

Article

# QSAR Models for Human Carcinogenicity: An Assessment Based on Oral and Inhalation Slope Factors

Cosimo Toma <sup>1,2</sup> , Alberto Manganaro <sup>3</sup>, Giuseppa Raitano <sup>1</sup>, Marco Marzo <sup>1</sup>, Domenico Gadaleta <sup>1</sup>, Diego Baderna <sup>1,\*</sup> , Alessandra Roncaglioni <sup>1</sup>, Nynke Kramer <sup>2</sup> and Emilio Benfenati <sup>1,\*</sup>

<sup>1</sup> Laboratory of Chemistry and Environmental Toxicology, Department of Environmental Health Sciences, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Via Mario Negri 2, 20156 Milan, Italy; cosimo.toma@marionegri.it (C.T.); giuseppa.raitano@marionegri.it (G.R.); marco.marzo@marionegri.it (M.M.); domenico.gadaeta@marionegri.it (D.G.); alessandra.roncaglioni@marionegri.it (A.R.)

<sup>2</sup> Institute for Risk Assessment Sciences (IRAS), Utrecht University, P.O. Box 80177, 3508 TD Utrecht, The Netherlands; N.I.Kramer@uu.nl

<sup>3</sup> Kode Chemoinformatics s.r.l., 56125 Pisa, Italy; a.manganaro@kode-solutions.net

\* Correspondence: diego.baderna@marionegri.it (D.B.); emilio.benfenati@marionegri.it (E.B.)

**Abstract:** Carcinogenicity is a crucial endpoint for the safety assessment of chemicals and products. During the last few decades, the development of quantitative structure–activity relationship ((Q)SAR) models has gained importance for regulatory use, in combination with in vitro testing or expert-based reasoning. Several classification models can now predict both human and rat carcinogenicity, but there are few models to quantitatively assess carcinogenicity in humans. To our knowledge, slope factor (SF), a parameter describing carcinogenicity potential used especially for human risk assessment of contaminated sites, has never been modeled for both inhalation and oral exposures. In this study, we developed classification and regression models for inhalation and oral SFs using data from the Risk Assessment Information System (RAIS) and different machine learning approaches. The models performed well in classification, with accuracies for the external set of 0.76 and 0.74 for oral and inhalation exposure, respectively, and  $r^2$  values of 0.57 and 0.65 in the regression models for oral and inhalation SFs in external validation. These models might therefore support regulators in (de)prioritizing substances for regulatory action and in weighing evidence in the context of chemical safety assessments. Moreover, these models are implemented on the VEGA platform and are now freely downloadable online.

**Keywords:** cancer slope factor; in silico method; QSAR; prioritization



**Citation:** Toma, C.; Manganaro, A.; Raitano, G.; Marzo, M.; Gadaleta, D.; Baderna, D.; Roncaglioni, A.; Kramer, N.; Benfenati, E. QSAR Models for Human Carcinogenicity: An Assessment Based on Oral and Inhalation Slope Factors. *Molecules* **2021**, *26*, 127. <https://doi.org/10.3390/molecules26010127>

Received: 2 December 2020

Accepted: 25 December 2020

Published: 29 December 2020

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2020 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

Every day, people are exposed to numerous environmental chemical stressors that can have adverse health effects during their life. Exposure to toxic chemicals or mixtures comes from the environment, living places and workplaces, but diet, drugs and lifestyle are important concurrent sources as well [1–4]. Adverse effects include chronic diseases and cancer. Nowadays, cancer is a major public health issue with more than 3 million new cases per year in the European Union [5,6].

Experimentally, the carcinogenic potential of a substance is evaluated by long-term in vivo carcinogenicity studies with laboratory animals. The conventional test for carcinogenicity is the two-year rodent carcinogenicity bioassay as described by the Organization for Economic Co-operation and Development (OECD) Test Guidelines 451 and 453 [7–9]. Although the procedure is expensive and time-consuming, animal models are still the most widely used method of investigation. In the last decade, the validity of the rodent bioassay was debated because of uncertainty related to extrapolating results to humans and ethical concerns about the numbers of animals needed [6]. Various non-animal methods have recently been proposed as alternative or complementary methods to assess carcinogenicity

with the aim of reducing animal testing, time and cost of the evaluation. These methods include *in vitro* bioassays (such as cell transformation assays and toxicogenomics) and *in silico* methods, such as (Q)SARs models and expert systems [10–20].

Most of the *in silico* models for carcinogenicity are classifier tools used to predict whether or not chemicals are carcinogens in animal models [15,21–24]. Only a few continuous models are used to quantitatively assess carcinogenicity, specifically, to predict the potency dose *in vivo* as the endpoint (TD50) [25–32]. Most of these models are already implemented in license-based or freely available software tools (Table S1).

To our knowledge, no model has yet been developed for oral and inhalation slope factors (SFs) used in human quantitative risk assessment (HRA) of environmental pollutants. The SF is the upper-bound estimate of the slope of the dose–response curve in the low-dose region for carcinogens and is used to assess the increase over a lifetime in incidence of cancers in humans from oral or inhalation exposure to a dose of a carcinogenic chemical [33–36]. In the HRA framework, the cancer risk for each chemical (CR or Incremental Lifetime Cancer Risk, ILCR) is calculated using the chronic daily intake (CDI, mg/kg-day) and the slope factor (SF, (mg/kg-day)<sup>−1</sup>); the SF provides the chemical-specific carcinogenic potency. With these two values, the cancer risk (CR, dimensionless) is obtained by multiplying them, as in Equation (1):

$$\text{CR} = \text{CDI} \times \text{SF} \quad (1)$$

Here we propose an integrated *in silico* approach for the qualitative and quantitative assessment of chemical carcinogenic potency, which includes classification and quantitative models for inhalation and oral human carcinogenicity based on slope factors.

## 2. Results

We developed both classification and regression models for carcinogenicity expressed as oral or inhalation slope factors. Data were collected from the Risk Assessment Information System (RAIS) Toxicity values database (<https://rais.ornl.gov>). For each exposure route (oral or inhalation), chemicals with a defined value for SF were considered carcinogenic, while compounds with no value were considered non-carcinogenic. This binary dataset was used to develop the classification models; however, the dataset including the SF values to describe carcinogenic potency was used for the regression models. Thus, in our proposed approach, the classifier models indicate if the substance is carcinogenic or not, and the regression model should be used in cascade to assess the substance's potency if it is labelled as carcinogenic.

