

Article

The System of Self-Consistent Models: QSAR Analysis of Drug-Induced Liver Toxicity

Alla P. Toropova *, Andrey A. Toropov, Alessandra Roncaglioni and Emilio Benfenati *

Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Via Mario Negri 2, 20156 Milano, Italy; andrey.toropov@marionegri.it (A.A.T.); alessandra.roncaglioni@marionegri.it (A.R.)

* Correspondence: alla.toropova@marionegri.it (A.P.T.); emilio.benfenati@marionegri.it (E.B.); Tel.: +39-02-3901-4595 (E.B.); Fax: +39-02-3901-4735 (E.B.)

Abstract: Removing a drug-like substance that can cause drug-induced liver injury from the drug discovery process is a significant task for medicinal chemistry. *In silico* models can facilitate this process. Semi-correlation is an approach to building *in silico* models representing the prediction in the active (1)—inactive (0) format. The so-called system of self-consistent models has been suggested as an approach for two tasks: (i) building up a model and (ii) estimating its predictive potential. However, this approach has been tested so far for regression models. Here, the approach is applied to building up and estimating a categorical hepatotoxicity model using the CORAL software. This new process yields good results: sensitivity = 0.77, specificity = 0.75, accuracy = 0.76, and Matthew correlation coefficient = 0.51 (all compounds) and sensitivity = 0.83, specificity = 0.81, accuracy = 0.83 and Matthew correlation coefficient = 0.63 (validation set).

Keywords: drug-induced liver injuries; hepatotoxicity; Monte Carlo method; index of ideality of correlation (IIC); CORAL software

Citation: Toropova, A.P.; Toropov, A.A.; Roncaglioni, A.; Benfenati, E. The System of Self-Consistent Models: QSAR Analysis of Drug-Induced Liver Toxicity. *Toxics* **2023**, *11*, 419. <https://doi.org/10.3390/toxics11050419>

Academic Editor: Jong-Choon Kim

Received: 15 March 2023

Revised: 11 April 2023

Accepted: 25 April 2023

Published: 29 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The liver is highly susceptible to drug insults: around 5–10% of adverse drug reactions result in liver injuries [1]. Naturally, this stimulates the search for reliable models to anticipate and avoid this dangerous toxicity [2]. More than 1100 chemical substances applied daily have been identified as potentially causing liver injuries [3–5]. The clinical impact may be vary, provoking oxidative stress, an increase in the level of liver enzymes (cytochromes P450), and a dangerous impact on metabolism [5–7].

In silico models can help predict adverse effects and plan safer drugs before their complete development. Of course, these models have limits. This is a general issue since experimental studies also have limits of different types, such as the time and costs needed and ethical concerns regarding the use of animals.

“The idea of approximation dominates all exact science” (Bertrand Russell). Quantitative structure–activity relationships (QSARs) are an example of science where approximation is relevant. QSAR should be considered a surrogate of a real experiment with some limits. Even though “all models are wrong” [8], “some of them are useful” [9]. Therefore, the point is to develop “useful models”. This refers to purpose and ambition, and how far we go with a model. For screening purposes, for instance, models for an initial evaluation are acceptable even if they have greater uncertainty. However, models for the final evaluation require much less uncertainty.

We aim to develop some simple, fast models for the first evaluation of large collections of substances. This is suitable for the endpoint we are addressing in the present case: drug-induced liver injuries (DILI). This relates to many toxicological mechanisms involving complexity. At the basis of our model, as with QSAR models in general, there are data collections with experimental values. These data serve to extract the correct information,

but as we said, we must verify that the model is “good”, which is achieved with other data not used to build the model. In practice, the original collection of data is split into a training and a validation set, and avoiding the QSAR model becoming too good depends on the precise distribution of substances between these sets. The present paper describes a study on a group of random models to gain a balanced and robust model representative of multiple conditions. This is achieved with our system of self-consistent models applied to DILI. This study is pertinent from the methodological and practical points of view, as the results can be used both to predict DILI and to assess the reliability of the approach employed to build up the group of models.

