



The Cosmetics Europe strategy for animal-free genotoxicity testing: Project status up-date



S. Pfuhler^{a,*}, R. Fautz^b, G. Ouedraogo^c, A. Latil^d, J. Kenny^e, C. Moore^f, W. Diembeck^g, N.J. Hewitt^h, K. Reisingerⁱ, J. Barroso^j

^a Procter & Gamble Co., 1 Procter and Gamble Plz, Cincinnati, OH, USA

^b Kao Germany GmbH, Pfungstädterstraße 92 – 100, D-64297 Darmstadt, Germany

^c L'Oreal Life Sciences Research, Aulnay sous Bois, France

^d Pierre Fabre, 3 Rue des satellites, 31432 Toulouse, France

^e GSK, Park Road, Ware, SG12 ODP, UK

^f Unilever, Colworth House, Sharnbrook, Bedford, MK44 1LQ, UK

^g Beiersdorf AG, Unnastrasse 4, D-20253 Hamburg, Germany

^h SWS, Wingertrasse 25, D-64390 Erzhausen, Germany

ⁱ Henkel AG & Co., KGaA, Henkelstraße 67, D-40191 Duesseldorf, Germany

^j Cosmetics Europe, Avenue Herrmann Debroux 40, B-1160 Auderghem, Brussels, Belgium

ARTICLE INFO

Article history:

Received 11 October 2012

Accepted 18 June 2013

Available online 27 June 2013

Keywords:

Genotoxicity

Cosmetics

Alternatives to animal testing

3D

Skin

Metabolism

ABSTRACT

The Cosmetics Europe (formerly COLIPA) Genotoxicity Task Force has driven and funded three projects to help address the high rate of misleading positives in *in vitro* genotoxicity tests:

The completed “False Positives” project optimized current mammalian cell assays and showed that the predictive capacity of the *in vitro* micronucleus assay was improved dramatically by selecting more relevant cells and more sensitive toxicity measures.

The on-going “3D skin model” project has been developed and is now validating the use of human reconstructed skin (RS) models in combination with the micronucleus (MN) and Comet assays. These models better reflect the in use conditions of dermally applied products, such as cosmetics. Both assays have demonstrated good inter- and intra-laboratory reproducibility and are entering validation stages.

The completed “Metabolism” project investigated enzyme capacities of human skin and RS models. The RS models were shown to have comparable metabolic capacity to native human skin, confirming their usefulness for testing of compounds with dermal exposure.

The program has already helped to improve the initial test battery predictivity and the RS projects have provided sound support for their use as a follow-up test in the assessment of the genotoxic hazard of cosmetic ingredients in the absence of *in vivo* data.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

The focus of many researchers has been to develop better *in vitro* tools to replace animal tests, especially in the light of regulations such as the 7th amendment to the Cosmetics Directive (EU, 2003) and REACh (European Commission, 2006). The use of *in vitro* models is especially relevant to the cosmetics industry which is banned from using animal tests for a number of endpoints, including genotoxicity. Thus, positive outcomes in standard

in vitro assays evaluating the genotoxic potential of chemicals can no longer be followed-up with *in vivo* assays. If genotoxicity is assessed using only *in vitro* assays, this may well result the de-selection of many safe new products since these assays have a high rate of positive results that do not correlate with *in vivo* genotoxicity or carcinogenicity (Kirkland et al., 2005). This problem is recognized as a critical issue and has led to a number of working groups investigating improved approaches for assessing genotoxicity (e.g., International Life Sciences Institute – Human and Environmental Sciences Institute’s (ILSI-HESI) committee on The Relevance and Follow-up of Positive Results in *in vitro* Genotoxicity Testing (IVGT); www.hesiglobal.org). As part of an international and multi-laboratory collaboration, Cosmetics Europe (formerly “COLIPA”) has funded and driven numerous projects aimed to address the lack of adequate alternatives to traditional *in vivo* tests and help validate successful models.

* Corresponding author. Tel.: +1 513 319 7468.

E-mail addresses: pfuhler.s@pg.com (S. Pfuhler), Rolf.Fautz@kao.com (R. Fautz), gouedraogo@rd.loreal.com (G. Ouedraogo), alain.latil@pierre-fabre.com (A. Latil), julia.d.kenny@gsk.com (J. Kenny), Claire.Moore@Unilever.com (C. Moore), Walter.Diembeck@Beiersdorf.com (W. Diembeck), nickyhewittld@yahoo.co.uk (N.J. Hewitt), kerstin.reisinger@henkel.com (K. Reisinger), jbarroso@cosmeticseurope.eu (J. Barroso).

The Cosmetics Europe Genotoxicity Task Force was set up to coordinate and drive three main projects. The aim of the “False Positives” project was to optimize current mammalian cell assays by focusing on two aspects of the micronucleus (MN) test, namely the cell type employed and the method of cytotoxicity measurement. The “3D skin model” project aimed to develop and validate a new assay incorporating the use of human reconstructed skin (RS) models with the genotoxicity endpoints, MN and Comet assays. Since the exposure of many compounds is dermal, especially for cosmetics, these models better reflect their in use conditions. In order to interpret outcomes from the 3D model assays, knowledge of the metabolic capacity of these RS models is an advantage, especially in comparison with native human skin. Therefore, the “Metabolism” project investigated the enzyme capacities of human skin and compared them with that in RS models and with 2D monolayer cultures of skin cells.

Here, we review the outcomes of the Genotoxicity Task Force projects, outline future research aims and apply the knowledge gained to a decision tree approach to assess the genotoxic potential of chemicals used in the cosmetics industry.

