This is the published version of a paper published in *Journal of Chemometrics*.

Citation for the original published paper (version of record):

Ghorbanzadeh, M., Zhang, J., Andersson, P L. (2016)  
Binary classification model to predict developmental toxicity of industrial chemicals in zebrafish.  
*Journal of Chemometrics*, 30(6): 298-307  
http://dx.doi.org/10.1002/cem.2791

Access to the published version may require subscription.  
N.B. When citing this work, cite the original published paper.

Permanent link to this version:  
http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-125560
Binary classification model to predict developmental toxicity of industrial chemicals in zebrafish

Mehdi Ghorbanzadeh*, Jin Zhang and Patrik L. Andersson

The identification of industrial chemicals, which may cause developmental effects, is of great importance for an early detection of hazardous chemicals. Accordingly, categorical quantitative structure-activity relationship (QSAR) models were developed, based on developmental toxicity profile data for zebrafish from the ToxCast Phase I testing, to predict the toxicity of a large set of high and low production volume chemicals (H/LPVCs). QSARs were created using linear (LDA), quadratic, and partial least squares-discriminant analysis with different chemical descriptors. The predictions of the best model (LDA) were compared with those obtained by the freely available QSAR model VEGA, created based on a dataset with a different chemical domain. The results showed that despite similar accuracy (AC) of both models, the LDA model is more specific than VEGA and shows a better agreement between sensitivity (SE) and specificity (SP). Applying a 90% confidence level on the LDA model led to even better predictions showing SE of 0.92, AC of 0.95, and geometric mean of SE and SP (G) of 0.96 for the prediction set. The LDA model predicted 608 H/LPVCs as toxicants among which 123 chemicals fall inside the AD of the VEGA model, which predicted 112 of those as toxicants. Among the 112 chemicals predicted as toxic H/LPVCs, 23 have been previously reported as developmental toxicants. The here presented LDA model could be used to identify and prioritize H/LPVCs for subsequent developmental toxicity assessment, as a screening tool of potential developmental effects of new chemicals, and to guide synthesis of safer alternative chemicals. © 2016 The Authors Journal of Chemometrics Published by John Wiley & Sons Ltd.

Keywords: classification; QSAR; developmental toxicity; industrial chemicals; zebrafish

1. INTRODUCTION

Developmental toxicity refers to adverse effects of chemicals or physical agents on organismal development. Major manifestations of developmental toxicity include death of the developing organism, structural abnormality, altered growth, or functional deficiency. Unfortunately, many developmental toxicants have been detected after human exposure; thus, means for early detection of such chemicals are warranted to prevent or reduce the risk of exposure to developmental hazards. This could be carried out by screening potential toxicants in animal models, such as zebrafish (Danio rerio). Data on critical environmental and human health effects are lacking for a large share of the thousands of chemicals used in commerce and industry including developmental effects. The field of developmental toxicity has recently adopted the zebrafish as a vertebrate toxicity screening model for effect assessment of chemicals on human health. Zebrafish has the advantage of being easy to breed and maintain, producing a large number of offspring per week and providing a vertebrate model for studying mammalian disease. Despite its advantages, acquiring data on zebrafish developmental toxicity is complex, time-consuming, labor intensive and expensive. Besides, there is a large number of existing chemicals for which there is very little information on developmental toxicity. Thus, rapid and inexpensive methods to prioritize potent chemicals for further testing are warranted. Non-testing techniques such as read-across methodologies and quantitative structure activity relationships (QSARs) are approaches that could potentially be used to screen large chemical inventories and identify the most potent industrial chemicals. The data obtained from in silico QSAR models, as recommended by the European Union’s chemicals legislation (Registration, Evaluation, Authorisation and Restriction of Chemicals), can help predict lacking experimental data and also screen and prioritize chemicals for toxicity testing in animal models. QSAR models potentially save cost and time and overcome the complexity of experimental methods in addition to reducing experimental animals, which is a critical ethical concern in toxicity testing.

A limited number of QSAR models has been established to predict developmental toxicity of chemicals. These are frequently based on information from the Teratogen Information System (TERIS) and US Food and Drug Administration (FDA). There also exist some commercially available models such as MultiCASE (MC4PC) and TOPKAT, which are typically based on classification models that categorize unknown chemicals according to developmental toxicity data for tested chemicals.

