Cent. Eur. J. Chem. • 9(1) • 2011 • 75-85
DOI: 10.2478/s11532-010-0116-x

VERSITA

Central European Journal of **Chemistry**

QSAR modelling toxicity toward rats of inorganic substances by means of CORAL

Research Article

Alla P. Toropova[1], Andrey A. Toropov[1,*], Emilio Benfenati[1], Giuseppina Gini[2]

[1]*Institute of Pharmacologic Researches by Mario Negri, 20156 Milan, Italy*

[2]*Departmen of Electronics and Information, Polytechnic Institute of Milan, 20133 Milan, Italy*

Received 04 June 2010; Accepted 17 September 2010

Abstract: CORAL ('CORrelation And Logic') is freeware available on the Internet www.insilico.eu/coral The aim of this program is to establish a correlation between an endpoint and descriptors calculated with a simplified molecular input line entry system (SMILES). Three models calculated by CORAL for toxicity towards rat (-pLD50) of inorganic substances (three random splits) have shown that CORAL could be a good tool to model this endpoint. The average statistical characteristics for the CORAL models are the following: n=38, r^2=0.8461, q^2=0.8298, s=0.273, F=198 (subtraining set); n=37, r^2=0.8144, s=0.322, F=154 (calibration set); and n=10, r^2=0.8004, $R_{m}^{2}{}_{(test)}$ =0.7815, s=0.240, F=32 (validation set).

Keywords: *QSAR • SMILES • Inorganic substance • Rat toxicity • Balance of correlation*

1. Introduction

Inorganic compounds are an important class of chemicals, because many of them have vital biochemical features. Typically studies on the quantitative structure–property/activity relationships (QSPR/QSAR) address organic substances [1-11], but only a few articles are related to QSPR/QSAR for inorganic compounds [12-15].

It is well-known that the majority of QSPRs/QSARs are correlations of molecular descriptors calculated with molecular graphs with endpoint values. A major problem with inorganic compounds is that many of the software tools for calculating chemical descriptors are not suitable for inorganic molecules, because molecular graphs for these substances are not defined.

Simplified molecular input line entry system (SMILES) is an alternative molecular graph for the representation of molecular structures [16-19]. Optimal SMILES-based descriptors have been examined as tools to establish QSPR/QSAR for several endpoints and for several classes of chemical substances [20-35].

The typical way to obtain the predictive QSPR/QSAR model involves two steps: 1) the building up of a model (for the training set) which is characterized by an 'ideal' statistical quality; and 2) the estimation of the model for an external validation set. However, this 'idealization' of the model *for the training set* can lead to overtraining. Balance of correlations is an approach that could help avoiding the overtraining [36]. The CORAL freeware can be used to carry out these calculations [37,38] (one can download it from www.insilico.eu/coral).

The aim of the present QSAR analysis is an estimation of the balance of correlations (which is carried out with the CORAL freeware) as a tool for the QSAR modelling of toxicity toward rats of an eclectic collection of inorganic substances.

2. Calculating Details

2.1. Data

The numerical data on toxicity toward rats (oral exposure, LD50, mg kg^{-1}) for 85 inorganic substances were taken from the US National Library of Medicine (at http://toxnet.nlm.nih.gov/). As an endpoint the -pLD50 (*i.e.*, the $\log_{10}$[1/LD50]) has been examined.

* *E-mail: andrey.toropov@marionegri.it*

Table 1. Options used in the CORAL calculations for split A, split B, and split C.

The classic scheme (i.e., training – validation system)

	Split A	Split B	Split C
Descriptor	**α = 1** **β = 1** **γ = 1**	**α = 1** **β = 1** **γ = 0**	**α = 1** **β = 1** **γ = 1**
N_{epoch}	30	20	15
D_{start}	0.5	0.5	0.5
$d_{precession}$	0.01	0.01	0.01
Start threshold	0	0	0
Maximal threshold	5	5	5
Number of probes of the Monte Carlo optimization	3	3	3

The balance of correlations with ideal slopes

	Split A	Split B	Split C
Descriptor	**α = 1** **β = 1** **γ = 1**	**α = 1** **β = 1** **γ = 0**	**α = 1** **β = 1** **γ = 1**
N_{epoch}	30	20	15
dR_{weight}	0.1	0.1	0.1
dC_{weight}	0.01	0.05	0.01
D_{start}	0.5	0.5	0.5
$d_{precession}$	0.01	0.01	0.01
Start threshold	0	0	0
Maximal threshold	5	5	5
Number of probes of the Monte Carlo optimization	3	3	3

Three random splits into the subtraining, calibration, and validation sets were examined.

2.2. Optimal descriptors

Optimal descriptors were calculated with CORAL [37] as follows

$$DCW(Threshold) = \alpha \sum CW(s_k) + \beta \sum CW(ss_k) + \gamma \sum CW(sss_k) \quad (1)$$

where, s_k, ss_k, and sss_k are one-, two-, and three-element SMILES attributes (SA); $CW(s_k)$, $CW(ss_k)$, and $\sum CW(sss_k)$ are the correlation weights. The **α**, **β**, and **γ** are coefficients which can have two values: 0 or 1. The threshold is a parameter for classification of the SA into two categories: rare and not rare. If the threshold is 3, then SA, which take place in less than 3 SMILES of the subtraining set, should be defined as rare. Their correlation weights are fixed equal to zero. Thus, these rare SA are blocked, and they have no influence upon the model.

Each SA is recorded by a string of 12 symbols. The first four symbols are used for recording of one-element SA (*i.e.*, the s_k); positions from 1 to 8 are used to record a two-element SA (*i.e.*, the ss_k), in order to record the three-element SA (*i.e.*, the sss_k) all 12 positions are involved. Vacant positions are indicated by 'x'.

For instance, SMILES = 'C(#N)[Cu]' is represented by lists of SA

s_k: C, (, #, N, (, [, Cu, [

ss_k: C(, (#, N#, N(, [(, [Cu, [Cu

sss_k: C(#, N#(, (N#, [(N, Cu[(, [Cu[

The brackets ')' and ']' are replaced by '(' and '[', because they are indicators of the same phenomena (branching or individual chemical element or functional group). In order to avoid situations where the same molecular attribute is represented by two versions, the ss_k and the sss_k are combined according to their ASCII code (*i.e.*, 'C(' and not '(C', 'N#' and not '#N', *etc.*). Finally for structuring of the calculation, each SA is represented by a string of twelve symbols, such as Cxxxxxxxxxxx (*i.e.*, 'C'); Cxxx(xxxxxxx (*i.e.*, 'C('); and Cxxx(xxx#xxx (*i.e.*, 'C(#').

The Monte Carlo method optimization of the correlation weights has been calculated with the CORAL [37]. Both, the classic scheme and the balance of correlations were used. In the case of the balance of correlation three sets are involved in the modeling process: subtraining, calibration, and validation sets. In the case of the classic scheme, subtraining and calibration sets are united in the common training set. Table 2 contains options which were used for these calculations. An additional parameterization for the balance of correlations, so-called ideal slopes [37], is aimed to improve similarity of the models for the subtraining and calibration sets.

