Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

04-Aug-2015

English - Or. English

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND
THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

PERFORMANCE STANDARDS FOR THE ASSESSMENT OF PROPOSED SIMILAR OR MODIFIED TEST METHODS FOR SKIN CORROSION TESTING AS DESCRIBED IN TG 430

Series on Testing & Assessment
No. 218

JT03380686

ENV/JM/MONO(2015)25
Unclassified
English - Or. English

OECD Environment, Health and Safety Publications

Series on Testing and Assessment

No. 218

PERFORMANCE STANDARDS FOR THE ASSESSMENT OF PROPOSED SIMILAR OR MODIFIED *IN VITRO* TRANSCUTANEOUS ELECTRICAL RESISTANCE (TER) TEST METHODS FOR SKIN CORROSION TESTING

AS DESCRIBED IN TG 430[1]

(Intended for the developers of new or modified similar test methods)

IOMC
INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS
A cooperative agreement among FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD

Environment Directorate
ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT
Paris 2015

[1] Proposed new similar or modified test method following the PS of this Test Guideline should be submitted to the OECD for adoption and inclusion into the Test Guideline before being used for regulatory purposes.

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 34 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in eleven different series: **Testing and Assessment; Good Laboratory Practice and Compliance Monitoring; Pesticides; Biocides; Risk Management; Harmonisation of Regulatory Oversight in Biotechnology; Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; and Safety of Manufactured Nanomaterials.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (http://www.oecd.org/chemicalsafety/).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organisations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

This publication is available electronically, at no charge.

Also published in the Series on Testing and Assessment link

For this and many other Environment, Health and Safety publications, consult the OECD's World Wide Web site (www.oecd.org/chemicalsafety/)

or contact:

OECD Environment Directorate,
Environment, Health and Safety Division
2 rue André-Pascal
75775 Paris Cedex 16
France

Fax: (33-1) 44 30 61 80

E-mail: ehscont@oecd.org

FOREWORD

This document contains the Performance Standards (PS) for the validation of similar or modified test methods to the Transcutaneous Electric Resistance (TER) for skin corrosion as described in TG 430. In the past, PS were usually annexed to TGs. However, in view of separating information on the *use* of a test method as contained in the TG from information needed to *validate* test methods as contained in the PS, TGs and PS will now both be stand-alone documents. This approach had been agreed by the Working Group of the National Coordinators of the Test Guidelines Programme (WNT). In case of the current PS for skin *in vitro* corrosion methods according to TG 430, the text was reviewed in regard to harmonising with other relevant documents addressing skin irritation and skin corrosion. The PS were reviewed by the OECD Expert Group on Skin Irritation/Corrosion in November 2014. The PS are intended for the developers of new or modified similar test methods to the validated reference method. The present document was approved by the WNT in April 2015, declassified and published under the responsibility of the Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides, and Biotechnology on 10 July 2015.

INTRODUCTION

1. This document contains Performance Standards which allow, in accordance with the principles of Guidance Document No.34 (1), determining the validation status (reliability and relevance) of similar and modified skin corrosion test methods that are structurally and mechanistically similar to the TER test method in OECD Test Guideline 430 (2).

2. These PS include the following sets of information: (i) Essential Test Method Components that serve to evaluate the structural, mechanistic and procedural similarity of a new similar or modified proposed test method, (ii) a list of 24 Reference Chemicals to be used for validating new or modified test methods and (iii) defined target values of reproducibility and predictive capacity that need to be met by proposed test methods in order to be considered similar to the validated reference method.

3. The purpose of Performance Standards (PS) is to provide the basis by which new similar or modified test methods, both proprietary (*i.e.* copyrighted, trademarked, registered) and non-proprietary, can be deemed to be structurally and mechanistically similar to a Validated Reference Method (VRM) and demonstrate to have sufficient reliability and relevance for specific testing purposes (i.e., scientifically valid), in accordance with the principles of Guidance Document No. 34 (1). The PS, based on scientifically valid and accepted test method(s), can be used to evaluate the reliability and relevance of test methods that are based on similar scientific principles and measure or predict the same biological or toxic effect (1). Such methods are referred to as *similar* or *"me-too"* test methods. Moreover, the PS may be used to evaluate *modified* test methods, which may propose potential improvements in comparison to approved earlier versions of a method. In such cases the PS can be used to determine the effect of the proposed changes on the test method's performance and the extent to which such changes may affect the information available for other components of the validation process (e.g. relating to Essential Test Method Components). However, depending on the number and nature of the proposed changes as well as the data and documentation available in relation to these changes, modified test methods may: i) either be found unsuitable for a PS-based validation (e.g. if the changes are so substantial that the method is not any longer deemed sufficiently similar with regard to the PS), in which cases they should be subjected to the same validation process as described for a new test method (1); or ii) suitable for a limited assessment of reliability and relevance using the established PS (1). Similar or modified new test methods (i.e., "me-too" tests) successfully validated according to Performance Standards can be added to TG 430. However, Mutual Acceptance of Data (MAD) will only be guaranteed for those test methods reviewed and adopted by the OECD. Proposed similar or modified test methods validated according to these PS should therefore be submitted to the OECD for adoption and inclusion into TG 430 before being used for regulatory purposes.

4. These PS are based on the ICCVAM PS (3) for evaluating the validity of new or modified TER test methods. The PS consists of: (i) Essential Test Method Components; (ii) Recommended Reference Chemicals, and; (iii) Defined Reliability and Predictive Capacity Values that the proposed similar or modified test method should meet or exceed. The VRM used to develop the present PS is the Transcutaneous Electrical Resistance test as described in TG 430 (2). Definitions are provided in Annex I.

