Food for Thought ...

Ready for Regulatory Use: NAMs and NGRA for Chemical Safety Assurance

Paul L. Carmichael[1,2], Maria T. Baltazar[1], Sophie Cable[1], Stella Cochrane[1], Matthew Dent[1], Hequn Li[1], Alistair Middleton[1], Iris Muller[1], Georgia Reynolds[1], Carl Westmoreland[1] and Andrew White[1]

[1]Safety & Environmental Assurance Centre (SEAC), Unilever, Sharnbrook, Bedfordshire, UK; [2]Toxicology, Wageningen University & Research, Wageningen, The Netherlands

Abstract

New approach methodologies (NAMs) that do not use experimental animals are, in certain settings, entirely appropriate for assuring the safety of chemical ingredients, although regulatory adoption has been slow. In this opinion article we discuss how scientific advances that utilize NAMs to certify systemic safety are available now and merit broader acceptance within the framework of next generation risk assessments (NGRA).

1 Introduction to the advancements

The science of assuring the safe use of chemicals in products for the consumers who use them and the people in factories who work with them has advanced a great deal since the 1950s when the only tools for determining the safety of chemicals were experimental animals. Over the intervening decades, an inexorable scientific progression has taken place towards the elimination of animal testing in chemical safety assessments and the introduction of advances from biomedical sciences to focus on better protecting human and environmental health.

During the 1980s and 1990s, considerable investment into "alternatives to animal testing" resulted in the development of several new tests for topical toxicity that could reliably predict the results of animal tests. A regulatory framework then grew around the animal tests and these newer "animal replacements". The pace of change has, however, recently quickened as advanced methodologies that can characterize fundamental biological alterations have been evaluated and implemented in toxicology in a concerted effort to assure safety in a more human-focused and more exposure-relevant way. A key landmark precipitating the change was the publication of the principles outlined in *Toxicity Testing in the 21st Century: A Vision and a Strategy* from the National Academy of Sciences and National Research Council of the USA (NRC, 2007). The book opened the door to a new paradigm, whereby advances that have been made in systems biology and associated computational modelling could be used to transform toxicity testing from its historical animal test base to one founded on *in vitro* and *in silico* methods that evaluate pathway changes in cells of human origin.

Today, many industry, academic, and regulatory toxicologists involved in this evolution face a decision point: whether to adopt these new ways, evaluate their use and seek to improve them or to remain wedded to using traditional animal tests (with the associated ethical and scientific issues they pose in modern society). This decision on whether to adopt or not occurs against the backdrop of generally slow uptake from regulatory agencies in many geographies that make it difficult to allow such a transition.

Indeed, the new approach methodologies (NAMs[1]) that are increasingly used to assure consumer safety of chemicals in cosmetics (Bernauer, et al., 2021; Dent et al., 2018, 2021) are generally not yet recognized as a valid route to provide the regulatory safety data in many other cases, e.g., to fulfil REACH information requirements for occupational and environmental safety of the same chemicals in the EU.

2 The unexploited science

This is at a time when enormous investment in non-animal approaches has occurred – it is estimated that almost a billion euros have been spent in the last decade by the EU Commission, inter-

Received April 28, 2022;
Epub July 4, 2022;

ALTEX 39(3), 359-366. doi:10.14573/altex.2204281

Correspondence: Paul L. Carmichael, PhD
Safety & Environmental Assurance Centre (SEAC)
Unilever, Sharnbrook, Bedfordshire, MK44 1LQ, UK
(paul.carmichael@unilever.com)

[1] https://www.epa.gov/sites/default/files/2018-06/documents/epa_alt_strat_plan_6-20-18_clean_final.pdf

national trade associations, and by personal care product companies responding to the challenges of the animal testing bans of the EU Cosmetics Regulation (Donnellan, 2020).

Research projects funded by the European Commission, and participated in by EURL-ECVAM, included integrated projects within the 6th and 7th Framework Programmes (FP6 and FP7) such as AcuteTox, carcinoGENOMICS, COMICS, ESNATS, INVITROHEART, Predictomics, ReProTect, TOXDROP, and VITROCELLOMICS[2]. It is worth noting that during FP7, some €200 million was dedicated to animal-free toxicology projects, supported mainly from the "Health" theme. As part of this effort, six large projects were co-financed for a total of €140 million by Cosmetics Europe and the European Federation of Pharmaceutical Industries and Associations within the context of public-private partnerships (SEURAT-1 and the Innovative Medicines Initiative). Furthermore, the Horizon 2020 Framework Programme for Research and Innovation was the biggest ever of its kind, covering seven years to 2020. The topic in the Work Programme 2015 for Health had a budget of €30 million on animal-free predictive safety testing of chemical substances.

Yet, despite the considerable scientific output from initiatives such as EUToxRisk and the regulatory engagement with these programs, it is unclear how the results from much of this remarkable research investment have been evaluated for regulatory use. Science from the FP6 project Sens-it-iv[3] and other research activities developed into non-animal approaches for assessments of skin sensitization and formed the basis of progress towards regulatory acceptance[4]. However, how the results of all the research funded into complex systemic toxicity have progressed into regulation, is much less obvious. Perhaps, the research has become a self-sustaining entity in its own right, moving from FP6 to FP7 to SEURAT-1 to EUToxRisk[5] to RiskHunt*3R*[6] with no clear plan for application or acceptance in the regulatory world of the science that addresses complex systemic human health effects. This situation may be compounded further in the future, depending on how the EU's Chemical Strategy on Sustainability (Fentem et al., 2021) chooses to embrace modern safety science.

