

ИССЛЕДОВАНИЕ СТРАТЕГИЙ МЕЖКЛАССОВОГО ПРОГНОЗА АКТИВНОСТИ БИФАРМАКОФОРНЫХ ИНГИБИТОРОВ БУТИРИЛХОЛИНЭСТЕРАЗЫ НА ОСНОВЕ QSAR МОДЕЛИРОВАНИЯ

© 2024 г. В. Ю. Григорьев^{1,*}, А. Н. Раздольский¹, В. П. Казаченко¹

¹ Институт физиологически активных веществ Федерального исследовательского центра проблем химической физики и медицинской химии Российской академии наук, Черноголовка, 142432 Россия

*e-mail: beng@ipac.ac.ru

Поступило в редакцию 4 октября 2024 г.

После доработки 12 декабря 2024 г.

Принято к печати 21 декабря 2024 г.

В работе исследованы три схемы межклассового прогноза активности ряда бифармакофорных ингибиторов бутирилхолинэстеразы с использованием QSAR моделирования. С помощью методов машинного обучения (множественная линейная регрессия, случайный лес, машина опорных векторов и гауссовский процесс) сконструированы QSAR модели с удовлетворительными статистическими характеристиками. На их основе изучены схемы рационального и случайного межклассового прогноза активности. Установлено, что эти схемы взаимно дополняют друг друга и оценена их относительная эффективность.

Ключевые слова: межклассовый прогноз активности, бутирилхолинэстераза, QSAR

DOI: 10.31857/S0044460X24100058, **EDN:** REXKZX

ВВЕДЕНИЕ

Разработка новых лекарственных средств является сложной и актуальной задачей, для решения которой требуются большие материальные и временные затраты. Достаточно сказать, что для того, чтобы вывести новое лекарственное средство на рынок требуется до 4.54 миллиардов долларов США [1] и 15 лет [2]. С учетом этого на фоне известных высокозатратных методов поиска новых лекарств, включая метод проб и ошибок, химическую модификацию известных лекарств и природных продуктов, высокопроизводительный скрининг, привлекательным выглядит использование компьютерного молекулярного дизайна [3, 4], который в значительной степени минимизирует вышеуказанные издержки. В основе этого подхода лежат методы молекулярного моделирования и машинного обучения. С их помощью были разработаны такие известные лекарственные

средства как каптоприл, индинавир, алискирен и другие [5].

Ключевыми этапами при использовании компьютерного молекулярного дизайна является конструирование модели биологической активности и ее применение для поиска новых физиологически активных веществ [3–6]. Для создания модели биологической активности используются различные подходы. В частности, если трехмерная структура биомолекулы известна, то для создания модели используются такие методы как молекулярный докинг, молекулярная динамика и дизайн *de novo*. При отсутствии такой информации, но при наличии ряда химических соединений с доказанной и измеренной биологической активностью, применяются различные варианты подхода известного как количественная связь структура–активность (QSAR) и фармакофорное моделирование.

Одной из основных концепций при разработке новых физиологически активных веществ является поиск соединений, обладающих различной структурой, но проявляющих одинаковый вид активности. В дальнейшем с учетом цели нашей работы для обозначения этой концепции будем использовать термин «межклассовый прогноз активности» (МПА). В рамках этой концепции сформирован подход, который в англоязычной научной литературе обозначается как «scaffold hopping» (SH) [7]. По сути, он представляет собой прогноз активности соединений одного класса с использованием модели, созданной на основе соединений другого класса, путем модификации/замены ядра (scaffold) активной молекулы. При этом в качестве ядра могут рассматриваться различные структурные элементы, включая фармакофоры, «отпечатки пальцев» и др. В качестве основных направлений использования SH можно выделить виртуальный скрининг [8] и целевую модификацию свойств физиологически активных веществ [9]. SH используется в сочетании с различными методами, в том числе с QSAR [10]. При этом указанный метод может использоваться в различном качестве. Например, для построения модели активности известных соединений с последующей визуализацией и SH анализом [11]. В другой работе [12] на основе сгенерированных ядер определенным образом формируются обучающая и тестовая выборки соединений и после создания QSAR модели анализируются ядра активных соединений тестовой выборки с использованием SH подхода. Наконец, в качестве другого примера можно привести работу [13], в которой QSAR модель используется просто для подтверждения того, что все исследуемые соединения обладают одинаковым типом активности по отношению к выбранной биомишени. На фоне различных сочетаний QSAR+SH выглядит малоисследованным и представляет определенный интерес изучение в рамках концепции технологии МПА, в которой QSAR модель непосредственно используется для оценки активности соединений с новыми ядрами.

В последнее время одним из перспективных направлений в медицине стала полифармакология [14, 15], которая занимается разработкой или использованием фармацевтических агентов, одновременно действующих на несколько биомишеней или на пути распространения болезней. В частности, такой

подход применяется при поиске новых химических соединений для лечения болезни Альцгеймера [16]. Известные лекарства для лечения болезни Альцгеймера, такие как донепезил, ривастигмин, являются мультитаргетными соединениями и активны, в частности, по отношению к AChE и BChE [17, 18]. Таким образом, полифармакофорные мультитаргетные соединения являются перспективными объектами для дальнейших исследований.

