

БИОИНФОРМАТИКА

УДК 541.69+519.25+518.5

©Коллектив авторов

ЛИНЕЙНЫЕ И НЕЛИНЕЙНЫЕ КССА МОДЕЛИ ОСТРОЙ ТОКСИЧНОСТИ ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ ПРИ ИХ ВНУТРИВЕННОМ ВВЕДЕНИИ

О.А. Раевский^{1}, Е.А. Липлавская¹, А.В. Ярков¹, О.Е. Раевская¹, А.П. Ворте²*

¹Институт физиологически активных веществ РАН, Московская область,
г. Черноголовка, Северный проезд, 1, 142432; факс: 8(49652)49-508;
эл. почта: raevsky@ipac.ac.ru

²Institute for Health and Consumer Protection, European Commission - Joint Research
Centre, Via Enrico Fermi 2749, 21027 Ispra (Va), Italy.

Представлен КССА (количественная связь структура-активность) анализ острой токсичности для 68 монофункциональных органических соединений введенных мышам внутривенно. Указанные соединения относятся к семи классам: углеводороды (6 соединений), спирты (13), амиды (22), амины (12), эфиры (5), кетоны (7) и нитрилы (3). Предварительный анализ данных для этих соединений показал, что необходимо рассматривать не столько линейные соотношения между дескрипторами и токсичностью, сколько нелинейные. Для каждого из указанных классов химических соединений были рассмотрены как линейные, так и нелинейные КССА модели. Аналогичные модели были построены для всей выборки монофункциональных соединений. Статистические параметры и устойчивость нелинейных моделей оказались существенно лучше линейных соотношений. Замена параметра липофильности на дескрипторы, характеризующие молекулярный вес и способность образовывать водородные связи молекулы в рамках нелинейных моделей, позволила еще улучшить статистические критерии. Сделан вывод, что если для монофункциональных соединений наблюдается нелинейная зависимость внутривенной токсичности от липофильности, то можно ожидать, что характер этой зависимости для соединений, содержащих несколько функциональных групп, также может быть нелинейным. Это было проверено построением линейных, параболических и билинейных моделей для ряда небольших кластеров, содержащих структурно-родственные соединения с несколькими функциональными группами. Установлен целый ряд таких кластеров, где нелинейные КССА модели имеют лучшие статистические критерии по сравнению с линейными.

Ключевые слова: линейные и нелинейные КССА модели, HVBOT, острая токсичность по отношению к мышам.

* - адресат для переписки

ВВЕДЕНИЕ. Экспериментальное определение биологической активности и токсичности химических соединений сопровождается использованием огромного числа животных. В настоящее время существует острая потребность значительно уменьшить количество таких экспериментов за счёт использования альтернативных методов, таких как *in vitro* методы, количественные соотношения между структурой и активностью, информация о структурно-родственных соединениях и др. Необходимость применения подходов без использования животных, включая *in silico* методы, особенно подчеркивается в правилах и методах REACH (Registration, Evaluation, Authorisation, Restriction of Chemicals) [1, 2].

Ключевым аспектом в оценке токсикологических данных является определение острой токсичности по отношению к млекопитающим, в частности, грызунам. Острая токсичность описывает отрицательный эффект, причиненный введением (однократным или многократным в течение 24 часов) химического вещества в организм. Обычно, воздействие осуществляется через оральный, внутривенный, ингаляционный и кожный пути. В основе регуляторной оценки химических соединений лежит летальная доза для 50% подвергшихся воздействию животных (значения LD_{50}). Несмотря на широкое применение, этот параметр создает трудности при КССА моделировании. Оценка такого сложного явления как острая токсичность единственной величиной (LD_{50}) приводит к потере большей части информации. Кроме этого, уже имеющиеся данные широко варьируют вследствие определения их в разных лабораториях, к животным разного пола, возраста и условий содержания. Указанные факторы затрудняют процесс моделирования и могут объяснить факт наличия лишь относительно небольшого числа КССА моделей и экспертных систем для предсказания токсичности [3-5]. Недавний детальный анализ 150 опубликованных в литературе КССА моделей острой токсичности по отношению к мышам и крысам показал, что они имеют ограниченную пользу для предсказаний значений LD_{50} вследствие их плохого статистического качества или небольшого количества соединений, на основе которых они получены [6].

Данная статья посвящена оценке вклада различных функциональных групп органических соединений в острую токсичность по отношению к мышам и построению соответствующих линейных и нелинейных КССА моделей.

МЕТОДИКА. Для 68 химических соединений, содержащих одну функциональную группу, нами были отобраны из литературы [7-11] данные по острой токсичности при внутривенном введении мышам. Для органических соединений, содержащих несколько функциональных групп, данные были взяты из SYMYX Toxicity database [12]. В качестве количественной меры острой токсичности использовали значения $\log(1/LD_{50})$ (ммоль/кг веса). Здесь необходимо отметить, что выбранные данные содержали значения LD_{50} для мышей различного пола, возраста, условий содержания. Влияние этих факторов на токсичность может быть весьма существенно. В собранных нами литературных данных указанные различия в токсичности химических соединений по отношению к одному и тому же биологическому объекту при одном и том же способе введения иногда доходили до $\pm 0,50$ в логарифмической шкале $\log(1/LD_{50})$ (ммоль/кг).

Выбранные 68 соединений относились к семи классам: углеводороды (6 соединений), спирты (13), амиды (22), амины (12), эфиры (5), кетоны (7), нитрилы (3). Данные о токсичности и значения дескрипторов приведены в таблице 1.

Физико-химические дескрипторы рассчитывались программой HYBOT [13]. Для оценки коэффициентов распределения веществ в системе 1-октанол-вода использовалась программа DRAGON [14].

Построение небольших кластеров соединений, имеющих несколько функциональных групп, осуществлялось на основе индекса структурного сходства (индекс Танимото). Вначале отбиралось кластеробразующее соединение, затем все остальные соединения базы данных ранжировались в порядке уменьшения значения индекса Танимото по отношению к рассматриваемому соединению. И в кластер включались соединения, имеющие значение данного индекса $T_s \geq 0,50$ по отношению к рассматриваемому соединению.

Таблица 1. Физико-химические дескрипторы и экспериментальные значения острой внутривенной токсичности по отношению к мышам для 68 монофункциональных соединений.

