



© А.В. Тарасов, С.К. Абилев,
Р.М. Велибеков, В.А. Тарасов

Институт общей генетики
им. Н.И. Вавилова РАН, Москва

✿ Представлен новый принцип описания зависимости канцерогенной активности химических соединений от их структуры (QSAR-анализ), в основе которого лежит использование компаундных дескрипторов, представляющих собой сочетание несвязанных между собой структурных фрагментов молекулы. Разработана компьютерная программа, позволяющая генерировать всевозможные структурные фрагменты молекул химических соединений и проводить отбор компаундных дескрипторов, статистически значимо влияющих на их биологическую активность. На примере выборки из 89 галогенпроизводных углеводов, для которых исследована их канцерогенная активность на грызунах, показано, что использование компаундных дескрипторов значительно увеличивает прогностическую эффективность QSAR-анализа.

✿ **Ключевые слова:** QSAR-анализ; галогенпроизводные углеводородов; канцерогенная активность; компаундные дескрипторы

УВЕЛИЧЕНИЕ ЭФФЕКТИВНОСТИ QSAR-АНАЛИЗА ПРИ ОЦЕНКЕ КАНЦЕРОГЕННОЙ АКТИВНОСТИ ГАЛОГЕНПРОИЗВОДНЫХ УГЛЕВОДОРОДОВ

ВВЕДЕНИЕ

Галогенпроизводные углеводов широко применяются в нефтеперерабатывающей, авиационной, строительной, электронной промышленности, в сельском и коммунальном хозяйстве, медицине. Вместе с тем в настоящее время известно, что многие хлор- и бромпроизводные обладают выраженной токсичностью, в том числе и генотоксичностью. Некоторые галогенпроизводные углеводов являются безусловными канцерогенами для человека (винилхлорид, бис(хлорметил)эфир, сернистый иприт, диоксин, хлорнафазин, хлорамбуцил и др.), несколько десятков галогенпроизводных проявляют канцерогенную активность на грызунах [1]. В этой связи особый интерес представляет исследование связи между химической структурой и канцерогенной активностью галогенсодержащих соединений. Было показано, что расстояние равное 6,5Å между атомами хлора у полихлорированных бифенилов и 5,2Å между атомами хлора и азота у хлорированных гетероциклических соединений способствует проявлению канцерогенной активности [2]. Эти особенности структуры соединений характеризуют их способность с Ah-рецептором. Не исключено, что взаимодействие с этим рецептором определяет активность таких негенотоксичных канцерогенов, как эстрогены и антиэстрогены [3]. С другой стороны, ряд галогенсодержащих канцерогенов относится к категории генотоксичных соединений. К ним можно отнести в основном галогензамещенные алифатические соединения с короткой цепью. Анализ связи между структурой и мутагенной активностью в ряду галогензамещенных алифатических соединений показывает четкую корреляцию активности с параметрами, характеризующими реакцию способность соединений в реакциях метаболической трансформации и реакционную способность образующихся метаболитов в реакциях с ДНК [4].

Результаты, указывающие на связь между конкретными особенностями структуры молекул химических соединений и их канцерогенной активностью, немногочисленны. Это связано, в первую очередь, с недостаточной эффективностью используемых в настоящее время методов QSAR (Quantitative Structure Activity Relationships) для анализа зависимости канцерогенной активности от структуры соединения. Изначально методы QSAR были разработаны и развиты для повышения эффективности синтеза и поиска новых лекарственных препаратов и пестицидов [5, 6]. В этих условиях QSAR-анализ проводится на выборке близкородственных соединений, имеющих сходную химическую структуру.

Показано, что эффективность прямого использования методов QSAR для оценки потенциальной канцерогенной и мутагенной опасности химических загрязнений окружающей среды, оказывается достаточно низкой и не превосходит 60–65% правильно предсказанных соединений [7, 8]. Такая низкая эффективность QSAR-анализа связана с разнообразием химических

структур у соединений с канцерогенной и мутагенной активностью. И это указывает на необходимость поиска новых нетрадиционных подходов к QSAR-анализу при оценке канцерогенной и генетической опасности химических соединений, представляющих собой потенциальные или реальные загрязнители окружающей среды.

В настоящей работе нами проведен анализ связи структура — канцерогенная активность для 79 галогензамещенных соединений с использованием новой разработанной нами методологии генерации и отбора компандных дескрипторов, представляющих собой наличие двух независимых структурных фрагментов или наличия одного структурного фрагмента при обязательном отсутствии второго фрагмента.