* Correspondence to: Mehdi Ghorbanzadeh, Department of Chemistry, Umeå University, SE-901 87 Umeå, Sweden. E-mail: ghorbanzade@gmail.com

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

M. Ghorbanzadeh, J. Zhang, P. L. Andersson
Department of Chemistry, Umeå University, SE-901 87, Umeå, Sweden
Developmental toxicity modeling

In the present study, we developed QSAR classification models using a well-defined dataset from one data source (ToxCast Phase I) to predict developmental toxicity of a large set of high and low production volume chemicals (H/LPVCs) with the aim of improving existing categorical models and to address the need for a non-testing method to prioritize H/LPVCs for subsequent toxicity testing. The models were developed using fast machine learning approaches and different sets of chemical descriptors. To our knowledge, these toxicity data have not yet been used for QSAR modeling. The performance of the created models was assessed by internal (cross-validation (CV)) and external validation procedures. The influence of potential outlying chemicals and significant molecular properties, which may be responsible for inducing developmental toxicity in zebrafish, are discussed. The best binary classification model selected after evaluating the performance of all models was applied to predict the developmental toxicity of H/LPVCs falling inside its applicability domain (AD). The predictions of the model were then compared with those obtained using the freely available model VEGA. Finally, the model was used to set a priority list of possibly hazardous H/LPVCs suggested for subsequent empirical assessments.

2. MATERIALS AND METHODS

2.1. Data sources

The present study was conducted using developmental toxicity data collected from the Environmental Protection Agency Phase I ToxCast library [5]. Briefly, the library is structurally diverse, with over 40 functional classes and more than 24 pesticidal modes of action, and consists of 309 environmental chemicals assessed using a zebrafish screen for developmental toxicity. In a concentration response study in developing zebrafish embryos, 191 chemicals were identified as toxic and 118 as nontoxic. Since molecular descriptors can only be calculated from well-defined molecular structures, the data set used for modeling was reduced to 291 chemicals (185 toxic and 106 nontoxic) by removing compounds with multiple components, such as mixtures and salts, and other chemicals that are not unique substances and do not hold unique Chemical Abstract Service (CAS) registry numbers. The advantage of using such a structurally diverse set of chemicals for QSAR modeling is that the developed model covers a wide AD and can therefore predict a variety of untested compounds’ toxicity. The industrial H/LPVCs dataset initially contained 10 614 chemicals [15]. After applying a number of filtrations, as described in reference [16], the final list contained 1341 HPVCs and 5316 LPVCs. The developed models were applied to the H/LPVCs to identify compounds potentially toxic to developing zebrafish.

2.2. Dataset splitting

Splitting a dataset into training and prediction (external test) sets is required to obtain a decision rule allowing the prediction of an untested compound’s toxicity. The Kennard–Stone (KS) algorithm [17] was applied to the ToxCast dataset to select training and prediction sets. To keep the same ratio of toxic to nontoxic compounds in training and prediction sets, the KS algorithm was separately applied to each class of chemicals. That is, 80% of each class of compounds was merged to form a training set and the remaining 20% to form a prediction set. Accordingly, the training set consisted of 233 (148 toxic (80% of toxic compounds) and 85 nontoxic (80% of nontoxic compounds)) members and the prediction set of 58 (37 toxic and 21 nontoxic) members. The training set was used to develop binary classification models and the prediction set to assess the predictive performance of the developed models.

2.3. Calculation, screening, and selection of molecular features

Each chemical’s molecular structure was represented in line notation using the simplified molecular-input line-entry system notation and subsequently imported into the MOE [18] and Dragon [19] programs to calculate its 2D molecular descriptors. All descriptors were cleaned up by removing missing values, constant, and near-constant variables. A pool of 2933 descriptors, 2745 from Dragon and 188 from MOE, were employed for further analysis. Because the whole molecular descriptors may not be appropriate for classification analysis, we removed redundant descriptors to increase the correctness of prediction and also to simplify the interpretation of the developed model (by focusing on the most appropriate descriptors). Using Pearson correlation analysis to identify highly correlated molecular descriptors (correlation coefficient 70% and more), the descriptor that was easier to interpret was retained for classification analysis. The resulting descriptor sets are referred to as the pretreated Dragon, MOE, or combined (Dragon and MOE) descriptor sets. To further reduce the effect of irrelevant descriptors, and to extract the features relevant for distinguishing between toxic and nontoxic compounds, appropriate descriptors were separately chosen from the pretreated MOE, Dragon, and a combined MOE and Dragon descriptor set by performing leave-one-out (LOO) linear classifications in a stepwise manner using Wilk’s lambda method [20]. At each step, the descriptor minimizing the overall Wilk’s lambda was added to the model. The F values for inclusion and exclusion of descriptors were set to 3.84 and 2.71, respectively. A descriptor with an F value greater than the inclusion value was added to the model. Subsequently, F values were recalculated for the rest of the descriptors and those with an F value lower than the exclusion value were removed. The procedure was continued until the F values of the remaining variables were all less than the inclusion F value. For each set of pretreated descriptors, classification parameters were calculated for the training and prediction sets to compare the performance of the created classification models together. The descriptor set, which improved classification performance, was retained for further modeling analysis.

