PERFORMANCE STANDARDS FOR APPLYING HUMAN SKIN MODELS TO *IN VITRO* SKIN IRRITATION TESTING

CONTENT

1 GENERAL PURPOSE

The purpose of Performance Standards (PS) is to communicate the basis by which new test methods, both proprietary (*i.e.*, copyrighted, trademarked, registered) and non-proprietary, can be determined to have sufficient accuracy and reliability for specific testing purposes (1). These PS, based on validated and accepted test methods, can be used to evaluate the accuracy and reliability of other analogous test methods (also referred to as “me-too” tests) that are based on similar scientific principles and measure or predict the same biological or toxic effect.

PS should adequately address to what extent the validation and acceptance criteria have been met. PS should be provided by the Management Team (MT) of a Validation Study and be used in Test Guidelines for new test methods.

2 SPECIFIC PURPOSE AND BACKGROUND

After an ECVAM pre-validation study (2) and following test optimization phases (3, 4, 5, 6), the ECVAM Skin Irritation Validation Study (SIVS) has been conducted and completed between December 2003 and August 2006. From the five in vitro methods initially evaluated, only the EPISKIN and EpiDerm skin irritation tests went through the whole pre-validation and validation process. A common test protocol was used for both skin models, and both skin models provided sufficient intra- and inter-laboratory reproducibility. In addition, the EPISKIN method showed sufficient sensitivity and specificity for a stand alone method that will be able to fully replace the in vivo skin irritation test in rabbits. The EpiDerm skin irritation test can currently be used only within a tiered testing strategy as described in the OECD TG 404 (an improvement, leading to increase of sensitivity is being evaluated in a follow-up study).

This document describes the PS that should be met by *in vitro* skin models proposed for testing the potential skin irritation hazard of chemicals. The criteria for evaluating these tests include:

- biological relevance of the *in vitro* skin model compared to human epidermis *in vivo*
- reproducibility of the results obtained when applying *in vitro* skin models for skin irritation testing
- performance of the *in vitro* skin models to classify substances according to their skin irritation potential

3 INTRODUCTION

Prior to the acceptance of new test methods for regulatory testing, validation studies are conducted to assess reliability (i.e., the extent of intra- and inter-laboratory reproducibility) and relevance (i.e., the ability of the test method to correctly predict or measure the biological effect of interest) (7)(8)(9). The purpose of the proposed PS document is to communicate the basis on which new test methods have proved to provide sufficient accuracy and reliability for specific testing purposes. PS may be used to evaluate the reliability and accuracy of new human skin models, which are based on similar scientific principles and which measure or predict the same biological or toxic effect as the validated and accepted method.

The three elements of the proposed PS are:

i. Minimum procedural standards that identify essential structural, functional, and procedural components (e.g. morphologic structure and integrity of the test system, proper controls, biological identity of key components, and expected biological responsiveness) of the validated test method. Adherence to the minimum procedural standards will help to assure that the proposed test method is based on the same concepts as the validated test method.

ii. A list of recommended reference chemicals that should be used to assess the reliability and predictivity of the proposed test method. A list including 20 commercially available compounds, tested in the ECVAM SIVS and optimisation studies is proposed.

iii. Specific test performance requirements: reliability and predictivity that should be achieved by the proposed test method when testing the proposed Reference Chemicals.

3.1 Regulatory Rationale for Using *In Vitro* Test Methods to Assess Skin Irritation

Dermal irritation is the production of reversible damage to the skin following the application of a test substance for up to 4 hours (10). Skin irritation is assessed by applying the test substance in a single dose to the skin of an experimental animal; untreated skin areas of the test animal serve as the control. The degree of irritation is read and scored at specified intervals in order to provide a complete evaluation of the effects.

Prevalidation and validation studies (2, 3, 4, 5, 6, 11, 12, 13, (14) have reported that in vitro tests employing reconstructed human skin models are able to reliably discriminate between known skin irritants and non-irritants according to EU classification system; R38, no label (15).

The test described in this PS allows the hazard identification of irritating chemicals and mixtures of isomers with high purity. It does not provide adequate information on skin corrosion, nor does it allow the sub-categorization of irritating substances as permitted in the Globally Harmonized Classification System (GHS) (16).

