

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



K.R. Mewes ^{a,*}, A. Fischer ^a, N.N. Zöller ^b, V. Laubach ^b, A. Bernd ^b, A. Jacobs ^c, A. van Rompay ^c, M. Liebsch ^d, R. Pirow ^d, D. Petersohn ^a

^a Henkel AG & Co. KGaA, Henkelstr. 67, 40589 Düsseldorf, Germany

^b Goethe-University, Medical School, Department of Dermatology, Venereology and Allergology, Theodor-Stern-Kai 7, 60590 Frankfurt, Germany

^c VITO NV Flemish Institute for Technological Research, Environmental Risk and Health Unit, Boeretang 200, 2400 Mol, Belgium

^d Bundesinstitut für Risikobewertung (BfR), Max-Dohrn-Straße 8–10, 10589 Berlin, Germany

ARTICLE INFO

Article history:

Received 15 January 2016

Received in revised form 30 June 2016

Accepted 13 July 2016

Available online 16 July 2016

Keywords:

Skin irritation test
Reconstructed human epidermis
Open source concept
Catch-up validation study
Cross-contamination
OECD performance standards

ABSTRACT

We have developed a new in vitro skin irritation test based on an open source reconstructed epidermis (OS-REp) with openly accessible protocols for tissue production and test performance. Due to structural, mechanistic and procedural similarity, a blinded catch-up validation study for skin irritation according to OECD Performance Standards (PS) was conducted in three laboratories to promote regulatory acceptance, with OS-REp models produced at a single production site only. While overall sensitivity and predictive capacity met the PS requirements, overall specificity was only 57%. A thorough analysis of the test results led to the assumption that some of the false-positive classifications could have been evoked by volatile skin-irritating chemicals tested in the same culture plate as the non-irritants falsely predicted as irritants. With GC/MS and biological approaches the cross-contamination effect was confirmed and the experimental set-up adapted accordingly. Retesting of the affected chemicals with the improved experimental set-up and otherwise identical protocol resulted in correct classifications as non-irritants. Taking these re-test results into account, 93% overall sensitivity, 70% specificity and 82% accuracy was achieved, which is in accordance with the OECD PS. A sufficient reliability of the method was indicated by a within-laboratory-reproducibility of 85–95% and a between-laboratory-reproducibility of 90%.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

In order to protect consumers from undesired side effects or even serious health damage all chemical ingredients of so-called consumer

products must undergo an extensive toxicological assessment. Human skin is the first contact site for numerous chemicals. Thus in vitro test methods, which aim at reliably identifying reactions of irritating chemicals with dermal tissue, have been developed. Moreover, since 2009 only in vitro tests are still allowed in the European Union to demonstrate the safety of new cosmetic ingredients for the skin (EU, 2009). During the last decade three-dimensional skin models which closely mimic native human epidermis in respect of tissue architecture and physiology have been recognized as suitable tools to test a wide variety of chemicals (de Wever et al., 2013). However, validation and acceptance by organisations like EUR-ECVAM and the OECD are mandatory for any in vitro test method before being used in the framework of the EU Cosmetics Regulation (EC) No. 1223/2009 (EU, 2009).

The latest version of the OECD TG 439 lists 4 commercially available in vitro test methods for skin irritation testing, based on different epidermal equivalents, which have undergone formal validation: the EpiSkin™ and EpiDerm™ models, the Skin Ethic™ RHE and the LabCyte EPI-MODEL24SIT (Spielmann et al., 2007; EUR-ECVAM, 2008, 2009a, 2009b; Katoh et al., 2009; Kojima et al., 2012, 2014; OECD, 2011, 2015a).

Some other pure epidermal models – Sterlab Epidermis, StratiCELL RHE, epiCS (formerly EST-1000), and KeraSkin™-VM have entered the

Abbreviations: 1-BH, 1-bromohexane; ALI, Air-Liquid Interface; BLR, Between-laboratory-reproducibility; CyA, Cyclamen aldehyde; D-PBS, Dulbecco's phosphate-buffered saline; ECETOC, European Centre for Ecotoxicology and Toxicology; EURL-ECVAM, European Union Reference Laboratory for the Validation of Alternative Methods; ET₅₀, effective time at 50% viability; EU, European Union; GC/MS, gas chromatography/mass spectroscopy; MITT, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazoliumbromide; NC, negative control; OD, optical density; OECD, Organization for Economic Co-operation and Development; OS-REp, open source reconstructed epidermis; PC, positive control; PS, performance standards; REACH, Registration, Evaluation, Authorisation and Restriction of Chemical Substances; RC, reference chemical; RHE, reconstructed human epidermis; SD, standard deviation; SDS, sodium dodecyl sulfate; SOP, standard operating procedure; SIT, skin irritation testing; SIVT, skin irritation validation test; TG, testing guideline; VRM, validated reference method; WLR, within-laboratory-reproducibility; ZEBET, Center for Documentation and Evaluation of Alternative Methods to Animal Experiments at the BfR (Federal Institute for Risk Assessment).

* Corresponding author at: Henkel AG & Co. KGaA, Biological & Clinical Research, Henkelstr. 67, 40589 Düsseldorf, Germany.

E-mail address: karsten.mewes@henkel.com (K.R. Mewes).

market, too, although formal regulatory acceptance based on a validation study for skin irritation testing is still missing (<http://www.sterlab.com>; <http://www.straticell.com>; <http://reconstructed-human-epidermis.com>, Jung et al., 2014).

All human epidermal equivalents mentioned above have one aspect in common: They are produced at the respective company's production site only and partly based on confidential and legally protected protocols. For the user of one of the accepted skin irritation tests the advantage of working with standardized and quality-approved tissue equivalents from one supplier must be balanced against tissue costs. In addition, several non-European countries have set up high customs barriers regarding the import of living human tissues, which hinder interested laboratories in these countries from using the tissue equivalents commercially produced abroad. Hence, these countries are in part cut off from innovation in the framework of in vitro alternative methods. Another disadvantage of using commercial skin models is its possible quality loss after shipment over longer periods of times and distances. In order to circumvent this restricted situation on the skin model market we decided to develop an in vitro test method for skin irritation assessment based on an open source reconstructed epidermis (OS-REp). The Open Source Concept was introduced first in information technology (IT) in the early 1980s with LINUX as the most prominent example for free and open source software (see e.g. Nicolosi and Ruivenkamp, 2012).

What has become a story of success in IT is still largely unknown in the field of alternatives to animal testing. Until today, nearly all regulatory accepted and/or scientifically valid in vitro test methods based on 3D tissue models depend on a few companies only, which produce and distribute their tissue equivalents for commercial purposes. This situation is illustrated by the circumstance that the 4 epidermal models which are currently included in the OECD test guideline 439 are produced by 3 companies only.

By introducing the open source concept into the framework of in vitro testing for regulatory purposes we intend to accelerate the development, acceptance and dissemination of innovative alternatives to animal testing worldwide. Our concept means that all protocols for cell isolation and proliferation, for the construction of the 3-dimensional tissue equivalent and the performance of the irritation test are herewith made publicly available without any legal restrictions due to intellectual property or licenses. Thus the OS-REp model culture is intended to be established by any potential user at any laboratory worldwide to be used either for skin irritation tests according to the protocol depicted in this paper or for other purposes. The open source concept will eventually lead to greater independence from commercial suppliers, cost reduction for the users and a suitable strategy to circumvent trade hurdles for living tissue models in several countries, e.g. in India and Brasil. However, it does not exclude the option to start a business with the newly-developed methods.

The OS-REp model is based on a protocol for epidermal equivalent culture originally developed and published by Poumay et al. (2004) for the unrestricted use in any laboratory. It consists of normal, human-derived epidermal keratinocytes, which have been cultured at the air-liquid interface (ALI) to form a multilayered, highly differentiated epidermal tissue. The tissue exhibits organized basal, spinous and granular layers, and a multi-layered stratum corneum and thus closely mimics native human epidermis. Based on the PS and taking the specific tissue properties of the OS-REp into account, a protocol for testing the skin-irritating potential of topically applied chemicals was developed. However, in order to gain regulatory acceptance as alternative method its reliability and high predictive capacity must be proven in a validation process, generally consisting of a pre-validation study followed by a multicentric validation study. Because of the structural, mechanistic and procedural similarity of the OS-REp model and the OS-REp SIT protocol with already accepted test methods, the acceptance process can be short-cut by following the Performance Standards for in vitro Skin Irritation Testing (OECD, 2015b).

The OS-REp catch-up validation study was conducted in a two-tier approach. In the first phase (Phase I), which is the subject of this paper, robustness and predictive capacity of the OS-REp skin irritation test were assessed in line with the OECD Performance Standards in 3 independent laboratories. All OS-REp models used in this study were produced exclusively at the developer's laboratory and subsequently shipped to the validation partners. This strategy enabled us to focus the analysis of the study results exclusively on the design of the SIT without being jeopardized by possible variations regarding tissue quality and properties of OS-REp models cultured at different sites.

To complete the catch-up validation study under open source conditions, a second study phase (Phase II) was initiated, where it was mandatory for the study partners to produce the OS-REp models for skin irritation testing by themselves, including keratinocyte isolation from human donors and selection of the best suited cell lots for OS-REp production. The results of phase II study will be reported in a second publication in this journal (Groeber et al., 2016-in this volume).

In this paper we present the results of the first phase of the OS-REp SIT catch-up validation study, independently conducted in 3 laboratories and in accordance with the OECD Performance Standards for in vitro skin irritation testing (2015b). Sensitivity and accuracy matched the PS thresholds, whereas specificity was clearly below the threshold of 70%. Those test data which led to the misclassification of non-irritant reference chemicals as irritants were retrospectively analyzed. We hypothesized that the volatile fractions of skin-irritating reference chemicals had damaged those tissues which had been tested in parallel with the non-irritating chemicals in the same culture plate. Using gas chromatography/mass spectroscopy techniques, the damaging effect of volatile irritating chemicals, namely heptanal and cyclamen aldehyde, on the OS-REp equivalents was confirmed, which subsequently resulted in an optimized testing protocol. Based on the optimized protocol which prevented any kind of cross-contamination, the previously misclassified chemicals were re-tested on the OS-REp models. Taking the re-test results into account, specificity and accuracy now both matched the PS.

In addition to the test results this paper provides detailed protocols which will allow trained researchers to establish reconstructed epidermal models and to conduct skin irritation testing under open source conditions independently from commercial suppliers and country. With this approach we contribute worldwide to the principles of refinement, reduction and replacement (3R) of animal experiments (Russel and Burch, 1959).

2. Materials & methods

2.1. Materials

If not indicated otherwise, all chemicals were purchased from Sigma-Aldrich, Germany. EpiLife® medium, Human Keratinocyte Growth Supplement (HKGS) supplement, penicillin/streptomycin solution and trypsin/EDTA solution were purchased from Life Technologies GmbH, Germany. Accutase® was delivered by PAA, Germany. The cell strainers were from BD Falcon, USA, the co-culture inserts from Merck Millipore, Germany (Millicell-PVF, 0.4 µm pore size, 12 mm outer insert diameter, membrane area ≈ 0.63 cm²). Fetal Clone II serum was purchased from Thermo Scientific, Germany, the fetal calf serum from Biochrom GmbH, Germany.

2.2. Culture media

Transport medium (TP medium): Dulbecco's Modified Eagle Medium (DMEM, incl. 4.5% Glucose, GlutaMAX™, Pyruvate) was supplemented with L-ascorbic acid 2-phosphate (1 mM), penicillin/streptomycin (100 UI/mL and 100 µg/mL) and fetal calf serum (10% v/v).

Keratinocyte medium for freshly isolated keratinocytes (passage 0) and feeder cell culture (K medium): A blend of DMEM and Ham's F-12

medium (volume ratio 1: 0.54) was supplemented with cholera toxin (10^{-1} nM), tri-*iodo* thyronin (5 µg/mL), L-ascorbic acid 2-phosphate (1 mM), penicillin (100 UI/mL)/streptomycin (100 µg/mL), recombinant human epidermal growth factor (10 ng/mL), adenin (24.3 µg/mL), hydrocortisone (0.4 µg/mL), bovine insulin (0.12 U/mL) and Fetal Clone II (10% v/v).

Primary keratinocytes in all passages higher than P0 were cultured in EpiLife® Basal medium supplemented with 1% HKGS consisting of bovine pituitary extract (0.2% v/v end concentration), recombinant human insulin-like growth factor-I (1 µg/mL), hydrocortisone (0.18 µg/mL), bovine transferrin (5 µg/mL) and human epidermal growth factor (0.2 ng/mL) and with penicillin (100 U/mL)/streptomycin (100 µg/mL). During the submersed phase the OS-REp models were cultured with EpiLife® Submerse medium supplemented with 1% HKGS, Penicillin (100 U/mL)/Streptomycin (100 µg/mL), and CaCl₂ solution (1.5 mM final concentration). After being lifted to the air liquid interface (ALI), the models were cultured with EpiLife® ALI medium consisting of EpiLife® Submerse medium supplemented with ascorbic acid phosphate solution (73 µg/mL end concentration) and keratinocyte growth factor (KGF, 10 ng/mL).

