environments

MDPI

Article

QSAR Models for Non-Effective Levels of Chemical Substances Using the CORAL Software

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

Department of Environmental Health Sciences, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Via Mario Negri 2, 20156 Milan, Italy; andrey.toropov@marionegri.it (A.A.T.); emilio.benfenati@marionegri.it (E.B.)

* Correspondence: alla.toropova@marionegri.it; Tel.: +39-02-3901-4595

Abstract

A new predictive model is proposed for determining the non-effective levels (NELs) of chemical substances across seven different training and validation splits. The model utilizes optimal molecular descriptors calculated from SMILES (Simplified Molecular Input-Line Entry System) strings, with correlation weights optimized via the Monte Carlo method. To improve predictive accuracy, we implemented this randomized optimization process using a Las Vegas algorithm integrated with the Correlation Intensity Index (CII) as a novel predictive criterion. A comparative analysis shows that incorporating the CII significantly improves model performance: the average coefficient of determination (R^2) for the validation set increased from 0.57 ± 0.07 (without CII) to 0.69 ± 0.07 (with CII), underscoring the advantage of this approach, and a similar improvement occurs on the calibration set. This shows that the adopted methodology is useful, which allows us to develop more general models.

Keywords: QSAR; non-effective level; SMILES; CORAL software; Las Vegas algorithm

1. Introduction

Chemical substances can induce a wide array of adverse health effects, varying significantly based on exposure duration and the biological target. These range from acute toxicity, which manifests within hours to days, to organ-specific damage such as cardiotoxicity, nephrotoxicity, neurotoxicity, and pulmonotoxicity [1–3]. Furthermore, research often focuses on specific toxic pathways, particularly those involving mutagenic or carcinogenic mechanisms [4,5]. To ensure safety, it is critical to determine exposure thresholds that prevent adverse outcomes; consequently, repeated, non-periodic exposure regimens are commonly used to evaluate the absence of toxicity after subchronic exposure. To address this need, researchers at the US Food and Drug Administration (FDA) recently introduced the concept of Non-Effective Levels (NELs) [6]. The NEL is defined as the maximum threshold below which a chemical, drug, or pollutant exerts no discernible adverse impact on organisms, human health, or ecosystems.

In toxicology and pharmacology, "no-effect levels" typically refer to doses or concentrations of a substance that do not cause detectable, adverse, or biologically significant changes in the body. These levels are crucial for establishing safety thresholds for drugs, chemicals, and food components. Using these values, is it possible to establish admissible chemical exposure levels? The NEL concept culminates in extensive studies of several thousand chemical substances and their associated toxic mechanisms. Through this comprehensive

Check for updates

Received: 10 September 2026
Revised: 2 October 2026
Accepted: 9 October 2026
Published: 12 October 2026

Copyright: © 2026 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.

evaluation, the authors established six distinct substance categories distributed across multiple structural classes. This framework directly advances the traditional Toxic Threshold of Concern (TTC) approach, which originates from studies done decades ago [7–9]. While the conventional TTC scheme relies on only three classes derived from a historically limited pool of chemicals, it remains a standard screening tool in regulatory science. For example, the European Food Safety Authority (EFSA) applies it for food safety assessments, and cosmetic evaluations often use it with only two classes. Consequently, the NEL framework represents a significant evolution of legacy TTC schemes; it is anchored by a substantially larger database, incorporates more nuanced explanations of underlying toxicological endpoints, and offers a more granular classification system [10].

The developers of the NEL database also introduced chemical-structure-based rules to assign a substance to an NEL class. This is similar to the TTC approach, except that in the case of the NEL, the set of rules is much larger and more elaborate.

In this study, we do not address assigning NEL classes. We tried to predict the quantitative value, yielding a more sophisticated outcome suitable for application across different schemes and classification approaches, with thresholds defined according to different regulatory frameworks. Quantitative structure–activity relationships (QSARs) can be used to simulate the endpoint. In this study, we build NEL models using CORAL software (https://www.insilico.eu/coral, accessed on 10 October 2026). CORAL is quite a simple approach, since it does not require calculating molecular descriptors, as most QSAR models do.

Like recent QSPR/QSAR studies we are using SMILES (Simplified Molecular Input-Line Entry System) as a representation of the molecular structure [11–22]. The studies mentioned address toxicity [14,15,21] as well as other biochemical endpoints [17,18,20,22]. However, these studies are theoretical, not regulatory [23].

Importantly, like many biochemical endpoints, the NEL is highly sensitive to the physicochemical environment, where even minor environmental variations can significantly alter its value. Under these conditions, the NEL is best treated as a random variable exhibiting a substantial range of values. This inherent stochasticity devalues the predictive potential of overly rigid, self-correcting methods, compromising their accuracy for individual substances. Instead, it calls for a methodology applicable to broader chemical groups, which typically comprise many homologs. These factors make stochastic processes a highly attractive alternative to rigorous, deterministic forecasts generated by complex self-consistent models. Consequently, the proposed software provides a robust, flexible approach that captures intricate associations in experimental data while remaining fully verifiable across randomized training and validation splits.

