

УДК 541

АВТОМАТИЗИРОВАННАЯ СИСТЕМА ДЛЯ ПОИСКА РЕАКЦИЙ, РЕАКЦИОННЫХ ЦЕНТРОВ И АНАЛИЗА ЗАЩИТНЫХ ГРУПП С ИСПОЛЬЗОВАНИЕМ КАРТРИДЖА ДАННЫХ CGR-DB**А.А. Фатыхова, Р.И. Нугманов, Р.Н. Мухаметгалеев, С.И. Мусаева, В.А. Афолина, Т.И. Маджидов, А. Варнек***Казанский федеральный университет, Казань, Россия*

Был разработан картридж для управления базой данных химических молекул и реакций, который позволяет хранить и вводить новые данные о реакциях, а также осуществлять поиск молекул по структуре, по подструктуре и по подобию. Аналогичные виды поиска были адаптированы для химических реакций. В картридже реализованы алгоритмы, обеспечивающие высокую скорость поиска реакций по подструктуре, использующие особый тип молекулярных отпечатков, основанных на технологии конденсированного графа реакции (КГР) [1]. Использование КГР позволяет использовать единый поисковый механизм для молекул и реакций, а также численно оценить схожесть реакций с помощью индекса Танимото. Картридж показал высокую эффективность работы с химическими данными, что делает его применимым для самых различных нужд. Разработанный картридж был интегрирован в проект по извлечению и анализу активности функциональных групп в условиях каталитического гидрирования. Была поставлена задача оценки активности функциональных групп (в первую очередь - защитных) в различных условиях (например, с различными катализаторами). С использованием базы данных CGR-DB была разработана система, выполняющая автоматическое извлечение защитных и функциональных групп из базы данных Reaxys с последующим статистическим анализом их реакционной способности. Реализованный в картридже поиск реакций по подструктуре позволил классифицировать трансформацию функциональных групп. Для каждого набора условий вычисляется вероятность химической трансформации. Полученная информация о реакционной способности была сравнена с экспертными знаниями, собранными вручную в книге "Защитные группы в органической химии" Т. Грин [2]. Результаты показывают, что в ряде случаев имеется хорошее соответствие между собранной вручную информацией и результатами автоматического анализа, однако были обнаружены некоторые расхождения, и в большинстве случаев было найдено, что имеющейся информации недостаточно для надежных выводов о реакционной способности групп. Помимо приведенных в книге, анализ был проведен для большого числа защитных групп, найденных автоматическим анализом. Удобные методы хранения и поиска химических данных, а также простой интерфейс позволяют использовать данный картридж в научно-исследовательских проектах в химии, медицине, хемо- и биоинформатике.

Ключевые слова: база данных, картридж, поиск по подструктуре, конденсированный граф реакции, защитные группы.

Литература

1. Varnek A. Substructural fragments: an universal language to encode reactions, molecular and supramolecular structures / A. Varnek, D. Fourches, F. Hoonakker // Journal of Computer-Aided Molecular Design. – 2005. – V. 19, Is. 9. – P. 693-703.
2. Wuts P.G.M. Greene's Protective Groups in Organic Synthesis, 2014, Wiley.