**A Strategy to Define Applicability Domains for Read-Across**

Cynthia Pestana1,Steven J. Enoch,2 James W. Firman,2 Judith C. Madden,2 Nicoleta Spînu,2 Mark T.D. Cronin2\*

1Department of Genetics - Federal University of Paraná (UFPR), Curitiba, PR, Brazil

2School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK

\*Author for Correspondence:

Mark Cronin

Email: m.t.cronin@ljmu.ac.uk

ORCID Identifiers:

Cynthia Pestana: 0000-0002-0468-8303

Steven J. Enoch: 0000-0001-9111-5783

James W. Firman: 0000-0003-0319-1407

Judith C. Madden: 0000-0001-6142-5860

Nicoleta Spînu: 0000-0002-9465-3090

Mark T.D. Cronin: 0000-0002-6207-4158

**Abstract**

The definition, characterisation and assessment of the similarity between target and source molecules are cornerstones of the acceptance of a read-across prediction to fill a data gap for a toxicological endpoint. There is much guidance and many frameworks which are applicable in a regulatory context, but as yet no formalised process exists by which to determine whether or not the properties of an analogue (or chemicals within a category) fall within an appropriate domain from which a reliable read-across prediction can be made. This investigation has synthesised much of the existing knowledge in this area into a practical strategy to enable the domain of a read-across prediction to be defined, in terms of chemistry (structure and properties), toxicodynamics and toxicokinetics. The strategy is robust, comprehensive, flexible, and can be implemented readily. It enables the relative similarity and dissimilarity, between target and source molecules, for both the analogue and category approaches, to be analysed and provides a basis for alternative scenarios such as read-across based on formation of a common metabolite or biological profile to be defiend. Herein, the read-across domains for the repeated dose toxicity of a group of triazoles and imidazoles have been evaluated. The most challenging aspect to this approach will continue to be determining what is an “acceptable” degree of similarity when performing read-across for a specific purpose.

**Keywords:** Read-Across; Toxicity prediction; Domain; Analogue; Category

**Graphical Abstract**

****

**Highlights**

* A strategy to define the applicability domain of analogues and categories for read-across
* Domains characterised in terms of chemistry, toxicodynamics and toxicokinetics
* Assessment of domains enables many lines of evidence to be incorporated
* Domain assessment is a simple-to-use, yet powerful, approach

**Abbreviations**: ADME, absorption, distribution, metabolism excretion; BPt, boiling point; CAS, Chemical Abstracts Service; Cmax, maximum plasma concentration; Css, steady-state plasma concentration; ECHA, European Chemicals Agency; GSH, Glutathione; ICE, Integrated Chemical Environment; log P, logarithm of the octanol-water partition coefficient; MIE, Molecular Initiating Event; MW, molecular weight; MPt, melting point, NAM, New Approach Methodology; NEIHS, National Institute of Environmental Health Sciences; NOAEL, No Observed Adverse Effect Level; OECD, Organisation for Economic Cooperation and Development; PBPK; Physiologically-based pharmacokinetics; PC, Principal Component; QSAR, quantitative structure-activity relationship; RAAF, Read-Across Assessment Framework; US EPA, United States Environmental Protection Agency; VP, vapour pressure; UVCBs, unknown or variable composition, complex reaction products

1. **Introduction**

Read-across is the process of inferring the toxicity of a compound with few, or no, data from a similar compound with acceptable data [1]. Whilst possibly simple to conceive at the outset, achieving an adequate read-across prediction can be a lengthy process. This is especially true when attempting to justify read-across for a complex endpoint such that it may be an acceptable replacement for an *in vivo* assay for regulatory purposes. The key questions often relate to justifying the assessment of similarity between the molecules and demonstrating there is no significant difference in structure (or none that is predictable) that may adversely influence the prediction of toxicity [2]. To date, much guidance has been published to assist the development of a read-across prediction, such as the Read-Across Assessment Framework (RAAF) from the European Chemicals Agency [3], weight-of-evidence for assessing analogue suitability [4] and approaches to address the issue of uncertainty in predictions [5]. However, a definitive approach to determining the adequacy of an analogue and the chemical, toxicodynamic and toxicokinetic domain to which it would be applicable remains elusive [6].

Hence, an approach that may assist in rationalising analogue selection, the determination of the applicability domain for an analogue (or category) and, moreover, the definition of a strategy through which this may be achieved would be beneficial. The concept of applicability domain is well developed and recognised for quantitative structure-activity relationships (QSARs) [7] and is one of the core Organisation for Economic Cooperation and Development (OECD) Principles for the Validation of (Q)SARs [8]. There are many means of defining the applicability domain of a QSAR, but one of the most comprehensive was that proposed by Dimitrov et al. [9], who defined it as a four-stage process:

* Stage 1. General Parametric Requirements. i.e. properties
* Stage 2. Structural Domain. i.e. fragments / similarity
* Stage 3. Mechanistic Domain
* Stage 4. Metabolic Domain. i.e. similarity in (simulated) metabolites

Whilst not directly applicable to the situation for read-across, the elements of the applicability domain concept for QSARs can be translated into the chemical, toxicodynamic and toxicokinetic properties required to identify analogues for read-across [4, 10]. Development of this concept can be further informed by the requirements of ECHA’s RAAF and the assessment of uncertainties. Many researchers are routinely collecting information to support read-across e.g. structural and property data. Templates are available for doing this [11, 12], as well as for organising data for QSARs - the so-called QSAR Model Reporting Format [13]. However, a systematic approach to organise the collection and analysis of data, that is sufficiently flexible and adaptable for all endpoints and enables definition of the domain is still lacking.

In order to address these needs, the aim of this study was to determine what would be required to enable the applicability domain for a read-across prediction of toxicity to be defined. Specifically, the types of information that could be used, differential requirements for analogue and category approaches and whether informatics / statistical approaches could be applied. The study utilised previously collected data for triazoles [14] to demonstrate the definition of domain for the analogue approach, whilst concurrently extending the analysis to imidazoles for the category approach. In addition, existing literature studies for read-across scenarios involving the production of a common metabolite, for biological read-across and unknown or variable composition, complex reaction products (UVCBs) were considered. This builds on both well-established means, and more recent proposals, to compile information to define uncertainties [10, 12] and justify analogue selection for read-across [4].

1. **Methods: The Approach to Define Applicability Domains for Read-Across**

*2.1 Criteria to define the applicability domain*

Based on ECHA’s RAAF [3] and the assessment of uncertainties of read-across relating to similarity [12], three criteria were identified as being crucial to define the applicability domain of an analogue or category. These correlate with the four types of evidence required to support read-across that were previously defined by Suter and Lizarraga [4]: structural features and physico-chemical attributes (herein jointly considered as “chemistry”), toxicodynamics and toxickinetics. The three criteria are noted below and defined in more detail in Table 1.

1. Chemistry. In general, a prime consideration in read-across is the chemical structure and properties of the target and source chemicals and how their similarity may be defined. This may be in terms of fragments or chemical class, physico-chemical properties and a relevant metric for structural similarity.
2. Toxicodynamics. Analogues should, ideally, have a demonstrable common mechanism of action in eliciting toxicity or absence of a specific mechanism. Possession of similar structural fragments, known to be associated with a given toxicity, may provide evidence whilst the absence of such toxic fragments may also be informative. *In silico*, *in chemico* and / or *in vitro* New Approach Methodology (NAM) data may play a role in defining toxicodynamics. Whilst toxicodynamic data may be used in conjunction with other data sources, approaches to read-across based solely on biological similarity using NAM data, such as from those -omics are also becoming commonplace, see Nakagawa et al. [15] and Escher et al. [16] as excellent examples.
3. Toxicokinetics. Historically, a significant barrier to acceptance of read-across was the failure to establish similarity in toxicokinetic behaviour [17] and there are several factors to consider. A significant concern is where the target molecule may have greater bioavailability than the source chemical, in general, and more specifically a higher level of exposure at the target site. Another concern is potential differences in metabolic profile, such as the rate, extent and route of metabolism, and in particular whether the target chemical may produce toxic metabolites not produced (or produced to a lesser extent) than the source chemical.

Table 1 summarises the key factors to consider in determining the chemical, toxicodynamic and toxicokinetic domains of a read across prediction. The relevance of these domains for different read-across scenarios are specifically identified in Table 1 and the concepts are expanded on further below.

Table 1. Key applicability domain criteria for the relationship between target and source chemical(s) indicating which are most relevant for given read-across scenarios

|  |
| --- |
| **Key Applicability Domain Criteria** |
|  |
| **Chemistry** | **Toxicodynamics** | **Toxicokinetics** |
|  |
| *Chemical substructure:* - Presence of key (activity defining) structural fragment(s); additional fragments have no significant dissimilarities\* (i.e. unlikely to modify activity)*Molecular similarity:*- Determined by appropriate similarity metric*Physico-chemical properties:*- Relevant physico-chemical properties are within requisite boundaries | *Toxicological activity:*- Toxicological endpoint data*MIE:* - Ability to elicit same molecular initiating event*Mechanistic analysis:* - Evidence of similarity in mode/mechanism of action\*; (conversely no evidence (or likelihood) of specific toxicities that are elicited only by the target or the source\*) | *ADME:*- Evidence of similarity in key absorption, distribution, metabolism (see also below) and excretion (ADME) properties\*- Level of target organ exposure (indicated by bioavailability, plasma/organ concentration-time profile; maximal or steady-state concentrations)*Metabolism:*Similarity in metabolic profile (where relevant) |
|  |
| **Scenario-Specific Considerations** |
|  |
|  | **Biological Read-Across:** Focus on similarity of biological profile (Toxicodynamics) | **Common Metabolite / Degradation Product:**Is the transformation product formed; is rate and extent of formation similar (Toxicokinetics) |
| **UVCB / Mixtures:** Focus on similarity of chemical profile with regard to structures / proportion of structures and link to activity (Chemistry and Toxicodynamics) |  |
| **Analogue / Category Approach:** All aspects considered (Chemistry, Toxicodynamics and Toxicokinetics) |

*\*according to current knowledge*

*2.2 Applying the read-across domain criteria*

Each of the applicability domain criteria described in Table 1 are considered individually in Table 2 [18]. The nature of the information available for compounds influences how readily a compound can be determined as being within or outside of the domain (in relation to a specific criteria). Table 2 summarises typical information available, how a domain may be defined and, pragmatically, how compounds can be assigned as being in or out of domain.

