Russian Journal of General Chemistry, Vol. 71, No. 6, 2001, pp. 833–838. Translated from Zhurnal Obshchei Khimii, Vol. 71, No. 6, 2001, pp. 889–894.

Structure–Property Correlation Equation: VI.[1] Accessible Surface Areas of Hydrocarbon Molecules and Simple Procedures for Their Approximate Estimation

I. B. Golovanov and I. G. Tsygankova

Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, Pushchino, Moscow oblast, Russia

Received August 16, 1999

Abstract — The molecule surface area accessible for a solvent is considered as a property apparently related to the molecular conformation. With this "ideal" property as example, the methodological aspects are considered of choosing an analytical expression for the structure–property correlation equation and of determining the parameters of this equation.

At present, many molecular properties cannot be calculated without using empirical data, and in the overwhelming majority of cases they are calculated approximately using structure–property correlation equations chosen, at best, semiempirically. As a rule, such equations contain certain parameters to be determined from experimental data.

However, experimental data on systems under consideration are often insufficient, and it becomes necessary to make reasonable assumptions. For example, any data on molecular properties refer to an equilibrium mixture of various conformers. In most cases their relative content is unknown, and even if such data are available, the contributions of particular conformers to a property should be known, etc.

Such problems could be avoided if it were possible to determine a certain property for fixed conformation of any set of molecules. In this case, a directed experiment could be performed, based on reasonable assumptions, to reveal the physicochemical sense of parameters entering into the structure–property correlation equation and find the ways to evaluate them.

However, actually there are no such systems, but there are properties that can be calculated with a high accuracy for any fixed molecular conformation, e.g., accessible surface area (*S*). The accessible surface area is the area of the surface circumscribed by the center of a test sphere of radius 1.4 Å (simulating a solvent molecule) rolling over the surface of the test molecule. It can be calculated fairly accurately for specific molecular conformations and used for choosing the analytical form of the structure–property correlation equation and determining its parameters.

There are reasonable hypotheses that the accessible surface area determines such molecular characteristics as the heats of vaporization, boiling points, free energies of transfer from the gas phase to water, and distribution factors between the organic phase and water [2–9]. It turns out that the accessible surface area correlates well with a large number of molecular characteristics. Below are given as examples the factors *r* characterizing correlation of molecular properties of hydrocarbons with their accessible surface areas (the correlations include the molecules given in the first column of Table 1; data on properties are taken from [9]; designations: α', polarizability; χ, diamagnetic susceptibility; T_b, boiling point; P_c, critical pressure; V_c, critical volume; T_c, critical temperature, c_p, heat capacity; *S*, entropy of formation; ΔG, free energy of formation; and ΔH, enthalpy of formation). It is seen that data on the accessible surface areas of hydrocarbon molecules allow evaluation of a number of their characteristics; therefore, we chose as "ideal" property the accessible surface areas of molecules.

α'	0.9986	T_c	0.9443
χ	0.9822	c_p	0.9724
T_b	0.9722	S	0.9966
P_c	0.9544	ΔG	0.9551
V_c	0.9972	ΔH	–0.9513

The procedures for calculating accessible surface areas are described elsewhere [3, 10, 11]. Here we used the program based on the approach in [10].

For approximate reproduction of the *S* values, we will use the following relation [12, 13]:

[1] For communiction V, see [1].