Настоящая работа посвящена изучению различных стратегий межклассового прогноза активности ряда бифармакофорных ингибиторов BChE с использованием QSAR и SH моделирования.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

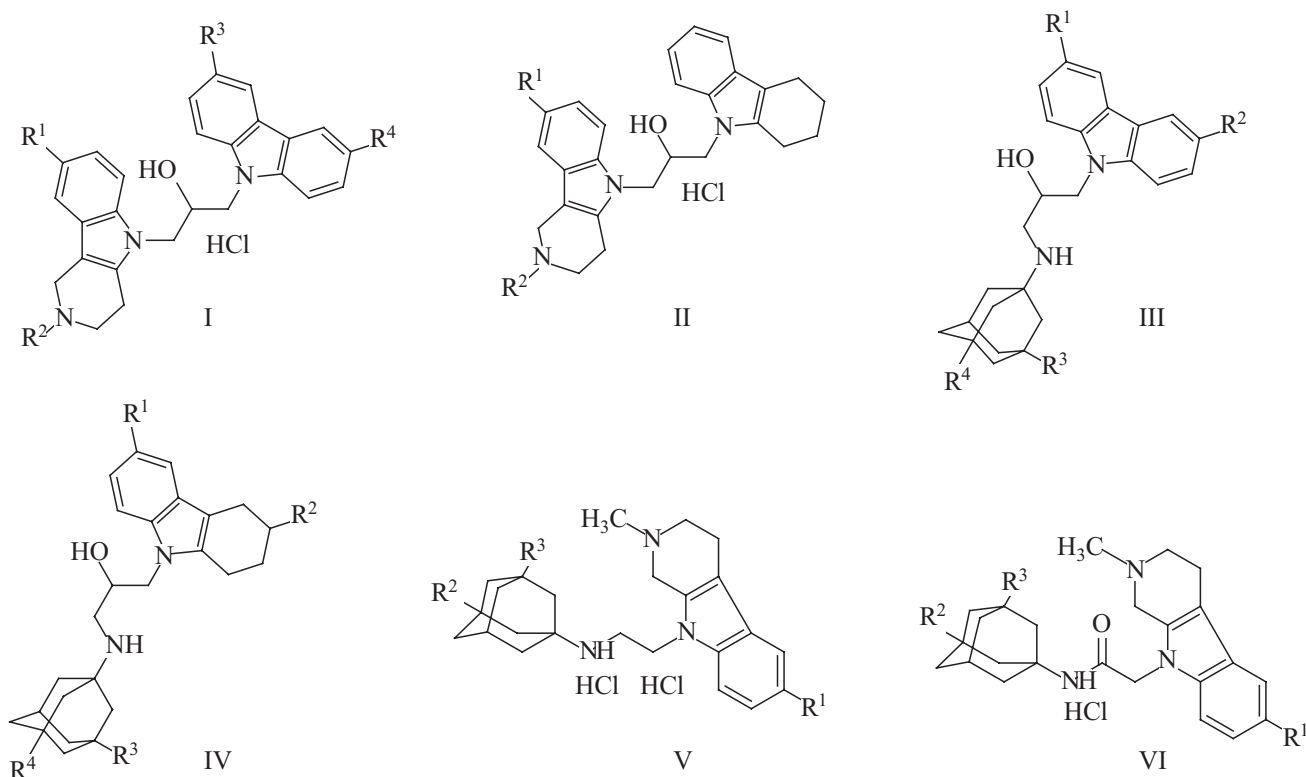
Для моделирования межклассового прогноза активности использовали ряд из 47 соединений, взятых из литературы [19, 20] (табл. 1, схема 1). Экспериментальные величины ингибиторной активности соединений по отношению к BChE (IC_{50} , мкМ./л) определены в одной лаборатории и по единой методике, что увеличивает точность прогноза. Минимальное и максимальное значение IC_{50} составляет 0.666 и 55.4 мкМ./л соответственно. Среднее значение равно 11.9 мкМ./л. Исследуемые соединения не только являются ингибиторами BChE, но и обладают ингибиторной активностью по отношению к AChE, т. е. относятся к классу мультитаргетных соединений. Учитывая, что величины IC_{50} (мкМ./л) по отношению к BChE таких известных лекарств для лечения болезни Альцгеймера, как донепезил и ривастигмин, составляют 4.94 и 1.71 [17] соответственно, можно констатировать, что диапазон изменения IC_{50} является вполне рабочим и перспективным для исследования. Для конструирования QSAR моделей величины IC_{50} преобразованы с использованием отрицательного логарифма $-\log(IC_{50})$. Следует отметить, что при этом интервал изменения активности составлял около двух порядков, а минимальное и максимальное значение были равны -1.744 и 0.177 соответственно.

Исследуемые соединения содержат по две фармакофорные группы, соединенные различными молекулярными спейсерами. Всего исследовано шесть рядов бифармакофорных соединений, включающих в себя конъюгаты γ -карболинов с производными карбазола (I) и тетрагидрокарбазола (II), конъюгаты

Таблица 1. Ингибирующая активность бифармакофорных соединений 1–47.

