



## Research Article

# Automated Integration of Structural, Biological and Metabolic Similarities to Improve Read-Across

Domenico Gadaleta, Azadi Golbamaki Bakhtyari, Giovanna J. Lavado, Alessandra Roncaglioni and Emilio Benfenati

Laboratory of Environmental Chemistry and Toxicology, Department of Environmental Health Sciences, Istituto di Ricerche Farmacologiche Negri IRCCS, Milan, Italy

### Abstract

Read-across (RAX) is a popular data gap-filling technique that uses category and analogue approaches to predict toxicological endpoints for a target. Despite its increasing relevance, RAX relies on human expert judgement and lacks a reproducible and automated protocol. It also relies only on structural similarity for identifying the analogues, while other aspects are often neglected. In this paper, we propose an automated procedure for the selection of analogues for data gap-filling. Analogues were identified by a decision algorithm that integrates three similarity metrics, each considering different toxicologically relevant aspects (i.e., structural, biological and metabolic similarity). Structural filters based on the presence of maximum common substructures (MCS) and common functional groups (FGs) were applied to narrow the chemical space for the analogue search. The procedure has been implemented as a workflow in KNIME and is freely available. The workflow provides informative tabular and graphical outputs to support toxicologists and risk assessors in drawing conclusions based on the RAX approach. The procedure has been validated for its predictive power on two datasets related to high-tier *in vivo* toxicological endpoints, i.e., human hepatotoxicity and drug-induced liver injury (DILI). The validation results gave good accuracy values (i.e., up to 0.79 for the binary hepatotoxicity classification and up to 0.67 for the three-class DILI classification) that were higher than those returned by RAX based on the sole use of structural similarity. Results confirmed the suitability of the procedure as a source of data to support regulatory decision-making.

## 1 Introduction

In the last 15 years, regulations in the field of chemical safety assessment have changed in that toxicological information for a large number of chemicals needs to be gathered prior to manufacture or import into the EU (EC, 2006). *In vivo* testing to meet these information requirements is clearly not feasible due to time, costs and the need to sacrifice an unacceptable number of animals. The EU Registration, Evaluation, Authorisation and restriction of CHemicals (REACH) calls for the use of non-testing approaches to be used in the assessment of chemical substances while vertebrate animal testing should be seen as a last resort (ECHA, 2014).

Among non-testing methods, read-across (RAX) has proven to be an effective and widely used approach to provide toxicological information without the need to undertake animal testing according to the 3Rs principle (Replace, Reduce, Refine) (Russell and

Burch, 1959), with clear benefits in terms of money, time and animal savings. RAX is a data gap-filling technique used to predict unknown toxicological endpoints for a chemical substance (target) by using the same endpoint information from one or more chemicals that are highly similar to the target (analogue(s)) (Patlewicz et al., 2013; Pradeep et al., 2017; OECD, 2014; ECHA, 2008). The first step of RAX consists in identifying potential analogue(s) that may serve to fill the target's toxicological data gaps. This can be done using quantitative metrics to evaluate the similarity between the target and potential analogue(s). The following steps comprise gathering relevant data to determine analogue(s) suitability for the RAX. The presence of functional group(s) (e.g., aldehyde, epoxide, ester, specific metal ion) shared with the target, common constituents or chemical classes, similar carbon range numbers or the likelihood of common precursors and/or breakdown products are typical considerations when evaluating

Received February 28, 2020; Accepted May 7, 2020;  
Epub May 8, 2020; © The Authors, 2020.

ALTEX 37(3), 469-481. doi:10.14573/altex.2002281

Correspondence: Domenico Gadaleta, PhD  
Laboratory of Environmental Chemistry and Toxicology  
Department of Environmental Health Sciences  
Istituto di Ricerche Farmacologiche Mario Negri IRCCS  
Via Mario Negri 2, 20156 Milan, Italy  
(domenico.gadaleta@marionegri.it)

This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is appropriately cited.



analogue(s) in the final RAX reasoning. Based on the retrieved data, one can assess the adequacy of the analogue(s) and use the most suitable ones to fill the data gaps for the target (Patlewicz et al., 2013, 2015; OECD, 2014; ECHA, 2008).

Two approaches for RAX exist. The “analogue approach” is based on a small number of structurally similar substances (usually a single analogue), while the “category approach” relies on a larger number of analogues included in the same chemical classes or showing some kind of trend in structures (e.g., an increasing number of carbon chain lengths).

Among other non-testing methods, RAX is applied in various regulatory programs such as the OECD High Production Volume Programme (Bishop et al., 2012) and REACH. Indeed, REACH encourages using a category/analogue approach (ECHA, 2008) to address regulatory requirements for chemical substances to the point that RAX has been used in up to 75% of analyzed registration dossiers for at least one endpoint (ECHA, 2014). This percentage is much higher than those related to other non-testing methods, e.g., quantitative structure-activity relationships (QSARs). This is especially true for high-tier toxicological endpoints (such as reproductive and repeated dose toxicities) that refer to a wide range of adverse effects on different target organs and tissues. RAX proved to be more intuitive for regulators and more appropriate than QSAR/*in vitro* in appreciating the composite nature of complex *in vivo* endpoints (Patlewicz et al., 2013).

Despite its increasing use for regulatory purposes, RAX remains largely subjective and relies on human expert judgement for both analogue(s) selection and data interpretation (Patlewicz et al., 2017). This prompted the European Chemicals Agency (ECHA) to publish a Read-Across Assessment Framework (RAAF) (ECHA, 2015), which complemented already existing regulatory technical guidance from ECHA (2008) and OECD (2014). These documents exemplify all relevant aspects that should be evaluated to assure the acceptability of RAX proposals included in REACH registrations. Despite this, the general lack of a well-defined and systematic decision workflow still hampers a consistent application of RAX (Patlewicz et al., 2017). Another limitation is that RAX traditionally relies on the chemical similarity principle. Indeed, REACH states that chemical structure should be the starting point for the definition of any category/analogue approach (ECHA, 2008). However, the accuracy of predictions based exclusively on structural similarity is often inadequate in handling complex mechanisms of toxicity (Ball et al., 2016). A pitfall of the sole use of structural similarity is, for example, the existence of activity cliffs, i.e., a group of compounds may have high structural similarity but unexpectedly high activity (or property) differences (Cruz-Monteagudo et al., 2014). Ideally, the overall RAX should be based on a weight of evidence (WoE) assessment of many different pieces of information, i.e., not only structures, but also metabolism, biology and physicochemical properties should be considered to substantiate the similarity between target and analogue(s) (Patlewicz et al., 2014, 2015).

These aspects highlight the need to define an automated RAX framework able to consider both structural and biological profiles of chemicals, still keeping the entire process transparent and easy to understand for regulators (Low et al., 2013). Several attempts to exploit hazard (Luechtefeld et al., 2018) or high throughput screening (HTS) *in vitro* data to include biological similarity in RAX reasoning have been reported (Petrone et al., 2012; Russo et al., 2017; Shah et al., 2016; Grimm et al., 2016; Low et al., 2013). Large databases exist that include thousands of HTS assays for a broad range of chemical substances. ToxCast by US EPA<sup>1</sup> and PubChem<sup>2</sup> are well-known examples. Biological phenotyping derived by combining results from HTS assays can serve to characterize the biological profile of target and analogue(s) in terms of a fingerprint that can be used to determine a quantitative biological similarity (Patlewicz et al., 2014). The increased availability of cheminformatics tools represents another possible improvement of traditional RAX. Automated tools exist that can be applied to analogue(s) identification (e.g., QSAR-Toolbox<sup>3</sup>), data retrieval or, if no experimental data are available, for the prediction of physicochemical, toxicokinetic and metabolic parameters for substances under study (Ball et al., 2016). Moreover, the same tools can be used to automate the evaluation of similarity between target and analogue(s) and other steps of the RAX, facilitating the final expert judgement.

