Environ Sci Pollut Res
DOI 10.1007/s11356-014-3977-2

RESEARCH ARTICLE

QSAR as a random event: a case of NOAEL

Alla P. Toropova · Andrey A. Toropov · Jovana B. Veselinović · Aleksandar M. Veselinović

Received: 17 October 2014 / Accepted: 9 December 2014

Abstract Quantitative structure–activity relationships (QSAR) for no observed adverse effect levels (NOAEL, mmol/kg/day, in logarithmic units) are suggested. Simplified molecular input line entry systems (SMILES) were used for molecular structure representation. Monte Carlo method was used for one-variable models building up for three different splits into the "visible" training set and "invisible" validation. The statistical quality of the models for three random splits are the following: split 1 n=180, r^2=0.718, q^2=0.712, s=0.403, F=454 (training set); n=17, r^2=0.544, s=0.367 (calibration set); n=21, r^2=0.61, s=0.44, r_m^2=0.61 (validation set); split 2 n=169, r^2=0.711, q^2=0.705, s=0.409, F=411 (training set); n=27, r^2=0.512, s=0.461 (calibration set); n=22, r^2=0.669, s=0.360, r_m^2=0.63 (validation set); split 3 n=172, r^2=0.679, q^2=0.672, s=0.420, F=360 (training set); n=19, r^2=0.617, s=0.582 (calibration set); n=21, r^2=0.627, s=0.367, r_m^2= 0.54 (validation set). All models are built according to OCED principles.

Keywords QSAR · Monte Carlo technique · Optimal descriptor · Chronic toxicity · NOAEL

Responsible editor: Michael Matthies

Electronic supplementary material The online version of this article (doi:10.1007/s11356-014-3977-2) contains supplementary material, which is available to authorized users.

A. P. Toropova · A. A. Toropov (✉)
IRCCS, Istituto di Ricerche Farmacologiche Mario Negri, 20156Via La Masa 19, Milano, Italy
e-mail: andrey.toropov@marionegri.it

J. B. Veselinović · A. M. Veselinović
Department of Chemistry, Faculty of Medicine, University of Niš, Niš, Serbia

Introduction

Humans encounter enormous amount of various chemical compounds daily in different exposure manners. In modern civilization, there are over five million man-made chemicals currently known (of which 70,000 are in use today) and about 100,000 naturally occurring chemicals of known structure (Mazzatorta et al. 2008). The evaluation of their potential influence to human health is very difficult since there are or no toxicological information available for most of compounds.

Toxicological studies are used for identification and characterization of the potential toxicity of chemical compounds. In many of them, laboratory animals are used to obtain key toxicology informations like acute or chronic toxicity, establishment of dose-response curves, and the determination of the no observed adverse effect level (NOAEL) and lowest observed adverse effect level (LOAEL). The United States Environmental Protection Agency (US EPA) defines NOAEL as an exposure level at which there are no statistically or biologically significant increases in the frequency or severity of adverse effects between the exposed population and its appropriate control; some effects may be produced at this level, but they are not considered as adverse, or as precursors to adverse effects. In an experiment with several NOAELs, the regulatory focus is primarily on the highest one, leading to the common usage of the term NOAEL as the highest exposure without adverse effects. [http://www.epa.gov/OCEPAterms/nterms.html].

NOAEL is the important information from point of view of risk assessment of various substances which take place in food industry, cosmetics, and in everyday life in general (Contrera, et al. 2004; Zhu et al. 2009; Rupp et al. 2010; Wang et al. 2011).

Also, toxicity testing of new compounds is essential for drug development process (Lewis et al. 2002; Parasuraman

Published online: 19 December 2014

2011; Park and Cho 2011). The determination of NOAEL for chemical compounds is time-consuming and a very costly process. Also, basically all laboratory testing is performed on animals. For this reasons, the information of NOAELs for hundreds of thousands or probably millions of substances is unknown. Mathematical chemistry, especially QSAR/QSPR modeling, is a field in chemistry which can help to overcome stated problems. QSPR/QSAR models build as a lineal equation of certain descriptors can be used for prediction of various endpoints (Furtula and Gutman 2011; Ojha and Roy 2011; Afantitis et al. 2011; Roy and Mitra 2012). This mathematical modeling approach can be applied for chemical compounds NOAEL and LOAEL toxicological profile prediction or estimation.

