Enoch, SJ, Hasarova, Z, Cronin, MTD, Bridgwood, K, Rao, S, Hueser, A, Kluxen, FM and Frericks, M

Matched molecular pairs-driven read-across for the prediction of genotoxicity of plant protection product residues

Regulatory Toxicology and Pharmacology 163 (2025) 105915

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

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

Matched molecular pairs-driven read-across for the prediction of genotoxicity of plant protection product residues

S.J. Enoch [a,*], Z. Hasarova [a], M.T.D. Cronin [a], K. Bridgwood [b], S. Rao [c], A. Hueser [d], F.M. Kluxen [e], M. Frericks [e]

[a] School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Liverpool, England, UK
[b] Syngenta, Bracknell, England, UK
[c] Gowan Company, Yuma, AZ, USA
[d] Bayer CropScience, Monheim am Rhein, Germany
[e] BASF SE, Limburgerhof, Germany

ARTICLE INFO

Handling Editor: Dr. Daniele Wikoff

Keywords:
Matched molecular pairs
Genotoxicity
Hazard assessment
Read-across
Plant protection product residues

ABSTRACT

In dietary risk assessment of plant protection products, residues of active ingredients and 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, a matched molecular pairs approach has been developed and applied to datasets of sulphonylurea herbicides, strobilurins fungicides and α-chloroacetamide herbicides for which either Ames, chromosomal aberration or micronucleus test results are publicly available. The approach is exemplified with four case studies illustrating how matched molecular pairs analysis can be used to identify analogues that cover the structural domain of the chemical for which a data-gap exists. The method is a robust and reproducible approach to such read-across predictions, with the potential to reduce unnecessary testing.

1. Introduction

In the framework of the new European Food Safety Authority (EFSA) requirements an assessment of the genotoxicity potential for active ingredient-associated residues of plant protection products (EFSA, 2016). The general approach outlined in the available (in the unadopted) EFSA guidance (EFSA, 2011) is that such residues should not increase the hazard to humans (and livestock). Thus, within a set of 'similar' residues, a category, a representative number need to have *in vitro* and/or *in vivo* data for gene mutation as well as structural and numerical chromosomal aberration. The availability of such data enables data-gaps within the category to be filled by read-across within this framework, with the minimal data requirements coming from the Ames test (gene mutation) and an *in vitro* micronucleus test (structural and numerical chromosomal aberration). The availability of additional negative *in vivo* data (frequently from the micronucleus test) adds further weight of evidence to the read-across prediction. If a genotoxicity prediction is negative, considering evidence from a quantitative structure-activity model and read-across, then no further experimental testing is required under the EFSA guidance. In contrast, a positive read-across prediction for genotoxicity requires further experimental data to be generated in a tiered approach. For example, if an initial *in vitro* micronucleus test confirms the positive read-across prediction for chromosome damage, an *in vivo* micronucleus test would be triggered.

The key step in the above process being the identification of analogues that are 'similar' to the target substance (for which genotoxicity data are lacking). This is relatively straightforward for substances that contain a structural alert for genotoxicity as the presence of the alert in both target and analogues define their ability to react via a common molecular initiating event (Enoch et al., 2011; Benigni et al., 2011; Mekenyan et al., 2004; Enoch et al., 2010; OECD, 2007; Sakuratani et al., 2008; Yamada et al., 2013). However, grouping substances, such as most plant protection metabolites or residues, that lack such alerts is more challenging. To this end, previous research by the current authors has outlined the Structural Space Alert (SPA) concept (Enoch et al., 2022a, 2022b, 2023, 2024). In this approach expert analysis of publicly available metabolism data from the Draft Assessment Reports (DAR) and/or Renewal Assessment Reports (RAR) (available from the EFSA website (www.efsa.europe.eu) are utilised to develop plant protection specific metabolic maps that summarise the common metabolism for the

* Corresponding author.
E-mail address: s.j.enoch@ljmu.ac.uk (S.J. Enoch).

https://doi.org/10.1016/j.yrtph.2025.105915
Received 19 May 2025; Received in revised form 21 July 2025; Accepted 27 July 2025

compound class. The information contained with each metabolic map is then developed into a set of structural alerts suitable for grouping metabolites and/or residues. The approach has the advantage of ensuring that the grouped substances are metabolically similar, an aspect that has been suggested to be an important consideration when making read-across predictions (Schultz et al., 2017; Yordanova et al., 2021). The key challenge with the SPA approach being the need for extensive expert judgement in the development of the metabolic maps and subsequent structural alert development.

Matched molecular pairs analysis (MMPA) has been outlined in the literature as an alternative way to identify analogues for read-across within the personal care sector (Hussain et al., 2010; Lester et al., 2021). This analysis involved the fact that MMPA identifies pairs of substances featuring common core with a so-called transformation point, at which varying substituents are attached. The MMPA approach has been most widely used in drug discovery as it enables the contribution to biological activity (or toxicity) of a series of substituents attached to a common core (or scaffold) to be understood in a systematic manner (for an overview of MMPA applied in drug discovery see reference (Yang et al., 2023)). Thus, in comparison to the expert driven SPA approach, MMPA offers an analysis that is automated and free from the need for expert judgement. Thus, the aim of this study was to outline the usage of MMPA for the identification of analogues for the prediction of the genotoxicity of plant protection products residues.