1070-3632/01/7106-0833 $25.00

Table 1. Accessible surface areas of some hydrocarbon molecules

Molecule	Conformation	Experimental value	Approximate estimate[a]	Molecule	Conformation	Experimental value	Approximate estimate[a]
CH_4		153.9	154.1	C_7H_{16}	*tttt*	338.0	337.6
C_2H_6		188.6	188.3	C_7H_{16}	*tttg*	331.8	331.6
C_3H_8		218.1	218.2	C_7H_{16}	*tgtt*	332.5	330.7
$(CH_3)_3CH$		243.5	243.7	C_7H_{16}	*tgtg*	327.3	325.3
$(CH_3)_4C$		264.9	264.8	C_7H_{16}	*g*tgt*	328.3	326.3
$H_3C–C(CH_3)_2–C_2H_5$	*t*	286.7	286.9	C_7H_{16}	*gttg*	326.6	325.5
$H_3C–C(CH_3)_2–C_2H_5$	*c*	286.1	286.8	C_7H_{16}	*gtg*t*	327.9	326.3
$H_3C–C(CH_3)_2CH(CH_3)CH_3$	*t*	304.9	304.6	C_7H_{16}	*ttgg*	326.2	324.3
$H_3C–C(CH_3)_2CH(CH_3)CH_3$	*c*	303.6	304.4	C_7H_{16}	*gggg*	315.5	314.2
$H_3C–C(CH_3)_2C(CH_3)_2CH_3$	*t*	319.9	317.9	C_7H_{16}	*g*ttg*	326.3	325.0
$H_3C–C(CH_3)_2C(CH_3)_2CH_3$	*c*	317.5	317.7	C_7H_{16}	*tggt*	324.8	323.3
C_4H_{10}	*t*	248.6	248.0	C_7H_{16}	*tgg*t*	317.9	320.8
C_4H_{10}	*g*	242.9	244.2	$C_2H_5–C(Et)_2C_2H_5$	*t*	346.7	348.4
C_5H_{12}	*tt*	278.1	277.9	$C_2H_5–C(Et)_2C_2H_5$	*g*	348.5	348.6
C_5H_{12}	*tg*	273.0	273.5	C_8H_{18}	*ttttt*	368.5	367.5
C_5H_{12}	*gg*	266.4	269.0	C_8H_{18}	*tggtg*	344.2	343.0
C_5H_{12}	*gg**	259.2	268.1	C_9H_{20}	*tttttt*	398.0	397.4
C_6H_{14}	*ttt*	308.5	307.8	C_9H_{20}	*tggtgg*	356.1	358.3
C_6H_{14}	*ttg*	302.9	302.5	C_9H_{20}	*ttggtt*	383.6	376.0
C_6H_{14}	*tgt*	303.4	302.4	C_9H_{20}	*ttgg*tt*	368.9	368.0
C_6H_{14}	*gtg*	296.9	297.5	C_9H_{20}	*gtggtg*	362.8	360.6
C_6H_{14}	*tgg*	296.3	296.9	C_9H_{20}	*gtgg*tg*	364.6	361.3
C_6H_{14}	*g*tg*	298.4	298.1	$C_{12}H_{26}$	*tttttt*	488.5	486.9
C_6H_{14}	*ggg*	291.5	292.4	$C_{12}H_{26}$	*tggtgg*	392.8	394.8
C_6H_{14}	*gg*t*	308.5	307.8				

[a] The accessible surface area was calculated by the equation $S = n\alpha + (n-1)\beta + n_{1\cdots3}\gamma + b_1\sum_{i=1}^{n-4}(3.86 - R_{i,i+4}) + b_2\sum_{i=1}^{n-5}(5.0 - R_{i,i+5}) + \sum_{i=1}^{n-6}(6.32 - R_{i,i+6}) + \sum_{i=7}^{n-1}\sum_{i=1}^{n-1}(7.5 - R_{i,i+l})$; α 154.1 ± 0.2812, β –119.8 ± 0.3525, γ –4.37 ± 0.0775, b_1 –4.21 ± 0.015, b_2 –0.81 ± 0.0085. Correlation factor r 0.9993, rms deviation s 0.2990.

$$S = \Sigma S_{ii} + \Sigma\Sigma S_{ij}, \quad (1)$$

where S_{ii} is the contribution of the ith molecular fragment and S_{ij} is the contribution corresponding to interaction of the ith and jth fragments. The goal of this study is to find simple methods for evaluating these contributions.

In what follows we will consider accessible surface areas for hydrocarbon molecules with "erased" H atoms. The C atoms, CC bond lengths, and CCC bond angles will be assumed to be equivalent. Then $S_{ii} = S_{jj} = \alpha$, $S_{i,i+1} = S_{j,j+1} = \beta$, $S_{i,i+2} = S_{j,j+2} = \gamma$ will be constant, and the contributions corresponding to interaction of more remote fragments $\delta_{i,i+k}$ (k = 3, 4, 5...) will depend on the molecular conformation, i.e., on the distance between the fragments i and $i + k$.

Apparently, having calculated S for a series of molecules C_n is the *trans* conformations and having found the first differences $\Delta_1 = S_{C_{n+1}} - S_{C_n}$, we can estimate various contributions S_{ij}. Table 2 gives as example the data for a series of linear hydrocarbons in the *trans* conformation. Since the first differences starting from C_3 are practically equal, all the contributions of type S_{1m} starting from m = 4 are zero. Hence, to estimate S for the *trans* conformers, it is sufficient to determine α, β, and γ. The set of test examples for determining these parameters should include CH_4, C_2H_6, C_3H_8, $CH(CH_3)_3$, and $C(CH_3)_4$. In the two latter molecules the number of contributions γ is increased relative to the linear isomers, and their inclusion into the set of test examples allows more accurate evaluation of γ. From these data we obtained α = 154.06, β = –119.82, and γ = 4.37.

