

УДК: 575.112

Разработка алгоритма для поиска соответствия атомов в биохимических реакциях

Сергушичев А.А.¹, Трезубов К.А.²

¹к.т.н., доцент факультета информационных технологий и программирования национального исследовательского университета информационных технологий, механики и оптики, г. Санкт - Петербург

²магистрант факультета программной инженерии и компьютерной техники национального исследовательского университета информационных технологий, механики и оптики, г. Санкт - Петербург

На данный момент существует множество алгоритмов для прогнозирования сопоставления атомов в метаболических реакциях, самые популярные из них - RDT, DREAM, AutoMapper, CLCA, MWED, ICMAP.

Мы рассматривали статью «Comparative evaluation of atom mapping algorithms for balanced metabolic reactions: application to Recon 3D», в которой для исследования использовали 4000 метаболических реакций библиотеки Recon 3D и 512 реакций Curated - ручное сопоставление атомов.

В ходе эксперимента авторы статьи выявили, что пять из 6 алгоритмов дали точные прогнозы для более чем 90% реакций, катализируемых оксидоредуктазами, причем наиболее точным является RDT. Однако точность всех шести алгоритмов была низкой для реакций, катализируемых лигазами. DREAM, CLCA и ICMAP были одинаково точны при прогнозировании отображений атомов для изомераз. CLCA была наиболее точной для гидролаз и лиаз. Наконец, алгоритм DREAM был наиболее точной для трансфераз.

Результаты работ ученых были опубликованы около двух лет назад, за это время алгоритмы стали лучше, метаболических реакций, которые они могут предсказывать стало намного больше.

Цель работы: проверить актуальность данных, полученных в статье, используя один из алгоритмов и разработать основу новых методов для поиска соответствия в биохимических реакция.

Мы остановились на алгоритме RDT, так как он является наиболее точным, легко поддается установке и имеет консольное приложение для запуска с использованием команд. RDT имеет в своем составе «AAM Tool», который представляет из себя .jar – архив (ReactionDecoder.jar). Для того чтобы, пользоваться данным инструментом мы использовали команду к нему «java -jar ReactionDecoder.jar -Q RXN -q reaction_file.rxn -g -j AAM -f BOTH», где «reaction_file.rxn» - это определенная метаболическая реакция, закодированная в RXN формате. На выходе мы получали изображение .png, которое позволяло детально увидеть работу алгоритма в действие и файлы с классами ECBLAST и расширениями. rxn, .txt и .xml.

Оставалась задача запустить «AAM Tool» на реакциях из библиотека Recon3D (современная Virtual Metabolic Human). Для этого мы написали код на Python 3, который мог бы запускать «AAM Tool» на всех доступных реакциях.

В ходе работы алгоритма RDT были получены результаты, которые подтверждали результаты работы ученых, которые работали над исследованием: German A. Preciat Gonzalez, Lemmer R. P. El Assal, Alberto Noronha, Ines Thiele, Hulda S. Haraldsottir и Ronan M. T. Fleming.

Также мы проверили работу алгоритма на 10% процентах реакций, для которых все 6 алгоритмов, включая RDT показывали ошибки в поиске соответствия в статье от 2017 года. Результаты работы алгоритма показали, что он был улучшен с течением времени. Так, для реакции аланинглиоксилаттрансаминазы (AGTim), которая катализирует химическое превращение L- аланина (ala_L) и глиоксилата (glx) в пируват (pyr) и глицин (gly) алгоритмом RDT было найдено верное соответствие и предсказан верный путь реакции, которая протекает с отщеплением амина группы у L – аланина и кислорода у глиоксилата. В статье, в 2017 году указывалось, что пять алгоритмов, за исключением MWED предсказывали, что разрыв C - C (σ - связи) у L – аланина энергетически наиболее выгоден. Таким образом алгоритмы, предсказывали неверный ход реакции и неверное соответствие атомов. Учитывая полученные нами результаты, можно сделать вывод, что сейчас с прогнозированием реакций, где происходит трансформации аминокислоты RDT справляется с высокой точностью

Однако, остаются все еще отдельно взятые группы реакций (реакции изомераз, гидролаз и лиаз), на которых RDT работает плохо. После подтверждения актуальности данных, полученных учеными в статье мы можем поставить перед собой новые задачи: необходимо разработать метод - алгоритм для поиска соответствия атомов для реакций, катализируемых лигаз и другие реакций на основании данных нашего эксперимента и данных из статьи.

В качестве основы нового метода может послужить алгоритм RDT, который является наиболее точным, показатели свободной энергии Гиббса. Для получения данных об энергии Гиббса мы можем использовать программу Морас.

Для того чтобы, последовательно проследить как атомы переходят из субстрата в реагент и как распадаются и создаются новые C - C (σ - связи) можно предложить подход, основанный на траектории перехода атомов из одного состояния в другое. Дальнейшая проработка этого подхода и поиски новых методов для сопоставления атомов остаются дальнейшей приоритетной задачей авторов.