Application of In Vitro Methods in
Preclinical Safety Assessment of
Skin Care Products

Gertrude-Emilia Costin and Kimberly G. Norman

Contents

Introduction ........................................ 1504
Preclinical Safety Assessment Framework for Skin Care Products: Integrating In Vitro Testing Strategies ................. 1504
In Vitro Methods Used in the Preclinical Safety Assessment of Skin Care Products .......... 1506
Skin Irritation ...................................... 1506
Skin Sensitization .................................. 1516
Conclusions ....................................... 1518
References ....................................... 1519

Abstract

One of the critical responsibilities of cosmetic and personal care industry is to determine the safety profile of the ingredients and/or formulations before launching new products on the market for consumers’ use. While products manufactured by other industries are thoroughly regulated (pharmaceuticals, pesticides, etc.), the safety assessment of cosmetic and personal care products seems to be less strictly integrated in the regulatory framework, despite the fact that the type of testing methods allowed for use became more restrictive in recent years. As such, a ban on animal testing of cosmetic ingredients and final formulations in the European Union (EU) took effect between 2009 and 2013. Thereon, industry used testing strategies based on nonanimal methods that were often designed to assess the safety profile of specific product lines. A diverse range of in vitro methods is now available and considered suitable to provide reliable interpretation of the safety data regarding ingredients used in finished cosmetic and personal care products. These methods range from simple cell monoculture test systems to more complex such as explants or three-dimensional reconstructed organotypic tissue models. This chapter discusses the use of several in vitro methods in the preclinical safety assessment of skin care products with special emphasis on skin irritation and sensitization endpoints.

G.-E. Costin (✉)
The Institute for In Vitro Sciences, Inc. (IIVS), Gaithersburg, MD, USA
E-mail: ecostin@iivs.org

K.G. Norman
Norman Institute for In Vitro Sciences, Gaithersburg, MD, USA
E-mail: knorman@iivs.org

© Springer-Verlag Berlin Heidelberg 2017
M.A. Farage et al. (eds.), Textbook of Aging Skin,
DOI 10.1007/978-3-662-47398-6_130
Introduction

Cosmetic industry faces nowadays a significant increase in the consumers’ demand for safe, efficacious, novel, and organic products. The consumers also became more conscious of the means industry uses to assess the safety and efficacy of such products. Thus, claims of using green chemistry or nonanimal testing methods are now noticed on the products labels. Skin care formulations represent a large percentage of the personal care industry’s production, and antiaging products seem to be situated at the luxury end of the cosmetics line. Regardless of the intended use, complexity of the formulation process, or consumer demand, industry is responsible of assessing the safety of all products launched to the public using the most modern available laboratory techniques.

The current EU legislation bans all types of animal testing for cosmetic ingredients, final formulations, and also the marketing of animal-tested cosmetics [1]. Therefore, their safety profile should be assessed using appropriate alternative methods that replace animal testing. Consequently, there is an increasing interest in using in vitro methods because they are scientifically more relevant to consumer exposure (particularly when based on cells of human origin), less time consuming, and more cost-effective and contribute to reducing the participation of human volunteers in the safety assessment process.

In general, the safety assessment of cosmetic products is not necessarily integrated into a well-defined, strict regulatory framework. This is in part due to the wide diversity of the products manufactured in terms of composition, intended use, targeted population, etc. In time it became necessary to generate a “checklist” of endpoints that are critical for the safety assessment of cosmetic products. Of these endpoints, skin irritation and sensitization occupy a central place and are one of the first to be addressed. This chapter discusses in chemico, in vitro cell- and tissue-based test systems that are currently used for these endpoints in the preclinical safety assessment of skin care products. The critical positioning of these in vitro methods in the larger context of the manufacturing process is also analyzed. Besides their use for safety assessment, modern in vitro methods provide relevant and reliable biological data that bring a strong scientific added value to the products (e.g., mechanistic insights). This added value further supports the decision-making processes in the research and development and/or marketing phase of the complex process that leads to the launching of a new product for consumers’ use.

Preclinical Safety Assessment Framework for Skin Care Products: Integrating In Vitro Testing Strategies

Testing strategies addressing multiple endpoints are often used for the safety assessment of skin care products before launching in order to limit the risks associated with human exposure. Although it may often be viewed as a generic endpoint checklist, it is this absolute need to have confidence in moving forward with an ingredient or formulation that actually positions the safety assessment on a central place in the complex manufacturing process of skin care products. Furthermore, it is always a crossing point in any project considering new ingredients or innovative formulation processes if those are not determined by testing to be safe. In the production continuum initiated with an innovative idea and progressing up to the post-market follow-up studies, ensuring the safety is a constant requirement throughout the life of a product and plays a critical role in the decision-making process, often providing insights useful to efficacy and clinical testing in a matrix approach (Fig. 1).

For cosmetic and personal care products (including skin care lines), there is no formal, strictly regulated approach for the safety assessment. This evaluation often varies based on the type of product, concentration of the ingredient (s), novelty of the composition, population targeted, intended use, frequency and duration of contact, normal conditions of use, and any foreseeable misuse. In general, the toxicological profile of a cosmetic ingredient/formulation is obtained by analyzing data provided by in vitro
and clinical testing, as well as results of epidemiological studies or reports of intended or accidental human exposure experience (market follow-up studies, input from factory workers, beauticians, etc.) where available.