5. Similar (me-too) or modified test methods proposed for use under Test Guideline 430 (2) should be evaluated to determine their reliability and predictive capacity using Reference Chemicals representing the full range of the TG 404 *in vivo* corrosivity scores (Table 1) prior to their use for testing other chemicals, in order to ensure that these methods are able to identify correctly UN GHS Category 1 corrosive chemicals (which includes UN GHS Sub-categories 1A, 1B, and 1C) and non-corrosive chemicals (4) (5). The proposed similar or modified test methods should have reproducibility, sensitivity,

specificity and accuracy which are equal or better than those derived from the VRM (6) and as described in paragraphs 28 to 32 of these PS (table 2).

ESSENTIAL TEST METHOD COMPONENTS

6. The Essential Test Method Components consist of essential structural, functional, and procedural elements of scientifically valid reference method (the VRM) that should be included in the protocol of a proposed, mechanistically and functionally similar or modified test method. These components include unique characteristics of the test method, critical procedural details, and quality control measures. Adherence to essential test method components will help to assure that a similar or modified proposed test method is based on the same concepts as the corresponding VRM (1) (2). The essential test method components to be considered for similar or modified test methods related to TG 430 are described in detail in the following paragraphs.

7. For specific parameters or modified procedures, adequate values or procedures should be provided for the proposed similar or modified test method; these specific values or procedures may vary depending on the specific test method and/or its modification. For example in the TER test method, the cut-off value distinguishing corrosive from non-corrosive test chemical is highly dependent on the nature of the skin preparations (source animals) and the equipment used (7) (8).

Animals

8. Rats are the species of choice because the sensitivity of their skin to test chemicals in this test method has been previously demonstrated (9) and rat skin is the only skin source that has been formally validated (6) (10). The age (when the skin is collected) and strain of the rat is particularly important to ensure that the hair follicles are in the dormant phase before adult hair growth begins.

9. The dorsal and flank hair from young, approximately 22 day-old, male or female rats (Wistar-derived or a comparable strain), is carefully removed with small clippers. Then, the animals are washed by careful wiping, whilst submerging the clipped area in antibiotic solution (containing, for example, streptomycin, penicillin, chloramphenicol, and amphotericin, at concentrations effective in inhibiting bacterial growth). Animals are washed with antibiotics again on the third or fourth day after the first wash and are used within 3 days of the second wash, when the *stratum corneum* has recovered from the hair removal.

Preparation of the skin discs

10. Animals are humanely killed when 28-30 days old; this age is critical. The dorso-lateral skin of each animal is then removed and stripped of excess subcutaneous fat by carefully peeling it away from the skin. Skin discs, with a diameter of approximately 20-mm each, are removed. The skin may be stored before discs are used where it is shown that positive and negative control data are equivalent to that obtained with fresh skin.

11. Each skin disc is placed over one of the ends of a PTFE (polytetrafluoroethylene) tube, ensuring that the epidermal surface is in contact with the tube. A rubber 'O' ring is press-fitted over the end of the tube to hold the skin in place and excess tissue is trimmed away. The rubber 'O' ring is then carefully sealed to the end of the PTFE tube with petroleum jelly. The tube is supported by a spring clip inside a receptor chamber containing $MgSO_4$ solution (154 mM) (Figure 1). The skin disc should be fully submerged in the $MgSO_4$ solution. As many as 10-15 skin discs can be obtained from a single rat skin. Tube and 'O' ring dimensions of the VRM are shown in Figure 2.

12. Before testing begins, the TER of two skin discs are measured as a quality control procedure for each animal skin. Both discs should give electrical resistance values greater than 10 kΩ for the remainder of the discs to be used for the test method. If the resistance value is less than 10 kΩ, the remaining discs from that skin should be discarded.

Application of the test chemical and control substances

13. Concurrent positive and negative controls should be used for each run (experiment) to ensure adequate performance of the experimental model. Skin discs from a single animal should be used in each run (experiment). For the VRM, the suggested positive and negative control test chemicals are 10M hydrochloric acid and distilled water, respectively.

14. Liquid test chemicals are applied uniformly to the epidermal surface inside the tube (e.g., 150 µL in the VRM). When testing solid materials, a sufficient amount of the solid is applied evenly to the disc to ensure that the whole surface of the epidermis is covered. Deionised water may be added on top of the solid and the tube gently agitated (e.g., 150 µL in the VRM). In order to achieve maximum contact with the skin, solids may need to be warmed to 30^0 C to melt or soften the test chemical, or ground to produce a granular material or powder.

15. In the VRM, three skin discs are used for each test and control chemical in each testing run (experiment). Test chemicals are applied for 24 hours at 20-23^0 C. The test chemical is removed by washing with a jet of tap water at up to room temperature until no further material can be removed.

