Regulatory Toxicology and Pharmacology 64 (2012) 402–414

Contents lists available at SciVerse ScienceDirect

Regulatory Toxicology and Pharmacology

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

Applicability of *in vitro* tests for skin irritation and corrosion to regulatory classification schemes: Substantiating test strategies with data from routine studies

Susanne N. Kolle [a,*], Kristie M. Sullivan [b], Annette Mehling [c], Bennard van Ravenzwaay [a], Robert Landsiedel [a]

[a] BASF SE, Experimental Toxicology and Ecology, Carl-Bosch-Straße 38, 67056 Ludwigshafen, Germany
[b] Physicians Committee for Responsible Medicine, 5100 Wisconsin Ave. NW, Washington, DC 20016, USA
[c] BASF Personal Care and Nutrition GmbH, Henkelstraße 67, 40589 Düsseldorf, Germany

ARTICLE INFO

Article history:
Received 14 July 2012
Available online 30 August 2012

Keywords:
Skin irritation
Skin corrosion
Routine testing
In vitro
RhE
EpiDerm
OECD TG 431
OECD TG 439
Predictivity

ABSTRACT

Skin corrosion or irritation refers to the production of irreversible or reversible damage to the skin following the application of a test substance, respectively. Traditionally, hazard assessments are conducted using the *in vivo* Draize skin test, but recently *in vitro* tests using reconstructed human epidermis (RhE) models have gained regulatory acceptance. In this study, skin corrosion (SCT) and irritation tests (SIT) using a RhE model were implemented to reduce the number of *in vivo* tests required by regulatory bodies. One hundred and thirty-four materials were tested from a wide range of substance classes included 46 agrochemical formulations. Results were assessed according to UN GHS, EU-CLP, ANVISA and US EPA classification schemes. There was high correlation between the two *in vitro* tests. Assessment of the SCT sensitivity was not possible due to the limited number of corrosives in the data set; SCT specificity and accuracy were 89% for all classification systems. Accuracy (63–76%) and sensitivity (53–67%) were low in the SIT. Specificity and concordance for agrochemical formulations alone in both the SCT and SIT were comparable to the values for the complete data set (SCT: 91% vs. 89% specificity, 91% vs. 89% accuracy and SIT: 64–88% vs. 70–85% specificity, 56–75% vs. 63–76% accuracy).

1. Introduction

Assessment of the potential of a substance to cause damage to the skin is a basic endpoint evaluated in regulatory toxicology. This endpoint is used to predict the hazard, i.e. the intrinsic properties, of a substance upon accidental or intentional contact with the skin. Traditionally, the Draize skin irritation test has been used for many decades to predict skin irritation hazard (OECD, 2002; Draize et al., 1944). In this test, the test material is applied topically onto the shaved skin of rabbits. Skin corrosion or irritation refers to the production of irreversible or reversible damage to the skin over time following the application of a test material, respectively.

Within the EU, a harmonized approach to the classification and labeling of chemicals was implemented via the Dangerous Substances Directive (DSD,[1] 67/548/EEC (EU, 1967)) in 1967. The goal of this directive was to provide better protection for public health and the environment. This directive has now been replaced by the new European Chemical Regulation REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals (EU, 2006)). Worldwide, however, legislation found in various countries differs in the requirements instated for the purposes of the classification and labeling of substances. Labeling is also used to convey information on the hazards of a substance to users via the material safety data sheets. This non-harmonized approach is problematic not only in terms of transport and trade but can also hinder efforts to protect consumers and workers. An important step in worldwide harmonization was the adoption of the Globally Harmonized System (GHS) of Classification and Labeling of Chemicals (UN, 2011). This act has been/will be integrated into the national legislation of numerous countries around the world within the near future. Enactment of the full GHS has/will take place in a step-wise fashion, starting with chemical substances then moving

* Corresponding author. Fax: +49 621 60 58043.
E-mail address: susanne.kolle@basf.com (S.N. Kolle).

