Chemometrics and Intelligent Laboratory Systems 105 (2011) 215–219

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Short Communication

Co-evolutions of correlations for QSAR of toxicity of organometallic and inorganic substances: An unexpected good prediction based on a model that seems untrustworthy

A.P. Toropova [a], A.A. Toropov [a,*], E. Benfenati [a], G. Gini [b]

[a] Istituto di Ricerche Farmacologiche Mario Negri, 20156, Via La Masa 19, Milano, Italy
[b] Department of Electronics and Information, Politecnico di Milano, piazza Leonardo da Vinci 32, 20133 Milano, Italy

ARTICLE INFO

Article history:
Received 22 November 2010
Received in revised form 21 December 2010
Accepted 22 December 2010
Available online 6 January 2011

Keywords:
QSAR
SMILES
Toxicity towards rat
Optimal descriptor
Balance of correlation
Co-evolution of correlation

ABSTRACT

The simplified molecular input line entry system (SMILES) gives a representation of the molecular structure by a sequence of special characters indicating different chemical elements, double/triple covalent bonds, and other features. We used this representation to establish quantitative structure–activity relationships (QSAR) for toxicity (pLD50, minus decimal logarithm of 50% lethal dose) of organometallic and inorganic substances. The balance of correlations was used in the Monte Carlo optimization aimed to build up optimal descriptors. It should be noted, that there are few QSAR models in the literature which are dealing with organometallic and inorganic substances. We used CORAL (CORrelations And Logic) freeware, available on the Internet, for the modelling. Ten random splits into the sub-training, calibration, and test sets have been examined. Statistical characteristics of the model (for the split 1) are the following: $n = 57$, $r^2 = 0.6005$, $Q^2 = 0.5721$, $s = 0.448$, $F = 83$ (sub-training set); $n = 55$, $r^2 = 0.6005$, $R^2_{pred} = 0.5701$, $s = 0.501$ (calibration set); $n = 12$, $r^2 = 0.8296$, $R^2_{pred} = 0.7695$, and $s = 0.233$ $R^2_m = 0.8142$ (test set). Statistical quality of models for other examined splits is also reasonable well.

1. Introduction

The majority of quantitative structure–property/activity relationships (QSPR/QSAR) described in the literature deals with the different classes of organic substances [1–6]. Very few studies cover QSPR/QSAR analyses of inorganic, organometallic, or coordination compounds, because of the lack of a suitable tool for calculating descriptors for heavy atoms [7–11].

The simplified molecular input line entry system (SMILES) [12–15] is an alternative to molecular graph for representing the molecular structure. There are QSPR/QSAR models based on SMILES for organic compounds [16–19] and SMILES-based modelling has been used for organometallic [20–22] and inorganic compounds [23]. The increasing numbers of databases on the Internet using the SMILES to represent molecular structures, is an important argument for using SMILES-based approaches in QSPR/QSAR analyses not only for inorganic substances but also for organic compounds, in spite of the widespread use of the molecular graphs (for organic substances).

The CORAL (CORrelations And Logic) is a freeware (available on the Internet [24]) for designing SMILES-based QSPR/QSAR-models. The present study examined of the CORAL as a tool for QSAR modelling toxicity of organometallic and inorganic substances towards rats.

2. Method

SMILES and numerical data on the oral lethal dose for 50% of rats (LD50): we used figures in mg/kg for organometallic and inorganic substances ($n = 124$) from the US National Library of Medicine web site [25]. The pLD50, i.e., decimal log(1/LD50) was examined as the endpoint in the QSAR analysis. The substances were randomly split into the sub-training set, calibration set, and test set by ten ways: 57–55–12; 62–49–13; 71–39–14; 75–36–13; 70–43–11; 66–42–16; 62–43–19; 69–36–19; 71–38–15; and 73–37–13.

