Regulatory Toxicology and Pharmacology 57 (2010) 315–324

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

A tiered approach to the use of alternatives to animal testing for the safety assessment of cosmetics: Genotoxicity. A COLIPA analysis

Stefan Pfuhler [a], Annette Kirst [b], Marilyn Aardema [a,1], Norbert Banduhn [c], Carsten Goebel [d], Daisuke Araki [e], Margit Costabel-Farkas [f], Eric Dufour [g], Rolf Fautz [b], James Harvey [h], Nicola J. Hewitt [i], Jalila Hibatallah [j], Paul Carmichael [k], Martin Macfarlane [k], Kerstin Reisinger [c], Joanna Rowland [l], Florian Schellauf [m], Andreas Schepky [n], Julia Scheel [c,*]

[a] *Procter & Gamble, Miami Valley Laboratories, 11810 East Miami River Road, Cincinnati, 45253 OH, USA*
[b] *KPSS – Kao Professional Salon Services GmbH, Pfungstaedter Strasse 92-100, 64297 Darmstadt, Germany*
[c] *Henkel AG & Co. KGaA, Henkelstraße 67, 40589 Düsseldorf, Germany*
[d] *Procter & Gamble, Darmstadt Innovation Center, Berliner Allee 65, 64274 Darmstadt, Germany*
[e] *Kanebo Cosmetics, 89/91 Rue du Faubourg St Honoré, 75008 Paris, France*
[f] *Johnson & Johnson GmbH, Johnson&Johnson-Platz 2, D-41470 Neuss, Germany*
[g] *L'Oreal, River Plaza 25-29, quai Aulagnier 92600 Asnières-sur-Seine, France*
[h] *GlaxoSmithKline R&D, Ware, Hertfordshire, UK*
[i] *Scientific Writing Services, Wingertstrasse 25, 64390 Erzhausen, Germany*
[j] *Chanel Parfums Beaute 135, Avenue Charles de Gaulle, 92525 Neuilly sur Seine, France*
[k] *Unilever, Colworth Science Park, Sharnbrook, Bedford MK44 1LQ, UK*
[l] *GlaxoSmithKline Consumer HealthCare R&D, Weybridge, Surrey KT13 ODE, UK*
[m] *COLIPA, Avenue Herrmann Debroux 15A, B-1160 Auderghem, Brussels, Belgium*
[n] *Beiersdorf AG, BF 682, Unnastrasse 48, 20245 Hamburg, Germany*

ARTICLE INFO

Article history:
Received 19 January 2010
Available online 9 April 2010

Keywords:
Genotoxicity
Cosmetic ingredients
Alternative methods
Hazard assessment
Safety assessment

ABSTRACT

For the assessment of genotoxic effects of cosmetic ingredients, a number of well-established and regulatory accepted *in vitro* assays are in place. A caveat to the use of these assays is their relatively low specificity and high rate of false or misleading positive results. Due to the 7th amendment to the EU Cosmetics Directive ban on *in vivo* genotoxicity testing for cosmetics that was enacted March 2009, it is no longer possible to conduct follow-up *in vivo* genotoxicity tests for cosmetic ingredients positive in *in vitro* genotoxicity tests to further assess the relevance of the *in vitro* findings. COLIPA, the European Cosmetics Association, has initiated a research programme to improve existing and develop new *in vitro* methods. A COLIPA workshop was held in Brussels in April 2008 to analyse the best possible use of available methods and approaches to enable a sound assessment of the genotoxic hazard of cosmetic ingredients. Common approaches of cosmetic companies are described, with recommendations for evaluating *in vitro* genotoxins using non-animal approaches. A weight of evidence approach was employed to set up a decision-tree for the integration of alternative methods into tiered testing strategies.

1. Introduction

In vitro tests form an essential part of the assessment of genotoxicity and provide information on three major genetic endpoints, namely (1) mutagenicity at a gene level, (2) chromosome breakage and/or rearrangements (clastogenicity), and (3) numerical chromosome aberrations (CA)[2] (aneugenicity) (SCCP, 2006a,b; Mueller et al., 2003; Dearfield et al., 2002; COM, 2000). In the past, *in vivo*

* Corresponding author. Fax: +49 211 798 12413.
E-mail address: julia.scheel@henkel.com (J. Scheel).
[1] Present address: Consulting LLC, 5315 Oakbrook Dr. Fairfield OH, 45014, USA.
[2] *Abbreviations used:* ADME, absorption, distribution, metabolism and excretion; CA, chromosome aberration; CHO, Chinese hamster ovary; COLIPA, The European Cosmetic Association; DEREK, deductive estimation of risk from existing knowledge; ECVAM, European Centre for the Validation of Alternative Methods; HPRT, hypoxanthine-guanine phosphoribosyl transferase; ICH, The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use; K_M, Michaelis constant; MLA, mouse lymphoma assay; MNT, micronucleus test; MPs, misleading positives; OECD, organisation for economic co-operation and development; RDT, repeated dose toxicity; REACh, regulation, evaluation, authorisation of chemicals; RS, reconstructed skin; RSMN, reconstructed skin micronucleus test; SCCP, Scientific Committee on Consumer Products; SCCS, Scientific Committee on Consumer Safety (formerly SCCP); SHE, Syrian Hamster Embryo; TG, test guideline; TK, thymidine kinase; TTC, threshold of toxicological concern; UK NC3Rs, The National Centre for the Replacement, Refinement and Reduction of Animals in Research; WoE, weight of evidence.

0273-2300/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.yrtph.2010.03.012

genotoxicity studies were used to further assess the relevance of positive *in vitro* findings for cosmetic ingredients. Due to the 7th amendment to the EU Cosmetics Directive testing ban (EU, 2003) that was enacted March 2009, it is no longer possible to conduct follow-up *in vivo* genotoxicity in this area. To address this, a workshop of COLIPA (The European Cosmetics Association) Safety Assessment and Genotoxicity project teams was held in Brussels on 3rd April 2008. Participants included members from a number of global cosmetic companies. The aim of the meeting was to (a) discuss company perspectives and current practices for the safety assessment of genotoxicity of cosmetic ingredients and (b) design a decision tree approach to the safety assessment of their potential to cause genotoxicity, with emphasis on non-animal methods. The outcome of this meeting is reported herein, starting with a review of the specific scientific challenges related to *in vitro* testing and the impact of the 7th Amendment. We describe research projects that have been undertaken by the COLIPA Genotoxicity Project Team to address the major limitations of the current *in vitro* testing paradigm. A proposed strategy is provided for the use of the non-animal methods to enable a thorough assessment of the genotoxic hazard of cosmetic ingredients.

2. Genotoxicity testing of cosmetic ingredients – challenges and approaches

2.1. In vitro genotoxicity testing leads to high percentage of misleading positive results

Due to the diverse nature of the mechanisms involved in genotoxicity, it is known that no single mutagenicity test can detect all classes and examples of genotoxic carcinogens. As a result, international guidelines for assessing the genotoxic potential of chemicals recommend the use of a battery of mutagenicity tests to detect gene, chromosome or genome mutations (Eastmond et al., 2009; Regulation, Evaluation, Authorisation of Chemicals (REACh); International Conference on Harmonisation (ICH), 2008; SCCP, 2006a,b, COM, 2000). For cosmetic ingredients, the Scientific Committee on Consumer Safety (SCCS, formerly the Scientific committee on Consumer Products (SCCP)) is the expert panel mandated by the European Commission to develop opinions for testing, review dossier submissions, and provide opinions concerning all types of safety risks. The SCCS recommended basic test battery for testing cosmetic ingredients for their genotoxic potential is: (1) two tests for gene mutation, the bacterial reverse mutation or "Ames" test (Organisation for Economic Co-operation and Development Test Guidelines (OECD TG) 471, 1997) and an *in vitro* gene mutation assay in mammalian cells (OECD TG 476, 1997) and (2) a test for clastogenicity and aneugenicity using the *in vitro* micronucleus test (MNT) (OECD TG 487, in development) (SCCP, 2006a,b). If all these tests are negative then no further testing is required.

The sensitivity and specificity of a number of *in vitro* genotoxicity assays in terms of predicting rodent carcinogenicity are shown in Table 1. Kirkland et al. (2005a, 2006) evaluated the predictivity of four standard *in vitro* tests for rodent carcinogenicity. The sensitivity of the MNT (i.e. ability to give positive results with rodent carcinogens) was the highest of the four individual tests analysed (though the database was smaller than the other assays) and the addition of the Ames assay increased the sensitivity further. For example, the sensitivity of the Ames assay was increased from 58.8% to 85.9% and 75.3% when it was combined with the mouse lymphoma assay (MLA), MNT or CA assay, respectively. Unfortunately, the specificity (the ability to correctly identify non-carcinogens) is greatly decreased with the addition of *in vitro* tests in a battery. For instance, the specificity of the Ames assay ranges from 74% to 80% (Table 1) and combining the Ames test with two other tests as required in the SCCS battery decreases the specificity to a low as 5–23% (Table 1). It is important to note that for some chem-

Table 1
A comparison of the sensitivity and specificity of some of the *in vitro* genotoxicity assays currently available. Taken from Kirkland et al., 2005b.

