Quantitative relationships for the prediction of the vapor pressure of some hydrocarbons from the van der Waals molecular surface

TUDOR OLARIU[1]*, VICENŢIU VLAIA[1], CIPRIAN CIUBOTARIU[2], DAN DRAGOŞ[1], DAN CIUBOTARIU[1] and MIRCEA MRACEC[3]**

[1]*Department of Organic Chemistry, Faculty of Pharmacy, "Victor Babes" University of Medicine and Pharmacy, P-ta Eftimie Murgu No. 2, 300041, Timisoara, Romania,* [2]*Department of Computer Sciences, University "Politehnica", P-ta Victoriei No. 2, 300006, Timisoara, Romania and* [3]*Molecular Forecast Research Center, Prof. Dr. Aurel Păunescu-Podeanu str. 125, A 4, RO-300588 Timişoara, Romania*

Abstract: A quantitative structure–property relationship (QSPR) modeling of vapor pressure at 298.15 K, expressed as log (VP / Pa) was performed for a series of 84 hydrocarbons (63 alkanes and 21 cycloalkanes) using the van der Waals (vdW) surface area, S^W / Å^2, calculated by the Monte Carlo method, as the molecular descriptor. The QSPR model developed from the subset of 63 alkanes (C_1–C_{16}), deemed as the training set, was successfully used for the prediction of the log (VP / Pa) values of the 21 cycloalkanes, which was the external prediction (test) subset. A QSPR model was also developed for a series composed of all 84 hydrocarbons. Both QSPR models were statistically tested for their ability to fit the data and for prediction. The results showed that the vdW molecular surface used as molecular descriptor (MD) explains the variance of the majority of the log (VP / Pa) values in this series of 84 hydro-carbons. This MD describes very well the intermolecular forces that hold neut-ral molecules together. The clear physical meaning of the molecular surface values, S^W / Å^2, could explain the success of the QSPR models obtained with a single structural molecular descriptor.

Keywords: MD – molecular descriptor; van der Waals molecular surface; Monte Carlo method; QSPR.

INTRODUCTION

Equilibrium vapor pressure is defined as the pressure exerted by a vapor in thermodynamic equilibrium with its condensed phase (solid or liquid) in a closed system under standard conditions (298.18 K and p = 100 kPa).

659

Vapor pressure, VP, plays an important role in the study of the environmental fate, transport and distribution of the compounds in water, air and soil, as well as for engineers designing chemical processes.[1] The distribution of environmental pollutants between the soil and the atmosphere is determined by their VP. It is also used for calculating other physicochemical properties, such as liquid viscosity, enthalpy of vaporization, air–water partition coefficients, rates of evaporation and distribution coefficients for adsorption on soil.[2]

Experimental vapor pressure data are abundant for low molecular weight hydrocarbons, but the measurements can be difficult to realize, and even unreliable in certain ranges. If the experimental data are missing, the only option is to predict the vapor pressure of various compounds using a quantitative structure–property relationship (QSPR).[3] Another difficulty is related to the fact that experimental VP measurement of actual chemical products is both time-consuming and expensive. The equations of state for the calculation of vapor pressure exist, but these typically require experimental determinations at a minimum of two temperatures. Most reported QSPR studies to predict vapor pressure still depend on various empirical data, such as boiling points, critical pressures, and critical temperatures[1] but, unfortunately, boiling points as well as critical properties are often lacking, especially for heavy hydrocarbons many of them are thermally unstable in the critical region.[2] Therefore, QSPR is an alternative approach for estimating vapor pressure based on the premise that physicochemical properties can be correlated with molecular structural characteristics expressed in terms of appropriate molecular descriptors. When a quantitative structure–property relationship is found, it might also provide insight into which aspect of the molecular structures influences the property. Such insight may facilitate a systematic approach to the design of new molecules with more desirable properties.[1]

Earlier, a QSPR study of boiling points of a series of alkanes based on comparative analysis of the molecular vdW space and topological space of these compounds with the aid of 36 molecular descriptors: 16 generalized topological distance matrix indices (GTDIs), 11 topological distance indices (TDIs) known in the literature (seven obtained from eigenvalues/eigenvectors of distance matrix), and 9 van der Waals molecular descriptors has been reported.[4]

Herein, it is supposed that the molecular van der Waals surfaces of hydrocarbons are responsible of their vapor pressure values. Intermolecular forces hold neutral molecule to each other. These forces involve attraction of charges of opposite signs. The attractive forces between the molecules of a non-polar compound, such as alkanes and cycloalkanes, are called van der Waals (vdW) forces. They are of the induced dipole–induced dipole type (dispersion forces),[4] and they act only at very short distances, between the parts of different molecules that are in close contact, *i.e.*, between the surfaces of molecules. These vdW forces holding non-polar molecules together are weak and of very short range. Therefore,

within the class of hydrocarbons, the larger is the molecule, the larger is its surface, the stronger are the intermolecular forces, and, hence, the lower will the VP value be.

