# [SC5] State-of-the-Art in Chemical Reaction Characteristics Prediction Using Condensed Graph of Reaction

Timur Madzhidov[1], Ramil Nugmanov[1], Timur Gimadiev[2], Marta Glavatskikh[1,2], Tagir Akhmetshin[1], Igor Baskin[1,3], Igor Antipin[1], Alexandre Varnek[1,2]

[1] *Kazan Federal University, Kazan, Russia*
[2] *University of Strasbourg, Strasbourg, France*
[3] *M.V. Lomonosov Moscow State University, Moscow, Russia*

Technology for prediction of reaction characteristics is to a large extent terra incognita in modern chemoinformatics. A the same time reaction rate and equilibrium constants are the most important characteristics of chemical reaction influencing its output (yield, conversion, time etc).
A workflow for reaction characteristics modeling was proposed. It includes the following steps: (i) reaction data collection and curation, (ii) encoding reactions by Condensed Graph of Reaction (CGR) [1], (iii) duplicate filtering and data verification, (iv) dataset analysis, (v) modeling, (vi) external validation of the model. The workflow for automatic reaction cleaning includes chemical structure standardization and reaction condition verification (temperature of reaction is valid number, above melting point of solvent and less than boiling point, etc). The Condensed Graph of Reaction (CGR) approach is a cornerstone of the approach developed. CGR signature was used for duplicated reaction identification with following rate constant verification incorporating Arrhenius equation and some expert rules. Matched Molecular Pairs approach was proposed to use in conjunction with CGR to assess substituent effect on reaction rate. Different types of fragment descriptor were used in the modeling of the reactivity parameters: Marked Atoms-based, CGR-based and mixture-based descriptors. The models predicting reaction rate of bimolecular substitution (SN2) [2], elimination (E2) [3], cycloaddition [4] as well as tautomeric equilibrium constants [5] were built and validated. The pitfalls in model validation and applicability domain application will be discussed.

Bibliography:

1. Varnek A. et al. J.Comput.-Aided. Mol. Des. 19 (2005) 693–703. doi: 10.1007/s10822-005-9008-0
2. Madzhidov T.I. et al Russ. J. Org. Chem. 50 (2014) 459–463.
3. Polishchuk P. et al. J.Comput.-Aided. Mol. Des. 31 (2017) 829–839. doi: 10.1007/s10822-017-0044-3
4. Madzhidov T.I. et al J. Struct. Chem. 58 (2017) 685–691.
5. Gimadiev T.R. et al J.Comput.-Aided. Mol. Des. (2018) In press. doi: 10.1007/s10822-018-0101-6

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