

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УДК 51-76:577.322:539.19:004.94

НИКОЛАЕВ
Григорий Игоревич

Разработка алгоритмов, основанных на технологии глубокого обучения и молекулярном моделировании, для конструирования потенциальных ингибиторов проникновения ВИЧ-1

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

по специальности 05.13.18 – математическое моделирование, численные методы и комплексы программ

Минск, 2022

Научная работа выполнена
в ГНУ «Объединенный институт проблем информатики НАН Беларуси».

Научный руководитель – Андрианов Александр Михайлович,
доктор химических наук,
главный научный сотрудник лаборатории
белковой инженерии
ГНУ Институт биоорганической химии
НАН Беларуси».

Официальные оппоненты: Головки Владимир Адамович,
доктор технических наук, профессор,
заведующий кафедрой интеллектуальных
информационных технологий
УО «Брестский государственный
университет имени А.С. Пушкина»;

Ковалев Василий Алексеевич,
кандидат технических наук,
заведующий лабораторией анализа
биомедицинских изображений
ГНУ «Объединенный институт проблем
информатики НАН Беларуси».

**Оппонирующая организация – УО «Белорусский государственный
университет информатики и
радиоэлектроники».**

Защита состоится **«23» июня** 2022 г. в **14.00** часов на заседании совета
по защите диссертаций Д 02.01.02 при Белорусском государственном
университете по адресу: 220030, г. Минск, ул. Ленинградская, 8
(корпус юридического факультета), ауд. 407.

Телефон ученого секретаря – (017) 209-57-09.

Почтовый адрес: пр-т Независимости 4, Минск, 220030.

С диссертацией можно ознакомиться в Фундаментальной библиотеке
Белорусского государственного университета.

Автореферат разослан « 19 » мая 2022 г.

Ученый секретарь
совета по защите диссертаций
кандидат физ.-мат. наук доцент



Е.С. Чеб

ВВЕДЕНИЕ

Современные методы компьютерного конструирования потенциальных лекарств значительно расширяют возможности фармацевтической индустрии, позволяя существенно сократить время и затраты, необходимые для создания новых терапевтических средств. Для реализации *de novo* дизайна весьма перспективным представляется применение методологии клик-химии, позволяющей генерировать большое число структур-кандидатов фармацевтических препаратов путем соединения между собой отдельных структурных блоков.

Несмотря на то, что эффективность традиционных методов компьютерного конструирования лекарств в настоящее время является общепризнанной, разработка новых математических подходов способствует их постоянному совершенствованию. Среди этих подходов важное место занимают методы машинного обучения (Machine Learning) и, в частности, методы глубокого обучения (Deep Learning), которые имеют большой потенциал для дальнейшего прогресса в данной области исследований.

В настоящей работе разработан алгоритм для компьютерного конструирования лекарственных препаратов на основе данных о структуре молекулярной мишени, использующий технологии глубокого обучения, *in silico* клик-химии и молекулярного моделирования. Алгоритм апробирован в процессе генерации и компьютерного тестирования потенциальных ингибиторов проникновения ВИЧ-1, блокирующих CD4-связывающий участок белка gp120. В рамках концепции *in silico* клик-химии осуществлен компьютерный дизайн новых ингибиторов проникновения ВИЧ-1 и идентифицированы соединения, перспективные для разработки новых эффективных противовирусных препаратов. Методами фармакофорного моделирования обнаружены химические соединения, имитирующие структурно-функциональные свойства анти-ВИЧ антитела N6, которое нейтрализует до 98% протестированных штаммов ВИЧ-1, включая 16 из 20 штаммов, резистентных к другим антителам, взаимодействующим с CD4-связывающим участком белка gp120.

Идентифицированные соединения формируют перспективные базовые структуры для создания эффективных анти-ВИЧ препаратов нового поколения.

Актуальность, научная и практическая значимость работы обусловлены тем, что в настоящее время нет лицензированных анти-ВИЧ препаратов, ингибирующих участок белка gp120 оболочки вируса, критически важный для его проникновения в клетку-мишень.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Связь работы с научными программами (проектами), темами

Диссертационная работа выполнена в Объединенном институте проблем информатики Национальной академии наук Беларуси в рамках следующих проектов: задание 3.2.05 «Исследование структурных свойств и мутаций вируса СПИДа на основе компьютерного моделирования, химического синтеза и медицинского тестирования» Государственной программы научных исследований на 2011-2015 годы «Конвергенция» (№ г.р. 20130278); задание 3.08 «Молекулярное конструирование и компьютерный скрининг потенциальных лекарственных препаратов против ВИЧ, блокирующих CD4-связывающий участок белка gp120 оболочки вируса» Государственной программы научных исследований на 2016-2020 годы «Конвергенция-2020» (№ г.р. 20162002); проект X15-022 Белорусского республиканского фонда фундаментальных исследований «Скрининг и молекулярное моделирование новых потенциальных ингибиторов ВИЧ-1 на основе моноклонального антитела 10E8, обладающего широким спектром нейтрализующей активности» (№ г.р. 20150985); проект X17МС-004 Белорусского республиканского фонда фундаментальных исследований «Молекулярный *de novo* дизайн потенциальных ингибиторов проникновения ВИЧ-1 на основе методологии клик химии» (№ г.р. 20171266).

Цель и задачи исследования

Цель исследования: Методами глубокого обучения разработать генеративный состязательный автоэнкодер для генерации молекулярных дескрипторов химических соединений, способных блокировать участок белка gp120 ВИЧ-1, критический для связывания вируса с клеточным рецептором CD4. С помощью комплексного подхода, включающего разработанную нейронную сеть, виртуальный скрининг, *in silico* клик-химию и методы молекулярного моделирования, идентифицировать молекулы, перспективные для создания эффективных анти-ВИЧ препаратов нового поколения.

В соответствии с поставленной целью в диссертации были сформированы следующие *задачи*.

1. Разработать методами глубокого обучения генеративный состязательный автоэнкодер, включая следующие этапы:
 - а) создать архитектуру автоэнкодера;

б) сформировать молекулярную библиотеку потенциальных ингибиторов белка gp120 ВИЧ-1 для обучения и тестирования нейронной сети;

в) провести докинг молекул из обучающей библиотеки с белком gp120 и расчет свободной энергии их связывания с молекулярной мишенью;

г) сгенерировать «молекулярные отпечатки пальцев» (фингерпринты) химических соединений обучающего набора данных;

д) провести обучение нейронной сети;

е) оценить результаты обучения и работы автоэнкодера.

2. Применить разработанную нейронную сеть для генерации молекулярных дескрипторов химических соединений с заданными физико-химическими параметрами и компьютерный скрининг потенциальных ингибиторов проникновения ВИЧ-1 методами молекулярного моделирования.

3. Разработать алгоритм для конструирования потенциальных ингибиторов белка gp120 ВИЧ-1 в рамках методологии *in silico* клик-химии и оценить их анти-ВИЧ активность с помощью методов молекулярного докинга, машинного обучения, квантовой химии и молекулярной динамики.

4. Разработать методику для предсказания потенциальных ингибиторов ВИЧ-1, имитирующих нейтрализующее антитело N6 широкого спектра действия, использующую методы виртуального скрининга, молекулярного докинга, квантовой химии и молекулярной динамики.

Объект исследования: низкомолекулярные соединения – ингибиторы ВИЧ-1.

Предмет исследования: математические методы и алгоритмы конструирования потенциальных ингибиторов ВИЧ-1 белка gp120 оболочки вируса.