TER measurements

16. In the VRM, the skin impedance is measured as TER by using a low-voltage, alternating current Wheatstone bridge (11). General specifications of the bridge are 1-3 Volt operating voltage, a sinus or rectangular shaped alternating current of 50 - 1000 Hz, and a measuring range of at least 0.1 -30 kΩ. The databridge used in the validation study measured inductance, capacitance and resistance up to values of 2000H, 2000 µF, and 2 MΩ, respectively at frequencies of 100Hz or 1kHz, using series or parallel values. For the purposes of the TER VRM corrosivity assay measurements are recorded in resistance, at a frequency of 100 Hz and using series values. Prior to measuring the electrical resistance, the surface tension of the skin is reduced by adding a sufficient volume of 70% ethanol to cover the epidermis. After a few seconds, the ethanol is removed from the tube and the tissue is then hydrated by the addition of 3 mL $MgSO_4$ solution (154 mM). The databridge electrodes are placed on either side of the skin disc to measure the resistance in kΩ/skin disc (Figure 1). Electrode dimensions and the length of the electrode exposed below the crocodile clips are shown in Figure 2. The clip attached to the inner electrode is rested on the top of the PTFE tube during resistance measurement to ensure that a consistent length of electrode is submerged in the $MgSO_4$ solution. The outer electrode is positioned inside the receptor chamber so that it rests on the bottom of the chamber. The distance between the spring clip and the bottom of the PTFE tube is maintained as a constant (Figure 2), because this distance affects the resistance value obtained. Consequently, the distance between the inner electrode and the skin disc should be constant and minimal (e.g., 1-2 mm in the VRM).

17. In the VRM, if the measured resistance value is greater than 20 kΩ, this may be due to the remains of the test chemical coating the epidermal surface of the skin disc. Further removal of this coating can be attempted, for example, by sealing the PTFE tube with a gloved thumb and shaking it for approximately 10 seconds; the $MgSO_4$ solution is discarded and the resistance measurement is repeated with fresh $MgSO_4$.

18. The properties and dimensions of the test apparatus and the experimental procedure used may influence the TER values obtained. The 5 kΩ corrosive threshold was developed from data obtained with the specific apparatus and procedure here described. Different threshold and control values may apply if the test conditions are altered or a different apparatus is used. Therefore, it is necessary to calibrate the methodology and resistance threshold values by testing a series of Proficiency Substances chosen from the substances used in the validation study (6) (10), or from similar chemical classes to the test chemicals being investigated. A set of suitable Proficiency Substances is identified in TG 430 (2).

Dye Binding Procedures

19. Exposure of certain non-corrosive materials can result in a reduction of resistance below the cut-off (i.e. 5 kΩ in the VRM) allowing the passage of ions through the *stratum corneum*, thereby reducing the electrical resistance (6). For example, neutral organics and test chemicals that have surface-active properties (including detergents, emulsifiers and other surfactants) can remove skin lipids making the barrier more permeable to ions. Thus, if TER values produced by such chemicals are less than or around the cut-off (i.e. 5 kΩ in the VRM) in the absence of visually perceptible damage of the skin discs, an assessment of dye penetration should be carried out on the control and treated tissues to determine if the TER values obtained were the result of increased skin permeability, or skin corrosion (6) (12). In case of the latter where the *stratum corneum* is disrupted, the dye sulforhodamine B, when applied to the skin surface rapidly penetrates and stains the underlying tissue. This particular dye is stable to a wide range of test chemicals and is not affected by the extraction procedure described below.

20. Following TER assessment, the magnesium sulphate is discarded from the tube and the skin is carefully examined for obvious damage. If there is no obvious major damage (*e.g.* perforation), in the VRM a total of 150 µL of a 10% (w/v) dilution in distilled water of the dye sulforhodamine B (Acid Red 52; C.I. 45100; CAS number 3520-42-1), is applied to the epidermal surface of each skin disc for 2 hours. The skin discs are then washed with tap water at up to room temperature for approximately 10 seconds to remove any excess/unbound dye. Each skin disc is carefully removed from the PTFE tube and placed in a vial (*e.g.* a 20-mL glass scintillation vial) containing deionised water (8 mL in the VRM). The vials are agitated gently for 5 minutes to remove any additional unbound dye. This rinsing procedure is then repeated, after which the skin discs are removed and placed into vials containing 5ml of 30% (w/v) sodium dodecyl sulphate (SDS) in distilled water and are incubated overnight at 60^0 C.

21. After incubation, in the VRM, each skin disc is removed and discarded and the remaining solution is centrifuged for 8 minutes at 21^0 C (relative centrifugal force ~175 x g). A 1mL sample of the supernatant is diluted 1 in 5 (v/v) [*i.e.* 1mL + 4mL] with 30% (w/v) SDS in distilled water. The optical density (OD) of the solution is measured at 565 nm.

22. The sulforhodamine B dye content per disc is calculated from the OD values (6) (sulforhodamine B dye molar extinction coefficient at 565nm = 8.7 x $l0^4$; molecular weight = 580). The dye content is determined for each skin disc by the use of an appropriate calibration curve and mean dye content is then calculated for the replicates.

Acceptability Criteria

23. The mean TER results are accepted if the concurrent positive and negative control values fall within the acceptable ranges for the method in the testing laboratory. The acceptable resistance ranges for the VRM methodology and apparatus described above are given in the following table:

Control	Substance	Resistance range (kΩ)
Positive	10M Hydrochloric acid	0.5 - 1.0

Negative	Distilled water	10 - 25

24. The mean dye binding results are accepted on condition that concurrent control values fall within the acceptable ranges for the method. The VRM suggested acceptable dye content ranges for the control substances based on the methodology and apparatus described above are given in the following table:

Control	Substance	Dye content range (μg/disc)
Positive	10M Hydrochloric acid	40 - 100
Negative	Distilled water	15 - 35

Interpretation of results

25. The cut-off TER value distinguishing corrosive from non-corrosive test chemicals for the VRM was established during test method optimization, tested during a pre-validation phase, and confirmed in a formal validation study.

