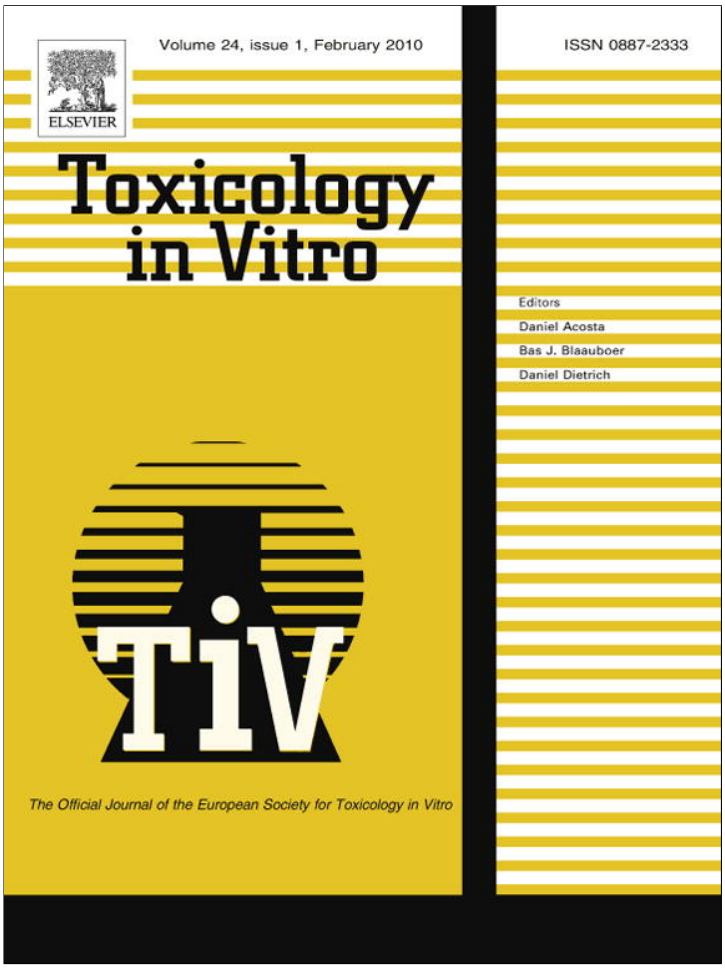

Toxicology in Vitro 24 (2010) 257–266

Contents lists available at ScienceDirect

Toxicology in Vitro

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

A catch-up validation study on reconstructed human epidermis (SkinEthic™ RHE) for full replacement of the Draize skin irritation test

Nathalie Alépée [a,*], Carine Tornier [b], Cécile Robert [c], Carole Amsellem [d], Marie-Hélène Roux [e], Olivier Doucet [c], Jean Pachot [e], Marisa Méloni [f], Anne de Brugerolle de Fraissinette [b]

[a] L'Oréal, 1 Avenue Eugène Schueller, 93600 Aulnay-Sous-Bois, France
[b] SkinEthic Laboratories, 45 rue Saint Philippe, 06000 Nice, France
[c] Coty, Athos Palace, 2 rue de la Lugernetta, 98000 Monaco, Monaco
[d] EPISKIN SNC, 4 rue Alexander Fleming, 69007 Lyon, France
[e] Oroxcell, Parc Biocitech, 102 Avenue Gaston Roussel, 93230 Romainville, France
[f] Vitroscreen, Via Mosè Bianchi, 103, 20145 Milano, Italy

ARTICLE INFO

Article history:
Received 11 May 2009
Accepted 26 August 2009
Available online 3 September 2009

Keywords:
Skin irritation
In vitro reconstructed human tissue (RhT)
Reconstructed human epidermis (RHE)
SkinEthic
"42 bis" Test method
Validation ECVAM

ABSTRACT

Efforts to fully replace the *in vivo* Draize skin irritation test, according to the Directive 67/548/ECC or OECD TG 404, were reinforced with the seventh Amendment of the Cosmetic Directive and the REACh regulation. In 2007, the EpiSkin™ test method was scientifically validated and recognized as the stand alone method to discriminate skin irritants (R38) from non-irritants (no label) according to the definition of the EU risk phrases. An ECVAM performance standards (PS) document was defined to evaluate the accuracy and reliability of other analogous test methods (ECVAM SIVS, May 2007). The present test was designed to determine the reliability and relevance of the Reconstructed Human Epidermis (RHE) model commercialized by SkinEthic™. The RHE skin irritation test method consisted to topically apply topically the test substances for 42 min followed by a 42 h post-incubation. The main selected endpoint was the cell viability (MTT reduction), with a threshold of 50% viability. The RHE test method showed a good intra and inter-laboratory reproducibilities in a multicentric study involving three independent laboratories. The SkinEthic™ RHE test method showed to be relevant and reliable with a sensitivity of 90% and a specificity of 80% (MTT only) and was not improved by integrating another endpoint such as IL-1α. The overall accuracy was 85% resulting in the recognition of the SkinEthic™ RHE test method, by the ECVAM Scientific Advisory Committee in November 2008, as a stand alone replacement test method for the Draize rabbit *in vivo* test, as a screen, or as part of a sequential testing strategy in a weight of evidence approach, for classifying non-irritant and irritant test substances, depending on country requirements.

1. Introduction

The potential for a test substance to cause skin irritation is an important consideration for safety procedures and is therefore required by the various regulatory authorities for notification and import of test substances. The OECD Guideline 404 for acute dermal irritation and corrosion testing of chemicals, and the guideline for non-clinical dermal tolerance testing of medicinal products are currently based on the method described by Draize et al. (1944) (European Commission: Method B.4 of Annex V to Directive 67/548/ECC). However, special attention is carried to be given to possible improvements in animal welfare. In addition, the European regulation for chemicals (REACh: Registration, Evaluation, Authorization, and Restriction of Chemicals, June 2007) prescribed the use of alternative methods to animal testing, as well as the seventh Amendment to the EU Cosmetics Directive, decrees a complete ban on animal testing for cosmetic ingredients (European Commission: Directive 76/768/ECC).

Since the 1980s, various reconstructed human epidermis models have been developed and used for skin irritancy testing such as EpiSkin™ (Tinois et al., 1991; Bell et al., 1981), LabCyte EPI-MODEL™ (Niwa et al., 2009), CellSystems EST-1000 (Hoffmann et al., 2005), EpiDerm™ (Cannon et al., 1994), OS-Rep Phenion® skin model technologies (Poumay et al., 2004) and SkinEthic™ RHE (Kandárová et al., 2006a). The importance of developing and validating methods for the prediction of skin irritation has been well documented by ECVAM which has funded prevalidation studies on five *in vitro* tests for acute skin irritation (Fentem and Botham, 2002). From 2003 to 2006, out of the five *in vitro* test methods

* Corresponding author. Tel.: +33 (0) 148689102; fax: +33 (0) 148689180.
E-mail address: nalepee@rd.loreal.com (N. Alépée).

0887-2333/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tiv.2009.08.024

initially evaluated, only the EpiSkin™ and EpiDerm™ skin irritation tests went through the whole validation process (Spielmann et al., 2007). A common test protocol was used for both reconstructed human epidermis models, and Episkin™ test method was described as the only stand alone test method for full replacement of the Draize *in vivo* skin irritation test showing sufficient sensitivity and specificity (Hoffmann, 2007; Eskes et al., 2007; ECVAM Statement, 2007). With the completion of the validation study, performance standards for applying human skin models to *in vitro* skin irritation testing were defined (ECVAM SIVS, 2007). These performance standards can be then used to evaluate the accuracy and reliability of other analogous test methods (also referred as "me-too" tests) that are based on similar scientific principles and measure or predict the same biological or toxic effect.

