Appendix C

Final Background Review Document:
Use of the Murine Local Lymph Node Assay for Potency Categorization of Chemicals Causing Allergic Contact Dermatitis in Humans

This page intentionally left blank

Final Background Review Document: Use of the Murine Local Lymph Node Assay for Potency Categorization of Chemicals Causing Allergic Contact Dermatitis in Humans

Interagency Coordinating Committee on the Validation of Alternative Methods

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

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

2011

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

This page intentionally left blank

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Acronyms

ACD	Allergic contact dermatitis
ACE	Acetone
AOO	Acetone: olive oil (4:1 by volume)
BRD	Background review document
BT	Buehler test
Conc.	Concentration tested
CPSC	U.S. Consumer Product Safety Commission
DEP	Diethyl phthalate
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
2,4-DNCB	2,4-Dinitrochlorobenzene
dpm	Disintegrations per minute
DSA	Dose per skin area
DSA_{05}	Induction dose per skin area, in $\mu g/cm^2$, in a human repeat-insult patch test or human maximization test that produces a positive response in 5% of the tested population
EC3	Estimated concentration of a substance expected to produce a stimulation index of 3, the threshold value for a substance to be considered a sensitizer in the LLNA
ECETOC	European Centre for Ecotoxicology and Toxicology
ECPA	European Crop Protection Association
ECVAM	European Centre for the Validation of Alternative Methods
EPA	U.S. Environmental Protection Agency
EtOH	Ethanol
FDA	U.S. Food and Drug Administration
FHSA	Federal Hazardous Substances Act
FR	*Federal Register*
GCP	Good Clinical Practices
GHS	Globally Harmonized System of Classification and Labelling of Chemicals
GLP	Good Laboratory Practice
GPMT	Guinea pig maximization test
HCA	Hexyl cinnamic aldehyde
HMT	Human maximization test
HRIPT	Human repeat-insult patch test
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
ICH	International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use
IPCS	International Programme on Chemical Safety
IWG	Immunotoxicity Working Group
ISO	International Organization for Standardization

JaCVAM	Japanese Center for the Validation of Alternative Methods
K_{ow}	Estimated log octanol-water partition coefficient
LLNA	Murine local lymph node assay
LOEL	Lowest observed effect level
MEK	Methyl ethyl ketone
n	Number
NA	Not available
NC	Not calculated
NICEATM	National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods
NIEHS	National Institute of Environmental Health Sciences
NOEL	No observed effect level
NTP	National Toxicology Program
OECD	Organisation for Economic Co-operation and Development
OPPTS	Office of Prevention, Pesticides and Toxic Substances
PG	Propylene glycol
r	Correlation coefficient
R^2	Coefficient of determination
REACH	Regulation on Registration, Authorisation and Restriction of Chemicals
SACATM	Scientific Advisory Committee on Alternative Toxicological Methods
SI	Stimulation index
SLS	Sodium lauryl sulfate
TG	Test Guideline
WHO	World Health Organization

Interagency Coordinating Committee on the Validation of Alternative Methods: Agency Representatives

Agency for Toxic Substances and Disease Registry
* Moiz Mumtaz, Ph.D.
Bruce Fowler, Ph.D.
Edward Murray, Ph.D.
Eric Sampson, Ph.D.

Consumer Product Safety Commission
* Marilyn L. Wind, Ph.D. (Chair, through July 2010)
+ Kristina Hatlelid, Ph.D.
Joanna Matheson, Ph.D.

Department of Agriculture
* Jodie Kulpa-Eddy, D.V.M. (Acting Chair)
+ Elizabeth Goldentyer, D.V.M.

Department of Defense
* David Honey, Ph.D.
* Robert E. Foster, Ph.D. (to August 2010)
+ Patty Decot
Harry Salem, Ph.D. (to August 2010)
Terry Besch, D.V.M., M.S., DACLAM, DACVPM
Peter J. Schultheiss, D.V.M., DACLAM (to August 2010)

Department of Energy
* Michael Kuperberg, Ph.D.
+ Marvin Stodolsky, Ph.D.

Department of the Interior
* Barnett A. Rattner, Ph.D.
+ Sarah Gerould, Ph.D. (to Feb. 2009)

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

Environmental Protection Agency
Office of Pesticide Programs
* John R. "Jack" Fowle III, Ph.D., DABT
+ Vicki Dellarco, Ph.D.
+ Tina Levine, Ph.D.
Deborah McCall
Christine Augustyniak, Ph.D. (*U.S. Coordinator, OECD Test Guidelines Program)*
Office of Pollution Prevention and Toxics
Jerry Smrchek, Ph.D. (*U.S. Coordinator, OECD Test Guidelines Program,* to July 2009*)*
Office of Research and Development
Suzanne McMaster, Ph.D. (to Dec. 2008)
Julian Preston, Ph.D. (to July 2009)
Stephanie Padilla, Ph.D. (to July 2009)
Office of Science Coordination and Policy
Karen Hamernik, Ph.D. (to July 2009)

Food and Drug Administration
Office of the Commissioner
* Suzanne Fitzpatrick, Ph.D., DABT
Center for Biologics Evaluation and Research
Richard McFarland, Ph.D., M.D.
Ying Huang, Ph.D.
Center for Devices and Radiological Health
Melvin E. Stratmeyer, Ph.D. (to May 2010)
Vasant G. Malshet, Ph.D., DABT
Center for Drug Evaluation and Research
+ Abigail C. Jacobs, Ph.D.
Paul C. Brown, Ph.D.
Center for Food Safety and Applied Nutrition
David G. Hattan, Ph.D.
Neil Wilcox, D.V.M., M.P.H.
Robert L. Bronaugh, Ph.D. (to May 2010)
Center for Veterinary Medicine
Devaraya Jagannath, Ph.D.
M. Cecilia Aguila, D.V.M.
National Center for Toxicological Research
Paul Howard, Ph.D.
Donna Mendrick, Ph.D.
William T. Allaben, Ph.D. (to Jan. 2009)
Office of Regulatory Affairs
Lawrence D'Hoostelaere, Ph.D.

National Cancer Institute
* T. Kevin Howcroft, Ph.D.
Chand Khanna, D.V.M., Ph.D.
Alan Poland, M.D. (to Oct. 2008)

National Institute of Environmental Health Sciences
* William S. Stokes, D.V.M., DACLAM
+ Warren Casey, Ph.D., DABT
Raymond R. Tice, Ph.D.
Rajendra S. Chhabra, Ph.D., DABT
Jerrold J. Heindel, Ph.D.

National Institute for Occupational Safety and Health
* Paul Nicolaysen, V.M.D.

National Institutes of Health
* Margaret D. Snyder, Ph.D.

National Library of Medicine
* Pertti (Bert) Hakkinen, Ph.D.
+ Jeanne Goshorn, M.S.

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

* Principal agency representative
\+ Alternate principal agency representative

Acknowledgements

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Interagency Immunotoxicity Working Group (IWG)

U.S. Consumer Product Safety Commission
Joanna Matheson, Ph.D. (IWG Co-chair)
Marilyn L. Wind, Ph.D. (through July 2010)

U.S. Environmental Protection Agency
Office of Pesticide Programs
Jonathan Chen, Ph.D.
Masih Hashim, D.V.M., Ph.D.
Marianne Lewis
Deborah McCall
Timothy McMahon, Ph.D.
John Redden, M.S.
Jenny Tao, Ph.D.
Office of Pollution Prevention and Toxics
Elizabeth Margosches, Ph.D.
Ronald Ward, Ph.D.
Office of Research and Development
Marsha Ward, Ph.D.
Office of Science Coordination and Policy
Karen Hamernik, Ph.D.

U.S. Food and Drug Administration
Center for Devices and Radiological Health
Vasant G. Malshet, Ph.D., DABT
Jeffrey Toy, Ph.D.
Center for Drug Evaluation and Research
Ruth Barratt, Ph.D., D.V.M.
Paul C. Brown, Ph.D.
Abigail C. Jacobs, Ph.D. (IWG Co-chair)
Jiaqin Yao, Ph.D.
Office of Science and Health Coordination
Suzanne Fitzpatrick, Ph.D., DABT

National Institute of Environmental Health Sciences
Warren Casey, Ph.D., DABT
Dori Germolec, Ph.D.
William S. Stokes, D.V.M., DACLAM

National Institute for Occupational Safety and Health
B. Jean Meade, D.V.M., Ph.D.
Paul D. Siegel, Ph.D.

National Library of Medicine
Pertti Hakkinen, Ph.D.

European Centre for the Validation of Alternative Methods – Liaison
Silvia Casati, Ph.D.
Alexandre Angers, Ph.D.

Japanese Center for the Validation of Alternative Methods – Liaison
Hajime Kojima, Ph.D.

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

National Institute of Environmental Health Sciences

William Stokes, D.V.M., DACLAM
Director

Warren Casey, Ph.D., DABT
Deputy Director

Deborah McCarley
Special Assistant; Assistant Project Officer

NICEATM Support Contract Staff (Integrated Laboratory Systems [ILS], Inc.)

David Allen, Ph.D.
Thomas Burns, M.S.
Linda Litchfield
Steven Morefield, M.D.
Michael Paris
Eleni Salicru, Ph.D.
Catherine Sprankle
Frank Stack
Judy Strickland, Ph.D., DABT
Linda Wilson

Statistical Consultant for ILS, Inc.

Joseph Haseman, Ph.D.

Other Acknowledgements

ICCVAM and NICEATM gratefully acknowledge the following individuals and institutions for submitting data to NICEATM for the evaluation of the use of the LLNA to assess relative skin sensitization potency.

Anne Marie Api, Ph.D.
Research Institute for Fragrance Materials
Woodlake, NJ

David Basketter, Ph.D.[1]
Unilever Safety and Environmental Assurance Centre
Sharnbrook, U.K.

Phil Botham, Ph.D.
European Crop Protection Association
Brussels, Belgium

Eric Debruyne, Ph.D.
Bayer CropScience SA, Sophia Antipolis Cedex, France

G. Frank Gerberick, Ph.D.
The Procter & Gamble Company
Cincinnati, OH

Dori Germolec, Ph.D.
National Toxicology Program
Research Triangle Park, NC

Ian Kimber, Ph.D.[2]
Syngenta Central Toxicology Laboratory
Macclesfield, U.K.

Heidi Ott
Federal Institute for Occupational Safety and Health
Dortmund, Germany

Kirill Skirda, Ph.D.
TNO Quality of Life
Delft, The Netherlands

Peter Ungeheuer, Ph.D.
European Federation for Cosmetic Ingredients
Frankfurt, Germany

[1] Present affiliation: DABMEB Consultancy, Ltd., Sharnbrook, Bedfordshire, U.K.
[2] Present affiliation: The University of Manchester, Manchester, U.K.

Preface

Allergic contact dermatitis (ACD) is an adverse health effect that frequently develops in workers and consumers following repeated exposed to skin sensitizing chemicals and products. ACD results in lost workdays[3] and can significantly diminish quality of life (Hutchings et al. 2001; Skoet et al. 2003). To minimize the occurrence of ACD, regulatory authorities require testing to identify substances that may cause skin sensitization. Sensitizing substances must be labeled with a description of the potential hazard and the precautions necessary for workers and consumers to avoid development of ACD.

In 1999, the U.S. Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) recommended the murine (mouse) local lymph node assay (LLNA) as a valid test method to assess the skin sensitization potential of most types of substances (Dean et al. 2001; Haneke et al. 2001; ICCVAM 1999; Sailstad et al. 2001). ICCVAM concluded that the LLNA provides several advantages compared to guinea pig test methods, including elimination of potential pain and distress, use of fewer animals, less time required to perform, and availability of dose-response information. United States and international regulatory authorities subsequently accepted the LLNA as an alternative test method for ACD testing. It is now used around the world.

