

В.В.Важев, Н.В.Важева, С.Х.Шерметова
Костанайский государственный педагогический институт

ПРОГНОЗИРОВАНИЕ КАНЦЕРОГЕННОЙ АКТИВНОСТИ ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ ДЛЯ КРЫС ПО МАСС-СПЕКТРАМ

Әр түрлі класс органикалық қосылыстарының егеуқұйрықтар үшін TD_{50} канцерогенділікті болжауда молекулалық құрылымның дескрипторлары ретінде масс-спектрлерді пайдалану мүмкіндігі зерттелді. 103 органикалық заттан тұратын жиынтық қолданылды. Барлық жиынтық үшін корреляция коэффициенті 0,9850 болып сипаттайтын модель құрастырылды. Жылжымалы бақылау кезіндегі корреляция коэффициенті 0,9275 болды.

The possibility of mass spectra as descriptors of molecular structure with the prediction of carcinogenic activity of TD_{50} for rats organic compounds of various classes. Use a set of 103 organic substances. The model, characterized by a correlation coefficient of 0,9850 for the whole set. When the slip control, the correlation coefficient was 0,9275.

Актуальность развития QSAR (Quantitative Structure–Activity Relationships) методов прогнозирования канцерогенной активности органических соединений не нуждается в обосновании. Химические канцерогены ответственны за возникновение до 80–90 % всех злокачественных опухолей человека. Исследования канцерогенной активности химических соединений, которые в основном проводятся на грызунах (мыши, крысы), могут продолжаться в течение двух-трех лет и требуют значительного финансирования и высококвалифицированных специалистов. По данным [1] стоимость исследования одного соединения составляет 2 миллиона долларов. При разработке новых лекарственных веществ без учета отдаленной угрозы их канцерогенности ущерб может составлять более 700 миллионов долларов и потерю шести-восьми лет усилий научных и производственных коллективов.

Учитывая, что каждый год синтезируются миллионы новых соединений, необходимость развития QSAR становится очевидной. Обычно количественной мерой канцерогенной активности служит TD_{50} , определяемая как суточная доза (мг/кг), вызывающая появление определенного вида опухолей в конце стандартной продолжительности жизни для данного биологического вида. Часто при расчетах используют $\lg TD_{50}$ — логарифм величины, пересчитанной в ммоль/кг. Качество прогнозирования обычно характеризуют коэффициентом корреляции R между вычисленными и экспериментальными значениями исследуемого свойства и соответствующим стандартным отклонением s .

В работе [2] для небольшого набора ароматических и гетероциклических нитросоединений представлена нейросетевая модель $\lg TD_{50}$ с $R = 0,919$. Джини и другими [3] был исследован набор из 104 ароматических и азотсодержащих веществ в отношении канцерогенной активности для крыс. Полученная модель имеет характеристики: $R = 0,831$ для всего набора из 104 веществ и $R = 0,906$ для подмножества из 92 веществ. Врачко [4] для набора ароматических соединений, принадлежащих к и $s = 0,381$. Поройковым и другими [6] при использовании базы данных для потенциально канцерогенных веществ [7] составлена выборка из 512 органических соединений различных классов, для которых известны значения TD_{50} в отношении крыс (мужские и женские особи вместе). Полученная для этой выборки регрессионная модель имела статистические параметры $R = 0,764$ ($R^2 = 0,584$), $s = 0,965$. После исключения 43 веществ повторное моделирование показало $R = 0,838$ ($R^2 = 0,703$) и $s = 0,776$ для всего набора веществ. Для контрольной выборки из 10 соединений средняя величина $R = 0,714$. Подобные результаты в свете современного состояния прогнозирования канцерогенной активности считаются удовлетворительными. Для выявления канцерогенной активности химических соединений по системе «да-нет» широко применяется ряд коммерческих программ, среди которых [8] DEREK, OncoLogic, HazardExpert, TOPKAT, MCASE, CASE, CASETOX, ToxScope. Среднее значение удачных предсказаний находится на уровне 80 %.

