Fundamental Toxicological Sciences (Fundam. Toxicol. Sci.)
Vol.10, No.7, 287-293, 2023

The Japanese Society of Toxicology

Fundamental Toxicological Sciences
URL : http://www.fundtoxicolsci.org/index_e.html

Original Article

Development of an *in silico* consensus model for predicting the chemical reactivity to cysteine measured by the DPRA and its application to predict the skin sensitization potential of chemicals

Koji Jojima[1] and Yumi Akahori[2]

[1]Chemicals Assessment and Research Center, Chemicals Evaluation and Research Institute, Japan, 1600 Shimotakano, Sugito-machi, Kitakatsushika-gun, Saitama 345-0043 Japan
[2]Chemicals Assessment and Research Center, Chemicals Evaluation and Research Institute, Japan, 1-4-25 Koraku,Bunkyo-ku,Tokyo 112-0004 Japan

(Received November 24, 2023; Accepted December 5, 2023)

ABSTRACT — Covalent binding of chemicals to skin proteins represents the Molecular Initiating Event (MIE) of the skin sensitization process. We attempted to construct *in silico* models for predicting the reactivities of chemicals to cysteine measured by the Direct Peptide Reactivity Assay (DPRA) as a screening tool for skin sensitization potential of chemicals since there was no readily available *in silico* prediction model for reactivity classes of the DPRA. We used a dataset of 211 chemicals compiled based on the chemical reactivity to cysteine in the DPRA for model construction, and each chemical was classified as "Minimal-Low" or "Moderate-High" reactivity according to the percent cysteine depletion value in the DPRA. We constructed two independent classification models using two machine learning algorithms named Random Forest (RF) and Graph Convolutional Network (GCN), and a consensus model adopting prediction results when both of the GCN-based and the RF-based models were matched was also constructed. Performance evaluation showed that the RF-based model showed higher specificity than the GCN-based model and the GCN-based model showed higher sensitivity than the RF-based model. The consensus model showed high accuracy and high specificity of over 0.9. Comparison of the reactivity class predicted by the consensus model and the skin sensitization potential for humans revealed that all chemicals classified into the "Moderate-High" class were human skin sensitizers. In conclusion, the consensus model we constructed here may be a promising *in silico* screening tool to predict cysteine reactivity measured by the DPRA and skin sensitization potential of chemicals.

Key words: In silico prediction, Reactivity to cysteine, DPRA, Skin sensitization

INTRODUCTION

Covalent binding of chemicals to macromolecules is one of the so-called Molecular Initiating Event (MIE) of toxicological responses. Covalent binding of a chemical to proteins is well-recognized as the MIE or the first Key Event (KE1) of the Adverse Outcome Pathway (AOP) for skin sensitization. The AOP for skin sensitization consists of 4 Key Events (KEs): the covalently bound chemical with the carrier proteins located within the skin can activate response pathways in the keratinocytes (KE2) and/or is recognized by dendritic cells (KE3), leading to T-cell proliferation (KE4) in the sensitization phase (OECD, 2012). Therefore, evaluation of the potential of chemicals

Correspondence: Koji Jojima (E-mail: jojima-koji@ceri.jp)

for covalent binding to proteins would be useful for predicting their skin sensitization potential.

The Direct Peptide Reactivity Assay (DPRA) has been developed for evaluating the skin sensitization potential of chemicals by addressing the MIE of the AOP for skin sensitization (Gerberick *et al*., 2004). The DPRA was validated by the European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM) and was adopted as the OECD Guidelines for the Testing of Chemicals (TG) No. 442C (TG 442C: *In Chemico* Skin Sensitization) (OECD, 2022). In the DPRA, model synthetic peptides containing either cysteine or lysine are used as surrogates for the skin proteins targeted by skin sensitizers. The test chemical is incubated by these synthetic peptides, and the percent cysteine peptide depletion values (Cys % depletion) and percent lysine peptide depletion values (Lys % depletion) are calculated as parameters of the chemical reactivities of the test chemicals to cysteine and lysine, respectively. The Mean of the Cys % depletion and Lys % depletion values, or the Cys % depletion value alone, is used to classify into one of four reactivity classes used to support the discrimination between sensitizers and non-sensitizers under OECD TG442C (Table 1) (OECD, 2022).

