Workshop on the validation and regulatory acceptance of innovative 3R approaches in regulatory toxicology - Evolution versus revolution

Burgdorf, T; Piersma, A H; Landsiedel, R; Clewell, R; Kleinstreuer, N; Oelgeschläger, M; Desprez, B; Kienhuis, A; Bos, P; de Vries, R

Total number of authors:
18

Published in:
Toxicology in Vitro

Link to article, DOI:
10.1016/j.tiv.2019.03.039

Publication date:
2019

Document Version
Peer reviewed version

Workshop on the validation and regulatory acceptance of innovative 3R approaches in regulatory toxicology – Evolution versus Revolution

T. Burgdorf[1], A.H. Piersma[2,3], R. Landsiedel[4], R. Clewell[5], N. Kleinstreuer[6], M. Oelgeschläger[1], B. Desprez[7], A. Kienhuis[2], P. Bos[8], R. de Vries[9], L. de Wit[8], T Seidle[10], J. Scheel[11], G. Schönfelder[1,14], J. van Benthem[2], A.M. Vinggaard[12], C. Eskes[13] , J. Ezendam[2]

[1] German Centre for the Protection of Laboratory Animals (Bf3R), German Federal Institute for Risk Assessment, Berlin, Germany

[2] National Institute for Public Health and the Environment (RIVM), Center for Health Protection, Bilthoven, Netherlands

[3] Institute for Risk Assessment Sciences, Utrecht University, Netherlands

[4] BASF SE, Ludwigshafen, Germany

[5] ScitoVation, 6 Davis Drive, PO Box 12878, Research Triangle Park, NC 27709, United States

[6] NIH/NIEHS/DNTP/NICEATM, Research Triangle Park, NC, USA

[7] Cosmetics Europe, Brussels, Belgium

[8] National Institute for Public Health and the Environment (RIVM), Centre for Nutrition, Prevention and Health Services, Bilthoven, Netherlands

[9] Evidence-based Toxicology Collaboration, Johns Hopkins Bloomberg School of Public Health, Baltimore, USA & SYRCLE, Department for Health Evidence, Radboud Institute for Health Sciences, Radboudumc, Nijmegen, The Netherlands

[10] Humane Society International, Toronto, Canada

[11] Evonik Performance Materials GmbH, Darmstadt, Germany

[12] National Food Institute, Technical University of Denmark, Kemitorvet building 202, DK-2800 Kgs.Lyngby, Denmark

[13] SeCAM, Magliaso, Switzerland

[14] Charité Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health

Corresponding Author:
Michael Oelgeschläger,
BfR, Max-Dohrn-Str. 8-10, 10589 Berlin, Germany,
Email Michael.Oelgeschlaeger@bfr.bund.de,
Phone +49 30 18412-2292

Abstract

At a joint workshop organized by RIVM and BfR, international experts from governmental institutes, regulatory agencies, industry, academia and animal welfare organizations discussed and provided recommendations for the development, validation and implementation of innovative 3R approaches in regulatory toxicology. In particular, an evolutionary improvement of our current approach of test method validation in the context of defined approaches or integrated testing strategies was discussed together with a revolutionary approach based on a comprehensive description of the physiological responses of the human body to chemical exposure and the subsequent definition of relevant and predictive *in vitro, in chemico or in silico* methods. A more comprehensive evaluation of biological relevance, scientific validity and regulatory purpose of new test methods and assessment strategies together with case studies that provide practical experience with new approaches were discussed as essential steps to build up the necessary confidence to facilitate regulatory acceptance.

Keywords: workshop; validation, evolution, revolution, relevance, case studies

1. General Introduction

(Tanja Burgdorf, BfR)

In 2017, RIVM and BfR organized a joint workshop entitled, “Validation redefined? Workshop on validation and regulatory acceptance of alternative methods and test strategies”. During this workshop, regulatory authorities, test method developers, and users from academia, industry and CRO as well as experts participating in the Test Guidelines Programme of the Organization for Economic Cooperation and Development (OECD) discussed the drivers and barriers concerning implementation of innovative 3R approaches in regulatory toxicology (Piersma et al., 2018a). As a result of the first workshop, two major scenarios allowing the implementation of alternative testing strategies were identified, an evolutionary and a revolutionary one. The evolutionary approach aims at an optimization of the current system towards testing strategies for toxicological endpoints in a stepwise fashion as non-animal approaches become available. This approach relies on the definition of biological and chemical space as well as technical validation (e.g. reproducibility and transferability) of individual assays which is crucial for determining its usefulness and positioning within a testing strategy. The revolutionary approach begins with the question: Which parts of human physiology need to be addressed in test systems in order to cover relevant mechanisms of toxicity? Downstream follows the design of testing strategies and the selection of a series of appropriate complementary assays for critical key events comprising the network of toxicity pathways. The need for a comprehensive description of biological processes, for example, by the description of adverse outcome pathways (AOPs) and mapping of interacting AOP networks had been identified as a key element of future work. This activity could allow the identification of essential key events (KE) and the development and implementation of relevant and reliable test methods addressing these KE as well as the targeted validation of these assays in the context of defined approaches (DA) which are rule-based approaches relying on a fixed data interpretation procedure (DIP) (OECD, 2016c). The fixed structure of DAs should facilitate their use and regulatory acceptance (Casati et al., 2018). As AOPs and novel, innovative alternative methods further evolve, there is an increasing need to integrate information from multiple sources into testing strategies (Piersma et al., 2018b). The DAs for skin sensitisation are the first examples of how test strategies can be harmonized and standardized and how they could be integrated into legal frameworks such as the OECD test guidelines programme to be covered by the principle of Mutual Acceptance of Data (MAD) (OECD, 2016b). However, further discussion is needed on how to validate test strategies, to ensure that they comply with regulatory requirements and to warrant regulatory acceptance. A follow-up joint workshop was organized by RIVM and BfR in 2018 aiming at developing feasible solutions and recommendations that facilitate the development of an evolutionary as well as revolutionary approach taking into consideration that both have their merits and limitations and both are needed in concert to move innovation in this area forward.

In a series of introductory talks the results of the last workshop were summarized and different points of view in respect to the evolutionary and revolutionary approach were presented by experts from RIVM, BASF and ScitoVation. In three breakout groups different topics were addressed. The first group discussed how to facilitate the regulatory acceptance of DAs and Integrated Approaches to Testing and Assessment (IATAs) as part of the evolutionary approach, in particular in respect to

assessment criteria, validation and building confidence. The second group focused more on the revolutionary approach. Not only the question of how the transfer from test method development into validation can be promoted was discussed but also if new methods should always be validated as part of testing strategies. The third group addressed whether case studies can provide a way to innovate validation and to converge evolutionary and revolutionary activities.

2. Lecture summaries

2.1 Setting the Scene: Evolution versus Revolution in Innovating Regulatory Toxicity Testing (Aldert H. Piersma, RIVM)

Following up from the first BfR-RIVM workshop (Piersma et al., 2018a), the scene of this workshop was set by introducing the concept of evolution versus revolution in innovating toxicity testing (Scialli et al., 2018). The historic sequence of human safety testing has proceeded through the introduction of animal methods in the mid-20th century, followed by development of in vitro alternatives in later decades. For these alternatives, validation against existing animal studies was deemed necessary, to assess their performance relative to existing hazard assessment methodologies. It is important to recognize that the original animal studies have never been validated before formal introduction into regulatory toxicology. In addition, validation against existing animal test methods as the gold standard ignores the fact that the animal is not always predicting human risk, as study designs may lack sufficient power and may miss important end points, and relevant human mechanisms may not be reproduced in animals. In fact, alternative approaches should be evaluated against knowledge of human biology which of course is challenging.

Within the context of workshop discussions, evolution was characterized as the improvement of current methods, using the animal study as the gold standard, and focused within current legislation. Revolution, on the other hand, would be a process starting from scratch, using human biology as the gold standard, and independent of current legislation. In other words, revolutionary approaches will be considered relevant to the extent they cover the necessary elements of human biology to allow reliable toxicity prediction, and will possibly lead to a different system of regulatory acceptance of chemicals on the market.

The AOP approach was considered a useful way of pointing to a transition toward mechanistic toxicology rather than adverse end point driven toxicological hazard and risk assessment (Conolly et al., 2017). On the other hand, the current AOP approach has overly simplistic elements, as physiology is not one-directional, toxicological mechanisms are not linear, and AOPs do not work in isolation. Instead, complex quantitative network modelling is probably necessary for adequate toxicity prediction (Staal et al., 2017). The example of endocrine feedback homeostasis was given as an essential mechanism of sustaining life, and in that context toxicity can be described as an effect that over rides homeostatic control.

The idea of a toxicological ontology was described as the network of quantitative interactions between rate-limiting key events that lead from exposure to adverse health effect. As to practical application of

the ontology, coverage of the rate-limiting key events in quantitative alternative assays, combined and extrapolated to the intact human using quantitative in vitro to in vivo extrapolation (QIVIVE) modelling should in principle allow animal-free risk assessment. A current project at RIVM, aiming at a developmental ontology, makes use of chemistry, toxicological as well as fundamental developmental biology data to mechanistically map neural tube closure (Hessel et al., 2018), (Baker et al., 2018). This map will be used to design a computational model for neural tube closure, which will also allow the assessment of adverse effect by chemicals causing critical gene expression changes. The example of the retinoic acid pathway was presented as a key regulating pathway in vertebrate embryogenesis, which is likely affected by many dysmorphogenic compounds (Tonk et al., 2015), (Piersma et al., 2017) .

