Reproductive Toxicology 89 (2019) 124–129

Contents lists available at ScienceDirect

Reproductive Toxicology

journal homepage: www.elsevier.com/locate/reprotox

Update of the DevTox data database for harmonized risk assessment and alternative methodologies in developmental toxicology: Report of the 9th Berlin Workshop on Developmental Toxicity

ARTICLE INFO

Keywords:
DevTox update
Categorised anomalies
Developmental toxicology
Alternative methods
Harmonized terminology
Risk assessment

ABSTRACT

Representatives of applied science (e.g. governmental organizations, academia, and industry) met to discuss the progress towards a harmonized human health risk assessment in developmental toxicology of plant protection products, biocidal products, and other environmental chemicals at the 9th Berlin Workshop on Developmental Toxicity held in September 2018. Within the focus of the scientific discussion were the future of *in-vitro* methods for developmental and reproductive toxicology, the potential relevance of alternative species in testing of developmental effects, and risk and hazard assessment of developmental and endocrine effects.

Furthermore, the need for a harmonized terminology for classification of anomalies in laboratory animals in developmental toxicity studies aiming for human health risk assessment was determined. Here, the DevTox database was identified as an extremely valuable tool. Overall, the participants agreed that still one of the biggest challenges for testing developmental toxicity in the 21st century is the development of animal-free test strategies and alternatives to animal testing that could provide human-relevant information in a rapid, efficient, and mechanistically informative manner.

1. Introduction

Improvement of testing methods, more experimental evidence and accompanying mechanistic data from laboratory animals are required for a better understanding of developmental outcomes by using new methodological approaches for alternative, non-animal testing methods for risk assessment of developmental toxicants in the 21st century [1]. Additionally, there are scientific and regulatory needs for a harmonized terminology for categorisation of anomalies in laboratory animals in developmental toxicity studies conducted for human health risk assessment [2]. The scientific basis for classification of chemicals as developmental toxicants, comparison of developmental anomalies in laboratory animals with human data, exploring possibilities to enhance the scientific dialogue between stakeholders and other experts, the implementation of alternative methods for risk assessment in developmental toxicity testing and the progress towards harmonized approaches for risk assessment were the initial motives to launch the 9th Berlin Workshop in 2018. The workshop was a satellite to the 46th Annual Meeting of the European Teratology Society within the series of Berlin Workshops [1–4]. Results of the first workshop in 1995 to the 8th Berlin Workshop on DevTox Terminology are summarized on the DevTox-Website (version 3.0 – https://www.devtox.org/index_en.php), which now contains more than 2.500 images with examples for external, skeletal, visceral and maternal-fetal anomalies.

The current workshop aimed to intensify discussions among risk assessors, risk managers, industry and scientific representatives of academia and non-academic organizations in order to understand present needs and explore possibilities to support the progress towards a harmonized human health risk assessment of plant protection products, biocidal products and other chemicals. Therefore, the following main topics were discussed: harmonized terminology in an updated DevTox database; hazard and risk assessment of both developmental toxicity and endocrine-related effects; relevance of alternative species in testing of developmental effects and appropriate animal-free testing strategies for the future.

To address appropriately the future of developmental toxicology within the 21st century it is important to understand on the one side the genesis of this field and on the other side how original scientific approaches evolved. Therefore, Diether Neubert, a pioneer within the field of Developmental and Reproductive Toxicology [5], presented some notes on the history of establishing methods in developmental and reproductive toxicology after the thalidomide disaster. The biggest scientific challenge at this time was to establish a guideline for developmental toxicity testing designed to provide general information concerning the effects of a test substance on the developing human organism. The development of appropriate methods has been greatly accelerated by the many worldwide scientific studies that have now attempted to explore the toxic effects and toxicodynamics of thalidomide. At this time it was already obvious that only complex biological systems (animal models) have the possibility to detect corresponding developmental toxic effects. Even if in vitro models were already used at that time, Neubert knew about their strengths but also limitations. Acyclovir another suitable test substance leading to developmental effects was able to shed more light into the discussion how useful in-vitro methods can be for developmental toxicity testing. Neubert considers in-vitro systems as not useful for endpoints such as retardation, impaired parturition or postnatal development. However, for revealing mechanisms, in-vitro tests usually are better suited than in-vivo tests and in-vitro tests can be applied for screening of similar chemical structures of known teratogens. But in case of negative results, adverse effects cannot be excluded and subsequent testing in-vivo is essential. He finally concluded that in regulatory toxicology it is suggested to adhere to the precautionary principle, e.g. on the basis of simpler test systems.

https://doi.org/10.1016/j.reprotox.2019.07.003
Received 24 April 2019
Available online 06 July 2019
0890-6238

2. Terminology in developmental toxicology

2.1. Update of the DevTox Data base Chinese version of the terminology

Following the 8th Berlin DevTox Workshop, the collaboration with Chinese DevTox Project partners was developed and the translation of the DevTox database into Chinese, together with a comprehensive revision of the database, took place [6]. Finally, the Chinese DevTox database was officially launched in 2016 on the DevTox website (devtox.org).

In the context of the improvement of the harmonized terminology in an updated DevTox database, 25 images of rat visceral anomalies and 20 images of mouse visceral anomalies have been uploaded to the DevTox database by Chinese researchers. In total, 212 images of visceral anomalies are currently available. The "Atlas of Common Malformations in Laboratory Rabbit and Rat", with more than 500 original abnormality images, was published in Chinese, Japanese and English languages [7]. These additional images will be integrated into the DevTox website. In a further project, the "Atlas of Developmental Anomalies in Experimental Animals" [8] was published in Japan. The implementation of this atlas on the DevTox website was considered as a possible further improvement of this tool for basic research, laboratory testing and teaching at universities.

