
Chapter 3
Carcinogenicity

Compiled by Workgroup 3
14 July 2010

Jan van Benthem¹, Susan Felter², Tuula Heinonen³, Albrecht Poth⁴, Rositsa Serafimova⁵, Joost van Delft⁶, Emilio Benfenati⁷, Pascal Phrakonkham⁵, Andrew Worth⁵, Raffaella Corvi⁵.

¹National Institute of Public Health and the Environment, Laboratory of Health Protection Research, Bilthoven, The Netherlands.
²Procter and Gamble Company, Central Product Safety, Cincinnati OH, USA.
³FICAM, Medical School, University of Tampere, Finland.
⁴Harlan Cytotest Cell Research GmbH, Roßdorf, Germany.
⁵ECVAM, Institute for Health & Consumer Protection, European Commission Joint Research Centre, 21020 Ispra (VA), Italy.
⁶Department of Health Risk Analyses and Toxicology, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands.
⁷Istituto di Ricerche Farmacologiche “Mario Negri”, Laboratory of Environmental Chemistry and Toxicology, Milan, Italy.
Executive summary

- Carcinogenesis is a complex long-term multi-factorial process, and consists of a sequence of stages.

- Carcinogens have conventionally been divided into two categories according to their presumed mode of action: genotoxic carcinogens that affect the integrity of the genome by interacting with DNA and/or the cellular apparatus and, non-genotoxic carcinogens that exert their carcinogenic effects through other mechanisms.

- The two-year cancer bioassay in rodents is widely regarded as the gold-standard to evaluate cancer hazard and potency, however this test is rarely done on cosmetic ingredients. Rather, a combination of shorter-term in vitro and in vivo studies have been used including in vitro and in vivo genotoxicity assays to assess genotoxic potential and repeat-dose (typically 90-day) toxicity studies to assess the risk non-genotoxic chemicals.

- It is clear that the animal testing bans under the 7th amendment of the Cosmetics Directive (EU 2003) will have a profound impact on the ability to evaluate and conduct a quantitative risk assessment for potential carcinogenicity of new cosmetic ingredients. This impact is not only due to the ban on the cancer bioassay itself (after 2013), but is also due to the ban of in vivo genotoxicity testing (effective since 2009), any repeat-dose toxicity testing (after 2013), and other tests such as toxicokinetics studies and in vivo mechanistic assays which are currently be used to aid safety assessment.

- This report is a critical evaluation of the available non-animal test methods and their ability to generate information that could be used to inform on cancer hazard identification.

- Although several in vitro short term tests are available beyond the standard in vitro genotoxicity assays to support conclusions on cancer hazard identification, the in vitro short-term tests will not be sufficient to fully replace the animal tests needed to confirm the safety of cosmetic ingredients. Furthermore, those that are available to assess potential carcinogenicity are focused on hazard evaluation only and cannot currently be used to support a risk assessment.

- Taking into consideration the present state of the art of the non-animal methods, the experts were unable to suggest a timeline for full replacement of animal tests currently needed to fully evaluate carcinogenic risks of chemicals. Although a timeline for full replacement cannot be developed, clearly the timeline is expected to extend past 2013.
**General considerations**

**Introduction**

Substances are defined as carcinogenic if after inhalation, ingestion, dermal application or injection they induce tumours (malignant), increase their incidence or malignancy, or shorten the time of tumour occurrence. It is generally accepted that carcinogenesis is a multi-hit/multi-step process from the transition of normal cells into cancer cells via a sequence of stages and complex biological interactions and strongly influenced by factors such as genetics, age, diet, environment, hormonal balance, etc.

Since the induction of cancer involves genetic alterations which can be induced directly or indirectly, carcinogens have conventionally been divided into two categories according to their presumed mode of action: genotoxic carcinogens and non-genotoxic carcinogens. Genotoxic carcinogens have the ability to interact with DNA and/or the cellular apparatus (such as e.g. the spindle apparatus and topoisomerase enzymes) and thereby affect the integrity of the genome, whereas non-genotoxic carcinogens exert their carcinogenic effects through other mechanisms that do not involve direct alterations in DNA.

The two-year cancer bioassay in rodents is widely regarded as the gold-standard to evaluate cancer hazard and potency. However, this test is rarely done on cosmetic ingredients because of expense, time, and animal welfare issues. Rather, a combination of shorter-term *in vitro* and *in vivo* studies have been used including *in vitro* and *in vivo* genotoxicity tests to assess genotoxic potential and repeat dose (typically 90-day) toxicity studies to assess non-genotoxic chemicals.

It is clear that the animal testing bans under the 7th amendment of the Cosmetics Directive (EU 2003) will have a profound impact on the ability to evaluate and conduct a risk assessment for potential carcinogenicity of new cosmetic ingredients. This impact is not only due to the ban on the cancer bioassay itself, but is also to that on *in vivo* genotoxicity testing (effective since 2009), any repeat-dose toxicity testing (after 2013), and other tests such as toxicokinetics studies and *in vivo* mechanistic assays that currently can be used to aid safety assessment.

The challenge will be to find/develop alternative tests for both genotoxic and non-genotoxic carcinogens. The complexity of the carcinogenicity process makes it difficult to develop *in vitro* alternative test models that mimic the full process, especially for non-genotoxic chemicals. The challenge in developing *in vitro* alternatives is also heightened because of the complexity of the number of potential target organs. Some key events of the carcinogenesis process, can be investigated *in vitro*. However, it is expected that an integrated approach involving multiple *in vitro* models will be needed, but a better understanding of the entire process is needed before this will be possible. Eventually, an integrated approach involving multiple *in vitro* models may be needed. Scientific research is ongoing to try to achieve this goal.

---

1 It is noted that not all genotoxic events lead to mutagenicity, and that some prefer the terminology “mutagenic mode of action.” However, genotoxicity assays are still commonly used to distinguish those chemicals with the potential to directly impact affect the integrity of DNA from those that don’t, so for the sake of simplicity, the text throughout refers to genotoxic versus non-genotoxic carcinogens.
Information Requirements for the carcinogenic Safety Assessment of Cosmetics Ingredients until March 2009 (Ref. SCCP Notes of guidance)

The EU Scientific Committee on Consumer Products (SCCP) issued the 6th revision of the “Notes of Guidance for the Testing of Cosmetic Ingredients and their Safety Evaluation” in 2006 (SCCP, 2006). This Guidance document lists the general toxicological requirements for the submission of a cosmetic ingredient dossier to the SCCP as follows: acute toxicity; irritation and corrosivity; skin sensitisation; dermal/percutaneous absorption; repeated dose toxicity; and genotoxicity. These are considered the minimal base set requirements. However, when considerable oral intake is expected or when the data on dermal absorption “indicate a considerable penetration of the ingredients through the skin (taking into account the toxicological profile of the substance and its chemical structure), additional data/information on carcinogenicity, reproductive toxicity and toxicokinetics may become necessary, as well as specific additional genotoxicity data”. It is noted that the SCCP Notes of Guidance does not define what is meant by “considerable oral intake” or “considerable penetration of the ingredients through the skin”. Tools such as the Threshold of Toxicologic Concern (TTC) may be helpful in determining which exposures warrant toxicologic evaluation.

Historically, the strategy for addressing the carcinogenicity endpoint for cosmetic ingredients has been three-fold:

1) First compounds are evaluated for genotoxicity. The first step in this evaluation was a battery of in vitro genotoxicity tests. A positive finding in an in vitro test (e.g. a chromosome aberration study, OECD 473) could then be followed-up with an in vivo study (e.g. mouse micronucleus, OECD 474), which is deemed to have more relevance to human exposures. Positive in vivo tests for their part could trigger carcinogenicity testing.

In general, compounds that have been shown to have genotoxic potential in vivo are not used in the formulation of cosmetics, and materials testing positive in these tests have rarely been pursued further as this might require the conduct of a carcinogenicity bioassay (OECD 452) or a combined chronic toxicity / carcinogenicity test (OECD 453). These studies take several years to run (the in-life portion alone lasts 24 months) and cost around one million euros. For these reasons and given the potential for genotoxic compounds to be positive in a rodent bioassay, new cosmetic ingredients are almost never tested in a carcinogenicity bioassay. Rather, these compounds are generally not progressed further.

2) For those chemicals shown to lack genotoxicity potential, it is generally assumed that there is a threshold and that the carcinogenic risk can be avoided based on data from repeat-dose toxicity studies. Prior to the formations of tumours (generally seen only after long-term exposures) non-genotoxic carcinogens cause changes in normal physiological function and these adverse effects, if relevant to the exposure, would be determined in a sub-chronic study. Accordingly, the risk assessment generally involves the identification of a NOAEL from an appropriate repeat-dose toxicity study (e.g. 90-day study) and the application of appropriate safety factors.

The methods use in such quantitative risk assessments, are regarded as being sufficiently conservative such that even if the chemical was later shown to be a non-genotoxic carcinogen, the exposure would be so low that there would be no risk to consumers. This is consistent with the risk assessment practices of virtually all regulatory bodies, including those inside and outside of Europe. Although this approach does not explicitly address the potential for non-genotoxic carcinogenicity, the conservatism in the approach is thought to be such that if the compound was later tested and shown to cause tumours via a non-genotoxic mode of action, the risk assessment would be sufficiently protective to cover this endpoint.
In addition to repeat-dose toxicity studies, other \textit{in vivo} studies are sometimes used to better understand the human relevance of findings in rodents (e.g., related to toxicokinetic handling, or species-specific effects) or the mechanism and associated dose-response for a chemical.