МАТЕРИАЛЫ И МЕТОДЫ

Выборка химических соединений

Источником информации при создании выборки галогенсодержащих химических соединений, канцерогенная активность которых была исследована на грызунах, служила компьютерная база данных CPDB — Carcinogenic Potential Data Base (<http://potency.derkeley.edu/cpdb.html>). Для ретроспективного анализа были выбраны 79 химических соединений, в том числе 36 канцерогенов и 43 неканцерогена (табл. 1). Из этой выборки 28 соединений применяются в сельском хозяйстве в качестве пестицидов, 17 веществ используются в химическом син-

Таблица 1

Перечень химических соединений, использованных для анализа

N	CAS	Название	Акт	N	CAS	Название	Акт
1	106-89-8	Эпихлоргидрин	1	41	75-34-3	1,1-Дихлорэтан	0
2	126-72-7	Трис (2,3-дибромпропил) фосфат	1	42	78-87-5	Пропилен дихлорид	0
3	51-21-8	5-Флуорурацил	0	43	115-32-2	Дикофол	0
4	3778-73-2	Изофосфамид	1	44	60-57-1	Дизлдрин	0
5	51-75-2	Азотистый иприт	1	45	13366-73-9	Фотодизлдрин	0
6	100-44-7	Бензилхлорид	0	46	115-29-7	Тиосульфан	0
7	106-93-4	Этилендибромид	1	47	72-20-8	Эндрин	0
8	107-06-2	1,2-Дихлорэтан	1	48	22966-79-6	Эстрадиол иприт	0
9	309-00-2	Алдрин	0	49	72-56-0	р,р'-Этил-ДДД	0
10	563-47-3	Изобутенилхлорид	1	50	54143-56-5	Флеканид ацетат	0
11	115-96-8	Трис (2-хлорэтил) фосфат	1	51	363-17-7	N-(2-Флуоренил)-2,2,2-трифлуороацетамид	1
12	305-03-3	Хлорамбуцил	1	52	69112-98-7	2-Фторэтилнитрозо мочевины	1
13	6959-47-3	2-(Хлорметил)пиридин гидрохлорид	0	53	96-24-2	Глицерол альфа-монохлоргидрин	0
14	6959-48-4	3-(Хлорметил)пиридин гидрохлорид	1	54	76-44-8	Гептахлор	0
15	96-12-8	1,2-Дибромо-3-хлорпропан	1	55	87-68-3	Гексахлорбутадиев	1
16	488-41-5	Дибромманнитол	1	56	67-72-1	Гексахлорэтан	1
17	542-75-6	1,3-Дихлорпропен	1	57	96806-35-8	1-(2-гидрокси-пропил)-нитрозо-3-хлорэтил мочевины	0
18	5634-39-9	Иодированный глицерол	1	58	144-48-9	Иодоацетамид	0
19	148-82-3	Мельфалан	1	59	143-50-0	Кепон	1
20	76-01-7	Пентахлорэтан	0	60	576-68-1	Маннитол азотистый иприт	0
21	630-20-6	1,1,1,2-Тетрахлорэтан	0	61	72-43-5	Метоксихлор	0

Продолжение таблицы 1

Перечень химических соединений, использованных для анализа

N	CAS	Название	Акт	N	CAS	Название	Акт
22	79-34-5	1,1,2,2-Тетрахлорэтан	0	62	2385-85-5	Мирекс	1
23	140-57-8	Арамит	1	63	79-11-8	Хлоруксусная кислота	0
24	108-60-1	Бис-хлороизопропиловый эфир	0	64	126-85-2	N-оксид иприта	1
25	109-69-3	1-Хлорбутан	0	65	83335-32-4	N-нитрозо-бис-(4,4,4-трифтор- <i>n</i> -бутил) амин	1
26	77-65-6	Кабромал	0	66	96806-35-8	1-нитрозо-1-(2-гидроксипропил)-3-хлорэтилмочевина	1
27	115-28-6	Хлореновая кислота	1	67	82018-90-4	N-нитрозо (2,2,2-трифлуорэтил) этиламин	1
28	56-75-7	Хлорамфеникол	0	68	625-89-8	N-нитрозо-бис (2,2,2-трифтороэтил) амин	0
29	57-74-9	Хлордан	0	69	29082-74-4	Октахлорстирен	0
30	75-88-7	2-Хлоро-1,1,1-трифторэтан	1	70	110559-85-8	1-(2-Оксипропил) нитрозо-3-(2-хлорэтил) мочевины	0
31	532-27-4	Альфа-хлороацетофенон	0	71	09.10.3546	Фенестрин	1
32	140-49-8	4'-(Хлорацетил)-ацетонилдид	0	72	63-92-3	Феноксibenзамин гидрохлорид	1
33	75-00-3	Этилхлорид	0	73	29069-24-7	Преднимустин	1
34	999-81-5	Хлорхолин хлорид	0	74	95-06-7	Сульфалат	1
35	54749-90-5	Хлорзотоцин	1	75	569-57-3	Хлортрианизен	0
36	56-72-4	Кумафос	0	76	76-13-1	1,1,2-Трихлоро-1,2,2-трифторо-этан	0
37	72-54-8	Дихлордифенилди-хлор-этан	0	77	79-00-5	1,1,2-Трихлорэтан	0
38	50-29-3	ДДТ	1	78	42011-48-3	2,2,2-Трифтор-N-[4-(5-нитро-2-фурил)-2-тиазолил] ацетамид	1
39	10318-26-0	Дибромдулцитол	1	79	1582-09-8	Трифлуралин	0
40	3296-90-0	Дибромнеопентил гликоль	0	—	—	—	—

тезе, 6 — в лабораторных исследованиях в качестве химических реактивов, 19 соединений применяются в медицине и ветеринарии, 9 из которых являются противоопухолевыми препаратами.