2.2. Japanese survey on terminology and atlas of developmental abnormalities

In 2009, "grey zone anomalies" were defined as abnormalities that did not fit readily into the categories of malformation or variation [5]. Currently, 73 external, 579 skeletal, 370 visceral and all 19 maternal-fetal anomalies are categorised as grey zone and 94 external, 271 skeletal, 157 visceral anomalies are categorised as malformations. The remaining 173 skeletal anomalies and 6 visceral anomalies are categorised as variations [9]. The Japanese Teratology Society (JTS) conducted a survey limited to external "grey zone" anomalies. The questionnaire was completed by 20 laboratories, including 12 companies active in pharmaceutical industries, 6 contracting research laboratories, 1 company active in chemical industry and 1 environmental research laboratory. In a first step, re-categorised anomalies that were approved by more than 80% in the survey or recommended by the Terminology Committee were selected. In a second step, intense discussion took place in the JTS annual meetings in 2015, 2016 and 2017. As a result, a possible new categorisation of external findings was suggested by the JTS. It was proposed that, within the category "malformation", distinction should be made between uncommon structural changes (deviations from the normal morphology of the species or strain) and non-structural abnormalities (deviations from normal function without structural change). The JTS Terminology Committee categorised the 73 external grey zone anomalies as follows: 38 findings were categorised as "malformation", 3 findings as "non-structural abnormality" and 1 finding as "malformation" or "non-structural abnormality". The remaining 31 findings were considered to be "not applicable" for fetal examination in rodent/rabbit embryo-fetal development, because observations were indistinguishable from changes in fetal growth (e.g. large or small fetus) or easily changeable under physiological condition during examination (e.g. discolored or pale fetus). In the future, re-categorisation of visceral findings will be performed as well. In order to reduce 'grey zone' categorisation of anomalies in the context of international harmonization, it was proposed that consideration of this survey could also enhance the DevTox database, providing that a more detailed overview of the survey results on teratogenic terms and criteria used by JTS could be undertaken. If, based on the survey results, the Europeans and the Japanese can deepen their mutual understanding regarding the criteria/terminology for anomaly or variation, a re-assessment of the European categorisation could be considered.

2.3. Evaluation of post-natal anomalies and research on mechanisms

According to Chahoud et al. [2], a malformation is defined as a "permanent structural change that is likely to adversely affect the survival or health of the species under investigation". It is noteworthy that the permanence of anomalies can only be evaluated in postnatal studies. Up to today, there is still a lack of data from postnatal evaluations of skeletal anomalies. Therefore, a study was presented in pregnant Wistar rats treated on gestation day 11 with 5-Flouro-2´-deoxyuridine (FUDR), a well-characterized teratogenic agent. Most of the skeletal anomalies observed in this study were considered to be variations since they did not persist up to postnatal day (PND) 21, e.g. unossified or asymmetric ossification of vertebrae, bipartite ossification of vertebrae, hemicentric vertebrae, and misshapen vertebrae in thoracic, lumbar and sacral region. Anomalies persisting up to PND 21 were classified as malformations, e.g. dumbbell-shaped cervical vertebral centrum and supernumerary lumbar centrum. As ossification of cervical vertebrae continues after gestation day (GD) 21, assessment of anomalies in this region is only reliable at a later postnatal time point. Currently, no guideline is available for studies on the postnatal fate of anomalies. It was noted that a guideline is urgently needed for well-founded classification, which may be highly relevant for classifying, or not, a substance as a high-concern developmental toxicant. Furthermore, the appropriate time point, i.e., the age of the animal at evaluation, should be considered carefully.

Possible mechanisms of thoracolumbar supernumerary rib development after birth, using CT scanning at various time points (up to PND 62) after birth in the same animal, were presented. The goal of the project was to investigate the toxicological significance of short thoracolumbar supernumerary ribs (STSR) after birth and to provide more data to discuss the importance of this finding, which is at present a controversial issue. Currently, STSR are categorised as variations. They are observed with relatively high incidence in rodent studies. Historical control data were collected from 24 Japanese laboratories, 15 pharmaceutical and chemical companies and 9 contract research organizations [10] between 2011 and 2015. Contrary to the previous assumption, the STSRs persisted and did not disappear postnatally, if STSRs were induced in rat fetuses by treating dams orally with 5-flucytocine (5-FC) [11].

An image analysis of external and skeletal malformations induced by retinoic acid in SD rats at an optimized dose for rat teratogenicity tests was presented. A dose dependent increase in multiple external and skeletal malformations was observed in all treated groups. As significant resorptions or stillbirths were observed at high doses of 100 and 150 mg/kg bw/d, it was concluded that 50 mg/kg bw/d is an optimized dose for rat teratogenicity tests.

3. Risk / hazard assessment of developmental, including endocrine-related effects

The presentation focused on recent evolutions in the European Union (EU) chemical legislation. Recent changes in regulatory toxicology have been driven by EU regulation for pesticides and biocides in particular and are still subject to a large body of debate.