Implications for carcinogenic safety assessment after the 7\textsuperscript{th} amendment

The 7\textsuperscript{th} Amendment to the Cosmetics Directive (EU 2003) banned the \textit{in vivo} genotoxicity tests for cosmetic ingredients starting March 2009, and the performance of the repeated dose toxicity and carcinogenicity by March 2013. The consequence of these bans for carcinogenicity assessment are that: (1) for genotoxic substances no \textit{in vivo} genotoxicity tests as follow up of \textit{in vitro} positive tests are allowed; and (2) risk from non-genotoxic carcinogens cannot not be sufficiently evaluated since the repeated dose toxicity (and carcinogenicity) test is not allowed anymore. Since both modes of action are important and need to be covered, for both alternative methods and approaches should be considered.

Assessment of genotoxic carcinogens

Until the 7\textsuperscript{th} Amendment both \textit{in vitro} and \textit{in vivo} tests played an important role in the recognition of carcinogenicity in cosmetic ingredients. The \textit{in vivo} tests were used to clarify whether positive results from \textit{in vitro} tests were relevant under conditions of \textit{in vivo} exposure. A number of well-established and regulatory accepted \textit{in vitro} tests are in place but a caveat to the use of these tests is the relatively low specificity and high rate of misleading positive results (i.e., the results are not indicative of an increased cancer risk associated with DNA reactivity, as generally assumed from these tests). Kirkland \textit{et al.} (2005, 2006) evaluated the predictivity of standard \textit{in vitro} tests for rodent carcinogens. The combination of at least two genotoxicity tests increases the sensitivity of the test battery but the specificity (ability to identify non-carcinogens) decreased drastically. The low level of specificity means that unacceptably high numbers of positives are generated, which, before March 2009 could be evaluated and often overruled with \textit{in vivo} genotoxicity tests. Thus the ban of \textit{in vivo} tests will have a negative impact on the development of new cosmetic ingredients. This is clearly demonstrated by the evaluation of 26 hair dyes by the SCCP. Nineteen hair dyes had to be further assessed due to a clastogenic effect found in the \textit{in vitro} systems. For these compounds 37 \textit{in vivo} genotoxicity tests were submitted and 35 turned-out to be negative and 2 (Comet assays) were at the time considered equivocal. These data indicate that the performance of at least 26 \textit{in vivo} tests was deemed necessary for the appropriate characterisation of the genotoxic potential. Without the performance of \textit{in vivo} tests at least 17 of these hair dyes might have been abandoned without full scientific justification (Speit, 2009).

Based on that the weak performance of existing \textit{in vitro} tests and/or the development of new \textit{in vitro} tests with better predictivity for cosmetic ingredients are in focus. An ECVM workshop on “how to reduce false positive results” when undertaking \textit{in vitro} genotoxicity testing was held in 2006 and identified, among others, the following factors as being important for the improvement of the current tests (Kirkland \textit{et al.}, 2007):

1. Identify chemicals to be used in the evaluation of modified or new tests (Kirkland \textit{et al.}, 2008)
2. Use of cell types with higher relevance (e.g. human origin, p53 proficient)
3. The current measures in cytotoxicity in the standard genotoxicity tests
4. The current maximum concentration in the standard genotoxicity tests

Although \textit{in vivo} genotoxicity tests are not allowed anymore, until 2013 they can still be carried out
under some circumstances. The micronucleus test and the Comet assay can be performed if included in a repeated dose study (Pfuhler et al., 2009). The in vivo Comet assay gained growing scientific acceptance over the last years, whereas the inclusion of the micronucleus test is a long and well established concept and is already represented in the current OECD guideline (OECD 474). While the integration of a micronucleus test into repeated dose toxicity studies can be accomplished without the addition of a positive control, this may be problematic for some studies. For example, the problem with the Comet assay may be that this currently requires the inclusion of an endpoint specific positive control group (Pfuhler et al., 2009). The use of positive control reference slides, however, could be an alternative to control for technical variations during study performance. In case the repeated dose studies which include a measure of genotoxicity but not an endpoint specific positive control will not be accepted by the regulatory bodies and thus cannot be used, the only way to further evaluate positive findings from in vitro genotoxicity studies in order to clarify the possible carcinogenic potential of a compound is the performance of a carcinogenicity study. This underscores the value of having a control group scientifically to answer questions that could only otherwise be answered by a bioassay.

The ability to investigate the relevance of positive in vitro genotoxicity results for prediction of carcinogenicity in humans without the use of animals is a significant scientific and technical challenge. In addition to an improvement of the existing in vitro genotoxicity tests, a range of accepted and scientifically accepted tools should be available to allow appropriate experiments for in vitro follow-up testing.

Assessment of non-genotoxic carcinogens

Although it is generally accepted that major carcinogenic risk is related to genotoxic compounds which can well be detected by in vitro methods, the potential risk related to non-genotoxic compounds must also be evaluated. Despite the fact that some of the major mechanisms behind non-genotoxic carcinogenicity are known, the multiple mechanisms of action and the insufficient knowledge of the cellular and molecular events has not yet allowed for the implementation of a battery of in vitro tests that could predict and/or explain their carcinogenic potential to man.

The mechanisms by which non-genotoxic carcinogens cause tumours are in most cases related to tissue- and species-specific disturbances in normal physiological control, gene expression patterns implicated in cellular proliferation, survival and differentiation (Widschwendter and Jones, 2002, Baylin and Ohmi, 2006, Esteller, 2007). Numerous examples exist where the mechanism is animal species specific and thus effects found in animals are not predictive for humans (Shanks et al., 2009).

The mechanisms behind non-genotoxic carcinogenicity can be manifold, many of which are still not completely understood/known. From the ones we know, the induction of tissue-specific toxicity (cytotoxicity) resulting in inflammation and regenerative hyperplasia appears to be a major one. As a result of cell death by cytotoxic agents persistent regenerative growth may occur with increasing probability for spontaneous mutations (Ames and Gold, 1991) which may lead to accumulation and proliferation of mutated cells giving rise to pre-neoplastic foci and, ultimately, to tumors via further clonal expansion. For the same reason, trophic hormones (e.g. estrogen, progestin, gonadotrophins, thyroid hormone) are known as non genotoxic carcinogens since they induce cell proliferation at their target organs (Lima and Van der Laan, 2000).
Many non-genotoxic mechanisms act via binding to receptors thus effecting proliferation, apoptosis and intercellular communication. For example, many endocrine modifiers are human non-genotoxic carcinogens by binding to receptors such as to the estrogen, the progesterone, the aryl hydrocarbon or the thyroid hormone receptor. Relevant roles are also granted for tyrosine kinase (TK), ion channel-coupled and G-protein-coupled receptors (Lima and Van der Laan, 2000). Induction of immunosuppression by chemicals may be another non-genotoxic mechanism associated with the progression (although not the induction, per se) of cancer. The results from immunosuppressant drugs like cyclosporine A have shown that they can elicit direct cellular effects that can lead to promotion of cancer, independent of immune reactivity (Hernández et al., 2009). Oxidative stress in cells also results in non-genotoxic carcinogenesis as it is shown that cancer cells commonly have increased levels of reactive oxygen species (ROS) and that ROS can induce cell malignant transformation (López-Lázaro, 2010). Oxidative stress has been suggested to have some involvement in the mode of carcinogenic activity of peroxisome proliferators in rodent livers (Doull et al., 1999, Hernández et al., 2009). Finally, chemical substances may cause tumours by non-genotoxic mechanisms by altering the chromatin structure by hyper or hypomethylation of DNA, histone modifications and nucleosomal remodelling (Lo and Sukumar, 2008; Sadikovic et al., 2008).

Different research methods, including in vitro methods using several cell types, are available to study a number of these potential mechanisms. For example, tests are available to measure oxidative stress or to measure the inhibition of gap junction intercellular communication (GJIC), both of which been associated with a number of non-genotoxic carcinogens. However, these methods cannot currently be used to reliably predict carcinogenic potential, but rather are focused on better understanding the mechanism for effects elicited by a chemical.

Up to now the safety of non-genotoxic compounds in man has mainly been concluded from repeat dose toxicity tests (mainly 90 days), toxicokinetics, rodent bioassays, if available, and by using threshold principle. Opposite to genotoxic carcinogens for non-genotoxic carcinogens the threshold principle is used for risk assessment. At this moment, no in vitro test battery is recommended to test non-toxicogenic carcinogenic potential of chemical substances. To avoid animal-specific and biased results a testing battery based on human cell based in vitro tissue models with relevant biomarkers is seen as the most optimal way to replace animal tests in non-genotoxic carcinogenic risk assessment. Indeed, it is expected that there will be significant synergies between work to develop replacement tests for repeat dose toxicity studies with tests to predict non-genotoxic carcinogens and quantitative response thresholds.

**Inventory of Alternative Methods Currently Available**

Non-testing methods

Non-testing methods include (quantitative) structure-activity relationships ([Q]SARs) and the formation of chemical categories to facilitate the application of read-across between similar chemicals. Non-testing methods are based on the assumption that the information of a certain compound can be extracted from the analysis of the effects of similar compounds. Such methods are generally computer-based (in silico) approaches.

Quantitative structure-activity relationship (QSAR) models link toxicity to continuous parameters (molecular descriptors) associated with the chemical structure. In case the relationship is simply
(Q)SARs are often incorporated, possibly in conjunction with databases, into expert systems. An expert system is any formalised system that is often, but not necessarily, computer based, and that can be used to make predictions on the basis of prior information (Dearden et al., 1997). Expert systems (and their implementation in software tools) are based on three main modelling approaches referred to rule-based, statistically-based, or hybrid methods. Rule-based methods codify the human rules which identified certain chemical fragments responsible for the effect. Statistical models are built by using data mining methods to extract the information from a set of compounds.

