkane C_nH_{2n+2}), which describes the hyperbola in the general case, can be determined to an accuracy of an arbitrary nonzero factor M using the a, b, n_0, P_0, and α values and the following relations

$$A = M[(\cos^2\alpha)/a^2 - (\sin^2\alpha)/b^2], \quad (9)$$

$$B = -2M(1/a^2 + 1/b^2)\sin\alpha\cos\alpha, \quad (10)$$

$$C = M(\sin^2\alpha/a^2 - \cos^2\alpha/b^2), \quad (11)$$

$$D = -(BP_0 + 2An_0), \quad (12)$$

$$E = -(Bn_0 + 2CP_0), \quad (13)$$

$$F = An_0^2 + Bn_0P_0 + CP_0^2 - M \quad (14)$$

The desired branch of the hyperbola is given by

$$P_{n,\mathrm{calc}} = [-Bn_1 - E - \sqrt{(Bn_1 + E)^2 - 4C(An_1^2 + Dn_1 + F)}\,]/(2C), \quad (15)$$

$$n_1 = [-BP_{n,\mathrm{calc}} - D - \sqrt{(BP_{n,\mathrm{calc}} + D)^2 - 4A(CP_{n,\mathrm{calc}}^2 + EP_{n,\mathrm{calc}} + F)}\,]/(2A). \quad (16)$$

Table 1. Modeling of ONs for n-alkanes*

n_1	P_{exp}	P_{calc}	$\lvert P_{\mathrm{exp}} - P_{\mathrm{calc}}\rvert$
0	107.5	107.3	0.2
1	107.1	107.2	0.1
2	105.7	106.7	1.0
3	93.6	93.1	0.5
4	61.8	62.4	0.6
5	31.0	31.1	0.1
6	0.0	–0.3	0.3

* $R^2 = 0.99984$, $s = 0.49603$, $\lvert\Delta_{\mathrm{max}}\rvert = 1.0$, $a = 0.9702$, $b = 1.0016$, $n_0 = 2.5756$, $P_0 = 107.5061$, $\alpha = -0.8013$.

The results of modeling are listed in Table 1 and shown in Fig. 1.

Now we will transform expression (16) in such a manner that its left part be a linear function of n_1 and independent of $P_{\mathrm{n,calc}}$:

$$2An_1 + D = -BP_{n,\mathrm{calc}} - \sqrt{(BP_{n,\mathrm{calc}} + D)^2 - 4A(CP_{n,\mathrm{calc}}^2 + EP_{n,\mathrm{calc}} + F)}\,. \quad (17)$$

Let us introduce a quantity

$$n_{P,\mathrm{exp}}(g_i) = -BP_{\mathrm{exp}} - \sqrt{(BP_{n,\mathrm{exp}} + D)^2 - 4A(CP_{n,\mathrm{exp}}^2 + EP_{n,\mathrm{exp}} + F)}\,, \quad (18)$$

obtained by replacing the calculated ON values in the right side of expression (17) by the experimental one, and call it the topological equivalent of the ON. The topological equivalent of any other physicochemical property can be defined by analogy, if there is one-to-one correspondence between the magnitude of the property and n_1, *i.e.*, the function $P(n_1)$ has no extrema. In our case the topological equivalent only depends on the ON, but unlike the ON it can well be approximated by a linear function of n_1, as follows from expressions (17) and (18) (this also holds for the topological equivalents of any other physicochemi-

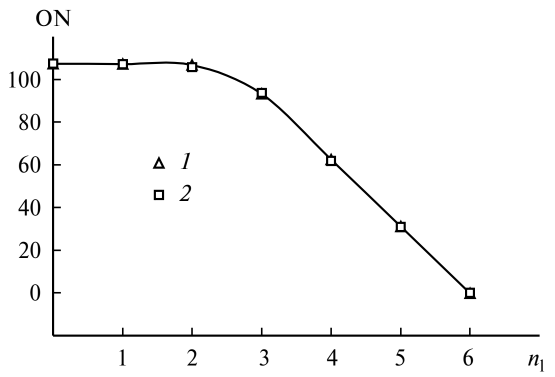

Fig. 1. Modeling of ONs for n-alkanes: experimental data (*1*) and results of modeling (*2*).

Table 2. Modeling of topological equivalents of ONs for *n*-alkanes (set S_1)*

n-Alkane	$-n_{P,\exp}$	$-n_{P,\mathrm{calc}}$	$\lvert n_{P,\exp} - n_{P,\mathrm{calc}}\rvert$
Methane	70.35693	70.35686	0.00007
Ethane	70.28866	70.28826	0.00040
Propane	70.21862	70.21966	0.00104
Butane	70.15361	70.15104	0.00257
Pentane	70.08021	70.08246	0.00225
Hexane	70.01312	70.01386	0.00074
Heptane	69.94612	69.94526	0.00086

* $R^2 = 0.99989$, $S = 0.00142$, $\lvert\Delta_{\max}\rvert = 0.00257$, $a = 1.68444$, $b = 1.83242$, $n^0 = 3.62115$, $P^0 = 107.94588$, $\alpha = -0.77465$

cal properties). Of course, in this case the parameters determined by the least squares method will differ from those listed in Table 1.

Since $n - n_1 = 1$, the theoretical topological equivalents, $n_{P,\mathrm{calc}}(g_i)$, of ONs (they are equal to the left part of Eq. (17)) can be transformed as follows:

$$n_{P,\mathrm{calc}}(g_i) = 2An_1 + D = 2An_1 + D(n - n_1) =$$
$$= Dn + (2A - D)n_1 = I_1. \quad (19)$$

From Eq. (19) it follows that the topological equivalent can be expanded over only two column- vectors of the CSM, which correspond to the methane and ethane graphs. Thus, for *n*-alkanes the quantity $n_{P,\mathrm{calc}}$ can be treated as a linear function of *n*. This holds for the topological equivalents of all properties (not only for the ON).

The results of modeling of the dependence of $n_{P,\exp}(g_i)$ on $n_{P,\mathrm{calc}}(g_i)$ for *n*-alkanes are listed in Table 2 and shown in Fig. 2.

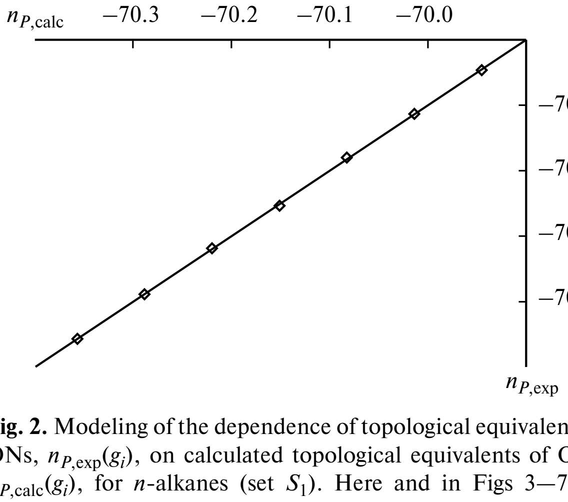

Fig. 2. Modeling of the dependence of topological equivalents of ONs, $n_{P,\exp}(g_i)$, on calculated topological equivalents of ONs, $n_{P,\mathrm{calc}}(g_i)$, for *n*-alkanes (set S_1). Here and in Figs 3—7 the straight line corresponds to equal calculated and experimental data.

Now we will express n_1 through $n_{P,\mathrm{calc}}$ using Eq. (19):

$$n_1 = [n_{P,\mathrm{calc}}(g_i) - D]/(2A), \quad (20)$$

and then substitute in relation (15). One gets

$$P_{n,\mathrm{calc}} = \left\{-B\frac{n_{P,\mathrm{calc}}(g_i) - D}{2A} - E - \right.$$
$$\left. -\sqrt{\left[B\frac{n_{P,\mathrm{calc}}(g_i) - D}{2A} + E\right]^2 - C\left[\frac{n^2_{P,\mathrm{calc}}(g_i) - D^2}{A}\right] + 4F}\right\}/(2C). \quad (21)$$

Method of topological equivalents: application for alkanes and cycloalkanes

Relation (18) has sense for all hydrocarbons rather than *n*-alkanes only, because the right side of this expression depends only on the ONs. This also holds for all physicochemical properties and their topological equivalents. The only limitation is that the expression under the radical should be non-negative. Therefore, we will impose some additional limitations on the parameters relating $n_{P,\exp}(g_i)$ to $P_{\exp}$:

$$b > a, \quad (22)$$

$$-\mathrm{arctg}(b/a) < \alpha < -\mathrm{arctg}(a/b). \quad (23)$$

The $n_{P,\exp}(g_i)$ values can be expanded using relationship (3); however, here we face some difficulties.

