

НЕЙРОННАЯ СЕТЬ ДЛЯ ПРЕДСКАЗАНИЯ ПОТЕНЦИАЛЬНЫХ ИНГИБИТОРОВ ПРОНИКНОВЕНИЯ ВИЧ-1

Александр Андрианов,
главный научный сотрудник Института
биоорганической химии НАН Беларуси, доктор
химических наук

Григорий Николаев,
научный сотрудник Объединенного института
проблем информатики НАН Беларуси

Никита Шульдов,
магистрант Белорусского государственного
университета

Иван Босько,
младший научный сотрудник Объединенного
института проблем информатики НАН Беларуси

Александр Тузиков,
генеральный директор Объединенного
института проблем информатики НАН Беларуси,
член-корреспондент

Современные методы компьютерного конструирования потенциальных лекарств значительно расширяют возможности фармацевтической индустрии, позволяя существенно сократить время и затраты, необходимые для их создания. Несмотря на то, что эффективность этих методов в разработке новых препаратов общепризнана, новые математические подходы в сочетании с доступностью мощных и дешевых вычислительных ресурсов способствуют их постоянному совершенствованию. Важное место при этом занимают методы машинного обучения (Machine Learning) и, в частности, глубокого обу-

чения (Deep Learning), которые обладают большим потенциалом для дальнейшего прогресса в данной области исследований [1, 2]. Компьютерное конструирование потенциальных лекарственных препаратов с их помощью – одна из наиболее важных и быстро развивающихся областей био- и хемоинформатики [1, 2]. Современные методы машинного обучения могут использоваться для моделирования количественной взаимосвязи «структура-активность» (QSAR; Quantitative Structure-Activity Relationship) или количественных соотношений «структура-свойство» (QSPR) и разработки интеллектуальных инструментов, способных точно предсказывать влияние химических модификаций соединения на его биологическую активность, фармакокинетические и токсикологические характеристики [1, 2]. Кроме этого, данные методы могут применяться для автоматического и прямого извлечения дескрипторов из химических структур [3], перепрофилирования лекарств [4], предсказания структуры белков [5], виртуального скрининга кандидатов в препараты и энергии их связывания с молекулярной мишенью [6]. Машинное обучение также можно использовать для создания новых соединений с заданными структурными и функциональными свойствами [7–9]. За последние 5 лет генеративные модели глубокого обучения стали инновационными методами разработки лекарств *de novo*. Благодаря огромному прогрессу в данной области созданы генеративные модели с различной архитектурой, типами данных и методами обучения, которые представляют перспективные решения для конструирования новых лекарств [7–9]. Недавние исследования с применением генеративных моделей глубокого обучения показали их способность генерировать молекулы, которые могут быть синтезированы, активны, метаболически стабильны и проявляют активность *in vivo*. В частности, ингибитор Янус-киназы 3, представляющий новый класс иммуномодулирующих агентов, разработан с помощью условно-состязательного автоэнкодера [10]. Кроме того, активные *in vivo* ингибиторы рецепторов 1 и 2 дискоидинового домена созданы с использованием генеративного обучения, а их фармакокинетический профиль подтвержден *in vivo* [11]. В связи с этим разработка и применение генеративных методов глубокого обучения для автоматизированного конструирования потенциальных лекарств имеет большое научное и практическое значение.

В последние годы появилось большое число работ по применению методов машинного обучения для

предсказания потенциальных ингибиторов ВИЧ-1 и резистентности вируса к анти-ВИЧ препаратам [12]. Однако все эти исследования сконцентрированы на вирусных ферментах – обратной транскриптазе и протеазе. Соединения, блокирующие их, не могут предотвращать проникновение вируса в клетку-мишень, что повышает внимание к ингибиторам ВИЧ-1, способным вмешиваться в ранние стадии жизненного цикла вируса путем блокирования процессов адсорбции и слияния мембран. Попадание вирусного генома в клетку хозяина – первый этап репликационного цикла ВИЧ-1 – представляет собой перспективную мишень для нескольких типов противовирусных препаратов, таких как ингибиторы связывания белка gp120 с первичным рецептором CD4, антагонисты корецепторов CCR5 и CXCR4 и ингибиторы слияния оболочки вируса с мембраной чувствительной клетки [13].