The aim of this paper is to propose a statistically validated QSPR model for vapor pressure prediction of 84 saturated hydrocarbons (63 alkanes and 21 cycloalkanes) and to develop a QSPR model for the series 63 alkanes that will be used to predict the log (VP / Pa) values for the series of 21 cycloalkanes. The models were based mainly of the vdW molecular surface, S^W / Å^2, which was used as a unique molecular descriptor. Regression analysis[5] was applied for the development of the QSPR models log (VP / Pa) *vs.* S^W / Å^2. The S^W values of these hydrocarbons were estimated with IRS software.[6] IRS (Investigation Receptor Space) implements numerical algorithms developed based on the Monte Carlo (MC) method for computation of various vdW molecular characteristics – the so-called molecular descriptors (MDs) – used for the study of relationships between the molecular structure and a variety of properties of chemical compounds.[4,7–17]

EXPERIMENTAL DATA

The experimental data used in this QSPR study were taken from the literature.[18] The values of experimental vapor pressures at 298.15 K were expressed as log (VP / Pa). The structures included 63 alkanes and 21 cycloalkanes. The data set containing the 63 alkanes was used as a training set (see Table S-I of the Supplementary material to this paper). In order to validate the predictive ability of the training set model (see, below, relation (7)), the series of 21 cycloalkanes (Table S-I) was used as an external prediction set. The values of log (VP / / Pa) for all hydrocarbons are systematized in Table S-I, together with the corresponding vdW surface, S^W / Å^2, values calculated by means of the Monte Carlo method with IRS software.[6] Table S-I also contains the predicted log (VP / Pa) values with the linear Model (7) and the errors, Δ, calculated as difference between experimental and the corresponding predicted log (VP / Pa) values.

Methods

The analysis of the relationship between the vapor pressure of the hydrocarbons in Table S-I and their molecular structure was performed using regression analysis as a statistical tool for the development of reliable QSPR models.

Below, the Monte Carlo integration algorithm implemented in IRS and used for the calculation of hydrocarbon surface areas (S^W) in Table S-I is presented, together with some general aspects of the regression analysis and statistics used for the validation of the developed QSPR models.

Molecular van der Waals surface. The space occupied by an alkane molecule (interacting with the other molecules) can be described within the frame of the "hard sphere approximation" as follows: each atom of a molecule is represented by an isotropic sphere having the centre in the equilibrium position of the atom and a radius equal to its vdW radius, r^W. Consequently, a molecular vdW envelope, **Γ**, can be defined as the external surface resulting from the intersection of all vdW spheres. This envelope embeds a 3D space of

volume V^W and surface S^W. The points (x,y,z) disposed within the envelope Γ satisfy one and only one of the following inequalities:

$$(X_i - x)^2 + (Y_i - y)^2 + (Z_i - z)^2 \leq (r_i^W)^2 \quad i = 1, 2, \ldots, m \tag{1}$$

The point (X_i, Y_i, Z_i) represent the center of an atomic sphere i ($i = 1, \ldots, m$), and m represents the number of atoms in a molecule M.

The molecular van der Waals envelope, Γ, defined by relation (1), is a surface. The calculation of the area of this surface should be made by means of the MC method.[7,11,19]

The MC algorithm[7,11] used for the calculation of the S^W value for the alkanes and cycloalkanes in Table S-I implies the random generation of a uniform grid (composed of n_t points) on each sphere i ($i = 1, 2, \ldots, m$) of the molecule M. After the generation of the total number of points on the surface (n_t), the detection of those points (n_e) that do not satisfy the inequalities in (1) follows. For every "hard sphere" i, the outer part of its surface, S_i^W, is computed:

$$S_i^W = \frac{(n_e)_i}{n_t} 4\pi \left(r_i^W \right)^2 \tag{2}$$

The final surface of hydrocarbon molecules in Table S-I is computed as the sum of the exterior surface, S_i^W, of each sphere:

$$S^W = \sum_{i=1}^{m} S_i^W \tag{3}$$

As was seen,[7-10,16,17] the vdW radius is a successful concept for the computation of molecular size and shape descriptors, even if in a quantum chemical description the electron cloud has no well-defined boundary surface. The values of the surface area of alkanes and cycloalkanes in Table S-I were computed using the Bondi vdW radii measured from various physicochemical properties.[20]

Regression analysis. Regression analysis is generally realized by a correlation equation of the following type:[6]

$$y_n = \beta_1 x_1 + \beta_2 x_{n2} + \cdots + \beta_p x_{p1} + z_n = \left(x_{n1}, \cdots, x_{np} \right) \beta + z_n \tag{4}$$

where y_n are the VP values and the matrix $\mathbf{X} = (x_{n1}, \ldots, x_{np})$ contains the predictor variables, *i.e.*, the structural parameters (MDs) of the compounds from the studied series.

