

НАУЧНАЯ СТАТЬЯ

УДК 544.165:615.917

QSAR-АНАЛИЗ ИНГИБИТОРОВ HDAC6

Олег Викторович Тиньков¹, Вениамин Юрьевич Григорьев², Людмила Дмитриевна Григорьева³

¹ Приднестровский государственный университет имени Т.Г. Шевченко, кафедра фармакологии и фармацевтической химии медицинского факультета, Тирасполь, Приднестровье, Молдова

² Институт физиологически активных веществ Российской академии наук, Черноголовка, Россия

³ Факультет фундаментальной физико-химической инженерии МГУ имени М.В. Ломоносова, Москва, Россия

Автор, ответственный за переписку: Олег Викторович Тиньков, oleg.tinkov.chem@mail.ru

Аннотация. Ингибиторы гистондеацетилаз – важнейший класс лекарственных средств для лечения онкологических и других заболеваний из-за их влияния на рост, дифференцировку и апоптоз клеток. Среди известных восемнадцати гистондеацетилаз важнейшей является гистондеацетилаза 6 (Histone deacetylase 6, HDAC6), участвующая в онкогенезе, выживаемости клеток, а также в метастазировании раковых клеток. С использованием молекулярных 2D-дескрипторов RDKit, симплексных дескрипторов, а также методов «Случайного леса», «Градиентного бустинга», «Опорных векторов» предложен ряд адекватных классификационных моделей количественной связи «структура – активность» (Quantitative Structure – Activity Relationship, QSAR). Для моделей, построенных с помощью симплексных дескрипторов, была проведена структурная интерпретация, позволившая выделить молекулярные фрагменты, повышающие и понижающие активность ингибиторов HDAC6. Результаты структурной интерпретации были использованы для рационального молекулярного дизайна потенциальных ингибиторов HDAC6, для которых также была проведена оценка ADMET (Absorption, Distribution, Metabolism, Excretion, Toxicity – всасывания, распределения, метаболизма, выведения, токсичности) свойств. Модели, построенные с помощью 2D-дескрипторов RDKit, находятся в свободном доступе на платформе github по адресу <https://github.com/ovttiras/HDAC6-inhibitors>.

Ключевые слова: ингибиторы гистондеацетилазы 6, QSAR, молекулярные дескрипторы, машинное обучение, структурная интерпретация

Финансирование. Часть работы выполнена в рамках Государственного задания ИФАВ РАН 2022 года (тема № FFSN-2021-0004).

Для цитирования: Тиньков О.В., Григорьев В.Ю., Григорьева Л.Д. QSAR-анализ ингибиторов HDAC6 // Вестн. Моск. ун-та. Сер. 2. Химия. Т. 64. № 1. С. 35–48.

ORIGINAL ARTICLE

QSAR ANALYSIS OF HDAC6 INHIBITORS

Oleg V. Tinkov¹, Veniamin Yu. Grigorev², Lyudmila D. Grigoreva³¹Department of Pharmacology and Pharmaceutical Chemistry, Medical Faculty, Shevchenko Transnistria State University, Tiraspol, Moldova²Institute of Physiologically Active Compounds of the Russian Academy of Sciences, Chernogolovka, Russia³Department of Fundamental Physical-Chemical Engineering, Lomonosov Moscow State University**Corresponding author:** Oleg V. Tinkov, oleg.tinkov.chem@mail.ru

Abstract. Histone deacetylase inhibitors are the most important class of drugs for the treatment of oncologies and other diseases due to their effect on cell growth, differentiation and apoptosis. Among the known eighteen histone deacetylases, Histone deacetylase 6 (HDAC6), which is involved in oncogenesis, cell survival, and cancer cell metastasis, is of high importance. Using 2D molecular descriptors RDKit, simplex descriptors, as well as methods of Random Forest (RF), Gradient Boosting (GBM), Support vectors (SVM), a number of adequate classification models of Quantitative Structure–Activity Relationship (QSAR) are proposed. For the models constructed using simplex descriptors, a structural interpretation was carried out, which made it possible to describe molecular fragments that increase and decrease the activity of HDAC6 inhibitors. The results of the structural interpretation were used for the rational molecular design of potential HDAC6 inhibitors, for which ADMET properties were also evaluated. Models built using 2D RDKit descriptors are freely available on the github platform (<https://github.com/ovttiras/HDAC6-inhibitors>).

Keywords: histone deacetylase 6 inhibitors, QSAR, molecular descriptors, machine learning, structural interpretation

Financial Support. Part of the work was carried out within the framework of the State Task of the IPAC RAS 2022 (topic No. FFSN-2021-0004).

For citation: Tinkov O.V., Grigorev V.Y., Grigoreva L.D. QSAR Analysis of HDAC6 Inhibitors // Vestn. Mosk. un-ta. Ser. 2. Chemistry. T. 64. N 1. P. 35–48.

Введение

На сегодняшний день химиотерапия онкологических больных продолжает оставаться важнейшим методом лечения. Одним из ограничений к повышению эффективности химиотерапии опухолей является лекарственная устойчивость. Утрата программы клеточной гибели дает возможность опухолевой клетке сохранять жизнеспособность в условиях высокой концентрации противоопухолевых лекарств и формировать опухоль, резистентную к химиотерапии. Достигнуть реактивации апоптоза в подобных опухолевых тканях становится практически невозможным. Таким образом, актуально создание новых селективных противоопухолевых препаратов.

Одной из важнейших мишеней противоопухолевых препаратов являются гистондеацетила-

зы (HDAC, EC 3.5.1.98). Данные ферменты катализируют удаление ацетильной группы ε-N-ацетил-лизина гистонов. Модифицируя гистоны и изменяя конформацию хроматина, гистондеацетилазы играют важную роль в регуляции экспрессии генов. Среди известных восемнадцати гистондеацетилаз наиболее значима гистондеацетилаза 6 (Histone deacetylase 6, HDAC6), участвующая в онкогенезе, выживаемости клеток, а также в метастазировании раковых клеток [1]. Ингибиторы HDAC в настоящее время успешно применяются для лечения онкологических заболеваний разной этиологии. Известно, что эффективными ингибиторами гистондеацетилаз (HDAC) являются производные гидроксамовой кислоты, например вориностат (Vorinostat) (SAHA), панобиностат (Panobinostat) (LBH589)

и белиноSTAT (Belinostat) (PXD101). Они одобрены Управлением по санитарному надзору за качеством пищевых продуктов и медикаментов США (Food and Drug Administration, FDA) для лечения Т-клеточной лимфомы кожи (CTCL) и множественной миеломы [2].

Классические ингибиторы гистондеацетилаз функционируют путем связывания иона цинка в активном центре фермента, инактивируя систему смены зарядов. Эти ингибиторы, как правило, имеют трехкомпонентную структуру (рис. 1), состоящую из цинк-связывающего участка, линкера, способного занимать канал фермента, и функциональной группы (так называемой «крышки», CAP группы, Capping group), взаимодействующей с аминокислотными остатками у входа в активный центр HDAC.

