Enoch, SJ, Hasarova, Z, Cronin, MTD, Bridgwood, K, Rao, S, Kluxen, F and Frericks, M

Sub-structure-based category formation for the prioritisation of genotoxicity hazard assessment for pesticide residues: Sulphonyl ureas

Sub-structure-based category formation for the prioritisation of genotoxicity hazard assessment for pesticide residues: Sulphonyl ureas

S. J. Enoch[1], Z. Hasarova[1], M. T. D. Cronin[1], K. Bridgewood[2], S. Rao[3], F. M. Kluxen[4], M. Frericks[5]

[1]School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Liverpool, England

[2]Syngenta, Bracknell, England

[3]Gowan Company, Yuma, Arizona, USA

[4]ADAMA Deutschland GmbH, Cologne, Germany

[5]BASF, Limburgerhof, Germany

Corresponding author: s.j.enoch@ljmu.ac.uk

Abstract

In dietary risk assessment, residues of pesticidal ingredients or their metabolites need to be evaluated for their genotoxic potential. The European Food Safety Authority recommend a tiered approach focussing assessment and testing on classes of similar chemicals. To characterise similarity and to identify structural alerts associated with genotoxic concern, a set of chemical sub-structures was derived for an example dataset of 74 sulphonyl urea agrochemicals for which either Ames, chromosomal aberration or micronucleus test results are publicly available. This analysis resulted in a set of seven structural alerts that define the chemical space, in terms of the common parent and metabolic scaffolds, associated with the sulphonyl urea chemical class. An analysis of the available profiling schemes for DNA and protein reactivity shows the importance of investigating the predictivity of such schemes within a well-defined area of structural space. Structural space alerts, covalent chemistry profiling and physico-chemistry properties were combined to develop chemical categories suitable for chemical prioritisation. The method is a robust and reproducible approach to such read-across predictions, with the potential to reduce unnecessary testing. The key challenge in the approach was identified as being the need for pesticide-class specific metabolism data as the basis for structural space alert development.

Keywords: read-across; mutagenicity; sulphonyl urea; metabolism; category formation

1. Introduction

The European Food Safety Authority (EFSA) guidance on the establishment of residue definition for the dietary risk assessment for genotoxicity specifically outlines the usage of category formation and read-across [1]. However, at the time of writing, this guidance has not been agreed between stakeholders and EFSA. Thus, there is currently no established legal framework for the use of read-across (or quantitative/qualitative structure-activity relationship methods, more generally). Therefore, a robust scientific weight-of-evidence needs to be established for such methods to become commonplace for the prediction of genotoxicity of pesticide residues. For these purposes, the term "residue" is defined as any compound associated with the active ingredient that may result in risk to human and/or livestock following the application of a pesticide. EFSA have published a workflow to enable the use of read-across to predict either the presence or absence of genotoxicity within a category of similar chemicals where data may be missing or incomplete [1]. In cases where genotoxicity is predicted, further testing is required to confirm the read-across prediction. The test strategy needs to ensure that a representative number of the chemicals in the category are tested for gene mutation as well as structural and numerical chromosomal aberration. A battery of *in vitro* and *in vivo* tests is recommended by EFSA to cover the three key genotoxicity endpoints with minimal animal usage [2]. The initial battery is typically the Ames test (gene mutation) and an *in vitro* micronucleus test (structural and numerical chromosomal aberration). This combination of testing is considered state of the art within most regulatory guidelines, with *in vivo* testing only being conducted as a higher tier to evaluate positive *in vitro* micronucleus assay results [2]. In contrast to this process for positive predictions, the absence of genotoxicity within a category (with data gaps being filled via read-across) requires no further genotoxicity testing.

The key step in the use of the category approach is the ability to confidently define 'similarity' between chemicals [3-5]. In terms of the use of category formation in the EFSA genotoxicity workflow noted above, defining similarity is relatively straightforward for potentially genotoxic chemicals. This is due to the key molecular initiating event for DNA-reactive genotoxicity being the formation of a covalent bond between nucleophilic centres in DNA and a chemical capable of behaving as an electrophile (either directly or after metabolic activation) [6-12]. The associated chemistry can be encoded easily as structural alert-based *in silico* profilers that enable chemicals to be assigned to a category based on the presence of a common alert. One well-established (others are available) computational tool for this type of analysis being the OECD QSAR Toolbox which contains a range of profilers of this type. In contrast, defining similarity between chemicals that lack an alert for DNA reactivity is more challenging [13]. One of the most common options (after identifying simple structural analogues) being to utilise molecular fingerprint-based methods coupled with a similarity metric such as the Tanimoto coefficient [14-18].

As part of an extensive study into the use of *in silico* methods for the prediction of genotoxicity, a recent publication outlined the use of Atom Centred Pair molecular fingerprints and the Dice similarity metric to identify analogues suitable for read-across for a series of target chemicals [19]. The study focussed on predicting the Ames and *in vitro* chromosomal aberration tests, with the results showing significantly better predictivity for the Ames test. The authors suggest better predictivity is due to greater number of Ames test results in the dataset (this increased predictivity is also partly due to the numerous variants of the chromosomal aberration assay, some of which lack DNA repair capability). Interestingly, this work also looked at both one-to-one and one-to-many read-across predictions – the results showing that the inclusion of additional chemicals added weight of evidence to the predictions. However, this analysis highlights the key challenge of using fingerprint methods in that the authors had to identify the value of the similarity metric at which chemicals were no longer considered similar enough to be within the same category as the target chemical (defined initially as 70% and then lowered to 60%). In addition to the fingerprint method, the use of sub-structure searching as a method to identify similar chemicals was also investigated [18]. This involved defining a key sub-structure within the target chemical and using the presence of this sub-structure as an initial screening tool to identify category members. The final category membership was determined via elimination of chemicals with a similarity score of less than 70% (as determined via Atom Centred Pair fingerprints and the Dice similarity metric). This analysis showed the potential for the development of key sub-structures to improve the structural space of the category (in terms of relevance to the target chemical).

Whilst the development of key sub-structures is vital for the proper definition of categories – and hence better read-across for residues – this has not been undertaken in a formal and systematic approach. The aim of the current investigation was to develop a methodology to derive chemical sub-structures considered to be the key drivers of chemical similarity (for chemicals lacking an alert for DNA reactivity) which could be applied to pesticide residues. The novelty in this investigation was rather than defining a sub-structure based on the target chemical, sub-structures were used to define the structural space associated with Ames, chromosomal aberration, and micronucleus test results. Importantly, the structure-toxicity relationships within the resulting sub-structure-based categories were investigated to establish their utility for use within the read-across step in the EFSA framework for residue risk assessment. Once a theoretical approach to derive the sub-structures was developed, this was implemented in a workflow with the example of the sulphonyl urea pesticides being used as a case study.