Hence, it is not surprising the development of commercial software for the NOAEL prediction (Rupp et al. 2010) to solve manifold tasks of cosmetics (Gajewska et al. 2014). As a rule, the programs aimed to solve these tasks involve many various types of descriptors (e.g. topological descriptors, quantum mechanics parameters, and physicochemical characteristics) which are able to be the basis for NOAEL models (Dobchev et al. 2013). Also, the correlations between the NOAEL and various kinds of toxicity including of the LOAEL can be a tool to predict NOAEL for new unknown substances (Zhu et al. 2009). On the one hand, these approaches owing to ability to involve all abovementioned data on new substances (to build up model) often have good predictive potential. However, on the other hand, finding or experimental definition of all necessary data sometimes can become an additional complex task. Thus, a possibility to build up model using solely the information on molecular structures together with experimental data on NOAEL is an attractive alternative for approaches based on knowledge on quantum mechanics, physicochemical parameters, and/or LOAEL.

Establishing specific one-variable QSPR/QSAR models using the Monte Carlo method can be achieved using optimal descriptors (García et al. 2011; Garro-Martinez et al. 2011; Mullen et al. 2011; Ibezim et al. 2012). Recently, the optimal descriptors calculations become available with CORAL software (http://www.insilico.eu/coral), where simplified molecular input line entry system (SMILES) (Weininger 1988; Weininger et al. 1989; Weininger 1990) were used for the representation of the molecular structure (Veselinović et al. 2013a, b; Nesmerak et al. 2013; Toropova and Toropov 2013, 2014). In fact, the NOAEL model built up with the CORAL software should be estimated as model based on information on the 2D molecular structures and the numerical experimental data on NOAEL (Goto 2013).

The main goal of the present study is the estimation of the SMILES-based optimal descriptors calculated with the CORAL software as a tool to predict of the NOAEL of various organic compounds.

Method

Data

The numerical data on no observed adverse effect levels (NOAEL, mmol/kg/day, in logarithmic units) and SMILES representation for 218 compounds are taken in the literature (Goto 2013). Splitting into the training, calibration, and validation sets has been done according to principles (i) the splits are random; (ii) these splits are different; and (iii) percentage of compounds in the calibration and validation sets is about 15 %. The examined set of 218 substances includes very diverse organic compounds which are containing fluorine, chlorine, bromine, nitrogen, oxygen, sulfur, as well as different cycles comprising of 3, 5, and 6 atoms. Supplementary materials section contains SMILES of the abovementioned compounds (Table S1).

Optimal descriptor

Optimal descriptor of correlation weights (DCW) used in this study is calculated as the following:

$$\mathrm{DCW}(T,\ N) = \sum_{k=1}^{\mathrm{NA}} \mathrm{CW}(\mathrm{SA}_k) \tag{1}$$

where SA_k are attributes of SMILES (Toropova et al. 2011) the NA is the number of SA_k which are extracted from the given SMILES. The SA_k extracted from SMILES are represented in Table 1; the $\mathrm{CW}(\mathrm{SA}_k)$ are correlation weights which are calculated by the Monte Carlo optimization gives maximal value for the correlation coefficient between NOAEL and DCW(T,N); T is the threshold used to classify SA_k into two classes: (i) rare (noise) if number SMILES in the training set which contain the SA_k is less than T, and vice versa; (ii) not rare (involved in building up model) if the number of SA_k in the training set is equal to or larger than T; the N is the number of epochs of the Monte Carlo optimization which provide maximal correlation coefficient between NOAEL and DCW(T,N) (Toropov et al. 2013). Having data on the optimal $\mathrm{CW}(\mathrm{SA}_k)$, one can develop a QSAR model, using compounds from the training set, shown as Eq. 2:

$$\mathrm{NOAEL} = C_0 + C_1 \times \mathrm{DCW}(T^*, N^*) \tag{2}$$

Table 1 SMILES attributes (SA) involved in the build-up model for NOAEL

No.	SA_k description	SA_k example	Comments and notes
1	SMILES atom	'C', '#', 'Br'	One or two symbols (which cannot be examined separately) from SMILES
2	A composition of two SMILES atoms	'CC', 'C#', 'CN'	These composition of SMILES atoms are ordered according to their ASCII codes
3	A composition of three SMILES atoms	'CNC', 'C=C', 'C#N',	These composition of SMILES atoms are ordered according to their ASCII codes of the first and third participants (Toropov et al. 2009)
4	Presence/absence of nitrogen, oxygen, sulfur, and phosphorus	'NOSP0000', 'NOSP1000', 'NOSP1100', etc.	Compact representation of information about four chemical elements (nitrogen, oxygen, sulfur, and phosphorus). NOSP0000 is indicator absence of all these chemical elements; NOSP0010 is indicator of presence of sulpur and absence of nitrogen, oxygen, and phosphorus.
6	Presence/absence of fluorine, chlorine, bromine, and iodine	'HALO0000', 'HALO1000', 'HALO1100', etc.	Compact representation of information about four chemical elements (fluorine, chlorine, bromine, and iodine). HALO0000 is indicator absence of all these chemical elements; HALO1010 is indicator of presence of fluorine and bromine and absence of chlorine and iodine.
7	Presence/absence of double, triple, and stereochemical bonds	'BOND000', 'BOND100', 'BOND010', etc.	Compact representation of information about three kinds of covalent bonds: double ('='), triple ('#') and stereochemical ('@'). BOND010 is indicator of presence of triple bond and absence of double and stereochemical bonds.
8	Presence/absence of pairs of the above-mentioned SMILES atoms	'++++O…=…', '++++#…N…', etc.	Presence of pairs of all possible SMILES atoms from (i) NOSP, (ii) HALO, and (iii) BOND

The model which is calculated with Eq. 2 should be (i) satisfactory for the "visible" calibration set and (ii) checked up with the external "invisible" validation set. Compounds from the validation set are not involved in the build-up model.

The calculations were carried out by means of the CORAL software according to method described in the literature (Veselinović et al. 2013a, b; Nesmerak et al. 2013; Toropova and Toropov 2013, 2014). The parameters T and N have

