Comparison of the predictive nature of the Genomic Allergen Rapid Detection (GARD) assay with mammalian assays in determining the skin sensitisation potential of agrochemical active ingredients.

William Masinja[1,3]*, Claire Elliott[1,2], Sandeep Modi[1], Steven J. Enoch[3], Mark T.D. Cronin[3], Elizabeth F. McInnes[1], Richard A. Currie[1]

[1]Syngenta, International Research Centre, Jealott's Hill, Bracknell, Berks, RG42 6EY, United Kingdom

[2]Penman Consulting Limited, Aspect House, Waylands Avenue, Wantage, Oxon, OX12 9FF, United Kingdom

[3]School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF, United Kingdom

*Corresponding Author: William Masinja

Email: william.masinja@syngenta.com

Tel: +44 1344 424701

Alternatives to mammalian testing are highly desirable to predict the skin sensitisation potential of agrochemical active ingredients (AI). The GARD assay, a stimulated, dendritic cell-like, cell line measuring genomic signatures, was evaluated using twelve AIs (seven sensitisers and five non-sensitisers) and the results compared with historical results from guinea pig or local lymph node assay (LLNA) studies. Initial GARD results suggested 11/12 AIs were sensitisers and six concurred with mammalian data. Conformal predictions changed one AI to a non-sensitiser. An AI identified as non-sensitising in the GARD assay was considered a potent sensitiser in the LLNA. In total 7/12 GARD results corresponded with mammalian data. AI chemistries might not be comparable to the GARD training set in terms of applicability domains. Whilst the GARD assay can replace mammalian tests for skin sensitisation evaluation for compounds including cosmetic ingredients, further work in agrochemical chemistries is needed for this assay to be a viable replacement to animal testing. The work conducted here is, however, considered exploratory research and the methodology needs further development to be validated for agrochemicals. Mammalian and other alternative assays for regulatory safety assessments of AIs must provide confidence to assign the appropriate classification for human health protection.

Key words: GARD assay, skin sensitisation prediction, agrochemicals, chemical domain, mammalian testing

Introduction

Allergic contact dermatitis is caused by an adverse immune response to chemical haptens (Rustemeyer et al., 2012, Kaplan et al., 2012). For compounds such as agrochemicals, the identification of skin sensitising properties is an important part of regulatory hazard assessment to ensure safety during manufacture and use. Currently there is no globally harmonised position on the use of *in vitro* alternatives for regulatory purposes. Consequently, agrochemicals are routinely tested for skin sensitisation using *in vivo* tests such as the guinea pig tests and the local lymph node assay (LLNA) (Basketter et al., 2012, Gwaltney-Brant, 2014). Recently attempts have been made to identify non animal-based methods with good predictive power for chemical hazard identification in a bid to reduce laboratory animal use (Alloul-Ramdhani et al., 2014, Doe and Botham, 2019, Reisinger et al., 2015, Ivan de Ávila et al., 2019). In accordance with Article 62 of the European Regulation (EC) No. 1107/2009, concerning the placing of plant protection products on the market; the use of *in vivo* mammalian test methods should only be used as a last resort. Where available non-animal test methods should be used and promoted ((EC), 2009) and several such *in vitro* assays have been developed for skin sensitisation. The Genomic Allergen Rapid Detection (GARD) assay is one of the more recent assays with as yet unknown potential for agrochemicals and therefore it was selected for evaluation in this investigation.

The GARD assay is a cell-based, *in vitro* alternative to animal testing which assesses skin sensitisation by measuring the biomarker signature in chemical-stimulated, human MUTZ-3 cells (Johansson et al., 2011). The MUTZ-3 cell line serves as a surrogate for dendritic cells (DC) and changes in transcription in the genes can be linked to processes involved in skin sensitisation (Rovida et al., 2013, Masterson et al., 2002). The GARD assay measures transcriptional changes in 200 genes associated with sensitisation (Johansson et al., 2011). The 200 gene biomarker signature includes transcripts involved in oxidative stress, dendritic

cell maturation and cytokine responses (Johansson et al., 2011). In particular, genes in pathways involved in dendritic cell maturation and activation, associated with key event three of the skin sensitisation adverse outcome pathway (AOP), which is also measured by the h-CLAT assay (OECD 2014, OECD 2018a), are included in the GARD assay. The Nrf-2 mediated oxidative response (Uruno and Motohashi, 2011), which is also the pathway measured in the KeratinoSens and LuSens assays (OECD 2018b, DB-ALM (INVITTOX), 2013) is included in the GARD assay. The results are then classified by a support vector machine (SVM) model trained on a set of reference chemicals (Forreryd et al., 2016).

During the validation process of alternative methods for skin sensitisation, a wide array of test materials from different industrial sectors have been tested using the GARD (Johansson et al., 2019), and other, assays (OECD, 2018a). This has aided in ascertaining limitations and, more specifically, chemical types that do not fall within the applicability domain of each method. The GARD assay consistently reports accuracies of close to 90 to 95% compared to *in vivo* data (Johansson et al., 2017, Johansson et al., 2014, Johansson et al., 2013; Zeller et al., 2017). The evaluation of the GARD assay in a blind study using cosmetics ingredients (from Cosmetics Europe) demonstrated a predictive performance of 83% (Johansson et al., 2017). Whilst the GARD assay has shown good performance in evaluation studies, it is worth noting that *in vitro* assays for skin sensitisation are not intrinsically standalone assays and none of them are perfectly predictive. However, they can be used a part of a weight of evidence approach, and as such, knowledge about the chemical and property domain in which an assay works is crucial.

The aim of this study was to assess the *in vitro* GARD assay's skin sensitisation predictivity in comparison with mammalian skin sensitisation tests on agrochemical active ingredients. To achieve this, agrochemical compounds for which sensitising potential had been previously established through GLP *in vivo* studies (OECD 429 murine local lymph node assays, OECD 406 guinea pig maximisation test and Buehler assays) were tested in the GARD assay. As a weight of evidence approach is advocated by the European Chemicals Agency (ECHA) when using *in vitro* data for the purpose of classification (ECHA 2017), a quantitative structure-activity relationship (QSAR) analysis of each of the test materials was also performed. Human data are available elsewhere for some of the active ingredients, however for the purposes of this evaluation these were not included as the comparison was with the available animal data. The mammalian studies are considered to be an appropriate standardised data set for comparison purposes.

Materials and Methods

GARD assay cell line

The GARD assays were conducted by Senzagen (Lund, Sweden) on behalf of Syngenta according to the protocol as described in Forreryd et al., 2016 and Johansson et al., 2013. The human myeloid leukemia-derived cell line SenzaCell (available through American Type Culture Collection (ATCC)) was used. This was maintained in α-minimum essential medium (Thermo Scientific Hyclone, USA) supplemented with 20% (volume/volume) foetal calf serum (Life Technologies, US) and 40 ng/ml recombinant human Granulocyte Macrophage Colony Stimulating Factor (rhGM-CSF) (Miltenyi Biotec, Germany). A medium change during cell expansion was performed every three to four days. Working stocks of cultures were grown for a maximum of 16 passages or two months after thawing. The chemically exposed cells were incubated for 24h at 37°C, 5% CO_2 and 95% humidity.

GARD Assay

Test substances (Table 1) were dissolved in dimethylsulfoxide (DMSO) or water, based on physico-chemical properties. The cytotoxic effects of test substances were monitored, as a concentration leading to 90% relative cell viability (Rv90) demonstrating the test substance's toxicity, was used in the assay. The assayed test substances were titrated to concentrations ranging from 1 µM to the maximum soluble concentration in cell media. For freely soluble test substances, 500 µM was set as the upper limit of the titration range. For test substances dissolved in DMSO, the in-well concentration of DMSO was 0.1%. After incubation with the test substance for 24 hours, harvested cells were stained with the viability marker Propidium Iodide (PI) (BD Bioscience, USA) and analysed by flow cytometry. For non-toxic test substances, a concentration of 500 µM was used, if possible. When test substances were poorly dissolved in cell medium or insoluble at the 500 µM concentration, the highest soluble concentration was assessed and used. The concentration to be used for any given chemical is referred to as the GARD input concentration.

Once the input concentration had been established, the cells were exposed solely to this concentration. A set of positive and negative controls were included as reference and quality controls. The test substances and controls were assayed in biological triplicates, performed at different timepoints and using different cell cultures. After incubation for 24h at 37°C, 5% CO_2 and 95% humidity, the cell cultures were lysed in TRIzol reagent (Life Technologies, Carlsbad, California, USA) and stored at -20°C until RNA had been extracted. In parallel, stimulated cells were propidium iodide (PI) stained and analysed using flow cytometry to verify the expected relative viability (Johansson et al., 2019).

RNA extraction and microarray hybridisation

RNA extraction and cDNA hybridisation were conducted on NanoString measurements as described by Johansson and co-workers (2019).