2.4. Methods for model development

Three classification-based QSAR modeling methods, namely linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and partial least squares-discriminant analysis (PLS-DA) were employed to discover relationships between molecular structures, encoded by calculated molecular features, and developmental toxicity of the studied chemicals. Each method is briefly described in the following sections.

2.4.1. Linear discriminant analysis

Linear discriminant analysis looks for a discriminant function of the variables, which maximizes the ratio of between-class variance and minimizes the ratio of within-class variance [21]. The discriminant function is provided by dividing an n-dimensional descriptor space into two regions separated by a hyperplane. These two regions correspond to two classes to which individual compounds belong.
2.4.2. Quadratic discriminant analysis

In QDA, which is closely related to the LDA, the hyperplane dividing the classes is quadratic and the covariance matrix can be different for each class [22].

2.4.3. Partial least squares-discriminant analysis.

Partial least squares-discriminant analysis is a linear classification method that involves partial least squares regression where the response variable is binary class membership expressing the class to which each compound belongs [23,24]. The relationship between the molecular descriptors and binary variables is obtained by calculating latent variables (LVs). The number of LVs is determined using classification parameters after CV.

2.5. Evaluation of models

Internal and external validation tests were performed on the LDA, QDA, and PLS-DA based binary classification models. For internal validation, LOO CV and fivefold CV tests were performed by excluding each compound (fold) once and creating a classification model without this compound (fold). The created model was subsequently asked to predict the class of the excluded compound (fold). Thus, in each round, a classification model was built and tested on the unseen compound (fold). For external validation of the binary classification models we used the compounds from the prediction set. To select the superior classification model we estimated the performance of all models using selected classification quality parameters applied to the training set, LOO-CV and fivefold CV tests, and prediction set. In addition to accuracy (AC), which is commonly used to measure a classification model’s overall prediction performance and refers to the ratio of correctly classified compounds, we calculated the alternative parameters sensitivity (SE) and specificity (SP), which measure class AC separately on toxic and non-toxic classes, respectively. We added the SE and SP parameters because although the AC parameter is commonly used to measure a classification model’s performance, AC is dependent on the data balance and may be biased toward the majority class [25]. This can cause confusing predictions in this study because the data set (with respect to the number of compounds in each class) is imbalanced. We used the following equations to calculate the parameters of the classification models’ predictive abilities:

\[
\text{AC} = \frac{(TP + TN)}{(TP + FN + TN + FP)} \quad (1)
\]

\[
\text{SE} = \frac{TP}{(TP + FN)} \quad (2)
\]

\[
\text{SP} = \frac{TN}{(TN + FP)} \quad (3)
\]

where TP and TN denote the number of true positives (a toxic compound predicted as toxic) and true negatives (a non-toxic compound predicted as non-toxic) and FP and FN denote the number of false positives (a non-toxic compound predicted as toxic) and false negatives (a toxic compound predicted as non-toxic), respectively. In addition to calculating the SE and SP parameters, the geometric mean of SE and SP (G) [26], which is not biased towards the majority class as it takes both SE and SP into consideration, was applied to measure the prediction performance on the imbalanced data set. Matthew’s Correlation Coefficient (MCC) [27,28], a measure of the quality of the binary classification models, was also calculated to further evaluate the balanced prediction of the classification models. The equations for computing the parameters G and MCC are

\[
G = \sqrt{SE \times SP} \quad (4)
\]

\[
MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FN)(TP + FP)(TN + FP)(TN + FN)}} \quad (5)
\]

