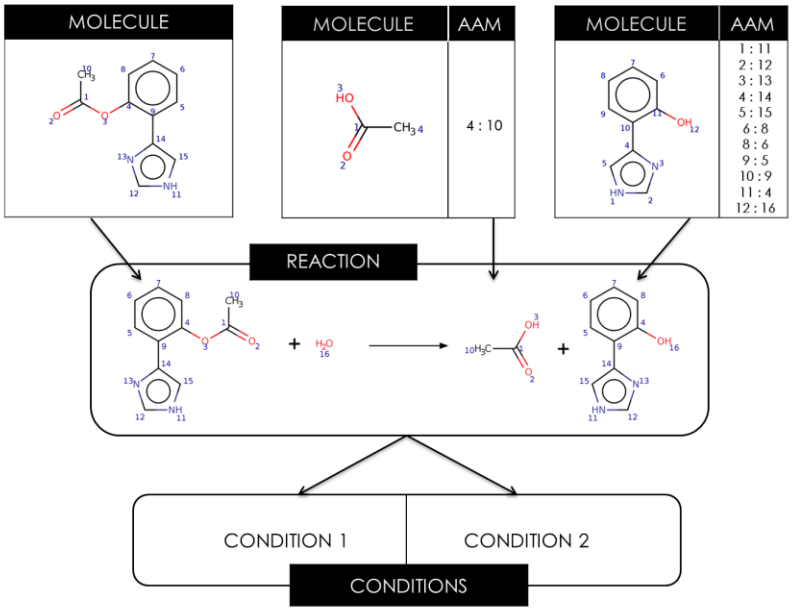
A. Fatykhova[1]
R. Nugmanov[1]
T. Madzhidov[1]
A. Varnek[1,2]

On-line reaction database cartridge with built-in condensed graph of reaction-based search engine

[1] A.M. Butlerov Insitute of Chemistry, Kazan Federal University, Russia;
[2] University of Strasbourg, Strasbourg, France.
nougmanoff@hotmail.com

Today, a large volume of data on chemical reactions has accumulated among various research teams around the world, but there is no known commercially available and convenient database cartridge to combine and store this data. Consequently, it is a good idea to create an online database of chemical reactions that will not only provide free access to information and search, but also provide researchers with a user-friendly interface for adding and modifying data.

To store reactions, we use the representation of reactions as a combination of molecules with a molecule-to-reaction mapping(AAM). Storing links to individual molecules instead of storing each molecule for each reaction saves a huge amount of memory. This approach allows to modify the structure of the molecule only once, without having to rewrite each record in different reactions. It also greatly simplifies the re-standardization. Also, now the most popular among organic chemists "reaction-by-molecule" search does not require scanning all reactions or applying complex indices to speed up the search. To simplify the structure of the data, the properties of molecules and reactions are stored in separate table in the "json" format, which can be indexed if necessary.



Using the technology of Condensed Graph of Reaction (CGR) allowed us to implement substructure and similarity searches that use the same engine for molecules and reactions. Using CGR, it became possible to calculate the Tanimoto similarity coefficient for reactions. Thanks to the use of a special type of fingerprint generation, substructure and similarity search for small and/or common fragments has become faster and more efficient.

In the framework of database management, a user-friendly mechanism was implemented to delete / add / change molecules or reactions, requiring minimal user actions and automated updating all dependencies in the database.

The available Application Programmer's Interface (API) on Python language provides an opportunity for users to integrate CGRdb database cartridge into their project. This will allow them to modify SQL-queries, manage flexible search/storage settings and many other custom settings.

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