Test	OECD TG	Sensitivity to rodent carcinogens (%)	Specificity to rodent carcinogens (%)	Reference
Bacterial reverse mutation test, Ames test	471[a]	58.8	73.9	Kirkland et al. (2005a)
		45		Tennant et al. (1987)
		54		Zeiger (1998)
		49.4	80.3	Matthews et al. (2006)
In vitro micronucleus test (MNT)	Draft 487[b]	78.7	30.8	Kirkland et al. (2005a)
		87.3	23.1	Matthews et al. (2006)
		89.2	55.0	Corvi et al. (2008)
In vitro mammalian cell gene mutation test Mouse lymphoma assay (MLA) and hypoxanthine–guanine phosphoribosyl transferase (HPRT) test	476[c]	MLA/TK:	MLA/TK:	
		73.1	39.0	Kirkland et al. (2005a)
		70.9	57.8	Zeiger (1998)
		62.8, 70.9	44.2	Matthews et al. (2006)
		HPRT/CHO:		
		48.4	65.2	Matthews et al., 2006
In vitro mammalian chromosomal aberration assay	473[d]	65.6	44.9	Kirkland et al. (2005a)
		55.3	63.3	Matthews et al. (2006)
Ames + MNT		85.9	121.0	Kirkland et al. (2005a)
Ames + CA		75.3	34.6	Kirkland et al. (2005a)
Ames + MLA + MNT		90.7	5.0	Kirkland et al. (2005a)
Ames + MLA + CA		84.7	22.9	Kirkland et al. (2005a)
In vitro Syrian hamster embryo (SHE) cell transformation assay	Draft 495[e]	87	83	LeBoeuf et al. (1996) (24 h and 7 day)
		66	85	OECD DRP (pH 6.7; 7 day only)
		92	66	OECD DRP (pH 7.0; 7 day only)

[a] OECD TG 471 (1997).
[b] OECD TG 487 (2009).
[c] OECD TG 476 (1997).
[d] OECD TG 473 (1997).
[e] OECD TG 495.

ical classes such as aromatic amines (ingredients in hair dyes) the specificity of the standard *in vitro* genotoxicity tests is even lower, rendering them essentially uninformative. For instance, specificity for rodent non-carcinogenic aromatic amines in the Ames assay is only 29% (Ashby and Tennant, 1991), indicating 71% misleading positives (MPs) in this single assay. Combining the Ames assay with other *in vitro* mammalian cell assays for aromatic amines will therefore result in almost all of these chemicals yielding a positive result in one or more assays.

Clearly, protection of consumers by employing sensitive tests to evaluate the potential genotoxicity of compounds is paramount, but the use of tests with such low specificities means that unacceptably high numbers of MPs are generated. This leads (for cosmetic ingredients: did lead before March 2009) to the necessity to conduct numerous *in vivo* genotoxicity assays, as well as mechanistic studies, and complex risk assessments. The complexity of risk assessments for compounds positive in standard *in vitro* genotoxicity assays is illustrated by the many recent publications (e.g. Elespuru et al., 2009; Eastmond et al., 2009), external workgroups, and meetings devoted to this topic. MP *in vitro* results will not only trigger (for cosmetic ingredients: have triggered) the use of large numbers of animals, but also require extensive time and personnel within regulatory agencies as well as industry. Because of the enormous resources involved in clarifying positive results from standard *in vitro* testing, cosmetic companies often eliminate such ingredients from use, thereby losing potentially safe and useful ingredients.

2.2. Impact of the Cosmetics Directive

The 7th Amendment to the Cosmetics Directive (EU, 2003) banned *in vivo* genotoxicity assays for cosmetic ingredients as of March 2009, and the performance of repeated dose toxicity (RDT) tests by March 2013. Therefore, the standard approach in all testing strategies of conducting *in vivo* genotoxicity assays as a follow-up for *in vitro* genotoxins is no longer possible for these chemicals. This lack of *in vivo* follow-up options is considered to have a negative impact on the evaluation of the mutagenic potential of cosmetic ingredients (Speit, 2009); the author concludes that in many cases such an evaluation will be impossible. Likewise, the SCCP issued an official opinion statement in January 2009 "Consequently, after 11 March 2009, in many cases, it will not be possible to evaluate the mutagenic potential of cosmetic ingredients on a sound scientific basis. Because the potential mutagenicity of these ingredients is of major concern, an important part of the toxicological evaluation of cosmetic ingredients cannot be accomplished" (SCCP, 2009). In addition, the conduct of *in vivo* genotoxicity assays is not compatible with the general direction within the scientific community to reduce animal testing and with large scale programs such as REACh (European Commission, 2006). Tens of thousands of chemicals will need to be evaluated or re-evaluated for their toxicological properties in this program, and *in vivo* follow-up testing will be required for all chemicals that show positive effects in any of the three standard *in vitro* genotoxicity assays, as described in the endpoint specific guidance of the REACh Information Requirements (European Chemicals Agency, 2008). Not only is there a focused effort to reduce animal use and improve the performance of the existing *in vitro* tests and/or to develop new *in vitro* assays with better specificities for cosmetics in Europe (Tweats et al., 2007), there is also a global need across many product categories for improved non-animal genotoxicity testing approaches and ways to reduce animal use (Pfuhler et al., 2009).

2.3. The COLIPA genotoxicity program – towards improved in vitro capabilities

To address the challenges created by the 7th Amendment to the Cosmetics Directive (EU, 2003) ban on *in vivo* genotoxicity assays for cosmetic ingredients, a COLIPA Genotoxicity Project Team was established in 2004. The team conducted a review of factors that influence the generation of MPs in the standard genotoxicity assays (Kirkland et al., 2007a; see Table 2). This review, together with the outcome of a workshop at European Centre for the Validation of Alternative Methods (ECVAM) held in 2006 (Kirkland et al., 2007b), triggered the start of two COLIPA projects. The first project aims at improving existing *in vitro* standard genotoxicity tests to avoid or minimise generation of MPs, while the second project focuses on the development of new assays that may allow follow-up of positive results from standard genotoxicity testing.

2.3.1. Reduction of "false positive" results

The first project is entitled "Reduction in the "false positive" rate of *in vitro* mammalian cell genotoxicity assays" and is co-sponsored by ECVAM and The National Centre for the Replacement, Refinement and Reduction of Animals in Research (UK NC3Rs). The main areas of focus are:

(a) The use of cell types with higher relevance:
The effects of the cell type on the outcome of the MNT are being studied to determine if there are cell lines which are less susceptible to produce MPs. Cell lines that do not have a high metabolic capacity (especially phase II enzymes) may lack the ability to detoxify chemicals, thereby causing MPs or "false positives". Moreover, cell lines that are not p53 competent or have a low proficiency in repairing DNA (Chaung et al., 1997; Oka et al., 2006) are considered to be hypersensitive and do not reflect the behaviour of normal mammalian cells where DNA damage might either be repaired or lead to apoptosis (Kirkland et al., 2007b). In the project that started in October 2007 and is performed at

Table 2
Explanations for "misleading negative" and "misleading positive" results in *in vitro* assays.

	Biological reasons for misleading results *in vitro*	Technical reasons for misleading results *in vitro*
Misleading negatives	Epigenetic mode of action	Inadequate test conditions (e.g. concentration, pH, solubility, stability)
	Relevant genetic endpoint not measured (e.g. aneuploidy, and some CA)	Volatile compounds
	Differences in metabolic activation *in vitro*/*in vivo*	
	Organ specific compounds (*in vivo*)	
	Very reactive (short lived) compounds	
	Toxic compounds	
Misleading positives	The test compound will not react with DNA *in vivo* (e.g. not absorbed, very reactive)	Contaminated test agent (bacterial tests)
		Chemical impurities in test agent
	Differences in metabolic activation	Inadequate test conditions (e.g. concentration, pH)
	Defence mechanism *in vivo* (e.g. detoxification, repair, apoptosis)	Unstable in media, causing production of hydrogen peroxide

Covance Laboratories in Harrogate, UK, p53 compromised cell lines (Chinese hamster ovary (CHO), Chinese hamster lung (CHL and V79) cells) are compared with primary human lymphocytes, the human lymphoblastoid cell line, TK6, and the human hepatocellular carcinoma cell line, HepG2, which has been shown to have similar drug metabolising enzyme activities to those in primary human hepatocytes (Hewitt and Hewitt, 2004; Westerink and Schoonen, 2007). The initial results of the COLIPA project show that human cells (primary human lymphocytes and TK6, and HepG2) appear to have better specificity than hamster cell lines (Fowler et al., 2009, 2010).

(b) Consideration of measures of toxicity:
There has been debate for some time over the appropriate measures of toxicity to be used in the mammalian cell assays. It is long known that excessive cytotoxicity can cause a positive result in genotoxicity testing which may not be biologically relevant (Hilliard et al., 1998; Scott et al., 1991; Kirkland, 1992; Greenwood et al., 2004). In particular, the methods to determine cytotoxicity are known to have an impact on the selection of the highest test concentration and thus on the genotoxicity outcome (Hilliard et al., 2007), and this topic has recently been discussed intensively in conjunction with the adoption of the OECD TG 487 MNT guideline (Lorge et al., 2008; Fellows et al., 2008). Experimental data from the COLIPA project indicate that toxicity measures that determine cell proliferation should be preferred over methods that do not consider this aspect (e.g. relative cell count) as they are less likely to generate MPs (Williams et al., 2009).