### 2.1. Classification Models

Binary classification models were built by the Classification and Regression Tree (CART) modelling approach. CART models for inhalation and oral carcinogenicity performed well for sensitivity and specificity. The structure of inhalation and oral CART models are included in the Supplementary Materials as Figures S1 and S2. In order to increase access to the models, we implemented them within the platform VEGA (Virtual models for property Evaluation of chemicals within a Global Architecture, [www.vegahub.eu](http://www.vegahub.eu)), our online, freely available platform that contains a series of QSAR models for regulatory purposes. There are negligible differences between the original models in CART and those in VEGA. Balancing the dataset between models led, as expected, to lower sensitivity for the inhalation model than for the oral model (Table 1).

### 2.2. Regression Models

We developed the regression models using descriptor-based multi-layer perceptron–artificial neural networks (MLP–ANNs). In the proposed strategy to assess carcinogenicity, the regression model should be run when the classification model indicates carcinogenicity. For each regression model, the performance is reported as a determination coefficient ( $r^2$ ), root-mean-square error (RMSE), mean absolute error (MAE) and percentage of predicted

compounds (this percentage is called coverage of the model). We split the substances into two approaches. As detailed in the Material and Methods section, in one case (split A) we used a test set (TeS) to select the best model, and then an external validation set (ES) was used to evaluate the performance of new substances. In approach B, there was only a training and validation set, and the model selection was done using the 10-fold cross-validation method, using substances from the training set. Performances are reported in Tables 2 and 3 for the two splitting approaches, with values for the training set (TrS), TeS and ES.

**Table 1.** Statistics for the final oral and inhalation classification models (as implemented in VEGA).

Endpoint	Statistics	Training Set	Test Set
OSF	Accuracy	0.81	0.76
	Sensitivity	0.82	0.76
	Specificity	0.79	0.76
ISF	Accuracy	0.81	0.76
	Sensitivity	0.73	0.72
	Specificity	0.86	0.79

**Table 2.** Performance of OSF and ISF regression models derived with split scheme A.

Endpoint	Statistics	TrS	TeS	ES
OSF	$r^2$	0.709	0.708	0.569
	MAE	0.651	0.784	0.946
	RMSE	0.875	0.945	1.255
	coverage	100%	81%	88%
ISF	$r^2$	0.745	0.654	0.510
	MAE	0.639	0.875	0.933
	RMSE	0.876	1.035	1.176
	coverage	100%	85%	89%

**Table 3.** Performance of OSF and ISF regression models derived with split scheme B. Coverage is the percentage of compounds retained after applying the applicability domain (AD).

Endpoint	Statistics	TrS		ES
		Goodness-of-Fit	10-Fold Cross-Validation	
OSF	$r^2$	0.756	0.608	0.515
	MAE	0.646	0.819	1.012
	RMSE	0.814	1.031	1.284
	coverage	100%	100%	76%
ISF	$r^2$	0.745	0.591	0.647
	MAE	0.639	0.838	0.853
	RMSE	0.876	1.088	1.023
	coverage	n.a.	n.a.	81%

Split scheme A returned similar results for both OSF and ISF. The predictive power of both models was confirmed on both validation sets, with  $r^2$  values from 0.70 to 0.65 on the TeS, and from 0.57 to 0.51 on the ES. Coverage of the models on the two validation sets was always greater than 80%. Models derived from split scheme B showed similar results,

though performance was slightly better for the ISF model that returned an  $r^2$  of 0.65 on the ES, while the OSF model returned an  $r^2$  of 0.52. Moreover, the coverage on the TeS for the OSF was lower than 80%. Thus, split scheme A can be considered preferable for model OSF, while split scheme B gave better results for model ISF.

One disadvantage of DRAGON descriptors is that they cannot be used to develop a completely free and open source QSAR model software, even though they are widely used and robust. There are some examples of QSAR models retrained with DRAGON-like descriptors that give similar results [37,38]. With a view of implementing all of the developed models in the VEGA platform [39], we retrained them using the best scheme for each endpoint (scheme A for OSF and scheme B for ISF).

The model for OSF was replicated without modifications because VEGA already has the same descriptors as DRAGON models; however, the selected descriptors were not available as DRAGON-like descriptors for the inhalation models, so retraining the model led to different performances.

Table 4 shows the statistics of the models implemented in VEGA. The implementation of OSF with DRAGON gave higher  $r^2$  in training (TrS = 0.709, TeS = 0.708) than VEGA did (TrS + Te = 0.62), which gave a better performance in ES (0.569 vs. 0.839). However,  $r^2$  values for the TrS and ES of ISF (TrS = 0.745, ES = 0.647) were both certainly higher than with VEGA implementation (TrS = 0.586, ES = 0.566). Within the feature selection, we reduced the number of descriptors, using tools such as genetic algorithms, as described in the Materials and Methods section. One possible explanation is that a genetic algorithm using a larger number of descriptors in DRAGON had more starting combinations and a greater chance of selecting the best pool of descriptors.

**Table 4.** Performance of OSF and ISF regression models after VEGA implementation. Performances are reported for the training and test sets. Since the applicability domain is evaluated with the ADI index (53), coverage is not reported.

Endpoint	Statistics	Train <sup>1</sup> Full	Train without Outliers	Test Full	Test in AD
OSF	$r^2$	0.658	0.642	0.573	0.636
	MAE	0.713	0.725	0.951	0.881
	RMSE	0.96	0.969	1.28	1.131
	$r^2m$	0.653	0.64	0.489	0.58
	$r^2mavg$	0.655	0.633	0.501	0.595
	$r^2mDelta$	0.004	0.015	0.023	0.029
	k	1.017	1.02	0.985	1.028
	ki	0.647	0.63	0.57	0.619
	F	536.384	477.804	40.315	48.974
	CCC	0.791	0.778	0.722	0.77
ISF	$r^2$	0.586	0.577	0.566	0.592
	MAE	0.862	0.868	0.923	0.91
	RMSE	1.119	1.13	1.121	1.115
	$r^2m$	0.584	0.574	0.524	0.525
	$r^2mavg$	0.568	0.562	0.484	0.482
	$r^2mDelta$	0.034	0.024	0.08	0.085
	k	0.993	0.996	0.916	0.936
	ki	0.595	0.582	0.63	0.635
	F	330.968	302.773	32.615	34.808
	CCC	0.74	0.732	0.729	0.738

<sup>1</sup> The training set for the OSF is TrS + TeS.