Была поставлена цель – разработать и применить генеративный состязательный автоэнкодер для идентификации потенциальных ингибиторов ВИЧ-1, способных блокировать гидрофобную полость Phe-43 белка gp120, критически важную для адсорбции вируса на поверхности CD+ T-клеток и последующего проникновения вирусного генома в клетку-мишень [13]. Выполнены исследования, включающие следующие этапы: создание архитектуры автоэнкодера; формирование молекулярной библиотеки потенциальных лигандов белка gp120 ВИЧ-1, предназначенной для обучения и тестирования нейронной сети; молекулярный докинг лигандов с белком gp120 и расчет свободной энергии связывания; генерация молекулярных дескрипторов химических соединений обучающего набора данных; обучение разработанного автоэнкодера и его тестирование на наборе данных, включающем более 21 млн молекул из библиотеки химических соединений ZINC15 [14].

РАЗРАБОТКА ГЕНЕРАТИВНОЙ СОСТЯЗАТЕЛЬНОЙ НЕЙРОННОЙ СЕТИ

Архитектура разработанного состязательного автоэнкодера состоит из двух нейросетей – автоэнкодера и дискриминатора, работающих во время обучения в соревновательном режиме. Автоэнкодер представляет собой 7-слойную нейронную сеть, имеющую входной, выходной и латентный слои, а также 4 полносвязных (рис. 1). На первый слой подаются молекулярные дескрипторы химических соединений, которые обрабатываются на двух полносвязных слоях (энкодер) и попадают на латент-

ный слой, где к полученному результату добавляется численная оценка энергии связывания с молекулярной мишенью. Далее молекулярные дескрипторы проходят 2 полносвязных слоя (декодер) и попадают на выход, который, как и вход, представляет собой вектор дескриптора. Работающая в таком режиме сеть формирует сжатое представление входного вектора на латентном слое с последующим его расширением на выходе декодера. Латентный слой состоит из трех нейронов, два из которых получают значения от энкодера, а третий – от энергии связывания с молекулярной мишенью. В генеративном режиме автоэнкодера на латентный слой, содержащий наиболее важную информацию об объекте, подаются случайные числа и требуемая оценка величины свободной энергии связывания, затем они проходят через декодер, генерирующий дескрипторы молекул с требуемыми свойствами. Для генерации таких молекул важно, чтобы данные, поступающие на латентный слой после прохождения энкодера, имели нормальное распределение, на которое обучен дискриминатор. Для обеспечения этого условия в процессе состязательного обучения энкодера и дискриминатора добивались того, чтобы энкодер был способен кодировать на латентный слой данные с нормальным распределением, а дискриминатор – отличать стандартное нормальное распределение (сгенерированные данные) от поступающего на латентный слой.

Разработанный автоэнкодер основан на модели нейронной сети, предназначенной для генерации химических соединений с противоопухолевой активностью [8], но имеет ряд особенностей (рис. 1).

1. *Вместо нейрона, ответственного за процентное ингибирование роста клеток [8], к латентному слою добавляется нейрон, отвечающий за значения свободной энергии связывания. Этот нейрон не взаимодействует с энкодером и подается только на вход декодера вместе с данными, полученными с помощью энкодера. На этапе обучения значения этого нейрона присваиваются в соответствии с оценкой величины свободной энергии связывания для тестируемых соединений. На генеративной стадии его значения имеют смысл порога свободной энергии связывания. Латентный слой содержит 3 нейрона.*
2. *Энкодер состоит из двух полносвязных слоев L1 и L2 с 32 и 16 нейронами соответственно. Декодер включает два слоя L'1 и L'2, содержащие 16 и 32 нейрона соответственно.*

