National Institute
for Public Health
and the Environment

Letter report 320016002/2009
J.G.M. Bessems

Opinion on the usefulness of in vitro data for human risk assessment

Suggestions for better use of non-testing approaches

Opinion on the usefulness of *in vitro* data for human risk assessment

Suggestions for better use of non-testing approaches

Author:
JGM Bessems

Date:
16 June 2009

VWS Kennisvraag 5.5.7

Project V/320016 ‘Nieuwe methodieken in de risicobeoordeling’

RIVM/SIR PORS nr: 12045

Reviewed by:
AJAM Sips and MTM van Raaij

Centre for Substances and Integrated Risk assessment (SIR), National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

Summary

Goal

This document focuses on the use of *in vitro* alternatives for **full replacement** of animal bioassays with respect to chemicals **risk assessment**. This focus will be put into several perspectives. One perspective reflects to risk assessment in contrast to classification and labelling as well as screening and prioritisation. The other perspective reflects to full replacement in contrast to reduction and refinement. All this will be discussed within the area of safety testing as it is required under the Cosmetics Directive and the REACH Regulation. Some recommendations are presented to advance the 3Rs in the regulatory use of animals.

Problems

The availability of *in vitro* full replacement methods is limited.

1. For screening/priority setting purposes, **several** *in vitro* tests are available for local as well as systemic toxicity endpoints.
2. For classification and labelling (C&L) purposes, *in vitro* tests are **only** available for **local** toxicity endpoints.
3. For quantitative risk assessment purposes (QRA), **no (!)** *in vitro* alternatives are available at all.

Conclusions

Current *in vitro* tests are not useful as full replacement of bioassays for systemic toxicity endpoints (C&L and QRA purposes) mainly because of the following reasons:

(A) The current *in vitro* tests are too reductionistic in nature. They do not sufficiently represent the complexity of most systemic endpoints, neither do currently available integrated battery approaches of *in vitro* tests.

(B) The current state of the art of quantitative *in vitro in vivo* extrapolation (QIVIVE) does not allow stand alone use of *in vitro* replacements for chemicals risk assessment.

Solutions

(I) *In vitro* approaches including *in vitro* test batteries deserve a more important role (i.e. more easily achievable in near future) in reduction and refinement than in full replacement. Reduction/refinement can be achieved with regard to species selection, selection of the most relevant route, assessment of the urgency (priority setting) or need at all (waiving) to perform *in vivo* bioassays.

(II) The aspect of back-calculation of an assumed equipotent *in vitro* and *in vivo* concentration to an *in vivo* dose (reverse dosimetry) needs much more attention. Development of QIVIVE cannot do without increased focus on toxicokinetics[1]. By using just a few animals, toxicokinetic data can be prosperous for further use of *in vitro* approaches (reduction instead of replacement). Notably, integrated kinetic modelling approaches called physiologically-based kinetic (PBK) modelling that can be fed significantly with *in vitro* data, could help to generate kinetic data using as few animals as possible.

[1] Toxicokinetics is the integration of absorption, distribution, metabolism and excretion (ADME) and describes 'what the body does to the chemical' whereas toxicology or better toxicodynamics describes 'what the chemical does to the body'.

Samenvatting

Doel
Dit briefrapport beschrijft het gebruik van *in vitro* alternatieven voor **volledige vervanging** van dierproeven met betrekking tot risicobeoordeling van chemische stoffen. Dit gebeurt vanuit diverse perspectieven. Een daarvan betreft de risicobeoordeling als tegenstelling van classificatie en etikettering zowel als screening en prioritering. Het andere perspectief betreft volledige vervanging ten opzichte van vermindering en verfijning. Dit wordt besproken vanuit het testen van de veiligheid zoals verplicht onder de Cosmetica Richtlijn en de REACH Wetgeving. Er worden enkele aanbevelingen gegeven ter bevordering van de 3V's in het regulatoire gebruik van proefdieren.

Problemen
De beschikbaarheid van volledig vervangende *in vitro* methoden is beperkt.

1. Voor screening/prioritering zijn er **diverse** *in vitro* testen beschikbaar voor lokale en systemische eindpunten.
2. Voor classificatie en etikettering ('C&L') doelen zijn *in vitro* testen **alleen** beschikbaar voor **lokale** eindpunten.
3. Voor kwantitatieve risico beoordeling ('QRA') zijn er überhaupt **geen (!)** *in vitro* alternativen beschibaar.

Conclusies
De huidige *in vitro* testen zijn niet geschikt als volledige vervanging van dierproeven voor systemische eindpunten ('C&L' en 'QRA') om de volgende redenen:

(A) De huidige *in vitro* testen zijn te simpel van aard. Ze reflecteren onvoldoende de complexiteit van de meeste systemische eindpunten. De thans beschikbare geïntegreerde benaderingen van batterijen van *in vitro* tests kunnen dat nog evenmin.
(B) De huidige stand van zaken met betrekking tot kwantitatieve *in vitro in vivo* extrapolatie (QIVIVE) is onvoldoende voor één op één vervanging voor risicobeoordeling van chemische stoffen

Oplossingen
(I) *In vitro* benaderingen inclusief *in vitro* test batterijen verdienen een grotere rol (d.w.z. eerder haalbaar) ten aanzien van vermindering en verfijning dan van volledige vervanging. Vermindering/verfijining is haalbaar met betrekking tot de keuze van het proefdier, keuze van de meest relevante route van blootstelling, inschatting van urgentie (prioritering) of überhaupt de noodzaak (evt. afzien) van het doen van dierproeven.
(II) Het terugrekenen van een equipotente *in vitro* en *in vivo* concentratie naar een *in vivo* blootstelling ('omgekeerde dosimetrie') verdient veel meer aandacht. Ontwikkeling van QIVIVE is onmogelijk zonder extra aandacht voor toxicokinetiek[2]. Door slechts een beperkt aantal dieren te gebruiken kunnen kinetiek gegevens *in vitro* aanpakken faciliteren (vermindering i.p.v. vervanging). Nota bene, geïntegreerde benaderingen t.a.v. de kinetiek genaamd 'physiologically-based kinetic' (PBK) modellering, kunnen grotendeels gevoed worden met *in vitro* data, en zo kinetiek gegevens genereren met zo min mogelijk proefdieren.

[2] Toxicokinetiek betreft de integratie van absorptie, distributie, metabolisme en excretie (ADME) en beschrijft 'wat het lichaam doet met de chemische stof' terwijl toxicologie of beter gezegd toxicodynamie beschrijft 'wat de stof doet met het lichaam'.

Introduction

The field of *in vitro* alternatives for animal bioassays in safety testing of chemicals is complex (throughout this document, the term chemicals will be used for non-pharma substances, with emphasis on cosmetic ingredients and REACH chemicals). A clear focus and definition of the domain for which an *in vitro* alternative is developed is often absent. Within safety testing three important purpose or applicability domains can be discerned.