Selected substances fall into the following categories: 1. Organic fragment–Metal–Organic fragment; 2. Organic fragment–Metal–Inorganic fragment; 3. Inorganic fragment–Metal–Inorganic fragment, where the Metal can be Li, Na, K, Cs, Mg, Ca, Ba, Cr, Mn, Fe, Co, Ni, Cu, Zn, Al, Si, As, Sb, Bi, Hg, Cd, Ag, and Au.

Optimal descriptors were calculated as the following

$$DCW(T) = \alpha\Sigma\, CW\left({}^1SA_k\right) + \beta\Sigma\, CW\left({}^2SA_k\right) + \gamma\Sigma\, CW\left({}^3SA_k\right) \quad (1)$$

where ${}^1SA_k$, ${}^2SA_k$, ${}^3SA_k$ are SMILES attributes. The ${}^1SA_k$, ${}^2SA_k$, and ${}^3SA_k$ contain one, two, and three SMILES elements, respectively. The SMILES element can be one (e.g., 'C', 'c', 'N', 'S', etc.) or two characters (e.g., 'Cl', 'Br', etc.). The order of elements in depicting the ${}^2SA_k$ or ${}^3SA_k$ is defined by the ASCII characters. In other words only one version of an AB-sequence or ABC-sequence is possible in the list of SMILES-attributes

* Corresponding author.
E-mail address: andrey.toropov@marionegri.it (A.A. Toropov).

0169-7439/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2010.12.007

(not AB together with BA, or ABC together with CBA). The CW(^1SA_k), CW (^2SA_k), and CW($^3SA^k$) are so-called correlation weights for the ^1SA_k, ^2SA_k, ^3SA_k. The correlation weights are calculated by the Monte Carlo method optimization procedure. The α, β, and γ are (0,1)-coefficients for selection of a preferable version of the DCW(T). In the present study we have used $\alpha = 1$, $\beta = 1$, and $\gamma = 0$.

The target function for this optimization procedure is

$$TF = R + R' - abs(R - R')^*dR\text{-}weight - abs(C0 + C0' + C1 - C1')^*dC\text{-}weight \quad (2)$$

where R and R' are correlation coefficients between endpoint and optimal descriptor for the sub-training set and calibration sets. The role of the calibration set is a preliminary validation of the model, as an attempt to avoid overtraining. In other words, in the case of balance of correlations [26] (i.e., $R \approx R'$), the training set is split into two sets: sub-training and calibration. The dR-weight is an empirical parameter; C0 and C0' are intercepts for the sub-training set and calibration set; C1 and C1' are slopes for the sub-training set and calibration set. The T is a threshold for the definition of rare SA_k. The total number of the SA_k involved in the modelling can be very large. However, some SA_k are rare (in the sub-training set), and these can lead to overtraining. The threshold is a parameter for defining rare attributes. For instance, if $T = 3$ and an SA_k takes place in only one or only two SMILES notations of the sub-training set, then SA_k is a rare attribute. The correlation weight of this SA_k must be fixed as zero, i.e., $CW(SA_k) = 0$ [24,26,27]. The advantage of scheme of the balance of correlations in comparison with the 'classic' scheme (i.e. training-test system) has been checked [26], so the 'classic' scheme was not used in this study.

We used for CORAL the Monte Carlo optimization for the range of thresholds from 1 to 5 [24]. We also studied how the number of epochs of the optimization influences the statistical quality of the model for the external test set.

3. Results and discussion

Fig. 1 shows co-evolutions of correlations between the DCW(4) and pLD50 for the sub-training, calibration, and test sets, for split 1. We used 50 epochs of the Monte Carlo optimization which involved three phases. In the first phase the correlation coefficient between DCW(X) and pLD50 is undefined and has a value near zero for the sub-training, calibration, and test sets. In the second phase the correlation coefficient increases for the sub-training, calibration, and test sets. In the third phase the correlation coefficient increases for sub-training and calibration sets, but decreases for the test set. Thus, the range of transition of the second to third phase is an indicator of the model with maximum predictive potential.