В связи с необходимостью интенсификации поиска новых физиологически активных соединений большое внимание привлекают методы количественной связи «Структура – Активность / Свойство» (QSAR/QSPR), которые оказывают существенную помощь в уменьшении как финансовых, так и временных затрат при создании лекарственных средств [3]. С учетом важности данной проблемы рядом исследователей [4–7] разработаны удовлетворительные QSAR-модели, описывающие влияние структуры химических соединений на их способность ингибировать HDAC6. Существующие QSAR-модели ингибиторов HDAC6 позволяют рассматривать молекулярную структуру с разных точек зрения, используя различные статистические подходы. Разнообразная информация, полученная с их помощью, представляет интерес для исследователей. В большинстве вышеуказанных исследований использованы методы 3D-QSAR (CoMFA, CoMSIA, GRIND

descriptors) в сочетании с молекулярным докингом (molecular docking). Однако в ряде случаев существующие QSAR-модели ингибиторов HDAC6 имеют ограниченное применение. Например, в методе CoMFA [8] существуют определенные трудности при анализе разнородных по структуре соединений из-за неоднозначности трехмерного выравнивания структур. Применяемые в большинстве существующих QSAR-моделей регрессионные методы анализа данных (MLR, PLS) [9] могут давать хорошие результаты только при наличии линейной связи между структурой и активностью. Один из способов преодоления вышеуказанного недостатка регрессионного анализа заключается в использовании непараметрических методов, в частности различных методов машинного обучения (machine learning, ML). Некоторые существующие модели разработаны с использованием достаточно скромных обучающих и контрольных выборок, что в значительной степени ограничивает структурное пространство, описываемое моделью. Так, в работе [5] обучающая и контрольная выборки были сформированы из набора, представленного всего 32 ингибиторами HDAC6, производными гидроксамовой кислоты.

В связи с вышесказанным задача настоящего исследования заключалась в следующем:

- 1) построение QSAR-моделей ингибиторов HDAC6, содержащих гидроксамовые кислоты;
- 2) структурная интерпретация разработанных QSAR-моделей и рациональный молекулярный дизайн ингибиторов HDAC6;
- 3) выполнение многопараметрического анализа для выявления сконструированных структур, обладающих оптимальным балансом ADMET-свойств.

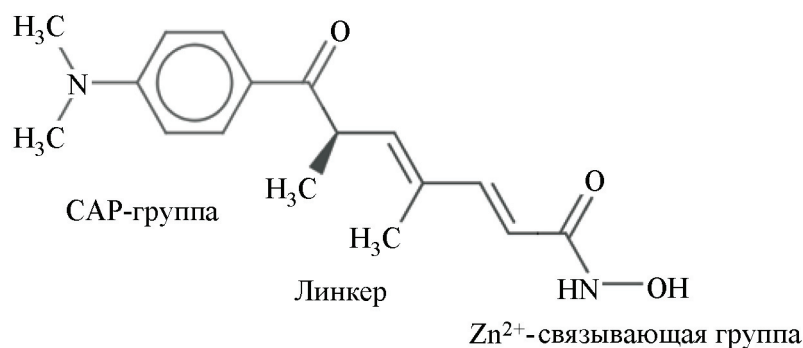


Рис. 1. Фармакофорное строение ингибиторов гистондеацетилаз на примере соединения Трихостатин А

Методы исследований

Источником формирования общей выборки для QSAR-моделирования послужили публикации [10–25], в каждой из которых приведены экспериментальные значения IC_{50} для ингибиторов HDAC6, содержащих в структуре соединений фрагменты гидроксамовой кислоты. При формировании выборки в нее не были включены полимеры, смеси, соединения в солевой форме. С нашей точки зрения, объединение экспериментальных данных вышеуказанных публикаций допустимо, поскольку среднее значение IC_{50} для референсного ингибитора HDAC Вориностата составляет $10,4 \pm 3$ нМ. С учетом того, что общая выборка, представленная 131 соединением, была интегрирована из разных источников, целесообразно построение классификационных QSAR-моделей, которые могут в некоторой степени снизить уровень межлабораторных ошибок при исследовании ингибиторов HDAC6. Среднее значение IC_{50} референсного ингибитора HDAC Вориностат (10,4 нМ) было использовано в качестве порога разделения изучаемых соединений на два класса: вещества со значением IC_{50} меньше или равным 10,4 нМ относили к классу активных соединений, если IC_{50} было больше 10,4 нМ, вещество считалось неактивным. Введение данного порога активности позволит выявлять соединения, обладающие более высокой активностью, чем известный ингибитор HDAC Вориностат, широко используемый в химиотерапии. В целях оценки предсказательной способности моделей общий набор соединений был разделен на обучающую и контрольную выборки. Для этого соединения общего набора были упорядочены по величине

IC_{50} , при этом каждое пятое соединение помещали в контрольную выборку, а оставшиеся соединения – в обучающую выборку. Таким образом, обучающая выборка была представлена 105 соединениями (38 активных и 67 неактивных), в то время как контрольный набор был представлен 26 соединениями (соответственно 9 активных и 17 неактивных). Для стандартизации структур использовали «ChemAxon Standardizer» [26]. Обучающая и контрольная выборки представлены в свободном доступе на ресурсе github по адресу <https://github.com/ovttiras/HDAC6-inhibitors/tree/main/datasets>.

Для описания молекулярной структуры использовали 2D-симплексные дескрипторы, которые рассчитываются в рамках симплексного представления молекулярной структуры (СПМС, Simplex representation of molecular structure, SiRMS) [27]. В рамках SiRMS молекула рассматривается как система различных симплексов – четырехатомных молекулярных фрагментов фиксированной структуры (рис. 2).

Дескриптором в этом случае является число симплексов определенного вида. На 2D-уровне атомы (вершины симплекса) дифференцируют не только по природе (метке) атома, но и с учетом разных физико-химических свойств – частичного заряда на атоме, липофильности, рефракции, способности выступать донором или акцептором водорода при образовании водородной связи.

Структурную интерпретацию при использовании симплексных дескрипторов проводили в соответствии с подходом [28], при котором вклад исследуемого фрагмента (С) рассчитывали

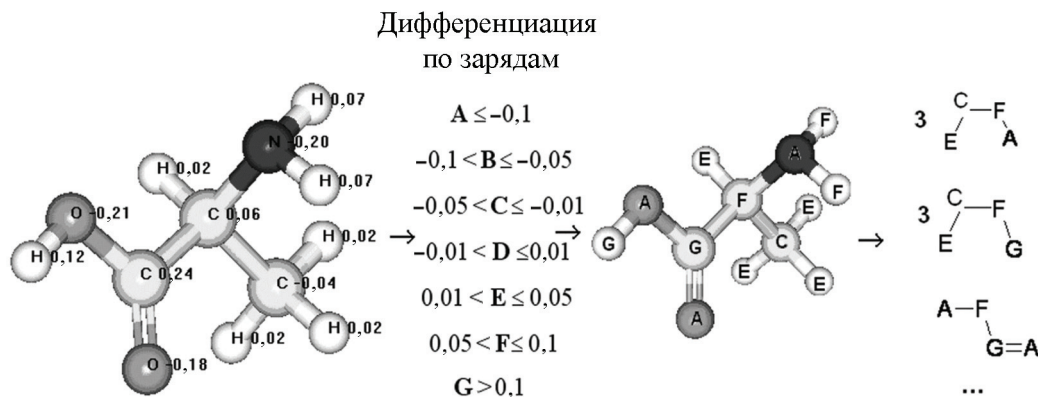


Рис. 2. Пример 2D-генерации симплексных дескрипторов для аланина на 2D-уровне при использовании дифференциации атомов по их частичным зарядам