3 Global regulatory activity on implementation of NAMs

In examining where we find ourselves today, one can only conclude that the bravery of the European Seventh Amendment to the Cosmetics Directive has perhaps left us short – the law told us what we must not do, but it didn't tell us what we should do.

And yet, in North America we see the US Environmental Protection Agency (EPA) and Health Canada actively embracing bespoke use of NAMs in their chemical safety assessments and the EPA aggressively reducing the use of animals in toxicity testing[7,8]. The agencies are growing their own confidence in the new approaches by investing and pursuing the science, not least through the regulator and agency-only APCRA activity (Accelerating the Pace of Chemical Risk Assessment), an international collaboration that brought together governmental entities (including EPA, Health Canada, ECHA, EFSA, and the JRC) to engage in the development of new hazard, exposure and risk assessment methods and approaches for their own chemical evaluation activities[9].

However, Europe seems to be much more reticent in grasping the NAM opportunity presented by the work of APCRA. This is particularly pertinent since the publication of the remarkable work led by Katie Paul Friedman (Paul Friedman et al., 2020) that clearly established the safe and conservative value of non-animal NAMs compared to traditional animal tests. The EPA, through their NAMs Workplan[10], have sought to support the development of the science and computational tools but in the EU, despite the best efforts of the JRC (Zuang et al., 2021), there is currently no roadmap for transition to non-animal methodologies.

The Next Generation Blueprint of Toxicology at the US EPA (Thomas et al., 2019) has clearly outlined how the multiple NAMs in the ToxCast platform can be transitioned to high-content, broad-coverage *in vitro* assays in multiple cell types with and without metabolic competence. Inspired by the Blueprint of Toxicology, some have shown how such tools, e.g., high-throughput transcriptomics, stress pathway characterization, and pharmacological profiling in multiple cell types, can likely cover the relevant biological pathways of potential toxicological concern. Possible perturbations of these pathways can be used to set a NAM point of departure (POD) that can be compared with relevant exposures of a given chemical ingredient in order to craft a next generation risk assessment (NGRA) decision (see details below) (Baltazar et al., 2020; Middleton et al., 2022).

4 Protection not prediction

To transition to this new way of assuring chemical safety requires the toxicologist to be open to the computational methods that can predict *human exposure* and the NAMs that seek to characterize the potential *bioactivity* of new and existing chemicals. In

2 https://cordis.europa.eu/projects/en

3 https://cordis.europa.eu/project/id/18681

4 https://echa.europa.eu/-/non-animal-methods-now-a-default-for-skin-sensitisation-submit-correct-information

5 www.eutoxrisk.eu

6 www.risk-hunt3r.eu

7 https://www.canada.ca/content/dam/eccc/documents/pdf/pded/bioactivity-exposure-ratio/Science-approach-document-bioactivity-exposure-ratio.pdf

8 EPA Seeks 'Meaningful' Goal As It Drops Plan To End Animal Testing By 2035 | InsideEPA.com

9 Accelerating the Pace of Chemical Risk Assessment (APCRA) | US EPA

10 EPA New Approach Methods Work Plan: Reducing Use of Vertebrate Animals in Chemical Testing | US EPA

many ways, animals have been used as *in vivo* bioactivity measures for many decades; a point of departure (POD) can be set on nothing more than factors such as decreased weight gain or rodent-specific histological changes. So why is it so hard to contemplate a POD set on more conservative, broad biological coverage NAMs? Comparing the NAM-PODs with estimated tissue exposure levels can provide a margin of safety determination or bioactivity-exposure ratio (BER) that allows us to gauge whether the use of that chemical in a product is safe, i.e., the new tools and approaches provide confidence of protection not prediction. Protection from harm of workers, consumers, and the environment; not prediction of apical endpoint pathology in rodents.

The aforementioned work of Paul Friedman et al. (2020) has been the "game changer" for many in this regard. The study encompassed a large and diverse enough group of chemicals to clearly establish that the NAMs in question (the broad coverage ToxCast *in vitro* tests and phenotypic profiling) could be sufficiently conservative, i.e., lower than the PODs set on the traditional animal data (with some notable and defined exceptions e.g., organophosphate insecticides). Comparison of these more conservative PODs with the specific chemical exposures encountered provides the unlock towards the novel risk assessment principles that are now identified as NGRA and have a clear practical application in certain settings. The old "gold standard" of rodent biology and high-dose, chemical-induced apical endpoint pathology can be replaced by broad biological coverage *in vitro* tools that establish a POD on the basis of bioactivity, and comparison of those PODs with exposure measures (or predictions, e.g., using physiological based kinetic modelling, PBK) can establish a pragmatic NGRA decision that allows for consumer and worker protection, which is not attempting to predict levels of toxicity that would never occur at the levels of real human exposure.

Indeed, casting the net wide with early-tier, broad coverage bioactivity NAMs can establish safe exposure levels of chemicals without animals but also without having to resort to a plethora of (currently impractical) adverse outcome pathway (AOP)-driven integrated approaches to testing and assessment (IATAs)[11]. It is worth contemplating why that is the case: There are 78 major human organs; let's say there are five different ways in which chemicals could be toxic to each one (an underestimate); and let's say we need five key events (including a molecular initiating event) measured across each IATA with new *in vitro* tests. That's around 2000 assays conducted at just one dose and at one time point for complete human AOP-driven biological coverage. Planning for a regulatory world with that many unvalidated *in vitro* assays is certainly impractical. The only way forward is a first tier of broad coverage NAMs, and then higher tier (AOP-driven, IATA) assays for specific pathways when margins of safety are not established at the lower tier.