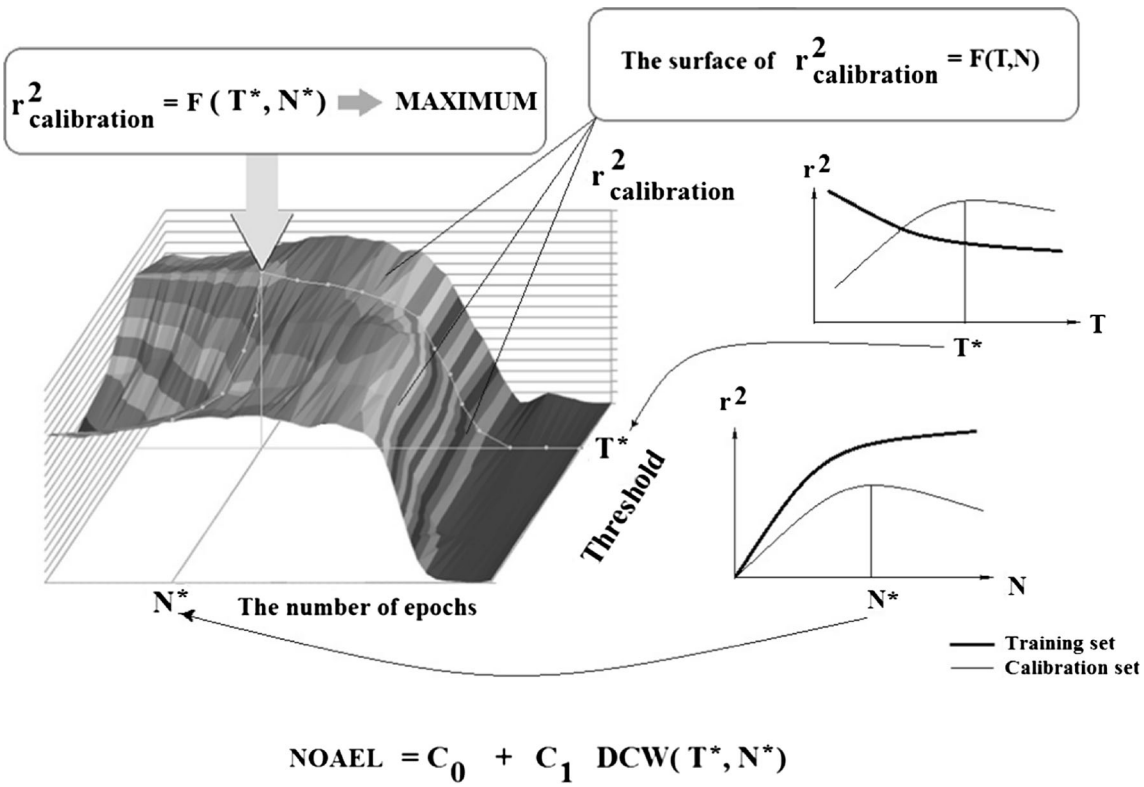

Fig. 1 Graphical representation of the selecting preferable T^* and N^*

Table 2 The definition of the T^* and N^* for three random splits in order to build up models

		Probe 1		Probe 2		Probe 3				
Split	T	N	r^2	N	r^2	N	r^2	$\overline{r^2}$ [a]	Δr^2	
1	1	5	0.3906	4	0.4048	4	0.3751	0.3902	0.0121	
	2	5	0.3823	5	0.4140	4	0.3918	0.3960	0.0133	
	3	13	0.5459	13	0.5471	13	0.5555	0.5495	0.0043	$T^*=3$, $N^*=13$
	4	8	0.5375	8	0.4951	9	0.5223	0.5183	0.0175	
	5	7	0.5020	7	0.5161	11	0.4954	0.5045	0.0086	
2	1	5	0.3590	4	0.3727	6	0.3527	0.3615	0.0084	
	2	4	0.4103	4	0.4110	15	0.4501	0.4238	0.0186	
	3	15	0.4505	15	0.5201	15	0.5236	0.4981	0.0337	$T^*=3$, $N^*=15$
	4	6	0.4358	6	0.4475	6	0.4236	0.4356	0.0097	
	5	3	0.3479	5	0.3785	6	0.4059	0.3774	0.0237	
3	1	9	0.6000	8	0.3868	8	0.3868	0.4934	0.1066	
	2	10	0.5822	9	0.5572	9	0.5572	0.5697	0.0125	
	3	9	0.6063	9	0.5970	9	0.5970	0.6016	0.0046	
	4	13	0.6605	10	0.5415	10	0.5415	0.6010	0.0595	
	5	9	0.6289	9	0.6371	9	0.6371	0.6330	0.0041	$T^*=5$, $N^*=9$

[a] The $\overline{r^2}$ and Δr^2 are average and dispersion for the square of correlation coefficient

influence upon the statistical characteristics of the model for both the training set and calibration set. One should expect that model which is good for the calibration set is better than a model which is good for the training set. Consequently, one can select T^* and N^* which are preferable for the calibration set (Toropov et al. 2013). Figure 1 contains the graphical interpretation of such selection for T^* and N^*.

QSAR model validation

The main goal of any QSPR modeling is developing a robust model capable of predicting the property of new molecules in objective, reliable, and precise manner (Golbraikh and Tropsha 2002). The robustness and reliability of developed QSAR models was estimated using r_m^2 metric described in the literature (Roy et al. 2008). Stated methodology for QSAR model validation was successfully applied to various QSAR models based on Monte Carlo method (Veselinović et al. 2013a, b; Nesmerak et al. 2013; Toropova and Toropov 2013). Also, an important note to this research is that all built QSAR models based on Monte Carlo method and SMILES-based optimal descriptors are in accordance with OCED principles (OECD 2007).

Results and discussion

The preferable values T^* and N^* (Toropova et al. 2011) together with statistical characteristics of preliminary models are represented in Table 2.

