



In silico prediction models for thyroid peroxidase inhibitors and their application to synthetic flavors

Mihyun Seo¹ · Changwon Lim² · Hoonjeong Kwon^{1,3}

Received: 13 October 2021 / Revised: 22 January 2022 / Accepted: 2 February 2022
© The Author(s) 2022

Abstract Systematic toxicity tests are often waived for the synthetic flavors as they are added in a very small amount in foods. However, their safety for some endpoints such as endocrine disruption should be concerned as they are likely to be active in low levels. In this case, structure–activity–relationship (SAR) models are good alternatives. In this study, therefore, binary, ternary, and quaternary prediction models were designed using simple or complex machine-learning methods. Overall, hard-voting classifiers outperformed other methods. The test scores for the best binary, ternary, and quaternary models were 0.6635, 0.5083, and 0.5217, respectively. Along with model development, some substructures including primary aromatic amine, (enol)ether, phenol, heterocyclic sulfur, and heterocyclic nitrogen, dominantly occurred in the most highly active compounds. The best predicting models were applied to synthetic flavors, and 22 agents appeared to have a strong inhibitory potential towards TPO activities.

Keywords Synthetic flavor · Quantitative structure–activity relationship (QSAR) · Thyroid peroxidase inhibitor (TPO) · Toxicity prediction · Machine learning

Introduction

Flavors are a type of food additives that are intentionally added to foods in order to enhance or fortify their original flavor. However, since flavors are used only in a small amount and their toxicological data are usually scarce, their safety is often assessed with an initial screening method without properly conducting systematic toxicity tests (MFDS, 2017). Among those methods, threshold of toxicological concern (TTC) method evaluates the safety of chemical compounds and sets an acceptable level of intake based on their structure and exposure levels (World Health Organization et al., 2016). The Joint FAO/WHO Expert Committee on Food Additives (JECFA) categorizes the flavors into three groups in accordance with Cramer class and evaluates their safety based upon TTC method. Cramer class was proposed by Cramer et al. in (1976), and it classifies the compounds into three classes based upon their potential of oral toxicity (Cramer et al., 1976). However, concerning that some endpoints such as endocrine disruptors could be active in low doses (Vandenberg et al., 2012; Vandenberg, 2014) and hormonal modulation is becoming a more important health issues nowadays, it is important to deal with those flavors having hormone-modulating activities even if they are used only in small amount. Thyroid hormones, for example, play an important role in many developmental process and regulate metabolic homeostasis (De Coster and van Larebeke, 2012), so the thyroid hormone modulation during developmental stage can lead to serious defect in neurogenesis as well as metabolic disturbance (Gore et al., 2015). Therefore, this study has attempted to conduct a quick screening of potential thyroid peroxidase (TPO)-inhibiting synthetic flavors using in silico prediction models.

✉ Changwon Lim
clim@cau.ac.kr

✉ Hoonjeong Kwon
hjkwon@snu.ac.kr

¹ Department of Food and Nutrition, Seoul National University, Seoul, Republic of Korea

² Department of Applied Statistics, Chung-Ang University, Seoul, Republic of Korea

³ Research Institute of Human Ecology, Seoul National University, Seoul, Republic of Korea

Currently, many studies and programs on machine-learning-based quantitative structure–activity relationship (QSAR) and *in silico* toxicity prediction were conducted (Fan et al., 2018; Idakwo et al., 2018; Jiang et al., 2019; Li et al., 2017; Zhang et al., 2018; Zhang et al., 2020). Results from high-throughput experiments may be employed to build a model. They are available from various databases, such as PubChem or EPA Chemistry Dashboard (Kim et al., 2019; Williams et al., 2017). Tox21, one of these databases, is a collaboration program of several federal agencies in US that aims to improve pre-existing testing strategies and develops many high-throughput screening methods that can be applied to toxicity prediction models (National Toxicology Program, 2020). In many QSAR practices, however, too many features compared to the size of the dataset, and class imbalance have been considered troublesome. Many attempts were made to address these problems, such as dimensionality reduction and under- or over-sampling. A previous study showed that the class-imbalance problem can be solved to some extent by the application of the so-called synthetic minority over-sampling technique (SMOTE) to low-dimensional models (Blagus and Lusa, 2013). In addition, learning methods also play a significant role in model performance. Simple but powerful learning methods, such as support vector machines (SVMs), random forests (RFs), and artificial neural networks (ANNs), demonstrated good performance and were used in many previous QSAR studies (Fan et al., 2018; Fan et al., 2018; Jiang et al., 2019; Li et al., 2014; Li et al., 2017). Recently, many ensemble-learning approaches were also introduced to improve the performance of models (Ai et al., 2019; Sheffield and Judson, 2019; Zhang et al., 2018).

Hence, in this study, prediction models for TPO inhibition were designed using various learning and dimensionality-reduction methods, and SMOTE. In addition, along with the over-sampling technique, multi-categorization of the dataset was attempted to improve the classification performance for minor classes, and their performance was compared with those of binary models. As well as the prediction model development, typical substructures frequently found in highly active compounds were analyzed using substructure frequency analysis (Jensen et al., 2007) and food-related compounds such as food additives or contact materials that feature active substructures were identified. The best models for each grouping method selected according to their test scores were applied to synthetic flavors currently used in South Korea. The flavors predicted by the models in this study to have strong inhibition effects towards TPO were then compared with their Cramer classes.

Materials and methods

Data curation and categorization

AC₅₀ and IC₅₀ values of various TPO inhibitors were collected from The Simmons Lab at the EPA National Center for Computational Toxicology's Amplex® Ultra-Red (AUR) assay data, and existing articles (Carvalho et al., 2000; Divi and Doerge, 1996; Habza-Kowalska et al., 2019; Lee, 2015). After removing salts, mixtures, and duplicated ones, 587 data items, including drugs, natural compounds, and environmental chemicals, were used to build machine-learning models. Given that AC₅₀ is a relative value that is calculated from each chemical's maximum inhibition rate, it had to be converted to the absolute IC₅₀ value, which is the concentration at which the chemicals inhibit TPO activities to 50% of maximal activity. The conversion is conducted via the following expression (Sebaugh, 2011):

$$\log \text{IC}_{50} = \log \text{AC}_{50} \times \left(\frac{a - 50\% \text{ response}}{50\% \text{ response} - b} \right)^{1/c} \quad (1)$$

a: min response; *b*: max response; *c*: slope factor.