[1] AAALAC, Association for Assessment and Accreditation of Laboratory Animal Care; ANVISA, Agência Nacional de Vigilância sanitária (Brazilian National Health Surveillance Agency); C, corrosive; Cat, category; CAS, Chemical Abstract Service; DSD, EU Dangerous Substance Directive Classification 67/548/EEC (EU, 1967); ECVAM, European Centre for the Validation of Alternative Methods; ESAC, ECVAM Scientific Advisory Committee; EU, European Union; EU-CLP, Classification, Labelling, and Packaging, European GHS Regulation (EC) No. 1272/2008 (EU, 2008); FN, false negative; FNR, false negative rate; FP, false positive; FPR, false positive rate; I, irritant; GHS, Globally Harmonized System of Classification and Labeling of Chemicals (UN, 2011); GLP, Good Laboratory Practice; HET-CAM, hen's egg chorioallantoic membrane; MTT, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, thiazolyl blue, CAS No. 298-93-1; n/a, not applicable; NPV, negative predictive value; not cl, not classified; NC, not corrosive; nd, not determined; NI, not irritant; OECD, Organisation for Economic Co-operation and Development; PBS, phosphate buffered saline; PPV, positive predictive value; REACH, EU Regulation 190/2006 on the Registration, Evaluation, Authorisation, and Restriction of Chemicals (EU, 2006); RN, real negative; RP, real positive; RhE, reconstructed human epidermis; SCT, skin corrosion test; SDS, sodium dodecyl sulfate; SIT, skin irritation test; TG, test guideline; UN, United Nations; US EPA, United States Environmental Protection Agency.

0273-2300/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.yrtph.2012.08.015

Table 1
In vivo categorization cut-offs for the four major classification systems discussed.

US EPA/ANVISA	Category/Class IV	Category/Class III	Category/Class II	Category/Class I
Reaction Scores	≤ 2[a]	$> 2 \leq 3$[a]	> 3[a]	$> 0 \leq 4$[c]
UN GHS	Not Classified	Category 3	Category 2	Category 1 A/B/C
Reaction Scores	< 1.5[b]	$\geq 1.5 < 2.3$[b]	$\geq 2.3 \leq 4.0$[b]	$> 0 \leq 4$[c]
EU-CLP	Not Classified	Category 2		Category 1 A/B/C
Reaction Scores	<2.3[b]	$\geq 2.3 \leq 4.0$[b]		$> 0 \leq 4$[c]

[a] For US EPA categories II, III, and IV: reactions observed in more than 1 animal from gradings at 72 h. For ANVISA toxicity classes II, III, and IV: reactions observed in more than 1 animal from gradings at any observed timepoint.
[b] For UN GHS and EU-CLP categories 3 and 2: reactions in at least 2 of 3 tested animals from gradings at 24, 48 and 72 hours.
[c] For all classification systems discussed here, a corrosive substance is a test material that produces destruction of skin tissue, namely, visible necrosis through the epidermis and into the dermis, in at least 1 tested anima after exposure up to a 4 hour duration.

to mixtures. Within the EU, GHS came into force in 2009 via the legislation referred to as Classification, Labelling and Packaging System (CLP; European GHS Regulation (EC) No. 1272/2008 (EU, 2008)) and is an integral part of REACH. EU-CLP has been in effect for substances since December 2010, and will apply to mixtures as of June 2015.

Many of the evaluation schemes used for the identification of a health hazard were developed prior to GHS and some variations in classification can take place. In the case of skin irritation and corrosion, EU-CLP differentiates between nonirritating or slightly irritating substances (no classification needed), skin irritating substances (Category 2) and corrosive substances (Category 1 with three subcategories: 1A, 1B and 1C). GHS Category 3 (mild irritant classification) is optional (i.e. not classified according to EU-CLP). The cut-off scores used to differentiate between irritants and non-irritants have changed from an *in vivo* score of 2.0 to 2.3, which in turn has led to changes in classification. Substances with an *in vivo* score between 2.0 and 2.3, which were classified as being irritants under DSD, are nonirritants according to EU-CLP. A comparison of the cut-off values for all four systems used here is provided in Table 1. Further, while the UN GHS and EU-CLP systems calculate *in vivo* scores by averaging scores at observation time points, the US EPA and ANVISA (Brazilian National Health Surveillance Agency) systems use the highest single score. Finally, while the UN GHS, EU-CLP, and US EPA systems allow a period of time for recovery, the ANVISA system does not.