2. Materials and Methods

2.1. Data

The numerical data for the NEL (expressed in mmol) were taken from [10] and comprise a total of 1.751 substances. The US FDA gathered and curated a much larger set of substances by collecting and organizing data from multiple sources, including the European Food Safety Authority (EFSA), the U.S. Environmental Protection Agency (US EPA), the European Chemicals Agency (ECHA), etc. The NEL encompasses "both no observed effect levels and no observed adverse effect levels". It has been obtained using a conservative approach. The values span 11 orders of magnitude and cover multiple adverse effects. Using this dataset, the US FDA assigned six toxicity classes; however, in our study, we preferred to model the quantitative value. The dataset was randomly partitioned into four distinct subsets: active training (~25%), passive training (~25%), calibration (~25%), and validation (~25%). Notably, while the active training, passive training, and calibration

sets collectively correspond to the traditional training set, their individual roles differ significantly due to the dynamic evolution of the CORAL algorithm.

Figure 1 shows the scheme that CORAL uses. The active training set provides correlation weights that link the optimal descriptor to the endpoint within that set; the passive training set monitors model quality for molecules not included in the active training set; and the calibration set acts as the model's "expert and consumer," optimizing the determination coefficient.

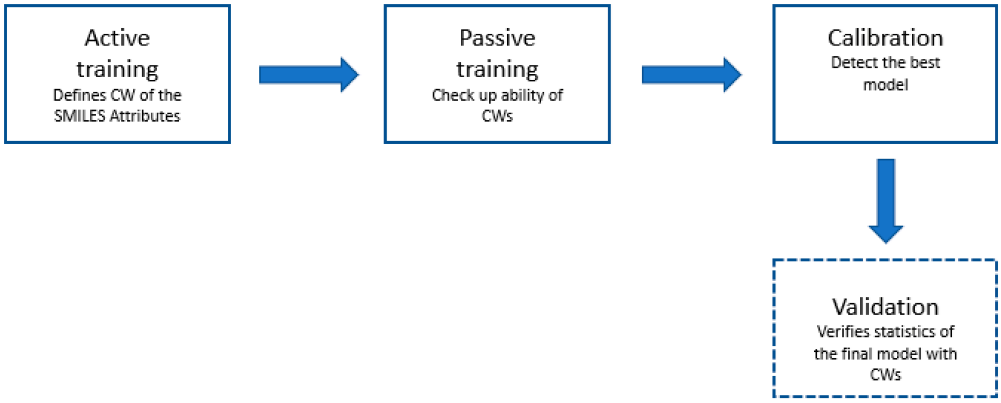


Figure 1. General scheme of model building.

Unlike classical QSAR approaches, which process all training compounds simultaneously, the CORAL model optimizes descriptors dynamically. In practical terms, the calibration set within the CORAL framework serves a function closest to that of a conventional QSAR training set: the main task of the system is to get the highest value of the determination coefficient for the calibration set, hoping that it will be accompanied by good statistics on the invisible validation set.

2.2. Model

The equation to calculate NEL values is the following

$$NEL = C_0 + C_1 \times DCW(T, N) \quad (1)$$

The calculation of the optimal descriptor integrates correlation weights of isolated SMILES atoms (S_k) and their connected pairs (SS_k):

$$DCW(T,N) = \sum CW(S_k) + \sum CW(SS_k) \quad (2)$$

The S_k is a SMILES atom, i.e., one symbol ('C', 'O', 'N') or a group of symbols in SMILES notation that cannot be considered separately ('Cl', 'Br'). The SS_k is a pair of described SMILES atoms that are neighbours in SMILES notation. CW(x) is a correlation weight (CW) for the SMILES attribute (S_k or SS_k). The CWs are calculated by Monte Carlo optimization to maximize a target function (described below). T and N are optimization parameters. T is a threshold that defines the minimum number of occurrences of a SMILES attribute in the active training set. If a SMILES attribute occurs in the active training set fewer than T times, it is considered "rare". The 'rare' SMILES attributes have correlation weights equal to zero; i.e., they are removed from the simulation process. N is the number of optimization epochs.

The Monte Carlo method performs the optimization of correlation weights and serves as the core stochastic engine for descriptor generation.

2.3. Monte Carlo Optimization

To compute the numerical values of the correlation weights required by Equation (2), a Monte Carlo optimization protocol was employed. The stochastic adaptation of these weights was driven by the following target functions:

$$TF_0 = r_A + r_P - |r_A - r_P| \times 0.1 \quad (3)$$

$$TF_1 = r_A + r_P - |r_A - r_P| \times 0.1 + CII \times 0.3 \quad (4)$$

The r_A and r_P are correlation coefficients between the observed and predicted endpoints for the active and passive training sets, respectively.

The correlation intensity index (CII) was developed as a tool to improve the quality of Monte Carlo optimization for building QSPR models [24].

The CII is calculated as follows

$$CII_C = 1 - \sum Protest_k \quad (5)$$

$$Protest_k = \begin{cases} R^2_k - R^2, & \text{if } R^2_k - R^2 > 0 \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

R^2 is the correlation coefficient for a set containing n substances. $R^2{}_k$ is the correlation coefficient for a set of $n - 1$ substances after removing the k-th substance. Hence, if $(R^2{}_k - R^2)$ is larger than zero, the k-th substance is an "oppositionist" to the correlation between experimental and predicted values for the set. A small sum of "protests" means a more "intensive" correlation.

Incorporating CII shifts the influence of correlation weights—originally focused on model formulation for training sets—toward those for the calibration set.