Based on their maximum inhibition and the converted IC₅₀ values, the collected and purified data were categorized into two (binary), three (ternary), or four (quaternary) groups. For the binary model, the chemicals that inhibited TPO activities more than or equal to 20% at the highest concentration were defined as group 'A' (active) as maintained by EPA (Friedman et al., 2016), and the remaining chemicals whose maximum inhibition rate was less than 20% were labeled as group 'C' (inactive). For the ternary model, the chemicals with maximum inhibition rate higher than 50%—in which their IC₅₀ could be calculated, between 20 and 50%, and less than 20% were labeled as group 'A', 'B', and 'C', respectively. The chemicals in group 'A' in the ternary model were subdivided into group 'A1' and 'A2' in the quaternary model based on their IC₅₀ values; the chemicals whose IC₅₀ values were lower than or equal to 10 μM were defined as group 'A1', and those with IC₅₀ higher than 10 μM as group 'A2', which was based upon previous studies on enzyme inhibition (Auld et al., 2008; MFDS, 2013; Lindström et al., 2019).

Feature generation and structure

Molecular descriptors and fingerprints (FPs) were used as features in the machine-learning models. In this study, a 'topology-substructure concatenated FP' was employed to consider various features and prevent underfitting. It consists of a concatenation of each topological FP with a substructure key-based FP to consider the atomic

connectivity, substructures, and their interactions within a molecule. RDKit, Morgan, and Atom Pair Count (APC) FPs were selected for topological FP, and the substructure count and ToxPrint FP were selected for the substructure key-based FP. RDKit and Morgan FPs were calculated using the RDKit library (Landrum, 2020), and APC, substructure count FP, and descriptors were calculated with the PaDEL-Descriptor software (Yap, 2011). ToxPrint FP was generated using the ChemoTyper program (Mn-Am, 2020).

Simple and complex ensemble learning methods

Both simple learning methods (SLMs) and complex ensemble learning methods (ELMs) were used to predict TPO inhibitors. As SLMs, a RF, a SVM, and an ANN were used given that they demonstrated to perform well on QSAR tasks in previous studies. Boosting and voting were utilized for the ELMs; adaptive boosting (AdaB) and extreme gradient boosting (XGB) were employed for boosting; and hard- and soft-voting were used as voting models. In the case of voting classifiers, the four best models among RF, SVM, ANN, AdaB, and XGB were selected for each feature and grouping method. The combinations of the four best models selected for the voting classifiers for each feature and grouping method are listed in Table S1.

Model selection and evaluation

Through the overall process conducted in this study, the Scikit-Learn library was used for feature processing and machine-learning modeling (Pedregosa et al., 2011). The entire dataset was randomly split into an 80:20 ratio, and the 20% subset was used as a test set subsequently. The ‘stratify’ parameter was applied to retain the compositional consistency of each category in both training and test sets. The number of compounds in each category is presented in Table 1.

A grid search with fivefold cross-validation on the training dataset was conducted for hyperparameter tuning and model validation and to prevent overfitting. The hyperparameter grid of each learning method is given in Table S2. Min–max feature normalization, variance thresholder, feature extraction for dimensionality reduction, SMOTE, and model training steps were implemented for each iteration of the fivefold cross-validation. Dimensionality reduction was conducted to solve the overfitting problem that may occur due to the large number of features. The features whose variance was less than 0.01 were removed. Either principal component analysis (PCA) or linear discriminant analysis (LDA) were applied, and their performance was compared. To address the class-imbalance problem, SMOTE was applied to each model. Through this process, the best combination of hyperparameters for each model was selected, and cross-validation scores (CV scores) were calculated as a F1-score value. F1 score, i.e., the harmonic mean of precision and recall, was employed owing to the class-imbalance problem. The formulae of precision, recall, and F1 score are as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$F1 \text{ score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2)$$

[TP: True positive; FP: False positive; FN: False negative]

The best models for each grouping and FP were selected based on the CV score, and their performance was evaluated by applying the models on the test sets. Then, the best models that produced the highest F1 score on the test set was singled out from each grouping as the final best-performing models. All combinations of models, FPs, and dimensionality-reduction methods (or voting methods) for each grouping method are listed in Table S3. After carrying

Table 1 Number of compounds in each subset and their classification criteria

Max inhibition	IC ₅₀	Binary	Number of compounds		Ternary	Number of compounds		Quaternary	Number of compounds	
Higher than 50%	≤ 10 μM	A	Train	388	A	Train	283	A1	Train	100
			Test			Test	71		Test	25
	> 10 μM		Train	97	B	Train	105	B	Train	105
			Test			Test	26		Test	26
Between 20 and 50%		C	Train	81	C	Train	81	C	Train	81
			Test	21		Test	21		Test	21

out model selection and evaluation, the effects of the type of features, feature extraction method, and learning method on model performance were evaluated and compared.

Analysis of active substructures in TPO inhibitors

The substructures that were frequently seen in the most active groups were analyzed using the substructure frequency analysis method previously introduced by Jensen et al. (2002). The frequency of a substructure in group A1 (the most highly active) and C (inactive) in the quaternary grouping was calculated as $\frac{f_x}{C} = \frac{f_x \times C}{f \times C_x}$, where f_x and C_x refer to the number of fragments and compounds in group x , respectively, and f and C refer to the number of fragments and compounds in the whole dataset, respectively. A substructure with a frequency greater than 1.2 in group A1, and simultaneously with a ratio of the frequency of group A1 to that of group C greater than 1.2, was defined as an active substructure.

Application to flavors

Among 2465 synthetic flavors listed in the Ministry of Food and Drug Safety (MFDS)'s Food Additive Code, 1774 compounds available in the EPA's CompTox Chemistry Dashboard database were obtained, excluding salts and mixtures. The best classification models selected were applied to these flavoring agents for screening. The Toxtree software (Patlewicz et al., 2008) was employed to classify selected flavor compounds which showed high TPO inhibitory activity, according to Cramer.