3. Epidermal model culture

3.1. Origin and isolation of keratinocytes

The OS-REp models were produced according to a protocol published by Poumay et al. (2004) with some modifications. Keratinocytes were isolated from juvenile foreskin biopsies of boys not older than 7 years. Use of the foreskin tissues after circumcision was in accordance with German law and took place only after informed written consent by the parents.

The enzymatic separation of epidermis and dermis started within 24 h after surgery. The biopsies were transported to the laboratory in transport medium (TP medium). Adipose tissue and adhering blood vessels were removed, and the biopsy was cut into small pieces of approx. 5–8 mm lengths with a scalpel. The pieces were placed in a petri dish with the epidermis facing upwards, followed by the addition of thermolysin solution (500 µg/mL in 0.01 M HEPES buffer). The volume of thermolysin was adjusted to a height of approx. half of the tissue thickness without covering the epidermal surface. After incubation for 15–18 h at 4 °C the epidermis was removed from the underlying dermis with tweezers, the dermis was discarded. The epidermal sheets were transferred into 20 mL of a Trypsin-EDTA solution (0.05% trypsin in 0.2 g/L EDTA) and incubated for 20 min at 37°. Repeated shaking during the incubation period increased the dissociation success. The enzymatic reaction was stopped with an equal volume of keratinocyte medium (K medium), and the cells were isolated by repeated aspiration of the epidermal pieces into and released from a 10 mL pipette. The cell suspension was applied onto a cell strainer (70 µm mesh size, BD Falcon, USA), and the resulting suspension was centrifuged at 178 g to collect the cells. The pellet was resuspended in 10 mL K medium, and after determination of the cell number the keratinocytes were seeded into cell culture flasks (175 cm² area) at a density of 2–4 × 10⁶ cells per flask. In order to support optimal cell proliferation the keratinocytes were grown on a feeder layer consisting of mitomycin C-inactivated human skin fibroblasts. The feeder cell culture was started in K medium 2–3 days before keratinocyte isolation to enable proper medium conditioning, and the keratinocytes were added without previous medium change. The keratinocytes were then cultured at 37 °C and 5% (v/v) CO₂ in an incubator (passage 0). The first medium exchange took place 24 h after inoculation, followed by 2–3 exchanges per week, depending on the growth rate. After having reached a confluence degree of approx. 75–80%, the keratinocytes were harvested. The culture was rinsed with PBS w/o Ca²⁺ and Mg²⁺ (Invitrogen) once, and the cells were detached by incubation with a Trypsin-EDTA solution for a maximum of 8 min. They were sedimented by subsequent centrifugation at

178 g for 5 min. All P0 keratinocytes were immediately frozen at –80 °C in keratinocyte medium supplemented with 10% (v/v) DMSO at a cell density of 3 × 10⁶ cells/mL and stored in liquid nitrogen until needed.

In order to multiply the cell number, cryostored cells of P0 were quickly thawed and seeded into large culture flasks (175 cm²) at a density of 750,000 cells each in 25 mL of prewarmed EpiLife® Basal medium. It is crucial for the culture process that for any passage higher than P0 exclusively media based on EpiLife® without feeder cells must be used. Culture medium was exchanged 3 times a week until the culture had reached approx. 80–90% confluence. Then the cells were harvested with Accutase® (10 mL/culture flask for 8–10 min) and subsequently frozen in liquid nitrogen again (passage 1).

3.2. OS-REp model culture

For the OS-REp model production, keratinocytes of passage 3 were needed. Thus, keratinocytes which had been cryopreserved at passage 1 were thawed, seeded into cell culture flasks at a density of 750,000 cells/flask and cultured with EpiLife® Basal medium until ~80–90% confluence at 37 °C/5% CO₂ (passage 2). Cells were harvested with Accutase®, counted and seeded into Millicell co-culture inserts at a seeding density of 315,000 cells (~500,000 cells/cm² membrane area; passage 3). After 24 h culture under submersed conditions with EpiLife® Submerse medium the cultures were lifted at the ALI by simultaneously discarding the culture medium from the inserts' inner cylinder and lowering the medium level in the culture vessel, e.g. a 6-well plate or a petri dish, to the level of the insert membrane. At least 1.5 mL of medium per tissue model for 2 days is mandatory for the ALI culture. The models were cultured at the ALI for 19 days with EpiLife® ALI medium. Epidermal models whose surface did not dry at any time during the ALI phase were discarded due to compromised terminal differentiation.

4. Skin irritation test

4.1. Study design

The OS-REp skin irritation test catch-up validation study was conducted according to the performance standards (PS) which had been previously defined by the EURL-ECVAM, later became an integral part of the OECD Test Guideline 439, and have been recently adopted as a separate OECD PS Document (OECD, 2015b). The Reference Chemicals were coded and distributed by ZEBET and the method was assessed independently in three laboratories comprising the lead laboratory (Henkel, Germany) and two 'naive' laboratories (University Frankfurt, Germany, and VITO NV, Belgium). Each laboratory tested the same set of 20 coded reference chemicals (Table 1) under blinded conditions in 3–5 independent runs at different occasions (Fig. 1). In each run, each test chemical was concurrently tested in three tissue replicates. In addition, for the set of test chemicals tested on that day three replicate tissues were treated with the negative control agent (NC) and three replicate tissues were treated with the positive control agent (PC). Only test data which met all test acceptance criteria defined in the SOP were considered valid and hence could be used in the data analysis. However, in case of invalid data, a maximum of 5 repeats for a single test chemical, or five repeats for a complete test run were allowed. For quality control, experimental data were collected in Excel spreadsheets and submitted to ZEBET/BfR for independent biostatistical analysis. To complete these data with additional auditable information as required by quality assurance systems like GLP or ISO 17025, for each testing day a "method documentation sheet (MDS)" was used for recording all test conditions other than the experimental test data. A flowchart of the pivotal working steps for conducting the OS-REp model-based in vitro skin irritation assay is depicted in Fig. 2.

Table 1

Reference chemicals for the skin irritation test and their respective code numbers. Given are the number and the name of the chemicals, the corresponding Chemical Abstracts Service (CAS) registration number, and the codes of the blinded samples which were independently generated for each of the participating laboratories by ZEBET.

No.	Chemical Name	CAS Number	GHS <i>in vivo</i>	Code numbers		
				Henkel	Univers. Frankfurt	VITO
1	methyl stearate	112-61-8	No Cat.	A242	B165	C050
2	di-n-propyl disulphide	629-19-6	Cat. 2	A169	B007	C123
3	hexyl salicylate	6259-76-3	No Cat.	A252	B167	C116
4	heptanal	111-71-7	Cat. 2	A243	B045	C114
5	diethyl phthalate	84-66-2	No Cat.	A063	B003	C106
6	heptyl butyrate	5870-93-9	No Cat.	A039	B180	C177
7	1-bromohexane	111-25-1	Cat. 2	A295	B236	C251
8	naphthalene acetic acid	86-87-3	No Cat.	A006	B119	C229
9	allyl phenoxy-acetate	7493-74-5	No Cat.	A222	B238	C028
10	4-methyl-thio-benzaldehyde	3446-89-7	No Cat.	A047	B008	C005
11	1-bromo-4-chlorobutane	6940-78-9	No Cat.	A171	B027	C193
12	isopropanol	67-63-0	No Cat.	A136	B234	C264
13	1-decanol	112-30-1	Cat. 2	A099	B239	C049
14	potassium hydroxide (5% aq.)	1310-58-3	Cat. 2	A187	B118	C031
15	1-methyl-3-phenylpiperazine	5271-27-2	Cat. 2	A122	B200	C011
16	2-chloromethyl-4-methoxy-3,5-dimethylpyridine HCl	86604-75-3	Cat. 2	A276	B096	C227
17	cinnamaldehyde	104-55-2	No Cat.	A284	B223	C250
18	2-methyl-5-tert-butylthiophenol	7340-90-1	Cat. 2	A121	B078	C030
19	tetrachloroethylene	127-18-4	Cat. 2	A120	B248	C014
20	cyclamen aldehyde	103-95-7	Cat. 2	A210	B032	C257

4.1.1. Substance application

Prior to the substance testing, all tissue models were transferred into fresh prewarmed EpiLife® ALI medium. The models were then topically exposed to the reference chemicals (see Table 1) for 35 min each. For liquids, 25 µL were dispensed directly atop the tissue surface using a pipette without touching the surface. If the liquid did not spread evenly over the surface in the first instance, it was gently taken up and dispensed again with the pipette until spreading was achieved. For solids, an equivalent of 25 µL (approx. 25 mg) was distributed evenly over the tissue surface with a sharp application spoon. Then 25 µL of sterile D-PBS was added in order to wet the surface of the test material and to increase surface contact. In every test run each substance was applied to 3 tissue models.

4.1.2. Controls

As negative control (NC) 25 µL of D-PBS was topically applied to the tissue models and tested concurrently with the test chemicals. 25 µL of a 5% aqueous SDS solution was used as positive control (PC). All chemicals

were tested for their ability to reduce MTT in the absence of living cells. No MTT reducer was observed within the set of reference chemicals.

4.1.3. Tissue viability assay

After the 35 min exposure period at room temperature the tissues were thoroughly washed 8 times with 600 µL D-PBS each, dipped 5 times into 500 mL of D-PBS, and transferred to 6-well plates with 1.5 mL of fresh prewarmed EpiLife® ALI medium. After 42 h of post-exposure incubation at 37 °C/5% CO₂ the tissues were transferred to 24-well plates containing 0.2 mL of fresh MTT solution (1 mg/mL in D-PBS), where they remained for another 3 h. The blue formazan dye, caused by intracellular metabolic MTT reduction, was extracted with 2 mL of 2-propanol each for 2 h at room temperature or overnight at 4 °C. The optical density of the extracted formazan solution was determined in a multiplate spectrophotometer at a single wavelength between 540 and 600 nm, depending on the filters available in the respective laboratories. The relative tissue viability was calculated for each tissue as the percentage of the mean optical density of the negative control, which was defined as 100% viability.

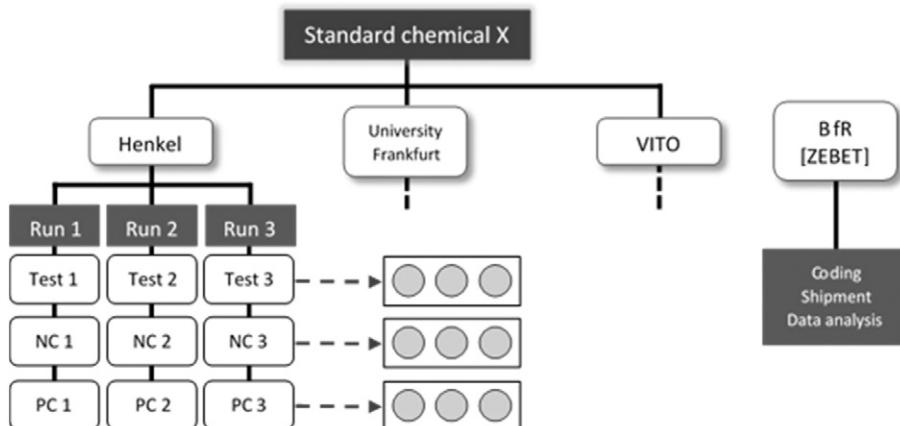


Fig. 1. Study design of the OS-REp SIT catch-up validation study according to the EUR-L-ECVAM Performance Standards (OECD, 2014). NC: negative control; PC: positive control; Test X: chemical to be tested. [Diagram adapted and amended from a previous version of the Performance Standards for in vitro skin irritation testing; EUR-L-ECVAM, 2009c].

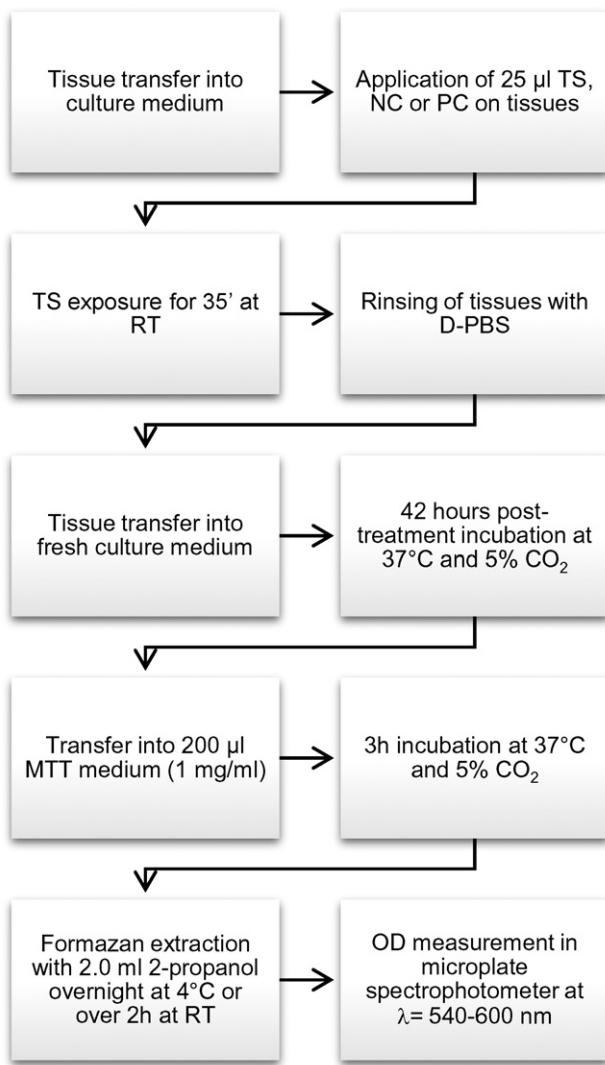


Fig. 2. Flowchart of the pivotal working steps in order to conduct the skin irritation assay. TS: test substance; NC: negative control; PC: positive control; RT: room temperature. The process is described in more detail in the respective Materials & methods section.