26. The prediction model for the VRM rat skin TER skin corrosion test method (6) (7), associated with the UN GHS (4) classification system, is given below:

The test chemical is considered to be non-corrosive to skin:

i) if the mean TER value obtained for the test chemical is greater than (>) 5 kΩ,
OR
ii) the mean TER value obtained for the test chemical is less than or equal to (≤) 5 kΩ, **AND**
- the skin discs show no obvious damage(*e.g.* perforation), **AND**
- the mean disc dye content is less than (<) the mean disc dye content of the 10M HCl positive control obtained concurrently (see paragraph 22 for positive control values).

The test chemical is considered to be corrosive to skin:

i) if the mean TER value obtained for the test chemical is less than or equal to (≤) 5 kΩ **AND** the skin discs are obviously damaged(*e.g.* perforated),
OR
ii) the mean TER value obtained for the test chemical is less than or equal to (≤) 5 kΩ, **AND**
- the skin discs show no obvious damage(*e.g.* perforation), **AND**
- the mean disc dye content is greater than or equal to (≥) the mean disc dye content of the 10M HCl positive control obtained concurrently (see paragraph 22 for positive control values).

27. A testing run (experiment) composed of at least three replicate skin discs should be sufficient for a test chemical when the classification is unequivocal. However, in cases of borderline results, such as non-concordant replicate measurements and/or mean TER equal to 5 ± 0.5 kΩ, a second independent testing run (experiment) should be considered, as well as a third one in case of discordant results between the first two testing runs (experiments).

MINIMUM LIST OF REFERENCE CHEMICALS

28. Reference Chemicals are used to determine if the reliability and predictive capacity of a proposed similar or modified test method, proven to be structurally and functionally sufficiently similar to the VRM, or representing a minor modification of the VRM, are equal or better than those derived from the VRM (6). The 24 recommended Reference Chemicals listed in Table 1 include chemicals representing different chemical classes (*i.e.* chemical categories based on functional groups), and are representative of the full

range of TG 404 *in vivo* skin corrosion scores. The chemicals included in this list comprise the following UN GHS (Sub-)categories: 5 Sub-category 1A chemicals, 7 chemicals of Sub-categories 1B and 1C (the *in vivo* data do not permit distinction between the two sub-categories) as well as 12 non-corrosive chemicals. The Reference Chemicals were selected from the test chemicals used in the validation study of the VRM (6) (10) using the selection criteria as described in Table 1 (foot-note 1), with due regard to e.g., chemical functionality and physical state.

29. The 24 Reference Chemicals listed in Table 1 represent the minimum number of chemicals that should be used to evaluate the reliability and predictive capacity of a proposed similar or modified test method. The exclusive use of these Reference Chemicals for the development/optimization of new similar test methods should be avoided to the extent possible. In situations where a listed Reference Chemical is unavailable, or cannot be used for other justified reasons, another chemical could be used provided it fulfils the selection criteria as described in Table 1 (foot-note 1) and for which adequate *in vivo* reference data are available could be used (5), e.g. primarily from the test chemicals used in the validation study of the VRM (6). To gain further information on the predictive capacity of the proposed test method, additional chemicals representing other chemical classes and for which adequate *in vivo* reference data are available may be tested in addition to the minimum list of Reference Chemicals.

Table 1: Minimum list of Reference Chemicals for determination of Reproducibility and Predictive Capacity of similar or modified *in vitro* TER skin corrosion test methods

Chemical[1]	CASRN	Chemical Class[2]	UN GHS Cat. based on *In Vivo* Results (4)[3]	VRM Cat. based on *In Vitro* Results	Physical State	pH[4]
***In Vivo* Corrosives**						
Phosphorus tribromide	7789-60-8	inorganic acid	1A	6 x C	L	1.0
Boron trifluoride dihydrate	13319-75-0	inorganic acid	1A	6 x C	L	1.5
Phosphorus pentachloride	10026-13-8	inorganic acid	1A	6 x C	S	ND
N,N'-Dimethyl dipropylenetriamine	10563-29-8	organic base	1A	6 x C	L	8.3
1,2-Diaminopropane	78-90-0	organic base	1A	6 x C	L	8.3
Sulfuric acid (10%)	7664-93-9	inorganic acid	(1A/)1B/1C	5 x C 1x NC	L	1.2
Potassium hydroxide (10% aq.)	1310-58-3	inorganic base	(1A/)1B/1C	6 x C	L	13.2
Hexanoic acid	142-62-1	organic acid	(1A/)1B/1C	6 x C	L	3.9
Octanoic (Caprylic) acid	124-07-2	organic acid	1B/1C	4 x C 2 x NC	L	3.6
N,N-Dimethyl isopropylamine	996-35-0	organic base	1B/1C	6 x C	L	8.3
n-Heptylamine	111-68-2	organic base	1B/1C	6 x C	L	8.4
2-tert-Butylphenol	88-18-6	phenol	1B/1C	4 x C 2 x NC	L	3.9
***In Vivo* Non-corrosives**						
Sulfamic acid	5329-14-6	inorganic acid	NC	5 x C 1 x NC	S	1.5
Sodium carbonate (50% aq.)	497-19-8	inorganic base	NC	6 x C	L	11.7
Isostearic acid	2724-58-5	organic acid	NC	6 x NC	L	3.6
Dodecanoic acid (Lauric acid)	143-07-7	organic acid	NC	6 x NC	S	ND
4-Amino-1,2,4-	584-13-4	organic base	NC	6 x NC	S	5.5

