

Молекулярно-биологические проблемы создания лекарственных средств и изучение механизма их действия

© Коллектив авторов, 2013

А. С. Кабанкин, Л. А. Радкевич

ПРИМЕНЕНИЕ СТРАТЕГИИ КОЛЛЕКТИВНОГО РАСПОЗНАВАНИЯ ПРИ ОЦЕНКЕ ГЕПАТОПРОТЕКТОРНОЙ АКТИВНОСТИ, ПОВЫШАЮЩЕЙ РЕПАРАТИВНЫЙ ПОТЕНЦИАЛ ПЕЧЕНИ, ДЛЯ РАЗНОРОДНЫХ ХИМИЧЕСКИХ СОЕДИНЕНИЙ

Федеральное государственное бюджетное учреждение науки Центр теоретических проблем физико-химической фармакологии Российской академии наук, Москва, Россия

Для разнородных химических соединений исследована связь между молекулярными характеристиками и гепатопротекторной активностью, повышающей репаративный потенциал печени. Для обучающей выборки, составленной из 44 соединений (23 — активные, 21 — неактивное), методом ближайшего соседа с использованием геометрических дескрипторов получено множество комбинаций дескрипторов, корректно классифицирующих 79–91 % соединений обучающей выборки. Показано, что использование процедуры простого голосования результатов, полученных для различных комбинаций дескрипторов, позволяет построить классификатор, корректно классифицирующий все соединения обучающей выборки. Методом перекрестного контроля с исключением групп соединений из обучающей выборки доказана эффективность предлагаемого подхода для предсказания активности нетестированных соединений.

Ключевые слова: гепатопротекторная активность; репаративный потенциал печени; предсказание активности; геометрические дескрипторы.

В предыдущих сообщениях [1–5] проанализирована зависимость гепатопротекторной детоксицирующей активности производных адамантана и индола от разнообразных дескрипторов, отображающих структуру молекул. Указанная активность, выраженная с помощью индекса выживаемости (ИВ), была оценена на первом этапе трехэтапной системы первичного скрининга, разработанной с использованием традиционной экспериментальной модели тетрахлометанового гепатита [6]. На втором этапе скрининга оценивалась способность соединений, отобранных на первом этапе, стимулировать репаративную активность в печени крыс на модели субтоксического поражения CCl_4 по данным морфометрии пораженной гепатотоксином печени животных. Результаты оценки сохранности гистоструктуры паренхимы печени у контрольных и опытных животных были выражены в виде индекса поражения печени (ИПП) в баллах от 0 до 2. Активными считали химические соединения, которые снижали ИПП до 0,5 баллов и ниже. В процессе экспериментального скрининга обнаружено, что прямой корреляции между детоксицирующей активностью, снижающей летальность, и активностью, стимулирующей репаративные процессы в паренхиме, у исследованных химических соединений нет. Поэтому изучение зависимости гепатопротекторной активности, повышающей репаративный потенциал печени, от молекулярной структуры соединений представляет теоретический и практический интерес.

Обучающая выборка содержала 44 соединения (23 активных в отношении повышения репаративного потенциала печени и 21 неактивное) из различных химических классов (производные адамантана, индола и др.). В табл. 1 представлены названия исследованных соединений. Поскольку рассматриваемые соединения обладали достаточно вариабельной пространственной структурой, для их описания использовали геометрические дескрипторы ($M = 35$) [7], вычисленные с помощью программы DRAGON [8]. Для расчета геометрических дескрипторов молекулярные структуры были предварительно оптимизированы полумпирическим методом самосогласованного поля в валентном приближении AM1.