Despite an apparent loosely regulated environment, it remains industry’s utmost responsibility toward the consumer to conduct safety testing of the formulations and ingredients contained therein in order to avoid adverse reactions with general use. While animals are still used particularly in the pharmaceutical industry for the preclinical safety testing, the personal care and cosmetic industry was open to rapidly adopt alternative in vitro methods, which reduce, refine, and replace the use of animals. The use of validated in vitro assays is of particular interest in the effort to comply with the most recent legislative measures banning the marketing of the products that contain ingredients tested in animal models (i.e., Regulation (EC) No 1223/2009). Furthermore, industry was the initiator and driver of the development of a wide array of in vitro assays routinely used for internal safety evaluations. Thus, industry created cutoff values of relevant endpoints for specific lines of products and immediately acknowledged the need to qualify through repeated testing benchmark (reference) materials that could be used for prototype exploration. It became the companies’ decision as of how to design custom-made testing strategies to ensure the safety of their products and to optimize the effectiveness of the process. Some companies may use in vitro assays as a screening tool before initiating a clinical test, while others may use a single assay or a tiered testing strategy to assess a single or multiple endpoints without further clinical confirmation (often used to address reformulations). The strategy of choice often depends on historical data available for the ingredients or formulations, company’s general approach as far as the production timeline is concerned, sequence of steps leading to launching, on budget and experience with available in vitro assays, and data interpretation (Fig. 1).

Safety testing strategies based on validated regulatory or non-regulatory in vitro assays serve a multitude of purposes within the cosmetic and personal care industry and are flexible enough to address a large variety of questions. For example, complex in vitro methods are often used by R&D, New Technology, and Discovery groups very early in the products’ design process to gain knowledge on certain ingredients’ mechanisms of action, synergistic effects, and new pathways

---

**Fig. 1** Integration of the in vitro preclinical safety testing methods in the skin care products manufacturing framework. In gray: concepts and methods discussed in the chapter. Dashed lines indicate indirect contribution to other components of the framework.
to explore for cosmetic use. Product Development and Innovation groups use in vitro assays as a tool for screening libraries of compounds that may have been previously considered of interest to another industry (e.g., pharmaceuticals). Following this approach, often used when the ingredients are novel and never used for cosmetic purposes, a so-called short list of ingredients that are determined to be safe for consumer’s use can be generated and can advance to further testing as necessary. In this case, the in vitro assays are used upstream in the production framework as a first informative step that provides not only safety information but also insights regarding any restrictions (concentration, skin area of interest) that could support the efficacy or clinical testing employed downstream for further confirmatory testing (Fig. 1). Often times, industry may decide to bypass the clinical testing entirely and to rely on data provided by in vitro assays for ingredients that have a relatively known, well-characterized safety profile or when considering minor changes of basic formulations that may have been previously tested. Since the in vitro tests used for safety assessment are usually performed during the early phases of new product development, the results are generated relatively fast and allow performing (re)formulation(s) and ingredient characteristic screening within ranges that could not be otherwise accommodated by clinical studies.

For cosmetic product manufacturers, various in vitro test systems of increasing complexity can be used to address numerous endpoints of interest as part of an integrated testing strategy. For example, cocultures of different cell types or ex vivo and/or three-dimensional (3D) organotypic tissue models can be used for confirmation or further in-depth assessment of effects initially identified in monolayer cell culture systems. The following sections will focus on in chemico and in vitro assays that can be used for skin irritation and sensitization assessment. The advantages and limitations of the in vitro methods detailed in the chapter will be analyzed, and guidance on data interpretation and subsequent use in the decision-making process will be provided.

### In Vitro Methods Used in the Preclinical Safety Assessment of Skin Care Products

#### Skin Irritation

The safety assessment for skin corrosion/irritation potential of ingredients and finished products is an essential part of the toxicological evaluation prior to manufacture, transport, or marketing. The intent is to protect consumers from toxicity associated with normal product use and reasonably foreseeable misuse exposures in the marketplace.

Traditionally, the in vivo corrosion/irritation test using rabbits introduced by Draize in the 1940s has been used to predict hazardous effects of substances coming in contact with human skin [2]. Throughout years of use, it was recognized that the visual grading of the effects is highly subjective and that the test overpredicts irritating effects of substances in relation to human skin [3–6]. Furthermore, the use of animals to assess skin safety has been criticized on inhumane grounds and unnecessary suffering. Thus, the combined needs to protect humans without exposing volunteers to potentially irritating products, to comply with regulations, and to reduce animal testing have led to significant efforts to develop alternative test methods that are predictive, rapid, and reproducible.

Over the years, a variety of cell culture systems have been developed and evaluated for prediction of skin corrosion and irritation potential. Simple monolayer keratinocyte cultures were investigated for their capacity to predict skin irritation using cell viability endpoints. They are useful in large-scale, first-line screening tests and are informative for potential hazards and to eliminate materials that are significantly cytotoxic. However, the monolayer systems are limited by materials’ solubility in the cell culture medium and lack the 3D architecture and epidermal–dermal interactions that are critical for skin homeostasis. This limitation was addressed by the use of explants (skin organ cultures) and, most importantly, by terminally differentiated in vitro reconstructed human skin equivalents (epidermal and full thickness) [7].
The 3D skin models can be designed to have only an epidermal compartment (containing mainly keratinocytes but also melanocytes or Langerhans cells) or can gain complexity by addition of a dermal compartment (containing fibroblasts and endothelial cells). The models are generated by growing the cultures at the air–liquid interface on de-epidermized dermis, acellular or fibroblast-populated dermal substrates such as inert filters, or collagen matrix. The cultures exhibit a stratified and cornified epidermis, with basal, spinous, and granular layers along with a functional stratum corneum, mimicking the architecture of the normal human skin (Figs. 2b1 and 3a) and allowing the direct topical application of ingredients or formulations. Although they reached a high level of architectural complexity, the models are yet to exhibit the degree of competency found in native human skin (Fig. 2b2–b6) [7–9].

Variability in the reconstructed human skin equivalents’ response to irritants and other insults is expected and may be influence, for example, by the source of the keratinocytes used to develop the...
models (neonatal foreskin or adult skin) [10, 11] or by variations in the barrier function. The models have been determined to have a higher permeability [12] compared to the native human skin. While this could be an advantage when testing mild formulations (typical for cosmetic industry), it could also result in overpredictions due to an increased penetration rate and thus higher availability to the viable keratinocytes [12].

The reconstructed human skin equivalents offer a relevant model to study basic skin biology processes such as wound repair, regulation of melanogenesis, phototoxicity, drug transfer, metabolism of topically applied products, pathogenesis of skin diseases, etc. [11, 13]. The 3D skin models also found clinical applications as grafts.