2.4. Las Vegas Algorithm

The Las Vegas algorithm and Monte Carlo methods are both derived from the idea of harnessing randomness to establish regularity. However, it should be noted that the Las Vegas algorithm and Monte Carlo methods differ significantly in their heuristics. Monte Carlo is a chaotic dynamic leading to an optimal combination of random data (correlation weights of molecular fragments encoded in SMILES and extracted for the aforementioned chaotic computer experiments), while the Las Vegas algorithm is a process of data summarization, favoring a subset of the data (the calibration set) and infringing on the interests of the active and passive training sets. The active and passive training sets are forced to serve the calibration set's statistical interests at the expense of their own statistical interests (the coefficient of determination and the standard deviation).

Following the partitioning of the initial dataset into active training, passive training, calibration, and validation subsets (Section 2.1), each Monte Carlo optimization run selects the split that yields optimal statistical performance on the calibration set. This selection assumes that a split that performs well on calibration data will show comparable predictive robustness on the validation set. Throughout the model-building pipeline, the validation set compounds remain completely external; they are strictly excluded from model development and optimization and serve exclusively as an independent test bed to evaluate predictive accuracy on unseen chemical structures. The target function TF_1 is controversial as an optimization problem. All information resources of the active (and passive) training sets must ensure good statistics on the calibration set. This is only possible by disregarding the statistical quality of the model in the training sets, because the apparently high statistical quality of the results achieved with the active and passive training sets may lead to overtraining; indeed, the model is not yet optimized using these sets.

Indeed, the scatter of points in the experimental versus calculated values for the training set is quite large, with both underestimation and overestimation. Considering Figure 1, we observe that the values for the substances on the training set are divided into pairs of clusters. Calculating the correlation coefficient for these two clusters simultaneously yields low values (≈0.2).

First, however, the green (underestimates) and red (overestimates) correlations are characterized by higher "local" coefficients of determination when considered separately (Figure 1).

Second, the regression line (i.e., the C0 and C1 values) fits both the training and calibration sets. Hopefully, the regression line will also fit the validation set.

Figure 2 graphically shows this situation.

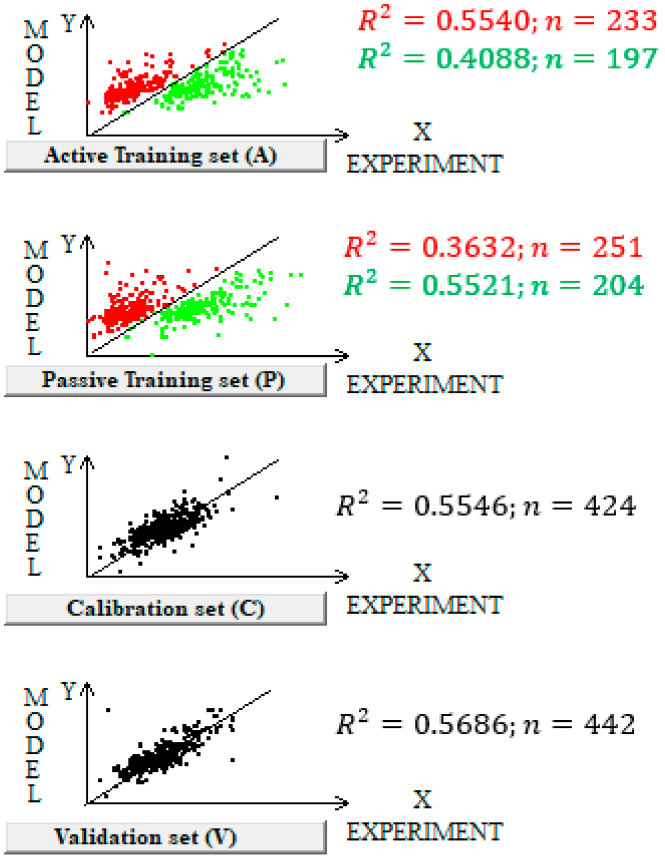


Figure 2. Predicted (y) versus experimental (x) values for the four subsets: active training, passive training, calibration, and validation. In the active and passive training sets, clusters of values are overestimated (red) or underestimated (green), whereas this is less pronounced in the calibration and validation sets.

2.5. Applicability Domain

The applicability domain (AD) of the developed model is established via Equation (7), which quantifies the statistical defects associated with individual SMILES attributes.

These defects are defined as follows:

$$d_k = \frac{|P(A_k) - P'(A_k)|}{N(A_k) + N'(A_k)} + \frac{|P(A_k) - P''(A_k)|}{N(A_k) + N''(A_k)} + \frac{|P'(A_k) - P''(A_k)|}{N'(A_k) + N''(A_k)} \quad (7)$$

where $P(A_k)$, $P'(A_k)$, and $P''(A_k)$ are the probabilities of A_k in the active training, passive training, and calibration sets, respectively; $N(A_k)$, $N'(A_k)$, and $N''(A_k)$ are the frequencies of A_k in the active training, passive training, and calibration sets, respectively.

The statistical SMILES-defects (D_j) are calculated as

$$D_j = \sum_{k=1}^{NA} d_k \quad (8)$$

where NA is the number of non-blocked SMILES attributes.

