UDC: 575.112

Development of algorithm for matching atoms in biochemical reactions

I. Anokhina [1], K. Trezubov [2]

[1]senior lecturer, ITMO University, St. Petersburg, Russia

[2] master student of the Faculty of Software Engineering and Computer Engineering of the National Research University of Information Technologies, Mechanics and Optics, St. Petersburg

Currently, there are many algorithms to predict matching of atoms in metabolic reactions, RDT, DREAM, AutoMapper, CLCA, MWED, ICMAP being the most popular.

Our study is based on the results represented in the article «Comparative evaluation of atom mapping algorithms for balanced metabolic reactions: application to Recon 3D». In this article German A. Preciat Gonzalez et al used 4000 metabolic reactions of Recon 3D library and 512 curated reactions, which are manual matching of atoms.

The experiment revealed that 5 of the 6 algorithms accurately predicted for more than 90% of the reactions which were catalyzed by oxidoreductases, with RDT giving the most accurate results. However, the accuracy of all six algorithms was very low for reactions which were catalyzed by ligases. The results of the research were published in 2017. Since then the algorithms have improved and, therefore, the number of metabolic reactions to predict has increased.

Our research aims to check the relevance of the data obtained in the article mentioned using one of the algorithms and to develop a base for new methods to find matchings in a biochemical reaction.

We have chosen RDT algorithm, as it is the most accurate, easy to install and has a console application for running by using commands. RDT also includes «AAM Tool», which allows to run commands through the console. To use this tool, we needed a special command. The algorithm gives access to the .png image, which allows to see the algorithm in action in detail, and to files with ECBLAST classes and extensions. rxn, .txt and .xml.

Moreover, there was a task to launch «AAM Tool» on reactions from Recon3D library (modern Virtual Metabolic Human). To do this, we wrote Python 3 script that could run «AAM Tool» on all available reactions.

Eventually, due to RDT algorithm operation, we obtained the results which confirmed the results of the 2017 study. We also tested the algorithm on 10% of reactions for which all 6 algorithms, including the RDT, showed errors in matching in the article 2017. Therefore, we can conclude the algorithm has improved since then.

However, there are still separate reaction groups on which RDT works with errors. After confirming the relevance of the data which were obtained by scientists in the article, we can set the following new tasks: to develop a method, which is an algorithm for finding matches of atoms for reactions based on our experimental data and results from the article.

As a base of the new method we can use RDT algorithm which is the most accurate. It is also possible to use the Gibbs free energy. To obtain data on the Gibbs energy, we can use MOPAC program.

In order to track sequentially how atoms are transferred from a substrate to a reagent and how new C – C (σ - bonds) are created, we can use a new approach. The approach should be based on the trajectory of the transition of atoms from one state to another. Further elaboration of this approach and the search for new methods for matching atoms remain a priority for the authors.

Author: K. Trezubov _____

Scientific adviser: I. Anokhina_____

Director of the Center for the Study of Foreign Languages: K. Ivanova _____