Научная новизна

Новизна полученных в диссертации результатов заключается в следующем:

несмотря на то, что методы глубокого обучения и молекулярного моделирования в последние годы широко используются для компьютерного конструирования потенциальных лекарств, в настоящее время в литературе отсутствуют данные по их применению в разработке потенциальных лекарственных препаратов против ВИЧ-1, терапевтическое действие которых основано на блокаде белка gp120 оболочки вируса. В связи с этим, для решения этой проблемы в диссертационной работе разработаны новые алгоритмы для конструирования потенциальных ингибиторов проникновения ВИЧ-1, основанные на технологии глубокого обучения и молекулярного

моделирования.

Положения, выносимые на защиту

1. Генеративный состязательный автоэнкодер на основе технологий глубокого обучения и его применение для генерации и идентификации потенциальных ингибиторов проникновения ВИЧ-1 совместно с методами виртуального скрининга и молекулярного моделирования.

2. Алгоритм для компьютерного конструирования лекарственных препаратов против ВИЧ-1, использующий методологию клик-химии, методы молекулярного моделирования и машинного обучения, с помощью которого предсказаны пять химических соединений, обладающих потенциальной анти-ВИЧ активностью.

3. Методика компьютерного моделирования потенциальных ингибиторов ВИЧ-1, имитирующих нейтрализующее антитело №6 широкого спектра действия, основанная на использовании методов виртуального скрининга, молекулярного докинга, квантовой химии и молекулярной динамики, которая позволила идентифицировать 9 перспективных химических соединений.

Личный вклад соискателя ученой степени

Все основные результаты, изложенные в диссертационной работе, получены автором самостоятельно. Определение темы и целей диссертационной работы, выбор методов исследования, анализ и обобщение полученных результатов проведены совместно с научным руководителем д.х.н. Андриановым А.М. Соавторами оказана консультационная (д.ф.-м.н., профессор, чл.-корр. Тузиков А.В.) и техническая (к.х.н. Кашин И.А., к.х.н. Корноушенко Ю.В., Шульдов Н.А, Босько И.П., Анищенко А.И.) помощь в выполнении работы. Остальные соавторы занимались вопросами, не вошедшими в диссертацию.

Апробация диссертации и информация об использовании ее результатов

Основные результаты диссертации были представлены на Международных конференциях: XII Международная научная конференция молодых ученых «Молодежь в науке – 2015» (Минск, 01-04 декабря 2015), «Молекулярные, мембранные и клеточные основы функционирования биосистем» XII съезд Белорусского общественного объединения фотобиологов и биофизиков (Беларусь,

Минск, 28-30 июня 2016), 1 Белорусский Биохимический Конгресс (Беларусь, Гродно, 5-6 июля 2016), XXX Международная научная конференция «Математические методы в технике и технологиях» (Минск, 30 мая – 02 июня 2017), «Moscow Conference on Computational Molecular Biology» (Россия, Москва, 27-30 июля 2017, 27-30 июля 2019), VI International Conference “Chemistry, Structure and Function of Biomolecules” (Беларусь, Минск, 22-25 мая 2018), 4th International Turkic World Conference on Chemical Sciences and Technologies (Украина, Киев, 7-10 сентября 2018), 2nd International Symposium on Virus Entry and Entry Inhibitors (China, Shenzhen, October 28-30, 2018), 2nd International Symposium on Infectology and Application (China, Shanghai, October 21-23, 2019), Национальный суперкомпьютерный форум (НСКФ’2019) (Россия, Переславль-Залесский, 26–29 ноября 2019), II Международная научно-практическая конференция «Компьютерные технологии и анализ данных» (Беларусь, Минск, 23-24 апреля 2020), Международная конференция «Математическая биология и биоинформатика» (Пушино: ИМПБ РАН, Россия, 12-19 октября 2020; онлайн конференция), 16th International Symposium “Bioinformatics Research and Applications” (Moscow, December 1-4, 2020).

Победитель республиканского молодежного инновационного проекта «100 идей для Беларуси» в номинации «Информационно-коммуникационные и авиакосмические технологии».

Имеется 2 акта о практическом использовании результатов исследования.

Опубликованность результатов диссертации

Основные результаты диссертации опубликованы в 23 научных работах, из которых 9 статей в научных изданиях в соответствии с п. 18 Положения о присуждении ученых степеней и присвоении ученых званий в Республике Беларусь (общим объемом 11,9 авторского листа), 1 статья в журнале «Наука и инновации», 10 статей в сборниках материалов научных конференций, 3 тезисов.

Структура и объем диссертации

Диссертация состоит из перечня условных обозначений, введения, пяти глав, заключения, библиографического списка и трех приложений. Полный объем диссертации составляет 135 страниц, в том числе 30 рисунков занимают 29 страниц, 15 таблиц на 15 страницах, три приложения занимают 17 страниц. Библиографический список содержит 209 наименований, включая собственные публикации соискателя ученой степени (на 19 страницах).

ОСНОВНАЯ ЧАСТЬ

В **первой главе** приведен обзор литературных данных о методах компьютерного конструирования лекарств, методологии клик-химии и вирусе иммунодефицита человека. Представлена информация о методах моделирования потенциальных лекарств, основанных на данных о структуре молекулярной мишени (SBDD; Structure-Based Drug Design) или структуре известного лиганда, обладающего биологической активностью (LBDD; Ligand-Based Drug Design). Подробно описана концепция клик-химии - нового подхода к синтезу молекул, который может ускорить процесс обнаружения лекарств с помощью нескольких практических и надежных реакций. Рассмотрены современные методы машинного обучения. Такие методы могут применяться для построения полносвязных нейронных сетей, автоматического и прямого извлечения дескрипторов из химических структур, перепрофилирования лекарств, прогнозирования структуры белка, виртуального скрининга потенциальных лекарств и предсказания аффинности их связывания с молекулярной мишенью. Приведена информация о структуре ВИЧ, механизме его проникновения в клетку-хозяина.

Вторая глава диссертации посвящена описанию методов компьютерного конструирования лекарств, которые использовались в работе для разработки генеративного состязательного автоэнкодера для генерации молекулярных дескрипторов химических соединений, способных блокировать участок белка gp120 ВИЧ-1 и идентификации молекул, перспективных для создания эффективных анти-ВИЧ препаратов нового поколения.

В диссертационной работе применен ряд методов машинного и глубокого обучения, включающий использование активационных функций, технологий дропаута, функций потерь, метода оптимизации Adam и алгоритма tSNE.

Основные шаги алгоритма, использованного для компьютерного дизайна потенциальных лекарственных препаратов на основе методологии клик-химии, приведены на рисунке 1.

В диссертационной работе использовались следующие методы молекулярного моделирования: молекулярный докинг, полуэмпирическое квантово-химическое моделирование и молекулярная динамика.

Для проведения молекулярного докинга наборов химических соединений с молекулами-мишенями применяли программный пакет QuickVina 2 с учетом конформационной подвижности лиганда, перебирая все его возможные ориентации относительно молекулы-рецептора.