Results and discussion

Model evaluation and selection

All the models were evaluated in terms of their test scores. The APC_Sub FP was the best-performing FP, whereas the best-performing models in the binary, ternary, and quaternary models were hard-voting, XGB with LDA, and soft-voting classifier, respectively. The test scores for the aforementioned best models were 0.6635, 0.5083, and 0.5217, respectively (Fig. 1).

In previous studies, Morgan FP is generally known to be one of the best-performing FPs (Idakwo et al., 2018; Riniker and Landrum, 2013). Morgan FP, however, hardly recognizes comprehensive characteristics such as molecular shape or size, and fails to perceive constitutional differences between isomers in large molecules. By contrast, Atom Pair FPs demonstrated to recognize the molecular shape and distinguish the constitutional differences better

(Capecchi et al., 2020). Therefore, in this study, APC_Sub FP performed well for molecular shape, substructures and their interactions were better recognized in APC_Sub FP than in Morgan_Sub FP.

To compare the performance of the models by feature extraction (PCA or LDA) or learning methods (RF, SVM, ANN, AdaB, XGB, or Voting), the CV and test scores of each feature extraction or learning methods were pooled and averaged to evaluate each method. For the feature extraction methods, there was no remarkable difference in the performance between the models using PCA and LDA, which is contrary to the expectation that LDA would outperform PCA when it comes to the classification model (Figure S1). Although LDA is generally considered to be a better feature extraction method for classification tasks, PCA may outperform LDA when the dataset is small (Martinez and Kak, 2001). Additionally, it can be inferred that LDA did not perform much better than PCA, given that the categorization of toxicity in this study was done from continuous maximum inhibition (%) and IC₅₀ values. Therefore, it seems that no significant difference in performance between LDA and PCA was observed because of the small size of the dataset and the linear characteristic of toxicity class data.

Meanwhile, learning methods were shown to have a relatively greater impact on model performance than feature extraction methods (Figure S2).

For all groupings, hard-voting classifiers remarkably improved the model performance, especially in the ternary and quaternary models. In the case of the ternary model, the hard-voting classifier showed approximately 16, 15, and 14% enhancements in average model performance compared to SLMs–RF, SVM, and ANN–respectively. For quaternary models, the hard-voting classifiers showed 15, 13, and 14% improvement with respect to the aforementioned SLMs–RF, SVM, and ANN–respectively. Among all the models, the worst-performing learning method was the AdaB classifier. For the binary, ternary, and quaternary models, the model performance of the hard-voting classifier was 13, 23, and 31% greater than that of AdaB, respectively. Compared to a previous study that has developed binary models for TPO inhibitors (Rosenberg et al., 2017), the result shows that the binary models in this study have similar levels of F1 scores to those of previous study by employing hard voting classifiers, in spite of smaller dataset.

Along with F1 scores, confusion matrices for the best-performing models indicated in Fig. 1 were generated to evaluate the classification performance in detail (Fig. 2). Based on confusion matrices, negative data were better classified in the ternary and quaternary models, even though the test score was the highest in the binary model. The specificities of the best binary, ternary, and quaternary

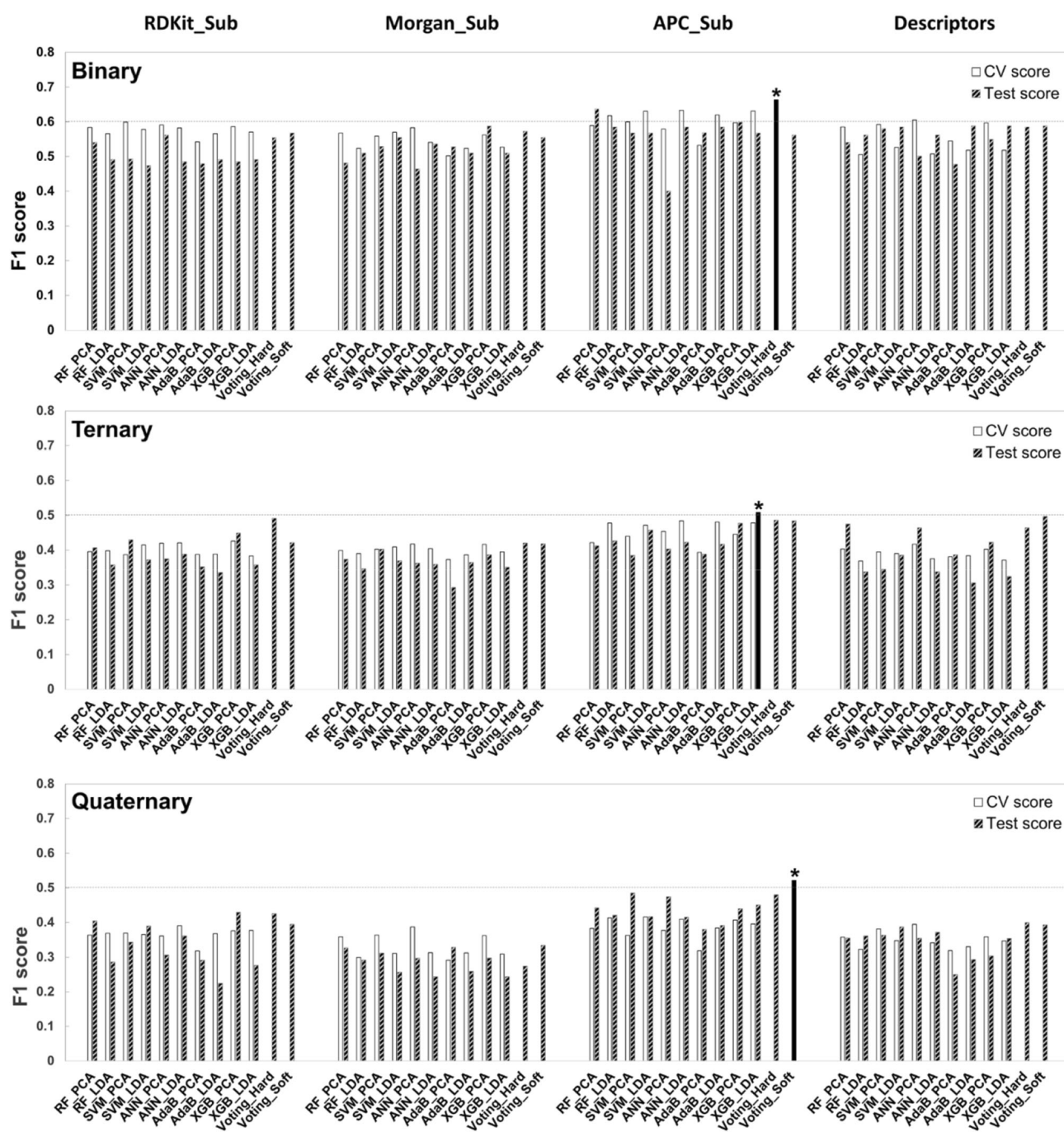


Fig. 1 Cross-validation (CV) and test scores for each feature, learning method, and grouping method (binary, ternary, or quaternary). For each grouping method, the black bar marked above with an asterisk indicates the test scores of the best-performing model. Each label is shown in “model name_feature extraction method (or voting