### 3. Discussion

Here we propose classification and regression models for the carcinogenicity risk assessment of organic chemicals. Classification models are used to detect potential carcinogens and assume that negatively predicted compounds are non-carcinogenic. Meanwhile, the regression models quantify the potency of each chemical as a slope factor. We developed models for both inhalation and oral carcinogenicity.

The carcinogenicity models performed well: accuracies in the test set were 0.76 and 0.74 for oral and inhalation models, respectively, and  $r^2$  values were 0.57 and 0.65 in the regression models, respectively, for oral and inhalation SFs in ES.

Our results suggest that these models could be useful to support regulators in chemical safety assessments, providing information not only on the carcinogenic potential of chemicals but also as a measure of their potency. This latter information is fundamental to establish threshold concentrations of each chemical carcinogen, and it also gives a quantitative estimate of the risk of adverse health effects in exposed recipients.

#### 3.1. Focusing on Selected Descriptors

Even though the models were trained with different split schemes and different datasets, some descriptors were selected in more than one model, and others came from the same descriptor block (Table S2).

The cyclomatic number (nCIC), for example, was selected for both classification and regression slope factors and refers to the number of rings. The descriptor is related to the high carcinogenicity potency and to the large number of rings in the same molecule and is typically seen in polycyclic aromatic hydrocarbons (PAHs), which are carcinogenic through the formation of epoxides [40].

Another important class of descriptors selected for the regression models relates to the presence of chlorine. Several non-genotoxic mechanisms are influenced by the presence of halogens. For example, polychlorinated biphenyl (PCB) interaction with the aryl hydrocarbon receptor (AHR) plays a major role in breast cancer. It has also been reported that the affinity is related to the planar conformation of the molecule [41–43]. Though the developed model does not take account of 3D information, the position of halogens on the ring (B02[Cl-I], B04[O-Cl], B07[Cl-Cl], B08[Cl-Cl], B08[Cl-Cl] and F04[O-Cl]) heavily influences the planarity of the molecule. This is well known for dioxin-like PCBs [44].

Descriptors like nRNNNO<sub>x</sub> (number of nitroso groups) and nN-N identify several indirect alkylating agents, such as hydrazine or N-nitroso groups, that can form DNA adducts after metabolic activation.

#### 3.2. Usefulness of the Model

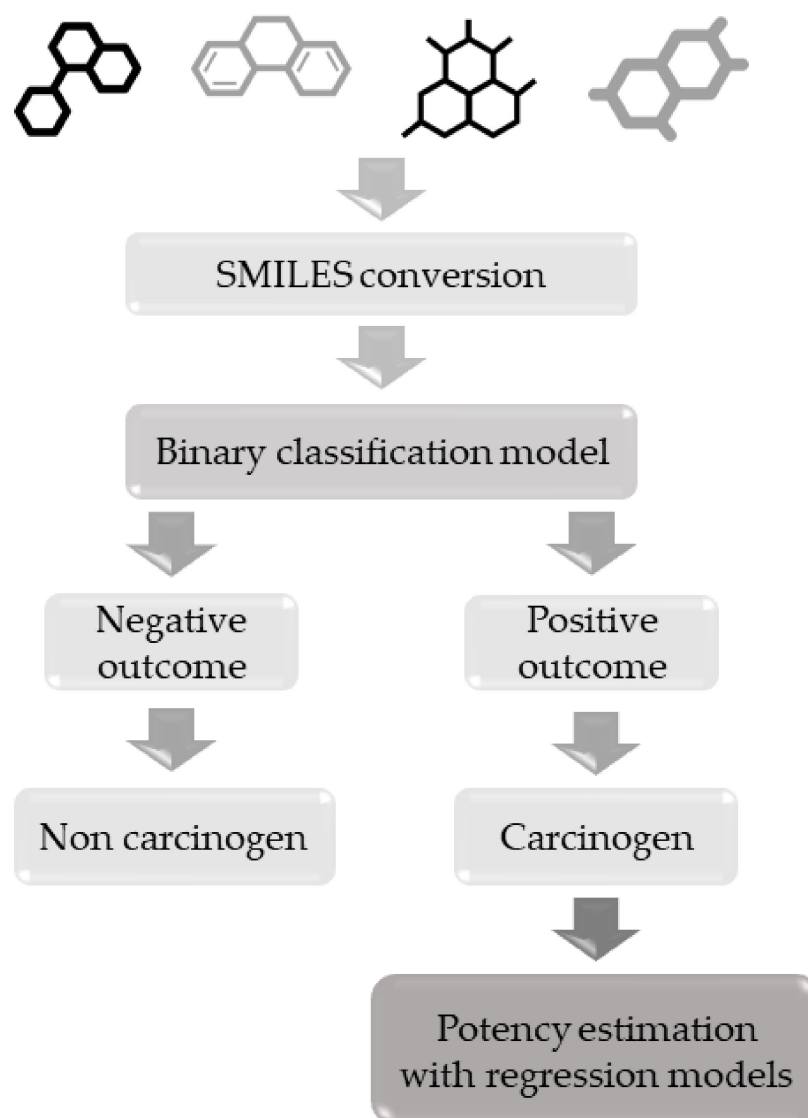
The classification models can help spot uncertain data from the original dataset. For example, vinyl chloride, 1,3-butadiene and chloromethane are predicted as non-carcinogens by the oral classification model, even though they have an oral slope factor value. If we look at the origin of the data, we see that the results for these three substances were extrapolated from inhalation tests on rats because they have a gaseous state at 20 °C [45].

The same information holds with the inhalation classification model. Thiourea, methylthiouracil and acetamide are classified as non-carcinogenic by the inhalation classification model, but there is an inhalation slope factor for them. The inhalation risk arises only if the particles are smaller than 5 µm [46,47]. The three misclassified substances have a particle size reported in the OECD QSAR toolbox [48] above this threshold, making them unlikely to have a carcinogenic effect via inhalation.

#### 3.3. Model Integration

The use of these models should follow a hierarchical pipeline. Since the regression models are based on a subset of compounds included in the classification dataset, a smaller chemical space will be covered. For this reason, substances should first be screened with the classification-based model in order to evaluate any carcinogenic effect, before then

evaluating the potency with the relative regression model. This suggested workflow is outlined in Figure 1.