Table 3 contains the statistical characteristics of models for NOAEL calculated with three different splits into the training set, calibration set, and validation set using preliminary

Table 3 The statistical characteristics of NOAEL models calculated with Eqs. 3–5

		Training set					Calibration set			Validation set			
Split	Descriptor	n	r^2	q^2	s	F	n	r^2	s	n	r^2	s	r_m^2
1	DCW(3,13)	180	0.718	0.712	0.40	454	17	0.54	0.37	21	0.61	0.44	0.61
2	DCW(3,15)	169	0.711	0.705	0.41	411	27	0.51	0.46	22	0.67	0.36	0.63
3	DCW(5,9)	172	0.679	0.672	0.42	360	19	0.62	0.58	21	0.63	0.37	0.54

n number of compounds in a set, r^2 square of correlation coefficient, q^2 leave-one-out r^2, s standard error of estimation, F Fischer F-ratio, r_m^2 metric of predictability according to the literature: a model has predictive potential if $r_m^2 > 0.5$ (Roy et al. 2012a, b)

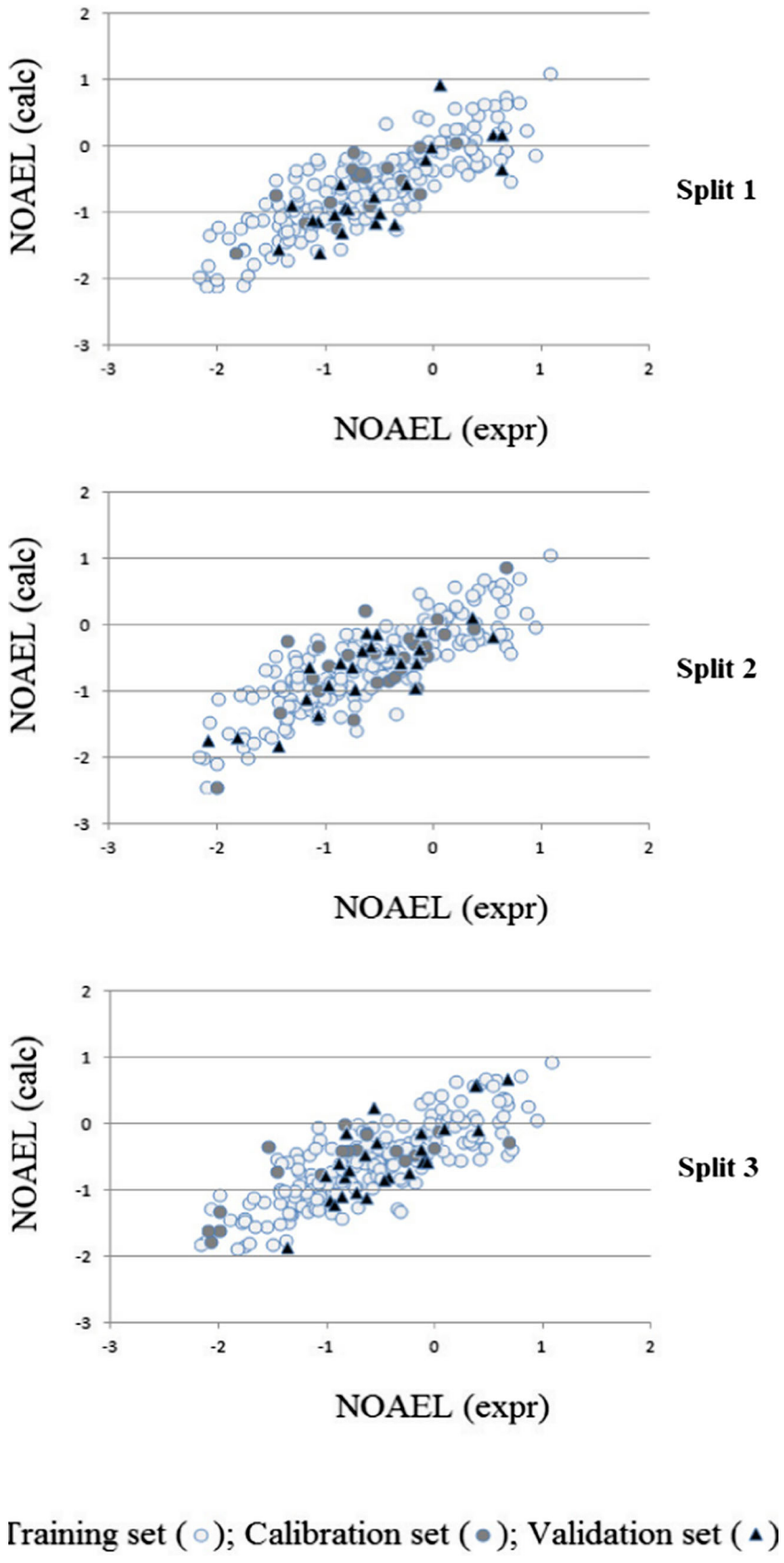

Fig. 2 Graphical representation of models for NOAEL. Training set (*white circle*); calibration set (*grey circle*); validation set (*black triangle*)

Table 4 Examples of SA_k which have mechanistic interpretation

No	SA_k	Comment
		Promoter of NOAEL decrease
1	'('	Branching of the molecular skeleton
2	'='	Double bonds
3	'1'	Presence of cycles
		Promoter of NOAEL increase
1	HALO0000	Absence of halogens
2	'O='	Presence of oxygen atoms connected via double bond
3	++++N…O…	Presence in the molecular structure nitrogen and oxygen

defined the above mentioned T^* and N^* (Fig. 1). These models are the following:

The optimal descriptor gives the following models for NOAEL:

Split 1

$$\mathrm{NOAEL} = -1.8563\ (\pm 0.0039) + 0.0934\ (\pm 0.0003) \times \mathrm{DCW}(3,13) \quad (3)$$

Split 2

$$\mathrm{NOAEL} = -1.6436\ (\pm 0.0038) + 0.0903\ (\pm 0.0003) \times \mathrm{DCW}(3,15) \quad (4)$$

Split 3

$$\mathrm{NOAEL} = -1.7654\ (\pm 0.0044) + 0.0661\ (\pm 0.0002) \times \mathrm{DCW}(5,9) \quad (5)$$

Figure 2 contains graphical representation of the model calculated with Eqs. 3–5.

Supplementary materials section contains (i) SMILES, experimental and predicted NOAEL values (logarithmic scale) for 218 compounds examined in this work; (ii) three random splits into the training set, calibration set, and test set; and (iii) list of SMILES attributes and their correlation weights (Table S2).