2. Methods

2.1. Datasets

Three previously published datasets were utilised. These consisted of plant protection product active ingredients and metabolite data previously harvested from the publicly available DAR/RAR documents. The datasets were as follows (*in vitro* assays with S9 fraction, Ames tests in the standard battery).

- 81 sulphonyl urea agrochemical active ingredients and metabolites, termed the 'sulphonyl urea genotoxicity dataset' contained the following test results (data taken from reference (Enoch et al., 2022a)):
  - Ames - 79 chemicals (all negative)
  - *in vitro* chromosomal aberration - 59 chemicals (45 negative, 14 positive)
  - *in vivo* chromosomal aberration - 6 chemicals (all negative)
  - *in vitro* micronucleus – 4 chemicals (3 negative, 1 positive)
  - *in vivo* micronucleus - 29 chemicals (27 negative, 2 positive)
- 46 strobilurin agrochemical active ingredients and metabolites, termed the 'strobilurin genotoxicity dataset' contained the following test results (data taken from reference (Enoch et al., 2023)):
  - Ames: 46 chemicals (all negative)
  - In vitro chromosomal aberration: 30 chemicals (23 negative, 7 positive)
  - In vivo chromosomal aberration: 1 chemical (negative)
  - In vitro micronucleus: 14 chemicals (12 negative, 2 positive)
  - I*n vivo* micronucleus: 24 chemicals (all negative)
- 68 α-chloroacetamide agrochemical active ingredients and metabolites, termed the 'α-chloroacetamide genotoxicity dataset' contained the following test results (data taken from reference (Enoch et al., 2024)):
  - Ames: 68 compounds (all negative)
  - In vitro chromosomal aberration: 55 compounds (38 negative, 16 positive, 1 equivocal)
  - In vivo micronucleus: 38
  - compounds (all negative)

Case study target chemicals were taken from the rat metabolism within the DAR/RAR documents of rimsulfuron, mandestrobin, azoxystrobin, and propisochlor. Metabolites were selected based on the absence of available data and are intended as illustrative examples. All chemical structures and associated genotoxicity data are available in the Supplementary Information.

2.2. Matched molecular pairs analysis

Matched molecular pairs analysis was carried out using a custom KNIME (V5.2.5) workflow, using the freely available MMPA RDKit nodes - 'MMP Molecule Multi-cut Fragment' and 'Fragments to MMPs'. The steps in this workflow are summarised in the following pseudo-code (KNIME workflow available from https://github.com/Enoch-chemoinformatics/MMPA-genotox.git).

- Step 1: The target and potential analogues chemicals are concatenated into a single table.
- Step2: The 'MMP Molecule Multi-cut Fragment' node fragments the table created in step 1 into 'Fragmentation Keys' and 'Fragmentation Values' – where the 'Fragmentation Keys' represent the 'Fixed' part of the molecule, whilst the 'Fragmentation Values' represent the part of the molecule attached to the 'Key' that is changing (where the 'Key' is usually larger than the 'Value'). This process results in a table of 'Fragmentation Keys' that are used to identify the structural overlap between the target chemical and the analogues.
- Step 3: The table created in step 2 is filtered so that only chemicals with a 'Fragmentation Key' in common with the target chemical are considered analogues. In effect, this step identifies analogues that have a chemical substructure in common with the target (the size of this common substructure being determined by the criteria set in the 'MMP Molecule Multi-cut fragment' node).
- Step 4: The target chemical and analogues identified in step 3 are concatenated into a single table containing the chemical ID and structural information (encoded as SMILES). This table represents the MMP category for the target chemical.
- Step 5: The final step in the workflow retrieves the toxicological data associated with the MMPA category developed in step 4, this table forms the output to the workflow.

The following settings were used in the RDKit MMP nodes within the KNIME workflow – these setting required a level of expert judgement in their usage (the rationale for which is as outlined).

- Fragmentation type: this parameter sets the algorithm used to fragment the dataset. In the current study this was set to 'all acyclic single bonds'.
  - Rationale: This algorithm is the most exhaustive in terms of bond fragmentation and so create the maximum number of 'Fragmentation Keys' (chemical substructures) with which to identify target-analogue pairs of molecules.
- Number of cuts: this value determines how many bonds can be broken ('cut') at a time. This was set to 'one' in the current study to ensure that a given molecule could only be split into two fragments linked via a single transformation point.
  - Rationale: This ensure a chemically interpretable fragmentation in which the effect of changing a single functional group (or moiety) can be investigated.
- Addition of hydrogen prior to fragmentation: this was set to 'add' as is recommended as the default settings within the RDKit node.
  - Rationale: This is the recommended setting in the node and so was left as default.
- Fragment filtering settings: all fragment filtering settings were left unchecked. The exception being the 'heavy atom ratio filter' which was set to 0.5. This value determines the ratio of the size of the changing fragment compared to the size of the common substructure (in terms of the number of heavy atoms) allowed when determining a matched molecular pair.