На основе указанных молекулярных дескрипторов проведен поиск классификационных правил, позволяющих с наименьшей ошибкой классифицировать по активности соединения из обучающей выборки. На начальном этапе исследований использовали линейный дискриминантный анализ (ЛДА) и метод ближайшего соседа (БС). Индивидуальная информативность геометрических дескрипторов при распознавании обучающей выборки в обоих методах оказалась примерно одинаковой (максимальное распознавание 64–70 %), однако при использовании пошаговых процедур для определения наилучшей комбинации дескрипторов с максимальной распознавательной способностью результаты для обоих методов существенно различаются. В методе ЛДА наилучшее распознавание (82 %)

Состав обучающей выборки соединений

№	Название соединения
	Активные соединения (ИПП ≤ 0,5)
1	1-(1-адамантил)-3-(нафтил-2-амино)-3-(<i>o</i> -фторфенил)пропан-1-он
2	1-тиоцианатоацетиламиноадамтан
3	3-(2-аминоэтил)-1-бензил-3Н-индол-2-он
4	4-амино-2-(диметиламино)хинолин
5	N ² -(4-бром-2-цианофенил)-N ¹ ,N ¹ -диметилацетамидин
6	1-бензил-2',2'-диметил-2-оксоиндолин-3-спиро-3'-пирролидин
7	4-амино-2-диметиламино-6,8-диметилхинолин
8	6-амино-1-метил-2,2,4,5,8,9-гексагидро-1Н,7Н-азелино[2,3- <i>b</i>]циклопента[<i>e</i>]пиридин
9	2-(3-метил-1-адамантилимино)-4-хлорметил-1,3-окситиолан
10	9-амино-2,3,5,6,7,8-гексагидро-1Н-циклопента[<i>b</i>]хинолин
11	1-(1-адамантил)-3-(4-пиридин)-2-пропен-1-он
12	1-бензил-2-[(<i>m</i> -нитробензилиден)амино]-3-метилиндол
13	1-нитро-4-(N-ацетил-Е-аминогексаноил)аминобензол
14	9-амино-3,3-диметил-3,4-дигидроакридин
15	1-изоникотиноиламиноадамтан
16	4-метил-4-антипиритиосемикарбазид
17	Диметилди(<i>n</i> -фениламинофеноксид)силан
18	1-(3-N-морфаминопропионил)аминоадамтан
19	1-бензил-2',2',5-триметил-2-оксоиндолил-3-спиро-3'-пирролидин
20	1-(1-адамантил)-3-(хинолил-6-амино)-3-фенилпропан-1-он
21	2-(1-адамантилимино)-4-бромметил-1,3-оксатиолан
22	2-(5,3-диметил-1-адамантилимино)-4-хлорметил-1,3-окситиолан
23	2,2,6,6-тетраметил-4,4-бис(пропилтио)пиперидин
	Неактивные соединения (ИПП > 0,5)
24	9-амино-1,2,3,4-тетрагидроакридин
25	N', N'-(оксидиэтилен)тиоуред-3-карбоксихумарин
26	2-морфолино-5-фенил-6Н-1,3,4-тиадиазин
27	5-аминонафталиноктилсульфамид
28	1-(1-адамантил)-3-анилино-3-(<i>n</i> -бромфенил)-пропан-1-он
29	3,3'-(4,4'-бифенилилендисульфони)дипропановая кислота
30	N'-аллиптиоуред-3-карбоксихумарин
31	1-(1-адамантил)-3-(<i>n</i> -фторанилино)-3-(<i>n</i> -фторфенил)пропан-1-он
32	5-оксо-L-пирролидин-2-карбоновая кислота
33	2,2'-(этилендитио)дипирролин-1
34	2-(додецилсульфинил)пропан-1,3-диол
35	N-метил-N-гомовератриламиотиокарбонилтиогликолевая кислота
36	9-аминометилен-4-азафлуорен
37	5-аминонафталиндиизобутилсульфамид
38	1-бензил-2-[(<i>o</i> -гидроксibenзалиден)амино]-3-метилиндол
39	4,4-дифторадамтан-1-ол
40	3-амино-6-фенил-4-трифторметил-2-хлорпиридин
41	N,N'-(сукцинил)бис(2-пирролидинтион)
42	10-(N-ацетил-γ-аминобутирил)-2-хлорфенотиазин
43	N-ацетилглутаминовая кислота
44	5-амино-7-бром-1-метил-1,2,3,4-тетрагидробензо[<i>b</i>]-1,8-нафтридин