5 Using NAMs, not rodents

The specific NAMs for characterizing bioactivity PODs will continue to evolve and improve. For example, the use of the ToxCast assays, exemplified in Paul Friedman et al. (2020), are, as previously mentioned, currently transitioning at the EPA towards the methods indicated in the EPA Blueprint of Toxicology (Thomas et al., 2019) that give broader biological coverage across fewer technologies. Those specific techniques are high-throughout transcriptomics (HTTr) (Harrill et al., 2019, 2021) and high-throughput phenotypic profiling (HTPP) (Nyffeler et al., 2020).

Work under a Unilever and EPA joint co-operative research and development agreement (CRADA) has been facilitating the experiments needed to establish that same level of conservatism from HTTr and HTPP as found with the ToxCast assays, but experiments are also being performed to increase the confidence across many more case study chemicals and to determine whether additional NAMs and human (and environmentally-relevant) cell types are required for greater safety assurance[12].

We recently showed how bioactivity PODs, determined using three such core broad biological coverage NAMs, together with PBK models of consumer exposure, could be used to arrive at an NGRA decision for a case study chemical in a cosmetic product (Baltazar et al., 2020). Those NAMs were HTTr, a recently developed cell stress panel (CSP) (Hatherell et al., 2020), and *in vitro* pharmacological profiling (IPP) (Middleton et al., 2022). Additional NAMs (ToxTracker[13] and BioMap[14]) were also used to evaluate specific targets of concern highlighted by cheminformatics tools. This exemplified how a core framework of principal broad biological coverage NAMs could be supplemented with other targeted NAMs for safety concerns such as immunotoxicity or genetic toxicity, or even extended to address cheminformatic flags of concern (or specific regulatory requests) for developmental and reproductive toxicity, with accompanying additional and relevant exposure predictions by PBK for the developing fetus (Rajagopal et al., 2022).

Baltazar et al. (2020) showed how, in principle, an early, low-tier toolbox could be sufficient for assessing safety in a large proportion of consumer safety assessments. Safety decisions could be based on the BER, which allows risk assessors to gauge whether a given chemical exposure scenario is safe or not. However, to be confident that this approach would, in general, lead to robust and reproducible safety decisions that are protective of human health, the overall approach has been undergoing a systematic evaluation (Middleton et al., 2022).

Notably, as with risk assessments based on animal data, there are uncertainties associated with each tool within the NAM toolbox. For example, how can we be confident that the assays have enough biological coverage? Or that the PBK models provide

11 https://www.oecd.org/chemicalsafety/risk-assessment/iata-integrated-approaches-to-testing-and-assessment.htm

12 EPA and Unilever Announce Major Research Collaboration to Advance Non-animal Approaches for Chemical Risk Assessment | US EPA

13 https://toxys.com/toxtracker/

14 https://www.eurofinsdiscoveryservices.com/services/phenotypic-assays/biomap

a reasonable estimate of the internal exposure level? One way to address this issue has been to use a benchmarking-based approach, whereby the toolbox is used to estimate the BER for a wide range of curated chemical-exposure scenarios, some of which are high-risk in the sense that there is documented evidence of them causing adverse health effects in humans (e.g., some drugs), and some of them with a documented safe use in consumer goods and foods, etc. The BERs are then estimated in the same way as they would be as part of an *ab initio* risk assessment (Berggren et al., 2017), and any of the relative shortcomings of the tools can be ascertained from the evaluation. Appropriate Bayesian statistical modelling approaches are used to quantify the relevant uncertainties contributing to the BER estimate, allowing risk assessors to have an indication of the relative *certainty* they can attach to various aspects of the safety decision (Middleton et al., 2022).

Many have now begun to adopt these principles, and examples of the use of NAMs in NGRA decisions are increasingly seen in the literature and considered in Notes of Guidance (Bernauer et al., 2021). *In vitro* bioactivity assays are regularly employed in the early tier screening phases of agricultural and pharmaceutical chemicals due to their ability to differentiate potential specific hazards. Indeed, Bowes et al. (2012) published a review of the most common targets associated with adverse effect-based drug attrition that has led to the development of various *in vitro* pharmacological profiling panels used widely today.

Examples of the use of NAMs in hazard identification and screening include assessment of hepatoxicity and genotoxicity potential of new anti-inflammatory actives using a Cellomics approach (Rampa et al., 2021), which concluded a potential risk for steatosis and genotoxicity, and a review of per- and polyfluoroalkyl substances (PFAS), looking at cytotoxicity, biotransformation data, and cellular responses at the gene level in placental cell lines, determined the potential for developmental toxicity (Solan and Lavado, 2021). Another recently published case study has been that of phenoxyethanol in cosmetics; this study utilized both specific and non-specific assays, including IPP, to inform on the bioactivity of phenoxyethanol and its metabolites. This example also emphasized the need for a tiered and iterative approach, with uncertainties being adequately documented to allow for an assessment of confidence in the interpretation of a calculated margin of exposure/BER.[11]

Whilst the coumarin and phenoxyethanol case studies were constrained by the limits of assuming no availability of existent data, NAMs can also be used in the context of existing toxicological information to fill gaps in risk assessments or to support other alternative approaches such as read-across. Read-across relies on the demonstration of equivalence between chemicals or chemical groups with respect to their toxicokinetic and toxicodynamic properties. However, read-across cases frequently fail regulatory acceptance because the available *in vitro* and *in silico* tools are not considered to adequately demonstrate kinetic and dynamic equivalence.