При проведении расчетов значений апостериорной вероятности по формуле Байеса, значений информации и уровня доказанности использовались специально созданные в нашей лаборатории компьютерные программы с помощью языка VBA для Excel и Access, входящего в пакет Microsoft Office. Математическое обоснование этих показателей нами описано ранее [9, 10]. При проведении дискриминантного анализа использовали пакет статистических программ «Statistica 6,0».

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Идентификация структурных дескрипторов, статистически значимо влияющих на канцерогенную активность химических соединений.

В работе использована выборка, включающая 79 соединений, у которых определена канцерогенная активность грызунов. Используя дескрипторный блок программного комплекса NASAWIN [11, 12], было получено более 2000 структурных дескрипторов, которые встречались не менее чем в 4 х соединениях анализируемой выборки. Далее, используя созданную нами компьютерную программу, для каждого из этих дескрипторов проводилась оценка с помощью χ^2 -критерия значимости влияния дескриптора на мутагенную активность. В результате было отобрано 87 структурных дескрипторов, значимо влияющих на исследуемую нами активность химических соединений.

Эти дескрипторы не являются статистически независимыми; набор из 87 дескрипторов несет избыточную информацию. Для снижения количества дескрипторов по критерию их значимости использовали дискриминантный анализ и были отобраны 7 структурных дескрипторов, 3 из которых присутствуют только в активных, 4 — только в

Таблица 2

Одиночные однозначные дескрипторы

№	Структура дескриптора	Встречаемость		
		N _к	N _{нк}	№ соединений
Д1	C-C-C-C=C-C=C-N-C	4	0	51, 19, 12, 73
Д2	$\begin{array}{c} \\ \text{C}-\text{C}-\text{Br} \\ \end{array}$	4	0	39, 16, 15, 2
Д3	C-C-C-C-C=O	6	0	59, 12, 73, 19, 27, 35
Д4	$\begin{array}{c} \\ \text{C}-\text{Cl} \\ \end{array}$	0	11	77, 49, 45, 42, 41, 37, 29, 28, 22, 20, 54
Д5	$\begin{array}{c} \quad \\ \text{C}-\text{C}=\text{O} \\ // \end{array}$	0	5	50, 32, 31, 3, 36
Д6	C-C-C-C=C-C-Cl	0	6	47, 44, 46, 9, 54, 29
Д7	$\begin{array}{c} \text{—O—C}=\text{C} \\ \\ \text{C} \\ \end{array}$	0	5	75, 48, 50, 61, 36

Примечание: N_к и N_{нк} — число канцерогенов и неканцерогенов, содержащих данный дескриптор.

неактивных соединений, то есть являются, соответственно, биофорами и биофобами. Таким образом, отобранные

нами дескрипторы являются однозначными идентификаторами канцерогенных или неканцерогенных соединений.

Характеристика одиночных однозначных дескрипторов

Структуры отобранных одиночных дескрипторов и их встречаемость в молекулах канцерогенов и неканцерогенов приведены в таблице 2.

Дескриптор Д1. Представляет собой фрагмент бензольного кольца, соединенного с углеродной цепью, с одной стороны, и атомом азота, с другой. Обнаружен в молекулах хлорамбуцила, мелфалана, преднимустина и N-2-фторенил-2, 2, 2-трифторацетамида.

Дескриптор Д2. Короткий фрагмент, содержащий атом брома. Содержится в молекулах дибромдуситола, дибромманнитола, трис-2, 3 (дибромпропил) фосфата и 1,2-дибром-3-хлорпропана.

Дескриптор Д3. Цепь из 5 углеродных атомов, связанных между собой одинарной связью, и атома кислорода, связанного с этой цепью двойной связью. Обнаружен в молекулах хлорамбуцила, мелфалана, преднимустина, хлореновой кислоты, хлорзотоцина и кепона.

Дескриптор Д4. Представляет собой короткий хлорсодержащий фрагмент. Биофоб и характеризует следующие неканцерогены: 1,1,1,2-тетрахлорэтан, 1,1,2,2-тетрахлорэтан, хлорамфеникол, хлордан, дихлордифенилдихлорэтан, 1,2-дихлорэтан, фотодиэльдрин, р,р-этил-ДДД, гептахлор, 1,1,2-трихлорэтан, пропилен дихлорид.

Дескриптор Д5. Фрагмент, содержащий ацетильный остаток, является биофобом. Обнаружен в молекулах 5-фторурацила, Ко Ралла, флекаинид ацетата, альфа-хлорацетофенона, 4-(хлорацетил)ацетамида.

Дескриптор Д6. Фрагмент из 6 атомов углерода, содержащий атом хлора в качестве терминального атома. Биофоб и характеризует следующие неканцерогены: хлордан, алдрин, диэльдрин, эндрин, гептахлор, тиосульфат. Все эти соединения относятся к группе полихлорированных пестицидов.