2.6. Applicability domain

The AD of the developed models was investigated as recommended by the Organization for Economic Cooperation and Development (OECD) [29]. The AD is defined as the chemical space on which the model has been developed and for which it is applicable to make predictions for new compounds with high reliability. For a more confident AD two value ranges were used, namely the value range of the first five principal components (PCs) of the whole set of pretreated descriptors explaining more than 70% of the variation in the chemical space, and the value ranges of the most influential descriptors. The first five PCs were extracted using principal component analysis (PCA), which is a statistical method that uses orthogonal transformation to generate linearly uncorrelated variables (PCs) from a set of possibly correlated variables. The range is defined as the interval between minimum and maximum values of PCs and descriptors in the training set. According to this method, a compound with both PC and descriptor values within the range of those of the training set compounds was considered to be inside the AD.

3. Results and discussion

3.1. Chemical space of studied data sets

To investigate the chemical variation of the H/LPVCs and ToxCast chemicals a PCA was calculated using 137 MOE and Dragon molecular descriptors (with no constant or near constant values) covering a wide variety of descriptor types, e.g., physicochemical properties (such as log P, molecular weight (MW), and molar refractivity), atom counts and bond counts, partial charge descriptors, topological indices, functional group counts, connectivity indices, and 2D-matrix based descriptors. Figure 1 shows the score plot, explaining 61% of variation by the first two PCs (51% by PC1 and 10% by PC2) with a cross-validated explained variation (Q2) of 60%. The ToxCast data set, which is structurally diverse and includes mainly pesticides and antimicrobials [5], covers a portion of the H/LPVCs in terms of molecular features. The analysis encircles the chemical domain of the H/LPVCs that could be covered by models developed using the ToxCast dataset.

3.2. Feature selection

Accurate classification models need an appropriate number of relevant molecular descriptors to attain a high prediction performance. A stepwise feature selection process using predefined F values for inclusion or exclusion of descriptors and the classification probabilities of 0.5 for the two classes resulted in a selection of three descriptors from the 66 pretreated MOE descriptors (Wilk’s lambda = 0.82), 10 descriptors from the 127 pretreated Dragon descriptors (Wilk’s lambda = 0.56), and 15 descriptors...
from the pretreated combined (MOE and Dragon) descriptor sets (Wilk’s lambda = 0.50). We evaluated the SE, SP, and AC of the discriminant models derived using the selected descriptors (Supporting Information, Figure S1). The parameters SE, SP, and AC for the models based on the combined (ranging from 0.82 to 0.84) and Dragon (ranging from 0.77 to 0.79) descriptor sets were superior to those of the MOE descriptor set-based model (ranging from 0.59 to 0.68). The MOE-based model also displayed lower MCC and G values indicating a poorer quality for classification and an improper balance between SE and SP. Accordingly, the selected descriptors from the Dragon and combined set were used to develop and validate initial LDA, QDA, and PLS-DA classification models.

### 3.3. Model development, performance, and validation

#### 3.3.1. Initial models

The initial binary classification models were developed using the LDA, QDA, and PLS-DA methods with the training set chemicals using the selected descriptors from Dragon and combined descriptor sets. We then examined which set of selected descriptors results in the most predictive model. The three initial models created using each set of descriptors predict toxicity similarly according to the LOO-CV test (Figure 2), although Dragon descriptor-based models were not as good as the combined descriptor-based models. The best balance between SE and SP (G) and the highest MCC value in the LOO-CV test were obtained for the LDA and PLS-DA models created with the combined descriptor set; showing G-mean ≥ 0.81 and MCC ≥ 0.62. In summary, the initial LDA and PLS-DA models using the combined descriptor set predicted compound toxicity better than the other initial models.

To further assess the robustness of the initial models, a fivefold CV (internal test sets) test was performed using the training set. Classification parameters calculated for the training set, LOO-CV, and fivefold CV for all initial models using the Dragon and combined descriptor sets are shown in Table I. The LDA and PLS-DA models created with the combined descriptor set display better fivefold CV classification parameters; e.g., these models generate a smaller number of false negatives and false positives after a fivefold CV leading to higher values of SE and SP. The two models showed similar classification quality (MCC values of 0.62 for LDA vs 0.60 for PLS-DA) and balance between SE and SP (G values of 0.81 for both models). Nevertheless the LDA model showed a slightly higher value for SE, an important parameter in classification models, because it correctly assigns more toxic chemicals as toxic (more true positives). Briefly, the results obtained for the training set, and LOO, and fivefold CV tests showed that the LDA and PLS-DA models built using the combined descriptor set outperformed the other models regarding classification quality (showing MCC values >60%) and balanced prediction (achieving G values >81%) (Table I).