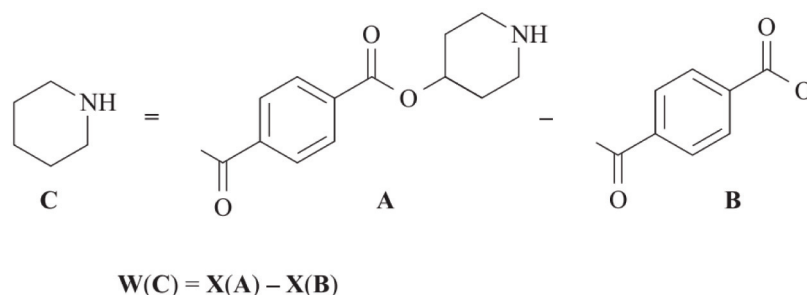


Рис. 3. Использованный принцип структурной интерпретации. $W(C)$ – вклад фрагмента (C), $X(A)$ – предсказанное значение активности родительской структуры (A); $X(B)$ – предсказанное значение активности для гипотетической структуры (B)

путем нахождения разницы между рассчитанными значениями активности для родительской структуры (A) и гипотетической структуры (B), полученной удалением исследуемого фрагмента (C) из родительской структуры (A) (рис. 3).

При моделировании с использованием симплексных дескрипторов применяли библиотеку «Scikit-learn» [29] для языка программирования Python, в которой реализованы методы «Случайного леса» (Random Forest, RF), «Опорных векторов» (Support Vector Machine, SVM), «Градиентного бустинга» (Gradient Boosting Method, GBM) для построения моделей QSAR.

При QSAR-моделировании с использованием симплексных дескрипторов для контрольной выборки учитывали вхождение соединений в область применимости (applicability domain, AD) [30]. При этом, если значение хотя бы одного дескриптора выходило за пределы его минимального или максимального значения для обучающей выборки, то соединение контрольной выборки, содержащее данный дескриптор, не входило в область применимости (bounding box). Вышеописанный подход для QSAR-моделирования с использованием симплексных дескрипторов реализован в виде программного обеспечения «SPCI», которое находится в свободном доступе по адресу http://qsar4u.com/pages/sirms_qsar.php.

Кроме того, для описания молекулярной структуры использовали физико-химические молекулярные 2D-дескрипторы, рассчитываемые с помощью библиотеки RDKit на языке программирования Python [31]. Для построения моделей с помощью молекулярных дескрипторов 2D RDKit использовали библиотеку «Scikit-learn» [29], при этом применяли методы RF, GBM, SVM. Матрица использован-

ных молекулярных дескрипторов 2D RDKit доступна по адресу https://github.com/ovttiras/HDAC6-inhibitors/blob/main/matrix_descriptors.csv. Вышеуказанные физико-химические молекулярные дескрипторы 2D RDKit принято делить на различные классы, которые подробно освещены по ссылке <https://www.rdkit.org/docs/GettingStartedInPython.html#list-of-available-descriptors>.

Вхождение соединений контрольной выборки в область применимости рассчитывали с помощью «расстояния сходства» (similarity distance) [32]. Считается, что соединение контрольной выборки принадлежит области применимости QSAR-модели, если его расстояние сходства не превышает порогового значения D_c , вычисляемого по формуле (1):

$$D_c = Z\sigma + \bar{y}, \quad (1)$$

где \bar{y} и σ – соответственно среднее значение и стандартное отклонение величин евклидова расстояния в химическом пространстве дескрипторов между всеми объектами из обучающей выборки и их ближайшими соседями в ней; Z – константа, которая, как правило, принимается равной 0,5.

Для оценки устойчивости моделей была использована пятикратная внутренняя кросс-валидация (CV, 5-fold cross validation). Для этого все соединения обучающей выборки случайным образом разделяются на 5 частей. Затем на 4 частях данных, объединенных в обучающую выборку, производится построение (обучение) QSAR-модели, а оставшаяся часть данных используется как внутренняя контрольная выборка. Эта процедура повторяется 5 раз, в результате чего каждая из 5 частей данных последовательно используется для внутренней валидации.

Важно подчеркнуть, что исследуемые соединения никогда не являются одновременно частью как обучающей, так и внутренней контрольной выборки. Поскольку обучающая выборка была достаточно несбалансированная, нами использовалась стратифицированная пятикратная внутренняя кросс-валидация [33].

QSAR-модели, разработанные с помощью симплексных дескрипторов 2D RDKit, интегрировались соответственно в две консенсусные QSAR-модели. Отнесение исследуемых соединений к классу активных или неактивных консенсусной моделью происходило исходя из того, какой из двух классов чаще прогнозировался частными моделями, разработанными с использованием методов RF, GBM, SVM.

Для валидации QSAR-моделей, построенных с помощью 2D RDKit дескрипторов, также применяли метод у-рандомизации [32] с 500 итерациями.

Оценку точности и прогностической способности предложенных в работе моделей и их сравнение с другими QSAR-моделями выполняли с помощью следующих статистических критериев.

1. Чувствительность (Sensitivity, SEN):

$$SEN = \frac{TP}{TP + FN}. \quad (2)$$

2. Специфичности (Specificity, SPC):

$$SPC = \frac{TN}{TN + FP}. \quad (3)$$

3. Сбалансированная точность (Balanced accuracy, BA):

$$BA = \frac{SEN + SPC}{2}. \quad (4)$$

4. Коэффициент каппа Коэна (Cohen's kappa coefficient, Каппа):

$$\text{Каппа} = \frac{2 \times (TP \times TN - FN \times FP)}{(TP + FP) \times (FP + TN) + (TP + FN) \times (FN + TN)}. \quad (5)$$

О б о з н а ч е н и я: TN (True Negative) – число верно классифицированных неактивных соединений, TP (True Positive) – число верно классифицированных активных соединений, FN (False Negative) и FP (False Positive) – число неверно классифицированных неактивных и активных соединений соответственно.

Результаты и обсуждение

Результаты QSAR-моделирования приведены в табл. 1. Все разработанные модели имеют удовлетворительные статистические характеристики и обладают сопоставимой предсказательной способностью. Модели, разработанные с использованием RDKit-дескрипторов, находятся в свободном доступе по ссылке <https://github.com/ovttiras/HDAC6-inhibitors> и могут быть использованы для виртуального скрининга. Доля соединений (coverage), вошедших в области применимости QSAR-моделей с использованием симплексных дескрипторов, составляет 0,42. Все соединения контрольной выборки вошли в области применимости соответствующих QSAR-моделей, построенных с помощью 2D RDKit-дескрипторов (coverage = 1).

Статистические характеристики QSAR-моделей, разработанных с использованием 208 RDKit-дескрипторов при 500 раундах у-рандомизации, представлены в табл. 2. Из данных этой таблицы видно, что при построении QSAR-моделей в условиях у-рандомизации максимальные значения сбалансированной точности не превышают BA моделей, сконструированных без использования указанной процедуры, что подтверждает отсутствие случайной корреляции в предлагаемых QSAR-моделях (табл. 1).