2. Materials and Methods

The experimental data ($n = 1274$) on DILI in the active (denoted by 1) and inactive formats (denoted by 0) were taken from the literature [1]. The compounds were randomly distributed into active training ($\approx 25\%$), passive training ($\approx 25\%$), calibration ($\approx 25\%$), and validation sets ($\approx 25\%$). Each set has a defined task. The active training set is used to build the model: molecular features extracted from SMILES of the active training set are used in Monte Carlo optimization with the CORAL software (<http://www.insilico.eu/coral>) (accessed on 25 April 2023) to provide correlation weights (CW) for these features, which give the largest correlation coefficient between descriptors (the sum of the CW) and the endpoint of the active training set. The passive training set serves to check whether the model for the active training set is satisfactory for substances that were not involved in the active training set. The calibration set should detect the start of overtraining (or overfitting). At the beginning of the optimization, the correlation coefficients between the experimental values of the endpoint and the descriptor rise for all sets, but the correlation coefficient for the calibration set reaches the maximum (this is the start of the overtraining), and further optimization leads to a decrease in the correlation coefficient for the calibration set. The optimization should be stopped when overtraining starts.

After stopping the Monte Carlo optimization procedure, the validation set is used to assess the predictive potential of the model. The present study applied semi-correlations [10–12]. The essence of this approach is the construction of a regression model for a set of compounds characterized by 1 (if the compound is active) or 0 (if the compound is inactive). Ideally, all active and inactive ones are above the regression line.

2.1. Optimal SMILES-Based Descriptors

The optimal descriptor forms the basis of the model suggested here. The descriptor is calculated as follows:

$$DCW(T, N) = \sum CW(S_k) + \sum CW(SS_k) + \sum CW(SSS_k) \quad (1)$$

T is an integer to separate SMILES attributes into rare and non-rare. The non-rare SMILES are applied to build up the model. The rare SMILES are not used for this. The selection of a value for T is empirical, according to the results of preliminary probes of building up the model (usually, it can be 1, 2, or 3). N is the number of epochs of the correlation weight (CW) optimization. S_k is a SMILES atom, i.e., one symbol of a SMILES line (e.g., C, O) or a group of symbols that cannot be examined separately (e.g., Cu, %11). $CW(S_k)$ are the correlation weights of the SMILES attributes. SS_k and SSS_k are compositions of, respectively, two and three SMILES atoms. $CW(SS_k)$ and $CW(SSS_k)$ are the correlation weights of the SMILES fragments. The CW are obtained through the Monte Carlo method [10–12].

The optimal SMILES-based descriptor $DCW(T, N)$ is applied for a model to predict DILI using the equation:

$$y = C_0 + C_1 \times DCW(T, N) \quad (2)$$

y defines the category of a compound:

$$\text{Category}(\text{SMILES}) = \begin{cases} 1 \text{ (active)} & \text{if } y \geq 0.5 \\ 0 \text{ (inactive)} & \text{if } y < 0.5 \end{cases} \quad (3)$$

2.2. Monte Carlo Optimization

Equation (1) needs the numerical data on the CW, and the Monte Carlo optimization is used to calculate these. Here, two target functions (TF_0 and TF_1) for the Monte Carlo optimization are examined:

$$TF_0 = r_{AT} + r_{PT} - |r_{AT} - r_{PT}| \times 0.1 \quad (4)$$

$$TF_1 = TF_0 + IIC_c \times 0.5 \quad (5)$$

r_{AT} and r_{PT} are correlation coefficients between the observed and predicted endpoints for the active and passive training sets. IIC_c is the index of ideality of correlation [13]. Recent computational experiments have shown [13] that considering the IIC_c -value in the Monte Carlo optimization may be useful. IIC_c is calculated with data on the calibration set as follows:

$$IIC_c = r_c \frac{\min(-MAE_c, +MAE_c)}{\max(-MAE_c, +MAE_c)} \quad (6)$$

$$\min(x, y) = \begin{cases} x, & \text{if } x < y \\ y, & \text{otherwise} \end{cases} \quad (7)$$