We next searched for class specific outliers to refine the initial models (to discard irrelevant portions of information), decrease the complexity by reducing the number of descriptors (to increase interpretability), and improve predictivity. For this purpose we calculated the Mahalanobis distance separately for toxic and non-toxic classes. The Mahalanobis distance is a measure of the distance from the centroid whereby the larger a compound’s distance, the smaller is its probability of belonging to the set and the larger its likelihood of being an outlier. In total, 26 chemicals were identified as potential outliers (with probability values less than 0.05), of which 18 were toxic and 8 nontoxic (Supporting Information, Figure S2 and Table S1). The potential outliers were excluded from the data set followed by rebuilding each model and reassessment of the importance of each descriptor. The resultant models were compared with the initial

![Figure 1](image1.png) **Figure 1.** Principal component analysis calculated for the studied compounds (Environmental Protection Agency Phase I ToxCast chemical library and industrial high/low production volume chemicals). The first two principal components (PCs) are shown as a score plot of PC1 versus PC2 explaining 61% of variation.

![Figure 2](image2.png) **Figure 2.** Discrimination ability of the initial models between toxic and nontoxic compounds in terms of Matthews Correlation Coefficient (MCC) and geometric mean of sensitivity and specificity (G) derived from the training set and leave-one-out cross-validation (LOO-CV) test.
The classification results of the final models are listed in Table II. The classification parameters calculated for the reduced models are given in Table II. In some cases the model did not successfully recognize specific structural features even with reasonable similar compounds in the training set. For instance, compound MCPA (CAS: 94-74-6) was wrongly predicted as a toxicant and compounds 2-Phenylphenol (CAS: 90-43-7) and Methyl hydroxybenzoate (CAS: 64-31-2) were misclassified as nontoxic. To understand the reasons for these misclassifications, the Euclidean distance between each misclassified prediction set compound and the training set compounds was calculated. The purpose was to find out whether the final model failed to predict toxicity because of an inability to recognize structural features of the model or whether the training set lack compounds with similar characteristics as the misclassified prediction set compounds. The structures of the incorrectly predicted compounds along with the three most similar training set compounds are given in Table III. In some cases the model did not successfully recognize specific structural features even with reasonable similar compounds in the training set. For instance, compound d-cis,trans-Allethrin (CAS: 584-79-2), Primisulfuron-
Table III. Incorrectly predicted compounds, each grouped with the three most similar training set compounds

<table>
<thead>
<tr>
<th>False predictions</th>
<th>Training set compounds most similar to the misclassified compound</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CAS</strong></td>
<td><strong>Name</strong></td>
</tr>
<tr>
<td>94-74-6</td>
<td>MCPA</td>
</tr>
<tr>
<td>81777-89-1</td>
<td>Clomazone</td>
</tr>
<tr>
<td>111988-49-9</td>
<td>Thiacloprid</td>
</tr>
<tr>
<td>90-43-7</td>
<td>2-Phenylphenol</td>
</tr>
<tr>
<td>584-79-2</td>
<td>d-cis,trans-Allethrin</td>
</tr>
<tr>
<td>4376-18-5</td>
<td>Methyl hydrogen phthalate</td>
</tr>
</tbody>
</table>

(Continues)
methyl (CAS: 23031-36-9), and Thiacloprid (CAS: 11988-49-9), two out of the three similar training set compounds showed false predictions, which would mean improper training of the model and therefore, incorrect assignment of prediction set compounds. For example, if Tetramethrin (CAS: 7696-12-0) and Imazamox (CAS: 114311-32-9) were predicted as toxicants, then compounds d-cis,trans-Allethrin (CAS: 584-79-2) and Primisulfuron-methyl (CAS: 23031-36-9) would also be more likely to be predicted as toxic. In some cases, there were insufficient number of training set compounds structurally similar to a prediction set compound, e.g., Clomazone (CAS: 81777-89-1), making it difficult for the model to predict the true class of the prediction set compound.