Соединение	Группа	R ¹	R ²	R ³	R ⁴	IC ₅₀ , мкМ./л	–log(IC ₅₀)
1	I	H	CH ₃	H	H	3.40	–0.531
2	I	H	C ₂ H ₅	H	H	2.19	–0.340
3	I	CH ₃	CH ₃	H	H	3.44	–0.537
4	I	CH ₃	C ₂ H ₅	H	H	6.11	–0.786
5	I	F	CH ₃	H	H	3.14	–0.497
6	I	F	C ₂ H ₅	H	H	4.05	–0.607
7	I	H	CH ₃	Cl	Cl	3.48	–0.542
8	I	CH ₃	CH ₃	Cl	Cl	16.4	–1.215
9	I	CH ₃	C ₂ H ₅	Cl	Cl	12.5	–1.097
10	I	F	CH ₃	Cl	Cl	1.93	–0.286
11	I	F	C ₂ H ₅	Cl	Cl	33.4	–1.524
12	I	H	CH ₃	Br	Br	2.69	–0.430
13	I	H	C ₂ H ₅	Br	Br	21.6	–1.334
14	I	CH ₃	C ₂ H ₅	Br	Br	3.16	–0.500
15	I	F	CH ₃	Br	Br	1.18	–0.072
16	II	H	CH ₃			1.06	–0.025
17	II	H	C ₂ H ₅			2.77	–0.442
18	II	CH ₃	CH ₃			3.90	–0.591
19	II	CH ₃	C ₂ H ₅			5.23	–0.719
20	II	F	CH ₃			5.03	–0.702
21	II	F	C ₂ H ₅			5.83	–0.766
22	III	H	H	H	H	15.9	–1.201
23	III	H	H	CH ₃	CH ₃	7.6	–0.881
24	III	Br	Br	H	H	20.7	–1.316
25	III	Br	Br	CH ₃	CH ₃	55.4	–1.744
26	III	Cl	Cl	H	H	23.7	–1.375
27	III	Cl	Cl	CH ₃	CH ₃	40.7	–1.610
28	IV	H	H	H	H	7.29	–0.863
29	IV	H	H	CH ₃	CH ₃	6.13	–0.787
30	IV	CH ₃	H	H	H	20.02	–1.301
31	IV	CH ₃	H	CH ₃	CH ₃	33.0	–1.519
32	IV	CH ₃	CH ₃	H	H	9.17	–0.962
33	IV	CH ₃	CH ₃	CH ₃	CH ₃	8.24	–0.916
34	IV	F	H	H	H	8.66	–0.938
35	IV	F	H	CH ₃	CH ₃	5.43	–0.735
36	V	H	CH ₃	CH ₃		0.666	0.177
37	V	CH ₃	CH ₃	CH ₃		5.07	–0.705
38	V	F	CH ₃	CH ₃		0.729	0.137
39	V	H	H	H		2.48	–0.394
40	V	CH ₃	H	H		9.05	–0.957
41	V	F	H	H		2.53	–0.403
42	VI	H	CH ₃	CH ₃		9.66	–0.985
43	VI	CH ₃	CH ₃	CH ₃		33.7	–1.528
44	VI	F	CH ₃	CH ₃		6.00	–0.778
45	VI	H	H	H		26.5	–1.423
46	VI	CH ₃	H	H		43.2	–1.635
47	VI	F	H	H		17.1	–1.233

Схема 1.



аминоадамтанов с производными карбазола (III) и тетрагидрокарбазола (IV) и конъюгаты аминоадамтанов с γ -карболинами (V, VI). Для увеличения статистической значимости результатов исследования введем некоторые допущения: (1) будем рассматривать карбазолы и тетрагидрокарбазолы в качестве одной фармакофорной группы; (2) будем считать, что все фармакофорные группы соединены с помощью одного и того же молекулярного спейсера. С учетом вышесказанного общее число фармакофорных групп, а также число их парных комбинаций будет равно 3. Для создания QSAR

моделей активности использовали три комбинации соединений: I+II, III+IV и V+VI, каждая из которых содержала две фармакофорные группы. Общая идея симуляции рационального межклассового прогноза активности бифармакофорных соединений состояла в конструировании QSAR модели с использованием обучающего ряда, включающего 1 или 2 комбинации соединений, и предсказании на ее основе активности для других комбинаций соединений.

Рассмотрим две стратегии межклассового прогноза активности с использованием трех комбинаций соединений: I+II, III+IV и V+VI (рис. 1, 2). В первой

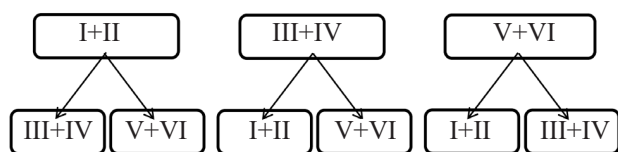


Рис. 1. Первая стратегия межклассового прогноза (МПА-1).

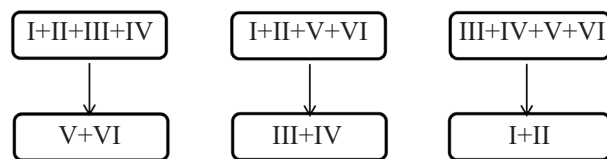


Рис. 2. Вторая стратегия межклассового прогноза (МПА-2).

стратегии (МПА-1) для создания рабочих моделей активности используются соединения, входящие в одну комбинацию соединений, а прогноз осуществляется для двух оставшихся комбинаций. Во второй стратегии (МПА-2) конструируются модели на основе двух комбинаций, а межклассовый прогноз проводится для третьей комбинации соединений.

В табл. 2–4 представлены статистические характеристики сконструированных QSAR моделей. Следует подчеркнуть, что большинство разработанных QSAR моделей удовлетворяют минимальным требованиям, которые к ним предъявляются [21], в частности, $R^2_{cv} > 0.5$, $R^2_p > 0.5$. Они также вполне соответствуют принципам OECD, связанным с

QSAR валидацией [22]. При этом в лучших сконструированных моделях величина R^2_{cv} достигает 0.8–0.9, что вполне сопоставимо с опубликованными в литературе данными по ингибиторам BChE. Так, например, в работе [23] для ряда из 68 соединений разработана QSAR модель, имеющая $R^2_{cv} = 0.873$, а в публикации [24] приведена модель, созданная на основе 46 ингибиторов (что соответствует размеру наших данных) BChE/AChE, имеющая величины $R^2 = 0.883$ и $R^2 = 0.881$ для обучающей и тестовой выборки соответственно.

В табл. 2 приведены статистические параметры моделей при использовании стратегии МПА-1. Следует отметить, что не все примененные алгоритмы

Таблица 2. Статистические характеристики QSAR моделей (МПА-1).