The models gained regulatory recognition through validations and can therefore be used to address the demands of regulatory authorities, animal welfare organizations, and consumers. Thus, one immediate use of the in vitro methods based on 3D models was in the regulatory setting for hazard identification and labeling of chemicals, transport of dangerous substances, and occupational safety/industrial hygiene. The methods used for these purposes usually employ a single exposure time/dose and endpoint to address a specific regulatory query (e.g., skin corrosion/irritation) rather than to provide the possibility for a comparative analysis of irritation potential between ingredients or formulations, which is more relevant for cosmetic products [14].

**Fig. 3** Use of the EpiDerm™ model for in vitro toxicity tests: the irritation effect of different materials was examined histologically after topical application on the surface of the tissues for various exposure times. The tissues have been fixed and stained with hematoxylin and eosin (H&E), and pictures were taken (400x magnification) (b), analyzed, and compared to the structure of the native human skin (a). Sterile, deionized water was used as the negative control, and it showed no adverse effect on the tissue’s architecture at either the short (4 h) or the long exposure time used (16 h). One surfactant, Triton X-100, with known irritation potential was also tested as a 1 % (w/v) dilution prepared in sterile, deionized water. The tissues treated with the surfactant showed early signs of toxicity at the short exposure time (4 h) as demonstrated by the disorganized structure of the tissue layers, sloughing of the stratum corneum and decreased thickness of the tissue. The long exposure time (16 h) to the 1 % Triton X-100 induced a progressive degradation of the tissues.
Cosmetic and personal care products frequently undergo reformulations in order to improve performance or to adjust concentrations of ingredients to levels that can be tolerated by the skin without adverse effects. Skin irritation that can be induced by minor formulation adjustments is reflected in effects of a range of severity, from near corrosive to cumulative or only sensory irritation or inflammatory reaction [15]. As a result, the assays using test systems based on 3D skin models and developed initially to examine corrosion or severe degrees of irritation were later optimized to be able to capture more subtle effects that are typical to cosmetic and personal care products of low cytotoxicity profiles. Since the current legislation imposes increasing measures of safety as well as the limitation of animal use for such testing [16], the in vitro methods based on available 3D models became almost the default test system to use as a cautious, convenient, and efficient alternative before considering the conduct of human clinical testing.

The primary goal of safety testing is to predict the results of human exposure to ingredients or formulations that is ultimately addressed in clinical studies using human volunteers (panelists). As an inherent variability exists for in vitro models, the same stands true for clinical studies, and thus a direct comparison of safety testing results may prove challenging. For example, in clinical testing the degree of response to a given substance may vary with the age of the subject as well as the location of the test site even if from the same individual [17]. The results of the clinical studies may also be affected by the mode of application (patch, occlusive or nonocclusive, rinse-off, etc.), the frequency of the application or implement used to apply the test material, the test material’s physical properties (solubility, melt point, etc.), and its concentration (tested at the full strength of dilution) [18]. As with in vivo studies, one criticism of the human skin irritation test resides with the subjective nature of visual grading of the endpoints (erythema and edema) and a high degree of inter-subject variability. Most importantly, clinical irritation studies can result in unforeseen, dangerous reactions for the individuals tested.

The human skin irritation test methods range from short, graduated skin patch tests (up to 4 h) developed to meet the regulatory labeling requirements [19, 20] to 14-day (or longer) cumulative irritation skin patch tests for assessing the irritancy of very mild products (e.g., cosmetics). Of the many clinical protocols available, the 4-h human patch test (4-h HPT) seems to be better developed and described in the literature to obtain controlled human acute skin irritation information and to further correlate with the in vitro skin irritation test data [19, 21–23]. The 4-h HPT provides the opportunity to identify single substances, mixtures, or formulations with significant skin irritation potential without recourse to the use of animals [24]. The method also provides gold standard data for future validations of alternative in vitro methods to replace the in vivo rabbit test for classification and labeling purposes in regulatory toxicology. The 14-day cumulative irritation patch test is used to assess and rank-order the skin irritation potential of milder cosmetics after repetitive exposure. For assessment of the skin sensitization potential of cosmetics, the human repeat insult patch test (HRiPT) is commonly used. Skin sensitization assessment differs from skin irritation due to the involvement of an immune response. The substance is applied to the skin, and then there is a rest period, followed by a repeated exposure to the substance after which skin reactions are scored by a dermatologist. There are different HRiPT protocols available, but the general principles of the assay remain the same.

Given the wide range of irritation potential of cosmetic ingredients or the more narrow irritation potential of final formulations, the clinical tests and in vitro methods developed for preclinical screening purposes became highly customized to specifically address effects of human exposure. Subsequently, the correlation of these data became necessary and proved rather challenging given the limited sets of data publicly available. This is the first attempt to an extended analysis that intends to provide a platform for future similar studies and to increase confidence in the capability of in vitro methods to predict human responses to ingredients and finished products of cosmetic and personal care industry.
In Vitro Assays Validated for Regulatory Purposes

The Skin Irritation Test (SIT)

Assays using 3D skin models represent one of the most promising alternatives to animal testing for regulatory labeling purposes of potential skin corrosives and irritants. Several commercially developed 3D models have been validated by the European Centre for the Validation of Alternative Methods (ECVAM) for this purpose (Fig. 2b2–b6). The validated SIT (OECD TG 439) is based on the measurement of mitochondrial activity using the 3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide (MTT) endpoint as a measure of tissue viability (Fig. 2c2). While this method can separate an irritant substance from a nonirritant, one of its major limitations is that it cannot rank the potency of an irritant chemical [25]. Despite this limitation, the SIT found its applications in the assessment of skin irritation for cosmetic products.