(c) Maximum concentration to be tested:
Concentrations of 10 mM in mammalian cell tests, as recommended in guidelines, may well be in excess of the physiologically-relevant concentrations (Kirkland et al., 2007b). Moreover, testing very high concentrations of chemicals in the presence of rat liver S9 may lead to "inappropriate" metabolism as such concentrations may far exceed the K_M of enzymes involved in their metabolism (e.g. the K_M for phenacetin metabolism by CYP1A2 is only 9 μM (Tassaneeyakul et al., 1993)). For this reason the revised draft guideline of the ICH S2 (R1) limits the highest concentration of pharmaceuticals to be tested for genotoxicity *in vitro* to 0.5 mg/mL or 1.0 mM, whichever is lower. The COLIPA "false positives" project will compare the dose–response of approximately 20 genotoxic carcinogens with a similar number of MPs and may therefore help to decide which impact the reduction of the maximum test concentration may have on the sensitivity and specificity of *in vitro* assays using mammalian cells. In addition, analysis of the maximum concentrations required in mammalian cell tests to give positive results with rodent carcinogens that are negative in the Ames test has indicated that substantial reduction from 10 mM can be achieved without losing sensitivity (Parry et al., 2010; Kirkland and Fowler, 2010)

2.3.2. 3D human reconstructed human skin models for genotoxicity testing

The second pillar of the COLIPA program is the development of new assays that may allow the follow-up of positive results from *in vitro* assays. Most cosmetic products, as well as many other types of products/chemicals, are exclusively or predominantly in contact with the skin; therefore, the skin will usually be the organ with the highest potential exposure. Moreover, the skin is an effective barrier to chemicals (Potts et al., 1992) and metabolically active (Oesch et al., 2007; Eilstein et al., 2009) such that compounds applied to the skin may not enter the systemic circulation, or are detoxified prior to systemic circulation. Based on this, assays using 3D human reconstructed skin (RS) models offer the potential for a more physiologically relevant approach to test dermal exposure. RS models are prepared from primary human cells and are expected to have normal DNA repair and cell cycle control. They are also expected to exhibit a human metabolic capability that is more relevant than the exogenous rodent metabolizing enzymes currently used in standard *in vitro* genotoxicity assays. Similar xenobiotic metabolism gene expression in EpiSkin™ (Luu-The et al., 2009) and EpiDerm™ (Hu et al., 2009a) to human skin has been reported. It has also recently been shown that in EpiDerm™ and EpiSkin™ tissues, the aromatic amines, *p*-phenylenediamine and *p*-aminophenol, are detoxified by N-acetylation in a similar manner to human skin (Hu et al., 2009a,b; Nohynek et al., 2005).

RS models are now being used to develop two different tests for *in vitro* genotoxicity:

(a) The RS micronucleus test (RSMN) (Curren et al., 2006; Mun et al., 2009; Hu et al., 2009a,b). The intra- and inter-laboratory reproducibility of the RSMN in EpiDerm™ was reported by Hu et al. (2009a,b) and Mun et al. (2009). In these studies, a number of chemicals were evaluated (including model genotoxins which exhibit a range of genotoxic mechanisms e.g. cross-linking (mitomycin C), alkylation (methyl methanesulfonate) and aneuploidy (vinblastine) and the ability of this model to respond to different types of genetic damage was ascertained. Moreover, dermal non-carcinogens were determined to be negative, showing that the assay seems to have good specificity as well as sensitivity. This was of particular importance because three of these substances (4-nitrophenol, trichloroethylene and 2-ethyl-1,3-hexanediol) were shown to be positive in one or more other *in vitro* genotoxicity models (Mun et al., 2009). There was a good reproducibility between different tissue constructs containing keratinocytes isolated from human foreskin tissue from four different donors (http://www.mattek.com) of the EpiDerm™ model, which is important since these primary cells have a finite lifespan and must be replaced by a new donor when batches of cells are depleted. Initial results with chemicals that require metabolism are promising (Kaluzhny et al., 2009). Thus, the 3D skin tissue used in the RSMN reflects exposure conditions of dermally applied cosmetic ingredients allowing for more realistic exposure conditions and hopefully better predictive performance.

(b) The RS Comet assay. The Comet assay detects a broad spectrum of DNA damage, including such that leads to gene mutation (Hartmann et al., 2003). This fact is underlined by the high sensitivity of the *in vivo* Comet assay for carcinogens with a gene mutation signature, which it detects more efficiently than other assays like transgenic mutation (TG) or unscheduled DNA synthesis (UDS) (Kirkland and Speit, 2008). It therefore seems well suited as additional endpoint to be incorporated into RS models. Initial results generated with this assay using several standard mutagens are very promising (Flamand et al., 2006; Zeller et al., 2007).

To extend supporting data, a COLIPA project entitled "Development of genotoxicity assays in 3D human skin models", co-sponsored by ECVAM, on the validation of genotoxicity assays in RS models was initiated in 2007. This project follows a modular validation approach (Hartung et al., 2004) and is divided into three phases: Phase 1 – transferability, optimization, and within laboratory reproducibility with two model genotoxins; Phase 2 – between laboratory reproducibility with three coded chemicals and Phase 3 – increasing domain of chemicals tested for predictive capacity and further evaluation or reproducibility. To date, the RSMN and the RS Comet assay using EpiDerm™ have been transferred to other US and European laboratories and the protocols were harmonized as part of the COLIPA project. The RS Comet assay is in the development stage and is undergoing initial evaluation of intra- and inter-laboratory reproducibility. The RSMN using Epi-

Derm™ has completed Phase 1 and Phase 2 testing showing excellent reproducibility of coded chemicals (Aardema et al., 2010) and is now undergoing further intra- and inter-laboratory reproducibility assessment in Phase 3. These are very promising models to consider incorporating into the *in vitro* genotoxicity endpoints such as the MNT and Comet assays as described in the proposed test strategy below.

3. Decision trees and testing strategies for genotoxicity assessment of cosmetics

Results from the COLIPA projects described above, along with an analysis of other non-animal approaches for assessing genotoxicity were discussed and used to design the following decision tree for the safety assessment of cosmetics (see Fig. 1). Because there are no currently available methods to fully replace the standard *in vivo* genotoxicity tests at this time, this decision tree is based in part on use of non-standard methods. The validity and acceptance of such an approach is discussed in chapter 3.3.

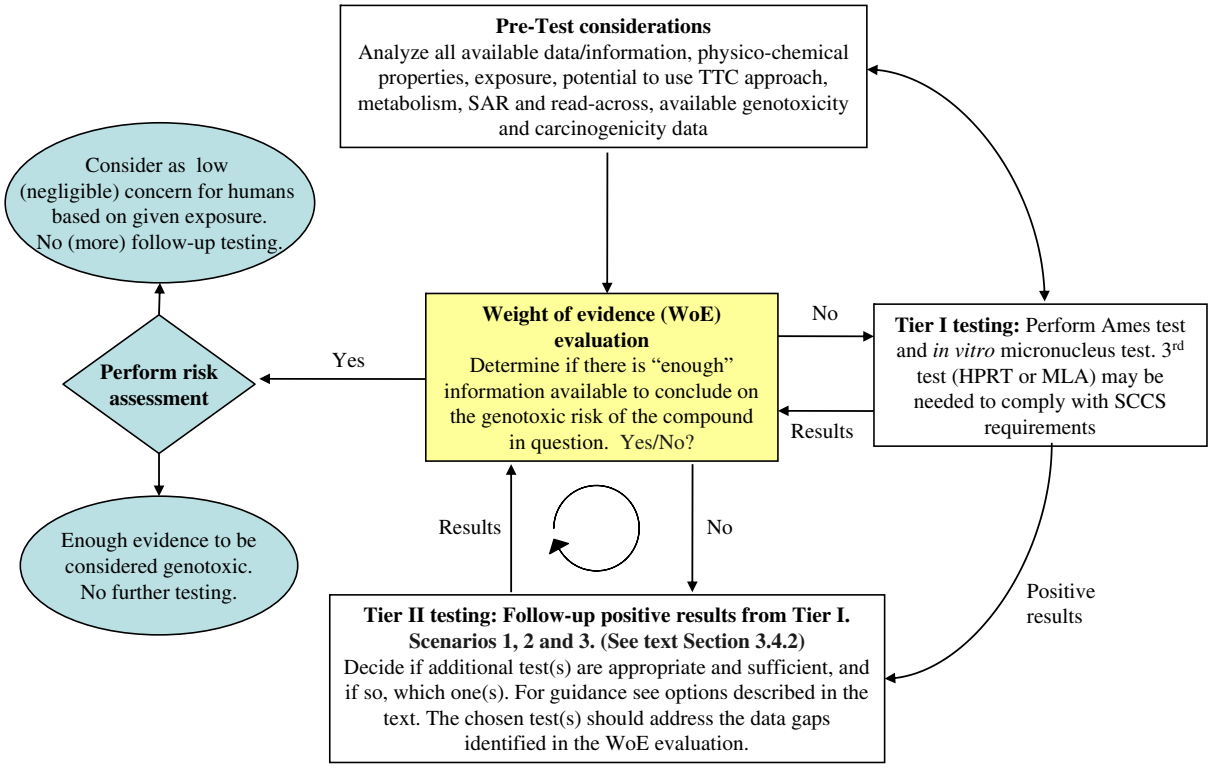
3.1. Weight of evidence as the basic principle

Perhaps the most important non-animal approach for cosmetic ingredients (as well as general chemicals/drugs) is to make use of available data and predicted results for the chemical of interest, using information on the chemical itself, structural/functional analogues, metabolites, etc. in so-called weight of evidence (WoE) assessments. WoE can help determine if testing is needed at all, or help design the most appropriate testing in those cases where additional testing is required to fill data gaps. WoE is broadly accepted by legislation and safety assessors as a basic principle in risk assessment and is explicitly mentioned in European chemicals and cosmetics legislation, e.g. in the current draft of the Recast of the European Cosmetics directive (EU, 2009); REACh Regulation (EC) No. 1907/2006, Annex XI, 1.2 (European Commission, 2006); and in the regulation on classification, labelling and packaging of substances and mixtures ('CLP regulation'; EU, 2008). The procedure usually implies an expert assessment of the relevance, i.e. scientific validity or suitability of the purpose of a method or approach before a decision is taken as to how to weigh individual pieces of information. Consequently, WoE is equally used in pre-test considerations as well as for the choice of follow-up tests at later stages, e.g. when positive results are obtained in Tier I tests, and to judge whether there is sufficient evidence for the presence or absence of a genotoxic potential or effect. Elements of WoE, the practical use in an industrial setting and specifically for cosmetic ingredients are described in the following section.