Дескриптор Д7. Представляет собой фрагмент бензольного кольца, соединенного с атомом кислорода, и является биофобом. Обнаружен в молекулах Ко Ралла, флекаинид ацетата, эстрадиол иприта, метоксихлора и хлортрианизина.

Одиночные дескрипторы описывают только 34 соединений из 79.

Выявление компаундных дескрипторов — однозначных идентификаторов канцерогенной активности

Как уже отмечалось выше, большинство компьютерных методов QSAR-анализа в генетической токсикологии основаны на генерации всевозможных фрагментов структуры и статистического анализа вклада этих фрагментов на генотоксичность химических соединений [13, 14]. Вме-

сте с тем, свойство генотоксичности может определяться не только одним фрагментом, но и сочетанием отдельных фрагментов структуры. В этой связи нами была разработана программа генерации и отбора компаундных дескрипторов, то есть сочетанных фрагментов структуры. Программа включает в себя несколько этапов (рис. 1).

Первый этап связан с генерацией компаундных дескрипторов, которые представляют собой всевозможные сочетания первичных структурных дескрипторов, представляющих собой различные фрагменты структуры молекул химических соединений. В отношении каждого дескриптора химические соединения можно рассматривать в бинарном пространстве событий, то есть соединение по отношению к конкретному дескриптору находится в двух состояниях, характеризующихся наличием данного дескриптора либо его отсутствием. В случае выборки, можно обозначить через p_i долю соединений, содержащих этот дескриптор, то очевидно:

$$p_i + q_i = 1,$$

где q_i — доля соединений, не содержащих i -й дескриптор.

Тогда распределение соединений в зависимости от содержания двух дескрипторов определяется:

$$(p_i + q_i) \times (p_j + q_j) = 1,$$

где q_i и q_j — доля соединений, не содержащих i -й и j -й дескрипторы, соответственно. Или

$$p_i p_j + p_i q_j + q_i p_j + q_i q_j = 1,$$

где $p_i p_j$ — доля соединений, содержащих i -й и j -й дескрипторы;

$p_i q_j$ — доля соединений, содержащих i -й дескриптор при условии отсутствия j -го дескриптора;

$q_i p_j$ — доля соединений, содержащих j -й дескриптор при условии отсутствия i -го дескриптора;

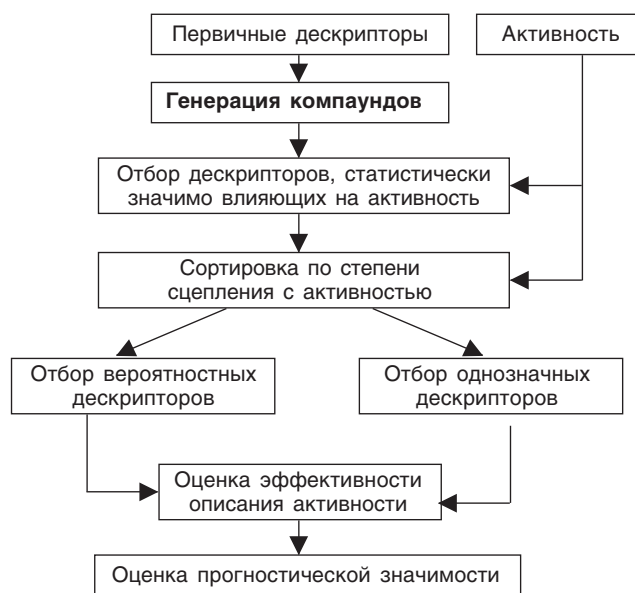


Рис. 1. Схема генерации и отбора компаундных дескрипторов

q_1q_2 — не рассматривается в качестве дескриптора, поскольку не несет значимую информацию.

Фактически, мы разделяем общую выборку на четыре подвыборки по наличию или отсутствию в них двух одиночных дескрипторов. Исходя из этого компаундные дескрипторы (двойные) можно разделить на 3 типа: pp — наличие в структуре молекулы обоих фрагментов; pq — наличие в структуре молекулы первого фрагмента при обязательном отсутствии второго фрагмента; qp — отсутствие в структуре молекулы первого фрагмента при обязательном наличии второго фрагмента.

На втором этапе для каждого из полученных таким образом компаундных дескрипторов определялось его присутствие в активных и неактивных соединениях. При этом отбирались с помощью критерия χ^2 те из них, для которых соотношение активных и неактивных соединений, содержащих данный дескриптор, статистически значимо отличалось от такового в исходной выборке.

В результате формировался набор компаундных дескрипторов, статистически значимо влияющих на канцерогенную активность химических соединений. Были выделены 19 компаундных дескрипторов, 9 из которых были однозначными идентификаторами канцерогенной активности, то есть биофорами. 10 компаундов присутствовали только в структурах неканцерогенных соединений и, соответственно, являлись биофобами.

Характеристика компаундных дескрипторов биофоров

В таблице 3 приведены структуры 9 компаундных дескрипторов являющихся однозначными индикаторами канцерогенной активности.

