REPORT FOR INTERNATIONAL COOPERATION ON COSMETICS REGULATION
Regulators & Industry Joint Working Group (JWG):
Integrated Strategies for Safety Assessment of Cosmetic Ingredients: Part 2

Authors:
Renata Amaral (1), Jay Ansell (2), Fanny Boisleve (3), Richard Cubberley (4), Matt Dent (4)#, Masato Hatao (5), Akihiko Hirose (6), Yutaka Kasai (7), Hajime Kojima (6)#, Petra Kern (8), Reinhard Kreiling (9), Beta Montemayor (10) Julcemara Gresselle de Oliveira (11), Andrea Richarz (12), Rob Taalman (13), Eric Vaillancourt (14), Rajeshwar Verma (15), Nashira Vieira O'Reilly Cabral Posada (11), Craig Weiss (16)

1) Brazilian Association of the Cosmetic, Toiletry and Fragrance Industry (ABIPHEC), Av. Paulista, 1313 Cerqueira César, São Paulo, SP 01311-000, Brazil
2) US Personal Care Products Council (PCPC), 1620 L St. NW, Suite 1200, Washington, D.C. 20036, USA
3) Johnson & Johnson Santé Beauté France, Domaine de Maigremont, CS 10615, F-27106 VAL DE REUIL Cedex, France
4) Unilever Safety and Environmental Assurance Centre, Colworth Science Park, Sharnbrook, Bedfordshire MK44 1LQ, UK
5) Japan Cosmetic Industry Association (JCIA), Metro City Kamiyacho 6F, 5-1-5, Toranomon, Minato-ku, Tokyo, 105-0001 Japan
6) National Institute of Health Sciences, 3-25-26 Tonomachi, Kawasaki-ku, Kawasaki, 210-9501, Kanagawa, Japan
7) Kao corporation, External Relations & Government Affairs 2-1-3, Bunka, Sumida-Ku, Tokyo, 131-8501 Japan
8) Procter and Gamble Services Company NV, Temselaan 100, B-1853 Strombeek-Bever, Belgium
9) Clariant Produkte (DE) GmbH, Global Toxicology and Ecotoxicology, Am Unisys-Park 1, 65843 Sulzbach, Germany
10) Cosmetics Alliance Canada, 420 Britannia Road East Suite 102, Mississauga, ON L4Z 3L5, Canada
11) Brazilian Health Regulatory Agency (ANVISA), Gerência de Produtos de Higiene, Perfumes, Cosméticos e Saneantes, Setor de Indústria e Abastecimento (SIA) Trecho 5, Area Especial 57, CEP 71205-050, Brasília-DF, Brazil
12) European Commission, Joint Research Centre (JRC), Directorate for Health, Consumers and Reference Materials, Chemical Safety and Alternative Methods Unit, Via E. Fermi 2749, 21027 Ispra (VA), Italy
13) Cosmetics Europe, Avenue Herrmann-Debroux 40, 1160 Auderghem, Belgium
14) Health Canada (HC), Consumer Products Safety Directorate, Healthy Environments and Consumer Safety Branch, 269 Laurier Ave. W., Ottawa, ON, K1A 0K9, Canada
15) US Food and Drug Administration (US FDA), Office of Cosmetics and Colors, Center for Food Safety and Applied Nutrition, 5001 Campus Drive, College Park, MD 20740, USA
16) Independent Cosmetic Manufacturing and Distributors (ICMAD), 21925 Field Parkway, Suite 2015, Deer Park, IL 60010, USA

Co-chairs

Contents

1. Abbreviations

ADME. . .	Absorption, distribution, metabolism, and excretion
AOP. . .	Adverse outcome pathway
CoMFA.. .	Comparative molecular field analysis
CRISPR. . .	Clustered regularly interspaced short palindromic repeats
DPRA. . .	Direct peptide reactivity assay
DST. . .	Dermal sensitization threshold
ECHA. . .	European Chemicals Agency
ES. . .	Embryonic stem
GFP. . .	Green fluorescence protein
GRAP. . .	Good read across practice
HPLC. . .	High-performance liquid chromatography
HRIPT. . .	Human repeat insult patch test
HTS. . .	High throughput screening
IATA. . .	Integrated approaches to testing and assessment
ICCR. . .	International Cooperation on Cosmetics Regulation
iPS. . .	induced pluripotent stem
JWG. . .	*Ad hoc* joint regulators-industry working group
LC-HRMS. .	Liquid chromatography coupled with high resolution mass spectrometry
Met ID. . .	Metabolite identification
MIE. . .	Molecular initiating event
MoA. . .	Mode of action
MoS. . .	Margin of safety
NAM. . .	New approach methodology
NGRA. . .	Next generation risk assessment
NMR. . .	Nuclear magnetic resonance
NOEL. . .	No observed effect level
NOTEL. . .	No observed transcriptional effect level
OC. . .	Organs-on-chips
OECD. . .	Organisation for Economic Co-operation and Development
PBBK. . .	Physiologically based biokinetic modelling

PBK. . .	Physiologically based kinetic
PBPK. . .	Physiologically based pharmacokinetic modelling
(Q)IVIVE. .	(Quantitative) *in vitro* to *in vivo* extrapolation
QMRF. . .	QSAR model reporting format
QSAR. . .	Quantitative structure-activity relationship
RAAF. . .	Read-across assessment framework
REACH. . .	Registration, Evaluation, Authorisation and Restriction of Chemicals, a regulation of the European Union (Regulation (EC) No 1907/2006).
SAR. . .	Structure-activity relationship
SEURAT. .	Safety Evaluation Ultimately Replacing Animal Testing
TEF. . .	Toxicity-equivalency factor
TG. . .	Test Guideline
TTC. . .	Threshold of toxicological concern
UN GHS. .	United Nations Globally Harmonised System of Classification and Labelling of Chemicals

2. Background

The International Cooperation on Cosmetics Regulation (ICCR) held its ninth annual meeting (ICCR-9) November 5, 2015 in Brussels, Belgium. At this meeting, discussions related to the evolution of the working group on *in silico* methods/quantitative structure activity relationships (QSARs) and updates on alternatives to animal testing approaches took place. These led to the formation of a new joint working group, covering a holistic approach to identify modern methods and Integrated Approaches to Testing and Assessment (IATA), relevant to the safety assessments of ingredients used in cosmetics.

At the ICCR annual meeting (ICCR-11) July 2017 in Brasilia, Brazil, the Joint Regulators-Industry Working Group (JWG) issued its report “Integrated Strategies for Safety Assessments of Cosmetic Ingredients - Part I” (ICCR, 2017).

In brief, JWG Integrated Strategies for Safety Assessments of Cosmetic Ingredients – Part I summarizes major overarching principles for incorporating new approach methodologies (NAMs) into an integrated strategy for risk assessment of cosmetic ingredients (or ‘Next Generation’ risk assessment : NGRA), along with examples showing their usefulness to safety evaluation.

In the context of cosmetics safety evaluation, a NGRA is defined as an exposure led, hypothesis-driven risk assessment approach that incorporates one or more NAMs to ensure that the use of a cosmetic product does not cause harm to consumers.

The 4 main overriding principles are articulated in this definition:

- The overall goal is a human safety risk assessment
- The assessment is exposure led
- The assessment is hypothesis driven
- The assessment is designed to prevent harm (*i.e.* distinguish between adaptation and adversity)

The following 3 principles describe how a NGRA should be conducted:

- Following an appropriate appraisal of existing information
- Using a tiered and iterative approach
- Using robust and relevant methods and strategies

Finally, 2 principles for documenting NGRA are described:

- Sources of uncertainty should be characterized and documented
- The logic of the approach should be transparently and explicitly documented

3. Purpose

Part 2 of the ICCR Integrated Strategies for Safety Assessment of Cosmetic Ingredients JWG report is intended to provide some additional guidance to safety assessors on the types of NAMs that may be used in a NGRA.

4. Introduction

The NAMs/technologies described in this report do not represent an exhaustive list. However, the NAMs, techniques and approaches that are likely to be particularly useful in the development of a NGRA have been highlighted, along with a brief description of the value they could add, where they could be used, and what their strengths and limitations are.

In deploying any NAM, the overarching principles described in the Part 1 report should be considered to ensure that the risk assessment remains useful in the decision making process. The SEURAT-1 *ab initio* safety assessment workflow (Berggren *et al.,* 2017) was developed to illustrate how NAMs may be used to arrive at a safety decision for repeated systemic exposure to a cosmetic ingredient. Importantly, as descibed here, the workflow can be used in a way which exemplifies the principles described in the Part 1 report. This workflow represents a tiered and iterative approach, comprising the following:

Tier 0: Identification of use scenario and collection of existing information on ingredientto be assessed for risk.

Tier 1: Hypothesis formation for *ab initio* approach

Tier 2: Application of *ab initio* approach

It should be noted that this workflow (illustrated in Figure 1) constitutes neither regulatory guidance nor a prescriptive risk asssessment approach, but is rather intended to provide a useful framework in which to consider how some NAMs may be used alongside more conventional approaches and/or data to arrive at a risk assessment for a cosmetic ingredient, with particular emphasis on exposure considerations. The workflow is therefore similar to IATA (Worth and Patlewicz, 2016) and how different NAMs may be incorporated into this workflow as part of a decision making progress is

discussed below. Where approaches are mentioned in bold further information on the definition, strengths and limitations of these approaches are provided in Section 5.

As described in Figure 1, the first priority (Tier 0) is to identify not only the composition of the ingredient in question but also the use scenario. This enables the routes of exposure to be determined, the extent of exposure (which is very different between *e.g.* leave-on and rinse-off cosmetics), and the identification of consumer groups that may experience high levels of exposure. The reason these considerations are important is due to the exposure-driven nature of the risk assessment, whereby exposure information not only is central to the final risk assessment but also helps to determine the relevant types of data that may be required. Without a full understanding of the use conditions of the ingredient, problem formulation is therefore not possible. Once the ingredient to be assessed for risk and its use scenarios are identified, all existing information on that ingredient is collected to ensure that all available relevant knowledge and information is used. Indeed, following a comprehensive literature search, sufficient data may be identified to complete a risk assessment using more traditional data and approaches. Where existing information is not sufficient to complete a risk assessment, established approaches such as **read-across (Section 5.1)** based on available data for similar compounds or **exposure-based waiving (Section 5.2)** based on structural information should be considered. As stated in Section 5, it is imperative that the rationale for the read-across and the similarity of the analogues be justified and clearly documented. This may require substantiation with ***in silico* tools (Section 5.3)** (*e.g.* to predict physico-chemical properties or the common metabolites between the ingredient being evaluated and its analogue candidates). Depending on the level of confidence in these predictions, biological data (such as *in vitro* substantiation of these metabolite predictions or 'omics approaches to assess biological similarity) may also be used.

Where it is not possible to complete a risk assessment using either existing data, exposure-based waiving or read-across alone, a Tier 1 evaluation is triggered. The purpose of Tier 1 is to formulate hypotheses regarding the biologically relevant modes of action (MoA) that may be associated with the specific chemical exposure, recognizing that different effects or MoA of concern will require different hypotheses and assessments. It should be stressed that information that is appropriate to one evaluation may not be appropriate to others, so the NAMs that will be useful in establishing these hypotheses should be considered on a case-by-case basis. However, tools that are likely to be common across many risk assessments involve generating data to further understand the extent of exposure to the parent molecule and any relevant metabolites, and which tissues or organs are likely to experience the highest levels of exposure. Both **metabolite identification (Section 5.4)** and physiologically-based kinetic **(PBK) modelling (Section 5.5)** play a fundamental role in determining the exposure to possible toxicants (parent or metabolite), and may also enable the values used in exposure-based waiving (*e.g.* thresholds of toxicological concern (TTC) values) to be refined using approaches such as the internal TTC. If application of the internal TTC is not possible, the methods used to develop MoA hypotheses include ***in silico* tools (Section 5.3)** (*e.g.* to predict likely molecular initiating events (MIEs)) and ***in chemico* assays (Section 5.6)**. The findings of these lower-tier assessments can be confirmed, refined, or refuted and broadened to provide more biological coverage, as necessary, using relevant high throughput screening (HTS) assays. Furthermore, **'omics** data **(Section 5.7)** will also be critical at this level, and especially transcriptomics data coupled with informatic approaches to identify critical biological pathways of concern, especially those that were not predicted from the available data or *in silico* alerts. **Reporter gene assays (Section 5.8) and *in vitro* pharmacological profiling (Section 5.9)** may also provide valuable information on specific interactions that should be further assessed.