Applicability domain can most readily be defined using chemical structure, properties and/or similarity; such approaches are often applied as a first step as these criteria lend themselves to being quantified easily. Other criteria are associated with the presence or absence of one or more structural features or chemical / biological properties. At best, these types of criteria can be only semi-quantitative and are most likely to be qualitative at the current time. They do, however, serve as a useful means of establishing differences between molecules. Defining the domain using toxicodynamics or toxicokinetics is associated with a higher level of complexity and it is generally more difficult to assign chemicals to a domain using this type of information, particularly where qualitative descriptions are given, such as “rapid” absorption. Thus, it is inevitable that the domain cannot be described quantitatively for a proportion of the criteria, making overtly parametric approaches to describe the domain difficult to apply.

Table 2. Features of the key applicability domain criteria including metrics and typical information, as well as how to determine whether compounds are within our outside of the domain

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Metric**  | **Type of information / examples**  | **Ease of assigning to domain?** | **Defining the domain and assessing whether compounds are in or out of domain** | **Comments** |
|  |
| **Chemistry** |
|  |
| **Chemical substructure / fragment(s)**  | Presence, absence or analysis of differences in fragments e.g. analogue(s) or member of mechanistic domain. | Readily assignable - indicated by presence or absence of functional groups or molecular substructures | Define key structural fragments of target and source molecule(s):Where differences in key fragments exist between source and target, determine whether these may potentially impact toxicity. | May not be defined for mixtures / UVCBsEffect of additional fragments may need to be investigated with e.g. *in silico* profilers, NAM data, etc |
| **Molecular similarity**  | Comparison of fingerprint(s) using a suitable metric (e.g. Tanimoto, Dice coefficient) | Readily assignable - determined by value of similarity metric | Calculate score using chosen similarity metric(s):Higher scores within domain, lower scores out of domain – a cut-off value can be arbitrarily determined | Scores should not be used in isolation as they are relative to the fingerprint and metric – may show consistency within a defined group. |
| **Physico-chemical properties** | Log P, boiling point, aqueous solubility, vapour pressure etc. | Readily assignable - If experimental data are available or property is readily calculable | Obtain experimental value or calculate property using a suitable method:Consider the magnitude in difference between target and source relative to: the range of the property; experimental variability and; context of application. For example: log P within 1 log unit; aqueous solubility within 1.5 log units; vapour pressure within 2 log units or M wt/boiling point/vapour pressure to be within “X” Da/oC/mmHg of target respectively (where “X” is proportional to the magnitude of the value for the target). | Consider appropriateness experimental versus calculated values; boundary of domain determined by property range, variability and context |
|  |
| **Toxicodynamics** |
|  |
| **Toxicology: Existing toxicity data** | Toxicity data for the endpoint to read-across (for the source molecule); other information for both target and source e.g. data for analogous / associated endpoints; data from non-standard assays (e.g not performed to Test Guideline); output from in silico studies (e.g. predictions from QSARs or software such as QSAR Toolbox, Toxtree etc). | Moderately complex to assign – may require expert interpretation of existing toxicity data, or profiles | Obtain relevant experimental data or perform *in silico* assessment.Assess whether there are significant differences in individual data or overall toxicological profile that would indicate the compound is out of domain. | Comparative data may not be available and data quality / reliability of predictions needs to be considered. |
| **Toxicology: Molecular Initiating Event** (Description of the MIE which may be part of an AOP) | Knowledge of the MIE may define it as being receptor mediating, reactive, acute enzyme inhibition, blockade of biochemical pathways, etc. Information on grouping should reflect this knowledge | Variable – MIE may be well-established or may be an unknown, ill-defined or non-specific MIE | Define MIE e.g. according to Cronin and Richarz [18]Determine if the MIE is the same and if similarity in molecular structure reflects the MIE? If not, the compound will be out of the MIE-defined domain. |  |
| **Toxicology: Mechanistic Analysis** | Biological data e.g. data from NAM, -omics or high content assays used to confirm similarity in mechanism of toxic action | Variable – data from NAMs may be directly associated with mechanism of action; however difficult to definitively assign using “high content” / -omics data | Obtain relevant data from biological assays, NAMs, high content or -omics assays.Establish if NAM assay results and/or high content /-omics assays indicate a commonality in mechanism of action  | Analysis of high content /-omics data can be intensive, although data are becoming more accessible; application of NAMs is rapidly expanding at present |
|  |
| **Toxicokinetics** |
|  |
| **Toxicokinetic (TK) parameters** Relating to absorption, distribution, metabolism\* and excretion (ADME) | Data available from toxicokinetic studies or predicted using software, includes overall bioavailability (via relevant route), half-life, area under concentration-time curve (AUC), rate/percentage absorbed; volume of distribution; plasma or tissue protein binding; clearance rate or percentage excreted by kidneys, liver, lung etc | Variable – Some TK parameters may be directly comparable; others require expert analysis of toxicokinetic profile | Obtain data from experimental toxicokinetic studies or *in silico* predictionsDetermine if there are significant differences in toxicokinetic data (individual values or AUC profile) that may indicate overall differences in bioavailability at relevant site, potential to accumulate or metabolism\*. | Differences in TK parameters such as bioavailability or accumulation may be accommodated if sufficiently understood, by adjusting the read-across value or using a safety factor |
| **Metabolism\*:** Metabolite identity, rate or extent of metabolism | Experimentally determined metabolites, metabolites predicted using software or assumption that target and source have same primary metabolite that is responsible for toxicity; metabolism rate data (usually experimentally derived) |  | Identify metabolites from experimental studies or *in silico* prediction; (rate data may be available from experimental studies).Determine whether target and source have common metabolites (especially where the metabolite is associated with toxicity); determine if either have additional metabolites that are associated with other toxic effects. Compounds are in domain if the same key metabolite(s) are formed and other metabolites formed by either target or source are deemed inconsequential to the toxicity. Comparison of rate of metabolism may also be relevant when comparing bioavailability or half-life. | Metabolism may be key consideration for certain read-across scenarios (e.g. formation of a common metabolite associated with toxicity) . Differences in metabolites or rate of metabolism can have a significant impact on read-across. |

One approach for read-across is the analogue approach, whereby the target and source molecules are compared directly. In reality, it is not possible to make a direct and objective comparison of the *absolute* similarity or dissimilarity of two molecules; this can only be considered in relation to specific properties. It may be possible to consider whether the two molecules are highly similar in terms of relevant physico-chemical properties such as molecular weight (MW), logarithm of the octanol-water partition coefficient (log P) etc. The category approach, however, lends itself to a more quantitative approach i.e. the property range within a category can be considered. Analysis of this sort has two distinct advantages, firstly it enables the determination of whether the target molecule fits within the chemical and toxicological space of the category, but also takes account of the heterogeneity of the category itself. Herein, the concepts of domain for an analogue and a category, and how the domain may be applied, are illustrated through the re-analysis of a grouping and read-across study proposed by Pestana et al. [14]. This study aimed to read-across the repeat dose toxicity of triazole fungicides, more specifically tetraconazole to hexaconazole as the analogue approach, and then beyond to a larger category. Pestana et al. [14] compiled a large amount of supporting data, including for NAMs, which were utilised and added to in this study. As indicated in Table 1, the analogue or category approaches require consideration of all three types of domain criteria (chemistry, toxicodynamic and toxicokinetic). For other read-across approaches there is greater focus on specific types of data, for example toxicodynamic data for biological read-across or toxicokinetic data for the common metabolite approach. The primary focus of this study was to demonstrate the practical application of domain definition, as relevant to the analogue and category approaches, however, application of the approach to other read-across methods (formation of a common metabolite or degradant, biological read-across and read-across for UVCBs or mixtures) are also discussed below.

*2.3 Data compilation to support the definition of domains*

The compounds in the data set considered as part of the analogue and category approaches are summarised in Table 3, including their identifiers and experimental No Observed Adverse Effect Levels (NOAELs). All values in Table 3 represent 90 day rat oral NOAEL values [19-36], which were considered to be the most representative publicly available value, as published previously by Pestana et al. [14].

Table 3. Name and Chemical Abstract Service (CAS) Number of the triazole and imidazole categories, with associated NOAEL values, considered in this study.

|  |  |  |  |
| --- | --- | --- | --- |
| Name and CAS Number | Structure | NOAEL values (mg/kg/day) | Reference for NOAEL value |
|  |
| **Target Molecule** |
|  |
| Tetraconazole112281-77-3 |  | 4 | EFSA [19] |
|  |
| **Source Molecules (Triazole category)** |
|  |
| Hexaconazole79983-71-4 |  | 2.5 | US EPA [20] |
| Bitertanol55179-31-2 |  | 8 | JMPR [21] |
| Cyproconazole94361-06-5 |  | 6.4 | EFSA [22] |
| Difenoconazole119446-68-3 |  | 20 | EFSA [23] |
| Epoxiconazole133855-98-8 |  | 7 | EFSA [24] |
| Fenbuconazole114369-43-6 |  | 5.7 | EFSA [25] |
| Myclobutanil88671-89-0 |  | 51.5 | JMPR [26] |
| Paclobutrazol76738-62-0 |  | 20 | EFSA [27] |
| Penconazole66246-88-6 |  | 25 | EFSA [28] |
| Propiconazole60207-90-1 |  | 15.9 | ECHA [29] |
| Prothioconazole178928-70-6 |  | 100 | JMPR [30] |
| Tebuconazole107534-96-3 |  | 9 | EFSA [31] |
| Triadimenol55219-65-3 |  | 9 | EFSA [32] |
| Triadimefon43121-43-3 |  | 13.6 | FAO [33] |
|  |
| Source Molecules (Imidazole category) |
|  |
| Imazalil35554-44-0 |  | 5 | EFSA [34] |
| Prochloraz67747-09-5 |  | 6 | EFSA [35] |
| Triflumizole68694-11-1 |  | 4.1 | EFSA [36] |
| Climbazole38083-17-9 |  | 45 | ECHA (extracted from the US EPA CompTox Chemicals Dashboard) |
| Ketoconazole65277-42-1 |  | 6.5 | NITE (extracted from the US EPA CompTox Chemicals Dashboard) |

The sources of information drawn upon to address the criteria describing read-across domain are detailed in Table 4 [18, 37] and were all publicly available. They represented either databases or *in silico* software accessible without charge. It is acknowledged that there are many other such sources, both freely and commercially available, that the user may wish to use.