In this study, we present a novel automated workflow for analogue(s) selection for RAX based on a WoE approach that systematically computes and combines three similarity metrics between target and potential analogue(s). Given a target, the workflow automatically lists potential analogue(s) that are selected independently based on three similarity criteria (i.e., structural, biological, and metabolic similarity). A large collection of data retrieved from on-line databases or calculated with cheminformatics tools is provided in the final output of the workflow to aid experts in RAX reasoning. Finally, compound(s) included in multiple similarity lists (e.g., structural and biological similarity lists) are suggested as the most suitable analogue(s), and their activity is used to infer the activity of the target chemical. Finally, we present examples to evaluate the suitability of the described procedure to predict high-tier endpoints for the chemicals.

The entire workflow is implemented in KNIME (version 3.4) (Berthold et al., 2008) and made freely available to the scientific community and to regulators to aid expert reasoning and decision-making.

## 2 Methods

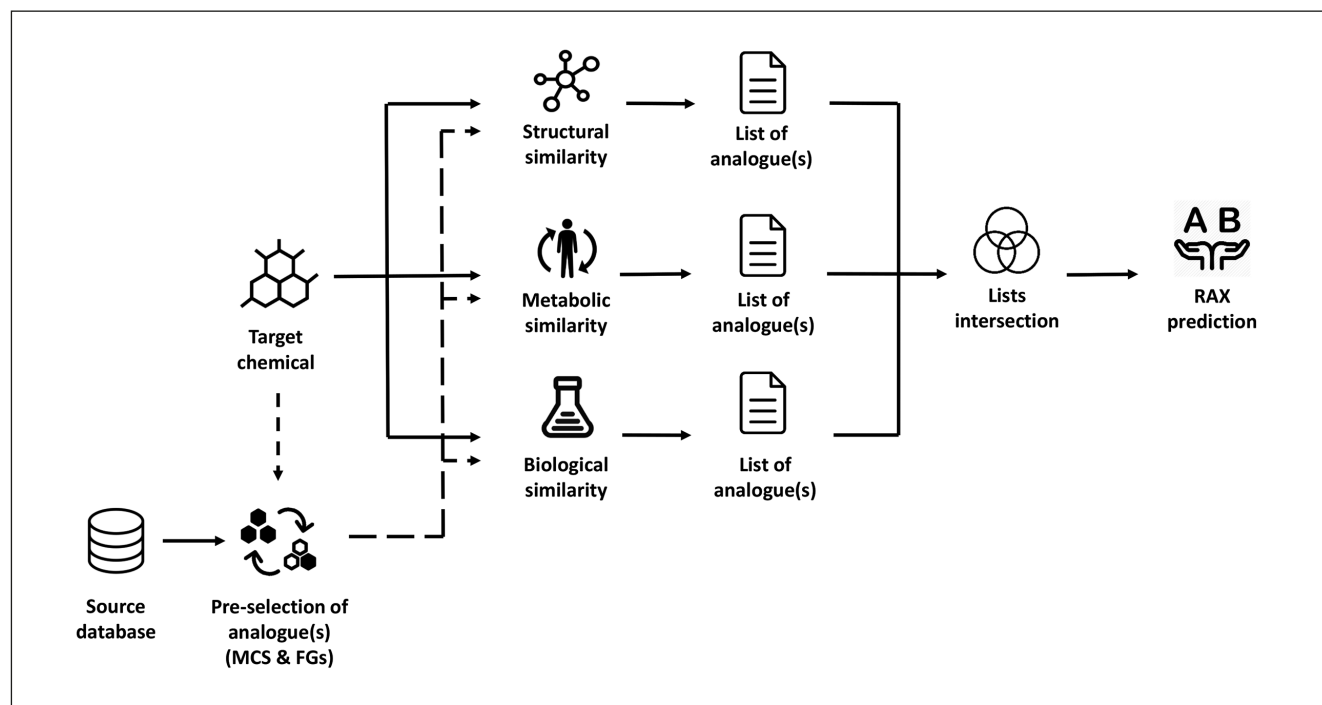
### 2.1 RAX workflow

An automated procedure to compute similarities between chemicals was implemented as a KNIME workflow (Berthold et al., 2008). The workflow accepts SMILES notation of the target chemical as input and is connected to a pre-loaded dataset in

<sup>1</sup> <https://www.epa.gov/chemical-research/toxicity-forecasting>

<sup>2</sup> <https://pubchem.ncbi.nlm.nih.gov/>

<sup>3</sup> <https://qsartoolbox.org/>



**Fig. 1: Conceptual scheme of the RAX workflow**

Similarities between the target and chemicals from the source dataset are computed to produce three independent lists of analogue(s). The activity of chemicals included in one (or multiple) list(s) are used to infer the RAX prediction. A preliminary selection of source chemicals can be applied based on the presence of maximum common substructures (MCS) and common functional group(s) (FG) with the target.

which the analogue(s) search is performed (source dataset). Three separate lists of possible analogue(s) are retrieved from the source dataset based on three different and independent methods to compute similarity, i.e., 1) structural (StrS), 2) metabolic (MS) and 3) biological similarity (BS). Chemicals in the source dataset are ranked for each of the three similarity metrics, and then three different lists of top-ranked similar compounds are returned as output. The number of analogue(s) returned in each list can be customized by the user in order to meet specific endpoint requirements (e.g., increase the MS analogues if metabolism is known to be particularly relevant). By default, up to 10 analogues are retrieved based on StrS, as a primary requisite for any RAX prediction (ECHA, 2015). Up to 5 analogues each are retrieved based on MS and BS. Figure 1 shows the logical scheme of the RAX workflow, while a more detailed depiction of the KNIME implementation is included in Figure S1<sup>4</sup>. The KNIME workflow is freely available for download<sup>5</sup>. The user's guide is available at the same GitHub link and in the supporting information<sup>6</sup>.

The workflow also offers rich tabular and graphical (pie-plot) outputs including relevant chemical and toxicological information

for both the target and the analogue(s) (e.g., metabolites, number of biological assays with the same outcome, common functional groups) that can be used by a toxicologist to support the RAX process and help regulators in decision-making. The pie plots that graphically describe RAX results are created using the bokeh library in Python<sup>7</sup>. Some examples are shown in Figure S2<sup>4</sup>.

#### Structural similarity

StrS was based on MACCS fingerprints (166 bits) (Durant et al., 2002) that were calculated for both target and analogue(s) with the KNIME implementation of the CDK toolkit<sup>8</sup>. MACCS keys are commonly used to calculate structural similarity (Maggiore et al., 2014; Sánchez-Cruz and Medina-Franco, 2018). Bits codify for the presence of a given substructure in the molecule (e.g., a phenyl ring or a functional group). They are useful to disclose analogies in terms of the chemical features relevant for biological and toxicological activity. StrS was computed by means of Tanimoto coefficient (Willett et al., 1998). In this regard, chemicals are identified as similar if they share a high number of biologically relevant moieties.

<sup>4</sup> doi:10.14573/altex.2002281s1

<sup>5</sup> [https://github.com/DGadaleta88/RAX\\_tool](https://github.com/DGadaleta88/RAX_tool)

<sup>6</sup> doi:10.14573/altex.2002281s2

<sup>7</sup> <http://bokeh.pydata.org/>

<sup>8</sup> <https://cdk.github.io/>



### Metabolic similarity

“SyGMA metabolite” KNIME implementation was used to simulate metabolites of the target and the analogue(s). SyGMA (Systematic Generation of possible Metabolites) is a freely available tool to simulate metabolism of chemicals. Metabolic rules implemented in SyGMA are derived from combining expert knowledge and empirical analysis of proprietary data (i.e., MDL dataset) and cover 70% of all known human metabolic reactions. Predictions made by SyGMA are associated with an empirical probability score that identifies more likely metabolic routes and reduces the number of false positives that are often generated by tools based only on expert rules (Ridder and Wagener, 2008).