In recent years, *in silico* methodologies have been widely applied to the prediction of chemical reactivity for evaluating skin sensitization (Wilm *et al*., 2018). As described above, the DPRA evaluates the reactivity of a chemical to a cysteine and lysine that triggers skin sensitization, therefore *in silico* prediction of the reactivity measured by DPRA may be a cost and time saving application to predict sensitization potential of chemicals. However, there is no readily available *in silico* prediction model for reactivity classes of the DPRA.

In the early stage of development of the DPRA, it was reported that the maximal sensitivity to detect skin sensitizers was obtained when cysteine peptides were used with a cutoff value of 10% depletion (Gerberick *et al*., 2004), suggesting that the use of cysteine peptides is promising for prediction of the chemical reactivities of test chemicals. Therefore, we selected chemical reactivity to cysteine as the primary target for developing our prediction system for skin sensitization.

In this study, we attempted to construct an *in silico* model for predicting the chemical reactivities of test chemicals to cysteine measured by the DPRA as a screening tool for the skin sensitization potential of the test chemicals. At first, we compiled a dataset of the Cys % depletion values determined by DPRA for a variety of chemicals from publications. Second, we developed two prediction models to classify the reactivities of the chemicals to cysteine by using two machine learning algorithms called the Random Forest (RF) (Breiman, 2001), which consists of numerous decision trees and can interpret the importance of each descriptor, and the Graph Convolutional Network (GCN) (Kipf and Welling, 2017), which is a graph-based neural network algorithm that can handle chemical structures as a graph with atom-based descriptors. Then, we also constructed a consensus model adopting the prediction results that were matched between the GCN-based and RF-based models. We then evaluated the individual performances of the GCN-based and RF-based models, as also the performance of the consensus model. Finally, the predicted class of reactivity of the test chemicals to cysteine by the consensus model and the skin sensitization potential of the chemicals in humans was compared to demonstrate the usefulness of the constructed consensus model.

MATERIALS AND METHODS

Dataset of Cys % depletion values

The Cys % depletion values determined by DPRA and CAS Registry Number (CAS RN) of a variety of chemicals were collected from published articles (Lalko *et al*., 2013; Yamashita *et al*., 2015; Dik *et al*., 2016; Kawakami *et al*., 2020; Hoffmann *et al*., 2022; Urbisch *et al*., 2015).

Duplicated data from the primary source (Natsch *et al*., 2013) were found from two secondary sources (Hoffman *et al*., 2022 and Urbisch *et al*., 2015), and these duplicat-

Table 1. Reactivity classes based on DPRA described in OECD TG 442C.

Reactivity class	Mean of Cys % depletion and Lys % depletion	Cys % depletion	Prediction for skin sensitization by DPRA*
High	42.47 < and ≤ 100	98.24 < and ≤ 100	
Moderate	22.62 < and ≤ 42.47	23.09 < and ≤ 98.24	Skin sensitizer
Low	6.38< and ≤ 22.62	13.89 < and ≤ 23.09	
Minimal	0 ≤ and ≤ 6.38	0 ≤ and ≤ 13.89	Non-sensitizer

*According to OECD TG 442C, if both the Cys % depletion and Lys % depletion values are determined, then the mean of these two values is used for predicting the skin sensitization potential. If only the Cys % depletion value is determined, but not the Lys % depletion value, the Cys % depletion value alone is used for predicting the skin sensitization potential.