2. The “False Positives” project

One of the drawbacks of the current *in vitro* clastogenicity assays is that they produce many positive results when compared with negative rodent carcinogenicity data. Indeed, an analysis of published data revealed the rate of misleading positives to be at least 80% when three *in vitro* assays were combined into a test battery (Kirkland et al., 2005), as requested by several guidelines (such as Reach European Commission, 2006 and the SCCS notes of guidance SCCP, 2009). Therefore, existing and new assays need to show better specificity (i.e. correctly identify non-carcinogens) without compromising sensitivity (i.e. still detecting *in vivo* genotoxins and DNA-reactive carcinogens). The “False positive” project was set up to optimize the *in vitro* MN test such that the methodology described in the OECD guideline 487 (OECD, 2010) was retained.

A European Center for the Validation of Alternative Methods (EC-VAM) workshop held in 2006 discussed ways to reduce the frequency of misleading positive results. Several suggestions for possible improvements/modifications to existing tests were identified, including the cell lines used and the method of cytotoxicity measurement (Kirkland et al., 2007). Following from these recommendations we started a program which was executed at Covance Laboratories, UK, where we investigated how the predictive capacity of the *in vitro* MN test is impacted by the choice of cell lines (Fowler et al., 2012a) and toxicity parameters (Fowler et al., 2012b). Others have compared different cell types (Sofuni et al., 1990; Hilliard et al., 2007; Erexson et al., 2001) but this is the first comparison that includes six cell types within the same project and using consistent, GLP-like experimental conditions including the same batches of chemicals, media and formulations (Fowler et al., 2012a). These investigations revealed that certain cell types were more prone to misleading positive responses than others, particularly rodent cell lines (Fig. 1). The p53 compromised rodent cell lines, CHL, CHO and V79, demonstrated poorer specificity than the p53 functional human cell types, TK6, HepG2 and human peripheral blood lymphocytes (HuLy) in response to compounds accepted as producing misleading positive results in *in vitro* clastogenicity assays (Kirkland et al., 2008). Cells of human origin may also be more favorable than rodent cells since these are more representative of human responses and may contain more human-specific metabolising enzymes and transporter proteins than rodent-derived cell types. Improving specificity cannot be at the expense of compromised sensitivity and thus it was necessary to make sure that the choice of cells did not lead to a decrease in sensitivity. Seventeen carcinogens that

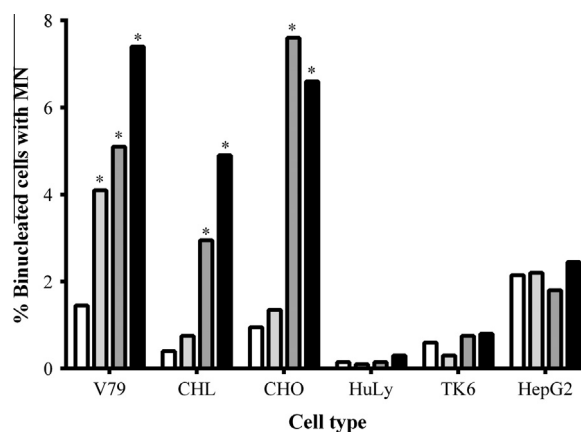


Fig. 1. A comparison of the formation of MN in different cell types. Each cell type was tested with solvent (□) and low (◻), medium (◼) and high (■) concentrations of resorcinol for 3 h in the presence of S9. Resorcinol caused statistically significant increases (denoted as “*”) in MN formation in V79, CHL and CHO cells but not in human lymphocytes (HuLy), TK6 cells or HepG2 cells. (Reproduced from Fowler et al., 2012a).

are thought to act via genotoxic mechanism (from the Group 1 chemicals from Kirkland et al., 2008) were re-tested in human lymphocytes and TK6 cells. Overall the data show that out of a panel of 17 genotoxic chemicals, TK6 and HuLy detect the majority of them as positive (15 out of 17; 88% accuracy), confirming the high sensitivity of these cells (Fowler et al., 2013).

Investigations of the cell type employed the replication index (RI) as a measure of cytotoxicity; however, it was considered that cytotoxicity may also be a parameter that could influence the outcome of the assay since this is used to select the top concentration tested (Kirkland et al., 2007). For example, one study showed that Relative Cell Counts (RCC) underestimated the toxicity of a number of direct and indirect genotoxins and thus selected higher concentrations for subsequent MN analysis (Fellows et al., 2008). Therefore, additional measures of cytotoxicity of a number of misleading positives in CHO, CHL and TK6 cells were investigated in the “False Positives” project, namely Relative Population Doubling (RPD), Relative Increase in Cell Counts (RICC), RI and RCC. Results revealed that estimation of toxicity based on relative proliferation increases (RPD and RICC) tended to select concentrations in the target toxicity range (50–60%) that give mainly negative MN responses (Fig. 2). Conversely, measurements of RCC and RI selected concentrations in the target toxicity range gave mainly positive MN responses. Therefore, using RICC and RPD to select concentrations for MN analysis reduces the number of misleading positive results, regardless of cell type, by selecting lower concentrations for analysis than when RCC or RI are used. The use of RICC and RPD does not result in a lower sensitivity of the assay. Others have shown that when RPD and RICC were used as measures of cytotoxicity of 14 known genotoxic agents, all selected concentrations for MN analysis gave rise to expected positive responses in a range of commonly used cell types (Kirkland, 2010). Another consideration is the effect of apoptosis inducing chemicals, such as curcumin and ethyl acrylate, which could contribute to the positive MN responses. In p53-competent TK6 cells, concentrations selected for MN analysis using the RCC and RI toxicity measures also caused increased levels of caspase activity, suggesting apoptosis had occurred. By contrast, when RPD and RICC were used to select concentrations, caspase levels were not significantly elevated, thus apoptosis was avoided and MN frequencies were close to background levels (see Fowler et al., 2012b for more detail on all of the results discussed above).

