Toxicology in Vitro 36 (2016) 254–261

Contents lists available at ScienceDirect

Toxicology in Vitro

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

Catch-up validation study of an in vitro skin irritation test method based on an open source reconstructed epidermis (phase II)

F. Groeber [a], L. Schober [b], F.F. Schmid [a], A. Traube [b], S. Kolbus-Hernandez [a], K. Daton [c], S. Hoffmann [d], D. Petersohn [c], M. Schäfer-Korting [e], H. Walles [a,f], K.R. Mewes [c,*]

[a] Fraunhofer Institute of Interfacial Engineering and Biotechnology, Stuttgart, Germany
[b] Fraunhofer Institute for Manufacturing Engineering and Automation, Stuttgart, Germany
[c] Henkel AG & Co. KGaA, Düsseldorf, Germany
[d] Seh consulting + services, Paderborn, Germany
[e] Institute of Pharmacy (Pharmacology and Toxicology), Freie Universität Berlin, Berlin, Germany
[f] Chair of Tissue Engineering and Regenerative Medicine, University Hospital Wuerzburg, Wuerzburg, Germany

ARTICLE INFO

Article history:
Received 15 January 2016
Received in revised form 30 June 2016
Accepted 13 July 2016
Available online 17 July 2016

Keywords:
In vitro skin irritation testing
Open source reconstructed epidermis
Validation
Alternative test methods

ABSTRACT

To replace the Draize skin irritation assay (OECD guideline 404) several test methods based on reconstructed human epidermis (RHE) have been developed and were adopted in the OECD test guideline 439. However, all validated test methods in the guideline are linked to RHE provided by only three companies. Thus, the availability of these test models is dependent on the commercial interest of the producer. To overcome this limitation and thus to increase the accessibility of in vitro skin irritation testing, an open source reconstructed epidermis (OS-REp) was introduced. To demonstrate the capacity of the OS-REp in regulatory risk assessment, a catch-up validation study was performed. The participating laboratories used in-house generated OS-REp to assess the set of 20 reference substances according to the performance standards amending the OECD test guideline 439. Testing was performed under blinded conditions. The within-laboratory reproducibility of 87% and the inter-laboratory reproducibility of 85% prove a high reliability of irritancy testing using the OS-REp protocol. In addition, the prediction capacity was with an accuracy of 80% comparable to previous published RHE based test protocols. Taken together the results indicate that the OS-REp test method can be used as a standalone alternative skin irritation test replacing the OECD test guideline 404.

1. Introduction

The epidermis forms the outermost barrier between the human body and the environment and is thus exposed to various potentially harmful substances. To test the potential of substances to cause skin irritation, the Draize assay was developed and implemented as test guideline 404 of the Organization for Economic Co-operation and Development (OECD) (Draize et al., 1944). However, due to the limited relevance of results from this animal assay for humans and animal welfare concerns, in vitro models have been developed to assess the skin irritation potential of chemicals (Fentem et al., 1998; Stobbe et al., 2003).The legislation worldwide endorses the introduction of alternative test methods that comply with the concept of reduction, refinement and replacement of animal experimentation introduced by Russell and Burch (Russell et al., 1959). In the European Union, clear priority is given for animal free tests due to the provisions of the legislation for the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH; EU, 2006) and the European Union Cosmetics Regulation (EU, 2009). The latter regulation put a stepwise ban of animal tests for cosmetic products and ingredients into force. The final step was enforced in

Abbreviations: EURL-ECVAM, European Union Reference Laboratory for Alternatives to Animal Testing; FHG, Fraunhofer Gesellschaft; FUB, Institute of Pharmacy at the Freie Universität Berlin; UN GHS, United Nations Globally Harmonized System of Classification and Labelling of Chemicals; hEK, Human epidermal keratinocytes; HEN, Henkel AG &Co. KGaA; MTT, 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazoliumbromid; OECD, Organization for Economic Co-operation and Development; OS-REp, Open source reconstructed epidermis; PBS, Phosphate buffered saline; RHE, Reconstructed human epidermis; SDS, Sodium dodecyl sulfate; SOP, Standard operating procedure.

* Corresponding author at: Henkel AG & Co. KGaA, Biological & Clinical Research, Henkelstr 67, 40589 Düsseldorf, Germany.
E-mail addresses: florian.groeber@igb.fraunhofer.de (F. Groeber), lena.schober@ipa.fraunhofer.de (L. Schober), freia.schmid@igb.fraunhofer.de (F.F. Schmid), andrea.traube@ipa.fraunhofer.de (A. Traube), Silvia.Kolbus-Hernandez@igb.fraunhofer.de (S. Kolbus-Hernandez), karolinadaton@gmx.de (K. Daton), sebastian.hoffmann@seh-cs.com (S. Hoffmann), dirk.petersohn@henkel.com (D. Petersohn), Monika.Schaefer-Korting@fu-berlin.de (M. Schäfer-Korting), heike.walles@igb.fraunhofer.de (H. Walles), karsten.mewes@henkel.com (K.R. Mewes).

http://dx.doi.org/10.1016/j.tiv.2016.07.008

March 2013 and bans the marketing of cosmetic products containing ingredients tested on animals after this deadline.

The prediction of skin irritation using an alternative approach is feasible by employing in vitro test methods based on reconstructed human epidermis (RHE), which is based on human epidermal keratinocytes (hEK). The models emulate the morphology, structure and metabolism of the human epidermis accurately (Rosdy and Clauss, 1990). Currently several RHE models are commercially available from various producers and academic institutes (De Wever et al., 2013; Groeber et al., 2011) and have been used for different research questions such as, skin corrosion (Kandárová et al., 2006), skin irritation (Cotovio et al., 2005), skin barrier formation (Thakoersing et al., 2013) and skin absorption (Schäfer-Korting et al., 2008). Currently the epiCS™ model by CellSystems (Pratt et al., 2014), the SkinEthic RHE (Alépée et al., 2014) and EpiSkin ™ (Alépée et al., 2016) model by L'Oréal, the EpiDerm™ model by Mattek (Chapman et al., 2014) and the ATERA-RHE by ATERA can be purchased worldwide. Additionally, the LabCyte EPI model by J-Tech (Katoh et al., 2009) and the KeraSkin-VM™ by MCTT Inc. (Jung et al., 2014) are available in Japan and Korea. Of the commercial RHE, two models, namely the EpiDerm™ and the EpiSkin™, were part of the first skin irritation validation study sponsored by the European Union Reference Laboratory for Alternatives to Animal testing (EURL-ECVAM). After optimization, the methods were found to be scientifically valid as a standalone method to identify skin irritants and were implemented by the OECD in the test guideline 439 (OECD, 2015a). Moreover, the results are the basis for the performance standards for in vitro skin irritation test methods based on reconstructed human epidermis (OECD, 2015b; Spielmann et al., 2007). Following the initial validation study, the SkinEthic™ RHE (Alépée et al., 2010) and the LabCyte EPI-MODEL™ (Kojima et al., 2013) were validated in catch-up validation studies and integrated in the test guideline 439.