Осложняющим обстоятельством для прогнозирования является комплексный механизм канцерогенеза, рассматриваемый в настоящее время большинством исследователей как многостадийный процесс, в котором следует различать 3 главные стадии: инициацию, промоцию и прогрессию. Принято считать, что существуют 2 типа агентов, различающихся по механизмам своего действия: инициаторы и промоторы. Действие инициаторов необратимо, действие промоторов до определенного момента обратимо. Оказалось, что большинство «сильных» канцерогенов обладает и иницирующими, и промоторными свойствами, а все промоторы, за редкими исключениями, проявляют канцеро-

генную активность, если их применять в высоких дозах и достаточно долго. Деление на инициаторы и промоторы в определенной степени соответствует делению канцерогенов на генотоксические и негенотоксические. В конечном счете, канцерогенность, как и другие свойства, определяется структурой молекулы, но зависит от большого числа факторов, которые не могут быть эффективно отражены обычно используемыми дескрипторами молекулярной структуры.

В наших работах [9] рассматривается возможность использования преобразованных ИК- и масс-спектров в качестве дескрипторов молекулярной структуры для QSAR/QSPR исследований. Спектры органических веществ обладают высокой индивидуальностью. Они однозначно (при стандартизации условий) определяются строением молекул и могут оказаться более информативными в отражении ее структуры, чем традиционно используемые дескрипторы. Ранее [10] нами было показано, что использование ИК-спектров как дескрипторов структуры молекул органических соединений позволяет прогнозировать канцерогенную активность в отношении крыс для всего набора из 81 соединения с коэффициентом корреляции $R = 0,9809$ и $s = 0,264$. Для контрольной выборки из 21 веществ $R = 0,9806$ и $s = 0,251$.

В данной работе исследована возможность использования масс-спектров для прогнозирования способности органических соединений различных классов вызывать рак печени у крыс (мужские и женские особи вместе). Использованы масс-спектры низкого разрешения веществ, имеющиеся в NIST (США) [11]. Показатели канцерогенной активности TD_{50} для крыс (рак печени) взяты из [7]. При пересечении множества органических веществ двух баз был сформирован набор из 103 веществ, присутствующих в обеих базах. Перед вычислениями все спектры были нормированы по формуле

$$d_{ij} = \frac{h_{ij}}{\sum_{k=1}^{k=n} h_{ik}}, \quad (1)$$

где d_{ij} — рабочие значения дескрипторов, используемые в расчетах; n — количество пиков; h_{ij} — их относительные интенсивности; i — номер вещества; j — номер пика с соответствующим m/z — отношением массы к заряду k -го иона в молекуле с номером i . Во всей совокупности масс-спектров для использованного набора веществ имеется 409 значений, отличных от нуля интенсивностей пиков в диапазоне отношений m/z от 12 до 550. Элементы d_{ij} формируют матрицу дескрипторов \mathbf{D} , состоящую из субматриц дескрипторов тренировочной — \mathbf{A} и контрольной — \mathbf{C} выборок. Тренировочная выборка состояла из 80, а контрольная — из 23 веществ. Расчеты выполнены с помощью разработанной нами [12] компьютерной программы PROGROC (PROGgram RObustness Calculation). Алгоритм программы отражен также в [13].

На рисунке 1 изображена зависимость коэффициента корреляции между экспериментальными и вычисленными значениями $\lg TD_{50}$ для всего набора веществ от величины $r(\mathbf{A})$ оценки ранга субматрицы дескрипторов тренировочной выборки.