Agrochemicals

Twelve agrochemical AIs were chosen to assess the GARD assay's suitability to evaluate skin sensitisation of technical active ingredients alone. Seven of the AIs were recognised skin sensitisers based on *in vivo* data and where harmonised classifications have been assigned to them, these are presented in Table 1. The remaining five were considered to be non-sensitising substances (Table 1). For the purpose of this evaluation, no new mammalian tests were conducted and it was recognised that results from both guinea pig (Buehler or Maximisation Tests (GPMT)) and the (reduced) local lymph node assay (r)LLNA would be used to assess the skin sensitisation potential of the 12 agrochemical AIs under evaluation.

Conformal predictions

Whilst not a standard part of the GARD assay, as an additional analysis, conformal predictions were generated to demonstrate the similarity of the individual agrochemical AI test substances to the GARD training set of confirmed skin sensitisers and non-sensitisers. These were performed after the initial GARD skin predictions were generated for the test set and used to overrule or confirm the GARD prediction. Conformal prediction for the test substances examined were obtained by comparing the mean Decision Values to mean Decision Values of all reference items in the calibration dataset. Conformal predictions were used to demonstrate the similarity of the individual agrochemical AI test substances to the training set of confirmed sensitisers and non-sensitisers (Forreryd et al., 2018). The *p*-values generated were used

to score the difference (strangeness) between the result for any of the agrochemical AIs in the test set (as shown in Table 1) against the results of the sensitisers or non-sensitisers in the GARD training set chemicals. So, for example, where the *p*-value for chlorothalonil against the training set positive sensitiser results (P_{sens}) is 0.76, we can say its non-conformity to that sensitising group was only 24%. Whereas its *p*-value associated with the non-sensitising test set chemical results ($P_{non\text{-}sens}$) of 0.05 indicated a non-conformity to that group of 95%. This demonstrated a higher confidence that the GARD result obtained for chlorothalonil should indeed sit within the cluster of results for sensitising training set chemicals (Vovk, 2005).

Performance criteria

In order to understand the GARD assay's predictive power and the accuracy of its performance in comparison with the *in vivo* laboratory animal test data, statistical parameters were calculated between *in vivo* experimental and GARD *in vitro* assay result data. Sensitivity, specificity, total success/accuracy, positive and negative predictivity as well as the Cohen's kappa coefficient were calculated to evaluate the performance of the GARD assay. These parameters were all calculated using the method described by Modi et al (2012).

The differences in chemical characteristics between the GARD training set and agrochemical AI test set were also examined. This was performed by assessing the molecular weight, logarithm of the octanol-water partition coefficient (log P) and numbers of hydrogen bond donors (HBD), hydrogen bond acceptors (HBA) and rotatable bonds (RB) present in each compound in the two chemical sets. AlogP was used to calculate log P in accordance with the previous work by Guziałowska-Tic (Guziałowska-Tic, 2017) who demonstrated that AlogP provided the optimum conformity for this chemical property. It should be noted

that Kathon CG/ICP is present in the GARD training set. In order to best capture the physicochemical properties of this preservative mixture, its two active components (Methylisothiazolinone and Methylchloroisothiazolinone) were entered individually into the data set for this evaluation.

Evaluation of structural alerts for protein binding and skin sensitisation

Structural alerts for protein binding and skin sensitisation (Aptula and Roberts, 2006, Enoch et al., 2011) were identified from the OECD QSAR Toolbox version 4.3 (https://www.oecd.org/chemicalsafety/oecd-qsar-toolbox.htm) for the chemicals in both the training and test sets. The following profilers were applied:

- Protein binding alerts for skin sensitisation by OASIS
- Protein binding alerts for skin sensitisation by OASIS with skin metabolism
- Protein binding alerts for skin sensitisation according to GHS
- Protein binding alerts for skin sensitisation according to GHS with skin metabolism

The alerts were assessed for their association with *in vivo* skin sensitisation. A compound was considered to be identified as a skin sensitiser if the OASIS/GHS profiler gave an outcome of 1A or 1B, or if the OASIS with skin metabolism profiler gave a 1A result. A non-sensitiser was concluded if the OASIS/GHS profiler identified no alert or if the OASIS/GHS with skin metabolism profiler gave a 1B result. The GARD assay does not encompass the metabolic system, consequently this was the rationale for rating OASIS with metabolism 1B as a non-sensitiser. This evaluation scheme is shown in Table 2.

Results

GARD Assay

The aim of this research was to compare the results of the GARD assay to the available *in vivo* skin sensitisation study outcomes for twelve agrochemical AIs. The results from the GARD assay are summarised in Table 3.

Analysis of Applicability Domain of the GARD Assay and Agrochemical AIs Tested

Table 4 details the *in vivo* assay predictions compared to those of the GARD assay for this study's test set. The GARD assay correctly predicted the six sensitisers, however, the negative predictivity of the GARD assay for the test set was not concordant with that of the *in vivo* results. Performance analysis of these data using the statistical parameters was conducted as shown in Table 5 and illustrated in Figure 1. When compared to the *in vivo* results, the negative predictivity of the GARD assay was mainly nonconcordant for this test set, with a positive predictivity of 55%, sensitivity of 86% and a total accuracy of 50%. Cohen's Kappa coefficient provided a statistical measure of inter-rater agreement for categorical items (sensitiser/non-sensitiser) and the value for Cohen's Kappa value was -0.16 indicating poor agreement between sensitisers and non-sensitisers.

In order to determine possible reasons for the nonconcordant results between the GARD assay and *in vivo* test results for the agrochemical AI test set, the ranges of physicochemical properties of the GARD training set and the AIs tested were compared. A broad overview of the range relative physicochemical properties which may affect solubility and uptake is provided as a plot in Figure 2. A range of physicochemical properties (i.e. calculated log P (AlogP) against molecular weight) associated with the AIs were plotted

against the published training set of the GARD assay. Figure 3 shows the comparison of the distribution of molecular weight for the GARD training set and the agrochemical AIs tested. The chemicals in the GARD training set had molecular weights of approximately 150Da and only two had molecular weights above 300Da. The molecular weights of the agrochemical AIs tested were higher, with many approximately 400Da and only two agrochemicals (dicamba and chlorothalonil) with molecular weights below 300Da. A comparison of the distribution of AlogP for the GARD training set and the agrochemical AIs tested is given in Figure 4. The majority of chemicals in the GARD training set have AlogP values in the range of 1, and only two were above 4. The majority of agrochemicals tested had an AlogP of approximately 4 and none had an AlogP value below 2. In terms of the ranges of the two physicochemical properties considered, there is a difference between those of the GARD training set and the AIs tested.

Following the conformal prediction analysis, the GARD assay result for dicamba was changed from a being a skin sensitiser to a non-sensitiser as shown in Table 6. This was due to the derived $_{Pnon\text{-}sens}$ value of 0.16, indicating that dicamba had non-conformity to the non-sensitising group of 84% as opposed to the 86% it demonstrated for the sensitising group.

An evaluation of the physico-chemical similarities between the GARD training set (Forreryd et al., 2018) and the test set of 12 agrochemical AIs was performed and the results shown in Table 7. The evaluation focused upon the test materials' log P, MW, HBA, HBD and RB, which are molecular descriptors often associated with membrane permeability and included in the defined rules for pesticide likeness (Avram et al., 2014). The HBA and RB also demonstrated a noticeable difference in recorded median values with

little difference between the two sets seen in HBD. Table 7 shows the differences between physicochemical values of the training and test set.

An assessment of the chemical domains covered by both sets of chemicals was performed. The Venn diagram shown in Figure 5 indicates that the training set covers all the chemical domains identified in the agrochemical AI test set and also covers bimolecular nucleophilic substitution (S_N2) which was not present in the test set.

In Silico Evaluation

Further evaluation using the OECD QSAR Toolbox comparing the *in vivo* study experimental results against the *in silico* profiling of both the agrochemical AI test set and GARD assay training set was performed. The structural alerts in the profilers were predictive of the skin sensitisation *in vivo* experimental outcome for the training set (Table 8).

Discussion

This study compared the predictions of the GARD assay to the results of previously conducted *in vivo* animal assays testing the skin sensitisation potential of 12 agrochemical AIs. The GARD assay identified ten of the test materials as skin sensitisers and two as non-sensitisers. The results from the GARD assay were not in agreement with the *in vivo* data for five of the 12 agrochemical AI materials tested. In order to ensure the veracity of the outcome of the GARD assay, conformal prediction analysis was performed, and this changed the outcome of the GARD assay for dicamba from being a sensitiser to a non-sensitiser.

In Silico Evaluation – Mechanistic Chemistry and Physicochemical Property Domains

In order to understand the performance of the GARD Assay compared to *in vivo* results for the AIs, their coverage in terms of mechanistic and chemical applicability domains was examined. First, an assessment was undertaken to comprehend the change in sensitisation outcome for dicamba using conformal prediction analysis (as shown in Table 6) and whether this could give an insight into domains. In this instance, it appears that this conformity exercise does not necessarily indicate if a chemical was within the appropriate applicability domain of the GARD assay, but rather, it indicates using the model's own training set, within which of the two groups of potential outcomes, the test compound is most likely to fall. Thus, the conformal method would not necessarily be able to indicate how appropriate the GARD assay is for a chemical that falls outside of the chemical space of the training set used. As such it can be determined that the use of conformal predictions is not an appropriate method to ascertain whether the agrochemical AI test materials in this study fall within the current applicability domain of the GARD assay.