methods in voting classifier” form. [RF: Random forest; SVM: Support vector machine; ANN: Artificial neural network; AdaB: Adaptive boosting; XGB: Extreme gradient boosting; PCA: Principal component analysis; LDA: Linear discriminant analysis]

models were 0.3810, 0.5238, and 0.6190, respectively. Furthermore, continuous toxicity data such as IC₅₀ would allow more precise prediction of toxicity strength of substances, but they are more difficult to be obtained in large

numbers. Therefore, this work is significant in that multi-classification models were developed with small numbers of IC₅₀ data by adopting ensemble learning methods.

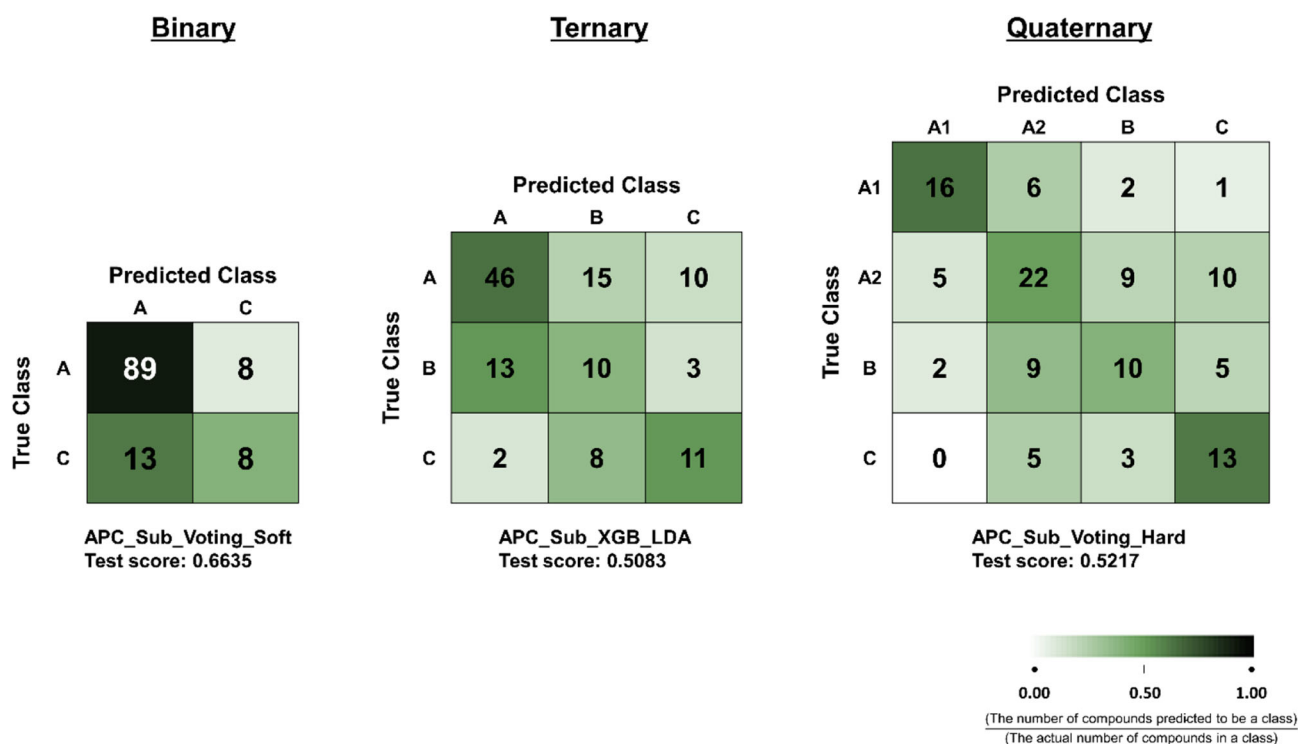


Fig. 2 Confusion matrices for the best-performing models of each grouping. The color intensity of each cell represents the proportion of the number of compounds predicted to be a class with respect to the actual number of compounds in a class

Active substructure analysis

An active substructure analysis was conducted using the substructure frequency analysis method. The active substructures included amines (especially primary aromatic amines and heterocyclic nitrogen), sulfur-containing compounds (carbothioic S-ester, thioether, thiocyanate, sulfenic derivatives, carbodithioic ester, heterocyclic sulfur, etc.), phenols, ethers (enoether and thioether), and vinylogous compounds. The most frequent and dominant substructure was primary aromatic amine with a frequency ratio of 35.90, meaning that it appears 35.90 times more frequently in group A1 than in group C. The inhibition mechanism of aromatic amines towards TPO was previously demonstrated by Doerge et al. (1994) The aromatic amines inhibit TPO activity by interacting with the compound I (an oxyferryl cation radical of iron-containing heme cofactor) instead of iodine, i.e., an endogenous substrate of TPO (Doerge and Decker, 1994).

Other frequent substructures were enol (enoether), sulfenic derivatives, and phenols, for which the frequency ratios were 9.79 (8.16), 8.16, and 4.45, respectively. Some examples of active compounds in group A1 and their corresponding active substructures are listed in Table 2. Among these highly potential compounds, food-related chemicals can be pointed out, including natural food substances, food additives or contact materials, and pesticides.