2. Methods

To assess chemical similarity and to identify structural alerts associated with genotoxicity concern, several key steps were required. The first was the identification of a dataset of sulphonyl urea pharmaceutical and agrochemical active ingredients with associated genotoxicity data. These structures were placed onto a metabolic map of the key transformations present for the sulphonyl urea chemical class, enabling common metabolic scaffolds to be developed into structural space alerts. The key steps required for this type of analysis being as follows:

2.1 Creation of a dataset for read-across and chemical space analysis

An initial dataset of 21 sulphonyl urea pesticide active ingredients was identified based on the availability of EFSA Draft Assessment Report (DAR)/Renewal Assessment Report (RAR) documents (www.efsa.europa.eu/en). Only chemicals for which a DAR/RAR document was available were included in the initial dataset. The chemical space of this initial dataset was expanded through a data harvesting exercise from the DAR/RAR documents in which any metabolite for which experimental genotoxicity data had been generated was added to the dataset. For the current study genotoxicity data were defined as either Ames, chromosomal aberration, or micronucleus test results. This data harvesting exercise resulted in a final dataset of 74 sulphonyl urea agrochemical active ingredients and metabolites. All structural space alerts were developed from this dataset. The dataset, termed the 'sulphonyl urea genotoxicity dataset' contained the following test results (*in vitro* assays with S9 fraction, Ames tests in the standard battery):

- Ames - 73 chemicals (all negative)
- *in vitro* chromosomal aberration - 53 chemicals (41 negative, 12 positive)
- *in vivo* chromosomal aberration - 5 chemicals (all negative)
- *in vitro* micronucleus – 4 chemicals (3 negative, 1 positive)
- *in vivo* micronucleus - 26 chemicals (26 negative, 0 positive)

A second dataset containing 697 additional genotoxicity test results from a wider range of agrochemicals was also utilised [20]. This dataset was denoted as the 'EFSA genotoxicity dataset'. Importantly, this dataset was used to assess the predictivity of the profiling schemes within the OECD QSAR Toolbox (detailed in section 2.4) and as a source of potential analogues for the read-across case studies only (section 3.3) – it was not used to define develop the structural space alerts. Full datasets are available in the Supplementary Information.

Chemical structures for both data sets were recorded with relevant identifiers e.g., name (where applicable), CAS numbers if available and SMILES strings. This information is also available in the Supplementary Information.

2.2 Structural space alert development

The development of the structural space alerts utilised the following protocol (the overall process is summarised in Figure 1):

1. Definition of the metabolic map for the sulphonylurea pesticides: This analysis involved inspection of the available metabolism data in the 21 DAR/RAR documents to identify metabolic transformations common to the sulphonyl urea active ingredients (www.efsa.europa.eu/en). In terms of the sulphonyl urea pesticides, these metabolic transformations were hydrolysis reactions resulting in the cleavage of the sulphonyl urea moiety and hydroxylation reactions on the aromatic rings. These reactions were common to all the sulphonyl urea pesticides in the dataset. The metabolic map is as shown in Figure 2.
2. Metabolic scaffold identification: Common metabolic scaffolds were then identified from the metabolic map developed in step 1. This involved applying the metabolic transformations defined in the metabolic map to the sulphonyl urea active ingredients. The resulting structures from this analysis were defined as metabolic scaffolds. These metabolic scaffolds were grouped together based on maximum common sub-substructures for development into structural space alerts.
3. Structural space alert development: The metabolic scaffolds identified in step 2 were used to profile the sulphonyl urea genotoxicity dataset – metabolic scaffolds that had genotoxicity data associated with them (either Ames, chromosomal aberration, or micronucleus test data) were denoted as structural space alerts and encoded as SMARTS patterns. No additional physico-chemical boundaries were imposed on the structural space alerts (enabling then to identify any chemical containing the alert sub-structure).

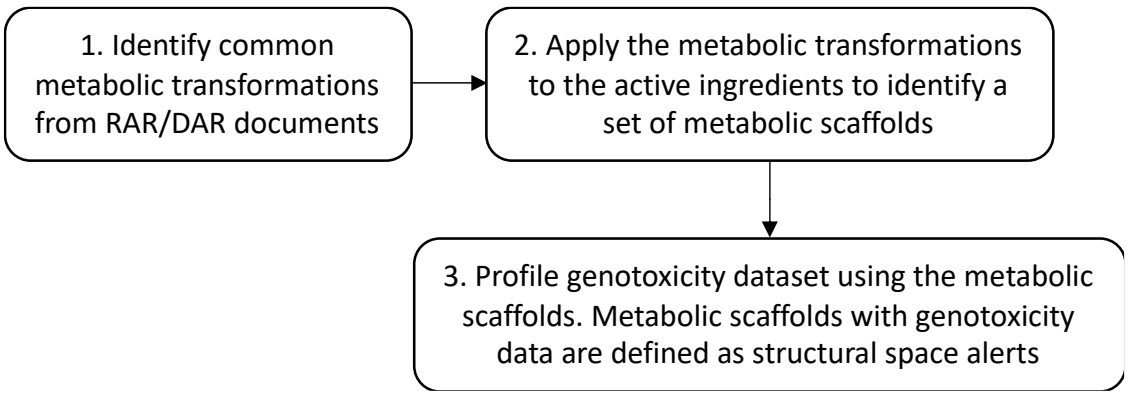


Figure 1: Flow chart outlining the protocol for the development of structural space alerts

2.3 Chemoinformatics analysis

All chemoinformatics analyses were carried out using the KNIME data analytics platform (V4.2.2). Structural space alerts were encoded as SMARTS patterns. These were developed and tested using the RDkit substructure node. The logarithm of the octanol-water partition coefficient (SlogP) and molecular weight (MW) were calculated using the RDKit Descriptor Calculation node in KNIME for all chemicals in the sulphonyl urea genotoxicity and EFSA genotoxicity datasets. Chemical similarity analysis using MACCS structural keys and Morgan fingerprints, coupled to the Tanimoto similarity metric, was performed using the RDKit Fingerprint node (fingerprint analysis was utilised as part of the read-across case studies in section 3.3).

2.4 Chemical profiling

Chemicals in both datasets were profiled using computational tools for the potential for covalent interactions relevant to genotoxicity within the OECD QSAR Toolbox (V4.1.1). These profilers being (CA is chromosomal aberration and MNT is the micronucleus test):

- OECD: DNA binding by OECD
- OASIS: DNA binding by OASIS; DNA alerts for AMES, CA and MNT by OASIS; protein binding alerts for CA by OASIS
- ISS: *in vitro* mutagenicity (Ames test) alerts by ISS, *in vivo* micronucleus alerts by ISS

2.5 Read-across

Read-across was attempted for two metabolites of the herbicide prosulfuron (SYN547308 and SYN542604 in [21]) with no data. Specifically, predictions of Ames, chromosomal aberration and micronucleus test results were attempted from suitable analogues. Each metabolite was assigned to a relevant structural space category based on the presence of a common structural space alert. The domain of the category was then defined based on similarity between the metabolite and category members in terms of DNA/protein binding and key physico-chemical properties. Subsequent read-across predictions were made using a majority rules basis.