Table 4. Sources of information defining the criteria addressed in the read-across domain.

|  |  |  |  |
| --- | --- | --- | --- |
| Criteria | Type of information | Resource | Reference and / or URL |
|  |
| **Chemistry** |
|  |
| Chemical substructure | Key structural fragment(s) for grouping  | OECD QSAR Toolbox: Profilers for mechanisms or endpoints | [www.qsartoolbox.org](http://www.qsartoolbox.org) |
| Additional fragment(s) defining structure | OECD QSAR Toolbox: Profilers for functional groups | [www.qsartoolbox.org](http://www.qsartoolbox.org) |
| Molecular similarity  | Tanimoto coefficient calculated from PubChem fingerprints | PubChem | https://pubchem.ncbi.nlm.nih.gov/ |
| Physico-chemical properties | Various physico-chemical properties  | United States Environmental Protection Agency (US EPA) Computational Toxicology (CompTox) Chemicals Dashboard | <https://comptox.epa.gov/dashboard> |
|  |
| **Toxicodynamics** |
|  |
| Toxicology | Primary toxicity data | EFSA OpenFoodTox database\*US EPA Chemicals Dashboard | <https://www.efsa.europa.eu/en/data/chemical-hazards-data>; <https://comptox.epa.gov/dashboard> |
| Non-standard data | US EPA Chemicals DashboardPubChem | <https://comptox.epa.gov/dashboard>; https://pubchem.ncbi.nlm.nih.gov/compound/ |
| Data for other (related) endpoints | PubChem | https://pubchem.ncbi.nlm.nih.gov/compound/ |
|  |  |  |
| Molecular Initiating Event | Definition according to type of interaction  | Scheme to classify MIEs according to the nature of the interaction | Cronin and Richarz [18] |
| Mechanistic analysis | *In silico* screening  | OECD QSAR Toolbox (ver 4.4.1) mechanistic profilers | [www.qsartoolbox.org](http://www.qsartoolbox.org) |
| ToxCast Data | US EPA Chemicals Dashboard | <https://comptox.epa.gov/dashboard> |
| Other mechanistic data | PubChem | https://pubchem.ncbi.nlm.nih.gov/compound/ |
|  |
| **Toxicokinetics** |
|  |
| Absorption, distribution, metabolism and excretion | Absorption, Distribution, Metabolism, Half-life | EFSA OpenFoodTox database; PubChem | <https://www.efsa.europa.eu/en/data/chemical-hazards-data>; <https://pubchem.ncbi.nlm.nih.gov/compound/> |
| Physiologically-Based Pharmacokinetics (PBPK) model predictions of distribution  | National Institute of Environmental Health Sciences Integrated Chemical Environment (NEIHS ICE) | <https://ice.ntp.niehs.nih.gov/Tools>; Bell et al. [37]  |
| Metabolite formation | Significant metabolites | OECD QSAR Toolbox (ver 4.4.1) | [www.qsartoolbox.org](http://www.qsartoolbox.org) |

* 1. *Applying the Applicability Domain Criteria to Different Types of Read-Across*

The criteria for the definition of an analogue or category approach, based on structurally-similar source chemicals, can also be applied to other types of read-across with some adaptations. The following read-across approaches were also considered:

1. formation of a common metabolite or degradant [38]
2. so-called “biological read-across” [16, 39]
3. UVCBs or mixtures [40]
4. **Results**

*3.1 The example of the analogue approach*

To illustrate the analogue approach for read-across, tetraconazole was selected as the target molecule and hexaconazole as the data-rich source molecule. The endpoint to be read across was repeated-dose toxicity and data and information were sourced, in the first instance, from Pestana et al. [14]. The authors developed a read-across approach supported by NAM data, to illustrate how uncertainty could be reduced by their inclusion. The purpose of the current study is not to determine uncertainty as such, although it may be derived from the approach, but rather to quantify the possible domain of the analogues.

Using the data summarised in Table 3, Pestana et al. [14] illustrated that within a group of triazoles (considered in Section 3.2), hexaconazole was an appropriate read-across candidate due to structural and toxicological similarity. The analysis detailed in Table 5 demonstrates that the two analogues are very similar in terms of their chemical and structural features and properties. In order to evaluate the chemical and other properties, some of the boundary criteria defined in Table 2 were assessed.

Table 5. Illustration of how the characteristics of two close analogues (tetraconazole and hexaconazole) can be compared according to the properties within each domain

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Applicability Domain Criteria  | Description | Target: Tetraconazole | Source: Hexaconazole | Comment | Conclusion |
|  |
| **Chemistry** |
|  |
| Chemical substructure | Key structural fragment(s) for grouping  | Triazole  | Triazole  | No significant differences | In domain |
| Additional fragment(s) defining structure | Alkane, ether moiety, aryl, aryl halide, perfluorocarbon | Alcohol, alkane, aryl, aryl halide, perfluorocarbon | Minor difference from ether to alcohol, which is unlikely to affect toxicity significantly  | In domain |
| Molecular similarity  | Tanimoto coefficient calculated from PubChem fingerprints | Tanimoto coefficient between tetraconazole and hexaconazole = 0.83 | N/A | High similarity | No reason to consider these compounds are not similar |
| Physico-chemical properties | MW (Da) | 372 | 314 | Relatively minor difference in molecular weight (58 Da) | In domain |
| Log P | 3.53 | 3.83 | Very minor difference in log P (0.3 log units) | In domain |
| Water Solubility (log solubility in moles/litre) | -3.03 | -4.26 | Moderate difference in water solubility (1.13 log units), unlikely to affect toxicity significantly | In domain |
| Vapour Pressure (log VP in mmHg) | -5.76 | -6.69 | Moderate difference in vapour pressure solubility (1.07 log units), unlikely to affect toxicity significantly | In domain |
| Boiling point (°C) | 339 | 321 | Very minor difference in boiling point (18 °C) | In domain |
| Conclusion and Justification (Chemistry Domain):  | The target and source molecules are highly similar with regard to the chemistry domain. The molecules share the same key structural fragment that is responsible for the dominant mode of action. There are no structural fragments that would be expected to affect toxicity significantly. All relevant physico-chemical properties are similar. |
|  |
| **Toxicodynamics** |
|  |
| Toxicology | Effect responsible for toxicity  | Liver toxicity | Liver toxicity | No significant difference | In domain |
| Non-standard data | None | None | N/A | In domain |
| Data for other (related) endpoints | None | None | N/A | In domain |
| Molecular Initiating Event | Interaction between the molecule and biological system causing the toxicity (i.e. the MIE) | Receptor binding (to nuclear receptors) | Receptor binding (to nuclear receptors) | No significant difference | In domain |
| Mechanistic analysis | *In silico* screening for oestrogen receptor binding | No alert | No alert | No significant difference | In domain |
| *In silico* screening for thyroid receptor binding | No alert | No alert | No significant difference | In domain |
| *In silico* screening for non-genotoxic carcinogenicity | No alert | No alert | No significant difference | In domain |
| *In silico* screening for Development and Reproductive Toxicity | Alerts for i) polyhalogenated benzene derivative and ii) thiazole derivative | Alerts for i) polyhalogenated benzene derivative and ii) thiazole derivative | No significant difference | In domain |
| *In silico* screening for protein binding | No alert | No alert | No significant difference | In domain |
| *In silico* screening for DNA binding | No alert | No alert | No significant difference | In domain |
| ToxCast FXR, RARa, AHR, CAR, PXR | Active | Active | No significant difference in nuclear receptor binding  | In domain |
| ToxCast CYP3A4, CYP1A2, CYP2C19, CYP2C9 | Active | Active | No significant difference in CYP activation  | In domain |
| Omics data | None | None | N/A | N/A |
| Conclusion and Justification (Toxicodynamics Domain):  | The target and source molecules are highly similar with regard to the toxicodynamic domain. There is strong evidence that the molecules share a common mode of action.  |
|  |
| **Toxicokinetics** |
|  |
| Absorption, distribution, metabolism and excretion  | Absorption | Rapid | Rapid | No significant difference | In domain |
| Distribution | Liver, kidneys, gonads, brain, bones | Kidney, liver and pancreas | Source compound not distributed as widely a target although high concentrations observed in the liver | In domain |
| Metabolism | Extensive | Extensive | No significant difference | In domain |
| Half-life (hours) | < 24 | 10 – 27 | Half-life is short, but potentially longer in source molecule, unlikely to significantly affect toxicity  | In domain |
| PBPK (liver maximum plasma concentration (Cmax)) (ng/mL) | 37.1 | 31.2 | Similar maximal concentrations in the liver | In domain |
| PBPK (Steady-state plasma concentration (Css)) (μM) | 50.1 | 1.43 | A large difference in Css is observed. Css is calculated to be significantly higher in the target molecule | Low confidence of being in domain, further information required |
| Metabolite formation | Major and significant metabolites | 1,2,4‐triazole; triazole alanine; triazole acetic acid | 1,2,4‐triazole; triazole alanine; triazole acetic acid | No significant difference in major and defining metabolites | In domain |
| Conclusion and Justification (Toxicokinetics Domain):  | The target and source molecules are highly similar with regard to the toxicokinetics domain. There is strong evidence that the molecules have similar absorption, distribution and excretion properties and no difference in major and significant metabolites although the effect of higher Css in the target should be acknowledged (although no significant differences were observed in the liver).  |
|  |  |
| Overall Conclusion and Justification: | There is strong evidence that target and source the molecules are highly similar with regard to the read-across domain indicating low uncertainty with regard to their similarity. |

*3.2 The example of the category approach*

The same principles to define the domain of an analogue can be applied to groups or categories of molecules that can be used for read-across. For a category, rather than there being a direct comparison of the information for two molecules, the ranges of descriptors (if continuous variables), or the presence or absence of specific properties or molecular features (if categorical variables), can be utilised. Table 6 provides a summary of domain, indicating that the target molecule falls within the domain of the azole category and hence would be a good source of information for read-across (the full data for the ranges for all compounds are provided in Supplementary Information Table S1). The ranges and properties that comprise the imidazole domain are more diverse, indicating a broader domain for this category. The breadth and diversity of a category, as defined by the domain, could potentially result in a category that is a less robust, without further evidence, source of read-across analogues.

The quantitative aspects of the read-across can be shown graphically. Figure 1 shows the scores for the first two principal components calculated from the continuous data in Table 6 / Supplementary Information Table S1 (logarithmically transformed values for MW, log P, BPt, VP, Css and Cmax were used where available). Whilst this gives a good understanding of the chemistry and toxicokinetic properties, it is not able to describe the toxicodynamic properties due to the lack of continuous data.