N	Название	CAS	Молек. масса (MW)	α	$\Sigma C_{a(i)}$	MlogP	$\log(1/LD_{50})$ (ммоль/кг)
1	water	7732-18-5	18,02	1,411	1,76	-2,058	-3,14
2	methanol	67-56-1	32,04	3,246	1,63	-0,814	-2,17
3	ethanol	64-17-5	46,07	5,081	1,66	-0,172	-1,63
4	propanol	71-23-8	60,10	6,916	1,61	0,347	-1,06
5	2-methyl-1-propanol	78-83-1	74,12	8,751	1,57	0,8	-0,75
6	1-butanol	71-36-3	74,12	8,751	1,6	0,8	-0,71
7	3-methyl-1-butanol	123-51-3	88,15	10,586	1,62	1,209	-0,42
8	1-pentanol	71-41-0	88,15	10,586	1,61	1,209	-0,32
9	4-methyl-1-pentanol	1320-98-5	102,18	12,421	1,6	1,587	-0,22
10	1-hexanol	111-27-3	102,18	12,421	1,59	1,587	0,00
11	test-heptanol	28929-86-4	116,20	14,256	1,6	1,94	-0,13
12	1-octanol	111-87-5	130,23	16,091	1,61	3,186	0,28
13	4,7-methanindan, hexahydro-2- hydroxymethyl	64644-36-6	166,26	19,274	1,55	2,794	-0,03
14	1-aminobutane	109-73-9	73,14	9,465	2	0,8	-0,43
15	cyclohexanemethylamine	3218-02-8	113,20	14,196	2,01	1,548	0,31
16	1-aminooctane	111-86-4	129,25	16,805	1,99	3,186	0,86
17	N-methyl-N-butylamine	110-68-9	87,16	11,3	2,03	1,209	-0,15
18	N-methyl-N- cyclohexylethylamine	62141-38-2	141,26	17,866	1,98	2,199	0,52
19	N-methyl-N-(1-methyl- 2-cyclohexylethylamine	101-40-6	155,28	19,701	1,94	2,502	0,69
20	dihexylamine	143-16-8	185,35	24,145	1,99	3,467	1,27
21	trimethanamine	75-50-3	59,11	7,63	1,87	0,347	-0,18
22	N-methyl-2- azabicyclo(2.2.1)heptane	4524-95-2	111,19	13,422	1,93	1,548	0,30
23	1-ethylpiperidine	766-09-6	113,20	14,196	1,93	1,548	0,31
24	N-ethyl-2-azabicyclo (2.2.1)heptane	4492-38-0	125,21	15,257	1,98	1,882	0,84
25	1-methyl-2- propylpiperidine	35305-13-6	141,26	17,866	2	2,199	0,84
26	formamide	75-12-7	45,04	4,24	2,58	-1,434	-2,05
27	acetamide	60-35-5	59,07	6,075	2,89	-0,792	-2,23
28	N-methylformamide	123-39-7	59,07	6,075	2,69	-0,792	-1,43
29	N-methylacetamide	79-16-3	73,09	7,91	3	-0,273	-1,74
30	2-piperidone; delta-valerolactam	675-20-7	99,13	10,806	3,25	0,198	-0,78
31	epsilon-caprolactam	105-60-2	113,16	12,641	3,25	0,575	-0,63
32	N-cyclohexylformamide	766-93-8	127,19	14,476	2,14	0,929	-0,40
33	N-nonylacacetamide	45108-98-3	185,31	22,59	2,53	3,478	0,76

ЛИНЕЙНЫЕ И НЕЛИНЕЙНЫЕ КССА МОДЕЛИ

Таблица 1. Продолжение.

N	Название	CAS	Молек. масса (MW)	α	$\Sigma C_{a(i)}$	MlogP	$\log(1/LD_{50})$ (ммм/кг)
34	N-decylacetamide	16423-52-2	199,34	24,425	2,53	3,759	0,55
35	N-dodecylacetamide	3886-80-4	227,39	28,095	2,53	4,295	0,48
36	N,N-dimethyl formamide	68-12-2	73,09	7,91	2,87	-0,273	-1,53
37	N,N-dimethyl acetamide	127-19-5	87,12	9,745	3,11	0,18	-1,54
38	N,N-dimethyl propionamide	758-96-3	101,15	11,58	3,21	0,59	-0,91
39	N,N-dimethyl butyramide	760-79-2	115,18	13,415	2,94	0,967	-1,15
40	N,N-dimethyl valeramide	6225-06-5	129,20	15,25	2,94	1,321	-0,25
41	N,N-dimethyl hexanamide	5830-30-8	143,23	17,085	2,94	1,654	0,20
42	N,N-dimethyl decanamide	14433-76-2	199,34	24,425	2,94	3,759	0,70
43	N,N-dimethyl dodecanamide	3007-53-2	227,39	28,095	2,94	4,295	0,48
44	N,N-diethyl dodecanamide	3352-87-2	255,44	31,765	3,25	4,8	-0,05
45	N,N-dimethyl tetradecanamide	3015-65-4	255,44	31,765	2,94	4,8	0,23
46	N,N-dimethyl hexadecanamide	3886-91-7	283,50	35,435	2,94	5,281	0,11
47	N,N-dimethyl octadecanamide	3886-90-6	311,55	39,105	2,94	5,741	0,31
48	diethyl ether	60-29-7	74,12	8,751	1,54	0,8	-1,13
49	dipropyl ether	111-43-3	102,18	12,421	1,5	1,587	-0,30
50	dihexyl ether	629-64-1	214,39	27,101	1,53	4,002	-0,34
51	dioctyl ether	629-82-3	242,45	30,771	1,53	5,42	-0,69
52	ethylcyclohexanacetate	5452-75-5	170,25	18,778	1,65	2,289	0,48
53	acetonitrile	75-05-8	41,05	4,469	1,73	-0,317	-1,17
54	2-methylheptanonitrile	54321-41-4	125,21	15,479	1,65	2,129	0,35
55	1-heptanenitrile, 2-methyl-2-pentyl	49827-48-7	195,35	24,654	1,68	3,594	0,04
56	3-pentanone	96-22-0	86,13	10,103	1,88	1,064	-0,77
57	2-norbornanone	497-38-1	110,16	12,225	1,88	1,403	-0,21
58	3-methyl-2-octanone	6137-08-2	142,24	17,443	1,96	2,446	-0,10
59	5-nonanone	502-56-7	142,24	17,443	1,96	2,446	0,01
60	1-camphor	464-48-2	152,24	17,73	1,9	2,357	-0,32
61	6-undecanone	927-49-1	170,30	21,113	1,96	3,041	0,16
62	8-pentadecanone	818-23-5	226,40	28,453	1,96	5,025	-0,72
63	heptane	142-82-5	100,20	13,619	0	3,869	-0,35
64	undecane	1120-21-4	156,31	20,959	0	5,115	-0,52
65	pentane	109-66-0	72,15	9,949	0	3,138	-0,79
66	nonane	111-84-2	128,26	17,289	0	4,52	-0,23
67	tridecane	629-50-5	184,37	24,629	0	5,668	-0,80
68	pentadecane	629-62-9	212,42	28,299	0	6,187	-1,22

Таблица 2. Линейные и нелинейные КССА модели для соединений, содержащих несколько химических функциональных групп.