It was concluded that a combination of careful selection of the cell type and toxicity measurement can significantly increase the

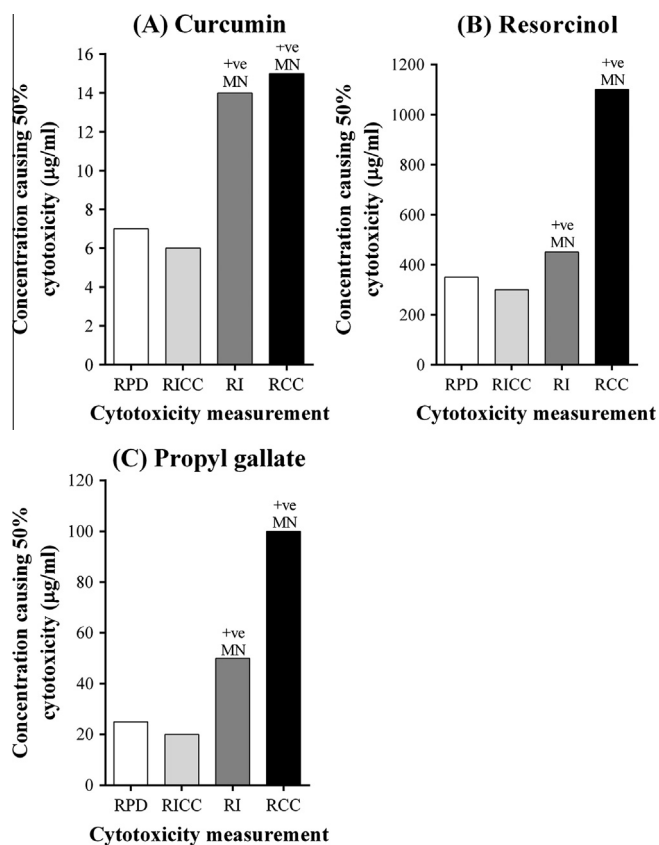


Fig. 2. Concentrations of misleading positive compounds resulting in 50–60% toxicity in TK6 cells according to different measurements. RPD and RICC tended to result in lower concentrations causing 50–60% toxicity and subsequently, fewer misleading positive responses. Bars with “+ve MN” indicate that the concentration indicated caused a statistically significant increase in MN formation (Reproduced from Fowler et al., 2012b).

specificity of clastogenicity assays (see also conclusions from an IWGT workshop reviewed by Pfuhrer et al., 2011). The now completed “False Positives” project has undoubtedly led to an improvement in the predictive capacity of the *in vitro* MNT. The higher predictive power of the *in vitro* assays will mean that fewer promising chemicals will be dropped from development without compromising safety of these ingredients.

3. The “3D skin model” project

The first site of exposure to many cosmetic ingredients is the skin and, as a result, these chemicals may not enter the systemic circulation due to the barrier properties of the stratum corneum, or may be metabolized in the skin prior to entering the systemic circulation. Therefore, the most relevant model in which to test these chemicals is human skin. However, the availability of fresh *ex vivo* human skin is sporadic which, together with the donor variability and differences in tissue quality, makes the use of this model impractical for routine testing. Alternatives to native human skin are RS models which are prepared from primary human keratinocytes. These have a structure (Ponec et al., 2001) and metabolic capacity (Luu-The et al., 2009; Hu et al., 2010; Götz et al., 2012a,b; van Eijl et al., 2012) similar to that of native skin, making RS models a relevant and predictive model with which to test the genotoxic potential of dermally exposed chemicals. Indeed, the use of RS models in genotoxicity risk assessment has been described recently (Pfuhrer et al., 2010). The *in vitro* human reconstructed skin micronucleus (RSMN) and Comet assays were developed for evaluating the genotoxicity of

dermally applied chemicals. The project has been run according to a modular approach:

- Phase 1: Optimization and transferability of method across different laboratories.
- Phase 2: Inter- and intra-laboratory reproducibility.
- Phase 3: Increasing the domain of chemicals tested for predictive capacity and further evaluation of reproducibility.

The two endpoints have so far produced very promising results, as described below and summarized in Table 1.

4. The reconstructed skin micronucleus (RSMN) assay

The *in vitro* MN assay was adopted for use with RS models to assess the genotoxic potential of a number of chemicals selected by an independent Chemical Selection Expert team, including positive genotoxins with different mechanisms of action, true negatives and misleading positives (Kirkland et al., 2008). The initial protocols were defined (Curren et al., 2006) and then tested in three US laboratories (Mun et al., 2009; Hu et al., 2009). In Phase 1, the RSMN assay method was transferred to Henkel and L'Oréal, both European-based (Germany and France, respectively), which was of significance because the RS models were supplied from a US provider (MatTek, MA). As part of this process, two training workshops were held to standardize the protocol and harmonize scoring of micronuclei, both of which were subsequently described and published by Dahl et al. (2011). The workshops outlined three key issues impacting on the assay, namely the shipping (which should be overnight and under cooled conditions); the solvents used (avoid those that interfere with the air–liquid interface of the EpiDerm™ model); and the subjective nature of the scorer (solved by creating a scoring atlas described by Dahl et al., 2011).