1. The *use domain*, i.e. screening and prioritisation, classification and labelling (C&L) or quantitative risk assessment (QRA).
2. The *biological domain*, i.e. is the test for a local toxicity endpoint, or does it aim at investigation of a systemic toxicity endpoint.
3. The *time domain*, i.e. does the test reveal effects upon single acute exposure or after repeated dose exposure.

Being a multidimensional issue, (sub)domains can overlap. For each (sub)domain, more or less specific animal bioassays were developed in the past century. Some bioassays can suit several (sub)domains. A one-generation study can serve C&L as well as QRA purposes, both for systemic endpoints as well as local toxicity (when the test substance is applied on the skin the test can reveal dermal toxicity upon repeated exposure). In the last few decades, *in vitro* alternatives were or are being developed to replace, reduce or refine these animal bioassays. For a proper assessment of the usefulness of *in vitro* alternatives, a clear understanding of these three applicability domains is of utmost importance as highlighted by the two following contrasts:

1. Classification and Labelling (C&L) versus Quantitative Risk Assessment (QRA), i.e. qualitative versus quantitative. Classification is about the inherent capacity of a chemical to cause an adverse effect at legally predefined dose levels, mostly in experimental animals. Human risk assessment is the assessment whether a known or expected real life exposure is expected to result in an adverse effect, including an estimate of incidence and severity.
2. Local vs systemic toxicity: C&L entails both groups of endpoints. For local toxicity, C&L is the only relevant use purpose, as no QRA follows. Classification for a systemic endpoints, however, may trigger further work, such as quantitative assessment of the hazard that the compound is classified and labelled for (the compound is toxic for reproduction, a C&L conclusion, but how potent is it for that hazard, a conclusion to be used in QRS) or banning of certain substances such as CMR substances in consumer products.

This document will focus on the usefulness of *in vitro* data for quantitative human risk assessment (QRA), i.e. translation of quantitative *in vitro* concentrations reflecting systemic toxicity effects to *in vivo* doses assumed to present equal effects. Systemic endpoints will be emphasized, local toxicity endpoints are mostly yes/no endpoints (eye irritating or not, skin sensitising or not). These are generally not used in a quantitative risk assessment, albeit local toxicity is a risk management issue (e.g. labelling of a substance and/or product and use of personal protection equipment). Site-specific or (inter)national measures are available to prevent local effects to occur during regular handling (personal protection equipment) or during accidents, e.g. during transport (transport packaging requirements), respectively.

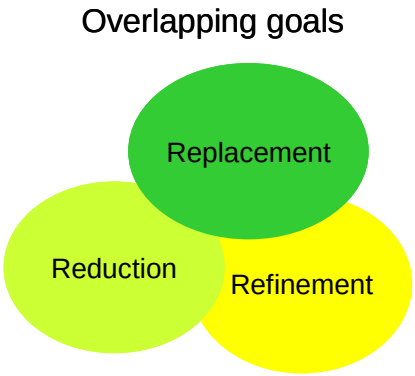
The systemic toxicity endpoints for which usually a QRA is performed are repeated dose toxicity (28-days, 90-days, two-years study), reproductive toxicity, developmental toxicity (including teratogenicity) and carcinogenicity. Skin and eye irritation/corrosion, skin sensitisation, phototoxicity and acute toxicity will be mentioned only shortly as these are only C&L endpoints. Albeit for the endpoint of skin sensitization there are ongoing efforts to perform a quantitative assessment.

As the issue of 3Rs with respect to chemical safety testing is very complex, for the reminder of this short document, the current discussions with respect to the usefulness of *in vitro* alternatives for QRA will be presented by a series of contrasts. There is no intention in the order these are presented, i.e. there is no causal, legislative, time nor importance reasoning in the way these contrasts are ordered. The document will conclude on some important general reasons why current attempts to reach the 3R goals largely fail as well as on some more specific lacunes in the road to 3R. Finally, recommendations are presented for a more efficient approach toward these goals.

Contrasts

Replacement versus reduction/refinement

More animals be saved and more animal suffering may be avoided in near future by pragmatism rather than by reasons of principle. Developments with respect to reduction and refinement are very likely capable of saving more animals than those aiming at full replacement. Although Russel and Burch[3] advocated three possibilities to increase animal welfare, full replacement has received most attention so far. Of course, if animal bioassays can be skipped, it is the best we can do to increase animal welfare. Unfortunately, results so far regarding full replacement are not as good as was hoped for in the past few decades. Regarding local endpoints, many full replacement methods have been technically validated[4], e.g. by ECVAM[5], some of which have been approved, e.g. by OECD, and some of which have made been implemented by the European Commission via Annex V to 67/548/EEC (EEC, 1967). Importantly, however, regarding systemic toxicity endpoints for risk assessment purposes, the conclusion is simple and clear, there is not a single full replacement assay. A recent thesis with cosmetics as scope provides a comprehensive overview for further in-depth information regarding the state-of-the-art regarding many replacement alternatives (Pauwels, 2008). Pauwels noted that no replacement alternatives for *in vivo* studies with regard to systemic toxicity endpoints are foreseen as all intensive efforts have not even generated a first perspective for the area or repeated dose toxicity (Pauwels, 2008). See Annex A for a short overview of ongoing initiatives with respect to alternative approaches for various general and specific repeated dose toxicity endpoints. A combined ECVAM/NICEATM[6] initiative just aims at prediction of the *in vivo* starting dose with respect to acute toxicity testing by using *in vitro* cytotoxicity measurements (SCCP, 2007).



[3] The so-called "three R's" of animal research - replacement, reduction and refinement - were first proposed by W. Russell and R. Burch. In the mid-1950s, they were hired by a scientific animal welfare organization based in the UK, to conduct a study of humane techniques for laboratory animal experiments (http://www.cnprc.ucdavis.edu/pages/alternativeshistory.html).

[4] Technical validation covers issues such as repeatability and robustness and for tests aiming at C&L purposes, specificity (what is the level of accurate positive predictions in comparison with the gold standard, usually an animal bioassay) and sensitivity (what is the percentage of accurate negative predictions compared to the gold standard). It may ignore issues such as metabolic bioactivation.

[5] European Center for the Validation of Alternative Methods

[6] US NTP Interagency Center for the Evaluation of Alternative Toxicological Methods

However, there are many opportunities for reduction and refinement. In integrated screening and testing strategies, intelligent use of *in vitro* tests can help to waive (and thus reduce) or refine animal testing.