In 2007, the U.S. Consumer Product Safety Commission (CPSC) asked ICCVAM and the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) to evaluate the scientific validity of the LLNA as a stand-alone assay for potency determinations for classification purposes. CPSC based the nomination on their interest in assessing the usefulness and limitations of the LLNA for identifying chemicals and products likely to be strong human sensitizers. ICCVAM assigned the nomination a high priority and established the ICCVAM interagency Immunotoxicity Working Group (IWG). The interagency IWG and NICEATM reviewed the current literature and evaluated available data to assess the application of the LLNA for this purpose.

The interagency IWG and NICEATM prepared a comprehensive draft background review document (BRD) that provided information, data, and analyses supporting the validation status of the LLNA for potency determinations for classification purposes. ICCVAM prepared draft test method recommendations, which included the usefulness and limitations, test method protocol, and future studies relevant to this application of the LLNA. Both documents were provided to an independent international scientific peer review panel (Panel) for their consideration at a public meeting on March 4–6, 2008.

A report of the Panel meeting was subsequently published on the NICEATM-ICCVAM website.[4] The Panel and ICCVAM concluded that the LLNA should not be used as a stand-alone assay for categorizing skin sensitizers based on potency but that it can be used as part of a weight-of-evidence evaluation for this purpose. The Panel recommended that NICEATM perform additional analyses using alternative human reference values that might be more appropriate for evaluating the accuracy of the LLNA for correctly determining skin sensitization potency categories. NICEATM performed these analyses for the final BRD.

ICCVAM considered the conclusions and recommendations of the Panel, along with comments from the public and the Scientific Advisory Committee on Alternative Toxicological Methods, and then finalized the BRD and test method recommendations. These will be forwarded to Federal agencies for their consideration and acceptance decisions, where appropriate.

---

[3] http://www.bls.gov/IIF

[4] http://iccvam.niehs.nih.gov/methods/immunotox/llna_PeerPanel08.htm

We gratefully acknowledge the organizations and scientists who provided data and information for this document. We would also like to recognize the efforts of the individuals who contributed to its preparation, review, and revision. We especially recognize the Panel members for their thoughtful evaluations and generous contributions of time and effort. Special thanks are extended to Dr. Michael Luster for serving as Panel Chair and to Dr. Michael Woolhiser, Dr. Michael Olson, and Kim Headrick for their service as Evaluation Group Chairs during the March 2008 Panel meeting. We thank Drs. Abigail Jacobs (U.S. Food and Drug Administration) and Joanna Matheson (CPSC) for serving as co-chairs of the interagency IWG, as well as the members of the interagency IWG and ICCVAM representatives who reviewed and provided comments throughout the process leading to the final BRD.

Integrated Laboratory Systems, Inc., the NICEATM support contractor, provided excellent scientific and operational support, for which we thank Dr. David Allen, Thomas Burns, Linda Litchfield, Dr. Steven Morefield, Michael Paris, Dr. Eleni Salicru, Catherine Sprankle, Frank Stack, and Dr. Judy Strickland. Finally, we thank Dr. Silvia Casati and Dr. Alexandre Angers from the European Centre for the Validation of Alternative Methods and Dr. Hajime Kojima from the Japanese Center for the Validation of Alternative Methods for their participation and contributions as interagency IWG liaisons.

Jodie A. Kulpa-Eddy, D.V.M.
APHIS-VS-NCIE-Agricultural Select Agent Program
U.S. Department of Agriculture
Acting Chair, ICCVAM

Rear Admiral William S. Stokes, D.V.M., DACLAM
Assistant Surgeon General, U.S. Public Health Service
Director, NICEATM
Executive Director, ICCVAM

Executive Summary

Background

Allergic contact dermatitis (ACD) is an adverse health effect that frequently develops in workers and consumers following repeated exposed to skin sensitizing chemicals and products. ACD results in lost workdays[5] and can significantly diminish quality of life (Hutchings et al. 2001; Skoet et al. 2003). To minimize the occurrence of ACD, regulatory authorities require testing to identify substances that may cause skin sensitization. Sensitizing substances must be labeled with a description of the potential hazard and the precautions necessary for workers and consumers to avoid development of ACD.

In 1999, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) recommended to U.S. Federal agencies that the murine local lymph node assay (LLNA) is a valid substitute for currently accepted guinea pig test methods to assess the ACD hazard potential of many, but not all, types of substances (Dean et al. 2001; Haneke et al. 2001; ICCVAM 1999; Sailstad et al. 2001). The recommendation was based on a comprehensive evaluation that included assessment of the LLNA's validation status by an independent international scientific peer review panel (Panel). The Panel and ICCVAM recommendations (ICCVAM 1999) are available at the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)-ICCVAM website.[6] The LLNA was subsequently incorporated into national and international test guidelines for the assessment of skin sensitization: U.S. Environmental Protection Agency Health Effects Test Guidelines on Skin Sensitization (EPA 2003), Organisation for Economic Co-operation and Development Test Guideline 429 (OECD 2002), and International Organization for Standardization (ISO) 10993-10: Tests for Irritation and Delayed-type Hypersensitivity (ISO 2002).[7]

In 2007, the U.S. Consumer Product Safety Commission formally requested that ICCVAM and NICEATM evaluate several activities related to the LLNA.[8] One of the nominated activities was an assessment of the validation status of the LLNA as a stand-alone assay for potency determinations for regulatory classification purposes. The information described in this background review document (BRD) was compiled by ICCVAM and NICEATM in response to this nomination. The BRD provides a comprehensive review of data and information regarding the usefulness and limitations of the LLNA as a stand-alone assay for determining potency (including severity) for the purpose of hazard classification.

Test Method Protocol

The LLNA test method identifies potential skin sensitizers by quantifying lymphocyte proliferation in the draining auricular lymph nodes during the induction phase of skin sensitization. The magnitude of lymphocyte proliferation then correlates with the extent to which sensitization develops after topical exposure to the potential skin sensitizer. For the purposes of this analysis, relative potency in the LLNA is defined as the concentration of a fixed volume of a substance that is required for the induction phase of a skin sensitization reaction to occur. The more potent the substance the smaller the quantity needed.

The recently updated ICCVAM-recommended test method protocol for the LLNA describes the conduct of the assay in detail (**Appendix B**). A test substance-induced increase in lymphocyte proliferation in the draining lymph nodes of the ear, the site of application, is used to identify chemical sensitizers. Mice are injected with radiolabeled thymidine (or an analogue of thymidine),

---

[5] http://www.bls.gov/IIF

[6] http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna/llnarep.pdf

[7] The OECD and ISO test guidelines were updated in 2010 (ISO 2010; OECD 2010a).

[8] http://iccvam.niehs.nih.gov/methods/immunotox/llnadocs/CPSC_LLNA_nom.pdf

which is incorporated into the DNA of proliferating cells. The stimulation index (SI), the ratio of incorporated radioactivity in the auricular lymph nodes of treated versus control mice, is used to assess the sensitizing potential of the test substance. An SI of 3 or greater is used to classify a test substance as a skin-sensitizing agent. In the LLNA, a volume of 25 μL of the test substance is applied to each ear, and the estimated concentration expected to produce an SI of 3 (i.e., the EC3) is used as the metric for predicting skin-sensitization potency.

Validation Database

The information summarized in this BRD is based on a review of data from the LLNA. Data were obtained from published reports and unpublished data submitted to NICEATM in response to a *Federal Register* notice (72 FR 27815).[9] The information includes LLNA, guinea pig, and human data derived from a database of over 600 substances, 196 of which have LLNA data with comparative guinea pig and/or human data. These 196 substances include 136 substances with comparative human data (76 sensitizers, 60 nonsensitizers), 116 substances with comparative guinea pig data (64 sensitizers, 52 nonsensitizers), and 56 substances with comparative human and guinea pig data (35 sensitizers, 21 nonsensitizers).

The reference database for this evaluation consisted of human clinical studies, the human maximization test (HMT) and the human repeat-insult patch test (HRIPT) and, for nonsensitizers, other published reports. In the HMT and the HRIPT, potency information is determined from the no observed effect level (NOEL), the lowest observed effect level (LOEL), or the induction dose per skin area (DSA) that produces a positive response in 5% of the tested population (DSA_{05}). The third revised edition of the Globally Harmonized System of Classification and Labelling of Chemicals (GHS) classifies skin sensitizers as Category 1 (UN 2009). Category 1 substances are further subcategorized into Subcategory 1A ("strong" skin sensitizers) or Subcategory 1B ("other" skin sensitizers) based on results from human and/or animal studies (i.e., the LLNA and guinea pig tests). According to the GHS, substances with positive responses in the HMT or HRIPT at induction thresholds $\leq$500 μg/cm^2 are classified as Subcategory 1A, and substances with positive responses at >500 μg/cm^2 are classified as Subcategory 1B.

The GHS also includes criteria to use the LLNA to subcategorize sensitizers based on the EC3 value. Substances that produce an EC3 $\leq$ 2% are classified as Subcategory 1A (strong sensitizers), and substances with EC3 > 2% are classified as Subcategory 1B (other sensitizers) (UN 2009). Nonsensitizers are not classified.

Usefulness of the LLNA in Predicting Skin Sensitization Potency in Humans

The extent to which the LLNA correctly classifies strong versus other than strong human skin sensitizers was evaluated using the database of 136 substances for which both LLNA and human data were available. First, linear regression analyses using LLNA EC3 versus DSA_{05} values were conducted to establish a positive correlation, and to determine the optimum comparison based on either the most potent LLNA EC3 and DSA_{05} or the geometric mean LLNA EC3 and DSA_{05} for substances with multiple test results. Based on the higher R^2 value (0.448 versus 0.382) achieved when geometric means of multiply tested substances were used, this approach was carried forward in the performance analyses.

The correct, under- and overclassification rates of the LLNA versus human data for these 136 substances were initially calculated using the GHS criteria: EC3 $\leq$ 2% to classify substances as strong sensitizers and EC3 > 2% to classify substances as other sensitizers. The LLNA correctly identified 52% (14/27) of the strong human sensitizers using EC3 $\leq$ 2% but underclassified 48% (13/27). Among the 21 substances that produced an EC3 $\leq$ 2%, 67% (14/21) were strong human skin sensitizers (GHS Subcategory 1A), but the remaining 33% (7/21) were either other human skin

[9] http://iccvam.niehs.nih.gov/SuppDocs/FedDocs/FR/FR_E7_9544.pdf

sensitizers (GHS Subcategory 1B, n = 3) or substances not classified as human skin sensitizers (n = 4).

Of the 13 strong human sensitizers that were not categorized as strong sensitizers using the LLNA EC3 ≤ 2%, 11 produced an EC3 > 2% (10/11 had an EC3 between 2% and 10% and one produced an EC3 of 30.9%), and two were negative in the LLNA. The commonalities among these 13 substances with regard to physicochemical characteristics include molecular weights within a range of 100 g/mole (12/13 substances had molecular weights of 98.15 to 192.3 g/mole); 8/13 substances were liquids; and, of the six substances for which peptide reactivity information was available, all had high (n = 5) or moderate (n = 1) peptide reactivity.

As noted above, most (77% [10/13]) of the strong human sensitizers that are underclassified by the LLNA produced EC3 values from 2% to 10%. Using LLNA EC3 ≤ 10% to classify substances as strong sensitizers correctly classified 89% (24/27) of the strong sensitizers compared with the 52% (14/27) of the strong sensitizers correctly classified using EC3 ≤ 2%.

