Strategies in genotoxicology: Acceptance of innovative scientific methods in a regulatory context and from an industrial perspective.

Guy Steiblen[1*], Jan van Benthem[2], Georges Johnson[3]

1. Solvay, Genas, France guy.steiblen@solvay.com + (33) 6 07 99 28 60
   *to whom correspondence should be addressed
2. RIVM-National Institute for Public Health and the Environment, Bilthoven, The Netherlands jan.van.benthem@rivm.nl, + (31) 6 5248 6578
3. Swansea University Medical School, Swansea University, UK g.johnson@swansea.ac.uk, + (44) 1792 295158

Abstract

The tests used and the general principles behind test strategies are now often over 30 years old. It may be time by now, given that our knowledge of genetic toxicology has improved and that we also technically are better able to investigate DNA damage making use of modern molecular biological techniques, to start thinking on a new test strategy. In the present paper, it is demonstrated that the time is there to consider a new approach for genotoxicity assessment of substances. A fit for all test strategy was discussed making use of the most recent technological methods and techniques.

It was also indicated that *in silico* tools should be more accepted by regulatory institutes/bodies as supporting information to better conclude which tests should be required for each separate substance to demonstrate its genotoxic potency. Next to that there should be a good rationale for performing *in vivo* studies. Finally, the need for germ cell genotoxicity testing, essential when classification and labeling of substances is mandatory, was discussed. It was suggested to change the GHS for genotoxicity classification and labelling from *in vivo* tests in germ cells into *in vivo* tests in somatic cells.

Quantitative genotoxicology was also discussed. It appeared that we are currently at a transition, where the science developing to justify carrying out human health risk assessments based on genetic toxicology data sets supported by mechanistic data and exposure data. However, implementation will take time, and acceptance will be supported through the development of numerous case studies. Major remaining questions are: is genetic damage a relevant endpoint in itself, or should the risk assessment be carried out on the apical endpoint of cancer and which genotoxic endpoint should be used to derive the point of departure (PoD) for the human exposure limit?

Keywords

Strategy in genotoxicology; Adverse Outcome Pathway (AOP); Germ cells and Global Harmonized System (GHS); Quantitative genotoxicity.

Introduction

This paper is a report from a workshop “Strategies in genotoxicology: Acceptance of innovative scientific methods in a regulatory context: Strengths and Weaknesses. From a regulatory point of view to an industrial perspective” at the 47th Annual Meeting of the European Environmental Mutagenesis and Genomics Society, held in Rennes, France in May 2019. Registrations of products always require the assessment of their genotoxic potential in order to protect humans and the environment. During this workshop strategies in genotoxicity were observed from a regulatory, an industrial and a quantitative perspective. The views expressed in these presentations are always exclusively those of the speakers.

Discussion

All new and existing chemicals, drugs, food ingredients, cosmetic ingredients, etc. are tested before marketing to be sure that they do not pose a risk for public health. These substances are tested on every toxicological endpoint, including genotoxicity. Testing for genotoxicity has a dual-purpose: i) hazard assessment and ii) classification and labelling according to the Global Harmonisation System (GHS).
Often, for instance for chemicals in Europe, genotoxicity has a special place in this assessment since for genotoxic substances that are also carcinogenic a non-thresholded risk assessment is mandatory.

Genotoxicity testing has some main principles. The maximum level of information should be obtained with a minimum number of tests. The tests therefore have to be very adequate and efficient. Next to that the extent of genotoxicity testing should be related to the extent of (potential) human exposure: the more human exposure the more testing. And finally the number or use of animals should be reduced. This means that *in vitro* tests are strongly preferred.

Genotoxicity testing has three genotoxic endpoints gene mutations (mutagenicity), structural chromosome aberrations (clastogenicity) and numerical chromosome aberrations (aneuploidy). For a complete assessment of genotoxicity of substances all endpoints have to be covered. Unfortunately, there is no genotoxicity test to cover all endpoints, indicating that a combination of tests is necessary to cover genotoxicity. The result is a strategy for genotoxicity testing.
For testing, a number of both *in vitro* and *in vivo* tests are available. Since 1982, recommendations for the conduct of the commonly used tests are provided in the Organization for Economic Co-operation and Development (OECD) test guidelines describing the major steps of the tests. However, many tests are older and were already developed (long) before the OECD started its test guidelines program. Since 1982 not much has changed. A few tests were deleted in 2014 e.g. the Sister Chromatid

Exchange (SCE) test, the *Saccharomyces* and *Drosophila* tests as well as the *in vitro* Unscheduled DNA Synthesis (UDS) test. The decision to delete some test guidelines was made based on the observation that these tests are rarely used in the various legislative jurisdictions, and on the availability of newer tests showing a better performance for the same endpoint. The deleted test guidelines should not be used for new testing and are no longer a part of the set of OECD recommended tests; however, data previously generated from these deleted test guidelines can still be used in a risk assessment or other types of regulatory decisions.
Only a few new tests were introduced: the *in vitro* mammalian micronucleus test (OECD 487 [1]), the transgenic gene mutation test (OECD 488 [2]) and the *in vivo* comet assay (OECD 489 [3]). The former test guideline (OECD 476) was divided in two test guidelines: one covering hprt/xprt mutations (revised as and a new guideline covering tk mutations in recommended TK6 cells and L5171Y cells (OECD 490 [4]). All other tests, were several times updated or revised. However, it may be clear that the possibilities to keep updating or revising these tests are limited. There comes a point that a revision becomes nit-picking.

At least within Europe, the genotoxicity testing strategy between regulatory bodies (ECHA, EFSA, SCCS, EMA, COM) is rather similar. The strategy consists of a stepwise approach starting with *in vitro* tests. In the event of negative *in vitro* results, it can be concluded that the substance has no genotoxic potential. Positive *in vitro* results trigger appropriate *in vivo* studies to assess whether the genotoxic potential observed *in vitro* is expressed *in vivo*. A negative result in these confirming *in vivo* tests overrules the *in vitro* positive result and the substance is considered not genotoxic. Positive *in vivo* results make the substance tested a genotoxicant. For substances which are genotoxic in somatic cells *in vivo*, the potential for germ cell mutagenicity should be considered. It is recognised that standard reproduction studies do not cover all germ cell effects. Next to that genotoxicity testing is more or less a pre-screen for carcinogenicity and as such many legislations ask for a carcinogenicity test when a substance is a mutagen.
Still, consideration should be given to whether specific features of the test substance might require substitution of one or more of the recommended *in vitro* tests by other *in vitro* or *in vivo* tests in the basic battery. The *in vivo* test to be performed is dependent on the genotoxic endpoint that is positive *in vitro*, e.g. a positive *in vitro* micronucleus test should be followed by an *in vivo* micronucleus test.

