Safety & Environmental Assurance Centre

Workshop on the Implementation of Next Generation Risk Assessment (NGRA) for Systemic Toxicity

Authors: Mabel Cotter[1], Maria Baltazar[1], Alistair Middleton[1], Sophie Cable[1], Georgia Reynolds[1], Predrag Kukic[1], Gavin Maxwell[1], Paul Russell[1], Matthew Dent[1], Steve Gutsell[1], Ans Punt[2], Bob Van de Water[3], Stephen Ferguson[4], Josh Harrill[5], Richard Judson[5], Imran Shah[5], Carl Westmoreland[1]

[1]SEAC, Unilever, Colworth Science Park, Sharnbrook, Bedford MK44 1LQ, UK; [2] Wageningen Food Safety Research, Wageningen University & Research, Akkermaalsbos 2, 6708 WB Wageningen, Netherlands; [3] Division of Drug Discovery and Safety, Leiden University, Einsteinweg 55, 2333 CC Leiden, Netherlands; [4] National Toxicology Program, National Institute of Environmental Health Sciences, P.O. Box 12233, Durham, NC USA 27709; [5] U.S. Environmental Protection Agency, 109 TW Alexander Drive, Research Triangle Park, North Carolina 27709

Introduction

Next Generation Risk Assessment (NGRA) uses non-animal new approach methodologies (NAMs) as part of an exposure-led, hypothesis-driven risk assessment to ensure safety of consumers. Unilever's Safety and Environmental Assurance Centre (SEAC) has recently carried out a hypothetical NGRA case study on 0.1% coumarin in shampoo, face cream and body lotion. This has provided practical experience in applying NAMs as part of an NGRA framework (shown in Fig.1), as proposed by the International Cooperation on Cosmetics Regulation (ICCR)[1] and the SEURAT-1 project[2]. A two-day workshop was held in October 2019, which focussed on this systemic toxicity risk assessment case study and the underpinning mechanistic science. There were over 50 participants, of which half were from international scientific partners*.

During the first day of the workshop, the coumarin case study was presented and the proposed approach was reviewed by participants. Breakout groups focused on the scientific methods and techniques presented, and key themes were identified. From these initial breakout sessions, six key areas were discussed in detail on day two of the workshop and are summarised below.

Effective Communication & Framework Development | Incorporating Metabolism | Choosing the Right Point of Departure (PoD) | When is Enough, Enough? | Making the Most of Benchmarking | Optimal Assay Design

[1] Dent et al., 2018. Computational Toxicology 7, 20–26.
[2] Berggren et al, 2017 Computational Tocology 4 31-44
*Participant organisations: Chinese Academy of Military and Medical Sciences, Beijing Proteomic Research Centre, Cyprotex, XCellR8, UK NC3Rs, National Toxicology Program (NTP, US Department of Health and Human Services), US Environmental Protection Agency (EPA), University of Nanjing, University of Cambridge, Emory University, Brown University , University of Leiden & University of Wageningen