The aim of the present study was to demonstrate the similarity/equivalence of the *in vitro* SkinEthic™ RHE skin irritation assay in a formal inter-laboratory study, on the basis of the validated EpiSkin™ reference test method. The RHE tissue which mimics the three-dimensional structure of the human epidermis was first released by Martin Rosdy in 1989 (Rosdy and Clauss, 1990; Fartasch and Ponec, 1994; Kandárová, 2006; Kandárová et al., 2006a). The reconstructed epidermis test system was used in a toxicity approach to the prediction of skin irritation potential of pharmaceutical products (de Brugerolle de Fraissinette et al., 1999), and a specific protocol was developed for the testing of finished products applicable for cosmetic formulations (De Wever and Charbonnier, 2002). Usually a prediction model with a cut-off value of 50% viability, was defined to classify test substances as irritant (I) or non-irritant (NI) (Kandárová et al., 2006b; Tornier et al., 2006). In addition, previous works have demonstrated that the culture environment allowed the detection of cytokines released or secreted by the epidermis in response to topical application of test substances (Doucet et al., 1996; Coquette et al., 1999). An optimized test method based on a 42 min application for both liquids and solids followed by a 42 h post-incubation period before assessment of tissue viability was established by Tornier et al. (unpublished data) and was assessed in the present work in three laboratories, under blind conditions.

2. Materials and methods

2.1. Functional reconstructed human epidermis (RHE) model conditions

The SkinEthic™ reconstructed human epidermis (RHE) model consists of normal, human-derived keratinocytes, which have been cultured to form a fully differentiated three-dimensional epidermis cultured for 17-days on an inert polycarbonate filter at the air–liquid interface in a chemically defined medium (Rosdy and Clauss, 1990; Rosdy et al., 1993) (Fig. 1). Characterization and comparison of SkinEthic™ RHE model and the reference validated EpiSkin™ model were performed, demonstrating that these tissues test systems reproduced many of characteristics of normal human epidermis (Boelsma et al., 2000). The expression of epidermis differentiation markers like cytokeratin 1, filaggrin, transglutaminase, involucrin in RHE tissues has historically been described by Rosdy and Clauss (1990). The RHE model, which express cytokeratin 10 (CK10), a specific marker of keratinizing epidermis, also presents a differentiation pattern comparable to the EpiSkin™ model and to the normal human epidermis. The synthesis of specific lipids including ceramides, and particular acylceramides, which are known to be responsible for the water barrier of stratum corneum was detected (Ponec et al., 1988; Cumpstone et al., 1989; Kandárová, 2006).

The reconstructed 0.5 cm^2 epidermis was received on day 18, and maintained overnight in a nutrient medium at 37 °C, 5% CO_2, according to SkinEthic's procedures (SkinEthic Laboratories, France). Different batches of the RHE tissues were used by all participating laboratories during the multicentric study (runs 1–5). A quality control data sheet was provided with every batch of tissue including histology, viability and safety data. Each RHE batch is tested for MTT reduction viability test using negative control tissues. The mean optical density (OD) value for 41 cell donors (from 2000 to 2007) is 1.05 ± 0.10, demonstrating the reproducibility of this assay over the years. A resulting OD value threshold acceptance limit of 0.7 has been established. Prior testing, the barrier function property of the tissue should also meet the historical acceptability range defined by the producer (SkinEthic, ISO 9001 version 2008). This property is estimated by the exposure time required to reduce cell viability by 50% (ET-50) upon application of 1% Triton X-100. The reproducibility of the ET-50 on RHE model has been evaluated on 21 cell donors from 2005 to 2007, with an overall mean for $ET\text{-}50_{TritonX\text{-}100}$ values of 5.58 ± 0.83 h. An upper ET-50 threshold acceptance limit has been established for the $ET\text{-}50_{TritonX\text{-}100}$ criteria ranging from 4 to 9 h. The third SkinEthic quality control criteria, is based on histological examination. The tissues exhibit four to seven viable layers comprising at least basal, suprabasal, spinous and granular cell layers, and a stratum corneum. The overall mean of living cell layers over 41 cell donors from 2000 to 2007 is 5.51 ± 0.55.

2.2. Selection of test substances

The reference test substances (10 non-irritants, 10 irritants), used in this study were selected in accordance with the Performance Standard document (ECVAM SIVS, 2007). Both liquid and solid test substances were included in the experiments (Table 1). It is important to note that three misclassified test substances were selected for improvement assessment of the performance with a new skin model, or modified test protocol compared to the *in vitro* validated method (Spielmann et al., 2007). The 20 test substances were coded by Vitroscreen and tested under blind conditions in three laboratories: L'Oréal (Lead Lab, France), Coty (Monaco), and Oroxcell (France).

Sodium dodecyl sulfate (SDS 5% W/V aqueous solution) was used as reference irritant (positive control) and phosphate buffer

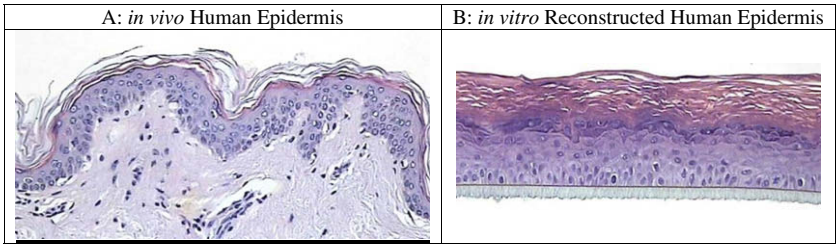

Fig. 1. Comparison between *in vivo* human epidermis (A) and the *in vitro* reconstructed human epidermis model (B) (SkinEthic Laboratories). Histological cross section (hematoxylin/eosin staining, magnification 20×).

Table 1
Current reference test substances of the performance standards (ECVAM SIVS, May 2007) and their respective characteristics.

Test substance number	Test substance name	CAS Number	Purity (%)	Substance type	Supplier	S/L	*In vivo* score[a]	EU *in vivo* Class.	GHS *in vivo* Class.[b]
1	1-Bromo-4-chlorobutane	6940-78-9	99.0	Brominated derivative	Aldrich	L	0	No	No Cat.
2	Diethyl phthalate	84-66-2	99,5	Ester	Aldrich	L	0	No	No Cat.
3	Di-propylene glycol	25265-71-8	99	Alcohol	Aldrich	L	0	No	No Cat.
4	Naphthalene acetic acid	86-87-3	97	Aromatic compound	Aldrich	S	0	No	No Cat.
5	Allyl phenoxy-acetate	7493-74-5	99	Ester	Aldrich	L	0.3	No	No Cat.
6	Isopropanol	67-63-0	99,5	Alcohol	Aldrich	L	0.3	No	No Cat.
7	4-Methyl-thio-benzaldehyde	3446-89-7	95	Aldehyde	Aldrich	L	1	No	No Cat.
8	Methyl stearate	112-61-8	99	Ester	Aldrich	S	1	No	No Cat.
9	Allyl heptanoate	142-19-8	>97	Ester	Aldrich	L	1.7	No	Cat 3
10	Heptyl butyrate	5870-93-9	98	Ester	Aldrich	L	1.7	No	Cat 3
11	Hexyl salicylate	6259-76-3	>99	Ester	Aldrich	L	2	R38	Cat 3
12	Terpinyl acetate	80-26-2	90	Ester	Alpha Aesar	L	2	R38	Cat 3
13	Tri-isobutyl phosphate	126-71-6	>99	Ester	Aldrich	L	2	R38	Cat 3
14	1-Decanol	112-30-1	99	Alcohol	Aldrich	L	2.3	R38	Cat 2
15	Cyclamen aldehyde	103-95-7	90	Aldehyde	Aldrich	L	2.3	R38	Cat 2
16	1-Bromohexane	111-25-1	98	Brominated derivative	Aldrich	L	2.7	R38	Cat 2
17	α-Terpineol	98-55-5	95	Alcohol	Alpha Aesar	L	2.7	R38	Cat 2
18	Di-*n*-propyl disulphide	629-19-6	99	*S*-containing compound	Alpha Aesar	L	3	R38	Cat 2
19	Butyl methacrylate	97-88-1	99	Ester	Aldrich	L	3	R38	Cat 2
20	Heptanal	111-71-7	95.0	Aldehyde	Aldrich	L	3.4	R38	Cat 2

