Subgroup 2. Skin Irritation/Corrosion

Valérie Zuang[1], Marie-Ange Alonso[2], Philip A. Botham[3], Chantra Eskes[1], Julia Fentem[4], Manfred Liebsch[5], Johannes J.M. van de Sandt[6],

[1]ECVAM, Institute for Health & Consumer Protection, European Commission Joint Research Centre, 21020 Ispra (VA), Italy; [2]EVIC FRANCE,33290 Blanquefort, France; [3]Syngenta, Central Toxicology Laboratory, Alderley Park, Macclesfield, Cheshire SK10 4TJ, UK; [4]SEAC, Unilever Colworth Laboratory, Sharnbrook Bedfordshire MK44 1LQ,,UK; [5]ZEBET BfR, Diedersdorfer Weg 1, 12277 Berlin, Germany; [6]TNO Nutrition & Food Research, Department of Explanatory Toxicology, Utrechtseweg 48, 3700 AJ Zeist, The Netherlands.

Skin irritation

1. Inventory of methods currently available

Since the skin is often exposed, either intentionally or unintentionally, to cosmetic products, it is clear that the potential for a particular product/ingredient to cause skin irritation or for a particular ingredient to cause skin corrosion needs to be carefully evaluated as part of the overall safety assessment process.

Dermal irritation is defined as the production of "reversible damage of the skin following the application of a test substance for up to 4 hours" (OECD TG 404, 2002). It is generally assessed by the potential of a certain substance to cause erythema/eschar and/or oedema after a single topical application on rabbit skin and based on the Draize score (OECD TG 404, 2002).
Recently some general refinement provisions were included in the OECD testing guideline 404. This updated version recommends, prior to undertake the described *in vivo* animal test for corrosion/irritation of the substance, to perform a sequential testing strategy. This strategy is based in a stepwise order on: a weight of evidence analysis, pH considerations, use of validated and accepted *in vitro* tests, and finally refinement of the animal testing. Some examples of decision-making in the testing strategy are the following:

1) A substance with a pH below 2.0 or above 11.5 should not be tested, due to its suspected corrosivity.

2) A substance found to be corrosive in one of the alternative corrosivity tests (taken up in Annex V of Dir. 67/548/EEC), should not be tested in the Draize test.

Although the OECD TG 404 was recently refined, there are to date no validated alternative methods replacing the classical Draize test for predicting the acute skin irritation of cosmetic ingredients or chemicals (Botham *et al.*, 1998).

2. Inventory of alternative methods currently available

Whereas relatively few QSAR studies for skin irritation have been reported (see chapter on SAR below), several prevalidation efforts were carried out for *in vitro* alternative methods, some of which are currently pursued in the on-going ECVAM validation study (Zuang *et al.*, 2002). *In vitro* alternatives in the field vary from more simple models such as keratinocyte cultures to more complex organotypic cultures and reconstituted human skin models (Botham *et al.*, 1998; van de Sandt *et al.*, 1999). During 1999-2001, some of the most promising *in vitro* methods were evaluated in prevalidation studies (see table 1; Fentem *et al.*, 2001). The five tests included were: EpiDerm™, EPISKIN™, Prediskin™, the non-perfused pig ear model, and the *in vitro* mouse skin integrity function test (SIFT). The outcome of these studies was that none of the tests was ready for progression to formal validation. Various follow-up activities took place where appropriate modifications were made to enable some test protocols to meet the criteria for inclusion in a formal validation study (Zuang *et al.*, 2002). On the basis of additional work, the EPISKIN™, EpiDerm™ and SIFT test protocols and/or prediction models will be evaluated in a formal validation study funded by ECVAM during 2004 (Fentem and Botham, 2002). The main overall objective is to identify those *in vitro* alternatives capable of discriminating skin irritants from non-irritants.

Several testing strategies for evaluating the skin irritation potential of ingredients and products have been described (Botham *et al.*, 1998; OECD, 2002; Robinson *et al.*, 2002), some of which also involve human volunteer studies (human 4-hour patch test; Basketter *et al.*, 1997). In such strategies human testing is generally used to determine the "relative skin compatibility" and "exposure" parts of the risk assessment of products/formulations (and occasionally ingredients), to confirm that there are no harmful effects when applying a cosmetic product for the first time to human skin or mucous membranes. There are however some specific occasions when cosmetic ingredients can be ethically tested using human volunteers to provide additional data for risk assessment purposes, but this is rather a case-by-case situation where the hazard has been characterised (sensitisation, corrosion, mutagenicity, etc.), than a generically applicable testing approach.

In addition for the hazard assessment of single entities, many cosmetics companies conduct *in vitro* and/or dermatological (human skin compatibility, often following repeated application; COLIPA, 1995; COLIPA, 1997) comparative studies on new formulations. In such studies the results for the new formulation are interpreted relative to those obtained for other similar formulations (typically those with a long history of safe consumer use and satisfaction).

Table 1. Summary of the *In Vitro* Methods which participated in the ECVAM pre-validation study for Skin Irritation

Method	Test System	Endpoint	Applicability	Validation Authority	Status	References
EPISKIN™ human skin model (commercial system)	reconstructed human epidermal equivalent	cell viability (MTT reduction assay)	general; a few materials may interfere with MTT reduction	ECVAM	protocol modification following prevalidation study; revised protocol and prediction model proposed for validation	Fentem *et al.* 2001 Zuang *et al.* 2002 Portes *et al.* 2002
EpiDerm™ human skin model (commercial system)	reconstructed human epidermal equivalent	cell viability (MTT reduction assay)	general; a few materials may interfere with MTT reduction	ECVAM	protocol modification following prevalidation	Fentem *et al.* 2001 Zuang *et al.* 2002
PREDISKIN™ human skin model (commercial system)	reconstructed human epidermal equivalent	histology and cell viability (MTT reduction assay)	general; a few materials may interfere with MTT reduction	ECVAM	did not meet criteria for progression to phase 3 of prevalidation	Fentem *et al.* 2001
Pig ear test	pig ear	trans-epidermal water loss (TEWL)	general	ECVAM	further development; did not meet criteria for progression to phase 3 of prevalidation	Fentem *et al.* 2001 Zuang *et al.* 2002
Mouse skin integrity function test (SIFT)	excised mouse skin	TEWL and electrical resistance (ER)	general; a few materials may interfere with either TEWL or ER determination	ECVAM	prediction model modification following prevalidation study; revised protocol and prediction model proposed for validation	Fentem *et al.* 2001 Zuang *et al.* 2002 Heylings *et al.* 2003