4.1.4. ET₅₀ determination

In order to get information about the barrier function of the OS-REp models, an ET₅₀ test was conducted with every OS-REp batch used in this study. Briefly, the tissues were topically treated with a 1% TRITON X-100 solution, and tissue viability was assessed with the MTT assay at defined time intervals after the start of the incubation. The ET₅₀ was calculated as the respective point of time where overall tissue viability had decreased to 50%.

4.1.5. Data analysis

An EXCEL spreadsheet was developed at ZEBET for data submission and analysis. It was password-protected and provided to the participating laboratories. At the end of the study the spreadsheets containing the test data were returned to ZEBET for final analysis. The analysis and graphical representation of the data were performed in the R statistical computing environment (R Development Core Team, 2010) in combination with the R lattice package (Sarkar, 2010).

Data analysis after cross-contamination tests and retesting were conducted at the Henkel laboratory with the respective spreadsheet. However, this analysis was conducted under unblinded conditions.

The mean ODs of the blanks were calculated and subsequently subtracted from the OD data of all other samples (OD_{TS}, OD_{NC}, OD_{PC}). For each sample the relative tissue viability in comparison to the NC was calculated using the following formulae:

$$\text{Relative viability TS}(\%) = [\text{OD}_{\text{TS}}/\text{mean of } \text{OD}_{\text{NC}}] \times 100$$

$$\text{Relative viability NC}(\%) = [\text{OD}_{\text{NC}}/\text{mean of } \text{OD}_{\text{NC}}] \times 100$$

$$\text{Relative viability PC}(\%) = [\text{OD}_{\text{PC}}/\text{mean of } \text{OD}_{\text{NC}}] \times 100$$

with TS: test substance, NC: negative control and PC: positive control.

For each TS, NC and PC the mean relative viability of the three individual tissues was calculated. These data were the basis for the subsequent substance classification.

4.1.6. Acceptance criteria

Based on historical data the mean optical density (OD) at $\lambda = 540-600$ nm for the NCs should be between 0.8 and 1.5 relative units. Mean viability of the PCs should be below 10% of the NC. According to the Performance Standards (OECD, 2015b) the standard deviations for triplicate measurements within a run must be equal or smaller than 18% ($SD \leq 18\%$) in order to become accepted as a valid run.

4.1.7. Prediction model

The prediction model (PM), i.e. the algorithm or rule for interpretation of the test data, is an integral part of any toxicological test method. In the current study, we used the PM developed with the validated reference method: a test substance is classified as a skin irritant (GHS category 2) if the mean relative tissue viability of the respective tissues is $\leq 50\%$ (OECD, 2015b). If the relative tissue viability is $> 50\%$, the chemical is considered a non-irritant (which may include "mild irritants, GHS category 3").

5. Physico-chemical analysis of cross contamination

Four OS-REp models were placed in 4 adjacent wells of a 6-well plate, where they were cultured with EpiLife® ALI medium. Two tissue models were topically treated with 25 µL of the volatile test substance each (donor models), whereas the surface of the other 2 models were covered with 25 µL of D-PBS (receptor models). In addition, two co-culture inserts, filled with Tenax-TA® 60/80 MESH (2,6-diphenylene-oxide polymer resin; Buchem BV, The Netherlands), an adsorbent powder, were placed in the 2 remaining wells of the 6-well plate (Fig. 3). After test chemical application the 6-well plates were covered by a lid which was removed only at the end of the respective exposure period or for the necessary working steps according to the protocol, respectively. The lid did not seal the wells individually, but allowed the diffusion of volatile substances into all wells of the respective 6-well plate. After 35 min of incubation with the chemicals, the donor models were thoroughly washed with D-PBS and then incubated for another 24 or 42 h under standard conditions. The D-PBS from the receptor models was carefully aspirated after 35 min of incubation, followed by a post-treatment incubation period of 42 h, too.

Tissues were collected after 0, 15 and 35 min of substance treatment, followed by sampling after 24 and 42 h of post-treatment incubation. Donor models were collected at 35 min only. At each indicated time point tissue models were removed from the 6-well plates, transferred to 20 mL glass vials and immediately frozen at -80°C in order to fix them until they were used for GC/MS analysis. In addition to the tissue models the respective culture media and the TENAX-filled inserts were also collected for analysis. For each of the indicated sampling time points an individual 6-well plate containing the respective number of tissue models and TENAX probes was cultured.

6. Gas chromatography–mass spectrometry (GC/MS)

Detection and quantification of chemicals in the samples was carried out with gas chromatography, followed by mass spectroscopy (GC/MS). Depending on the nature of the samples, different GC/MS processes were used (Fig. 3).

The culture medium was analyzed with the headspace GC/MS method on a Perkin Elmer Headspace “HS40” and Perkin Elmer GC “Autosystem XL” with an Agilent/J&W DB-5MS column (30 m, 0.25 mm internal diameter (ID), 1.00 μm film thickness (FT)).

To analyze the compounds trapped in the TENAX adsorbent powder thermal desorption GC/MS analysis was conducted on an Agilent GC 6890 and Agilent MSD 5973 Gerstel TDS2 (Thermal Desorption System), using an Agilent/J&W DB-5MS column (30 m, 0.25 mm ID, 1.00 μm FT).

The tissue model samples were analyzed with the dynamic headspace method, using the same Agilent system as above, but with a different Agilent/J&W DB-5MS column (25 m, 0.20 mm ID, 0.33 μm FT).

7. Histology

OS-REp models were cut out of the inserts with the insert membrane still adhering to the tissue, fixed in 4% buffered formaldehyde solution (Roti-Histofix, Roth, Germany), and embedded in paraffin. Sections of 5 μm thickness were subsequently stained with hematoxylin/eosin solution (H&E staining).

8. Results

8.1. Catch-up validation study

8.1.1. Morphological characteristics of the OS-REp model

All epidermal models consisted of 3–5 layers of keratinocytes, revealing the tissue-specific differentiation pattern (Fig. 4). Palisade-shaped basal keratinocytes closely adhered to the polycarbonate membrane, followed by cells of the s. spinosum and granulosum, respectively. At the apical surface a cornified layer had developed. The ET_{50} values, which provide an estimation about the barrier function of the OS-REp models, varied between 2.58 and 9.49 h at day 19 of air-liquid interface culture for all tissue batches constructed so far at the Henkel laboratory,

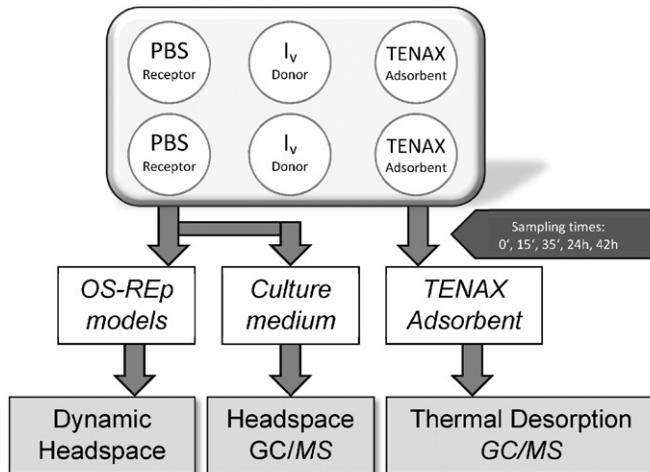


Fig. 3. Illustration of the experimental setup underlying the physico-chemical analysis of cross contamination due to volatile chemicals. OS-REp tissues incubated with the volatile skin-irritating chemical (I_v), with D-PBS only and co-culture inserts filled with the adsorbent powder (TENAX) were incubated together in a 6-well plate. After the indicated periods of time (gray arrow), tissue models, the respective culture media and the TENAX-filled inserts were sampled and frozen. The samples were subsequently subjected to gas chromatography, followed by mass spectroscopy (GC/MS) for physico-chemical analysis. For each sort of sample a different GC/MS process was used in order to achieve best resolution.

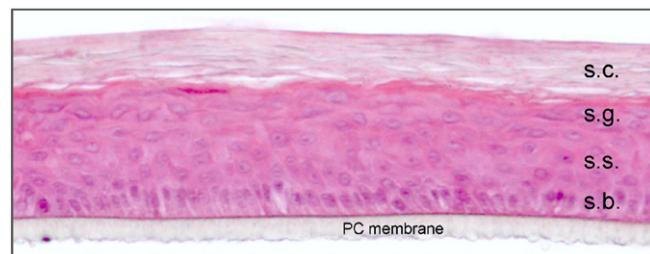


Fig. 4. Tissue architecture of the open source reconstructed epidermis at day 19 of air-liquid interface culture. The tissue model was formalin-fixed, embedded in paraffin, cut and stained with hematoxylin/eosin. s.b.: stratum basale; s.s.: stratum spinosum; s.g.: stratum granulosum; s.c.: stratum corneum. The tissue model was prepared for histology while still adhering to the polycarbonate (PC) insert membrane.

with a mean of 5.66 ± 1.66 h. Based on these historical data an ET_{50} value of 3.0 h was defined as minimum value for a batch to be qualified for phase I of the catch-up validation study. The ET_{50} values of the 3 batches employed in the validation study were 3.97, 5.97 and 5.96 h, those for the OS-REp models needed for the additional runs at VITO and Frankfurt University were 5.38, 4.65 and 3.38 h, respectively. Thus, all batches were qualified.

8.1.2. Test acceptance criterion for the negative control (NC)

Tissue viability of PBS-treated OS-REp models was determined with the MTT assay. In all 3 laboratories the mean optical density was confined to a range from 0.97 to 1.22 units (Fig. 5). All data fell within the acceptance range of 0.8–1.5 OD units which was defined in the SOP of the OS-REp SIT based on historical data collected during the developmental phase, and thus all runs clearly fulfilled this acceptance criterion.

8.1.3. Test acceptance criterion for the positive control (PC)

In each run 3 tissues were treated with a 5% aqueous SDS solution as a positive control. The relative tissue viability observed for all runs in all

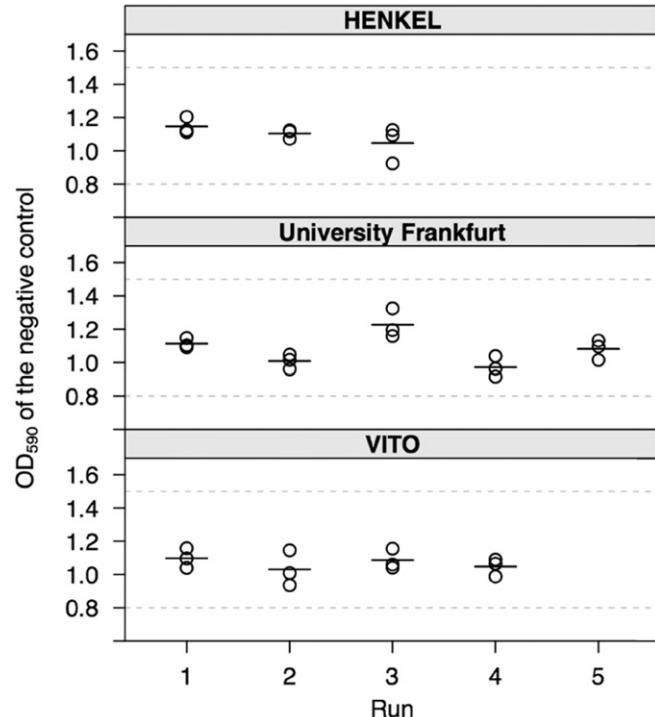


Fig. 5. Response of the negative controls (NC) in the three laboratories. Given is the background-corrected optical density (OD) for each NC tissue replicate in relation to the experimental run. Horizontal bars indicate the mean OD for each run. All mean OD values fell within the acceptance range of ≥ 0.8 and ≤ 1.5 (dashed lines).

