

**ICCVAM TEST METHOD EVALUATION REPORT:
IN VITRO OCULAR TOXICITY TEST METHODS FOR
IDENTIFYING SEVERE IRRITANTS AND
CORROSIVES**

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

[This Page Intentionally Left Blank]

ICCVAM TEST METHOD EVALUATION REPORT:

***IN VITRO* OCULAR TOXICITY TEST METHODS FOR
IDENTIFYING SEVERE IRRITANTS AND CORROSIVES**

**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**

**September 2006
NIH Publication No.: 06-4511**

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

[This Page Intentionally Left Blank]

Printed: _____

TABLE OF CONTENTS

LIST OF TABLES vi
LIST OF ABBREVIATIONS AND ACRONYMSvii
INTERAGENCY COORDINATING COMMITTEE ON THE VALIDATION OF ALTERNATIVE METHODS –AGENCY REPRESENTATIVESviii
ACKNOWLEDGEMENTSix
PREFACExiii
EXECUTIVE SUMMARYxvii

1.0 INTRODUCTION 1
2.0 THE BCOP TEST METHOD 5
 2.1 BCOP Technical Summary 5
 2.1.1 Test Method Description 5
 2.1.2 Validation Database..... 6
 2.1.3 Test Method Accuracy..... 6
 2.1.4 Test Method Reliability (Inter- and Intra-Laboratory Reproducibility)..... 8
 2.2 ICCVAM Recommendations for the BCOP Test Method 9
 2.2.1 Use of the BCOP Test Method..... 9
 2.2.2 BCOP Test Method Protocol 10
 2.2.3 Optimization of the Current BCOP Test Method Protocol..... 10
3.0 THE ICE TEST METHOD..... 13
 3.1 ICE Technical Summary 13
 3.1.1 Test Method Description 13
 3.1.2 Validation Database..... 13
 3.1.3 Test Method Accuracy..... 14
 3.1.4 Test Method Reliability (Inter- and Intra-Laboratory Reproducibility)..... 16
 3.2 ICCVAM Recommendations for the ICE Test Method..... 16
 3.2.1 Use of the ICE Test Method 16
 3.2.2 ICE Test Method Protocol 17
 3.2.3 Optimization of the Current ICE Test Method Protocol 18
4.0 THE IRE TEST METHOD 19
 4.1 IRE Technical Summary 19
 4.1.1 Test Method Description 19
 4.1.2 Validation Database..... 19
 4.1.3 Test Method Accuracy..... 20
 4.1.4 Test Method Reliability (Inter- and Intra-Laboratory Reproducibility)..... 20
 4.2 ICCVAM Recommendations for the IRE Test Method..... 22
 4.2.1 Use of the IRE Test Method 22
 4.2.2 IRE Test Method Protocol 23
 4.2.3 Optimization of the Current IRE Test Method Protocol 23
5.0 THE HET-CAM TEST METHOD 25

5.1 HET-CAM Technical Summary..... 25

5.1.1 Test Method Description 25

5.1.2 Validation Database..... 25

5.1.3 Test Method Accuracy..... 26

5.1.4 Test Method Reliability (Inter- and Intra-Laboratory
Reproducibility)..... 29

5.2 ICCVAM Recommendations for the HET-CAM Test Method 30

5.2.1 Use of the HET-CAM Test Method 30

5.2.2 HET-CAM Test Method Protocol..... 30

5.2.3 Optimization of the Current HET-CAM Test Method Protocol..... 31

**6.0 GENERAL RECOMMENDATIONS AND COMPARISON OF
PERFORMANCE CHARACTERISTICS FOR FOUR EVALUATED *IN
VITRO* TEST METHODS..... 33**

**7.0 ICCVAM RECOMMENDATIONS ON SUBSTANCES FOR
VALIDATION OF *IN VITRO* OCULAR TOXICITY TEST METHODS
FOR THE EVALUATION OF OCULAR CORROSIVES AND SEVERE
IRRITANTS..... 37**

8.0 REFERENCES 39

APPENDIX A EXPERT PANEL REPORTS A-1

A1 Expert Panel Report: Evaluation of the Current Validation Status of *In
Vitro* Test Methods for Identifying Ocular Corrosives and Severe
Irritants A-3

A2 Expert Panel Report: Evaluation of the Current Validation Status of *In
Vitro* Test Methods for Identifying Ocular Corrosives and Severe
Irritants – Addendum A-135

A3 Minutes from Expert Panel Meeting on January 11-12, 2005..... A-153

A4 Minutes from Expert Panel Teleconference on September 19, 2005 A-221

**APPENDIX B COMPARISON OF PERFORMANCE
CHARACTERISTICS OF FOUR *IN VITRO* TEST
METHODS FOR THE IDENTIFICATION OF SEVERE
OCULAR IRRITANTS OR CORROSIVES..... B-1**

B1 Comparison of Performance Characteristics of Four *In Vitro* Test
Methods for Identification of GHS Severe Ocular Irritants or
Corrosives..... B-3

B2 Comparison of Performance Characteristics of Four *In Vitro* Test
Methods for Identification of EPA Severe Ocular Irritants or
Corrosives..... B-7

B3 Comparison of Performance Characteristics of Four *In Vitro* Test
Methods for Identification of EU Severe Ocular Irritants or Corrosives.... B-11

**APPENDIX C RELEVANT FEDERAL OCULAR IRRITATION
REGULATIONS AND TESTING GUIDELINES C-1**

C1 Table of Relevant Ocular Irritation Regulations..... C-3

C2 16CFR1500.42: Test for Eye Irritants..... C-9

C3 29CFR1910.1200: Hazard Communication C-13

C4 Health Effects Test Guidelines OPPTS 870.2400: Acute Eye Irritation C-39

APPENDIX D **ICCVAM RECOMMENDED BCOP TEST METHOD
PROTOCOL.....D-1**

APPENDIX E **ICCVAM RECOMMENDED ICE TEST METHOD
PROTOCOL..... E-1**

APPENDIX F **ICCVAM RECOMMENDED IRE TEST METHOD
PROTOCOL..... F-1**

APPENDIX G **ICCVAM RECOMMENDED HET-CAM TEST METHOD
PROTOCOL.....G-1**

APPENDIX H **ICCVAM RECOMMENDED REFERENCE
SUBSTANCES LISTH-1**

APPENDIX I **FEDERAL REGISTER NOTICES AND PUBLIC
COMMENTS..... I-1**

 I1 Federal Register Notices I-3

 I2 ICCVAM Consideration of Public Comments Received in Response to
 Federal Register Notices I-27

LIST OF TABLES

Table 2-1 False Positive and False Negative Rates of the BCOP Test Method, by Chemical Class and Properties of Interest, for the GHS Classification System 7

Table 3-1 False Positive and False Negative Rates of the ICE Test Method, by Chemical Class and Properties of Interest, for the GHS Classification System 15

Table 4-1 False Positive and False Negative Rates of the IRE Test Method, by Chemical Class and Properties of Interest, for the GHS Classification System (Analysis Based on the Pooled Data Set)..... 21

Table 5-1 False Positive and False Negative Rates of the HET-CAM Test Method, by Chemical Class and Properties of Interest, for the GHS Classification System 27

Table 6-1 Comparison of Performance Characteristics of Four *In Vitro* Ocular Test Methods for the Identification of Severe Ocular Irritants or Corrosives, for Three Hazard Classification Systems..... 34

LIST OF ABBREVIATIONS AND ACRONYMS

°C	Degrees Centigrade
BCOP	Bovine Corneal Opacity and Permeability
BRD	Background Review Document
CAM	Chorioallantoic Membrane
CV	Coefficient of Variation
ECVAM	European Center for the Validation of Alternative Methods
EPA	U.S. Environmental Protection Agency
EU	European Union
FR	<i>Federal Register</i>
g	Gram
GHS	Globally Harmonized System
HET-CAM	Hen's Egg Test-Chorioallantoic Membrane
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
ICE	Isolated Chicken Eye
IRE	Isolated Rabbit Eye
IS	Irritation Score
MeSH	Medical Subject Headings
mL	Milliliter
NICEATM	National Toxicology Program Center for the Evaluation of Alternative Toxicological Methods
NTP	U.S. National Toxicology Program
OTWG	Ocular Toxicity Working Group
SACATM	Scientific Advisory Committee on Alternative Toxicological Methods
UN	United Nations
UV/VIS	Ultraviolet/Visible

**INTERAGENCY COORDINATING COMMITTEE ON THE VALIDATION
OF ALTERNATIVE METHODS - AGENCY REPRESENTATIVES**

Agency for Toxic Substances and Disease Registry

- Moiz Mumtaz, Ph.D.

Consumer Product Safety Commission

- Marilyn L. Wind, Ph.D. (Vice-Chair)
- * Kailash C. Gupta, D.V.M., Ph.D.
- * Patricia Bittner, M.S.
- * Kristina Hatlelid, Ph.D.

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.

Department of Energy

- Michael Kuperberg, 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

- Karen Hamernik, Ph.D.

Office of Research and Development

- ◇ Julian Preston, Ph.D.
- * Suzanne McMaster, Ph.D.

OECD Test Guidelines Program

- * Jerry Smrcek, Ph.D.

Office of Pesticides Programs

- * Amy Rispin, Ph.D.
- * Deborah McCall

- Principal Agency Representative
- ◇ Alternate Principal Agency Representative
- * Other Designated Agency Representatives

8-25-2006

Food and Drug Administration

- Leonard M. Schechtman, Ph.D. (Chair)

Office of Science

- ◇ Suzanne 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.
- * Ying Huang, Ph.D.

Center for Food Safety and 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.

Office of Regulatory Affairs

- * Lawrence A. D'Hoostelaere, Ph.D.

National Cancer Institute

- Alan Poland, M.D.
- ◇ T. Kevin Howcroft, Ph.D.

National Institute of Environmental Health Sciences

- William S. Stokes, D.V.M., D.A.C.L.A.M.
- ◇ 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.
- ◇ K. Murali Rao, MD, Ph.D.

National Institutes of Health

- Margaret D. Snyder, Ph.D.

National Library of Medicine

- Vera Hudson, M.S.
- ◇ Jeanne Goshorn, M.S.

Occupational Safety and Health Administration

- Surender Ahir, Ph.D.

ACKNOWLEDGMENTS

The following individuals are acknowledged for their contributions to the in vitro ocular test method review process

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Ocular Toxicity Working Group (OTWG)

Consumer Product Safety Commission

Kailash Gupta, Ph.D.
Cassandra Prioleau, Ph.D.
Marilyn Wind, Ph.D.

Department of Defense

Harry Salem, Ph.D.

Department of Transportation

Steve Hwang, Ph.D.

Environmental Protection Agency

Karen Hamernik, Ph.D. (OTWG, Co-Chair)
Meta Bonner, Ph.D.
Andrew Geller, Ph.D.
Karen Hicks
Marianne Lewis
Deborah McCall
Mark Perry, Ph.D.
John Redden, Ph.D.
Amy Rispin, Ph.D.

Food and Drug Administration

Jill Merrill, Ph.D. (OTWG, Co-Chair)
Robert Bronaugh, Ph.D.
Paul Brown, Ph.D.
Wiley Chambers, M.D.
Abigail Jacobs, Ph.D.
Donnie Lowther
Leonard Schechtman, Ph.D. (ICCVAM, Chair)

Occupational Safety and Health Administration

Surender Ahir, Ph.D.

National Institute of Environmental Health Sciences

Raymond Grissom, Ph.D.
William Stokes, D.V.M., D.A.C.L.A.M.
(Executive Director, ICCVAM)
Raymond Tice, Ph.D.

***In Vitro* Ocular Test Methods Expert Panel**

Robert Scala, Ph.D. (Panel Chair), Consultant, Tucson, Arizona

Sally Atherton, Ph.D., Professor, Medical College of Georgia, Augusta, Georgia

Roger Beuerman, Ph.D., Professor, Louisiana State University, New Orleans, Louisiana

June Bradlaw, Ph.D., International Foundation for Ethical Research, Rockville, Maryland

Ih Chu, Ph.D., Health Canada, Ottawa, Canada

Henry Edelhauser, Ph.D., Professor, Emory University, Atlanta, Georgia

Nancy Flournoy, Ph.D., Professor, University of Missouri, Columbia, Missouri

Donald Fox, Ph.D., Professor, University of Houston, Houston, Texas

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

Shayne Gad, Ph.D., D.A.B.T., A.T.S., Consultant, Gad Consulting Services, Cary, North Carolina

Sidney Green, Ph.D., A.T.S., Graduate Professor, Howard University, Washington, DC

Frederick Guerriero, M.S., Senior Occupational Toxicologist, GlaxoSmithKline, King of Prussia, Pennsylvania

A. Wallace Hayes, Ph.D., D.A.B.T., F.A.T.S., F.I.Biol., F.A.C.F.E., E.R.T., Scientist, Harvard School of Public Health, Andover, Massachusetts

Hiroshi Itagaki, Ph.D., Principal Scientist, Shiseido Co., Ltd., Japan

David Lovell, Ph.D., Reader in Medical Statistics, University of Surrey, United Kingdom

Yasuo Ohno, Ph.D., D.J.S.T.S., Director of Japanese Society of Alternatives to Animal Experiments and Director of Division of Pharmacology, National Institute of Health Sciences, Japan

Robert Peiffer, D.V.M., D.A.C.V.O., Senior Investigator, Merck Research Laboratories, West Point, Pennsylvania

Lionel Rubin, V.M.D., D.A.C.V.O., Emeritus Professor of Ophthalmology, University of Pennsylvania, Philadelphia, Pennsylvania

Horst Spielmann, Dr. Med., Director and Professor, ZEBET at the BfR, Berlin, Germany

Martin Stephens, Ph.D., Vice President for Animal Research, Humane Society of the United States, Washington DC

Katherine Stitzel, D.V.M., Consultant, West Chester, Ohio

Peter Theran, V.M.D., D.A.C.V.I.M., Vice President Animal Science, Massachusetts Society for the Prevention of Cruelty to Animals, Novato, California

Scheffer Tseng, M.D., Ph.D., Director, Ocular Surface Research and Education Foundation, Miami, Florida

Philippe Vanparys, Ph.D., Senior Research Fellow, Johnson and Johnson, Belgium

**National Toxicology Program Interagency Center for the Evaluation
of Alternative Toxicological Methods**

David Allen, Ph.D.

ILS, Inc.

Frank Deal, M.S.

ILS, Inc.

Bradley Blackard, M.S.P.H.

ILS, Inc.

Linda Litchfield

ILS, Inc.

Sue Brenzel

ILS, Inc.

Deborah McCarley

NIEHS

Thomas Burns, M.S.

ILS, Inc.

Michael Paris

ILS, Inc.

Patricia Ceger, M.S.

ILS, Inc.

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

(Director)

NIEHS

Jeff Charles, Ph.D., M.B.A., D.A.B.T.

ILS, Inc.

Judy Strickland, Ph.D., D.A.B.T.

ILS, Inc.

Neepta Choksi, Ph.D.

ILS, Inc.

Raymond Tice, Ph.D. (Deputy Director)

NIEHS

James Truax, M.A.

ILS, Inc.

Additional Contributors

Chantra Eskes, Eng., Ph.D.

ECVAM
Ispra, Italy

Robert L Guest Bsc, CBiol, MIBiol

SafePharm Laboratories, Ltd.
Derby, United Kingdom

John Harbell, Ph.D.

Institute for *In Vitro* Sciences
Gaithersburg, Maryland

Joe Haseman, Ph.D.

Consultant
Raleigh, North Carolina

Penny Jones

Unilever Research
Sharnbrook, United Kingdom

Menk Prinsen

TNO Nutrition & Food Research Institute
The Netherlands

Horst Spielmann, Dr. med.