To determine the possible role of mechanistic chemistry with regard to domain alerts flagged by the OECD QSAR Toolbox, each set of chemicals were investigated (note structural alerts are discussed in more detail below). No significant differences between the GARD assay training set and the agrochemical AI test set were observed during our evaluation. All of the chemical mechanisms of action important for skin sensitisation *(*Aptula and Roberts, 2006, Enoch et al., 2011) have been identified in the GARD assay training set. Thus, differences in the responses from the GARD assay and *in vivo* rodent skin sensitisation test results for the AIs tested, are not as a result of any specific chemical mechanism of action for skin sensitisation being absent in the GARD training set, as it encompassed all those identified in the agrochemical AI test set. Therefore, in order to further understand why the difference in results between

the GARD assay and *in vivo* experimental tests was observed, the physico-chemical properties of the chemicals in the GARD training set and the agrochemical test set were examined.

The physicochemical property domains of the GARD training set and the 12 AIs tested were compared to provide further understanding of the differences observed between the predicted GARD results and the *in vivo* experimental results. The purpose of this analysis was to determine if the sets represented different areas of chemical space, as defined by the physicochemical properties considered. Such properties are a key component of the "applicability domain" of a test assay or QSAR and other components of the applicability domain (where relevant) include structural similarity, mechanism of action, metabolism, reactivity and toxicokinetics (Dimitrov et al., 2005, Netzeva et al., 2005, van der Laan et al., 2012). This analysis was not intended to be a full determination of the applicability domains of the AI test set and GARD assay training set. For skin sensitisation, a full analysis of applicability domain would include an analysis of the mechanistic reactivity domain associated with each chemical (Aptula et al., 2005, Aptula and Roberts, 2006, Roberts et al., 2007). However, definition and consideration of the physico-chemical property ranges, such as compound solubility, is a key step in the assessment of technical limitations to assist in the evaluation and ultimate validation of an *in vitro* assay (Bruner et al., 1996, Worth and Balls, 2004) and assists in its correct usage. Following the evaluation of molecular descriptors of the test and training set chemicals, an apparent difference in molecular weights was observed between the test and training set, indicating that a higher molecular weight range is present in the agrochemical AI test set compared to that seen in the training set. Whilst the training set contained molecules with a molecular weight of predominantly 50-200Da, one further compound, Tween 80 with a molecular weight 833Da, was included in the training set. This compound is, however, benign, a non-sensitiser and is used regularly as a vehicle in toxicity studies. Thus, the inclusion of Tween 80 in the training set has expanded the

molecular weight range of this set and this range may not be representative of all the compounds contained within it. This is demonstrated by the median of the training set.

There is a substantial difference between the molecular weight of Tween 80 and the nearest training set neighbour (penicillin). This indicated that the molecular weight of the test set of agrochemicals was not adequately represented within the GARD training set, however, it is acknowledged that these were well within the limits of absorption and skin penetration (Lipsinki et al, 2001). This means that there is a domain of chemical reactivity unaccounted for concerning the molecular weight of the penicillin compound in the training set. As many of the agrochemicals fall within this domain, confidence in the accuracy with which the GARD assay will be able to give the correct prediction may not be strong.

In addition to the difference in molecular weight there is also a separation between the training set log P (with a range of -4.77 to 5.74 and median of 1.12) and the test set log P values (2.78 to 5.02 with a median of 3.94). In this context the initial GARD predictions cannot be considered robust based on the current test data used in the assay. It is well reported that molecular weight and log P have an influence on the rate of dermal absorption of chemicals (Potts and Guy, 1992). These chemical parameters have not been used in this study to aid in the evaluation of skin sensitisation potential, instead they have been used here to identify potential differences in the chemical space between the two sets of chemicals. It has been previously reported that the most marked difference in physico-chemical properties between pharmaceuticals and agrochemicals is the lower number of hydrogen bond donors (Clarke and Delaney, 2003, Tice, 2001). Consequently the hydrogen bond donors (HBD), hydrogen bond acceptors (HBA) and rotating bonds (RB) in the training and test set groups have been compared (Clarke and Delaney, 2003). The addition of these three physico-chemical properties to this study's evaluation enabled the complete

comparison of the chemical sets in accordance with Lipinski's "rule of five" and Hao and coworkers's rules for pesticide likeness (Avram et al., 2014, Clarke and Delaney, 2003, Barret, 2018, Lipinski et al., 2001, Hao et al., 2011). A clear difference in distribution can be observed in four of the five physico-chemical properties of the chemical sets that have been reviewed here. Thus, at this time there is insufficient evidence to suggest the GARD training set offers the width in range necessary to capture the agrochemical AI test set properties.

Review of Structural Alerts

The assessment of the presence of structural alerts for skin sensitisation, as identified from the OECD QSAR Toolbox, in the chemical structures of the GARD assay training set and agrochemical AI test set, also provided predictions that, in comparison to the *in vivo* experimental data, overestimated the skin sensitisation potential of the test set. For the test set, there was 71% agreement between experimental sensitisation and predicted sensitisation. However, only 60% of the test set agrochemical AIs with *in vivo* non-sensitising results, were associated with structural alerts for protein binding (and hence skin sensitisation) by the OECD QSAR Toolbox. This overestimation of the sensitisation potential of the agrochemical AI test set is largely in keeping with the trend observed with the GARD assay results. It should be remembered that structural alerts for protein binding (related to skin sensitisation) in the OECD QSAR Toolbox have been developed from many sources including historical skin sensitisation data. For instance, Enoch et al (2008) developed a set of structural alerts for skin sensitisation based on historical LLNA data compiled by Gerberick et al (2005). These data, and a subsequent expanded LLNA data set (Kern et al 2010), are predominantly for small, low molecular weight compounds, the majority of which are relevant as cosmetics ingredients or represent the chemical of cosmetic ingredient space with few, or no compounds representative of agrochemicals.

The results seen in the OECD QSAR Toolbox profiling suggest that differences in chemical space can also influence skin sensitisation outcome. It may be hypothesised that the structural alerts are more informative of the skin sensitisation potential of low molecular weight, cosmetic-like compounds than the potential for adverse outcomes in agrochemicals and specifically for our test set. In addition, the shift towards increased hydrophobicity and molecular weight in the agrochemical AIs compared to the training set values, indicates a potential for lower skin penetration which is not accounted for. This is in line with a previous publication by Basketter et al (1992) suggesting that an important factor governing the skin sensitisation potential of halogenated chemicals, such as bromoalkanes, is their skin penetration rate (Basketter et al., 1992). To attain a more predictive set of structural alerts for agrochemicals these additional physicochemical factors and skin penetration need to be accounted for, or a factor may need to be applied to account for the dermal absorption differences. This is also an important consideration for all *in vitro* assays for skin sensitisation and is often accommodated within the weight of evidence or as part of the risk assessment.

To illustrate the issue of the assessment of halogenated compounds, dicamba is a chlorinated benzoic acid that has been used widely on a variety of crops as an effective herbicide for more than 50 years (Wang et al., 2016, Yao et al., 2015). Whilst some acids are included in the training set e.g. salicylic acid, lactic acid, benzoic acid, the GARD assay was unable to make an accurate prediction for dicamba. The GARD assay predicted dicamba to be a skin sensitiser, whilst the *in vivo* study and ECHA harmonised classification have not classified it as such. The acids present in the training set were not halogenated and the only compound present in the GARD training set that was halogenated was methylchloroisothiazolinone. As expected with agrochemicals (Jeschke, 2010), nine of the 12 compounds in the agrochemical AI test set

were halogenated. This further indicates the difference between the chemistry of the two chemical sets evaluated in this study. In particular, there was discordance between the GARD and *in vivo* results for pinoxaden, which had an identified EC3 value from a previously conducted LLNA corresponding to a harmonised classification Skin sens. Category 1A, H317 (ECHA, 2015, EFSA, 2013). This compound is outside of the applicability domain, however this does not fully explain why this assay was unsuccessful at predicting a potent sensitiser. A potential limitation of the *in vivo* methods may also have been a factor in the differences in results seen between the GARD assay and *in vivo* methods results. The highest test material dose that could be selected for the guinea pig or LLNA skin sensitisation tests is the maximum soluble concentration that does not induce systemic toxicity and/or excessive local irritation (OECD, 1992, OECD, 2010). Where observed toxicity of a given test material may have limited the highest concentration that could be tested in the *in vivo* experiments, the GARD assay was still able to use the high concentrations and investigate skin sensitisation potential at these levels. For these test substances, solubility and cytotoxicity were not limiting factors in the methodology as the maximum exposure concentration was used for negative outcomes.