3. Results and Discussion

The aim of this study was to develop a workflow to implement a set of structural alerts to define the structural space associated with a set of sulphonyl urea agrochemical active ingredients and metabolites for which genotoxicity data exist. The development of these structural space alerts was driven by the need to prioritise potential metabolites (coming from plants and other animals) for further testing using a read-across approach. This study also

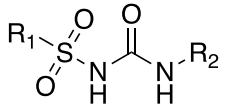
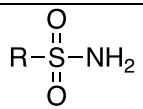
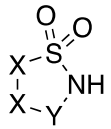
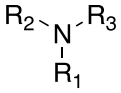
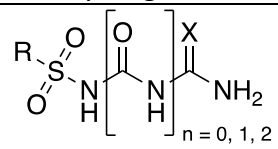
demonstrated how these structural space alerts can be used in conjunction with existing *in silico* profilers designed to identify DNA and/or protein reactive chemicals.

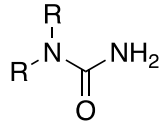
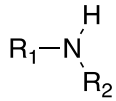
3.1 Structural space definition

The structural space associated with the sulphonyl urea genotoxicity dataset was defined through a set of seven structural space alerts as shown in Table 1. This set of structural space alerts identified 64 of the 74 chemicals in the sulphonyl urea dataset from which they were developed. Inspection of the remaining 10 chemicals showed them to have no common sub-structures suitable for further alert development. The sub-structures of these seven structural space alerts are depicted in Table 1. In addition, a summary of the genotoxicity data for the sulphonyl ureas, associated with the seven alerts, is provided in Table 1 (denoted as "SU"). The data in Table 1 show that only structural space alerts 1 and 2 have more than a single *in vivo* assay result associated with them, with only structural space alert 1 having data from both *in vivo* assays. All of these *in vivo* test results were negative. In terms of *in vitro* data, all the structural space alerts had Ames and chromosomal aberration data, with only alert 1 and having data from the micronucleus test. Interestingly, despite most of the *in vitro* and *in vivo* assay results being negative, *in vitro* chromosomal aberration test showed positive results for at least one chemical associated for nearly every alert.

The chemical domain associated with the eight structural space alerts was expanded by profiling the larger EFSA genotoxicity dataset (697 chemicals), with the results summarised in Table 1 (denoted as "EFSA"). This analysis resulted in the identification of an additional 46 chemicals with data (ten *in vivo* micronucleus, one *in vitro* micronucleus, five *in vivo* chromosomal aberration, 33 *in vitro* chromosomal aberration and 44 Ames assay test results). The inclusion of these chemicals increased the numbers of relevant chemicals with Ames and *in vitro* chromosomal aberration data for most of the structural space alerts. However, the inclusion of this dataset only added further *in vivo* data to alerts 1 - 4.

Table 1: Structural space alerts developed from the sulphonylurea genotoxicity dataset with number of compounds associated with the various mutagenicity endpoints from the sulphonyl urea (SU) and EFSA data sets (N.B. structural space alerts were developed from the SU dataset).

			In vitro						*In vivo*			
			Ames		CA		MNT		CA		MNT	
Alert	Structural space alert	Dataset	+	-	+	-	+	-	+	-	+	-
1	 R_1 = aromatic carbon R_2 = pyrimidine or triazine	SU	0	27	6	17	0	3	0	5	0	17
		EFSA	0	11	2	8	0	1	0	5	0	4
		Total	**0**	**38**	**8**	**25**	**0**	**4**	**0**	**10**	**0**	**21**
2	 R = aromatic carbon	SU	0	12	4	3	0	0	0	0	0	6
		EFSA	0	10	1	6	0	0	0	0	0	3
		Total	**0**	**22**	**5**	**9**	**0**	**0**	**0**	**0**	**0**	**9**
3	 X = aromatic carbon Y = aliphatic carbon (sp^3 or sp^2 hybridised)	SU	0	3	0	2	0	0	0	0	0	0
		EFSA	0	1	0	1	0	0	0	0	0	0
		Total	**0**	**4**	**0**	**3**	**0**	**0**	**0**	**0**	**0**	**0**
4	 R_1 = pyrimidine or triazine R_2, R_3 = hydrogen or methyl	SU	0	6	0	2	1	0	0	0	0	1
		EFSA	0	16	1	10	0	0	0	0	1	1
		Total	**0**	**22**	**1**	**12**	**1**	**0**	**0**	**0**	**1**	**2**
5	 n = 0, 1, 2	SU	0	8	1	8	0	0	0	0	0	0
		EFSA	0	2	0	2	0	0	0	0	0	0

	R = aromatic carbon X = O, NH	**Total**	**0**	**10**	**1**	**10**	**0**	**0**	**0**	**0**	**0**	**0**
6	 R_1 = aromatic carbon R_2 = pyrimidine or triazine	SU	0	5	0	4	0	0	0	0	0	0
		EFSA	0	0	0	0	0	0	0	0	0	0
		Total	**0**	**5**	**0**	**4**	**0**	**0**	**0**	**0**	**0**	**0**
7	 R_1 = aromatic carbon R_2 = pyrimidine or pyrazine ring	SU	0	3	1	0	0	0	0	0	0	1
		EFSA	0	4	0	2	0	0	0	0	0	1
		Total	**0**	**7**	**1**	**2**	**0**	**0**	**0**	**0**	**0**	**2**

Abbreviations: CA = chromosomal aberration; MNT = micronucleus test, SU = sulphonyl urea genotoxicity dataset, EFSA = EFSA genotoxicity dataset.

3.2 In silico profiler analysis

The chemicals with relevant genotoxicity data identified by the structural space alerts outlined in Table 1 were profiled using a set of DNA and protein binding profilers within the OECD QSAR Toolbox applicable to genetic toxicity. This set of profilers being those outlined in the EFSA genotoxicity read-across guidance documents [1, 2]. These profiling schemes are split into two classes: general mechanistic and endpoint specific. The general mechanistic profiling schemes outline a broad spectrum of structural alerts developed from a knowledge of the potential organic chemistry associated with DNA and/or protein binding. Importantly, toxicological data are not necessarily associated with these alerts. In contrast, the endpoint specific profiling schemes were developed from an analysis of toxicological data making them more focussed. A summary of the results of this profiling, in terms of the numbers of compounds hit, is shown in Table 2.

