HTML

NIH Publication No: 04-4510

Recommended Performance Standards for *In Vitro* Test Methods for Skin Corrosion

Prepared by the
Interagency Coordinating Committee on the
Validation of Alternative Methods (ICCVAM)
and the
National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)

National Institute of Environmental Health Sciences
National Institutes of Health
U.S. Public Health Service
Department of Health and Human Services

About ICCVAM and NICEATM

The National Institute of Environmental Health Sciences (NIEHS) of the National Institutes of Health (NIH) established the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) in 1997 to coordinate the interagency technical review of new, modified, and alternative test methods of interagency interest and to coordinate cross-agency issues relating to the validation, acceptance, and national and international harmonization of toxicological testing methods. ICCVAM was established as a permanent interagency committee of the NIEHS under the National Toxicology program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) on December 19, 2000, by the ICCVAM Authorization Act of 2000 (Public Law 106-545; Appendix E).

The Committee is comprised of representatives from the fifteen U.S. Federal regulatory and research agencies that use or generate toxicological information. ICCVAM promotes the scientific validation and regulatory acceptance of toxicological test methods that more accurately assess the safety or hazards of chemicals and products and that refine (i.e., decrease or eliminate pain and distress), reduce, and replace animal use. NICEATM administers the ICCVAM and provides operational and scientific support for ICCVAM and ICCVAM-related activities. NICEATM and ICCVAM work collaboratively to evaluate new and improved test methods applicable to the needs of U.S. Federal agencies. More information about ICCVAM and NICEATM can be found at *http://iccvam.niehs.nih.gov,* by contacting NICEATM at (919) 541-2384 or by email to *iccvam@niehs.nih.gov*.

The U.S. Federal regulatory and research agencies that participate in this effort are the:

- Consumer Product Safety Commission
- Department of Agriculture
- Department of Defense
- Department of Energy
- Department of Health and Human Services
  - Agency for Toxic Substances and Disease Registry
  - Food and Drug Administration
  - National Cancer Institute, NIH
  - National Institute for Occupational Safety and Health, CDC
  - National Institute of Environmental Health Sciences, NIH
  - National Institutes of Health, Office of the Director
  - National Library of Medicine, NIH
- Department of the Interior
- Department of Labor
  - Occupational Safety and Health Administration
- Department of Transportation
- Environmental Protection Agency

On the Cover

The ICCVAM/NICEATM graphic symbolizes the important role of new and alternative toxicological methods in protecting and advancing the health of people, animals, and our environment.

Recommended Performance Standards for *In Vitro* Test Methods for Skin Corrosion

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM)

National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)

National Toxicology Program
P.O. Box 12233
Research Triangle Park, NC 27709

May 2004
NIH Publication No. 04-4510

National Institute of Environmental Health Sciences
National Institutes of Health
U. S. Public Health Service
Department of Health and Human Services

This document is available electronically at:
http://iccvam.niehs.nih.gov/methods/ps/ps044510.pdf

TABLE OF CONTENTS

Appendices

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

°C	degrees centigrade
CASRN	Chemical Abstracts Service Registry Number
CDS	chemical detection system
Cm	centimeter
CV	coefficient of variation
DCIWG	Dermal Corrosivity and Irritation Working Group (ICCVAM)
DABT	Diplomate, American Board of Toxicology
DACLAM	Diplomate, American College of Laboratory Animal Medicine
DACVP	Diplomate, American College of Veterinary Pathology
DOT	U.S. Department of Transportation
ECVAM	European Centre for the Validation of Alternative Methods
EPA	U.S. Environmental Protection Agency
EPI-200	EpiDerm™ (MatTek Corp., Ashland, MA, USA.)
EU	European Union
FR	*Federal Register*
g	gram
GHS	Globally Harmonized Classification System
GLP	Good Laboratory Practice
Hz	hertz
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
kg	kilogram
kΩ	killiOhm
L	liter
M	molar
mg	milligrams
mL	milliliter
mm	millimeter
mM	millimolar
NICEATM	NTP Interagency Center for the Evaluation of Alternative Toxicological Methods
NIEHS	National Institute of Environmental Health Sciences
NIH	National Institutes of Health
nm	nanometer
NTP	National Toxicology Program
OD	optical density

LIST OF ACRONYMS AND ABBREVIATIONS (Cont.)

OECD	Organisation for Economic Co-operation and Development
PETA	People for the Ethical Treatment of Animals
pH	A measure of the negative logarithm of the H^+ ion concentration
P.L.	Public Law
PTFE	polytetrafluoroethylene
QC	quality control
R34	EU classification for chemicals that cause dermal corrosion following a 4-hour application, includes chemicals classified as UN Transportation Packing Group II or III
R35	EU classification for chemicals that cause dermal corrosion following a 3-minute application, analogous to UN Transportation Packing Group I
3Rs	Refinement, Reduction, and Replacement (of animal use)
SAR	Structure activity relationships
SDS	sodium dodecyl sulfate
SOP	Standard Operating Procedure
TER	transcutaneous electrical resistance
TG	Test Guideline
TM	Trademark
®	Registered Trademark
µg	microgram
µL	microliter
UN	United Nations
V	volt
v/v	volume to volume ratio
w/v	weight to volume ratio
ZEBET	Zentralstelle zur Erfassung und Bewertung von Ersatz- und Ergänzungsmethoden zum Tierversuch (German Centre for Documentation and Evaluation of Alternatives to Animal Experiments)

INTERAGENCY COORDINATING COMMITTEE ON THE VALIDATION OF ALTERNATIVE METHODS (ICCVAM)
Designated Agency Representatives

Agency for Toxic Substances and Disease Registry
William Cibulas, Ph.D.*
Moiz Mumtaz, Ph.D.**

Consumer Product Safety Commission
Marilyn L. Wind, Ph.D. (Vice-Chair)*
Susan Aitken, Ph.D.
Kailash C. Gupta, D.V.M., Ph.D.
Patricia Bittner, M.S.

Department of Agriculture
Jodie Kulpa-Eddy, D.V.M.*
Elizabeth Goldentyer, D.V.M.**

Department of Defense
Robert E. Foster, Ph.D.*
Patty Decot**
Harry Salem, Ph.D.
John M. Frazier, Ph.D.

Department of Energy
Marvin E. Frazier, Ph.D.*
Marvin Stodolsky, Ph.D.**

Department of the Interior
Barnett A. Rattner, Ph.D.*
Sarah Gerould, Ph.D.**

Department of Transportation
George Cushmac, Ph.D.*
Steve Hwang, Ph.D.**

Environmental Protection Agency
Office of Science Coordination and Policy
Joseph J. Merenda, Jr. *
Karen Hamernik, Ph.D.**
Office of Pollution Prevention and Toxics
Angela Auletta, Ph.D
Office of Pesticides Programs
Amy Rispin, Ph.D.
Deborah McCall
Office of Research and Development
Harold Zenick, Ph.D.**
Suzanne McMaster, Ph.D.
OECD Test Guidelines Program
Maurice Zeeman, Ph.D.

Food and Drug Administration
Leonard M. Schechtman, Ph.D. (Chair)*
Suzanne C. Fitzpatrick, Ph.D., D.A.B.T.**
Center for Drug Evaluation and Research
Abigail C. Jacobs, Ph.D.
Center for Devices and Radiological Health
Raju Kammula, D.V.M., Ph.D., D.A.B.T.
Melvin E. Stratmeyer, Ph.D.
Center for Biologics Evaluation and Research
Richard McFarland, Ph.D., M.D.
Center for Food Safety and Applied Nutrition
David G. Hattan, Ph.D.
Robert L. Bronaugh, Ph.D.
Center for Veterinary Medicine
Devaraya Jagannath, Ph.D.
M. Cecilia Aguila, D.V.M..
National Center for Toxicological Research
William T. Allaben, Ph.D.
Martha M. Moore, Ph.D.
Office of Regulatory Affairs
Atin R. Datta, Ph.D.

National Cancer Institute
Alan Poland, M.D.*
Marjorie Strobel, Ph.D.**

National Institute of Environmental Health Sciences
William S. Stokes, D.V.M., D.A.C.L.A.M.*
(Executive Director)
John R. Bucher, Ph.D., D.A.B.T.**
Rajendra S. Chhabra, Ph.D., D.A.B.T
Jerrold J. Heindel, Ph.D.

National Institute for Occupational Safety and Health
Paul Nicolaysen, V.M.D.*
Douglas Sharpnack, D.V.M., M.S., D.A.C.V.P.**

National Institutes of Health
Margaret D. Snyder, Ph.D.*
Nelson Garnett, D.V.M., D.A.C.L.A.M.**

National Library of Medicine
Vera Hudson, M.S.*
Jeanne Goshorn, M.S.**

Occupational Safety and Health Administration
Surender Ahir, Ph.D.*

---

* Principal Agency Representative
** Alternate Principal Agency Representative

ICCVAM DERMAL CORROSIVITY AND IRRITATION WORKING GROUP (DCIWG)
Members

Consumer Product Safety Commission
Marilyn Wind, Ph.D.
Susan Aitken, Ph.D.
Kailash Gupta, D.V.M., Ph.D.

Department of the Interior
Barnett Rattner, Ph.D.

Department of Transportation
George Cushmac, Ph.D.
Steve Hwang, Ph.D.

Environmental Protection Agency
Karen Hamernik, Ph.D. (Co-Chair)
Angela Auletta, Ph.D
Masih Hashim, Ph.D.
Leonard Keifer, Ph.D..
Marianne Lewis
Deborah McCall
Hari Mukhoty, Ph.D.
Mark Perry, Ph.D.
John Redden
Amy Rispin, Ph.D.

Food and Drug Administration
Abigail Jacobs, Ph.D. (Co-Chair)
Leonard Schechtman, Ph.D.
Jill Merrill, Ph.D.
Robert Bronaugh, Ph.D.
Donnie Lowther

National Institute of Environmental Health Sciences
William S. Stokes, D.V.M., D.A.C.L.A.M.
(ICCVAM Executive Director)
William Eastin, Ph.D.
Raymond Grissom, Jr., Ph.D.
Rajendra S. Chhabra, Ph.D., D.A.B.T

National Institute for Occupational Safety and Health
Douglas Sharpnack, D.V.M., M.S., D.A.C.V.P.

National Library of Medicine
Vera Hudson, M.S.

Occupational Safety and Health Administration
Surender Ahir, Ph.D.

NATIONAL TOXICOLOGY PROGRAM INTERAGENCY CENTER FOR THE EVALUATION OF ALTERNATIVE TOXICOLOGICAL METHODS (NICEATM)

Dave Allen, Ph.D.
Integrated Laboratory Systems, Inc.

Brad Blackard, M.S.P.H.
Integrated Laboratory Systems, Inc.

Sue Brenzel
Integrated Laboratory Systems, Inc.

Neepa Choksi, Ph.D.
Integrated Laboratory Systems, Inc.

Christina Inhof, M.S.P.H.
Integrated Laboratory Systems, Inc.

Linda Litchfield
Integrated Laboratory Systems, Inc.

Debbie McCarley
NIEHS

Raymond Tice, Ph.D.
Integrated Laboratory Systems, Inc.

William S. Stokes, D.V.M., D.A.C.L.A.M.
NIEHS

PREFACE

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) previously reviewed and recommended four *in vitro* test methods for assessing the dermal corrosivity potential of chemicals (ICCVAM 1999, 2002). Because three of these methods were proprietary, ICCVAM was asked by the U.S. Environmental Protection Agency (EPA) to develop and recommend performance standards that could be used to evaluate the acceptability of test methods that are based on similar scientific principles and that measure or predict the same biological or toxic effect. ICCVAM, in collaboration with the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), subsequently proposed and sought public comment on performance standards for these three types of test methods. Comments were also obtained on the draft standards from the ICCVAM Scientific Advisory Committee on Alternative Toxicological Methods (SACATM) and the EPA Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP). Following consideration of public and advisory committee comments, ICCVAM revised and approved recommended performance standards for the three types of *in vitro* corrosivity test methods. Those performance standards are based on validated and accepted proprietary (i.e., copyrighted, trademarked, registered) and nonproprietary *in vitro* test methods for assessing skin corrosivity that have been determined to have sufficient accuracy and reliability for specific testing purposes. The performance standards should assist other test developers in the validation of test methods that are similar in structure and function, and facilitate acceptance of test methods that adhere to the applicable performance standards.

This document is available online at *http://iccvam.niehs.nih.gov*; printed copies are available on request from the NICEATM (NIEHS, P.O. Box 12233, MD EC-17, Research Triangle Park, NC 27709; telephone: 919-541-3398, fax: 919-541-0947, e-mail: *iccvam@niehs.nih.gov*).

We gratefully acknowledge the ICCVAM agency representatives and members of the ICCVAM Dermal Corrosivity and Irritation Working Group, who contributed to the preparation of this document, and the NICEATM staff that assisted throughout the process. We also appreciate the constructive suggestions from interested stakeholders in response to a *Federal Register* notice and comments from the SACATM and the EPA FIFRA SAP.

Leonard M. Schechtman, Ph.D.
Chair, ICCVAM

William S. Stokes, D.V.M., Diplomate, A.C.L.A.M.
Director, NICEATM
Executive Director, ICCVAM

EXECUTIVE SUMMARY

The purpose of performance standards is to communicate the basis by which validated new proprietary (e.g., copyrighted, trademarked, registered) and nonproprietary test methods have been determined to have sufficient accuracy and reliability for specific testing purposes. Performance standards can then be used to evaluate the accuracy and reliability of other test methods that are based on similar scientific principles and that measure or predict the same biological or toxic effect. The three elements of performance standards are: 1) essential test method components (i.e., structural, functional, and procedural elements of a validated test method that a proposed, mechanistically and functionally similar test method should adhere to); 2) a minimum list of reference chemicals that is used to assess the accuracy and reliability of the proposed test method; and 3) the accuracy and reliability values that should be achieved by the proposed test method when evaluated using the minimum list of reference chemicals.

ICCVAM previously evaluated and recommended four validated test methods for assessing the dermal corrosivity hazard potential of chemicals: Corrositex®, EPISKIN™, EpiDerm™ (EPI-200), and the rat skin transcutaneous electrical resistance (TER) Assay. Subsequently, the EPA requested that ICCVAM establish performance standards for the three proprietary dermal corrosivity test methods (Corrositex®, EPISKIN™, EpiDerm™ [EPI-200]) and the non-proprietary rat skin TER test method. In response, the ICCVAM Dermal Corrosivity and Irritation Working Group (DCIWG) developed proposed performance standards based on these validated *in vitro* test methods. In a *Federal Register* Notice published on July 1, 2003**,** NICEATM invited public comment on the proposed performance standards for the three types of validated *in vitro* test methods for assessing dermal corrosivity hazard potential of chemicals. Comments on the draft document were also obtained during a public meeting of the NICEATM/ICCVAM Scientific Advisory Committee on Alternative Toxicological Methods (SACATM) in August 2003, and the EPA Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) in October 2003. All comments were considered by the DCIWG and ICCVAM during development of this final document.

This document describes the performance standards that should be met by *in vitro* corrosivity test methods that utilize membrane barrier test systems, cultured human skin model systems, or the rat skin TER test method. These three types of *in vitro* corrosivity test methods have been recommended by ICCVAM as screening assays to identify corrosive substances based on data from the respective validated reference test method. The extent to which proposed test methods that are mechanistically and functionally similar to the validated reference test methods must demonstrate comparable performance should be considered on a case-by-case basis. While it would be desirable for such test methods to have reliability and accuracy values at least as good as that of the corresponding validated reference test method, some flexibility might be acceptable to the extent that it would not compromise the ultimate protection of human and animal health. For example, a test method with lower specificity will have a higher false positive rate, which may be undesirable because this results in erroneous classification into a more hazardous category, but does not result in lowered protection of human health. A test method that has lower sensitivity will result in higher false negative rates. However, because these test methods are recommended as screening tests, this will simply result in a greater number of positive corrosivity test results in the first animal tested for dermal irritancy. For future test methods proposed as replacements

for existing test methods, minimum acceptable false positive and false negative rates will likely be recommended by ICCVAM, based on what is necessary to provide for an equivalent or better protection of human and animal health or the environment.

In Vitro Membrane Barrier Test Systems for Skin Corrosion

Validation studies have been completed for an *in vitro* membrane barrier test system commercially available as Corrositex®. Based on its scientific validity, this test method has been recommended for use as part of a tiered testing strategy for assessing the dermal corrosion hazard potential of chemicals, whereby any substance that qualifies for testing can be evaluated. In addition, this test method may be used to make decisions on the corrosivity and noncorrosivity of specific classes of chemicals (e.g., organic and inorganic acids, acid derivatives[1], and bases) for certain transport testing circumstances. The basis of this test system is that it detects membrane damage caused by corrosive test substances. The test substance is first evaluated to determine if it is compatible with the test procedure. If compatible, the substance is evaluated for category of acid or base (strong or weak) to determine the appropriate time scale to use to classify the potential corrosivity of the test substance. Finally, a compatible substance is applied to the surface of the artificial membrane barrier. The time it takes for the test substance to penetrate through the membrane barrier to an underlying indicator solution determines the corrosivity classification of that test substance. Penetration of the barrier might be measured by a number of procedures, including a color change in a pH indicator dye or other properties of the solution below the barrier (e.g., electrical conductivity).

Investigators using *in vitro* membrane barrier test systems for skin corrosion must be able to demonstrate that the assay is valid for its intended use. This includes demonstrating that different preparations are consistent in barrier properties, capable of maintaining a barrier to noncorrosive substances, and able to categorize the corrosive properties of chemicals across the various subcategories of corrosivity described by the United Nations (UN) Packing Group classification system. A sample protocol for the validated reference test method is available at *http://iccvam.niehs.nih.gov*.

Essential test method components have been developed for *in vitro* membrane barrier test systems for corrosivity. These include the physical components of the test method (e.g., membrane barrier, categorization solutions, indicator solution); a test substance categorization system; processes for determining test substance compatibility and test substance categorization; procedures for assembly of the physical components of the test method; proocedures for application of a test substance; appropriate control substances (solvent controls, positive [corrosive] controls, negative [noncorrosive] controls, benchmark controls); procedures to measure membrane barrier penetration; interpretation of results; classification of test substances with regard to corrosivity potential; and elements of the test report. The test report provides the following information: test and control substances, justification of the test method and protocol used, test method integrity, criteria for an acceptable test, test conditions, results, description of other effects observed, discussion of the results, and conclusion.

[1] "Acid derivative" is a non-specific class designation and is broadly defined as an acid produced from a chemical substance either directly or by modification or partial substitution. This class includes anhydrides, haloacids, salts, and other types of chemicals.

ICCVAM recommends the use of a minimum list of 40 reference chemicals to evaluate the reliability and the accuracy of test methods similar to Corrositex®. The distribution of chemicals in this list by corrosivity and UN Packing Group classification are 12 noncorrosive chemicals and 28 corrosive chemicals (9 UN Packing Group I, 9 UN Packing Group II, 10 UN Packing Group III). When evaluated using this minimum list of recommended reference chemicals, the reliability and accuracy of the proposed *in vitro* membrane test method should be, at a minimum, comparable to that of the validated reference test method. ICCVAM also recommends that 12 of the reference chemicals (3 noncorrosives and 3 in each UN Packing Group classification) be used by laboratories to evaluate their proficiency in the appropriate use of Corrositex®.

In Vitro Human Skin Model Systems for Skin Corrosion

Pre-validation and validation studies have been completed for an *in vitro* human skin cell culture model system commercially available as EPISKIN™. Based on its scientific validity, this test method has been recommended for the testing of all classes of chemicals and for inclusion in tiered testing strategies as part of a tiered or weight-of-evidence evaluation. In addition to EPISKIN™, a related human skin cell culture model corrosivity test method marketed as EpiDerm™ (EPI-200) has been validated and recommended for the same use as EPISKIN™. Neither test method has been validated for categorizing the corrosive properties of chemicals across the three UN Packing Group subcategories of corrosivity.

The test material is applied topically to a three-dimensional human keratinocyte culture model, comprised of at least a reconstructed epidermis with a functional stratum corneum. Corrosive substances are identified by their ability to induce a decrease in cell viability below defined threshold levels at specified exposure periods. The principle of the human skin model assay is based on the premise that corrosive chemicals are able to penetrate the stratum corneum by diffusion or erosion, and are cytotoxic to the keratinocytes in the underlying layers. The use of test systems that include human-derived cells or tissue should be in accordance with applicable national and international laws, regulations, and policies.

Investigators using an *in vitro* human skin cell culture model system for skin corrosion must be able to demonstrate that the assay is valid for its intended use. This includes demonstrating that different preparations are consistent in barrier properties (i.e., capable of maintaining a barrier to noncorrosive substances, able to respond appropriately to weak and strong corrosive substances) and/or that any modification to the existing validated reference test method does not adversely affect its performance characteristics.

Essential test method components have been developed for *in vitro* human skin model test methods for skin corrosivity, based on Organisation for Economic Co-operation and Development (OECD) Test Guideline 431. The components are essentially the same as those described for Corrositex® with the additional inclusion of components related to *in vitro* human skin model systems. Human skin models can be obtained commercially (e.g., EPISKIN™, EpiDerm™ [EPI-200]) or they can be developed or constructed in the testing laboratory.

ICCVAM recommends the use of a minimum list of 24 reference chemicals (12 noncorrosives, 12 corrosives) to evaluate the reliability and accuracy of test methods similar to EPISKIN™.

When evaluated using this minimum list of recommended reference chemicals, the reliability and accuracy of the proposed *in vitro* membrane test method should be, at a minimum, comparable to that of the validated reference test method. ICCVAM also recommends that 12 of the reference chemicals (6 noncorrosives and 6 corrosives varying in corrosive potency) be used by laboratories to evaluate their proficiency in the appropriate use of Corrositex®.

In Vitro Skin Transcutaneous Electrical Resistance (TER) Tests for Skin Corrosion

Prevalidation and validation studies have been completed for the rat skin TER assay. Based on its scientific validity, this test method has been recommended for the testing of all classes of chemicals and for inclusion in tiered testing strategies as part of a tiered or weight-of-evidence evaluation.

The test substance is applied for up to 24 hours to the epidermal surface of skin discs in a two-compartment test system in which the skin discs function as the separation between the compartments. The skin discs are prepared from humanely killed 28 to 30 day-old rats. Corrosive substances are identified by their ability to produce a loss of normal stratum corneum integrity and barrier function, which is measured as a reduction in the TER below a specified level. For rat skin TER, a cutoff value of 5 kΩ has been selected based on extensive data for a wide range of substances where the majority of values were either clearly well above or well below this value. Generally, substances that are noncorrosive but irritating in animals do not reduce the TER below this cutoff value. However, the use of other skin preparations or other equipment to measure resistance may necessitate the use of a different cutoff value. In such situations, more extensive validation would be required. A dye-binding step is incorporated into the test procedure to confirm positive results. The dye-binding step determines if the increase in ionic permeability is due to physical destruction of the stratum corneum.

Investigators using an *in vitro* skin TER corrosivity test must be able to demonstrate that the assay is valid for its intended use. This includes demonstrating that different preparations are consistent in barrier properties (i.e., capable of maintaining a barrier to noncorrosive substances, able to respond appropriately to weak and strong corrosive substances) and/or that any modification to the existing validated reference test method does not adversely affect its performance characteristics.