Statistical characteristics of model for NOAEL described in the literature are the following: r^2=0.35 (training set) and r^2=0.21 (test set) (Goto 2013). The statistical characteristics for LOAEL in logarithmical units described in the literature (Mazzatorta et al. 2008) are the following: r^2=0.54, s=0.700. The statistical characteristics of the best NOAEL model calculated with using quantum mechanics descriptors together with physicochemical parameters are n=90 and r^2=0.78 for training set and n=15 and r^2=0.68 for test set (Dobchev et al. 2013). Thus, the statistical characteristics of NOAEL models suggested in this work are comparable with statistical characteristics of QSAR models described in the literature (Mazzatorta et al. 2008; Goto 2013; Dobchev et al. 2013).

In the case of several probes of the Monte Carlo optimization, one can obtain data for classify SA_k into the following classes: (i) stable promoters of endpoint increase if correlation weight is positive in several probes of the optimization; (ii) stable promoters of endpoint decrease if correlation weight is negative in several probes of the optimization; and (iii) SA_k which have unclear role. Partially, SA_k can be interpreted (e.g. 'C' is representation of carbon atom; 'CN' is fragment that contains carbon and nitrogen, etc.), but unfortunately SA_k with complex interpretation (or without interpretation at all) also take places in the model. Thus, one can extract only probabilistic and partial interpretation for the model. Taken into account these circumstances, the list of SA_k which can be useful to search for mechanistic interpretation of the model calculated with Eqs. 3–5 is represented in Table 4. Naturally, it is necessary to take into account frequencies of SA_k in the training set. The lists of promoters of increase or decrease of

Table 5 The compliance to the OECD principles

	OECD principles	How a principle is taken into account in this work?
1	A defined endpoint	No observed adverse effect levels (NOAEL) mmol/kg/day, in logarithmic units.
2	An unambiguous algorithm	The Monte Carlo optimization which is represented by the CORAL software available on the Internet.
3	A defined domain of applicability	The domain of applicability is defined by means of the percentage of molecular features with defined role (promoters of increase or decrease for the endpoint).
4	Appropriate measures of goodness-of-fit, robustness and predictivity	The first, r_m^2 metric (Roy et al. 2012a, b) of predictive potential is involved as an addition criterion of robustness for models, and the second, the approach has been checked up with three different random splits into "visible" training set and "invisible" validation set.
5	A mechanistic interpretation, if possible	Lists of stable promoters of endpoint increase and stable promoters of endpoint decrease are represented in Table 4.