As an example, under the REACH regulation, the required *in vitro* assays for genotoxicology are always the same. The Ames test (OECD 471) [5] is required for a substance production or importation from 1 to 10 tons/year; the upper band tonnage between 10-100 tons/year requires two additional tests (e.g. the *in vitro* micronucleus test (OECD 487) [1] and *in vitro* mammalian cell gene mutation tests using the thymidine kinase gene (OECD 490) [4]. If all these three assays are found to be negative, i.e. do not generate any genotoxic response, then the substance is considered to be non-genotoxic and no additional *in vitro* or *in vivo* assays are then required. Under REACH, the use of animals for genotoxicity assessment is possible but needs to be justified.
Scientifically, this testing battery proposed by REACH regulation sounds acceptable according to literature for assessing substance genotoxicity. In a review of Kirkland et al. (2011) [6], the authors demonstrated that a battery of only two *in vitro* tests, the Ames

test (OECD 471) [5] and the *in vitro* micronucleus assay (OECD 487) [1]**,** is enough as both are covering all three endpoints of genotoxicology, i.e. gene mutations, clastogenicity and aneuploidy. Nearly 100% (958 out of 962) of rodent carcinogens or *in vivo* genotoxins were correctly detected with these two tests, which make this battery a particular high sensitive combination.

Consequently, it may be justified to wonder whether *in vitro* tests alone are sufficient for both single substance or mixture genotoxic potency assessment. Such philosophy would be also in line with United States Environmental Protection Agency's (US-EPA) intention which targets to ban animal use in toxicology studies by 2035 [7].

The tiered approach required by most regulations could then focus on *in vitro* tests and on (Q)SAR in order to avoid unnecessary use of animals.

Table 1: predictivity of 3 common genotoxicity tests: Gene mutation test in bacteria (Ames test); MLA: mouse lymphoma assay (gene mutation test in mammalian cells); CA: chromosome aberration test. (modified or calculated from Kirkland et al., 2005 [8])

	Ames	**MLA**	**CA**	**Ames + CA**	**Ames + CA + MLA**
Sensitivity (%)	**58.8**	**73.1**	**65.6**	**82.2**	**84.7**
Specificity (%)	**73.9**	**39.0**	**44.9**	**33.1**	**22.9**
Concordance (%)	**62.5**	**62.9**	**59.8**	**71.1**	**64.8**

However, the sensitivity (percentage of correctly predicted genotoxic substances) as well as previously reported the specificity (percentage of correctly predicted non-genotoxic substances) of the current test is rather poor. In table 1 the specificity and sensitivity of the most commonly used *in vitro* genotoxicity tests are shown. It is clear that particularly the specificity of some tests is much too low for tests that are used for regulatory decisions [8], [9]). As the possibilities to update these tests are limited as was discussed before, the chances to improve the specificity and sensitivity are limited as well. That is why such a proposed standalone *in vitro* battery of tests can still be challenged regarding its lack of specificity.

Indeed, there is a strong ethical pressure and policy against the use of experimental animals. For example, there is already legislation for the testing of cosmetic ingredients where the use of *in vivo* tests is prohibited. That leads to carcinogenicity assessment being dependent on *in vitro* genotoxicity tests only, for which we just concluded that they are not optimal.

Since 2005, several initiatives have been undertaken to improve the specificity of the *in vitro* tests without compromising sensitivity, e.g. preference for p53-competent human cells, cytotoxicity measures based on cell proliferation, carefully controlling the upper limit of cytotoxicity or the highest non-toxic concentration, all of which have been incorporated into the OECD guideline revisions [4].

When *in vivo* studies are required, registrants are often sensitive to the 3R principles. This has contributed to an increase in the assessment of multiple genotoxic endpoints from the same animals . As an example, it is common that data from assays which combine a subacute toxicology study in the rodent, with micronucleus assessment in the blood, as well as comet assay in the liver [10] (figure 1) are presented in submitted dossiers. Combined studies are even recommended by regulations like the ICH S2 (R1) (step 4 version, 2011) [11] for assessing the potential genotoxicity of a substance under pharmaceutical regulation.

There is also be a good rationale to perform *in vivo* assays in rats rather than in mice for the *in vivo* micronucleus test. The logic behind is that there is usually more kinetic data available on rats compared to mice (e.g. absorption, distribution, metabolism, and excretion (ADME) studies). Toxicokinetic (TK) information is required in most regulations and is performed mainly on rats compared to mice, so when TK information is already available from previous TK studies, it is not required to repeat the generation of TK data during an *in vivo* micronucleus assay.

In addition, if the absorption has to be assessed during a micronucleus test to prove animal's exposure, rats should be also prioritized, if possible, since it is possible to use one animal for assessing the two different endpoints, i.e. bioanalysis and micronucleus assessment. In mice, assessment of both parameters would require double the animals because there is not enough blood available per animal to assess both endpoints.

Figure 1: Combine multi-assay in *in vivo* rodent

It is generally accepted that standard genotoxicity tests are rather old. The test strategy may use some refreshment. Most regulatory guidelines contain a list of tests which have to be ticked off, and they do not cover new endpoints or new groups of chemicals. It may also be time to think about a new approach for genotoxicity testing. A requisite for a new strategy is that the three genotoxic endpoints, gene mutations, structural chromosome aberrations and numerical chromosome aberrations are still covered. Moreover, there should be room for classic as well as new tests, but this will required a paradigm shift in genotoxicity thinking. A new harmonised and more advanced approach, should include options to use mechanism-based assays, and additionally, both toxicokinetics and exposure have to be included. Of course also the Adverse Outcome Pathway (AOP)

science will be a substantial part of it. However, the most dramatic change in thinking, will be that the test substance and not the list of tests (the strategy) should be the main consideration. On a substance by substance basis the real strategy for each specific substance has to be determined. The idea is that there is a general framework of a strategy that will be fine-tuned for each single substance, and chemical group.