Essential test method components have been developed for *in vitro* skin TER test methods for skin corrosivity, based on OECD Test Guideline 430. The components are essentially the same as those described for Corrositex® except that this test method includes essential components related to the use of animals, the physical measurement of transcutaneous electrical resistance, and dye binding procedures.

ICCVAM recommends that laboratories use a minimum list of 12 calibration chemicals (6 noncorrosives and 6 corrosives varying in corrosive potency) to evaluate their proficiency and to determine that the *in vitro* rat skin TER test method is performing as expected. A minimum list of 24 reference chemicals is recommended to determine if the reliability and accuracy of a new or modified *in vitro* skin TER test for skin corrosion is comparable to that of the validated reference test method. When a modified TER test method is evaluated using this minimum list of recommended reference chemicals, the reliability and accuracy of the proposed *in vitro* test method should be, at a minimum, comparable to that of the validated reference test method.

This Page Intentionally Left Blank

1.0 PURPOSE AND BACKGROUND OF PERFORMANCE STANDARDS

1.1 Introduction

Prior to the acceptance of a new test method for regulatory testing applications, validation studies are conducted to assess its reliability (i.e., the extent of intra- and inter-laboratory reproducibility) and its accuracy (i.e., the ability of the test method to correctly predict or measure the biological effect of interest) (ICCVAM 1997, 1999, 2002; OECD 1996, 2002a). The purpose of performance standards is to communicate the basis by which new proprietary (i.e., copyrighted, trademarked, registered) and nonproprietary test methods have been determined to have sufficient accuracy and reliability for specific testing purposes. These performance standards, based on test methods accepted by regulatory agencies, can be used to evaluate the reliability and accuracy of other test methods that are based on similar scientific principles and measure or predict the same biological or toxic effect. EpiDerm™, a human skin model system for skin corrosion, is an example of a test method that underwent an expedited validation process because it was mechanistically and functionally similar to a previously validated human skin model system for skin corrosion called EPISKIN™ (see Section 3.0). Another example of their application would be to evaluate the use of mouse or human skin rather than rat skin in the rat skin transcutaneous electrical resistance (TER) assay. This section describes the three elements of performance standards identified by ICCVAM (ICCVAM 2003), the ICCVAM process for developing performance standards during test method evaluations, and the ICCVAM process to retrospectively develop performance standards for previously reviewed test methods. A retrospecitive process was used to develop performance standards for three types of validated *in vitro* corrosivity test methods (a noncellular membrane barrier test system, cultured human skin model systems, the rat skin TER test method), and those perfomance standards are provided in **Sections 2.0**, **3.0**, and **4.0**.

1.2 Elements of ICCVAM Performance Standards

The three elements of performance standards are:

- **Essential test method components:** These consist of essential structural, functional, and procedural elements of a validated test method that should be included in the protocol of a proposed mechanistically and functionally similar test method. Essential test method 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 proposed test method is structurally and functionally similar to the corresponding validated test method.

- **A minimum list of reference chemicals:** Reference chemicals are used to assess the accuracy and reliability of a proposed mechanistically and functionally similar test method. These chemicals are a representative subset of those used to demonstrate the reliability and the accuracy of the validated test method. To the extent possible, this subset of chemicals should:
  - be representative of the range of responses that the validated test method is capable of measuring or predicting
  - have produced consistent results in the validated test method and the *in vivo* reference test method and/or the species of interest

- have well-defined chemical structures
- be readily available
- not be associated with excessive hazard or prohibitive disposal costs
- have performance characteristics (e..g., accuracy, sensitivity, specificity, false negative and false positive rates) in the validated test method that approximate the performance values obtained for all appropriate substances during the validation process

These reference chemicals are the minimum number that should be used to evaluate the performance of a proposed mechanistically and functionally similar test method. Reference chemicals should not be used to develop the prediction model for the proposed test method. If any of the recommended reference chemicals are not available, other chemicals for which adequate reference data are available could be substituted. To the extent possible, the substituted chemical(s) should be of the same chemical class as the original reference chemical(s). If desired, additional chemicals representing other chemical or product classes and for which adequate reference data are available can be used to more comprehensively evaluate the accuracy of the proposed test method. However, none of these additional chemicals should have been used to develop the proposed test method.

- **Accuracy and reliability values:** These are the accuracy and reliable characteristics that the proposed test method should be comparable to when evaluated using the minimum list of reference chemicals.

1.3 ICCVAM Process for the Development of Performance Standards

The process followed by ICCVAM for developing performance standards for new test methods is as follows (ICCVAM, 2003):

- NICEATM and the appropriate ICCVAM working group develop proposed performance standards for consideration during the ICCVAM evaluation process. If performance standards are proposed by a test method sponsor, they will be considered by ICCVAM at this stage. Generally, the proposed performance standards are based on the information and data provided in the test method submission or other available applicable data.
- The ICCVAM/NICEATM Peer Review Panel evaluates the proposed performance standards for completeness and appropriateness during its evaluation of the validation status of the proposed test method. The proposed performance standards, as well as the test method submission, are made available to the public for comment prior to and during the Peer Review Panel meeting.
- The appropriate ICCVAM working group, with the assistance of NICEATM, prepares the final performance standards for ICCVAM approval, taking into consideration the recommendations of the Peer Review Panel and public comments.

Performance standards recommended by ICCVAM are incorporated into ICCVAM test method evaluation reports, which are then provided to U.S. Federal agencies and made available to the public. Regulatory authorities can then reference the performance standards in the ICCVAM report when they communicate their acceptance of a new test method. In addition, performance standards adopted by U.S. Federal regulatory authorities can be provided in guidelines issued for new test

methods. Availability of ICCVAM test method evaluation reports are announced routinely in the *Federal Register*, NTP Newsletters, and by e-mail to ICCVAM/NICEATM listserve groups.

1.4 ICCVAM Development of Recommended Performance Standards for *In Vitro* Test Methods for Skin Corrosion

Skin corrosion refers to the visible destruction or irreversible alteration of skin following exposure of the skin to a chemical substance. Skin corrosivity has traditionally been assessed by applying the test substance to the skin of living animals and evaluating the extent of tissue damage after a fixed period of time (OECD 2002b; EPA 1998). Some U.S. regulatory authorities require determination of corrosivity using three categories of responses, as provided in **Table 1-1** (EPA 1998; DOT 2003a, 2003b).

Table 1-1 Skin Corrosive Category and Subcategories

Corrosive Category (category 1) (applies to authorities not using subcategories)	Potential Corrosive Subclasses[1] (UN Packing Group Classification[2])	Corrosive in at least 1 of 3 animals	
		Exposure	Observation
Corrosive	Corrosive subcategory 1A (I)	≤3 minutes	≤1 hour
	Corrosive subcategory 1B (II)	>3 minutes / ≤1hour	≤14 days
	Corrosive subcategory 1C (III)	>1 hour / ≤4 hours	≤14 days

[1] Classifications designated by the United Nations (UN) Globally Harmonised System for the Classification and Labelling of Chemical Substances and Mixtures (GHS) (UN 2003a).

[2] Corresponding UN packing group classifications to be used for the transport of dangerous goods (UN 2003b).

The EPA test guideline (EPA 1996), a DOT exemption (DOT 2002), and a globally-harmonized tiered testing strategy (UN 2003a) for the assessment of skin corrosivity allow for the use of validated and accepted *in vitro* methods. In both the EPA guideline and the tiered testing strategy, positive results from *in vitro* test methods can be used to classify a substance as corrosive without the need for animal testing. Substances that are negative *in vitro* should undergo additional testing in accordance with the tiered testing strategy. The DOT exemption allows for the determination of corrosivity and noncorrosivity of specific classes of chemicals for certain transport testing circumstances. The use of *in vitro* methods to identify corrosive substances can therefore avoid pain and distress that may result from the application of corrosive substances to animals.

A number of *in vitro* test methods have been proposed as alternatives for the standard *in vivo* rabbit skin procedure to identify corrosive substances. Generally, these test methods have involved the use of a noncellular membrane barrier test system, cultured human skin model systems, or isolated rat skin (Fentem et al. 1998). ICCVAM previously evaluated and recommended four validated test methods for assessing the dermal corrosivity hazard potential of substances:

Corrositex®, EPISKIN™, EpiDerm™ (EPI-200), and the rat skin TER Assay (ICCVAM 1999, 2002). Subsequently, the EPA requested that ICCVAM establish performance standards for the three proprietary *in vitro* dermal corrosivity test methods (Corrositex®, EPISKIN™, EpiDerm™ [EPI-200]) and the non-proprietary rat skin TER test method. In response, the ICCVAM Dermal Corrosivity and Irritation Working Group (DCIWG) drafted proposed performance standards based on these validated *in vitro* test methods. As described earlier in this section, for future test methods evaluated by ICCVAM, performance standards will be included as part of the test method recommendations forwarded to U.S. Federal regulatory authorities.

In a *Federal Register* notice published on July 1, 2003**,** NICEATM announced the availability of and invited public comment on the proposed performance standards for the three types of validated *in vitro* test methods for assessing the dermal corrosivity hazard potential of chemicals (**Appendix A**). Public comments were received from individuals representing five organizations (**Appendix B**). Comments on the draft document were also obtained during public meetings of the ICCVAM Scientific Advisory Committee on Alternative Toxicological Methods (SACATM) in August, 2003, and the EPA Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) in October, 2003 (Appendix C). These comments, which are discussed briefly here, were considered by the DCIWG and ICCVAM during development of these recommended performance standards.

One commenter (Dr. Roland Roguet, L'ORÉAL Research) agreed with the importance of establishing test method performance standards and provided several suggestions for revising the performance standards for *in vitro* human skin model systems. In addition to minor editorial changes, the suggestions included noting that EPISKIN™ is currently commercially available for skin corrosivity assessments and that the test method can discriminate between United Nations (UN) transportation Packing Group I and Packing Group II/III substances. ICCVAM incorporated the suggested editorial changes and revised the text to note that EPISKIN™ is commercially available. However, the performance standards were not revised to reflect the ability of EPISKIN™ to discriminate between UN Packing Group I and Packing Group II/III substances. The ability of a test method to classify corrosive substances by UN Packing Group is applicable to the U.S. Department of Transportation (DOT) and not to other Federal agencies, and current U.S. DOT transportation regulations require that test method be able to classify the corrosive potential of substances across all three UN subcategories of corrosivity.

Two commenters (Drs. John Harbell and Rodger Curren, Institute for In Vitro Sciences, Inc.) commended the efforts of NICEATM and the DCIWG for drafting performance standards for the three types of *in vitro* skin corrosivity test methods and stated that these standards represent a substantial step forward in regulatory toxicology. Suggestions for improving the proposed performance standards included adding to the test report specifications the phrase "if relevant to the conduct of the study" where appropriate to make them less prescriptive and including specifications for test acceptance criteria (e.g., acceptable range of positive responses). The performance standards for all three test method systems were revised to include these suggested additions.

Another commenter (Mr. Troy Seidle, People for the Ethical Treatment of Animals [PETA]) stated that PETA appreciated the effort involved in the development of these performance standards and

that they were "hopeful that they would not only satisfy the needs of U.S. regulatory agencies, given their inability to lawfully require or recommend the use of proprietary test methods, but will also be useful in preventing future bottlenecks in the validation pipeline both domestically and internationally." The commenter indicated that PETA was in general agreement with the content of these documents. Additional comments were provided that did not pertain to the proposed performance standards but rather were directed at the manner in which *in vitro* test methods results are used in making decisions about the corrosivity/non-corrosivity of a test substance. Specifically, the commenter disagreed with ICCVAM's recommendation that *in vitro* human skin model systems (i.e., EpiDerm™ and EPISKIN™) (ICCVAM 2002) could be used as screening methods where substances inducing a positive result would be classified as corrosive, while substances inducing a negative result should undergo additional testing in accordance with the globally-harmonized tiered testing strategy (UN 2001). Rather, the commenter recommended that these *in vitro* tests should be considered as full replacements for the *in vivo* rabbit skin test method. ICCVAM's recommendation was based on the high false negative rates for these test methods (17% for EPISKIN™, 13% for EpiDerm™ [EPI-200], and 12% for TER) for identifying corrosive substances, and the irreversible permanent damage that could result from exposure to corrosive substances that were not properly classified and labelled as corrosive hazards. The commenter also noted that Worth et al. (1998) reported that the frequency of false negatives in the human skin model test systems, when combined with pH measurements and computerized structure-activity relationship (SAR) modeling in a sequential testing approach, could be reduced to zero. In response, ICCVAM noted that Worth et al. (1998) acknowledged that the SAR models used in this evaluation had not yet been validated. In addition, while eliminating test substances from further consideration based on pH alone appears useful, NICEATM could not locate a published standardized protocol for preparing solutions (in terms of the amount of the substance dissolved/mixed with water) for determining the pH. Furthermore, the approach also results in a relatively high percentage of false positives. Lacking a standardized protocol and validation of this approach, ICCVAM concluded that it is premature to formally evaluate this sequential testing strategy.

ICCVAM recognizes that, to date, a careful evaluation of the reliability and accuracy of the current *in vivo* rabbit skin corrosivity test has not yet been conducted. To correct this deficiency, NICEATM is reviewing *in vivo* rabbit skin corrosivity data that has been generated using current test method procedures. These data are being extracted from the published literature, and from data submitted by U.S. Federal regulatory agencies and commercial organizations to ICCVAM. Once these data have been tabulated and the appropriately analyzed, scientifically sound estimates of the reliability and under-prediction rates of the *in vivo* rabbit skin corrosivity test should be available which can be used to compare the performance of the *in vitro* corrosivity test methods.

A fourth commenter (Dr. Manfred Liebsch, Zentralstelle zur Erfassung und Bewertung von Ersatz- und Ergänzungsmethoden zum Tierversuch [German Centre for Documentation and Evaluation of Alternatives to Animal Experiments]; ZEBET) commented that ZEBET very much welcomed the general concept and the definition of performance standards for the future development of test systems that claim to be scientifically equivalent to existing validated systems. However, he recommended that ICCVAM adopt only the 12 chemicals provided in the proposed OECD test guidelines for the TER and *in vitro* human skin model systems as reference chemicals, rather

than the proposed 24 chemicals. He proposed that the other 12 reference chemicals could be recommended for test refinement if the first set of 12 reference chemicals were not classified 100% correctly.

ICCVAM considered these comments and decided to retain the current list of 24 chemicals based in part on the following reasons:

- In terms of the *in vitro* human skin model systems, ICCVAM agrees with the European Centre for the Validation of Alternative Methods (ECVAM) that 24 reference chemicals are needed, at a minimum, to adequately evaluate the reliability and accuracy of a test method that is mechanistically and functionally similar to EPISKIN™ (Liebsch et al. 2000). In addition, including substances that tested false positive or false negative in the validated reference test method in this minimum list allows the developer of a proposed test method to potentially demonstrate that their method provides a greater level of accuracy than the validated test method. However, ICCVAM concluded that 12 chemicals (6 noncorrosive and 6 corrosive) could be used by a naïve laboratory to demonstrate proficiency with the validated reference test methods (i.e. EPISKIN™, EpiDerm™ [EPI-200]) and should therefore be referred to as proficiency chemicals.
- In terms of the rat skin TER assay, the list of 24 reference chemicals was recommended for use only when a substantial protocol change was incorporated into the test method (e.g., skin from an animal of a different age, strain, and/or species than that the validated reference test method. In any case, 12 calibration chemicals would still be used to calibrate a rat skin TER assay in the hands of new investigator.
- ICCVAM did not include acrylic acid (Chemical Abstract Services Registry Number [CASRN] 79-10-7), one of the twelve reference chemicals recommend in the OECD test guidelines, in its list of reference chemicals for these two test methods because no data are available for this chemical in EPISKIN™ and the chemical was not used in the validation of EpiDerm™ (EPI-200).

The proposed ICCVAM performance standards for *in vitro* corrosivity test methods were discussed at the August 12-13, 2003 meeting of the SACATM (information about this meeting can be found at *http://iccvam.niehs.nih.gov/about/sacatm.htm*). The rationale and process for the development of the performance standards for these validated test methods was discussed, as was the recommended essential test method components, the minimum list of reference chemicals, and the specified levels of accuracy and reliability that should be achieved by a mechanistically and functionally similar test method. There was public comment on the proposed performance standards prior to the general discussion by the SACATM. The single public commenter at the meeting expressed opposition to the ICCVAM recommended application of these test methods as screening assays in a tiered testing strategy. NICEATM responded by reiterating that the recommended application was the consensus decision of an independent Peer Review Panel, the ICCVAM Corrosivity Working Group, and ICCVAM, and was based on the relatively high false negative rates of the *in vitro* test methods.

During the general discussion, the SACATM discussed the scientific validity of employing the proposed performance standards in the validation of new test methods that were mechanistically and functionally similar to the validated reference test method. The SACATM generally agreed that the approach should expedite the development and implementation of alternative test methods,

but that new (and unrelated) test methods would require more extensive validation, as outlined in the ICCVAM guidelines for test method validation. In addition, the SACATM emphasized that all test methods should be validated against the original animal-based reference test method and not the validated *in vitro* reference test method.

On October 28 and 29, 2003, an EPA Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) Meeting was held to review scientific issues being considered by the EPA regarding data quality for *in vitro* tests used as alternatives to animal studies for regulatory purposes (information about this meeting can be obtained at *http://www.epa.gov/scipoly/sap/*). During this meeting, the SAP reviewed the draft Performance Standards developed by ICCVAM for three types of *in vitro* corrosivity test systems. Based on its evaluation, the SAP:

- endorsed the Performance Standards approach to identify and validate *in vitro* test methods that are structurally and functionally similar to a validated *in vitro* reference test method
- concurred that the Performance Standards prepared by ICCVAM were very well described for each of the three tests
- concluded that the information generated using Performance Standards should provide a basis to determine whether a test is mechanistically and functionally similar to a validated *in vitro* test method

The SAP expressed the view that the strength of the Performance Standards approach to validating a new test method or those structurally and functionally similar to a validated reference test method derives from the stated selection criteria for the Reference Chemical set. The Panel also concluded that the approach of specifying a known level of accuracy and reliability for new test methods to meet in order to be considered equivalent to the validated reference test system was acceptable.

In terms of the Performance Standards, the EPA asked the SAP to comment on:

- the extent to which the essential test method components for each method adequately described the unique characteristics of the method necessary to determine whether a proposed test is mechanistically and functionally similar
- the strengths or weaknesses of this approach and any modifications to the criteria that should be considered
- whether test methods that are mechanistically and functionally similar to a validated reference test method should be demonstrated to be accurate for all of the chemicals in the Performance Standards

In responding to these questions, the SAP made a number of recommendations, several of which have been incorporated into the ICCVAM Performance Standards document. These include:

- provide examples of test methods that would be considered mechanistically and functionally similar to a validated reference test method
- expand the discussion on the use of benchmark controls
- specify the minimum replicate requirements for positive, negative, and benchmark controls
- provide general criteria for acceptance of concurrent positive controls in relation to historical positive control data

- expand the accuracy and reliability sections to include information on an effective approach for establishing intra- and inter-laboratory reproducibility
- clarify the concept of comparable performance characteristics for these three types of *in vitro* corrosivity screening assays

Several of the SAP recommendations were considered but could not be practically implemented within the scope of this document. One recommendation was that a single standard list of reference chemicals should be developed for validating all *in vitro* corrosivity test methods, while another was that a chemical repository should be established for reference samples/positive controls to be used by laboratories for developing/conducting *in vitro* skin studies. ICCVAM appreciates the value of a single standard list of reference chemicals for the validation of any test method proposed as an alternative to the traditional *in vivo* rabbit skin corrosivity test method. A similar approach was used by ICCVAM in its development of a proposed list of reference chemicals for the future validation of *in vitro* estrogen and androgen receptor binding and transcriptional activation assays (*http://iccvam.niehs.nih.gov/methods/endodocs/edfinrpt/edfinrpt.pdf*). Due to the sequence of events that led to the development of the Performance Standards for these three types of *in vitro* corrosivity test systems, this approach was not used. However, in the future, ICCVAM plans to routinely develop reference chemical sets for new test methods, which should eliminate the need for different reference chemical sets for test methods that measure or predict the same test endpoint. In response to the recommendation that ICCVAM establish a chemical repository, ICCVAM concluded that there was no need at the present time, considering that the recommended reference chemicals are commercially available.

The SAP recommended that laboratories should be allowed to determine their own positive control(s) and that the Performance Standards should not suggest specific examples. ICCVAM agreed, but considered that the inclusion of example positive controls for each test system would be useful to developers of new test methods. However, the appropriate sections in the Performance Standards document were revised to clarify that the indicated positive control chemical(s) is only provided as an example, .

The SAP raised concerns about the chemical classes represented in the list of reference chemicals, noting that some classes of potentially corrosive chemicals (e.g., hydrocarbons and halogenated hydrocarbons) were not represented. In addition, the Panel recommended including substances in the list of reference chemicals that would challenge the technical skill of laboratory staff. ICCVAM was unable to identify substances in the chemical classes of concern to the SAP that met the criteria established for inclusion as reference chemicals (see Section 1.2). Furthermore, although ICCVAM agrees that a wide range of chemicals, including those that are difficult to work with, would be ideal for determining limitations of a test method, the reference chemical list is limited to substances that meet the criteria described in Section 1.2. However, each reference chemical list includes substances that produced false negative and false positive responses in the validated reference test method (i.e., substances potentially difficult to work with).

The SAP questioned whether the numbers of chemicals in each list were sufficient for adequately demonstrating the reliability and accuracy of a proposed test method, even if it was mechanistically and functionally similar to a validated reference test method. ICCVAM agrees that this issue needs to be addressed further, especially for alternative test methods that are proposed as total

replacements for a traditional *in vivo* test method. However, since these *in vitro* dermal corrosivity test methods are recommended as screening tests, the minimum set of reference chemicals is considered appropriate for this purpose. Nevertheless, as more data regarding the limitations of these test methods for certain classes of chemicals or products are generated, or if improved methods are proposed as replacements, it may be desirable to revise the minimum reference list to ensure that these classes are better represented.

With regard to Quality Control issues, the SAP considered:

- the utility of and necessity for training or calibration sets in assuring data quality
- aspects of the quality control criteria that are necessary for assuring the integrity of such systems over time and from lot-to-lot
- the advantages and disadvantages of including concurrent positive and negative controls with *in vitro* assays when used as alternatives to animal testing
- whether benchmark controls serve a useful purpose to demonstrate the level of response that can be expected for each chemical class for each lot of proprietary test method assays

The SAP noted that individual test facilities may detect failures or out-of-specification performance of a proprietary test method and proceed according to their operating procedures, but the lack of a Good Manufacturing Process-like regulatory authority does not require these failures to be reported to and addressed by the vendor. Other facilities may unknowingly use an inadequate/underperforming proprietary test method without benefit of the experiences of the first facility. Thus, the SAP recommended that proprietary test method quality control reports should be compiled by the vendor and reported to purchasers of that test method. This issue was not specifically addressed in the Performance Standards, but will likely be addressed in an Organisation for Economic Co-operation and Development (OECD) Advisory Document on the Application of Good Laboratory Practices to *In Vitro* Studies that is currently under development.

The SAP commented on the essential role that concurrent positive control(s) have in ensuring the adequacy of *in vitro* studies, and on the appropriate role and properties that benchmark controls might have in such studies. The Performance Standards now provide greater emphasis on these topics.