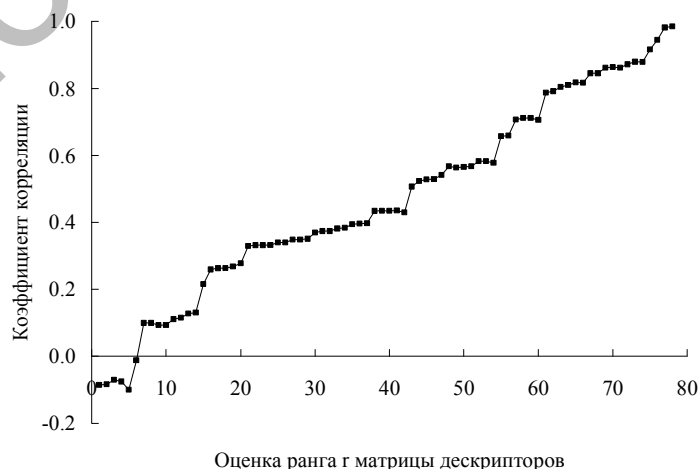


Рис. 1. Зависимость коэффициента корреляции между экспериментальными и вычисленными значениями $\lg TD_{50}$ от величины $r(\mathbf{A})$ оценки ранга субматрицы дескрипторов

Рисунок дополнительно свидетельствует о многофакторности прогнозируемого свойства и необходимости привлечения большого числа дескрипторов. Максимальное значение $R = 0,9850$ достигается при $r(\mathbf{A}) = 78$. Результаты прогнозирования $\lg\text{TD}_{50}$ при этом ранге приведены в таблице 1 и на рисунке 2.

Т а б л и ц а 1

Экспериментальные [7] и рассчитанные по масс-спектрам значения $\lg\text{TD}_{50}$ для крыс (рак печени), CAS — регистрационный номер по Chemical Abstracts

№	Соединение*	CAS	Эксп.	Рассч.	Разн.
1	2	3	4	5	6
1	Dimethoxane	828-00-2	0,61	0,58	0,03
2	N,N-dimethyl-4-[(3-methylphenyl)azo]-Benzenamine	55-80-1	-1,86	-1,86	0,00
3	Furfural	98-01-1	0,85	0,85	0,00
4	Clofibrate	637-07-0	-0,16	-0,16	0,00
5	2-Methyl-, 7-[[2,3-dihydroxy-2-(1-methoxyethyl)-3-methyl-1-oxobutoxy]methyl]-2,3,5,7a-tetrahydro-1H-pyrrolizin-1-yl ester, [1S-[1a(Z),7(2S*,3R*),7aa]]-2-butenic acid	303-34-4	-3,02	-2,92	-0,10
6	3,3'-Diaminobenzidine	7411-49-6	0,04	0,04	0,00
7	Carbamic acid, methyl ester	598-55-0	-0,12	-0,11	-0,01
8	1,2-Dimethoxy-4-(2-propenyl)-benzene	93-15-2	-0,95	-0,97	0,02
9	11-Amino-undecanoic acid	2432-99-7	0,74	0,75	-0,01
10	Cyclopentanone, oxime	1192-28-5	-0,38	-0,39	0,01
11	4-(N-Nitroso-N-methylamino)-1-(3-pyridyl)-1-butanone	64091-91-4	-3,32	-3,32	0,00
12	4-(4-chlorophenoxy)-benzenamine	101-79-1	-0,77	-0,80	0,03
13	Acetaminophen	103-90-2	0,51	0,50	0,01
14	Tetrafluoro-ethene	116-14-3	0,03	0,02	0,01