RECONSTRUCTED HUMAN SKIN MODELS

Reconstructed human skin models are able to mimic the human skin to a large extend. They are three-dimensional models generated by growing keratinocyte cultures at the air-liquid interface on various substrates, and enable the topical application of either neat or diluted test materials (Botham *et al.*, 1998; van de Sandt *et al.*, 1999; Faller *et al.*, 2002). Several human skin models are manufactured commercially, such as the EPISKIN ™ model (EPISKIN-SNC, Gerland, France), the EpiDerm™ model (MatTek, Ashland, MA, USA) and the SkinEthic™ model (SkinEthic, Nice, France). Moreover, some in-house models were also developed and evaluated for their potential in detecting skin irritation such as the one developed by Cosmital, Wella (Faller *et al.*, 2002), or the one developed by Ponec and co-workers (Ponec and Kempenaar, 1995).

1. The EPISKIN™ model

Short description , scientific relevance and purpose

EPISKIN™ (EPISKIN-SNC, Gerland, France) is a three-dimensional human skin model consisted of a type 1 bovine collagen matrix, representing the dermis, surfaced with a type IV human collagen, upon which is laid after 13 days in culture stratified differentiated epidermis derived from second passage human keratinocytes (Tinois *et al.*, 1991). Its use for skin irritation testing involves topical application of test materials to the surface of the skin, and the subsequent assessment of their effects on cell viability by using the MTT assay. The endpoint used to distinguish between potential skin irritants and non-irritants is the percentage (%) of cell viability (Roguet *et al.*, 1994, 1998). However, the use of other, more mechanistic, endpoints such as Interleukin 1? (IL-1?) or lactate dehydrogenase (LDH) has also been evaluated (Roguet *et al.*, 1998; Faller *et al.*, 2002). Following results from a prevalidation study, the protocol was refined in order to improve the specificity of the method (Fentem et al., 2001; Zuang *et al.*, 2002). This refinement consisted in reducing the exposure time of epidermis with chemicals. Sensitivity, specificity and accuracy of the new method were improved and the EPISKIN™ model is now ready to enter a validation study of *in vitro* tests for acute skin irritation (Portes *et al.*, 2002).

References

Faller C., Bracher M., Dami N., Roguet R. (2002). Predictive ability of reconstructed human epidermis equivalents for the assessment of skin irritation of cosmetics. *Toxicology in Vitro* 16, 557-572.

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

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.

Roguet R., Cohen C., Dossou K.G., Rougier A. (1994). EPISKIN, a reconstituted epidermis for assessing *in vitro* the irritancy of topically applied compounds. *Toxicology in Vitro* 8, 49-59.

Roguet R., Cohen C., Robles C., Courtellemont P., Tolle M., Guillot J.P., Pouradier Duteil X. (1998). An interlaboratory study of the reproducibility and relevance of EPISKIN™, a reconstructed human epidermis, in the assessment of cosmetics irritancy. *Toxicology in Vitro* 12, 295-304.

Tinois E., Tillier J., Gaucherand M., Dumas H., Tardy M., Thivolet J. (1991). *In vitro* and post-transplantation differentiation of human keratinocytes grown on the human type IV collagen film of a bilayered dermal substitute. *Experimental Cell Research* 193, 310-319.

Zuang V., Balls M., Botham P.A., Coquette A., Corsini E., Curren R.D., Elliott G.R., Fentem J.H., Heylings J.R., Liebsch M., Medina J., Roguet R., van de Sandt J.J.M., Wiemann C., Worth A.P. (2002). Follow-up to the ECVAM prevalidation study on *in vitro* tests for acute skin irritation. *ATLA* 30, 109-129.

Producer

EPISKIN-SNC, Gerland, France.

Known users

Industry and contract testing laboratories use this method.

Status of validation and/or standardisation

Since the method has already undergone prevalidation and will undergo validation by ECVAM, a defined/optimised protocol is available, as well as data on intra-and inter-laboratory variation, protocol transferability and in vitro/in vivo comparisons.

Which efforts are needed to complete validation of the method?

The method is currently part of the ECVAM validation study of *in vitro* methods for acute skin irritation, expected to be finished in 2005. The possible use of more mechanistically based endpoints, such as adenylate kinase release and IL-1? will also be evaluated.

2. The EpiDerm™ model

Short description , scientific relevance and purpose

In the EpiDerm™ reconstituted human skin model (MatTek, Ashland, MA, USA), human skin-derived keratinocytes are grown on specially prepared Millicell cell culture inserts, forming a multi-layered, differentiated, model of the human epidermis *in vitro* (Cannon *et al.*, 1994; Earl *et al.*, 1999). The original EpiDerm™ test protocol was based on the comparison of the time of exposure required to decrease tissue viability by 50 % (ET50) of the test chemical with the ET50 of a reference standard (Fentem *et al.*, 2001). With this protocol the intralaboratory and interlaboratory reproducibilities were found to be acceptable. On the other hand the test method gave a too high percentage of false negatives during phase III of the prevalidation study. As a follow-up activity to this prevalidation study, the test protocol and prediction model were modified, which improved the predictive ability of the model (Zuang *et al.*, 2002).

References

Cannon CL, Neal PJ, Southee JA, Kubilus J, Klausner M. (1994). New epidermal model for dermal irritancy testing. *Toxicology in Vitro* 8, 889-891.