Chemical[1]	CASRN	Chemical Class[2]	UN GHS Cat. based on *In Vivo* Results (4)[3]	VRM Cat. based on *In Vitro* Results	Physical State	pH[4]
triazole						
Eugenol	97-53-0	phenol	NC	1 x C 5 x NC	L	3.6
2-Methoxyphenol	90-05-1	phenol	NC	6 x NC	L	3.9
Phenethyl bromide	103-63-9	electrophile	NC	6 x NC	L	3.6
4-(Methylthio)-benzaldehyde	3446-89-7	electrophile	NC	6 x NC	L	6.8
1,9-Decadiene	1647-16-1	neutral organic	NC	6 x NC	L	3.9
Tetrachloroethylene	127-18-4	neutral organic	NC	6 x NC	L	4.5
Sodium lauryl sulfate (20% aq.)	151-21-3	surfactant	NC	6 x C	L	3.9

Abbreviations: aq = aqueous; CASRN = Chemical Abstracts Service Registry Number; UN GHS = United Nations Globally Harmonised System (1); VRM = Validated Reference Method; ND = Not Determined.

[1]The Reference Chemicals, sorted first by corrosives versus non-corrosives, then by corrosive sub-category and then by chemical class, were selected from the test chemicals used in the ECVAM validation study of the rat skin TER test method (6)(10). Unless otherwise indicated, these chemicals were tested at the purity level obtained when purchased from a commercial source (10). The selection included, to the extent possible, chemicals that: (i) are representative of the range of corrosivity responses (*e.g*. non-corrosives; weak to strong corrosives) that the VRM is capable of measuring or predicting; (ii) are representative of the chemical classes used in the validation study; (iii) reflect the performance characteristics of the VRM; (iv) have chemical structures that are well-defined; (v) induce definitive results in the *in vivo* reference test method; (vi) are commercially available; and (vii) are not associated with prohibitive disposal costs.

[2]Chemical class assigned by Barratt *et al.* (10).

[3]The corresponding UN Packing groups are I, II and III, respectively, for the UN GHS 1A, 1B and 1C.

[4]The pH values were obtained from Fentem *et al.* (6) and Barratt *et al.* (10).

DEFINED RELIABILITY AND PREDICTIVE CAPACITY VALUES

30. For purposes of establishing the reliability (i.e., within- and between-laboratory reproducibility) and predictive capacity (i.e., sensitivity, specificity and accuracy) of proposed similar or modified TER test methods to be used by several independent laboratories, all 24 Reference Chemicals listed in Table 1 should be tested in at least three laboratories. In each laboratory, all 24 Reference Chemicals should be tested in three independent runs performed with skin discs obtained from different animals and at sufficiently spaced time points. Each testing run should consist of at least three concurrently tested skin discs for each test chemical, negative and positive control, all obtained from the same animal.

31. The calculation of the within-laboratory reproducibility, between-laboratory reproducibility, predictive capacity (i.e. sensitivity, specificity and accuracy) values of the proposed test method should be done according to the rules described below to ensure that a predefined and consistent approach is used:

1. Within-laboratory reproducibility (WLR) should be calculated based on concordance of classifications using at least two qualified testing runs from Reference Chemicals.

2. For the calculation of between-laboratory reproducibility (BLR) the final classification for each Reference Chemical in each participating laboratory should be obtained by using the arithmetic mean TER and dye binding values over the different qualified testing runs performed. BLR should be calculated based on concordance of classifications using only qualified testing runs obtained with the Reference Chemicals for which at least one qualified testing run per laboratory is available.

3. The calculation of predictive capacity (i.e. sensitivity, specificity and accuracy values) should be done using all qualified testing runs obtained for each Reference Chemical in each laboratory. The calculations should be based on the individual predictions of each qualified testing run for each Reference Chemical in each laboratory and not on the arithmetic mean TER and dye binding values over the different qualified tests performed.

In this context, a qualified testing run consists of at least three replicates tested concurrently within a qualified run that meets the acceptance criteria for the negative and positive control, as defined in the corresponding SOP. Otherwise, the testing run is considered as non-qualified.

Within-laboratory reproducibility

32. An assessment of within-laboratory reproducibility should show in every laboratory a concordance of predictions (corrosive or non-corrosive) obtained in different, independent runs of the 24 Reference Chemicals equal or higher (≥) than 90% (actual for rat skin TER: 87.5%, 91.7% and 100% in each laboratory, respectively).

Between-laboratory reproducibility

33. An assessment of between-laboratory reproducibility should show a concordance of predictions (corrosive or non-corrosive) between a minimum of three laboratories, obtained for the 24 Reference Chemicals, equal or higher (≥) than 80% (actual for rat skin TER: 95.8 to 79.2%, 1 to 5 chemicals non-concordant).

Predictive capacity

34. The predictive capacity of the proposed similar or modified TER test method should be equal or better than the target values derived from the VRM (Table 2). The sensitivity, specificity and accuracy obtained with the 24 relevant Reference Chemicals listed in Table 1 should be equal or higher (≥) than 90%, 75% and 82.5% respectively (Table 2).

