Iranian Journal of Mathematical Chemistry, Vol. 5, No.2, November 2014, pp. 127–142 **IJMC**

Prediction of the Adsorption Capability onto Activated Carbon of Liquid Aliphatic Alcohols using Molecular Fragments Method

SAADI SAAIDPOUR*

(COMMUNICATED BY IVAN GUTMAN)

Chemistry Department, Faculty of Science, Sanandaj Branch, Islamic Azad University, Sanandaj, Iran

ABSTRACT. Quantitative structure-property relationship (QSPR) for estimating the adsorption of aliphatic alcohols onto activated carbon were developed using substructural molecular fragments (SMF) method. The adsorption capacity of activated carbon (gr/100grC) for 150 aliphatic alcohols onto activated carbon (AC) is studied under equilibrium conditions. Forward and backwards stepwise regression variable selection and multilinear regression (MLR) are combined to describe the effect of molecular structure on the adsorption capacity of activated carbon according to the QSPR method. To quantitatively relate adsorption capacity (Q_e) with the molecular structure MLR analysis is performed on the set of 15 substructural molecular fragments (SMF) provided by the software ISIDA. The five fragments selected by variable subset selection, all belonging to the subfragments, adequately represent the structural factors influencing the affinity of alcohols to AC in the adsorption process. Finally, a QSPR model is selected based on leave-one-out cross-validation and its prediction ability is further tested on 30 representative compounds excluded from model calibration. The prediction results from the MLR models are in good agreement with the experimental values. By applying MLR method we can predict the test set (30 compounds) with squared cross validated correlation coefficient (Q^2_{ext}) of 0.9538 and root mean square error (RMSE) of 2.0832.

Keywords: Activated carbon; Adsorption; Aliphatic alcohols; QSPR; SMF method.

1. INTRODUCTION

Adsorption phenomena have been used to perform separation and purification processes for the organic compounds. Such processes usually use a suitable porous solid adsorbent with a

*Corresponding author: Tel.: +98 871 228 4495; Fax: +98 871 228 4301
E-mail address: sasaaidpour@iausdj.ac.ir (S. Saaidpour)
Received: October 2, 2014; Accepted: November 10, 2014.

high surface area and or a high micro-pore volume. Activated carbons are excellent and versatile adsorbents. Their important applications relate to their use in the adsorptive removal of colour, odour, taste, and other undesirable organic and inorganic impurities from drinking waters; in the treatment of urban ground and industrial waste water; solvent recovery; air purification in inhabited spaces such as restaurants, food processing, and chemical industries; for the removal of colour from various types of sugar syrups, oils, and fats; in the purification of many chemical, food, and pharmaceutical products; in respirators for work under hostile environments; and in various other gas-phase applications[1]. They are increasingly being used in hydrometallurgy for the recovery of gold, silver, and other inorganic and as catalytic and catalyst supports. Their use in medicine and health applications to combat certain types of bacterial ailment and for the removal of certain toxins is well-known. These applications of activated carbon are of interest to most economic sectors and concern areas as diverse as the food, pharmaceutical, chemical, petroleum, mining, nuclear, automobile, and vacuum industries [2]. The adsorption data are generated by carrying out isotherms in batch reactors. The Freundlich classical model is used to fit the isotherms:

$$Q_e = K \times C_e^{1/n} \quad (Freundlich) \tag{1}$$

$$LogQ_e = logK + \left(\frac{1}{n}\right) logC_e \tag{2}$$

where C_e is the solution concentration at equilibrium (mgL^{-1}), Q_e the adsorption capacity at equilibrium ($mg.g^{-1}$), K a Freundlich parameter ($mg^{1-1/n}L^{1/n}g^{-1}$) and 1/n a Freundlich parameter. Different compounds have different adsorption capabilities onto activated carbon [3]. Like other properties, an adsorption capability is highly related to the molecular structures. Therefore, whether the treatment of a compound can be performed by activated carbon or not will largely rely on its structure. Since there are many compounds in the environment, it is difficult and time-consuming to experimentally determine the adsorption capability for all compounds. Thus, accurate prediction of adsorption capability of a compound on to activated carbon will be very interesting and useful. Some works have been published to predict the activated carbon adsorption capability by quantitative structure-activity relationship (QSAR) [4-17]. However, considering the large number of chemicals identified as alcohols, it would be interesting to develop relationships between the adsorbability of a great number of compounds and their structural fragments to predict the adsorption of new compounds without experimentation and to understand the adsorption mechanism. These quantitative structure property relationships (QSPR) are generally used to correlate the biological, chemical, or physical property of a compound with its physico-chemical characteristics. In our previous papers, we reported on the application of QSPR techniques in to develop a new, simplified approach to prediction of compounds properties [18-23].

For the first time, we applied the sub-structural molecular fragment (SMF) method for modeling the equilibrium adsorption of aliphatic alcohols onto activated carbon. The goal of this study is to develop a SMF method and the related software tools, to model relationships between the structure of 150 aliphatic alcohols and their adsorbability onto activated carbon. This method is based on to represent a molecule by its fragments and on to calculate their contributions to a given property. It uses two types of fragments: (i) the sequences of atoms and/or bonds (atom and/or bond paths up to specified maximal length) and (ii) "augmented" represented by a selected atom and/or bonds with its environment. In fact, it represents an extension of empirical methods used to calculate physical or chemical properties of molecules using atomic or bond increments.

2. DATA AND METHODS

The QSPR model for the estimation of the equilibrium adsorption of various aliphatic alcohol compounds is established in the following six steps: the molecular structure input and generation of the files containing the chemical structures is stored in a computer-readable format; quantum mechanics geometry is optimized with a semi-empirical (AM1) method; sub-structural molecular fragments are computed; molecular fragments are selected; and the molecular fragments–Q_e model is generated by the multilinear regression analysis (MLRA), and statistical approval techniques and prediction analysis.

