Introduction

The potential for chemicals to cause skin corrosion or skin irritation is an important consideration in establishing procedures for the safe handling, packing and transport of chemicals. The determination of skin corrosion and irritation potential is therefore included in international regulatory requirements for the testing of chemicals. The standard approach used involves applying the test material to the shaved skin of albino rabbits (1). Testing for skin corrosion and irritation in laboratory animals has the potential to cause them severe discomfort or pain. For this reason, the prediction of chem-
ical-induced acute dermal toxicity is an area in which considerable effort has been directed toward the development and evaluation of alternative test methods in recent years (2–5).

Various types of alternative methods are being used for the assessment of potential dermal corrosion (for example, the rat skin transcutaneous electrical resistance [TER] assay, CORROSTEX™, structure-activity relationship [SAR] models, and tests employing human skin models, such as EPISKIN™ and EpiDerm™ [6, 7]) and irritation (for example, in vitro cell and tissue culture assays based on cell lines, primary monolayer cultures, skin slices, organ cultures, and reconstituted human skin equivalents [8, 9], and SAR models). The in vitro tests incorporate numerous endpoints, including cell morphology, viability, membrane damage, metabolic effects, and the release of various inflammatory mediators (5, 8).

It is becoming increasingly apparent that the development and implementation of stepwise (hierarchical) testing strategies, combining experimental data derived from a range of alternative methods (physicochemical techniques, SAR, in vitro tests, etc.) and animal tests, if necessary, provides the most effective way forward for trying to predict toxicity while at the same time reducing the number of laboratory animals used for testing purposes. Testing strategies provide a means to: a) improve the scientific basis of toxicity testing; b) implement the Three Rs, in terms of minimising the use and suffering of laboratory animals; c) maximise the use of existing knowledge; and d) optimise the use of resources. Hierarchical testing schemes have been described in the literature for eye irritation (10), skin irritation/corrosion (11), phototoxicity (12) and skin sensitisation (13).

During the updating of OECD testing guidelines, special attention is meant to be given to possible improvements in animal welfare. In this respect, the 1992 revision of OECD Guideline 404 on acute dermal irritation/corrosion (14) incorporates two major changes compared with the original, 1981, guideline: a) the use of data derived from in vitro tests for determining whether to proceed to an in vivo test; and b) the possibility of using a single animal in the first step of the in vivo procedure, thereby enabling certain chemicals to be classified as corrosive on the basis of a positive result in one animal. Thus, the current guideline allows for a tiered approach to skin irritation and corrosion testing.

The issue of testing strategies was one of the topics discussed at an OECD workshop on Harmonisation of Validation and Acceptance Criteria for Alternative Toxicological Test Methods, held in Solna, Sweden, in January 1996 (15). The testing strategy for skin irritation/corrosion proposed at the workshop is reproduced in Figure 1. This testing strategy has been considered by the OECD Advisory Group on Harmonization of Classification and Labelling, and has been included, with minor modifications (principally the inclusion of consideration of historical human and animal data as a first step), in the draft US/German proposal for a harmonised system for the classification of chemicals which cause skin irritation/corrosion (16).

The ECVAM Skin Irritation Task Force was established in November 1996 under the chairmanship of Philip Botham (ZENECA CTL, Macclesfield, UK), a member of the ECVAM Scientific Advisory Committee (ESAC). The remit given to the task force was to prepare a report for ECVAM and the ESAC on the current status of alternative test development and validation in the area of skin irritation/corrosion and, in particular, to identify any appropriate non-animal tests for predicting human skin irritation which were sufficiently well-developed to warrant ECVAM supporting their prevalidation/validation.

Three meetings of the task force were held between 1996 and 1998, on 17 December 1996, 24 February 1997, and 16 January 1998. In addition, at its second meeting, the task force specifically requested that ECVAM supported a workshop on the use of human keratinocytes and human skin models for predicting skin irritation. This was held in Utrecht, The Netherlands, on 9–11 November 1997, under the co-chairmanship of Han van de Sandt (TNO, Zeist, The Netherlands) and Roland Roguet (L’Oréal, Aulnay-sous-Bois, France). The aims of the workshop, as defined by the task force, were to review the applicability of human keratinocytes (in monolayer culture, in multilayer skin equivalents, or in organ culture) for predicting human skin irritation, and to formulate recommendations for ECVAM.
with regard to the status of development and validation of in vitro test methods incorporating human skin for irritancy testing. The conclusions and recommendations of the workshop participants were considered by the task force during the production of this report.

The task force based its discussions around the proposed testing strategy for skin irritation/corrosion emanating from the OECD workshop (15; Figure 1). In particular, the following areas were reviewed: a) SAR and structure-property relationships (SPR) for skin corrosion and irritation; b) the use of pH and acid/alkaline reserve measurements in predicting skin corrosivity; c) in vitro tests for skin corrosion; d) in vitro tests for skin irritation (keratinocyte cultures, organ cultures, and reconstituted human skin models); and e) human patch tests for skin irritation.

