J. Pestic. Sci. **47**(4), 184–189 (2022)
DOI: 10.1584/jpestics.D22-043

JOURNAL OF PESTICIDE SCIENCE

Regular Article

Reconfiguring the online tool of SkinSensPred for predicting skin sensitization of pesticides

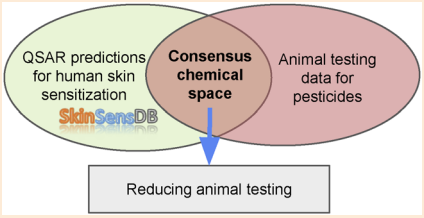
Chia-Chi Wang,[1] Shan-Shan Wang,[2] Chun-Lin Liao,[3] Wei-Ren Tsai[3,*] and Chun-Wei Tung[2,4,*]

[1] *Department and Graduate Institute of Veterinary Medicine, School of Veterinary Medicine, National Taiwan University, Taipei 10617, Taiwan*
[2] *Institute of Biotechnology and Pharmaceutical Research, National Health Research Institutes, Miaoli County 35053, Taiwan*
[3] *Taiwan Agricultural Chemicals and Toxic Substances Research Institute, Council of Agriculture, Taichung 41358, Taiwan*
[4] *Graduate Institute of Data Science, College of Management, Taipei Medical University, Taipei 10675, Taiwan*

(Received July 7, 2022; Accepted August 22, 2022)

Supplementary material

Adverse outcome pathway (AOP)-based computational models provide state-of-the-art prediction for human skin sensitizers and are promising alternatives to animal testing. However, little is known about their applicability to pesticides due to scarce pesticide data for evaluation. Moreover, pesticides traditionally have been tested on animals without human data, making validation difficult. Direct application of AOP-based models to pesticides may be inappropriate since their original applicability domains were designed to maximize reliability for human response prediction on diverse chemicals but not pesticides. This study proposed to identify a consensus chemical space with concordant human responses predicted by the SkinSensPred online tool and animal testing data to reduce animal testing. The identified consensus chemical space for non-sensitizers achieved high concordance of 85% and 100% for the cross-validation and independent test, respectively. The reconfigured SkinSensPred can be applied as the first-tier tool for identifying non-sensitizers to reduce. animal testing for pesticides by 19.6%.



Keywords: adverse outcome pathway, skin sensitization, machine learning, SkinSensPred, pesticides, 3R.

Introduction

Skin sensitization is a key endpoint considered by chemical and pesticide regulatory authorities. Small molecules acting as haptens can form a complex with proteins that may trigger T cell-mediated immune reactions and lead to allergic contact dermatitis.[1)] Until the European Union's 2013 ban on animal testing for cosmetics, assessment protocols for skin sensitization have long relied on animal tests, such as the guinea pig maximization test (GPMT) and the murine local lymph node assay (LLNA). Following the 3R (replacement, reduction, and refinement) principle, alternative non-animal assays are currently the preferred method for assessing the skin sensitization of chemicals.[2–4)]

Among the alternative methods, computational models such as structural alerts (SAs) and quantitative structure–activity relationship (QSAR) models are the most cost-effective methods and are generally acceptable for regulatory use.[4)] Novel algorithms and software are continuously proposed for improving the prediction of skin sensitizers.[5,6)] While SA-based methods are useful, chemicals without SAs may not be considered non-sensitizers. Also, traditional end-to-end QSAR models provide no mechanism information. For improving regulatory acceptance, the emerging adverse outcome pathway (AOP) concept has been integrated into the prediction models of skin sensitization for mechanism interpretation and performance improvement.

Tung *et al.* proposed the first AOP-based computational method[7)] for predicting skin sensitizers by aggregating individual read-across results from SkinSensDB[8)] with three key events—protein binding, keratinocyte activation, and the activation of dendritic cells. The read-across method offers improved prediction with the predicted activation status of the

* To whom correspondence should be addressed.
E-mail: sftsai@tactri.gov.tw, cwtung@nhri.edu.tw
Published online November 1, 2022

three key events for mechanism interpretation. SkinSensPred[9] introduced an advanced multitask ExtraTree algorithm to simultaneously train the target task of human skin sensitization and three relevant learning tasks corresponding to the three key events. The knowledge-sharing mechanism among the target and three relevant tasks along with the incorporation of SA information achieved state-of-the-art performance for human skin sensitizers.[9] Similarly, the Pred-Skin 3.0 online tool provides improved performance for human skin sensitizers over previous models by developing a Bayesian-based meta classifier to integrate five QSAR models representing the three key events, LLNA, and human skin sensitization.[10]