3.2.1 General mechanistic profilers

The DNA binding by OASIS and OECD profiling schemes both identified more than 20 chemicals in the combined dataset featuring at least one structural alert associated with DNA reactivity (24 and 55 chemicals respectively). Inspection of the profiling results showed two alerts to be triggered in these chemicals: an alert for Schiff base formation for 20 of the 24 chemicals identified by the OASIS profiler and an alert for acylation for 45 of the 55 chemicals identified by the OECD profiler. Interestingly, the two profilers did not agree with one another for these predictions – with no alert being identified by the second profiler. In contrast to the DNA profiling schemes, the Protein binding by OASIS and OECD profiling schemes identified relatively few chemicals in the combined dataset (six and nine chemicals respectively). Inspection of the structural alerts identified showed there to be no common alerts within the chemicals that triggered an alert.

3.2.2 Endpoint specific profilers

The OASIS profilers in the OECD QSAR Toolbox showed only three chemicals (in the combined dataset) to contain structural features associated with a positive result in either the Ames, chromosomal aberration, or micronucleus tests. In contrast, 31 out of 110 chemicals have at least one structural feature associated with a positive Ames result in the *in vitro* mutagenicity (Ames test) alerts by ISS. Furthermore, the *in vivo* micronucleus alerts by ISS profilers identified 102 out of 110 chemicals as having a structural feature associated with a positive micronucleus test result. Inspection of the poor performance of the ISS profiler for the micronucleus test showed it to be due to the "H-acceptor-path3-H-acceptor" structural alert. This alert was identified in 84 of the 110 chemicals in the combined datasets. The predictivity of this alert is of concern given that there are 35 chemicals within the combined datasets with an *in vivo* micronucleus test result, 34 of which are negative (Table 3). This confirms the analysis of Benigni et al. [21] that showed this alert to have significantly lower predictivity

than the other alerts in the profiler [22]. In addition, both ISS profilers consistently identify 2-aminopyrimidines and 2-amino-1,3,5-triazines as being potentially genotoxic (four in the SU dataset and 13 in the EFSA dataset). Inspection of the experimental data shows that these chemicals are not genotoxic suggesting that these ring systems do not undergo the same metabolic transformation into a nitrenium ion as the benzene equivalent [6]. This is hypothesis is supported by the profiling results from the OASIS profilers which did not flag these chemicals as being potentially DNA-reactive.

3.2.3 Profiling summary

Overall, the profiling results summarised in Table 2 suggest that the endpoint specific schemes are more applicable than the general mechanistic profilers for the domain defined by structural space alerts in Table 1. The reason for this being that the general mechanistic profilers, by design, cover a very broad range of covalent chemistry – much of which is not associated with a given endpoint (meaning these profilers tend to be overly predictive). Such over-predictivity is, in certain circumstances advantageous e.g., in the analysis of new compound classes, although it often requires subsequent sub-categorisation. However, when available, the endpoint specific profilers offer a more tightly defined set of alerts making them significantly more useful within a read-across scheme.

In addition to the above, the analysis also showed that for the sulphonyl urea chemical space the endpoint specific OASIS profilers were significantly more applicable than the ISS profilers – this being due to several of the alerts within the ISS schemes being significantly over-predictive. Therefore, in terms of the current study, these results indicated that it is the endpoint specific OASIS profilers that should be utilised in conjunction with the structural space structural alerts to build the weight of evidence that a chemical is unlikely to be genotoxic.

Table 2: A summary of *in silico* profiling results showing the number of chemicals identified as having a structural alert using the four endpoint specific profilers for available in the OECD QSAR Toolbox (V4.4.1).

Profiling scheme	SU (64 chemicals)	EFSA (46 chemicals)	Combined (110 chemicals)
DNA binding by OASIS	19	5	24
DNA binding by OECD	38	17	55
Protein binding by OASIS	2	4	6
Protein binding by OECD	6	3	9
DNA alerts for AMES, CA and MNT by OASIS	0	1	1
Protein binding alerts for Chromosomal aberration by OASIS	1	1	2
In vitro mutagenicity (Ames test) alerts by ISS	9	22	31
In vivo mutagenicity (Micronucleus) alerts by ISS	59	43	102

Abbreviations: CA = chromosomal aberration; MNT = micronucleus test, SU = sulphonylurea genotoxicity dataset, EFSA = EFSA genotoxicity dataset

Table 3: Summary of the genotoxicity test results within the domain defined by the structural space alerts outlined in Table 1.

		In vitro			*In vivo*	
Dataset	Result	Ames	CA	MNT	CA	MNT
SU	Number of positive results	0	12	1	0	0
	Number of negative results	64	36	3	5	25
EFSA	Number of positive results	0	4	0	0	1
	Number of negative results	44	29	1	5	9
Total	**Number of positive results**	**0**	**16**	**1**	**0**	**1**
	Number of negative results	**108**	**65**	**4**	**10**	**34**

Abbreviations: CA = chromosomal aberration; MNT = micronucleus test, SU = sulphonyl urea genotoxicity dataset, EFSA = EFSA genotoxicity dataset

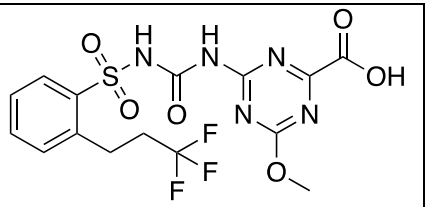
3.3 Read-across usage examples

An important part of the current study was to outline how the structural space alerts and endpoint specific profiling information can be used to identify chemicals that are unlikely to be genotoxic. In this case, the definition provided by EFSA that the absence of genotoxicity is a lack of gene mutation, structural (clastogenic) and numerical (aneugenic) chromosomal

aberration was applied [2]. The intended use of the structural space structural alerts outlined in Table 1 is to locate analogues suitable for a read-across prediction for data poor compounds. As an example, consider the metabolite SYN547308 from prosulfuron (identified from reference [21]). The structural space and endpoint specific profiling results for this chemical are as shown in Table 4. These results show this chemical to be within the structural space defined by structural alert 1 (alert detailed in Table 1). In addition, it does not feature any structural alerts associated with either covalent DNA or protein binding (as determined using the OASIS profilers). Finally, this chemical has an SlogP value of 1.58 and molecular weight of 449.06 g/mol, these fall within the ranges defined by the other category members (ranges as outlined in Table 4).

The data outlined for the structural space category shown in Table 4 suggest that chemicals within this category do not cause gene mutation or structural/numerical chromosomal aberration. This is based on an extensive set of data from the Ames assay, *in vitro* and *in vivo* chromosomal aberration and micronucleus tests, the majority of which were negative in these assays. Despite this, eight of the category members did show positive results in the *in vitro* chromosomal aberration assay (out of 33 tested). However, inspection of the chemicals associated with these positive data within the category showed two to be negative in the *in vivo* chromosomal aberration assay and a further four to be negative in the *in vivo* micronucleus assay (one of the remaining two chemicals showed an inconclusive result in the *in vivo* chromosomal aberration and the second had not been tested in either *in vivo* assay). This suggests that these six chemicals do not cause structural chromosomal aberration *in vivo* (as predicted by the *in vitro* assay results). These structure-toxicity relationships within the category enable SYN547308 to be predicted as non-genotoxic via read-across. Importantly, the available experimental data give the maximum confidence that members of this category do not cause gene mutation or structural/numerical chromosomal aberration. In addition, the category data presented in Table 4 suggest that no further genotoxicity testing is required for SYN547308 (all category data, including chemical structures, are available in the Supplementary Information).