$$\max(x, y) = \begin{cases} x, & \text{if } x > y \\ y, & \text{otherwise} \end{cases} \quad (8)$$

$$-MAE_c = \frac{1}{-N} \sum |\Delta_k|, \quad -N \text{ is the number of } \Delta_k < 0 \quad (9)$$

$$+MAE_c = \frac{1}{+N} \sum |\Delta_k|, \quad +N \text{ is the number of } \Delta_k \geq 0 \quad (10)$$

$$\Delta_k = \text{observed}_k - \text{calculated}_k \quad (11)$$

The corresponding formulae accompany all values; r_c is the correlation coefficient between the observed and calculated values of the endpoint in the calibration set and \textcircled{c} indicates that it belongs to the calibration set. Observed and calculated are the corresponding values of y .

2.3. The System of Self-Consistent Models

The system of self-consistent models [14] for five random splits into the training (visible) and validation (invisible) sets confirms the high quality (predictive potential) of the models. The training set here is divided into active, passive training, and calibration sets. Thus, the difference between models reflects the difference in training sets. However, the key attribute of the system of self-consistent models is the unified method for validating these models; each i th model has an i th validation set. The validation sets are far from identical (Table S1, Supplementary Materials). This supports the statistical fact that we explore multiple conditions, and the results are representative of a set of cases, each obtained by chance, and their overall results should be evaluated jointly.

The measure of self-consistency is based on the average and dispersion of the Matthews correlation coefficient (MCC) in different validation sets. The corresponding computational experiments are represented by the following matrix:

$$\begin{bmatrix} (M_1(V_1): V_1 \rightarrow MCCv_{11}) & \cdots & (M_5(V_5): V'_1 \rightarrow MCCv_{51}) \\ \vdots & & \vdots \\ (M_1(V_1): V'_5 \rightarrow MCCv_{15}) & \cdots & (M_5(V_5): V_5 \rightarrow MCCv_{55}) \end{bmatrix} \quad (12)$$

M_i is an i th model, V'_j is the list of compounds employed as the validation set in the case of the j th split, and $MCCv_{ij}$ is the Matthews correlation coefficient for the j th validation set if applied to the i th model. Figure S1, in the Supplementary Materials, shows the general scheme of validation of model 1 with validation set 3.

3. Results

We applied different methods using the CORAL software. Some models were developed with the classical approach, as in Equation (4), while others used a more recent approach, using Equation (5). Figure 1 shows the pattern of the Monte Carlo optimization of the correlation coefficient between the experimental and calculated y -values for the active training, passive training, calibration, and validation sets in the case of the Monte Carlo optimization without IIC (Figure 1a) or optimized with IIC (Figure 1b).

The determination coefficients of the training set increase slowly and continuously. Nevertheless, the patterns for the values of the other two sets, calibration and validation, are different, in particular for the validation. In Figure 1a, they peak earlier. As a consequence, depending on the epoch number, the results vary. The values for the two training sets (passive and active) are not a good indicator of the results when new substances are predicted. On increasing the number of epochs, overfitting starts (Figure 1a). It is also clear that in our conditions, the training set values often differ from those of the other sets, and depending on the epoch, the highest or lowest values of the training sets appear.

Figure 1b shows the Matthews correlation coefficients for the active training, passive training, calibration, and validation sets in the case of the optimization with IIC (Equation (5)). In this case, the training set provides general features that are useful for the model, and the calibration set optimizes these features. The overall process is successful since the results for the validation set in Figure 1b are better than those in Figure 1a.