The study by Jirova et al. [16] analyzed paired skin irritation data generated in in vivo, in vitro, and clinically for 25 substances (from the ECVAM validation study and several other commercially available chemicals) [16]. Among the 15 substances predicted as skin irritants by the in vivo test, only five were found to be irritating by the 4 h HPT. These results confirmed the findings of an extensive study published previously by Basketter et al. [6] showing that about 40 % of the substances classified as skin irritants by the rabbit test did not trigger skin irritation when tested in the 4 h HPT. The in vitro EpiDerm™-based SIT using a 15-min exposure correlated better with the clinical data (four of the five irritants were correctly classified), while the optimized (60 min) version reached higher concordance with the rabbit test. This study was the first to advance the concept of using 4-h HPT data for the validation of alternative methods because of its relevance in detecting acute skin irritation potential to humans. While the substances analyzed in this study may not be immediately relevant to cosmetic formulations, the study provided an introductory view of the possibility to use the SIT to predict irritation in humans.

The study by Hoffmann et al. [26] investigated how in silico, in vitro (EpiDerm™- and EpiSkin™-based SIT), and in vivo data could be used in a combined fashion for the assessment of skin irritation hazard relevant to human exposure. The authors analyzed a database of 100 chemicals of which only 31 had 4-h HPT data available (2 irritants and 29 nonirritants). The data showed that the EpiSkin™-based SIT correctly predicted 19 chemicals as nonirritants and overpredicted 12, while the EpiDerm™-based SIT correctly predicted 22 nonirritants, overpredicted 8, and underpredicted 1 skin irritant. The results of this study further advanced the in vitro-clinical correlation that encouraged the industry to consider alternative assays for the prediction of skin irritation potential of ingredients and formulations.

A recent study [27] took the data analysis one step further by performing a comparative analysis of in vitro (EpiDerm™-based SIT), clinical (exaggerated human occlusive patch test), and controlled consumer usage data for whole-body, leave-on skin care products containing 1–5 % behentrimonium chloride (BTC). The BTC-containing formulations were predicted to be nonirritant to skin by the in vitro method, and these results were confirmed by clinical and post-marketing data.

In a study published last year [28], a new in vitro method using human viable skin obtained during surgery was reported. The results obtained for four chemicals (and based on histology analysis) were compared to the in vitro (EpiDerm™- and EpiSkin™-based SIT, 15-min exposure) and 4-h HPT results. The data analysis revealed an accuracy of 100 % compared to the human test, and a 75 % accuracy with the other validated in vitro methods. This study advanced yet another in vitro method to be considered and further investigated for the safety assessment of ingredients and formulations.

Even though still limited in their predictivity for irritation induced by complex cosmetic products particularly when of reduced irritation potential, the regulatory assays are occasionally used as a preclinical screening tool to assess acute skin irritation and to avoid unnecessary animal and clinical tests.
**In Vitro Assays Not Validated for Regulatory Purposes**

**The Time-to-Toxicity Assay**

Contact irritants vary significantly with regard to the relative potency to induce skin irritation. To address the irritancy ranking and to meet the typical needs of product development groups charged with the design of increasingly milder cosmetic and personal care products, the in vitro methods have been adapted to use multiple endpoints and exposure times and to allow for interpretation of potency. The assays are based on the effective time (ET₅₀ value) representing the exposure time that decreases the tissue viability by 50 % as measured by the tissue’s ability to reduce MTT (time-to-toxicity methods). While this method has not yet been validated for regulatory purposes, it is frequently used as a rapid screening tool able to rank-order the irritation potential of mild cosmetic formulations.

The time-to-toxicity method was the first one to be investigated for its potential to predict the acute skin irritation in humans, with the goal of validating an in vitro assay for such purposes [29]. A pre-validation study was performed for ten known skin irritants and ten nonirritants tested on three different tissue models: EpiSkin™, EpiDerm™, and PREDISKIN. The data were analyzed in comparison to unambiguous skin irritation classifications derived from the rabbit data included in the ECETOC (European Centre for Ecotoxicology and Toxicology of Chemicals) database [15, 30]. Of the 20 chemicals investigated, 4 h HPT human data were found only for four (dl-citronellol; 10-undecenoic acid; methylpalmitate; hydroxycitronellal) and allowed for a limited analysis of the predictive capacity when correlating the data with the human test results. All four chemicals were predicted as nonirritant to skin (nonclassified) by the 4-h HPT; hydroxycitronellal was the only one correctly predicted by the animal model, while the other three were overpredicted. The in vitro assays based on the EpiSkin™- and EpiDerm™ models correctly predicted only methylpalmitate as nonirritant and overpredicted the other three chemicals; hydroxycitronellal and dl-citronellol were correctly predicted as nonirritants to the skin even though intra-lab and inter-lab variability was reported. While the pre-validation study concluded that the method investigated was not ready for inclusion in a large-scale formal validation study, the assay was appealing to industry for preclinical screening purposes and started to be used quite frequently.

Despite not advancing through validation, the time-to-toxicity assay was used in several studies that investigated the capacity of various reconstructed tissue models to predict the irritation potential of a wide variety of cosmetic and personal care products. One such model was Skin²® ZK1301 (ATS; La Jolla, CA, USA), comprised of human dermal fibroblasts cocultured on a nylon mesh with partially stratified epidermal keratinocytes; the model is not produced anymore. Developed originally for autologous grafting in burned patients and diabetics, the Skin²® ZK1301 model was also used in two separate studies to analyze the skin irritation potential of surfactants used to formulate household cleaning products, laundry products, shampoos, and other personal care and cosmetic products [31, 32]. Both studies considered a wide range of exposure times (from 30 min to 24 h) and correlated the data with the 24-h human clinical patch test [31] and a human clinical modified repeat application soap chamber test [32], respectively. In both studies, the percentage tissue viability results obtained for the 30-min exposure correlated best with the clinical data, thus supporting the use of a relatively short exposure time for rinse-off products.

The study by Doyle et al. [32] also investigated the EpiDerm™ model and used the ET₅₀ value for correlation with the clinical data and for further comparison with results generated using the Skin²® ZK1301 model [32]. The EpiDerm™ model is a human tissue model comprised of human-derived epidermal keratinocytes cultured on permeable cell culture inserts, which allows for differentiation and formation of a multilayered, highly differentiated model of the human epidermis (Fig. 2b3). The complexity of the EpiDerm™ model was reflected in higher percentage viability values of the tissues treated with eight complex
shampoo formulations and when compared to those obtained for the SkinZK1301 after 1 h exposure. Furthermore, the data from the EpiDerm™-based test system resulted into a better correlation with relevant human clinical data.