1. *Determination of acyclic subgraph of a cycle-containing graph.* We will illustrate this problem taking hexane and cyclohexane as examples. Theoretically, the cyclohexane graph includes a total of six hexane subgraphs having the same number of vertices, but this is impossible for alkanes (in this case, the graph does not include such subgraphs at all). This requires a rule to be introduced, namely, if selection of acyclic subgraphs of a cycle-containing graph requires break of one and only one cycle-constituting edge, these subgraphs will be ignored.

2. *Geometric inhomogeneity.* According to preliminary calculations, almost all a_j values in the expansion for branched alkanes were rather large; therefore, neglect of any coefficient a_j introduced large errors in calculations. We assumed that the geometric parameters of the branched alkane molecules are usually considerably different from those calculated in the "staggered" conformation approximation (d(C—C) = 1.54 Å; d(C—H) = 1.09 Å; the C—C—C, C—C—H, and H—C—H angles are 109°28′; and the C—C—C—C, C—C—C—H, and H—C—C—H angles ∈ (60°, 180°, 300°)), distortions for the atoms and bonds near the tertiary and quaternary carbon atoms being most pronounced. Cycloalkanes significantly differ from *n*-alkanes in that they do not adopt certain conformations. Therefore, the entire set

Table 3. Modeling of topological equivalents of ONs for cyclopentanes (set S_5)*

Cycloalkane	$-n_{P,\mathrm{exp}}$	$-n_{P,\mathrm{calc}}$	$\lvert n_{P,\mathrm{exp}} - n_{P,\mathrm{calc}}\rvert$
Cyclopentane	70.18007	70.17526	0.00481
Methylcyclopentane	70.14267	70.14520	0.00253
Ethylcyclopentane	70.09211	70.09582	0.00371
1,1-Dimethylcyclopentane	70.15020	70.15560	0.00540
1,3-Dimethylcyclopentane	70.12057	70.11514	0.00543
n-Propylcyclopentane	70.01355	70.01653	0.00298
Isopropylcyclopentane	70.12332	70.12603	0.00271
1-Methyl-2-ethylcyclopentane	70.07099	70.06576	0.00523
1,1,3-Trimethylcyclopentane	70.12470	70.12554	0.00084
1,2,3-Trimethylcyclopentane	70.14242	70.14536	0.00294
Isobutylcyclopentane	70.01832	70.01683	0.00149
1,1,2,4-Tetramethylcyclopentane	70.16082	70.15666	0.00481

* The coefficients a_j for cyclopentane, *n*-propylcyclopentane, methane, 2-methylpropane, 3-methylpentane, and 2,2-dimethylbutane molecules corresponding to the subgraph were −70.7235, −0.03037, 0.10966, −0.07960, −0.03014, and 0.05937, respectively; $R^2 = 0.99442$, $S = 0.00381$, $|\Delta_{\mathrm{max}}| = 0.00543$.

including 46 branched and 32 cyclic alkanes was divided into subclasses in accordance with the number of Me groups in the molecules, namely, 7 alkanes having two Me groups (set S_1); 11 alkanes having three Me groups (set S_2); 17 alkanes having four Me groups (set S_3); 11 alkanes having five and six Me groups (set S_4); 12 cyclopentanes (set S_5); and 20 cyclohexanes (set S_6).

The division of cycloalkanes into two sets is due to the fact that the graph of any cyclohexane includes no cyclopentane graphs and *vice versa*. Note that the division of alkanes into subclasses with respect to the number of Me groups was proposed more than four decades ago.[7] The results of modeling for cyclopentanes are listed in Table 3 and shown in Fig. 3.

The quantities $n_{P,\mathrm{exp}}(g_i)$ were expanded only for cycloalkanes. In other cases for the set S_i we studied the difference between $n_{P,\mathrm{exp}}(g_i)$ and $n_{P,\mathrm{calc}}(g_i)$ values calculated using the model for the alkanes included in the S_{i-1} set.

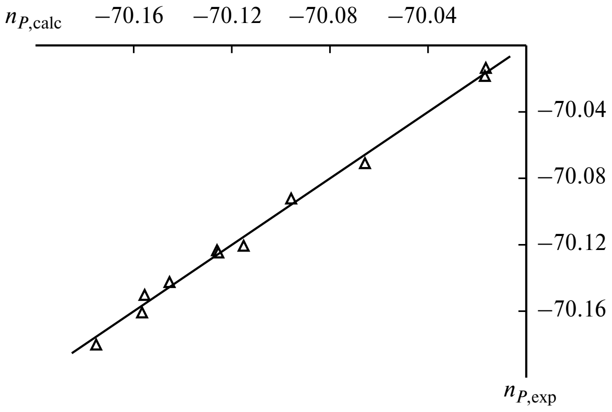

Fig. 3. Modeling of topological equivalents of ONs for cyclopentanes (set S_5).

3. *Incompleteness of measurements*. Not all subgraphs of the molecular graphs of the compounds included in the sets used in this work correspond to molecules of compounds with measured ON. For instance, the ON of 3,3,5-trimethylheptane was measured (86.4), whereas for 2,4,4-trimethylhexane this was not done. The problem is complicated by the division of the set, because the molecular graph of a branched alkane having *N* methyl groups *a priori* includes the graphs of the alkanes having 1, 2, ..., $N-1$ Me groups, as well as the methane graphs. In this case the number of variables exceeds the number of equations, which precludes obtaining of unambiguous solution.

Therefore, the following algorithm can be proposed. From the entire set of vectors (number of vectors is N_v and the number of compounds is N_c, $N_\mathrm{v} \geq N_\mathrm{c}$) we will consequently form subsets including 1, 2, ..., $N_\mathrm{c} - 1$ vectors (number of vectors in the subsets is denoted by N). For the subsets including the same number of vectors we will perform least-squares calculations of parameters using the following model equations

$$\mathbf{n}_{P,\mathrm{calc}}(g_i) = \sum_{j=1}^{N} a_j \mathbf{g}_j \qquad (24)$$

for the S_5 and S_6 sets and

$$\mathbf{n}_{P,\mathrm{calc}}(g_i)|_{S_i} - \mathbf{n}_{P,\mathrm{calc}}(g_i)|_{S_{i-1}} = \sum_{j=1}^{N} a_j \mathbf{g}_j \qquad (25)$$

for the S_2, S_3, and S_4 sets. Clearly, the accuracy of calculations expressed through $\max(|n_{P,\mathrm{calc}}(g_i) - n_{P,\mathrm{exp}}(g_i)|)$ and much exceeding the analogous quantity for the S_1 set used to determine $n_{P,\exp}(g_i)$, seems to be of no sense. Now we will introduce the quantities

$$R^2_{i,\text{opt}} = 1 - \frac{\max\left\{\left[n_{P,\text{calc}}(g_i) - n_{P,\exp}(g_i)\right]^2\right\}\Big|_{S_1}}{D\left[n_{P,\exp}(g_i)\right]\Big|_{S_i}} \qquad (26)$$

for the S_5 and S_6 sets and

$$R^2_{i,\text{opt}} = 1 - \frac{\max\left\{\left[n_{P,\text{calc}}(g_i) - n_{P,\exp}(g_i)\right]^2\right\}\Big|_{S_1}}{D\left[n_{P,\exp}(g_i) - n_{P,\text{calc}}(g_i)\Big|_{S_{i-1}}\right]\Big|_{S_i}} \qquad (27)$$

for the S_2, S_3, and S_4 sets (D is dispersion). If for the set of subsets including N vectors one gets $\max(R^2) < R_{i,\text{opt}}{}^2$, the subsets including $N + 1$ vectors should be investigated. If $\max(R^2) \geq R_{i,\text{opt}}{}^2$, the expansion over the N-vector basis corresponding to the maximum value of correlation coefficient R^2 is the optimal topological index for the property under study. The optimal topological index obtained for the modeled property of the compounds included in the S_i set will be denoted I_i.

