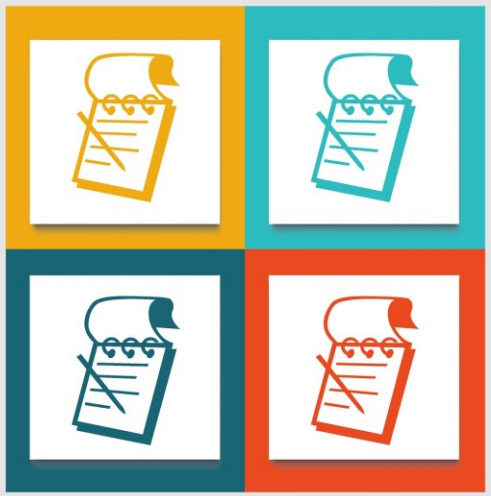
This presentation summarises the results obtained by the participants of

Workshop 3

New approaches for the characterisation of genotoxic effects based on *in vitro* and *in silico* data

It does not necessarily reflect the opinion of the BfR.



RESULTS WORKSHOP 3

New approaches for the characterisation of genotoxic effects based on *in vitro* and *in silico* data

Do the existing standard assays sufficiently cover hazard ID, or do we need additional (new) tests?

- **Sufficient but:**
  - **Nature of the compounds is often not considered in the respective strategy**
  - **MOA needed to get more information which is important for regulatory aspects**
  - **The relevance of in vitro results is often not sufficiently questioned (positive results immediately trigger in vivo follow up)**

- **The current tests are sufficient as long as a negative/positive answer is sufficient (hazard ID). If more is required, e.g. quantitative approach, the battery has to be adjusted.**

- **Adjustments can be on the current tests but also NAMs can play a role.**

2-test versus 3-test strategy; what is the best option?

- **It was clear that the performance of the standard in vitro test battery has not been sufficiently evaluated.**
  - **Results indicate that combinations of different assays are less predictive than the results of single assays.**
  - **Sensitivity increases but specificity decreases dramatically with increasing number of tests.**

- **There is no consensus for a 2- or a 3-test battery.**

- **There is agreement that a more flexible approach, depending on the chemical that is assessed, may be desirable**
  - **Try to find as much information on the compound and thereafter make a plan for testing, particularly when a quantitative approach is desirable.**

- **Converging of strategies between sectors would be beneficial but is challenging, but there may be an opportunity to do this via the EU "one substance, one assessment' approach.**

Is there still a role for the Ames test?

Jeroen Pennings, IWGT inititive:
Assessment of the predictivity of different combinations of *in vitro* genotoxicity tests on qualitative test results (yes/no outcomes) using Bayesian modelling.

Identical genotoxicity endpoints have been combined, with the aim to obtain data for three classes of *in vitro* genotoxicity tests:

1. **Bacterial mutagenicity test (Ames)**
2. **Mammalian cell gene mutation test (mammalian mutagenicity; hprt, MLA)**
3. **Mammalian *in vitro* clastogenicity test (*in vitro* MN, *in vitro* CA)**

***In vivo* genotoxicity data, and not carcinogenicity data, were used as reference point for comparison with the *in vitro* genotoxicity data.**

Is there still a role for the Ames test?
Is a standard battery of mammalian cell tests a better option?

Results:

1. **Each of the three assay types commonly used for genotoxicity testing were shown to be reasonably concordant with results from the same endpoint *in vivo*.**
2. **When using a battery of the three types of genotoxicity tests, combinations of two mammalian cell tests showed the highest predictive value for *in vivo* genotoxicity. Adding Ames test results had no impact on the prediction of in vivo genotoxicity.**
3. **Combination of the 2 mammalian cell tests had a better prediction than the Ames - micronucleus combination**

Discussion in the workgroup:

- **There was no consensus answer to the question "is there still a role for the Ames test".**
- **Nor for the suggestion that a standard battery of mammalian cell tests would be a better option**
- **A voting among the participants showed that 6 wanted to keep the Ames test and 8 wanted a standard battery of a gene mutation test and a mammalian clastogenicity test as standard option. 25 participants did not vote.**

Newly developed *in vitro* test methods

- NAMs = new approach methodologies

- Current standard *in vitro* battery gives a yes/no answer & provides limited mechanistic information
- A variety of NAMs is available and already used (2/3 workshop participants have experience with at least 1 NAM)
- Limited experience by reg authorities. EFSA has considered NAM data in submitted dossiers as supportive information, often from literature as they are not a requirement
- Different situation across regulatory frameworks and the globe
- NAMs have different status (from development to validation) and some are closer to regulatory acceptance

NAMs Role?