- o Rationale: This setting required the most expert judgement with the 'heavy atom ratio filter' having the most effect on the size of the 'Fragmentation Key' (i.e., the size of the substructure common to two molecules needed for them to be a 'pair'). In practice setting the value two low results in increasingly small fragments (for example, pairs of molecules being matched by a common methyl group). In contrast, setting it too high has the opposite effect to the point where the fragment requirements become too big for any pairs to be identified. This setting is likely to require optimisation for different datasets/endpoints depending on the amount of data available.

2.2.1. *In-silico profiling for DNA reactivity*

Chemicals were profiled using the profiling schemes within the OECD QSAR Toolbox (V4.7.1). A subset of the available profilers was utilised based on the results of a previous study into their suitability for read-across predictions within the plant protection chemical space (Enoch et al., 2022a, 2022b, 2023, 2024). These profilers were (CA is chromosomal aberration and MNT is the micronucleus test).

- DNA alerts for AMES, CA and MNT by OASIS
- Protein binding alerts for CA by OASIS

2.3. *Fingerprint similarity*

Chemical similarity analysis was carried out using AtomPair fingerprints, coupled to the Tanimoto similarity metric, was performed using the RDKit Fingerprint node as implemented in KNIME V5.2.5.

3. Results and discussion

The aim of this study was to outline the use of MMPA for the identification of analogues for the prediction of genotoxicity of plant protection product residues. This was achieved through the development of a KNIME workflow enabling read-across case studies to be developed for three plant protection product classes: sulphonylureas, strobilurins, and α-chloroacetamides. The presented method is applicable to any class of plant protection product, without the need for the *a-prior* development of class-specific structural alerts. In addition, the method offers a robust and repeatable identification of analogues for read-across, an important aspect for regulatory acceptance.

3.1. *Case study 1: sulphonylureas*

Twenty-two analogues were identified by MMPA for target 1, nine of these are as shown in Table 1, ranked by their Tanimoto similarity calculated using AtomPair fingerprints (the remaining 13 structures are parent sulphonylureas, with structure analogous to SU-125, and are not shown for clarity). The MMPA identified six substructures with a single transformation point (Table 1). The substructures show the common overlap between the target and each Analogue identified by the MMPA (with the transformation point, denoted by '*', indicating where this overlap ends). In most cases there was a single substructure linking the target to each Analogue, the exception being for Analogue 3, for which two substructures were identified. The analogues identified via MMPA enable the following structure-toxicity relationship to be outlined for the group.

- Analogues 1, 2, 5 and 6 feature a nitrogen-linked pyridine-pyrimidine ring system, with analogue 5 containing a fused ring system containing this system (substructures 1, 2, and 5). All analogues showed that the presence of this ring system not to be associated with genotoxicity. In addition, analogue 4 features a closely related pyrazole-pyrimidine ring system which did not display any genotoxicity in the Ames assay. Importantly, analogue 2 also has a negative *in vivo* micronucleus test result.
- Analogues 1, 3 and 8 feature the di-substituted pyridine ring system in which an ethyl sulfonyl group is also present (substructures 1, 3, and 6). All three analogues have negative Ames and *in vitro* chromosomal aberration test results. In addition, analogue 3 also has a negative *in vivo* micronucleus test result.
- Analogues 1–7 and 9 feature the 1,3-dimethoxy moiety that is also present in the target (substructures 4 and 5). In addition, a further 12 parent compounds also feature this moiety (not shown in Table 1). Except for analogue 2, all the available data for was negative for substances with this structural feature (analogue 2 has a positive *in vitro* chromosomal aberration test).
- None of the chemicals in the category triggered an alert for DNA or protein binding using the genotoxicity endpoint specific OASIS profilers in the OECD QSAR Toolbox V4.7.

These different transformation points allow the analogues to cover the structural space of the target chemical, whilst remaining related to the target in a transparent (and repeatable) manner. Interestingly, only analogue 1 would be identified as being similar using the commonly cited Tanimoto cut-off value of 0.7 (Enoch et al., 2009, 2022a), highlighting the difficulty in using such methods for the identification of analogues for read-across (similar trends in the Tanimoto similarity measures were seen with the other case studies in this study – Tables 2–4).

Overall, the weight of evidence outlined above enables the Ames, *in vitro* chromosomal aberration and *in vivo* micronucleus test results to be predicted as negative via read-across for the target chemical.

