NAM-Supported Read-Across: From Case Studies to Regulatory Guidance in Safety Assessment

The use of new approach methodologies (NAMs) in support of read-across (RAx) approaches for regulatory purposes is a main goal of the EU-ToxRisk project. To bring this forward, EU-ToxRisk partners convened a workshop in close collaboration with regulatory representatives from key organizations including European regulatory agencies, such as the European Chemicals Agency (ECHA) and the European Food Safety Authority (EFSA), as well as the Scientific Committee on Consumer Safety (SCCS), national agencies from several European countries, Japan, Canada and the USA, as well as the Organisation for Economic Cooperation and Development (OECD). More than a hundred people actively participated in the discussions, bringing together diverse viewpoints across academia, regulators and industry. The discussion was organized starting from five practical cases of RAx applied to specific problems that offered the opportunity to consider real examples.

There was general consensus that NAMs can improve confidence in RAx, in particular in defining category boundaries as well as characterizing the similarities/dissimilarities between source and target substances. In addition to describing dynamics, NAMs can be helpful in terms of kinetics and metabolism that may play an important role in the demonstration of similarity or dissimilarity among the members of a category. NAMs were also noted as effective in providing quantitative data correlated with traditional no observed adverse effect levels (NOAELs) used in risk assessment, while reducing the uncertainty on the final conclusion.

An interesting point of view was the advice on calibrating the number of new tests that should be carefully selected, avoiding the allure of "the more, the better". Unfortunately, yet unsurprisingly, there was no single approach befitting every case, requiring careful analysis delineating the optimal approach. Expert analysis and assessment of each specific case is still an important step in the process.

1 Introduction

1.1 Definitions and aim of the workshop

Read-across (RAx) is a data-filling technique that allows endpoint information for one chemical to be predicted by using data for the same endpoint from (an)other chemical(s) considered to be similar in a significant way. This technique permits new *in vivo* tests on the target substance(s) to be waived by using available data on the source substance(s) within an analogue or category approach (Escher et al., 2019; Rovida et al., 2020) (Box 1). The technique is used in both regulatory and non-regulatory applications, from hazard classification to safety assessments. The analogue approach (one-to-one) is the direct transposition of one substance to extrapolate the properties of another substance, while the category approach (many-to-many or many-to-one) is applied to many substances as source that may fill gaps for one or many other substances included in the same group. In this case, it is also important to define whether the extrapolated value for the specific endpoint is the same or whether there is a defined and measurable trend.

The overarching RAx principle is the assumption that similar chemicals will exhibit similar biological activities. Unfortunately, this is not always the case, and apparent chemical similarity may result in different biological/toxicological activities. Consequently, to support the RAx hypothesis, a robust scientific justification including analysis of the structural, chemical, physico-chemical, toxicokinetic, biological and toxicological similarities together with uncertainty analysis that considers the probability of any unexpected biological/toxicological activity is needed (Schultz et al., 2015). The biological activity among the

category members may be equivalent or may show a trend that depends on many factors, such as physicochemical, toxicokinetic or toxicodynamic properties.

Despite its potential, RAx often fails to be accepted by regulatory agencies, as the scientific justification for its use is inadequate or incomplete, and hence the associated uncertainties are perceived as being too great. The issue of RAx justifications has been actively investigated within the EU-ToxRisk project through the combination of RAx procedures with other new approach methodologies (NAMs), i.e., by combining existing *in vivo* data with novel *in vitro* and *in silico* approaches (Box 1) (see also ECHA, 2016). This approach is expected to effectively fill gaps in RAx justification (Rovida et al., 2020), but there remains a lack of experience in its application that continues to thwart routine uptake within the regulatory community.

In order to discuss ways in which regulatory confidence in RAx can be strengthened by application of NAMs, EU-ToxRisk partners organized a dedicated workshop to discuss five different case studies with stakeholders from academia, industry and regulatory representatives from key organizations including European regulatory agencies, such as the European Chemicals Agency (ECHA) and the European Food Safety Authority (EFSA), as well as the Scientific Committee on Consumer Safety (SCCS), national agencies from several European countries, Japan, Canada and the USA, as well as the Organisation for Economic Cooperation and Development (OECD). More than a hundred people actively participated in the discussion, bringing forth viewpoints across academia, regulators and industry.

The workshop was structured around case studies that had been evaluated within the EU-ToxRisk scientific advisory board and regulatory advisory board and provided to EU regulators for technical feedback, plus two additional case studies from the OECD Integrated Approaches for Testing and Assessment (IATA) Case Studies Project. At the workshop, regulators outlined their needs, while case study developers explained the scientific basis of the case studies. This set the scene for subsequent discussion among the workshop participants, who were split into breakout groups to address a set of common questions that had been prepared in advance by the scientific committee of the EU-ToxRisk project and the workshop organizing committee. Ensuing wrap-up sessions aimed at reaching a shared conclusion and agreement on main developing points.

1.2 Regulatory / cross-regulatory perspective with ECHA and EFSA as examples

Box 1: Definitions of domain-specific terminology to ensure a common understanding of key concepts

– *Read-across (RAx)* describes a category or analogue approach as defined in the Read-Across Assessment Framework (RAAF) (ECHA, 2017). Compounds with relevant data are named *source compounds*, whereas the compounds that lack data are named *target compounds*. Within a RAx problem formulation, endpoint data of source compounds are used to estimate the same end-point for the target compound in a qualitative and quantitative way.
– *Analogue approach* refers to RAx from one or only a small number of structurally similar source compounds to the target compounds. Typically, there is no trend or regular pattern on the properties.
– *Category approach* refers to a grouping in which data from many source compounds are used to predict the hazard of one or more target compounds ("many-to-one" or "many-to-many" RAx) or many target compounds (many-to-many RAx). Properties of compounds within a category may be constant or follow a consistent trend.
– *New approach methodologies (NAMs)*, within EU-ToxRisk, encompass novel *in vitro* methodologies, for example high-throughput screening and high-content imaging methods or microphysiological systems, along with *in silico* methods, such as QSAR and PBTK modelling, that are used not only for data generation, but also for data interpretation and integration. The acronym NAM is also used as an adjective to indicate something that is related to this area. The inclusion of *in vivo* tests in the definition of NAMs is a matter of lively debate.
– *Lead effect* is a critical effect that is likely to determine the point of departure (PoD) for risk assessment (e.g., NOAEL/LOAEL) in the *in vivo* study.