However, the validated test protocols are linked to a limited number of models produced and marketed by a few companies only. Hence, all validated tests are dependent on the availability and the commercial strategies of the suppliers. An unrestricted approach has become possible by an open source reconstructed epidermis (OS-REp) established by the Henkel AG &Co. KGaA (De Wever et al., 2013, 2015).

Comparable to the open source concept in information technology, in which the source code of software is openly accessible, the OS-REp model will include the publication of the production protocol without legal restrictions (Bagozzi and Dholakia, 2006; Hertel et al., 2003; Lakhani and Von Hippel, 2003). Thus, any laboratory can use OS-REp for irritation testing and other research purposes without being dependent on commercially available models. Thereby the open source concept could foster the dissemination and implementation of skin models especially in countries with customs regulations that restrict the import of tissue models or budget restricted research laboratories. Moreover, open source allows end users to amend the source code, to improve the program or adapt it to a specific purpose. In the framework of alternative test methods it is envisioned that this philosophy will encourage a constant refinement of the model. However, it should be noted that once a model is implemented in a test procedure the model cannot be changed without again demonstrating the validity of the test methods.

The OS-REp is based on an initial protocol (Lemper et al., 2013; Poumay et al., 2004), refined by Henkel (HEN) and employed in an initial skin irritation test. In order to become accepted by regulatory authorities, we performed and describe a catch-up validation study with the OS-REp according to the EURL-ECVAM performance standards for in vitro skin irritation testing, previously an integral part of the OECD TG 439 (OECD, 2015b). In Phase I of the two-tiered approach a blinded study was conducted, in which all OS-REp models were exclusively produced at HEN and shipped to all participating laboratories. This study resulted in an overall accuracy of 75%. In addition, the skin irritation testing process was refined regarding the handling of volatile irritating test substances while leaving the procedural details of the testing protocol unchanged.

Here we present the results of the Phase II of the catch-up validation study to achieve regulatory acceptance for skin irritation testing. Due to structural, mechanistic and procedural similarity of the OS-REp model and the OS-REp skin irritation testing protocol with accepted test methods, the performance standards for in vitro skin Irritation testing are applicable to reduce the acceptance process (OECD, 2015b). In the present study HEN distributed only the standard operating procedures (SOP) for hEK isolation, OS-REp production and skin irritation testing to two additional independent laboratories. After a transfer phase all laboratories established their own hEK batches from diverse donors, generated their own OS-REp models and then only used the in-house produced models for the validation runs.

2. Material and methods

2.1. Study design

The catch up validation study was conducted according to the performance standards (OECD, 2015b) at the lead laboratory HEN, the Fraunhofer Institute for Interfacial Engineering and Biotechnology as part of the Fraunhofer Gesellschaft (FHG) and the Institute of Pharmacy at the Freie Universität Berlin (FUB). Each laboratory established own hEK batches and produced its own OS-REp in the respective facilities (Fig. 1). Subsequently, the models were employed to test the same set of 20 reference substances in each laboratory in three to five independent runs at different occasions. Test substances were coded and distributed by the Biotesys GmbH (Esslingen, Germany), safeguarding that the testing in the laboratories was conducted under blind conditions. The reference substances comprised ten irritants, i.e. category 2 substances according to the United Nations Globally Harmonized System of Classification and Labelling of Chemicals (UN GHS), and ten non-irritants, i.e. no category substances (Table 1) (United Nations, 2015).

In addition to the blinded test substances, two controls were included in each test run. As a negative control that controls for insufficient tissue viability, phosphate buffered saline (PBS, Life Technologies) was applied to the models. A 5% aqueous solution of sodium dodecyl sulfate (SDS, Sigma-Aldrich; Steinheim, Germany) was used as the positive control.

Per test run, each substance and each control was tested on three tissue replicates. Only test data, which met all acceptance criteria specified in the SOPs were considered valid for data analysis. In case the standard deviation between run replicates exceeded 18%, the respective substance was re-tested up to two times. However, if the acceptance criteria for the negative control and the positive control were not met,

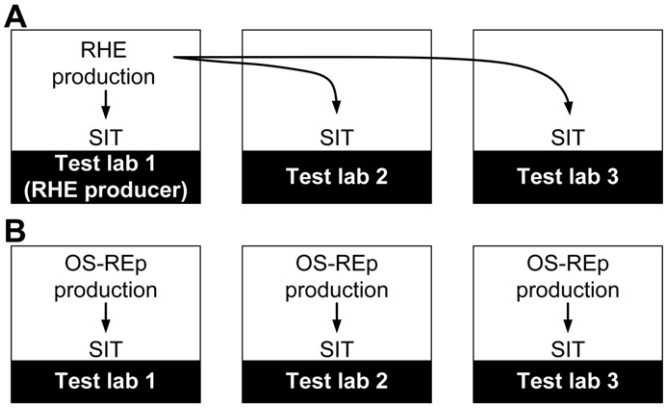

Fig. 1. Procedure for the conduction of a validation study for skin irritation testing (SIT). In the classical approach (A) the employed reconstructed human epidermis (RHE) is provided by the test developer to the additional two laboratories, where the models are used in the SIT validation test runs. In the present study (B) all participating laboratories established their own OS-REp production and then used the in-house produced models in the SIT.

Table 1
List of the 20 reference chemicals defined in the OECD performance standards (OECD, 2015a, 2015b), which were used to assess the reliability and predictive capacity of the OS-REp based skin irritation test. The chemicals comprise 10 non-irritants (no category) and 10 irritants (category 2) according to the UN GHS classification.