The chemical space disparity that has been identified between the GARD training set and the agrochemical AI test set may have occurred because the predictive model is a machine learning classifier (a support vector machine model) that has been trained on gene signals mainly for compounds used as, or similar to, cosmetics ingredients. The gene signal in relation to cosmetics ingredients has been learned by the model and chemicals of all domains are classified in this way. This gives each chemical a biological fingerprint relevant to cosmetics but not to agrochemicals.

<u>Future work and opportunities for further improvement</u>

The results from the GARD assay indicate that the biological fingerprint (i.e. the changes in transcription in the genes in the Mutz-3 cells (surrogates for dendritic cells)) for skin sensitisation is not consistent across all chemicals. The GARD assay performed in the manner expected of it, in that it provided predictions of skin sensitisation potential for the agrochemical AI test set using the machine learnt, biological fingerprint provided by its training set. In an attempt to improve skin sensitisation predictivity for agrochemical compounds in the GARD assay in the future, additional compounds should be added to the GARD training set with molecular weights between 300- 800Da and ALogP values of 3-5. Halogenation has not been identified as a cause of sensitisation, however unlike cosmetics, agrochemicals are intentionally biologically active and frequently halogenated. This may skew or change the biological fingerprint of this chemical set in a manner that affects the prediction produced by the GARD assay in comparison to the *in vivo* experimental results. As noted above, halogenation of compounds is important when considering skin sensitisation (Basketter et al., 1992), and thus addition of halogenated compounds to the GARD assay training set may improve its capability to predict skin sensitisation of agrochemicals. The compounds added to the training set should include an increased number of different chemistries i.e. biocides and agrochemicals and this may aid to further investigate this hypothesis.

The design of the training set for machine learning in the GARD assay is a key component to adequately establish an understanding of biological outputs and how they apply to the individual domains of the AIs tested. The application of structural alerts delivers a clear understanding of the applicability domain and is required to be able to identify limitations to mechanistic chemistry in the *in vitro* assay being evaluated. However, we have observed in this study that all reactivity domains present in the test set are covered by the training set, and yet a nonconcordant result is observed between the *in vitro* and *in vivo* test methods. We hypothesise that the physico-chemical parameters of the test and training chemical sets examined in this study also play a role in the setting of an applicability domain. This is in line with the previously made

assumption that similar predictivity can be achieved for substances that are similar to those in the training set and that the applicability domain of a model would depend on the structural, physico-chemical and response information in the data used for training a model (Wilm et al., 2018). It is noteworthy that 12 compounds is a small test set to evaluate the GARD assay. The lack of overlap between the test and training set indicates that additional work needs to be conducted to address false positive and negative outcomes.

Whilst the GARD assay is not an approved OECD test guideline, it has the potential to replace mammalian testing in a number of different chemistries as part of a weight of evidence. However, the above work has demonstrated the need for the GARD training set to be expanded, in particular to include agrochemical compounds that occupy a different chemical space in terms of size and hydrophobicity. Additional confidence needs to be demonstrated or limitations to the assay identified, before the GARD assay can become a standalone replacement to animal testing. Validation of new alternative methods using different chemistries to ensure robustness of *in vitro* assays and scientifically reliable results across chemical domains, is crucial. This is exploratory research and the GARD assay needs further development to be validated for agrochemicals in our endeavour to confidently replace mammalian studies.

Conflicts of Interest

The authors declare no conflicts of interest

References

Alloul-Ramdhani, M., Tensen, C.P. & El Ghalbzouri, A. (2014) Chapter 3.2 - Chemical Sensitization. *In:* Kleinjans, J. (ed.) *Toxicogenomics-Based Cellular Models.* San Diego: Academic Press pp. 67-80.

Aptula, A.O. & Roberts, D.W. (2006) Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity. *Chemical Research in Toxicology,* 19 (8), pp. 1097-105. DOI: 10.1021/tx0601004

Aptula, A.O., Patlewicz, G. & Roberts, D.W. (2005) Skin sensitization: reaction mechanistic applicability domains for structure-activity relationships. *Chemical Research in Toxicology*, 18 (9), pp. 1420-6. DOI: 10.1021/tx050075m

Avram, S., Funar-Timofei, S., Borota, A., Chennamaneni, S.R., Manchala, A.K. & Muresan, S. (2014) Quantitative estimation of pesticide-likeness for agrochemical discovery. *Journal of Cheminformatics.* 6 (1), pp. 42. DOI: 10.1186/s13321-014-0042-6.

Barret R. (2018) Lipinski's Rule of Five. In: Barret, R. (ed.) *Therapeutical Chemistry*. Therapeutical Chemistry: Elsevier. pp. 97-100. [online] Available through: https://www.sciencedirect.com/science/article/pii/B9781785482885500068 [Accessed: 22nd September 2020]

Basketter D., Crozier J., Hubesch B., Manou I., Mehling A. & Scheel J. (2012) Optimised testing strategies for skin sensitization – The LLNA and beyond. *Regulatory Toxicology and Pharmacology.* 64 (1), pp. 9-16. DOI: 10.1016/j.yrtph.2012.06.003.

Basketter D., Roberts, D. W., Cronin, M. & Scholes, E. W. (1992) The value of the local lymph node assay in quantitative structure-activity investigations. *Contact Dermatitis.* 27 (3), pp. 137-42. DOI: 10.1111/j.1600-0536.1992.tb05241.x.

Bruner, L. H., Carr, G. J., Chamberlain, M. & Curren, R. D. (1996) Validation of alternative methods for toxicity testing. *Toxicology in Vitro*. 10 (4), pp. 479-501. DOI: 10.1016/0887-2333(96)00028-8

Clarke, E.D. & Delaney, J.S. (2003) Physical and Molecular Properties of Agrochemicals: An Analysis of Screen Inputs, Hits, Leads, and Products. *CHIMIA International Journal for Chemistry*. 57 (11), pp. 731-734. DOI: 10.2533/000942903777678641

Dimitrov, S., Dimitrova, G., Pavlov, T., Dimitrova, N., Patlewicz, G., Niemela, J. & Mekenyan, O. A. (2005) A stepwise approach for defining the applicability domain of SAR and QSAR models. *J Chem Inf Model*. 45 (4), pp. 839-849. DOI: 10.1021/ci0500381

Doe, J. & Botham, P. (2019) Chapter 3.3 - Chemicals and Pesticides: A Long Way to Go. In*:* Balls, M., Combes, R. & Worth, A. (eds.). *The History of Alternative Test Methods in Toxicology*. Academic Press. pp. 177-184. [online] Available through: https://www.sciencedirect.com/science/article/pii/B9780128136973000214 [Accessed: 22nd September 2020]

DB-ALM (INVITTOX). (2013) Protocol 155: KeratinoSensTM [online]. Available at: https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/EURL-ECVAM/datasets/DBALM/LATEST/online/DBALM_docs/155_P_%20KeratinoSens.pdf [Accessed: 22nd September 2020]

European Commission. (2009) Regulation (EC) No 1107/2009 of the European Parliament and of the council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC. [online] *Official Journal of the European Union.* Available at: https://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:309:0001:0050:EN:PDF [Accessed: 22nd August 2019]

European Chemical Agency. (2017) Guidance on the Application of the CLP Criteria. [online] Available at: https://echa.europa.eu/documents/10162/23036412/clp_en.pdf/58b5dc6d-ac2a-4910-9702-e9e1f5051cc5 [Accessed: 22nd September 2020]

European Chemical Agency. (2015) CLH Report for Pinoxaden. [online] Available at: https://echa.europa.eu/documents/10162/13626/clh_report_pinoxaden_en.pdf [Accessed: 22nd September 2020]

Enoch S.J., Ellison C.M., Schultz T.W. & Cronin M.T. (2011) A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity. *Crit Rev Toxicol*. 41 (9), pp. 783-802. DOI: 10.3109/10408444.2011.598141

Enoch, S.J., Madden, J.C. & Cronin M.T. (2008) Identification of mechanisms of toxic action for skin sensitisation using a SMARTS pattern based approach. *SAR and QSAR in Environmental Research.* 19 (5-6), pp. 555-78. DOI: 10.1080/10629360802348985

European Chemicals Agency. (2008a) Harmonised classification - Annex VI of Regulation (EC) No 1272/2008 (CLP Regulation). Available at: https://echa.europa.eu/information-on-chemicals/registered-substances [Accessed 19 June 2020]

European Chemicals Agency. (2008b) Dicamba Harmonised classification - Annex VI of Regulation (EC) No 1272/2008 (CLP Regulation) Available at: https://echa.europa.eu/information-on-chemicals/cl-inventory-database/-/discli/details/30170 [Accessed 19 June 2020]

European Food Safety Authority. (2011a). Conclusion on the peer review of the pesticide risk assessment of the active substance dicamba. *European Food Safety Authority Journal.* 9 (1): 1965. DOI: 10.2903/j.efsa.2011.1965

European Food Safety Authority. (2011b) Conclusion on the peer review of the pesticide risk assessment of the active substance difenoconazole. *European Food Safety Authority Journal.* 9 (1): 1967. DOI: 10.2903/j.efsa.2011.1967.