One of the major changes will be that, based on the strong ethical pressure and policy against the use of experimental animals, the framework predominantly or even exclusively consists of *in vitro* genotoxicity tests. Fortunately, there is already experience with what can be expected when *in vivo* tests are not allowed. Since the marketing ban of March 11, 2013 [12], in Europe *in vivo* testing is prohibited for cosmetic products or ingredients, irrespective of the availability of alternative non-animal tests. This major change, lead to evaluations on genotoxicity of cosmetic ingredients being done exclusively in silico, i.e QSAR and read across, or on *in vitro* tests, i.e. the gene mutation test in bacteria and the *in vitro* micronucleus test. A conclusion when one of the two tests was positive (considered genotoxic) or negative (considered non-genotoxic) may be clear but difficulties were expected when one or the 2 tests was positive. In the end the problems were excellently tackled by collaboration between all stakeholders, including industry and regulatory institutes. It resulted in a revised Notes of Guidance [13] in which it is stated that the test should be critically evaluated for e.g. false positive results. The latter may, for example, be the result of the quality of a study, excessive cytotoxicity, positive results inside the historical control data or by the presence of impurities. For specific tests like the gene mutation test in bacteria, the reason for a false positive result may be bacterial toxicity or specific metabolism in bacteria.

If after checking for false negatives the positive result remains, then further testing may be considered. Further tests can be classic genotoxicity tests like the mammalian chromosomal aberration test or the gene mutation test in mammalian cells, but also tests for which an OECD test guideline is not available like the 3D-skin micronucleus test, the 3D-skin comet assay, γH2AX assay, the hen's egg test (Het) micronucleus test, indicator approaches like Multiflow or Toxtracker and even techniques which are very new like array approaches, next generation sequencing or recombinant cell models. A prerequisite is that a justification for the test is available and that the final conclusion is made on weight of evidence. Until now, excessive problems in deciding whether a substance is genotoxic or not did not occur. Industry screens their new substances also for genotoxicity and uses sensitive new techniques including systems biology and high throughput tests. Why is it not possible to use new approaches for decision making as well, rather than sticking with the rather old fashioned tests low throughput and low content assays as described before!

A common comment on why these new approaches are not suitable, is that they are not validated and that an OECD test guideline for these tests is not available. Of course this is true, and the lack of protocol standardisation comes with its own issues. On the other hand, adequate scientific justification of why a specific test is used along with a clear study design and test protocol, should be sufficient to overrule the comment that an OECD test guideline is not available. Moreover, the tests mentioned are already known

in the genotoxic community as well as their characteristics. Concerning validation, that is a more difficult story. The present validation project is a long, time consuming and expensive course. Today, there is much discussion regarding the simplification of the validation process [14, 15]. Hopefully this will lead in the near future to a shorter process without losing the same level of quality.

It may be time to think about a new strategy for genotoxicity testing [16], or at least introducing increased flexibility for inclusion of new approaches. When developing a new strategy, it would be worthwhile considering an overall strategy for genotoxicity testing that fits for all substances: chemicals, drugs, cosmetic ingredients, food etc. It is clear that all strategies are similar, which is expected since they are all designed for detecting hazard. Determination of genotoxic hazard can be performed identically according to one single strategy for all groups of substances. However, divergences in regulatory guidelines become important when the risk of the hazard is considered: e.g. no genotoxic substance in food or a risk/benefit approach for drugs/chemicals.

Such a new framework is already described by the “clean sheet” workgroup of the Genetic Toxicology Technical Committee (GTTC) of the Health and Environmental Sciences Institute (HESI) [16]. In this framework the substance is leading. Figure 2 clearly shows the different steps to be taken in such a new approach. It is clear that next to the planning and scoping, also exposure, knowledge base, test selection and appropriate Point of Departure (PoD) selection play a very important role. In the paper by Dearfield et al., 2017 [16] all steps are clearly explained. In the end it should lead to estimated acceptable levels for endpoint of human exposure and risk characterization. This “Clean Sheet framework” is now tested with (industrial) case studies.

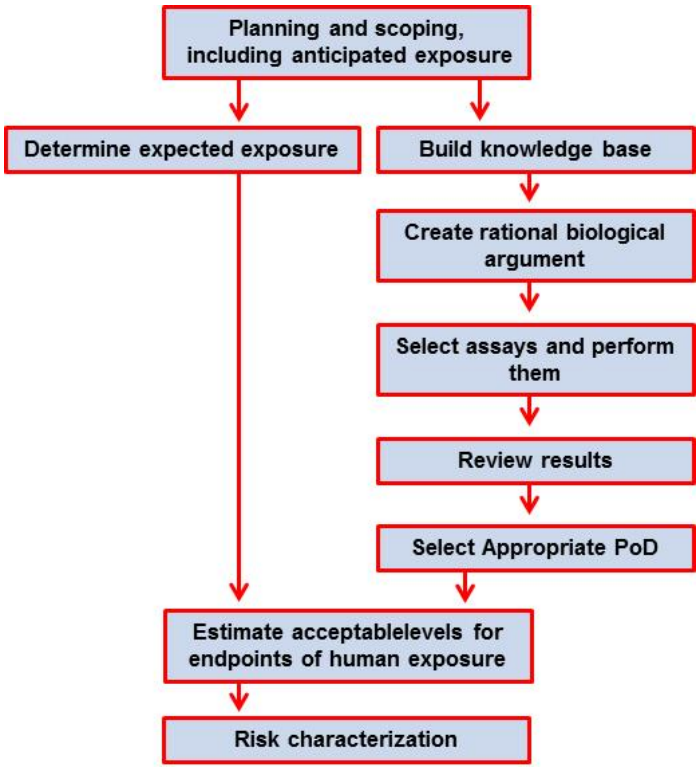

Figure 2: Strategy for examining genomic damage (from Dearfield et al., 2017 [16])

In figure 3, this theoretical framework is filled in in a more practical way. Still the substance is leading and it starts with a planning and scoping including which factors are important for a specific risk assessment of a specific substance. Relevant questions are, for example, how humans are exposed, are some population groups more sensitive for the substance, which legislation is applicable. The first step in a putative framework could deal with consideration of physical chemical data and/or existing data. Theoretically this may give enough information for a conclusion on genotoxicity of the substance, e.g. when there is no exposure, testing is redundant. Other follow up levels may be *in silico* methods, like quantitative structure activity relationship (QSAR) and read across as well as screening tests.