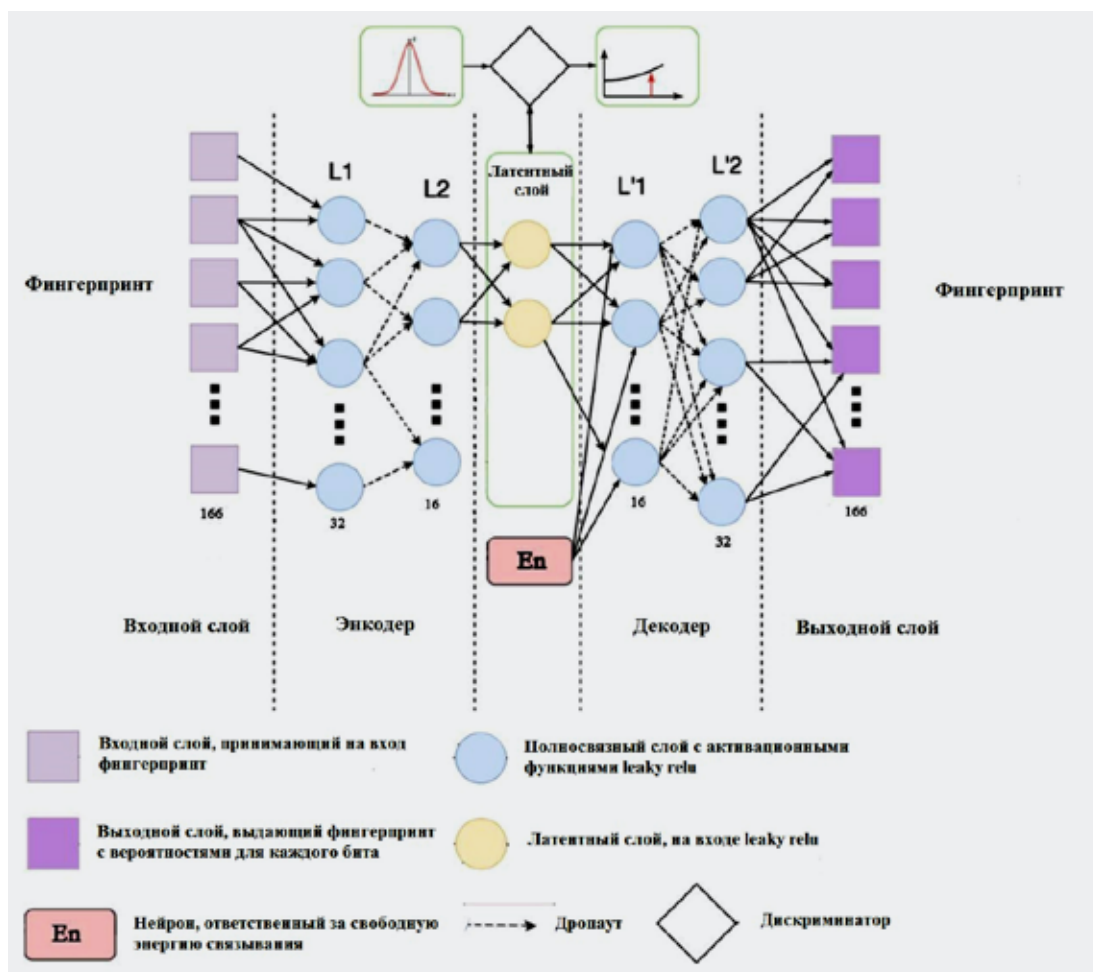


Рис. 1. Архитектура нейронной сети для генерации потенциальных ингибиторов ВИЧ-1, блокирующих CD4-связывающий сайт белка gp120 оболочки вируса

3. Дискриминатор состоит из четырех слоев, включающих 2, 16, 3 и 1 нейрон соответственно.
4. На промежуточных слоях автоэнкодера используется активационная функция leaky relu [15].
5. На всех слоях дискриминатора имеются сигмоидные активационные функции [16].
6. Для обеспечения дополнительного уровня защиты от переобучения между двумя полносвязными слоями энкодера и декодера добавлен слой дропаута, наличие которого позволяет нейронной сети полностью задействовать все свои параметры (веса) за счет отключения случайных из них во время обучения.

Основное назначение дропаут-слоя заключается в том, чтобы вместо одной сети обучить ансамбль из нескольких, а затем усреднить полученные результаты. Для этого использовали трехступенчатый итерационный процесс: обучение дискриминатора различать заданное нормальное распределение от закодированного, полученного энкодером на латентном слое; совместное обучение энкодера

и декодера как автоэнкодера; обучение энкодера сжимать данные таким образом, чтобы они представляли нормальное распределение.

СОЗДАНИЕ МОЛЕКУЛЯРНОЙ БИБЛИОТЕКИ ДЛЯ ОБУЧЕНИЯ АВТОЭНКОДЕРА

Формирование обучающего набора данных выполнено в рамках подхода, использующего методологию клик-химии [17] для генерации наиболее вероятных структур-кандидатов биологически активных соединений. Для конструирования потенциальных лигандов с помощью программы DataWarrior [18] были созданы две молекулярные библиотеки. Первая (библиотека 1) включала отобранные из кластера Drug-Like базы данных ZINC15 [14] небольшие молекулы (масса < 250 Да) с азидной или алкиновой группами, содержащие ароматические фрагменты – элементы структуры, которые, согласно данным об известных ингибиторах про-