**Figure 1.** Hierarchical workflow to apply classification and regression-based models for carcinogenicity.

#### 4. Materials and Methods

##### 4.1. Datasets and Data Curation

Regression and classification models were developed using data from the RAIS Toxicity values database [49]. Data cover different pollutant categories including organic and inorganic compounds, such as dioxins, PAHs, pesticides and metals frequently found in contaminated sites. We retrieved 1110 and 990 values for the oral slope factor (OSF,  $(\text{mg}/\text{kg}\cdot\text{day})^{-1}$ ) and for the inhalation unit risk (IUR,  $\mu\text{g}/\text{m}^3$ ), respectively.

In accordance with the United States Environmental Protection Agency (US EPA) [50], IUR data were converted to inhalation slope factor (ISF) using the formula

$$\text{ISF} = (\text{IUR} * \text{BW} * \text{CF})/\text{IR} \quad (2)$$

where IUR = inhalation unit risk [ $(\mu\text{g}/\text{m}^3)^{-1}$ ], BW = average body weight [70 kg], IR = inhalation rate [20  $\text{m}^3/\text{day}$ ] and CF = conversion factor [1000  $\mu\text{g}/\text{mg}$ ]. Both OSF and ISF values were then converted in logarithmic units for modelling purposes.



Canonical simplified molecular-input line-entry systems (canonical SMILES) were retrieved for each chemical from JChem for Office [51] and ChemID plus [52], and chemicals showing incongruency between the various sources were rejected. Most QSAR models cannot handle inorganic compounds, metals and metal complexes or organic salts, so data related to these compounds or mixtures were rejected. Ionized structures were neutralized and counterions eliminated. The datasets were further checked for duplicates.

Chemicals with a defined value (in our case SF) were considered carcinogenic, and compounds with no value were considered non-carcinogenic.

The final datasets for the classification models included 745 and 750 compounds, respectively, for OSF and ISF. For the regression models, only compounds with continuous data were used for modelling. This led to two final datasets comprising 315 compounds with OSF data and 263 with ISF data. Datasets are available on Zenodo [53] and also in the Supplementary Materials (Table S5: Dataset OSF and Table S6: Dataset ISF).

#### 4.2. Classification Models

We applied the same modelling workflow to OSF and ISF datasets. 2D molecular descriptors were calculated using DRAGON version 7.0.6 [54]. All available descriptors were selected, then pruned within DRAGON, removing those with missing values, constant and semi-constant values and redundant descriptors (those with a pairwise correlation over 0.95 with another descriptor).

For the training/test split, we used constitutional and ring descriptor blocks, together with the experimental class value, as input for principal component analysis (PCA). The first principal component (PC) was used to rank the compounds, then a venetian blind approach was used to split training and test set compounds in an 80–20% ratio.

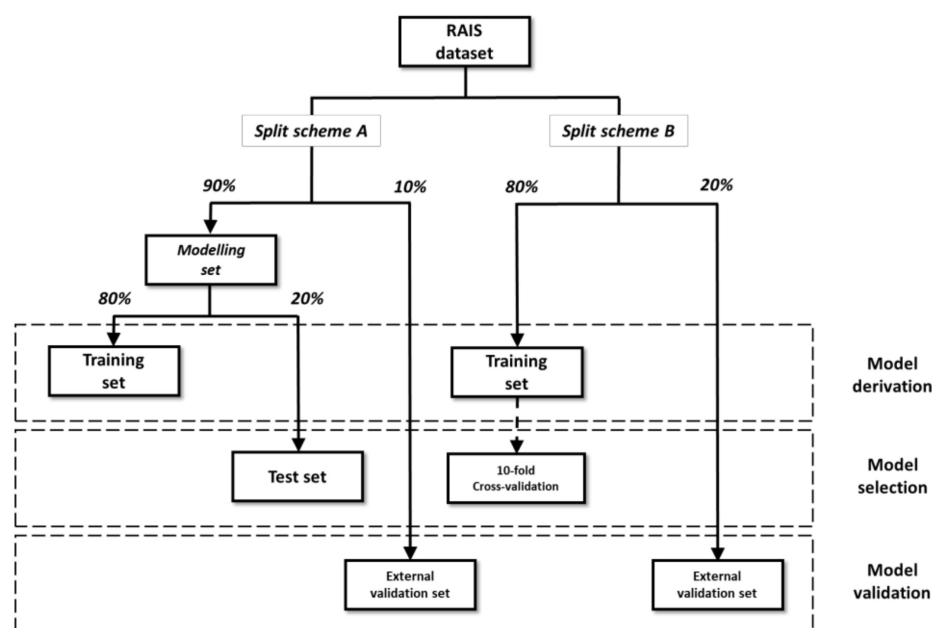
An in-house tool developed in the R statistical platform [55] was used to select the best descriptors set and size to be employed for the final model. The approach was based on a forward selection technique, a well-known general strategy previously used by our group to build models for toxicological endpoints [56,57]. In this approach, the descriptor leading to the best model was added at each iteration, starting from the descriptor most closely correlated with the experimental data, until the final number of 25 descriptors. Models were built with Linear Discriminant Analysis (LDA) modelling and applied with a bootstrap cross-validation approach (100 iterations). The fitness function was calculated for each model as a linear combination of the means for accuracy, sensitivity and specificity from the models built in each bootstrap iteration. This function was used to select the best descriptor to be added in the process to proceed to the next iteration. The set of descriptors with the best cross-validation values was used for the final model.

The “best” values were defined on the basis of their trend: by progressively adding descriptors to the model, cross-validation performance increases up to a plateau, meaning that the optimal number of descriptors has been reached, and adding further descriptors would lead to over-fitting.

The optimal set of descriptors was used to build a CART model in the R statistical environment (using the “rpart” package [58]) for both datasets to improve the performance of the naïve LDA approach used in the first step. The CART modelling implemented in R includes an internal cross-validation to reduce the complexity of the model; as a result, the final trees contain fewer descriptors than the set provided as input (while still leading to better performance than with the LDA models).

#### 4.3. Regression Models

Two different split schemes were applied on the OSF and ISF datasets (Figure 2). In split scheme A, 10% of the entire dataset was randomly extracted from the ES for external validation. The modeling set, consisting of the remaining chemicals, was split into a TrS and a TS [59] containing 80% and 20% of the modeling set, respectively. In split scheme B, only two datasets were created: TrS and ES, containing 80% and 20% of the entire dataset, respectively.



**Figure 2.** Graphic representation of the two schemes applied for dataset split. For each dataset, the function in model development (i.e., derivation, selection and validation) is reported.