The correlation coefficient between the experimental LD50 and calculated LD50 is a mathematical function of the threshold and N_{epoch}. Table 1 shows statistical characteristics of the models with $N_{epoch} = 50$ and optimal values of the N_{epoch}. One can see, first, the optimal N_{epoch} is individual for each split; and second, the optimal N_{epoch} improves the statistical quality of the prediction in comparison with $N_{epoch} = 50$ (Table 1).

Analysis of the surface for the mathematical function $r^2_{test} = F$ (Threshold, N_{epoch}) shows that there is a maximum of the r^2_{test} for each split. Fig. 2 shows the surface for the case of split 1. One can use the surface in order to define the preferable number of epochs for the Monte Carlo optimization.

The majority of the substances has an 'average' behavior and is the basis for building up the pLD50 model *in the second phase*. However, there are substances with 'atypical' behavior in both the sub-training and calibration sets (Fig. 3). During *the second phase* of the Monte Carlo optimization the main contribution for building up of the model is from extraction of knowledge from the substances with 'average' behavior. When the real information contained in the substances of

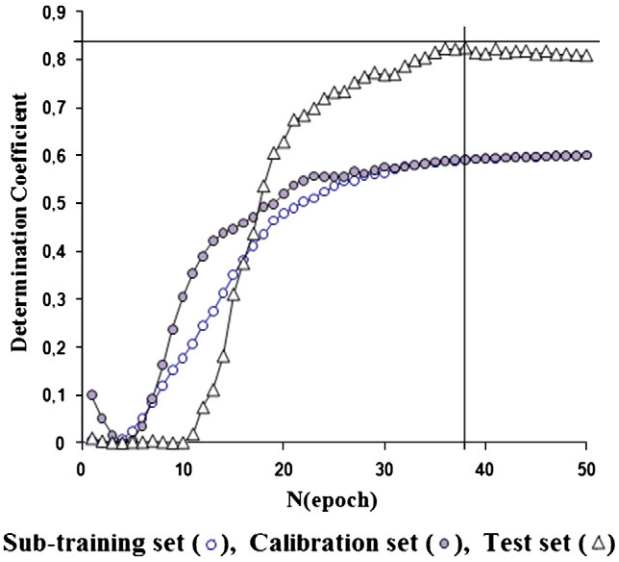

Fig. 1. Co-evolution of correlations between experimental pLD50 and the calculated pLD50 for split 1. The best prediction (i.e. maximum of correlation coefficient for the external test set) takes place if the $N(epoch) \approx 38$.

'average' behavior runs out, overtraining starts. The essence of overtraining is modification of the correlation weights of available attributes for improving only the model *for the sub-training set*. Unfortunately, that reduces the predictive potential of the model for the external test set. However, the preferable N_{epoch} can be selected by analysis of the co-evolutions of correlations (Fig. 1), the function $r^2_{test} = F(\text{Threshold}, N_{epoch})$ serves to select both the preferable N_{epoch} and the preferable Threshold (Fig. 2).

One can see (Table 1, Figs. 1 and 2) that for split 1 the preferable $N_{epoch} \approx 38$ and the preferable threshold is 4. The QSAR model for pLD50 obtained with CORAL freeware under such conditions is the following:

$$\begin{aligned} &pLD50 = -2.562(\pm 0.0122) + 0.0547(\pm 0.0008)^*DCW(4) \\ &n = 57, r^2 = 0.6005, Q^2 = 0.5721, s = 0.448, F = 83 \text{ (sub-training set)}; \\ &n = 55, r^2 = 0.6005, R^2_{pred} = 0.5701, s = 0.501 \text{(calibration set)}; \\ &n = 12, r^2 = 0.8296, R^2_{pred} = 0.7695, s = 0.233 \text{(test set)} \end{aligned} \quad (3)$$

where

$$Q^2 = 1 - \frac{\sum [Ypred - Y]^2}{\sum [Y - \overline{Y}(sub-training)]^2} \qquad \left(Y \text{ and } Y_{pred} \text{ on } sub\text{-}training \text{ set}\right)$$

$$R^2_{pred} = 1 - \frac{\sum [Ypred - Y]^2}{\sum [Y - \overline{Y}(sub-training)]^2} \qquad \left(Y \text{ and } Y_{pred} \text{ on } calibration \text{ or } test \text{ set}\right)$$