1. *Replacement*: Use of *in vitro* approaches to waive basically required effect tests for specicif toxicological endpoints under the Cosmetics Directive as well as under REACH, could effectively turn out to be an effective replacement approach, although by definition, it is not a one-in-one replacement. Moreover, a one-in-one replacement for the whole (future) chemical universe (all different kinds of chemical, from inorganics to organics, from simple atoms like silver in nanosize form as preservative to more complex soluble organic chemicals) in principle is at least for the next decade a theoretical, i.e. impracticable goal. This renders the requirements as laid down in the 7th Amendment to the Cosmetics Directive more or less a theoretical amendment.
2. *Reduction*: If a battery of *in vitro* tests can confirm what has been hypothesised for a specific chemical, i.e. no or negligible risk for a specific hazard, because of the absence of any structural alert for e.g. developmental toxicity, waiving of *in vivo* testing for developmental toxicity should be seriously considered. When used in such a strategy, the result of the strategy is in fact reduction of the number of actual *in vivo* tests still required (reduction).
3. *Refinement*: Positive predictions based on a set of *in vitro* assays, could be used to focus the *in vivo* assay and use less animals (just for confirmation) or finish the study at an earlier time-point by using more specific, and/or earlier and or more sensitive endpoints or biomarkers.

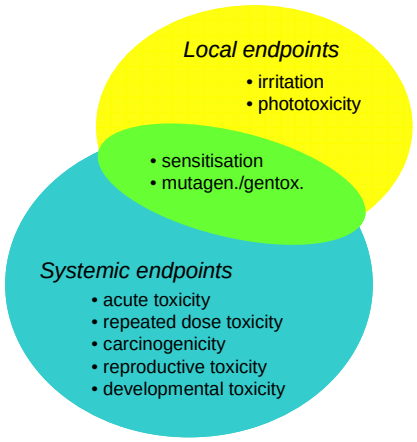
Conclusively, no full one-in-one replacement alternatives for QRA purposes will be available in short or medium term. Therefore it is recommended to shift focus from aiming at full replacement alternatives to methods, strategies and approaches that can help towards reduction and refinement with respect to *in vivo* animal bioassays.

Local versus systemic endpoints

In vitro replacement test protocols are available, technically validated and even regulatory accepted for several local endpoints. *In vitro* replacement tests as well as multi-test approaches for systemic toxicity endpoints have not escaped the stage of primary research. With respect to acute toxicity, the European Commission funded FP6 project ACuteTox may come up with interesting results that may be of use for C&L purposes with respect to the endpoint acute systemic toxicity (http://www.acutetox.org/, 2009.03.25). Other projects aim at repeated dose systemic endpoints which are by far much more complex and more difficult to mimic in *in vitro* models. Several ongoing European projects try to establish reduction/refinement alternatives for the endpoints skin sensitisation, repeated dose toxicity, carcinogenicity and reproductive and developmental toxicity.

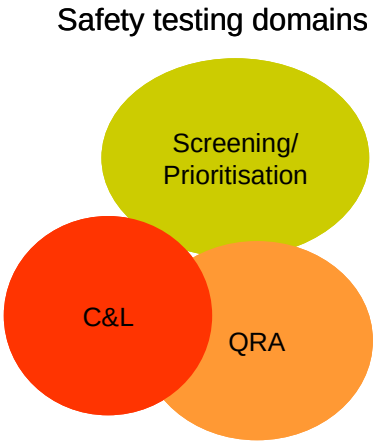
Concrete results for replacement alternatives for systemic toxicity are not expected in the nearby future. For a short overview see Annex A.

Local endpoints versus systemic endpoints



C&L versus QRA

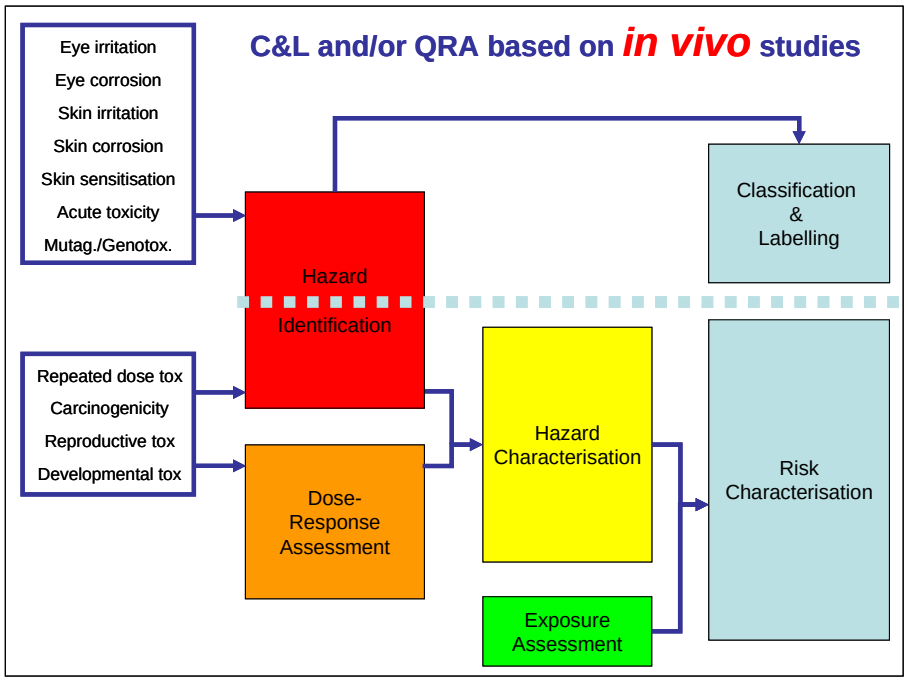
As for C&L purposes for the endpoint mutagenicity/genotoxicity, positive results from a battery of *in vitro* tests have to be followed by a confirmatory *in vivo* test. This *in vivo* test can