достигается для комбинаций, содержащих 7 – 8 дескрипторов, и далее, при добавлении других дескрипторов, не меняется. В методе БС максимальное распознавание (91 %) достигается для комбинаций, включающих 4 – 5 дескрипторов, а далее, при добавлении других дескрипторов, степень распознавания обучающей выборки, как правило, постепенно уменьшается. Метод БС имеет определенные преимущества перед методом ЛДА в том отношении, что, во-первых, не требуется, чтобы распределение вычисленных значе-

ний дескрипторов для рассматриваемых соединений подчинялось нормальному закону; во-вторых, не требуется, чтобы соединения, относящиеся к различным классам активности, были линейно разделимы. В этой связи в последующих исследованиях использовали только метод БС.

В методе БС, являющемся самым простым алгоритмом классификации, классифицируемое соединение относится к тому классу активности, к которому принадлежит ближайшее в используемом пространстве

Индивидуальная распознающая способность наиболее информативных дескрипторов (объем обучающей выборки — 44 соединения)

Обозначение в программе DRAGON [7]	Физический смысл [6]	Расознающая способность, %	Число неверно классифицируемых соединений
<i>W3D</i>	Число Винера на основе геометрической матрицы расстояний	68,2	14
<i>G1</i>	Гравитационный индекс	70,5	13
<i>G2</i>	Гравитационный индекс	65,9	15
<i>G(N...O)</i>	Сумма геометрических расстояний между атомами N и O	65,9	15
<i>ADDD</i>	Среднее значение отношения геометрического расстояния к топологическому из соответствующих матриц	61,4	17
<i>DELS</i>	Электротопологическая вариация для молекулы в целом	63,6	16
<i>FDI</i>	Индекс, характеризующий степень фолдинга	61,4	17

дескрипторов соединения из обучающей выборки. Для исключения доминантного влияния дескрипторов с большими числовыми значениями все исходные данные были стандартизированы. В качестве меры близости в методе БС обычно используют евклидову метрику. В настоящей работе кроме евклидовой метрики в качестве меры близости молекулярных структур использовали также коэффициент Танимото для непрерывных переменных [9]:

$$S_{AB} = (\sum x_{jA} \cdot x_{jB}) / [\sum (x_{jA})^2 + \sum (x_{jB})^2 - \sum (x_{jA}x_{jB})],$$

где x_{jA} и x_{jB} — значения j -го дескриптора для молекул A и B соответственно, суммирование производится по совокупности выбранных дескрипторов.

В целом, результаты применения обеих метрик близки, хотя и наблюдаются некоторые различия. Дальнейшее обсуждение основано на результатах, полученных с использованием указанного коэффициента Танимото.

В табл. 2 представлены геометрические дескрипторы с максимальной индивидуальной распознающей способностью. Эти дескрипторы далее были использованы в качестве стартовых в различных пошаговых процедурах для выявления комбинаций дескрипторов с максимальным распознаванием обучающей выборки. Получено большое количество различных комбинаций, содержащих 3–7 дескрипторов. Как уже сказано выше, наилучшее распознавание обучающей выборки (91 %) найдено для некоторых комбинаций, включающих 4 или 5 дескрипторов. Последовательное добавление к этим комбинациям других дескрипторов либо не изменяет, либо постепенно уменьшает долю правильно классифицируемых соединений.