European Food Safety Authority. (2013) Conclusion on the peer review of the pesticide risk assessment of the active substance pinoxaden. *European Food Safety Authority Journal.* 11 (8): 3269 DOI: 10.2903/j.efsa.2013.3269.

European Food Safety Authority. (2014) Conclusion on the peer review of the pesticide risk assessment of the active substance cyantraniliprole. *European Food Safety Authority Journal.* 12 (9): 3814. DOI: 10.2903/j.efsa.2014.3814.

European Food Safety Authority. (2015) Conclusion on the peer review of the pesticide risk assessment of the active substance benzovindiflupyr. *European Food Safety Authority Journal.* 13 (3): 4043, DOI: 10.2903/j.efsa.2015.4043

European Food Safety Authority., Arena, M., Auteri, D., Barmaz, S., Bellisai, G., Brancato, A., Brocca, D., Bura, L., Byers, H., Chiusolo, A., Court Marques, D., Crivellente, F., De Lentdecker, C., Egsmose, M., Erdos, Z., Fait, G., Ferreira, L., Goumenou, M., Greco, L., Ippolito, A., Istace, F., Jarrah, S., Kardassi, D., Leuschner, R., Lythgo, C., Magrans, J.O., Medina, P., Miron, I., Molnar, T., Nougadere, A., Padovani, L., Parra Morte, J.M., Pedersen, R., Reich, H., Sacchi, A., Santos, M., Serafimova, R., Sharp, R., Stanek, A., Streissl, F., Sturma, J., Szentes, C., Tarazona, J., Terron, A., Theobald, A., Vagenende, B., Verani, A. & Villamar-Bouza, L. (2018a.) Peer review of the pesticide risk assessment of the active substance chlorothalonil. *European Food Safety Authority Journal., 16 (1): 5126.* DOI: 10.2903/j.efsa.2018.5126.

European Food Safety Authority., Arena, M., Auteri, D., Barmaz, S., Brancato, A., Brocca, D., Bura, L., Carrasco Cabrera, L., Chaideftou, E., Chiusolo, A., Court Marques, D., Crivellente, F., De Lentdecker, C., Egsmose, M., Fait, G., Greco, L., Ippolito, A., Istace, F., Jarrah, S., Kardassi, D., Leuschner, R., Lostia, A.,

Lythgo, C., Magrans, O., Mangas, I., Miron, I., Molnar, T., Padovani, L., Parra Morte, J.M., Pedersen, R., Reich, H., Santos, M., Serafimova, R., Sharp, R., Stanek, A., Sturma, J., Szentes, C., Terron, A., Tiramani, M., Vagenende, B. & Villamar-Bouza, L. (2020b). Peer review of the pesticide risk assessment of the active substance clodinafop (variant evaluated clodinafop-propargyl). *European Food Safety Authority Journal.* 18 (7): 6151. DOI: 10.2903/j.efsa.2020.6151.

Food and Agriculture Organisation of the United Nations. (2014) Pesticide residues in food – 2013: 3: toxicological evaluations / Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and the WHO Core Assessment Group on Pesticide Residues, Geneva, Switzerland, 17–26 September 2013. World Health Organization. Available at: https://apps.who.int/iris/bitstream/handle/10665/129593/9789241665292_eng.pdf?sequence=1&isAllowed=y [Accessed 22 September 2020]

Food and Agriculture Organisation of the United Nations. (2016) Pesticide residues in food 2016. Joint FAO/WHO meeting on pesticide residues. Pinoxaden (293). Available at: http://www.fao.org/fileadmin/templates/agphome/documents/Pests_Pesticides/JMPR/Report2016/PINOXADEN.pdf [Accessed 22 September 2020]

Forreryd, A., Norinder, U., Lindberg, T. & Lindstedt M. (2018) Predicting skin sensitizers with confidence - Using conformal prediction to determine applicability domain of GARD. *Toxicology In Vitro*. 48, pp. 179-187. DOI: 10.1016/j.tiv.2018.01.021

Forreryd, A., Johansson, H., Lindberg, T., Zeller, K.S. & Lindstedt, M. (2019) Letter to the editor regarding the article “Is a combination of assays really needed for non-animal prediction of skin sensitization potential? Performance of the GARD™ (Genomic Allergen Rapid Detection) assay in comparison with OECD guideline assays alone and in combination.” *Regulatory Toxicology and Pharmacology*. 107: 104439. DOI: 10.1016/j.yrtph.2019.104439

Forreryd, A., Zeller, K.S., Lindberg, T., Johansson, H. & Lindstedt, M. (2016) From genome-wide arrays to tailor-made biomarker readout – Progress towards routine analysis of skin sensitizing chemicals with GARD. *Toxicology in Vitro*. 37, pp. 178-188, DOI: 10.1016/j.tiv.2016.09.013

Gerberick, G.F., Ryan, C.A., Kern, P.S., Schlatter, H., Dearman, R.J., Kimber, I., Patlewicz, G.Y., & Basketter, D.A. (2005) Compilation of historical local lymph node data for evaluation of skin sensitization alternative methods. *Dermatitis*. 16 (4), pp. 157-202. DOI: 10.2310/6620.2005.05040

Guziałowska-Tic, J. (2017) The use of QSAR methods for determination of n-octanol/water partition coefficient using the example of hydroxyester HE-1. *E3S Web of Conferences*. 19: 02034. DOI: 10.1051/e3sconf/20171902034

Gwaltney-Brant, S. (2014) Chapter 22 - Immunotoxicity biomarkers. In*:* Gupta, R. C. (ed.) *Biomarkers in Toxicology*. Boston: Academic Press. pp. 373 -385. [online] Available through: http://www.sciencedirect.com/science/article/pii/B9780124046306000221

[Accessed: 22nd September 2020]

Harp, P.R. (2010). Chapter 85 - Dicamba. In*:* Krieger, R. (ed.) *Hayes' Handbook of Pesticide Toxicology (Third Edition).* New York: Academic Press.

Hao, G., Dong, Q. & Yang, G., (2011) A Comparative Study on the Constitutive Properties of Marketed Pesticides. *Molecular informatics*. [First published online July 2011], DOI: 10.1002/minf.201100020

Ivan de Ávila, R., Lindstedt, M. & Valadares, M.C. (2019) The 21st Century movement within the area of skin sensitization assessment: From the animal context towards current human-relevant in vitro solutions. *Regulatory Toxicology and Pharmacology*. 30 (6-7), pp. 614-22. DOI: 10.1016/j.yrtph.2019.104445

Jeschke, P. (2010) The unique role of halogen substituents in the design of modern agrochemicals. *Pest Management Science*. 66 (1), pp. 10-27. DOI: 10.1002/ps.1829

Johansson, H., Albrekt, A.S., Borrebaeck, C.A. & Lindstedt, M. (2013) The GARD assay for assessment of chemical skin sensitizers. *Toxicology In Vitro*. 27 (3), pp. 1163-9. DOI: 10.1016/j.tiv.2012.05.019

Johansson, H., Gradin, R., Forreryd, A., Agemark, M., Zeller, K., Johansson, A., Larne, O., van Vliet, E., Borrebaeck, C. & Lindstedt, M. (2017). Evaluation of the GARD assay in a blind Cosmetics Europe study. *ALTEX*. 34 (4), pp. 515-523. DOI: 10.14573/altex.1701121

Johansson, H., Gradin, R., Johansson, A., Adriaens, E., Edwards, A., Zuckerstätter, V., Jerre, A., Burleson, F., Gehrke, H. & Roggen, E.L. (2019) Validation of the GARD™skin Assay for Assessment of Chemical Skin Sensitizers: Ring Trial Results of Predictive Performance and Reproducibility. *Toxicological Sciences*. 170 (2), pp. 374-381. DOI: 10.1093/toxsci/kfz108

Johansson, H., Rydnert, F., Kühnl, J., Schepky, A., Borrebaeck, C. & Lindstedt, M. (2014) Genomic allergen rapid detection in-house validation--a proof of concept. *Toxicol Sci*. 139 (2), pp. 362-70. DOI: 10.1093/toxsci/kfu046

Johansson, H., Lindstedt, M., Albrekt, A.S. & Borrebaeck, C.A. (2011) A genomic biomarker signature can predict skin sensitizers using a cell-based in vitro alternative to animal tests. *BMC Genomics*. 12 (1), pp. 399. DOI: 10.1186/1471-2164-12-399

Kaplan, D.H., Igyártó, B.Z. & Gaspari, A.A. (2012) Early immune events in the induction of allergic contact dermatitis. *Nature reviews. Immunology.* 12 (2), pp. 114-124. DOI: 10.1038/nri3150

Kern, P.S., Gerberick, G.F., Ryan, C.A., Kimber, I., Aptula, A. & Basketter, D.A. (2010) Local lymph node data for the evaluation of skin sensitization alternatives: a second compilation. *Dermatitis*. 21 (1), pp. 8-32. DOI: 10.2310/6620.2009.09038

Lipinski, C.A., Lombardo, F., Dominy, B.W. & Feeney, P.J (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*. 23 (1-3), pp. 3-25. DOI: 10.1016/s0169-409x(00)00129-0.