ECHA's main responsibilities lie with Regulation (EC) 1907/2006 (Registration Evaluation, Authorisation and restriction of Chemicals, REACH), Regulation (EC) 1272/2008 (Classification, Labelling and Packaging, CLP) and Regulation (EU) 528/2012 (Biocidal Products Regulation, BPR), three prominent European regulations. REACH and BPR lay out legal data requirements for chemicals, including requests for *in vivo* toxicological tests. Arguably, REACH was the first piece of legislation worldwide to mention the possibility of using alternative approaches like RAx to waive new *in vivo* studies for the preparation of registration dossiers. RAx was extensively used in REACH registration dossiers, but in many cases the approach failed the completeness check of the authority.

In order to increase transparency on the evaluation of RAx in registration dossiers, ECHA published a document describing the so-called Read-Across Assessment Framework (RAAF) in 2015, which was extended for environmental toxicity endpoints in 2017 (ECHA, 2017). The RAAF was not intended as guidance for applicants, but provides, in a structured way, assessment elements to evaluate the RAx approach in order to understand whether the RAx hypothesis is supported by the available data through providing key assessment elements to identify strengths and weaknesses. The RAAF illustrates 6 scenarios to describe the type of RAx approach (Tab. 1), considering whether source and target substance(s) are bio-transformed into common compounds or whether they share the same type of effect as a result of their structural similarity. The RAAF strengthens the importance of the quality of the RAx justification as a basis for acceptance in REACH registration dossiers, categorizing it as accept-

Tab. 1: Different types of RAx as considered in the ECHA RAAF document (ECHA, 2017)

Scenario	Approach	READ-across hypothesis	Based on quantitative variations
1	Analogue	(Bio)transformation to common compound(s)	Properties of the target substance predicted to be quantitatively equal to those of the source substance or prediction based on a worst-case approach.
2	Analogue	Different compounds have qualitatively similar properties	Properties of the target substance predicted to be quantitatively equal to those of the source substance or prediction based on a worst-case approach.
3	Category	(Bio)transformation to common compound(s)	Variations in the properties observed among source substances. Prediction based on a regular pattern or on a worst-case approach.
4	Category	Different compounds have qualitatively similar properties	Variations in the properties observed among source substances. Prediction based on a regular pattern or on a worst-case approach.
5	Category	(Bio)transformation to common compound(s)	No relevant variations in properties observed among source substances and the same strength predicted for the target substance.
6	Category	Different compounds have qualitatively similar properties	No relevant variations in properties observed among source substances and the same strength predicted for the target substance

able with high confidence (score 5), acceptable with medium confidence (score 4), acceptable with just sufficient confidence (score 3), not acceptable in its current form (score 2) or not acceptable (score 1).

EFSA is the agency in Europe that produces scientific opinions and advice that form the basis for European policies and legislation related to the food chain. One of the areas of interest is within Regulation (EC) 1107/2009 on Plant Protection Products (PPPs), together with Regulation (EU) 283/2013 that lists the data requirements for active substances. An extensive array of *in vivo* tests is always mandatory for pesticide active substances, but there is an openness to alternative approaches, for example in the evaluation of metabolites, assessment of endocrine disruptive properties (EFSA et al., 2018), for the genotoxic assessment of impurities (Benigni et al., 2019) or the dietary risk assessment of PPP residues (EFSA, 2016).

Legislation on food additives is not as precise as the legislation on PPPs in terms of data requirements, and preparation of a risk assessment dossier for EFSA should follow the prevailing EFSA guidance documents. Here, quantitative structure activity relationships (QSARs) and RAx are mentioned as possibilities, mainly restricted for the selection of the most suitable *in vivo* test to be performed or in support of the available data.

There are other examples of EFSA endorsing the application of RAx, such as the assessment of substances being present at very low concentrations in food, like an impurity derived from a new manufacturing process of a food supplement or as trace substances migrating from food contact materials. At present, there are no guidelines on this topic, and these are generally assessed on a case-by-case basis.

The conclusion from the perspective of EFSA is that they are aware of the new possibilities coming from NAMs and RAx for

regulatory risk assessment. Once accepted by risk assessors and risk managers, they can help drive innovation in regulatory chemical risk assessment.

Many other international organizations and national regulatory agencies are applying RAx in several areas. Rovida et al. (2020) presented a general overview of the international application of RAx.

1.3 The EU-ToxRisk NAM-based RAx strategy

The EU-ToxRisk Project (Daneshian et al., 2016) started in 2016 as an Integrated European "Flagship" program aiming to study the possibility of replacing systemic *in vivo* test methods with integrated systems using QSARs and *in vitro* methods in IATA.

EU-ToxRisk developed specific RAx case studies to evaluate the applicability of NAMs to provide evidence and support for RAx approaches, in particular by using mechanistic data for hazard characterization and state-of-the-art physiologically-based toxicokinetic (PBTK) modelling to address differences in human kinetics/bioavailability of compounds. The ultimate goal was to i) increase confidence in the final RAx conclusion, ii) understand and document remaining uncertainties, and iii) build acceptance from a regulatory perspective.

The RAx hypothesis is based on the available data for the source compounds, mainly *in vivo* endpoint data. For hazard identification, the generation of NAM data is oriented to this RAx hypothesis and distinguishes three different scenarios (Escher et al., 2019), namely:

1. *Known adverse outcome pathway (AOP) or mode of action.* Source substances share a well-defined AOP/AOP network or mode of action that is represented by a specific molecular initiation event (MIE) and subsequent downstream key events (KEs) that can be modelled with *in vitro* tests.
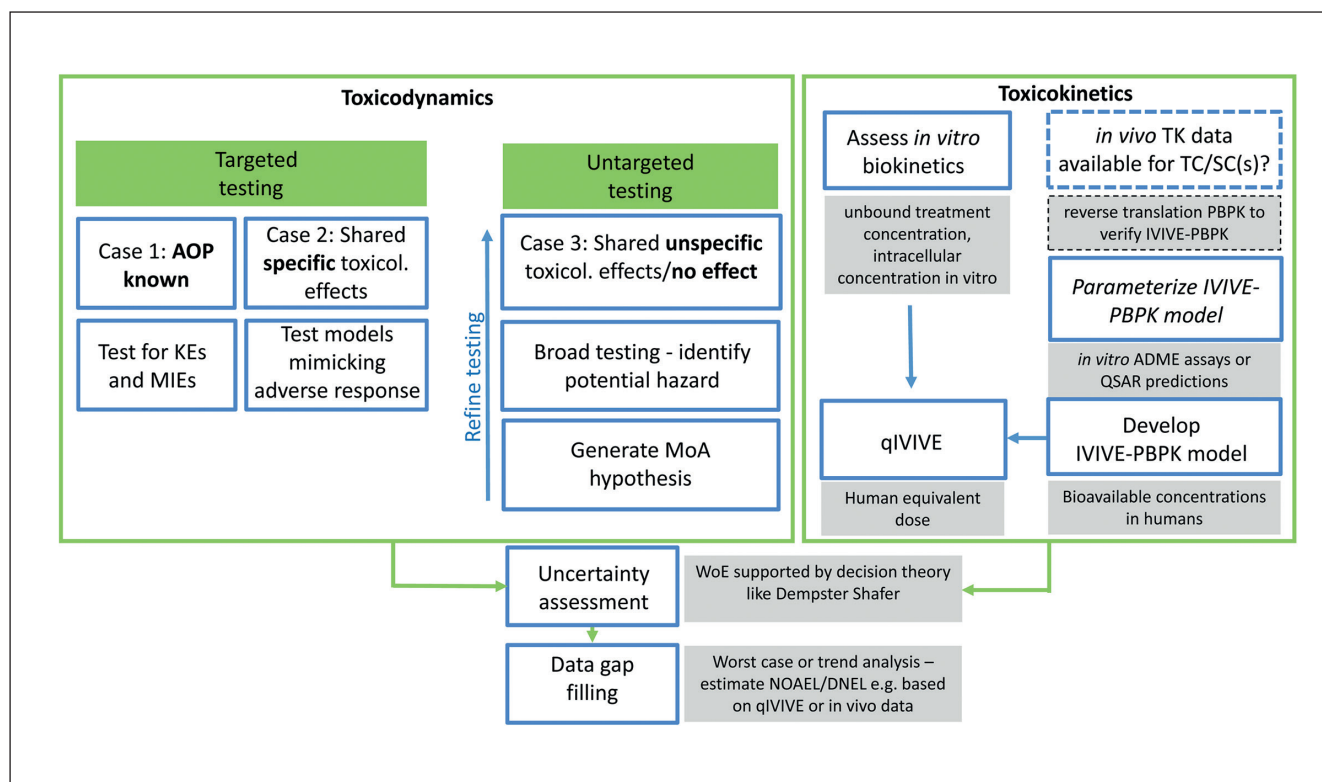
Fig. 1: Generation of NAM data based on the RAx hypothesis (adopted from Escher et al., 2019)

2. *Shared specific toxicological effects*. Source substances share a specific and known lead effect on humans or target organisms. In this case, the applied NAMs are focused on specific organ toxicities.

3. *Shared unspecific toxicological effects/no effect*. One example of unspecific general effects is significant loss of organ or body weight, which are considered adverse, but cannot be directly tested with *in vitro* models. This is the most challenging scenario, because it includes the absence of either a measurable specific toxicological effect or target organ. It requires the application of broad experimental testing to reduce the uncertainty of missing a toxicity endpoint that can be present in the target substance.

NAM data generated within the three scenarios should be complemented with toxicokinetic data that are necessary for the confirmation of any RAx hypothesis, defining category boundaries and characterizing dissimilarities between target and source substances. This is also necessary to derive a point of departure (PoD) for risk assessment, e.g., a human equivalent dose using a quantitative *in vitro* to *in vivo* extrapolation (IVIVE) model. Quantitative IVIVE needs an assessment of *in vitro* biokinetics and an estimate of the bioavailability per compound in the human organism (Fig. 1). The evaluation of the PoD may also show that the group behaves in a similar manner or trend, providing evidence on the place of the target substance within the group.

An important step before concluding the RAx with the final data gap-filling is to perform an uncertainty analysis to outline remaining limitations with regard to the RAx hypothesis and justification. In the best case, uncertainty analysis eventually quantifies the impact of the limitations on both the hazard prediction and determination of the PoDs, e.g., NOAEL or benchmark dose lower confidence limit(s) (BMDL) for the target compound(s). Depending on the problem formulation and therewith the regulatory area, conclusions might be different.

The EU-ToxRisk project is also preparing a template to organize the necessary data for a good RAx justification in a reporting document (still under evaluation). This template is based on the template used by the OECD IATA Case Studies Project (OECD, 2016, 2017, 2018a) as well as the ECHA RAAF document (ECHA, 2017). The project's external regulatory advisory board (RAB) has provided valuable input after the definition of the first case studies. The RAB comprises regulatory representatives from several EU member states in addition to other international agencies such as EFSA, ECHA, US Environmental Protection Agency (US EPA), US Food and Drug Administration (US FDA) and others.

At the time of the workshop, the EU-ToxRisk project had been working on nine case studies, three of which were presented for further discussion in breakout groups. The other two case studies discussed here were selected from the OECD IATA Case Studies Project[1] (OECD, 2018b, 2019).

---

[1] http://www.oecd.org/chemicalsafety/risk-assessment/iata-integrated-approaches-to-testing-and-assessment.htm

2 Case studies presented at the workshop

2.1 EU-ToxRisk case study – Prediction of a 90-day repeated dose toxicity study (OECD TG 408) for 2-ethylbutyric acid using a RAx approach to other branched carboxylic acids

This case study considers a group of aliphatic carboxylic acids, some of which are known to cause liver toxicity and steatosis. The aim was to use RAx to fill the information gap for a 90-day repeated dose toxicity study (OECD TG 408) for 2-ethylbutyric acid (2-EBA) using other branched carboxylic acids as source substances. KEs were combined in an AOP network for liver steatosis, and corresponding *in vitro* assays were applied to support the RAx hypothesis. The information was integrated as part of a weight of evidence (WoE) approach backed by Dempster-Shafer decision theory to aid uncertainty analysis. Subsequently, calculation of the human equivalent dose was performed to derive a point of departure (PoD) for risk assessment (publication in preparation).

Briefly, NAMs confirmed the *in vivo* results for the three analogues having *in vivo* animal data. The different tested human liver models confirmed the hypothesis of decreasing activity with side chain length, while IVIVE with PBTK modelling was successfully applied to derive a human equivalent dose for 2-EBA (data not yet published).