3 laboratories were always below 2% of the negative control (Fig. 6). Thus all runs clearly matched the acceptance criterion of $PC < 10\%$, as defined in the respective SOP and adopted from the OECD Performance Standards.

8.1.4. Variability of the tests

In order to find out whether all OS-REp models concurrently tested with a substance reacted similarly, the standard deviations associated with the mean values of every single test were analyzed (Table 2). Only 9 out of 198 tests (i.e. 4.5%), which were obtained by the three laboratories in the first three runs, revealed standard deviations larger than 18% which is the defined acceptance cut-off value, possibly indicating towards defect tissue models or inappropriate dosing. Those 9 non-qualified tests were repeated in up to 2 additional runs. The SDs of the re-tests all fulfilled the acceptance criterion of $SD \leq 18\%$ in accordance with the OECD Performance Standards.

8.1.5. Within-laboratory reproducibility (WLR)

The reproducibility of the test performance within each laboratory was assessed by analyzing the WLR, or concordance, between the independent runs. A test result is considered concordant if tissue viabilities of the 3 valid runs are either all above or all below (or equal) 50% cut-off value and hence result in the same classification. If one run ends up in a deviating result, the test result is considered discordant. The tissue viabilities for all runs and all chemicals are presented in Table 3 and in Fig. 7.

In the Henkel laboratory, 17 out of 20 substance treatments were concordant, resulting in a WLR of 85%. Di-n-propyl disulphide, heptyl butyrate and 1-bromohexane led to discordant results. The Frankfurt laboratory reached a concordance of 90%, with di-n-propyl disulphide and 5% KOH solution leading to discordant results. At the VITO laboratory, only one discordant result was observed after isopropanol treatment, which equals 95% WLR.

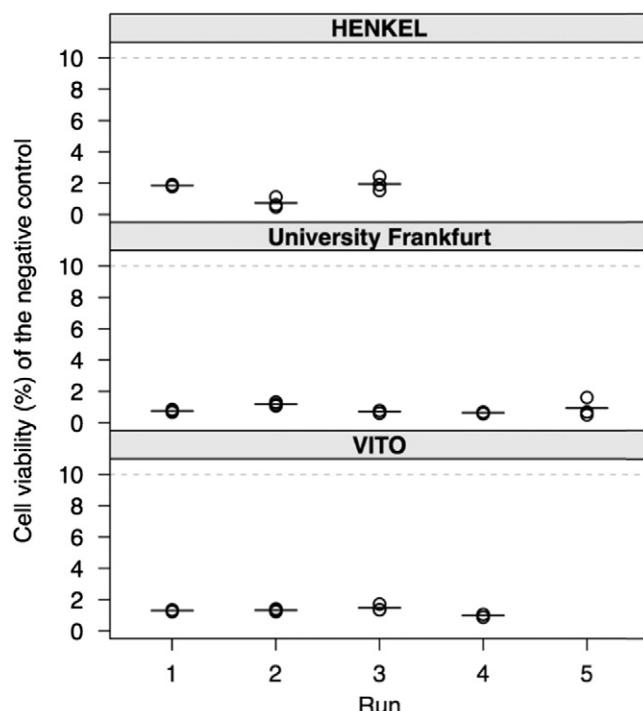


Fig. 6. Cell viability of the positive controls (PC) in the three laboratories. Given is the cell viability (normalized to negative control) for each PC tissue replicate in relation to the experimental run. Horizontal bars indicate the mean cell viability for each run. All mean viability values met the acceptance criterion of $< 10\%$ (dashed line).

8.1.6. Between-laboratory reproducibility (BLR)

The BLR was determined by comparing the classification results for every tested substance between all 3 laboratories. The final classification for a chemical in one laboratory was obtained by calculating the mean value of relative tissue viability for the 3 valid runs. Fourteen out of 20 substance treatments gave concordant classifications (Table 4). Methyl stearate, diethyl phthalate, heptyl butyrate, 1-bromohexane, isopropanol, and the 5% aqueous solution of potassium hydroxide led to discordant classifications. This study resulted in a 70% BLR, which is below the defined acceptance criterion of 80%.

8.1.7. Predictive capacity

In order to determine the predictive capacity of the OS-REp-based skin irritation test, the in vitro classification was compared with the in vivo classification according to the GHS system (Table 5). This was done with contingency tables for every laboratory and for the overall classification. Sensitivity was best at Henkel and VITO with 100%, meaning that all skin-irritating cat. 2 chemicals were predicted correctly. The Frankfurt laboratory achieved 80% sensitivity, misclassifying 1-bromohexane and 5% KOH solution as non-irritants. With an overall sensitivity of 93% the study met the acceptance criterion of $\geq 80\%$. All laboratories correctly classified 1-decanol, a so-called borderline chemical (in vivo score 2.3) and di-n-propyl disulphide, a chemical which gave false-negative results in the validated reference method (VRM).

Specificity was best at the Frankfurt laboratory with 70%, whereas Henkel and VITO only achieved 50%. Although Frankfurt met the acceptance criterion of 70% specificity, with 57% the overall specificity across the three laboratories was too low, strongly suggesting additional investigations on the false positive chemicals as described below.

The accuracy of 75%, for the single laboratories as well as for the whole study, met the acceptance criterion of 75% as defined in the OECD PS (OECD, 2015b).

8.2. Cross contamination

8.2.1. Re-testing of the falsely predicted chemicals

When analyzing the results of the validation study in more detail, it became evident that in each laboratory different chemicals were misclassified: methyl stearate and heptyl butyrate at Henkel, diethyl phthalate and isopropanol at VITO, and 1-bromohexane and 5% KOH at Frankfurt, respectively (see above). In order to find out whether the misclassification was based on general limitations of the OS-REp SIT, on mistakes during test performance, or on lab-specific effects, the 6 chemicals mentioned above were retested in all 3 laboratories. A new batch of OS-REp models produced at Henkel, which originated from the same keratinocyte batch as in the validation runs and matched the quality standards, was used.

Each lab performed one additional skin irritation test with its own 2 misclassified chemicals including negative and positive controls similar to the original validation runs on triplicate tissues for every test item. In addition, VITO and Frankfurt each sent aliquots of their respective misclassified chemicals to Henkel, where they were tested on non-shipped OS-REp models with Henkel's own chemical samples as a benchmark (Fig. 8).

Henkel as well as VITO now correctly classified their own 2 chemicals, respectively. In addition, diethyl phthalate and isopropanol samples from VITO were classified correctly, too, when tested at the Henkel lab. The same results were achieved with the Henkel's own diethyl phthalate and isopropanol samples. All 4 chemicals are non-irritating according to the Draize test, and in all cases the OS-REp models treated with these substances retained high viabilities with $SDs < 18\%$.

The brominated alkane 1-bromohexane was classified as a non-irritant at the Frankfurt and the Henkel laboratory, irrespective of its in vivo score of 2.7 (irritant; OECD, 2015b). The strongly irritating substance 5% KOH was correctly predicted at the Henkel lab with its own sample as well as with the sample delivered from the Frankfurt lab. However,

Table 2

Standard deviations (SD) of the normalized tissue viability [%] for all substance exposures and all runs in the 3 laboratories. Included are the SD values for the negative (NC) and positive (PC) controls. SD values which did not meet the PS acceptance criterion of $\leq 18\%$ are highlighted in gray.

Test substance		Henkel			University Frankfurt					VITO			
No.	Name	Run 1	Run 2	Run 3	Run 1	Run 2	Run 3	Run 4	Run 5	Run 1	Run 2	Run 3	Run 4
NC	PBS	4.4	2.5	10.3	2.7	4.3	7.1	6.4	5.5	5.4	10.3	5.7	5.1
PC	SDS (5% aq.)	0.1	0.3	0.4	0.1	0.1	0.1	0.1	0.6	0.1	0.1	0.2	0.1
1	Methyl stearate	11.6	14.7	16.6	8.6	2.0	4.9	—	—	11.9	3.3	9.9	—
2	Di-n-propyl disulphide	9.3	7.9	14.7	17.6	30.7	22.4	3.4	0.3	7.7	2.5	3.7	—
3	Hexyl salicylate	1.5	3.2	8.5	4.6	5.2	9.5	—	—	14.6	10.0	5.0	—
4	Heptanal	1.0	0.6	0.5	0.4	0.3	0.6	—	—	0.4	0.2	0.4	—
5	Diethyl phthalate	11.8	5.4	6.7	9.5	14.8	8.1	—	—	1.7	0.4	0.3	—
6	Heptyl butyrate	11.8	15.2	16.8	15.5	3.8	13.9	—	—	6.5	7.1	13.0	—
7	1-Bromohexane	8.1	4.2	8.6	10.0	25.1	23.9	7.6	4.2	27.0	1.0	4.8	2.6
8	Naphthalene acetic acid	4.8	5.9	4.5	7.3	19.9	4.2	2.3	—	15.3	8.6	9.3	—
9	Allyl phenoxy-acetate	8.2	13.1	7.5	10.9	7.0	10.6	—	—	8.7	7.1	12.8	—
10	4-Methyl-thio-benzaldehyde	0.3	4.3	8.4	15.6	12.1	0.8	—	—	0.5	5.8	4.1	—
11	1-Bromo-4-chlorobutane	4.6	0.5	10.0	1.2	2.3	37.3	—	5.3	1.1	3.3	1.7	—
12	Isopropanol	1.9	7.7	6.2	2.2	5.9	2.4	—	—	23.9	7.3	4.0	2.6
13	1-Decanol	0.5	0.9	0.8	5.1	0.1	0.4	—	—	7.5	0.4	0.2	—
14	Potassium hydroxide (5% aq.)	0.4	4.8	0.1	11.1	11.7	18.5	—	0.5	0.2	0.7	1.0	—
15	1-Methyl-3-phenylpiperazine	0.5	1.1	0.3	2.5	2.9	2.8	—	—	0.2	0.8	0.5	—
16	2-Chloromethyl-4-methoxy-3,5-dimethylpyridine HCl	0.5	0.9	0.5	0.3	0.1	0.3	—	—	0.3	0.2	0.1	—
17	Cinnamaldehyde	0.2	0.1	0.2	0.2	0.5	0.3	—	—	0.5	0.4	0.3	—
18	2-Methyl-5-tert-butylthiophenol	1.8	1.1	0.9	1.4	0.8	0.7	—	—	1.7	2.5	2.8	—
19	Tetrachloroethylene	1.2	0.0	5.2	3.9	11.5	4.0	—	—	7.3	1.3	5.5	—
20	Cyclamen aldehyde	0.0	0.1	0.1	0.3	0.2	0.3	—	—	0.1	0.1	0.4	—

this substance was again misclassified as a non-irritant at the Frankfurt lab (81% rel. viability).

8.2.2. Biological analysis of cross contamination

These results, together with a thorough investigation of the experimental conditions during the validation runs provided evidence that the test results in the study were negatively influenced by the volatile fractions of skin-irritating substances, especially heptanal and cyclamen aldehyde, which had been tested in parallel with a non-irritating

chemical at the same time in the same culture plate. Although the vapor pressure of the concerned irritating chemicals are quite low it was concluded that under the given experimental conditions small amounts of the substances evaporated from the surface of the tissues and reached the tissues in the neighbouring wells of the same plates by diffusion via the air phase (cross contamination). It seemed that even low concentrations of said chemicals may be sufficient to damage the tissue equivalents irreversibly, resulting in false positive predictions for the non-irritating chemicals tested in parallel. Each laboratory faced

Table 3

Tissue viability (% of NC) of the individual runs at the 3 laboratories. Gray highlighted values indicate non-valid runs with an SD $\leq 18\%$. Additionally the mean tissue viability of the 3 valid runs is revealed.