Research programs such as EU-ToxRisk have been developing case studies that use many available NAMs to demonstrate that compounds can be grouped based on *in vitro* and *in silico* data (Escher et al., 2022). A fully worked example of using NAM-based read-across approaches to conclude on safety was also published looking at the risk presented by caffeine to consumers (Bury et al., 2021), and a recent publication addressed the use of high-throughput data to facilitate grouping of food-related substances to allow gap-filling through read-across of available *in vitro* data from a variety of targeted and broad coverage assays. Punt et al. (2020) reviewed the data generated through the ToxCast and Tox21 programs for over 500 food-related substances and concluded that the data could be useful for regulatory submissions in addressing gaps in read-across assessments for additives and nutrients.

Other aspects that case studies have focused on include addressing some of the uncertainties that remain when looking to extrapolate PODs from *in vitro* data, such as the use of nominal test concentrations for dose response analysis and whether this is an accurate dose metric for all substances. Various groups have reviewed potentially difficult-to-test materials that would perhaps not normally be suitable for testing under standard *in vitro* conditions and have considered ways in which this can be accounted for when interpreting the data. These have included volatile phenols, highly oxidizable and reactive compounds, and nanomaterials amongst others (Tolosa et al., 2021; Drasler et al., 2017; Proença et al., 2021) and have highlighted the need to increase the understanding of identity and quantity of a compound exposed in *in vitro* systems, accurately interpreting the dose at which an effect occurs. The documentation and quantification of such uncertainties will be key to increasing regulatory acceptance of NAMs.

A recent example of the application of NAMs in an actual regulatory decision was the EPA assessment of chlorothalonil, which used a new approach to address the call for an unethical and non-feasible 90-day inhalation study of an irritant[15]. The approach submitted PODs derived from a human airway cell system in conjunction with advanced computational fluid dynamic modelling of the deposition of particles within the respiratory tract and demonstrated the lack of activity at exposures relevant to residential and occupational exposures. The opinion detailed the remaining uncertainties in the derivation of safety factors for the assessment and concluded that the quality of the data provided, the nature of the chemical itself, and the relevance of the test systems can indeed address many typical uncertainties, and therefore a margin of exposure (MOE) > 3 was deemed appropriate for the inhalation endpoint. The report highlighted the increased utility of using human exposure modelling and human cell-based assays where species differences exist in physiology and functionality of the respiratory tract between humans and the traditional rat models.

Such regulatory decisions, based on NAMs, can hopefully be confidently made in the future through the generation of more data to allow comparisons with chemicals of known human tox-

[15] https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=OPP&dirEntryID=347130

icity, which can aid interpretation and contextualization. These larger scale benchmarking approaches, collating information on reference compounds to inform safety decisions, have already been published in the context of skin allergy assessment and in the prediction of drug-induced liver injury (DILI) in candidate drugs (Reynolds et al., 2021; Williams et al., 2020). The utility of a benchmarking approach was also demonstrated in the context of interpreting specific activity assays such as the investigation of androgenic activity (Dent et al., 2019).

6 Making decisions: The NGRA approach for systemic safety

Next generation risk assessment (NGRA) is defined as an exposure-led, hypothesis-driven risk assessment approach that integrates NAMs to assure safety without the use of animal testing. It is based on the concept that NAMs can be used to identify exposure levels that are, as outlined above, protective of potential adverse health effects in humans in the context of exposure in either a consumer or occupational setting. The overarching premise of NGRA is that if *in vitro* bioactivity is not seen in a suitable test panel at human-relevant concentrations, then the risk of adverse effects is low. As in any safety assessment, successive refinements are possible to increase certainty in the outcome.

As a starting point, exposure assessments (route and level of human exposure) and *in silico* tools are used to generate predictions on hazard alerts (flags for concern) and physicochemical properties. Tools used in this initial phase can include ToxTree, OECD Toolbox, Derek Nexus, VITIC, Meteor Nexus, TIMES, and Molecular Initiating Events (MIE) ATLAS (Allen et al., 2014). Assuming methods of exposure-based waiving cannot be applied (e.g., the threshold for toxicological concern (TTC) is exceeded (Yang et al., 2017), nor can simple read-across methods be used, and all topical toxicity endpoints and potential genotoxicity have been covered, as appropriate, by "non-animal" OECD assays (skin allergy (Reynolds et al., 2021), skin and eye irritation, skin mutagenicity, phototoxicity), then the next phase involves determining a BER, which, as described above, quantifies the differences between relevant internal exposure levels (e.g., C_{max}) in humans (given the use case of the chemical) and the concentration required to trigger bioactivity in a broad range of *in vitro* assays in terms of PODs.