**Structure-activity and Structure-property Relationships**

Quantitative structure-activity relationships (QSARs) have been derived which relate skin corrosivity data on organic acids, organic bases, phenols and electrophiles to their log octanol/water partition coefficients, molecular volumes, melting points and pKa/pKb values (6, 17–19). These have proved valuable in the selection of test chemicals for use in the recent ECVAM skin corrosivity validation study (7, 19). A PC-based QSAR program for predicting the skin corrosivity potential of organic acids and bases, phenols and electrophiles is used routinely for screening purposes by Unilever (Safety & Environmental Assurance Centre [SEAC] Toxicology Unit; L.K. Earl, personal communication), and this could be considered for prevalidation/validation. There are plans to incorporate the program into the Standardised Argument Report (StAR) PC/Windows-based system for the assessment of toxicological hazard and risk (20, 21), which has been launched recently by LHASA UK.

QSARs for predicting the skin irritation potential of neutral and electrophilic organics are under development at Unilever’s SEAC Toxicology Unit (L.K. Earl, personal communication), and the UK Health and Safety Executive are currently investigating the use of SAR/SPR for predicting the irritation potential of single chemical entities (22).

Several expert systems for predicting toxicity incorporate skin irritation/corrosion as one of the endpoints (21). For example, the DEREK (Deductive Estimation of Risk from Existing Knowledge) rulebase has nine rules for the prediction of irritancy, although none of these is specific to skin irritancy or corrosivity. At a recent ECVAM workshop on the development and validation of expert systems for predicting toxicity, it was concluded by the participants that “the predictive capability of DEREK for irritancy and corrosivity is poor . . .” and that, since DEREK cannot take physicochemical properties into account, it is questionable whether this particular expert system is appropriate for such predictions (21). Skin irritation is also one of the endpoints covered by the Toxicity Prediction by Computer Assisted Technology (TOPKAT) and Hazarexpert systems (21). The validation of SARs, expert systems, and other “computational prediction techniques” has recently been discussed by Worth et al. (23).

With respect to both the limited applicability (use for restricted chemistries) and availability of SAR/SPR models for the prediction of skin corrosivity and irritation potentials, the task force considers that these are not a priority for any validation activities supported by ECVAM. Nevertheless, the valuable contributions to compound selection and the identification of potential toxic effects made by such systems when used as in-house screens is recognised. Since validated in vitro tests for skin corrosivity are now available (see later section on In Vitro Tests for Skin Corrosion), the need for SAR models for this particular endpoint is not as great as, for example, that for scientifically validated in vitro tests for skin irritation. However, it is recommended that efforts are made to encourage the wider use of the PC-based QSAR program for predicting the skin corrosivity potentials of organic acids, organic bases, phenols and electrophiles, by making it more readily available (for example, through LHASA UK).

**pH and Acid/Alkaline Reserve Measurements**

According to OECD Guideline 404 (14), substances should not be tested in animals for...
Figure 1: Proposed testing strategy for skin irritation/corrosion

1a) SAR/SPR \[\rightarrow\] skin corrosive \[\rightarrow\] STOP
   no or don’t know

1b) SAR/SPR \[\rightarrow\] skin irritant \[\rightarrow\] STOP
   no or don’t know

2) pH or acid/alkaline reserve \[\rightarrow\] > 11.5 or < 2 \[\rightarrow\] STOP
   < 11.5 or > 2

3) Other information indicating the material is a dermal corrosive \[\rightarrow\] yes \[\rightarrow\] STOP
   no

4) Is a valid in vitro test available to assess skin corrosion potential? \[\rightarrow\] no \[\rightarrow\] go to step 5
   yes

4a) In vitro test for skin corrosion \[\rightarrow\] skin corrosive \[\rightarrow\] STOP
   not a skin corrosive

5) Is a valid in vitro test for skin irritation available? \[\rightarrow\] no \[\rightarrow\] go to step 6
   yes

5a) In vitro skin irritation test \[\rightarrow\] skin irritant \[\rightarrow\] STOP
   not a skin irritant

6) Can we ethically approve human patch testing? \[\rightarrow\] no \[\rightarrow\] go to step 7
   yes

6a) Human patch test \[\rightarrow\] skin irritant \[\rightarrow\] STOP
   not a skin irritant

7) In vivo skin corrosion test using one rabbit \[\rightarrow\] skin corrosive \[\rightarrow\] STOP
   not a skin corrosive

8) Complete three rabbit skin irritation tests with two more rabbits \[\rightarrow\] skin irritant \[\rightarrow\] STOP
   not a skin irritant
skin irritation or corrosivity if they can be predicted to be corrosive on the basis of their physicochemical properties. In particular, substances exhibiting strong acidity or alkalinity should not be tested; these are predicted to be corrosive (C) from the following prediction model (PM), given in the OECD guideline: if pH ≤ 2 → C; if pH ≥ 11.5 → C.

In the recent ECVAM skin corrosivity validation study, the prediction of skin corrosivity from pH was evaluated (19). With the subset of 12 chemicals which had extreme pH values (that is, pH ≤ 2 or ≥ 11.5) and thus were predicted to be corrosive, there were three false positives when the predicted (from pH) and observed (rabbit in vivo) classifications were compared (i.e. the PM had a sensitivity of 75%). Since not all corrosive chemicals have a mechanism of action directly related to pH (the PM was only applicable to 12 of the 52 test chemicals for which pH data were obtained; 19), it is not appropriate to use pH data in isolation for classification purposes; other physicochemical properties should be taken into account when trying to predict corrosivity. For example, the need to consider acid/alkaline reserve values, which are a measure of the buffering capacities of chemicals, in addition to their pH values, is recognised in OECD Guideline 404 (14) and in the proposed OECD testing strategy for skin irritation/corrosion (15). This could be achieved by using a PM such as the one developed by Young and colleagues (24, 25).