For the 56 substances that had LLNA, guinea pig (i.e., the guinea pig maximization test and/or the Buehler test), and human skin sensitization data, the overall correct classification rate produced by the LLNA, using EC3 ≤ 2% to classify substances as strong sensitizers and EC3 > 2% to classify substances as other sensitizers, was similar to that for the guinea pig tests. The overall correct classification rate of human sensitizers and nonsensitizers was 61% (34/56) for the LLNA versus 59% (33/56) for the guinea pig tests. The LLNA correctly classified more strong sensitizers and other sensitizers than did guinea pig tests; however, the LLNA correctly classified fewer nonsensitizers. The LLNA correctly classified 71% (10/14) of the strong human sensitizers versus 57% for the guinea pig tests, 67% (14/21) of the other human sensitizers versus 52% (11/21) for the guinea pig tests, and 48% (10/21) of the nonsensitizers versus 67% (14/21) for the guinea pig tests.

Test Method Reliability

Basketter and Cadby (2004) evaluated the intralaboratory variability associated with 29 individual EC3 concentrations for isoeugenol, which ranged from 0.5% to 2.6%. These data were used to support the “often-mentioned perspective that the biological variation associated with the estimation of EC3 values means that any particular EC3 can be halved or doubled” (Basketter and Cadby 2004). Basketter et al. (2007a) evaluated the interlaboratory reproducibility of EC3 data for 17 sensitizers tested in at least two laboratories using the same vehicle. The authors concluded that, although variability exists, it is less than an order of magnitude.

A number of analyses included in this BRD highlight the potential impact of the LLNA vehicle on EC3 values and potency classification. Forty-five substances in the NICEATM LLNA database had data from tests in multiple vehicles. Evaluation revealed that the vehicle-specific values for only 9% (4/45) of the substances varied by more than an order of magnitude. The GHS potency classifications differed by vehicle for 18% (8/45) of these substances. Another 24% (11/45) of the substances were classified differently as either sensitizers or nonsensitizers by the LLNA. Also, for 16% (7/45) of the substances, LLNA results from the same vehicle resulted in discordant sensitizer or nonsensitizer outcomes.

In a separate analysis, a correlation was calculated for EC3 values from two vehicles (dimethylformamide [DMF] and acetone) when compared to the EC3 values for the same substance obtained with acetone: olive oil (AOO; 4:1 by volume) as the vehicle. These data indicate that EC3 values for substances tested in acetone and AOO are similar, while EC3 values for substances tested in DMF are consistently lower than those obtained with AOO (i.e., the sensitizers are more potent in DMF than in AOO).

While vehicle may be an important determinant of the calculated EC3 value, it had no impact on the relationship of the LLNA EC3 with DSA_{05} values for the 63 substances that were sensitizers in the LLNA and in the HMT and/or HRIPT (see **Annex IV**).

Animal Welfare Considerations

The proposal for using the LLNA for potency determinations does not impact its requirement for using animals or the number of animals that are required. However, this application could broaden the use of the LLNA protocol in place of guinea pig tests and thereby further reduce the number of guinea pigs that are being used to assess skin sensitization potential. The LLNA test method protocol requires a minimum of only four mice per treatment group, whereas currently recommended guinea pig tests require at least 10 guinea pigs per group for the Buehler test and at least 5 guinea pigs per group for the guinea pig maximization test. The LLNA is also a refinement compared with guinea pig tests because it avoids the pain and distress that occurs in guinea pig tests when substances cause ACD.

Test Method Transferability

No changes to the LLNA protocol are being proposed. Therefore, the transferability, training requirements, and time and cost considerations for the LLNA remain unchanged from the previous ICCVAM evaluations (ICCVAM 1999, 2010c).

1.0 Introduction and Rationale for the Proposed Use of the Murine Local Lymph Node Assay for Potency Assessment

1.1 Introduction

1.1.1 Allergic Contact Dermatitis

Allergic contact dermatitis (ACD) is an adverse health effect that frequently develops in workers and consumers exposed to skin-sensitizing chemicals and products. ACD results in lost workdays[10] and can significantly diminish quality of life (Hutchings et al. 2001; Skoet et al. 2003). ACD develops in two phases: induction and elicitation. The induction phase occurs when a susceptible individual is exposed topically to a skin-sensitizing substance. During induction, the substance passes through the epidermis, where it forms a hapten complex with dermal proteins. The Langerhans cells, the resident antigen-presenting cells in the skin, process the hapten complex. The processed hapten complex then migrates to the draining lymph nodes. Antigen presentation to T-lymphocytes follows, which leads to the clonal expansion of these cells. At this point, the individual is sensitized to the substance (Basketter et al. 2003; Jowsey et al. 2006). Studies have shown that the magnitude of lymphocyte proliferation correlates with the extent to which skin sensitization develops (Kimber and Dearman 1991; Kimber and Dearman 1996).

The elicitation phase occurs when the individual is topically exposed to the same substance again. As in the induction phase, the substance penetrates the epidermis, is processed by the Langerhans cells, and is presented to circulating T-lymphocytes. The antigen-specific T-lymphocytes are then activated, which causes release of cytokines and other inflammatory mediators. This release produces a rapid dermal immune response that can lead to ACD (Basketter et al. 2003; ICCVAM 1999; Jowsey et al. 2006; Sailstad et al. 2001).

1.1.2 Historical Background

In 1999, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) recommended the murine local lymph node assay (LLNA) as a valid substitute for currently accepted guinea pig test methods (i.e., the guinea pig maximization test [GPMT] and the Buehler test [BT]) to assess the ACD hazard potential of many, but not all, types of substances (Dean et al. 2001; Haneke et al. 2001; ICCVAM 1999; Sailstad et al. 2001). The recommendation was based on a comprehensive evaluation that included an assessment of the validation status of the LLNA by an independent scientific peer review panel (Panel). The Panel and ICCVAM recommendations (ICCVAM 1999) are available at the website of the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)-ICCVAM.[11]

ICCVAM recommended to U.S. Federal agencies that the LLNA should be considered for regulatory acceptance or other nonregulatory applications for assessing the ACD hazard potential of substances, while recognizing that some testing situations would still require the use of traditional guinea pig test methods (ICCVAM 1999; Sailstad et al. 2001). The LLNA was subsequently incorporated into national and international test guidelines for the assessment of skin sensitization:

- U.S. Environmental Protection Agency Health Effects Test Guidelines on Skin Sensitization (EPA 2003)
- Organisation for Economic Co-operation and Development Test Guideline 429 (OECD 2002)

[10] http://www.bls.gov/IIF
[11] http://iccvam.niehs.nih.gov/

- International Organization for Standardization 10993-10: Tests for Irritation and Delayed-type Hypersensitivity (ISO 2002).[12]

On January 10, 2007, the U.S. Consumer Product Safety Commission (CPSC) formally requested that ICCVAM and NICEATM evaluate several activities related to the LLNA. One of the nominated activities was an assessment of the validation status of the LLNA as a stand-alone assay for determining the potency (including severity) of skin sensitizers for regulatory classification purposes. In response to this nomination, ICCVAM and NICEATM compiled a comprehensive draft background review document (BRD). ICCVAM and its interagency Immunotoxicity Working Group (IWG) evaluated the validation status of the LLNA as a stand-alone assay to determine the potency of skin sensitizers for regulatory classification purposes, and ICCVAM developed draft test method recommendations based on this initial evaluation.

An independent scientific peer review panel (Panel) reviewed the draft BRD in March 2008 to evaluate the extent to which the information in the draft BRD supported the draft test method recommendations. The Panel concluded that the LLNA should not be used as a stand-alone assay for categorizing skin sensitizers based on potency but that it could be used as part of a weight-of-evidence evaluation for this purpose. The Panel recommended that NICEATM perform additional analyses using alternative human reference values that might be more appropriate for evaluating the use of the LLNA for determining skin sensitization potency categories.

ICCVAM and the interagency IWG considered the conclusions and recommendations of the Panel, as well as comments from the public and the Scientific Advisory Committee on Alternative Toxicological Methods, in developing this final BRD. ICCVAM will provide the final BRD for consideration by U.S. regulatory agencies as part of the ICCVAM test method evaluation report.

1.1.3 Classification of Skin Sensitizers Based on Potency

Allergens are known to vary significantly in the potency with which they can induce skin sensitization. It has been suggested that skin-sensitizing chemicals vary as much as 10,000-fold in relative sensitization potency (Kimber et al. 2003). For the purposes of this BRD, *potency* is defined as a function of the concentration of a substance that is required for either induction or elicitation of skin sensitization. For induction, potency refers to the concentration of a substance needed to induce a skin sensitization response. The more potent the substance the smaller the quantity needed for induction. Likewise, for elicitation, potency refers to the concentration of a substance needed to elicit a response in a previously sensitized individual. The more potent the substance the smaller the quantity needed for elicitation (ECETOC 2003, 2008).

Interestingly, it has been widely reported that a defined test substance concentration does not necessarily result in a similar level of sensitization, or frequency of sensitization, every time. Actually, the key factor in the induction of skin sensitization is the dose of the substance per unit area of skin (Friedmann 1990; White et al. 1986). Kimber et al. (2008) discuss the evidence that in most typical situations it is dose per unit area that determines the effectiveness and extent of skin sensitization. Thus, it is recommended that sensitization thresholds obtained from animal and human data be expressed as the dose per unit area of skin (Boukhman and Maibach 2001).

The observed dose-response relationships associated with induction and elicitation allow thresholds for each phase to be determined (ECETOC 2003, 2008; Kimber et al. 2003). This includes thresholds for the applied dose of a substance below which (1) skin sensitization will not be induced in a naïve individual or (2) an elicitation reaction will not occur in a previously sensitized subject (Kimber et al. 1999). Although these thresholds are largely determined by the potency of a particular allergen, they

[12] The OECD and ISO test guidelines were updated in 2010 (ISO 2010; OECD 2010a).

vary due to vehicle effects and the extent of dermal exposure (Lea et al. 1999; Marzulli and Maibach 1976). Additionally, it has been suggested that

- Induction thresholds for particular substances differ from the elicitation threshold for the same substance (i.e., in general, higher levels are needed for induction in a naïve individual than for elicitation in a previously sensitized individual) (Griem et al. 2003).
- Interindividual variability in thresholds for elicitation exists and is attributed largely to the extent to which individuals have been previously exposed (Basketter et al. 2003; ECETOC 2003, 2008; Kimber et al. 1999).

Most authorities do not currently regulate products based on skin sensitization potency, instead using simple "yes" or "no" designations of skin sensitization hazard. The CPSC, under the Federal Hazardous Substances Act,[13] currently requires hazard labeling of only those products that are considered to be strong skin sensitizers based on a weight-of-evidence approach that considers frequency of responses in exposed human populations, severity of responses, and the doses at which allergic reactions occur. The following substances meet the CPSC's definition of strong sensitizers:

- 4-Phenylenediamine and products containing it
- Powdered orris root and products containing it
- Epoxy resins systems containing, in any concentration, ethylenediamine, diethylenetriamine, and diglycidyl ethers with molecular weight less than 200
- Formaldehyde and products containing 1% or more of formaldehyde
- Oil of bergamot and products containing 2% or more of oil of bergamot

In December 2008, the third revised edition of the Globally Harmonized System of Classification and Labelling of Chemicals (GHS) was adopted; it was published in July 2009 (UN 2009). This edition of the GHS introduced two new subcategories for skin sensitizers:

- Subcategory 1A— strong skin sensitizers, for substances that occur frequently in humans and/or have high potency in animals
- Subcategory 1B— "other" skin sensitizers, for substances that show low frequency of occurrence in humans and/or a low to moderate potency in animals (see **Table C-1**)

Skin sensitizers are classified as Category 1 when the relevant regulatory authority does not require subcategorization or when data are insufficient for subcategorization. Nonsensitizers are not classified.