1 - $\log(1/LD_{50})_{\text{exper}} = A_0(\pm \text{er}_{A_0}) + A_1(\pm \text{er}_{A_1}) \times \text{MlogP}; 2 - \log(1/LD_{50})_{\text{exper}} = A_0(\pm \text{er}_{A_0}) + A_1(\pm \text{er}_{A_1}) \times \text{MlogP} + A_2(\pm \text{er}_{A_2}) \times \text{MlogP}^2;$
 3 - $\log(1/LD_{50})_{\text{exper}} = A_0(\pm \text{er}_{A_0}) + A_1(\pm \text{er}_{A_1}) \times \text{MlogP} + A_2(\pm \text{er}_{A_2}) \times \log(\text{Beta} \times P + 1)$

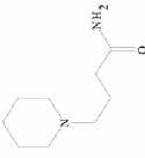
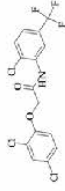
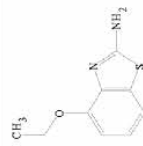
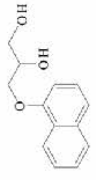
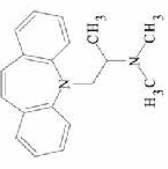
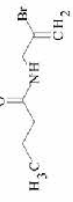
Кластеробразующее соединение	Модель	N	R ²	SD	F	Q ²	SD _{cv}	A ₀	er _{A0}	A ₁	er _{A1}	A ₂	er _{A2}	exp	calc	MlogP	Beta
	1	26	0,134	0,72	3,7	-0,330	0,89	-0,19	0,26	0,22	0,11			-0,34	-0,02		
	2	26	0,686	0,44	25,1	0,449	0,59	-0,87	0,19	1,30	0,18	-0,27	0,04	-0,34	-0,05	0,743	
	3	26	0,686	0,44	25,1	0,579	0,51	-0,69	0,18	0,64	0,10	-2,74	0,43	-0,34	-0,22		3,0709 E-04
	1	14	0,326	0,26	5,8	-0,206	0,34	-0,20	0,25	0,22	0,09			0,10	0,77		
	2	14	0,556	0,22	6,9	0,386	0,26	-1,87	0,73	1,38	0,49	-0,19	0,08	0,10	0,52	4,409	
	3	14	0,555	0,22	6,9	0,331	0,27	-7,98	3,27	7,78	3,17	-7,78	3,26	0,10	0,59		7,9019 E-02
	1	14	0,433	0,25	9,2	0,229	0,29	-0,62	0,30	0,48	0,16			0,39	0,24		
	2	14	0,643	0,21	9,9	0,477	0,25	-3,61	1,20	4,11	1,43	-1,04	0,41	0,39	0,41	1,783	
	3	14	0,647	0,2	10,1	0,480	0,25	-4,92	1,68	6,92	2,50	-7,53	2,92	0,39	0,41		1,2390 E-01
	1	23	0,523	0,43	23,0	0,335	0,50	-0,33	0,15	0,33	0,07			0,39	0,25		
	2	23	0,721	0,33	25,8	0,628	0,38	-0,46	0,12	0,69	0,11	-0,11	0,03	0,39	0,42	1,765	
	3	23	0,737	0,32	28,1	0,691	0,35	-0,52	0,12	0,83	0,13	-0,86	0,21	0,39	0,49		4,0931 E-02
	1	26	0,569	0,18	31,8	0,395	0,21	-0,04	0,15	0,24	0,04			0,93	0,83		
	2	26	0,774	0,13	39,3	0,688	0,16	-1,14	0,27	1,05	0,18	-0,13	0,03	0,93	0,96	3,652	
	3	26	0,773	0,13	39,2	0,702	0,15	-0,82	0,21	0,63	0,09	-0,71	0,16	0,93	0,92		1,1520 E-03
	1	36	0,293	0,39	14,1	0,159	0,42	0,10	0,14	0,15	0,04			1,06	0,36		
	2	36	0,526	0,32	18,3	0,401	0,36	-0,43	0,18	0,62	0,12	-0,08	0,02	1,06	0,40	1,707	
	3	36	0,743	0,24	47,6	0,688	0,26	-0,79	0,14	2,12	0,26	-2,12	0,28	1,06	0,61		1,9849 E-01

Таблица 2. Продолжение.

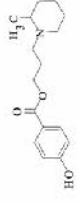
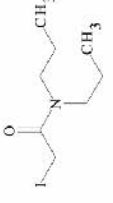



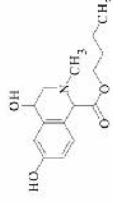
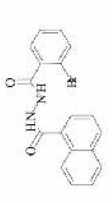





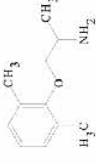
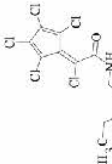
Кластеробразующее соединение	Мо-дель	N	R ²	SD	F	Q ²	SD _{cv}	A ₀	er _{A₀}	A ₁	er _{A₁}	A ₂	er _{A₂}	exp	calc	MlogP	Beta
	1	86	0,329	0,34	41,1	0,256	0,36	0,19	0,11	0,22	0,03	-0,08	0,01	1,19	0,75		
	2	86	0,525	0,29	45,9	0,428	0,32	-0,43	0,14	0,69	0,09	-0,08	0,01	1,19	0,80	2,500	6,8993 E-04
	3	86	0,473	0,29	37,2	0,430	0,30	-0,27	0,14	0,46	0,06	-0,60	0,11	1,19	0,83		
	1	10	0,009	0,98	0,07	-0,85	1,34	-0,14	0,49	0,05	0,18			1,28	-0,03		
	2	10	0,761	0,52	11,2	0,392	0,82	-0,77	0,29	1,25	0,27	-0,26	0,05	1,28	0,73	2,275	
	3	10	0,799	0,47	13,9	0,639	0,63	-0,76	0,26	0,66	0,15	9,07	1,73	1,28	0,74		-7,9100 E-06
	1	20	0,64	0,44	31,9	0,536	0,50	-0,86	0,32	0,56	0,10			1,33	1,04		
	2	20	0,792	0,35	32,4	0,713	0,41	-2,86	0,62	2,16	0,46	-0,28	0,08	1,33	1,24	3,386	
	3	20	0,824	0,32	39,7	0,782	0,36	-1,66	0,30	0,94	0,11	-4,10	0,97	1,33	1,34		4,5230 E-05
	1	17	0,298	0,89	6,38	-0,30	1,21	1,41	0,42	-0,25	0,10			1,68	0,98		
	2	17	0,626	0,67	11,7	0,414	0,84	2,88	0,53	-1,07	0,25	0,08	0,02	1,68	1,26	1,737	
	3	17	0,623	0,68	11,6	0,425	0,84	2,34	0,42	-0,59	0,12	0,91	0,26	1,68	1,32		4,9998 E-06
	1	22	0,354	0,69	10,9	0,161	0,79	-0,09	0,26	0,54	0,16			1,89	0,42		
	2	22	0,595	0,56	14,0	0,511	0,62	-0,28	0,22	1,53	0,32	-0,43	0,13	1,89	0,78	0,939	
	3	22	0,602	0,56	14,4	0,524	0,61	0,03	0,21	2,36	0,55	-2,55	0,74	1,89	0,81		3,0519 E-01
	1	41	0,295	0,43	16,9	0,152	0,48	0,31	0,11	0,17	0,04			0,24	0,51		
	2	41	0,533	0,36	21,7	0,473	0,38	-0,06	0,12	0,54	0,09	-0,06	0,01	0,24	0,50	1,195	
	3	41	0,544	0,35	22,6	0,485	0,38	-0,03	0,12	0,41	0,06	-0,52	0,11	0,24	0,46		7,7996 E-04
	1	13	0,182	0,28	2,5	-0,22	0,35	-0,24	0,40	0,19	0,12			0,31	0,57		
	2	13	0,607	0,21	7,7	0,447	0,24	-4,00	1,18	2,71	0,77	-0,40	0,12	0,31	0,24	4,324	
	3	13	0,63	0,2	8,5	0,479	0,24	-6,63	1,86	3,72	1,02	-4,08	1,17	0,31	0,24		8,5585 E-03

