Full Articles

Modeling of physicochemical characteristics of alkanes*

E. A. Smolenskii, A. N. Ryzhov, A. L. Lapidus, and N. S. Zefirov*

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

A new type of descriptors is considered, which allows for actual molecular structure in contrast to topological indices. In order to calculate these descriptors, a total of about 600,000 alkane conformational isomers up to $C_{16}H_{34}$ were analyzed. Using the new descriptors, the densities, molar volumes, and the boiling and melting points of alkanes were modeled. The relations derived can be used for prediction of characteristics of non-studied and non-existing compounds. The results of calculations of the melting points of all non-studied *n*-alkanes up to $C_{100}H_{202}$ are reported. The possibility of calculations and prediction of the melting points of some polyethylene-type polymers is shown.

Key words: mathematical modeling, physicochemical properties, structure—property relations, alkanes, descriptors, topological indices, indices topographic, skeletal molecular form.

At present, a rather large number of schemes for the description of structure—property relations are available. They are based on applications of graph theory and permit highly accurate calculations of a broad spectrum of physicochemical characteristics of organic compounds.[1—4] The major problem in this field consists in correct numerical representation of the molecular structure. To this end, a molecule is described using one or more parameters (descriptors), which can be calculated exactly (*e.g.*, the number of subgraphs of a particular type containing in the molecular graph). A particular case of descriptors is provided by the topological indices (TIs). Usually, degeneration occurs in this case, *i.e.*, the same TI value can correspond to different structures. In this case an important characteristic of the dependence between the parameter under study and the TIs is informativity of the TI.[5] Informativity depends on the ratio of the dispersion of the characteristic under study over the classes into which the set of compounds is divided (a class comprises compounds characterized by the same TI values) to the total dispersion of this characteristic. In fact, informativity is the upper limit of the squared correlation coefficient for any theoretically possible analytical dependence of a given property on the TI chosen, which is calculated *a priori*. It should be noted that informativity similar to unity is a necessary but not always sufficient condition for construction of a correct computational scheme.

* Dedicated to Academician N. K. Kochetkov on the occasion of his 90th birthday.

Published in Russian in *Izvestiya Akademii Nauk. Seriya Khimicheskaya,* No. 5, pp. 1056—1071, May, 2005.

1066-5285/05/5405-1084

Yet another, at least equally important, problem is to choose the function relating the physicochemical characteristic under study to the descriptors. It does not always happen that a function, which correctly describes the dependence and at the same time includes a small number of empirical parameters, can be found.[6] Rather often, the desired function is represented by the sum of elementary functions. But in this case the number of variable parameters can be too large, which may require an increase in the number of compounds in the set and thus make the prediction of the characteristics of non-investigated compounds less reliable.

In this work we modeled the densities (ρ) molar volumes (they depend on density), V_M, at 20 °C, and the boiling and melting points (T_b and T_m, respectively) of alkanes at 760 Torr. The last-named characteristic is the well-known "stumbling block" for mathematical modeling.[7,8] Except for a few particular cases there are no quite exact methods for calculating the alkane melting points as yet. For instance, highly accurate calculations were reported,[9] but only for monoalkylalkanes.

In this work the correlation coefficients were calculated using the Microsoft Excel® program. Experimental data were taken from handbooks.[10,11]

Descriptors used in this work. Topographic indices

We began modeling the physicochemical characteristics of alkanes with the use of the simplest TI, namely, the number, n, of carbon atoms as a descriptor. The parameter n can also be treated as the topological distance between terminal C atoms in the n-alkane molecule, expressed through the number of C—C bonds and equal to $n - 1$. Since inclusion of branched alkanes into the set appreciably reduces informativity of the models constructed using the n descriptor, this TI should be used along with other descriptors.

This complicates comparison of the models for branched alkanes. Therefore, we also modeled the same characteristics using the Wiener index, W,[12,13] which equals the sum of the topological distances, d_{ik}, between all pairs of carbon atoms in a molecule:

$$W = \sum_{i=1}^{n-1} \sum_{k=i+1}^{n} d_{ik} \qquad (1)$$

(hereafter i and k denote the number of atoms). The W indices of n-alkanes are calculated as follows

$$W(n) = (n^3 - n)/6. \qquad (2)$$

Since these TIs ignore the actual spatial molecular structure, we employed other descriptors, *e.g.*, the maximum distance, l_{max}, between carbon atoms in n-alkane molecules, which is given by the equation

$$l_{max} = \sqrt{\frac{8}{9}\left(\left[\frac{n-1}{2}\right]\right)^2 + \left(n - 1 - \frac{2}{3}\left[\frac{n-1}{2}\right]\right)^2}, \qquad (3)$$

where the square brackets denote the integer part of the number (the result is expressed in units of the C—C bond length in ethane molecule, 1.54 Å).

This maximum is attained only for TT...T-conformers (for the conformation encoding system, see Ref. 14). Similar relations for the conformations containing *gauche*-fragments were used for calculating the topographic Wiener index, W_{3D},[15] which is analogous to the Wiener index W, but here we deal with the geometrical distances expressed in units of the C—C bond length in ethane molecule rather than the topological distances. The W_{3D} index is a conformation-dependent descriptor calculated as follows

$$W_{3D} = \sum_{j=1}^{N} \omega_j \left(\sum_{i=1}^{n-1} \sum_{k=i+1}^{n} r_{ik,j} \right) \qquad (4)$$

(j is the number of conformer and ω_j is the mole fraction of the conformer). The average physical distance, l_f, between terminal carbon atoms is calculated analogously (only for n-alkanes)

$$l_f = \sum_{j=1}^{N} \omega_j r_{fj} \qquad (5)$$

(r_{fj} is the distance between terminal carbon atoms for the conformation j) and the average distance, l_a, between all carbon atoms in the molecule (which can be expressed with ease through the topographic Wiener index,[16] which makes it possible to reduce the computational resources)

$$l_a = \frac{2}{n(n-1)} \sum_{j=1}^{N} \omega_j \left(\sum_{i=1}^{n-1} \sum_{k=i+1}^{n} r_{ik,j} \right) = \frac{2W_{3D}}{n(n-1)}. \qquad (6)$$

But here a new type of problems arises. The case in point is that the geometry of any n-alkane, starting with C_4H_{10}, is determined ambiguously, because the molecule now has new degrees of freedom associated with internal rotation. This problem can be solved in the framework of the rotational isomeric approximation assuming that molecules adopt for long time only those conformations, which correspond to minima of the internal rotation energy, and ignoring the existence of other forms. These forms are called conformers of a molecule. The conformational composition of a substance depends on the energy difference and statistical weight difference between conformers and on temperature. But calculations of geometric parameters of all conformers of higher alkanes were impossible without employing considerable computational resources, because this requires enumeration of a

Table 1. Number of conformers of *n*-alkane molecules

n-Alkane	Number of conformers	*n*-Alkane	Number of conformers
C_4H_{10}	2	$C_{12}H_{26}$	5002
C_5H_{12}	4	$C_{13}H_{28}$	14884
C_6H_{14}	10	$C_{14}H_{30}$	44530
C_7H_{16}	25	$C_{15}H_{32}$	133225
C_8H_{18}	70	$C_{16}H_{34}$	399310
C_9H_{20}	196	$C_{17}H_{36}$	1196836
$C_{10}H_{22}$	574	$C_{18}H_{38}$	3589414
$C_{11}H_{24}$	1681		

total of 3^{n-3} conformers of a higher alkane C_nH_{2n+2}, exclusion of identical conformers (and optical isomers, if necessary), and calculations of the geometries of the remaining isomers. Finally, it is also necessary to determine the mole fraction, ω_j, of each conformer in the mixture. The ω_j values were calculated as follows

$$\omega_j = g_j \Big/ \sum_{j=1}^{N} g_j, \qquad (7)$$

where N is the number of conformers of a given alkane (N values for *n*-alkanes are listed in Table 1) and g_j is the contribution of the jth conformer to the statistical sum in the denominator of the fraction. The g_j parameters were calculated in the following approximations.

1. High-temperature approximation implies that the reaction temperature is sufficiently high and the energy difference between conformers can be ignored. In this case the contribution of the jth conformer (g_j) equals its statistical weight s_j. The statistical weight of a conformer is called the number of variants of encoding the conformer structure using the encoding system[14] assuming that the optical isomers are indiscernible.

$$g_j = s_j \qquad (8)$$

Note that degeneration of conformers is possible due to the existence of a molecular symmetry plane, or "inversion" of its encoding, *i.e.*, replacement of the numbers of the topologically indiscernible atoms. Therefore, the s_j parameters of *n*-alkanes can take the values 1 (no degeneration), 2 (one symmetry operation), or 4 (both symmetry operations simultaneously) and only these values.

2. Steric approximation takes into account the fact that the $G^+G^-G^+G^-$ conformation of heptane molecule is unrealizable due to superimposition of terminal carbon atoms (heptane effect). This means that if an encoded conformation contains such a fragment, this conformation is unrealizable and its statistical weight is zero. Strictly speaking, the statistical weights of any conformers with partial superimposition of carbon atoms (*excluded volume effect,* V_{excl}) must be set to zero

$$g_j = \begin{cases} s_j, & \text{if } V_{\text{excl}} = 0 \\ 0, & \text{if } V_{\text{excl}} \neq 0. \end{cases} \qquad (9)$$

3. Additive energy approximation implies that transition from the *trans*- to *gauche*-conformation of a molecule is accompanied by an increase in the energy of the molecule by about $E_{T/G} = 0.6$ kcal mol^{-1}.[17] Then one has

$$g_j = s_j \exp[-nE_{T/G}/(RT)], \qquad (10)$$

where n is the number of *gauche*-fragments in the conformer.