In Phase 2, three coded compounds (N-ethyl-N-nitrosourea (ENU), MMC (both genotoxic carcinogens) and cyclohexanone (non-carcinogen and non-genotoxic)) were tested by three laboratories (Aardema et al., 2010). In addition to a good reproducibility between experiments within each laboratory, there was also a good inter-laboratory reproducibility for all three chemicals tested. Moreover, the genotoxic activity of each chemical was correctly identified in each laboratory.

In Phase 3, the number of coded chemicals was increased to 29 as part of the validation process. All results were sent to ECVAM for decoding and evaluation according to specific pre-determined criteria. Results demonstrated an excellent specificity such that

Table 1

Summary of the validation of the RS Comet and RSMN assays.

	RSMN assay	RS Comet assay
Phase 1	Completed: Optimization of incubation conditions. Correct prediction of 5 dermal non-carcinogens and 7 model genotoxins including genotoxic dermal carcinogens	Completed: Assay readily adapted and transferred to different laboratories. Good intra- and inter-laboratory reproducibility with 2 model genotoxins
Phase 2	Completed: Transfer to different laboratories. Good intra- and inter-laboratory reproducibility of MN responses to 3 different coded chemicals. Correct identification of positive and negative genotoxicants	Completed: Good intra- and inter-laboratory reproducibility of responses to 5 different coded chemicals. Correct identification of positive and negative genotoxicants. Optimization of RS model transport
Phase 3	On-going: Number of chemicals tested increased to 29. Initial results show improved specificity of the RSMN assay. Requires testing of additional positive chemicals to confirm sensitivity of the assay	Not started. Project is merging with a project sponsored by the “Bundesministerium für Bildung und Forschung” (BMBF) to more efficiently test the validation chemicals

approximately 90% of the experiments predicted *in vivo* non-genotoxic non-carcinogens correctly (Fautz et al., 2012). Of the 29 chemicals tested, only 8 chemicals were classified as carcinogens with a suggested genotoxic mode of action and 21 were non-carcinogens. Therefore, the current dataset is biased towards non-carcinogens with the total number of carcinogens in the dataset considered too low to draw a final conclusion about the sensitivity of the RSMN assay. More coded compounds will be tested in the next project phase with a focus on carcinogens.

During the testing, it was noted that a couple of compounds precipitated, which was not considered in the initial criteria, but should be included in future evaluations. A misleading negative may arise if the intended concentration is not reached or false positives can be caused by precipitation due to practical issues such as difficulties in scoring (especially with compounds that fluoresce at the same wavelengths as acridine orange) or disruption of the air-liquid interface which can cause MN formation (Dahl et al., 2011).

In order to speed up the scoring of the MN, efforts towards automation are on the way. This should enable analysis of a greater number of cells, resulting in a higher statistical power of the assay and incorporate an automatic cytotoxicity measurement as part of the analysis. Initial results show good comparisons between manual scoring and flow cytometric methods (unpublished data).

Some genotoxins require metabolic activation; therefore, we have investigated a number of chemicals that fall into this category, namely 4-nitroquinoline-n-oxide (4NQO), cyclophosphamide, dimethylbenzanthracene (DMBA), dimethylnitrosamine (DMN), dibenzanthracene (DBA) and benzo[a]pyrene (BaP). Since the skin has been shown to have a very low phase 1 (normally bioactivating) capacity, it was considered that these chemicals may require a longer incubation duration in order to generate sufficient levels of the ultimate genotoxin. However, extending the dosing regimen from 48 h and two doses to 72 h and three doses did not always change the outcome of the assay, such that cyclophosphamide and DMBA were positive and DBA and DMN were negative using both dosing regimen. The outcome of the assay was only changed for 4NQO, which was negative in the standard 48 h dosing regimen, but positive with the 72 h treatment (Aardema et al., 2012). BaP gave mixed results, possibly due to this chemical precipitating at high concentrations and to alterations in metabolic enzyme levels caused by BaP (Götz et al., 2012c). The results for DBA and DMN may be expected since, for DBA, the level of CYP1A1/2 (which is needed for bioactivation of this molecule) is very low in both native skin and RS models (including EpiSkin[®] and EpiDerm[®] models (van Eijl et al., 2012; Hu et al., 2010) such that the initial activation step is not possible or efficient enough in this organ. The low CYP1A1/2 expression is common to a number of RS skin models and to native human skin (van Eijl et al., 2012; Götz et al., 2012a) and is not a shortcoming of the EpiDerm[®] model or 3D cultures in general. The lack of a genotoxic effect by DMN is reflected in the cancer bioassay, in which this compound did not cause tumours in the skin after application to the skin of rats (although it did result in tumours in the liver, lung and kidney) (Benemanskii and Levina, 1985). This suggests that DMN is not bioactivated in the skin or, at least to a sufficient extent to cause MN. Based on the result observed for 4NQO, it is recommended that a 48 h treatment is used for general testing and a longer treatment period is used when the outcome of the standard 48 h treatment is negative or questionable. This practice would not cost extra time, since a negative result in the first 48 h experiment would be repeated using the 72 h dosing regimen instead of 48 h.

These data support the conclusion that the RSMN assay in EpiDerm[™] is a valuable *in vitro* method for genotoxicity assessment of dermally applied chemicals. The above mentioned global validation project sponsored by Cosmetics Europe and ECVAM is contin-

uing to collect data with a goal of demonstrating the strengths and limitations of this method.