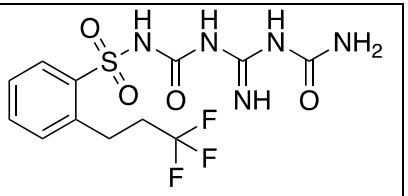
Table 4: *In silico* profiling results for metabolite SYN547308 of prosulfuron (metabolite identified in reference [21]).

Target structure (SYN547308)				
Structural space	Target	Alert 1	MACCS	Morgan
Number of category members	N/A	39	39	0
DNA alerts for AMES, CA and MNT	No alert	No alerts	No alerts	N/A
Protein binding alerts for CA	No alert	No alerts	No alerts	N/A
SlogP	1.58	-1.07 -> 2.43	-1.07 -> 3.64	N/A
MW	449.06	350.07 -> 492.96	256.01 -> 493.14	N/A
Ames	Negative (R/A from alert 1)	38 negatives 0 positives	38 negatives 0 positives	N/A
In vitro CA	Negative (R/A from alert 1)	25 negatives 8 positives	26 negatives 8 positives	N/A
In vivo CA	Negative (R/A from alert 1)	10 negatives 0 positives	10 negatives 0 positives	N/A
In vitro MNT	Negative (R/A from alert 1)	4 negatives 0 positives	3 negatives 0 positives	N/A
In vivo MNT	Negative (R/A from alert 1)	21 negatives 0 positives	20 negatives 0 positives	N/A

Abbreviations: CA = chromosomal aberration; MNT = micronucleus test; R/A = read-across; N/A = not applicable.

The structural space category approach can also be utilised to prioritise a chemical for further testing. Consider metabolite SYN542604 (also from prosulfuron), this chemical falls into the domain of structural space alert 5. This category contains 11 chemicals, ten with Ames data and 11 with *in vitro* chromosomal aberration test results (Table 5). All but one of these *in vitro* results are negative, the sole positive results being for *in vitro* chromosomal aberration. Taking a weight of evidence approach, this enables these two endpoints to be predicted via read across as being negative for SYN542604. In this example, only two of the three regulatory endpoints can be predicted for this chemical. This highlights the potential for the category approach to identify where targeted testing of several category members could enable weight of evidence to be employed to fill the data gap. In this example this would be to utilise the *in vitro* micronucleus assay to investigate the ability of category members to cause numerical chromosomal aberration. This test would also add further evidence to the *in vitro* chromosomal aberration data (as the micronucleus it also tests for structural chromosomal aberration). Full data for the category, including all chemical structures are available in the Supplementary Information.

Table 5: *In silico* profiling results for metabolite SYN542604 of prosulfuron (metabolite identified in reference [21]).

Target structure (SYN542604)				
Structural space	Target	Alert 5	MACCS	Morgan
Number of category members	N/A	11	8	1
DNA alerts for AMES, CA and MNT	No alert	No alerts	No alerts	No alert
Protein binding alerts for CA	No alert	No alerts	No alerts	No alert
SlogP	0.77	-1.68 -> 1.06	-1.68 -> 1.83	1.06
MW	381.07	258.03 -> 393.04	253.04 -> 393.04	125.14
Ames	Negative (R/A from alert 6)	10 negatives 0 positives	5 negatives 3 positives	1 negative
In vitro CA	Negative (R/A from alert 6)	10 negatives 1 positive	No data	1 negative
In vivo CA	No prediction	No data	No data	No data
In vitro MNT	No prediction	No data	No data	No data
In vivo MNT	No prediction	No data	1 negative 1 positive	No data

Abbreviations: CA = chromosomal aberration; MNT = micronucleus test; R/A = read-across; N/A = not applicable.

3.4 Metabolic space and structural space alerts

The key advantage of the structural space alert approach is that structural space alerts are developed from an analysis of parent and metabolite structures for a given chemical class. This results in the alerts defining a set of common metabolic scaffolds that define metabolic space for which genotoxicity data exist. This ensures that chemicals grouped together using a specific structural space alert are likely to have a degree of metabolic similarity – an important factor in category formation [23]. The metabolic space associated with the sulphonyl ureas can be summarised as follows (denoted A, B, C and D in Figure 2):

A. Hydroxylation reactions of the parent sulphonyl urea (relevant to structural space alerts 1 and 2 in Table 1). These reactions typically occur on the six-membered aromatic rings present in this class of chemical. The benzene (or pyridine) ring, which is common to all sulphonyl ureas, has been shown to undergo a hydroxylation reaction; however, the exact position of this reaction is not typically determined experimentally. In contrast, the pyrimidine ring has been shown to undergo an equivalent reaction in the 4-postion. Replacement of this ring with a triazine prevents this reaction due to the presence of a nitrogen in the 4-postion (pyrimidine and triazines defined by alert 1). Finally, the majority of the sulphonyl urea chemical class feature methoxy groups in the 3- or 5- positions of the pyrimidine (or triazine) ring. The experimental data show that one of these groups undergoes a demethylation reaction producing a hydroxyl group.
B. Cleavage via hydrolysis of the sulphonyl urea bond (structural space alerts 2, 3 and 4). These reactions involve cleavage via hydrolysis of either one of the carbonyl-nitrogen bonds present in the sulphonyl urea moiety. The products of this reaction produce either a sulphonamide (defined by alert 2) and a carbamic acid or a primary aromatic amine (defined by alert 3) and a sulphonyl carbamic acid. The carbamic acids have been suggested to potentially undergo a methylation reaction; however, further functionalisation reactions of these species have not been observed experimentally (presumably due to the polar nature of the carbamic acid). In contrast, the benzene ring of the sulphonamide and the pyrimidine ring of the aromatic amine have been shown to undergo hydroxylation reactions (no reaction is possible when a triazine ring is present instead of a pyrimidine moiety). In the case of the pyrimidine ring, this reaction occurs in the 4-position, whereas the exact position of the hydroxyl group on the benzene ring of the sulphonamide is not typically defined in the available experimental data. In certain circumstances, the sulphonamide has been shown to undergo cyclisation reactions (defined by alert 4).
C. Cleavage of the six-membered heterocyclic ring (structural space alert 5). This reaction involves the cleavage of either the pyrimidine or triazine ring to produce a guanidine moiety. No further reactions have been noted for this species.
D. Re-arrangement of the sulphonyl urea bond (structural space alerts 6 and 7). This reaction involves the re-arrangement of the sulphonyl urea moiety into a urea, driven by the loss of sulphur dioxide (defined by alert 6). Experimental evidence has shown

that the urea species undergoes subsequent loss of the amide functional group resulting in a secondary amine (defined by alert 7). Finally, the secondary amine is either hydroxylated in the 4-position of the pyrimidine ring (no such reaction is possible when the pyrimidine is replaced by a triazine) or, when present, a methoxy group on the pyrimidine ring undergoes demethylation.