2.1. Experimental Data

The 150 aliphatic alcohols and corresponding properties used in this study are available from Yaws' Handbook of thermodynamic and physical properties of chemical compounds [24]. The adsorption capacities of these compounds to activated carbon in liquid phase were obtained as capacities of equilibrium adsorption (g/100 g Carbon). The values were used as dependent variable in the following analyses and the values ranged from 7.39 to 44.60. The names of the compounds used in this study with their adsorption capabilities are listed in the Table 1.

2.2. Computer Hardware and Software

All calculations were run on a Dell Inspiron N5010 laptop computer with Intel® Core™ i7 processor with Windows 7 operating system. The molecular structures of all compounds were drawn into the HyperChem 7.5 program (Hypercube, Inc., Gainesville, 2003) and preoptimized using MM^+ molecular mechanics method (Polak–Ribiere algorithm). The final geometries of the minimum energy conformation were obtained by more precise optimization with the semi-empirical AM1 method, applying a root mean square gradient limit of 0.01 ($Kcal.mol^{-1}.\ Å^{-1}$), as a stopping criterion for optimized structures. Then, the

resulted geometries were put in to ISIDA/QSPR (version 5.76.003, 2010) to calculate substructural molecular fragments. The ISIDA/QSPR program realizes the substructural molecular fragments (SMF) method for QSPR/MLRA modeling. The SMF method is based on the splitting of molecular graph into fragments, and on the calculation of their contributions to a given property.

2.3. Computational Procedure

2.3.1. Substructural Molecular Fragments

The ISIDA/QSPR program realizes the substructural molecular fragments (SMF) method [25-31], which is based on the splitting of a molecular graph on fragments (subgraphs), and on the calculation of their contributions to a given property Y. Two classes of fragments are used: "sequences" (**I**) and "augmented" (**II**). Three sub-types **AB**, **A** and **B** are defined for each class. For the fragments **I**, they represent sequences of atoms and bonds (**AB**), of atoms only (**A**), or of bonds only (**B**). Shortest or all paths from one atom to the other are used. For each type of sequences, the minimal (n_{min}) and maximal (n_{max}) number of constituted atoms must be defined. Thus, for the partitioning **I**(**AB**, n_{min} - n_{max}), **I**(**A**, n_{min} - n_{max}) and **I**(**B**, n_{min} - n_{max}), the program generates "intermediate" sequences involving n atoms ($n_{min} \le n \le n_{max}$). In the current version of ISIDA/QSPR, $n_{min} \ge 2$ and $n_{max} \le 15$. An "augmented" represents a selected atom with its environment including both neighbouring atoms and bonds (AB), or atoms only (A, without taking hybridization of neighbours into account, or Hy, where hybridization of neighbours is accounted for), or bonds only (B).

2.3.2. Variable Selection Procedures

Generally, generated pool of descriptors is much larger than the number of compounds in the training set; therefore procedures for selecting variables should be applied to build statistically significant multilinear regressions. In ISIDA, a combination of forward and backward stepwise variable selection procedures is used.

1). *Filtering stage.* The program eliminates variables X_i which have small correlation coefficient with the property, $R_{y,i} < R^0_{y,i}$, and those highly correlated with other variables X_j ($R_{i,j} > R^0_{i,j}$), which were already selected for the model. In this work, the values $R^0_{y,i} = 0.001$, … and $R^0_{i,j} = 0.75$, … were used. Fragments always occurring in the same combination in each compound of the training set (concatenated fragments) are treated as one extended fragment.

2). *Forward stepwise pre-selection stage*. The suite of forward and backward stepwise algorithms has been used for variable pre-selection in ISIDA/QSPR studies by the variable selection suite (VSS) program. Three algorithms for forward stepwise variable selection are based on the calculations of correlation

coefficients and subtractions. This is an iterative procedure, on each step of which the program selects one X_i (two X_i and X_j or three variables X_i , X_j and X_k) maximizing the correlation coefficient $R_{y,j}$ ($R_{y,ij}$ or $R_{y,ijk}$) between X_i (X_i and X_j or X_i, X_j and X_k) and dependent variable Y. At the first step (s = 1), the modeled property for each compound is taken as its experimental one $Y_s = Y$. At each next step s, as the property value Y_s is used residual $Y_s = Y_{s-1} - Y_{calc}$, where $Y_{calc} = c_0 + c_iX_i$ ($Y_{calc} = c_0 + c_iX_i + c_jX_j$ or $Y_{calc} = c_0 + c_iX_i + c_jX_j + c_kX_k$) is calculated property by the one-variable (two- or three-variables) model with selected variable X_i (variables X_i and X_j or X_i, X_j and X_k). This loop is repeated until the number of variables k reaches a user-defined value; in this work, k was analyzed from 0.1n to 0.9n, where n is the number of the molecules in the training set.

3). *Backward stepwise selection stage.* The final selection is performed using backward stepwise variable selection procedure based on the t statistic criterion. Here, the program eliminates the variables with low $t_i = c_i/s_i$ values, where s_i is standard deviation for the coefficient c_i at the i-th variable in the model. First, the program selects the variable with the smallest $t < t_0$, then it performs a new fitting excluding that variable. This procedure is repeated until $t \geq t_0$ for selected variables or if the number of variables reaches the user's defined value. Here, t_0, the tabulated value of Student's criterion, is a function of the number of data points, the number of variables, and the significance level. Default value of the t_0 is 1.96, it can be analyzed from 1.96 to 3.9.