Earl L.K., Hall-Manning T.J., Holland G.H. (1999). The determination of acute chemically induced skin irritation using a human skin model. In: Clark D.G., Lisansky S.G., Macmillan R. (Eds.), Alternatives to Animal Testing II. Proceedings of the Second International Scientific Conference Organised by the European Cosmetic Industry. CPL Press, Newbury, p. 227.

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

Zuang V., Balls M., Botham P.A., Coquette A., Corsini E., Curren R.D., Elliott G.R., Fentem J.H., Heylings J.R., Liebsch M., Medina J., Roguet R., van de Sandt J.J.M., Wiemann C., Worth A.P. (2002). Follow-up to the ECVAM prevalidation study on *in vitro* tests for acute skin irritation. *ATLA* 30, 109-129.

Producer

MatTek, Ashland, MA, USA.

Known users

Industry and contract testing laboratories use this method.

Status of validation and/or standardisation

As for Episkin, the method had undergone prevalidation by ECVAM and the protocol has been optimised to be included in the ECVAM validation study on acute skin irritation.
Thus, data on intra-and inter-laboratory variation, protocol transferability and in vitro/in vivo comparisons are available.

Which efforts are needed to complete validation of the method?

In order to have one common protocol applicable to different skin models, it was suggested to apply the revised EPISKIN™ protocol (measure of the % of cell viability by using the MTT assay after a reduced time of exposure) to the Epiderm™ model. This protocol showed promising results and was chosen to be used for the ECVAM validation study of *in vitro* methods for acute skin irritation (Zuang *et al.*, 2002).

3. The SkinEthic™ model

Short description , scientific relevance and purpose

The Skinethic™ cultures are obtained by culturing normal human keratinocytes on a polycarbonate culture inserts lifted to the air-liquid interface, and a stratum corneum forms from growth at the air-liquid interface. The method is based on the model developed by Rosdy and Clauss (1990). Novartis Pharma has proposed a two –step acute and cumulative irritation screening system based on this approach (de Brugerolle de Fraissinette *et al.*, 1999). The most used endpoints with this model are the modulation of cell viability, release of IL-1? , and morphological changes.

References

de Brugerolle de Fraissinette A., Picarles V., Chibout S., Kolopp M., Medina J., Burtin P., Ebelin M.E., Osborne S., Mayer F.K., Spake A., Rosdy M., De Wever B., Ettlin R.A., Cordier A. (1999). Predictivity of an *in vitro* model for acute and chronic skin irritation (SkinEthic) applied to the testing of topical vehicles. *Cell Biology and Toxicology* **15**, 121-135.

Rosdy M., Clauss LC (1990). Complete human epidermal cell differentiation in chemical defined medium at the air-liquid interface on inert filter substrates. *Journal of Investigative Dermatology* **95**, 409-414.

Producer
SkinEthic, Nice, France.

Known users
Industry and contract testing laboratories use this method.

Status of validation and/or standardisation
- Is the protocol defined/optimised? **YES**
- Is there data on intra-laboratory variation? **YES**
- Is there data on the transferability? **Limited**
- Is there data on the inter-laboratory variability? **Limited**
- *In vivo* /*in vitro* comparisons? **Limited**
- Is the method validated? **NO**

Which efforts are needed to complete validation of the method?
There is a need to determine the transferability and inter-laboratory variability of the method before including it in a validation study.

4. Other models

Other models exist where the keratinocyte cultures grown at air-liquid interface are cultured in different substrates. Amongst the well-characterised models are:
- the Cosmital in-house model (Wella, Switzerland),
- the Apligraf® model (Organogenesis Inc., Novartis Pharmaceuticals Corporation, Canton, MA USA),
- the reconstituted epidermis on de-epidermized dermis (RE-DED) used by Ponec and co-workers (Ponec and Kempenaar, 1995; Robinson *et al.*, 2000; Boelsma *et al.*, 1996, 2000).

The predictive ability of the cosmital in-house model and the comparison to *in vivo* data have been assessed (Faller *et al.*, 2002). The use of Apligraf® in a tiered strategy for the evaluation of cumulative skin irritation potential of chemicals has also been assessed (Medina *et al.*, 2000). However, none of these models are at the same stage as the previous examples. They would be able to undergo catch-up validation and/or be used within the context of any future test guideline as long as they meet defined performance criteria (as with the skin corrosion TG).

References
Boelsma E., Tanojo H., Boddé H.E., Ponec M. (1996). Assessment of the potential irritancy of oleic acid on human skin: evaluation *in vitro* and *in vivo*. *Toxicology in Vitro* **10**, 729-742.

Boelsma E., Gibbs S., Faller C. Ponec M. (2000). Characterization and comparison of reconstructed skin models: morphological and immunohistochemical evaluation. *Acta Derm Venereol* **80**, 82-88.

Faller C., Bracher M., Dami N., Roguet R. (2002). Predictive ability of reconstructed human epidermis equivalents for the assessment of skin irritation of cosmetics. *Toxicology in Vitro* **16**, 557-572.

Medina J., de Brugerolle de Fraissinette A., Chibout S.D., Kolopp M., Kammermann R., Burtin P., Ebelin M.E., Cordier A. (2000). Use of human skin equivalent Apligraf for *in vitro* assessment of cumulative skin irritation potential of topical products. *Toxicology and Applied Pharmacology* **164**, 38-45.

Ponec M. and Kempenaar J. (1995). Use of human skin recombinants as an *in vitro* model for test the irritation potential of cutaneous irritants. *Skin Pharmacol* **8**, 49-59.

Robinson M.K., Cohen C., de Brugerolle de Fraissinette A., Ponec M., Whittle E., Fentem J.H. (2000). Non-animal testing strategies for assessment of the skin corrosion and skin irritation potential of ingredients and finished products. *Food and Chemical Toxicology* **40**, 573-592.