1. Quantitative structure-activity relationship (QSAR)

Short description, scientific relevance and purpose
To date, hundreds of (Q)SAR models have been published in the literature for predicting genotoxicity and carcinogenicity. The most commonly modelled endpoint for genotoxicity has been Ames test mutagenicity. The application of the Ames test to large numbers of chemicals has shown that this test has a high positive predictivity for chemical carcinogens (around 80%) (Benigni, 2010). Most models are qualitative (SARs), i.e. coarse-grain classifiers that predict a chemical compound as genotoxic or carcinogenic or not. Relatively few models are quantitative (QSARs) which provide a more precise means of assessing genotoxicity and carcinogenicity, mainly for congenic sets of chemicals.

QSARs for epigenetic carcinogenicity are still in an early stage of development. A number of structural alerts (SAs) and characteristics of several types of non-genotoxic carcinogens have been summarised (Woo and Lai, 2003). Relatively few models are available for identifying non-genotoxic carcinogens or for predicting carcinogenic potency (Toropov et al., 2009).

There exist several commercial as well as free available expert systems for predicting genotoxicity and carcinogenicity (Benfenati et al., 2009, Serafimova et al., 2010). Freely available models in the public domain include CAESAR, Toxtree, Oncologic, and LAZAR. Commercial models requiring license fees include MultiCase, Topkat, HazardExpert, DEREK, and ToxBoxes.

Rule-based systems contain “if-then-else” rules that combine toxicological knowledge, expert judgment and fuzzy logic. Commonly used software tools based on this approach include OncoLogic (Woo et al. 1995), Derek (Sanderson & Earnshaw 1991; Ridings et al., 1996) and HazardExpert (Smithing and Darvas, 1992). Derek and HazardExpert can be used in conjunction with their sister programs Meteor and Metabolexpert to predict the genotoxicity and carcinogenicity potential of metabolites as well as parent compounds. In addition to these commercial tools, models included in Toxtree and the OECD Toolbox (OECD, 2010) are rule-based.

Statistically-based systems use a variety of statistical, rule-induction, artificial intelligence, and pattern recognition techniques to build models from non-congeneric databases. Statistically based systems are included in the commercial tools MultiCASE (Klopman and Rosenkranz, 1994) and TOPKAT (Enslein et al., 1994) and the publicly available Lazar (Helma, 2006) and CAESAR (Ferrari et al., 2009) models. In addition, many models published in the literature and not implemented in software are statistically based.

Hybrid models are based on a combination of knowledge-based rules and statistically-derived models. These are based on the general idea that, within the structural space of a single SA
(considered to represent a single interaction mechanism), statistically derived models can quantitatively predict the variation in the reactivity of the alert conditioned by the rest of the molecular structure. Examples of the hydrid approach include models implemented in the OASIS TIMES (Mekenyan et al., 2004, 2007; Serafimova et al., 2007) as well as some literature-based models not implemented in software (Purdy, 1996).

The accuracy of QSARs for potency for both Ames test mutagenicity and rodent carcinogenicity (applicable only to toxic chemicals) is 30-70%, whereas the accuracy of QSARs for discriminating between active and inactive chemicals is 70 - 100% depending on the (Q)SAR and dataset used. Usually accuracy of the models for carcinogenicity is relatively lower than what Ames test gives. This is reasonable taking into account the complexity of the carcinogenicity endpoint, and the fact that models do not explicitly include ADME properties, which could be critical steps in the carcinogenic process. It has been argued that QSARs for carcinogenicity classification are of comparable performance to the Ames test (Benigni et al., 2010).

Status of validation and/or standardisation

The validation process for a (Q)SAR model does not follow the validation procedures of in vitro test methods (http://ecb.jrc.ec.europa.eu/qsar/background/). It is fast and unofficial approach for characterising models and documenting them according to an internationally harmonised format, the QSAR Model Reporting Format (QMRF: http://ecb.jrc.ec.europa.eu/qsar/qsar-tools/index.php?c=QRF). Since the usefulness of QSAR estimates is highly context-dependent, there is no official acceptance or adoption process at the EU or OECD levels.

The REACH legislation allows the use of QSAR models that are scientifically valid, applicable to the chemical of interest and that give results that are adequate for the regulatory purpose.

The validation procedure includes an assessment of model performance based on different statistical analyses (Eriksson et al., 2003). For models which are classifiers, statistical parameters, such as accuracy (concordance), sensitivity and specificity, are used. For continuous (regression) models, a range of other parameters are typically used (e.g. the coefficient of determination, $R^2$, and the standard error of the estimate, $s$). The ultimate proof of the predictivity of a QSAR is the demonstration that, when applied to a new set of chemicals not used for the modelling (an independent test set), it predicts reliably their biological activity (Benigni and Bossa, 2008).

The validation procedure also includes an assessment of the applicability domain (AD) of the model. The key is to understand whether it is appropriate to make a prediction for a given query chemical. Different chemometric approaches can be used to describe the AD of a model, and thus to assess model applicability. Some AD methods are based on the structural similarity of the chemical of interest to the training set chemicals, whereas others are based on mechanistic similarity. The program AMBIT for example can evaluate this information (CEFIC, 2010). Other approaches explored the possible use of other pieces of information for the AD. For example, software based on the CAESAR model for carcinogenicity takes into consideration not only chemical but also the toxicological information (CAESAR, 2010). This method evaluates not only the input of the model (the chemical descriptors), but also the output, which is the toxicity property.

Unfortunately, there is no single and harmonised way of evaluating chemical similarity and defining applicability domains (Jaworska and Nikolova, 2010), which means that the assessment of model applicability is not straightforward and needs to rely to some extent on expert judgement.

Field of application and limitations
(Q)SAR models are used by industry in the upstream process, for a fast screening. Generally, the carcinogenicity models are not used for the final assessment.

It is interesting to notice that the results from different models may not agree, because they are based on different chemical information and rules. Indeed, any model is incomplete in its knowledge. Thus, the recommendation is to use more than one model, and then compare results. If all models agree that a chemical is safe, this can be a good basis that the general knowledge, based on known carcinogens, do not identify major reasons of concern. If a model identifies critical aspects, this information can be used in the evaluation of the carcinogenicity. The results should be analysed considering the knowledge at the basis of the software. Indeed, several programs are quite transparent and show for instance the fragment which is supposed to trigger the toxic effect, or similar carcinogenic compounds, *etc.* It is important to read carefully this information, and not to simply limit the information on the carcinogenicity predicted class. This information should be critically evaluated, and can also show that the assumptions at the basis of the predictions are not correct. For this reason, it is very useful to evaluate the AD of the model, when available.

**Ongoing developments**

There are some interesting perspectives for the integration of (Q)SAR with results of other tests. (Q)SAR models can offer advantages in the organisation and exploration of the data and information. This will be more powerful in case of the availability of a huge number of data, arising, for instance, from the ToxCast initiative.

### 2. Read-across and grouping of chemicals

**Short description, scientific relevance and purpose**

A chemical category is a group of chemicals whose physicochemical and human health and/or ecotoxicological properties and/or environmental fate properties are likely to be similar or follow a regular pattern, usually as a result of structural similarity (OECD, 2007). The grouping approach represents a move away from the traditional substance-by-substance evaluation to a more robust approach based on a family of related chemicals. Within a chemical category, data gaps may be filled by read-across, trend analysis and QSARs (van Leeuwen *et al.*, 2009).

The OECD (2007) guidance document on toxicological grouping of chemicals, which is based on the REACH guidance for grouping, proposes a stepwise approach for analogue read across. The steps include: (1) identifying potential analogues, (2) gathering data on these potential analogues, (3) evaluating the adequacy of data for each potential analogue, (4) constructing a matrix with available data for the target and analogue(s), (5) assessing the adequacy of the analogue(s) to fill the data gap, and (6) documenting the entire process. The guideline also indicates the importance of comparing the physicochemical properties of the analogue and target chemicals as well as assessing the likely toxicokinetics of the substances, including the possibility that divergent metabolic pathways could be an important variable. Using the OECD guidance as a foundation, Wu *et al.* (2010) have recently published a framework for using similarity based on chemical structure, reactivity, and metabolic and physicochemical properties to specifically evaluate the suitability of analogues for use in read-across toxicological assessments.

Read-across interpolates or extrapolates the property of one or more compounds. For a given category endpoint, the category members are often related by a trend (*e.g.* increasing, decreasing or constant) in an effect, and a trend analysis can be carried out using a model based on the data for the members of the category. Data gaps can also be filled by an external QSAR model, where the
category under examination is a subcategory of the wider QSAR. All of these approaches can be performed in a qualitative or quantitative manner. In other words, using of a category approach means to extend the use of measured data to similar untested chemicals, and reliable estimates that are adequate for classification and labelling and/or risk assessment can be made without further testing.

Status of validation and/or standardisation

By its very nature, the grouping and read-across approach is an ad hoc, non-formalised approach based on a number of steps including expert choices. Thus, the term “validation” is not meaningful in this context. Instead, estimated properties obtained the grouping and read-across approach need to be assessed in terms of their adequacy, and the justification needs to be clearly documented according to an accepted format (ECHA, 2010). The critical issues of chemical category formation procedure are quality of the existing data for known chemicals and defining of the similarity. The similarities may be based on the following:

- structural features e.g. common substructure, functional group, chemical elements
- physico-chemical, topological, geometrical, surface, quantum chemical properties
- behaviour (eco)toxicological response underpinned by a common Mechanism of Action
- toxicokinetics properties, including metabolic pathways

At present, there are several software tools that can be used to build a category and fill data gaps related with genotoxicity and carcinogenicity (Serafimova et al., 2010). In the current version of the OECD QSAR Application Toolbox, five mechanistically based profilers connected with genotoxicity and carcinogenicity are implemented. Also in the software are included 5 databases which contain genotoxic and carcinogenic experimental data. The Toolbox gives the possibility to form a category using also other criteria for similarity included metabolism. Toxmatch (JRC, 2010b) is another software tool that encodes several chemical similarity indices to facilitate the grouping of chemicals into categories and the application of read-across.

