Toxicology in Vitro 94 (2024) 105726

Contents lists available at ScienceDirect

Toxicology in Vitro

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

Assessment of the utility of the novel Phenion® full thickness human skin model for detecting the skin irritation potential of antimicrobial cleaning products

Kathryn Page [a,*], Walter Westerink [b], Kristie Sullivan [c,1], Thomas McDonald [a], Clive Roper [d]

[a] Clorox Services Company, 121 Broadway, Oakland, CA 94612, USA
[b] Charles River Laboratories, Hambakenwetering 7, 's, Hertogenbosch 5231 DD, The Netherlands
[c] Physicians Committee for Responsible Medicine, 5100 Wisconsin Ave NW Ste 400, Washington 20016, USA
[d] Roper Toxicology Consulting Limited, 6 St Colme Street, Edinburgh EH3 6AD, UK

ARTICLE INFO

Editor: Dr. P Jennings

Keywords:
Skin irritation
In vitro
Formulations
Antimicrobial cleaning products
Disinfectants
Phenion® FT

ABSTRACT

The skin is a potential route of exposure to antimicrobial cleaning products (ACP). Skin irritation, reversible damage to the skin, is an endpoint for protecting consumers and operators accidently exposed to these complex mixtures. To assess skin irritation of 24 ACP formulations, a new protocol was developed and adapted from OECD Test Guideline No. 439 with EpiDerm™ (epidermis model) replaced by Phenion® FT (full thickness tissue, including epidermis and dermis) as the test system. A full thickness tissue was utilized to provide a more human *in vivo-like* model. Formulations were applied to Phenion® FT and cell viability measured by MTT reduction after a 15-min exposure and 42 h post exposure period. A prediction model was applied, and results compared with *in vivo* rabbit skin irritation data. Concordance between *in vivo* and *in vitro* was demonstrated to be suitable (*i.e.*, sensitivity 78%, specificity 83%, and accuracy 79%) using this modified OECD Test Guideline No. 439 method with a 70% cell viability selected as the most reasonable cut off for discriminating non-irritants (EPA Class IV). These results were considered suitable to develop a draft IATA *i.e.*, with any ACP formulation identified as EPA Category IV in this test. The method will be further refined to distinguish irritant categories.

1. Introduction

The skin is a potential route of exposure to chemicals that are present in commonly used household and professional products, such as antimicrobial cleaning products (ACP). Skin irritation, *i.e.*, reversible damage to the skin, is an important endpoint for protecting consumers and operators exposed to these complex mixtures. Therefore, assessment of skin irritation toxicity potential is an important requirement for ACP manufacturers, as well as government regulatory agencies charged with overseeing the safety of chemicals and consumer products. Skin irritation testing is required in the U.S. and Europe for hazard identification and product labelling for pesticides and other industrial chemicals, and to determine levels of personal protective equipment for safe handling (US EPA, 2016).

The irritation tests are based on the experience that irritant chemicals show cytotoxic effects following short term exposure to the *stratum corneum* of the epidermis. Methods for assessing skin irritation, which are currently accepted by regulatory authorities either rely on use of rabbits *in vivo,* a test first proposed by Draize et al. (1944) and subsequently refined as OECD Test Guideline (TG) No. 404 (OECD, 2015), or

Abbreviations: 3D, 3-Dimensional; ACP, Antimicrobial Cleaning Products; ADBAC, Alkyl Dimethyl Benzyl Ammonium Chloride; CO_2, Carbon Dioxide; CoA, Certificate of Analysis; CTAC, Cetyl Trimethyl Ammonium Chloride; DDAC, Didecyl Dimethyl Ammonium Chloride; EPA, US Environmental Protection Agency; FT, Full thickness; GHS, **Globally Harmonized System of Classification and Labeling of Chemicals**; GLP, Good Laboratory Practice; IATA, Integrated Approaches to Testing and Assessment; LDH, Lactate Dehydrogenase; MSDS, Material Safety Data Sheet; MTT, 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyl Tetrazolium Bromide; OECD, Organisation for Economic Cooperation and Development; PBS, Phosphate Buffered Saline; RhE, Reconstructed Human Epidermis; SDS, Sodium Dodecyl Sulphate; TG, Test Guideline; UN, United Nations.

* Corresponding author at: The Clorox Company, 4900 Johnson Dr, Pleasanton, CA 94588, USA
E-mail addresses: kathryn.page@clorox.com (K. Page), walter.westerink@crl.com (W. Westerink), ksullivan@iivs.org (K. Sullivan), thomas.mcdonald@clorox.com (T. McDonald), clive@ropertcl.com (C. Roper).
[1] Current affiliation: Institute for *In Vitro* Sciences, 30 West Watkins Mill Road, Suite 100, Gaithersburg, MD 20878, USA.

https://doi.org/10.1016/j.tiv.2023.105726
Received 15 June 2023; Received in revised form 23 October 2023; Accepted 1 November 2023
Available online 4 November 2023

by various reconstructed human epidermis (RhE) *in vitro* models (OECD, 2021). The results of these tests are currently utilized to assign chemicals to specific hazard categories with global regulatory labelling implications according to the Globally Harmonized System (GHS) of Classification and Labeling of Chemicals (UN GHS, 2021) and the US EPA classification system (US EPA, 2018). The US EPA categorizes irritant substances into Category II, III, or IV (Category IV being equivalent to UN GHS No Category). Most global regulatory authorities classify irritants as GHS Category 2 only. The US EPA identify Category II (moderate irritant) and Category III (mild irritant). All other substances do not require classification for labelling for skin irritation hazard (GHS No category and EPA Category IV).

In an initial proof-of-concept test (unpublished) with MatTek EpiDerm™ using the OECD TG No. 439 (OECD, 2021) and 10 ACP formulations, compared with the rabbit *in vivo* classification, the sensitivity, specificity, and accuracy were 100, 33 and 60%, respectively. The specificity was considered too low with this set of ACP formulations for further evaluation. Optimization of existing prediction models within OECD TG No. 439 (*e.g.*, modifications of exposure time or cut off values) may also improve productivity and these are under investigation in similar efforts. Other reconstructed human epidermis models approved for use in OECD TG No. 439 (OECD, 2021) were expected to perform similarly. All these test systems are reconstructed human epidermal skin models, have been cross validated against the reference test model and are assumed to have a similar skin barrier function. Therefore, a novel approach was investigated using the Phenion® FT skin model.

The Phenion® FT skin model consists of both a fully differentiated epidermis, the outermost layer of the skin, and the underlying dermis that contains a natural, collagen-based connective tissue. In consequence, the model resembles human native skin in a variety of anatomical and physiological properties, making it an ideal tool for research projects as well as for the development of innovative, replacement test methods. Each batch of the Phenion® FT model is well characterised for quality and is supplied with a Certificate of Analysis (CoA) and a Material Safety Data Sheet (MSDS). The CoA includes the production lot, matrix lot, fibroblast lot, keratinocyte lot, date of shipment, a contamination and quality assessment and a lactate dehydrogenase (LDH) release test. The serological test results identify that the batch is negative for Hepatitis B, Hepatitis C, HIV-1/HIV-2, and mycoplasma. The LDH release assay confirms that the batch has passed a 24 h LDH release assay confirming that the non-stressed cells display an LDH activity of <5% of the Triton-X-100 (1%) stressed cells so confirming that they have passed the quality control test. The MSDS contains a description of the model (*i.e.*, *in vitro* full thickness, epidermis and dermis, skin model), state of differentiation (*e.g.*, Air Liquid Interface Day 10), storage conditions (*i.e.*, sterile, 37 °C, 5% CO_2 (*v*/v)), production batch, matrix batch and cell batches. The virus status of the cells, the cell type (*e.g.*, non-modified human basic primary keratinocytes and fibroblasts obtained from foreskin from the same donor and a confirmation that all four typical layers (including *stratum corneum*) are present.