Y and Y_{pred} are experimental and predicted values of the pLD50, respectively; $\overline{Y}(sub\text{-}training)$ is an average of the experimental values of the pLD50 over the sub-training set.

In addition, we have checked the predictability of the model calculated with Eq. (3) for the test set, according to criterions of Golbraikh and Tropsha [28] and P.P. Roy and K. Roy [29]:

$$\begin{aligned} &n = 12 \\ &R^2 = 0.8296 \\ &R^2_0 = 0.8292 \\ &R'^2_0 = 0.8033 \\ &\left(R^2 - R^2_0\right)/R^2 = 0.0004 \text{ should be } < 0.1 \text{ [28]} \\ &\left(R^2 - R'^2_0\right)/R^2 = 0.0317 \text{ should be } < 0.1 \text{ [28]} \\ &k = 0.9944 \text{ should be } 0.85 < k < 1.15 \text{ [28]} \\ &k' = 0.9999 \text{ should be } 0.85 < k' < 1.15 \text{ [28]} \\ &R^2_m = R^2\left(1 - abs\left(R^2 - R^2_0\right)^{0.5}\right) = 0.8142 \text{ should be } > 0.5 \text{ [29]} \end{aligned}$$

Table 1
Statistical characteristics of the pLD50 model (toxicity toward rats) obtained with thresholds from 1 to 5 for ten random splits. The N_{act} is the number of not blocked (i.e. active) SMILES attributes. Best models obtained according to co-evolution of correlations (Fig. 1) are indicated by bold.