Regulatory aspects of both developmental toxicology and endocrine effects were discussed within the current regulatory framework of the European legislation. The new legislation has introduced hazard cut-off criteria in the EU Pesticide regulation [12] and the EU Biocide regulation [13] which are based on classification and labelling according to the classification, labelling and packaging (CLP) regulation [14]. The hazard cut-off criteria approach is expected to be faster and more protective, considering especially the fact that existing exposure models may underestimate risks. In addition, this approach is an incentive to the development of safer chemicals. However, opponents consider that such an approach is unscientific, which might needlessly exclude much needed Plant Protection Products (PPPs) from the market, and refute

that risk assessment is less protective than the hazard cut-off criteria approach. The new system raises a number of difficult to resolve issues, in particular, how to define negligible exposure. Therefore, there is an urgent need for a good practical definition or technical guidance to explain the meaning of negligible exposure/risk.

It was summarized that current approaches for classification of developmental toxicity according to the CLP regulation and identification of endocrine disrupters (EDs) according to the commission proposal may be overlapping but there are no harmonized agreements to assess the degree of potency for EDs [15]. Therefore, it was suggested that classification of developmental or reproductive hazards should also be considered if these hazards may be induced by disruption of endocrine pathways. It was concluded that a grading of hazard categories if used in classification/labelling of developmental toxicity and identification of EDs could improve the confidence for appropriate risk assessments and support the options for risk management decisions.

Classification of developmentally toxic pesticides and negligible exposure were critically disputed from the regulatory side with the objective to discuss the term negligible and its implications. It could not be answered, what does negligible in practical considerations mean. Based on several examples from hazard characterization and exposure assessment, it was concluded that regulatory developments can highlight new aspects of hazard identification/characterization. Newly introduced regulatory endpoints in reproductive toxicity studies, if applied and implemented in a robust way, can change classification and/or the no-observed-effect-level (NOAEL), e.g. endocrine-related developmental toxicity. Likewise, a more robust and consistent tiered approach to developmental neurotoxicity (DNT), such as the strategy recommended by the European Food Safety Authority (EFSA) for DNT [16,17], might lead to increase in the number of active substances classified as developmental toxicants and could have a significant impact on the level of safety parameters such as Acceptable Daily Intake or Acute Reference Dose. In addition, new approaches could help to define toxicologically relevant exposures. It was concluded that what has been previously considered negligible may not be negligible anymore, especially in the case of cumulative exposure to substances with similar effects. Although the “negligible exposure approach” may seem to be a reasonable approach, it needs to be handled with care since new insights in risk assessment may influence the definition of negligible exposure. Progress in regulatory science and cumulative effects need to be taken into consideration.

The industry perspective on low dose effects and mixed exposure, regarding classification of developmentally toxic pesticides, was presented: A series of reproductive toxicity studies was conducted, comparing single compounds and several mixed exposures at low doses [18]. Single- and mixed-exposures could be compared from molecular to pathological levels in the same animals. Based on the critical review of several pesticidal active substances, it was concluded that for the antiandrogens tested in this project the NOAELs and Reference Values were protective even when considering potential mixture effects and there was no evidence for an interaction of the compounds at the individual NOAEL or lower doses. Finally, it was noted that it is not possible to derive any general conclusions from this study, since the results apply only to those antiandrogens that were tested.

Following one recommendation of the 8th workshop, the relevance of human data for assessment of developmental effects was again highlighted. Indeed, many human teratogens were identified by clinicians when they observed a small number of patients with birth defects. However, after the thalidomide scandal one has to realize that developmental toxicity in humans has been mostly prevented by preclinical toxicological studies using laboratory animals. Therefore, despite species differences, the value of preclinical studies using laboratory animals cannot be underestimated. However, the extrapolation of reproductive toxicity data in laboratory animals to the human is one of the most difficult parts in developmental toxicology, because of several reasons. First of all, human embryos seem to be several to ten times more sensitive to teratogenicity of exogenous agents than small animal species (compiled from different sources). Thus, it is reasonable to assume that the human embryo is more susceptible to developmental toxicants than the embryo of laboratory animals and there is a need to apply safety factors. Second, embryos are vulnerable to teratogenic agents during the critical period of organogenesis. The length of this critical period is much shorter in small animal species than in humans, suggesting that the human embryo can be exposed to deleterious exogenous agents significantly longer than small animals. Third, science still lacks information about developmental effects in humans of over 2000 chemicals, which have been shown to be embryotoxic/teratogenic up to now only in one or more animal species. This is mostly due to the fact that only application observations of approved and registered drugs are able to provide that information. And embryotoxic/teratogenic effects can hardly be detected in humans, because the reproductive loss rate is extremely high in humans (approximately 10–15% of clinically recognized conceptions end in spontaneous abortions) [19]. In addition, a considerably large proportion of human conceptions seem to end in subclinical abortions. It seems that defective development occurs frequently in early human pregnancy and more than 90% of malformed embryos die in utero [20,21].Spontaneous abortion was considered to be a natural scavenging process that reduces the birth of abnormal babies. This is also a reason why the extrapolation of reproductive toxicity data in laboratory animals to the human is one of the most difficult tasks in developmental toxicology.

4. Alternative species in testing of developmental effects

This session focused on the relevance of alternative species to screen chemicals for potential developmental effects, i.e. on *Caenorhabditis (C.) elegans and* Zebrafish (*Danio rerio*) in developmental toxicity studies, as well as *Biomphalaria (B.) glabrata* snails.