Алгоритм ^а	Дескрипторы ^б	<i>n</i>	R^2_{cv}	s_{cv}	FIT_{cv}	R^2_p	Комбинация ^в
МЛР	256; 534; 614; 636; 747	21	0.560	0.25	0.41	0.620	I+II(III+IV)
МЛР	256; 336; 618; 636; 717	21	0.541	0.25	0.38	0.614	I+II(V+VI)
СЛ	–	21	<0.5	–	–	–	I+II(III+IV)
СЛ	–	21	<0.5	–	–	–	I+II(V+VI)
МОВ	37; 149; 240; 717	21	0.635	0.23	0.75	0.794	I+II(III+IV)
МОВ	256; 534; 636; 747	21	0.544	0.25	0.52	0.519	I+II(V+VI)
ГП	37; 122; 156; 256; 336	21	0.611	0.23	0.51	0.536	I+II(III+IV)
ГП	256; 534; 636; 747; 777	21	0.575	0.24	0.44	0.535	I+II(V+VI)
МЛР	143; 634; 741	14	0.811	0.14	1.87	0.758	III+IV(V+VI)
МЛР	143; 634; 741	14	0.811	0.14	1.87	0.758	III+IV(I+II)
СЛ	–	14	<0.5	–	–	–	III+IV(V+VI)
СЛ	146; 214; 619	14	0.511	0.22	0.45	0.580	III+IV(I+II)
МОВ	143; 634; 741	14	0.818	0.13	1.95	0.617	III+IV(V+VI)
МОВ	143; 634; 741	14	0.818	0.13	1.95	0.617	III+IV(I+II)
ГП	143; 634; 741	14	0.798	0.14	1.72	0.669	III+IV(V+VI)
ГП	143; 634; 741	14	0.798	0.14	1.72	0.669	III+IV(I+II)
МЛР	125; 440; 534	12	0.677	0.33	0.80	0.664	V+VI(I+II)
МЛР	133; 208; 240	12	0.895	0.19	3.26	0.788	V+VI(III+IV)
СЛ	–	12	<0.5	–	–	–	V+VI(I+II)
СЛ	–	12	<0.5	–	–	–	V+VI(III+IV)
МОВ	240; 313; 624	12	0.769	0.28	1.27	0.717	V+VI(I+II)
МОВ	125; 133; 208	12	0.885	0.20	2.94	0.747	V+VI(III+IV)
ГП	125; 440; 534	12	0.615	0.36	0.61	0.589	V+VI(I+II)
ГП	125; 133; 208	12	0.903	0.18	3.53	0.679	V+VI(III+IV)

^а МЛР – множественная линейная регрессия [25], СЛ – случайный лес [26], МОВ – машина опорных векторов [27], ГП – гауссовский процесс [28].

^б Дескрипторы (тип): 1÷100 (AAE); 101÷200 (AAF); 201÷300 (Q[–]); 301÷400 (Q⁺); 401÷500 (Q⁺⁺); 501÷600 (DAE); 601÷700 (DAF); 701÷800 (VDW).

^в Комбинация X1+X2(Y1+Y2): X – группы соединений, использованных для создания QSAR модели; Y – группы соединений, использованных для межклассового прогноза.

Таблица 3. Статистические характеристики QSAR моделей (МПА-2).

Алгоритм ^а	Дескрипторы ^б	<i>n</i>	R^2_{cv}	s_{cv}	FIT_{cv}	R^2_p	Комбинация ^в
МЛР	62; 311; 320; 444; 528	35	0.681	0.24	1.03	0.695	I, II, III, IV(V,VI)
	—	35	—	—	—	<0.5	I, II, III, IV(V,VI)
случайного леса	—	—	—	—	—	—	—
МОВ	311; 320; 528; 722	35	0.638	0.26	1.04	0.511	I, II, III, IV(V,VI)
ГП	152; 311; 320; 528; 722	35	0.652	0.25	0.90	0.620	I, II, III, IV(V,VI)
МЛР	55; 240; 247; 408; 768	33	0.595	0.30	0.68	0.631	I, II, V, VI(III,IV)
СЛ	—	33	<0.5	—	—	—	I, II, V, VI(III,IV)
МОВ	152; 243; 617; 768	33	0.549	0.31	0.70	0.604	I, II, V, VI(III,IV)
ГП	240; 242; 408; 416; 768	33	0.612	0.29	0.74	0.581	I, II, V, VI(III,IV)
МЛР	36; 117; 320; 407; 440	26	0.851	0.19	2.24	0.801	III, IV, V, VI(I,II)
СЛ	10; 247; 318; 407; 639	26	0.653	0.29	0.74	0.567	III, IV, V, VI(I,II)
МОВ	36; 109; 117; 348	26	0.856	0.19	2.96	0.691	III, IV, V, VI(I,II)
ГП	36; 109; 117; 348; 407	26	0.888	0.16	3.10	0.735	III, IV, V, VI(I,II)

^а МЛР – множественная линейная регрессия [25], СЛ – случайный лес [26], МОВ – машина опорных векторов [27], ГП – гауссовский процесс [28].

^б Дескрипторы (тип): 1÷100 (AAE); 101÷200 (AAF); 201÷300 (Q^{-}); 301÷400 (Q^{+}); 401÷500 (Q^{++}); 501÷600 (DAE); 601÷700 (DAF); 701÷800 (VDW).

^в Комбинация X1,X2,X3,X4(Y1,Y2): X – группы соединений, использованных для создания QSAR модели; Y – группы соединений, использованных для межклассового прогноза.

обеспечивают получение удовлетворительных QSAR моделей. В частности, метод случайного леса показывает худший результат, поэтому он не использовался для межклассового прогноза активности. Для выявления значимости дескрипторов был проведен расчет частоты их появления при проведении МПА-1 (рис. 3). При этом можно отметить, что наиболее часто в качестве независимых переменных в QSAR моделях фигурируют AAF и DAF дескрипторы. Это может свидетельствовать о том, что при использовании стратегии МПА-1 наиболее значимыми внутримолекулярными парными атомными взаимодействиями являются взаимодействия типа Н-акцептор–Н-акцептор и Н-донор–Н-акцептор.