15	Dichloroacetic acid	79-43-6	0,10	0,10	0,00
16	Piperonyl butoxide	51-03-6	0,27	0,27	0,00
17	N-2-Fluorenylacetamide	53-96-3	-2,26	-2,27	0,01
18	4,4'-Thiobis-Benzenamine	139-65-1	-1,76	-1,77	0,01
19	1-Methyl-2-nitro-Benzene	88-72-2	-1,47	-1,36	-0,11
20	p, p'-DDT	50-29-3	-0,62	-0,63	0,01
21	2-Amino-3-methyl-3H-Imidazo(4,5-f)quinoline	76180-96-6	-2,39	-2,38	-0,01
22	1-Nitroso-Azacyclotridecane	40580-89-0	-1,29	-1,30	0,01
23	3a,12c-Dihydro-8-hydroxy-6-methoxy-, (3aR-cis)-7H-furo[3',2':4,5]furo[2,3-c]xanthen-7-one	10048-13-2	-3,33	-3,34	0,01
24	3-Chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone	77439-76-0	-2,57	-2,51	-0,06
25	Prednisolone	50-24-8	-2,37	-1,50	-0,87
26	1-Ethenyl-2-pyrrolidinone	88-12-0	-0,97	-0,97	0,00
27	Acetamide	60-35-5	0,48	0,45	0,03
28	2H-1-Benzopyran-2-one	91-64-5	-0,57	-0,58	0,01
29	Chloro-benzene	108-90-7	0,34	0,31	0,03
30	Diazene, dimethyl-, 1-oxide	25843-45-2	-3,20	-3,14	-0,06
31	Renardine	2318-18-5	-2,33	-2,32	-0,01
32	N'-(4-Chlorophenyl)-N,N-dimethyl-urea	150-68-5	-0,18	-0,19	0,01
33	2-Nitro-butane	600-24-8	-1,56	-1,62	0,06
34	1-Nitroso-pyrrolidine	930-55-2	-2,10	-2,10	0,00
35	Bis(2-ethylhexyl) phthalate	117-81-7	0,26	0,25	0,01
36	1,4-Dioxane	123-91-1	0,48	0,41	0,07
37	Gemfibrozil	25812-30-0	-0,01	-0,16	0,15
38	Octahydro-1-nitroso-azocine	20917-49-1	-3,58	-3,48	-0,10
39	Nitro-Benzene	98-95-3	-0,68	-0,70	0,02
40	1-(Phenylazo)-2-naphthalenol	842-07-9	-0,93	-0,93	0,00
41	Mirex	2385-85-5	-2,49	-2,49	0,00
42	Triamcinolone Acetonide	76-25-5	-3,91	-3,95	0,04
43	S-Ethyl-DL-homocysteine	67-21-0	-1,22	-1,20	-0,02
44	2-Methyl-4-[(2-methylphenyl)azo]-benzenamine	97-56-3	-1,75	-1,73	-0,02