С целью увеличить долю корректного распознавания соединений в обучающей выборке были исследованы возможности так называемого ансамблевого подхода (коллективного распознавания), при котором решение о классификации каждого соединения в ту или иную группу принимается не по одному классификационному правилу, пусть даже наилучшему из найденных, а с использованием совокупности полученных правил. За последние годы разные варианты ансамблевых методов были использованы для решения ряда задач в области хемо- и биоинформатики [10–19]. По-

казано, что применение ансамбля умеренных и даже слабых классификаторов дает возможность построить достаточно сильный классификатор. В настоящее время известны различные алгоритмы построения ансамблей: простое и взвешенное голосование, бустинг (усиление), бутстреп (случайная выборка с возвратом), бэггинг (bootstrap aggregating), случайный лес (random forest) и др. [20–22]. Выбранная нами стратегия принятия окончательного решения по классификации каждого соединения основана на простом голосовании результатов, полученных для различных комбинаций дескрипторов и считающихся равноценными, и в выборе группы с наибольшим числом совпадающих оценок. Чтобы более четко определить наличие или отсутствие ансамблевого эффекта при классификации рассматриваемой обучающей выборки, в каждой отдельной процедуре голосования принимали участие только те комбинации дескрипторов, которые классифицировали обучающую выборку с одинаковой ошибкой. При этом были отброшены комбинации дескрипторов, для которых ошибка классификации была сильно ассиметричной, т.е. неверно классифицированные соединения принадлежали преимущественно к одной из групп. Полученные результаты представлены в табл. 3.

Таблица 3

Классификация обучающей выборки методом коллективного распознавания (объем обучающей выборки — 44 соединения)

Классификация до голосования			Классификация после голосования	
Расознающая способность отдельной комбинации, %	Число соединений, неверно классифицируемых одной комбинацией дескрипторов	Число комбинаций дескрипторов, участвующих в голосовании	Расознающая способность, %	Число неверно классифицируемых соединений
91	4	13	93	3
88,6	5	37	95,5	2
86,4	6	61	100	0
84,1	7	112	100	0
81,2	8	107	100	0
79,5	9	163	97,7	1

Оценка активности соединений в тестовых выборках с помощью ансамблей комбинаций дескрипторов (1 — активное соединение, 2 — неактивное соединение)

№ вы- борки	Соединение	Экспери- ментальная активность	Результаты голосования и финальная оценка активности							
			Ансамбль 1* (61 комбинация)		Ансамбль 2** (112 комбинаций)		Ансамбль 3*** (107 комбинаций)		Ансамбль 4# (163 комбинации)	
1	2	1	61:0	1	112:0	1	103:4	1	157:6	1
	3	1	61:0	1	112:0	1	107:0	1	163:0	1
	7	1	61:0	1	112:0	1	105:2	1	163:0	1
	18	1	61:0	1	105:7	1	95:12	1	117:46	1
	24	2	12:49	2	33:79	2	42:65	2	81:82	2
	30	2	1:60	2	3:109	2	0:107	2	5:158	2
	31	2	3:58	2	15:97	2	16:91	2	29:134	2
	35	2	1:60	2	2:110	2	1:106	2	10:153	2
2	5	1	45:16	1	83:29	1	68:39	1	109:54	1
	6	1	61:0	1	112:0	1	107:0	1	162:1	1
	15	1	61:0	1	110:2	1	97:10	1	158:5	1
	17	1	59:2	1	111:1	1	101:6	1	156:7	1
	25	2	2:59	2	6:106	2	4:103	2	17:146	2
	27	2	2:59	2	15:97	2	20:87	2	31:132	2
	37	2	12:49	2	31:81	2	35:72	2	70:93	2
	44	2	12:49	2	38:74	2	45:62	2	66:97	2
3	14	1	59:2	1	101:11	1	95:12	1	133:30	1
	16	1	52:9	1	96:16	1	76:31	1	117:46	1
	19	1	49:12	1	87:25	1	77:30	1	128:35	1
	21	1	40:21	1	82:30	1	83:24	1	126:37	1
	29	2	4:57	2	7:105	2	4:103	2	9:154	2
	36	2	9:52	2	17:95	2	33:74	2	40:123	2
	42	2	25:36	2	34:78	2	37:70	2	36:127	2
	43	2	3:58	2	5:107	2	5:102	2	10:153	2
4	1	1	57:4	1	98:14	1	95:12	1	137:26	1
	8	1	61:0	1	112:0	1	105:2	1	154:9	1
	11	1	54:7	1	98:14	1	100:7	1	153:10	1
	23	1	40:21	1	75:37	1	74:33	1	88:75	1
	26	2	27:34	2	52:60	2	37:70	2	60:103	2
	28	2	19:42	2	43:69	2	34:73	2	46:117	2
	39	2	20:41	2	34:78	2	25:82	2	43:120	2
	41	2	8:53	2	20:92	2	16:91	2	42:121	2
5	4	1	2:59	2	15:97	2	11:96	2	33:130	2
	10	1	0:61	2	0:112	2	1:106	2	3:160	2
	13	1	40:21	1	81:31	1	70:37	1	107:56	1
	22	1	61:0	1	112:0	1	107:0	1	162:1	1
	32	2	0:61	2	0:112	2	1:106	2	6:157	2
	34	2	25:36	2	45:67	2	48:59	2	70:93	2
	38	2	5:56	2	16:96	2	14:93	2	36:127	2
	40	2	6:55	2	23:89	2	36:71	2	48:115	2
6	9	1	61:0	1	112:0	1	107:0	1	161:2	1
	12	1	41:20	1	69:43	1	55:52	1	98:65	1
	20	1	46:15	1	74:38	1	76:31	1	127:36	1
	33	2	7:54	2	12:100	2	19:88	2	49:114	2