4. Additive energy approximation with inclusion of steric effects allows for the effects described in the approximations 2 and 3.

$$g_j = \begin{cases} s_j \exp[-nE_{T/G}/(RT)], & \text{if } V_{\text{excl}} = 0 \\ 0, & \text{if } V_{\text{excl}} \neq 0 \end{cases} \qquad (11)$$

5. Pentane energy approximation assumes that the energies of the G^+G^-- and G^-G^+-conformers of pentane molecule are 2.0 kcal mol^{-1} higher than the energies of the G^+G^+- and G^-G^--conformers of this molecule ($E_{\text{as/s}} =$ 2.0 kcal mol^{-1}).[18]

$$g_j = s_j \exp[-(nE_{T/G} + mE_{\text{as/s}})/(RT)], \qquad (12)$$

where m is the number of such fragments.

6. Pentane energy approximation with allowance for steric effects is in essence the preceding approximation with zero statistical weights of those conformers which cannot be realized for steric reasons.

$$g_j = \begin{cases} s_j \exp[-(nE_{T/G} + mE_{\text{as/s}})/(RT)], & \text{if } V_{\text{excl}} = 0 \\ 0, & \text{if } V_{\text{excl}} \neq 0 \end{cases} \qquad (13)$$

Using all these approximations, we calculated the ω_j values and the 3D-descriptors mentioned above for branched alkanes (up to $n = 10$) and *n*-alkanes (up to $n = 16$). The temperature was taken to be 510 °C, the C—C bond length was set to 1.54 Å, the angles between adjacent C—C bonds were considered tetrahedral, and the angles between C—C bonds separated by a bond were odd multiples of 60°.

In this work we introduce two new 3D-descriptors, namely, the average surface area and the average visualization volume of a skeletal molecular form (SMF). By analogy with the molecular topological form,[19,20] a skeletal molecular form is called a set of $3n$ spatial coordinates

$$F = \{x_i, y_i, z_i\} \ (i = 1, \ldots, n) \qquad (14)$$

of a molecule of a chemical compound containing n nonhydrogen atoms. Construction of this set is similar to the

construction of a molecular graph with omitted H atoms. (Indeed, the omitted H atoms can be reconstructed in the alkane molecular graphs with allowance for the valence; approximate coordinates of these atoms can be determined with ease from the coordinates of the remaining atoms and the bond angles). Of course, the x_i, y_i, z_i ($i = 1, ..., n$) coordinates of a Z_i atom are not arbitrary, being governed by many conditions dependent on the bond lengths and bond angles.

Now we will divide a set of n atoms Z_i into three subsets. If for an atom Z_i there exists a plane P_i passing through Z_i, such that all other non-hydrogen atoms lie on one side of this plane (strictly speaking, in this case we have an infinite set of such planes), the Z_i atom is called an external atom and denoted as an element of the subset A_e (if the Z_i atom is the only non-hydrogen atom, as in the molecules of methane and volatile hydride compounds, it is *a priori* treated as the external atom). If there is no P_i plane, but for the atom Z_i there is a plane E_i such that this atom belongs to E_i and any of the remaining non-hydrogen atoms lies on one side of E_i or belongs to E_i, the atom Z_i is called a frontier atom A_s. If for the Z_i atom there is no E_i plane, this atom will be called an internal atom A_i. Apparently, these definitions divide the set of non-hydrogen atoms in a molecule into three subsets. The meaning of these definitions consists in that the external atoms (for the molecules in which the non-hydrogen atoms do not lie in the same plane) form a convex polyhedron, lying at its vertices, and the internal atoms lie inside of this polyhedron, the frontier atoms being located exactly at the faces of the polyhedron. The geometric figure formed by the external atoms will be called the SMF visualization.

Let us exemplify these definitions. The cases $n = 1—3$ are trivial (all carbon atoms are external atoms, but the corresponding SMF visualizations are fundamentally different, namely, a point, a linear segment, and a triangle, respectively). The case for $n = 4$ (n-butane) has two variants. In the *trans*-conformation all four carbon atoms are external atoms, while in the *gauche*-conformation these atoms are also external atoms but do not lie in the same plane, as in the former case, and form an irregular tetrahedron. This also concerns isobutane (all carbon atoms are external atoms and form a regular tetrahedron). The case for $n = 5$ provides the first examples of internal atoms (quaternary C atom in neopentane molecule) and frontier atoms (central carbon atom in the *trans*-conformation of n-pentane molecule) of the SMF. But internal atoms also exist in certain conformations of n-alkane molecules. A simplest example is provided by the central C atom in the G^+TTG^+-conformation of n-heptane molecule (and in its optical isomer, G^-TTG^--conformation) (these conformations can be called crab conformations). Some examples of the SMFs and SMF visualizations are given in Fig. 1.

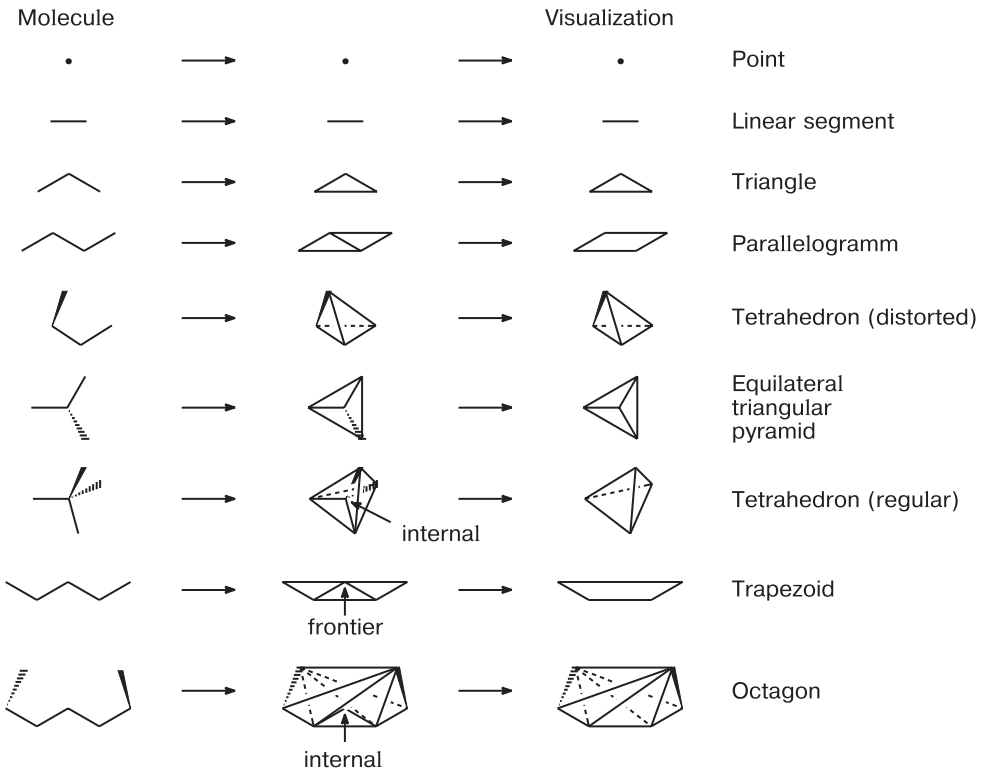

Fig. 1. The SMF visualizations of different molecules.

Table 2. Parameters (a, b, c, d) and squared correlation coefficients (r^2) for the models of T_b of n-alkanes constructed using the n, W, and l_{max} descriptors

Relation	Descriptor	a	b	c	d	r^2
(17)	n	316.2287	3.967426	−593.186	—	0.99992
	l_{max}	317.5157	3.986708	−663.034	—	0.999879
	W	83.65084	4.541555	−248.792	—	0.99870
(18)	n	−25755.1	19.67019	1.079617	899.9872	0.99862
	l_{max}	1040.575	1.768962	−0.16601	−1391.59	0.998796
	W	1138.396	1.478301	−0.05169	−1303.2	0.99603
(19)	n	608.8013	732.3515	0.067757	—	0.99680
	l_{max}	610.366	772.3777	0.054894	—	0.996835
	W	468.9681	453.7791	0.001314	—	0.91

The SMF visualization has the surface area and volume. But since the visualizations of different conformers differ from one another, it is meaningful to consider only the $\bar{V}$ and $\bar{S}$ values averaged over all conformations, which can be considered as new topographic indices:

$$\bar{V} = \sum_{j=1}^{N} \omega_j V_j, \qquad (15)$$

$$\bar{S} = \sum_{j=1}^{N} \omega_j S_j. \qquad (16)$$

The $\bar{S}$ and $\bar{V}$ values are measured in units of the squared and cubed C—C bond length in ethane molecule.