Linear regression provides estimates and other inferential results for the (statistical) parameters $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^T$ in the model (4). In this model, the random variable y_n, which represents the response for the case n, $n = 1, 2, \ldots, N$, which has a deterministic part and a stochastic part. The deterministic part, $(x_{n1}, \ldots, x_{np})\beta$, depends on the parameter β and on the predictor or regression variables x_{np}, $p = 1, 2, \ldots, P$. The stochastic part, represented by the random variable z_n, is a disturbance that perturbs the response for that case. The superscript ^{T} denotes the transpose of a matrix.

The model for N cases can be written as follows:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z} \tag{5}$$

where $\mathbf{Y}$ is the vector of random variables represented the experimental data one may obtains, $\mathbf{X}$ is the $N{\times}P$ matrix of regression variables, *i.e.*, the molecular structural parameters and/or the physical and chemical properties (especially for QSAR studies) and $\mathbf{Z}$ is the vector of

variables representing the disturbances; one assumes that $\mathbf{Z}$ is normally distributed. The maximum likelihood estimate $\hat{\beta}$ is the value of β that minimizes $S(\beta)$:

$$S\left(\beta\right) = \mathbf{Y} - \mathbf{X}\beta^2 = \sum_{n=1}^{N}\left[y_n - \left(\sum_{p=1}^{P}x_{np}\beta_p\right)\right]^2 \qquad (6)$$

This β is called the least squares estimate and can be written: $\beta = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$.

Once the QSPR model is built, its goodness of fit is evaluated by means of the following statistical parameters: the correlation coefficient (r) and the coefficient of determination (r^2), adjusted for the degree of freedom r^2_{adj}, which is also called explained variance, (EV). The uncertainty in the model was noted as the standard error (s), and the reliability in the model was expressed by F- (Fisher) and t-(Student) statistics. The t-test was used to determine the 95 % confidence limits of the developed QSPR model.[7]

The statistical fit should not be confused with the ability of a model to make predictions. Therefore, the leave-one-out (LOO) and the leave-n-out (L-n-O) cross-validation methods to estimate the predictive ability of the obtained QSPR model, using the cross-validation coefficient (also called coefficient of predictions), q^2, and the squared correlation coefficient between the predicted and experimental properties, R^2 (see Fig. 2) was used.[21,22] In the LOO procedure, a value of $q^2 > 0.5$ is acceptable.[21]

Finally, the chance correlation in the developed QSPR was been checked by scrambling the VP response values (Y-scrambling) and trying to build a model using the scrambled data. This procedure was then repeated, say, 100 times and the r^2 values were checked against that for the actual QSPR: if only one of the r^2 values from the scrambled data is as high as that from actual QSPR, then there is 1 % risk that the real QSPR is a chance correlation.[23]

In this study, only S^W was used as a MD of the molecular size of hydrocarbon molecules. Consequently, the matrix $\mathbf{X}$ of the predictor variables contains a single MD, $\mathbf{X} = (S^W)$, and the above Eq. (4) is a simple linear model.

RESULTS AND DISCUSSIONS

Hitherto, the proposed vapor pressure QSPRs were limited to predicting VP at a constant temperature, using five or more linear correlation models based on topostructural, topochemical, and geometrical type descriptors.[18] The interpretation of these QSTRs was difficult in terms of an interaction model between the molecular entities because, commonly, the physical meaning of topological indices was not clear.

Here, a simple linear model is presented, which was developed based on the vdW surface area of hydrocarbons, S^W. It is known that vdW forces act at very short distances and are responsible for the intermolecular forces manifested between non-polar molecules, such as hydrocarbons. Consequently, it may be supposed that these forces act between the surfaces of molecules. S^W is a molecular feature that could explain the intermolecular interactions that occur in the physical process of the vaporization of a given chemical substance at a specified temperature.

All statistical calculations were realized by means of MobyDigs software.[23] First, the *VP* / Pa values for the hydrocarbons in Table S-I of the Supplementary material were used as experimental values in the training set. The cycloalkanes in Table S-I of the Supplementary material were used as an external prediction (test) set. The values of surface area, S^W / Å^2, of alkanes and cycloalkanes in Table S-I were calculated with in house developed IRS software.[6] Secondly, the entire set of alkanes and cycloalkanes (in Table S-I) was used as test set, and the obtained QSPR model was analyzed for its predictive ability by the cross-validation procedures described above.

The range of vapor pressure values measured at 298.15 K was about 8 logarithmic Pa units (see Table S-I). The values of vdW surface varied from 48.54 Å^2 (the smallest studied hydrocarbon molecule, methane) to 377.78 Å^2 (the largest studied alkane molecule, *n*-hexadecane). The range of S^W values was about 330 Å^2, and the ratio between the maximal and minimal values was also about 8. Consequently, one logarithmic unit corresponds to about (40 Å^2). Therefore, the discrimination capability of the S^W molecular descriptor for this QSPR study was good.