NOAEL can be extended if additional experimental data will become available in the future.

Organization for economic cooperation and development (OECD) has suggested five principles for estimation of a QSPR/QSAR model (OECD 2007). Table 5 indicates that models for NOAEL calculated with Eqs. 3–5 reflect these principles (OECD 2007). Thus, the optimal descriptor gives reasonable well model for the chronic toxicity expressed via logarithm scale of the NOAEL.

Conclusion

Optimal descriptors can be a tool for predicting the chronic toxicity of chemical compounds. The robust QSAR model for NOAEL build with the application of Monte Carlo method is presented. The statistical quality is reproduced for a group of different splits. Consequently, the described approach of the QSAR analysis is organized in accordance with OECD principles (OECD 2007).

References

Afantitis A, Melagraki G, Koutentis PA, Sarimveis H, Kollias G (2011) Ligand-based virtual screening procedure for the prediction and the identification of novel β-amyloid aggregation inhibitors using Kohonen maps and Counterpropagation Artificial Neural Networks. Eur J Med Chem 46(2):497–508

Contrera JF, Matthews EJ, Kruhlak NL, Benz RD (2004) Estimating the safe starting dose in phase I clinical trials and no observed effect level based on QSAR modeling of the human maximum recommended daily dose. Regul Toxicol Pharm 40(3):185–206

Dobchev DA, Tulp I, Karelson G, Tamm T, Tämma K, Karelson M (2013) Subchronic oral and inhalation toxicities: a challenging attempt for modeling and prediction. Mol Inf 32(9–10):793–801

Furtula B, Gutman I (2011) Relation between second and third geometric-arithmetic indices of trees. J Chemometr 25(2):87–91

Gajewska M, Worth A, Urani C, Briesen H, Schramm K-W (2014) Application of physiologically-based toxicokinetic modelling in oral-to-dermal extrapolation of threshold doses of cosmetic ingredients. Toxicol Lett 227(3):189–202

García J, Duchowicz PR, Rozas MF, Caram JA, Mirífico MV, Fernández FM, Castro EA (2011) A comparative QSAR on 1,2,5-thiadiazolidin-3-one 1,1-dioxide compounds as selective inhibitors of human serine proteinases. J Mol Graph Model 31:10–19

Garro-Martinez JC, Duchowicz PR, Estrada MR, Zamarbide GN, Castro EA (2011) QSAR study and molecular design of open-chain enaminones as anticonvulsant agents. Int J Mol Sci 12(12):9354–9368

Golbraikh A, Tropsha A (2002) Beware of q2! J Mol Graph Model 20(4):269–276

Goto, T., 2013. QSAR modeling using a set of intermediate-duration oral NOELs (2013), PhD thesis, https://etd.library.emory.edu/view/record/pid/emory:d724n

Ibezim E, Duchowicz PR, Ortiz EV, Castro EA (2012) QSAR on arylpiperazine derivatives with activity on malaria. Chemometr Intell Lab Syst 110(1):81–88

Lewis RW, Billington R, Debryune E, Gamer A, Lang B, Carpanini F (2002) Recognition of adverse and nonadverse effects in toxicity studies. Toxicol Path 30(1):66–74

Mazzatorta P, Estevez MD, Coulet M, Schilter B (2008) Modeling oral rat chronic toxicity. J Chem Inf Model 48(10):1949–1954

Mullen LMA, Duchowicz PR, Castro EA (2011) QSAR treatment on a new class of triphenylmethyl-containing compounds as potent anticancer agents. Chemometr Intell Lab Syst 107(2):269–275

Nesmerak K, Toropov AA, Toropova AP, Kohoutova P, Waisser K (2013) SMILES-based quantitative structure-property relationships for half-wave potential of N-benzylsalicylthioamides. Eur J Med Chem 67:111–114