2.2 EU-ToxRisk case study – RAx based filling of developmental and reproductive toxicity data gap for 2-methyl hexanoic acid

This case study considers a set of substances similar to those analyzed in the above case study, i.e., aliphatic carboxylic acids, but the RAx is applied for a different endpoint, namely developmental and reproductive toxicity (DART). The target chemical, 2-methylhexanoic acid (MHA), is assumed to lack this specific toxicity data. Structural analogues were sought for which these associated data were known in order to explore the possibility to read across information of these source chemicals to MHA. The following structurally-related aliphatic carboxylic acids, which have *in vivo* DART data, were selected: 2-ethylhexanoic acid (EHA), 2-propylpentanoic acid (valproic acid, VPA), 2-propylheptanoic acid (PHA), 2-ethylbutanoic acid (EBA), 4-pentenoic acid (PA), 2-propyl-4-pentenoic acid (4-ene-VPA), and 2-dimethylpentanoic acid (DMPA). Some of these analogues are known to be developmental toxicants, including VPA, PHA, EHA, and 4-ene-VPA, whilst others had been identified as non-toxic to development, like EBA, PA, and DMPA, i.e., some of the analogues induced neural tube defects upon *in vivo* exposure, while others did not. Thus, structural similarity alone cannot allow a conclusive decision on the developmental toxicity of MHA. MHA and all selected source chemicals were tested in a battery of *in vitro* tests with clear relevance to DART, i.e., the zebrafish embryo test (ZET), the mouse embryonic stem cell test (mEST), the induced pluripotent stem cell (iPSC) based neurodevelopmental model (UKN1 test method), and a series of

CALUX reporter assays, which were combined with toxicokinetic models to calculate effective cellular concentrations and associated *in vivo* exposure doses. The aim was to explore whether they could correctly predict the respective *in vivo* developmental toxicity potentials of these aliphatic carboxylic acids, and thus could also be used to select the appropriate source chemical for RAx to predict the *in vivo* developmental toxicity of MHA. To further explore the relationship between structure and developmental toxicity within this series of aliphatic carboxylic acids, MPA was tested in the NAMs despite the fact that it lacked *in vivo* data. It is unclear what AOP underlies the *in vivo* neural tube defects observed with some of the analogue chemicals, but probably multiple AOPs are involved. Histone deacetylase inhibition was postulated as a critical, even initial, step in neural tube defects, so the carboxylic acid analogues were investigated for their potential to inhibit this enzyme. Analysis of the NAM results showed that VPA, PHA, EHA, and 4-ene-VPA were correctly predicted as *in vivo* developmental toxicants, and EBA and DMPA as non-developmental toxicants. The NAM results for MHA, for which *in vivo* data were negative, do not help to identify an appropriate source chemical for RAx, as they fail to fully resemble the NAM results of the non-developmental source chemicals.

2.3 EU-ToxRisk case study – Mitochondrial complex-III-mediated neurotoxicity of azoxystrobin. RAx to other strobilurins

Strobilurins are a family of fungicides that are active as mitochondrial Complex III (CIII) inhibitors. According to the PPP regulation, neurotoxicity assessment with OECD TG 424 is only triggered when relevant clues are observed during systemic toxicity testing or in case of structural analogy with a known neurotoxic compound. However, from literature there are some signals of potential neurotoxicity from *in vitro* studies by a mitochondrial CIII-mediated mechanism. In this case study, RAx is applied to support the hypothesis that the target substance is not a neurotoxicant via a CIII-mediated mechanism.

The group of strobilurins shares the strobilurin core structure but varies with respect to the substitution pattern of the aromatic ring. As a consequence, they also show some variation with regard to their physico-chemical properties. Their common feature is that all are presumed to undergo extensive metabolism and fast excretion. The formation of the category was based on a similar pesticidal mode of action, the presence of a common toxophore, and the availability of *in vivo* neurotoxicity data. The hypothesis is that they also have similar neurotoxic potential and similar toxicokinetics to justify RAx for the target substance azoxystrobin.

Existing regulatory *in vivo* data was collected for the source and target compounds with a focus on ADME, neurotoxicity, and target organ toxicity data. The source compounds show no signs of activity either in neurotoxicity studies or in other repeat dose toxicity studies. The question was whether the absence of a neurotoxic potential (as detected with a protocol based on OECD TG 424 as a guideline for neurotoxicity study in rodents) mediated by in-

hibition of mitochondrial CIII could be predicted by toxicodynamic and toxicokinetic NAM data. The hypothesis was supported by mechanistic data, anchored to a putative AOP and underpinning kinetic PBTK data.

In spite of the big difference in chemical structure, initial 3D evaluation revealed the importance of the methoxyacrylate group in the components of the category. By using test methods addressing key events of an AOP focused on neuronal degeneration as well as mitochondrial dysfunction due to CIII inhibition, it was demonstrated that the group had no activity on these endpoints. As a consequence, neurotoxicity can be excluded with enough justification to waive a new *in vivo* test. Kinetic data and simulations confirmed comparable kinetics with limited exposure of the brain to the strobilurins.

2.4 EPA/Health Canada case study – Case study on the use of integrated approaches for testing and assessment (IATA) for estrogenicity of the substituted phenols

This case study summarized the outcome of a collaboration between US EPA and Health Canada as a contribution to the OECD IATA Case Studies Project (OECD, 2018b). The idea was to demonstrate that *in silico* and *in vitro* data can be used to screen for estrogenic potential of chemical substances, and that these data sources provide a good proxy for estimating the *in vivo* PoD dose. A bi-directional approach was used: RAx between target and source analogues was conducted in the horizontal dimension (inter-chemical) analysis of the data matrices. Furthermore, data from many different streams (traditional and alternative) were tabulated, integrated and compared in the vertical dimension (intra-chemical).

The compounds under study were hindered/partially hindered and unhindered phenols. The hypothesis was that non-hindered phenols were expected to be estrogenic, whereas hindered phenols were not. Three phenols that fell into these categories were selected as target substances, and candidate source analogues were identified from a large inventory of collected substances. Candidate source analogues were identified using two complementary approaches: a local similarity method (LSM) as well as a global similarity method (GSM). The estrogenic potentials of the three target chemicals were determined using an IATA that combined (Q)SAR approaches and data from *in vitro* and *in vivo* studies. (Q)SAR predictions were generated using selected publicly available and commercial models. The *in vitro* high throughput screening (HTS) data from multiple assays were combined into a consensus prediction of estrogenic potential. Extrapolation of HTS bioactivity to an estimated applied dose equivalent was performed through the application of reverse dosimetry. For the target substance that was predicted to show estrogenic potential, the applied dose equivalent (ADE) was compared to effect levels from traditional *in vivo* animal studies to demonstrate the utility of these HTS data for use during prioritization and assessment. The methods and application of the

IATA can broadly support prioritization setting for further evaluation as well as hazard characterization for risk assessment.