3.2. *Case study 2: strobilurins*

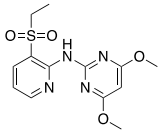
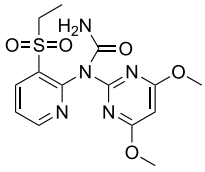
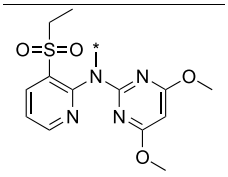
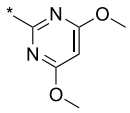
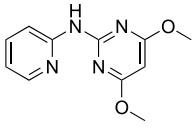
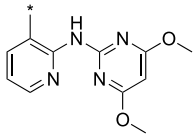
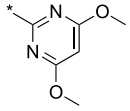
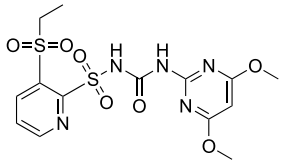
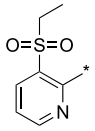
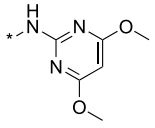
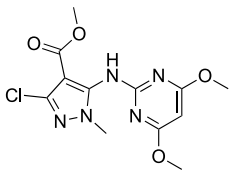
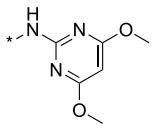
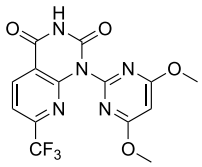
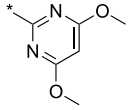
Nine analogues were identified by MMPA for target 2, a metabolite present in the rat, for which no data exist. These analogues are as shown in Table 2 ranked by their Tanimoto similarity calculated using AtomPair fingerprints. The MMPA analysis identified three substructures with transformation points (marked by '*' in Table 2). The presence of these substructures enabled the following structure-toxicity relationship to be outlined.

- Analogues 1, 4–8 feature the 2-methoxy-*N*-methyl-2-phenylacetamide moiety (substructure 1) for which seven Ames, three *in vitro* chromosomal aberration and two *in vivo* micronucleus test results are available. All of these are negative, except for a single positive *in vitro* chromosomal aberration result for analogue 4.
- Analogues 2–5 and 9 all feature a benzene ring with a carboxylic acid in the ortho position (substructure 2) for which five Ames, four *in vitro* chromosomal aberration, one *in vitro* micronucleus test and four *in vivo* micronucleus test results. These are all negative, expect for two *in vitro* chromosomal aberration and one *in vitro* micronucleus test results.
- None of the chemicals in the category triggered an alert for DNA or protein binding using the genotoxicity endpoint specific OASIS profilers in the OECD QSAR Toolbox.

Overall, the weight of evidence outlined above enables the Ames, *in vitro* chromosomal aberration and *in vivo* micronucleus test results to be predicted as negative via read-across.

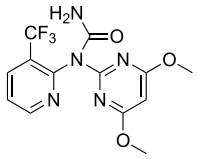
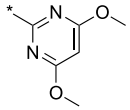
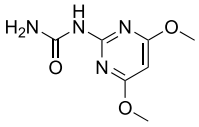
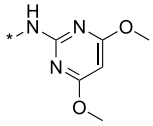
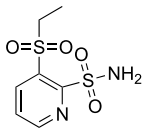
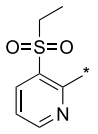
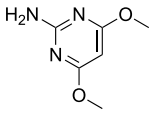
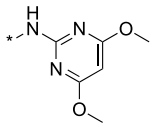
In contrast, MMPA analysis for target 3, highlighted a situation in which a read-across prediction could not be made due to a lack of structural coverage between the analogues and the target. In this example, two analogues were identified, both feature the 4,6-diphenoxy-pyrimidine substructure with an ortho substituent as the transformation point (substructure 1 in Table 3). However, neither analogue features the presence of the hydroxy propanoate moiety at the transformation point, making the assessment of its genotoxicity potential

Table 1
Analogues identified via MMPA for the target 1 (rat metabolite from rimsulfuron). Substance identifiers refer to the numbering in the sulphonylurea dataset available as part of the supplementary information. Abbreviations: MMPA = Matched Molecular Pairs Analysis; CA = Chromosomal Aberration; MNT = Micro Nucleus Test; R/A = Read Across; * = transformation point; N/A = Not Applicable.

ID	Structure	MMPA Common Substructure	Similarity (AtomPair/Tanimoto)	Genotoxicity data
Target 1 (IN-70942)		N/A	N/A	Ames: ve (R/A) *In vitro* CA: ve (R/A) *In vivo* MNT: ve (R/A)
Analogue 1 (SU-127)			0.708	Ames: ve
		Substructure 1 		*In vitro* CA: ve
		Substructure 5		
Analogue 2 (SU-85)			0.552	Ames: ve
		Substructure 2 		*In vitro* CA: +ve
		Substructure 5		*In vivo* MNT: ve
Analogue 3 (SU-125)			0.471	Ames: ve
		Substructure 3 		*In vitro* CA: ve
		Substructure 4		*In vivo* MNT: ve
Analogue 4 (SU-100)			0.457	Ames: ve
		Substructure 4		
Analogue 5 (SU-95)			0.454	Ames: ve *In vitro* CA: ve
		Substructure 5		

(*continued on next page*)

Table 1 (*continued*)

ID	Structure	MMPA Common Substructure	Similarity (AtomPair/Tanimoto)	Genotoxicity data
Analogue 6 (SU-84)			0.449	Ames: ve
		Substructure 5		*In vitro* CA: ve
Analogue 7 (SU-83)			0.357	Ames: ve
		Substructure 4		*In vitro* CA: ve *In vivo* MNT: ve
Analogue 8 (SU-126)			0.290	Ames: ve
		Substructure 6		*In vitro* CA: ve
Analogue 9 (SU-68)			0.226	Ames: ve
		Substructure 4		

impossible. Thus, within the suggested MMPA approach to read-across the target chemical is considered out of domain of the available analogues.