For both split schemes, uniform distribution of the endpoint values among the subsets was ensured by applying an activity sampling method: compounds were sorted on the basis of their activity and divided into bins of equal size in terms of activity range. For each bin, chemicals were randomly assigned to datasets based on the percentages [60–62]. Table S3 reports the size of the datasets with the two split schemes, OSF and ISF.

As for the classification models, molecular descriptors were calculated for each compound with DRAGON 7 software, and the same pruning procedure was applied for the regression models.

Features were selected with the “gaselect” R package [63] that implements a partial least-squares genetic algorithm (PLS-GA) and repeated double cross-validation [64] for statistical analysis of subsets of descriptors. The following settings were applied for the PLS-GA: initial population 2000; number of iterations 5000; minimum number of variables 5; maximum number of variables 12. Optimal subsets of descriptors returned by the final iteration of the run were used for model derivation, using  $r^2$  as a fitness function.

Regression models were derived from each optimal subset of descriptors with multi-layer perceptron–artificial neural networks (MLP–ANNs) [65], as implemented in KNIME software. The MLP node is an implementation of the RPROP algorithm for multilayer feed-forward networks. This is a multi-layer perceptron trained with backpropagation that performs a local adaptation of the weight-updates according to the behavior of the error function. This solution has been found useful to overcome the inherent disadvantages of pure gradient-descent. In this algorithm, weights near the input layer have an equal chance to grow and learn as weights near the output layer. In addition, the descriptors were normalized in a range from 0 to 1 to be used in the algorithm.

MLP–ANNs were trained over 100 iterations and had a standard architecture formed by one input layer (with the same number of nodes as input descriptors), one hidden layer (with 10 neurons) and one output layer.

The applicability domain (AD) [66–68] was defined in order to identify less reliable predictions that were more likely to be wrong. The AD was evaluated by PCA. TrS chemicals were projected on a chemical space defined by the first two principal components (PCs) calculated using model descriptors. TeS and ES chemicals whose PC1 or PC2 values fell outside the range defined by the 5th and 95th percentiles of the distribution for TrS compounds were considered outside the AD and excluded from the statistical analysis of the models. This restrictive approach was preferred to avoid the inclusion of underrep-



resented areas within the chemical space defined by the first two PCs. The AD was also defined using the standardization approach [69].

Models for OSF and ISF were derived for each of the two split schemes. The selection of the best models was different based on the splitting scheme applied. In scheme A, models were ranked according to their  $r^2$  value on the TeS, without considering compounds outside the AD. The single best-performing OSF and ISF models were then evaluated for their external predictivity on the ES. In scheme B, single best models were selected according to internal performance (i.e.,  $r^2$  in 10-fold cross-validation) and then were evaluated for their external predictivity on the TeS.

All the models were then implemented in VEGA [39,70]. The OSF model has the same descriptors as the original DRAGON models since the VEGA engine is already able to calculate them. After recalculation of the descriptors with VEGA, the model was retrained using TrS + TeS, then validated on ES. Since selected descriptors for the inhalation regression model were not available as free descriptors in the VEGA platform, it was necessary to replicate the model using the same approach but with the descriptors already implemented in VEGA.

#### 4.4. Statistical Analysis

Accuracy, sensitivity and specificity were calculated according to Baratloo et al. [71] to evaluate the classification models. For regression models, we calculated the  $r^2$ , the RMSE and the MAE:

$$r^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - y_{avg})^2} = 1 - \frac{RSS}{TSS} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (4)$$

$$MAE = |y_i - \hat{y}_i| \quad (5)$$

where  $y_i$  is the observed dependent variable (the experimental response),  $\hat{y}_i$  is the calculated value,  $y_{avg}$  is the mean of the dependent variable, RSS is the residual sum of squares and TSS is the total sum of squares for  $n$  elements of the modeled dataset.

Finally,  $r^2m$  metrics (including  $r^2m$ , average  $r^2m$  and Delta  $r^2m$ ) were calculated for validation purposes according to previously published approaches [72,73].

## 5. Conclusions

The protection of human health is the most important goal of public health management. The need to characterize the effects of chemicals is now considered a priority research area for environmental protection agencies and national institutes of health in different countries. In this study, we propose four QSAR models to assess chemical carcinogenicity, based on inhalation and oral slope factors, which are key parameters for health risk assessment, especially in the investigation of contaminated sites. To our knowledge, few models are available to quantitatively assess carcinogenicity, and most of them only predict *in vivo* oral carcinogenicity in animal models. Our combined approach can classify a compound as potentially carcinogenic or not, and it can estimate its carcinogenic potency for humans in terms of oral and inhalation slope factors in cases of carcinogenic activity.

The proposed models could help regulators to evaluate chemical substances for carcinogenicity in humans. Making a version of the models freely available will permit easy screening of chemicals, which will greatly support health risk assessments.

**Supplementary Materials:** The following are available online. Table S1. Non-exhaustive list of software tools to predict carcinogenicity, adapted from Bossa et al., 2018 [32]. Table S2. Descriptors of the OSF and ISF models. Table S3. Sizes of datasets for the regression model. Table S4. QSAR Model Reporting Format of the carcinogenicity. Table S6 Dataset OSF. Table S7. Dataset ISF. Figure S1 Inhalation CART model. Figure S2 Oral CART model.