3.2. Elements of WoE

The major sources of information that are used within a WoE assessment performed prior to testing (see Fig. 1) encompass exposure considerations as well as information on the intrinsic/hazardous properties of the chemical under evaluation:

- *Exposure:* Most cosmetics involve a cutaneous exposure with skin providing an effective barrier and metabolic detoxification (Oesch et al., 2007; Eilstein et al., 2009). If there is no systemic exposure and sufficient information for low/no inherent genotoxicity, no further testing may be needed. For chemicals with limited systemic exposure, a Threshold of Toxicological Concern (TTC) approach can be used (see below).
- *Assessment of structural alerts:* Expert systems and databases such as DEREK, MultiCASE, and ADMEWorks (Hayashi et al., 2005) and the Vitic database (Lhasa Limited, Leeds, UK) are



* „Result" refers to a positive or negative outcome of an assay which can be used in a WoE evaluation

Fig. 1. Schematic representation of the decision making process described in Section 3. For details see text.

valuable at identifying chemical moieties that could lead to genotoxicity. An example for this and the above bullet point is high molecular weight chemicals such as polymers which may lack structural alerts, and exposure through the dermal route may be negligible. Such chemicals may not need to be tested for genotoxic properties, as long as impurities or unconverted monomers are not of concern. The presence of structural alerts is also a component of the TTC approach described below.

- *TTC (Threshold of Toxicological Concern):* If very low systemic exposure to a chemical can be demonstrated, the TTC concept can be used and genotoxicity testing may not be needed. The TTC risk assessment tool is based on the principle of establishing a human exposure threshold value for chemicals, below which there is a very low probability of an appreciable risk to human health (Kroes et al., 2004). The application of the TTC approach as a decision tree has been extended to cosmetic ingredients and impurities (Kroes et al., 2007). The TTC decision tree starts with the identification and evaluation of possible structural alerts for genotoxicity and high potency carcinogenicity. With the exclusion of high potency genotoxins (e.g. aflatoxin-like compounds, N-nitroso-compounds and azoxy compounds), a TTC of 0.15 μg/person/day (0.0025 μg/kg body weight/day) can be used. For substances without structural alerts that raises concern for potential genotoxicity, higher thresholds could be used depending e.g. on their structural class (Cramer Class) and according to the follow-up of the decision tree (Kroes et al., 2007; Cramer et al., 1978; Munro, 1996, 1999).
- *Metabolism:* If there is sufficient data from *in vivo* ADME or *in vitro* metabolism studies indicating detoxification of the test compound to non-genotoxic species further testing may not be necessary. However, if detoxification does not occur in the first tissue of contact (e.g. skin), tissues may be exposed to the direct actions of the chemical before detoxification occurs. This has to be taken into account in the risk assessment. The aromatic amine 4-amino-2-hydroxytoluene, for example, was shown to be converted into its N-acetylated metabolite in the skin when the compound is applied via the dermal route (Goebel et al., 2009). This metabolite was shown to be non-genotoxic in the Ames test and the *in vitro* micronucleus assay in peripheral human lymphocytes (Pfuhler, S., unpublished data).
- *The quality of in vitro and/or in vivo genotoxicity data* available from scientific literature or previous testing needs to be considered in an initial assessment. The available studies should be checked for relevance, compliance with good scientific practice, current guidelines, consistency with literature data for the assay in question, etc.
- If there is *reliable evidence of lack of carcinogenicity* for the cosmetic ingredient under evaluation, genotoxicity testing would not be needed. For instance, carcinogenicity results may be available from testing of the same ingredient under other regulatory regimes, such as the rules governing pharmaceuticals. This also applies to complex raw materials when there are carcinogenicity data on all or some of the constitutive ingredients available, as is the case for some botanical products.
- *Structure–activity-relationship (SAR) and read-across:* Information on DNA binding, electrophilicity and intended functionality for the chemical of interest, as well as information on structural/functional analogues, metabolites, etc. can be used to predict genotoxicity results. Read-across considerations may be based on the chemical domain (e.g. using the OECD Application Toolbox (www.oecd.org)) and when data are available on similar substances. In some cases, sufficient read-across data on structurally similar materials will be available and indicate a low/no probability of genotoxic potential. In other cases, this information may help to pick the appropriate follow-up test(s).
- *Physiological occurrence:* Endogenously occurring substances (in the human) may be considered to be non-hazardous if the physiological amount is not exceeded. It should be kept in mind that the route of exposure may have an impact on the concentrations achieved in the tissues of concern and that not all tissues may be protected equally against reactive chemicals.
- *History of safe use:* Chemicals used in cosmetics most often have other uses and in some cases have a long history of safe use. This information cannot be used as a stand-alone decision-making tool because the history of safe use, in the absence of comprehensive epidemiological data, does not usually provide conclusive evidence of the absence of mutagenic or carcinogenic activity. However, the knowledge that a chemical has been used safely for years can be used in the WoE approach.

3.3. Tier I in vitro testing

In order to test a cosmetic ingredient for its genotoxic potential, it is common practice in the cosmetic industry to conduct the Ames assay and an *in vitro* MNT or CA assay. Most companies are using the MNT since it is recommended by the SCCS and recognised as an excellent tool to detect clastogenic and aneugenic effects (Kirsch-Volders, 1997; Corvi et al., 2008). The MNT has demonstrated a very high concordance with the *in vitro* CA assay (Corvi et al., 2008; Matsushima et al., 1999; Miller et al., 1997), and is considered a valid alternative to the CA test (Corvi et al., 2008). In the ongoing COLIPA genotoxicity program the *in vitro* MNT is being used as the mammalian cell test of choice, and initial results indicate that p53 competent, human cells perform best with regards to specificity of this assay (see Section 2.3.1).

The participants of the Workshop concluded that a two-test battery may in most cases be sufficient for the purposes of decision-making within cosmetics companies. One of the reasons for this common practice of conducting two *in vitro* genotoxicity tests is the high sensitivity of such a battery along with the desire to maintain the specificity of the prediction (see Section 2.1 and Table 1). However, to comply with the SCCS requirements for testing of cosmetic ingredients, an *in vitro* gene mutation assay in mammalian cells will be included as a third test. This third test can also be part of the below described follow-up testing strategy, i.e. when the Ames test shows positive results.

3.4. Decision tree for follow-up assessment of Tier I results (see Fig. 1)

In this section the use of test options which are linked to a specific data constellation is described, and some of these options include methods which are not (yet) formally validated. The WoE evaluation implies an expert assessment of the scientific validity or suitability of a method or approach. In REACh it is stated that "results obtained from suitable *in vitro* methods may indicate the presence of a certain dangerous property or may be important in relation to a mechanistic understanding, which may be important for the assessment." In this context, 'suitable' means 'sufficiently developed' according to internationally agreed test development criteria (e.g. ECVAM criteria for entry into pre-validation process). There has been some debate about the appropriate definition and use of the terms "valid" and "suitable" (e.g. Balls, 2009; Combes, 2007) and whether these terms can be applied to cosmetics (SCCP, 2007). In general, it seems justified to include information that has been generated with methods that have not (yet) undergone formal validation as a part of the overall assessment when it can be demonstrated that the data are relevant. A positive impact on regulatory acceptance of such an assessment is expected when transparent criteria, reliable documentation and plausibility of the scientific reasoning are provided.

3.4.1. Follow up when Ames test and in vitro MNT or CA are negative

In this case, to fulfil the SCCS requirements (SCCP, 2006a), a third test, namely a gene mutation assay in mammalian cells, will be performed. If this assay is also negative this confirms the absence of a genotoxic hazard for the test item and no further testing will be needed. If the assay is positive Scenario 1, as described below, should be followed.

3.4.2. Follow up when there are positive Tier I in vitro results

When a conclusion about the genotoxic potential cannot be drawn from the data generated in Tier I testing and/or the application of a WoE approach that takes into consideration all data including the aspects described in Section 3.2, further testing will be needed. The different scenarios and testing follow-up options are described below and summarised in Fig. 1.

The "options" described below for each of the scenarios in which positive results may arise are considered the most promising approaches in light of each respective scenario (outcome of the Tier I testing), however, the choice of suitable assays may vary depending on the chemical class and the specific data constellation.

3.4.2.1. Scenario 1: In the case that the outcome of the Ames test is positive and in vitro MNT or CA is negative. There are a number of common options for the follow-up of this scenario:

Option 1: If not already done in Tier 1, carry out a mammalian gene mutation assay, that either detects mutations at the *hprt* (hypoxanthine–guanine phosphoribosyl transferase) or *tk* (thymidine kinase) gene locus. Being a mandatory part of the SCCS requirements (SCCP, 2006a) a mammalian gene mutation assay will have to be performed in conjunction with options 2–4 also.