The main difference between the reconstructed human epidermal test systems; MatTek EpiDerm™, EpiSkin SkinEthic™ RHE and Phenion epiCS®, and the reconstructed human full thickness test system, Phenion® FT, is that the EpiDerm, SkinEthic™ RHE and epiCS® are epidermis only and the Phenion® FT test system contains dermis underneath the epidermis. For the 3 reconstructed human epidermis test systems, the epidermis is constructed from human keratinocytes and are grown on the polycarbonate membrane at the air liquid interface on Transwell® plates. For the Phenion® FT full thickness model, human keratinocytes are also used to build the epidermis, and primary human fibroblasts are used to populate the collagen sponge and generate the dermis (Mewes et al., 2007). All 4 test systems are sold in 24-well plate formats and develop a *stratum corneum* with similar structural properties to human skin. The presence of the dermis in the Phenion® FT test system provides this model with a more *in vivo*-like structure since the human epidermis (*in vivo*) is embedded onto the dermis. The structural similarities between human skin (Fig. 1A) and the Phenion® FT model (Fig. 1B) can be seen clearly seen.

The aim of this proof-of-concept test was to determine if a modified protocol, based on the EpiDerm™ protocol identified in OECD TG No. 439 (OECD, 2021), using the Phenion® FT test system could accurately predict ACP formulations with known non-irritant animal hazard categorizations.

2. Materials and methods

2.1. In vivo reference data

No new animal tests were performed; historically, generated conventional *in vivo* data were used. These *in vivo* studies were conducted under US EPA Good Laboratory Practice regulations, or equivalent.

2.2. In vitro skin irritation test protocol

The design of the *in vitro* test protocol was based on OECD TG No. 439 (OECD, 2021). The modifications to this standard test are described and highlighted in this manuscript. Since the work was conducted over a period of 2 years and the test guideline was updated frequently, there were different versions of the test guideline issued during the work, although none of them resulted in changes to the method initiated at the start of the programme. The latest version included additional test systems that have been validated for this purpose. All variations to the test guideline are described in detail below and summarised in Table 1. The application volume and exposure time comply with the guideline but are highlighted as they do differ across the individual RhE compliant protocols.

The test system, Phenion® FT skin model, was supplied by Phenion, Düsseldorf, Germany. The accompanied CofAs demonstrated that the models complied with the acceptance criteria set by Phenion. All Phenion® FT production lot numbers complied with the acceptance criteria applied by Phenion and these are identified in Table 2. The use of this full thickness human-derived tissue construct was a critical change from the OECD TG No. 439 (OECD, 2021) which uses reconstructed human epidermal test systems. All other changes were a consequence of the decision to use this test system. A cross section through normal human skin is shown in Fig. 1(A) and is compared with a cross section of the Phenion® FT model in Fig. 1(B). The Phenion® FT model is produced on a Transwell® insert (Fig. 2). The Phenion® FT skin irritation tests were performed at Charles River Laboratories, 's-Hertogenbosch, The Netherlands. The study did not claim GLP compliance, but the experimental phase of the study was conducted in the quality assured environment of Charles River Laboratories 's-Hertogenbosch GLP Test Facility.

The ACP formulations were prepared by The Clorox Company, Pleasanton, CA, USA. The formulations were prepared to have the same composition to formulations used to generate hazard classifications in the conventional animal test. Formulation details, including their main active ingredient, are summarised in Table 3.

The ACP formulations, which were all supplied as liquids, were stored at room temperature, or maintained at *ca* 4 °C in a refrigerator, according to the manufacturer's (The Clorox Company) instructions prior to use. The ACP formulations were not supplied with a CoA. Expiry dates were assigned as a default of 1 year from preparation, and all were evaluated within this timeframe. The negative control was phosphate buffered saline (PBS) obtained from Merck KGaA, Darmstadt, Germany. The positive control was sodium dodecyl sulphate (SDS) obtained from Sigma-Aldrich Chemie GmbH, Steinheim, Germany diluted in PBS (5%, *v*/v).

A formulation or chemical may interfere with the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) endpoint if it

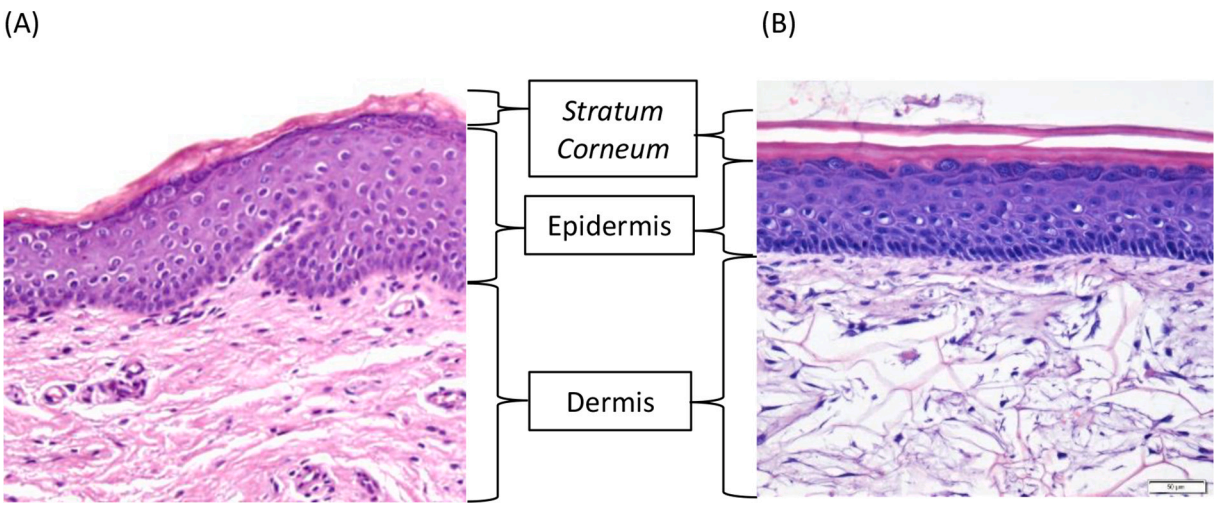

Fig. 1. Cross Section Through (A) Human Skin and (B) Phenion® FT Skin Model.
Photo supplied by https://www.phenion.com

Table 1
Modifications to OECD TG No. 439 (OECD, 2021).

Model Step	OECD TG No. 439 Requirement	Adaptation
Test System	Approved RhE Models	Phenion® FT Model
Application Volume	10–40 µL	39 µL
Exposure Time	15–60 min	15 min
Cutting Tissue Prior to MTT Test	Not Applicable	Additional Step
Tissue Viability for Irritant Classification	≤50% (GHS 2)	≤70% (EPA I/II/III)[¥]
Tissue Viability for Non-Irritant Classification	>50% (GHS No Category)	>70% (EPA IV)[¥]

[¥] Derived from the results generated herein.

Table 2
Phenion FT Production Lot No. Acceptance Data.