The tool allows calculation both of specific metabolites and the type of metabolic pathways in humans that the parent undergoes to generate the metabolite. In addition, SyGMA provides a free Python code and a KNIME implementation that allows easy integration into the workflow presented here.

Only one cycle of Phase I metabolism was considered for the present simulation. AMS score was calculated based on the number of common and exclusive metabolic pathways of the two compared chemicals.

$$MS(A,B) = \frac{P(A,B)}{P(A,B)+P(A)+P(B)}$$

$P(A,B)$  are metabolic pathways that are shared by compounds A and B, while  $P(A)$  and  $P(B)$  are metabolic pathways that are used by only one of the two compounds being compared.

The presence of shared, structurally identical metabolites and/or parent compounds between the two compared chemicals is also reported in the final output of the workflow.

### Biological similarity

BS calculation was based on HTS assays from PubChem. Assays were used to compile biological binary fingerprints to compare target and analogue(s), with each bit of the fingerprint codifying the outcome of a specific assay. The REST version of the PUG (Power User Gateway) interface for accessing PubChem data was implemented in KNIME to automatically retrieve assay information for target and analogue(s)<sup>9</sup>. In particular, “Active” or “Probe” outcomes were flagged with 1, while “Inactive” was flagged with 0. “Inconclusive” and “Unspecified” outcomes were ignored. Duplicate assays resulting in different outcomes for the same chemical (e.g., “Active” and “Inactive”) were considered “Inconclusive”.

BS was calculated as proposed by Russo et al. (2017):

$$BS(A,B) = \frac{|A_a \cap B_a| + |A_i \cap B_i| \cdot w}{|A_a \cap B_a| + |A_i \cap B_i| \cdot w + |A_a \cap B_i| + |A_i \cap B_a|}$$

$A_a$  and  $A_i$  are active and inactive assays for the target A, while  $B_a$  and  $B_i$  are active and inactive assays for the analogue B.  $\cap$  indicates assays in common between the two compounds under in-

vestigation, while  $w$  is a weight assigned to common inactive assays that accounts for the ratio of active to inactive bits in the target compound’s biological fingerprint:

$$w = \frac{A_a}{A_i}$$

The weight ranges from 0 to 1, giving to inactive data a fraction of the weight of active data. This variable was adopted by Russo et al. (2017) because biosimilarity should rely on active data more than on inactive data although there is a far higher number of inactive assays reported in public HTS repositories compared to the active ones. Given the unbalanced nature of HTS data, values lower than 1 are usually assigned to  $w$ , and thus a lower weight is assigned to inactive assays compared to the active ones.

A confidence  $Conf_{BS(A,B)}$  index was assigned to BS to account for missing assays. The equation proposed by Russo et al. (2017) was modified in order to normalize the confidence in a 0-1 range as follows:

$$Conf_{BS(A,B)} = \frac{|A_a \cap B_a| + |A_i \cap B_i| \cdot w + |A_a \cap B_i| + |A_i \cap B_a|}{A_a + A_i \cdot w}$$

A lower weight was given to assays that are negative for both compounds as explained above.

A final weighted BS ( $BS_{weight(A,B)}$ ) is calculated as the product of  $B(A,B)$  and  $Conf_{BS(A,B)}$ , which accounts for both the degree of similarity of the two compared biological fingerprints and the number of bits on which the comparison is based:

$$BS_{weight(A,B)} = B(A,B) \cdot Conf_{BS(A,B)}$$

Assays having no active compounds in the source dataset were not considered in the final fingerprints. In the same way, if either the target or the analogue had no positive bits in their fingerprints, the final similarity value was imposed to be equal to 0 to avoid large similarity values resulting from the exclusive comparison of negative assays.

### Structural and functional group(s) filters

Pre-filters were implemented in the KNIME workflow to limit the search for potential analogue(s) to chemicals sharing relevant common structural features with the target. The user can decide to activate or deactivate each filter by modifying settings. This becomes particularly relevant when studying endpoints that are known to be related to a well-defined substructure and/or chemical category. Two independent filters were implemented:

1) *Maximum common substructures (MCS)*. Chemicals in the source dataset are filtered based on the presence of a MCS with respect to the target. MCS is calculated using the RDKit MCS code<sup>10</sup> implemented in KNIME. If the size of the MCS (i.e., the number of atoms in the common structural moiety) is greater than a given percentage of the size (i.e., number of at-

<sup>9</sup> <https://pubchemdocs.ncbi.nlm.nih.gov/pug-rest>

<sup>10</sup> <https://www.rdkit.org/>



**Tab. 1: Source datasets implemented in the RAX workflow**

For each dataset, the categories (C) and the number of chemicals included in each category are indicated. The range of (and the mean) values of relevant physico-chemical properties for the two datasets are reported. Properties were calculated with the Chemistry Developmental Kit (CDK) "Molecular Properties" node available in KNIME<sup>a</sup>.

Dataset	Categories	Total	C1	C2	C3	logP	TPSA	MW
DILIRank	No-DILI (C1) Less-DILI (C2) Most-DILI (C3)	691	252	260	179	-2.06 to 14.44 (2.8)	0.0 to 3115.35 (126.54)	60.02 to 7049.04 (431.39)
ToxRef	Non-Hepatotox (C1) Hepatotox (C2)	663	454	209	-	0.03 to 4.76 (2.37)	0.0 to 474.9 (58.9)	42.02 to 972.32 (272.57)

<sup>a</sup> <https://cdk.github.io/>

oms) of both the target and the analogue, the analogue is retained for the following searches. The default value is set at 50%, which the user can manually customize in the settings.

- 2) *Functional groups (FG)*. Chemicals in the source dataset are filtered based on the presence of chemical functional groups in common with the target. The presence of 22 functional groups codified as SMARTS is verified for both the target and the possible analogue. Those functional groups codify for general chemical classes/categories (e.g., carboxylic acids or amine) and are relevant to describe the reactivity of chemicals. The collection of SMARTS codifying functional groups was retrieved and adapted from RDKit Functional Group Filter KNIME node and is available in Table S1<sup>4</sup>. If the percentage of common functional groups compared to the number of those present in the target is greater than a given percentage, the analogue was considered for the following searches. A threshold of 65% is the default value; this can be customized by the user.

### Integration of similarities

The three independent lists of analogue(s) are integrated to identify a narrower range of chemicals for target data gap-filling. If a given chemical is found in multiple similarity lists, it is considered a more suitable analogue compared to those found in a lower number of lists. For example, if an analogue is found in all three similarity lists, a maximum suitability is assigned to that compound.

The prediction of activity for the target is made by averaging activities of analogue(s) included in the similarity lists. Chemicals included in multiple lists are prioritized (e.g., chemicals included in at least two out of three lists, or only those included in all three lists). The number of analogues used for prediction may vary based on the user's decision and the degree of overlap of the similarity lists. It can be up to 20 (i.e., 10 from StrS, 5 from BS, 5 from MS), but the number is often reduced by the presence of chemicals that appear in multiple similarity lists or by the application of a threshold for the selection of analogues (e.g., using only analogues that are included in multiple similarity lists).

## 2.2 Source datasets

Two datasets (source datasets) including toxicological data related to high-tier *in vivo* endpoints were compiled from the literature and used to validate the proposed methodology:

- 1) The DILIRank dataset (Chen et al., 2011, 2016) is a collection of 1,036 FDA-approved drugs divided into four classes according to their potential for causing drug-induced liver injury (DILI). The DILI classification is the result of analysis of FDA-approved drug labeling documents and literature. Drugs are classified into three groups of DILI concern (Most-, Less- and No-DILI concern), and one group (Ambiguous-DILI-concern) with undetermined causality. For the present work, the DILIRank dataset<sup>11</sup> was downloaded. Compounds with Ambiguous-DILI-concern label were discarded.
- 2) The Liu et al. (2015) dataset includes data for 667 compounds retrieved from ToxRef<sup>12</sup>. These data refer to three major groups of hepatic histopathological effects, i.e., hypertrophy, injury, and proliferative lesions. In the present work, if at least one of the three histopathological effects was positive, the chemical was classified as hepatotoxic, otherwise as non-hepatotoxic.