The sensitivity of the *hprt* test is reported to be good (in a similar range to the sensitivity of the MLA) for a large number of pharmaceuticals (Matthews et al., 2006) and is considered to be equally acceptable as the MLA for assessing gene mutation activity in mammalian cells (Pfuhler et al., 2007; European Chemicals Agency, 2008). Some safety assessors would consider the *hprt* gene locus to be preferred as the third *in vitro* test, rather than the *tk*, as the clastogenic potential of the substance will have already been evaluated in the MNT or CA; and avoiding this overlap should usually lower the MP rate. A negative outcome from an *hprt* assay, in conjunction with other data that point towards the lack of relevance of the Ames positive result for mammalian systems may be sufficient to address a positive outcome in an Ames assay.

Option 2: Carry out the Syrian hamster embryo (SHE) cell transformation assay. This assay is currently the preferred test because it is listed in the "Notes of Guidance for the Testing of Cosmetic Ingredients" (SCCP, 2006a) and the strategy for testing of oxidative hair dyes (SCCP, 2006b). This assay measures morphological cell transformation of primary SHE cells (evident as changes in behaviour and growth control of cultured cells occurring in the early steps of carcinogenesis) and has been reported to have a high predictivity (~80% concordance) of rodent carcinogenicity (LeBoeuf et al., 1990, 1996) as was recently reviewed by the OECD (OECD, 2007). The review reported the results from tests on 203 organic chemicals and 61 inorganic chemicals of which 78% of the rodent organic carcinogens and 94% of the inorganic carcinogens were positive in the SHE assay. 9% of the carcinogens were not detected by this assay. This assay is especially useful in a strategy for testing Ames positive chemicals such as aromatic amines, important components of hair dyes, since the SHE assay accurately distinguishes between carcinogenic and non-carcinogenic aromatic amines 90% of the time (Kerckaert et al., 1998; Isfort et al., 1996). Unlike other *in vitro* assays, this assay detects both genotoxic and non-genotoxic rodent carcinogens (LeBoeuf et al., 1996; Myhr and Zhang, 2000). Its high predictivity has resulted in some regulatory authorities (e.g. FDA) requesting this test to be carried out as a follow-up test for positive outcomes in initial standard test batteries (Bigger, 2003). The SHE assay has been evaluated in an ECVAM pre-validation exercise and shows good intra- and inter-laboratory reproducibility resulting in the development of a standard protocol and the conclusion that, in combination with the extensive database summarized in the OECD DRP31 (OECD, 2007), the assay can be used for the assessment of carcinogenicity potential (VanParys et al., 2010). A negative outcome from SHE may be sufficient to address a positive outcome in an Ames assay.

Other cell transformation assays might also be useful for this purpose if they can be shown to discriminate between genotoxic carcinogens and genotoxic non-carcinogens.

Option 3: Carry out the *in vivo* Comet assay which is integrated into repeat dose toxicology (RDT) studies without the need of additional animals ("3R" strategy). The *in vivo* Comet assay has growing scientific acceptance not only as a stand-alone assay but also for integration into RDT studies (European Chemicals Agency, 2008). A caveat is that this assay is not currently fully accepted from a scientific or regulatory viewpoint (Pfuhler et al., 2009). A further point that should be kept in mind is that, for cosmetic ingredients, no endpoint-specific positive control group can be included as the EU Cosmetics Directive prohibited the use of animals for the purpose of genotoxicity testing in March of 2009, and positive control animals would not otherwise be needed for the RDT study. This might prove challenging for the Comet assay as, at this point in time, not many laboratories have sufficient historical data and experience to support exclusion of the positive control group. In addition, the suitability of integration of the Comet assay into RDT studies is currently under investigation in the pharmaceutical industry, in conjunction with the revision of the ICH S2 (R1) guideline (ICH, 2008). In general the option of integration of genotoxicity endpoints into RDT studies will only be available for cosmetic ingredients until 2013 when the ban on these studies comes into force.

Once further developed, the 3D skin Comet assay could also be a valid choice for dermally exposed chemicals under scenario 1. As mentioned earlier, the Comet assay detects a broad spectrum of DNA damage, including effects that lead to gene mutation and seems therefore suited to add useful information in case the Ames test and/or mammalian cell gene mutation test are positive. However, the assay is still undergoing development (see COLIPA project in Section 2.3.2) and will be considered for use depending on the outcome of the ongoing studies.

Option 4: It may be that a threshold mode of action can be envisaged for the observed genotoxic effect. Although threshold responses for genotoxicity and carcinogenicity have been related to non-DNA-reactive chemicals (epigenetic mechanisms (Fukushima et al., 2005)), they may also apply to DNA-reactive chemicals (Jenkins et al., 2005; Henderson et al., 2000; Doak et al., 2007) despite other findings that they are linear in nature (i.e. non-threshold) (Sanner and Dybing, 2005). This is likely to be due to the fact that organisms can tolerate low level DNA damage due to protective mechanisms such as DNA repair (Jenkins et al., 2005; Doak et al., 2008). Examples of this are the alkylating agents, such as ethyl methane sulfonate (EMS), for which a threshold has been accepted *in vitro* and *in vivo* (Gocke and Müller, 2009). Although it is not easy to prove experimentally, if a threshold can be shown to exist as part of a mode of action approach, the exposure level can be taken together with the threshold concentration in order to make an assessment on the risk potential of a chemical (Kirkland et al., 2007a; Carmichael et al., 2009; Kirsch-Volders et al., 2009). Evidence for thresholds typically requires the conduct of extensive *in vivo* tests. In the case of EMS the i*n vitro* assays were able to correctly predict that it exhibited a threshold *in vivo*; however, more examples are needed to be able to judge the reliability of *in vitro*

assays for defining such thresholds *in vivo*. It is proposed that the scientific community should give increased attention in the future to developing *in vitro* approaches for the establishment of thresholds.

3.4.2.2. Scenario 2: In the case that the outcome of the Ames test is negative and that the outcome of in vitro MNT or CA is positive. Option 1: Carry out the *in vivo* MNT by integrating it into RDT studies without the need for additional animals ("3R" strategy) – this is the straightforward approach but for cosmetic ingredients it is only permitted until 2013. The integration of the micronucleus (MN) endpoint into RDT testing is a long and well established concept (Wakata et al., 1998; Hayashi et al., 2000, 2007; Hamada et al., 2001) and is already represented in the current OECD guideline, which came into effect in 1997 (OECD TG 474). The fact that micronuclei can be scored in peripheral blood using flow cytometry does simplify this approach, and the above mentioned studies did show that peripheral blood is equally as reliable as the bone marrow for measuring micronuclei even in rats that exhibit an efficient splenic filtration function.

The integration of the MNT is also recognised by the draft ICH S2 (R1) guideline, which states that "the *in vivo* genotoxicity assays can often be integrated into RDT studies when the doses are sufficient" (ICH, 2008). At an ECVAM workshop in June 2008, the possibility to integrate the MNT and Comet assays into RDT studies was thoroughly discussed and the MNT was considered as scientifically credible and therefore ready for use in RDT studies (Pfuhler et al., 2009). The results of further study by the pharmaceutical industry were presented at the IWGT meeting in Basel Switzerland in August 2009 and led to consensus agreement on integration of MNT in RDT (Rothfuss, A., unpublished report from the International Workshop on Genotoxicity Tests (IWGT) in Basle 2009).

Option 2: Another possibility to deal with this data constellation would be to carry out the RSMN or RS Comet assay. Until more supporting data are generated, the RSMN is not sufficient as a standalone follow-up of a positive result from standard *in vitro* MNT or CA tests but will add value to the WoE assessment. The RS Comet assay is still undergoing development (see COLIPA project in Section 2.3.2) and will be considered for use depending on the outcome of the ongoing studies.

Option 3: Carry out the SHE cell transformation assay. See scenario 1, option 2.

Option 4: Investigate the application of a threshold for the observed genotoxic effect. See scenario 1, option 4.

3.4.2.3. Scenario 3: In the case that the outcomes of the Ames test and the in vitro MNT or CA are positive. In the case that both tests are positive, then a combination of assays is needed (see respective options from scenarios 1 and 2). The choice of assays differs according to the rates of MPs that are associated with each endpoint; however, one individual assay will usually not be sufficient to resolve this data constellation. If the WoE from the *in vitro* follow-up tests is not sufficient to come to a conclusion as to the genotoxic potential of such a chemical, then integration of genotoxicity endpoints into RDT studies could serve as an option, although only until 2013 for cosmetic ingredients.

4. Conclusions

In this paper we provide a framework for genotoxicity assessment of cosmetic ingredients based on common practice within the industry with an emphasis on non-animal approaches. Features of this strategy include (a) an extensive pre-testing evaluation using existing data and exposure considerations, (b) *in vitro* screening assays, and (c) validated *in vitro* genotoxicity tests that can be used independently or in combination with a WoE approach to assess the genotoxic potential of cosmetic ingredients. These tests are sensitive enough to detect rodent carcinogens; however, their specificities are not sufficient to replace fully the *in vivo* rodent assays. Currently, in the case that an *in vitro* test battery shows a positive outcome, *in vivo* genotoxicity endpoints can be integrated into RDT studies but will only be carried out on the most promising/relevant ingredients and only after extensive *in vitro* testing. In the near future (2013), *in vitro* assays will be the only available tools with which to follow-up after positive outcomes from initial *in vitro* testing. Therefore, vigorous efforts are ongoing to improve existing tests (and reduce MPs) and to develop new models in order to reduce the high attrition rate of new and potentially safe chemicals.