ed data were handled as one data. Chemicals not included in these articles were further collected from two databases, namely, Chemical Reactivity by COLIPA (from OECD QSAR Toolbox ver. 4.5) (Dimitrov *et al*., 2016) and Reference Data Matrix and Comparison (Annex 2) in Series on Testing and Assessment No.336 (RDMC) (OECD, 2021), and the Cys % depletion and CAS RN of these chemicals were also added to the primary dataset for this study. The primary dataset consisted of the Cys % depletion values for 527 data containing unique 351 CAS RNs. For structure information, Simplified Molecular Input Line Entry System (SMILES) strings of chemicals were collected by reference to the CAS RN using Power User Gateway (PUG) REST in PubChem (Kim *et al*., 2015). In order to create the final dataset for building models targeting the cysteine reactivity of chemicals, chemicals that met at least one of the following five conditions were removed from the dataset; (1) chemicals with no SMILES string; (2) multi-component chemicals; (3) metal-containing chemicals; (4) chemicals with potential for autoxidation (predicted using Autoxidation Simulator ver. 3.5 from OECD QSAR Toolbox ver. 4.5); and (5) acid anhydrides. Condition (1) was included as an exclusion criterion, since SMILES was essential to generate descriptors used as input into the prediction models. Conditions (2) and (3) allowed focusing the prediction target on mono-constituent organic chemicals. Conditions (4) and (5) were set to remove unstable chemicals which are readily oxidized or hydrolyzed. In cases with multiple reported Cys % depletion values, the maximum reported value was used for subsequent classification from the perspective of safety. Finally, we complied a final dataset consisting of the Cys % depletion values and structural information for 211 chemicals (Supplementary Table S1).

Classification of the chemical reactivity to cysteine

All the chemicals included in the final dataset were classified into 2 classes based on their Cys % depletion values, adopting the threshold of 23.09% described in OECD TG 442C. Chemicals with Cys % depletion values of more than 23.09% were classified as "Moderate-High" (Moderate or High reactivity) class and chemicals with Cys % depletion values of 23.09% or less were classified into the "Minimal-Low" (Minimal or Low reactivity) class. In this way, of the 211 chemicals included in the dataset, 78 were classified into the "Moderate-High" class, and the remaining 133 were classified into the "Minimal-Low" class.

Descriptors

Semi-empirical quantum chemical calculations were performed for each chemical with PM6 using Gaussian16 (Frisch *et al*., 2016), to calculate the Highest Occupied Molecular Orbital (HOMO) energy, second HOMO (HOMO-1) energy, third HOMO (HOMO-2) energy, Lowest Unoccupied Molecular Orbital (LUMO) energy, second LUMO (LUMO+1) energy, and third LUMO (LUMO+2) energy. The molecular weight and Crippen's logP of each chemical were calculated using RDKit (ver. 2022.09.4). These 8 descriptors were used as molecular-based descriptors for both the RF-based and GCN-based models. In addition to these descriptors, for the RF-based model, counts of each fragment generated by using the Morgan fingerprint algorithm (radius = 4) were calculated by RDKit. For the GCN-based model, atom-based descriptors (Supplementary Table S2) were obtained from the results of calculation of the PM6 using Gaussian 16 or calculated by RDKit.

Model construction

By performing stratified sampling based on clustering of chemical structure with k-medoids (partition count =5) using Morgan fingerprint (1024 bits, radius = 4) as the clustering descriptor, the final dataset was divided into training, validation and test datasets (n = 131, n = 40, and n = 40, respectively), while maintaining the structural diversity of the chemicals. For model construction, the training dataset was used for learning process. Also, the validation dataset was used for hyperparameter optimization of the RF-based model and early stopping of the learning process of the GCN-based model, to avoid overfitting. For external evaluation of the constructed models, the test dataset was used. Of the 131, 40 and 40 chemicals included in the training, validation, and test datasets, respectively, 35.9% (47/131), 35.0% (14/40), and 42.5% (17/40), respectively, were classified into the "Moderate-High" class. The Morgan fingerprint generation, clustering and splitting processes were conducted using the KNIME Analytics platform (ver. 4.3.2) (Berthold *et al*., 2008).