4. *Linear dependence of vectors*. A classical example of this dependence is provided by the numbers of occurrence of the molecular graphs of methane (number of carbon atoms) and ethane (number of C—C bonds) in the molecular graphs of cycloalkanes (these numbers are equal to each other). In such cases the optimal topological indices differing only in these vectors will have the same statistical characteristics (squared correlation coefficient, mean standard deviation, and maximum deviation) and it is impossible to determine which model is characterized by better optimality. One can calculate the differences $n_{P,\exp}(g_i)|_{S_i} - n_{P,\text{calc}}(g_i)|_{S_{i-1}}$ for all optimal topological indices for the set S_{i-1} and choose the optimal topological indices that provide a minimum dispersion for the property in question. But this is not a common approach, because in the example given above there is only one solution, namely, the models with one (arbitrary) vector should be excluded.

Construction of a general model

Now we will construct a general model by combining the partial models for the topological equivalents of the ONs based on the optimal topological indices constructed for the S_1—S_6 sets. To this end, we will proceed as follows. Each set, S_5 and S_6, corresponds to its own topological index $\mathbf{L}_5$ and $\mathbf{L}_6$, respectively, which is equal to the number of occurrence of the cyclopentane and cyclohexane subgraphs, respectively, in the molecular graph. These topological indices are equal to unity for the compounds included in a given set and to zero for the compounds included in other sets. For the S_i sets ($i = 2, 3, 4$) one can derive the indices L_i that are equal to zero for the compounds included in the S_j sets ($j < i$) and differ from zero for the compounds included in the S_k sets ($i \leq k \leq 4$) (at $k = i$, $L_i = 1$). The column vector for the index $\mathbf{L}_2$ corresponds to isobutane, the column vector for the index $\mathbf{L}_3$ corresponds to the sum of the vectors corresponding to neopentane, 2,3-dimethylbutane, 2,4-dimethylpentane, *etc.*, the column vector $\mathbf{L}_4$ corresponds to the sum of the vectors corresponding to 2,2,3-trimethylbutane, 2,2,4-trimethylpentane, 2,3,4-trimethylpentane, *etc.* Then the general model for all sets can be written in the form

$$\mathbf{n}_{P,\text{calc}}\Big|_j = \sum_{i=5}^{6} L_{ij} I_{ij} + \left(I_{1j} + \sum_{i=2}^{4} L_{ij} I_{ij}\right)\left(1 - \sum_{i=5}^{6} L_{ij}\right), \qquad (28)$$

where $L_{ij} = (\mathbf{L}_i)_j$, $I_{ij} = (\mathbf{I}_i)_j$.

From topological equivalents to octane numbers

Once determined the quantities $n_{P,\text{calc}}$ for all 78 hydrocarbons under study, relation (21) becomes sensible for all alkanes rather than n-alkanes only. Using this relation, we can now go from topological equivalents of the octane numbers to the octane numbers of hydrocarbons. It should be noted that expansion (19) of the I_1 index appeared in the final expression (28) was done using the parameters A and D in Eq. (8) and depends on all parameters of the hyperbola (a, b, n_0, P_0, α).

Substituting the calculated $n_{P,\text{calc}}(g_i)$ values for alkanes into expression (28) gives $R^2 = 0.974$ (model *1*). This is a rather poor result; nevertheless, it is better than those obtained earlier for the same set (*cf.* $R^2 = 0.956$,**[5]** 0.942,**[8]** 0.937,**[9]** and 0.940 (see Ref. 10)). The maximum errors were obtained for 2-methyl-3-ethylpentane ($P_{n,\text{calc}} = 105.0$, $P_{n,\exp} = 87.3$), and 3-ethyl-3-methylpentane ($P_{n,\text{calc}} = 103.5$, $P_{n,\exp} = 80.8$). It follows that expansion (25) for the indices I_3 and I_4 is sufficient, which gives $R^2 = 0.99671$ (model *2*). The largest deviations from the previously obtained values were only found for the parameters corresponding to the molecular graphs of ethylmethylpentanes, diethylpentane (in the expansion of the I_3 index), and 2,4-dimethyl-3-ethylpentane (in the expansion of the I_4 index), which was mentioned above. The use of these parameters led to $R^2 = 0.99669$ (model *3*). The model augmented with the coefficients in the expansion of the I_2 index gave $R^2 = 0.99677$ (model *4*). Finally, we obtained gives $R^2 = 0.99715$ using all coefficients (model *5*). The difference between the largest and the smallest values of the R^2 coefficients for models *2*—*5* is rather small, being only 0.00046. The results are listed in Tables 4 and 5 and shown in Fig. 4.

Substitution of the results of $n_{P,\text{calc}}$ calculations (see above) for cycloalkanes into expression (28) gives