For a full evaluation of local skin effects after single dermal exposure, it is recommended to follow the sequential testing strategy as appended to Test Guideline 404 (10) and provided in the Globally Harmonized System (16). This testing strategy includes the conduct of in vitro tests for skin corrosion (17) and skin irritation (as described in this document) before considering the necessity of any exceptional or complimentary testing in living animals.

3.2 Principle of *In Vitro* Human Skin Model Systems for Skin Irritation Testing

The principle of the in vitro skin model irritation assay is based on the premise that irritant chemicals are able to penetrate the stratum corneum by diffusion and are cytotoxic to the cells in the underlying layers. Moreover, if the cytotoxic effect is absent or weak, a quantifiable amount of inflammatory mediators is released by the epidermis and may be used in a tiered approach to increase the sensitivity of the test.

The test material is applied topically to a three-dimensional human epidermal model, comprised of at least a reconstructed epidermis with several epidermal cells layers and a functional stratum corneum. Irritant materials are identified by their ability to decrease cell viability below defined threshold levels (e.g. 50%). As an additional measure of skin irritation, release of inflammatory mediators (e.g. Interleukin 1alpha) may be determined.

In vitro human skin model systems for skin irritation testing may be used to test solids, liquids, semi-solids and waxes. The liquids may be aqueous or non aqueous; solids may be soluble or insoluble in water. Solids should be ground to a powder before application; no other prior treatment of the sample is required. Since 60 carefully selected chemicals representing a wide spectrum of chemical classes were included in the validation of the in vitro human skin model test system for skin irritation, the method is expected to be generally applicable across chemical classes except for gases and aerosols.

4 GENERAL SKIN MODEL PERFORMANCE

4.1 *In Vitro* Human Skin Model Characteristics

Human skin models can be obtained commercially or be developed or constructed in the testing laboratory. Any new model should be validated and at least comply with the following performance standards:

General Model Conditions

Human keratinocytes should be used to construct the epithelium. Multiple layers of viable epithelial cells (basal layer, stratum spinosum, stratum granulosum) should be present under a functional stratum corneum. Stratum corneum should be multilayered containing the essential lipid profile to produce a functional barrier with robustness to resist rapid penetration of cytotoxic markers chemicals, e.g. Sodium Dodecyl Sulphate (SDS) or Triton X-100. This property may be estimated by the determination of IC_{50} or ET_{50} after application of an established cytotoxic marker chemical. The containment properties of the model should prevent the passage of material around the stratum corneum to the viable tissue, which would lead to poor modelling of the exposure to skin. The skin model should be free of contamination by bacteria, mycoplasma, or fungi.

Functional Model Conditions

1. Viability:
   the magnitude of viability is usually quantified by using MTT (18) or other metabolically converted vital dyes. In these cases the optical density (OD) of the extracted (solubilised) dye from the negative control tissue should be at least 20 fold greater than the OD of the extraction solvent alone. It should be documented that the negative control tissue is be stable in culture (provide similar viability measurements) for the duration of the test exposure period.
2. Barrier function: The stratum corneum (SC) and its lipid composition should be sufficient to resist the rapid penetration of cytotoxic marker chemicals, e.g. SDS or Triton X-100. This property can be estimated either by determination of the concentration at which a marker chemical reduces the viability of the tissues by 50% (IC_{50}) after a fixed exposure time, or by determination of the exposure time required to reduce cell viability by 50% (ET_{50}) upon application of the marker chemical at a specified, fixed concentration.
3. Morphology: An on-going histological examination of the reconstructed skin/epidermis should be performed, showing human skin/epidermis-like structure (including functional stratum corneum).
4. Reproducibility: The results of the method using a specific model should demonstrate reproducibility over time and between laboratories. The model must be capable to demonstrate correct prediction of Reference Chemicals over an extended time period.
5. Quality controls (QC) of the model: Each batch of the epidermal model used must meet defined production release criteria, among those for *viability* (cf. 1.) and for *barrier function* (cf. 2.) are most relevant. An acceptability range (upper and lower limit) for the IC_{50} or the ET_{50} must be established by the skin model supplier (or investigator when using an in-house model). Only results produced with qualified tissues can be accepted for reliable prediction of irritation effects. As an example, the acceptability ranges for EPISKIN and EpiDerm are given below:

Table 1: Examples of QC batch release criteria

	lower acceptance limit	**mean of acceptance range**	**upper acceptance limit**
EPISKIN (18 h SLS)	IC50 = 1.0 mg/ml	IC50 = 2.32 mg/ml	IC50 = 3.0 mg/ml
EpiDerm (1% Triton X100)	ET50 = 4.8 hr	ET50 = 6.7 hr	ET50 = 8.7 hr

5 TEST ACCEPTANCE CRITERIA

5.1 Test Substances

Each test substance is tested concurrently on three tissue replicates, of which the arithmetic mean relative viability compared to the mean of the negative control replicates is used to classify the test substance as R38 or no-label. A high Standad Deviation (SD) indicates defects in single tissues or inappropriate dosing. Therefore, the test is only valid, if the Standard Deviation obtained from the three concurrently tested tissues is ≤ 18%.

5.2 Control Substances

A negative control (NC) and a positive control (PC) should be tested concurrently with the test substances to demonstrate that viability (NC), barrier function and resulting tissue sensitivity (PC) of the tissues are within a defined historical acceptance range.

5.2.1 Negative Control (NC)

A non irritating NC (e.g. PBS or water) must be tested concurrently with the test substance. The negative control tissues should be stable in culture and provide similar viability measurements throughout the test chemical exposure and post-incubation periods. A minimum viability (e.g. expressed as absolute OD of the vital dye) must be established as a test acceptance criterion.

5.2.2 Positive Control (PC)

An appropriate PC should be used in the assay (e.g. 5% SDS) to avoid complete “knock-out” of the model. The range of responses to the PC must be developed and based on data obtained a sufficient number of independent experiments. In each assay, the positive control must

- be correctly classified as irritant,
- be within the established range of responses,
- and the SD of the three tissue replicates must be below a defined maximum

…or the assay must be repeated.

An example of typical ranges for two skin models used in the ECVAM SIVS (EPISKIN and EpiDerm) is given in Table 2.

Table 2: Example for range of model responses to PC (5% SDS)

	viability	**range (95% prediction interval)**	**SD**
EPISKIN	< 40%	1.5 – 32.2 (1.3 – 41.6)	≤ 18%
EpiDerm	< 20 %	3.7 – 13.8 (4.7 – 13.6)	≤ 18%

6 TEST PROCEDURE AND DATA INTERPRETATION

6.1 Test Substances

Three tissue replicates are used for each treatment and for the controls. For liquid as well as solid materials, sufficient amount of test substance must be applied to uniformly cover the skin surface, a minimum of 25 µL/cm2 or (25 mg/cm2) should be used. Solid substances should be moistened with deionised or distilled water after application to ensure good contact with the skin. If appropriate, solids should be ground to a powder before application. At the end of the exposure period, the test material must be carefully washed from the skin surface with an appropriate buffer or 0.9% NaCl.

6.2 Viability Measurements

The most important element of the test procedure is that viability measurements are not performed immediately after the exposure to the test chemicals, but after a sufficiently long post-incubation period of the rinsed tissues in fresh medium. This period allows for recovery from weakly irritant effects. During the test optimisation phase (3)(4)(5)(6), a 42 hr post-incubation period proved to be optimal and was therefore used in the ECVAM SIVS.

Only quantitative, validated methods can be used to measure cell viability. Furthermore, the measure of viability must be compatible with use in a three-dimensional tissue construct. Non-specific dye binding must not interfere with the viability measurement. Protein binding dyes and those, which do not undergo metabolic conversion (e.g. neutral red), are therefore not appropriate.

The most frequently used assay is MTT [3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Thiazolyl blue; CAS number 298-93-1)] reduction (18), which has been shown to give accurate and reproducible results. The skin sample is placed in MTT solution of appropriate concentration (e.g. 0.3 – 1 mg/mL) for 3 hours. The precipitated blue formazan product is then extracted using a solvent (isopropanol), and the concentration of formazan is measured by determining the OD at a wavelength between 540 and 595 nm.

Chemical action by the test material on the vital dye may mimic that of cellular metabolism leading to a false estimate of viability. This may occur when a test material is not completely removed from the skin by rinsing. If the test material acts directly on the vital dye, additional controls should be used to detect and correct for test substance interference with the viability measurement.

