Food and Chemical Toxicology 125 (2019) 361–369

Contents lists available at ScienceDirect

Food and Chemical Toxicology

journal homepage: www.elsevier.com/locate/foodchemtox

Applications of *in silico* methods to analyze the toxicity and estrogen receptor-mediated properties of plant-derived phytochemicals

K. Kranthi Kumar[a], P. Yugandhar[b], B. Uma Devi[a], T. Siva Kumar[a], N. Savithramma[b], P. Neeraja[a,*]

[a] *Department of Zoology, Sri Venkateswara University, Tirupati, 517502, India*
[b] *Department of Botany, Sri Venkateswara University, Tirupati, 517502, India*

ARTICLE INFO

Keywords:
Phytochemicals
QSAR modeling
Toxicity
Nuclear hormone receptor binding
Self – Organizing maps
Clustering and classification schemes

ABSTRACT

A myriad of phytochemicals may have potential to lead toxicity and endocrine disruption effects by interfering with nuclear hormone receptors. In this examination, the toxicity and estrogen receptor – binding abilities of a set of 2826 phytochemicals were evaluated. The endpoints mutagenicity, carcinogenicity (both CAESAR and ISS models), developmental toxicity, skin sensitization and estrogen receptor relative binding affinity (ER_RBA) were studied using the VEGA QSAR modeling package. Alongside the predictions, models were providing possible information for applicability domains and most similar compounds as similarity sets from their training sets. This information was subjected to perform the clustering and classification of chemicals using Self – Organizing Maps. The identified clusters and their respective indicators were considered as potential hotspot structures for the specified data set analysis. Molecular screening interpretations of models were exhibited accurate predictions. Moreover, the indication sets were defined significant clusters and cluster indicators with probable prediction labels (precision). Accordingly, developed QSAR models showed good predictive abilities and robustness, which observed from applicability domains, representation spaces, clustering and classification schemes. Furthermore, the designed new path could be useful as a valuable approach to determine toxicity levels of phytochemicals and other environmental pollutants and protect the human health.

1. Introduction

Phytochemicals are bioactive non – nutritive compounds of plants. They provide a wide range of antioxidant activities that can relieve chronic diseases such as cancer, diabetes, and cardiovascular diseases (Rahal et al., 2014; Girish and Pradhan, 2011; Kumpulainen, 1999). Phytochemicals have been identified as the secondary metabolites from several biological – pathways of plants (Chu et al., 2011). Numerous studies have shown that a large group of phytochemicals can be metabolized through specific enzymes (CYP450 enzymes) of food and drug metabolism (Wanwimolruk and Prachayasittikul, 2014; Fenton et al., 2015). It evidenced that the phytochemicals can stimulate and/or restrain the CYP450 enzymes through drug – like activities (Korobkova, 2015). Accordingly, a myriad of phytochemicals are used as potent anticancer (epigallocatechin gallate (EGCG), resveratrol and quercetin), antioxidant (defending against oxidative stress), antidiabetic, antibacterial and antifungal agents (Kennedy and Wightman, 2011; Kasote et al., 2015; Newell-Mcgloughlin, 2008; Bonofiglio et al., 2016; Aras et al., 2014). Many phytochemicals have been reported with adverse toxic effects such as phytoestrogens (xenoestrogens and isoflavones), dietary carcinogens (capsaicin, caucasian and ptaquiloside) and mutagens (caffeine, theobromine and theophylline), particularly showing nuclear hormone receptor binding abilities (Lavecchia et al., 2013; Ekor, 2014; Bode and Dong, 2014; Monteiro et al., 2016). Because of structural similarities such as hormones, phytoestrogens are bound to estrogen and peroxisome proliferator – activated receptors (PPARs). Hence, they act as endocrine disrupting compounds (EDCs), and can lead to cancer and numerous health effects. Accordingly, phytoestrogens can alter the function of normal hormone receptors leads to disrupt physiological development of humans and wildlife. Consequently, the binding of phytochemicals with nuclear hormone receptors has been the focus of research (Casanova, 1999; Guerrero-Bosagna and Skinner, 2014; Morito et al., 2001, 2002). Although only some phytochemicals (e.g., isoflavones) were consistent with the experimental toxicological results (Doerge and Sheehan, 2002; Chen and Rogan, 2004; Messina, 2014), much remains to be recognized and debated. In this situation, escalating demands are urgently required to the determination of toxic phytochemicals, which regulated by the Food and Drug Administration (FDA). Likewise, potential toxicity assessments have yet to be continued. Nevertheless, people use plant – based

[*] Corresponding author. Department of Zoology, Sri Venkateswara University, Tirupati, 517502, A.P. India.
E-mail address: pneeraja.bio2015@gmail.com (P. Neeraja).

https://doi.org/10.1016/j.fct.2018.12.033
Received 6 April 2018; Received in revised form 17 December 2018; Accepted 20 December 2018
Available online 21 January 2019

hygiene products in their daily nourishment propensities, but they lack proper knowledge about the toxicity of phytochemicals, including in food products (Bode and Dong, 2014; Fink-Gremmels, 2010).

The Kyoto Encyclopedia of Genes and Genomes (KEGG) is free sourced database prominently used for biological – pathways and chemical compound analyses found with large – scale molecular annotations of pathways and diseases (Kanehisa and Goto, 2000). In addition, KEGG provides chemical names, structures and their related data in literature from the outside of databases. Moreover, more than 3000 phytochemicals related structural classification data from various medicinal herbs and plants have been deposited in KEGG. They organized by KEGG COMPOUND, KEGG BRITE and KEGG PATHWAY modules. Herewith, the KEGG database was selected as the most eminent source for phytochemical screening. Notwithstanding, due to the lack of experimental evidence about the adverse mutagenicity, carcinogenicity, developmental toxicity, skin sensitization and nuclear hormone receptor – mediated potency of a significant number of phytochemicals, earnest scrutiny is required in this current situation. However, there are a large number of phytochemicals and because the experimental bioassay tests are unfeasible to determine the toxicity and nuclear receptor binding properties of each phytochemical compound due to time-consuming, financial burden and ethical considerations. At present, computational toxicology and systems biology approaches are progressively assumed an indispensable part of toxicogenomic examinations because of the short time requirements and low financial burden with reduced laboratory wastages. Furthermore, the escalating demand to reduce the rigorous usage of animals in chemical screening strategies contributes the propagation and development of *in silico* tools. Consequently, numerous industrial stakeholders and government authorities extensively use these potential in – house system biology tools for toxicity evaluations to protect human populations from environmental pollutants and natural phytotoxicants.

The structure activity relationship (SAR) and quantitative – structure activity relationship (QSAR) approaches have been virtually used by some regulatory authorities such as agrochemical, academic, food and pharmaceutical industries to address the toxicological examination of an intrinsic chemical with an understandable annotation (Winkler, 2002). Consequently, SAR/QSAR programs defined as efficient, enormous and enhanced core prediction tools in systems biology, and they can provide potential approaches to extending our understanding of the unsafe effects of phytochemicals. Several commercial (CoMFA/CoMSIA, ADMET – Predictor™ and MetaDrug™) and free academic (Virtual models for property Evaluation of chemicals within a Global Architecture – Non – Interactive Client (VEGA – NIC), Toxicity Estimation Software Tool (T.E.S.T) and the Online Chemical Modeling Environment (OCHEM)) SAR/QSAR modeling packages and programs used to define a wide assortment of toxicological endpoints in plant – flavonoids and environmental pollutants have been reported (Rasulev et al., 2005; Passerini, 2003; Sushko et al., 2011). Compared with the addressed QSAR models, VEGA gives reasonable, reproducible, transparent and verifiable outcomes based on the read across strategy. Evidently, VEGA allows for screening the target chemicals based on the experimental results of similar (structurally related) compounds from implicated independent algorithms of QSAR models. Additionally, VEGA provides comprehensive information on applicability domains, similarity sets, structural alert fragments and reasoning, etc., which can be indicated as potential sources of the decision-making process about a chemical substance. In this examination, highly plausible VEGA QSAR model endpoints and toxicogenomic studies were analyzed against a dataset of 2826 phytochemicals optimized from the KEGG database. Most endocrine – disrupting phytoestrogens/phytochemicals mimic the action of estrogen by acting as an endogenous ligand for estrogen receptor, which stimulates the adverse toxic effects rather than beneficial properties. Therefore, nuclear hormone receptor – mediated effects of the dataset molecules were analyzed against estrogen receptors using standard QSAR models. This investigation extended its scope by implementing structural clustering approach, and a molecular representation space analysis was used to define and understand the similarity in behavior of eminent dataset scaffolds, such as drug – like candidates, for future wet biology analyses. A simultaneous evaluation of adverse toxicity and nuclear hormone receptor – mediated properties of large – scale plant phytochemical datasets has not yet been performed. Therefore, every step was carefully inspected before progressing to the next step. Consequently, these findings could provide critical viewpoints that reveal toxicity and endocrine disruption possibilities of phytochemicals. Altogether, these consequences could contribute to a flexible and high – grade virtual path to reach potential phytotoxicants and provide awareness to society and researchers about plant – derived chemical toxicants.