Table 2. Accessible surface areas of linear hydrocarbon molecules

Molecule	Representation of S by relation (3)	Δ_1	S_{exp}, Å^2	Δ_1
CH_4	α		153.9	
C_2H_6	$2\alpha + \beta$	$\alpha + \beta$	188.6	34.7
C_3H_7	$3\alpha + 2\beta + \gamma$	$\alpha + 2\beta + \gamma$	218.1	29.5
n-C_4H_{10}	$4\alpha + 3\beta + 2\gamma + \delta_{14}^{trans}$	$\alpha + \beta + \gamma + \delta_{14}^{trans}$	248.6	30.5
n-C_5H_{12}	$5\alpha + 4\beta + 3\gamma + 2\delta_{14}^{trans} + \delta_{15}^{trans}$	$\alpha + \beta + \gamma + \delta_{14}^{trans} + \delta_{15}^{trans}$	278.1	29.5
n-C_6H_{14}	$6\alpha + 5\beta + 4\gamma + 3\delta_{14}^{trans} + 2\delta_{15}^{trans} + ...$	$\alpha + \beta + \gamma + \delta_{14}^{trans} + \delta_{15}^{trans}$	308.0	30.4
n-C_7H_{16}	$7\alpha + 6\beta + 5\gamma + 4\delta_{14}^{trans} + 3\delta_{15}^{trans} + ...$	$\alpha + \beta + \gamma + \delta_{14}^{trans} + \delta_{15}^{trans} + \delta_{16}^{trans}$	338.0	29.5
n-C_8H_{18}	$8\alpha + 7\beta + 6\gamma + 5\delta_{14}^{trans} + 4\delta_{15}^{trans} + ...$	$\alpha + \beta + \gamma + \delta_{14}^{trans} + \delta_{15}^{trans} + \delta_{16}^{trans} + \delta_{17}^{trans}$	368.5	29.5

To calculate S by relation (1), it is necessary to find the dependence of S_{1m} on the distance between the fragments. To maximally simplify the problem, we can first obtain separate expressions for each m. For example, to determine S_{14}, it is appropriate to consider the following set of test examples: *trans*-*n*-C_4H_{10}, *gauche*-*n*-C_4H_{10}, 2,2-dimethylbutane, 2,2,3-trimethylbutane, and 2,2,3,3-tetramethylbutane. Although the latter three molecules under equilibrium conditions exist in the form of a single (*staggered*) conformation, it is quite feasible to calculate the accessible surface areas for the other (e.g., *eclipsed*) conformations produced by rotation around the central CC bond. Experimental data for these molecules show that the conformers that are more compact than the *trans* conformers have smaller accessible surface areas. For butane, according to (1), the following relations are valid:

$$S^{trans} = 4\alpha + 3\beta + 2\gamma + \delta_{14}^{trans},$$

$$S^{gauche} = 4\alpha + 3\beta + 2\gamma + \delta_{14}^{gauche}.$$

Since the value of S_{14}^{trans} corresponding to the maximal distance between fragments 1 and 4 (R_{14}) is zero, in any other conformation with smaller R_{14} a negative S_{14} should be observed. The plot of S_{14} vs. R_{14} is shown in the figure. At $b = -4.209$, the experimental values are reproduced to the third decimal place.

Similar data treatment with a larger set of parameters (and experimental data for their determination) will result in a very good agreement of the calculated values with the experiment.

However, the experimental data required to estimate particular properties are few, and therefore the parameters involved should also be few. For estimating S, let us assume that the contributions corresponding to interaction of remote (m = 4, 5...) fragments are defined similarly:

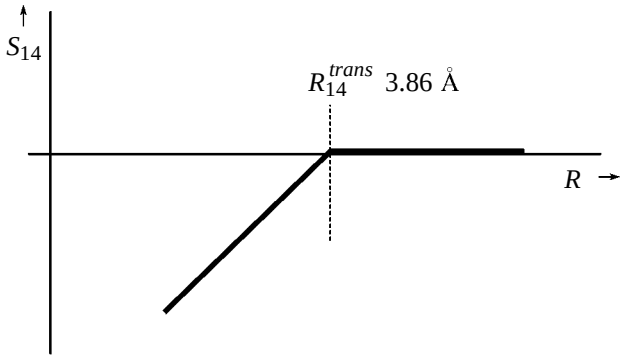
$$S_{1m} = \begin{cases} k_m(R_{1m}^{trans} - R_{1m}) \text{ at } R_{1m} < R_{1m}^{trans}, \\ 0 \text{ at } R_{1m} > R_{1m}^{trans}. \end{cases} \quad (2)$$