SKIN EXPLANTS AND ORGAN CULTURES

1. The mouse skin integrity function test (SIFT)

Short description , scientific relevance and purpose

The SIFT (protocol of Syngenta CTL, Macclesfield, UK; Heylings *et al.*, 2001) is based on assessment of mouse skin integrity following exposure to test material.Two methods are used to assess stratum corneum integrity: trans-epidermal water loss (TEWL), and electrical resistance (ER). The SIFT protocol was developed as a prescreen for assessing the skin irritation potential of industrial chemicals. It combines elements of *in vitro* percutaneous absorption, where skin integrity assessment is fundamental to the method, together with knowledge from models used for the assessment of skin emolliency using intact skin. Mouse skin provides such a model since it is sufficiently thin to use in static glass diffusion cells as intact whole skin, and it displays relevant biochemical activity and highly reproducible and robust barrier properties (Fentem *et al.*, 2001). The basis of the SIFT prediction model is if the ratios of the pre- and post-application values for either TEWL or ER are greater or smaller than five-fold, then the test chemical is deemed irritant or non irritant respectively (Heylings *et al.*, 2003). The prediction model that failed in the prevalidation exercise was modified by using new statistical means of evaluating the data that were more suited for the model. Such modification improved the specificity and sensitivity of the model (Heylings *et al.*, 2003; Zuang *et al.*, 2002).

References

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

Heylings, J. R. Clowes, H.M. Hughes L. (2001). Comparison of tissue sources for the skin integrity function test (SIFT). *Toxicology in Vitro* 13, 597-600.

Heylings J.R., Diot S., Esdaile D.J., Fasano W.J., Manning L.A., Owen H.M. (2003). A prevalidation study on the *in vitro* skin irritation function test (SIFT) for prediction of acute skin irritation *in vivo*: results and evaluation of ECVAM Phase III. *Toxicology in Vitro* 17, 123-138.

Zuang V., Balls M., Botham P.A., Coquette A., Corsini E., Curren R.D., Elliott G.R., Fentem J.H., Heylings J.R., Liebsch M., Medina J., Roguet R., van de Sandt J.J.M., Wiemann C., Worth A.P. (2002). Follow-up to the ECVAM prevalidation study on *in vitro* tests for acute skin irritation. *ATLA* 30, 109-129.

Developer of the method

Syngenta Central Toxicology Laboratory, Macclesfield, UK

Known users - **YES**

Status of validation and/or standardisation

- Is the protocol defined/optimised? **YES**
- Is there data on intra-laboratory variation? **YES**
- Is there data on the transferability? **YES**
- Is there data on the inter-laboratory variability? **YES**
- *In vivo* /*in vitro* comparisons? **YES**
- Is the method validated? **NO**

Which efforts are needed to complete validation of the method?

In order to improve the low inter-laboratory reproducibility observed in phase III of the prevalidation study with the revised protocol, the future validation protocol for SIFT will reinforce that the empirical values for TEWL and ER in positive and negative controls must be within certain ranges (Heylings *et al.*, 2003).

2. The Prediskin™ model

Prediskin™ (Biopredic, Rennes, France) improved test protocol involves exposure of human skin cultures (obtained from patients undergoing plastic surgery) to test materials for 20 hours, and subsequent assessment of effects on the percentage of cell viability by using the MTT assay (Fentem *et al.*, 2001). In the initial protocol, histology was included if cell viability was greater than 35%. The Prediskin did not perform sufficiently well in phase II for it to progress to phase III; the protocol was overly sensitive, resulting in the prediction of all the non-irritants as irritant (Fentem *et al.*, 2001). In additional studies (a repeat of phase I), the Prediskin protocol was modified by using non-stripped human skin and by increasing the threshold for discriminating between irritants and non-irritants (MTT assay). The change in the prediction model considerably improved the ability of the test to distinguish irritant from non-irritant chemicals (Fentem *et al.*, 2001). However, further work on Prediskin™ has not been followed up for commercial reasons.

3. The pig ear test

The non-perfused pig ear test (protocol of TNO-PML, Rijkswijk, NL) is based on determination of the absolute increase in trans-epidermal water loss (TEWL) from the skin surface, following exposure of the pig ear to test material, as the endpoint to distinguish between irritants and non-irritants. The pig ear test did not perform sufficiently well in phases I and II for it to progress to phase III; the variability in the results obtained was too great, indicating that the test would be of limited predictive value (Fentem *et al.*, 2001). In additional studies (a repeat of phase I), attempts to improve the intralaboratory reproducibility of the pig ear test were unsuccessful.

Structure-activity relationships for skin irritation

Relatively few QSAR studies for skin irritation have been reported in the literature.
The most recent reviews were made by Cronin et al. (in press), Hulzebos et al., (2003) and Patlewicz et al. [in press (b)].

Barratt (1996) reported a QSAR for predicting the primary irritation index (PII) of organic chemicals, but this had little predictive value (r^2 =0.42). In the same study, discriminant analysis was shown to discriminate between irritant and non-irritant chemicals, as defined by EU classification criteria, with an accuracy of 67%.

Hayashi *et al.* (1999) reported two QSARs for predicting the molar-weighted PII of phenols. One model, based on absolute hardness, was proposed for chemicals with negative lowest unoccupied molecular orbital (LUMO) energies, whereas the other model, based on the logarithm of the octanol-water partition coefficient (logP), was proposed for chemicals with positive LUMO energies. These models had correlation coefficients of 0.72 and 0.82, respectively (i.e. r^2 values of 0.52 and 0.67).

Smith *et al*. (2000) analysed a data set of 42 esters, for which human skin irritation data were available, and for which 19 physicochemical properties had been calculated. Best subsets regression was used to select variables for subsequent inclusion in discriminant models. The best variables were water solubility (lower for irritants than non-irritants), a dispersion parameter (higher for irritants), a hydrogen-bonding parameter (higher for irritants), the sum of partial positive charges (lower for irritants), and density (lower for irritants). A discriminant model based on all five parameters had a sensitivity of 85% and a specificity of 92%.

Expert systems such as DEREK for Windows, HazardExpert and TOPKAT have been reviewed in Cronin et al. (2003).