С помощью библиотеки «SHAP» на языке программирования Python (<https://github.com/slundberg/shap>) нами была проведена оценка важности дескрипторов при построении QSAR-моделей с использованием RDKit-дескрипторов. Оценка важности признаков в библиотеке «SHAP» рассчитывается с использованием значений Шепли [34]. Из представленных на рис. 4–6 результатов проведенного анализа можно отметить, что значительное влияние оказывают дескрипторы PEOE_VSA12, VSA_EState7, Minimum absolute EState index, BCUT2D_LOGPLOW, BCUT2D_CHGLO. Дескриптор PEOE_VSA12 рассчитывается как часть ван-дер-ваальсовой поверхности, внутри которой сумма частичных зарядов находится в интервале 0,23–0,30. Дескриптор VSA_EState7 также рассчитывается как часть ван-дер-ваальсовой поверхности, при этом индекс электротопологического состояния EState должен находиться в интервале 1,81–2,05. Индексы электротопологического состояния учитывают влияние топологического окружения атома на его электронные свойства

Т а б л и ц а 1

Статистические характеристики QSAR-моделей

Дескрипторы	Метод	Обучающая выборка (5-fold CV)				Контрольная выборка			
		BA	SEN	SPC	Каппа	BA	SEN	SPC	Каппа
SIRMS	RF	0,71	0,76	0,66	0,39	0,94	1,0	0,88	0,84
	GBM	0,73	0,76	0,70	0,44	0,94	1,0	0,88	0,84
	SVM	0,71	0,82	0,61	0,39	0,85	1,0	0,71	0,62
	консенсусная модель	0,72	0,78	0,66	0,41	0,94	1,0	0,88	0,84
RDKit	RF	0,81	0,71	0,91	0,64	0,92	0,89	0,94	0,83
	GBM	0,81	0,76	0,87	0,63	0,89	0,89	0,88	0,75
	SVM	0,81	0,74	0,88	0,62	0,92	0,89	0,94	0,83
	консенсусная модель	0,83	0,74	0,93	0,68	0,92	0,89	0,94	0,83

Т а б л и ц а 2

Статистические характеристики QSAR-моделей в условиях у-рандомизации

Дескрипторы	Метод	BA	BA_y_random	Max	Min	STD
RDKit	RF	0,81	0,50	0,67	0,36	0,05
	GBM	0,81	0,50	0,67	0,35	0,06
	SVM	0,81	0,50	0,67	0,33	0,06

П р и м е ч а н и я. BA – сбалансированная точность без использования процедуры у-рандомизации; BA_y_random – средняя сбалансированная точность в условиях у-рандомизации; Max – максимальное значение сбалансированной точности в условиях у-рандомизации; Min – минимальное значение сбалансированной точности в условиях у-рандомизации; STD – стандартное отклонение среднего значения сбалансированной точности в условиях у-рандомизации.

[35]. Дескриптор Minimum absolute EState index определяется как минимальное абсолютное значение индекса электротопологического состояния EState. Дескрипторы BCUT2D_LOGPLOW, BCUT2D_CHGLO относятся к группе дескрипторов BCUT, которые были созданы для кодирования как топологии, так и различных физико-химических свойств атомов [36]. Эти дескрипторы базируются на модифицированной матрице Бюрдена, в которой диагональные элементы описывают заряд атома, величины, связанные с поляризуемостью атома, а также способность к образованию водородных связей. Недиагональные элементы этой матрицы учитывают тополо-

гические и геометрические расстояния между атомами. Шесть (или восемь) значений, соответствующих максимальным и минимальным собственным значениям этих матриц, являются дескрипторами BCUT.

При интерпретации консенсусной модели, построенной с помощью симплексных дескрипторов, были определены вклады молекулярных фрагментов в величину активности ингибиторов HDAC6 (рис. 7). Этот набор молекулярных фрагментов сформирован исходя из вышеуказанной концепции трехкомпонентной структуры классических ингибиторов гистондеацетилаз и представленного химического

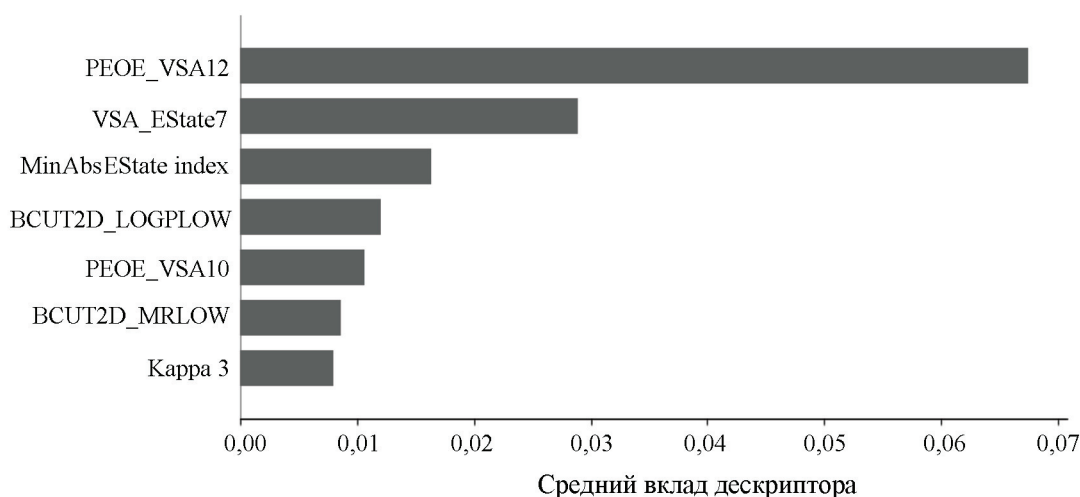


Рис. 4. ТОП-7 наиболее значимых дескрипторов RDKit при построении QSAR-модели с помощью метода RF

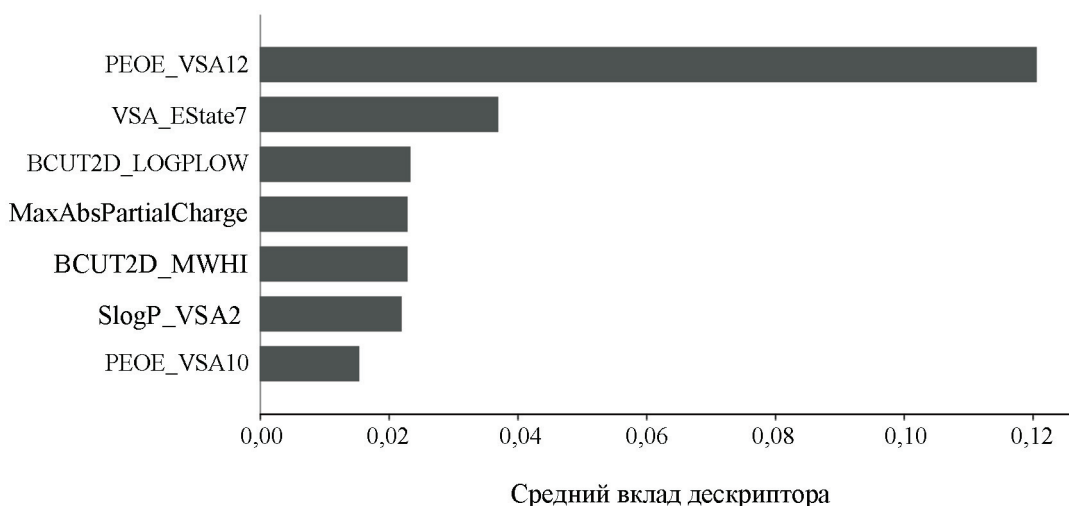


Рис. 5. ТОП-7 наиболее значимых дескрипторов RDKit при построении QSAR-модели с помощью метода GBM

разнообразия соединений обучающей выборки. Интерпретации подверглись только те молекулярные фрагменты, которые были обнаружены в пяти и более соединениях, что, с нашей точки зрения, позволило сконцентрировать внимание на фрагментах, устойчиво влияющих на ингибирующую активность и избежать в некоторой степени влияния случайных факторов, например ошибок в экспериментальных данных или предсказанных значениях активности и вкладах фрагментов. Полный список идентифицированных молекулярных фрагментов, записанных в виде SMARTS с рассчитанными вкладами в активность, приведен в дополнительных материалах (Таблица 1S и Рисунок 1S), которые

размещены по адресу <https://github.com/ovttiras/HDAC6-inhibitors/tree/main/supplementary%20material>.