2. Materials and methods

2.1. Data set

The examined data set is a group of 2826 plant – derived phytochemicals retrieved from the KEGG (http://www.genome.jp/kegg/) database (Kanehisa and Goto, 2000). The defined data set compounds were selected by implementing the structure cleaning and optimization algorithms of the ChemAxon Standardizer module. Compounds without repetitions and with clearly defined structures were considered. Accordingly, the data set compounds were classified as follows: alkaloids, flavonoids, phenylpropanoids, skimate/acetate – malonate pathway – derived compounds, terpenoids, polyketides, fatty acid – related compounds and amino acid – related compounds. SMILES (Simplified Molecular – Input Line Entry System codes) notations and structures for 2826 compounds were obtained from the PubChem database (http://pubchem.ncbi.nlm.nih.gov/search/search.cgi) for further investigation (Table S1). In this report, we did not reflect on the individual predictions or assess the accuracy of expectations for the complete set of phytochemicals. Rather, the objective was to examine the structural fingerprints of chemically assorted phytochemicals and characterize the adverse toxicity with the training sets of the model. Next query was to define a classification scheme or clustering of large data sets, which can reveal the significant cluster indicators for phytochemical dataset investigation.

2.2. QSAR models

CAESAR models are particularly permitted to develop QSAR models to support the Registration, Evaluation, Authorization and Restriction of Chemicals legislation (REACH). In CAESAR, five models (bioconcentration factor (BCF), skin sensitization, carcinogenicity, mutagenicity and developmental toxicity) were characterized with high significance for REACH and judged according to the OECD principles for the legalization of (Q)SAR models for regulatory purposes (Benfenati, 2010). The BCF model is expressed as a real number and other four endpoints are expressed as a binary classification form by following the OECD or US EPA rules. The Benigni and Bossa ISS (Istituto Superiore di Sanita) rule – based model is widely used to extend the significance of mutagenicity and carcinogenicity endpoints. Likewise, ISS provides the predictions as a binary classification, in addition which reveal the structural alert fragments of compounds. Estrogen receptors (ER) are one of extensively studied molecular targets for EDCs screening. Hence, the ER relative binding affinity (RBA) model was allowed for the qualitative prediction of EDCs with the classification mode against the estrogen receptors. Overall, the described models were implicated in the VEGA – NIC v1.1.4 program (http://www.vega-qsar.eu/), which allows QSAR models to investigate the toxicity, ecotoxicity, environmental toxicity and physicochemical properties of chemical substances (Benfenati et al., 2013). VEGA QSAR is performed in a batch mode and provides reproducible, transparent and reliable results for chemical compounds. Alongside the classification, VEGA provides the applicability domain (AD) to assess the reliability of predictions by using several inbuilt parameters. The regard parameters of individual models are shown below. The AD is shown with six criteria: *similar molecules with known experimental value* (similarity index), *accuracy of prediction for similar molecules* (accuracy index), *concordance for similar molecules* (concordance index), *atom – centered fragments similarity check* (ACF index), *model descriptors range check* (descriptors range check) and *global AD Index* (AD index). For carcinogenicity, CAESAR provides two additional criteria: *model assignment reliability* (positive/non – positive difference) and *neural map neurons concordance* (neurons concordance); however, ISS offers only five of defined criteria for carcinogenicity. Consequently, VEGA has classified a compound as either positive or negative with one of the three explanations: prediction has low reliability (compound out of the AD), moderate reliability (compound could be out of the AD), and high reliability (compound into the AD). Moreover, VEGA lists the six most similar compounds for specified molecules by comparing with its internal dataset. Similarities between represented molecules and most similar compounds are indicated by the similarity index. For each similar compound, VEGA indicates the predicted, experimental and similarity values. An algorithm analyzes the similar compounds, which is not a part of the QSAR modeling process. In VEGA, results are revealed by the independent methods, and they can guide as special perceptions into an elucidated chemical structure. The reports are reflected for model descriptors, similarity sets, AD space, structural alerts, reasoning, etc. Overall, the obtained data are represented as a potential source that provide awareness on the QSAR model prediction, and may help the expert to make the decision about a chemical (Plošnik et al., 2015; Kumar et al., 2018).

2.2.1. Mutagenicity

The mutagenicity model is defined in the literature (Ferrari and Gini, 2010; Benigni et al., 2008). The model comprises a set of 4204 compounds and their qualitative *Salmonella typhimurium* (Ames test) test results. This integrated model consists of two models. Model A is a classifier of trained Support Vector Machine (SVM), and model B is useful for removal of false negative (FNs) predictions based on structural alerts (SAs). The original model is working with MDL and dragonX software for the initial descriptors calculation and in – house programming modules for the overall classification. The mutagenicity is shown as binary: 'non – mutagen' or 'mutagen'.

2.2.2. Carcinogenicity (CAESAR)

The carcinogenicity model is described in the literature (Fjodorova et al., 2010). The model comprises a set of 806 compounds from the carcinogenic potency database (CPDB). The model is built on values through the neural network analysis (Counter Propagation Artificial Neural Network (CPANN)). The neural network results provide two values (0, 1) that respectively indicate positive and non-positive prediction. Moreover, it represented that how much the neuron supports in which predicted compound falls belongs to the classification either positive or negative. The predictions are shown as binary, 'non – - carcinogenic' or 'carcinogenic' activity of a compound.

2.2.3. Carcinogenicity (ISS)

In addition, the ISS model was studied to extend the scope of carcinogenic activity of data set compounds. The carcinogenicity model is described in the literature (Benigni and Bossa, 2011). The model is built on a set of 797 compounds of ToxTree (http://toxtree.sourceforge.net) database. The predictions are indicated as binary, 'carcinogenic' or 'non – carcinogenic', with specific SAs (Benigni and Bossa, 2008). Generally, experimental sets of CAESAR and ISS models are not equal; however, both model sets comprise some similar compounds within their data sets.

2.2.4. Developmental toxicity

The developmental toxicity model is explained in the literature (Cassano et al., 2010). The model comprises an Arena data set of 292 compounds. This data set was sorted by the FDA for qualitative prediction of developmental toxicity. The model was performed by the Random Forest Method and run with WEKA (Waikato Environment for Knowledge Analysis) open source libraries. The prediction was performed as binary, 'non – toxicant' or 'toxicant', and the FDA category was defined as A or B and C or D, respectively.

2.2.5. Skin sensitization

The skin sensitization model is defined in the literature (Chaudhry et al., 2010). The model comprises a data set of 209 local lymph node assay model compounds obtained from the Gerberick data set. The model has an Adaptive Fuzzy Partition (AFP) based on 8 descriptors indicated that two values, such as O(positive) and O(negative), which suggests a prediction as toxic and non-toxic classes. The expression is shown as binary, 'inactive' or 'active' with a class index like carcinogenicity model.