The plot of scores for the first two principal components, as shown in Figure 1, is a very simplistic representation of the chemical and toxicokinetic space of the triazole and imidazole categories. As such it should be used for illustrative purposes only. However, it does provide some useful information regarding the categories. In terms of chemical and toxicokinetic space, the two categories are relatively homogenous and, for the most part, the imidazoles fall into the same space as the triazoles. Hexaconazole, the analogue chosen by Pestana et al. [14], is relatively close in terms of this space. However, others are closer - the nearest being cyproconazole. As noted above, the imidazole category is broader than that of the triazoles. It is vital to note that whilst the two categories show consistency, no information regarding toxicodynamics or the significant metabolites is included and these issues should be considered by expert judgement.



Figure 1. Plot of the scores for the first two principal components (PCs) based on the continuous data for the triazole and imidazole categories’ chemical and toxicokinetics domains. Data for tebuconazole were not available so it is excluded. Green diamonds represent triazole compounds and blue circles imidazoles. The target compound (tetraconazole) is identified.

Table 6. Domain of the triazole category, with the possible extension to include imidazoles, as defined by the criteria set out in Table 2.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Applicability Domain Criteria  | Description | Target: Tetraconazole | Category Range: Triazoles | Conclusion | Category Range: Imidazoles | Conclusion |
|  |
| **Chemistry** |
|  |
| Chemical substructure | Key structural fragment(s) for grouping  | Triazole  | Triazole | Identical | Imidazole |  |
| Additional fragment(s) defining structure | Alkane, ether moiety, aryl, aryl halide, perfluorocarbon | Aryl, aryl halide in all compounds; alkanes in most compounds | Target in domain of the triazole category | Aryl, aryl halide in most compounds | Similarity of triazoles and imidazoles need to be assessed to determine if it will affect toxicity |
| Molecular similarity  | Tanimoto coefficient calculated from PubChem fingerprints | N/A | Most Tanimoto coefficients (with target) > 0.75 | Target in domain of the triazole category | Tanimoto coefficient of target between 0.50 - 0.70  | Similarity of target to imidazole category needs further evidence to be strongly in domain |
| Physico-chemical properties | MW (Da) | 372 | 292 - 406  | Target in domain of the triazole category | 293 - 531 | Target in domain of the imidazole category (which is broader than that of triazoles) |
| Log P | 3.53 | 2.75 - 4.26  | Target in domain of the triazoles category | 1.40 – 4.35 | Target in domain of the imidazole category, (which is broader than that of triazoles) |
| Water Solubility (log solubility in moles/litre) | -3.03 | -3.03 - -5.89 | Target in domain of the triazole category | -3.80 to -5,98 | Target in domain of the imidazole category (which is similar to that of triazoles) |
| Vapour Pressure (log VP in mmHg) | -5.76 | -5.73 - -9.21 | Target in domain of the triazole category |  -2.98 to - 9.57 | Target in domain of the imidazole category (which is broader than that of triazoles) |
| Boiling point (°C) | 339 | 321 - 350 | Target in domain of the triazole category | 336 -753 | Target in domain of the imidazole category (which is broader than that of triazoles) |
| Conclusion and Justification for the Triazoles Category (Chemistry Domain):  | The target and source molecules in the triazole category are highly similar with regard to the chemistry domain. The molecules share the same key structural fragment that is responsible for the dominant mode of action. There are no structural fragments that would be expected to affect toxicity significantly. All relevant physico-chemical properties are similar. |
| Conclusion and Justification for the Imidazoles Category (Chemistry Domain):  | The target and source molecules in the imidazole category are moderately similar with regard to the chemistry domain. The molecules share a highly related key structural fragment that is responsible for the dominant mode of action. There are few structural fragments that would be expected to affect toxicity significantly in the category. The target molecule falls within the range of relevant physico-chemical properties, but it should be noted that this range is relatively broad. |
|  |
| **Toxicodynamics** |
|  |
| Toxicology | Effect responsible for toxicity  | Liver toxicity | Liver toxicity | Target in domain of the triazole category | Liver toxicity | Target in domain of the imidazole category |
| Non-standard data | None | None | N/A | None | N/A |
| Data for other (related) endpoints | None | None | N/A | None | N/A |
| Molecular Initiating Event | Type of interaction between the molecule and biological system causing the toxicity (i.e. the MIE) | Receptor binding (to nuclear receptors) | Receptor binding (to nuclear receptors) | Target in domain of the triazole category | Receptor binding (to nuclear receptors) | Target in domain of the imidazole category |
| Mechanistic analysis | *In silico* screening for oestrogen receptor binding | No alert | No alert | Target in domain of the triazole category | No alert | Target in domain of the imidazole category |
| *In silico* screening for thyroid receptor binding | No alert | No alert | Target in domain of the triazole category | No alert | Target in domain of the imidazole category |
| *In silico* screening for non-genotoxic carcinogenicity | No alert | No alert | Target in domain of the triazole category | No alert | Target in domain of the imidazole category |
| *In silico* screening for Development and Reproductive Toxicity | Alerts for i) polyhalogenated benzene derivative and ii) thiazole derivative | Alerts for i) polyhalogenated benzene derivative and ii) thiazole derivative | Target in domain of the triazole category | Alerts for i) polyhalogenated benzene derivative and ii) imidazole derivative | Target in domain of the imidazole category |
| *In silico* screening for protein binding | No alert | No alert  | Target in domain of the triazole category | No alert  | Target in domain of the imidazole category |
| *In silico* screening for DNA binding | No alert | No alert except epoxy group in epoxiconazole | Target in domain of the triazole category | No alert except for triflumizole (Schiff base formation) | Target in domain of the imidazole category |
| ToxCast CYP3A4, CYP1A2, CYP2C19, CYP2C9 | Active | Active | Target in domain of the triazole category | Active | Target in domain of the imidazole category |
| Omics data | None | None | N/A | None | N/A |
| Conclusion and Justification for the Triazoles Category (Toxicodynamics Domain):  | The target and source molecules in the triazole category are highly similar with regard to the toxicodynamic domain. There is strong evidence that the molecules share a common mode of action.  |
| Conclusion and Justification for the Imidazoles Category (Toxicodynamics Domain):  | The target and source molecules in the imidazole category are highly similar with regard to the toxicodynamic domain. There is strong evidence that the molecules share a common mode of action.  |
|  |
| **Toxicokinetics** |
|  |
| Absorption, distribution, metabolism and excretion | Absorption | Rapid | Rapid | Target in domain of the triazole category | Rapid | Target in domain of the imidazole category |
| Distribution | Liver, kidneys, gonads, brain, bones | Liver and kidneys  | Target and category both show significant distribution to the liver, therefore in domain | Widely distributed; plasma and thyroids, liver and kidneys | Target and category both show significant distribution to the liver, however the wide distribution to other organs may be significant and would require further evidence |
| Metabolism | Extensive | Extensive | Target in domain of the triazole category | Extensive | Target in domain of the imidazole category |
| Half-life (hours) | < 24 | < 96 | Target in domain of the triazole category | < 96 (not available for all) | Target in domain of the imidazole category |
| PBPK (liver maximum plasma concentration (Cmax)) (ng/mL) | 37.1 | 10.25 – 55.12 | Target in domain of the triazole category | 9.75 – 46.48  | Target in domain of the imidazole category |
| PBPK (Steady-state plasma concentration (Css)) (μM) | 50.1 | 0.43 - 83.66 | Target in domain of the triazole category | 0.33 – 77.86 | Target in domain of the imidazole category |
| Metabolite formation | Major and significant metabolites | 1,2,4‐triazole; triazole alanine; triazole acetic acid | 1,2,4‐triazole; triazole alanine; triazole acetic acid | Target in domain of the triazole category | Cleavage to the imidazole ring in some of the compounds  | Target is not in the metabolic domain of the imidazoles. Further evidence would be required to assess the potential for the different imidazoles to affect the toxicity outcome |
| Conclusion and Justification for the Triazoles Category (Toxicokinetics Domain):  | The target and source molecules in the triazole category are highly similar with regard to the toxicokinetics domain. There is strong evidence that the molecules have similar absorption, distribution and excretion properties and no difference in major and significant metabolites.  |
| Conclusion and Justification for the Imidazoles Category (Toxicokinetics Domain):  | The target and source molecules in the imidazole category are moderately similar with regard to the toxicokinetics domain. There is evidence that some, but not all, molecules in the category have similar absorption, distribution and excretion properties. There is a difference in major and significant metabolites. Further evidence of the toxicokinetics properties may be required to reduce uncertainty. |
|  |  |
| Overall Conclusion and Justification for the Triazole Category: | There is strong evidence that target and source the molecules are highly similar with regard to the read-across domain indicating low uncertainty with regard to their similarity. The target molecule is highly likely to be a member of this category and the conservative read-across of a NOAEL is justified. |
| Overall Conclusion and Justification for the Imidazole Category: | There is moderate evidence that target and source the molecules are similar with regard to the read-across domain indicating low uncertainty with regard to their similarity. The target molecule is highly likely to share the same mode of toxicity action but possible differences in toxicokinetics must be investigated in order to ascertain whether a conservative read-across of a NOAEL is justified. |

*3.3 Consideration of other types of read-across*

The definition of applicability domain can be adapted for other types of read-across, as indicated in Table 1. Table 7 summarises the criteria identified in Tables 1 and 2 that would be important for each type of read-across.

For read-across associated with a common metabolite or degradant, as exemplified by the study reported by Ball et al. [38], the fundamental aspects are to ensure that the identical metabolite or degradant is formed, and that the rate of formation is comparable between the target and source molecules. As shown in Table 7, this should be demonstrated in terms of the chemical structure. Should a common structure (for the relevant metabolite) be identified then the structural similarity and physico-chemical properties will be assumed to be identical. Likewise, toxicodynamics need not be considered as the read-across molecule will be assumed to be the same. The only factors that will need to be demonstrated relate to toxicokinetics, specifically the formation of the key metabolite.

Biological read-across and read-across for UVCBs / mixtures share many commonalities, indeed biological read-across is likely to be a key tool to fill data gaps for such assortments of substances. For the purposes of this discussion, biological read-across is taken to be an approach where chemical structure may not be known and the read-across is performed on the basis of similar biological data. Such data may typically be from –omics outputs or the results of high content assays, as exemplified by a number of studies e.g. Escher et al [16], Grimm et al. [39], Nakagawa et al. [15] and van Ravenzwaay et al. [41]. As reported in Table 7, for read-across candidates with known chemical structure, such similarity can form part of the read-across argument, although strictly speaking is not required. For biological read-across to be acceptable, it is the toxicodynamics that must be demonstrated to be similar through similar (experimental) biological responses.

Consideration of read-across for UVCBs and mixtures in many ways applies and extends the process of biological read-across [42]. However, as noted in Table 7, for UVCBs and mixtures it may still be possible to categorise the structures of chemical constituents [43-45], although this is not theoretically required. As with biological read-across, similarity in mixtures and UVCBs may be assumed by similarity in biological responses including NAMs data, as exemplified by House et al. [40].