Compared to (Q)SAR methods, the experience on the use of these methods for carcinogenicity is limited. A limitation of these methods is that their reproducibility can be low, because the definition of the similar compounds and their number is not standardised. Different results are expected if the toxicity prediction is based on different compounds. More experience should be produced comparing results obtained from different users.

Field of application and limitations

Read-across is typically used when very similar compounds are present. Assessors rely on the property of these similar compounds. It is obvious that the reliability of these non-testing methods is highly related to toxicity values of the similar compounds. If the information is extracted from one or two chemicals, this information has to be very carefully checked. This applies to all non-testing methods, but in case of a large population of compounds the presence of errors is less critical.

Furthermore, interpolation should be preferred, compared to extrapolation.

Ongoing developments

Interesting perspective exist in the development of more robust methods for similarity evaluation. For this, some of the tools above discussed for the applicability domain can be used for the evaluation of correctness of read-across and grouping.

3. TTC approach
Cosmetics are typically mixtures of different ingredients added at varying levels, some of which are associated with very low exposure to consumers. For these ingredients, the TTC approach may offer a conservative, transparent, and pragmatic way to assure safety. However, because of the conservative assumptions associated with TTC, it will be limited in its general applicability to cosmetics, and will likely not be useful for ingredients used at higher levels or in products that involve higher exposures (e.g., body lotions).

Short description, scientific relevance and purpose

The Threshold of Toxicologic Concern (TTC) is a scientifically-based approach to establish acceptable exposure limits when sufficient chemical-specific toxicologic information is lacking. It is a pragmatic risk assessment tool that relies on the broad grouping of chemicals based on structural features and then assumes that an untested chemical is potentially as toxic as the most toxic chemicals in the group. As a consequence, the TTC exposure limits are by design quite conservative. Furthermore, it is likely that if chemical specific data were available that the risk assessment would support higher exposure levels. The intent of this approach is to provide a framework that minimizes the time and effort spent on assessing low level exposures by providing a means to develop scientifically supported exposure limits for these materials without the need to generate additional toxicity data. It is noted that TTC is different from some of the other alternative approaches described herein because the focus of TTC is on risk assessment (i.e., establishing an acceptable exposure limit) rather than being limited to hazard identification.

The origins of TTC as a risk assessment tool can be found in the US FDA’s Threshold of Regulation (ToR), which was developed as a pragmatic way to assess the safety of low-level food packaging migrants (U.S. FDA, 1995). The ToR established an exposure level of 1.5 µg/day as being protective for chemicals lacking structural alerts for genotoxicity. This number was based on an analysis of the distribution of potencies of chemical carcinogens in the Carcinogen Potency Database (CPDB), which had 477 carcinogens in it at the time. Importantly, a re-analysis of a later update of the Gold’s CPDB (1995) that included more than 700 chemicals showed a similar distribution of cancer potencies (Cheeseman et al., 1999; Kroes et al., 2004). Although the ToR was based on an evaluation of cancer potencies, the exposure limit of 1.5 µg/day was not intended to be used with genotoxic carcinogens. This is because the Delaney Clause in U.S. law prohibits the use of carcinogens as indirect food additives. Therefore, 1.5 µg/day was established as a limit that was would not be used for chemicals with structural alerts or other reason for concern for genotoxicity, but would still be considered to be protective in the event that later testing revealed that the chemical did have some carcinogenic potential.

Status of validation and/or standardization

The TTC methodology and scientific underpinnings continue to be expanded upon such that its utility and acceptance are growing. Since the initial work of the FDA on the ToR, the TTC methodology has been expanded into a tiered approach that has the potential for much broader applicability. Most notably, Kroes et al. (2004) describes the work of an Expert Group of ILSI Europe that culminated in the development of a decision tree that is now widely cited as providing the foundation for a tiered TTC approach. This publication describes a step-wise process in which it is first determined whether TTC is an appropriate tool (proteins, heavy metals and polyhalogenated-dibenzodioxins and related compounds have been so far excluded from use with TTC), and then follows a series of questions to determine the appropriate TTC tier. The initial step is the identification of high potency carcinogens that have currently been excluded from the TTC approach ( aflatoxin-like, azoxy and N-nitroso compounds). After that, the chemical would be analyzed for structural alerts for possible genotoxicity. Those with alerts would be assigned to the
lowest TTC tier of 0.15 µg/day (an order of magnitude lower than the FDA’s ToR). Organophosphates have then been assigned the next highest tier, followed by three higher tiers for non-genotoxic substances. These three tiers are based on the work of Munro et al. (1996), who established a non-cancer database for TTC consisting of repeat-dose oral toxicity data of 613 substances. These substances were divided into three chemical classes of toxic potential on the basis of their structure using the decision tree of Cramer et al. (1978), and the distribution of NOEL’s was established for each of the three Cramer Classes. The 5th percentile NOEL was then calculated for each Cramer Class distribution, and an uncertainty factor of 100 was applied to establish human exposure thresholds of 1800, 540 and 90 mg per person per day (30, 9 and 1.5 mg/kg bw/day) for the Cramer structural classes III, II and I. A new chemical lacking repeat-dose toxicity data could then be assigned a Cramer class based on structure, and the appropriate TTC value assigned.

*Fields of application and limitations*

Since its origins as a tool for food packaging materials in the mid-1990s, the acceptance and utility of TTC has been expanded such that it has been used extensively to assess food flavouring agents (JECFA, 1996, 1997; EFSA, 2004; Renwick, 2004), and genotoxic impurities in pharmaceuticals (EMEA, 2006, 2008; Muller et al., 2006). The TTC decision tree has also been recommended as a tool to evaluate low-level exposures associated with personal and household products (Blackburn et al., 2005) and cosmetic ingredients and impurities in the absence of chemical-specific toxicology data (Kroes et al., 2007). Whereas the TTC databases are oral repeat-dose studies, cosmetic exposures are predominantly dermal. Therefore, in addition to considerations of the chemical domain, application of the TTC approach to cosmetics requires consideration of route-to-route extrapolation, including differences in absorption and first-pass metabolism. Kroes et al. (2007) published an analysis showing that the oral TTC values are in fact valid for use with dermal exposures. Furthermore, they recommended conservative default adjustment factors based on *in silico* prediction tools that could be used to estimate an absorbed dose following dermal exposure.

*Ongoing developments*

In addition to work ongoing to further expand the applicability and acceptance of TTC as a risk tool for cosmetic ingredients, other projects have aimed to expand the tool itself. For example, additional refinements have been recommended by Felter et al. (2009) that allow for the inclusion of genotoxicity data as a way to refine the TTC limit for chemicals that have structural alerts for genotoxicity, and to support higher exposure limits for less-than-lifetime exposures. Also, work is ongoing to develop TTC as a tool to evaluate inhalation exposures (Carthew et al., 2009) from cosmetics, and also as a tool for safety assessment of sensitisation following dermal exposure (Safford, 2008). While initial recommendations have been made in these areas, work is ongoing in each case.

More recently, Bercu et al. (2010) proposed the use of TTC in combination with QSAR tools to establish safe levels for genotoxic impurities (GTIs) in drug substances. The single TTC limit of 0.15 µg/day is highly conservative and intended to be protective for the more potent end of the distribution of potencies for genotoxic chemicals (after excluding highly potent categories such as the N-nitroso carcinogens), and as such can be very restrictive in the development of new drug substances. To address this, Bercu et al. developed a tiered approach to use *in silico* tools to predict the cancer potency (TD50) of a compound based on its structure. Structure activity relationship (SAR) models were developed from the CPDB using two software packages: MultiCASE and VISDOM (Eli Lilly proprietary software). MultiCASE was used to predict a carcinogenic potency class, while VISDOM was used to predict a numerical TD50. For those compounds not categorized as “potent” by MultiCASE, TD50 values were predicted by VISDOM that could then be used in
establishing acceptable exposure levels. For those that were categorized as “potent”, the previously established TTC value of 0.15 µg/day would be used.

**In vitro methods**

1. Classical genotoxicity tests

*Short description, scientific relevance and purpose*

Originally, *in vitro* genotoxicity tests are used to predict the intrinsic potential of substances to induce mutations. The rationale behind using genotoxicity tests for identifying potential carcinogens is that mutations and/or chromosomal aberrations are strongly associated with the carcinogenesis process. For this task only *in vitro* genotoxicity test which measure a mutation endpoint (gene or chromosomal mutation) are qualified for this task: the gene mutation test in bacteria (OECD 471), the gene mutation test in mammalian cells (OECD 487), the chromosome aberration test (OECD 473) and the *in vitro* micronucleus test (OECD 476).

The tests rely on the fixation of initial DNA damage (DNA adducts or chromosomal damage) or damage to the cellular apparatus like the spindle figure into stable irreversible DNA modifications or changes in chromosome number. These modifications may again result in the induction of diseases like cancer or genetic inheritable diseases. The tests are used to predict the potential of chemical substances to induce the former diseases.

*Known users*

Academics for mechanistic studies, all industries for screening purpose but also for regulatory application.

*Status of validation and/or standardization*

With the exception of the *in vitro* micronucleus test (Corvi *et al.*, 2006) none of the genotoxicity tests are formally validated but nonetheless established, scientifically accepted and used tests. For all the tests suggested OECD guidelines exist.

*Fields of application and limitations*

The problem with *in vitro* genotoxicity tests, particularly for the tests measuring chromosome aberrations, is the high number of misleading positives, i.e. positive test results for known non-carcinogens, as was discussed before. Improvement of existing *in vitro* standard genotoxicity tests is under investigation. Preliminary data generated in a project sponsored by ECVAM and predominantly the cosmetic industry show that misleading positive results can be reduced if: 1) *p53*-competent cells (e.g. human lymphocytes, TK6) instead of *p53*-compromised rodent cells (Fowler *et al.*, in press); 2) cytotoxicity measures based on proliferation during treatment instead of measures based simply on cell count (Kirkland and Fowler, in preparation); 3) The top test substance concentration is reduce from 10mM to 1 mM (Parry *et al.*, in press; Kirkland and Fowler, in prep). These modifications are in line with the OECD TG, except for the reduction of the top concentration which would need revision of the TGs for *in vitro* genotoxicity testing.