No.	Test substances	CAS Number	Supplier	Physical state	In vivo UN GHS category
1	1-bromo-4-chlorobutane	6940-78-9	Sigma-Aldrich	liquid, clear	No cat.
2	4-methyl-thio-benzaldehyde	3446-89-7	Sigma-Aldrich	liquid, clear, yellow	No cat.
3	allyl phenoxy-acetate	7493-74-5	Sigma-Aldrich	liquid, clear, colorless	No cat.
4	cinnamaldehyde	104-55-2	Sigma-Aldrich	liquid, clear, yellow	No cat.
5	diethyl phthalate	84-66-2	Sigma-Aldrich	liquid, clear	No cat.
6	heptyl butyrate	5870-93-9	Sigma-Aldrich	liquid, clear, colorless	No cat.
7	hexyl salicylate	6259-76-3	Sigma-Aldrich	liquid, clear, colorless	No cat.
8	isopropanol	67-63-0	AppliChem	liquid, clear, colorless	No cat.
9	methyl stearate	112-61-8	Sigma-Aldrich	solid, white	No cat.
10	naphthalene acetic acid	86-87-3	Sigma-Aldrich	solid, white	No cat.
11	1-bromohexane	111-25-1	Sigma-Aldrich	liquid, clear, colorless	Cat. 2
12	1-decanol	112-30-1	Sigma-Aldrich	liquid, clear, colorless	Cat. 2
13	1-methyl-3-phenyl-1-piperazine	5271-27-2	Sigma-Aldrich	solid, yellowish	Cat. 2
14	2-chloromethyl-3,5-dimethyl-4-methoxypyridine HCL	86604-75-3	Sigma-Aldrich	solid, white	Cat. 2
15	benzenethiol,5-(1,1-dimethylethyl)-2-methyl	7340-90-1	Acros-Organics	liquid, clear, colorless	Cat. 2
16	cyclamen aldehyde	103-95-7	Sigma-Aldrich	liquid, clear, colorless	Cat. 2
17	di-n-propyl disulphide	629-19-6	Sigma-Aldrich	liquid, clear, colorless	Cat. 2
18	heptanal	111-71-7	Sigma-Aldrich	liquid, clear, colorless	Cat. 2
19	potassium hydroxide (5% aq)	1310-58-3	Carl Roth	liquid, clear, colorless	Cat. 2
20	tetrachloroethylene	127-18-4	Sigma-Aldrich	liquid, clear, colorless	Cat. 2

the test run was considered as invalid. These runs had to be repeated in full. Thus, a maximum of five repeats for a single test substance, or five repeats for a complete test run were allowed. For quality control, the participating laboratories submitted the results to Seh consulting services (Paderborn, Germany) for independent statistical analysis. Additionally, auditable information was provided in a method documentation sheet as required by quality assurance systems such as GLP or ISO 17025.

2.2. Open source reconstructed epidermis

The OS-REp is based on a protocol published by Poumay et al. (Poumay et al., 2004) which was further refined in order to improve lifetime and architecture of the tissues. In the protocol, hEK were isolated from foreskin sample after receiving an informed consent and ethical approval. Subsequently, hEKs were seeded on the polycarbonate membranes of respective cell culture inserts (Millicell-PVF, pore size 0.4 μm, Merck Millipore; Darmstadt, Germany), in the third passage with a cell density of 5×10^5 hEK/cm², in 500 μL EpiLife® Submerse Medium based on EpiLife® supplemented with 0.2% v/v Bovine pituitary extract, 1 μg/mLml Recombinant human insulin-like growth factor-I, 0.18 μg/mL Hydrocortisone, 5 μg/mL Bovine transferrin, 0.2 ng/mL Human epidermal growth factor (all from Life Technologies; Darmstadt, Germany) and 1.5 mM $CaCl_2$ (Sigma-Aldrich Chemie; Munich, Germany). To ensure sufficient supply to the hEK 20 OS-REp were cultured in 38 mL medium in a 145 mm petri-dish (Greiner Bio One, Frickenhausen, Germany). Medium was changed after 24 h to EpiLife® air–liquid–interface (ALI) medium, additionally containing 73 μg/mL L-ascorbic acid 2-phosphate and 10 ng/mL keratinocyte growth factor (both Sigma-Aldrich). During the air–liquid–interface culture 10 OS-REp were cultured per petri-dish. Within a 19 days culture phase, in which the medium was replaced by fresh air–liquid–interface medium three times a week, a well-stratified epidermal layer was formed and the models could be used for skin irritation testing. Cell and tissue models were cultured at 37 °C and 5% CO_2 in a humidified incubator.

2.3. Quality control

2.3.1. Histology

Tissue samples were fixated in Bouin's fixation agent (Sigma-Aldrich; Steinheim, Germany) for 1 h, washed in tab water for 2 h and then embedded in paraffin. Subsequently, histological cross-sections of 3 μm were generated. For an overview of the general histological features tissue slides were stained with hematoxylin and eosin.

2.3.2. Assessment of barrier function

To determine the skin barrier, triplicates of OS-REp were challenged with 25 μL of a 1% solution of Triton-X 100 in PBS for 0, 1.5, 3, 4.5 and 6 h. After washing the models 8 times with 600 μL PBS tissue viability was assessed using the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazoliumbromid (MTT) assay (OECD, 2015a). For that OS-REp were placed in 200 μL MTT solution (1 mg/mL) (Sigma-Aldrich) for 3 h in a humidified atmosphere with 37 °C and 5% CO_2. The MTT reduction was then quantified by extracting the precipitated blue formazan salt with 2 mL 2-propanol and measuring the optical density of the extract at a wavelength of between 540 and 600 nm using a spectrophotometer. After correcting the data using 2-propanol as blank the relative tissue viability was calculated for each OS-REp by normalizing the corrected optical density values to the models that have not been treated with Triton-X 100 (0 h; Spielmann et al., 2007). From these data the ET_{50}-value, meaning the incubation time in Triton-X 100 solution that reduced tissue viability to 50%, was calculated. Based on historical data and the outcome of Phase I of the validation study the ET_{50} must be equal to or bigger than 2 h.

2.4. Skin irritation test

2.4.1. Test protocol

The skin irritation test was conducted according to the protocol as described in Phase I of the validation study. Briefly, 25 μl of liquid test substances or 25 μg of solid substances (moistened with 25 μ μl PBS) were applied to the surface of the models. After a treatment time of 35 minutes (Mewes et al., 2016) at room temperature the OS-REp were carefully rinsed 8 times with 600 μL PBS each. Additionally, the tissues were immersed 5 times in 500 mL fresh PBS. Following a 42 h post-exposure incubation at 37 °C and 5% CO_2 tissue viability was assessed using the MTT assay. Tissue viability of treated models was normalized to the negative control, which was set to 100%.

2.4.2. Prediction model

In the current study, the prediction model defined in the performance standards was applied (OECD, 2015a, 2015b). Thus, to predict skin irritation from the viability data a 50% threshold was used. In case, the mean viability after the skin irritation testing was above 50%, a substance was considered as non-irritating or 'no category' according

to the UN GHS system. Substances that reduced the viability to equal or below 50% were classified as irritating or 'Category 2' according to the UN GHS system.