Таблица 2. Продолжение.

Мо- дель	Кластероброу- ющее соединение	N	R ²	SD	F	Q ²	SD _{cv}	A ₀	er ₋ A ₀	A ₁	er ₋ A ₁	A ₂	er ₋ A ₂	exp	calc	MlogP	Beta
1		12	0,193	0,52	2,4	-0,31	0,67	-0,22	0,36	0,16	0,10			0,52	0,27		
2		12	0,799	0,28	17,8	0,397	0,48	-1,96	0,39	1,49	0,26	-0,20	0,04	0,52	0,75	3,163	
3		12	0,955	0,13	96,4	0,916	0,18	-3,69	0,29	2,94	0,23	-3,19	0,26	0,52	0,76		2,2058 E-02
1		20	0,216	0,24	5,0	-0,09	0,28	0,29	0,28	0,16	0,07			1,01	0,88		
2		20	0,609	0,17	13,2	0,366	0,22	-2,97	0,82	1,97	0,44	-0,24	0,06	1,01	1,04	3,751	
3		20	0,652	0,16	15,9	0,450	0,21	-0,70	0,29	0,47	0,08	-2,39	0,52	1,01	0,99		1,2756 E-05
1		16	0,783	0,23	50,4	0,697	0,27	-0,25	0,14	0,33	0,05			1,21	0,52		
2		16	0,926	0,14	81,6	0,848	0,20	-0,81	0,14	1,03	0,14	-0,15	0,03	1,21	0,77	2,303	
3		16	0,934	0,13	92,0	0,886	0,17	-0,56	0,10	0,58	0,05	-1,04	0,19	1,21	0,72		6,0086 E-04
1		16	0,595	0,48	20,6	0,430	0,57	-0,58	0,19	0,38	0,08			0,83	0,35		
2		16	0,836	0,32	33,1	0,757	0,38	-0,77	0,13	0,89	0,13	-0,14	0,03	0,83	0,57	2,467	
3		16	0,859	0,29	39,4	0,801	0,35	-0,85	0,13	0,91	0,12	-0,99	0,20	0,83	0,59		1,8606 E-02
1		12	0,135	0,33	1,60	-0,772	0,55	-0,05	0,42	0,21	0,17			0,83	0,35		
2		12	0,72	0,2	11,5	0,539	0,25	-2,50	0,62	2,83	0,61	-0,63	0,14	0,83	0,61	1,921	
3		12	0,722	0,2	11,7	0,570	0,25	-1,76	0,47	1,54	0,32	-2,67	0,61	0,83	0,62		7,7797 E-03
1		12	0,184	0,38	2,3	-0,50	0,52	0,05	0,37	0,28	0,19			0,89	0,67		
2		12	0,651	0,26	8,4	0,506	0,31	-2,03	0,65	2,66	0,70	-0,62	0,18	0,89	0,82	2,224	
3		12	0,654	0,26	8,5	0,541	0,30	-1,22	0,44	1,30	0,32	-2,83	0,81	0,89	0,85		5,6882 E-03
1		12	0,404	0,58	6,8	-0,19	0,83	-0,37	0,32	0,42	0,16			0,34	1,30		
2		12	0,728	0,42	12,1	0,595	0,51	-0,45	0,23	1,16	0,25	-0,30	0,09	0,34	-0,51	3,919	
3		12	0,733	0,41	12,3	0,585	0,51	-0,55	0,23	1,21	0,26	-1,68	0,50	0,34	-0,08		4,2101 E-02

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ.

КССА модели монофункциональных химических соединений. На рисунке 1 представлена зависимость токсичности от молекулярной массы для всех исследуемых 68 соединений.