6.3 Viability Data Interpretation Procedure / Prediction Model

The optical density (OD) values obtained with each test sample can be used to calculate the percentage of viability compared to the negative control, which is set at 100%. The cut-off value of percentage cell viability distinguishing irritating from non-irritating test materials and the statistical procedure(s) used to evaluate the results and identify irritating materials, must be clearly defined and documented, and proven to be appropriate. The cut-off values were established during prevalidation and test optimisation studies and confirmed in the ECVAM SIVS. As an example, the prediction of irritation associated with the EPISKIN and EpiDerm models is given below:

> "The test substance is considered to be irritating to skin (R38), if the tissue viability after exposure and post incubation is less or equal (≤) to 50%".

6.4 Complementary endpoints

In response to physical or chemical stress, keratinocytes produce and release inflammatory cytokines interleukins [IL-1α, tumor necrosis factor α (TNF-a)], chemotactic cytokines [IL-8, interferon, e.g. induced protein 10 (IP-10)], growth-promoting factor [IL-6, IL-7, IL-15, granulocyte/macrophage

colony-stimulating factor GM-CSF)], transforming growth factor [TGF], cytokines regulating humoral versus cellular immunity [IL-10, IL-12] and other signalling factors which rapidly generate cutaneous inflammation, suggesting that measurement of such keratinocyte responses may allow the evaluation of toxicological properties of chemicals in order to identify irritants and/or sensitizers (19).

In the first and second phase of the ECVAM SIVS, IL-1α release into the assay medium was evaluated as a promising complimentary endpoint to the classic MTT cytotoxicity test (18). It was proven during the study that MTT is a more robust endpoint than IL-1 alpha (13). Although IL-1 alpha might be useful to acquire additional information on the irritating potency of chemicals, only results from the MTT assay are currently used for classification and labelling according to EU classification system. Further investigations are required to improve the reproducibility of the IL-1 alpha assay to allow combination of two endpoints for reliable prediction of irritancy.

6.4.1 Example of Interleukin 1 alpha (IL-1α) measurements in EPISKIN model

For epidermis tissues showing a cell viability > 50%, the amount of IL-1α released into the tissue culture medium at the end of the post incubation period (after 42h post-incubation) is measured in the medium (immediately or frozen) using ELISA kits DLA 50 from R&D (20) (21) (22).

The test substance is considered to be an **irritant** if the viability after 15 minutes of exposure and 42 hours of post incubation is more (>) than 50%, and the amount of IL-1α release is more (>) than 60pg/ml The test substance is considered to be **non irritant** to skin: if the viability after 15 minutes of exposure and 42 hours of post incubation is more (>) than 50%, and the amount of IL-1α release is less or equal (≤) to 60pg/ml

6.5 Test Reporting

The test report should include the following information**:**

Test and Control Substances

- Chemical name(s) such as Chemical Abstract Services (CAS) preferred name and Registry Number (RN), followed by other names, if known
- Purity and composition of the substance or preparation (in percentage(s) by weight)
- Physicochemical properties such as physical state, volatility, pH, stability, chemical class, water solubility relevant to the conduct of the study
- Treatment of the test/control substances prior to testing, if applicable (e.g., warming, grinding)
- Stability, if known

Justification of the Skin Model and Protocol Used

Test Conditions

- Cell system used
- Calibration information for measuring device used for measuring cell viability (e.g., spectrophotometer)
- Complete supporting information for the specific skin model used including its validity
- Details of test procedure used
- Test doses used
- Description of any modifications of the test procedure
- Reference to historical data of the model
- Description of evaluation criteria used

Results

- Per each treatment, data from individual replicate test samples (e.g., OD values and calculated percentage cell viability data for each test chemical, including positive and negative classification) must be reported in tabular form, including data from repeat experiments as appropriate. In addition means ± standard deviation for each trial shall be reported. Results of tests of interactions with MTT reagent and eventually IL-1α-values, if appropriate, must be reported for each tested chemical.

Description of Other Effects Observed

Discussion of the Results

Conclusion

7 REFERENCE CHEMICALS

7.1 Selection Criteria

Reference Chemicals are used to determine if the performance of a proposed *in vitro* human skin model system for skin irritation testing is comparable to that of the validated *in vitro* test method.