Chemical data included in each dataset were curated by means of a semi-automated in-house procedure described by Gadaleta et al. (2018). The procedure addresses the identification and removal of inorganic and organometallic compounds and mixtures, the neutralization of salts, and the removal of duplicates (also checking for tautomeric forms). Finally, the resulting SMILES are converted to a standardized format. The procedure is implemented in KNIME and is freely available for download<sup>13</sup>. Entries with unspecified SMILES and compounds with ambiguous classification were removed.

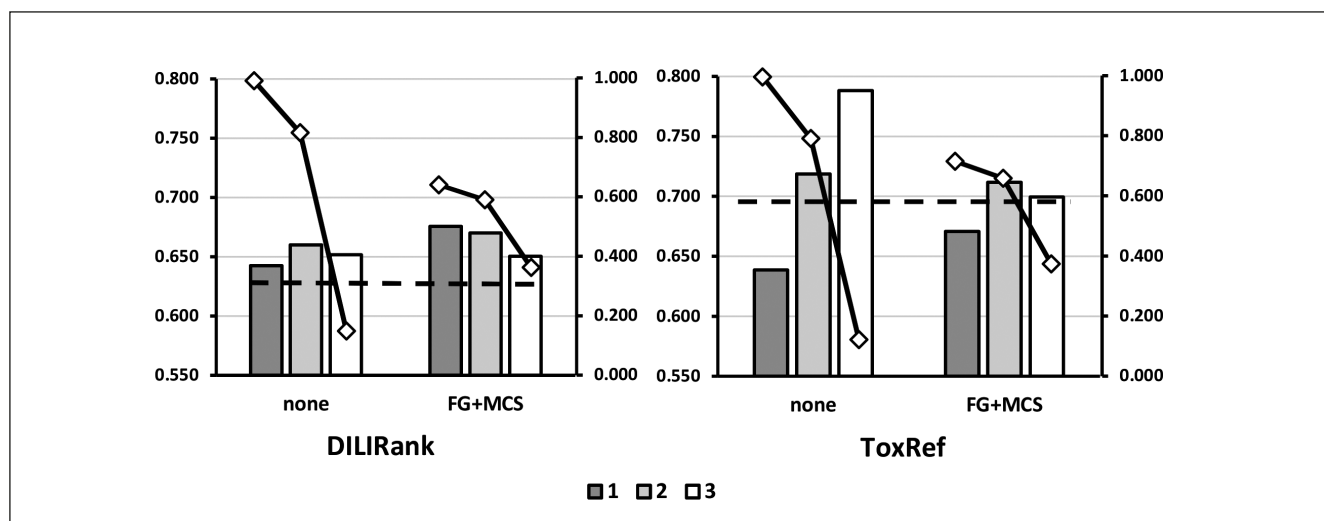
Information related to the stereoisomery was ignored because it is statistically not relevant. Indeed, out of the 15 couples of stereoisomers found in the DILIRank, only two cases (i.e., levofloxacin vs ofloxacin and amphetamine vs dextroamphetamine) showed differences in biological activity. No cases of stereoisomers with different activities were observed in the ToxRef dataset.

The final number of compounds included in each dataset and the distribution of activities as well as details on the chemical

<sup>11</sup> <https://www.fda.gov/science-research/liver-toxicity-knowledge-base-ltkb/drug-induced-liver-injury-rank-dilirank-dataset>

<sup>12</sup> <http://www.epa.gov/nct/toxcast/data.html>

<sup>13</sup> [https://github.com/DGadaleta88/data\\_curation\\_workflow](https://github.com/DGadaleta88/data_curation_workflow)



**Fig. 2: Validation results of the RAX integrated method applied to the DILIrank (three-class) and ToxRef (binary classification) dataset**

Grey bars report balanced accuracies (on the left y-axis) while solid black lines are the ratios of predicted compounds on the total (on the right y-axis). Results refer to predictions inferred from analogue(s) included in at least one, two or all three similarity lists (i.e., StrS, MS and BS). Dashed lines refer to BAs obtained by using only the closest structural analogue to make predictions.

space covered by the two datasets in terms of relevant physico-chemical properties (i.e., molecular weight, MW; octanol-water partition coefficient, logP; and topological polar surface area, TPSA) are given in Table 1, while detailed information on single activity categories are given in Table S2<sup>4</sup>. The datasets and the information retrieved with the RAX workflow used to compute similarities are included in the supporting information<sup>14</sup>.

### 3 Results

#### 3.1 Overall RAX strategy output

It should be kept in mind that this workflow is not primarily designed for batch calculation on large datasets, and therefore one cannot expect to reach prediction accuracies at the same level of other methodologies (e.g., QSARs) specifically tailored for predicting large databases. One of the major strengths of this approach is that the information used to infer predictions is explicitly reported to the user and, unlike the complex theoretical chemical descriptors used in QSAR modeling, it consists of sound chemical and toxicological data to facilitate use for regulators and scientists. In this regard, the authors propose to use the workflow for single chemical RAX predictions so that experts can take case-by-case decisions on the suitability of the identified analogue(s) based on the evaluation of the gathered data.

With this in mind, the large-scale evaluation described here is intended to provide an indication of the overall predictivity of the approach. The RAX workflow was validated by predicting compounds included in the two source datasets (i.e., DILIrank and ToxRef). For each chemical, analogue(s) were identi-

fied among remaining compounds in the source dataset. Then, a prediction was returned as a majority vote of activities of selected analogue(s). Separate predictions were generated by considering analogue(s) included in at least one, two or three similarity lists (i.e., structural, metabolic and biological). The effect of reducing the searchable list of analogue(s) by applying MCS and FG pre-filters on accuracy was also evaluated. In order to determine a benchmark in prediction quality, chemicals in the two source datasets were also predicted based on the sole use of the single closest analogue in terms of structural similarity. This was done to evaluate whether the combined use of multiple pieces of information and similarities represents an added value with respect to the traditional use of structural similarity. Full details on predictions are included in the supporting information<sup>14</sup>.

Figure 2 and Table 2 report balanced accuracies (BA) and the ratio of predicted compounds (i.e., coverage) for the two source datasets for each combination of pre-filtering options and minimum number of similarity lists that should contain a chemical to take it into account for the prediction. Table 2 and Tables S3-S4<sup>4</sup> include detailed statistics on the validation performed.

For the DILIrank source dataset, the “benchmark” BA obtained by using the closest structural analogue for predictions was equal to 0.632. For this dataset, the combined use of StrS, MS and BS slightly improved this result (i.e., BA = 0.642). BA was further improved when only analogue(s) matching at least two similarity lists were used (BA = 0.660), at the cost of a slight loss in the number of predictions (i.e., 0.18 ratio loss). The integration of all three similarity lists, on the other hand, did not provide further improvement. This is likely related to the increased unbalancing of the dataset when considering only chemicals that

<sup>14</sup> doi:10.14573/altex.2002281s3

**Tab. 2: Prediction statistics of the RAX integrated approach applied to ToxRef and DILIrank datasets**

For each combination of similarity lists (i.e., number of lists including a single analogue) and pre-filtering method, the sensitivity (SEN), the specificity (SPE), the balanced accuracy (BA) and ratio (%) of predictions are reported. For the multi-category DILIrank database,  $SEN_{avg}$  and  $SPE_{avg}$  are the average of values computed separately for each class, while  $BA_{avg}$  is the arithmetic mean of  $SEN_{avg}$  and  $SPE_{avg}$ . The first row of the table refers to the benchmark performance related to the sole use of the closest structural neighbor to infer the prediction.