A draft OECD guideline (physicochemical methods) on the measurement of acid/alkaline reserve has recently been prepared (G. Holland & A.P. Walker, personal communication). Subsequently, ECVAM has established a contract with BIBRA International (Surrey, UK) to evaluate the applicability of a test protocol based on this draft guideline for the prediction of skin corrosivity potential from acid/alkaline reserve measurements. Initially, the chemicals to be tested are those used in the ECVAM skin corrosivity validation study (7, 19).

The task force recommends that a critical appraisal of the usefulness of acid/alkaline...
reserve measurements for predicting skin corrosivity should be undertaken by the Management Team/Chemicals Selection Sub-Committee for the recent ECVAM skin corrosivity validation study once the data from BIBRA International are available. In particular, the following points need to be addressed: a) whether the current draft guideline is acceptable in relation to the methodology and the interpretation of the results, and whether a simpler, more cost-effective method could be developed; b) whether the measurement of acid/alkaline reserve enables test materials to be labelled as corrosive which would not normally be labelled as such, solely on the basis of their pH values; and c) for which chemical types the measurement of acid/alkaline reserve is appropriate. If this appraisal concludes that no additional information on corrosivity potential can be gained by determining the acid/alkaline reserves of test materials, this stage should be omitted from the testing strategy proposed at the OECD workshop (Figure 1).

It is suggested that the proposed OECD testing strategy is modified prior to its incorporation either directly in any updated version of Guideline 404 or in a guidance document to accompany the test guideline. It should be revised to indicate that historical data, SAR/SPR information, and pH or acid/alkaline reserve values and other relevant physicochemical data, where applicable, should be grouped together as initial considerations for the classification of chemicals as corrosive (step 1), prior to progressing to an in vitro test for skin corrosivity, if necessary (Figure 2).

In Vitro Tests for Skin Corrosion

A prevalidation study on in vitro skin corrosivity testing was conducted during 1993 and 1994 (6), as a first step toward defining those alternative tests which could be used within the context of the updated version of OECD Guideline 404 (14). The updated guideline includes the following statements: “... it may not be necessary to test in vivo materials for which corrosive properties are predicted on the basis of results from in vitro tests”; and “If an in vitro test is performed before the in vivo test, the description or reference of the test, including details of the procedure, must be given together with results obtained with the test and reference substances”. Similarly, the 18th amendment to Directive 67/548/EEC (26) states that: “... classification can be based on the results of validated in vitro tests”.

Three tests were included in the prevalidation study (6): a) the rat skin TER assay; b) CORROSITEX (In Vitro International, Irvine, CA, USA); and c) the Skin™ ZK1350 corrosivity test (Advanced Tissue Sciences, La Jolla, CA, USA). Fifty coded chemicals (25 corrosives [C] and 25 non-corrosives [NC]) were tested. In accordance with the conclusions and recommendations of the report on the prevalidation study (6), a formal international validation study on alternative methods for skin corrosivity testing was coordinated and funded by ECVAM during 1996 and 1997 (7, 19). The main objectives of the study were to: a) identify tests capable of discriminating C from NC for selected types of chemicals and/or all chemicals; and b) determine whether these tests could correctly identify known R35 (United Nations [UN] packing group I) and R34 (UN packing groups II & III) chemicals (7). The tests evaluated were: a) the rat skin TER assay; b) CORROSITEX; c) the Skin™ ZK1350 corrosivity test; and d) EPISKIN (EPISKIN, Chaponost, France; the production rights to this model now belong to L’Oréal). Each test was conducted in three independent laboratories. Sixty coded chemicals were tested (19), including organic acids (6C/5NC), organic bases (7C/3NC), neutral organics (9NC), phenols (2C/3NC), inorganic acids (6C/1NC), inorganic bases (2C/2NC), inorganic salts (1C/2NC), electrophiles (3C/5NC) and soaps/surfactants (3NC).

All of the tests evaluated showed acceptable intralaboratory and interlaboratory reproducibilities, and the TER, Skin™ and EPISKIN tests proved applicable to testing a diverse group of chemicals of different physical forms (7). Two of the four tests evaluated, the TER assay and EPISKIN, met the criteria agreed by the Management Team concerning acceptable underprediction and overprediction rates for them to be considered scientifically validated for use as replacements for the animal test for distinguishing between C and NC chemicals for all of the chemical types studied (objective a). EPISKIN was the only test able to distinguish between known R35 (UN packing
Figure 2: Modified strategy for skin irritation/corrosion testing of chemicals

Initial considerations, as appropriate:
- a) historical animal or human data
- b) SAR/SPR information
- c) pH or acid/alkaline reserve values
- d) other physicochemical data

**SAR** = structure-activity relationship; **SPR** = structure-property relationship; **C** = corrosive; **I** = irritant.