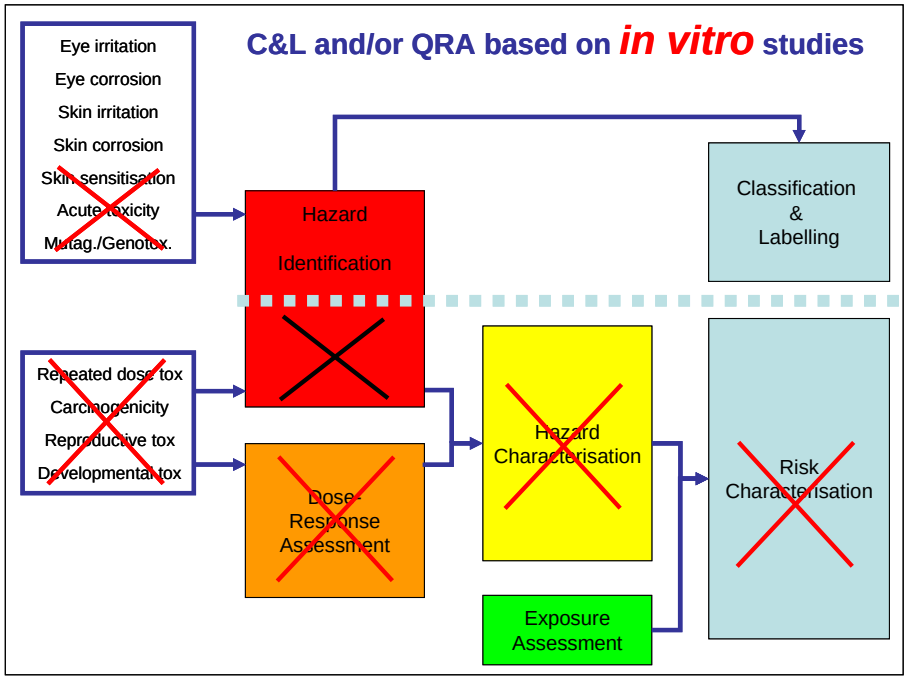


be relatively short as it aims only at a classification endpoint, not at a quantitative dose descriptor. Alternatively, one could perform some kind of a toxicokinetic (TK) study to investigate whether the chemical can enter the systemic circulation (ECHA, 2008; ILSI, 2009)[7]. It is noted that even if an *in vivo* test confirms the *in vitro* results, i.e. that the chemical is a mutagen, for risk assessment purposes, a quantitative *in vivo* dose descriptor like a NOAEL or a Benchmark Dose is necessary as a point of departure (ECHA, 2008; Van Benthem, 2007). In case of a mutagen, this would be a carcinogenicity study. Lastly and importantly, quantitative dose descriptors can not be based directly on *in vitro* effect tests only as the endpoint of an *in vitro* test is a concentration, not an *in vivo* dose. Thus, a translation to an equivalent or equipotent *in vivo* dose is needed for QIVIVE.

In conclusion, some *in vitro* approaches have potencies to develop into useful tools with respect to C&L purposes. *In vitro* approaches to be used for QRA are far away as tools for QIVIVE are lacking. Increased efforts are necessary to widen the scope and applicability of C&L approaches to QRA purposes by taking TK including biotransformation into account.

[7] It is noted that for assessment of this systemic bioavailability, all possible human exposure routes (oral, dermal, via inhalation) should be considered.


C&L and/or QRA based on in vivo studies
Eye irritation
Eye corrosion
Skin irritation
Skin corrosion
Skin sensitisation
Acute toxicity
Mutag./Genotox.
Repeated dose tox
Carcinogenicity
Reproductive tox
Developmental tox
Hazard Identification
Dose-Response Assessment
Hazard Characterisation
Exposure Assessment
Classification & Labelling
Risk Characterisation


C&L and/or QRA based on in vitro studies
Eye irritation
Eye corrosion
Skin irritation
Skin corrosion
Skin sensitisation
Acute toxicity
Mutag./Genotox.
Repeated dose tox
Carcinogenicity
Reproductive tox
Developmental tox
Hazard Identification
Dose-Response Assessment
Hazard Characterisation
Exposure Assessment
Classification & Labelling
Risk Characterisation

Metabolic incompetent versus metabolic competent systems

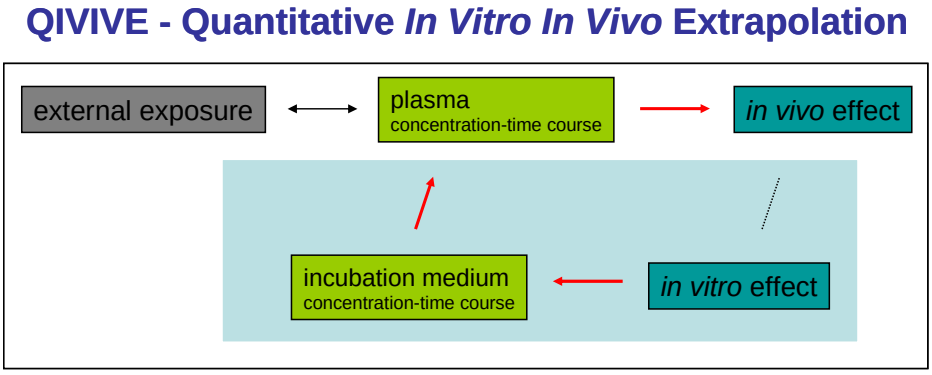
Most if not all alternatives that are under development for systemic toxicity testing lack a metabolic system. Some chemicals exert toxicity only after metabolic activation. This seriously hampers the applicability domain and thus the reliability of these assays as possible metabolic activation may result in hazardous metabolites. Many attempts are being undertaken for co-culturing where target cells are co-cultivated in the presence of metabolic competent cells. So far, very little progress has been made. A recommended solution to this problem would be to investigate metabolic activation in isolated *in vitro* systems that are developed so far. If the chemical turns out to be metabolically stable, this would increase the reliability of *in vitro* effect tests that lack metabolism. Another option is to investigate all possible and relevant metabolites that result from the *in vitro* biotransformation assay in the *in vitro* effect test. An example is described by Janer et al. (2008). It is noted that, currently, for inclusion of many metabolic systems in *in vitro* assays, still animals are needed as these 'systems' are isolated from animals (hepatocytes, microsomes, S9 mix). For cosmetics, this means that the metabolic incompetence of current *in vitro* systems as such is a serious problem as this implies at most reduction and refinement rather than replacement.

In conclusion, increased efforts are needed to include metabolically active systems in *in vitro* alternatives for systemic toxicity endpoints.

Quantitative in vitro in vivo extrapolation for QRA: Facts and myths

Quantitative *in vitro in vivo* extrapolation for QRA purposes is currently a myth. It is, by definition quantitative and in need of TK data, as explained above. Second and by definition as well, it refers to repeated dose systemic toxicity. The second issue is discussed shortly in paragraph B. The first issue mentioned regarding the need of TK data is discussed here. For QIVIVE, it is necessary to assume that *in vitro* concentrations around the target (cells, tissue) are related to *in vivo* concentrations at the target site (e.g. liver cells or the brain). As *in vivo* target sites are often difficult to measure, blood or plasma (any tissue in a living organism is more or less surrounded by blood) is chosen as surrogate for the target tissue. This leads to the first assumption that an *in vitro* effect concentration is related to an *in vivo* blood or plasma concentration that would result in the same effect (size). So the working hypothesis needed is that the *in vitro* effective concentration is equivalent or at least linearly related to the *in vivo* effective concentration in serum. An example may be that the *in vitro* and the *in vivo* concentrations causing LDH leakage (a marker for liver cell damage) in an hepatocyte culture and a living organism, respectively, are the same. By accepting this assumption, there is still one step to go, i.e. internal to external extrapolation. Human limit values such as an ADI or TDI (as being part of human risk assessment in a wider perspective) are expressed as external exposures. The obvious solution is to relate the effective *in vivo* internal blood concentration to an external dose/exposure by using use TK information for back-calculation of the value. This is what is called reverse dosimetry.

In conclusion: Any future use of *in vitro* effect data for human risk assessment (i.e. once any *in vitro* replacement approach is ready for QRA) is in need of TK data on the chemical to perform quantitative *in vitro in vivo* extrapolation (QIVIVE). This means that the study of the TKs of a chemical is a prerequisite before any *in vitro* effect data can be used at the basis of quantitative risk assessment and thus should obtain much more attention.