никновения ВИЧ-1 [19, 20], играют ключевую роль для специфического взаимодействия с Phe-43-полостью CD4-связывающего сайта белка gp120. В библиотеку 2 были отобраны все низкомолекулярные соединения с молекулярной массой < 250 Да, имеющие азидную или алкиновую группы. В результате работы программы DataWarrior библиотека 1 включала 1388 соединений, а библиотека 2 – 3769. На следующем этапе они были использованы в качестве исходных реагентов для имитации реакции азид-алкинового циклоприсоединения с помощью программы AutoClickChem [21], рассматривающей все возможные комбинации молекул из библиотек 1 и 2. Это позволило получить набор из 1 655 301 гибридной молекулы, из которого 120 тыс. соединений, удовлетворяющих «правилу пяти» Липинского [22], были использованы для формирования обучающего набора данных. Оценку энергии их связывания с белком gp120 проводили методом молекулярного докинга – процедуры виртуального скрининга, позволяющей предсказать наиболее вероятные ориентации лиганда в активном центре белка и рассчитать свободную энергию образования комплексов лиганд/белок.

Генерацию молекулярных дескрипторов MACCS [23] в обучающем наборе данных осуществляли с помощью программного пакета RDKit [24] с открытым исходным кодом. Молекулярный докинг лигандов с белком gp120 выполняли в программе QuickVina 2 [25] с учетом конформационной подвижности лиганда. Трехмерную структуру белка gp120 выделяли из комплекса этого гликопротеина

с рецептором CD4 и антителом 17b (код 1GC1 в Банке данных белков [26]). Атомы водородного рода добавляли к структуре белка gp120 с использованием программного пакета AutoDockTools. Ячейка для докинга представляла фрагмент белка gp120 с координатами $x \in (24 \text{ \AA}; 34 \text{ \AA})$, $y \in (-15 \text{ \AA}; -5 \text{ \AA})$, $z \in (78 \text{ \AA}; 88 \text{ \AA})$, включающий Phe-43-полость гликопротеина; то есть ее объем составлял $10 \times 10 \times 10 = 1000 \text{ \AA}^3$. Параметр, характеризующий полноту поиска (охват конформационного пространства), был задан равным 50.

НЕЙРОННАЯ СЕТЬ ДЛЯ ПРЕДСКАЗАНИЯ ПОТЕНЦИАЛЬНЫХ ЛЕКАРСТВЕННЫХ ПРЕПАРАТОВ ПРОТИВ ВИЧ-1

Применение разработанного автоэнкодера для прогнозирования ингибиторов ВИЧ-1 предполагает работу этой нейронной сети в сочетании с виртуальным скринингом и традиционными методами молекулярного моделирования. Общая схема алгоритма прогнозирования анти-ВИЧ-1 препаратов показана на рис. 2.

Автоэнкодер генерирует набор структурных ключей MACCS при заданном пороговом значении энергии связывания для скрининга библиотеки молекулярных дескрипторов из базы данных ZINC15. Соединения с молекулярными дескрипторами, подобными тем, что генерируются автоэнкодером, связываются с белком gp120 оболочке ВИЧ-1 с использованием

методов молекулярного докинга. Затем комплексы лиганд/gp120 оптимизируются полуэмпирическим квантово-химическим методом PM7 для повышения точности предсказания типов взаимодействий, полученных на основе алгоритмов, базирующихся на классических силовых полях. На заключительном этапе выполняется моделирование молекулярной динамики комплексов идентифицированных соединений с белком gp120 с последующим анализом MD-траекторий, включая предсказание межмолекулярных взаимодействий и аффинности связывания для динамических структур лиганд/gp120.