Several computational systems have been developed at US EPA in its Virtual Embryo Project, e.g. regarding prediction of effects on blood vessel development, showing proof of principle of such approaches (Kleinstreuer et al., 2013). Bringing computational toxicology to the integrative level of the intact virtual human is the ultimate challenge, which will require big data analysis. Interestingly, the description of neural networks is reminiscent of the AOP scheme (Koutsoukas et al., 2017). Input nodes, hidden nodes, connections and output nodes in neural networks can be replaced by initiating events, key events, key event relationships and adverse outcomes, respectively, to arrive at an AOP network. In the era of big data and artificial intelligence, toxicity prediction can benefit from machine learning (Wu and Wang, 2018). The challenge is in feeding the system with sound data to allow machine learning to the level necessary to cover all essentials of toxicological pathways.

Given the toxicodynamic focus of AOPs, they should be coupled to kinetic models that describe the fate of a compound in the body after external exposure (Wetmore et al., 2015). This fate determines which initiating event(s) will be challenged by the exposure, and what will be the corresponding concentration-time characteristics. This is essential for appropriate quantitative risk assessment.

Getting the revolutionary approach to work requires a substantial to do list. Human physiology needs to be mapped to the level of detail fit for purpose for toxicity testing. Chemistry, biology and toxicology knowledge needs to be integrated in an ontology that will drive defining the quantitative key event network of toxicity pathways. This ontology will then drive computational modeling of the system, and defining the necessary rate-limiting key events that require quantitative in vitro testing. The integration of this toxicodynamic model with kinetic models for compound absorption, distribution, metabolism and excretion (ADME), and quantitative in vitro to in vivo extrapolation (Fragki et al., 2017) is necessary to complete the model for computational hazard and risk assessment.

The challenges are great, but novel tools in computational systems open doors to unseen landscapes that are still to be discovered. Toxicological risk assessment cannot afford not to embark on these discoveries. From a pragmatic point of view many novel revolutionary methods may already be usefully employed in our current regulatory landscape, which will facilitate ultimate transition to a novel, animal-free human-based system for chemical hazard and risk assessment. This also points to the necessity to involve all stakeholders in the process, be they from academia, government, industry, NGOs or the political arena. It also requires intensified interdisciplinary collaboration among scientists, be they biologists, toxicologists, clinicians, data scientists, software developers, artificial intelligence

experts, etc. The stakes for leaving the experimental animal and embarking on human-relevant paradigms are high and worth taking the challenge.

2.2 On the ongoing evolution of toxicological assessment and testing

(Robert Landsiedel, BASF SE)

A change in paradigms and methodology for risk assessment and validation (Zurlo, 1994), (Davis et al., 2013), (Hartung, 2008), (Leist et al., 2008), (Scialli et al., 2018) is already well on the way – even without the need for a revolution (Monosson, 2005). Classical toxicological assessment, especially for regulatory purposes, is subjected to rigid rules and relies heavily on data obtained from pre-defined lists of standarized (mainly animal) studies described in OECD test guidelines (http://www.oecd.org/env/ehs/testing/section4-health-effects.htm) as exemplified in the REACh Annexes (EU, 2006), or the Classification, Labelling and Packaging (CLP) regulation (EU, 2008)). These are, however, no stagnant systems. In fact there are periodical Adaptations to Technical Progress (ATP). Yet, there is a disconnect between the development of a mechanistic, modern toxicology and the rigid approaches in the regulatory field (Hartung et al., 2009). In the past, toxicology has largely been driven by the demand for protocols for regulatory actions (Lotti and Nicotera, 2002). In fact, biomedical sciences are developing and using a plethora of new methods (MacGregor, 2003), (Um et al., 2018) and toxicological science is taking advantage of this and has adopted new methods as well as developed new concepts. Some of these are listed in Figure 1. Yet, their adoption in regulatory toxicology has just started.

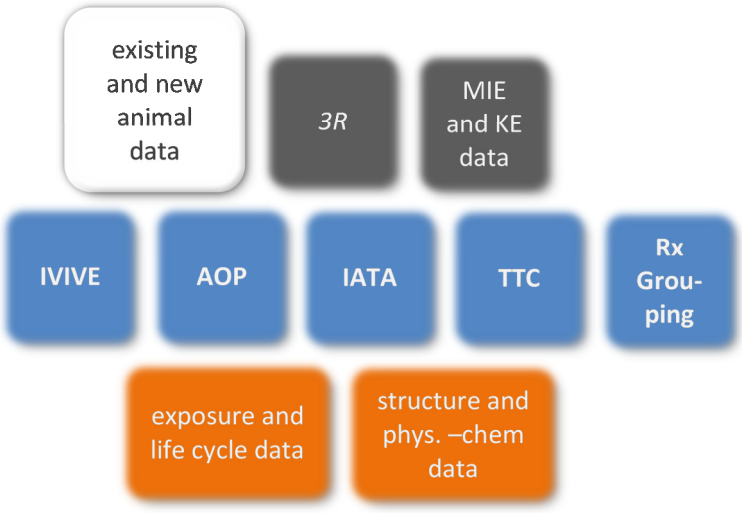

Figure 1 Elements of current toxicological assessments (IVIVE: in vitro to in vivo extrapolation; AOP: adverse outcome pathway, IATA: integrated approach to testing and assessment, TTC: threshold of toxicological concern, Rx: read-across, MIE: molecular initiating event, KE: key event)

Biological evolution is driven by new traits (as a phenotype of genetic mutations) and by their competitive advantage within the given environment which is leading to greater reproductive success (natural selection). In toxicological assessments there is a wealth of new traits (methods, concepts)

available but the (regulatory) success is hardly there yet. The question is whether the environment of regulations wants or demands new traits (and which ones). The natural selection and the evolution of species (Darwin, 1895) is complex. The tree of life (Haeckel, 1878) can be described as a development from simple prokaryotic cells to complex organisms consisting of trillions of eukaryotic cells with specific functions. Interestingly, toxicological testing evolved in just the opposite direction: from testing on humans to more simple models. Models based on target biomacromolecules, bacteria, and mammalian cells or tissues are generally less complex than humans or whole animal models. Mostly, rather simple models are combined to account for the complexity of events leading to adverse effects in whole organism (such approaches are termed IATA, ITS, STS and DA; the terms are explained by Sauer *et al*., 2016). In fact, there may be more sophistication in combining and interpreting data than from actually generating data by new experimental methods.

It should be understood, that evolution has no long-term goal. Rather, Charles Darwin made clear: "It is not the strongest of the species that survives, nor the most intelligent that survives. It is the one that is the most adaptable to change." The change of requirements drives the evolution of toxicological methods just as much as the development of new methods. These requirements are shaped by public's desire for rapid and precautionary solutions to potential hazards and by industry and regulatory need for unambiguous, easily comprehensible assessments and foreseeable regulatory decisions. In the current 'environment' different factors demand changes of the toxicological assessment praxis (Table 1) whereas other factors favour the *status quo* (Table 2). Unlike in biological evolution, it is our task to shape the environment that prefers one toxicological method or concept over another. In the evolution of toxicological assessments, we will eventually have those methods in use, which are the best fit for the requirements which we have defined.

Table 1 Environmental factors demanding adaptation and thus driving evolution

Capabilities and diversity of available methods
Success of new approaches (e.g. in screening)
Animal welfare
Disasters due to the lack of predictivity
Credibility crisis

Table 2 Environmental factors favouring the *status quo*

Hazard-based classification
Success of current assessments
Validation and Mutual acceptance of data (MAD)
Urge for unambiguity and ignoring uncertainty
Hesitant translation from science to regulation

In biological evolution species vanished (like the *Homo neanderthalensis*) and other – fitter - species (like the *Homo sapiens*) survived. In toxicological testing and assessment, numerous new approaches have been developed and are currently in use. The question is: Which ones are further evolving and which ones may have a dead-end?

In vitro methods to test for skin and eye irritation were developed relatively early and are in regulatory use – albeit not quite on equal footage with the animal studies (Sauer et al., 2016). The new

approaches to skin sensitization testing (Gabbert et al., 2017) are currently regarded as the best example for 'modern' toxicity testing and assessment: There are numerous new testing methods available, three have been regulatorily accepted (OECD test guidelines no. 442C, D and E) and there are several AOP-based IATAs (OECD, 2016b), three of which are currently evaluated to draft the first defined approach (DA)-based OECD test guideline. New concepts of toxicological testing and assessment have also been proposed and are being implemented for new materials such as nanomaterials (Burden et al., 2017), (Landsiedel et al., 2017). Methods which are termed 'non-testing methods' (methods using existing data rather than generating new data from experiments) such as grouping and read-across are widely used to fulfil REACH information requirements and are continuously improved and refined (Teubner and Landsiedel, 2015) including the use of computational toxicology (Luechtefeld et al., 2018), (Myatt et al., 2018). They all may be more '*sapiens*-like*'* methods but could turn out to be rather '*neanderthalensis'*. A relatively old set of methods, mutagenicity testing, may actually have all the traits we are (or should be) asking for future toxicological assessments: There are numerous methods available (*in vivo*, *in vitro* and *in silico* as well as successful read-across), there are flexible, hypothesis-driven testing schemes and the toxicological assessment is actually based on the molecular or cellular event rather than the adverse effect in the end (which is carcino- or teratogenicity).