5. The reconstructed skin Comet assay

As with the RSMN assay, the RS Comet project was based on EpiDerm[™] tissues. In contrast to the RSMN assay, in which the compounds are applied for at least 48 h, this assay involves the topical exposure to the chemicals for at least 3 h, followed by cell isolation and assessment of DNA damage using the Comet assay.

Phase 1 studies have been completed. The method using 3D EpiDerm[™] tissues was readily adapted and transferred to different laboratories, showing good intra- and inter-laboratory reproducibility with two model genotoxins, methyl methane sulfonate (MMS) and 4-NQO, in accordance with *in vivo* data. In Phase 2, all laboratories tested five coded chemicals, and the reproducibility with these chemicals was generally good (Reuss et al., 2013).

Phase 3 of this project has not yet started as we are still exploring opportunities to optimize the assay. Considerable intra- and inter-experimental variability was observed in some experiments, in which the solvent control values were greater than 30% (measured as % tail DNA). This variability was not attributable to a single factor but was thought to be a stress-induced negative impact of transport on the quality of the tissues. The most promising way forward along with the below described use of full thickness models is the use of “underdeveloped” (EPI-201) skin tissues rather than the standard MatTek epidermal (EPI-200) tissues. The underdeveloped EPI-201 tissues consist of the same normal, human-derived epidermal keratinocytes and are cultured in the same manner as standard EPI-200 tissues, but the underdeveloped tissues are shipped 4 days earlier than the normal tissues (on the week preceding the Comet assay). The EPI-201 tissues are then cultured further for 4 days in the receiving laboratory to produce tissues which are equivalent to the normal EPI-200 tissues. The additional time in culture after shipping may allow the tissues to recover from the transport-induced stress effects. The use of underdeveloped tissues changed the rate of invalid experiments due to a high background (>30% tail DNA in the untreated or vehicle treated group) in one laboratory from 4 in 8 experiments to 1 in 9 experiments. Despite the use of these tissues lowering the background, the response of the tissues to the positive control, MMS, was unaltered, suggesting that this modification has not decreased the sensitivity of the assay (Reuss et al., 2013).

The Cosmetics Europe RS Comet project is merging with a project sponsored by the German Federal Ministry of Education and Research (“Bundesministerium für Bildung und Forschung”, BMBF) in order to combine efforts and more efficiently test the validation chemicals in Phase 3. In the combined project we will be investigating the usefulness of full-thickness skin models, aside from EPI 200/201, since they have been shown to have a higher metabolic capacity than epidermal models (Jäckh et al., 2011) and may better suited to address the genotoxic properties of pro-mutagens.

6. The “metabolism” project

This project investigated xenobiotic metabolising enzymes (XMEs) in native human skin, RS and monolayer cultures of skin cells using both a proteomic approach and measurement of substrate metabolism. The proteomic methods included immunoblotting and a technique involving LC-MS/MS analysis of peptides and subsequent software analysis (van Eijl et al., 2012). The latter method has the advantage of a much higher sensitivity than traditional immunoblotting techniques and it allows for a comprehensive analysis of over 2000 XMEs. Phase 1 and 2 XME activities were measured using enzyme-selective substrates for cytochrome P450s (CYPs), sul-

fotransferases (SULTs) and UDPGA-glucuronosyltransferases (UGTs) and glutathione S-transferases (GSTs) (Götz et al., 2012a, b,c). CYP 1–3 family proteins were not detected in native whole human skin or any of the *in vitro* models tested, which reflects the low mRNA expression of these CYPs reported by others (Luu-The et al., 2009) and the low or lacking metabolism of CYP-selective substrates in our studies (Götz et al., 2012a). The abundance of CYP1–3 proteins in human skin was estimated to be at levels at least 300-fold lower than that of liver. However, there were multiple other phase 1 XME proteins that were present in significant levels, such as alcohol dehydrogenases, aldehyde dehydrogenases, amine oxidases and epoxide hydrolases. GST proteins were the most abundant of the phase 2 enzymes investigated, and were present in both native human skin and EpiDerm™ models. GST Pi was also identified as the most abundant isoform (van Eijl et al., 2012), which correlates with the high mRNA expression of this enzyme (Luu-The et al., 2009; Hu et al., 2010). The GST substrate, CDNB (Sherratt and Hayes, 2002), was metabolised at appreciable rates in whole skin (~20 nmol/min/mg), although this is still lower than that reported to be present in human liver (van Eijl et al., 2012; Götz et al., 2012b). The potential routes of metabolism in human skin and liver, based on their proteomic XME profiles, are depicted in Fig. 3. The overall results from this project supported the view that skin tended to be more of a detoxification than a bioactivation organ (in contrast to the liver) and that the levels of XMEs were all generally much lower than the liver.