2.2.6. Estrogen receptor relative binding affinity (ER_RBA)

The ER_RBA model is reported in the literature (Roncaglioni et al., 2008). The model comprises a set of 806 compounds obtained from the cited work of Roncaglioni et al. (2008). This is a QSAR classification model, which works based on a classification and regression tree (CART) algorithm. The qualitative prediction of estrogen – mediated activity is characterized as binary, defined as 'active' or 'inactive'.

Altogether, the CAESAR mutagenicity and carcinogenicity models were organized with well – balanced data sets, specifically non – toxic and toxic compounds are characterized equivocally. The carcinogenicity (ISS), developmental toxicity and skin sensitization models are more subjective, which means most of the training set compounds belong to a toxic class. However, the ER_RBA model is deferentially distributed, i.e., most of the training set compounds belong to a non – toxic class (inactive).

2.3. The representation space analysis, clustering and classification

In our classification scheme or clustering analysis, vectors indicated compounds in the representation space. It was built with similarity sets, which are revealed from QSAR model predictions. The representation space is an association of similarity sets related to a particular endpoint. In fact, it is a part of the models training set and recommends a specific dataset to evaluate chemical compounds. A molecule is reflected by a multi – dimensional vector for further investigation, where each vector component shows a molecule from the portrayal space (Vračko and Bobst, 2013; Plošnik et al., 2015; Kumar et al., 2017, 2018).

Self – Organizing Maps (SOM) or the Kohonen neural network is an unsupervised machine learning method used to compute common neural networks. The architecture of SOM is designed to reduce the dimensionality of multivariate data to low – dimensional spaces, i.e., a two – dimensional array of neurons. Usually, they are vector (objective) weights. SOM employs a non-linear iterative approach with a two – step algorithm. Initially, the observations are consecutively assembled with the nodes of all neurons, and the algorithm defines most similar objective. Next, the weights of the object and its neighborhood are altered to evolve into the same vector; both steps repeated until stabilized weights are generated. Afterwards, the nodes are scattered on a two – dimensional map, and the most similar objects are indicated as one cluster. Hence, SOM is a network of vector weights (Rallo et al., 2011; Vračko and Bobst, 2013; Plošnik et al., 2015; Kumar et al., 2018). In our case, the vector weights were resolved from the representation space. When the compounds were found in one neuron, the most comparable vector components were close together due to anticipated model identification structures situated in the same neuron, and they were characterized as cluster indicators for one group. The technical parameters such as network dimensions and the number of training epochs must be identified. They were fixed and determined by two criteria: the initial one was the top map that comprising the largest dimension, which does not show empty neurons and, the next one was the average error at one vector (Jezierska et al., 2004). The grid dimension was set to 10 × 10, hexagonal topology, Gaussian neighborhood function, 100 presentation times for the SOM algorithm (number of iterations) and 1000 learning epochs (Vračko and Bobst, 2013; Plošnik et al., 2015; Kumar et al., 2018). For the SOM technical parameters the data components were elucidated with R – programing as implemented in XLSTAT – R software (https://cran.r-project.org/web/packages/kohonen/). In our report, the clusters with more than 50 compounds were specified with cluster indicators. Herewith, we examined the cluster weights, where the indicators with highest value were judged, which can help to make a decision about the investigated data set.

2.4. Model evaluation

After each endpoint prediction, compound set of data with the priority of the representation space for each endpoint was taken into consideration. Purposely, this space contributes a linear order to explain how the phytochemical data set is projected with the experimental training set compounds. In addition, this can reflect how well the phytochemical compounds fit into the representation space of each endpoint. Thereafter, the performances of the QSAR models were evaluated with representation space compounds (training sets) using traditional Cooper statistics in an implemented confusion matrix. This is a general task to explore a number of correct and incorrect predictions from the generated output. The consequences were analyzed by comparing with the experimental (actual) values in the representation space data (Fjodorova et al., 2010; Cooper et al., 1979). The representation of such metrics is N × N, where N is the number of positive and negative predictive values in optimum space. Thus, a highly efficient 2 × 2 confusion matrix was used to evaluate all of the QSAR models.

Confusion Matrix	Actual value		
Predicted Positive value	True Positive (TP) ***Correctly predicting a label***	False Positive (FP) ***Falsely Predicting a label***	*Precision* or *Positive Predictive Value (PPV)*
Predicted Negative value	False Negative (FN) ***Missing and incoming label***	True Negative (TN) ***Correctly predicting the other label***	*Negative Predictive Value (NPV)*
	Sensitivity (SN)	***Specificity (SP)***	***Accuracy***

Measure Derivations

- Sensitivity (SN): TP/(TP + FN) (True Positive Rate (TPR))
- Specificity (SP): TN/(FP + TN) (True Negative Rate (TNR))
- Precision (PPV): TP/(TP + FP)
- Negative Predictive Value (NPV): TN/(TN + FN)
- Accuracy (ACC): (TP + TN)/(TP + FP + FN + TN)
- Matthews Correlation Coefficient (MCC): TP × TN–FP × FN/sqrt((TP + FP) × (TP + FN) × (TN + FP) × (TN + FN))

3. Results and discussion

3.1. Toxicity and nuclear receptor mediated assets of phytochemicals

The predicted statistics for a set of 2826 phytochemicals are shown in Table 1. For mutagenicity 790 (27.9%) compounds were predicted to be mutagenic, for carcinogenicity (CAESAR) 989 (34.9%) compounds were predicted to be carcinogenic, for carcinogenicity (ISS) 1146 (40.5%) compounds were predicted to be carcinogenic, for developmental toxicity 2300 (81.3%) compounds were predicted to be toxic and for skin sensitization 1692 (59.8%) compounds were predicted to be sensitizer. These estimations were not overwhelmed with negative predictions. Because of the mutagenicity and carcinogenicity (CAESAR), models were constructed with well – balanced training sets. The carcinogenicity (ISS), developmental toxicity and skin sensitization models are more biased, i.e., most of the training set compounds belong to toxic class. Thus, the carcinogenicity (ISS), developmental toxicity and skin sensitization models were anticipated most of the compounds as positive (toxic) than mutagenicity and carcinogenicity (CAESAR) models. For the ER_RBA model, 930 (32.9%) compounds were predicted to be active estrogen receptor binders possibly exhibiting endocrine – disrupting activity by an interaction with estrogen receptors. In the ER_RBA model, most of the training set compounds belonged to inactive class. Interestingly, the prediction revealed that 930 compounds have estrogen receptor mediators, which is completely different from other QSAR models. For 358 investigated data set compounds, the QSAR models were provided the experimental results for one of the explored toxic endpoints (Table 1). The accuracy of this experimental data set compounds predicted to be 84%. Consequently, the predictive ability of the tool shows better performance and effort for experimental compounds (Kumar et al., 2017, 2018). In future, VEGA can be defined as an efficient (Q)SAR modeling tool for toxicologists and pharmaceutical developers, which could be a valuable aid to the wet-biology explanations for molecular screening elucidations of environmental chemicals and pollutants.

