CGR-DB CARTRIDGE IN CHEMICAL SYNTHESIS PLANNING

Fatykhova A.A.[a], Nugmanov R.I.[a], Madzhidov T.I.[a], Varnek A.[b]

[a] *A.M. Butlerov Insitute of Chemistry, Kazan Federal University, Russia*
[b] *University of Strasbourg, Strasbourg, France*
adelik21979@gmail.com

The CGR-DB cartridge is a toolkit for management a database of chemical molecules and reactions that allows you to store and input new reaction data, to perform the basic types of molecular search: structure, substructure and similarity searches. The same types of search was implemented for chemical reactions. The cartridge use algorithms that provide a high speed of reaction substructure search, that use a special type of fingerprints based on the technology of the Condensed Graph of Reaction (CGR) [1]. Using CGR allowed us to implement substructure and similarity searches that use the same engine for molecules and reactions; and to calculate the Tanimoto similarity coefficient for reactions. The cartridge also supports a substructure search for molecules by Markush simplified structures — i.e. it is possible to search for molecules with additional restrictions on atoms and their neighbors. The information about the atom type, number of neighbors, atom hybridization is supported.

This cartridge showed high performance in working with chemical data, and it makes it possible to apply its functionality to the computer planning of the chemical synthesis. The high efficiency of molecular similarity search makes it possible to efficiently organize the creation of a sequential synthetic pathway from reagents to products. The cartridge allows quickly searching for possible reactants from a databases of commercially available chemical building blocks. If necessary, the searching for a second reactant for two-component reactions is based on the analysis of the functional groups involved in the transformation using searching by Markush simplified structures.

Search algorithms have been further optimized. The search time now does not depend on the number of structures to be found. The search results became more accurate, new algorithms made it possible to distinguish even small and/or common fragments. It also became possible to extract from the database all available structures without restrictions. This is also useful in planning chemical synthesis. The available Application Programmer's Interface (API) on Python language provides an opportunity for users to integrate cartridge into their project and use it in various scientific research in the field of chemistry, medicine, chemo- and bioinformatics.

This work was supported by a subsidy provided as part of government support for Kazan Federal University to make it more competitive with leading international scientific and educational centers, and to complete the government mission in the field of scientific activity (projects № 4.1493.2017/4.6 u № 4.5151.2017/6.7).