Therefore, the aim of the reference chemical selection process was to include, to the extent possible, qualifying chemicals that:

1. are commercially available
2. are representative of the range of irritant responses (from negative to strong positives) that the validated *in vitro* test method is capable of predicting
3. have a well-defined chemical structure
4. are representative of the validated method's **reproducibility** and **predictive capacity** as determined in the ECVAM validation study
5. include classification based on both endpoints (MTT and IL-1α release)
6. are representative of the chemical classes used in the validation process
7. are not associated with an extremely toxic profile (e.g. carcinogenic or toxic to the reproductive system)
8. and are not associated with prohibitive disposal costs

Because the Reference Chemicals are a sub-set of the chemicals used in the SIVS, several additional selection criteria were applied by the ECVAM Chemical Selection Sub Committee (CSSC) in the selection process of test chemicals used in the ECVAM SIVS (14). These comprise e.g. exclusion of rapidly polymerizing and hydrolyzing chemicals, chemical gases and aerosols.

7.2 Proposed Reference Chemicals

The 20 commercially available Reference Chemicals (10 non-irritants, 10 irritants) listed in Table 3 provide a representative distribution of the 58 chemicals used in the ECVAM SIVS. It is important to note that they also mirror the predictive performance of the validated EPISKIN test: two *false positive* (*1-bromo-4-chlorobutane* and *4-methyl-thio-benzaldehyde*s) and three *false negative* compounds (*hexyl salicylate, terpinyl acetate* and *dipropyl disulphide*) are contained in Table 3. Thus, assessment of improvements of the performance obtained with a new skin model, or modified test protocol may become possible.

Table 3 Commercially available Reference Chemicals recommended for validation of new *in vitro* human skin model irritation test methods:

Predictive performance in the MTT assay obtained with the EPISKIN model (13).

				***In vitro* Prediction**			
				MTT assay			
Chemical Name	**CAS Number**	**EU label**	***In vivo* score§**	**Lab A**	**Lab B**	**Lab C**	**Over all**
1-bromo-4-chlorobutane	6940-78-9	**no**	**0**	1	1	1	**1**
diethyl phthalate	84-66-2	**no**	**0**	0	0	0	**0**
di-propylene glycol	25265-71-8	**no**	**0**	0	0	0	**0**
naphthalene acetic acid	86-87-3	**no**	**0**	0	0	0	**0**
allyl phenoxy-acetate	7493-74-5	**no**	**0.3**	0	0	0	**0**
isopropanol	67-63-0	**no**	**0.3**	0	0	0	**0**
4-methyl-thio-benzaldehyde	3446-89-7	**no**	**1**	0	1	1	**1**
methyl stearate	112-61-8	**no**	**1**	0	0	0	**0**
allyl heptanoate	142-19-8	**no**	**1.7**	0	0	0	**0**
heptyl butyrate	5870-93-9	**no**	**1.7**	0	0	0	**0**
hexyl salicylate	6259-76-3	**R38**	**2**	0	0	0	**0**
terpinyl acetate	80-26-2	**R38**	**2**	0	1	0	**0**
tri-isobutyl phosphate	126-71-6	**R38**	**2**	1	1	1	**1**
1-decanol	112-30-1	**R38**	**2.3**	1	1	1	**1**
cyclamen aldehyde	103-95-7	**R38**	**2.3**	1	1	1	**1**
1-bromohexane	111-25-1	**R38**	**2.7**	1	1	1	**1**
a-terpineol	98-55-5	**R38**	**2.7**	1	1	1	**1**
di-n-propyl disulphide	629-19-6	**R38**	**3**	0	1	0	**0**
butyl methacrylate	97-88-1	**R38**	**3**	1	1	1	**1**
heptanal*	111-71-7	**R38**	**4**	1			**1**

* = tested only in the "optimization studies" in one laboratory (5,6)
In vitro prediction: 1 = irritant (R38) / 0 = non irritant (no label)
§ = dominant median *in vivo* score (13)
grey: classified false negative or false positive with the endpoint MTT in the reference method (EPISKIN)

8 SPECIFIC TEST PERFORMANCE: RELIABILITY AND PREDICTIVITY

When using the list of recommended Reference Chemicals (Table 3), the proposed test method should provide performance characteristics (reliability and predictivity) that are equal or better than the performance of the validated reference method. Non-irritant and irritant chemicals, ranging in their activity from strong to weak, and representing relevant chemical classes are included to allow that the performance of the proposed test method can be determined and compared to that of the validated *in vitro* test method.