C. elegans is a free living hermaphrodite nematode and represents a long established model system in developmental biology with many advantages, including a short life cycle (approx. 3 days, at 20 °C), a large brood size, and a plethora of available mutant strains and genetic and molecular tools [22]. Due to its small size it also fulfils crucial criteria for high-throughput-screening applications. The nematode possesses reproductive, metabolically active digestive, endocrine, neuromuscular and sensory systems, while it lacks other mammalian organs. However, between *C. elegans* and *homo sapiens*, many genes and signaling pathways are highly conserved [23]. Therefore, especially for substance screening and mechanistic toxicity studies, *C. elegans* might prove valuable. Complex endpoints like e.g. fertility, fecundity, germline function, and developmental toxicity can be assessed fast and subsequently studied on a mechanistic level. *C. elegans* has already been used for toxicity screening, e.g. for the EPA’s ToxCast™ Phase I and Phase II libraries, which contain 292 and 676 chemicals, respectively, by using the endpoints of decreased larval development and growth. A significant overlap in the activity of chemicals in the ToxCast™ libraries between *C. elegans* and Zebrafish developmental screens was found [24]. It was suggested that including *C. elegans* toxicological assays as part of a battery of *in-vitro*, *in-vivo* and *in-silico* assays can provide additional information to predict the potential toxicity of chemicals to humans.

One of the advantages of using Zebrafish in developmental toxicity studies is its easily visible and rapid development from fertilization to larval stages. A full range of Cyp-genes demonstrate a strong evolutionary relationship with mammals, including humans. The fish is a vertebrate with high fecundity, has a small size and is suitable for high-throughput screening with chemicals under investigation being added to the medium/water [25]. As in mammalian developmental toxicity studies, external, visceral and skeletal findings can be differentiated. Known teratogenic substances show a good concordance when tested in mammals and in Zebrafish. Work still needs to be done in the Zebrafish test systems focusing on the distinction between systemic (i.e. related to

general toxicity) and teratogenic effects. Systemic testing of different exposure levels in Zebrafish test systems is part of the Programme in the US ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods) [26]. Thus, the Zebrafish could be a good candidate model organism for developmental toxicity testing.

Developmental, multigeneration reproduction and dominant lethal toxicity tests have been more and more frequently conducted using freshwater snails, such as *B. glabrata*. Like *C. elegans*, the snail is hermaphrodite and self-fertilization takes place. In 2016, the OECD issued an assay for the general reproductive performance of the snail *Lymnaea stagnalis* [27]. Endpoints to be evaluated are survival and the cumulated number of egg-clutches produced per snail during exposure; with individual growth and the number of eggs per clutch per snail as additional endpoints. This assay is primarily intended to assess the potential impact of chemicals on ecosystems and not to screen for developmental toxicity, however some information about the impact of a test chemical on the development of the snails can obtained, e.g. individual growth, production of abnormal eggs. The "Snail DevTox Assay", conducted on *B. glabrata* and developed by the National School of Public Health, Brazil, seems to be a feasible alternative test system for screening chemicals for developmental toxicity [28]. The assay is cheap, quick and easy to perform, it requires only simple laboratory equipment and it is possible to test many substances over a wide concentration range in a relatively short time period. Further research steps are necessary to implement the Snail DevTox Assay in the testing strategy of chemicals and a comparative study with tests on other species is urgently needed. The *B. glabrata* genome includes xenobiotic biotransformation enzyme genes. Snails have a more complex metabolic system than *C. elegans* and are equipped with Cyp-genes comparable to those present in humans; however, the limited metabolic competence is one of the greatest disadvantages.

Finally, it was concluded that a large number of chemicals can be tested quickly and at a wide range of concentrations with the alternative test systems using the worm *C. elegans*, Zebrafish and the snail *B. glabrata*. There is an urgent need to better understand the advantages and limitations of alternative methods before invertebrates and fish embryos are suitable for substitution of mammalian test systems. It remains questionable how robust and effective those systems are. The predictability of non-mammalian species is still limited but also, in mammalian species, exposure to certain substances did not lead to the expected outcome.

5. Methodology for the future

This session focused on the improvement of testing methods for a better understanding of developmental outcomes using new methodological approaches and non-animal alternative testing methods for risk assessment of developmental anomalies in the 21st century.

A presentation on Tox21 summarized its origins in the publication of a National Research Council (NRC) report on "Toxicity Testing in the 21st Century" [29], the formation of a Federal Partnership of four U.S. agencies in 2008 (National Institute of Environmental Health Sciences [NIEHS], National Center for Advancing Translational Science [NCATS], Environmental Protection Agency [EPA], and Food and Drug Administration [FDA]), and its significant collaborative goals and accomplishments over the subsequent 10 years. A Tox21 strategic and operational plan, published in 2018 [30], is focused on developing and deploying alternative test systems that are predictive of human toxicity and dose response; addressing key technical limitations of current high-throughput (HTP) screening systems; consolidating chemical library management; characterizing legacy animal toxicity studies for comparison to HTP screening results; validating HTP assays, integrated assay batteries, computational models, 3-D organ-like model systems, and other emerging Tox21 approaches; and refining HTP methods for characterizing pharmacokinetics to better predict the relationship between target tissue concentrations and external doses of chemicals.