Среди AAF дескрипторов наиболее часто встречается дескриптор 143, при этом он изменяется в пределах от 0.000 до 0.122. Минимальное и максимальное значение наиболее значимого DAF дескриптора 634 составляет 0.000 и 0.674 соответственно.

Результаты исследования стратегии МПА-2 приведены в табл. 3 и на рис. 4. Обращает на себя внимание тот факт, что, как и в случае МПА-1, наиболее слабым алгоритмом машинного обучения является метод случайного леса. QSAR модели, полученные с помощью других алгоритмов, имеют сопоставимые статистические характеристики и могут быть применены для межклассового прогноза. Молекулярные дескрипторы, использованные для конструирования

Таблица 4. Статистические характеристики QSAR моделей (МПА-3).

Алгоритм ^а	Дескрипторы ^б	<i>n</i>	R^2_{cv}	s_{cv}	FIT_{cv}	R^2_p
МЛР	125; 142; 260; 351; 517	24	0.756	0.21	1.14	0.749
МОВ	324; 351; 443; 517	24	0.767	0.21	1.57	0.898
ГП	57; 125; 351; 517; 626	24	0.754	0.21	1.13	0.675

^а МЛР – множественная линейная регрессия [25], СЛ – случайный лес [26], МОВ – машина опорных векторов [27], ГП – гауссовский процесс [28].

^б Дескрипторы (тип): 1÷100 (AAE); 101÷200 (AAF); 201÷300 (Q^{-}); 301÷400 (Q^{+}); 401÷500 (Q^{++}); 501÷600 (DAE); 601÷700 (DAF); 701÷800 (VDW).

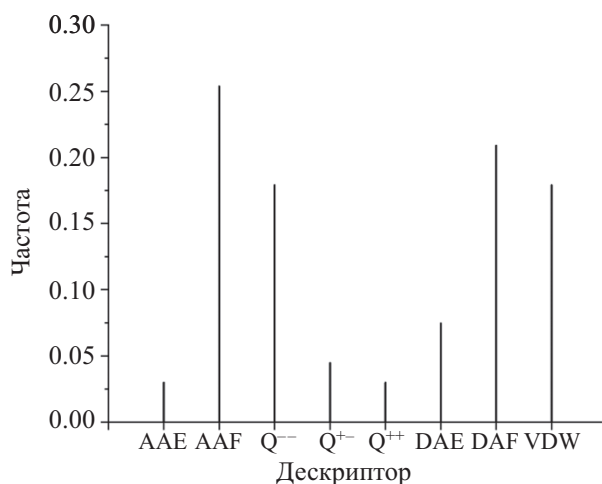


Рис. 3. Частота появления групповых дескрипторов в QSAR моделях (МПА-1).

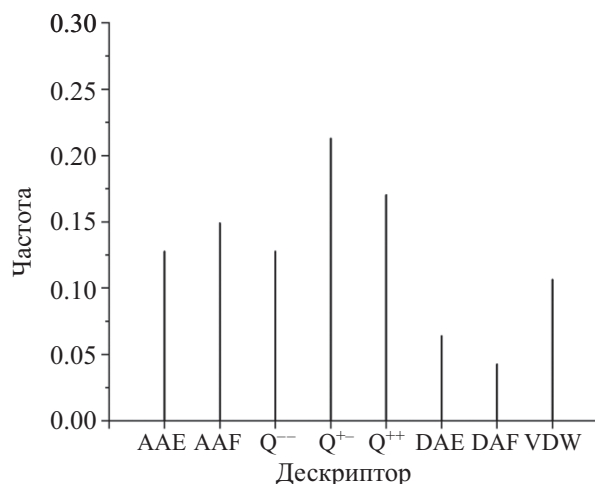


Рис. 4. Частота появления групповых дескрипторов в QSAR моделях (МПА-2).

этих моделей, имеют по сравнению с МПА-1 другое распределение. В частности, наиболее часто встречаются дескрипторы двух типов: Q^{+-} и Q^{++} . Вероятно, это указывает на более значимую роль внутримолекулярных электростатических взаимодействий при проведении стратегии МПА-2. При этом минимальные значения ведущих дескрипторов 320 (тип Q^{+-}) и 407 (тип Q^{++}) равны 0.008 и 0.000, а максимальные величины составляют 0.023 и 0.0001 соответственно.

Для улучшения статистической значимости результатов межклассового прогноза активности мы применили консенсусный подход. При этом в качестве предсказанной активности анализируемого соединения рассматривали среднее арифметическое из рассчитанных значений активности на основе всех моделей соответствующей стратегии прогноза. Необходимым условием включения соединения в прогнозный список было вхождение в область применимости всех QSAR моделей. Результаты такого подхода отражены на рис. 5. Среднеквадратичное отклонение между экспериментальными и прогнозируемыми величинами активности (s_{pr}) составило 0.68 и 0.49 для МПА-1 и МПА-2 соответственно. Полученные величины в 2–3 раза превышают величины s_{cv} , но вполне соответствуют приводимым в литературе межлабораторным ошибкам определения IC_{50} [29]. Только для трех соединений (1, 2 и 28) из 17 величина $\log(IC_{50})$ рассчитана с использованием обеих схем межклассового прогноза. Наибольшие

отклонения между экспериментальными и прогнозируемыми величинами ингибиторной активности, приближающиеся к $2s_{pr}$, наблюдались для соединений 1 и 2 (МПА-1) и для соединений 2, 24 и 45 (МПА-2). В целом можно отметить, что в отношении диапазона прогнозируемых активностей стратегии МПА-1 и МПА-2 дополняют друг друга.