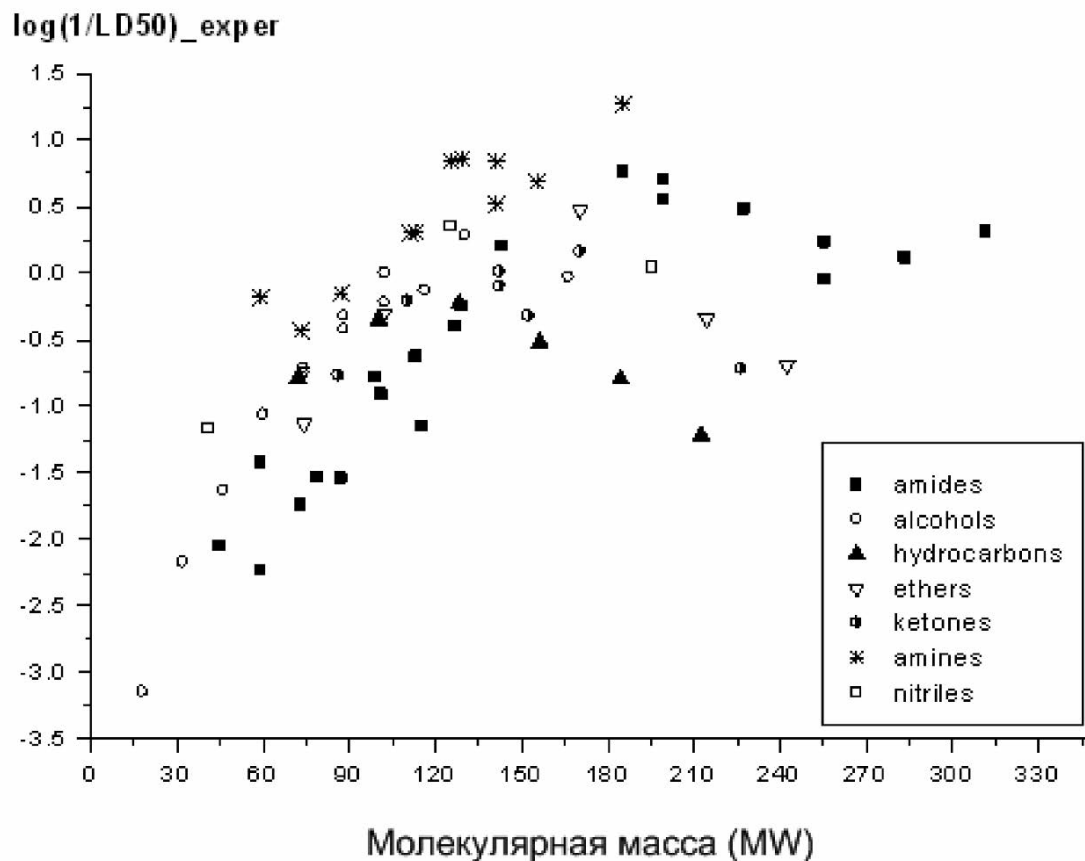


Рисунок 1.

Графическое представление зависимости внутривенной токсичности по отношению к мышам от молекулярной массы.

Предварительный анализ данных как для всех исследуемых соединений, так и для отдельных классов показывает, что необходимо рассматривать не только традиционные линейные корреляции между дескрипторами и токсичностью, но и нелинейные соотношения. Здесь необходимо упомянуть, что нелинейные модели взаимосвязи структура - биологическая активность, были представлены уже в первых публикациях пионеров QSAR Ганча (Hansch) [15] и МакФарланда (McFarland) [16], и были развиты Кубиньи (Kubinyi) [17]. Такой вид соотношений может быть результатом многих физиологических процессов в биологических системах, включая кинетический контроль транспорта веществ, равновесный контроль распределения веществ, стерический фактор, различную фармакокинетику, метаболизм, растворимость и другие эффекты [17]. С этой точки зрения нелинейные соотношения должны рассматриваться в качестве общей платформы при КССА моделировании. А линейные соотношения могут проявляться для небольших рядов структурно-родственных соединений в рамках узкого интервала свойств (активностей).

КССА для амидов. Мы начали нашу работу с рассмотрения данных для самого большого ряда из рассматриваемых веществ – амидов. Указанная выборка содержала 22 соединения. Лучшие однопараметровые соотношения

на основе молекулярной массы (MW), молекулярной поляризуемости (α) и рассчитанного коэффициента распределения в системе 1-октанол-вода (MlogP) представлены ниже:

$$\log(1/LD_{50})_{\text{exper}} = -2,00(\pm 0,25) + 0,00984(\pm 0,00147) \text{ MW} \quad (1)$$

$n=22, R^2=0,692, sd=0,54, Q^2=0,620, sd_{cv}=0,60, F=45,0$

$$\log(1/LD_{50})_{\text{exper}} = -1,86(\pm 0,24) + 0,075(\pm 0,011) \alpha \quad (2)$$

$n=22, R^2=0,689, sd=0,54, Q^2=0,618, sd_{cv}=0,61, F=44,3$

$$\log(1/LD_{50})_{\text{exper}} = -1,20(\pm 0,14) + 0,36(\pm 0,05) \text{ MlogP} \quad (3)$$

$n=22, R^2=0,758, sd=0,48, Q^2=0,707, sd_{cv}=0,53, F=62,7,$

где n - число соединений, включенных в корреляцию; R – коэффициент корреляции, sd – стандартное отклонение, Q – коэффициент корреляции в процедуре перекрестного контроля (cross-validation) с исключением по одному, sd_{cv} – стандартное отклонение в процедуре перекрестного контроля (cross-validation) с исключением по одному, F – критерий Фишера. В скобках указаны ошибки оценки параметров в пределах 95% доверительного интервала.

Необходимо заметить, что указанные три дескриптора являются высокозакоррелированными ($R^2 \approx 0,98$). И поэтому не могут одновременно быть использованными в рамках двух- или трёхпараметровых уравнений.

В то же время корреляции с теми же дескрипторами для 12 амидов, имеющих $MW < 130$ существенно лучше:

$$\log(1/LD_{50})_{\text{exper}} = -3,08(\pm 0,27) + 0,0206(\pm 0,0029) \text{ MW} \quad (4)$$

$n=12, R^2=0,835, sd=0,27, Q^2=0,760, sd_{cv}=0,33, F=50,5$

$$\log(1/LD_{50})_{\text{exper}} = -2,83(\pm 0,25) + 0,160(\pm 0,024) \alpha \quad (5)$$

$n=12, R^2=0,817, sd=0,29, Q^2=0,739, sd_{cv}=0,34, F=44,5$

$$\log(1/LD_{50})_{\text{exper}} = -1,29(\pm 0,10) + 0,66(\pm 0,12) \text{ MlogP} \quad (6)$$

$n=12, R^2=0,745, sd=0,34, Q^2=0,641, sd_{cv}=0,41, F=29,2$

Для оставшихся 10 амидов, имеющих $MW > 130$, лучшим дескриптором оказалась сумма акцепторных факторов водородной связи ($\Sigma Ca(o)$):

$$\log(1/LD_{50})_{\text{exper}} = 2,61(\pm 0,78) - 0,78(\pm 0,27) \Sigma Ca(o) \quad (7)$$

$n=10, R^2=0,516, sd=0,19, Q^2=0,311, sd_{cv}=0,23, F=8,5$

При анализе вышеуказанных уравнений становится очевидным тот факт, что для амидов, имеющих $MW < 130$, вес молекулы вносит положительный вклад в токсичность и является доминирующим фактором. Для амидов, имеющих $MW > 130$, доминирующей является акцепторная способность водородной связи, которая уменьшает величину токсичности. Необходимо заметить, что явление липофильности определяется двумя факторами: стерическим взаимодействием и водородным связыванием, действующими на токсичность в противоположных направлениях [18]. Поэтому нами были опять объединены 22 рассматриваемых амида и построены двухпараметровая линейная модель с участием MW и $\Sigma Ca(o)$ (взаимная корреляция этих дескрипторов чрезвычайно мала) и параболические модели с использованием только одного дескриптора. Соответствующие уравнения для всех исследуемых амидов представлены ниже:

$$\log(1/LD_{50})_{\text{exper}} = -0,52(\pm 1,22) + 0,010(\pm 0,01) MW - 0,52(\pm 0,14) \Sigma Ca(o) \quad (8)$$

$n=22, R^2=0,716, sd=0,53, Q^2=0,631, sd_{cv}=0,62, F=23,9$

ЛИНЕЙНЫЕ И НЕЛИНЕЙНЫЕ КССА МОДЕЛИ

По сравнению с однопараметровым уравнением статистические критерии последнего уравнения улучшились незначительно. Переход же к параболическим моделям на основе рассматриваемых трёх дескрипторов привели к гораздо лучшим результатам.

$$\log(1/LD_{50})_{\text{exper}} = -3,83(\pm 0,31) + 0,03753(\pm 0,00422)MW - 0,000081(\pm 0,000012)MW^2 \quad (9)$$

$n=22, R^2=0,907, sd=0,30, Q^2=0,885, sd_{cv}=0,33, F=195,3$

Статистические параметры параболической модели, описываемой уравнением (9), значительно лучше по сравнению со статистическими параметрами соответствующего линейного уравнения (6).