2.3.3. Multilinear Regression Model

The modeled physical or chemical property Y can be quantitatively calculated accounting for contributions of fragments using linear (3) fitting equations.

$$Y = A_0 + \sum_i A_i \times N_i + \Gamma, \text{ Additive Model} \quad (3)$$

where a_i is a fragment contributions, N_i is the number of fragments of i type. The a_o term is fragment independent and Γ term is external descriptors (*e.g.*, topological, electronic, etc.) by default $\Gamma = 0$. Contributions of a_i are calculated by minimizing a functional

$$U[\,a_i] = \sum_{i=1}^{n} w_i \,(Y_{\text{exp},i} - Y_{\text{calc},i})^2 \Rightarrow \min \quad (4)$$

where n is the number of the compounds in the training set, w_i the weight accounting for the accuracy of the experimental data, Y_{exp} and Y_{calc} are, respectively, experimental and calculated according to (3) property values. The equation (3) represents calculation of

property *Y* by using additive contributions of fragments. The coefficients of the equation (3) being optimized at the training stage are then used to estimate *Y* values of the compounds from the test set or to screen external databases of real or virtual compounds.
Using singular value decomposition method (SVD), ISIDA/QSPR fits the a_i terms in equations (3) and calculates corresponding statistical characteristics (correlation coefficient (*R*), standard deviation (*s*), Fischer's criterion (*F*), cross-validation correlation coefficient (*Q*), standard deviation of predictions (s_{PRESS}), Kubyni's criterion (*FIT*), R_H-factor of Hamilton and matrix of pair correlations (covariation matrix) for the terms a_i) and performs statistical tests to select the best models. The prediction ability of the models is characterized by leave-one-out correlation coefficient Q^2 and by leave-one-out standard deviation s_{press}, as well as by dispersions of predicted values of $<Y_{pred}>$ averaged over several models.

2.3.4. Validation of QSPR Models

In ISIDA/QSPR calculations, each initial data set was split into two sub-sets: training and test sets. The QSPR models were built on the training set followed by "prediction" calculations for the test set. Before a QSPR model is used to predict the properties for new compounds, it should be validated both internally and externally to ensure that the built model is robust, reliable, stable and predictive. In the current work, several statistic terms such as squared correlation coefficient R^2 for the training set fitness and Q^2_{ext} for the external predictive ability, leave-one-out (LOO) cross-validated Q^2_{LOO} and root mean square error (RMSE) were used to assess the internal and external predictive ability of the proposed model. The corresponding statistical parameters were defined as:

$$\boldsymbol{R^2 = 1 - \frac{\sum_i^n (y_{ip} - y_{ie})^2}{\sum_i^n (y_{ie} - y_{mean}^{training})^2}} \quad (5)$$

$$\boldsymbol{Q^2_{Loo} = 1 - \frac{\sum_i^n (y_{ip} - y_{icv})^2}{\sum_i^n (y_{ie} - y_{mean}^{training})^2}} \quad (6)$$

$$\boldsymbol{Q^2_{ext} = 1 - \frac{\sum_i^n (y_{ip} - y_{ie})^2}{\sum_i^n (y_{ie} - y_{mean}^{test})^2}} \quad (7)$$

$$\boldsymbol{RMSE = \sqrt{\frac{\sum_i^n (y_{ip} - y_{ie})^2}{n}}} \quad (8)$$

Where i represents i_{th} molecule, y_{ie} is the desired output (experimental property),y_{ip} the actual output, y_{icv} is the output of leave-one-out cross-validation, $y_{training}$ ^{mean} and y_{test} ^{mean} are the mean values of y_{ip} for the training and test sets, respectively. N is the number of compounds in the training or test set. In addition, the built model was also validated

externally using the test set compounds due to the fact that the best way to evaluate the predictive ability of a QSPR model is its validation using compounds not included in the training set with known properties.

3. RESULTS AND DISCUSSION

The ISIDA program has been developed to establish structure-property relationships based on the SMF partitioning. The program inputs data in the SDF format [32] containing structural and properties information. The graphical interface of ISIDA allows to attribute data to the learning or to the validation sets and to set up the parameters of calculations (type of fragments, minimal and maximal number of atoms/bonds in the sequences, type of equation). A QSPR is a mathematical relationship between a property of a chemical, in this case adsorbability onto activated carbon, and molecular fragments of the chemical. The fragments are obtained from the structure of the chemical. First a training set of adsorption data is used to statistically establish the relationship between adsorbability and the molecular fragments. The QSPR can then be used to predict the adsorbability of untested chemicals for which the fragments are known. Thus the fragments selected to describe this process in a QSPR should be able to describe the relative affinities of chemicals for activated carbon. To establish relationships between the structure of aliphatic alcohols and their adsorption properties, we used the recently developed substructural molecular fragments (SMF) method which is based on the representation of the molecular graph by fragments and on the calculation of their contributions to a given property. The sequences fragments represent sequences of atoms and bonds (AB), of atoms only (A), or of bonds only (B). The length of sequences varies from 2 to 15 atoms. For any sequence containing from n_{min} to n_{max} atoms, all fragments of n_{max}, n_{max-1}, n_{max-2},..., n_{min} length are considered. In this work, the **I**(**AB**, **2-6**) decomposition scheme corresponds to five sequences containing 6, 4, 4, 3, and 2 atoms and linking bonds are selected. To select the most relevant fragments to the adsorption capabilities, 15 fragments calculated by ISIDA for each compound were used as the inputs for stepwise regression. The 120 alcohols were selected for the training set and 30 alcohols for the test set. The optimum subset size was reached when adding another fragment did not improve the performance of the model significantly. Through this procedure, the 5-parameter model was selected as the best model. It can be described at **Table 2**. The quality of a QSPR model is generally expressed by its fitting ability and prediction ability, and the latter one is more important. The statistical parameters for the test set were Q^2_{ext} of 0.9538 and standard deviation error of prediction (SDEP) of 1.8183.
When a compound is split into constitutive fragments, the fragments contributions to the adsorption capacity (Q_e) (gr/100grC) or to any other physical or chemical property are calculated using linear fitting equation:

$$Q_e = A_0 + \sum(A_i \times N_i) \quad (9)$$

Here, A_i is contribution of fragment, and N_i is the number of fragments of i type. The A_o term is fragment independent. The fragments contributions as fitted coefficients in the equation (9) at the learning stage are used to predict Q_e for the compounds from the validation set. Set of fragments, coefficients of the equation, standard deviations for coefficients and their t-test for equation (9) are shown in Table 2. This shows that the adsorbability increases as C-C, C-C-C-C, C-C-C-C-C-C, increase, and C-C-O, C-C-C-O decrease, respectively. The correlation matrix between selected fragments and Q_e are significant (see Table 3). On the other hand, the signs of the coefficients were used in order to determine the influence of each variable, positive or negative, on the adsorbability Qe. The experimental, predicted and residuals data for training set (120 compounds) and test set (30 compounds) are shown in Tables 4 and 5. The statistical results of training and external validation of model are shown in Table 6.

Carbon surface has a unique character. It has a porous structure which determines its adsorption capacity, it has a chemical structure which influences its interaction with polar and non-polar adsorbates, it has active sites in the form of edges, dislocations and discontinuities which determine its chemical reactions with other atoms. The determination of a correct model for adsorption on activated carbon adsorbents with complex chemical structure is therefore, a complicated problem. A proper model must take into consideration both the chemical and the porous structure of the carbon, which includes the nature and concentration of the surface chemical groups, the polarity of the surface, the surface area, and the pore size distribution, as well as the physical and chemical characteristics of the adsorbate, such as its chemical structure, polarity, and molecular dimensions. Adsorption arises as a result of the unsaturated and unbalanced molecular forces that are present on every solid surface. Thus, when a solid surface is brought into contact with a liquid or gas, there is an interaction between the fields of forces of the surface and that of the liquid or the gas. The solid surface tends to satisfy these residual forces by attracting and retaining on its surface the molecules, atoms, or ions of the gas or liquid. The adsorption involves two types of forces: physical forces that may be dipole moments, polarization forces, dispersive forces, or short-range repulsive interactions and chemical forces that are valency forces arising out of the redistribution of electrons between the solid surface and the adsorbed atoms. Depending upon the nature of the forces involved, the adsorption is of two types: physical adsorption and chemical adsorption. In the case of physical adsorption, the adsorbate is bound to the surface by relatively weak van der Walls forces, which are similar to the molecular forces of cohesion and are involved in the condensation of vapors into liquids. Chemisorption, on the other hand, involves exchange or sharing of electrons between the adsorbate molecules and the surface of the adsorbent resulting in a chemical reaction. The bond formed between the adsorbate and the adsorbent is essentially a chemical bond and is thus much stronger than in the physisorption. Physical adsorption is nonspecific and occurs between any adsorbate-adsorbent systems, but chemisorption is

specific. The type of adsorption that takes place in a given adsorbate-adsorbent system depends on the nature of the adsorbate, the nature of the adsorbent, the reactivity of the surface, the surface area of the adsorbate, and the temperature and pressure of adsorption. This study shows that linear molecules seem to be better adsorbed onto activated carbon than bulky molecules, because of an adsorption which is located between the micro-graphitic planes of AC. The adsorption process occurs via a donor-acceptor interaction between the surface of the activated carbon and the solute. Physical adsorption onto activated carbon mainly involves Van der Waals forces (dispersion–repulsion). In general, molecular fragments correlate well with physical properties that are dependent on molecular volume as each index incorporates a summation of terms representing fragments of the molecules. Thus in the QSPR here, one sees a general increase in absorbability as molecular size increase, reflected in increases in fragments. Thus in a homologous series such as the alcohols, adsorption increases with increasing chain length data and Van der Waals forces.

4. CONCLUSION

In this work, MLR modeling method was used to study the quantitative structure-property relationship of adsorption capability into activated carbon for an aliphatic alcohol data set. We can conclude that: firstly, the prediction results indicate that the multi-linear regression modeling method can improve the prediction accuracy significantly for this large data set; secondly, the models developed in this work provide an accurate model that can be used to predict the adsorption capability to activated carbon from the molecular structure only. In this case the physical adsorption occurs between molecular structures and activated carbon. Physical adsorption onto activated carbon mainly involves Van der Waals forces. In this paper, new QSPR models have been developed for predicting the Qe of a diverse set of alcohols from the molecular structure alone. The obtained results show that MLR method could model the relationship between Qe and their sub-structural fragmental. By performing model validation, it can be concluded that the presented model is a valid model and can be effectively used to predict the Qe of alcohols with an accuracy approximating the accuracy of experimental Qe determination. It can be reasonably concluded that the proposed model would be expected to predict Qe for new aliphatic alcohols for which experimental values are unknown. The main advantages of fragment descriptors lie in the simplicity of their computation, the easiness of their interpretation as well as in efficiency of their applications in similarity searches and SAR/QSAR/QSPR modeling.