3.2. Applicability domain (AD) analysis

For the elucidation of QSAR model's prediction, an assessment of the AD is proposed as a noteworthy and essential objective. As indicated by the OECD's third principle for the examination of (Q)SAR models, *'a (Q) SAR model should be associated with one of the applied domains'* (OECD, 2004). *QSAR models are inexorably related with confines in terms of the varieties of chemical structures, physicochemical properties and mechanisms of actions*. Here, CAESAR indicated six criteria (eight criteria for carcinogenicity model), and ISS provided five criteria (descriptor range check of the model not provided) for the exploration of the AD. The models AD space analysis is shown in Fig. 1(A-F). Therefore, the essentials of different criteria for AD indicate how well the investigated compounds fit into the domains of the model. Indeed, the global AD index was considered for the AD space analysis. Accordingly, for mutagenicity 34% of the compounds were in the domain, 28% were outside of the domain and others were not clearly found in the domain (Fig. 1(A)). For carcinogenicity (CAESAR) only 13% of the compounds were found in the domain, 67% were outside of the domain and others were not clearly found in the domain (Fig. 1(B)). For carcinogenicity (ISS) only 11% of the compounds were found in the domain, 55% were found outside of the domain and others were not clearly found in the domain (Fig. 1(C)). For developmental toxicity 34% of the compounds were found in the domain, approximately 48% were found outside of the domain and others were not clearly found in the domain (Fig. 1(D)). For skin sensitization only 20% of the compounds occupied the domain, 70% were outside of the domain and others were not clearly found in the domain (Fig. 1(E)). In the case of ER_RBA 30% of the phytochemicals were distributed in the domain, 49% were outside of the domain and others are not clearly found in the domain (Fig. 1(F)). Interestingly, a significant number of compounds have been found outside the AD space. However, the limited phytochemicals were consistent with the AD. Therefore, the mutagenicity, developmental toxicity and ER_RBA endpoints AD spaces suggested approximately 35% of toxic phytochemicals. Moreover, only 10–20% of the compounds were predicted to be carcinogens. This is consistent with the skin sensitization (20%) model. For 305 predicted mutagens and 1146 carcinogens, the CAESAR and ISS models were individually revealed the structural alert fingerprints, which indicate the toxic possibilities of phytochemicals (Table

Table 1
QSAR models prediction statistics for a set of 2826 phytochemicals.

Endpoints	Pred. positive[a]	Pred. negative[a]	Pred. (exp.) positive[b]	Pred. (exp.) negative[b]
Mut.	790 (27.9%)	2036 (72%)	51 (45)	140 (127)
Car.1	989 (34.9%)	1837 (65%)	18 (14)	23 (18)
Car.2	1146 (40.5%)	1680 (59.4%)	23 (13)	10 (6)
Dev. Tox.	2300 (81.3%)	526 (18.6%)	15 (13)	6 (4)
Skin Sens.[c]	1692 (59.8%)	1132 (40%)	18 (18)	2 (1)
ER_RBA	930 (32.9%)	1896 (67%)	27 (26)	25 (17)

[a] Compounds with predicted positive and negative classes.
[b] Compounds with known experimental values, c Two compounds were not classifiable, Mut.: Mutagenicity, Car. 1: Carcinogenicity (CAESAR), Car. 2: Carcinogenicity (ISS), Dev Tox.: Developmental toxicity, Skin Sen.: Skin sensitization, ER_RBA; Estrogen receptor relative binding affinity models.

S2). As shown in Table S2, SA6 (propiolactones and propiosultones), SA7 (epoxides and aziridines), SA8 (aliphatic halogens), SA12 (quinones), SA14 (aliphatic azo and azoxy), SA18 (polycyclic aromatic hydrocarbons), SA19 (heterocyclic and polycyclic aromatic hydrocarbons), SA27 (nitro aromatic) and SA28 (bis aromatic mono − and dialkylamine) SAs were commonly reported for mutagenicity and carcinogenicity (ISS) endpoints. Thus, the majority of polyketides, moderate number of alkaloids and flavonoids, and low number of other compounds were anticipated as toxicants. The endpoints incorporate mutagenicity and carcinogenicity (ISS) was demonstrating strong agreement for the observed outcomes by characterizing SAs. Generally, mycotoxins of fungi are produced as polyketides (e.g., Aflatoxin B1).

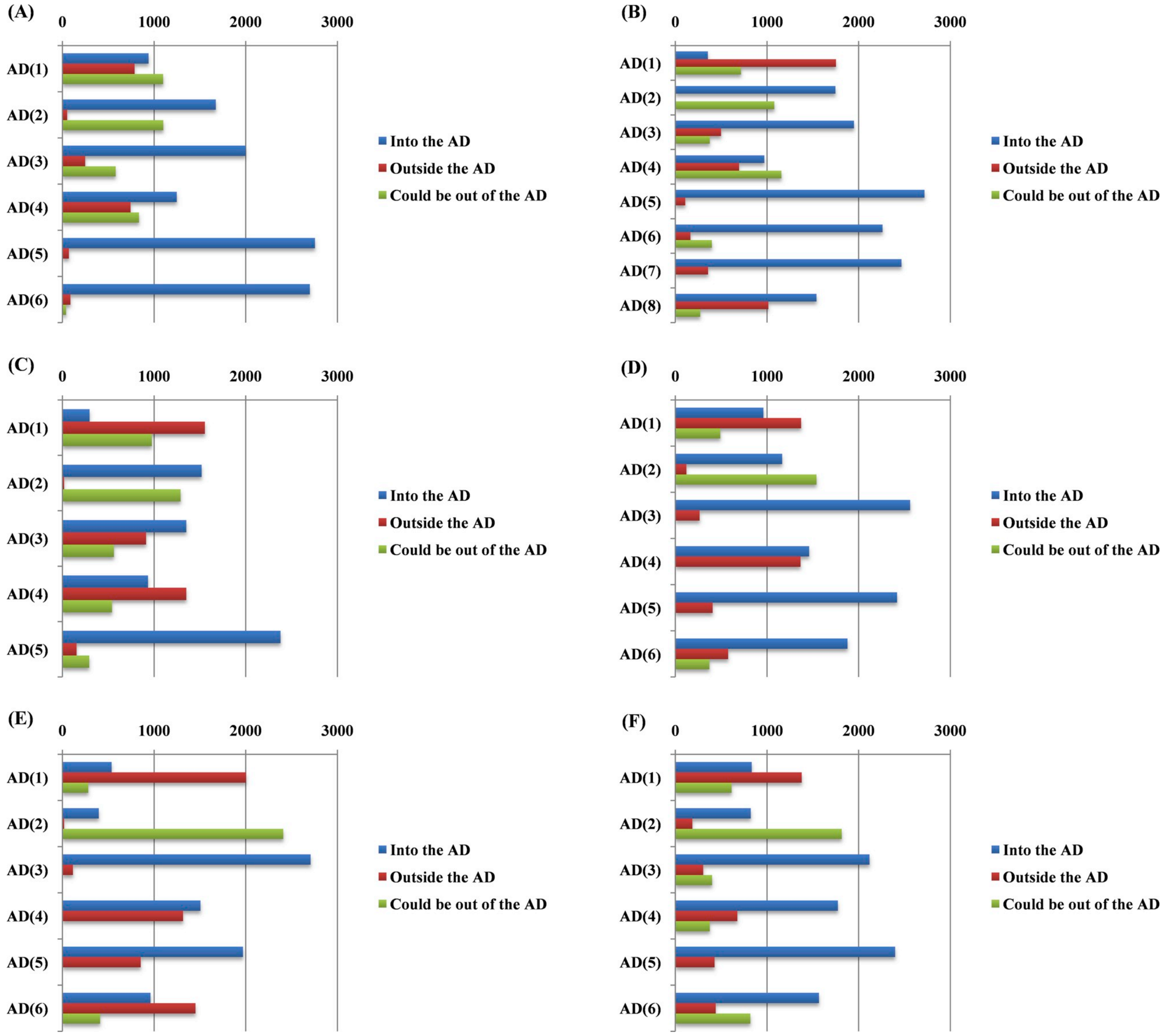

Fig. 1. Applicability domain analysis and measured scores of regarded six endpoints. (A): Mutagenicity; (B): Carcinogenicity (CAESAR); (C): Carcinogenicity (ISS); (D): Developmental toxicity; (E): Skin sensitization; (F): Estrogen receptor relative binding affinity. AD(1): AD index; AD(2): Similarity index; AD(3): Accuracy index; AD(4): Concordance index; AD(5): Descriptor range check; AD(6): ACF index; AD(7): Pos/Non − Pos difference; AD(8): Neuron concordance.

Table 2
The top 10 clusters and their specific indicators of the more than 50 elements predicted from QSAR models (The cluster indicators are defined from model's training sets.).