[a] *In vivo* score in accordance with the OECD Test Guideline 404.
[b] Substances with an *in vivo* score <2.3 (=non-irritants under GHS). Double black line: threshold under EU and GHS classifications (Spielmann et al., 2007).

solution (PBS without Ca^{++} and Mg^{++}) served as negative control, in each series of experiments. The negative control data meet the acceptance criteria if the mean optical density (OD) value of the three tissues is $\geqslant$1.2 at 570 nm according to the historical database. The standard deviation (SD) value is considered as valid if it is $\leqslant$18%, according to the performance standards (PS: ECVAM SIVS, 2007). The positive control data meet the acceptance criteria if the mean viability, expressed as % of the NC, is <40% and the standard deviation value is $\leqslant$18%. Therefore, the run is only valid if both positive and negative controls meet their respective acceptance criteria. According to the PS document, the SD obtained from three concurrently tested tissues and in three independent runs is $\leqslant$18% for a test substance.

2.3. 42 min exposure + 42 h post-incubation protocol ("42 bis")

RHE tissues were topically exposed to undiluted liquids (16 ± 0.5 μL i.e. 32 μL/cm^2) or solids (16 ± 2 mg i.e. 32 mg/cm^2) for 42 min at room temperature. Prior applying solids, 10 ± 0.5 μL (20 μL/cm^2) of distilled water was spread on the whole surface of RHE tissue. A nylon mesh (7.5 mm diameter, Sefar Fyltis, France) was applied onto the test substance as a spreading support for all liquid and viscous test substances. RHE tissues were then rinsed 25 times with 1 mL each time of sterile PBS without Ca^{++} and Mg^{++} using a multistep pipette. Treated tissues were incubated for 42 h at 37 °C, 5% CO_2, with 2 mL growth medium. At the end of the incubation, in addition to the MTT reduction test on the tissues samples, the test media underneath were collected to determine the cytokine IL-1α release. Each experiment was performed at least in triplicate on one tissue production batch.

2.4. Cell viability measurement by MTT reduction

The cytotoxicity was determined by measuring the dehydrogenase activity of viable keratinocytes following a 42 h post-incubation. The activity was determined after incorporation of 3-[4,5-dimethyl-thiazol-2-yl]-2,5-diphenyl tetrazolium bromide (MTT, Sigma–Aldrich) (Mosmann, 1983; Korting et al., 1994). RHE tissues were placed in 1 mg/mL MTT solution for 3 h. The blue formazan crystals were then extracted from the tissue using isopropanol for at least 2 h. The concentration of formazan was measured by determining the optical density (OD) at 570 nm by spectrophotometer.

After subtracting the blank OD from all raw data, mean OD values ± standard deviations (SD) were calculated using nine measurements per test substance (three RHE tissues, three replicates/tissue) and the percentage of cell viability was expressed relative to negative control as: 100 × mean $OD_{treated}$/mean $OD_{control.}$ The negative control value was set at 100%. For each test substance, calculations of the OD values were performed for each of the three replicate tissues in each laboratory.

The ability of the test substance to directly reduce MTT was also assessed. For MTT-interacting test substances a direct interaction between test substance and MTT reagent was assumed and quantitatively measured on frozen-killed RHE tissues. The percentage of MTT reduction of a test substance was expressed relatively to OD value of the negative untreated killed tissues and considered as incompatible with the test when the MTT-reducing test substance value was above 30%.

2.5. IL-1α release

The inflammatory response was evaluated by the quantification of the specific human cytokine, interleukin-1 alpha (IL-1α, Dinarello, 1991; Corsini et al., 1996; Doucet et al., 1996; Coquette et al., 1999) using a commercial enzyme linked immuno-absorbent assay (ELISA) kit (Quantikine® Human IL-1α/IL-1F1 Immunoassay, from R&D systems UK). The analyses were performed according to manufacturer's instructions. The detection limit of the assay was 1 pg/mL. Mean concentrations (pg/mL) ± standard deviations (SD) were calculated (one replicate per tissue and three tissues per test

substance in each laboratory). IL-1α released was expressed either as absolute IL-1α data, negative control (NC) corrected data, or relative to untreated controls as: mean concentration$_{treated}$/mean concentration$_{control}$.

2.6. Prediction model

Based on EpiSkin™ validated reference test method, 39 commercially available test substances were tested in the laboratory of the skin model producer, SkinEthic Laboratory prior to the catch-up study. The prediction model (PM1) was based on the relative test substance tissue viability compared to the negative control-treated viability (100%). The test substance was considered to be a skin irritant, in accordance to European classification system (ECS), if the tissue viability was less than or equal (⩽) to 50%.

In order to increase the sensitivity of the test method, three additional prediction models (PM 2–4) combining MTT and IL-1α endpoints were proposed before the catch-up study based on the absolute data of IL-1α; negative control (NC) corrected data, and fold increase, respectively (Table 2).

2.7. Statistical analysis

Data submission templates (MTT and IL-1α spreadsheets) were password-protected before passing them to the three testing facilities. An independent statistical analysis was performed by Effi-Stat (Paris, France) to assess intra and inter-laboratory reproducibility. The main analyses undertaken on the results were related to the determination of means, standards deviation (SD), and coefficient of variation (CV). In addition, statistical analyses of variance (1-way ANOVA, significance difference level of 1%), and the correlation according to Bravais–Pearson (comparison of pairs of runs) were completed. Contingency table statistics were used for evaluating the reproducibility and predictive ability of the RHE method.

3. Results

3.1. Transfer phase

Having never used the SkinEthic's RHE test method, Oroxcell Laboratory was considered as "naïve" since, though trained in conduct of the RHE test method, they had not used the assay outside this training context. Coty Laboratory routinely used the RHE model but was not familiar with this new test method. Training on the conduct of the SkinEthic™ RHE test method and use of the laboratory documentation was provided by SkinEthic's Laboratories to the three participating laboratories. Practical and technical transfer of the RHE test method to the laboratories was feasible within a short period of training (no more than 1–3 days training). The spectrophotometer conditions are equipment dependent and were set up accordingly in each laboratory. This assay is a straightforward assay to be conducted. When comparing the data obtained at L'Oréal and those obtained at Oroxcell (naïve laboratory), the results showed the ease of the transferability of this method (narrow inter-laboratory variability not shown data).

3.2. Quality control of RHE tissues

According to SkinEthic's quality control criteria, functionality model conditions (viability, barrier function, and histology) were evaluated on each RHE batch. A RHE tissue batch was considered as conform if the following criteria were met: the mean OD values of the negative control obtained at day 17 were above 0.7 (viability MTT test), the ET-50$_{1\%TritonX\text{-}100}$ values were ranged from 4 to 9 h, and normal histology (at least four viable cell layers and a stratum corneum) was observed.

In this study, the acceptance criteria for both viability and barrier functions were met with a mean OD of 1.38 ± 0.14 (CV 10.33) and a mean ET-50_value of 5.47 ± 0.36 (CV 6.61), respectively (Table 3). An absence of significant histological alteration was observed.