Table 2: Required sensitivity, specificity and accuracy for similar or modified TER skin corrosion test methods to be considered valid to discriminate corrosive from non-corrosive chemicals (C vs. NC) but not able to sub-categorize corrosive chemicals

Sensitivity	Specificity	Accuracy
≥ 90% (actual for rat skin	≥ 75% (actual for rat skin	≥ 82.5% (actual for rat skin

TER[1]: 93%)	TER[1]:75%)	TER[1]: 84%)

[1] Values based on the results for the VRM TER for the 24 Reference Chemicals listed in Table 1.

LITERATURE

(1) OECD (2005), OECD Series on Testing and Assessment No. 34. Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Available at: [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2005)14&doclanguage=en].

(2) OECD (2015), OECD Guideline for the Testing of Chemicals No. 430. *In Vitro* Skin Corrosion: Transcutaneous Electrical Resistance (TER). OECD, Paris.

(3) ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods) (2004), Recommended Performance Standards for *In Vitro* Test Methods for Skin Corrosion. NIH Publication Number 04-4510. Research Triangle Park, NC: National Institute of Environmental Health Sciences.

(4) United Nations (UN) (2013), Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Second revised edition, UN New York and Geneva, 2013.

(5) OECD (2015), OECD Guideline for Testing of Chemicals. No. 404: Acute Dermal Irritation, Corrosion, OECD, Paris.

(6) Fentem, J.H., Archer, G.E.B., Balls, M., Botham, P.A., Curren, R.D., Earl, L.K., Esdaile, D.J., Holzhütter, H.-G., and 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.

(7) TER SOP, *INVITTOX* Protocol No. 115 (December 2008), Rat Skin Transcutaneous Electrical Resistance (TER) Test. Available at: [http://ecvam-dbalm.jrc.ec.europa.eu/].

(8) Davies DJ, Ward RJ, Heylings (2004), Multi-species assessment of electrical resistance as a skin integrity marker for *in vitro* percutaneous absorption studies, JR.Toxicol *In Vitro*.;18(3):351-8.

(9) ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods) (2002), ICCVAM evaluation of EpiDerm™ (EPI-200), EPISKIN™ (SM), 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. National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA.

(10) Barratt, M.D., Brantom, P.G., Fentem, J.H., Gerner, I., Walker, A.P., and 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.

(11) Botham, P.A., Hall, T.J., Dennett, R., McCall, J.C., Basketter, D.A., Whittle, E., Cheeseman, M., Esdaile, D.J., and Gardner, J. (1992),The skin corrosivity test *in vitro:* results of an interlaboratory trial. *Toxicology in Vitro* 6,191-194.

(12) Botham, P.A., Chamberlain, M., Barratt, M.D., Curren, R.D., Esdaile, D.J., Gardner, J.R., Gordon, V.C., Hildebrand, B., Lewis, R.W., Liebsch, M., Logemann, P., Osborne, R., Ponec, M., Regnier, J.F., Steiling, W., Walker, A.P., and Balls, M. (1995), A prevalidation study on *in vitro* skin corrosivity testing. The report and recommendations of ECVAM Workshop 6. *ATLA* 23, 219-255.