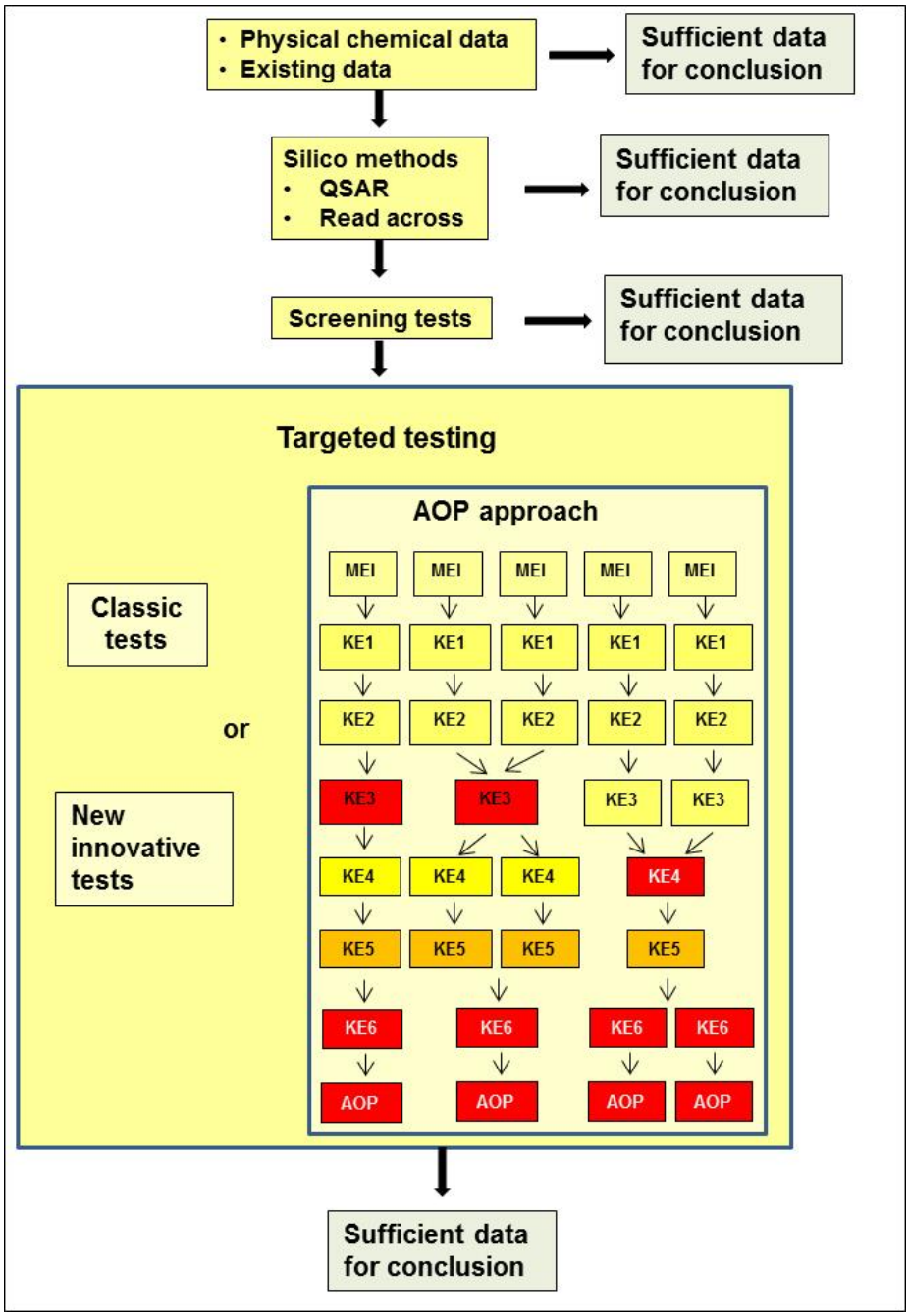

Figure 3: Putative strategy to determine the genotoxicity of substances. Following exposure with a substance, a Molecular Initiating Event (MEI) will be generated followed by Key Event(s) (KE) leading to the characterization of an AOP in a dedicated matrix (cell, protein-chips, etc.).

Most of EU regulations like REACH regulation EC [17], pesticide (PPP) [18], biocide (BPR) [19] and many others have a common philosophy in line with the above suggestion. The philosophy is to reduce the use of animals for experiments by using a

tiered approach, requesting at first alternative, assays like *in vitro* tests and/or *in silico* tools. The cosmetic regulation (1223/2009) [12] is even more drastic, because animal use is completely forbidden for experimental use.

For screening purposes, these tendencies or requirements have enhanced an adaptation of industrials from the 2000's in terms of assay development or alternative method developments. As an example *in silico* tools have been well developed and applied in genotoxicology for many years, giving a wide range of (Q)SAR tools with a powerful ability to predict genotoxic potency of a substance [20]. It is nevertheless not possible to replace completely *in vitro* or *in vivo* assays by these *in silico* tools. In fact, (Q)SAR are currently considered for providing alerts on hazardous substances. However, they are also tools which can be used to replace some classic assays for some regulations like REACH particularly for well studied groups of compounds, and this option is possible for many endpoints required in REACH, i.e. from ecotoxicology to toxicology assays (Annex XI of REACH regulation) [17]. Currently, it looks like nevertheless that these *in silico* tools are essentially accepted as "supporting studies" by regulators, so the "key study", i.e. the main assay (*in vitro* or *in vivo*) is still essential for a substance submission under REACH. Thus far, the authors are unaware whether genotoxicity data obtained by *in silico* tools only was allowed to address the requirement for a substance submission under REACH.

Concerning screening tests, one may think of all kinds of high throughput tests that already are commonly used for decision making purposes within the industrial setting. Again the combined information from these different assays may be good enough to conclude on genotoxicity. However, this conclusion can probably only be taken after targeted testing and information on a putative mode of action becomes clear. For targeted testing one should think of using an already available AOP in full in first instance or a tiered number of classical or new innovative genotoxicity tests. Until now six genotoxicity AOPs are available [21-23].