The XME profiles, using Affymetrix gene analysis, of different donors of EpiDerm™ models has been reported to be very similar (Hu et al., 2010) but there are no reports on how and if XMEs change during the course of an assay. Therefore, measurement of XMEs was adapted to determine how genotoxic compounds affect XME activities in EpiDerm™ models used under the conditions of the RSMN assay (Götz et al., 2012c). The genotoxins used were BaP and cyclophosphamide, both of which require bioactivation to their ultimate genotoxic metabolites. BaP caused a marked increase in ethoxyresorufin O-deethylase (reflecting CYP1A and CYP1B activity, also responsible for the bioactivation of BaP Shimada and Fujii-Kuriyama, 2004) activities even 24 h after the first dose; and these raised levels continued until the final dose was applied (three doses in total and 72 h of exposure). Since the CYP1A/1B pathway was significantly increased, this would suggest

that BaP may well cause an increase in MN formation in the EpiDerm™ models in the RSMN assay. This was indeed the case for two experiments but not a third (Aardema et al., 2012), suggesting that there is an even balance of bioactivation via CYPs and detoxification by GSTs (which were not induced by BaP) which can be tipped either way in different experiments. Unlike BaP, cyclophosphamide did not alter any of the enzymes measured (Götz et al., 2012c), despite their involvement in its metabolism. Unlike BaP, cyclophosphamide was consistently positive in all experiments in both laboratories that tested this compound (Aardema et al., 2012). This suggests that the balance of metabolism of cyclophosphamide in EpiDerm™ models is towards bioactivation. Overall, the measurement of certain XME activities in EpiDerm™ models treated under the conditions of an endpoint assay helped in the interpretation of the outcome.

7. Summary

The Cosmetics Europe Genotoxicity Task Force projects, which have been running over the course of the last five years, have helped improve predictive capacity of *in vitro* clastogenicity assays, and resulted in an increased understanding of the methods used to predict the *in vivo* genotoxic potential of dermally applied chemicals. Important protocol modifications, namely choice of the cell type and cytotoxicity measurement, have resulted in improved specificity of the MN test such that over 60% irrelevant positive findings could be prevented by using the optimized methods. This in itself will lead to an increase of the predictive capacity of the initial test battery of *in vitro* tests and therefore reduce the number of chemicals de-selected due to misleading positives. The “3D skin model” project has shown that genotoxic endpoints, such as the MN and Comet assays, can be adapted to RS models and that the protocols developed are robust, as demonstrated by the high degree of reproducibility between and within laboratories when coded compounds were tested. By establishing good predictivity of the RS MN and Comet models, together with the confirmation that the RS models mimic native human skin in terms of their metabolic capacity (demonstrated in the “Skin Metabolism project”), our results will support their use in follow-up tests in the assessment of the genotoxic hazard of cosmetic ingredients in the absence of *in vivo* data.

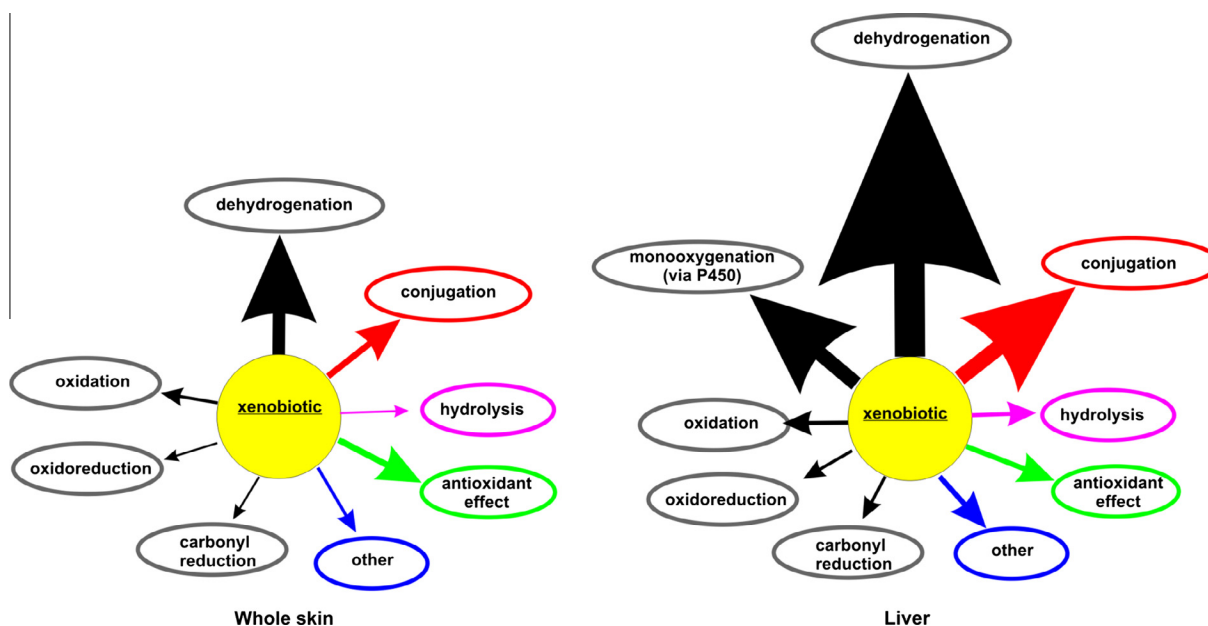


Fig. 3. Potential routes of xenobiotic metabolism in skin and liver. The size of each arrow is proportional to the number of XMEs detected that may catalyze each bioconversion indicated. (Taken from van Eijl et al., 2012 with kind permission from PLoS ONE, <http://dx.doi.org/10.1371/journal.pone.0041721.g004>.)

Conflict of interest

There are no conflicts of interests for any of the authors.

Acknowledgements

This work was funded by the Cosmetics Europe. The “3D skin model” project received input from ECVAM. Work by Covance and TNO was co-funded by ECVAM and UK NC3Rs. We would like to thank David Kirkland and Paul Fowler for their dedicated contributions to the “False Positives” project. We would also like to thank Marilyn Aardema for her role in the RSMN project and Raffaella Corvi for her help with steering the program.