The BgVV database has been used to develop specific SAR models for predicting skin irritation/corrosion. These models have been incorporated into a DSS (Gerner et al., 2000a, 2000b; Zinke er al., 2000). The DSS is mainly a rule-based approach, the rules being developed based not only on substructural molecular features but also on physicochemical properties such as molecular weight, acqueous solubility, and log K_{ow}. The rules have been developed and validated on a total of 1508 compounds (of which 199 are classified as being hazardous). The DSS is designed to predict EU risk phrases.

(Q)SARs might not enable replacement by themselves, but they are a valuable tool for screening and for prioritisation. Further develop (Q)SARs and/or expert system rulebases for skin irritation is recommended.

References

Barratt, M.D. (1996a). Quantitative structure-activity relationships for skin irritation and corrosivity of neutral and electrophilic organic chemicals. *Toxicology in Vitro* **10**, 247-256.

Hayashi, H., Nakamura, Y., Higashi, K., Kato, H., Kishida, F. & Kaneko, H. (1999). A quantitative structure-Activity relationship study of the skin irritation potential of phenols. *Toxicology in Vitro* **13**, 915-922.

Smith, J.S., Macina, O.T., Sussman, N.B., Luster, M.I. & Karol, M.H. (2000). A robust structure-activity relationship (SAR) model for esters that cause skin irritation in humans. *Toxicological Sciences* **55**, 215-222.

Cronin, M.T.D, Dearden J.C., Worth, A.P., Walker, J.D. In press. QSARs for predicting environmental-human health interactions. Environmental & Toxicological Chemistry.

Hulzebos, E.M., Maslankiewicz, L., Walker, J.D. (2003) Verification of literature-derived SARs for skin irritation and corrosion. QSAR Comb Sci 22, 351-363.

Patlewicz, G., Rodford, R., Walker, J.D. In press. QSARs for predicting skin and eye irritation. Environmental & Toxicological Chemistry.

Cronin, M.T.D., Jaworska, J.S., Walker, J.D., Comber, M.H.I., Watts, C.D. & Worth, A.P. (2003) Use of QSARs in International Decision-Making Frameworks to predict health effects of chemical substances. Environmental Health Perspectives 10, 1391-1402.

Gerner, I., Graetschel, G., Kahl, J., Schlede, E. (2000a) Development of a decision support system for the introduction of alternative methods into local irritancy/corrositivity testing strtaegies. Development of a relational database. ATLA 28, 11-28.

Gerner, I., Zinke, S., Graetschel, G., Schlede, E. (2000b) Development of a decision support system for the introduction of alternative methods into local irritancy/corrosivity testing strategies. Creation of fundamental rules for a decision support system. ATLA 28, 665-698.

Zinke, S., Gerner, I., Schlede, E.(2000) Local irritation/corrosion testing strategies: development of a decision support system for the introduction of alternative methods. ATLA 28, 29-40.

3. Identified steps or tests with no alternative methods available

The *in vitro* methods under validation only enable acute (single application) effects to be studied. Models/protocols have yet to be developed for evaluating adverse effects following repeated exposure and for determining the reversibility of these effects.
Furthermore, methods/protocols which measure the inflammatory response as well as the time course of an irritant reaction are only at research level.
Information on these aspects is key to the overall risk assessment and at present can only be derived from animal experiments, or is based on knowledge about skin irritation in humans which is still rather limited.

4. Summary of alternative methods which are currently available and foreseeable time to achieve peer reviewed validation

The human skin model assays (e.g. EpiDerm™ and EPISKIN™) and the mouse SIFT appear to be the most promising *in vitro* methods for skin irritation testing (Zuang *et al.*, 2002; Portes *et al.*, 2002; Heylings *et al*., 2003) (see table 2). The forthcoming ECVAM validation study may determine whether any of these methods can adequately distinguish acute skin irritants from non-irritants for classification and labelling purposes (2004-2005).

However, there is a need to develop mechanistically based endpoints that are more predictive of skin irritation than are simple cytotoxicity determinations. The existing *in vitro* models also need to be improved, so that they are better representative of the skin *in vivo*. For this purpose, more resources should be provided for research aimed at the identification of new markers for skin irritation, and on-going activities should be coordinated, with a view to identify new promising toxicological endpoints and develop new toxicity tests for validation. This research should be undertaken in parallel with the validation of existing test protocols for hazard identification. One approach to this research is through the application of genomics and proteomics which should facilitate the identification of specific biomarkers and the development of new, more predictive, *in vitro* methodologies. This promising approach is being progressed in the on-going COLIPA skin irritation research programme.

Key references

Basketter, D. A., Chamberlain, M., Griffiths, H. A., Rowson, M., Whittle, E., York, M. (1997). The classification of skin irritants by human patch test. *Food and Chemical Toxicology* **35**, 845-852.

Botham, P. A., Earl, L. K., Fentem, J. H., Roguet, R., van de Sandt, J. J. M. (1998) Alternative methods for skin irritation testing: the current status. ECVAM Skin Irritation Task Force Report 1. *ATLA* **26**, 195-211.

COLIPA (1995) Cosmetic Product Test Guidelines for the Assessment of Human Skin Compatibility.

COLIPA (1997) Test Guidelines for the Assessment of Human Skin Tolerance of Potentially Irritant Cosmetic Ingredients.

Faller C., Bracher M., Dami N., Roguet R. (2002). Predictive ability of reconstructed human epidermis equivalents for the assessment of skin irritation of cosmetics. *Toxicology in Vitro* 16, 557-572.

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

Fentem, J.H. and Botham, P.A. (2002). ECVAM's activities in validating alternative tests for skin corrosion and irritation. *ATLA* **30**, Suppl. 2, 61-67.

Heylings, J. R. Clowes, H.M., Hughes L. (2001). Comparison of tissue sources for the skin integrity function test (SIFT). *Toxicology in Vitro* **13**, 597-600.

Heylings J.R., Diot S., Esdaile D.J., Fasano W.J., Manning L.A., Owen H.M. (2003). A prevalidation study on the *in vitro* skin irritation function test (SIFT) for prediction of acute skin irritation *in vivo*: results and evaluation of ECVAM Phase III. *Toxicology in Vitro* **17**, 123-138.