* Каждая комбинация дескрипторов в этом ансамбле неверно классифицирует 6 соединений в обучающей выборке;

** каждая комбинация дескрипторов в этом ансамбле неверно классифицирует 7 соединений в обучающей выборке;

*** каждая комбинация дескрипторов в этом ансамбле неверно классифицирует 8 соединений в обучающей выборке;

каждая комбинация дескрипторов в этом ансамбле неверно классифицирует 9 соединений в обучающей выборке.

Найдено, что итоговое число комбинаций дескрипторов с наибольшей распознавательной способностью (91 %) невелико (чуть более 10), и применение вышеуказанной стратегии коллективного распознавания к классификационным правилам на основе этих комби-

наций дескрипторов дает небольшой эффект (увеличение распознавания обучающей выборки от 91 до 93 %). С другой стороны, показано, что, если применить стратегию коллективного распознавания к комбинациям дескрипторов, каждая из которых обладает го-

раздо меньшей распознавательной способностью (80 – 86 %), но число которых достаточно велико, суммарный эффект весьма ощутим (табл. 3). Доля правильно распознаваемых соединений в обучающей выборке в ряде случаев составляет 100 %. Например, совместное использование 107 комбинаций, полученных с помощью пошаговых процедур из исходного набора геометрических дескрипторов, причем распознавательная способность каждой из этих комбинаций составляет 81 % (неправильная классификация 8 соединений в обучающей выборке), приводит к корректной классификации всех соединений обучающей выборки, по-видимому, за счет взаимной компенсации ошибок, свойственных каждому отдельному классификатору. То же самое справедливо для ансамблей, состоящих из 61 комбинации дескрипторов (неправильная классификация 6 соединений в обучающей выборке) и из 112 комбинаций дескрипторов (неправильная классификация 7 соединений). В случае участия в голосовании 163 комбинаций дескрипторов, каждая из которых неверно классифицирует 9 соединений (табл. 3), соединение 24 из второй группы ошибочно классифицируется в первую группу с перевесом всего в 1 голос (82:81), что фактически означает отказ от голосования. Детальный анализ показал, что классификация подавляющего большинства соединений в ту или иную группу в рассмотренных ансамблях происходит, как правило, за счет заметного преобладания голосов в пользу одной из групп. Во всяком случае из представленных данных следует, что можно подобрать такой набор комбинаций, содержащих различные дескрипторы, которые при их совместном использовании будут корректно классифицировать все соединения обучающей выборки. Возможно, для построения наилучшего ансамбля в данном случае требуется некоторая оптимизация как общего числа используемых в ансамбле комбинаций, так и числа дескрипторов в комбинациях.