As mentioned previously, bioactivity assays can currently include the CSP (Hatherell et al., 2020), HTTr (Harrill et al., 2021), IPP (e.g.,[16]), and HTPP (Nyffeler et al., 2020). All assays involve generating concentration-response data, and PODs can be estimated from the data using different approaches such as BMDexpress (Phillips et al., 2018), Gaussian process regression (Reynolds et al., 2020), or TCPL2.0. (Harrill et al., 2021). It is worth reemphasizing that the various PODs derived from these assays do not necessarily determine a level at which adverse health effects would occur; rather, they reflect perturbations observed in *in vitro* assays covering a broad biological range of possible biochemical and cellular targets that may form the basis of events in an AOP but are not indicative on their own of an adverse health effect. It is expected that these initial biological perturbations will occur at lower concentrations than the concentrations at which downstream adverse health effects manifest following longer term *in vivo* exposures.

The effect of metabolism on the bioactivity of the compound (e.g., through the formation of toxicologically relevant metabolites) is most effectively addressed using metabolically proficient cell models to form metabolites *in situ* to ensure transient or reactive metabolites reach the sites of concern. Alternatively, *in vitro* assays can be supplemented with sub-cellular fractions (e.g., liver microsomes, S9 or other advanced methods (Deisenroth et al., 2020; Hopperstad et al., 2022) to metabolize the test chemical within the assay or predicted metabolites can be produced externally (by chemical or biological means) and added at dosing.

The internal human exposures are estimated using a physiologically based kinetic (PBK) model for the chemical of interest for a given exposure scenario based on the habits and practices of the relevant populations (Hall et al., 2007; Bernauer et al., 2021). The physiological structure of PBK models offers a scientifically sound framework for integrating information on absorption, distribution, metabolism, and excretion (ADME) to predict the kinetics of chemicals or metabolite(s) in plasma or target tissues of the exposed human.

To link to DART endpoints, a pregnancy-PBK model may need to be built to predict PK behavior (e.g., internal maternal/fetal plasma/tissue concentrations) of the chemical of interest, describing the potential impact of anatomical and physiological changes during pregnancy on the chemical's PK profile. In general, however, PBK tools used in NGRA can include commercial software, e.g., Gastroplus®[17] and Simcyp™[18], or publicly available platforms, e.g., PK-Sim® and MoBi®[19] and HTTK (Pearce et al., 2017).

Physicochemical and kinetic parameter values as inputs in PBK models may be obtained using other NAMs, i.e., *in silico* QSAR approaches or *in vitro* techniques, e.g., *ex vivo* skin penetration assays for dermal absorption, Caco-2 permeability assay for intestinal absorption, hepatocyte assays or microsome assays for metabolism and clearance, ultrafiltration assays or rapid equilibrium dialysis assays for plasma protein binding.

A framework has been developed for PBK models in NGRA, with NAMs to estimate internal exposure. It can be split into levels, with increasing complexity and increasing extents of refinement, by moving from parameters derived solely from *in silico* predictions to *in vitro* determined parameters, or even to parame-

16 https://www.eurofinsdiscoveryservices.com/services/safety-and-efficacy/safety-pharmacology/safety-panels

17 https://www.simulations-plus.com/software/gastroplus/

18 https://www.certara.com/software/simcyp-pbpk/

19 http://www.systems-biology.com/products/pk-sim/

ters calibrated using human PK data. At each level the predicted systemic exposure can be compared with *in vitro* toxicity POD values for BER derivation for decision-making and may need to progress to the next level of refinement if there is insufficient certainty in the PBK model output for a safety decision to be made (Moxon et al., 2020; Li et al., 2022; Punt et al., 2022).

Once the PBK internal estimates and PODs are estimated as described above, the BER can be calculated. A large BER means it is unlikely that internal exposure levels are high enough to trigger any bioactivity and hence give rise to any adverse health effects; the chemical can therefore be considered safe in that product-use scenario at the specific concentration of inclusion. A small BER may indicate a need to reject a chemical as potentially unsafe or, in some circumstances, generate a more refined safety assessment using higher-tier tools. In these situations, a more physiological system, such as an appropriate and targeted microphysiological system (MPS) (Rusyn et al., 2022), could be employed to further explore the mode of action (MoA) of a compound (e.g., establish POD based on a specific pathway (Dent et al., 2019) or adverse effect using a target organ MPS) or to refine the exposure (e.g., chronic low dose exposure or to link multiple compartment models).

Specific toxicological endpoints of concern, such as developmental and reproductive toxicity or systemic immunotoxicity can be also addressed by supplementing the framework with additional specific assays, for example, using cells that enable biological coverage of germ-line lineages in differentiating iPSC, as in Reprotracker®[20] and devTOXqp™[21] *in vitro* assays for DART (Rajagopal et al., 2022). Regarding immunotoxicity, due to the complexity and multi-component nature of the immune system, several additional assays may be required to cover immunotoxicity, and a tiered approach has been proposed for their application (Hartung and Corsini, 2013). Such an approach would begin with an assessment of potential direct immunotoxicity by investigating myelotoxicity, then lymphotoxicity, before progressing to targeted functional assays to investigate the mechanism(s) involved and to potentially derive a POD. Myelotoxicity and lymphotoxicity *in vitro* assays are well described (e.g., Pessina et al., 2007), and *in vitro* functional assays that have been used to assess immunotoxicity include, but are not limited to, the human lymphocyte activation (HuLA) assay, an antigen recall assay similar to the *in vivo* T-cell-dependent antibody response (TDAR) (Collinge et al., 2020), lymphocyte proliferation assays, natural killer (NK) cell assays, human whole blood cytokine release assays (HWBCRA), the IL2-Luc /IL2 Luc LTT assays (Kimura et al., 2020), and the MITA (Multi-ImmunoTox Assay), which uses a combination of three luciferase reporter cell lines. In Unilever's Safety and Environment Assurance Centre, immune modulation data for NGRA has been previously gleaned from BioMAP Diversity 8 and Plus Panels (Singer et al., 2019), PBMC-based assays covering the adaptive (B and T cells) and innate arms of the immune system, using both stimulated and unstimulated cells and measuring a range of parameters including viability, proliferation, cytokine and prostaglandin release, and activation markers.