OECD (2002). *OECD test guidelines 404: acute dermal irritation/corrosion.* Paris, France: OECD.

Ponec M. and Kempenaar J. (1995). Use of human skin recombinants as an *in vitro* model for test the irritation potential of cutaneous irritants. *Skin Pharmacol* **8**, 49-59.

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.

Robinson, M. K., Osborne, R., Perkins, M. A. (2000). *In vitro* and human testing strategies for skin irritation. *Annals of the New York Academy of Sciences* **919**, 192-204.

van de Sandt J., Roguet R., Cohen C., Esdaile D., Ponec M., Corsini E., Barker C., Fusenig N., Liebsch M., Benford D., de Brugerolle de Fraissinette A., Fortasch M. (1999). The use of keratinocytes and human skin models for predicting skin irritation. *ATLA* **27**, 723-743.

Worth, A.P. & Balls, M. (2002). Editors. *Alternative (Non-animal) Methods for Chemicals Testing: Current Status and Future Prospects.* A report prepared by ECVAM and the ECVAM Working Group on Chemicals. *ATLA* **30**, Suppl. 1.

Zuang, V. Balls, M., Botham, P.A., Coquette, A., Corsini, E., Curren, R.D., Elliott, G.R., Fentem, J.H., Heylings, J.R., Liebsch, M., Medina, J., Roguet, R., van de Sandt, H., Wiemann, C., Worth, A.P. (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.

SKIN CORROSION

1. Inventory of methods currently available

Skin corrosion tests assess the potential of a substance to cause irreversible damage to skin; namely, visible necrosis through the epidermis and into the dermis, following the application of a test substance for up to four hours. Corrosivity is not a feature one expects to occur with cosmetics, but occasionally could occur after a manufacturing mistake or misuse by the consumer. On the other hand, a cosmetic ingredient that has the intrinsic property to be corrosive, is not necessarily excluded for use in cosmetics. It very much depends on its final concentration in the cosmetic product, the presence of "neutralising" substances, the excipients used, the exposure route, the conditions of use, etc.

In the past skin corrosion was assessed using animal studies such as OECD TG 404, but since recently three alternative methods have been included in the Annex V of the Dangerous Substances Directive and are proposed as an OECD draft testing guideline. Moreover, recent refinement provisions included in the OECD testing guideline 404 (OECD, 2002a), recommend, prior to undertake the described *in vivo* test for corrosion/irritation of the substance, to perform a sequential testing strategy (see point 1 under skin irritation).

2. Alternative methods currently available

In vitro methods for skin corrosion

The current status of *in vitro* alternative methods for skin corrosion is summarised in table 1. Three methods were validated (Fentem et al., 1998; Liebsch et al., 2000; ECVAM, 1998, 2000) and included in Annex V of the Dangerous Substances Directive (Directive 67/548/EEC, 2000). These are :

- The *in vitro* skin corrosion rat skin transcutaneous electrical resistance (TER) test, which uses excised rat skin as a test system and its electrical resistance as an endpoint.

- The human skin model tests such as EPISKIN™ and EpiDerm™, which are reconstructed human epidermal equivalents and use the cell viability (MTT-test) as an endpoint.

The Corrositex™ test, which uses penetration of test substances through a hydrogenated collagen matrix (biobarrier) and supporting filter membrane, represents another corrosivity test which was considered to be useful only for acids, bases and their derivates (ECVAM, 2001; NIH, 1999). Although it passed the ECVAM Scientific Advisory Committee (ESAC), it has not been taken up in the EU legislation. It is nevertheless a legal test adopted by the US Department of Transport (US DOT) and a draft Test Guideline "In vitro Membrane Barrier Test Method for Skin Corrosion", which is based on the Corrositex™ test method, has been submitted to the OECD for Member Countries considerations.

A draft Test Guideline (TG) on *in vitro* tests for skin corrosion was submitted to the OECD in late 1998, for consideration by the OECD Member Countries. Following a number of commenting rounds, an expert meeting, held on 1-2 November 2001 in Berlin, agreed that the draft TG on *in vitro* skin corrosion should be divided into two separate TGs: a draft proposal for a new TG 430 on the TER test (not restricted to the rat skin TER test) and a draft proposal for a new TG 431 on the human skin model test. The new TG 430 and TG 431 were accepted by OECD Member Countries in June 2003 and have been sent to the OECD Council for final approval.

Structure-activity relationships for skin corrosion

Various SARs for skin corrosion have been reported by Barratt and colleagues (Barratt, 1996a; Barratt, 1996b; Whittle *et al.*, 1996; Barratt *et al.*, 1998). On the whole, the SARs presented in these studies take the form of principal component (PC) plots, which are based on physicochemical properties and show a separation between corrosive (C) and non-corrosive (NC) chemicals. Explicit classification models were not presented. Rather than modelling a heterogeneous group of chemicals, separate analyses were performed for acids, bases, electrophiles and neutral organics (defined as uncharged molecules which lack the potential to react covalently and which do not ionise under biological conditions [Martin Barratt, personal communication]). The most recent presentation of this approach is given in Barratt *et al.* (1998). In addition to PC analyses, discriminant analysis and neural network analysis were also applied to a group of neutral and electrophilic chemicals (Barratt, 1996a), and to the acids, bases and phenols (Barratt, 1996b). Finally, in another study (Barratt *et al.*, 1996), PC plots for acids were based not only on physicochemical properties, but also on *in vitro* cytotoxicity measurements in mouse 3T3 cells. More recently, it was shown that a heterogeneous set of organic chemicals could be predicted as C or NC on the basis of melting point (Mpt) and molecular weight (MW) (Worth, 2000), according to the following PM:

If Mpt ? 37 ?C and MW ? 123 g/mol, predict as C; otherwise predict as NC.