Following issuance of the proposed performance standards for public comment, the DCIWG and ICCVAM revised some of the terminology in order to eliminate potential confusion. Specifically, “minimum performance standards” was revised to “performance standards” and “minimum procedural standards” was revised to “essential test method components”.

The following Sections describe the performance standards that should be met for three types of *in vitro* corrosivity test methods proposed for testing the skin corrosion hazard potential of chemicals (membrane barrier test systems, human skin model systems, and the rat skin TER test method). Validated versions of these three types of *in vitro* corrosivity test methods have been recommended by ICCVAM as screening assays for the detection of corrosive substances. ICCVAM recommends that proposed test methods that are mechanistically and functionally similar to the validated reference test methods must demonstrate comparable performance using the minimum list of reference substances included in these performance standards, and that decisions on comparable performance should be handled on a case-by-case basis. While it would be desirable for such

test methods to have reliability and accuracy values at least as good as that of the corresponding validated reference test method, some flexibility might be acceptable to the extent that it would not compromise the ultimate protection of human and animal health. For example, slightly higher false positive rates, while undesirable because they result in erroneous classification in a more hazardous category, do not result in lowered protection of human health. Because these test methods are used as *screening* tests, negative results will be followed with testing at least one animal as part of a dermal irritancy assessment, where false negatives should be detected by the presence of a corrosive skin lesion on the treated animal. Thus, a test method with a higher false negative rate will simply result in more positive corrosivity test results in the first animal tested for dermal irritancy. However, for future test methods proposed as *replacements* for existing test methods, minimum acceptable false positive and false negative rates will likely be recommended by ICCVAM, based on what is necessary to provide for equivalent or better protection of human or animal health or the environment.

2.0 *IN VITRO* MEMBRANE BARRIER TEST SYSTEMS FOR SKIN CORROSION

2.1 Background

Validation studies have been completed for an *in vitro* membrane barrier test system commercially available as Corrositex® (ICCVAM 1999; Fentem et al. 1998; Barratt et al. 1998; Gordon et al. 1994; InVitro Intl. 1995). Based on its scientific validity, this test method has been recommended for use as part of a tiered testing strategy for assessing the dermal corrosion hazard potential of chemicals, whereby any substance that qualifies for testing can be evaluated (ICCVAM 1999; ECVAM 2001). The use of an *in vitro* membrane barrier test method as part of a tiered approach reduces and refines the use of animals in testing and provides a basis for deciding on the adequacy of information for hazard classification or the need for further testing. In addition, such a test method may be used to make decisions on the corrosivity and noncorrosivity of specific classes of chemicals (e.g., organic and inorganic acids, acid derivatives[1], and bases) for certain transport testing circumstances (DOT 2002). This chapter briefly describes the principles of *in vitro* membrane barrier test systems for corrosivity followed by the recommended performance standards, which consists of essential test method components, reference chemicals, and comparison of accuracy and reliability.

2.2 Principles of *In Vitro* Membrane Barrier Test Systems for Skin Corrosion

The basis of this test system is that it detects membrane damage caused by corrosive test substances (ICCVAM 1999). The test substance is first evaluated to determine if it is compatible with the test procedure. If compatible, the substance is evaluated for category of acid or base (strong or weak) to determine the appropriate time scale used to classify the potential corrosivity of the test substance. Finally, a compatible substance is applied to the surface of the artificial membrane barrier. The time it takes for the test substance to penetrate through the membrane barrier to an underlying indicator solution determines the corrosivity classification of that test substance. Penetration of the barrier might be measured by a number of procedures, including a color change in a pH indicator dye or other properties of the solution below the barrier (e.g., electrical conductivity).

Investigators using *in vitro* membrane barrier test systems for skin corrosion must be able to demonstrate that the assay is valid for its intended use. This includes demonstrating that different preparations are consistent in barrier properties, capable of maintaining a barrier to noncorrosive substances, and able to categorize the corrosive properties of chemicals across the various subcategories of corrosivity described by the UN Packing Group classification system. For *in vitro* membrane barrier test systems, the UN Packing Group classification assigned is based on the time it takes the test substance to penetrate through the membrane barrier. For Corrositex®, the validated *in vitro* reference test method, a color change in the underlying Chemical Detection System (CDS) indicates that the membrane barrier has been penetrated. The CDS changes color when a chemical or chemical mixture changes the pH of the solution to less than 4.5 or greater than 8.5.

[1] "Acid derivative" is a non-specific class designation and is broadly defined as an acid produced from a chemical substance either directly or by modification or partial substitution. This class includes anhydrides, haloacids, salts, and other types of chemicals.

In vitro membrane barrier test systems may be used to test solids, liquids, and emulsions. The liquids can be aqueous or nonaqueous; solids can be soluble or insoluble in water. The samples may be pure chemicals, dilutions, formulations, or waste. No prior treatment of the sample is required. A limitation of the validated *in vitro* membrane barrier test method is that many noncorrosive chemicals and chemical mixtures and some corrosive chemicals and chemical mixtures do not qualify for testing. Test chemicals and chemical mixtures are considered nonqualifying if they do not cause a color change in the CDS. Aqueous substances with a pH in the range of 4.5 to 8.5 often do not qualify for testing; however, 85% of chemicals tested in this pH range were noncorrosive in animal tests (ICCVAM 1999).

2.3 Essential Test Method Components

The following is a description of the essential test method components of *in vitro* membrane barrier test systems for corrosivity. A sample protocol for the validated reference test method is available at *http://iccvam.niehs.nih.gov*.

2.3.1 Test Method Components (Membrane Barrier, Categorization Solutions, Indicator Solution)

Membrane Barrier: The membrane barrier consists of two components -- a proteinaceous macromolecular aqueous gel and an underlying, permeable supporting membrane. The proteinaceous gel, composed of protein (e.g., keratin, collagen, or mixtures of proteins) forming a gel matrix, serves as the target for the test substance. It should be impervious to liquids and solids but able to be corroded and made permeable, presumably by the same mechanism(s) of corrosion that operates on living skin. The permeable supporting membrane provides mechanical support to the proteinaceous gel during the gelling process and exposure to the test substance, preventing sagging or shifting of the gel. The supporting membrane should be readily permeable to test substances so as not to interfere with its passage through to the indicator solution. The proteinaceous material is placed on the surface of the supporting membrane and allowed to gel prior to placing the membrane barrier over the indicator solution. The proteinaceous gel should be of equal thickness and density throughout, and with no air bubbles or defects that could affect its permeability or response to a corrosive test substance. The fully constructed membrane barrier should be stored under predetermined conditions shown to preclude deterioration of the gel (drying, microbial growth, etc) or loss of uniformity (shifting or cracking), which would degrade its performance. The acceptable storage period should be determined and membrane barrier preparations not used after that period.

Test Substance Categorization System: Experience with the validated reference system has shown that “strong” acids or bases and “weak” acids or bases behave somewhat differently in the time required to breakthrough the barrier membrane relative to their corrosive potential *in vivo*. Scoring of all test substances on a scale appropriate for strong acids and bases led to an over prediction of corrosivity for weak acids and bases. Thus, two scoring scales of breakthrough times are used to determine corrosivity (and UN Packing Group classification) or noncorrosivity for strong acids and bases and one for weak acids and bases. If a categorization system is used, objective criteria must be developed to place test substances into the appropriate categories for scoring. Changes in the pH of calibrated buffer solutions (one for acids and one for bases) could be used for this purpose. Specific ranges for strong and weak acids or bases should be defined.

Indicator Solution: An indicator solution responds to the presence of a test substance. This response can be assessed as an observable color change in a pH indicator dye, or by other types of chemical or electrochemical reactions. A pH-specific indicator dye or combination of dyes (e.g., cresol red and methyl orange) that will show a color change in response to the presence of the test substance can be used. The measurement system could be visual or electronic. Test substances must be determined to be capable of causing a measurable response in the indicator solution before they are considered qualified for evaluation in the test system.

2.3.2 Test Procedure

Test Substance Compatibility: Prior to testing, a qualification or compatibility test is performed to determine if the test substance can be detected by the indicator solution. The indicator system and the conditions of exposure used for the compatibility test must reflect the exposure in the subsequent corrosivity test. If the test substance is not detectable by the indicator solution, then the test system cannot be used to evaluate the corrosivity of that test substance.

Test Substance Categorization: If appropriate for the assay, a test substance that has been qualified by the compatibility test should be subjected to a categorization test (i.e., a screening test to distinguish between weak and strong acids or bases) to determine the appropriate breakthrough timescale to use for determining corrosivity and GHS skin corrosivity subcategory.

Assembly of the Test Method Components: The membrane barrier is positioned in a vial (or tube) containing the indicator solution so that the supporting membrane is in full contact with the indicator solution and with no air bubbles present. Care should be taken to ensure that barrier integrity is maintained.

Application of Test Substances: The assay is performed at room temperature (17-25°C), and a test substance is at room temperature when applied. A suitable amount of the test substance (e.g., 500 µL of liquid or 500 mg finely powdered solid) for the validated reference test method (InVitro Intl. 1995) is carefully layered onto the upper surface of the membrane barrier and distributed evenly. An appropriate number of replicates (e.g., four, as is used in the validated reference method) are prepared for each test substance and the concurrent controls. The time of addition of the test substance is recorded. To ensure that short corrosion times can be accurately recorded, the application times of the test substance to the replicate vials are staggered.

2.3.3 Control Substances

Solvent Controls: In tests that involve the use of a vehicle or solvent with the test substance, the vehicle or solvent must be compatible with the barrier system (i.e., not alter the integrity of the membrane barrier system) and should not alter the corrosivity of the test substance. When applicable, solvent (or vehicle) controls should be tested concurrently with the test substance to demonstrate the compatibility of the solvent with the barrier system.

Positive (Corrosive) Controls: A positive control chemical should be tested concurrently with the test substance to demonstrate that the *in vitro* membrane barrier test system is functioning properly. The positive control should be well characterized for its corrosive activity and should generate a response that is low to intermediate within the range of corrosive responses for the assay. Thus, extremely corrosive (UN Packing Group I) or noncorrosive chemicals are of limited utility, while

a Packing Group II substance would allow detection of a too rapid or too slow breakthrough time. To measure performance of the test method close to the cut off time between corrosive and noncorrosive, a weak Packing Group III substance might be employed. An acceptable positive control response range must be developed based on the historical range of breakthrough times for the positive control(s) employed. In each study, the positive control should be evaluated to determine if the breakthrough time is within the acceptable positive control range. For the validated reference test method, the acceptable breakthrough time for sodium hydroxide pellets, a Packing Group II positive control, ranges from 10.6 to 15.9 minutes.

Negative (Noncorrosive) Controls: A noncorrosive substance should also be tested concurrently with the test substance as another quality control measure to demonstrate the functional integrity of the membrane barrier. Examples of noncorrosive substances used as negative controls in the validated reference test method include 10% citric acid or 6% propionic acid.

Benchmark Controls: Benchmark controls may be useful to demonstrate that the test method is functioning properly for detecting the dermal corrosivity potential of chemicals of a specific chemical class or a specific range of responses, or for evaluating the relative corrosivity potential of a corrosive test substance. Appropriate benchmark controls should have the following properties:

- consistent and reliable source(s) for the chemical
- structural and functional similarity to the class of the substance being tested
- known physical/chemical characteristics
- supporting data on known effects in animal models
- known potency in the range of response (including moderate response)

2.3.4 Measurement of Membrane Barrier Penetration

Each vial is appropriately monitored and the time of the first change in the indicator solution (i.e., barrier penetration) is recorded. The difference in time between application of the test substance and penetration of the membrane barrier is determined.

2.3.5 Interpretation of Results

According to the established time parameters for each UN Packing Group, the time (in minutes) elapsed between application of the test substance and barrier penetration is used to predict the corrosivity of a test substance. For a test to be considered acceptable, the concurrent positive control must give the expected penetration response time, and, when included, the concurrent solvent control must not be corrosive.

2.3.6 Classification of Test Substances

The time (in minutes) elapsed between application and appearance of a color change in the CDS is used to classify the test substance in terms of corrosivity and, if applicable, UN Packing Group.

2.3.7 Test Report

The test report should include the following information, if relevant to the conduct of the study:

Test and Control Substances

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

Justification of the Test Method and Protocol Used

Test Method Integrity

- The procedure used to ensure the integrity (i.e., accuracy and reliability) of the test method over time
- If the test method employs proprietary components, the procedure used to ensure their integrity from "lot-to-lot" and over time
- The procedures that the user may employ to verify the integrity of the proprietary components

Criteria for an Acceptable Test

- Acceptable concurrent negative control ranges based on historical data
- Acceptable concurrent positive control ranges based on historical data

Test Conditions

- Apparatus and preparation procedures used
- Source and composition of the biological membrane barrier
- Composition and properties of the qualification and detection solutions
- Method of measurement of effect
- Details of test procedure used (e.g., test substance amounts, number of replicates, method of application, observation times)
- Description of any modifications of the test procedure
- Reference to historical data of the model
- Description of the evaluation and classification criteria used

Results

- Tabulation of test results from individual test samples; (i.e., the time in minutes elapsed between application and barrier penetration for the test substance and the positive, negative, solvent, and benchmark controls reported as individual replicate data, as well as means ± the standard deviation for each trial)

Description of Other Effects Observed

Discussion of the Results

Conclusion

2.4 Reference Chemicals

To ensure that a proposed *in vitro* membrane barrier test method possesses reliability and accuracy characteristics that are comparable to the validated reference test method, the 40 reference chemicals listed in **Table 2-1** must be used. However, to demonstrate technical proficiency, users of the validated reference test method or other similar validated test method that adhere to these

performance standards may want to evaluate their ability to correctly identify the dermal corrosivity classification of a subset of twelve of the chemicals (e.g., 3 noncorrosives, 3 from each Packing Group subcategory) that were correctly identified by the reference test method (see **Table 2-1**). The 40 reference chemicals represent relevant chemical classes and the range of corrosivity responses (i.e., noncorrosives; Packing Group I, II, and III corrosives) and were selected from the 163 chemicals used for the validation of the *in vitro* reference test method. These 40 chemicals consist of eight acid derivatives, eight inorganic acids, eight organic acids, seven organic bases, two acid esters, four inorganic bases, one electrophile, one quaternary ammonium, and one surfactant. They represent the minimum number of reference chemicals that should be used to evaluate the performance of a mechanistically and functionally similar, proposed test method. These chemicals should not be used to develop the prediction model for the proposed test method. If any of the recommended chemicals are unavailable, other chemicals for which adequate *in vivo* reference data are available could be substituted. To the extent possible, the substituted chemical(s) should be of the same chemical class as the original chemical(s). If desired, additional chemicals representing other chemical or product classes and for which adequate reference data are available can be used to more comprehensively evaluate the accuracy of the proposed test method. However, these additional chemicals should not include any that had been used to develop the prediction model for the proposed test method.

The distribution of chemicals in this list by corrosivity and UN Packing Group classification are:

- 12 Noncorrosive Chemicals
- 28 Corrosive Chemicals
  - 9 UN Packing Group I
  - 9 UN Packing Group II
  - 10 UN Packing Group III

2.5 Accuracy and Reliability

When evaluated using the minimum list of recommended reference chemicals in **Table 2-1**, the reliability and accuracy (i.e., sensitivity, specificity, false positive rates, and false negative rates) of the proposed *in vitro* membrane test method should be at least comparable to that of the validated *in vitro* membrane barrier test method (ICCVAM 1999). Noncorrosive and corrosive chemicals, ranging in activity from strong to weak, and representing relevant chemical classes are included so that the performance of the proposed test method can be determined and compared to that of the validated reference test method. For purposes of transportation hazard classification, the list of corrosive chemicals also covers the range of UN Packing Group classifications (ICCVAM 1999; ECVAM 2001). Including these substances will allow for the determination of whether the breakthrough times used to assign test substances to different UN Packing Groups are appropriate.

The penetration times associated with the assignment of each UN Packing Group (or other classification) must be determined for each composition of barrier, indicator, and categorization system. The reliability of the proposed *in vitro* test system, as well as its ability to over- and under-predict known corrosive substances, should be determined prior to testing new chemicals. Based on experience with the validation of different *in vitro* test methods, one effective approach used to establish intra- and inter-laboratory reproducibility for a test method not previously validated is

Table 2-1 Recommended Chemicals for Validation of New *In Vitro* Membrane Corrosivity Test Methods

Chemical[1]	CASRN	Chemical Class[2]	Conc[3] (%)	UN *In Vivo* PG[4]	Validated Test Method PG	pH[3]
Fluorosulfonic acid	7789-21-1	inorganic acid	neat	I	I	0
Nitric acid	7697-37-2	inorganic acid	90	I	I	0
Phosphorus pentachloride	10026-13-8	inorganic acid	98	I	I	0
Selenic acid	7783-08-6	inorganic acid	95	I	I	0
Boron trifluoride dehydrate	13319-75-0	inorganic acid	96	I	I	0.4
Phosphorus tribromide	7789-60-8	inorganic acid	97	I	I	1.0
Sulfuric acid, 10% wt.	7664-93-9	inorganic acid	10	I	I	1.2
Benzyl chloroformate	501-53-1	acid derivative	95	I	NC	2.5
1,2-Diaminopropane	78-90-0	organic base	NA	I	II	8.3
Phosphoric acid	7664-38-2	inorganic acid	85	II	II	0.4
Valeryl chloride	638-29-9	acid derivative	98	II	II	0.5
Acetic acid	64-19-7	organic acid	99+	II	II	1.9
Caprylic acid	124-07-2	organic acid	95	II	NC	2.7
Capric:caprylic acid (45:55)	68937-75-7	organic acid	95	II	NC	3.0
Ammonium hydrogen difluoride	1341-49-7	acid derivative	98	II	II	5.2
1-(2-Aminoethyl) piperazine	140-31-8	organic base	99	II	II	11.8
Ethanolamine	141-43-5	organic base	99+	II	II	11.8
Sodium hydroxide	1310-73-2	inorganic base	100	II	II	13.8
Cyanuric chloride	108-77-0	acid derivative	99	III	III	1.7
Benzenesulfonyl chloride	98-09-9	acid derivative	Neat	III	III	1.8
Crotonic acid	107-93-7	organic acid	99+	III	III	2.3
Butyric anhydride	106-31-0	acid derivative	99	III	III	3.1
Hydroxylamine sulfate	10039-54-0	organic acid	97+	III	III	3.6
2-Methylbutyric acid	600-07-7	organic acid	NA	III	III	3.6
Dicyclohexylamine	101-83-7	organic base	99	III	III	9.6
N,N-Dimethyl benzylamine	103-83-3	organic base	99	III	III	10.7
Tetraethylenepent-amine	112-57-2	organic base	neat	III	III	11.9
2-Ethylhexylamine	104-75-6	organic base	98	III	III	12.0
Maleic acid	110-16-7	organic acid	99	NC	II	1.3

Chemical[1]	CASRN	Chemical Class[2]	Conc[3] (%)	UN *In Vivo* PG[4]	Validated Test Method PG	pH[3]
Copper(II) chloride	7447-39-4	acid derivative	97	NC	II	3.0
Eugenol	97-53-0	organic acid	NA	NC	NC	3.7
Chromium(III) fluoride	7788-97-8	acid derivative	97	NC	NC	3.9
Cinnamaldehyde	14371-10-9	electrophile	100	NC	NC	3.9
Ethyl triglycol methacrylate	39670-09-2	acid ester	neat	NC	NC	4.5
Nonyl acrylate	2664-55-3	acid ester	neat	NC	NC	6.9
Benzalkonium chloride	8001-54-5	quaternary ammonium	100	NC	NC	7.6
Sodium acid carbonate	144-55-8	inorganic base	100	NC	NC	8.3
Sodium undecylenate	3398-33-2	surfactant	33	NC	NC	8.3
Sodium carbonate, 50% aqueous	497-19-8	inorganic base	100	NC	II	11.7
Calcium carbonate	471-34-1	inorganic base	neat	NC	NC	12.6

Abbreviations: CASRN = Chemical Abstracts Service Registry Number; Conc = concentration; NA = not available; NC = noncorrosive; PG = Packing Group; UN = United Nations.

[1]These chemicals, sorted first by *in vivo* rabbit skin corrosivity response and then by pH, represent the range of chemical classes and corrosivity responses [e.g., noncorrosives; UN Packing Groups I, II, and III corrosives] used to validate Corrositex® (ICCVAM 1999). The goal of the selection process was to include, to the extent possible, chemicals that: were representative of the range of corrosivity responses (e.g., noncorrosives; UN Packing Groups I, II, and III corrosives) that the validated reference test method is capable of measuring or predicting; were representative of the chemical classes used during the validation process; reflected the overall performance characteristics of the validated reference test method; have chemical structures that were well-defined; induced reproducible results in the validated reference test method; induced definitive results in the *in vivo* reference test; were commercially available; and were not associated with prohibitive disposal costs.

[2]Chemical class assigned by Barratt et al. (1998) and InVitro International, as provided to ICCVAM (1999).

[3]The concentration tested and the pH values were obtained from the original sources as indicated in ICCVAM (1999).

[4] Within the UN Globally Harmonized System of Classification and Labeling of Chemicals (GHS), the PG classifications correspond as follows: PG I = 1A, PG II = 1B, PG III = 1C (UNECE 2003).

to test each of the reference chemicals three times in each of three independent laboratories. The accuracy of the validated *in vitro* membrane barrier test method for the 40 reference chemicals, and the corresponding values obtained for the complete database considered by ICCVAM in its evaluation of this test method are summarized in **Table 2-2**. The accuracy of the validated *in vitro* membrane barrier test method for the reference chemicals and the corresponding values obtained for the total database compiled during the ICCAM evaluation process are not identical due to constraints associated with the chemical selection process.

The reliability of the proposed test method should also be comparable to that of the validated reference method. However, an assessment of inter-laboratory reproducibility is not essential if the test method is to be used in one laboratory only. The overall inter-laboratory reproducibility of the proposed *in vitro* membrane barrier test method for correctly classifying the UN Packing

group of a test substance detected as corrosive should be at least 93% (ICCVAM 1999; Fentem et al. 1998). In terms of membrane breakthrough times, the overall median coefficient of variation (CV) should not exceed 30% for studies conducted in different laboratories and should not exceed 5% for replicate measurements within an experiment (ICCVAM 1999; Fentem et al. 1998).

Table 2-2 Accuracy of the Validated *In Vitro* Membrane Barrier Test System (Corrositex®) for Skin Corrosion[1]

Source	# of Chemicals	Sensitivity[2]	Specificity[2]	False Negative Rate[2]	False Positive Rate[2]	UN Packing Group Accuracy[2]
Reference Chemicals	40	89% (25/28)	75% (9/12)	11% (3/28)	25% (3/12)	96% (24/25)
ICCVAM (1999)	163	85% (76/89)	70% (52/74)	15% (13/89)	30% (22/74)	Not Determined

Definitions: Sensitivity is defined as the proportion of all positive chemicals or chemical mixtures that are correctly classified as positive in a test. Specificity is defined as the proportion of all negative chemicals or chemical mixtures that are correctly classified as negative in a test. False positive rate is defined as the proportion of all negative chemicals or chemical mixtures that are falsely identified as positive. False negative rate is defined as the proportion of all positive chemicals or chemical mixtures that are falsely identified as negative. UN Packing Group Accuracy reflects the frequency with which Corrositex® correctly assigned the UN Packing Group classification to a substance the *in vitro* test method correctly classified as corrosive.