Modeling of alkane boiling points at 760 Torr

The n-alkane boiling points, T_b, were modeled using the following relationships (here x is the descriptor):

$$f(x) = a\ln(b + x) + c, \qquad (17)$$

$$f(x) = a/(b + x)^c + d, \qquad (18)$$

$$f(x) = a - b\exp(-cx) \qquad (b, c > 0). \qquad (19)$$

Modeling of the n-alkane boiling points can be performed with ease using these relations and the n descriptor (see above). The use of the Wiener index, W, as the descriptor leads to somewhat worse results. The use of the l_{max} descriptor calculated using Eq. (3) gives similar and sometimes better results compared with the calculations based on n. Parameters of the models are listed in Table 2. It may appear that the squared correlation coefficients for these descriptors are virtually the same. But analysis of the differences between the squared correlation coefficients and unity shows that each succeeding descriptor is about an order of magnitude worse than the preceding one.

The approximations 3—6 used for calculating the W_{3D}, l_a, and l_f descriptors cannot be employed to model the boiling point. These calculations are carried out using the temperature as parameter; therefore, we arrive at vicious circle. The results of calculations using a set of fourteen n-alkanes C_3H_8—$C_{16}H_{34}$ are listed in Table 3. For comparison, we also present the results of calculations using the n and W descriptors.

Table 3. Squared correlation coefficients (r^2) for modeling of the dependence of the boiling points of n-alkanes on different descriptors

Descriptor	Approximation	r^2		
		Relation (17)	Relation (18)	Relation (19))
n	—	0.9999937	0.9994754	0.9998310
W	—	0.9994817	0.9901738	0.9748393
W_{3D}	1	0.9999998	0.9999569	0.9997352
	2	0.9999995	0.9999548	0.9997136
l_f	1	0.9998625	0.9998390	0.9999921
	2	0.9999990	0.9999022	0.9999976
l_a	1	0.9999958	0.9999978	0.9999983
	2	0.9999992	0.9999993	0.9999996
$\bar{V}$	1	0.9972467	0.9853407	0.9795877
	2	0.9973095	0.9847709	0.9786674
$\bar{S}$	1	0.9992639	0.9991727	0.9990297
	2	0.9991594	0.9990439	0.9989717

The best agreement with experimental data was obtained using relationships (17)—(19) and the W_{3D} and l_f descriptors (see Table 3). Eventually, we modeled T_b using the $\bar{V}$ and $\bar{S}$ descriptors calculated for the same set comprising fourteen n-alkanes. The corresponding results are also listed in Table 3. Analysis of these data shows that the model based on the $\bar{S}$ descriptor is comparable in quality with the models constructed using TIs, being at the same time much worse than the models based on the topographic indices.

The changes in T_b for the series of isomeric n-alkanes and methylalkanes are plotted in Fig. 2 (x_b is the topological coordinate of the branching point of alkane, counted from the end of the medium branch). It is clearly seen that the slopes of the straight lines are similar for all these series. Consider the dependence of T_b on the length, l_{min}, of the shortest substituent. Figure 3 clearly demonstrates similarity of the slopes of all straight lines. This

Table 4. Parameters (a—i) and squared correlation coefficients (r^2) for the models of T_b of n-alkanes and monoalkylalkanes, constructed using independent topological indices and the Wiener index W as descriptors

Descriptor	a	b	c	d	e	f	g	h	i	r^2
n, x_b, l_{min}	1038.977	1.782663	−0.16677	22.48799	−1.74566	15.95611	−49.8801	63.93199	−1393.13	0.999825
W (18)	1136.388	1.686365	−0.05189	−1304.9	—	—	—	—	—	0.99862

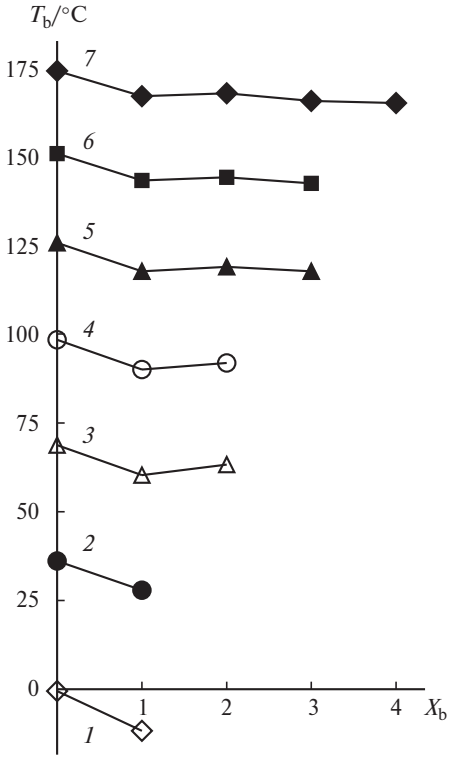

Fig. 2. Changes in the boiling points in series of isomeric methylalkanes: n = 4 (*1*), 5 (*2*), 6 (*3*), 7 (*4*), 8 (*5*), 9 (*6*), and 10 (*7*).

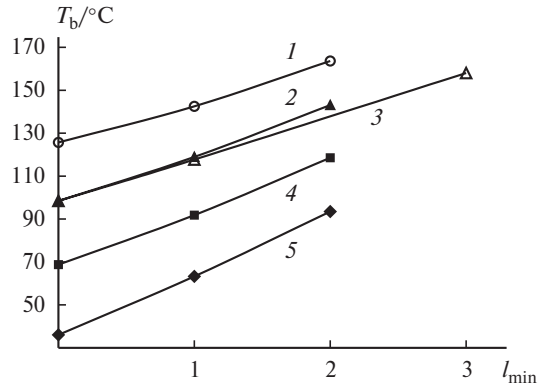

Fig. 3. Boiling points of monoalkylalkanes plotted *vs*. length of the shortest branch: $n - l_{min}$ = 8, x_b = 2 (*1*); $n - l_{min}$ = 7, x_b = 2 (*2*); $n - l_{min}$ = 7, x_b = 3 (*3*); $n - l_{min}$ = 6, x_b = 2 (*4*); $n - l_{min}$ = 5, x_b = 2 (*5*).

means that T_b of monoalkylalkanes increase linearly as the shortest branch is lengthened.

Thus, the boiling points of n-alkanes and monoalkylalkanes can be approximated as follows:

$$T_b = f(n) + el_{min} + x_b(fx_b^3 + gx_b^2 + hx_b + i), \quad (20)$$

where n is the number of carbon atoms (for n-alkanes) or the total number of carbon atoms in the long branches plus 2 (for monoalkylalkanes); and $f(n)$ is one of the model functions (17)—(19). Note that modeling of T_b of monoalkylalkanes using three independent TIs (n, x_b, l_{min}) as descriptors has informativity equal to unity. For comparison, we constructed yet another model based on the Wiener index (calculations were performed using relation (18)), which appeared to be much worse than the preceding model. The parameters of both models are listed in Table 4.

Now we will compare the models based on other descriptors including $\bar{V}$ and $\bar{S}$ (see above) with the model (20). (Because in this case informativity of the descriptor, n, differs from unity and l_f is senseless, they were not considered.) A set of twenty-seven monoalkylalkanes was used. The results are listed in Table 5. The best agreement was achieved by modeling based on the 3D-descriptors and even the use of theoretically necessary and sufficient (with respect to informativity) number of independent topological parameters does not allow to obtain better results.

Modeling of the density of liquid alkanes

Similarly to the case for T_b, we will begin modeling of the density of liquid n-alkanes with the use of the n de-

Table 5. Squared correlation coefficients (r^2) for modeling of T_b of monoalkylalkanes using different descriptors

Descriptor	Approximation	r^2		
		Relation (17)	Relation (18)	Relation (19)
n, x_b, l_{min}	—	0.995474	0.995348	0.995176
W	—	0.998327	0.889852	0.974537
W_{3D}	1	0.999767	0.998803	0.997384
	2	0.999737	0.998765	0.997364
l_a	1	0.998549	0.998721	0.858329
	2	0.998508	0.998589	0.857632
$\bar{V}$	1	0.997842	0.997843	0.985581
	2	0.998334	0.922281	0.985124
$\bar{S}$	1	0.997289	0.997037	0.998208
	2	0.997810	0.924665	0.997661

Table 6. Squared correlation coefficients (r^2) for modeling of the density of *n*-alkanes using different descriptors (listed are the maximum out of six values obtained in calculations using different approximations; the number of approximation is given in parentheses)

Descriptor	r^2		
	Relation (17)	Relation (18)	Relation (19)
n	0.9993849	0.9999865	0.9991867
W	0.9977767	0.9999754	0.9802286
l_{max}	0.9991412	0.9999544	0.9993132
W_{3D}	0.9982547 (2)	0.9999950 (1)	0.9866265 (1)
l_a	0.9997225 (3)	0.9999961 (1)	0.9999060 (1)
l_f	0.9998386 (5)	0.9999910 (6)	0.9998873 (3)
$\bar{V}$	0.9991331 (1)	0.9999885 (2)	0.9860594 (1)
$\bar{S}$	0.9988062 (1)	0.9992289 (6)	0.9996461 (1)

scriptor. The model based on the W descriptor gives somewhat worse results while the model involving the l_{max} descriptor provides with intermediate results. The correlation coefficients and parameters of the models constructed using these conformation-independent descriptors are listed in Tables 6 and 7.

We modeled the densities of *n*-alkanes C_5H_{12}—$C_{16}H_{34}$ using three 3D-descriptors considered above and the $\bar{V}$ and $\bar{S}$ descriptors and all six approximations (unlike the modeling of T_b, here we can use all approximations) using relationships (17)—(19). The correlation coefficients and parameters of the models are listed in Tables 6 and 8.