L-ascorbic acid, L-tryptophan, and polyphenols such as quercetin and genistein are typical examples of natural food substances that inhibit TPO activities. Food additives or contact materials such as indole (synthetic flavor) or 4,4'-methylenedianiline (food contact material) were also found to be highly active. In addition, some sulfur-containing or organophosphorus pesticides were included in the A1 group. While the frequency of phosphate substructures in group A1 was less than 1.2, their frequency in group A2 was 1.34. Furthermore, the frequency ratios of groups A1 and A2 to group C were 2.72 and 3.14, respectively. Hence, it can be stated that phosphate substructures in organophosphorus pesticides could also be considered as active substructures.

Application to flavors

In addition to the substructure analysis, 1774 synthetic flavors currently listed in South Korean Food Code were applied to the three best performing models. Before the application, the chemical spaces of these synthetic flavors and of the dataset used for model development were analyzed with three principal components extracted from their molecular descriptors (Fig. 3). Most flavor compounds showed a comparable chemical space with that of the dataset employed in this study.

Table 2 Examples of compounds in the highest activity group (group A1) shown with active substructures studied by means of substructure frequency analysis (depicted in red)

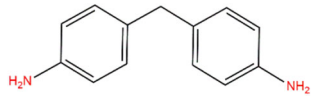
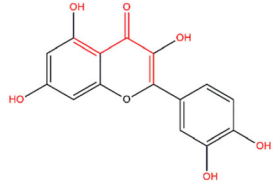
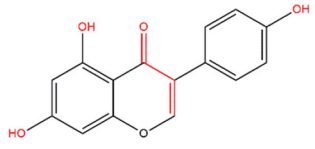
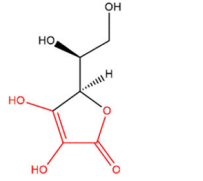
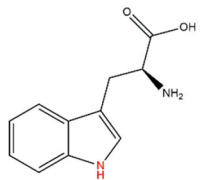
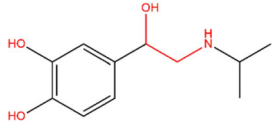
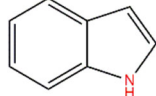
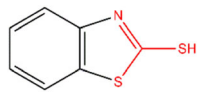
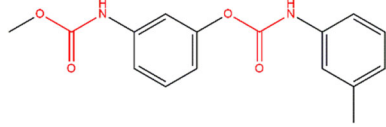
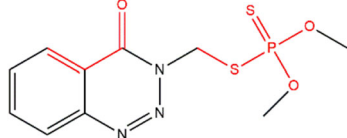
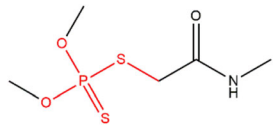
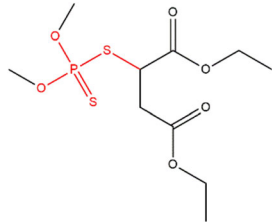
Name of compounds	Structures of compounds and corresponding active substructure	Purpose of use or distribution in nature	Ref
4,4'-methylenedianiline		Food contact material	(U.S. Food and Drug Administration, 2020a)
Quercetin		A flavonoid (flavonol) in edible plants (e.g., onion)	(Chandra, 2010)
Genistein		A flavonoid (isoflavone) in edible plants (e.g., soybean)	(Chandra, 2010)
L-ascorbic acid		Natural food substance; food additive	(U.S. Food and Drug Administration, 2020b)
L-Tryptophan		Natural food substance (amino acid)	(U.S. Food and Drug Administration, 2020c)
Isoproterenol		Drug; β -adrenergic receptor agonist (analogue of epinephrine)	(U.S. National Library of Medicine, 2020)
Indole		Food additive (flavoring agent)	(Food and Agriculture Organization of the United States (FAO), 2020)
2-Mercaptobenzothiazole		Pesticide	(United States Environmental Protection Agency (EPA), 2020)
Phenmedipham		Pesticide	(European Commission, 2020)
Azinphos-methyl		Pesticide	(European Commission, 2020)
Dimethoate		Pesticide	(European Commission, 2020)

Table 2 continued

Name of compounds	Structures of compounds and corresponding active substructure	Purpose of use or distribution in nature	Ref
Malathion		Pesticide	(European Commission, 2020)

Chemical Space Analysis

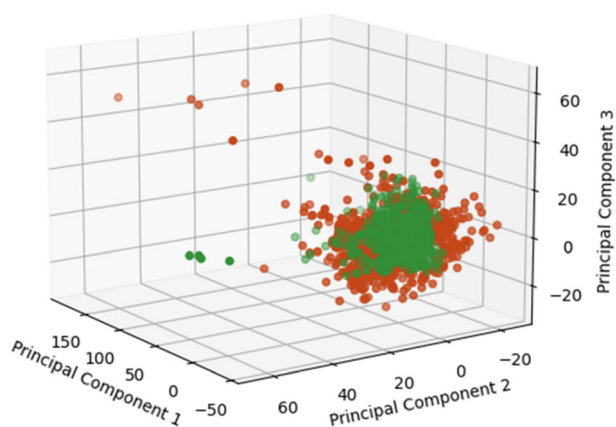


Fig. 3 Chemical space analysis between the input data for model development and synthetic flavors (red dots: synthetic flavors; green dots: input data for model development)

As a result, 22 out of 1774 were found to be in the most active group both in the ternary (group A) and quaternary (group A1) models and active in the binary model (group A). The molecular structures of these 22 compounds include primary aromatic amines, vinylogous carbonyls, and sulfur-containing substances (Table 3). These synthetic flavors, predicted to be highly active by the models, were compared with their Cramer classes predicted by the Toxtree software.