Threshold	N_{act}	N_{epoch}	n	r^2	s	F	n	r^2	s	n	r^2	s	R^2_m
Split 1													
1	175	50	57	0.8729	0.253	378	55	0.7564	0.479	12	0.7476	0.576	0.1748
2	130	50	57	0.8440	0.280	297	55	0.7069	0.591	12	0.5251	0.711	−0.0000
3	105	50	57	0.7237	0.373	144	55	0.6761	0.505	12	0.5297	0.503	0.2293
4	83	50	57	0.6007	0.448	83	55	0.6007	0.503	12	0.8131	0.248	0.7412
4	**83**	**38**	**57**	**0.6005**	**0.448**	**83**	**55**	**0.6005**	**0.501**	**12**	**0.8296**	**0.233**	**0.8142**
5	71	50	57	0.5723	0.463	74	55	0.5772	0.516	12	0.8135	0.247	0.7599
Split 2													
1	183	50	62	0.7428	0.415	173	49	0.9221	0.452	13	0.6809	0.305	0.6268
1	**183**	**37**	**62**	**0.7301**	**0.425**	**162**	**49**	**0.8840**	**0.401**	**13**	**0.6892**	**0.301**	**0.6590**
2	136	50	62	0.6766	0.466	126	49	0.8489	0.369	13	0.6842	0.324	0.5492
3	112	50	62	0.6413	0.490	107	49	0.7575	0.371	13	0.4894	0.409	0.3251
4	89	50	62	0.6090	0.512	93	49	0.6432	0.403	13	0.4501	0.474	0.2076
5	78	50	62	0.5776	0.532	82	49	0.6448	0.398	13	0.3727	0.481	0.1909
Split 3													
1	189	50	71	0.8153	0.343	305	39	0.9598	0.274	14	0.3057	0.912	0.0757
2	130	50	71	0.6989	0.438	160	39	0.9067	0.280	14	0.4842	0.646	0.2983
3	111	50	71	0.6320	0.484	119	39	0.8720	0.271	14	0.5872	0.517	0.4058
4	96	50	71	0.5884	0.512	99	39	0.8312	0.338	14	0.5922	0.487	0.4541
4	**96**	**37**	**71**	**0.5583**	**0.531**	**87**	**39**	**0.7997**	**0.330**	**14**	**0.6860**	**0.403**	**0.6106**
5	84	50	71	0.6030	0.503	105	39	0.7407	0.320	14	0.6317	0.533	0.3924
Split 4													
1	196	50	75	0.8238	0.312	341	36	0.9702	0.217	13	0.3180	0.599	0.1717
2	133	50	75	0.6514	0.439	136	36	0.9469	0.194	13	0.8490	0.302	0.7260
3	112	50	75	0.6056	0.467	112	36	0.8933	0.348	13	0.8998	0.236	0.8399
3	**112**	**45**	**75**	**0.5941**	**0.474**	**107**	**36**	**0.8893**	**0.347**	**13**	**0.9186**	**0.234**	**0.8779**
4	97	50	75	0.5687	0.489	96	36	0.8975	0.310	13	0.8516	0.265	0.8010
5	86	50	75	0.5313	0.510	83	36	0.8468	0.335	13	0.8072	0.304	0.6763
Split 5													
1	191	50	70	0.8000	0.338	272	43	0.9619	0.256	11	0.3795	0.450	0.2944
1	**191**	**29**	**70**	**0.7358**	**0.388**	**189**	**43**	**0.8849**	**0.262**	**11**	**0.8303**	**0.227**	**0.7305**
2	134	50	70	0.7228	0.398	177	43	0.9222	0.234	11	0.5044	0.510	0.2032
3	113	50	70	0.6220	0.464	112	43	0.8862	0.263	11	0.1614	1.015	−0.0684
4	100	50	70	0.5193	0.524	73	43	0.8801	0.303	11	0.2758	0.639	0.0521
5	88	50	70	0.4952	0.537	67	43	0.8718	0.305	11	0.2190	0.644	0.0482
Split 6													
1	191	50	66	0.8064	0.359	267	42	0.9509	0.354	16	0.5959	0.590	0.3205
2	143	50	66	0.7017	0.445	151	42	0.9354	0.337	16	0.7331	0.405	0.6032
2	**143**	**32**	**66**	**0.6566**	**0.478**	**122**	**42**	**0.8658**	**0.259**	**16**	**0.8019**	**0.324**	**0.7701**
3	119	50	66	0.5759	0.531	87	42	0.9349	0.429	16	0.7301	0.360	0.7050
4	94	50	66	0.5741	0.532	86	42	0.8591	0.348	16	0.7572	0.343	0.7125
5	73	50	66	0.5538	0.545	79	42	0.7250	0.316	16	0.6487	0.421	0.5614
Split 7													
1	179	50	62	0.8344	0.319	302	43	0.9488	0.184	19	0.0430	0.987	−0.0005
2	123	50	62	0.7083	0.423	146	43	0.8501	0.274	19	0.8159	0.320	0.7709
2	**123**	**40**	**62**	**0.6824**	**0.442**	**129**	**43**	**0.8231**	**0.296**	**19**	**0.8387**	**0.312**	**0.7636**
3	100	50	62	0.6044	0.493	92	43	0.8221	0.302	19	0.7011	0.387	0.6200
4	91	50	62	0.5567	0.522	75	43	0.7966	0.314	19	0.5205	0.480	0.4982
5	80	50	62	0.5446	0.529	72	43	0.7260	0.360	19	0.4307	0.537	0.4148
Split 8													
1	203	50	69	0.7566	0.430	208	36	0.9861	0.364	19	0.6245	0.339	0.5765
2	143	50	69	0.6278	0.532	113	36	0.9720	0.346	19	0.7031	0.310	0.6044
2	**143**	**45**	**69**	**0.6176**	**0.539**	**108**	**36**	**0.9666**	**0.336**	**19**	**0.7137**	**0.302**	**0.6324**
3	119	50	69	0.5905	0.558	97	36	0.9771	0.348	19	0.6218	0.340	0.5879
4	98	50	69	0.5320	0.597	76	36	0.9217	0.319	19	0.5757	0.366	0.4752
5	84	50	69	0.4795	0.629	62	36	0.8753	0.306	19	0.6049	0.362	0.4701
Split 9													
1	204	50	71	0.8244	0.366	324	38	0.8849	0.163	15	0.6793	0.554	0.2937
2	146	50	71	0.6801	0.494	147	38	0.9166	0.235	15	0.7301	0.387	0.5318
2	**146**	**35**	**71**	**0.6362**	**0.526**	**121**	**38**	**0.8660**	**0.227**	**15**	**0.8600**	**0.257**	**0.8099**
3	123	50	71	0.6011	0.551	104	38	0.8448	0.264	15	0.6149	0.389	0.5846
4	101	50	71	0.5384	0.593	81	38	0.8358	0.364	15	0.3021	0.481	0.2584
5	84	50	71	0.5078	0.612	71	38	0.8285	0.354	15	0.4285	0.444	0.3556
Split 10													
1	205	50	74	0.8254	0.335	340	37	0.9811	0.338	13	0.6488	0.538	0.4366
2	143	50	74	0.6861	0.449	157	37	0.9532	0.382	13	0.7368	0.333	0.6491
3	116	50	74	0.6321	0.486	124	37	0.9043	0.333	13	0.8517	0.277	0.7631
3	**116**	**50**	**74**	**0.6389**	**0.482**	**127**	**37**	**0.8979**	**0.320**	**13**	**0.8604**	**0.279**	**0.7594**
4	96	50	74	0.5523	0.536	89	37	0.8838	0.387	13	0.4697	0.413	0.4545
5	86	50	74	0.5573	0.533	91	37	0.8755	0.396	13	0.3107	0.468	0.2957