In addition to the metabolic reactions described above, several conjugated products have also been identified experimentally. These products typically involved the hydroxylated species outlined being conjugated with glutathione, sulphate or, to a lesser extent, glucuronic acid. These conjugated species are not shown in Figure 2 for clarity.

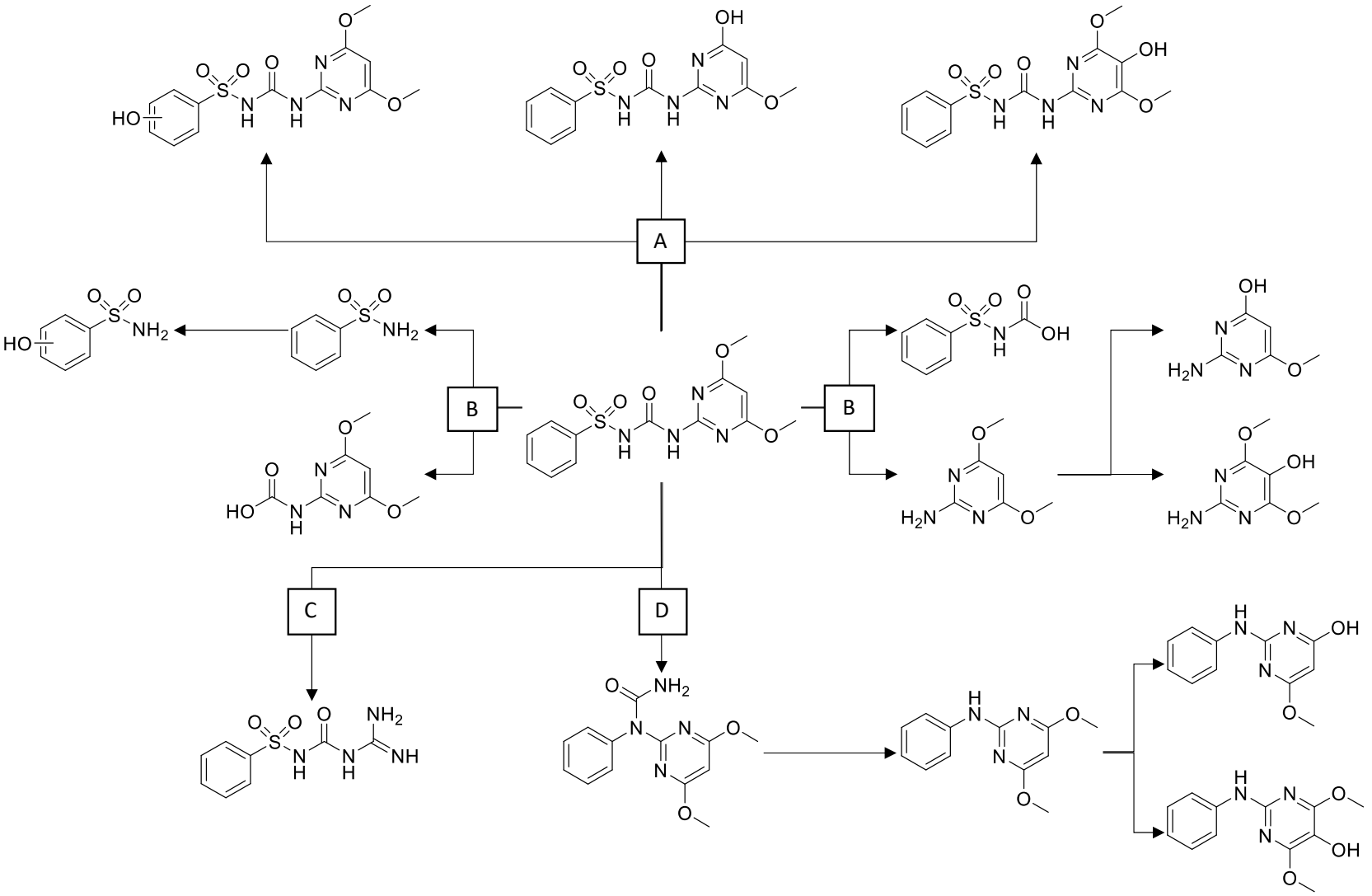
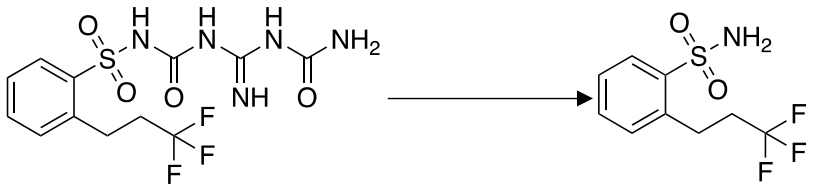


Figure 2: Overview of the common metabolic transformations for the sulphonyl urea chemical class

3.5 Structural space alerts versus fingerprint approaches

The analysis presented above demonstrates how the structural space alerts can be used to develop structural space categories within which read-across predictions can be made. However, it is also possible to utilise chemical fingerprint methods to identify similar chemicals to a target chemical. The key challenge with this approach being the need to identify a suitable fingerprint method (of which there are many) and to select an appropriate similarity coefficient by which to decide if a chemical is similar or not [24]. By way of comparison chemical categories were developed for the two metabolites using MACCS keys and Morgan fingerprint methods. Chemical analogues were defined as being within the fingerprint categories if they were within a Tanimoto distance of 0.7 of the Target chemical [16, 17]. Utilising the MACCS keys fingerprints resulted in a similar sized categories for both Target metabolites when compared to the categories defined by the respective structural space alerts (39 chemicals for both for metabolite SYN547308 and eight compared to 11 for metabolite SYN542604 - Tables 4 and 5 respectively). In contrast, the Morgan fingerprint method failed to identify any analogues within the 0.7 similarity cut-off for metabolite SYN547308, and only a single analogue for metabolite SYN542604. This highlights the variability that using different fingerprint methods has upon the resulting category membership.