Test substance		Henkel			University Frankfurt					VITO						
No.	Name	Run 1	Run 2	Run 3	Mean (3 valid runs)	Run 1	Run 2	Run 3	Run 4	Run 5	Mean (3 valid runs)	Run 1	Run 2	Run 3	Run 4	Mean (3 valid runs)
1	Methyl stearate	41.8	43.1	45.0	43.3	98.2	107.5	98.0	—	—	101.2	106.5	92.9	80.5	—	93.3
2	Di-n-propyl disulphide	52.2	28.6	57.6	46.2	22.3	52.1	28.7	98.8	12.7	44.6	26.1	8.8	6.5	—	13.8
3	Hexyl salicylate	94.9	91.5	100.3	95.6	100.2	103.8	90.8	—	—	98.3	117.1	110.3	88.7	—	105.3
4	Heptanal	2.1	2.3	3.7	2.7	2.2	2.9	2.6	—	—	2.6	2.7	3.0	2.8	—	2.8
5	Diethyl phthalate	79.4	80.6	95.3	85.1	95.2	104.0	89.2	—	—	96.1	2.3	2.0	2.4	—	2.2
6	Heptyl butyrate	43.8	61.3	38.8	48.0	73.0	84.0	71.1	—	—	76.1	96.6	84.6	98.0	—	93.1
7	1-Bromohexane	67.2	13.9	25.4	35.5	71.6	80.5	46.9	75.8	86.2	77.9	68.8	6.9	10.4	27.9	15.1
8	Naphthalene acetic acid	94.0	89.1	108.0	97.0	97.8	97.5	92.5	98.5	—	96.3	109.7	119.4	94.4	—	107.8
9	Allyl phenoxy-acetate	52.0	57.5	61.4	57.0	89.3	99.7	92.1	—	—	93.7	83.2	104.2	75.3	—	87.6
10	4-Methyl-thio-benzaldehyde	1.7	11.8	15.6	9.7	16.2	22.4	3.2	—	—	13.9	4.0	15.2	8.9	—	9.4
11	1-Bromo-4-chlorobutane	7.1	0.8	10.1	6.0	12.4	22.7	32.2	—	7.3	14.1	2.5	4.6	3.5	—	3.6
12	Isopropanol	100.2	99.2	102.8	100.8	87.4	103.7	93.7	—	—	94.9	46.9	28.8	27.0	88.2	48.0
13	1-Decanol	1.7	1.9	2.9	2.2	4.4	1.7	1.8	—	—	2.6	6.3	2.4	1.9	—	3.5
14	Potassium hydroxide (5% aq.)	1.3	9.0	1.7	4.0	57.1	49.0	37.6	—	68.7	58.3	1.3	2.1	1.8	—	1.8
15	1-Methyl-3-phenylpiperazine	2.7	2.2	2.9	2.6	3.4	3.3	2.5	—	—	3.1	1.1	2.1	1.1	—	1.5
16	2-Chloromethyl-4-methoxy-3,5-dimethylpyridine HCl	1.4	1.0	2.0	1.5	0.7	1.2	0.8	—	—	0.9	0.9	1.3	1.1	—	1.1
17	Cinnamaldehyde	1.2	0.1	2.8	1.4	1.8	2.9	2.0	—	—	2.2	2.4	3.7	2.7	—	2.9
18	2-Methyl-5-tert-butylthiophenol	10.9	4.5	4.4	6.6	4.5	4.0	3.9	—	—	4.1	6.3	11.2	10.0	—	9.2
19	Tetrachloroethylene	2.1	1.1	27.9	10.4	40.6	19.8	6.5	—	—	22.3	37.2	3.5	5.2	—	15.3
20	Cyclamen aldehyde	1.1	0.3	2.7	1.4	1.0	2.0	1.1	—	—	1.4	1.9	2.3	2.5	—	2.3

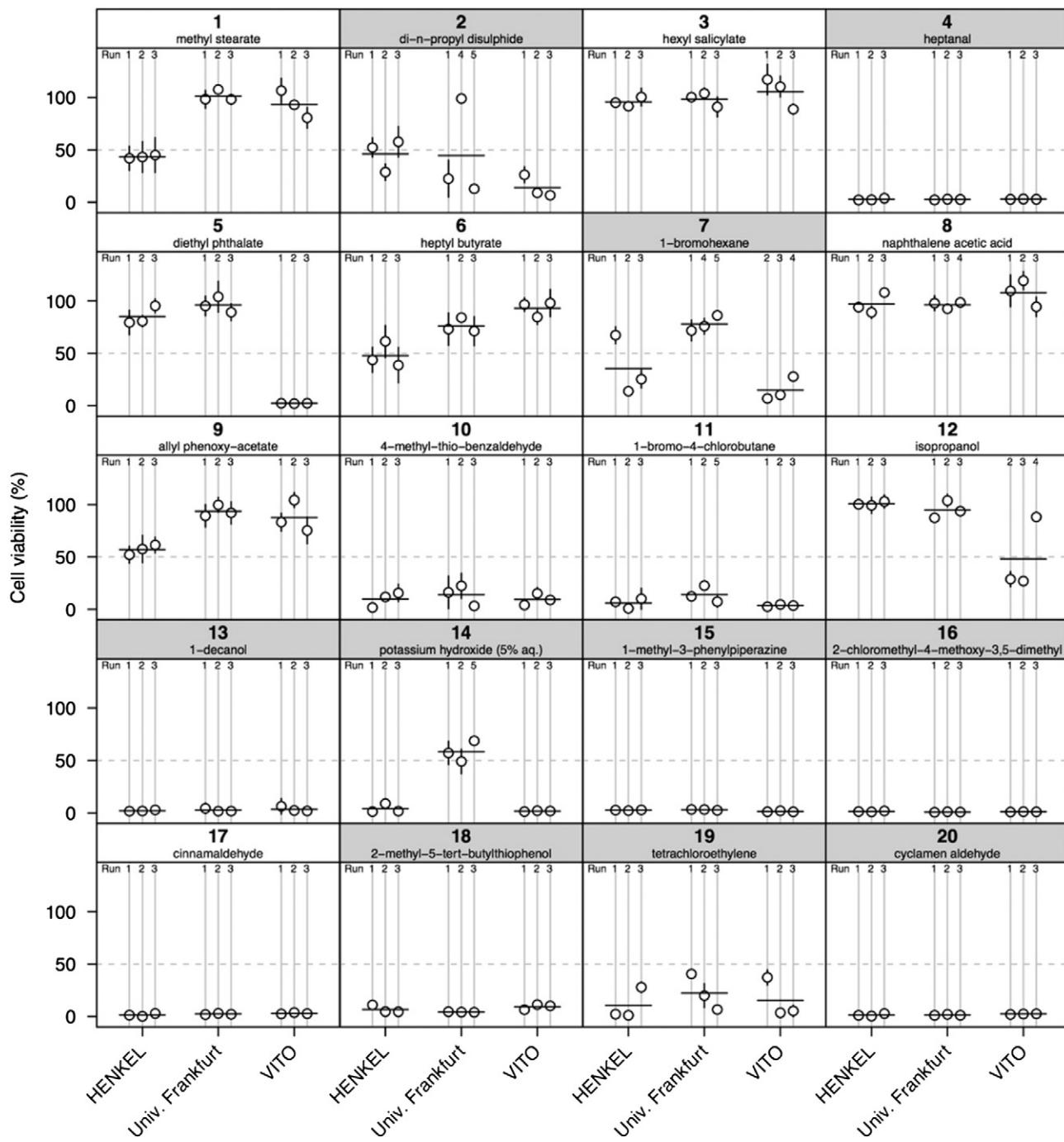


Fig. 7. Multipanel conditional display of the cell viability. The mean cell viability (circle symbols) of each valid run is plotted against the laboratory for all 20 test substances. Vertical solid lines refer to the three valid runs. The horizontal dashed line at 50% viability represents the cut-off value for the prediction of non-irritant No-Category substances (>50%) and irritant Category-2 substances ($\leq 50\%$). The in vivo classification of the substances is indicated by strip backgrounds in white (non-irritants) and gray (irritants).

other misclassified substances because the test chemicals were coded differently for each laboratory and thus were tested in different combinations.

In order to support the above outlined hypothesis of cross contamination due to volatile irritating chemicals, 2 additional analyses, one looking again at tissue viability, and one based on gas chromatographical methods, were performed. A biological approach (viability assay) was conducted in 2 independent experiments with a new batch of OS-REp models at the Henkel laboratory, which recapitulated the experimental conditions at the beginning of the catch-up validation study. OS-REp models treated with the 4 misclassified non-irritating

chemicals were confronted with models treated with the respective volatile irritating substances in the same culture plate, followed by viability assessment. Also models were treated with the misclassified non-irritants in individual culture plates only and subsequently analyzed in respect of tissue viability.

Treating the tissue models with either heptanal or cyclamen aldehyde for the indicated period of time reduced its viabilities to values of $<2\%$ of the negative control, indicating towards completely killed tissues (Fig. 9). The viabilities of tissues treated with methylstearate and diethylphthalate clearly depended on the culture conditions. When the tissues were cultivated in one 6-well plate for a single substance

Table 4

Predictive capacity of the OS-REp model. Given is the (final) classification of each test substance for the laboratories HENKEL, University Frankfurt, and VITO. Additionally given is the in vivo classification according to the UN GHS system as well as the in vitro classification of the Validated Reference Method (VRM; Spielmann et al., 2007; OECD, 2015a). NI: non-irritant (No Category); Cat. 2: irritant (Category 2). The last column indicates the concordance between the 3 laboratories (yes: concordant results, no: discordant results). White fields: classification as no category; gray fields: classification as category 2.

No.	Chemical Name	GHS in vivo	Henkel	Univers. Frankfurt	VITO	VRM in vitro	Concordance
1	Methyl stearate	No Cat.	Cat. 2	NI	NI	NI	no
2	Di-n-propyl disulphide	Cat. 2	Cat. 2	Cat. 2	Cat. 2	NI	yes
3	Hexyl salicylate	No Cat.	NI	NI	NI	NI	yes
4	Heptanal	Cat. 2	Cat. 2	Cat. 2	Cat. 2	Cat. 2	yes
5	Diethyl phthalate	No Cat.	NI	NI	Cat. 2	NI	no
6	Heptyl butyrate	No Cat.	Cat. 2	NI	NI	NI	no
7	1-Bromohexane	Cat. 2	Cat. 2	NI	Cat. 2	Cat. 2	no
8	Naphthalene acetic acid	No Cat.	NI	NI	NI	NI	yes
9	Allyl phenoxy-acetate	No Cat.	NI	NI	NI	NI	yes
10	4-Methyl-thio-benzaldehyde	No Cat.	Cat. 2	Cat. 2	Cat. 2	Cat. 2	yes
11	1-Bromo-4-chlorobutane	No Cat.	Cat. 2	Cat. 2	Cat. 2	Cat. 2	yes
12	Isopropanol	No Cat.	NI	NI	Cat. 2	NI	no
13	1-Decanol	Cat. 2	Cat. 2	Cat. 2	Cat. 2	Cat. 2	yes
14	Potassium hydroxide (5% aq.)	Cat. 2	Cat. 2	NI	Cat. 2	Cat. 2	no
15	1-Methyl-3-phenylpiperazine	Cat. 2	Cat. 2	Cat. 2	Cat. 2	Cat. 2	yes
16	2-Chloromethyl-4-methoxy-3,5-dimethylpyridine HCl	Cat. 2	Cat. 2	Cat. 2	Cat. 2	Cat. 2	yes
17	Cinnamaldehyde	No Cat.	Cat. 2	Cat. 2	Cat. 2	Cat. 2	yes
18	2-Methyl-5-tert-butylthiophenol	Cat. 2	Cat. 2	Cat. 2	Cat. 2	Cat. 2	yes
19	Tetrachloroethylene	Cat. 2	Cat. 2	Cat. 2	Cat. 2	Cat. 2	yes
20	Cyclamen aldehyde	Cat. 2	Cat. 2	Cat. 2	Cat. 2	Cat. 2	yes

only, the mean viabilities were around 100% of the negative control for both substances. According to the underlying prediction model these chemicals were unambiguously classified as non-irritants.

However, the situation changed when the models treated with these non-irritating chemicals were cultivated together with heptanal-

treated models in the same 6-well plate. The mean viability of the methylstearate-treated models decreased to 48% of NC, tissue viability of the diethylphthalate-treated models even dropped to 17% of NC only.

The mean viabilities of the tissues covered with heptylbutyrate and isopropanol and cultured in individual 6-well plates reached 72% and 100% of NC, respectively. Thus both were correctly classified as non-irritants. Cultivating tissue models treated with heptylbutyrate and isopropanol in a 6-well plate together with cyclamen aldehyde-treated models elicited no visible effect in this study.

Table 5

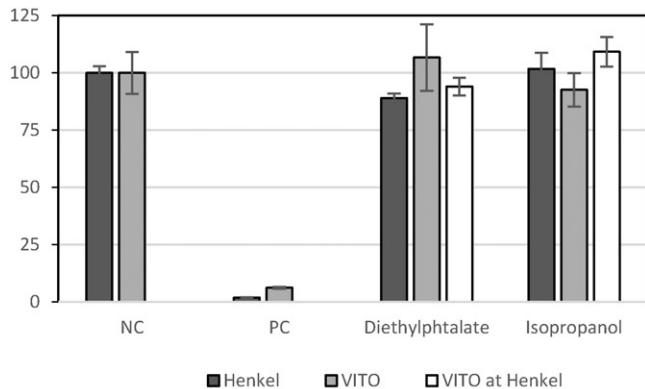
Summary of the laboratory-specific and the overall predictive capacity of the OS-REp SIT according to the GHS classification system, based on 3 valid test runs each. The fields highlighted in light gray indicate those parameters which did not match the OECD 439 Performance Standards (OECD, 2015b).