REFERENCES

1. R. Ch. Bansal and M. Goyal, Activated Carbon Adsorption, Taylor & Francis Group, CRC Press, London, 2005.
2. R. H. Ottewill, C. H. Rochester and A. L. Smith, in Adsorption from Solution, Acad Press, New York, 1983.
3. J. S. Mattson and H. B. Mark, Activated Carbon: Surface Chemistry and Adsorption from Solution, Marcel Dekker, New York, 1971.
4. Z. Yu, S. Peldszus and P. M. Huck, Adsorption of selected pharmaceuticals and an endocrine disrupting compound by granular activated carbon. 2. Model prediction, *Environ. Sci. Technol.* **43** (2009) 1474–1479.
5. C. Brasquet and P. Le Cloirec**,** QSAR for organics adsorption onto activated carbon in water: what about the use of neural networks, *Wat. Res.* **33** (17) (1999) 3603–3608.
6. C. Brasquet, B. Bourges and P. Le Cloirec, Quantitative structure-property relationship (qspr) for the adsorption of organic compounds onto activated carbon cloth: comparison between multiple linear regression and neural network, *Environ. Sci. Technol.* **33** (1999) 4226–4231.
7. P. Pre, F. Delage, C. Brasquet and P. Le Cloirec, Quantitative structure–activity relationships for the prediction of VOCs adsorption and desorption energies onto activated carbon, *Fuel Process. Technol.* **77–78** (2002) 345–351.
8. D. J. W. Blum, I. H. Suffet and J. P. Duguet, Quantitative structure-activity relationship using molecular connectivity for the activated carbon adsorption of organic chemicals in water, *Wat. Res.* **28** (3) (1994) 687–699.
9. B. Lei, Y. Ma, J. Li, H. Liu, X. Yao and P. Gramatica, Prediction of the adsorption capability onto activated carbon of a large data set of chemicals by local lazy regression method, *Atmos. Environ.* **44** (2010) 2954–2960.
10. L. A. Jonas, Y. B. Tewari and E. B. Sanson, Prediction of adsorption rate constants of activated carbon for various vapors, *Carbon* **11** (1979) 345–349.
11. L. Jantschi, Water Activated Carbon Organics Adsorption Structure – Property Relationships, *Leonardo J. Sci.* **3** (5) (2004) 63–73.
12. H. M. Pignon, C. Faur and P. L. Cloirec, Adsorption of dyes onto activated carbon cloth: Using QSPRs as tools to approach adsorption mechanisms, *Chemosphere* **66** (2007) 887–893.
13. A. A. Darchivio, A. Incani, P. Mazzeo and F. Ruggieri, Adsorption of s-triazines onto polybenzimidazole: A quantitative structure–property relationship investigation, *Anal. Chim. Acta,* **650** (2009) 175–182.

14. B.Yan and X. Yang, Adsorption prediction for three binary supercritical gas mixtures on activated carbon based on a NDFT/PSD approach, *Chem. Eng. Sci.* **60** (2005) 3267 – 3277.
15. T. C. Keener and D. Zhou, Prediction of activated carbon adsorption performance under high relative humidity conditions, *Environ. Prog.* **9** (1990) 40–46.
16. A. Jedrzejak and M. Jablofiski, Prediction of Adsorption of Organic Component and Water Vapour Mixture onto Activated Carbon, *Monatshefte fur Chemie*, **119** (1988) 187–194.
17. M. Huggahalli and J. R. Fair, Prediction of Equilibrium Adsorption of Water onto Activated Carbon, *Ind. Eng. Chem. Res.* **35** (1996) 2071–2074.
18. J. Ghasemi and S. Saaidpour, QSPR prediction of aqueous solubility of drug–like organic compounds, *Chem. Pharm. Bull.* **55** (2007) 669–674.
19. J. Ghasemi, S. Saaidpour and S. D. Brown, QSPR study for estimation of acidity constants of some aromatic acids derivatives using multiple linear regression (MLR) analysis, *J. Mol. Struct. (Theochem.)* **805** (2007) 27–32.
20. J. Ghasemi and S. Saaidpour, Quantitative structure–property relationship study of n-octanol- water partition coefficients of some of diverse drugs using multiple linear regression, *Anal. Chim. Acta* **604** (2007) 99–106.
21. J. Ghasemi and S. Saaidpour, QSRR prediction of the chromatographic retention behavior of painkiller drugs, *J. Chromatogr. Sci*. **47** (2009) 156–163.
22. J. Ghasemi and S. Saaidpour, Artificial neural network-based quantitative structural property relationship for predicting boiling points of refrigerants, *QSAR Comb. Sci.* **28** (2009) 1245–1254.
23. S. Saaidpour, Prediction of drug lipophilicity using back propagation artificial neural network modeling, *Orient. J. Chem.*, **30**(2) (2014) 793–802.
24. C. L. Yaws, Handbook of Thermodynamic and Physical Properties of Chemical Compounds: Physical, Thermodynamic and Transport Properties for 5,000 Organic Chemical Compounds. Lamar University, Beaumont, Texas, Norwich, New York, 2004.
25. V. P. Solovev, A. Varnek and G. Wipff, Modeling of Ion Complexation and Extraction Using Substructural Molecular Fragments, *J. Chem. Inf. Comput. Sci.* **40** (2000) 847–858.
26. A. Varnek, G. Wipff and V. P. Solovev, Towards an Information System on Solvent Extraction, *Solvent Extr. Ion Exc.* **19** (2001) 791–837.
27. A. Varnek, G. Wipff, V. P. Solovev and A. F. Solotnov, Assessment of the macrocyclic effect for the complexation of crown-ethers with alkali cations using the substructural molecular fragments method, *J. Chem. Inf. Comput. Sci.* **42** (4) (2002) 812–829.
28. V. P. Solovev and A. Varnek, Anti–HIV activity of hept, tibo and cyclic urea derivatives: structure-property studies, focused combinatorial library generation and hits