Проведенная интерпретация позволила количественно описать и ранжировать влияние молекулярных фрагментов на изменение активности ингибиторов HDAC6, детализировать молекулярное окружение известных функциональных групп, выделив производные фрагменты, повышающие и понижающие вклады в указанный вид активности. Так, например, при детализации 1,2,3-триазола выделен молекулярный фрагмент 1,4-дифенил-1H-1,2,3-триазол (**f2** на рис. 7), который в значительной степени увеличивает активность ингибиторов

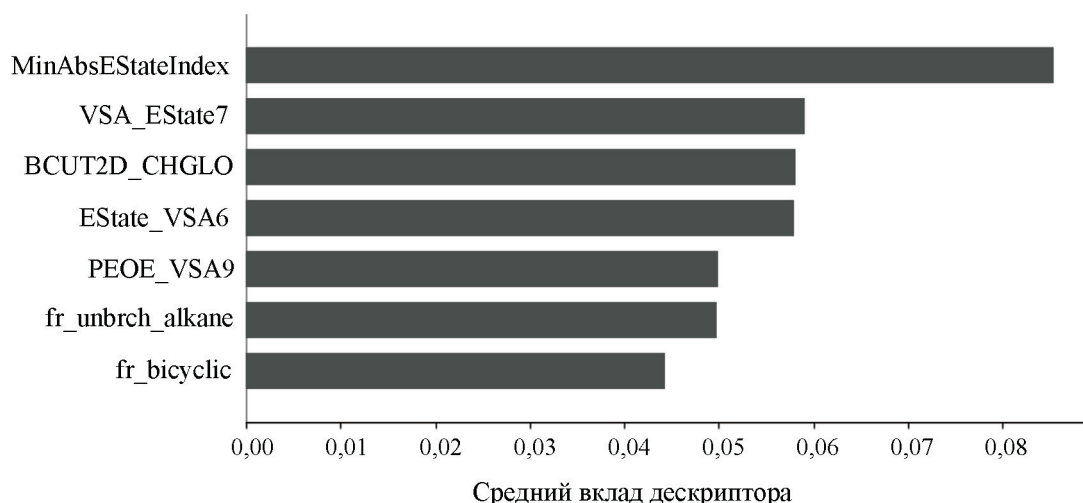


Рис. 6. ТОП-7 наиболее значимых дескрипторов RDKit при построении QSAR-модели с помощью метода SVM

HDAC6. Фенильные производные 1,2-оксазола, хиназолина и индазола (фрагменты **f10**, **f11** и **f12** на рис. 7), наоборот, уменьшают активность ингибиторов HDAC6. Результаты интерпретации указывают на то, что наиболее предпочтительными линкерами для ингибиторов HDAC6 являются фрагменты гептанамида (фрагмент **f1** на рис. 7), при этом уменьшение длины линкера на две метильные группы (фрагмент **f9**) приводит к существенному снижению активности ингибиторов HDAC6. К росту активности ингибиторов HDAC6 приводит также использование в качестве линкера фрагментов фенилэтилена (фрагмент **f3**).

Принимая во внимание тенденции влияния структуры соединений на изменение активности, выявленные в ходе интерпретации, был проведен рациональный молекулярный дизайн и предложен ряд перспективных ингибиторов HDAC6. Конструирование ингибиторов HDAC6 проводили путем комбинирования линкеров (фрагменты **f1** и **f3**) и CAP-групп (фрагменты **f2** и **f4**), которые согласно результатам интерпретации повышают активность ингибиторов HDAC6. В результате были предложены три гипотетических соединения (рис. 8, вещества **1**, **2**, **3**), обладающие более высокой ингибирующей активностью, рассчитанной консенсусной моделью, чем известный ингибитор HDAC6 Вориностат, и входящие в области применимости QSAR-моделей, разработанных с помощью дескрипторов RDKit. Для всех приведенных на рис. 8 потенциальных ингибиторов HDAC6 была проведена проверка наличия структур в

базе данных PubChem (<https://pubchem.ncbi.nlm.nih.gov/>), в ходе которой указанные вещества не были обнаружены. Необходимо отметить, что при молекулярном дизайне в данных примерах учитывались результаты интерпретации QSAR модели, построенной с помощью симплексных дескрипторов, в то время как прогнозирование активности ингибиторов проводилось с помощью QSAR-модели, построенной с использованием RDKit-дескрипторов.

При определении стратегии синтеза и испытаний помимо целевой активности важно оценивать ADMET-свойства, которые являются существенными факторами при принятии решения о рекомендации к использованию соединения в качестве действующего вещества лекарственного средства. В этих целях для соединений **1–3** была проведена оценка острой токсичности (LD_{50}) при пероральном введении крысам с помощью программы «Т.Е.С.Т. v.4.2.», разработанной специалистами Агентства по охране окружающей среды США [37]. Для этих соединений с помощью интернет-платформы швейцарского Института биоинформатики «swissADME» (<http://www.swissadme.ch/>) [38] оценивалась также липофильность ($\text{Log } P_{o/w}$), ряд важнейших ADME-свойств, соответствие правилам Липински [39], синтетическая доступность по десятибалльной шкале (0 – максимальная степень синтетической доступности, 10 – минимальная степень синтетической доступности) [40], что крайне важно для предложенных, но еще не синтезированных соединений. Результаты прогнозирования приведены