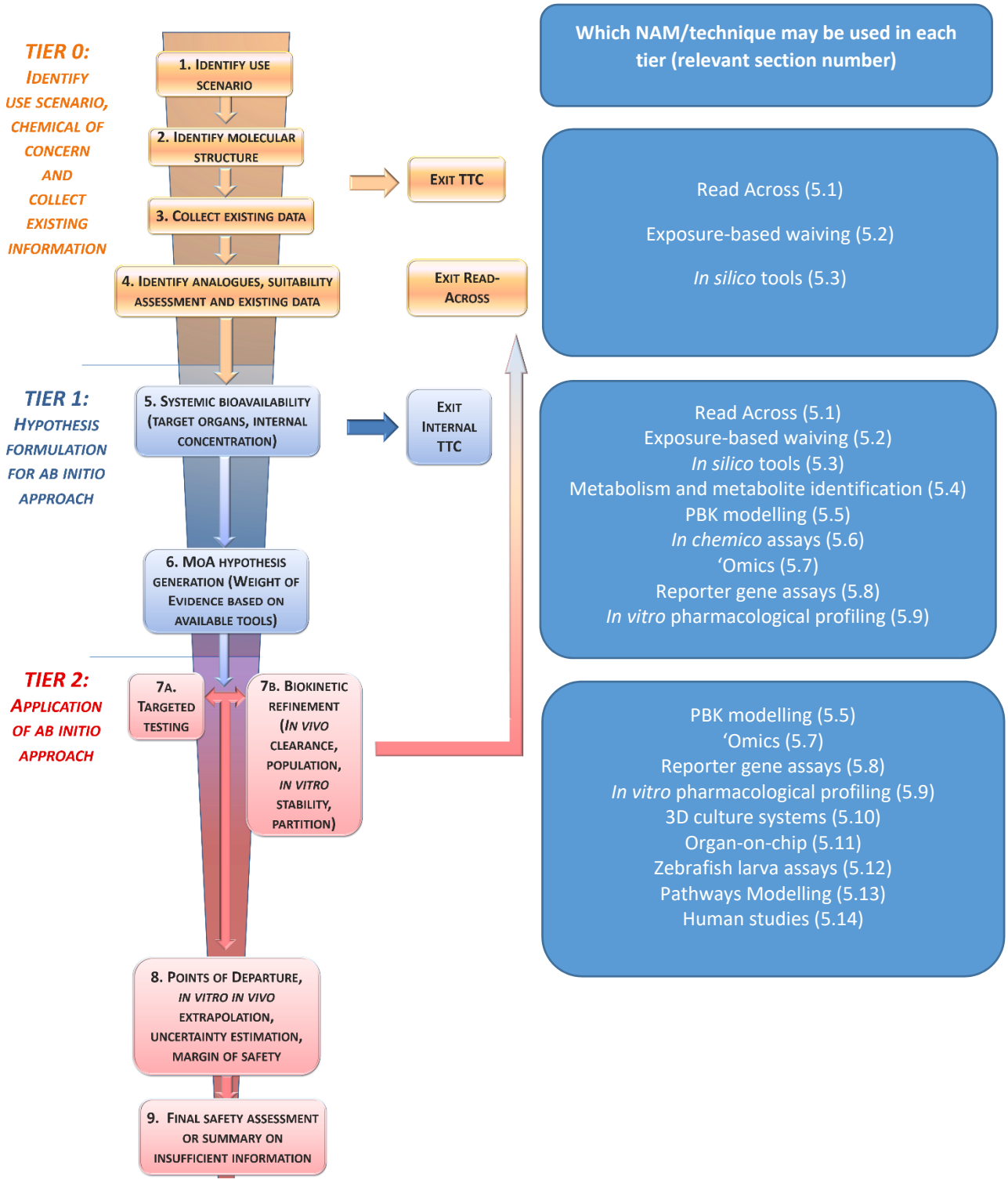

Figure 1: Aligning NAMs to different tiers of the SEURAT-1 *ab initio* workflow for systemic (repeat-dose) toxicity (*Berggren et al.,* 2017, Figure from OECD 2017 permission pending). When using this tiered approach not every NAM will be needed for every risk assessment, and some NAMs may only very rarely be used (see corresponding sections). NAMs may be used at more than one Tier; for example, NAMs such as *in vitro* pharmacological profiling or transcriptomics may be used to help identify critical modes of action for the risk assessment in Tier 1, but may also provide the critical point(s) of departure at Tier 2.

If existing data and/or HTS allow for critical pathways associated with the substance or its metabolites to be identified, appropriate dose-response assays relevant to those pathways can be used in Tier 2. The approaches used in Tier 2 will therefore be highly dependent on whether a specific MoA is identified at Tier 1, and whether the Tier 1 data are considered sufficient to provide the point of departure for the risk assessment. For many cosmetic ingredients, critical effects are likely to be due to generalized cellular stress pathways rather than a specific (*e.g.* receptor-mediated) MoA. Therefore, if very little biological activity is predicted at relevant concentrations, it may be sufficient to base the risk assessment on a weight of evidence (WoE) evaluation culminating in the use of a no-observed-transcriptional-effect level (NOTEL) derived from the transcriptomics dataset (Lobenhofer *et al*., 2004). Conversely, if specific pathways are identified for further evaluation, Tier 2 may include any number of tailored cell-based assays using 2D or **3D culture systems (Section 5.10)** or **Organ-on-Chip (Section 5.11)** technologies using approaches able to detect perturbations in components of those pathways. Where these cell culture systems are not appropriate to address the pathway/system of concern, whole organism models such as **zebrafish larva assays (Section 5.12)** may be required, although it should be recognized that many of these higher-tier approaches are not routinely used for characterizing systemic effects of cosmetic ingredients. Depending on the complexity of the pathway being evaluated, further interpretation using computational systems biology **pathway modelling (Section 5.13)** may be needed. The final step of the risk assessment is integrating the different lines of evidence to arrive at a conclusion.

Whether from an *in vitro* experiment or a computational model, the ultimate goal of these integrated approaches is to identify a quantitative point of departure for pathway perturbation that can be compared with the predicted human internal dose derived from PBK modelling to be used in risk assessment. Some examples showing how this can be done for dermally-applied substances have been documented (Adeleye *et al.,* 2015; Dancik *et al.,* 2015). Box and whisker plots are proving a useful tool to document the overlap of predicted exposures with relevant effect levels, as illustrated in the example shown in Figure 2 (Wetmore *et al.*, 2015), which compares the human oral dose that would be required to reach the AC_{50}s from approximately 700 *in vitro* assays for 163 chemicals (in green) with exposure predictions (in blue) for those chemicals. The chemicals that showed overlapping exposure and activity concentrations are indicated.

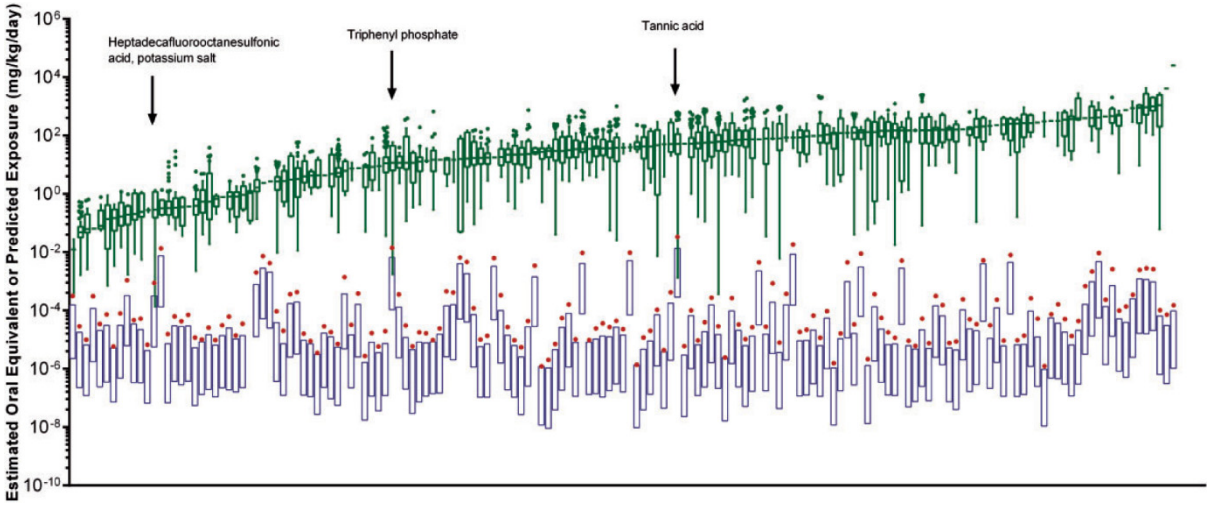

Figure 2: Box and whisker plot illustrating (Q)IVIVE (taken from Wetmore *et al.,* 2015, used with permission).

Consideration of the relevant dose metric is important in these plots, to assess whether the free concentration of the substance *in vitro* is similar to the free concentration *in vivo* and, if not, some correction should be included (Groothius *et al.*, 2015). Once a risk assessment has been completed, **human studies (Section 5.14)** may be needed to confirm tolerability of the formulation. Data from these studies can also be used to refine the models built and improve the overall integrated risk assessment.

Lastly, it is important that the safety assessment process, argumentation and collected data are transparently documented, and that sources of uncertainty, including data and methodological quality, and their possible impact on the assessment are adequately characterised and transparently documented.

5. New Approach Methodologies that may be used in the risk assessment of cosmetic ingredients

5.1 Read-across

Description

Although read-across is a well-established technique in toxicological risk assessment, it is included in this section because it is likely to prove to be a critical tool for performing NGRA and forming a weight of evidence to drive the formation of Mode of Action hypotheses. Read-across is one of the major data gap-filling approaches typically based on the "chemical similarity principle" to predict the toxicity of untested or poorly characterized chemicals. Chemical similarity, in read-across, is mainly established by "structural similarity", in a first instance, because chemical structure determines physico-chemical properties and therefore potential reactivity, both of which are key determinants of chemical interactions with biological systems. Thus, chemical similarity includes considerations about the structure and reactivity, the formation of common metabolites, as well as similarity of physico-chemical property(ies). Overall, toxicological similarity has to be considered. Using jointly both "*chemical similarity*" and "*biological similarity*" improves read-across predictions. The current state of the art of the read-across therefore contains a variety of well-defined and available scenarios to establish the *chemical similarity* and *biological similarity* between the *source* and *target* chemicals for a toxicological endpoint or health effect of interest (Ball *et al.,* 2016; Zhu *et al.,* 2016; Schultz *et al.,* 2015; OECD, 2014; Patlewicz *et al.,* 2013; Low *et al*., 2013; ECETOC, 2012; ECHA, 2008). Biological similarity comprises considerations of the degree of congruence or convergence of mode/mechanism of action (MoA) or adverse outcome pathway (AOP), or similar toxicokinetics.

In practice, the read-across technique is used within either analogue (one-to-one) or category (many-to-one) approaches.

The European Chemicals Agency (ECHA) has published a *Read-Across Assessment Framework (RAAF)* to facilitate the effective use of read-across for regulatory submissions (ECHA, 2017). The RAAF was developed to be used by ECHA in evaluating the acceptability of read-across proposals included in REACH[1] registrations. Therefore, REACH registrants can make use of the RAAF as a tool to help them assess the quality of their read-across assessments prior to dossier submission. In addition, an article "*Toward Good Read-Across Practice (GRAP) Guidance"* was published to assist the read-across practitioners, it presents the state of the art of "read-across", summarizes existing tools/approaches,

[1] REACH stands for Registration, Evaluation, Authorisation and Restriction of Chemicals, a regulation of the European Union (Regulation (EC) No 1907/2006).

possible prediction uncertainties, the use of biological data, and the impact of the ECHA RAAF (Ball *et al.,* 2016). Similar approaches have also long been applied and recognized in new and existing chemicals assessment frameworks in both Canada (*e.g.* Chemicals Management Plan) and the US (*e.g.* US TSCA Program). Recently, these topics have been exhaustively reviewed (Patlewicz *et al*., 2017, 2018).

Strengths

If performed correctly and documented transparently, read-across can be a powerful non-testing tool. It allows the assessment of groups of chemicals together, if appropriate similarity is determined, instead of all compounds individually, saving time and resources. Category approaches can maximize the experimental data (chemical space) to similar untested chemicals, thus resulting in the enhancement of prediction reliability. Therefore, there has been intensive effort in recent years to exploit the use of High Throughput/High Content screening to substantiate read across (Zhu, *et al*., 2016; Schultz and Cronin, 2017). Many practical tools are also available to facilitate the application of read across (Patlewicz *et al*., 2017), and a number of case studies exist illustrating its use (Blackburn *et al.,* 2011; Ball *et al.,* 2014; Schultz *et al.,* 2015).

Limitations

Chemical toxicity normally arises from complex biological mechanisms, so depending on chemical similarity alone may not be sufficient to fully justify the applicability or reliability of the read-across. This is especially the case when chemically similar compounds exhibit dissimilar toxicity profiles (Zhu *et al.,* 2016). Structural similarity can be subjective, and even minimal structural differences may sometimes lead to a very different biological activity. Failure to consider the implications of structural changes to the overall reactivity or more specific molecular interactions may lead to a flawed read-across justification. The careful consideration of overall toxicological similarity as discussed above, as well as transparent identification and documentation of possible uncertainties, should mitigate this risk.

Read-across is limited by the quality of toxicity data for the source chemicals, and can be difficult to be justified in the absence of relevant mechanisms (source compounds with non/low toxicity) (Schultz *et al.,* 2015). Uncertainty can be reduced by providing additional supporting information, by selecting appropriate *in chemico* methods, *in vitro* tests and/or *in silico* tools, (*i.e.* targeted testing) (Blackburn and Stuard, 2014).

Integration in the risk assessment

In Tier 0, read-across may serve two purposes. First, if it is possible to place an untested chemical within a category of tested chemicals, it may be possible to make a safety decision at Tier 0. Because the risk assessment is exposure led, under circumstances whereby the resulting margin of safety (MoS) is large enough to address the uncertainties described above, it may be possible to complete the risk assessment with sufficient confidence without generating further data. Overall toxicological similarity and thus also biological and biokinetic similarity should always be considered together with structural similarity (Blackburn *et al.,* 2011). On the other hand, even if there is no structural similarity, consideration can be given (depending on the regulatory context) as to whether purely mechanistic or biological similarity can be established. Techniques that could be useful include 'omics approaches (Zhu *et al.,* 2016), to allow chemically dissimilar molecules to be grouped together if they are biological similar (*i.e.,* if they display the same MoA or activate a particular AOP).