The EpiDerm™ model is also sensitive enough to capture effects of surfactants applied topically for multiple exposure times (Costin, unpublished data and Fig. 3b). These results and the studies by Demetrulias et al. [31] and Doyle et al. [32] are very encouraging and demonstrate that the in vitro test systems based on reconstructed tissue models can serve as reliable screening tools prior to clinical studies.

The Fixed Exposure(s) Assays Using Cytokine Endpoints

Personal care and cosmetic products are often purposely formulated to be mild or extremely mild to the skin while preserving their efficacy for the intended end result. To address subtle formulation modifications, the in vitro assays based on 3D skin models have been adapted to address multiple biological inquiries by assessing additional endpoints besides the tissue viability.

The skin is a fully immunocompetent organ, capable of initiating an inflammatory reaction in response to irritants primarily driven by the release of primary cytokines such as interleukin (IL)-1α and tumor necrosis factor (TNF)-α from keratinocytes [33]. The inflammatory cascade further involves multifunctional secondary and chemotactic cytokines, growth factors, etc. that either escalate and drive the inflammatory cascade or repress the reaction through negative feedback [33–35]. IL-1α was routinely used in addition to MTT as a secondary (but more revealing) endpoint in in vitro assays based on 3D tissue models, particularly to provide insights on the irritation potential of materials that otherwise would be considered nonirritants based on simple viability assays. When using in vitro assays based on 3D skin models, the culture media can be collected and analyzed for the IL-1α released by the tissues following exposure to various materials. IL-1α can result either from an active secretion from viable cells (de novo synthesis or activated by irritants) or a passive release from damaged cells [36]. By far, the best-described mechanism by which chemicals induce skin irritation is that triggered by surfactants which can disrupt the cell membrane resulting in the release of IL-1α into the cytoplasm [37, 38].

The in vitro assays based on 3D skin models and using cytokine analysis as a secondary endpoint tend to vary widely regarding the exposure time, dosing volume, cytokines analyzed, etc. The variations are necessary for the assessment of the irritation potential of a large variety of products (leave-on or wash-off, surfactants or other complex formulations, etc.). The correlation with clinical data thus becomes even more challenging as the clinical tests are also adapted to address the respective classes of products investigated and their effect on consumers. Table 1 provides an overview of the in vitro and clinical methods analyzed in this section of the chapter for correlation purposes and captures all the variations of these methods.

In most of the studies reported thus far and using IL-1α as a skin irritation indicator (Table 1), the 3D tissues were exposed to the test materials for long exposure times (up to 24 h or longer) thus extending the detection limits of the in vitro test system to make it capable to capture irritation events induced by very mild products [12, 22, 38–42]. In these studies, the culture media was collected immediately after the completion of the exposure time(s). This strategy intends to mimic the effects of leave-on products on the skin and seems to provide a better correlation with the clinical tests that address either long human exposures or repeated applications of the test materials (24- or 48-h patch test, or 14-day or longer tests, respectively) [12, 38–40].

For example, the study by Roguet et al. [39] showed that the concordance between the in vitro tissue viability results and the associated IL-1α release data and the human irritation data was 74 %. Overall, only one out of six oils tested was overpredicted by the in vitro assay, while from the 17 emulsions tested, three were underpredicted and one was overpredicted. Of the tested mascaras, four out of nine were underpredicted and one was overpredicted as irritant to the skin. The IL-1α levels obtained from
### Table 1  Overview of reports correlating in vitro and clinical data – focus on inflammatory cytokines

<table>
<thead>
<tr>
<th>In vitro test system</th>
<th>Test material</th>
<th>Number (total)</th>
<th>Dosing volume/weight</th>
<th>Endpoint Type</th>
<th>Clinical test</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D tissue model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Clinical test</td>
<td>References</td>
</tr>
<tr>
<td><strong>EpiSkin™</strong></td>
<td>Gel, oil, emulsion, mascara</td>
<td>38</td>
<td>150 µl or 150 mg</td>
<td>Tissue viability IL-1α Others (specify)</td>
<td>48-h occlusion test</td>
<td>[39]</td>
</tr>
<tr>
<td><strong>Skin equivalent</strong></td>
<td>Lotion, mascara, shower gel, deodorant, cream, serum, toothpaste, bath oil, lip stick, makeup, shower cream</td>
<td>14</td>
<td>10 µl</td>
<td>Tissue viability IL-1α Others (specify)</td>
<td>48-h patch test (occlusive or semioclusive)</td>
<td>[40]</td>
</tr>
<tr>
<td><strong>EpiDerm™; EpiDerm™ cocultured with human dermal fibroblasts; SKIN2™ model ZK1350</strong></td>
<td>Raw materials representing the major classes of surfactants used in consumer products; prototype facial creams</td>
<td>16</td>
<td>100 µl</td>
<td>Tissue viability IL-1α Others (specify)</td>
<td>IL-6, IL-8, IL-10, SCF, c-Kit, GM-CSF, TNF-α, TNF-R1, IFN-γ, TGFβ, ICAM-1 Modified human cumulative irritation test (48-h patch test – semioclusive); Human clinical closed patch test (24 h)</td>
<td>[43]</td>
</tr>
<tr>
<td><strong>EpiDerm™</strong></td>
<td>Facial cleansing bar formulations, liquid facial cleansers</td>
<td>6</td>
<td>100 µl</td>
<td>Tissue viability IL-1α Others (specify)</td>
<td>IL-1ra Exaggerated human arm wash test</td>
<td>[44]</td>
</tr>
<tr>
<td><strong>EpiDerm™</strong></td>
<td>Surfactants, cosmetics, antiperspirants, deodorants</td>
<td>7</td>
<td>100 µl</td>
<td>Tissue viability IL-1α Others (specify)</td>
<td>AST, histology 24-h patch test (surfactants); human 14-day cumulative irritation test (cosmetics)</td>
<td>[12]</td>
</tr>
<tr>
<td><strong>SkinEthic</strong></td>
<td>Vaseline; SDS (0.2 %, 0.4 %, 0.8 %); calcipotriol; all-trans-retinoic acid</td>
<td>7</td>
<td>10 mg/cm²</td>
<td>Tissue viability IL-1α Others (specify)</td>
<td>IL-8, RNA, histology Repeated 24-h application over 3 weeks under Finn chamber patches</td>
<td>[38]</td>
</tr>
</tbody>
</table>