Production Lot No.	*LDH Release (%)
Phe-HM-21-14	0.8
Phe-HM-22-01	1.6
Phe-HM-19-24	1.0
Phe-HM-19-23	0.8

[*] The acceptance criterion for LDH release for non-stressed tissues was <5% of the Triton-X-100 stressed tissues.

is coloured and/or is able to directly reduce MTT. The cell viability measurement is affected only if the chemical or formulation is present on the tissues when the MTT viability test is performed. The ACP formulations were checked for possible colour interference before the study was started. A non-coloured chemical or formulation may change into a coloured solution in aqueous conditions and thus stain the skin tissues during the exposure. To assess the colour interference, an ACP formulation (10 µL) was added to Milli-Q water (90 µL). This was mixed for *ca* 15 min. A negative control, Milli-Q water (10 µL), was evaluated concurrently. At the end of the shaking period a visual colour check was performed. The formulation was checked for possible direct MTT reduction before the study was started. To assess the ability of the ACP formulations to reduce MTT, each ACP formulation (39 µL) was added to MTT solution (2 mL, 0.5 mg/mL in PBS). The mixture was incubated for 3 h at 37 °C. A negative control, sterile Milli-Q water (25 µL) was evaluated concurrently. At the end of the incubation period a visual colour check was performed.

On the day of tissue receipt, the Phenion® FT tissues were

Fig. 2. Phenion® FT Skin Model in Transwell® Insert Format.
Photo supplied by https://www.phenion.com

transferred to 12-well plates and pre-incubated with pre-warmed Maintenance Medium (Phenion, Düsseldorf, Germany) for 18 to 24 h at 37 °C. The production lot numbers used were identified against the ACP formulations assessed. When a sample failed any of the acceptance criteria, it was repeated with the next available batch.

Three replicate tissues were treated (39 µL) with either the negative control (PBS), positive control (SDS; 5%, *v*/v) or with an ACP formulation. These were topically applied, undiluted, on to the *stratum corneum* surface of the Phenion® FT skin tissue. The positive control was respread after *ca* 7 min contact time. Negative and positive controls were shared with parallel studies. After the exposure period (15 ± 0.5 min) at room temperature, the tissues were washed with PBS to remove residual control or ACP formulation. After rinsing, the cell culture inserts were each dried carefully and moved to a new well on pre-warmed Maintenance Medium (2 mL) until all tissues were dosed and rinsed. Subsequently, the skin tissues were incubated for 42 h under a standard controlled environment (standard conditions). The standard conditions were a humid atmosphere (80–100%) containing carbon dioxide (CO_2; 5.0 ± 0.5%) in air in the dark maintained at an optimal temperature (37.0 ± 1.0 °C). The temperature and humidity were continuously monitored throughout the experiment. The CO_2 percentage was

Table 3
ACP Formulation Information.

ACP Formulation	Batch	Appearance/ Colour	Active Ingredient(s)
F19950660	PS# 2019.199	Clear Light Yellow	Sodium Hypochlorite
82-65C	PS# 2019.200	Clear Light Yellow	Sodium Hypochlorite
F20040111	PS# 2019.201	Clear Colourless	Hydrogen Peroxide
83-48C	PS# 2019.202	Clear Light Yellow	Sodium Hypochlorite
TA214–88	PS# 2019.203	Clear Colourless	Quaternary Ammonium Compounds
F20010053	PS# 2019.204	Clear Blue	Quaternary Ammonium Compounds
F20100130	PS# 2019.205	Clear Amber	Glycolic Acid
FIS20160164	PS# 2019.206	Clear Colourless	Quaternary Ammonium Compounds
F19950616	PS# 2019.207	Clear Light Yellow	Sodium Hypochlorite
FIS20190037	PS# 2019.208	Turbid White	Citric Acid
F19950198	2021,041	Thick Green	CTAC
F19950406	2021,045	Thin Pale Yellow	Hydrogen Peroxide
FIS20170076	PS2021,040	Thin Clear	Sodium Hypochlorite
F20140084	PS2021,038	Thin Amber	ADBAC
FIS20150266	PS2021,036	Thin Blue	ADBAC & DDAC
F20100130	PS2021,037	Thin Amber	Glycolic Acid
FIS20170080	PS2021,039	Thin Purple	ADBAC & DDAC
F20040124	PS2021,035	Thick Blue	Benzalkonium Chloride
F19950616	2021,043	Clear Pale Yellow	Sodium Hypochlorite
F19950010	2021,046	Slightly Thick Green	Sodium Hypochlorite
F19950354	2021,042	Very Thick Green	CTAC
F19950166	2021,044	Thin Brown	Isopropanol
F19950114	20,200,112–1	Clear Pale Yellow	Sodium Hypochlorite
F19950380	20,200,112–2	Clear Pale Yellow	Sodium Hypochlorite

monitored once on each working day. Temporary deviations from the temperature, humidity and CO_2 percentage may occur due to opening and closing of the incubator door. Any variations to these conditions in any single experimental run were evaluated and were not considered to have resulted in an adverse impact on the study integrity.

After incubation, cell culture inserts were dried carefully to remove excess medium. Tissues were cut twice with a sharp scalpel in a roughly crucifix cross shape (Fig. 3). This procedure enabled the MTT solution to better penetrate the complete tissue and allow for improved extraction of reduced MTT (formazan) out of the skin during the MTT tissue viability assay. This additional step was a change from OECD TG No. 439 (OECD, 2021).

Cytotoxicity is expressed as the reduction of mitochondrial dehydrogenase activity measured by formazan production from MTT at the end of the treatment. The cut tissues were transferred into wells of a 12- or 24-well plate, filled with fresh MTT working solution (1–2 mL) prepared from MTT concentrate (Sigma Aldrich, Zwijndrecht, The Netherlands) diluted in Assay Medium (final concentration 0.5 mg/mL). The wells with tissues were placed into an incubator and incubated for 3 h under standard conditions. After this incubation time, the tissues were carefully washed eight times with Dulbecco's PBS (600 µL). Excess liquid was removed while placing the tissues on to dry Kleenex filter paper. The dried tissues were transferred into new wells of a 12- or 24-well plate filled with 2-propanol (1–2 mL). Plates were closed with the lid and sealed with Parafilm® or adhesive foils. The formazan was eluted, refrigerated, and protected from light for 18 to 72 h. Before analysing the formazan extracts, the extracts were diluted four times with isopropanol. The extracts were mixed to obtain homogeneous solutions. From each tissue sample formazan extraction solution (2 × 200 µL) were transferred to the wells of a translucent 96-well flat-bottom microtiter plate. The blank was 2-propanol. The amount of extracted formazan was determined spectrophotometrically at 570 nm, in duplicate, with a TECAN Infinite® M200 Pro Plate Reader.

Each *in vitro* skin irritation test run was considered acceptable if it met all the following three acceptability criteria; (i) The absolute mean OD_{570} of the three tissues of the negative control should reasonably be within the acceptance limits based on OECD TG No. 439 (OECD, 2021) or modified to a lower acceptance limit ($\geq$0.6) and upper acceptance limit ($\leq$2) and the standard deviation (SD) of the % viability was $\leq$18%. (ii) The mean relative tissue viability of the positive control was $\leq$40% relative to the negative control and the SD of the % viability was $\leq$18%. (iii) The SD calculated from individual % tissue viabilities of the three identically treated replicates were $\leq$ 18%.

Cell viability was calculated using the optical density (OD) readings and calculations. The corrected OD (OD_c) for each sample or control (OD_{raw}) was calculated by subtracting the value of the blank mean (OD_{bl}) from each reading (OD_{raw}). The OD value representing 100% cell viability was the average OD of the negative controls (OD_{lt_u+MTT}). The percentage viability for each sample and positive control was calculated by OD_c divided by mean OD_{lt_u+MTT} x 100%.