Second, where the risk assessment cannot be completed (*e.g.* no adequate analogues with sufficient data are identified, or the MoS between consumer exposure and the points of departure for the read-across candidates is too small), the considerations undertaken for the read-across may still be useful in helping to identify potential modes of action and therefore develop a hypothesis on how the chemical may interact with biological systems. Read-across can also be reconsidered for reaching a decision at a later tier, when more evidence becomes available.

5.2 Exposure-Based Waiving

Description

The TTC is a pragmatic approach that can be used in the absence of chemical-specific toxicity data to establish a level of exposure below which there is expected to be no appreciable risk to human health. It should be noted that use of TTC is not an alternative to searching for existing substance-specific data, but should be considered when scientific literature searches fail to identify adequate data upon which to base a decision. TTC is based on structure-activity relationships, and allows the comparison of a substance of known chemical structure and exposure to a group of compounds that are structurally similar. The approach does not involve QSARs for specific endpoints but rather is based on the distribution of potencies for chemicals that share similar broad structural characteristics with the chemical under evaluation (Dewhurst and Renwick, 2013).

Originally proposed as a pragmatic way to assess food packaging materials, Frawley (1967) began with the starting premise that there must be some uses of food packaging materials that do not present and appreciable risk to the health of the consumers of food. He then set about defining a baseline dose that could be considered to be "without harm" *via* the analysis of 2-year, chronic toxicity studies on 220 different chemicals given *via* the diet. The concept was further developed (Munro *et al.,* 1996), based on an extensive analysis of available chronic oral toxicity data of substances, which were divided into three chemical classes of toxic potential on the basis of their structure using the decision tree of Cramer *et al*. (1978). This decision tree categorizes chemicals, mainly on the basis of chemical structure and reactivity, into three classes indicating a high (Class III), medium (Class II) or low (Class I) level of concern.

TTC values were defined for the three classes using toxicity data for 613 compounds based on the following procedure: The cumulative distribution of the chronic oral no-observed-effect levels (NOELs) for the compounds for the most sensitive systemic toxicological endpoint in animal models were plotted and fitted with a log-normal distribution (Munro *et al.,* 1996), and the 5th percentile value for each NOEL distribution was selected. This provides a 95% confidence that the NOEL of any compound of unknown toxicity within a given structural class would be greater or equal to the 5th percentile value and, consequently, of an inferior toxicity potency. For each structural class (Cramer class), a TTC value was determined by dividing the 5th percentile value by an uncertainty factor (UF) of 100 (UFs of 10 each for interspecies and intraspecies variabilities). A TTC value is therefore defined as the maximal oral dose for which there is no appreciable risk to human health following lifetime daily exposure, and the TTC values for each Cramer class are: 30 µg/kg bw/day (Cramer class I), 9 µg/kg bw/day (Cramer class II), 1.5 µg/kg bw/day (Cramer class III).

For a given substance, the TTC concept applies equally whether it is intentionally added or is present in the form of inadvertent impurities/contaminants in a formulation, because they are toxicologically equivalent.

It should be noted that the aforementioned TTC values were derived from oral exposure data, because insufficient dermal data exist to develop a robust database. However, for substances with low dermal absorption, it may be more relevant to make comparisons using values derived from internal exposure (SCCS, 2016). Therefore, "*internal* TTC" values have also been defined for the same three Cramer classes (see details in Partosch *et al.,* 2015) and the internal TTC values based on the 5th percentile values are: 6.9 µg/kg bw/day (Cramer class I) and 0.1 µg/kg bw/day (Cramer class II/III). Similarly, because exposure to cosmetic ingredients can also occur from aerosol products, "*inhalation* TTC" values have been defined for two Cramer classes for both local and systemic effects (due to the paucity of chemicals falling under class 2 (*n*=4), no TTC values could be derived for that class) (Carthew *et al.,* 2009). The inhalation TTC values for systemic effects based on the 5th percentile values are: 16.4 µg/kg bw/day (Cramer class I) and 2.8 µg/kg bw/day (Cramer class III).

Useful *in silico* tools (*e.g.* Toxtree; OECD QSAR Toolbox) are publicly available to automate the task of determining the Cramer class of compounds of interest, based on the input of the chemical structure. However, their intent is to guide (not replace) expert judgment. Software may have difficulty interpreting some Cramer Rules, and this may lead to the misclassification of some compounds that may be out of the domain of applicability of specific models. It should also be noted that there may also be discrepancies in classification outcomes between different software packages. In cases where a Cramer classification (and its associated TTC value) is pivotal to a risk assessment, expert judgment should always be used to confirm the software-assigned classification (Bhatia *et al.,* 2015), and justification for the expert determinations documented.

Strengths

Even though the TTC approach was initially developed to assess the safety of dietary chemical contaminants, this method may similarly be appropriate to evaluate or prioritize the necessity for toxicity testing and assessment for compounds found in cosmetic ingredients or final formulations (including impurities or degradation products) (Kroes *et al.,* 2007).

A preliminary analysis of the applicability of the TTC approach to cosmetics concluded that the existing approach is generally applicable to cosmetics and personal care products (Worth *et al.,* 2012). Also, as demonstrated through the analysis of a dataset expanded from 477 to 709 rodent carcinogens (Cheeseman *et al.,* 1999), the TTC approach is robust. Similarly, a thorough analysis of the original database compiled by Munro *et al*. (1996) regarding the robustness of TTC values for non-cancer endpoints has been performed by the European Food Safety Authority (EFSA, 2012). Therefore, in theory, the TTC approach can be applied to evaluate the safety of any known chemical structure for which human exposure can be accurately and robustly estimated. The TTC approach has been improved through peer-reviewed scientific publications (*e.g.* Cheeseman *et al.,* 1999; Cramer *et al.,* 1978; Kroes *et al.,* 2004, 2007; Munro *et al*., 1996), and the original Munro dataset has now been expanded to include 552 cosmetics-relevant substances, and the TTC approach and values have been re-evaluated accordingly (Yang *et al.,* 2017). Further, the applicability of the TTC concept has been demonstrated to be an effective approach for risk assessment of skin sensitization with the Dermal Sensitization Threshold (DST) (Safford *et al*., 2015). In summary, TTC is now a globally recognized approach, and uses a transparent decision tree with explicitly documented datasets.

Overall, the strength of the TTC approach lies in its applicability at an early stage of the safety assessment, allowing a decision based only on known chemical structure and exposure.

Limitations

The three Cramer structural classes specifically exclude heavy metals (*e.g.* cadmium, lead; due to bioaccumulation), proteins (due to their potential for allergenicity, and due to the potent biological activities of some peptides), and the following compounds due to bioaccumulation, high-potency, genotoxic carcinogenicity, and an already established toxicity-equivalency factor (TEF)-based risk assessment methodology: aflatoxin-like compounds, azoxy compounds, *N*-nitroso compounds, and polyhalogenated-dibenzo-*p*-dioxins, -dibenzofurans and -biphenyls (EFSA, 2012; Kroes *et al.,* 2004; Renwick, 2005). Consequently, excluded compounds require substance-specific toxicity data. Therefore, as a first step, the target substance has to be evaluated for applicability of the TTC approach, and checked for potential genotoxicty or high-potency carcinogenicity, for example by identification of structural alerts. Apart from the exclusion criteria, specific, more conservative thresholds apply for potentially genotoxic substances or organophosphates (Kroes *et al.,* 2004).

Despite the conclusion that the TTC approach could be broadly applied to cosmetics and personal care products, as with all approaches continuous improvements are still possible, particularly regarding quality control of the original TTC datasets (in the interest of toxicological consistency and transparency), moderate revisions/extensions of the Cramer classification scheme (*e.g.* retinoids; aliphatic halides; aromatic amines; aromatic azo compounds; sulphonates; silicon-containing compounds; organometallics; non-ionic and cationic surfactants), and development of unambiguous guidance on how to apply the TTC approach (*e.g.* explicit inclusion/exclusion criteria; including respective adaption of specific software to be used; the use of pre-Cramer filters such as structural alerts or QSARs) (Barlow, 2005; Worth *et al.,* 2012). As outlined above, the interpretation of the Cramer rules can be ambiguous and can lead to deviating decisions by different software tools which may need to be resolved through expert judgement.

In addition, TTC values are derived from toxicological endpoints in animal models and, therefore, may not necessarily be representative of human biology or have been derived from specific, human-relevant, mode of action hypotheses. In particular, the original Munro dataset is based on oral toxicity data, and as the dermal exposure route plays an important role for cosmetics-related chemicals, differences of bioavailability may apply (see Yang *et al.,* 2017 and Williams *et al.,* 2016, respectively)

Furthermore, the described TTC approach is based on external exposure. For further refinement, the internal exposure at the specific relevant site in the body should also be considered.

Integration in the risk assessment

The TTC approach can be applied at an early stage of the risk assessment, based on exposure considerations. By modelling the chronic systemic dose for a given cosmetic ingredient (or impurity/degradation product), the outcome of an exposure scenario could be compared to the TTC value for that compound. In cases where the chronic systemic dose is lower than the TTC value, toxicological concerns could be excluded through this simple comparison. However, when the chronic systemic dose is greater than the TTC value, concerns for toxicity would *not* be excluded, and would thus require model refinement through higher tier approaches, prior to integration in a risk assessment designed to differentiate between adaptation and an adverse effect.

5.3 *In silico* tools

Description

In silico approaches can be used in many different parts of the risk assessment for different reasons, and comprise a wide array of common computational models, methods and tools. In addition to the specific approaches mentioned elsewhere in this document (*e.g.* tools for metabolite prediction, Section 5.4 or PBK modelling, Section 5.5) *in silico* tools include:

i) Structure-activity relationships (SARs) describe *qualitative* relationships between a chemical structure and biological activity. A SAR can simply take the form of a structural alert which identifies a unique feature in a compound that is associated with its biological activity. Structural alerts can be combined in profilers used for screening list of chemicals for potential activities;

ii) Decision trees are a specific form of SARs, describing for example discrete responses such as active/inactive, weak/moderate/strong. They can also be used in the random forest method, or to classify compounds in "Cramer classes" for application of the TTC (Section 5.2);

iii) QSARs describe *quantitative* relationships between a chemical structure and biological activity. These relationships can be expressed by either linear or non-linear algorithms, such as multiple linear regressions, partial least squares, artificial neural networks, support vector machine, clustering algorithms, etc. The acronym (Q)SAR is being used frequently in the literature and represents collectively both the SAR and QSAR models;

iv) Molecular modelling/docking approaches are used to model how small molecule ligands interact with target protein active sites and are based on the 3D-structure and conformation of molecules;

v) Grouping and read-across allows to fill data gaps for a specific toxicological endpoint when data are available for one or multiple structurally similar compounds for that same endpoint (see Section 5.1);

vi) Expert systems are formalised systems integrating diverse combinations of read-across, decision trees, structural alerts, QSAR, expert knowledge, and molecular fragment and descriptors;

vii) Databases (for physico-chemical properties and biological effects) as well as data mining tools can support safety assessment.

The relevance and use of computational approaches in cosmetic ingredient safety assessments have previously been reviewed by the ICCR (ICCR, 2014).

Many *in silico* approaches are collected in model databases or available as software tools and platforms, both freely available and commercial. Efforts have been undertaken to make models available in a harmonised format, in order to increase their transparency, reproducibility and improving their uptake. QSAR formats and databases include the QSAR Model Reporting Format (QMRF)[2] and EC JRC QSAR database, and the QsarDB Repository[3] (see Tetko *et al*., 2017). Quality considerations are included, *e.g.* the QMRF is structured according to the Organisation for Economic Co-operation and Development (OECD) validation principles (OECD 2004). To facilitate the consideration of QSAR models for regulatory purposes, OECD (2004) proposed that QSAR models should provide ***i***) a defined endpoint; ***ii***) an unambiguous algorithm; ***iii***) a defined domain of

[2] https://qsardb.jrc.ec.europa.eu/qmrf/

[3] http://qsardb.org

applicability; ***iv***) appropriate measures of goodness-of-fit, robustness and predictivity; and ***v***) if possible, a mechanistic interpretation. Various examples of (Q)SAR software tools are described in the ICCR report (2014).

Strengths

In silico approaches cover a wide range of chemical space and toxicological endpoints considered for cosmetics safety assessment (ICCR 2014). They are reasonably easy to apply, especially if implemented in software tools. However, it must be noted that not only their development but also their use requires some degree of expert knowledge and consequently should not be used indiscriminately as a black box.

Approaches such as SARs can be automated and allow for the screening of large datasets and alert on specific structural features, properties or potential biological activities which could pose concern during a human health risk assessment. They can thus be used as high-throughput tools. Furthermore, the use of techniques to mine large datasets may also enable the identification of relevant candidate analogues associated with the biological activity of interest and existing data.

Several approaches can also easily be combined, *e.g.* as a consensus model, or applied as a battery of models. This extends the applicability domain, and potentially serves to overcome possible limitations of individual models, or increase confidence in the prediction. Furthermore, combining models may allow screening for several structural features or biological activities at the same time.

Overall, computational models, ideally based on mechanistic understanding of the biological activity, can help to inform on the MoA of a target chemical and to formulate relevant hypotheses, and may allow quantitative predictions of toxic potency.

Limitations

The reliability and robustness of computational models inherently depend on the quality of the underlying data used to build and validate the models and which may be subject to biological variability or differences in experimental settings, species differences, *etc.* Similarly, if the dataset to build the model is too small, the model might not be representative.