Size of the cluster	CAS No.	Name of the compound	Classification	Experimental value	Predicted value
(a) Mutagenicity					
288	**23255–69–8**	Fusarenon X	Mycotoxin	Non-mutagenic	Suspect mutagenic
266	2270-40-8	Diacetoxyscirpenol	Mycotoxin	Non-mutagenic	Suspect mutagenic
261	21259-20-1	T-2 toxin	Mycotoxin	Non-mutagenic	Suspect mutagenic
186	119525-97-2	2-hexenoic acid	Fatty acid	Mutagenic	Mutagenic
183	152-84-1	Ruberythric acid	Dye	Non-mutagenic	Mutagenic
169	77-06-5	Gibberellic acid	Hormone	Non-mutagenic	Non-mutagenic
157	57817-89-7	Stevioside	Glycoside	Mutagenic	Mutagenic
151	87625-62-5	Ptaquiloside	Glucoside	Mutagenic	Mutagenic
148	1405-86-3	Glycyrrhizin	Emulsifier	Non-mutagenic	Non-mutagenic
141	**520–18–3**	Kaempferol	Flavonoid	Mutagenic	Mutagenic
(b) Carcinogenicity (CEASAR)					
462	**51333–22–3**	Budesonide	Steroid	Carcinogen	Non-carcinogen
459	**23255–69–8**	Fusarenon X	Mycotoxin	Non-carcinogen	Non-carcinogen
423	83-79-4	Rotenone	Insecticide	Non-carcinogen	Non-carcinogen
375	**520–18–3**	Kaempferol	Flavonoid	Non-carcinogen	Non-carcinogen
365	**517–28–2**	Hematoxylin	Dye	Carcinogen	Carcinogen
360	7681-93-8	Natamycin	Antibiotic	Non-carcinogen	Non-carcinogen
359	50-23-7	Hydrocortisone	Steroid	Non-carcinogen	Non-carcinogen
354	**53–43–0**	Prasterone	Steroid	Carcinogen	Carcinogen
342	**10048–13–2**	Sterigmatocystin	Mycotoxin	Carcinogen	Carcinogen
297	65176-75-2	5, 6-dimethoxysterigmatocystin	Mycotoxin	Carcinogen	Carcinogen
(c) Carcinogenicity (ISS)					
941	**51333–22–3**	Budesonide	Steroid	Carcinogen	Carcinogen
659	17673-25-5	Phorbol	Diterpenes	Carcinogen	Carcinogen
648	60102-37-6	Petasitenine	Alkaloid	Carcinogen	Carcinogen
590	434-07-1	Oxymetholone	Steroid	Carcinogen	Carcinogen
533	**53–43–0**	Prasterone	Steroid	Carcinogen	Non-carcinogen
466	**517–28–2**	Hematoxylin	Dye	Carcinogen	Non-carcinogen
448	117-39-5	Quercetin	Flavonoid	Carcinogen	Carcinogen
412	29069-24-7	Prednimustine	Ester	Carcinogen	Carcinogen
409	**10048–13–2**	Sterigmatocystin	Mycotoxin	Carcinogen	Non-carcinogen
399	23246-96-0	Riddelline	Alkaloid	Carcinogen	Carcinogen
(d) Developmental toxicity					
651	**50–04–4**	Cortisone acetate	Steroid	Toxicant	Toxicant
539	630-60-4	Ouabain	Glycoside	Non-toxicant	Non-toxicant
480	50-24-8	Prednisolone	Steroid	Toxicant	Toxicant
473	23214-92-8	Doxorubicin	Antibiotic	Toxicant	Toxicant
462	53-06-5	Cortisone	Steroid	Toxicant	Toxicant
418	20830-81-3	Daunorubicin	Antibiotic	Toxicant	Toxicant
413	152-72-7	Acenocoumarol	Coumarin	Toxicant	Toxicant
411	7683-59-2	Isoprenaline	Catechol	Toxicant	Toxicant
394	81-81-2	Warfarin	Anticoagulant	Toxicant	Toxicant
386	**59–01–8**	Kanamycin a	Antibiotic	Toxicant	Toxicant
(e) Skin sensitization					
1274	1166-52-5	Dodecyl gallate	Food additive	Sensitizer	Sensitizer
1105	1675-54-3	Bisphenol a diglycidyl ether	Food contaminant	Sensitizer	Sensitizer
947	514-10-3	Abietic acid	Plant-organic compound	Sensitizer	Sensitizer
881	135099-98-8	1345trimethoxyphenyl4dimethylpentane13dione	Preparation products and raw materials	Non-sensitizer	Non-sensitizer
829	3326-32-7	Fluorescein 5-isothiocyanate	Dye	Sensitizer	Sensitizer
827	**59–01–8**	Kanamycin a	Antibiotic	Non-sensitizer	Non-sensitizer
820	13557-75-0	Sodium lauroyl lactylate	Surfactant	Sensitizer	Sensitizer
726	94612-91-6	Sodium 4-(3,5,5-trimethylhexanoyloxy)benzenesulfonate	Preparation products and raw materials	Sensitizer	Sensitizer
663	31906-04-4	Lyral	Cosmetic	Sensitizer	Sensitizer
639	3810-74-0	Streptomycin sulfate	Antibiotic	Non-sensitizer	Non-sensitizer
(f) ER_RBA					
539	2203-97-6	Hydrocortisone sodium succinate	Steroid	Inactive	Inactive
506	4319-56-6	Desoxycorticosterone glucoside	Steroid	Inactive	Inactive
310	50-03-3	Hydrocortisone acetate	Steroid	Inactive	Inactive
307	57524-89-7	Hydrocortisone valerate	Steroid	Inactive	Inactive
303	751-94-0	Fusidate sodium	Antibiotic	Inactive	Inactive
272	50-55-5	Reserpine	Alkaloid	Inactive	Inactive
239	**50–04–4**	Cortisone acetate	Steroid	Inactive	Inactive
229	62507-01-1	3′-benzyloxy-5,7-dihydroxy-3,4′-dimethoxyflavone	Preparation products and raw materials	Active	Active
211	**490–46–0**	Epicatechin	Flavonoid	Active	Active
181	**855–96–9**	Eupatorin	Flavonoid	Inactive	Active

Bold: Common compounds that are frequently found in all the endpoints.

Furthermore, structural properties of alkaloids, flavonoids and other phytochemicals such as number of aromatic ring structures and related side chains comprising compounds were predicted as toxicants (Table S2). As per these explorations, one can conclude that the data set compounds represented accurately predicted positive compounds. Prognostic positives of the investigated data set were expected to comprise the phytochemicals, which are closer to the training set of one of the endpoints. Moreover, the other AD criterias were satisfied for all the developed models (Fig. 1(A-F)).