Concentrations and their time course provide link between the effects observed *in vitro* and the effects predicted (or observed) *in vivo*. More precise, the only link between in vitro systems and the in vivo situation is the **unbound concentration** in medium (in vitro) and in extracellular fluids (in vivo). Extracellular fluid unbound concentrations can be approximated by plasma unbound concentrations.

TK Information for reverse dosimetry as such: Uses and needs

For many chemicals, the needs with respect to TK data suitable for reverse dosimetry are not fulfilled in reality. There are two approaches possible with respect to fill in the needs regarding TK information, being key in QIVIVE:

- If *in vivo* TK data are present, TK parameters can be derived using classical compartmental modelling. This was illustrated by a recent ECVAM-funded project at RIVM/SIR (Noorlander et al., 2008). By using compartmental modelling, external doses equivalent to an *in vitro* effect concentration can be calculated in an iterative process as long as the kinetics stays linear. It is noted that in this approach, *in vitro* data on TK can not be incorporated as classical compartmental modelling is an empirical technique.
- When TK data are insufficient or absent anyway, a first assessment with respect to reverse dosimetry is feasible using only *in silico* or *in vitro* TK data (Hagens et al., 2008). These non-*in vivo* data could be integrated for reverse dosimetry purposes by using whole-body physiologically-based (PBK) models. However, this is more complicated in most cases and a very limited number of *in vivo* TK data (limited blood sampling = minor animal discomfort) would be needed for verification reasons. Human PBK models can be generated and subsequently verified by using limited blood samples from human volunteer studies.

The following should be noted with respect to the issue of PBK-based reverse dosimetry as such. So far it has been tested with disputable results for a few chemicals (Verweij et al., 2006). In the meantime, others have done important work for both approaches on taking into account protein-binding aspects and differences thereof *in vitro* and *in vivo* (Gülden and Seibert, 2003; Gülden et al., 2006). A theoretical framework as well solutions for practical problems are needed.

In conclusion, actual implementation of TK-based reverse dosimetry in QIVIVE needs further focus and research such as actual case studies followed by (pre) validation.

In silico and in vitro assays to inform TK/ADME for reverse dosimetry: Needs versus reality
At the moment, the needs for useful *in silico* and *in vitro* assays to provide data for integrated PBK modelling approaches for industrial chemicals by far exceed the availability. However, various ongoing European FP6 projects aim at *in vitro* technologies for the assessment of pharmacokinetics, i.e. kinetics at pharmacologically (not per se equivalent to toxicologically) relevant dose. The project Liintop aims at optimisation of liver and intestine *in vitro* models for amongst others pharmacokinetic studies (http://www.liintop.cnr.it/index.php, 2009.02.12). Issues under research are intestinal absorption, metabolic modification in the intestinal cells, absorption and metabolism in the hepatocytes. Another FP6 project is MEMTRANS. Its goals are to optimize and pre-validate *in vitro* cultured cell models to predict oral absorption and pharmacokinetics of efflux systems substrates (http://www.acrossbarriers.de/34+M52087573ab0.html , 2009.02.12). Although no results are published yet, it indicates that at least some efforts are ongoing to investigate absorption and metabolism *in vitro* systems. Unfortunately there is a large bias towards pharmacologically relevant chemicals because of the scope of these projects (pharmacokinetics). The transferability to industrial chemicals of any outcome would be very unclear and thus would need further investigation. The two other important processes in ADME that have obtained significantly less attention so far are distribution and excretion. In PBK models, distribution of many chemicals can be described by partitioning between blood and tissues, or more specific, partition coefficients that describe the steady state ratio of concentrations between two tissues, e.g. fat:blood. Up scaling of existing approaches as well as future approaches to medium throughput *in vitro* systems or *in silico* prediction models for the determination of partition coefficients would be very favourable for increased use of PBK modelling. This applies to its use in QIVIVE as well as in various other extrapolations usually needed in QRA (interspecies extrapolation, high-to-low dose extrapolation, route-to-route extrapolation). Preferably, these *in silico* and/or *in vitro* models should be applicable and valid for a wide range of physicochemical properties in order to cover most chemicals under the Cosmetics Directive and REACH.

In conclusion, *in vivo* TK information that is already present could be taken up in QIVIVE by using classical compartmental modelling. When (almost) no TK information is present, various combined *in silico* and /or *in vitro* techniques are available in theory. However, significant efforts are needed for development and/or expansion of the applicability domain of these techniques to non-pharmacologically relevant chemicals

Omics – helpful or hype?

Omics techniques such as toxicogenomics, proteomics, metabonomics are not usefull for QRA. Moreover, they are not alternative methods. However, since there have been large expectations as to what omics techniques in the long run could do for the 3R (including the use for QRA) a few words will be spent here. Principally, omics analytics are tools, and yes, sometimes powerful tools, to investigate biological processes at various levels of cell biochemistry. In this respect, it is a tool for screening for potentially hazardous properties. Omics does not investigate apical endpoints[8] such as relative liver weight increase, or delayed type neurotoxicity. Omics can help to reveal biological mechanisms within homeostasis as well as once homeostasis is disturbed (Fig. xx in Annex). Furthermore, in future, omics observations such as disturbance of gene pathways might be used as qualitative early biomarkers. However, to fulfil this expectation, much more research is needed regarding biological mechanisms and adverse effects with respect to causal relationships, time-dependency, time window and dose-response relationships. This could help to change QRA as a science based largely on apical endpoints and black box extrapolation to a systems biology based science incorporating biological mechanisms at human relevant exposures.

Conclusively, in future omics could be usefull for QRA once the multidimensional relationships (cause-effect, time-effect, dose-response) have been elucidated to a large extent and for a significant number of chemicals.

[8] Apical end point. An observable outcome in a whole organism, such as a clinical sign or pathologic state, ….. that can result from exposure to a toxicant (US NRC/NAS, 2007).

Human versus animal

By using *in vitro* approaches, the species of interest, i.e. man, can be investigated directly. Use of the human species for safety testing is an ethical issue. However, the increasing number of human cells-based assays, in combination with omics techniques, offers opportunities (US NRC/NCA, 2007). It will create possibilities to investigate animal-human differences using a systems biology approach, offering new possibilities for parallel approaches (animal *in vitro* – man *in vitro* – animal *in vivo* – man *in vivo*). Although no use for quantitative risk assessment is foreseen in short or medium term, the increased use of human biomaterial in *in vitro* tests may have some relevance in establishment of the relevance or irrelevance of a specific hazard that is observed in a laboratory animal species (hazard identification purpose). Also, increased exploration of human biology that is responsible for identified hazards may finally help to develop human *in vitro* models representing specific targets in human biology responsible for specific hazards (US NRC/NCA, 2007). The recently launched ASAT[9]-programme is mainly following this strategy with an important addition, i.e. the increased use of human clinical data (http://www.asat-initiative.eu/launchevent.htm, 2009.02.24).