2.4.3. Acceptance criteria

In concordance with the performance standards for in vitro skin irritation testing, each OS-REp batch used in the skin irritation test had to meet predefined acceptance criteria based on historical data of the method developers and on the results of Phase I of the validation study (Mewes et al., 2016). All PBS exposed OS-REp (negative control) had to show a high viability, with an absolute optical density between 0.6 [relative units] and 1.8 [relative units], and the relative viability following the exposure to 5% aqueous solution of SDS (positive control) had to be below 10%. The variability of the relative viabilities of the tissue replicates is used as an acceptance criterion for the test procedure, since problems such as defects in single tissues, effects of specific substances with inherently variable responses, or inappropriate application or dosing may result in excessive replicate variability. The standard deviation of the viability is used to measure this variability. A standard deviation of 18% or higher was considered as not acceptable and thus required a re-testing.

3. Results

3.1. Transfer phase

A vital aspect of this validation study was that all test laboratories produced their own OS-REp models with comparable properties and quality. Thus, the SOPs for the hEK isolation, cell culture, OS-REp assembly and skin irritation testing were transferred by the test developer HEN to the two participating laboratories FUB and FHG. The transfer process was organized in two phases. First, the technical staff from the partner laboratories was trained by HEN employees at HEN laboratories according to the SOPs. Next, the participating laboratories received hEK batches from the tests developer for the production of OS-REp and the morphology was then compared to OS-REp grown by HEN in parallel. The histological architectures of these tissue models were nearly indistinguishably from each other. In the second phase, the participating laboratories established own hEK batches and subsequently generated OS-REp from these cells. The quality of the OS-REp was assessed by histology, tissue viability (absolute optical density, MTT assay) and the ET_{50}-values indicative for the barrier function. The mean value of the absolute optical density produced at FUB was 0.68 [relative units] +/- 0.15 [relative units] and at FHG 1.28 [relative units] +/- 0.13 [relative units]. The barrier function at FUB and FHG, too, was well above the acceptance limit of 2 h of the OS-REp skin irritation SOP (data not shown).

3.2. Acceptance criteria

Generally the histological architecture of the OS-REp is highly comparable to the human epidermis (Fig. 2). This includes a clear single-layered Stratum basale, consisting of hEK arranged in a palisade-like pattern, at least 2-3 cell layers in the stratum spinosum, at least 2-3 layers of Stratum granulosum cells, characterized by its flattened cell polarity, absence of nuclei and presence of darkly-stained granules and a Stratum corneum, consisting of several layers of cornified hEK. Additionally, no histological differences are detectable between OS-REp models that were generated in the 3 laboratories (HEN, FUB and FHG).

In each experiment, a negative control was included in order to demonstrate appropriate viability of the tissue model. Three tissues were treated with PBS and their viability assessed using the MTT assay. The mean optical density in all test runs in the three laboratories fell within the acceptable optical density interval ranging from 0.6 [relative units] to 1.8 [relative units] as defined in the OS-REp quality criteria (Fig. 3, A). Additionally, each experiment included a positive control (5% SDS solution) in order to demonstrate the sensitivity of the tissue model to a known irritant. Fig. 3 B depicts the relative mean cell viabilities of the three tissue replicates and the respective standard deviation for each experiment of all laboratories. The mean viability of the positive control tissue replicates ranging from 0.3 to 3.1% was always clearly below the maximum acceptable value of 10% as defined for OS-REp. The mean value of the ET_{-50} of all test runs and re-tests was above the acceptable lower limit of 2 h (Fig. 3, C).

In 188 tests, the standard deviation values exceeded the acceptance cut-off value of 18% only 8 times (Fig. 3, C) and thus required a re-testing. To compensate for not acceptable variability in the first three experiments, the laboratories re-tested substances up to twice. All re-tests had acceptable standard deviation, so that for all 20 substances a complete data set in all three laboratories was available (Supplementary Table 1).

3.3. Reliability

Concordance of classifications between independent runs in laboratories was used to assess the within-laboratory reproducibility. Classifications were derived from relative cell viabilities, which are presented in Fig. 4, by applying the prediction model (non-irritant, if >50% viability; irritant, if ≤50% viability). Furthermore, Supplementary Table 2 gives an overview over the results of the three valid test runs and the two re-rest at the laboratories of HEN, FUB and FHG. A substance was considered reproducible within a laboratory if the three runs resulted in a concordant classification. At HEN, 16 of the 20 test substances gave concordant results resulting in a within-laboratory reproducibility of 80%. FHG had a within-laboratory reproducibility of 95%, with di-n-propyl disulphide being the only substance with discordant classifications. At FUB three substances had discordant classifications resulting in a within-laboratory concordance of 85%. The between-laboratory reproducibility was assessed by the concordance of classifications between the laboratories. Therefore, the final classification for each test substance per laboratory was obtained by using the mean viability over the three valid runs. 17 of the 20 substances gave concordant classifications which resulted in a between laboratory reproducibility of 85%. Table 2 contains a summary of the obtained results of the OS-REp skin irritation validation study in comparison to the acceptance limits.

3.4. Predictive capacity

The predictive capacity in this catch-up validation study was determined by comparing the in vitro classification obtained in the OS-REp

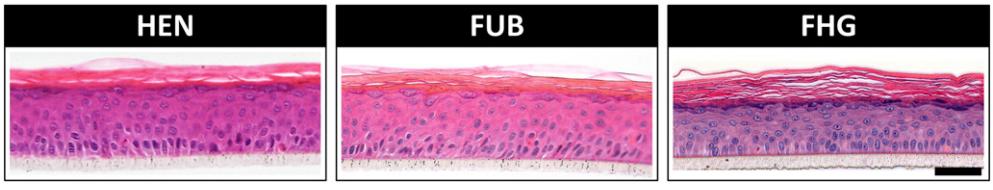

Fig. 2. Histological morphology of OS-REp. Hematoxylin and eosin stained cross section of OS-Rep generated in the 3 participated laboratories (HEN, FUB and FHG). The models were generated by seeding primary human epidermal keratinocytes (hEK) on a polycarbonate membrane. hEK differentiated into all layers, observed in native epidermis. The basal layer (stratum basale) with cubic shaped hEK, the spinous layer (stratum spinosum) containing cells with a more flatten morphology and the granular layer (stratum granulosum) where granula were additionally visible. After these viable cell layers, a corneous layer of non-viable hEK was formed (stratum corneum). Scale bar equals 50 μm.