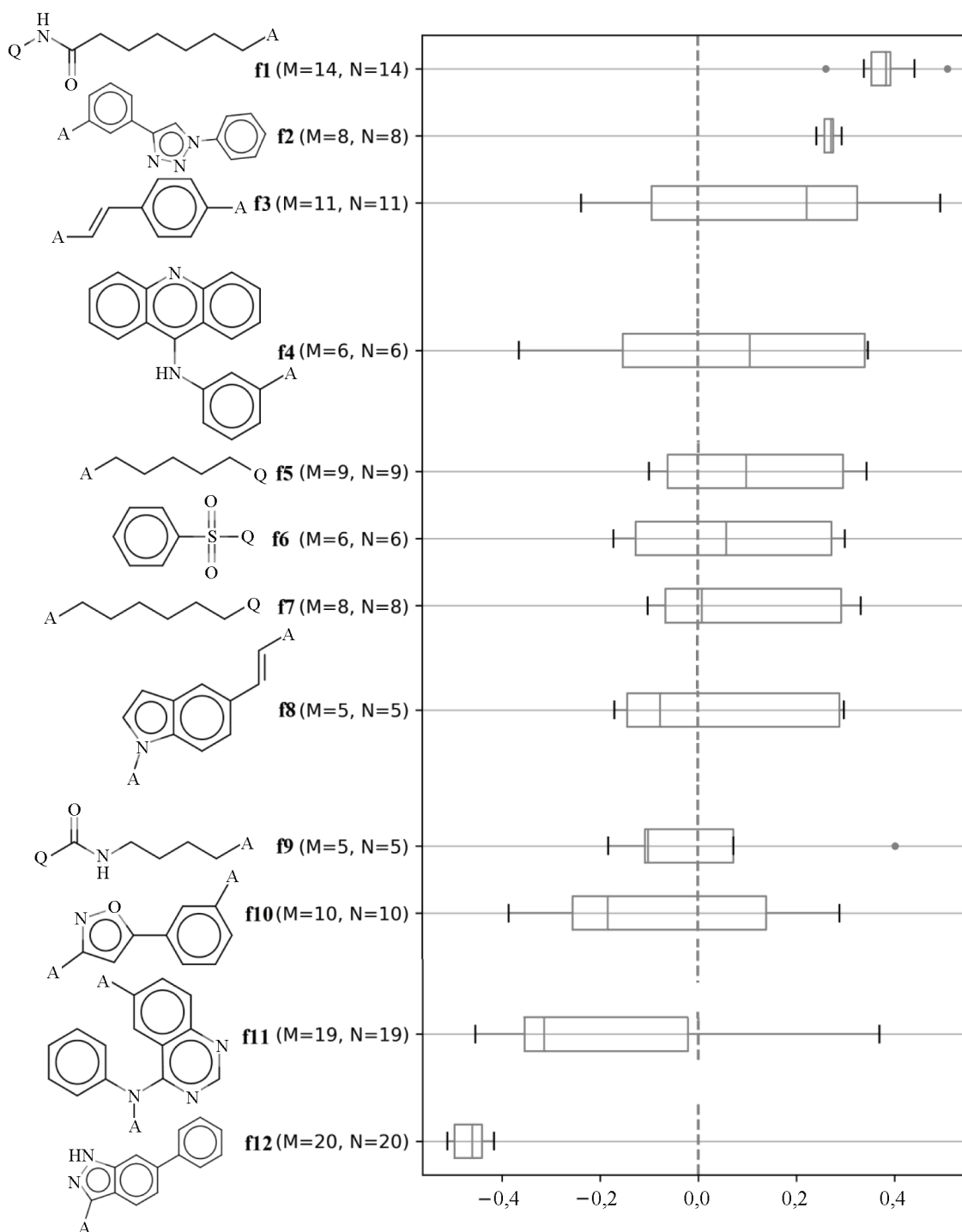


Рис. 7. Вклады молекулярных фрагментов в способность соединений ингибировать HDAC6 (A – место присоединения фрагмента к оставшейся части молекулы, Q – место присоединения фрагмента к ароматической системе оставшейся части молекулы, f – порядковый номер фрагмента, M – число соединений, содержащих данный фрагмент, N – число обнаружений соответствующего фрагмента в выборке)

в табл. 3, из которых видно, что предложенные в ходе молекулярного дизайна соединения **1–3** обладают приемлемой синтетической доступностью.

По результатам проведенного анализа (табл. 3), изучаемые соединения **2** и **3** обладают высокой всасываемостью в кишечнике чело-

века, но не способны проникать через гематоэнцефалический барьер.

Референсное соединение Вориностат в настоящее время используется в качестве лекарственного средства при терапии Т-клеточной лимфомы [41]. В этой связи крайне важно оценить

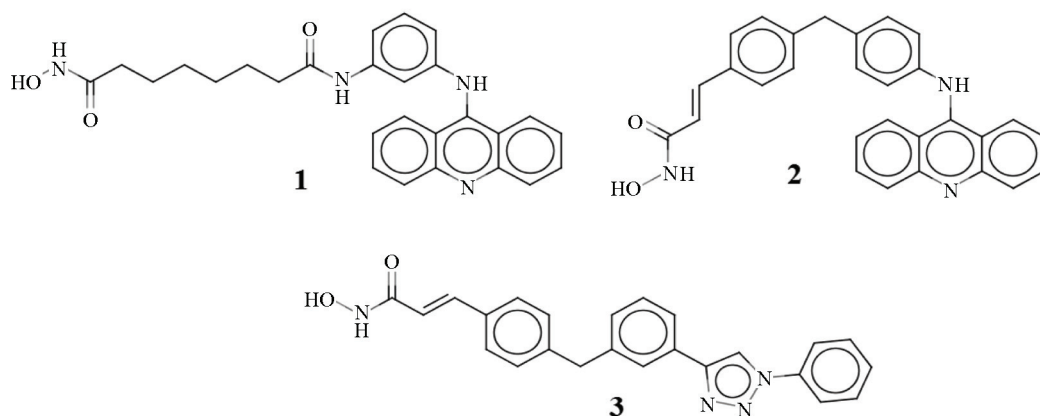


Рис. 8. Результаты молекулярного дизайна. Соединение **1** – N'-{3-[(акридин-9-ил)амино]фенил}-N-гидроксиоктандиамид; соединение **2** – (2E)-N-гидрокси-3-(4-{[3-(1-фенил-1H-1,2,3-триазол-4-ил)фенил]метил}фенил)проп-2-енамид; соединение **3** – (2E)-3-[4-{[4-[(акридин-9-ил)амино]фенил]метил}фенил]-N-гидроксипроп-2-енамид

Т а б л и ц а 3

Оценка некоторых физико-химических свойств, видов токсичности и синтетической доступности соединений, исследуемых в рамках молекулярного дизайна (структуры соединений представлены на рис. 8)

Свойство/вид токсичности	Соединение			
	1	2	3	Вориностат
Log $P_{o/w}$	4,28	3,71	5,00	1,92
Всасываемость в кишечнике человека	низкая	высокая	высокая	высокая
Способность проникать через гематоэнцефалический барьер	нет	нет	нет	нет
Субстрат P-гр	нет	нет	нет	нет
Ингибитор CYP1A2	да	нет	нет	нет
Ингибитор CYP2C19	да	да	да	нет
Ингибитор CYP2C9	да	да	нет	нет
Ингибитор CYP2D6	да	нет	да	нет
Ингибитор CYP3A4	да	нет	нет	нет
Log Kp	-5,72	-5,78	-4,62	-6,59
Число несоответствий правилам Липински	0	0	1 (Log $P_{o/w} \geq 5$)	0
Синтетическая доступность	3,36	3,57	3,41	1,91
LD ₅₀ , крысы при пероральном введении, мг/кг	4338	2046	1171	4886

П р и м е ч а н и е. Единицы измерения Kp – см/с. Чем более отрицательно значение log Kp , тем медленнее вещество проникает через кожу.

способность изучаемых соединений проникать через кожу. Данные табл. 3 указывают на то, что соединения **1–3** демонстрируют более высокую способность проникать через кожу по сравнению с Вориностатом.

Сведения о соединениях, не являющихся субстратом P-гликопротеина (P-glycoprotein, P-gr), важны для оценки потенциальных лекарств, поскольку P-gr сверхэкспрессируется в некоторых опухолевых клетках и приводит к раку с

множественной лекарственной устойчивостью. P-gp транспортирует различные субстраты через клеточную мембрану, тем самым повышая лекарственную устойчивость при лечении опухолей [42]. Согласно проведенным расчетам (табл. 3), изучаемые соединения **1–3** не являются субстратами P-gp, что дает определенную надежду на невысокий уровень лекарственной устойчивости при использовании этих соединений в химиотерапии онкологических заболеваний.