The training process involves using a dataset with known values, and learning a model from that dataset. However, it is possible that models fit the training dataset very well, but they may fail to predict new data points. Such over-fitting of the training data will most likely yield a model that cannot be generalized and, therefore, would not be useful. Thus, an algorithm and its associated parameters must be validated before they are used to predict new data. This process involves segmenting the entire data set into two subsets. One set is used for training and the other for testing the model.

By correlating the vapor pressures at 298.15 K ($A = \log$ (*VP* / Pa), with the vdW surface areas (S^W / Å^2) for the 63 alkanes in Table S-I, used as the training set, the following linear QSPR model was obtained:

$$\log\left(VP/\,\mathrm{Pa}\right) = 8.237\left(\pm 0.0779\right) - 0.0242\left(\pm 0.0004\right) S^{\,W} \tag{7}$$

$$N = 63;\ r^2 = 0.986;\ r^2_{\mathrm{adj}} = 0.985;\ s = 0.173;\ F = 4072.5$$

The QSPR Model (7) has good statistical quality for fitting the calculated VP values to the experimental ones. Thus, 98.5 % of the variance of experimental *VP* values is explained by this model, the confidence limits of the model parameters are less than 4 %, and the standard error is less than 3 % of the domain of log (*VP* / / Pa). The parameters of model (7) are significantly different from the zero value. This model was tested to evaluate its predictive ability, as shown below.

The robustness of Model (7) and its internal predictive ability were evaluated by both the q^2 cross validation coefficient based on the leave-one-out (LOO) and bootstrap procedures. In the bootstrap technique, *M n*-dimensional groups are

randomly selected from the original data set in Table S-Ia. The model obtained on the first selected objects is used to predict the remaining values, and then q^2 is calculated for each model. The bootstrapping was repeated 100 times for each validated model. The values of cross validation coefficients for LOO $\left(q^2_{\text{LOO}}\right)$ and bootstrapping $\left(q^2_{\text{BOOT}}\right)$ procedures are given in Table I.

TABLE I. Statistics corresponding to the QSPR Linear Model (7); they were obtained by LOO $\left(q^2_{\text{LOO}}\right)$, bootstrapping $\left(q^2_{\text{BOOT}}\right)$, and Y-scrambling $\left(r^2_{\text{Y-scr}}, q^2_{\text{Y-scr}}\right)$ procedures, and the standard deviations (errors) of the predicted (*SDEP*) and calculated (*SDEC*) values; the values were obtained by cross-validation with MobyDigs software[23]

Eq.	q^2_{LOO}	q^2_{BOOT}	$r^2_{\text{Y-scr}}$	$q^2_{\text{Y-scr}}$	*SDEP*	*SDEC*
(7)	0.983	0.982	–0.037	–0.108	0.181	0.170
(8)	0.988	0.987	–0.039	–0.106	0.144	0.140
(9)	0.929	0.927	–0.035	–0.085	0.341	0.331

Model (7) was also checked for reliability, robustness and chance correlation by applying the Y-scrambling method of the MobyDigs computer programs.[23] The basis of the Y-scrambling technique is the permutation concept, in which new models are recalculated for randomly reordered experimental vapor pressure values $A_i = \log(VP_i / \text{Pa})$, $i = 1,2,\ldots,63$, and the corresponding values of the cross-validation coefficients, $q^2_{\text{Y-scr}}$, were computed. These new models should have significantly lower $q^2_{\text{Y-scr}}$ values than the proposed ones because of the linear relationship between the predictor variables that describe the molecular structure (molecular surface S^{W} of the hydrocarbons in Table I) and the response A is broken. This is proof of the validity of the proposed model because it can be reasonably concluded that the originally proposed model was not obtained by chance correlation. For Model (7), the A *vs.* S^{W} Y-scrambling procedure was performed by A scrambling with 100 iterations. The values of the statistical parameters obtained by applying the *Y*-scrambling method to the data in Table S-I are summarized in Table I.