ZEBET
Berlin, Germany

PREFACE

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) is charged by the ICCVAM Authorization Act of 2000 (42 U.S.C. § 2851-2, 2851-5 (2000); available at <http://iccvam.niehs.nih.gov/about/PL106545.pdf>) with evaluating the scientific validity of new, revised, and alternative toxicological test methods applicable to U.S. Federal agency safety testing requirements. Following such evaluations, ICCVAM is required to provide recommendations to U.S. Federal agencies regarding the usefulness and limitations of such methods.

In October 2003, the U.S. Environmental Protection Agency (EPA) formally nominated several ocular toxicity test method activities to ICCVAM. ICCVAM determined that four *in vitro* test methods proposed for identifying potential ocular corrosives and severe irritants in a tiered-testing strategy should have the highest priority for evaluation. This was based on the availability of existing validation data for all four methods and the fact that determining the adequacy of validation¹ is a prerequisite for test methods to be considered for regulatory acceptance (ICCVAM 1997, 2003). The four test methods were the Bovine Corneal Opacity and Permeability (BCOP) assay, the Hen's Egg Test - Chorioallantoic Membrane (HET-CAM) assay, the Isolated Chicken Eye (ICE) assay, and the Isolated Rabbit Eye (IRE) assay.

An ICCVAM Ocular Toxicity Working Group (OTWG) was established to work with the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) to carry out the test method evaluations. ICCVAM and NICEATM also collaborated closely with the European Centre for the Validation of Alternative Methods (ECVAM) in conducting the evaluations, with Drs. Chantra Eskes and Valérie Zuang serving as ECVAM liaisons to the OTWG.

NICEATM, in conjunction with the OTWG, prepared four comprehensive background review documents (BRDs) reviewing the available data and information for each of the four *in vitro* test methods. Each BRD described the current validation status of the *in vitro* test method, including its reliability and accuracy, the scope of the substances tested, and the availability of a standardized protocol. The BRDs were based on published studies using the respective test method, and other data and information submitted in response to a 2004 public call for information. The draft BRDs were made available to the public for comment on November 1, 2004, and a public independent expert panel meeting also was announced.

The ICCVAM organized an international independent Expert Panel meeting on January 11-12, 2005, to assess the validation status of these four *in vitro* test methods for identifying ocular corrosives or severe irritants. While a comprehensive review was conducted, public comments at the meeting revealed that additional relevant data were available that had not yet been provided in response to earlier requests for data. Accordingly, the Expert Panel recommended that if such data could be obtained, a reanalysis of each test method should be

¹Validation is the process by which the reliability and relevance of a test method are established for a specific purpose (ICCVAM 1997, 2003).

performed. Availability of the Expert Panel's independent report was announced on March 21, 2005.

In response to the Expert Panel's recommendation, a second public request for *in vitro* data was published on February 28, 2005. In response to this request, additional *in vitro* test method data and corresponding *in vivo* rabbit eye test results were submitted for the BCOP, HET-CAM, and ICE test methods. The additional data, together with clarified rules for hazard classification and reclassification of the chemical classes of the test substances necessitated a reanalysis of the accuracy and reliability of all four test methods. The accuracy and reliability reanalyses and a revised reference substances list for validation of *in vitro* tests to detect ocular corrosives and severe irritants were provided in a BRD Addendum released on July 26, 2005.

The Expert Panel was subsequently reconvened via teleconference on September 19, 2005 to discuss the BRD Addendum. The Expert Panel provided final conclusions regarding the effects of the information in the BRD Addendum on their original evaluation from the January 11-12, 2005 meeting. The report of this meeting also was published and public comments requested.

The draft BRDs, draft BRD Addendum, Expert Panel report and addendum, and all public comments were subsequently made available to the Scientific Advisory Committee on Alternative Toxicological Methods (SACATM) for comment at their meeting on December 12, 2005. The SACATM concurred with the consensus conclusions of the Expert Panel.

The ICCVAM and OTWG considered the Expert Panel report and addendum, the revised accuracy and reliability analyses, all public comments, and the comments of SACATM in preparing the final ICCVAM test method recommendations provided in this report. This report will be made available to the public and provided to U.S. Federal agencies for consideration, in accordance with the ICCVAM Authorization Act of 2000 (42 U.S.C. § 2851-2, 2851-5 [2000]) (Available at <http://iccvam.niehs.nih.gov/about/PL106545.pdf>). Agencies with applicable testing regulations and/or guidelines must respond to ICCVAM within 180 days after receiving the ICCVAM recommendations. These responses will be made available to the public on the ICCVAM website (<http://iccvam.niehs.nih.gov>) as they are received.

In this Test Method Evaluation Report, ICCVAM states that there are sufficient data to substantiate the use of BCOP and ICE test methods, with certain limitations, as a screening test to identify substances as ocular corrosives and severe irritants in a tiered-testing strategy, using a weight-of-evidence approach. When used in this manner, these methods should reduce the number of animals needed for ocular toxicity testing and refine animal use by avoiding the pain and distress associated with testing severely irritating and corrosive substances. Since ocular irritancy testing may involve more than slight or momentary pain or distress, alternative test methods must be considered prior to the use of animals, as required by U.S. Federal animal welfare regulations and policies. Accordingly, *in vitro* alternative test methods should be considered prior to *in vivo* ocular testing and used where determined appropriate for a specific testing situation. Consistent with the mission of ICCVAM,

appropriate use of these methods will support improved animal welfare while ensuring the continued protection of human health.

Acknowledgments

The efforts of many individuals who contributed to the preparation, review, and revision of this report are gratefully acknowledged. We especially recognize all of the Expert Panel members for their thoughtful evaluations and generous contributions of time and effort. Special thanks are extended to Dr. Robert Scala for serving as both the Panel Chair and a Group Chair and to Drs. Kathy Stitzel, James Freeman, and Shayne Gad for their service as Group Chairs. The efforts of the OTWG were invaluable for assuring a meaningful and comprehensive review. We especially would like to thank the co-chairs of the OTWG, Drs. Jill Merrill (U.S. Food and Drug Administration) and Karen Hamernik (U.S. Environmental Protection Agency) for their leadership. The efforts of the NICEATM staff in preparing the BRDs, organizing the Expert Panel meeting and teleconference, and preparing this final report are greatly appreciated. We also acknowledge Drs. David Allen, Jeff Charles, and Neepa Choksi and Messrs. Bradley Blackard, Thomas Burns, and James Truax of Integrated Laboratory Systems (ILS), Inc., the NICEATM Support Contractor, and Dr. Raymond Tice for his initial contributions as a member of the ILS, Inc. support contract and later as a member of NICEATM.

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

Leonard Schechtman, Ph.D.
U.S. Food and Drug Administration
National Center for Toxicological Research
Chairman, ICCVAM

[This Page Intentionally Left Blank]

EXECUTIVE SUMMARY

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) recently completed the technical evaluation of the validation status of four *in vitro* ocular irritation test methods proposed as screening tests² for identifying potential ocular corrosives and severe irritants in a tiered-testing strategy³, as part of a weight-of-evidence approach. The four test methods are the Bovine Corneal Opacity and Permeability (BCOP) assay, the Hen's Egg Test - Chorioallantoic Membrane (HET-CAM) assay, the Isolated Chicken Eye (ICE) assay, and the Isolated Rabbit Eye (IRE) assay. The U.S. Environmental Protection Agency (EPA) formally nominated these test methods for evaluation by ICCVAM in October 2003. In addition to evaluating their current usefulness and limitations as screening tests for identifying ocular corrosives and severe irritants, ICCVAM developed a recommended standardized protocol for each test method; made recommendations, where considered appropriate, for further research and development, optimization, and/or validation efforts; and developed a list of reference substances for such activities.

None of the four *in vitro* test methods evaluated can be considered to be replacements for the *in vivo* rabbit eye test. However, based on the available data, BCOP and ICE can be used, in appropriate circumstances and with certain limitations, as screening tests for the detection of ocular corrosives and severe irritants in a tiered-testing strategy, as part of a weight-of-evidence approach. At the present time, HET-CAM, using the decision criteria of Luepke (1985), and IRE are not recommended as screening tests for the identification of ocular corrosives and severe irritants for regulatory hazard classification purposes. Before HET-CAM and IRE can be recommended for this purpose, the protocol and the decision criteria for the identification of ocular corrosives and severe irritants need to be optimized and undergo further validation.

This evaluation provides validation information that should be helpful to various stakeholders (e.g., applicable U.S. Federal regulatory agencies, the international regulatory community, the pharmaceutical, pesticide, and commercial chemical industries) in determining when these test methods might be useful and which test method might be the most appropriate for a specific testing situation. These *in vitro* test methods, when used appropriately, will reduce and refine animal use for ocular safety testing.

²According to the *ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods*, a **screen** or **screening test** is "a rapid, simple test conducted for the purposes of a general classification of substances according to general categories of hazard. The results of a screen generally are used for preliminary decision making and to set priorities for more definitive tests. A screening test may have a truncated response range (e.g., be able to reliably identify active chemicals but not inactive chemicals)" (ICCVAM 2003).

³A tiered-testing strategy approach may not be applicable to purposes other than regulatory classification and labeling.

Specific Test Method Recommendations

BCOP Test Method

There are sufficient data to support the use of the BCOP test method, in appropriate circumstances and with certain limitations, as a screening test to identify substances as ocular corrosives and severe irritants (i.e., EPA Category I, United Nations [UN] Globally Harmonized System of Classification and Labelling of Chemicals [GHS] Category 1, European Union [EU] R41) in a tiered-testing strategy, as part of a weight-of-evidence approach. The identified limitations for this test method are based on the false negative and false positive rates observed for certain chemical and physical classes. Based on the available database, the false negative rates for alcohols and solids range from 67% (2/3)⁴ to 100% (2/2) and 42% (5/12) to 50% (5/10), respectively, depending on the hazard classification system. Additionally, the false positive rates for alcohols, ketones, and solids range from 50% (7/14) to 56% (9/16), 40% (4/10), and 10% (2/20 to 2/21), respectively, depending on the hazard classification system. When substances within these chemical and physical classes are excluded from the database, the accuracy of BCOP across the EU, EPA, and GHS classification systems ranges from 87% (72/83) to 92% (78/85) and the false negative and false positive rates range from 0% (0/27) to 12% (3/26) and 12% (7/58) to 16% (9/56), respectively.

Coefficient of variation (CV) analysis of BCOP test method intralaboratory repeatability data (*In Vitro* Irritancy Scores) from two studies ranged from 11.8% to 14.2% for 16 substances of varying irritancy and from 1.1% to 13% for five substances predicted as severe irritants. Intralaboratory reproducibility evaluations indicated mean and median CV values for permeability values were 33.4% and 29.0%, respectively, for 25 surfactant-based personal care cleaning formulations in one study. Mean CV values of *In Vitro* Irritancy Scores for 16 substances tested two or more times in three laboratories ranged from 12.6% to 14.8%, while the median CV values ranged from 6.7% to 12.4%.

In a qualitative assessment of interlaboratory reproducibility of hazard classification category, 67% to 94% of the substances were classified the same by the participating laboratories. Substances with less than complete agreement in the testing laboratories include those representing such chemical classes as alcohols, ketones, and heterocyclic compounds, and such product classes as solvents, surfactants, chemical intermediates, and pesticides.

A quantitative evaluation of interlaboratory reproducibility was conducted for three studies by performing a CV analysis of *In Vitro* Irritancy Scores obtained for substances tested in multiple laboratories. In these studies, the mean and median CV values were (a) 36% and 17%, respectively, for results obtained in either 11 or 12 laboratories, (b) 25% and 22%,

⁴The numbers in parentheses represent the numbers used to calculate the percentages. For the false negative or false positive rates, the numerators represent the total number of substances incorrectly identified as negatives or positives, respectively, by the *in vitro* test method, while the denominators represent the total number of substances identified as negatives or positives, respectively, by the *in vivo* rabbit eye test method.

respectively, for results obtained in five laboratories, and (c) 32.4% and 22.8%, respectively, for results obtained in three laboratories.

When studies are conducted using the BCOP test method, the study protocol should be based on the recommended standardized test method protocol provided in **Appendix D**. Exceptions and/or changes to the standardized test method protocol should be accompanied by a scientific rationale.

Users should be aware that BCOP's performance characteristics and the standardized test method protocol could be revised as additional data become available. For example, the current validation database did not allow for adequate evaluation of all chemical or product classes (e.g., formulations). Additional data may allow for further evaluation of this, as well as other chemical and product classes. Therefore, prior to initiation of BCOP studies, investigators are encouraged to consult the ICCVAM/National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) website (see <http://iccvam.niehs.nih.gov/methods/eyeirrit.htm>) to review the most current validation database, overall performance characteristics, chemical and physical class performance characteristics, and the recommended standardized test method protocol. Evaluation of the most current information will allow users to determine the appropriateness of this test method for evaluating substances that are within a specific chemical, physical, or product classes.

To further characterize and potentially improve the usefulness of the BCOP test method for identifying ocular corrosives and severe irritants, and to evaluate its possible future use for the identification of mild and moderate ocular irritants (e.g., EPA Category II, III, and IV; GHS Category 2; EU R36), the following evaluations are recommended:

1. A histopathological evaluation of the corneal tissue, using a standardized scoring scheme, should be conducted. Such data will allow for the development of standardized decision criteria and a more comprehensive evaluation of the usefulness of this endpoint for classifying and labeling substances, especially those that may otherwise produce borderline or false negative results.
2. Studies should be conducted to evaluate the impact of using a corneal holder that maintains normal corneal curvature (e.g., the corneal mounting system designed by Ubels et al. 2002) on accuracy and/or reliability of the BCOP test method.
3. The effect of modifying various test method protocol components (e.g., changing the duration of exposure) on the accuracy and/or reliability of the BCOP test method should be evaluated.

ICE Test Method

There are sufficient data to support the use of the ICE test method, in appropriate circumstances and with certain limitations, as a screening test to identify substances as ocular corrosives and severe irritants (i.e., EPA Category I, UN GHS Category 1, EU R41) in a tiered-testing strategy, as part of a weight-of-evidence approach. The identified limitations

for this method are based on the false negative and false positive rates that are observed for certain chemical and physical classes. Based on the available database, the false negative rates for alcohols, surfactants and solids range from 33% (1/3) to 50% (1/2), 44% (4/9) to 57% (4/7), and 46% (6/13) to 70% (7/10), respectively, depending on the hazard classification system. Additionally, the false positive rates for alcohols range from 27% (3/11) to 50% (5/10), depending on the hazard classification system evaluated. When substances within these chemical and physical classes are excluded from the database, the accuracy of ICE across the EU, EPA, and GHS classification systems ranges from 91% (72/79 to 75/82) to 92% (69/75) and the false negative and false positive rates range from 29% (2/7) to 33% (3/9) and 5% (4/73) to 6% (4/68 to 4/70), respectively.

The range of CV values for the corneal thickness measurement, when results were compared within experiments, was from 0.9% to 6.1%. The other endpoints evaluated produced ranges of CV values that were larger, with variability most prominent with the nonirritating substance. The range of CV values for the corneal thickness measurement, when results were compared across experiments, was from 1.8% to 6.3%. The CV values for the remaining endpoints had a larger range (e.g., corneal swelling CV = 13.9% to 138.7%). However, if the nonirritating substance is removed, the range of CV values is reduced (e.g., corneal swelling CV = 13.9% to 22.4%).

One interlaboratory comparative study involving four laboratories contained test data on 59 substances for an assessment of interlaboratory reproducibility. Based on a qualitative analysis, 60% to 70% of the substances classified as ocular corrosives or severe irritants, depending on the regulatory classification system employed (i.e., EPA 1996, GHS [UN 2003], EU 2001), were correctly identified by all four participating laboratories. A CV analysis of these same data indicated that the mean and median CV for severe substances tested was less than 35% for all test method endpoints, with the exception of corneal swelling.

When studies are conducted using this test method, the study protocol should be based on the recommended standardized ICE test method protocol provided in **Appendix E**. Exceptions and/or changes to the standardized test method protocol should be accompanied by a scientific rationale.