Scikit-learn (ver. 1.2.1) (Pedregosa *et al*., 2011) was used for construction of the RF-based model. Hyperparameter optimization of the RF-based model was conducted using Optuna (ver. 3.1.1) (Akiba *et al*., 2019). DGL (ver. 1.0.0) (Wang, 2019) was used for construction of the GCN-based model. For input into the GCN-based model, the structure of each chemical was represented as a graph, and atom-based descriptors were applied to each node of the graph. As the structure of the GCN-based model, 4 units with a graph convolution, a dropout, and

a ReLU layer as one unit were arranged first, followed by setting of a concatenate layer, linear transformation layer, and sigmoid function layer. In the GCN-based model, the graph with atom-based descriptors was transformed into 1 feature per chemical. This 1 feature was concatenated with the 8 molecular-based descriptors and converted to 1 output value by a linear transformation layer. A sigmoid function was applied to convert the output value in the range of 0 to 1. In the prediction of the reactivity of chemicals to cysteine by the GCN-based model, chemicals for which the converted output value was equal to or greater than 0.5 were assigned into the "Moderate-High" class, and others into the "Minimal-Low" class. Furthermore, we also set a combination model based on the RF-based and GCN-based models (named as the "Consensus model"). When the prediction results of the RF-based and GCN-based models were consistent, the results were adopted into the Consensus model. In the case of discrepancy of the prediction results between the two models, the result was termed "Inconclusive". Python (ver. 3.8.10) was used for each process of model construction.

Performance evaluations of the constructed models

The sensitivity (the rate of "Moderate-High" prediction outcomes for chemicals classified into the "Moderate" or "High" reactivity classes by the DPRA), specificity (the rate of "Minimal-Low" prediction outcomes for chemicals classified into the "Minimal" or "Low" reactivity classes by the DPRA), and accuracy (overall rate of correct prediction) of the models were calculated using the Cooper statistics (Cooper *et al*., 1979) for performance evaluation of constructed models. For evaluation of the Consensus model, chemicals with an "Inconclusive" result were excluded from the calculation, and the applicability (rate of chemicals with prediction outcomes of "Moderate-High" or "Minimal-Low") was calculated.

Comparison of the chemical reactivity to cysteine and the skin sensitization potential of chemicals

The prediction results by the Consensus model were compared to the known skin sensitization classification information for humans to demonstrate the usefulness of this model. Expert-derived skin sensitization hazard for humans (referring to the column named "Basketter_human_Call") provided in the RDMC (OECD, 2021) were used for referring to the known skin sensitization classification for each chemical. A final dataset of 56 chemicals for which information on the skin sensitization classification in the RDMC was known were used for this comparison. These 56 chemicals included 29 (16 skin sensitizers and 13 non-sensitizers), 14 (11 skin sensitizers and 3 non-sensitizers), and 13 (9 skin sensitizers and 4 non-sensitizers) chemicals of the training, validation, and test datasets, respectively.

RESULTS AND DISCUSSION

Performance of the constructed models for classifying the reactivity of chemicals to cysteine

The performances of the RF-based, GCN-based and the Consensus models for the training, validation, and test datasets are shown in Table 2. External validation using the test dataset revealed a sensitivity and specificity of 0.765 and 0.913, respectively, for the RF-based model, and 0.882 and 0.826, respectively, for the GCN-based model. The accuracy of both the RF-based and the GCN-based models was 0.850. The RF-based model showed a higher specificity as compared with that of the GCN-based model for all the datasets. On the other hand, the GCN-based model was characterized by a higher sensitivity as compared with that of the RF-based model for all the datasets. These results suggest that the Consensus model developed by combining the RF- and GCN-based models yields a better performance than either of the two models alone.

Table 2. Performances of each model for the training, validation, and test datasets.