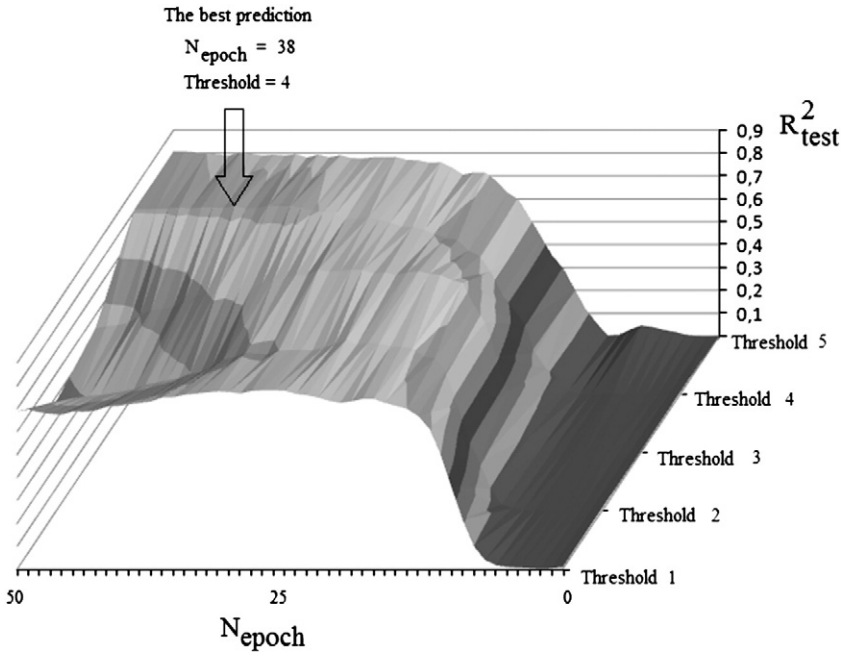

Fig. 2. Split 1: The correlation coefficient (R^2test) between experimental and calculated pLD50 for external test set is a mathematical function of the Threshold and N (epoch).