Numerical data for the $CW(s_k)$, $CW(ss_k)$, and $CW(sss_k)$ are calculated by the Monte Carlo method. In the case of the classic 'training-test' system the Monte Carlo approach is aimed to provide the above numerical data which produce the maximum of the correlation

Table 2. Correlation weights of SMILES attributes (SA) which have been obtained for split A using the threshold 2. ID contains two integers: the first is numbering for the given category; the second is the number of SA in the general list of SA for all 85 substances. The NSs, NSc, and NSv are the numbers of SMILES which contain the SA in the subtraining, calibration, and validation sets, respectively. The 'x' denotes a vacant place in the string (twelve symbols) assigned for registration of SMILES attributes

ID	SA* Promoters of LD50 increase	CWs in Probe 1	CWs in Probe 2	CWs in Probe 3	NSs	NSc	NSv
1 104	[xxx-xxxxxxx	0.0576600	0.9450661	0.9805466	26	28	5
2 155	.xxx[xxx-xxx	2.2038857	1.4593387	1.0653456	21	26	5
3 49	=xxx(xxxxxxx	0.8785747	0.4167887	1.9569084	17	23	5
4 89	Oxxx=xxxxxxx	0.5704458	1.1432888	0.4797992	17	23	5
5 161	=xxxOxxx(xxx	1.8312771	2.5764785	0.9802907	16	23	5
6 235	Oxxx(xxx(xxx	2.2892888	4.9882807	3.3545376	16	21	5
7 103	[xxx+xxxxxxx	0.7251218	2.1692506	1.8466593	15	16	2
8 146	-xxx[xxx(xxx	0.1593465	1.3834500	1.4022906	15	18	5
9 6	2xxxxxxxxxx	0.2212588	2.6820747	1.5589226	14	18	6
10 106	[xxx2xxxxxx	2.3273179	1.9409162	3.8107024	14	18	6
11 263	[xxx(xxx[xxx	1.6282372	1.1187021	1.6592683	13	18	5
12 27	Hxxxxxxxxxx	1.6362790	0.1364932	0.3766129	11	6	0
13 269	[xxx(xxxOxxx	2.3628014	4.1277704	4.0550755	8	11	0
14 95	Sxxx(xxxxxx	1.6343216	3.0914364	1.5875177	6	5	2
15 131	[xxxNxxxxxx	0.6501815	0.3422517	1.1503681	6	11	2
16 238	Oxxx.xxxOxxx	1.5412490	2.3695221	0.2346976	6	6	2
17 250	Sxxx(xxx=xxx	3.3260978	4.2938686	4.1748031	6	5	2
18 28	Ixxxxxxxxxx	5.3780692	9.4330953	6.0279275	5	1	0
19 63	Cuxx+xxxxxxx	0.5457627	0.5604974	0.1759676	5	7	2
20 312	[xxxCuxx+xxx	0.7031505	0.5207120	0.9682433	5	7	2
21 23	Cxxxxxxxxxx	2.6677394	6.0098955	4.3791818	4	4	1
22 83	Nxxx+xxxxxx	1.2765710	1.7811092	0.0211702	4	10	2
23 125	[xxxIxxxxxx	1.2516285	1.6676506	1.7775596	4	1	0
24 162	=xxxOxxx.xxx	6.9662245	6.7135914	4.8806802	4	4	1
25 313	[xxxCuxx[xxx	3.3476776	4.4779790	4.8251339	4	0	1
26 329	[xxxNxxx+xxx	1.6964687	1.8026253	1.2459789	4	10	2
27 10	Agxxxxxxxxx	0.6259372	3.6372047	1.2260232	3	2	0
28 33	Nixxxxxxxxx	1.0463609	1.0083731	0.5367132	3	3	1
29 109	[xxxAgxxxxx	4.4563879	4.8684290	4.2778627	3	2	0
30 118	[xxxCoxxxxx	0.4304015	1.5635457	1.9491328	3	6	3
31 130	[xxxNixxxxx	0.2766059	0.3393253	0.4259146	3	3	1
32 225	Naxx[xxx.xxx	1.1593849	2.4382307	1.5887773	3	2	0
33 327	[xxxNixx+xxx	0.9350159	0.4334318	1.0911294	3	1	1
34 1	#xxxxxxxxxx	1.1688509	2.1089683	1.0470773	2	1	0
35 13	Asxxxxxxxxx	3.0703976	2.1357512	0.4031646	2	1	1
36 20	Crxxxxxxxxx	2.5816063	4.1724909	2.6515589	2	1	0
37 21	Csxxxxxxxxx	3.0341249	3.8600413	3.4368597	2	0	0
38 37	Pxxxxxxxxxx	4.4720316	4.7135088	5.4724131	2	2	0
39 48	4xxx+xxxxxx	0.4032275	1.6432239	1.0220531	2	1	0
40 73	HxxxClxxxxx	0.3014049	0.5040726	0.9728434	2	3	0
41 81	Nxxx#xxxxxx	1.3767994	2.1385912	1.6689809	2	1	0
42 112	[xxxAsxxxxx	2.0393594	1.3680096	1.2031513	2	1	1
43 117	[xxxClxxxxx	1.2765075	0.7194067	0.0523193	2	3	1
44 119	[xxxCrxxxxx	3.3997797	2.9798493	2.8453707	2	1	0
45 123	[xxxFexxxxx	2.6523098	3.1928159	2.9419725	2	4	0
46 140	(xxxClxx(xxx	4.7925082	7.6678350	3.6217257	2	2	0
47 159	3xxx[xxx.xxx	0.7187899	1.9203502	2.5581630	2	4	1
48 188	ClxxHxxx-xxx	0.2615864	0.7932372	2.0120621	2	3	0
49 214	Hxxx4xxx+xxx	0.1557204	1.5824918	0.4619864	2	1	0
50 221	Kxxx[xxx.xxx	3.8356035	6.3419081	4.2673328	2	3	0
51 233	Oxxx(xxx=xxx	2.1498905	4.2666163	2.1126859	2	1	0
52 298	[xxxAsxx[xxx	0.5108754	4.1844898	3.1221218	2	1	0
53 304	[xxxBrxxHxxx	1.6350513	0.7549200	0.7938617	2	0	0
54 328	[xxxNxxxHxxx	1.5426633	0.5105789	0.4810251	2	1	0
1 5	.xxxxxxxxxx	-0.3535555	-1.3587912	-0.6760902	28	28	6

Contiuned **Table 2.** Correlation weights of SMILES attributes (SA) which have been obtained for split A using the threshold 2. ID contains two integers: the first is numbering for the given category; the second is the number of SA in the general list of SA for all 85 substances. The NSs, NSc, and NSv are the numbers of SMILES which contain the SA in the subtraining, calibration, and validation sets, respectively. The 'x' denotes a vacant place in the string (twelve symbols) assigned for registration of SMILES attributes