The values of R_{1m}^{trans} are determined on the basis of geometric considerations: R_{14}^{trans} = 3.86, R_{15}^{trans} = 5.0, R_{16}^{trans} = 6.32 Å; for all $m \geq 7$ R_{1m}^{trans} = 7.50 Å. The parameter $b_1 = k_4$ for the contribution 1···4 is –4.209. All the contributions from more remote fragments can be reproduced with a single parameter $b_2 = -0.81$.

Apparently, such an approach (estimation of the contributions α, β, and γ from data for a series of small molecules and then choice of the contribution $P_{1\cdots m}$ in the form of a function of distance between the first and mth fragments) is general. Practically all the problems are related to the choice of the function $P_{1\cdots m} = f(R_{1m})$ and determination of its parameters. Available separate experimental data for "large" molecules suggest the type of this function. An example is given in [13].

Let us illustrate the possibilities of this approach to estimation of the accessible surface areas of complexes of hydrocarbon molecules.

Complexes consisting of hydrocarbon molecules can be considered as supermolecules; however, in this case it is difficult to determine the corrections for intermolecular interactions which are complex functionals depending on the mutual arrangement of the interacting fragments and their neighbors. Since, when a complex **AB** is formed, the accessible surface area decreases by the quantity $\Delta S_{\mathbf{AB}} = S_{\mathbf{AB}} - S_{\mathbf{A}} - S_{\mathbf{B}}$ (where $S_{\mathbf{AB}}$, $S_{\mathbf{A}}$, and $S_{\mathbf{B}}$ are the accessible surface areas of the complex and its components, respectively), it is reasonable to look for correlation between ΔS and the sum of contributions from intermolecular interactions of the fragments of molecules composing the complex.



Contribution S_{14} vs. distance between fragments 1 and 4.

$$S_{14} = \begin{cases} b(R_{14} - 3.86) \text{ at } R_{14} < 3.86, \\ 0 \text{ at } R_{14} > 3.86. \end{cases}$$

The value of $\Delta S_{\mathbf{AB}}$ will be different for different types of complexes. First, data for complexes $C_1 \cdots C_1$ with various distances between the centers show that the corrections for intermolecular interactions of type 1–2, hereinafter denoted as β*, have a form similar to that shown in the figure:

$$\beta^* = \begin{cases} 22(R_{ij} - 7.0) \text{ at } R_j < 7.0 \text{ Å}, \\ 0 \text{ at } R_{ij} > 7.0 \text{ Å}. \end{cases}$$

At $R_{12} \geq 7.0$ Å, a water molecule can be accommodated between the centers, and $\Delta S = 0$. As R_{12} decreases by 1 Å, $\Delta S_{\mathbf{AB}}$ decreases by ~22 Å^2. The same result is obtained as R_{12} decreases to zero (when one molecule is "pressed" into the other), and $\Delta S_{\mathbf{AB}}$ becomes equal to the accessible surface area of the methane molecule.

In all linear complexes of types $C_1 \cdots C_n$ and $C_n \cdots C_n$ (n = 1–8) ΔS is the same; therefore, we should assume that only the contributions β* (of type 1–2) are significant, and the more remote contributions (γ*, corresponding to 1–3 contribution) at such distances can be neglected.

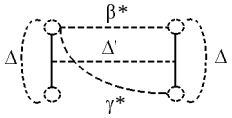
In perpendicular complexes of type $C_1 \cdots C_n$ the values of $\Delta S_{\mathbf{AB}}$ differ from those in linear complexes. These differences can be described as contributions from interaction with the second C atom of the C_2 molecule and correction for the fact that these C atoms are not independent and form a bond. The differences between these complexes can be schematically represented as follows.