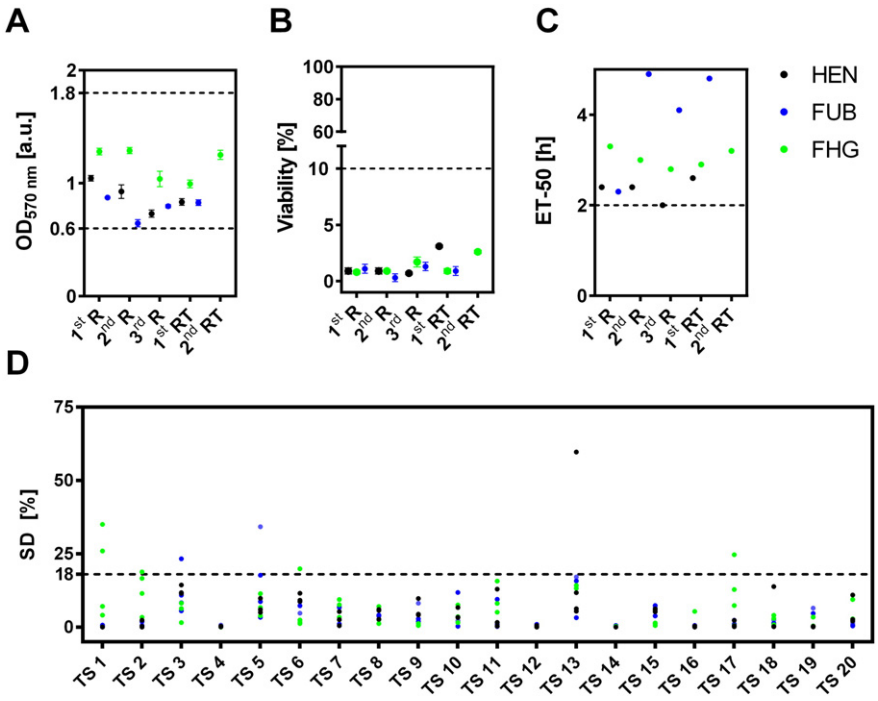

Fig. 3. Acceptance criteria of the OS-REp during the validation runs. The study comprises three full test runs (1st R, 2nd R and 3rd R) and two re-tests (1st RT and 2nd RT) for substances where the standard deviation of the mean results exceeded the 18% limit and was conducted at HEN, the FUB and FHG. All OS-REp batches met the required absolute optical density (OD) of the negative control (NC) (A),the viability of the positive control (PC) (B), and the ET_{50} value (C). Dotted lines indicate the defined limits of the respective acceptance criterion. It is to be noted that at the HEN and FUB laboratories only one re-test was necessary to achieve a comprehensive data set. Image D depicts the distribution of the standard deviation (SD) in the test runs for all test substances (TS 1–TS 20).

skin irritation test with the reference in vivo UN GHS classifications based on the Draize skin irritation test (OECD, 2015b). Three of the non-classified substances, namely 1-bromo-4-chlorobutane, 4-methylthio-benzaldehyde and cinnamaldehyde were classified as irritating by all three laboratories (Fig. 4, A) resulting in a specificity of 70%. The sensitivity was at least 80% in all laboratories. While HEN correctly identified all category 2 substances (100% sensitivity), FHG misclassified two category 2 substances (80% sensitivity) and FUB one category 2 substance (90% sensitivity) as non-irritating (Fig. 4, B). Consequently, HEN reached an accuracy of 85%, FHG of 80% and FUB of 75% (Table 2).

4. Discussion

Four epidermal models produced by three companies have been included in OECD TG 439 based on validation exercises (OECD, 2015a). The validation of a model includes the publication of the protocols for the skin irritation test, yet the exact manufacturing process for the respective tissue models is generally not disclosed by the manufacturer and hence not openly accessible, which might limit the scientific progress and dissemination of animal-free test methods. As an alternative approach, the open source concept was introduced by HEN into the field of alternative test methods that also includes the publication of manufacturing SOPs (De Wever et al., 2013, 2015). The availability of the models will be increased and the respective test can be conducted independently from commercial suppliers. This is of special interest in countries with customs hurdles for the import of commercially available living tissues e.g. Brazil, produced in the United States, Japan or the EU. Moreover, a broad scientific community can contribute to the further refinement of the model without legal restrictions.

Even though the open source concept allows and even encourages the constant refinement of the models, a vital aspect of regulatory accepted testing is that the results must be reproducible irrespective of the tissue model producer and applicant, respectively. Therefore, if the OS-REp skin irritation method is used for regulatory purposes, introduction of changes to the OS-REp manufacturing or skin irritation testing protocols may render the method unsuitable for these purposes as a changed protocol might have unintended side effects on the respective test method. In this regard, the open source concept in the framework of alternative methods differs substantially from open source in information technology as changes need to be approved and the validity of the test methods has to be demonstrated again.

Since for the irritation testing in the OS-REp model not only the test protocols are transferred but also the whole manufacturing process has been performed by the end user, this study evaluated if a naïve laboratory can perform the entire process at a high standard, which means the production of OS-REp models of sufficient quality in addition to the conduct of skin irritation testing. Although the within-laboratory-reproducibility in two laboratories was below the acceptance limit of 90%, the between-laboratory-reproducibility was with 85% above the limit of 80% specified in the OECD performance standards (OECD, 2015a, 2015b). The outcome of the transfer phase thus proved that the SOPs for OS-REp model production and skin irritation testing have been successfully transferred to the laboratories at the FUB and the FHG. However, to ensure the transferability of the protocols and thus reliable test results, a qualified training for a naïve laboratory that wishes to use the OS-REp model is proposed in order to implement all SOPs diligently.

Following this initial qualification a fully blinded catch-up validation study according to the performance standards (OECD, 2015b) was conducted, in which each laboratory solely employed OS-REp generated in the respective facility. Although the models were produced at independent facilities, all test runs met the predefined quality criteria showing the good transferability of the method.