Targeted testing may be repeated as one or more modes of action may be applicable. When *in silico* and *in vitro* analysis does not give a reliable answer, *in vivo* testing may be considered, but as a last resort. In the end it should lead to a final conclusion on genotoxicity for this specific substance according to this specific strategy.

Most of the available results in genetic toxicology are focused on bacterial or somatic cells. Currently, the Global Harmonized System (GHS) which is used in most of substance/mixture evaluation regulations (i.e. pesticide, biocide, drug…) [24] contains a germ cell mutagen classification. Therefore, there seems to be a gap between what is most available, i.e. data from bacterial and somatic cells and the requirements to conclude on substance classification, i.e. germ cells.

Today, most EU regulations try to avoid the use of *in vivo* tests in genotoxicology as much as possible, only some still require such performance independently of *in vitro* results. For example, pesticide regulation [18] required the systematic performance of an *in vivo* rodent assay. The consequence is that most results generated for a substance in genotoxicity assays are mainly available in *in vitro* tests and more seldom in *in vivo* tests that, are predominantly performed on somatic cells. The impact is not neutral,

particularly for the substance classification under the Global Harmonization System (GHS) [25], because GHS requires germ cell genotoxicity data for classification(Table 2). With data available mainly on somatic cells, it is therefore difficult to conclude on a substance classification, because nearly no information is available on germ cells [26]. In the GHS, the category 1A or 1B will lead to a market ban of substance or mixtures.

Another problem is that currently, regulations recognize only animal assays to have the ability to assess the genotoxic potential in germ cells. Among the *in vivo* assays available to assess germ cell potent genotoxicity are:

- The transgenic mice test (OECD 488) [2] which has the ability to detect gene mutations in the mouse or rat.
- The rodent dominant lethal test (OECD 478) [27], the mammalian spermatogonial chromosomal aberration test (OCDE 483) [28] and the mouse heritable translocation assay (OCDE 485) [29] these last three allow to detect chromosomal aberration endpoint.

These tests are very demanding, use a lot of animals and are therefore far from the 3R principles. In addition, these tests are not performed by a lot of laboratories which make their use very difficult for most of industrials.

Moreover, there is no example of a substance positive only in germ cells and not in somatic cells in genotoxicity. This statement is clearly indicated in the guidance on genotoxicity testing and data interpretation for pharmaceuticals intended for human use (ICHS2 (R1): 2.4 Germ cells paragraph) [11]. Based on the present knowledge and the presently available tests, which are insensitive and uses large numbers of animals, germ cell genotoxicity testing may be considered redundant. However, in a recent review Marchetti et al. (2019, in press) [30] report on indications that many substances may also be a risk for human germ cells, justifying germ cell testing.

To focus on the germ cells is not necessarily to claim that germ cells and somatic cells are different, it can be more driven by the fact that the kinetics of substances are not the same between substances. Consequently a substance could be positive in somatic cells, whilst it is negative in germ cells, for instance because it is not able to reach the germ cells. Marchetti et al. [30] indicate that if mutation is considered a toxicological endpoint and mutation will be used in human risk assessment, the PoD for germ cells surely will be different from PoD's in somatic cells which makes germ cell mutagenicity testing essential. Therefore, the idea to focus in somatic cells only for the GHS classification purpose instead of germ cells is not always applicable.

Table 2: Hazard categories for germ cells mutagens extracted from ECHA guidance (v.5, 2017)

Annex I: ***3.5.2.2.*** *For the purpose of classification for germ cell mutagenicity, substances are allocated to one of two categories as shown in Table 3.5.1.*

Table 3.5.1

Hazard categories for germ cell mutagens

Categories	*Criteria*
CATEGORY 1:	*Substances known to induce heritable mutations or to be regarded as if they induce heritable mutations in the germ cells of humans.* *Substances known to induce heritable mutations in the germ cells of humans.*
Category 1A:	*The classification in Category 1A is based on positive evidence from human epidemiological studies.* *Substances to be regarded as if they induce heritable mutations in the germ cells of humans.*
Category 1B:	*The classification in Category 1B is based on:* – *positive result(s) from in vivo heritable germ cell mutagenicity tests in mammals; or* – *positive result(s) from in vivo somatic cell mutagenicity tests in mammals, in combination with some evidence that the substance has potential to cause mutations to germ cells. It is possible to derive this supporting evidence from mutagenicity/genotoxicity tests in germ cells in vivo, or by demonstrating the ability of the substance or its metabolite(s) to interact with the genetic material of germ cells; or* – *positive results from tests showing mutagenic effects in the germ cells of humans, without demonstration of transmission to progeny; for example, an increase in the frequency of aneuploidy in sperm cells of exposed people.*

CATEGORY 2:	*Substances which cause concern for humans owing to the possibility that they may induce heritable mutations in the germ cells of humans.*
	The classification in Category 2 is based on:
	– *Positive evidence obtained from experiments in mammals and/or in some cases from in vitro experiments, obtained from:*
	  – *Somatic cell mutagenicity tests in vivo, in mammals; or*
	  – *Other in vivo somatic cell genotoxicity tests which are supported by positive results from in vitro mutagenicity assays.*
	Note: Substances which are positive in in vitro mammalian mutagenicity assays, and which also show chemical structure activity relationship to known germ cell mutagens, shall be considered for classification as Category 2 mutagens.

Until now the genotoxicity of chemicals has been evaluated in a yes/no fashion. Genotoxicity is hazard assessment. But there is already a paradigm shift ongoing from this qualitative approach into a quantitative approach. Dose response modelling to establish a PoD was introduced in genotoxicity testing. Assessing the potency and PoD in addition to the positive/negative effect, leads to the use of mutations for human health risk assessment. In fact there are already examples where this approach was successfully used [31-33]. In this approach gene mutation has to be considered as a toxicological endpoint. For more detail on this consideration see the review by Heflich et al. (2019, in-press) [34].