Kimber et al. (2003) proposed a four-level classification scheme for skin sensitization potency based on a log scale of EC3 values (see **Table C-2**). The Task Force on Contact Sensitization of the European Centre for Ecotoxicology and Toxicology also proposed a four-level classification scheme for assessing skin sensitization potency (see **Table C-3**) (ECETOC 2003). However, the evaluation in this BRD focuses on the usefulness of the LLNA as a stand-alone assay for determining skin sensitization potency based on the GHS classification scheme (see **Table C-1**). These other classification schemes are provided for reference only.

[13] Federal Hazardous Substances Act. 15 U.S.C. 1261, 16 C.F.R. 1500.

Table C-1 GHS Classification Categories for Skin Sensitization

Category	Classification Criteria	LLNA EC3	Human Evidence	GPMT Response	BT Response
1 (Skin sensitizer)	Evidence that skin sensitization occurs in a substantial number of people OR Positive results from an appropriate animal test	NA	NA	NA	NA
1A (Strong skin sensitizer)	High frequency of occurrence in humans AND/OR High potency in animals May consider severity	≤2%	Positive response at ≤500 μg/cm^2 (HRIPT or HMT induction threshold[1])	EITHER: ≥30% responders at ≤0.1% intradermal induction dose OR ≥60% responders at >0.1% to ≤1% intradermal induction dose	EITHER: ≥15% responders at ≤0.2% topical induction dose OR ≥60% responders at >0.2% to ≤20% topical induction dose
1B (Other skin sensitizer)	Low to moderate frequency of occurrence in humans AND/OR Low to moderate potency in animals May consider severity	>2%	Positive response at >500 μg/cm^2 (HRIPT or HMT induction threshold[2])	EITHER: ≥30 to <60% responders at >0.1% to ≤1% intradermal induction dose OR ≥30% responders at >1% intradermal induction dose	EITHER: ≥15% to <60% responders at >0.2% to ≤20% topical induction dose OR ≥15% responders at >20% topical induction dose

Abbreviations: BT = Buehler test; EC3 = estimated concentration of a substance expected to produce a stimulation index of 3, which is the threshold value for a substance to be considered a sensitizer in the LLNA; GHS = Globally Harmonized System of Chemical Classification and Labelling (UN 2009); GPMT = guinea pig maximization test; HMT = human maximization test; HRIPT = human repeat-insult patch test; LLNA = murine local lymph node assay; NA = not applicable.

[1] Human evidence for strong skin sensitizers can also include (1) diagnostic patch test data where there is a relatively high and substantial incidence of reactions in a defined population in relation to relatively low exposure or (2) other epidemiological evidence where there is a relatively high and substantial incidence of allergic contact dermatitis in relation to relatively low exposure.

[2] Human evidence for other skin sensitizers can also include (1) diagnostic patch test data where there is a relatively low but substantial incidence of reactions in a defined population in relation to relatively high exposure or (2) other epidemiological evidence where there is a relatively low but substantial incidence of allergic contact dermatitis in relation to relatively high exposure.

Table C-2 Potency Categorization of Skin Sensitizers Based on LLNA EC3 Values[1]

Potency Category	EC3 Value (%)
Extreme	<0.1
Strong	≥0.1 to <1
Moderate	≥1 to <10
Weak	≥10 to ≤100

Abbreviations: EC3 = estimated concentration of a substance expected to produce a stimulation index of 3, which is the threshold value for a substance to be considered a sensitizer in the LLNA; LLNA = murine local lymph node assay.

[1] Proposed by Kimber et al. (2003).

Table C-3 Proposed Skin Sensitization Potency Categories Based on Guinea Pig Data[1]

Induction Concentration (%)	GPMT Incidence (%)		BT Incidence (%)	
	30 to <60	≥60	15 to <60	≥60
<0.1	Strong	Extreme	Strong	Extreme
≥0.1 to <1	Moderate	Strong	Moderate	Strong
≥1 to <10	Weak	Moderate	Weak	Moderate
≥10 to ≤100	Weak	Weak	Weak	Weak

Abbreviations: BT = Buehler test; GPMT = guinea pig maximization test.

[1] Proposed by the ECETOC Task Force on Contact Sensitization (ECETOC 2003).

1.1.4 Use of the LLNA as a Stand-Alone Method for Potency Determinations

Traditional regulatory test methods for skin sensitization (i.e., GPMT, BT, LLNA) have focused on "yes" or "no" determinations of sensitization hazard. In recent years, the LLNA has been proposed as an effective method for determining skin sensitization potency because of the dose-response information that is generated. Originally suggested by Kimber and Basketter (1997), this concept was based on their characterization of the large difference in LLNA threshold response between 2,4-dinitrochlorobenzene (DNCB) and hexyl cinnamic aldehyde (HCA). A number of studies have been conducted in an attempt to support the use of the LLNA for this purpose (see **Section 9.0** for the review articles on this topic).

However, the LLNA had yet to be adequately validated for classifying skin sensitizers based on potency. Consequently, a number of workshops on skin sensitization reviewed the use of the LLNA to assess skin sensitization potency:

- CPSC Sensitizer Scientific Panel — July 2005 (Matheson 2006)
- World Health Organization International Programme on Chemical Safety (IPCS) International Workshop on Skin Sensitization in Chemical Risk Assessment — October 2006 (WHO/IPCS 2007)
- OECD Expert Group on Sensitization — February 2007 and March 2008

In each case, the participants concluded that the LLNA should be used in a weight-of-evidence approach to determine skin sensitization potency categories. The independent scientific peer review panel convened by ICCVAM in March 2008 was the first public independent peer review of the use of the LLNA as a stand-alone assay to assess the skin sensitization potency of test substances.

1.2 Validation of the LLNA for Skin Sensitization Potential

The ICCVAM Authorization Act of 2000 (Sec. 4(c)) mandates that "[e]ach Federal Agency … shall ensure that any new or revised … test method … is determined to be valid for its proposed use prior to requiring, recommending, or encouraging [its use]" (Public Law 106-545. 42 U.S.C. 285*l*-3).

Validation is the process by which the reliability and relevance of a test method for a specific purpose are established. *Relevance* is the extent to which a test method will correctly predict or measure the biological effect of interest (ICCVAM 1997). *Reliability* is defined as the reproducibility of a test method within and among laboratories. Reliability should be assessed by testing a diverse set of substances that represent (1) the types of chemical and product classes expected to be tested and (2) the range of responses that needs to be identified. This validation process is intended to provide data and information to allow U.S. Federal agencies to develop guidance on the use of test methods in evaluating the potential of substances to cause skin sensitization.

The validation process begins with preparation of a BRD that provides a comprehensive review of a test method, including its mechanistic basis, proposed uses, data quality, and performance characteristics (i.e., relevance and reliability) (ICCVAM 1997). This BRD summarizes the available information on the use of the LLNA for potency categorization of chemicals causing ACD. It will also help to identify any additional studies that should be considered during future development and validation activities.

1.3 Selection of Citations for the BRD

The test method data summarized in this BRD are based on information obtained from the peer-reviewed scientific literature identified through online searches via PubMed and Scopus, through citations in publications, and in response to a *Federal Register* notice requesting LLNA, guinea pig, and/or human skin sensitization data and experience (72 FR 27815).[14] **Annex I** contains a document provided by Basketter et al. for consideration by ICCVAM and the European Centre for the Validation of Alternative Methods (ECVAM) during their evaluations of the LLNA for potency determinations. The NICEATM potency database includes 191 references relevant to this evaluation. Key words used in the online searches for this evaluation were ("LLNA" OR "Local Lymph Node" OR "Local lymph node" OR "local lymph node") AND "potency."

[14] http://iccvam.niehs.nih.gov/SuppDocs/FedDocs/FR/FR_E7_9544.pdf

2.0 LLNA Test Method Protocol Components

The original ICCVAM-recommended LLNA test method protocol describes the conduct of the assay in detail (Dean et al. 2001; ICCVAM 2001). A test-substance-induced increase in lymphocyte proliferation in the draining lymph nodes of the ear, the site of test substance application, is used in the LLNA to identify chemical sensitizers. The test substance is first applied to the dorsum of the mice ears on three consecutive days. After 48 hours, mice are injected with radiolabeled compound (^{3}H-methyl thymidine or ^{125}I-iodeoxyuridine), which is incorporated into the DNA of proliferating cells. Five hours later, mice are euthanized and the auricular lymph nodes are harvested and processed so that the incorporated radioactivity can be measured. The stimulation index (SI), which is the ratio of incorporated radioactivity (measured as disintegrations per minute [dpm]) in the auricular lymph nodes of treated versus vehicle control mice, is used to assess the sensitizing potential of the test substance. Based on collecting lymph nodes from individual animals within each group, the formula for calculating the SI is:

$$SI = \frac{mean\ dpm\ from\ lymph\ nodes\ in\ substance\text{-}treated\ group}{mean\ dpm\ from\ lymph\ nodes\ in\ vehicle\text{-}treated\ group}$$

A test substance with an SI of 3 or greater is classified as a skin-sensitizing agent. The estimated concentration of a substance expected to produce an SI of 3 (i.e., the EC3) is the metric for determining skin sensitization potency using the LLNA. The method for determining the EC3 is a simple linear interpolation of the points in the dose-response curve that lie immediately above and below an SI of 3, the classification threshold for sensitizers in the LLNA. This method was chosen from an evaluation of a variety of statistical approaches to derive EC3 values from LLNA dose-response data (Basketter et al. 1999b). When there are no data points that fall below an SI value of 3, a more complex log-linear extrapolation may be applied as described in Ryan et al. (2007) in which the two lowest test concentrations from the dose-response curve are used, provided the lowest SI value approaches the value of 3 and that a linear dose-response exists.

The LLNA procedure recommended by ICCVAM (Dean et al. 2001; ICCVAM 2001) differs from the protocol described in OECD Test Guideline (TG) 429 (OECD 2002) in that the ICCVAM protocol requires a concurrent positive control and the collection and analysis of individual animal data rather than pooled animal data. The ICCVAM-recommended protocol and OECD TG 429 were recently updated to allow the use of four animals per dose group, rather than the minimum of five that was required previously, when individual animal data are collected (ICCVAM 2009; OECD 2010a). Most recently, ICCVAM evaluated and recommended variations of the LLNA that do not employ radioactivity (ICCVAM 2010b, 2010a). These were adopted as OECD test guidelines (OECD 2010c, 2010b). However, these nonradioactive LLNA methods have not been evaluated for potency determination.

3.0 Substances Used for Validation of the LLNA for Potency Determinations

No new LLNA, guinea pig, or human skin sensitization studies were conducted for this evaluation. Rather, data from available studies were evaluated retrospectively. Data were obtained from 141 different sources, including published reports as well as unpublished data submitted to NICEATM in response to a *Federal Register* notice (72 FR 27815)[15] requesting LLNA, guinea pig, and human skin sensitization study data.

The information included in this BRD is derived from a database of over 600 substances, 196 of which have LLNA data with comparative guinea pig and/or human data. Among these 196 substances are 136 substances with comparative human data (76 sensitizers, 60 nonsensitizers), 116 substances with comparative guinea pig data (64 sensitizers, 52 nonsensitizers), and 56 substances with comparative human and guinea pig data (35 sensitizers, 21 nonsensitizers) (see **Annex II**). Two of the five substances that meet the CPSC's definition of strong sensitizers, 4-phenylenediamine and formaldehyde, are among the 136 substances with comparative human data and the 56 substances with comparative human and guinea pig data (see **Section 1.1.3**).