With increased interest in replacing animal testing with computational methods for pesticide regulation, the AOP-based computational methods are promising alternatives to animal testing. However, most of the datasets are of general chemicals, pharmaceuticals, and cosmetics. The predictive performance of AOP-based models on pesticides is largely unknown. A previous evaluation of skin sensitization for pesticides using traditional SA and end-to-end QSAR models showed a low-to-moderate predictive performance.[11] Therefore, it is interesting to investigate the predictive performance of AOP-based QSAR methods on pesticides. Since the existing skin sensitization information for pesticides is mostly based on animal testing, which is different from the human response endpoint predicted by modern AOP-based tools, the validation and interpretation of the predictive results and determining the way to incorporate the tools for regulatory use are challenging tasks.

This study first evaluated a list of 143 pesticides using SkinSensPred. As expected, only 11 pesticides fell within the applicability domain (AD) of SkinSensPred. Among them, three pesticides with corresponding human data from the Hazardous Substances Data Bank (HSDB) are all correctly predicted by SkinSensPred. However, other predictions without human data are difficult to validate, which impedes their regulatory adaptation. To solve this issue, we proposed to identify the consensus chemical space with concordant animal testing data and predicted human responses by reconfiguring the AD of SkinSensPred. The 143 pesticides were randomly divided into a training dataset and a test dataset at a 2:1 ratio. Exclusion rules for chemicals outside of the consensus chemical space were derived by using a decision tree-based algorithm[12,13] with nine different fingerprints based on the training dataset, and their predictive performance was evaluated using the test dataset. The test result showed that the proposed method using standard fingerprints is effective in identifying non-sensitizers with a high concordance of 100%. The overall coverage of the consensus chemical space of 143 pesticides is 19.6%. With the consensus chemical space reconfigured, SkinSensPred can be applied as the first-tier tool for identifying non-sensitizers, thus reducing animal testing for pesticides.

Table 1. The type and skin sensitization category of the utilized dataset

Type	Category of skin sensitization	Number of chemicals
Acaricide, miticide insecticide	Sensitizer	19
	Non-sensitizer	48
Fungicide	Sensitizer	22
	Non-sensitizer	40
Herbicide	Sensitizer	4
	Non-sensitizer	10

Materials and methods

1. Dataset

In total, 143 active ingredients of pesticides with skin sensitization data based on animal testing mostly were collated from five sources, including 125 pesticides from The Joint FAO/WHO Meeting on Pesticide Residues (JMPR, https://www.fao.org/agriculture/crops/thematic-sitemap/theme/pests/lpe/en/), six pesticides from the Food Safety Commission of Japan (FSCJ, https://www.fsc.go.jp/english/evaluationreports/agrichemicalsl_e1.html), 11 pesticides from the European Food Safety Authority (EFSA, https://www.efsa.europa.eu/en/microstrategy/openfoodtox), and one pesticide from the International Programme on Chemical Safety (IPCS) INCHEM Environmental Health Criteria Monographs (EHCs, https://inchem.org/pages/ehc.html). The detailed numbers for different types of pesticides are shown in Table 1, and detailed information for the dataset is shown in Table S1.

2. Fingerprint

Nine types of chemical fingerprints—including circular (ECFP6), standard (StandardFP), extended (ExtendedFP), graph (GraphFP), hybridization (HybridizationFP), Molecular ACCess System (MACCSFP), estate (EStateFP), pubchem (PubChemFP), and KlekotaRoth (KRFP)—were evaluated in this study to identify the consensus chemical space with concordant animal testing data and predicted human responses. Among them, StandardFP, ExtendedFP, GraphFP, HybridizationFP, and ECFP6 are hashed fingerprints with a default length of 1024 bits, where each bit represents the existence of specific substructures. ExtendedFP, GraphFP, and HybridizationFP are all similar to the StandardFP. ExtendedFP takes rings and atomic properties into account. GraphFP and HybridizationFP consider only connectivity and hybridization states, respectively. ECFP6 is widely utilized for QSAR model development. Due to bit collision issues, the direct structural interpretation of a specific bit is not possible. In contrast, the bit vectors encoded by MACCSFP, EStateFP, PubChemFP, and KRFP, with lengths of 166, 79, 881, and 4860, respectively, are interpretable. RCDK v3.4.7.2[14] based on the Chemistry Development Kit (CDK) library[15] was utilized to generate the nine types of chemical fingerprint.

