ISSN 0012-5008, Doklady Chemistry, 2011, Vol. 440, Part 2, pp. 299–301. © Pleiades Publishing, Ltd., 2011.
Original Russian Text © A.A. Kravtsov, P.V. Karpov, I.I. Baskin, V.A. Palyulin, N.S. Zefirov, 2011, published in Doklady Akademii Nauk, 2011, Vol. 440, No. 6, pp. 770–772.

Prediction of Rate Constants of S_N2 Reactions by the Multicomponent QSPR Method

A. A. Kravtsov, P. V. Karpov, I. I. Baskin, V. A. Palyulin, and Academician N. S. Zefirov

Received April 19, 2011

DOI: 10.1134/S0012500811100107

One of the most important parameters of any reaction is its rate constant. However, when dealing with a chemical reaction, a researcher faces the necessity to consider the contributions of many components that form the reaction system: the solvent, substrate, and reagent. The reaction rate also dramatically depends on pressure and temperature.

The problem of prediction of chemical reaction rate constants has long attracted the attention of researchers, and considerable progress has been achieved in this respect. For many reaction series, specific models have been constructed that consider the effect of different reagent parameters. Such models are characterized by high correlation coefficients but cannot be thought of as universal [1, 2].

Therefore, it is of interest to construct a unified model that could predict, with an acceptable accuracy, the rate constants of nucleophilic substitution at a saturated carbon atom on the basis of the structures of all the reagents involved, no matter what class of compounds they belong to.

Consideration of such a complicated problem from the standpoint of QSPR (quantitative structure–property relationship) methodology necessitates using the multicomponent QSPR (MQSPR) approach, which was previously successfully applied to predicting the solvation free energy, which depends on the structure of two substances, the solute and the solvent [3].

The present work was aimed at constructing structure–property models in the framework of MQSPR methodology for predicting the rates of S_N2 nucleophilic substitution at a saturated carbon atom.

To solve these problems, a database was formed containing voluminous information on reactions [4]: the structures of a solvent, nucleophile, and substrate (separately of its electrophilic moiety and nucleofuge); reaction temperature; and reaction rate constant. The database was supplemented with information on the solvatochromic parameters of solvents [5], which made it possible to considerably improve the model quality [1]. Altogether, the database contained information on 3451 S_N2 reactions. Hereinafter, the reactions that occur in individual solvents rather than in a mixture of several solvents were used for constructing the models. There was a total of 1924 such reactions in the database.

For the description of the structures of the reagents to predict reaction rate constants, fragmental descriptors of two types were used. Descriptors of the first type were obtained with inclusion of the number of substructures constituting a molecule; such substructures consist of one to fifteen atoms and can be atom chains, rings, bi- and tricyclic moieties, and some branched fragments. In the description of types of atoms in the fragments, several levels of generalization were used, which made it possible to discern similar backbone segments in rather different structures and thoroughly consider the hybridization of atoms and the number of the bonds with hydrogen atoms. Fragmental descriptors of the other type, LFA [3], are based on applying lexical analysis to the consideration of the structural features of a chemical compound. In this method, the atoms with regard to their hybridization and environment are put in correspondence with lexemes, which are later combined in “words” following a specified grammar (a set of rules defining how to combine lexemes); the words are complex fragments, and their presence in a molecule is of interest to a researcher. As previously [3], we determined whether a molecule had functional groups with known contributions to the total acidity and basicity of the molecule [8]. However, the models constructed with the use of only the above fragmental descriptors had a low predictive power. In this context, we also calculated descriptors taking into account partial atomic charges determined by the Gasteiger scheme [9, 10] and the Fukui indices [11].

Moscow State University, Moscow, 119992 Russia

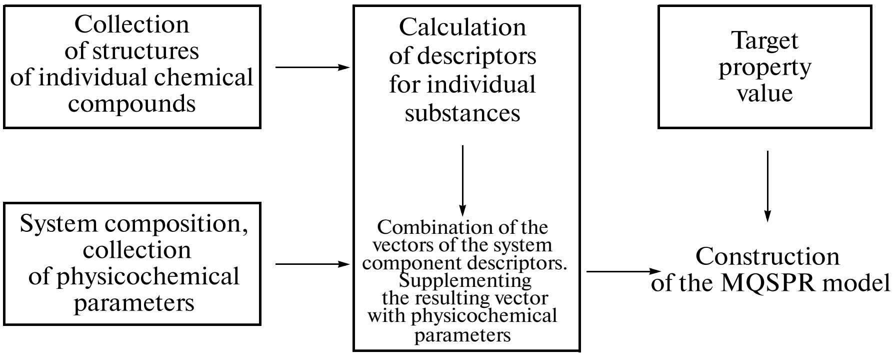

Collection of structures of individual chemical compounds
Calculation of descriptors for individual substances
Target property value
System composition, collection of physicochemical parameters
Combination of the vectors of the system component descriptors. Supplementing the resulting vector with physicochemical parameters
Construction of the MQSPR model

Fig. 1. Sequence of operations for constructing an MQSPR model.