The idea of quantitative genetic toxicology is to advance beyond the standard hazard assessment of substances, to a risk based approach. This changes the assessment from whether substances are genotoxic or not genotoxic to calculations of exposure levels that are of negligible concern to the exposed population. Advancements have been made in defining PoD metrics from genetic toxicity data, and this has progressed the consideration of these data for risk assessment purposes. The standard genetic toxicity models of the *in vivo* micronucleus assay and the transgenic gene mutation tests are suitable for dose response analysis and derivation of PoD. These *in vivo* assays can be assessed using Benchmark Dose (BMD) analysis. The BMD approach uses statistical models to define a small but measurable increased above the background, and it is a more advanced statistical approach than pairwise testing which is used to derive No Observed Effect Levels (NOEL).

Some recent developments with the BMD approach have led to increase precision, with the major ones being the covariate BMD approach put forward by Slob and Setzer [35] which uses conserved shape parameters to increase precision in the analyses. Another improved part of BMD analysis comes from adjusting the Critical Effect Sizes (CES) from the default of 5% or 10% to a higher part of the dose response. Slob [36] and Zeller et al. [37], used extensive data sets to calculate that for genetic toxicity data sets, a CES of 50% should be used. Both modifications to the BMD analysis, lead to decreased range in BMD confidence interval and increased precision in the BMD.

There are numerous different approaches for calculating human exposure limits using genetic toxicity PoD metrics. They are the same as those derived using PoD from other toxicological endpoints, or even from the cancer bioassay. The calculations involve an extrapolation to a human comparable dose, and then division by a number of uncertainty/adjustment factors to cover the diversity of the human population, the study duration, severity of endpoint, animal model being used, and which metric was being

used. This provides a number such as a Permissible Daily Exposure (PDE), Acceptable Daily Intake (ADI) or other comparable approaches [38]. There is also the less complex method, where the PoD is divided by the human exposure level, to calculate the Margin of Exposure (MOE). If the MOE is over 10,000, then it is considered to be of negligible concern, but a cause for concern if the MOE is below this value.

The HESI GTTC has been tasked with publishing case studies, where example substances are assessed using these different approaches. In the most case examples, PDEs were calculated and compared based on genetic toxicity data and cancer bioassay data for certain alkylnitrosamines. These have recently been identified as impurities in some pharmaceutical products, and the products were recalled. It is therefore important to determine whether the exposed population has an increased risk of mutation and cancer due to this exposure. The PDE derived from the cancer bioassay data were excellent due to the high power and high-quality cancer bioassay study for the alkylnitrosamines N-nitrosodiethylamine (NDEA) and N-nitrosodimethylamine (NDMA). Data from the transgenic gene mutation studies were also good, but the studies included short term dosing, so the adjustment factors were larger and potentially required further adjustment. However, in this case, the PDEs were comparable and the mutation PDEs supported the cancer bioassay derived PDEs. For N-Nitrosodibenzylamine (NMBA), no suitable cancer bioassay data were available, however the gene mutation data were relevant and usable. These case studies show that PDE can be derived, but there are still ongoing discussions about when they would be suitable for use in a risk assessment. The major considerations are that the dosing strategy and study design should be suitable, and that multiple data sets are available to derive the PoD. Extensive mechanistic understanding for a defined non-linear or threshold dose response are required for the PDE or MOE type approach.

Benzene has also become an interesting case study. In 2018, ECHA carried out a risk assessment for benzene, used the *in vivo* micronucleus data to define Occupational Exposure Levels (OELs), and these were compared to those from OELs using human exposure data. They were very comparable, and when the *in vivo* MN data were re-analysed using the new best practice of covariate BMD and a CES of 50%, the data were of great interest and provided a higher utility than those derived from the previous analysis [39].

We are currently at a transition, where the science used to justify carrying out human health risk assessments based on genetic toxicology data sets is suitable. However, there needs to be a development of case studies in order to overcome the numerous hurdles that are appearing. This will take time, and acceptance will be supported through the development of numerous case studies in the different areas from pharmaceutical impurities, food contaminants, industrial chemicals, cosmetics etc.

Resistance to using genetic toxicology data for human health risk assessments is common within the field. The main arguments are;

1. Is genetic damage a relevant endpoint in itself, or should the risk assessment be carried out on the apical endpoint of cancer?
2. Which endpoint should be used to derive the PoD for the human exposure limit?
3. Which study design, including acute, sub-chronic or chronic dosing, number of days, replicate number, and more.
4. The adjustment factors are also a huge area. Issues include which factors to use for the exposure time, 1-10 or even higher? Severity of endpoints, is mutation as severe as cancer?
5. How much mechanistic data is required to support the use of a PDE through the 'threshold mechanism' type approach, outlined in the ICH M7 framework?

Future perspectives for quantitative genetic toxicology

Once the issues have been overcome, the result will be that genetic toxicity data can be used for risk assessment purposes. *In vivo* data from relevant endpoints of mutation and/or chromosome damage will be used to derive BMDL metrics from which Health Based Guidance Values (HBGV) will be calculated. These will be compared with those derived from other toxicological endpoints, and the most relevant and most conservative metric will be used as the final value from which the risk assessment is carried out. This is based on the recognition that mutation is a relevant endpoint and that the human population should be protected from increased risk of mutation [34]. In addition to this, mutation is linked with cancer, and the development of case studies will potentially show that genetic toxicity based HBGV will be comparable to those from the cancer bioassay, obviously ensuring exposure, tissue, study design, adjustment factors are suitable. Such a major advancement will lead to a reduction in animal testing through less cancer bioassay studies being required, and additionally through the increased use of combined genetic toxicity assays.

Additional advances that could improve risk assessment based on genetic toxicity data, would be the use of next generation sequencing in place of transgenic gene mutation data. This would lead to mutations in cancer genes and other disease related genes forming the assessment, which would increase relevance of the mutation endpoint for use in risk assessment.

Summary and perspectives

The field of genetic toxicity testing is strongly in motion. There are many new developments and approaches, including the development of high throughput and screening tests. *In silico* approaches like QSAR and read across are generally accepted. The idea for quantitative genotoxicity, in which genotoxicity is a toxicological endpoint, gains more and more ground. Taking into account the quality of the current classical tests as compared to the new developments, it may be clear that genotoxicity testing is ready for an update.

There is no need to immediately go to a clean sheet approach but why not start with considering the chemical as the lead and not the tests, and think of, next to the classical strategy, an alternative new approach and framework when preparing a legislation dossier. Hopefully this will trigger a new framework in the near future.