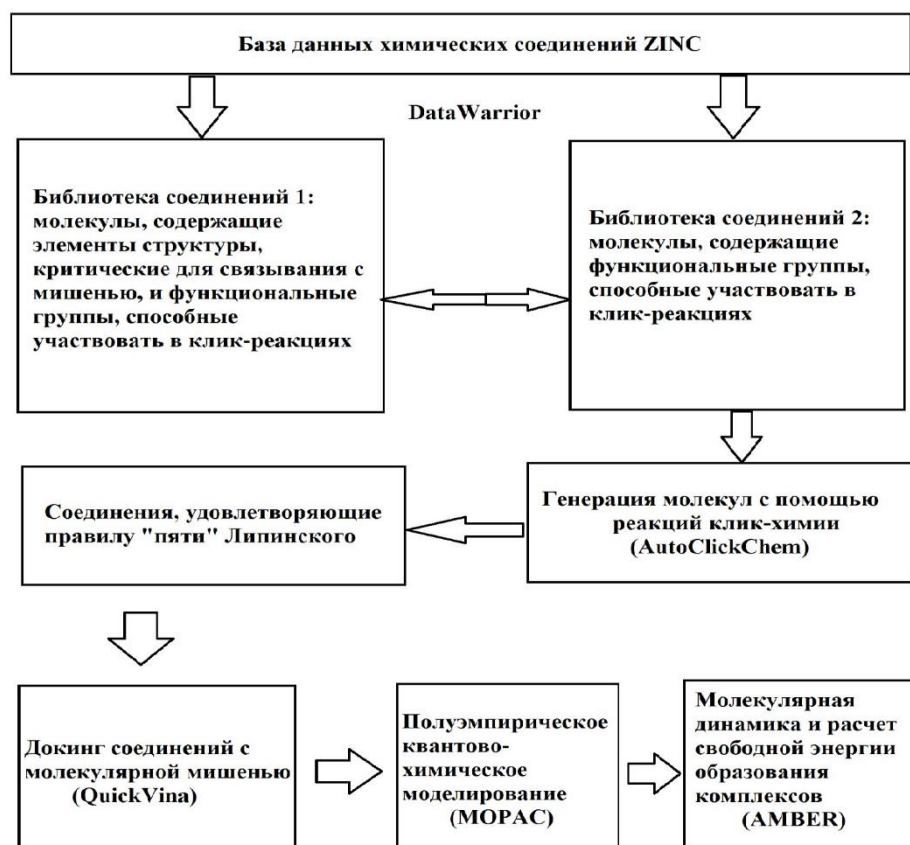


Рисунок 1. - Основные шаги алгоритма для компьютерного дизайна потенциальных лекарственных препаратов

Для проведения полуэмпирического квантово-химического моделирования наборов химических соединений с молекулами-мишенями применяли программный пакет MOPAC с методом PM7 для уточнения ориентации лиганда относительно молекулы рецептора.

Молекулярной динамикой (МД) называется метод компьютерного моделирования, основанный на численном решении уравнений движения для системы с большим (но конечным) количеством степеней свободы. Для моделирования белковых взаимодействий используются специальные силовые поля с экспериментально подобранными параметрами. Одним из таких силовых полей является поле Amber, которое было применено в настоящей работе для анализа структурных комплексов лигандов с молекулярными мишенями.

Расчеты проводили на суперкомпьютере «СКИФ-ОИПИ» Объединенного института проблем информатики НАН Беларуси.

Для формирования начальной выборки потенциальных лигандов, миметиков кросс-реактивного антитела N6 ВИЧ-1, применяли веб-сервис

Pharmit, который позволяет проводить поиск потенциальных лигандов на основе заданной трехмерной структуры молекул.

В третьей главе описана разработка генеративного состязательного автоэнкодера для идентификации потенциальных ингибиторов ВИЧ-1 методами глубокого обучения.

Для решения поставленной задачи были выполнены исследования, включающие следующие этапы:

- 1) создание архитектуры состязательного автоэнкодера;
- 2) формирование молекулярной библиотеки потенциальных лигандов белка gp120 ВИЧ-1 для обучения нейронной сети;
- 3) молекулярный докинг лигандов с белком gp120 и расчет свободной энергии связывания;
- 4) генерацию молекулярных дескрипторов химических соединений обучающего набора данных;
- 5) обучение нейронной сети;
- 6) оценку результатов обучения и работы состязательного автоэнкодера.

Архитектура разработанного состязательного автоэнкодера состоит из двух нейросетей – автоэнкодера и дискриминатора, работающих во время обучения в соревновательном режиме. Автоэнкодер представляет собой семислойную нейронную сеть, имеющую входной и выходной слои, латентный слой, а также 4 полносвязных слоя (рисунок 2). На входной слой подаются молекулярные дескрипторы химических соединений, данные о которых проходят энкодер и попадают на латентный слой, где к полученному результату добавляется численная оценка энергии связывания с молекулярной мишенью. Далее молекулярные дескрипторы проходят декодер и попадают на выход, который представляет собой вектор молекулярного дескриптора. Латентный слой состоит из 3 нейронов, два из которых получают значения от энкодера, а третий получает значение энергии связывания с молекулярной мишенью. В рабочем режиме автоэнкодера на латентный слой подаются случайные числа, которые затем проходят через декодер, генерирующий молекулярные дескрипторы молекул с требуемыми свойствами.

Для обучения автоэнкодера использовались следующие параметры:

- количество эпох для главной модели, используемой для генерации, – 400;
- скорость обучения всего автоэнкодера на первой ступени итерации – 0,005;
- скорость обучения дискриминатора на второй ступени итерации – 0,001;
- скорость обучения энкодера на третьей ступени итерации – 0,005;
- параметр Batch size – 128;
- оптимизатор – метод Adam.