For the current environment, descriptive human and animal data remain entrenched in regulation, yet if we want more humane methods providing data of higher human-relevance and including mechanistic understanding, regulatory requirements need to change. They are currently based on adverse effects which manifested in animals after substance administration and human data; whereas molecular or cellular events are merely used as an approximation of animal data or their *post hoc* mechanistic explanation.

Like any evolution, also the one in toxicological assessment will bring an unknown future, be rather slow (compared to revolution) and may be painful (there will be extinction!). Eventually, it is unimportant whether there will be evolution or revolution – the need for a change is indisputable. And we should do both, provide sufficient methods with different traits and shape the scientific, societal and regulatory environment. These will require beneficial collaboration of the academia, industry, regulators and society. As frightening as this evolution may seem, Charles Darwin provided some reassurance: "In the long history of humankind (and animal kind, too) those who learned to collaborate and improvise most effectively have prevailed".

2.3 On Revolution: A new paradigm for risk assessment needs a new paradigm for validation?
(Rebecca Clewell, Scitovation, Member of ESAC)

Validation of new approach methods is currently based on comparisons to in vivo animal data, assessing how well these alternative methods reproduce past results related to predicting organism level toxicity in intact animals. Several key tenets are relied on to bolster confidence in test results: libraries of control compounds that elicit known outcomes in a "gold standard" in vivo model (positive, negative, false positive), historical databases with well-defined ranges of acceptability for model readouts, and standardized protocols that ensure reproducibility across laboratories (e.g., confirmation

through ring trials). Success with this approach is demonstrated by the recent progress in defining in vitro testing strategies for prioritization of estrogen and androgen disruptors in the USEPA's endocrine disruptor screening program, and the use of Defined Approaches for skin irritation and sensitization as a replacement for in vivo tests in the European Union and the United States. However, the path to acceptance of these testing strategies also highlights the pitfalls of the current approaches to validation. First, the cost and time associated with these traditional validation approaches are prohibitive to rapid adoption of new approaches. Only 28 in vitro alternatives have been sufficiently validated to be accepted for regulatory use (McMullen et al., 2018). Second, as more alternatives are developed, it becomes less likely that large libraries of reference compounds that have been tested in animal models will be available for use as gold standards. Finally, given the limited correlation between different animal models and human response, the pertinent question becomes – "Is it appropriate to validate in vitro human models against in vivo rodent models when the goal is to predict in vivo human response?". Instead, it may be time to consider a revolution not only in approaches to chemical testing, but also in the approaches to validation of new tests.

Such a revolution would be a biological validation process taking advantage of the advances in our understanding of biology and medicine. The underlying assumption to a biologically-based approach is that assays designed to recapitulate the key events within an AOP, will inherently contain the necessary biological fidelity to predict toxicological outcome. This idea calls for us to use the best of available science to identify the key components of biological pathways and to incorporate these components into the test system. Thus, to focus not on the number of compounds tested, but on the design of the system and targeted testing of the system components with meaningful control compounds. The revolution is the shift in focus from finding comfort in comparing new approaches to traditional animal tests to trusting that a biologically relevant system will provide a human relevant estimate of chemical response. Case studies are described to demonstrate the value in using AOP-structured approaches to increase the utility and predictivity of in vitro assays, not only for potential hazard identification, but also for quantitative risk assessment, e.g., setting a point of departure (PoD).

The first case study addressed estrogen mediated proliferation in the uterus. High throughput screening (HTS) assays for estrogenic activity exist, but are focused on chemical classification rather than quantitative dose-response and they generally rely on simple model systems to test very early events in the estrogen pathway (receptor binding, receptor dimerization, transactivation) or measure responses in non-uterine cells. However, the development of an in vitro alternative to the in vivo uterotrophic assay should rather be based on the best current knowledge of the estrogen-mediated proliferation response in the uterus in order to be capable of predicting in vivo dose-response. First an AOP-like framework for the estrogen-mediated proliferation pathway was developed, focused on defining key events governing cellular outcome in order to identify the most important characteristics of the required *in vitro* assays (Miller et al., 2017) (Figure 2).

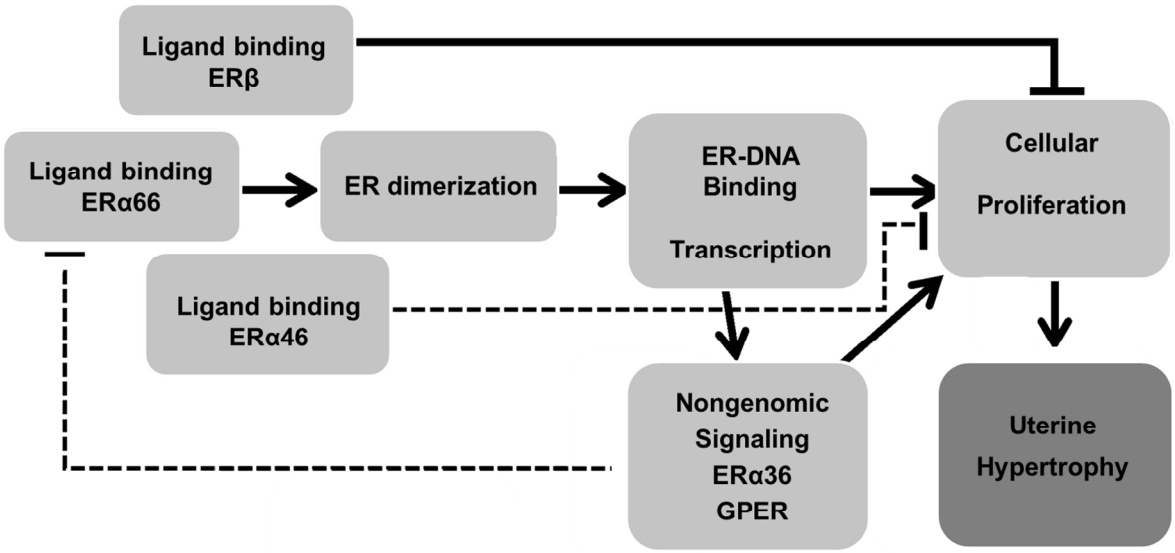

Figure 2. Proposed Adverse Outcome Pathway for Estrogen Receptor-mediated Uterine Hypertrophy Including Key Events and Key Event Relationships that Underpin Tissue Dose-Response. Figure adapted from Miller et al. (2017). In order to describe tissue dose-response, both negative (cross bars) and positive (shown with arrows) regulators of proliferative response must be accounted for. From this AOP framework, several estrogen receptors (ERs) that have unique signaling contributions to estrogen mediated cellular response and are important determinants of downstream cellular and tissue response were identified: ERβ, full-length ERα (ERα66), two short isoforms of ERα (ERα46, ERα36), and an estrogen binding G-coupled protein receptor (GPER) (Miller et al., 2016), (Filardo et al., 2000), (Penot et al., 2005), (Wang et al., 2006).

The Ishikawa human uterine adenocarcinoma cell line expresses in vivo relevant levels of these receptors and has the ability to recapitulate the phenotypic responses to estrogen treatment at the RNA, protein and cellular level, including cellular proliferation. However, the more important goal was to ensure that the model was quantitatively predictive of human response. To this end, several known estrogenic compounds including the endogenous ligand (17ß- estradiol) were tested in the Ishikawa model and the points of departure were compared to human in vivo data from clinical studies, as well as results from simple HTS assays (ToxCast/Tox21). The biologically based in vitro assay was consistently able to predict safe levels of exposure to human estrogens and was consistently more protective than the simple HTS assays (Figure 3), indicating that an AOP-based approach to assay design does improve utility for quantitative dose-response assessment.