From the data in Table 6 it follows that the results of density modeling using almost all 3D-descriptors (including $\bar{V}$) are comparable with (and often, better than) the results of modeling using the TIs. At first glance the only exception is the $\bar{S}$ descriptor. But this is only ostensibility. The case in point is that the $\bar{S}$ and $\bar{V}$ descriptors are related by a nonlinear relation

$$\bar{S} = k\bar{V}^n + m \qquad (21)$$

(parameters of this model are listed in Table 9). Therefore, the $\rho(\bar{S})$ dependence must have the form (from Table 6 it follows that only relationship (18) correctly reproduces the $\rho(x)$ dependences)

$$\rho = \frac{a}{\left(\sqrt[n]{\bar{S} - m} + b\right)^c} + d. \qquad (22)$$

The results of modeling are listed in Table 10. In this case the results obtained were only slightly improved, *viz.*, the r^2 value approached unity by ~0.0005.

Now we will consider yet another approach to the modeling of density. From the definitions of the density, molar weight (M), and molar volume it follows that

$$V_M = M/\rho. \qquad (23)$$

Modeling of the dependence of V_M on the descriptors described above is a much simpler task, because the molar volumes of *n*-alkanes continuously and monotonically increase with the increase in the molecular sizes and tend to $+\infty$.

Analysis of the $V_M(x)$ dependences (Fig. 4) shows that

1) if $x = n$ or l_{max}, one gets a linear dependence;

2) if $x = l_a$, l_f, or $\bar{S}$, the dependence can be considered to some extent linear only for calculating the concentrations of conformers using the approximations 5 and 6;

3) is $x = \bar{V}$, W, or W_{3D}, one gets a nonlinear dependence.

Table 7. Parameters of different models of the density of *n*-alkanes constructed using conformation-independent descriptors

Descriptor	Relation (17)			Relation (18)				Relation (19)		
	a	$-b$	c	$-a$	b	c	d	a	b	c
n	0.071095	3.480113	0.595466	1.662062	1.318899	1.099153	0.845503	0.78837	0.023824	0.204781
W	0.033302	12.9849	0.560189	0.51564	−1.37544	0.239757	0.881657	0.764646	0.155592	0.009889
l_{max}	0.071964	1.99527	0.607737	1.26389	1.82313	1.068188	0.848262	0.788685	0.363775	0.249574

Table 8. Parameters of different models of density of *n*-alkanes, constructed using the conformation-dependent descriptors obtained in the best approximations (see Table 6)

Descriptor	Relation (17)			Relation (18)				Relation (19)		
	a	$-b$	c	$-a$	b	c	d	a	b	c
W_{3D}	0.037276	9.82906	0.55958	0.60165	1.72231	0.326454	0.863325	0.766739	0.170148	−0.01637
l_a	0.086389	1.2333	0.706249	0.47743	−0.03069	1.58297	0.869161	0.811342	1.350036	−1.27351
l_f	0.090222	1.61476	0.600183	0.83352	0.840372	0.816566	0.907673	0.818404	0.512877	−0.34825
$\bar{V}$	0.038627	0.09859	0.690193	0.23795	0.126365	0.222933	0.916325	0.767247	0.055024	0.543411
$\bar{S}$	0.085251	0.44841	0.518282	0.55916	1.260052	0.374031	0.936628	0.799978	$6.26 \cdot 10^{-6}$	0.109387

Table 9. Results of modeling of the $\bar{S}(\bar{V})$ dependence

Approximation	m	k	n	r^2
1	0.026105	0.010348	2.235339	0.999664
2	0.02192	0.010603	2.22334	0.999641
3	0.00936	0.008387	2.229975	0.999694
4	0.040572	0.007583	2.255479	0.99818
5	−0.048143	0.012167	2.082952	0.997804
6	−0.006016	0.010265	2.133081	0.999725

Nonlinear dependences were described by the following relationship

$$V_M = a + bx^c, \quad (24)$$

where a, b, abd c are free parameters. The results of modeling are listed in Table 11.

From Table 11 it follows that the nonlinear model describes the $V_M(x)$ dependences better than the linear model and that the models based on the 3D-descriptors are better, with respect to r^2 values, than the models constructed using conventional TIs (except for the Wiener index). Failures of the modeling with the $\bar{S}$ descriptor are due to nonlinear relation between this descriptor and the $\bar{V}$ descriptor (see above).

Now consider the $V_M(x)$ ($x = W$, W_{3D}, l_a, $\bar{V}$, $\bar{S}$) dependences for a set comprising a total of twenty-five monoalkylalkanes. (Because in this case informativity of the n descriptor differs from unity and the l_f and l_{max} descriptors are senseless, they were leaved out of consideration.)

In this case analysis of the $V_M(x)$ dependences (Fig. 5) shows that

1) at $x = W$, W_{3D}, or $\bar{V}$, the dependence monotonically increases;

2) at $x = \bar{S}$, the dependence for monoalkylalkanes can be considered linear, although this is not the case for n-alkanes;

3) at $x = l_a$, one has a family of parallel lines, all points corresponding to isomers being at the same line. In the last case one must model using a function of two descriptors, l_a for monoalkylalkane and l_{an} for corresponding isomeric n-alkane:

$$V_M = k(l_a - l_{an}) + V_M(l_{an}). \quad (25)$$

The results are listed in Table 12.

By and large, this siuation is similar to the cases considered above. The 3D-descriptors, except for $\bar{S}$, correlate with the molar volume much better than the Wiener

Table 10. Results of modeling of the $\rho(\bar{S})$ dependence using relationship (22)

Approximation	$-m$	n	$-a$	b	c	d	r^2
1	0.948458	1.702778	0.904706	1.179283	0.361373	1.221027	0.999567
2	0.923701	1.695705	0.858105	1.15853	0.382901	1.178117	0.999536
3	0.510738	1.593561	0.630565	0.76639	0.367549	1.069276	0.99942
4	0.465742	1.571243	0.611458	0.72711	0.362086	1.060628	0.999365
5	0.288243	1.488645	0.544149	0.573688	0.348104	1.027609	0.999234
6	0.271913	1.48745	0.538818	0.560814	0.347658	1.024989	0.999441

Table 11. Squared correlation coefficients (r^2) for modeling of the dependences of the molar volumes of n-alkanes at 20 °C on different descriptors* in approximations 1—6

Descriptor	Model	r^2					
		1	2	3	4	5	6
W_{3D}	Nonlinear	0.999996	0.999995	0.999997	0.999995	0.999998	0.999998
l_a	Nonlinear	0.999995	0.999987	0.999997	0.999998	0.999997	0.999997
	Linear	0.992706	0.994101	0.996447	0.996949	0.998664	0.998684
l_f	Nonlinear	0.999987	0.999905	0.99998	0.999987	0.999957	0.999959
	Linear	0.990581	0.990159	0.989374	0.990404	0.994765	0.994795
$\bar{V}$	Nonlinear	0.999990	0.999996	0.999995	0.999988	0.999994	0.999993
$\bar{S}$	Nonlinear	0.999559	0.999478	0.999408	0.999124	0.998527	0.999381
	Linear	0.992706	0.994101	0.996447	0.996949	0.998664	0.998684

* The descriptors n (linear model), W (nonlinear model), and l_{max} (linear model) are conformation-independent; therefore, they are characterized by single values of the squared correlation coefficient: 0.999959, 0.999999, and 0.999929, respectively.