In 2018, the U.S. Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) published a Strategic Roadmap [31]. This document states that Federal agencies and stakeholders will collaborate to "develop, establish confidence in, and encourage the use of new approaches to toxicity testing that improve human health relevance and reduce or eliminate the need for testing in animals." The scope and charge of the ICCVAM Developmental and Reproductive Toxicology (DART) Working Group include identifying agency needs for developmental toxicity assessment, working with international partners to identify global developmental toxicity regulatory testing requirements, identifying endpoints needed by federal agencies, examining the importance of specific endpoints and study types in product development and regulatory decision-making, cataloguing existing and emerging technologies and mapping measured endpoints to known mechanisms of developmental toxicity and their use in fulfilling regulatory testing requirements, and establishing a stakeholder group of government and non-government scientists to coordinate efforts towards developing and implementing integrated strategies for developmental toxicity testing. Several completed and ongoing programs and activities that align with these goals were described. One notable example was the validation for using HTP and computational toxicology data to replace the *in vivo* uterotrophic and Hershberger assays in EPA endocrine disruptor screening and testing requirements. Other efforts include the Integrated Chemical Environment (ICE: https://ice.ntp.niehs.nih.gov), global quantitative structure-activity relationship (QSAR) modelling collaborations, automation of reference data identification through machine learning, mechanistic mapping of HTP assays to bioactivity, the use of pluripotent stem cells to predict developmental toxicity, and the use of *in vitro* or *in silico* models for developmental risk prioritization. A large body of legacy prenatal developmental toxicity studies from the U.S. National Toxicology Program are currently being digitized for use in developing reference datasets, and based on the discussions at the workshop; it was decided to apply the DevTox harmonized nomenclature to the study extractions.

A presentation on developmental neurotoxicity (DNT) evaluation emphasized the need for sensitive, biologically-based and mechanistically relevant *in vitro* assays to replace the current *in vivo* testing paradigm. Consideration has been given to the process for developing and validating assays, and for implementation in a regulatory context [32]. Collaborative global efforts are focused on developing a DNT *in vitro* testing battery that addresses neurodevelopmental processes and timing. *in vitro* testing for DNT is based on the core processes of brain development: proliferation, apoptosis, differentiation/migration, growth/synaptogenesis, myelination, and functional network formation [33]. Examples include assays that assess neural progenitor cell (NPC) proliferation; migration of neural crest cells (NCC), neuroepithelial cells (NEP), or radial glial cells during brain development; and oligodendrocyte differentiation. Assays that target synaptogenesis and neural network formation use primary rat cortical cultures or human induced pluripotent stem cell-derived systems. DNT *in vitro* assays have contributed critical information to the development of several Adverse Outcome Pathways (AOPs - an organizational framework that facilitates application of alternative methods for regulatory application) that are associated with DNT outcomes and that are published on the OECD AOP Knowledge Base (https://aopkb.oecd.org/).

The German Centre for the Protection of Laboratory Animals (Bf3R) at the BfR presented a novel model used to detect skeletal anomalies and the application of improved embryonic bodies for studying development, embryotoxicity, and placental function *in-vitro*.

In the late 1950s and early 1960s, the Contergan scandal initiated a broad public discussion about teratogenic substances such as thalidomide and their approval and use as pharmacological drugs. The scientific discussion about Thalidomide's mode of action also initiated a debate about the predictive value of animal testing. One important aspect concerns species-specific differences in physiology and

metabolism. For example, the teratogenicity of Thalidomide was not observed in rats or mice. Only experiments in rabbits hinted on skeletal abnormalities in the offspring after maternal exposure during pregnancy. Even the use of animals that are phylogenetically closer to humans (such as non-human primates) does not guarantee the detection of all substances that exhibit teratogenic hazards for humans. Animal testing in the context of teratogenicity is thus limited regarding sensitivity and specificity and does therefore not allow for full translation of results to humans. Further, animal testing is connected with high cost and time investments. In light of these arguments, alternatives to animal testing are discussed that are 1) more predictive for the human system 2) cost and time effective [34].

When thinking about the implementation of such alternative methods, it needs to be considered that teratogenic effects might be induced directly or indirectly. A direct effect can be defined as the specific alteration of organ or tissue development, such as the inhibition of bone matrix mineralization through tetracycline that is incorporated instead of calcium. An indirect effect is present in case the teratogenic effect and the connected phenotype are the result of secondary mechanisms initiated by the substance in question. For example, the inhibition of CYP26 by azoles leads to a secondary increase of retinoic acid that disturbs the spatial anterior/posterior patterning during embryogenesis through alterations in the expression profile of Hox genes and others [35,36].

While direct effects can be detected in suitable *in vitro* experiments, indirect effects might be extremely difficult or even impossible to predict without using a whole organism. The complexity of embryogenesis might thus demand the ongoing use of animal testing for the detection of teratogenic substances. Yet, the supplementation of classical animal testing with advanced *in vitro* testing methods might help to address known limitations and increase the predictive value of animal testing. Here, organ-on-a-chip technology is a promising approach to faithfully recreate a given organ or tissue [37]. A well characterized and physiologic relevant bone-on-a-chip system can help to identify substances that directly affect the process of bone formation. In this regard, we develop a bone-on-a-chip system that allows for the determination and control of oxygen levels and mechanical loading. Using human cells, the system will have a higher predictive value for substances that directly alter bone formation during human embryogenesis. A bone-on-a-chip has therefore high potential to replace animal experiments that aim to elucidate direct effects of a given substance on *de novo* bone formation [38].