References

Aardema, M., Barnett, B., Khambatta, Z., Reisinger, K., Ouedraogo, G., Faquet, B., Ginestet, A. C., Mun, G., Dahl, E., Hewitt, N. J., Corvi, R., Curren, R., 2010. International Prevalidation Studies of the EpiDermTM 3D Human Reconstructed Skin Micronucleus (RSMN) Assay: Transferability and Reproducibility. Mutat. Res., submitted for publication.

Ashby, J., Tennant, R.W., 1991. Definitive relationships among chemical structure, carcinogenicity and mutagenicity for 301 chemicals tested by the U.S. NTP. Mutat. Res. 257 (3), 229–306 (Erratum in: Mutat. Res. 317 (2) (1994) 175).

Balls, M., 2009. Animal experimentation and the three Rs: time for honest answers to some leading questions (Editorial). ATLA 37, 229–232.

Bigger, A., 2003. CDER position on the use of the SHE cell transformation assay in regulatory determinations. Regulatory and Safety evaluation Speciality Section Newsletter of the Society of Toxicology. http://www.toxicology.org/ISOT/SS/regulatorysafety/NLSummer_2003.pdf.

Carmichael, P., Kirsch-Volders, M., Vrijhof, H., 2009. Thresholds for genotoxins and their application in risk assessment. Mutat. Res. 678 (2), 71.

Chaung, W., Mi, L.-J., Boorstein, R.J., 1997. The p53 status of Chinese hamster V79 cells frequently used for studies of DNA damage and DNA repair. Nucleic Acid Res. 25, 992–994.

COM (UK Committee on Mutagenicity of Chemicals in Food, Consumer Products and the Environment), 2000. Guidance on a strategy for Testing of Chemicals for Mutagenicity. http://www.advisorybodies.doh.gov.uk/COM/guidance.pdf.

Combes, R., 2007. Why 'suitable' in vitro methods, as defined in the final EU REACh legislation, are an inappropriate basis for risk assessment. ATLA 35 (2), 289–293.

Corvi, R., Albertini, S., Hartung, T., Hoffmann, S., Maurici, D., Pfuhler, S., van Benthem, J., Vanparys, P., 2008. ECVAM retrospective validation of in vitro micronucleus test (MNT). Mutagenesis 23 (4), 271–283.

Cramer, G.M., Ford, R.A., Hall, R.L., 1978. Estimation of toxic hazard – a decision tree approach. Food Cosmet. Tox. 16, 255–276.

Curren, R.D., Mun, G.C., Gibson, D.P., Aardema, M.J., 2006. Development of a method for assessing micronucleus induction in a 3D human skin model (EpiDerm). Mutat. Res. 607 (2), 192–204.

Dearfield, K.L., Cimino, M.C., McCarroll, N.E., Mauer, I., Valcovic, L.R., 2002. Genotoxicity risk assessment: a proposed classification strategy. Mutat. Res. 521, 121–135.

Doak, S.H., Jenkins, G.J., Johnson, G.E., Quick, E., Parry, E.M., Parry, J.M., 2007. Mechanistic influences for mutation induction curves after exposure to DNA-reactive carcinogens. Cancer Res. 67 (8), 3904–3911.

Doak, S.H., Brüsehafer, K., Dudley, E., Quick, E., Johnson, G., Newton, R.P., Jenkins, G.J., 2008. No-observed effect levels are associated with up-regulation of MGMT following MMS exposure. Mutat. Res. 648 (1–2), 9–14.

Eastmond, D.A., Hartwig, A., Anderson, D., Anwar, W.A., Cimino, M.C., Dobrev, I., Douglas, G.R., Nohmi, T., Phillips, D.H., Vickers, C., 2009. Mutagenicity testing for chemical risk assessment: update of the WHO/IPCS harmonized scheme. Mutagenesis 24 (4), 341–349.

Eilstein, J., Lereaux, G, Daronnat, E., Dichich, S., Meunier, J.R., Leclair, J., Duche, D., 2009. Characterisation of N-acetyl transferase activity in reconstructed human skin models. Drug Metab. Rev. 41 (1) (Abstract 68).

Elespuru, R.K., Agarwal, R., Atrakchi, A.H., Bigger, C.A., Heflich, R.H., Jagannath, D.R., Levy, D.D., Moore, M.M., Ouyang, Y., Robison, T.W., Sotomayor, R.E., Cimino, M.C., Dearfield, K.L., 2009. Current and future application of genetic toxicity assays: the role and value of in vitro mammalian assays. Toxicol Sci. 109 (2), 172–179.

European Chemicals Agency, 2008. Guidance on information requirements and chemical safety assessment. CH. R.7a: Endpoint Specific Guidance. See Chapter R.7.7.1 Mutagenicity Testing. pp. 374–402; http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_r7a_en.pdf?vers=20_08_08.

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

Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32006R1907:EN:NOT.

EU, 2003. Directive 2003/15/EC of the European Parliament and the Council of 27 February 2003 amending Council Directive 76/768/EEC on the approximation of the laws of the Members States relating to cosmetic products. Official J. Eur. Union L66, 26–35.

EU, 2008. CLP-Regulation (EC) No. 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No. 1907/2006.

EU, 2009. Cosmetic products (repeal. 'Cosmetics Directive' 76/768/EEC). Recast. European Parliament legislative resolution of 24 March 2009 on the proposal for a regulation of the European Parliament and of the Council on cosmetic products (recast). COD/2008/0035.

Fellows, M.D., O'Donovan, M.R., Lorge, E., Kirkland, D., 2008. Comparison of different methods for an accurate assessment of cytotoxicity in the in vitro micronucleus test: II. Practical aspects with toxic agents. Mutat. Res. 655 (1–2), 4–21.

Flamand, N., Marrot, L., Belaidi, J.P., Bourouf, L., Dourille, E., Feltes, M., Meunier, J.R., 2006. Development of genotoxicity test procedures with Episkin, a reconstructed human skin model: towards new tools for in vitro risk assessment of dermally applied compounds? Mutat. Res. 606 (1–2), 39–51.

Fowler, P., Williams, K., Jeffrey, L., Young, J., Carmichael, P., Aardema, M., Diembeck, W., Fautz, R., Harvey, J., Hewitt, N., Latil, A., Pfuhler, S., Quedraogo, G., Reisinger, K., Marrec-Fairley, M., Kirkland, D. 2009. Reduction of misleading (False") positive results in mammalian cell genotoxicity assays: I. Choice of cell type. VII World Congress on Alternatives & Animal Use in the Life Sciences, 30th August–3rd September, 2009. Rome, Italy. Altex 26, special issue, p. 190. http://altweb.jhsph.edu/altex/26_special/index.html.

Fowler, P., Young, J., Jeffrey, L., Hand, T., Smith, K., Kirkland, D., Carmichael, P., Pfuhler, S., 2010. Reduction of misleading ("false") positive results in mammalian cell genotoxicity assays. I. Choice of cell type. Mutat. Res., submitted for publication.

Fukushima, S., Kinoshita, A., Puatanachokchai, R., Kushida, M., Wanibuchi, H., Morimura, K., 2005. Hormesis and dose–response-mediated mechanisms in carcinogenesis: evidence for a threshold in carcinogenicity of non-genotoxic carcinogens. Carcinogenesis 26 (11), 1835–1845.

Gocke, E., Müller, L., 2009. In vivo studies in the mouse to define a threshold for the genotoxicity of EMS and ENU. Mutat. Res. 678 (2), 101–107.

Goebel, C., Hewitt, N.J., Kunze, G., Wenker, M., Hein, D.W., Beck, H., Skare, J., 2009. Skin metabolism of aminophenols: human keratinocytes as a suitable in vitro model to qualitatively predict the dermal transformation of 4-amino-2-hydroxytoluene in vivo. Toxicol. Appl. Pharmacol. 235 (1), 114–123.

Greenwood, S.K., Hill, R.B., Sun, J.T., Armstrong, M.J., Johnson, T.E., Gara, J.P., Galloway, S.M., 2004. Population doubling: a simple and more accurate estimation of cell growth suppression in the in vitro assay for chromosomal aberrations that reduces misleading positive results. Environ. Mol. Mutagen. 43, 36–44.

Hamada, S., Sutou, S., Morita, T., Wakata, A., Asanami, S., Hosoya, S., Ozawa, S., Kondo, K., Nakajima, M., Shimada, H., Osawa, K., Kondo, Y., Asano, N., Sato, S., Tamura, H., Yajima, N., Marshall, R., Moore, C., Blakey, D.H., Schechtman, L.M., Weaver, J.L., Torous, D.K., Proudlock, R., Ito, S., Namiki, C., Hayashi, M., 2001. Evaluation of the rodent micronucleus assay by a 28-day treatment protocol: Summary of the 13th Collaborative Study by the Collaborative Study Group for the Micronucleus Test (CSGMT)/Environmental Mutagen Society of Japan (JEMS)-Mammalian Mutagenicity Study Group (MMS). Environ. Mol. Mutagen. 37 (2), 93–110.

Hayashi, M., Kamat, E., Hirose, A., Takahashi, M., Morita, T., Ema, M., 2005. In silico assessment of chemical mutagenesis in comparison with results of Salmonella microsome assay on 909 chemicals. Mutat. Res./Genet. Toxicol. Environ. Mutagen. 588 (2), 129–135.

Hartung, T., Bremer, S., Casati, S., Coecke, S., Corvi, R., Fortaner, S., Gribaldo, L., Halder, M., Hoffmann, S., Roi, A.J., Prieto, P., Sabbioni, E., Scott, L., Worth, A., Zuang, V., 2004. A modular approach to the ECVAM principles on test validity. ATLA 32 (5), 467–472.