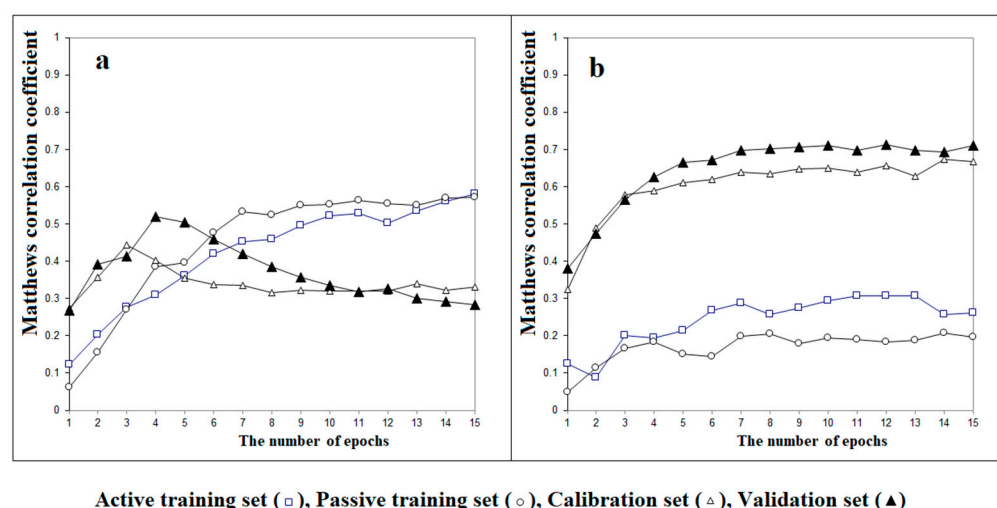


Figure 1. The evolution histories of the Matthews correlation coefficients during the Monte Carlo optimization for the active training set, the passive training set, the calibration set, and the validation set: (a) the optimization without IIC and (b) the optimization with IIC .

Table 1 illustrates the acceptable statistical quality of the approach, with the reproducibility of the predictive potential for all variations, expressed as

$$M_i(V_i): V'_j \rightarrow MCCv_{ij}, i \neq j \quad (11)$$

since the average $\overline{MCCv_{ij}} = 0.7634 \pm 0.0528$.

Table 1. The statistical characteristics of the DILI model observed for split 1.

Set	Observed Classification Quality					Statistical Characteristics			
	TP	TN	FP	FN	N	Sensitivity	Specificity	Accuracy	MCC
Active training	119	99	49	47	314	0.7169	0.6689	0.6983	0.3861
Passive training	123	101	32	63	319	0.6613	0.7594	0.7022	0.4150
Calibration	160	100	28	31	319	0.8377	0.7813	0.8150	0.6167
Validation	181	86	20	35	322	0.8380	0.8113	0.8292	0.6300
Total	583	386	129	176	1274	0.7681	0.7495	0.7606	0.5116

The comparison of the models described here with the statistical quality of DILI models from the literature (Table 2) confirms the practical and heuristic potential of the developed approach.

Table 2. The statistical quality of the DILI models reported in the literature.

N	Sensitivity	Specificity	Accuracy	Sensitivity (Validation Set)	Specificity (Validation Set)	References
-	0.73	0.73	-	-	-	[2]
6853	0.91	0.53	0.79	-	-	[3]
1550	0.76	0.71–0.92	-	-	-	[4]
1148	-	-	-	0.68–0.76	0.83–0.99	[4]
1254	0.82	0.75	0.78	-	-	[5]
83	-	-	-	0.818	0.748	[5]
1036	0.82–0.90	0.55–0.64	0.71–0.75	-	-	[6]
1274	0.77	0.75	0.76	-	-	This work (split 1)
322	-	-	-	0.838	0.8113	This work (split 1)

4. Discussion

The approach considered here is based on the use of semi-correlations. The latter are specific kinds of correlations where one variable takes only two values (for example, 0 and 1), expressing the presence (1) or the absence (0) of some modeled activity. The concept of categorical (binary classification) simulation has been successfully tested for several types of biological activity [15–21]. The convenience of the practical application of the concept of semi-correlation lies in the possibility of using SMILES to represent the molecular structure without requiring additional descriptors. The hidden analogy of semi-correlations with the usual linear regression helps the perception and interpretation of the resulting models. The necessary CORAL software is available on the internet (<http://www.insilico.eu/coral>) (accessed on 25 April 2023). The models considered here are comparable in their predictive potential with models obtained through more complex calculations using random forest procedures, support vector machines, gradient methods, and others [2–6].