**Author Contributions:** Conceptualization, C.T., D.B. and A.R.; Data curation, C.T., G.R., M.M. and D.B.; Methodology, C.T., A.M., M.M. and D.G.; Project administration, E.B.; Software, A.M. and D.G.; Supervision, E.B. and N.K.; Validation, C.T., G.R., M.M. and A.R.; Writing—original draft, C.T.; Writing—review and editing, A.M., D.G., E.B., D.B., A.R. and N.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** We acknowledge the JANUS project (Z6 80710/20 Research (FKZ): 3716654140, German UBA) and the LIFE-VERMEER project (LIFE 16ENV/IT7000167) for their financial support.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are openly available in Zenodo at doi:10.5281/zenodo.4385768 and in VEGA HUB website at <https://www.vegahub.eu/portfolio-item/repository-of-models-dataset/>.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. De Rosa, C.; El-Masri, H.; Pohl, H.; Cibulas, W.; Mumtaz, M. Implications of chemical mixtures in public health practice. *J. Toxicol. Environ. Health Part B* **2004**, *7*, 339–350. [[CrossRef](#)] [[PubMed](#)]
2. Espina, C.; Straif, K.; Friis, S.; Kogevinas, M.; Saracci, R.; Vainio, H.; Schüz, J. European Code against Cancer 4th Edition: Environment, occupation and cancer. *Cancer Epidemiol.* **2015**, *39*, S84–S92. [[CrossRef](#)] [[PubMed](#)]
3. Li, D.; Suh, S. Health risks of chemicals in consumer products: A review. *Environ. Int.* **2019**, *123*, 580–587. [[CrossRef](#)] [[PubMed](#)]
4. Bhat, S.A.; Hassan, T.; Majid, S.; Ashraf, R.; Kuchy, S. Environmental Pollution as Causative Agent for Cancer—A Review. *Cancer Clin. Res. Rep.* **2017**, *1*, 3.
5. Ferlay, J.; Soerjomataram, I.; Ervik, M.; Dikshit, R.; Eser, S.; Mathers, C.; Rebeiro, M.; Parkin, D.M.; Forman, D.; Bray, F. *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. 2013*; International Agency for Research on Cancer: Lyon, France, 2015.
6. Madia, F.; Worth, A.; Corvi, R. *Analysis of Carcinogenicity Testing for Regulatory Purposes in the European Union*; European Commission: Luxembourg, 2016.
7. OECD. *Test No. 451: Carcinogenicity Studies*; OECD Publishing: Paris, France, 2018.
8. OECD. *Test No. 453: Combined Chronic Toxicity/Carcinogenicity Studies*; OECD Publishing: Paris, France, 2018.
9. Schechtman, L.M. Rodent cell transformation assays—A brief historical perspective. *Mutat. Res. Genet. Toxicol. Environ. Mutag.* **2012**, *744*, 3–7. [[CrossRef](#)]
10. Combes, R.; Balls, M.; Curren, R.; Fischbach, M.; Fusenig, N.; Kirkland, D.; Lasne, C.; Landolph, J.; LeBoeuf, R.; Marquardt, H.; et al. Cell transformation assays as predictors of human carcinogenicity. *Altern. Lab. Anim.* **1999**, *27*, 745–768. [[CrossRef](#)]
11. Cronin, M.T.; Jaworska, J.S.; Walker, J.D.; Comber, M.H.; Watts, C.D.; Worth, A.P. Use of QSARs in international decision-making frameworks to predict health effects of chemical substances. *Environ. Health Perspect.* **2003**, *111*, 1391. [[CrossRef](#)]
12. Vasseur, P.; Lasne, C. Oecd detailed review paper (drp) number 31 on “Cell transformation assays for detection of chemical carcinogens”: Main results and conclusions. *Mutat. Res. Genet. Toxicol. Environ. Mutagenesis* **2012**, *744*, 8–11. [[CrossRef](#)]
13. Benigni, R.; Bossa, C.; Tcheremenskaia, O.; Giuliani, A. Alternatives to the carcinogenicity bioassay: In silico methods, and the in vitro and in vivo mutagenicity assays. *Expert Opin. Drug Metab. Toxicol.* **2010**, *6*, 809–819.
14. Rohrbeck, A.; Salinas, G.; Maaser, K.; Linge, J.; Salovaara, S.; Corvi, R.; Borlak, J. Toxicogenomics applied to in vitro carcinogenicity testing with Balb/c 3T3 cells revealed a gene signature predictive of chemical carcinogens. *Toxicol. Sci.* **2010**, *118*, 31–41. [[CrossRef](#)]
15. Milan, C.; Schifanella, O.; Roncaglioni, A.; Benfenati, E. Comparison and possible use of in silico tools for carcinogenicity within REACH legislation. *J. Environ. Sci. Health Part C* **2011**, *29*, 300–323.
16. Carrasquer, C.; Malik, N.; States, G.; Qamar, S.; Cunningham, S.; Cunningham, A. Chemical structure determines target organ carcinogenesis in rats. *SAR QSAR Environ. Res.* **2012**, *23*, 775–795. [[CrossRef](#)] [[PubMed](#)]
17. Corvi, R.; Aardema, M.J.; Gribaldo, L.; Hayashi, M.; Hoffmann, S.; Schechtman, L.; Vanparys, P. ECVAM prevalidation study on in vitro cell transformation assays: General outline and conclusions of the study. *Mutat. Res. Genet. Toxicol. Environ. Mutag.* **2012**, *744*, 12–19.
18. Golbamaki, A.; Benfenati, E. In silico methods for carcinogenicity assessment. In *Silico Methods for Predicting Drug Toxicity*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 107–119.
19. Golbamaki, A.; Benfenati, E.; Golbamaki, N.; Manganaro, A.; Merdivan, E.; Roncaglioni, A.; Gini, G. New clues on carcinogenicity-related substructures derived from mining two large datasets of chemical compounds. *J. Environ. Sci. Health Part C* **2016**, *34*, 97–113. [[CrossRef](#)]
20. Yamane, J.; Aburatani, S.; Imanishi, S.; Akanuma, H.; Nagano, R.; Kato, T.; Sone, H.; Ohsako, S.; Fujibuchi, W. Prediction of developmental chemical toxicity based on gene networks of human embryonic stem cells. *Nucleic Acids Res.* **2016**, *44*, 5515–5528. [[CrossRef](#)]
21. Benfenati, E.; Benigni, R.; Demarini, D.M.; Helma, C.; Kirkland, D.; Martin, T.M.; Mazzatorta, P.; Ouedraogo-Arras, G.; Richard, A.M.; Schilter, B.; et al. Predictive models for carcinogenicity and mutagenicity: Frameworks, state-of-the-art, and perspectives. *J. Environ. Sci. Health Part C* **2009**, *27*, 57–90.