It is clear therefore that aspects of the NAM-based NGRA can, within the general NGRA framework, become increasingly bespoke, as additional NAMs may be added to address specific concerns, but the desire to assure safety using the best available tools with an investigative mindset should not be inhibited by an intransigent regulatory framework entrenched in animal tests. Furthermore, this entire strategy is very closely aligned with the aforementioned US EPA Blueprint for Toxicology (Thomas et al., 2019) and its tiered guidance on how to characterize the MoA of a chemical at consumer-relevant concentrations. In cases when a chemical elicits non-specific effects, which is particularly relevant to cosmetic ingredients and industrial chemicals, a POD is derived using the most sensitive pathway or phenotypic effect. Such a derived POD does not aim to identify a specific adverse outcome or pathology but, rather, aims to be protective of human health by estimating an exposure at which no biological response is expected (Wetmore et al., 2015). This should be the primary and overriding aim: safety, not compliance with archaic animal tests for the sake of dossier completeness.

7 Concluding remarks: NAMs and NGRA are ready for regulatory use

The desire of consumers and society for non-animal-based safety assurances, geographical bans on animal testing of cosmetic ingredients, and inspirational texts suggesting that the safety of humans and the environment could be achieved through better ways than tonnage-driven rodent studies, has galvanized thinking on how to make the best of available and developing scientific methods to inform our safety decisions. Those methods, and the approaches by which they are drawn together to make confident and reliable decisions on safety, *are ready for use*.

Understandably, many will seek affirmation that safety assessments using non-animal approaches can provide the same level of protection as current animal tests. This is undoubtedly because the traditional animal-based testing paradigm has a long history of use, and the scientific and regulatory community has confidence that safety assessments based on these data are protective of human health. However, evidence is emerging that similar (or better) protection of human health can be provided using NAMs and NGRA approaches, and the greater understanding of human biology they impart can be especially valuable, without necessarily predicting the effects seen in high-dose rodent studies.

Twenty years ago, Olson et al. (2000) established that the true positive human drug toxicity concordance rate with rodents was a mere 43%, and yet many still hold that the rodent tests are the "gold standard". Most money launderers will give a better return than 43 cents on the dollar. The gold standard was a monetary system by which the economic unit of account was based on a gold reserve of the same value, held by central banks in

[20] https://toxys.com/reprotracker/

[21] https://stemina.com/products-and-services/devtox-quickpredict/

such vaults as Fort Knox and the Old Lady of Threadneedle Street. It was largely abandoned during the Great Depression that began in 1929. The Acute Toxicity Test (LD50) was introduced two years earlier than that, 1927, and yet, remarkably, it is still requested by regulators around the world. Needless to say, much more can be done to improve and build experience and confidence in NAMs/NGRA and to create new regulatory frameworks around this science, but bigger and braver steps now need to be taken.

References

Allen, T., Goodman, J., Gutsell, S. et al. (2014). Defining molecular initiating events in the adverse outcome pathway framework for risk assessment. *Chem Res Toxicol 27*, 2100-2112. doi:10.1021/tx500345j

Baltazar, M. T., Cable, S., Carmichael, P. L. et al. (2020). A next-generation risk assessment case study for coumarin in cosmetic products. *Toxicol Sci 176*, 236-252. doi:10.1093/toxsci/kfaa048

Berggren, E., White, A., Ouedraogo, G. et al. (2017). Ab initio chemical safety assessment: A workflow based on exposure considerations and non-animal methods. *Comput Toxicol 4*, 31-44. doi:10.1016/j.comtox.2017.10.001

Bernauer, U., Bodin, L., Chaudhry, Q. et al. (2021). The SCCS notes of guidance for the testing of cosmetic ingredients and their safety evaluation, 11th revision, 30-31 March 2021, SCCS/1628/21. *Regul Toxicol Pharmacol 127*, 105052. doi: 10.1016/j.yrtph.2021.105052

Bowes, J., Brown, A., Hamon, J. et al. (2012). Reducing safety-related drug attrition: The use of in vitro pharmacological profiling. *Nat Rev Drug Discov 11,* 909-922. doi:10.1038/nrd3845

Bury, D., Alexander-White, C., Clewell, H. et al. (2021). New framework for a non-animal approach adequately assures the safety of cosmetic ingredients – A case study on caffeine. *Regul Toxicol Pharmacol 123*, 104931. doi:10.1016/j.yrtph.2021.104931

Collinge, M., Schneider, P., Dingzhou Li, D. et al. (2020). Cross-company evaluation of the human lymphocyte activation assay. *J. Immunotoxicol 17*, 51-58. doi:10.1080/1547691X.2020.1725694

Deisenroth, C., DeGroot, D., Zurlinden, T. et al. (2020). The alginate immobilization of metabolic enzymes platform retrofits an estrogen receptor transactivation assay with metabolic competence. *Toxicol Sci 178*, 281-301. doi:10.1093/toxsci/kfaa147