A SMILES falls within the applicability domain [25,26] if

$$Dj < 2 * \overline{D} \quad (9)$$

2.6. The Procedure for the Mechanistic Interpretation

The developed QSPR models allow for a mechanistic interpretation of the studied endpoints. By analyzing the numerical correlation weights of specific molecular features across multiple Monte Carlo optimization runs, these features can be classified into three distinct categories:

- Promoters of endpoint increase: Features that consistently maintain positive correlation weights across all optimization runs, indicating a stable contributing effect to the endpoint.
- Promoters of endpoint decrease: Features that consistently exhibit negative correlation weights across all runs, indicating a stable diminishing effect.
- Features with indefinite roles: Features that display fluctuating (both positive and negative) correlation weights across different runs, rendering their structural contribution to the endpoint ambiguous.

3. Results

3.1. Statistics

Table 1 summarizes the statistical characteristics of the developed models and Monte Carlo optimization (MCO) using the TF_0 and TF_1 target functions.

Table 1. Statistical quality of models obtained from different MCO splits using TF_0 and TF_1. Bold values are the highest obtained using TF_1. Using CII significantly improves the model's predictive potential; the average determination coefficient for the validation set with TF_0 is 0.573 ± 0.147, whereas with TF_1 it is 0.687 ± 0.075.

Target Function	Split		*n* *	R^2	CCC	CII	Q^2	RMSE	MAE	F	Na
TF_0	1	C	424	0.5186	0.7178	0.7679	0.5074	0.633	0.456	455	
		V	442	0.6780	-	-	-	0.49	0.40	-	157
	2	C	424	0.4465	0.6574	0.7570	0.4268	0.749	0.517	340	
		V	442	0.4688	-	-	-	0.69	0.51	-	154
	3	C	424	0.4676	0.6806	0.7572	0.4614	0.650	0.501	371	
		V	442	0.6688	-	-	-	0.54	0.42	-	147
	4	C	424	0.4873	0.6975	0.7550	0.4805	0.625	0.475	401	
		V	442	0.6154	-	-	-	0.55	0.44	-	153
	5	C	424	0.4822	0.6936	0.7558	0.4745	0.642	0.497	393	
		V	442	0.6874	-	-	-	0.48	0.40	-	151
	6	C	424	0.4702	0.6836	0.7600	0.4625	0.676	0.509	375	
		V	442	0.2531	-	-	-	1.15	0.88	-	150
	7	C	424	0.3437	0.5694	0.7187	0.2981	0.859	0.541	221	
		V	442	0.6361	-	-	-	0.54	0.44	-	153
TF_1	1	C	424	0.5546	0.7439	0.8231	0.5484	0.622	0.485	525	
		V	442	**0.5686**	-	-	-	0.62	0.46	-	137
	2	C	424	0.6297	0.7901	0.8046	0.6247	0.604	0.462	717	
		V	442	**0.6058**	-	-	-	0.61	0.44	-	155
	3	C	424	0.6122	0.7735	0.8102	0.6078	0.617	0.478	666	

Table 1. *Cont.*

Target Function	Split		*n* *	R^2	CCC	CII	Q^2	RMSE	MAE	F	Na
		V	442	**0.7234**	-	-	-	0.54	0.42	-	148
	4	C	424	0.6625	0.8120	0.8417	0.6582	0.574	0.449	828	
		V	442	**0.7629**	-	-	-	0.50	0.41	-	154
	5	C	424	0.6297	0.7827	0.8138	0.6256	0.603	0.468	718	
		V	442	**0.6958**	-	-	-	0.57	0.44	-	155
	6	C	424	0.6558	0.8007	0.8310	0.6522	0.563	0.452	804	
		V	442	**0.7890**	-	-	-	0.48	0.39	-	154
	7	C	424	0.5581	0.7387	0.7943	0.5521	0.649	0.507	533	
		V	442	**0.6607**	-	-	-	0.58	0.44	-	160

* C = calibration set; V = validation set; R^2 = determination coefficient; CCC = concordance correlation coefficient; CII = correlation intensity index; Q^2 = cross-validated R^2; RMSE = root mean squared error; F = Fischer F-ratio; Na = the number of non-blocked SMILES attributes.

One can see that the target function TF_1 yields better models than TF_0. This holds across all statistical parameters. For instance, for R^2, the range on the validation set is 0.47–0.69 for TF_0, while it is 0.57–0.79 for TF_1.

3.2. Applicability Domain (AD)

Outlier identification is based on substances containing molecular moieties that are rare, i.e., uncommon in the training set. For these substances, predictions are more uncertain.

Table 2 compares determination coefficients for validation sets for MCO with TF_1. Removing compounds with statistical defects twice the average improves model quality on the validation set.

Table 2. Determination coefficients for the validation set overall and for AD. The number of removed compounds is indicated in brackets.

	1	2	3	4	5	6	7
All validation sets	0.57	0.61	0.72	0.76	0.70	0.79	0.66
AD only	0.65 (−57)	0.73 (−54)	0.75 (−54)	0.76 (−52)	0.75 (−56)	0.79 (−51)	0.74 (−48)

3.3. The Results of the Mechanistic Interpretation Process

Table 3 presents a collection of SMILES attributes that promote increases or decreases in NEL. One can see that promoters of increased NEL are carbon (sp^3) indicated by 'C', branching indicated by brackets, pairs of oxygen and carbon ('O. . .C'), sulfur ('S'), as well as pairs of sulfur and carbon ('S. . .C'). Promoters for decreasing NEL are the presence of several cycles (digits), nitrogen ('N. . .'), and double bonds ('='). The prevalence of SMILES attributes in the active and passive sets, as well as in the calibration set, should also be considered.