One can see that the Eq. (3) is confirmed by above-mentioned criteria [28,29]. Fig. 3 shows the model calculated with Eq. (3) graphically.

In spite of the modest statistics for the sub-training and calibration sets, one can trust these predictions, since the Monte Carlo optimization gave satisfactory statistics for ten random splits into the sub-training, calibration, and test sets (Table 1). A unique situation takes place for split 10. Preferable N(epoch) is 50. We have checked further increase of the N(epoch) is not accompanied by the increase of the statistical quality of the prediction.

To characterize the applicability domain of this model, one can consider appropriate the substances represented by the SMILES *without rare attributes* (i.e. without the rare attributes defined according to the selected threshold). Having the list of attributes extracted from a given SMILES, one can detect rare attributes by analysis of the prevalence of attributes in the sub-training set.

The molecular structures of substances, their CAS numbers, the ten random splits, the correlation weights to calculate the optimal descriptors are presented in *Supplementary Materials* section.

4. Conclusions

Analysis of co-evolution of correlation between experimental and calculated pLD50 shows that there are three phases in the Monte Carlo optimization. In the first phase there is uncertainty about the correlation coefficient for the sub-training, calibration, and test sets. In the second phase the correlation coefficient is higher for the sub-training, calibration, and test sets. In the third phase there is a further increase in the correlation coefficient for the sub-training and calibration sets, accompanied by a decrease in the correlation coefficient for the test set. The transition of the second to the third phase is an indicator of the model with maximum predictive potential. The best predictability for different splits takes place under different conditions: the range of threshold is from 1 to 4; the range of N(epoch) is approximately from 30 to 50.

Thus, the optimal SMILES-based descriptors calculated with CORAL freeware can be robust predictors for toxicity towards rats (pLD50) of organometallic and inorganic substances when appropriate thresholds are used. The modest statistical quality of the model for the sub-training and calibration sets provides good prediction for an external

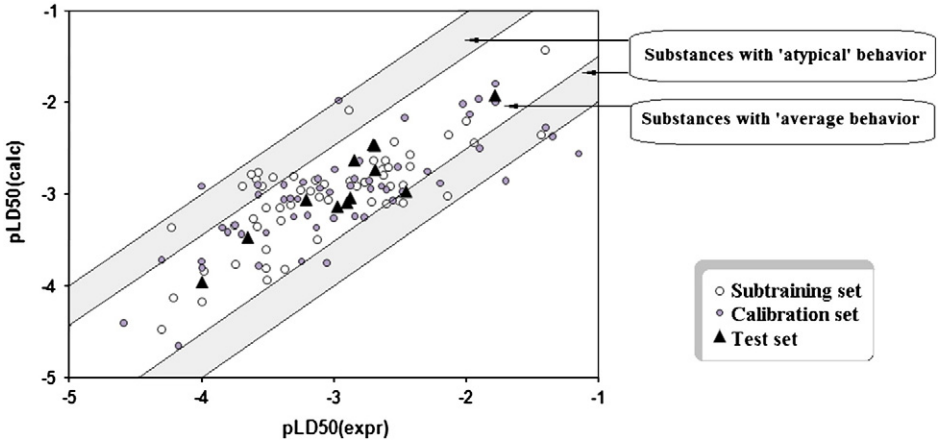

Fig. 3. Split 1: Experimental and calculated with Eq. (3) pLD50 values.

test set, while the excellent statistical quality for the sub-training and calibration sets can be an indicator of the overtraining.

Acknowledgements

The authors thank OSIRIS for financial support, and express gratitude to Dr. L. Cappellini (*Istituto di Ricerche Farmacologiche Mario Negri, Milano*) for technical assistance.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.chemolab.2010.12.007.

