T.R. Gimadiev[1,2]
I.V. Casciuc[1]
T.I. Madzhidov[2]
R.I. Nugmanov[2]
I.I. Baskin[2, 3]
I.S. Antipin[2]
O. Klimchuk[1]
A.A. Varnek[1,2]

CHEMICAL SPACE ANALYSIS OF S_N2 REACTIONS IN SOLUTION BASED ON THE CONDENSED REACTION GRAPH APPROACH

[1]University of Strasbourg, France;
[2]Kazan Federal University, Russia;
[3]Lomonosov Moscow State University, Russia

Timur.gimadiev@gmail.com

A chemical reaction is rather complex object in chemoinformatics because it involves several molecular species and which yield and other properties depend on experimental conditions. This complexity can be reduced in the framework of Condensed Graph of Reaction (CGR) approach [1] in which a reaction can be represented by one sole graph. In turn, a CGR can be encoded by ensemble of molecular descriptors which can be further used in any data analysis or modeling exercise.

Here, the CGR methodology has been used for chemical space analysis and building predictive models for the reaction rate (lg*k*) of S_N2 reactions in solution. A set of 4867 lg*k* values corresponding to 1394 reactions in 22 solvents and their mixtures in wide range of temperature has been carefully selected from the literature and manually curated. Then, the reactions were transformed into CGRs for which ISIDA fragment descriptors were generated.

The data analysis has been performed with the help of Generative Topographic Mapping (GTM) algorithm representing reactions as data points projected on 2D map. Figure 1 shows that the obtained map well separates reactions involving neutral and anion nucleophiles. The map also allows to identify the areas populated by substrates with particular chemotypes.

Predictive models were obtained using Support Vector Machine machine-learning method and a combination of ISIDA fragments [3], some solvent parameters [4] and inverse temperature as descriptors. 10 best individual models corresponding to different types of ISIDA descriptors and/or SVM kernels were selected in 10*5-fold cross validation and assembled in one consensus model. Its performance (RMSE=0.65 lg*k* units) is close to the experimental error of 0.5 lg*k* units. We also built several local models for different solvents and nucleophile types, most of which performed similarly to the global consensus model. However, relatively low performance has been achieved for the local model for S_N2 reactions involving anion nucleophiles. This could be explained by formation in solution of ionic pairs, which implicitly is proved by variation of lg*k* with the nucleophile concentration observed experimentally for some reactions.

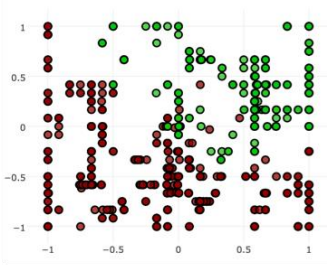

Figure 1. Map of S_N2 reaction space. Green and red dots denote reactions with anionic and neutral nucleophile respectively.

1. Madzhidov T. I. et al. *Russian Journal of Organic Chemistry*, 2014, **50 (4):** 459-463.
2. Pal`m V.A. *Usp. Khim.*, 1961, **30 (9):** 1069.
3. Varnek A. et al. *Current Computer-Aided Drug Design*, 2008, **4 (3):** 191-198.
4. Madzhidov T.I. et al. *Russian Journal of Organic Chemistry*, 2014, **50 (4):** 459-463.
5. Bishop C.M. et al. *Neural Computation*, 1998, **10(1):** 215–234.

The research was supported by Russian Scientific Foundation, grant 14-43-00024.