As synthetic flavors are added to foods only in a small amount, they are often regulated with threshold of toxicological concern (TTC) method, which is based on the chemical structures of substances. In JECFA, synthetic flavors are categorized into one of the structural classes proposed by Cramer et al. (1976) and their structures, metabolic fate and intake levels are assessed (Joint FAO/WHO Expert Committee on Food Additives, 2002). Cramer decision tree classifies compounds into three groups—Class I (low concern of oral toxicity), Class II (moderate concern of oral toxicity), and Class III (high concern of oral toxicity), according to the predicted intensity of their

oral toxicity (Cramer et al., 1976; Roberts et al., 2015). When the Cramer class of these 22 active flavor compounds was determined using the Toxtree software, 16 of them were classified into Class III, while 6 were classified into Class I. These six substances were butyl anthranilate, linalyl anthranilate, geranyl tiglate, allyl ionone, cedr-8(15)-en-4-ol, and 2'-aminoacetophenone. Aniline substructures stand out among the compounds that show high TPO inhibitory potential but low overall oral toxicity potential. The TPO-inhibiting effect and mechanism of action of substances with aromatic amines are clearly reported (Doerge et al., 1994). Moreover, as it is known that endogenous hormones work at very low concentrations in the body, even a slight modulation in the endocrine system in their low-dose range might significantly change the hormonal effect (Vandenberg et al., 2012).

Although the prediction models in this study are based on *in vitro* experimental data, it may be necessary to ameliorate Cramer's decision tree with recent toxicity result in light of the current situation that hormonal modulation is becoming a more important health issue. Further, in order to support the findings of this study, a follow-up study is necessary that determines whether the 22 flavors actually have TPO-inhibition activity through a laboratory experiment.

Summing up, the present study raises the possibility that some synthetic flavors may have a health effect on TPO activities through *in silico* toxicity prediction models. Binary, ternary, and quaternary prediction models were designed using various machine-learning methods. The best models for the binary, ternary, and quaternary models were 'hard-voting classifier with APC_Sub FP', 'XGB with APC_Sub FP and LDA', and 'soft-voting classifier with APC_Sub FP', respectively, and the F1 scores on the test set for each best model were 0.6635, 0.5083, and 0.5217, respectively. Although the test score was the highest in the binary model, the minor class (group C) was better predicted in the ternary and quaternary models. However, it should be concerned that these prediction models are based

Table 3 Synthetic flavors found to be active in the predictive models and their corresponding Cramer classes predicted by the Toxtree software

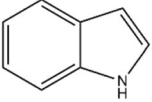
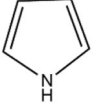
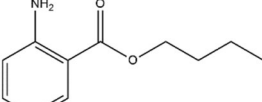
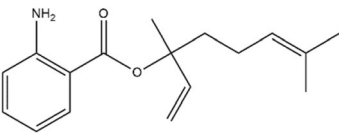
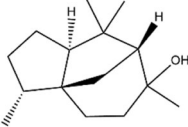
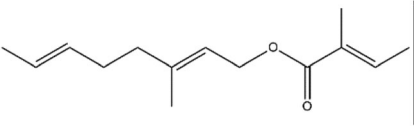
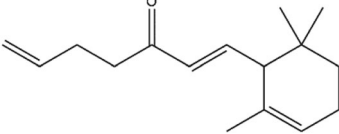
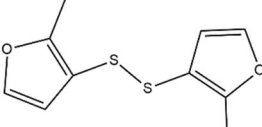
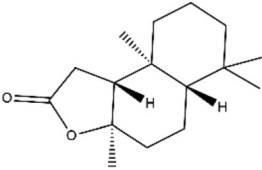
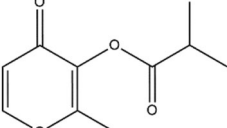
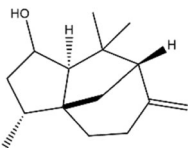
Name of compound	Structure	Cramer class ^a		
		Class I	Class II	Class III
Indole				✓
Pyrrole				✓
Butyl anthranilate		✓		
Linalyl anthranilate		✓		
(+)-Cedrol				✓
Geranyl tiglate		✓		
Allyl Ionone		✓		
Bis(2-methyl-3-furyl)disulfide				✓
(3aR)-(+)-Sclareolide				✓
Maltol isobutyrate				✓
Cedr-8(15)-en-4-ol		✓		

Table 3 continued

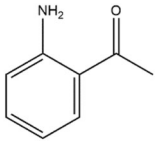
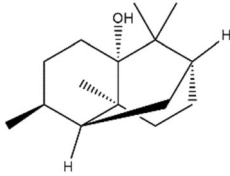
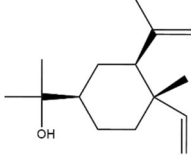
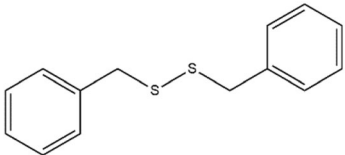
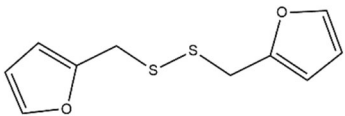
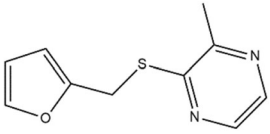
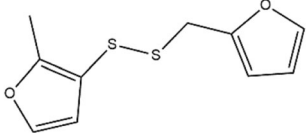
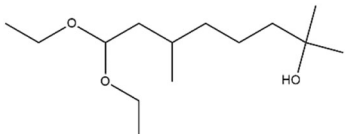
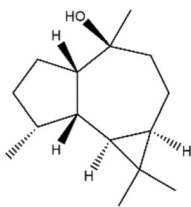
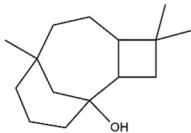
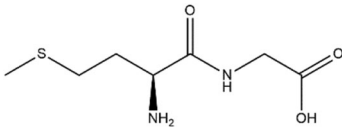
Name of compound	Structure	Cramer class ^a		
		Class I	Class II	Class III
2'-Aminoacetophenone		✓		
Patchouli alcohol				✓
Elemol				✓
Dibenzyl disulfide				✓
Difurfuryl disulfide				✓
2-Furfurylthio-3-methylpyrazine				✓
Furfuryl-2-methyl-3-furyl disulfide				✓
8,8-diethoxy-2,6-dimethyl-2-Octanol				✓
Viridiflorol				✓
Caryolan-1-ol				✓

Table 3 continued

Name of compound	Structure	Cramer class ^a		
		Class I	Class II	Class III
L-Methionylglycine				✓

^aCramer classes are defined as follows:

Class I: Chemical structures with small potential of oral toxicity (low class)

Class II: Chemical structures that are more potentially toxic than Class I substances, but without typical structures suggestive of toxicity (intermediate class)

Class III: Chemical structures that have no solid evidence of safety or may even have functional groups that suggest strong toxicity or reactivity (high class)

on in vitro experimental data, which hardly considers toxicokinetics of chemicals in the body.