Of particular note are actually new principles for testing the predictive potential of models that can be used for any similar models aimed at developing categorical binary models of biological activity. Criticism of cross-validation has a long history. The essence of claims of Q^2 is its weak relationship with the predictive ability of models [22]. The Q^2 is present as a characteristic of the model; however, this criterion has lost confidence. The first attempt to achieve a plausible test of the predictive potential of the model was to use an external set of test substances unknown at the time of model development called the validation set [23]. However, if we assume that the model can be built on an

arbitrary distribution of data in the training and testing sets, then any model being constructed based on a random split should be considered as some random event described by statistical criteria (most significant ones, are related to the validation set) [24].

In view of what has been said, the system of self-consistent models looks very attractive from the point of view of justifying the use of an appropriate approach to constructing the models. A test of this scheme for the case of ordinary regression models showed the convenience of using the approach for linear regression models [14,25–29]. Since the system of self-consistent models implies the consideration of groups of models with the allocation of average values of statistical parameters, this approach provides more objective information about the statistical significance of the approach used.

The index of ideality of correlations is an important component of the considered scheme for constructing self-consistent models. This criterion for the predictive potential of linear regression models has found a number of applications for the development of various endpoints related to organic substances [30–34], polymers [26], and nanomaterials [35–40]. The universality and attractiveness of *IIC* applications are probably due to the fact that this index contains information on correlation and average absolute error. It should be noted that the self-consistency methodology of models and the use of *IIC* are two innovations in modeling that harmoniously complement each other.

Thus, the predictive potential of corresponding models is similar for all random splits. However, the model applying the *IIC*, i.e., the Monte Carlo optimization using the target function calculated with Equation (5), is better. Table 2 compares models for the DILI reported in the literature.

The advantage of our model is that it is simple: it does not require the calculation of molecular descriptors since only the SMILES are used. The results are optimized towards the prediction of new substances, remaining far from overfitting. Even if this may reduce the results of the training set, it is a good approach for predicting other substances. The detection of outliers for models based on semi-correlations is carried out using the values of statistical defects described in the literature [10].

5. Conclusions

We have introduced some new models for DILI using the CORAL software. The statistical quality of models developed here confirms (i) the suitability of the index of ideality of correlation as the criterion for the predictive potential and (ii) the expediency of the system of self-consistent models as the rational method for the validation of QSAR models for DILI.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/toxics11050419/s1>, Table S1 contains MCC values for validation set after removing compounds present in the training set; and Table S2 contains technical details on split 1. Figure S1 demonstrated the general scheme of checking up of model-1 with validation set-3.