OECD TG No. 439 (OECD, 2021) uses these irritation prediction interpretations to identify the categorization of the chemical or formulation. A chemical is categorized to be an irritant (*i.e.*, GHS Category 1 or 2) in the *in vitro* skin irritation test using the EpiSkin™ model, if the relative mean tissue viability of three individual tissues after 15 min of

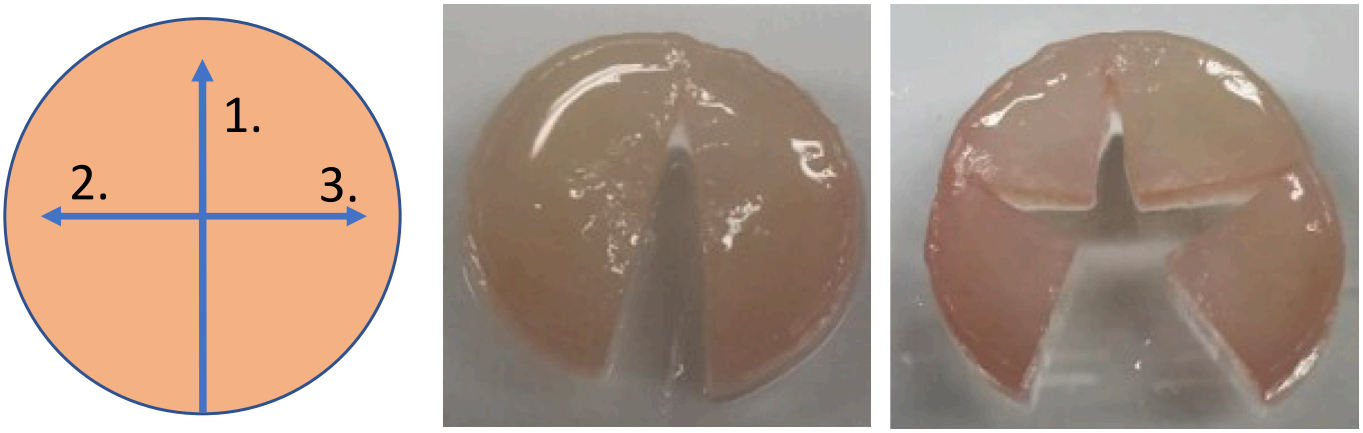

Fig. 3. Cutting of Phenion® FT Skin Tissue Prior to MTT Test. The Schematic Picture Shows the Cuts (1, 2, and 3). Cuts are shown in an example Phenion FT® unit with Cut 1. and then with All 3 Cuts.

exposure to the chemical and 42 h of post incubation, is ≤50% of the mean viability of the negative controls. A chemical is categorized to be non-irritant (*i.e.*, No Category) in the *in vitro* skin irritation test if the relative mean tissue viability of three individual tissues after 15 min of exposure to the chemical and 42 h of post incubation is >50% of the mean viability of the negative controls. This prediction model was not suitable as the Phenion® FT model is a full thickness model, compared to the RhE models accepted for use in the guideline. There is more tissue to damage in a full thickness skin model than in an epidermal skin model. Consequently, the irritation prediction model interpretations were adjusted using the current study results. To assess the irritation prediction interpretations with the best fit to identify the categorization of the ACP formulations, concordance between the *in vivo* and *in vitro* irritation data was determined using the following three methods. Method A used a ≤ 60% cell viability to identify combined GHS Category 2/ EPA Category II and GHS Category 3/EPA Category III and > 60% cell viability to identify GHS No Category/EPA Category IV. Method B used a ≤ 70% cell viability to identify combined GHS Category 2/ EPA Category II and GHS Category 3/EPA Category III and > 70 to identify GHS No Category/EPA Category IV. Method C used a ≤ 50% cell viability to identify GHS Category 2/ EPA Category II and > 50% and ≤ 70% to identify GHS Category 3/EPA Category III and > 70% cell viability to identify GHS No Category/EPA Category IV.

3. Results

The ACP formulations were checked for colour interference in aqueous conditions and possible direct MTT reduction by mixing the ACP formulation with MTT medium. Because no colour changes were observed for any of these ACP formulations, it was considered that none of these formulations would interact with MTT, therefore, these ACP formulations were not considered to impact on the measured endpoint.

Based on the tissue viability (% of negative control) results, the concordance of the cell viability cut off value of 50% defined in OECD TG No. 439 (OECD, 2021) was compared to the three proposed cut off values for irritation prediction. The OECD TG No.439 (OECD, 2021) cut off value of 50% gave a sensitivity of 78%, specificity of 73% and an overall accuracy of 75%. The method for irritation prediction interpretation criteria was selected based on best fit of the three evaluated methods, as well as the standard OECD TG No. 439 (OECD, 2021) criteria (Table 4). The three methods were: Method A gave a sensitivity of 72% (13/18) when GHS Category 2 and 3 were combined, a specificity of 100% (6/6) and accuracy of 75%. Method A also predicted 7/9 GHS Category 2 and 2/3 EPA Category II concordant with *in vivo*, which was considered moderately protective for severe irritants identified *in vivo*. Method B gave a sensitivity of 78% (14/18), a specificity of 83% (5/6) and accuracy of 79% (19/24). Method B also predicted 8/9 GHS Category 2 and 2/3 EPA Category II concordant with *in vivo*, which was considered highly protective for severe irritants identified *in vivo*. Due to the non-concordance with EPA and GHS Categories II/2 and III/3, Method C was evaluated separately for the EPA and GHS Categories. Method C, using EPA Categories, gave a sensitivity of 100% (2/2) when identifying EPA Category II and 19% (3/16) when identifying Category III. Method C, using GHS Categories, gave a sensitivity of 78% (7/9) when identifying GHS Category 2 and 22% (2/9) when identifying Category III. Method C for both EPA and GHS gave an 83% specificity when predicting EPA IV/ GHS No Category (5/6). The accuracy of Method C, when predicting EPA Categories, was 42% (10/24), and 58% (14/24) when predicting GHS Categories.

These models were not *a priori* expectations, but instead were identified by the results generated in this project. Indeed, all cut off values are derived from observations from experiments including the *in vivo* rabbit assay classifications.

Therefore, based on the concordance between *in vivo* and *in vitro* irritation data in the assessed models, 70% cell viability (Model B, with an overall accuracy of 79%) was selected as the most suitable cut off value for discriminating non-irritants (Class IV) from mild and moderate irritants (EPA Class III and II). The logic behind the 70% cut off was that the Phenion® FT test system was full thickness skin, whereas the other test systems used in the OECD Test Guideline No. 439 (OECD, 2021) test are epidermis only. Therefore, there are more cells that would need to be killed by any components of an antimicrobial formulation to result in an irritant prediction. Using the 50% cut off (OECD, 2021) would have resulted in a test less able to identify irritants.

Method B was chosen for the prediction model in this study.

Tissue viability was expressed as the remaining cell viability after exposure to the formulation and normalised against the negative control viability. The mean tissue viability for the ACP formulations compared to the negative control tissues is presented in Fig. 4. ACP formulations, F19950660, 82-65C, F20040111, 83-48C and F20010053, were tested in Phenion® FT Lot No. Phe-HM-19-23. ACP formulations, TA214–88, F20100130, FIS20160164, F19950616 and FIS20190037, were tested in Phenion® FT Lot No. Phe-HM-19-24. ACP formulations, FIS20170076, F20140084, FIS20150266, F20100130, FIS20170080, F20040124, F19950616, F19950010, F19950354, F19950166, F19950198, F19950406, were tested in Phenion® FT Lot No. Phe-HM-21-14. ACP formulations, F19950114 and F19950380, were tested in Phenion® FT Lot No. Phe-HM-22-01. All positive and negative controls fulfilled the acceptance criteria on each testing day. Fig. 4 shows two horizontal lines; the first is at 70% which was the prediction model (Method B). The second is at 100% which shows the mean negative control ($n = 3$) for each batch.