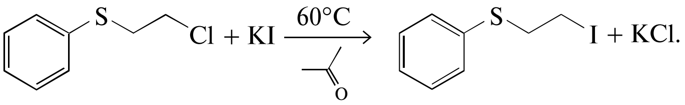
To construct the models, three reaction sets were used: the training set (1539 reactions), the validation (193 reactions), and the set for independent prediction (192 reactions).

The fragmental descriptors were calculated for the molecules of nucleophilic substitution substrates as a whole and for the nucleofuges and electrophilic moieties of the substrate molecules, for the nucleophile molecules, and for the solvents. For the substrate molecules, the quantum descriptors and descriptors based on the charge distribution were calculated. For each reaction system, descriptor vectors were obtained by combining the descriptors of its components, individual compounds. The resulting vector was augmented with temperatures, solvatochromic parameters, and Palm parameters for the solvent used in the reaction (Fig. 1).

Significant descriptors were selected using the fast stepwise linear regression procedure built into the NASAWIN program package [12]. For predicting the reaction rate constants, 142 descriptors were selected.

Structure–property relationships were searched for with the use of artificial neural networks, a mathematical algorithm that makes it possible to study nonlinear dependences with an a priori unknown character of the influence of the input data on the output ones. We chose a three-layer neural network architecture with eight neurons in the hidden layer. The resulting models are characterized by rather high correlation coefficient, maximal values for some models being as high as 0.97.

It is worth noting that, despite a large total number of descriptors used for constructing the model, only some of them turn out to be significant (nonzero) in prediction of a specific reaction rate. Below is the scheme of the S_N2 reaction of thiophenylethyl chloride with potassium iodide in acetone; the experimental log*k* value is –4.39 and the predicted, –4.49.



Among the descriptors describing the substrate (thiophenylethyl chloride), nonzero descriptors (parenthesized) are those reflecting the number of chains of three atoms connected by simple bonds (3), the number of chains of two saturated carbon atoms and the chlorine atom (1), the electrophilic superdelocalizability 1 (–0.2335), and the LUMO energy (0.1378). Separately for the electrophilic substrate moiety (thiophenylethyl), the following descriptors are nonzero: the number of chains of four aromatic carbon atoms, the sulfide S atom, and two sp^3-hybridized carbon atoms (1), the number of saturated carbon atoms at the reaction site (1), the number of chains of two carbon atoms connected by simple bonds involving the reaction site (1), the mean bond dipole moment (0.033), the smallest charge on the carbon atom (–0.0394), and the ratio of the maximal charge to the sum of all positive charges (0.106). For the nucleofuge (chloride), the nonzero descriptors are the number of chloride atoms (1) and the Sanderson

Statistical parameters of the models for prediction of S_N2 reaction rates

Set	Number of reactions	R^2		q^2	*RMSE*
		best	mean		
Training	1539	0.946	0.888	–	0.42
Validation	192	0.849	0.798	0.80	0.56
Test	193	0.794	0.783	0.79	0.58