8.1 Reliability

The reproducibility (within one laboratory, and over time, and preferably also between laboratories) of the proposed test method obtained with the set of Reference Chemicals should be at least comparable to that of the validated reference method. The assessment of reliability must include the use of different, independent production batches of the skin model so to assure that the performance is stable over time.

8.1.1 Reproducibility within one laboratory (and over time)

The concordance of classifications (R38 / no label) obtained in different, independent test runs of the 20 Reference Chemicals within one laboratory must be equal or better (≥) than 90%.

For example, in the ECVAM SIVS, with EPISKIN a 94% within-laboratory concordance of classifications was obtained, and a 96% within-laboratory concordance was obtained with EpiDerm.

8.1.2 Reproducibility between laboratories (and over time)

To avoid full scale repetition of validation studies, assessment of inter-laboratory reproducibility of a "me-too test" is not mandatory, if the within-laboratory reproducibility has been thoroughly established. However, any significant procedural change may cause new transferability problems with impact on between-laboratory reproducibility. In this case, assessment of the between laboratory reproducibility is highly recommended.

The concordance of classifications (R38 / no label) obtained in different, independent test runs of the 20 Reference Chemicals between three laboratories must be equal or better (≥) than 80%.

For example, with the validated EPISKIN test, 3 of 19 Reference Chemicals (Table 3) were classified discordantly across three laboratories (84% concordance), and in the ECVAM SIVS 8 of 58 test chemicals were classified discordantly (82% concordance).

8.2 Predictivity

The predictions obtained by testing the 20 Reference Chemicals (Table 3) with validated endpoint (MTT) must reveal equal or better parameters, i.e. sensitivity and specificity, than the reference method. A maximum of 30% false negative classifications and 20% false positive classifications must not be exceeded. In other words, the sensitivity of a "me-too" test must be equal or higher (≥) than 70%, and the specificity must be equal or higher (≥) than 80%.

For comparison, the overall performance obtained with the EPISKIN test in the ECVAM SIVS with 58 substances tested in three laboratories (n = 174 predictions) revealed a sensitivity of 74.7% (25,3% false negative classifications) and a specificity of 80.8 % (19.2% false positive classifications)

9 COMPLEMENTARY ENDPOINT (IL-1α RELEASE)

Complementary endpoints may be added to the assay after their reliability and reproducibility has been confirmed. Experience has shown that the combination of MTT and IL-1α in a tiered strategy seems promising (22). However, inter-laboratory reproducibility could only be obtained with a Prediction Model developed post-hoc in the SIVS (12)(13). Additional studies are required to confirm the robustness of the secondary endpoint IL-1α. Technical challenges linked to the reproducibility of commercially available ELISA kits have to be investigated and solved before the tiered strategy employing 2 endpoints will be acceptable for regulatory purposes.

To allow best use of the information from the ECVAM SIVS, Table 5 shows the increase in sensitivity obtained with the EPISKIN test if the 20 Reference Chemicals (Table 3) are tested in a tiered strategy according to section 6.4.1.

Table 5: Commercially available Reference Chemicals recommended for validation of new *in vitro* human skin model irritation test methods:

pha to increase the sensitivity of the EPISKIN assay.

				***In vitro* pretiction**						
				MTT alone			**MTT + IL-1α**			**Over all**
Chemical Name	**CAS Number**	**EU label**	***In vivo* score§**	**Lab A**	**Lab B**	**Lab C**	**Lab A**	**Lab B**	**Lab C**	
1-bromo-4-chlorobutane	6940-78-9	**no**	**0**	1	1	1	1	1	1	**1**
diethyl phthalate	84-66-2	**no**	**0**	0	0	0	0	0	0	**0**
di-propylene glycol	25265-71-8	**no**	**0**	0	0	0	0	0	0	**0**
naphthalene acetic acid	86-87-3	**no**	**0**	0	0	0	0	0	0	**0**
allyl phenoxy-acetate	7493-74-5	**no**	**0.3**	0	0	0	0	0	0	**0**
isopropanol	67-63-0	**no**	**0.3**	0	0	0	0	0	0	**0**
4-methyl-thio-benzaldehyde	3446-89-7	**no**	**1**	0	1	1	0	1	1	**1**
methyl stearate	112-61-8	**no**	**1**	0	0	0	0	0	0	**0**
allyl heptanoate	142-19-8	**no**	**1.7**	0	0	0	0	0	0	**0**
heptyl butyrate	5870-93-9	**no**	**1.7**	0	0	0	0	0	0	**0**
hexyl salicylate	6259-76-3	**R38**	**2**	0	0	0	0	0	0	**0**
terpinyl acetate	80-26-2	**R38**	**2**	0	1	0	1	1	1	**1**
tri-isobutyl phosphate	126-71-6	**R38**	**2**	1	1	1	1	1	1	**1**
1-decanol	112-30-1	**R38**	**2.3**	1	1	1	1	1	1	**1**
cyclamen aldehyde	103-95-7	**R38**	**2.3**	1	1	1	1	1	1	**1**
1-bromohexane	111-25-1	**R38**	**2.7**	1	1	1	1	1	1	**1**
a-terpineol	98-55-5	R38	**2.7**	1	1	1	1	1	1	**1**
di-n-propyl disulphide	629-19-6	R38	**3**	0	1	0	1	1	1	**1**
butyl methacrylate	97-88-1	**R38**	**3**	1	1	1	1	1	1	**1**
heptanal*	111-71-7	**R38**	**4**	1			1			**1**