Green and red bars are the negative and positive control tissue viability (% of negative control) for each batch of formulations tested. The different shades of blue bars identify formulations tested in the same Phenion® FT Production Lot No. Formulations F19950660, 82-65C, F20040111, 83-48C and F20010053 were tested on Phenion® FT Production Lot No. Phe-HM-19-23. Formulations TA214–88, F20100130, FIS20160164, F19950616 and FIS20190037 were tested on Phenion® FT Production Lot No. Phe-HM-19-24. Formulations FIS20170076, F20140084, FIS20150266, F20100130, FIS20170080, F20040124, F19950616, F19950010, F19950354, F19950166, F19950198 and

Table 4
Accuracy of tested irritation prediction models.

Prediction Method	Irritant		Non-Irritant	Method Concordance		
GHS:	2	3	None			
EPA:	II	III	IV	Sensitivity	Specificity	Accuracy
OECD TG (GHS only)	≤50%	N/A	>50%	78%	73%	75%
Method A	≤60%		>60%	72%	100%	75%
Method B	≤70%		>70%	78%	83%	79%
Method C (EPA)	≤50%	>50% ≤70%	>70%	100%[1] 19%[2]	83%	42%
Method C (GHS)	≤50%	>50% ≤70%	>70%	78%[1] 22%[2]	83%	58%

[1] Denotes the accuracy of predicting the *in vivo* Category 2/II.
[2] Denotes the *in vivo* Category 3/III.

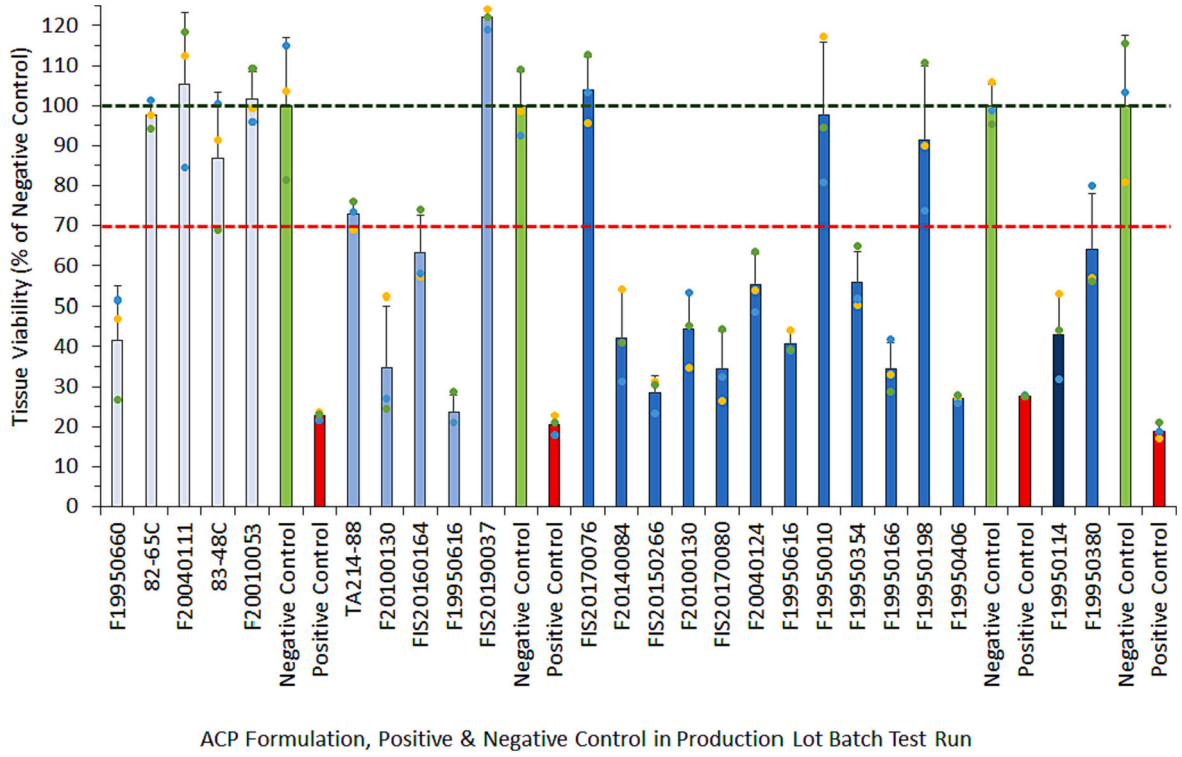

Fig. 4. Tissue Viability (% of Negative Control) in the Phenion® FT Model for 24 ACP Formulations and Respective Positive and Negative Controls.

F19950406 were tested on Phenion® FT Production Lot No. Phe-HM-21-14. Formulations F19950114 and F19950380 were tested on Phenion® FT Production Lot No. Phe-HM-22-01.

The ACP formulations tested passed all OECD TG No. 439 (OECD, 2021) acceptance criteria except for ACP formulation No. TA214–88 (Batch No. PS# 2019.203) where the variation (SD) for the test item group was 20%, *i.e.*, greater than the 18% acceptance criterion. Therefore, this ACP formulation was tested again, and the repeat data set was accepted as it fulfilled all acceptance criteria and is reported here. The absolute mean OD_{570} of the three tissues of the negative control was within the acceptance limits (lower acceptance limit $\geq$0.6 and upper acceptance limit $\leq$2). The mean cell viability after 15 $\pm$ 0.5 min exposure of the positive control was $\leq$40% for all tissue lots. The SD value of the percentage viability of the three tissues treated identically with the positive controls were $\leq$ 18% for all tissue lots. Therefore, the positive control irritant chemical, SDS, caused toxicity to the Phenion® FT cells

Table 5
Comparison of Phenion® FT Category with EPA and GHS Categories for 24 ACP Formulations.

Formulation	GHS Category (OECD TG 404)	Phenion® FT Category (OECD 439)*	EPA Category (OPPTS 870.2500)¥	Phenion® FT EPA Category	EPA Prediction
82-65C	No Category	No Category	IV	IV	Yes
F20040111	No Category	No Category	IV	IV	Yes
83-48C	No Category	No Category	IV	IV	Yes
TA214–88	No Category	No Category	IV	IV	Yes
FIS20190037	No Category	No Category	IV	IV	Yes
FIS20160164	No Category	Category 1/2	IV	I, II, III	No
F20010053	Category 3*	No Category	III	IV	No
FIS20170076	Category 3*	No Category	III	IV	No
F19950198	Category 3*	No Category	III	IV	No
F19950010	Category 2	No Category	III	IV	No
F20140084	Category 3*	Category 1/2	III	I, II, III	Yes
FIS20170080	Category 3*	Category 1/2	III	I, II, III	Yes
F20040124	Category 3*	Category 1/2	III	I, II, III	Yes
F19950354	Category 3*	Category 1/2	III	I, II, III	Yes
F20100130	Category 3*	Category 1/2	III	I, II, III	Yes
F20100130	Category 3*	Category 1/2	III	I, II, III	Yes
F19950616	Category 2	Category 1/2	III	I, II, III	Yes
F19950616	Category 2	Category 1/2	III	I, II, III	Yes
FIS20150266	Category 2	Category 1/2	III	I, II, III	Yes
F19950166	Category 2	Category 1/2	III	I, II, III	Yes
F19950406	Category 2	Category 1/2	III	I, II, III	Yes
F19950380	Category 2	Category 1/2	III	I, II, III	Yes
F19950114	Category 2	Category 1/2	II	I, II, III	Yes
F19950660	Category 2	Category 1/2	II	I, II, III	Yes

* OECD TG No. 439 (OECD, 2021) has not been validated to predict GHS Category 3.
¥ U.S. EPA Health Effects Test Guidelines, OPPTS 870.2500.

within the acceptable variability limits of the assay. The SD value of the percentage viability of the three tissues treated identically with the negative controls were ≤ 18% for all tissue lots. Therefore, the test system was confirmed to be functioned within the acceptability criteria set out for this assay.