3. Consensus chemical space with concordant animal testing data and predicted human response

The concept of identifying consensus chemical space with concordant animal testing data and predicted human responses is similar to determining the applicability domain (AD) of quantitative structure–activity relationship (QSAR) models. In this study, we used the state-of-the-art SkinSensPred method[9] to generate skin sensitization predictions in humans. The web server is freely available at https://cwtung.nhri.edu.tw/skinsensdb/predict. The AD of the original model aims to identify the chemical space with a reliable prediction of the target endpoint of a QSAR model, *i.e.*, a human response. Since our aim is to generate skin sensitization predictions that can be validated in animal models, the idea is to identify the consensus chemical space that can maximize the concordance between animal testing data and predicted human responses.

The consensus chemical space was determined by using a decision tree-based method[12,13] to derive classification rules for the exclusion of chemicals outside of the consensus chemical space. To avoid overfitting problems, the determination of the consensus chemical space was derived solely from the training dataset and independently tested using the test dataset. First, all testing chemicals were randomly divided into a training dataset and a test dataset with a 2 : 1 ratio for each pesticide class. Second, the human skin sensitization score and category (sensitizer or non-sensitizer) for each chemical in the training datasets were predicted with SkinSensPred without considering the original AD of the model. The scores ranged from 0 to 1. A chemical with a score greater than 0.5 is considered to be a sensitizer. In contrast, non-sensitizers are chemicals with scores of less than or equal to 0.5. Third, the concordance between SkinSensPred results and animal testing data was calculated and utilized as the label column. Fourth, the classification and regression tree (CART) algorithm[16] was applied to generate rules with corresponding concordance for each fingerprint. Fifth, rules for conflicts between SkinSensPred and animal testing data were utilized as exclusion rules for identifying chemicals outside of the consensus chemical space. Finally, the test dataset was applied to evaluate the exclusion rules. R environment v3.6.0 and packages of rpart v4.1-15[17] and partykit v1.2-6[18] were used to implement the decision tree-based method.

Results and discussion

1. Application of SkinSensPred to pesticides

The applicability of SkinSensPred to pesticides was first studied by utilizing the tool to predict skin sensitizers of the 143 collected pesticides. Due to the data availability issue, most prediction tools, including SkinSensPred, were trained on chemicals that are mostly general chemicals, pharmaceuticals, and cosmetics and were expected to have limited applicability to pesticides. The analysis results showed that only 11 pesticides fell within the applicability domain (AD) of SkinSensPred, with a very low coverage of 7.7%. Since SkinSensPred was designed to predict human skin sensitizers, it is reasonable to compare its predictions to human data rather than to animal data. Among the 11 pesticides, only three pesticides of thiophanate-methyl (CASN: 23564-05-8), ferbam (CASN: 14484-64-1), and carbaryl (CASN: 63-25-2) were associated with human data based on HSDB. All three pesticides are sensitizers and were all correctly predicted, which shows the usefulness of SkinSensPred prediction. However, the other eight predictions without human data conflicted with animal testing data. Both species differences and prediction errors could be responsible for the conflicts. For example, the abovementioned carbaryl is annotated as a non-sensitizer based on guinea pig data that is different from human data. Due to potential ethical issues for validating the predicted effects on human beings,[19] it may not be appropriate to directly adopt a human-based prediction for regulatory use. The scarcity of human data for pesticides makes regulatory adaptation difficult. To achieve a compromise between the advanced prediction model for human responses and a validatable outcome based on animal data, we proposed identifying the consensus chemical space with concordant animal testing data and predicted human responses by reconfiguring SkinSensPred.