- Validated tests (rat skin TER and EPISKIN™ assays) are now available.
- No validated tests exist at the present time, so the options are to conduct a human 4-hour patch test (if appropriate), or to progress to an in vivo test with three rabbits.
- If appropriate, and subject to strict ethical and quality controls.

In vitro test for skin corrosion

Not a skin corrosive

- In vitro test for skin irritation
- Human 4-hour patch test

Not likely to be a skin irritant

Animal test for skin irritation (three rabbits)

Classify as I or not classified

Label as C

Label as I

Classify as I or not classified

SAR = structure-activity relationship; SPR = structure-property relationship; C = corrosive; I = irritant.

*Validated tests (rat skin TER and EPISKIN™ assays) are now available.*

*No validated tests exist at the present time, so the options are to conduct a human 4-hour patch test (if appropriate), or to progress to an in vivo test with three rabbits.*

*If appropriate, and subject to strict ethical and quality controls.*
group I) and R34 (UN packing groups II & III) chemicals, for all of the chemical types included, on an acceptable number of occasions (objective b). The corrosive potentials of about 40% of the test chemicals could not be assessed with CORROSITEX, and the assay did not meet all of the criteria for it to be considered acceptable as a replacement test. However, CORROSITEX may be valid for testing specific classes of chemicals, such as organic bases and inorganic acids (7). The Skin2 assay did not meet the criteria for it to be considered scientifically validated.

Due to the situation which arose concerning the commercial availability of the Skin2 and EPISKIN assays while the validation study was in progress (7), ECVAM is currently supporting the prevalidation of the EpiDerm human skin model (MatTek Corporation, Ashland, MA, USA) with respect to its subsequent validation for use for skin corrosivity testing. The initial results of this prevalidation are highly encouraging (M. Liebsch, personal communication).

Since scientifically validated in vitro tests for skin corrosivity are now available, the task force proposes that measures are taken to press for the updating of OECD Guideline 404 and, if necessary, Directive 67/548/EEC (to specify those Guideline 404 and, if necessary, taken to press for the updating of OECD the task force proposes that measures are tests for skin corrosivity are now available, Liebsch, personal communication). prevalidation are highly encouraging (M. corrosivity testing. The initial results of this prediction of skin corrosivity, as indicated require for a proper assessment of the applicabilities of human keratinocyte cultures for testing purposes. Furthermore, it appears that new endpoints, probably based on a mechanistic understanding of the events occurring at the cellular level which are critical for subsequent expression of der-
matotoxic responses \textit{in vivo}, and assays for their determination, need to be developed for use with both monolayer and more complex, multilayer, human skin models (see later section on Reconstituted Human Skin Models).

Irrespective of these considerations, monolayer keratinocyte cultures are not really suitable for use in \textit{in vitro} systems for the routine testing of chemicals for skin irritation potential. This is because of the importance of skin penetration as a determinant of any subsequent irritation response. Thus, the barrier function of the stratum corneum is essential in a model to predict skin irritation. Monolayer keratinocyte cultures are a very sensitive test system because of the lack of a stratum corneum, which means that their main applications should relate to: \textit{a}) research into mechanisms of skin irritation; and \textit{b}) demonstration of the lack of irritation potential of water-soluble, non-cytotoxic, chemicals. The task force proposes that monolayer keratinocyte cultures should not be a high priority for any prevalidation/validation studies supported by ECVAM.

\textit{Organ cultures}

The effects of chemical irritants in human and animal skin organ cultures have been investigated. For example, a two-compartment skin organ culture model is used at the TNO Nutrition and Food Research Institute (Zeist, The Netherlands). This model contains all of the dermal and epidermal cell types and structures involved in the irritation response, and the presence of air-exposed stratum corneum enables topical application of test materials (36, 37). The skin can be cultured for several days without loss of its differentiation characteristics, enabling studies to be undertaken of some aspects of tissue recovery following initial damage. Several endpoints of toxicity have been evaluated, including: \textit{a}) epidermal cell proliferation (38); \textit{b}) release of eicosanoids and cytokines (39); \textit{c}) histomorphology; and \textit{d}) cell viability, as determined by measuring MTT conversion (40).

The studies conducted with the two-compartment skin organ culture model to date indicate that rabbit skin cultures are generally more sensitive to chemical-induced toxicity than are human skin cultures (37). Good \textit{in vitro/in vivo} correlations have been observed when using diluted chemicals, but not when applying neat compounds to the \textit{in vitro} model. The toxic effects of various chemicals (irritants and non-irritants), at doses similar to those normally used \textit{in vivo}, have been studied in human skin organ cultures. Known irritants \textit{in vivo} caused histopathological changes and inhibited cell proliferation in the skin cultures; in contrast, these effects were not observed with non-irritants. Some, but not all, of the known irritants inhibited the reduction of MTT and caused the release of hydroxy fatty acids (37).

A similar procedure with skin explants taken from young rats has been performed at Shell (UK) and Rhône-Poulenc (France). The explants were maintained on cell culture medium and exposed to test chemicals for 18 hours. This method has also shown that irritant chemicals can cause a reduction in cell viability (as measured by MTT reduction) and an increase in the release of cytosolic and mitochondrial enzymes into the culture medium (D.J. Esdaile, personal communication).