Similarity lists	Pre-filtering	DILIrank				ToxRef			
		$SEN_{avg}$	$SPE_{avg}$	$BA_{avg}$	% <sup>a</sup>	SEN	SPE	BA	% <sup>b</sup>
-	-	0.510	0.754	0.632	1.000	0.571	0.825	0.698	1.000
1	none	0.523	0.762	0.642	1.000	0.332	0.945	0.639	1.000
2	none	0.546	0.774	0.660	0.825	0.596	0.841	0.719	0.798
3	none	0.533	0.770	0.652	0.156	0.697	0.879	0.788	0.141
1	FG+MCS	0.570	0.781	0.676	0.649	0.438	0.904	0.671	0.724
2	FG+MCS	0.563	0.777	0.670	0.599	0.527	0.897	0.712	0.668
3	FG+MCS	0.535	0.766	0.650	0.371	0.567	0.832	0.699	0.389

<sup>a</sup> 58 out of 691 SMILES in the dataset were unread from the RAX workflow and were not considered for statistical calculation.

<sup>b</sup> 18 out of 663 SMILES in the dataset were unread from the RAX workflow and were not considered for statistical calculation.

**Tab. 3: Percentage of single analogues having the same activity as the target**

The number of single selected source compounds (#scmpds), the number (#m\_scmpds) and the ratio (%m\_scmpds) of those matching their target's activity are grouped based on the similarity lists in which they are included.

Similarity	ToxRef			DILIrank		
	#scmpds	#m_scmpds	%m_scmpds	#scmpds	#m_scmpds	%m_scmpds
STR, MET, BIO	100	82	0.820	142	86	0.606
STR, MET	1186	895	0.755	1007	575	0.571
STR, BIO	145	108	0.745	204	103	0.505
MET, BIO	38	24	0.632	79	36	0.456
STR	5324	3506	0.659	5077	2046	0.403
MET	2795	1690	0.605	2140	829	0.387
BIO	2250	1350	0.600	4849	1824	0.376

are predicted under these conditions. Indeed, the “MostDILI” category showed the highest variation in ratio with respect to the initial distribution of activities (from 0.38 to 0.22 of the entire dataset). This category is associated with the highest reduction in classification performance and also affects the performance of the whole dataset.

For the ToxRef database, the simple integration of multiple similarities ( $BA = 0.639$ ) does not improve the benchmark performance (i.e.,  $BA = 0.698$ ). On the other hand, BA is improved considerably when only analogue(s) included in two ( $BA = 0.719$ ) or three ( $BA = 0.788$ ) similarity lists are considered, even if in the last case the coverage is severely reduced to 0.141.

Low sensitivity values were observed in some cases. The reason is likely related to the degree of unbalance of the datasets (see Tab. 1). The issue of handling unbalanced datasets is also commonly observed for other *in silico* methodologies for predicting toxicity

(i.e., QSAR), so that advanced strategies specifically designed to solve this problem have been proposed (Zakharov et al., 2014).

For the DILIrank dataset, improvements were observed when the analogue search was restricted to chemicals sharing common MCSs and FGs, with BA reaching a value of 0.670. On the other hand, the application of these filters on the ToxRef dataset did not improve the statistical performance. In this regard, the inspection of performance for single chemical categories revealed disappointing statistics for some of them (Tab. S5<sup>4</sup>). In particular, aromatic amines ( $BA = 0.58$ ) and aromatic alcohols ( $BA = 0.51$ ) are relatively well-represented in the dataset, and the low sensitivities associated with these classes are likely the reason for the drop in performance observed in the entire dataset. On the other hand, the methodology was found to perform better on aliphatic compounds such as carboxylic acids ( $BA = 0.88$ ), alcohols ( $BA = 0.82$ ) and halogens ( $BA = 0.71$ ). Overall, good performances



**Fig. 3: RAX examples from the ToxRef source dataset**

(A) 2-chlorophenol, (B) hydrazobenzene. HT, hepatotoxic; nHT, non-hepatotoxic

**A**

NAME	Structure	Rank	Activity	Structural Similarity	Metabolic Similarity	Common metabolic pathways	Biological Similarity	Common biological outcomes
2-Chlorophenol		Target	non Toxic	-	-	-	-	-
Phenol		STR5, MET1, BIO4	non Toxic	0.765	1.000	2 out of 2	0.553	299 NEG 2 POS
Triclosan		STR3, MET1	Toxic	0.810	1.000	2 out of 2	0.011	180 NEG 1 POS
p-Bromodiphenyl ether		STR6, MET1	non Toxic	0.750	1.000	2 out of 2	0.020	124 NEG
Pentachloro phenol		STR1	Toxic	1.000	0.000	0 out of 2	0.040	282 NEG 2 POS

⋮  
14 additional analogue(s):  
6 HT, 8 nHT

**B**

NAME	Structure	Rank	Activity	Structural Similarity	Metabolic Similarity	Common metabolic pathways	Biological Similarity	Common biological outcomes
Hydrazo benzene		Target	Toxic	-	-	-	-	-
Azobenzene		STR5, MET3, BIO3	Toxic	0.696	0.667	2 out of 3	0.590	224 NEG 7 POS
Benzidine		STR3, BIO2	non Toxic	0.708	0.250	1 out of 4	0.621	206 NEG 5 POS
Diphenyl amine		STR6, MET1	Toxic	0.682	1.000	2 out of 2	0.357	181 NEG 1 POS
4-Aminoazo benzene		STR1	non Toxic	0.792	0.400	2 out of 5	0.244	80 NEG 2 POS

⋮  
17 additional analogue(s):  
5 HT, 12 nHT

were observed for the majority of the chemical categories in the DILIrank dataset, with aliphatic carboxylic acids being predicted with the highest accuracy (BA = 0.80). Lower values were observed for aliphatic amines (BA = 0.57) and aromatic halogens (BA = 0.59). Detailed statistics for individual chemical classes are shown in Table S5<sup>4</sup>.

Generally speaking, results confirmed that the use of analogue(s) with different types of similarity improves performance compared to the sole use of structural similarity. This is especially true when source compounds used for RAX belong to at least two orthogonal similarity lists of analogue(s). In this re-

gard, they can be considered “close” to the target under multiple relevant aspects (e.g., chemical, toxicological, kinetic). In order to understand the contribution of different similarities to the final RAX prediction, the activity of single analogues was evaluated for their coherence with the activity of the relative target. In Table 3, analogues are grouped based on the similarity list(s) in which they appear. For each combination, the number of analogues exactly matching the relative target activity is reported. As expected, analogues included in multiple similarity lists are more likely to match the activity of the target compound. Indeed, the combination of all three similarity lists shows the highest ra-



A

NAME	Structure	Rank	Activity	Structural Similarity	Metabolic Similarity	Common metabolic pathways	Biological Similarity	Common biological outcomes
Amlodipine		Target	Less-DILI	-	-	-	-	-
Felodipine		STR2, MET1, BIO3	Less-DILI	0.754	0.818	9 out of 11	0.271	231 NEG 8 POS
Clevidipine		STR1, MET3	No-DILI	0.769	0.600	9 out of 15	c.ca 0	5 NEG
Nifedipine		MET4, BIO1	Less-DILI	0.514	0.500	7 out of 14	0.302	212 NEG 14 POS

13 additional analogue(s)  
4 No-DILI, 8 Less-DILI, 1 Most-DILI

B

NAME	Structure	Rank	Activity	Structural Similarity	Metabolic Similarity	Common metabolic pathways	Biological Similarity	Common biological outcomes
Dobutamine		Target	No-DILI	-	-	-	-	-
Isoproterenol		STR7, MET5, BIO2	No-DILI	0.688	0.455	5 out of 11	0.331	194 NEG 12 POS
Dopamine		STR2, MET1	No-DILI	0.756	0.625	5 out of 8	0.243	177 NEG 6 POS
Epinephrine		STR7, MET3	No-DILI	0.688	0.500	5 out of 10	0.061	26 NEG 3 POS
Propafenone		STR1	Less-DILI	0.827	0.211	4 out of 19	0.258	166 NEG 16 POS
Labetalol		STR3	Most-DILI	0.745	0.385	5 out of 13	0.186	42 NEG 13 POS

11 additional analogue(s)  
6 No-DILI, 4 Less-DILI, 1 Most-DILI

**Fig. 4: RAX examples from the DILrank source dataset**  
(A) amlodipine;  
(B) dobutamine.

tios of source compounds having the same activity as the target, i.e., 0.82 for ToxRef and 0.60 for DILrank, reinforcing the initial hypothesis of this study.