3.3. Case study 3: α-chloroacetamides

Thirteen analogues were identified by MMPA for target 4, a metabolite present in the rat, for which no data exist. These analogues are as shown in Table 4, ranked by their Tanimoto similarity calculated using AtomPair fingerprints. The MMPA analysis identified three substructures with transformation points (marked by '*' in Table 4). The presence of these substructures enabled the following structure-toxicity relationship to be outlined.

- Analogues 1, 3, 4 and 7 all feature an ether linked iso-propyl chain (substructure 1) connected to an amide for which four Ames, three *in vitro* chromosomal aberration and two *in vivo* micronucleus test results. These are all negative, expect for one *in vitro* chromosomal aberration for analogue 1.
- Analogues 2, 5 and 10 all feature a di-carbonyl carboxylic acid moiety (substructure 3) for which three Ames, *in vitro* chromosomal aberration and two *in vivo* micronucleus test results. All of which are negative.
- All 13 analogues feature the di-ortho substituted benzene moiety (substructure 2). This substructure adds five additional analogues (analogues 6, 8, 9, 11–13) that adds weight of evidence that this feature is not associated with genotoxicity.
- Analogues 3 and 8 triggered alerts for DNA and protein binding due to the presence of the α-chloroacetamide moiety using the genotoxicity endpoint specific OASIS profilers in the OECD QSAR Toolbox. The remaining analogues did not trigger any alerts.

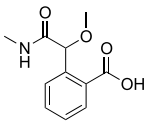
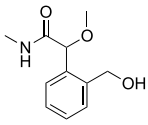
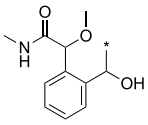
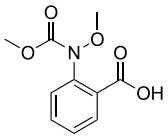
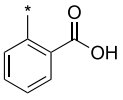
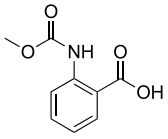
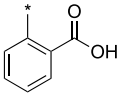
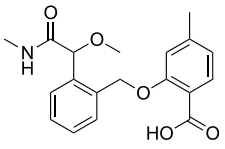
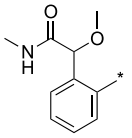
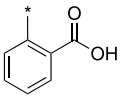
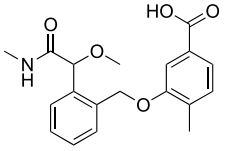
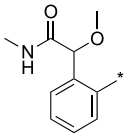
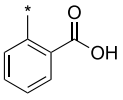
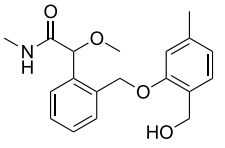
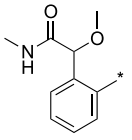
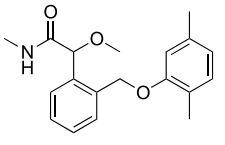
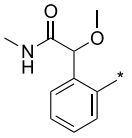
Overall, despite removing analogues 3 and 8 due to their protein/DNA reactivity, the weight of evidence outlined above enables the Ames, *in vitro* chromosomal aberration and *in vivo* micronucleus test results to be predicted as negative via read-across.

3.4. Matched molecular pairs versus structural space alerts

Matched molecular pairs analysis utilises structural similarity as the basis for analogue identification. The current study extends the previous work in this area (Hussain et al., 2010) to the plant protection product space showing how it can be used to identify analogues from datasets of metabolites and residues. Importantly, the outlined approach does not require any expert judgement or the *α-priori* development of compound class specific structural alerts or profiling schemes. Thus, it can be applied to any plant protection product residue, with the only limitation being the availability of suitable analogues within a given dataset. In addition, the ease with which the applicability domain of the target chemical to potential analogues can be defined using the MMPA approach has also been shown – this stemming from the definition of the substructures and their transformation points. The MMPA approach also offers a significantly robust and repeatable approach to analogue identification when compared to fingerprint similarity, as there is no need for an evaluation of the Tanimoto cut-off values or different fingerprint methods (Enoch et al., 2009; Yang et al., 2021; Hewitt et al., 2013).