Users should be aware that ICE's performance characteristics and the standardized test method protocol could be revised as additional data become available. For example, the current validation database did not allow for adequate evaluation of all chemical or product classes (e.g., formulations). Additional data may allow for further evaluation of this, as well as other, chemical and product classes. Therefore, prior to initiation of ICE studies, investigators are encouraged to consult the ICCVAM/NICEATM website (see <http://iccvam.niehs.nih.gov/methods/eyeirrit.htm>) to review the most current validation database, overall performance characteristics, chemical and physical class performance characteristics, and the recommended standardized test method protocol. Evaluation of the most current information will allow users to determine the appropriateness of this test

method for evaluating substances that are within a specific chemical, physical, or product classes.

To further characterize and potentially improve the usefulness of the ICE test method for identifying severe ocular irritants and corrosives and its possible future use for the identification of mild and moderate ocular irritants (e.g., EPA Category II, III, and IV; GHS Category 2; EU R36), the following evaluations are recommended:

1. Additional optimization studies/evaluations should be conducted in an attempt to decrease the 29% to 33% false negative rate of the ICE test method. After optimization, additional studies to further assess the reliability and accuracy of the test method are recommended.
2. A histopathological evaluation of the corneal tissue, using a standardized scoring scheme, should be included when the ICE test method is conducted. Such data will allow for development of decision criteria and future assessments on the usefulness of this endpoint for classifying and labeling substances, especially those that may otherwise produce borderline or false negative results.

ICCVAM also recommends that centering lights be installed on the optical pachymeter, which is used to measure corneal thickness, to ensure consistent central corneal thickness measurements across laboratories.

IRE Test Method

Based on the accuracy (64% [68/107] to 69% [79/114]), false negative (24% [12/49] to 31% [14/45]), and false positive (35% [23/65] to 40% [25/62]) rates across the EU, EPA, and GHS classification systems, the use of the IRE test method for screening and identifying ocular corrosives and severe irritants (i.e., EPA Category I, GHS Category 1, EU R41) in a tiered-testing strategy, as part of a weight-of-evidence approach, is not recommended. There also are insufficient data using all four recommended IRE endpoints (corneal opacity, fluorescein penetration, corneal swelling, and observations of significant effect on corneal epithelium) to assess test method accuracy and reliability when all these endpoints are evaluated in a single study.

Based on a qualitative analysis of available data, 100% of the 12 to 18 substances were correctly identified as severe irritants or ocular corrosives in the IRE by four laboratories participating in a validation study, when compared to *in vivo* rabbit eye test data classification dependent on the regulatory classification system employed (i.e., EPA 1996, GHS [UN 2003], EU 2001). Substances with less than complete agreement in the testing laboratories include those representing such chemical classes as alcohols, ketones, and heterocyclic compounds; and such product classes as organic solvents, surfactants, chemical intermediates, and pesticides.

A quantitative evaluation of interlaboratory reproducibility was conducted for two studies by performing a CV analysis of corneal opacity, swelling, and, for the second study, fluorescein penetration measurements for substances tested in multiple laboratories. The CV analysis of

the first study indicated that the median CV for 59 substances tested was between 43.4% and 49.7% for the 4-hour corneal opacity and swelling endpoints, respectively. The CV was between 33.6% and 35.5% when only severe irritants are considered. In the second study using corneal opacity, swelling, and fluorescein penetration, the median CV for all substances ranged from 24.0% to 40.0% and from 15.4% to 35.5% when only severe irritants were considered.

When non-regulatory, validation, or optimization studies are conducted using the IRE test method, the protocol should be based on the standardized protocol provided in **Appendix F**. Exceptions and/or changes to the test method protocol should be accompanied by a scientific rationale.

Users should be aware that IRE's performance characteristics and the standardized test method protocol could be revised as additional data become available. Therefore, prior to initiation of IRE studies, investigators are encouraged to consult the ICCVAM/NICEATM website (see <http://iccvam.niehs.nih.gov/methods/eyeirrit.htm>) to review the most current validation database, overall performance characteristics, chemical and physical class performance characteristics, and the recommended standardized test method protocol. Evaluation of the most current information will allow users to determine the appropriateness of this test method for evaluating substances that are within a specific chemical, physical, or product classes.

To potentially improve the usefulness of the IRE test method for identifying severe ocular irritants and corrosives and its possible future use for the identification of mild and moderate ocular irritants (e.g., EPA Category II, III, and IV; GHS Category 2; EU R36), the following evaluations should be conducted:

1. The IRE test method decision criteria should be optimized. Once optimized, additional validation studies should be conducted to further evaluate the relevance and reliability of the IRE test method.
2. A histopathological evaluation of the corneal tissue, using a standardized scoring scheme, of the corneal tissue should be included when the IRE test method is conducted. Such data will allow for development of decision criteria and future assessments on the usefulness of this endpoint for classifying and labeling substances, especially those that may otherwise produce borderline or false negative results.

ICCVAM also recommends that centering lights be installed when an optical pachymeter is used to measure corneal thickness, to ensure consistent central corneal thickness measurements across laboratories.

HET-CAM Test Method

ICCVAM evaluated several HET-CAM analysis methods proposed for identifying substances that are ocular corrosives or severe irritants. These included one analysis method termed Irritation Score (IS)(B)-10 and another analysis method termed IS(B)-100. The range of hazard classification accuracy rates across the EU, EPA, and GHS classification systems

for these two analysis methods ranged from 65% (64/98) to 68% (69/101) for IS(B)-10 and 52% (69/133) to 57% (94/164) for IS(B)-100, when the decision criteria of Luepke (1985) were used. The overall false negative and false positive rates of the IS(B)-10 analysis method range from 30% (10/33 to 12/40) to 32% (10/31) and 33% (20/61) to 36% (24/67), respectively, depending on the classification system. The overall false negative and false positive rates for the IS(B)-100 analysis method range from 6% (2/33) to 13% (5/39) and 52% (68/131) to 59% (58/99), respectively, depending on the classification system. Based on these rates, the use of these analyses methods and decision criteria for screening and identifying ocular corrosives and severe irritants (i.e., EPA Category I, GHS Category 1, EU R41) in a tiered-testing strategy, as part of a weight-of-evidence approach, is not recommended.

The analysis of intralaboratory repeatability was evaluated using data from two different publications for the IS(B) analysis method. In both studies, the hemorrhage endpoint had the highest CV value (109.10%-117.56%). Similar results were obtained from the analysis of intralaboratory reproducibility.

A qualitative analysis of interlaboratory reliability for the also was conducted for the IS(B) analysis method. For the IS(B)-10 analysis method, the participating laboratories were in 100% agreement for 84 to 85 (79% to 81%) of 104 to 107 substances evaluated, when compared to all three hazard classification systems. For the IS(B)-100 analysis method, the participating laboratories in a study were in 100% agreement for 80 to 81 (82% to 84%) of the 95 to 99 substances evaluated, when compared to all three hazard classification systems.

A quantitative evaluation of interlaboratory reproducibility for 14 substances, evaluated at 100% concentration (IS(B)-100), indicated that the mean and median CV values were 31.86% and 33.04%, respectively. For 12 substances evaluated at 10% concentration (IS(B)-10), the mean and median CV values were 66.29% and 60.75%, respectively. For the substances evaluated in another study which used the IS(B) analysis method, the mean and median CV values for substances tested at 10% concentration were 60.17% and 42.65%, respectively. For substances tested at 100% concentration in the same study, the mean and median CV values were lower: 35.21% and 26.22%, respectively.

When non-regulatory, validation, or optimization studies are conducted using the HET-CAM test method, the protocol should be based on the standardized protocol provided in **Appendix G**. Exceptions and/or changes to the test method protocol should be accompanied by a scientific rationale.

Users should be aware that HET-CAM's performance characteristics and the standardized test method protocol could be revised as additional data becomes available. Therefore, prior to initiation of HET-CAM studies, investigators are encouraged to consult the ICCVAM/NICEATM website (see <http://iccvam.niehs.nih.gov/methods/eyeirrit.htm>) to review the most current validation database, overall performance characteristics, chemical and physical class performance characteristics, and the recommended standardized test method protocol. Evaluation of the most current information will allow users to determine

the appropriateness of this test method for evaluating substances that are within a specific chemical, physical, or product classes.

To potentially improve the usefulness of the HET-CAM test method for identifying severe ocular irritants and corrosives and its possible future use for the identification of mild and moderate ocular irritants (e.g., EPA Category II, III, and IV; GHS Category 2; EU R36), additional studies should be conducted to further optimize the HET-CAM prediction models and the decision criteria (e.g., mtc10) that would be used to identify ocular corrosives and severe irritants for the EPA, GHS, or EU classification systems.

General Recommendations and Comparison of Performance Characteristics for Four *In Vitro* Test Methods

Results from appropriately validated *in vitro* ocular toxicity test methods are recommended for use in a weight-of-evidence decision making process in accordance with the EPA and EU ocular testing regulations (EPA 1998, EU 2004) and the GHS tiered-testing strategy (UN 2003). In these testing schemes, when a positive result is obtained in an appropriately validated *in vitro* test, a test substance may be classified as an ocular hazard without testing in rabbits. A substance that tests negative in the *in vitro* ocular toxicity test would need to be tested in the *in vivo* ocular test to identify possible *in vitro* false negatives and to identify moderate and mild ocular irritants. As is appropriate for any test system, there is the opportunity for confirmatory testing if false positive results are indicated based on a weight-of-evidence evaluation of supplemental information (e.g., structure-activity relationships, other testing data). Use of a weight-of-evidence decision making process and a tiered-testing strategy for classification of substances as ocular corrosives or severe irritants may eliminate the pain and distress that might be experienced by rabbits who otherwise would have been administered these test substances.

The comparative accuracy and false positive/false negative rates of these four *in vitro* ocular toxicity test methods in identifying ocular corrosives and severe irritants using the EU, EPA, and GHS classification systems are summarized in **Table 6-1**. Exclusion of specific chemical and physical classes increases the accuracy and decreases the false positive and false negative rates for BCOP and ICE. ICCVAM recommends that users consider, to the extent possible, the chemical and physical structures of the substances to be tested to determine whether either of these test methods would be appropriate to use as a screening test for ocular corrosion or severe irritation. Additional studies with each test method are recommended to determine if modification of the test method standardized protocol and/or the decision criteria for classification of a test substance as a corrosive/severe irritant or as a nonsevere irritant/nonirritant can improve test method sensitivity and specificity.

Additional research and development, optimization, and/or validation efforts should use reference substances with existing rabbit data. Additional rabbit studies should be conducted only if important data gaps are identified. If such studies are conducted, they should be designed to minimize the number of rabbits tested, to minimize or avoid pain and distress, and to maximize the information collected. Design and conduct of such studies should be in

accordance with the recommendations from the Scientific Symposium on Mechanisms of Chemically-Induced Ocular Injury and the Scientific Symposium on Minimizing Pain and Distress in Ocular Safety Testing (see <http://iccvam.niehs.nih.gov/methods/ocudocs/ocumeet/sympinfo.htm>). These symposia were organized by ICCVAM, NICEATM, and the European Centre for the Validation of Alternative Methods.

All raw data generated using any of the recommended standardized *in vitro* ocular testing protocols and the *in vivo* rabbit eye test on the same substance should be submitted to NICEATM to expand the available validation database for these four test methods. The availability of such data will allow for additional retrospective evaluations of test method accuracy and/or reliability. Ideally, all substances should be completely identified (e.g., chemical name, chemical class, physicochemical properties). However, if this is not possible for proprietary reasons, data may be submitted using coded labels for each substance tested. If such coding is used, as much information as possible on physical and chemical properties should be provided to NICEATM.

Although the IRE and HET-CAM test methods cannot currently be recommended for meeting regulatory testing requirements, there may be non-regulatory uses for these two test methods. Accordingly, the four *in vitro* test methods should be considered prior to conducting *in vivo* ocular testing and an alternative test method should be used where determined appropriate for the specific testing situation. Since ocular irritancy testing frequently involves more than slight or momentary pain or distress, consideration of alternative test methods prior to the use of animals is necessary to comply with provisions of U.S. Animal Welfare Act regulations (9 CFR, Part 2, Section 2.31 and 9 CFR, Part 2, Section 2.32), the Public Health Service Policy on the Humane Care and Use of Laboratory Animals (PHS 2002), and the U.S. Government Principles for the Utilization and Care of Vertebrate Animals Used in Testing, Research, and Training (National Research Council 1996).

The potential usefulness of combining two or more *in vitro* test methods in a battery to identify ocular corrosives and severe irritants should be evaluated. Currently, there is insufficient guidance on the utility of a battery approach for such determinations.

Interested stakeholders are encouraged to support research and development of alternative test methods and technologies that may provide for a more accurate assessment of ocular toxicity and/or advantages in terms of time and cost.

ICCVAM Recommended Substances for Validation of *In Vitro* Ocular Toxicity Test Methods for the Evaluation of Ocular Corrosives and Severe Irritants

ICCVAM developed a list of reference substances recommended for the development of alternative ocular toxicity test methods and for evaluating the performance of any optimized test method protocol (**Appendix H**). Use of this standardized list of reference substances will aid in evaluating the comparative performance of different alternative test methods and, thus, in the selection of the most appropriate test method(s) to be used for a particular testing purpose. In accordance with ICCVAM procedures, once an adequate validation database is

available for any of these test methods, performance standards will be developed that can be used to evaluate the performance of other test methods that are structurally and functionally similar. These performance standards will include essential test method components, a minimum list of reference chemicals (i.e., a subset of the recommended list in this report), and comparable performance that should be achieved.

1.0 INTRODUCTION

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) is charged by the ICCVAM Authorization Act of 2000 (42 U.S.C. § 2851-2, 2851-5 [2000]; available at <http://iccvam.niehs.nih.gov/about/PL106545.pdf>) to evaluate the scientific validity of new, revised, and alternative toxicological test methods applicable to U.S. Federal agency safety testing requirements. Following such evaluations, ICCVAM is required to provide recommendations to U.S. Federal agencies regarding the usefulness and limitations of such methods.

In August 2003, the ICCVAM Scientific Advisory Committee on Alternative Toxicological Methods (SACATM) recommended that ICCVAM give high priority to reviewing the validation status of existing *in vitro* test methods proposed for identifying ocular corrosives and severe irritants. In October 2003, the U.S. Environmental Protection Agency (EPA) formally nominated four *in vitro* ocular irritation test methods and related activities for evaluation by ICCVAM. This included review of the current validation status of four *in vitro* test methods proposed for identifying potential ocular corrosives and severe irritants in a tiered-testing strategy, since validation⁵ of a test method is a prerequisite for it to be considered for regulatory acceptance (ICCVAM 1997, 2003). The four test methods were the Bovine Corneal Opacity and Permeability (BCOP) assay, the Hen's Egg Test - Chorioallantoic Membrane (HET-CAM) assay, the Isolated Chicken Eye (ICE) assay, and the Isolated Rabbit Eye (IRE) assay. Within Europe, the European Commission has concluded that positive results from these four methods can be used to classify and label substances as severe ocular irritants and corrosives (EU 2004). However, the policy specifically states:

“These tests are not yet validated, and therefore not included in Annex V. Positive results can be used to consider a substance a severe irritant and R41 applied with no further testing. Where a negative result is obtained, an *in vivo* test should subsequently be required, as the *in vitro* tests have not been shown to adequately discriminate between eye irritants and non-irritants.”

ICCVAM unanimously agreed that the four nominated *in vitro* test methods should have a high priority for evaluation. An ICCVAM Ocular Toxicity Working Group (OTWG) was established to work with the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) to carry out these evaluations. ICCVAM and NICEATM also collaborate closely with the European Centre for the Validation of Alternative Methods (ECVAM), a component of the European Commission's Joint Research Centre. Accordingly, ECVAM liaisons were designated for the ICCVAM OTWG to ensure input and contributions during the evaluation and review process.