3.3. The representation space analysis

The representation space investigation was defined as follows. A representation space for the mutagenicity contains 972 compounds with an aggregate size of 14051, i.e., 23% of the total training set was allotted. Overall, the clusters and their indicators are reported in Table S3. For mutagenicity, 7009 compounds were collected into 71 clusters with more than 50 elements. The largest cluster had 288 chemicals, i.e. Fusarenone X (mycotoxin), and its indicator is an experimental non − mutagen. However, the structure consists of the epoxide and aziridine relevant chemical fragments and moieties (SA7). Therefore, it is suspected as a 'mutagenic'. From Tables S3(a), 47 clusters related compounds had non − mutagenic indicators, other 24 clusters had experimental mutagenic indicators. For carcinogenicity CAESAR, the representation space comprises 319 compounds with an aggregate size of 13116, i.e., 39.5% of the total training set was equipped. According to Tables S3(b), 10727 compounds were gathered into 65 clusters with more than 50 elements. The largest cluster comprised 462 chemicals, i.e., budesonide (glucocorticoid), and its indicator was an exploratory carcinogen. Because of anticipated compound is outside the AD of model (AD index). Thus, the CAESAR has shown as false negative prediction (non-carcinogenic). As revealed in Table S3(b), the 36 clusters related compounds comprise non − carcinogenic indicators, other 29 clusters have experimental carcinogenic indicators. On account of the ISS carcinogenicity model the representation space indicated 398 compounds with a total cluster size of 20693, i.e., half of the aggregate training set was adopted. According to Tables S3(c), 17469 compounds were collected into 82 clusters with an excess of 50 elements. Interestingly, the largest cluster comprised 941 chemicals, i.e., budesonide, and its indicator was an experimental carcinogen likewise CAESAR model. The classification scheme defined 13 cluster compounds having non − carcinogenic indicators, other 69 clusters comprised experimental carcinogenic indicators. Here, one can observe that, as per the earlier description CAESAR and ISS models are not equal; however, both models comprise some similar compounds within their training sets. Hence, most clusters of both models have shown some similar molecules. Additionally, ISS demonstrated that more efficiency for AD index. For developmental toxicity, the representation space identified 168 compounds with an aggregate size of 12806, i.e., comprising 57.5% of the total training set. As shown in Tables S3(d), 11171 compounds were gathered into 48 clusters with more than 50 elements. Interestingly, the largest cluster had 651 compounds, i.e., cortisone acetate (steroid hormone), its indicator was an experimental toxicant. As per classification scheme, 9 cluster − related compounds showed non − developmental toxicity indicators while 39 clusters had experimental developmental toxicity indicators. In the case of skin sensitization, the representation space specified 127 compounds with a total cluster size of 13706, i.e., 60% of the aggregate training set was dispensed. As shown in Tables S3(e), 12679 chemicals were assembled into 41 clusters with more than 50 elements. The largest cluster had 1274 compounds, i.e., dodecyl gallate (food additive), and its indicator is an experimental sensitizer. In total, 8 cluster − associated compounds were non − sensitizer indicators, other 33 clusters had experimental positive indicators. For the ER_RBA the representation space had 297 compounds with an aggregate size of 11141, i.e., 37% of the total training set was provided. As shown in Tables S3(f), 8462 compounds were grouped into 67 clusters with an excess of 50 elements. The largest cluster had 539 compounds, i.e, hydrocortisone succinate (corticosteroid), and its indicator was experimentally inactive for the estrogen receptor. A total of 45 clusters depicted compounds with inactive indicators, other 22 clusters comprised experimentally active indicators. As per Table 1, an extensive number of phytochemicals were anticipated as toxicants. Now, one can observe that most of the elucidated clusters associated compounds were experimental toxicants (Table S3). This kind of results was revealed due to phytochemicals showing close resemblances with distributed similarity sets of the models (Vračko and Bobst, 2013; Plošnik et al., 2015; Kumar et al., 2018).

3.4. Most studied indicators

Contingent upon the representation space indicating cluster sizes, one can conclude that the indicated data set is larger than the considered data set (2826 phytochemicals) and comprises different compound classes. Altogether, the anticipated toxic indicators were shown reliable experimental evidences. The classification scheme of the top 10 clusters and their indicators are reported in Table 2(a − f). Accordingly, the cluster indicators can mostly be defined from plants, fungi, synthetic chemicals, and the secretions of living organisms characterized as mycotoxins, fatty acids, dyes, hormones, glycosides, glucosides, emulsifying agents, flavonoids, steroids, insecticides, antibiotics, diterpenes, alkaloids, ester coumarins, catechols, anticoagulants, food additives, food contaminant, plant − derived organic compounds, prepared products and raw materials, surfactants, and cosmetics, etc. In our case, this was expected for the data set analysis. The representation spaces revealed cluster indicators, and the classification schemes are distinctively scattered for the respective toxic endpoints. In this study, the clusters above with more than 50 elements are depicted as the larger ones. An intact classification scheme and the clustering structures for six endpoints are shown in Fig. 2. Approximately 20–35% of the data set compounds are found in the largest clusters while another cluster indicator was collected into small groups or diverse components; this result shows some distinctions in the endpoints. Interestingly, the highly articulated clustering structures were for skin sensitization, carcinogenicity (ISS), and developmental toxicity where we categorized 1274, 941 and 651 chemicals as an enormous cluster, respectively. In contrast, the data set clusters of mutagenicity, carcinogenicity (CAESAR) and ER_RBA mostly scattered into smaller groups. In fact, this outcome is different, because the models were behaving differently due to their different training sets (Vračko and Bobst, 2013; Plošnik et al., 2015; Kumar et al., 2018). Table 3 shows the statistically predicted values of entire cluster indicators and defined experimental and predicted values as accounted for the regarded QSAR models. Consequently, the assigned data set shows significant correctly and incorrectly predicted labels. By considering precision (positive predictive value), which is related, most of the cluster indicators predictions are good, and they were 82%, 86%, 85%, 100%, 95% and 78% for the mutagenicity, carcinogenicity (CEASAR), carcinogenicity (ISS), developmental toxicity, skin sensitization and ER_RBA models, respectively. The other statistically significant assets such as sensitivity, specificity, negative predictive values, accuracy and Matthews correlation coefficient were found in adaptable regions and showed a strong agreement with precision (Fig. 3). Thus, most of the predictions of investigating data set were correct.

4. Conclusions

In this study, integrated *in silico* methods and computational systems biology approaches were used as potential tools to determine the toxicity and endocrine disruption activities of a set of 2826 phytochemicals. The endpoints such as mutagenicity, carcinogenicity (both CAESAR and ISS models), developmental toxicity, skin sensitization and relative binding affinity models of estrogen receptors were assessed. For some of the phytochemicals (358 compounds) described models were indicated experimental values, and a most part of them was correctly predicted with 84% accuracy. It was found that, compared to the endpoints of mutagenicity and carcinogenicity (ISS) majority of toxic chemicals contains structural alerts fragments, which are revealed by inbuilt parameters of the models. QSAR models provide comprehensive information on applicability domains and most similar compounds from training sets were considered for clustering and

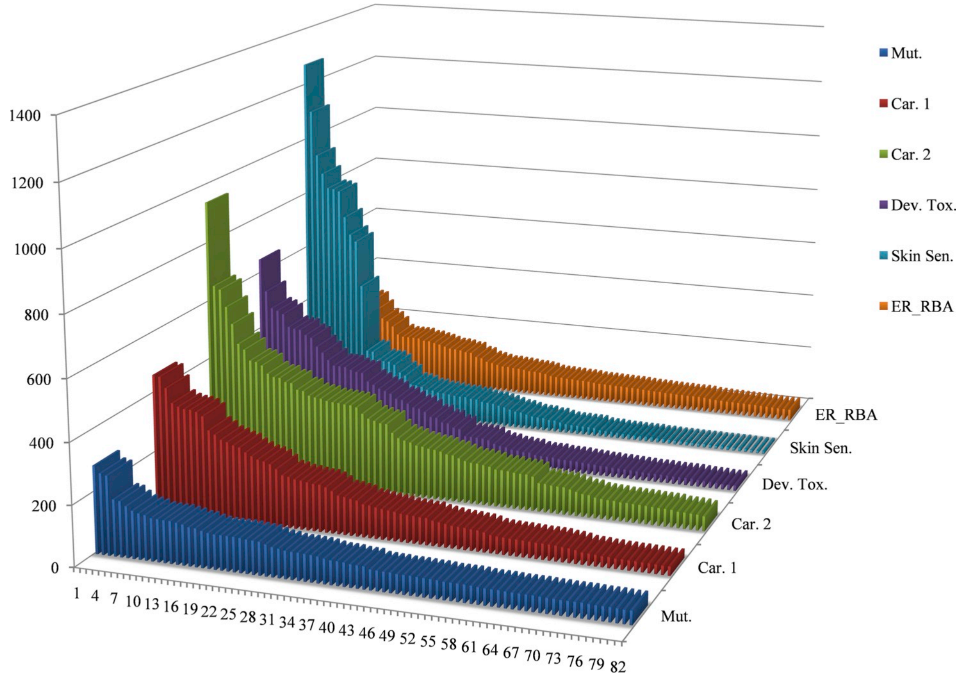

Fig. 2. Sizes of clusters for regarded six endpoints. Mut: Mutagenicity; Car. 1: Carcinogenicity (CAESAR); Car. 2: Carcinogenicity (ISS); Dev Tox: Developmental toxicity; Skin Sen.: Skin sensitization; ER_RBA; Estrogen receptor relative binding affinity.

Table 3
Number of clusters, and statistics of the number of predicted positive and negative labeled indicators from model's training sets.