(continued)
Table 1 (continued)

<table>
<thead>
<tr>
<th>In vitro test system</th>
<th>Test material</th>
<th>Endpoint</th>
<th>Clinical test</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D tissue model</td>
<td>Exposure time (h)</td>
<td>Type</td>
<td>Number (total)</td>
<td>Dosing volume/weight</td>
</tr>
<tr>
<td>eEpiDerm™</td>
<td>3</td>
<td>Soaps (commercially available or experimental)</td>
<td>5</td>
<td>100 μl</td>
</tr>
<tr>
<td>eEpiDerm™</td>
<td>Up to 24</td>
<td>Antiperspirants, deodorants</td>
<td>11</td>
<td>NP</td>
</tr>
<tr>
<td>eEpiDerm™, Episkin™, Cosmital</td>
<td>1–16</td>
<td>Surfactant-based formulations, shampoos, mascaras, emulsions, gels, oils, creams</td>
<td>22</td>
<td>26 μl; 50 μl (EpiSkin only)</td>
</tr>
<tr>
<td>iSkinEthic</td>
<td>4</td>
<td>Chemicals from ECVAM pre-validation</td>
<td>50</td>
<td>100 μl</td>
</tr>
</tbody>
</table>

AST aspartate aminotransferase, ECVAM European Centre for the Validation of Alternative Methods, GM-CSF granulocyte-macrophage colony-stimulating factor, ICAM intercellular adhesion molecule, IFN interferon, IL interleukin, LDH lactate dehydrogenase, MTT 3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide, MTS 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium, inner salt, NP not provided, RNA ribonucleic acid, SCF stem cell factor, SDS sodium dodecyl sulfate, TGF transforming growth factor, TNF tumor necrosis factor, TNF-R tumor necrosis factor receptor

aCulture media collected immediately after exposure
bIMEDEX (Chaponost, France)
cLaboratoire des Substituts Cutanés
dMTT and/or LDH assessment
eMatTek Corporation (Ashland, MA, USA)
fAdvanced Tissue Sciences (ATS; La Jolla, CA, USA)
gMTS and/or Alamar Blue
hCulture media collected 24 h post-exposure
iSkinEthic, Nice, France
jEpiskin SNC (L’Oréal, Chaponost, France)
kWella/Cosmital SA (Research Company of Wella AG, Germany), Marly, Switzerland
the tissues exposed to the test materials included in this study (Table 1) varied widely for correctly predicted irritants (111.6–565.1 μg/ml), overpredicted materials (72.0–188.1 μg/ml), or underpredicted materials (15.8–215.1 μg/ml), while the range was tighter for correctly predicted nonirritants to human skin (1.65–38.4 μg/ml). The wide range of IL-1α response likely resides with the nature of the ingredient or formulation, bioavailability, kinetics of penetration, etc. and makes the correlation with human clinical data difficult for some of the products tested.

The study by Augustin et al. [40] used an in vitro test system based on the skin equivalent (SE) which mimics the skin architecture better than the dermal equivalent (DE) model. The results obtained for the SE test system and using the tissue viability endpoint for the prediction model with a single cutoff value of 80 % (and no apparent cutoff values established for IL-1α) had a 79 % concordance with the human data (Table 1). The materials tested covered clinically five irritation categories (eight nonirritants, one very slightly irritating, three slightly irritating, one moderately irritating, and one strong irritant). When based on the single cutoff viability value used for the data analysis, the in vitro method had a mixed predictive capacity for the materials with intermediate irritation potential. The method was however capable of predicting correctly seven out of the eight nonirritating materials and the single strong irritant. The difficulty to address the materials with intermediate irritation potential by the in vitro assays revealed the need to either introduce more cutoff values for the viability endpoint, create cutoff values for the IL-1α endpoint, or further refine the assay to address the middle-range potential.

The study by Perkins et al. [12] correlated the in vitro response of 3D tissues to human skin data for a variety of materials (Table 1). Of particular interest were the antiperspirants and deodorants for which the IL-1α release data showed the greatest capacity to distinguish irritancy over a broad range and correlated well with consumer reported irritation (14-day cumulative irritancy test). A single study analyzed the IL-1α production by tissues exposed to antiperspirants and deodorants in comparison to reported incidence (in subject diaries) data of adverse skin effects. Despite the semiquantitative nature of this analysis, a remarkable correlation between the IL-1α data and adverse responses reported with human use was observed across the entire range of clinical responses [22].

The design of the in vitro methods described thus far revealed technical aspects that could be improved to further increase the predictive capacity of the test system when relying on the IL-1α endpoint. As such, the analysis of the IL-1α after long exposure times may not be particularly relevant to certain classes of materials (surfactants) [42] or may have limited reliability for other classes, especially those of mild irritant profile [20]. Addressing the middle irritation range is critical for comparison and ranking purposes as correctly isolating nonirritants can be achieved by many other assays without the need for a secondary endpoint.