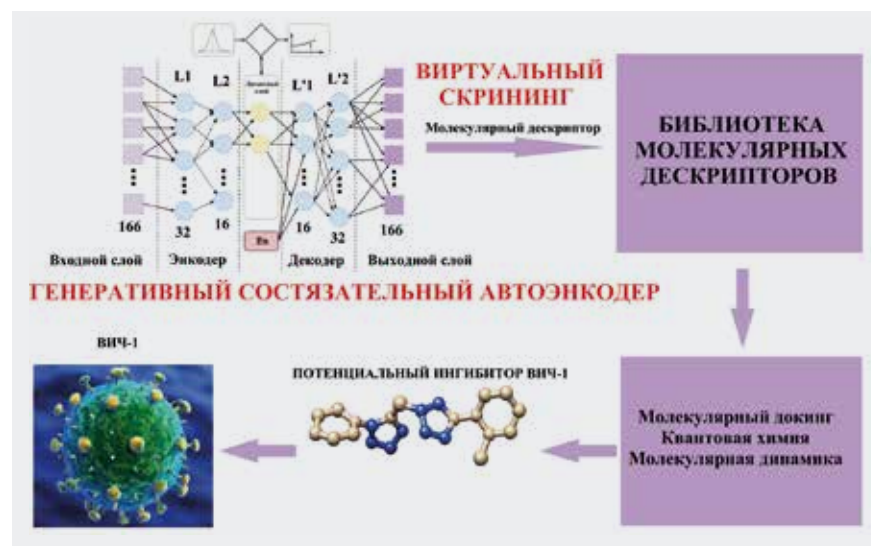


Рис. 2. Схема работы разработанной нейронной сети для предсказания потенциальных ингибиторов белка gp120 ВИЧ-1

ОЦЕНКА РЕЗУЛЬТАТОВ ОБУЧЕНИЯ И РАБОТЫ АВТОЭНКОДЕРА

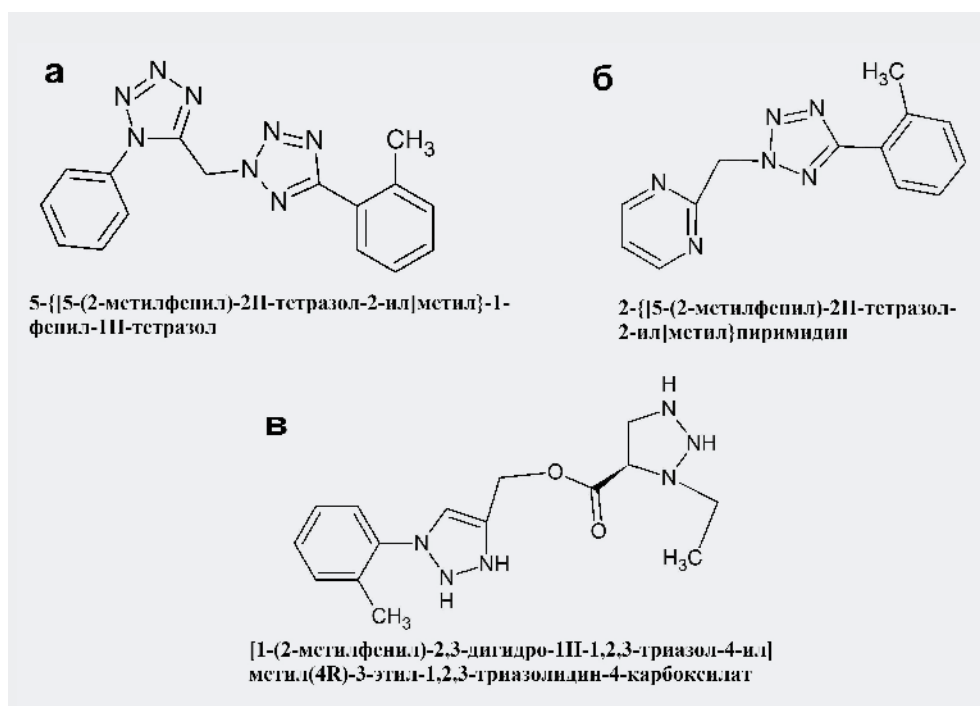
Для тестирования автоэнкодера с помощью программного пакета RDKit [24] была создана библиотека структурных ключей MACCS [23] для 21 325 567 соединений из базы данных ZINC15 [14] и рассчитаны 5 молекулярных дескрипторов для сгенерированных молекул при пороговых значениях энергии связывания, равных 5 ккал/моль и 8 ккал/моль. В результате виртуального скрининга для каждой из этих молекул были найдены лиганды с подобными дескрипторами. При этом в качестве меры подобия использовали расстояние Хэмминга, определяемое в теории кодирования как число пар несовпадающих компонент сравниваемых векторов [27], и коэффициент Танимото, который вычисляли по формуле [28]:

$$T(a, b) = \frac{N_c}{N_a + N_b - N_c},$$

где T – коэффициент Танимото, принимающий значения от 0 до 1; N_a – количество элементов в первом векторе; N_b – количество элементов во втором векторе; N_c – количество одинаковых элементов в двух векторах. В процессе скрининга библиотеки молекулярных дескрипторов из базы данных ZINC15 отбирали соединения, для которых коэффициент Танимото удовлетворял условию $T > 0,85$ [28].