Endpoints	Mut.	Car. 1	Car. 2	Dev Tox.	Skin Sen.	ER_RBA
Size of the clusters	972	319	398	168	127	297
TP	402	133	224	117	99	112
TN	473	151	63	51	20	144
FP	84	20	39	0	5	31
FN	13	15	72	0	3	10
Sensitivity	0.9687	0.8986	0.7568	1	0.9706	0.918
Specificity	0.8492	0.883	0.6176	1	0.8	0.8229
Precision	0.8272	0.8693	0.8517	1	0.9519	0.7832
Negative Predictive Value	0.9733	0.9096	0.4667	1	0.8696	0.9351
Accuracy	0.9002	0.8903	0.7211	1	0.937	0.862
Matthews Correlation Coefficient	0.8091	0.7803	0.3453	1	0.7956	0.7295

Mut.: Mutagenicity, Car. 1: Carcinogenicity (CAESAR), Car. 2: Carcinogenicity (ISS), Dev Tox.: Developmental toxicity, Skin Sen.: Skin sensitization, ER_RBA; Estrogen receptor relative binding affinity, True Positive: experimental positive, predicted positive. True Negative: experimental negative, predicted negative. False Positive: experimental negative, predicted positive. False Negative: experimental positive, predicted negative.

classification schemes examination. The applicability domain estimations have different criterias. According to consider global AD index, one can conclude that in mutagenicity 34% of the compounds were found in the domain, and 28% were outside of the domain. For carcinogenicity, 13% of the compounds were found in the domain, and 67% were outside of the domain. For carcinogenicity (ISS), 11% of the compounds were found in the domain, and 55% were outside of the domain. For developmental toxicity, 34% of the compounds were found in the domain, and approximately 48% were outside of the domain. For skin sensitization, 20% of the compounds were found in the domain, and 70% were outside of the domain. In the case of ER_RBA, 30% of the compounds were found in the domain, and 49% were outside of the domain. According to the regarded toxic endpoints, some phytochemicals were not clearly found in the domain. The representation space revealed clustering and classification scheme interpretations provided that significant cluster indicators. Considering clusters with more than 50 elements, one concludes that the sizes of cluster indicators are larger than the size of the investigated data set. This is an effective reason for the aggregate data set prediction. As expected, models behave differently due to different training sets. Consequently, clusters for respective endpoints are different. Accordingly, largest clusters were found in three of the endpoints such as skin sensitization, carcinogenicity (ISS), and developmental toxicity. For the other endpoints, mutagenicity, carcinogenicity (CAESAR) and ER_RBA were mostly scattered into

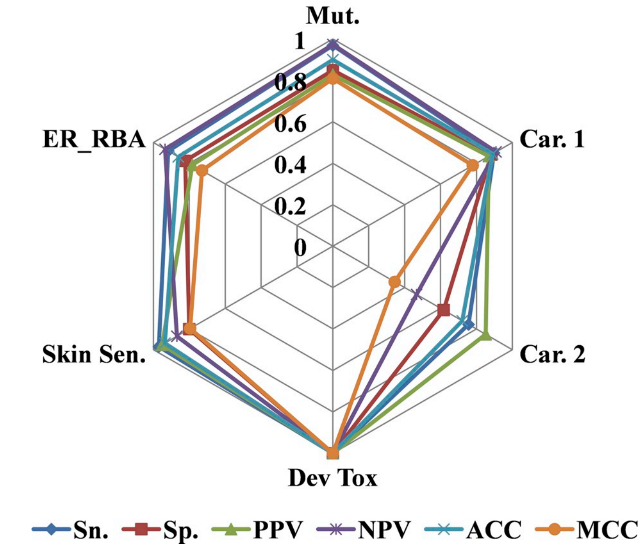

Fig. 3. Radial wheel diagram of predicted statistics for representation space resulting cluster indicators revealed by using a 2 × 2 confusion matrix. Mut: Mutagenicity; Car. 1: Carcinogenicity (CAESAR); Car. 2: Carcinogenicity (ISS); Dev Tox: Developmental toxicity; Skin Sen.: Skin sensitization; ER_RBA; Estrogen receptor relative binding affinity. Sn: Sensitivity; Sp: Specificity; PPV: Precision or positive predictive value; NPV: Negative predictive value; ACC: Accuracy; MCC: Matthews Correlation Coefficient.

smaller groups. Interestingly, predicted positive and negative labels of clusters demonstrated good precision and other significant scores. According to the predicted positive compounds, we suggest that care should be taken towards phytochemicals specifically in the use of food industries, food and beverage processing, biofilms preparation for food packaging and textile industries. Additionally, we emphasize that the QSAR modeling, clustering and classification schemes are indispensable approaches to determine potential toxic chemicals (i.e., environmental pollutants or persistent organic pollutants) of large data sets for effortless comprehension, examination and utilization. Altogether, our new directions could useful as a promising way to evaluate small molecules (i.e., phytochemicals) before they are used and processed.

Conflicts of interest

All authors declare no conflicts of interest.

Acknowledgements

K. Kranthi Kumar (F1 − 17.1/2013 − 14/RGNF − 2013 − 14 − SC − AND − 40670/(SAIII/Website)) and B. Uma Devi (F1 − 17.1/2012 − 13/RGNF − 2012 − 13 − SC − AND − 32761/(SA − III/WEBSITE)) are grateful to University Grants Commission New Delhi, India, for awarding the research fellowship as the Rajiv Gandhi National Fellowship − Scheme (RGNF/SRF) to carry out this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fct.2018.12.033.

References

Aras, A., Khokhar, A.R., Qureshi, M.Z., Silva, M.F., Sobczak-Kupiec, A., Pineda, E.A.G., Hechenleitner, A.A.W., Farooqi, A.A., 2014. Targeting cancer with nano-bullets: curcumin, EGCG, resveratrol and quercetin on flying carpets. Asian Pac. J. Cancer Prev. APJCP 15, 3865–3871.

Benfenati, E., 2010. The CAESAR project for in silico models for the REACH legislation. Chem. Cent. J. 4. https://doi.org/10.1186/1752-153x-4-s1-i1.

Benfenati, E., Manganaro, A., Gini, G., 2013. VEGA-QSAR: AI inside a platform for predictive toxicology. CEUR Workshop Proceedings, vol. 1107. CEUR-WS, pp. 21–28.

Benigni, R., Bossa, C., Jeliazkova, N.G., Netzeva, T.I., Worth, A.P., 2008. The Benigni/Bossa Rulebase for Mutagenicity and Carcinogenicity - a Module of Toxtree. Technical Report EUR 23241 EN. European Commission - Joint Research Centre.

Benigni, R., Bossa, C., 2008. Structure alerts for carcinogenicity, and the Salmonella assay system: a novel insight through the chemical relational databases technology. Mutat. Res. Rev. Mutat. Res. 659, 248–261.

Benigni, R., Bossa, C., 2011. Mechanisms of chemical carcinogenicity and mutagenicity: a review with implications for predictive toxicology. Chem. Rev. 111, 2507–2536.

Bode, A.M., Dong, Z., 2014. Toxic phytochemicals and their potential risks for human cancer. Cancer Prev. Res. 8, 1–8.

Bonofiglio, D., Giordano, C., Amicis, F.D., Lanzino, M., Andò, S., 2016. Natural products as promising antitumoral agents in breast cancer: mechanisms of action and molecular targets. Mini Rev. Med. Chem. 16, 596–604.

Casanova, M., 1999. Developmental effects of dietary phytoestrogens in Sprague-Dawley rats and interactions of genistein and daidzein with rat estrogen receptors alpha and beta in vitro. Toxicol. Sci. 51, 236–244.

Cassano, A., Manganaro, A., Martin, T., Young, D., Piclin, N., Pintore, M., Bigoni, D., Benfenati, E., 2010. CAESAR models for developmental toxicity. Chem. Cent. J. 4 (Suppl. 1), S4.

Chaudhry, Q., Piclin, N., Cotterill, J., Pintore, M., Price, N.R., Chrétien, J.R., Roncaglioni, A., 2010. Global QSAR models of skin sensitisers for regulatory purposes. Chem. Cent. J. 4 (Suppl. 1), S5.