Figure 1: Apparatus for the rat skin TER assay

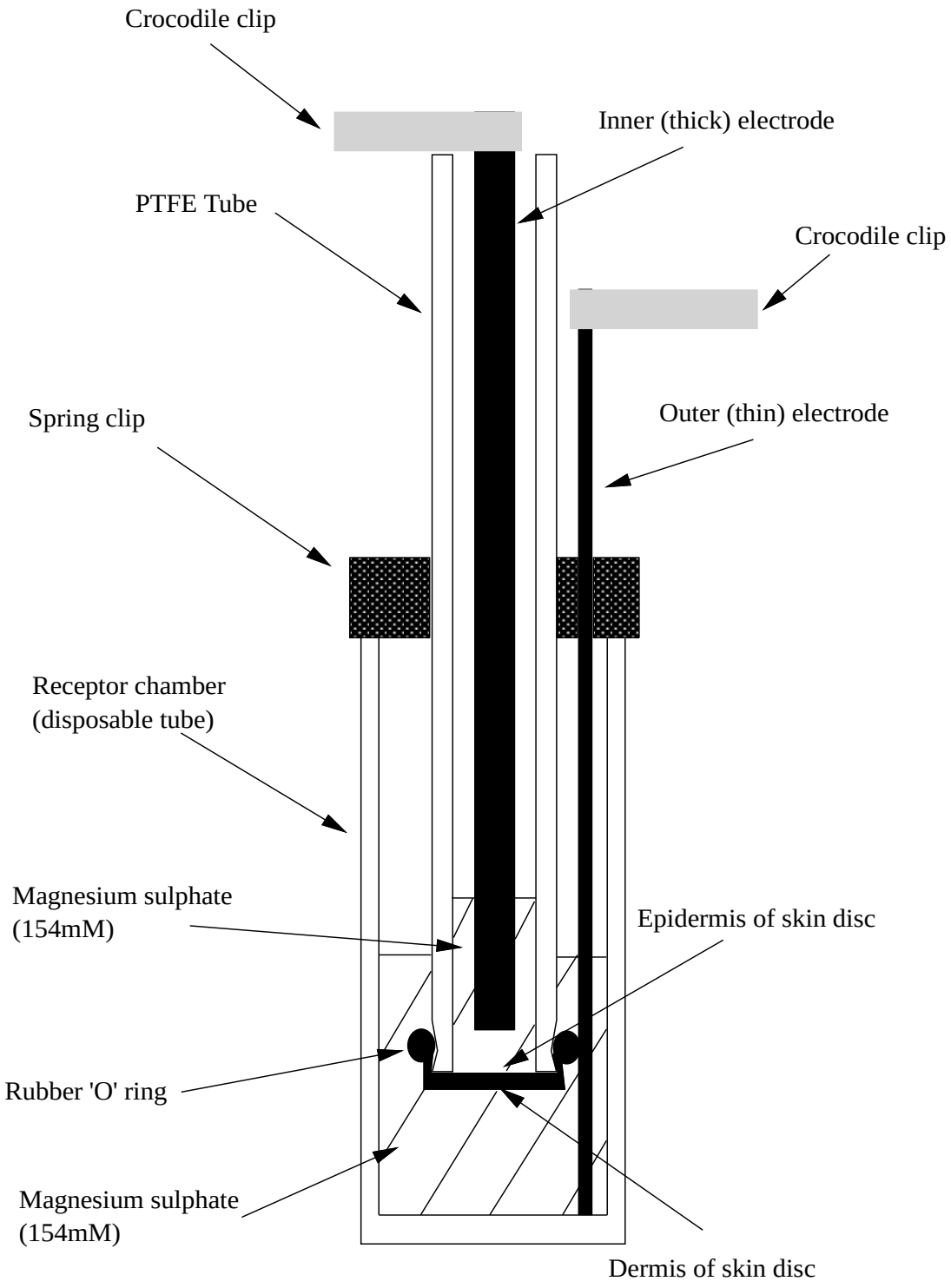
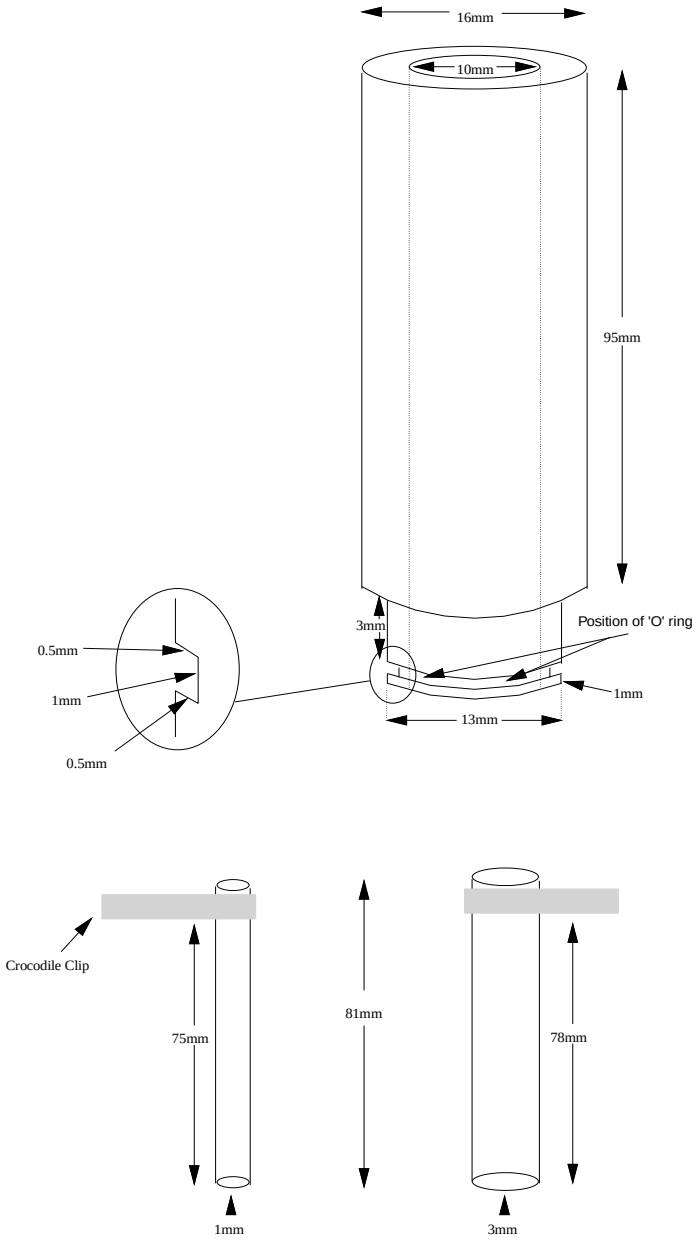


Figure 2: Dimensions of the polytetrafluoroethylene (PTFE) and receptor tubes and electrodes used in the VRM



Critical factors of the apparatus shown above:

- The inner diameter of the PTFE tube,
- The length of the electrodes relative to the PTFE tube and receptor tube, such that the skin disc should not be touched by the electrodes and that a standard length of electrode is in contact with the $MgSO_4$ solution,
- The amount of $MgSO_4$ solution in the receptor tube should give a depth of liquid, relative to the level in the PTFE tube, as shown in Figure 1,
- The skin disc should be fixed well enough to the PTFE tube, such that the electrical resistance is a true measure of the skin properties.

ANNEX 1

DEFINITIONS

Accuracy: The closeness of agreement between test method results and accepted reference values. It is a measure of test method performance and one aspect of relevance. The term is often used interchangeably with "concordance" to mean the proportion of correct outcomes of a test method (1).

Between-laboratory reproducibility: A measure of the extent to which different qualified laboratories, using the same protocol and testing the same substances, can produce qualitatively and quantitatively similar results. Between-laboratory reproducibility is determined during the prevalidation and validation processes, and indicates the extent to which a test can be successfully transferred between laboratories, also referred to as inter-laboratory reproducibility (1).