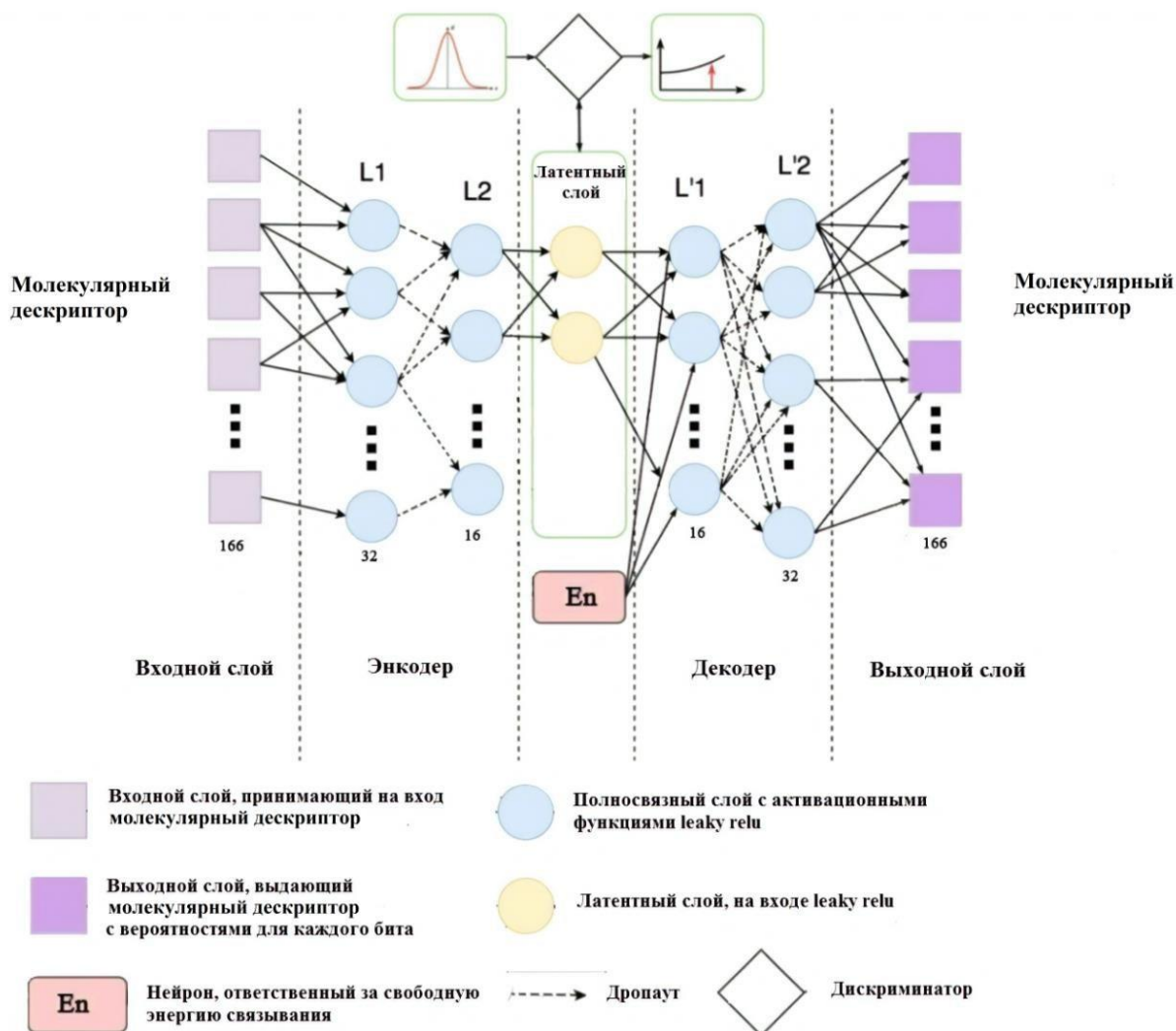


Рисунок 2. - Архитектура нейронной сети для генерации потенциальных ингибиторов ВИЧ-1, блокирующих CD4-связывающий сайт белка gp120 оболочки вируса

Для тестирования работы автоэнкодера с помощью программного пакета RDKit была создана библиотека молекулярных дескрипторов MACCS для 21 325 567 соединений из библиотеки Drug-Like базы данных ZINC15 и рассчитаны пять молекулярных дескрипторов для сгенерированных автоэнкодером молекул при пороговом значении энергии связывания с белком gp120, равном -5 ккал/моль. В результате виртуального скрининга для каждой из этих молекул были найдены лиганды с подобными дескрипторами.

Анализ результатов молекулярного докинга найденных соединений с белком gp120 показал, что совместное использование нейронной сети с виртуальным скринингом библиотеки молекулярных дескрипторов позволяет идентифицировать лиганды с более низкой по сравнению с заданным пороговым значением энергией связывания (рисунок 3).

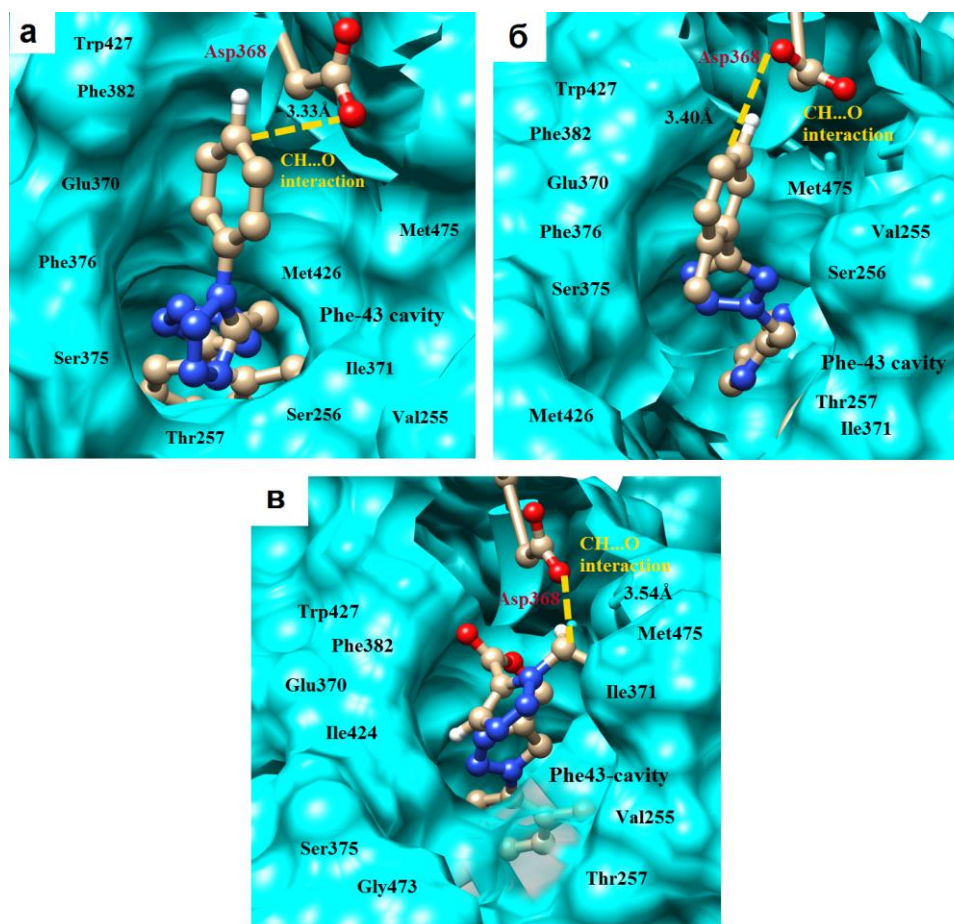


Рисунок 3. - Комплексы соединений ZINC000026430653 (а), ZINC000191389930 (б) и ZINC000293423658 (в) с белком gp120, построенные методом молекулярного докинга и уточненные с помощью квантово-химических расчетов методом PM7

Моделирование молекулярной динамики подтверждает данные о высоком сродстве связывания идентифицированных соединений с белком gp120, полученные в результате анализа статических моделей комплексов лиганд/gp120. Эти комплексы обнаруживают относительную стабильность в течение МД расчетов, о чем свидетельствуют средние значения свободной энергии связывания и их стандартные отклонения.

В четвертой главе описывается разработка алгоритма для компьютерного конструирования лекарственных препаратов против ВИЧ-1 на основе методологии клик-химии, молекулярного моделирования и машинного обучения. Для решения поставленной задачи были выполнены исследования, включающие следующие этапы:

1) с помощью компьютерного скрининга базы данных ZINC15 идентифицированы низкомолекулярные химические соединения, которые содержат функциональные группы, участвующие в реакции азид-алкинового циклоприсоединения – основной реакции клик-химии;

2) осуществлен компьютерный дизайн структур-кандидатов с привлечением найденных ранее соединений в качестве исходных структурных блоков;

3) проведен отбор молекул, удовлетворяющих правилу “пяти” Липинского;

4) методами молекулярного докинга построены структурные комплексы полученных сконструированных молекул с белком gp120 ВИЧ-1;

5) построены МД траектории комплексов этих соединений с белком gp120 ВИЧ-1 и рассчитаны свободные энергии их образования;

6) определены вклады индивидуальных аминокислотных остатков белка gp120 в энтальпию связывания и выполнен анализ межмолекулярных водородных связей, реализующихся в динамических структурах комплексов.

На основе полученных данных отобраны пять соединений-лидеров, характеризующихся низкими значениями свободной энергии связывания (рисунок 4). Данные молекулярной динамики показывают, что сконструированные соединения могут эффективно блокировать функционально консервативные элементы структуры CD4-связывающего белка gp120 участка ВИЧ-1, что указывает на высокую вероятность проявления ими противовирусной активности.

Пятая глава посвящена описанию разработки компьютерной методики для предсказания потенциальных ингибиторов ВИЧ-1 (рисунок 4), которая включает в себя следующие этапы: построение компьютерной фармакофорной модели на основе антитела, виртуальный скрининг и отбор перспективных химических соединений, проведение компьютерного моделирования, состоящего из молекулярного докинга, квантово-химического моделирования и молекулярной динамики. В результате использования данной методики идентифицированы молекулы, перспективные для создания новых эффективных противовирусных препаратов широкого спектра действия.