3.3. Test method acceptance criteria

PBS negative control (NC) and 5% SDS positive control (PC) were tested in parallel to the test substances to demonstrate that viability, and the resulting tissue sensitivity of the SkinEthic™ RHE model were within the defined acceptance range. Similar viability measurements (e.g. expressed as absolute OD of the vital dye based on the tissue replicates of the NC) throughout all runs in the three laboratories were observed for both negative and positive controls (Fig. 2). The OD values of the negative control, PBS (2.22 ± 0.20 OD units) were within the range of the measured historical values and met the acceptance criteria of OD > 0.7. A marked decrease of cell

Table 2
Four prediction models based on both the MTT and IL-1α endpoints.

Prediction model (PM)	Criteria for *in vitro* interpretation		IL-1α analyses criteria	Classification system
	MTT endpoint	IL-1α endpoint		
For all PMs	Mean tissue viability is ⩽50%	No impact on classification	No impact on the classification	Irritant (I)
PM1	Mean tissue viability is >50%	Not applicable		Non-irritant (NI)
PM 2		⩾60 pg/mL	Absolute data: mean $_{[IL\text{-}1\alpha\ treated\ tissues]}$	I
		<60 pg/mL		NI
PM 3		⩾50 pg/mL	NC corrected data: mean $_{[IL\text{-}1\alpha\ treated\ tissues]}$/	I
		<50 pg/mL	mean $_{[IL\text{-}1\alpha\ negative\ control\ tissues]}$	NI
PM 4		⩾5	Fold increase: mean $_{[IL\text{-}1\alpha\ treated\ tissues]}$/	I
		<5	mean $_{[IL\text{-}1\alpha\ negative\ control\ tissues]}$	NI

Table 3
Quality control data corresponding to SkinEthic RHE™ tissues used by partners for experiments 1 to 5.

Tissues characteristics	WLV – BLV studies				
	Run 1	Run 2	Run 3	Run 4	Run 5
Batch	PK2 HRS 14′	PK2 HRS 15	PK2 HRS 16	PK2 HRS 17	PK2 DL 19
Lot number	07 022A 110	07 022A 1110	07 022A 1204	07 022A 1208	08 022A 0603
OD	1.266 ± 0.043	1.506 ± 0.053	1.348 ± 0.085	1.235 ± 0.083	1.555
ET_{50} (h)	5.72	5.64	4.90	5.32	5.76

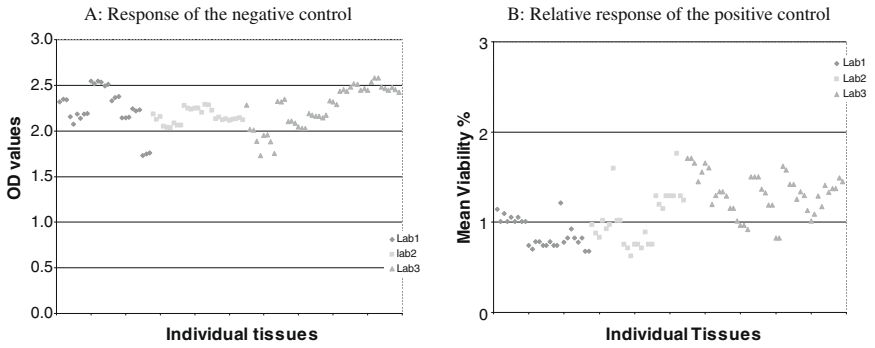

Fig. 2. A comparison of the response of the negative (PBS treated tissues) and positive (5% SDS) controls was performed. Each laboratory provided at least three independent runs (three tissues per run) for each control. The responses of both the negative (A) and the positive (B) controls fell in the same range through the three laboratories. Both OD and cell viability values threshold acceptance limits of >0.7% and <40% were met for negative and positive controls, respectively.

viability was observed with the positive control SDS 5% ranging from 0.62% to 1.76% (1.12% ± 0.30). The standard deviation (SD) associated with the mean of the tissue replicates of a substance may also indicate defects in single tissues or inappropriate dosing. In this study, 204 of 206 SD obtained within the three laboratories were between 0.01 and 12.83 range, and significantly below the acceptance cut-off value of 18% (Table 4). Taking into account the good reproducibility of all applied measures for both positive and negative controls, the intra- and inter-laboratory variability of the twenty test substances was evaluated.

3.4. Intra and inter-laboratory variability

The main analyses undertaken on the results for the twenty substances, were related to the determinations of means cell viability, and standard deviations (SD) for the three valid runs per test substance (Table 5). None of those test substances showed a SD > 18%, demonstrating the reproducibility of the SkinEthic™ RHE test method for both laboratories 1 and 2. For laboratory 3, one test substance (number 2, diethyl phthalate) showed a mean cell viability of 66.99% with a SD > 18% when considered the third run (invalid run). Focusing on the three valid runs (run 1, 2, and 5), the diethyl phthalate showed a SD of about 8%. Only one out of 20 test substances (number 5, allyl phenoxy-acetate) showed a SD > 18% as unacceptable, in one laboratory. The standard deviation means value of 18 was also applied as an inter-laboratory (9 valid runs) variability criterion. None of the 20 test substances had a standard deviation larger than 18%, indicating a high reproducibility between the three laboratories (Table 5).

Analyses of variance (1-way ANOVA, significance level of 1%), coefficient of variation (CV), and Bravais–Pearson's correlation were also performed (data not shown). Briefly, the Bravais–Pearson correlation, which calculates the correlation of the mean cell viability for the pairs of runs for the 20 test substances, was above 0.979 in the three laboratories. Limited ANOVA/*t*-test *p*-values were below 0.01 and no ANOVA sum of squares was above 2605 for all test substances. The proportion of identically classified test substances derived from the prediction model was 100% for both laboratories 1 and 2 and 95% for laboratory 3, when considering alls runs.

Regardless the analyses, low intra and inter-runs variability for all laboratories was observed with the negative and positive controls, and the 20 reference test substances indicating a significant

Table 4
Standard deviation of the variability of all runs in the three laboratories of the 20 substances in addition to negative (NC) and positive (PC) controls. Runs, which were not considered in the three valid runs, are marked grey (dark grey = non-valid runs (SD > 18 for run 3 and only 2 accepted tissues for run 4), light grey = 'extra' runs). nd: not done.

Test substance number	Laboratory 1			Laboratory 2			Laboratory 3				
	Run 1	Run 2	Run 3	Run 1	Run 2	Run 3	Run 1	Run 2	Run 3	Run 4	Run 5
NC	0.10	0.09	0.22	0.05	0.03	0.01	0.16	0.13	0.07	0.03	0.06
PC	0.05	0.03	0.16	0.22	0.07	0.18	0.19	0.17	0.27	0.20	0.13
1	0.08	0.06	1.91	2.65	0.25	0.38	0.07	12.83	0.29	nd	nd
2	4.34	2.10	0.84	7.22	7.69	1.35	9.72	9.21	19.75	2.17	2.68
3	4.74	3.68	9.66	6.64	0.85	11.31	3.80	7.02	12.23	nd	nd
4	3.88	5.34	1.42	1.92	0.95	7.75	3.50	2.19	4.28	nd	nd
5	7.28	0.24	3.95	4.48	4.96	4.00	4.73	4.50	4.96	5.01	nd
6	1.90	3.68	9.60	3.21	1.80	5.79	7.68	6.36	3.30	nd	nd
7	0.07	0.12	0.34	0.40	2.71	2.09	0.36	0.24	0.12	nd	nd
8	5.55	2.10	5.09	2.24	3.92	0.42	4.79	3.21	8.78	nd	nd
9	1.68	4.99	5.69	5.38	9.93	9.76	5.57	1.31	8.60	4.63	nd
10	4.13	2.41	7.04	6.41	1.30	2.14	1.55	2.42	7.37	nd	nd
11	1.57	1.53	7.60	5.17	1.38	5.34	2.43	6.66	5.24	nd	nd
12	0.14	0.56	0.01	1.20	0.41	0.13	1.26	4.82	0.34	nd	nd
13	0.11	0.13	0.08	0.20	0.28	0.22	0.12	0.33	0.30	nd	nd
14	0.09	0.05	0.04	0.05	0.23	0.24	0.18	0.17	0.21	nd	nd
15	0.10	0.09	0.07	0.38	0.36	0.18	0.07	0.14	0.15	nd	nd
16	0.01	0.31	0.07	0.04	0.10	0.23	0.39	0.25	0.16	nd	nd
17	0.44	11.14	0.02	0.30	0.24	1.14	0.25	0.27	0.35	nd	nd
18	0.13	0.03	0.56	0.04	0.10	0.18	0.19	0.18	0.14	nd	nd
19	0.97	1.46	0.07	0.18	0.17	0.30	0.33	0.23	0.22	nd	nd
20	0.08	0.04	0.05	0.30	0.65	0.21	0.13	0.16	0.14	nd	nd