* = tested only in the "optimization studies" in one laboratory (5,6)
In vitro prediction: 1 = irritant (R38) / 0 = non irritant (no label) / - not tested according to strategy
§ = dominant median *in vivo* score (13)
grey: classified false negative / false positive with the reference method (EPISKIN)

Table 5 shows for the 20 Reference Chemicals that the overall rate of false negative classifications is reduced from 30% to 10% (i.e. increase of the sensitivity from 70% to 90%) when the release of IL-1α is used as complementary endpoint for chemicals classified negative with MTT alone. At the same time, the increase in sensitivity has no impact on the 80% specificity (20% false positives)

For comparison, when the strategic combination of MTT plus IL-1α release was used the ECVAM SIVS a similar increase of the overall performance was obtained. The 58 substances tested in three laboratories (n = 174 predictions) revealed overall results shown in Table 5.

Table 5: Overall performance characteristics obtained with the EPISKIN assay in the ECVAM SIVS for 58 test chemicals tested in 3 Laboratories (n = 174)

	n	MTT	MTT + IL-1α
Sensitivity	75	**74.7%**	**90.7%**
Specificity	99	**80.8%**	**78.8%**

Although the results obtained with the strategic combination of MTT + IL-1α in the ECVAM SIVS are very promising, due to issues that need to be resolved with regard to reproducibility and definition of the IL-1α endpoint, currently performance standard requirements for the combined use of the two endpoints cannot be defined

10 REFERENCES

1 Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment . OECD Series on Testing and Assessment, N° 34, Environment Directorate, ENV/JM/Mono(2005)14.

2 Fentem, J.H., Briggs, D., Chesné, C., Elliot, G.R.,Harbell, J.W., Heylings, J.R., Portes, P., Roguet, R.,van de Sandt, J.J.M. & Botham, P. (2001). A prevalidationstudy on in vitro tests for acute skin irritation results and evaluation by the Management Team. Toxicology in Vitro 15, 57-93

3 Portes, P., Grandidier, M.H., Cohen, C. & Roguet, R.(2002). Refinement of the EPISKIN protocol for the assessment of acute skin irritation of chemicals: follow-up to the ECVAM prevalidation study. Toxicology in Vitro 16, 765–770.

4 Kandarova, H., Liebsch, M., Genschow, E., Gerner, I., Traue, D., Slawik, B. & Spielmann, H. (2004). Optimisation of the EpiDerm test protocol for the upcoming ECVAM validation study on in vitro skin irritation tests. ALTEX 21, 107–114.

5 Kandarova, H., Liebsch, M., Gerner, I., Schmidt, E., Genschow, E., Traue, D. & H. Spielmann (2005). The EpiDerm Test Protocol fort the Upcoming ECVAM Validation Study on In Vitro Skin Irritation Tests – An Assessment of the Performance of the Optimised Test. ATLA 33, 351-367

6 Cotovio, J., Grandidier, M.-H., Portes, P., Roguet, R. & G. Rubinsteen. (2005). The In Vitro Acute Skin Irritation of Chemicals: Optimisation of the EPISKIN Prediction Model within the Framework of the ECVAM Validation Process. ATLA 33, 329-249.