Dent, M., Amaral, R., Da Silva, P. et al. (2018). Principles underpinning the use of new methodologies in the risk assessment of cosmetic ingredients. *Comput Toxicol 7*, 20-26. doi:10.1016/j.comtox.2018.06.001

Dent, M., Li, H., Carmichael P. L. et al. (2019). Employing dietary comparators to perform risk assessments for anti-androgens without using animal data. *Toxicol Sci 167*, 375-384. doi:10.1093/toxsci/kfy245

Dent, M., Vaillancourt P. E., Thomas R. S. et al. (2021). Paving the way for application of next generation risk assessment to safety decision-making for cosmetic ingredients. *Regul Toxicol Pharmacol 125*, 105026. doi:10.1016/j.yrtph.2021.105026

Donnellan, L. (2020). Animal welfare in the European Union: The cosmetics regulation. In R. Moldovan (ed.), *The European Union: Policies, Perspectives and Politics* (Book Chapter, 187-215). Hauppauge, NY USA: Nova Science Publishers. https://novapublishers.com/shop/the-european-union-policies-perspectives-and-politics/

Drasler, B., Sayre, P., Steinhäuser, K. et al. (2017). In vitro approaches to assess the hazard of nanomaterials. *NanoImpact 8*, 99-116. doi:10.1016/j.impact.2017.08.002

Escher, S. E., Aguayo-Orozco, A., Benfenati, E. et al. (2022). Integrate mechanistic evidence from new approach methodologies (NAMs) into a read-across assessment to characterise trends in shared mode of action. *Toxicol In Vitro 79*, 105269. doi:10.1016/j.tiv.2021.105269

Fentem, J., Malcomber, I., Maxwell, G. et al. (2021). Upholding the EU's commitment to 'animal testing as a last resort' under REACH requires a paradigm shift in how we assess chemical safety to close the gap between regulatory testing and modern safety science alternatives to laboratory animals. *Altern Lab Anim 49*, 122-132. doi:10.1177/02611929211040824

Hall, B., Tozer, S., Safford, B. et al. (2007). European consumer exposure to cosmetic products, a framework for conducting population exposure assessments. *Food Chem Toxicol 45*, 2097-2108. doi:10.1016/j.fct.2007.06.017

Harrill, J., Shah, I. and Woodrow Setzer, R. W. (2019). Considerations for strategic use of high-throughput transcriptomics chemical screening data in regulatory decisions. *Curr Opin Toxicol 15*, 64-75. doi:10.1016/j.cotox.2019.05.004

Harrill, J. A., Logan, J. and Everett, D. E. (2021). High-throughput transcriptomics platform for screening environmental chemicals. *Toxicol Sci 181*, 68-89. doi:10.1093/toxsci/kfab009

Hartung, T. and Corsini, E. (2013). Immunotoxicology: Challenges in the 21st century and in vitro opportunities. *ALTEX 30,* 411-426. doi:10.14573/altex.2013.4.411

Hatherell, S., Baltazar M. T., Reynolds, J. et al. (2020). Identifying and characterizing stress pathways of concern for consumer safety in next-generation risk assessment. *Toxicol Sci 176*, 11-33. doi:10.1093/toxsci/kfaa054

Hopperstad, K., DeGroot, D. E., Zurlinden, T. et al. (2022). Chemical screening in an estrogen receptor transactivation assay with metabolic competence. *Toxicol Sci 187*, 112-126. doi:10.1093/toxsci/kfac019

Kimura, Y., Yasuno, R., Watanabe, M. et al. (2020). An international validation study of the IL-2 Luc assay for evaluating the potential immunotoxic effects of chemicals on T cells and a proposal for reference data for immunotoxic chemicals. *Toxicol In Vitro 66*, 104832. doi:10.1016/j.tiv.2020.104832

Li, H., Reynolds, J., Sorrell, I. et al. (2022). PBK modelling of topical application and characterisation of the uncertainty of C_{max} estimate: A case study approach. *Toxicol Appl Pharmacol 442*, 115992. doi:10.1016/j.taap.2022.115992

Middleton, A. M., Reynolds, J., Cable, S. et al. (2022). Are non-animal systemic safety assessments protective? A toolbox and evaluation strategy. *Toxicol Sci Jul 13,* kfac068. doi:10.1093/toxsci/kfac068

Moxon, T. E., Li, H., Lee, M. Y. et al. (2020). Application of physiologically based kinetic (PBK) modelling in the next generation risk assessment of dermally applied consumer products, *Toxicol In Vitro 63*, 104746. doi:10.1016/j.tiv.2019.104746

NRC – National Research Council (2007). *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Washington, DC, USA: The National Academies Press. doi:10.17226/11970

Nyffeler, J., Willis, C., Lougee, R. et al. (2020). Bioactivity screening of environmental chemicals using imaging-based high-throughput phenotypic profiling. *Toxicol Appl Pharmacol 389*, 114876. doi:10.1016/j.taap.2019.114876

Olson, H., Betton, G., Robinson, D. et al. (2000). Concordance of the toxicity of pharmaceuticals in humans and in animals, *Regul Toxicol Pharmacol 32*, 56-67. doi:10.1006/rtph.2000.1399

Paul Friedman, K., Gagne, M., Loo, L.-H. et al. (2020). Utility of in vitro bioactivity as a lower bound estimate of in vivo adverse effect levels and in risk-based prioritization. *Toxicol Sci 173*, 202-225. doi:10.1093/toxsci/kfz201