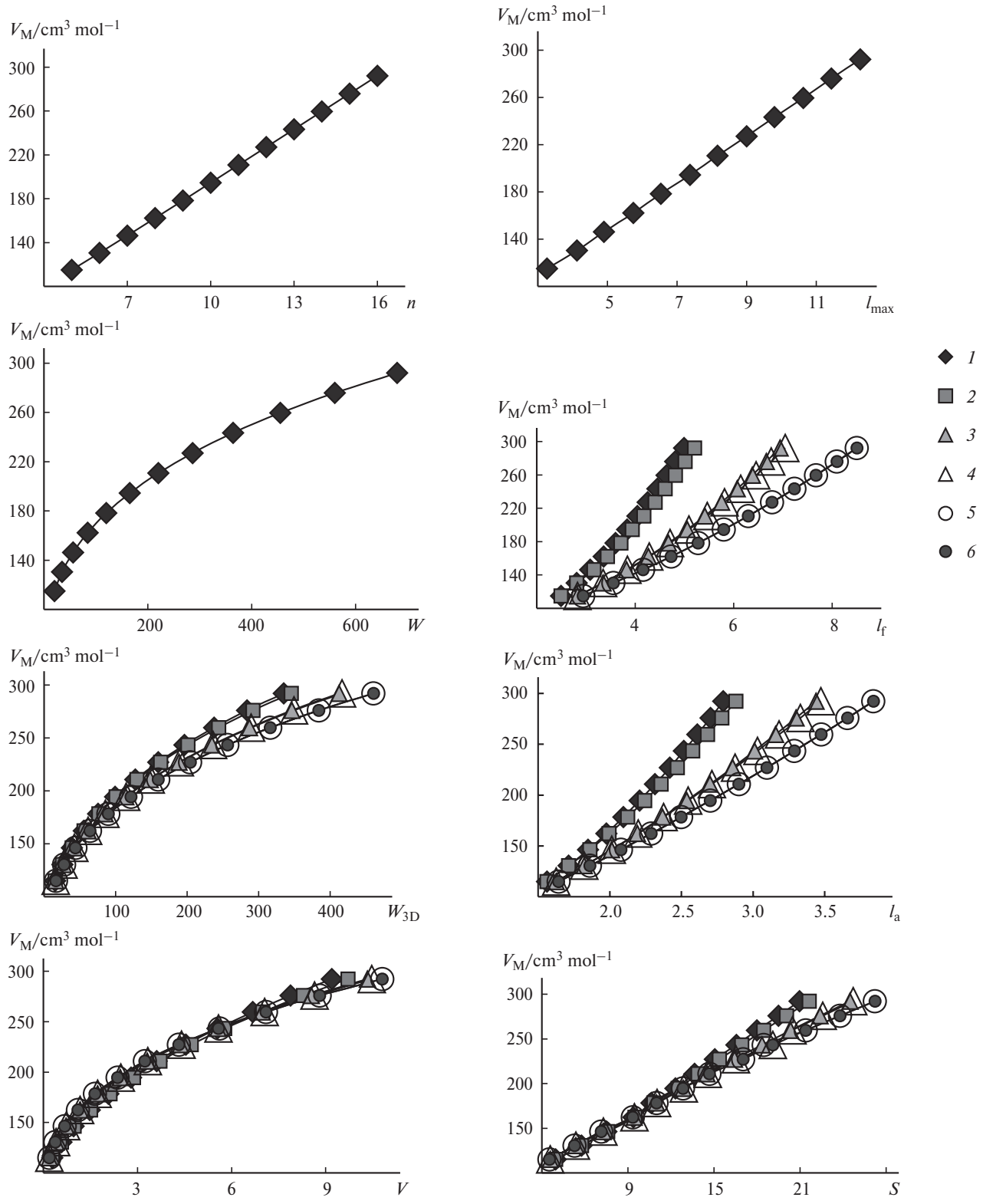

Fig. 4. Molar volumes of *n*-alkanes at 20 °C plotted *vs.* different descriptors. Approximation number: 1 (*1*), 2 (*2*), 3 (*3*), 4 (*4*), 5 (*5*), and 6 (*6*).

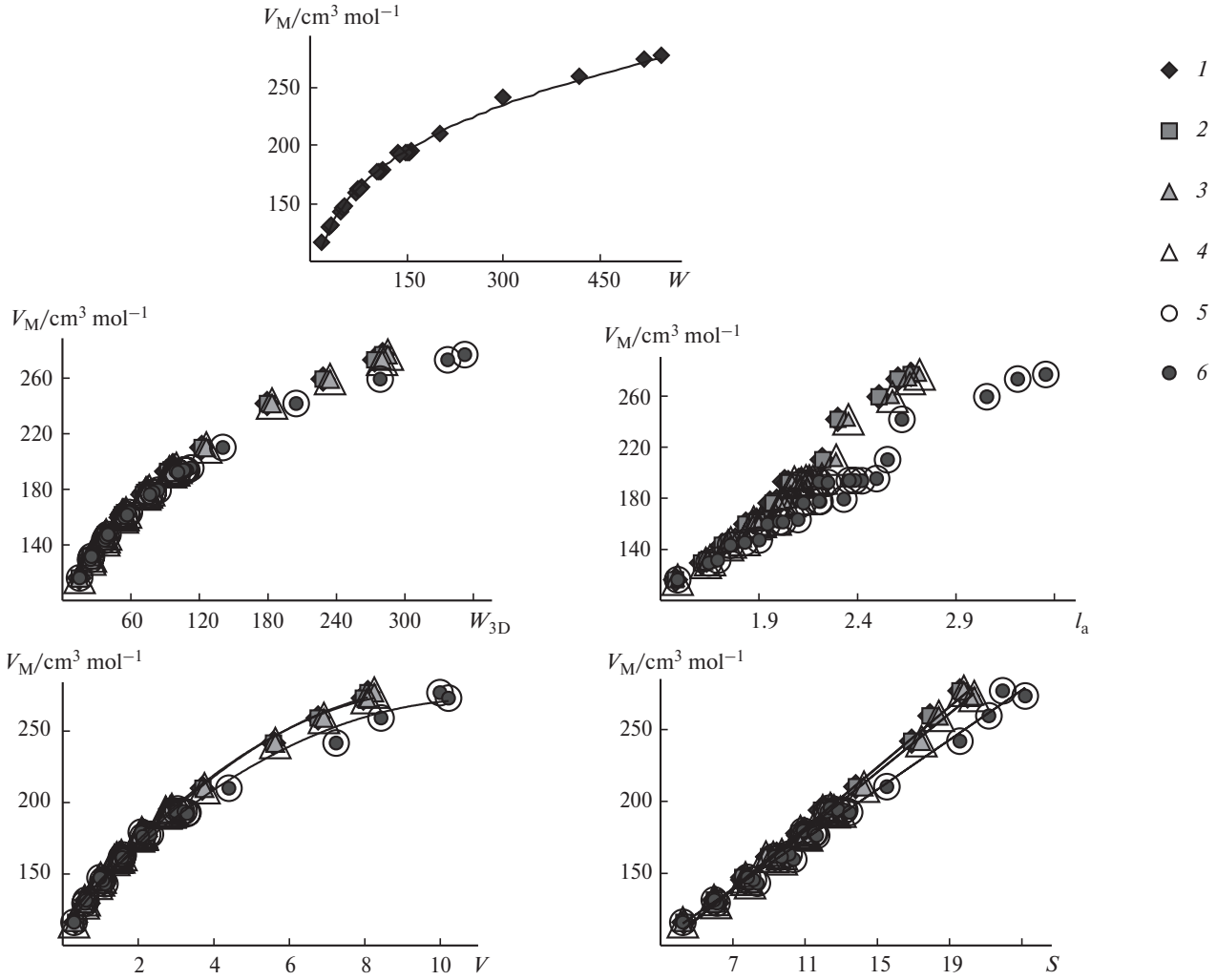

Fig. 5. Molar volumes of monoalkylalkanes at 20 °C plotted *vs*. different descriptors. Approximation number: 1 (*1*), 2 (*2*), 3 (*3*), 4 (*4*), 5 (*5*), and 6 (*6*).

Table 12. Squared correlation coefficients (r^2) for modeling of the dependences of the molar volume of monoalkylalkanes at 20 °C on different descriptors using approximations 1—6*

Descriptor	Model	r^2					
		1	2	3	4	5	6
W_{3D}	Nonlinear	0.99935	0.99937	0.99931	0.99933	0.99845	0.99845
l_a	Nonlinear	0.99947	0.99952	0.99955	0.99956	0.99960	0.99960
	Linear	0.99096	0.99159	0.99436	0.99492	0.99796	0.99798
$\bar{V}$	Nonlinear	0.99875	0.99887	0.99933	0.99931	0.99542	0.99542
$\bar{S}$	Nonlinear	0.99647	0.99659	0.99407	0.99368	0.98700	0.98700
	Linear	0.99582	0.99543	0.99252	0.99177	0.98677	0.98677

* Modeling using the nonlinear model with the W descriptor gave $r^2 = 0.99819$.

index, the best results being achieved using the l_a descriprot and relationship (24).

In order to model the molar volumes of polymethylalkanes, we will first consider how they depend on the descriptors under study (see the plots in Fig. 6). From these plots it follows that the $\bar{S}$ descriptor cannot be used to model this dependence. Modeling with the W and W_{3D} descriptors can be performed using relationship (24), while the l_a and $\bar{V}$ descriptors are applicable only in combination with the corresponding parameters