Table 4. Modeling of ONs for alkanes using models *1—5*

Hydrocarbon	P_{exp}	*1* [a]		*2* [b]		*3* [c]		*4* [d]		*5* [e]	
		P_{calc}	$\lvert P_{exp} - P_{calc}\rvert$	P_{calc}	$\lvert P_{exp} - P_{calc}\rvert$	P_{calc}	$\lvert P_{exp} - P_{calc}\rvert$	P_{calc}	$\lvert P_{exp} - P_{calc}\rvert$	P_{calc}	$\lvert P_{exp} - P_{calc}\rvert$
Methane	107.5	107.5	0.0	107.5	0.0	107.5	0.0	107.5	0.0	106.6	0.9
Ethane	107.1	107.1	0.0	107.1	0.0	107.1	0.0	107.1	0.0	106.2	0.9
Propane	105.7	105.7	0.0	105.7	0.0	105.7	0.0	105.7	0.0	104.9	0.8
n-Butane	93.6	92.6	1.0	92.6	1.0	92.6	1.0	92.6	1.0	93.4	0.2
Methylpropane	102.1	102.7	0.6	102.7	0.6	102.7	0.6	102.9	0.8	102.3	0.2
n-Pentane	61.8	62.8	1.0	62.8	1.0	62.8	1.0	62.8	1.0	63.4	1.6
Methylbutane	93.0	92.6	0.4	92.6	0.4	92.6	0.4	93.1	0.1	93.4	0.4
Dimethylpropane	85.5	88.0	2.5	88.1	2.6	88.0	2.5	87.5	2.0	87.4	1.9
n-Hexane	31.0	31.3	0.3	31.3	0.3	31.3	0.3	31.3	0.3	31.2	0.2
2-Methylpentane	73.4	70.5	2.9	70.5	2.9	70.5	2.9	70.9	2.5	71.3	2.1
3-Methylpentane	74.5	74.5	0.0	74.5	0.0	74.5	0.0	74.3	0.2	74.8	0.3
2,2-Dimethylbutane	91.8	88.3	3.5	88.5	3.3	88.3	3.5	88.9	2.9	88.9	2.9
2,3-Dimethylbutane	104.3	103.9	0.4	103.9	0.4	103.9	0.4	104.2	0.1	103.4	0.9
n-Heptane	0.0	−0.4	0.4	−0.4	0.4	−0.4	0.4	−0.4	0.4	−1.3	1.3
2-Methylhexane	46.4	46.9	0.5	46.9	0.5	46.9	0.5	47.4	1.0	47.4	1.0
3-Methylhexane	52.0	51.0	1.0	51.0	1.0	51.0	1.0	50.8	1.2	51.0	1.0
Ethylpentane	65.0	65.0	0.0	65.0	0.0	65.0	0.0	65.0	0.0	65.0	0.0
2,2-Dimethylpentane	92.8	92.7	0.1	92.5	0.3	92.7	0.1	92.7	0.1	92.8	0.0
2,3-Dimethylpentane	91.1	93.2	2.1	93.1	2.0	93.2	2.1	93.5	2.4	93.4	2.3
2,4-Dimethylpentane	83.1	79.5	3.6	79.4	3.7	79.5	3.6	79.7	3.4	79.9	3.2
3,3-Dimethylpentane	80.8	82.0	1.2	82.2	1.4	82.0	1.2	82.4	1.6	82.4	1.6
Trimethylbutane	112.1	112.0	0.1	112.0	0.1	112.0	0.1	111.9	0.2	111.3	0.8
2-Methylheptane	21.7	23.1	1.4	23.1	1.4	23.1	1.4	23.5	1.8	23.1	1.4
3-Methylheptane	26.8	27.2	0.4	27.2	0.4	27.2	0.4	26.9	0.1	26.7	0.1
4-Methylheptane	26.7	27.2	0.5	27.2	0.5	27.2	0.5	26.9	0.2	26.7	0.0
Ethylhexane	33.5	33.5	0.0	33.5	0.0	33.5	0.0	33.5	0.0	33.5	0.0
2,2-Dimethylhexane	72.5	72.3	0.2	72.0	0.5	72.3	0.2	72.1	0.4	72.2	0.3
2,3-Dimethylhexane	71.3	72.9	1.6	72.8	1.5	72.9	1.6	72.9	1.6	72.9	1.6
2,4-Dimethylhexane	65.2	64.3	0.9	64.2	1.0	64.3	0.9	64.1	1.1	64.2	1.0
2,5-Dimethylhexane	55.5	58.0	2.5	57.8	2.3	58.0	2.5	57.9	2.4	57.9	2.4
3,3-Dimethylhexane	75.5	75.5	0.0	75.6	0.1	75.5	0.0	75.5	0.0	75.4	0.1
3,4-Dimethylhexane	76.3	75.6	0.7	75.5	0.8	75.6	0.7	74.8	1.5	74.7	1.6
2-Methyl-3-ethylpentane	87.3	105.0	17.7	87.3	0.0	87.3	0.0	87.3	0.0	87.3	0.0
3-Methyl-3-ethylpentane	80.8	103.5	22.7	80.9	0.1	80.8	0.0	80.7	0.1	80.6	0.2
2,2,3-Trimethylpentane	109.6	109.6	0.0	109.6	0.0	109.6	0.0	109.7	0.1	109.0	0.6
2,2,4-Trimethylpentane	100.0	106.2	6.2	106.2	6.2	106.2	6.2	106.2	6.2	105.5	5.5
2,3,3-Trimethylpentane	106.1	103.4	2.7	103.8	2.3	103.4	2.7	104.0	2.1	103.9	2.2
2,3,4-Trimethylpentane	102.7	102.6	0.1	103.1	0.4	102.7	0.0	103.0	0.3	103.2	0.5
Tetramethylbutane	103.0	106.7	3.7	106.8	3.8	106.7	3.7	106.9	3.9	106.0	3.0
2,2-Dimethylheptane	50.3	50.6	0.3	50.3	0.0	50.6	0.3	50.0	0.3	50.0	0.3
Diethylpentane	84.0	84.0	0.0	84.0	0.0	84.0	0.0	84.0	0.0	84.0	0.0
2,2-Dimethyl-3-ethylpentane	112.1	112.1	0.0	112.0	0.1	111.9	0.2	112.4	0.3	111.9	0.2
2,4-Dimethyl-3-ethylpentane	105.3	105.3	0.0	105.3	0.0	105.3	0.0	105.3	0.0	105.3	0.0
2,2,3,3-Tetramethylpentane	116.8	116.9	0.1	116.9	0.1	116.9	0.1	117.3	0.5	117.3	0.5
3,3,4-Trimethylheptane	86.4	86.4	0.0	86.4	0.0	86.6	0.2	86.4	0.0	86.5	0.1
2,2,3,3-Tetramethylhexane	112.8	112.7	0.1	112.8	0.0	112.7	0.1	113.3	0.5	113.0	0.2

[a] $R^2 = 0.97410$, $s = 4.54533$, $|\Delta_{max}| = 22.7$.

[b] $R^2 = 0.99671$, $s = 1.61911$, $|\Delta_{max}| = 6.2$.

[c] $R^2 = 0.99669$, $s = 1.62451$, $|\Delta_{max}| = 6.2$.

[d] $R^2 = 0.99677$, $s = 1.60438$, $|\Delta_{max}| = 6.2$, $a = 1.68444$, $b = 1.83242$, $n_0 = 3.62115$, $P_0 = 107.94588$, $\alpha = -0.77465$.

[e] $R^2 = 0.99715$, $s = 1.50696$, $|\Delta_{max}| = 5.5$, $a = 1.75745$, $b = 1.90891$, $n_0 = 3.70678$, $P_0 = 107.03371$, $\alpha = -0.77464$.

$R^2 = 0.9958$ (model *1*) (*cf.* $R^2 = 0.987$,[5] 0.992,[8] 0.937,[9] and 0.940 (see Ref. 10)). Various types of expansion of the I_5 and I_6 indices led to a minor variation of R^2 by $1 \cdot 10^{-9}$, while the use of all parameters gave $R^2 = 0.9959$ (model *2*). The results are listed in Tables 6 and 7 and shown in Fig. 5.

Table 5. Expansion coefficients for models *1—5* (alkanes)

I_i	Molecule corresponding to subgraph	Coefficients in expansion (3)				
		1	*2*	*3*	*4*	*5*
I_2	Methylpropane	−0.03485	−0.03603	−0.03091	—	—
I_2	3-Methylpentane	0.00790	0.00938	0.00870	—	—
I_2	Ethylpentane	−0.02949	−0.03317	−0.02913	—	—
I_2	Ethylhexane	0.01701	0.01706	0.01435	—	—
I_2	*n*-Butane	−0.01689	−0.01683	−0.01560	—	—
I_3	3,3-Dimethylhexane	0.02578	0.02511	0.02578	0.02534	0.02310
I_3	2-Methyl-3-ethylpentane	−0.04450	0.02389	0.02410	0.02814	0.02351
I_3	3-Methyl-3-ethylpentane	−0.06706	0.00194	0.00154	0.00302	0.00145
I_3	Methane	−0.00389	−0.00383	−0.00389	−0.00334	−0.00322
I_3	2-Methylbutane	−0.00498	−0.00502	−0.00498	−0.00605	−0.00521
I_4	2,2,3-Trimethylpentane	−2.15639	−2.15634	−2.15639	−2.33645	−2.38132
I_4	2,2-Dimethyl-3-ethylpentane	−6.11672	−6.11752	−6.11672	−6.15386	−6.16201
I_4	2,2,3,3-Tetramethylpentane	−12.47688	−12.47620	−12.47688	−12.51707	−12.52603
I_4	Methane	−37.04599	−36.95574	−37.04599	−35.97841	−35.83373
I_4	Ethane	42.33401	42.23042	42.33394	41.11382	40.94757

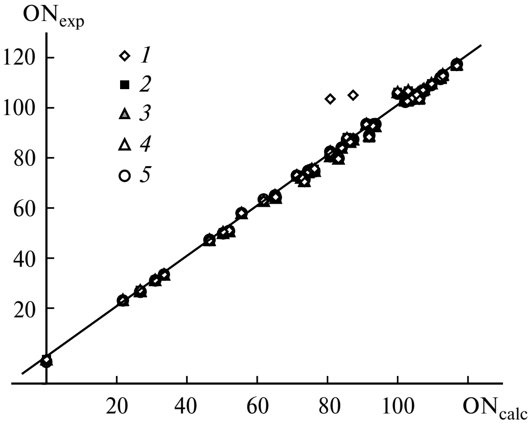

Fig. 4. Modeling of ONs for alkanes using models *1—5*.