Of note, in this study hEK from 4 different donors were employed. At HEN and FUB one cell donor and at FHG two donors were employed to generate models whereas cell batches were not pooled. Thereby the results reflect individual variability towards skin irritants within the population to a greater extent than previous work, which becomes evident in the discordant classification of few category 2 test substances. In the valid test runs 1-bromohexane (substance 11) and di-n-propyl disulphide (substance 17) were misclassified at FHG and 1-methyl-3-

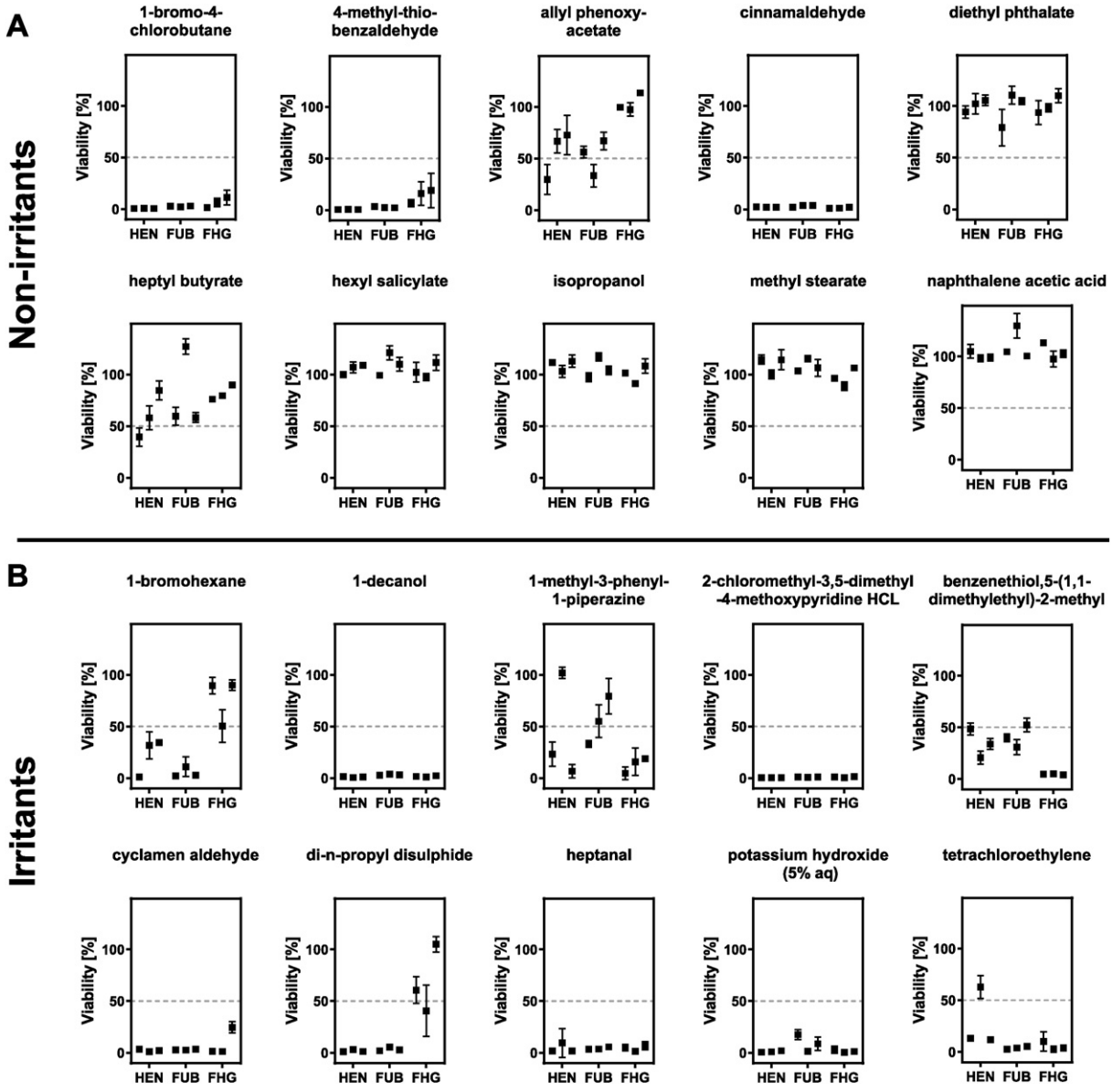

Fig. 4. Results of the blinded skin irritation testing. The effect of 10 non-irritants (A) and 10 irritants (B) on OS-REp produced at the respective test facility was investigated using the MTT assay, viability of the negative control (PBS) was set to 100%. The figures show the results of the three valid test runs at the laboratories of HEN, FUB and FHG. In each test run 3 OS-REp were used per test substance and the result of each test run is depicted as mean viability (back squares) and standard deviation (error bars). In case the standard deviation exceeded 18%, re-testing (up to twice) was conducted for the respective substance. A substance was classified as a non-irritant if the mean viability was higher than 50%. In case the mean viability was equal to or below 50% the substance was categorized as an irritant.

phenyl-1-piperazine (substance 13) was misclassified at FUB (Fig. 4). Employing the EpiDerm™ model, 1-methyl-3-phenyl-1-piperazine was also misclassified in the initial validation study in one out of three test runs (Spielmann et al., 2007). Even though this problem could be omitted in the subsequent optimization phase, it shows a tendency of the test substance to yield in discordant results. However, as the sensitivity was 90%, all laboratories met the OECD acceptance criterion of ≥80%.

Table 2
Summary of the OS-REp skin irritation validation study. The table shows the results for the three test laboratories HEN, FUB and FHG in comparison to the acceptance limits defined in the OECD performance standards (OECD, 2015a, 2015b). Furthermore, the overall mean values for all participating laboratories are presented (Mean).

Parameter	Acceptance limit	HEN	FUB	FHG	Mean
Sensitivity [%]	80	100	80	90	90
Specificity [%]	70	70	70	70	70
Accuracy [%]	75	85	75	80	80
Within-laboratory-reproducibility [%]	90	80	85	95	87
Inter-laboratory-reproducibility [%]	80	85			

Misclassification of 1-bromohexane was seen in the OS-REp model as well as in the KeraSkin™ and Labcyte model (Jung et al., 2014; Kojima et al., 2013). Furthermore, human in-vivo-data suggest that the irritating effect of 1-bromohexane is dependent on the individual genetic background of a human as only 16 out of 30 test persons reacted positively in a 4 hour patch test (Jírová et al., 2010). At FHG two test runs conducted with hEK from one donor yielded in a viability of well above 85%, whereas in models generated from the second donor the viability was decreased to 50.4% only. A similar donor dependency is observed for di-n-propyl disulphide that was classified as an irritant in only 6 out of 30 test persons in the 4 hour patch test study (Jírová et al., 2010). These ambiguous results are also reflected by the

misclassification of di-n-propyl disulphide in the Labcyte™ catch-up validation study and the outcomes of the EURL-ECVAM validation study, where the substance was false-negative in 4 of 6 test runs employing the EpiDerm™ and EpiSkin™ model (Kojima et al., 2013; Spielmann et al., 2007).