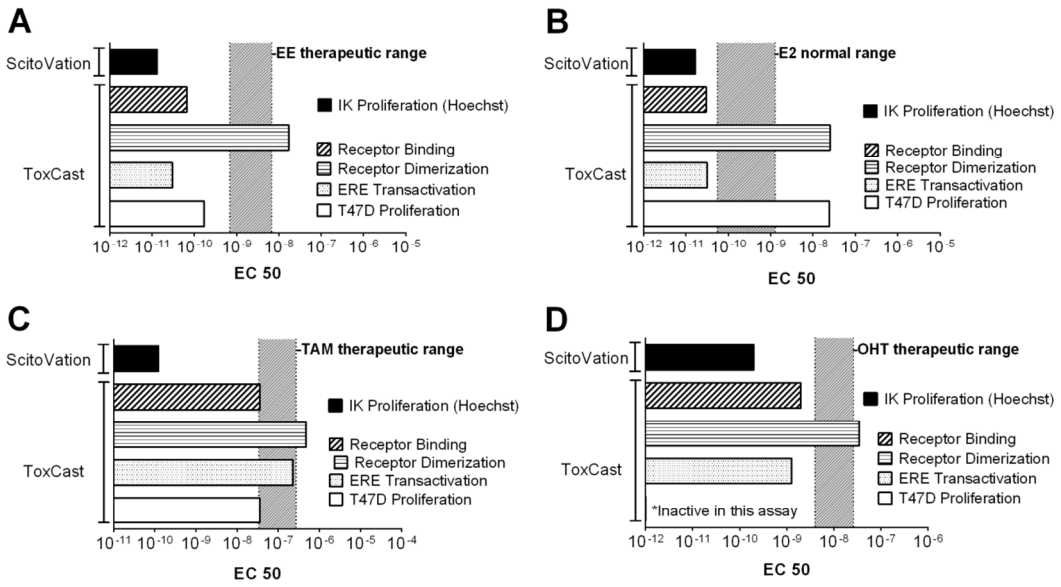

Figure 3. Comparison of EC50 values for estrogenic compounds across ToxCast assays or Ishikawa proliferation assay. A-D. For each compound, EC50s for fit-for-purpose (ScitoVation) assay proliferation data (IK proliferation) were calculated and are shown for comparison with published EC50 values from the ToxCast suite of assays (version 2, http://actor.epa.gov/dashboard/). Therapeutic levels for each compound are shown as a hatched overlay. ToxCast assay identifications: T47D Proliferation (ACEA_T47D_80hr_Positive), ERE Transcription (OT_ERα_EREGFP_0120), Receptor Binding (NVS_NR_hER), Receptor Dimerization (OT_ER_ERαERα_0480). Figure reproduced from Miller et al., (2016).

The second case study addressed development of an in vitro assay for a toxicological endpoint for which there is no “gold standard” in vivo (or in vitro) assay: adipogenesis (Sargis et al., 2010), (Lyssimachou et al., 2015), (den Broeder et al., 2017), (Foley et al., 2015). Adipogenesis is a tightly controlled program of gene expression that drives the differentiation of committed pre-adipocytes to mature functional adipocytes (Fig. 4) (Moseti et al., 2016). The obesogen hypothesis suggests that chemical induction of adipogenesis in prenatal development may predispose individuals to metabolic disease later in life by increasing their adipose stores. A novel human adipose-derived stem cell (hASC) assay to assess the adipogenic effects of enviromental chemicals has been developed (Foley et al., 2017), (Hartman et al., 2018) (Foley et al., 2015) based on the current knowledge of the key biological pathways that drive adipocyte differentiation applying a hormone cocktail of Peroxisome proliferator-activated receptor gamma (PPARγ), Glucocorticoidrezeptor (GR), and CCAAT/enhancer-binding-protein (C/EBP) agonists and insulin. This cocktail drives differentiation of human adipocyte-derived stem cells to a fully mature phenotype capable of lipid accumulation and adipokine secretion. This assay was found to be useful for evaluating chemical-induced adipogenesis via activation of PPARγ and GR (Foley et al., 2017), (Hartman et al., 2018). Testing GR agonists and antagonists that are also used for various clinical applications in the human population (Hartman et al., 2018) demonstrated that the assay was able to predict concentrations at which GR activity would be expected for natural and synthetic glucocorticoids. The assay also predicted the absence of GR activity at therapeutic concentrations for compounds targeting other hormone pathways, and the off-target effects of these compounds on GR at higher concentrations (i.e., loss of specificity above clinical concentrations).

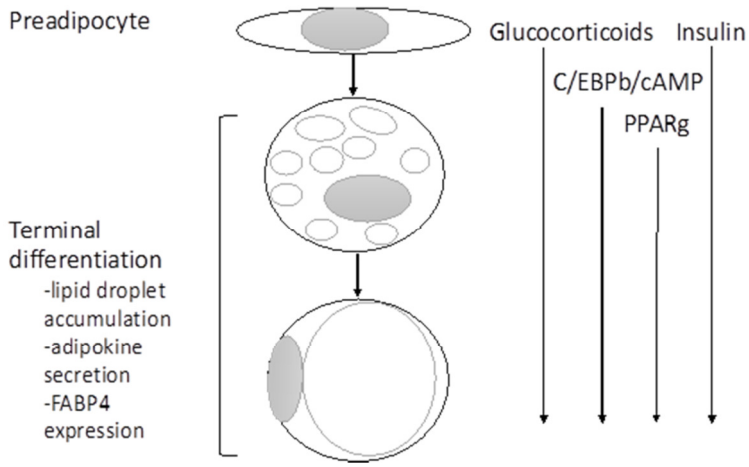

Figure 4. Adipogenesis pathway defined in vitro assay. An adipogenic stimulator activates the glucocorticoid receptor, leading to transcriptional activation of CEBPD and the downstream signaling cascade, including expression of PPARγ, which is the master regulator of adipocyte differentiation. To fully differentiate, a committed pre-adipocyte undergoes activation of adipogenic gene expression, followed by terminal differentiation, lipid accumulation, and ultimately adipokine secretion. To develop an in vitro model of adipogenesis, hASCs are treated with a cocktail that can induce adipogenesis (dexamethasone, rosiglitazone, insulin, and 3-Isobutyl-1-methylxanthin). The cells are simultaneously treated in test compound and differentiation is phenotypically assessed in a single well using three distinct phenotypic endpoints: adipogenic gene activation (FABP4 protein expression), lipid accumulation (AdipoRed neutral lipid staining), and adipokine secretion (adiponectin secretion). Figure adapted from Hartman et al., 2018).

These in vitro results indicate that the biologically-based design of the assay allows accurate prediction of the therapeutic index in humans. The favorable comparison of the in vitro predicted potency with clinical data supports the potential utility of these assays for predicting human response. By building the assay around the biological pathway of interest, ensuring that the system maintains the key biological factors that drive phenotypic responses, and framing the chemical responses' in the context of human therapeutic or environmental exposures, we also increase confidence in the interpretation of the resulting data.

3. Breakout discussion group summaries

3.1. Breakout group 1: Accelerating evolution: how to facilitate the regulatory acceptance of defined approaches and/or IATA

(Nicole Kleinstreuer (NICEATM) and Michael Oelgeschläger (BfR))

The group discussed challenges and opportunities on the path to a regulatory acceptance of defined approaches (DAs) and assessment strategies, like IATAs of which DAs may be part of. Developing DAs to a point comparable to test guidelines that fall under MAD was identified as an important step to ensure regulatory acceptance of the data worldwide. However, this raises the question of how to evaluate DAs or to what extent DAs can be assessed in an equivalent manner as single test methods in the development of test guidelines, and which performance criteria they are requested to meet.

The group agreed that the assessment criteria developed during the ICATM workshop on DAs in 2016 (Casati et al., 2018) and the work done in the OECD over the last two years (OECD, 2016c), (OECD, 2016b), (OECD, 2016a) constitute an excellent starting point. As has been defined for single method test guidelines, the purpose and applicability of the DA for the various regulatory systems must first be defined. It needs to be clear if the DA is for hazard identification only, for potency classification and labelling, or could also be used to derive safe no-effect-levels, in combination for example with in vitro-in-vivo extrapolations (IVIVE). Here, a deep understanding of regulatory needs, decision contexts and legal mandates is a prerequisite to define appropriate acceptance criteria. A survey in different regions might help to define the potential use of DAs under different regulations, with the UN GHS system as a common basis. Similarly, the limitations and applicability domain with respect to different chemical groups and DA performance need to be evaluated, which also requires a thorough selection of relevant reference chemicals and high quality reference data. Again, similar to single test methods, the structure of the DA needs to be biologically plausible and its DIP robust, transparent and reliable. Thus, in general the acceptance criteria hardly differ between DAs and single test methods. One difference might be that the use of multiple, complementary or even redundant methods might help to confirm the biological and toxicological relevance of single assays in a DA. In general, as true for single test methods, DA reproducibility is much more straightforward to define than its predictivity, given the dependence of sensitivity and specificity calculations on the number as well as the relevance of tested chemicals and the quality of in vivo reference data. A DA will need to be part of an IATA that describes its use in a risk assessment process. IATAs are never fully harmonized because they always include weight-of-evidence (WoE) that allows flexibility and some degree of expert judgment, depending on the specific case and regulatory context. This necessary level of flexibility, however, should not go too far because that might weaken acceptance. Regardless, a DA per definition is not flexible since it is based on a fixed DIP, and it needs to be clear to what extent the DA outcomes suffice for a regulatory decision. In addition, the DA based-IATA not only needs to facilitate acceptance of the DA data but also the acceptance of resulting regulatory decisions. With respect to performance evaluations, given the degree of uncertainty and variability associated with in vivo studies (e.g. (Kleinstreuer et al., 2016), (Hoffmann et al., 2018), (Browne et al., 2018), it might be better not to compare DA data directly with animal data to calculate values of predictivity. Rather, the ultimate

decision after DA-based IATA evaluations should be compared with current, accepted WoE decision matrixes to calculate the statistical performance.