The second innovative approach presented by the Bf3R is the generation of synthetic embryos. So far, animal-based embryotoxicity testing is a central part of identifying and characterizing potential hazards of chemicals (prenatal development toxicity study, OECD TG 414), but it comes with ethical concerns. Over the past decades, several in vitro alternative approaches for reducing, refining and replacing animal testing were developed and some of them, such as the Embryonic Stem Cell Test (EST), the limb bud micromass test, and the rat whole embryo culture assay have already been validated by the European Centre for the Validation of Alternative Methods (EVCAM). However, existing models like the EST have several drawbacks as experimental tool since they show intrinsic limitations (for instance, lack of extra-embryonic tissues). By combining three cell types as surrogates for the corresponding component tissues of the early mouse blastocyst - embryonic stem cells, trophoblast stem cells and extra-embryonic endoderm cells – in three-dimensional aggregates, these synthetic embryos generate structures whose morphogenesis is similar to that of natural embryos with all embryonic and extra-embryonic compartments [39]. If the methodologies to generate such improved stem cell-based models are further optimized and verified in their suitability for high throughput screening they may complement animal models with better reproducibility and predictability within the future. Furthermore, synthetic embryos could serve as a tool in basic research, the area in which most laboratory animals are employed. Hence, cell models of mammalian embryogenesis are needed to identify signaling pathways, cell-cell communication, and further molecular processes during the peri-implantational period.

6. Conclusions

- If the classical regulatory developmental *in-vivo* toxicity studies for legally required risk assessment of chemical substances should be replaced in the future, more research and funding is considered necessary.
- Postnatal evaluation of anomalies, which cannot be categorised in the rat at GD20/21, should be regularly employed. These evaluations should be part of other regulatory studies, e.g. the extended-one-generation reproductive toxicity study, so that no additional animals are necessary.
- Studies with alternative species are useful for screening purposes or for additional clarification of selected questions. For risk assessment, tests with these species are not sufficiently reliable to replace the classical developmental toxicity studies with laboratory mammals.
- International cooperation is highly important for further harmonization of evaluation and categorisation of anomalies. In this sense, it was considered valuable to organize a meeting to discuss the issues with representative scientists from the Berlin workshop, the Japanese Teratology Society (JTS) and the Food Safety Commission of Japan (FSCJ) before the next workshop in Berlin.
- The investigation of comparative aspects of reproductive toxicology was again considered as an important factor to extend the knowledge on the relationship between developmental anomalies, prenatal effects and postnatal consequences in humans and laboratory animals as a template for human health risk assessment.

Disclaimer

This manuscript has been reviewed by the above mentioned institutions and approved for publication. Approval does not signify that the contents reflect the views of the institutions, nor does mention of trade names or commercial products constitute endorsement or recommendation for use. The findings and conclusions in this report have not been formally disseminated by the U.S. EPA and should not be construed to represent any agency determination or policy.

Acknowledgments

Special thanks Diether Neubert for presenting the key note lecture via skype, to Andrea Gall and Martina Rauch and the other co-rapporteurs for composing the publication and to Ruth Clark for the final editing of the text. The authors would also like to thank the BfR Academy, especially Frederic Müller and Anne- Katrin Hermann for their excellent technical assistance in the organization of the Workshop as well as to the Council of the European Teratology Society (ETS) for cooperation in organizing the 9th Berlin Workshop on Developmental Toxicity as a satellite to the 46th Annual Meeting of the ETS.

Presentations from the 9th Berlin Workshop on Developmental Toxicology as a Satellite event to the 46th Annual Meeting of the European Teratology Society are available on the *DevTox* website (www.devtox.org).

References

[1] R. Solecki, M. Rauch, A. Gall, J. Buschmann, R. Clark, A. Fuchs, et al., Continuing harmonization of terminology and innovations for methodologies in developmental toxicology: report of the 8th Berlin Workshop on Developmental Toxicity, 14–16 May 2014, Reprod. Toxicol. 57 (2015) 140–146, https://doi.org/10.1016/j.reprotox.2015.06.046.

[2] I. Chahoud, J. Buschmann, R. Clark, A. Druga, H. Falke, A. Faqi, et al., Classification terms in developmental toxicology: need for harmonization. Report of the Second Workshop on the Terminology in Developmental Toxicology Berlin, 27-28 August