Consensus

- **Follow-up standard battery to identify MoA, rather than replace the it** (e.g. distinguish between direct & inderect MoA, examples of biomarker tests that provide additional mechanistic information).
- **Take an informed decision on follow-up evaluation**
  - **Crucial to refine HazID assessment**
  - **Inform on quantitative follow-up analysis**

Such discussions are already happening, but often data are not sufficient and assessors do not have the means to ask for further data in the current framework.

- „High tier“ tests (based on 3D models), if sufficiently validated, have potential to replace *in vivo* tests

Opportunities

- Apply NAMs for assessment of impurities, less regulated, more flexibility
- Develop NAMs with adequate metabolic capacity
- Learn from cosmetics sector experience & SCCS assessments (ref. Note of Guidance, 2021)

- Several NAMs have the potential to be used for quantitative assessment. Case studies would be needed.
- Promising results from transcriptomics-based signature assays, possibility of integrating signatures into whole gnome analyses and obtain additional mechanistic information

Recommendations

Transitioning from Hazard ID to Hazard characterisation (HC) approach

- Add check points after standard battery and define criteria for the need of more data leading to HC, before moving *in vivo*

- Guidance on weight of evidence to consider more information to help registrants and assessors to take more informed decisions (consider integrated approaches for testing and assessment-IATAs)

In silico

- Approaches have improved and matured due to extensive use in the last 10 years in the context of, e.g. ICH M7 for pharmaceutical impurities, or read across within REACH
- For (Q)SAR models, quality of the models is strongly dependent on data density, and quality. Ames/reactivity-based models are very mature but for other apical endpoints they need to be improved
- Models that base on in vivo data would be highly desirable – concerted effort across industries could fuel this. It was suggested that a model could be built on overall in vivo genotoxiciity which would increase numbers. Further improvements are expected though improved AI technologies.
- Regulatory uptake is variable and data are typically used for decision making only for contaminants/impurities and metabolites (QSAR) but is broader for read across. This requires stringent approaches, e.g, RAAF in REACH.

Quantitative assessment

- Basing risk decisions on quantitative modeling on in vitro datasets is a step change and we recognize that regulatory acceptance of this concept may need a paradigm change, away from a hazard focus.
- A lot of progress has been made in the last few year with IVIVE modeling from genotoxicity data, pioneered by the HESI GTTC with leadership from Health Canada. With constant refinement the modeling has been getting closer and closer to in vivo estimates and we expect continued improvements.
- The WG agreed that, while promising, there are still many open questions that need to be addressed and further standardization should be targeted for both the biology and the modeling aspects of IVIVE. For modeling it is expected that it will profit from advanced in PBPK modelling which is a broader need across toxicity endpoints.
- While IVIVE modeling can be done from any genotoxicity dataset that has enough datapoints we suggest that the choice of the best cell type and genotoxicity endpoint should be informed by existing knowledge about the respective substance, e.g, its metabolic fate, genotoxicity MoA, etc. Standard in vitro tests with the possible exception of the Ames test, as well as NAMs could be used.

Quantitative assessment

- Several different cell types should be used. Ideal datasets would have more dose levels with tighter spacing,
- Case studies will be very helpful to further IVIVE approaches and it was suggested that good examples would be in the contaminants/impurities space, e.g., PAHs for which extensive datasets are available to validate against, and nitrosamines which are currently under intensive investigation.
- It was agreed that the use of potency as a guiding principle for substance evaluations could be a 'low hanging fruit' for chemicals withing a substance class, e.g, pyrrolizidine alkaloids, PAHs)
- Quantitative AOPs were discussed as an alternative concept for quantitative assessment. The WG agreed that these at a proof of concept stage and that it is premature to assess their utility.