Для достижения поставленной цели были выполнены следующие исследования:

1) построены модели фармакофоров, описывающих совокупность структурно-функциональных свойств анти-ВИЧ антитела N6, обеспечивающих специфичность его взаимодействия с CD4-связывающим участком белка gp120 оболочки вируса;

2) проведен виртуальный скрининг молекулярных библиотек веб-сервера Pharmit;

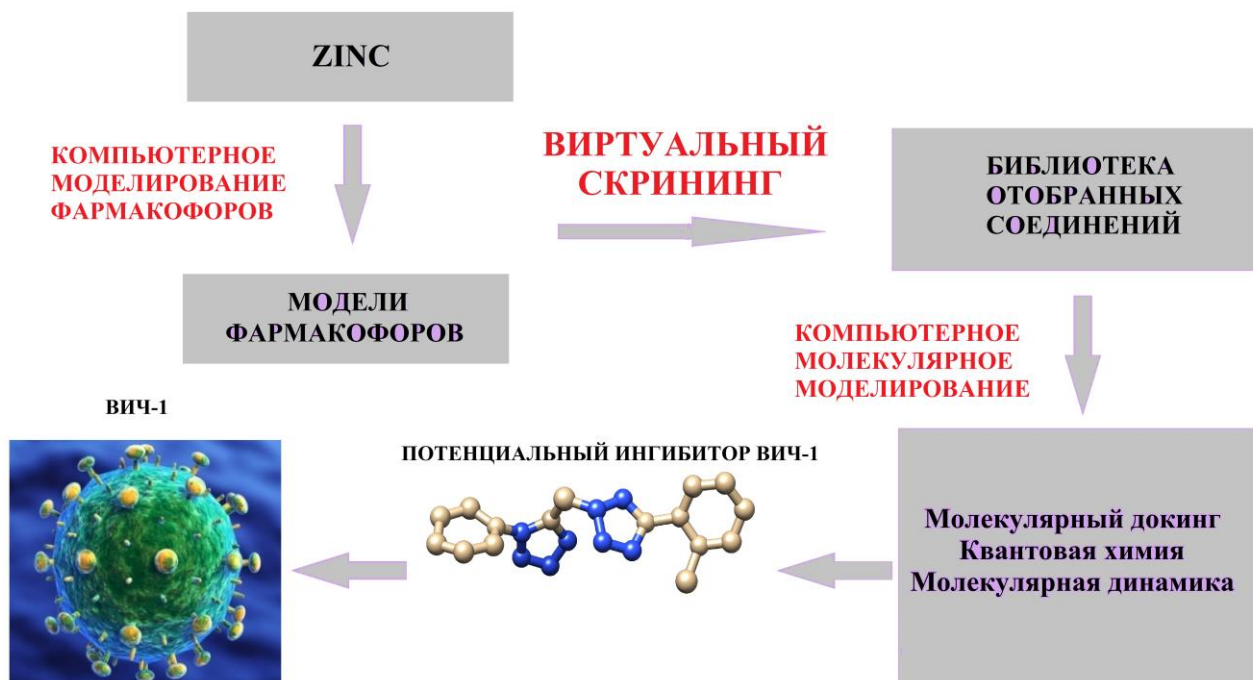
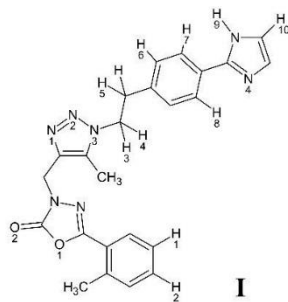


Рисунок 4. - Методика компьютерного моделирования для предсказания потенциальных ингибиторов ВИЧ-1, имитирующих нейтрализующее антитело N6

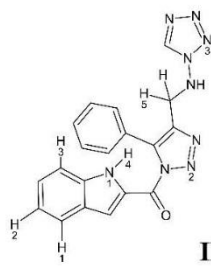
3) выполнен молекулярный докинг белка gp120 ВИЧ-1 с соединениями, согласующимися с заданными фармакофорными моделями моноклонального антитела N6;

4) проведена молекулярная динамика (МД) комплексов потенциальных ингибиторов антитела N6 с белком gp120 и осуществлен расчет свободной энергии их образования.

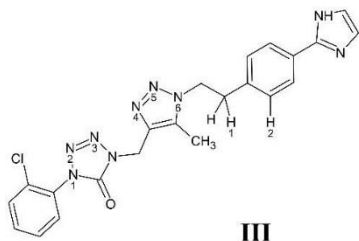
В результате проведенных исследований идентифицированы потенциальные пептидомиметики кросс-реактивного нейтрализующего анти-ВИЧ-1 антитела N6, способные имитировать фармакофорные свойства этого иммуноглобулина путем специфических и эффективных взаимодействий с CD4-связывающим сайтом белка gp120 оболочки вируса (рисунок 5). По данным молекулярной динамики, комплексы найденных потенциальных N6-миметиков с белком gp120 энергетически стабильны и характеризуются более низкими значениями свободной энергии связывания с белком по сравнению с известными ингибиторами ВИЧ-1 NBD-11021 и (+)-DMJ-II-121, использованными в расчетах в качестве контрольных соединений.



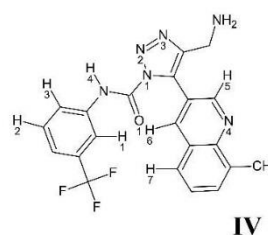
3 - [(1- [2- [4-(1H-имидазол-2-ил) фенил] этил] -5-метил-1H-1,2,3-триазол-4-ил) метил] -5-(2-метилфенил) -1,3,4-оксадиазол-2 (3H)-он



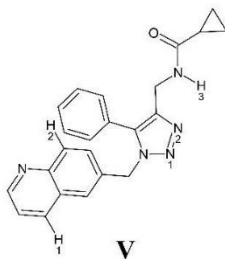
(1H-индоп-2-ил) (5-фенил-4 - [[(1H-тетразол-1-ил) амино] метил] -1H-1,2,3-триазол-1-ил)метанол



1 - [(1- [2- [4-(1H-имидазол-2-ил) фенил] этил] -5-метил-1H-1,2,3-триазол-4-ил) метил] -4-(2-хлорфенил)-1,4-дигидро-5H-тетразол-5-он



4- (аминометил)-5-(8-метилхинолин-3-ил) -N- [3- (трифторметил) фенил] -1H-1,2,3-триазол-1-карбоксамид



N-[(5-фенил-1-[(хинолин-6-ил)метил] -1H-1,2,3-триазол-4-ил)метил] циклопропан карбоксамид

Приведены систематические названия соединений

Рисунок 5. - Химические структуры потенциальных ингибиторов проникновения ВИЧ-1

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

1. Методами глубокого обучения разработан генеративный состязательный автоэнкодер для рационального дизайна потенциальных ингибиторов проникновения ВИЧ-1, способных блокировать участок белка gp120 оболочки вируса, критический для его связывания с клеточным рецептором CD4. Проведено тестирование нейронной сети на широком наборе соединений из молекулярной библиотеки ZINC15. Показано, что совместное использование нейронной сети с виртуальным скринингом баз данных

химических соединений формирует продуктивную платформу для идентификации базовых структур, перспективных для создания новых противовирусных препаратов, ингибирующих ранние стадии развития ВИЧ-инфекции [8, 9, 10, 18, 20].