The predictions of the final LDA model were compared with those of the publicly available VEGA model [30] using those ToxCast compounds falling inside the AD of the VEGA model (to assess the performance of the models using the same compounds being within the AD of both models). Details on the VEGA model can be found in the supporting information. In total only 52 (35 toxic and 17 nontoxic) compounds from the ToxCast data set fall inside the AD of the VEGA model, which indicates that VEGA was developed on a dataset with a different chemical domain as compared with ToxCast. Of these 52 compounds, VEGA correctly predicted 31 out of 35 as toxic (SE = 0.89) and only 5 out of 17 as nontoxic (SP = 0.29), while the final LDA model correctly predicted 27 toxic (SE = 0.77) and 15 nontoxic (SP = 0.88) chemicals (Supporting Information, Table S2). Because 38 out of the 52 chemicals were used in the training process of the final LDA model, a set of 14 chemicals belonging to the prediction set was considered as a better validation set of the final LDA model, which also showed a better agreement between SE and SP compared with the VEGA model (G of 0.73 vs 0.45). While a classification model’s SE (true classification rate of toxic class) reflects its ability to identify chemicals more likely to be toxic, a model’s SP (true classification rate of nontoxic class) reflects its ability to detect chemicals that are more likely nontoxic, which when removed from further studies saves time and cost.

It is worth noting that the developed model could not be reliably applicable to the H/LPVCs with very complex structures and high MWs such as decapeptides, e.g., Cyclosporin A (MW = 1203, CAS: 59865-13-3), Zoladex (MW = 1269, CAS: 65807-02-5), Cetrorelix (MW = 1431, CAS: 120287-85-6), and Ganirelix (MW = 1570, CAS: 124904-93-4), which were out of the developed model’s AD. The reason for being outside the AD could be their high MWs, which are out of the MW range of the training set compounds (ranging from 40.0 to 510.7). In addition, the molecular structures of some of these compounds include a large number of double bonds (for instance 21 and 27 double bonds for Cyclosporin A and Zoladex, respectively), which is more than the number of double bonds of the training set chemicals and could be another reason for them to be outside the AD. The other groups of industrial chemicals whose toxicity could not be truly predicted by the model are Monoazo pigments, exemplified by pigment yellow 97 (CAS: 12225-18-2) and pigment violet 32 (CAS: 12225-08-0), which are used, for example, for PVC coloring, long chain alcohols, such as 1-Dodecanol (CAS: 112-42-5), 1-Octadecene (CAS: 112-88-9), and polycyclic diones, such as Dibromopyranthrone (CAS: 1324-35-2) and Dichloroisovalianthrone (CAS: 1324-55-6).

### 3.4. Interpretation of model descriptors

The selected molecular descriptors (eight from Dragon and three from MOE) listed in Table IV take into account hydrophobicity, molecular polarity, branching, cyclicity, bond multiplicity, and
and MlogP reflects hydrophobicity[31]. The MOE descriptors are molecular descriptor corresponding to functional group counts bond multiplicity in a molecule [31,37]. Then Cconj is a simple chi matrix, which is an atom connectivity matrix accounting for [35,36]. SpPosA_X is a 2D matrix-based descriptor derived from matrix defined to account for heteroatoms and bond multiplicity value derived from the Burden matrix, a weighted adjacency matrix, which represents the whole set of connections between adjacent pairs of atoms giving information about branching [32–34]. These topological indices are weighted by dipole moments and are therefore sensitive to charge separation in a molecule. SpMin2_Bh(s) is a topochemical Burden eigenvalue derived from the Burden matrix, a weighted adjacency matrix defined to account for heteroatoms and bond multiplicity [35,36]. SpPosA_X is a 2D matrix-based descriptor derived from chi matrix, which is an atom connectivity matrix accounting for bond multiplicity in a molecule [31,37]. Then Cconj is a simple molecular descriptor corresponding to functional group counts and MlogP reflects hydrophobicity[31]. The MOE descriptors include VAdjMa from the atom counts and bond count group and a_don/a_count and rings/a_count, which represent the number of hydrogen bond donors and number of rings weighted by number of atoms, respectively [38].

To identify the most contributing descriptors the absolute difference between class means of each descriptor was calculated (Figure 3), with higher absolute differences indicating a stronger contribution to the developed LDA model. Thus, the most contributing descriptors were nDB, SpMin2_Bh(s), SM06_EA (dm), MlogP, and a_don/a_count. This observation agrees well with the results obtained from the PLS-DA loading plot (Figure S4, Supporting Information), where a_don/a_count is positively correlated with non-toxic and nDB, SpMin2_Bh(s), SM06_EA (dm), and MlogP are all correlated with toxic compounds. It should be noted that some of the selected descriptors, such as nDB, MlogP, and a_don/a_count, are easier to interpret and understand than complex indices such as SM06_EA (dm), which conveys a lot of information in a single number. However, the ability of the LDA model to distinguish toxic from non-toxic depends mainly on an aggregate of structural information derived from all selected descriptors combined. Therefore, using a single descriptor to describe toxicity is misleading.