Similar rules have been developed by Gerner and colleagues, who have incorporated a system of decision rules into an expert system used by the German BgVV (Gerner *et al.*, 2000a, 2000b; Zinke *et al.*, 2000). An example is the following PM:

If MW > 1200 g/mol, then the substance has no local toxic effects.

Tiered testing strategies for skin corrosion

In 2002, a supplement to the OECD test guideline 404 for dermal irritation and corrosion was included. In this supplement a sequential (stepwise) testing strategy for hazard identification is recommended which allow for the best practice and ethical benchmark before undertaking *in vivo* experimentation. Some general refinement provisions have been introduced (OECD, 2002a).

The evaluation of a two-step strategy, based on the sequential use of pH measurements and *in vitro* data, indicated that the use of pH data in addition to TER or EPISKIN data, improves the ability to predict corrosion potential (Worth and Cronin, 2001a). An evaluation of a three-step strategy, based on the sequential use of QSARs, pH measurements and *in vitro* data, indicated that tiered approaches provide an effective means of classifying chemicals, while at the same time reducing and refining the use of animals (Worth *et al.*, 1998). A study carried out by ECVAM confirmed the usefulness of pH as a predictor of skin corrosion potential, and provided a new PM for identifying chemicals that are corrosive by a pH-dependent mechanism (Worth and Cronin, 2001b).

In the EU, a tiered testing strategy for skin corrosion/irritation is being proposed for incorporation into Annex V of *Directive 67/548/EEC*. This could be achieved by means of the 29th Adaptation to Technical Progress (ATP) of the directive (Juan Riego-Sintes, personal communication).

3. Future prospects and recommendations

Alternative methods for skin corrosion have been validated and accepted for regulatory use in the EU and the OECD Member Countries, so animal testing should not be performed for this endpoint. The hazard identification (classification and labelling) of skin corrosives should be based on the use of a pH test, where appropriate, and an *in vitro* test (rat skin TER assay, human skin model assay or, for qualifying test chemicals, CORROSITEX®). For risk assessment (dose-response investigations, coupled with assessments of skin irritation potential at doses negative in skin corrosion tests), the rat skin TER or a human skin model assay are recommended for use.

As a further recommendation is the validation of QSARs and/or expert system rulebases for skin corrosion.

Key references

Barratt, M.D. (1996a). Quantitative structure-activity relationships for skin irritation and corrosivity of neutral and electrophilic organic chemicals. *Toxicology in Vitro* **10**, 247-256.

Barratt, M.D. (1996b). Quantitative structure-activity relationships (QSARs) for skin corrosivity of organic acids, bases and phenols: principal components and neural network analysis of extended datasets. *Toxicology in Vitro* **10**, 85-94.

Barratt, M.D., Dixit, M.B., Jones, P.A. (1996). The use of *in vitro* cytotoxicity measurements in QSAR methods for the prediction of the skin corrosivity potential of acids. *Toxicology in Vitro* **10**, 283-290.

Barratt, M.D., Brantom, P.G., Fentem, J.H., Gerner, I., Walker, A.P., Worth, A.P. (1998). The ECVAM international validation study on *in vitro* tests for skin corrosivity. 1. Selection and distribution of the test chemicals. *Toxicology in Vitro* **12**, 471-482.

Directive 67/548/EEC (2000). Annex I to Commission Directive 2000/33/EC adapting to technical progress for the 27th time Council Directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labeling of dangerous substances. *Official Journal of the European Communities* **L136**, 91-97.

EC (2000). Annex I to *Commission Directive 2000/33/EC* adapting to technical progress for the 27th time *Council Directive 67/548/EEC* on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labeling of dangerous substances. *Official Journal of the European Communities* **L136**, 91-97.

ECVAM (1998). ECVAM News & Views. *ATLA* **26**, 275-280.

ECVAM (2000). Statement on the application of the Epiderm™ human skin model for skin corrosivity testing. *ATLA* **28**, 365-366.

ECVAM (2001). Statement on the application of the CORROSITEX® assay for skin corrosivity testing. *ATLA* **29**, 96-97.

Gerner, I., Graetschel, G., Kahl, J., Schlede, E. (2000a). Development of a decision support system for the introduction of alternative methods into local irritancy/corrosivity testing strategies. Development of a relational database. *ATLA* **28**, 11-28.

ICCVAM (2002) OECD Guideline for the Testing of Chemicals. Proposal for a New Guideline. *In Vitro* Membrane Barrier Test System for Skin Corrosion. 16pp. NC, USA: ICCVAM, NIEHS.

Liebsch, M., Traue, D., Barrabas, C., Spielmann, H., Uphill, P., Wilkins, S., Wiemann, C., Kaufmann, T., Remmele, M. & Holzhütter, H. G. (2000). The ECVAM prevalidation study on the use of EpiDerm for skin corrosivity testing. *ATLA* **28**, 371-401.

Gerner, I., Zinke, S., Graetschel, G., Schlede, E. (2000b). Development of a decision support system for the introduction of alternative methods into local irritancy/corrosivity testing strategies. Creation of fundamental rules for a decision support system. *ATLA* **28**, 665-98.

Fentem, J. H., Archer, G. E. B., Balls, M., Botham, P. A., Curren, R. D., Earl, L. K., Esdaile, D. J., Holzhütter, H. G. & Liebsch, M. (1998). The ECVAM international validation study on *in vitro* tests for skin corrosivity. 2. Results and evaluation by the Management Team. *Toxicology in Vitro* **12**, 483-524.

NIH (1999). *Corrositex®: an in vitro test method for assessing dermal corrosivity potential of chemicals*. NIH Publication No. 99-4495. Research Triangle Park, NC, USA: NIEHS.

NIH (2002). ICCVAM evaluation of EPISKIN™, EpiDerm™ (EPI-200), and the rat skin transcutaneous electrical resistance (TER) assay: in vitro test methods for assessing dermal corrosivity potential of chemicals. NIH Publication No. 02-4502. Research Triangle Park, NC, USA: NIEHS.

OECD (2002a). *OECD test guidelines 404: acute dermal irritation/corrosion.* Paris, France: OECD.