References

[1] A. Afantitis, G. Melagraki, H. Sarimveis, P.A. Koutentis, J. Markopoulos, O. Igglessi-Markopoulou, QSAR Comb. Sci. 25 (2006) 928–935.
[2] A. Afantitis, G. Melagraki, H. Sarimveis, P.A. Koutentis, J. Markopoulos, O. Igglessi-Markopoulou, Polymer 47 (2006) 3240–3248.
[3] E. Vicente, P.R. Duchowicz, E.A. Castro, A. Monge, J. Mol. Graph. Model. 28 (2009) 28–36.
[4] T. Puzyn, A. Mostrag, N. Suzuki, J. Falandysz, Atmos. Environ. 42 (2008) 6627–6636.
[5] T. Puzyn, N. Suzuki, M. Haranczyk, Environ. Sci. Technol. 42 (2008) 5189–5195.
[6] T. Puzyn, N. Suzuki, M. Haranczyk, J. Rak, J. Chem. Inform. Model. 48 (2008) 1174–1180.
[7] I. Gutman, B. Furtula, A.A. Toropov, A.P. Toropova, MATCH Commun. Math. Comput. Chem. 53 (2005) 225–230.
[8] L. Mu, C. Feng, H. He, MATCH Commun. Math. Comput. Chem. 53 (2007) 111–134.
[9] L.-L. Mu, H.-M. He, C.-J. Feng, Chin. J. Chem. 24 (2006) 855–861.
[10] P. Duchowicz, E.A. Castro, Russ. J. Gen. Chem. 72 (2002) 1867–1873.
[11] A.A. Toropov, A.P. Toropova, Russ. J. Coord. Chem. 24 (1998) 81–85.
[12] D. Weininger, J. Chem. Inform. Comput. Sci. 28 (1988) 31–36.
[13] D. Weininger, A. Weininger, J.L. Weininger, J. Chem. Inform. Comput. Sci. 29 (1989) 97–101.
[14] D. Weininger, J. Chem. Inform. Comput. Sci. 30 (1990) 237–243.
[15] ACD/ChemSketch Freeware, version 11.00, Advanced Chemistry Development, Inc., Toronto, ON, Canada, http://www.acdlabs.com, 2007.
[16] D. Vidal, M. Thormann, M. Pons, J. Chem. Inform. Model. 45 (2005) 386–393.
[17] A.A. Toropov, D. Leszczynska, J. Leszczynski, Chem. Phys. Lett. 441 (2007) 119–122.
[18] A.A. Toropov, A.P. Toropova, I. Raska Jr., Eur. J. Med. Chem. 43 (2008) 714–740.
[19] A.A. Toropov, E. Benfenati, Bioorg. Med. Chem. 16 (2008) 4801–4809.
[20] A.A. Toropov, A.P. Toropova, E. Benfenati, Chem. Phys. Lett. 461 (2008) 343–347.
[21] A.A. Toropov, A.P. Toropova, E. Benfenati, Cent. Eur. J. Chem. 7 (2009) 846–856.
[22] A.A. Toropov, A.P. Toropova, E. Benfenati, D. Leszczynska, J. Leszczynski, J. Comput. Chem. 31 (2010) 381–392.
[23] A.P. Toropova, A.A. Toropov, S.Kh. Maksudov, Chem. Phys. Lett. 428 (2006) 183–186.
[24] CORAL 2010 at http://www.insilico.eu/coral.
[25] US National Library of Medicine (2009), at http://www.toxnet.nlm.nih.gov/.
[26] A.A. Toropov, B.F. Rasulev, J. Leszczynski, Bioorg. Med. Chem. 16 (2008) 5999–6008.
[27] A.A. Toropov, A.P. Toropova, E. Benfenati, Int. J. Mol. Sci. 10 (2009) 3106–3127.
[28] A. Golbraikh, A. Tropsha, J. Mol, Graph. Model. 20 (2002) 269–276.
[29] P.P. Roy, K. Roy, QSAR Comb. Sci. 27 (2008) 302–313.