A pronounced donor dependency was also described for the expression of metabolic enzymes in native human skin and several skin models. These results indicate that the individual genetic background of a cell donor is reflected in skin models generated from cells of this donor (Oesch et al., 2014; Wiegand et al., 2014). Currently it cannot be decided if the genetic background of the cell donors is the only reason underlying the variation of the irritant response observed here. However, since di-n-propyl disulphide and 1-bromohexane show discordant results between different cell donors, future validation studies may profit from replacement of the critical test substances in the OECD test guideline by less ambiguous substances.

Of the 10 non-irritating substances, 1-bromo-4-chlorobutane (substance 1), cinnamaldehyde (substance 4) and 4-methylthio-benzaldehyde (substance 2) were falsely classified as irritants in all laboratories. This resulted in a specificity of 70% (Fig. 4), which is within the acceptance limit of the performance standards. Despite their in vivo classification as non-irritants, these substances are known to yield false positive results in epidermal model-based skin irritation tests (Alépée et al., 2010; Jung et al., 2014; Kojima et al., 2013; Spielmann et al., 2007). Taken together, with an accuracy of 80% the OS-REp skin irritation test is sufficiently accurate to meet the acceptance criterion of ≥75%. The predictive capacity demonstrated here is comparable or even exceeding previous validation studies — which is a remarkable result as all participating laboratories established independent cell lots and manufactured OS-REp model batches on their own.

After receiving a training, the OS-REp skin irritation test could be used by laboratories for regulatory accepted skin irritation testing independent of commercially available epidermis models. To increase the within-laboratory-reproducibility, special care has to be taken to train the responsible staff in the procedures. In addition, for guaranteed reproducibility and predictive capacity of the results, strict quality criteria have to be achieved. Besides compliance with the quality criteria listed in the performance standards, we suggest that a naïve laboratory that intends to use the OS-REp protocol for regulatory purposes has to demonstrate sufficient capability in a proficiency exercise employing the 20 reference test substances listed in the performance standards (OECD, 2015b).

In conclusion, the present catch-up validation study demonstrates the capacity of the OS-REp-based skin irritation test to discriminate between skin irritating and non-irritating substances with a high overall accuracy of 80%. With the exception of the within-laboratory-reproducibility, all performance standards of the OECD test guideline were met or even exceeded, even though the study was the first to comprise the independent manufacturing of tissue models in all three participating laboratories. Therefore, after further optimization the test method is moving in the direction of being a full standalone replacement of the Draize in vivo skin irritation test, and the OS-REp skin irritation test might set a precedent for further open source models in the framework of alternative testing.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.tiv.2016.07.008.

Conflict of interest statement

The authors indicate no conflict of interest.

Transparency document

The Transparency document associated with this article can be found, in online version.

Acknowledgments

The authors kindly thank Brigitte Höhl for technical support and the "Fraunhofer Zukunftsstiftung" for financially supporting the study. Furthermore, the authors would like to express their gratitude to the Biotesys GmbH (Esslingen, Germany) for the coding and shipping of the test substances. They especially thank Hannelore Gonska from the FUB for her excellent technical assistance in the validation study.

References

Alépée, N., Tornier, C., Robert, C., Amsellem, C., Roux, M.-H., Doucet, O., Pachot, J., Méloni, M., de Brugerolle de Fraissinette, A., 2010. A catch-up validation study on reconstructed human epidermis (SkinEthic RHE) for full replacement of the Draize skin irritation test. Toxicol. In Vitro 24, 257–266. http://dx.doi.org/10.1016/j.tiv.2009.08.024.

Alépée, N., Robert, C., Tornier, C., Cotovio, J., 2014. The usefulness of the validated SkinEthic™ RHE test method to identify skin corrosive UN GHS subcategories. Toxicol. In Vitro 28, 616–625. http://dx.doi.org/10.1016/j.tiv.2013.12.013.

Alépée, N., Hibatallah, J., Klaric, M., Mewes, K.R., Pfannenbecker, U., McNamee, P., 2016. Assessment of cosmetic ingredients in the in vitro reconstructed human epidermis test method EpiSkin™ using HPLC/UPLC-spectrophotometry in the MTT-reduction assay. Toxicol. In Vitro 33, 105–117. http://dx.doi.org/10.1016/j.tiv.2016.02.006.

Bagozzi, R.P., Dholakia, U.M., 2006. Open source software user communities: a study of participation in Linux user groups. Manag. Sci. 52, 1099–1115. http://dx.doi.org/10.1287/mnsc.1060.0545.

Chapman, K.E., Thomas, A.D., Wills, J.W., Pfuhler, S., Doak, S.H., Jenkins, G.J.S., 2014. Automation and validation of micronucleus detection in the 3D EpiDerm™ human reconstructed skin assay and correlation with 2D dose responses. Mutagenesis 29, 165–175. http://dx.doi.org/10.1093/mutage/geu011.

Cotovio, J., Grandidier, M., Portes, P., Roguet, R., Rubinstenn, G., 2005. The in vitro skin irritation of chemicals: optimisation of the EPISKIN prediction model within the framework of the ECVAM validation process. Altern. Lab. Anim 33, 329–349.

De Wever, B., Petersohn, D., Mewes, K.R., 2013. Overview of human three-dimensional (3D) skin models used for dermal toxicity assessment — part 1. Househ. Pers. Care Today 8, 18–23.

De Wever, B., Goldberg, A., Eskes, C., Roggen, E., Vanparys, P., Schro, K., 2015. "Open source"-based engineered human tissue models : a new gold standard for nonanimal testing through. Appl. In Vitro Toxicol. 1, 5–9. http://dx.doi.org/10.1089/aivt.2014.0011.

Draize, J., Woodard, G., Calvery, H., 1944. Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. J. Pharmacol.

Fentem, J.H., Archer, G.E.B., Balls, M., Botham, P.A., Curren, R.D., Earl, L.K., Esdaile, D.J., Holzhütter, H.-G., Liebsch, M., 1998. The ECVAM international validation study on in vitro tests for skin corrosivity. 2. Results and evaluation by the management team. Toxicol. In Vitro 12, 483–524. http://dx.doi.org/10.1016/S0887-2333(98)00019-8.

Groeber, F., Holeiter, M., Hampel, M., Hinderer, S., Schenke-Layland, K., 2011. Skin tissue engineering—in vivo and in vitro applications. Adv. Drug Deliv. Rev. 63, 352–366. http://dx.doi.org/10.1016/j.addr.2011.01.005.