In this way, the QSPR Model (7) was statistically validated. It was then used to predict the log (*VP* / Pa) values not only for the alkanes, but also for the log (*VP* / Pa) values of the cycloalkanes in Table S-I. It is important to emphasize that the results of cross-validation and bootstrapping procedures presented above only assess the internal predictive ability of the developed Model (7), and this fact validates the use of this model to predict the values of alkanes (see A_{pred} values in Table S-I). Analyses of the deviations of the calculated (predicted) log (*VP* / Pa) values in Table S-I allowed the consideration of the point represented by methane (C1) as an outlier. The linear equation resulting when this point was eliminated from correlation is the following:

$$\log(VP/\text{Pa}) = 8.106(\pm 0.137) - 0.0236(\pm 0.0007)S^{\text{W}} \qquad (8)$$

$$n = 62; \; r^2 = 0.988; \; r_{adj}^2 = 0.988; \; s = 0.143; \; F = 5075.3$$

The results of the predictive ability of the QSPR Model (8) are also presented in Table I. By elimination of this point, the goodness of fit, the reliability and the robustness of the QSPR Model (7) were not significantly improved. The point corresponding to the methane is not a real outlier because there is no reason to eliminate it from the correlation. An outlier usually indicates a data point that is not at all typical for the rest of the data set and/or does not operates by the same mechanism. In this case, methane is the first term of the homologous series of alkanes and all hydrocarbon molecules are hydrophobic. Probably methane is an outlier because it is not polar, while the other molecules are slightly polar. Therefore, one may suppose that the intermolecular forces responsible of their vapor pressure are, in essence, of the same type.

An external predictive ability of the linear Model (7) would provide a more rigorous evaluation of the capability of the model to predict log (VP / Pa) values for chemicals with unknown VP values. In this investigation, Model (7) was externally validated on a series of structurally different compounds, namely the cyclic hydrocarbons presented in Table S-I. The predicted log (VP / Pa) values for the series of 21 cycloalkanes calculated with Eq. (7) are also given in Table S-I, together with their deviations from the corresponding measured log (VP / Pa) values.

It is known that alkane molecules are characterized by staggered conformations, which is most stable conformation because the carbon–hydrogen bonds of the vicinal non-bonded atoms are as far removed from each other as possible and, on the contrary, the eclipsed conformation is the least stable conformation because the C–H bonds are close to one another. In fact, the eclipsed conformations should be viewed as transition states between the staggered conformations, and they are very short lived. The supplementary energy of the eclipsed conformations is called torsional strain. The most stable of the staggered conformers is called the ANTI (A) conformer. However, beginning with *n*-butane, other staggered conformers exist, which are called GAUCHE (G) conformers. ANTI and GAUCHE conformers do not have the same energy because of steric strain (Fig. 1).

At room temperature, the barrier energy between staggered and eclipsed conformations is very small. Therefore, there is enough thermal energy in the surroundings to allow the conformational isomers to interconvert many times per second. Obviously, the more stable the conformation, the greater the number of molecules that will be in that conformation.

The S^W / Å^2 values of alkane molecules are calculated only for the most extended *anti* conformation. This fact could explain the small differences between the experimental and predicted (calculated) values of log (VP / Pa) for

the same alkane molecule in Table S-I. Thus, neglecting the smallest molecule, CH_4 (C1), the maximum deviation of the predicted log (VP / Pa) values was –0.446, which represents less than 6 % from the range of experimentally measured log (VP / Pa) values. It could be concluded that Model (7) predicts well the log (VP / Pa) values of alkanes, within the limits of the accuracy and precision of experimental measurements. This fact may be very well observed in Fig. 2 that contains the linear relationship between the experimental and the predicted log (VP / Pa) values for the 63 hydrocarbons in Table S-Ia, calculated with Eq. (7).

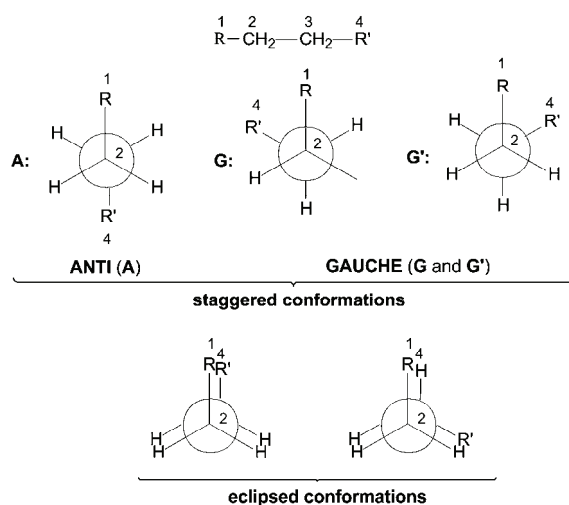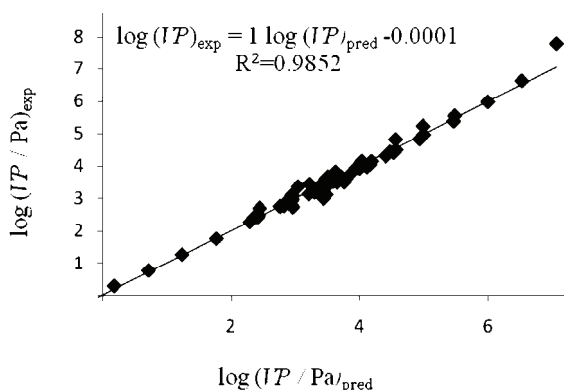Fig. 1. Newman projection of the staggered and eclipsed conformation of an alkane molecule.