Author Contributions: Conceptualization, A.P.T., A.A.T., A.R., E.B.; data curation, A.P.T., A.A.T., A.R., E.B.; writing—original draft preparation, A.P.T., A.A.T., A.R., E.B.; writing—review and editing, A.P.T., A.A.T., A.R., E.B.; supervision, A.R., E.B.; project administration, E.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by ONTOX, grant agreement 963845 of the European Commission under the Horizon2020 research and innovation framework program.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This work was supported by ONTOX, grant agreement 963845 of the European Commission under the Horizon2020 research and innovation framework program.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Liew, C.Y.; Lim, Y.C.; Yap, C.W. Mixed learning algorithms and features ensemble in hepatotoxicity prediction. *J. Comput. Aided Mol. Des.* **2011**, *25*, 855–871. <https://doi.org/10.1007/s10822-011-9468-3>.
- Zhu, X.; Kruhlak, N.L. Construction and analysis of a human hepatotoxicity database suitable for QSAR modeling using post-market safety data. *Toxicology* **2014**, *321*, 62–72. <https://doi.org/10.1016/j.tox.2014.03.009>.
- Huang, S.-H.; Tung, C.-W.; Fülöp, F.; Li, J.-H. Developing a QSAR model for hepatotoxicity screening of the active compounds in traditional Chinese medicines. *Food Chem. Toxicol.* **2015**, *78*, 71–77. <https://doi.org/10.1016/j.fct.2015.01.020>.
- Kim, M.T.; Huang, R.; Sedykh, A.; Wang, W.; Xia, M.; Zhu, H. Mechanism profiling of hepatotoxicity caused by oxidative stress using antioxidant response element reporter gene assay models and big data. *Env. Environ. Health Perspect.* **2016**, *124*, 634–641. <https://doi.org/10.1289/ehp.1509763>.
- He, S.; Ye, T.; Wang, R.; Zhang, C.; Zhang, X.; Sun, G.; Sun, X. An in silico model for predicting drug-induced hepatotoxicity. *Int. J. Mol. Sci.* **2019**, *20*, 1897. <https://doi.org/10.3390/ijms20081897>.
- Ancuceanu, R.; Hovanet, M.V.; Anghel, A.I.; Furtunescu, F.; Neagu, M.; Constantin, C.; Dinu, M. Computational models using multiple machine learning algorithms for predicting drug hepatotoxicity with the dilrank dataset. *Int. J. Mol. Sci.* **2020**, *21*, 2114. <https://doi.org/10.3390/ijms21062114>.
- Bajželj, B.; Drgan, V. Hepatotoxicity modeling using counter-propagation artificial neural networks: Handling an imbalanced classification problem. *Molecules* **2020**, *25*, 481. <https://doi.org/10.3390/molecules25030481>.
- Box, G.E.P. Science and statistics. *J. Am. Stat. Assoc.* **1976**, *71*, 791–799. <https://doi.org/10.1080/01621459.1976.10480949>.
- Curchoe, C.L. All Models Are Wrong, but Some Are Useful. *J. Assist. Reprod. Genet.* **2020**, *37*, 2389–2391. <https://doi.org/10.1007/s10815-020-01895-3>.
- Toropov, A.A.; Toropova, A.P.; Raitano, G.; Benfenati, E. CORAL: Building up QSAR models for the chromosome aberration test. *Saudi J. Biol. Sci.* **2019**, *26*, 1101–1106. <https://doi.org/10.1016/j.sjbs.2018.05.013>.
- Toropova, A.P.; Toropov, A.A.; Benfenati, E. Semi-correlations as a tool to model for skin sensitization. *Food Chem. Toxicol.* **2021**, *157*, 112580. <https://doi.org/10.1016/j.fct.2021.112580>.
- Toropov, A.A.; Toropova, A.P.; Roncaglion, A.; Benfenati, E. The system of self-consistent semi-correlations as one of the tools of cheminformatics for designing antiviral drugs. *New J. Chem.* **2021**, *45*, 20713–20720. <https://doi.org/10.1039/d1nj03394h>.