When available, chemical classes for each substance were retrieved from the National Library of Medicine Medical Subject Headings database.[16] If chemical classes were unavailable, they were assigned using a standard classification scheme based on the Medical Subject Headings classification system. A substance could be assigned to more than one chemical class; however, no substance was assigned to more than three classes. Chemical class information is presented only to provide an indication of the variety of structural elements that are present in the structures that were evaluated in this analysis. Classification of substances is not intended to indicate the impact of structure on biological activity with respect to sensitization potential.

Table C-4 shows the chemical classes represented by the 196 substances tested in the LLNA with human and/or guinea pig skin sensitization data. If *inorganic* is considered to be one class, the 196 substances represent 30 chemical classes. Fifty-five substances are classified in more than one chemical class. The classes with the highest number of substances are carboxylic acids (33 substances) and aldehydes (18 substances). In the entire NICEATM LLNA database of more than 600 substances, 22 chemical classes are represented by at least five substances, thereby providing a sufficiently large representation for further analyses. Twenty of those classes had at least 60% of the LLNA results identified as positive, and these 20 classes were identified as those most likely to be associated with skin sensitization. In comparison, 19 of these 20 classes were also represented in the database of 196 substances included in this evaluation (i.e., the LLNA potency database); only macromolecular substances were not included. Further, some of the chemical classes that have been previously identified as containing common skin allergens (aldehydes, ketones, quinones, acrylates) (Gerberick et al. 2004) were represented in this LLNA potency evaluation. **Annex III** provides the chemical class, information on the physicochemical properties (e.g., K_{ow} [estimated log octanol-water partition coefficient]), Chemical Abstracts Service Registry Number, and uses for each of the 196 substances. This information was obtained from the published reports, submitted data, or through online literature.

[15] http://iccvam.niehs.nih.gov/SuppDocs/FedDocs/FR/FR_E7_9544.pdf
[16] http://www.nlm.nih.gov/mesh/meshhome.html

Table C-4 Chemical Classes[1] Represented in the LLNA Potency Database

Chemical Class	# of Substances[2]
Inorganic chemicals	**11**
Aluminum compounds	1
Chromium compounds	1
Elements	1
Gold compounds	1
Manganese compounds	1
Mercury compounds	1
Metals	5
Sulfur compounds	1
Zinc compounds	1
Organic chemicals	**185**
Alcohols	15
Aldehydes	18
Amides	5
Amines	16
Anhydrides	2
Azo compounds	5
Carbohydrates	6
Carboxylic acids	33
Cyanates	1
Esters	5

Chemical Class	# of Substances[2]
Organic chemicals (continued)	**---**
Ethers	6
Formulations[3]	16
Heterocyclic compounds	15
Hydrocarbons, acyclic	5
Hydrocarbons, cyclic	12
Hydrocarbons, halogenated	1
Hydrocarbons, other	9
Ketones	3
Lactones	1
Lipids	15
Natural complex substances[3]	15
Nitriles	2
Nitro compounds	2
Onium compounds	1
Phenols	14
Polycyclic compounds	4
Quinones	1
Sulfur compounds	16
Ureas	2
Unknown[4]	**3**

Abbreviations: LLNA = murine local lymph node assay.

[1] Chemical classifications are based on the Medical Subject Headings classification for chemicals and drugs, developed by the National Library of Medicine (http://www.nlm.nih.gov/mesh/meshhome.html).

[2] The total number of substances assigned to each chemical class does not equal the total number of substances evaluated because some substances were assigned to more than one chemical class, and some substances were not assigned to a specific chemical class.

[3] Substances assigned to these classes were mixtures of two or more components. In some cases, another chemical class was also assigned based on the active ingredient (for formulations) or the principal component (for natural complex substances).

[4] The proprietary substances (fatty acid glutamate, fatty acid alcohol #1, and fatty acid alcohol #2) were not identified sufficiently for a chemical class to be assigned.

4.0 Comparative *In Vivo* Reference Data

4.1 Human Reference Data

The human reference data for this evaluation, which were obtained from 44 sources, were from human predictive skin sensitization tests (i.e., the human maximization test [HMT], the human repeat-insult patch test [HRIPT]) and, for nonsensitizers, other published reports. Protocols for the HMT include that of Kligman (1966) and Kligman and Epstein (1975). The HRIPT protocol is a modification of the Draize test (Marzulli and Maibach 1974; Politano and Api 2008). Both the HMT and HRIPT involve an induction phase of repeated applications of test substance to the skin for 24 (HRIPT) to 48 hours (HMT) with an occlusive dressing. This sensitization phase is followed by a rest period of 10 to 14 days and then a challenge phase that includes an additional application of test substance to the skin in an occlusive dressing for 24 (HRIPT) to 48 (HMT) hours. The application sites are then evaluated 24 (HRIPT), 48, and 72 hours after application; and the incidence and severity of a skin sensitization reaction are reported. The major differences between the HMT and HRIPT protocols are:

- The application of sodium lauryl sulfate (SLS) 24 hours before the induction test and one hour before the challenge test when testing nonirritating substances in the HMT
- The smaller number of subjects in the HMT (i.e., 25) versus approximately 100 for the HRIPT
- The length of patch application: five 48-hour patches for the HMT versus nine or ten 24-hour patches for the HRIPT

Human predictive test data were obtained from the published literature and from the Research Institute for Fragrance Materials (A. Api, personal communication). Skin sensitization potency in humans was identified as the threshold concentration inducing a sensitizing response in either the HMT or HRIPT. For the purposes of this evaluation, the threshold for induction of skin sensitization in humans was the induction dose per skin area in a HMT or HRIPT that produced a positive response in 5% of the tested population (i.e., the dose per skin area leading to a sensitization incidence of 5% [DSA_{05}]). The DSA_{05} value, which represents a defined low level of a positive response (i.e., 5%), was used as the human threshold response because it corresponds best (compared with the no observed effect levels [NOEL] or lowest observed effect level [LOEL]) to the LLNA EC3, which is also a threshold positive response. The DSA_{05} was used if it was reported in the literature, otherwise it was calculated from the LOEL and the incidence of a positive response in the study:

$$DSA_{05} = \frac{LOEL\ (\mu g/cm^2) \times 5\%}{incidence\ (\%)}$$

If the LOEL was not reported in $\mu g/cm^2$ skin area, it was calculated using the concentration of test substance applied, the weight or volume applied, and the size of the patch:

$$Dose\ applied = \frac{fractional\ concentration\ (\mu g\ substance/\mu g\ solution) \times weight\ applied\ (\mu g\ solution)}{patch\ size\ (cm^2)}$$

The volume applied is often reported in µL and must be multiplied by 1000 µg/µL to convert it to a weight. Substances in tests that resulted in only NOELs at the highest dose tested (i.e., no LOELs) were considered nonsensitizing.

The potency for each human skin sensitizer was determined, with DSA_{05} values as the metric for the positive responses, using the GHS criteria in **Table C-1**. Substances with $DSA_{05} \leq 500\ \mu g/cm^2$ were considered strong human skin sensitizers (GHS Subcategory 1A); and substances with $DSA_{05} > 500\ \mu g/cm^2$ were considered other human skin sensitizers (GHS Subcategory 1B).

Annex II-2 provides the available human data for each substance, which includes, where available, the induction dose, vehicle, the NOEL and/or LOEL and DSA_{05} values for human sensitizers. Because they fail to elicit a positive response, nonsensitizers have no DSA_{05} values.

It is important to discuss some of the limitations associated with the human data, much of which come from older studies. First, the HMT and the HRIPT have differences in sensitivity. For instance, the HMT tends to give lower LOEL values than the HRIPT (Griem et al. 2003). Further, even when using the same human predictive test, the protocols often differ between laboratories in the application frequency, amount applied, and skin site used (Griem et al. 2003). In addition, the intraspecies variability of human susceptibility to skin sensitization (Friedmann 1990) may further confound the results from human predictive tests.

4.2 Guinea Pig Data

The guinea pig data for this evaluation were used not as reference data for the LLNA but as comparative data for usefulness in determining human skin sensitization potency categories. The guinea pig data, from 26 different sources, were obtained from the published literature or submitted reports and were generated using the currently accepted guinea pig test methods for skin sensitization (i.e., the GPMT and the BT). National and international test guidelines are available for these test methods (EPA 2003; OECD 1992).

Both test methods involve induction and elicitation phases. The GPMT requires intradermal injections, with and without Freund's complete adjuvant, followed by topical induction on Days 5 through 8. Induction concentrations should be systemically well tolerated but high enough to produce mild to moderate skin irritation. The challenge concentration, which is applied to the skin on Days 20 through 22, must be the highest nonirritating concentration.

The BT requires topical application of an induction concentration high enough to produce mild irritation on Days 0, 6 through 8, and 13 through 15. The challenge concentration, applied on Days 27 and 28, is the highest nonirritating concentration.

In both the GPMT and the BT, the challenge sites are evaluated 24 and 48 hours after removal of the challenge dose. The incidence and severity of skin sensitization reactions are reported. For the purposes of this evaluation, the potency for each guinea pig skin sensitizer (GHS Subcategory 1A—strong or GHS Subcategory 1B—other) is based on the percentage of responding guinea pigs and the associated induction concentration in accordance with the GHS criteria in **Table C-1**. Substances that produce positive responses in less than 30% of the test group for the GPMT and 15% of the test group for the BT are considered to be nonsensitizers in the guinea pig tests. **Annex II-3** provides the guinea pig test data for each substance, including, where available, the induction dose (intradermal for GPMT and topical for BT), the percentage of animals exhibiting a positive response, and the corresponding data source.

4.3 Availability of Original Records for Human and Guinea Pig Data

NICEATM was unable to obtain the original records and/or reports for the human and guinea pig reference data used in this evaluation. All animal data supporting the validity of a test method should be obtained and reported from studies conducted in accordance with Good Laboratory Practice (GLP) guidelines, which are internationally recognized principles designed to produce high-quality laboratory records (EPA 2006b, 2006a; FDA 2009; OECD 1998). Human studies should conform to Good Clinical Practice (GCP) guidelines (ICH 1996). GLP and GCP guidelines provide an internationally standardized procedure for study conduct, reporting requirements, archiving study data and records, and information about the test protocol in order to ensure the integrity, reliability, and accountability of a study.

The extent to which the human or guinea pig studies complied with GLP or GCP guidelines, respectively, is based on the information provided in published and submitted reports. Information on GLP compliance was available for data from guinea pig studies submitted by E. Debruyne (Bayer CropScience SA) and P. Botham (European Crop Protection Association [ECPA]). None of the published references from which human or guinea pig data were obtained have GLP or GCP information specified.

5.0 LLNA Data and Results

5.1 Description of the LLNA Test Method Protocol Used to Generate Data

The majority of studies included in this evaluation were reportedly conducted according to the original ICCVAM protocol (Dean et al. 2001; ICCVAM 1999) or following OECD TG 429 (OECD 2002). Where OECD TG 429 was the reference protocol, specifics on the number of animals per dose group tested, whether or not lymph nodes were pooled within dose groups, and/or whether a concurrent positive control was used were generally not available. In addition, in order to increase the LLNA database, NICEATM determined that data from nonstandard LLNA protocols could be used in the analyses without affecting the LLNA outcomes (see **Annex IV**). The nonstandard protocol deviations included use of a different strain of mouse, use of both sexes of mice, different dose schedule for topical application of test substance, different duration between the last topical application and the injection of radioactive marker, and pretreatment with SLS prior to topical application of the test substance.