Table 7. Relevance of the criteria for defining the applicability domain of a read-across (as described in Tables 1 and 2) for different types of read-across applications.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Analogue or Category | Common Metabolite / Degradant | Biological Read-Across | UVCB / Mixture |
|  |
| **Chemistry** |
|  |
| Chemical substructure | Essential – similarity in analogues and category members must be demonstrated | Essential – common metabolite or degradant is assumed | Non-essential. If structures are known, this may be helpful but is not a vital part of the read-across process | Non-essential. If structures are known, this may be helpful but is not a vital part of the read-across process |
| Molecular similarity  | Essential – similarity in analogues and category members must be demonstrated | Not relevant - common metabolite or degradant is assumed | Non-essential. If structures are known, then information may be provided | Not relevant – similarity metrics cannot currently be calculated for UVCBs or mixtures |
| Physico-chemical properties | Essential – similarity in analogues and category members must be demonstrated | Not relevant - common metabolite or degradant is assumed | Non-essential. If structures are known, then information may be provided | Non-essential. If properties are known, then information may be provided |
|  |
| **Toxicodynamics** |
|  |
| Toxicology | Essential – similarity in analogues and category members must be demonstrated | Not relevant - common metabolite or degradant is assumed | Essential – similarity must be demonstrated | Essential – similarity must be demonstrated |
| Molecular Initiating Event | Essential – similarity in analogues and category members must be demonstrated | Not relevant - common metabolite or degradant is assumed | Essential – similarity must be demonstrated | Non-essential – it may not be possible to demonstrate a common MIE |
| Mechanistic analysis | Essential – similarity in analogues and category members must be demonstrated | Not relevant - common metabolite or degradant is assumed | Essential – similarity must be demonstrated | Essential – similarity must be demonstrated |
|  |
| **Toxicokinetics** |
|  |
| Absorption, distribution, metabolism and excretion  | Essential – similarity in analogues and category members must be demonstrated | Not relevant - common metabolite or degradant is assumed | Non-essential. If ADME information is known, then information may be provided | Non-essential. If ADME information is known, then information may be provided |
| Metabolite formation | Essential – similarity in analogues and category members must be demonstrated | Essential – demonstrating the formation of a common metabolite / degradant, with similar rates of formation, is crucial to ensuring successful read-across | Non-essential. If metabolic information is known, then information may be provided | Non-essential. If metabolic information is known, then information may be provided |

1. **Discussion**

This study has defined the domains of analogues and categories for read-across with a particular focus on deriving a suitable conclusion for the question of whether two, or more, compounds are “similar”. The complexity of answering this question is well known and is the cornerstone of acceptance of read-across predictions for data gap filling [2]. Until this time, the answer to this question has either been a subjective response built on weight of evidence [4, 46], or a more parametric approach based solely on molecular structure and properties [47]. The approach described in this paper attempts to draw on the strengths of both types of analyses, in turn reducing the weaknesses. Specifically, in this study we have derived a scheme that incorporates all properties of analogues identified by ECHA’s RAAF [3], as well as those associated with known uncertainties [5]. Going beyond the identification of these properties, it has attempted to quantify differences between molecules providing a means of defining the similarity between analogues.

The definition of the domain of read-across is defined in three steps relating to chemistry (structure and properties), toxicodynamics and toxicokinetics. The significance of each is described in more detail below.

*4.1 Chemistry Domain*

The simplest step in defining the domain of a read-across analogue relates to chemical structure and properties. This is the most fundamental for similarity with regard to read-across and (with the possible exception of consideration of UVCBs, mixtures and for biological read-across) the structural basis of the read-across should be stated explicitly. Should a molecule (or a constituent of a UVCB or mixture) be deemed dissimilar in terms of structure or properties, then it would be considered to fall out of domain. There are, of course, many subtleties here. Whilst category members may possess quite different global structures, provided that the key fragment driving toxicity is preserved, it may be considered similar. However, it is also recognised that even very minor structural changes can be associated with “activity cliffs” [48]. These are, in essence, dramatic shifts in bioactivity resulting from apparently minor structural modifications, which would lead to the discounting of an analogue.

To illustrate the issue of activity cliffs, the use of read-across for skin sensitisation is a case in point. The MIE for skin sensitisation is the covalent reaction of a hapten with a skin immunoprotein [49]. Therefore, for this endpoint, analogues can potentially be read across from fragments associated with reactivity (many of these being publicly available, e.g. Enoch et al. [50, 51]. A small change in the environment about such a fragment may significantly alter reactivity. For instance, as shown in Figure 2, the susceptibility of the *α,β*-unsaturated carbonyl moiety towards conjugate addition renders it a standard alert for protein alkylation – potentially enabling read-across over a range of molecules broader than close structural analogues. However, the reactivity may be significantly influenced by the identity of the substituents present at either alkene carbon atom. The simple addition of a methyl substituent (an electron donating group) at these positions shall impair reactivity, whereas a halogen atom (electron withdrawing) shall enhance it [52]. Thus, there is a need to accurately identify and describe the structural basis for grouping in order that such issues may be accounted for.



Figure 2. An example of an “activity cliff” where small changes in chemical structure may have a significant influence on toxicity. The example shown illustrates the α-substituent effect upon the reactivity of the cinnamaldehyde derivatives towards reduced glutathione (GSH).

Table 2 illustrates the range of properties that can be considered to define the chemistry domain for a read-across analogue. For a single chemical (as opposed to a UVCB), the starting point is the definition of the structural fragment(s) on which the grouping is based. For the purposes of the triazole fungicides, it is the triazole structure that is the fundamental key unit for read-across. This is on the basis of it promoting a recognisable mode of toxic action [53]. It is noted that different types of fragments will be required for different endpoints. The molecule is then defined further in the fragments that describe the structure – such that any differences in structure may be determined. Chemical similarity can be ensured by a metric such as provided by the Tanimoto score derived from a suitable molecular fingerprint. The score can be provided for the one-to-one analogues (see Table 3), but in this case would be taken in isolation – it being difficult to demonstrate similarity on a single score alone [54]. However, if a low similarity score was determined, this may need to be investigated and understood. Similarity scores may be more useful for the category approach (Table 6) and may give an indication of the “consistency” of analogues, in other words whether all scores are in a similar range. Thus, for the triazoles, the individual Tanimoto score shows a high degree of similarity (0.83) between the target and source molecules, with an overall highly consistent range of scores (0.62 to 0.85).

There is a large range of physico-chemical properties that can be calculated for the molecules included in the read-across. A small selection is given in Tables 5 and 6. Some are rather fundamental, e.g. log P and molecular weight, and will probably be appropriate regardless of the chemistry or endpoints. Others will be more context dependent: for instance measures of volatility, such as vapour pressure and Henry’s Law constant, may be useful for inhalation effects. In addition, measures of stability (or degradation) could be included here and may be essential for environmental endpoints. The triazoles show similar ranges of properties with regard to those selected. The issue of how close a compound should be in terms of absolute properties is discussed in Section 4.4.

*4.2 Toxicodynamics Domain*

A significant means of defining similarity for read-across is through reference to the toxicodynamics. Indeed, this is the fundamental premise of biological-read across (and its application for UVCBs and mixtures) - namely that source and target substances can be demonstrably related in terms of a common mechanism or mode of action. Since many NAMs data are related to the molecular initiating event (MIE) of the adverse outcome pathway (AOP), this will become a critical aspect of determining similarity [55]. In order to form the relationship between MIEs and chemical structure and properties, Cronin and Richarz [18[ categorised MIEs in terms of relevant molecular features that could be defined. Thus, MIE-derived structural similarity could be based around fragments related to reactivity for genotoxicity, sensitisation, corrosion etc., but may further relate to toxicophores if the activity is as a result of receptor mediated effects. If the mechanism, or MIE, is unknown, then a structural analogue approach will be preferred. For biological read-across, activation of, or association with, an MIE will be on the basis of appropriate experimental data. Naturally, confirmation by both chemical and biological data will increase confidence.

Toxicodynamics, specifically mechanisms of action, are essential to determining if a specific toxicity is being read across. The challenging case of “no or low” toxicity obviously removes this requirement, and accordingly increases complexity by ensuring no specific mechanisms may be apparent for the target molecule [56]. Proving a mechanism of action definitively is a lengthy and complex process which is likely to require experimental evidence, for instance -omics analyses [41] or *in vitro* data [57, 58]. *In silico* profiling will go some way to providing a weight of evidence, and is rapid, practical and potentially cost-effective. It will, however, seldom provide as reliable a result as experimental determination.

At this point it must be recognised that it may not be possible to assign a particular mode or mechanism of action to a compound. Examples of this occur when the mode of action may be unknown, or uncharacterisable. In addition, there will be many chemicals which are classified as “no or low” toxicity with regard to chronic effects, e.g. a high NO(A)EL value may be determined for the source molecule(s) [56]. Since it is not strictly possible to confirm the absence of specific modes or mechanisms, a more realistic approach would be to gather information and lines of evidence that confirm, through an overall WoE, that there are no specific toxicities i.e. no differences in NAMs associated with the target and source molecule.

The triazoles represent a data-rich class of compounds for which a specific mode of toxicity is identifiable. Thus, the toxicodynamics domain is founded on a nuclear receptor-centred mechanism of action, for which evidence is given by a variety of NAM data incorporating *in silico* screening and *in vitro* assays. These, indirectly at least, provide evidence for the toxicodynamic domain. It must be noted that similar *in silico* and *in vitro* profiles do not necessarily provide a definitive confirmation of similar mechanism, but, when combined with structural evidence, they will provide reasonable assurance that no other mechanisms of action may be present.

*4.3 Toxicokinetics Domain*

The third aspect to the definition of read-across domain is the consideration of toxicokinetics. Potential differences in toxicokinetics should be addressed if they are likely to be significant i.e. greater systemic bioavailability or concentration at the site of action, for the target as compared to the source molecule(s). It is possible that toxicokinetics can be ignored if they are not relevant to the endpoint being read across. For instance, *in vitro* endpoints such as the Ames Test, or local effects such as eye or skin corrosion may be much less susceptible to the effects of uptake and systemic absorption – although aqueous solubility may also be an issue. For most, if not all, systemic effects toxicokinetics should be at least considered, if not quantified in some manner.