OECD (2002b) OECD Guideline for the Testing of Chemicals 430. *In Vitro* Skin Corrosion: Transcutaneous Electrical Resistance Test (TER). 11pp. Paris, France: OECD.

OECD (2002c) OECD Guideline for the Testing of Chemicals 431. *In Vitro* Skin Corrosion: Human Skin Model Test. 8pp. Paris, France: OECD.

Whittle, E.G., Barratt, M.D., Carter, J.A., Basketter, D.A., Chamberlain, M. (1996). The skin corrosivity potential of fatty acids: *in vitro* rat and human skin testing and QSAR studies. *Toxicology in Vitro* **10**, 95-100.

Worth A.P., Fenterm J.H., Balls M., Botham P.A., Curren R.D., Earl L.K., Esdaile D.J., Liebsch M. (1998). An evaluation of the proposed OECD testing strategy for skin corrosion. *ATLA* **26**, 709-720.

Worth A.P. and Cronin, M.T.D. (2000). Embedded cluster modelling: a novel QSAR method for generating elliptic models of biological activity. In *Progress in the Reduction, Refinement and Replacement of Animal Experimentation* (eds. M. Balls, A-M. van Zeller, M.E. Halder). pp. 479-491. Elsevier Science, Amsterdam.

Worth, A.P. and Cronin, M.T.D. (2001a). The use of bootstrap resampling to assess the uncertainty of Cooper statistics. *ATLA* **29**, 447-459.

Worth, A.P. and Cronin, M.T.D. (2001b). The use of pH measurements to predict the potential of chemicals to cause acute dermal and ocular toxicity. *Toxicology* **169**, 119-131.

Zinke, S., Gerner, I., Graetschel, G., Schlede, E. (2000). Local irritation/corrosion testing strategies: development of a decision support system for the introduction of alternative methods. *ATLA* **28**, 29-40.

4. Summary of the currently available alternative methods and foreseeable time to achieve ESAC endorsement

Table 2. Summary of validated and most advanced alternative methods for skin corrosion and skin irritation, respectively

Current endpoints addressed in the animal test	Alternative tests available	In vitro endpoints measured	Purpose	Area of application	Validation Status	Regulatory acceptance	Comments	Estimated time to have the method validated (ESAC endorsement)*
Skin corrosion: Full thickness destruction of skin tissue	Rat skin transcutaneous electrical resistance (TER) assay	Stratum corneum integrity and barrier function	Replacement	general; additional dye binding step for surfactants and solvents	Validated	accepted by regulatory authorities as a replacement test in the EU (B.40), but only as a screen for positives in the US; OECD TG 430		0 years
	EPISKIN™ human skin model (commercial system)	Cytotoxicity	Replacement	general; a few materials may interfere with MTT reduction	Validated	accepted by regulatory authorities as a replacement test in the EU (B.40), but only as a screen for positives in the US; OECD TG 431		0 years
	EpiDerm™ human skin model (commercial system)	Cytotoxicity	Replacement	general; a few materials may interfere with MTT reduction	Validated	accepted by regulatory authorities as a replacement test in the EU (B.40), but only as a screen for positives in the US; OECD TG 431		0 years

	CORROSITEX™ (commercial system)	//	Replacement (only for acids and bases and their derivatives)	mainly acids, bases and derivatives	Validated	validated and endorsed (US and EU) method for skin corrosion testing of acids and bases; proposal for a new OECD TG prepared by the US		0 years
Skin irritation: Erythema/eschar and/or oedema	EPISKIN™ human skin model (commercial system)	Cytotoxicity	Replacement [a]/ Partial replacement[b]	general; a few materials may interfere with MTT reduction	Under validation (ECVAM)		Covers one component of skin irritation (cytotoxicity)	2 years
	EpiDerm™ human skin model (commercial system)	Cytotoxicity	Replacement [a]/ Partial replacement[b]	general; a few materials may interfere with MTT reduction	Under validation (ECVAM)		Covers one component of skin irritation (cytotoxicity)	2 years
	Mouse skin integrity function test (SIFT)	Stratum corneum barrier function and integrity	Replacement [a]/ Partial replacement[b]	General; a few materials may interfere with the test system	Under validation (ECVAM)			2 years

* This table estimates the time needed to achieve ESAC endorsement for individual alternative tests assuming optimal conditions. It does not indicate the time needed to achieve full replacement of the animal test, nor does it include the time needed to achieve regulatory acceptance. “Optimal conditions” means that all necessary resources, for example technical, human, financial and coordination, are met at all times in the process and that the studies undertaken have successful outcomes.

[a] Replacement for hazard identification (classsification & labelling)

[b] Partial replacement (Reduction) for risk assessment purposes,

Conclusions from Table 2

The ECVAM validation study, if successful, may result in validated alternative methods which will enable partial replacement (reduction) for risk assessment purposes. If the study runs 2004-2005, we may have an international regulatory position by 2007/2008 (based on recent experience of how long these things take).

A more technical problem we might face, is the lack of clarity on when the GHS for classification will be introduced and how this could further adversely impact timelines for regulatory acceptance.

Moreover, for total replacement, we need to address the reversibility aspect, and dose-response/risk assessment based solely on in vitro/non-animal methods. The experts are not aware of much research on these elements. The current R&D activities are focused on improving the human skin models (closer to human skin) and identifying better endpoints/predictive markers of human skin irritation using molecular techniques (genomics, proteomics, etc.). These are also important building blocks for a new, integrated approach, to risk assessment of chemicals. This represents a considerable challenge which will keep us occupied well beyond 2009.

Finally, we should consider discussing the opportunity side of needing to use alternative techniques for skin irritation. Recently, a lot of thinking has gone into how to improve risk assessments for skin sensitisation by investigating dose-response relationships in hazard tests (e.g. the local lymph node assay) as a measure of potency. The same principle should be considered for irritation, at the moment, we're still working mainly on what happens to a neat test chemical . We should perhaps use the "higher throughput" potential of in vitro assays to do more dose-response studies.