In general, the OECD seems the best platform for discussion, since it provides the necessary opportunities for communication and harmonization between different regions. In addition, OECD expert groups that focus on specific case studies might provide important information on the applicability of DAs as well as experience to foster confidence. Here we need a transparent discussion of the qualitative and quantitative performance of current methods, as well as the levels of uncertainty we already accept by using them. Although difficult and requiring a comprehensive understanding of the endpoint, uncertainty evaluations and establishment of biological relevance are key components to build confidence in the DA. However, acceptance of uncertainty is not only a scientific but also a matter of social and, subsequent, political acceptance, which is particularly true for the acceptance of potential false-negative results. A frank discussion about already existing uncertainty associated with animal studies and acknowledging shortcomings of accepted methods is urgently needed to set appropriate expectations for DA evaluations.

The group felt that communication within and among regulatory authorities in various regions is of key importance to ensure that distinct legal requirements, but also differences in levels of awareness and understanding are considered. These include discussions at all levels: local, regional, national, international and should involve multi-stakeholder groups including regulators and regulated industry across all relevant sectors to discuss practical implementation and case examples. In addition, authorities and regions outside of the OECD should be approached to get them involved in transparent communication about relevant applications and regulatory contexts. Since regulators need as much information as possible to be able to follow the rationale behind a DA more communication and training using hands-on tools and real data, in particular for complex approaches like machine learning or IVIVE, would facilitate progress. The ever-increasing complexity and diversity of specific knowledge needed and technologies applied had resulted in an increased demand for continuing education and to empower regulators to understand and utilize novel methods and approaches. Here, as has been discussed for decades in basic research, education and training on science communication is needed. Additionally professional translators for conversations would help overcome language barriers as well as standardized, easy-to-communicate information material. Finally, it will always be more convincing if leaders in the field set an example and communicate experiences and success in the implementation of novel approaches with other authorities.

There is already a plethora of activities to address mechanisms of toxicity. However, these activities often lack focus with respect to toxicological endpoint to be addressed and regulatory purpose. It seems of utmost importance to identify and agree on areas where animal approaches might be insufficient or problematic, including the fact that studies may not have been designed to provide much mechanistic information, and where DAs can potentially provide more insight. As mentioned, a survey on needs with respect to the various decision contexts would be helpful, and work is ongoing in that area (e.g. (Strickland et al., 2018), (Daniel et al., 2018), (Choksi et al., 2018)). The group agreed that for the future it will be important to define frameworks that allow the regulation of substances based on comprehensively described mechanisms with a clearly defined level of disturbance that lead to an

adverse effect. Currently, this is most easily implemented for methods addressing substances that damage DNA or interfere with the endocrine system, due to the in-depth mechanistic understanding and excellent assay coverage of the relevant pathways. To ensure human relevance, the identification of mechanisms and pathway perturbations could be supported by the use of clinical data, since in the clinical markers that are indicative for relevant changes in physiology or specific organ toxicity are partially already known.

3.2 Breakout group 2: Towards revolution: a new paradigm for risk assessment needs a new paradigm for validation

(Bertrand Desprez (CE), Anne Kienhuis (RIVM))

The second group debated the statement that a revolutionary paradigm for risk assessment based on human biology and physiology needs a new paradigm for validation. The breakout group acknowledged that it is hard to give a definition of validation as a tool in a revolutionary framework without knowing how the framework would look like. The revolutionary framework for safety assessment is however expected to be constructed as a tiered approach. It may start with *in silico* and less complex *in vitro* models and proceed with models increasing in complexity that also include kinetics. In later tiers of the revolutionary framework more complex models may be included, such as organs-on-chips.

The breakout group identified two steps of validation that are important for the vast amount of less to more complex new technologies that are being developed: scientific validation and regulatory validation. It was discussed that the first step for new technologies and methods is '**scientific validation**', in which the technology/method is characterized, the physiological processes are modelled, and reproducibility of models of disease is assessed. After scientific validation, methods need to be validated for regulatory purposes in a '**regulatory validation**'. Here, it is assessed whether methods are fit-for-purpose. An example are the various models for hepatic metabolic clearance that are available (Krause and Goss, 2018): all models may be equal in characterizing the physiological process but may differ to the extent they are fit for the different purposes of safety assessment. For the lower tier *in vitro* high throughput screening methods, scientific validation may be sufficient. For the more complex approaches mimicking human physiology regulatory validation of the corresponding testing strategies (IATA or DA) should be performed, although the way this validation is done has to change (Piersma et al. 2018a), and focus on reliability.

Although validation is a slow process and often not effective regarding regulatory acceptance, it is still an important step in both the evolutionary and revolutionary approach in order to demonstrate relevance and reliability of in vitro methods and innovative technologies such as high-throughput assays, organ-on-a-chip, and mathematical models. However, traditional validation of in vitro tests is typically performed against an in vivo database that originates from animal experiments, and based on comparisons between new in vitro data and existing in vivo data, sensitivity and specificity are calculated. These values only make sense if the animal model has a good predictivity for human responses, but this knowledge is often lacking. Therefore, rather than using the predictive capacity of

a method described by specificity and sensitivity to demonstrate relevance, we should focus on the biological plausibility or biological statistical relevance of a method from the revolutionary point of view.

In this process, the extent to which the test method represents human relevant mechanisms of toxicity as well as the limitations of the test method has to be clearly defined. This should not be assessed by testing numerous chemicals in a, potentially fruitless attempt to cover the chemical universe, but by composing a panel of chemicals representing relevant chemical structures and expected toxicodynamic and toxicokinetic characteristics of the chemicals in the assay. Concentration-responses and kinetics should also be considered and it should be elucidated whether the method is able to predict adverse effects *in vivo*?

Still, the current evolutionary hazard-based framework is perceived as a pragmatic approach that works at least to some extent. However, artefacts are introduced: we depend on NOAELs, EC values' and uncertainty factors that are rarely based on data to assess safety. For the revolutionary framework, a change of mind set is needed. Also, validation needs a different way of thinking. We need to move away from using animals as a gold standard, or from using gold standards at all. The revolutionary approach may address different adverse effects in various organs, but may shift the focus from organs to metabolic pathways. Research should focus on building IATA's, DAs, toxicity pathways, AOPs, and, finally, AOP networks. An evidence-based system should be used to build AOPs, with focus on the right keywords. Key-event relationships should be reported as building blocks for AOPs in order to facilitate development of test methods addressing the relevant key events.

In addition to new models and technologies, *in silico* approaches such as machine learning, neuronal networks and artificial intelligence may be applied for safety assessment. To take advantage of these new scientific developments, flexibility in the regulatory system is needed. On the other hand, when we move towards *in silico*, we should take care to correlate *in silico* data with experimental or clinical *in vitro* or *in vivo* data. To increase this understanding, we could work on registries with *in vivo* and *in vitro* data/studies. To be able to get confidence in the how in silico data, toxicologists should better understand how data is produced.

The current regulatory arena relies on standards without flexibility. In the case of the OECD test guideline for DAs for skin sensitization, we replace one standard with another standard based on testing strategies instead of individual assays. This might not be the way to go in the long run. The breakout group discussed how to be more flexible. It was noted that the specific performance of each individual assay in the AOP, pathway or IATA may become less important; confidence can be increased by integration of data from several methods (tiered approach, QSARs, high throughput models, more complex models) that elucidate the same toxicity pathway from various perspectives.

In a revolutionary framework, validation should be regarded as a review process. It should be knowledge-based, flexible and iterative. It was suggested to review the credibility of the methods by an independent, multi-disciplinary body involving regulators, clinicians, toxicologists as well as experts from the field of basic or biomedical sciences. Case studies may be a platform for discussions between different stakeholders (method developers, regulators, risk assessors and end-users) as a starting point to change the safety assessment paradigm including the process of validation.

3.3 Breakout group 3: Innovating validation: are case studies the answer?

(Chantra Eskes (SeCAM), Janine Ezendam (RIVM))

Case studies have been suggested, during the previous BfR-RIVM workshop, to have the potential to serve as the new validation or 'quality assurance' tool in order to accelerate the validation process of alternative methods and testing strategies (Piersma et al., 2018a). The breakout group addressed the potential advantages, challenges, uses and considerations of case studies for innovating the validation of novel 3R approaches in regulatory toxicology.

It was noted that different interpretations might exist on what is meant by a case study. For example, case studies may be used as proof of concepts to demonstrate confidence in novel approaches as within the Integrated European Flagship Programme, Driving Mechanism-based Toxicity Testing and Risk Assessment for the 21st century (EU-ToxRisk (Daneshian et al., 2016)). In this project a broad spectrum of case studies are used to test the applicability of new strategies based on human cell responses and on a comprehensive mechanistic understanding of cause-consequence relationships of chemical adverse effects (www.eu-toxrisk.eu). Case studies may also be used to assess the applicability of a novel approach for a certain regulatory purpose such as for example within the ongoing work from the OECD on the use of IATA (http://www.oecd.org/chemicalsafety/risk-assessment/iata-integrated-approaches-to-testing-and-assessment.htm), including the OECD work on establishing DAs for skin sensitization assessment (OECD, 2016b). In this case, performing case studies allows increasing experience with combined methodologies within the regulatory context, and to create a common understanding as well as the generation of considerations regarding novel approaches. Finally, case studies may be used for specific qualitative and semi-quantitative assessment such as the US-EPA case studies on endocrine disruptors using high-throughput assays combined with computational tools (www.epa.gov/endocrine-disruption/use-high-throughput-assays-and-computational-tools-endocrine-disruptor).