This combination is followed by those integrating two out of three different similarities; in particular combinations including structural similarity (i.e., StrS and MS or StrS and BS) were always characterized by higher percentages of concordant analogues than the one combining MS and BS. As for source compounds included in a single similarity list, those included in the StrS list were more often concordant with their target's activity than those included in the BS or MS lists.

### 3.2 RAX examples

Figures 3 and 4 provide four examples that show how the output of the RAX workflow should be interpreted and how the integrated similarities represent an advantage over the StrS alone.

A total of 18 analogue(s) were identified from ToxRef to predict the RAX of 2-chlorophenol (CAS 95-57-8) (Fig. 3A), a non-hepatotoxic compound. Phenol (CAS 108-95-2) was the only analogue that was included in all three similarity lists. Despite not being ranked as a top analogue for StrS, it may be considered the most suitable analogue for RAX; indeed, the chemical has a very high BS to the target (BS = 0.553, 299 negative assays and 2 positive



assays shared with the target). In addition, it undergoes the same metabolic biotransformations observed for the target, i.e., aromatic hydroxylation ortho- and para- to oxygen (MS = 1.00). In case of 2-chlorophenol, these biotransformations generate 3-chlorocatechol and 2-chlorobenzene-1,4-diol, respectively, while phenol is converted to catechol and hydroquinone, respectively. Hydroxylation may plausibly play a role in the detoxification of both chemicals, because it increases the number of hydroxyl reactive centers that can undergo phase II reactions (i.e., conjugation), making the chemicals more easily excreted from the body. Phenol has the same activity as the target, leading to a correct RAX prediction. On the other hand, using the activity of the top structural analogue (pentachlorophenol, CAS 87-86-5) leads to an incorrect prediction. Indeed, pentachlorophenol does not undergo hydroxylation, which is observed for 2-chlorophenol and is a key process for detoxification. Moreover, despite sharing a good amount of common biological assay outcomes, pentachlorophenol also activates 41 additional assays that are not activated by the target (data not shown), drastically lowering the BS (i.e., 0.040).

Hydrazobenzene (CAS 122-66-7) (Fig. 3B) is a hepatotoxic chemical from ToxRef. As in the previous example, the sole use of the top structural analogue (4-aminoazobenzene, CAS 60-09-3) leads to an underestimation of the target's toxicity. Indeed, 4-aminoazobenzene has a relatively low BS (i.e., 0.244) that leads to its exclusion from the top five biological analogues. It also differs from the target from a metabolic point of view because it undergoes additional biotransformations that are likely responsible for its detoxification. On the other hand, azobenzene (CAS 103-33-3) appears in all three similarity lists and is more suitable as a RAX analogue despite it being less similar to the target from a structural point of view. Benzidine (CAS 92-87-5) and diphenylamine (CAS 122-39-4) are further analogues of hydrazobenzene that are included in two out of three similarity lists. Benzidine has a lower MS score (only one shared pathway out of four observed for the target and the analogue) while diphenylamine shows a different biological behavior (only one positive assay shared with the target). Overall, when analogue(s) that are in two or more lists are used, a prevalence of toxic chemicals (two out of three) is observed, leading to a correct prediction of hydrazobenzene as hepatotoxic.

Amlodipine (CAS 88150-42-9) (Fig. 4A) is classified as "Less-DILI" in the DILIRank dataset. Out of the 16 analogues identified in the RAX procedure, three are included in two or more similarity lists. Two of them (i.e., fenlodypine, CAS 72509-76-3, and nifedipine, CAS 21829-25-4) have the same "Less-DILI" classification as the target, while the third (i.e., clevidipine, CAS 167221-71-8) has a "No-DILI" classification. This is also the top structural analogue of amlodipine in the dataset. The unsuitability of clevidipine as a source compound for the RAX is related to its very low BS with respect to amlodipine (i.e., close to zero). Indeed, this chemical only shares five negative assay outcomes with the target, while the other two analogues are characterized by more than 200 common assay responses. Overall, fenlodypine is the most suitable analogue for RAX, as it is included in all three similarity lists, while nifedipine shows a relatively low StrS value (StrS = 0.514). Metabolism is also relevant; indeed, clevidipine

is rapidly metabolized to its primary high-clearance metabolite (H152/81) after cleavage of the ester group at position 3 of the dihydropyridine core (Ericsson et al., 1999). Due to absence of this rapid cleavage, amlodipine has a much longer half-life than clevidipine and a higher toxicity.

The second example from DILIRank is dobutamine (CAS 34368-04-2) (Fig. 4B), a "No-DILI" chemical. Two of the top structural analogues, i.e., propafenone (CAS 54063-53-5) (rank 1) and labetalol (36894-69-6) (rank 3), have discordant activity with respect to the target, and are indeed considered not suitable for RAX. The reason is that they show a very different metabolic profile in comparison to dobutamine, which shares only 4 out of 19 pathways with propafenone, and 5 out of 13 with labetalol. It is likely that some of the non-shared biotransformations are responsible for the detoxification of the target and/or the increased toxicity of the analogues. In particular, propafenone and labetalol do not share the same catechol moiety that is present in dobutamine. This is the main site of the metabolic transformations (i.e., aromatic hydroxylation and conjugations) that are responsible for the compound's clearance. On the other hand, the three analogues isoproterenol (CAS 7683-59-2), dopamine (CAS 51-61-6) and epinephrine (51-43-4) share the same catechol moiety as dobutamine. They share a higher number of detoxification pathways and show higher MS values with dobutamine than propafenone and labetalol. They are indeed characterized by the same activity as dobutamine. Isoproterenol is also one of the closest analogues in terms of biological profile (i.e., 194 negative assay responses and 12 positive ones in common with the target), despite it being ranked only as the seventh closest structural analogue.

In addition to tabular outputs, Figure S2<sup>4</sup> shows pie-plots generated by the workflow to graphically describe RAX results for the four compounds. In these plots, all the putative analogue(s) are included as separate slices in clockwise order, starting from those included in multiple similarity lists (i.e., outlined in dark blue for three lists and light blue for two lists) with background colors that are descriptive of the experimental activity of each analogue(s).

## 4 Discussion

### 4.1 Integrated RAX strategy

This paper describes an automated approach to identify suitable analogue(s) for RAX. As in traditional RAX, the selection of analogue(s) is based on similarity with the target under study. However, while StrS is in many cases the only piece of information used to identify neighbors, this automated tool mathematically combines different approaches to evaluate similarity between the target and a series of putative analogue(s) from a source dataset to strengthen the evidence in supporting analogue(s) selection. Results in Tables 2-3 confirm the beneficial role of this strategy that improves predictivity in comparison to the sole use of the StrS.

The use of MCS and common FGs as pre-filters to narrow the list of candidate analogues does not seem to always carry evident improvements for the ToxRef dataset. Hepatotoxicity is a complex endpoint, which is related to multiple mechanisms of

action. Consequently, restricting the analogue(s) search to a single chemical category may be less effective than for other endpoints with a clear mechanistic link to specific substructures and/or functional groups (e.g., mutagenicity).

The methodology led to good predictive performance on the two validation datasets, with BA values in the range of 0.632–0.670 for the DILIrank and 0.639–0.788 for the ToxRef dataset. The approach suffers from the use of unbalanced data for validation, sometimes leading to low sensitivity values. However, the reader should keep in mind that the methodology presented here does not aspire to reach the same level of predictive performance of other statistical approaches that are specifically tailored to large-scale toxicity predictions. The main strength of this integrated RAX approach is its ability to provide results that are easy to understand, as well as a data-rich output for users to evaluate the final toxicity outcome based on their expertise.