2. С помощью комплексного подхода, включающего разработанный генеративный состязательный автоэнкодер, виртуальный скрининг и методы молекулярного моделирования, идентифицированы три низкомолекулярных химических соединения, проявляющие высокую аффинность связывания с белком gp120 ВИЧ-1. Согласно данным молекулярного докинга, машинного обучения, квантово-химических расчетов и молекулярной динамики, эти соединения характеризуются низкими значениями свободной энергии связывания в комплексах с белком gp120, близкие к значениям, рассчитанным с использованием идентичных вычислительных протоколов для ингибиторов ВИЧ-1 NBD-11021 и NBD-14010, представляющих новое поколение антагонистов клеточного рецептора CD4 [8, 9, 18, 20].

3. Разработан алгоритм для компьютерного конструирования лекарственных препаратов против ВИЧ-1, использующий методологию клик-химии, методы молекулярного моделирования и машинного обучения, с помощью которого осуществлен компьютерный дизайн новых потенциальных ингибиторов ВИЧ-1, специфически взаимодействующих с CD4-связывающим сайтом белка gp120. Методами молекулярного моделирования проведена оценка нейтрализующей активности сконструированных молекул, в результате которой обнаружены пять соединений-лидеров, перспективных для синтеза и биологических испытаний. Как и молекула CD4, эти соединения образуют многочисленные ван-дер-ваальсовы контакты и водородные связи с аминокислотами белка gp120, непосредственно участвующими в ключевом для связывания взаимодействии с остатком Phe-43_{CD4}. Структурные комплексы идентифицированных соединений с белком gp120 не претерпевают значительных структурных перестроек в течение МД расчетов и проявляют высокий процент заселенности межмолекулярных водородных связей, которые вносят существенный вклад в энергетическую стабильность комплексов [1, 2, 3, 5, 11, 12, 13, 14, 15, 21, 22, 23].

4. Разработана методика компьютерного моделирования для предсказания потенциальных ингибиторов ВИЧ-1, имитирующих нейтрализующее антитело N6 широкого спектра действия, основанная на использовании методов виртуального скрининга, молекулярного докинга, квантовой химии и молекулярной динамики. В результате ее применения идентифицированы девять пептидомиметиков моноклонального антитела N6, блокирующих CD4-связывающий участок ВИЧ-1 путем образования

водородных связей и ван-дер-ваальсовых контактов с аминокислотными остатками белка gp120, ответственными за связывание вируса с первичным рецептором CD4. При этом ключевую роль играют ван-дер-ваальсовы контакты идентифицированных соединений с остатками Phe⁴³-полости gp120, ответственными за взаимодействие ВИЧ-1 с Phe-43_{CD4}, а также водородная связь с остатком Asp-368_{gp120}, образование которой увеличивает аффинность связывания без активации нежелательного аллостерического эффекта [4, 6, 7, 16, 17, 19].

В результате компьютерного тестирования идентифицированных соединений на потенциальную анти-ВИЧ активность, проведенного методами молекулярного моделирования, квантовой химии и машинного обучения, показано, что все эти молекулы формируют перспективные базовые структуры для создания новых противовирусных препаратов – ингибиторов проникновения ВИЧ-1, блокирующих CD4-связывающий сайт белка gp120 оболочки вируса.

Рекомендации по практическому использованию

Полученные в диссертационном исследовании результаты могут быть использованы в работах по разработке более эффективных и безопасных анти-ВИЧ препаратов. Поэтому, дальнейшее развитие настоящей работы предусматривает тестирование обнаруженных молекул на противовирусную активность, идентификацию соединений-лидеров, оптимизацию структур базовых соединений, синтез наиболее перспективных из них с последующими биологическими испытаниями.

Применение разработанных алгоритмов компьютерного конструирования лекарственных препаратов может способствовать конструированию новых потенциальных противовирусных препаратов, что позволит сформировать наборы соединений различных классов – перспективных кандидатов для разработки эффективных препаратов широкого спектра действия.

СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ

Статьи в научных изданиях в соответствии с п. 18 Положения о присуждении ученых степеней и присвоении ученых званий в Республике Беларусь

1. In silico дизайн и оценка потенциальной активности новых ингибиторов ВИЧ-1 – миметиков первичного рецептора CD4 белка gp120 оболочки вируса / А. М. Андрианов, И. А. Кашин, Г. И. Николаев, А. В. Тузиков // Доклады НАН Беларуси. – 2017. – Т. 61, № 3. – С. 47-57.
2. Молекулярная динамика структурных комплексов потенциальных ингибиторов ВИЧ-1 с белком gp120 оболочки вируса / И. А. Кашин, Г. И. Николаев, А. В. Тузиков, А. М. Андрианов // Доклады Национальной академии наук Беларуси. – 2018. – Т. 62, № 5. – С. 576-584.
3. Разработка потенциальных ингибиторов ВИЧ-1 методами in silico клик-химии и молекулярного моделирования [Электронный ресурс] / А. М. Андрианов, Г. И. Николаев, И. А. Кашин, А. В. Тузиков // Математическая биология и биоинформатика. – 2018. – Т. 13, № 2. – С. 507-525. – Режим доступа: https://www.matbio.org/2018/Andrianov_13_507.pdf. – Дата доступа: 10.08.2021.
4. Виртуальный скрининг и идентификация потенциальных ингибиторов ВИЧ-1 на основе кросс-реактивного нейтрализующего антитела N6 / А. М. Андрианов, Г. И. Николаев, Ю. В. Корноушенко, Дж. Хуанг, Ш. Дзян, А. В. Тузиков // Доклады Национальной академии наук Беларуси. – 2019. – Т. 63, № 4. – С. 445-456.
5. In Silico Identification of Novel Aromatic Compounds as Potential HIV-1 Entry Inhibitors Mimicking Cellular Receptor CD4 / А. М. Andrianov, G. I. Nikolaev, Y. V. Kornoushenko, W. Xu, S. Jiang, A. V. Tuzikov // Viruses. – 2019. – Vol. 11, № 746. – P. 1-21.
6. In silico идентификация высокоаффинных лигандов белка gp120 ВИЧ-1 – потенциальных пептидомиметиков нейтрализующего антитела N6 [Электронный ресурс] / А. М. Андрианов, Г. И. Николаев, Ю. В. Корноушенко, Дж. Хуанг, Ш. Дзян, А. В. Тузиков // Математическая биология и биоинформатика. – 2019. – Т. 14, № 2. – С. 430-449. - Режим доступа: https://www.matbio.org/2019/Andrianov_14_430.pdf. – Дата доступа: 10.08.2021.
7. Идентификация функциональных миметиков нейтрализующего анти-ВИЧ антитела N6 методами виртуального скрининга и молекулярного моделирования / А. М. Андрианов, Г. И. Николаев, Ю. В. Корноушенко,

А. Д. Карпенко, Дж. Хуанг, Ш. Дзян, А. В. Тузиков // Доклады Национальной академии наук Беларуси. – 2019. – Т. 63, № 5. – С. 561-571.

8. Разработка генеративной состязательной нейронной сети для идентификации потенциальных ингибиторов ВИЧ-1 методами глубокого обучения / Г. И. Николаев, Н. А. Шульдов, А. И. Анищенко, А. В. Тузиков, А. М. Андрианов // Информатика. – 2020. – Т. 17, № 1. – С. 7-17.

9. Application of deep learning and molecular modeling to identify small drug-like compounds as potential HIV-1 entry inhibitors [Electronic esource] / A. M. Andrianov, G. I. Nikolaev, N. A. Shuldov, I. P. Bosko, A. I. Anischenko, A. V. Tuzikov // The Journal of Biomolecular Structure and Dynamics. – 2021. - Mode of access: <https://www.tandfonline.com/doi/full/10.1080/07391102.2021.1905559?scroll=top&needAccess=true>. – Date of access: 22.04.2021.