The applicability domain always needs to be taken into account. The model should not be used for a prediction if it does not cover the chemical and toxicological space of the substance in question. There are often several models and/or software tools available for any given endpoint, and different tools may give conflicting results. Therefore, careful consideration is needed to decide which model or models may be most appropriate for the assessment in question, based on the adequate applicability domain and expert judgement. Generally, some level of chemical and toxicological expertise is needed for application of *in silico* models.

Another limitation of models derived from 2D structures is that the 3D structure could be relevant for the MoA. Models based on 2D structures cannot distinguish enantiomeric forms, and dealing with tautomeric forms is also an issue. However, there are specialist *in silico* systems, such as Comparative Molecular Field Analysis (CoMFA) (Cramer *et al.,* 1988), that can consider chirality and therefore deal with stereochemistry issues.

To make the models usable and reproducible in practice, the models, descriptors, the clearly defined applicability domain, algorithm, all limitations and uncertainties as well as the predictions obtained need to be completely and transparently documented.

Integration in the risk assessment

In silico models are ideally based on mechanistic understanding and, therefore, can help to inform on the MoA, for example if they are related to a MIE.

As a first step, data mining and structural alert profilers can be used to pre-screen libraries of target substances and give first indications of biological activity as well as find existing data for possible similar chemicals. Targeted predictions support the formulation of the MoA hypothesis and can also be used to support the further verification of putative adverse-outcome pathways.

In silico approaches also contribute to the assessment of other toxicity modulating factors, such as absorption, distribution, metabolism and excretion (ADME), or setting parameters for PBK models.

5.4 Metabolism and metabolite identification

Description

Following absorption, any xenobiotic is likely to undergo some form of metabolism, which may result in more or less toxic metabolites. Metabolism may occur in different sites; therefore although most focus is usually placed on the liver as the major organ of metabolism for xenobiotics, for dermally-applied cosmetics the effect of skin metabolism may need to be considered. Various approaches can be used to identify the rate of metabolism of a test item (clearance of the parent molecule) and formation of metabolites. The rate of metabolism is an important parameter to be used in PBK modelling, and can be determined experimentally by analysing the disappearance of the test item over time in hepatocyte or microsomal systems.

Guidance is available for metabolite identification and safety assessment in drug development (US FDA, 2016) but, to date, no such guidance is available for cosmetic ingredients.

In a tiered approach, metabolite identification may begin with *in silico* predictions using commercially available packages (see Section 5.3). The *in silico* approaches predict which metabolites may be formed and to some extent how likely they are to be formed. There may be the need for follow up in *in vitro* test systems using appropriate analytical techniques to confirm which metabolites are most likely to be formed in a biological system, which might depend on several parameters. The metabolism rate is also of particular interest, especially if a metabolite is the toxicant to be considered for safety assessment.

In vitro approaches to study metabolism range in complexity from subcellular fractions of liver cells (microsomes and S9) to a suspension of hepatocytes and on to 2D and 3D cultures of hepatocytes (sometimes combined with other cells) right up to slices of human liver. It is also possible to consider pre-systemic metabolism of dermally applied cosmetics using skin S9 preparations, although low metabolic activity means it can be difficult to characterize metabolic products. The most common approaches are therefore the use of microsomes and hepatocyte suspensions for which established scaling factors exist to estimate *in vivo* clearance rates.

To measure clearance, a physiologically relevant dose (typically 1 μM) is used and the concentration of parent chemicals is monitored over time using a suitable analytical method. The rate of depletion of the parent or the calculated half-life is used in clearance calculations. The formation of metabolites can also be monitored and metabolites quantified, if suitable analytical standards are available.

For metabolite identification (Met I.D.), it is not unusual to use a higher dose than in clearance work, to maximise the likelihood of detecting metabolites. Metabolite analysis is typically performed using liquid chromatography coupled with high resolution mass spectrometry (LC-HRMS) which enables very accurate masses to be determined, thereby providing important information on the molecular formula of metabolites. The combination of this formula along with knowledge of the structure of the parent chemical helps to determine the identity of the metabolites, although the exact site of metabolism is not always discernible. Further HRMS analysis to study the fragmentation pattern of the metabolite compared to the parent chemical can help further refine identification, as can the relative LC retention time. In some cases, isolation and enrichment of individual metabolites may be required with analysis by NMR for unequivocal structural I.D.

Strengths

Well established *in vitro* test methods are available to characterize both the clearance of parent material and the formation of metabolic products from Phase I metabolism. Identification of relevant metabolites would make a risk assessment based only on *in vitro* data much more robust, by ensuring that relevant metabolites are present in the *in vitro* systems used to assess safety.

The use of subcellular fractions such as microsomes and S9 provides a relatively cheap, simple approach to obtaining metabolic information on a given test chemical. The use of microsomes, in particular, is established in terms of scaling from a protein concentration to a whole liver. Because this is an acellular system, the toxicity of the test chemical and metabolites is not a concern and this means higher doses can be applied which may be particularly useful for Met I.D. work.

The use of human hepatocytes is generally favoured for *in vitro* metabolism work because the system is considered more representative of the liver *in vivo*, with applicability for medium and fast metabolic rates. For slow clearance chemicals, hepatocytes in culture can extend the duration of metabolism. Some of the more sophisticated cultures can introduce other cell types to better represent the environment in the liver and even the formation of bile canaliculi, adding complexity and cost but with the added benefit of being more representative of *in vivo*-like conditions.

Limitations

In silico models to predict metabolism provide a large number of possible metabolites, and it can be difficult to determine which are likely to be formed in significant quantities and therefore be relevant to the risk assessment.

For microsomes and S9, it is important to bear in mind that essentially these are just solutions of enzymes at a physiologically relevant ratio. The structural aspects of the cell which may influence the accessibility of these enzymes to the test chemical are not present, and this may have a significant influence on the metabolism observed. Additionally, neither microsomes nor S9 contain all the relevant enzymes. Microsomes contain Phase I enzymes and also those for glucuronidation, but little else. S9 contains most of the Phase I (albeit at a lower level than in microsomes) and Phase II enzymes, but is still missing certain pathways such as glycine conjugation. In both cases, the relevant co-factors are added at a level designed to maximize metabolism (*i.e.* in excess) and may therefore not truly represent the *in vivo* situation.

The limitation with hepatocytes in suspension is that metabolic activity is only maintained for up to 4 hours and there may be chemicals for which metabolism is too slow for a clearance rate to be determined, despite the formation of metabolites being observed. For longer duration metabolism in culture, more sophisticated co-cultures and 3D cultures are available (Godoy *et al*., 2013). Toxicity to

substrate and major metabolites should be considered and the dose adjusted accordingly with a means to monitor cell health added to the assay. For more complicated systems, the scaling of the measured *in vitro* clearance rate to the *in vivo* value is less well established.

With all the *in vitro* systems described above, interindividual variability is an important consideration. Although there is a good correlation between *in vitro* and *in vivo* clearance based on primary hepatocytes taken from the same donor (Ponsoda *et al.*, 2001), there is a significant variation between donors (Gómez-Lechón *et al.,* 2007). This should be considered when applying measured clearance rates to safety assessments.

Integration in the risk assessment

Depending on the risk assessment, metabolism and metabolite identification may be integrated at different points in the workflow. For example, if the hypothesis being tested is that the cosmetic ingredient will be converted *in vivo* to a well-studied chemical structure, some characterization of metabolic products could be used in Tier 0 to complete a read-across risk assessment. If an *ab initio* risk assessment is being performed, it is important to ensure the role of metabolic products are considered in the hypothesis setting (Tier 1) and testing (Tier 2), to ensure that any metabolites that could be responsible for toxic effects are not missed. It is critical to consider the rate and extent of metabolism of the parent chemical to use in PBK modelling in Tier 2.

5.5 Physiologically based kinetic (PBK) modelling

Description

PBK, also called physiologically based pharmacokinetic modelling (PBPK) or physiologically based biokinetic modelling (PBBK), is the use of mathematical models to predict the ADME profile of a specific systemic chemical exposure. Different organs are described in different compartments of the model, and the physiological parameters associated with the compartments (*e.g.* the rate of blood flowing in and out of the organs) are described by a set of equations (Campbell *et al*., 2012). PBK models are one of the most critical tools in NGRA as their output is ultimately used in the quantitative *in vitro* to *in vivo* extrapolation (QIVIVE), comparing *in vitro* effect concentrations with human internal doses as well as at other points in the workflow.

Strengths

PBK models can be developed in a tiered manner. For example, a preliminary model can be developed using very little input data (*in silico* predictions, clearance of the test item in hepatocytes or microsomes). In an exposure-led risk assessment, PBK models allow different exposure conditions (*e.g.* inclusion levels, product types, consumer habits) and the impact on internal exposure to be modelled. Because PBK modelling is a fairly mature science, guidance is available (*e.g.* World Health Organization, 2010) to ensure the quality and transparency of the used methods, model documentation, and the output.

Limitations

In common with most approaches described in this document, specialist knowledge and equipment is needed. In the case of PBK modelling, the specialist equipment means access to suitable PBK software and sufficient computing power to resolve the equations. Each software platform carries its own

strengths and limitations and these need to be thoroughly understood by the operator to ensure sources of uncertainty can be characterized and documented. Many PBK models lack a skin compartment, which presents a significant limitation for dermally applied cosmetics and therefore requires a modular approach to be adopted. In addition, physiological processes such as renal clearance are often more complex than the descriptions used to represent them in the model. Therefore, although models can be built with very little input data, to provide confidence that the model is robust a certain amount of experimental data is likely to be needed to support these approaches. This could include human exposure (kinetic) studies which are needed to verify the model output, which are costly and time consuming.

Integration in the risk assessment

PBK modelling may be used at various points in the risk assessment workflow, depending on the structure of the risk assessment and the hypothesis being tested. For example, in a read-across risk assessment, PBK models may be used to confirm the validity of a read-across, *i.e.* by demonstrating similar kinetics between the test item and comparators (Tier 0 Step 4). Where the internal TTC is used clearly, a prediction of internal exposure is needed at Tier 1 Step 5, and it is likely that this information will be provided by a PBK model. Finally, where neither read-across nor TTC is an appropriate method of completing the risk assessment, PBK modelling can give indications of the target organs concerned, while establishing the MoA hypothesis in Tier 1. Ideally preliminary PBK models should be used early in the risk assessment to guide dose selection in some *in vitro* tests (where this is scientifically appropriate). Finally, PBK predictions are critical in the final QIVIVE (Tier 2 Step 8).

5.6 *In chemico* tools

Description

Any toxic effect is the result of the interaction of a chemical substance or metabolite with a biochemical process, and the concept of MIE is well established in safety science (Allen *et al.,* 2014). *In chemico* analysis (experimental work not involving the use of cells or tissues) can aid in the understanding of the MIEs that may be associated with a specific chemical exposure and, therefore, provide insights into the possibility that adverse effects will occur. Examples of *in chemico* analyses that are relevant to human health risk assessment are the Direct Peptide Reactivity Assay (DPRA) for skin sensitization (OECD Test No. 442C), receptor binding assays (*i.e.* analysis of whether a chemical will bind to isolated receptors) and characterization of the phototoxicity hazard (production of Reactive Oxygen Species (ROS) following UV absorption and molecular excitation) of chemicals. In addition, other approaches described in this document (*e.g.* metabolite identification using isolated cellular components rather than whole cells) could be described as *in chemico.*

The DPRA is proposed to address the MIE of the skin sensitisation AOP, namely protein reactivity, by quantifying the reactivity of test chemicals towards model synthetic peptides containing either lysine or cysteine (Gerberick, *et al.,* 2004; OECD Test No. 442C). Cysteine and lysine percent peptide depletion values are then used to categorise a substance in one of four classes of reactivity for supporting the discrimination between skin sensitisers and non-sensitisers. The DPRA test method proved to be transferable to laboratories experienced in high-performance liquid chromatography (HPLC) analysis.

The correlation of protein reactivity with skin sensitisation potential is well established. Nevertheless, because protein binding represents only one key event (albeit the molecular initiating event of the

skin sensitisation AOP), protein reactivity information generated with testing and non-testing methods may not be sufficient on its own to conclude on the absence of skin sensitisation potential of chemicals.

Furthermore, when evaluating non-animal methods, it should be kept in mind that existing animal tests may not fully reflect the situation in the species of interest, *i.e.* humans. It is therefore important to understand the mechanistic basis of the health effect being evaluated. In the skin sensitisation AOP, the MIE is represented by the DPRA, which has been shown to be applicable to test chemicals covering a variety of organic functional groups, reaction mechanisms, skin sensitisation potency (as determined in both human and animal studies) and physico-chemical properties. Taken together, this information indicates the usefulness of the DPRA to contribute to the identification of skin sensitisation hazard. Therefore, DPRA data should be considered in the context of integrated approaches such as IATA, combining them with other complementary information *e.g.*, derived from *in vitro* assays addressing other key events of the skin sensitisation AOP as well as non-testing methods including read-across from chemical analogues (OECD, 2014; Hoffmann *et al.,* 2018).

Other *in chemico* analyses may be similarly relevant to help characterize the MIEs or key events (KEs) related to other AOPs.

Strengths

In chemico approaches can provide relevant data to help inform the likely behaviour of a test item or its metabolites. Results of *in chemico* assays tend to be more reproducible and demonstrate less variability than the results of biological (*e.g.* cell-based) assays.

Limitations

Depending on the method used, *in chemico* methods do not generally provide information which can be extrapolated directly to the *in vivo* human situation for risk assessment. It is therefore likely that *in chemico* methods will form part of the tiered approach and require *in vitro* follow up using appropriate models. For example, receptor binding assays, DPRA, or UV absorbance assays all require follow up in other experimental systems to determine the likelihood that an adverse health effect will manifest at relevant exposures.