Идентифицированные в базе данных ZINC15 соединения подвергали процедуре докинга с белком gp120 и рассчитывали энергию связывания с Phe-43-полостью CD4-связывающего сайта оболочки ВИЧ-1. Молекулярный докинг проводили с помощью вычислительного протокола программы QuickVina 2 [25], идентичного тому, что был применен при создании обучающего набора данных. 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. При этом найденные в базе данных ZINC15 [14] соединения с кодами ZINC000026430653, ZINC000191389930 и ZINC000293423658 (рис. 3) характеризуются величинами энергии связывания с белком gp120, сопоставимыми со значением $-9,5 \pm 0,1$ ккал/моль, измеренным для комплекса CD4-gp120 методом изотермической титрационной калориметрии [29]. Эти величины близки к значениям оценочной функции QuickVina 2 [25], полученным нами ранее [30–33] для высокоаффинных лигандов белка gp120, сконструированных методами молекулярного моделирования, а также для ингибиторов ВИЧ-1 NBD-11021 и NBD-14010, представляющих новое поколение полных функциональных антагонистов клеточного рецептора CD4 [34]. Кроме того, приведенные выводы согла-

Рис. 3. Химические структуры ZINC000026430653 (а), ZINC000191389930 (б) и ZINC000293423658 (в). Для этих молекул четыре параметра Липинского [22] – молекулярная масса, липофильность (LogP), число доноров и акцепторов водородной связи – равны соответственно: (а) 318,344 Да, 1,673, 0 и 8; (б) 252,281 Да, 1,487, 0 и 6; (в) 312,333 Да, 1,544, 0 и 8



суются со значениями свободной энергии связывания анализируемых молекул с белком gp120, рассчитанными на основе предсказанных значений констант диссоциации (K_d) комплексов по формуле $\Delta G = R \cdot T \cdot \ln(K_d)$, где ΔG – энергия связывания, R – универсальная газовая постоянная, T – абсолютная температура, равная 310 К [35]. Полученные данные показывают, что идентифицированные соединения могут проявлять сильное связывание с гидрофобной полостью Phe-43 белка gp120 ВИЧ-1 в соответствии с низкими значениями энергии, сравнимыми с предсказанными для ингибиторов ВИЧ NBD-11021 и NBD-14010.

Исследование комплексов лиганд/gp120 с помощью полуэмпирического метода квантовой химии PM7 свидетельствует о наличии многочисленных ван-дер-ваальсовых контактов между идентифицированными соединениями и Phe-43-полостью белка gp120, которая играет ключевую роль в связывании ВИЧ-1 с клеточным рецептором CD4 (рис. 4) [26]. По аналогии с молекулой CD4 [26] одна из ароматических систем этих соединений погружена в гидрофобный карман Phe-43-полости белка gp120, имитируя основные взаимодействия бензольного кольца Phe-43_{CD4} с функционально важными остатками гликопротеина. Этими остатками gp120 являются Val-255, Thr-257, Glu-370, Ile-371, Asp-368, Ser-375, Phe-382, Trp-427 и Met-475 (рис. 4). Согласно данным о структуре белка gp120 в комплексе с молекулой CD4 в кристалле [26], все эти аминокислоты участвуют в прямых межатомных контактах, а консервативные остатки Asp-368, Glu-370 и Trp-427 доминируют в интерфейсе gp120/CD4.

Наряду с ван-дер-ваальсовыми контактами анализ комплексов лиганд/gp120 указывает на высокую вероятность реализации особого типа слабой водородной связи C–H...O между группами CH ароматических колец соединений ZINC000026430653, ZINC000191389930, ZINC000293423658 и карбоксильной группой остатка Asp-368, расположенного у входа в Phe-43-полость белка gp120. Об этом свидетельствуют значения расстояния d между атомами С

и угла между атомами С, Н и О (рис. 4), которые подтверждают предположение о том, что указанные атомы могут быть вовлечены в диполь-диполь взаимодействие C–H...O [37]. Эти результаты представляют значительный интерес, так как водородная связь с остатком Asp-368 белка gp120 очень важна для потенциальных миметиков CD4, чтобы проявлять свойства низкомолекулярных антагонистов рецептора CD4 [20].