selection using substructural molecular fragments method, *J. Chem. Inf. Comp. Sci.* **43** (5) (2003) 1703–1719.
29.V. P. Solovev and A. Varnek, Structure-property modeling of metal binders using molecular fragments, *Russ. Chem. Bull.* **53** (2004) 1434–1445.
30. A. Varnek and V. P. Solovev, "In Silico" design of potential anti-hiv actives using fragment descriptors, *Comb. Chem. High T. Scr.* **8** (5) (2005) 403–416.
31. A. Varnek, D. Fourches, F. Hoonakker and V. P. Solovev, Substructural fragments: an universal language to encode reactions, molecular and supramolecular structures, *J. Comput. Aided Mol. Des.* **19** (2005) 693–703.
32. A. Dalby, J. G. Nourse, W. D. Hounshell, A. K. I. Gushurst, D. L. Grier, B. A. Leland and J. Laufer, Description of several chemical structure file formats used by computer programs developed at molecular design limited, *J. Chem. Inf. Comput. Sci*. **32** (1992) 244–255.

Table 1. Experimental data of equilibrium adsorption at maximum concentration of alcohols on activated carbon.

NO	Name	Qe (gr/100grC)	NO	Name	Qe (gr/100gr C)
1	butan-1-ol	10.16	51	2,2-dimethylpentan-3-ol	25.99
2	butan-2-ol	7.39	52	2,3-dimethylpentan-3-ol	27.75
3	2,2-diethylbutan-1-ol	42.41	53	2-ethylbutan-1-ol	26.35
4	3,3-dimethylbutan-1-ol	26.05	54	4-ethylhexan-1-ol	39.52
5	2,3-dimethylbutan-1-ol	28.92	55	3-ethylhexan-1-ol	39.75
6	2,2-dimethylbutan-1-ol	24.30	56	2-ethylhexan-1-ol	40.19
7	4,4-dimethylhexan-1-ol	40.41	57	2-ethylpentan-1-ol	37.91
8	4,5-dimethylhexan-1-ol	40.41	58	3-ethylpentan-1-ol	37.85
9	5,5-dimethylhexan-1-ol	40.41	59	4-ethylhexan-2-ol	39.45
10	2,3-dimethylhexan-1-ol	36.95	60	3-ethylhexan-2-ol	39.52
11	3,4-dimethylhexan-1-ol	39.98	61	3-ethylpentan-2-ol	32.14
12	3,3-dimethylhexan-1-ol	40.34	62	3-ethylpentan-2-ol	35.99
13	2,4-dimethylhexan-1-ol	37.40	63	4-ethylhexan-3-ol	38.01
14	2,2-dimethylhexan-1-ol	36.34	64	3-ethylpentan-3-ol	28.77
15	3,5-dimethylhexan-1-ol	40.71	65	heptan-2-ol	34.97
16	2,4-dimethylpentan-1-ol	34.40	66	heptan-1-ol	40.21
17	2,3-dimethylpentan-1-ol	37.27	67	heptan-3-ol	33.38
18	2,2-dimethylpentan-1-ol	32.34	68	hexan-1-ol	29.47
19	3,3-dimethylpentan-1-ol	37.44	69	hexan-3-ol	23.50
20	3,4-dimethylpentan-1-ol	37.41	70	hexan-2-ol	23.78
21	4,4-dimethylpentan-1-ol	34.51	71	2-methylpropan-1-ol	9.78
22	2,2-dimethylpropan-1-ol	13.31	72	3-methylbutan-1-ol	17.60
23	2,3-dimethylbutan-2-ol	18.38	73	2-methylbutan-1-ol	17.25
24	3,3-dimethylbutan-2-ol	18.67	74	3-methylheptan-1-ol	40.19
25	2-ethyl-2,3-dimethylbutan-1-ol	42.35	75	2-methylheptan-1-ol	41.41
26	2-ethyl-3,3-dimethylbutan-1-ol	42.39	76	4-methylheptan-1-ol	40.05
27	2,3-dimethylhexan-2-ol	36.54	77	5-methylheptan-1-ol	41.97
28	4,5-dimethylhexan-2-ol	37.48	78	6-methylheptan-1-ol	42.57
29	4,4-dimethylhexan-2-ol	37.48	79	2-methylhexan-1-ol	36.26
30	3,5-dimethylhexan-2-ol	36.75	80	5-methylhexan-1-ol	40.00
31	3,4-dimethylhexan-2-ol	41.33	81	4-methylhexan-1-ol	40.85
32	3,3-dimethylhexan-2-ol	36.80	82	3-methylhexan-1-ol	40.52
33	2,5-dimethylhexan-2-ol	32.76	83	4-methylpentan-1-ol	29.56
34	2,4-dimethylhexan-2-ol	32.06	84	3-methylpentan-1-ol	30.18
35	5,5-dimethylhexan-2-ol	38.95	85	2-methylpentan-1-ol	27.54
36	2,4-dimethylpentan-2-ol	24.60	86	3-methylbutan-2-ol	14.43
37	4,4-dimethylpentan-2-ol	26.26	87	2-ethyl-3-methylbutan-1-ol	36.28
38	2,3-dimethylpentan-2-ol	27.16	88	2-ethyl-2-methylbutan-1-ol	33.82
39	3,4-dimethylpentan-2-ol	32.66	89	2-ethyl-4-methylpentan-1-ol	43.20
40	3,3-dimethylpentan-2-ol	30.08	90	2-ethyl-3-methylpentan-1-ol	43.49
41	5,5-dimethylhexan-3-ol	33.53	91	2-ethyl-2-methylpentan-1-ol	43.62
42	2,2-dimethylhexan-3-ol	34.75	92	5-methylheptan-2-ol	40.33
43	4,4-dimethylhexan-3-ol	36.05	93	2-methylheptan-2-ol	34.16
44	2,3-dimethylhexan-3-ol	35.74	94	6-methylheptan-2-ol	40.03
45	2,5-dimethylhexan-3-ol	35.30	95	4-methylheptan-2-ol	39.72
46	3,4-dimethylhexan-3-ol	33.59	96	3-methylheptan-2-ol	38.18