⁵Validation is the process by which the reliability and relevance of a test method are established for a specific purpose (ICCVAM 1997, 2003).

NICEATM, in conjunction with the OTWG, subsequently prepared four comprehensive background review documents (BRDs) reviewing the available data and information for each of the four *in vitro* test methods. Each BRD described the current validation status of the *in vitro* test method, including what is known about its reliability and accuracy, the scope of the substances tested, and the availability of a standardized protocol.

The BRDs were based on published studies using the respective test method, and other data and information submitted in response to a 2004 public call for information, which was published in a *Federal Register* (FR) notice (FR Vol. 69, No. 57, pp. 13859-61; available at <http://iccvam.niehs.nih.gov/methods/eyeirrit.htm>). On November 3, 2004, the availability of the draft BRDs was announced in an FR notice (Vol. 69, No. 212, pp. 64081-2; available at <http://iccvam.niehs.nih.gov/methods/eyeirrit.htm>). The BRDs were made available in electronic format on the ICCVAM/NICEATM website (Available at <http://iccvam.niehs.nih.gov/methods/eyeirrit.htm>) and from NICEATM on request.

The ICCVAM convened an international independent Expert Panel on January 11-12, 2005, to assess the validation status of these four *in vitro* test methods for identifying ocular corrosives or severe irritants. Comments from the public and scientific community on the BRDs were provided to the Expert Panel and made available on the ICCVAM/NICEATM website (<http://iccvam.niehs.nih.gov/methods/ocudocs/ocucomm.htm>). Public comments at the meeting revealed that additional relevant data was available that had not yet been provided in response to earlier requests for data. The Expert Panel recommended that the additional data be requested and that a reanalysis of the accuracy and reliability of each test method be conducted, where appropriate. On March 21, 2005, the availability of *The ICCVAM Expert Panel Evaluation of the Current Validation Status of In Vitro Test Methods for Identifying Ocular Corrosives and Severe Irritants* was announced via an FR notice (Vol. 70, No. 53, pp. 13513-4; available at <http://iccvam.niehs.nih.gov/methods/eyeirrit.htm>).

In response to the Expert Panel's recommendation, an FR notice was published on February 28, 2005 (Vol. 70, No. 38, pp. 9661-2; available at <http://iccvam.niehs.nih.gov/methods/eyeirrit.htm>). The notice requested all available *in vitro* data on these four *in vitro* ocular irritancy test methods and corresponding *in vivo* rabbit eye test method data, as well as any human exposure data (either via ethical human studies or accidental exposure). A request for relevant data was re-sent directly to the primary developers or users of each test method. In response to these requests, additional *in vitro* test method data and corresponding *in vivo* rabbit eye test results were submitted for the BCOP, HET-CAM, and ICE test methods, which were used for reanalysis of test method performance.

Further clarification of hazard classification rules for severe irritants also was obtained subsequent to the release of the four draft BRDs. This change resulted in a small number of substances previously classified as nonsevere irritants now being classified as severe irritants, and necessitated a reanalysis of the accuracy and reliability of all four test methods.

The original draft BRDs also provided an evaluation of the accuracy of each test method by chemical class. The chemical classes assigned to each test substance were revised based on a

chemical classification system consistent with the U.S. National Library of Medicine's Medical Subject Headings (MeSH; available at <http://www.nlm.nih.gov/mesh>), an internationally recognized standardized classification scheme. This scheme was used to ensure consistency in classifying substances by chemical class among all the *in vitro* ocular test methods under consideration, and resulted in some chemicals being re-classified into different chemical classes. As a result, the accuracy of each test method by chemical class was reanalyzed.

Finally, an additional accuracy analysis was conducted. In this analysis, the accuracy of each *in vitro* ocular irritancy test method for detecting ocular corrosives or severe irritants, depending on whether the *in vivo* rabbit classification was based on the severity of the response and/or its persistence to day 21 post-treatment, was determined.

A list of proposed reference substances for validation of *in vitro* tests to detect ocular corrosives and severe irritants was included in the draft BRDs released on November 3, 2004. A revised list of proposed reference substances was prepared after consideration of the following:

- Recommendations of the Expert Panel that resulted from their deliberations on January 11-12, 2005
- Submission of additional Draize rabbit eye test results for approximately 300 substances
- Clarification regarding the United Nations (UN) Globally Harmonized System (GHS) rules for classification of severe irritants (UN 2003) that resulted in the reclassification of two proposed reference substances from nonsevere to severe irritants
- Reassignment of the candidate reference substances to chemical classes using MeSH (NLM 2005)

The accuracy and reliability reanalyses and the revised reference substances list for validation of *in vitro* tests to detect ocular corrosives and severe irritants were presented in a BRD Addendum that was released on July 26, 2005, with notification of its release through the ICCVAM electronic mailing list and via an *FR* notice (Vol. 70, No. 142, p. 43149; available at <http://iccvam.niehs.nih.gov/methods/eyeirrit.htm>). The BRD Addendum was made available in electronic format on the ICCVAM/NICEATM website and from NICEATM on request.

The Expert Panel was subsequently reconvened via teleconference on September 19, 2005 to discuss the BRD Addendum. Prior to this meeting, public comments on the Addendum were received from three organizations and provided to the Expert Panel for their consideration (<http://iccvam.niehs.nih.gov/methods/ocudocs/addendcomm.htm>). The Expert Panel provided formal comment on each of the four *in vitro* test methods, as well as the proposed list of reference substances. In addition, the public were provided time at the public meeting to comment (although no public comments were provided). The Expert Panel then provided final endorsement regarding the impact, if any, of the information in the BRD Addendum on their original evaluation from the January 11-12, 2005 meeting. The availability of *The ICCVAM Expert Panel Evaluation of the Draft Background Review Document for In Vitro*

Test Methods For Identifying Ocular Corrosives and Severe Irritants - Addendum was announced via an *FR* notice (Vol. 70, No. 211, p. 66451; available at <http://iccvam.niehs.nih.gov/methods/eyeirrit.htm>) on November 2, 2005.

Subsequently, the draft BRDs and the draft BRD Addendum, the Expert Panel report and its addendum, and all public comments were made available to the SACATM for their consideration at their meeting on December 12, 2005. The SACATM agreed with the conclusions of the Expert Panel.

The ICCVAM and OTWG considered the Expert Panel report and its addendum (**Appendix A**), the revised accuracy and reliability analyses (see **Appendix B** for accuracy analyses results), all public comments, and the comments of SACATM in preparing the final test method recommendations that are provided in this report. This report will be made available to the public and provided to U.S. Federal agencies for consideration, in accordance with the ICCVAM Authorization Act of 2000 (42 U.S.C. § 2851-2, 2851-5 [2000]; available at <http://iccvam.niehs.nih.gov/about/PL106545.pdf>). Agencies with applicable testing regulations and guidelines (**Appendix C**) must respond to ICCVAM within 180 days of receiving the ICCVAM recommendations. These responses will be made available to the public on the ICCVAM website (<http://iccvam.niehs.nih.gov>) as they are received.

2.0 THE BCOP TEST METHOD

2.1 BCOP Technical Summary

The following technical summary provides a synopsis of the performance analysis described in the BCOP BRD, which reviewed the available data and information for the test method.⁶ The BRD describes the current validation status of the BCOP test method, including what is known about its reliability and accuracy, the scope of the substances tested, and a standardized protocol. The BRD may be obtained from the ICCVAM/NICEATM website (<http://iccvam.niehs.nih.gov/>).

2.1.1 Test Method Description

The BCOP test method is an organotypic model that provides short-term maintenance of normal physiological and biochemical function of the bovine cornea in an isolated system. In this test method, damage by the test substance is assessed by quantitative measurements of changes in corneal opacity and permeability with an opacitometer and an ultraviolet/visible (UV/VIS) spectrophotometer, respectively. Both measurements are used to calculate an *In Vitro* Irritancy Score, which is used to assign an *in vitro* irritancy classification for prediction of the *in vivo* ocular irritation potential of a test substance. Although histopathological data could not be formally evaluated by ICCVAM, a histopathological assessment can be included on a case-by-case basis to discriminate borderline cases (i.e., substances that produce results that preclude assignment to a single category) or to identify ocular damage that does not produce opacity or permeability changes in the isolated cornea.⁷ Histopathology also is used for chemical classes or formulations that are not well characterized in the BCOP assay, where the mode of action cannot be easily predicted, when delayed effects might be anticipated, or when a more complete characterization of damage is needed.

The BCOP test method protocols used in the various studies are similar, but not identical.⁸ Variations in the publicly available BCOP protocols include different instrumentation to evaluate opacity, different decision criteria (i.e., prediction models) or *in vitro* classification systems, and differences in the use of positive controls, among other methodological variations. The essential principles of the test method protocol include isolating and culturing the bovine cornea, treating the isolated cornea with a test substance, collecting opacity and permeability data, and evaluating the data in relation to a prediction model. However, given the various uses and applications of the BCOP test method by different investigators and laboratories, and the evolution of the test method over time, a number of laboratory-specific differences have been noted regarding the conduct of the test method.

⁶Comparison of the performance analysis for BCOP to the other three *in vitro* test methods evaluated can be reviewed in **Section 6.0** and **Appendix B**.

⁷For the studies discussed here, histopathological endpoints were not evaluated or incorporated into the accuracy assessment.

⁸For additional information on this evaluation, please see the BCOP BRD (http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm#bcop).

2.1.2 Validation Database

A total of 158 substances in eight studies were used to evaluate BCOP test method accuracy. These substances represented a variety of chemical and product classes (ICCVAM 2006a). The chemical classes tested included alcohols, heterocyclic compounds, carboxylic acids, ketones, esters, inorganic salts, ethers, hydrocarbons, amines, and onium compounds. The product classes tested included solvents, surfactants, chemical/synthetic intermediates, drugs/pharmaceuticals/therapeutic agents, petroleum products, cleaners, personal care cleansers, hair shampoos, pesticides, plasticizers, reagents, bactericides, and insect repellents.

2.1.3 Test Method Accuracy

Based on all available data, the BCOP test method has an overall accuracy of 79% (113/143)⁹ to 81% (119/147), when compared to *in vivo* rabbit eye test method data classified according to the EPA (1996), European Union (EU; 2001), or GHS (UN 2003) classification systems. Furthermore, the BCOP test method has an overall false positive rate of 19% (20/103) to 21% (22/103) and an overall false negative rate of 16% (7/43) to 25% (10/40), when compared to *in vivo* rabbit eye test method data classified according to the EPA (1996), EU (2001), or GHS (UN 2003) classification systems.

There were some notable trends in the performance of the BCOP test method among substances grouped according to chemical class and/or physicochemical properties (**Table 2-1**). The chemical classes of substances that were most consistently overpredicted (i.e., were false positives) by the BCOP test method, according to the GHS classification system are alcohols (53%, 8/15) and ketones (40%, 4/10). With regard to physical form, liquids (26%, 18/68) appear more likely than solids (10%, 2/20) to be overpredicted by the BCOP test method.

Alcohols (67%, 2/3) also were most often underpredicted (i.e., were false negatives) by the BCOP test method, according to the GHS classification system. With regard to physical form, solids (42%, 5/12) appear more likely than liquids (4%, 1/24) to be underpredicted by the BCOP test method. There was no definitive difference among the underpredicted substances for which pH information was available.

BCOP test method performance statistics also were evaluated when substances from the classes that gave the most discordant results were excluded (i.e., alcohols, ketones, solids). When using the GHS classification system, exclusion of alcohols and ketones individually resulted in small changes in the performance statistics. However, exclusion of solids from the data set caused a four-fold decrease in the false negative rate from 16% (7/43) to 4% (1/29). When both alcohols and ketones were excluded, the accuracy increased from 81% (119/147) to 88% (103/117) and the false positive rate decreased from 20% (21/104) to 12% (9/77). The largest changes were observed when all three discordant classes were excluded from the data set; accuracy increased to 92% (78/85), the false positive rate decreased to 12% (7/58), and the false negative rate decreased to 0% (0/27).

⁹The numbers in parentheses represent the data used to calculate the percentages noted.

Table 2-1 False Positive and False Negative Rates of the BCOP Test Method, by Chemical Class and Properties of Interest, for the GHS Classification System

Category	N ¹	False Positive Rate ²		False Negative Rate ³	
		%	No. ⁴	%	No.
Overall	147	20	21/104	16	7/43
Chemical Class⁵					
Alcohol	18	53	8/15	67	2/3
Amine/Amidine	8	0	0/4	0	0/4
Carboxylic acid	15	38	3/8	14	1/7
Ester	12	12	1/8	0	0/4
Ether/Polyether	6	0	0/5	0	0/1
Heterocyclic	12	33	2/6	17	1/6
Hydrocarbon	12	8	1/12	-	0/0
Inorganic salt	5	0	0/3	0	0/2
Ketone	10	40	4/10	-	0/0
Onium compound	11	0	0/3	0	0/8
Properties of Interest					
Liquids	92	26	18/68	4	1/24
Solids	32	10	2/20	42	5/12
Pesticide	8	33	1/3	40	2/5
Surfactant – Total ⁶	35	5	1/21	7	1/14
-nonionic	5	0	0/4	0	0/1
-anionic	3	0	0/2	100	1/1
-cationic	6	0	0/1	0	0/5
pH – Total ⁷	28	-	-	21	5/24
- acidic (pH < 7.0)	11	-	-	18	2/11
- basic (pH > 7.0)	15	-	-	23	3/13
- equals 7	2	-	-	-	-
Category 1 Subgroup ⁸ - Total	38 ¹⁰	-	-	18	7/38
- 4 (CO=4 at any time)	20	-	-	15	3/20
- 3 (severity/persistence)	1	-	-	0	0/1
- 2 (severity)	4	-	-	25	1/4
- 2-4 combined ⁹	25	-	-	16	4/25
- 1 (persistence)	13	-	-	23	3/13

Abbreviations: BCOP = Bovine Corneal Opacity and Permeability; CO = corneal opacity; GHS = Globally Harmonized System (UN 2003).

¹N = number of substances.

²False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*.

³False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.

⁴Data used to calculate the percentage.

⁵Chemical classes included in this table are represented by at least five substances tested in the BCOP test method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).

⁶Combines single chemicals labeled as surfactants along with surfactant-containing formulations.

⁷Total number of GHS Category 1 substances for which pH information was obtained.

⁸NICEATM-defined subgroups assigned based on the lesions that drove classification of a GHS Category 1 substance. 1: based on lesions that are persistent; 2: based on lesions that are severe (not including CO=4); 3: based on lesions that are severe (not including CO=4) and persistent; 4: CO=4 at any time.

⁹Subcategories 2 to 4 combined to allow for a direct comparison of GHS Category 1 substances classified *in vivo* based on some lesion severity component and those classified based on persistent lesions alone.

¹⁰The number of substances evaluated in the Category 1 subgroup analysis may be less than the total number of *in vivo* Category 1 substances evaluated since some substances could not be classified into the subgroups used in the evaluation.

Finally, the underpredicted substances were more likely to be classified *in vivo* (according to the GHS classification system) based on persistent lesions, rather than on severe lesions. However, three substances that caused severe lesions *in vivo* (corneal opacity=4) were false negatives in BCOP.

The performance statistics for the EPA and EU classification systems are similar to those discussed for the GHS classification system. Additional information on the performance characteristics of the BCOP test method for the EPA and EU classification systems can be obtained from **Section 6.0, Appendix B**, and the BCOP BRD.