In contrast, the previously published Structural Space Alert (SPA) concept uses metabolism as the key measure of chemical similarity (Enoch et al., 2022a, 2022b, 2023, 2024). The SPA concept requires extensive expert analysis of multiple metabolic maps to define the structural space alerts – with this analysis being required for each class of plant protection products for which read-across predictions are required. Clearly, this is a time-consuming process that potentially limits the applicability of the SPA concept to residues for which SPAs have already been defined. However, the use of metabolism as the basis of chemical similarity is in keeping with previous studies which have outlined the importance of considering such similarity when defining the applicability domain of a read-across prediction (Schultz et al., 2017; Yordanova et al., 2021). This being the key advantage of the SPA approach for the prediction of genotoxicity of plant protection product

Table 2
Analogues identified via MMPA for target 2 (rat metabolite from mandestrobin). Substance identifiers refer to the numbering in the strobilurin dataset available as part of the supplementary information. Abbreviations: MMPA = Matched Molecular Pairs Analysis; CA = Chromosomal Aberration; MNT = Micro Nucleus Test; R/A = Read Across; * = transformation point.

ID	Structure	MMPA Common Substructure	Similarity (AtomPair/Tanimoto)	Genotoxicity data
Target 2 (DX-CA-S-2200)		N/A	N/A	Ames: ve (R/A) *In vitro* CA: ve (R/A) *In vivo* MNT: ve (R/A)
Analogue 1 (STB-68)		 Substructure 1	0.706	Ames: ve
Analogue 2 (STB-109)		 Substructure 2	0.636	Ames: ve *In vitro* CA: +ve *In vitro* MNT: +ve *In vivo* MNT: ve
Analogue 3 (STB-108)		 Substructure 2	0.554	Ames: ve *In vitro* CA: ve *In vivo* MNT: ve
Analogue 4 (STB-64)		 Substructure 1  Substructure 2	0.437	Ames: ve *In vitro* CA: +ve *In vivo* MNT: ve
Analogue 5 (STB-65)		 Substructure 1  Substructure 2	0.412	Ames: ve *In vitro* CA: ve
Analogue 6 (STB-66)		 Substructure 1	0.411	Ames: ve
Analogue 7 (STB-10)		 Substructure 1	0.393	Ames: ve *In vitro* CA: ve

(*continued on next page*)

Table 2 (*continued*)

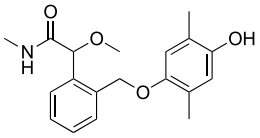
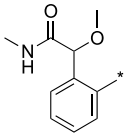
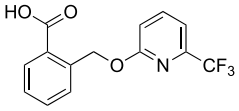
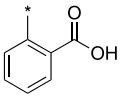
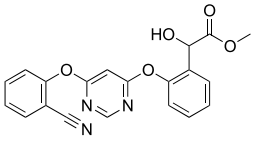
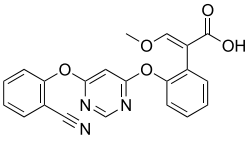
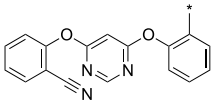
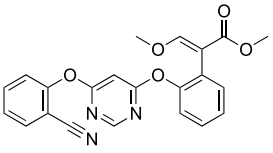
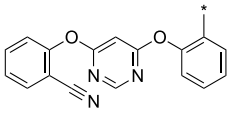
ID	Structure	MMPA Common Substructure	Similarity (AtomPair/Tanimoto)	Genotoxicity data
				In vivo MNT: ve
Analogue 8 (STB-67)			0.389	Ames: ve
		Substructure 1		
Analogue 9 (STB-25)			0.318	Ames: ve
		Substructure 2		*In vivo* MNT: ve

Table 3
Analogues identified via MMPA for target 3 (plant metabolite from azoxystrobin). Substance identifiers refer to the numbering in the strobilurin dataset available as part of the supplementary information. Abbreviations: MMPA = Matched Molecular Pairs Analysis; CA = Chromosomal Aberration; MNT = Micro Nucleus Test; R/A = Read Across; * = transformation point.

ID	Structure	MMPA Common Substructure	Similarity (AtomPair/Tanimoto)	Genotoxicity data
Target 3 (Compound 24)		N/A	N/A	R/A = out of domain
Analogue 1 (STB-6)			0.798	Ames: ve
		Substructure 1		*In vitro* CA: +ve *In vivo* MNT: ve
Analogue 2 (STB-36)			0.749	Ames: ve
		Substructure 1		

residues via read-across.

As outlined above, the key advantage that the SPA concept has over MMPA is that the substructures used as the basis for analogue identification are explicitly linked to metabolism. However, it is possible to build a weight of evidence that metabolism will not create any reactive metabolites within a category developed from MMPA. Consider the data in Table 4, each of the substructure identified in the target is present in at least one other analogue for which either a negative *in vitro* or *in vivo* genotoxicity test results are available. This suggests that none of the substructures has undergone metabolism that leads to a reactive species, as if they had, then (the majority of) these assays would be positive (as they are all metabolically capable).