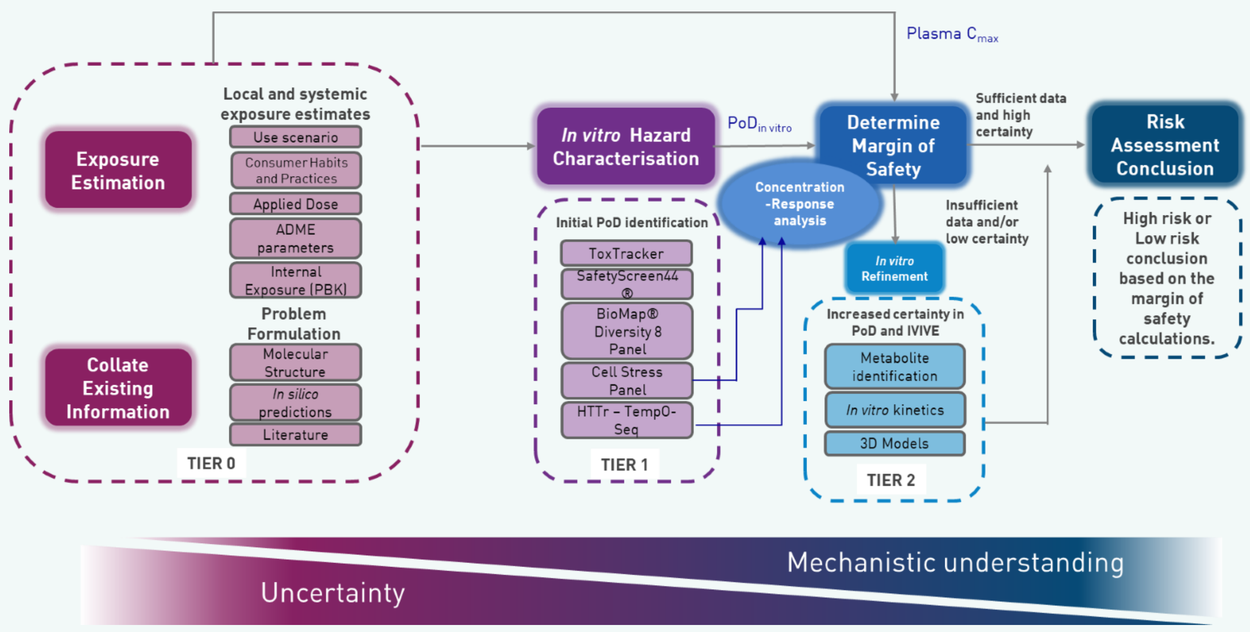

Fig.1 Framework for ab initio next generation risk assessment as presented at Centre for Alternatives to Animal Testing (CAAT) on 11th December 2019

Effective Communication & Framework Development

Discussion focused on the need for a more general and universal framework for approaching an *ab initio* NGRA, with clear decision criteria for progression through framework tiers and 'exit' points described. In addition, communication of the inherent uncertainties and the distributions generated for points of departure and margins of safety are essential in the context of justifying a risk assessment decision based on NAMs.

- There is a lack of goal setting and decision gates within the framework (Fig. 1) – each tier should have an explicit goal e.g. hypothesis formulation.
- Decision points should be defined throughout, for example at the end of each tier, where expert analysis takes place and the strategy for the subsequent tier is defined based upon the available evidence.
- There are likely to be a variety of users and audiences for an NGRA framework and the decisions made within. Different requirements, and therefore tools, may be present at each tier depending on the user or audience.

When is Enough, Enough?

Confidently being able to make decisions on the data available is a key part of safety assessment, but knowing when we have enough data to be confident can be difficult. This group discussed points including the importance of confident Physiologically Based Kinetic (PBK) models and exposure estimates incorporating evaluation of specific software packages and characterisation of the uncertainty in the input parameters and the important challenge of characterising the biological coverage of the assays.

- An increase in the number of case study chemicals studied through the use of NAMs is required in order to systematically evaluate and quantify the associated uncertainties. Specifically, assays for determining relevant MIEs along with the broader cell stress and transcriptomic approaches require additional scrutiny to understand their utility across a broader chemical space of known responders and those with limited biological effects, as well as across multiple laboratories.
- This will require understanding the correlation with respective human *in vivo* studies, and hence the choice of chemicals to study should be carefully thought out to ensure examples are impactful without creating a bias through 'cherry picking'.
- Robust intra-assay variability measures will be required to better characterise the confidence with which they can be applied. Increasing confidence in the *in vitro* assay data is equally critical for ensuring that the computational models and predications which rely on such experimental data can be of real benefit within a tiered risk assessment framework.

Incorporating Metabolism

This breakout group focused on forming pragmatic decision criteria that would trigger an in-depth assessment of the metabolism and toxicity of the metabolites of a chemical. The usefulness of *in silico* tools for the prediction of metabolites was emphasised and in vitro assays described in the case study e.g. HepatoPac, were discussed with regards to their metabolic competency. Furthermore, the appropriate use of metabolism ID work was discussed.

- The Coumarin case study highlighted how metabolism and reactive metabolites can be a major source of uncertainty and significantly impact the risk assessment decisions made for a chemical when using NAMs.
- As with risk assessment of a parent chemical, it was agreed that *in silico* tools were the best starting point for determining a chemical's potential for metabolism into a reactive metabolite(s). Discussion touched on the most appropriate *in silico* tools, which included METEOR and ADMET-Predictor.
- After determining predictions for parent metabolites *in silico*, the next logical step is to calculate *in vitro* parameters including clearance. An initial experimental phase using primary hepatocytes should provide sufficient information to populate a PBK model for the parent compound, and gauge the relative importance of predicted metabolites to systemic circulation. It was noted that further work could be completed in the HepatoPac model, for example.