Masterson, A. J., Sombroek, C. C., De Gruijl, T. D., Graus, Y. M., Van Der Vliet, H. J., Lougheed, S. M., Van Den Eertwegh, A. J., Pinedo, H. M. & Scheper, R. J. (2002) MUTZ-3, a human cell line model for the cytokine-induced differentiation of dendritic cells from CD34+ precursors. *Blood.* 100 (2), pp. 701-703. DOI: 10.1182/blood.v100.2.701.

Modi, S., Li, J., Malcomber, S., Moore, C., Scott, A., White, A. & Carmichael, P. (2012) Integrated in silico approaches for the prediction of Ames test mutagenicity. *Journal of Computer-Aided Molecular Design*. 26 (9), pp. 1017-33. DOI: 10.1007/s10822-012-9595-5

Netzeva, T. I., Worth, A., Aldenberg, T., Benigni, R., Cronin, M. T., Gramatica, P., Jaworska, J. S., Kahn, S. Klopman. G., Marchant, C. A., Myatt, G., Nikolova-Jeliazkova, N., Patlewicz, G. Y., Perkins, R., Roberts, D. W., Schultz, T. W., Stanton, D. T., van de Sandt, J. J. M., Tong, W., Veith, G. & Yang, C. (2005) Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. The report and recommendations of ECVAM Workshop 52. *ATLA - Altern Lab Anim,* 33 (2), pp. 155-73. DOI: 10.1177/026119290503300209.

The Organisation for Economic Co-operation and Development (OECD). (2014) *The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins* [online]. Available at: https://www.oecd.org/env/the-adverse-outcome-pathway-for-skin-sensitisation-initiated-by-covalent-binding-to-proteins-9789264221444-en.htm [accessed: 22 September 2020]

The Organisation for Economic Co-operation and Development (OECD). (1992) *OECD Guideline for testing of chemicals, Test guideline number 406: Skin Sensitisation*. [online] Available at: https://www.oecd-ilibrary.org/docserver/9789264070660

en.pdf?expires=1600791260&id=id&accname=guest&checksum=F852CDE0B1AD8B9AFEB2A75F3F75D1DE [accessed: 22 September 2020]

OECD. (2010) *OECD Guideline for testing of chemicals, Test guideline number 429*: *Skin Sensitisation: Local Lymph Node Assay.* [online] Available at: https://www.oecd-ilibrary.org/docserver/9789264071100-en.pdf?expires=1600791579&id=id&accname=guest&checksum=52641FF1890D13CA60A4A09D473A631A [accessed: 22 September 2020]

OECD. (2018a) *Key Event-Based Test Guideline 442E, in vitro skin sensitisation assays addressing the key event on activation of dendritic cells on the adverse outcome pathway for skin sensitisation*.[online] Available at: https://www.oecd-ilibrary.org/docserver/9789264264359-en.pdf?expires=1600791789&id=id&accname=guest&checksum=6696CBB5584A600E833E396625AD3AF2 [accessed: 22 September 2020]

OECD. (2018b) *Key Event Based Test Guideline 442d, in vitro skin sensitisation assays addressing the AOP key event on keratinocyte activation*. [online] Available at: https://www.oecd-ilibrary.org/docserver/9789264229822-en.pdf?expires=1600792322&id=id&accname=guest&checksum=679A5E3517F761BC124DABF6455A5E79 [accessed: 22 September 2020]

OECD. (2020) *The OECD QSAR Toolbox*. [online] Available at: https://www.oecd.org/chemicalsafety/oecd-qsar-toolbox.htm. [Accessed 22 April 2019].

Potts, R. O. & Guy, R. H. (1992) Predicting Skin Permeability. *Pharmaceutical Research,* 9 (5), pp. 663-9. DOI: 10.1023/a:1015810312465

Reisinger, K., Hoffmann, S., Alépée, N., Ashikaga, T., Barroso, J., Elcombe, C., Gellatly, N., Galbiati, V., Gibbs, S., Groux, H., Hibatallah, J., Keller, D., Kern, P., Klaric, M., Kolle, S., Kuehnl, J., Lambrechts, N., Lindstedt, M., Millet, M., Martinozzi-Teissier, S., Natsch, A., Petersohn, D., Pike, I., Sakaguchi, H., Schepky, A., Tailhardat, M., Templier, M., van Vliet, E. & Maxwell, G. (2015) Systematic evaluation of non-animal test methods for skin sensitisation safety assessment. *Toxicology in Vitro*. 29 (1), pp. 259-70. DOI: 10.1016/j.tiv.2014.10.018

Roberts, D. W., Patlewicz, G., Kern, P. S., Gerberick, F., Kimber, I., Dearman, R. J., Ryan, C. A., Basketter, D. A. & Aptula, A. O.(2007) Mechanistic applicability domain classification of a local lymph node assay dataset for skin sensitization. *Chem Res Toxicol,* 20 (7), pp. 1019-30. DOI: 10.1021/tx700024w

Roberts, D.W. (2018) Is a combination of assays really needed for non-animal prediction of skin sensitization potential? Performance of the GARD (Genomic Allergen Rapid Detection) assay in comparison with OECD guideline assays alone and in combination. *Regul Toxicol Pharmacol*. 98, pp. 155-160. DOI: 10.1016/j.yrtph.2018.07.014

Roberts, D.W. (2019) Response to the comments of Natsch and Landsiedel on my paper "Is a combination of assays really needed for non-animal prediction of skin sensitization potential? Performance of the GARD (genomic allergen rapid detection) assay in comparison with OECD guideline assays alone and in combination." *Regul Toxicol Pharmacol*. 102, pp. 95-97. DOI: 10.1016/j.yrtph.2019.01.006.

Rovida, C., Martin, S. F., Vivier, M., Weltzien, H. U. & Roggen, E. (2013) Advanced tests for skin and respiratory sensitization assessment. *ALTEX.* 30 (2), pp. 231-52. DOI: 10.14573/altex.2013.2.231

Rustemeyer, T., Van Hoogstraten, I. M. W., Von Blomberg, B. M. E. & Scheper, R. J. (2012) Mechanisms of Allergic Contact Dermatitis. In*:* Rustemeyer, T., Elsner, P., John, S.-M. & Maibach, H. I. (eds.) *Kanerva's Occupational Dermatology.* Berlin, Heidelberg: Springer Berlin Heidelberg. [online] Available through:

https://link.springer.com/referenceworkentry/10.1007%2F978-3-642-02035-3_14 [Accessed: 22nd September 2020]

Stevenson, M., Czekala, L., Simms, L., Tschierske, N., Larne, O. & Walele, T. (2019) The use of Genomic Allergen Rapid Detection (GARD) assays to predict the respiratory and skin sensitising potential of e-liquids. *Regulatory Toxicology and Pharmacology.* 103, pp. 158-165. DOI: 10.1016/j.yrtph.2019.01.001

Tice, C.M. (2001) Selecting the right compounds for screening: does Lipinski's Rule of 5 for pharmaceuticals apply to agrochemicals? *Pest Management Science.* 57 (1), pp. 3-16. DOI: 10.1002/1526-4998(200101)57:1<3::AID-PS269>3.0.CO;2-6.

United States Environmental Protection Agency. (2006) Reregistration Eligibility Decision for Dicamba and Associated Salts. Available at: https://nepis.epa.gov/Exe/ZyPDF.cgi/P10049M3.PDF?Dockey=P10049M3.PDF [Accessed: 22nd September 2020]

Uruno, A. & Motohashi, H. (2011) The Keap1–Nrf2 system as an in vivo sensor for electrophiles. *Nitric Oxide. 25 (2)*, pp. 153-60. DOI: 10.1016/j.niox.2011.02.007

Van Der Laan, J. W., Chapin, R. E., Haenen, B., Jacobs, A. C. & Piersma, A. (2012) Testing strategies for embryo-fetal toxicity of human pharmaceuticals. Animal models vs. *in vitro* approaches: a workshop report. *Regul Toxicol Pharmacol.* 63 (1), pp. 115-23. DOI: 10.1016/j.yrtph.2012.03.009

Vovk, V. Gammerman, A. & Shafer, G. (2005) *Algorithmic Learning in a Random World*. Boston, MA: Springer US. pp. 17-51.