Таким образом, данные молекулярного докинга в сочетании с квантово-химическими расчетами показывают, что идентифицированные соединения могут проявлять высокую аффинность связывания с гидрофобной полостью Phe-43 белка gp120 путем имитации ключевых взаимодействий CD4 с этим гликопротеином оболочки ВИЧ-1.

Молекулярная динамика подтверждает данные о высоком сродстве связывания идентифицированных соединений с gp120, полученные в результате

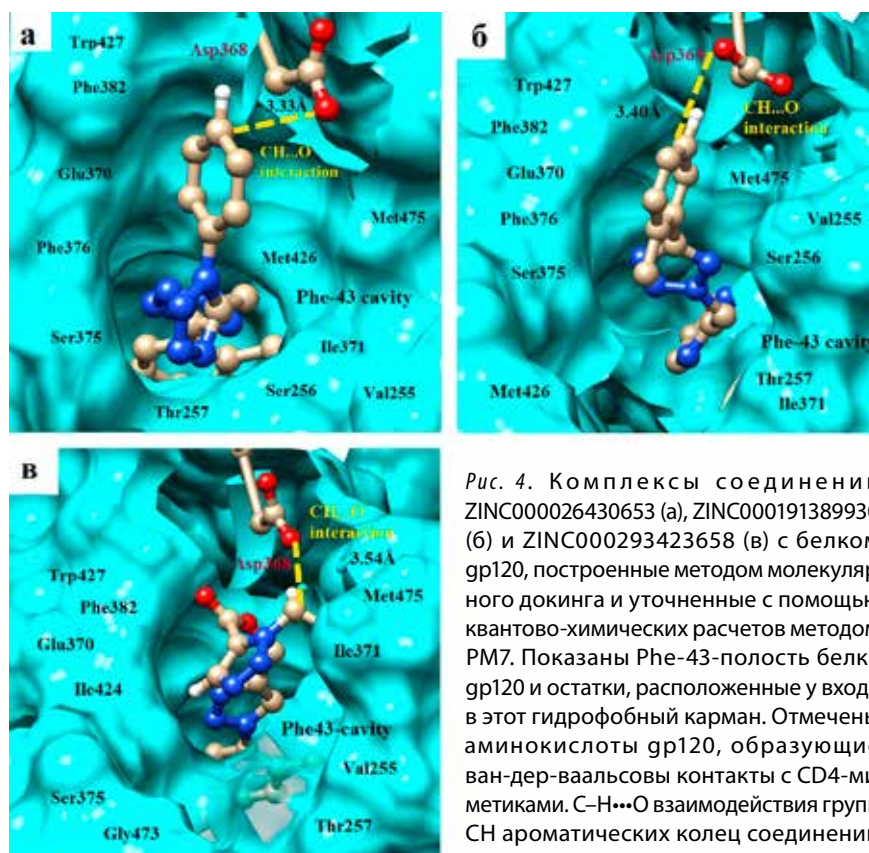


Рис. 4. Комплексы соединений ZINC000026430653 (а), ZINC000191389930 (б) и ZINC000293423658 (в) с белком gp120, построенные методом молекулярного докинга и уточненные с помощью квантово-химических расчетов методом PM7. Показаны Phe-43-полость белка gp120 и остатки, расположенные у входа в этот гидрофобный карман. Отмечены аминокислоты gp120, образующие ван-дер-ваальсовы контакты с CD4-миметиками. C–H...O взаимодействия групп CH ароматических колец соединений ZINC000026430653, ZINC000191389930 и ZINC000293423658 с карбоксильной группой Asp-368 белка gp120 отмечены пунктирными линиями

Лиганд	$\langle \Delta H \rangle \pm (\Delta H)_{\text{STD}}$ ккал/моль	$\langle T\Delta S \rangle \pm (T\Delta S)_{\text{STD}}$ ккал/моль	$\langle \Delta G \rangle \pm (\Delta G)_{\text{STD}}$ ккал/моль
ZINC000026430653	-40,5 ± 3,1	-20,3 ± 5,4	-20,2 ± 6,1
ZINC000191389930	-31,3 ± 2,9	-20,3 ± 6,1	-11,0 ± 6,9
ZINC000293423658	-42,6 ± 3,9	-20,9 ± 6,2	-21,7 ± 7,3
NBD-11021	-42,2 ± 3,7	-22,9 ± 6,9	-19,3 ± 7,9
NBD-14010	-39,7 ± 2,9	-21,8 ± 6,1	-18,0 ± 6,4