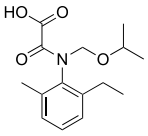
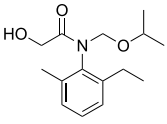
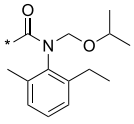
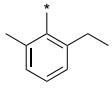
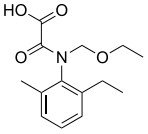
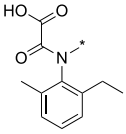
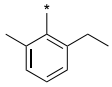
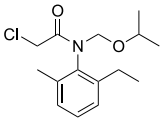
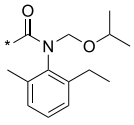
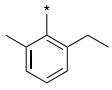
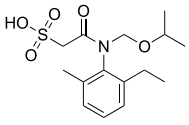
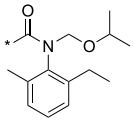
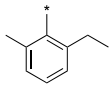
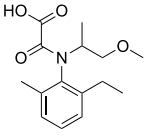
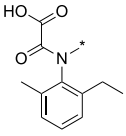
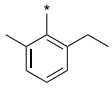
4. Conclusions

The aim of this study was to outline the use of matched molecular pairs analysis to identify analogues suitable for the prediction of the genotoxicity of plant protection product residues via read-across. The presented case studies have shown that matched molecular pairs analysis offers a systematic way of exploring the structure-toxicity relationships between a target and a set of analogues. The ability of matched molecular pairs analysis to locate analogues that cover different functional groups (and combinations of functional groups) is a particular advantage as it enables multiple lines of evidence to be drawn about the likely genotoxicity of each moiety. In addition, the method enables a transparent way in which to define the applicability domain of a read-across prediction. Whilst the current study has focused on genotoxicity the method is applicable to any hazard endpoint for which sufficient toxicological data exist, including those driven by receptor effects, such as endocrine disruption and neurotoxicity. Finally, the current study adds to the previous work in demonstrating the matched molecular pairs analysis offers a repeatable and transparent method that does not rely on expert judgment for grouping substances for read-across.

CRediT authorship contribution statement

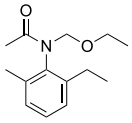
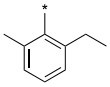
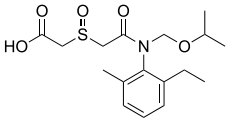
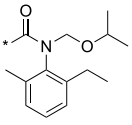
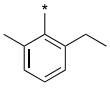
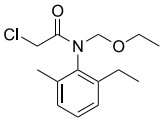
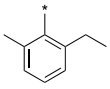
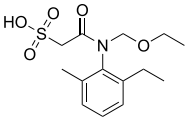
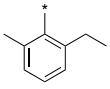
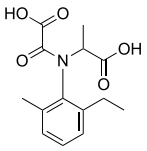
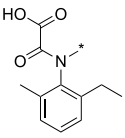
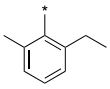
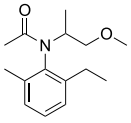
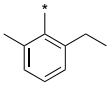
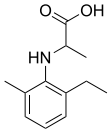
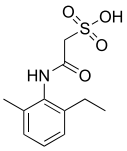
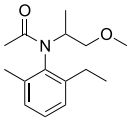
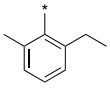
S.J. Enoch: Writing – review & editing, Writing – original draft, Funding acquisition, Formal analysis, Conceptualization. **Z. Hasarova:** Formal analysis, Data curation, Conceptualization. **M.T.D. Cronin:** Writing – review & editing. **K. Bridgwood:** Writing – review & editing. **S. Rao:** Writing – review & editing. **A. Hueser:** Writing – review & editing. **F.M. Kluxen:** Writing – review & editing. **M. Frericks:** Writing – review & editing, Conceptualization.

Table 4
Analogues identified via MMPA for target 4 (rat metabolite from propisochlor). Substance identifiers refer to the numbering in the α-chloroacetamide dataset available as part of the supplementary information. Abbreviations: MMPA = Matched Molecular Pairs Analysis; CA = Chromosomal Aberration; MNT = Micro Nucleus Test; R/A = Read Across; * = transformation point.

ID	Structure	MMPA Common Substructure	Similarity (AtomPair/Tanimoto)	Genotoxicity data
Target 4 (HRAC-15-65)		N/A	N/A	Ames: ve (R/A) *In vitro* CA: ve (R/A) *In vivo* MNT: ve (R/A)
Analogue 1 (HRAC-15-67)			0.784	Ames: ve
		Substructure 1 		*In vitro* CA: +ve
		Substructure 2		*In vivo* MNT: ve
Analogue 2 (HRAC-15-29)			0.781	Ames: ve
		Substructure 3 		*In vitro* CA: ve
		Substructure 2		*In vivo* MNT: ve
Analogue 3 (HRAC-15-63)			0.661	Ames: ve
		Substructure 1 		*In vitro* CA: ve
		Substructure 2		*In vivo* MNT: ve
Analogue 4 (HRAC-15-64)			0.621	Ames: ve
		Substructure 1 		
		Substructure 2		
Analogue 5 (HRAC-15-43)			0.600	Ames: ve
		Substructure 3 		*In vitro* CA: ve
		Substructure 2		*In vivo* MNT: ve

(*continued on next page*)

Table 4 (*continued*)