ID	SA* Promoters of LD50 decrease	CWs in Probe 1	CWs in Probe 2	CWs in Probe 3	NSs	NSc	NSv
2 105	[xxx.xxxxxxx	-0.6839461	-0.2141445	-0.5525999	28	28	6
3 244	Oxxx[xxx(xxx	-0.4044131	-1.3559237	-0.0554347	19	24	5
4 9	=xxxxxxxxxx	-2.6686830	-3.1966241	-2.2396317	18	24	5
5 43	(xxx(xxxxxxx	-1.3271455	-2.4838085	-1.6506776	18	24	5
6 86	Oxxx(xxxxxxx	-1.4546362	-1.9268053	-2.2868945	18	26	5
7 267	[xxx(xxx(xxx	-0.3695109	-0.3062094	-1.6214723	14	18	5
8 287	[xxx2xxx+xxx	-1.2254924	-0.3789960	-1.6175192	14	18	6
9 22	Cuxxxxxxxxx	-0.5032979	-1.2040271	-0.6523617	9	7	3
10 157	2xxx[xxx.xxx	-2.5476360	-1.8641632	-2.4247987	9	10	3
11 39	Sxxxxxxxxxx	-6.3234222	-6.1747179	-6.5410924	8	6	2
12 154	.xxx[xxx+xxx	-0.1775732	-1.4562227	-0.7983506	8	3	0
13 258	[xxx(xxx=xxx	-3.3058990	-4.5570755	-2.4409187	7	9	2
14 285	[xxx.xxxOxxx	-1.5021631	-1.6261899	-2.0359670	7	9	3
15 58	Clxx(xxxxxxx	-2.1562036	-3.0918407	-1.4449027	5	4	0
16 64	Cxxx(xxxxxxx	-3.0125926	-3.1282304	-3.9221050	4	3	1
17 200	Cuxx[xxx.xxx	-2.2156539	-1.5935031	-1.4802805	4	7	2
18 201	Cuxx[xxx(xxx	-1.3687126	-3.2664801	-1.9332554	4	0	0
19 277	[xxx+xxxNxxx	-2.1201789	-0.3356208	-0.3933820	4	10	2
20 7	3xxxxxxxxxx	-2.9512852	-0.0426882	-2.0348928	3	4	1
21 19	Coxxxxxxxxx	-0.0752838	-1.5533739	-1.5734505	3	6	3
22 30	Kxxxxxxxxxx	-0.5752487	-1.2640758	-0.4883962	3	3	0
23 75	IxxxHxxxxxx	-0.3292639	-1.8107167	-1.4233345	3	0	0
24 107	[xxx3xxxxxx	-1.6928732	-4.1176172	-1.5219211	3	4	1
25 187	Clxx(xxx=xxx	-3.4733406	-2.1388202	-4.4016123	3	2	0
26 216	IxxxHxxx-xxx	-0.9666807	-1.6052409	-0.2782676	3	0	0
27 218	Ixxx[xxx.xxx	-11.0594602	-12.7841647	-12.7507689	3	0	0
28 232	Nxxx[xxx.xxx	-1.2050965	-2.3928255	-0.8795547	3	6	2
29 276	[xxx+xxxKxxx	-2.5189914	-0.9624966	-0.7764334	3	3	0
30 319	[xxxIxxxHxxx	-2.3263736	-1.2460704	-1.1759973	3	0	0
31 322	[xxxKxxx+xxx	-0.9905894	-1.5673520	-2.3822990	3	3	0
32 24	Fexxxxxxxxx	-3.0204790	-3.3142088	-3.7803395	2	4	0
33 42	(xxx#xxxxxx	-1.9295562	-0.6882164	-2.7079777	2	0	0
34 46	3xxx+xxxxxx	-1.0251195	-3.2775900	-2.7317117	2	4	0
35 50	Agxx+xxxxxx	-2.3464099	-2.9035978	-2.2291225	2	2	0
36 116	[xxxBrxxxxx	-3.4040792	-1.5758190	-1.1030577	2	0	1
37 120	[xxxCsxxxxx	-8.6513903	-11.1181316	-9.0752663	2	0	0
38 144	(xxxOxxx(xxx	-3.5068409	-4.5554151	-2.9473162	2	4	0
39 164	Agxx[xxx.xxx	-2.0971701	-3.3142411	-2.2106100	2	2	0
40 184	Brxx[xxx.xxx	-0.3686379	-1.7866277	-3.1877083	2	0	0
41 186	Clxx(xxxClxx	-1.6681200	-0.3010030	-1.8067953	2	2	0
42 189	Clxx[xxx.xxx	-3.0920568	-3.7173343	-2.4599040	2	3	0
43 202	Cxxx(xxx#xxx	-2.8335161	-1.7235983	-0.9424401	2	0	0
44 229	Nxxx#xxx(xxx	-0.5453921	-1.3583182	-2.5121414	2	0	0
45 261	[xxx(xxxClxx	-0.4054346	-0.4258772	-0.4853416	2	2	0
46 262	[xxx(xxxNxxx	-3.2418617	-1.0825639	-1.2393210	2	0	0
47 273	[xxx+xxxAgxx	-2.1388921	-3.1519504	-1.2937646	2	2	0
48 283	[xxx.xxxClxx	-0.8780951	-1.4658470	-1.1918423	2	0	0
49 290	[xxx3xxx+xxx	-0.3766330	-1.1289534	-0.0995489	2	4	0
50 292	[xxxAgxx+xxx	-1.3084432	-2.6984509	-2.5808398	2	2	0
	Undefined						
1 41	[xxxxxxxxxx	-0.3962186	0.2094313	-0.6616779	38	37	10
2 3	+xxxxxxxxxx	0.6039544	-0.9842311	-0.8816147	28	28	6
3 4	-xxxxxxxxxx	1.1201566	0.3938436	-0.2923927	28	28	6
4 2	(xxxxxxxxxx	0.2109716	-0.6453440	0.1836858	27	31	7
5 102	[xxx(xxxxxxx	-0.3786440	0.7499205	0.4261079	27	31	7

Continued **Table 2.** Correlation weights of SMILES attributes (SA) which have been obtained for split A using the threshold 2. ID contains two integers: the first is numbering for the given category; the second is the number of SA in the general list of SA for all 85 substances. The NSs, NSc, and NSv are the numbers of SMILES which contain the SA in the subtraining, calibration, and validation sets, respectively. The 'x' denotes a vacant place in the string (twelve symbols) assigned for registration of SMILES attributes