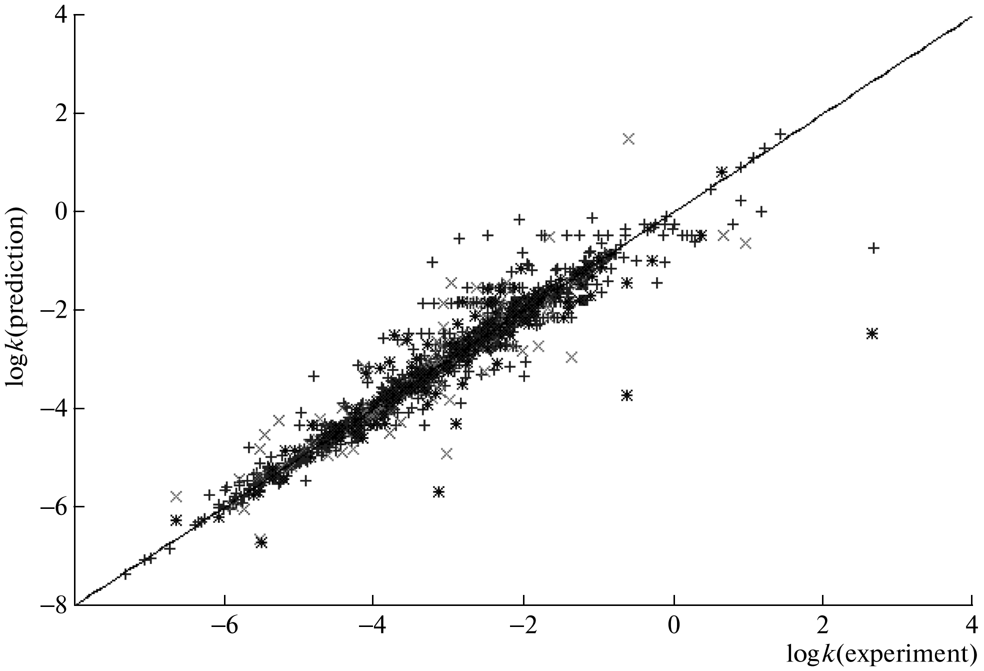


Fig. 2. Scatter of the predicted and experimental values of S_N2 reaction rate constants.

equalized electronegativity [10] (3.475). The significant descriptors for the description of the nucleophile (potassium iodide) are the sum of charges on the halogen atoms (−0.692), the sum of the charge magnitudes of all atoms of the molecule (1.38), and the n umber of iodine atoms (1). The medium is characterized by the temperature (60), the solvent dielectric constant 2 (20.74), the solvent polarizability (0.297), the total acidity of the solvent E (2.1), the number of sp^2-hybridized oxygen atoms in the solvent molecule (1), and the total number of atoms in the solvent molecule (4).

The statistical quality of the constructed models was confirmed by the cross-validation procedure. The mean and best R^2 values for different sets, q^2 values, and root-mean-square errors *RMSE* are listed in the table. A typical scatter diagram of the predicted and experimental values for the resulting models is shown in Fig. 2.

Thus, in the framework of the MQSPR methodology, we constructed the models for predicting the S_N2 reaction rate constants, which have a satisfactory predictive power and are most universal among the available models.

REFERENCES

1. Halberstam, N.M., Baskin, I.I., Palyulin, V.A., and Zefirov, N.S., *Mendeleev Commun.,* 2002, vol. 12, no. 5, pp. 185–186.

2. Hiob, R. and Karelson, M., *Comput. Chem.*, 2002, vol. 26, pp. 237–243.

3. Kravtsov, A.A., Karpov, P.V., Baskin, I.I., Palyulin, V.A., and Zefirov, N.S., *Dokl. Chem.*, 2007, vol. 414, part 1, pp. 128–131 [*Dokl. Akad. Nauk,* 2007, vol. 414, no. 3, pp. 339–342].

4. *Itogi Nauki Tekh., Ser. Obshch. Vopr. Org. Khim.*, 1977, vol. 2, part 2.

5. Palm, V.A., *Osnovy kolichestvennoi teorii organicheskikh reaktsii* (Fundamentals of Quantitative Theory of Organic Reactions), Leningrad: Khimiya, 1977.

6. Artemenko, N.V., Baskin, I.I., Palyulin, V.A., and Zefirov, N.S., *Dokl. Chem.*, 2001, vol. 381, nos. 1–3, pp. 317–320 [*Dokl. Akad. Nauk*, 2001, vol. 381, no. 2, pp. 203–206].

7. Zefirov, N.S. and Palyulin, V.A., *J. Chem. Inf. Comput. Sci.,* 2002, vol. 42, pp. 1112–1122.

8. Abraham, M.H. and Platts, J.A., *J. Org. Chem.,* 2001, vol. 66, no. 10, pp. 3484–3491.

9. Mortier, W.J., Genechten, K.V., and Gasteiger, J., *J. Am. Chem. Soc.*, 1985, vol. 107, pp. 829–835.

10. Sanderson, R.T., *J. Am. Chem. Soc.*, 1983, vol. 105, pp. 2259–2261.

11. Fukui, K., Yonezawa, T., and Shingu, H., *J. Chem. Phys.*, 1952, vol. 20, pp. 722–725.

12. Baskin, I.I., Halberstam, N.M., Artemenko, N.V., et al., in *EuroQSAR 2002 Designing Drugs and Crop Protectants: Processes, Problems and Solutions*, Ford, M. et al., Eds., Melbourne: Blackwell, 2003, pp. 260–263.

SPELL: 1. superdelocalizability, 2. polarizability