2.1.4 Test Method Reliability (Inter- and Intra-Laboratory Reproducibility)

Quantitative BCOP test method data were available for replicate corneas within individual experiments or for replicate experiments within an individual laboratory for three studies. Therefore, an evaluation of the intralaboratory repeatability and reproducibility of the BCOP test method could be conducted. Intralaboratory repeatability of *In Vitro* Irritancy Scores was assessed by analyzing two studies for substances predicted as severe eye irritants (*In Vitro* Scores ≥ 55.1). For 16 substances of varying irritancy evaluated in one study, the median coefficient of variation (CV) for *In Vitro* Irritancy Scores for replicate corneas (n=3) ranged from 11.8% to 14.2%. In a second study, the range of mean and median CV values for *In Vitro* Irritancy Scores for replicate corneas (n=4) was 1.1% to 13% for five substances predicted as severe irritants.

A CV analysis of intralaboratory data (*In Vitro* Irritancy Scores) from two studies indicated the following intralaboratory reproducibility of the BCOP test method for substances predicted as severe eye irritants. In one study, the between experiment (n=3) mean and median CV values for permeability values were 33.4% and 29.0%, respectively, for 25 surfactant-based personal care cleaning formulations. In the second study, the between experiment mean CV values of *In Vitro* Irritancy Scores for 16 substances tested two or more times in three laboratories ranged from 12.6% to 14.8%, while the median CV values ranged from 6.7% to 12.4%.

Additionally, comparable BCOP data were available for multiple laboratories within each of three comparative validation studies, which allowed for an evaluation of the interlaboratory reproducibility of the BCOP test method. For these studies, interlaboratory reproducibility was evaluated qualitatively based on the ocular irritancy classification assigned to each substance by each laboratory, and quantitatively using *In Vitro* Irritancy Scores. In the qualitative assessment of interlaboratory reproducibility of hazard classification category, 67% to 94% of the substances were classified the same by the participating laboratories. Substances with less than complete agreement in the testing laboratories include those representing such chemical classes as alcohols, ketones, and heterocyclic compounds, and such product classes as solvents, surfactants, chemical intermediates, and pesticides. A quantitative evaluation of interlaboratory reproducibility was conducted for these three studies by performing a CV analysis of *In Vitro* Irritancy Scores obtained for substances tested in multiple laboratories. In one study, the 17 substances predicted as severe in the BCOP assay had mean and median CV values of 36% and 17%, respectively, for results obtained in either 11 or 12 laboratories. In a second study, the 32 substances predicted as

severe in the BCOP assay had mean and median CV values of 25% and 22%, respectively, for results obtained in five laboratories. In a third study, the mean and median CV values for the *In Vitro* Irritancy Scores of the 16 substances were 32.4% and 22.8%, respectively, for results obtained in three laboratories.

Finally, the interlaboratory correlation between BCOP test method endpoint data generated by each laboratory was determined for 60 substances, as well as for various subsets of test substances (water-soluble, water-insoluble, surfactants, solids, solutions, and liquids). This analysis yielded a range of correlation coefficients for the subsets of test substances. Interlaboratory correlation coefficients for the *In Vitro* Irritancy Score generally spanned a range of 0.867 to 0.958 depending on the specific subsets of substances being evaluated.

2.2 ICCVAM Recommendations for the BCOP Test Method

2.2.1 Use of the BCOP Test Method

ICCVAM recognizes that the BCOP test method is not proposed as a stand alone replacement for the *in vivo* rabbit eye test method currently used for regulatory classification and labeling. ICCVAM concludes that there are sufficient data to support the use of the BCOP test method, in appropriate circumstances and with certain limitations, as a screening test to identify substances as ocular corrosives and severe irritants (i.e., EPA Category I, UN GHS Category 1, EU R41) in a tiered-testing strategy, as part of a weight-of-evidence approach.¹⁰

The identified limitations for this test method are based on the false negative and false positive rates that are observed for certain chemical and physical classes. Based on the available database, the false negative rates for alcohols and solids range from 67% (2/3) to 100% (2/2) and 42% (5/12) to 50% (5/10), respectively, depending on the hazard classification system. Additionally, the false positive rates for alcohols, ketones, and solids range from 50% (7/14) to 56% (9/16), 40% (4/10), and 10% (2/20 to 2/21), respectively, depending on the hazard classification system. When substances within these chemical and physical classes are excluded from the database, the accuracy of BCOP across the EU, EPA, and GHS classification systems ranges from 87% (72/83) to 92% (78/85) and the false negative and false positive rates range from 0% (0/27) to 12% (3/26) and 12% (7/58) to 16% (9/56), respectively.

A tiered-testing strategy for ocular irritation/corrosion (e.g., as described in the Globally Harmonized System of Classification and Labelling of Chemicals; UN 2003) allows for the use of validated and accepted *in vitro* methods prior to the use of animals for ocular safety testing. In a tiered testing strategy, when a positive result is obtained in an appropriately validated *in vitro* test, a test substance may be classified as an ocular hazard without testing in rabbits. A substance that tests negative in the *in vitro* ocular toxicity test would need to be tested in the *in vivo* ocular test to identify possible *in vitro* false negatives and to identify moderate and mild ocular irritants. As is appropriate for any test system, there is the

¹⁰The recommendations are based on the performance results for BCOP without the use of histopathology for decision making purposes.

opportunity for confirmatory testing if false positive results are suggested based on a weight-of-evidence evaluation of supplemental information (e.g., pH, structure-activity relationships, other testing data). Using *in vitro* data in a tiered-testing strategy with a weight-of-evidence decision process to classify substances as ocular corrosives or severe irritants will avoid the potential pain and distress that might be experienced by rabbits who otherwise would have been administered these test substances. A tiered-testing strategy may not be applicable to purposes other than regulatory classification and labeling.

Users should be aware that BCOP's performance characteristics could be revised as additional data become available. For example, the current validation database did not allow for adequate evaluation of all chemical or product classes (e.g., formulations). Additional data may allow for further evaluation of this, as well as other, chemical and product classes. Therefore, prior to initiation of BCOP studies, investigators are encouraged to consult the ICCVAM/NICEATM website (see <http://iccvam.niehs.nih.gov/methods/eyeirrit.htm>) to review the most current validation database, overall performance characteristics, and chemical and physical class performance characteristics. Evaluation of the most current information will allow users to determine the appropriateness of this test method for evaluating substances that are within a specific chemical, physical, or product classes.

2.2.2 BCOP Test Method Protocol

ICCVAM recommends that when testing is conducted, the BCOP test method protocol should be based on the BCOP standardized test method protocol provided in **Appendix D**. This will facilitate collection of consistent data and expand the current validation database. Exceptions and/or changes to the proposed standardized test method protocol should be accompanied by a scientific rationale. Users should be aware that the test method protocol could be revised based on future optimization and/or validation studies. ICCVAM, therefore, recommends that test method users consult the ICCVAM/NICEATM website to ensure use of the most current recommended test method protocol (<http://iccvam.niehs.nih.gov/methods/eyeirrit.htm>).

2.2.3 Optimization of the Current BCOP Test Method Protocol

The current ICCVAM recommendations are focused on the use of the BCOP test method as a screening test for ocular corrosives and severe irritants (see **Section 2.2.1**). For that use, the current test method protocol should be sufficient. To further the use of this test method and to evaluate the use of the BCOP test method as a potential replacement for the *in vivo* rabbit eye test method or for the identification of mild and moderate ocular irritants (e.g., EPA Category II, III, and IV; GHS Category 2; EU R36), ICCVAM recommends additional studies be considered and undertaken to decrease the false positive rate of this test method.

A histopathological evaluation of the corneal tissue, using a standardized scoring scheme, should be conducted. Such data will allow for the development of standardized decision criteria and a more comprehensive evaluation of the usefulness of this endpoint for classifying and labeling substances, especially those that may otherwise produce borderline or false negative results

Studies should be conducted to evaluate the impact of using a corneal holder that maintains normal corneal curvature (e.g., the corneal mounting system designed by Ubels et al. 2002) on accuracy and/or reliability of the BCOP test method.

ICCVAM also recommends that an evaluation be conducted on the effect of modifying various test method protocol components (e.g., duration of test substance exposure) on the accuracy and/or reliability of the BCOP test method.

[This Page Intentionally Left Blank]

3.0 THE ICE TEST METHOD

3.1 ICE Technical Summary

The following technical summary provides a synopsis of the performance analysis described in the ICE BRD, which reviewed the available data and information for the test method.¹¹ The BRD describes the current validation status of the ICE test method, including what is known about its reliability and accuracy, the scope of the substances tested, and a standardized protocol. The BRD may be obtained from the ICCVAM/NICEATM website (<http://iccvam.niehs.nih.gov/>).

3.1.1 Test Method Description

The ICE test method is an organotypic model that provides short-term maintenance of the chicken eye in an isolated system. In this test method, damage by the test substance is assessed by determination of corneal swelling, opacity, and fluorescein retention. While the latter two parameters involve a subjective assessment, analysis of corneal swelling provides an objective measurement. This objective measure potentially provides improved precision and reduced interlaboratory variability compared to the traditional *in vivo* rabbit eye test, which relies only on subjective measurements. Each measurement is either converted into a quantitative score used to calculate an overall Irritation Index, or assigned a qualitative categorization that is used to assign an *in vitro* irritancy classification. Either of these outcomes can then be used to predict the *in vivo* ocular irritation potential of a test substance. A histopathological assessment also can be included on a case-by-case basis to discriminate borderline cases (i.e., substances that produce results that preclude assignment to a single category).

The ICE test method protocols used in the various studies are similar, but not identical.¹² The primary difference among these protocols was the number of treated eyes per test substance. Acceptable ranges for negative control responses, historical data used to establish these ranges, and procedures to determine the optimum quantity of test substance to be applied have not been published.

3.1.2 Validation Database

A total of 154 substances in five studies were used to evaluate ICE test method accuracy. These substances represent a variety of chemical and product classes (ICCVAM 2006b). The chemical classes tested included, but were not limited to, acyl halides, alcohols, alkalis, amines/amidines, carboxylic acids, esters, heterocyclic, hydrocarbons, inorganic salts, ketones, onium compounds, and organophosphates. Commercial products or formulations tested included, but were not limited to, detergents, pesticides, silicone powder, ink, solvents, surfactants, toilet cleaners, and thermal paper coatings.

¹¹Comparison of the performance analysis for ICE to the other three *in vitro* test methods evaluated can be reviewed in **Section 6.0** and **Appendix B**.

¹²For additional information on this evaluation, please see the ICE BRD (http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm#ice).

3.1.3 Test Method Accuracy

Based on all available data, the ICE test method has an overall accuracy of 83% (120/144) to 87% (134/154), an overall false positive rate of 6% (7/122) to 8% (9/114 to 9/116), and an overall false negative rate of 41% (13/32) to 50% (15/30), when compared to *in vivo* rabbit eye test method data classified according to the EPA (1996), EU (2001), or GHS (UN 2003) classification systems.

There were some notable trends in the performance of the ICE test method among substances grouped according to chemical class and/or physicochemical properties (**Table 3-1**). The chemical class of substances that was most consistently overpredicted (i.e., were false positives) by the ICE test method according to the GHS classification system is alcohols (50%, 5/10). With regard to physical form, liquids (10%, 9/90) appear more likely than solids (0%, 0/24) to be overpredicted by the ICE test method.

No single chemical class was prominently represented among 15 substances that were underpredicted. Five of the 15 underpredicted substances were unclassified coded substances and three were carboxylic acids. No other chemical class was represented more than twice. However, these studies do suggest that surfactants or formulations containing surfactants (e.g., detergents) (56%, 5/9) may be underpredicted by the ICE test method. They also suggest that pesticides (60%, 3/5) may be underpredicted.

With regard to physical form, eight of the 15 underpredicted substances were liquids while seven were solids. However, considering that the total number of solids (36) in the database is much smaller than the number of liquids (108), solids, with a false negative rate of 58% (7/12), appear more likely to be underpredicted than liquids, with a false negative rate of 44% (8/18).

ICE test method performance statistics also were evaluated when substances from the classes that gave the most discordant results were excluded (i.e., alcohols, surfactants, solids). When using the GHS classification system, exclusion of surfactants and solids individually resulted in small changes in the performance statistics. However, exclusion of alcohols from the data set caused a two-fold decrease in the false positive rate from 8% (9/114) to 4% (4/104). When both alcohols and surfactants were excluded, the false positive rate decreased from 8% (9/114) to 4% (4/92). The largest changes were observed when all three discordant classes were excluded from the data set; accuracy increased from 83% (120/144) to 92% (69/75), the false negative rate decreased from 50% (15/30) to 29% (2/7), and the false positive rate decreased from 8% (9/114) to 6% (4/68).

Among the eight underpredicted substances for which pH information was available, four were acidic (pH <7.0) and four were basic (pH >7.0). Basic substances (8) occupy a smaller proportion of the total database than acidic substances (12), and were more often underpredicted (50% vs. 33%). However, pH information was obtained for only 20 of the 30 total Category 1 substances.

Finally, the underpredicted substances were more likely to be classified *in vivo* based on persistent lesions (according to the GHS classification system) than on severe lesions.

Table 3-1 False Positive and False Negative Rates of the ICE Test Method, by Chemical Class and Properties of Interest, for the GHS Classification System

Category	N ¹	False Positive Rate ²		False Negative Rate ³	
		%	No. ⁴	%	No.
Overall	144	8	9/114	50	15/30
<i>Chemical Class⁵</i>					
Alcohol	12	50	5/10	50	1/2
Amine/Amidine	5	0	0/2	33	1/3
Carboxylic acid	10	0	0/3	43	3/7
Ester	9	13	1/8	0	0/1
Heterocyclic	9	0	0/3	33	2/6
Onium compound	8	0	0/2	33	2/6
<i>Properties of Interest</i>					
Liquids	108	10	9/90	44	8/18
Solids	36	0	0/24	58	7/12
Pesticide	11	0	0/6	60	3/5
Surfactant – Total	21	0	0/12	56	5/9
-nonionic	4	0	0/3	100	1/1
-anionic	2	0	0/1	100	1/1
-cationic	7	0	0/1	33	2/6
pH – Total ⁶	20	-	-	40	8/20
- acidic (pH < 7.0)	12	-	-	33	4/12
- basic (pH > 7.0)	8	-	-	50	4/8
Category 1 Subgroup ⁷					
- Total	23 ⁹	-	-	35	8/23
- 4 (CO=4 at any time)	12	-	-	33	4/12
- 3 (severity/persistence)	2	-	-	50	1/2
- 2 (severity)	4	-	-	0	0/4
- 2-4 combined ⁸	18	-	-	28	5/18
- 1 (persistence)	5	-	-	60	3/5

Abbreviations: CO = corneal opacity; GHS = Globally Harmonized System (UN 2003); ICE = Isolated Chicken Eye.

¹N = number of substances.

²False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.

³False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*.

⁴Data used to calculate the percentage.

⁵Chemical classes included in this table are represented by at least five substances tested in the ICE test method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).

⁶Total number of GHS Category 1 substances for which pH information was obtained.

⁷NICEATM-defined subgroups assigned based on the lesions that drove classification of a GHS Category 1 substance. 1: based on lesions that are persistent; 2: based on lesions that are severe (not including CO=4); 3: based on lesions that are severe (not including CO=4) and persistent; 4: CO=4 at any time.

⁸Subcategories 2 to 4 combined to allow for a direct comparison of GHS Category 1 substances classified *in vivo* based on some lesion severity component and those classified based on persistent lesions alone.

⁹The number of substances evaluated in the Category 1 subgroup analysis may be less than the total number of *in vivo* Category 1 substances evaluated since some substances could not be classified into the subgroups used in the evaluation.

However, four substances that caused severe lesions *in vivo* (corneal opacity=4) were false negatives in ICE.

The performance statistics for the EPA and EU classification systems are similar to those discussed for the GHS classification system. Additional information on the performance characteristics of the ICE test method for the EPA and EU classification systems can be obtained from **Section 6.0, Appendix B**, and the ICE BRD.