ID	SA* Promoters of LD50 increase	CWs in Probe 1	CWs in Probe 2	CWs in Probe 3	NSs	NSc	NSv
6 284	[xxx.xxx[xxx	-0.5431463	-2.4469889	0.2294511	25	28	5
7 35	Oxxxxxxxxxx	-0.6299027	0.1758056	-0.5951530	23	28	5
8 132	[xxxOxxxxxx	0.0571956	-0.0093109	-0.0360964	20	24	5
9 87	Oxxx-xxxxxx	-0.7915673	0.0067693	-0.5830652	19	24	5
10 279	[xxx-xxxOxxx	0.9212547	-0.6703821	0.0131939	19	24	5
11 333	[xxxOxxx-xxx	-0.6840899	-0.5220567	0.0550842	19	24	5
12 239	Oxxx=xxx(xxx	0.1681475	-0.5290606	0.3026886	17	23	5
13 44	2xxx+xxxxxxx	1.5033106	-0.7186941	0.1442114	14	18	6
14 160	=xxx(xxx(xxx	-0.3029422	0.2055669	-1.1206689	11	11	2
15 34	Nxxxxxxxxxx	0.5230718	-2.2226487	0.9580387	9	12	2
16 121	[xxxCuxxxxx	-0.5546086	0.2770756	-1.1822262	9	7	3
17 88	Oxxx.xxxxxx	0.3394429	-0.0383585	0.6537916	8	10	3
18 18	Clxxxxxxxxx	0.1421387	0.7536233	-0.0553613	7	7	1
19 68	Hxxx-xxxxxx	-0.2985598	1.3858692	0.9736300	7	5	0
20 278	[xxx-xxxHxxx	-0.5263933	0.0527898	0.9534665	7	5	0
21 151	.xxxOxxx.xxx	-0.4890381	-0.4959921	0.6174209	6	6	2
22 32	Naxxxxxxxxx	-0.6050326	0.4184994	0.4307683	5	2	0
23 79	Naxx+xxxxxx	-1.7734936	0.3547931	-0.5814195	5	2	0
24 129	[xxxNaxxxxx	0.7316306	-2.3661233	-2.6744980	5	2	0
25 198	Cuxx+xxx2xxx	0.3539778	-0.7385148	0.0048348	5	7	2
26 272	[xxx+xxxNaxx	-0.7193167	-3.2654965	0.2561318	5	2	0
27 325	[xxxNaxx+xxx	-0.6420081	0.4640281	-0.6530044	5	2	0
28 145	+xxx[xxx(xxx	0.1634558	0.9774663	-1.9463970	4	10	2
29 77	Kxxx+xxxxxx	1.1837322	-1.1284629	-1.2142130	3	3	0
30 80	Nixx+xxxxxx	-0.3298605	0.8139675	-0.2377991	3	1	1
31 127	[xxxKxxxxxx	-0.9667531	-0.0906606	0.7555277	3	3	0
32 226	Nixx+xxx2xxx	0.5006789	-0.0510994	-0.6290999	3	1	1
33 228	Nixx[xxx.xxx	-0.8205333	0.3910435	0.0922058	3	0	1
34 8	4xxxxxxxxxx	0.3920360	0.6440884	-1.4630758	2	1	0
35 17	Brxxxxxxxxx	-0.8868612	0.6447100	-0.8258790	2	4	3
36 59	Clxx.xxxxxx	0.0462615	-0.6234592	-0.3970997	2	0	0
37 71	Hxxx4xxxxxx	0.5281011	1.4342059	-0.5568255	2	1	0
38 72	HxxxBrxxxxx	-0.7305051	1.1366902	-0.9261132	2	0	0
39 82	Nxxx(xxxxxx	0.1993662	-1.8315847	-2.3695151	2	0	0
40 85	NxxxHxxxxxx	1.0974556	-0.3459387	-1.3554573	2	1	0
41 96	Sxxx.xxxxxx	1.8002577	-0.3601657	0.7063291	2	2	0
42 143	(xxxNxxx#xxx	-0.1372832	-2.7136462	0.2748450	2	0	0
43 148	.xxxClxx(xxx	0.4023198	-0.0286432	1.8112045	2	0	0
44 153	.xxxSxxx(xxx	0.5342443	-0.4274236	2.0303805	2	2	0
45 173	Asxx[xxx(xxx	1.9596292	-0.0911699	3.6284417	2	1	0
46 183	BrxxHxxx-xxx	-1.2778241	-1.8622282	0.6508208	2	0	0
47 231	NxxxHxxx4xxx	-1.9796093	-0.7927365	2.3454573	2	1	0
48 268	[xxx(xxxSxxx	-0.1163417	2.2845051	2.1262967	2	1	0
49 270	[xxx+xxx4xxx	-0.8635547	2.0217190	0.2334416	2	1	0
50 305	[xxxClxxHxxx	0.3862709	2.0820154	-0.7775611	2	3	0
51 306	[xxxCoxx[xxx	0.2277696	1.4450168	-1.0777830	2	0	1

*) *The list of blocked SA is omitted*

coefficient between pLD50(expr) and DCW(Threshold) for the training set.

In the case of the balance of correlations, the target function of the Monte Carlo method optimization is the following:

$$BC=R+R' - abs(R-R')*dR_{weight} \quad (2)$$

balance of correlations: **R** and **R'** are correlation coefficient between endpoint and optimal descriptors for the subtraining set and calibration set. The role of the calibration set is a preliminary validation of the model. This approach is an attempt to avoid overtraining. In other words, in the case of balance of correlations, the training set is split into two sets: subtraining and

Table 3. Example of DCW(2) calculation. SMILES: C(#N)[Cu]. CAS 544-92-3. Number of Monte Carlo optimization probe 1. Threshold 2. DCW (2) = -6.1470623

SA	CW(SA)
Cxxxxxxxxxx	2.6677394
(xxxxxxxxxx	0.2109716
#xxxxxxxxxx	1.1688509
Nxxxxxxxxxx	0.5230718
(xxxxxxxxxx	0.2109716
[xxxxxxxxxx	-0.3962186
Cuxxxxxxxxx	-0.5032979
[xxxxxxxxxx	-0.3962186
Cxxx(xxxxxxx	-3.0125926
(xxx#xxxxxxx	-1.9295562
Nxxx#xxxxxxx	1.3767994
Nxxx(xxxxxxx	0.1993662
[xxx(xxxxxxx	-0.3786440
[xxxCuxxxxxx	-0.5546086
[xxxCuxxxxxx	-0.5546086
Cxxx(xxx#xxx	-2.8335161
Nxxx#xxx(xxx	-0.5453921
(xxxNxxx#xxx	-0.1372832
[xxx(xxxNxxx	-3.2418617
Cuxx[xxx(xxx	-1.3687126
[xxxCuxx[xxx	3.3476776

calibration. The $\mathbf{dR_{weight}}$ is an empirical parameter. This optimization is the following: **BC → maxBC**

Table 1 contains the options for CORAL software, which were used for the modeling process.

3. Results and Discussion

Fig. 1 shows the influence of the threshold value upon predictability (determination coefficient between experimental and calculated –pLD50 values). One can see that for all splits, balance of correlations gave better prediction than the classic scheme. Each split has a particular threshold value that gives the most reliable prediction (for a given split). These values are the following: for split A Threshold =2, for split B threshold =0, and for split C threshold =1. It is to be noted that the classic scheme also has preferable thresholds (these are 4, 2, and 1 for splits A, B, and C, respectively). However there is no correlation of these thresholds for the classic scheme and the balance of correlations.

The preferable –pLD50 model for split A (threshold=2, probe 1) is the following

$$\mathbf{-pLD50 = -2.4234(\pm 0.01139) + 0.1190(\pm 0.00133) * DCW(2)} \quad (3)$$

n=38, r^2=0.8461, q^2=0.8298, s=0.273, F=198 (subtraining set)
n=37, r^2=0.8144, s=0.322, F=154 (calibration set)
n=10, r^2=0.8004, s=0.240, F=32 (validation set)

An additional check of predictability of Eq. 3 according to [39] and [40] gives

$(r^2-r_0^2)/r^2$ = 0.0007 (according to [39] should be < 0.1)

$(r^2-r'^2_0)/r^2$ = 0.0781 (according to [39] should be < 0.1)

k = 1.0224 (according to [39] should be 0.85 < k <1.15)

k' = 0.9726 (according to [39] should be 0.85 < k' < 1.15)

$R^2_{m\ (test)}$ = 0.7815 (according to [40] should be > 0.5)

Thus, the –pLD50 model calculated with Eq. 3 has satisfactory predictability [39,40]. Table 2 contains correlation weights for the calculation of DCW(2). In particular, the correlation weights of SA which have been obtained in the probe 1 are used for Eq. 3. Table 3 shows an example of the DCW(2) calculations. Table 4 contains the numerical data on the value of -pLD50 experimental and calculated with Eq. 3.