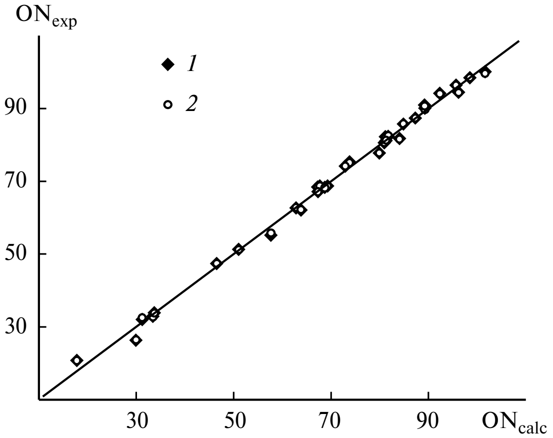

Fig. 5. Modeling of ONs for cycloalkanes using models *1* (*1*) and *2* (*2*).

The model including all hydrocarbons studied ($R^2 = 0.9966$) has the following statistical characteristics: $R^2 = 0.99660$, $s = 1.53050$, and $|\Delta_{max}| = 6.23228$; the corresponding results are shown in Fig. 6.

Construction of an extrapolation model

Now we can make an attempt to determine the parameters of the model starting from the minimum amount of input data and to predict the remaining ON values. The difficulty consists in that no experimental data on the ON are available for a number of graphs used in the model. This concerns acyclic subgraphs in the I_5 and I_6 indices and graphs with the number of vertices smaller than i for the I_i indices ($2 \le i \le 4$). Therefore, one should use the ONs of other compounds whose molecular graphs include at least one subgraph of the necessary type. The number of such compounds is of course much larger than the number of such graphs; in this connection, selection should be performed, *e.g.*, as follows. In the test set, the compound characterized by the maximum absolute value of the deviation of the calculated value from experimental one was chosen. Then, the models obtained by replacing this compound by any admissible compound included in the training set were considered. If, for a model characterized by the smallest $|\Delta_{max}|$ value, the $|\Delta_{max}|$ value was smaller than that obtained in the preceding step, the procedure was continued; otherwise, the preceding model was treated as final one. The training set included the minimum possible number of compounds (41) and the test set included 30 compounds.

We succeeded in constructing such a model, which has the following statistical characteristics: $R^2 = 0.99894$, $s = 0.82896$, and $|\Delta_{max}| = 3.0$ for the training set and $R^2 = 0.99012$, $s = 2.62038$, and $|\Delta_{max}| = 5.6$ for the test set. The results of modeling are listed in Tables 8 and 9 and

Table 6. Modeling of ONs for cycloalkanes using models *1* and *2*

Cycloalkane	P_{exp}	*1* [a]		*2* [b]	
		P_{calc}	$\lvert P_{\mathrm{exp}} - P_{\mathrm{calc}}\rvert$	P_{calc}	$\lvert P_{\mathrm{exp}} - P_{\mathrm{calc}}\rvert$
Cyclopentane	101.6	100.1	1.5	99.6	2.0
Methylcyclopentane	89.3	90.1	0.8	89.8	0.5
Ethylcyclopentane	67.2	68.4	1.2	68.5	1.3
1,1-Dimethylcyclopentane	92.3	94.1	1.8	94.0	1.7
1,3-Dimethylcyclopentane	79.9	77.7	2.2	77.7	2.2
n-Propylcyclopentane	31.2	32.1	0.9	32.5	1.3
Isopropylcyclopentane	81.1	82.2	1.1	81.9	0.8
1-Methyl-2-ethylcyclopentane	57.6	55.2	2.4	55.7	1.9
1,1,3-Trimethylcyclopentane	81.7	82.3	0.6	82.5	0.8
1,2,3-Trimethylcyclopentane	89.2	90.9	1.7	90.6	1.4
Isobutylcyclopentane	33.4	32.9	0.5	32.7	0.7
1,1,2,4-Tetramethylcyclopentane	96.2	94.5	1.7	94.4	1.8
Cyclohexane	84	81.7	2.3	81.5	2.5
Methylcyclohexane	73.8	75.3	1.5	75.2	1.4
Ethylcyclohexane	46.5	47.4	0.9	47.4	0.9
1,1-Dimethylcyclohexane	87.3	87.3	0.0	87.4	0.1
1,2-Dimethylcyclohexane	80.9	80.6	0.3	80.5	0.4
1,3-Dimethylcyclohexane	69.3	68.7	0.6	68.7	0.6
1,4-Dimethylcyclohexane	67.7	68.7	1.0	68.7	1.0
n-Propylcyclohexane	17.8	20.7	2.9	20.8	3.0
Isopropylcyclohexane	62.8	62.7	0.1	62.7	0.1
1-Methyl-1-ethylcyclohexane	68.7	68.3	0.4	68.2	0.5
1,1,2-Trimethylcyclohexane	95.7	96.4	0.7	96.4	0.7
1,1,3-Trimethylcyclohexane	81.3	81.1	0.2	81.2	0.1
1,2,3-Trimethylcyclohexane	84.8	85.7	0.9	85.8	1.0
1,2,4-Trimethylcyclohexane	72.9	74.1	1.2	74.1	1.2
1,3,5-Trimethylcyclohexane	63.8	62.1	1.7	62.1	1.7
Isobutylcyclohexane	33.7	33.9	0.2	33.8	0.1
sec-Butylcyclohexane	51	51.3	0.3	51.2	0.2
tert-Butylcyclohexane	98.5	98.4	0.1	98.5	0.0
1-Isopropyl-4-methylcyclohexane	67.3	67.2	0.1	67.2	0.1
1-Methyl-2-*n*-propylcyclohexane	29.9	26.3	3.6	26.4	3.5

[a] $R^2 = 0.9958$, $s = 1.41765$, $\lvert\Delta_{\mathrm{max}}\rvert = 3.6$, $a = 1.68444$, $b = 1.83242$, $n_0 = 3.62115$, $P_0 = 107.94588$, $\alpha = -0.77465$.

[b] $R^2 = 0.9959$, $s = 1.40723$, $\lvert\Delta_{\mathrm{max}}\rvert = 3.5$, $a = 1.68376$, $b = 1.83292$, $n_0 = 3.62115$, $P_0 = 107.93995$, $\alpha = -0.77464$.

Table 7. Expansion coefficients for models *1* and *2* (cycloalkanes)

I_i	Molecule corresponding to subgraph	Coefficients in expansion (3)	
I_5	Cyclopentane	−70.72324	−70.72521
I_5	Methane	0.10995	0.10893
I_5	2-Methylpropane	−0.08101	−0.08037
I_5	3-Methylpentane	−0.03003	−0.02967
I_5	2,2-Dimethylbutane	0.06131	0.06052
I_6	Cyclohexane	−70.12467	−70.13217
I_6	1,1-Dimethylcyclohexane	0.03418	0.03421
I_6	1,2-Dimethylcyclohexane	0.03931	0.03935
I_6	1,3-Dimethylcyclohexane	0.05785	0.05880
I_6	1,4-Dimethylcyclohexane	0.02449	0.02460
I_6	Methylpropane	−0.05331	−0.05432
I_6	1,2-Dimethylbutane	0.04796	0.04876

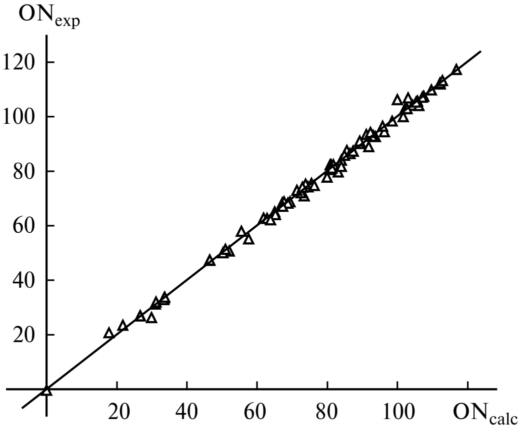

Fig. 6. Results of modeling of ONs for saturated hydrocarbons using the general model.