- Toropov, A.A.; Toropova, A.P. The index of ideality of correlation: A criterion of predictive potential of QSPR/QSAR models? *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* **2017**, *819*, 31–37. <https://doi.org/10.1016/j.mrgentox.2017.05.008>.
- Toropova, A.P.; Toropov, A.A.; Roncaglion, A.; Benfenati, E. The system of self-consistent models for vapour pressure. *Chem. Phys. Lett.* **2022**, *790*, 139354. <https://doi.org/10.1016/j.cplett.2022.139354>.
- Toropov, A.A.; Toropova, A.P.; Rasulev, B.F.; Benfenati, E.; Gini, G.; Leszczynska, D.; Leszczynski, J. CORAL: Binary classifications (active/inactive) for liver-related adverse effects of drugs. *Curr. Drug. Saf.* **2012**, *7*, 257–261. <https://doi.org/10.2174/157488612804096542>.
- Toropova, A.P.; Toropov, A.A. CORAL: Binary classifications (active/inactive) for drug-induced liver injury. *Toxicol. Lett.* **2017**, *268*, 51–57. <https://doi.org/10.1016/j.toxlet.2017.01.011>.
- Toropov, A.A.; Toropova, A.P.; Benfenati, E.; Dorne, J.L. SAR for gastro-intestinal absorption and blood-brain barrier permeation of pesticides. *Chem-Biol. Interact.* **2018**, *290*, 1–5. <https://doi.org/10.1016/j.cbi.2018.04.030>.
- Toropova, A.P.; Toropov, A.A.; Benfenati, E.; Leszczynska, D.; Leszczynski, J. Prediction of antimicrobial activity of large pool of peptides using quasi-SMILES. *BioSystems* **2018**, *169–170*, 5–12. <https://doi.org/10.1016/j.biosystems.2018.05.003>.
- Toropova, A.P.; Toropov, A.A.; Veselinović, A.M.; Veselinović, J.B.; Leszczynska, D.; Leszczynski, J. Semi-correlations combined with the index of ideality of correlation: A tool to build up model of mutagenic potential. *Mol. Cell. Biochem.* **2019**, *452*, 133–140. <https://doi.org/10.1007/s11010-018-3419-4>.
- Toropova, A.P.; Toropov, A.A.; Carnesecchi, E.; Benfenati, E.; Dorne, J.L. The index of ideality of correlation: Models for flammability of binary liquid mixtures. *Chem. Pap.* **2020**, *74*, 601–609. <https://doi.org/10.1007/s11696-019-00903-w>.
- Carnesecchi, E.; Toropov, A.A.; Toropova, A.P.; Kramer, N.; Svendsen, C.; Dorne, J.L.; Benfenati, E. Predicting acute contact toxicity of organic binary mixtures in honey bees (*A. Mellifera*) through innovative QSAR models. *Sci. Total. Environ.* **2020**, *704*, 135302. <https://doi.org/10.1016/j.scitotenv.2019.135302>.
- Majumdar, S.; Basak, S.C. Beware of naïve q2, use true q2: Some comments on QSAR model building and cross validation. *Curr. Comput. Aided Drug. Des.* **2018**, *14*, 5–6. <https://doi.org/10.2174/15734099140118032112006>.
- Tuppurainen, K.; Korhonen, S.-P.; Ruuskanen, J. Performance of multicomponent self-organizing regression (MCSOR) in QSAR, QSPR, and multivariate calibration: Comparison with partial least-squares (PLS) and validation with large external data sets. *SAR QSAR Environ. Res.* **2006**, *17*, 549–561. <https://doi.org/10.1080/10629360601033390>.
- Toropov, A.A.; Toropova, A.P.; Benfenati, E.; Fanelli, R. QSAR as a random event: Selecting of the molecular structure for potential anti-tuberculosis agents. *Anti-Infect. Agents* **2016**, *14*, 3–10. <https://doi.org/10.2174/221135251401160302120554>.
- Toropova, A.P.; Meneses, J.; Alfaro-Moreno, E.; Toropov, A.A. The system of self-consistent models based on quasi-SMILES as a tool to predict the potential of nano-inhibitors of human lung carcinoma cell line A549 for different experimental conditions. *Drug. Chem. Toxicol.* **2023**, *in press*. <https://doi.org/10.1080/01480545.2023.2174986>.