Hartmann, A., Plappert, U., Poetter, F., Suter, W., 2003. Comparative study with the alkaline Comet assay and the chromosome aberration test. Mutat. Res. 20, 27–38.

Hayashi, M., MacGregor, J.T., Gatehouse, D.G., Adler, I.-D., Blakey, D.H., Dertinger, S.D., Krishna, G., Morita, T., Russo, A., Sutou, S., 2000. In vivo rodent erythrocyte micronucleus assay: II. Some aspects of protocol design including repeated treatments, integration with toxicity testing, and automated scoring. Environ. Mol. Mutagen. 35, 234–252.

Hayashi M., MacGregor J.T., Gatehouse D.G., Blakey D.H., Dertinger S.D., Abramsson-Zetterberg L., Krishna G., Morita T., Russo A., Asano N., Suzuki H., Ohyama W., Gibson D., 2007. In vivo Micronucleus Assay Working Group, IWGT. In vivo erythrocyte micronucleus assay III. Validation and regulatory acceptance of automated scoring and the use of rat peripheral blood reticulocytes, with discussion of non-hematopoietic target cells and a single dose-level limit test. Mutat. Res. 627 (1), 10–30.

Henderson, L., Albertini, S., Aardema, M., 2000. Thresholds in genotoxicity responses. Mutat. Res. 464 (1), 123–128.

Hewitt, N.J., Hewitt, P., 2004. Phase I and II enzyme characterization of two sources of HepG2 cell lines. Xenobiotica 34 (3), 243–256.

Hilliard, C.A., Armstrong, M.J., Bradt, C.I., Hill, R.B., Greenwood, S.K., Galloway, S.M., 1998. Chromosome aberrations in vitro related to cytotoxicity of nonmutagenic chemicals and metabolic poisons. Environ. Mol. Mutagen. 31, 316–326.

Hilliard, C., Hill, R., Armstrong, M., Fleckenstein, C., Crowley, J., Freeland, E., Duffy, D., Galloway, S.M., 2007. Chromosome aberrations in Chinese hamster and human cells: a comparison using compounds with various genotoxicity profiles. Mutat. Res. 616 (1–2), 103–118.

Hu, T., Bailey, R.E., Morrall, S.W., Aardema, M.J., Stanley, L.A., Skare, J.A., 2009a. Dermal penetration and metabolism of *p*-aminophenol and *p*-phenylenediamine: application of the EpiDerm human reconstructed epidermis model. Toxicol. Lett. 188 (2), 119–129.

Hu, T., Kaluzhny, Y., Mun, G.C., Barnett, B., Karetsky, V., Wilt, N., Klausner, M., Curren, R.D., Aardema, M.J., 2009b. Intralaboratory and interlaboratory evaluation of the EpiDerm 3D human reconstructed skin micronucleus (RSMN) assay. Mutat. Res. 673 (2), 100–108.

ICH, 2008. Guidance on Genotoxicity Testing and Data Interpretation for Pharmaceuticals Intended for Human Use S2(R1). Current Ste 2 version dated 6 March 2008. http://www.ich.org.

Isfort, R.J., Kerckaert, G.A., LeBoeuf, R.A., 1996. Comparison of the standard and reduced pH Syrian hamster embryo (SHE) cell in vitro transformation assays in predicting the carcinogenic potential of chemicals. Mutat. Res. 356, 11–63.

Jenkins, G.J., Doak, S.H., Johnson, G.E., Quick, E., Waters, E.M., Parry, J.M., 2005. Do dose–response thresholds exist for genotoxic alkylating agents? Mutagenesis 20 (6), 389–398.

Kaluzhny, Y., Letasiova, S., Hayden, P., Bolmarcich, J., Jackson, G., Klausner, M., 2009. Drug metabolizing enzyme activity of in vitro human dermal (EpiDerm(TM)) and airway (EpiAirway(TM)) epithelial models: relationship to genotoxicity of chemicals as determined by in vitro skin micronucleus assays (Abstract). Toxicol. Lett. 189 (Supplement 1), p. 86.

Kerckaert, G.A., LeBoeuf, R.A., Isfort, R.J., 1998. Assessing the predictiveness of the Syrian Hamster embryo cell transformation assay for determining the rodent carcinogenic potential of single ring aromatic/nitroaromatic compounds. Toxicol. Sci. 41, 189–197.

Kirkland, D., 1992. Chromosomal aberration tests in vitro: problems with protocol design and interpretation of results. Mutagenesis 7, 95–106.

Kirkland, D., Aardema, M., Henderson, L., Müller, L., 2005a. Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens: I. Sensitivity, specificity and relative predictivity. Mutat. Res. 584 (1–2), 1–256 (Erratum in: Mutat. Res. 588 (1), 70).

Kirkland, D.J., Henderson, L., Marzin, D., Müller, L., Parry, J.M., Speit, G., Tweats, D.J., Williams, G.M., 2005b. Testing strategies in mutagenicity and genetic toxicology: an appraisal of the guidelines of the European Scientific Committee for Cosmetics and Non-Food Products for the evaluation of hair dyes. Mutat. Res. 588 (2), 88–105.

Kirkland, D., Aardema, M., Müller, L., Makoto, H., 2006. Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens: II. Further analysis of mammalian cell results, relative predictivity and tumour profiles. Mutat. Res. 608 (1), 29–42.

Kirkland, D., Aardema, M., Banduhn, N., Carmichael, P., Fautz, R., Meunier, J.R., Pfuhler, S., 2007a. In vitro approaches to develop weight of evidence (WoE) and mode of action (MoA) discussions with positive in vitro genotoxicity results. Mutagenesis 22 (3), 161–175.

Kirkland, D., Pfuhler, S., Tweats, D., Aardema, M., Corvi, R., Darroudi, F., Elhajouji, A., Glatt, H., Hastwell, P., Hayashi, M., Kasper, P., Kirchner, S., Lynch, A., Marzin, D., Maurici, D., Meunier, J.R., Müller, L., Nohynek, G., Parry, J., Parry, E., Thybaud, V., Tice, R., van Benthem, J., Vanparys, P., White, P., 2007b. How to reduce false positive results when undertaking in vitro genotoxicity testing and thus avoid unnecessary follow-up animal tests: report of an ECVAM Workshop. Mutat. Res. 628 (1), 31–55.

Kirkland, D., Speit, G., 2008. Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens: III. Appropriate follow-up testing in vivo. Mutat. Res. 654 (2), 114–132.

Kirkland, D., Fowler, P., 2010. Further analysis of Ames-negative rodent carcinogens that are only genotoxic in mammalian cells at concentrations exceeding 1 mM, including retesting of compounds of concern. Mutagenesis, submitted for publication.

Kirsch-Volders, M., 1997. Towards a validation of the micronucleus test. Mutat. Res. 392 (1–2), 1–4.

Kirsch-Volders, M., Gonzalez, L., Carmichael, P., Kirkland, D., 2009. Risk assessment of genotoxic mutagens with thresholds: a brief introduction. Mutat. Res. 678 (2), 72–75.

Kroes, R., Renwick, A.G., Cheeseman, M., Kleiner, J., Mangelsdorf, I., Piersma, A., Schilter, B., Schlatter, J., van Schothorst, F., Vos, J.G., Würtzen, G., 2004. Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. Food Chem. Toxicol. 42 (1), 65–83.

Kroes, R., Renwick, A.G., Feron, V., Galli, C.L., Gibney, M., Greim, H., Guy, R.H., Lhuguenot, J.C., van de Sandt, J.J., 2007. Application of the threshold of toxicological concern (TTC) to the safety evaluation of cosmetic ingredients. Food Chem. Toxicol. 45 (12), 2533–2562.

LeBoeuf, R.A., Kerckaert, G.A., Poiley, J.A., Raineri, R., 1990. An interlaboratory comparison of enhanced morphological transformation of Syrian hamster embryo cells cultured under conditions of reduced bicarbonate concentration and pH. Mutat. Res. 222 (3), 205–218.

LeBoeuf, R.A., Kerckaert, G.A., Aardema, M.J., Gibson, D.P., Brauninger, R., Isfort, R.J., 1996. The pH 6.7 Syrian hamster embryo cell transformation assay for assessing the carcinogenic potential of chemicals. Mutat. Res. 356(1), 85-127.

Lorge, E., Hayashi, M., Albertini, S., Kirkland, D., 2008. Comparison of different methods for an accurate assessment of cytotoxicity in the in vitro micronucleus test: I. Theoretical aspects, Mutat. Res. 655 (1–2), 1–3.

Luu-The, V., Duche, D., Ferraris, C., Meunier, J.R., Leclaire, J., Labrie, F., 2009. Expression profiles of phases 1 and 2 metabolizing enzymes in human skin and the reconstructed skin models Episkin and full thickness model from Episkin. J. Steroid Biochem. Mol. Biol. 116 (3–5), 178–186.

Matsushima, T., Hayashi, M., Matsuoka, A., Ishidate Jr., M., Miura, K.F., Shimizu, H., Suzuki, Y., Morimoto, K., Ogura, H., Mure, K., Koshi, K., Sofuni, T., 1999. Validation study of the in vitro micronucleus test in a Chinese hamster lung cell line (CHL/IU). Mutagenesis 14 (6), 569–580.

Matthews, E.J., Kruhlak, N.L., Cimino, M.C., Benz, R.D., Contrera, J.F., 2006. An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: I. Identification of carcinogens using surrogate endpoints. Regul. Toxicol. Pharmacol. 44 (2), 83–96.

Miller, B., Albertini, S., Locher, F., Thybaud, V., Lorge, E., 1997. Comparative evaluation of the in vitro micronucleus test and the in vitro chromosome aberration test: industrial experience. Mutat. Res. 392 (1–2), 45–59.