Важна также информация о взаимодействии целевых соединений с цитохромами P450 (CYP). Ингибирование пяти основных изоформ P450 (CYP1A2, CYP2C19, CYP2C9, CYP2D6, CYP3A4) является одной из причин фармакокинетических межлекарственных взаимодействий, приводящих к токсическим или другим нежелательным побочным эффектам из-за более низкого клиренса и накопления препарата или его метаболитов в организме человека [43]. Согласно проведенным расчетам, соединения **2** и **3** способны ингибировать две из пяти основных изоформ цитохромов P450.

Результаты прогнозирования LD₅₀ при пероральном введении крысам дают основание полагать, что рассматриваемые соединения относятся к III категории токсичности (незначительно токсичные) в соответствии с классификацией Агентства по охране окружающей среды США (United States Environmental Protection Agency, U.S. EPA) [44]. За исключением соединения **3** все вещества удовлетворяют правилам пяти Липински.

Таким образом, соединение **2** может быть рекомендовано для дальнейших исследований в качестве перспективного ингибитора HDAC6 при пероральном приеме, в то время как соединение **3** целесообразно исследовать в качестве ингибитора HDAC6 при трансдермальном введении.

Заключение

В ходе проведения вычислительных экспериментов с применением концептуально различных дескрипторов и методов машинного обучения были разработаны приемлемые QSAR-модели ингибиторов HDAC6.

Проведенная структурная интерпретация QSAR-моделей позволила выявить общие закономерности влияния структуры химических соединений на их ингибирующую активность, выделив молекулярные фрагменты, которые повышают и понижают активность ингибиторов HDAC6. Результаты структурной интерпретации использованы при проведении рационального молекулярного дизайна, в ходе которого предложен ряд перспективных ингибиторов HDAC6. Для сконструированных структур проведена *in silico* оценка основных ADMET-свойств. Выявленные в ходе оценки приемлемые фармакокинетические показатели и относительно невысокая острая токсичность позволяют рекомендовать приведенные структуры для дальнейших доклинических исследований в качестве перспективных противоопухолевых препаратов.

Предложенные QSAR-модели размещены в свободном доступе на сайте github (<https://github.com/ovttiras/HDAC6-inhibitors>) и могут быть использованы для виртуального скрининга новых перспективных ингибиторов HDAC6. Кроме того, открытый исходный код на языке программирования Python, доступный по вышеуказанной интернет-ссылке, позволяет проводить QSAR-моделирование других видов биологической активности.

Полученные в ходе исследования результаты будут способствовать снижению финансовых, временных и трудовых затрат при синтезе и испытании новых противоопухолевых препаратов.

СПИСОК ЛИТЕРАТУРЫ

1. Aldana-Masangkay G.I., Sakamoto K.M. // J Biomed Biotechnol. 2011. Vol. 2011. P. 875824 (DOI: 10.1155/2011/875824).
2. Mottamal M., Zheng S., Huang T.L., Wang G. // Molecules. 2015. Vol. 20. P. 3898. (DOI: 10.3390/molecules20033898).
3. Muratov E.N., Bajorath J., Sheridan R.P., Tetko I.V., Filimonov D., Poroikov V., Oprea T.I., Baskin I.I., Varnek A., Roitberg A., Isayev O., Curtalolo S., Fourches D., Cohen Y., Aspuru-Guzik A., Winkler D.A., Agrafiotis D., Cherkasov A., Tropsha A. // Chem Soc Rev. 2020. QSAR without borders. Vol. 49. N 11. P. 3525 (DOI: 10.1039/d0cs00098a).
4. Ruzic D., Djokovic N., Nikolic K. // Methods Mol. Biol. 2021. Vol. 2266. P. 155 (DOI: 10.1007/978-1-0716-1209-5_9).
5. Ruzic D., Petkovic M., Agbaba D., Ganesan A., Nikolic K. // Mol Inform. 2019. Vol. 38. P. e1800083 (DOI:10.1002/minf.201800083).
6. Hu C., Hong L., Du W. // Lett. Drug. Des. Discov. 2017. Vol. 14. P. 798 (DOI: 10.2174/1570180813666161028165151).
7. Vijayarathay S., Chatterjee J. // Int. J. Curr. Res. 2016. Vol. 8. P. 39618.
8. Mittal L., Kumari A., Srivastava M., Singh M., Asthana S. // J. Biomol. Struct. Dyn. 2020. Advance