Two important aspects of toxicokinetics relate to systemic bioavailability and production of similar metabolites. These have been identified as significant areas of uncertainty in read-across [2], not least due to the lack of suitable *in vivo* data, cost and complexity of obtaining even *in vitro* data and the paucity of reliable *in silico* approaches. However, there has been a greater realisation of the need for this information and it is more frequently being utilised. In assessing both systemic bioavailability and production of significant metabolites, if suitable experimental data are absent or insufficient, *in silico* tools may be used. In this study a PBPK prediction was used in conjunction with SARs and QSARs for ADME properties. Likewise, *in silico* profilers for metabolism can be applied to predict the metabolites that may be produced. Where there is uncertainty in the predictions themselves, it is often better not to consider them as authoritative values, but rather as a means to identify significant differences in overall toxicokinetics between two, or more, molecules. As noted above, for read-across based on a common metabolite or degradant, this aspect of the domain (i.e. formation of identical metabolites or degradants at a similar extent and rate) is essential to enable justification of similarity [38].

Many types of toxicokinetic information were available for the triazoles. All compounds were shown to have similar toxicokinetic properties. It is appreciated that reliable toxicokinetic data are likely to be sparse and that much reliance will be placed on NAMs. *In silico* and *in vitro* approaches will certainly support the concept of toxicokinetic similarity, as well as highlighting any potential differences in toxicokinetic behaviour. For the triazoles considered in this study, it was also possible to run a generic PBPK model, in this case a freely available on-line model (https://ice.ntp.niehs.nih.gov/Tools). Data from such models should also be used to identify dissimilarities as well as similarities, and the potentially high level of uncertainty should be borne in mind. For the triazole category, all are seen to accumulate predominantly in the liver (although this is likely for any compound) and at similar time points. This would be consistent with the promotion of liver toxicity above adverse effects to other organs.

The measurement or prediction of metabolites is one of the most difficult aspects of defining the domain for read-across. At the current time, unless detailed experimental data are available, this is likely to involve checking that the major metabolites of the target and source molecules are similar (especially if one is known to be responsible for the toxic effect) and that there are no potential additional metabolites that may cause concern for toxicity. For the triazoles, the analysis demonstrates that there is likely to be a consistent set of metabolites, which thus supports the overall hypothesis of similarity.

*4.4 Boundaries of the read-across domains*

If required, the most difficult aspect of employing the applicability domain approach for read-across will be the definition of bounds, or upper and lower limits, both for analogues and for categories. Taking the one-to-one analogue approach, potential cut-offs for similarity could be utilised for all numerical properties, as proposed in Table 2 and implemented in Table 5 and, to a lesser extent, for the categories in Table 6. The logic behind these boundaries will need to be explored, bringing into question: what is an acceptable difference for analogues? It should also be noted that many descriptors will be correlated and there should be consistency between the boundaries, for instance since log P and aqueous solubility are often strongly related, there is no need to have a wide boundary for one and a narrow boundary for the other. The dependency of toxicokinetics properties on log P is likely to be of particular significance here.

One means by which to determine acceptable boundaries may be to consider what makes a one-to-one analogue appropriate. If this is considered in terms of the addition of a certain number of atoms or functional groups, then values could be calculated e.g. MW and log P values for the addition or removal of one, two or more carbon atoms, hydroxy groups etc. For instance, in Table 2 it is stated that analogues should be within “X” Da (MW) or 1 log unit (log P). These are arbitrary boundaries created for convenience, in reality the decision must be whether there will be significant difference in the toxicodynamics and toxicokinetics of the source and target molecule. Identifying whether two substances are implicitly similar is difficult and would rely on quantitative measures or properties. Boundaries of applicability domain for analogues defined by properties are likely to be most relevant for toxicokinetics. For instance, a difference of 1 log unit of log P will impact absorption and distribution, and if these are lower in the target molecule then this is justifiable. If absorption and distribution are greater in the target, then 1 log unit is likely to be “reasonable” and within experimental limits.

The key points to making the definition of applicability domain truly quantitative are firstly assessing the impact of differences in properties for a particular endpoint, secondly determining whether the impact is likely to make the target molecule more or less toxic (with lower toxicity being the preferred and precautionary option) and thirdly being able to agree what is a significant difference between two molecules. The final point could, for instance, be addressed through an understanding of the error associated with the experimental data. Thus, if the likely difference between molecules is within experimental error, this could be used as a criterion that there is high confidence of the compounds being within domain. At the current time, we lack an understanding of the error associated with many toxicity data, although analyses of large data sets with multiple values for the same compound may be able to give insight into this [59, 60]. The use of an empirically derived similarity index may, at first sight, seem an ideal solution to the issue of defining domains. However, as stated above, it must be remembered that metrics such as the Tanimoto index are dependent on the quality of information on which they are based [54]. Attempts have been made to identify cut-offs for identifying similar chemicals [61], however the reality is that they must be intelligently and flexibly applied to be successful [62]. The use of cut-offs based on Tanimoto indices as definitive measures of similarity, category membership or falling within an applicability domain is not recommended at this time without considerable care and expertise.

Given the difficulty in defining the confidence of being within applicability domains, it is probable that flexible and context-dependent boundaries may need to be decided upon, which become part of the similarity justification itself. In order to facilitate the use of such an approach, a set of boundary conditions have been defined for the one-to-one read across example for tetraconazole, as shown in Table 5. It is vital to stress that these were developed for this read-across alone and were developed through a problem formulation process. In reality, such boundaries are likely to be arbitrary, but they could be justifiable. There may be occasions where closely related analogues are essential, other situations where greater flexibility may be justifiable in analogue selection.

The situation for categories becomes slightly less challenging. The category implicitly defines the range of properties and effects, and analogues should be checked as to whether they fall within these ranges. For the triazoles, in terms of chemical properties and descriptors, the category is quite broad but it does demonstrate that the target molecule is representative. Indeed, given sufficient relevant quantitative information, it is possible to determine the closest analogues in “read-across relevant category space”.

*4.5 Overall Assessment for Triazoles*

In order to demonstrate how the approach for domain assessment could be applied, the information required in Table 2 was compiled. Read-across starts with problem formulation, which gives the opportunity to state the problem to be addressed and identify which data are required. At this point boundaries could be set for quantifiable analogues. Following this, is the data gathering phase, as described for the triazoles in Tables 5 and 6. Following compilation of the data, a series of questions clarify whether the domains are met and what possible consequences arise should they not have been. In some cases there may be an opportunity to obtain further information, either by measurement or from predictions. There are opportunities in the domain definition to use alternative sources of information, for instance where definitive chemical structures cannot be achieved e.g. if similarity can be obtained from biological read-across. This biological read-across concept implies that similar substances have similar toxicological profiles when measured experimentally [16, 39, 40].

The properties listed in Table 2 have been compiled for both the one-to-one analogue and for the triazole and imidazole categories (Supplementary Information Table S1). Thus, in this study, the use of the applicability domain has confirmed the suitability of hexaconazole to be an analogue for tetraconazole. Obviously, the criteria to define the domain can be tightened or adjusted as is required. The values to confirm similarity are arbitrary at this time and are based solely on the authors’ experience and what may be considered to have little significant effect in terms of changing properties, or be within expected limits of experimental error. It must also be stated that the domain must be considered in a flexible context, with different properties being important for different endpoints and effects e.g. for dermal toxicity, internal organ concentrations would not be required, whilst for environmental effects, other factors such as fate and persistence in the different media (e.g. water, soil, air, etc) may be of importance.

The applicability domain concept can also be applied to potentially expand the categories of molecules. Analysis of the domain through descriptors has at least two applications: firstly it will determine whether a new molecule is within the domain of a category; secondly it may assist in the better identification of one or more similar analogues to the target from within the group. Therefore in this study, the domain of the triazoles is defined. The target and source molecules exemplified by the analogue approach are shown to be in close proximity, and as such to be suitable analogues. This approach also provides an opportunity to determine whether the triazole group can be expanded to include the imidazoles. From a chemistry perspective it would appear that such an expansion is feasible. However, the key question that would need to be addressed is whether there is sufficient similarity in mechanism. This final question may be resolved by *in silico* profiling or, more likely, by examining existing *in vivo* data e.g. for commonality in adverse effects or measuring further in vitro data.

There are several advantages to the analysis of ranges of variables when describing the domain of a category. They provide the opportunity to determine if the target molecule falls within the range – which can be confirmatory of being in domain. Also, if a target molecule appears to fall out the range, this can be considered in the context of the read-across problem formulation e.g. is this a significant issue and / or can further information be obtained? This further information may be arise in the form of additional NAM data, or it may instead be measured rather than estimated. In addition, they may give an insight into the “quality” of a category in terms of its suitability i.e. a narrow range of descriptors may indicate a “tight” domain and could be indicative of closely related molecules. Furthermore, making the domain quantitative allows for the selection of highly related analogues, if a choice is available.

Consideration of the data in Table 6 demonstrates how the definition of domains for read-across assists in our understanding of the quality of read-across analogues. The target molecule (tetraconazole) is seen to fit closely in the domain of the azole category, as would be expected. The domain of the category is quite consistent and relatively restricted in terms of the descriptors considered. The multivariate analysis of the domain e.g. Figure 1, can allow for a smaller number of closely related analogues. This is analogous to the process of “sub-categorisation” [54], albeit in this case it would be on the basis of predefined domain characteristics within a relatively small category. Such sub-categorisation would allow for the identification of a small number of potentially strong and justifiable analogues from the category and may make further data collection more streamlined.

Attempts to expand the triazole domain to imidazoles demonstrate that, whilst there are many commonalities in the applicability domain, there are also differences. The data in Table 6 indicate key differences in the domains between the triazoles and imidazoles. Firstly, the imidazole category is broader than for the triazoles in terms of the physico-chemical properties, therefore there may be fewer high quality analogues. There is good evidence that the toxicodynamics of the triazoles and imidazoles may be similar, however the most striking difference is in the toxicokinetic domain. As well as some differences in distribution of the imidazoles, as predicted by PBPK modelling, the most important difference is in identities of the significant metabolites formed. Further assessment of the domain of the imidazoles with regard to the PBPK modelling and metabolism is required. This may be in terms of better optimisation of the models or other measurements, or alternatively a review of existing knowledge to increase understanding. Thus, the ability to organise the information regarding the category, as shown in Table 6, is a useful means to identify potential uncertainties in the read-across argument and justification. Overall, the extension of the read-across category to include the imidazoles should not be dismissed, but the information provided by the analysis of the domain suggests that this category will provide lower quality analogues. Whilst potentially of lower quality, the imidazole analogues may, however, provide additional weight of evidence - in this case in terms of consistency within toxicodynamics and NOAEL values.