Complex $C_1 \cdots C_1$ C--β*--C; $\nu^* = 0$ if $R_{13} > R$

Linear complex $C_1 \cdots C_2$ C---C—C, R_0

Perpendicular complex $C_1 \cdots C_2$ C, β*, C, Δ, β*, C

In the perpendicular complex $C_1 \cdots C_2$ the two centers of the C_2 molecule are linked with each other; taking into account the correction for their interaction, $\Delta S_{C_1 \cdots C_2} = 2\beta^* + \Delta$. To estimate the value of $\Delta S_{C_2 \cdots C_2}$ in the parallel complex, these contributions are insufficient. Then, it is appropriate to introduce another correction Δ' for interaction between the bonds, determined by the distance between the bond centers.



This is already sufficient for describing $\Delta S_{\mathbf{AB}}$ in parallel complexes $C_n \cdots C_n$. For such complexes we obtain

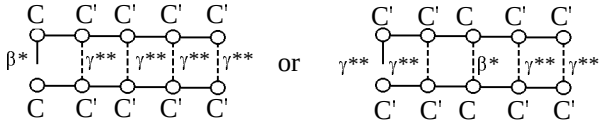
$$\Delta S_{C_n \cdots C_n} = n\beta^* + 2(n - 31)\gamma^* + 2(n - 1)\Delta + (n - 1)\Delta'. \quad (3)$$

It can be readily seen that the first differences for the complexes $C_n \cdots C_n$ are equal; this means that the value of $\Delta S_{\mathbf{AB}}$ is determined by only two significant parameters. It follows from relation (3) that $\Delta S_{C_n \cdots C_n}$ can be represented as

$$\Delta S_{C_n \cdots C_n} = \beta^* + (n - 1)[\beta^* + 2\gamma^* + 2\Delta + \Delta'] = \beta^* + (n - 1)\gamma^{**}. \quad (3')$$

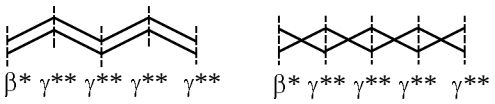
Here $\gamma^{**} = \beta^* + 2\gamma^* + 2\Delta + \Delta'$, and for equilibrium distances (R_{CC} = 4.0 Å) γ^{**} is equal to –20 Å^2.

Relation (3') can be interpreted as follows: (1) The molecule C_n is represented by two types of centers: center C and (n – 1) centers C' (the position of C in the chain is arbitrary); (2) in the parallel complex the position can be represented as



In this case, $\Delta S_{\mathbf{AB}}$ is determined by the sum of independent interactions C···C and C'···C''.

Since the quantities β* and γ** can be readily determined, in terms of this approach it is easy to estimate $\Delta S_{\mathbf{AB}}$. Indeed, the parallel complexes $C_n \cdots C_n$ can be of two types:



However, since the respective distances $C_i \cdots C_i$ (denoted by dashed lines in the scheme) are approximately equal in both types of complexes, the contributions β* and γ** and hence the values of $\Delta S_{\mathbf{AB}}$ will also be equal. The difference between the $\Delta S_{\mathbf{AB}}$ values for these complexes is very small (for $C_8 \cdots C_8$, it is about 1 Å^2). In many cases such differences can be neglected, and it can be assumed that in any parallel complexes $C_n \cdots C_n$ the values of $\Delta S_{\mathbf{AB}}$ are equal and can be calculated by relation (3').

We can formulate also some other useful rules based on relation (3'). For example, it is obvious that $\Delta S_{\mathbf{AB}}$ for a parallel complex $C_n \cdots C_n$ can be estimated from data on complexes $C_{n-1} \cdots C_{n-1}$ and $C_{n+1} \cdots C_{n+1}$:

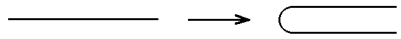
$$\Delta S_{C_n \cdots C_n} = 0.5(\Delta S_{C_{n-1} \cdots C_{n-1}} + \Delta S_{C_{n+1} \cdots C_{n+1}}). \quad (4)$$

For example, from data for complexes $C_1 \cdots C_1$ and $C_2 \cdots C_2$ we obtain $\Delta S_{C_1 \cdots C_2} = -78$ Å^2, whereas estimation by Eq. (4) gives 74 Å^2. In many other cases the agreement is still better. For example, $\Delta S_{C_4 \cdots C_4}$ is equal to –125 Å^2, and estimation by Eq. (4) gives –126 Å^2, etc.

These simple approaches supplemented by obvious reasonings allow estimation of ΔS for complexes of arbitrary configuration.