An isolated perfused bovine udder skin model, developed initially for studying the percutaneous absorption of drugs, has recently been evaluated for its suitability for skin irritation testing (41). Results with only two test materials, both surfactants, have been reported to date; thus, it is too early to comment on the value of this particular \textit{ex vivo} system for predicting skin irritation potential.

In relation to the use of the human skin organ culture model, it is important that the ethical considerations of using human skin for testing purposes are addressed satisfactorily, and that suitable, viable human skin is available, as surgical waste, on a reasonably regular basis. Where human skin is not readily available, animal skin may be a realistic alternative. However, it is recognised that the barrier function of most animal skin is less than that of human skin. Hence, \textit{in vitro} animal skin models are likely to over-predict human skin irritation (as do the \textit{in vivo} animal models). The inclusion of animal skin explant models in a prevalidation study would provide useful information, both as a link between \textit{in vitro/in vivo} and human/animal comparisons, and as a potential alternative to human skin.

It is concluded that human and animal skin organ culture models should be consid-
Reconstituted human skin models

Since the 1980s, various human skin models have been developed, primarily for use clinically in the treatment of severe burns (42–44). Subsequently, some of these models have been used for testing the potential dermatotoxic effects of chemicals and products (29, 45–51). In contrast to simple monolayer keratinocyte cultures, the skin models are raised to the air–liquid interface, enabling the topical application of either neat or diluted test materials. The applications of \textit{in vitro} human skin models for pharmacotoxicological studies and dermal irritation testing have been reviewed recently by Roguet & Schaefer (45) and Lawrence (9), respectively.

Several human skin models are manufactured commercially. Currently, there are three reconstituted human skin equivalents which are being used for skin irritancy testing: a) EPISKIN; b) EpiDerm; and c) SKINETHIC\textsuperscript{TM} (Skinetic Laboratories, Nice, France). The production of Testskin\textsuperscript{TM} (Organogenesis, Canton, MA, USA), one of the first human skin equivalents available commercially, was discontinued in early 1993, while that of Skin\textsuperscript{2} ceased in October 1996. EPISKIN, EpiDerm and SKINETHIC are all reconstructed human epidermal models. EPISKIN is based on the skin model developed by Tinois \textit{et al.} (44). The dermal support comprises human collagens I and III covered with a fine layer of collagen IV. The EpiDerm model has been developed by Canon \textit{et al.} (51) and incorporates normal human keratinocytes cultured on permeable Millipore membranes. The SKINETHIC human skin equivalent is based on the model developed by Rosdy \& Claus (43). The keratinocyte support is either a cellulose acetate or a polycarbonate filter.

Test protocols have been developed for use with these human skin models which enable materials to be evaluated for their skin irritation potentials. The endpoint measurements used (9, 52) include histology, cell viability (typically MTT conversion), release of inflammatory mediators (particularly IL-1\textalpha), and effects on barrier function (for example, measurement of sodium fluorescein leakage). The models reportedly show close structural and biochemical resemblance to human skin \textit{in vivo} (9, 45, 48), but are typically at least 2–3-times more permeable than normal human skin (53).

A recent literature review on the use of commercial skin models specifically for skin irritation testing indicated that few studies have been published on this topic (29, 46, 50, 51, 53–55). A limited number of materials have been tested (typically surfactants and surfactant-containing products), mainly using histology, MTT conversion, and IL-1\textalpha release as the endpoints, either alone or in combination. The few \textit{in vitro/in vivo} comparisons which have been reported suggest that, for surfactants and cosmetic formulations, there is a very good agreement between the skin irritancy potentials observed \textit{in vivo} (either in humans or in rabbits) and \textit{in vitro} (29, 50, 51, 55; L.K. Earl \textit{et al.}, personal communication).

At the ECVAM workshop on the use of human keratinocytes and human skin models for predicting skin irritation, held in November 1997, several of the presentations related to on-going investigations with human skin models. Maja Ponec (University Hospital of Leiden, The Netherlands) showed that measurement of intraacellular IL-1\textalpha, rather than IL-1\textalpha release, was a more sensitive endpoint for determining the irritation potentials of sodium dodecyl sulphate (SDS) and oleic acid. Anne De Fraissinette (Novartis Pharma, Basel, Switzerland) summarised the outcome of \textit{in vitro/in vivo} comparisons with 75 dermal formulations, undertaken using the SKINETHIC model with MTT conversion as the endpoint. The \textit{in vivo} data had been generated with a 14-day application protocol. In \textit{vitro}, 72% of the formulations were classified as “innocuous” and 28% as “mild irritants”; \textit{in vivo}, 80% were classified as “innocuous” and 20% as “mild irritants”.

Before the decision was taken to cease production of Skin\textsuperscript{2}, prevalidation and validation studies on a protocol for the Skin\textsuperscript{2} model ZK1301 human skin equivalent are known to have been completed. The validation stage involved seven laboratories testing 21 materials (16 surfactants and five surfactant blends). The objective of the study was to determine whether the Skin\textsuperscript{2} test results were both reproducible and predictive of the
data generated concurrently for the same test materials in standard 14-day cumulative irritation studies in human volunteers. The *in vitro* endpoint used in the PM was IL-1α release. The outcome of the validation study, which is thought to be reasonably promising, has yet to be published.