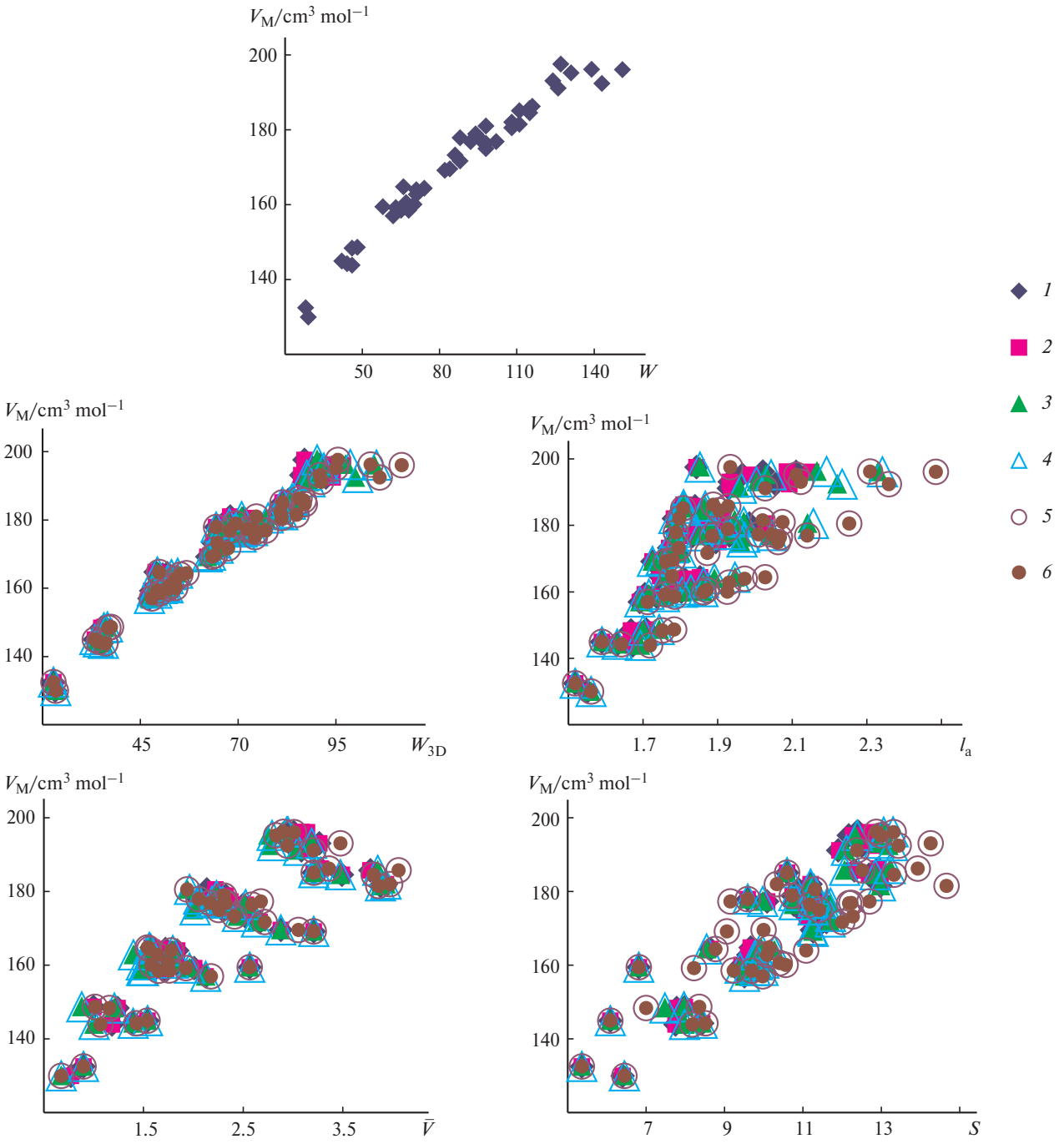

Fig. 6.* Molar volumes of polymethylalkanes at 20 °C plotted *vs.* different descriptors. Approximation number: 1 (*1*), 2 (*2*), 3 (*3*), 4 (*4*), 5 (*5*), and 6 (*6*).

for *n*-alkanes and descriptor *n* (Table 13, a set of 43 compounds was used):

$$V_M = a(n + b)(l_a - l_{an}) + cl_{an}{}^d + e, \quad (26)$$

$$V_M = a(n + b)(\bar{V} - \bar{V}_n) + c\bar{V}_n^d + e. \quad (27)$$

The model constructed using three descriptors ($\bar{V}$, $\bar{V}_n$ and n) provides the best results for V_M. The final relationship for calculating the density has the form:

$$\rho(n, \bar{V}) = \frac{14n + 2}{a(n + b)(\bar{V} - \bar{V}_n) + c\bar{V}_n^d + e}. \quad (28)$$

* Figure 6 is available in full color in the on-line version of the journal (http://www.springeronline.com) and on the web site of the journal (http://russchembull.ru).

Table 13. Squared correlation coefficients (r^2) for modeling of the dependences of the molar volumes of polymethylalkanes at 20 °C on different descriptors in approximations 1—6*

Descriptor	r^2					
	1	2	3	4	5	6
W_{3D}	0.977702	0.973928	0.978215	0.970645	0.963348	0.963409
l_a	0.979586	0.978615	0.977973	0.977872	0.976768	0.976767
$\overline{V}$	0.984692	0.982366	0.977921	0.977347	0.979670	0.979671

* For the W descriptor, one gets $r^2 = 0.979378$.

Table 14. Squared correlation coefficients (r^2) for the models of the dependence of the molar volume on the descriptors using relationship (27) in approximations 1—6

Model	r^2					
	1	2	3	4	5	6
Prediction for *n*-alkanes	0.999634	0.999261	0.988766	0.990199	0.996330	0.996314
Prediction for monoalkylalkanes	0.998119	0.987861	0.993704	0.994972	0.980674	0.980236
Common model	0.994717	0.996640	0.996129	0.996033	0.995083	0.995046

Table 15. Parameters of models for the dependence of the molar volume and density of alkanes on descriptors using relationships (27) and (28)

Approximation	*a*	*b*	*c*	*d*	*e*
1	2.773602	−5.9354	66.45504	0.522693	81.19923
2	1.222619	−2.00362	67.17523	0.499136	80.45182
3	−0.085233	−87.32328	86.7056	0.39554	71.84099
4	−0.085701	−86.34729	87.62494	0.390412	70.85427
5	0.026873	159.9975	90.5954	0.383787	69.57025
6	0.027126	157.402	90.5168	0.384211	69.63168

Now let us predict the molar volumes of *n*-alkanes and monoalkylalkanes using relation (27) and compare the data with the results of modeling using a training set comprising polymethylalkanes. Here at $r^2 \approx 0.977$—0.985 for the test sets one gets $r^2 \approx 0.980$—0.998 for monoalkylalkanes and $r^2 \approx 0.990$—0.999 for *n*-alkanes (Table 14).

Parameters of the models that are common to all alkanes are listed in Table 15. The density itself should be calculated using relationship (28). The set comprises a total of eighty-four alkanes.

Modeling of alkane melting points

Undoubtedly, the melting points (T_m) of organic compounds depend on their molecular structures in the most complex fashion in spite of apparent simplicity of experimental measurement of this parameter. At present, this is the one and only physicochemical characteristic, for which no reasonable structure—property model was proposed so far. This fact was pointed out earlier[7] and then analyzed in more detail.[5] The exceptions are a few particular cases, *e.g.*, calculations of T_m of monoalkylalkanes.[9]

In spite of a few attempts to construct computational schemes for predicting the T_m of organic compounds,[9,21] solving the problem even for alkanes seems to be unreal at

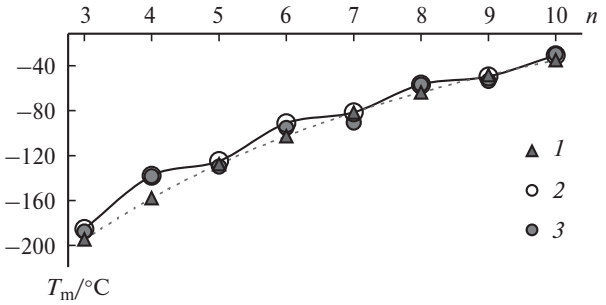

Fig. 7. Predicted melting points of *n*-alkanes C_3H_8—$C_{10}H_{22}$ plotted *vs.* *n* (*1*) and *x* (*2*) and experimental data (*3*). Here and in Fig. 8 the abscissa axis for plot (*2*) is given in *n* values corresponding to the *x* values.

Table 16. Results of prediction of the $T_m = f(n)$ dependence using the parameters n (I) and x (II)

n	x	T_m/°C		
		Exper-iment	Calculations	
			I	II
	Prediction for C_3H_8—$C_{10}H_{22}$			
3	1.33	−187.65	−183.436	−188.38
4	2.33	−138.35	−150.278	−141.24
5	2.67	−129.72	−122.653	−128.02
6	3.67	−95.32	−99.3699	−94.08
7	4.00	−90.61	−79.5458	−84.38
8	5.00	−56.80	−62.5131	−59.12
9	5.33	−53.54	−47.7599	−51.79
10	6.33	−29.67	−34.8881	−32.46
	Prediction for higher n-alkanes			
25	16.00	53.30	53.10	52.42
26	17.00	56.20	55.65	56.29
27	17.33	59.50	58.00	57.48
28	18.33	61.28	60.19	60.81
29	18.67	63.90	62.23	61.84
30	19.67	65.88	64.12	64.73
31	20.00	67.25	65.89	65.63
32	21.00	70.16	67.55	68.16
33	21.33	71.80	69.10	68.95
34	22.33	72.65	70.55	71.17
35	22.67	74.45	71.92	71.86
36	23.67	75.83	73.20	73.82
37	24.00	77.40	74.42	74.44
38	25.00	79.30	75.56	76.18
39	25.33	80.10	76.64	76.73
40	26.33	81.30	77.65	78.28
41	26.67	81.70	78.62	78.77
42	27.67	82.90	79.53	80.16
43	28.00	85.30	80.40	80.60
44	29.00	86.40	81.22	81.85
50	33.00	93.00	85.39	86.02
52	34.33	94.00	86.54	87.16
54	35.67	95.00	87.59	88.21
60	39.67	98.90	90.26	90.86
62	41.00	100.50	91.02	91.61
64	42.33	102.00	91.72	92.30
66	43.67	103.60	92.37	92.95
67	44.00	104.10	92.67	93.10
70	46.33	105.25	93.54	94.10
71	46.67	105.50	93.80	94.23
82	54.33	110.40	96.24	96.74
100	66.33	115.30	98.86	99.27

present. Note that even the T_m of n-alkanes exhibit a complex behavior. Figure 7 and Table 16 show that the melting points of the even and odd n-alkanes are changed in significantly different manner.