Table 1. Continued

NO	Name	Qe (gr/100gr C)	NO	Name	Qe (gr/100gr C)
47	3,5-dimethylhexan-3-ol	33.03	97	2-methylhexan-2-ol	27.92
48	2,4-dimethylhexan-3-ol	36.23	98	4-methylhexan-2-ol	31.21
49	4,5-dimethylhexan-3-ol	34.80	99	3-methylhexan-2-ol	31.68
50	2,4-dimethylpentan-3-ol	27.04	100	5-methylhexan-2-ol	30.88
101	3-methyl-2-(propan-2-yl)butan-1-ol	42.41	126	octan-3-ol	41.92
102	3-methylpentan-2-ol	23.37	127	octan-1-ol	44.60
103	4-methylpentan-2-ol	23.18	128	pentan-3-ol	14.75
104	2-methylpentan-2-ol	18.86	129	pentan-1-ol	18.78
105	3-ethyl-2-methylpentan-1-ol	42.63	130	pentan-2-ol	15.21
106	3-ethyl-3-methylpentan-1-ol	42.63	131	2-propylpentan-1-ol	39.14
107	3-ethyl-4-methylpentan-1-ol	42.63	132	2,2,3,3-tetramethylbutan-1-ol	42.39
108	3-ethyl-4-methylpentan-2-ol	38.37	133	2,3,3-trimethylbutan-1-ol	35.19
109	3-ethyl-2-methylpentan-2-ol	35.70	134	2,2,3-trimethylbutan-1-ol	34.56
110	3-ethyl-3-methylpentan-2-ol	37.52	135	2,3,3-trimethylpentan-1-ol	43.60
111	3-ethyl-2-methylpentan-3-ol	36.41	136	3,4,4-trimethylpentan-1-ol	40.88
112	2-methylheptan-3-ol	39.00	137	3,3,4-trimethylpentan-1-ol	40.88
113	3-methylheptan-3-ol	36.53	138	2,4,4-trimethylpentan-1-ol	40.84
114	4-methylheptan-3-ol	33.23	139	2,2,4-trimethylpentan-1-ol	39.90
115	5-methylheptan-3-ol	33.13	140	2,2,3-trimethylpentan-1-ol	42.55
116	6-methylheptan-3-ol	35.37	141	2,3,4-trimethylpentan-1-ol	41.85
117	2-methylhexan-3-ol	29.00	142	2,3,3-trimethylbutan-2-ol	24.87
118	5-methylhexan-3-ol	30.18	143	2,3,3-trimethylpentan-2-ol	36.00
119	3-methylhexan-3-ol	28.27	144	3,4,4-trimethylpentan-2-ol	36.47
120	4-methylhexan-3-ol	30.46	145	2,3,4-trimethylpentan-2-ol	34.88
121	3-methylpentan-3-ol	19.55	146	3,3,4-trimethylpentan-2-ol	39.33
122	2-methylpentan-3-ol	20.70	147	2,4,4-trimethylpentan-2-ol	31.05
123	2-methylheptan-4-ol	37.87	148	2,3,4-trimethylpentan-3-ol	35.78
124	3-methylheptan-4-ol	38.20	149	2,2,3-trimethylpentan-3-ol	33.82
125	4-methylheptan-4-ol	36.32	150	2,2,4-trimethylpentan-3-ol	32.90

Table 2. Set of fragments, Coefficients (Ai) of the equation, standard deviations for coefficients and their t-Test for $Q_e = A_0 + \sum(A_i \times N_i)$.

NO	Variable[i]	Contribution (Ai)	Standard deviation (ΔA)	t-Test
0	A_0	-0.90751	0.05	17.90
1	C-C	5.9394	0.39	15.24
2	C-C-O	-3.5797	0.25	14.16
3	C-C-C-O	-0.8072	0.21	3.90
4	C-C-C-C	0.87059	0.19	4.61
5	C-C-C-C-C-C	1.16071	0.41	2.84

Table 3. Correlation matrix between fragments and Qe.

	C-C	C-C-O	C-C-C-O	C-C-C-C	C-C-C-C-C-C
C-C	1				
C-C-O	0.1152	1			
C-C-C-O	0.1739	0.2091	1		
C-C-C-C	0.7049	0.0689	0.2173	1	
C-C-C-C-C-C	0.2585	0.0384	-0.1462	-0.0503	1
Qe	0.8736	-0.2773	-0.0139	0.7583	0.2635

Table 4. Data of experimental, predicted and residual for training set of alcohol compounds.