Conclusively, *in vitro* approaches offer opportunities for investigations regarding hazard identification directly in human biomaterial, thereby circumventing interspecies extrapolation. However, there is a long way to go before this is practically useful in QRA.

[9] ASAT. Assuring Safety without Animal Testing

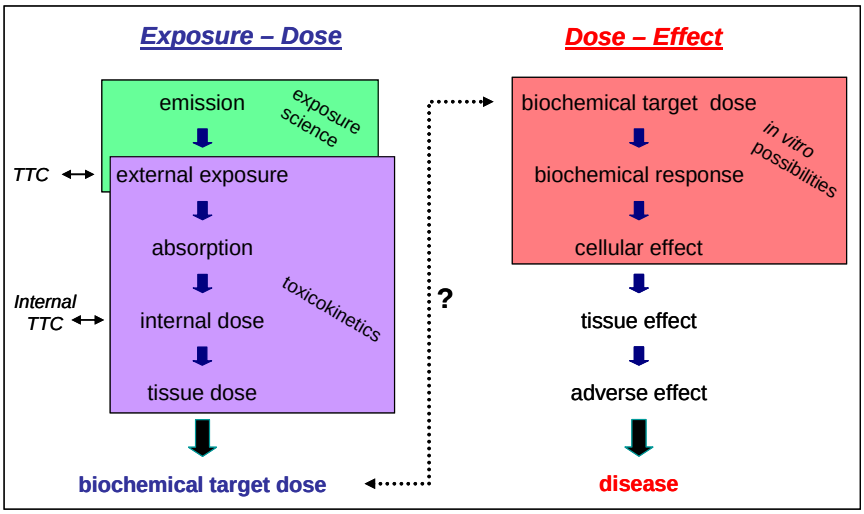
Current strategy versus intelligent testing strategies

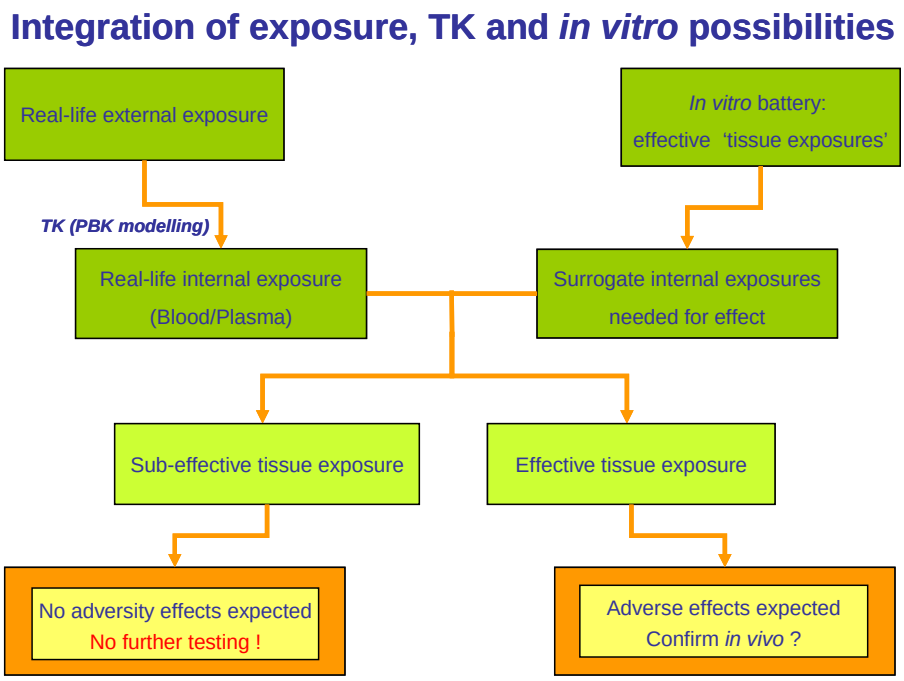
Current testing strategies focus too much on hazard whereas the real life issue is about risks. Short and medium term results regarding 3R can be expected partly from a combination of approaches including *in silico* and/or *in vitro* replacement and reduction/refinement of existing tests. However, often ignored but are strategies that could be used to prevent even thinking about hazards. E.g. are there situations where we do not need to know the hazard potential of a chemical because of very unlikely or very low human exposure? When do we think it is acceptable to waive some of the systemic toxicity bioassays? These questions are addressed in the FP6 project OSIRIS in which several national organisations such as TNO, WUR, VU, KWR and RIVM (using VROM funding) are involved (http://www.osiris-reach.eu/, 2009.03.25).

This is where possibilities of exposure and exposure assessment in its widest sense emerge. In the well known *exposure – dose – effect* continuum, one can see that exposure science, TK and *in vitro* toxicology offer possibilities, not just the one or the other. In a continuum approach, one could foresee the following.

– Exposure science and TK could investigate the 'exposure – dose' part and focus on one or more aspects in *emission – external (consumer/worker) exposure – absorption – internal dose – tissue dose – (maximum) target dose*.
– *In vitro* toxicology could investigate the 'dose – effect' part in a tiered approach by focussing only on earlier or sometimes including later steps in the *(effective) target dose – biochemical response – cellular effect – tissue effect – adverse effect – disease* continuum (toxicodynamics).

Reduction opportunities by using the *CONTINUUM* approach





In an intelligent screening strategy, using read-across and the category approache as well as SAR or QSAR tools, a chemical could show up with alerts for one or more toxicologically relevant endpoints. Subsequently, based on the threshold of toxicological concern (TTC) concept, an in-depth exposure assessment could result in the following: exclusion of the possibility of any relevant level of exposure that could result in any systemic toxicity. From an economic/financial viewpoint, the intelligent screening strategy including feed-back loops could even result in some risk management measures (emission reduction) which could be much cheaper and cost no animals for testing compared to performing the required set of bioassays for systemic toxicity endpoints. In other words, to reach the 3R goals, the attitude should change from hazard-driven testing strategies to possible risk-driven strategies.

Conclusively. The question underlying ITS should be: how do we use information on (internal) human exposure in order to guide testing strategies, so we can move away from animal high-dose toxicity testing to human relevant low-dose testing (US NRC, 2007; Hubal et al., 2008; OSIRIS; TTC references).

Conclusions

At the moment, *in vitro* full replacement alternatives for *in vivo* bioassays to be used for QRA purposes do not exist. Several ongoing initiatives try to establish reduction/refinement alternatives for the endpoints skin sensitisation, repeated dose toxicity, carcinogenicity and reproductive and developmental toxicity. Many of them use a systems biology approach (battery approach) and omics techniques. This may lead to major changes in the risk assessment paradigm as well as to results with respect to reduction and refinement, albeit not in the near future. Concrete results for complete replacement alternatives are very unlikely in the foreseeable future. In contrast, a change of hazard-driven to risk-driven testing strategies offers serious possibilities for approaches that lead to reduction in the number of bioassays.