2.5 NIHS Japan case study – Case study on the use of integrated approaches to testing and assessment for testicular toxicity of ethylene glycol methyl ether (EGME)-related chemicals

This case study originates from the OECD IATA Case Studies Project (OECD, 2019), with the collaboration of the National Institute of Health Sciences (NIHS) in Japan.

When performing a risk assessment under the Japanese Chemical Substances of Control Law (CSCL), a screening assessment is first carried out to select priority assessment of chemical substances. Category assessment, like for instance DART, is not currently utilized in the screening assessment if animal studies are not available, yet it is recommended (Ministry of Economy, Trade and Industry et al., 2012).

This case study was developed to demonstrate how RAx can be applied to fill data gaps regarding reproductive toxicity endpoints for screening assessments under CSCL. A category approach was used to assess the testicular toxicity of ethylene glycol methyl ether (EGME)-related chemicals. Based on toxicity information for EGME and related chemicals, and possible adverse outcome pathway information on the testicular toxicity of EGME, the category members were defined as chemicals that are metabolized to methoxy- or ethoxyacetic acid, which are responsible for testicular toxicity. A Japanese chemical inventory was screened using the metabolism simulator of the Hazard Evaluation Support System (HESS)[2] to obtain metabolism information for EGME-related chemicals. Published data showed that chemicals that produce methoxy- or ethoxyacetic acid as metabolites possess testicular toxicity potential, suggesting that untested chemicals that are predicted to produce these toxic metabolites may also have this effect. Although the overall uncertainty of the case study was low, some of the original compounds were structurally diverse, and metabolic hydrolysis or dealkylation could produce additional toxic compounds that need to be explicitly considered. However, a database search for toxicity and metabolism information suggested that these possible metabolites do not affect the toxicity levels through different mechanisms of action.

The category approach was further extended to produce NOAEL values that were derived for category members without *in vivo* data on testicular toxicity. The NOAEL values were subsequently used to derive hazard classification/prioritization under the CSCL in Japan.

3 Questions for the breakout groups

Workshop participants were divided into groups to discuss a set of common questions that had been prepared in advance. The aim of the breakout sessions was to discuss and propose recommendations to critical elements that would be needed in a guidance

---

[2] https://www.nite.go.jp/en/chem/qsar/hess-e.html

document on NAM-enhanced RAx in order to identify and fill knowledge gaps. The discussion was intended to map out guiding principles for circumstances, areas or problem formulations where NAM-enhanced RAx is acceptable.

The questions were:

– What are the generic requirements for a NAM assay/outcome to be acceptable in a RAx justification?
– What are the requirements under the following RAx conditions?
  ○ When an AOP is known for the shared apical effects and target organs.
  ○ When MoA or specific shared apical effects and target organs are known.
  ○ When a MoA is unknown.
  ○ When the source chemical(s) hardly show(s) toxic effects, i.e., only at very high doses, for example up to the testing limit of 1000 mg/kg bw/day.
  ○ When *in vivo* data is sparse, e.g., when there is uncertainty for source compound(s) that all required effects are covered by *in vivo* data.
  ○ Possible use of additional safety factors.
– What data integration and analysis are required?

During this breakout group session, the experts started with the case study-specific considerations, followed by a discussion on the general requirements for the use of NAMs to support RAx, even though each study was not always suitable for answering all questions.

Acknowledging the need to strengthen RAx justification, the following sections highlight the pivotal learnings from the working groups. A more extensive account will be published separately as a main deliverable of the EU-ToxRisk project.

4 Report from breakout groups

4.1 Working group from case study: Prediction of a 90-day repeated dose toxicity study (OECD TG 408) for 2-ethylbutyric acid using a RAx approach to other branched carboxylic acids

This working group analyzed the RAx concept (Fig. 1) and used the case study as an example to discuss four frequently observed situations with ideas on how NAMs can support the RAx hypothesis.

The first situation, which corresponds to Case 1 in Figure 1, is when the AOP/AOP-network is well-established for the adverse outcome(s), and NAMs are used to confirm that the grouped substances all share the same mechanism of the lead effect(s). There was a discussion on how many AOP events should normally be investigated in order to support the plausibility of the RAx hypothesis. In case of a complex AOP network, there is no need to verify all steps. Establishing similarity in the response by testing the MIEs and a few key events close to the apical endpoint could be enough to support the plausibility of the RAx hypothesis. In summary, it was suggested to look for concordance in the MIE signature and/or the final KE before the apical endpoint. In case of one linear AOP, the KEs close to the apical endpoint should be

the most relevant ones to demonstrate the plausibility of the RAx hypothesis.

The second situation, which corresponds to Case 2 in Figure 1, is when the mechanism of toxicity of the main adverse effect is not known even though the substances show a lead effect in a relevant target organ (system). The analysis of the group of substances should be focused on the biological fingerprint of source and target substances relevant to the target organ (system). A battery of NAM tests has to cover the whole biology of the target organ (e.g., start with omics, continue with *in vitro* organ simulating assays, etc.), whilst acknowledging that dissimilarities in the results may be difficult to interpret. Concordance of the biological fingerprint of source and target compounds is seen as mandatory to reassure the RAx hypothesis. A description of the remaining uncertainty is needed if the testing battery does not cover the whole organ biology/function. Beside qualitative comparison, potency needs to be addressed in the assays to demonstrate why and how the battery of tests proves that the potency of source and target compounds are similar or represents a worst-case approach.

The third situation, which is very common in practice, considers multiple effects caused by chemicals, i.e., a main effect to determine the lowest observed adverse effect level (LOAEL) and other effects at higher doses for which the mechanism is unknown. All effects seen in the *in vivo* study on the source chemical are to be read across for regulatory purposes. Hence the RAx justification must cover these as well as the lead effect(s). The level of assessment is, however, context-dependent. It might either be appropriate to use NAMs to evaluate non-lead effects (at higher dosing) or to consider them in a WoE approach based on, for example, the *in vivo* data.

The fourth situation, which is covered by Case 3 in Figure 1, is about substances for which no toxicological effects are observed up to the limit dose (1000 mg/kg bw/day). In this case, it is very important to understand the cause of the absence of effects in the *in vivo* studies. The first step is to examine all relevant existing *in vivo* and *in vitro* test results for signals or alerts that could indicate differences in the toxicological properties of target and source chemicals. If there is no evidence to refute the RAx hypothesis, further investigation may be necessary. There was discussion about using a broad testing battery to screen for "general" toxic effects to highlight any dissimilarity between source and target chemicals. It was acknowledged that it would be challenging to assess concordance, because dissimilarities, e.g., in the biological fingerprints, might be difficult to interpret, and an additional challenge is to present complex, extensive NAM data in an understandable and a meaningful way

A variation of the fourth situation, which is also covered by Case 3 in Figure 1, applies when source substances have sparse, unspecific effects. As above, the contribution of biological fingerprinting may help to unravel the situation even though it should be acknowledged that the interpretation of the results from a battery of *in vitro* tests can be really challenging.