Over the past several years considerable progress has been made in the development of non-animal test methods for hazard identification. Increasing concerns for animal welfare and the ethics of animal testing has been taken into consideration in REACH (EU, 2006) and even more so in the amendments of the European Cosmetics Directive/Regulation (76/768/EEC (EU, 1976) and 1223/2009 (EU, 2009)). A number of methods for the identification of skin corrosion and skin irritation have gained a certain degree of regulatory acceptance. The methods described in OECD TG 431 (skin corrosion (OECD, 2004)) and OECD TG 439 (skin irritation (OECD, 2010)) utilize reconstructed human epidermis (RhE) models to which the test material is applied. In this study, skin corrosion and/or irritation test protocols using a RhE model were integrated into testing strategies to reduce the number of *in vivo* Draize skin irritation tests to be performed due to the requirements of regulatory bodies. The 134 materials tested included a wide range of substances from different chemical classes as well as 46 agrochemical formulations. The results were assessed in a regulatory context according to the GHS (UN, 2011), EU-CLP (EU, 2008), ANVISA (AENDA, 1992) and US EPA guidelines (EPA, 2007).

2. Materials and methods

2.1. Materials

Over the several years, 134 materials were tested in the *in vivo* rabbit skin irritation tests (OECD TG 404) for registration purposes. Using a tiered approach the materials were tested in the *in vitro* skin corrosion test (SCT; OECD TG 431) before subjected to *in vivo* testing. Thirty-eight materials were also tested in the *in vitro* skin irritation test (SIT; OECD TG 439). Identical batches of all tested materials were used in both tests. Eighty-seven materials were liquids (three were viscous) and 47 solids (one a waxy solid). Agrochemical formulations ($n = 46$) made up the largest proportion of the tested materials; other materials included acrylates ($n = 8$), agrochemicals ($n = 5$), amines ($n = 3$), boron compounds ($n = 6$), emollients ($n = 3$), ionic liquids ($n = 4$), pharmaceutical compounds ($n = 5$), pigment/dyes ($n = 8$), polymers ($n = 7$), surfactants ($n = 5$), and a range of other materials (Fig. 1 and Table 2).

2.2. In vivo acute dermal irritation/corrosion

The *in vivo* skin irritation test, initially described by Draize et al. (1944), was performed according to OECD TG 404 (OECD, 2002) in an AAALAC certified BASF SE laboratory under Good Laboratory Practice (GLP) conditions and according to the provisions of the German animal welfare regulations. The potential of the test materials to cause acute dermal irritation or corrosion was assessed by a single topical application of 0.5 mL of the liquid test materials to the intact skin of three White New Zealand rabbits (Centre Lago S.A., Vonnas, France) for 4 h. A stepwise procedure was used in which the test material was initially applied to one animal using a patch with an area of 2.5 cm $\times$ 2.5 cm and covered with (semi)-occlusive dressing and, depending on the severity of the reactions, then to an additional two animals. After removal of the patch, residual material was removed from the application area via rinsing. The cutaneous reactions were assessed immediately after removal of the patch, approximately 1, 24, 48 and 72 h after removal of the patch, and then in weekly intervals until day 14. Skin reactions were evaluated by grading erythema, eschar formation and edema formation. Classifications presented here are based on the results of the Draize skin irritation test and the criteria of the different classification systems (Tables 3a and 3b). Here the optional Category 3 in the UN GHS system is used. All *in vivo* studies were performed as regulatory requirements. No additional animal testing was performed for the purpose of this study.

2.3. In vitro skin corrosion test (SCT)

The skin corrosion test using RhE was conducted in accordance with OECD TG 431 (OECD, 2004). Briefly, the potential of the test materials to cause dermal corrosion was assessed following a single topical application of 50 µL for liquids or 25 µL bulk volume for solids of the neat test material to a reconstructed three-dimensional human epidermis model (EpiDerm™, MatTek Corporation, Ashland, MA, USA). For this purpose, two EpiDerm™ tissues per treatment were incubated with the test material for 3 min and 1 h each. Tissue destruction was determined by measuring the