It remains unclear how carcinogen potency and acceptable human exposure levels will be estimated if a compound is found to be positive in *in vitro* genotoxicity tests and no animal tests are allowed as in the case of the recent ban of animal testing in the cosmetics industry. However, Kirkland demonstrated in a recent analysis of databases for over 950 compounds that when using data from only two *in vitro* genotoxicity tests all of the relevant *in vivo* carcinogens and *in vivo*
genotoxins for which data exist were detected by the 2-test battery, i.e. the sensitivities of the 2- and 3-test batteries are comparable (Kirkland et al., in prep).

**Ongoing developments**

The role of genotoxicity testing can be both qualitative (hazard assessment) and quantitative (risk assessment). A preliminary investigation on the applicability of *in vivo* genotoxicity tests to estimate cancer potency looked promising. (see also the paragraph on *in vivo* genotoxicity test and Hernandez et al. 2010, in prep). For a quantitative approach of *in vitro* genotoxicity tests, a foreseeable problem is the metrics comparison of the correlation, particularly how the dose of *in vitro* studies (in mM) translates to a dose in *in vivo* tests (mg/kg bw/day). For this reason, dose-response analysis of both *in vitro* and *in vivo* genotoxicity endpoints and carcinogenicity is essential. Unfortunately, dose-response analyses using sophisticated dose-response software such as PROAST (RIVM) or the BMDS (USEPA) have never been performed with *in vitro* genotoxicity tests. Given the promising results obtained between *in vivo* genotoxicity and carcinogenicity, it is worthwhile applying a similar approach to investigate whether *in vitro* genotoxicity tests are correlated to carcinogenic potency.

**2. *In vitro* Micronucleus test in 3D human reconstructed skin models (RSMN)**

**Short description, scientific relevance and purpose**

The micronucleus test in 3D human reconstructed skin models (RSMN) offers the potential for a more physiologically relevant approach to test dermal exposure and also including a more relevant exogenous human metabolizing system. It has been anticipated that these features of the reconstituted skin models could improve the predictive value of a genotoxicity assessment compared with that of existing *in vitro* tests and therefore could be used as a follow-up test in case of positive results from the standard *in vitro* genotoxicity testing battery (Maurici et al., 2005). Several 3D skin models are commercially available and are suitable for conducting such test, provided that sufficient cell proliferation is available.

**Status of validation and/or standardisation**

A RSMN protocol using the EpiDerm™ (MatTec Corporation, Ashland, MA, USA) model has been developed and evaluated with a variety of chemicals across three laboratories in the United States (Curren et al., 2006; Mun et al., 2009; Hu et al. 2009). A multi-laboratory prevalidation study was initiated in 2007 and is sponsored and coordinated by COLIPA. This study aims at establishing the reliability of the method and at increasing the domain of chemicals tested for predictive capacity (Aardema et al., 2010). Preliminary results suggest that the RSMN in EpiDerm™ is a valuable *in vitro* method of dermally applied chemicals.

**Fields of application and limitations**

The test is aimed for use at chemicals for which there is dermal exposure. The RSMN test must be seen as an addition to the standard battery of *in vitro* genotoxicity tests. It will be important to demonstrate if these tests have an equivalent sensitivity and a better specificity of the standard *in vitro* micronucleus test.

**On-going development**

Research on the metabolic capacity of the test (Hu et al., 2010) and investigation of the utility of more complex models, such as full-thickness skin models, are ongoing.
3. In vitro Comet assay in 3D human reconstructed skin models

Short description, scientific relevance and purpose

The Comet assay in 3D human reconstructed skin models is considered to be more relevant to evaluate the genotoxic potential of chemicals than when performed in cell cultures, because genotoxic effects can be evaluated under physiological conditions, especially regarding metabolic properties (Hu et al., 2010), and therefore be closer to the human situation than animal testing. This assay is a rapid and sensitive method to evaluate primary DNA damage and it could be used as a follow up test for chemicals that cause gene mutation in the in vitro standard tests (Maurici et al., 2005). Several 3D skin models are commercially available and are suitable for conducting such assay.

Status of validation and/or standardisation

Similarly to the RSMN test, a protocol using the EpiDerm™ model has been developed for the Comet assay in 3D human reconstructed skin models and is being optimised and evaluated across three laboratories in the United States and Europe. This study which aims at establishing the reliability of the method and at increasing the domain of chemicals tested for predictive capacity was initiated in 2007 and is sponsored and coordinated by COLIPA.

Fields of application and limitations

The test is aimed for use at chemicals for which there is dermal exposure. It must be seen as an addition to the standard battery of in vitro genotoxicity tests. Being the endpoint measures very sensitive to DNA damage, it is crucial that the quality of the tissues and good shipping conditions are ensured.

On-going development

Research on the metabolic capacity of the assay (Hu et al., 2010) and investigation of the utility of more complex models, such as full-thickness skin models, are ongoing. Moreover, application of the comet assay to freshly obtained human skin tissue that is generally obtained following cosmetic surgery is under investigation.

4. GreenScreen HC assay

Short description, scientific relevance and purpose

The GreenScreen HC (Gentronix Ltd, Manchester, UK) is a commercially available assay for genotoxicity testing, using human lymphoblastoid TK6 cells transfected with the GADD45a (Growth Arrest and DNA Damage) gene linked to a Green Fluorescent Protein (GFP) reporter (Hastwell et al., 2006). This assay is based on the up-regulation of GADD45a-GFP transcription and the subsequent increase in fluorescence, in response to genome damage and genotoxic stress. The test can be performed with or without metabolic activation by S9 liver fraction.

Status of validation and/or standardisation

Standard protocols have been developed for both methods, without (Hastwell et al., 2006) or with (Jagger et al., 2009) metabolic activation, and their transferability, within-laboratory reproducibility (Hastwell et al., 2006; Jagger et al., 2009), between-laboratory reproducibility (Billinton et al., 2008; 2010) have been evaluated.
Fields of application and limitations

This test is used by the pharmaceutical industry as an early screening tool in drug discovery. However, most pharmaceutical companies are still investigating the utility of the screen in their strategies, and how to interpret the data for internal decision making.

Some technical aspects have also to be taken into account for the conduct of the test: the protocol in the absence of metabolic activation only requires the use of a microplate spectrophotometer and is compatible with high throughput screening, whereas the accessibility and the automation of the S9 protocol are both limited by the necessity of a flow cytometer to avoid interference with the light-absorbing and fluorescent properties of S9 particulates.

On-going development

A variant of the S9 protocol has been developed, that was adapted for microplate readers by the use of a fluorescent cell stain and fluorescence (instead of absorbance) measurement to estimate cell number. Although flow cytometry remains the most sensitive method, this variant is more suitable for non-flow cytometer users and for high throughput screening.

The BlueScreen HC is a new assay under development that uses the same GADD45α reporter gene as the GreenScreen HC assay but linked to Gaussia luciferase gene, which leads to a greater signal-to-noise ratio than with GFP and full compatibility with S9 use and thus with high throughput screening capability.

5. Hens egg test for micronucleus induction (HET-MN)

Short description, scientific relevance and purpose

Another promising system as a follow-up for in vitro positive for cosmetic ingredient is the hens egg test for micronucleus induction (HET-MN; Wolf et al., 2008). The HET-MN combines the use of the commonly accepted genetic endpoint “formation of micronuclei” with the well-characterized and complex model of the incubated hen's egg, which enables metabolic activation, elimination and excretion of xenobiotics, including those that are mutagens or promutagens. This assay procedure is in line with demands for animal protection. The scientific rationale and methodological aspects for this assay as well as results for some well-characterized mutagens and promutagens is provided.

After 8 day of incubation at 37.5°C, the test compounds are applied to the air cell. After another 2.5 days of incubation blood is taken by incising in situ a major vessel. The blood is spread out on microscopic slides, stained and evaluated for micronuclei.

Status of validation and/or standardization

A prevalidation study is planned starting in September 2010 with at least three participating laboratories investigating the transferability and intra-laboratory reproducibility. Results of this study will most probably be available in 2012.

Fields of application and limitations

At present the HET-MN is not frequently used. Only few laboratories have established this test for screening purposes. Studies on metabolism indicate that certain important phase I and II enzymes are active and therefore the detection of liver mutagens is possible. Up to now the transferability and intra-laboratory reproducibility is not provided.

Ongoing developments

An improvement may be the inclusion of flow cytometric analysis where higher cell numbers can be evaluated in a shorter time.
6. Cell transformation assay

Short description, scientific relevance and purpose

Mammalian cell culture systems may be used to detect phenotypic changes in vitro induced by chemical substances associated with malignant transformation in vivo. Widely used cells include SHE, C3H10T1/2 Balb/3T3 and and Bhas 42 cells. The tests rely on changes in cell colony morphology and monolayer focus formation. Less widely used systems exist which detect other physiological or morphological changes in cells following exposure to carcinogenic chemicals. Cytotoxicity is determined by measuring the effect of the test material on colony-forming abilities (cloning efficiency) or growth rates of the cultures.