References

- Aardema, M.J., Barnett, B.C., Khambatta, Z., Reisinger, K., Ouedraogo-Arras, G., Faquet, B., Ginestet, A.C., Mun, G.C., Dahl, E.L., Hewitt, N.J., Corvi, R., Curren, R.D., 2010. International prevalidation studies of the EpiDerm 3D human reconstructed skin micronucleus (RSMN) assay: transferability and reproducibility. *Mutat. Res.* 701 (2), 123–131.
- Aardema, M.J., Barnett, B., Mun, G., Dahl, E., Curren, R., Hewitt, N.J., Pfuhrer, S., 2012. Evaluation of chemicals requiring metabolic activation in the EpiDerm™ 3D human reconstructed skin micronucleus (RSMN) assay. *Rev. Version Mut. Res.*
- Benemanskii, V.V., Levina, V.A., 1985. Carcinogenic effect of N-nitrosodimethylamine after application to rat skin. *Eksp. Onkol.* 7 (2), 20–21.
- 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.
- Dahl, E.L., Curren, R., Barnett, B.C., Khambatta, Z., Reisinger, K., Ouedraogo, G., Faquet, B., Ginestet, A.C., Mun, G., Hewitt, N.J., Carr, G., Pfuhrer, S., Aardema, M.J., 2011. The reconstructed skin micronucleus assay (RSMN) in EpiDerm™: detailed protocol and harmonized scoring atlas. *Mutat. Res.* 720 (1–2), 42–52.
- Erexson, G.L., Periago, M.V., Spicer, C.S., 2001. Differential sensitivity of Chinese hamster V79 and Chinese hamster ovary (CHO) cells in the in vitro micronucleus screening assay. *Mutat. Res.* 495 (1–2), 75–80.
- EU, 2003. EC – Directive 2003/15/EC of the European parliament and of the council of 27 February 2003 amending council directive 76/768/EEC on the approximation of the laws of the member states relating to cosmetic products. *Official Journal L66*, 11/03/2003, p. 26.
- 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>.
- Fautz, R., Curren, R., Krul, C., Reisinger, K., Ouedraogo, G., Corvi, R., Aardema, M., Reus, A., Barnett, B., Downs, T., Faquet, B., Hoffmann, S., Hewitt, N., Barroso, J., Pfuhrer, S., in: Poster presentation at the EPAA Annual Conference 2012 – Global Cooperation on alternatives (3Rs) to animal testing, Brussels, November, 2012: Pre-validation of the Reconstructed 3D Human Skin Micronucleus and Comet Assay. http://ec.europa.eu/enterprise/epaa/3_events/ann-conf-2012/poster-book.pdf.
- 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.
- Fowler, P., Smith, K., Young, J., Jeffrey, L., Kirkland, D., Pfuhrer, S., Carmichael, P., 2012a. Reduction of misleading (“false”) positive results in mammalian cell genotoxicity assays. I. Choice of cell type. *Mutat. Res.* 742 (1–2), 11–25.
- Fowler, P., Smith, R., Smith, K., Young, J., Jeffrey, L., Kirkland, D., Pfuhrer, S., Carmichael, P., 2012b. Reduction of misleading (“false”) positive results in mammalian cell genotoxicity assays. II. Importance of accurate toxicity measurement. *Mutat. Res.* 747 (1), 104–117.
- Fowler, P., Smith, R., Smith, K., Young, J., Jeffrey, L., Kirkland, D., Pfuhrer, S., Carmichael, P., 2013. Reduction of misleading (“false”) positive results in mammalian cell genotoxicity assays. III. Sensitivity of human cell types. *Mutat. Res.*, accepted for publication.
- Fowler, P., Smith, R., Smith, K., Young, J., Jeffrey, L., Kirkland, D., Pfuhrer, S., Carmichael, P., 2013. Reduction of misleading (“false”) positive results in mammalian cell genotoxicity assays. Accepted for publication in Mutation Research, III. Sensitivity of human cell types.
- Götz, C., Pfeiffer, R., Tigges, J., Hübenenthal, U., Ruwiedel, K., Freytag, E.-M., Merk, H.F., Krutmann, J., Edwards, R.J., Abel, J., Pease, C., Goebel, C., Hewitt, N.J., Fritsche, E., 2012a. Xenobiotic metabolism capacities of human skin in comparison to 3D-epidermis models and keratinocyte-based cell culture as in vitro alternatives for chemical testing: phase I. *Exp. Dermatol.* 21 (5), 358–363.
- Götz, C., Pfeiffer, R., Tigges, J., Hübenenthal, U., Ruwiedel, K., Freytag, E.-M., Merk, H.F., Krutmann, J., Edwards, R.J., Abel, J., Pease, C., Goebel, C., Hewitt, N.J., Fritsche, E., 2012b. Xenobiotic metabolism capacities of human skin in comparison to 3D-epidermis models and keratinocyte-based cell culture as in vitro alternatives for chemical testing: phase 2. *Exp. Dermatol.* 21 (5), 364–369.
- Götz, C., Hewitt, N.J., Jermann, E., Tigges, J., Kohne, Z., Hübenenthal, U., Krutmann, J., Merk, H.F., Fritsche, E., 2012c. Effects of the genotoxic compounds, benzo[a]pyrene and cyclophosphamide on phase 1 and 2 activities in EpiDerm™ models. *Xenobiotica* 42 (6), 526–537.