Дескриптор С1. Представляет сочетание в структуре химического соединения двух коротких фрагментов, один из которых является метилгалогеном, второй — представляет собой одинарную связь между атомами углерода и кислорода в цепи или в цикле. Такие фрагменты содержатся в структуре таких семи канцерогенов, как эпихлоргидрин, изофосфамид, трис-(2-хлорэтил) фосфат, арамит, феноксибензамин, преднимустин, трис-(2,3-дибромпропил) фосфат.

Дескриптор С2. Компаунд состоит из двух фрагментов, один из которых является цепочкой из трех углеродных атомов одинарными связями, второй — тройное разветвление у атома азота, два из которых связаны с атомами углерода. Такие фрагменты встречаются у изофосфамида, хлорамбуцила, нитрозо-бис-трифторбутиламина, фенестерина и преднимустина.

Дескриптор С3. Представляет сочетание хлорметильного фрагмента и фрагмента углерод—углерод с одинарной связью между ними. Однако у одного из атомов углерода обязательно должна быть двойная связь. Такие

фрагменты встречаются у 1,3-дихлорпропана, хлорзоточина, фенестерина и преднимустина.

Дескриптор С4. Компаунд состоит из атома азота и фрагмента бензольного кольца. Встречается у хлорамбуцила, хлорметилпиридина, мелфалана, фенестерина и преднимустина.

Дескриптор С5. Сочетание короткого ацетильного фрагмента и длинной цепи из семи разных атомов с одинарными связями между ними за исключением последних. Содержится компаунд в структуре N-фторметилтрифторацетамида, кепона, преднимустина, трифторнитрофурил тиазолила, ацетамида.

Дескриптор С6. Данный дескриптор типа pq характеризует наличие тройного разветвления у атома азота при обязательном отсутствии в структуре химического соединения сложной связи типа разветвления в четыре стороны. Дескриптор обнаруживается в молекулах азотистого иприта, хлорамбуцила, мелфалана, сулфуллата, феноксибензамина и преднимустина.

Дескриптор С7. Представляет сочетание в структуре химического соединения длинного фрагмента из 8 атомов, соединенных между собой одинарной связью и фрагмента с тройным разветвлением у атома углерода. Встречается в молекулах арамит, глицерола йодированного, фенестерина, феноксибензамина и 1-нитрозо-1-(2-гидроксипропил)-3-хлормочевина.


Дескриптор С8. Дескриптор типа qp отличается от других, подобно дескриптору С6, наличием одного характерного фрагмента при отсутствии в структуре соединения другого фрагмента. В данном случае отсутствие бензольного кольца сочетается с обязательным наличием этилгалогена в молекулах трис(2,3-дибромпропил) фосфата, дибромдуситола, 1,2-дибром-3-хлорпропана, дибромманитола.

Дескриптор С9. Представляет сочетание в структуре химического соединения фрагмента с тройным разветвлением у атома углерода и длинной цепи из 5 атомов углерода, терминальный атом которого связан двойной связью с атомом кислорода. Такие фрагменты имеются в молекулах хлорамбуцила, мелфалана, кепона, хлорзоточина, хлорденовой кислоты и преднимустина.

Анализ структур вышеописанных компаундов биофоров и их встречаемости в молекулах канцерогенных соединений позволяет разделить эти дескрипторы на три группы: азотсодержащие (С2, С4, С6), галогенсодержащие (С1, С3, С8) и дескрипторы с гетероатомами (С5, С7, С9). Азотсодержащие дескрипторы встречаются преимущественно в препаратах медицинского назначения, главным образом, в противоопухолевых соединениях (например, изофосфамид, азотистый иприт, хлорамбуцил, мелфалан и др.). Галогенсодержащие дескрипторы обнаруживаются в основном в молекулах пестицидов и в соединениях промышленного применения (эпихлоргидрин, 1,3-дихлорпропан, кепон, и др.). Гетероатомные

Таблица 3

Компаундные дескрипторы-биофоры

№	Тип дескриптора	Структура дескриптора		Встречаемость	
		Фрагмент 1	Фрагмент 2	N _к	№ соединений
C1	pp	$\begin{array}{c} \quad \\ \text{C} - \text{O} \end{array}$	$\begin{array}{c} \\ \text{C} - \text{Hal} \end{array}$	7	1, 2, 3, 11, 23, 72, 73
C2	pp	$\begin{array}{c} \quad \quad \\ \text{C} - \text{N} - \text{C} \end{array}$	$\begin{array}{c} \quad \\ \text{C} - \text{C} - \text{C} \end{array}$	5	4, 12, 65, 71, 73
C3	pp	$\begin{array}{c} \text{C} - \text{C} \\ \end{array}$	$-\text{C} - \text{Cl}$	4	17, 19, 35, 71, 73
C4	pp	N	$\begin{array}{c} \text{C} = \text{C} - \text{C} \\ \\ \text{C} \end{array}$	6	12, 14, 19, 71, 72, 73
C5	pp	$\begin{array}{c} \quad \\ \text{C} - \text{C} = \text{O} \\ \end{array}$	X — X — X — X — X — X — X = X	4	51, 59, 73, 79
C6	pq	$\begin{array}{c} \quad \quad \\ \text{C} - \text{N} - \text{C} \end{array}$		7	4, 5, 12, 19, 72, 73, 74
C7	pp	$\begin{array}{c} \\ \text{C} - \text{C} \\ \end{array}$	X — X — X — X — X — X — X — X	5	18, 23, 66, 71, 72
C8	qp	$\begin{array}{c} \\ = \text{C} \end{array}$	$\text{C} - \text{C} - \text{Cl}$	4	2, 15, 16, 39
C9	pp	$\begin{array}{c} \\ = \text{C} \end{array}$	$\text{C} - \text{C} - \text{C} - \text{C} - \text{C} = \text{O}$	6	12, 19, 25, 35, 59, 73