ID	Structure	MMPA Common Substructure	Similarity (AtomPair/Tanimoto)	Genotoxicity data
Analogue 6 (HRAC-15-31)			0.590	Ames: ve
		Substructure 2		*In vitro* CA: ve *In vivo* MNT: ve
Analogue 7 (HRAC-15-68)			0.568	Ames: ve
		Substructure 1		*In vitro* CA: ve
				
		Substructure 2		
Analogue 8 (HRAC-15-28)			0.548	Ames: ve
		Substructure 2		*In vitro* CA: ve *In vivo* MNT: ve
Analogue 9 (HRAC-15-32)			0.535	Ames: ve
		Substructure 2		*In vitro* CA: ve *In vivo* MNT: ve
Analogue 10 (HRAC-15-44)			0.517	Ames: ve
		Substructure 3		*In vitro* CA: ve
				
		Substructure 2		
Analogue 11 (HRAC-15-54)			0.462	Ames: ve
		Substructure 2		*In vitro* CA: +ve *In vivo* MNT: ve
Analogue 12 (HRAC-15-53)			0.383	Ames: ve
		Substructure 2		*In vitro* CA: +ve *In vivo* MNT: ve
Analogue 13			0.375	Ames: ve
HRAC-15-33/49/66		Substructure 2		*In vitro* CA: ve

Funding body

Crop Life Europe are gratefully acknowledged for funding this research.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Zuzana Hasarova reports financial support was provided by Crop Life

Europe.

Data availability

Data available as SI.

References

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

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

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

Enoch, S.J., et al., 2009. Formation of structural categories to allow for read-across for teratogenicity. QSAR Comb. Sci. 28 (6–7), 696–708.

Enoch, S.J., Cronin, M.T.D., 2010. A review of the electrophilic reaction chemistry involved in covalent DNA binding. Crit. Rev. Toxicol. 40, 728–748.

Enoch, S.J., Cronin, M.T.D., Ellison, C.M., 2011. The use of a chemistry based profiler for covalent DNA binding in the development of chemical categories for read-across for genotoxicity. ATLA, Altern. Lab. Anim. 39, 131–145.

Enoch, S.J., et al., 2022a. Sub-structure-based category formation for the prioritisation of genotoxicity hazard assessment for pesticide residues: sulphonyl ureas. Regul. Toxicol. Pharmacol. 129, 105115.

Enoch, S.J., et al., 2022b. Sub-structure-based category formation for the prioritisation of genotoxicity hazard assessment for pesticide residues (part 2): triazoles. Regul. Toxicol. Pharmacol. 134, 105237.

Enoch, S.J., et al., 2023. Metabolism-based category formation for the prioritisation of genotoxicity hazard assessment for plant protection product residues (part 3): strobilurins. Regul. Toxicol. Pharmacol. 144, 105484.

Enoch, S.J., et al., 2024. *Metabolism-based category formation for the prioritisation of genotoxicity hazard assessment for plant protection product residues (part 4): α-chloroacetamides*. Regul. Toxicol. Pharmacol. 150, 105641.

Hewitt, M., et al., 2013. Hepatotoxicity: a scheme for generating chemical categories for read-across, structural alerts and insights into mechansism(s) of action. Crit. Rev. Toxicol. 43 (7), 537–558.

Hussain, J., Rea, C., 2010. Computationally efficient algorithm to identify matched molecular pairs (MMPs) in large data sets. Journal Chemical Informatics and Modelling 50, 339–348.

Lester, C.C., Yan, G., 2021. A matched molecular pair (MMP) approach for selecting analogs suitable for structure activity relationship (SAR)-Based read across. Regul. Toxicol. Pharmacol. 124, 104966.

Mekenyan, O., et al., 2004. Identification of the structural requirements for mutagenicity by incorporating molecular flexibility and metabolic activation of chemicals I: TA100 model. Chem. Res. Toxicol. 17 (6), 753–766.

OECD, 2007. Guidance on grouping of chemicals. http://www.oecd.org/LongAbstract/0,3425,en_2649_34379_39404999_119829_1_1_37407,00.html.

Sakuratani, Y., et al., 2008. Category analysis of the substituted anilines studied in a 28-day repeat-dose toxicity test conducted on rats: correlation between toxicity and chemical structure. SAR QSAR Environ. Res. 19 (7–8), 681–696.

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

Yamada, T., et al., 2013. A category approach to predicting the repeated-dose hepatotoxicity of allyl esters. Regul. Toxicol. Pharmacol. 65, 189–195.

Yang, C., et al., 2021. Do similar structures have similar no observed adverse effect level (NOAEL) values? Exploring chemoinformatics approaches for estimating NOAEL bounds and uncertainties. Chem. Res. Toxicol. 34 (2), 616–633.

Yang, Z., et al., 2023. Matched molecular pairs analysis in drug discovery: methods and recent applications. J. Med. Chem. 66, 4361–4377.

Yordanova, D.G., et al., 2021. Assessing metabolic similarity for read-across predictions. Computational Toxicology 18.