2. Identification of the consensus chemical space

To determine the consensus chemical space, a decision tree-based method was applied to extract rules based on chemical fingerprints for identifying chemicals with skin sensitization results in which the predicted human response conflicted with the animal testing data. Nine chemical fingerprints were evaluated in this study. The 143 pesticides were randomly divided into a training dataset and a test dataset with 97 and 46 chemicals, respectively. An analysis of the concordance between the predicted human responses and animal testing data on the training dataset was conducted first. There are 52 and 45 chemicals predicted to be human skin sensitizers and non-sensitizers, respectively. Among the 52 predicted human skin sensitizers with a score greater than 0.5, only 17 (32.7%) pesticides are sensitizers based on animal data. In contrast, 31 of the 45 predicted human non-sensitizers with scores less than or equal to 0.5 are also non-sensitizers according to the animal data, with a higher concordance of 68.9%. The results showed that the negative predictions of SkinSensPred are more likely to have the same result as that by animal testing. Since pesticides could induce skin sensitization *via* different mechanisms, the lack of pesticide data in the training set of SkinSensPred could lead to the underprediction of sensitizers of pesticides. In contrast, non-sensitizers may share the same structural features among pesticides and other chemicals.

Based on the preliminary concordance analysis, the identification of the consensus chemical space was focused on the predicted human non-sensitizers. Nine fingerprints, along with a decision tree-based method, were applied to identify the rules for excluding chemicals with prediction that conflicted with animal data. Due to the uncertainty of SkinSensPred prediction scores close to the decision threshold (0.5), two cut-off values of 0.5 and 0.4 were applied to filter out chemicals with a predicted

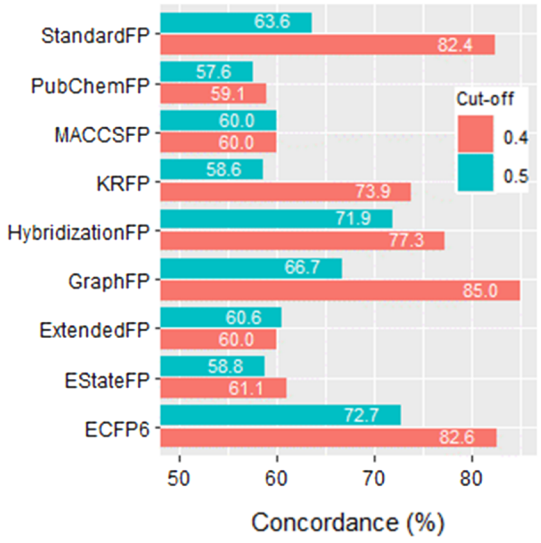


Fig. 1. The concordance between SkinSensPred prediction and animal data. The concordance of nine chemical fingerprints and two cut-off values were compared. GraphFP and the cut-off value of 0.4 had the highest concordance.

score greater than the values that are considered more likely to be sensitizers. The filtered datasets for the cut-off values of 0.5 and 0.4 consist of 70 and 38 chemicals, respectively. Each dataset was randomly divided into a training dataset and a test dataset. The numbers of chemicals in the training and test datasets are 48 and 22 for the cut-off value of 0.5, respectively. As for the cut-off value of 0.4, there are 29 and 9 chemicals in the training and test datasets, respectively.

To ensure the robustness of the identified consensus chemical space, a leave-one-out cross-validation (LOOCV) was applied to evaluate the rules defining the consensus chemical space. Each pesticide in the training dataset was utilized as a validation chemical to assess the performance of the rules derived from the remaining chemicals. In this study, the concordance of non-sensitizer predictions was utilized for the selection of cut-off values, since only negative predictions were considered by applying the cut-off values. Figure 1 shows the concordance based on the two cut-off values using LOOCV in the training dataset. Except for ECFP6 and HybridizationFP, the utilization of the other seven chemical fingerprints showed no improvement in concordance as compared to the original predictions without considering the consensus chemical space based on the cut-off value of 0.5. For the cut-off value of 0.4, StandardFP, KRFP, HybridizationFP, GraphFP, and ECFP6 showed improvement over the original predictions. The highest concordance (85.0%) was obtained by using GraphFP and a cut-off value of 0.4. A consensus chemical space was then determined by applying the decision tree-based method with GraphFP and a cut-off value of 0.4 to the whole training dataset. Four rules were generated for identifying chemicals outside of the consensus chemical space. Twenty of the 29 chemicals in the training dataset were found to be within the consensus chemical space with a coverage of 69.0%. The inferred decision tree and rules are shown in Fig. S1.

ExtendedFP, GraphFP, and HybridizationFP are specialized versions of StandardFP. The performance difference obtained from this study may imply that the rings and atomic properties covered by ExtendedFP and hybridization states considered by HybridizationFP are less relevant for defining the consensus chemical space, while GraphFP, which considers only atom connectivity without bond order information, is more useful for identifying chemicals with concordant predicted human responses and animal testing results.