The most frequent and dominant substructures within the highly active compounds (group A1) were primary aromatic amines, sulfur-containing, phenols, vinylogous compounds, and phosphates. When the best models selected from each grouping method were applied to 1774 synthetic flavors listed in South Korea, 22 out of 1774 agents were predicted to show inhibitory activity toward TPO. Sixteen out of 22 substances belonged to Cramer Class III, while six were Class I. Among these 6 compounds, which were predicted to have high TPO inhibitory activity but classified as low oral toxicity in Cramer class, three compounds had aniline substructures, butyl anthranilate, linalyl anthranilate, and 2'-aminoacetophenone, suggesting a revision of Cramer class to encompass hormonal modulation.

To suggest, follow-up study on the prediction model that involves the toxicokinetics of the TPO-inhibiting substances and on the determination of the 22 active flavors through actual laboratory experiments will support the major findings in this study.

Declarations

Conflict of interest The authors declare no conflicts of interest.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10068-022-01041-y>.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless

indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ai H, Wu X, Zhang L, Qi M, Zhao Y, Zhao Q, Zhao J, Liu H. QSAR modelling study of the bioconcentration factor and toxicity of organic compounds to aquatic organisms using machine learning and ensemble methods. *Ecotoxicology and Environmental Safety*. 179: 71-78 (2019)
- Auld DS, Southall NT, Jadhav A, Johnson RL, Diller DJ, Simeonov A, Austin CP, Inglese J. Characterization of chemical libraries for luciferase inhibitory activity. *Journal of Medical Chemistry*. 51: 2372-2386 (2008)
- Blagus R, Lusa L. SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics*. 14: 106. (2013)
- Capecchi A, Probst D, Reymond J. One molecular fingerprint to rule them all: drugs, biomolecules, and the metabolome. *Journal of Cheminformatics*. 12: 43 (2020)
- Carvalho DP, Ferreira AC, Coelho SM, Moraes JM, Camacho MA, Rosenthal D. Thyroid peroxidase activity is inhibited by amino acids. *Brazilian Journal of Medical and Biological Research*. 33(3): 355-361 (2000)
- Chandra AK. Chapter 42 - Goitrogen in food: Cyanogenic and flavonoids containing plant foods in the development of goiter. In: *Bioactive foods in promoting health*. Ronald RW, Victor RP. (Eds). Academic Press. San Diego (2010) pp. 691-716
- De Coster S, van Larebeke N. Endocrine-disrupting chemicals: associated disorders and mechanisms of action. *Journal of Environmental and Public Health*. 1687-9813 (Electronic) (2012)
- Cramer GM, Ford RA, Hall RL. Estimation of toxic hazard—A decision tree approach. *Food and Cosmetics Toxicology*. 16: 255-276 (1976)
- Divi RL, Doerge DR. Inhibition of thyroid peroxidase by dietary flavonoids. *Chemical Research in Toxicology*. 9: 16-23 (1996)
- Doerge DR, Decker CJ. Inhibition of peroxidase-catalyzed reactions by arylamines: mechanism for the anti-thyroid action of