Fig. 2 shows the model for split A (calculated with Eq. 3) graphically as well as analogical models for the split B (threshold =0) and split C (threshold =1). One can see from Fig. 2, that correlations between pLD50(expr) and pLD50(calc) for the subtraining set and calibration set are poorer than this correlation for the test set. However, probably, this conflict between the subtraining set and calibration set avoids the overtraining, whereas for the classic scheme these models are better for the general list of substances (*i.e.*, list of subtraining together with calibration sets) but poorer for the external test set.

One can see from the Table 2, that there are SA which are stable promoters of –pLD50 increase (*i.e.*, 54 SA with positive correlation weights for all three probes of the Monte Carlo optimization), there are SA which are stable promoters of –pLD50 decrease (*i.e.*, 50 SA with negative correlation weights for all three probes of the Monte Carlo optimization), there are 51 SA with an undefined role (both, positive and negative correlation weights were obtained in the probes of the Monte Carlo optimization). Finally there are 185 blocked SA (this list is omitted in Table 2). Analysis of these data can give hints for mechanistic interpretations of the model, which can be used for synthesis or definition of preferable substances. In order to obtain a substance with a high toxicity -pLD50 value, it is preferable that the the molecular structure fragments which are encoded by the following SMILES attributes (which are promoters of –pLD50) increase: '.xxx[xxx-xxx', '=xxxOxxx(xxx', and

Table 4. The -pLD50 experimental values calculated with Eq. 3

CAS	SMILES	DCW(2)	LD50(expr)	LD50(calc)
	Subtraining set			
544-92-3	C(#N)[Cu]	-6.1470623	-3.102	-3.155
1317-39-1	O([Cu])[Cu]	-1.0281763	-2.672	-2.546
1332-40-7	[OH-].[Cu+2].[ClH-].[OH-].[OH-].[Cu+2].[ClH-]	-5.1156994	-2.845	-3.032
7447-39-4	[Cu](Cl)Cl	-3.8480679	-2.766	-2.881
7733-02-0	S(=O)(=O)([O-])[O-].[Zn+2]	-0.5183244	-3.233	-2.485
7758-99-8	S(=O)(=O)([O-])[O-].O.O.O.O.O.[Cu+2]	-2.4532302	-2.477	-2.715
12643-19-5	[Cu+2].[Cu+2].[Cu+2].[PH3-3].[PH3-3]	-13.1839393	-4.000	-3.992
13933-17-0	[Cu+2].[ClH-].O.O.[ClH-]	-3.2520798	-2.462	-2.810
20427-59-2	[Cu](O)O	-7.1504416	-3.000	-3.274
73156-86-2	[Cu+2].[As]([O-])([O-])[O-].[Cu+2].[Cu+2].[As]([O-])([O-])[O-]	1.5595365	-2.452	-2.238
7758-94-3	[Fe](Cl)Cl	-1.9303772	-2.653	-2.653
7783-85-9	S(=O)(=O)([O-])[O-].O.[Fe+3].[NH4+].S(=O)([O-])([O-])=O.O.O.O.O.O.[NH4+]	-9.4525133	-3.512	-3.548
7785-20-8	OS(=O)(=O)[O-].OS([O-])(=O)=O.N.N.[Ni+2].O.O.O.O.O.O	-2.2153293	-2.601	-2.687
10101-97-0	S(=O)(=O)([O-])[O-].[Ni+2]	1.0899434	-2.422	-2.294
15699-18-0	S([O-])([O-])(=O)=O.S(=O)(=O)([O-])[O-].[Ni+2].[NH4+].[NH4+]	-1.9734627	-2.602	-2.658
1317-42-6	S=[Co]	-8.7712536	-3.699	-3.467
10026-17-2	[Co](F)F	0.2641508	-2.176	-2.392
10294-50-5	[Co+2].P(=O)([O-])([O-])[O-].O.[Co+2].[Co+2].P(=O)([O-])([O-])[O-].O.O.O.O.O.O.O	-3.3096627	-2.732	-2.817
506-64-9	C(#N)[Ag]	3.0252008	-2.090	-2.063
534-16-7	[O-]C([O-])=O.[Ag+].[Ag+]	-9.6161086	-3.572	-3.568
10124-50-2	[As]([O-])([O-])[O-].[K+]	5.9134602	-1.146	-1.720
10588-01-9	[Cr](O[Cr](=O)(=O)[O-])(=O)(=O)[O-].[Na+].[Na+]	5.9886027	-1.699	-1.711
50864-67-0	[Ba+2].[SH2-2]	-3.7639062	-2.574	-2.871
144-55-8	C(O)(=O)[O-].[Na+]	-7.7274492	-3.625	-3.343
7647-15-6	[Na+].[BrH-]	-8.7498995	-3.544	-3.465
7758-02-3	[K+].[BrH-]	-9.6095295	-3.487	-3.567
7775-09-9	Cl(=O)(=O)[O-].[Na+]	-7.1287623	-3.079	-3.272
10035-05-9	[Ca+2].Cl(=O)(=O)[O-].Cl(=O)(=O)[O-].O.O	-12.0301843	-3.653	-3.855
10326-21-3	Cl(=O)(=O)[O-].Cl(=O)(=O)[O-].[Mg+2]	-8.1976409	-3.803	-3.399
7757-79-1	[N+](=O)([O-])[O-].[K+]	-2.0235432	-3.574	-2.664
7761-88-8	[N+](=O)([O-])[O-].[Ag+]	-4.7998865	-3.069	-2.995
7789-02-8	[N+](=O)([O-])[O-].[N+](=O)([O-])[O-].[N+](=O)([O-])[O-].[Cr+3]	-10.2470066	-3.512	-3.643
7789-18-6	[N+](=O)([O-])[O-].[Cs+]	-7.6085617	-3.378	-3.329
7681-82-5	[Na+].[IH-]	-11.7691240	-3.637	-3.824
7774-29-0	[Hg](I)I	10.0070004	-1.255	-1.233
7789-17-5	I[Cs]	-8.4313952	-3.378	-3.427
7790-29-6	[Rb+].[IH-]	-8.7609036	-3.673	-3.466
7790-80-9	[IH-].[IH-].[Cd+2]	-0.1347951	-2.346	-2.439
	Calibration set			
1333-22-8	[Cu+2].[Cu+2].[Cu+2].[OH-].S(=O)(=O)([O-])[O-].[OH-].S(=O)(=O)([O-])[O-]	-7.3213041	-3.398	-3.295
7720-78-7	S(=O)(=O)([O-])[O-].[Fe+2]	-0.8864936	-2.504	-2.529
7758-98-7	S(=O)(=O)([O-])[O-].[Cu+2]	-2.1889938	-2.477	-2.684
10031-43-3	[N+](=O)([O-])[O-].[N+](=O)([O-])[O-].[Cu+2]	-5.2937244	-2.973	-3.053
12069-69-1	C(=O)([O-])[O-].[OH-].[OH-].[Cu+2].[Cu+2]	-3.6656192	-3.130	-2.860
13478-34-7	[As]([O-])([O-])([O-])=O.[As](=O)([O-])([O-])[O-].[Cu+2].[Cu+2].[Cu+2]	-4.7812976	-3.176	-2.992
17836-27-0	[Cr+3].[Cu+2].P(=O)([O-])([O-])[O-]	-4.8418769	-3.137	-3.000
61482-17-5	[NH4+].[Cu+2].[NH4+].[Cu+2].S([O-])([O-])(=O)=O.S(=O)(=O)([O-])[O-].S(=O)(=O)([O-])[O-]~	-6.3067973	-3.403	-3.174
7705-08-0	[Fe](Cl)(Cl)Cl	-2.2133398	-2.653	-2.687
7720-78-7	S(=O)(=O)([O-])[O-].[Fe+2]	-0.8864936	-2.504	-2.529
7782-61-8	O[N+](=O)[O-].O[N+](=O)[O-].O[N+](=O)[O-].[Fe+3].O.O.O.O.O.O.O.O.O	-7.5137521	-3.512	-3.318
7718-54-9	[Ni](Cl)Cl	-2.6149451	-2.833	-2.735