*4.6 General Comments and Context*

Read-across for the filling of toxicity data gaps remains a beguiling, yet somehow elusive, proposition. From humble beginnings, in what was little more than an analogue approach to fill data gaps in high production volume categories [63], it has been transformed into something of great expectations and now can be viewed as a potentially complex data gathering and interpretation exercise [14, 16, 57]. The complexities of “modern” read-across should not, however, hide the basic need, and intrinsic difficulty, of justifying similarity, and identifying significant dissimilarity, of the data-poor and data-rich molecules. The justification and description of similarity is one of the fundamental means of demonstrating the validity of a read-across for a particular purpose and hence its potential acceptability. This study has made an attempt to formalise the overall domain of a read-across into chemistry, toxicodynamic and toxicokinetic domains.

The concept of an applicability domain for *in silico* toxicology is not, of course, novel, having been applied extensively in QSAR. Its application, by name at least, to grouping and read-across has been limited, although it is a concept that is required to assist in the acceptance of predictions. The strategy outlined in this publication is a step in the direction of a standardised approach, as it covers all the main points required by e.g. ECHA RAAF and in identification of uncertainties. It continues ongoing work that extends previous, predominantly chemistry-based domain analyses by inclusion of information on toxicodynamics and toxicokinetics (e.g. Rice et al. [64]). It seems obvious that there is a solution to the question of whether a single analogue is within the domain of another molecule, without making prior assumptions of what is acceptable in term of differences. It is highly likely that mechanistic similarity is a pre-requisite for analogues, followed by the acceptance that toxicokinetics should not increase adversity beyond what may be predictable. In addition, it should be considered that applicability domains, however defined, will be context-dependent and adaptable for different endpoints and chemistries. The presentation of the domains in this study should be suitable for adaptation to any scenario.

The definition of applicability domain for read-across proposed in this study goes much further than those forwarded in previous studies [4, 46, 47]. Data gathering exercises address all significant uncertainties and ECHA RAAF criteria, and allows for a (semi-) quantitative estimation of applicability domain to be made. To our knowledge, it is the first published study that formally organises all of this information and brings the concept of domain boundary criteria to read-across. In addition, it is a simple, flexible and adaptable approach based on the transparent presentation and interpretation of the properties and effects of molecules.

1. **Conclusions**

This study has brought together a wealth of knowledge and experience from the scientific community spanning at least a decade [65] in order to derive an approach that will assist in the definition of the domain of a read-across analogue. The approach taken in this study has demonstrated that the key features for read-across domain are based around the chemistry, toxicodynamics and toxicokinetics. These features can be characterised by a variety of information, including *in silico* predictions and experimental measurements. The approach provides an opportunity to undergo a problem formulation phase before assessing analogues to define suitable information to be compiled as well as acceptable boundaries for quantitative values. This approach brings in the open question of what is meant by similarity and what will be considered acceptable. It gives an opportunity to derive acceptability criteria that are appropriate for both the endpoint and chemistry being considered. In addition to the consideration of similarity, a key component of this approach is to consider the dissimilarity of analogues and categories alike, and whether any differences may bring uncertainty into the assessment.

**Declaration of Interest**

The authors declare no conflicts of interest.

**References**

[1] Madden JC, Enoch SJ, Paini A, Cronin MTD (2020) A review of *in silico* tools as alternatives to animal testing: Principles, resources and applications. *Alt. Lab. Anim. (ATLA)* 48: 146-172.

[2] Schultz TW, Cronin MTD (2017) Lessons learned from read-across case studies for repeated-dose toxicity. *Regulat. Toxicol. Pharmacol.* 88: 185-191.

[3] European Chemicals Agency (ECHA) (2017) *Read-Across Assessment Framework* (RAAF). ECHA-17-R-01-EN, ISBN 978-92-9495-758-0.

[4] Suter GW, Lizarraga LE (2022) Clearly weighing the evidence in read-across can improve assessments of data-poor chemicals. *Regul. Toxicol. Pharmacol*. 129: 105111.

[5] Schultz TW, Richarz A-N, Cronin MTD (2019) Assessing uncertainty in read-across: Questions to evaluate toxicity predictions based on knowledge gained from case studies. *Comput. Toxicol*. 9: 1-11.

[6] Rovida C, Barton-Maclaren T, Benfenati E, Caloni F, Chandrasekera PC, Chesne C, Cronin MTD, De Knecht Joop, Dietrich DR, Escher SE, Fitzpatrick S, Flannery B, Herzler M, Bennekou Hougaard S, Hubesch B, Kamp H, Kisitu J, Kleinstreuer N, Kovarich S, Leist M, Maertens A, Nugent K, Pallocca G, Pastor M, Patlewicz G, Pavan M, Presgrave O, Smirnova L, Schwarz M, Yamada T, Hartung T (2020) Internationalization of read-across as a validated New Approach Method (NAM) for regulatory toxicology. *ALTEX* 37: 579-606.

[7] Netzeva TI, Worth A, Aldenberg T, Benigni R, Cronin MT, Gramatica P, Jaworska JS, Kahn S, Klopman G, Marchant CA, Myatt G, Nikolova-Jeliazkova N, Patlewicz GY, Perkins R, Roberts D, Schultz T, Stanton DW, van de Sandt JJ, Tong W, Veith G, Yang C (2005). Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. The report and recommendations of ECVAM Workshop 52. *Alt. Lab. Anim. (ATLA)* 33: 155-173.

[8] OECD (Organisation for Economic Cooperation and Development), 2007. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships. ENV/JM/MONO(2007)2. OECD, Paris, pp. 154.

[9] Dimitrov S, Dimitrova G, Pavlov T, Dimitrova N, Patlewicz G, Niemela J, Mekenyan O (2005) A stepwise approach for defining the applicability domain of SAR and QSAR models. *J. Chem. Inf. Model*. 45: 839-849.

[10] Blackburn K, Stuard SB (2014). A framework to facilitate consistent characterization of read across uncertainty. *Regul Toxicol Pharmacol.* 68: 353-362.

[11] Escher SE, Kamp H, Bennekou SH, Bitsch A, Fisher C, Graepel R, Hengstler JG, Herzler M, Knight D, Leist M, Norinder U, Ouedraogo G, Pastor M, Stuard S, White A, Zdrazil B, van de Water B, Kroese D (2019) Towards grouping concepts based on new approach methodologies in chemical hazard assessment: the read-across approach of the EU-ToxRisk project. *Arch. Toxicol.* 93: 3643-3667.

[12] Schultz TW, Amcoff P, Berggren E, Gautier F, Klaric M, Knight DJ, Mahony C, Schwarz M, White A, Cronin MTD (2015) A strategy for structuring and reporting a read-across prediction of toxicity. *Regul. Toxicol. Pharmacol*. 72: 586–601.

[13] Worth AP (2010) The role of QSAR methodology in the regulatory assessment of chemicals. In Puzyn T, Lesczynski J, Cronin MTD (Eds.) *Recent Advances in QSAR Studies: Methods and Applications*. Springer, Dordrecht, The Netherlands, pp. 367-382.

[14] Pestana CB, Firman, JW, Cronin MTD (2021) Incorporating lines of evidence from New Approach Methodologies (NAMs) to reduce uncertainties in a category based read-across: A case study for repeated dose toxicity. *Regulat. Toxicol. Pharmacol*. 120: 104855.

[15] Nakagawa S, Okamoto M, Yoshihara K, Nukada Y, Morita O (2021) Grouping of chemicals based on the potential mechanisms of hepatotoxicity of naphthalene and structurally similar chemicals using in vitro testing for read-across and its validation. *Regul. Toxicol. Pharmacol*. 121: 104874.

[16] Escher SE, Aguayo-Orozco A, Benfenati E, Bitsch A, Braunbeck T, Brotzmann K, Bois F, van der Burg B, Castel J, Exner T, Gadaleta D, Gardner I, Goldmann D, Hatley O, Golbamaki N, Graepel R, Jennings P, Limonciel A, Long A, Maclennan R, Mombelli E, Norinder U, Jain S, Santos Capinha L, Taboureau OT, Tolosa L, Vrijenhoek NG, van Vugt-Lussenburg BMA, Walker P, van de Water B, Wehr M, White A, Zdrazil B, Fisher C (2022) Integrate mechanistic evidence from new approach methodologies (NAMs) into a read-across assessment to characterise trends in shared mode of action. *Toxicol. in Vitro* 79: 105269.

[17] Laroche C, Aggarwal M, Bender H, Benndorf P, Birke B, Crozier J, Dal Negro G, De Gaetano F, Desaintes C, Gardner I, Hubesch B, Irizar A, John D, Kumar V, Lostia A, Manou I, Monshouwer M, Müller BP, Paini A, Reid K, Rowan T, Sachana M, Schutte M, Stirling C, Taalman R, van Aerts L, Weissenhorn R, Sauer UG (2018) Finding synergies for 3Rs – Toxicokinetics and read-across: Report from an EPAA partners’ Forum. *Reg. Tox. Pharm.* 99: 5-21.

[18] Cronin MTD, Richarz A-N (2017) Relationship between Adverse Outcome Pathways and chemistry-based *in silico* models to predict toxicity. *Appl. In Vitro Toxicol.* 3: 286-297.

[19] European Food Safety Authority (EFSA) (2008) Conclusion on the peer review of the pesticide risk assessment of the active substance tetraconazole. *EFSA J* 6, 152.

[20] United States Environmental Protection Agency (US EPA) (Federal Register): June 30, 1999 (Volume 64, Number 125 Available from: <https://www.govinfo.gov/content/pkg/FR-1999-06-30/pdf/99-16545.pdf>

[21] Joint FAO/WHO Meeting on Pesticide Residues, (JMPR) (1998) Bitertanol. JMPR Evaluations 1998 Part II Toxicological. Available from: http://www.inchem.org/documents/jmpr/jmpmono/v098pr04.htm

[22] European Food Safety Authority (EFSA), (2010) Conclusion on the peer review of the pesticide risk assessment of the active substance cyproconazole. *EFSA J* 8, 1897.

[23] European Food Safety Authority (EFSA) (2011) Conclusion on the peer review of the pesticide risk assessment of the active substance difenoconazole. *EFSA J* 9, 1967.

[24] European Food Safety Authority (EFSA) (2008) Conclusion on the peer review of the pesticide risk assessment of the active substance epoxiconazole. *EFSA J* 6, 138.

[25] European Food Safety Authority (EFSA) (2010) Conclusion on the peer review of the pesticide risk assessment of the active substance fenbuconazole. *EFSA J* 8, 1558.