Table 3. Lists SMILES attributes that increase or decrease NEL, as observed across three runs of the Monte Carlo optimization. Columns Run1, Run2, and Run3 contain numerical values of their correlation weights.

Category	S_k or SS_k	Run 1	Run 2	Run 3	NA *	NP	NC	d_k (Equation (7))
NEL increase	C.....	0.1998	0.2590	0.2148	425	450	420	0.0000
	C...=.....	0.2131	0.1248	0.3521	306	306	288	0.0001

Table 3. *Cont.*

Category	S_k or SS_k	Run 1	Run 2	Run 3	NA *	NP	NC	d_k (Equation (7))
	O...C.....	0.4538	0.1110	0.1647	217	235	191	0.0002
	Cl.....	1.8293	1.3319	0.5304	100	92	90	0.0002
	S.....	1.8616	1.1966	1.1820	98	93	113	0.0004
	S...C.....	0.5476	0.4453	0.6799	43	40	50	0.0005
NEL decrease	=.....	−0.3347	−0.3590	−0.4434	383	403	370	0.0000
	C...C.....	−0.3812	−0.3590	−0.3817	374	394	364	0.0000
	O.....	−0.9709	−0.4125	−0.4407	359	391	337	0.0001
	2.....	−0.1195	−0.5689	−0.3178	196	196	169	0.0002
	=...2.....	−1.3928	−0.5604	−1.1038	144	119	115	0.0004
	N...=.....	−0.5402	−0.3909	−0.7347	110	89	106	0.0004
	=...4.....	−3.6138	−1.9311	−3.0550	32	37	29	0.0003
	C...5.....	−0.4767	−0.0802	−1.3833	21	36	20	0.0008

* NA, NP, and NC are the numbers of compounds in the active training set, passive training set, and calibration set, respectively.

4. Discussion

The Toxic Threshold of Concern (TTC) approach is widely adopted in food and cosmetics regulations, but it originates from studies conducted decades ago [7,8]. The TTC approach uses a set of substances and derives three classes of substances, with different levels of expected adverse effects. The experimental values underlying the TTC are obsolete, and more data are now available under more recent protocols. These considerations prompted the US FDA to develop a new approach, called the expanded decision tree (EDT) [10]. Modeling this endpoint is uniquely challenging due to a multitude of compounding factors, including the necessity of evaluating diverse routes of exposure (oral, dermal, and inhalation), accounting for variable frequencies and durations, and addressing demographic heterogeneity across exposed populations (such as workers, consumers, children, and pregnant women). Consequently, quantitatively characterizing the NEL is highly valuable yet inherently complex.

To build the model for NEL, we implemented the Correlation Intensity Index (CII). Unlike the standard correlation coefficient, which measures data consistency, the CII primarily captures data diversity from a model-construction perspective. The good performance of the CII over the traditional coefficient of determination, R^2, stems from these fundamental semantic differences. Furthermore, as detailed in Section 2.4 and illustrated in Figure 1, the CII directly affects the regression line. While the positive impact of the CII is clear as a substantial improvement in the statistical quality of the calibration set, the precise mechanism driving this optimization remains a compelling phenomenon that warrants further theoretical investigation.

The CORAL model uses a procedure different from most other QSAR models. The classical procedure splits the initial set of available substances into two: the training set, used to build the model, and the validation set, used to verify whether the model is predictive when applied to new substances. The CORAL model further splits the substances in the training set, as explained in Section 2.1. The modelling procedure follows a sequential scheme, which progressively optimizes the parameters of the model. The calibration-set results indicate the final model, while the statistical values on the active and passive training sets help develop the models and select features but should not be considered representative of the final model. Figure 2 helps illustrate the model's evolution across phases. The first steps help identify the SMILES attributes useful for prediction. However, the model may focus on substances relevant to the active and passive training sets but that do not represent good examples for a more general model. In its initial phase,

the model learns from some substances but cannot yet generalize the main features. Thus, for some substances there are strong deviations between the predicted and experimental values. This is clear for the substances represented by green and red dots in Figure 2. For some substances, the values are overestimated; for others, they are underestimated. CORAL can select "good" and "bad" representatives of the model when it uses substances from the calibration set. Good substances help generate a predictive model that also works for substances in the calibration set; thus, they are valid for developing a general model. Conversely, bad substances generate a model that is too local, replicating certain substances without the possibility of providing valid results when new chemicals are used, such as those in the calibration set.

These tasks are those performed by the chosen objective function (the CII) and the partitioning using the Las Vegas algorithm (one that demonstrates a statistical situation "favorable" for the calibration set). Following this selection procedure, the resulting model is not entirely random; it represents a particular or special statistical case (event).

From a statistical point of view, the process aims to determine how rare or widespread the approach is. Analyzing the datasets in terms of these "good" and "bad" compounds can provide additional epistemological leverage to improve knowledge of the structure-activity relationship.

Knowledge of lists of "good" and "bad" compounds can refine the predictive potential of models and, potentially, rationalize model construction through such discrimination. Furthermore, this knowledge can point to the underlying drivers (mechanisms) of substance effects in relation to the endpoints studied in general and the endpoint under consideration (NEL) in particular. It should be noted that the SMILES lists of attributes contained (prevailing) in the structures of "bad" and "good" compounds are not associated with a mechanistic interpretation as an immediate interpretation; indeed, it is the combination of multiple features that provokes the final effect, and it would be wrong to get to the conclusion that a single, small series of atoms is sufficient to predict the effects of the whole molecule. Anyhow, these features may prove useful in the aforementioned epistemological aspect as sources of knowledge about endpoints, from the point of view of their behavior as participants in the stochastic processes used to develop QSAR models.