[1]The validation database is limited to those chemicals that qualified for testing in Corrositex®. The ability of the validated *in vitro* membrane barrier test system to correctly identify the corrosivity potential of the reference chemicals and the corresponding performance characteristics obtained for the complete database evaluated during the ICCVAM evaluation process are not identical due to the constraints associated with the reference chemical selection process. The goal of the selection process was to include chemicals that were representative of the range of corrosivity responses (e.g., noncorrosives; UN Packing Groups I, II, and III corrosives) that the validated reference test method is capable of measuring or predicting; were representative of the chemical classes used during the validation process; reflected the overall performance characteristics of the validated reference test method; have a chemical structure that was well-defined; induced reproducible results in the validated reference test method; induced definitive results in the *in vivo* reference test; were commercially available; and were not associated with prohibitive disposal costs.

[2]In this analysis (see ICCVAM [1999]), a substance is first classified as positive or negative for corrosivity within each laboratory based on the majority of test results obtained (when replicate testing was conducted). Next, the substance is classified as positive or negative for corrosivity based on the majority of test results obtained in multiple laboratories (when multiple laboratory studies were conducted). This approach was used due to the considerable variability in the database in the number of times a substance was tested.

This Page Intentionally Left Blank

3.0 *IN VITRO* HUMAN SKIN MODEL SYSTEMS FOR SKIN CORROSION

3.1 Background

Pre-validation and validation studies have been completed for an *in vitro* human skin model system commercially available as EPISKIN™ (ICCVAM 2002; Fentem et al. 1998; Botham et al. 1992; Botham et al. 1995; Barratt et al. 1998). Based on its scientific validity, this test method has been recommended for the testing of all classes of chemicals (ICCVAM 2002; Fentem et al. 1998; Balls and Corcelle 1998b) and for inclusion in tiered testing strategies as part of a tiered or weight-of-evidence evaluation (ICCVAM 2002). In addition to EPISKIN™, a related human skin model corrosivity test method marketed as EpiDerm™ (EPI-200) has been validated (Liebsch et al. 2000). Neither test method has been validated for categorizing the corrosive properties of chemicals across the three UN Packing Group subcategories of corrosivity (ICCVAM 2002; Liebsch et al. 2000; Balls and Hellsten 2000). This chapter briefly describes the principles of *in vitro* human skin model systems for corrosivity followed by the recommended performance standards, which consists of essential test method components, reference chemicals, and comparison of accuracy and reliability.

3.2 Principles of *In Vitro* Human Skin Model Systems for Skin Corrosion

The test material is applied topically to a three-dimensional human skin model, comprised of at least a reconstructed epidermis with a functional stratum corneum. Corrosive materials are identified by their ability to induce a decrease in cell viability below defined threshold levels at specified exposure periods. The principle of the human skin model assay is based on the premise that corrosive chemicals are able to penetrate the stratum corneum by diffusion or erosion, and are cytotoxic to the cells in the underlying layers. The use of test systems that include human-derived cells or tissue should be in accordance with applicable national and international laws, regulations, and policies.

Investigators using an *in vitro* human skin model system for skin corrosion must be able to demonstrate that the assay is valid for its intended use. This includes demonstrating that different preparations are consistent in barrier properties (i.e., capable of maintaining a barrier to noncorrosive substances, able to respond appropriately to weak and strong corrosive substances) and/or that any modification to the existing validated reference test method does not adversely affect its performance characteristics.

In vitro human skin model systems for skin corrosion may be used to test solids, liquids, and emulsions of any chemical or product class. The liquids can be aqueous or nonaqueous; solids can be soluble or insoluble in water. The samples may be pure chemicals, dilutions, formulations, or waste. Where appropriate, solids should be ground to a powder before application; no other prior treatment of the sample is required. In some chemical classes, relatively few chemicals were included in the validation of the accepted *in vitro* human skin model system for skin corrosion (Fentem et al. 1998). However, taking into account the limited mechanisms that result in corrosivity, this method is expected to be generally applicable across all chemical classes (ICCVAM 2002; Fentem et al. 1998; Balls and Corcelle 1998b).

3.3 Essential Test Method Components

The following is a description of the essential test method components for *in vitro* human skin model test methods for skin corrosivity, as provided in OECD Test Guideline 431 (OECD 2003a). Human skin models can be obtained commercially (e.g., EPISKIN™, EpiDerm™ [EPI-200]) or they can be developed or constructed in the testing laboratory (Ponec et al. 2000; Wilkins et al. 1994).

3.3.1 *In Vitro* Human Skin Model Conditions

Human keratinocytes should be used to construct the epithelium. Multiple layers of viable epithelial cells should be present under a functional stratum corneum. The skin model may also have a stromal component layer. Stratum corneum should be multilayered with the necessary lipid profile to produce a functional barrier with robustness to resist the rapid penetration of cytotoxic chemicals used as positive controls. The containment properties of the model should prevent the passage of material around the stratum corneum to the viable tissue, which would lead to poor modeling of the exposure to skin. The skin model should be free of contamination with bacteria, mycoplasma, or fungi.

The magnitude of viability is usually quantified by using MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, thiazolyl blue; CASRN 298-93-1) or other metabolically converted vital dyes (Marshall et al. 1995). The negative control tissue should be stable in culture (provide similar viability measurements) for the duration of the test exposure period. The stratum corneum should be sufficiently robust to resist the rapid penetration of positive control chemicals (e.g., 1% Triton X-100), which can be assessed by the exposure time required to reduce cell viability by 50%.

3.3.2 Application of the Test Substances

Two tissue replicates are used for each test and control substance. For liquid materials, sufficient test substance must be applied to uniformly cover the skin surface; a minimum of 25 $\mu L/cm^2$ should be used. For solid materials, sufficient test substance must be applied evenly to cover the skin surface, and it should be moistened with deionized or distilled water to ensure good contact with the skin. Where appropriate, solids should be ground to a powder before application. At the end of each exposure period (3 minutes to 1 or 4 hours), the test material must be carefully washed from the skin surface with an appropriate buffer or 0.9% NaCl.

3.3.3 Control Substances

Solvent Controls: In tests that involve the use of a vehicle or solvent with the test substance, the vehicle or solvent must be compatible with the barrier system (i.e., not alter the integrity of the membrane barrier system) and should not alter the corrosivity of the test substance. When applicable, solvent (or vehicle) controls should be tested concurrently with the test substance to demonstrate the compatibility of the solvent with the barrier system.

Positive (Corrosive) Controls: A positive control chemical should be tested concurrently with the test substance to demonstrate that the human skin membrane barrier is functioning properly. The positive control should be well characterized for its corrosive activity and should generate a response that is low to intermediate within the range of corrosive responses for the assay. An acceptable

positive control response range must be developed based on historical positive control(s) data. In each test, the positive control should be evaluated to determine if the value is within the acceptable positive control range. Typically, for biologically-based test methods, acceptable ranges are within 2 to 3 standard deviations of the historical mean response but developer of proprietary test methods may establish tighter ranges. Glacial acetic acid is an example of a positive control substance producing a low to intermediate response in the validated reference test method.

Negative (Noncorrosive) Controls: A noncorrosive substance should also be tested concurrently with the test substance as another quality control measure to demonstrate the functional integrity of the human skin membrane barrier. Examples of noncorrosive substances used as negative controls in the validated reference test method include 0.9% sodium chloride and water.

Benchmark Controls: Benchmark controls may be useful to demonstrate that the test method is functioning properly for detecting the dermal corrosivity potential of chemicals of a specific chemical class or a specific range of responses, or for evaluating the relative corrosivity potential of a corrosive test substance. Appropriate benchmark controls should have the following properties:

- consistent and reliable source(s) for the chemical
- structural and functional similarity to the class of the substance being tested
- known physical/chemical characteristics
- supporting data on known effects in animal models
- known potency in the range of response (including moderate response)

3.3.4 Viability Measurements

Only standardized, quantitative methods should be used to measure cell viability. Furthermore, the measure of viability must be compatible with use in a three-dimensional tissue construct. Non-specific dye binding must not interfere with the viability measurement. Protein binding dyes and those that do not undergo metabolic conversion (e.g., neutral red) are therefore not appropriate. The most frequently used assay is MTT reduction, which has been shown to give accurate and reproducible results (Fentem et al. 1998) but others may be used.

Chemical action by the test material on the vital dye may mimic that of cellular metabolism leading to a false estimate of viability. This has been shown to happen when such a test material is not completely removed from the reconstructed skin by rinsing (Liebsch et al. 2000). If the test material directly acts on the vital dye, additional controls should be used to detect and correct for test substance interference with the viability measurement (Liebsch et al. 2000; Fentem et al. 2001).

3.3.5 Interpretation of Results

The optical density (OD) values obtained for each test sample can be used to calculate percentage viability relative to the negative control, which is arbitrarily set at 100%. The cell viability criteria used to distinguish between corrosive and noncorrosive test chemicals (or to discriminate between different corrosive classes), or the statistical procedure(s) used to evaluate the results and identify corrosive materials must be clearly defined and documented, and be shown to be appropriate. In general, such criteria are established during test optimization, tested during a prevalidation phase,

and confirmed in a validation study. As examples, the predictions of corrosivity associated with EPISKIN™ (Fentem et al. 1998) and EpiDerm™ (EPI-200) (Liebsch et al. 2000) are:

EPISKIN™: The test substance is considered to be corrosive to skin:

i) if the viability after 3 minutes of exposure is less than 35%, or
ii) if the viability after 3 minutes of exposure is greater than or equal to 35% and the viability after 4 hour of exposure is less than 35%.

The test substance is considered to be noncorrosive to skin:

i) if the viability after 4 hours of exposure is greater than or equal to 35%.

EpiDerm™ (EPI-200): The test substance is considered to be corrosive to skin:

i) if the viability after 3 minutes of exposure is less than 50%, or
ii) if the viability after 3 minutes of exposure is greater than or equal to 50% and the viability after 1 hour of exposure is less than 15%.

The test substance is considered to be noncorrosive to skin:

i) if the viability after 3 minutes of exposure is greater than or equal to 50% and the viability after 1 hour of exposure is greater than or equal to 15%.

3.3.6 Test Report

The test report should include the following information, if relevant to the conduct of the study:

Test and Control Substances

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

Justification of the Skin Model and Protocol Used

Test Method Integrity

- The procedure used to ensure the integrity (i.e., accuracy and reliability) of the test method over time
- If the test method employs proprietary components, the procedure used to ensure their integrity from "lot-to-lot" and over time
- The procedures that the user may employ to verify the integrity of the proprietary components

Criteria for an Acceptable Test

- Acceptable concurrent negative control ranges based on historical data
- Acceptable concurrent positive control ranges based on historical data

Test Conditions

- Cell system used
- Calibration information for measuring device used for measuring cell viability (e.g., spectrophotometer)

- Complete supporting information for the specific skin model used including its validity
- Details of test procedure used
- Test doses used
- Description of any modifications of the test procedure
- Reference to historical data of the model
- Description of evaluation criteria used

Results

- Tabulation of data from individual test samples (e.g., OD values and calculated percentage cell viability data for the test substance and the positive, negative, and benchmark controls, reported in tabular form, including data from replicate repeat experiments as appropriate, and means and ± the standard deviation for each trial)

Description of Other Effects Observed
Discussion of the Results
Conclusion

3.4 Reference Chemicals

To demonstrate technical proficiency with the validated reference test method, the user should evaluate his/her ability to correctly identify the dermal corrosivity classification of twelve of the chemicals (6 noncorrosive and 6 corrosives varying in corrosive potency) listed in Table 3-1. However, to ensure that a proposed *in vitro* skin model system possesses reliability and accuracy characteristics that are comparable to the validated reference test method, the 24 reference chemicals listed in **Table 3-1** must be used. The 24 reference chemicals (12 noncorrosives, 12 corrosives) listed in **Table 3-1** provide a representative distribution of the 60 chemicals used in the ECVAM validation study of EPISKIN™ (Fentem et al. 1998; Barratt et al. 1998) and cover the range of corrosivity responses obtained for the *in vivo* rabbit skin reference test method. The 24 reference chemicals include 23 of the 24 chemicals used to validate EPIDERM™ (EPI-200), a test method structurally and functionally similar to EPISKIN™ (Liebsch et al. 2000). Included in this list are five organic bases, four inorganic acids, three inorganic bases, three organic acids, three electrophiles, three phenols, two neutral organics, and one surfactant. These reference chemicals are the minimum number that should be used to evaluate the performance of a mechanistically and functionally similar, proposed test method. These chemicals should not be used to develop the prediction model for a proposed test method. If any of the recommended chemicals are unavailable, other chemicals for which adequate reference data are available could be substituted. To the extent possible, the substituted chemical(s) should be of the same chemical class as the original chemical(s). If desired, additional chemicals representing other chemical or product classes and for which adequate reference data are available can be used to more comprehensively evaluate the accuracy of a proposed test method. However, these additional chemicals should not include any that had been used to develop the prediction model for the proposed test method.

3.5 Accuracy and Reliability

When evaluated using the minimum list of recommended reference chemicals (**Table 3-1**), the proposed test method should have reliability and performance (i.e., sensitivity, specificity, false positive rates, and false negative rates) characteristics that are comparable to the performance of the validated reference test method (ICCVAM 2002; Fentem et al. 1998). Noncorrosive and corrosive

chemicals, ranging in activity from strong to weak, and representing relevant chemical classes are included so that the performance of the proposed test method can be determined and compared to that of the validated reference test method. Eleven of the 12 chemicals mentioned in the OECD proposed Test Guideline 431 (*In vitro* skin corrosion human skin model system) (OECD 2003a) are included. Acrylic acid, proposed by the OECD as a severe corrosive, was not included because the comparative performance of this chemical in the validated reference test method (EPISKIN™) and the *in vivo* rabbit skin corrosivity test had not been demonstrated and thus the accuracy of the validated reference test method for this chemical was not established. Based on experience with the validation of different *in vitro* test methods, one effective approach used to establish intra- and inter-laboratory reproducibility for a test method not previously validated is to test each of the reference chemicals three times in each of three independent laboratories. The accuracy of the validated *in vitro* human skin model test system, EPISKIN™, for the 24 reference chemicals and the complete validation database considered by ICCVAM are provided in **Table 3-2**. Its accuracy for the reference chemicals and the corresponding values obtained for the total database compiled during the ICCVAM evaluation process are not identical due to constraints associated with the chemical selection process.

Table 3-1 Recommended Chemicals for Validation of New *In Vitro* Human Skin Model Corrosivity Test Methods

Chemical[1]	CASRN	Chemical Class[2]	UN *In Vivo* PG	pH[3]
***In Vivo* Corrosives**				
Phosphorus tribromide	7789-60-8	inorganic acid	I	1.0
Sulfuric acid (10%)	7664-93-9	inorganic acid	II/III	1.2
Boron trifluoride dihydrate	13319-75-0	inorganic acid	I	1.5
Glycol bromoacetate (85%)	3785-34-0	electrophile	II/III	2.0
Caprylic acid	124-07-02	organic acid	II/III	3.6
2-tert-Butylphenol	88-18-6	phenol	II/III	3.9
Dimethyldipropylenetriamine	10563-29-8	organic base	I	8.3
Dimethylisopropylamine	996-35-0	organic base	II/III	8.3
1,2-Diaminopropane	78-90-0	organic base	I	8.3
n-Heptylamine	111-68-2	organic base	II/III	8.4
2-Mercapoethanol, sodium salt (45% aq.)	37482-11-4	inorganic base	II/III	12.0
Potassium hydroxide (10% aq.)	1310-58-3	inorganic base	II	13.1
***In Vivo* Noncorrosives**				
Sulfamic acid	5329-14-6	inorganic acid	NC	1.5
Isostearic acid	30399-84-9	organic acid	NC	3.6
Phenethyl bromide	103-63-9	electrophile	NC	3.6
Eugenol	97-53-0	phenol	NC	3.7
1,9-Decadiene	1647-16-1	neutral organic	NC	3.9
o-Methoxyphenol	90-05-1	phenol	NC	3.9
Sodium lauryl sulfate (20% aq.)	151-21-3	surfactant	NC	3.9

Chemical[1]	CASRN	Chemical Class[2]	UN *In Vivo* PG	pH[3]
Tetrachloroethylene	127-18-4	neutral organic	NC	4.5
4-Amino-1,2,4-triazole	584-13-4	organic base	NC	5.5
4-(methylthio)-Benzaldehyde	3446-89-7	electrophile	NC	6.8
Sodium carbonate (50% aq.)	7664-93-9	inorganic base	NC	11.7
Dodecanoic acid (lauric acid)	143-07-7	organic acid	NC	ND

Abbreviations: aq = aqueous; CASRN = Chemical Abstracts Service Registry Number; PG = Packing Group; NC = Noncorrosive; ND = not determined (unable to measure); UN = United Nations.

[1]These chemicals, sorted first by corrosives versus noncorrosives and then by pH, were selected from among the 60 chemicals used by ECVAM to validate EPISKIN™ (Fentem et al. 1998; Barratt et al. 1998). Unless otherwise indicated, the chemicals were tested at the purity level obtained when purchased from a commercial source (Barratt et al. 1998). The goal of the selection process was to include, to the extent possible, chemicals that: were representative of the range of corrosivity responses (e.g., noncorrosives; weak to strong corrosives) that the validated reference test method is capable of measuring or predicting; were representative of the chemical classes used in the validation process; reflected the performance characteristics of the validated reference test method; have chemical structures that were well-defined; induced reproducible results in the validated reference test method; induced definitive results in the *in vivo* reference test; were commercially available; and were not associated with prohibitive disposal costs.

[2]Chemical class assigned by Barratt et al. (1998).

[3]The pH values were obtained from Fentem et al. (1998) and Barratt et al. (1998).

The reliability of the proposed test method for the reference chemicals should be comparable to that of the validated reference test method. An assessment of interlaboratory reproducibility is not essential if the test method is to be used in one laboratory only. In terms of cell viability measurements, the median CV should not exceed 35% for studies conducted in different laboratories (ICCVAM 2002; Fentem et al. 1998). The median CV for replicate studies conducted in the same laboratory should be appreciably less than the median CV for studies conducted in different laboratories.

Table 3-2 Accuracy of the Validated *In Vitro* Human Skin Model System Test Method (EPISKIN™) for Skin Corrosion[1]

Source	# of Chemicals	# of Tests[2]	Sensitivity	Specificity	False Negative Rate	False Positive Rate
Reference Chemicals	24	216	83% (90/108)	79% (85/108)	17% (18/108)	21% (23/108)
Fentem et al. (1998)	60	540	83% (201/243)	80% (237/297)	17% (42/243)	20% (60/297)

Definitions: Sensitivity is defined as the proportion of all positive chemicals that are correctly classified as positive in a test. Specificity is defined as the proportion of all negative chemicals that are correctly classified as negative in a test. False positive rate is defined as the proportion of all negative chemicals or chemical mixtures that are falsely identified as positive. False negative rate is defined as the proportion of all positive chemicals or chemical mixtures that are falsely identified as negative.

[1]The ability of the validated *in vitro* human skin model system to correctly predict the *in vivo* rabbit skin corrosivity potential of the 24 reference chemicals and the corresponding performance characteristics obtained by Fentem et al. (1998) are not identical due to the constraints associated with selection of the reference chemicals. The goal of the selection process was to include, to the extent possible, chemicals that: were representative of the range of corrosivity responses (e.g., noncorrosives; weak to strong corrosives) that the validated reference test method is

capable of measuring or predicting; were representative of the chemical classes used during the validation process; reflected the performance characteristics of the validated reference test method; have a chemical structure that was well-defined; induced reproducible results in the validated reference test method; induced definitive results in the *in vivo* reference test; were commercially available; and were not associated with prohibitive disposal costs.

[2]In the Fentem et al (1998) validation study, each chemical was tested three times in each of three laboratories. Due to the presence of a balanced design, the performance characteristics are based on individual tests rather than individual chemicals.

4.0 *IN VITRO* SKIN TRANSCUTANEOUS ELECTRICAL RESISTANCE (TER) TESTS FOR SKIN CORROSION

4.1 Background

Prevalidation and validation studies have been completed for the rat skin TER assay (ICCVAM 2002; Fentem et al. 1998; Oliver et al. 1986; Oliver et al. 1988; Botham et al. 1992; Botham et al. 1995; Barratt et al. 1998). Based on its scientific validity, this test method has been recommended for the testing of all classes of chemicals (ICCVAM 2002; Fentem et al. 1998; Balls and Corcelle 1998a) and for inclusion in tiered testing strategies as part of a tiered or weight-of-evidence evaluation (ICCVAM 2002). This chapter briefly describes the principles of the *in vitro* skin TER test for corrosivity followed by the recommended performance standards, which consists of essential test method components, reference chemicals, and comparison of accuracy and reliability.

4.2 Principles of the *In Vitro* Skin TER Test for Skin Corrosion

The test material is applied for up to 24 hours to the epidermal surfaces of skin discs in a two compartment test system in which the skin discs function as the separation between the compartments. The skin discs are prepared from humanely killed 28-30 day old rats. Corrosive materials are identified by their ability to produce a loss of normal stratum corneum integrity and barrier function, which is measured as a reduction in the TER below a threshold level (Oliver et al. 1986). For rat skin TER, a cutoff value of 5 kΩ has been selected based on extensive data for a wide range of chemicals where the vast majority of values were either clearly well above or well below this value (Oliver et al. 1986). Generally, chemicals that are noncorrosive but irritating in animals do not reduce the TER below this cutoff value. However, the use of other skin preparations or other equipment may alter the cutoff value, necessitating further validation. A dye-binding step is incorporated into the test procedure for confirmation testing of positive results in the TER. The dye-binding step determines if the increase in ionic permeability is due to physical destruction of the stratum corneum.

Investigators using an *in vitro* skin TER corrosivity test must be able to demonstrate that the assay is valid for its intended use. This includes demonstrating that different preparations are consistent in barrier properties (i.e., capable of maintaining a barrier to noncorrosive substances, able to respond appropriately to weak and strong corrosive substances) and/or that any modification to the existing validated reference test method does not adversely affect its performance characteristics.

The *in vitro* TER test for skin corrosion may be used to test solids, liquids, and emulsions of any chemical or product class. The liquids can be aqueous or nonaqueous; solids can be soluble or insoluble in water. The samples may be pure chemicals, dilutions, formulations, or waste. Where appropriate, solids can be heated to 300°C to melt or soften the test material or ground to a powder before application; no other prior treatment of the sample is required. In some chemical classes, relatively few chemicals were included in the validation of the accepted *in vitro* rat skin TER test for skin corrosion (Fentem et al. 1998). However, considering the limited mechanisms that result in corrosivity, this method is expected to be generally applicable across all chemical classes (ICCVAM 2002; Fentem et al. 1998; Balls and Corcelle 1998a).

4.3 Essential Test Method Components

The following is a description of the essential test method components of the *in vitro* skin TER test for skin corrosivity, as provided in the OECD Test Guideline 430 (OECD 2003b).