Table 8. Expansion coefficients for the general model*

I_i	Molecule corresponding to subgraph	Coefficients in expansion (3)
I_2	Methylpropane	−0.13768
I_2	3-Methylpentane	0.03664
I_2	Ethylpentane	−0.13510
I_2	Ethylhexane	0.05997
I_2	*n*-Butane	−0.06570
I_3	Dimethylpropane	0.43487
I_3	2,2-Dimethylpentane	−0.24809
I_3	3,3-Dimethylhexane	0.10240
I_3	2-Methyl-3-ethylpentane	0.13360
I_3	3-Methyl-3-ethylpentane	0.00408
I_3	Methane	−0.01524
I_3	2-Methylbutane	−0.02542
I_4	2,2,3-Trimethylpentane	−4.30253
I_4	2,2-Dimethyl-3-ethylpentane	−6.49473
I_4	2,4-Dimethyl-3-ethylpentane	−5.96745
I_4	2,2,3,3-Tetramethylpentane	−12.82695
I_4	Methane	−35.59253
I_4	Ethane	40.65145
I_5	Cyclopentane	−76.38207
I_5	*n*-Propylcyclopentane	−0.14130
I_5	Methane	0.45885
I_5	2-Methylpropane	−0.33128
I_5	2,2-Dimethylbutane	0.23695
I_6	Cyclohexane	−73.85972
I_6	Methylcyclohexane	0.55415
I_6	Ethylcyclohexane	0.38205
I_6	1,1-Dimethylcyclohexane	0.15148
I_6	1,4-Dimethylcyclohexane	0.10376
I_6	Methylpropane	−0.21985
I_6	Methylbutane	−0.13432
I_6	1,2-Dimethylbutane	0.19106

* $a = 1.51598$, $b = 1.90031$, $n_0 = 3.76424$, $P_0 = 105.36326$, $\alpha = -0.70399$.

Table 9. Results of odeling usinf the general model

Hydrocarbon	P_{exp}	P_{calc}	$\lvert P_{\text{exp}} - P_{\text{calc}} \rvert$
Training set[a]			
Methane	107.5	105.4	2.1
Ethane	107.1	104.9	2.2
Butane	93.6	93.7	0.1
Methylpropane	102.1	101.7	0.4
Methylbutane	93.0	93.3	0.3
Dimethylpropane	85.5	85.6	0.1
Hexane	31.0	31.7	0.7
3-Methylpentane	74.5	74.3	0.2
Heptane	0.0	−0.7	0.7
Ethylpentane	65.0	65.0	0.0
3,3-Dimethylpentane	80.8	80.9	0.1
2,2-Dimethylpentane	92.8	92.6	0.2
Trimethylbutane	112.1	109.1	3.0
Ethylhexane	33.5	33.6	0.1
3,3-Dimethylhexane	75.5	75.6	0.1
3,4-Dimethylhexane	76.3	76.3	0.0
2-Methyl-3-ethylpentane	87.3	87.2	0.1
3-Methyl-3-ethylpentane	80.8	80.8	0.0
2,2,3-Trimethylpentane	109.6	108.6	1.0
2,3,4-Trimethylpentane	102.7	103.0	0.3
Diethylpentane	84.0	84.0	0.0
2,2-Dimethyl-3-ethylpentane	112.1	112.7	0.6
2,4-Dimethyl-3-ethylpentane	105.3	105.5	0.2
2,2,3,3-Tetramethylpentane	116.8	117.4	0.6
3,3,4-Trimethylheptane	86.4	86.4	0.0
Cyclopentane	101.6	101.2	0.4
1,3-Dimethylcyclopentane	79.9	79.9	0.0
n-Propylcyclopentane	31.2	31.6	0.4
Isopropylcyclopentane	81.1	80.9	0.2
Isobutylcyclopentane	33.4	33.2	0.2
Test set[b]			
Propane	105.7	103.8	1.9
2-Methylpentane	73.4	71.0	2.4
2,2-Dimethylbutane	91.8	87.1	4.7
2,3-Dimethylbutane	104.3	102.7	1.6
Pentane	61.8	63.9	2.1
2-Methylhexane	46.4	46.7	0.3
3-Methylhexane	52	50.1	1.9
2,3-Dimethylpentane	91.1	94.5	3.4
2,4-Dimethylpentane	83.1	80.8	2.3
2-Methylheptane	21.7	22.1	0.4
3-Methylheptane	26.8	25.6	1.2
4-Methylheptane	26.7	25.6	1.1
2,2-Dimethylhexane	72.5	71.9	0.6
2,4-Dimethylhexane	65.2	65.1	0.1
2,5-Dimethylhexane	55.5	58.7	3.2
2,3-Dimethylhexane	71.3	74.3	3.0
2,3,3-Trimethylpentane	106.1	103.1	3.0
2,2,4-Trimethylpentane	100	104.4	4.4
Tetramethylbutane	103	104.8	1.8
2,2-Dimethylheptane	50.3	49.5	0.8
2,2,3,3-Tetramethylhexane	112.8	113.6	0.8
1,1-Dimethylcyclopentane	92.3	97.9	5.6
1,2,3-Trimethylcyclopentane	89.2	91.1	1.9
1-Methyl-3-ethylcyclopentane	57.6	54.0	3.6
Methylpentane	89.3	93.1	3.8
Ethylpentane	67.2	68.9	1.7
1,1,3-Trimethylcyclopentane	81.7	86.5	4.8
n-Propylcyclohexane	17.8	20.9	3.1
1,1,2-Trimethylcyclohexane	95.7	97.1	1.4
Isobutylcyclohexane	33.7	36.9	3.2

(*to be continued*)

Table 9 (*continued*)

Hydrocarbon	P_{exp}	P_{calc}	$\|P_{exp} - P_{calc}\|$
	Training set[a]		
1,1,2,4-Tetramethylcyclopentane	96.2	96.3	0.1
Cyclohexane	84.0	82.9	1.1
Methylcyclohexane	73.8	75.5	2.1
Ethylcyclohexane	46.5	46.8	2.2
1,1-Dimethylcyclohexane	87.3	87.5	0.7
1,2-Dimethylcyclohexane	80.9	81.0	0.4
1,3-Dimethylcyclohexane	69.3	68.0	0.2
1,4-Dimethylcyclohexane	67.7	68.0	0.0
Isopropylcyclohexane	62.8	62.8	0.1
1-Methyl-1-ethylcyclohexane	68.7	68.6	0.1
1-Isopropyl-3-methylcyclohexane	67.3	67.2	0.1

Hydrocarbon	P_{exp}	P_{calc}	$\|P_{exp} - P_{calc}\|$
	Test set[b]		
1,2,4-Trimethylcyclopentane	72.9	73.6	0.7
1,3,5-Trimethylcyclopentane	63.8	60.4	3.4
tert-Butylcyclohexane	98.5	99.2	0.7
sec-Butylcyclohexane	51	52.9	1.9
1,1,3-Trimethylcyclohexane	81.3	80.4	0.9
1,2,3-Trimethylcyclopentane	84.8	86.2	1.4
1-Methyl-2-*n*-propylcyclohexane	29.9	26.6	3.3

[a] $R^2 = 0.99894$, $s = 0.82896$, $|\Delta_{max}| = 3.0$.

[b] $R^2 = 0.99012$, $s = 2.62038$, $|\Delta_{max}| = 5.6$.