Table 5
MTT cell viability (individual runs and mean) and standard deviation (SD) at L'Oréal, Coty and Oroxcell Laboratories [a: Test substance 2 (diethyl phthalate) had two non-valid runs (run 3 with a SD > 18, and run 4 with technical problems with the treated tissue number 2), and three valid runs (runs 1, 2, and 5); test substances 5 and 9 (allyl phenoxy-acetate and allyl heptanoate, respectively) were re-tested in a fourth run (blind study) but the three first runs data met the acceptance criteria. The data of run 4 were taken into account only for statistical analyses of all runs]. Light grey cells: runs, which were not considered for the three valid runs (bold = non-valid runs, not bold = 'extra runs); Grey cells: chemical with SD > 18.

Test substance number	Laboratory 1 (L'Oréal)					Laboratory 2 (Coty)					Laboratory 3 (Oroxcell)[a]							SD all labs (9 valid runs)
	Viability (%)				SD	Viability (%)				SD	Viability (%)						SD	
	Run 1	Run 2	Run 3	Mean (3 valid runs)	3 Valid runs	Run 1	Run 2	Run 3	Mean (3 valid runs)	Three valid runs	Run 1	Run 2	Run 3	Run 4	Run 5	Mean (3 valid runs)	Three valid runs	
1	1.30	0.85	2.74	1.63	0.99	2.79	1.14	1.06	1.66	0.98	2.47	10.54	1.67	–	–	4.89	4.91	1.87
2	94.03	92.52	93.17	93.24	0.76	99.70	92.16	83.82	91.90	7.94	100.46	95.01	**66.99**	90.02	84.95	95.16	7.87	0.85
3	100.39	94.38	99.76	98.18	3.30	101.32	103.73	94.93	99.99	4.54	106.10	103.25	82.99	–	–	97.45	12.60	1.31
4	98.20	91.95	103.90	98.02	5.98	102.39	96.71	100.62	99.91	2.91	95.15	95.80	95.57	–	–	95.51	0.33	2.21
5	84.82	57.95	60.88	67.88	14.74	86.23	88.38	89.86	88.15	1.83	95.00	85.45	47.28	60.24	–	75.91	**25.25**	10.21
6	92.84	88.15	101.05	94.01	6.52	105.92	101.35	105.03	104.10	2.42	91.34	88.70	93.86	–	–	91.30	2.58	6.75
7	2.91	2.84	2.65	2.80	0.13	3.03	4.41	5.15	4.20	1.07	4.63	2.94	3.92	–	–	3.83	0.84	0.72
8	103.07	101.14	109.01	104.41	4.10	112.50	105.33	106.65	108.16	3.81	105.13	101.63	91.32	–	–	99.36	7.18	4.41
9	96.70	75.43	87.27	86.47	10.66	99.56	89.65	75.17	88.13	12.26	85.27	82.43	55.42	78.35	–	74.37	16.48	7.51
10	95.95	91.49	97.53	94.99	3.13	102.61	100.60	103.62	102.28	1.53	104.89	101.49	91.55	–	–	99.31	6.93	3.66
11	94.92	89.38	96.91	93.74	3.91	104.26	89.59	90.18	94.68	8.31	106.37	104.90	86.31	–	–	99.19	11.18	2.92
12	1.84	1.50	1.63	1.66	0.17	2.06	1.36	1.54	1.66	0.37	2.83	5.05	1.65	–	–	3.18	1.73	0.88
13	1.32	1.18	1.52	1.34	0.17	1.74	1.39	1.42	1.52	0.19	1.99	1.59	1.60	–	–	1.73	0.23	0.19
14	1.41	1.24	1.63	1.43	0.19	1.89	1.25	1.42	1.52	0.33	1.99	1.54	1.33	–	–	1.62	0.34	0.10
15	1.51	1.85	1.41	1.59	0.23	1.82	1.63	1.82	1.76	0.11	1.68	1.90	1.91	–	–	1.83	0.13	0.12
16	1.11	1.01	1.05	1.06	0.05	1.34	1.08	1.36	1.26	0.16	1.65	1.69	1.47	–	–	1.60	0.12	0.28
17	1.09	7.13	0.65	2.96	3.62	1.05	0.84	2.18	1.36	0.72	1.44	1.02	1.42	–	–	1.30	0.24	0.94
18	1.50	0.95	2.04	1.50	0.55	1.33	1.03	1.25	1.20	0.16	2.62	1.69	1.91	–	–	2.08	0.49	0.44
19	2.85	4.31	0.95	2.70	1.68	1.20	0.99	1.19	1.13	0.12	2.48	1.61	1.73	–	–	1.94	0.47	0.79
20	1.17	1.09	1.11	1.12	0.04	1.44	1.59	1.58	1.54	0.09	1.52	1.07	1.31	–	–	1.30	0.22	0.21

promising intra- and inter-laboratory variability using the SkinEthic™ RHE test method.

3.5. Predictive capacity

The specificity and sensitivity over all runs per laboratory were evaluated based on the four predictive models (PM) described in Table 2. All the PMs were firstly based on the MTT threshold of 50% cell viability and then IL-1α different response levels of the test substances were eventually considered. The means of the total IL-1α amount release (pg/mL) for all runs of all tested substances, in the three laboratories were summarized in Table 6. The substances classified as irritant had a tendency to induce higher IL-1α released as measured by all laboratories. When considering the median over all laboratories, none of IL-1α release of those test substances showed a SD > 18%.

Applying the proportion of identically classified test substances measure (non-irritants versus irritants), the results showed that all test substances were identically identified in the three laboratories with three valid runs (Table 7). In a pair-wise comparison of the laboratories, the concordance of classifications was 100%.

The specificity and sensitivity over laboratory were also evaluated and presented in Table 7. Considering the three valid runs ($n = 90$), a specificity of 80% and a sensitivity of 90% was achieved with an overall accuracy of 85% when using the PM1 (MTT endpoint only cut-off of 50%). When considering the mean either of IL-1α, [IL-1α] NC corrected or IL-1α fold releases of all runs on the median over all laboratories, the PMs 2–4 performed similar results. No additional correctly classified irritants as positives were identified, and none of the nine MTT non-irritants were falsely classified as irritants. The specificity and sensitivity were about 80% and 90%, respectively, independently of the added IL-1α parameter compared to those obtained using the MTT parameter only.