1	2	3	4	5	6
45	2-Propanone, oxime	127-06-0	-0,78	-0,77	-0,01
46	Budesonide	51333-22-3	-3,17	-3,56	0,39
47	Aflatoxin b1	1162-65-8	-4,99	-5,01	0,02
48	2-Nitro-9H-Fluorene	607-57-8	-2,87	-2,86	-0,01
49	1,4,5,6,7,7-Hexachloro-bicyclo[2,2,1]hept-5-ene-2,3-dicarboxylic acid	115-28-6	-0,98	-0,97	-0,01
50	2,4,5-Trimethyl-benzenamine	137-17-7	-0,60	-0,73	0,13
51	Furan	110-00-9	-2,24	-2,23	-0,01
52	N-Methyl-N-nitro-methanamine	4164-28-7	-2,22	-2,30	0,08
53	1,3-Dichloro-1-propene	542-75-6	-0,07	-0,14	0,07
54	N-Methyl-N-nitroso-ethanamine	10595-95-6	-3,24	-3,24	0,00
55	Hydrazine	302-01-2	-1,72	-1,72	0,00
56	Captafol	2425-06-1	-0,68	-0,66	-0,02
57	2,3,6a,9a-Tetrahydro-1-hydroxy-4-methoxy-, [1S-(1a,6aδ,9aδ)]-cyclopenta[c]furo[3',2':4,5]furo[2,3-h][1]benzopyran-11(1H)-one	29611-03-8	-5,10	-5,18	0,08
58	Disperse yellow 3, CI 11855	2832-40-8	0,15	0,14	0,01
59	4-Methyl-1,3-Benzenediamine	95-80-7	-1,69	-1,83	0,14
60	Methylene bis-4,4'-(2-chloro-aniline)	101-14-4	-1,14	-1,15	0,01
61	Acetic acid ethenyl ester	108-05-4	0,60	0,63	-0,03
62	Carbon Tetrachloride	56-23-5	-0,74	-0,73	-0,01
63	Tris(1,3-dichloroisopropyl)phosphate	13674-87-8	-0,97	-0,98	0,01
64	4-Nitroso-N-phenyl-benzenamine	156-10-5	0,00	0,00	0,00
65	1-Hydroxy-9,10-anthracenedione	129-43-1	-0,38	-0,40	0,02
66	Bis[4-(dimethylamino)phenyl]-methanone	90-94-8	-1,68	-1,72	0,04
67	2-Methyl-1-nitro-9,10-Anthracenedione	129-15-7	-0,50	-0,49	-0,01
68	2,2-Dichloro-1,1,1-trifluoro-ethane	306-83-2	1,19	1,18	0,01
69	2-Methoxy-5-methyl-benzenamine	120-71-8	-0,15	-0,15	0,00
70	2-Butanone, oxime	96-29-7	-0,07	-0,01	-0,06
71	N-Methyl-N-nitroso-1-dodecanamine	55090-44-3	-2,63	-2,63	0,00
72	Prasterone	53-43-0	-0,54	-0,55	0,01
73	Chloro-ethene	75-01-4	-1,01	-1,02	0,01
74	1-Methyl-2-nitroso-benzene	611-23-4	-0,38	-0,48	0,10
75	1-Methyl-2,4-dinitro-benzene	121-14-2	-1,47	-2,15	0,68
76	Isoniazid	54-85-3	0,04	0,01	0,03
77	2-Methyl-1,3-dinitro-Benzene	606-20-2	-2,80	-1,82	-0,98
78	5-(2-propenyl)-1,3-benzodioxole	94-59-7	0,43	0,41	0,02
79	Ethanethioamide	62-55-5	-0,82	-0,81	-0,01
80	4,4'-Methylenebis[2-methyl-Benzenamine]	838-88-0	-1,49	-1,50	0,01
81	1,2-Diphenyl-hydrazine	122-66-7	-1,52	-1,18	-0,34
82	N-9H-fluoren-2-yl-N-hydroxy-acetamide	53-95-2	-2,38	-2,40	0,02
83	(E)-2-Butenal	123-73-9	-1,22	-0,84	-0,38
84	Ethyne, dichloro-	7572-29-4	-1,42	-1,66	0,24
85	N-Nitrosodimethylamine	62-75-9	-2,89	-3,68	0,79
86	2-Amino-9,10-anthracenedione	117-79-3	-0,34	-0,38	0,04
87	N-Butyl-N-nitroso-1-butanamine	924-16-3	-2,36	-2,56	0,20
88	alpha-Lindane	319-84-6	-1,41	-1,53	0,12
89	Nitroglycerin	55-63-0	-0,09	-0,08	-0,01
90	S-Ethyl-L-homocysteine	13073-35-3	-1,52	-1,07	-0,45
91	Ethanol, 2,2'-(nitrosoimino)bis-	1116-54-7	-1,63	-1,18	-0,45
92	1-Amino-2-methyl-9,10-anthracenedione	82-28-0	-0,60	-1,10	0,50
93	Ethanol	64-17-5	2,30	2,70	-0,40
94	Dehydroisoandrosterone acetate	853-23-6	-0,82	-1,17	0,35
95	Monocrotaline	315-22-0	-2,54	-2,20	-0,34
96	Pirinixic acid	50892-23-4	-1,87	-1,29	-0,58
97	N-Ethyl-N-nitroso-ethanamine	55-18-5	-3,59	-3,48	-0,11
98	Vinyl bromide	593-60-2	-0,76	-0,97	0,21
99	1-Amino-2,4-dibromo-9,10-anthracenedione	81-49-2	-0,92	-0,95	0,03
100	4,4'-Oxybis-benzenamine	101-80-4	-1,32	-1,07	-0,25