Integration in the risk assessment

In chemico analysis is likely to be most useful at the hypothesis setting stage of the risk assessment (Tier 1), informed by *in silico* predictions and, where appropriate, followed-up with *in vitro* assays.

5.7 'Omics

Description

'Omics represent the collective array of molecular technologies that have been developed to characterize and explore the interactions and relationships of various types of molecules that make up the cells of a given organism (such as genes, proteins, metabolites, *etc*.). Given that molecular profiles may vary after exposures to chemical stressors, these technologies represent an important tool in helping to elucidate functional and/or structural alternations within a cell that may signal a molecular response that denotes specific perturbations in cellular function (Aardema and MacGregor,

2002). Techniques such as transcriptomics (a description of all the genes expressed within a cell or tissue), proteomics (a description of all the proteins expressed by a cell or tissue) or metabolomics (a description of the all metabolites produced by a cell or tissue) are the most common 'omics technologies that are likely to have an integral role in supporting NGRA.

Strengths

The strengths of transcriptomics, proteomics and metabolomics approaches is that they give a broad coverage of the genes, proteins or metabolites that are differentially expressed following exposure to a test chemical at different concentrations. Approaches are high density (collecting a lot of data) and can be high throughput. Bioinformatics tools can be used to interpret these changes and to propose which pathways (and therefore disease states) may potentially be affected by chemical exposures.

Limitations

Each technique carries its own limitations, but all require a significant level of expertise to interpret the data obtained. For example, when interpreting transcriptomics data it is important to recognize that changes in gene expression do not necessarily result in changes in protein expression, and post-translation modification of proteins will clearly not be detected using transcriptomics. Because changes in protein expression do not always correlate with changes in gene expression, lack of a coherent mechanistic link between protein and mRNA expression can hamper efforts to develop MoA hypotheses. Proteomic analyses tend to be less sensitive than transcriptomic analyses, because it is possible to measure very low levels of mRNA. Furthermore, although high throughput metabolomics approaches are available, they do require highly specialized equipment and expertise. In general, it needs to be considered whether changes are adaptive or adverse. Overall, because transcriptomics analyses are more standardized and cost-effective than are proteomics and metabolomics analyses, transcriptomics approaches tend to be more accepted and used.

Integration in the risk assessment

'Omics approaches may be integrated in a number of areas in the risk assessment of cosmetic ingredients. The concept of using gene, protein or metabolite expression to inform biological grouping and to cluster chemicals that appear to have the same MoA (Tier 0 Step 4 of the SEURAT-1 workflow) has already been discussed. However, these types of approaches may also be useful later in the workflow, for example in Tier 1 Step 6 where it is necessary to develop MoA hypotheses. As mentioned above, standardized approaches, ease of detecting small changes and cost effectiveness mean that transcriptomics in particular is likely to be critical in providing an early indication of the pathways that may be affected by chemical exposure and guide hypothesis generation and testing at Tiers 1 and 2 of the framework. Additionally, the concept of using transcriptomic data to identify NOTEL (Lobenhofer *et al.,* 2004) could be particularly useful at this level for cosmetic ingredients. A NOTEL is the concentration of the test chemical that does not cause any differential gene expression compared with control, or does not cause any meaningful differential gene expression from a systems biology point of view. In other words, isolated changes in individual genes may not represent a meaningful difference, whereas perturbation of several genes in the functional unit (the biological pathway) would be a meaningful change. The reason this is of particular interest to cosmetic ingredients is because there is a greater likelihood that, for ingredients which are not intended to be biologically active and have low systemic bioavailability, the NOTEL and systemic exposure levels will not overlap. This information in itself is critically important to the risk assessment.

5.8 Reporter gene assays

Description

Reporter genes are widely used to monitor cellular events, such as gene expression, protein-protein interactions, and receptor activation. The reporter gene is connected to marker promoter/enhancer element, and up- or down-regulation of the reporter gene expression is used to assess the effect of the tested compound on the expression/activation of the target element. Signals sent by the reporter genes can be bioluminescent (luciferase, Luc), fluorescent (green fluorescence protein (GFP), etc.), and colorimetric (β-galactosidase, etc.). The bioluminescent reporter luciferase emits light by oxidizing its substrate, luciferin, in a specific manner. Reporter gene assays using a bioluminescent reporter gene are considered the most suitable for quantitative monitoring of cellular events because their sensitivity and range of linear response are superior to those of other reporter gene assays. Fluorescent reporter genes emit a fluorescent signal when excited by an external source. Because their relative fluorescence intensity is overwhelmingly superior to other reporter genes, imaging analysis at single cell or subcellular levels are frequently performed. Colorimetric reporter genes, including β-galactosidase and chloramphenicol acetyltransferase, are the most traditional reporter genes. Recently, poor sensitivity and narrow linear dynamic range due to simple colorimetric measurement has been largely improved by replacing with chemiluminescence measurement methods. Similarly, reporter gene assays have become one of the conventional methods for assessing chemical hazards, including skin sensitization or endocrine activity (*e.g.* OECD Test Nos. 442D, 455).

In addition to traditional reporter gene assay techniques, similar detection methods can be used alongside automated microscopy to enable the collection of high content image-based data. Cellular assays with multiple readouts permit the evaluation of several biological outcomes at the single cell level. In recent years, this technology has evolved to the extent that it is now well established and widely used in basic research, drug discovery and toxicology (Zock, 2009).

Strengths

The most important strength of the reporter gene assay is simplicity of measurement of the reporter activity. Commercialized kits are readily used without a great deal of specific training. Depending on the assay, reporter gene assays can also provide a very specific readout for the molecular target being investigated, and it is possible to increase human relevance by ensuring recombinant protein targets are used. Reporter gene assays are therefore critical tools in defining or confirming the MIE associated with a particular chemical exposure. In addition, with these assays the number of samples can be high, and automated microscopy can provide the opportunity for high content or high throughput analysis (Jiang *et al.,* 2016). Furthermore, because bioluminescent and fluorescent signals from cells can be detected without disrupting cells, continuous or real-time measurement of the signals can be performed, thereby allowing kinetic analysis of the target, including marker gene expression or receptor activation.

Limitations

Although reporter gene assays are useful to help identify mechanisms of action, in most of cases, when used alone, they are insufficient to predict a biological and physiological effect. As is the case with other *in vitro* assays, a chemical that precipitates in culture medium during cell treatment cannot appropriately be assessed with reporter gene assays. Furthermore, in the case of reporter gene assays using bioluminescent and colorimetric reporter genes, chemicals that strongly inhibit the enzymatic reaction of the reporter protein can be problematic. Some chemicals interfere with the light signals

produced by reporter proteins owing to absorbance, or can stabilize the reporter protein in cells by binding to the protein, which can lead to misinterpretation of assay results.

Integration in the risk assessment

Reporter gene assays, including high throughput screening and high content imaging analysis, have emerged as a high reliable method mechanism-based chemical risk assessment. Reporter gene assays are typically used in the first steps of AOP. For example, in the KeratinoSens^{TM} skin sensitization assay (OECD Test No. 442D) and the IL-8 Luc assay OECD Test No. 442E), activation of the oxidative stress pathway and expression of interleukin-8 gene, respectively, are monitored using bioluminescent reporter genes. Furthermore, to more precisely analyze changes in cellular events associated with toxicity, development of multiplex assay platforms (in which multiple cellular events including MIEs and key events can be simultaneously monitored using multiple reporter genes) may be highly advantageous.

5.9 *In vitro* pharmacological profiling

Description

In vitro pharmacological profiling involves screening compounds against a broad range of targets (receptors, ion channels, enzymes and transporters) using a variety of assay technologies (Bowes *et al.,* 2012). This is used in the early drug discovery process to screen pharmaceutical candidates for safety liabilities. The diverse range of targets is selected to be predictive of adverse drug reactions in humans which have a well characterized MoA. Although these targets and assays were originally selected for detecting off-target interactions of drugs, this concept may have utility in identifying potential undesirable modes of action for other types of chemical and, therefore, help develop mode of action hypotheses for a given chemical exposure.

Strengths

A well-established, human-relevant panel of targets and assays has been developed and commercialized for use within the pharmaceutical industry. Using *in vitro* pharmacological screening can help to predict clinical side effects which may not be detected in *in vivo* animal testing or clinical trials. These studies are low cost and rapid, and provide mechanistic information. Used in combination with the other techniques described in this document, such a panel could provide a screen for activities that could result in non-pharmaceutical chemicals causing adverse effects in humans.

Limitations

Because these techniques were developed for early screening of pharmaceuticals, they are of uncertain relevance to less specific chemistries (*i.e.* those with a much lower biological activity). Application of these tools to the risk assessment of cosmetic ingredients is therefore needed to provide information on how often they contribute to the risk assessment.

Integration in the risk assessment

Pharmacological profiling could be useful at Step 6 (Tier 1) of the SEURAT-1 framework (hypothesis setting) in combination with the other techniques such as transcriptomics.

5.10 3D culture systems

Description

3D models have been developed to overcome some of the limitations of 2D cell cultures, and their use is well-established for local effects. Therefore, most 3D models already being applied to safety assessment are surrogates for external organs, including human dermis, epidermis, full-thickness and pigmented epidermis models. A number of these are now commercially-available worldwide and are useful both for dermal research and for the safety assessment of cosmetics. Moreover, the use of some of these models has been validated within OECD Test Guidelines for assessment of skin corrosion (OECD Test No. 431), skin irritation (OECD Test No. 439), dermal absorption (OECD Test No. 428) and recently, for eye irritation (OECD Test No. 492).

3D models of internal organs are also being investigated, but they are not yet widely used in the safety evaluation of cosmetic ingredients. There is a significant amount of research on 3D models of hepatocytes either grown on or within a matrix or in suspension (Edmonson *et al*., 2014). 3D cultures of primary hepatocytes or cell lines derived from the liver are useful for studying long-term culture effects, the maintenance of functional structure, the functional expression of the human liver, and to test the effects of repeated exposure to chemicals. Similar liver models from a variety of animal species are being considered for use in pharmaceutical screening. Other models have been developed (Heinonen, 2015; Jamin 2015), several of which are also commercially available and are used worldwide in research and for toxicological safety assessments.

Strengths

3D cultures provide a test system that is potentially more representative of tissue biology than are 2D cell cultures. Because normal tissue functioning is dependent on interactions between multiple cell types, 3D co-cultures provide possibilities for the development of biomarkers that could be useful in risk assessment. Some 3D culture systems are already mentioned in OECD TGs addressing skin corrosion, skin irritation, and eye irritation. These TGs describe *in vitro* procedures for identifying chemicals (substances and mixtures) not requiring classification and labelling for local toxicological damage, in accordance with the United Nation Globally Harmonised System of Classification and Labelling of Chemicals (UN GHS). The test methods using skin and eye 3D models are applicable to substances and mixtures, and to solids, liquids, semi-solids and waxes. The liquids may be aqueous or non-aqueous; solids may be soluble or insoluble in water. Whenever possible, solids should be ground to a fine powder before application; no other pre-treatment of the sample is required. This wide range of applicability regarding the physico-chemical properties of the tested compound is a great advantage of the 3D models compared to the *in vitro* cells culture extensively used in research and basic toxicity assessment, like cytotoxicity. Skin 3D models are also used for follow-up genotoxicity testing (Reisinger *et al*., 2018).

Limitations

Although the use of 3D models of skin and eye is well established, standardized models for internal organs are not yet in common use in the risk assessment of cosmetics. Many current models need significant further developments, and most are constructed with only one cell type. Therefore, their construction and functions are not comparable to *ex vivo* models. Further advances in these areas are expected, particularly because 3D epithelium models have advanced over the past decade. The development of 3D models of a wide variety of cell types can be expected, and a model constructed

with differentiated cells (including different types of stem cells) will be produced in the near future (for example, a full-thickness skin model that includes melanocytes, Langerhans cells and hair follicle cells derived from stem cells as well as models of key internal organs). The toxicological biomarker used with all of the current 3D models in OECD TGs for local toxicity, is cytotoxicity. Cytotoxicity is not specific and does not give any information on the mode of action of the toxicological process. For more mechanistic or organ-specific applications, specific toxicological biomarkers need to be developed for each 3D model. The long lead-time required to prepare 3D models is a limiting factor that drives up costs. In addition, the economic viability of developing a wide variety of small-scale 3D models remains challenging and a production platform that enhances efficiency is needed. Further study of 3D models using cells differentiated from ES (Embryonic Stem) or iPS (induced Pluripotent Stem) cells is impossible without the development of quality control criteria for the system used to differentiate the cells. It is difficult to coordinate the long-term maintenance of 3D models with combinations of cells, and it will be necessary to co-culture with organ-derived substances and reconstructed blood vessels, in order to promote the development of humans-on-chips or organs-on-chips.

Integration in the risk assessment

A range of TGs describing test methods that use epidermal and/or ocular models are already available worldwide for regulatory use. The quality of the procedures that use these models is maintained by the suppliers. The 3D-models used to evaluate primary irritancy and tolerance of ingredients, as well as formulated products, on reconstructed skin and ocular epithelium are widely used and recognized for their predictivity, particularly in the cosmetic industry. Moreover, these methods are usually considered as a first step of tolerance validation of cosmetic formula before testing on humans. In addition, 3D models can be used for absorption assessment and, thus, be a key tool for the PBK assessment. OECD Test No. 428 includes the use of an *ex vivo* skin model for assessing the effects of exposure to chemicals. In the future, *in vitro* full-thickness skin, intestine and alveolar models are expected to be used for assessing the effects of exposure to chemicals. Due to the limitations mentioned above, routine use of 3D model systems outside of these areas of local toxicity and bioavailability assessment is extremely challenging. They are only likely to be used to assess systemic toxicity at the highest tier (Tier 2) of the risk assessment where other methods cannot provide the information required.