Для сравнительного анализа сконструированы QSAR модели (табл. 4) и осуществлен межклассовый прогноз (МПА-3) (рис. 5) со случайным

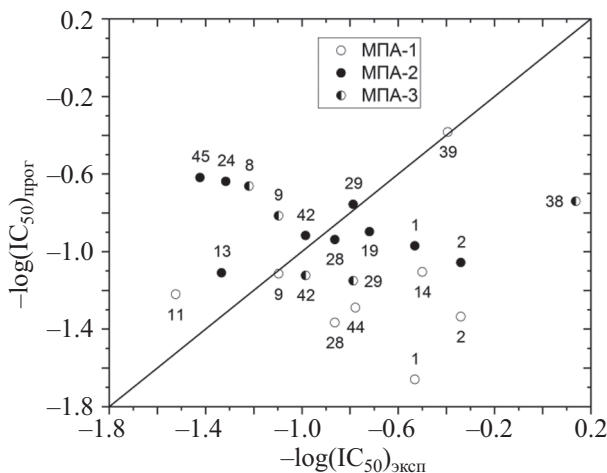


Рис. 5. Зависимость между экспериментальными и прогнозируемыми значениями активности соединений.

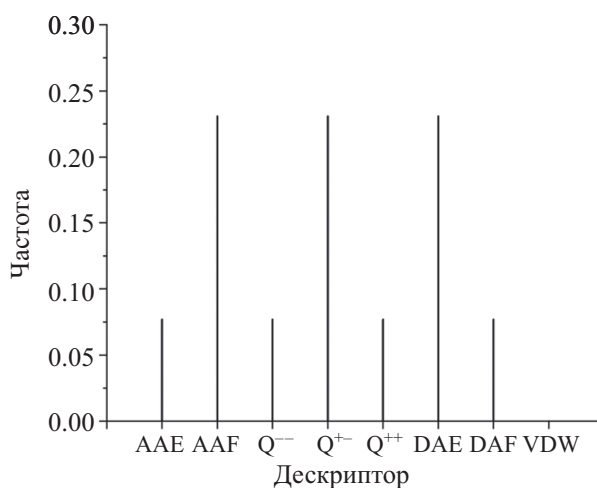


Рис. 6. Частота появления групповых дескрипторов в QSAR моделях (МПА-3).

выбором соединений. Размер обучающей выборки с учетом того, что в стратегиях МПА-1 и МПА-2 обучение проводили с использованием от 12 до 35 соединений, составил 24 соединения (среднее значение). Полученные модели имеют удовлетворительные статистические характеристики и могут быть использованы для межклассового прогноза активности. Распределение ведущих дескрипторов в полученных QSAR моделях (рис. 6) отличается от рассмотренных ранее частот появления в стратегиях МПА-1 и МПА-2 и представляет собой комбинацию из дескрипторов AAF, Q⁺ и DAE. В случае МПА-3 одно соединение (**38**) имеет значительную разницу между величинами экспериментальной и предсказанной активности. В целом для всех пяти соединений $s_{\text{пр}} = 0.57$, что вполне соответствует данным, полученным при проведении МПА-1 и МПА-2. Однако количество соединений, для которых была рассчитана активность, оказывается меньше: пять (МПА-3) против восьми (МПА-1) и девяти (МПА-2). При этом из этих пяти соединений для трех (**9**, **29** и **42**) предсказаны активности в стратегиях МПА-1 и МПА-2. Также следует отметить, что при использовании этих стратегий величины экспериментальных и прогнозируемых активностей соединений близки между собой: соединения **9** (–1.097, –1.114), **29** (–0.787, –0.756) и **42** (–0.985, –0.917).

Корреляция между экспериментальными и прогнозируемыми активностями для всех соединений, представленных на рис. 5, оказывается ожидаемо низкой ($n = 22$, $R^2 = 0.007$, $s = 0.42$). Очевидно, что это связано как с природой МПА, так и с теми допущениями, которые были положены в основу формирования трех исследуемых комбинаций соединений. Удаление из выборки соединений, у которых разность между $\log(\text{IC}_{50})_{\text{эксп}}$ и $\log(\text{IC}_{50})_{\text{прог}}$ по абсолютной величине превышала 0.5 логарифмических единиц, значительно улучшало корреляцию ($n = 12$, $R^2 = 0.461$, $s = 0.25$). Пять из десяти удаленных соединений принадлежало комбинации I+II, два относилось к комбинации III+IV и три входило в состав комбинации V+VI. Вероятно, наибольший вклад в ошибку прогноза вносит объединение производных карбазола и тетрагидрокарбазола в одну группу. Менее значим вклад, связанный с использованием различных спейсеров для формирования конъюгатов.

ВЫВОДЫ

В результате проведенного исследования установлено, что рациональные стратегии межклассового прогноза (МПА-1 и МПА-2) в отношении диапазона прогнозируемых активностей взаимно дополняют друг друга. При использовании стратегии случайного прогноза (МПА-3) три из пяти предсказанных соединений также обнаруживаются в прогнозах МПА-1 и МПА-2. Эффективность стратегий может быть выражена в виде ряда: МПА-2 > МПА-3 > МПА-1 в соответствии с величинами стандартных отклонений или в виде ряда МПА-2 > МПА-1 > МПА-3 при использовании в качестве критерия количества предсказанных соединений. Стратегии МПА, исследованные при анализе бифармакофорных ингибиторов BChE на основе замены фармакофорной группы, а также полученные при этом оценки средне-квадратичных отклонений межклассового прогноза биоактивности могут быть использованы при поиске новых бифармакофорных физиологически активных веществ.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

QSAR моделирование проводили на основе *in-house* компьютерных программ с использованием различных алгоритмов машинного обучения, включая множественную линейную регрессию (МЛР) [25],