Примечание: pp — наличие в структуре молекулы обоих фрагментов; pq — наличие в структуре молекулы первого фрагмента при обязательном отсутствии второго фрагмента; qp — отсутствие в структуре молекулы первого фрагмента при обязательном наличии второго фрагмента.

дескрипторы содержатся в соединениях самого различного назначения. Следует учесть, что имеет место частичная перекрываемость дескрипторов, то есть в соединениях со сложной структурой встречаются дескрипторы как азотсодержащие, так и галогенсодержащие. В этой связи деление дескрипторов на группы является условным. Однако оно может отражать особенности биотрансформации соединения, приводящей к образованию конечного канцерогенного метаболита. Так, азотсодержащие дескрипторы характеризуют те соединения, которые в процессе биотрансформации образуют электрофильный имин. Со стабильностью такого имина коррелирует мутагенная и канцерогенная активность ароматических и гетероциклических аминов [15]. У большинства соединений, охарактеризованных галогенсодержащими дескрипторами, биологическая активность (токсичность, мутагенность) коррелирует с квантово-химическими параметрами (например, разность энергии высшей заполненной и нижней свободной молекулярных орбиталей), характеризующими реакционную способность соединения [4, 16].

Прогностическая эффективность одиночных и компаундных дескрипторов и их комбинации

Как видно из табл. 4, семь одиночных дескрипторов, отобранных по критерию встречаемости их не менее чем в 4-х соединениях, могут прогнозировать активность только 34 соединений из 79, в том числе, 11 канцерогенов из 36. В то же время 19 компаундов прогнозируют актив-

ность 61 соединения из 79, в том числе 25 канцерогенов из 6.

Если одиночные дескрипторы, а именно 3 иофора, определяют активность 30% канцерогенов, то компаунды, а именно 9 биофоров, определяют активность 69% канцерогенов. При этом представляет интерес дополнение одиночных дескрипторов компаундами для тех соединений, активность которых не определяется одиночными дескрипторами. Оказалось, что при сочетании одиночных дескрипторов и компаундов (7 одиночных + 14 компаундов), доли определяемых канцерогенов и неканцерогенов равны долям этих категорий соединений, определяемыми только компаундами. Это свидетельствует о том, что компаунды перекрывают все те соединения, активность которых определяется одиночными дескрипторами. Проведенный анализ относительно «взаимозаменяемости» одиночных и компаундных дескрипторов показал, что фактически биофоры Д1, Д2 и Д3 заменяют компаунды С5, С8 и С9. Однако это не было связано со структурными сходствами этих одиночных биофоров с какими-либо составляющими фрагментами компаундов С5, С8, и С9. Это связано только с тем, что в соединениях, в которых встречаются указанные компаунды, обнаруживается тот или иной одиночный биофор. Все это зависит только от сложности или разветвленности структуры соединения. Из табл. 4 также следует, что при использовании комбинации одиночных + компаундных дескрипторов и только компаундных дескрипторов достигается один и тот же результат по числу правильно определяемых соеди-

Таблица 4

Результаты анализа встречаемости одиночных и компаундных дескрипторов в молекулах галогенсодержащих соединений

Категория	Соединения, содержащие дескрипторы		Из них правильно определенных		Средняя информация
	Число	Доля	Число	Доля	
Одиночные дескрипторы (7 дескр.)					
Канцерогены	11	0,3	11	1,0	0,14
Неканцерогены	23	0,53	23	1,0	0,27
Вся выборка	34	0,43	34	1,0	0,21
Одиночные + компаундные дескрипторы (7 + 14 дескр.)					
Канцерогены	25	0,69	25	1,0	0,35
Неканцерогены	36	0,84	36	1,0	0,48
Вся выборка	61	0,77	61	1,0	0,42
Компаундные дескрипторы (19 дескр.)					
Канцерогены	25	0,69	25	1,0	0,35
Неканцерогены	36	0,84	36	1,0	0,48
Вся выборка	61	0,77	61	1,0	0,42

нений. Однако одинаковый прогностический эффект является результатом использования разных по количеству дескрипторов — 21 комбинированных (7 одиночных + 14 компаундов) дескрипторов или 19 компаундных дескрипторов.