5.11 Organ-on-chip

Description

Organs-on-chips (OCs) are microengineered biomimetic systems that represent key functional units of living human organs. They often consist of transparent 3D polymeric microchannels lined by living human cells and replicate three important aspects of intact organs: the 3D microarchitecture defined by the spatial distribution of multiple tissue types; functional tissue-tissue interfaces; and complex organ-specific mechanical and biochemical microenvironments (Esch *et al.,* 2015). With the integration of many other technologies, OCs have the potential to more accurately simulate the *in vivo* microenvironment and replace animal models (Aziz *et al.,* 2017).

Strengths

The capacity to reproduce the function and structure of the organ, and the ability to evaluate simultaneously efficacy, toxicity, and pharmacokinetics with multi-organs-on-a-chip are some

advantages of OCs over traditional *in vitro* cytology experiment tools (such as 96-well plate and transwell chamber) and general cell culture systems based on microchip (An *et al.,* 2015). OCs therefore provide the potential for more human-relevant safety evaluations and the ability to reproduce *in vivo* exposure conditions. They are likely to become important tools for finding functional properties, pathological states, and developmental studies of organs (Aziz *et al.,* 2017), and thus allow the development of biomarkers predictive of adverse changes. Optical transparency of OC microdevices enables direct real-time visualization and quantitative high-resolution analysis of diverse biological processes in ways that have not been possible in animal models (Esch *et al.,* 2015).

Limitations

In general, the ultimate goal of OC systems is to build human-on-a-chip systems using human cells and tissues that are capable of predicting effects in humans. To achieve this goal, the feasibility, reliability, controllability, and observability of organ-on-a-chip systems must be improved, to make them comprehensive platforms for toxicological and metabolic tests (Wang *et al.,* 2015). The physico-chemical properties of synthetic materials, heavily used on cell culture substrates of existing organ-on-chip models, are not appropriate for mimicking extracellular matrices *in vivo* and new cell culture substrates are needed. Fabrication requires specialized micro-engineering capabilities. Another limitation is that OCs testing will be more costly than biochemical and cellular testing in high-throughput contexts, such as for screening a large compound library for target activity (Esch *et al.,* 2015; Bhatia and Ingber, 2014).

Integration in the risk assessment

Due to their complexity, these models would only be used for the most refined risk assessments at Tier 2 of the framework.

5.12 Zebrafish larva assays

Description

In most jurisdictions non-mammalian vertebrate larvae are not regulated as laboratory animals (*e.g.* Council Directive 2010/63/EU on the protection of animals used for scientific purposes), which has led to interest in their use as an 'alternative' test system. Although not commonly used in cosmetic risk assessment, zebrafish larvae are used in screening and drug discovery applications such as central nervous system development, cardiovascular diseases, hematopoietic development and disease, and they have emerged as a powerful biological system for drug development against hearing loss (Philip *et al*., 2018). Recent studies have begun to establish the capabilities and limitations of zebrafish for disease modelling, drug screening, target identification, pharmacology, and toxicology. As understanding increases and as the technologies for manipulating zebrafish improve, it is hoped that the zebrafish larvae will have a key role in accelerating the emergence of precision medicine. The most important strategies in pharmacogenomics are gene expression profiling and the network analysis of human disease models.

With the high cost and the long-term assessment of developmental or systemic toxicity testing in mammals, the vertebrate zebrafish larva has become a useful model organism for high-throughput toxicity testing. Some of the advantages of using zebrafish in developmental and repeat dose toxicity testing are that they allow to focus on central nervous system development and neurobehavior.

Zebrafish larva assays have not widely been used in the human safety assessment of cosmetic ingredients but do show potential as a non-mammalian test system.

Strengths

These models allow easy optical imaging, large sample size, and organ-specific studies and, hence, an increasing number of preclinical studies are employing zebrafish models. Zebrafish have a short reproductive cycle and produce a large number of offspring, meaning that they can be bred very efficiently and in a cost-effective manner, and have a long history of use in pharmaceutical screening and in (eco)toxicological safety evaluation. The original transgenic dyschromia zebrafish are transparent allowing internal organs and neurons to be visualized. Their genome can be modified using techniques such as CRISPR (Hwang *et al.,* 2013) to assist in the study of human pathogenesis (Liu *et al.,* 2017). Their rapid embryonic development allows the study of diseases at early stages of embryonic development.

Limitations

Although zebrafish embryos up to certain developmental stages are not regulated as laboratory animals in many regulations, beyond the larval stage they are regulated as animals. Zebrafish cannot therefore be considered as a non-animal model, and do not fulfil the criteria of being 'human derived'. To date, their use in risk assessment of cosmetics is extremely limited. The methodologies used by research groups vary greatly, and there are no harmonized protocols for their use in human safety assessment. Finally, a more comprehensive understanding of the similarities and differences between human and zebrafish biology would be needed to use them in a mechanism-based risk assessment.

Integration in the risk assessment

As an immature animal model, zebrafish are unlikely to be considered as a test system of first-choice for cosmetic risk assessment. However, where human-derived cell lines or primary cells do not express the processes required to test the mode of action hypothesis, they may be considered as a necessary tool to fill the gap between cell cultures and whole organisms for the targeted testing required at Tier 2 of the SEURAT-1 framework.

5.13 Pathways modelling

Description

Identifying which pathway(s) may be perturbed by a chemical exposure does not provide any information about how different chemical exposures will affect the dynamics of the pathway(s). Computational systems biology modelling can be used to not only describe the ADME characteristics of a substance exposure (PBK modelling, Section 5.5), it can also be used to describe the dynamics of a cellular signaling pathway. In computational systems biology pathways models, the components of the pathway and the network motifs (*e.g.* negative feedback loops, feed forward loops, etc.) which maintain pathway balance for normal cellular function are described. In this way, a pathway model may be built which describes how the overall pathway would be affected under different exposure conditions. For example, a mathematical model describing the basic interactions between follicle-stimulating hormone, luteinizing hormone, testosterone, 5-α-reductase and dihydrotestosterone has been published which simulated the levels of these hormones in testes and blood of adult rats and allowed accurate recapitulation of experimental data showing the extent to which the mass of the

prostate decreased following the daily administration of finasteride to rats for 21 days (Zager and Barton, 2012). These types of models, parameterized for human biology, can therefore provide valuable insights into the outcome of different interventions and be used to underpin safety decisions (Adeleye *et al.,* 2015).

Strengths

Pathways modelling can provide the mechanistic insight that may be needed to increase confidence in risk assessment decisions, and to help to determine the exposure conditions necessary to have an adverse effect on the pathway (*i.e.* to move it from a position of adaptation to adversity). Computational models can be informed by the experimental data and, in turn, inform the design of subsequent experiments, leading to a more refined outcome.

Limitations

Computational systems biology modelling requires a specialized skillset, and is a multidisciplinary effort, requiring close interactions between modellers, experimentalists, biologists and chemists. Each biological pathway requires a bespoke model to be built, and the amount of resources needed to develop a new model is significant. This type of modelling is therefore only likely to be employed where biological activity and exposure levels are close or overlapping and the benefit of performing the modelling justifies the cost.

Integration in the risk assessment

Pathway models represent a high level of refinement in the risk assessment and, where needed, would be used at Tier 2 of the SEURAT-1 workflow.

5.14 Human studies

Description

As discussed in the section on PBK modelling (Section 5.5), human kinetic studies play a valuable role in verifying exposure predictions and refining the model output. Human studies can also be useful in confirming that the risk assessment is indeed protective of consumers (*i.e.* to confirm a lack of effect). This clearly requires the completion of a risk assessment that supports the specific conditions of use in that study and ethical approval. For many ingredients, it may be possible to confirm a lack of adverse effects without creating a model product provided the ingredient is compatible with an innocuous diluent. For ingredients that are more difficult to assess (*e.g.* that show marked vehicle-dependent differences in skin penetration), it may be advantageous to create a model formulation at the maximum use concentration to confirm a lack of adverse effect. Historically, this type of human study has been restricted to assessing tolerability of formulations (*e.g.* human primary dermal irritancy) or dermal sensitization. These types of volunteer studies often involve exaggerated exposure (where the risk assessment allows) to demonstrate a lack of effect or compatibility and tolerability. In a NGRA context, the range of effects that needs to be investigated is likely to increase and may need to involve investigation of systemic biomarkers as well as local skin effects.

Strengths

The use of human studies in cosmetic risk assessment is well established, and generation of human data significantly decreases the uncertainty in the risk assessment. Furthermore, the results of human studies can be invaluable in contributing to the refinement of predictive models (Maxwell *et al.*, 2014).

Limitations

There are a number of limitations and barriers associated with performing human studies for cosmetic ingredients. These include ethical, logistical and scientific considerations. It is clearly important to ensure all studies are conducted in line with the appropriate ethical standards, for example, adherence to the principles of the Declaration of Helsinki (World Medical Association 2013), and the protection and welfare of volunteers is paramount. A key limitation is that perturbation of some biological pathways may not be amenable to examination in a human volunteer study. For example, the possibility that panellists in a Human Repeat Insult Patch Test (HRIPT) may become sensitized dictates that these studies should only be performed rarely and where the benefit overwhelmingly outweighs the risk (Basketter, 2009). Another key limitation is that perturbation of some biological pathways may not be amenable to examination in a human volunteer study. For example, if the risk assessment hypothesis relates to disrupted embryo-fetal development, a confirmatory human study would not be possible, and other methods of reducing uncertainty in the risk assessment need to be explored. Logistical challenges include the size and cost of human studies and the fact that, if systemic biomarkers are involved, techniques that are currently not routine in volunteer studies of cosmetic products will need to be used. Some of the scientific limitations of these studies are related to their cost, because the number of volunteers needed for robust analysis with relevance to the population may result in a study that is prohibitively expensive. For example, although the HRIPT is considered the most reliable test method by which confirmatory human data can be made available for skin sensitization (McNamee *et al*., 2008), small group sizes may hamper data interpretation. Typically, these studies are conducted in groups of 50 to 100 subjects, and have been conducted with as few as 25. Unless a large proportion of the population is sensitized, the probability of detecting a response in an HRIPT study on 25-50 volunteers is low (McNamee *et al.,* 2008). Thus, the number of participants is critical in designing a successful study.

Integration in the risk assessment

Human studies are most likely to fit into the workflow in Tier 2, *i.e.* once a mode of action hypothesis has been developed and tested using targeted *in vitro* studies and points of departure for the pathway perturbation have been set. This will enable verification that the points of departure are indeed protective of human health.

6. Discussion and Conclusion

Many non-animal approaches are available to conduct a NGRA, and different proof-of-concept examples showing how a NGRA may be done in practice are emerging. These range from high throughput evaluation using multiple toxicodynamic assays for the purpose of priority setting (Wetmore *et al.,* 2015; Wambaugh *et al.,* 2018) to *in vitro* to *in vivo* extrapolation for specific activities (Dancik *et al.,* 2015), and even the use of dietary comparators to put *in vitro* mode of action data into context (Becker *et al.,* 2015). Although all different, these approaches have several points in common in that their goal is a risk assessment or prioritization approach which is human relevant, exposure-led, hypothesis driven, and designed to prevent harm. It is clear that some of the higher-tier approaches that may be necessary to complete risk assessments where exposure and activity concentrations overlap do not yet exist or require further evaluation. For instance, 3D culture systems, organ-on-chip, and zebrafish models are not yet in routine use in the evaluation of systemic

effects of cosmetics. Other techniques, such as pathways modelling have been applied to proof of concept risk assessments but are highly tailored to individual assessments and as such require significant investment on a case-by-case basis (Adeleye *et al.,* 2015). A summary of the techniques/NAMs described in this report and the JWG's opinion of their readiness for use in the risk assessment of cosmetic ingredients is shown in Table 1.

Table 1: Summary of the new approach methodologies described herein and their proposed state of readiness for use in the risk assessment of cosmetic ingredients

Already in use in cosmetic risk assessment	Mature technology with likely utility in cosmetic risk assessment	May have utility but insufficiently developed for current use
Read across Exposure-based-waiving *In silico* tools Metabolism and metabolite identification PBK modelling *In chemico* assays Reporter gene assays 3D culture systems (local effects and genetic toxicity) Human studies	'Omics (especially transcriptomics) *In vitro* pharmacological profiling Pathways modelling 3D culture systems (for systemic effects)	Organ-on-chip Zebrafish larva assays

The SEURAT-1 *ab initio* workflow for systemic (repeat-dose) toxicity provides a flexible framework upon which to build an NGRA (Berggren *et al.,* 2017). Importantly, this framework can be used in a way that exemplifies the 9 principles outlined in Part 1 of this report (ICCR, 2017). It is therefore the opinion of the JWG that the novel approaches described in this document should be applied to risk assessments, using the workflow and the 9 principles as guidelines, to evaluate their utility. Using this exposure-led framework, it is possible that for cosmetic ingredients some of the higher-tier approaches may only rarely be needed.