The statistical values in Table 1 show variability and uncertainty. The calibration set values for the determination coefficient range from 0.55 to 0.66. This indicates that the model is predictive, but not perfect, and results may change depending on the substances considered. The validation set shows even greater variation, with values between 0.57 and 0.79. These data clearly indicate that the model works well even when tested on new substances. However, because the range is so large, more substances are needed to provide more accurate estimates.

A comparison of the models in Table 4 shows that, although this study considers one order of magnitude more organic substances than other works, the determination coefficient for the validation set is comparable to the corresponding values for other models. An advantage of our model is that, given the very limited number of substances in the models in [6,25–27], we expect our model to have broader applicability. Table 4 presents the statistical characteristics of the model obtained for the split #6, as this model possesses the greatest predictive potential (the highest coefficient of determination for the independent validation set).

Table 4. Comparison of models for non-effective levels. Because the approach under consideration is based on a structured training set (divided into active and passive learning, along with a calibration set), we also present the data for the red and green clusters here too, whose meaning is shown in Figure 2.

		Training Set		Validation Set	
Model	**Sub-Set**	***n***	**R^2**	***n***	**R^2**
As in [6]		140	0.51	140	0.55
As in [25]		51	0.88	10	0.88
As in [7]		134	0.82	44	Q^2_{F1} = 0.83
This study	Total	1309	0.23	391 *	0.79
Split #6	Red	225	0.56		
	Green	205	0.50		
	Red	253	0.55		
	Green	202	0.56		
	Calibration	424	0.66		

* Split #6 applicability domain (391) for validation set.

It should be noted that the question of the origin of the CII is quite complex. The value of the correlation ideality index, which is a logical precursor to the CII, is more or less clear [26]. The correlation ideality index combines the correlation coefficient and the standard deviation into a single measure. It is difficult to identify any logical justification for the CII. What is clear is that the CII works: it forces models for the studied endpoint to deliver acceptable statistical quality on external validation sets, though unfortunately at the expense of quality on training sets. Finally, CII has been successfully applied to develop models for other endpoints [28–34]. Nevertheless, the index of ideality correlation [28] and the CII generally compete as potential improvers of model statistical quality when the primary focus is external validation of modeled endpoints.

The considered modeling system can be used for rapid qualitative (non-quantitative) assessment of the NEL value for substances that have not undergone experimental study.

The application of our model is useful for regulations that apply TTC, i.e., for the food and cosmetic sectors. The TTC approach is typically used for impurities, and in the absence of other, more robust, toxicological studies. The much larger and more robust database used by the US FDA may ideally extend the applications to more situations in the food and cosmetics sectors. In addition, our study can be used within other regulatory frameworks, covering different sectors such as pesticides, industrial chemicals, biocides, etc. Because the NEL database covers multiple endpoints, including chronic toxicity, our model results can be used for other purposes, such as prioritizing environmental contaminants.

5. Conclusions

The CORAL approach demonstrated significant advantages for modelling non-effective levels (NELs), an endpoint inherently characterized by a study-dependent and potentially stochastic nature due to variations in experimental protocols. The approach under consideration fully complies with traditional QSPR/QSAR requirements: a measure of model quality (the coefficient of determination for the validation set), a defined applicability domain, and mechanistic interpretation. This approach also requires additional confirmation of performance across different partitions of the training and validation sets. Considering splits of the training and validation sets undoubtedly strengthens the reliability and plausibility of the results (i.e., the statistical quality of the model obtained using the method on the external validation set). Furthermore, certain unusual characteristics can be identified in the numerical prediction results and classified as “good” or “bad” compounds. “Good” structures are compounds which are more often included in the calibration and

validation sets (indicating high predictability and stability despite experimental endpoint uncertainty). "Bad" compounds are those which are more often included in the active and passive training sets. Mechanistically, the core novelty of this work lies in its optimization paradigm: rather than treating molecular features extracted from SMILES strings as the sole adjustable variables, the algorithm simultaneously optimizes the configuration of the compound lists across the training, calibration, and validation subsets. This optimization balances structural descriptors and data distribution, providing robust, reproducible QSAR models even for complex, study-dependent biological endpoints like NELs.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/environments13100554/s1: Table S1: Contains technical details on the model for split 1.

Author Contributions: A.A.T.: conceptualization, methodology, resources; writing—original draft preparation; visualization; formal analysis; A.P.T.: conceptualization, methodology, resources; writing—original draft preparation; visualization; formal analysis; writing—review and editing; E.B.: project administration; writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: We acknowledge EFSA for the financial contribution within the project sOFT-ERA, OC/EFSA/IDATA/2022/02.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Materials. Further inquiries can be directed to the corresponding author.