Status of validation and/or standardization

In 2007 the OECD published a Detailed Review Paper (DRP31) aiming at reviewing all the available data on the 3 main protocols for cell transformation assays and concluded that the performance of the SHE an Balb/c 3T3 were sufficiently adequate and that these assays should be developed into OECD test guidelines. A prevalidation study including the SHE (pH 6.7 and 7.0) and the Balb/c 3T3 was organised by ECVAM to address issues of standardisation of the protocols, transferability and reproducibility. The experimental work was finished in 2009. The data demonstrated that the SHE protocols and the assays system themselves are transferable between laboratories, and are reproducible within- and between laboratories. For the Balb/c 3T3 method an improved protocol has been developed, which allowed to obtain reproducible results. Further testing of this improved protocol is recommended in order to confirm its robustness. Overall, these results in combination with the extensive database summarized in the OECD DRP31 (OECD, 2007) will support the development of the OECD test guidelines for the assessment of carcinogenicity potential. This ongoing work should progress in the coming 3 years.

Fields of application and limitations

The in vitro cell transformation assays have been established in order to predict tumorigenicity (DiPaolo et al., 1969; Isfort et al., 1996; Matthews et al., 1993). Some of the test systems are capable of detecting tumour promoters (Rivedal and Sanner, 1982). Some cell types and substances may require an appropriate external metabolic activation system. When primary cell types are used that possess intrinsic metabolic activity, additional metabolic activation is not used. The scoring of transformed colonies and foci may require some training and experience.

The cell transformation assays are currently used in the pharmaceutical industry and occasionally in the cosmetic industry for clarification of in vitro positive results from genotoxicity assays to be used in the weight of evidence assessment. Data generated by cell transformation assay can be useful where genotoxicity data for a certain substance class have limited predictive capacity (e.g. aromatic amines) or for investigation of compounds with structural alerts for carcinogenicity (industrial chemicals under the REACH regulation) or to demonstrate differences or similarities across a chemical category (industrial chemicals under the REACH regulation). Also the tumour-promoting (non-genotoxic) activity of chemicals (agrochemicals, industrial chemicals, cigarette industry) and the chemopreventive activity (pharmaceuticals) are investigated by the cell transformation assays.

Known users

Academics, pharmaceutical and agrochemical industry for screening purpose, cosmetic industry, and chemical industry also for regulatory application. Academia is using it for mechanistic studies.
Ongoing developments

Certain improvements for investigating the transformed phenotype have been proposed. Transformed colonies can be detected by discrimination of the transformation phenotype by using ATR-FTIR spectroscopy (Walsh et al., 2009), by image analysis (Urani et al., 2009) or by the inclusion of molecular biomarkers (Poth et al., 2007). The technical performance of the SHE assay has been improved by avoiding the use of X-ray irradiated feeder layers (Pant et al., 2008). Systems biology is included for mechanistic investigation of cellular transformation (Ao et al., 2010; Rohrbeck et al., in prep) and also the throughput has been increased by using soft agar colony screening (Thierbach et al., 2009) and Bhas 42 96-well plate method (Ohmori et al., 2005).

7. In vitro toxicogenomics

Short description, scientific relevance and purpose

Since the introduction of genomic technologies ca 10 years ago, their application in toxicology – toxicogenomics – has developed enormously (Ellinger-Ziegelbauer et al., 2009; Guyton et al., 2009; Waters et al., 2010, in press) The unbiased analyses of global perturbations by chemicals in cells and organisms at the level of genes, transcripts, proteins and metabolites, in combination with powerful bioinformatic tools, provides an unprecedented wealth of information about the molecular processes and mechanisms that can be affected. This knowledge can be used for elucidating the mode-of-action of compounds, prediction of toxic properties, cross-species and in vitro-in vivo comparison, and even in epidemiological settings for assessment of exposure and (adverse) effects in humans. Predominantly transcriptomics (gene expression analyses at the level of mRNA) has received most attention and has proven to be promising. For in vitro hazard assessment in the area of genotoxicity and carcinogenicity TK6, HepG2, and liver cells are mostly used as cell models (Ellinger-Ziegelbauer et al., 2009; Li et al., 2007; Tsujimura et al., 2006; Le Fevre et al., 2007; Hu et al., 2004; Mathijs et al., 2010). These toxicogenomics approaches reach 80-90% accuracy for predicting in vivo toxicity in rodents.

Known users

Academics. Pharmaceutical and cosmetic industry in screening purposes only.

Status of validation and/or standardization

No formal validation of a method has been performed. For some tests based on gene expression analyses standard protocols are being developed and optimised (see “Ongoing developments”). This ongoing work will progress in the coming 3 years; depending on the results and conclusions of these studies, some tests might be ready to enter prevalidation.

Fields of application and limitations

Various tests based on gene expression analyses can be foreseen for the near future, which can be used for screening purposes and labelling of compounds, thus only for hazard assessment. Tests for genotoxicity and carcinogenicity in general, or for specific mechanisms therein, are under development [4-8] (Mathijs et al., 2010 in press). The transcriptomic biomarkers will be complex, consisting of profiles for multiple genes. As many genes have been annotated with respect to their function and sometimes to toxicological pathways (e.g. the DNA damage pathway), this will provide mechanistic information as well. Some cell types and substances may require an appropriate external metabolic activation system, but in general cells derived from liver are used that possess intrinsic metabolic activity and additional metabolic activation is not used. The limitations are many, such as risk assessment is problematic, each assay focuses on a specific aspect
of genotoxicity or carcinogenicity, limited public accessibility of raw data, the mechanism of not all
genes in the prediction sets are understood, lack of uniformity in study design (e.g. cell lines, dose
setting criteria, time points, repeats, etc.) and bioinformatic analyses, and the requirement of
expensive equipment and specialized staff.

_**Ongoing developments**_

Recently, a multi-laboratory project coordinated by the Health and Environmental Sciences Institute
demonstrated that expression analysis by RT-PCR of a relevant gene set derived from omics data, is
capable of distinguishing compounds that are DNA reactive genotoxins from those that non-DNA
reactive genotoxins (Ellinger-Ziegelbauer _et al._, 2009). RT-PCR provide a cheaper and faster test
for gene expression profiling, when limited to relative small gene sets. Furthermore, as part of the
EU-funded project _carcinoGENOMICS_, certain aspects related to reliability of the tests will be
addressed in 2010-2011(www.carcinogenomics.eu).

_**In vivo methods (Reduction / refinement)**_

Since no replacement methods are to date available in the area of carcinogenicity, we considered
also reduction and refinement methods. As alternatives we mention only those tests which are
considered as predictive tests for carcinogenicity and those which continue to undergo further
development. Rather old approaches like the liver foci assay and the neonatal mouse assay are for
this reason not described further.

1. _**In vivo genotoxicity tests**_

For the same reason as for _in vitro_ tests, also _in vivo_ genotoxicity tests may be a tool to predict
cancer risk assessment. _In vivo_ tests may even be a better predictor of carcinogenicity tests since in
these test the role ADME is implicated. As for the _in vitro_ genotoxicity tests, also for _in vivo_
genotoxicity test the specificity and predictivity have to be at a level that the prediction of
carcinogenicity is justified. This again may lead to modifications of test protocols, _i.e._ species, (top)
doses among others. But then a positive result in an _in vivo_ genotoxicity test may point to a
carcinogenic potency of the compound under investigation and further carcinogenicity testing may
not be needed. As for the _in vitro_ tests predominantly tests which measure irreversible genotoxic
damage may be considered: the chromosome aberration test (OECD 475), the micronucleus test
(OECD 474) and the gene mutation test with transgenic animals (OECD guideline in prep). However, as an exception the Comet assay (OECD guideline in prep), measuring single and double
strand breaks may be considered.

The role of genotoxicity testing can be both qualitative (hazard assessment) and quantitative (risk
assessment). A preliminary investigation on the applicability of genotoxicity tests to estimate
cancer potency was undertaken in the RIVM using the benchmark dose approach and positive
correlations between _in vivo_ genotoxicity (micronucleus test and transgenic rodent mutation test)
and carcinogenic potency were found (Hernandez _et al._ 2010, in prep). Dose-response analyses
using sophisticated dose-response software such as PROAST (RIVM) or the BMDS (USEPA) was
used. The results suggest that _in vivo_ genotoxicity tests may be used to estimate carcinogenic
potency.
2. Transgenic mouse models

Short-term tests with transgenic mouse models (p53⁺/-, rasH2, Tg.AC, Xpa⁻/+ and Xpa⁻/-p53⁺/) are a good alternative for the classical 2-year cancer bioassay (Ashby 2001). The rationale for using transgenic mice in regulatory carcinogenicity testing is that transgenic mouse models may be more sensitive predictors of carcinogenic risk to humans. Indeed these transgenic mouse models had a reduced tumor latency period (6-9 months) to chemically-induced tumors (Marx 2003). The increased sensitivity to tumor formation in transgenic mouse models is primarily due to modifications in the mouse genome by either removing or adding specific genetic material (Tennant 1995; 1999). Although not a complete replacement to the rodent 2-year cancer bioassay, transgenic mouse models are a refinement and result in a significant reduction in the use of experimental animals.

Several studies (ILSI/HESI ACT 2001; Eastin 1998; Bucher 1998; Pritchard 2003; de Vries et al., 2004) demonstrate that in all transgenic models a limited number of animals 20 – 25 animals/sex/treatment group can be used and that an exposure of 6 – 9 months is sufficient. However, wild type animals should be included in test battery to demonstrate that no genetic drift may affect interpretation of the results. Transgenic mouse models showed a high specificity given that all non-carcinogens tested gave negative results in all 5 transgenic models. These findings provide evidence against “oversensitivity” concerns associated with transgenic mouse models due to modifications in cancer-related genes. Transgenic models were able to discriminate not only between carcinogens and non-carcinogens but even between rodent carcinogens and putative human non-carcinogens to a high degree of accuracy.