3.1.4 Test Method Reliability (Inter- and Intra-Laboratory Reproducibility)

Data were received that allowed for a quantitative analysis of intralaboratory repeatability and reproducibility of ICE test method endpoints. The range of CV values for the corneal thickness measurement, when results were compared within experiments, was from 0.9% to 6.1%. The other endpoints evaluated produced ranges of CV values that were larger, with variability most prominent with the nonirritating substance. However, this could be an exaggeration of variability given the relatively small values that were produced from the nonirritating substance relative to the irritating and corrosive substances (i.e., corneal swelling values of 2, 0, and 3 yield a higher CV than values of 11, 14, and 18). A similar discussion also can be applied to the variability among the qualitative endpoints (i.e., corneal opacity and fluorescein retention) given the small dynamic range of their scores (0-4 or 0-3, respectively). The range of CV values for the corneal thickness measurement, when results were compared across experiments, was from 1.8% to 6.3%. The CV values for the remaining endpoints had a larger range (e.g., corneal swelling CV = 13.9% to 138.7%). However, if the nonirritating substance is removed, the range of CV values is reduced (e.g., corneal swelling CV = 13.9% to 22.4%).

One interlaboratory comparative study involving four laboratories contained sufficient ICE test data on 59 substances for an assessment of interlaboratory reproducibility. Based on a qualitative analysis, 60% to 70% of the substances classified as ocular corrosives or severe irritants, depending on the regulatory classification system employed (i.e., EPA 1996, GHS [UN 2003], EU 2001), were correctly identified by all four participating laboratories. A CV analysis of these same data indicated that the mean and median CV for severe substances tested was less than 35% for all test method endpoints, with the exception of corneal swelling.

3.2 **ICCVAM Recommendations for the ICE Test Method**

3.2.1 Use of the ICE Test Method

ICCVAM recognizes that the ICE test method is not proposed as a stand alone replacement for the *in vivo* rabbit eye test method currently used for regulatory classification and labeling. ICCVAM concludes that there are sufficient data to support the use of the ICE test method, in appropriate circumstances with certain limitations, as a screening test to identify substances as ocular corrosives and severe irritants (i.e., EPA Category I, UN GHS Category 1, EU R41) in a tiered-testing strategy, as part of a weight-of-evidence approach.

The identified limitations for this method are based on the false negative and false positive rates that are observed for certain chemical and physical classes. Based on the available

database, the false negative rates for alcohols, surfactants and solids range from 33% (1/3) to 50% (1/2), 44% (4/9) to 57% (4/7), and 46% (6/13) to 70% (7/10), respectively, depending on the hazard classification system. Additionally, the false positive rates for alcohols range from 27% (3/11) to 50% (5/10), depending on the hazard classification system evaluated. When substances within these chemical and physical classes are excluded from the database, the accuracy of ICE across the EU, EPA, and GHS classification systems ranges from 91% (72/79 to 75/82) to 92% (69/75) and the false negative and false positive rates range from 29% (2/7) to 33% (3/9) and 5% (4/73) to 6% (4/68 to 4/70), respectively.

A tiered-testing strategy for ocular irritation/corrosion (e.g., as described in the Globally Harmonized System of Classification and Labelling of Chemicals; UN 2003) allows for the use of validated and accepted *in vitro* methods prior to the use of animals for ocular safety testing. In a tiered testing strategy, when a positive result is obtained in an appropriately validated *in vitro* test, a test substance may be classified as an ocular hazard without testing in rabbits. A substance that tests negative in the *in vitro* ocular toxicity test would need to be tested in the *in vivo* ocular test to identify possible *in vitro* false negatives and to identify moderate and mild ocular irritants. As is appropriate for any test system, there is the opportunity for confirmatory testing if false positive results are suggested based on a weight-of-evidence evaluation of supplemental information (e.g., pH, structure-activity relationships, other testing data). Using *in vitro* data in a tiered-testing strategy with a weight-of-evidence decision process to classify substances as ocular corrosives or severe irritants will avoid the potential pain and distress that might be experienced by rabbits who otherwise would have been administered these test substances. A tiered-testing strategy may not be applicable to purposes other than regulatory classification and labeling.

Users should be aware that ICE's performance characteristics could be revised as additional data become available. For example, the current validation database did not allow for adequate evaluation of all chemical or product classes (e.g., formulations). Additional data may allow for further evaluation of this, as well as other, chemical and product classes. Therefore, prior to initiation of ICE studies, investigators are encouraged to consult the ICCVAM/NICEATM website (see <http://iccvam.niehs.nih.gov/methods/eyeirrit.htm>) to review the most current validation database, overall performance characteristics, and chemical and physical class performance characteristics. Evaluation of the most current information will allow users to determine the appropriateness of this test method for evaluating substances that are within a specific chemical, physical, or product classes.

3.2.2 ICE Test Method Protocol

ICCVAM recommends that when testing is conducted, the ICE test method protocol should be based on the ICE standardized test method protocol provided in **Appendix E**. This will facilitate collection of consistent data and expand the current validation database. Exceptions and/or changes to the proposed standardized test method protocol should be accompanied by a scientific rationale. Users should be aware that the test method protocol could be revised based on future optimization and/or validation studies. ICCVAM, therefore, recommends that test method users consult the ICCVAM/NICEATM website to ensure use of the most current recommended test method protocol (<http://iccvam.niehs.nih.gov/methods/eyeirrit.htm>).

3.2.3 Optimization of the Current ICE Test Method Protocol

The current ICCVAM recommendations are focused on the use of the ICE test method as a screening test for ocular corrosives and severe irritants (see **Section 3.2.1**). For that use, the current test method protocol should be sufficient. To further the use of this test method and to evaluate the use of the ICE test method as a potential replacement for the *in vivo* rabbit eye test method or for the identification of mild and moderate ocular irritants (e.g., EPA Category II, III, and IV; GHS Category 2; EU R36), ICCVAM recommends additional studies be considered and undertaken.

Additional optimization studies/evaluations should be conducted in an attempt to decrease the 29% to 33% false negative rate of the ICE test method. After optimization, additional studies to further assess the reliability and accuracy of the test method are recommended.

ICCVAM recommends that a histopathological evaluation of the corneal tissue, using a standardized scoring scheme, be included when the ICE test method is conducted. Such data will allow for development of decision criteria and future assessments on the usefulness of this endpoint for classifying and labeling substances, especially those that may otherwise produce borderline or false negative results.

ICCVAM also recommends that centering lights be installed on the optical pachymeter, which is used to measure corneal thickness, to ensure consistent central corneal thickness measurements across laboratories.

4.0 THE IRE TEST METHOD

4.1 IRE Technical Summary

The following technical summary provides a synopsis of the performance analysis described in the IRE BRD, which reviewed the available data and information for the test method.¹³ The BRD describes the current validation status of the IRE test method, including what is known about its reliability and accuracy, the scope of the substances tested, and a standardized protocol. The BRD may be obtained from the ICCVAM/NICEATM website (<http://iccvam.niehs.nih.gov/>).

4.1.1 Test Method Description

The IRE test is an organotypic model that provides short-term maintenance of normal physiological and biochemical function of the entire rabbit eye in an isolated system. In this test method, damage by the test substance is assessed by determination of corneal swelling, corneal opacity, fluorescein retention, and effects on the corneal epithelium. Identification of severe ocular irritants and corrosives is based on reaching or exceeding predetermined cut-off values in any one of the four endpoints (e.g., product of the corneal opacity and area scores ≥ 3 ; product of area and intensity scores for fluorescein penetration ≥ 4 ; corneal swelling $\geq 25\%$; or any significant effect on corneal epithelium (pitting, mottling, stippling, ulceration) (See **Appendix F** for details).

The IRE test method protocols used in the various studies are similar, but not identical.¹⁴ Examples of some of the test method components that differed among the IRE protocols used to generate data include:

- temperature of solution used to rinse solids from the eyes ranged from room temperature to 32°C,
- amount of substance applied as a solid ranged from 25 mg to 100 mg, and
- decision criteria used for classification of substances was based on scores from two to four endpoints.

4.1.2 Validation Database

A total of 149 substances were evaluated in three studies, of which 25 were commercial products or formulations (ICCVAM 2006c). The chemical classes tested included, but were not limited to, alcohols, amides, amines, carboxylic acids, esters, ethers, formulations, heterocyclic, ketones, onium compounds, and sulfur compounds. The commercial products or formulations tested were skin cleansers, soaps, shampoos, conditioners, surfactants, and solvents.

¹³Comparison of the performance analysis for IRE to the other three *in vitro* test methods evaluated can be reviewed in **Section 6.0** and **Appendix B**.

¹⁴For additional information on this evaluation, please see the IRE BRD (http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm#ire).

4.1.3 Test Method Accuracy

The overall accuracy (based on the pooled data set¹⁵) for the IRE test method ranged from 64% (68/107) to 69% (79/114) when compared to the *in vivo* test method data classified according to the GHS (UN 2003), EPA (1996), and EU (2001) regulatory classification systems. The overall false positive rates, when compared to these regulatory classification systems, ranged from 35% (23/65) to 40% (25/62). The overall false negative rates, when compared to the three regulatory classification systems, ranged from 24% (12/49) to 31% (14/45).

There were some trends in the performance of the IRE test method among substances grouped according to chemical class and/or physicochemical properties (**Table 4-1**). The chemical classes that were consistently overpredicted (i.e., false positives), when compared to classifications based on the GHS classification system, were alcohols (55%, 6/11), amines (50%, 3/6), and ketones (67%, 4/6). The chemical classes that were underpredicted (i.e., false negatives), when compared to classifications based on the GHS classification system, were carboxylic acids (67%, 4/6) and organic compounds (50%, 3/6).

With regard to physical form, liquids have a higher false positive rate (49%, 18/37) when compared to solids (22%, 5/23) for the IRE test method. The false negative rates for liquids and solids were relatively similar (29%, 8/28 vs. 32%, 6/19; respectively).

A subset of the substances evaluated had pH information available. For these substances, the overall false positive rate was 24% (4/17) and the overall false negative rate was 0% (0/10).

Of the surfactant-based formulations evaluated by this test method, the false positive rate was 25% (2/8) and the false negative rate was 38% (6/16). Comparatively, for substances identified as surfactants in the database, the false positive rate was 40% (2/5) and the false negative rate was 12% (1/8).

Finally, the underpredicted substances were more likely to be classified *in vivo* (according to the GHS classification) system based on persistent lesions, rather than severe lesions. However, three substances that caused severe lesion *in vivo* (corneal opacity=4) were false negatives.

The performance statistics for the EPA and EU classification systems are similar to those discussed for the GHS classification system. Additional information on the performance characteristics of the IRE test method for the EPA and EU classification systems can be obtained from **Section 6.0, Appendix B**, and the IRE BRD.

4.1.4 Test Method Reliability (Inter- and Intra-Laboratory Reproducibility)

Due to the lack of available quantitative IRE test method data for replicate eyes within individual experiments or for replicate experiments within an individual laboratory, an

¹⁵The pooled dataset represents the results from all the available studies combined, regardless of the number of endpoints evaluated by each of the individual studies. Additional information about this dataset can be obtained from the IRE BRD.

Table 4-1 False Positive and False Negative Rates of the IRE Test Method, by Chemical Class and Properties of Interest, for the GHS Classification System (Analysis Based on the Pooled Data Set)

Category	N ¹	False Positive Rate ²		False Negative Rate ³	
		%	No. ⁴	%	No.
Overall	107	38	23/60	30	14/47
Chemical Class⁵					
Alcohol	13	55	6/11	50	1/2
Amide	5	0	0/3	0	0/2
Amine	11	50	3/6	20	1/5
Carboxylic acid	12	33	2/6	67	4/6
Ester	10	30	3/10	-	0/0
Ether	9	33	2/6	0	0/3
Formulation	24	25	2/8	38	6/16
Heterocycle	18	44	4/9	11	1/9
Ketone	6	67	4/6	-	0/0
Onium compound	10	33	1/3	0	0/7
Organic	12	17	1/6	50	3/6
Sulfur compound	8	20	1/5	33	1/3
Properties of Interest					
Liquid/Solution	65	49	18/37	29	8/28
Solids	42	22	5/23	32	6/19
Surfactant-based formulation	24	25	2/8	38	6/16
Surfactants	13	40	2/5	12	1/8
-nonionic	4	33	1/3	0	0/1
-anionic	2	0	0/1	100	1/1
-cationic	7	100	1/1	0	0/6
pH – Total ⁶	27	24	4/17	0	0/10
-acidic	18	20	2/10	0	0/8
-basic	7	33	2/6	0	0/1
-equals 7	2	0	0/1	0	0/1
Category 1 Subgroup ⁷ - Total	37 ⁹	-	-	32	12/37
- 4 (CO=4 at any time)	11	-	-	27	3/11
- 3 (severity/persistence)	4	-	-	25	1/4
- 2 (severity)	3	-	-	33	1/3
- 2-4 combined ⁸	18	-	-	28	5/18
- 1 (persistence)	19	-	-	37	7/19

Abbreviations: CO = corneal opacity; GHS = Globally Harmonized System (UN 2003); IRE = Isolated Rabbit Eye.

¹N = number of substances.

²False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*.

³False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.

⁴Data used to calculate the percentage.

⁵Chemical classes included in this table are represented by at least five substances tested in the IRE test method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).

⁶Total number of GHS Category 1 substances for which pH information was obtained.

⁷NICEATM-defined subgroups assigned based on the lesions that drove classification of a GHS Category 1 substance. 1: based on lesions that are persistent; 2: based on lesions that are severe (not including CO=4); 3: based on lesions that are severe (not including CO=4) and persistent; 4: CO = 4 at any time.

⁸Subcategories 2 to 4 combined to allow for a direct comparison of GHS Category 1 substances classified *in vivo* based on some lesion severity component and those classified based on persistent lesions alone.

⁹The number of substances evaluated in the Category 1 subgroup analysis may be less than the number of *in vivo* Category 1 substances evaluated, since some substances could not be classified into the subgroups used in the evaluation.

evaluation of the intralaboratory repeatability and reproducibility of the IRE test method could not be conducted. However, two studies contained sufficient IRE test data (n=59 and 21 substances, respectively) for an assessment of interlaboratory reproducibility based on data reported for three or four different laboratories. For these studies, interlaboratory reproducibility was evaluated qualitatively based on the ocular irritancy classification assigned to each substance by each laboratory and quantitatively using corneal opacity, swelling in one study, and corneal opacity, corneal swelling and evaluation of fluorescein penetration in the second study.

Based on a qualitative analysis, 100% of the 12 to 18 substances were correctly identified as severe irritants or ocular corrosives in the IRE by all four participating laboratories, when compared to *in vivo* rabbit eye test data classification dependent on the regulatory classification system employed (i.e., EPA 1996, GHS [UN 2003], EU 2001). Substances with less than complete agreement in the testing laboratories include those representing such chemical classes as alcohols, ketones, and heterocyclic compounds; and such product classes as organic solvents, surfactants, chemical intermediates, and pesticides.

A quantitative evaluation of interlaboratory reproducibility was conducted for these two studies by performing a CV analysis of corneal opacity, swelling, and, for the second study, fluorescein penetration measurements for substances tested in multiple laboratories. The CV analysis of the first study indicated that the median CV for all 59 substances tested was between 43.4% and 49.7% for the 4-hour corneal opacity and the 4-hour swelling endpoints, respectively. The CV was between 33.6% and 35.5% when only severe irritants are considered. In the second study using corneal opacity, swelling, and fluorescein penetration, the median CV for all substances ranged from 24.0% to 40.0% (the largest variability was for corneal swelling) and from 15.4% to 35.5% when only severe irritants were considered.

4.2 ICCVAM Recommendations for the IRE Test Method

4.2.1 Use of the IRE Test Method

Based on the accuracy (64% [68/107] to 69% [79/114]), false negative (24% [12/49] to 31% [14/45]), and false positive (35% [23/65] to 40% [25/62]) rates across the EU, EPA, and GHS classification systems, the use of the IRE test method for screening and identifying ocular corrosives and severe irritants (i.e., EPA Category I, GHS Category 1, EU R41) in a tiered-testing strategy, as part of a weight-of-evidence approach, is not recommended. There also are insufficient data using all four recommended IRE endpoints (corneal opacity, fluorescein penetration, corneal swelling, and observations of significant effect on corneal epithelium) to assess test method accuracy and reliability when all these endpoints are evaluated in a single study.