Another topic of discussion was how to deal with "black swan" uncertainty in RAx, i.e. "unexpected" toxic effects that are not indicated by the *in vivo* data of source compound(s). It

may be that a well-designed battery of tests to screen for "general" toxic effects, as discussed above, would partly mitigate this concern.

In all cases, toxicokinetics play an important role in establishing (dis)similarity among chemicals. PBTK modelling was seen as a useful tool to support the RAx justification by demonstrating trends. Nevertheless, data gap-filling is often a standard data requirement, e.g., under REACH, therefore also rat PBTK models are necessary. Further investigation into the applicability of PBTK modelling to chemical risk assessment, the impact of *in vitro* parameters on the model estimates, and quantification of uncertainties like the confidence intervals of human equivalent doses need to be provided to gain confidence in this approach.

Metabolism has to be considered to investigate the difference between situations in which the parent substance and/or metabolite(s) trigger toxicity. Within this analysis, it is questionable to which extent the identification of metabolic pathways and/or a full spectrum of predicted metabolites is needed. The relevance of metabolism for RAx justification has to be explained, e.g., how metabolism is linked to structural similarity. Further tools and approaches are needed to better integrate and visualize metabolic pathways and first-generation metabolites in an understandable way.

4.2 Working group from case study: RAx based filling of developmental and reproductive toxicity data gap for methyl hexanoic acid

A conclusion of the group discussing the DART data gap for branched carboxylic acids was that the preliminary RAx hypothesis, source compound identification, and RAx justification should be performed in an iterative cycle until the required minimum level of certainty is reached. Characterization of the target substance should include detailed analytical characterization, physico-chemical properties, and any other data that is available. The rationale for selection of source chemicals should be transparent.

There are two possible roles that NAMs may play: biological justification of the category with demonstration of similarity and/or filling in missing information, representing in this case a real replacement of the *in vivo* test. A careful selection and use of NAMs is helpful to reduce the uncertainty and fill data gaps. The identification of the regulatory context and the associated information requirements form the basis to set up a strategy and define the acceptance threshold, which depends on the specific problem. For example, the level of acceptable uncertainty in the assessment can be higher when applied to the evaluation of impurities or minor components, while the highest possible accuracy may be required for demonstration of the possible toxicity of the main chemicals.

Assessment of uncertainty is essential, with transparent statements when quantification is not possible, covering all the different aspects, including category formation, metabolism predictions, choice of NAMs and so on. From this perspective, it is important that also newly generated data should contain confidence intervals. Trend analysis within a category should be backed by statistical analysis. In case this is not possible, it has to be as ob-

jective as possible with the demonstration that the input from expert judgement was unbiased. The RAx as such should be valid, and the use of an additional assessment factor (AF) for RAx to compensate for uncertainty is not scientifically justified at this stage. Analyses of the RAx hypothesis should be independent and reach a precise conclusion, while AFs are applied during a later stage of the risk assessment.

Beyond the scientific question, it was discussed to what extent a RAx justification is worthwhile, considering that it can be very demanding in terms of time and resources.

4.3 Working group from case study: Mitochondrial complex-III-mediated neurotoxicity of azoxystrobin RAx to other strobilurins

Identification of source compounds can allow either an analogue approach or a category approach. The analogue approach can be easier to justify, but the category approach can lead to a more robust scientific justification if the initial hypothesis is well demonstrated. In both cases, NAMs can support the suitability of the source chemical(s), even though no generic rules for acceptability can be made and which therefore needs to be defined case-by-case. NAMs are fundamental to prove that the source compounds belong to a category, i.e., to define the boundaries as well as the definition of trends and/or outliers.

In the case study of neurotoxicity of strobilurins, a shared toxicophore with the common propensity for the same activation of a MIE/KE in an AOP was the basis for the RAx. There are two aspects for the definition of the most suitable AOP to represent the category. If there is a clearly defined hypothesis, then testing along one single AOP or MoA is enough. If the hypothesis is broad, for example when there is a need to cover full systemic toxicity, NAMs should cover a broader range of mechanisms to represent the general toxicological profile of the substances. The selected NAMs have to be consistent with the biological mechanisms that need to be demonstrated, but they should also have a defined accuracy to help assess uncertainty.

The situation is different when the RAx is applied to substances with a very low toxicological profile, because there is no starting point for the selection of the most suitable NAM. Another difficulty is the decision on how many tests are required to exclude the presence of any missed or hidden effects. In this sense, it can be helpful to concentrate on dissimilarities among substances rather than similarities, with the use of positive controls in order to demonstrate that a possible effect would have been identified.

The assessment of toxicokinetics is important for the determination of the PoD, with data on the nominal versus extracellular concentrations used to generate the basis for the IVIVE and get to the final input for risk assessment. Regarding this specific example, the main source of uncertainty was the assessment of metabolites derived from different functional groups that are present in the molecule and that may exert another type of toxicity that was not considered in the selection of the source chemicals.

The strategy for NAM selection should be carefully calibrated, with little use for tests that may predominantly add noise to the statistical evaluation of the results. Assessment of the uncertainty

and its possible impact on the final results is fundamental in relation to the RAx hypothesis, and it should address aspects such as the influence of the source substance selection on the final conclusion. Quantitative analysis is often difficult if not impossible. Statistical tools can provide valid support, but they cannot cover all the different aspects (EFSA et al., 2017a,b, 2019). However, it is also true that, in many cases, reliability of the *in vivo* tests is not known, and another difficulty may derive from the presence of inconclusive *in vivo* data among source substances.

New NAMs have to be well-documented, for example by following OECD GD 211 (OECD, 2014), with suitable positive/negative controls that should be relevant to the specific issue, i.e., the RAx hypothesis that they are supposed to support.

4.4 Working group from case study: EPA/Health Canada Case study on the use of integrated approaches for testing and assessment (IATA) for estrogenicity of the substituted phenols

The selection of source compounds may derive from an incorrect hypothesis and can introduce some degree of bias, indicating the need for an appropriate problem formulation. Analogues should be evaluated with respect to a number of similarity contexts such as chemical structure, reactivity, metabolic pathways, physico-chemical properties or toxicokinetic information. Ideally, an approach to identify analogues should try to address all similarity conditions simultaneously to broaden the search and then filter analogues on the basis of the most relevant aspect. Quantifying the impact of each similarity context as it pertains to a toxicity endpoint under consideration is important, e.g., what weight/contribution do physico-chemical properties play in modulating the observed toxicity relative to structural or metabolic considerations.