1998, Reprod. Toxicol. (13) (1999) 77–82.
[3] R. Solecki, H. Bürgin, J. Buschmann, R. Clark, M. Duverger, O. Fialkowski, et al., Harmonization of rat fetal skeletal terminology and classification. Report of the Third Workshop on the Terminology in Developmental Toxicology. Berlin, 14-16 September 2000, Reprod. Toxicol. (15) (2001) 713–721.
[4] R. Solecki, B. Bergmann, H. Bürgin, J. Buschmann, R. Clark, A. Druga, et al., Harmonization of rat fetal external and visceral terminology and classification. Report of the Fourth Workshop on the Terminology in Developmental Toxicology, Berlin, 18-20 April 2002, Reprod. Toxicol. (17) (2003) 625–637.
[5] I. Chahoud, F.J.R. Paumgartten, An appreciation of Diether Neubert and Hans-Joachim Merker's contributions to Reproductive and Developmental Toxicology on their 80th birthday, Reprod. Toxicol. 29 (June (3)) (2010) 257–261.
[6] Hu Jingying, Kan Haidong, Fang Tian, Xia Minjie, Yuzhu Wang, Cuiqin Xiang, Xuncheng Ding, Li Weihua, Zhijun Zhou, The international collaboration on ATLAS of developmental abnormalities of laboratory animals, Chin. J. of Carcinogenesis, Teratogenesis Mutag. 30 (2) (2018) 155–157.
[7] L. Zhou, Z. Sun, Atlas of Common Malformations in Laboratory Rabbit and Rat, (2014) ISBN 978-7-5478-2339-2/R 782.
[8] Y. Ooshima, M. Fujiwara, K. Chihara, Y. Izumi, Y. Katsumata, et al., Atlas of Developmental Anomalies in Experimental Animals, (2015) Yakuji Nippo, Ltd., External anomalies –ISBN978-4-8408-1296-2, Visceral anomalies ISBN978-4-8408-1305-1, Skeletal anomalies – ISBN978-4-8408-1320-4.
[9] S.L. Makris, H.M. Solomon, R. Clark, K. Shiota, S. Barbellion, J. Buschmann, et al., Terminology of developmental abnormalities in common laboratory mammals (Version 2), Reprod. Toxicol. 28 (3) (2009) 371–434.
[10] M. Kuwagata, Y. Sakai, S. Tanaka, H. Takashima, R. Katagiri, T. Matsuoka, et al., Historical control data on developmental toxicity studies in rats, Congenit. Anom. (2018) 1–7, https://doi.org/10.1111/cga.12305 2018, Japanese Teratology Society.
[11] M. Horimoto, et al., Toxicological significance of lumber rib in rat fetuses, J. Toxicol. Sci. 41 (Suppl. S302) (2016) 2016, JSOT meeting, in Japanese.
[12] Regulation (EC) No. 1107/2009 of the European Parliament and of the Council Concerning the Placing of Plant Protection Products on the Market and Repealing Council Directives 79/117/EEC and 91/414/EEC, (2019).
[13] Regulation (EU) 528/2012 of the European Parliament and of the Council of 22 May 2012 concerning the making available on the market and use of biocidal products, Off. J. Eur. Union 27 (June) (2012) 1–123 L167.
[14] Regulation (EC) 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No. 1907/2006, Off. J. Eur. Union 31 (December) (2008) 1–1355 L353.
[15] Commission Regulation (EU) 2018/605 of 19 April 2018 amending Annex II to Regulation (EC) 1107/2009 by setting out scientific criteria for the determination of endocrine disrupting properties, Off. J. Eur. Union 20 (April) (2018) 33–36 L101.
[16] European Food Safety Authority (EFSA), OECD/EFSA Workshop on Developmental Neurotoxicity (DNT): The Use of Non-animal Test Methods for Regulatory Purposes, (2016) Brussels, 18. Oktober https://www.efsa.europa.eu/de/events/event/161018b.
[17] ALTEXOECD/EFSA Workshop on Developmental Neurotoxicity (DNT): The Use of Non-Animal Test Methods for Regulatory Purposes, Meeting Report3422017, OECD/EFSA Workshop on Developmental Neurotoxicity (DNT): The Use of Non-Animal Test Methods for Regulatory Purposes, Meeting Report 34 (2) (2017).
[18] S. Schneider, K.C. Fussell, S. Melching-Kollmuss, R. Buesen, S. Gröters, et al., Investigations on the dose–response relationship of combined exposure to low doses of three anti-androgens in Wistar rats, Arch. Toxicol. 91 (2017) 3961–3989, https://doi.org/10.1007/s00204-017-2053-3.
[19] J.L. Simpson, Incidence and timing of pregnancy losses: relevance to evaluating safety of early prenatal diagnosis, Am. J. Med. Genet. 35 (1990) 165–173.
[20] A.T. Hertig, J. Rock, E.C. Adams, A description of 34 human ova within the first 17 days of development, Am. J. Anat. 98 (1956) 435–493.
[21] A.T. Hertig, J. Rock, E.C. Adams, M.C. Menkin, Thirty-four fertilized human ova, good, bad and indifferent, recovered from 210 women of known fertility; a study of biologic wastage in early human pregnancy, Pediatrics 23 (1959) 202–211.
[22] P.H. Harlow, et al., The nematode Caenorhabditis elegans as a tool to predict chemical activity on mammalian development and identify mechanisms influencing toxicological outcome, Sci. Rep. 6 (2016) 22965.
[23] P.R. Hunt, The C. elegans model in toxicity testing, J. Appl. Toxicol. 37 (2017) 50–59.
[24] W.A. Boyd, M.V. Smith, C.A. Co, J.R. Pirone, J.R. Rice, et al., Developmental effects of the ToxCast phase I and phase II chemicals in Caenorhabditis elegans and corresponding responses in zebrafish, rats, and rabbits, Environ. Health Perspect. 124 (2016) 586–593.
[25] A.J. Udvadia, E. Linney, Windows into development: historic, current, and future perspectives on transgenic zebrafish, Dev. Biol. 256 (2003) 1–17.
[26] Hamm, et al., Zebrafish Developmental Screen NICEATM, Abstract: 2014 SOT Annual Meeting, (2013).
[27] OECD Guideline 243/2016: *Lymnaea stagnalis* Reproduction Test, (2016).
[28] Tallarico, et al., Developmental toxicity, acute toxicity and mutagenicity testing in freshwater snails Biomphalaria glabrata (Mollusca: gastropoda) exposed to chromium and water samples, Ecotoxicol. Environ. Saf. 110 (Dec) (2014) 208–215, https://doi.org/10.1016/j.ecoenv.2014.09.005 2014, Epub 2014 Sep 26.
[29] NRC, Toxicity Testing in the 21st Century: A Vision and a Strategy, National Academies Press, Washington, DC, 2007https://www.nap.edu/catalog/11970/toxicity-testing-in-the-21st-century-a-vision-and-a.
[30] R.S. Thomas, R.S. Paules, A. Simeonov, S.C. Fitzpatrick, K.M. Crofton, W.M. Casey, et al., The US Federal Tox21 Program: a strategic and operational plan for continued leadership, ALTEX 35 (2) (2018) 163–168.
[31] NIEHS, Strategic Roadmap for Establishing New Approaches to Evaluate the Safety of Chemicals and Medical Products in the United States, (2018) https://ntp.niehs.nih.gov/iccvam/docs/roadmap/iccvam_strategicroadmap_january2018_document_508.pdf.
[32] E. Fritsche, P. Grandjean, K.M. Crofton, M. Aschner, A. Goldberg, T. Heinonen, et al., Consensus statement on the need for innovation, transition and implementation of developmental neurotoxicity (DNT) testing for regulatory purposes, Toxicol. Appl. Pharmacol. 354 (September (1)) (2018) 3–6, https://doi.org/10.1016/j.taap.2018.02.004.
[33] A. Bal-Price, H.T. Hogberg, K.M. Crofton, M. Daneshian, R.E. FitzGerald, E. Fritsche, et al., Recommendation on test readiness criteria for new approach methods in toxicology: Exemplified for developmental neurotoxicity, ALTEX. 35 (3) (2018) 306–352, https://doi.org/10.14573/altex.1712081 Epub 2018 Feb 23.
[34] C. Zhang, J. Ball, J. Panzica-Kelly, K. Augustine-Rauch, In vitro developmental toxicology screens: a report on the progress of the methodology and future applications, Chem. Res. Toxicol. 29 (2016) 534–544, https://doi.org/10.1021/acs.chemrestox.5b00458.
[35] Piersma, et al., Reprod. Toxicol. 72 (Sep) (2017) 53–61, https://doi.org/10.1016/j.reprotox.2017.05.014 2017, Epub 2017 Jun 4. Review.PMID:28591664, Retinoic acid in developmental toxicity: Teratogen, morphogen an biomarker.
[36] J.F. Robinson, et al., A comparison of gene expression responses in rat whole embryo culture and in vivo: time-dependent retinoic acid-induced teratogenic response, Toxicol. Sci. 126 (March (1)) (2012) 242–254, https://doi.org/10.1093/toxsci/kfr342 2012, Epub 2012 Jan 18.
[37] R. Kodzius, F. Schulze, X.H. Gao, M.R. Schneider, Organ-on-chip technology: current state and future developments, GenesBasel 8 (2017) doi:ARTN 266.
[38] J. Scheinpflug, M. Pfeiffenberger, A. Damerau, F. Schwarz, M. Textor, A. Lang, et al., Journey into bone models: a review, GenesBasel 9 (2018) doi:ARTN 247.
[39] F. Knöspel, Z. Ban, G. Schönfelder, M.R. Schneider, Next milestone in understanding early life-blastoids mimic embryogenesis in vitro, Biol. Reprod. 100 (Jan (1)) (2019) 11–12, https://doi.org/10.1093/biolre/ioy182.