Mueller, L., Blakey, D., Dearfield, K.L., Galloway, S., Guzzie, P., Hayashi, M., Kasper, P., Kirkland, D., MacGregor, J.T., Parry, J.M., Schechtman, L., Smith, A., Tanaka, N., Tweats, D., Yamasaki, H., 2003. Strategy for genotoxicity testing and stratification of genotoxicity test results – report on initial activities of the IWGT Expert Group. Mutat. Res. 540, 177–181.

Mun, G.C., Aardema, M.J., Hu, T., Barnett, B., Kaluzhny, Y., Klausner, M., Karetsky, V., Dahl, E.L., Curren, R.D., 2009. Further development of the EpiDerm 3D reconstructed human skin micronucleus (RSMN) assay. Mutat. Res. 673 (2), 92–99.

Munro, I.C., Ford, R.A., Kennepohl, E., Sprenger, J.G., 1996. Correlation of structural class with no-observed effect levels: a proposal for establishing a threshold of concern. Food Chem. Toxicol. 34, 829–867.

Munro, I.C., Kennepohl, E., Kroes, R., 1999. A procedure for the safety evaluation of flavouring substances. Food Chem.Toxicol. 37, 207–232.

Myhr, B.C., Zhang, H., 2000. Application of the in vitro SHE cell assay to the prediction of rodent carcinogenesis. Abstract P/3 from EEMS-2000 (30th Annual Meeting of the European Environmental Mutagen Society), Budapest, Hungary. http://www.swansea.ac.uk/cget/ejgt/abstracts_eems.pdf.

Nohynek, G.J., Duche, D., Garrigues, A., Meunier, P.A., Toutain, H., Leclaire, J., 2005. Under the skin: biotransformation of *para*-aminophenol and *para*-phenylenediamine in reconstructed human epidermis and human hepatocytes. Toxicol. Lett. 158, 196–212.

OECD, 2007. Series on testing and assessment Number 31 Detailed review paper on cell transformation assays for detection of chemical carcinogens. August 13, 2007. http://www.olis.oecd.org/olis/2007doc.nsf/FREREFCORPLOOK/NT00002F0A/$FILE/JT03230941.PDF.

OECD TG 471, 1997. OECD Guidelines for the Testing of Chemicals, No. 471: Bacterial Reverse Mutation Test. Adopted 21st July 1997.

OECD TG 473, 1997. OECD Guidelines for the Testing of Chemicals, No. 473: In vitro Mammalian Chromosome Aberration Test. Adopted 21st July 1997.

OECD TG 474, 1997. OECD Guidelines for the Testing of Chemicals, No. 474: Mammalian Erythrocyte Micronucleus Test. Adopted 21st July 1997.

OECD TG 476, 1997. OECD Guideline for testing of chemicals, No. 476: In vitro Mammalian Cell Gene Mutation Test. Organization for Economic Cooperation and Development, Paris, adopted 4 April 1984, last updated 21 July 1997.

OECD TG 487, 2009. OECD Guideline for testing of chemicals, No. 487 DRAFT: In vitro Micronucleus Test. Organization for Economic Cooperation and Development, Paris, (draft version 5, 2 November 2009).

OECD TG 495. OECD Guideline for testing of chemicals, No. 495 DRAFT: In vitro Syrian Hamster Embryo (SHE) Cell Transformation Assay. Organization for Economic Cooperation and Development, Paris (draft approved).

Oesch, F., Fabian, E., Oesch-Bartlomowicz, B., Werner, C., Landsiedel, R., 2007. Drug-metabolizing enzymes in the skin of man, rat, and pig. Drug Metab. Rev. 39 (4), 659–698.

Oka, H., Ikeda, K., Yoshimura, H., Ohuchida, A., Honma, M., 2006. Relationship between p53 status and 5-fluorouracil sensitivity in 3 cell lines. Mutat. Res. 606, 52–60.

Parry, J. M., Parry, E., Phrakonkham, P., Corvi, R., 2010. Analysis of published data for top concentration considerations in mammalian cell genotoxicity testing. Mutagenesis, submitted for publication.

Pfuhler, S., Albertini, S., Fautz, R., Herbold, B., Madle, S., Utesch, D., Poth, A., 2007. Genetic toxicity assessment: employing the best science for human safety evaluation part IV: recommendation of a working group of the Gesellschaft fuer Umwelt-Mutationsforschung (GUM) for a simple and straightforward approach to genotoxicity testing. Toxicol. Sci. 97 (2), 237–240.

Pfuhler, P., Kirkland, D., Kasper, P., Hayashi, M., Vanparys, P., Carmichael, P., Dertinger, S., Eastmond, D., Elhajouji, A., Krul, C., Rothfuss, A., Schoening, G., Smith, A., Speit, G., Thomas, C., van Benthem, J., Corvio, R., 2009. Reduction of use of animals in regulatory genotoxicity testing: identification and implementation opportunities – report from an ECVAM workshop. Mutat. Res. 680, 31–42.

Potts, R.O., Bommannan, B., Guy, R.H., 1992. Percutaneous absorption. In: Mukhtar, H. (Ed.), Pharmacol. of the Skin. CRC Press, USA, pp. 12–27.

Sanner, T., Dybing, E., 2005. Comparison of carcinogenic and in vivo genotoxic potency estimates. Basic Clin. Pharmacol. Toxicol. 96 (2), 131–139.

SCCP, 2006a. SCCP/1005/06: The SCCP's Notes of Guidance for the Testing of Cosmetic Ingredients and their Safety Evaluation, Adopted by the SCCP During the 10th Plenary Meeting of 19 December 2006.

SCCP, 2006b. SCCP/0971/06 Updated Recommended Strategy for Testing Oxidative Hair Dye Substance for Their Potential Mutagenicity/Genotoxicity adopted 28, March 2006.

SCCP, 2007. SCCP/1111/07, Memorandum on Actual Status of Alternative Methods on the Use of Experimental Animals in the Safety Assessment of Cosmetic Ingredients in the European Union.

SCCP, 2009. SCCP/1212/09. Position statement on genotoxicity/mutagenicity testing of cosmetic ingredients without animal experiments.

Scott, D., Galloway, S.M., Marshall Jr, R., Ishidate Jr., M., Brusick, D., Ashby, J., Myhr, B.C., 1991. Genotoxicity under extreme culture conditions. A report from ICPEMC Task group 9. Mutat. Res. 257, 147–204.

Speit, G., 2009. How to assess the mutagenic potential of cosmetic products without animal tests? Mutat. Res. 678 (2), 108–112.

Tassaneeyakul, W., Birkett, D.J., Veronese, M.E., McManus, M.E., Tukey, R.H., Quattrochi, L.C., Gelboin, H.V., Miners, J.O., 1993. Specificity of substrate and inhibitor probes for human cytochromes P450 1A1 and 1A2. J. Pharmacol. Exp. Ther. 265 (1), 401–407.

Tennant, R.W., Margolin, B.H., Shelby, M.D., Zeiger, E., Haseman, J.K., Spalding, J., Caspary, W., Resnick, M., Stasiewicz, S., Anderson, B., Minor, R., 1987. Prediction of chemical carcinogenicity in rodents from in vitro genetic toxicity assays. Science 236, 933–941.

Tweats, D.J., Scott, A.D., Westmoreland, C., Carmichael, P.L., 2007. Determination of genetic toxicity and potential carcinogenicity in vitro – challenges post the seventh amendment to the European cosmetic directive. Mutagenesis 22, 5–13.

Vanpary, P., Corvi, R., Aardema, M., Gribaldo, L., Hayashi, M., Hoffmann, S., Schechtman, L., 2010. ECVAM prevalidation of SHE and Balb/3T3 cell transformation assays. ALTEX, accepted for publication.

Wakata, Y., Miyamae, S., Sato, T., Suzuki, T., Morita, N., Asano, T., Awogi, Kondo, K., Hayashi, M., 1998. Evaluation of the rat micronucleus test with bone marrow and peripheral blood: summary of the 9th collaborative study by CSGMT/JEMS/MMS. Environ. Mol. Mutagen. 32, 84–100.

Westerink, W.M., Schoonen, W.G., 2007. Cytochrome P450 enzyme levels in HepG2 cells and cryopreserved primary human hepatocytes and their induction in HepG2 cells. Toxicol. In Vitro 21 (8), 1581–1591.

Williams, K., Fowler, P., Jeffrey, L., Young, J., Carmichael, P., Aardema, M., Diembeck, W., Fautz, R., Harvey, J., Hewitt, N., Latil, A., Pfuhler, S., Quedraogo, G., Reisinger, K., Marrec-Fairley, M., Kirkland, D., 2009. Reduction of misleading ("False") positive results in mammalian cell genotoxicity assays. II: Importance of accurate toxicity measurement: VII. World Congress on Alternatives & Animal Use in the Life Sciences, Rome, Italy. 30th August–3rd September, 2009. Altex 26, special issue, p. 189. http://altweb.jhsph.edu/altex/26_special/index.ht.

Zeiger, E., 1998. Identification of rodent carcinogens and noncarcinogens using genetic toxicity tests: premises, promises and performance. Regul. Toxicol. Pharmacol. 28, 85–95.

Zeller, A., Richoz, L., Corbino, L., Pfuhler, S., 2007. Use of the comet assay to detect genotoxic effects in reconstituted human skin. Society of Toxicology 46th Annual Meeting, Charlotte, NC. Abstract #1171. The Toxicologist 96 (1), 243.