Fig. 2. Experimental log (VP / Pa) values *vs*. the predicted values for the alkanes in the training set.

A systematic deviation of the predicted log(VP / Pa) values of the cycloalkanes could be observed in Table S-I. This is, probably, due to the difference in the structural three-dimensional (3D) characteristics of the acyclic and cyclic hydrocarbons, respectively, which were reflected in the variation of their surface area values between the members of each homologous series. The deviations in the predicted log (VP / Pa) values of the cycloalkanes from the regression line

Eq. (7) are, obviously, greater than those for the alkanes. The regression line between the experimental values and predicted log (VP / Pa) values for the cycloalkanes, calculated using Eq. (7), is presented in Fig. 3.

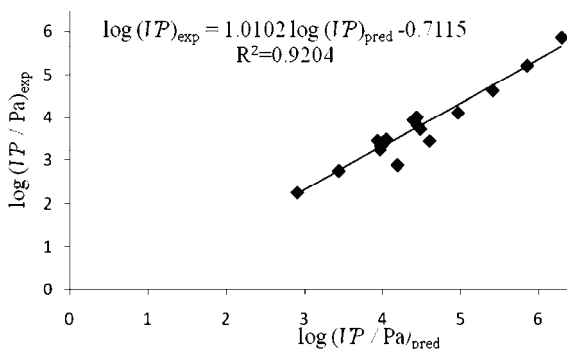Fig. 3. Experimental values *vs.* predicted log(VP / Pa) values for the cycloalkanes in the test set.

To explain these results, the energetic aspects of cycloalkane molecules have to be taken into consideration. There are three kinds of internal strain that can destabilize a cyclic compound: angle strain, torsional strain and steric strain. Deviation of bond angles from the tetrahedral bond angle (109°29′) generates angle strain. Torsional strain is caused by repulsion between the bonding electrons of adjacent non-bonded atoms or substituents. Steric strain is produced by the approach of atom groups to a distance less than to their vdW distance. Cyclic molecules bend and twist in order to achieve a final structure that minimizes these strains. These steric requirements could be a reason for the observed systematic deviations of the predicted log (VP / Pa) values. A short analysis of this subject is given in Table II.

The increments of a methylene group corresponding to alkane and cyclo-alkane molecules are approximately constant, as may be seen in Table II. This is just the difference in the surface area between the vicinal compounds from the corresponding homologous series. On the contrary, the difference in the surface area of a methylene group (CH_2) between the alkanes and cycloalkanes with the same number of carbon atoms in the molecule (see the corresponding values in Table II) varies considerably from cyclopropane 12.317 Å^2 to cyclooctane 34.812 Å^2. This fact could explain the constant deviations of all the cycloalkanes log (VP / Pa) values predicted with the QSPR Model (7), developed based on the surface area of alkanes in Table S-I. The cycloalkane molecules are more compact; this fact is reflected much more in their surface area, less than that the surface area of alkane molecules with the same number of carbon atoms.

Taking into consideration the entire series of alkanes and cycloalkanes in Table S-I, the obtained QSPR linear model was the following:

$$\log\left(VP/\,\mathrm{Pa}\right) = 7.7709\left(\pm 0.1309\right) - 0.0226\left(\pm 0.0007\right)S^{\mathrm{W}} \tag{9}$$

$$n = 84; \; r^2 = 0.933; \; r_{\text{adj}}^2 = 0.932; \; s = 0.3348; \; F = 1142.4$$

TABLE II. The variation of the surface area of a methylene group ($S_{\text{CH}_2}^{\text{W}}$): a) between the vicinal terms of the homologous series of alkanes; b) between the vicinal members of the homologous series of cycloalkanes; c) among the alkanes and cycloalkanes with the same number of carbon atoms in the molecule; $S_{\text{CH}_2}^{\text{W}} = S^{\text{W}}(Cn+1) - S^{\text{W}}(Cn)$, $n = 1,2,\ldots,16$; $S_{\text{CH}_2}^{\text{W}} = S^{\text{W}}(CyCn+1) - S^{\text{W}}(CyCn)$, $n = 1,2,\ldots,8$; $S_{\text{CH}_2}^{\text{W}} = S^{\text{W}}(Cn) - S^{\text{W}}(CyCn)$, $n = 3,4,\ldots,8$