3. *Independent test of the consensus chemical space*

The abovementioned test datasets were applied to assess the performance of the determined consensus chemical space. As shown in Fig. 2, an excellent 100% concordance was achieved by using the consensus chemical space based on GraphFP and a cut-off value of 0.4. The four rules of the consensus chemical space filtered out five chemicals from the corresponding test dataset, resulting in coverage of 44.4% (4/9). For comparison, the original prediction showed only 64% (16/25) concordance in the corresponding test dataset. ECFP6 and a cut-off value of 0.5 achieved an 80.0% (12/15) concordance with 68.2% (15/22) coverage in the corresponding test dataset.

To further validate this method for determining the consensus chemical space, 100 runs of random splitting of the dataset for the cut-off value of 0.4 were conducted. For each run, the corresponding training and test datasets were applied to derive rules for identifying chemicals outside of the consensus chemical space and independently testing the rules, respectively. Results are shown in Fig. 3. The median and mean concordance values are very promising, with values of 100% and 93.8%, respectively. Altogether, the proposed strategy for deriving rules is

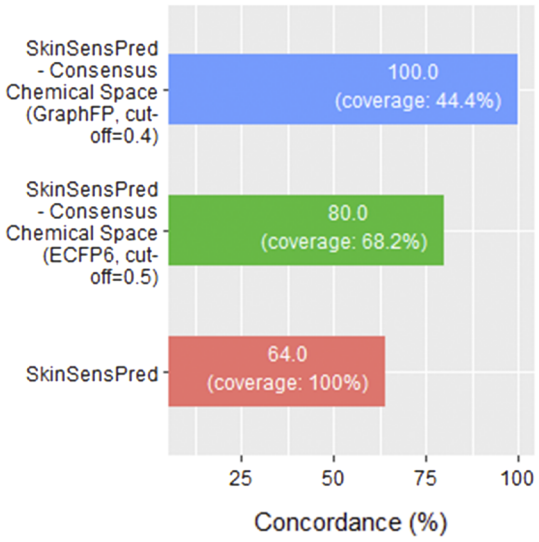


Fig. 2. Concordance comparison of the predictions in two consensus chemical spaces and original predictions. The cut-off values were applied to filter out positive or near-decision boundary predictions, resulting in a lower coverage of chemicals in the consensus chemical spaces.

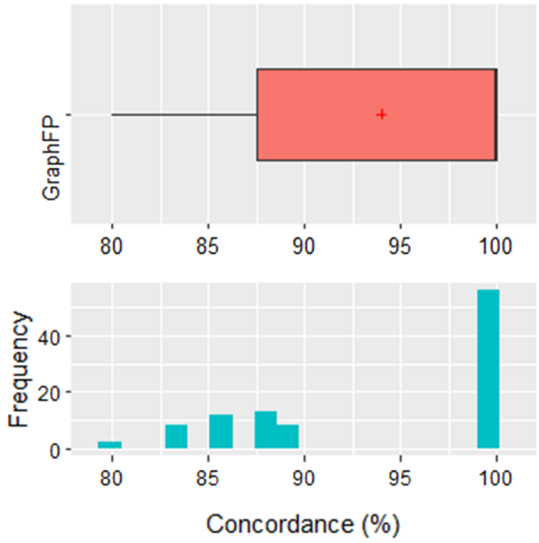


Fig. 3. Concordance based on 100 randomly divided test datasets. The box plot shows the minimum, lower quartile, median, upper quartile, and maximum of concordance from 100 experimental runs. The median, upper quartile, and maximum are all 100%. The red cross represents the mean of concordance.

effective in identifying the consensus chemical space.

4. Consensus chemical space based on the whole dataset

For machine learning algorithms, a bigger dataset often leads to more robust performance. Therefore, the whole dataset—consisting of 38 pesticides for the cut-off value of 0.4—was utilized to derive the final consensus chemical space. Five rules were obtained for filtering out chemicals outside of the consensus chemical space. The rules are shown in Table S2. The coverage of the consensus chemical space is 73.7% (28/38). As compared to the original dataset consisting of 143 pesticides, the proposed method is expected to reduce animal testing efforts by 19.6% (28/143). This method, with a low false negative rate, could be deployed as a first-tier tool to identify non-sensitizers without the need for animal testing. On the other hand, the remaining pesticides may require testing based on alternative testing methods.