Continued **Table 4.** The -pLD50 experimental values calculated with Eq. 3

CAS	SMILES Subtraining set	DCW(2)	LD50(expr)	LD50(calc)
7791-20-0	[Ni+2].[ClH-].O.O.O.O.O.O.[ClH-]	1.6222908	-2.021	-2.230
12054-48-7	[Ni](O)O	-5.9173188	-3.180	-3.128
513-79-1	C(O)([O-])[O-].[Co+2]	-4.6626638	-2.806	-2.978
7646-79-9	[ClH-].[Co+2].[ClH-]	2.1572621	-1.903	-2.167
7791-13-1	[Co+2].[ClH-].O.[ClH-].O.O.O.O.O	-1.8252699	-2.884	-2.641
10026-22-9	[N+](=O)([O-])[O-].[O-][N+](=O)[O-]. [Co+2].O.O.O.O.O.O	-2.9800656	-2.839	-2.778
10141-05-6	[N+]([O-])([O-])=O.[N+](=O)([O-])[O-].[Co+2]	-1.6959644	-2.637	-2.625
13455-36-2	P([O-])([O-])(=O)[O-].P(=O)([O-])([O-])[O-].[Co+2]. [Co+2].[Co+2]	-2.2477020	-2.588	-2.691
506-61-6	[C-]#N.[C-]#N.[Ag+].[K+]	5.9589523	-1.320	-1.714
7761-88-8	[N+](=O)([O-])[O-].[Ag+]	-4.7998865	-3.069	-2.995
1306-23-6	S=[Cd]	-9.7845424	-3.850	-3.588
7758-01-2	Br(=O)(=O)[O-].[K+]	-0.7116294	-2.196	-2.508
1309-42-8	[Mg](O)O	-7.5168915	-3.929	-3.318
10022-31-8	[N+]([O-])([O-])=O.[N+](=O)([O-])[O-].[Ba+2]	-2.0510821	-2.550	-2.667
554-13-2	C(=O)([O-])[O-].[Li+].[Li+]	-3.2833436	-2.720	-2.814
7727-15-3	[Al](Br)(Br)Br	-4.3149239	-3.204	-2.937
7789-41-5	[Ca](Br)Br	-2.5228604	-3.613	-2.724
7789-47-1	[Hg](Br)Br	-2.5228604	-1.602	-2.724
3811-04-9	Cl(=O)(=O)[O-].[K+]	-5.3121737	-3.272	-3.056
7601-89-0	Cl(=O)(=O)(=O)[O-].[Na+]	-7.3523948	-3.322	-3.298
7631-99-4	[N+](=O)([O-])[O-].[Na+]	-3.8401318	-3.103	-2.880
7784-27-2	[O-][N+](=O)[O-].[O-][N+](=O)[O-].[O-][N+](=O) [O-].[Al+3].O.O.O.O.O.O.O.O.O	-7.7813966	-3.565	-3.349
10022-68-1	[N+](O)(=O)[O-].[N+](O)(=O)[O-].[Cd+2]	-1.2872508	-2.477	-2.577
10035-06-0	[Bi+3].[N+]([O-])(=O)[O-].[N+]([O-])(=O)[O-].[N+] ([O-])(=O)[O-].O.O.O.O.O	-7.3248615	-3.607	-3.295
7790-30-9	[Tl]I	5.8372605	-1.382	-1.729
	Validation set			
3251-23-8	[N+]([O-])([O-])=O.[N+](=O)([O-])[O-].[Cu+2]	-3.7217515	-2.900	-2.866
7758-89-6	Cl[Cu]	2.3613715	-2.146	-2.142
12006-17-6	[Cu+2].[As-3].[Cu+2].[Cu+2].[As-3]	-9.2784761	-4.000	-3.528
13478-00-7	[N+](=O)([O-])[O-].[N+](=O)([O-]) [O-].O.[Ni+2].O.O.O.O.O	-7.6567841	-3.210	-3.335
7789-43-7	[Co](Br)Br	-1.5095716	-2.609	-2.603
10026-24-1	S(=O)(=O)([O-])[O-].O.[Co+2].O.O.O.O.O.O	-5.0575651	-2.765	-3.025
10124-43-3	S(=O)(=O)([O-])[O-].[Co+2]	-0.1632067	-2.627	-2.443
513-77-9	C(=O)([O-])[O-].[Ba+2]	0.7234577	-2.621	-2.337
7550-35-8	[Li]Br	-5.0833776	-3.255	-3.028
7789-42-6	[Cd](Br)Br	-2.5228604	-2.508	-2.724

'[xxx(xxx[xxx'. *Vice versa*, in order to obtain a substance with minimal toxicity, it is preferable that the presence in molecular structure fragmets which are encoded by attributes such as '.xxxxxxxxxxx', '=xxxxxxxxxxx', and 'Oxxx(xxxxxxx' increase. The criterions for the above-mentioned classifications of the SA can be defined as the following: first, the number of SMILES containing the SA (in the subtraining set) should be as large as possible, and second, the average absolute value of the correlation weights for the SA in three probes of the optimization also should be as large as possible.

It is to be noted that the analysis of the numbers of the promoters of endpoint values increasing/decreasing together with the list of undefined SMILES attributes may be used for constructive comparisons of different splits. For instance, split A has the numbers of promoters of –pLD50 increase, promoters of –pLD50 decrease, and undefined, which are equal to 54, 50, and 51, respectively. These numbers for split C are equal to 103, 72, and 85, respectively. Thus, the number of undefined SA for split A are considerably lower. This can be estimated as an advantage of the split A.