Wang, C., Glenn, K.C., Kessenich, C., Bell, E., Burzio, L.A., Koch, M.S., Li, B. & Silvanovich, A.(2016) Safety assessment of dicamba mono-oxygenases that confer dicamba tolerance to various crops. *Regulatory Toxicology and Pharmacology*. 81, pp. 171-182. DOI: 10.1016/j.yrtph.2016.08.014

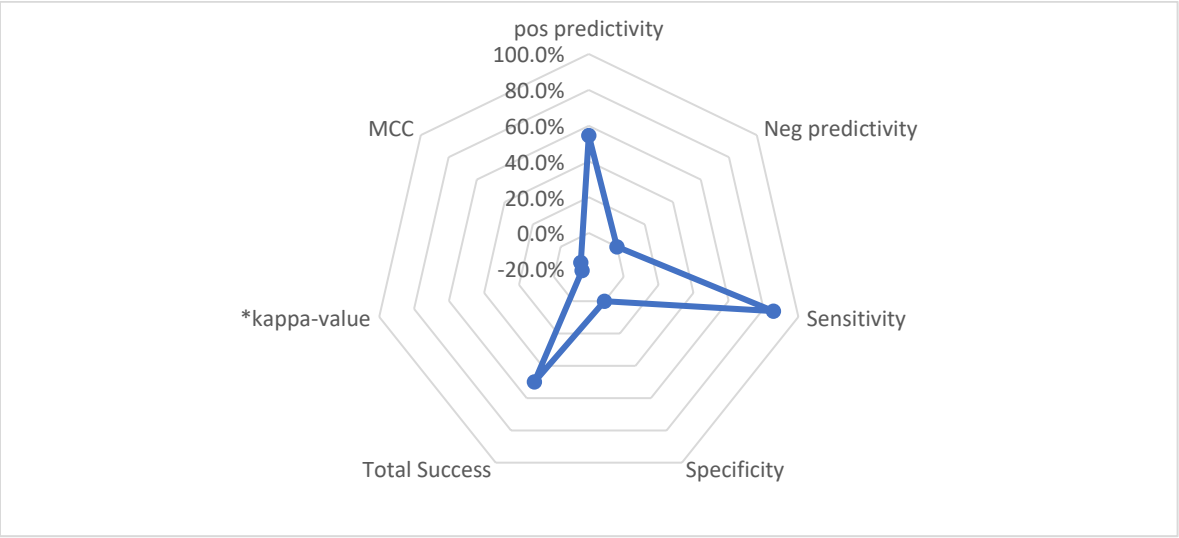
Wilm, A., Kühnl, J. & Kirchmair, J. (2018) Computational approaches for skin sensitization prediction. *Critical Reviews in Toxicology.* 48 (9), pp. 738-760. DOI: 10.1080/10408444.2018.1528207.

Worth, A. P. & Balls, M. (2004) The principles of validation and the ECVAM validation process. *Altern Lab Anim. 32 (1)*, pp. 623-629. DOI: 10.1177/026119290403201s105.

Yao, L., Jia, X., Zhao, J., Cao, Q., Xie, X., Yu, L., He, J. & Tao, Q. (2015) Degradation of the herbicide dicamba by two sphingomonads via different O-demethylation mechanisms. *International Biodeterioration & Biodegradation.* 104, pp. 324-332. DOI: 10.1016/j.ibiod.2015.06.016.

Zeller, K.S., Forreryd, A., Lindberg, T., Gradin, R., Chawade, A. & Lindstedt, M. (2017) The GARD platform for potency assessment of skin sensitizing chemicals. *ALTEX*, 34 (4), pp. 539-559. DOI: 10.14573/altex.1701101

Figure 1. Various statistical parameters adopted to evaluate prediction of skin sensitisation potential by the GARD assay conducted on the test set (12 agrochemical active ingredients)



Matthews Correlation Coefficient (MCC)

Figure 2. The molecular weight and Log P values of both the GARD training set (Forreryd et al., 2018) and agrochemical compounds tested

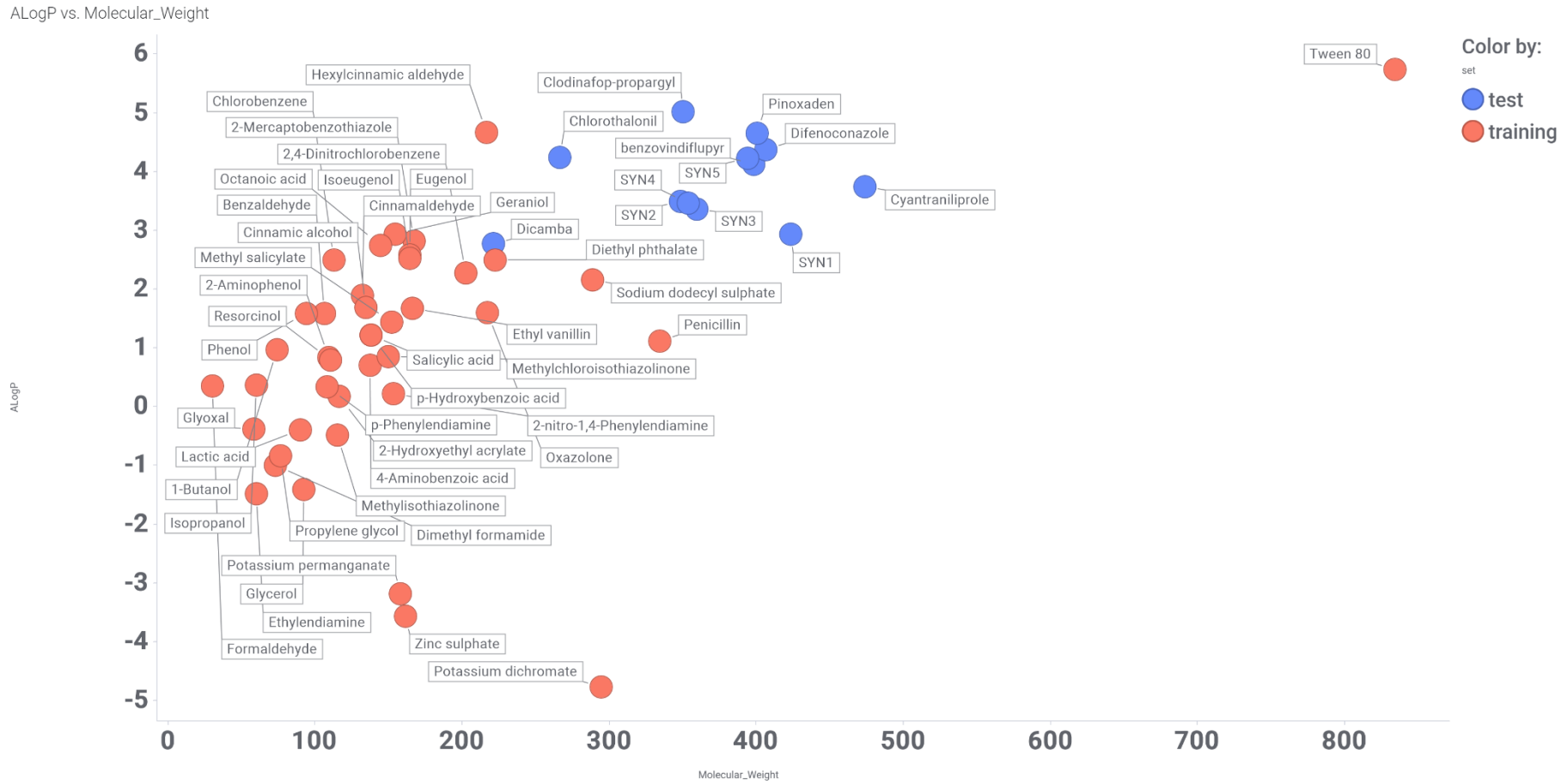
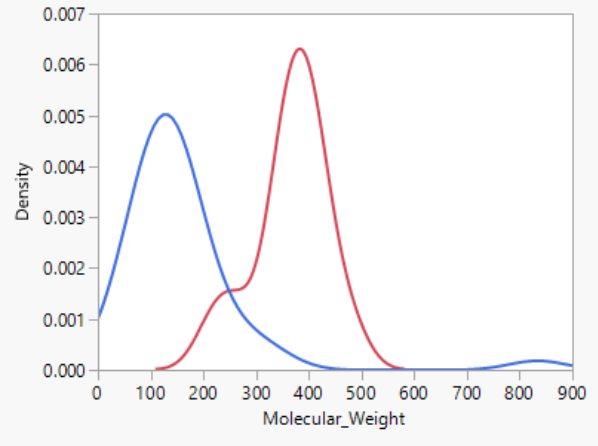
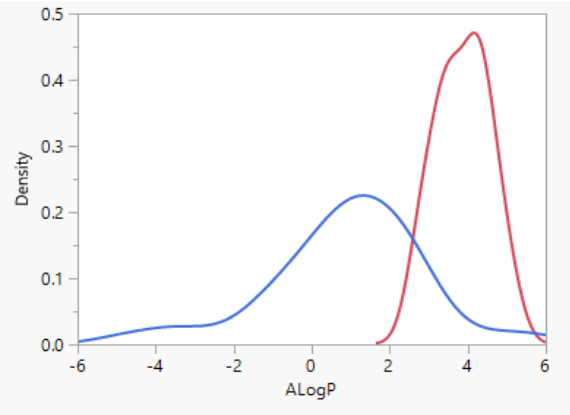


Figure 3. The distribution of molecular weights of the GARD training set of compounds (Forreryd et al., 2018) and the agrochemicals tested