Оценка эффективности QSAR-анализа

В качестве количественной меры эффективности QSAR-анализа нами использовалась селективная информация о наличии либо отсутствии канцерогенной активности и уровень доказанности канцерогенной активности химических соединений, целесообразность использования которых была ранее нами обоснована для оценки эффективности биотестирования [9, 10]. Величина уров-

ня доказанности после проведения i -го испытания равна полной информации о наличии у соединений активности, в данном случае канцерогенной.

$$\omega_i = \frac{1 + I_n \text{Sign}(p_{ai} - 0,5)}{2},$$

где ω_i — уровень доказанности канцерогенной активности;

I_n — полная информация о наличии канцерогенной активности.

В свою очередь при выборе между альтернативными возможностями (канцероген — неканцероген) полная информация равна:

$I_n = 1 - H_a$, а $H_a = p_a \log_2 1/p_a + (1 - p_a) \log_2 1/(1 - p_a)$, где H_a — апостериорная неопределенность, p_a — апостериорная вероятность наличия канцерогенной активности.

Видно, что значение уровня доказанности однозначно определяется апостериорной вероятностью, которая рассчитывалась в Бейсовском приближении.

В табл 4, наряду с долями правильно предсказанных соединений среди канцерогенов и неканцерогенов, приведены значения средней информации для каждой группы соединений и для всей выборки, рассчитанные на основе апостериорной вероятности наличия канцерогенной активности для каждого соединения в зависимости наличия в его структуре одиночных или компаундных дескрипторов. Видно, что уровень средней информации в случае однозначных компаундных дескрипторов намного выше, чем в случае одиночных дескрипторов.

На рис. 2 представлено распределение химических соединений в зависимости от величины уровня доказанности наличия у них канцерогенной активности в случае одиночных (А) и компаундных (Б) дескрипторов. Поскольку часть соединений, присутствующих в выборке, не содержит отобранных одиночных дескрипторов, информация об их активности равна нулю; значения апостериорной вероятности и уровня доказанности для них не изменяются в ходе анализа и остаются равными 0,5, то есть в области неопределенности.

В данной области находится большая доля как канцерогенов, так и неканцерогенов. Иная картина наблюдается в случае использования для прогноза канцерогенной активности компаундных дескрипторов — для основной массы канцерогенов $\omega = 1$, тогда как для большинства неканцерогенов $\omega = 0$ (рис. 2Б).

Таким образом, предложенный нами подход к изучению зависимости канцерогенной активности соединений от их структуры, основанный на использовании однозначных компаундных дескрипторов, заметно повышает его прогностическую эффективность.

Работа выполнена при финансовой поддержке гранта РФФИ 02-04-49849 и Программы фундаментальных исследований РАН «Динамика генофондов растений, животных и человека» (Проект № 24-ПИОГ-0302004).

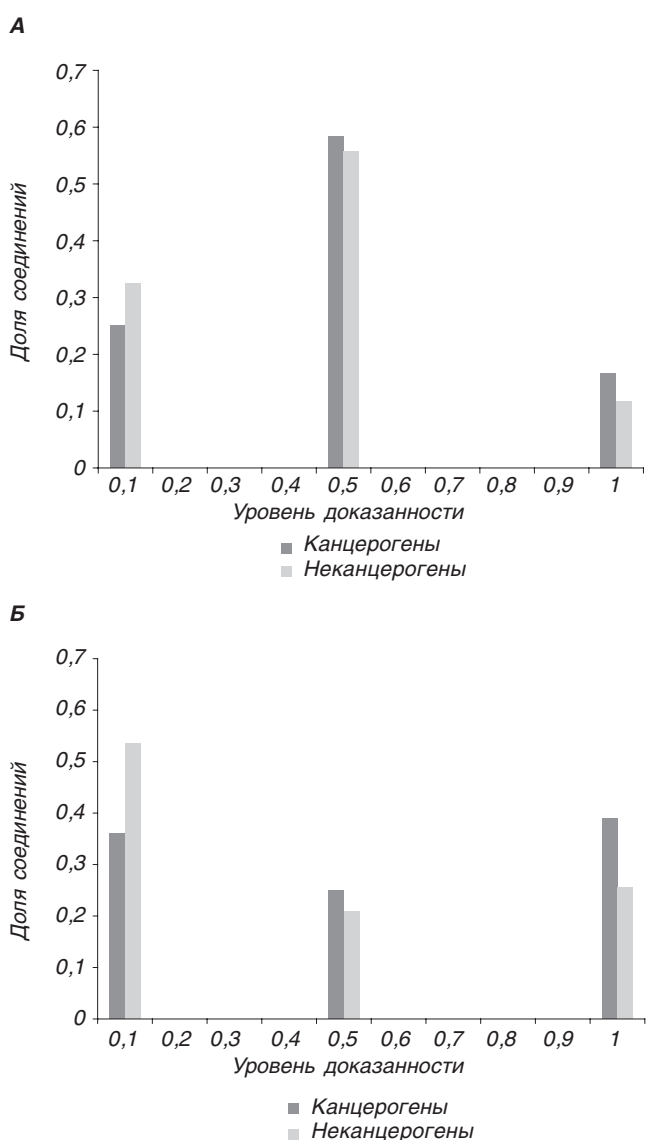


Рис. 2. Распределение химических соединений по уровням доказанности канцерогенной активности при использовании для прогноза их активности одиночных (А) и компаундных (Б) дескрипторов