Статьи в других научных изданиях

10. Нейронная сеть для предсказания потенциальных ингибиторов проникновения ВИЧ-1 / А. М. Андрианов, Г. И. Николаев, Н. А. Шульдов, И. П. Босько, А. В. Тузиков // Наука и инновации. – 2021. – №5. – С. 28-34.

Статьи в сборниках материалов научных конференций

11. Компьютерный “de-novo” дизайн потенциальных ингибиторов проникновения ВИЧ-1 на основе методологии клик-химии / Г. И. Николаев, И. А. Кашин, А. В. Тузиков, А. М. Андрианов // Современные проблемы биохимии = Current problems in biochemistry: к Первому Бел. Биохимич. Конгр., Гродно, 5-6 июля 2016 г. / НАН Бел., Отделение медицинских наук; Респ. НИУП «И-т биохимии биологически активных соединений»; редкол.: Л. И. Надольник (гл. ред.) [и др.]. – Гродно: ЮрСаПринт, 2016. – Ч.2. - С. 218-224.

12. Компьютерный дизайн потенциальных анти-ВИЧ агентов, блокирующих CD4-связывающий участок белка gp120 оболочки вируса / Г. И. Николаев, И. А. Кашин, А. В. Тузиков, А. М. Андрианов // Молекулярные, мембранные и клеточные основы функционирования биосистем: материалы Международной научной конференции и Двенадцатого съезда Белорусского общественного объединения фотобиологов и биофизиков, Минск, 28-30 июня 2016 г. / МО Респ. Бел.; БГУ; НАН Беларуси; И-т биофизики и клеточной инженерии; БРФФИ; редкол.: И. Д. Волотовский [и др.]. – Минск: Изд. центр БГУ, 2016. – С. 154-156.

13. De Novo Design of Potential HIV-1 Entry Inhibitors Based on the Click Chemistry Concept: A Computational Study [Electronic resource] / A. M. Andrianov, G. I. Nikolaev, I. A. Kashyn, A. V. Tuzikov // MCCMB 2017: Proceedings of the International Moscow Conference on Computational Molecular Biology, Moscow, July 27–30 / In-t for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute) [et al.] – Moscow, 2017. – Mode of access: <https://mccmb.belozersky.msu.ru/2017/proceedings/MCCMB.html>. – Date of access: 22. 04. 2021.

14. Компьютерный de novo дизайн потенциальных ингибиторов ВИЧ-1 на основе концепции клик-химии [Электронный ресурс] / Г. И. Николаев, И. А. Кашин, А. В. Тузиков, А. М. Андрианов // Математические методы в технике и технологиях: сб. тр. XXX Междунар. науч. конф., г. Санкт-Петербург, 30 мая - 2 июня 2017 г. в 12 т. / Мин. обр. и науки РФ; Inst. of Hydrodynamics Academy of Science of the Czech Rep.; Ангарская гос. технич. ак.; Астрахан. гос. технич. у-т; Белгород. гос. технол. у-т; БГАТУ; БГУ; БГТУ; БНТУ; БГУИР; Донской гос. технич. у-т; Иванов. гос. хим.-технол. у-т; И-т выч. матем. РАН; И-т проблем упр. РАН; Казан. науч. исслед. технол. у-т [и др.]; под общ. ред. А. А. Большакова. – СПб.: Изд-во Политехн. у-та, 2017. – Т. 9. - С. 104-111. – Режим доступа: <http://mmtt30.sstu.ru/mmtt-30.nsf/pages/Publication>. – Дата доступа: 22. 04. 2021

15. Identification of Novel Potential HIV-1 Entry Inhibitors Targeting CD4-Binding Site of the Envelope gp120: Click Chemistry In Silico, Structure-based Docking and Molecular Dynamics Studies [Electronic resource] / A. M. Andrianov, G. I. Nikolaev, Y. V. Kornoushenko, A. D. Karpenko, A. V. Tuzikov // MCCMB'2019: Proceedings of 9th Moscow Conference on Computational Molecular Biology, Moscow, July 27–30, 2019 / In-t for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute). – Moscow, 2019. – Mode of access: <https://mccmb.belozersky.msu.ru/2019/proceedings/MCCMB.html>. – Date of access: 22. 04. 2021.

16. In Silico Discovery of Novel HIV-1 Entry Inhibitor Scaffolds Based on Broad and Potent Neutralizing Antibody N6 [Electronic resource] / A. M. Andrianov, G. I. Nikolaev, Y. V. Kornoushenko, J. Huang, S. Jiang, A. V. Tuzikov // MCCMB'2019: Proceedings of 9th Moscow Conference on Computational Molecular Biology, Moscow, July 27–30, 2019 / In-t for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute). – Moscow, 2019. – Mode of access: <https://mccmb.belozersky.msu.ru/2019/proceedings/MCCMB.html>. – Date of access: 22. 04. 2021.

17. In Silico-Guided Discovery of Potential HIV-1 Entry Inhibitors Mimicking bNAb N6: Virtual Screening, Docking, Molecular Dynamics, and Post Molecular Modeling Analysis [Electronic resource] / A. M. Andrianov, G. I. Nikolaev, Y. V. Kornoushenko, A. D. Karpenko, I. P. Bosko, A. V. Tuzikov // ISBRA 2020: Bioinformatics Research and Applications: 16th Int. Symp., Moscow, December 1-4, 2020 / Sechenov University et al. – Moscow, 2020. – Mode of access: <https://isbra.confreg.org/>. – Date of access: 22. 04. 2021.

18. Development of a Neural Network-Based Approach for Prediction of Potential HIV-1 Entry Inhibitors Using Deep Learning and Molecular Modeling Methods [Electronic resource] / G. I. Nikolaev, N. A. Shuldov, A. I. Anischenko, A. V. Tuzikov, A. M. Andrianov // ISBRA 2020: Bioinformatics Research and Applications: Moscow, December 1-4, 2020 / Sechenov University et al. – Moscow, 2020. – Mode of access: <https://isbra.confreg.org/>. – Date of access: 22. 04. 2021.

19. Компьютерное моделирование потенциальных ингибиторов ВИЧ-1 на основе нейтрализующего антитела N6 / А. М. Андрианов, Ю. В. Корноушенко, Г. И. Николаев, А. Д. Карпенко, А. В. Тузиков // Материалы II Междунар. науч.-практ. конф. «Компьютерные технологии и анализ данных» (СТДА'2020), Минск, 23-24 апреля 2020 г. / БГУ; редкол. В. В. Скакун (отв. ред.) [и др.]. – Мн., 2020. – С. 135-138.

20. Автоэнкодерная нейронная сеть для генерации потенциальных ингибиторов ВИЧ-1 методами глубокого обучения / Г. И. Николаев, Н. А. Шульдов, А. И. Анищенко, А. М. Андрианов, А. В. Тузиков // Материалы II Междунар. науч.практ. конф. «Компьютерные технологии и анализ данных» (СТДА'2020), Минск, 23-24 апреля 2020 г. / БГУ; редкол. В. В. Скакун (отв. ред.) [и др.]. – Мн., 2020. – С. 154-157.

Тезисы

21. Computational development of novel HIV 1 entry inhibitors targeting CD binding site of the viral envelope gp120 protein [Electronic resource] / G. I. Nikolaev, I. A. Kashyn, Y. V. Kornoushenko, A. V. Tuzikov, A. M. Andrianov // VI Int. Conf. “Chemistry, Structure and Function of Biomolecules”, Minsk, May 22-25, 2018 / National academy of sciences of Belarus, The department of chemistry and earth sciences, Institute of bioorganic chemistry, Belarusian republican foundation for fundamental research. - Minsk, 2018. – Mode of access: <https://catalog.belstu.by/catalog/books/d/IDX710a/view/10586>. – Date of access: 22. 04. 2021.