Consider Figs 7 and 8. At first glance it seems that one should derive two equations for separate description of the T_m of n-alkanes with even and odd number of carbon atoms in the molecule. This is quite a natural desire, but it

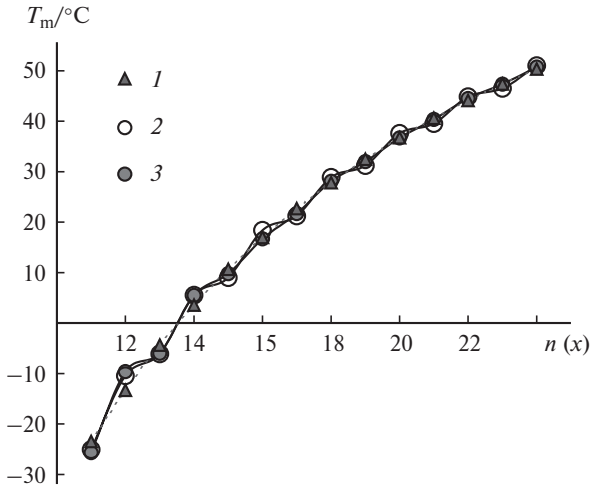

Fig. 8. Predicted melting points of n-alkanes $C_{11}H_{24}$—$C_{24}H_{50}$ plotted *vs.* n (*1*) and x (*2*) and experimental data (*3*).

is better to derive a general relationship using, *e.g.*, expression (18) and the n descriptor. Parameters of the models are given below.

Parameter	n	x
a	−25755	−25755
b	9.86	8.98
c	1.74	1.92
d	106.02	105.67

The results of calculations are listed in Table 17 and presented in Fig. 8. The statistical characteristics of the model are as follows: $r^2 = 0.9962$, $s = 1.45$, and cr.v.$r^2 = 0.9902$ (abbreviation "cr.v." denotes "cross validation").

Table 17. Results of modeling of the $T_m = f(n)$ using parameters n (I) and x (II)

n	x	T_m/°C		
		Exper-iment	Calculations	
			I	II
11	6.67	−25.65	−23.44	−25.09
12	7.67	−9.60	−13.30	−10.42
13	8.00	−6.00	−4.36	−6.08
14	9.00	5.50	3.58	5.55
15	9.33	9.81	10.65	9.02
16	10.33	16.70	16.98	18.40
17	10.67	21.72	22.68	21.22
18	11.67	28.10	27.82	28.90
19	12.00	32.00	32.48	31.22
20	13.00	36.70	36.72	37.59
21	13.33	40.35	40.87	40.56
22	14.33	44.50	44.54	46.21
23	14.67	47.35	47.92	47.95
24	15.67	50.75	51.05	52.75

Let us consider the $C_{11}H_{24}$—$C_{24}H_{50}$ *n*-alkanes as the training set and more light *n*-alkanes as the test set. We will "predict" the melting points of *n*-alkanes from C_3H_8 to $C_{10}H_{22}$; methane and ethane were leaved out of consideration, because their T_m do not obey this dependence, being outliers. The results of our prediction are shown in Fig. 7 and listed in Table 16. Experimental points readily begin to deviate from the predicted values, the deviation being as large as 12 °C at $n = 4$ at $r^2 = 0.9659$ and $s = 10.61$. From these data it follows that the model using relationship (18) has "good" statistical characteristics but is inappropriate for the description of the $T_m(n)$ dependence. The *n* descriptor ignores peculiar features of the molecular structure, which govern the characteristic under study.

Although *n*-alkanes have a large number of conformational isomers,[14] one can assume that they crystallize in one, most energetically favorable TT...T-conformation.

In this connection let us consider the following situation. Let the *X* axis begin at a terminal carbon atom (C_{Me}), being directed along the line of the C_{Me}—C_{CH_2} bond. Then the desired parameter is the coordinate, *x*, of the other terminal atom, C_{Me}, *i.e.*, the projection of the topographic distance between terminal carbon atoms in the molecule on the direction of the first C—C bond for the TT...T-conformation. In fact we now have the following parameter (rather than the *n* parameter)

$$x(n) = n - 1 - (2/3)[(n - 1)/2], \qquad (29)$$

where the square brackets denote the integer part of a number. This parameter allows for the character of the T_m dependence on the evenness of *n* for the first members of homologous series. We performed calculations using relationship (18), the *x* descriptor, and the same data for *n*-alkanes as for the preceding model. The results of calculations are listed in Table 17 and in Fig. 8. Statistical characteristics of the model ($r^2 = 0.9988$, $s = 0.80$, cr.v.$r^2 = 0.9984$) suggest that it offers considerable advantages compared to the model based on the *n* descriptor. The results of prediction (in fact, extrapolation) for the C_3H_8—$C_{10}H_{22}$ *n*-alkanes are shown in Fig. 7 and listed in Table 16. Indeed, the model involving relationship (18) and the *x* descriptor has much better statistical characteristics compared to the model based on the *n* descriptor ($r^2 = 0.9926$ and $s = 4.94$ *vs.* $r^2 = 0.9659$ and $s = 10.51$, respectively). Apparently, the parameter *x* better reflects those structural features of *n*-alkanes, which govern the T_m values. The *x* descriptor can be treated as a kind of the so-called topographic indices.[15] Because this parameter characterizes the TT...T-conformation, the results obtained substantiate the assumption that crystallization occurs in the minimum-energy conformation (see above).

The test set can be extended by including all known experimental data for the $C_{25}H_{44}$—$C_{100}H_{202}$ *n*-alkanes. Let us compare the predictive power of the models based on the *n* and *x* descriptors taking this set as an example (T_m values for a total of thirty-two hydrocarbons from this set are available[10,11]). The results are $r^2 = 0.8231$ and $s = 7.19$ for *n* and $r^2 = 0.8406$ and $s = 6.83$ for *x*. As can be seen, the *x* descriptor is somewhat better than the *n* descriptor (see Table 16). This is due to the fact that evenness of the *n* value plays a much less significant role for higher *n*-alkanes compared to lower *n*-alkanes.

The most interesting consequence of using the topographic index *x* is the possibility of constructing a model based on the *x* descriptor and relationship (18). The model allows the melting points of *n*-alkanes with even *n* ($n = 4, 6, \ldots, 14$, $r^2 = 0.9947$, $s = 3.97$) to be predicted from the melting points of *n*-alkanes with odd *n* ($n = 3, 5, \ldots, 15$; $r^2 = 0.9995$, $s = 1.66$) and *vice versa*. Similar modeling using the *n* descriptor is senseless, which again confirms that the *x* parameter much better reflects structural features of *n*-alkanes.

The final results of calculations of the *a*, *b*, *c* and *d* coefficients using relationship (18) and all fifty-four T_m values are listed in Table 18. Here, the model with the *x* descriptor gives $r^2 = 0.9988$ and $s = 2.36$ (*cf.* $r^2 = 0.9975$ and $s = 3.46$ for the model based on *n*). Of particular interest are the results of calculations of the limiting cases for expression (18) using the *x* descriptor. Unexpectedly, substituting $x = 0$ into this relationship (using the coefficients taken from Table 18) gives $T(0) = -274$ °C, which is very similar to the absolute zero of temperature (−273.15 °C). This allows one to reduce the number of parameters in relationship (18) using the following condition

$$f(x = 0) = -273.15 \text{ °C}. \qquad (30)$$

When using the *n* descriptor, relationship (18) gives a value $T(0) = -371.68$ °C, which has no physical meaning.

A more expected result is given below:

$$\lim f(x)|_{x \to +\infty} = d \approx 117.73 \text{ °C}. \qquad (31)$$

This means that the melting points of *n*-alkanes containing as large number of carbon atoms in the molecule as is wished will be no higher than 117.73 °C. One would assume that the melting point of polyethylene (PE) will approach this value, but T_m(PE) lies in the interval 105—115 °C. This is probably to the fact that half PE macromolecules contain the terminal double bonds and that individual PE macromolecules have different chain lengths (because of this, the compound can exhibit the properties of an eutectic mixture). Besides, experimentally produced PE macromolecules have long but finite chains. But even if one would obtain a pure *n*-alkane with $n > 1000$, such a compound would exhibit the properties of PE and have a T_m of about 118 °C. At the same time from this limiting relation it follows that there is no technological procedure suuitable for the synthesis of PE with

Table 18. Comparison of the results for modeling of the dependence of the melting points of *n*-alkanes on *x* with experimental data*

n	T_m/°C		*n*	T_m/°C	
	Experiment	Calculations		Experiment	Calculations
3	−187.65	−185.15	30	65.88	68.02
4	−138.33	−140.35	31	67.25	69.05
5	−129.72	−127.67	32	70.16	71.93
6	−95.32	−94.87	33	71.80	72.84
7	−90.60	−85.42	34	72.65	75.39
8	−56.80	−60.65	35	74.45	76.19
9	−53.53	−53.42	36	75.83	78.46
10	−29.67	−34.24	37	77.40	79.18
11	−25.65	−28.58	38	79.30	81.21
12	−9.60	−13.41	39	80.10	81.85
13	−6.00	−8.88	40	81.30	83.66
14	5.50	3.33	41	81.70	84.24
15	9.81	7.01	42	82.90	85.87
16	16.70	16.99	43	85.30	86.39
17	21.72	20.02	44	86.40	87.87
18	28.10	28.29	50	93.00	92.83
19	32.00	30.81	52	94.00	94.21
20	36.70	37.75	54	95.00	95.47
21	40.35	39.88	60	98.90	98.69
22	44.50	45.75	62	100.50	99.60
23	47.35	47.56	64	102.00	100.45
24	50.75	52.58	66	103.60	101.24
25	53.30	54.14	67	104.10	101.43
26	56.20	58.46	70	105.25	102.66
27	59.50	59.81	71	105.50	102.82
28	61.28	63.57	82	110.40	105.95
29	63.90	64.73	100	115.30	109.15

* Parameter *x* is calculated for a given *n* value using relationship (29), the parameters *a*, *b*, *c*, and *d* are equal to −25755, 9.73, 1.85, and 117.73, respectively.