- online publication. 1-19 (DOI: 10.1080/07391102.2020.1768151).
9. Gyebi G.A., Ogunro O.B., Adegunloye A.P., Ogunyemi O.M., Afolabi S.O. // *J. Biomol. Struct. Dyn.* 2020. Advance online publication. 1–13 (DOI: 10.1080/07391102.2020.1764868).
 10. Tapadar S., He R., Luchini D.N., Billadeau D.D., Kozikowski A.P. // *Bioorganic & medicinal chemistry letters.* 2009. Vol. 19. N 11. P. 3023 (DOI: 10.1016/j.bmcl.2009.04.058).
 11. He R., Chen Y., Ougolkov A.V., Zhang J.S., Savoy D.N., Billadeau D.D., Kozikowski A.P. // *J. Medicinal Chemistry.* 2010. Vol. 53 N 3. P. 1347 (DOI: 10.1021/jm901667k).
 12. Kozikowski A.P., Tapadar S., Luchini D.N., Kim K.H., Billadeau D.D. // *J. Medicinal Chemistry.* 2008. Vol. 51. N 15. P. 4370 (DOI: 10.1021/jm8002894).
 13. Conti P., Tamborini L., Pinto A., Sola L., Ettari R., Mercurio C., De Micheli C. // *European journal of medicinal chemistry.* 2010. Vol. 45 N 9. P. 4331 (DOI: 10.1016/j.ejmech.2010.06.035).
 14. Hutt D.M., Herman D., Rodrigues A.P., Noel S., Pilewski J.M., Matteson J., Hoch B., Kellner W., Kelly J.W., Schmidt A., Thomas P.J., Matsumura Y., Skach W.R., Gentsch M., Riordan J.R., Sorscher E.J., Okiyoneda T., Yates J.R., Lukacs G.L., Frizzell R.A., Balch W.E. // *Nature chemical biology.* 2010. Vol. 6. N 1. P. 25–33 (DOI: 10.1038/nchembio.275).
 15. Lai M.J., Huang H.L., Pan S.L., Liu Y.M., Peng C.Y., Lee H.Y., Yeh T.K., Huang P.H., Teng C.M., Chen C.S., Chuang H.Y., Liou J.P. // *J. Medicinal Chemistry.* 2012. Vol. 55. N 8. P. 3777 (DOI: 10.1021/jm300197a).
 16. Lee H.Y., Tsai A.C., Chen M.C., Shen P.J., Cheng Y.C., Kuo C.C., Pan S.L., Liu Y.M., Liu J.F., Yeh T.K., Wang J.C., Chang C.Y., Chang J.Y., Liou J.P. // *J. Medicinal Chemistry.* 2014. Vol. 57 N 10. P. 4009 (DOI: 10.1021/jm401899x).
 17. Yang Z., Wang T., Wang F., Niu T., Liu Z., Chen X., Long C., Tang M., Cao D., Wang X., Xiang W., Yi Y., Ma L., You J., Chen L. // *J. Medicinal Chemistry.* 2016. Vol. 59. N 4, P. 1455 (DOI: 10.1021/acs.jmedchem.5b01342).
 18. Ding C., Chen S., Zhang C., Hu G., Zhang W., Li L., Chen Y.Z., Tan C., Jiang Y. // *Bioorganic & medicinal chemistry.* 2017. Vol. 25. N 1. P. 27 (DOI: 10.1016/j.bmc.2016.10.006).
 19. Raji I., Ahluwalia K., Oyelere A. K. // *Bioorganic & Medicinal Chemistry Letters.* 2017. Vol. 27. N 4. P. 744 (DOI: 10.1016/j.bmcl.2017.01.044).
 20. Chen C., Hou X., Wang G., Pan W., Yang X., Zhang Y., Fang H. // *European Journal of Medicinal Chemistry.* 2017. Vol. 133. P. 11 (DOI: 10.1016/j.ejmech.2017.03.064).
 21. Yuan Z., Sun Q., Li D., Miao S., Chen S., Song L., Gao C., Chen Y., Tan C., Jiang Y. // *European Journal of Medicinal Chemistry.* 2017. Vol. 134. P. 281. (DOI: 10.1016/j.ejmech.2017.04.017).
 22. Yuan Z., Chen S., Sun Q., Wang N., Li D., Miao S., Gao C., Chen Y., Tan C., Jiang Y. // *Bioorganic & Medicinal Chemistry.* 2017. Vol. 25. N 15. P. 4100 (DOI: 10.1016/j.bmc.2017.05.058).
 23. Chao S. W., Chen L. C., Yu C. C., Liu C. Y., Lin T. E., Guh J. H., Wang C. Y., Chen C. Y., Hsu K. C., Huang W. J. // *European Journal of Medicinal Chemistry.* 2018. Vol. 143. P.792 (DOI: 10.1016/j.ejmech.2017.11.092).
 24. Liu J., Qian C., Zhu Y., Cai J., He Y., Li J., Wang T., Zhu H., Li Z., L. W., Hu L. // *Bioorganic & Medicinal Chemistry.* 2018. Vol. 26. N 3. P. 747 (DOI: 10.1016/j.bmc.2017.12.041).
 25. Chen J., Li D., Li W., Yin J., Zhang Y., Yuan Z., Gao C., Liu F., Jiang Y. // *Bioorganic & Medicinal Chemistry.* 2018. Vol. 26. N 14. P. 3958 (DOI: 10.1016/j.bmc.2018.06.016).
 26. Standardizer, version 5.4; ChemAxon: Budapest, Hungary; Available from: <https://docs.chemaxon.com/display/docs/Standardizer+Command-line+Application> (accessed: 25.03.2022).
 27. Simplex representation of molecular structure – a chemoinformatic tool for calculation of simplex (fragment) descriptors. Available from: <https://github.com/DrrDom/sirms> (accessed: 25.03.2022).
 28. Polishchuk P., Tinkov O., Khristova T., Ognichenko L., Kosinskaya A., Varnek A., Kuz'min V. // *J. Chem. Inf. Model.* 2016. Vol. 56. N 8. P. 1455 (DOI: 10.1021/acs.jcim.6b00371).
 29. Scikit-learn. Free software machine learning library for the Python programming language. Available from: <https://scikit-learn.org/stable/> (accessed: 25.03.2022).
 30. Jaworska J., Nikolova-Jeliazkova N., Aldenberg T. // *Altern Lab Anim.* 2005. Vol. 33. N 5. P. 445 (DOI: 10.1177/026119290503300508).
 31. RDKit – open source toolkit for cheminformatics. Available from: <https://www.rdkit.org/docs/index.html> (accessed: 25.03.2022).
 32. Alves V.M., Capuzzi S.J., Braga R.C., Korn D., Hochuli J.E., Bowler K.H., Yasgar A., Rai G., Simeonov A., Muratov E. N., Zakharov A.V., Tropsha A. // *J. Chem. Inf. Model.* 2020. Vol. 60 N 8. P. 4056–4063 (DOI: 10.1021/acs.jcim.0c00415).
 33. Krstajic D., Buturovic L.J., Leahy D.E., Thomas S. // *J. Cheminformatics.* 2014. Vol. 6. N 1. P. 10 (DOI: 10.1186/1758-2946-6-10).
 34. Lipovetsky S., Conklin M. // *Applied Stochastic Models in Business and Industry*, (2001). Vol. 17. N 4. P. 319 (DOI: 10.1002/asmb.446).
 35. Kier L.B., Hall L.H. // *Pharmaceut. Res.* 1990. Vol. 7. N 8. P. 801.
 36. Pearlman R.S., Smith K.M. // *Perspectives in Drug Discovery and Design.* 1998. Vol. 9. P. 339.
 37. Martin T., Harten P., Young D., TEST (Toxicity Estimation Software Tool) V 4.1, US Environmental Protection Agency, Washington DC, USA, 2012; Available from: <https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test> (accessed: 25.03.2022).
 38. Daina A., Michielin O., Zoete V. // *Sci Rep.* 2017. Vol. 7. P. 42717 (DOI: 10.1038/srep42717).
 39. Lipinski C.A., Lombardo F., Dominy B.W., Feeney P.J. // *Adv. Drug Deliv. Rev.* 2001. Vol. 46. P. 3 (DOI: 10.1016/S0169-409X(00)00129-0).
 40. Ertl P., Schuffenhauer A. // *J Cheminform.* 2009. Vol. 1. N 1. P. 8 (DOI: 10.1186/1758-2946-1-8).

41. Zolinza (vorinostat) Capsules. Full Prescribing Information. Merck & Co. Available from: https://www.merck.com/product/usa/pi_circulars/z/zolinza/zolinza_pi.pdf (accessed: 25.03.2022).
42. Váradi A., Szakács G., Bakos E., Sarkadi B. // Novartis Found Symp. 2002. Vol. 243. P. 54 (DOI: 10.1002/2F0470846356.ch5).
43. Hollenberg P.F. // Drug Metab. Rev. 2002. Vol. 34. P. 17 (DOI: 10.1081/DMR-120001387).
44. The Environmental Protection Agency. Office of Pesticide. Programs Label Review Manual; 2018. Available from: <https://www.epa.gov/sites/default/files/2018-04/documents/lrm-complete-mar-2018.pdf> (accessed: 25.03.2022).

Информация об авторах

Тиньков Олег Викторович – доцент кафедры фармакологии и фармацевтической химии медицинского факультета Приднестровского государственного университета им. Т.Г. Шевченко, канд. хим. наук, oleg.tinkov.chem@mail.ru;

Григорьев Вениамин Юрьевич – зав. отделом компьютерного молекулярного дизайна ИФАВ РАН, докт. хим. наук, beng@ipac.ac.ru;

Григорьева Людмила Дмитриевна – доцент факультета фундаментальной физико-химической инженерии МГУ имени М.В. Ломоносова, канд. физ.-матем. наук, ldg-msu@yandex.ru.

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Статья поступила в редакцию 16.03.2022;
одобрена после рецензирования 12.05.2022;
принята к публикации 14.05.2022.