V. Afonina[1]
A. Lin[1,2]
T. Madzhidov[1]
R. Nugmanov[1]
I. Antipin[1]
O. Klimchuk[2]
A. Varnek[2]

AUTOMATIC SYSTEM FOR PREDICTION OF OPTIMAL CONDITION FOR HYDROGENATION REACTIONS

[1]Chemoinformatics and Molecular Modeling Laboratory, A.M. Butlerov Institute of Chemistry, Kazan Federal University, Kremlyovskaya street, 18, Kazan, Russia;
[2]Laboratory of Chemoinformatics, University of Strasbourg, Blaise Pascal street, 1, Strasbourg, France

ValAAfonina@kpfu.ru

Previously a prototype of the expert system for prediction of optimal catalysts for protective groups was published [1]. Present work represents the development of the concept outlined in the above paper, the expansion of its application and the correction of some shortcomings.

The aim of our investigation is the development of the tool for prediction of optimal conditions for hydrogenation reactions. The tool uses the hydrogenation reactions database that annotates information about conditions and transformations of protective groups based on dataset downloaded from Reaxys [1]. The set of considered transformation types was expanded by hydrogenation reactions involving functional groups, such as nitro-, nitroso-, alkenyl-, alkynyl group and so on. The algorithm of the prediction includes three steps:

1. Processing of the user's query and the identification of protective and functional groups existing in the user's query and their transformations.
2. Prediction of catalysts carrying out the required transformations of protective and functional groups.
3. Prediction of the optimal catalysts for the user's query.

The first step includes the detection of protective and functional groups of interest and the type of their transformation, in particular, for protective groups - cleavage or retention and for functional groups – reduction or not. On the second step for every requested protective or functional group, it is determined what kinds of transformations are most likely carried out on each catalyst. This is done using the Condensed Graph of Reaction approach (CGR) [2], the structural fingerprints based on ISIDA fragments [3] and the k nearest neighbors algorithm for statistical model development. Finaly, for every protective or functional group of interest, it will be known on which catalysts transformation will occur or will not. On the third step using this information, the optimal catalyst is selected. Furthermore, not only a prediction is returned to the user, but also examples of reactions that have requested protective or functional groups in the most similar chemical environment.

This technology allows us to predict the optimal conditions for catalytic hydrogenation reactions. In this work some shortcomings of the system, published by us earlier [1], were eliminated, the scope of its application was expanded, and in addition, the performance was significantly improved.

1. Lin A.I. et al. *Journal of Chemical Information and Modeling*, 2016, **56(11)**: 2140-2148.
2. Hoonakker F. et al. *International Journal on Artificial Intelligence Tools*, 2011, **20**: 253-270.
3. Varnek A. et al. *Journal of Computer-Aided Molecular Design*, 2005, **19**: 693-703.

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