5.2 Availability of Copies of Original LLNA Data Used to Evaluate Accuracy and Reliability

Copies of original data for the LLNA studies considered during the earlier ICCVAM evaluation (ICCVAM 1999) were made available to NICEATM for that evaluation. For the current evaluation, individual animal data for some of the LLNA studies submitted to NICEATM earlier were included; however, the original data for the vast majority of the LLNA studies used in this evaluation are not available. Individual animal data were submitted by P. Botham (ECPA), D. Germolec (NTP), H. Ott (Federal Institute for Occupational Safety and Health, Germany), and P. Ungeheuer (European Federation for Cosmetic Ingredients).

5.3 Description of the Statistical Approach Used to Evaluate the Resulting Data

Section 2.0 describes the derivation of the SI and the EC3. The EC3 (typically expressed as %) is the metric used to evaluate the capacity of the LLNA to predict skin sensitization potency.

To evaluate the correlation between EC3 values and human DSA_{05} values (expressed in $\mu g/cm^2$), EC3 values (in %) were converted to $\mu g/cm^2$ by multiplying by a factor of 250 (based on an exposed area of 1 cm^2 and a dosing volume of 25 µL in the LLNA) (Griem et al. 2003). For all other comparisons between LLNA and human or guinea pig test results, the EC3 was expressed in its traditional units (%).

5.4 Summary of Results

NICEATM obtained LLNA data for this evaluation from 95 sources. The data are provided in **Annex II-1**. Where available, the SI values for each concentration tested, the calculated EC3 values, and the corresponding data sources are provided. The information provided with the submitted data was used, but no additional attempt was made to identify the source or purity of the test substance.

5.5 Use of Coded Chemicals

Coding of substances to avoid potential scoring bias did not occur for any of the LLNA test substances evaluated by ICCVAM in the original evaluation (ICCVAM 1999) or for any of the more recently obtained studies used in the current evaluation.

5.6 Lot-to-Lot Consistency of Test Substances

Ideally, a single lot of each substance is used during the validation of a test method. In situations where multiple lots of a substance must be used, lot-to-lot consistency must be evaluated to ensure the consistency of the substance evaluated over the course of the study. There was no available information in any of the reports included in this evaluation with which to assess lot-to-lot consistency.

5.7 Availability of Data for External Audit

The data for the LLNA test substances previously evaluated by ICCVAM (1999) were audited during that evaluation. Whether the other LLNA studies included in this evaluation are available for audit is unknown.

6.0 Test Method Accuracy

This section evaluates the capacity of the LLNA to accurately predict skin sensitization potency in humans, based on data generated by the HMT and HRIPT for sensitizers. Other published data for nonsensitizers are included. The comparative capacity of the LLNA and guinea pig tests to predict skin sensitization potency in humans is also examined for substances tested in mice (LLNA), guinea pigs, and humans.

6.1 Usefulness of the LLNA in Predicting Skin Sensitization Potency in Humans

Two approaches were used to evaluate the capacity of the LLNA to predict skin sensitization potency in humans. In the first approach, for each substance classified as a sensitizer in both the LLNA and in humans, the LLNA EC3 (expressed in $\mu g/cm^2$ skin surface and not as a percent) was correlated against the human threshold response, the DSA_{05} (expressed in $\mu g/cm^2$).

In the second approach, using the same LLNA/human sensitizers as the first approach, the human sensitizers were classified as strong (GHS Subcategory 1A) or other sensitizers (GHS Subcategory 1B) based on the GHS decision criteria (strong sensitizers had $DSA_{05} \leq 500$ $\mu g/cm^2$, and other sensitizers had $DSA_{05} > 500$ $\mu g/cm^2$; see **Table C-1**). Classification rate analyses were then performed to determine the correct, overclassification, and underclassification rates of EC3 cutoffs, including that specified in the GHS, for classifying substances in the human skin sensitization potency categories (i.e., strong and other than strong).

In a variant of the second approach, substances that were classified in the LLNA as false positives (i.e., sensitizers in the LLNA but nonsensitizers in humans), false negatives (i.e., nonsensitizers in the LLNA but sensitizers in human tests), or nonsensitizers in both the LLNA and human tests were included. Then the correct classification rate as well as the under- and overclassification rates were recalculated for each skin sensitization category (strong sensitizer, other than strong sensitizer, nonsensitizer).

Data from more than one LLNA test were available for many of the substances in the NICEATM potency database, and some substances had more than one DSA_{05} value. Before conducting correlation/regression analyses, single EC3 and DSA_{05} values were established for each substance. The correlation/regression analyses used two different approaches for combining EC3 values or human DSA_{05} values where multiple values existed for individual substances: (1) most potent EC3 (i.e., the lowest) versus most potent DSA_{05} (i.e., the lowest) and (2) geometric mean EC3 versus geometric mean DSA_{05}. The regression with the highest coefficient of determination, R^2, was used to determine which approach to use for combining multiple values in the classification rate analyses. The impact of variability in the EC3 on skin sensitization potency categorization is discussed in **Section 7.0**.

6.1.1 Regression Analyses for LLNA EC3 versus Human Threshold Concentrations

The current NICEATM potency database includes 136 substances with both LLNA and human data. Sixty-three of the 136 are classified as skin sensitizers in both the LLNA and in the HMT and/or the HRIPT (see **Annex II**). Although there were 65 substances with positive LLNA and HMT/HRIPT responses, nickel salts and streptomycin were not considered to be LLNA sensitizers because the most prevalent LLNA responses were negative (8/10 tests for nickel salts and 4/5 tests for streptomycin). The distribution of the 63 LLNA sensitizers by the number of studies conducted and the solvent used is provided in **Table C-5**.

Table C-5 Distribution of 63 LLNA/Human Sensitizers by the Number of LLNA Studies Conducted and the Solvent Used

Multiplicity of LLNA Studies for 63 Sensitizers					
1 Study	2 Studies	3 Studies	4 Studies	5 Studies	≥6 Studies
24 (38%)	12 (19%)	4 (6%)	5 (8%)	1 (2%)	17 (27%)
Number of Sensitizers Tested in Each Solvent[1]					
AOO	ACE	DMF	DMSO	EtOH-DEP	Other[2]
41 (65%)	9 (14%)	20 (32%)	11 (17%)	22 (35%)	49 (78%)

Abbreviations: ACE = acetone; AOO = acetone: olive oil (4:1 by volume); DMF = dimethylformamide; DMSO = dimethyl sulfoxide; EtOH-DEP = 1:3 or 3:1 ethanol: diethyl phthalate; LLNA = murine local lymph node assay.

[1] Numbers add up to more than 63 because 30 substances were tested in two or more solvents.

[2] Includes EtOH; DEP; methyl ethyl ketone; methyl ethyl ketone and olive oil; petrolatum; propylene glycol; Pluronic L92; hydroxypropyl cellulose in methanol; water; ACE and water; DMF and water; DMSO and water; unspecified solvents; and EtOH-DEP with additives such as tocopherol, Trolox C, butylated hydroxytoluene, and eugenol.

The analyses of the 63 LLNA/human sensitizers include both linear regressions and Spearman correlations (Steel and Torrie 1980) of the log-transformed LLNA EC3 values and human DSA_{05} values, both in units of $\mu g/cm^2$. **Annex IV** describes the analyses performed to evaluate a number of approaches to calculate the geometric mean EC3 in order to determine (1) the use of negative LLNA results for substances that also produced positive results (i.e., how to account for discordant negative results), (2) the use of vehicle-specific LLNA results for substances that had tests in multiple vehicles, and (3) the use of LLNA results from nonstandard protocols (**Section 5.1**). The preferred approach for calculating geometric mean EC3 values, presented in this section, ignores discordant negative results and vehicles (i.e., all EC3 values were pooled regardless of vehicle). It includes nonstandard protocols because these approaches had no impact on the EC3–DSA_{05} relationship (see **Annex IV**). Geometric mean DSA_{05} values were calculated using all available DSA_{05} values for each substance with multiple values.

Figure C-1 shows both the geometric mean and the most potent EC3–DSA_{05} regressions. Both regressions indicated a positive correlation between LLNA and human test results. The slopes for both regressions were significantly different from zero ($p < 0.001$). The geometric mean regression yielded $R^2 = 0.448$, and the most potent regression yielded $R^2 = 0.382$. The resulting regression equations are provided in **Table C-6** as regressions 1 (most potent) and 2 (geometric mean). Spearman correlations also indicated that the EC3–DSA_{05} relationship was statistically significant: $p < 0.0001$ for both correlations. The Spearman r (correlation coefficient) for the geometric mean regression was higher than that for the most potent regression ($r = 0.692$ versus 0.594). Because the geometric mean regressions produced a higher R^2 value than the most potent regression, and the Spearman r was also higher for the geometric mean regression, the geometric mean approach for calculating a single LLNA EC3 and DSA_{05} value for each substance was carried forward for the classification rate analyses in **Section 6.1.2** and for additional regressions that were performed using the NICEATM potency database (i.e., regressions 3 and 4).

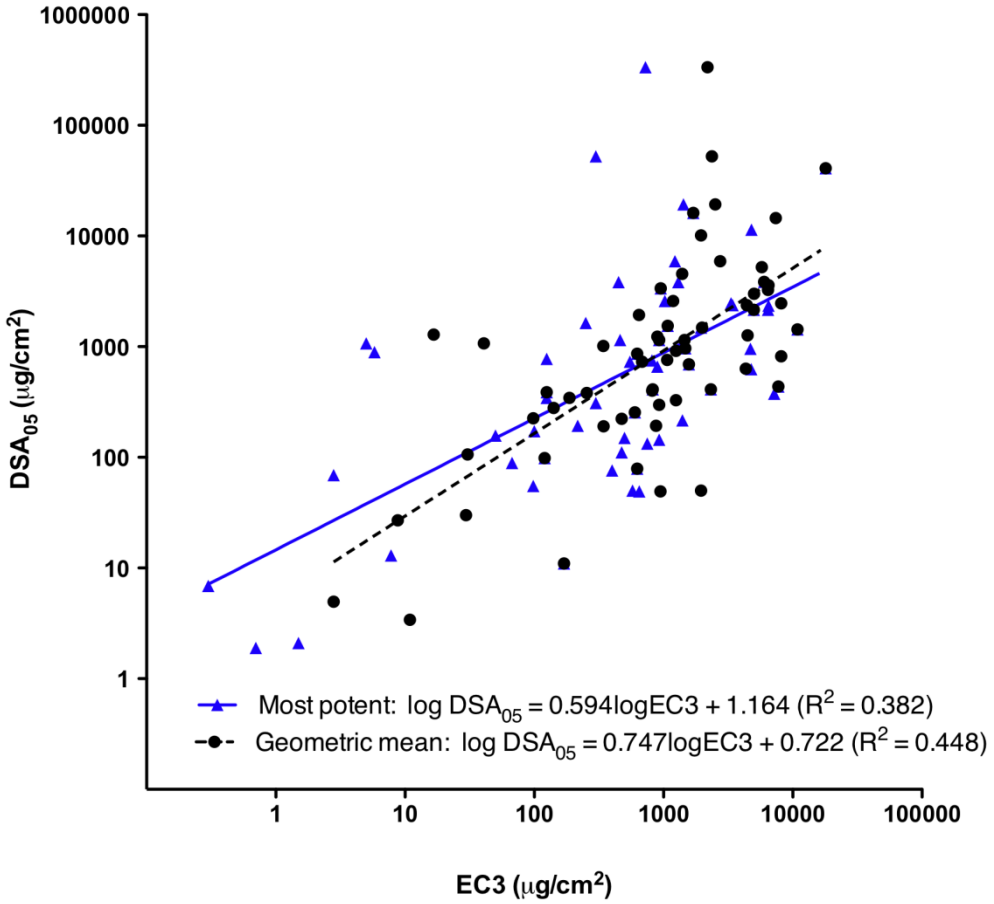
Table C-6 compares the correlation results obtained using the NICEATM potency database (see **Annex II**):

- When LLNA EC3 data were correlated against HMT threshold data, HMT NOEL data only, or HMT LOEL/10 data only
- When LLNA EC3 data were correlated against HRIPT threshold data, HRIPT NOEL data only, or HRIPT LOEL/10 data only

- For sensitizers tested in the LLNA using acetone: olive oil (AOO; 4:1 by volume), the most common solvent used, when correlated against human threshold data

For comparative purposes, **Table C-6** also provides linear regression data for LLNA EC3 values versus various sets of human threshold data published previously (Griem et al. 2003; Schneider and Akkan 2004) or submitted to NICEATM (Basketter et al. in **Annex I**). All of the sensitizers in these data sets are included in the NICEATM potency database (see **Annex II**).