Table 10. Results of prediction for hydrocarbons with unknown ONs

Hydrocarbon	ON
Octane	−33.3
Nonane	−65.8
2-Methyloctane	−2.5
3-Methyloctane	0.9
4-Methyloctane	0.9
3-Ethylheptane	8.9
4-Ethylheptane	1.8
2,3-Dimethylheptane	51.9
2,4-Dimethylheptane	42.5
2,5-Dimethylheptane	42.5
2,6-Dimethylheptane	36.1
2-Methyl-3-ethylhexane	51.7
3-Methyl-3-ethylhexane	56.8
3-Methyl-4-ethylhexane	60.8
2-Methyl-4-ethylhexane	46.4
2,2,3-Trimethylhexane	29.0
2,2,4-Trimethylhexane	−476.3
2,2,5-Trimethylhexane	−512.4
2,3,3-Trimethylhexane	−498.7
2,3,4-Trimethylhexane	−502.6
2,3,5-Trimethylhexane	−509.9
2,4,4-Trimethylhexane	−491.2
3,3,4-Trimethylhexane	24.2
2,3-Dimethyl-3-ethylpentane	−490.0
2,2,3,4-Tetramethylpentane	108.2
2,2,4,4-Tetramethylpentane	−360.9
2,3,3,4-Tetramethylpentane	−471.0
Decane	−98.4
2-Methylnonane	−27.3
3-Methylnonane	−23.8
4-Methylnonane	−23.8
5-Methylnonane	−23.8
3-Ethyloctane	−15.8
4-Ethyloctane	−22.9
4-Propylheptane	−30.0
2,2-Dimethyloctane	26.8
2,3-Dimethyloctane	29.2
2,4-Dimethyloctane	19.7
2,5-Dimethyloctane	19.7
2,6-Dimethyloctane	19.7
2,7-Dimethyloctane	13.3
2-Methyl-3-ethylheptane	28.9
3-Methyl-3-ethylheptane	34.1
4-Methyl-3-ethylheptane	38.1
3-Methyl-5-ethylheptane	30.1
2-Methyl-5-ethylheptane	30.7
2-Methyl-4-ethylheptane	16.5
3-Methyl-4-ethylheptane	23.9
4-Methyl-4-ethylheptane	32.1
4-Isopropylheptane	14.7
3,3-Diethylhexane	34.0
3,4-Diethylhexane	30.8
2,2,3-Trimethylheptane	−597.0
2,2,4-Trimethylheptane	−1102.8
2,2,5-Trimethylheptane	−1132.4
2,2,6-Trimethylheptane	−1138.9
2,3,3-Trimethylheptane	−1125.2
2,3,4-Trimethylheptane	−1129.2
2,3,5-Trimethylheptane	−1130.0
2,3,6-Trimethylheptane	−1136.5
2,4,4-Trimethylheptane	−1100.3
2,4,5-Trimethylheptane	−1134.4
2,4,6-Trimethylheptane	−1146.0
2,5,5-Trimethylheptane	−1135.1
3,3,4-Trimethylheptane	−601.9
3,4,4-Trimethylheptane	−596.8
3,4,5-Trimethylheptane	−1127.1
2,2-Dimethyl-3-ethylhexane	108.8
2,3-Dimethyl-3-ethylhexane	−1137.9
2-Methyl-3-isopropylhexane	−453.9

(*to be continued*)

Table 10 (*continued*)

Hydrocarbon	ON
2,4-Dimethyl-3-ethylhexane	−437.5
2,5-Dimethyl-3-ethylhexane	−1151.1
3,4-Dimethyl-4-ethylhexane	−607.4
2,3-Dimethyl-4-ethylhexane	−1129.5
2,4-Dimethyl-4-ethylhexane	−1133.6
3,3-Dimethyl-4-ethylhexane	108.8
2,2-Dimethyl-4-ethylhexane	−1106.0
2-Methyl-3,3-diethylpentane	−1118.4
2,2,3,4-Tetramethylhexane	−31.9
2,2,3,5-Tetramethylhexane	−577.5
2,2,4,4-Tetramethylhexane	−1008.1
2,2,4,5-Tetramethylhexane	−1073.8
2,2,5,5-Tetramethylhexane	−1105.9
2,3,3,4-Tetramethylhexane	−581.3
2,3,3,5-Tetramethylhexane	−1100.6
2,3,4,4-Tetramethylhexane	−51.2
2,3,4,5-Tetramethylhexane	−1104.6
3,3,4,4-Tetramethylhexane	129.0
2,2,3-Trimethyl-3-ethylpentane	133.8
2,2,4-Trimethyl-3-ethylpentane	129.5
2,3,4-Trimethyl-3-ethylpentane	−397.0
2,4-Dimethyl-3-isopropylpentane	123.2
2,2,3,3,4-Pentamethylpentane	129.3
2,2,3,4,4-Pentamethylpentane	117.6
1,2-Dimethylcyclopentane	100.4
1-Methyl-2-ethylcyclopentane	75.5
1-Methyl-2-ethylcyclopentane	93.0
1,1,2-Trimethylcyclopentane	102.2
1,2,3-Trimethylcyclopentane	104.1
Butylcyclopentane	−22.9
Isobutylcyclopentane	71.5
1-Methyl-1-propylcyclopentane	38.5
1-Methyl-2-propylcyclopentane	57.9
1-Methyl-3-propylcyclopentane	16.5
1,1-Diethylcyclopentane	76.3
1,2-Diethylcyclopentane	82.0
1,3-Diethylcyclopentane	27.5
tert-Butylcyclopentane	102.5
1-Isopropyl-1-methylcyclopentane	86.9
1-Isopropyl-2-methylcyclopentane	103.2
1-Isopropyl-3-methylcyclopentane	66.4
1,2-Dimethyl-1-ethylcyclopentane	102.3
1,1-Dimethyl-2-ethylcyclopentane	102.4
1,2-Dimethyl-3-ethylcyclopentane	103.2
1,3-Dimethyl-4-ethylcyclopentane	79.8
1,3-Dimethyl-2-ethylcyclopentane	104.2
1,1,2,2-Tetramethylcyclopentane	91.6
1,1,2,3-Tetramethylcyclopentane	104.8
1,1,3,3-Tetramethylcyclopentane	92.6
1,1,3,4-Tetramethylcyclopentane	96.5
1,2,2,3-Tetramethylcyclopentane	104.8
1,2,3,4-Tetramethylcyclopentane	104.9
Pentylcyclopentane	−77.5
1-Cyclopentyl-3-methylbutane	−38.0
1-Cyclopentyl-2-methylbutane	−7.3
1-Cyclopentyl-1-methylbutane	17.6
1-Butyl-1-methylcyclopentane	−15.9
1-Butyl-2-methylcyclopentane	3.7
1-Butyl-3-methylcyclopentane	−38.0
1-Propyl-1-ethylcyclopentane	39.3
1-Propyl-2-ethylcyclopentane	45.3
1-Propyl-3-ethylcyclopentane	−10.2
Neopentylcyclopentane	84.1
2-Cyclopentyl-3-methylbutane	96.9
1-Isobutyl-1-methylcyclopentane	40.1
1-Isobutyl-2-methylcyclopentane	59.6
1-Isobutyl-3-methylcyclopentane	18.1
2-Cyclopentyl-2-methylbutane	99.7
1-*sec*-Butyl-1-methylcyclopentane	90.4
1-*sec*-Butyl-2-methylcyclopentane	101.5
1-*sec*-Butyl-3-methylcyclopentane	56.7
1,2-Dimethyl-1-propylcyclopentane	77.8
1,3-Dimethyl-1-propylcyclopentane	23.4
1,1-Dimethyl-2-propylcyclopentane	78.1
1,2-Dimethyl-3-propylcyclopentane	83.5
1,3-Dimethyl-4-propylcyclopentane	43.0
1,3-Dimethyl-2-propylcyclopentane	95.1
1,1-Dimethyl-3-propylcyclopentane	23.7
1,2-Dimethyl-4-propylcyclopentane	29.1
1-Isopropyl-1-ethylcyclopentane	102.5
1-Isopropyl-2-ethylcyclopentane	100.7
1-Isopropyl-3-ethylcyclopentane	40.1
1,2-Diethyl-1-methylcyclopentane	102.5
1,3-Diethyl-1-methylcyclopentane	34.4
1,1-Diethyl-2-methylcyclopentane	104.5
1,2-Diethyl-3-methylcyclopentane	103.3
1,3-Diethyl-4-methylcyclopentane	53.9
1,3-Diethyl-2-methylcyclopentane	100.7
1,1-Diethyl-3-methylcyclopentane	61.6
1,2-Diethyl-4-methylcyclopentane	67.5
1-*tert*-Butyl-1-methylcyclopentane	92.5
1-Methyl-2-*tert*-butylcyclopentane	105.1
1-Methyl-3-*tert*-butylcyclopentane	97.2
1,2-Dimethyl-1-isopropylcyclopentane	104.8
1,3-Dimethyl-1-isopropylcyclopentane	72.8
1,1-Dimethyl-2-isopropylcyclopentane	104.8
1,2-Dimethyl-3-isopropylcyclopentane	104.6
1,3-Dimethyl-4-isopropylcyclopentane	99.8
1,3-Dimethyl-2-isopropylcyclopentane	105.2
1,1-Dimethyl-3-isopropylcyclopentane	73.5
1,2-Dimethyl-4-isopropylcyclopentane	78.6
1,1,2-Trimethyl-2-ethylcyclopentane	103.2
1,2,3-Trimethyl-3-ethylcyclopentane	104.8
1,3,4-Trimethyl-4-ethylcyclopentane	96.8
1,2,3-Trimethyl-2-ethylcyclopentane	105.3
1,1,3-Trimethyl-3-ethylcyclopentane	67.9
1,2,4-Trimethyl-4-ethylcyclopentane	73.1
1,1,3-Trimethyl-2-ethylcyclopentane	105.1
1,1,4-Trimethyl-2-ethylcyclopentane	97.0
1,2,2-Trimethyl-3-ethylcyclopentane	104.8
1,1,2-Trimethyl-3-ethylcyclopentane	104.4
1,1,3-Trimethyl-4-ethylcyclopentane	86.4