Based on the concordance between *in vivo* and *in vitro* irritation data in the assessed models, 70% cell viability (Method B, with an overall accuracy of 79%) was selected as the most suitable cut off for discriminating non-irritants (Class IV) from mild and moderate irritants (EPA Class III and II). The results are summarised in Table 5 against the EPA and GHS Categorizations using the original conventional animal results.

Since the objective was to be able to identify EPA Class IV (GHS No Category) ACP formulations, the sensitivity, specificity, and accuracy was calculated based on a combined EPA Category I (corrosive), II (moderate irritant) and III (mild irritant) against EPA Category IV (non-irritant) and compared with the conventional EPA *in vivo* categorizations. The sensitivity, specificity, and accuracy were 78% (14/18), 83% 5/6) and 79% (19/24), respectively.

Two of the test formulations were tested twice. F20100130 (*in vivo* GHS Category 3, EPA Class III) was identified as a GHS Category 1/2 (EPA Class I, II, III) on both occasions with mean viability of 35% and 44%. F19950616 (*in vivo* GHS Category 2, EPA Class III) was identified as a Category 1/2 (EPA Class I, II, III) on both occasions with mean viability of 24% and 41%. In total, there were 4 batches of Phenion® FT tested with the positive control, SDS. The mean viability (calculated as the mean of the mean of each of the 4 triplicate tests) was 23% (SD 4%) with a CV of 16%, and all values were well below the 70% viability cut off for the positive control. This confirmed that the cutting technique either supported the access of the MTT to the keratinocytes or egress of the reduced MTT back into the MTT solution, or there was no need for this additional step. Although more repeatability testing would be recommended as the assay gains wider use, these data and the reproducibility of the positive control samples demonstrates that the assay was repeatable, inferring the suitable assay robustness criterion for NAM confidence (van der Zalm et al., 2022) and outperforming the rabbit *in vivo* test (Rooney et al., 2021).

Table 6
Comparison of EpiDerm™ Categorization of Skin Irritation to Classification from an *In Vivo* Animal Test for 10 ACP Formulations.

Formulation	Active Ingredient	EPA Category (OECD TG 404)	GHS Category (OPPTS 870.2500)	EpiDerm™ Category (OECD 439)[¥]
82-65C	Sodium hypochlorite	IV	No Category	Category 1/2[¥]
F20040111	Hydrogen peroxide	IV	No Category	No Category
83-48C	Sodium hypochlorite	IV	No Category	Category 1/2[¥]
TA214–88	Quaternary Ammonium Compounds	IV	No Category	Category 1/2[¥]
FIS20160164	Quaternary Ammonium Compounds	IV	No Category	Category 1/2[¥]
FIS20190037	Citric acid	IV	No Category	No Category
F20010053	Quaternary Ammonium Compounds	III	Category 3[*]	Category 1/2[¥]
F20100130	Glycolic acid	III	Category 3[*]	Category 1/2[¥]
F19950660	Sodium hypochlorite	II	Category 2	Category 1/2[¥]
F19950616	Sodium hypochlorite	III	Category 2	Category 1/2[¥]

[*] OECD TG No. 439 (OECD, 2021) has not been validated to predict GHS Category 3.

[¥] When the EpiDerm™ test identifies a positive result, a definitive categorization is not possible. According to the guideline, this will be categorized as Category 1 or Category 2. An additional *in vitro* corrosion test, OECD Test Guideline No. 431 (OECD, 2014) is needed to distinguish between these categories and give a definitive categorization, *i.e.*, Category 1 or Category 2, and this is detailed in the OECD IATA for skin corrosion and irritation (OECD, 2017).

4. Discussion

The existing humanised 3D *in vitro* test method for skin irritation (OECD, 2021) has been extensively validated for single chemicals. However, while these models have been used for internal decision-making, there is less experience and data to support their use for ACP mixtures in a regulatory context. Therefore, a direct comparison between the EpiDerm™ protocol (EpiDerm™ (2009)) and historical data from studies performed according to OECD TG No. 404 (OECD, 2015) was initiated. The results for this first set of ACP formulations demonstrated that the EpiDerm™ model, using OECD TG No. 439 (OECD, 2021), resulted in skin irritation categorizations much higher (*i.e.*, indicating a higher hazard) than the corresponding *in vivo* test data using OECD TG No. 404 (OECD, 2015). The results are summarised in Table 6.

The same binary system *i.e.*, EPA Category I, II, III *versus* EPA Category IV was performed. The sensitivity, specificity, and accuracy were calculated to be 100% (4/4) 33% (2/6) and 60% (6/10). The specificity was considered too low for further evaluation. The reasons for the lower performance were not investigated further and should not be considered as a reason not to use the standard RhE test method for other formulations and mixtures. Other reconstructed human epidermis models approved for use in OECD TG No. 439 (OECD, 2021) were expected to perform similarly since their performance has been compared in their "me-too" validation tests (Zuang et al., 2002; Spielmann et al., 2007).

The ACP chemistry is important to consider. These formulations contain water, ethanol, short and long chain alcohols, and anionic, cationic, and non-ionic surfactants. Skin hydration has long been known to increase skin penetration (Behl et al., 1980). Ethanol is a good penetration enhancer for dermal absorption (Schmook et al., 2001; Gupta et al., 2020). Short and long chain alcohols and anionic, cationic, and non-ionic surfactants are also penetration enhancers (Lane, 2013). Since inert ingredients including solvents and surfactants have the potential for increasing cytotoxicity (Rhein, 2007), this can result in higher levels of cytotoxic chemicals reaching the keratinocytes resulting in lower viability scores in RhE models, which would not be expected in humans with a more robust dermal barrier function. As a full thickness model containing keratinocytes and fibroblasts with more cells overall than the epidermal RhE models, the Phenion® FT model is expected to have a more robust barrier function. In addition, the Phenion® FT model was deemed to express good levels of skin metabolising enzymes (Wiegand et al., 2014).

The *stratum corneum* barrier function of the approved RhE model test systems was considered important for the ACP formulations. In an early skin absorption pre-validation study (Schäfer-Korting et al., 2006), the absorption of caffeine and testosterone was compared between three RhE models, EpiDerm™, EpiSkin™ and SkinEthic™ and with human epidermis over a 6 h exposure period using methods based on OECD TG No. 428 (OECD, 2004). The lag time for the RhE models were shorter, or non-existent for EpiDerm™ with testosterone than with human epidermis for both chemicals. Caffeine absorption through EpiDerm™ (4.87 μg/cm^2) was 4.3-fold greater than through human epidermis (1.12 μg/cm^2). Testosterone absorption through EpiDerm™ (2.36 μg/cm^2) was 7.4-fold greater than through human epidermis (0.32 μg/cm^2). The absorption of these chemicals was overestimated when using RhE and the higher absorption values suggested that the barrier functions of the reconstructed tissues were considerably less developed than the human epidermis (Mertsching et al., 2008). Although the Phenion® FT model has not been extensively tested as a model for skin absorption, as this is a full thickness model and the manufacturers have developed a version of this for skin absorption testing, it is anticipated that this test system will have an enhanced barrier function when compared to the

RhE model. The Phenion® FT model showed an improved ratio of absorption for testosterone, caffeine, benzoic acid, and nicotine than the RhE models when standardized to pig skin (Ackermann et al., 2010).