4.3.1 Animals

All procedures involving the use of animals should be in compliance with relevant national animal welfare act regulations and policies, and the studies should be approved by the Institutional Animal Care and Use Committee or its equivalent. Rats are the species of choice because the sensitivity of their skin to chemicals in this test has been previously demonstrated (Oliver et al. 1986). 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. The use of skin from another species is possible as long as the test system is appropriately calibrated and the reliability and accuracy, using at the minimum, the provided list of reference chemicals (**Table 4-3**), is determined to be at least comparable to the performance characteristics of the validated reference test method.

If rat skin is used, 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, while 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 three days of the second wash, when the stratum corneum has recovered from the hair removal.

4.3.2 Preparation of Skin Discs

Animals are humanely killed when 28-30 days old; this age is critical to the performance of the assay. The dorsolateral 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 excised. The skin may be stored prior to use provided that positive and negative control data are equivalent to that obtained with fresh skin.

Each skin disc is placed over one of the ends of a polytetrafluoroethylene (PTFE) 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. Tube and 'O' ring dimensions are provided in OECD Test Guideline (OECD 2003b). 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) (OECD 2003b). 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.

Before testing begins, the electrical resistance of two skin discs is measured as a quality control procedure for each animal skin pelt. If both discs have resistance values greater than 10 kΩ then the remainder of the discs may be used for the test. If the resistance value is less than 10 kΩ, the remaining discs from that skin pelt should be discarded.

4.3.3 Application of Test Substances

Liquid test substances (150 µL) are applied uniformly to the epidermal surface inside the tube. 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. In order to achieve maximum contact with the skin, solids may need to be warmed to 300°C to melt or soften the test substance, or ground to produce a granular material or powder. Deionized water (150 µL) is added on top of the solid and the tube is gently agitated.

Three skin discs are used for each test and control substance; skin discs from a single animal should be used. Test substances are applied for 24 hours at 20-23°C. The test substance is removed by washing with a jet of tap water at temperatures up to 30°C, until no further material can be removed.

4.3.4 Control Substances

Solvent Controls: In tests that involve the use of a vehicle or solvent with the test substance, the vehicle or solvent must be compatible with the barrier system (i.e., not alter the integrity of the membrane barrier system) and should not alter the corrosivity of the test substance. When applicable, solvent (or vehicle) controls should be tested concurrently with the test substance to demonstrate the compatibility of the solvent with the barrier system.

Positive (Corrosive) Controls: A positive control chemical should be tested concurrently with the test substance to demonstrate that the *in vitro* skin TER test method is functioning properly. The positive control should be well-characterized for its corrosive activity and should generate a resistance value that is low to intermediate within the range of corrosive responses for this assay. An acceptable positive control response range must be developed based on historical positive control(s) data. In each test, the positive control should be evaluated to determine if the value is within the acceptable positive control range. Typically, for biologically-based test methods, acceptable ranges are within 2 to 3 standard deviations of the historical mean response but tighter ranges may be established by the developer of a proprietary test. 10 M Hydrochloric acid is an example of a positive control substance used in the rat skin TER assay.

Negative (Noncorrosive) Controls: A noncorrosive substance should also be tested concurrently with the test substance as another quality control measure to demonstrate the functional integrity of the human skin membrane barrier. An examples of a noncorrosive substance used as a negative control in the validated reference test method is distilled water.

Benchmark Controls: Benchmark controls may be useful to demonstrate that the test method is functioning properly for detecting the dermal corrosivity potential of chemicals of a specific chemical class or a specific range of responses, or for evaluating the relative corrosivity potential of a corrosive test substance. Appropriate benchmark controls should have the following properties:

- consistent and reliable source(s) for the chemical
- structural and functional similarity to the class of the substance being tested
- known physical/chemical characteristics
- supporting data on known effects in animal models
- known potency in the range of response (including moderate response)

4.3.5 TER Measurements

The skin impedance is measured as TER by using a low-voltage, alternating current Wheatstone bridge (Oliver et al. 1986). General specifications of the bridge are 1-3 V operating voltage, a sinus or rectangular shaped alternating current of 50–1000 Hz, and a measuring range of at least 0.1 -30 kΩ. For the TER 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 (OECD 2003b). Electrode dimensions and the length of the electrode exposed below the crocodile clips are provided in the OECD Test Guideline (OECD 2003b). 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 (Balls and Corcelle 1998a), because this distance affects the resistance value obtained. Consequently, the distance between the inner electrode and the skin disc should be constant and minimal (1-2 mm).

If the measured resistance value is greater than 20 kΩ, this may be due to the remains of the test substance 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 $MgSO_4$.

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 described by OECD in Test Guideline 430. 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 calibration chemicals (see Section 4.4).

4.3.6 Dye-Binding Methods

Exposure of certain noncorrosive materials can result in a reduction of resistance below the cutoff of 5 kΩ allowing the passage of ions through the stratum corneum, thereby reducing the electrical resistance (Fentem et al. 1998). For example, neutral organics and 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 the rat skin TER values of test substances are less than or around 5 kΩ in the absence of visual damage, 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 (Fentem et al. 1998; Botham et al. 1995). In the latter case where the stratum corneum is disrupted, the dye sulforhodamine B (Acid Red 52; Color Index 45100; CASRN 3520-42-1), when applied to the skin surface rapidly penetrates and stains the underlying tissue. This particular dye is stable to a wide range of chemicals and is not affected by the extraction procedure described below.

Sulforhodamine B Dye Application and Removal: Following TER assessment, the magnesium sulfate is discarded from the tube and the skin is carefully examined for obvious damage. If there is no obvious major damage, 150 µL of a 10% (w/v) dilution of sulforhodamine B in distilled water, is applied to the epidermal surface of each skin disc for two hours. These 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 deionized water (8 mL). 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 5 mL of 30% (w/v) sodium dodecyl sulfate (SDS) in distilled water and are incubated overnight at 60°C.

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

Calculation of Dye Content: The sulforhodamine B dye content per disc is calculated from the OD values (Fentem et al. 1998) (sulforhodamine B dye molar extinction coefficient at 565 nm = 8.7 x l04; molecular weight = 580). The dye content is determined for each skin disc by the use of an appropriate calibration curve and a mean dye content is then calculated for the replicates.

4.3.7 Interpretation of Results

The mean rat skin TER results are accepted if the concurrent positive and negative control values fall within the acceptable ranges for the testing laboratory. The acceptable resistance ranges for the rat skin TER methodology and apparatus described above are given in **Table 4-1**.

Table 4-1 Acceptable Resistance Ranges for the Rat Skin TER Methodology and Apparatus

Control	Substance	Resistance range (kΩ)
Positive	10 M Hydrochloric acid	0.5 - 1.0
Negative	Distilled water	10 - 25

The mean dye-binding results are accepted on condition that concurrent control values fall within the acceptable ranges for the method. Suggested acceptable dye content ranges for the control substances for the rat skin TER methodology and apparatus described above are provided in **Table 4-2**.

Table 4-2 Suggested Acceptable Dye Content Ranges for the Control Substances for the Rat Skin TER Methodology and Apparatus

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

The test substance is considered to be noncorrosive to skin:

i) if the mean TER value obtained for the test substance is greater than 5 kΩ, or
ii) the mean TER value is less than or equal to 5 kΩ, and
    - the skin disc is showing no obvious damage, and
    - the mean disc dye content is well below the mean disc dye content of the 10 M HCl positive control obtained concurrently.

The test substance is considered to be corrosive to skin:

i) if the mean TER value is less than or equal to 5 kΩ and the skin disc is obviously damaged, or
ii) the mean TER value is less than or equal to 5 kΩ, and
    - the skin disc is showing no obvious damage, but
    - the mean disc dye content is greater than or equal to the mean disc dye content of the 10 M HCl positive control obtained concurrently.

4.3.8 Test Report

The test report should include the following information, if relevant to the conduct of the study:

Test and Control Substances

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

Test Animals

- Strain and sex used
- Age of the animals when used as donor animals
- Source, housing condition, diet, etc.
- Details of the skin preparation

Justification of the Skin Model and Protocol Used

Test Method Integrity

- The procedure used to ensure the integrity (i.e., accuracy and reliability) of the test method over time
- If the test method employs proprietary components, the procedure used to ensure their integrity from "lot-to-lot" and over time
- The procedures that the user may employ to verify the integrity of the proprietary components

Criteria for an Acceptable Test

- Acceptable concurrent negative control ranges based on historical data
- Acceptable concurrent positive control ranges based on historical data

Test Conditions

- Calibration curves for test apparatus
- Calibration curves for dye-binding test performance

- Details of the test procedure used for TER measurements
- Details of the test procedure used for the dye-binding assessment, if appropriate
- Description of any modification of the test procedures
- Description of evaluation criteria used
- Reference to historical data of the model
- Description of evaluation criteria used

Results

- Tabulation of data from the TER and dye-binding assay (if appropriate) for individual animals and individual skin samples for the test material, as well as for positive and negative controls (individual trial data and means ± S.D.), including data for replicates/repeat experiments, mean and individual values
- Description of any effects observed
- Tabulation of data from individual test samples (e.g., resistance values [kΩ] and mean dye content values [µg/disc], where appropriate)

Description of Other Effects Observed
Discussion of the Results
Conclusion

4.4 Reference Chemicals

Calibration chemicals are used to demonstrate that the validated *in vitro* rat skin TER test method is performing as expected; reference chemicals are used to determine if the performance of a new or modified *in vitro* skin TER test for skin corrosion is comparable to that of the validated reference test method. The 24 reference chemicals (12 noncorrosives, 12 corrosives) listed in **Table 4-3** provide a representative distribution of the 60 chemicals used in the ECVAM validation study of the rat skin TER assay (Fentem et al. 1998; Barratt et al. 1998) and the range of corrosivity responses obtained for the *in vivo* rabbit skin reference test method. These reference chemicals are the minimum number that should be used to evaluate the performance of a mechanistically and functionally similar, proposed test method. These chemicals should not be used to develop the prediction model for a proposed test method. If any of the recommended chemicals are unavailable, other chemicals for which adequate reference data are available could be substituted. To the extent possible, the substituted chemical(s) should be of the same chemical class as the original chemical(s). If desired, additional chemicals representing other chemical or product classes and for which adequate reference data are available can be used to more comprehensively evaluate the accuracy of a proposed test method. However, these additional chemicals should not include any that had been used to develop the prediction model for the proposed test method.

Included in this list are five organic bases, four organic acids, four inorganic acids, three electrophiles, three neutral organics, two inorganic bases, two phenols, and one surfactant. A subset of the 24 reference chemicals (12 total; 6 noncorrosives, 6 corrosives) serve as calibration chemicals for the rat skin TER assay; the names of these chemical are bolded in **Table 4-3**.

Table 4-3 Recommended Chemicals for Validation of New *In Vitro* TER Corrosivity Test Methods

Chemical[1]	CASRN	Chemical Class[2]	UN *In Vivo* PG	pH[3]
***In Vivo* Corrosives**				
Phosphorus tribromide	7789-60-8	inorganic acid	I	1.0
Sulfuric acid (10%)	**7664-93-9**	**inorganic acid**	**II/III**	**1.2**
Boron trifluoride dehydrate	13319-75-0	inorganic acid	I	1.5
Glycol bromoacetate (85%)	3785-34-0	electrophile	II/III	2.0
Caprylic acid	**124-07-02**	**organic acid**	**II/III**	**3.6**
2-tert-Butylphenol	**88-18-6**	**phenol**	**II/III**	**3.9**
60/40 Caprylic/decanoic acids	68937-75-7	organic acid	II/III	3.9
Dimethyldipropylenetriamine	**10563-29-8**	**inorganic base**	**I**	**8.3**
Dimethylisopropylamine	996-35-0	organic base	II/III	8.3
1,2-Diaminopropane	**78-90-0**	**organic base**	**I**	**8.3**
n-Heptylamine	111-68-2	organic base	II/III	8.4
Potassium hydroxide (10% aq.)	**1310-58-3**	**inorganic base**	**II**	**13.1**
***In Vivo* Noncorrosives**				
Sulfamic acid	5329-14-6	inorganic acid	NC	1.5
Isostearic acid	**30399-84-9**	**organic acid**	**NC**	**3.6**
Phenethyl bromide	**103-63-9**	**electrophile**	**NC**	**3.6**
Eugenol	**97-53-0**	**phenol**	**NC**	**3.7**
1,9-Decadiene	1647-16-1	neutral organic	NC	3.9
Benzyl acetone	2550-26-7	neutral organic	NC	3.9
Sodium lauryl sulfate (20% aq.)	151-21-3	surfactant	NC	3.9
Tetrachloroethylene	**127-18-4**	**neutral organic**	**NC**	**4.5**
4-Amino-1,2,4-triazole	**584-13-4**	**organic base**	**NC**	**5.5**
4-(methylthio)-Benzaldehyde	**3446-89-7**	**electrophile**	**NC**	**6.8**
Sodium carbonate (50% aq.)	7664-93-9	inorganic base	NC	11.7
Dodecanoic acid (lauric acid)	143-07-7	organic acid	NC	ND

Abbreviations: aq = aqueous; CASRN = Chemical Abstracts Service Registry Number; PG = Packing Group; NC = Noncorrosive; ND = not determined (unable to measure); UN = United Nations. Recommended calibration chemicals are indicated in bold type.

[1]These chemicals, sorted first by corrosives versus noncorrosives and then by pH, were selected from among the 60 chemicals used by ECVAM to validate TER (Fentem et al. 1998; Barratt et al. 1998). Unless otherwise indicated, the chemicals were tested at the purity level obtained when purchased from a commercial source (Barratt et al. 1998). The goal of the selection process was to include, to the extent possible, chemicals that: were representative of the range of corrosivity responses (e.g., noncorrosives; weak to strong corrosives) that the validated reference test method is capable of measuring or predicting; were representative of the chemical classes used during the validation process; reflected the overall performance characteristics of the validated reference test method; have chemical structures that were well-defined; induced reproducible results in the validated reference test method; induced definitive results in the *in vivo* reference test; were commercially available; and were not associated with prohibitive disposal costs.

[2]Chemical class assigned by Barratt et al. (1998).

[3]The pH values were obtained from Fentem et al. (1998) and Barratt et al. (1998).

These 12 calibration and the 24 reference chemicals are the minimum number that should be used to calibrate the validated reference test method or to evaluate the performance of a new or modified *in vitro* skin TER test for skin corrosion, respectively. While not sufficient to allow for an assessment of the ability of an *in vitro* skin TER test to accurately predict the UN Packing Group classification for a test chemical, these chemicals are adequate to assess if a rat skin TER test is functioning appropriately and to assess the extent that a modified or new skin TER test can correctly identify corrosive and noncorrosive substances. These chemicals should not be used to develop the prediction model for an alternative skin TER test method. If any of the recommended chemicals are unavailable, other chemicals for which adequate reference data are available could be substituted. To the extent possible, the substituted chemical(s) should be of the same chemical class as the original chemical(s). If desired, additional chemicals representing other chemical or product classes and for which adequate reference data are available can be used to more comprehensively evaluate the accuracy of an alternative skin TER test method. However, these additional chemicals should not include any that had been used to develop the prediction model for the alternative skin TER test method.

4.5 Accuracy and Reliability

When calibrating the performance of the rat skin TER test, 100% concordance is required for the 12 calibration chemicals (6 corrosive, 6 noncorrosive) listed in **Table 4-3**. With one exception, these 12 chemicals are the same as those listed in OECD Test Guideline 430 (*In vitro* skin corrosion: transcutaneous electrical resistance test [TER]) (OECD 2003b). Acrylic acid, proposed by the OECD as a severe corrosive, was not included because the comparative performance of this chemical in EPISKIN™ and the *in vivo* rabbit skin corrosivity test had not been demonstrated and thus the accuracy of the validated reference test method for this chemical was not established.

When evaluated using the minimum list of recommended reference chemicals in **Table 4-3**, the reliability and accuracy (i.e., sensitivity, specificity, false positive rates, and false negative rates) of the proposed *in vitro* skin TER assay should be at least comparable to that of the validated *in vitro* rat skin TER test method (ICCVAM 2002). Noncorrosive and corrosive chemicals, ranging in activity from strong to weak, and representing relevant chemical classes are included so that the performance of the proposed test method can be determined and compared to that of the validated reference test method. Based on experience with the validation of different *in vitro* test methods, one effective approach used to establish intra- and inter-laboratory reproducibility for a test method not previously validated is to test each of the reference chemicals three times in each of three independent laboratories.

The accuracy of the validated *in vitro* rat skin TER test method for the 24 reference chemicals, and the corresponding values obtained for the complete database considered by ICCVAM in its evaluation of this assay, are summarized in **Table 4-4**. The accuracy of the validated *in vitro* rat skin TER test method for the reference chemicals and the corresponding values obtained for the total database compiled during the ICCVAM evaluation process are not identical due to constraints associated with the chemical selection process.

The reliability of the proposed test method for the reference chemicals should be comparable to that of the validated *in vitro* rat skin TER test method. An assessment of interlaboratory

reproducibility is not essential if the test method is to be used in one laboratory only. In terms of cell viability measurements, the median coefficient of variation (CV) should not exceed 35% for studies conducted in different laboratories (ICCVAM 2002; Fentem et al. 1998). The median CV for replicate studies conducted in the same laboratory should be appreciably less than median CV for studies conducted in different laboratories.

Table 4-4 Accuracy of the Validated *In Vitro* Rat Skin TER Test for Skin Corrosion[1]

Source	# of Chemicals	# of Tests[2]	Sensitivity	Specificity	False Negative Rate	False Positive Rate
Reference Chemicals	24	144	86% (62/72)	75% (54/72)	14% (10/72)	25% (18/72)
Fentem et al (1998)	60	355	88% (140/159)	72% (142/196)	12% (19/159)	28% (54/196)

Definitions: Sensitivity is defined as the proportion of all positive chemicals that are correctly classified as positive in a test. Specificity is defined as the proportion of all negative chemicals that are correctly classified as negative in a test. False positive rate is defined as the proportion of all negative chemicals or chemical mixtures that are falsely identified as positive. False negative rate is defined as the proportion of all positive chemicals or chemical mixtures that are falsely identified as negative.

[1]The ability of the validated *in vitro* rat skin TER test method to correctly predict the *in vivo* rabbit skin corrosivity potential of the 24 reference chemicals and the corresponding performance characteristics obtained by Fentem et al. (1998) are not identical due to the constraints associated with selection of the reference chemicals. The goal of the selection process was to include, to the extent possible, chemicals that: were representative of the range of corrosivity responses (e.g., negative; weak to strong positive corrosives) that the validated reference test method is capable of measuring or predicting; were representative of the chemical classes used in the validation process; reflected the performance characteristics of the validated reference test method; have a chemical structure that was well-defined; induced reproducible results in the validated reference test method; induced definitive results in the *in vivo* reference test; were commercially available; and were not associated with prohibitive disposal costs.

[2]In the Fentem et al (1998) validation study, each chemical was tested twice in each of three laboratories (with five failed tests). Due to the presence of a balanced design, the performance characteristics are based on individual tests rather than individual chemicals.

5.0 REFERENCES

Balls M, and Corcelle G. 1998a. Statement on the scientific validity of the rat skin transcutaneous electrical resistance (TER) test (an *in vitro* test for skin corrosivity). ATLA 26: 275-277.

Balls M, and Corcelle G. 1998b. Statement on the scientific validity of the EPISKIN™ test (an *in vitro* test for skin corrosivity). ATLA 26: 278-280.

Balls M, and Hellsten E. 2000. Statement on the Application of the EpiDerm human skin model for corrosivity testing. ATLA 28: 365-367.

Barratt MD, Brantom PG, Fentem JH, Gerner I, Walker AP, and Worth AP. 1998. The ECVAM international validation study on *in vitro* test for skin corrosivity. 1. Selection and Distribution of the Test Chemicals. Toxicol In Vitro 12:471-482.

Botham PA, Chamberlain M, Barratt MD, Curren RD, Esdaile DJ, Gardner JR., et al. 1995. A Prevalidation study on *in vitro* skin corrosivity testing: The Report and recommendations of ECVAM Workshop 6. ATLA 23:219-255.

Botham PA, Hall TJ, Dennett R, McCall JC, Basketter DA, Whittle E, et al. 1992. The skin corrosivity test *in vitro*. Results of an inter-laboratory trial. Toxicol In Vitro 6:191-194.

Cannon CL, Neal PJ, Southee JA, Kubilus J, and Klausner M. 1994. New epidermal model for dermal irritancy testing. Toxicol In Vitro 8: 889-891.

DOT. 2002. Exemption DOT-E-10904 (Fifth Revision). Available:
http://hazmat.dot.gov/exemptions/E10904.pdf [accessed 16 March 2004]

DOT. 2003a. Shippers--General Requirements for Shipments and Packagings--Class 8 Definitions. 49 CFR 173.136. Available: *http://www.access.gpo.gov/nara/cfr/waisidx_03/49cfr173_03.html* [accessed 16 March 2004]

DOT. 2003b. Shippers--General Requirements for Shipments and Packagings--Class 8 Assignment of Packing Group. 49 CFR 173.137.
Available: _ *http://www.access.gpo.gov/nara/cfr/waisidx_03/49cfr173_03.html*
[accessed 16 March 2004]

ECVAM. 2001. Statement on the application of the CORROSITEX assay for skin corrosivity testing. 15th Meeting of European Centre for the Validation of Alternative Methods Scientific Advisory Committee (ESAC), Ispra, Italy. ATLA 29(2), 96-97.

EPA. 1996. Method 1120, Dermal Corrosion. Washington, DC: U.S. Environmental Protection Agency. Available: *http://www.epa.gov/epaoswer/hazwaste/test/pdfs/1120.pdf*
[accessed 16 March 2004].

EPA. 1998. Health Effects Test Guidelines: Acute Dermal Irritation. OPPTS 870.2500. Washington, DC. U.S. Environmental Protection Agency. Available: *http://www.epa.gov/docs/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/870-2500.pdf* [accessed 16 March 2004].

Fentem JH, Archer GEB, Balls M, Botham PA, Curren RD, Earl LK, et al. 1998. The ECVAM international validation study on *in vitro* tests for skin corrosivity. 2. Results and evaluation by the management team. Toxicol In Vitro 12:483-524.

Fentem JH, Briggs D, Chesné C, Elliot GR, Harbell JW, Heylings JR., et al, 2001. A prevalidation study on *in vitro* tests for acute skin irritation: results and evaluation by the Management Team. Toxicol In Vitro 15:57-93.

Gordon VC, Harvell JD, and Maibach, HI. 1994. Dermal corrosion, the Corrositex® System: A DOT accepted method to predict corrosivity potential of test materials. In: *In vitro Skin Toxicology - Irritation, Phototoxicity, Sensitization. Alternative Methods in Toxicology.* Rougier A, Goldberg AM, Maibach, HI, eds. 10: 37-45. Mary Ann Liebert, New York.

ICCVAM. 1997. Validation and Regulatory Acceptance of Toxicological Test Methods. A Report of the *ad hoc* Interagency Coordinating Committee on the Validation of Alternative Methods. NIH Publication No. 97-3981. Research Triangle Park, NC. National Institute of Environmental Health Sciences. Available: *http://iccvam.niehs.nih.gov/docs/guidelines/validate.pdf*
[accessed 16 March 2004].

ICCVAM. 1999. Corrositex®: An *In Vitro* Test Method for Assessing Dermal Corrosivity Potential of Chemicals. NIH Publication No. 99-4495. Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available: *http://iccvam.niehs.nih.gov/docs/reports/corprrep.pdf* [accessed 16 March 2004].