	Training dataset (n = 131)			Validation dataset (n = 40)			Test dataset (n = 40)		
	RF	GCN	Consensus (n = 119)	RF	GCN	Consensus (n = 33)	RF	GCN	Consensus (n = 34)
Sensitivity	0.936	0.979	1.000	0.786	0.857	0.909	0.765	0.882	0.867
Specificity	0.976	0.929	1.000	0.923	0.846	0.955	0.913	0.826	0.947
Accuracy	0.962	0.947	1.000	0.875	0.850	0.939	0.850	0.850	0.912
Applicability	1.000	1.000	0.908	1.000	1.000	0.825	1.000	1.000	0.850

Table 3. Predicted and experimental chemical reactivity to cysteine and known human skin sensitization potential for each chemical.

Chemical Name	CAS RN	Prediction	Experimental [a]		Skin Sens [b]
		Reactivity class by the Consensus model	Reactivity class	Cysteine % depletion in DPRA	
Training dataset					
Benzoyl peroxide	94-36-0	Moderate-High	High	100	S
Diethyl maleate	141-05-9	Moderate-High	High	100	S
Iodopropynyl butylcarbamate	55406-53-6	Moderate-High	High	99.7	S
Methyl heptine carbonate	111-12-6	Moderate-High	High	99.2	S
Diphenylcyclopropenone	886-38-4	Inconclusive	High	98.83	S
Butyl glycidyl ether	2426-08-6	Inconclusive	Moderate	67.3	S
Formaldehyde	50-00-0	Moderate-High	Moderate	60.41	S
Methylmethacrylate	80-62-6	Moderate-High	Moderate	55.25	S
Tetrachlorosalicylanilide	1154-59-2	Moderate-High	Moderate	36.8	S
Thioglycerol	96-27-5	Inconclusive	Moderate	27.3	S
3-Propylidenephthalide	17369-59-4	Inconclusive	Low	14.3	S
Penicillin G	61-33-6	Minimal-Low	Low	14.3	S[c]
4-Aminobenzoic acid	150-13-0	Minimal-Low	Minimal	10.7	N
3-Dimethylamino propylamine	109-55-7	Minimal-Low	Minimal	10.18	S[c]
Salicylic acid	69-72-7	Minimal-Low	Minimal	8.71	N
Diethyl toluamide	134-62-3	Minimal-Low	Minimal	6.7	N
Benzyl salicylate	118-58-1	Minimal-Low	Minimal	6.325	N
Isopropyl myristate	110-27-0	Minimal-Low	Minimal	5.9	N
Benzyl cinnamate	103-41-3	Minimal-Low	Minimal	4.3	S[c]
Cinnamyl nitrile	1885-38-7	Inconclusive	Minimal	4.04	S
Hexyl salicylate	6259-76-3	Minimal-Low	Minimal	3.9	S[c]
Lactic acid	50-21-5	Minimal-Low	Minimal	3	N
1-Butanol	71-36-3	Minimal-Low	Minimal	0.7	N
Triethanolamine	102-71-6	Minimal-Low	Minimal	0.63	N
Propylene glycol	57-55-6	Minimal-Low	Minimal	0.35	N
Hexane	110-54-3	Minimal-Low	Minimal	0.33	N
Glycerol	56-81-5	Minimal-Low	Minimal	0	N
Isopropanol	67-63-0	Minimal-Low	Minimal	0	N
Pentachlorophenol	87-86-5	Minimal-Low	Minimal	0	N
Validation dataset					
Methyl 2-nonynoate	111-80-8	Moderate-High	High	100	S
Methyldibromo glutaronitrile	35691-65-7	Moderate-High	High	100	S
2-Mercaptobenzothiazole	149-30-4	Moderate-High	High	99.88	S
Methylisothiazolinone	2682-20-4	Moderate-High	Moderate	97.9	S
1,2-Benzisothiazolin-3-one	2634-33-5	Moderate-High	Moderate	97.65	S
Ethyleneglycol dimethacrylate	97-90-5	Moderate-High	Moderate	93.47	S
Benzocaine	94-09-7	Minimal-Low	Moderate	29.2	S[c]
5-Methyl-2,3-hexanedione	13706-86-0	Inconclusive	Moderate	25.79	S
Amylcinnamyl alcohol	101-85-9	Minimal-Low	Low	23	S[c]
Ethylene diamine	107-15-3	Minimal-Low	Low	18.6	S[c]
Diethyl phthalate	84-66-2	Minimal-Low	Minimal	0.76	N
DMSO	67-68-5	Minimal-Low	Minimal	0.4	N
Anethole	104-46-1	Minimal-Low	Minimal	0	N
Aniline	62-53-3	Minimal-Low	Minimal	0	S[c]
Test dataset					
2,4-Dinitrochlorobenzen	97-00-7	Moderate-High	High	100	S
Dimethyl fumarate	624-49-7	Moderate-High	High	100	S
Ethyl acrylate	140-88-5	Moderate-High	High	100	S
2-Hydroxyethyl acrylate	818-61-1	Moderate-High	Moderate	92.64	S
Imidazolidinyl urea	39236-46-9	Moderate-High	Moderate	59	S
Glyoxal	107-22-2	Moderate-High	Moderate	56.5	S
Bisphenol A-diglycidyl ether	1675-54-3	Inconclusive	Moderate	42.5	S
Hydrocortisone	50-23-7	Minimal-Low	Moderate	39.1	N
Coumarin	91-64-5	Minimal-Low	Minimal	7	S[c]
Pyridine	110-86-1	Minimal-Low	Minimal	1.5	N
Benzyl benzoate	120-51-4	Minimal-Low	Minimal	0.88	N
Allyl phenoxyacetate	7493-74-5	Minimal-Low	Minimal	0.61	S[c]
Octanoic acid	124-07-2	Minimal-Low	Minimal	0	N