26. Toropov, A.A.; Toropova, A.P.; Kudyshkin, V.O. The system of self-consistent QSPR-models for refractive index of polymers. *Struct. Chem.* **2022**, *33*, 617–624. <https://doi.org/10.1007/s11224-021-01875-y>.
27. Toropov, A.A.; Toropova, A.P.; Achary, P.G.R.; Raškova, M.; Raška, I. The searching for agents for Alzheimer's disease treatment via the system of self-consistent models. *Toxicol. Mech. Methods* **2022**, *32*, 549–557. <https://doi.org/10.1080/15376516.2022.2053918>.
28. Toropova, A.P.; Toropov, A.A. The System of Self-Consistent of Models: A New Approach to Build Up and Validation of Predictive Models of the Octanol/Water Partition Coefficient for Gold Nanoparticles. *Int. J. Env. Environ. Res.* **2021**, *15*, 709–722. <https://doi.org/10.1007/s41742-021-00346-w>.
29. Toropov, A.A.; Toropova, A.P. The system of self-consistent models for the uptake of nanoparticles in PaCa2 cancer cells. *Nanotoxicology* **2021**, *15*, 995–1004. <https://doi.org/10.1080/17435390.2021.1951387>.
30. Toropova, A.P.; Toropov, A.A.; Roncaglioni, A.; Benfenati, E. The index of ideality of correlation improves the predictive potential of models of the antioxidant activity of tripeptides from frog skin (*Litoria rubella*). *Comput. Biol. Med.* **2021**, *133*, 104370. <https://doi.org/10.1016/j.combiomed.2021.104370>.
31. Ghiasi, T.; Ahmadi, S.; Ahmadi, E.; Talei Bavi Olyai, M.R.; Khodadadi, Z. The index of ideality of correlation: QSAR studies of hepatitis C virus NS3/4A protease inhibitors using SMILES descriptors. *SAR QSAR Environ. Res.* **2021**, *32*, 495–520. <https://doi.org/10.1080/1062936X.2021.1925344>.
32. Javidfar, M.; Ahmadi, S. QSAR modelling of larvicidal phytocompounds against *Aedes aegypti* using index of ideality of correlation. *SAR QSAR Environ. Res.* **2020**, *31*, 717–739. <https://doi.org/10.1080/1062936X.2020.1806922>.
33. Kumar, A.; Kumar, P. Prediction of power conversion efficiency of phenothiazine-based dye-sensitized solar cells using Monte Carlo method with index of ideality of correlation. *SAR QSAR Environ. Res.* **2021**, *32*, 817–834. <https://doi.org/10.1080/1062936X.2021.1973095>.
34. Kumar, A.; Kumar, P.; Singh, D. QSRR modelling for the investigation of gas chromatography retention indices of flavour and fragrance compounds on Carbowax 20 M glass capillary column with the index of ideality of correlation and the consensus modelling. *Chemom. Intell. Lab. Syst.* **2022**, *224*, 104552. <https://doi.org/10.1016/j.chemolab.2022.104552>.
35. Jafari, K.; Fatemi, M.H.; Toropova, A.P.; Toropov, A.A. The development of nano-QSPR models for viscosity of nanofluids using the index of ideality of correlation and the correlation intensity index. *Chemom. Intell. Lab. Syst.* **2022**, *222*, 104500. <https://doi.org/10.1016/j.chemolab.2022.104500>.
36. Toropova, A.P.; Toropov, A.A.; Viganò, E.L.; Colombo, E.; Roncaglioni, A.; Benfenati, E. Carcinogenicity prediction using the index of ideality of correlation. *SAR QSAR Environ. Res.* **2022**, *33*, 419–428. <https://doi.org/10.1080/1062936X.2022.2076736>.
37. Ahmadi, S.; Lotfi, S.; Kumar, P. Quantitative structure–toxicity relationship models for predication of toxicity of ionic liquids toward leukemia rat cell line IPC-81 based on index of ideality of correlation. *Toxicol. Mech. Methods* **2022**, *32*, 302–312. <https://doi.org/10.1080/15376516.2021.2000686>.
38. Duhan, M.; Sindhu, J.; Kumar, P.; Devi, M.; Singh, R.; Kumar, R.; Lal, S.; Kumar, A.; Kumar, S.; Hussain, K. Quantitative structure activity relationship studies of novel hydrazone derivatives as α -amylase inhibitors with index of ideality of correlation. *J. Biomol. Struct. Dyn.* **2022**, *40*, 4933–4953. <https://doi.org/10.1080/07391102.2020.1863861>.
39. Kumar, A.; Kumar, P. Cytotoxicity of quantum dots: Use of quasi SMILES in development of reliable models with index of ideality of correlation and the consensus modelling. *J. Hazard. Mater.* **2021**, *402*, 123777. <https://doi.org/10.1016/j.jhazmat.2020.123777>.
40. Kumar, P.; Kumar, A. Unswerving modeling of hepatotoxicity of cadmium containing quantum dots using amalgamation of quasi SMILES, index of ideality of correlation, and consensus modelling. *Nanotoxicology* **2021**, *15*, 1199–1214. <https://doi.org/10.1080/17435390.2021.2008039>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.