The comparison of split B with split A and/or split C cannot be done by the scheme above because the total number of SA for split B (138) is considerably less than the number for the split A and the split C (340). However, one can agree that the smaller total number of parameters of the Monte Carlo optimization most probably is an advantage for split B.

The probabilistic analysis of the CORAL data (the numbers of SMILES attributes which are promoters of –pLD50 increase/decrease together with the number

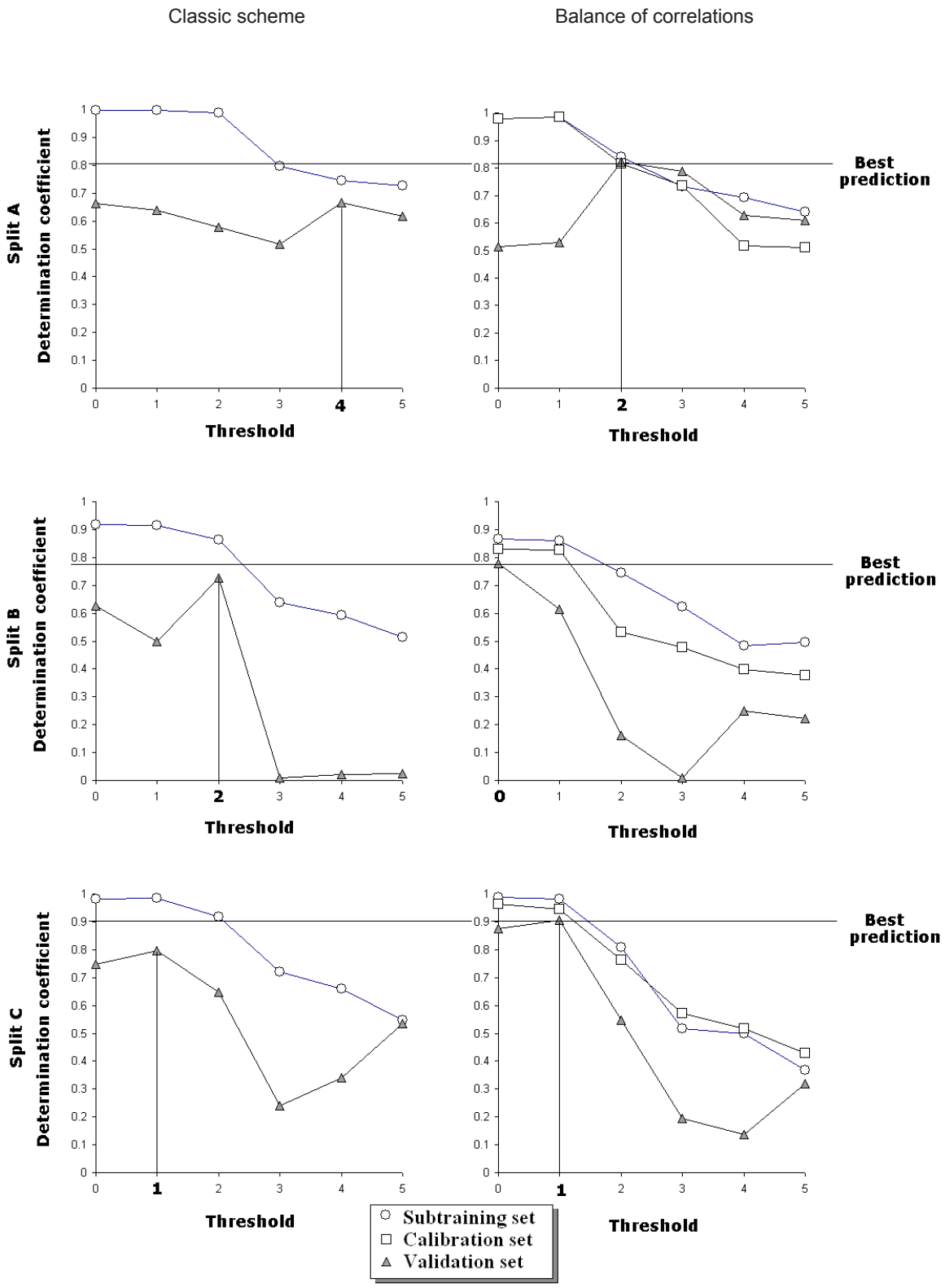

Figure 1. Determination of coefficient values which were obtained with different thresholds (0-5) for the split A, split B, and split C.

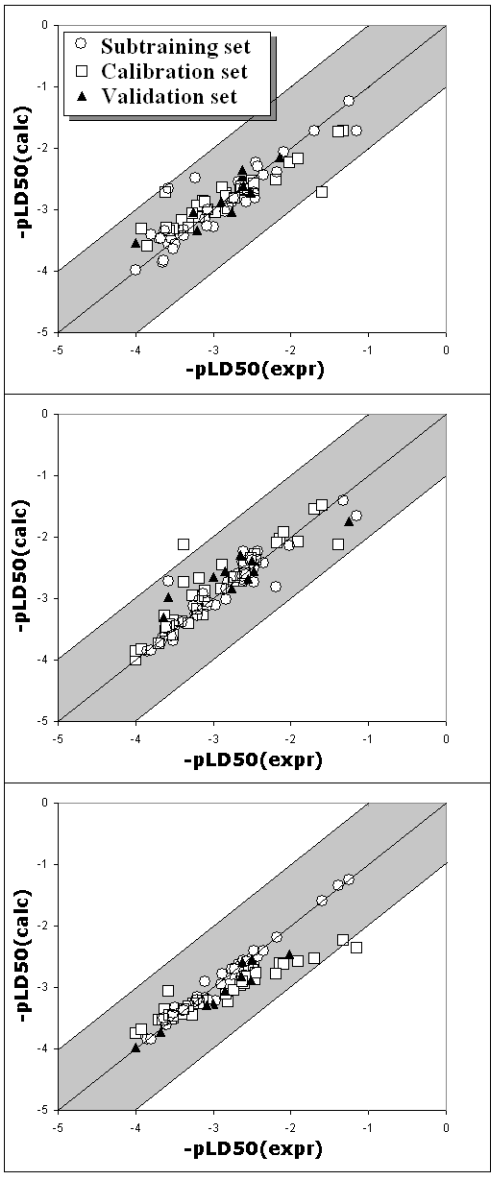

Figure 2. Graphical representation of –pLD50 models for split A, split B, and split C

of undefined SMILES attributes) can be used for the definition of robust hypotheses related to the influence of molecular fragments upon the endpoint –pLD50 value.

The applicable domain for a CORAL model of an endpoint can also be formulated using the above-mentioned data. In a first approximation it can be the following: *structures with SMILES which do not contain rare attributes.*

Thus, the numerical data on the correlation weights are defined by the given endpoint and by a list of training and calibration sets. The modification of this list can lead to considerable changes in the correlation weights and moreover it can lead to considerable changes of statistical characteristics of the models. It is clear that subtraining and calibration sets should be larger than the test set.

Since we have not found in the literature reports dedicated to QSAR modelling of rat toxicity for inorganic substances, we have compared the model calculated with Eq. 3 with analogous models for organic compounds.