22. Fjodorova, N.; Vračko, M.; Tušar, M.; Jezierska, A.; Novič, M.; Kühne, R.; Schüürmann, G. Quantitative and qualitative models for carcinogenicity prediction for non-congeneric chemicals using CP ANN method for regulatory uses. *Mol. Divers* **2010**, *14*, 581–594.
23. Wu, X.; Zhang, Q.; Wang, H.; Hu, J. Predicting carcinogenicity of organic compounds based on CPDB. *Chemosphere* **2015**, *139*, 81–90.
24. Zhang, L.; Ai, H.; Chen, W.; Yin, Z.; Hu, H.; Zhu, J.; Zhao, J.; Zhao, Q.; Liu, H. CarcinoPred-EL: Novel models for predicting the carcinogenicity of chemicals using molecular fingerprints and ensemble learning methods. *Sci. Rep.* **2017**, *7*, 2118.
25. Matthews, E.J.; Contrera, J.F. A New Highly Specific Method for Predicting the Carcinogenic Potential of Pharmaceuticals in Rodents Using EnhancedMCASEQSAR-ES Software. *Regul. Toxicol. Pharmacol.* **1998**, *28*, 242–264. [[CrossRef](#)]
26. Gini, G.; Lorenzini, M.; Benfenati, E.; Grasso, P.; Bruschi, M. Predictive carcinogenicity: A model for aromatic compounds, with nitrogen-containing substituents, based on molecular descriptors using an artificial neural network. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1076–1080. [[CrossRef](#)] [[PubMed](#)]
27. Toropova, A.; Toropov, A.; Diaza, R.; Benfenati, E.; Gini, G. Analysis of the co-evolutions of correlations as a tool for QSAR-modeling of carcinogenicity: An unexpected good prediction based on a model that seems untrustworthy. *Open Chem.* **2011**, *9*, 165–174. [[CrossRef](#)]
28. Wang, N.C.Y.; Venkatapathy, R.; Bruce, R.M.; Moudgal, C. Development of quantitative structure–activity relationship (QSAR) models to predict the carcinogenic potency of chemicals. II. Using oral slope factor as a measure of carcinogenic potency. *Regul. Toxicol. Pharmacol.* **2011**, *59*, 215–226. [[CrossRef](#)] [[PubMed](#)]
29. Kar, S.; Deeb, O.; Roy, K. Development of classification and regression based QSAR models to predict rodent carcinogenic potency using oral slope factor. *Ecotoxicol. Environ. Saf.* **2012**, *82*, 85–95. [[CrossRef](#)]
30. Piparo, E.L.; Maunz, A.; Helma, C.; Vorgrimmler, D.; Schilter, B. Automated and reproducible read-across like models for predicting carcinogenic potency. *Regul. Toxicol. Pharmacol.* **2014**, *70*, 370–378. [[CrossRef](#)]
31. Raitano, G.; Goi, D.; Pieri, V.; Passoni, A.; Mattiussi, M.; Lutman, A.; Romeo, I.; Manganaro, A.; Marzo, M.; Porta, N.; et al. (Eco) toxicological maps: A new risk assessment method integrating traditional and in silico tools and its application in the Ledra River (Italy). *Environ. Int.* **2018**, *119*, 275–286. [[CrossRef](#)]
32. Bossa, C.; Benigni, R.; Tcheremenskaia, O.; Battistelli, C.L. (Q) SAR Methods for Predicting Genotoxicity and Carcinogenicity: Scientific Rationale and Regulatory Frameworks. In *Computational Toxicology*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 447–473.
33. USEPA. *Guidelines for Carcinogen Risk Assessment*; U.S. Environmental Protection Agency: Washington, DC, USA, 2005.
34. USEPA. *Definitions of Key Terms Related to the Risk-Screening Environmental Indicators Model (RSEI)*; U.S. Environmental Protection Agency: Washington, DC, USA, 2013.
35. USEPA. *Basic Information about the Integrated Risk Information System (IRIS)*; U.S. Environmental Protection Agency: Washington, DC, USA, 2016.
36. USEPA. *Health Effects Assessment Summary Tables (Heast)*; U.S. Environmental Protection Agency: Washington, DC, USA, 2017.
37. Martin, T. *User's Guide for TEST (Version 4.2) (Toxicity Estimation Software Tool): A Program to Estimate Toxicity from Molecular Structure*; U.S. Environmental Protection Agency: Washington, DC, USA, 2016.
38. Gramatica, P.; Cassani, S.; Chirico, N. QSARINS-chem: Insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS. *J. Comput. Chem.* **2014**, *35*, 1036–1044.
39. Benfenati, E.; Manganaro, A.; Gini, G.C. (Eds.) VEGA-QSAR: AI Inside a Platform for Predictive Toxicology. In *CEUR Workshop Proceedings*; CEUR-WS.: Berlin, Germany, 2013; Volume 1107, pp. 21–28.
40. Benigni, R.; Bossa, C. Mechanisms of chemical carcinogenicity and mutagenicity: A review with implications for predictive toxicology. *Chem. Rev.* **2011**, *111*, 2507–2536. [[CrossRef](#)]
41. Kafafi, S.A.; Afeefy, H.Y.; Ali, A.H.; Said, H.K.; Kafafi, A.G. Binding of polychlorinated biphenyls to the aryl hydrocarbon receptor. *Environ. Health Perspect.* **1993**, *101*, 422. [[CrossRef](#)]
42. Vondráček, J.; Machala, M.; Bryja, V.; Chramostová, K.; Krčmář, P.; Dietrich, C.; Hampl, A.; Kozubík, A. Aryl hydrocarbon receptor-activating polychlorinated biphenyls and their hydroxylated metabolites induce cell proliferation in contact-inhibited rat liver epithelial cells. *Toxicol. Sci.* **2004**, *83*, 53–63.
43. Feng, S.; Cao, Z.; Wang, X. Role of aryl hydrocarbon receptor in cancer. *Biochim. Et Biophys. Acta (Bba)-Rev. Cancer* **2013**, *1836*, 197–210. [[CrossRef](#)] [[PubMed](#)]
44. Van den Berg, M.; Birnbaum, L.; Bosveld, A.T.; Brunström, B.; Cook, P.; Feeley, M.; Giesy, J.P.; Hanberg, A.; Hasegawa, R.; Kennedy, S.W.; et al. Toxic equivalency factors (TEFs) for PCBs, PCDDs, PCDFs for humans and wildlife. *Environ. Health Perspect.* **1998**, *106*, 775–792. [[CrossRef](#)] [[PubMed](#)]
45. OEHHA. *Air Toxics Hot Spot Program Technical Support Document for Cancer Potencies. Appendix B. Chemical-Specific Summaries of the Information Used to Derive Unit Risk and Cancer Potency Values*; Updated 2011; California Environmental Protection Agency, Office of Environmental Health Hazard Assessment: California, CA, USA, 2011.
46. Rothe, H.; Fautz, R.; Gerber, E.; Neumann, L.; Rettinger, K.; Schuh, W.; Gronewold, C. Special aspects of cosmetic spray safety evaluations: Principles on inhalation risk assessment. *Toxicol. Lett.* **2011**, *205*, 97–104. [[CrossRef](#)] [[PubMed](#)]
47. DFG. *MAK-Und BAT-Werte-Liste 2010. Senatskommission Zur Prüfung Gesundheitsschädlicher Arbeitsstoffe*; Deutsche Forschungsgemeinschaft eV: Bonn, Germany, 2010.