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

Abbreviations

The following abbreviations are used in this manuscript:

SMILES	Simplified molecular input line entry system
DCW	Descriptor of correlation weights
CII	Correlation intensity index
NEL	Non-effective level

References

1. Pai, P.; D'Mello, R.S.; Nayak, S.; Rao, P.; Oruganti, S.; Satyamoorthy, K.; Sundara, B.K. In vivo acute oral toxicity assessment of novel histone deacetylase 2 inhibitor. *BMC Pharmacol. Toxicol.* **2026**, *27*, 3. [CrossRef] [PubMed]
2. Viganò, E.L.; Iwan, M.; Colombo, E.; Ballabio, D.; Roncaglioni, A. Mixture of experts for multitask learning in cardiotoxicity assessment. *J. Cheminform.* **2025**, *17*, 135. [CrossRef] [PubMed]
3. Bassan, A.; Alves, V.M.; Amberg, A.; Anger, L.T.; Beilke, L.; Bender, A.; Bernal, A.; Cronin, M.T.D.; Hsieh, J.-H.; Johnson, C.; et al. In silico approaches in organ toxicity hazard assessment: Current status and future needs for predicting heart, kidney and lung toxicities. *Comput. Toxicol.* **2021**, *20*, 100188. [CrossRef] [PubMed]
4. Basant, N.; Gupta, S. QSAR modeling for predicting mutagenic toxicity of diverse chemicals for regulatory purposes. *Environ. Sci. Pollut. Res.* **2017**, *24*, 14430–14444. [CrossRef] [PubMed]
5. Adiga Perdur, G.; AJ, Z.; Krishnappa, M.; Jurowski, K.; Ahuja, V. In silico approaches using CASE Ultra and QSAR Toolbox for predicting genotoxicity and carcinogenicity on diverse groups of chemicals. *Comput. Toxicol.* **2025**, *36*, 100380. [CrossRef]
6. Selvestrel, G.; Lavado, G.J.; Toropova, A.P.; Toropov, A.A.; Gadaleta, D.; Marzo, M.; Baderna, D.; Benfenati, E. Monte Carlo models for sub-chronic repeated-dose toxicity: Systemic and organ-specific toxicity. *Int. J. Mol. Sci.* **2022**, *23*, 6615. [CrossRef] [PubMed]
7. Cramer, G.M.; Ford, R.A.; Hall, R.L. Estimation of toxic hazard—A decision tree approach. *Food Cosmet. Toxicol.* **1978**, *16*, 255–276. [CrossRef] [PubMed]

