

Автоматизация Fragment-based подхода для рационального конструирования молекул в среде KNIME

Пац К. М.

Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики»

Научный руководитель – Порозов Ю. Б., к.м.н., доц. каф. ИТГС

Введение

Поиск новых лекарственных средств характеризуется высокими показателями отсева соединений на каждом из этапов исследований и разработки. Возникновение этой проблемы объясняется, в том числе, и качеством соединений-лидеров. Fragment-based подход к разработке лекарств (Fragment-Based Drug Discovery, FBDD) все чаще и чаще используется в фармацевтической индустрии, чтобы сократить количество отсеиваемых структур [1].

KNIME Analytics Platform – open-source платформа для автоматизированного многопоточного анализа данных. Данная платформа очень популярна для решения задач, связанных с разработкой лекарственных соединений, о чем свидетельствуют многочисленные расширения, поддерживающие работу с химическими данными и интеграция некоторого специализированного ПО [2]. Однако анализ литературных данных говорит о том, что KNIME чаще используется как инструмент решения узких задач по автоматизации того или иного шага [3-5]. При этом KNIME представляет богатые возможности по созданию самостоятельных инструментов. Их настройка является очень гибкой, что позволяет с легкостью оптимизировать эти инструменты под интересующую задачу.

Цель работы

Разработать пайплайн для платформы KNIME, который позволял бы автоматизировать основные этапы разработки лекарств на основе фрагментов с учетом заданной белковой мишени.

Основные положения исследования

Разрабатываемый автоматизированный пайплайн для KNIME должен осуществлять следующие процедуры.

Во-первых, это обработка данных, полученных из базы DrugBank в формате SDF. Из этих данных необходимо извлечь SMILES-код молекулы и ее ID в базе данных. Извлеченные данные записываются в новый файл, создавая таким образом новую базу, которая в дальнейшем будет использована для анализа. На этом этапе применяются стандартные узлы KNIME, а также узлы, содержащие собственный код на Python.

После формирования базы данных с SMILES и ID структур, данные отправляются на цепочку узлов KNIME, в каждом из которых находится требуемая процедура Schrödinger для проведения молекулярного докинга (подготовка мишени, подготовка лигандов, непосредственно докинг).

Далее, результаты докинга разделяются на несколько (предположительно, 2-3) категорий, в зависимости от значения GScore. Каждая категория затем обрабатывается с помощью узлов KNIME с RDKit на предмет анализа скаффолдов и функциональных групп.

Финальные этапы предполагают статистический анализ полученных результатов и проверку альтернативных способов выполнения тех или иных шагов, подбор параметров и т.д.

В результате, готовый инструмент будет выдавать пользователю набор скаффолдов и функциональных групп, которые являются перспективными с точки зрения связывания с заданной мишенью.

Результаты

В ходе работы построен пайплайн для платформы KNIME, позволяющий выявлять набор фрагментов химических соединений (скаффолдов и функциональных групп), предпочтительных для дизайна новых лекарственных средств на основе Fragment-based подхода.

С помощью данного пайплайна также возможно формировать новые базы данных, содержащие сведения, удобные для использования в других программах для поиска новых химических структур (например, ScaffoldHunter).

Список литературы

1. Murray, C. W., Rees, D. C. The rise of fragment-based drug discovery // *Nat Chem*. 2009. №1(3). P. 187–92.
2. Berthold, M. R. et al. KNIME: The Konstanz Information Miner // *Data Analysis, Machine Learning and Applications. Proceedings of the 31st Annual Conference of the Gesellschaft für Klassifikation e.V., Albert-Ludwigs-Universität Freiburg, March 7–9, 2007*. P. 319–326.
3. Geldenhyus, W. J. et al. Identification of novel monoamine oxidase B inhibitors by structure-based virtual screening // *Bioorganic & Medicinal Chemistry Letters*. 2010. №20(17). P. 5295–5298.
4. Poongavanam, V., Kongsted, J. Virtual Screening Models for Prediction of HIV-1 RT Associated RNase H Inhibition // *PLoS One*. 2013. №8(9): e73478.
5. Melagraki, G., Afantits, A. Enalos in silico nano platform: an online decision support tool for the design and virtual screening of nanoparticles // *RSC Advances*. №4(92). 2014. P.50713–50725.