Making the Most of Benchmarking

Use of appropriate benchmarking chemicals was discussed for both assay evaluation and for determining a MoS derived from NAM-based points of departure. A reference benchmark database of compounds of varying MoAs and *in vitro* data from the currently available assays was mentioned as a useful tool for future assessments.

- Historically, benchmarking is how we build confidence in our interpretations of both individual assays and techniques and at the risk assessment level.
- For assays, benchmarking is frequently used and is part of good practice as it confirms that the assay is capable of differentiating between known examples of potent, weak and negative responses. It is also possible to use this benchmarking approach to investigate the relationship between the results of an assay and the *in vivo* biological response, i.e. what is the false positive/false negative ratio for this assay.
- At the risk assessment level, it allows a comparison of the unknown to the known and relies on having reference information for chemicals that you know something about e.g. their hazard/toxicity profile in humans. This gives context to a safety decision e.g this chemical at this exposure should have a higher, lower or comparable risk to what is seen for chemical X in use. A large database of benchmarks that span potency levels and effects is required to be able to confidently find a comparator.

Choosing the Right Point of Departure (PoD)

The assays used in this case study all generated dose-response data *in vitro* from which a PoD is derived to calculate margins of safety (MoS). A key topic in this breakout session surrounded the differentiation of adverse and adaptive effects as measured in the *in vitro* assays. The potential use of new approaches such as phosphoproteomics and changes in the epigenome were considered.

- The strategy adopted in the Coumarin case study was to select the lowest PoD that was measured among the different assays, this being protective of human health by ensuring that substance exposures are below levels expected to trigger a biological response.
- Refinements to this approach were discussed, including *in vitro* testing strategies to 1) determine whether a substance has a specific toxicological mode of action (MoA), or acted through non-specific effects and 2) differentiate between adaptive and adverse responses if the MoA was non-specific effect.
- One approach posited by the group was to perform transcriptomics using a broad range of cell types and complement this with specific biological target screens. A key element of this strategy involved using cell types with sufficiently broad coverage of receptors, enzymes and other potential targets of toxicological concern, and thereby ensuring the data are sufficiently protective. If no MoA could be established from these data, then it could be expected that the chemical acts through a non-specific MoA, which could be evaluated further using cellular stress panel or phenotypic screening assays.

Optimal Assay Design

Three themes emerged during the discussion of assay design: technical aspects of assay design, *in vitro* exposure considerations and *in vitro* to *in vivo* extrapolation (IVIVE), and the impact of assay decision making.

Technical aspects of assay design:
- Ensure that experimental design addresses potential sources of biases such as plate bias and pipetting.
- Experiment with different plate layouts to decide which one is more suitable and bias can be modelled.
- Ensure sufficient biological repeats to assess the robustness of the assays

***In vitro* exposure considerations and IVIVE:**
- *In vitro* disposition in 3D cultures (kinetics of penetration).
- *In vitro* dosing: nominal vs free and IVIVE (Cmax/AUC) – what is the impact on decision? How reliable are the existent models that predict free concentration?
- Repeat dosing using complex 3D cultures as a possible mechanisms to assess the impact of metabolites.
- Investigate role of transporters in *in vitro* kinetics and impact on risk assessment.

Impact of assay design on decision-making: tiered approach
- Importance of problem formulation and hypothesis generation.
- Selection of assays in a tiered approach: when do we use 3D tissues versus 2D tissues.
- Investigate population variability using cells from multiple donors
- Investigate redundancy in assays.

Summary

Overall, the presented NGRA framework for the coumarin case study was well received, and discussion on areas for development was encouraging. This example illustrates how case studies are an impactful method for communicating the current capabilities of NGRA, ultimately driving conversations that will lead to change in the understanding and acceptance of non-animal approaches to safety assessment globally. Therefore, more examples of NGRA in various exposure scenarios are crucial.