[a] Reactivity class based on the percent cysteine depletion (Cys % depletion) values by the DPRA using the criteria shown in Table 1.

[b] Skin sensitization potential in humans (Basketter *et al*., 2014) collected from the Reference Data Matrix and Comparison (Annex 2) in Series on Testing and Assessment No.336. Skin sensitizers are represented as "S," and non-sensitizers are represented as "N."

[c] Skin sensitizer predicted as the "Minimal-Low" class by the Consensus model.

In general, there is a problem of trade-off between the applicability and prediction performance in model selection. The applicability values of the Consensus model to the training, validation and test datasets were 0.908, 0.825 and 0.850, lesser than the corresponding values for either of the RF-based and GCN-based models, because the prediction outcomes for several chemicals were inconsistent between the RF-based and GCN-based models. However, although the sensitivity of the Consensus model (0.867) was slightly lower than that of the GCN-based model in the test dataset, both the specificity (0.947) and accuracy (0.912) of the Consensus model were higher than those of either the RF-based or GCN-based model alone.

Comparison of the predicted chemical reactivity to cysteine and the skin sensitization potential for human

Table 3 shows the predicted and experimental chemical reactivity of chemicals to cysteine and the known skin sensitization potential in humans of 56 chemicals of the final dataset. Among the 56 chemicals, the result for 7 chemicals was "Inconclusive" by the Consensus model, and these chemicals were excluded from this comparison.

Among the remaining 49 chemicals, 19 chemicals were classified into the "Moderate-High" class, and all of these 19 chemicals were known skin sensitizers for humans. The remaining 30 chemicals were classified into the "Minimal-Low" class by the Consensus model; while 20 of these 30 chemicals were known non-sensitizers for humans, the remaining 10 classified into the "Minimal-Low" class were known skin sensitizers for humans. Among these 10 chemicals, 9 were classified into the "Minimal" or "Low" class, according to the Cys % depletion value experimentally determined by the DPRA, according to the criteria shown in Table 1. This means that the results for these 9 skin sensitizers were consistent between their reactivity classes to cysteine determined by the DPRA ("Minimal" or "Low") and the predicted reactivity class by Consensus model ("Minimal-Low"). The remaining one skin sensitizer, benzocaine (CAS RN: 94-09-7), was classified into the "Moderate" reactivity class according to the Cys % depletion value of 29.2 in the DPRA, but was underestimated as belonging to the "Minimal-Low" class by the Consensus model. However, it is notable that the experimental Cys % depletion value (29.2%) of benzocaine was close to the lower limit of the Cys % depletion range for the "Moderate" class ($23.09 < \text{Cys \% depletion} \leq 98.24$) in DPRA (Table 1).