3. In vivo toxicogenomics

In the paragraph on in vitro toxicogenomics an introduction is given on relevance of genomics technologies and their application in toxicology. Also in case of in vivo toxicogenomics, gene expression analyses (transcriptomics) has been developed furthest. Most in vivo toxicogenomics studies on assessment of carcinogenicity, focus – but do not limit – on short term rat studies and non-genotoxic hepato-carcinogenicity (Ellinger-Ziegelbauer et al., 2008; Nie et al., 2006; Fielden et al., 2007; Stemmer et al., 2007; Nioi et al., 2008; Uehara et al., 2008; Jonker et al., 2009). These toxicogenomics approaches reach 80-90% accuracy for predicting rodent carcinogenicity. Pharmaceutical and sometimes the chemical industry are the main users in screening purposes. In rare cases the pharmaceutical industry also uses in vivo toxicogenomics for mechanistic purposes.

Various assays based on gene expression analyses can be foreseen for the near future, which can be used for screening/prioritization purposes and for labelling of compounds. These assays can also be helpful for the understanding of modes of action. Especially assays for non-genotoxic hepato-carcinogenicity are under development (Ellinger-Ziegelbauer et al., 2008; Nie et al., 2006; Fielden et al., 2007; Stemmer et al., 2007; Nioi et al., 2008; Uehara et al., 2008; Jonker et al., 2009). Like for in vitro toxicogenomics, the limitations are many, such as quantitative risk assessment is in its infancy, strong focus on non-genotoxic carcinogens (mainly pharmaceuticals) and on the liver as target organ, limited public accessibility of raw data, the mechanism of not all genes in the prediction sets are understood, lack of uniformity in study design (e.g. rodent species and strain, dose setting criteria, time points, repeats, etc.) and bio-informatic analyses, and the requirement of expensive equipment and specialized staff.
The in vivo toxicogenomics assays could be very helpful for hazard assessment and by that may lead to a reduction in the number of bioassays and the number of animals in the remaining in vivo tests. The number of animals required for toxicogenomics based assays are at least 10-fold smaller than for the rodent bio-assay, and the exposure periods last maximally four weeks instead of two years.

So far, no formal validation of a method has been performed. A gene expression profile for rat hepato-carcinogenicity is being investigated for some aspects of reliability. This ongoing work will progress in the coming 3 years; depending on the results and conclusions of these studies, some tests may might be ready to enter pre-validation. The Predictive Safety Testing Consortium of the Critical Path Institute evaluated the predictivity of two published hepatic gene expression signatures when sharing the data (Fielden et al., 2008). Based on the results, a QPCR-based signature has been derived and is currently evaluated for inter-laboratory precision, sensitivity and specificity, time-dependency, and non-genotoxic versus genotoxic mechanisms.

Identify Areas with no Alternative Methods Available and Related Scientific/Technical Difficulties

This report has highlighted a number of in vivo studies that have historically been used in the safety evaluation of cosmetics with respect to carcinogenicity. For that endpoint, the gold standard of a two-year bioassay is not commonly used, but several shorter-term studies are conducted, including in vitro and in vivo genotoxicity studies, repeat-dose studies, and other mechanistic studies used in the safety assessment of non-genotoxic compounds. Therefore, the animal testing ban has implications well beyond the two-year bioassay for the evaluation of cancer hazard and risk. The first challenge is for in vitro genotoxic compounds, where the strategy until 2009 allowed for in vivo testing. That was important to fully characterize the potential genotoxicity and with that indicate to carcinogenicity. The second challenge lies beyond 2013 where all in vivo testing is banned. From that year on, and in case nothing would change in available test methods/strategies, we can only rely on in vitro tests which test for the intrinsic genotoxic potential with a substantial rate of misleading positives in the classical tests. Consequently, the characterization of the carcinogenic potential of substances will not be of the same quality as today. The difficulties with the in vitro tests have been described by Kirkland et al. (2007) and work is on going to improve these assays.

Especially for non-genotoxic chemicals, a ban on repeat-dose toxicity testing raises new questions safety can be assured. It is noted that this is the same challenge posed to those charged with developing alternatives for target organ toxicity. Some in vitro tests have been developed to better understand non-genotoxic modes of action, such as the induction of oxidative stress and the inhibition of gap junction intercellular communication; however, these tests are primarily used for evaluating mechanism and cannot currently be used to predict whether a chemical will be a carcinogen and under what conditions. Cell transformation assays have been developed as a tool to identify both genotoxic and non-genotoxic carcinogens, but are not fully validated. Toxicogenomics is an emerging area and also offers promises for the detection of non-genotoxic carcinogens, but still stands in its infancy. The challenges of extrapolation from in vitro to in vivo are also that most in vitro studies are limited to use in hazard identification and cannot yet be used in risk assessment. These have been described by Blaauboer (2010) and are also covered by the working group which develops the alternatives for toxicokinetics.
In silico methods such as QSAR have proven successful at predicting genotoxic potential and rationalising the chemical basis in terms of DNA reactivity. Such QSARs can be as reliable and informative as the Ames test, provided that their predictive algorithms and applicability domains are well characterized (Benigni et al., 2010). More research is needed to understand how the applicability domains relates to the chemical classes used in cosmetics. An advantage of the QSAR approach is that the models can be tuned to meet user-defined performance criteria such as low false positive or negative rates, depending on their foreseen use in a testing strategy. Relatively few QSARs are available for non-genotoxic carcinogenicity and carcinogenic potency, and this represents a knowledge/development gap. The category and read-across approach provides a promising means of filling qualitative and quantitative data gaps (van Leeuwen et al., 2009). However, specialized knowledge and tools are needed to build the category and draw conclusions on the adequacy of the read-across. It is not possible to validate this approach a priori. In addition to being a stand-alone approach, read-across can also be used to add confidence to a prediction generated by QSAR. Additional confidence can be provided by in vitro data. At present, it is recommended to apply QSAR and read-across within the context of a WoE or TTC approach. This implies the use of multiple QSARs in combination with each other (e.g. Matthews et al, 2008) and if possible with in vitro tests (e.g. Peer Consultation on Health Canada Draft Weight of Evidence Framework for Genotoxic Carcinogenicity, 2005).

In order to test a cosmetic ingredient for the evaluation of its carcinogenic potential a tier-testing approach COLIPA (Pfuhler et al, 2010) has proposed a tiered testing strategy focused on genotoxic carcinogens for cosmetic ingredients for beyond 2013. However, this strategy does not take into account non-genotoxic carcinogens. No strategy is currently in place to detect non-genotoxic carcinogens. In the past the safely assessment for non-genotoxic chemicals has been based on identification of a NO(A)EL from repeat dose toxicity studies. If in the future it will no longer be possible to conduct those studies for cosmetics ingredients, there will be a lack of information requirements that cover the non-genotoxic endpoint. Given that some non-genotoxic carcinogens are known human carcinogens and the potential hazard associated with them, there is a need for the development of alternative methods for the detection and risk assessment of non-genotoxic carcinogens. Cell transformation assays have the potential to be used as a tool to identify non-genotoxic carcinogens, whereas the emerging field of toxicogenomics might also offer some promise for the detection of non-genotoxic carcinogens. However, neither of these can currently be used to support quantitative risk assessment, as that currently relies on in vivo data.

Conclusions

As a general limitation, cancer is a long-term process that is to day despite best efforts impossible to mimic with relatively short-term in vitro tests.

An in vivo testing is no longer possible, the safety of many potential new cosmetic ingredients will not be able to be substantiated and will therefore not be allowed.

The process of carcinogenesis is recognised as resulting from a sequence of stages and complex biological interactions. It is also recognized that there are many different modes of action that can contribute to the carcinogenic process and that even for one chemical, the mode of action can be different in different target organs, or in different species. Despite best efforts, the modeling of such complex adverse effects cannot fully be accomplished at present by the use of non-animal tests.
For genotoxic chemicals, a number of in vitro genotoxicity tests are available that are currently used to screen chemicals for activity that is considered to be predictive of potential carcinogenicity. While these tests have good sensitivity, some (especially the in vitro mammalian cell tests) also have a high rate of misleading positives. Prior to 2009, a positive finding in an in vitro study was commonly followed by an in vivo study that was of critical importance for clarifying the in vitro results. Indeed, the vast majority of compounds tested in vivo were negative. Because of the 2009 ban on in vivo genotoxicity testing, the situation is now problematic and many potential cosmetic ingredients may be lost because of an inability to clarify misleading positive results from an in vitro genotoxicity test. Work is ongoing to improve these in vitro tests.

Cell transformation assays are to date the only in vitro tests that have reached a certain level of standardisation and have the potential to detect both genotoxic and non-genotoxic carcinogens. However, at the moment these assays cannot be considered as a stand alone solution to detect human carcinogens, but have the potential to contribute to a weight of evidence approach. Importantly, these assays are currently useful only in the hazard identification of carcinogens; there are no methods yet to use data from these tests to support a risk assessment.

For non-genotoxic chemicals, the standard approach for risk assessment (globally and for all sectors, not limited to cosmetics) has been to assume a threshold; for cosmetic ingredients in general on repeated dose toxicity studies. A NOAEL from a repeat dose toxicity study, along with appropriate conservative safety factors, has been used to support a risk assessment for these chemicals, including the risk for carcinogenicity. When repeat-dose toxicity testing is banned in 2013, methods for quantitative assessment of non-genotoxic carcinogenic risks will be limited to tools such as read-across, (Q)SAR and TTC. They are, however, not sufficiently well developed and/or validated to be able to serve as a suitable basis for risk assessment for most new chemicals. Because of limitations, some methods such as TTC and read-across may default to using very conservative assumptions that further limit the utility of these approaches.

Although many in vitro short term tests are available beyond the standard in vitro genetox tests to support conclusions on cancer hazard identification, the in vitro short-term tests will not be sufficient to fully replace the animal tests needed to perform risk assessment for carcinogenicity for cosmetic ingredients.