1	2	3	4	5	6
101	Hexachloro-benzene	118-74-1	-1,87	-2,09	0,22
102	Auramine o	2465-27-2	-1,44	-1,05	-0,39
103	N-Nitroso-N-pentyl-1-pentanamine	13256-06-9	-1,67	-1,17	-0,50

Примечание. Вещества с 1 по 80 составляют тренировочную выборку, с 81 по 103 — контрольную. Используются названия веществ, принятые в базе NIST.

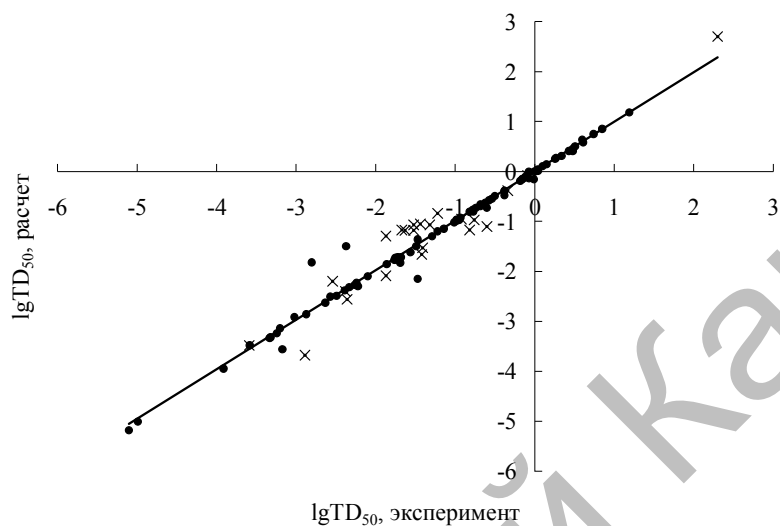


Рис. 2. Корреляция между экспериментальными и вычисленными значениями $\lg TD_{50}$: ● — тренировочная выборка; × — контрольная выборка

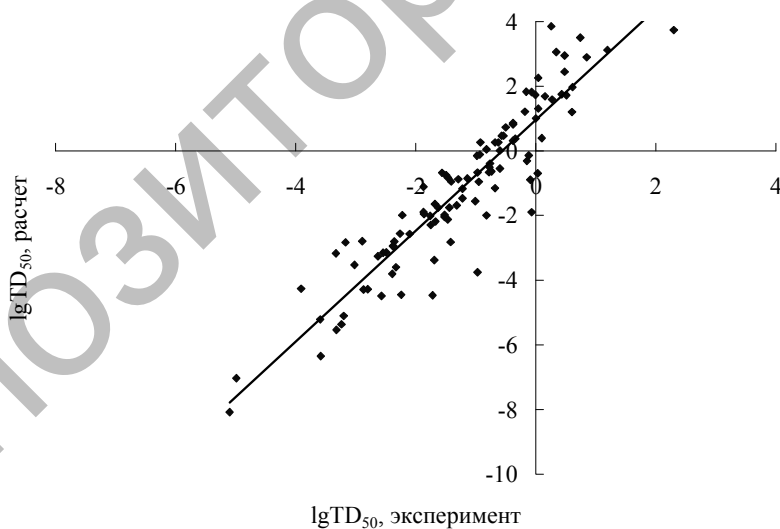


Рис. 3. Корреляция между экспериментальными и вычисленными значениями $\lg TD_{50}$ при скользящем контроле

Количественные показатели соответствующих корреляционных зависимостей приведены в таблице 2.

Данные таблиц 1 и 2 свидетельствуют, что полученные нами результаты по статистическим параметрам превосходят лучшие примеры прогнозирования других авторов. Для проверки адекватности корреляционных моделей и дескрипторов в QSAR/QSPR-исследованиях иногда применяется прием, носящий название перекрестной проверки — leave-one-out (удаление одного из), или скользящий контроль. Из исследуемого набора поочередно извлекается каждое вещество, модель строится по оставшимся веществам, затем выполняется оценка свойства удаленного вещества. Нами было выпол-

нено [14] повторное прогнозирование $IgTD_{50}$ для крыс с контролем по спарринг-партнеру — молекулярной массе. Результаты, приведенные на рисунке 3, характеризуются коэффициентом корреляции $R = 0,9275$ и стандартным отклонением $s = 0,92$.