Species differences in the skin permeability of animal and human skin have been identified to be related to skin structure; *stratum corneum* thickness and number of cell layers, epidermal and dermal thickness, and area of hair follicle openings (Scott et al., 1991). These differences are also observed for rabbit skin's higher hair follicle density and more permeable *stratum corneum* (Bartek et al., 1972). Therefore, the rabbit test is already overly conservative with relation to human hazard/ irritation potential. Further, the Draize skin irritation test has been shown to have especially poor reproducibility (<50%) for mild and moderate irritation predictions (Rooney et al., 2021). The Draize rabbit test accurately predicted severe human skin irritants and non-irritants but failed to separate the mild and moderate skin irritants when compared to a similar protocol with human volunteers (Phillips et al., 1972). This was confirmed more recently, when the Draize rabbit skin irritation test was evaluated for its robustness (Rooney et al., 2021). The authors stated "Chemicals classified as moderate irritants at least once were classified as mild or non-irritants at least 40% of the time when tested repeatedly. Variability was greatest between mild and moderate irritants, which both had less than a 50% likelihood of being replicated". They also suggested "variability present in the rabbit skin irritation test should be considered when evaluating nonanimal alternative methods as potential replacements".

van der Zalm et al. (2022) suggests that fitness for purpose, human biological and, if appropriate, mechanistic relevance, technical characterization, data integrity and transparency, and independent review are all essential when establishing scientific confidence in new approach methodologies for regulatory use. The Phenion® FT test system, with a new % viability cut off, does fulfil the fitness for purpose for ACP formulations and human biological relevance criteria (Phenion® FT is derived from human keratinocytes and fibroblasts). The model is also aligned mechanistically with skin irritation processes. The toxic components of the ACP formulations must first cross the skin barrier, *i.e.*, the *stratum corneum* and once in the viable epidermis, these toxicants can produce their cytotoxic effects on the dermal keratinocytes. It is the viability of keratinocytes as well as fibroblasts that are measured in this assay.

A 2-tier testing integrated approaches to testing and assessment (IATA) can be visualised with all formulations evaluated using the modified OECD TG No. 439 (OECD, 2021) using Phenion® FT as the test system and an EPA Category IV acceptance criterion of >70% cell viability. In this proposed IATA, any ACP formulation identified as EPA Category IV would not require further testing. However, if the formulation was identified as an irritant, since this model does not yet discriminate between classes, further evaluation would be required, or an EPA Category I/II waiver could be applied. Although it is not possible to identify a cut-off value to distinguish between a Category 1/2 or Category 3, an arbitrary mean viability of, for example, <50% could be used to choose the latter option. This <50% cut off is used in OECD (2021), which adds an additional weight to this suggestion. This work could fit into the current skin corrosion and irritation IATA (OECD, 2017) taking into consideration the effects that antimicrobial formulation mixtures have on the epidermal *in vitro* test systems with the improved skin barrier function properties of the Phenion FT® test system. This approach will reduce animal use whilst maintaining a rigorous hazard classification and minimising harm to animals in the Draize test.

The differences observed between some of the classifications between the *in vitro* and *in vivo* results were further considered. Other than *in vitro-in vivo*, the most obvious difference between the tests is the species difference, *i.e.*, rabbit *versus* human and this has been long known in skin irritation testing (Phillips et al., 1972). In this work, the rabbit test accurately predicted severe human skin irritation, but failed to discriminate the mild and moderate skin irritants. Dermal irritation potential for antimicrobial formulations range from EPA category IV (mild or slight irritation), to EPA category III (moderate irritation), and EPA category II (severe irritation); EPA Category I denotes dermal corrosion potential. The *in vivo* studies were performed to regulatory standards using relevant EPA/OECD test guidelines. However, *in vivo* tests are subjective, *i.e.*, analysis is performed by a pathologist, albeit using a standardized pathology scaling, however, the interpretation of the scale is subjective for any individual pathologist. Conversely, the *in vitro* tests are objective as they use analytical measurements, *i.e.*, the spectrophotometric colour changes measuring the MTT reduction by the viable cells using the TECAN Infinite® M200 Pro Plate Reader. Vast variability is known to be present in the *in vivo* tests (Rooney et al., 2021). For example, even regulatory compliant tests may use animals of a slightly different age or weight range, or different strain. The food provided to the animals may be different and there can be differences in water quality and housing conditions (*e.g.*, European housing have been traditionally more spacious than other geographies). These test method differences were considered to be attributable to the differences observed between *in vitro* human and *in vivo* rabbit data.

5. Conclusions

In conclusion, concordance between *in vivo* and *in vitro* irritation data was demonstrated to be suitable (*i.e.*, sensitivity 76%, specificity 83%, and accuracy 78%) using the modified OECD Test Guideline No. 439 using the Phenion® FT test system with a 70% cell viability selected as the most appropriate cut off for discriminating non-irritants (Class IV). These results were considered suitable to develop a draft IATA *i.e.*, with any ACP formulation identified as EPA Category IV in this test, then no further testing would be required and if identified as an irritant, then a waiver would be used to accept an EPA Category I, II, III.

Funding

This work was conducted in part with funding from the Carroll Petrie Foundation with the remainder funded by Clorox.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

KEP and TM are employees of The Clorox Company. CSR is a consultant working with The Clorox Company. KEP is a Special Government Employee for the National Institutes of Health (NIH) and the current chair of the Scientific Advisory Committee on Alternative Toxicological Methods (SACATM). WW is an employee of Charles River Laboratories, subcontracted to perform the in-life testing. KS was an employee of PCRM during the development, conduct and reporting of the project and is now an employee of IIVS.

Data availability

Data will be made available on request.

Acknowledgements

The authors thank the following people. Dipl. Biol. Lars Vierkotten (Head of Phenion Skin Model Production, Henkel AG & Co. KGaA, Düsseldorf, Germany) and Dr. Karsten R. Mewes (Senior Manager Alternatives to Animal Testing/Phenion Business Development, Henkel AG & Co. KGaA, Düsseldorf, Germany) for creating the Transwell® plate version of the Phenion® FT model for this project, supplying histology photos of human skin and the Phenion™ FT model and in providing advice and enthusiasm for this work. Jet Gijsbrechts, the Study Director at Charles River Laboratories, Hambakenwetering 7, 's-Hertogenbosch 5231 DD, The Netherlands, responsible for the scientific conduct of this series of tests. Jessica Ponder, Physicians Committee for Responsible

Medicine, 5100 Wisconsin Ave NW Ste 400, Washington 20016, USA for contributions to the development of the prediction model.

References

Ackermann, K., Lombardi Borgia, S., Korting, H.C., Mewes, K.R., Schäfer-Korting, M., 2010. The Phenion® full thickness skin model for percutaneous absorption testing. Skin Pharmacol. Physiol. 23, 105–112. https://doi.org/10.1159/000265681.

Bartek, M.J., Labudde, J.A., Maibach, H.I., 1972. Skin permeability *in vivo*: comparison in rat, rabbit, pig and man. J. Investig. Dermatol. 58 (3), 114–123. https://doi.org/10.1111/1523-1747.ep12538909.