Test laboratory	Henkel	University Frankfurt	VITO	All
Sensitivity [%]	100	80	100	93
Specificity [%]	50	70	50	57
Accuracy [%]	75	75	75	75
No. of chemicals	20	20	20	60

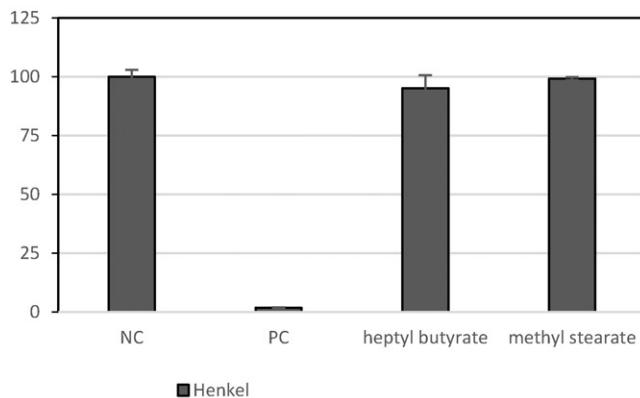
8.2.3. Physico-chemical analysis of cross contamination

A physico-chemical analytical approach was designed in order to detect small quantities of heptanal and cyclamen aldehyde in affected neighbouring OS-REp models. The method relied on the dynamic head-space analysis, followed by a coupled GC/MS detection, which is able to detect small organic molecules.

Retesting of chemicals falsely classified at VITO



Retesting of chemicals falsely classified at Henkel



Retesting of chemicals falsely classified at Frankfurt

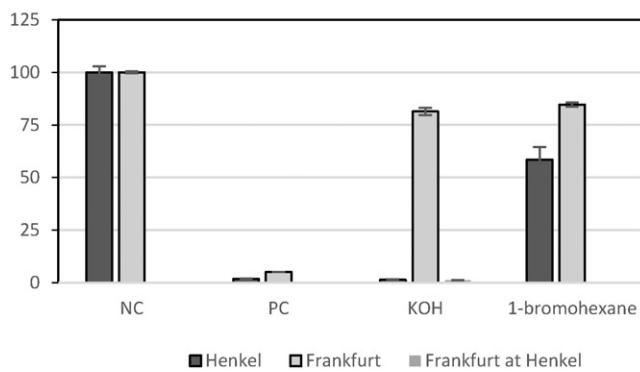


Fig. 8. Results of the repeated skin irritation tests with the falsely classified chemicals after protocol optimization. Relative viabilities of the OS-REp models topically treated with the substances indicated in the diagrams. The tissue models were incubated with 25 µL of substance each for 35 min and post-incubated for 42 h at 37 °C. Viability was determined with the MTT assay. The PBS-treated negative controls (NC) are set as 100% viability, all other data are calculated relative to the NC. Positive controls: tissues treated with 5% SDS solution. Tests were performed either with the Henkel reference chemicals (RC) at Henkel, the VITO and Frankfurt RC at the respective labs (VITO, Frankfurt) or the VITO and Frankfurt RC at the Henkel lab (VITO or Frankfurt at Henkel, respectively).

8.3. Heptanal

Heptanal accumulated in the D-PBS-treated receptor OS-REp models from not detectable amounts at the beginning of the exposure period up to 90 and 120 ng, respectively, measured after 35 min. The highest accumulation was seen in the tissues 24 h after the test substance had been

washed away from the treated models, reaching values of >200 ng/tissue. At the end of the post-treatment incubation period, after 42 h, the amount of heptanal had decreased to values of 70 and 140 ng/tissue, respectively; Table 6).

Interestingly, parallel to the accumulation of heptanal, also heptanol (CAS-No. 111-70-6), the reduced form of heptanal, was identified in the untreated samples, exceeding the amount per tissue found for the original aldehyde. Even after 35 min of incubation values were found with peak heights beyond the detector limit (200 ng), and according to the peak width heptanol enrichment proceeded throughout the whole culture period.

Heptanal rapidly accumulated in TENAX, an adsorbent powder which had been placed in the 6-well plate, too. Only after 15 min of tissue exposure heptanal levels exceeded the upper limit of the detection system, and after 24 h post-treatment incubation its amount could be assumed to be in the microgram range. Heptanol was absent in TENAX. No heptanol could be detected in the culture medium from underneath the tissue models.

8.4. Cyclamen aldehyde (CyA)

Cyclamen aldehyde in the D-PBS-treated receptor OS-REp models was detected first after 35 min of exposure, reaching values of 3 and 30 ng/tissue, respectively. Throughout the post-treatment incubation period the aldehyde level in the tissue remained constant at a quite low level (3–7 ng/tissue). Already after 15 min of cultivation cyclamen alcohol (CAS-No. 4756-19-8), the reduced form of cyclamen aldehyde, could be detected in the receptor models, reaching levels between 80 and >200 ng/tissue, respectively. Alcohol amounts of >200 ng/tissue were observed after 35 min and, after removal of the remaining aldehyde on the treated models, during the whole post-treatment incubation period. Cyclamen aldehyde also accumulated in TENAX, which had been placed in the 6-well plate. After 35 min of tissue exposure the aldehyde reached levels of 50 and 100 ng/sample, respectively. Throughout the whole post-treatment incubation period the amount of CyA in the adsorbent powder could be assumed to be in the microgram range, according to the peak width. The reduced form, cyclamen alcohol, was not found in TENAX. No cyclamen aldehyde could be detected in the culture medium, which had been sampled from underneath the tissue models.

8.4.1. Recalculation of predictive capacity

The biological analysis revealed that those 4 chemicals which had been classified false-positive in the original test series were predicted correctly as non-irritants when they were tested without any possible interference with volatile skin-irritating compounds, confirming that the misclassification was not inherent to the OS-REp models, but to an experimental detail. Since apart from separating the tissues during incubation to avoid cross-contamination, no procedural detail of the testing protocol had been changed in this special additional re-testing exercise of a sub-set of chemicals, we decided to merge the supplementary data for diethylphthalate, methylstearate, heptylbutyrate and isopropanol with the original data and to recalculate the predictive performance parameters for the OS-REp SIT catch-up validation study. While the overall sensitivity remained unchanged at 93% (80/100/100%), the overall specificity increased to 70% (70/70/70%) as required by the OECD PS (OECD, 2015b), and the accuracy increased to 82% (75/85/85%). The recalculated predictive parameters are presented in Table 7.

9. Discussion

The catch-up validation study presented here was conducted according to the rules depicted in the OECD Performance Standards (2015b). These standards provide a guideline for developers of me-too methods for in vitro skin irritation testing based on human reconstructed epidermal models. Derived from the Validated Reference Methods

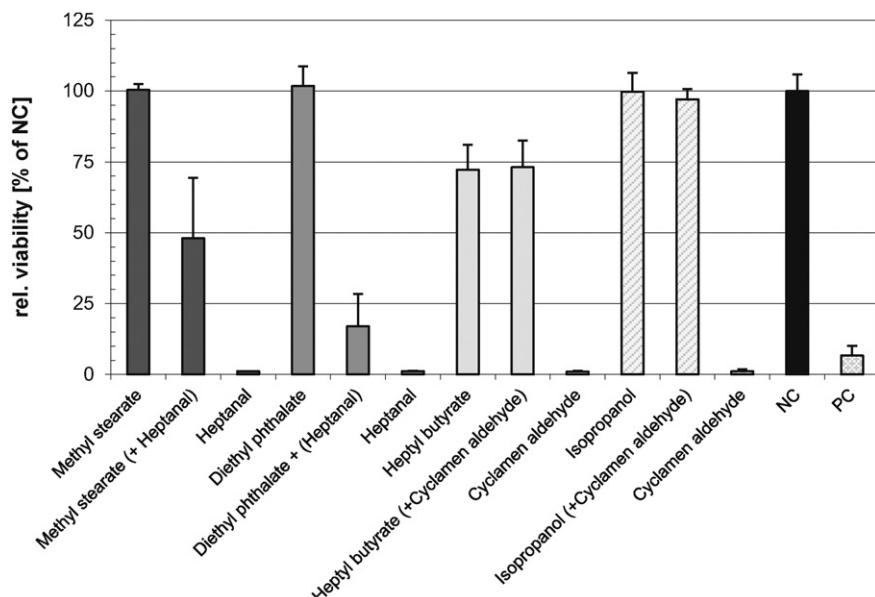


Fig. 9. Results of the biological cross contamination analysis. OS-REp models treated with a volatile irritant and a non-irritant (with the potentially interfering irritant indicated in brackets), respectively, were cultured together in a 6-well plate. OS-REp models treated with the non-irritating chemicals only were cultured in separate plates. The columns represent the mean relative viabilities of 3 OS-REp models \pm SD, expressed as % of the negative control.

(Spielmann et al., 2007) they include information about the essential test method components, a minimum list of reference chemicals to be tested, and defined reliability and accuracy values.

For the OS-REp model production the protocol for epidermal model culture, published first by Poumay et al. (2004), was modified in order to enhance tissue quality and lifespan of the tissue equivalents. Instead of adult skin, juvenile foreskin was used for keratinocyte isolation, the culture medium was supplemented with keratinocyte growth factor (KGF), a member of the fibroblast growth factor family, and the period of time in which the tissues were cultured at the air-liquid interface was extended up to 19 days before they are used in the skin irritation test.

KGF supplementation resulted in increased tissue thickness due to a higher number of viable keratinocyte layers compared to KGF-free culture conditions. Thus the OS-REp architecture resembled more closely that of native human skin. KGF has been shown to modulate proliferation and differentiation of keratinocytes in a 3D human skin model grown on decellularized dermis (Andreadis et al., 2001) and also to enhance migration and hyaluronan synthesis in rat epidermal keratinocytes (Karvinen et al., 2003). In both studies delayed or even inhibited expression of cytokeratin 10 (CK10), an early marker of epithelial differentiation (Stark et al., 1999), was observed. Therefore delayed CK10 expression is a clear indicator for a prolonged proliferative state on the one hand

and for a delayed or compromised terminal differentiation of the keratinocytes on the other hand, which coincides with increased tissue thickness and prolonged lifespan of the OS-REp model. KGF also exhibited growth-stimulating effects in reconstituted human oral epithelium by increasing the rate of proliferating keratinocytes without influencing terminal differentiation (Costea et al., 2003).

The OECD Performance Standards (2015b) define the experimental set-up which was successfully applied in the validated reference methods with the EpiSkin™ and the EpiDerm™ models as a blueprint for the development of any me-too method for in vitro skin irritation testing. However, a certain degree of freedom in defining the range of values for different quality-related parameters is foreseen in the PS, taking into account the specific inherent morphological and physiological properties of reconstructed epidermal models of different origin. For the OS-REp SIT the exposure time and temperature as well as the washing procedure was adapted in order to maximize reliability and predictivity of our test system.

The OS-REp SIT phase I catch-up validation study, simultaneously conducted in 3 independent laboratories with 3 batches of OS-REp models produced independently at Henkel, fulfilled the OECD study acceptance criteria, meaning that for every test chemical and every laboratory complete run sequences were generated, which allowed a sound data analysis.

In the first instance, not all reliability and accuracy values were met. One out of 3 laboratories missed the $\geq 90\%$ threshold for within-laboratory reproducibility, and the between-laboratory reproducibility of 70%

Table 6

Results of the physico-chemical analysis of cross contamination. The numbers represent the amounts of chemicals identified in each sample: OS-REp + chemical: concentration of indicated chemical in untreated tissue models; TENAX + chemical: concentration of indicated chemical in TENAX adsorber powder (ng/sample). CyA/OH: cyclamen aldehyde/alcohol. >>>: beyond the upper detection limit of the GC/MS devices.

Time	OS-REp	OS-REp	TENAX	OS-REp	OS-REp	TENAX
	Heptanal	Heptanal	Heptanal	CyA	CyOH	CyA
0'	<1	<1	<1	<1	<1	<1
15'	4	15	3	<1	<1	3
15'	4	40	>> 200	<1	<1	2
35'	90	>200	>200	3	80	50
35'	120	>200	>200	30	>200	100
24 h	200	>> 200	>> 200	7	>200	>> 200
24 h	>200	>> 200	>> 200	3	>200	>> 200
42 h	140	>> 200	>> 200	5	>200	>> 200
42 h	70	>> 200	>> 200	4	>200	>> 200

Table 7

Summary of the laboratory-specific and the overall predictive capacity of the OS-REp SIT according to the GHS classification system. The values have been recalculated on the basis of the results which were achieved after having solved the cross-contamination issue due to volatile irritating substances within the reference chemicals panel. All parameters matched or exceeded the OECD 439 Performance Standards (OECD, 2015b).

Test laboratory	Henkel	University Frankfurt	VITO	All
Sensitivity [%]	100	80	100	93
Specificity [%]	70	70	70	70
Accuracy [%]	85	75	85	82
No. of chemicals	20	20	20	60

was clearly below the acceptance criterion of $\geq 80\%$. While sensitivity always fulfilled the acceptance criterion of 80%, in two laboratories only 50% specificity was achieved and thus failed expectations. Interestingly, each of the three laboratories contributed exactly 2 different chemicals to the sum of 6 discordantly predicted substances, which included 4 non-irritating, but falsely positive classified chemicals, and 2 skin-irritating agents, which were classified as non-irritants. This observation eventually resulted in a series of additional experiments and analyses discussed in the chapter below.