Thus, we confirmed that 19 chemicals predicted as the "Moderate-High" class by the Consensus model were known human sensitizers, and 29 out of 30 chemicals classified into the "Minimal-Low" class by the Consensus model were either known non-sensitizers or classified into the "Minimal" or "Low" reactivity class by the DPRA, based on the Cys% depletion value, except for one chemical, benzocaine, for which the Cys % depletion was close to the lower limit of the range corresponding to the "Moderate" class.

In conclusion, we constructed two independent classification models for predicting the reactivity of chemicals to cysteine measured by the DPRA using the RF and GCN algorithms, and then a Consensus model based on the two algorithms since there was no readily available *in silico* prediction model for reactivity classes of the DPRA. Comparison of the constructed RF-based and GCN-based models revealed a higher specificity of the RF-based model as compared with that of the GCN-based model, and a higher sensitivity of the GCN-based model as compared with that of the RF-based model. Both the accuracy and specificity of the Consensus model, constructed based on a combination of the GCN-models and RF-based models, were high and over 0.9. Although the applicability was somewhat limited, we believe that the Consensus model developed in this study may provide useful information to predict the skin sensitization potential of chemicals in humans. Especially, all of the chemicals predicted as the "Moderate-High" class by the Consensus model were known skin sensitizers for humans. Therefore, our Consensus model may be a promising *in silico* screening tool to predict the reactivity of chemicals to cysteine measured by the DPRA and their skin sensitization potential in humans.

Conflict of interest---- The authors declare that there is no conflict of interest.

REFERENCES

Akiba, T., Sano, S., Yanase, T., Ohta, T. and Koyama, M. (2019): Optuna: A next-generation hyperparameter optimization framework. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2623-2631.

Basketter, D.A., Alépée, N., Ashikaga, T., Barroso, J., Gilmour, N., Goebel, C., Hibatallah, J., Hoffmann, S., Kern, P., Martinozzi-Teissier, S., Maxwell, G., Reisinger, K., Sakaguchi, H., Schepky, A., Tailhardat, M. and Templier, M. (2014): Categorization of chemicals according to their relative human skin sensitizing potency. Dermatitis, **25**, 11-21.

Berthold, M., Cebron, N., Dill, F., Gabriel, T., K"otter, T., Meinl, T., Ohl, P., Sieb, C., Thiel, K., and Wiswedel, B. (2008): KNIME: The Konstanz Information Miner. In Data Analysis, Machine Learning and Applications (Preisach, C., Burkhardt, H., Schmidt-Thieme, L., and Decker, R., eds.), pp. 319-326, Springer Berlin Heidelberg.

Breiman, L. (2001): Random forests. Mach. Learn., **45**, 5-32.

Cooper, J.A. 2nd, Saracci, R. and Cole, P. (1979): Describing the validity of carcinogen screening tests. Br. J. Cancer, **39**, 87-89.

Dik, S., Rorije, E., Schwillens, P., van Loveren, H. and Ezendam, J. (2016): Can the direct peptide reactivity assay be used for the identification of respiratory sensitization potential of chemicals? Toxicol. Sci., **153**, 361-371.

Dimitrov, S.D., Diderich, R., Sobanski, T., Pavlov, T.S., Chankov, G.V., Chapkanov, A.S., Karakolev, Y.H., Temelkov, S.G., Vasilev, R.A., Gerova, K.D., Kuseva, C.D., Todorova, N.D., Mehmed, A.M., Rasenberg, M. and Mekenyan, O.G. (2016): QSAR Toolbox - workflow and major functionalities. SAR QSAR Environ. Res., **27**, 203-219.