Case studies may therefore be useful to test and assess the practical applicability of novel approaches, as well as to demonstrate confidence in novel approaches. These approaches may be based on AOPs and AOP networks. Furthermore, they can provide insight in the predictivity of a novel approach and may be used for quantitative versus qualitative assessment as well as for risk versus hazard assessment. It is important to clearly define the purpose(s) of the case study, as this determines the design of the case study as well as its evaluation.

A number of challenges were identified when case studies are being used for validation purposes. In the different ongoing projects, case studies are often used to get experience with a novel approach using a (limited) set of reference chemicals. Assessment of the reproducibility is in most case studies not the primary aim, but it is an important aspect for the ‘classical’ validation studies. It was therefore considered important to define at which time point, how and to which extent reproducibility of the new approach described in the case studies needs to be assessed. It was agreed that case studies have more value earlier in the process of development and evaluation of novel approaches, and it was questioned whether case studies are fit for the purpose to validate the reproducibility of such an approach. Case studies in contrast, can help to present the complexity of combined mechanistic information from novel methods in an understandable way for e.g., regulators and end users. In that

sense, case studies can help to get different stakeholders familiarized with approaches they are not acquainted to. It was considered important that case studies are developed and evaluated in a multi-stakeholder group consisting of regulators, risk assessors and end-users.

In designing a case study, the following factors were considered important to be taken into account. First, mechanistic know-how should drive the design of the case study as well as the prioritization of information sources used. Second, there is a need to take into account physiologically based pharmacokinetic (PBPK) modelling when selecting chemicals. The inclusion of information on kinetics of the selected chemicals, e.g. metabolism and clearance was considered important. Third, the technical, chemical and biological applicability domain of the information sources composing the approach need to be considered as well in the design of the case studies. Regarding the selection of chemicals for a case study, it was considered more relevant to select the number and types of chemicals fit for purpose for a certain case study rather than defining a set of minimum number of chemicals. Chemicals included should preferentially cover a range of different potencies to enable a quantitative assessment. Furthermore, inclusion of positive and negative chemicals known to reach the target organ is important to consider as well. The existing know-how on chemistry that is available within the different industry sectors (plant protection products, industrial chemicals, cosmetic ingredients, pharmaceuticals, etc.) was deemed important to take into account when selecting chemicals. Finally, the selected chemicals should allow for a sufficient mechanistic coverage of the case study.

Case studies can be used to build confidence and trust in novel technologies and approaches. Systematic reviews are important to get confidence in the importance of key events and key event relationships in triggering a certain adverse outcome. In an innovative approach for risk assessment, apical endpoints will not be assessed anymore. Therefore, it is important to understand the so-called 'point of no return' in the pathway and individual mechanistic assays, i.e., the point after which an effect is no longer expected to reverse but to lead to an adverse outcome. This information is required to understand if a chemical dose induces an adverse or an adaptive effect in a mechanistic assay. Finally, a description of uncertainties associated with the different elements of a case study is also considered necessary to build confidence on the respective case study.

Performing case studies is relatively new in the field of regulatory toxicology, so that a practical approach is warranted in order to get more experience with them. Case studies can be used to go through different tiers as suggested in different frameworks for innovative safety assessment. The tiers go from simple to complex, depending on the final regulatory need to be addressed, and may start for instance with *in silico* tools and high-throughput *in vitro* assays, followed by moderate-throughput assays and more complex quantitative models that can be used to establish a point of departure for safety assessment in the final tier. Case studies allow evaluating whether all tiers within an approach are useful. Furthermore, by running the different tiers in a case study, the framework may be modified if needed. Case studies can also help in defining the decision criteria needed to proceed to a next tier within the framework. Such an exercise may also be helpful in defining the mechanistic, chemical and technical applicability domain of the information sources used.

Overall, case studies were considered to be of value in the development of a new approach, in demonstrating how safety assessment is conducted with a new approach and in helping to underpin if the predictions obtained are sufficient to reach a conclusion on the safety of a chemical or whether additional testing is needed. Regarding the validation of novel approaches, case studies were considered to be of value mainly for the soft aspects of validation, e.g. by providing trust, experience and confidence. Other advantages of case studies that were identified are that:

- They can help with the iterative development of novel (r)evolutionary approaches;
- They allow demonstrating how novel approaches are built and work;
- They allow combining different assays / information sources in a framework that may ultimately revolutionize chemical safety assessment;
- They allow demonstrating new ways of testing (new components, test strategies);
- They allow learning by doing and developing a common language;
- They allow to assess how well a novel approach works in different contexts;
- They can inspire confidence by allowing to get familiar with the novel approaches and facilitating their acceptance and use by different stakeholders.

Outlook

The validation process described in the OECD guidance document 34 (OECD, 2005) had been established to provide the necessary confidence in a test method to ensure regulatory acceptance. In recent years, it has become clear that although validation of single test methods in particular in respect to transferability and reproducibility is essential, the determination of relevance remains a difficult and often controversial issue that should rather be addressed in the context of integrated, defined approaches comprising several test methods that complement and support each other in order to provide sufficient data for regulatory decision making.

Empirical validation based on predictive capacity is often troubled by a limited number of sufficiently characterized and relevant reference chemicals as well as feasibility. These problems could be tackled in a revolutionary approach embracing a knowledge-based validation that focuses on coherence of data and physiological relevance. These new approaches attempt to explain responses of the whole (human) organism by understanding processes occurring at lower levels of biological organization and building testing strategies around these processes. To be able to define human relevance, comprehensive descriptions of human biologically and physiology as well as toxicological perturbations are needed. This can only be achieved by developing ontologies and (quantitative) AOP networks that will guide the development of mechanism-based methods and approaches. As such, key events that are indicative in a quantitative or at least qualitative way for the adverse outcome need to be identified. Dose-response data need to be generated in order to assess the 'tipping point' of the key events, enabling the distinction between adaptive and adverse effects. It is of utmost importance to discuss, in particular, with regulators appropriate relevant key events and assays to be taken into account upfront to ensure that the approach can and will be applied in regulatory risk assessment.

In the context of integrative approaches the definition of the applicability domain is another important issue. The applicability domain is meant to provide context to the prediction. For QSARs the similarity to the training set determines the level of uncertainty for the analysis of a new substance. For in vitro assays, the applicability domain is generally defined by the chemical groups tested during the validation process as well as by technical limitations. Since alternative approaches are being developed, it has always been a matter of debate how a new method can be applied to chemicals (or even mixtures) that have not been tested during validation. In general it is not feasible to cover the whole chemical universe during validation and, in addition, applicability domains have never been defined for the currently standard in vivo tests. In sum, it seems more important to define limitations of an assay or chemical groups that can be tested with a new approach.

Following an integrative, (r)evolutionary approach, an applicability domain might rather be defined via a biological domain and technical applicability that describes the technical limitations of the assay. Technical validation in respect of the definition of a suitable SOP, transferability and reproducibility is challenging. However, it was agreed that clearly defined and described technical limitations are in any case of key importance to allow acceptance. For novel in vitro assays, it was also agreed that it is not about in- or exclusion of chemicals but about biological relevance i.e. whether and to what extent human relevant mechanisms of toxicity are covered.

The regulatory use of (r)evolutionary approaches might be context and region specific and therefore these discussions need to be placed in an international multi-stakeholder forum, for example the OECD EHS program. Similar to scientific acceptance, regulatory acceptance is built on confidence. Here, case studies can play an important role to foster understanding of regulators and users. Case studies can be quite diverse in the combination different information sources (e.g. *in silico*, phys.-chem. properties, *in vitro*) and the application in various chemistries evaluated in different sectors (PPP, BC, REACH) or regions. There are different types of case studies, e.g. proof of concept studies for new approach methodologies and risk assessment case studies, which means that the place along the line of method development, validation and regulatory acceptance depends on the type of case study. They can be of particular value during the development phase of a new approach and support the iterative development of novel both evolutionary and revolutionary approaches, since they allow learning by doing, developing a common language and built confidence. It can still be quite challenging, however, to use new approaches that people are not familiar with and to present the complexity of combined mechanistic information in an understandable way, e.g. for regulators and end users, and case studies might also be particularly important to enhance communication between stakeholders.

In this workshop communication has been a central point of discussion. Especially in the revolutionary approach we are moving from a fixed to a more flexible mechanism-based, iterative approach. As a result, the definition of data requirements, validation processes and the interpretation of data are not as defined anymore as for the evolutionary approach based on predictivity. This asks for clear and comprehensive information exchange between test method developers, end-users, risk assessors and regulators. Frequent bilateral communication between regulators from all relevant sectors and developers can help to facilitate mutual understanding of new approaches and procedures for data interpretation as well as the use of the data for regulatory risk assessment. Thus, multi-stakeholder involvement is a pre-requisite for the design of new approaches and in particular for the establishment of a knowledge-based, revolutionary approach in risk assessment. It is important to define a common language for all different stakeholders and across the different regulations to prevent repetitive discussions on single toxicological endpoint and regulatory areas. On the other hand it has to be emphasized that also the shortcomings of the current systems are not fully understood which means that uncertainties of the current system based *in vivo* methods have to be described and communicated in order to be able to assess uncertainties associated with new methods and approaches. These discussions will help to identify areas where current in vivo testing is difficult to translate in regulatory relevant adverse effects in humans. These areas will also the ones where new, innovative and even revolutionary approaches could be most easily implemented.