Behl, C.R., Flynn, G.L., Kurihara, T., Harper, N., Smith, W., Higuchi, W.I., Ho, N.F., Pierson, C.L., 1980. Hydration and percutaneous absorption: I. Influence of hydration on alkanol permeation through hairless mouse skin. J. Investig. Dermatolol. 75 (4), 346–352. https://doi.org/10.1111/1523-1747.ep12531118.

Draize, J., Woodard, G., Calvery, H., 1944. Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. J. Pharmacol. Exp. Ther. 82, 377–390.

EpiDerm™, 2009. SOP, Version 7.0, Protocol for: *In vitro* EpiDerm™ Skin Irritation Test (EPI-200-SIT), for Use with MatTek Corporation's Reconstructed Human Epidermal Model EpiDerm (EPI-200). Available at. https://www.mattek.com/wp-content/uploads/EPI-200-SIT-Skin-Irritation-MK-24-007-0023.pdf.

Gupta, R., Badhe, Y., Rai, B., Mitragotri, S., 2020. Molecular mechanism of the skin permeation enhancing effect of ethanol: a molecular dynamics study. RSC Adv. 10, 12234–12248. https://doi.org/10.1039/D0RA01692F.

Lane, M.E., 2013. Skin penetration enhancers. Int. J. Pharm. 447 (1), 12–21. https://doi.org/10.1016/j.ijpharm.2013.02.040.

Mertsching, H., Weimer, M., Kersen, S., Brunner, H., 2008. Human skin equivalent as an alternative to animal testing. GMS Krankenhhyg Interdiszip 3 (1). Doc11. PMID: 20204113; PMCID: PMC2831516.

Mewes, K.R., Raus, M., Bernd, A., Zöller, N.N., Sättler, A., Graf, R., 2007. Elastin expression in a newly developed full-thickness skin equivalent. Skin Pharmacol. Physiol. 20 (2), 85–95. https://doi.org/10.1159/000097655.

OECD, 2004. Test No. 428: Skin Absorption: *In Vitro* Method. OECD Guidelines for the Testing of Chemicals. Section 4,. OECD Publishing, Paris. https://doi.org/10.1787/9789264071087-en.

OECD, 2014. Test No. 431: *In Vitro* Skin Corrosion: Reconstructed Human Epidermis (Rhe) Test Method. OECD Publishing, Paris. https://doi.org/10.1787/9789264224193-en.

OECD, 2015. Test No. 404: Acute Dermal Irritation/Corrosion. OECD Guidelines for the Testing of Chemicals. Section 4,. OECD Publishing, Paris. https://doi.org/10.1787/9789264242678-en.

OECD, 2017. Guidance Document on an Integrated Approach on Testing and Assessment (IATA) for Skin Corrosion and Irritation. OECD Series on Testing and Assessment, No. 203. OECD Publishing, Paris. https://doi.org/10.1787/9789264274693-en.

OECD, 2021. Test No. 439: *In Vitro* Skin Irritation: Reconstructed Human Epidermis Test Method. OECD Guidelines for the Testing of Chemicals. Section 4,. OECD Publishing, Paris. https://doi.org/10.1787/9789264242845-en.

Phillips, L., Steinberg, M., Maibach, H.I., Akers, W.A., 1972. A comparison of rabbit and human skin response to certain irritants. Toxicol. Appl. Pharmacol. 21 (3), 369–382. https://doi.org/10.1016/0041-008X(72)90157-3.

Rhein, L., 2007. Chapter C.3 surfactant action on skin and hair: cleaning and skin reactivity mechanisms. In: Johansson, I., Sonasundaran, P. (Eds.), Handbook for Cleaning/Decontamination of Surfaces. Elsevier B.V, pp. 305–369.

Rooney, J.P., Choksi, N.Y., Ceger, P., Daniel, A.B., Truax, J., Allen, D., Kleinstreuer, N., 2021. Analysis of variability in the rabbit skin irritation assay. Regul. Toxicol. Pharmacol. 122, 104920 https://doi.org/10.1016/j.yrtph.2021.104920.

Schäfer-Korting, M., Bock, U., Gamer, A., Haberland, A., Haltner-Ukomadu, E., Kaca, M., Kamp, H., Kietzmann, M., Korting, H.C., Krächter, H.U., Lehr, C.M., Liebsch, M., Mehling, A., Netzlaff, F., Niedorf, F., Rübbelke, M.K., Schäfer, U., Schmidt, E., Schreiber, S., Schröder, K.R., Spielmann, H., Vuia, A., 2006. Reconstructed human epidermis for skin absorption testing: results of the German prevalidation study. Altern. Lab. Anim 34 (3), 283–294. https://doi.org/10.1177/026119290603400312.

Schmook, F.P., Meingassner, J.G., Billich, A., 2001. Comparison of human skin or epidermis models with human and animal skin in in-vitro percutaneous absorption. Int. J. Pharm. 215, 51–56. https://doi.org/10.1016/s0378-5173(00)00665-7.

Scott, R.C., Corrigan, M.A., Smith, F., Mason, H., 1991. The influence of skin structure on permeability: an intersite and interspecies comparison with hydrophilic penetrants. J. Investig. Dermatol. 96 (6), 921–925. https://doi.org/10.1111/1523-1747.

Spielmann, H., Hoffmann, S., Liebsch, M., Botham, P., Fentem, J.H., Eskes, C., Roguet, R., Cotovio, J., Cole, T., Worth, A., Heylings, J., Jones, P., Robles, C., Kandárová, H., Gamer, A., Remmele, M., Curren, R., Raabe, H., Cockshott, A., Gerner, I., Zuang, V., 2007. The ECVAM international validation study on *in vitro* tests for acute skin irritation: report on the validity of the EPISKIN and EpiDerm assays and on the skin integrity function test. Altern. Lab. Anim 6, 559–601. https://doi.org/10.1177/026119290703500614.

UN, 2021. Globally Harmonized System of Classification of Labeling of Chemicals (GHS) 9th Revision. https://unece.org/sites/default/files/2021-09/GHS_Rev9E_0.pdf.

US EPA, 2016. Office of Pesticide Programs: Label Review Manual. Chapter 10: Worker Protection Labeling.

US EPA, 2018. Label Review Manual - Chapter 7: Precautionary Statements.

van der Zalm, A.J., Barroso, J., Browne, P., Casey, W., Gordon, J., Henry, T.R., Kleinstreuer, N.C., Lowit, A.B., Perron, M., Clippinger, A.J., 2022. A framework for establishing scientific confidence in new approach methodologies. Arch. Toxicol. 96, 2865–2879. https://doi.org/10.1007/s00204-022-03365-4.

Wiegand, C., Hewitt, N.J., Merk, H.F., Reisinger, K., 2014. Dermal xenobiotic metabolism: a comparison between native human skin, four *in vitro* skin test systems and a liver system. Skin Pharmacol. Physiol. 27 (5), 263–275. https://doi.org/10.1159/000358272.

Zuang, V., Balls, M., Botham, P.A., Coquette, A., Corsini, E., Curren, R.D., Elliott, G.R., Fentem, J.H., Heylings, J.R., Liebsch, M., Medina, J., Roguet, R., van de Sandt, J.J., Wiemann, C., Worth, A.P., 2002. Follow-up to the ECVAM prevalidation study on *in vitro* tests for acute skin irritation. The European Centre for the Validation of alternative methods skin irritation task force report 2. Altern. Lab. Anim 30 (1), 109–129. https://doi.org/10.1177/026119290203000110.