[26] Joint FAO/WHO Meeting on Pesticide Residues, (JMPR) (2014) Report of the Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and the WHO Core Assessment Group on Pesticide Residues, Rome, Italy, 16-25 September 2014.

[27] European Food Safety Authority (EFSA) (2010) Conclusion on the peer review of the pesticide risk assessment of the active substance paclobutrazol. *EFSA J* 8, 1876.

[28] European Food Safety Authority (EFSA) (2008) Conclusion on the peer review of the pesticide risk assessment of the active substance penconazole. *EFSA J* 6, 175.

[29] European Chemicals Agency (ECHA) (2016) *Committee for Risk Assessment. Opinion proposing harmonised classification and labelling at EU level of propiconazole*. 2016, 1-51, CLH-O-0000001412-86-139/F.

[30] Joint FAO/WHO Meeting on Pesticide Residues, (JMPR) (2008) Prothioconazole. Plant Protection Paper 232. Available from: <http://www.fao.org/fileadmin/templates/agphome/documents/Pests_Pesticides/JMPR/Report08/Prothioconazole.pdf>.

[31] European Food Safety Authority (EFSA) (2014) Conclusion on the peer review of the pesticide risk assessment of the active substance tebuconazole. *EFSA J* 12, 3485.

[32] European Food Safety Authority (EFSA) (2008) Conclusion on the peer review of the pesticide risk assessment of the active substance triadimenol. *EFSA J* 6, 177.

[33] Food and Agriculture Organization of the United Nations (FAO) (2011) FAO Specifications and Evaluations for Agricultural Pesticides: Triadimefon. Available from: <http://www.fao.org/fileadmin/templates/agphome/documents/Pests_Pesticides/Specs/Triadimefon_2011.pdf>.

[34] European Food Safety Authority (EFSA) (2010) Conclusion on the peer review of the pesticide risk assessment of the active substance imazalil. *EFSA J* 8, 1526.

[35] European Food Safety Authority (EFSA) (2011) Conclusion on the peer review of the pesticide risk assessment of the active substance prochloraz. *EFSA J* 9, 2323.

[36] European Food Safety Authority (EFSA) (2009) Conclusion on the peer review of the pesticide risk assessment of the active substance triflumizole. *EFSA J* 7, 1415.

[37] Bell S, Abedini J, Ceger P, Chang X, Cook B, Karmaus AL, Lea I, Mansouri K, Phillips J, McAfee E, Rai R, Rooney J, Sprankle C, Tandon A, Allen D, Casey W, Kleinstreuer N (2020) An integrated chemical environment with tools for chemical safety testing. *Toxicol. in Vitro* 67: 104916.

[38] Ball N, Bartels M, Budinsky R, Klapacz J, Hays S, Kirman C, Patlewicz G (2014) The challenge of using read-across within the EU REACH regulatory framework; how much uncertainty is too much? Dipropylene glycol methyl ether acetate, an exemplary case study. *Regul. Toxicol. Pharmacol*. 68: 212-221.

[39] Grimm FA, House JS, Wilson MR, Sirenko O, Iwata Y, Wright FA, Ball N, Rusyn I (2019) Multi-dimensional *in vitro* bioactivity profiling for grouping of glycol ethers. *Regul. Toxicol. Pharmacol*. 101: 91-102.

[40] House JS, Grimm FA, Klaren WD, Dalzell A, Kuchi S, Zhang S-D, Lenz K, Boogaard PJ, Ketelslegers HB, Gant TW, Wright FA, Rusyn I (2021) Grouping of UVCB substances with new approach methodologies (NAMs) data. *ALTEX* 38: 123-137.

[41] van Ravenzwaay B, Sperber S, Lemke O, Fabian E, Faulhammer F, Kamp H, Mellert W, Strauss V, Strigun V, Peter E, Spitzer M, Walk T (2016) Metabolomics as read-across tool: A case study with phenoxy herbicides. *Regul. Toxicol. Pharmacol.* 81: 288-304.

[42] Boberg J, Dybdahl M, Petersen A, Hass U, Svingen T, Vinggaard AM (2019) A pragmatic approach for human risk assessment of chemical mixtures. *Curr. Opin. Toxicol*. 15: 1-7.

[43] Dimitrov SD, Georgieva DG, Pavlov TS, Karakolev YH, Karamertzanis PG, Rasenberg M, Mekenyan OG (2015) UVCB substances: Methodology for structural description and application to fate and hazard assessment. *Environ. Toxicol. Chem*. 34: 2450-2462.

[44] Kutsarova SS, Yordanova DG, Karakolev YH, Stoeva S, Comber M, Hughes CB, Vaiopoulou E, Dimitrov SD, Mekenyan OG (2019) UVCB substances II: Development of an endpoint-nonspecific procedure for selection of computationally generated representative constituents. *Environ. Toxicol. Chem*. 38: 682-694.

[45] Yordanova DG, Patterson TJ, North CM, Camenzuli L, Chapkanov AS, Pavlov TS, Mekenyan OG (2021) Selection of representative constituents for Unknown, Variable, Complex, or Biological origin substance assessment based on hierarchical clustering. *Environ. Toxicol. Chem*. 40: 3205-3218.

[46] Benfenati E, Chaudhry Q, Gini G, Dorne JL (2019) Integrating *in silico* models and read-across methods for predicting toxicity of chemicals: A step-wise strategy. *Environ. Int*. 131: 105060.

[47] Gadaleta D, Golbamaki BA, Lavado GJ, Roncaglioni A, Benfenati E (2020) Automated integration of structural, biological and metabolic similarities to sustain read-across. *ALTEX* 37: 469-481.

[48] Guo Y, Zhao L, Zhang X, Zhu H (2019) Using a hybrid read-across method to evaluate chemical toxicity based on chemical structure and biological data. *Ecotoxicol. Environ. Saf*. 178: 178-187.

[49] Gilmour N, Kern PS, Alepee N, Boisleve F, Bury D, Clouet E, Hirota M, Hoffmann S, Kuehnl J, Lalko JF, Mewes K, Miyazawa M, Nishida H, Osmani A, Petersohn D, Sekine S, van Vliet E, Klaric M (2020) Development of a next generation risk assessment framework for the evaluation of skin sensitisation of cosmetic ingredients. *Regul. Toxicol. Pharmacol*. 116: 104721.

[50] Enoch SJ, Cronin MTD, Schultz TW, Madden JC (2008) Quantitative and mechanistic read across for predicting the skin sensitization potential of alkenes acting via Michael addition. *Chem. Res. Toxicol.* 21: 513-520.

[51] Enoch SJ, Ellison CM, Schultz TW, Cronin MTD (2011) A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity. *Crit. Rev. Toxicol*. 41: 783-802.

[52] Koleva YK, Madden JC, Cronin MTD (2008) Formation of categories from structure-activity relationships to allow read-across for risk assessment: Toxicity of ,-unsaturated carbonyl compounds. *Chem. Res. Toxicol.* 21: 2300-2312.

[53] Marx-Stoelting, P., Knebel, C., Braeuning, A., 2020. The connection of azole fungicides with xeno-sensing nuclear receptors, drug metabolism and hepatotoxicity. Cells 9: 1192.

[54] Mellor CL, Marchese Robinson RL, Benigni R, Ebbrell D, Enoch SJ, Firman JW, Madden JC, Pawar G, Yang C, Cronin MTD (2019) Molecular fingerprint-derived similarity measures for toxicological read-across: Recommendations for optimal use. *Regul. Toxicol. Pharmacol*. 101: 121-134.

[55] Rovida C, Escher SE, Herzler M, Bennekou SH, Kamp H, Kroese DE, Maslankiewicz L, Moné MJ, Patlewicz G, Sipes N, van Aerts L, White A, Yamada T, van de Water B (2021) NAM-supported read-across: From case studies to regulatory guidance in safety assessment. *ALTEX* 38: 140-150.

[56] Schultz TW, Przybylak KR, Richarz A-N, Mellor CL, Escher SE, Bradbury SP, Cronin MTD (2017) Read-across for 90-day rat oral repeated-dose toxicity for selected n-alkanols: A case study. *Comput. Toxicol*. 2: 12-19.

[57] Bury D, Alexander-White C, Clewell HJ 3rd, Cronin M, Desprez B, Detroyer A, Efremenko A, Firman J, Hack E, Hewitt NJ, Kenna G, Klaric M, Lester C, Mahony C, Ouedraogo G, Paini A, Schepky A (2021) New framework for a non-animal approach adequately assures the safety of cosmetic ingredients - A case study on caffeine. *Regul. Toxicol. Pharmacol*. 123: 104931.

[58] Firman JW, Punt A Cronin MTD, Boobis AR, Wilks MF, Hepburn, PA, Thiel, A, Fussell KC (2021) Exploring the potential of ToxCast data in supporting read-across for evaluation of food chemical safety. *Chem. Res. Toxicol.* 34: 300-312.

[59] Firman JW, Cronin MTD, Rowe PH, Semenova EP, Doe JE (2022) The use of Bayesian methodology in the development and validation of a tiered assessment approach towards prediction of rat acute oral toxicity. *Arch. Toxicol*. In press.

[60] Steinmetz FP, Madden JC, Cronin MTD (2015) Data quality in the human and environmental health sciences: using statistical confidence scoring to improve QSAR/QSPR modelling. J. Chem. Inf. Model 55: 1739-1746.

[61] Enoch SJ, Cronin MTD, Madden JC, Hewitt M (2009) Formation of structural categories to allow for read-across for teratogenicity. *QSAR and Comb. Sci*. 28: 696-708.

[62] Enoch SJ, Hasarova Z, Cronin MTD, Bridgewood K, Rao S, Kluxen FM, Frericks M (2022) Sub-structure-based category formation for the prioritisation of genotoxicity hazard assessment for pesticide residues: Sulphonyl ureas. *Regul. Toxicol. Pharmacol*. 129: 105115.

[63] Hanway RH, Evans PF (2000) Read-across of toxicological data in the notification of new chemicals. *Toxicol. Lett.* 116 (Suppl. 1): 61

[64] Rice PA, Cooper J, Koh-Fallet SE, Kabadi SV (2021) Comparative analysis of the physicochemical, toxicokinetic, and toxicological properties of ether-PFAS. *Toxicol. Appl. Pharmacol*. 422: 115531.

[65] Patlewicz G, Cronin MTD, Helman G, Lambert JC, Lizarraga LE, Shah I (2018) Navigating through the minefield of read-across frameworks: A commentary perspective. *Comput. Toxicol*. 6: 39-54.

Supplementary Information

Table S1. Excel file containing full data matrix for the domains of the triazole and imidazole categories.