Acknowledgements

The authors would like to acknowledge the contributions of the remaining workshop participants: Joao Barroso (JRC) Nikki Gellatly (NC3R), Betty Hakkert (RIVM) Emily McVey (Ctgb), Jeanette Paulussen (Charles River), Kristie Sullivan (Physicians Committee for Responsible Medicine), Bob van de Water (Leiden University), Jacqueline van Engelen (RIVM), Carl Westmoreland (Unilever)

References

Baker, N., Boobis, A., Burgoon, L., Carney, E., Currie, R., Fritsche, E., Knudsen, T., Laffont, M., Piersma, A.H., Poole, A., Schneider, S., Daston, G., 2018. Building a developmental toxicity ontology. Birth Defects Res 110, 502-518.

Browne, P., Kleinstreuer, N.C., Ceger, P., Deisenroth, C., Baker, N., Markey, K., Thomas, R.S., Judson, R.J., Casey, W., 2018. Development of a curated Hershberger database. Reprod Toxicol 81, 259-271.

Burden, N., Aschberger, K., Chaudhry, Q., Clift, M.J.D., Fowler, P., Johnston, H., Landsiedel, R., Rowland, J., Stone, V., Doak, S.H., 2017. Aligning nanotoxicology with the 3Rs: What is needed to realise the short, medium and long-term opportunities? Regul Toxicol Pharmacol 91, 257-266.

Casati, S., Aschberger, K., Barroso, J., Casey, W., Delgado, I., Kim, T.S., Kleinstreuer, N., Kojima, H., Lee, J.K., Lowit, A., Park, H.K., Regimbald-Krnel, M.J., Strickland, J., Whelan, M., Yang, Y., Zuang, V., 2018. Standardisation of defined approaches for skin sensitisation testing to support regulatory use and international adoption: position of the International Cooperation on Alternative Test Methods. Arch Toxicol 92, 611-617.

Choksi, N.Y., Truax, J., Layton, A., Matheson, J., Mattie, D., Varney, T., Tao, J., Yozzo, K., McDougal, A.J., Merrill, J., Lowther, D., Barroso, J., Linke, B., Casey, W., Allen, D., 2018. United States regulatory requirements for skin and eye irritation testing. Cutan Ocul Toxicol, 1-58.

Conolly, R.B., Ankley, G.T., Cheng, W., Mayo, M.L., Miller, D.H., Perkins, E.J., Villeneuve, D.L., Watanabe, K.H., 2017. Quantitative Adverse Outcome Pathways and Their Application to Predictive Toxicology. Environ Sci Technol 51, 4661-4672.

Daneshian, M., Kamp, H., Hengstler, J., Leist, M., van de Water, B., 2016. Highlight report: Launch of a large integrated European in vitro toxicology project: EU-ToxRisk. Archives of Toxicology 90, 1021-1024.

Daniel, A.B., Strickland, J., Allen, D., Casati, S., Zuang, V., Barroso, J., Whelan, M., Regimbald-Krnel, M.J., Kojima, H., Nishikawa, A., Park, H.K., Lee, J.K., Kim, T.S., Delgado, I., Rios, L., Yang, Y., Wang, G., Kleinstreuer, N., 2018. International regulatory requirements for skin sensitization testing. Regul Toxicol Pharmacol 95, 52-65.

Darwin, C., 1895. On the origin of species by means of natural selection or the preservation of favoured races in the struggle for live., 1st edition ed. John Murray, Albermale Street, London.

Davis, M., Boekelheide, K., Boverhof, D.R., Eichenbaum, G., Hartung, T., Holsapple, M.P., Jones, T.W., Richard, A.M., Watkins, P.B., 2013. The new revolution in toxicology: The good, the bad, and the ugly. Annals of the New York Academy of Sciences 1278, 11-24.

den Broeder, M.J., Moester, M.J.B., Kamstra, J.H., Cenijn, P.H., Davidoiu, V., Kamminga, L.M., Ariese, F., de Boer, J.F., Legler, J., 2017. Altered Adipogenesis in Zebrafish Larvae Following High Fat Diet and Chemical Exposure Is Visualised by Stimulated Raman Scattering Microscopy. Int J Mol Sci 18.

European, C., 2006. 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). EUR-Lex Document 02006R1907-20140410.

European, C., 2008. Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures. EUR-Lex.

Filardo, E.J., Quinn, J.A., Bland, K.I., Frackelton, A.R., Jr., 2000. Estrogen-induced activation of Erk-1 and Erk-2 requires the G protein-coupled receptor homolog, GPR30, and occurs via trans-activation of the epidermal growth factor receptor through release of HB-EGF. Mol Endocrinol 14, 1649-1660.

Foley, B., Clewell, R., Deisenroth, C., 2015. Development of a Human Adipose-Derived Stem Cell Model for Characterization of Chemical Modulation of Adipogenesis. Applied In Vitro Toxicology 1, 66-78.

Foley, B., Doheny, D.L., Black, M.B., Pendse, S.N., Wetmore, B.A., Clewell, R.A., Andersen, M.E., Deisenroth, C., 2017. Editor's Highlight: Screening ToxCast Prioritized Chemicals for PPARG Function in a Human Adipose-Derived Stem Cell Model of Adipogenesis. Toxicol Sci 155, 85-100.

Fragki, S., Piersma, A.H., Rorije, E., Zeilmaker, M.J., 2017. In vitro to in vivo extrapolation of effective dosimetry in developmental toxicity testing: Application of a generic PBK modelling approach. Toxicol Appl Pharmacol 332, 109-120.

Gabbert, S., Maria, L., Robert, L., 2017. A Critical Review of Adverse Outcome Pathway-Based Concepts and Tools for Integrating Information from Nonanimal Testing Methods: The Case of Skin Sensitization. Applied In Vitro Toxicology 3, 250-264.

Haeckel, E., 1878. Anthropogenie oder Entwicklungsgeschichte des Menschen. Verlag Wilhelm Engelmann, Leipzig.

Hartman, J.K., Beames, T., Parks, B., Doheny, D., Song, G., Efremenko, A., Yoon, M., Foley, B., Deisenroth, C., McMullen, P.D., Clewell, R.A., 2018. An in vitro approach for prioritization and evaluation of chemical effects on glucocorticoid receptor mediated adipogenesis. Toxicol Appl Pharmacol 355, 112-126.

Hartung, T., 2008. Toward a new toxicology - evolution or revolution? Altern Lab Anim 36, 635-639.

Hartung, T., Blaauboer, B., Leist, M., 2009. Food for thought ... on education in alternative methods in toxicology. ALTEX 26, 255-263.

Hessel, E.V.S., Staal, Y.C.M., Piersma, A.H., 2018. Design and validation of an ontology-driven animal-free testing strategy for developmental neurotoxicity testing. Toxicol Appl Pharmacol 354, 136-152.

Hoffmann, S., Kleinstreuer, N., Alepee, N., Allen, D., Api, A.M., Ashikaga, T., Clouet, E., Cluzel, M., Desprez, B., Gellatly, N., Goebel, C., Kern, P.S., Klaric, M., Kuhnl, J., Lalko, J.F., Martinozzi-Teissier, S., Mewes, K., Miyazawa, M., Parakhia, R., van Vliet, E., Zang, Q., Petersohn, D., 2018. Non-animal methods to predict skin sensitization (I): the Cosmetics Europe database(). Crit Rev Toxicol 48, 344-358.

Kleinstreuer, N., Dix, D., Rountree, M., Baker, N., Sipes, N., Reif, D., Spencer, R., Knudsen, T., 2013. A computational model predicting disruption of blood vessel development. PLoS Comput Biol 9, e1002996.

Kleinstreuer, N.C., Ceger, P.C., Allen, D.G., Strickland, J., Chang, X., Hamm, J.T., Casey, W.M., 2016. A Curated Database of Rodent Uterotrophic Bioactivity. Environ Health Perspect 124, 556-562.

Koutsoukas, A., Monaghan, K.J., Li, X., Huan, J., 2017. Deep-learning: investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data. J Cheminform 9, 42.

Krause, S., Goss, K.U., 2018. In Vitro- in Vivo Extrapolation of Hepatic Metabolism for Different Scenarios - a Toolbox. Chem Res Toxicol 31, 1195-1202.

Landsiedel, R., Ma-Hock, L., Wiench, K., Wohlleben, W., Sauer, U.G., 2017. Safety assessment of nanomaterials using an advanced decision-making framework, the DF4nanoGrouping. J Nanopart Res 19, 171.

Leist, M., Hartung, T., Nicotera, P., 2008. The dawning of a new age of toxicology. ALTEX 25, 103-114.

Lotti, M., Nicotera, P., 2002. Toxicology: a risky business. Nature 416, 481.

Luechtefeld, T., Marsh, D., Rowlands, C., Hartung, T., 2018. Machine Learning of Toxicological Big Data Enables Read-Across Structure Activity Relationships (RASAR) Outperforming Animal Test Reproducibility. Toxicol Sci 165, 198-212.