Необходимо заметить, что нелинейные корреляции с молекулярной массой и молекулярной поляризуемостью также имеют очень хорошие статистические параметры:

$$\log(1/LD_{50})_{\text{exper}} = -1,18(\pm 0,09) + 0,78(\pm 0,09) M\log P - 0,094(\pm 0,19) (M\log P)^2 \quad (10)$$

$n=22, R^2=0,892, sd=0,33, Q^2=0,849, sd_{cv}=0,39, F=78,2$

$$\log(1/LD_{50})_{\text{exper}} = -3,37(\pm 0,26) + 0,273(\pm 0,030)\alpha - 0,0048(\pm 0,0007) \alpha^2 \quad (11)$$

$n=22, R^2=0,907, sd=0,30, Q^2=0,886, sd_{cv}=0,33, F=194,0$

В ряду рассматриваемых 22 амидов изменения структуры связаны преимущественно с изменением молекулярной массы, которая вносит существенный вклад в неспецифическую токсичность. Модели (9-11) имеют близкие статистические критерии. Однако модель, основанная на использовании $M\log P$, представляется предпочтительней с позиции интерпретируемости транспорта веществ в организме [15-17].

КССА для спиртов. В данную выборку были включены 12 спиртов и вода. Интервал токсичности для них равен 3,42 логарифмическим единицам.

Однопараметровые корреляции с физико-химическими дескрипторами имеют довольно хорошие статистические параметры и являются сходными с соответствующими уравнениями для амидов:

$$\log(1/LD_{50})_{\text{exper}} = -2,60(\pm 0,32) + 0,0213(\pm 0,0034) MW \quad (12)$$

$n=13, R^2=0,778, sd=0,48, Q^2=0,575, sd_{cv}=0,71, F=38,7$

$$\log(1/LD_{50})_{\text{exper}} = -2,55(\pm 0,28) + 0,176(\pm 0,025) \alpha \quad (13)$$

$n=13, R^2=0,820, sd=0,44, Q^2=0,666, sd_{cv}=0,61, F=51,9$

$$\log(1/LD_{50})_{\text{exper}} = -1,43(\pm 0,10) + 0,67(\pm 0,06) M\log P \quad (14)$$

$n=13, R^2=0,921, sd=0,29, Q^2=0,866, sd_{cv}=0,38, F=129,1$

Однако параболическая модель, основанная на параметре липофильности, имеет ещё лучшие статистические критерии:

$$\log(1/LD_{50})_{\text{exper}} = -1,28(\pm 0,06) + 0,78(\pm 0,04) M\log P - 0,093(\pm 0,018) (M\log P)^2 \quad (15)$$

$n=13, R^2=0,978, sd=0,16, Q^2=0,942, sd_{cv}=0,32, F=226,6$

В случае кетонов линейные модели острой внутривенной токсичности не имели достаточно значимых статистических критериев. Однако для этого небольшого ряда соединений была выявлена явная параболическая зависимость:

$$\log(1/LD_{50})_{\text{exper}} = -1,58(\pm 0,40) + 1,06(\pm 0,29) M\log P - 0,176(\pm 0,046) (M\log P)^2 \quad (16)$$

$n=7, R^2=0,785, sd=0,20, Q^2=0,682, sd_{cv}=0,22, F=7,3$

Другие подвыборки не имеют достаточного количества соединений для КССА моделирования. Тем не менее, параболическая зависимость токсичности от молекулярной массы для углеводородов очевидна из рисунка 1.

На рисунке 2 представлена графическая зависимость токсичности от параметра MlogP для 63 соединений.

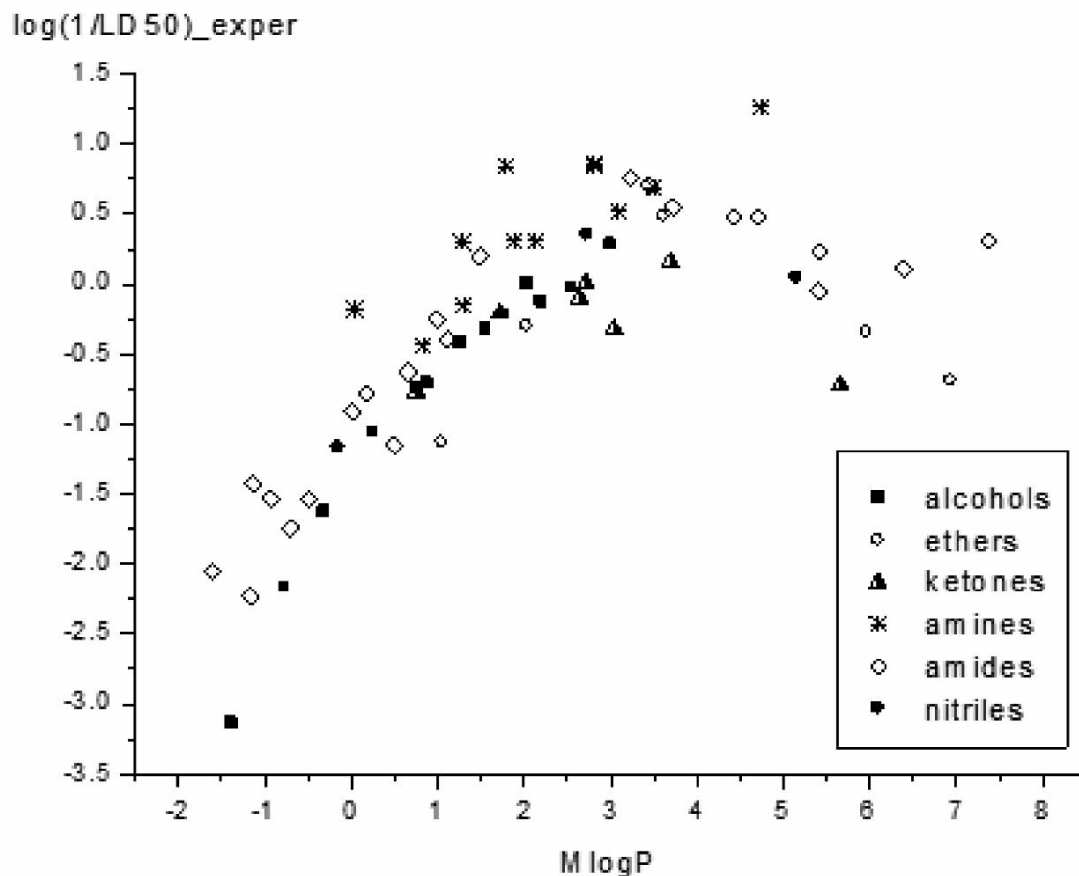


Рисунок 2.