Other examples that combine different approaches to compute similarity for RAX have been described. Notably, Wu et al. (2010) proposed and validated a robust evaluation framework to determine analogue(s) suitability for RAX that used several cheminformatics tools, e.g., to evaluate physicochemical properties of substances and simulate their metabolic pathway. Even though the decision framework proposed by Wu and coworkers takes into consideration all relevant aspects necessary for a RAX evaluation, no automated procedure is proposed and the intervention of various experts is required at various levels. Low et al. (2013) proposed an automated Chemical Biological Read-Across (CBRA) that used an integrated StrS and BS to predict the toxicity of chemicals. The approach was adapted by Shah et al. (2016) in their Generalized Read-Across (GenRA) that compares the efficiency of structural and biological fingerprints (and a combination of both) to search for analogue(s) for RAX. In both cases, only StrS and BS are considered, while other highly relevant information (e.g., physicochemical properties, metabolism, reactivity and pharmacokinetics) is not addressed.

Results reported in Table 3 confirm that in our approach StrS maintains a predominant role. It is accepted that the chemistry should be the starting point for the definition of similarity (ECHA, 2008; Ball et al., 2016) owing to the strong correlation between the structure of compounds and their biological effects (Bender and Glen, 2004). For this reason, a higher number of neighbors based on StrS was selected (i.e., up to 10 compared to up to 5 analogue(s) for other similarities). StrS identifies common behaviors between the target and the analogue(s) in terms of physicochemical properties and, consequently, in terms of toxicokinetics. Indeed, the toxicity of a chemical depends on its absorption and excretion rates and the time that it effectively spends in the organism. Differences in these parameters could affect *in vivo* toxicity (due to differences in bioavailability) or *in vitro* toxicity (due to differences in solubility). On the contrary, high structural similarity values will result in analogies for most important physicochemical properties (e.g., molecular weight, lipophilicity, solubility) and, as a consequence, in a similar toxicokinetic profile.

The role of MS becomes relevant when the metabolism of non-toxic chemicals leads to the production of harmful metabolites or, alternatively, when a toxic substance is detoxified by

a metabolic process. Metabolism of the analogue(s) and of the target can have a significant impact on the overall RAX assessment. Indeed, the potential for two chemicals to diverge in their bioactivation pathway may result in a different toxicological profile, and it may therefore affect the conclusions drawn when using structural similarity alone (Patlewicz et al., 2013).

The use of BS has been explored extensively in the last years (Petrone et al., 2012; Russo et al., 2017; Shah et al., 2016; Grimm et al., 2016; Low et al., 2013; Patlewicz et al., 2017). An approach that found broad consensus, and was also applied in this paper, was to use outcomes (i.e., positive or negative) from a large number of HTS assays to build a binary biological fingerprint of the target and of the analogue(s). Fingerprints are well suited to compute similarity; indeed one can use the collective set of results from different assays to compare the target and the analogue(s) using classical mathematical methods, e.g., Tanimoto or Euclidean distances (Willet et al., 1998). Two chemicals characterized by a similar behavior on a large number of different biological assays are also likely to share a common toxicological profile.

#### 4.2 Evaluation of uncertainty

The importance of an explicit strategy to characterize the uncertainty associated with RAX data gap-filling has been highlighted (OECD, 2017; Patlewicz et al., 2013; Blackburn and Stuard, 2014). The number and the degree of suitability of analogue(s), the quality and quantity of the data considered, the nature and severity of the identified toxic effects, and the potency of the analogue(s) for those effects should be ideally evaluated in order to assess the effectiveness of the RAX and make transparent decisions.

The RAX tool offers several elements to quantify the uncertainty associated with RAX, e.g., the final number of analogue(s) used for the data gap-filling, the criteria for selection (e.g., use of FG and MCS filters), and the number of categories into which each analogue falls. Ideally, a RAX based on a higher number of analogue(s) included in multiple lists is considered more reliable than a RAX based on few analogue(s) sharing few similarity lists. Another element that can be used to evaluate uncertainty associated with RAX is the consistency of activities across analogue(s) used. The greater the percentage of analogue(s) having the same activity, the more the final prediction can be considered reliable.

Currently, this RAX workflow was validated for qualitative prediction of toxicity (i.e., classification), because quantitative RAX was recognized as more challenging for the higher number of potential areas of uncertainty to address (Ball et al., 2016). As a drawback, increasing the number of similarities introduces further requirements for analogue(s) selection, which consequently reduces coverage of the method.

#### 5 Conclusions

In this paper, we present an automated tool that implements good practices for toxicological data-gap filling in RAX that have been described in recent literature and technical guidelines. Much emphasis is put on the combined use of different types of similari-



ty to identify suitable analogue(s) with a high level of reliability. The workflow provides a rich output in form of tables/graphs that provide a strong basis to support RAX conclusions and to aid toxicologists and risk assessors in decision-making. The data (i.e., simulated metabolites, positive HTS assays, common substructures and/or functional groups) used to selected analogue(s) are made available in the workflow output and represent a relevant resource for users to assess the reliability of the conclusions drawn by the tool, e.g., by manually evaluating the consistency of toxicological and biological data gathered across the analogue(s).

A limitation of the tool is that it does not identify *a priori* specific information relevant for different types of toxicity. Indeed, some specific properties can be key elements for making assessments for some endpoints (e.g., reactivity for mutagenicity, lipophilicity for bioaccumulation), while being less important for other endpoints. Each endpoint should be justified case-by-case, and RAX should be ideally endpoint-specific (Patlewicz et al., 2015). In this regard, we invite users to consider the workflow presented here mainly in combination with other sources of evidence, as a visualization tool, and a source of relevant data to aid the reasoning, more than as an autonomous predictive tool. Further work is needed to develop variations of the workflow that are specific for a given endpoint.

Adverse outcome pathways (AOPs) represent a promising resource that has been proposed to address endpoint specificities (Tollefsen et al., 2014; Patlewicz et al., 2015). Although the use of AOPs is currently limited by the low number of validated AOPs, some efforts have been made to apply computational tools to AOPs (Gadaleta et al., 2018). In the future, the integration of computational predictions for molecular initiating events and key events in an AOP could be used to demonstrate that a set of chemicals has analogous biological behavior that is relevant for the toxicological endpoint of concern, providing new evidence for improving RAX results (Leist et al., 2017).

The presented automated workflow for analogue(s) selection for RAX can reduce animal experiments and improve the process of extracting all relevant information from existing data in a more efficient and organized way where multiple features of heterogeneous nature are integrated.