ICCVAM. 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. National Institute of Environmental Health Sciences.
Available: *http://iccvam.niehs.nih.gov/methods/epiderm.htm* [accessed 16 March 2004].

ICCVAM. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods. NIH Publication No. 03-4508. Research Triangle Park, NC. National Institute of Environmental Health Sciences.
Available: *http://iccvam.niehs.nih.gov/docs/guidelines/subguide.htm* [accessed 16 March 2004].

InVitro International. 1995. Corrositex Instruction Manual. Irvine, CA.
Available: *http://www.invitrointl.com/* [accessed 16 March 2004].

Liebsch, M, Traue, D, Barrabas, C, Spielmann, H, Uphill, P, Wilkins, S, et al. 2000. The ECVAM prevalidation study on the use of EpiDerm for skin corrosivity testing. ATLA 28:371-401.

Marshall, NJ, Goodwin, CJ, and Holt, SJ. 1995. A critical assessment of the use of microculture tetrazolium assays to measure cell growth and function. Growth Regulation 5:69-84.

OECD. 1996. Final Report of the OECD Workshop on Harmonisation of Validation and Acceptance Criteria for Alternative Toxicological Methods. (Solna Report) Paris: Organisation for Economic Co-operation and Development. Available: *http://www.oecd.org* [accessed 16 March 2004].

OECD. 2002a. Report of the Stockholm Conference on Validation and Regulatory Acceptance of New and Updated Methods in Hazard Assessment. Paris: Organisation for Economic Co-operation and Development. Available: *http://www.oecd.org* [accessed 16 March 2004].

OECD. 2002b. Test Guideline 404. OECD Guideline for Testing of Chemicals: Acute Dermal Irritation/Corrosion. Paris: Organisation for Economic Co-operation and Development. Available: *http://www1.oecd.org/ehs/test/health.htm* [accessed 16 March 2004].

OECD. 2003a. OECD Guideline for the Testing of Chemicals: Draft Proposal for a New Guideline 431, *In Vitro* Skin Corrosion Human Skin Model System. OECD ENV/JM/TG(2002)8. Paris: Organisation for Economic Co-operation and Development. Available: *http://www.oecd.org* [accessed 16 March 2004].

OECD. 2003b. OECD Guideline for the Testing of Chemicals: Draft Proposal for a New Guideline 430, *In Vitro* Skin Corrosion: Transcutaneous Electrical Resistance Test (TER). Paris: Organisation for Economic Co-operation and Development. Available: *http://www.oecd.org* [accessed 16 March 2004].

Oliver GJA, Pemberton MA, and Rhodes C. 1986. An *in vitro* skin corrosivity test-modifications and validation. Fd Chem Toxicol 24:507-512.

Oliver GJA, Pemberton MA, and Rhodes C. 1988. An *in vitro* model for identifying skin corrosive chemicals. I. Initial validation. Toxicol In Vitro 2: 7-17.

Parenteau NL, Bilbo P, Nolte CJ, Mason VS, and Rosenberg M. 1992. The organotypic culture of human skin keratinocytes and fibroblasts to achieve form and function. Cytotechnology 9:163-171.

Ponec M, Boelsma E, Weerheim A, Mulder A, Bouwstra J, and Mommaas, M. 2000. Lipid and ultrastructural characterization of reconstructed skin models. Intl J Pharmaceutics 203:211-225.

UN. 2003. Globally Harmonized System of Classification and Labeling of Chemicals (GHS). [ST/SG/AC.10/30]. United Nations, New York and Geneva. Available: *http://www.unece.org/trans/danger/publi/ghs/officialtext.html* [accessed 16 March 2004]

UN. 2003b. Recommendations on the Transport of Dangerous Goods. Model Regulations. Thirteenth Revised Edition. United Nations, New York and Geneva. Available: *http://www.unece.org/trans/danger/publi/unrec/rev13/13files_e.html* [accessed 16 March 2004]

Wilkins LM, Watson SR, Prosky SJ, Meunier, SF, and Parenteau, NL. 1994. Development of a bilayered living skin construct for clinical applications. Biotechnology and Bioengineering 43: 747-756.

Worth AP, Fentem JA, Balls M, Botham PA, Curren RD, Earl LK, et al. 1998. An evaluation of the proposed OECD testing strategy for skin corrosion. ATLA 26: 709-721.

APPENDIX A

Federal Register Notice

Vol. 68, No. 126, pp. 39104-5, July 1, 2003

Notice of Availability of Proposed Minimum Performance Standards (MPS) for Three Types of *In Vitro* Methods for Assessing the Dermal Corrosivity Hazard Potential of Chemicals; Request for Comments

This Page Intentionally Left Blank

39104 **Federal Register** / Vol. 68, No. 126 / Tuesday, July 1, 2003 / Notices

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Public Health Service

National Toxicology Program (NTP); National Institute of Environmental Health Sciences (NIEHS); National Institutes of Health; Notice of Availability of Proposed Minimum Performance Standards (MPS) for Three Types of *In vitro* Methods for Assessing the Dermal Corrosivity Hazard Potential of Chemicals; Request for Comments

Summary

The NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) announces the availability of and invites public comment on proposed MPS for three types of *in Vitro* methods for assessing the dermal corrosivity hazard potential of chemicals. The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Dermal Corrosivity and Irritation Working Group (DCIWG) developed these proposed MPS. The ICCVAM developed the proposed MPS to communicate criteria which could be used to determine if similar test methods have comparable accuracy and reliability.

Availability of the Proposed MPS

Copies of the MPS are available electronically in PDF format on the ICCVAM/NICEATM web site at *http://iccvam.niehs.nih.gov* or in printed form by contacting Dr. William Stokes, NICEATM Director, NIEHS, P.O. Box 12233, MD EC–17, Research Triangle Park, NC, 27709, (phone) 919–541–3398, (fax) 919–541–0947, (e-mail) *iccvam@niehs.nih.gov*.

Request for Comments

NICEATM invites the submission of written comments on the proposed MPS. When submitting written comments, please refer to this **Federal Register** notice and provide applicable contact information (name, affiliation, mailing address, phone, fax, e-mail and sponsoring organization). Written comments should be sent by mail, fax, or e-mail to NICEATM (contact information provided above) by noon on August 15, 2003. All written comments received by this date will be posted on the ICCVAM/NICEATM web site and will be considered by the DCIWG and ICCVAM during development of the final ICCVAM MPS for these assays. Final ICCVAM MPS will be published as addendums to previously published ICCVAM reports on these test methods and will be forwarded to Federal agencies for their consideration. Availability of the final MPS will be announced via a **Federal Register** notice. Copies of the MPS will be made available electronically in PDF format on the ICCVAM/NICEATM web site or can be obtained in printed form by contacting NICEATM (contact information provided above).

Supplemental Information about the Proposed MPS

ICCVAM previously evaluated and recommended four validated test methods for assessing the dermal corrosivity hazard potential of chemicals: Corrositex®, EPISKIN™, EpiDerm™ (EPI–200), and the rat skin transcutaneous electrical resistance (TER) Assay (NIEHS 1999 and NIEHS 2002). Subsequently, the U.S. Environmental Protection Agency (EPA) requested that ICCVAM establish MPS for the three proprietary dermal corrosivity test methods (Corrositex®, EPISKIN™, EpiDerm™) and the non-proprietary rat skin TER test method. In response, the ICCVAM DCIWG drafted proposed MPS based on the validated reference test methods for these three types of *in vitro* dermal corrosivity assays: membrane barrier test methods, human skin model system test methods, and skin TER test methods.

The purpose of the MPS is to communicate the basis on which a validated and accepted proprietary (*e.g.*, copyrighted, trademarked, registered) or non-proprietary test method has been determined to have sufficient accuracy and reliability for a specific testing purpose. Accuracy refers to the ability of the test method to correctly predict or measure the biological effect of interest (also referred to as relevance) while reliability refers to the extent of intra- and inter-laboratory reproducibility. MPS also provide the criteria that should be met by other proposed test methods that are based on similar scientific principles and that measure or predict the same biological or toxic effect.

The three elements of MPS are:

- Minimum procedural standards that identify essential structural, functional, and procedural components of the validated reference test method (*e.g.*, procedural details, proper controls, morphologic structure and integrity of the test system, biological identity of key components, and expected biological responsiveness). Adherence to the minimum procedural standards will help to assure that the proposed test method is based on the same concepts as the referenced test method.
- A list of recommended reference chemicals that can be used to assess the accuracy and reliability characteristics of the proposed test method. The list includes substances that are representative of the chemical and product classes for which the validated test method is considered applicable, as well as substances that are representative of the range of responses (*e.g.*, negative, weak to strong positive) that the validated test method is capable of measuring or predicting.
- The accuracy and reliability that should be achieved by the proposed test method when evaluated using the minimum list of reference chemicals.

Background Information on ICCVAM and NICEATM

The NIEHS established the ICCVAM in 1997 to coordinate the interagency technical review of new, revised, and alternative test methods of interagency interest, and to coordinate cross-agency issues relating to the validation, acceptance, and national/international harmonization of toxicological testing methods. ICCVAM was established as a permanent interagency committee of the NIEHS under the NICEATM on December 19, 2000, by the ICCVAM Authorization Act of 2000 (Pub. L. 106–545, available at *http://iccvam.niehs.nih.gov/about/PL106545.pdf*. The Committee is composed of representatives from fifteen Federal regulatory and research agencies that use or generate toxicological information. Its purpose is to promote the scientific validation and regulatory acceptance of toxicological test methods that will improve the agencies' ability to accurately assess the safety or hazards of chemicals and various types of products, while refining, reducing, and replacing animal use wherever possible. NICEATM provides operational and scientific support for ICCVAM and ICCVAM-related activities. NICEATM and ICCVAM work collaboratively to evaluate new and improved test methods applicable to the needs of Federal agencies. Additional information about ICCVAM and NICEATM can be found at the following web site: *http://iccvam.niehs.nih.gov.*

References

NIEHS. 1999. Corrositex® An *In Vitro* Test Method for Assessing Dermal Corrosivity Potential of Chemicals. NIH Publication No. 99–4495. Available at: *http://iccvam.niehs.nih.gov/methods/corrode.htm.*

NIEHS. 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. Available at *http://iccvam.niehs.nih.gov/methods/epiderm.htm.*

Dated: June 12, 2003.

Samuel Wilson,

Deputy Director, National Institute of Environmental Health Sciences.

[FR Doc. 03–16506 Filed 6–30–03; 8:45 am]

BILLING CODE 4140B01BP

APPENDIX B

Public Comments in Response to the *Federal Register* Notice (July 1, 2003)

This Page Intentionally Left Blank

Institute for **I**n **V**itro **S**ciences, Inc.

Advancing **Science &** Animal **Welfare** Together

21 Firstfield Road
Suite 220
Gaithersburg, MD 20878
tel: 301.947.6523
fax: 301.947.6538
www.iivs.org

August 14, 2003

Dr. William Stokes, D.V.M.
NICEATM Director, NIEHS
P.O. Box 12233, MD EC-17
Research Triangle Park, NC, 27709

Dear Dr. Stokes:

The members of the NICEATM and ICCVAM DCIWG committee are to be commended for the considerable effort that went into preparing the draft Minimum Performance Standards (MPS) documents for the three in vitro corrosivity assays. In requesting that these documents be prepared, the EPA has helped us all. Conceptually, the MPS documents are a substantial step forward in regulatory toxicology. They link the validation of an assay system (test system, protocol, endpoint determinations, controls, and prediction model) not only to the application of the assay system, but also to the production of data for regulatory review. These documents, and those that follow, will serve several purposes that are discussed in more detail below.

To start with some background, new test methods undergo several stages of maturation. A newly developed test will first be subject to prevalidation where the effectiveness of the technology transfer process and final protocol development occurs. The final protocol will be used to develop the prediction model that will allow the data from the new test to be calibrated against the desired toxicological action. A training set of reference test materials is used to develop the prediction model. Finally, the new test is subjected to formal validation, usually in several laboratories. For the validation study, a new set of reference test materials is employed. From the validation study, the performance characteristics of the new test (test system, protocol, and prediction model) are determined. Part of this process involves the identification of the essential elements of the test and test system; those elements (independent variables) that must be maintained/controlled to make the test reliable and predictive. This analysis should be performed with both proprietary and nonproprietary tests.

The ICCVAM submission guidelines require the use of controls (specifically positive controls). Performance norms for the positive control are established as part of acceptance criteria for a given "run" of the assay. The acceptable result obtained with the positive control helps to assure that the test system and test execution are functioning properly. While the concept of controls is not new to toxicology, the specification that the controls be performed concurrently with each unknown (or group of unknowns in a single batch) is new and tremendously important. The positive control provides a measure of consistency over time and across laboratories. The MPS documents also identify an important selection criterion for the positive control. The positive control must be able to demonstrate both over and under prediction (sensitivity) relative to the historical performance of the test. Using a 9-pound hammer (i.e., concentrated nitric acid) as a positive control is unlikely to effectively measure assay response. The discussion of benchmark controls (either chemical or formulation) is very helpful. While the positive

control selected for a given assay should remain constant over time, benchmark materials tested concurrently with the unknowns would be selected to match the chemical class of the unknowns. The response of the benchmark controls facilitates interpretation of the results for the unknowns.

Validation studies are complex, time consuming and expensive. They serve to validate the complete test (test system [target tissue], protocol, endpoint measures, and prediction models). The successful validation of the test also tends to validate the mode of action measured by that test. For example, the mode of action for many corrosive chemicals is to penetrate the stratum corneum of the skin and rapidly kill the underlying keratinocytes. Conceptually, it is not hard to imagine modeling such a mode of action with an engineered human skin construct. However, modeling the quantitative (kinetic) aspects of the action is much more difficult. How much test material must be applied and for how long? How to measure the viability of the keratinocytes? How to translate the assay endpoint (e.g., percent viability) to a prediction of corrosive action? The test developer produces an assay protocol to address all of these parameters. The protocol may be based on a proprietary test system (e.g., skin construct) or assay endpoint (e.g., company X's ATP assay). Are those proprietary components of the test absolutely essential or could substantial equivalence be established for another test system or endpoint measure? By identifying the essential structural and functional elements of the test, the MPS approach will allow us (collectively) to draw on validation studies where a successful mode of action has been identified.

There are several additional reasons that the MPS approach is important:

1) In the original ECVAM-sponsored validation of in vitro assays for corrosivity, two skin constructs were tested. At the end of the validation program, neither skin construct was available commercially. Therefore, ZEBET conducted a study to show that the EpiDerm (MatTek, Ashland MA) construct was substantially equivalent to the validated tissue. Thus, the effort and expense of the validation study was not lost.

2) The Organization of Economic Cooperation and Development (OECD) prepares test guidelines for review and acceptance by its 30 member nations (including the United States). Their policy precludes specification of a proprietary test in OECD guidelines. As a result, the OECD has begun to specify structural and functional characteristics of a test (or test system) so that the guideline can draw on validation programs that employ proprietary methods or components.

3) Some proprietary test developers may have made substantial investments in the validation programs for their test. Do the MPS guidelines diminish the economic value of that investment? We believe that they do not. The MPS guidelines provide a controlled mechanism for entry of a new test system or endpoint measure so that the field can grow, but they maintain and codify the standards for that assay. The MPS guidelines assume that the new or modified

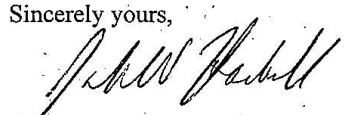
"component" of the test will show substantial equivalence to that component of the validated test. For example, it is reasonable to expect that the substantially equivalent test will use the same prediction model as the validated test. Otherwise, a new set of training test materials will be needed to develop the model. Clearly, one can not use the chemicals provided in the MPS to develop and then validate a prediction model! At some point in the number or degree of changes, a more complete validation of a modified method could be necessary.

Once a new test is accepted for regulatory use, additional laboratories are likely to begin using the method. The MPS documents provide the guidance needed to help demonstrate that the new test is being conducted properly. Successful execution of the test with the reference chemicals will help show that the equipment and reagents used in the new laboratory are within "normal limits" for the assay as it was validated. It will also help assess proper assay execution. The MPS guidelines are not a barrier to entry for a new laboratory but a means to link its performance with that of the validation laboratories. Data developed on unknown test materials would then be more credible for both the producers and users of such data.

Again, the authors of these documents are to be commended for developing the MPS concept and creating the subsequent documents. The format is well designed. I would ask however, that the authors become less prescriptive in their specifications for the report contents. Not every test substance will fit into the box that they have built. Perhaps more of the bullet points could include "if relevant to the conduct of the study". One item missing from the list is designation of the acceptance criteria (i.e, range acceptable positive control responses). For ease, the report section might have its own number (rather than being part of section 3).

The Minimum Performance Standards guidelines are an important step forward and NICEATM and ICCVAM DCIWG deserve a great deal of credit for their contribution.

Sincerely yours,



John W. Harbell, Ph.D.
Chief Scientific Officer

Rodger D. Curren, Ph.D.
President

L'ORÉAL
RECHERCHE
RECHERCHE AVANCÉE
SCIENCES DU VIVANT

Roland Roguet
Business development EPISKIN
Department of reconstructed human skin models
L'ORÉAL RECHERCHE
90, rue du général Roguet
92583 Clichy Cedex France
tel +33 1 47 56 42 98
fax + 33 1 47 56 79 65

Comments on the draft
« ICCVAM Minimum Performance Standards : In Vitro Human Skin Model Systems for Skin Corrosion »

The definition of minimum criteria of biological systems and performances on their uses is an essential guarantee to obtain relevant in vitro data and recognition of the method by regulatory authorities.
The initiative of the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Dermal Corrosivity and Irritation Working Group (DCIWG) corresponds to a real need for academics and industrial companies.

We have some suggestions in order to implement the draft.

The EPISKIN™ model has been validated trough the ECVAM validation process and is on the market today. We suggest **to mention that EPISKIN™, besides EpiDerm™,** has a commercially available models for skin corrosivity assessment, *page 4 line 20 and p7 line 17.* In the same way we suggest to suppress the part of sentence " **and recommended as an alternative to EPISKIN™** *page 3 line 17.*

The ECVAM validation process demonstrated the ability of the EPISKIN™ test to discriminate corrosives from non-corrosives but also its ability to identify correctly known R35/I and R34II&III chemicals. This important advantage as regard to the UN packing groups classification should be mentionned in the draft. The addition of a sentence such as '**including the discrimination between R35 (UN packing group I) and R34 (UN packing groups II&III)** after the citations (2,12,16) *page 3 line 14* would complete the information provided on Table 1 concerning the corrosive subclasses recognized by some authorities.

CENTRE CHARLES ZVIAK
90, rue du Général Roguet - 92583 Clichy Cedex FRANCE
Tél. : (33) 01 47 56 82 00 - Fax : (33) 01 47 56 75 27

L'OREAL
RECHERCHE

Some taping errors have been found in the draft:

- *page 3 line 18*: the **citation 2** should be suppressed, since the ECVAM publication do not mention EpiDerm.
- *page 4 line 21*: citation number **22** instead of 221
- *page 5 line 19*: the exposure period of **4 hours** should be omitted since the draft described the method and predictive model for only two classes (corrosives/non-corrosives)(see prediction model *page 7*)
- *page 6 line 9*: add '**or**' between glacial acetic acid and 8N KOH
- *page 6 line 22,* the word '**Cell**' should be suppressed
- *page 12:* the title of the Table 3 should be corrected as " Accuracy of the Validated Human Skin Model **EPISKIN™** Test Method for Skin Corrosion assessment"

We insist on the relevance of the draft proposed by the ICCVAM Dermal Corrosivity and Irritation Working Group and thank you for the opportunity to comment.
We hope the proposed modifications would help users of the method to better evaluate chemicals.

Sincerely,

Roland Roguet

PEOPLE FOR THE ETHICAL TREATMENT OF ANIMALS

HEADQUARTERS
501 FRONT STREET
NORFOLK, VA 23510
TEL 757-622-PETA
FAX 757-622-0457

August 14, 2003

Dr. William Stokes
Director, NICEATM
National Institute of Environmental Health Sciences
P.O. Box 12233, MD EC-17
Research Triangle Park, NC 27709

Via electronic transmission to: iccvam@niehs.nih.gov

Dear Dr. Stokes:

These comments are submitted on behalf of People for the Ethical Treatment of Animals (PETA) and our more than 750,000 members and supporters in response to a July 1 notice in the *Federal Register* inviting public comment on three sets of "Minimum Performance Standards" for *in vitro* skin corrosivity tests proposed by the Dermal Corrosivity and Irritation Working Group of the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM). We appreciate the work that has gone into the development of these documents and are hopeful that they will not only satisfy the needs of U.S. regulatory agencies, given their inability to lawfully require or recommend use of proprietary test methods, but will also be useful in preventing future bottlenecks in the validation pipeline both domestically and internationally.

PETA is in general agreement with the content of ICCVAM's proposed Minimum Performance Standards, with one notable exception: we strongly disagree with ICCVAM's recommendation that fully-validated *in vitro* human skin model systems (i.e., EpiDerm™ and EPISKIN™) be relegated to the status of merely "positive screens," whereby "substances that are negative *in vitro* might undergo additional testing in accordance with the tiered testing strategy" (*In Vitro* Human Skin Model MPS, p. 3), or, as articulated in ICCVAM's official recommendations to federal agencies: "Negative *in vitro* corrosivity responses shall be followed by *in vivo* dermal corrosion/irritation testing" (66 *Fed. Reg.* 49685).

As you know, both the European Union and the 30-member-country Organization for Economic Cooperation and Development (OECD) have accepted these validated *in vitro* human skin model systems either as stand-alone methods or as part of a purely *non-animal* weight-of-evidence strategy. Given ICCVAM's statutory mandate to promote the replacement, reduction, or refinement of animal-based testing and to strive for the elimination of unnecessary and duplicative efforts (42 *U.S.C.* Sec. 2851-3(b)), we cannot comprehend why ICCVAM persists in advocating a testing paradigm that is so clearly out-of-step with the international consensus on this issue.

It is also worth reiterating a point that was raised several times during the August 12-13 meeting of the National Toxicology Program's Scientific Advisory Committee on Alternative Toxicological Methods: that only a miniscule number (estimates range from two to six percent) of chemicals in commerce today are believed to possess irritating or corrosive properties. Thus, if regulatory agencies adhere to ICCVAM's testing recommendations (i.e., 66 *Fed. Reg.* 49685) and accept *in vitro* skin corrosivity assays as merely "positive screens," only a tiny handful of chemicals would likely be classified on the basis of *in vitro* data, while the overwhelming majority would still be required to undergo animal testing, ostensibly to "confirm" *in vitro* findings of non-corrosivity. From this perspective, ICCVAM's testing recommendations not only squander a golden opportunity for replacement, they promise to be equally meaningless and ineffectual from a reduction standpoint as well.

Even recognizing ICCVAM's stated concern regarding the potential for "false-negative" results *in vitro*, we should not need to remind the committee or its member agencies that the animal-based

reference data against which *in vitro* assays are so often compared have themselves seldom, if ever, been formally validated to demonstrate either their intra- or inter-laboratory reproducibility, much less their relevance to human beings. As just one example, we call your attention to a comparison of data from skin irritation tests on rabbits and skin patch tests on human volunteers for 65 substances, which found that nearly half—fully 45 percent—of classifications of chemical irritation potential based on animal tests were incorrect (MK Robinson et al. *Food Chem Toxicol* 40, 573-592, 2002).