OECD, 2007, Guidance document on the validation of (quantitative) structure-activity relationships [(Q)Sar] models, http://www.oecd.org/dataoecd/55/35/38130292.pdf

Ojha PK, Roy K (2011) Comparative QSARs for antimalarial endochins: importance of descriptor-thinning and noise reduction prior to feature selection. Chemometr Intell Lab Syst 109(2):146–161

Parasuraman S (2011) Toxicological screening. J Pharmacol Pharmacoth 2(2):74–79

Park Y-C, Cho M-H (2011) A new way in deciding NOAEL based on the findings from GLP-toxicity test. Toxicol Res 27(3):133–135

Roy K, Mitra I (2012) Electrotopological state atom (E-state) index in drug design, QSAR, property prediction and toxicity assessment. Curr Comput-Aid Drug Des 8(2):135–158

Roy PP, Leonard JT, Roy K (2008) Exploring the impact of size of training sets for the development of predictive QSAR models. Chemometr Intell Lab Syst 90(1):31–42

Roy K, Mitra I, Ojha PK, Kar S, Das RN, Kabir H (2012a) Introduction of r_m^2 (rank) metric incorporating rank-order predictions as an additional tool for validation of QSAR/QSPR models. Chemometr Intell Lab Syst 118:200–210

Roy K, Mitra I, Kar S, Ojha PK, Das RN, Kabir H (2012b) Comparative studies on some metrics for external validation of QSPR models. J Chem Inf Model 52(2):396–408

Rupp B, Appel KE, Gundert-Remy U (2010) Chronic oral LOAEL prediction by using a commercially available computational QSAR tool. Arch Toxicol 84(9):681–688

Toropov AA, Toropova AP, Benfenati E (2009) QSPR modeling bioconcentration factor (BCF) by balance of correlations. Eur J Med Chem 44(6):2544–2551

Toropov AA, Toropova AP, Puzyn T, Benfenati E, Gini G, Leszczynska D, Leszczynski J (2013) QSAR as a random event: modeling of nanoparticles uptake in PaCa2 cancer cells. Chemosphere 92(1):31–37

Toropova AP, Toropov AA (2013) Optimal descriptor as a translator of eclectic information into the prediction of membrane damage by means of various TiO2 nanoparticles. Chemosphere 93(10):2650–2655

Toropova AP, Toropov AA (2014) CORAL software: prediction of carcinogenicity of drugs by means of the Monte Carlo method. Eur J Pharm Sci 52(1):21–25

Toropova AP, Toropov AA, Benfenati E, Gini G, Leszczynska D, Leszczynski J (2011) CORAL: quantitative structure-activity relationship models for estimating toxicity of organic compounds in rats. J Comput Chem 32(12):2727–2733

Veselinović AM, Milosavljević JB, Toropov AA, Nikolić GM (2013a) SMILES-based QSAR model for arylpiperazines as high-affinity 5-HT1A receptor ligands using CORAL. Eur J Pharm Sci 48(3):532–541

Veselinović AM, Milosavljević JB, Toropov AA, Nikolić GM (2013b) SMILES-Based QSAR models for the calcium channel-antagonistic effect of 1,4-dihydropyridines. Arch Pharm 346:134–139

Wang Y-J, Dou J, Cross KP, Valerio LG (2011) Computational analysis for hepatic safety signals of constituents present in botanical extracts

widely used by women in the United States for treatment of menopausal symptoms. Regul Toxicol Pharm 59(1):111–124

Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci 28:31–36

Weininger D (1990) Smiles. 3. Depict. Graphical depiction of chemical structures. J Chem Inf Comput Sci 30(3):237–243

Weininger D, Weininger A, Weininger JL (1989) SMILES. 2. Algorithm for generation of unique SMILES notation. J Chem Inf Comput Sci 29:97–101

Zhu H, Ye L, Richard A, Golbraikh A, Wright FA, Rusyn I, Tropsha A (2009) A novel two-step hierarchical quantitative structure-activity relationship modeling work flow for predicting acute toxicity of chemicals in rodents. Environ Health Persp 117(8):1257–1264