случайный лес (СЛ) [26], машину опорных векторов (МОВ) [27] и гауссовский процесс (ГП) [28]. С учетом малых размеров обучающих выборок применяли только внутреннее тестирование на основе скользящего контроля с выбором по 5 (10 итераций). В качестве статистических характеристик моделей использовали: n – число соединений; R^2 – квадрат коэффициента линейной корреляции; R^2_{cv} – квадрат коэффициента линейной корреляции в условиях скользящего контроля; s – среднее квадратичное отклонение; s_{cv} – среднее квадратичное отклонение в условиях скользящего контроля; s_{pr} – среднее квадратичное отклонение межклассового прогноза; R^2_p – рандомизационный параметр [30] (100 итераций), FIT_{cv} – модифицированный критерий Фишера [31] в условиях скользящего контроля. Для оценки области применимости (ОП) модели использовали три величины: X-ОП (интервал изменения дескрипторов), Y-ОП (интервал изменения активности) и S-ОП (структурное сходство между молекулами на основе индекса Танимото (T_c) [32], рассчитанного на основе дескрипторных 10-битовых «отпечатков пальцев» с пороговым значением $T_c = 0.7$ и первым ближайшим соседом). Расчет T_c проводили исключительно на основе дескрипторов QSAR модели. При проведении межклассового прогноза активности рассматривали только молекулы, попадающие в ОП. Минимальная, максимальная и средняя величина индексов Танимото, рассчитанных с использованием всех 800 дескрипторов и первого ближайшего соседа, для комбинаций соединений (I+II)/(III+IV+V+VI), (III+IV)/(I+II+V+VI) и (V+VI)/(I+II+III+IV) составила (0.395, 0.496, 0.442), (0.440, 0.595, 0.519) и (0.499, 0.568, 0.527) соответственно.

Для описания структуры молекул, учитывая положительный опыт применения для мультитаргетных соединений [33], использовали интегралы интенсивностей спектров межатомных внутримолекулярных взаимодействий в диапазоне от 0 до 20 ангстрем с шагом 0.2 ангстрема, что приводило к появлению 100 дескрипторов, которые рассчитывали с помощью программы MOLTRA [34]. Расчет проводили с учетом 8 типов парных внутримолекулярных атомных взаимодействий: Н-акцептор–Н-акцептор (AAE, AAF), Н-донор–Н-акцептор (DAE, DAF), отрицательно заряженных атомов (Q^{-}), положительно заряженных атомов (Q^{+}), положительно и отрицательно заряженных атомов (Q^{+-}) и ван-дер-ваальсовых взаи-

модействий (VDW). Таким образом, для описания пространственной структуры каждого соединения с помощью спектров межатомных внутримолекулярных взаимодействий использовали $8 \times 100 = 800$ дескрипторов. После процедуры анализа коэффициентов корреляции дескрипторов с активностью при пороге 0.5 их число сокращалось до 15–45. Выбор дескрипторов для QSAR моделирования проводили путем перебора всех возможных комбинаций из 1–5 дескрипторов с учетом того, чтобы в модели на каждую переменную приходилось не менее четырех соединений обучающей выборки. Отбор лучших моделей осуществляли на основе максимального значения FIT_{cv} при условии, что $R^2_{cv} > 0.5$, $R^2_p > 0.5$.

ИНФОРМАЦИЯ ОБ АВТОРАХ

Григорьев Вениамин Юрьевич, ORCID: <https://orcid.org/0000-0002-5288-3242>

Раздольский Александр Николаевич, ORCID: <https://orcid.org/0000-0002-3389-4659>

Казаченко Владимир Павлович, ORCID: <https://orcid.org/0000-0003-1424-1895>

ФИНАНСОВАЯ ПОДДЕРЖКА

Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации в рамках государственного задания Института физиологически активных веществ РАН 2024 года (тема № FFSG-2024-0019).

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Schlander M., Hernandez-Villafuerte K., Cheng C.Y., Mestre-Ferrandiz J., Baumann M. // *Pharmacoeconomics*. 2021. Vol. 39. P. 1243. doi 10.1007/s40273-021-01065-y
2. Sadybekov A.V., Katritch V. // *Nature*. 2023. Vol. 616. P. 673. doi 10.1038/s41586-023-05905-z
3. Doytchinova I. // *Molecules*. 2022. Vol. 27. P. 1496. doi 10.3390/molecules27051496
4. Niazi S.K., Mariam Z. // *Pharmaceuticals*. 2024. Vol. 17. P. 22. doi 10.3390/ph17010022
5. Baig M.H., Ahmad K., Roy S., Ashraf J.M., Adil M., Siddiqui M.H., Khan S., Kamal M.A., Provaznik I., Choi