Roland Solecki[a], Martina Rauch[a], Andrea Gall[a], Jochen Buschmann[b], Rupert Kellner[d], Olena Kucheryavenko[a], Anne Schmitt[a,*], Nathalie Delrue[e], Weihua Li[f], Jingying Hu[f], Michio Fujiwara[g], Makiko Kuwagata[h], Alberto Mantovani[i], Susan L. Makris[j], Francisco Paumgartten[k], Gilbert Schönfelder[a,c], Steffen Schneider[m], Silvia Vogl[a], Nicole Kleinstreuer[n], Marlon Schneider[a], Frank Schulze[a], Ellen Fritsche[o], Ruth Clark[p], Kohei Shiota[l], Ibrahim Chahoud[c]

[a] *German Federal Institute for Risk Assessment, Germany*
[b] *Hannover, Germany*
[c] *Charité Universitätsmedizin Berlin, Germany*
[d] *Fraunhofer Institute for Toxicology and Experimental Medicine, Hannover, Germany*
[e] *Organisation for Economic Co-operation and Development (OECD), France*
[f] *NHC Key Lab. of Reproduction Regulation (Shanghai Institute of Planned Parenthood Research), Fudan University, China*
[g] *Astellas Pharma Inc., Tokyo, Japan*
[h] *Hatano Research Institute, FDSC, Kanagawa, Japan*
[i] *Istituto Superiore di Sanità (ISS), Rome, Italy*
[j] *U.S. Environmental Protection Agency, Washington, D.C., USA*
[k] *Oswaldo Cruz Foundation Brazil, Brazil*
[l] *Shiga University of Medical Sciences, Japan*
[m] *BASF SE, Ludwigshafen, Germany*
[n] *National Toxicology Program, Durham, USA*
[o] *Leibniz Research Institute for Environmental Medicine, Düsseldorf, Germany*
[p] *Ruth Clark Associates Ltd., United Kingdom*
E-mail address: devtox@bfr.bund.de (A. Schmitt).

* Corresponding author at: Bundesinstitut für Risikobewertung, Max-Dohrn Strasse 8-10, D-10589, Berlin, Germany.