7. References

Aardema, M.J. and J.T. MacGregor (2002) Toxicology and genetic toxicology in the new era of "toxicogenomics": Impact of "-omics" technologies. *Mutation Research.* 499(1): 13-25. https://doi.org/10.1016/S0027-5107(01)00292-5

Adeleye, Y., Andersen, M., Clewell, R., Davies, M., Dent, M., Edwards, S., Fowler, P., Malcomber, S., Nicol, B., Scott, A., Scott, S., Sun, B., Westmoreland, C., White, A., Zhang, Q., and Carmichael, P. L. (2015) Implementing Toxicity Testing in the 21st Century (TT21C): Making safety decisions using toxicity pathways, and progress in a prototype risk assessment. *Toxicology*. 5: 102-111. https://doi.org/10.1016/j.tox.2014.02.007

Allen, T., Goodman, J., Gutsell, S., and Russell, P. (2014) Defining Molecular Initiating Events in the Adverse Outcome Pathway Framework for Risk Assessment. *Chemical Research in Toxicology* 27(12): 2100-2112. https://dx.doi.org/10.1021/tx500345j

An, F., Qu, Y., Liu, X., Zhong, R., and Luo, Y. (2015) Organ-on-a-Chip: New Platform for Biological Analysis. *Analytical Chemistry Insights* 10: 39-45. https://dx.doi.org/10.4137/ACI.S28905

Aziz, A.U.R., Geng, C., Fu, M., Yu, X., Qin, K., and Liu, B. (2017) The Role of Microfluidics for Organ on Chip Simulations. *Bioengineering* 4(2): 39. https://dx.doi.org/10.3390/bioengineering4020039

Ball, N., Bartels, M., Budinsky, R., Klapacz, J., Hays, S., Kirman, C. and Patlewicz, G. (2014) The challenge of using read-across within the EU REACH regulatory framework; how much uncertainty is too much? Dipropylene glycol methyl ether acetate, an exemplary case study. *Regulatory Toxicology and Pharmacology*. 68: 212–221. https://doi.org/10.1016/j.yrtph.2013.12.007

Ball, N., Cronin, M. T. D., Shen, J., Blackburn, K., Ewan D. Booth, E. D., Bouhifd, M., Donley, E., Egnash, L., Hastings, C., Juberg, D. R., Kleensang, A., Kleinstreuer, N., Kroese, E. D., Lee, A. C., Luechtefeld, T., Maertens, A., Sue Marty, S., Naciff, J. M., Palmer, J., Pamies, D., Penman, M., Richarz, A. N., Russo, D. P., Stuard, S. B., Patlewicz, G., Ravenzwaay, B. V., Wu, S., Zhu, H., and Hartung, T. (2016) Toward good read-across practice (GRAP) guidance. *Alternatives to Animal Experimentation (ALTEX)* 33: 149-166. http://dx.doi.org/10.13/altex.1601251

Barlow, S. (2005) Threshold of Toxicological Concern (TTC): A tool for assessing substances of unknown toxicity present at low levels in the diet. ILSI EUROPE Concise Monograph Series. http://ilsi.eu/wp-content/uploads/sites/3/2016/06/C2005Thres_Tox.pdf

Becker, R.A., Friedman, K.P., Simon, T.W., Marty, M.S., Patlewicz, G. and Rowlands, J.C. (2015) An exposure:activity profiling method for interpreting high-throughput screening data for estrogenic activity—Proof of concept, *Regulatory Toxicology and Pharmacology*. 71(3): 398–408. https://dx.doi.org/10.1016/J.YRTPH.2015.01.008

Berggren, E., White, A., Ouedraogo, G., Paini, A., Richarz, A., Bois, F., Exner, T., Leite, S., Grunsven, A.v., Worth, A., and Mahony, C. (2017) *Ab initio* chemical safety assessment: a workflow based on exposure considerations and non-animal methods. *Computational Toxicology* 4: 31-44. https://doi.org/10.1016/j.comtox.2017.10.001

Bhatia, S.N. and Ingber, D.E. (2014) Microfluidic organs-on-chips. *Nature Biotechnology.* 32(8): 760-772. https://dx.doi.org/10.1038/nbt.2989

Bhatia, S., Schultz, T., Roberts, D., Shen, J., Kromidas, L., and Api, A.M. (2015) Comparison of Cramer classification between Toxtree, the OECD QSAR Toolbox and expert judgment. *Regulatory Toxicology and Pharmacology* 71(1): 52-62. https://doi.org/10.1016/j.yrtph.2014.11.005

Blackburn, K., Bjerke, D., Daston, G., Felter, S., Mahony, C., Naciff, J., Robison, S., and Wu., S. (2011) Case studies to test: A framework for using structural, reactivity, metabolic and physico-chemical similarity to evaluate the suitability of analogs for SAR-based toxicological assessments, *Regulatory Toxicology and Pharmacology.* 60: 120-35. https://doi.org/10.1016/j.yrtph.2011.03.002

Blackburn, K., and Stuard, S.B. (2014) A framework to facilitate consistent characterization of read across uncertainty, *Regulatory Toxicology and Pharmacology*. 68: 353-62. https://doi.org/10.1016/j.yrtph.2014.01.004

Bowes, J., Brown, A., Hamon, J., Jarolimek, W., Sridhar, A., Waldron, G., and Whitebread, S. (2012) Reducing safety-related drug attrition: the use of *in vitro* pharmacological profiling. *Nature Reviews* 11: 909-922. https://dx.doi.org/10.1038/nrd3845

Campbell, J.L., Clewell, R.A., Gentry, P.R., Andersen, M.E. and Clewell, H.J. (2012) Physiologically based pharmacokinetic/toxicokinetic modeling. In: Reisfeld B., Mayeno A. (eds) Computational Toxicology. Methods in Molecular Biology (Methods and Protocols), vol 929, pp. 439-499, Humana Press, Totowa, NJ https://dx.doi.org/10.1007/978-1-62703-050-2_18

Carthew, P., Clapp, C., and Gutsell, S. (2009) Exposure based waiving: The application of the toxicological threshold of concern (TTC) to inhalation exposure for aerosol ingredients in consumer products. *Food and Chemical Toxicology* 47(6): 1287-1295. https://dx.doi.org/10.1016/j.fct.2009.02.024

Cheeseman, M.A., Machuga, E.J., and Bailey, A.B. (1999) A tiered approach to threshold of regulation. *Food and Chemical Toxicology* 37(4): 387-412. https://dx.doi.org/10.1016/S0278-6915(99)00024-1

Council Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. OJ L276 http://data.europa.eu/eli/dir/2010/63/oj

Cramer, G.M., Ford, R.A., and Hall, R.L. (1978) Estimation of toxic hazard – A decision tree approach. *Food and Cosmetics Toxicology* 16(3): 255-276. https://dx.doi.org/10.1016/S0015-6264(76)80522-6

Cramer, R.D. III, Patterson, D.E., and Bunce, J.D. (1988) Comparative molecular field analysis (CoMFA). 1. Effect on shape and binding of steroids to carrier proteins. *Journal of the American Chemical Society* 110: 5959-5967. https://pubs.acs.org/doi/pdf/10.1021/ja00226a005

Dancik, Y., Troutman, J. A. and Jaworska, J. (2015) Estimation of *in vivo* dose of dermally applied chemicals leading to estrogen/androgen receptor-mediated toxicity from *in vitro* data—Illustration with four reproductive toxicants, *Reproductive Toxicology* 55: 50–63. https://dx.doi.org/10.1016/J.REPROTOX.2015.01.002

Dewhurst, I. and Renwick, A.G. (2013) Evaluation of the Threshold of Toxicological Concern (TTC) – Challenges and approaches. Regulatory Toxicology and Pharmacology 65(1): 168-177. https://doi.org/10.1016/j.yrtph.2012.03.007

ECETOC (2012) Category approaches, Read-across, (Q)SAR. Technical Report No 116. European Centre for Ecotoxicology and Toxicology of Chemicals. http://www.ecetoc.org/publications

ECHA (2008) Guidance on information requirements and chemical safety assessment, Chapter R.6: QSARs and grouping of chemicals. http://echa.europa.eu/documents/10162/13632/information_requirements_r6_en.pdf

ECHA (2017) Read-Across Assessment Framework (RAAF). https://echa.europa.eu/documents/10162/13628/raaf_en.pdf%0A

Edmondson, R., Broglie, J.J., Adcock, A.F., and Yang, L. (2014) Three-dimensional cell culture systems and their applications in drug discovery and cell-based biosensors. *ASSAY and Drug Development Technologies* 12(4): 207-218. https://dx.doi.org/10.1089/adt.2014.573

[EFSA] European Food Safety Authority (2012) Scientific opinion on exploring options for providing advice about possible human health risks based on the concept of Threshold of Toxicological Concern (TTC). *ESFA Journal* 10(7): 2750. https://dx.doi.org/10.2903/j.efsa.2012.2750

Esch, E.W., Bahinski, A., and Huh, D. (2015) Organs-on-chips at the frontiers of drug discovery. *Nature Reviews Drug Discovery* 14(4): 248-260. https://dx.doi.org/10.1038/nrd4539

Frawley, J.P. (1967) Scientific evidence and common sense as a basis for food-packaging regulations. *Food and Cosmetics Toxicology* 5(3): 293-308. https://doi.org/10.1016/S0015-6264(67)83055-4

Gerberick, G.F., Vassallo, J.D., Bailey, R.E., Chaney, J.G., Morrall, S.W., and Lepoittevin, J.P. (2004) Development of a peptide reactivity assay for screening contact allergens. *Toxicological Sciences* 81(2): 332-343. https://dx.doi.org/10.1093/toxsci/kfh213

Godoy, P., Hewitt, N.J., Albrecht, U., *et al.* (2013) Recent advances in 2D and 3D in vitro systems using primary hepatocytes, alternative hepatocyte sources and non-parenchymal liver cells and their use in investigating mechanisms of hepatotoxicity, cell signaling and ADME. *Archives of Toxicology* 87(8): 1315-1530. https://rd.springer.com/article/10.1007%2Fs00204-013-1078-5

Gómez-Lechón, M-J., Castell, J.V., Donato, M.T (2007) Hepatocytes – the choice to investigate drug metabolism and toxicity in man: *in vitro* variability as a reflection of *in vivo*. *Chemico-Biological Interactions* 168: 30-50. https://doi.org/10.1016/j.cbi.2006.10.013

Groothius, F.A., Heringa, M.B., Nicol, B., Hermens, J.L., Blaauboer, B.J., Kramer, N.I. (2015) Dose metric considerations in in vitro assays to improve quantitative in vitro-in vivo dose extrapolations. *Toxicology* 332: 30-40. https://doi.org/10.1016/j.tox.2013.08.012

Heinonen, T. (2015) Better science with human cell-based organ and tissue models. *Alternatives to Laboratory Animals (ATLA)* 43: 29–38.

Hoffmann, S., Kleinstreuer, N., Alépée, N., *et al.* (2018) Non-animal methods to predict skin sensitization (I): the Cosmetics Europe database *Critical Reviews in Toxicology.* 48(5): 344-358. https://doi.org/10.1080/10408444.2018.1429385

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Kaini, P., Sander, J.D., Joung, J.K., Peterson, R.T., and Yeh, J.-R.J., (2013) Heritable and precise zebrafish genome editing using a CRISPR-Cas system. *PLoS One* 8(7): e68708. https://dx.doi.org/10.1371/journal.pone.0068708

ICCR (2014) Report: *In silico* approaches for safety assessment of cosmetic ingredients. http://www.iccr-cosmetics.org/files/5314/1407/7607/2014-07_In-silico_Approaches_for_Cosmetic_Product_Safety_Assessments.pdf

ICCR (2017) Report: Joint Regulators-Industry Working Group: Integrated Strategies for Safety Assessments of Cosmetic Ingredients - Part I. http://www.iccr-cosmetics.org/files/4715/0824/0761/ICCR-11_JWG_Integrated_Strategies_Integrated_Strategies_for_Safety_Assessments_of_Cosmetic_Ingredients_-_Part_I.pdf

Jamin, A., Sr (2015) Predicting respiratory toxicity using a human 3D airway (EpiAirway™) model combined with multiple parametric analysis. *Applied In Vitro Toxicology* 1: 55–65. https://dx.doi.org/10.1089/aivt.2014.0003

Jiang, X., Wink, S., van de Water, B. and Kopp-Schneider, A. (2016) Functional analysis of high-content high-throughput imaging data. *Journal of Applied Statistics.* 44(11): 1903-1919. https://dx.doi.org/10.1080/02664763.2016.1238048

Kroes, R., Renwick, A.G., Cheeseman, M., Kleiner, J., Mangelsdorf, I., Piersma, A., Schilter, B., Schlatter, J., van Schothorst, F., Vos, J.G., and Würtzen, G. (2004) Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. *Food and Chemical Toxicology* 42(1): 65-83. https://dx.doi.org/10.1016/j.fct.2003.08.006

Kroes, R., Renwick, A.G., Feron, V., Galli, C.L., Gibney, M., Greim, H., Guy, R.H., Lhuguenot, J.C., and van de Sandt, J.J.M. (2007) Application of the threshold of toxicological concern (TTC) to the safety evaluation of cosmetic ingredients. *Food and Chemical Toxicology* 45(12): 2533-2562. https://dx.doi.org/10.1016/j.fct.2007.06.021

Lobenhofer, E.K., Cui, X. , Bennett, L. , Cable, P.L. , Merrick, B.A. , Churchill, G.A. , and Afshari, C.A. (2004) Exploration of low-dose estrogen effects: identification of No Observed Transcriptional Effect Level (NOTEL). *Toxicologic Pathology* 32(4) 482-492. https://dx.doi.org/10.1080/01926230490483324

Liu J., Zhou Y., Qi X., Chen J., Chen W., Qiu G., Wu Z., and Wu N., (2017) CRISPR/Cas9 in zebrafish: an efficient combination for human genetic diseases modelling. *Human Genetics* 136(1): 1-12. https://dx.doi.org/10.1007%2Fs00439-016-1739-6