8. Ghosh, S.; Roy, K. Quantitative read-across structure-activity relationship (q-RASAR): A novel approach to estimate the sub-chronic oral safety (NOAEL) of diverse organic chemicals in rats. *Toxicology* **2024**, *505*, 153824. [CrossRef] [PubMed]
9. Munro, I.C.; Ford, R.A.; Kennepohl, E.; Sprenger, J.G. Correlation of structural class with no-observed-effect levels: A proposal for establishing a threshold of concern. *Food Chem. Toxicol.* **1996**, *34*, 829–867. [CrossRef] [PubMed]
10. Stice, S.; Adams, T. The expanded decision tree: Predicting chronic toxic potential and safe intake levels. *Food Chem. Toxicol.* **2026**, *212*, 115903. [CrossRef] [PubMed]
11. Askarzade, A.; Ahmadi, S.; Almasirad, A. SMILES-based QSAR and molecular docking studies of chalcone analogues as potential anti-colon cancer. *Sci. Rep.* **2025**, *15*, 6573. [CrossRef] [PubMed]
12. Akrom, M.; Al Azies, H.; Herowati, W.; Sutojo, T.; Rustad, S.; Dipojono, H.K.; Kasai, H. SMILES-driven machine learning for high-throughput investigation of anti-corrosion materials. *Chemom. Intell. Lab. Syst.* **2025**, *263*, 105441. [CrossRef]
13. He, Y.; Liu, F.; Min, W.; Liu, G.; Wu, Y.; Wang, Y.; Yan, X.; Yan, B. De novo design of biocompatible nanomaterials using quasi-SMILES and recurrent neural networks. *ACS Appl. Mater. Interfaces* **2024**, *16*, 66367–66376. [CrossRef] [PubMed]
14. Salarzaei, S.; Shiri, F.; Ahmadi, S. A global quasi-SMILES model based on the Monte Carlo algorithm for assessing the multi-organism aquatic ecotoxicity of personal care products. *Ecotoxicol. Environ. Saf.* **2026**, *312*, 119948. [CrossRef] [PubMed]
15. Ouabane, M.; Zaki, K.; Tabti, K.; Alaqarbeh, M.; Sbai, A.; Sekkate, C.; Bouachrine, M.; Lakhlifi, T. Molecular toxicity of nitrobenzene derivatives to Tetrahymena pyriformis based on SMILES descriptors using Monte Carlo, docking, and MD simulations. *Comput. Biol. Med.* **2024**, *169*, 107880. [CrossRef] [PubMed]
16. Chauhan, K.; Ranolia, A.; Kiran; Kumar, P.; Singh, D.; Sindhu, J. Monte Carlo-based SMILES-driven QSPR modeling of Organic Chromophores across the visible-to- NIR spectral range. *J. Mol. Graph. Model.* **2026**, *148*, 109533. [CrossRef] [PubMed]
17. Chopdar, K.S.; Dash, G.C.; Mohapatra, P.K.; Nayak, B.; Raval, M.K. Monte-Carlo method-based QSAR model to discover phytochemical urease inhibitors using SMILES and GRAPH descriptors. *J. Biomol. Struct. Dyn.* **2022**, *40*, 5090–5099. [CrossRef] [PubMed]
18. Gupta, S.; Kashyap, M.; Bansal, Y.; Bansal, G. In silico insights into design of novel VEGFR-2 inhibitors: SMILES-based QSAR modelling, and docking studies on substituted benzo-fused heteronuclear derivatives. *SAR QSAR Environ. Res.* **2024**, *35*, 265–284. [CrossRef] [PubMed]
19. Latifi, N.; Ahmadi, S.; Lotfi, S.; Kolahi, S.A. SMILES-based QSAR analysis of carbamate derivatives targeting butyrylcholinesterase. *Sci. Rep.* **2026**, *16*, 3910. [CrossRef] [PubMed]
20. Lotfi, S.; Nooraei, F.; Ahmadi, S. QSAR-based evaluation of plant-derived larvicidal agents against Zika vector Aedes aegypti via Monte Carlo optimization and SMILES-based descriptors and molecular docking. *SAR QSAR Environ. Res.* **2026**, *37*, 423–447. [CrossRef] [PubMed]
21. Rezaie-keikhaie, N.; Shiri, F.; Ahmadi, S.; Salahinejad, M. QSTR based on Monte Carlo approach using SMILES and graph features for toxicity toward Tetrahymena pyriformis. *J. Iran. Chem. Soc.* **2023**, *20*, 2609–2620. [CrossRef]
22. Kumar, A.; Bagri, K.; Nimbhal, M.; Kumar, P. In silico exploration of the fingerprints triggering modulation of glutaminyl cyclase inhibition for the treatment of Alzheimer's disease using SMILES based attributes in Monte Carlo optimization. *J. Biomol. Struct. Dyn.* **2021**, *39*, 7181–7193. [CrossRef] [PubMed]
23. Kreider, M.; Williams, E. Interpreting REACH guidance in the determination of the derived no effect level (DNEL). *Regul. Toxicol. Pharmacol.* **2010**, *58*, 323–329. [CrossRef] [PubMed]
24. Toropov, A.A.; Toropova, A.P. Correlation intensity index: Building up models for mutagenicity of silver nanoparticles. *Sci. Total Environ.* **2020**, *737*, 139720. [CrossRef] [PubMed]
25. Toropov, A.A.; Toropova, A.P. QSAR as a random event: Criteria of predictive potential for a chance model. *Struct. Chem.* **2019**, *30*, 1677–1683. [CrossRef]
26. Toropov, A.A.; Toropova, A.P.; Roncaglioni, A.; Benfenati, E.; Leszczynska, D.; Leszczynski, J. The system of self-consistent models: The case of Henry's law constants. *Molecules* **2023**, *28*, 7231. [CrossRef] [PubMed]
27. Dobchev, D.A.; Tulp, I.; Karelson, G.; Tamm, T.; Tämma, K.; Karelson, M. Subchronic oral and inhalation toxicities: A challenging attempt for modeling and prediction. *Mol. Inform.* **2013**, *32*, 793–801. [CrossRef] [PubMed]
28. Toropova, A.P.; Toropov, A.A.; Mescieri, S.; Roncaglioni, A.; Benfenati, E. Simulation of the impact of pesticides on pollinators under different conditions using correlation weighting of quasi-SMILES components together with the Index of Ideality of Correlation (IIC). *J. Xenobiotics* **2026**, *16*, 10. [CrossRef] [PubMed]
29. Kumar, P.; Kumar, A. CORAL: Predictions of quality of rice based on retention index using a combination of Correlation Intensity Index and consensus modelling. In *QSPR/QSAR Analysis Using SMILES and Quasi-SMILES*; Toropova, A.P., Toropov, A.A., Eds.; Challenges and Advances in Computational Chemistry and Physics; Springer: Cham, Switzerland, 2023; Volume 33, pp. 421–462. [CrossRef]
30. Bhawna; Kumar, A.; Singh, V.; Kumar, P.; Kumar, S. System of self-consistent models synergistic or antagonistic with correlation intensity index: A combined perusal using MAO-A inhibitors. *SAR QSAR Environ. Res.* **2025**, *36*, 1209–1231. [CrossRef] [PubMed]

31. Singh, R.; Kumar, P.; Sindhu, J.; Kumar, A.; Lal, S. CORAL: Probing the structural requirements for α-amylase inhibition activity of 5-(3-arylallylidene)-2-(arylimino)thiazolidin-4-one derivatives based on QSAR with correlation intensity index, molecular docking, molecular dynamics, and ADMET studies. *J. Biomol. Struct. Dyn.* **2024**, *42*, 11861–11878. [CrossRef] [PubMed]
32. Kumar, P.; Kumar, A.; Sindhu, J.; Lal, S. Quasi-SMILES as a basis for the development of QSPR models to predict the CO2 capture capacity of deep eutectic solvents using correlation intensity index and consensus modelling. *Fuel* **2023**, *345*, 128237. [CrossRef]
33. Bhawna; Kumar, S.; Kumar, P.; Kumar, A. Correlation intensity index-index of ideality of correlation: A hyphenated target function for furtherance of MAO-B inhibitory activity assessment. *Comput. Biol. Chem.* **2024**, *108*, 107975. [CrossRef] [PubMed]
34. Toropova, A.P.; Toropov, A.A.; Roncaglioni, A.; Benfenati, E. Using the Correlation Intensity Index to build a model of cardiotoxicity of piperidine derivatives. *Molecules* **2023**, *28*, 6587. [CrossRef] [PubMed]

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.