Литература

1. *Худолей В.В.* Канцерогены: характеристики, закономерности, механизмы действия. — СПб.: НИИ Химии СПбГУ, 1999. — 419 с.
2. *Cunningham A.R., Klopman G., Rosenkranz H.S.* Identification of structural features and associated mechanisms of action for carcinogens in rats // *Mutation Res.* — 1998. — Vol. 405. — P. 9–28.
3. *Rosenkranz H.S., Cunningham A.R., Klopman G.* Identification of 2D geometric descriptor associated with nongenotoxic carcinogens and estrogens and antiestrogens // *Mutagenesis.* — 1996. — Vol. 11. — P. 95–100.
4. *Харчевникова Н.В., Жолдакова З.И., Журков В.С.* Теоретическое обоснование связи структура–генотоксичность для млекопитающих и бактерий в ряду галогензамещенных алифатических соединений с короткой цепью // *Вестн. РАМН.* — 1997. — № 12. — С. 8–13.
5. *Hansch C., Leo A.* QSAR. Part I. Fundamentals and applications in chemistry and biology // *American Chem. Soc.* — Washington, D, 1995.
6. *Fujita T.* Recent success stories leading to commercializable bioactive compounds with the aid of traditional QSAR procedures // *QSAR.* — 1997. — Vol. 16. — P. 107–112.
7. *Benigni R.* The first US National Toxicology Program comparative exercise on the prediction of rodent carcinogenicity: definitive results // *Mutation Res.* — 1997. — Vol. 387. — P. 35–45.
8. *Benigni R., Zito R.* The second National Toxicology Program comparative exercise on the prediction of rodent carcinogenicity: definitive results // *Mutation Res.* — 2004. — Vol. 566. — P. 49–63.
9. *Тарасов В.А., Асланян М.М., Абилев С.К.* Принципы формализованной количественной оценки опасности химических соединений для человека // *Генетика.* — 1999. — Т. 35. — С. 1585–1599.
10. *Тарасов В.А., Тарасов А.В., Любимова И.К., Асланян М.М.* Проблема количественной оценки опасности химических соединений в генетической токсикологии // *Усп. совр. биол.* — 2002. — Т. 122, № 2. — С. 136–147.
11. *Баскин И.И., Палюлин В.А., Зефилов Н.С.* Применение искусственных нейронных сетей в химических и биохимических исследованиях // *Вестн. Моск. ун-та. Сер. 2: Химия.* — 1999. — Т. 40, № 5. — С. 323–326.
12. *Баскин И.И., Гальберштам Н.М., Палюлин В.А., Зефилов Н.С.* NASAWIN — программный комплекс для изучения соотношения структура–свойство в химии // *Тр. VII Всерос. конф. «Нейрокомпьютеры и их применение» с Международным участием.* — М., 2001. — С. 419–422.
13. *Klopman G., Frierson M., Rosenkranz H.* The structural basis of the mutagenicity of chemicals in *Salmonella typhimurium*: The Gene-Tox Data Base // *Mutation Res.* — 1990. — Vol. 128. — P. 1–50.
14. *Klopman G., Rosenkranz H.S.* Approaches to SAR in carcinogenesis and mutagenesis. Prediction of carcinogenicity/mutagenicity using MULTICASE // *Mutation Res.* — 1994. — Vol. 305. — P. 33–46.
15. *Benigni R., Passerini L.* Carcinogenicity of the aromatic amines: from the structure–activity relationships to mechanisms of action and risk assessment // *Mutation Res.* — 2002. — Vol. 511. — P. 191–203.
16. *Tuppurainen K., Lotjonen S., Laatikainen R., Vartiainen T., Maran U., Strandberg M., Tamm T.* About the mutagenicity of chlorine-substituted furanones and halopropenales. A QSAR study using molecular orbital indices // *Mutation Res.* — 1991. — Vol. 247. — P. 97–102.

Increase of efficiency of QSAR — the analysis of carcinogenic activity of halogenated hydrocarbons

Tarasov A.V., Abilev S.K., Velibekov R.M., Tarasov V.A
Vavilov Institute of General Genetics, Russian Academy of Sciences

✿ **SUMMARY:** The new principle of analysis of structureactivity of carcinogenic compounds is offered. It is basis on use a compound structural descriptors representing a combination untied among themselves of structural fragments of a molecule The computer program is developed, allowing to generate every possible structural fragments of molecules of chemical compounds and their combinations and to carry out selection of the compound descriptors statistically significantly influencing their biological activity. In work it is used the samples consisting of 89 halogenated hydrocarbons for which their carcinogenic activity on rodents is investigated.

✿ **KEY WORDS:** QSAR; carcinogenic activity; compound descriptors; halogenated hydrocarbons