Databases using a significant number of substances, indicate that *in vitro* test batteries using the Ames test as well as the micronucleus assay would have a sensitivity of nearly 100% making the use of animals test not necessary. The drawback is that there is a lack of specificity of such tests that could make their use, as standalone, critical.

Currently, although many data are available on somatic cells, on substances or mixtures, they will not allow to classify a substance according to the GHS which is mainly based on germ cells results. As there is no unique germ cell mutagen, it would therefore be more appropriate to change the GHS for the genotoxicity classification and propose instead *in vivo* assays in somatic cells instead of germ cells.

We are currently at a transition, where the science is developing in order to justify carrying out human health risk assessments based on genetic toxicology data sets, which is supported by mechanistic data and exposure data. With this is mind, in vitro genetic toxicity data should be used for potency ranking and mechanistic understanding which can support the *in vivo* derived BMDL and HBGV. Future advances in this area, in line with the efforts from EPA [40] will lead to in vitro to *in vivo* comparisons to allow calculation of usable BMDL from in vitro studies. However, there are numerous issues that need to be overcome in order to address the shortcomings of the current approaches in this area. In silico analysis combined with potency ranking, chemical grouping and chemical categorization also offer opportunities to enhance the application of in vitro data.

584

585

586

587 **Bibliographical review**

588

589 [1] OECD, OECD Guideline for the Testing of Chemicals. In Vitro Mammalian Cell Micronucleus Test. 487,
590 In Vitro Micronucleus Test2010.
591 [2] OECD, OECD Guidelines for Testing of Chemicals. Section 4: Health Effects: Test No.488: Transgenic
592 rodent somatic and germ cell mutation assays, 2011.
593 [3] OECD Guidelines for the Testing of Chemicals, Section 4: Test No. 489: In Vivo Mammalian Alkaline
594 Comet Assay, 2016.
595 [4] OECD Guidelines for Testing of Chemicals. Section 4: Health Effects: Test No.490: In vitro mammalian
596 cell gene mutation test using the Thymidine Kinase gene, 2015.
597 [5] OECD, OECD Guideline for the Testing of Chemicals. Bacterial Reverse Mutation Test. 471, in: OECD
598 (Ed.)1997.
599 [6] Kirkland D, Reeve L, Gatehouse D, V. P., A core in vitro genotoxicity battery comprising the Ames test
600 plus the in vitro micronucleus test is sufficient to detect rodent carcinogens and in vivo genotoxins. ,
601 Mutagenesis, 721 (2011) 27-73.
602 [7] United States Environmental Protection Agency: Directive to prioritize efforts to reduce animal
603 testing, 2019.
604 [8] D. Kirkland, M. Aardema, L. Henderson, L. Muller, Evaluation of the ability of a battery of three in
605 vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens I. Sensitivity, specificity
606 and relative predictivity (vol 584, pg 1, 2005), Mutat. Res. Genet. Toxicol. Environ. Mutagen., 588 (2005)
607 70-70.
608 [9] D. Kirkland, E. Zeiger, F. Madia, R. Corvi, Can in vitro mammalian cell genotoxicity test results be used
609 to complement positive results in the Ames test and help predict carcinogenic or in vivo genotoxic
610 activity? II. Construction and analysis of a consolidated database, Mutat. Res. Genet. Toxicol. Environ.
611 Mutagen., 775 (2014) 69-80.
612 [10] S. Khanal, P. Singh, S. Avlasevich, D.K. Torous, J. Bemis, D. SD, Integration of liver and blood
613 micronucleus and Pig-a gene mutation endpoints into rat 28-day repeat-treatment studies: Proof-of-
614 principle with diethylnitrosamine., Mutat. Res., (2018) 30-35.
615 [11] ICH, ICH Topic S2 (R1). Guidance on Genotoxicity Testing and Data Interpretation for
616 Pharmaceuticals Intended for Human Use, in: CPMP/ICH (Ed.)2008, pp. 1-13.
617 [12] EC, Regulation (EC) No 1223/2009 of the European parliament and of the Council of 30 November
618 2009 on cosmetic products (2009).
619 [13] Scientific Committee on Consumer Safety 2018. The SCCS notes of guidance for the testing of
620 cosmetic ingredients and their safety evaluation, 10th revision. , 2018.
621 [14] A.H. Piersma, T. Burgdorf, K. Louekari, B. Desprez, R. Taalman, R. Landsiedel, J. Barroso, V. Rogiers,
622 C. Eskes, M. Oelgeschlager, M. Whelan, A. Braeuning, A.M. Vinggaard, A. Kienhuis, J. van Benthem, J.
623 Ezendam, Workshop on acceleration of the validation and regulatory acceptance of alternative methods
624 and implementation of testing strategies, Toxicology in Vitro, 50 (2018) 62-74.
625 [15] T. Burgdorf, A.H. Piersma, R. Landsiedel, R. Clewell, N. Kleinstreuer, M. Oelgeschlager, B. Desprez, A.
626 Kienhuis, P. Bos, R. de Vries, L. de Wit, T. Seidle, J. Scheel, G. Schonfelder, J. van Benthem, A.M.