It is important to address metabolism, as it is estimated that chemicals can exert their effect following metabolic transformation, and to establish whether/which NAMs are applicable in these cases. *In silico* tools will be useful to support the application of NAMs by identifying potential transformation products and to narrow the scope from many potential metabolites to the specific relevant ones. Sufficient coverage of transformation rules and applicability for that purpose need to be established.

NAMs representing the KEs of a known AOP are useful to represent a prediction model, but the risk is to miss unexpected behavior. In some cases, a WoE approach using a battery of different assays might be more fit-for-purpose; in this case, the selection of the most suitable set of tests should be made carefully to yield a good coverage of endpoints with an acceptable level of uncertainty with a minimum number of experiments.

Investigation around a specific AOP or MoA is very useful when the question is related to that mechanism, as is the case for prioritization of chemicals in view of a defined concern. However, the approach needs to be adapted for cases of unspecific effects or in the absence of effects, which might prevail in practice, depending on the regulatory context. In some cases, the identification of a minimum or median PoD threshold based on NAM data, a sort of threshold of toxicological concern (TTC)-like approach, can help in the risk assessment of a substance.

Quantifying uncertainty is fundamental to overcome qualitative and subjective levels of confidence (Schultz et al., 2019). If the data are robust, the assessment could start from a Dempster-Shafer analysis. Characterizing the variability of *in vivo* toxicity data is key to benchmark the performance of a RAx approach. For example, it can be challenging to apply RAx on an *in vivo* 90-day repeated toxicity study when there is no explanation to link the effects reported in the available *in vivo* study and the real PoD.

In practice, there is a need to improve the interpretation of NAMs, and this may come from the adoption of biomarker qualification, even though the problem is not necessarily solved through the application of transcriptomics techniques that, on the other hand, may add complexity.

4.5 Working group from case study: NIHS Japan on the use of integrated approaches to testing and assessment for testicular toxicity of ethylene glycol methyl ether (EGME)-related chemicals

In this case study, NAMs were used for the prediction of metabolism. The proposed metabolic pathway/AOP formed the basis for the justification of the chemical category hypothesis. Strong aspects were clear hypothesis formulation, specific purpose, well-defined inclusion/exclusion criteria for category members, well-defined applicability domain, and good discussion on uncertainties, with further NOAEL or benchmark dose (BMD) suggested. It is important to note here that the amount of data sufficed to justify the case without an overload of unused and confusing supplementary information. The identified weakness was the lack of toxicokinetic data.

Specific generic requirements were identified for a NAM assay/outcome to be acceptable in a RAx justification. If an AOP is used, a clear description of the pathways involved, a list of KEs for testing with the reason for selection as well as data on positive controls are necessary to justify their use for the investigated endpoint. The assays applied to represent the KEs should adequately mimic the *in vivo* MoA, with specific emphasis on the relevance of the AOP. The *in vitro* assays should be carefully chosen, justified, and optimized. Strength in numbers may not be the most appropriate approach, but reproducible assays are essential. Finally, assays need to be appropriate for the test substances, i.e., the substance needs to be within the applicability domain of the assay.

It is also important to understand the requirements for the negative prediction of RAx. Acceptance of negative predictions tends to encounter more skepticism, whereas positive predictions are often more readily accepted. Nonetheless, well-known and studied chemical categories allow higher confidence in negative predictions. Developing new case studies with low toxicity substances may lead to better acceptance, although there are already studies with plausible justification but their acceptance yet remains poor. Kinetic information indicating low absorption, rapid metabolic degradation or rapid excretion could be important elements for justifying low toxicity. Further, NAMs can be used to reduce safety factors. Recognizing uncertainties in non-benchmarked "gold standard" *in vivo* tests is needed to provide more accurate comparisons and hopefully increase acceptance.

Relevant data integration and analysis approaches are also important considerations for utilizing NAMs in risk assessment approaches. Transparency is needed, and often already available through suitable documentation (OECD, 2018a,b, 2019), but method acceptability should not be limited to data availability alone. "Acceptable" for one framework does not imply universal acceptability. In most cases, NAM approaches are considered "acceptable" when they are compliant with regulatory needs and data requirements of the specific regulatory program/process.

An overabundance of available NAM data without context makes the acceptance and understanding of new methodologies more difficult and confusing. Instead, tailored data including the proper reasoning for the proposed selection is preferred. Only relevant information should be provided with proper justification of the assays and tools that were used.

One should realize that we are working on replacing *in vivo* studies, which have unknown but accepted uncertainties. NAMs also have unknown uncertainties, but as these methods are new, their concomitant uncertainties are not (yet) generally accepted. One should be aware that there are also uncertainties connected to the GLP/OECD/*in vivo* studies in the source data. Acknowledging this would help understand and accept NAM uncertainties.

5 Conclusion

Workshop discussions were valuable in many respects, leading to shared conclusions on few, but important, issues for RAx. There is no single solution befitting all problems, therefore each case needs careful evaluation and definition of the most suitable strategy.

Biological similarity among chemicals is a scientifically acceptable concept, but its application requires robust justification. Participants acknowledged the importance of NAMs in supporting the RAx hypothesis and reducing the uncertainty of the final conclusion. NAMs can be highly useful in defining category boundaries and supporting the absence of activity cliffs, which is necessary to exclude toxicological concern. Another important area is the role that NAMs play in terms of kinetics and metabolism, which may have a strong impact on the biological effects of chemicals in organisms.

It was also recognized that there are a number of specific points that should be considered. One is the selection of relevant new tests. There was common agreement that too many tests are not helpful and instead may increase complexity and difficulty of data interpretation. It is best to select fewer reliable and relevant assays from a precise problem formulation and with a proper justification of use. Demonstration of low toxicity is more challenging in terms of regulatory acceptance, and confirmation is required for testing with assays embracing the assessment of apical endpoints related to multiple organ response. Kinetic information indicating low absorption, rapid metabolic degradation or rapid excretion could be important elements for justifying low toxicity.

For risk assessment, there is the additional need to define the PoD of the most sensitive test and correlate it with the NOAEL that is measured in an *in vivo* study, representing the starting point for derived no effect levels (DNELs). This procedure requires a careful IVIVE analysis.

Finally, the quality of the new tests is fundamental, and results should be provided with confidence limits for a trustworthy conclusion.