In addition, the use of fingerprint can result in groups of highly similar chemicals that are not always suitable for read-across. For example, the single analogue with *in vivo* MNT data for Target SYN542604 (identified using MACCS key fingerprints) was identified as similar due to overlap in the skeletal parts of the molecules (Figure 3 - similarity based on a Tanimoto coefficient of 0.71). However, profiling of the Target and analogue structures with the structural space alerts showed them to have differing metabolic scaffolds (SYN542604: alert 5, sulphonamide A: alert 2 - Table 1). These different scaffolds are likely to undergo different metabolic transformations (as outlined in Figure 2), meaning that they are metabolically diss-similar. This highlights the key advantage of the structural space approach in that category membership is determined by the presence of well-defined chemical sub-structures derived from knowledge of the common scaffolds present within the metabolic space for a given chemical class.



SYN542604 Sulphonamide A

Figure 3: Analogue identified using MACCS keys fingerprint method for SYN542604

3.6 Proposed read-across workflow for genotoxicity

The analysis presented above demonstrates how read-across can be used for the prediction of genotoxicity for agrochemical residues. The approach requires the description of a set of structural space alerts that define a set of chemical categories for which experimental genotoxicity data exist. These structural space alerts are developed from an expert analysis of genotoxicity data and are agrochemical-class specific. The development of such alerts requires a data harvesting exercise to identify relevant chemical structures (for parent and metabolites) and associated genotoxicity test results. The availability of genotoxicity data, coupled with a set of structural space alerts, enables a read-across workflow to be outlined (Figure 4). This series of steps uses structural and physico-chemical parameters to ensure the target chemical can be confidently assigned to the chemical space category. However, the potential need to assess the ability of (representative) members of the category being able to reach the bone marrow is one of the key areas of uncertainty/difficulty in the proposed scheme. This becomes important when using existing negative *in vivo* micronucleus test data as the basis for a read-across prediction, as EFSA guidance states that bone marrow exposure must be demonstrated [1]. Such data on bone marrow exposure are typically available from the ADME studies in the DAR/RAR documents. However, these data are generated using the parent compound and thus do not enable metabolite exposure levels to be quantified (although in newer versions of the *in vivo* MNT quantification of bone marrow or, as a minimum, plasma levels are determined). An alternative solution, potentially capable of quantifying metabolite exposure levels, could be to use PBPK modelling to demonstrate exposure to the bone marrow [25-28]. It is important to note that this scheme does not imply the need for additional *in vivo* testing, it outlines how existing *in vitro* and/or *in vivo* data can be used to make read-across predictions. As stated previously, any additional testing to support a category would be either in the Ames or *in vitro* micronucleus tests.

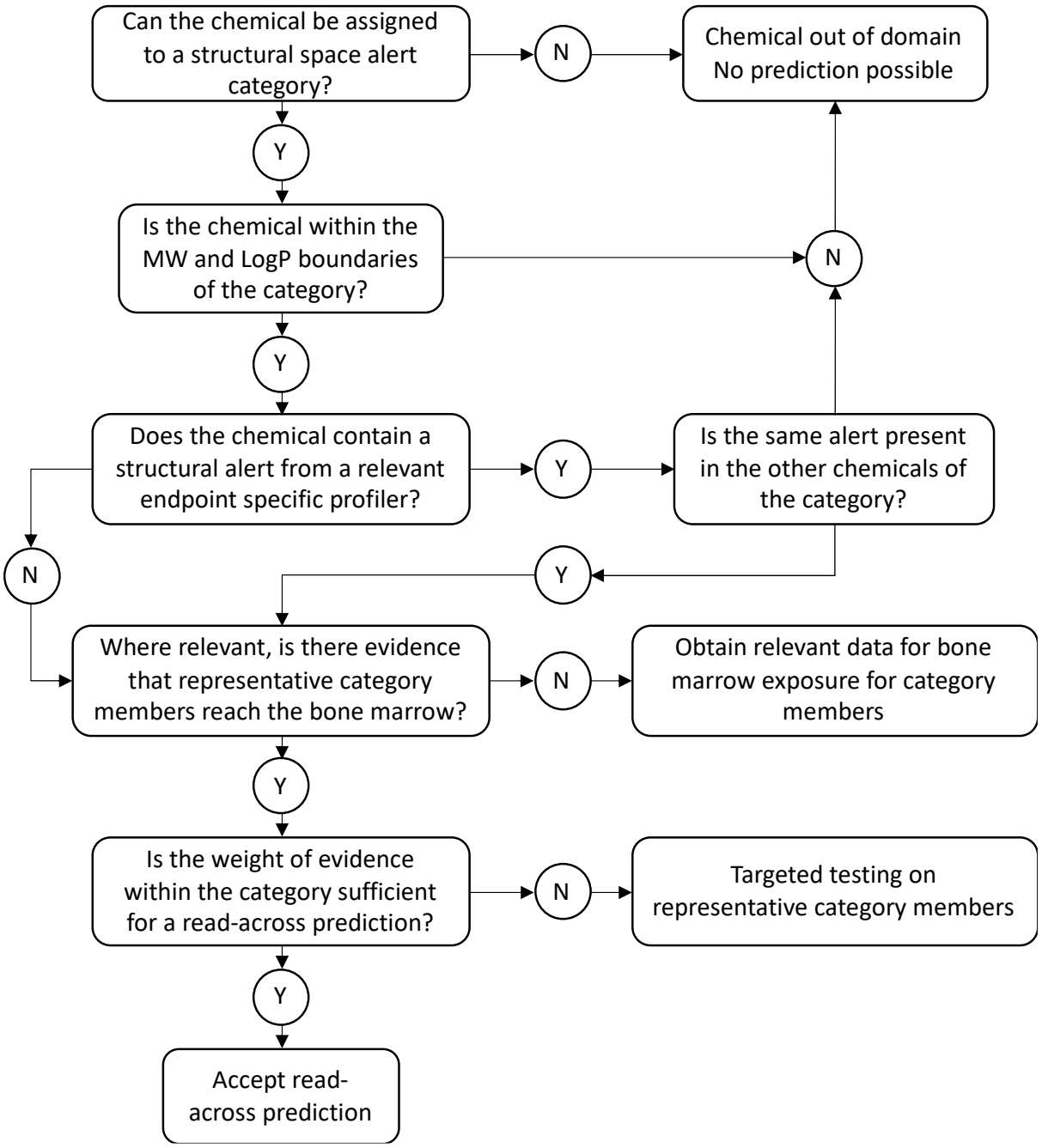


Figure 4: Proposed read-across workflow for genotoxicity prediction for agrochemical residues

4. Conclusions

The aim of this study was to develop a set of structural alerts to define the structural space associated with a set of sulphonyl urea agrochemicals. This analysis resulted in a set of seven structural space alerts developed from a dataset of 74 chemicals for which either Ames, chromosomal aberration of micronucleus test results were publicly available. The structural space alerts were developed based on a maximum common sub-structure linked to the common metabolic transformations for the sulphonyl urea chemical class. This linkage meaning the structural space alerts implicitly take metabolism into account. However, the

approach relies on the availability of metabolism information (typically from DAR/RAR documents), without such information structural space alerts cannot be developed. The study has also outlined how these structural space alerts could be used to identify further chemicals with similar chemistries from additional datasets. In addition, an analysis of the available profiling schemes for DNA and protein reactivity showed the importance of investigating the predictivity of such schemes within a well-defined area of structural space. Finally, an outline of how a combination of structural space alerts, covalent chemistry profiling and physical-chemistry properties could be used to develop chemical categories suitable for chemical prioritisation was presented. These results showed that genotoxicity for pesticide residues and/or metabolites could be predicted via read-across, including for the *in vivo* micronucleus test. The method presented represents a robust and repeatable approach to such read-across predictions, with the potential to reduce unnecessary testing.