7 Balls, M., Blaauboer, B., Brusick, D., Frazier, J., Lamb, D., Pemberton, M., Reinhardt, C., Roberfroid, M., Rosenkranz, H., Schmid, B., Spielmann, H., Stammati, A.L. & E. Walum (1990). Report and recommendations of the CAAT/ERGATT workshop on the validation of toxicity test procedures. ATLA 18, 313-337

8 Balls, M., Blaauboer, B.J., Fentem, J.H., Bruner, L., Combes, R.D., Ekwall, B., Fielder, R.J., Guillouzo, A., Lewis, R.W., Lovell, D.P., Reinhardt, C.A., Repetto, G., Sladowski, D., Spielmann, H., & F. Zucco (1995).

Practical aspects of the validation of toxicity test procedures. The report and recommendation of ECVAM Workshop 5. ATLA 23, 129-147

9 Hartung, T., Bremer, S., Casati, S., Coecke, S., Corvi, R., Fortaner, S., Gribaldo, L., Halder, M., Hoffmann, S., Janusch Roi, A., Prieto, P., Sabbioni, E., Scott, L., Worth, A. & Valerie Zuang (2004). A Modular Approach to the ECVAM Principles on Test Validity. ATLA 32, 467-472.

10 OECD (2002). Guideline for the Testing of Chemicals,No. 404: Acute Dermal Irritation/Corrosion.13pp. Paris, France: OECD

11 Zuang, V., Balls, M., Botham, P.A., Coquette, A.,Corsini, E., Curren, R.D., Elliot, G.R., Fentem, J.H.,Heylings, J.R., Liebsch, M., Medina, J., Roguet, R.,van De Sandt, J.J.M., Wiemann, C. & Worth, A.(2002). Follow-up to the ECVAM prevalidation study on in vitro tests for acute skin irritation. ECVAM Skin Irritation Task Force Report 2. ATLA 30,109–129.

12 In preparation: Validation Management and participants of the ECVAM International Validation Study on In Vitro Tests for Acute Skin Irritation: Results and Evaluation.

13 Hoffmann, S. (2006) Skin Irritation Validation Study. Phase II: Analysis of the primary endpoint MTT and the secondary endpoint IL1-□. Version 6 (14 September 2006)

14 Eskes, C., Cole, T., Hoffmann, S., Worth, A., Cockshott, A., Gerner, I. & V. Zuang (2007). ECVAM International Validation Study on In Vitro Tests for Acute Skin Irritation: Selection of Test Chemicals. In preparation: SIVS Chemical Selection Draft v.3 (07.02.07)

15 EC, (1984). Commission Directive 84/449/EEC of 25 April 1984 adapting to technical progress for the sixth time Council Directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packing and labelling of dangerous substances; skin irritation. Official Journal of the European Communities L251, 106-108.

16 United Nations Economic Commission for Europe (2003). The Globally Harmonized System of Classification and Labelling of Chemicals (GHS).UN ST/SG/AC.10/30, United Nations, New York, NY, USA/Geneva, Switzerland: United NationsOrganisation. http://www.unece.org/trans/danger/publi/ghs (Accessed 15.8.03).

17 OECD (2004). Guidelines for the Testing of Chemicals, No. 431: In Vitro Skin Corrosion: Human Skin Model Test. 8pp. Paris, France: OECD.

18 Mosman, T. (1983) Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. Journal of Immunological Methods **65**, 55-63

19 Wiliams & Kupper (1996). Immunity at the surface: Homeostatic mechanisms of the skin immune system. Life Sci. **58**, 1485 - 1507

20 R&D Systems, Human IL-1a/IL-1F1 Immunoassay, Catalog Number DLA50. For the quantitative determination of human interleukin 1 alpha (IL-1α) concentrations in cell culture supernates, serum, and plasma. http://www.rndsystems.com/pdf/dla50.pdf (accessed 04 April 2007)

21 EPISKIN SOP, Version 1.2 (September 2005). Validation of the EPISKIN Irritation Test - 42 Hour Assay for the prediction of acute Skin Irritation of chemicals. Determination of IL-1α concentration in the culture medium. Provided *by L'ORÉAL 13 February 2007*

22 Roguet, R. & J. Cotovio (2007) Interleukin 1 alpha (IL-1α). Rationale and use Role in inflammation/irritation processes Measurement conditions specificities.
Document provided 30 March 2007 for consideration by the SIVS MT and PRP