With an overall sensitivity of 93% the PS acceptance criterion of $\geq 80\%$ was clearly met. Only the misclassification of 2 chemicals in one laboratory prevented a 100% correct prediction of the skin-irritating chemicals. It should be mentioned that all laboratories correctly classified 1-decanol (a borderline chemical) and di-*n*-propyl disulphide, a false negative of the validated reference method, which both were irritant in the rabbit test, but non-irritant in the human patch test (Baskettter et al., 2004, 2012; Jírová et al., 2010; EUR-ECVAM, 2009a, 2009b).

In order to achieve more insight into the possible reasons for the misclassifications of distinct chemicals in the study laboratories, the results were analyzed in more detail. Impaired tissue models could be excluded as a possible culprit as all OS-REp models used in this study were produced at Henkel, quality-checked, and then shipped to the study partners, making sure that every partner received tissues from the same production lot. Hence, it would have been plausible in terms of tissue conditions if all partners had misclassified the same chemicals. However, this explanation failed in view of the actual test results.

A thorough investigation of the experimental conditions during the validation runs provided convincing evidence that some test results in the study were negatively influenced by the volatile fractions of some skin-irritating reference chemicals, especially heptanal and cyclamen aldehyde. They had been tested together with non-irritating chemicals in the same culture plate according to our initial SOP. Although the vapor pressures of the concerned irritating chemicals are quite low (e.g. <http://echa.europa.eu/>) we hypothesized that under the given experimental conditions, especially culture of the treated tissue models at 37 °C, small amounts of the irritant substances evaporated from the surface of the these tissues and eventually reached the tissues in the neighbouring wells of the same plates via diffusion. In order to support the above outlined hypothesis of cross contamination due to volatile irritating chemicals, we designed an analytical approach, using GC/MS methods, to detect even small quantities of heptanal and cyclamen aldehyde in affected neighbouring OS-REp models. Additionally a biological approach was chosen in 2 independent experiments which recapitulated the experimental situation being prevalent during the catch-up validation study.

The time course of aldehyde accumulation in both the tissue equivalents and the TENAX adsorbent powder is indicative of the dynamic within the culture plates. It can be assumed that already shortly after starting the tissue incubation with the volatile compounds the whole space of the 6-well plate was filled with the respective molecules. Certainly aldehyde evaporation was facilitated by the incubation temperature of 37 °C. Thus, under the given test and cultivation conditions, the aldehydes rapidly accumulated in the neighbouring receptor tissue equivalents via diffusion. Unexpectedly, even after the donor tissues had been thoroughly rinsed and placed in the 6-well plate again, aldehyde accumulation in the receptor models continued, leading to even higher aldehyde levels during the post-treatment incubation period. This observation can be explained only if we assume that the donor tissues became massively impregnated with the neat chemicals during the exposure period. Even after removal of the substances from the tissue surface the molecules then gradually evaporated from the inner part of the tissues in the course of the post-treatment incubation period, boosted by the culture temperature of 37 °C.

The accumulation rate of heptanal in the receptor models was higher compared to that of cyclamen aldehyde. This result can be presumably

attributed to differences in vapor pressure and flash point, which both point at heptanal as the more volatile agent (e.g. <http://echa.europa.eu/>).

It also became evident that both aldehydes were metabolically reduced into their respective alcohols within the epidermal tissues. This is a clear indication for at least basal metabolic activity of the keratinocytes at the beginning of the post-treatment incubation period, as far as the cells were not yet irreversibly damaged. In the inert TENAX adsorber material, the aldehydes accumulated over time to extremely high values. However, due to the lack of any metabolic capacity of the adsorbent no reduction into their alcohols or other metabolites occurred.

With the biological approach the immediate toxic effect of volatile heptanal on juxtaposed tissue models could be demonstrated, finally proving the correctness of our hypothesis. In contrast to heptanal, in this approach cyclamen aldehyde was inconspicuous in respect of damaging neighbouring epidermal equivalents, although this chemical, when applied topically, led to massively decreased viabilities of the respective tissue. The question why in our experiment the massive accumulation of cyclamen aldehyde according to GC/MS analysis did not also result in the expected viability decrease remains open at that point.

However, when tested alone in a culture plate and protected from any possibly interfering volatile irritants, all 4 previously falsely classified non-irritants were predicted correctly. The observation that in different laboratories different chemicals were falsely classified can be considered a consequence of the double-blind coding of the reference chemicals. As the chemicals samples were coded differently for every laboratory, they were tested in different orders and hence in different pairwise combinations on the 6-well plate, too.

To prevent cross contamination and thus misclassification of chemicals due to volatile irritating chemicals in the atmosphere of the test plate, we highly recommend precautionary measures which should be implemented in SOPs, Performance Standards and the OECD TG 439.

First of all it should be guaranteed that a single substance is tested in a single 6-well plate only. Special care should be taken regarding the D-PBS-treated negative controls, too. Any cross contamination must be avoided e.g. by placing them far away from any volatile substances in the incubator or even in a separate device. Only then the uncompromised maximum tissue viability can be determined, which, because this value is defined as 100% relative viability, influences all other results in a given test. Covering or sealing the 6-well plates with a plastic film cannot be recommended throughout the whole test period. While it might work during the 35 min of incubation time to withhold toxic vapors, sealing the tissue models for 42 h is inadequate, possibly leading to unpredictable culture conditions in the vessel. The above-mentioned recommendations were adopted in the latest version of the respective SOP, which was the basis for the second phase of the OS-REp SIT catch-up validation study (Groeber et al., 2016-in this volume).

In addition the recent findings should ignite a discussion about the question whether those 20 substances stipulated by the OECD TG439 are the best, in respect of their physico-chemical properties, to reliably validate new in vitro skin irritation tests. Although cyclamen aldehyde and heptanal, together with the other reference chemicals (RC), were considered non-volatile when the RC list was originally defined (Eskes et al., 2007), our results, but also their use in industry, point into another direction. Cyclamen aldehyde, for example, serves as a fragrance in a plethora of cosmetic and household products which requires a certain degree of volatility (http://en.wikipedia.org/wiki/Cyclamen_aldehyde). Heptanal, as a precursor of perfume components, is already described as a bad-smelling chemical in publicly available databases (<http://en.wikipedia.org/wiki/Heptanal>), which again indicates at its volatile property. Thus, in order to avoid any of the above mentioned side effects when conducting a catch-up validation study, especially the physical properties of the RCs like vapor pressure and flash point should be thoroughly revised and eventually critical chemicals replaced.

On the other hand the analysis of cross contamination presented in this work might open the door to testing gaseous chemicals and

aerosols in vitro in respect of their skin irritation potential. It became evident that epidermal equivalents exposed to gaseous chemicals only via diffusion in a lid-covered 6-well plate reacted similarly to equivalents treated directly with the respective liquids. Thus, one could imagine a quite simple experimental set-up where the skin models and either a small reservoir of volatile compounds or a defined volume of a gas or aerosol are enclosed in a vessel under static conditions. To our knowledge no validated method exists so far for testing gases and aerosols (OECD TG 439, 2015a; OECD Guidance Document No. 203, 2014).

During the catch-up validation study 2 chemicals were classified false negative in 1 laboratory, an effect which cannot be explained by cross contamination as described above. No explanation was found for 5% KOH, because when retested in the other 2 laboratories identical samples of this highly alkaline solution always reduced the viability to values below 10% (data not shown). For 1-bromohexane (1-BH) the situation appears quite different, and its selection as a reference chemical bears some unpredictabilities. It was not only tested false negative in 1 laboratory, but also in 1 out of 3 valid runs in a second laboratory.

Other companies faced this problem, too. Straticell (Belgium) has presented results of a skin irritation study with their own epidermal model, where 1-bromohexane was not always predicted correctly (Eeman et al., 2010). In addition, they observed a high degree of variability in the viability data between the single runs, which could not be eliminated even after optimizing the testing process. Similar results were shown in a multicenter study performed by the Japanese company J-TEC Co. Ltd. with the epidermal model "LabCyte EPI-MODEL24" in preparation of a catch-up validation study (Katoch et al., 2009; OECD, 2011; Kojima et al., 2014). In the Japanese study 1-BH was predicted as a non-irritant according to the MTT data, and only after taking supplementary IL1- α release data into account it was considered to be an irritant. Increasing the volume of test substance did not affect the test results at all. Raising the temperature to 37 °C during substance treatment resulted in a correct prediction of 1-BH, but other non-irritating chemicals were predicted false positive instead. Only standardizing the washing protocol in order to remove chemical remnants from the tissue surface resulted in sufficiently low relative viabilities for a classification concordant with the in vivo score. However, it is not obvious why a more gentle washing procedure lowered tissue viability after substance application. Any other modifications revealed no effect regarding the 1-BH challenge.

During the EURL-ECVAM validation study 1-BH was misclassified in three laboratories when using the EpiDerm™ model (Spielmann et al., 2007). No clear reasons were provided for the misclassification. Misclassification of 1-BH as a non-irritant was also recently reported in a skin irritation study with the KeraSkin™-VM epidermal model (Jung et al., 2014).

The varying behaviour of the different epidermal models after 1-bromohexane treatment conspicuously corresponds with the effects observed on the native human skin as revealed in a study published by Jírová et al. (2010). In a 4-h patch test 16 out of 30 test persons exhibited a positive skin irritation reaction (53%), whereas the other 14 test persons did not show any sign of a reaction. Therefore the chemical with an in vivo score of 2.7 (OECD, 2015b) must be discussed as a "borderline" substance, of which the reaction cannot be determined unambiguously. Based on these data it can be hypothesized that the high degree of variability seen in the patch test is also mirrored in the tissue models when they are constructed from cells of different donors. Depending on the individual genetic and physiological backgrounds of their donors, keratinocytes and epidermal models, respectively, will then react the one or other way when confronted with 1-BH. A uniform reaction cannot be expected on this premise.

Another case of low Draize test -human correlation was revealed with di-n-propyl disulphide with only 20% of the human volunteers exhibiting positive reactions (Jírová et al., 2010). Interindividual differences in the in vivo test and thus high donor variability in epidermal equivalents are the likely cause for the observed differences in the in vitro skin irritation tests. Using chemicals, where the skin irritation

reaction differs significantly between rabbit and men and even within a human population, as a standard for validation of an in vitro method must be seen very critically. Especially against the background of the open source concept, which explicitly allows flexibility in respect of cell donor origin, these standards can erect massive hurdles on the way to establish the new method at any laboratory. It must also be questioned whether a new in vitro method should not better mirror the human skin physiology rather than the animal situation.

Differences based on the ethnic background of the epidermal models are discussed as a possible reason for misclassification, too (Jung et al., 2014). Taken together these observations should initiate a discussion about the value of 1-bromohexane and di-n-propyl disulphide as SIT reference substances. More emphasis should be laid on human data if they are available for a certain chemical.

After having re-analyzed those chemicals which were falsely predicted in the first instance the predictive capacity of the OS-REP SIT has been recalculated for each of the participating laboratories, eliminating those discordant results clearly linked to the cross-contamination effect.

On this premise all laboratories met the EURL-ECVAM acceptance criteria of ≥80% sensitivity and ≥70% specificity. Whereas the degree of sensitivity was not affected by the recalculation, the specificity was markedly increased. Only the classification for 5% KOH and 1-bromohexane, both considered skin-irritant according to the Draize skin test, remained unchanged due to the arguments presented above.

Taken together, the OS-REP SIT method has proven its value to distinguish non-irritants (below GHS category 2) from skin-irritating substances (GHS category 2) with high reliability.

With the OS-REP a new epidermal equivalent has been introduced as a valuable tool to assess the skin irritation potential of chemicals in vitro. The most recent version of the OECD TG 439 (2015a) lists four commercially available epidermal equivalents which have undergone the complete formal validation process for its use in in vitro skin irritation testing. Based on the data and experiences achieved during the EpiSkin™, EpiDerm™ SIT (EPI-200) and SkinEthic™ RHE validation studies, the so-called Validated Reference Methods (VRM), performance standards were defined which are mandatory to be matched by new methods considered similar to the VRM (me-too methods). They define the basis for every catch-up validation study for skin irritation tests with reconstructed epidermal models (EURL-ECVAM, 2009a, 2009b; OECD, 2015b).

With 74.4% sensitivity and 80.8% specificity the EpiSkin™ SIT predictive capacity was considered sufficiently high in order to endorse this method as a full replacement of the Draize skin test. According to the lower sensitivity of 56.3% in the course of the original ECVAM SIVT, the protocol for the EpiDerm™ SIT was optimized (Kandárová et al., 2009). Specificity was then 76.3%, and sensitivity increased to 86.1% with an overall accuracy of 80.6% for a set of 55 test chemicals, which qualified the EpiDerm™ SIT as a stand-alone test, too.

The validation study with the SkinEthic™ RHE model resulted in 90% sensitivity, 80% specificity and an overall accuracy of 85% (Alépée et al., 2010). While sensitivity of the OS-REP SIT was even better, specificity was lower, which might be attributed to changes in the reference chemicals list after the SkinEthic™ SIT study had been conducted.