To address the ranking of very mild, wash-off products, the in vitro methods based on 3D skin models were further refined in two studies published by Bernhofer et al. [43, 44] (Table 1) by introducing a short exposure time (1 h) and a postexposure period (24 h) to allow the secretion of IL-1α. The first study showed that the full thickness model used (Skin2™ Model ZK1350, containing fibroblasts) separated irritating formulations (facial creams) from milder ones based on the expression of several cytokines [IL-1 receptor antagonist (ra), granulocyte-macrophage colony-stimulating factor (GM-CSF), and IL-8] [43]. However, these endpoints were not as efficient in identifying differences between mild irritants of this category. To address this issue, the EpiDerm™ model (without fibroblasts) was used and the IL-1α and IL-1ra responses were analyzed for the tissues with >80 % viability. The data showed that the lack of fibroblasts increased the sensitivity of the test system and enabled the detection of differences between surfactants of mild irritation potential [44]. This was particularly relevant for the IL-1ra endpoint which is part of a negative feedback system protecting against potential damage from an excess of IL-1α secretion. IL-1ra is secreted in response to topically
applied irritants but is otherwise suppressed in the presence of fibroblasts. The reliability and relevance of the test system was confirmed in the second study [44] that analyzed cleansing bars and cleansing lotions that exhibited varying degrees of irritation potential in the mild range and correlated well with clinical data (Table 1).

A recent study [45] used a similar protocol to investigate the correlation of IL-1α release and clinical skin irritation results for various classes of surfactants. The manuscript advances the field significantly by analyzing the data from the structure/toxicity relationship perspective and introduces several new parameters specific to surfactants in the attempt to explain the cytotoxicity of this class of ingredients very frequently used in the cosmetic industry. The manuscript opens the door to similar investigations that could address other chemistries used for personal care and skin care products.

The efforts to model the in vitro methods to best predict human exposure continued by introducing an in vitro patch test to correlate with the similar clinical test [46]. The study showed that the in vitro patch test met the specificity, sensitivity, and overall accuracy performance criteria defined for the ECVAM pre-validation study [29] and even proposed a prediction model for the data analysis. This in vitro protocol was very promising for the advancement through a formal pre-validation; however, no progress was reported in recent years.

In an effort to allow for better correlations between in vitro and clinical data, attempts have been made to refine the clinical tests as well. For example, cytokine profiling of different skin layers was investigated in hopes of providing a mechanistic view of toxicity induced by various materials. Noninvasive methods to collect samples directly from human skin in clinical studies and using simple dermal tape adsorption technique were investigated [12, 47]. This procedure has been used to investigate baseline cytokine levels in the skin, to assess normal skin condition, and to evaluate changes due to chemical insult, existing dermatitis, or sun exposure by direct comparison between cytokines (IL-1α, IL-1ra, IL-8) and classic clinical endpoints such as erythema or transepidermal water loss (TEWL) [48–50]. It is of consideration to further compare the cytokine expression in the stripped human skin and in the culture media collected from 3D skin models treated with various materials and to generate relevant and reliable prediction models.

In conclusion, multiple in vitro assays are available for the preclinical safety assessment of skin care products. They are relied upon by industry before performing clinical studies or as stand-alone; some became validated, while others are undergoing further improvements to best address the safety testing needs of cosmetic and personal care products.

### Skin Sensitization

Determination of skin sensitization potential is a critical toxicological endpoint in the safety assessment of cosmetic ingredients and formulations. Preservatives, hair dyes, and natural extracts are the most commonly implicated cosmetic ingredients in inducing an allergic reaction. These low-molecular-weight chemicals are known to cause a type IV delayed hypersensitivity reaction in the skin, also known as allergic contact dermatitis (ACD). ACD is the most prevalent form of immunotoxicity in humans with an occupational health-related adverse response cost of up to one billion per year estimated by NIOSH in 2009. Although in vivo tests including the guinea pig maximization test (GPMT) and the local lymph node assay (LLNA) have traditionally been used to assess skin sensitization, recent activity has focused on the development of novel nonanimal assays for the endpoint. European legislation and US research activities toward “twenty-first-century toxicology” are setting the standards for animal-free toxicological assessments, and the cosmetics industry has been the first to feel the effects. As per EU Regulation 1223/2009, animal testing of cosmetic ingredients and finished products is currently banned in the EU, as is the marketing of finished cosmetic products and the ingredients contained therein which were tested on animals for cosmetic safety assessment purposes. The marketing ban is creating a global
ripple effect as non-EU-based companies that employ animal testing are no longer able to sell their cosmetics in the EU. The ban on animal testing does not imply that safety testing is unnecessary but rather that scientifically sound, proven methods using cellular and computational models serve as a replacement.

Hundreds of skin sensitizers have been identified, and research has accordingly focused on the mechanistic understanding of allergic contact dermatitis and the development of assays for detection of skin sensitizers. The adverse outcome pathway (AOP) for skin sensitization describes each key event in the complex cascade leading to allergic contact dermatitis [51]. The AOP concept supports a toxicity testing paradigm focused on toxicity pathways. The molecular initiating event (MIE) and the cellular, organ, and organism responses are defined, and assays which address each of these key responses are considered for inclusion in a testing strategy for that endpoint. Specifically for skin sensitization, a chemical must first penetrate into the viable epidermis and covalently interact with proteins in the skin (MIE), leading to cellular activation and the expression of genes, cell surface markers and cytokines (cellular response), then proliferation of T cells in lymph nodes (organ response), and ultimately the clinical manifestation of ACD (organism response). Three methods, each modeling a key event in the AOP, have reached an advanced regulatory status: the direct peptide reactivity assay (DPRA) modeling protein reactivity, the KeratinoSens™ assay modeling keratinocyte activation, and the human cell line activation test (h-CLAT) modeling Langerhans cell activation and expression of related cell surface markers. Each assay has been evaluated for transferability, reproducibility, and accuracy, and following successful interlaboratory investigations has been endorsed by EURL ECVAM (European Union Reference Union Reference Laboratory for Alternatives to Animal Testing) as part of an integrated testing strategy to assess skin sensitization potential. As a culmination of these efforts, the Organisation for Economic Co-operation and Development (OECD) published the following test guidelines for nonanimal skin sensitization testing on 5 February 2015: direct peptide reactivity assay (DPRA) (OECD TG 442C) [52] and ARE-Nrf2 luciferase test method (also referred to as the KeratinoSens™ Assay) (OECD TG 442D) [53]. The human cell line activation test (h-CLAT) has reached an advanced stage of pre-validation, and OECD TG is expected to be published soon. Although each assay was initially designed for the assessment of pure substances, ongoing research aims to understand how they may also be used for assessing formulations.