Conclusion

AOP-based prediction models provide promising performance for predicting human skin sensitizers. However, their applicability to pesticides is largely unknown due to scarce pesticide data for model training. This study first evaluated the AOP-based SkinSensPred method for predicting human skin sensitization for 143 pesticides. Only a limited number of the pesticides are associated with human data. Among them, three pesticides within the original AD of SkinSensPred were all correctly predicted by SkinSensPred. While excellent predictions were obtained, the AD of SkinSensPred was designed to maximize the reliability of human response prediction based on its training dataset with little information of pesticides, which, therefore, led to limited coverage. As for the other pesticides with only animal testing data, it is difficult to assess the reliability of predictions for human responses made by SkinSensPred. Therefore, this study proposed identifying the consensus chemical space with concordant results between predicted human responses and animal testing data.

As pesticides may have some skin sensitization properties different from other types of chemicals, the prediction of skin sensitizers based on current models with little information of pesticides could be unreliable. However, non-sensitizers could still share similar structure features. Our evaluation of the concordance between predicted human responses and animal testing data concluded a low concordance (32.7%) for predicted human skin sensitizers and a high concordance (68.9%) for predicted human skin non-sensitizers. The results showed that the AOP-based model could still catch the patterns of non-sensitizers in both species.

Since species differences could still affect concordance, this study utilized a decision tree-based method along with nine fingerprints to identify the consensus chemical space for predicted human skin non-sensitizers. The method extracted exclusion rules for identifying chemicals outside of the consensus chemical space. GraphFP with a stringent cut-off value of 0.4 was found to be most useful for deriving exclusion rules based on the training dataset, with a high concordance of 85.0% (LOOCV). A 100% concordance on the test dataset showed the usefulness of the proposed method. A final set of five rules was derived from the whole dataset to give a more comprehensive description of the consensus chemical space that could be useful for further reducing animal testing.

While the proposed method effectively identifies non-sensitizers without the need for animal testing, the proposed method is based on GraphFP, whose encoding is a hashed bit vector. Due to potential bit collision issues, the identified rules are not interpretable. To make the results useful for the scientific community, the code implementing the exclusion of chemicals outside of the consensus chemical space has been made publicly available. Altogether, this study demonstrated a useful method for reconfiguring the AOP-based SkinSensPred for pesticides that is expected to reduce animal testing by 19.6%. The proposed method potentially could be applied to reconfigure other computational methods for pesticides.

To date, non-animal testing methods for skin sensitization have not gained routine regulatory use for the registration of pesticide products in most countries.[4,20] Animal testing methods such as LLNA, GPMT, and the Buehler test remain the standard tests required for active ingredients or final products for an informed hazard assessment of pesticides. Alternative methods may be considered in a case-by-case manner or when the data meet some acceptable regulatory criteria.[4] Furthermore, none of the alternative assays is recommended as a stand-alone replacement for animal tests. As a recently published approach for defining skin sensitization,[21] the developed method could be further integrated with other alternative methods to inform regulatory use.

Availability

The code for applying the exclusion rules for identifying pesticides outside of the consensus chemical space is available at https://github.com/sswang33/SkinSensPred-pesticides.

Acknowledgements

This work was supported by the National Science and Technology Council of Taiwan (MOST-107-2221-E-400-004-MY3, MOST-110-2221-E-400-004-MY3, MOST-110-2313-B-002-051-) and Taiwan Agricultural Chemicals and Toxic Substances Research Institute (109AS-24.1.2-PI-P3, 110AS-16.1.1-PI-P2, 111AS-13.1.1-PI-P2).

Conflict of interest

The authors declare that they have no competing interests.

Electronic supplementary materials

The online version of this article contains supplementary material (Supplemental Tables S1 and S2 and Supplemental Fig. S1), which is available at https://www.jstage.jst.go.jp/browse/jpestics/.