Statistical characteristics of a model for rat toxicity (567 compounds) based on two-dimensional chemical descriptors reported in [41] are s=0.73 (in logarithmic scale) and q^2=0.64. According to [42], statistical characteristics of the partial least squares QSAR models for rat toxicity (28 benzene derivatives) are R^2 = 0.81-0.92; q^2 = 0.64-0.83; R^2(test) =0.84 -0.87. Statistical characteristics of the logLD50 model described in [43] are the following: n=49, r^2=0.854, s=0.477.

Thus, the comparison of statistical characteristics of the model that is calculated with Eq. 3 with the above-mentioned models [41-43] indicates that Eq. 3 is an acceptable predictor for the logLD50.

The molecular structure for 85 inorganic compounds and the CORAL models for split B and split C are given in the *Supplementary materials*.

4. Conclusions

CORAL can be used as a tool for the satisfactory prediction of the –pLD50 for 85 examined inorganic substances; the balance of correlations gave preferable models for the –pLD50 in comparison with the classic scheme.

Acknowledgements

The authors thank Federchimica - AISPEC for financial support.

References

[1] G. Melagraki, A. Afantitis, H. Sarimveis, P.A. Koutentis, G. Kollias, O. Igglessi-Markopoulou, Mol. Divers 13, 301 (2009)

[2] G. Melagraki, A. Afantitis, O. Igglessi-Markopoulou, A. Detsi, M. Koufaki, C. Kontogiorgis, D.J. Hadjipavlou-Litina, Eur. J. Med. Chem. 44, 3020 (2009)

[3] A. Afantitis, G. Melagraki, H. Sarimveis, P.A. Koutentis, J. Markopoulos, O. Igglessi-Markopoulou, QSAR and Comb. Sci. 27, 432 (2008)

[4] E. Vicente, P.R. Duchowicz, E.A. Castro, A. Monge, J. Mol. Graphics Model. 28, 28 (2009)

[5] P.R. Duchowicz, A. Talevi, L.E. Bruno-Blanch, E.A. Castro, Bioorg. Med. Chem. 16, 7944 (2008)

[6] P.R. Duchowicz, E.A. Castro, Int. J. Mol. Sci. 10, 2558 (2009)

[7] T. Puzyn, A. Mostrag, J. Falandysz, Y. Kholod, J. Leszczynski, J. Hazard. Mater. 170, 1014 (2009)

[8] F. Torrens, G. Castellano, Stud. Comput. Intell. 205, 243 (2007)

[9] F. Torrens, G. Castellano, SAR QSAR Environ. Res. 19, 643 (2008)

[10] G. Fath-Tabar, B. Furtula, I. Gutman, J. Math. Chem. 47, 477 (2009)

[11] I. Gutman, B. Furtula, M. Petrović, J. Math. Chem. 46, 522 (2009)

[12] S.P. Niculescu, M.A. Lewis, J. Tigner, SAR QSAR Environ. Res. 19, 735 (2008)

[13] L. Mu, C. Feng, H. He, MATCH Commun. Math. Comput. Chem. 58, 591 (2007)

[14] A.A.Toropov, A.P. Toropova, E. Benfenati, Cent. Eur. J. Chem. 7, 846 (2009)

[15] A.A.Toropov, A.P. Toropova, E. Benfenati, J. Math. Chem. 46, 1060 (2009)

[16] D. Weininger, J. Chem. Inf. Comput. Sci. 28, 31 (1988)

[17] D. Weininger, A. Weininger, J.L. Weininger, J. Chem. Inf. Comput. Sci. 29, 97 (1989)

[18] D. Weininger, J. Chem. Inf. Comput. Sci. 30, 237 (1990)

[19] ACD/ChemSketch Freeware, version 11.00 (Advanced Chemistry Development, Inc., Toronto, ON, Canada, 2007) www.acdlabs.com

[20] A.A. Toropov, A.P. Toropova, D.V. Mukhamedzhanova, I. Gutman, Indian J. Chem. 44A, 1545 (2005)

[21] A. Toropov, K. Nesmerak, I. Raska Jr, K. Waisser, K. Palat, Comput. Biol. Chem. 30, 434 (2006)

[22] A.A.Toropov, E. Benfenati, Comput. Biol. Chem. 31, 57 (2007)

[23] K. Roy, A. Toropov, I. Raska Jr. QSAR Comb. Sci. 26, 460 (2007)

[24] A.A. Toropov, E. Benfenati, Curr. Drug. Dis. Technol. 4, 77 (2007)

[25] A.A. Toropov, A.P. Toropova, I. Raska Jr., Eur. J. Med. Chem. 43, 714 (2008)

[26] A.A. Toropov, B.F. Rasulev, D. Leszczynska, J. Leszczynski, Chem. Phys. Lett. 457, 332 (2008)

[27] A.A. Toropov, A.P. Toropova, E. Benfenati, A. Manganaro, Mol. Divers 13, 367 (2009)

[28] A.A. Toropov, A.P. Toropova, E. Benfenati, Eur. J. Med. Chem. 44, 2544 (2009)

[29] A.A. Toropov, A.P. Toropova, E. Benfenati, Int. J. Mol. Sci. 10, 3106 (2009)

[30] A.A. Toropov, A.P. Toropova, E. Benfenati, Chem. Biol. Drug Des. 73, 515 (2009)

[31] A.A. Toropov, A.P. Toropova, E. Benfenati, Chem. Biol. Drug Des. 73, 482 (2009)

[32] A.A. Toropov, A.P. Toropova, E. Benfenati, Chem. Biol. Drug Des. 73, 301 (2009)

[33] A.A. Toropov, A.P. Toropova, E. Benfenati, D. Leszczynska, J. Leszczynski, J. Comput. Chem. 31, 381 (2010)

[34] A.A. Toropov, A.P. Toropova, E. Benfenati, D. Leszczynska, J. Leszczynski, J. Math. Chem. 47, 647 (2010)

[35] A.A. Toropov, A.P. Toropova, E. Benfenati, D. Leszczynska, J. Leszczynski, J. Math. Chem. 47, 355 (2010)

[36] A.A. Toropov, B.F. Rasulev, J. Leszczynski, Bioorg. Med. Chem. 16, 5999 (2008)

[37] A.P. Toropova, A.A. Toropov, E. Benfenati, D. Leszczynska, J. Leszczynski, J. Math. Chem. 2010 (In press)

[38] A.P. Toropova, A.A. Toropov, E. Benfenati, G. Gini, D. Leszczynska, J. Leszczynski, Mol. Divers. 2010 (In press)

[39] A. Golbraikh, A. Tropsha, J. Mol. Graphics Model. 20, 269 (2002)

[40] P.P. Roy, K. Roy, QSAR Comb. Sci. 27, 302 (2008)

[41] P. Mazzatorta, M.D. Estevez, M. Coulet, B. Schilter, J. Chem. Inform. Model. 48, 1949 (2008)

[42] V.E. Kuz'min, E.N. Muratov, A.G. Artemenko, L. Gorb, M. Qasim, J. Leszczynski, Chemosphere 72, 1373 (2008)

[43] A.P. Freidig, S. Dekkers, M. Verwei, E. Zvinavashe, J.G.M. Bessems, J.J.M. van de Sandt, Toxicol. Lett. 170, 214 (2007)