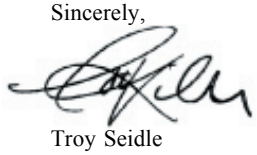
As we have also pointed out in previous correspondence, a 1998 study by Worth and colleagues (*ATLA* 26, 709-720) determined that "false-negative" results from human skin equivalent models ***can be reduced to zero*** when combined with pH measurements and computerized structure-activity relationship modeling. The fact that this study is based on modeling data as opposed to a multi-chemical, multi-laboratory validation exercise should not, in itself, be seen to diminish the significance of the study's findings. Indeed, ICCVAM has already established a precedent for the acceptance of modeling data for validation purposes through its endorsement of the revised Up-and-Down Procedure for acute toxicity, the "validation" of which was based *entirely* on computer modeling.

Nonetheless, if ICCVAM and/or its constituent agencies had lingering doubts regarding the findings of Worth *et al.* (1998), they have had ample opportunity in the more than four years since this study was published to either confirm or refute its assertions. However, to the best of our knowledge, no such study has been undertaken by any ICCVAM member agency, which calls into question ICCVAM's continued resistance to a non-animal weight-of-evidence approach and its inexplicable insistence on "confirmatory" testing *in vivo*. Clearly, the former scenario is not only more humane, but also fully in harmony with the international consensus on this issue—both considerations being directly relevant to ICCVAM's statutory mandate.

With these considerations in mind, we strongly urge ICCVAM to revise its proposed Minimum Performance Standards and testing recommendations for *in vitro* human skin corrosivity systems to bring them into line with international regulations (e.g., EU Annex V) and testing guidelines (e.g., OECD 431).

Thank you for your attention and responsiveness to these comments.

Sincerely,



Troy Seidle
Science Policy Advisor

Federal Institute for Risk Assessment

Bundesinstitut für Risikobewertung

BfR

zebet

ZEBET at the BfR, Diedersdorfer Weg 1, 12277 Berlin, Germany

Zentralstelle zur Erfassung und Bewertung von Ersatz- und Ergänzungsmethoden zum Tierversuch

Centre for Documentation and Evaluation of Alternatives to Animal Experiments

Diedersdorfer Weg 1
D-12277 Berlin
Germany
☎ +49-1888-412-2270
Fax: +49-1888-412-2958
E-mail: zebet@bfr.bund.de

P.O. Box 33 00 13
D-14191 Berlin
Germany

☎: (0)1888-412-0
Fax: (0)1888-412-4741
Internet:
http://www.bfr.bund.de

Dr. William Stokes
Director, NICEATM
NIEHS, MD EC-17
P.O. Box 12233
Research Triangle Park, NC, 27709
USA

your message date & initials	our document number & initials	Phone	Berlin,
	ZEBET-2 / B4122	+49-1888-412- 2275	15 August 2003

Comments on ICCVAM Minimum Performance Standards on three types of *In Vitro* Tests for Skin Corrosion (Federal Register Notice Vol. 68, No. 126 / Tuesday, July 1, 2003, page 39104)

Dear Dr. Stokes

The institutions ZEBET and ECVAM have in 1997 already worked on the concept of a general use of skin models for regulatory toxicology. We have developed test protocols and prediction models that were generally applicable to different commercial skin models. For example, our skin model phototoxicity test developed with the full thickness skin model Skin_ [*Liebsch et al. Toxic. in Vitro* 9, 557 – 562, 1994] could later be applied without any change to the epidermis model EpiDerm [*Liebsch et al. Altex 14: 165 – 174, 1997*], and was just recently successfully applied to the epidermis model SkinEthic [*Jones et al. Toxic. In Vitro 17, 471-480, 2003*]. Taking into account that experience and a comparable experience in the field of skin corrosion tests Michael Balls wrote in 1997 an ATLA editorial about definition of structural and performance criteria (copy enclosed) to facilitate the use of equivalent biological test systems in validated robust test methods. Finally, as you will recall, in the year 2002 we have internationally agreed on that concept in the OECD Workshop on Validation and Acceptance in Stockholm.

With this detailed introduction we want to emphasise that ZEBET very much welcomes the general concept and the definition of Minimum Performance Standards for the future use of "me too" test systems that claim to be equivalent to validated systems. In November 2001 this concept has been intensively discussed in the two OECD Extended Nominated Expert Consultations for the revision of Draft Test Guideline proposals on new Guidelines for Skin Corrosion and Phototoxicity, that finally resulted in accepted new OECD TG 430 and 431 on *Skin Corrosion*, and TG 432 on *Phototoxicity*. The Experts (incl. an ICCVAM representative) defined, for example, in TG 431 functional and performance criteria for new skin models in paragraphs 9, 10 and 11. In addition, 12 Reference Chemicals were defined that should be correctly classified if a new skin model was used or the test protocol modified. The Experts agreed that meeting these criteria is a sufficient proof of equivalency for a new skin model, and this was later confirmed by the National Co-ordinators of the OECD Member Countries. For TG 430 (TER Test), the same Reference Chemicals were defined to address the problem that the TER is sensitive to the rat strain used and the dimensions of the apparatus used. Here the twelve chemicals function as re-calibration chemicals rather than as a confirmation of the usability of the biological test system.

Because international consensus has been reached on OECD Test Guidelines 430 and 431, we welcome that the wording of these Guidelines has been used unchanged also in the ICCVAM MPS documents. **However, ZEBET is opposing the additional mandatory requirement to test a**

Berlin-Dahlem
Thielallee 88 - 92
D-14195 Berlin
Tel.: +49-1888-412-0

Berlin-Marienfelde
Diedersdorfer Weg 1
D-12277 Berlin
Tel.: +49-1888-412-0

Dessau
Jahnstraße 8
D-06846 Dessau
Tel.: +49-340-64000-0

larger set of chemicals with the TER and Skin Model Corrosion Test, since it results in mandatory re-validation of validated methods.

If testing a new skin model or a modified TER technology provides correct and reproducible results for the 12 OECD Reference Chemicals, then there is no need for testing additional chemicals, if we accept the robustness and general applicability of the new corrosion methods.

However, if not all of the 12 OECD Reference Chemicals are correctly classified additional refinement work and additional data is needed (depending on whether it looks promising). In that case, a list of well selected and easily available chemicals like the ones defined in the MPS documents can be very helpful. **We therefore ask ICCVAM to accept the 12 OECD Reference Chemicals* and make it a mandatory requirement. The second set of 12 Test Chemicals should be recommended for test refinement when the 12 OECD Reference Chemicals have not 100% correctly been classified**.

(* ICCVAM has deleted one of the twelve OECD Reference Chemicals (Acrylic Acid) from the list, because this was not included in the ECVAM Validation studies. However, the OECD experts had intentionally selected this chemical as a challenge for the skin model test, because it has a clear *in vivo* database as a strong corrosive.)

To emphasise our statement I can inform you that ZEBET and L'ORÉAL are currently very successfully co-operating on the generation of a common skin model test for *Skin Irritation Testing* that can be applied both to EPISKIN and EpiDerm models and that provides the same results in both models.

We do not comment in detail on the MPS document of the third Skin Corrosion Test (Barrier Test), since the situation is totally different: Because no OECD Test Guideline has been adopted, the ICCVAM MPS on the Barrier Test is not in conflict with international consensus. Moreover, to date the Barrier Method is still more a "black box" than the well validated and characterised skin models. Therefore, we support the definition of a sufficient number of reference chemicals, as suggested by the MPS document.

We do hope ICCVAM re-considers the TER and Skin Model MPS documents accordingly

On behalf of ZEBET

Sincerely yours

Dr. Manfred Liebsch

PS: We would like to put your attention to a few minor points (typos etc.):

Skin Model MPS:

Page 3, 3rd para: Although historically EpiDerm has been validated as an alternative to EPISKIN because it was not available any more, it was the catch up validation concept, only to show that EpiDerm was equivalent to EPISKIN. Delete that sentence, as EPISKIN is available again.

Page 4, 3rd para: Change reference (221) into (22)

Page 6, 4th para: Delete "cell"

Page 10, Table 2: As a strong MTT reducer that accumulates in the tissues n-Heptylamine is now correctly classified in all skin models (including SkiEthic), if the killed tissue control procedure is applied (see paragraph 15 of TG 431 and Liebsch et al ATLA 28, 371-401, 2000)

ENCLOSURE

ATLA **25**, 483–484, 1997

Editorial

Defined Structural and Performance Criteria would Facilitate the Validation and Acceptance of Alternative Test Procedures

The developers of new test procedures tend to want them to be tightly defined, so that they can gain their specific acceptance in the face of real or imagined competition, either for commercial reasons or to ensure that they gain the personal recognition they may deserve. However, it has become clear that this attitude is not in the interests of *in vitro* toxicology in general and may delay, or even prevent, the acceptance and application of scientifically relevant and reliable new approaches.

Three examples will illustrate the point. Firstly, Advanced Tissue Sciences withdrew their reconstituted human skin product, Skin^{2TM}, from the market, *after* it had been accepted by the US Department of Transport as a basis for classifying chemicals in terms of their skin corrosivity. Secondly, the withdrawal of Skin^2 and of EPISKIN^{TM}, a similar product made by Imedex, took place *during* a formal international study on *in vitro* tests for skin corrosivity, funded by ECVAM. Thirdly, Skin^2 was also in the process of being evaluated in the EU/COLIPA international validation study on *in vitro* tests for photoirritancy. As in the case of the withdrawal of a human skin product by Organogenesis a few years earlier, these developments led to annoyance and frustration since, whatever the manufacturers themselves had invested, and while one must sympathise with them, many other companies and laboratories had themselves invested considerable time and effort in evaluating the use of these systems for their own particular purposes. The results they had obtained had been most encouraging, which added to their sense of frustration.

This kind of problem could be avoided if, rather than validating and accepting particular kinds of commercial products, or methods involving particular cell lines, endpoints or endpoint assays, clearly laid down structural and performance criteria were to be defined and agreed for test systems to be used for particular purposes, then themselves subjected to prevalidation and formal validation. Any new test system which could meet these criteria would then be considered to be scientifically valid and acceptable, albeit after a small and independent confirmatory study in some circumstances.

It is for this reason that ECVAM and ZEBET are supporting studies on the applicability for *in vitro* corrosivity and photoirritancy testing of another human reconstituted human skin equivalent, EpiDerm^{TM}, made by MatTek, which, happily, promises to survive longer than its competitors. We are using our experience with Skin^2 and EPISKIN to speed up the acceptance of EpiDerm, not because we have any particular interest in MatTek or its products, but because we do not want much valuable experience to be wasted or the undoubted promise of this kind of test system to be lost.

At the same time, in order to provide one possible route of escape from the current impasse in the case of the acceptance of *in vitro* systems for percutaneous absorption, ECVAM has commissioned a study to define the structural and performance criteria which would be needed in such systems. Clearly, the structural characteristics required would include an effective barrier sufficiently similar to that found in the skin *in vivo*, and the performance criteria would include an ability to prevent the passage of certain standard test materials, while permitting the passage of others. Ideally, some of the *in vitro* systems should have the capacity to metabolise those kinds of test materials which would be likely to be metabolised by the human skin *in vivo*.

This having been done, ECVAM would be willing to support a prevalidation/validation study on *in vitro* systems which might meet the structural and performance criteria defined for percutaneous absorption testing.

This approach could be linked to the benchmarking concept as a possible route of escape from another impasse, namely, the absence of sufficient chemicals representative of the spectrum of chemicals to be tested in terms of type and scale of toxicity, backed by knowledge of sufficiently high quality. For example, an appropriate, and relatively small, set of standard materials which met *these* criteria, could be used to provide a standard curve, not only to establish the performance of the system on a particular occasion, but also as a means of expressing the result of the test on a novel material under investigation.

The structural and performance criterion approach could be taken further, since new tests could be developed to provide knowledge which is needed (i.e. to provide what Björn Ekwall has called "missing tests"). For example, it would be much more intelligent to devise realistic new tests for identifying *human* carcinogens, rather than merely speeding up the rodent bioassay or finding alternative methods for identifying chemicals which might be carcinogenic at high doses in *rodents*.

Michael Balls

This Page Intentionally Left Blank

APPENDIX C

Minutes of the EPA Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) Meeting (October 28 and 29, 2003)

This Page Intentionally Left Blank

SAP Minutes No. 2003-03

October 28 and 29, 2003
FIFRA Scientific Advisory Panel Meeting,
Held at the Holiday Inn Hotel,
Arlington, Virginia

A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding:

Ensuring Data Quality for In Vitro Tests Used as Alternatives to Animal Studies for Regulatory Purposes: A Consultation

Myrta R. Christian, M.S.
Designated Federal Official
FIFRA Scientific Advisory Panel
Date: January 23, 2004

Steven G. Heeringa, Ph.D.
FIFRA SAP, Session Chair
FIFRA Scientific Advisory Panel
Date: January 23, 2004

NOTICE

These meeting minutes have been written as part of the activities of the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP). These meeting minutes represent the views and recommendations of the FIFRA SAP, not the United States Environmental Protection Agency (Agency). The content of these meeting minutes do not represent information approved or disseminated by the Agency. These meeting minutes have not been reviewed for approval by the Agency and, hence, the contents of this report do not necessarily represent the views and policies of the Agency, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names or commercial products constitute a recommendation for use.

The FIFRA SAP is a Federal advisory committee operating in accordance with the Federal Advisory Committee Act and was established under the provisions of FIFRA, as amended by the Food Quality Protection Act FQPA of 1996. The FIFRA SAP provides advice, information, and recommendations to the Agency Administrator on pesticides and pesticide-related issues regarding the impact of regulatory actions on health and the environment. The Panel serves as the primary scientific peer review mechanism of the EPA, Office of Pesticide Programs (OPP) and is structured to provide balanced expert assessment of pesticide and pesticide-related matters facing the Agency. Food Quality Protection Act Science Review Board members serve the FIFRA SAP on an ad hoc basis to assist in reviews conducted by the FIFRA SAP. Further information about FIFRA SAP reports and activities can be obtained from its website at *http://www.epa.gov/scipoly/sap/* or the OPP Docket at (703) 305-5805. Interested persons are invited to contact Larry Dorsey, SAP Executive Secretary, via e-mail at *dorsey.larry@.epa.gov.*

In preparing these meeting minutes, the Panel carefully considered all information provided and presented by the Agency presenters. This document addresses the information provided and presented within the structure of the charge by the Agency.

CONTENTS

October 28 and 29, 2003

Ensuring Data Quality for In Vitro Tests Used as Alternatives to Animal Studies for Regulatory Purposes: A Consultation

PARTICIPANTS

FIFRA SAP, Session Chair
Steven G. Heeringa, Ph.D., Research Scientist & Director for Statistical Design, University of Michigan, Institute for Social Research, Ann Arbor, MI

Designated Federal Official
Mrs. Myrta R. Christian, FIFRA Scientific Advisory Panel Staff, Office of Science Coordination and Policy, EPA

FIFRA Scientific Advisory Panel Members
Stephen M. Roberts, Ph.D. (FIFRA SAP Chair), Professor & Program Director, University of Florida, Center for Environmental & Human Toxicology, Gainesville, FL

Stuart Handwerger, M.D., Director, Division of Endocrinology, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati, OH

Gary E. Isom, Ph.D., Professor of Toxicology, School of Pharmacy and Pharmacological Sciences, Purdue University, West Lafayette, IN

Mary Anna Thrall, D.V.M., Professor, Department of Pathology, College of Veterinary & Biomedical Sciences, Colorado State University, Fort Collins, CO

FQPA Science Review Board Members
Joseph P. Bressler, Ph.D., Department of Neurobiology, The Kennedy Krieger Research Institute, Baltimore, MD

George L. DeGeorge, Ph.D., Chief Scientific Officer, MB Research Laboratories, Spinnerstown, PA

James Freeman, Ph.D., ExxonMobil Biomedical Sciences, Inc., Annandale, New Jersey

Michael Luster, Ph.D., Chief, Toxicology and Molecular Biology Branch, National Institute for Occupational Safety and Health, Morgantown, WV

James F. McCormack, Ph.D., Director of Nonclinical Laboratory Compliance, Food & Drug Administration, Rockville, MD

Kenneth Portier, Ph.D., Associate Professor of Statistics, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL

Gregory A. Reed, Ph.D., Associate Professor, Department of Pharmacology, Toxicology, and Therapeutics, University of Kansas Medical Center, Kansas City, KS

INTRODUCTION

The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP) has completed its review of the set of scientific issues being considered by the Agency pertaining to the processes for regulatory acceptance of and ensuring the quality of data from *in vitro* tests used as alternatives to animal studies for regulatory purposes. Advance notice of the meeting was published in the *Federal Register* on September 22, 2003. The review was conducted in an open Panel meeting held in Arlington, Virginia, on October 28 and 29, 2003. Dr. Steven G. Heeringa chaired the meeting. Mrs. Myrta R. Christian served as the Designated Federal Official.

The FIFRA SAP was asked to review issues concerned with processes for regulatory acceptance of and ensuring the quality of data from *in vitro* tests used as alternatives to animal studies for regulatory purposes, including performance standards, essential test method components, and quality control of test methods, in the context of three new *in vitro* assays for dermal corrosivity which will be incorporated into its OPTS 870.2500 test guideline for Acute Dermal Irritation.

In preparing these meeting minutes, the Panel carefully considered all information provided and presented by the Agency presenters, as well as information presented by public commenters. These meeting minutes address the information provided and presented at the meeting, especially the response to the charge by the Agency.

CHARGE

Performance Standards

The Agency plans to adopt the Performance Standards developed by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) as a means of communicating the basis by which each of three validated *in vitro* test methods, Corrositex®, EPISKIN™/EpiDerm™, and Transcutaneous Electrical Resistance (TER), are deemed acceptable for providing dermal corrosivity data. Performance Standards consist of descriptions of (1) essential test method components, which are the essential structural, functional, and procedural elements of a validated test method that should be included in the protocol of a proposed mechanistically and functionally similar test method; (2) a minimum list of Reference Chemicals, which is used to assess the accuracy and reliability of the similar test method; and (3) comparable accuracy and reliability values that should be achieved by the proposed test method when evaluated using the minimum set of Reference Chemicals.

Question 1

Please comment on the provisions in the Performance Standards for each of the three methods to demonstrate mechanistic similarity of "me-too" methods. Do the essential test method components for each method adequately describe the unique characteristics of the method necessary to determine whether a test is mechanistically and functionally similar?

Question 2

In its evaluation of any mechanistically similar test system, the Agency plans to use the generic criteria used by ICCVAM for selecting subsets of the Reference Chemicals for all three ICCVAM Performance Standards documents. The criteria specify that chemicals should be selected in such a way that the subset: includes representatives of applicable chemical classes, measures a range of corrosive strengths, includes well-defined chemicals that are currently available commercially, and has unequivocal animal or other *in vivo* evidence. Please comment on the strengths or weaknesses of this approach and identify and discuss any modifications to the criteria that should be considered.

Question 3

The ICCVAM approach for demonstrating functional similarity of "me-too" test methods to validated methods includes the use of well-characterized Reference Chemicals and specifies the accuracy and reliability that should be achieved by "me-too" test systems when tested in intra- and inter-laboratory studies. Please comment on whether "me-too" test systems should be demonstrated to be effective for evaluating the testing endpoint for all of the chemicals in the Performance Standard. Please comment on the value of including chemicals with range of potencies in the Performance Standard. Under what circumstances might testing of "me-too" systems within one laboratory ever be sufficient to demonstrate functional equivalence?

Quality Control

The Agency is proposing quality control measures that should be considered when evaluating the reliability of test kits for regulatory purposes. Please address the following specific issues.

Question 4

Subsets of the Reference Chemicals used in test method validation may be used as training or calibration sets by testing laboratories using *in vitro* systems. Please discuss the utility of and necessity for training or calibration sets in assuring data quality. Please comment on the chemicals selected by ICCVAM for use as a calibration set for TER for this purpose. Please comment on the ranges of chemical classes and potencies of these chemicals. How might other chemicals be selected for possible use in the calibration sets? Please comment on the value of identifying chemicals that might be used by laboratories as training sets to demonstrate proficiency in performing the test.

Question 5

Anticipating the use of systems using tissue constructs, *ex vivo* systems, microarrays or genetically modified cells, please discuss aspects of the quality control criteria that are necessary for assuring the integrity of such systems over time and from lot-to-lot. Please comment on whether and how the type of system - tissue constructs, *ex vivo* systems, or genetically modified cells or animals - should affect the criteria for quality control for assuring the integrity of such systems, both over time and from lot-to-lot.

Question 6

Please comment on the advantages and disadvantages of including concurrent positive and negative controls with *in vitro* assays when used as alternatives to animal testing. What are the important characteristics of positive and negative controls for *in vitro* studies? What aspects of positive control characteristics allow them to be used as part of the quality control process? When might confirmation that positive controls are performing within expected or historical limits be sufficient to demonstrate that the Proprietary Test Method or non-proprietary assay system is functioning properly? When might additional quality control measures be needed?

Question 7

Does the Panel agree that the benchmark controls serve a useful purpose to demonstrate the level of response that can be expected for each chemical class for each lot of Proprietary Test Method assays? Can the Panel suggest criteria for choice of appropriate benchmark controls?

PANEL DELIBERATIONS AND RESPONSE TO CHARGE

The specific issues addressed by the Panel are keyed to the Agency's background documents, and the Agency's charge questions.

Response to Charge

I. Performance Standards

The Agency plans to adopt the Performance Standards developed by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) as a means of communicating the basis by which each of three validated *in vitro* test methods, Corrositex®, EPISKIN™/EpiDerm™, and Transcutaneous Electrical Resistance (TER), are deemed acceptable for providing dermal corrosivity data. Performance Standards consist of descriptions of (1) essential test method components, which are the essential structural, functional, and procedural elements of a validated test method that should be included in the protocol of a proposed mechanistically and functionally similar test method; (2) a minimum list of Reference Chemicals, which is used to assess the accuracy and reliability of the similar test method; and (3) comparable accuracy and reliability values that should be achieved by the proposed test method when evaluated using the minimum set of Reference Chemicals.

Question 1

- Please comment on the provisions in the Performance Standards for each of the three methods to demonstrate mechanistic similarity of "me-too" methods. Do the essential test method components for each method adequately describe the unique characteristics of the method necessary to determine whether a test is mechanistically and functionally similar?

Panel's comments:

The Panel endorsed the Performance Standards (PS) approach to identify and validate "me-too" and "unique" *in vitro* assays. The following paragraphs summarize the Panel's response for each of the three major components of the ICCVAM performance standards for *in vitro* tests.

Structural/functional components:

The Panel concurred that the PS prepared by ICCVAM are very well described for each of the three tests, and the information should provide a basis to determine whether a test is mechanistically and functionally similar to a validated *in vitro* test method. The Panel stated that it would be helpful for the submitting laboratories if the Agency provided examples of what they would consider as a "me-too" assay or a new assay, based upon the essential structural and functional elements (e.g., human skin TER vs. rat skin TER). There was some concern among the Panel members that identification of a "me-too" assay could be a somewhat subjective process rather than one based entirely on objective criteria. However,

with the limited tests that have been evaluated to date (one "me-too" and three unique), there was consensus that this approach of using structural and functional equivalence to determine a "me too" test is conceptually feasible.