NO	Qe_exp	Predicted	Residual	NO.	Qe_exp	Predicted	Residual	NO.	Qe_exp	Predicted	Residual
5	7.39	6.49	0.90	52	27.75	27.62	0.13	102	23.37	24.20	-0.83
3	42.41	42.79	-0.38	53	26.35	27.92	-1.57	103	23.18	23.68	-0.50
4	26.05	26.19	-0.14	54	39.52	42.83	-3.31	104	18.86	19.96	-1.10
5	28.92	27.32	1.60	55	39.75	42.83	-3.08	105	42.63	42.86	-0.23
7	40.41	41.68	-1.27	57	37.91	35.07	2.84	107	42.63	43.46	-0.83
8	40.41	42.23	-1.82	58	37.85	36.85	1.00	108	38.37	39.15	-0.78
9	40.41	39.33	1.08	59	39.45	39.11	0.34	109	35.7	35.43	0.27
10	36.95	41.63	-4.68	60	39.52	38.51	1.01	110	37.52	39.07	-1.55
12	40.34	41.68	-1.34	62	35.99	34.20	1.79	112	39.00	37.86	1.14
13	37.40	40.51	-3.11	63	38.01	37.92	0.09	113	36.53	34.74	1.79
14	36.34	38.14	-1.80	64	28.77	28.22	0.55	114	33.23	37.92	-4.69
15	40.71	41.11	-0.40	65	34.97	32.50	2.47	115	33.13	38.46	-5.33
17	37.27	35.65	1.62	67	33.38	31.90	1.48	117	29.00	29.58	-0.58
18	32.34	32.21	0.13	68	29.47	27.94	1.53	118	30.18	30.18	0.00
19	37.44	35.64	1.80	69	23.50	23.63	-0.13	119	28.27	27.64	0.63
20	37.41	36.25	1.16	70	23.78	24.22	-0.44	120	30.46	30.76	-0.30
22	13.31	14.43	-1.12	72	17.60	19.06	-1.46	122	20.70	22.49	-1.79
23	18.38	19.88	-1.50	73	17.25	18.46	-1.21	123	37.87	38.46	-0.59
24	18.67	21.28	-2.61	74	40.19	42.77	-2.58	124	38.20	37.86	0.34
25	42.35	42.18	0.17	75	41.41	42.17	-0.76	125	36.32	34.80	1.52
27	36.54	34.20	2.34	77	41.97	42.77	-0.80	127	44.60	44.49	0.11
28	37.48	38.51	-1.03	78	42.57	42.77	-0.20	128	14.75	15.35	-0.60
29	37.48	37.96	-0.48	79	36.26	33.90	2.36	129	18.78	19.66	-0.88
30	36.75	36.79	-0.04	80	40.00	34.49	5.51	130	15.21	15.95	-0.74
32	36.80	36.77	0.03	82	40.52	35.67	4.85	132	42.39	40.98	1.41
33	32.76	32.44	0.32	83	29.56	27.39	2.17	133	35.19	34.45	0.74
34	32.06	33.67	-1.61	84	30.18	28.51	1.67	134	34.56	33.85	0.71
35	38.95	35.61	3.34	85	27.54	26.80	0.74	135	43.60	42.78	0.82
37	26.26	29.69	-3.43	87	36.28	35.65	0.63	137	40.88	43.38	-2.50
38	27.16	28.21	-1.05	88	33.82	34.45	-0.63	138	40.84	38.31	2.53
39	32.66	31.93	0.73	89	43.20	40.51	2.69	139	39.90	37.71	2.19
40	30.08	30.73	-0.65	90	43.49	42.81	0.68	140	42.55	41.07	1.48
42	34.75	33.82	0.93	92	40.33	39.05	1.28	142	24.87	26.41	-1.54
43	36.05	36.18	-0.13	93	34.16	35.33	-1.17	143	36.00	34.75	1.25
44	35.74	33.60	2.14	94	40.03	39.05	0.98	144	36.47	37.94	-1.47
45	35.30	34.96	0.34	95	39.72	39.11	0.61	145	34.88	34.83	0.05
47	33.03	33.08	-0.05	97	27.92	27.06	0.86	147	31.05	31.47	-0.42
48	36.23	35.60	0.63	98	31.21	31.95	-0.74	148	35.78	33.64	2.14
49	34.80	37.32	-2.52	99	31.68	31.36	0.32	149	33.82	33.03	0.79
50	27.04	28.50	-1.46	100	30.88	30.78	0.10	150	32.90	32.80	0.10

Table 5. Predicted and residual equilibrium adsorption at maximum concentration for test set of alcohol compounds.

NO.	Qe(exp)	Predicted	Residual	NO.	Qe(exp)	Predicted	Residual
1	10.16	10.21	-0.05	76	40.05	42.83	-2.78
6	24.30	25.00	-0.70	81	40.85	35.67	5.18
11	39.98	43.4	-3.42	86	14.43	14.75	-0.32
16	34.40	33.41	0.99	91	43.62	40.49	3.13
21	34.51	33.41	1.10	96	38.18	38.46	-0.28
26	42.39	41.66	0.73	101	42.41	42.26	0.15
31	41.33	39.09	2.24	106	42.63	43.98	-1.35
36	24.60	26.57	-1.97	111	36.41	34.24	2.17
41	33.53	35.02	-1.49	116	35.37	38.46	-3.09
46	33.59	34.78	-1.19	121	19.55	20.48	-0.93
51	25.99	27.90	-1.91	126	41.92	40.18	1.74
56	40.19	42.17	-1.98	131	39.14	42.23	-3.09
61	32.14	32.53	-0.39	136	40.88	42.26	-1.38
66	40.21	36.21	4.00	141	41.85	42.26	-0.41
71	9.78	7.89	1.89	146	39.33	38.47	0.86

Table 6. Statistical parameters of QSPR-MLRA model.	
Multiple correlation coefficient (train set)	R=0.9813, R^2=0.9628
Fischer's criterion (train set)	F=334.2918
Standard deviation (train set)	SD=1.5835
Root mean-squared error (train set)	RMSE=1.7265
Mean absolute error (train set)	MAE=1.2992
Squared correlation coefficient of leave-one-out cross-validation	Q^2-LOO=0.9408
Standard deviation error of prediction	SDEP=1.8183
Squared correlation coefficient of test set	Q^2-Ext=0.9538
Root mean-squared error of test set	RMSE=2.0820