Chen, A., Rogan, W.J., 2004. ISOFLAVONES IN SOY INFANT FORMULA: a review of evidence for endocrine and other activity in infants. Annu. Rev. Nutr. 24, 33–54.

Chu, H.Y., Wegel, E., Osbourn, A., 2011. From hormones to secondary metabolism: the emergence of metabolic gene clusters in plants. Plant J. 66, 66–79.

Cooper, J.A., Saracci, R., Cole, P., 1979. Describing the validity of carcinogen screening tests. Br. J. Canc. 39, 87–89.

Doerge, D.R., Sheehan, D.M., 2002. Goitrogenic and estrogenic activity of soy isoflavones. Environ. Health Perspect. 110, 349–353.

Ekor, M., 2014. The growing use of herbal medicines: issues relating to adverse reactions and challenges in monitoring safety. Front. Pharmacol. 4.

Fenton, T.R., Armour, B., Thirsk, J., 2015. Comment on "modulation of metabolic detoxification pathways using foods and food-derived components: a scientific review with clinical application. J Nutr Metab 1–2.

Ferrari, T., Gini, G., 2010. An open source multistep model to predict mutagenicity from statistical analysis and relevant structural alerts. Chem. Cent. J. 4 (Suppl. 1), S2.

Fink-Gremmels, J., 2010. Defense mechanisms against toxic phytochemicals in the diet of domestic animals. Mol. Nutr. Food Res. 54, 249–258.

Fjodorova, N., Vračko, M., Novič, M., Roncaglioni, A., Benfenati, E., 2010. New public QSAR model for carcinogenicity. Chem. Cent. J. 4 (Suppl. 1), S3.

Girish, C., Pradhan, S.C., 2011. Indian herbal medicines in the treatment of liver diseases: problems and promises. Fundam. Clin. Pharmacol. 26, 180–189.

Guerrero-Bosagna, C.M., Skinner, M.K., 2014. Environmental epigenetics and phytoestrogen/phytochemical exposures. J. Steroid Biochem. Mol. Biol. 139, 270–276.

Jezierska, A., Vračko, M., Basak, S.C., 2004. Counter-propagation artificial neural network as a tool for the independent variable selection: structure-mutagenicity study on aromatic amines. Mol. Divers. 8, 371–377.

Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27–30.

Kasote, D.M., Katyare, S.S., Hegde, M.V., Bae, H., 2015. Significance of antioxidant potential of plants and its relevance to therapeutic applications. Int. J. Biol. Sci. 11, 982–991.

Kennedy, D.O., Wightman, E.L., 2011. Herbal extracts and phytochemicals: plant secondary metabolites and the enhancement of human brain function. Adv Nutr 2, 32–50.

Korobkova, E.A., 2015. Effect of natural polyphenols on CYP metabolism: implications for diseases. Chem. Res. Toxicol. 28, 1359–1390.

Kumar, K.K., Devi, B.U., Neeraja, P., 2017. Integration of in silico approaches to determination of endocrine-disrupting perfluorinated chemicals binding potency with steroidogenic acute regulatory protein. Biochem. Biophys. Res. Commun. 491, 1007–1014.

Kumar, K.K., Devi, B.U., Neeraja, P., 2018. Elucidation of endocrine − disrupting polychlorinated biphenyls binding potency with steroidogenic genes: integration of in silico methods and ensemble docking approaches. Ecotoxicol. Environ. Saf. 165, 194–201.

Kumpulainen, J., 1999. Preface. Natural antioxidants and anticarcinogens in nutrition. Health and Disease v-vi. https://doi.org/10.1016/b978-1-85573-793-8.50003-0.

Lavecchia, T., Rea, G., Antonacci, A., Giardi, M.T., 2013. Healthy and adverse effects of plant- derived functional metabolites: the need of revealing their content and bioactivity in a complex food matrix. Crit. Rev. Food Sci. Nutr. 53, 198–213.

Messina, M., 2014. Soy foods, isoflavones, and the health of postmenopausal women. Am. J. Clin. Nutr. 100. https://doi.org/10.3945/ajcn.113.071464.

Monteiro, J., Alves, M., Oliveira, P., Silva, B., 2016. Structure-bioactivity relationships of methylxanthines: trying to make sense of all the promises and the drawbacks. Molecules 21, 974.

Morito, K., Aomori, T., Hirose, T., Kinjo, J., Hasegawa, J., Ogawa, S., Inoue, S., Muramatsu, M., Masamune, Y., 2002. Interaction of phytoestrogens with estrogen receptors α and β (II). Biol. Pharm. Bull. 25, 48–52.

Morito, K., Hirose, T., Kinjo, J., Hirakawa, T., Okawa, M., Nohara, T., Ogawa, S., Inoue, S., Muramatsu, M., Masamune, Y., 2001. Interaction of phytoestrogens with estrogen receptors .ALPHA. And .BETA. Biol. Pharm. Bull. 24, 351–356.

Newell-Mcgloughlin, M., 2008. Nutritionally improved agricultural crops. Plant Physiol. 147, 939–953.

OECD, 2004. Environmental Directorate. Report from the Expert Group on Quantitative Structure – Activity Relationships (Q)SAR on the Principles of Validation of (Q)SARs. (ENV/JM/MONO), Series on Testing and Assessment, No. 49.

Passerini, L., 2003. QSARs for Individual Classes of Chemical Mutagens and Carcinogens. Quantitative Structure-Activity Relationship (QSAR) Models of Mutagens and Carcinogens. https://doi.org/10.1201/9780203010822.ch3.

Plošnik, A., Zupan, J., Vračko, M., 2015. Evaluation of toxic endpoints for a set of cosmetic ingredients with CAESAR models. Chemosphere 120, 492–499.

Rahal, A., Kumar, A., Singh, V., Yadav, B., Tiwari, R., Chakraborty, S., Dhama, K., 2014. Oxidative stress, prooxidants, and antioxidants: the interplay. BioMed Res. Int. 1–19.

Rallo, R., France, B., Liu, R., Nair, S., George, S., Damoiseaux, R., Giralt, F., Nel, A., Bradley, K., Cohen, Y., 2011. Self-organizing map analysis of toxicity-related cell signaling pathways for metal and metal oxide nanoparticles. Environ. Sci. Technol. 45, 1695–1702.

Rasulev, B.F., Abdullaev, N.D., Syrov, V.N., Leszczynski, J., 2005. A quantitative structure- activity relationship (QSAR) study of the antioxidant activity of flavonoids. QSAR Comb. Sci. 24, 1056–1065.

Roncaglioni, A., Piclin, N., Pintore, M., Benfenati, E., 2008. Binary classification models for endocrine disrupter effects mediated through the estrogen receptor†. SAR QSAR Environ. Res. 19, 697–733.

Sushko, I., Pandey, A., Novotarskyi, S., Körner, R., Rupp, M., Teetz, W., Brandmaier, S., Abdelaziz, A., Prokopenko, V., Tanchuk, V., Todeschini, R., Varnek, A., Marcou, G., Ertl, P., Potemkin, V., Grishina, M., Gasteiger, J., Baskin, I., Palyulin, V., Radchenko, E., Welsh, W., Kholodovych, V., Chekmarev, D., Cherkasov, A., Aires-De-Sousa, J., Zhang, Q.Y., Bender, A., Nigsch, F., Patiny, L., Williams, A., Tkachenko, V., Tetko, I., 2011. Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. J. Comput. Aided Mol. Des. 25, 533–554.

Vračko, M., Bobst, S., 2013. Performance evaluation of CAESAR-QSAR output using PAHs as a case study. J. Chemometr. 28, 100–107.

Wanwimolruk, S., Prachayasittikul, V., 2014. Cytochrome P450 enzyme mediated herbal drug interactions (Part 1). EXCLI J 13, 347–391.

Winkler, D.A., 2002. The role of quantitative structure–activity relationships (QSAR) in biomolecular discovery. Briefings Bioinf. 3, 73–86.