Figure C-1 Most Potent and Geometric Mean Regressions for LLNA EC3 Values versus Human DSA_{05} Values for 63 LLNA/Human Skin Sensitizers



Abbreviations: DSA_{05} = induction dose per skin area, in $\mu g/cm^2$, in a human repeat-insult patch test or human maximization test that produces a positive response in 5% of the tested population; EC3 = estimated concentration of a substance expected to produce a stimulation index of 3, which is the threshold value for a substance to be considered a sensitizer in the LLNA; LLNA = murine local lymph node assay.

The solid line shows the regression line for the geometric mean LLNA EC3 versus the corresponding geometric mean human DSA_{05} (both in $\mu g/cm^2$) for 63 sensitizers. Geometric mean EC3 and DSA_{05} values were used for substances that had multiple values. The dashed line shows the regression line for the most potent LLNA EC3 versus the corresponding most potent human DSA_{05} for the same substances. The lowest EC3 and DSA_{05} values were used for substances that had multiple values. Human results were obtained from the human maximization test and/or human repeat-insult patch test.

Table C-6 Linear Regressions Obtained for LLNA EC3 Values versus Human Threshold Values

No.	Comparison	N	Regression Coefficient ($\mu g/cm^2$)[1]	Y-intercept[1]	R^2	p-value
1	NICEATM LLNA EC3 versus human DSA_{05} for sensitizers using most potent value	63	0.594 ± 0.097	1.164 ± 0.275	0.382	<0.0001
2	NICEATM LLNA EC3 versus human DSA_{05} for sensitizers using geometric mean values for multiply tested substances	63	0.747 ± 0.106	0.722 ± 0.322	0.448	<0.0001
3	NICEATM LLNA EC3 versus HMT DSA_{05}	36	0.579 ± 0.111	1.076 ± 0.344	0.441	<0.0001
4	NICEATM LLNA EC3 versus HRIPT DSA_{05}	42	0.832 ± 0.152	0.578 ± 0.455	0.427	<0.0001
5	Basketter et al. submission (see **Annex I**) reported EC3 data versus HMT/HRIPT NOEL, LOEL, and DSA_{05} values	66	0.896 ± 0.108	0.211 ± 0.335	0.519	<0.0001
6	Schneider and Akkan (2004) reported EC3 data versus HMT DSA_{05}	38	0.586 ± 0.115	0.936 ± 0.347	0.419	<0.0001
7	Schneider and Akkan (2004) reported EC3 data versus HRIPT DSA_{05}	24	0.765 ± 0.122	0.818 ± 0.355	0.641	<0.0001
8	Basketter et al. (2005) reported EC3 data versus HRIPT NOEL and LOEL data	25	1.121 ± 0.147	-0.533 ± 0.463	0.717	<0.0001
9	Griem et al. (2003) reported EC3 data versus HMT/HRIPT NOEL data	18	0.959 ± 0.129	0.111 ± 0.424	0.776	<0.0001
10	Griem et al. (2003) reported EC3 data versus HMT/HRIPT LOEL data	23	0.783 ± 0.123	0.682 ± 0.365	0.657	<0.0001
11	Griem et al. (2003) reported EC3 data versus HMT/HRIPT LOEL and NOEL data	41	0.854 ± 0.087	0.466 ± 0.271	0.711	<0.0001

Abbreviations: DSA_{05} = induction dose per skin area, in $\mu g/cm^2$, in a human repeat-insult patch test or human maximization test that produces a positive response in 5% of the tested population; EC3 = estimated concentration of a substance expected to produce a stimulation index of 3, which is the threshold value for a substance to be considered a sensitizer in the LLNA; HMT = human maximization test; HRIPT = human repeat-insult patch test; LLNA = murine local lymph node assay; LOEL = lowest observed effect level; N = number of studies included in analyses; No. = number of the analysis presented in the table; NOEL = no observed effect level; R^2 = coefficient of determination.

[1] Shown as parameter estimate with standard error.

Table C-6 shows that separating the HMT (regression 3; $R^2 = 0.441$) and HRIPT data (regression 4; $R^2 = 0.427$) did not increase the R^2 compared with the regression that combined HMT and HRIPT (regression 2) DSA_{05} values ($R^2 = 0.448$). For comparative purposes, linear regression data for LLNA EC3 values versus various sets of human threshold data published previously (Basketter et al. 2005; Griem et al. 2003; Schneider and Akkan 2004) and the Basketter et al. submission to NICEATM (see **Annex I)** are also provided in **Table C-6**. All of the sensitizers in these data sets are included in the NICEATM potency database (see **Annex II**).

As demonstrated in **Table C-6**, there are differences in R^2 values (which is a measure of the strength of the linear association between the LLNA EC3 and the DSA_{05}) among the various regressions. These differences presumably reflect differences in the number of substances with both LLNA EC3 and human skin sensitization threshold data; which human test is considered (HMT or HRIPT); whether NOEL, LOEL, and/or DSA_{05} values are used; and how data for substances tested multiple

times are collapsed into a single value. For example, the R^2 value generated with the NICEATM potency database (n = 63) increased from 0.382 to 0.448 when geometric mean threshold values were used for multiply tested sensitizers (regression 2) instead of the most potent value (regression 1). The R^2 values generated from data in the Basketter et al. submission (see **Annex I**), Basketter et al. (2005), Schneider and Akkan (2004), and Griem et al. (2003) are generally higher than the R^2 values from the NICEATM potency database. There may be several reasons for this apparent discordance including the following facts:

- The Basketter et al. submission to NICEATM (see **Annex I**) (regression 5) and the NICEATM analysis (regression 2) used data from both the HMT and the HRIPT for a similar number of chemicals (66 versus 63, respectively), but the Basketter et al. submission used NOELs, LOELs, and DSA_{05} values while the NICEATM analysis used only DSA_{05} values. The NICEATM analysis combined HMT and/or HRIPT data when multiple results were available, but the Basketter et al. submission had only one HMT or HRIPT result for each substance. The parameter estimates (i.e., the regression coefficients and the y-intercepts) for the two regressions are close enough that the standard errors overlap.
- The NICEATM analyses (regressions 1 and 2) represent a larger set of substances (n = 63) than the published datasets (n = 18 to 41).
- Schneider and Akkan (2004) and the NICEATM analysis used DSA_{05} values. Schneider and Akkan (2004) performed separate regressions for the HMT (regression 6) and the HRIPT (regression 7). The HMT analysis (regression 6) was similar in the number of substances, regression coefficient, slope, and R^2 value to the NICEATM HMT analysis (regression 3). The HRIPT analysis (regression 7) was less similar to the NICEATM HRIPT analysis (regression 4), but the standard errors for the parameter estimates did overlap. The NICEATM HRIPT regression (4) contained 75% (18/24) more chemicals than the Schneider and Akkan regression (7).
- Basketter et al. (2005) (regression 8) used only the highest NOELs available (preferred) and LOELs (if NOELs were unavailable and sensitization incidence was <8%) from HRIPT data, while the NICEATM regression (regression 2) used DSA_{05} values from both the HMT and the HRIPT. NICEATM combined multiple HMT/HRIPT results for single substances using a geometric mean.
- Griem et al. (2003) (regressions 9-11) and the NICEATM analysis (regression 2) each included threshold doses from both HMT and HRIPT data. However, NICEATM used DSA_{05} values, while Griem et al. (2003) used NOELs (regression 9), LOELs (regression 10), or NOELs and LOELs combined (regression 11). Griem et al. considered incidences of positive responses below 10% to be LOELs. To derive LOELs for other incidences below 50%, uncertainty factors were applied: 10 for incidences between 25% and 50% and three for incidences between 10% and 25%.

6.1.2 Correct, Underclassification, and Overclassification Rates for EC3 Value Predictions of Human Skin Sensitization Potency Categories

In this analysis, the extent that the LLNA EC3 value correctly distinguished between strong and other sensitizers in humans was evaluated using the criteria for human thresholds (see **Table C-1**) recently published in the third revised edition of the GHS (UN 2009). The GHS criteria for human skin sensitization classifies sensitizers as strong (Subcategory 1A) if the positive response in an HMT or HRIPT test occurs at $\leq$500 μg/cm^2. The GHS criteria categorizes a sensitizer as other (Subcategory 1B) if the positive response in an HMT or HRIPT occurs at >500 μg/cm^2. Substances that do not produce a positive response are not classified (i.e., nonsensitizers). Similarly, positive LLNA responses can be divided into Subcategory 1A or 1B with an EC3 $\leq$ 2% or >2%, respectively. Substances with negative LLNA responses are not classified (i.e., nonsensitizers).

Substances with multiple EC3 values were assigned a geometric mean EC3 value calculated from all of the available positive LLNA tests regardless of vehicle (see **Annex II-4**). Forty-seven of the 98 substances with positive LLNA results had multiple EC3 values; the number of values per substance ranged from 2 to 66. Individual EC3 values ranged from 0.0007% to 98.5%. If a majority of the LLNA tests for a substance were negative, however, it was not assigned an EC3 value. LLNA results for nickel salts and streptomycin were designated as negative because the most prevalent LLNA responses were negative (8/10 tests for nickel salts and 4/5 tests for streptomycin).

Substances with multiple DSA_{05} values were assigned a geometric mean DSA_{05} value calculated from all of the available DSA_{05} values (see **Annex II-4**). Thirty-two of the 76 substances with positive human results had multiple DSA_{05} values; the number of values per substance ranged from 2 to 8. Individual DSA_{05} values ranged from 1.9 $\mu g/cm^2$ to 335545 $\mu g/cm^2$. **Table C-7** shows the distribution of substances into the GHS potency categories using LLNA and human results.

Classification rates to determine the extent that the EC3 could predict strong and other human skin sensitizers were calculated from the results of receiver-operator characteristic calculations (Fawcett 2006), which provide sensitivity and 1-specificity results from the EC3 values divided into those associated with human $DSA_{05} \leq 500$ $\mu g/cm^2$ and those associated with $DSA_{05} > 500$ $\mu g/cm^2$. Two approaches were used to estimate the classification rates (correct, underclassification, and overclassification) for the EC3 classification of strong and other human skin sensitizers. In the first approach, the classification analysis considered only the 63 substances classified as sensitizers in both the LLNA and humans based on the HMT and/or HRIPT. In the second approach, the analysis took into consideration those substances that were LLNA false positives (35) and false negatives (13) against human skin sensitization data, as well as those classified as nonsensitizers in both the LLNA and in humans (25) (see **Table C-7**).

