Multicomponent QSPR in Estimation of Rate of Nucleophilic Substitution at Saturated Carbon

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

Department of Chemistry, Moscow State University, 119899, Moscow, Russia

Properties of complex multicomponent systems at different conditions depend on molecular structures of several compounds. Rate constant of the reaction of nucleophilic substitution at saturated carbon is one of such properties.

The application of strict physical theory for predicting this property for real reaction systems is limitated by neccesity of using empirical parameters unknown in advance and time consuming quantum-chemical molecular simulations. To solve this problem in the QSPR framework it is necessary to describe all components of the system, including reactants, solvents, and physical conditions. Such modification of the standard QSPR procedure, which usually deals with properties and structures of individual compounds, results in the MQSPR (Multicomponent Quantitative Structure-Property Relationships) methodology.

Experimental data consisted of 3451 records on the rate constants of the nucleophilic substitution reaction, involving 360 electrophiles, 18 leaving groups, 188 reagents and 45 solvents. Substrates of the reaction were characterized with fragmental descriptors FRAGMENT and LFA (lexical fragment analysis), along with indicator variables for denoting nearest neighborhood of reaction’s center. Reagents were described by means of the same fragmental descriptors. The FRAGMENT block considers tiny details of the structure, while LFA describes large fragments, e.g. subrings of fused systems. Taken together, they provide the most comprehensive description of molecular structures.

To take into consideration solvent's effect, Palm's parameters (B, E, Y, P) were used. Temperature of reaction was taken into account as the most important physical parameter. Physical-chemical descriptors of great significance, such as B (basicity), E (polarity), T (temperature), were mapped into higher space by means of RBF transformation. To construct the quantitative model relating the above-mentioned descriptors to the kinetic rate constants, we used the backpropagation artificial neural networks. Final model can be characterized with correlation coefficient exceeding 0,979. So, in the framework of the MQSPR approach, we have succeeded in constructing the universal computational model for estimating the rate constants of the nucleophilic substitution reaction at saturated carbon for arbitrary structures of substrates, reagents, any solvents and reaction conditions.