C: Corrosive.

Chemical: means a substance or a mixture.

Concordance: This is a measure of test method performance for test methods that give a categorical result, and is one aspect of "relevance". The term is sometimes used interchangeably with "accuracy", and is defined as the proportion of all chemicals tested that are correctly classified as positive or negative. Concordance is highly dependent on the prevalence of positives in the types of test chemical being examined (1).

GHS (Globally Harmonized System of Classification and Labelling of Chemicals (UN)): A system proposing the classification of chemicals (substances and mixtures) according to standardized types and levels of physical, health and environmental hazards, and addressing corresponding communication elements, such as pictograms, signal words, hazard statements, precautionary statements and safety data sheets, so that to convey information on their adverse effects with a view to protect people (including employers, workers, transporters, consumers and emergency responders) and the environment (4).

Me-too test: A colloquial expression for a test method that is structurally and functionally similar to a validated and accepted reference test method. Such a test method would be a candidate for catch-up validation (1). The term is interchangeably used with similar test method.

Mixture: means as a mixture or a solution composed of two or more substances in which they do not react (4).

NC: Non corrosive.

OD: Optical Density.

PC: Positive Control, a replicate containing all components of a test system and treated with a substance known to induce a positive response. To ensure that variability in the positive control response across time can be assessed, the magnitude of the positive response should not be excessive.

Performance standards (PS): Standards, based on a validated test method, that provide a basis for evaluating the comparability of a proposed test method that is mechanistically and functionally similar. Included are (i) essential test method components; (ii) a minimum list of Reference Chemicals selected from among the chemicals used to demonstrate the acceptable performance of the validated test method;

and (iii) the comparable levels of accuracy and reliability, based on what was obtained for the validated test method, that the proposed test method should demonstrate when evaluated using the minimum list of Reference Chemicals (1).

Prediction Model: a formula or algorithm (*e.g.,* formula, rule or set of rules) used to convert the results generated by a test method into a prediction of the (toxic) effect of interest. Also referred to as decision criteria. A prediction model contains four elements: (i) a definition of the specific purpose(s) for which the test method is to be used; (ii) specifications of all possible results that may be obtained, (iii) an algorithm that converts each study result into a prediction of the (toxic) effect of interest, and (iv) specifications as to the accuracy of the prediction model (*e.g.,* sensitivity, specificity, and false positive and false negative rates). Prediction models are generally not used in *in vivo* ecotoxicological tests (1).

Predictive Capacity: The predictive capacity reflects the test method performance in terms of correct and incorrect predictions in comparison to reference data. It gives quantitative information (e.g. correct prediction rate) on the relevance of the test method. It comprises, amongst others, the sensitivity and specificity of the test method.

Qualified run (experiment): A run that meets the acceptance criteria for the negative and positive controls, as defined in the corresponding SOP. Otherwise, the run is considered as non-qualified.

Reference Chemicals: Chemicals selected for use in the validation process, for which responses in the *in vitro* or *in vivo* reference test system or the species of interest are already known. These chemicals should be representative of the classes of chemicals for which the test method is expected to be used, and should represent the full range of responses that may be expected from the chemicals for which it may be used, from strong, to weak, to negative. Different sets of reference chemicals may be required for the different stages of the validation process, and for different test methods and test uses (1).

Relevance: Description of relationship of the test method to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test method correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test method (1).

Reliability: Measures of the extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol (1). It is assessed by calculating Within- and Between-Laboratory Reproducibility.

Reproducibility: The agreement among results obtained from testing the same substance using the same test protocol (1).

Sensitivity: The proportion of all positive/active chemicals that are correctly classified by the test method. It is a measure of accuracy for a test method that produces categorical results, and is an important consideration in assessing the relevance of a test method (1).

Skin corrosion *in vivo*: The production of irreversible damage of the skin; namely, visible necrosis through the epidermis and into the dermis, following the application of a test chemical for up to four hours. Corrosive reactions are typified by ulcers, bleeding, bloody scabs, and, by the end of observation at 14 days, by discoloration due to blanching of the skin, complete areas of alopecia, and scars. Histopathology should be considered to evaluate questionable lesions (5).

Specificity: The proportion of all negative/inactive chemicals that are correctly classified by the test method. It is a measure of accuracy for a test method that produces categorical results and is an important consideration in assessing the relevance of a test method (1).

Substance: means chemical elements and their compounds in the natural state or obtained by any production process, including any additive necessary to preserve the stability of the product and any impurities deriving from the process used, but excluding any solvent which may be separated without affecting the stability of the substance or changing its composition (4).

(Testing) run: A single test chemical concurrently tested in a minimum of three replicate skin discs.

Transcutaneous Electrical Resistance (TER): is a measure of the electrical impedance of the skin, as a resistance value in kilo Ohms. A simple and robust method of assessing barrier function by recording the passage of ions through the skin using a Wheatstone bridge apparatus.

Validated Reference Method(s) (VRM(s)): one (or more) test method(s) officially endorsed as scientific valid that was(were) used to develop the related official Test Guidelines and Performance Standards (PS). The VRM is considered the reference test method to compare new proposed similar or modified test methods in the framework of a PS-based validation study.

Within-laboratory reproducibility: determination of the extent that qualified people within the same laboratory can successfully replicate results using a specific protocol at different times, also referred to as intra-laboratory reproducibility (1).