Графическая зависимость внутривенной токсичности по отношению к мышам от MlogP.

Нелинейный характер данной зависимости позволяет построить объединённую параболическую КССА модель для этих разнообразных соединений:

$$\log(1/LD_{50})_{\text{exper}} = -1,11(\pm 0,08) + 0,87(\pm 0,06) \text{ MlogP} - 0,134(\pm 0,010) (\text{MlogP})^2 \quad (17)$$

$n=68$, $R^2=0,755$, $sd=0,43$, $Q^2=0,729$, $sd_{cv}=0,45$, $F=100,1$

В случае использования доступных экспериментальных данных по коэффициенту распределения в системе октанол-вода ($\log P_{o/w, \text{exp}}$) полученное уравнение является сходным с уравнением (17):

$$\log(1/LD_{50})_{\text{exper}} = -1,10(\pm 0,09) + 0,69(\pm 0,08) (\log P_{o/w, \text{exp}}) - 0,104(\pm 0,020) (\log P_{o/w, \text{exp}})^2 \quad (18)$$

$n=30$, $R^2=0,775$, $sd=0,42$, $Q^2=0,773$, $sd_{cv}=0,44$, $F=46,4$

Это свидетельствует о возможности использования параметра MlogP для расчёта значений токсичности разнообразных органических соединений вместо экспериментально определенных значений коэффициентов распределения в системе октанол-вода.

Кроме вышеуказанной параболической модели (ур. (17)) для всего ряда соединений была также построена и билинейная модель, основанная на том же рассчитанном параметре липофильности $M\log P$ (ур. (19)):

$$\log(1/LD_{50})_{\text{exper}} = -1,27(\pm 0,08) + 0,81(\pm 0,06)M\log P - 1,20(\pm 0,11)\log(0,0031844 \cdot P + 1) \quad (19)$$

$n=68, R^2=0,768, sd=0,42, Q^2=0,743, sd_{cv}=0,44, F=107,3$

Статистические критерии уравнений (17, 19) являются практически одинаковыми и весьма хорошими, а стандартное отклонение находится на уровне ошибки экспериментального определения.

Как уже было отмечено, $\log P$ в системе октанол-вода хорошо описывается с помощью молекулярной поляризуемости и акцепторного фактора водородной связи. Поэтому мы решили заменить параметр липофильности на физико-химические дескрипторы, которые вносят вклад в распределение соединений в системе октанол-вода. В случае параболической модели помимо указанных двух дескрипторов необходимо вводить три дополнительных дескриптора: квадрат молекулярной поляризуемости (α^2), квадрат акцепторного фактора водородной связи (ΣCa_o)², и произведение молекулярной поляризуемости на акцепторный фактор водородной связи ($\alpha \cdot \Sigma Ca_o$). Полученное уравнение представлено ниже:

$$\log(1/LD_{50})_{\text{exper}} = -2,03(\pm 0,36) + 0,215(\pm 0,024) \alpha - 0,17(\pm 0,24) \Sigma Ca_o - 0,0069(\pm 0,0006) \alpha^2 - 0,162(\pm 0,049) (\Sigma Ca_o)^2 + 0,047(\pm 0,008) \alpha \cdot \Sigma Ca_o \quad (20)$$

$n=68, R^2=0,815, sd=0,37, Q^2=0,802, sd_{cv}=0,38, F=289,5$

Так как ошибка в расчётах коэффициента при ΣCa_o находится на уровне значения коэффициента, данный дескриптор может быть исключён из этой корреляции:

$$\log(1/LD_{50})_{\text{exper}} = -2,23(\pm 0,22) + 0,220(\pm 0,023) \alpha - 0,0068(\pm 0,0006) \alpha^2 - 0,190(\pm 0,030) (\Sigma Ca_o)^2 + 0,043(\pm 0,005) \alpha \cdot \Sigma Ca_o \quad (21)$$

$n=68, R^2=0,812, sd=0,37, Q^2=0,800, sd_{cv}=0,38, F=286,0$

Последнее уравнение имеет лучшие статистические критерии по сравнению с уравнениями (17, 18, 19) и позволяет выявить положительный по знаку вклад стерических взаимодействий и отрицательный вклад водородного связывания в нелинейную зависимость внутривенной токсичности по отношению к мышам для соединений, имеющих углеводородный остов и какую-либо одну функциональную группу.

Токсичность химических функциональных групп. Зависимость токсичности исследуемых соединений от молекулярного веса даёт возможность приблизительно оценить сравнительные вклады одиночных функциональных групп в различных интервалах молекулярной массы. Например, таблица 1 содержит данные для трёх самых небольших соединений в соответствующих рядах: метанол ($\log(1/LD_{50})_{\text{exper}} = -2,14$), ацетамид (-2,23), ацетонитрил (-1,17). Эти данные показывают, что амидная и гидроксильная группы являются скорее слаботоксичными, а токсичность нитрильной группы на одну логарифмическую единицу больше. В интервале MW 72,15-74,12 имеется шесть соединений: 2-метил-1-пропанол ($\log(1/LD_{50})_{\text{exper}} = -0,75$), бутанол (-0,71), 1-аминобутан (-0,43), N,N-диметилформамид (-1,53), диэтиловый эфир (-1,13), пентан (-0,79). Очевидно, что в данном подряде более токсичным является аминобутан. Токсичность двух спиртов и углеводорода является несколько меньшей, эфир ещё менее токсичен, и наименьшей токсичностью в этом ряду обладает амид. Шесть соединений, имеющих молекулярную массу в интервале 125,21-130,23, в порядке уменьшения токсичности располагаются следующим образом: 1-аминооктан ($\log(1/LD_{50})_{\text{exper}} = 0,86$), 2-метилгептанитрил (0,35), октанол (0,28),

нонан (-0,23), N,N-диметилвалерамил (-0,25), N-циклогексилформамид (-0,40). В данном ряду соединений аминогруппа также оказалась наиболее токсичной, а амидная – наименее. Однако гидроксильное производное оказалось более токсичным по сравнению с углеводородом. Токсичность гидроксильного производного здесь приблизительно такая же, как и у нитрила. Что касается соединений, имеющих молекулярную массу MW=200, токсичность N-додецилацетамида (MW=227,39, $\log(1/LD_{50})_{\text{exper}} = 0,48$) уже существенно больше по сравнению с дигептиловым эфиром (MW=214,39, $\log(1/LD_{50})_{\text{exper}} = -0,34$). Вероятно, это является результатом некоторого различия в форме нелинейных соотношений между токсичностью и физико-химическими дескрипторами. Конечно, такая сложная картина создает трудности в построении общей КССА модели и её интерпретации.