Users should be aware that IRE's performance characteristics could be revised as additional data become available. Therefore, prior to initiation of non-regulatory, validation, or optimization IRE studies, investigators are encouraged to consult the ICCVAM/NICEATM website (see <http://iccvam.niehs.nih.gov/methods/eyeirrit.htm>) to review the most current validation database, overall performance characteristics, and chemical and physical class performance characteristics. Evaluation of the most current information will allow users to

determine the appropriateness of this test method for evaluating substances that are within a specific chemical, physical, or product classes.

4.2.2 IRE Test Method Protocol

When non-regulatory, validation, or optimization studies are conducted using the IRE test method, the protocol should be based on the standardized protocol provided in **Appendix F**. This will facilitate collection of consistent data and expand the current validation database. Exceptions and/or changes to the test method protocol should be accompanied by a scientific rationale.

Users should be aware that IRE's standardized test method protocol could be revised as additional data become available. Therefore, prior to initiation of IRE studies, investigators are encouraged to consult the ICCVAM/NICEATM website (see <http://iccvam.niehs.nih.gov/methods/eyeirrit.htm>) to review the most current recommended standardized test method protocol.

ICCVAM recommends that, for all studies, raw data be collected and maintained. The availability of such data will allow for further retrospective evaluation of test method accuracy and/or reliability.

4.2.3 Optimization of the Current IRE Test Method Protocol

ICCVAM recommends that additional evaluation studies be conducted to increase the current IRE database and optimize the IRE test method decision criteria. Once these studies are conducted, ICCVAM recommends that additional validation studies be conducted to further evaluate the relevance and reliability of the IRE test method.

ICCVAM recommends that a histopathological evaluation of the corneal tissue, using a standardized scoring scheme, be included when the IRE test method is conducted. Such data will allow for development of decision criteria and future assessments on the usefulness of this endpoint for classifying and labeling substances, especially those that may otherwise produce borderline or false negative results.

ICCVAM also recommends that centering lights be installed when an optical pachymeter is used to measure corneal thickness, to ensure consistent central corneal thickness measurements across laboratories.

[This Page Intentionally Left Blank]

5.0 THE HET-CAM TEST METHOD

5.1 HET-CAM Technical Summary

The following technical summary provides a synopsis of the performance analysis described in the HET-CAM BRD, which reviewed the available data and information for the test method.¹⁶ The BRD describes the current validation status of the HET-CAM test method, including what is known about its reliability and accuracy, the scope of the substances tested, and a standardized protocol. The BRD may be obtained from the ICCVAM/NICEATM website (<http://iccvam.niehs.nih.gov/>).

5.1.1 Test Method Description

The HET-CAM test method uses the chorioallantoic membrane (CAM), which is a vascular fetal membrane, composed of the fused chorion and allantois. It was assumed that acute effects induced by a test substance on the small blood vessels and proteins of this soft tissue membrane are similar to effects induced by the same test substance in the eye of a treated rabbit. The CAM has been proposed as a model for a living membrane (such as the conjunctiva) since it comprises a functional vasculature. Additionally, evaluation of coagulation (i.e., protein denaturation) may reflect corneal damage that may be produced by the test substance. The CAM is evaluated for the development of irritant endpoints (hyperemia, hemorrhage, and coagulation). Depending on the method used to collect data on the endpoints (time to development, severity of observed effect) qualitative assessments of the irritation potential of test substances are made.

The HET-CAM test method protocols used in the various studies evaluated are similar, but not identical. Examples of some of the test method components that differed among the HET-CAM protocols used to generate data include:

- relative humidity during egg incubation ranged from 52.5% to 62.5%,
- volume or quantity of the test substance applied to the CAM (when reported) was either 0.1 mL or 0.3 mL for liquids and 0.3 g for solids,
- number of replicate eggs per test substance ranged from three to six, and
- some studies included concurrent positive control substances, while others did not.

5.1.2 Validation Database

There were several HET-CAM analysis methods used by the various studies.¹⁷ For the Irritation Score (IS)(A)¹⁸ and IS(B)¹⁹ analysis methods, data were available to conduct additional sub-analyses (ICCVAM 2006d). For these sub-analyses, substances tested at a 10% concentration or 100% concentration *in vitro* were compared to responses observed at a 100% concentration tested *in vivo* (e.g., IS(A)-10, IS(B)-10, IS(B)-100).

¹⁶Comparison of the performance analysis for HET-CAM to the other three *in vitro* test methods evaluated can be reviewed in **Section 6.0** and **Appendix B**.

¹⁷For additional information on this evaluation, please see the HET-CAM BRD (http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm#hetcam).

¹⁸Analysis method described in Luepke (1985).

¹⁹Analysis method described in Kalweit et al. (1987).

A total of 24 and 20 substances were evaluated for the IS(A)-10 and IS(A)-100 analysis methods, respectively, using the decision criteria of Luepke (1985). For the IS(B)-10 and IS(B)-100 analysis methods, using the decision criteria of Luepke (1985), 101 and 138 substances were evaluated, respectively. The chemical classes tested included, but were not limited to, alcohols, amines, esters, ethers, formulations, heterocyclic compounds, inorganic salts, ketones, and organic salts. The product classes tested included, but were not limited to, cosmetics, solvents, shampoos, flavor ingredients, and pharmaceutical synthetics.

5.1.3 Test Method Accuracy

For the IS(A) analysis method, accuracy increased when substances were evaluated at *in vitro* were tested at 100% concentration compared to the 10% concentration and where *in vivo* data were classified according to the EPA (1996), EU (2001), and GHS (UN 2003) classification systems. The opposite pattern was observed for the IS(B) analysis method; test method accuracy increased when substances were evaluated *in vitro* at 10% concentration (IS(B)-10) compared to the 100% concentration (IS(B)-100) and where *in vivo* data were classified according to the EPA (1996), EU (2001), and GHS (UN 2003) classification systems.

Chemical classes that were overpredicted by the HET-CAM IS(B) analysis methods, when testing substances at either a 10% or at 100% concentration, include alcohols (IS(B)-10: 89% [8/9]; IS(B)-100: 88% [14/16]), ethers (IS(B)-10: 50% [5/10]; IS(B)-100: 50% [6/12]), amines (IS(B)-10: 60% [3/5]; IS(B)-100: 83% [5/6]), organic salts (IS(B)-10: 57% [4/7]; IS(B)-100: 86% [6/7]), and heterocyclic compounds (IS(B)-10: 86% [6/7]; IS(B)-100: 78% [7/9]). Formulations appeared to have the lowest false positive rates for both IS(B)-10 and IS(B)-100 (Table 5-1). Chemical classes that were underpredicted by both analysis methods were amines and ethers.

An evaluation based on the physical form of the test substance *in vivo* depended on the analysis method being evaluated. For the IS(B)-100 analysis method, substances tested as solids *in vivo* had a false positive rate of 67% (16/24) and substances tested as liquids *in vivo* had a false positive rate of 65% (33/51) (Table 5-1). For the IS(B)-100 analysis method, substances tested as liquids *in vivo* had a false negative rate of 0% (0/9) and substances tested as solids *in vivo* had a false negative rate of 24% (4/17). For the IS(B)-10 analysis method, liquids had a false positive rate of 19% (3/16) and false negative rate of 37% (7/19) while solids had false positive and false negative rates of 58% (11/19) and 13% (1/8), respectively.

An analysis of the ability of the HET-CAM test method to identify ocular corrosives and severe irritants, depending on the nature of the *in vivo* ocular lesions (i.e., severity and/or persistence) responsible for classification of a substance as an ocular corrosive/severe irritant, indicated that, for IS(B)-10, the underpredicted substances were more likely to be substances classified as corrosive or severely irritating *in vivo* based on persistent lesions, with a false negative rate of 37% (10/27) compared to 15% (2/13) for substances classified as corrosive or severely irritating *in vivo* based on severity. For the IS(B)-100 analysis method, the underpredicted substances were more likely to be substances classified as corrosive or severely irritating *in vivo* based on severe lesions, with a false negative rate of 11% (2/19)

Table 5-1 False Positive and False Negative Rates of the HET-CAM Test Method, by Chemical Class and Properties of Interest, for the GHS Classification System

Category	N ¹	False Positive Rate ²		False Negative Rate ³	
		%	No. ⁴	%	No.
Overall IS(B)-10 (Entire database)	101	33	20/61	30	12/40
Overall IS(B)-100 (Entire database)	138	59	58/99	13	5/39
<i>Chemical Class-IS(B)-10⁵</i>					
Alcohol	16	89	8/9	25	2/7
Aldehyde	5	0	0/4	100	1/1
Amine	7	60	3/5	50	1/2
Ether	14	50	5/10	50	2/4
Formulation	24	0	0/8	44	7/16
Heterocyclic Compound	7	86	6/7	-	0/0
Organic salt	7	57	4/7	-	0/0
<i>Chemical Class-IS(B)-100⁵</i>					
Alcohol	24	88	14/16	13	1/8
Aldehyde	6	80	4/5	0	0/1
Amine	9	83	5/6	33	1/3
Carboxylic acid/Carboxylic acid salt	11	60	3/5	17	1/6
Ester	12	90	9/10	0	0/2
Ether	16	50	6/12	25	1/4
Formulation	27	26	6/23	0	0/4
Heterocyclic Compound	12	78	7/9	33	1/3
Inorganic salt	5	100	2/2	0	0/3
Ketone	6	67	4/6	-	0/0
Organic salt	9	86	6/7	0	0/2
<i>Properties of Interest</i>					
Physical Form: IS(B)-10					
Liquids/Solutions	35	19	3/16	37	7/19
Solids	27	58	11/19	13	1/8
Unknown	39	23	6/26	31	4/13
Physical Form: IS(B)-100					
Liquids	60	65	33/51	0	0/9
Solids	41	67	16/24	24	4/17
Unknown	37	38	9/24	8	1/13
Surfactant – Total IS(B)-100	2	50	1/2	-	0/0
-nonionic	2	50	1/2	-	0/0
-anionic	0	-	-	-	-
-cationic	0	-	-	-	-
Surfactant-Based Formulation – IS(B)-10	24	0	0/8	44	7/16
pH – IS(B)-10⁶ - acidic (pH < 7.0)	35	58	11/19	13	2/16

Category	N ¹	False Positive Rate ²		False Negative Rate ³	
		%	No. ⁴	%	No.
- basic (pH > 7.0)	24	50	7/14	20	2/10
	11	80	4/5	0	0/6
pH – IS(B)-100 ⁶	35	68	13/19	13	2/16
- acidic (pH < 7.0)	23	69	9/13	10	1/10
- basic (pH > 7.0)	12	67	4/6	17	1/6
Category 1 Subgroup- IS(B)-10⁷					
- Total	40	-	-	30	12/40
- 4 (CO=4 at any time)	13	-	-	15	2/13
- 3 (severity/persistence)	0	-	-	-	-
- 2 (severity)	0	-	-	-	-
- 2-4 combined ⁸	13	-	-	15	2/13
- 1 (persistence)	27	-	-	37	10/27
Category 1 Subgroup- IS(B)-100⁷					
- Total	38 ⁹	-	-	11	4/38
- 4 (CO=4 at any time)	19	-	-	11	2/19
- 3 (severity/persistence)	1	-	-	100	1/1
- 2 (severity)	2	-	-	0	0/2
- 2-4 combined ⁸	22	-	-	14	3/22
- 1 (persistence)	16	-	-	6	1/16

Abbreviations: CO = corneal opacity; GHS = Globally Harmonized System (UN 2003); HET-CAM = Hen’s Egg Test – Chorioallantoic Membrane.

¹N=number of substances.

²False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*.

³False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.

⁴Data used to calculate percentages.

⁵Chemical classes included in this table are represented by at least five substances tested in the HET-CAM test method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).

⁶Total number of GHS Category 1 substances for which pH information was obtained.

⁷NICEATM-defined subgroups assigned based on the lesions that drove classification of a GHS Category 1 substance. 1: based on lesions that are persistent; 2: based on lesions that are severe (not including CO=4); 3: based on lesions that are severe (not including CO=4) and persistent; 4: CO = 4 at any time.

⁸Subcategories 2 to 4 combined to allow for a direct comparison of GHS Category 1 substances classified *in vivo* based on some lesion severity component and those classified based on persistent lesions alone.

⁹The number of substances evaluated in the Category 1 subgroup analysis may be less than the number of *in vivo* Category 1 substances evaluated, since some substances could not be classified into the subgroups used in the evaluation.

compared to 6% (1/16) for substances classified as corrosive or severely irritating *in vivo* based on persistence. However, two substances that were classified based on severe lesions (i.e., CO=4) were underpredicted by the HET-CAM IS(B)-10 and IS(B)-100 analysis methods.

The performance statistics for the EPA and EU classification systems are similar to those discussed for the GHS classification system. Additional information on the performance characteristics of the HET-CAM test method for the EPA and EU classification systems can be obtained from **Section 6.0, Appendix B**, and the HET-CAM BRD.

5.1.4 Test Method Reliability (Inter- and Intra-Laboratory Reproducibility)

The analysis of intralaboratory repeatability was evaluated using data from two different publications for the IS(B) analysis method. In both studies, the hemorrhage endpoint had the highest CV value (109.10%-117.56%). The CV values for the coagulation endpoint ranged from 41.78% to 95.69%. The difference in the numbers may be due to several factors including test substances evaluated and differences in the test method protocols used between the two studies. The calculated variability for the endpoints and the overall test method may be exaggerated because of the relatively small dynamic ranges for each of the endpoints (0.02 to 5 for hemorrhage, 0.02 to 7 for lysis, and 0.03 to 9 for coagulation). Similar results were obtained from the analysis of intralaboratory reproducibility.

A qualitative analysis of interlaboratory reliability also was conducted. For the IS(B)-10 analysis method, the participating laboratories were in 100% agreement for 84 to 85 (79% to 81%) of 104 to 107 substances evaluated, when compared to all three hazard classification systems. For the IS(B)-100 analysis method, the participating laboratories in a study were in 100% agreement for 80 to 81 (82% to 84%) of the 95 to 99 substances evaluated, when compared to all three hazard classification systems. There was 100% agreement in regard to the ocular irritancy classification for 11 (64% to 69%) of the 16 to 17 substances evaluated in five laboratories using the IS(A) analysis method, when compared to all three hazard classification systems.

The overall reliability statistics, arranged by HET-CAM data analysis method, were consistent with what was observed for the individual studies evaluated. For the IS(B)-10, the statistics were identical to what was discussed previously. For the IS(A) and IS(B)-100 analysis methods, additional data laboratory data was available for a subset of the substances tested for each analysis method. For both of these analysis methods, the addition of the results from additional testing laboratories yielded a concordance pattern consistent with that described above.

A quantitative evaluation of interlaboratory reproducibility was conducted for the same analysis methods. For one study, two different evaluations were conducted based on the concentration tested *in vitro* using the IS(B) analysis method. For 14 substances evaluated at 100% concentration, the mean and median CV values were 31.86% and 33.04%, respectively. In the same study, for 12 substances evaluated at 10% concentration, the mean and median CV values were 66.29% and 60.75%, respectively. For the substances evaluated in another study which used the IS(B) analysis method, the mean and median CV values for substances tested at 10% concentration were 60.17% and 42.65%, respectively. For substances tested at 100% concentration in the same study, the mean and median CV values were lower: 35.21% and 26.22%, respectively. When substances that were tested in three different testing laboratories (instead of two) were removed from the assessment, little change was seen in the mean and median CV values for both concentrations tested. For a study using the IS(A) analysis method, the mean and median CV for substances classified as GHS Category 1 (UN 2003) were 26.09% and 27.08%, respectively. The mean and median CV for substances classified as EPA Category I (EPA 1996) were 25.86% and 26.43%, respectively.