Low, Y., Sedykh, A., Fourches, D., Golbraikh, A., Whelan, M., Ivan Rusyn, I., and Tropsha, A. (2013)Integrative Chemical−Biological Read-Across Approach for Chemical Hazard Classification. *Chemical Research in Toxicology*26: 1199−1208. http://pubs.acs.org/doi/pdf/10.1021/tx400110f

Maxwell, G., Mackay, C., Cubberley, R., Davies, M., Gellatly, N., Glavin, S., Gouin, T., Jacquoilleot, S., Moore, C., Pendlington, R., Saib, O., Sheffield, D., Stark, R., and Summerfield, V. (2014) Applying the skin sensitisation adverse outcome pathway (AOP) to quantitative risk assessment. *Toxicology in Vitro* 28(1): 8-12. https://dx.doi.org/10.1016/j.tiv.2013.10.013

McNamee, PM., Api, AM., Basketter, DA., Frank Gerberick, G., Gilpin, DA., Hall, BM., Jowsey, I., and Robinson, MK. (2008) A review of critical factors in the conduct and interpretation of the human repeat insult patch test. *Regulatory Toxicology and Pharmacology* 52(1): 24-34. https://dx.doi.org/10.1016/j.yrtph.2007.10.019

Munro, I.C., Ford, R.A., Kennepohl, E., and Sprenger, J.G. (1996) Correlation of structural class with no-observed-effect levels: A proposal for establishing a threshold of concern. *Food and Chemical Toxicology* 34(9): 829-867. https://dx.doi.org/10.1016/S0278-6915(96)00049-X

OECD QSAR Toolbox. https://www.qsartoolbox.org/

OECD (2004) Report: Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure-Activity Relationship Models. http://www.oecd.org/env/ehs/risk-assessment/37849783.pdf

OECD (2014) Environment Directorate Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology. *Guidance on Grouping of Chemicals, Second Edition. Series on Testing & Assessment No. 194*. Vol. OECD Environment, Health and safety Publications. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2014)4&doclanguage=en

OECD (2017) Chemical Safety Assessment Workflow based on Exposure Considerations and Non-Animal Methods. Series on Testing & Assessment No. 275. ENV/JM/MONO (2017)27. OECD

Publishing, Paris. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2017)27&doclanguage=en

OECD Test No.428: Skin Absorption: *In Vitro* Method. Organisation for Economic Cooperation and Development, Paris. https://www.oecd-ilibrary.org/environment/test-no-428-skin-absorption-in-vitro-method_9789264071087-en

OECD Test No. 431: *In Vitro* skin corrosion: reconstructed human epidermis (RHE) test method. Organisation for Economic Cooperation and Development, Paris. https://www.oecd-ilibrary.org/environment/test-no-431-in-vitro-skin-corrosion-reconstructed-human-epidermis-rhe-test-method_9789264264618-en

OECD Test No. 439: *In Vitro* Skin Irritation - Reconstructed Human Epidermis Test Method . Organisation for Economic Cooperation and Development, Paris. https://www.oecd-ilibrary.org/environment/test-no-439-in-vitro-skin-irritation-reconstructed-human-epidermis-test-method_9789264242845-en

OECD Test No. 442C: *In Chemico* Skin Sensitisation: Direct Peptide Reactivity Assay (DPRA). Organisation for Economic Cooperation and Development, Paris. http://dx.doi.org/10.1787/9789264229709-en

OECD Test No. 442D: *In Vitro* Skin Sensitisation : *ARE-Nrf2 Luciferase Test Method.* Organisation for Economic Cooperation and Development, Paris. http://dx.doi.org/10.1787/9789264229822-en

OECD Test No. 442E: *In Vitro* Skin Sensitisation : *In Vitro Skin Sensitisation assays addressing the Key Event on activation of dendritic cells on the Adverse Outcome Pathway for Skin Sensitisation.* Organisation for Economic Cooperation and Development, Paris. http://dx.doi.org/10.1787/9789264264359-en

OECD Test No. 455: *Performance-Based Test Guideline for Stably Transfected Transactivation In Vitro Assays to Detect Estrogen Receptor Agonists and Antagonists.* Organisation for Economic Cooperation and Development, Paris. http://dx.doi.org/10.1787/9789264265295-en

OECD Test No. 492: Reconstructed human Cornea-like Epithelium (RhCE) test method for identifying chemicals not requiring classification and labelling for eye irritation or serious eye damage. Organisation for Economic Cooperation and Development, Paris. https://www.oecd-ilibrary.org/environment/test-no-492-reconstructed-human-cornea-like-epithelium-rhce-test-method-for-identifying-chemicals-not-requiring-classification-and-labelling-for-eye-irritation-or-serious-eye-damage_9789264242548-en

Partosch, F., Mielke, H., Stahlmann, R., Kleuser, B., Barlow, S., and Gundert-Remy, U. (2015) Internal threshold of toxicological concern values: enabling route-to-route extrapolation. *Archives of Toxicology* 89(6): 941-948. https://dx.doi.org/10.1007/s00204-014-1287-6

Patlewicz, G., Ball, N., Booth, E. D., Hulzebos, E., Zvinavashe, E., and Hennes, C. (2013) Use of category approaches, read-across and (Q)SAR: General considerations. *Regulatory Toxicololgy Pharmacology* 67: 1-12. http://dx.doi.org/10.1016/j.yrtph.2013.06.002

Patlewicz, G., Helman, G., Pradeep, P., and Shah, I. (2017) Navigating through the minefield of read-across tools: A review of in silico tools for grouping. *Computational Toxicology* 3: 1-18. https://doi.org/10.1016/j.comtox.2017.05.003

Patlewicz, G., Cronin, M.T.D., Helman, G., Lambert, J.C., Lizarraga, L.E., and Shah, I. (2018) Navigating through the minefield of read-across frameworks: A commentary Perspective. *Computational Toxicology* 6: 39-54. https://doi.org/10.1016/j.comtox.2018.04.002

Philip, R.C., Rodriguez, J.J., Niihori, M., Francis, R.H., Mudery, J.A., Caskey, J.S., Krupinski, E., and Jacob, A., (2018) Automated high-throughput damage scoring of zebrafish lateral line hair cells after ototoxin exposure. *Zebrafish* 15(2): 145-155. https://dx.doi.org/10.1089/zeb.2017.1451

Ponsoda, X., Pareja, E., Gómez-Lechón, M-J., Fabra, R., Carrasco, E., Trullenque, R., and Castell, J.V. (2001) Drug biotransformation by human hepatocytes. *In vitro/in vivo* metabolism by cells from the same donor. *Journal of Hepatology* 34 (19-25) https://doi.org/10.1016/S0168-8278(00)00085-4

Reisinger, K., Blatz, V., Brinkman, J., Downs, T., Fischer, A., Henkler, F., Hoffmann, S., Krul, C., Liebsch, M., Luch, A., Pirow, R., Reus, A., Schultz, M., and Pfuhler, S. (2018) Validation of the 3D Skin Comet assay using full thickness skin models: Transferability and reproducibility. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* 827: 27-41. https://dx.doi.org/10.1016/j.mrgentox.2018.01.003

Regulation (EC) No 1907/2006 Of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). OJ L 396 https://eur-lex.europa.eu/legal-content/en/TXT/PDF/?uri=CELEX:02006R1907-20180301&from=EN

Renwick, A.G. (2005) Structure-based thresholds of toxicological concern – guidance for application to substances present at low levels in the diet. *Toxicology and Applied Pharmacology* 207(2):S585-S591. https://doi.org/10.1016/j.taap.2005.01.034

Safford, R.J., Api, A.M., Roberts, D.R., and Lalko, J.F. (2015) Extension of the Dermal Sensitisation Threshold (DST) approach to incorporate chemicals classified as reactive. *Regulatory Toxicology and Pharmacology* 72: 694–701. https://doi.org/10.1016/j.yrtph.2015.04.020

SCCS (Scientific Committee on Consumer Safety), SCCS Notes of Guidance for the Testing of Cosmetic Ingredients and their Safety Evaluation 9th revision, 29 September 2015, SCCS/1564/15, revision of April 2016. https://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_o_190.pdf

Schultz, T. W., Amcoff, P., Berggren, E., Gautier, F., Klaric, M., Knight, D. J., Mahony, C., Schwarz, M., White, A., and Cronin, M. T. D. (2015) A strategy for structuring and reporting a read-across prediction of toxicity. *Regulatory Toxicology and Pharmacology* 72: 586-601. http://dx.doi.org/10.1016/j.yrtph.2015.05.016

Schultz, T.W. and Cronin, M.T.D. (2017) Lessons learned from read-across case studies for repeated-dose toxicity, *Regulatory Toxicology and Pharmacology* 88: 185–191. http://dx.doi.org/10.1016/j.yrtph.2017.06.011.

Tetko, I.V., Maran, U. and Tropsha, A. (2017) Public (Q)SAR services, integrated modeling environments, and molecular repositories on the web: state of the art and perspectives for future development. *Molecular Informatics* 36(3), 1600082. http://dx.doi.org/10.1002/minf.201600082

Toxtree. http://toxtree.sourceforge.net/

US FDA (2016) Safety Testing of Drug Metabolites: Guidance for Industry. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER). https://www.fda.gov/downloads/Drugs/.../Guidances/ucm079266.pdf

Wambaugh, J.F. , Hughes, M.F., Ring, C.L., MacMillan, D.K., Ford, J., Fennell, T.R., Black, S.R., Snyder, R.W., Sipes, N.S., Wetmore, B.A., Westerhout, J., Setzer, R.W., Pearce, R.G., Simmons, J.E. and Thomas, R.S. (2018) Evaluating *in vitro-in vivo* extrapolation of toxicokinetics. *Toxicological Sciences* 163(1): 152-169. https://dx.doi.org/10.1093/toxsci/kfy020

Wang, Z., Samanipour, R., Koo, K., and Kim, K. (2015) Organ-on-a-Chip platforms for drug delivery and cell characterization: a review. *Sensors and Materials* 27(6): 487-506. https://dx.doi.org/10.18494/SAM.2015.1136

Wetmore, B.A., Wambaugh, J.F., Allen, B., Ferguson, S.S., Sochaski, M.A., Setzer R.W., Houck K.A., Strope C.L., Cantwell K., Judson R.S., LeCluyse E., Clewell H.J., Thomas R.S., and Andersen M.E. (2015) Incorporating High-Throughput Exposure Predictions With Dosimetry-Adjusted *In Vitro* Bioactivity to Inform Chemical Toxicity Testing. *Toxicological Sciences* 148(1): 121-136. https://doi.org/10.1093/toxsci/kfv171

Williams, F.M., Rothe, H., Barrett, G., Chiodini, A., Whyte, J., Cronin, M.T.D., Monteiro-Riviere, N.A., Plautz, J., Roper, C., Westerhout, J., Yang, C., and Guy, R.H. (2016) Assessing the safety of cosmetic chemicals: Consideration of a flux decision tree to predict dermally delivered systemic dose for comparison with oral TTC (Threshold of Toxicological Concern). *Regulatory Toxicology and Pharmacology* 76: 174-186. https://doi.org/10.1016/j.yrtph.2016.01.005

World Health Organization (2010) Characterization and Application of Physiologically Based Pharmacokinetic Models in Risk Assessment. IPCS Harmonization Project Document No. 9 http://www.who.int/ipcs/methods/harmonization/areas/pbpk_models.pdf?ua=1)

World Medical Association (2013) World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *Journal of the American Medical Association*. 310(20): 2191-4. https://doi.org/10.1001/jama.2013.281053

Worth, A., Cronin, M., Enoch, S., Fioravanzo, E., Fuart-Gatnik, M., Pavan, M., and Yang, C. (2012) Applicability of the Threshold of Toxicological Concern (TTC) approach to cosmetics – preliminary analysis. COSMOS project. https://dx.doi.org/10.2788/5059

Worth, A., and Patlewicz, G., (2016) Integrated approaches to testing and assessment. *Advances in Experimental Medicine and Biology* 856: 317-342. https://doi.org/10.1007/978-3-319-33826-2_13

Yang, C., Barlow, S.M., Muldoon Jacobs, K.L., Vitcheva, V., Boobis, A.R., Felter, S.P., Arvidson, K.B., Keller, D., Cronin, M.T.D., Enoch, S., Worth, A., and Hollnagel, H.M. (2017) Thresholds of toxicological concern for cosmetics-related substances: new database, thresholds, and enrichment of chemical space. *Food and Chemical Toxicology* 109(Part 1): 170–193. https://doi.org/10.1016/j.fct.2017.08.043

Zager, M.G. and Barton, H.A. (2012) A multiscale, mechanism-driven, dynamic model for the effects of 5α-reductase inhibition on prostate maintenance. *PLoS One* 7(9): e44359. https://dx.doi.org/10.1371/journal.pone.0044359

Zock, J.M. (2009) Applications of high content screening in life science research. *Combinatorial Chemistry and High Throughput Screening* 12(9): 870–876. https://dx.doi.org/10.2174%2F138620709789383277

Zhu, H., Bouhifd, M., Donley, E., Egnash, L., Kleinstreuer, N., Kroese, E. D., Liu, Z., Luechtefeld, T., Palmer, J., Pamies, D., Shen, J., Strauss, V., Wu, S., Hartung, T. (2016) Supporting read-across using biological data. *Alternatives to Animal Experimentation (ALTEX)* 33: 167-182. http://dx.doi.org/10.14573/altex.1601252