A preliminary analysis has been undertaken on the data generated for the EPISKIN model evaluated in the ECVAM skin corrosivity validation study (7), with respect to the predictive ability of the *in vitro* viability results (MTT reduction assay; 4-hour exposure) for the rabbit primary irritation index (PII) values (56, 57) of the 33 NC chemicals included in the test set (these were initially selected to cover a range of skin irritation potentials based on their PII values (19)). With this diverse group of chemicals, a reasonable correlation was observed between the *in vitro* and *in vivo* values (statistically significant only at the 10% level), and it was concluded that, while the EPISKIN viability data (generated with the protocol for corrosivity testing) alone would not be sufficiently predictive of PII values, a useful PM could possibly be developed by combining them with information from additional relevant endpoints measured in the EPISKIN model (G. Archer, personal communication).

At the ECVAM workshop in November 1997, Catherine Cohen (L’Oréal, Aulnay-sous-Bois, France) summarised the outcome of a multicentre study on EPISKIN, in which the *in vitro* data (IL-1α release, MTT reduction) had been compared with human PII values for various cosmetic products (55). A concordance of 74% between the IL-1α release and human PII data was reported.

The EpiDerm model (Epi-200) has been evaluated with respect to its potential to identify acute skin irritants. The assay involves estimating the loss of viability (by measuring MTT conversion) of the human epidermal keratinocyte cultures over time when substances are applied topically, with the endpoint being the time taken to reduce cell viability by 50%. A range of organic compounds, including some simple binary mixtures of surfactants, have been tested by using a standard protocol. The *in vitro* results have been compared with acute dermal irritancy data obtained in volunteers with a human 4-hour patch test protocol (58). A test protocol and PM for the EpiDerm assay, which employ a concurrent reference standard to enable prediction of the acute dermal irritancy of test materials, have been proposed (L.K. Earl, personal communication). This reference standard is 20% SDS, which was chosen because it is also the reference standard for the interpretation of the human 4-hour patch test data and is classified as “irritant to skin” (R38) in the European Union (EU). The test protocol, PM and supporting data have been submitted to ECVAM for independent evaluation, with a view to including the EpiDerm assay in a prevalidation study.

The European Commission (DGXII) is currently supporting a 3-year project on various reconstructed skin models under the Standards, Measurements and Testing Programme. The collaborative study involves three industrial companies and one academic laboratory, and its objective is to develop standardised test protocols for the use of human skin models for the safety testing of cosmetics. Specifically, protocols for assessing percutaneous absorption, skin metabolism and skin irritation will be established, and the reproducibility and relevance of the data obtained will be determined.

It is clear from the literature that it is too early to draw any general conclusions about the predictive ability of test protocols employing human skin models for the skin irritation potentials of chemicals *in vivo*. The use of XTT rather than MTT for determining cell viability with human skin models should be considered. Further investigations directed toward evaluating the widespread applicability of human skin models for skin irritancy testing, including the definition of appropriate endpoints (possibly multiple endpoints encompassing different putative mechanisms of dermal irritation), test protocols and PMs are needed prior to any prevalidation studies being undertaken.

The task force considers that human skin models are relevant and appropriate *in vitro* systems for skin irritation testing, and that ECVAM should support the prevalidation/validation of such tests as a matter of high priority. It is recommended that initiatives are taken immediately by ECVAM to encourage the development of standard protocols and preliminary PMs for tests employing human skin models, so that these can be reviewed by the ECVAM Skin Irritation Task Force prior to the planning and design of a prevalidation study (see later section on *Preparation of Tests for Prevalidation*).
Human Patch Tests for Skin Irritation

Human volunteers have been used in the evaluation of skin irritation for many years, both for research and safety assessment purposes. However, only recently has an approach been developed which is specifically designed to address the issue of the classification and labelling (irritant/non-irritant) of pure chemicals and preparations. The human 4-hour patch test protocol developed by Unilever is designed to avoid the production of any response greater than a mild irritant reaction; it was conceived as part of a strategy to replace the use of animals for the identification of skin irritation/corrosion (11). Its development has been based on the principle that if corrosive (as demonstrated in an in vitro skin corrosivity test) and other toxicologically unacceptable hazards do not exist for the chemical or preparation, it is possible to conduct a human test, performed to the highest ethical standards, which is similar to the Draize rabbit skin irritation test but is designed to limit the intensity of any skin reactions (58).

The human 4-hour patch test protocol has been evaluated with about 65 materials for which EU classifications (R34, R38 or non-classified) were available (5, 58; D.A. Basketter, personal communication). The results indicate that the method provides an accurate “gold standard” assessment of acute irritation potential to human skin (58). The data generated are always compared with those for a standard positive control, 20% SDS, to enable their proper interpretation. The impacts of ethnic, inter-individual and seasonal variations have been studied (reviewed in reference 58), and the reproducibility of the test protocol has been demonstrated in an interlaboratory evaluation (59).