22. Development of novel small-molecule HIV-1 entry inhibitors: a computational study [Electronic resource] / G. I. Nikolaev, I. A. Kashyn, Y. V. Kornoushenko, A. V. Tuzikov, A. M. Andrianov // 4th International Turkic World Conference on Chemical Sciences and Technologies, Kiev, September 7-10, 2018. - Kiev, 2018. – Mode of access: <https://conferencealerts.com/show-event?id=197433>. – Date of access: 22. 04. 2021.

23. Click chemistry in silico and molecular simulation studies to identify novel HIV-1 entry inhibitor scaffolds targeting CD4-binding site of the envelope gp120 protein / A. M. Andrianov, G. I. Nikolaev, Y. V. Kornoushenko, A. V. Tuzikov // Journal of Biomolecular Structure and Dynamics. – 2019. – Vol. 37: Book of Abstracts. Albany 2019: The 20th Conversation, Albany, 27.05.2019. – P. 22-23.

A handwritten signature in black ink, appearing to be 'A. M. Andrianov', located in the lower right quadrant of the page.

РЕЗЮМЕ

Николаев Григорий Игоревич

Разработка алгоритмов, основанных на технологии глубокого обучения и молекулярном моделировании, для конструирования потенциальных ингибиторов проникновения ВИЧ-1

Ключевые слова: методы глубокого обучения, генеративно-состязательный автоэнкодер, белок gp120, ингибиторы проникновения ВИЧ-1, методы молекулярного моделирования, ВИЧ-1, клик-химия, антитело N6.

Цель работы: методами глубокого обучения разработать генеративный состязательный автоэнкодер для генерации молекулярных дескрипторов химических соединений, способных блокировать участок белка gp120 ВИЧ-1, критический для связывания вируса с клеточным рецептором CD4. С помощью комплексного подхода, включающего разработанную нейронную сеть, виртуальный скрининг, *in silico* клик-химию и методы молекулярного моделирования, идентифицировать молекулы, перспективные для создания эффективных анти-ВИЧ препаратов нового поколения.

Методы исследования и использованная аппаратура: методы молекулярного моделирования, анализ данных, глубокое обучение, технологии программирования и оптимизации.

Полученные результаты и их новизна: в диссертационной работе разработаны новые алгоритмы для конструирования потенциальных ингибиторов проникновения ВИЧ-1, основанные на технологии глубокого обучения и молекулярного моделирования.

Рекомендации по использованию и область применения: полученные в диссертационном исследовании результаты могут быть использованы в работах по разработке анти-ВИЧ препаратов нового поколения. Развитие настоящей работы предусматривает тестирование обнаруженных молекул на противовирусную активность, идентификацию соединений-лидеров, оптимизацию структур базовых соединений, синтез наиболее перспективных из них с последующими биологическими испытаниями.

РЭЗІЮМЭ

Нікалаеў Рыгор Ігаравіч

Распрацоўка алгарытмаў, заснаваных на тэхналогіі глыбокага навучання і малекулярным мадэляванні, для канструявання патэнцыйных інгібітараў пранікнення ВІЧ-1

Ключавыя словы: метады глыбокага навучання, генератыўных-спаборны аўтаэнкодар, бялок gp120, інгібітары пранікнення ВІЧ-1, метады малекулярнага мадэлявання, ВІЧ-1, клік-хімія, антыцела N6.

Мэта даследванняў: метадамі глыбокага навучання распрацаваць генератыўных спаборны аўтаэнкодар для генерацыі малекулярных дэскрыптараў хімічных злучэнняў, здольных блакаваць участак бялку gp120 ВІЧ-1, крытычны для звязвання віруса з клеткавым рэцэптарам CD4. З дапамогай комплекснага падыходу, які ўключае распрацаваную нейронную сетку, віртуальны скрынінг, in silico клік-хімію і метады малекулярнага мадэлявання, ідэнтыфікаваць малекулы, перспектыўныя для стварэння эфектыўных анты-ВІЧ прэпаратаў новага пакалення.

Метады даследавання і выкарыстаная апаратура: метады малекулярнага мадэлявання, аналіз даных, глыбокае навучанне, тэхналогіі праграмавання і аптымізацыі.

Атрыманыя вынікі і іх навізна: ў дысертацыйнай рабоце распрацаваны новыя алгарытмы для канструявання патэнцыйных інгібітараў пранікнення ВІЧ-1, заснаваныя на тэхналогіі глыбокага навучання і малекулярнага мадэлявання.

Рэкамендацыі па выкарыстанні і вобласць прымянення: атрыманыя ў дысертацыйным даследаванні вынікі могуць быць скарыстаны ў працах па распрацоўцы анты-ВІЧ прэпаратаў новага пакалення. Развіццё сапраўднай працы прадугледжвае тэсціраванне выяўленых малекул на супрацьвірусную актыўнасць, ідэнтыфікацыю злучэнняў-лідэраў, аптымізацыю структур базавых злучэнняў, сінтэз найбольш перспектыўных з іх з наступнымі біялагічнымі выпрабаваннямі.

SUMMARY

Nikalayeu Ryhor

Development of algorithms based on deep learning technology and molecular modeling for the construction of potential HIV-1 entry inhibitors

Keywords: deep learning methods, generative adversarial autoencoder, gp120 protein, HIV-1 entry inhibitors, molecular modeling techniques, HIV-1, click chemistry, antibody N6.

Objective: using deep learning to develop a generative adversarial autoencoder to generate molecular descriptors for chemical compounds capable of blocking a region of the HIV-1 gp120 protein that is critical for viral binding to the cellular CD4 receptor. Using an integrated approach, including a developed neural network, virtual screening, in silico click chemistry and molecular modeling methods, identify molecules that are promising for the creation of effective anti-HIV new generation drugs.

Research methods and used facilities: molecular modeling methods, data analysis, deep learning, programming and optimization technologies.

Obtained results and their novelty: new algorithms for constructing potential inhibitors of HIV-1 entry, based on deep learning technology and molecular modeling were developed in the dissertation.

Application area and usage recommendations: the results obtained in the dissertation research can be used in the development of a new generation of anti-HIV drugs. The development of this work involves testing the detected molecules for antiviral activity, identifying the leading compounds, optimizing the structures of basic compounds, and synthesizing the most promising of them with subsequent biological tests.



ЛИСТ ИЗМЕНЕНИЙ

в автореферате диссертации Николаева Г.И.
«Разработка алгоритмов, основанных на технологии глубокого обучения
и молекулярном моделировании, для конструирования
потенциальных ингибиторов проникновения ВИЧ-1»
на соискание ученой степени **кандидата** технических наук
по специальности 05.13.18 – Математическое моделирование,
численные методы и комплексы программ

На оборотной стороне титульного листа следует читать:

«Официальные оппоненты: Головко Владимир Адамович, доктор
технических наук, профессор, заведующий кафедрой интеллектуальных
информационных технологий УО «Брестский государственный технический
университет»

(вместо «Официальные оппоненты: Головко Владимир Адамович, доктор
технических наук, профессор, заведующий кафедрой интеллектуальных
информационных технологий УО «Брестский государственный университет
имени А.С. Пушкина»).

Соискатель



Г.И. Николаев

Ученый секретарь совета
по защите диссертаций Д 02.01.02 при БГУ
кандидат физ.-мат. наук доцент

Е.С. Чеб