These approaches include increased focus on all parts of the exposure – dose – response continuum. For the first part, i.e. external exposure and internal exposure (ADME/TK) this is can be based completely and mainly on non-animal testing approaches, respectively. For the second part (dose – response), this reflects to combined use of *in silico* predictions and *in vitro* test batteries to mimic the complex whole organism. Important in this respect is integration of the results of the latter. Further, for quantitative *in vitro in vivo* extrapolation (QIVIVE), proper TK information is indispensable, either from *in vivo* assays or from combined *in vitro / in silico* approaches. An *in vitro* concentrations assumed to be equipotent to an *in vivo* blood concentration has to be back-calculated to an *in vivo* dose (reverse dosimetry). The only methodology capable to perform this task is physiologically-based kinetic (PBK) modelling. Notably, PBK modelling is the only approach that can integrate *in silico* and *in vitro* predictions of absorption, distribution, metabolism and excretion.

Stepwise recommendations to achieve 3R results in the near future

A shift from full replacement *tests* to alternative *approaches* incorporating

1. systems biology opportunities including omics techniques for hazard testing
2. increased awareness of exposure science for waiving and targeted testing
3. focus on developments in toxicokinetics science for integration of hazard and exposure by reverse dosimetry as needed in quantitative *in vitro in vivo* extrapolation

Ad 1: Systems biology including omics techniques should be used to find new opportunities mainly in hazard testing and to bridge various gaps that occur when animal testing is reduced and replaced by batteries of individual simple, tests that each mimick parts of the biology. Most important challenges are the gap between current animal bioassays and new *in vitro* assays, between human clinical findings (biomarkers) and new human *in vitro* assays as well as the interspecies gaps (parallel approach).

Ad 2: Developments in exposure sciences, especially increasing the level of prediction models in the low to very low exposure range should allow the following: conclude that toxicity testing is not necessary (waiving) based on the threshold of toxicological concern (TTC) concept as the exacerbation of toxicity is a matter of dose. Better exposure estimates are needed to facilitate risk- and exposure driven testing strategies.

Ad 3: Developments in toxicokinetic sciences should focus on internal exposure assessment and reverse dosimetry by using whole body PBK models. Recommended in this respect is:

- further development and validation of *in vitro* absorption tools (e.g. Caco2 system)
- development of medium throughput systems to establish partitioning coefficients
- harmonisation of *in vitro* metabolism test protocols
- development of *in vitro* models for renal excretion
- further development of *in silico* tools for first tier estimates of ADME

As implementation of these recommendations requires substantial effort, increased focus on concerted actions in an international setting is recommended, such as European Framework Programs.

Acknowledgements

The authors thank dr JGM van Engelen and dr AJAM Sips (RIVM, Centre for Substances and Integrated Risk Assessment), dr BC Hakkert (RIVM, Expertise Centre for Substances) and dr J van Benthem (RIVM, Laboratory for Health Protection Research) for their input and valuable discussions.

References

ASAT (2008) Assuring Safety without Animal Testing. ASAT Initiative http://www.asat-initiative.eu/index.htm (9 February 2009)

EEC (1967) Council Directive 67/548/EEC on the approximation of laws regulations and administrative provisions relating the the classification, packaging and labelling of dangerous substances. As adapted to techincal progress (ATP) several times.

EC (2003) Directive 2003/15/EC of the European Parliament and of the Council of 27 February 2003 amending Council Directive 76/768/EEC on the approximation of the laws of the Member States relating to cosmetic products. Official Journal L66, 26-35, 11 March 2003

EC (2006) Regulation (EC No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC

ECHA (2008) Guidance on information requirements and chemical safety assessment. Chapter R.7a: Endpoint specific guidance. Guidance for the implementation of REACH. European Chemicals Agency

Gülden M, Seibert H (2003) *In vitro-in vivo* extrapolation: estimation of human serum concentrations of chemicals equivalent to cytotoxic concentrations *in vitro*. Toxicology. 2003 Aug 1;189(3):211-22. Erratum in: Toxicology. 2003 Nov 5;192(2-3):265.

Gülden M, Dierickx P, Seibert H (2006) Validation of a prediction model for estimating serum concentrations of chemicals which are equivalent to toxic concentrations *in vitro*. Toxicol In Vitro. 2006 Oct;20(7):1114-24.

Hagens WI, Brandon EFA, Van Eijkeren J, Sips AJAM, Bessems JGM (2008). Non-*in vivo* techniques to assess model parameters for physiologically-based kinetic (PBK) modelling - An overview. RIVM-SIR Advisory report nr: 11985. Centre for Substances and Integrated Risk Assessment, National Institute for Public Health and the Environment, The Netherlands.

Prepared under the 'Alternatieven voor Dierproeven' project as financed by the Netherlands ministry of Public Health, Welfare and Sports (Knowledge question 5.5.5 - 2007)

Hartung T, Bremer S, Casati S, Coecke S, Corvi R, Fortaner S, Gribaldo L, Halder M, Hoffmann S, Roi AJ, Prieto P, Sabbioni E, Scott L, Worth A and Zuang V (2004) A modular approach to the ECVAM principles on test validity. *Altern Lab Anim*, 32, 467-72.

ILSI (2009) ILSI-HESI "The relevance and follow-up of positive results in in vitro genetic toxicity (IVTG) testing", Workshop of the ILSI-HESI "in vitro genetic toxicity testing" (IVTG) Committee, Madison Hotel, Washington DC, USA, 2-6 February 2009

Janer G, Verhoef A, Gilsing HD, Piersma AH (2008) Use of the rat postimplantation embryo culture to assess the embryotoxic potency within a chemical category and to identify toxic metabolites. Toxicol In Vitro. 2008 Oct;22(7):1797-805.

INVITTOX No 113. Embryonic Stem Cell Test for Embryotoxicity / Teratogenicity http://ecvam-dbalm.jrc.ec.europa.eu/public_view_doc.cfm?id=DC5ABDF7AC30F1B7DF7EF27E87D68AAC7180BB0BC12CB10496CDA74B54630A05A3291B895581F634 (5 February 2009)

Kabinetsvisie (2008) "Alternatieven voor Dierproeven". Brief van de Minister van Volksgezondheid, Welzijn en Sport aan de Tweede Kamer, VGP-VV 2855846 (6 juni 2008), Kamerstuk 9 juni 2008. http://www.minvws.nl/kamerstukken/vgp/2008/kabinetsvisie-alternatieven-voor-dierproeven.asp?rss

Noorlander C, Zeilmaker M, Van Eijkeren J, Bourgeois F, Beffers R, Brandon E and Bessems J (2008) Data collection on kinetic parameters of substances. Pilot phase – A methodological report, CCR.IHCP.C432921.XO, RIVM, December 2008

OECD (1984) OECD Guideline for testing of chemicals, Test No 417, Toxicokinetics, 4 April 1984