Pearce, R. G., Setzer, R. W., Strope, C. L. et al. (2017). httk: R package for high-throughput toxicokinetics. *J Stat Softw 79*, 1-26. doi:10.18637/jss.v079.i04

Pessina, A., Bonomi, A., Casati, S. et al. (2007). Mitochondrial function, apoptosis and cell cycle delay in the WEHI-3B leukaemia cell line and its variant ciprofloxacin-resistant WEHI-3B/CPX. *Cell Prolif 40*, 568-579. doi:10.1111/j.1365-2184.2007.00456.x

Phillips, J. R., Svoboda, D. L., Tandon, A. et al. (2018). BMDExpress 2: Enhanced transcriptomic dose-response analysis workflow. *Bioinformatics 35*, 1780-1782. doi:10.1093/bioinformatics/bty878

Proença, S., Escher, B. I., Fischer, F. C. et al. (2021). Effective exposure of chemicals in in vitro cell systems: A review of chemical distribution models. *Toxicol In Vitro 73*, 105133. doi:10.1016/j.tiv.2021.105133

Punt, A., Firman, J., Boobis, A. et al. (2020). Potential of ToxCast data in the safety assessment of food chemicals. *Toxicol Sci 174*, 326-340. doi:10.1093/toxsci/kfaa008

Punt, A., Louisse, J., Beekmann, K. et al. (2022). Predictive performance of next generation human physiologically based kinetic (PBK) model predictions based on in vitro and in silico input data. *ALTEX 39*, 221-234. doi:10.14573/altex.2108301

Rajagopal, R., Baltazar, M. T., Carmichael, P. L. et al. (2022). Beyond AOPs: A mechanistic evaluation of NAMs in DART testing. *Front Toxicol 4*, 838466. doi:10.3389/ftox.2022.838466

Rampa, K. M., Van De Venter, M., Koekemoer, T. C. et al. (2021). Exploring four South African Croton species for potential anti-inflammatory properties: In vitro activity and toxicity risk assessment. *J Ethnopharmacol 10*, 282. doi:10.1016/j.jep.2021.114596

Reynolds, J., Malcomber, S. and White, A. (2020). A Bayesian approach for inferring global points of departure from transcriptomics data. *Comput Toxicol 16*, 100138. doi:10.1016/j.comtox.2020.100138

Reynolds, G., Reynolds, J., Gilmour, N. et al. (2021). A hypothetical skin sensitisation next generation risk assessment for coumarin in cosmetic products. *Regul Toxicol Pharmacol 127*, 105075. doi:10.1016/j.yrtph.2021.105075

Rusyn, I., Sakolish, C. and Kato, Y. (2022). Microphysiological systems evaluation: Experience of TEX-VAL tissue chip testing ConsortiumTox sciences. *Toxicol Sci*, in press. doi:10.1093/toxsci/kfac061

SCCS – Scientific Committee on Consumer Safety (2021). SCCS Notes of Guidance for the Testing of Cosmetic Ingredients and their Safety Evaluation, 11th revision, 30-31 March 2021, SCCS/1628/21. doi:10.1016/j.yrtph.2021.105052

Singer, J. W., Al-Fayoumi, S., Taylor, J. et al. (2019). Comparative phenotypic profiling of the JAK2 inhibitors ruxolitinib, fedratinib, momelotinib, and pacritinib reveals distinct mechanistic signatures. *PLoS One 27*, 14. doi:10.1371/journal.pone.0222944

Solan, M. E. and Lavado, R. (2021). The use of in vitro methods in assessing human health risks associated with short-chain perfluoroalkyl and polyfluoroalkyl substances (PFAS). *J Appl Toxicol 6*, 4270. doi:10.1002/jat.4270

Thomas, R. S., Bahadori, T., Buckley T. J. et al. (2019). The next generation blueprint of computational toxicology at the U.S. Environmental Protection Agency. *Toxicol Sci 169*, 317-332. doi:10.1093/toxsci/kfz058

Tolosa, L., Martínez-Sena, T., Schimming, J. P. et al. (2021). The in vitro assessment of the toxicity of volatile, oxidisable, redox-cycling compounds: Phenols as an example. *Arch Toxicol 95*, 2109-2121. doi:10.1007/s00204-021-03036-w

Wetmore, B. A., Wambaugh, J. F., Allen, B. et al. (2015). Incorporating high-throughput exposure predictions with dosimetry-adjusted in vitro bioactivity to inform chemical toxicity testing. *Toxicol Sci 148*, 121-136. doi:10.1093/toxsci/kfv171

Williams, D. P., Lazic, S. E., Foster, A. J. et al. (2020). Predicting drug-induced liver injury with Bayesian machine learning. *Chem Res Toxicol 33*, 239-248. doi:10.1021/acs.chemrestox.9b00264

Yang, C., Barlow, S. B., Muldoon Jacobs, K. L. et al. (2017). Thresholds of toxicological concern for cosmetics-related substances: New database, thresholds, and enrichment of chemical space. *Food Chem Toxicol 109*, 170-193. doi:10.1016/j.fct.2017.08.043

Zuang, V., Dura, A., Asturiol Bofill, D. et al. (2021). Non-animal methods in science and regulation. EUR 30553 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-28396-6. doi:10.2760/163367

Conflict of interest

All authors are members of the Safety and Environmental Assurance Centre of Unilever.