4. Discussion

In this study, we described a SkinEthic™ RHE test method for determining the irritant potency of a test substance, using basically the same protocol than the validated reference EpiSkin™ test method with only minor refinements. Likewise, an exposure time period of 42 min was optimal for the present test method to achieve very good within and between-laboratory reproducibility and comparable reliability and accuracy results. Only the exposure volumes and exposure times were needed to be optimized for SkinEthic's epidermis test method while the basic elements of the method, a post-incubation period of 42 h and the prediction model were identical. Similarly, the same parameters were optimized for other models such as EpiDerm™, LabCyte EPI-MODEL™, EST-1000 or Leiden human epidermal models (Spielmann et al., 2007; El Ghalbzouri et al., 2008).

To compare the performance of the SkinEthic™ RHE test method to the validated available epidermis test methods, twenty coded test substances, which were considered to be essential for comparison by ECVAM, were applied onto the *in vitro* model according to the "42 bis" protocol. From the eleven independent runs with different RHE batches, the results obtained with the positive and negative controls showed that the test method had a high repeatability. Therefore, both controls met the acceptance criteria indicating that all runs were valid. This reproducibility was not only seen for positive and negative controls on repeated experiments, but also on measured individual endpoint for each test substance. In our experiments, the standard deviation was below 18% (except for one non-irritating test substance, the diethyl phthalate), and the Pearson correlation contained between 0.980 and 0.995. Similar variability was observed for the diethyl phthalate when using the EpiDerm™ test method (Spielmann et al., 2007). Regarding the twenty test substances, the variability of the SkinEthic™ RHE test method was lower compared to the validated test methods Episkin™ and EpiDerm™ test methods which have provided a standard deviation above 18% for 3 (4-methyl-thio-benzaldehyde, terpinyl acetate and di-n-propyl disulphide) and 5 (1-bromo-4-chlorobutane, 4-methyl-thio-benzaldehyde, α-terpineol, 1-decanol, and terpinyl acetate) out of 20 test substances, respectively. Based on the 20 test substances, the intra-laboratory concordance was 95% as the non-irritant allyl phenoxy-acetate test substance had provided a positive result amongst the 11 runs (9% discordance) in which it has been tested. Based on SkinEthic™ RHE and the endpoint MTT, the inter-laboratory variability of the classification of the test substances showed that the reproducibility was above 99% in which 163 of 164 tests had provided similar classification of the test substances. For both non-irritating and irritating test substances, the overall inter-laboratory concordance

Table 6
Mean IL1-α (pg/mL) data measurements and its variability within the three laboratories. Italic font: chemicals classified as irritant according to the mean viability of MTT over all runs per laboratory.

Test substance number	EU class	Mean IL1-α (pg/mL)				
		Lab1	Lab 2	Lab 3	Mean all labs	SD all labs
1	*No label*	*40.34*	*52.51*	*57.73*	*50.19*	*8.93*
2	No label	9.72	18.36	25.53	17.87	7.91
3	No label	12.60	13.35	16.45	14.13	2.04
4	No label	7.92	9.46	5.10	7.50	2.21
5	No label	12.72	14.20	24.84	17.25	6.61
6	No label	17.68	19.48	19.30	18.82	0.99
7	*No label*	*32.91*	*34.56*	*44.42*	*37.30*	*6.22*
8	No label	10.18	13.17	16.52	13.29	3.17
9	No label	13.58	25.39	23.04	20.67	6.25
10	No label	24.49	24.56	30.35	26.47	3.37
11	R38	10.04	10.47	11.18	10.56	0.58
12	*R38*	*91.34*	*73.34*	*73.43*	*79.37*	*10.36*
13	*R38*	*38.44*	*44.99*	*61.24*	*48.22*	*11.74*
14	*R38*	*39.92*	*33.40*	*52.36*	*41.89*	*9.64*
15	*R38*	*28.95*	*29.75*	*44.94*	*34.54*	*9.01*
16	*R38*	*51.49*	*47.72*	*59.37*	*52.86*	*5.94*
17	*R38*	*55.94*	*48.57*	*54.68*	*53.06*	*3.94*
18	*R38*	*44.52*	*42.40*	*61.19*	*49.37*	*10.29*
19	*R38*	*51.62*	*41.88*	*75.09*	*56.19*	*17.07*
20	*R38*	*37.02*	*29.19*	*48.46*	*38.22*	*9.69*

Table 7
Predictive capacity for the four PMs. For the PM1, the classification was based on mean viability over all runs per laboratory. For PM2-4, the combined classifications (MTT and IL1-α endpoints) represented in this Table were based on the median over all laboratories. NI, non-irritant (no label); I, irritant (R38).

Test substance number	Test substance name	EU class	PM1 MTT classification only			PM2 MTT + IL1-α classification (cut-off: 60 pg/mL)	PM3 MTT + IL1-α NC corrected classification (cut-off: 50 pg/mL)	PM4 MTT + IL1-α fold increase classification (cut-off: 5)
			Lab 1	Lab 2	Lab 3			
1	1-Bromo-4-chlorobutane	No label	I	I	I	I	I	I
2	Diethyl phthalate	No label	NI	NI	NI	NI	NI	NI
3	Di-propylene glycol	No label	NI	NI	NI	NI	NI	NI
4	Naphthalene acetic acid	No label	NI	NI	NI	NI	NI	NI
5	Allyl phenoxy-acetate	No label	NI	NI	NI	NI	NI	NI
6	Isopropanol	No label	NI	NI	NI	NI	NI	NI
7	4-Methyl-thio-benzaldehyde	No label	I	I	I	I	I	I
8	Methyl stearate	No label	NI	NI	NI	NI	NI	NI
9	Allyl heptanoate	No label	NI	NI	NI	NI	NI	NI
10	Heptyl butyrate	No label	NI	NI	NI	NI	NI	NI
11	Hexyl salicylate	R38	NI	NI	NI	NI	NI	NI
12	Terpinyl acetate	R38	I	I	I	I	I	I
13	Tri-isobutyl phosphate	R38	I	I	I	I	I	I
14	1-Decanol	R38	I	I	I	I	I	I
15	Cyclamen aldehyde	R38	I	I	I	I	I	I
16	1-Bromohexane	R38	I	I	I	I	I	I
17	α-Terpineol	R38	I	I	I	I	I	I
18	Di-*n*-propyl disulphide	R38	I	I	I	I	I	I
19	Butyl methacrylate	R38	I	I	I	I	I	I
20	Heptanal	R38	I	I	I	I	I	I

was 100%. The concordance between laboratories for the SkinEthic™ RHE was comparable to the reference validated EpiSkin™ test method since only the di-*n*-propyl disulphide and terpinyl acetate were classified differently amongst the three laboratories. The overall inter-laboratory concordance was 90% and 75% for both validated EpiSkin™ and EpiDerm™ test method, respectively (Spielmann et al., 2007). The high reproducibility of the results obtained with the SkinEthic™ RHE test method indicated its ease of use since this test method was successfully transferred to laboratories that had never been using the test method protocol before.