The list of epidermal models for SIT was only recently complemented by the LabCyte EPI-MODEL24 SIT (OECD, 2015a). The overall predictive parameters of the Labcyte catch-up validation study with the optimized protocol were 70% specificity, 90–100% sensitivity and 80–85% accuracy (Kojima et al., 2014). In a skin irritation test carried out with the KeraSkin™-VM epidermal equivalent following the Performance Standards 70% specificity and 80% sensitivity were achieved, too (Jung et al., 2014).

Thus, the predictive capacity of our OS-REP SIT is comparable with or even better than the already validated skin irritation assays listed above. On this basis the second tier of the catch-up validation study was organized, where also OS-REP model production was completely conducted

under open source conditions. The results are published in another paper in this journal (Groeber et al., 2016-in this volume).

To our knowledge this is the first time that an *in vitro* skin irritation test based on the open source concept undergoes a Performance Standard-driven validation study. The open source concept encompasses both the production of the respective epidermal equivalent (OS-REp) and the test performance, which gives any potential user the freedom to operate independently of commercial or legal restrictions.

As already indicated in the introduction chapter, the open source concept entered public consciousness first as an innovative strategy in information technology (Nicolosi and Ruivenkamp, 2012). In the meantime the open source debate has also been opened e.g. in the fields of nanotechnology (Frangioni, 2012), drug discovery (Ardal, 2012) and biomedical technology (Pearce, 2012), both aiming at more flexibility, transparency, and product availability in the respective markets. The increasing scientific and public interest in open source concepts is also mirrored in the rapid increase in the number of open source-related publications during the last few years (<http://www.ncbi.nlm.nih.gov/pubmed>).

Many of the features characterizing the open source concept for software can be applied for animal alternatives, too. All underlying protocols and quality criteria for the tissue production as well as for the *in vitro* test are made publicly available without any restrictions. Everyone will be free and in the legal position to perform all working steps and experiments by his/her own. However, the right to alter the source code, or, in the case of an *in vitro* method, the underlying SOP's, must be restricted for the following reasons. In the case of validated and regulatory accepted tissue models and alternative methods no deviations from the protocols as well as from the respective quality and performance standards must be allowed. Any change in the protocols can lead to unexpected effects which, in the end, can compromise the reliability and predictivity of the whole test system. Therefore, once a method has undergone the complete validation and acceptance process, the quality criteria for tissue model production and test performance are mandatory for any user when the test is used in a regulatory framework. Monitoring of the quality standards at the production sites/test labs is recommended on a regular basis.

Nevertheless, the development/optimization of tissue models and new *in vitro* test systems are dependent on the active participation of interested and experienced users and scientists. Only improvements in the protocols and adaptations of the quality standards will pave the way to innovation. However, in the end the optimized test method or tissue model will differ from the original one, for which the validation had been achieved. Therefore a new validation study or at least a proficiency exercise will be required in order to prove the equivalence of the new and old test method.

10. Conclusion

Taken together, the OS-REp SIT method presented in this paper has proven its value to distinguish non-irritants from skin-irritating substances with high reliability, thereby matching the stipulations of the respective Performance Standards. The predictive capacity is comparable to those achieved with the four test methods listed in the OECD TG 439 (OECD, 2015a).

The OS-REp mimics anatomical and structural features of the human epidermis and thus can be employed to predict skin irritation. Several quality criteria must be met before OS-REp models become qualified to be used in the skin irritation assay. Beside those criteria as described in the Performance Standards (OECD, 2015b) it is of utmost importance to carefully assess the suitability of the keratinocytes in order to ensure that the tissue models constructed of these cells have the appropriate predictive capacity. Only after a proficiency exercise with all PS reference chemicals has been successfully conducted, the open source-based SIT can be used to test unknown chemicals for their skin-irritating potential.

The test protocol had to be modified after it became evident that some of the chemicals were classified false-positive in the first instance, resulting in a too low overall specificity. It was proven that some of the skin-irritating reference chemicals used in the validation study were volatile and thus able to contaminate and damage tissue models cultured in the same culture vessel or even in the same incubator. When the non-irritating overpredicted reference chemicals were retested with the slightly modified protocol, in which every chemical was tested in an individual culture vessel, they were all predicted correctly. These results were then included into the final calculation of predictivity.

From these analyses we recommend to implement specific measures in order to avoid any cross contamination effects. The tissue models should be incubated with a single substance in a single 6-well plate only, and models treated with non-irritants should be kept separately from those treated with volatile irritants, at best in separate CO₂ incubators, if possible. Negative controls should be kept as far away as possible from any volatile irritants.

We also highly recommended that these measures should be implemented in SOPs, Performance Standards and the respective OECD testing guideline in order to raise consciousness for challenging physical properties of the chemicals to be tested and eventually to avoid cross contamination and thus misclassification of chemicals.

A few reference chemicals, especially 1-bromohexane, but also di-*n*-propyl disulphide, appear not to be suitable to validate *in vitro* skin irritation tests due to their poor animal – human correlation and high impact of donor variability on the test outcome. A replacement of these chemicals in the TG is highly recommended.

Transparency document

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

Acknowledgements

The authors would like to thank Rolf Herrmann and Daniel Armanazi (Henkel AG & Co. KGaA, Düsseldorf, Germany) for conducting the GC/MS analyses. We also thank Julian Tharman (at that time BfR, ZEBET, Berlin, Germany) for coding and distributing the 20 Reference Chemicals and also Dr. Dirk Dressler (BioTeSys GmbH, Esslingen, Germany) for the re-coding and shipping of the 6 test chemicals, used for re-testing.

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.
- Andreadis, S.T., Hamoen, K.E., Yarmush, M.L., Morgan, J.R., 2001. Keratinocyte growth factor induces hyperproliferation and delays differentiation in a skin equivalent model system. *FASEB J.* 15, 898–906.
- Ardal, C., Röttingen, J.A., 2012. Open source drug discovery in practice: a case study. *PLoS Negl. Trop. Dis.* 6, 1–11.
- Baskett, D.A., York, M., McFadden, J.P., Robinson, M.K., 2004. Determination of skin irritation potential in the human 4-h patch test. *Contact Dermatitis* 51, 1–4.
- Baskett, D., Jirova, D., Kandárová, H., 2012. Review of skin irritation/corrosion hazards on the basis of human data: a regulatory perspective. *Interdiscip. Toxicol.* 5, 98–104.
- Costea, D.E., Loro, LL., Dimba, E.A.O., Vintermyr, O.K., Johannessen, A.C., 2003. Crucial effects of fibroblasts and keratinocyte growth factor on morphogenesis of reconstituted human oral epithelium. *J. Invest. Dermatol.* 121 (1479), 1486.
- De Wever, B., Petersohn, D., Mewes, K.R., 2013. Overview of human three-dimensional (3D) skin models used for dermal toxicity assessment - part 1. *H&PC Today* 8, 18–23.
- Eeman, M., Bastin, C., Leroy, M., Belot, N., Salmon, M., 2010. Prediction of Skin Irritation Potential of Chemicals Using StraticeLL Reconstituted Human Epidermal Model. I-SUP2010. Innovation for Sustainable Production, Bruges, Belgium.
- 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. *ATLA Altern. Lab. Anim.* 35, 603–619.
- EU, 2009. Regulation (EC) No 1223/2009 of the European Parliament and of the council of 30 November 2009 on cosmetic products. Download under. <http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02009R1223-20150416&from=EN>.
- EURL-ECVAM, 2008. Statement on the scientific validity of *in vitro* tests for skin irritation testing, issued by the ECVAM scientific advisory committee (ESAC29), 5 November

2008. <https://eurl-ecvam.jrc.ec.europa.eu/validation-regulatory-acceptance/docs-skin-irritation-1/DOC7-ESAC2008.pdf>.
- EUR-ECVAM, 2009a. Statement on the “performance under UN GHS of three in vitro assays for skin irritation testing and the adaptation of the innovation for sustainable ProductionReference chemicals and defined accuracy values of the ECVAM skin irritation performance standards”, issued by the ECVAM Scientific Advisory Committee (ESAC30), 9 April 2009. https://eurl-ecvam.jrc.ec.europa.eu/about-ecvam/archive-publications/publication/ESAC30_statement_skin-irritation-methods-PS_20090409.pdf.
- EUR-ECVAM, 2009b. ESAC statement on the performance standards (PS) for in vitro skin irritation testing using reconstructed human epidermis, issued by the ECVAM scientific advisory committee (ESAC31), 8 July 2009. https://eurl-ecvam.jrc.ec.europa.eu/validation-regulatory-acceptance/docs-skin-irritation-1/DOC9-ESAC_statement_updated-skin-irritation-PS_200909-22.pdf.
- EUR-ECVAM, 2009c. Performance Standards for In-Vitro Skin Irritation Test Methods Based on Reconstructed Human Epidermis (RhE).
- Frangioni, J.V., 2012. Nonprofit foundations for open-source biomedical technology development. *Nat. Biotechnol.* 30, 928–932.
- Groeber, F., Schober, L., Schmid, F., Traube, A., Hernandez-Kolbus, S., Daton, K., Gonska, H., Hoffmann, S., Petersohn, D., Traube, A., Schäfer-Korting, M., Walles, H., Mewes, K.R., 2016. Catch-up validation study of an in vitro skin irritation test method based on an open source reconstructed epidermis (phase II). *J. Dermatol.* 36, 254–261 (in this volume).
- 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.
- Jung, K.M., Lee, S.H., Jang, W.H., Jung, H.S., Heo, Y., Park, Y.H., Bae, S., Lim, K.M., Seok, S.H., 2014. KeraSkin-VM: a novel reconstructed human epidermis model for skin irritation tests. *Toxicol. In Vitro* 28, 742–750.
- Kandárová, H., Hayden, P., Klausner, M., Kubilus, J., Kearney, P., Sheasgreen, J., 2009. In vitro skin irritation testing: improving the sensitivity of the EpiDerm skin irritation test protocol. *ATLA Altern. Lab. Anim.* 37, 671–689.
- Karvinen, S., Pasonen-Seppänen, S., Hyttinen, J.M., Pienimäki, J.P., Törrönen, K., Jokela, T.A., Tammi, M.I., Tammi, R., 2003. Keratinocyte growth factor stimulates migration and hyaluronan synthesis in the epidermis by activation of keratinocyte hyaluronan synthases 2 and 3. *J. Biol. Chem.* 278, 49495–49504.
- 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.
- Kojima, H., Ando, Y., Idehara, K., Katoh, M., Kosaka, T., Miyaoka, E., Shinoda, S., Suzuki, T., Yamaguchi, Y., Yoshimura, I., Yuasa, A., Watanabe, Y., Omori, T., 2012. Validation study of the in vitro skin irritation test with the LabCyte EPI-MODEL24. *ATLA Altern. Lab. Anim.* 40, 33–50.
- Kojima, H., Katoh, M., Shinoda, S., Hagiwara, S., Suzuki, T., Izumi, R., Yamaguchi, Y., Nakamura, M., Kasahawa, T., Shiba, A., 2014. A catch-up validation study of an in vitro skin irritation test method using reconstructed human epidermis LabCyte EPI-MODEL24. *J. Appl. Toxicol.* 34, 766–774.
- Nicolosi, G., Ruivenkamp, G., 2012. Re-skilling the social practices: open source and life-towards a commons-based peer production in agro-biotechnology? *Sci. Eng. Ethics* <http://dx.doi.org/10.1007/s11948-012-9405-4>.
- OECD, 2011. Peer Review Report of the Validation of the Skin Irritation Test Using Labcyte EPI-MODEL24 Series on Testing and Assessment. No. 155.
- OECD, 2014. New Guidance Document on an Integrated Approach on Testing and Assessment (IATA) for Skin Corrosion and Irritation. OECD Environment, Health and Safety Publications, Series on Testing and Assessment No. 203.
- OECD, 2015a. OECD Guidelines for the Testing of Chemicals no. 439: In Vitro Skin Irritation: Reconstructed Human Epidermis Test Method. Organisation for Economic Co-operation and Development, 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. [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO\(2015\)27&docLanguage=En](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2015)27&docLanguage=En).
- Pearce, J.M., 2012. Physics: make nanotechnology research open-source. *Nature* 491, 519–521.
- 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.
- R Development Core Team, 2010. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (<http://www.R-project.org>).
- Russel, W.M.S., Burch, R.L., 1959. *The Principles of Humane Experimental Technique*. Methuen, London, pp. 69–154.
- Sarkar, D., 2010. Lattice: lattice graphics. R package version 0.18-8. <http://CRAN.R-project.org/package=lattice>.
- 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., Remmeli, 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 Altern. Lab. Anim.* 35, 559–601.
- Stark, H.J., Baur, M., Breitkreutz, D., Mirancea, N., Fusenig, N.E., 1999. Organotypic keratinocyte cocultures in defined medium with regular epidermal morphogenesis and differentiation. *J. Invest. Dermatol.* 112, 681–691.