Т а б л и ц а 2

Показатели корреляции между экспериментальными и вычисленными значениями $IgTD_{50}$ для крыс

Показатели корреляции	Весь набор	Тренировочная выборка	Контрольная выборка
R	0,9850	0,9916	0,9577
s	0,23	0,18	0,37

Можно отметить, что впервые в области прогнозирования канцерогенной активности осуществлен скользящий контроль с качеством, не уступающим простым корреляционным моделям, описанным в литературе.

Разрабатываемый нами подход может быть использован для оперативной оценки канцерогенной опасности неизвестных веществ с использованием только масс-спектров.

Список литературы

1. *Contrera J.F., Laughlin P.M., Hall L.H., Kier L.B.* QSAR Modeling of Carcinogenic Risk Using Discriminant Analysis and Topological Molecular Descriptors // *Current Drug Discovery Technologie*. — 2005. — Vol. 2. — P. 55–67.
2. *Ghoshal N., Mukhopadhyay S.N., Ghoshal T.K., Achari B.* Quantitative structure-activity relationship studies of aromatic and heteroaromatic nitro compounds using neural network // *Bioorganic & Medicinal Chemistry Letters*. — 1993. — № 3. — P. 329–332.
3. *Gini G., Lorenzini M., Benfenati E., Grasso P., Bruschi M.* Predictive Carcinogenicity: A Model for Aromatic Compounds, with Nitrogen-Containing Substituents, Based on Molecular Descriptors Using an Artificial Neural Network // *J. Chem. Inf. Comput. Sci.* — 1999. — Vol. 39. — P. 1076–1080.
4. *Vracko M.* A study of structure-carcinogenic potency relationship with artificial neural networks. The using of descriptors related to geometrical and electronic structures // *J. Chem. Inf. Comput. Sci.* — 1997. — Vol. 37. — P. 1037–1043.
5. *Benigni R., Giuliani A., Franke R., Gruska A.* Quantitative structure-activity relationships of mutagenic and carcinogenic aromatic amines // *Chem. Rev.* — 2000. — Vol. 100. — P. 3697–3714.
6. *Zakharov A., Lagunin A., Filimonov D., Poroikov V.* QSAR Modelling Of Rat's Carcinogenic Toxicity // *Fourth International Symposium Computational Methods in Toxicology and Pharmacology Integrating Internet Resources (CMTPI-2007) Abstract Book*. — М.: Russia, 2007. — P. 170.
7. Carcinogenic Potency Project (CPDB) // Режим доступа: <http://potency.berkeley.edu/cpdb.html>
8. *Cronin M.T.D., Jaworska J.S. et al.* Use of QSARs in International Decision-Making Frameworks to Predict Health Effects of Chemical Substances // *Environmental Health Perspectives*. — 2003. — Vol. 111. — № 10. — P. 1391–1401.
9. *Важев В.В., Алдабергенов М.К.* Прогнозирование технически важных свойств органических соединений по их ИК- и масс-спектрам // *Вестн. КазНУ. Сер. хим.* — 2007. — № 2. — С. 79–95.
10. *Важев В.В., Алдабергенов М.К., Важева Н.В.* Оценка канцерогенной активности органических веществ для крыс по ИК-спектрам // *Вестн. КазНУ. Сер. хим.* — 2005. — № 4 (40). — С. 84–89.
11. National Institute of Standards and Technology // Режим доступа: <http://www.nist.gov/index.html>
12. *Важев В.В.* Использование ИК- и масс-спектров в QSAR/QSPR- исследованиях. — Костанай: Изд-во КГУ, 2003. — 114 с.
13. *Важев В.В.* Прогнозирование ИК-спектров алкенов по их масс-спектрам // *Журн. структ. химии*. — 2005. — Т. 46. — № 2. — С. 251–256.
14. *Важев В.В., Бектурганова Г.К. и др.* Прогнозирование и экспертиза значений температуры вспышки углеводородов. I. Алкены // *Метрология*. — 2007. — № 4. — С. 13–18.