Hertel, G., Niedner, S., Herrmann, S., 2003. Motivation of software developers in open source projects: an Internet-based survey of contributors to the Linux kernel. Res. Policy 32, 1159–1177. http://dx.doi.org/10.1016/S0048-7333(03)00047-7.

Jírová, D., Basketter, D., Liebsch, M., Bendová, H., Kejlová, K., Marriott, M., Kandárová, H., 2010. Comparison of human skin irritation patch test data with in vitro skin irritation assays and animal data. Contact Dermatitis 62, 109–116. http://dx.doi.org/10.1111/j.1600-0536.2009.01640.x.

Jung, K.-M., Lee, S.-H., Jang, W.-H., Jung, H.-S., Heo, Y., Park, Y.-H., Bae, S., Lim, K., Seok, S.H., 2014. KeraSkin™-VM: a novel reconstructed human epidermis model for skin irritation tests. Toxicol. In Vitro 28, 742–750. http://dx.doi.org/10.1016/j.tiv.2014.02.014.

Kandárová, H., Liebsch, M., Spielmann, H., Genschow, E., Schmidt, E., Traue, D., Guest, R., Whittingham, A., Warren, N., Gamer, A.O., Remmele, M., Kaufmann, T., Wittmer, E., De Wever, B., Rosdy, M., 2006. Assessment of the human epidermis model SkinEthic RHE for in vitro skin corrosion testing of chemicals according to new OECD TG 431. Toxicol. In Vitro 20, 547–559. http://dx.doi.org/10.1016/j.tiv.2005.11.008.

Katoh, M., Hamajima, F., Ogasawara, T., Hata, K., 2009. Assessment of human epidermal model LabCyte EPI-MODEL for in vitro skin irritation testing according to European Centre for the Validation of Alternative Methods (ECVAM)-validated protocol. J. Toxicol. Sci. 34, 327–334. http://dx.doi.org/10.2131/jts.34.327.

Kojima, H., Katoh, M., Shinoda, S., Hagiwara, S., Suzuki, T., Izumi, R., Yamaguchi, Y., Nakamura, M., Kasahawa, T., Shibai, A., 2013. A catch-up validation study of an in vitro skin irritation test method using reconstructed human epidermis LabCyte EPI-MODEL24. J. Appl. Toxicol. http://dx.doi.org/10.1002/jat.2937.

Lakhani, K.R., Von Hippel, E., 2003. How open source software works: "free" user-to-user assistance. Res. Policy 32, 923–943. http://dx.doi.org/10.1016/S0048-7333(02)00095-1.

Lemper, M., De Paepe, K., Rogiers, V., 2013. Practical problems encountered during the cultivation of an open-source reconstructed human epidermis model on a polycarbonate membrane and protein quantification. Skin Pharmacol. Physiol. 27, 106–112. http://dx.doi.org/10.1159/000351814.

Mewes, K.R., Fischer, A., Zöller, N.N., Laubach, V., Bernd, A., Jacobs, A., van Rompay, A., Liebsch, M., Pirow, R., Petersohn, D., 2016 Jul 15. Catch-up validation study of an in vitro skin irritation test method based on an Open Source Reconstructed Epidermis (Phase I). Toxicol. In Vitro 36, 238–253.

OECD, 2015a. OECD Guidelines for the Testing of Chemicals No. 439: In vitro Skin Irritation: Reconstructed Human Epidermis Test Method (Paris).

OECD, 2015b. Performance Standards For The Assessment Of Proposed Similar Or Modified In Vitro Reconstructed Human Epidermis (RhE) Test Methods For Skin Irritation Testing As Described in TG 439.

Oesch, F., Fabian, E., Guth, K., Landsiedel, R., 2014. Xenobiotic-metabolizing enzymes in the skin of rat, mouse, pig, guinea pig, man, and in human skin models. Arch. Toxicol. 88, 2135–2190. http://dx.doi.org/10.1007/s00204-014-1382-8.

Poumay, Y., Dupont, F., Marcoux, S., Leclercq-Smekens, M., Hérin, M., Coquette, A., 2004. A simple reconstructed human epidermis: preparation of the culture model and utilization in in vitro studies. Arch. Dermatol. Res. 296, 203–211. http://dx.doi.org/10.1007/s00403-004-0507-y.

Pratt, L.F., Troese, M.J., Fuchs, H.W., Engelking, O., DeGeorge, G.L., 2014. The highly differentiated 3D epidermal skin model (epiCS®) to characterize skin sensitizers in mixtures. Toxicol. Lett. 229, S141. http://dx.doi.org/10.1016/j.toxlet.2014.06.497.

Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products.

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. J. Invest. Dermatol. 95, 409–414.

Russell, W.M.S., Burch, R.L., Hume, C.W., 1959. The Principles Of Humane Experimental Technique.

Schäfer-Korting, M., Bock, U., Diembeck, W., Düsing, H.J., Gamer, A., Haltner-Ukomadu, E., Hoffmann, C., Kaca, M., Kamp, H., Kersen, S., et al., 2008. The use of reconstructed human epidermis for skin absorption testing: results of the validation study. Altern. Lab. Anim 36, 161–187.

Spielmann, H., Hoffmann, S., Liebsch, M., Botham, P., Fentem, J.H., Eskes, C., Roguet, R., Cotovio, 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. Altern. Lab. Anim 35, 559–601.

Stobbe, J.L., Drake, K.D., Maier, K.J., 2003. Comparison of in vivo (Draize method) and in vitro (Corrositex assay) dermal corrosion values for selected industrial chemicals. Int. J. Toxicol. 22, 99–107. http://dx.doi.org/10.1080/10915810305094.

Thakoersing, V.S., van Smeden, J., Mulder, A.A., Vreeken, R.J., El Ghalbzouri, A., Bouwstra, J.A., 2013. Increased presence of monounsaturated fatty acids in the stratum corneum of human skin equivalents. J. Invest. Dermatol. 133, 59–67. http://dx.doi.org/10.1038/jid.2012.262.

The European Parliament and the Council of the European Union, 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 1999/4. European Union.

United Nations, 2015. Globally Harmonized System of Classification and Labelling of Chemicals (GHS).

Wiegand, C., Hewitt, N.J., Merk, H.F., Reisinger, K., 2014. Dermal xenobiotic metabolism: A comparison between native human skin, four in vitro skin test systems and a liver system. Skin Pharmacol. Physiol. 27, 263–275. http://dx.doi.org/10.1159/000358272.