No.	Cn	S^{W} / Å^2	$S_{\text{CH}_2}^{\text{W}}$ / Å^2	No.	$CyCn$	S^{W}/ Å^2	$S_{\text{CH}_2}^{\text{W}}$ / Å^2	$S_{\text{CH}_2}^{\text{W}}$ / Å^2
56	C1	48.541	–	1	CyC3	80.531	–	12.317
1	C2	70.896	22.355	2	CyC4	98.905	18.374	15.926
2	C3	92.848	21.952	3	CyC5	117.025	18.120	19.591
3	C4	114.831	21.983	4	CyC6	135.333	18.308	23.256
5	C5	136.616	21.785	5	CyC7	150.489	15.156	29.966
8	C6	158.589	21.973	6	CyC8	167.417	16.928	34.812
21	C8	202.229	21.774					
38	C9	224.104	21.875					
57	C10	246.018	21.914					
58	C11	267.742	21.724					
59	C12	289.892	22.150					
60	C13	311.469	21.577					
61	C14	333.608	22.139					
62	C15	355.375	21.767					
63	C16	377.377	22.002					

Model (9) has the ability to fit well the experimental vapor pressure values of both the alkanes and cycloalkanes in Table S-I (84 compounds). The predictive power of this model was estimated with the statistics summarized in Table I. The values of the coefficients of prediction, q_{LOO}^2 and q_{BOOT}^2, are significantly greater than the limit value, $q^2 = 0.500$. These facts are reflected in Fig. 4, which presents the regression line between the experimental vapor pressure values, log (VP / Pa)$_{\text{exp}}$, and the corresponding predicted values, log (VP / Pa)$_{\text{pred}}$, which were calculated using Eq. (9).
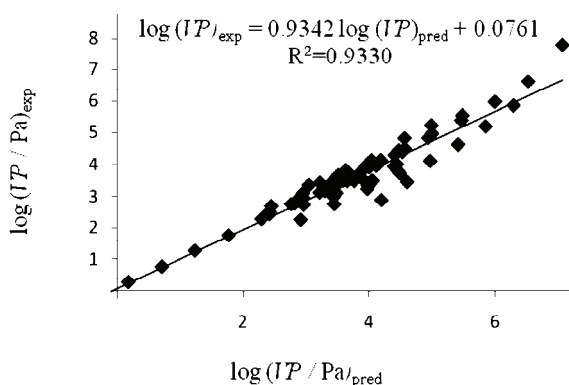

Fig. 4. Experimental log (VP / Pa) values *vs*. predicted values for the entire series of hydrocarbons in Table S-I.

CONCLUSIONS

The most important objective of this study was to investigate the usefulness of the vdW molecular surface area, S^W / Å^2, as a molecular descriptor for QSPR studies.

A QSPR approach using S^W values as the MD was successfully applied to a series of 84 hydrocarbons (alkanes and cycloalkanes) with vapor pressure values measured at 298.15 K, log (VP / Pa). QSPR equations with this single theoretical molecular descriptor of the vdW molecular space, S^W, were obtained for two subsets comprised of 63 alkanes (training set) and 21 cycloalkanes (prediction set). The S^W / Å^2 values were calculated solely from the three-dimensional chemical structures of the compounds with a Monte Carlo algorithm implemented in in-house developed software Investigation of Receptor Space (IRS), and have a clear physical meaning corresponding to intermolecular interactions. All the developed QSPR models were tested with good results for their ability of prediction and fitting the data according to the corresponding statistics.

The log (VP / Pa) values of cycloalkanes were predicted well by the QSPR model developed for the subset of alkanes as the training set. The observed small systematic deviations of these values from the linear equation obtained for the training set were explained by the difference in the structural 3D characteristics in the series of cyclic and acyclic hydrocarbons, respectively. This difference was measured by the S^W / Å^2 values between the members of each homologous series. A general QSPR model for all 84 hydrocarbons was also proposed.

This study shows that a QSPR model could be obtained by means of a single molecular descriptor when this MD has a definite physical meaning.

SUPPLEMENTARY MATERIAL

Vapor pressure data sets for alkanes and cycloalkanes are available electronically from http://www.shd.org.rs/JSCS/, or from the corresponding author on request.

ИЗВОД

КВАНТИТАТИВНЕ РЕЛАЦИЈЕ ЗА НАПОН ПАРЕ НЕКИХ УГЉОВОДОНИКА ИЗ ПОЗНАВАЊА ВАН ДЕР ВАЛСОВЕ МОЛЕКУЛСКЕ ПОВРШИНЕ

TUDOR OLARIU[1], VICENŢIU VLAIA[1], CIPRIAN CIUBOTARIU[2], DAN DRAGOŞ[1], DAN CIUBOTARIU[1] и MIRCEA MRACEC[3]

[1]*Faculty of Pharmacy, "Victor Babes" University, Timisoara, Romania,* [2]*Department of Computer Sciences, University "Politehnica", Timisoara, Romania и* [3]*Molecular Forecast Research Center, Timişoara, Romania*