- sulfamethazine. *Chemical Research in Toxicology*. 7: 164-169 (1994)
- European Commission. EU Pesticide Database. Available from: https://ec.europa.eu/food/plant/pesticides/eu-pesticides-database/public/?event=activesubstance.detail&language=EN&selecte_dID=914. Accessed Oct 29 2020.
- Fan D, Yang H, Li F, Sun L, Di P, Li W, Tang Y, Liu G. In silico prediction of chemical genotoxicity using machine learning methods and structural alerts. *Toxicological Research*. 7: 211-220 (2018)
- Fan T, Sun G, Zhao L, Cui X, Zhong R. QSAR and Classification Study on Prediction of Acute Oral Toxicity of N-Nitroso Compounds. *International Journal of Molecular Sciences*. 19:3015 (2018)
- Food and Agriculture Organization of the United States (FAO). Food safety and quality. Online Edition: "Specification for Flavourings". Available from <http://www.fao.org/food/food-safety-quality/scientific-advice/jecfa/jecfa-flav/details/en/c/1185/>. Accessed Oct. 29, 2020.
- Friedman KP, Watt ED, Hornung MW, Hedge JM, Judson RS, Crofton KM, Houck KA, Simmons SO. Tiered high-throughput screening approach to identify thyroperoxidase inhibitors within the ToxCast Phase I and II chemical libraries. *Toxicological Sciences*. 151: 160-180 (2016)
- Gore AC, Chappell VA, Fenton SE, Flaws JA, Nadal A, Prins GS, Toppari J, Zoeller RT. EDC-2: The Endocrine Society's Second Scientific Statement on Endocrine-Disrupting Chemicals. *Endocrine reviews*. 36, E1-E150 (2015)
- Habza-Kowalska E, Kaczor AA, Żuk J, Matosiuk D, Gawlik-Dziki U. Thyroid Peroxidase Activity is Inhibited by Phenolic Compounds-Impact of Interaction. *Molecules*. 24: 2766 (2019)
- Idakwo G, Luttrell J, Chen M, Hong H, Zhou Z, Gong P, Zhang C. A review on machine learning methods for in silico toxicity prediction. *Journal of Environmental Science and Health, Part C*. 36: 169-191 (2018)
- Jensen BF, Vind C, Padkjær SB, Brockhoff PB, Refsgaard HHF. In Silico Prediction of Cytochrome P450 2D6 and 3A4 Inhibition Using Gaussian Kernel Weighted k-Nearest Neighbor and Extended Connectivity Fingerprints, Including Structural Fragment Analysis of Inhibitors versus Noninhibitors. *Journal of Medicinal Chemistry*. 50: 501-511 (2007)
- Jiang C, Yang H, Di P, Li W, Tang Y, Liu G. In silico prediction of chemical reproductive toxicity using machine learning. *Journal of Applied Toxicology*. 39: 844-854. (2019)
- Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, Zaslavsky L, Zhang J, Bolton EE. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Research*. 47(D1): D1388-D1395 (2019)
- Landrum GA. RDKit: Open-Source Cheminformatics. Available from : <https://www.rdkit.org>. Accessed Nov. 11, 2020.
- Lee J. Conversion of the organic breakdown products of glucosinolate to thiocyanate anions and their effects on thyroid hormone production. PhD thesis, Seoul National University, Seoul, South Korea (2015)
- Li X, Chen L, Cheng F, Wu Z, Bian H, Xu C, Li W, Liu G, Shen X, Tang Y. In silico prediction of chemical acute oral toxicity using multi-classification methods. *Journal of Chemical Information and Modeling*. 54: 1061-1069 (2014)
- Li F, Fan D, Wang H, Yang H, Li W, Tang Y, Liu G. In silico prediction of pesticide aquatic toxicity with chemical category approaches. *Toxicological Research*. 6: 831-842 (2017)
- Lindström H, Mazari AMA, Musdal Y, Mannervik B. Potent inhibitors of equine steroid isomerase EcaGST A3-3. *PLoS One*. 14: e0214160 (2019)
- Martinez AM, Kak AC. PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 23: 228-233 (2001)
- MFDS. Guideline for drug metabolism evaluation. (2013)
- MFDS. Principles of establishing standards for foods. (2017)
- Mn-Am. ChemoTyper: Chemotype Your Molecular Datasets. Available from <https://www.mn-am.com/products/chemotyper>. Accessed Nov. 11, 2020
- National Toxicology Program. What We Study: Tox21. Available from : <https://ntp.niehs.nih.gov/whatwestudy/tox21/index.html>. Accessed Oct. 12, 2020.
- Patlewicz G, Jeliaskova N, Safford RJ, Worth AP, Aleksiev B. An evaluation of the implementation of the Cramer classification scheme in the Toxtree software. *SAR and QSAR in Environmental Research*. 19: 495-524 (2008)
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*. 12:2825–2830 (2011)
- Riniker S, Landrum GA. Open-source platform to benchmark fingerprints for ligand-based virtual screening. *Journal of Cheminformatics*. 5: 26 (2013)
- Roberts DW, Aptula A, Schultz TW, Shen J, Api AM, Bhatia S, Kromidas L. A practical guidance for Cramer class determination. *Regulatory Toxicology and Pharmacology*. 73: 971-984 (2015)
- Rosenberg SA, Watt ED, Judson RS, Simmons S, Friedman KP, Dybdahl M, Nikolov NG, Wedebye EB. QSAR models for thyroperoxidase inhibition and screening of US and EU chemical inventories. *Computational Toxicology*. 4: 11-21 (2017)
- Sebaugh JL. Guidelines for accurate EC50/IC50 estimation. *Pharmaceutical Statistics*. 10: 128-134 (2011)
- Sheffield TY, Judson RS. Ensemble QSAR Modeling to Predict Multispecies Fish Toxicity Lethal Concentrations and Points of Departure. *Environmental Science & Technology*. 53: 12793-12802 (2019)
- The Joint FAO/WHO Expert Committee on Food Additives (JECFA). Safety evaluation of certain food additives and contaminants / prepared by the fifty-seventh meeting of the Joint FAO/WHO Expert Committee on Food Additives. 57th Meeting, Rome, Italy. pp.131-133 (2002)
- U.S. Food and Drug Administration. Food Ingredient and Packaging Inventories. Indirect Additives Used in Food Contact Substances. Available from: <https://www.cfsanappsexternal.fda.gov/scripts/fdcc/index.cfm?set=IndirectAdditives&id=METHYLENEDIANILINE>. Accessed Oct 30 2020a.
- U.S. Food and Drug Administration. Food Ingredients & Packaging, Food Additives & Petitions. Food Additive Status List. Available from: <https://www.fda.gov/food/food-additives-petitions/food-additive-status-list>. Accessed Oct 30 2020b.
- U.S. Food and Drug Administration. Ingredients and Packaging, Food Ingredient and Packaging Inventories. Substances Added to Food (formerly EAFUS). Available from: <https://www.cfsanappsexternal.fda.gov/scripts/fdcc/index.cfm?set=FoodSubstances&id=TRYPTOPHAN>. Accessed Oct 30 2020c.
- U.S. National Library of Medicine. DailyMed. ISUPREL- isoproterenol hydrochloride injection, solution. Available from: <https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=a58f95fa-18f9-4938-8e3f-f33244f68298>. Accessed Oct 29 2020d.
- United States Environmental Protection Agency (EPA). OPP Pesticide Ecotoxicity Database. Available from: <https://ecotox.ipmcenters.org/details.cfm?recordID=6710>. Accessed Nov. 2, 2020.
- Vandenberg LN. Low-dose effects of hormones and endocrine disruptors. *Vitamins and hormones*. 94: 129-165 (2014)

- Vandenberg LN, Colborn T, Hayes TB, Heindel JJ, Jacobs DR Jr, Lee DH, Shioda T, Soto AM, vom Saal FS, Welshons WV, Zoeller RT, Myers JP. Hormones and endocrine-disrupting chemicals: low-dose effects and nonmonotonic dose responses. *Endocrine reviews*. 33: 378-455 (2012)
- Williams AJ, Grulke CM, Edwards J, McEachran AD, Mansouri K, Baker NC, Patlewicz G, Shah I, Wambaugh JF, Judson RS, Richard AM. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. *Journal of Cheminformatics*. 9: 61 (2017)
- World Health Organization (WHO), Food and Agriculture Organization of the United Nations (FAO), The Joint FAO/WHO Expert Committee on Food Additives (JECFA). Evaluation of certain food additives. In: Eighty-Second Report of the Joint FAO/WHO Expert Committee on Food Additives. WHO Technical Report Series; 1000. 82nd Meeting, Geneva, Switzerland (2016)
- Yap CW. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *Journal of Computational Chemistry*. 32: 1466-1474 (2011)
- Zhang L, Zhang H, Ai H, Hu H, Li S, Zhao J, Liu H. Applications of Machine Learning Methods in Drug Toxicity Prediction. *Current Topics in Medicinal Chemistry*. 18: 987-997 (2018)
- Zhang H, Mao J, Qi HZ, Ding L. In silico prediction of drug-induced developmental toxicity by using machine learning approaches. *Molecular Diversity*. 24: 1281-1290 (2020)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.