The validation status of the human 4-hour patch test protocol has been commented upon by Basketter et al. (58), who argue that, since the method is undisputably relevant (with respect to the species, organ and endpoint) and has been shown to be reliable (that is, it is reproducible within and between laboratories), it is a “valid alternative to the Draize rabbit skin irritation test for the classification of skin irritation hazard.” Some advantages of using such an approach for determining the skin irritation potentials of certain chemicals are that: a) the use of animals can be avoided; b) the resulting classification (in EU terms) of skin irritation hazard is accurate; and c) “gold standard” data are generated for use in the validation of in vitro tests for predicting skin irritation (which are needed because of the cost and time implications of conducting human studies, and because there are many chemicals which cannot be tested for irritation potential in humans due to other associated toxicological effects, etc.).

The data on the reproducibility/interlaboratory transferability of the human 4-hour patch test protocol (59) are promising. However, as can only be expected with a test involving human volunteers, the number of test chemicals (ten, including two different concentrations of SDS) was relatively small, and for five of these the test was only conducted in two laboratories. In terms of the further standardisation and improved transferability of the protocol, consideration could be given to the use of objective (for example, transepidermal water loss) rather than subjective (visual scoring) measurements, although it is recognised that, with extensive training, visual scoring can be sensitive, reliable and reproducible within a testing laboratory (60).

The human 4-hour patch test is currently being considered by the OECD Member Countries as a potential new test guideline (Acute Dermal Irritation Study in Human Volunteers [61]). Despite the concerns raised by some countries in relation to human testing, it is hoped that a revised guideline can be finalised and accepted during 1998. The task force is supportive of the use, subject to strict ethical and quality controls, of the human 4-hour patch test protocol for the identification of potential skin irritants. The current protocol would appear to be applicable to the acute dermal irritation testing of chemicals, pesticide formulations and pharmaceuticals, but the exposure period is too short for it to be used for testing cosmetic formulations. It is proposed that the testing strategy for skin irritation/corrosion developed at the OECD workshop (Figure 1) is revised to enable the human 4-hour patch test to be conducted at an earlier stage in the hierarchical scheme, if this is considered to be appropriate for the particular chemical or formulation being tested. A modification to the sug-
gested OECD testing strategy for skin irritation/corrosion has been proposed by the task force (Figure 2).

Preparation of Tests for Prevalidation

A standard protocol and preliminary PM for a test for skin irritation employing EpiDerm were made available to the task force, along with supporting data on a mixed set of 30 test chemicals for which 4-hour human patch test results (obtained according to a standardised protocol) were also available. Thus, the task force considers it appropriate to propose to ECVAM that this protocol with EpiDerm is put forward for prevalidation during 1998.

In addition, since it was felt preferable to be able to include other in vitro tests in such a prevalidation study, the task force recommends that a “challenge” is set to anyone interested in taking part. This should involve submitting data, obtained according to a specified test protocol with a preliminary PM, on 20% SDS (as a reference standard) and ten test chemicals selected by the task force (a subset of the 30 chemicals already tested with EpiDerm; Table I), for review by the task force by 31 May 1998. These “test development” studies are to be conducted without financial support from ECVAM, although it is envisaged that ECVAM funding will be made available for any subsequent prevalidation and validation studies.

From 21 January 1998 onwards, information about this “challenge” was circulated by ECVAM, and a list of the test chemicals was distributed on 5 February 1998. Those laboratories wishing to take part have been asked to provide the following information when submitting data for review by the task force.

1. A description of the basis of the method.
2. A definition of the scientific purpose and the proposed practical application of the method.
3. The case for the relevance of the method.
4. A test protocol/Standard Operating Procedures, including specification of: a) endpoint; b) derivation and expression of results; and c) their interpretation and application, via a PM. Appropriate controls, and historical data for these, should be included.

Table I: Test chemicals for use in standardising test protocols and developing prediction models

<table>
<thead>
<tr>
<th>Chemical</th>
<th>Supplier</th>
<th>Rabbit data</th>
<th>HPT data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decanol</td>
<td>Aldrich</td>
<td>56</td>
<td>58, 63</td>
</tr>
<tr>
<td>Isopropanol</td>
<td>Aldrich</td>
<td>56</td>
<td>58</td>
</tr>
<tr>
<td>Octanoic acid</td>
<td>Aldrich</td>
<td></td>
<td>58</td>
</tr>
<tr>
<td>Decanoic acid</td>
<td>Aldrich</td>
<td></td>
<td>58, 63</td>
</tr>
<tr>
<td>Dodecanoic (lauric) acid</td>
<td>Aldrich</td>
<td>56</td>
<td>63</td>
</tr>
<tr>
<td>Methyl caproate</td>
<td>Aldrich</td>
<td>56</td>
<td>58</td>
</tr>
<tr>
<td>Methyl laurate</td>
<td>Aldrich</td>
<td>56</td>
<td>63</td>
</tr>
<tr>
<td>Methyl palmitate</td>
<td>Aldrich</td>
<td>56</td>
<td>58</td>
</tr>
<tr>
<td>Isopropy palmitate</td>
<td>Aldrich</td>
<td>56</td>
<td>58</td>
</tr>
<tr>
<td>20% Dimethyldodecylaminobetaine</td>
<td>Albright &amp; Wilson</td>
<td>56</td>
<td>58</td>
</tr>
</tbody>
</table>

*aHPT = human patch test.*
5. Limitations of the method.
6. Data supporting the intralaboratory reproducibility and, if available, the interlaboratory transferability of the method (62).