Извршено је QSPR моделовање напона паре на 298.15 K, у облику log (VP / Pa), за 84 угљоводоника (63 алкана иу 21 циклоалкана). Као молекулски дескриптор употреб-љене су ван дер Валсове (*van der Waals*, vdW) површине, S^W / Å^2, израчунате методом Монте Карло. QSPR модел је развијен на подскупу од 63 алкана (C_1–C_{16}), а затим је ус-пешно примењен за предвиђање log (VP / Pa)-вредности 21 циклоалкана. Један други QSPR модел је конструисан на бази свих 84 угљоводоника. Оба QSPR модела су статис-тички тестирани. Резултати показују vdW молекулска површина може да објасни нај-већи део log (VP / Pa)-вредности, и на тај начин да објасни међумолекулске силе које де-

луjу између неутралних молекула. Jасан физички смисао молекулске површине, S^W / Å^2, може да обjасни успешност QSPR модела са jедним jединим молекулским дескриптором.

REFERENCES

1. C. Liang, D. A. Gallagher, *J. Chem. Inf. Comput. Sci.* **38** (1998) 321
2. D. Yaffe, Y. Cohen, *J. Chem. Inf. Comput. Sci.* **41** (2001) 463
3. H. E. McClelland, P. Jurs, *J. Chem. Inf. Comput. Sci.* **40** (2000) 967
4. D. Ciubotariu, M. Medeleanu, V. Vlaia, T. Olariu, C. Ciubotariu, D. Dragos, C. Seiman, *Molecules* **9** (2004) 1053
5. N. R. Draper, H. Smith, *Applied Regression Analysis*, Wiley, New York, 1981
6. IRS Computer program, http://irs.cheepee.ro/ (accessed in May, 2015)
7. D. Ciubotariu, *PhD Thesis*, Polytechnic Institute of Bucharest, Bucharest, 1987
8. I. Niculescu-Duvaz, D. Ciubotariu, Z. Simon, N. Voiculetz, in *Modeling of Cancer Genesis and Prevention*, N. Voiculetz, A. T. Balaban, I. Niculescu-Duvaz, Z. Simon, Eds., CRC Press, Boca Raton, FL, 1991, p. 157
9. A. Chiriac, D. Ciubotariu, Z. Simon, *Quantitative Structure – Activity Relationship (QSAR). The MTD Method*, Mirton, Timisoara, 1996
10. D. Ciubotariu, V. Gogonea, M. Medeleanu, in *QSPR/QSAR Studies by Molecular Descriptors*, M. V. Diudea, Ed., NOVA Science Publishers, Huntington, New York, 2000, p. 281
11. V. Gogonea, D. Ciubotariu, E. Deretey, I. Iorga, M. Medeleanu, *Rev. Roum. Chim.* **36** (1991) 465
12. D. Ciubotariu, V. Gogonea, E. Deretey, A. Chiriac, Z. Simon, *Reprint, Univ. Timisoara, Ser. Chim.* **3**, 1987
13. V. Gogonea, A. Motsenigos, D. Ciubotariu, E. Delete, A. Chiriac, Z. Simon, *Reprint, Univ. Timişoara, Fac. St. Nat. Ser. Chim.* **1**, 1987
14. M. Medeleanu, D. Dragoş, T. Olariu, V. Vlaia, C. Ciubotariu, C. Seiman, D. Ciubotariu, *Chem. Bull. Politehnica Univ.* **49** (2004) 8
15. C. Ciubotariu, M. Medeleanu, D. Ciubotariu, *Chem. Bull. Politehnica Univ. Timisoara*, **51** (2006) 13
16. D. Ciubotariu, V. Vlaia, C. Ciubotariu, T. Olariu, M. Medeleanu, in *Quantum Frontiers of Atoms and Molecules*, M. V. Putz, Ed., Nova Science Pub., Huntington, New York, 2010, p. 629
17. D. Ciubotariu, V. Vlaia, C. Ciubotariu, T. Olariu, M. Medeleanu, in *Carbon Bonding and Structures. Advances in physics and chemistry*, M. V. Putz, Ed., Springer, Dordrecht, 2011, p. 337
18. A. R. Katritzky, S. H. Slavov, D. A. Dobchev, M. Karelson, *Comput. Chem. Eng.* **31** (2007) 1123
19. S. M. Ermakov, *Metoda Monte Carlo şi probleme înrudite*, Ed. Tehnică, Bucureşti, Romania, 1976 (in Romanian)
20. A. J. Bondi, *J. Phys. Chem.* **68** (1964) 441
21. K. H. Esbensen, *Multivariate Data Analysis-In Practice*, CAMO Process AS, Oslo, 2004
22. P. Gedeck, B. Rohde, C. Bartel, *J. Chem. Inf. Model.* **46** (2006) 1924
23. MobyDigs v.1.1, Talete srl., *via* V. Pisani, 13-20124, Milano, http://www.talete.mi.it (accessed May, 2015).