Taking into consideration the present state of the art of the non-animal methods, the experts were unable to suggest a timeline for full replacement of animal tests currently needed to fully evaluate carcinogenic risk of chemicals. Although a timeline for full replacement cannot be developed, clearly the timeline is expected to extend well past 2013.

References


of branched- and straight-chain aliphatic saturated primary alcohols and branched- and straight-chain saturated aldehydes, and an orthoester of formic acid, from chemical groups 1 and 2.


Pfuhler, S., Kirst, A., Aardema, A., Banduhn, N., Goebel, C., Araki, D., Costabel-Farkas, F.,
Dufour, E., Fautz, R., Harvey, J., Hewitt, N.J., Hibatallah, J., Carmichael, P., Macfarlane, M.,
of alternatives to animal testing for the safety assessment of cosmetics: Genotoxicity. A COLIPA
analysis. *Regulatory Toxicology and Pharmacology*, Doi:10.1016/j.yrtph.2010.03.012

Peer Consultation on Health Canada Draft Weight of Evidence Framework for Genotoxic

Poth, A., Kunz, S. & Heppenheimer, A. (2007) Bhas cell transformation assay as a predictor of
carcinogenicity, abstract presented at the 6th World Congress on alternatives and animal use in the
life sciences, Tokyo.


Purdy, R. (1996). A Mechanism-mediated Model for Carcinogenicity: Model Content and
Prediction of the Outcome of Rodent Carcinogenicity Bioassays Currently Being Conducted on 25

Renwick, A.G., (2004). Toxicology databases and the concept of thresholds of toxicological
concern as used by the JECFA for the safety evaluation of flavouring agents. *Toxicol Lett* **149**, 223–
234.

Ridings, J.E., Barratt, M.D., Cary, R., Earnshaw, C.G., Eggington, C.E., Ellis, M.K., Judson, P.N.,
prediction of possible toxic action from chemical structure: an update on the DEREK system.
*Toxicology* **106**, 267-279.


Rohrbeck, A., Salinas, G., Maaser, K., Linge, J., Salovaara, S., Corvi, R. & Borlak, J. In vitro
carcinogenicity testing with Balb/c 3T3 cells treated with various chemical carcinogens. *In
preparation.*


Safford, R.J. (2008). The Dermal Sensitisation Threshold—A TTC approach for allergic contact

chemical structure; the DEREK system. *Human and Experimental Toxicology* **10** 261-273.

Scientific Committee on Consumer Products. The SCCP’s Notes of Guidance for the Testing of
Cosmetic Ingredients and Their Safety Evaluation. 6th Revision. Adopted by the SCCP during the
10th plenary meeting of 19 December.


### Table 1. Summary of identified alternative non-animal methods for carcinogenicity

<table>
<thead>
<tr>
<th>Current endpoints addressed in animal test</th>
<th>Alternative tests available</th>
<th>Part of mechanism covered</th>
<th>Area(s) of application</th>
<th>Status (R&amp;D, optimisation, prevalidation, validation, regulatory acceptance)</th>
<th>Comments</th>
<th>Estimated time until entry into pre-validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotoxicity and carcinogenicity</td>
<td>QSAR and expert systems</td>
<td>Mainly genotoxicity and carcinogenicity based on DNA reactivity. Mostly classification models</td>
<td>Mainly screening, but also provide a means of filling data gaps in hazard assessment</td>
<td>Models and software tools implementing them are subject to ongoing development</td>
<td>These are <em>in silico</em> tools that generate predictions based on chemical structure; they are not experimental test methods.</td>
<td>n/a</td>
</tr>
<tr>
<td>Genotoxicity and carcinogenicity</td>
<td>Category formation and read-across approach</td>
<td>Depends on choice of category members</td>
<td>Grouping and read-across provides a means of filling data gaps in both hazard and risk assessment</td>
<td>Various software tools are available to carry out grouping and read-across</td>
<td>These are <em>in silico</em> tools that support grouping and read-across.</td>
<td>n/a</td>
</tr>
<tr>
<td>Genotoxic and non-genotoxic carcinogenicity</td>
<td>TTC approach</td>
<td>TTC is a statistically-based approach to establish a conservative default risk value based on worst-case assumptions about the chemical in the absence of data. It is not an assay that is intended to replace current testing strategies.</td>
<td>Screening, Conservative risk assessment tool based on distribution of risk values (one distribution for genotoxic compounds and a separate distribution for non-genotoxic compounds)</td>
<td>TTC has regulatory acceptance as a risk tool in the U.S. for food packaging materials and in the U.S. and Europe to set acceptable exposure limits for genotoxic impurities in drugs. It has not yet been granted regulatory acceptance for use in cosmetics.</td>
<td>Although TTC has the potential to be a useful risk assessment tool for very low level exposures, it is not expected to be useful for cosmetic ingredients used at higher levels or in product types associated with higher consumer exposures (e.g., body lotion).</td>
<td>n/a</td>
</tr>
<tr>
<td>Genotoxicity</td>
<td>Assay Type</td>
<td>Mutations/Effects</td>
<td>Screening</td>
<td>Hazard Identification</td>
<td>OECD TG</td>
<td>Remarks</td>
</tr>
<tr>
<td>-------------</td>
<td>---------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>-----------</td>
<td>-----------------------</td>
<td>---------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Genotoxicity</td>
<td><strong>Bacterial Reverse Mutation Assay in</strong> <em>Salmonella typhimurium</em> and <em>Escherichia coli</em></td>
<td>Gene mutations (point mutations, base pair substitutions and frameshift mutations)</td>
<td>Screening</td>
<td>Hazard identification</td>
<td>OECD TG 471</td>
<td></td>
</tr>
<tr>
<td>Genotoxicity</td>
<td><strong>In Vitro Gene Mutation Assay in Mammalian Cells</strong></td>
<td>Gene mutations (point mutations, base pair substitutions, frameshift mutations) Structural and numerical chromosome damage (Mouse Lymphoma L5178Y cells)</td>
<td>Screening</td>
<td>Hazard identification</td>
<td>OECD TG 476</td>
<td>Problems related to low specificity of the test (exception of HPRT) Guideline available</td>
</tr>
<tr>
<td>Genotoxicity</td>
<td><strong>In vitro Micronucleus Assay in Mammalian Cells</strong></td>
<td>Structural and numerical chromosome damage in mammalian cells (i.e. clastogenicity and aneuploidy)</td>
<td>Screening</td>
<td>Hazard identification</td>
<td>OECD TG 487</td>
<td>Problems related to low specificity of the test Guideline available</td>
</tr>
<tr>
<td>Genotoxicity</td>
<td><strong>In vitro Chromosome Aberration Assay in Mammalian Cells</strong></td>
<td>Structural and numerical chromosome damage in mammalian cells (i.e. clastogenicity and polyploidy)</td>
<td>Screening</td>
<td>Hazard identification</td>
<td>OECD TG 473</td>
<td>Problems related to low specificity of the test Guideline available</td>
</tr>
<tr>
<td>Genotoxicity</td>
<td>Comet assay in human 3D skin model</td>
<td>DNA damage</td>
<td>Hazard identification</td>
<td>Clarification of in vitro genotoxic positives</td>
<td>Optimisation</td>
<td>Additional work is ongoing with further human 3D skin models 1-3 years</td>
</tr>
</tbody>
</table>

**OECD TG 471** available

**OECD TG 476** available

**OECD TG 477** available

**OECD TG 487** available

**Comet assay in human 3D skin model** additional work is ongoing with further human 3D skin models 1-3 years
<table>
<thead>
<tr>
<th>Genotoxicity</th>
<th>Micronucleus assay in human 3D skin model</th>
<th>Structural and numerical chromosome damage in mammalian cells (i.e. clastogenicity and aneuploidy)</th>
<th>Hazard identification</th>
<th>Prevalidation ongoing for one of the epidrmis human models</th>
<th>Additional work is ongoing with further human 3D skin models</th>
<th>0-3 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotoxicity</td>
<td>GreenScreen HC</td>
<td>p53-dependent up-regulation of GADD45a expression in response to DNA damage</td>
<td>Screening</td>
<td>Optimisation</td>
<td></td>
<td>1-3 years</td>
</tr>
<tr>
<td>Genotoxicity</td>
<td>Hens egg test for micronucleus induction (HET-MN)</td>
<td>Structural and numerical chromosome damage in mammalian cells (i.e. clastogenicity and aneuploidy)</td>
<td>Screening Hazard identification</td>
<td>Optimisation</td>
<td></td>
<td>1-4 years</td>
</tr>
<tr>
<td>Genotoxic and non-genotoxic carcinogenicity</td>
<td>Cell transformation assays (CTA)</td>
<td>Transformation</td>
<td>Screening Clarification of in vitro genotoxic positives Hazard identification Identifying promoters Chemopreventive activity Mechanistic studies</td>
<td>Prevalidation completed for SHE and Balb/c 3T3 Prevalidation ongoing for Bhas 42</td>
<td>Standard protocol developed for SHE, to be considered for drafting the OECD TG</td>
<td>0 year</td>
</tr>
<tr>
<td>Genotoxic and non-genotoxic carcinogenicity</td>
<td>In vitro toxicogenomics</td>
<td>Expression changes of genes and gene sets (biological pathways)</td>
<td>Screening Clarification of in vitro genotoxic positives Hazard identification Mechanistic studies</td>
<td>R&amp;D Optimisation</td>
<td></td>
<td>1-5 years</td>
</tr>
</tbody>
</table>

1639 n/a not applicable
1640
1641 This table shows an approximated estimation of the time needed to enter prevalidation according to the ECVAM criteria for individual non-animal tests, assuming optimal conditions. It does not indicate the time needed to achieve full replacement of the animal tests, nor does it include
the time to achieve regulatory acceptance. "Optimal conditions" means that all necessary resources are available and that the studies undertaken have successful outcomes.