OECD (2004a) OECD Guideline for testing of chemicals, Test No. 427: Skin Absorption: *In Vivo* Method, 13 April 2004

OECD (2004b) OECD Guideline for testing of chemicals, Test No. 428: Skin Absorption: *In Vitro* Method, 13 April 2004

OSIRIS (2007) Optimized Strategies for Risk Assessment of Industrial Chemicals through Integration of Non-Test and Test Information. EU Integrated Project. Sixth Framework Programme. Contract no. GOCE-CT-2007-037017. http://www.osiris.ufz.de/index.php?en=16314. (9 February 2009)

Ovigne JM, Martinozzi-Teissier S, Verda D, Abdou D, Piroird C, Ade N, Rousset F (2008) The MUSST for *in vitro* skin sensitization prediction: Applicability domains and complementary protocols to adapt to the physico-chemical diversity of chemicals. Abstracts / Toxicology Letters 180S (2008) S216, I27

Pauwels M (2008), Critical evaluation of the current EU regulatory framework for the safety assessment of cosmetics, Doctoral Thesis, Vrije Universiteit Brussels

Piersma AH, Janer G, Wolterink G, Bessems JGM, Hakkert BC and Slob W (2008) Quantitative extrapolation of *in vitro* whole embryo culture embryotoxicity data to developmental toxicity *in vivo* using the benchmark dose approach, Toxicological Sciences 101(1), 91–100 (2008)

SCCP (2007) Memorandum on the actual status of alternative methods on the use of experimental animals in the safety assessment of cosmetic ingredients in the European Union, SCCP/1111/07, European Commission, Directorate-General Health & Consumer Protection, http://ec.europa.eu/health/ph_risk/committees/04_sccp/docs/sccp_s_06.pdf

Slob W, Janer G, Bessems JGM, Hakkert BC, Sips AJAM, Verhoef A, Wolterink G, Piersma AH (2008) Quantitative *in vitro - in vivo* extrapolation. Analysis of 19 compounds of varying embryotoxic potency. RIVM Report 340720001/2008, Bilthoven, The Netherlands

SCCP (2007) Memorandum on the actual status of alternative methods on the use of experimental animals in the safety assessment of cosmetic ingredients in the European Union, EU Scientific Committee on Consumer Products, SCCP adopted this opinion at its 12th plenary on 19 June 2007, SCCP/1111/07, http://ec.europa.eu/health/ph_risk/committees/04_sccp/docs/sccp_s_06.pdf (2009.02.24)

SCCP (2009) Position Statement on Genotoxicity / mutagenicity testing of cosmetic ingredients without animal experiments, EU Scientific Committee on Consumer Products, SCCP adopted this opinion at its 19th plenary on 21 January 2009, SCCP/1212/09, http://ec.europa.eu/health/ph_risk/committees/04_sccp/docs/sccp_s_08.pdf (2009.02.24)

US NRC/NAS (2007) Toxicity Testing in the Twenty-first Century: A Vision and a Strategy. US National Research Council of the National Academy of Sciences.

Van Benthem J (2007) The effect of REACH implementation on genotoxicity and carcinogenicity testing. RIVM Report 601200008/2007.

Verwei M, van Burgsteden JA, Krul CA, van de Sandt JJ, Freidig AP (2006) Prediction of *in vivo* embryotoxic effect levels with a combination of *in vitro* studies and PBPK modelling. Toxicol Lett. 2006 Aug 1;165(1):79-87.

Annex A.

Where are we regarding 3R goals for QRA purpose assays?

Skin sensitisation

Although in most if not all legislative frameworks regarded as an endpoint for C&L purposes only, ongoing attempts try to use quantitative information from e.g. the LLNA assay to be used in quantitative risk assessment. Although the LLNA and the reduced LLNA result in reduction and refinement compared to the classical Buehler and the Magnusson-Kligmann Guinea-Pig Maximisation test, they are no replacement alternatives. Conclusively, there are no *in vitro* alternatives for skin sensitisation. Some results may be reached in the EU FP6 project Sens-it-iv. One of its goals is the establishment of *in vitro* conditions supporting communication between various relevant cell-types and the cascade of cellular and molecular events triggered in such a complex system by a test-compound (http://www.sens-it-iv.eu/, 2009.02.12).

Repeated dose toxicity

Regarding 28-days, 90-days and chronic toxicity no replacement alternatives do exist. No replacement alternatives are foreseen either, although some results regarding drugs are expected from the FP6 Project Predictomics (short term *in vitro* assays for long term toxicity). It aims at early biomarkers of chronic toxicity based on omics analysis of cells exposed to model hepatotoxins and nephrotoxins and establishment of a decision tree, based on the biomarkers, as well mathematical models to early anticipate the potential toxicity of drugs under development (http://www.predictomics.com/). As a spin-off or lead for future research, these findings could be used for hazard identification (but not risk assessment) of chemicals relevant for cosmetics and REACH (reduction and/or refinement, not replacement). Another initiative may provide some results in reducing animal experimentation in safety testing by human cardiomyocyte by using *in vitro* models derived from embryonic stem cells (http://er-projects.gf.liu.se/%7Einvitroheart).

Carcinogenicity

No replacement alternatives do exist regarding carcinogenicity, none are foreseen either. Although some results may come from the FP6 project carcinogenomics (http://www.carcinogenomics.eu/ , 2009.02.12). It aims to develop a battery of mechanism-based *in vitro* tests accounting for various modes of carcinogenic action. These tests will be designed to cover major target organs for carcinogenic action e.g. the liver, the lung and the kidney. The novel assays will be based on the application of "omics" technologies. The generated omics data will be integrated into a holistic understanding of systems biology.

Reproductive toxicity

No replacement alternatives do exist regarding reproductive toxicity, none are foreseen either. Some results regarding integrated approaches may result from the FP6 project ReProTect where the predictive power of a range of pioneering *in vitro* tests is explored to cover a large part of the reproductive cycle (http://www.reprotect.eu/, 2009.02.12).

Developmental toxicity

Although several *in vitro* tests have been developed and technically validated (ESAC Statement) to describe embryonic development (Luijten et al., 2007), none of them is ready for use in risk assessment. Every single test of a series of three tests, i.e. the EST (embryonic stem cell test), the rat WEC (whole embryo culture) assay and the MM (limb bud micromass) test only covers part of the prenatal growth and development. Further, none of them has been validated for QRA purposes, one of the reasons being a lack of suitable and reliable *in vivo* data (Slob et al., 2008; Piersma et al., 2008).

The above results in an intermediate conclusion: At the moment, there **are no *in vitro* replacement alternatives available for QRA purposes**. Secondly, various achievements with respect to 'omics' are helpful in hazard screening and intelligent testing strategies.

RIVM

National Institute
for Public Health
and the Environment

P.O. Box 1
3720 BA Bilthoven
The Netherlands
www.rivm.com