Acknowledgements

This research received funding from CropLife Europe.

Declaration of Interest

No conflicts of interest were declared by the authors.

References

1. EFSA, *Guidance on the establishment of the residue definition for dietary risk assessment.* EFSA Journal, 2016. **14**(12).

2. EFSA, *Scientific opinion on genotoxicity testing strategies applicable to food and feed safety assessment.* EFSA Journal, 2011. **9**(9).

3. Enoch, S.J., *Chemical category formation and read-across for the prediction of toxicity*, in *Recent advances in QSAR studies: Methods and applications*, T. Puzyn, J. Leszczynski, and M.T.D. Cronin, Editors. 2010, Springer. p. 209-219.

4. Enoch, S.J. and D.W. Roberts, *Approaches for grouping chemicals into categories*, in *Chemical Toxicity Prediction: Category Formation and Read-Across* M.T.D. Cronin, et al., Editors. 2013, The Royal Society of Chemistry: Cambridge, UK. p. 30-43.

5. OECD, *Guidance on Grouping of Chemicals.* (available from: http://www.oecd.org/LongAbstract/0,3425,en_2649_34379_39404999_119829_1_1_37407,00.html), 2007.

6. Enoch, S.J. and M.T.D. Cronin, *A review of the electrophilic reaction chemistry involved in covalent DNA binding.* Critical Reviews in Toxicology, 2010. **40**: p. 728-748.

7. Benigni, R. and C. Bossa, *Structure alerts for carcinogenicity, and the Salmonella assay system: A novel insight through the chemical relational databases technology.* Mutation Research, 2008. **659**: p. 248-261.

8. Benigni, R., et al., *Development of structural alerts for the in vivo micronucleus assay in rodents.* http://ecb.jrc.ec.europa.eu/DOCUMENTS/QSAR/EUR_23844_EN.pdf, 2009.

9. Mekenyan, O., et al., *Identification of the structural requirements for mutagenicity by incorporating molecular flexibility and metabolic activation of chemicals I: TA100 Model.* Chemical Research in Toxicology, 2004. **17**(6): p. 753-766.

10. Mekenyan, O., et al., *Identifying the structural requirements for chromosomal aberration by incorporating molecular flexibility and metabolic activation of chemicals.* Chemical Research in Toxicology, 2007. **20**(12): p. 1927-1941.

11. Serafimova, R., et al., *Identification of the structural requirements for mutagencitiy, by incorporating molecular flexibility and metabolic activation of chemicals. II. General Ames mutagenicity model.* Chemical Research in Toxicology, 2007. **20**(4): p. 662-676.

12. Enoch, S.J. and M.T.D. Cronin, *Development of new structural alerts for chemical category formation for assigning covalent and non-covalent mechanisms relevant to DNA binding.* Mutation Research – Genetic Toxicology and Environmental Mutagenesis, 2012. **743**: p. 10-19.

13. Schultz, T.W., et al., *The OECD QSAR Toolbox Starts Its Second Decade*, in *Computational Toxicology. Methods in Molecular Biology*, N. O., Editor. 2018, Humana Press, New York, NY. p. 55-77.

14. Jaworska, J. and N. Nikolova-Jeliazkova, *How can structural similarity analysis help in category formation?* SAR and QSAR in Environmental Research, 2007. **18**(3-4): p. 195-207.

15. Patlewicz, G., et al., *Toxmatch - a new software tool to aid in the development and evaluation of chemically similar groups.* SAR and QSAR in Environmental Research, 2008. **19**(3): p. 397-412.

16. Enoch, S.J., et al., *Formation of structural categories to allow for read-across for teratogenicity.* QSAR & Combinatorial Science, 2009. **28**(6-7): p. 696-708.

17. Hewitt, M., et al., *Hepatotoxicity: A scheme for generating chemical categories for read-across, structural alerts and insights into mechansism(s) of action* Critical Reviews in Toxicology, 2013. **43**(7): p. 537-558.

18. Yang, C., et al., *Do Similar Structures Have Similar No Observed Adverse Effect Level (NOAEL) Values? Exploring Chemoinformatics Approaches for Estimating NOAEL Bounds and Uncertainties.* Chem Res Toxicol, 2021. **34**(2): p. 616-633.

19. Benigni, R., et al., *Evaluation of the applicability of existing (Q)SAR models for predicting the genotoxicity of pesticides and similarity analysis related with genotoxicity of pesticides for facilitating of grouping and read across.* EFSA Supporting Publications, 2019. **16**(3).

20. Metruccio, F., et al., *Compilation of a database, specific for the pesticide active substance and their metabolites, comprising the main genotoxicity endpoints.* EFSA Supporting Publications, 2017. **14**(5).

21. EC/EFSA, *Renewal Assessment Report: Tritosulfuron (Toxicology & Metabolism).* 2009.

22. Benigni, R., et al., *The new ISSMIC database on in vivo micronucleus and its role in assessing genotoxicity testing strategies.* Mutagenesis, 2012. **27**(1): p. 87-92.

23. Schultz, T.W. and M.T.D. Cronin, *Lessons learned from read-across case studies for repeated-dose toxicity.* Regul Toxicol Pharmacol, 2017. **88**: p. 185-191.

24. Mellor, C.L., et al., *Molecular fingerprint-derived similarity measures for toxicological read-across: Recommendations for optimal use.* Regul Toxicol Pharmacol, 2019. **101**: p. 121-134.

25. Yuan, D., et al., *Physiologically Based Pharmacokinetic Modeling of Nanoparticles.* J Pharm Sci, 2019. **108**(1): p. 58-72.

26. Knutsen, J.S., et al., *A calibrated human PBPK model for benzene inhalation with urinary bladder and bone marrow compartments.* Risk Anal, 2013. **33**(7): p. 1237-51.

27. Xia, B., et al., *Novel physiologically based pharmacokinetic modeling of patupilone for human pharmacokinetic predictions.* Cancer Chemother Pharmacol, 2012. **69**(6): p. 1567-82.

28. Gerard, C., et al., *Links between cyclosporin exposure in tissues and graft-versus-host disease in pediatric bone marrow transplantation: analysis by a PBPK model.* Pharm Res, 2011. **28**(3): p. 531-9.