The catch-up validation study was in accordance with the defined validation process which includes the establishment of a prediction model. Interestingly, the results obtained in the three laboratories with an overall accuracy of 85% met the specificity (>80%), sensitivity (>70%) defined by ECVAM (ECVAM SIVS, 2007). The three same test substances were misclassified (1-bromo-4-chlorobutane, 4-methyl-thio-benzaldehyde and hexyl salicylate) with other epidermis test methods (Spielmann et al., 2007). Therefore, no clear difference in the physicochemical properties between the correctly and incorrectly classified test substances was identified (Zuang et al., 2007). In this manuscript, four prediction models were proposed with two endpoints (MTT and IL-1α). The overall predictivity of the combination of both endpoints was performed with the mean IL-1α amount, the absolute difference and the 5-fold increase. In this study, the detected differences for the test substances among the laboratories were not being statistically significant for the IL-1α. However, compared to the predictive capacity of the MTT alone (specificity: 80%; sensibility: 90%), no increase in sensitivity and specificity were achieved when using the three additional proposed combined prediction models. Thus, the IL-1α does not seem to offer any advantages, compared to the MTT only for the classification of those twenty given test substances using the SkinEthic™ RHE test method. For comparison with the EpiSkin™ test method, when the combination of MTT and release IL-1α was used, a similar overall performance (specificity: 80%; sensibility: 90%) was obtained using these 20 reference test substances (Cotovio et al., 2005; ECVAM statement, May 2007). The IL-1α endpoint, tested with the same 20 chemicals for the optimized EpiDerm™ test method, gave no contribution to the performance of the test method (ECVAM statement, November 2008). The outcome of SIVS and update validation studies stated that IL-1α endpoint was regarded as a useful adjunct to the MTT assay, as it has the potential to increase the sensitivity of the test, without reducing its specificity (Spielmann et al., 2007; ECVAM statement, November 2008). Interestingly, in a group of 10 irritants, two test substances (2-methyl-4-phenyl-2-butanol and 4,4-methylenebis-(2,6-di-tert-butyl)phenol) were classified as irritant using the IL-1α endpoint with the Leiden reconstructed human epidermal (LHE) test method compared to the MTT data only (El Ghalbzouri et al., 2008). Therefore, inclusion of the interleukin-1 alpha (IL-1α) release as an endpoint could be considered to confirm the test substances classified as negative when using the MTT endpoint and might

be of help to better classify the mild to moderate irritancy potential of some test substances.

The present catch-up study wad based on the ECVAM Performance Standard (PS) document defined in 2007 following the ECVAM Skin Irritation Validation Study (ECVAM SIVS, May 2007). Since, the SIVS was conducted prior to the adoption of the United Nations Globally Harmonized System for the Classification and Labeling of Substances (GHS, United Nations Economic Commission, 2003), the PS was based with respect to the two categories European Classification System (ECS) (Hoffman et al., 2005). Of the 20 test substances, 10 were irritants (labeled R38) and 10 were non-irritants (not labeled) with an *in vivo* cut-off score of 2.0. With respect to the GHS classification system, with a cut-off score of 2.3, the unbalanced distribution of the GHS categories (13 non-irritant versus 7 irritants) did not allow a representative determination of the predictive capacity of the twenty test substances. Consequently, our analysis of skin irritation presented here showed high prevalence of the threshold ECS dominant median 2.0 and further investigations directed towards evaluating of GHS classification might be evaluated on test substances selected for this purpose with a higher number of substances covering the range if irritancy. As described in the introduction for an evaluation of local acute skin effects, it is recommended to follow the sequential testing strategy as appended to OECD Test Guideline 404 (OECD 404, 2002), in which prior to progressing to an *in vitro* test for skin irritation test, the OECD testing strategy suggests initial consideration for skin corrosivity. The tiered approach for testing and evaluating both skin corrosion and irritation potential might incorporates the use of existing validated *in vitro* methodologies such as SkinEthic™ RHE test methods. This testing strategy includes the conduct of *in vitro* validated test method for skin corrosion (Kandárová et al., 2006b; Tornier et al., 2006; ECVAM statement, November 2006) and skin irritation here. To highlight this strategy, the EU test method B.46 on reconstructed human epidermis models (EpiSkin™, SkinEthic™ RHE and EpiDerm™) for skin irritation, was included into the EU Test Method Regulation (COM regulation 440/2008/EC).

To conclude, a prediction model was defined to classify a test substance as irritant when the mean viability was below 50%. As outlined in the present work, the intra- and inter-laboratory evaluations demonstrated high reliability and relevance of the SkinEthic™ RHE "42 bis" test method. Therefore, the catch-up validation study on the *in vitro* test method for skin irritation was endorsed by the ECVAM Scientific Advisory Committee in November 5th, 2008. This test method might be defined as a stand alone replacement test method for the Draize rabbit *in vivo* test, as a screen, or as part of a sequential testing strategy in a weight of evidence approach, for classifying non-irritant and irritant test substances, depending on country requirements.

Acknowledgments

The authors would like to thank the partners of the project, J.R. Meunier and O. De Silva for fruitful discussions and special advice and support. We are very grateful to M. Bonnet, B. De Servi, and I. Goulet for their laboratory technical support.

The excellent assistance on statistical analysis by J.C. Lemarié, and S. Gicquel from Effi-stat is gratefully acknowledged. Finally, the authors thank B. Dewilde and C. Capallere for high quality tissue manufacturing unless such results would be impossible.

References

Bell, E., Ehrlich, H.P., Buttle, D.J., Nakatsuji, T., 1981. Living tissue formed in vitro and accepted as skin-equivalent tissue of full thickness. Science 211, 1051–1054.

Boelsma, E., Gibbs, S., Faller, C., Ponec, M., 2000. Characterization and comparison of reconstructed skin models: morphological and immunohistochemical evaluation. Acta Dermato-Venereologica 80, 82–88.

Cannon, C.L., Neal, P.J., Southee, J.A., Kubilus, J., Klausner, M., 1994. New epidermal model for dermal irritancy testing. Toxicology In Vitro 8, 889–891.

Coquette, A., Berna, N., Vandenbosch, A., Rosdy, M., Poumay, Y., 1999. Differential expression and release of cytokines by an in vitro reconstituted human epidermis model following skin irritant and sensitizing compounds. Toxicology In Vitro 13, 867–877.

Corsini, E., Bruccoleri, A., Marinovich, M., Galli, C.L., 1996. Endogenous interleukin-1 alpha associated with skin irritation induced by tributylin. Toxicology and Applied Pharmacology 138, 268–274.

Cotovio, J., Grandidier, M.-H., Portes, P., Roguet, R., Rubinstenn, G., 2005. The *In Vitro* acute skin irritation of chemicals: optimisation of the EPISKIN prediction model within the framework of the ECVAM validation process. ATLA 33, 329–349.

Cumpstone, M.B., Kennedy, A.H., Harmon, C.S., Phil, D., Potts, R.O., 1989. The water permeability of primary mouse keratinocyte cultures grown at the air–liquid interface. Journal of Investigative Dermatology 92, 598–600.

de Brugerolle de Fraissinette, A., Picarles, V., Chibout, S., Kolopp, M., Medina, J., Burtin, P., Ebelin, M.E., Osborne, S., Mayer, F.K., Spake, A., Rosdy, M., De Wever, B., Ettlin, R.A., Cordier, A., 1999. Predictivity of an in vitro model for acute and chronic skin irritation (SkinEthic®) applied to the testing of topical vehicles. Cell Biology and Toxicology 15, 121–135.

De Wever, B., Charbonnier, V., 2002. Using tissue engineered skin to evaluate the irritation potential of skin care products. Cosmetics and Toiletries 117, 28–36.

Dinarello, C.A., 1991. Interleukin 1 and interleukin 1 antagonism. Blood 77, 1627–1652.

Doucet, O., Robert, C., Zastrow, L., 1996. Use of a serum-free reconstituted epidermis as a skin pharmacological model. Toxicology In Vitro 10, 305–313.

Draize, J.H., Woodand, G., Calvery, H.O., 1944. Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. Journal of Pharmacology and Experimental Therapeutics 82, 337–390.

ECVAM Scientific Advisory Committee (ESAC) Statement, 2006. Statement on the SkinEthic™ Application of the Skin Human Skin Model for Skin Corrosivity Testing. ESAC 25th Meeting, 16–17 November 2006. Available at: <http://ecvam.jrc.it/index.htm>.

ECVAM Scientific Advisory Committee (ESAC) Statement, 2007. Statement on the validity of *in-vitro* Tests for Skin Irritation. ESAC 26th Meeting, 26–27th April 2007. Available at: <http://ecvam.jrc.it/index.htm>.