Consider, for example, cross-shaped complexes in which the molecules are located in different planes and their axes are perpendicular. At large *n*, the molecule can be represented as a folded cylinder  with the diameter equal to 7 Å^2 ($2r_C$ + $2r_{H_2O}$), which coincides with the length of the C_7 molecule. Therefore, the maximal value of $\Delta S_{C_n \cdots C_n}$ in perpendicular complexes should be attained for molecules containing seven carbon atoms, and with increasing their number ΔS should not change, since the fragments that lie beyond the crossing site are accessible for interaction with water. Assuming that in such complexes $\Delta S_{\mathbf{AB}}$ varies linearly from –66 ($C_1 \cdots C_1$) to –140 Å^2 ($C_7 \cdots C_7$), we can estimate $\Delta S_{\mathbf{AB}}$ fairly accurately. For example, for $C_5 \cdots C_5$ the calculated value is –130 Å^2, whereas the estimation gives –118 Å^2.

Such reasoning is very useful when estimating ΔS for the case of transition of a polymeric molecule C_n from one conformation to another. Consider, e.g., transition of a C_n molecule from the *trans* conformation to a *hair-pin*-shaped conformation:



The curved part of the pin consists of six CH_2 groups, and the remaining groups form parallel sections structurally similar to the complexes $C_m \cdots C_m$ in which all the atoms are C' centers and the C center is located in the curved part. Apparently, the change in the accessible surface area in going from the *trans* to *hair-pin* conformation is approximately equal to $m\gamma^{**}$ (e.g., 93 Å^2 for C_{14}). The linear sections of the pin in the case of C_{14} contain four C fragments each; therefore, $\Delta S \sim 4\gamma^{**} \sim 80$ Å^2. For the dimer $C_4 \cdots C_4$, ΔS is 126 Å^2; however, in this dimer one of the interactions (β*) significantly differs from the other interactions, which is not the case in the pin. An additional empirical correction can be introduced for more accurate estimation of ΔS. In this case, for transition from the *trans* to *hair-pin* conformation we obtain

$$\Delta S = m\gamma^{**} + 13. \quad (5)$$

Using Eq. (5), we obtain a very good agreement with the accurately calculated ΔS values for C_{16}, C_{18}, C_{20}, and C_{22}.

Since the accessible surface areas of hydrocarbon molecules correlate with many of their properties, approximate approaches used for their estimation will be in many cases similar to approaches used for estimating the other properties. Such a property as the accessible surface area can be chosen as a criterion for testing various relations for determining the contributions of S_{ii} and S_{ij}.

REFERENCES

1. Golovanov, I.B. and Tsygankova, I.G., *Zh. Obshch. Khim.*, 2001, vol. 71, no. 5, pp. 759–764.
2. Hildebrand, J.H., Prausnitz, J.M., and Scott, R.L., *Regular and Related Solutions,* New York: Van Norstrand Reinhold, 1970.
3. Hermann, R.B., *J. Phys. Chem.*, 1972, vol. 76, no. 19, pp. 2754–2759.
4. Hermann, R.B., *J. Phys. Chem.*, 1975, vol. 79, no. 2, pp. 163–169.
5. Chotia, C., *Nature,* 1974, vol. 248, no. 5446, pp. 338–339.

6. Reinolds, J.A., Gilbert, D.B., and Tanford, C., *Proc. Natl. Acad. Sci. USA*, 1974, vol. 71, pp. 2925–2927.
7. Einsberg, D. and McLachlan, A.D., *Nature*, 1986, vol. 319, no. 6050, pp. 199–203.
8. Carlson, H.A. and Jorgensen, W.L., *J. Phys. Chem.*, 1995, vol. 99, no. 26, pp. 10667–10673.
9. Ashbaugh, H.S., Kaler, E.W., and Paulaitis, M.E., *Biophys. J.*, 1998, vol. 75, no. 2, pp. 755–768.
10. Lee, V. and Richards, F.M., *J. Mol. Biol.*, 1971, vol. 55, no. 3, pp. 379–400.
11. Nauchitel', V.V., Golovanov, I.B., and Vol'kenshtein, M.V., *Dokl. Akad. Nauk SSSR*, 1988, vol. 302, no. 1, pp. 138–142.
12. Golovanov, I.B., Ivanitskii, G.R., and Tsygankova, I.G., *Dokl. Ross. Akad. Nauk*, 1998, vol. 359, no. 2, pp. 258–262.
13. Golovanov, I.B. and Tsygankova, I.G., *Zh. Obshch. Khim.*, 1999, vol. 69, no. 8, pp. 1275–1282.