References

1) A.-T. Karlberg, M. A. Bergström, A. Börje, K. Luthman and J. L. G. Nilsson: Allergic contact dermatitis—Formation, structural requirements, and reactivity of skin sensitizers. *Chem. Res. Toxicol.* **21**, 53–69 (2008).
2) R. J. Vandebriel and H. van Loveren: Non-animal sensitization testing: State-of-the-art. *Crit. Rev. Toxicol.* **40**, 389–404 (2010).
3) A. Mehling, T. Eriksson, T. Eltze, S. Kolle, T. Ramirez, W. Teubner, B. van Ravenzwaay and R. Landsiedel: Non-animal test methods for predicting skin sensitization potentials. *Arch. Toxicol.* **86**, 1273–1295 (2012).
4) A. B. Daniel, J. Strickland, D. Allen, S. Casati, V. Zuang, J. Barroso, M. Whelan, M. J. Régimbald-Krnel, H. Kojima, A. Nishikawa, H.-K. Park, J. K. Lee, T. S. Kim, I. Delgado, L. Rios, Y. Yang, G. Wang and N. Kleinstreuer: International regulatory requirements for skin sensitization testing. *Regul. Toxicol. Pharmacol.* **95**, 52–65 (2018).
5) A. Wilm, J. Kühnl and J. Kirchmair: Computational approaches for skin sensitization prediction. *Crit. Rev. Toxicol.* **48**, 738–760 (2018).
6) G. H. Ta, C.-F. Weng and M. K. Leong: *In silico* prediction of skin sensitization: Quo vadis? *Front. Pharmacol.* **12**, 655771 (2021).
7) C.-W. Tung, C.-C. Wang and S.-S. Wang: Mechanism-informed read-across assessment of skin sensitizers based on SkinSensDB. *Regul. Toxicol. Pharmacol.* **94**, 276–282 (2018).
8) C.-C. Wang, Y.-C. Lin, S.-S. Wang, C. Shih, Y.-H. Lin and C.-W. Tung: SkinSensDB: A curated database for skin sensitization assays. *J. Cheminform.* **9**, 5 (2017).
9) C.-W. Tung, Y.-H. Lin and S.-S. Wang: Transfer learning for predicting human skin sensitizers. *Arch. Toxicol.* **93**, 931–940 (2019).
10) J. V. B. Borba, R. C. Braga, V. M. Alves, E. N. Muratov, N. Kleinstreuer, A. Tropsha and C. H. Andrade: Pred-Skin: A web portal for accurate prediction of human skin sensitizers. *Chem. Res. Toxicol.* **34**, 258–267 (2021).
11) C. Braeuning, A. Braeuning, H. Mielke, A. Holzwarth and M. Peiser: Evaluation and improvement of QSAR predictions of skin sensitization for pesticides. *SAR QSAR Environ. Res.* **29**, 823–846 (2018).
12) C.-C. Wang, P. Lin, C.-Y. Chou, S.-S. Wang and C.-W. Tung: Prediction of human fetal-maternal blood concentration ratio of chemicals. *PeerJ* **8**, e9562 (2020).
13) H.-L. Kan, C.-C. Wang, Y.-C. Lin and C.-W. Tung: Computational identification of preservatives with potential neuronal cytotoxicity. *Regul. Toxicol. Pharmacol.* **119**, 104815 (2021).
14) R. Guha: Chemical informatics functionality in R. *J. Stat. Softw.* **18**, 16 (2007).
15) E. L. Willighagen, J. W. Mayfield, J. Alvarsson, A. Berg, L. Carlsson, N. Jeliazkova, S. Kuhn, T. Pluskal, M. Rojas-Chertó, O. Spjuth, G. Torrance, C. T. Evelo, R. Guha and C. Steinbeck: The Chemistry Development Kit (CDK) v2.0: Atom typing, depiction, molecular formulas, and substructure searching. *J. Cheminform.* **9**, 33 (2017).
16) L. Breiman: "Classification and Regression Trees", Routledge, 2017.
17) T. Therneau and B. Atkinson: *Rpart: Recursive Partitioning and Regression Trees* (2019).
18) T. Hothorn and A. Zeileis: Partykit: A modular toolkit for recursive partytioning in R. *J. Mach. Learn. Res.* **16**, 3905–3909 (2015).
19) D. B. Resnik and C. Portier: Pesticide testing on human subjects: Weighing benefits and risks. *Environ. Health Perspect.* **113**, 813–817 (2005).
20) F. Pistollato, F. Madia, R. Corvi, S. Munn, E. Grignard, A. Paini, A. Worth, A. Bal-Price, P. Prieto, S. Casati, E. Berggren, S. K. Bopp, V. Zuang and E. U. Current: Regulatory requirements for the assessment of chemicals and cosmetic products: Challenges and opportunities for introducing new approach methodologies. *Arch. Toxicol.* **95**, 1867–1897 (2021).
21) OECD: "Guideline No. 497: Defined Approaches on Skin Sensitisation", Organisation for Economic Co-operation and Development, Paris, 2021.