ECVAM SIVS, May 2007. Performance Standards for Applying Human Skin Models to *in vitro* Skin Irritation Testing. Available at: <http://ecvam.jrc.it/index.htm>.

ECVAM Scientific Advisory Committee (ESAC) Statement, 2008. Statement on the Scientific Validity of in-vitro Tests for Skin Irritation Testing. ESAC 29th Meeting, 4–5th November 2008. Available at: <http://ecvam.jrc.it/index.htm>.

El Ghalbzouri, A., Siamari, R., Willemze, R., Ponec, M., 2008. Leiden reconstructed human epidermal model as a tool for the evaluation of the skin corrosion and irritation potential according to the ECVAM guidelines. Toxicology In Vitro 22, 1311–1320.

Eskes, C., Cole, T., Hoffmann, S., Worth, A., Cockshott, A., Gerner, I., Zuang, V., 2007. The ECVAM international validation study on *in vitro* tests for acute skin irritation: selection of test chemicals. Alternatives to Animal Testing 35, 603–619.

European Commission, 1976. Council Directive 76/768/EEC of 27 July 1976 on the Approximation of the Laws of the Member States Relating to Cosmetic Products. Official Journal of the European Communities L262, 169.

European Commission, 2001. Annex VI of the Directive 67/548/ECC. General Classification and Labeling Requirements for Dangerous Substances and Preparations. Official Journal of the European Communities L225, 263–314 (Method B.4 of Annex V to Council Directive 67/548/EEC).

European Commission, 2009. Test Method B.46 on Reconstructed Human Epidermis Models (EPISKIN, SKINETHIC, and EPIDERM) for Skin Irritation. EU Test Method Regulation (COM Regulation 440/2008/EC) through the 1st Adaptation to Technical Progress ("ATP").

Fartasch, M., Ponec, M., 1994. Improved barrier structure formation in air-exposed human keratinocyte culture systems. Journal of Investigative Dermatology 102, 366–374.

Fentem, J.H., Botham, P.A., 2002. ECVAM's activities in validating alternative tests for skin corrosion and irritation. ATLA 30, 61–67.

Hoffmann, J., Heisler, E., Karpinski, S., Losse, J., Thomas, D., Siefken, W., Ahr, H.J., Vohr, H.W., Fuchs, H.W., 2005a. Epidermal-skin-test (EST-1000) – a new reconstructed epidermis for in vitro skin corrosivity testing. Toxicology In Vitro 19, 925–929.

Hoffmann, S., Cole, T., Hartung, T., 2005b. Skin irritation: prevalence, variability, and regulatory classification of existing in vivo data from industrial chemicals. Regulatory Toxicology and Pharmacology 41, 159–166.

Hoffmann, S., 2007. ECVAM Skin Irritation Validation Study Phase II: Analysis of the Primary Endpoint MTT and the Secondary Endpoint IL-1α. Available at: <http://ecvam.jrc.it/index.htm>.

Kandárová, H., 2006. Dissertation: Evaluation and Validation of Reconstructed Human Skin Models as Alternatives to Animal Tests in regulatory Toxicology. Free University of Berlin.

Kandárová, H., Liebsch, M., Spielmann, H., Genshow, E., Schmidt, E., Traue, D., Guest, R., Whittingham, A., Warren, N., Gamer, AO., Remmele, M., Kaufmann, T., Wittmer, E., De Wever, B., Rosdy, M., 2006a. Assessment of the human epidermis model SkinEthic RHE for in vitro skin corrosion testing of chemical according to new OECD TG431. Toxicology In Vitro 20, 547–559.

Kandárová, H., Liebsch, M., Schmidt, E., Genshow, E., Traue, D., Meyer, K., Spielmann, H., Steinoff, C., Tornier, C., De Wever, B., Rosdy, M., 2006b. Assessment of the skin irritation potential of chemicals by using the SkinEthic reconstructed human epidermal model and the common skin irritation protocol evaluated in the ECVAM skin irritation study. Alternatives to Laboratory Animals 34, 393–406.

Korting, H.C., Schindler, S., Hartinger, A., Kerscher, M., Angerpointner, T., Maibach, H.I., 1994. MTT-assay and neutral red release (NRR)-assay: relative role in the prediction of the irritancy potential of surfactants. Life Sciences 55, 533–540.

Mosmann, T., 1983. Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. Journal of Immunology Methods 65, 55–63.

Niwa, M., Nagai, K., Oike, H., Kobori, M., 2009. Evaluation of the skin irritation using a DNA microarray on a reconstructed human epidermal model. Biological and Pharmaceutical Bulletin 32, 203–208.

OECD, 2002. Test Guideline 404. OECD Guideline for the Testing of Chemicals. Acute Dermal Irritation/Corrosion. Updated Guideline, Adopted 24 April 2002. Available at: <http://www.oecd.org/document/22/0,2340,en_2649_34377_1916054_1_1_1_1,00.html>.

Ponec, M., Weerheim, A., Kempenaar, J., Mommaas, A.-M., Nugteren, D.H., 1988. Lipid composition of cultured keratinocytes in relation to their differentiation. Journal of Lipid Research 29, 949–961.

Poumay, Y., Dupont, F., Marcoux, S., Leclerq-Smekens, M., Hérin, M., Coquette, A., 2004. A simple reconstructed human epidermis: preparation of the culture model and utilization in in vitro studies. Archives of Dermatological Research 296, 203–211.

Rosdy, M., Clauss, L.C., 1990. Terminal epidermal differentiation of human keratinocytes grown in chemically defined medium on inert filter substrates at the air–liquid interface. Journal of Investigative Dermatology 95, 409–414.

Rosdy, M., Pisani, A., Ortonne, J.-P., 1993. Production of basement membrane components by a reconstructed epidermis cultured in the absence of serum and dermal factors. Journal of Investigative Dermatology 129, 227–234.

Spielmann, H., Hoffmann, S., Liebsch, M., Botham, P., Fentem, J., Eskes, C., Roguet, R., Cotović, J., Cole, T., Worth, A., Heylings, J., Jones, P., Robles, C., Kandárová, H., Gamer, A., Remmele, M., Curren, R., Raabe, H., Cockshott, A., Gerner, I., Zuang, V., 2007. The ECVAM international validation study on in vitro tests for acute skin irritation: report on the validity of the EPISKIN and EpiDerm assays and on the skin integrity function test. ATLA 35, 559–601.

Tinois, E., Tiollier, J., Gaucherand, M., Dumas, H., Tardy, M., Thivolet, J., 1991. In vitro and post-transplantation differentiation of human keratinocytes grown on the human type IV collagen film of a bilayered dermal substitute. Experimental Cell Research 193, 310–319.

Tornier, C., Rosdy, M., Maibach, H.I., 2006. In vitro skin irritation testing on reconstituted human epidermis: reproducibility for 50 chemicals tested with two protocols. Toxicology In Vitro 20, 401–416.

United Nations Economic Commission for Europe, 2003. The Globally Harmonized System of Classification and Labelling of Chemicals (GHS).UN ST/SG/AC.10/30, United Nations, New York, NY, USA/Geneva, Switzerland: United NationsOrganisation. <http://www.unece.org/trans/danger/publi/ghs> (accessed 15.08.03).

Zuang, V., Eskes, C., Worth, A., Cole, T., Hoffman, S., Gallegos Saliner, A., Netzeva, T., Patlewicz, G., Cockshott, A., Gerner, I., 2007. Report from the Chemicals Selection Sub-Committee to the Management Team on Potential Reasons for the Misclassification of Chemicals in the EPISKIN® and EpiDerm® Assays. Available at <http://ecvam.jrc.it/index.htm>.