Таблица. Средние значения свободной энергии связывания $\langle \Delta G \rangle$ для предсказанных соединений, NBD-11021 и NBD-14010 в комплексах с белком gp120 и их стандартные отклонения $(\Delta G)_{\text{STD}}$ ¹

Примечание: 1 $\langle \Delta H \rangle$ и $\langle T\Delta S \rangle$ – средние значения энтальпийной и энтропийной составляющих свободной энергии соответственно; $(\Delta H)_{\text{STD}}$ и $(T\Delta S)_{\text{STD}}$ – стандартные отклонения, соответствующие этим значениям

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 (таблица). Низкие средние значения свободной энергии связывания, рассчитанные для динамических моделей идентифицированных соединений, связанных с gp120, подтверждают сделанное выше предположение, согласно которому все предсказанные кандидаты-миметики CD4 демонстрируют эффективное взаимодействие с CD4-связывающим сайтом белка gp120 ВИЧ-1. Полученные данные показывают, что для соединений ZINC000026430653 и ZINC000293423658 средние значения свободной энергии связывания сопоставимы с величинами, предсказанными для ингибиторов ВИЧ-1 NBD-11021 и NBD-14010 (таблица), а соответствующее среднее значение для соединения ZINC000191389930 близко к экспериментальной величине $9,5 \pm 0,1$ ккал/моль, полученной для комплекса белка gp120 с CD4 [29].

Тестирование разработанного генеративного состязательного автоэнкодера с использованием широкого набора химических соединений из базы данных ZINC15 показывает, что эта нейронная сеть в комбинации с виртуальным скринингом и традиционными методами молекулярного моделирования формирует продуктивную основу для обнаружения потенциальных анти-ВИЧ агентов, способных блокировать CD4-связывающий сайт белка gp120 оболочки вируса. Согласно данным молекулярного моделирования, соединения ZINC000026430653, ZINC000191389930 и ZINC000293423658 (рис. 3) демонстрируют высокую аффинность связывания с белком gp120, которое обеспечивается взаимодействиями, доминиру-

ющими в интерфейсе gp120/CD4 (рис. 4). Кроме того, они полностью удовлетворяют критериям, налагаемым на потенциальное лекарство «правилом пяти» Липинского [22], что подтверждается их физико-химическими параметрами, приведенными в подписи к рис. 3. Поэтому обнаруженные в базе данных ZINC15 молекулы могут быть использованы в качестве перспективных базовых структур для разработки новых эффективных противовирусных препаратов, ингибирующих ранние стадии развития ВИЧ-инфекции.

При анализе полученных результатов необходимо иметь в виду, что в данном исследовании были использованы только пять структурных ключей MACCS, сгенерированных нейронной сетью для каждого из двух пороговых значений энергии для проверки работы автоэнкодера, что привело к идентификации трех соединений, проявляющих сильное связывание с гидрофобной полостью Phe-43 белка gp120. Однако эта нейронная сеть может генерировать множество молекулярных дескрипторов при заданном максимально допустимом значении энергии связывания. Очевидно, что более широкий набор дескрипторов должен обеспечить обнаружение новых соединений – перспективных кандидатов в CD4-миметики – для создания новых эффективных и безопасных противовирусных средств широкого спектра действия. В связи с этим дальнейшее развитие этой работы предполагает скрининг различных химических баз данных с применением гораздо большего числа структурных ключей MACCS, полученных с помощью предложенной нейронной сети. ■

Работа поддержана Белорусским республиканским фондом фундаментальных исследований (проекты X20MC-006 и Ф21КОВИД-002).

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Applications of machine learning in drug discovery and development / J. Vamathevan [et al.] // Nat. Rev. Drug Discov. 2019. Vol. 18(6). P. 463–477.
2. Advances and perspectives in applying deep learning for drug design and discovery / C.F. Lipinski [et al.] // Front. Robot. AI. 2019. Vol. 6. P. 108.
3. Jaeger, S. Mol2vec: unsupervised machine learning approach with chemical intuition / S. Jaeger, S. Fulle, S. Turk // J. Chem. Inf. Model. 2018. Vol. 58(1). P. 27–35.
4. A Machine learning-based method to improve docking scoring functions and its application to drug repurposing / S.L. Kinnings [et al.] // J. Chem. Inf. Model. 2011. Vol. 51(2). P. 408–419.
5. Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13) / A.W. Senior [et al.] // Proteins. 2019. Vol. 87(12). P. 1141–1148.

Полный список использованных источников размещен

 SEE http://innosfera.by/2021/05/machine_learning