- I. // Curr. Pharm. Des.* 2016. Vol. 22. P. 572. doi 10.2174/1381612822666151125000550
6. *Зефирова О.Н., Зефиров Н.С. // Вестн. Московск. унив. Сер. 2. Химия.* 2000. Т. 41. С. 103.
7. *Hu Y., Stumpfe D., Bajorath J. // J. Med. Chem.* 2017. Vol. 60. P. 1238. doi 10.1021/acs.jmedchem.6b01437
8. *Stojanović L., Popović M., Tijanić N., Rakočević G., Kalinić M. // J. Chem. Inf. Model.* 2020. Vol. 60. P. 4629. doi 10.1021/acs.jcim.0c00622
9. *Acharya A., Yadav M., Nagpure M., Kumaresan S., Guchhait S.K. // Drug Discov. Today.* 2024. Vol. 29. Article no. 103845. doi 10.1016/j.drudis.2023.103845
10. *Wang Y., Jia S., Wang F., Jiang R., Yin X., Wang S., Jin R., Guo H., Tang Y., Wang Y. // Int. J. Mol. Sci.* 2024. Vol. 25. Article no. 7434. doi 10.3390/ijms25137434
11. *Floresta G., Rescifina A., Marrazzo A., Dichiaro M., Pistarà V., Pittalà V., Prezzavento O., Amata E. // Eur. J. Med. Chem.* 2017. Vol. 139. P. 884. doi 10.1016/j.ejmech.2017.08.053
12. *Škuta C., Cortés-Ciriano I., Dehaen W., Kříž P., van Westen G.J.P., Tetko I.V., Bender A., Svozil D. // J. Cheminform.* 2020. Vol. 12. P. 39. doi 10.1186/s13321-020-00443-6
13. *Zheng S., Lei Z., Ai H., Chen H., Deng D., Yang Y. // J. Cheminform.* 2021. Vol. 13. P. 87. doi 10.1186/s13321-021-00565-5
14. *Ryszkiewicz P., Malinowska B., Schlicker E. // Pharmacol. Rep.* 2023. Vol. 75. P. 755. doi 10.1007/s43440-023-00501-4
15. *Sánchez-Tejeda J.F., Sánchez-Ruiz J.F., Salazar J.R., Loza-Mejía M.A. // Front. Chem.* 2020. Vol. 8. P. 176. doi 10.3389/fchem.2020.00176
16. *Albertini C., Salerno A., de Sena Murteira Pinheiro P., Bolognesi M.L. // Med. Res. Rev.* 2021. Vol. 41. P. 2606. doi 10.1002/med.21699
17. *Zhou S., Huang G. // Biomed. Pharmacother.* 2022. Vol. 146. Article no. 112556. doi 10.1016/j.biopha.2021.112556
18. *Greig N.H., Lahiri D.K., Sambamurti K. // Int. Psychogeriatr.* 2002. Vol. 14. P. 77. doi 10.1017/s1041610203008676
19. *Makhaeva G.F., Shevtsova E.F., Boltneva N.P., Lushchekina S.V., Kovaleva N.V., Rudakova E.V., Bachurin S.O., Rudy J. Richardson R.J. // Chem. Biol. Interact.* 2019. Vol. 308. P. 224. doi 10.1016/j.cbi.2019.05.020
20. *Bachurin S.O., Makhaeva G.F., Shevtsova E.F., Aksinenko A.Y., Grigoriev V.V., Shevtsov P.N., Goreva T.V., Epishina T.A., Kovaleva N.V., Pushkareva E.A., Boltneva N.P., Lushchekina S.V., Gabrelyan A.V., Zamoyski V.L., Dubova L.G., Rudakova E.V., Fisenko V.P., Bovina E.V., Richardson R.J. // Molecules.* 2021. Vol. 26. P. 5527. doi 10.3390/molecules26185527
21. *Kiralj R., Ferreira M.M.C. // J. Braz. Chem. Soc.* 2009. Vol. 20. P. 770. doi 10.1590/S0103-50532009000400021
22. *Tropsha A., Gramatica P., Gombar V.K. // QSAR Comb. Sci.* 2003. Vol. 22. P. 69. doi 10.1002/qsar.200390007
23. *Kumar S., Manoharan A., Jayalakshmi J., Abdelgawad M.A., Mahdi W.A., Alshehri S., Ghoneim M.M., Papachen L.K., Zachariah S.M., Aneesh T.P., Mathew B. // RSC Adv.* 2023. Vol. 13. P. 9513. doi 10.1039/d3ra00526g
24. *Pang X., Fu H., Yang S., Wang L., Liu A.-L., Wu S., Du G.-H. // Molecules.* 2017. Vol. 22. P. 1254. doi 10.3390/molecules22081254
25. Fortran Numerical Library. <https://developer.nvidia.com/imsf-fortran-numerical-library?display=default>
26. Random Forest. http://www.stat.berkeley.edu/~breiman/RandomForests/cc_examples/prog.f
27. *Suykens J.A.K., Vandewalle J. // Neural Process. Lett.* 1999. Vol. 9. P. 293. doi 10.1023/A:1018628609742
28. Gaussian Processes for Machine Learning. <http://gaussian-process.org/gpml/>
29. *Landrum G.A., Riniker S. // J. Chem. Inf. Model.* 2024. Vol. 64. P. 1560. doi 10.1021/acs.jcim.4c00049
30. *Mitra I., Saha A., Roy K. // Mol. Simul.* 2010. Vol. 36. P. 1067. doi 10.1080/08927022.2010.503326
31. *Kubinyi H. // Quant. Struct. Act. Relat.* 1994. Vol. 13. P. 285. doi 10.1002/qsar.19940130306
32. *Willett P., Barnard J.M., Downs G.M. // J. Chem. Inf. Comput. Sci.* 1998. Vol. 38. P. 983. doi 10.1021/ci9800211
33. *Раздольский А.Н., Казаченко В.П., Страхова Н.Н., Григорьев В.Ю. // Современные наукоемкие технологии.* 2023. Вып. 10. С. 63. doi 10.17513/snt.39792
34. *Trepalin S.V., Razdolskii A.N., Raevskii O.A. // Pharm. Chem. J.* 2000. Vol. 34. P. 650. doi 10.1023/A:1010499601434

Investigation of Strategies for the Interclass Prediction of the Activity of Biparmacophore Butyrylcholinesterase Inhibitors Based on QSAR Modeling

V. Y. Grigorev^{1,*}, A. N. Razdolsky¹, and V. P. Kazachenko¹

¹ *Institute of Physiologically Active Compounds, Federal Research Center of Problems of Chemical Physics and Medicinal Chemistry of the Russian Academy of Sciences, Chernogolovka, 142432 Russia*

**e-mail: beng@ipac.ac.ru*

Received October 4, 2024; revised December 12, 2024; accepted December 21, 2024

Three schemes of interclass prediction of the activity of a number of biparmacophoric butyrylcholinesterase inhibitors were studied using QSAR modeling. Using machine learning methods (multiple linear regression, random forest, support vector machine and Gaussian process), QSAR models with satisfactory statistical characteristics were constructed. Based on them, rational and random interclass prediction schemes were studied. It was found that these schemes complement each other and their relative efficiency was assessed.

Keywords: interclass prediction, butyrylcholinesterase, QSAR