5.2 ICCVAM Recommendations for the HET-CAM Test Method

5.2.1 Use of the HET-CAM Test Method

ICCVAM evaluated several HET-CAM analysis methods proposed for identifying substances that are ocular corrosives or severe irritants. These included one analysis method termed the IS(B)-10 and another analysis method termed IS(B)-100. The range of hazard classification accuracy rates across the EU, EPA, and GHS classification systems for these two analysis methods ranged from 65% (64/98) to 68% (69/101) for IS(B)-10 and 52% (69/133) to 57% (94/164) for IS(B)-100, when the decision criteria of Luepke (1985) were used. The overall false negative and false positive rates of the IS(B)-10 analysis method range from 30% (10/33 to 12/40) to 32% (10/31) and 33% (20/61) to 36% (24/67), respectively, depending on the classification system. The overall false negative and false positive rates for the IS(B)-100 analysis method range from 6% (2/33) to 13% (5/39) and 52% (68/131) to 59% (58/99), respectively, depending on the classification system. Based on these rates, the use of these analyses methods and decision criteria for screening and identifying ocular corrosives and severe irritants (i.e., EPA Category I, GHS Category 1, EU R41) in a tiered-testing strategy, as part of a weight-of-evidence approach, is not recommended.

Users should be aware that HET-CAM's performance characteristics could be revised as additional data become available. Therefore, prior to initiation of non-regulatory, validation, or optimization HET-CAM studies, investigators are encouraged to consult the ICCVAM/NICEATM website (see <http://iccvam.niehs.nih.gov/methods/eyeirrit.htm>) to review the most current validation database, overall performance characteristics, and chemical and physical class performance characteristics. Evaluation of the most current information will allow users to determine the appropriateness of this test method for evaluating substances that are within a specific chemical, physical, or product classes.

5.2.2 HET-CAM Test Method Protocol

When non-regulatory, validation, or optimization studies are conducted using the HET-CAM test method, the protocol should be based on the standardized protocol provided in **Appendix G**. This will facilitate collection of consistent data and expand the current validation database. Exceptions and/or changes to the test method protocol should be accompanied by a scientific rationale.

Users should be aware that HET-CAM's standardized test method protocol could be revised as additional data become available. Therefore, prior to initiation of HET-CAM studies, investigators are encouraged to consult the ICCVAM/NICEATM website (see <http://iccvam.niehs.nih.gov/methods/eyeirrit.htm>) to review the most current recommended standardized test method protocol.

ICCVAM recommends that, for all studies, raw data be collected and maintained. The availability of such data will allow for further retrospective evaluation of test method accuracy and/or reliability.

5.2.3 Optimization of the Current HET-CAM Test Method Protocol

ICCVAM recommends that additional studies should be conducted to further optimize the HET-CAM prediction models and the decision criteria (e.g., mtc10) that would be used to identify ocular corrosives and severe irritants for the EPA, GHS, or EU classification systems. Such studies could potentially improve the usefulness of the HET-CAM test method for identifying severe ocular irritants and corrosives and its possible future use for the identification of mild and moderate ocular irritants (e.g., EPA Category II, III, and IV; GHS Category 2; EU R36).

[This Page Intentionally Left Blank]

6.0 GENERAL RECOMMENDATIONS AND COMPARISON OF PERFORMANCE CHARACTERISTICS FOR FOUR EVALUATED *IN VITRO* TEST METHODS

In addition to the test method specific recommendations discussed in Sections 2.0 through 5.0, ICCVAM also makes some general recommendations that relate to all the *in vitro* test methods discussed.

Table 6-1 provides a comparison of the accuracy, false positive, and false negative rates for all four *in vitro* ocular toxicity test methods evaluated for each of the regulatory hazard classification systems evaluated (EPA, EU, and GHS). As noted in the sections discussing each of the test methods individually (Sections 2.0 through 5.0), these performance characteristics are similar among the three hazard classification systems.

Although both BCOP and ICE can be used as screens for the detection of ocular corrosives and severe irritants in a tiered testing strategy, as part of a weight-of-evidence approach, both test methods as well as HET-CAM and IRE have limitations. As shown in Table 6-1, exclusion of specific chemical and physical classes increases the accuracy and decreases the false positive and false negative rates for BCOP and ICE. ICCVAM recommends that users consider, to the extent possible, the chemical and physical structures of the substances to be tested to determine whether either of these test methods would be appropriate to use as a screening test for ocular corrosion or severe irritation. Also, additional studies with each test method are recommended to determine if modification of the test method standardized protocol and/or the decision criteria for classification of a test substance as a corrosive/severe irritant or as a nonsevere irritant/nonirritant can improve test method sensitivity and specificity.

Results from appropriately validated *in vitro* ocular toxicity test methods are recommended for use in a weight-of-evidence decision making process in accordance with the EPA and EU ocular testing regulations (EPA 1996, EU 2004) and the GHS tiered-testing strategy (UN 2003)²⁰. In these testing schemes, when a positive result is obtained in an appropriately validated *in vitro* test, a test substance may be classified as an ocular hazard without testing in rabbits. A substance that tests negative in the *in vitro* ocular toxicity test would need to be tested in the *in vivo* ocular test to identify possible *in vitro* false negatives and to identify moderate and mild ocular irritants. As is appropriate for any test system, there is the opportunity for confirmatory testing if false positive results are indicated based on a weight-of-evidence evaluation of supplemental information (e.g., structure-activity relationships, other testing data). Use of a weight-of-evidence decision making process and a tiered-testing strategy for classification of substances as ocular corrosives or severe irritants will eliminate the pain and distress that might be experienced by rabbits who otherwise would have been administered these test substances.

²⁰ A tiered-testing strategy approach may not be applicable to purposes other than regulatory classification and labeling.

Table 6-1 Comparison of Performance Characteristics of Four *In Vitro* Ocular Test Methods for the Identification of Severe Ocular Irritants or Corrosives, for Three Hazard Classification Systems

Test Method	Database	EPA Classification System				EU Classification System				GHS Classification System			
		N ¹	Accuracy (%) ²	False Positive Rate ³ (%)	False Negative Rate ⁴ (%)	N ¹	Accuracy (%) ²	False Positive Rate ³ (%)	False Negative Rate ⁴ (%)	N ¹	Accuracy (%) ²	False Positive Rate ³ (%)	False Negative Rate ⁴ (%)
BCOP	All	143	79 (113/143)	19 (20/103)	25 (10/40)	143	80 (114/143)	21 (22/103)	18 (7/40)	147	81 (119/147)	20 (21/104)	16 (7/43)
	Excluding alcohols, ketones, and solids	83	87 (72/83)	14 (8/57)	12 (3/26)	82	88 (72/82)	16 (9/56)	4 (1/26)	85	92 (78/85)	12 (7/58)	0 (0/27)
ICE	All	145	84 (122/145)	8 (9/116)	48 (14/29)	154	87 (134/154)	6 (7/122)	41 (13/32)	144	83 (120/144)	8 (9/114)	50 (15/30)
	Excluding alcohols, surfactants, and solids	79	91 (72/79)	6 (4/70)	33 (3/9)	82	91 (75/82)	5 (4/73)	33 (3/9)	75	92 (69/75)	6 (4/68)	29 (2/7)
IRE	Pooled Data Set	107	64 (68/107)	40 (25/62)	31 (14/45)	114	69 (79/114)	35 (23/65)	24 (12/49)	107	65 (70/107)	38 (23/60)	30 (14/47)
HET-CAM	IS(B)-10	98	65 (64/98)	36 (24/67)	32 (10/31)	95	67 (64/95)	34 (21/62)	30 (10/33)	101	68 (69/101)	33 (20/61)	30 (12/40)
	IS(B)-100	133	52 (69/133)	58 (61/105)	11 (3/28)	164	57 (94/164)	52 (68/131)	6 (2/33)	138	54 (75/138)	59 (58/99)	13 (5/39)

Abbreviations: BCOP = Bovine Corneal Opacity and Permeability; EPA = U.S. Environmental Protection Agency; EU = European Union; GHS = Globally Harmonized System (UN 2003); HET-CAM = Hen's Egg Test - Chorioallantoic Membrane; ICE = Isolated Chicken Eye; IRE = Isolated Rabbit Eye.

¹N=number of substances.

²Numbers in parentheses represent data used to calculate percentages.

³False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*.

⁴False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.

Additional research and development, optimization, and/or validation efforts should use reference substances with existing rabbit data. Additional rabbit studies should be conducted only if important data gaps are identified. If such studies are conducted, they should be designed to minimize the number of rabbits tested, to minimize or avoid pain and distress, and to maximize the information collected. Designing and conducting such studies should be in accordance with the recommendations from the Scientific Symposium on Mechanisms of Chemically-Induced Ocular Injury and the Scientific Symposium on Minimizing Pain and Distress in Ocular Safety Testing (see <http://iccvam.niehs.nih.gov/methods/ocudocs/ocumeet/sympinfo.htm>). These symposia were organized by ICCVAM, NICEATM, and ECVAM.

All raw data generated using any of the recommended standardized *in vitro* ocular testing protocols and the *in vivo* rabbit eye test on the same substance should be submitted to NICEATM to expand the available validation database for these four test methods. The availability of such data will allow for additional retrospective evaluations of test method accuracy and/or reliability. Ideally, all substances should be completely identified (e.g., chemical name, chemical class, physicochemical properties). However, if this is not possible for proprietary reasons, data may be submitted using coded labels for each substance tested. If such coding is used, as much information as possible on physical and chemical properties should be provided to NICEATM.

Although the IRE and HET-CAM test methods cannot currently be recommended for meeting regulatory testing requirements, there may be non-regulatory uses for these two test methods. Accordingly, the four *in vitro* test methods should be considered prior to conducting *in vivo* ocular testing and an alternative test method should be used where determined appropriate for the specific testing situation. Since ocular irritancy testing frequently involves more than slight or momentary pain or distress, consideration of alternative test methods prior to the use of animals is necessary to comply with provisions of U.S. Animal Welfare Act regulations (9 CFR, Part 2, Section 2.31 and 9 CFR, Part 2, Section 2.32), the Public Health Service Policy on the Humane Care and Use of Laboratory Animals (PHS 2002), and the U.S. Government Principles for the Utilization and Care of Vertebrate Animals Used in Testing, Research, and Training (National Research Council 1996).

The potential usefulness of combining two or more *in vitro* test methods in a battery to identify ocular corrosives and severe irritants should be evaluated. Currently, there is insufficient guidance on the utility of a battery approach for such determinations.

Interested stakeholders are encouraged to support research and development of alternative test methods and technologies that may provide for a more accurate assessment of ocular toxicity and/or advantages in terms of time and cost.

[This Page Intentionally Left Blank]

7.0 ICCVAM RECOMMENDATIONS ON SUBSTANCES FOR VALIDATION OF *IN VITRO* OCULAR TOXICITY TEST METHODS FOR THE EVALUATION OF OCULAR CORROSIVES AND SEVERE IRRITANTS

In addition to evaluating the validation status of four *in vitro* ocular toxicity test methods for their ability to identify ocular corrosives and severe irritants, ICCVAM developed a list of reference substances for the optimization and/or validation of *in vitro* tests to identify ocular corrosives and severe irritants. This section provides ICCVAM's recommendations on these reference substances.

ICCVAM reviewed the Expert Panel's report and addendum (provided in **Appendix A**), the results of the analysis in the BRDs, and the public comments received to both. Based on these sources, ICCVAM makes the following recommendations with relation to the list of reference substances for the optimization and/or validation of *in vitro* ocular toxicity test methods for identification of ocular corrosives and severe irritants.²¹

ICCVAM endorses the reference substances list of 122 substances. The list of substances (see **Appendix H**) includes:

- 79 GHS Category 1 substances (UN 2003); 10 of which the Category 1 classification is based solely on human data
- 28 GHS Category 2 substances (UN 2003)
 - 15 GHS Category 2A substances (moderate irritants)
 - 13 GHS Category 2B substances (mild irritants)
- 15 GHS nonirritant substances (UN 2003)
- 34 chemical classes
- 24 product classes
- 79 liquids
- 43 solids

ICCVAM further endorses the use of the reference substance list as a source for generating a subset of substances to be used for evaluating *in vitro* ocular toxicity test methods on a scientifically sound case-by-case basis. It is recommended that the subset of substances that are developed from the reference substance list comprise a scientifically sound distribution of substances among various properties including, but not limited to, chemical class, product class, physical form, irritancy severity classification, mechanism of action, physical and chemical characteristics, and molecular weight. In situations where a listed substance is not available, other substances of the same class for which there is high quality *in vivo* reference data may be used. Following completion of optimization and/or validation studies, substances from this list can be selected for inclusion in performance standards and proficiency testing (ICCVAM 2003).

²¹The recommendations discussed here are based on the ability of the *in vitro* test method to identify *in vivo* classifications based on the GHS classification system.

[This Page Intentionally Left Blank]

8.0 REFERENCES

- EPA. 1996. Label Review Manual: 2nd Ed. EPA737-B-96-001. Washington, DC: U.S. Environmental Protection Agency.
- EU. 2004. Manual of Decisions for Implementation of the Sixth and Seventh Amendments to Directive 67/548/EEC on Dangerous Substances (Directives 79/831/EEC and 92/32/EEC). European Chemicals Bureau, IHCP, European Commission DG-JRC, report in. EUR 20519 EN, updated version of July 2004.
- EU. 2001. Commission Directive 2001/59/EC of 6 August 2001 adapting to technical progress for the 28th time Council Directive 67/548/EEC on the approximation of the laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. Official Journal of the European Communities L255:1-333.
- ICCVAM. 2006a. ICCVAM Evaluation of the Bovine Corneal Opacity and Permeability Test for the Identification of Severe Ocular Irritants and Corrosives. Available: http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm#bcop [accessed 10 Jan 2006].
- ICCVAM. 2006b. ICCVAM Evaluation of the Isolated Chicken Eye Test for the Identification of Severe Ocular Irritants and Corrosives. Available: http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm#ice [accessed 10 Jan 2006].
- ICCVAM. 2006c. ICCVAM Evaluation of the Isolated Rabbit Eye Test for the Identification of Severe Ocular Irritants and Corrosives. Available: http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm#ire [accessed 10 Jan 2006].
- ICCVAM. 2006d. ICCVAM Evaluation of the Hen's Egg Test – Chorioallantoic Membrane Test for the Identification of Severe Ocular Irritants and Corrosives. Available: http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm#hetcam [accessed 10 Jan 2006].
- ICCVAM. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods. NIH Publication No: 03-4508. Research Triangle Park: National Toxicology Program.
- 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: National Toxicology Program.
- Kalweit S, Gerner I, Spielmann H. 1987. Validation project of alternatives for the Draize eye test. *Mol Toxicol* 1:597-603.
- Luepke N. 1985. Hen's egg chorioallantoic membrane test for irritation potential. *Food Chem Toxicol* 23:287-291.

National Research Council. 1996. Guide for the Care and Use of Laboratory Animals. Washington, DC: National Academy Press.

NLM. 2005. Medical Subject Heading. Available: <http://www.nlm.nih.gov/mesh> [accessed 09 January 2006].

PHS. 2002. Public Health Service (PHS) Policy on Humane Care and Use of Laboratory Animals.

Ubels JL, Paauw JD, Casterton PL, Kool DJ. 2002. A redesigned corneal holder for the bovine cornea opacity and permeability assay that maintains normal corneal morphology. *Toxicol In Vitro* 16:621-628.

UN. 2003. Globally Harmonized System of Classification and Labelling of Chemicals (GHS). New York & Geneva: United Nations Publications. Available: http://www.unece.org/trans/danger/publi/ghs/ghs_rev00/00files_e.html. [accessed 16 October 2006].