- 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., Kaluzhny, Y., Mun, G.C., Barnett, B., Karetzky, V., Wilt, N., Klausner, M., Curren, R.D., Aardema, M.J., 2009. Intralaboratory and interlaboratory evaluation of the EpiDerm 3D human reconstructed skin micronucleus (RSMN) assay. *Mutat. Res.* 673 (2), 100–108.
- Hu, T., Khambatta, Z.S., Hayden, P.J., Bolmarcich, J., Binder, R.L., Robinson, M.K., Carr, G.J., Tiesman, J.P., Jarrold, B.B., Osborne, R., Reichling, T.D., Nemeth, S.T., Aardema, M.J., 2010. Xenobiotic metabolism gene expression in the EpiDerm™ 3D human epidermis model compared to human skin. *Toxicol. In Vitro* 24 (5), 1450–1463.
- Jäckh, C., Blatz, V., Fabian, E., Guth, K., van Ravenzwaay, B., Reisinger, K., Landsiedel, R., 2011. Characterization of enzyme activities of Cytochrome P450 enzymes, Flavin-dependent monooxygenases, N-acetyltransferases and UDP-glucuronyltransferases in human reconstructed epidermis and full-thickness skin models. *Toxicol. In Vitro* 25 (6), 1209–1214.
- Kirkland, D., Aardema, M., Henderson, L., Müller, L., 2005. 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.
- Kirkland, D., Pfuhrer, 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., 2007. 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., Kasper, P., Müller, L., Corvi, R., Speit, G., 2008. Recommended lists of genotoxic and non-genotoxic chemicals for assessment of the performance of new or improved genotoxicity tests: a follow-up to an ECVAM workshop. *Mutat. Res.* 653 (1–2), 99–108.
- Kirkland, D., 2010. Evaluation of different cytotoxic and cytostatic measures for the in vitro micronucleus test (MNVit): summary of results in the collaborative trial. *Mutat. Res.* 702 (2), 139–147.
- 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.
- Mun, G.C., Aardema, M.J., Hu, T., Barnett, B., Kaluzhny, Y., Klausner, M., Karetzky, 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.
- OECD, 2010. OECD guideline for the testing of chemicals draft proposal for a new guideline 487: in vitro micronucleus test. (adopted 22.07.10). <http://iccvam.niehs.nih.gov/SuppDocs/FedDocs/OECD/OECD-TG487.pdf>.
- Pfuhrer, S., Kirst, A., Aardema, M., Banduhn, N., Goebel, C., Araki, D., Costabel-Farkas, M., Dufour, E., Fautz, R., Harvey, J., Hewitt, N.J., Hibatallah, J., Carmichael, P., Macfarlane, M., Reisinger, K., Rowland, J., Schellau, F., Schepky, A., Scheel, J., 2010. A tiered approach to the use of alternatives to animal testing for the safety assessment of cosmetics: genotoxicity. A COLIPA analysis. *Regul. Toxicol. Pharmacol.* 57 (2–3), 315–324. *Erratum in: Regul. Toxicol. Pharmacol.* 58(3), 544.
- Pfuhrer, S., Fellows, M., van Benthem, J., Corvi, R., Curren, R., Dearfield, K., Fowler, P., Frötschl, R., Elhajouji, A., Le Hégarat, L., Kasamatsu, T., Kojima, H., Ouedraogo, G., Scott, A., Speit, G., 2011. In vitro genotoxicity test approaches with better predictivity: summary of an IWGT workshop. *Mutat. Res.* 723 (2), 101–107.
- Ponec, M., Gibbs, S., Pilgram, G., Boelsma, E., Koerten, H., Bouwstra, J., Mommaas, M., 2001. Barrier function in reconstructed epidermis and its resemblance to native human skin. *Skin Pharmacol. Appl. Skin Physiol.* 14 (Suppl. 1), 63–71.
- Reuss, A., Reisinger, R., Downs, T.R., Carr, G., Zeller, A., Corvi, R., Krul C.A.M., Pfuhrer, S., 2013. Comet assay in reconstructed 3D human epidermal skin models – investigation of intra- and inter-laboratory reproducibility with coded chemicals, *Mutagenesis*, accepted for publication.
- SCCP, 2009. SCCP/1212/09. Position statement on genotoxicity/mutagenicity testing of cosmetic ingredients without animal experiments. http://ec.europa.eu/health/ph_risk/committees/04_sccp/docs/sccp_s_08.pdf.
- Sherratt, P.J., Hayes, J.D., 2002. Glutathione S-transferases. In: Ioannides, C. (Ed.), *Enzyme Systems that Metabolise Drugs and Other Xenobiotics*. Wiley & Sons, Chichester, UK, pp. 319–352.
- Shimada, T., Fujii-Kuriyama, Y., 2004. Metabolic activation of polycyclic aromatic hydrocarbons to carcinogens by cytochromes P450 1A1 and 1B1. *Cancer Sci.* 95 (1), 1–6.
- Sofuni, T., Matsuoka, A., Sawada, M., Ishidate Jr., M., Zeiger, E., Shelby, M.D., 1990. A comparison of chromosome aberration induction by 25 compounds tested by two Chinese hamster cell (CHL and CHO) systems in culture. *Mutat. Res.* 241 (2), 175–213.
- van Eijl, S., Zhu, Z., Cupitt, J., Gierula, M., Götz, C., Fritsche, E., Edwards, R.J., 2012. Elucidation of xenobiotic metabolism pathways in human skin and human skin models by proteomic profiling. *PLoS One* 7 (7), e41721.