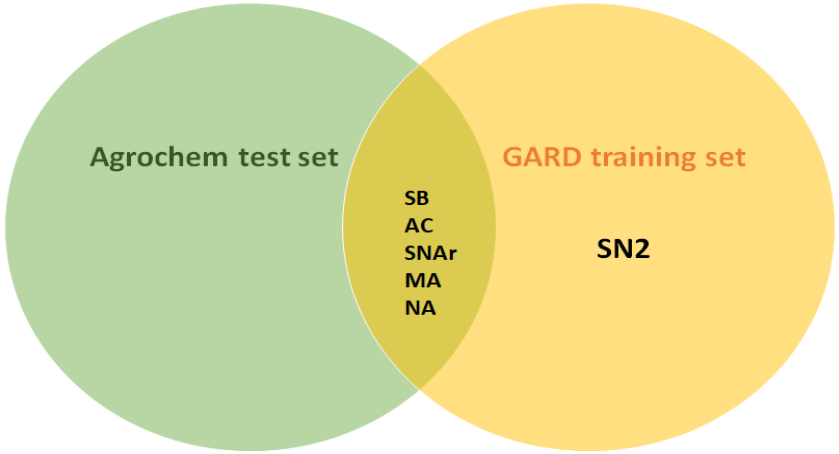
Red	Agrochemical test set
Blue	GARD training set

Figure 4. The distribution of Log P values of the GARD training set of compounds and the agrochemicals tested



Red	Agrochemical test set
Blue	GARD training set

The Venn diagram of the chemical domains identified in the two chemical sets i.e. the GARD training set and the agrochemicals. Training data set as indicated in Forreryd et al., 2018



MA	Michael addition
SB	Schiff base
NA	Nucleophilic addition
AC	Acylation
SNAr	Aromatic nucleophilic substitution
SN2	Bimolecular nucleophilic substitution

Table 1. The active ingredient, agrochemical use and *in vivo* skin sensitisation outcomes of agrochemical compounds used in this study.

Agrochemical Active Ingredient**	**Indication (F, H, I) ***	***In Vivo* Outcome**	***In Vivo* Study**	**Skin Sensitisation Harmonised Classification Labelling and Packaging (CLP) category**
benzovindiflupyr	F	Negative	LLNA	Not classified (EFSA, 2015, FAO, 2014)
chlorothalonil	F	Positive	Buehler	Skin Sens. 1, H317 (EFSA et al., 2018, O'Malley, 2010)
clodinafop-propargyl	H	Positive	GPMT	Skin Sens. 1, H317 (EFSA et al., 2020)
cyantraniliprole	I	Negative	LLNA	Not classified (FAO, 2014, EFSA, 2014)
dicamba	H	Negative	LLNA	Not classified (EFSA, 2011a, Harp, 2010, EPA, 2006, ECHA, 2008b)
difenoconazole	F	Negative	Buehler	Not classified (EFSA, 2011b)
pinoxaden	H	Negative Positive (EC3 =0.43%)	GPMT LLNA***	Skin Sens. 1A, H317 (EFSA, 2013, FAO, 2016)
SYN1	I	Positive (EC3 =0.13%)	LLNA	No harmonised classification
SYN2	I	Positive (EC3 =1.1%)	LLNA	No harmonised classification
SYN3	I	Positive	rLLNA	No harmonised classification
SYN4	I	Negative	rLLNA	No harmonised classification
SYN5	I	Positive	rLLNA	No harmonised classification

***F: fungicide, H: herbicide, I: insecticide**

****SYN1 - SYN5: anonymised agrochemical active ingredients**

***** The result corresponding with the harmonised classification (LLNA) has been used for the purposes of comparison. These studies were considered OECD & GLP compliant**

Table 2. OECD QSAR Toolbox prediction scheme

OECD QSAR Toolbox prediction scheme	
Sensitiser	OASIS GHS profiler - 1A or 1B OASIS w/metabolism profiler - 1A
Non-sensitiser	OASIS GHS profiler - No alert OASIS GHS w/metabolism profiler - 1B

Table 3. Protein binding alerts*, *in vivo* study results, Rv90, GARD input concentration, GARD skin results, GARD decision values**

Test material	Structural alert**	*in vivo* study	*in vivo* study result	Rv90***	GARD input concentration	GARD Decision Value (Mean ± SD)	GARD skin result
benzovindiflupyr	AC/SB	LLNA	negative	40 µM	40 µM	6.0±0.9	positive
chlorothalonil	SNAr	Buehler	positive	0.5 µM	0.5 µM	4.6±1.4	positive
clodinafop-propargyl	No alert	GPMT	positive	-	100 µM	6.1±0.9	positive
cyantraniliprole	AC/SB	LLNA	negative	-	100 µM	3.4±0.7	positive
dicamba	SB	LLNA	negative	-	500 µM	0.0±0.8	positive
difenoconazole	No alert	Buehler	negative	50 µM	50 µM	6.3±0.9	positive
pinoxaden	No alert	LLNA	positive (EC3 =0.43%)	-	500 µM	-0.5±0.5	negative
SYN1	SB/NA	LLNA	positive (EC3 =0.13%)	-	100 µM	1.1±0.7	positive
SYN2	No alert	LLNA	positive (EC3 =1.1%)	140 µM	140 µM	6.9±0.4	positive
SYN3	MA	rLLNA	positive	-	100 µM	3.4±0.6	positive
SYN4	No alert	rLLNA	negative	250 µM	250 µM	6.4±0.4	positive
SYN5	SNAr	LLNA	positive (EC3 = 0.9%)	50 µM	50 µM	4.4±0.5	positive

*** AC, Acylation; MA, Michael addition; NA, Nucleophilic addition; SB, Schiff base formation; SNAr, Aromatic nucleophilic substitution;**

****Reaction domains were assigned based on expert judgment using the chemistry defined in Enoch et al (2011)*****Rv90 - concentration of test substance inducing 90% relative viability.**

Positive control p-phenylendiamine,

Negative control dimethylsulfoxide

Table 4. Test results of agrochemical test set *in vivo* skin sensitisation results versus the GARD assay results

	Pred. Pos	Pred. Neg.
in vivo Obs. Pos.	6	1
in vivo Obs. Neg.	5	0

Table 5. Statistical parameters used for evaluation of the GARD assay predictions of the agrochemical test set results versus the in vivo skin sensitisation assay results

Positive predictivity	54.5%
Negative predictivity	0.0%
Sensitivity	85.7%
Specificity	0.0%
Total Success/Accuracy	50.0%
*kappa-value	-0.16
MCC	-0.14

***kappa-value: < 0.20 poor, 0.21 - 0.40 fair, 0.41 - 0.60 moderate, 0.61 - 0.80 substantial**

Table 6. GARD assay conformal predictions of the test items

Test material	*in vivo* Study result	GARD skin prediction	Psens*	Pnon-sens**	Conformal Prediction
benzovindiflupyr	negative	positive	0.84	0.02	sensitiser
chlorothalonil	positive	positive	0.76	0.05	sensitiser
clodinafop-propargyl	positive	positive	0.85	0.02	sensitiser
cyantraniliprole	negative	positive	0.63	0.05	sensitiser
dicamba	negative	positive	0.14	0.16	non-sensitiser
difenoconazole	negative	positive	0.85	0	sensitiser
pinoxaden	positive (EC values)	negative	0.11	0.4	non-sensitiser
SYN1	positive	positive	0.31	0.07	sensitiser
SYN2	positive	positive	0.85	0	sensitiser
SYN3	positive	positive	0.63	0.05	sensitiser
SYN4	negative	positive	0.85	0	sensitiser
SYN5	positive	positive	0.72	0.05	sensitiser

***A measure of the Test Item non-conformity compared to the calibration set. If the p-value is below the error level 0.15 the Test Item is strange compared to calibration sensitisers. A value of greater than 0.15 indicates that it belongs to the group sensitisers with 85% confidence.**

****A measure of the Test Item non-conformity compared to the calibration set. If the p-value is below the error level 0.15 the Test Item is strange compared to calibration non-sensitisers. A value above 0.15 is therefore proof that it belongs to the group non-sensitisers with a confidence of 85%.**

Table 7. Test set versus training set molecular properties

	Training set				Test set			
Properties	**Range**	**1st Quartile**	**Median**	**3rd Quartile**	**Range**	**1st Quartile**	**Median**	**3rd Quartile**
Molecular weight	30 to 834	106	138	164	221 to 473	349	377	402
Log P	-4.77 to 5.74	0.18	1.12	2.17	2.78 to 5.02	3.44	3.94	4.28
HBA	0 to 12	1	2	3	2 to 6	4.5	5	5
HBD	0 to 3	0	1	2	0 to 2	0	0	0.25
RB	0 to 19	0	1	3	0 to 7	3.75	6	7

Table 8. The OECD QSAR Toolbox prediction for skin sensitisation against the *in vivo* experimental results of the test and training sets

Number of compounds in each class	Training set	Predicted Sensitiser	Predicted non-sensitiser
20	Exp-sens	85%	15%
20	Exp-NS	20%	80%
40			

Number of compounds in each class	Test set	Predicted Sensitiser	Predicted non-sensitiser
7	Exp-sens	71%	29%
5	Exp-NS	60%	40%
12			