Lyssimachou, A., Santos, J.G., Andre, A., Soares, J., Lima, D., Guimaraes, L., Almeida, C.M., Teixeira, C., Castro, L.F., Santos, M.M., 2015. The Mammalian "Obesogen" Tributyltin Targets Hepatic Triglyceride Accumulation and the Transcriptional Regulation of Lipid Metabolism in the Liver and Brain of Zebrafish. PLoS One 10, e0143911.

MacGregor, J.T., 2003. The future of regulatory toxicology: impact of the biotechnology revolution. Toxicol Sci 75, 236-248.

McMullen, P.D., Andersen, M.E., Cholewa, B., Clewell, H.J., 3rd, Dunnick, K.M., Hartman, J.K., Mansouri, K., Minto, M.S., Nicolas, C.I., Phillips, M.B., Slattery, S., Yoon, M., Clewell, R.A.,

2018. Evaluating opportunities for advancing the use of alternative methods in risk assessment through the development of fit-for-purpose in vitro assays. Toxicol In Vitro 48, 310-317.

Miller, M.M., Alyea, R.A., LeSommer, C., Doheny, D.L., Rowley, S.M., Childs, K.M., Balbuena, P., Ross, S.M., Dong, J., Sun, B., Andersen, M.A., Clewell, R.A., 2016. Editor's Highlight: Development of an In vitro Assay Measuring Uterine-Specific Estrogenic Responses for Use in Chemical Safety Assessment. Toxicol Sci 154, 162-173.

Miller, M.M., McMullen, P.D., Andersen, M.E., Clewell, R.A., 2017. Multiple receptors shape the estrogen response pathway and are critical considerations for the future of in vitro-based risk assessment efforts. Crit Rev Toxicol 47, 564-580.

Monosson, E., 2005. Chemical mixtures: considering the evolution of toxicology and chemical assessment. Environ Health Perspect 113, 383-390.

Moseti, D., Regassa, A., Kim, W.K., 2016. Molecular Regulation of Adipogenesis and Potential Anti-Adipogenic Bioactive Molecules. Int J Mol Sci 17.

Myatt, G.J., Ahlberg, E., Akahori, Y., Allen, D., Amberg, A., Anger, L.T., Aptula, A., Auerbach, S., Beilke, L., Bellion, P., Benigni, R., Bercu, J., Booth, E.D., Bower, D., Brigo, A., Burden, N., Cammerer, Z., Cronin, M.T.D., Cross, K.P., Custer, L., Dettwiler, M., Dobo, K., Ford, K.A., Fortin, M.C., Gad-McDonald, S.E., Gellatly, N., Gervais, V., Glover, K.P., Glowienke, S., Van Gompel, J., Gutsell, S., Hardy, B., Harvey, J.S., Hillegass, J., Honma, M., Hsieh, J.H., Hsu, C.W., Hughes, K., Johnson, C., Jolly, R., Jones, D., Kemper, R., Kenyon, M.O., Kim, M.T., Kruhlak, N.L., Kulkarni, S.A., Kummerer, K., Leavitt, P., Majer, B., Masten, S., Miller, S., Moser, J., Mumtaz, M., Muster, W., Neilson, L., Oprea, T.I., Patlewicz, G., Paulino, A., Lo Piparo, E., Powley, M., Quigley, D.P., Reddy, M.V., Richarz, A.N., Ruiz, P., Schilter, B., Serafimova, R., Simpson, W., Stavitskaya, L., Stidl, R., Suarez-Rodriguez, D., Szabo, D.T., Teasdale, A., Trejo-Martin, A., Valentin, J.P., Vuorinen, A., Wall, B.A., Watts, P., White, A.T., Wichard, J., Witt, K.L., Woolley, A., Woolley, D., Zwickl, C., Hasselgren, C., 2018. In silico toxicology protocols. Regul Toxicol Pharmacol 96, 1-17.

OECD, 2005. Guidance Document on the validation and international acceptance of new or updated test methods for hazard assessment. OECD Series on Testing and Assessment No. 34.

OECD, 2016a. GUIDANCE DOCUMENT FOR THE USE OF ADVERSE OUTCOME PATHWAYS IN DEVELOPING INTEGRATED APPROACHES TO TESTING AND ASSESSMENT (IATA) OECD Series on Testing & Assessment No. 260

OECD, 2016b. Guidance Document on the Reporting of Defined Approaches and Individual Information Sources to be Used Within Integrated Approaches to Testing and Assessment (IATA) for Skin Sensitisation. Series on Testing and Assessment. OECD Series on Testing and Assessment No. 256.

OECD, 2016c. Guidance Document on the Reporting of Defined Approaches to be Used Within Integrated Approaches to Testing and Assessment. . OECD Series on Testing and Assessment No.255.

Penot, G., Le Peron, C., Merot, Y., Grimaud-Fanouillere, E., Ferriere, F., Boujrad, N., Kah, O., Saligaut, C., Ducouret, B., Metivier, R., Flouriot, G., 2005. The human estrogen receptor-alpha isoform hERalpha46 antagonizes the proliferative influence of hERalpha66 in MCF7 breast cancer cells. Endocrinology 146, 5474-5484.

Piersma, A.H., Burgdorf, T., Louekari, K., Desprez, B., Taalman, R., Landsiedel, R., Barroso, J., Rogiers, V., Eskes, C., Oelgeschlager, M., Whelan, M., Braeuning, A., Vinggaard, A.M., Kienhuis, A., van Benthem, J., Ezendam, J., 2018a. Workshop on acceleration of the validation and regulatory acceptance of alternative methods and implementation of testing strategies. Toxicol In Vitro 50, 62-74.

Piersma, A.H., Hessel, E.V., Staal, Y.C., 2017. Retinoic acid in developmental toxicology: Teratogen, morphogen and biomarker. Reprod Toxicol 72, 53-61.

Piersma, A.H., van Benthem, J., Ezendam, J., Kienhuis, A.S., 2018b. Validation redefined. Toxicol In Vitro 46, 163-165.

Sargis, R.M., Johnson, D.N., Choudhury, R.A., Brady, M.J., 2010. Environmental endocrine disruptors promote adipogenesis in the 3T3-L1 cell line through glucocorticoid receptor activation. Obesity (Silver Spring) 18, 1283-1288.

Sauer, U.G., Hill, E.H., Curren, R.D., Raabe, H.A., Kolle, S.N., Teubner, W., Mehling, A., Landsiedel, R., 2016. Local tolerance testing under REACH: Accepted non-animal methods are not on equal footing with animal tests. Altern Lab Anim 44, 281-299.

Scialli, A.R., Daston, G., Chen, C., Coder, P.S., Euling, S.Y., Foreman, J., Hoberman, A.M., Hui, J., Knudsen, T., Makris, S.L., Morford, L., Piersma, A.H., Stanislaus, D., Thompson, K.E., 2018. Rethinking developmental toxicity testing: Evolution or revolution? Birth Defects Res 110, 840-850.

Staal, Y.C.M., L.A., P.J., V.S., H.E., H., P.A., 2017. Advanced Toxicological Risk Assessment by Implementation of Ontologies Operationalized in Computational Models. Applied In Vitro Toxicology 3, 325-332.

Strickland, J., Clippinger, A.J., Brown, J., Allen, D., Jacobs, A., Matheson, J., Lowit, A., Reinke, E.N., Johnson, M.S., Quinn, M.J., Jr., Mattie, D., Fitzpatrick, S.C., Ahir, S., Kleinstreuer, N., Casey, W., 2018. Status of acute systemic toxicity testing requirements and data uses by U.S. regulatory agencies. Regul Toxicol Pharmacol 94, 183-196.

Teubner, W., Landsiedel, R., 2015. Read-across for hazard assessment: the ugly duckling is growing up. Altern Lab Anim 43, P67-71.

Tonk, E.C., Pennings, J.L., Piersma, A.H., 2015. An adverse outcome pathway framework for neural tube and axial defects mediated by modulation of retinoic acid homeostasis. Reprod Toxicol 55, 104-113.

Um, J., Jung, D.W., Williams, D.R., 2018. The future is now: cutting edge science and understanding toxicology. Cell Biol Toxicol 34, 79-85.

Wang, Z., Zhang, X., Shen, P., Loggie, B.W., Chang, Y., Deuel, T.F., 2006. A variant of estrogen receptor-{alpha}, hER-{alpha}36: transduction of estrogen- and antiestrogen-dependent membrane-initiated mitogenic signaling. Proc Natl Acad Sci U S A 103, 9063-9068.

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., Andersen, M.E., 2015. Incorporating High-Throughput Exposure Predictions With Dosimetry-Adjusted In Vitro Bioactivity to Inform Chemical Toxicity Testing. Toxicol Sci 148, 121-136.

Wu, Y., Wang, G., 2018. Machine Learning Based Toxicity Prediction: From Chemical Structural Description to Transcriptome Analysis. Int J Mol Sci 19.

Zurlo, J., Deborah Rudacille, and Alan M. Goldberg, 1994. Animals and alternatives in testing: History, science, and ethics. . Mary Ann Liebert.