КССА модели острой токсичности органических соединений, содержащих несколько функциональных групп. Наблюдаемая нелинейная зависимость внутривенной токсичности химических соединений с одной функциональной группой должна учитываться при КССА моделировании токсичности и для соединений с несколькими функциональными группами. Поэтому мы рассмотрели для небольших кластеров структурно-родственных соединений из базы данных SYMUX Toxicity database [12] как линейные, так и нелинейные модели взаимосвязи острой внутривенной токсичности от параметра липофильности, который рассматривается в качестве дескриптора, связанного с транспортом веществ в организме через ряд биологических мембран. Таблица 2 содержит яркие примеры таких кластеров, где преимущество нелинейных моделей по сравнению с линейными является очевидным.

ЗАКЛЮЧЕНИЕ. Очевидно, что для предсказания такого сложного многофакторного явления как токсичность целесообразно использование как линейных, так и нелинейных соотношений структура-токсичность.

В качестве стратегии КССА моделирования комбинирование линейных соотношений между острой токсичностью и структурными дескрипторами для небольших рядов соединений и нелинейных их соотношений для больших рядов с большими интервалами значений токсичности представляется более обоснованным по сравнению с использованием моделей, основанных только на линейных соотношениях. Линейные модели могут быть использованы как локальные модели для определения токсичности специфических функциональных групп, тогда как нелинейные модели могут быть использованы для гетерогенных рядов соединений или соединений, содержащих несколько функциональных групп.

Работа выполнена при частичной финансовой поддержке МНТЦ (проект №3777).

ЛИТЕРАТУРА

1. EC (2006). Regulation (EC) No. 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No. 793/93 and Commission Regulation (EC) No. 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Official Journal of the European Union L 396/1 of 30.12.2006. Office for Official Publications of the European Communities (OPOCE), Luxembourg.

2. ECHA (2008). Guidance on Information Requirements and Chemical Safety Assessment. Chapter R6. European Chemicals Agency, Helsinki, Finland. Available from: http://win.mail.ru/cgi-bin/link?check=1&url=http%3A%2F%2Fguidance.echa.europa.eu%2Fdocs%2Fguidance_document%2Finformation_requirements_en.htm%3Ftime%3D1252064523%23r6http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_en.htm?time=1252064523#r6
3. *Tsakovska I., Lessigiarska L., Netzeva T., Worth A.* (2008) *QSAR & Combi. Sci.*, **27**, 41-48.
4. *Jäckel H., Klein W.* (1991) *Quant. Struct.-Act. Relat.* **10**, 198-204.
5. *Juranic I., Drakulic B., Petrovic S., Mijin D., Stankovic M.* (2006) *Chemosphere*, **62**, 641-649.
6. *Devillers J., Devillers H.* (2009) *SAR QSAR in Environ. Res.*, **20**, 467-500.
7. *Tichy M., Trcka V., Roth Z., Krivucova M.* (1985) *Environ. Health Perspect.*, **61**, 321-328.
8. *Wittliff J.* (2005) in: *Encyclopedia of Toxicology* (Wexler P., ed.) Acad Press, 2Ed, pp. 377-381.
9. *Wiles J., Narcisse J.* (1971) *Am. Indust. Hyg. Assoc.* **32**, 539-545.
10. *Di Paolo Th.* (1978) *J. Pharm. Sc.* **67**, 566-568.
11. *Farrant J.* (1964) *J. Pharmacol.* **16**, 472-483.
12. SYMYX Toxicity database. Available at <http://www.symyx.com/products/databases/index.jsp>
13. *Raevsky O.* (1997) in: *Computer-Assisted Lead Finding and Optimization* (Waterbeemd H., Testa B., Folkers G., eds.) Verlag, Basel, pp. 367-378.
14. DRAGON. Talete srl, Italy. Software available at <http://win.mail.ru/cgi-bin/link?check=1&url=http%3A%2F%2Fwww.talete.mi.it%2Fdragon.htm>
<http://www.talete.mi.it/dragon.htm>
15. *Hansch C., Fujita T.* (1964) *J. Am. Chem. Soc.*, **86**, 1616-1626.
16. *McFarland J.* (1970) *J. Med. Chem.*, **13**, 1192-1186.
17. *Kubinyi H.* (1977) *J. Med. Chem.*, **20**, 625-629.
18. *Raevsky O.* (2004) *Med. Chem.*, **4**, 1041-1052.

Поступила: 14. 09. 2010.

LINEAR AND NONLINEAR QSAR MODELS OF ACUTE INTRAVENOUS TOXICITY TO MICE FOR ORGANIC CHEMICALS

O.A. Raevsky¹, E.A. Liplavskaya¹, A.V. Yarkov¹, O.E. Raevskaya¹, A.P. Worth²

¹Department of Computer-Aided Molecular Design, Institute of Physiologically Active Compounds of Russian Academy of Sciences, Chernogolovka, Moscow region, 142432 Russia;
e-mail: raevsky@ipac.ac.ru

²Institute for Health and Consumer Protection, European Commission - Joint Research Centre, Via Enrico Fermi 2749, 21027 Ispra (Va), Italy

QSAR analysis of acute intravenous toxicity to mice for 68 monofunctional chemicals is presented. There compounds represents seven classes of organic chemicals: hydrocarbons (6 chemicals), alcohols (13), amides (22), amines (12), ethers (5), ketones (7), nitriles (3). Preliminary consideration of data for these chemicals showed that it is necessary to consider not only linear toxicity – descriptors relationships, but also nonlinear models. The linear and nonlinear QSAR models were considered for each from indicated classes of organic chemicals. Analogical models were constructed for whole subset of monofunctional chemicals. The statistical parameters and robustness of nonlinear models are essential better then statistics of linear models. Replacing a lipophilicity descriptor with molecular polarizability and H-bond ability in nonlinear models permits also to improve statistical characteristics. Clearly, if relationships between the intravenous toxicity of compounds bearing only a single functional group and lipophilicity are nonlinear, then similar relationships must be considered with compounds containing more than one functional group. To check up this idea whole set of small clusters containing structure relative compounds with few functional groups was examined from position of linear and nonlinear relationships between toxicity and lipophilicity. It was estimated in most causes advantages of nonlinear models.

Key words: linear and nonlinear QSAR models, HYBOT, acute toxicity to mice.