Frisch, M. ea, Trucks, G., Schlegel, H., Scuseria, G., Robb, M., Cheeseman, J., Scalmani, G., Barone, V., Petersson, G., Nakatsuji, H. and others. (2016): Gaussian 16, revision C. 01. Gaussian, Inc., Wallingford CT.

Gerberick, G.F., Vassallo, J.D., Bailey, R.E., Chaney, J.G., Morrall, S.W. and Lepoittevin, J.-P. (2004): Development of a peptide reactivity assay for screening contact allergens. Toxicol. Sci., **81**, 332-343.

Hoffmann, S., Alépée, N., Gilmour, N., Kern, P.S., van Vliet, E., Boislève, F., Bury, D., Cloudet, E., Klaric, M., Kühnl, J., Lalko, J.F., Mewes, K., Miyazawa, M., Nishida, H., Tam Brami, M.T., Varçin, M., Api, A.M. and Europe, C. (2022): Expansion of the Cosmetics Europe skin sensitisation database with new substances and PPRA data. Regul. Toxicol. Pharmacol., **131**, 105169.

Kawakami, T., Isama, K., Ikarashi, Y. and Jinno, H. (2020): Evaluation of the sensitization potential of volatile and semi-volatile organic compounds using the direct peptide reactivity assay. J. Toxicol. Sci., **45**, 725-735.

Kim, S., Thiessen, P.A., Bolton, E.E. and Bryant, S.H. (2015): PUG-SOAP and PUG-REST: web services for programmatic access to chemical information in PubChem. Nucleic Acids Res., **43** (W1), W605-W611.

Kipf, T.N. and Welling, M. (2017). Semi-supervised classification with graph convolutional networks. ArXiv Preprint ArXiv:1609.02907.

Lalko, J.F., Kimber, I., Dearman, R.J., Api, A.M. and Gerberick, G.F. (2013): The selective peptide reactivity of chemical respiratory allergens under competitive and non-competitive conditions. J. Immunotoxicol., **10**, 292-301.

Natsch, A., Ryan, C.A., Foertsch, L., Emter, R., Jaworska, J., Gerberick, F. and Kern, P. (2013): A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation. J. Appl. Toxicol., **33**, 1337-1352.

OECD. (2012): The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins Part 1: Scientific Evidence. Series on Testing and Assessment No. 168. Available at: https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD. (2021): Supporting Document to the OECD Guideline 497 on Defined Approaches for Skin Sensitisation. Series on Testing and Assessment No.336. Available at: https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD. (2022): OECD Guideline for the Testing of Chemicals No. 442C: In Chemico Skin Sensitisation-Assays addressing the Adverse Outcome Pathway key event on covalent binding to proteins. Available at: https://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788

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. and Duchesnay, É. (2011): Scikit-learn: machine learning in Python. The JMLR, **12**, 2825-2830.

Urbisch, D., Mehling, A., Guth, K., Ramirez, T., Honarvar, N., Kolle, S., Landsiedel, R., Jaworska, J., Kern, P.S., Gerberick, F., Natsch, A., Emter, R., Ashikaga, T., Miyazawa, M. and Sakaguchi, H. (2015): Assessing skin sensitization hazard in mice and men using non-animal test methods. Regul. Toxicol. Pharmacol., **71**, 337-351.

Wang, M.Y. (2019). Deep graph library: Towards efficient and scalable deep learning on graphs. ICLR Workshop on Representation Learning on Graphs and Manifolds.

Wilm, A., Kühnl, J. and Kirchmair, J. (2018): Computational approaches for skin sensitization prediction. Crit. Rev. Toxicol., **48**, 738-760.

Yamashita, K., Shinoda, S., Hagiwara, S., Miyazaki, H. and Itagaki, H. (2015): Unsaturated fatty acids show clear elicitation responses in a modified local lymph node assay with an elicitation phase, and test positive in the direct peptide reactivity assay. J. Toxicol. Sci., **40**, 843-853.