References

Benigni, R., Battistelli, C. L., Bossa, C. et al. (2019). Evaluation of the applicability of existing (Q)SAR models for predicting the genotoxicity of pesticides and similarity analysis related with genotoxicity of pesticides for facilitating of grouping and read across. *EFSA Supporting Publications 16*, 1598E. doi:10.2903/sp.efsa.2019.EN-1598

Daneshian, M., Kamp, H., Hengstler, J. et al. (2016). Highlight report: Launch of a large integrated European in vitro toxicology project: EU-ToxRisk. *Arch Toxicol 90*, 1021-1024. doi:10.1007/s00204-016-1698-7

ECHA (2016). New Approach Methodologies in Regulatory Science. Proceedings of a scientific workshop. Helsinki, 19-20 April 2016. doi:10.2823/543644

ECHA (2017). Read-Across Assessment Framework (RAAF). doi:10.2823/619212

ECHA, EFSA, JRC et al. (2018). Guidance for the identification of endocrine disruptors in the context of Regulations (EU) No 528/2012 and (EC) No 1107/2009. *EFSA J 16*, e05311. doi:10.2903/j.efsa.2018.5311

EFSA (2016). Guidance on the establishment of the residue definition for dietary risk assessment. *EFSA J 14*, e04549. doi:10.2903/j.efsa.2016.4549

EFSA, Benford, D. Halldorsson, T. et al. (2017a). Guidance on Uncertainty Analysis in Scientific Assessments. *EFSA J 16*, e05123. doi:10.2903/j.efsa.2018.5123

EFSA, Benford, D., Halldorsson, T. et al. (2017b). The principles and methods behind EFSA's Guidance on Uncertainty Analysis in Scientific Assessment. *EFSA J 16*, e05122. doi:10.2903/j.efsa.2018.5122

EFSA, Hart, A., Maxim, L. et al. (2019). Guidance on Communication of Uncertainty in Scientific Assessments. *EFSA J 17*, e05520. doi:10.2903/j.efsa.2019.5520

Escher, S. E., Kamp, H., Bennekou, S. H. et al. (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. doi:10.1007/s00204-019-02591-7

Ministry of Economy, Trade and Industry (METI), Ministry of Health, Labour and Welfare (MHLW) and Ministry of the Environment (MOE) (2012). Technical guidance for risk assessment of priority assessment chemical substances based on Chemical Substances of Control Law. [Article in Japanese]. http://www.mhlw.go.jp/seisakunitsuite/bunya/kenkou_iryou/

iyakuhin/topics/dl/tp120725-01-07.pdf

OECD (2014). Guidance Document for Describing Non-Guide-line In Vitro Test Methods. *OECD Series on Testing and Assessment, No. 211*. OECD Publishing, Paris. doi:10.1787/9789264274730-en

OECD (2016). Report on Considerations from Case Studies on Integrated Approaches for Testing and Assessment (IATA). First Review Cycle (2015). *OECD Series on Testing and Assessment, No. 250*. OECD Publishing, Paris. doi:10.1787/9789264274815-en

OECD (2017). Report on Considerations from Case Studies on Integrated Approaches for Testing and Assessment (IATA). Second Review Cycle (2016). *Series on Testing and Assessment, No. 270*. OECD Publishing, Paris. https://bit.ly/2J5iGPw

OECD (2018a). Report on Considerations from Case Studies on Integrated Approaches for Testing and Assessment (IATA). Third Review Cycle (2017). *Series on Testing and Assessment, No. 289*. OECD Publishing, Paris. https://bit.ly/3m5VrmO

OECD (2018b). Case Study on the Use of Integrated Approaches for Testing and Assessment (IATA) for Estrogenicity of the Substituted Phenols. *Series on Testing and Assessment, No. 290*. OECD Publishing, Paris. https://bit.ly/36XPFgo

OECD (2019). Case Study on the Use of Integrated Approaches to Testing and Assessment for Testicular Toxicity of Ethylene Glycol Methyl Ether (EGME)-Related Chemicals. *Series on Testing and Assessment, No. 308*. OECD Publishing, Paris. https://bit.ly/2Ku8Vuq

Rovida, C., Barton-Maclaren, T., Benfenati, E. et al. (2020). Internationalisation of read-across as a validated new approach method (NAM) for regulatory toxicology. *ALTEX 37*, 579-606. doi:10.14573/altex.1912181

Schultz, T. W., Amcoff, P., Berggren, E. et al. (2015). A strategy for structuring and reporting a read-across prediction of toxicity. *Regul Toxicol Pharmacol 72*, 586-601. doi:10.1016/j.yrtph.2015.05.016

Schultz, T. W., Richarz, A. N. and Cronin, M. T. D. (2019). Assessing uncertainty in read-across: Questions to evaluate toxicity predictions based on knowledge gained from case studies. *Comput Toxicol 9*, 1-11. doi:10.1016/j.comtox.2018.10.003

Costanza Rovida[1], Sylvia E. Escher[2], Matthias Herzler[3], Susanne H. Bennekou[4], Hennicke Kamp[5], Dinant E. Kroese[6], Lidka Maslankiewicz[7], Martijn J. Moné[8], Grace Patlewicz[9], Nisha Sipes[10], Leon van Aerts[11], Andrew White[12], Takashi Yamada[13] and Bob van de Water[8]

[1]Center for Alternatives to Animal Testing, CAAT-Europe, University of Konstanz, Konstanz, Germany; [2]Fraunhofer Institute for Toxicology and Experimental Medicine (ITEM), Hannover, Germany; [3]German Federal Institute for Risk Assessment (BfR), Berlin, Germany; [4]Danish Technical University, FOOD, Lyngby, Denmark; [5]Experimental Toxicology and Ecology, BASF SE, Ludwigshafen, Germany; [6]Department of Risk Analysis of Products in Development, TNO, Utrecht, The Netherlands; [7]Centre for Safety of Substances and Products, National Institute of Public Health and the Environment (RIVM), Bilthoven, The Netherlands; [8]Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands; [9]Center for Computational Toxicology & Exposure (CCTE), US Environmental Protection Agency, Research Triangle Park, NC, USA; [10]Division of the National Toxicology Program (DNTP), National Institute of Environmental Health Sciences (NIEHS), Research Triangle Park, NC, USA; [11]Medicines Evaluation Board (CBG-MEB), Utrecht, The Netherlands; [12]Unilever PLC, Sharnbrook, United Kingdom; [13]National Institute of Health Sciences, Kawasaki, Japan

(costanza.rovida@chimici.it)