48. Dimitrov, S.D.; Diderich, R.; Sobanski, T.; Pavlov, T.S.; Chankov, G.V.; Chapkanov, A.S.; Karakolev, Y.H.; Temelkov, S.G.; Vasilev, R.A.; Gerova, K.D.; et al. QSAR Toolbox–workflow and major functionalities. *SAR QSAR Environ. Res.* **2016**, *27*, 203–219. [[CrossRef](#)] [[PubMed](#)]
49. RAIS. RAIS Toxicity Values and Physical Parameters Search. University of Tennessee, 2017. Available online: [https://rais.ornl.gov/cgi-bin/tools/TOX\\_search?select=chemtox](https://rais.ornl.gov/cgi-bin/tools/TOX_search?select=chemtox) (accessed on 23 December 2020).
50. USEPA. *Technical Appendix A. Toxicity Weights for TRI Chemicals and Chemical Categories*; U.S. Environmental Protection Agency: Washington, DC, USA, 2015.
51. Chemaxon. *JChem for Office (Excel), 17.22 ed2017*; ChemAxon: Budapest, Hungary, 2017.
52. Tomasulo, P. ChemIDplus-super source for chemical and drug information. *Med Ref. Serv. Q.* **2002**, *21*, 53–59. [[CrossRef](#)] [[PubMed](#)]
53. Toma, C.; Baderna, D.; Roncaglioni, A.; Benfenati, E. Datasets of the proposed models for carcinogenicity assessment based on slope factor. *Zenodo Repos.* **2020**. [[CrossRef](#)]
54. Kode. *DRAGON (Software for Molecular Descriptor Calculation) Version 7.0*; Kode: Pisa, Italy, 2016.
55. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
56. Xu, L.; Zhang, W.-J. Comparison of Different Methods for Variable Selection. *Anal. Chim. Acta* **2001**, *446*, 475–481. [[CrossRef](#)]
57. Manganelli, S.; Roncaglioni, A.; Mansouri, K.; Judson, R.S.; Benfenati, E.; Manganaro, A.; Ruiz, P. Development, Validation and Integration of in Silico Models to Identify Androgen Active Chemicals. *Chemosphere* **2019**, *220*, 204–215. [[CrossRef](#)]
58. Therneau, T.; Atkinson, B.; Ripley, B. *Rpart: Recursive Partitioning and Regression Trees*, R package version 4.1–10; Mayo Foundation for Medical Education and Research: Rochester, FL, USA, 2015.
59. Tropsha, A. Best practices for QSAR model development, validation, and exploitation. *Mol. Inform.* **2010**, *29*, 476–488. [[CrossRef](#)]
60. Worth, A.P.; Bassan, A.; Gallegos, A.; Netzeva, T.I.; Patlewicz, G.; Pavan, M.; Tsakovska, I.; Vracko, M. *The characterisation of (quantitative) Structure-Activity Relationships: Preliminary Guidance*; Institute for Health and Consumer Protection, European Chemical Bureau: Ispra, Italy, 2005.
61. Golbraikh, A.; Tropsha, A. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *Mol. Divers.* **2000**, *5*, 231–243. [[CrossRef](#)]
62. Golbraikh, A.; Shen, M.; Xiao, Z.; Xiao, Y.D.; Lee, K.H.; Tropsha, A. Rational selection of training and test sets for the development of validated QSAR models. *J. Comput. Aided Mol. Des.* **2003**, *17*, 241–253. [[CrossRef](#)] [[PubMed](#)]
63. Kepplinger, D. *Gaselect: Genetic Algorithm (GA) for Variable Selection from High-Dimensional Data R Package Version 1.0.5 ed2015*; CRAN Repository, Wirtschaftsuniversität Wien: Vienna, Austria, 2015.
64. Filzmoser, P.; Liebmann, B.; Varmuza, K. Repeated double cross validation. *J. Chemom. A J. Chemom. Soc.* **2009**, *23*, 160–171. [[CrossRef](#)]
65. Riedmiller, M.; Braun, H. (Eds.) A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In Proceedings of the IEEE International Conference on Neural Networks, San Francisco, CA, USA, 28 March–1 April 1993.
66. Jaworska, J.; Nikolova-Jeliazkova, N.; Aldenberg, T. QSAR applicability domain estimation by projection of the training set descriptor space: A review. *Altern. Lab. Anim.* **2005**, *33*, 445. [[CrossRef](#)] [[PubMed](#)]
67. Gadaleta, D.; Mangiatordi, G.F.; Catto, M.; Carotti, A.; Nicolotti, O. Applicability domain for QSAR models: Where theory meets reality. *Int. J. Quant. Struct. Prop. Relatsh. (IJQSPR)* **2016**, *1*, 45–63. [[CrossRef](#)]
68. Netzeva, T.I.; Worth, A.P.; Aldenberg, T.; Benigni, R.; Cronin, M.T.D.; Gramatica, P.; Jaworska, J.S.; Kahn, S.; Klopman, G.; Marchant, C.A.; et al. Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. *Altern. Lab. Anim.* **2005**, *33*, 155–173. [[CrossRef](#)]
69. Roy, K.; Kar, S.; Ambure, P. On a Simple Approach for Determining Applicability Domain of QSAR Models. *Chemom. Intell. Lab. Syst.* **2015**, *145*, 22–29. [[CrossRef](#)]
70. Benfenati, E.; Roncaglioni, A.; Lombardo, A.; Manganaro, A. Integrating QSAR, Read-Across, and Screening Tools: The VEGAHUB Platform as an Example. In *Advances in Computational Toxicology*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 365–381.
71. Baratloo, A.; Hosseini, M.; Negida, A.; El Ashal, G. Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity. *Emerg (Tehran)* **2015**, *3*, 48–49.
72. Roy, K.; Chakraborty, P.; Mitra, I.; Ojha, P.K.; Kar, S.; Das, R.N. Some Case Studies on Application of “Rm2” Metrics for Judging Quality of Quantitative Structure–Activity Relationship Predictions: Emphasis on Scaling of Response Data. *J. Comput. Chem.* **2013**, *34*, 1071–1082. [[CrossRef](#)]
73. Gajo, G.C.; De Assis, T.M.; Assis, L.C.; Ramalho, T.C.; da Cunha, E.F.F. Quantitative Structure-Activity Relationship Studies for Potential Rho-Associated Protein Kinase Inhibitors. Available online: <https://www.hindawi.com/journals/jchem/2016/9198582/> (accessed on 23 December 2020).