Для валидации построенных ансамблей классификационных правил использовали перекрестный контроль методом исключения группы соединений. С этой целью из обучающей выборки случайным образом были выделены 6 тестовых выборок, так что каждое соединение обучающей выборки оказалось в одной из тестовых выборок. Активность соединений в каждой тестовой выборке оценивали на основе соединений, оставшихся в обучающей выборке, с использованием полученных выше ансамблей комбинаций. Результаты классификации приведены в табл. 4 с указанием номеров соединений, включенных в каждую тестовую выборку, и числа голосов, поданных за каждое соединение. Как можно видеть, лишь 2 активных соединения (№ 4 и 10) из тестовой группы 5 ошибочно классифицируются как неактивные всеми указанными ансамблями. Прогноз активности для всех остальных соединений совпадает с экспериментально установленной активностью. При этом, несмотря на то, что при применении ансамбля 4 соединение 24 правильно классифицируется в неактивную группу перевесом

всего лишь в один голос, означая фактически отказ от голосования, учет результатов голосования по всем ансамблям дает надежную правильную классификацию этого соединения.

В целом, на основании полученных результатов можно заключить, что избранная стратегия одновременного использования коллектива (ансамбля) умеренных классификационных правил позволяет построить достаточно сильный классификатор для оценки гепатопротекторной активности химических соединений и дает веские основания для применения такого подхода в последующем виртуальном скрининге нетестированных соединений. Необходимо подчеркнуть, что применение ансамбля классификационных правил, каждое из которых включает небольшое число дескрипторов и статистически достоверно, дает возможность одновременно учитывать большое количество разнообразных молекулярных дескрипторов, кодирующих самую разную информацию о структуре молекулы, и обойти таким способом известную проблему “проклятия размерности”, что было бы невозможно, если использовать только единственное классификационное правило, даже наилучшее из найденных. Это особенно важно для обучающих выборок небольшого размера, когда число молекулярных дескрипторов, которые можно вычислить, существенно превышает число соединений в обучающей выборке. Более того, в последнем случае ансамблевый подход имеет большое преимущество, так как позволяет наиболее полно охарактеризовать топологические, стереохимические, электронные и другие особенности структуры классифицируемых молекул, необходимые для отбора перспективных с фармакологической точки зрения соединений. В то же время следует иметь в виду, что обратной стороной ансамблевого подхода, несмотря на его полезность в практическом плане для виртуального скрининга, является невозможность делать определенные выводы о механизмах действия рассматриваемых соединений именно из-за обилия и разнообразия одновременно используемых дескрипторов.

Для проведения исследований был составлен ряд рабочих программ на языке Фортран, реализующих метод ближайшего соседа, в том числе с автоматическим подсчетом результатов голосования по каждому соединению для любого количества введенных комбинаций дескрипторов.

ЛИТЕРАТУРА

1. Л. А. Пирузян, А. С. Кабанкин, Л. И. Габриелян и др., *Хим.-фарм. журн.*, **38**(3), 19 – 25 (2004).
2. А. С. Кабанкин, Л. И. Габриелян, *Хим.-фарм. журн.*, **39**(3), 23 – 27 (2005).
3. А. С. Кабанкин, Л. А. Радкевич, Л. И. Габриелян и др., *Хим.-фарм. журн.*, **39**(4), 24 – 28 (2005).
4. А. С. Кабанкин, Л. И. Габриелян, *Хим.-фарм. журн.*, **40**(6), 17 – 21(2006).
5. А. С. Кабанкин, Л. И. Габриелян, *Хим.-фарм. журн.*, **42**(5), 30 – 32 (2008).