Reference Chemicals:

The Panel recommended that NIEHS/EPA (thru ICCVAM) develop a standard list of reference chemicals for validating *in vitro* tests and establish a chemical repository for reference samples/positive controls available to laboratories for developing/conducting *in vitro* skin studies. The reference panel should contain sufficient numbers of different chemical classes (with a range of potency, solubility, etc.) to establish reasonable performance of that specific test for those particular classes of chemicals.

For the three validated test methods, members of the Panel recommended that the laboratories be allowed to determine their own positive control(s) and suggested that the PS not suggest specific examples such as NaOH pellets and 10 N HCl. The Panel felt that these particular examples may be too corrosive, and if suggested by the Agency as a positive control, could become the "gold standard." In lieu of citing specific examples for positive controls, the Panel suggested that the Agency PS provide general requirements (e.g., well characterized, results in a low-to-intermediate response, etc.) wanted in a positive control for a validated test.

For all three validated tests, the Agency PS would benefit from a more thorough discussion of appropriate benchmark controls (range of severity, classes of chemicals) and also how benchmark controls would be considered in the validation studies of the assay. The Panel also recommends that minimum replicate requirements be specified for positive, negative and benchmark controls, and that the PS be unambiguously stated.

Concordance and reliability values:

The Panel suggested that the Agency provide clear guidance on requirements necessary to establish test reliability for the PS for each validated *in vitro* test (how many labs for the inter-laboratory reliability and how many intra-laboratory replications?). The Panel also recommended that the Agency better define what is meant by comparable concordance for test accuracy – will this be statistically based? The Panel expressed the view that the PS should include specific guidelines for minimum achieved sensitivity and specificity of the test when applied to the reference chemical set.

One Panel member expressed the view that if there is no appreciable difference in performance, an *in vitro* assay should be recommended as the preferred alternative testing method for use over an ex-vivo assay (e.g., rat skin TER) as the former more directly addresses the goal of animal replacement.

Question 2

- In its evaluation of any mechanistically similar test system, the Agency plans to use the generic criteria used by ICCVAM for selecting subsets of the Reference Chemicals for all three ICCVAM Performance Standards documents. The criteria specify that chemicals should be selected in such a way that the subset includes representatives of applicable chemical classes, measures a range of corrosive strengths, includes well-defined chemicals that are currently available commercially, and has unequivocal animal or other *in vivo* evidence. Please comment on the strengths or weaknesses of this approach and identify and discuss any modifications to the criteria that should be considered.

Panel's comments:

The Panel expressed the view that the strength of the PS approach to validating a new or "me too" *in vitro* test derives from the stated selection criteria for the Reference Chemical set. By including a range of chemical classes in the Reference Chemical set the general applicability of the test is supported. Choosing Reference Chemicals exhibiting a broad range of corrosive strengths provides insight into the quantitative value of the test. This could be important for assignment of corrosive agents to packing groups. In addition, the inclusion of mildly corrosive agents supports estimation of the sensitivity of the test. The use of well-defined agents with unequivocal animal or other *in vivo* evidence in regard to skin corrosivity anchors the Reference Chemicals as valid "real world" representatives and allows for validated comparisons between the *in vitro* findings and the potential effects of actual environmental or occupational exposures. Limiting the Reference Chemical set to commercially available chemicals allows for the widespread use of this testing regimen.

The Panel identified a weakness of the approach in that it may be difficult to include a sufficient number of Reference Chemicals in each class, both corrosive and non-corrosive, which meet all of these criteria. The Episkin/Epiderm Reference Chemical set comes closest, with 6 of 8 classes containing both corrosive and non-corrosive agents. Although numerous classes of potentially corrosive chemicals are included in the various Reference Chemical sets, some classes are missing. This includes inorganic salts, such as FeCl3, which was reported by ECVAM to be corrosive. Also the Panel noted that hydrocarbons and halogenated hydrocarbons are common solvents and diluents for pesticides, and that these chemicals might be included for study either as individual agents or in combination with other chemicals. The question of how many "classes" the test methods (or "me-too" tests) are validated with, versus the number of classes which the test may be approved for, remains unanswered.

The Panel pointed out that a second weakness of the PS Reference Chemical descriptions for the validated in vitro tests is the lack of standardization of the list. Different groups of specific chemical agents are employed (or recommended) for the different *in vitro* tests. While this may not affect the validation of individual test systems, it does impact on comparisons between the available and proposed test systems.

Question 3

- The ICCVAM approach for demonstrating functional similarity of "me-too" test methods to validated methods includes the use of well-characterized Reference Chemicals and specifies the accuracy and reliability that should be achieved by "me-too" test systems when tested in intra- and inter-laboratory studies. Please comment on whether "me-too" test systems should be demonstrated to be effective for evaluating the testing endpoint for all of the chemicals in the Performance Standard. Please comment on the value of including chemicals with range of potencies in the Performance Standard. Under what circumstances might testing of "me-too" systems within one laboratory ever be sufficient to demonstrate functional equivalence?

Panel's comments:

The Panel agreed that a minimum number of Reference Chemicals (subset of the entire list) should be specified in the PS, to be used for validation procedures of existing alternative test methods, as well as "me-too" tests. It was noted, for example, that there was a large range in the number of Reference Chemicals used among the three test systems presented, with a low of 24 reference chemicals, depending upon the test method under consideration. Although the use of the entire original Reference Chemical set for a validated test method for validation of a "me-too" test might be considered excessive, it is nonetheless important to carry out a sufficiently broad characterization of a new test to validate its performance.

The approach of specifying a known level of accuracy and reliability for a "me-too" test to be considered equivalent to the validated test system was accepted by the Panel. Panel members suggested that Reference Chemicals be limited to those that have been tested with sufficient replication, such that the reliability and accuracy estimates themselves are considered sufficiently precise. The Panel recommended that the concordance of results from "me-too" tests be established by comparison to the unequivocal properties of the test chemicals in human or animal tests, rather than by comparison to an alternative test method. It was recognized that other alternative tests may have less than 100% accuracy (sensitivity and specificity) that would cloud the meaning of "me-too" test "accuracy" or concordance.

One Panel member considered it essential that if as few as 24 or less chemicals are specified in the PS then 100% concordance with *in vivo* test results should be required to demonstrate test equivalence and assure the public safety. Lower percentage concordance would be acceptable if a large enough subset of the Reference Chemicals were tested so as to include more than one chemical from all classes originally validated, with a range of potencies or responses for each class. In the case of the Corrositex validation, a minimal set of 40 reference chemicals were used, resulting in a 25% false positive and 11% false negative rate (Table 2 and Table 3, Section 4.0 and 5.0, respectively, of "ICCVAM Performance Standards: In Vitro Membrane Barrier Test Systems for Skin Corrosion," ICCVAM-DCIWG Proposed MPS; June 23, 2003).

While recognizing that the validated test provides the history (that is, the empirical criteria for acceptable sensitivity, selectivity, etc.), it remains questionable whether this is an

appropriate "bright-line." There may be important statistical or practical considerations to the choice of the subset of Reference Chemicals to be included in the PS. One Panel member queried whether the decision point for qualitative judgment of corrosive agents is sensitive enough to detect even weakly corrosive agents, stating that the judgment of sensitivity cannot be made without validation using known weakly corrosive agents. Thus, the "Performance Standard" should include: 1) a stated minimum number of diverse test chemicals, from all relevant chemical classes; (2) a requirement for Reference Chemicals with varying potencies, efficacies, or range of response, ideally within each chemical class; and 3) minimum standards for reliability and accuracy/concordance in the "me-too" test system results when compared to the known properties of the test chemicals for *in vivo* tests.

A majority of the Panel agreed that validation of a "me-too" test in a single laboratory should be acceptable, if that single laboratory is the only practitioner of the method. The criteria for acceptance should be as rigid as that for a multi-laboratory validation. This would involve at least a sufficient number of independent, repeated tests using the Reference Chemicals to establish the concordance of the "me too" test with a validated test, and to determine the intra-laboratory test reliability of the "me too" test.

The Panel noted the importance of using good experimental design in intra- and inter-laboratory studies, being concerned that there was little discussion of batch-to-batch (or pelt-to-pelt in the case of TER) variability in any of the test method protocols, data, or results. The implication is that this is a very small source of variability for these test systems, which may not be the case in future systems. The general procedures for evaluating "me-too" systems should take this into account.

II. Quality Control

The Agency is proposing quality control measures that should be considered when evaluating the reliability of test kits for regulatory purposes. Please address the following specific issues.

Question 4

- Subsets of the Reference Chemicals used in test method validation may be used as training or calibration sets by testing laboratories using *in vitro* systems. Please discuss the utility of and necessity for training or calibration sets in assuring data quality. Please comment on the chemicals selected by ICCVAM for use as a calibration set for TER for this purpose. Please comment on the ranges of chemical classes and potencies of these chemicals. How might other chemicals be selected for possible use in the calibration sets? Please comment on the value of identifying chemicals that might be used by laboratories as training sets to demonstrate proficiency in performing the test.

Panel's comments:

Given the nature of these *in vitro* systems, particularly in regard to lot-to-lot and day-to-day variability, the Panel felt it essential that test system performance be established and understood. A simple positive and negative control may not be sufficient to represent the

range of responses and the sensitivity required for detection of weakly corrosive agents. In the case of the TER test the twelve Calibration Chemicals suggested by ICCVAM meet the criterion of including strongly and weakly corrosive and non-corrosive agents. However, 12 chemicals constitute a limited test set. It also is incomplete; missing are potentially corrosive inorganic salts like $Fe(Cl)_3$, which is noted in the 60 chemical ECVAM list. Further, the ECVAM "60" list does not completely reflect the classes of chemicals that are important with regard to pesticide registration. Hydrocarbon solvents, for example, find use as diluents but are not included in the list. While most of these solvents are complex mixtures, toxicity profiles can be established both for the mixture and for suitable single-chemical surrogates (e.g., toluene, decane, etc). Clearly a balance must be struck between maintaining a manageable number of Reference Chemicals and assuring that all relevant mechanistic and chemical classes are included.

While the background documents discuss the need for a range of potencies for chemicals, it is important that Reference Chemicals that represent a range of implementation difficulties be included as well. Part of the calibration process for testing laboratories is that the technicians learn to be consistent in application so that reproducible results will be obtained for the Reference Chemicals over time. The potency of a chemical may not be the best measure of how difficult it is for a technician to get consistent results with that chemical. The Reference Chemical set should include some chemicals that are difficult to work with, thereby challenging the technical skill of the staff and forcing them to "stay skilled." Further, some chemicals (e.g., solvents) may destroy the test system; knowledge of this is important if such a chemical is tested in a formulated product.

The Panel noted that training in the use of the validated test is required to be documented under GLPs, presumably with Reference Chemicals. One panel member expressed caution regarding the use of the terms proficiency and calibration set. Proficiency implies a precision and accuracy as may be required by independent accreditation. The training to meet this objective is a laboratory management function. The term, "calibration set" implies traceability to some standard, e.g., a national standard. In the context of the Panel's discussion, Reference Chemicals are identified that can be used as control or benchmark chemicals to help standardize or validate a method in a laboratory and monitor its performance but may not, in the strictest sense, be a true calibration of the test results.

For a training set of chemicals to be used either initially or at some set intervals for the validation of an assay and its performance in a given laboratory, this balance between number of chemicals and inclusivity shifts to a higher number of individual chemicals. Whereas twelve might be an appropriate number for regular "calibration" a training and validation set could easily be 2-3 times this number. This would ensure coverage of relevant classes and potencies for corrosive agents and better test the abilities of a given laboratory to perform the assay accurately.

The Panel expressed the view that Reference Chemical testing:

- Provides relevant training and documentation of training as required by GLPs;
- Provides a means to evaluate technician competency for the test method;

- Permits comparison to a validation database and assessment of variability among labs;
- Identifies relative strengths/weaknesses of the lab and whether additional training is needed.

Question 5

- Anticipating the use of systems using tissue constructs, *ex vivo* systems, microarrays or genetically modified cells, please discuss aspects of the quality control criteria that are necessary for assuring the integrity of such systems over time and from lot-to-lot. Please comment on whether and how the type of system - tissue constructs, *ex vivo* systems, or genetically modified cells or animals - should affect the criteria for quality control for assuring the integrity of such systems, both over time and from lot-to-lot.

Panel's comments:

The use of PS, positive controls, negative controls and benchmark controls will provide the opportunity to achieve a degree of control over the quality of Proprietary Test Methods (PTMs). Two issues that have not been addressed in the PS are how drift in the PTMs will be monitored and how information about problems that arise from the use of these controls will be assimilated and evaluated by the vendor. Individual test facilities may detect failures or out-of-specification performance of the PTM and proceed according to their operating procedures, but the lack of GMP-like regulatory authority does not require these failures to be reported to and addressed by the vendor.

Other facilities may then use an inadequate/under-performing PTM or lot of PTM without benefit of the experiences of the first facility. There should be some consideration that PTM performance reports be compiled by the vendor and reported to purchasers of the PTM. Similar mechanisms are used by computer software vendors to alert purchasers of their products of problems or issues with their products.

The answer to the second part of this question goes beyond the immediate concerns of the Panel, which were *in vitro* tests for corrosive chemicals. Rather, the answer discusses general considerations for future *in vitro* tests that will incorporate the newest advances that are being made in molecular biology. All testing systems require quality control for assuring reproducibility, sensitivity, and specificity. Otherwise, results from the same test repeated in the same laboratory, or in different laboratories, could not be compared. Incorporating positive and negative controls, as well as benchmark samples, monitors quality control. The specific types of controls, the number of controls, the frequency of inserting these controls and the benchmark samples, however, will likely be different for different types of assays. The number of controls would be expected to increase in highly variable systems (e.g. those that require animals) but must be limited because of cost considerations. Hence, the development of newer testing systems that limit variability would have substantial benefit.

An example of the concern for variability is an *ex vivo* system, in which tissue is excised from a donor and cultured as either organ culture, explants, or dissociated cells. There will

be variability in each type of culture because of variability in the donor animals. In primary cultures, however, the variability can be greatly limited if large batches of cells are prepared from several animals and frozen. New testing systems for screening different types of toxic chemicals will likely be developed using genetically modified cell lines. The Ames assay is one example of an already established test that uses genetically modified bacteria to screen for mutagens, which are possible carcinogens. A more complex test system could be developed to establish tissue constructs. For example, a testing system might be developed that uses genetically modified skin stem cell lines that differentiate into skin. The currently available tissue construct uses skin epithelial cells from donors. The advantage of the stem cell line is that the lot-to-lot variability would be reduced because the source of variability, donor tissue, would be reduced.

Microarrays (gene arrays) are powerful endpoint assays that measure changes in the expression of hundreds or thousands of genes and will likely be used in different types of testing systems. One such use would be in classifying xenobiotics according to the patterns of genes that they induce. The pattern of gene expression has been termed a gene fingerprint, and testing systems might be developed for screening xenobiotics by measuring gene fingerprints. In measuring gene fingerprints, rather than one or two specific genes, the testing system has more power for statistical analysis and will likely produce more consistent data. Gene arrays also have the potential of reducing the number of required controls. For example, testing systems for determining gene fingerprints for xenobiotics must use cell lines that express enzymes that metabolize xenobiotics. Positive controls should be incorporated in the testing systems for screening xenobiotics to validate the presence of these enzymes. In using gene arrays, the positive controls might not be necessary because the expression of the activating enzymes, as well as the gene fingerprints, would be determined in the same gene array.

The Panel noted that microarrays and other related systems seem to have a long way to go toward producing reproducible responses among true replicates. In fact, very little true replication is being done, primarily due to the expense of each replicate. As the state of the art in microarray use becomes mature, true replication with demonstrated repeatability may become the standard. When that is truly the case, test systems based on this technology should provide useful tools for risk evaluations. As these new tests are put into practice, more attention must be focused on how drift in performance test standards will be monitored and how information about these problems will be assimilated and evaluated by the vendor.

Question 6

- Please comment on the advantages and disadvantages of including concurrent positive and negative controls with *in vitro* assays when used as alternatives to animal testing. What are the important characteristics of positive and negative controls for *in vitro* studies? What aspects of positive control characteristics allow them to be used as part of the quality control process? When might confirmation that positive controls are performing within expected or historical limits be sufficient to demonstrate that the Proprietary Test Method or non-proprietary assay system is functioning properly? When might additional quality control measures be needed?

Panel's comments:

The Panel commented that insufficient controls may preclude meaningful interpretation of *in vitro* test results. Despite the fact that positive and negative controls are not often used in *in vivo* studies, they are routinely included in *in vitro* studies and it is clearly advantageous and desirable that they be used in the test systems being discussed here. Positive and negative (and vehicle) controls provide needed checks within a study that tell the investigator that the test system appears to be intact and functional. Positive controls help identify performance variability between technicians, between laboratories and between lots of test system. Appropriate controls will likely be needed for some length of time until the Agency and practitioners are satisfied with the performance of the test over time and across laboratories. From a quality control perspective, temporal monitoring of controls across studies and laboratories will help establish consistency of response for the test system.

The Panel again pointed out that a single positive control per assay may not be sufficient and that it may be desirable to include positive control chemicals for one or more of the classification severities. At least one Panel member queried as to what actions should be taken when a negative control produces a positive response or a positive control produces a negative response. The answer will depend on the degree of replication assigned to controls and the specified minimum accuracy or concordance for the test system. Clearly, the developers of the test system should incorporate into the recommended protocols guidance on the degree of replication needed for controls, and what actions should be taken when unexpected results are observed with controls. The degree of replication should be based upon the expected variability and the levels of specificity and sensitivity displayed by the test systems for the Reference Chemicals used as controls. With adequate replication, the fact that positive controls (and negative and vehicle controls, for that matter) are performing within expected limits should be sufficient to demonstrate that the test system is functioning properly.

Question 7

- Does the Panel agree that the benchmark controls serve a useful purpose to demonstrate the level of response that can be expected for each chemical class for each lot of Proprietary Test Method assays? Can the Panel suggest criteria for choice of appropriate benchmark controls?

Panel's comments:

The Panel agrees that benchmark controls are an important mechanism to assess both the adequacy of the method as well as lot-to-lot variability and should be considered as a standard component of these test methods. Benchmark controls, as well as positive and negative controls, should be tested in each new lot to determine the viability and usability of each lot. Control charts could assess variability among lots, and provide a basis for acceptance/rejection. The Panel suggests that benchmark controls include several "classic" responders from different chemical classes/mode of actions.

Variability between different lots is a major concern of the Panel and must be assessed with negative and positive controls as well as benchmark samples. The number of controls and samples depends on several factors; many of which will be defined by the specific test. Certain tests are very consistent and require fewer positive controls and benchmark samples for assessing lot-to-variability whereas other tests are less consistent. The Panel agreed that there is concern regarding whether the lots are large enough to accommodate these types of controls. To address this concern, the Panel suggests that EPA establish the necessary controls and benchmark samples in the individual tests and consults with the manufacturer of the test. Accordingly, the manufacturer would be encouraged to change production so that the size of lots are sufficient for allowing adequate controls and benchmarks.

Appropriate benchmark controls should have the following properties:

- Consistent and reliable source(s) for the chemical
- Structural and functional similarity to the class of article being tested
- Known physical/chemical characteristics
- Supporting data on known effects in animal models
- Known potency in the range of response (including moderate response)

One Panel member stated that benchmark controls can serve a very useful purpose, especially in the situation where the test system demonstrates significant batch-to-batch variability in response. But this variability has not been directly addressed for the test systems being discussed here. If we assume that such variability is quite low, the benefit of re-running benchmark controls for each batch is reduced. In this case, the use of benchmark controls might be relegated to a supplier QC role with periodic running of benchmark chemical to ensure continued consistency of response over time. On the other hand, if the test system does demonstrate significant batch-to-batch variability, it would be important to run benchmark controls more often. Finally, it would seem that benchmark controls would be more important in calibrating a formal dose response model. The need for these controls then depends on the level of precision needed in the final model.

REFERENCES

1. Interagency Coordinating Committee for Validation of Alternative Methods. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised and Alternative Test Methods. NIH Publication 03-4508. September 2003.

2. Interagency Coordinating Committee for Validation of Alternative Methods. 2003. ICCVAM Minimum Performance Standards: In Vitro Membrane Barrier Test Systems for Skin Corrosion.

3. Interagency Coordinating Committee for Validation of Alternative Methods. 2003. ICCVAM Minimum Performance Standards: In Vitro Human Skin Model Systems for Skin Corrosion.

4. Interagency Coordinating Committee for Validation of Alternative Methods. 2003. ICCVAM Minimum Performance Standards: In Vitro Skin Transcutaneous Electrical Resistance (TER) Tests for Skin Corrosion.

5. Interagency Coordinating Committee for Validation of Alternative Methods. 2003. Attachment to memorandum from ICCVAM to OECD dated February 19, 2003. Justification for Development of An OECD Guidance Document on the Application of Good Laboratory Practice (GLP) Principles to *In Vitro* Testing.

6. Interagency Coordinating Committee for Validation of Alternative Methods. 1997. Validation and Regulatory Acceptance of Toxicological Methods of Test Methods. NIH Publication 97-3981. March 1997. See Chapter 2. Validation of Test Methods and Chapter 3. Regulatory Acceptance of Toxicological Test Methods.

7. Interagency Coordinating Committee for Validation of Alternative Methods. 1999. Corrositex®: An *In Vitro* Test Method for Assessing Dermal Corrosivity Potential of Chemicals. NIH Publication No. 99-4495. June 1999.

8. Interagency Coordinating Committee for Validation of Alternative Methods. 2002. ICCVAM Evaluation of EISKIN™, 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. June 2002. *http://iccvam.niehs.nih.gov/methods/epiddocs/cwgfinal/cwgfinal.pdf.*

9. Hartung, T., Balls, M., Bardouille, C., Blanck, O., Coecke, S., Gstraunthaler, G., and Lewis, D. (2003). Good Cell Culture Practice: ECVAM Good Cell Culture Practice Task Force Report 1. ATLA 30: 407-414.

10. Cooper-Hannan, R., Harbell, J.W., Coecke, S., Balls, M., Bowe, G., Cervinka, M., Clothier, R., Hermann, F., Klahn, L.K., de Lange, J., Lievsch, M. and Vanparys, P. (1999). The principles of Good Laboratory Practice: Application to *in vitro* toxicology studies. The report and recommendations of ECVAM workshop 37. ATLA 27, 539-577.

11. OPPTS Harmonized Test Guideline 870.2600 Acute Dermal Irritation.

12. Environmental Protection Agency. 40 CFR 160 Good Laboratory Practice Standards (Office of Pesticides) July 1, 2003.

13. Environmental Protection Agency. 40 CFR 792 Good Laboratory Practice Standards (Office of Prevention and Toxic Substances). July 1, 2002.

14. Environmental Protection Agency. Documents for FIFRA Scientific Advisory Panel Meeting on Regulatory Use of *In Vitro* Systems, October 28-29, 2003. (This document is a list of documents which appeared on the disk submitted to OSCP on October 2, 2003.

15. Microarray Gene Expression Data Society. Minimum Information About a Microarray Experiment_MIAME for Toxicogenomics (MIAME/TOX). Draft report February 12, 2003.