

УДК 004.94

ПРИМЕНЕНИЕ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ СОЗДАНИЯ ПОТЕНЦИАЛЬНЫХ ЛЕКАРСТВЕННЫХ ПРЕПАРАТОВ

А. Д. КАРПЕНКО

Объединённый институт проблем информатики НАН Беларуси
Минск, Беларусь

Введение. Компьютерное моделирование потенциальных лекарственных препаратов стало неотъемлемым этапом в разработке новых лекарств при нацеливании на конкретные биологические мишени. В частности, машинное обучение стало мощным инструментом для генерации и идентификации новых лекарственных соединений. Вычислительные подходы могут не только сэкономить время и средства, затрачиваемые на скрининг *in vitro*, путем предоставления списка потенциальных нецелевых объектов, но и дать представление о молекулярных механизмах взаимодействия белок-лекарство. Было показано, что потенциальные мишени могут быть идентифицированы *in silico* путем установления взаимосвязи структура-активность малых молекул [1].

Основная часть. Актуальным является поиск новых мультикиназных ингибиторов, обладающих меньшей токсичностью и снижающих риск возможного возникновения резистентности к используемым препаратам, связанной с их длительным применением, в частности для терапии хронического миелоидного лейкоза. Нами была разработана архитектура гетероэнкодера, которая состоит из трех энкодеров и двух декодеров. Для создания модели использовалась открытая библиотека для разработки нейронных сетей Keras (<https://keras.io/>).

Для обучения был сформирован обучающий набор данных из соединений базы PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) содержащих 2-ариламинопиримидин. Молекулярный докинг обучающего набора выполнялся программой AutoDock Vina (<https://vina.scripps.edu>) в приближении жесткого рецептора и гибких лигандов. В результате обучающий набор составил 120 000 соединений с энергией связывания с BCR-ABL тирозинкиназой.

На вход нашей модели подаются представления молекул в виде SMILES и канонический SMILES. Были реализованы следующие подмодели: в качестве энкодеров для SMILES формата и для Канонического SMILES формата была выбрана архитектура, состоящая из двух LSTM слоев. Входные данные обрабатываются первым LSTM слоем, и далее, результаты работы слоя передаются на второй LSTM слой, а состояния с первого слоя передаются в качестве эмбедингов. Второй слой также обрабатывает данные и после передает

состояния в качестве эмбедингов. Состояния из двух последовательных LSTM ячеек являются эмбедингами для строкового формата.

Для оценки результатов работы гетероэнкодера сгенерированные соединения подверглись процедуры молекулярного докинга. Было сгенерировано 1 117 оригинальных SMILES, которым соответствуют 1 117 оригинальных химических структур. Полученные данные были очищены от дубликатов, проверены на валидность, интерпретируемость и содержание 2-ариламинопиримидина с помощью модуля RDKit (<http://www.rdkit.org/>). Успешно были отобраны 1083 соединения.

С помощью программы AutoDock Vina (<https://vina.scripps.edu>) проводили молекулярный докинг успешно прошедших проверку соединений с рентгеновскими структурами Vcr-Abl тирозинкиназы (PDB ID: 3OXZ; <https://www.rcsb.org>) и ее мутантной формы Vcr-Abl^{T315I} (PDB ID: 3OY3; <https://www.rcsb.org>). Молекулярный докинг проводили в приближении жесткого рецептора и гибких лигандов с помощью программы AutoDock Vina (<https://vina.scripps.edu>). Ячейки для докинга охватывали АТФ-связывающий сайт Vcr-Abl тирозинкиназы.

Значение параметра, характеризующего полноту поиска, задавали равным 100. Была получена выборка соединений из 569 молекул, проявляющих высокое сродство к АТФ-связывающим сайтам Vcr-Abl тирозинкиназы и ее мутантной формы Vcr-Abl^{T315I}.

Согласно данным молекулярного докинга, отобранные соединения имеют низкие значения энергии связывания, варьирующие от $-9,0$ до $-13,8$ ккал/моль.

Заключение. Результаты исследования показывают, что разработанная нейронная сеть представляет собой эффективную математическую модель для генерации малых молекул с низкой энергией связывания с VCR-ABL тирозинкиназой и её мутантной формой VCR-ABL (T315I), и разработку на их основе новых противоопухолевых препаратов против острого миелоидного лейкоза.

Анализ полученных результатов показывает, что комбинированный подход с использованием методов машинного обучения и молекулярного докинга эффективен для генерации потенциальных лекарственных препаратов, терапевтическое действие которых основано на ингибировании ранних стадий поздних стадий миелоидного лейкоза.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. **Durrant, J.** Molecular dynamics simulations and drug discovery / J. Durrant, J. McCammon. – BMC Biology, 2011. – 9 p.