6. Л. А. Радкевич, *Автореф. дис. докт. биол. наук*, Москва (1998).
7. R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, Germany (2000).
8. Milano Chemometrics & QSAR Research Group. DRAGON Web version 2.1 from <http://www.disat.unimib.it/chm> (2001).
9. P. Willett, J. M. Barnard, and G. M. Downs, *J. Chem. Inf. Comput. Sci.*, **38**(6), 983 – 996 (1998).
10. D. K. Agrafiotis, W. Cedeno, and V. S. Lobanov, *J. Chem. Inf. Comput. Sci.*, **42**(4), 903 – 911 (2002).
11. V. Svetnik, A. Liaw, C. Tong, et al., *J. Chem. Inf. Comput. Sci.*, **43**(6), 1947 – 1958 (2003).
12. C. Merkwirth, H. Mauser, T. Schulz-Gasch, et al., *J. Chem. Inf. Model.*, **44**(6), 1971 – 1978 (2004).
13. R. Guha and P. C. Jurs, *J. Chem. Inf. Model.*, **44**(6), 2179 – 2189 (2004).
14. H. Hong, W. Tong, Q. Hie, et al., *SAR QSAR Environ. Res.*, **16**(4), 339 – 347 (2005).
15. V. Svetnik, T. Wang, C. Tong, et al., *J. Chem. Inf. Model.*, **45**(3), 786 – 799 (2005).
16. T. Arodz, D. A. Yuen, and A. Z. Dudek, *J. Chem. Inf. Model.*, **46**(1), 416 – 423 (2006).
17. Q. Zhang, J. M. Hughes-Oliver, and R. T. Ng, *J. Chem. Inf. Model.*, **49**(8), 1857 – 1865 (2009).
18. П. М. Васильев, *Автореф. дис. докт. биол. наук*, Волгоград (2009).
19. M. Culp, K. Johnson, and G. Michalidis, *J. Chem. Inf. Model.*, **50**(2), 309 – 316 (2010).
20. I. J. Izenman, *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*, Springer Science + Business Media, New York (2008), pp. 505 – 550.
21. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Science + Business Media, New York (2009).
22. А. Б. Мерков, *Распознавание образов: Введение в методы статистического обучения*, Едиториал УРСС, Москва (2011), сс. 183 – 205.

Поступила 07.12.11

COLLECTIVE RECOGNITION STRATEGY FOR ESTIMATING HEPATOPROTECTOR ACTIVITY OF VARIOUS CHEMICAL COMPOUNDS IN INCREASING LIVER REPAIR POTENTIAL

A. S. Kabankin and L. A. Radkevich

Center for Theoretical Problems of Physicochemical Pharmacology, Russian Academy of Sciences, Moscow, Russia

The relationship between molecular characteristics and hepatoprotector activity (increasing the liver repair potential) has been investigated for a series of diverse chemical compounds. Application of the nearest-neighbor method with the use of geometric descriptors to a training set including 44 compounds (23 active and 21 inactive) gave combinations of the descriptors that correctly classify 79 – 91% of compounds in the training set. It is shown that, by applying the procedure of simple voting to the results obtained for various combinations of descriptors, it is possible to construct a classifier that correctly classifies all compounds of the training set. Efficiency of the proposed approach to prediction of the activity of non-tested compounds is proved by the method of cross-validation with groups of compounds removed from the training set.

Keywords: hepatoprotector activity; repair potential of liver; activity prognosis; geometric descriptors



22.04.2013

Компании Bind Therapeutics и AstraZeneca объявили о международном сотрудничестве по разработке и коммерциализации противоопухолевого препарата

Компания Bind Therapeutics и AstraZeneca объявили о стратегическом сотрудничестве по разработке и коммерциализации инновационного препарата класса Accurins™, обладающего адресным и программируемым действием. В основе разработки будет лежать уникальная технологическая платформа Medicinal Nanoengineering™, созданная компанией Bind Therapeutics, и селективный ингибитор киназы, созданный AstraZeneca.

Совместный проект основывается на недавно полученных научных данных, свидетельствующих о том, что препараты класса Accurins™ селективно накапливаются в поврежденных тканях и клетках, приводя к повышению концентрации препарата в зоне опухоли и снижению воздействия на здоровые ткани.

Пресс-служба РОСНАНО

Тел. +7 (495) 988-56-77

E-mail: press@rusnano.com