Developmental toxicity in an embryo could be receptor-mediated whereby chemicals interact directly with an endogenous hormone or growth factor receptors [4]. Taking MlogP as one of the most contributing descriptors into consideration, it can be concluded that hydrophobic chemicals with a high MlogP value can easily pass through cell membranes and bind to receptors such as the estrogen receptor and the aryl hydrocarbon receptor. The ligand-receptor complex translocates to the nucleus where it may interact with DNA to activate or inactivate the expression of specific genes. It has been reported that the interaction of chemicals with these receptors may result in abnormal development [4]. Matrix-based descriptors, which correlate well with descriptors encoding chemical information related to branching, cyclicity, and molecular size [34], potentially reflect the size requirement at a receptor site.

### 3.5. Identification of potent H/LPVCs

To predict the toxicity of the H/LPVCs and to identify the potent ones we applied a 90% confidence level on the final LDA model to increase prediction performance (the confident final LDA

![Figure 3](image.png)

**Figure 3.** Absolute difference between class means for each descriptor.

<table>
<thead>
<tr>
<th>Selected descriptors</th>
<th>Abbreviation</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of double bonds</td>
<td>nDB</td>
<td>Dragon</td>
</tr>
<tr>
<td>Leading eigenvalue from edge adjacency matrix weighted by dipole moment</td>
<td>SpMax_EA(dm)</td>
<td>Dragon</td>
</tr>
<tr>
<td>Spectral diameter from augmented edge adjacency matrix weighted by dipole moment</td>
<td>SpDiam_AEA(dm)</td>
<td>Dragon</td>
</tr>
<tr>
<td>Spectral moment of order 6 from edge adjacency matrix weighted by dipole moment</td>
<td>SM06_EA(dm)</td>
<td>Dragon</td>
</tr>
<tr>
<td>The smallest eigenvalue of Burden matrix weighted by I-state</td>
<td>SpMin2_Bh(s)</td>
<td>Dragon</td>
</tr>
<tr>
<td>The normalized spectral positive sum from chi matrix</td>
<td>SpPosA_X</td>
<td>Dragon</td>
</tr>
<tr>
<td>Number of nonaromatic conjugated C (sp2)</td>
<td>nCconj</td>
<td>Dragon</td>
</tr>
<tr>
<td>Moriguchi octanol-water partition coefficient logP</td>
<td>MlogP</td>
<td>Dragon</td>
</tr>
<tr>
<td>Vertex adjacency information (magnitude)</td>
<td>VAdjMa</td>
<td>MOE</td>
</tr>
<tr>
<td>Number of hydrogen bond donor atoms weighted by the number of atoms</td>
<td>a_don/a_count</td>
<td>MOE</td>
</tr>
<tr>
<td>Number of rings weighted by the number of atoms</td>
<td>rings/a_count</td>
<td>MOE</td>
</tr>
</tbody>
</table>
model; Table II). Keeping in mind that a high probability for a compound indicates a high likelihood of toxicity, we set a probability of 90% as the final LDA model’s lower limit for classifying. Applying this limit on the ToxCast dataset, 67 chemicals (54 training and 13 prediction compounds) were predicted as developmental toxicants using the probability of ≥90% among which 51 and 12 chemicals were correctly classified in the training (SE = 0.94) and prediction (SE = 0.92) set, respectively. Furthermore, 42 chemicals were predicted as nontoxic using the probability of ≤10%, where 34 out of 36 compounds from the training set (SP = 0.94) and all six compounds from the prediction set (SP = 1.00) were truly predicted. The classification parameters for the confident final LDA model are shown in Table II. Compared with the final LDA model, the confident final LDA model’s classification parameter SE improved for both training (0.94 vs 0.88) and prediction (0.92 vs 0.85) sets. The improvement holds true for the other classification parameters; SP increased 19% and 17%, and G 13% and 14% for the training and prediction set, respectively, and AC 12% and MCC 33% for both training and prediction sets.