$T_m > 117.73$ °C. Thus, the limiting relations like expression (31) can be used to predict certain properties of polymers, *e.g.*, T_m of polyethylenes. This is a crucially important issue, because solving this problem will permit the synthesis of polymers with improved heat resistance. For instance, replacement of every ninth methylene unit in PE by a Me group reduces the melting point of the polymer down to −2 °C.[22]

Table 19 lists the predicted T_m values of those *n*-alkanes with $n < 100$, for which no experimental data are available in the literature,[10,11] and of a number of heavier alkanes. The results of calculations for $n = 1000$ and $n = 10000$ closely approach the limiting value (31).

Now we will plot the melting points of 2-methylalkanes *vs.* T_m of corresponding *n*-alkanes (Fig. 9). We get a linear plot (the absolute term equals −15.9061, the slope is 1.0609; and $r^2 = 0.998716$). Thus, itis possible to construct a model for prediction of T_m of the simplest alkanes.

Table 19. Prediction of T_m using relationship (18) and the *x* descriptor for *n*-alkanes with unknown T_m

n	T_m/°C	*n*	T_m/°C	*n*	T_m/°C
45	88.34	73	103.45	90	107.57
46	89.68	74	103.90	91	107.66
47	90.11	75	104.04	92	107.92
48	91.33	76	104.46	93	108.01
49	91.72	77	104.60	94	108.25
51	93.19	78	104.99	95	108.34
53	94.53	79	105.12	96	108.57
55	95.77	80	105.49	97	108.65
56	96.63	81	105.61	98	108.87
57	96.90	83	106.07	99	108.94
58	97.70	84	106.39	120	111.37
59	97.95	85	106.50	150	113.35
61	98.92	86	106.81	200	115.05
63	99.82	87	106.91	500	117.20
65	100.65	88	107.20	1000	117.57
68	101.97	89	107.30	10000	117.72
69	102.15				

Finally, we propose a hypothesis concerning the problem of abnormally high T_m values of some highly branched polymethylalkanes. The case in point is that the melting points of three out of four compounds are much higher than those of isomeric *n*-alkanes (Table 20).

The use of the $\bar{S}-\bar{V}$ diagram (Fig. 10) for alkanes allows one to clarify the problem. Two groups of compounds, which do not fall on the common curve, are clearly seen. One group includes the alkanes containing a dustant *tert*-butyl group. The corresponding points are grouped and form a "cluster" in the middle part of the plot. The other group of compounds comprises neopentane, its derivatives, and hexamethylethane; here, no "cluster" is formed, the corresponding points being outliers. Probably, this is a reason for abnormally high melting points of these compounds.

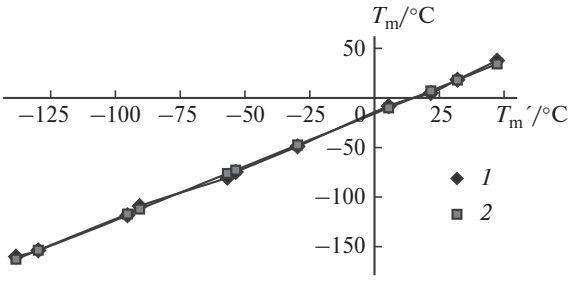

Fig. 9. Melting points of methylalkanes (T_m) plotted *vs.* melting points of corresponding *n*-alkanes (T_m'): experimental data (*1*) and results of calculations (*2*).

Table 20. Anomalies of alkane melting points (T_m)

Compound	T_m/°C		
	Alkane	Isomeric *n*-alkane	Other isomers
 Neopentane	−16.55	−129.723	−159.890
 Methylneopentane	−99.73	−95.320	−153.680——128.41
 Pentamethylethane	−24.93	−90.605	−134.46——118.27
 Hexamethylethane	+100.81	−56.798	−126.10——90.870

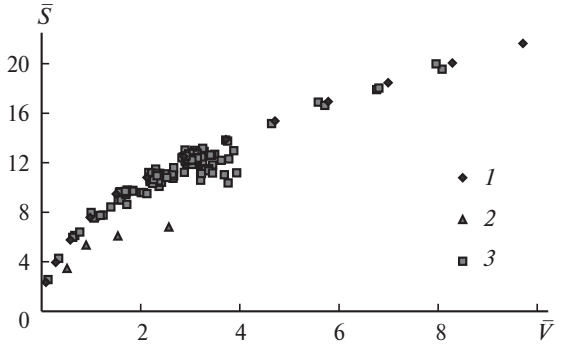

Fig. 10. The $\bar{S}—\bar{V}$ diagrams for *n*-alkanes (*1*), alkanes with abnormally high T_m (*2*), and branched alkanes (*3*) (calculated in the approximation No. 2).

* * *

Thus, it seems that the topographic indices (W_{3D}, l_a, l_f, l_{max}, $\bar{S}$, $\bar{V}$, and x) based on the physical distance more correctly describe the molecular structure compared to the W and n parameters based on the topological distance. Apparently, numerous topological indices, which ignore the spatial arrangement of atoms in molecules, are of limited use in the QSPR studies compared to the topographic indices.

This work was carried out with partial funancial support from the Chemistry and Materials Science Division of the Russian Academy of Sciences (in the framework of the Program No. 1).

References

1. D. H. Rouvray, *Math. Chem.* (*MATCH*), 1975, **1**, 61.
2. A. T. Balaban, *Chemical Applications of Graph Theory*, Academic Press, New York—London, 1979.
3. *Chemical Applications of Topology and Graph Theory*, Ed. R. B. King, Elsevier, Amsterdam—New York, 1983.
4. M. I. Stankevich, I. V. Stankevich, and N. S. Zefirov, *Usp. Khim.*, 1988, **57**, 337 [*Russ. Chem. Rev.*, 1988, **57** (Engl. Transl.)].
5. E. A. Smolenskii, *Dokl. Akad. Nauk*, 1999, **365**, 767 [*Dokl. Chem.*, 1999 (Engl. Transl.)].
6. J. G. Topliss and R. P. Edwards, *Med. Chem.*, 1979, **22**, 1238.
7. D. E. Needham, I. C. Wei, and P. C. Seybold, *J. Am. Chem. Soc.*, 1988, **110**, 4186.
8. E. A. Smolenskii, A. N. Ryzhov, A. L. Lapidus, and N. S. Zefirov, *Dokl. Akad. Nauk*, 2002, **387**, 69 [*Dokl. Chem.*, 2002 (Engl. Transl.)].
9. E. A. Smolenskii, L. K. Maslova, S. A. Shpil´kin, and N. S. Zefirov, *Dokl. Akad. Nauk*, 1999, **368**, 219 [*Dokl. Chem.*, 1999 (Engl. Transl.)].
10. R. D. Obolentsev, *Fizicheskie konstanty uglevodorodov* [*Physical Constants of Hydrocarbons*], Gos. Nauchno-Tekh. Izd. Neftyanoi i Gorno-Topl. Lit-ry, Moscow—Leningrad, 1953, 446 pp. (in Russian).
11. *Fiziko-khimicheskie svoistva individual´nykh uglevodorodov* [*Physicochemical Properties of Individual Hydrocarbons*], pod red. V. M. Tatevskogo, Gos. Nauchno-Tekh. Izd. Neftyanoi i Gorno-Topl. Lit-ry, Moscow, 1960, 412 pp. (in Russian).
12. H. Wiener, *J. Am. Chem. Soc.*, 1947, **69**, 17.
13. H. Wiener, *J. Am. Chem. Soc.*, 1947, **69**, 2636.
14. Yu. G. Papulov, *Stroenie molekul* [*Molecular Structure*], TGU, Tver´, 1994, 200 pp. (in Russian).
15. Z. Mihalic and N. Trinajstic, *J. Mol. Struct. (THEOCHEM)*, 1991, **232**, 65.
16. E. A. Smolenskii, *J. Chem. Inf. Comput. Sci.*, 2004, **44**, 522.
17. E. A. Smolenskii, *Dokl. Akad. Nauk SSSR*, 1976, **30**, 373 [*Dokl. Chem.*, 1976 (Engl. Transl.)].
18. A. Abe, R. L. Jernigan, and P. J. Flory, *J. Am. Chem. Soc.*, 1966, **88**, 631.
19. V. N. Drozd, N. S. Zefirov, V. I. Sokolov, and I. V. Stankevich, *Zh. Org. Khim.*, 1979, **15**, 1785 [*J. Org. Chem. USSR*, 1979, **15** (Engl. Transl.)].
20. N. S. Zefirov, V. N. Drozd, V. I. Sokolov, and I. V. Stankevich, *Zh. Org. Khim.*, 1981, **17**, 233 [*J. Org. Chem. USSR*, 1981, **17** (Engl. Transl.)].
21. A. R. Katritzky, U. Maran, M. Karelson, and V. S. Lobanov, *J. Chem. Inf. Comput. Sci.*, 1997, **37**, 913.
22. D. Bonchev, E. Markel, and A. Dekmezian, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 1274.

Received December 24, 2004;
in revised form April 1, 2005