## References

- Ball, N., Cronin, M. T., Shen, J., Blackburn, K. et al. (2016). Toward good read-across practice (GRAP) guidance. *ALTEX* 33, 149-166. doi:10.14573/altex.1601251
- Bender, A. and Glen, R. C. (2004). Molecular similarity: A key technique in molecular informatics. *Org Biomol Chem* 2, 3204-3218. doi:10.1039/b409813g
- Berthold, M. R., Cebon, N., Dill, F. et al. (2008). KNIME: The Konstanz information miner. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme and R. Decker (eds.), *Data Analysis, Machine Learning and Applications. Studies in Classification, Data Analysis, and Knowledge Organization* (319-326). Berlin, Germany: Springer. doi:10.1007/978-3-540-78246-9\_38
- Bishop, P. L., Manuppello, J. R., Willett, C. E. et al. (2012). Animal use and lessons learned in the US high production volume chemicals challenge program. *Environ Health Perspect* 120, 1631-1639. doi:10.1289/ehp.1104666
- Blackburn, K. and Stuard, S. B. (2014). A framework to facilitate consistent characterization of read across uncertainty. *Regul Toxicol Pharmacol* 68, 353-362. doi:10.1016/j.yrtph.2009.09.006
- Chen, M., Vijay, V., Shi, Q. et al. (2011). FDA-approved drug labeling for the study of drug-induced liver injury. *Drug Discov Today* 16, 697-703. doi:10.1016/j.drudis.2011.05.007
- Chen, M., Suzuki, A., Thakkar, S. et al. (2016). DILrank: The largest reference drug list ranked by the risk for developing drug-induced liver injury in humans. *Drug Discov Today* 21, 648-653. doi:10.1016/j.drudis.2016.02.015
- Cruz-Monteagudo, M., Medina-Franco, J. L., Perez-Castillo, Y. et al. (2014). Activity cliffs in drug discovery: Dr Jekyll or Mr Hyde? *Drug Discov Today* 19, 1069-1080. doi:10.1016/j.drudis.2014.02.003
- Durant, J. L., Leland, B. A., Henry, D. R. et al. (2002). Reoptimization of MDL keys for use in drug discovery. *J Chem Inf Comput Sci* 42, 1273-1280. doi:10.1021/ci010132r
- EC – European Commission (2006). Regulation (EC) of No 1907/2006 of the European parliament and of the council 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. doi:10.5771/9783845266466-1026
- ECHA (2008). Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.6: QSARs and Grouping of Chemicals.
- ECHA (2014). The Use of Alternatives to Testing on Animals for the REACH Regulation, ISBN 978-92-9244-593-5. Second report under Article 117(3) of the REACH Regulation. ECHA-14-A-07-EN. doi:10.2823/22471
- ECHA (2015). Read-across Assessment Framework (RAAF). ECHA-15-R-07-EN, 2015.
- Ericsson, H., Tholander, B., Björkman, J. A. et al. (1999). Pharmacokinetics of new calcium channel antagonist clevidipine in the rat, rabbit, and dog and pharmacokinetic/pharmacodynamic relationship in anesthetized dogs. *Drug Metab Dispos* 27, 558-564. <http://dmd.aspetjournals.org/content/27/5/558.long>
- Gadaleta, D., Manganelli, S., Roncaglioni, A. et al. (2018). QSAR modeling of ToxCast assays relevant to the molecular initiating events of AOPs leading to hepatic steatosis. *J Chem Inf Model* 58, 1501-1517. doi:10.1021/acs.jcim.8b00297
- Grimm, F. A., Iwata, Y., Sirenko, O. et al. (2016). A chemical-biological similarity-based grouping of complex substances as a prototype approach for evaluating chemical alternatives. *Green Chem* 18, 4407-4419. doi:10.1039/c6gc01147k
- Leist, M., Ghallab, A., Graepel, R. et al. (2017). Adverse outcome pathways: Opportunities, limitations and open questions. *Arch Toxicol* 91, 3477-3505. doi:10.1007/s00204-017-2045-3



- Liu, J., Mansouri, K., Judson, R. S. et al. (2015). Predicting hepatotoxicity using ToxCast in vitro bioactivity and chemical structure. *Chem Res Toxicol* 28, 738-751. doi:10.1021/tx500501h
- Low, Y., Sedykh, A., Fourches, D. et al. (2013). Integrative chemical-biological read-across approach for chemical hazard classification. *Chem Res Toxicol* 26, 1199-1208. doi:10.1021/tx400110f
- Luechtefeld, T., Marsh, D., Rowlands, C. et al. (2018). Machine learning of toxicological big data enables read-across structure activity relationships (RASAR) outperforming animal test reproducibility. *Toxicol Sci* 165, 198-212. doi:10.1093/toxsci/kfy152
- Maggiara, G., Vogt, M., Stumpfe, D. et al. (2014). Molecular similarity in medicinal chemistry: Miniperspective. *J Med Chem* 57, 3186-3204. doi:10.1021/jm401411z
- OECD (2014). Guidance on Grouping of Chemicals. *OECD Series on Testing and Assessment, No. 80*. OECD Publishing, Paris. doi:10.1787/9789264085831-en
- OECD (2017). Guidance on Grouping of Chemicals, Second Edition. *OECD Series on Testing and Assessment, No. 194*. OECD Publishing, Paris. doi:10.1787/9789264274679-en
- Patlewicz, G., Ball, N., Booth, E. D. et al. (2013). Use of category approaches, read-across and (Q) SAR: General considerations. *Regul Toxicol Pharmacol* 67, 1-12. doi:10.1016/j.yrtph.2013.06.002
- Patlewicz, G., Ball, N., Becker, R. A. et al. (2014). Food for thought: Read-across approaches – Misconceptions, promises and challenges ahead. *ALTEX* 31, 387-396. doi:10.14573/altex.1410071
- Patlewicz, G., Ball, N., Boogaard, P. J. et al. (2015). Building scientific confidence in the development and evaluation of read-across. *Regul Toxicol Pharmacol* 72, 117-133. doi:10.1016/j.yrtph.2015.03.015
- Patlewicz, G., Helman, G., Pradeep, P. et al. (2017). Navigating through the minefield of read-across tools: A review of in silico tools for grouping. *Comput Toxicol* 3, 1-18. doi:10.1016/j.comtox.2017.05.003
- Petrone, P. M., Simms, B., Nigsch, F. et al. (2012). Rethinking molecular similarity: Comparing compounds on the basis of biological activity. *ACS Chem Biol* 7, 1399-1409. doi:10.1021/cb3001028
- Pradeep, P., Mansouri, K., Patlewicz, G. et al. (2017). A systematic evaluation of analogs and automated read-across prediction of estrogenicity: A case study using hindered phenols. *Comput Toxicol* 4, 22-30. doi:10.1016/j.comtox.2017.09.001
- Ridder, L. and Wagener, M. (2008). SyGMA: Combining expert knowledge and empirical scoring in the prediction of metabolites. *ChemMedChem* 3, 821-832. doi:10.1002/cmdc.200700312
- Russell, W. M. S. and Burch, R. L. (1959). *The Principles of Humane Experimental Technique*. London, UK: Methuen.
- Russo, D. P., Kim, M. T., Wang, W. et al. (2017). CIIPPro: A new read-across portal to fill data gaps using public large-scale chemical and biological data. *Bioinformatics* 33, 464-466. doi:10.1093/bioinformatics/btw640
- Sánchez-Cruz, N. and Medina-Franco, J. L. (2018). Statistical-based database fingerprint: Chemical space dependent representation of compound databases. *J Cheminform* 10, 1-13. doi:10.1186/s13321-018-0311-x
- Shah, I., Liu, J., Judson, R. S. et al. (2016). Systematically evaluating read-across prediction and performance using a local validity approach characterized by chemical structure and bioactivity information. *Regul Toxicol Pharmacol* 79, 12-24. doi:10.1016/j.yrtph.2016.05.008
- Tollefsen, K. E., Scholz, S., Cronin, M. T. et al. (2014). Applying adverse outcome pathways (AOPs) to support integrated approaches to testing and assessment (IATA). *Regul Toxicol Pharmacol* 70, 629-640. doi:10.1016/j.yrtph.2014.09.009
- Willett, P., Barnard, J. M. and Downs, G. M. (1998). Chemical similarity searching. *J Chem Inf Comput Sci* 38, 983-996. doi:10.1021/ci9800211
- Wu, S., Blackburn, K., Amburgey, J. et al. (2010). A framework for using structural, reactivity, metabolic and physicochemical similarity to evaluate the suitability of analogs for SAR-based toxicological assessments. *Regul Toxicol Pharmacol* 56, 67-81. doi:10.1016/j.yrtph.2009.09.006
- Zakharov, A. V., Peach, M. L., Sitzmann, M. et al. (2014). A new approach to radial basis function approximation and its application to QSAR. *J Chem Inf Model* 54, 713-719. doi:10.1021/ci400704f

#### Conflict of interest

The authors declare that they have no conflicts of interest.

#### Acknowledgements

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 681002 (EU-ToxRisk). We acknowledge the anonymous reviewers for their comments that significantly improved the manuscript.