Conclusions and Proposals to ECVAM

1. SAR/SPR models for skin corrosivity or irritation are not considered to be a priority for any validation activities supported by ECVAM, due to the availability of validated in vitro tests for skin corrosivity and the recognised need to evaluate and validate in vitro tests for skin irritation. Nevertheless, it is recommended that efforts are made to encourage the wider use of the PC-based QSAR program for predicting the skin corrosivity potential of organic acids, organic bases, phenols and electrophiles, by making it more readily available to potential users (for example, through LHASA UK).

2. A critical appraisal of the usefulness of acid/alkaline reserve measurements for predicting skin corrosivity should be undertaken, by the Management Team/Chemicals Selection Sub-Committee for the recent ECVAM skin corrosivity validation study, once the values for the 60 test chemicals are available. If it is not clear that there is any extra valuable information to be readily gained by determining the acid/alkaline reserves of test materials, this stage should be omitted from the testing strategy proposed at the OECD workshop.

3. Since scientifically validated in vitro tests for skin corrosivity are now available, it is proposed that measures are taken to press for the updating of OECD Guideline 404 and, if necessary, Directive 67/548/EEC (to specify those in vitro tests which could be used; i.e. the rat skin TER assay, and EPISKIN or other human skin models which demonstrate a performance equivalent to that achieved by EPISKIN and the TER assay in the validation study).

4. The combined use of SAR/SPR information, pH data, acid/alkaline reserve values (if appropriate) and in vitro data for the prediction of skin corrosivity, as indicated in the proposed testing strategy emanating from the OECD workshop, should be evaluated by ECVAM. It is suggested that the OECD testing strategy is modified prior to its incorporation either directly in an updated version of Guideline 404 or in a guidance document to accompany the test guideline, to indicate that historical data, SAR/SPR information, and pH or acid/alkaline reserve values and other physicochemical data, where applicable, should be grouped together as initial considerations for the classification of chemicals as corrosives or irritants (Figure 2), prior to progressing to an in vitro test for skin corrosivity.

5. Other alternative tests for skin corrosivity, such as the human TER assay and new tests with human skin equivalents, should also be considered for prevalidation/validation by ECVAM if defined test protocols, PMs and supporting data are made available. In this respect, structural and performance criteria should be defined by the Management Team for the ECVAM skin corrosivity validation study, based on the results obtained with the EPISKIN and TER assays in the validation study, which other skin corrosivity tests must meet to be considered scientifically validated.

6. There is no current consensus on the general predictive ability of monolayer keratinocyte cultures for skin irritation in vivo, or on which in vitro endpoints are the most appropriate to use. Monolayer keratinocyte cultures are not really suitable for use in in vitro systems for the routine testing of chemicals for skin irritation potential due to the lack of a stratum corneum (skin penetration is a key determinant of any subsequent irritation response in vivo), and because of solubility considerations. The main applications of monolayer keratinocyte cultures should relate to research into mechanisms of skin irritation and demonstration of the lack of irritation potential of water-soluble, non-cytotoxic, chemicals. It is proposed that monolayer keratinocyte cultures should not be a high priority for any prevalidation/validation studies supported by ECVAM.
The human and animal skin organ culture methods should be considered as candidates for ECVAM-supported prevalidation activities if test protocols, preliminary PMs and supporting data are made available for review by the ECVAM Skin Irritation Task Force.

Reconstituted human skin models are relevant and appropriate in vitro systems for skin irritation testing. ECVAM should support the prevalidation/validation of such tests as a matter of high priority. Initiatives should be taken immediately by ECVAM to encourage the development of standard protocols and preliminary PMs for tests employing human skin models, so that these can be reviewed by the ECVAM Skin Irritation Task Force prior to the planning and design of a prevalidation study.

The task force is supportive of the use, subject to strict ethical and quality controls, of the human 4-hour patch test protocol for the identification of potential skin irritants. It is proposed that the testing strategy for skin irritation/corrosion developed at the OECD workshop is revised to enable the human 4-hour patch test to be conducted at an earlier stage in the hierarchical scheme (Figure 2), if this is considered to be appropriate for the particular chemical or formulation being tested.

In recognition of the pressure on ECVAM to make rapid progress in relation to the validation of in vitro tests for skin irritation testing, the ECVAM Skin Irritation Task Force proposes, at least as an interim measure, to act as a preliminary Management Team for ECVAM’s activities in this area and, in particular, to oversee the studies needed to prepare tests for prevalidation.

Acknowledgements

We would like to thank the following people for providing information for this report: Graeme Archer (ECVAM, Italy); David Basketter (SEAC Toxicology Unit, Unilever, UK); and David Esdaile (Rhône-Poulenc Agro, France). In addition, the contributions of the participants at the ECVAM workshop on the use of human keratinocytes and human skin models for predicting skin irritation (Utrecht, The Netherlands, 9–11 November 1997) are acknowledged.

References


Alternative methods for skin irritation testing