The confident final LDA predicted 608 H/LPVCs to cause developmental toxicity in zebrafish and 353 to be nontoxic. Among the 608 chemicals predicted as toxicants, 14 chemicals are listed as toxicants in TERIS and US FDA guidelines [13,14], 18 are listed as causing reproductive toxicity by the California Environmental Protection Agency’s Office of Environmental Health Hazard Assessment (OEHHA) [39], and 7 are listed in both sources. Out of the 608 chemicals 123 are within the AD of the VEGA model, and among those 112 were as well predicted to cause developmental toxicity by the VEGA model. Among these 112 chemicals, 23 compounds were listed in TERIS and US FDA guidelines and the OEHHA report (Table S4). Examples of predicted developmental toxicants are 2-naphthalenecarboxamide derivatives, such as 3-Hydroxy-N-(3-nitrophenyl)-2-naphthalenecarboxamide (CAS: 135-65-9) used in agrochemical, pharmaceutical, and dyestuff fields; phenothiazine derivatives such as 4-[3-(2-chloro-10H-phenothiazin-10-yl) propyl]-1-piperazineacetone (CAS: 58-39-9) used as intermediates in synthesis of pharmaceuticals and agrochemicals; benzodiazepine derivatives which are GABA modulators acting as anti-anxiety agents and sedative drugs, showing anxiolytic and anticonvulsant properties, such as 5-(2-fluorophenyl)-1,3-dihydro-1-methyl-7-nitro-2H-1, 4-benzodiazepin-2-one (rohypnol, CAS: 1622-62-4); benzamide derivatives, such as 5-chloro-N-(2-chloro-4-nitrophenyl)-2-hydroxybenzamide (Nicosamide CAS: 50-65-7), which is a salicylanilide compound with antihemorrhagic actions; benzimidazoles, such as 1-(1-(4-(4-fluorophenyl)-4-oxobutyl)-4-piperidinyl)-1,3-dihydro-2H-benzimidazol-2-one (CAS: 2062-84-2); and methanone derivatives applied as pharmaceutical intermediates, such as (2,4-difluorophenyl) phenyl-methanone (CAS: 85068-35-5). The 112 chemicals predicted as potential developmental toxicants by the confident final LDA model as well as the VEGA model could serve as a priority list for future developmental toxicity assessment of H/LPVCs (Table S4). These industrial chemicals, which have been in worldwide trade and commerce, are proposed to be experimentally assessed in developmental toxicity testing using zebrafish to ascertain whether they interfere with normal development. It is known that zebrafish share critical developmental processes with mammals, and thus, effects on zebrafish could be predictive for human health hazards. Data from the zebrafish developmental toxicity assay should thus be critically examined and used to trigger higher tier tests with, e.g., rodents or the replacement of hazardous chemicals with more benign alternatives.

4. CONCLUSION

In this study, new QSAR classification models were developed and validated, based on the OECD QSAR validation principles, to discriminate developmental toxic compounds from non-toxic ones in zebrafish using LDA, QDA, and PLS-DA methods and the ToxCast Phase I dataset. The final model showed that SE is comparable to the initial models for the training set, while it showed an improvement for the external SE. Applying a 90% confidence level on the final LDA model noticeably increased the performance of the predictive performance, showing an external SE of 0.92, SP of 1.00, and G of 0.96. The LDA model was applied to screen zebrafish developmental toxicity of over 7000 industrial chemicals, mostly without experimental data. Among the 112 H/LPVCs predicted as potential developmental toxicants by the LDA model and the VEGA model were 23 chemicals reported as toxicants in the TERIS and US FDA guidelines and the OEHHA report. According to structural information provided by the selected descriptors hydrophobicity and charge distribution were found to be influential properties on developmental toxicity in zebrafish. The proposed QSAR model developed based on high throughput screening data of 309 unique chemicals, could time- and cost-effectively be applied in further identification of hazardous chemicals regarding developmental toxicity as well as help predict developmental toxicity of newly synthesized compounds.

Acknowledgement

This study was financed by the Swedish Research Council (VR) (521-2011-6427) and the MISSE project through grants from the Swedish Research Council for the Environment, Agricultural Sciences and Spatial Planning (Formas) (210-2012-131).

REFERENCES

Developmental toxicity modeling


18. MOE (Molecular Operating Environment), 2008.10; Chemical Computing Group: Montreal, Quebec, Canada, 2012.


Additional supporting information can be found in the online version of this article at the publisher’s website.