

Применение методов искусственного интеллекта и компьютерного моделирования для создания потенциальных противоопухолевых препаратов

Войтко Т. Д.¹, Фурс К. В.², Карпенко А. Д.²

¹Белорусский государственный университет

²Объединенный институт проблем информатики НАН Беларуси

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].

Открытие природы ингибирования опухолевых процессов малыми органическими молекулами изменило принципы разработки лекарственных соединений для противоопухолевой терапии. Последние достижения в этой области связаны с созданием низкомолекулярных мультикиназных ингибиторов [2].

За последние три десятилетия было установлено, что множественные злокачественные опухоли человека обусловлены модуляцией и дисфункцией протеиновых и липидных киназ и дезактивированных фосфатаз из-за перестановок хромосом и генетических мутаций. С 2001 года было зарегистрировано 28 низкомолекулярных ингибиторов АТФ-связывающего сайта киназ, одобренных Управлением по санитарному надзору за качеством пищевых продуктов и медикаментов (Food and Drug Administration, FDA) для терапии онкологических заболеваний [2]. Однако все эти соединения проявляют высокую токсичность, вызывающую ряд гематологических и других разнообразных побочных эффектов. Кроме того, у большинства пациентов возникает резистентность к применяемым препаратам, приобретаемая после длительной химиотерапии [3].

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 3 энкодеров и 2 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 виду SMILES и канонический SMILES. Были реализованы следующие подмодели: в качестве энкодеров для SMILES формата и для Канонического SMILES формата была выбрана архитектура, состоящая из 2х 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}.

описать кросс ранжирование

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

Выводы. 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.

Литература.

1. Durrant, J. Molecular dynamics simulations and drug discovery / J. Durrant, J. McCammon. - BMC Biology, 2011. - 9 p.
2. Garuti L. Multi-kinase inhibitors / L. Garuti, M. Roberti, G. Bottegoni. - Medicinal Chemistry, 2015. - 6 p.
3. Krug, M. Recent Advances in the Development of Multi-Kinase Inhibitors / M. Krug, A. Hilgeroth. - Halle : Mini-Reviews in Medicinal Chemistry, 2008. - 16 p.