Vinggaard, C. Eskes, J. Ezendam, Workshop on the validation and regulatory acceptance of innovative 3R approaches in regulatory toxicology - Evolution versus revolution, Toxicology in Vitro, 59 (2019) 1-11.
[16] K.L. Dearfield, B.B. Gollapudi, J.C. Bemis, R.D. Benz, G.R. Douglas, R.K. Elespuru, G.E. Johnson, D.J. Kirkland, M.J. LeBaron, A.P. Li, F. Marchetti, L.H. Pottenger, E. Rorije, J.Y. Tanir, V. Thybaud, J. van Benthem, C.L. Yauk, E. Zeiger, M. Luijten, Next Generation Testing Strategy for Assessment of Genomic Damage: A Conceptual Framework and Considerations, Environmental and Molecular Mutagenesis, 58 (2017) 264-283.
[17] EC, No. 1907/29006 concerning the Registration, Evaluation, Authorisation and restriction of Chemicals (REACH), (2006).
[18] EC, Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council directives 79/117/EEC and 91/414/EEC, 2009.
[19] EC, Regulation (EU) No 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 2012.
[20] N.F. Cariello, J.D. Wilson, B.H. Britt, D.J. Wedd, B. Burlinson, V. Gombar, Comparison of the computer programs DEREK and TOPKAT to predict bacterial mutagenicity, Mutagenesis, 17 (2002) 321-329.
[21] C.L. Yauk, J. Bishop, K.L. Dearfield, G.R. Douglas, B.F. Hales, M. Luijten, J.M. O'Brien, B. Robaire, R. Sram, J. van Benthem, M.G. Wade, P.A. White, F. Marchetti, The development of adverse outcome pathways for mutagenic effects for the organization for economic co-operation and development, Environmental and Molecular Mutagenesis, 54 (2013) 79-81.
[22] F. Marchetti, A. Massarotti, C.L. Yauk, F. Pacchierotti, A. Russo, The adverse outcome pathway (AOP) for chemical binding to tubulin in oocytes leading to aneuploid offspring, Environmental and Molecular Mutagenesis, 57 (2016) 87-113.
[23] A.A. Sasaki JC, Bryce SM, Custer L, Dearfield K, Dietz Y, Elhajouji A, Escobar P, Fornace Jr AJ, Froetschl R, Galloway S, Hemmann U, Hendriks G, Li H-H, Luijten M, Ouedraogo G, Peel L, Pfuhlker S, Roberts D, Thybaud V, van Benthem J, Yaul CL, Schuler M., Understanding genotoxic mode of action to inform risk characterization: application of the adverse outcome pathway framework. In preparation.
[24] Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification and labelling and packaging of substances and mixtures, 2008.
[25] EC, Regulation (EC). No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification and labelling and packaging of substances and mixtures., 2008.
[26] M. Woutersen, M. Beekman, M.E.J. Pronk, A. Muller, J.A. De Knecht, B.C. Hakkert, Does REACH provide sufficient information to regulate mutagenic and carcinogenic substances? , Human and Ecological Risk Assessment, 0 (2018) 1-20.
[27] OECD, OECD Guidelines for Testing of Chemicals. Section 4: Health Effects: Test No.478: Rodent Dominant Lethal test, (1984).
[28] OECD Guidelines for Testing of Chemicals. Section 4: Health Effects: Test No.478: Rodent Dominant Lethal test, 2016.
[29] OECD, OECD Guidelines for Testing of Chemicals. Section 4: Health Effects: Test No.485: Mouse heritable translocation assay, (1986).
[30] D.G. Marchetti F, Yauk CL. 2019., A return to the origin of the EMGS: rejuvenating the quest for human germ cell mutagens and determining the risk to future generations., Environ Mol Mutagen. 2019 Aug 31.
[31] G.E. Johnson, L.G. Soeteman-Hernandez, B.B. Gollapudi, O.G. Bodger, K.L. Dearfield, R.H. Heflich, J.G. Hixon, D.P. Lovell, J.T. MacGregor, L.H. Pottenger, C.M. Thompson, L. Abraham, V. Thybaud, J.Y. Tanir, E. Zeiger, J. van Benthem, P.A. White, Derivation of Point of Departure (PoD) Estimates in Genetic

Toxicology Studies and Their Potential Applications in Risk Assessment, Environmental and Molecular Mutagenesis, 55 (2014) 609-623.
[32] P.A. White, G.E. Johnson, Genetic toxicology at the crossroads-from qualitative hazard evaluation to quantitative risk assessment, Mutagenesis, 31 (2016) 233-237.
[33] J.W. Wills, G.E. Johnson, S.H. Doak, L.G. Soeteman-Hernandez, W. Slob, P.A. White, Empirical analysis of BMD metrics in genetic toxicology part I: in vitro analyses to provide robust potency rankings and support MOA determinations, Mutagenesis, 31 (2016) 255-263.
[34] J.G. Heflich RH, Zeller A, Marchetti F, Douglas GR, Witt KL, Gollapudi BB, White PA. , Mutation as a toxicological endpoint for regulatory decision-making, Environ Mol Mutagen. 2019 Aug 312019.
[35] W. Slob, R.W. Setzer, Shape and steepness of toxicological dose-response relationships of continuous endpoints, Crit Rev Toxicol, 44 (2014) 270-297.
[36] W. Slob, A general theory of effect size, and its consequences for defining the benchmark response (BMR) for continuous endpoints, Crit Rev Toxicol, (2016) 1-10.
[37] A. Zeller, G. Duran-Pacheco, M. Guerard, An appraisal of critical effect sizes for the benchmark dose approach to assess dose-response relationships in genetic toxicology, Arch Toxicol, 91 (2017) 3799-3807.
[38] J.P. Bercu, S.M. Galloway, P. Parris, A. Teasdale, M. Masuda-Herrera, K. Dobo, P. Heard, M. Kenyon, J. Nicolette, E. Vock, W. Ku, J. Harvey, A. White, S. Glowienke, E.A. Martin, L. Custer, R.A. Jolly, V. Thybaud, Potential impurities in drug substances: Compound-specific toxicology limits for 20 synthetic reagents and by-products, and a class-specific toxicology limit for alkyl bromides, Regul Toxicol Pharmacol, 94 (2018) 172-182.
[39] M. Luitjen, N.S. Ball, K.L. Dearfield, B.B. Gollapudi, G.E. Johnson, F. Madia, F. Peel, S. Pfuhler, R.S. Settivari, W. ter Burg, J. van Benthem, Utility of a Next Generation Framework for Assessment of Genomic Damage: A Case Study Using the Industrial Chemical Benzene, Environ Mol Mutagen, (2019).
[40] R.S. Thomas, M.A. Philbert, S.S. Auerbach, B.A. Wetmore, M.J. Devito, I. Cote, J.C. Rowlands, M.P. Whelan, S.M. Hays, M.E. Andersen, M.E. Meek, L.W. Reiter, J.C. Lambert, H.J. Clewell, 3rd, M.L. Stephens, Q.J. Zhao, S.C. Wesselkamper, L. Flowers, E.W. Carney, T.P. Pastoor, D.D. Petersen, C.L. Yauk, A. Nong, Incorporating new technologies into toxicity testing and risk assessment: moving from 21st century vision to a data-driven framework, Toxicological sciences : an official journal of the Society of Toxicology, 136 (2013) 4-18.