(*to be continued*)

Table 10 (*continued*)

Hydrocarbon	ON
1,1,2-Trimethyl-4-ethylcyclopentane	73.1
1,2,3-Trimethyl-4-ethylcyclopentane	104.6
1,2,4-Trimethyl-3-ethylcyclopentane	104.9
1,1,2,2,3-Pentamethylpentane	104.9
1,1,2,2,4-Pentamethylpentane	78.0
1,1,2,3,3-Pentamethylpentane	105.3
1,1,3,3,4-Pentamethylpentane	100.0
1,1,2,3,4-Pentamethylpentane	105.2
1,2,2,3,4-Pentamethylpentane	105.4
1,2,3,4,5-Pentamethylpentane	105.9
1-Methyl-2-ethylcyclohexane	52.5
1-Methyl-3-ethylcyclohexane	39.0
1-Methyl-4-ethylcyclohexane	39.0
1,1,3-Trimethylгексан	80.4
Butylcyclohexane	20.9
1-Methyl-1-propylcyclohexane	42.9
1-Methyl-3-propylcyclohexane	13.1
1-Methyl-4-propylcyclohexane	13.1
1,1-Diethylcyclohexane	48.7
1,2-Diethylcyclohexane	23.2
1,3-Diethylcyclohexane	9.7
1,4-Diethylcyclohexane	9.7
1-Isopropyl-1-methylcyclohexane	101.0
1-Isopropyl-2-methylcyclohexane	80.2
1-Isopropyl-3-methylcyclohexane	67.2
1,2-Dimethyl-1-ethylcyclohexane	95.7
1,3-Dimethyl-1-ethylcyclohexane	78.3
1,4-Dimethyl-1-ethylcyclohexane	78.3
1,1-Dimethyl-2-ethylcyclohexane	71.8
1,2-Dimethyl-3-ethylcyclohexane	58.1
1,3-Dimethyl-4-ethylcyclohexane	44.7
1,4-Dimethyl-2-ethylcyclohexane	44.7
1,3-Dimethyl-2-ethylcyclohexane	58.1
1,1-Dimethyl-3-ethylcyclohexane	51.9
1,2-Dimethyl-4-ethylcyclohexane	44.7
1,3-Dimethyl-5-ethylcyclohexane	31.3
1,1-Dimethyl-4-ethylcyclohexane	51.9
1,1,2,2-Tetramethylcyclohexane	104.3
1,1,2,3-Tetramethylcyclohexane	100.0
1,1,2,4-Tetramethylcyclohexane	91.5
1,2,2,4-Tetramethylcyclohexane	101.4
1,2,2,3-Tetramethylcyclohexane	103.9
1,1,3,3-Tetramethylcyclohexane	101.5
1,1,3,4-Tetramethylcyclohexane	85.7
1,1,3,5-Tetramethylcyclohexane	73.0
1,1,4,4-Tetramethylcyclohexane	101.5
1,2,3,4-Tetramethylcyclohexane	91.2
1,2,3,5-Tetramethylcyclohexane	79.1
1,2,4,5-Tetramethylcyclohexane	79.1

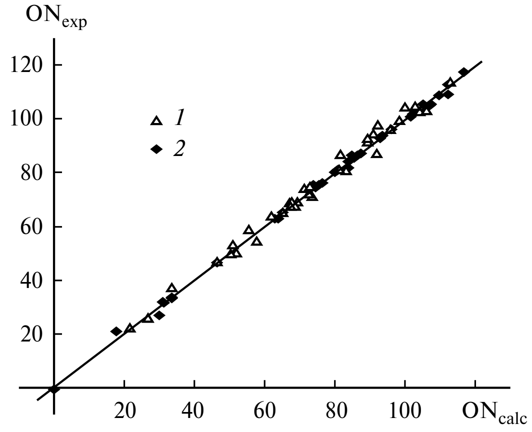

Fig. 7. Results of modeling of ONs using the extrapolation model for alkanes (*1*) and cycloalkanes (*2*).

shown in Fig. 7. This model was used for calculating the ONs of 231 hydrocarbons with $n \le 10$ (94 alkanes, 98 cyclopentanes, and 39 cyclohexanes), which have not been synthesized so far or whose octane numbers were not reported in the available literature.[5,6] The results of calculations are listed in Table 10.

This work was financially supported by the Council on Grants at the President of the Russian Federation (Program of State Support of Leading Scientific Schools in Russian Federation, Grant NSh-9071.2006.3).

References

1. M. I. Stankevich, I. V. Stankevich, and N. S. Zefirov, *Usp. Khim.*, 1988, **57**, 337 [*Russ. Chem. Rev.*, 1988, **57** (Engl. Transl.)].
2. T. G. Schmalz, D. J. Klein, and B. L. Sandleback, *J. Chem. Inf. Comput. Sci.*, 1992, **32**, 54.
3. E. A. Smolenskii, *Izv. Akad. Nauk. Ser. Khim.*, 2006, 1447 [*Russ. Chem. Bull., Int. Ed.*, 2006, **55**, 1501].
4. E. A. Smolenskii, G. V. Vlasova, D. Yu. Platunov, and A. N. Ryzhov, *Izv. Akad. Nauk. Ser. Khim.*, 2006, 1454 [*Russ. Chem. Bull., Int. Ed.*, 2006, **55**, 1508].
5. A. V. Sidorova, I. I. Baskin, D. E. Petelin, V. A. Palyulin, and N. S. Zefirov, *Dokl. Akad. Nauk*, 1996, **350**, 642 [*Dokl. Chem.*, 1996 (Engl. Transl.)].
6. R. D. Obolentsev, *Fizicheskie konstanty uglevodorodov zhidkikh topliv i masel* [*Physical Constants of Hydrocarbons, Liquid Fuels, and Oils*], Gos. Nauchno-Tekhn. Izd. Neftyanoi i Gorno-Topl. Lit., Moscow—Leningrad, 1953, 446 pp. (in Russian).
7. E. A. Smolenskii and A. L. Seifer, *Zh. Fiz. Khim.*, 1963, **37**, 2657 [*J. Phys. Chem. USSR*, 1963, **37** (Engl. Transl.)].
8. E. A. Smolenskii, G. V. Vlasova, and A. L. Lapidus, *Dokl. Akad. Nauk*, 2004, **397**, 219 [*Dokl. Chem.*, 2004 (Engl. Transl.)].
9. M. Randič, *J. Chem. Inf. Comput. Sci.*, 1997, **37**, 672.
10. A. T. Balaban, L. B. Kier, and N. Josh, *Match*, 1992, **28**, 13.

Received January 11, 2007;
in revised form May 3, 2007