A common characteristic of all chemical allergens is their covalent modification of proteins followed by immune activation. Chemical allergens are intrinsically electrophilic, or may be transformed into electrophiles, and react with nucleophilic amino acids within proteins. Therefore, research has focused on the ability to identify chemical allergens, specifically skin sensitizers, based on their reactivity with peptides/proteins. The direct peptide reactivity assay (DPRA) is an in chemico assay that identifies dermal sensitizers based on their reactivity with synthetic peptides containing either lysine (Lys) or cysteine (Cys) [54]. The assay models the first key event, protein reactivity, in the skin sensitization AOP. In the DPRA method, the test chemical is incubated with two peptides, one containing Cys and one containing Lys, for 24 h. Following this reaction period, peptide depletion is analyzed by HPLC-UV. Depletion of the peptide is then used as a quantitative measure of reactivity and may be correlated with skin sensitization potential, with minimal reactivity indicating a non-sensitizer and mean peptide depletion of the Cys- and Lys peptides of >6.38 % indicating skin sensitization potential.

The KeratinoSens™ assay is a cell-based reporter gene assay which identifies skin sensitizers by measuring the induction of a luciferase gene under the control of the antioxidant response element (ARE) derived from the human AKR1C2 gene [55]. Electrophilic chemicals are detected by inducing conformational changes in the Keap1 target protein, which in turn activates the ARE and leads to an upregulation of the luciferase reporter. The degree of upregulation is measured...
by relative luminescence, and cell viability is measured in parallel. In the adverse outcome pathway (AOP) leading to skin sensitization, this method addresses the second key event, gene expression in keratinocytes associated with the antioxidant/electrophile response element (ARE)-dependent pathway. In the KeratinoSens™ assay, cells are grown for 24 h in 96-well plates, then the medium is replaced with medium containing the test chemical at 12 dilutions ranging from 0.98 to 2000 μM, and cells are incubated for 48 h followed by measurement of luciferase activity and cell viability. The assay includes a full dose–response analysis for each chemical and is performed in three independent experiments, with each experiment being performed in triplicate. Luciferase gene induction and cell viability relative to the solvent controls included on each plate are calculated, and chemicals can be rated as positive or negative using the prediction model for the assay. A chemical is positive if it meets the following criteria: (1) it produces a statistically significant induction of luciferase activity >1.5-fold in at least 2/3 experiments, (2) the concentration where this induction is observed (EC1.5 value) is below 1000 μM, and (3) the viability at the EC1.5 value-determining concentration is greater than 70%. Alternatively, a chemical is rated as negative if it fails to meet these criteria.

The human cell line activation test (h-CLAT) is a cell-based skin sensitization assay which identifies skin sensitizers based on increased expression of CD86 and CD54 in THP-1 cells (a human leukemia cell line) in response to treatment with test chemical [56]. The h-CLAT method is proposed to address the third key event (dendritic cell activation) in the skin sensitization AOP. It is well known that dendritic cells in the skin, Langerhans cells, play a critical role in the induction of skin sensitization. Upon antigen capture, the Langerhans cells undergo maturation and migrate to the draining lymph nodes. This Langerhans cells maturation is characterized by the upregulation of cell surface markers, CD86 and CD54. In the assay, THP-1 cells are grown in 24-well plates and treated with eight concentrations of test chemical for 24 h, then cell staining with CD86 and CD54 antibodies is performed, and expression of cell surface antigens and cell viability (using propidium iodide) is analyzed by flow cytometry. Each chemical is evaluated in three independent experiments. For a test chemical to be rated as positive, the relative fluorescence intensity (RFI) of CD86 and CD54 should be greater than 150 or 200, respectively, for at least one concentration.

Thus far, over 100 chemicals have been evaluated using these assays, and the results indicate a good predictive value for each assay as compared to the available correlative in vivo and human clinical data. Data suggest that each assay may serve as a valuable preclinical screening tool to assess the skin sensitization potential of a wide range of ingredients. However, due to the complex cascade of events leading to skin sensitization, it is generally thought that an integrated testing approach combining multiple assays and in silico predictive tools will be needed to fully replace the animal-based methods. There is a significant effort underway to ascertain how the nonanimal assays may be combined to both qualitatively and quantitatively assess skin sensitization most effectively.

Conclusions

Although there has been significant focus on the regulatory applicability of nonanimal methods, the most routine use of the assays is screening during product development and evaluating products which do not require pre-market approval. Since cosmetics fall in this category, companies may choose how to design their testing strategies to ensure the safety of their products. Some companies may use the assays as a screening tool before initiating a clinical test, while others may use a single in vitro assay or a combination of assays to assess a single endpoint without further clinical confirmation. The in vitro assays presented herein address the first critical endpoints, skin irritation and skin sensitization, for assessing the safety of cosmetics and personal care ingredients and formulations. Use of these assays will be essential for understanding the skin
toxicity profile of novel ingredients and formulations before progressing into clinical testing or consumer use.

Beyond cosmetics, other industries not required by law to replace animal methods have proactively developed programs to phase out animal testing and only test on animals when required for regulatory purposes. Developing an in vitro program may be spurred by the changing regulatory landscape, ethical considerations for the reduction of animal use for safety assessments, the need for quicker and relatively inexpensive screening tools, or a combination of the above.

References


44. Bernhofer LP, Bakovic S, Appa Y, Martin KM. IL-1alpha and IL-1ra secretion from epidermal equivalents of chemicals tested in two protocols. Colloids Surf A Physicochem Eng Asp. 2015;469:166–79.


51. Mackay C, Davies M, Summerfield V, Maxwell G. From pathways to people: applying the adverse outcome pathway (AOP) for skin sensitization to risk assessment. ALTEX. 2013;30:473–86.