Table C-7 Distribution of 136 Substances for Classification Rate Analyses[1]

LLNA +/Human +		**LLNA +/Human -**	**LLNA -/Human +**	**LLNA -/Human -**
Strong[2]	**Other[3]**			
25 (14 EC3 ≤ 2%; 11 EC3 > 2%)	38 (3 EC3 ≤ 2%; 35 EC3 > 2%)	35 (4 EC3 ≤ 2%; 31 EC3 > 2%)	13 (2 strong; 11 other)[2,3]	25

Abbreviations: DSA_{05} = induction dose per skin area, in $\mu g/cm^2$, in a human repeat-insult patch test or human maximization test that produces a positive response in 5% of the tested population; EC3 = estimated concentration of a substance expected to produce a stimulation index of 3, which is the threshold value for a substance to be considered a sensitizer in the LLNA; LLNA = murine local lymph node assay.

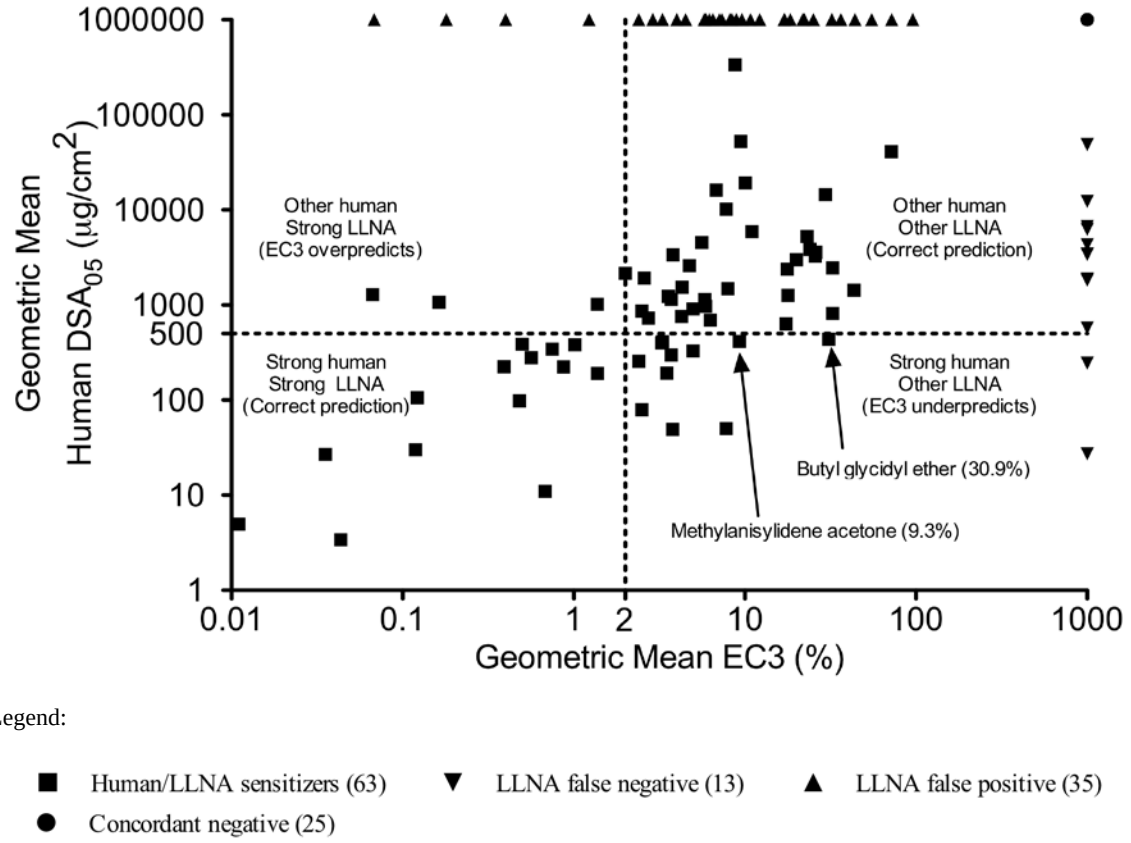
[1] Classification based on geometric mean EC3 (regardless of vehicle and discordant negative results) and DSA_{05} values.

[2] Human skin sensitizers were classified as strong sensitizers if the geometric mean DSA_{05} from HMT and/or HRIPT was ≤ 500 $\mu g/cm^2$.

[3] Human skin sensitizers were classified as other sensitizers if the geometric mean DSA_{05} from HMT and/or HRIPT was >500 $\mu g/cm^2$.

Figure C-2 shows geometric mean LLNA EC3 values plotted against the geometric mean DSA_{05} values for the 63 LLNA/human sensitizers. Also shown on the edges of the graph are concordant LLNA and human nonsensitizers (n = 25), LLNA false positives (n = 35), and LLNA false negatives (n = 13). The GHS cutoffs, 2% for EC3 and 500 $\mu g/cm^2$ for DSA_{05}, are marked to show the correspondence of the data with GHS Subcategories 1A and 1B.

Figure C-2 LLNA EC3 versus Human Results by GHS Potency Category for 136 Substances

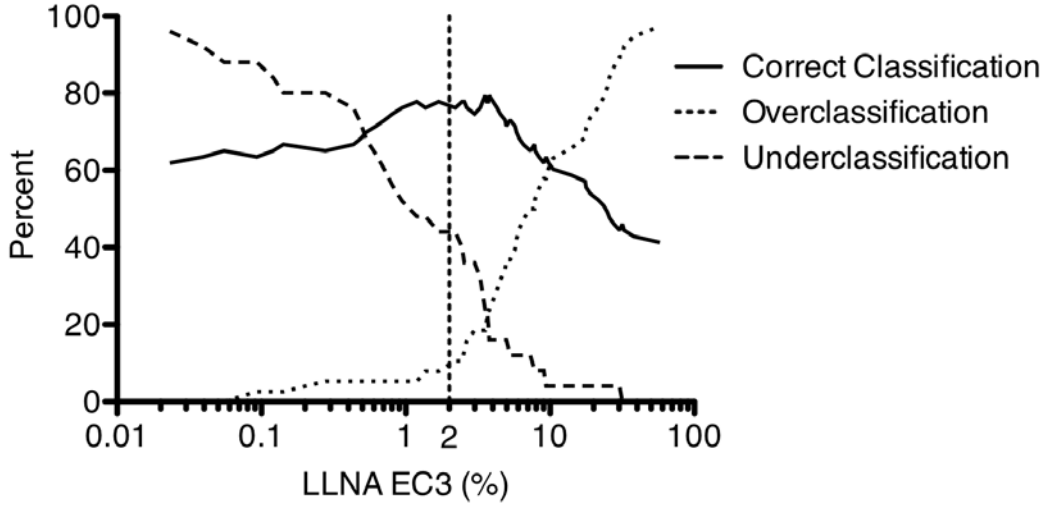


Abbreviations: DSA_{05} = induction dose per skin area, in $\mu g/cm^2$, in a human repeat-insult patch test or human maximization test that produces a positive response in 5% of the tested population; EC3 = estimated concentration of a substance expected to produce a stimulation index of 3, which is the threshold value for a substance to be considered a sensitizer in the LLNA; GHS = Globally Harmonized System of Classification and Labelling of Chemicals (UN 2009); LLNA = murine local lymph node assay.

The relationship between the LLNA EC3 value and the correct human skin sensitization potency classification as well as the under- (i.e., EC3-classified strong human skin sensitizers as other than strong skin sensitizers) and overclassification (i.e., EC3-classified other than strong human sensitizers as strong sensitizers) rates for the 63 substances that were sensitizers in both the LLNA and in human tests are shown in **Figure C-3**. From EC3 ≤ 0.8% to ≤4.5%, the correct classification rate changes little, ranging from 75% (47/63) to 79% (50/63) (see **Annex V**). However, the under- and overclassification rates change remarkably, ranging from 56% (14/25) to 16% (4/25) and from 5% (2/38) to 32% (12/38), respectively. For these data, the optimal EC3 value was ≤3.79%. This EC3 produced the highest correct classification rate, 79% (50/63), with an underclassification rate of 16% (4/25) and an overclassification rate of 24% (9/38). Although EC3 ≤ 3.54% also produced a correct classification rate of 79% (50/63), it yielded a higher underclassification rate (24% [6/25]) than EC3 ≤ 3.79%. The correct classification rate using EC3 ≤ 2% to classify substances as strong skin sensitizers, as prescribed by the GHS, yielded a correct classification rate of 78% (49/63) with an underclassification rate of 44% (11/25) and an overclassification rate of 8% (3/38).

At an EC3 ≤ 9.38%, one strong human skin sensitizer was underclassified as an other than strong sensitizer, in comparison to 11 when using EC3 ≤ 2% to classify substances as strong sensitizers. This suggests that substances with EC3 values in the range of 2% to 10% should be considered as having the potential to cause strong human responses unless there is evidence that indicates otherwise. Also, of note, no strong human sensitizers were underclassified as other sensitizers with an EC3 ≤ 31.00% (see **Figures B-2** and **B-3**).

Figure C-3 Classification Rates for the LLNA EC3 Prediction of Human Skin Sensitization Potency Categories for 63 Sensitizers



Abbreviations: EC3 = estimated concentration of a substance expected to produce a stimulation index of 3, which is the threshold value for a substance to be considered a sensitizer in the LLNA; LLNA = murine local lymph node assay.

Analysis based on 63 substances identified as sensitizers both in the LLNA and in humans using the human maximization test and/or the human repeat-insult patch test. In humans, substances were classified as strong sensitizers (n = 25) if the induction dose (in μg/cm^2 skin surface) in a human repeat-insult patch test or human maximization test that produced a positive response in 5% of the tested population was ≤500 μg/cm^2. Those that produced values >500 μg/cm^2 were classified as other than strong human skin sensitizers (n = 38).

The second approach for the classification analysis included data for the 35 human nonsensitizers that produced false positive results in the LLNA, the 13 human sensitizers that were false negative in the LLNA, and the 25 substances that were concordant nonsensitizers in the LLNA and in humans (see **Table C-7**). This increased the number of substances with comparative LLNA and human data from 63 to 136. Thus, the overall correct classification rate for this analysis includes the correct classification rate for human nonsensitizers, as well as strong and other than strong human sensitizers. Likewise, the overall underclassification rate includes the underclassification of all categories that can be underclassified (strong and other than strong human sensitizers); and the overall overclassification rate includes all categories that can be overclassified (other than strong human sensitizers and nonsensitizers).

The correct, underclassification, and overclassification rates of the LLNA versus human data were initially calculated using EC3 ≤ 2%. As indicated in **Tables C-8** and **C-9**, based on the NICEATM potency database, the LLNA correctly identified 52% (14/27) of the strong human skin sensitizers using EC3 ≤ 2%, but 48% (13/27) were underclassified by the LLNA. Among the 21 substances that produced an EC3 ≤ 2%, 67% (14/21) were strong human skin sensitizers. The remaining 33% (7/21) were either other than strong human sensitizers (n = 3) or substances not classified as human skin sensitizers (nonsensitizers; n = 4).

As shown in **Figure C-2**, most of the strong human sensitizers that were underclassified by the LLNA occurred between EC3 ≤ 2% and ≤10%. With this in mind, the classification rates for human sensitization categories obtained using incremental EC3 cutoff values up to 10% were also evaluated (see **Table C-8**). From EC3 ≤ 2% to ≤4%, the increase in the number of correctly classified strong sensitizers (14 to 21) is almost directly proportional to the decrease in the number of correctly classified other than strong sensitizers (35 to 29). The number of human nonsensitizers overclassified as strong sensitizers increases from four to seven when the LLNA EC3 cutoff moves from ≤2% to ≤4%. With each further increase of 2% in the LLNA EC3 cutoff, the number of correctly classified strong sensitizers increases by one substance. Using LLNA EC3 ≤ 10% to classify substances as strong sensitizers correctly classifies 89% (24/27) of the strong sensitizers compared with the 52% (14/27) of the strong sensitizers correctly classified using EC3 ≤ 2% (see **Table C-9**). However, the proportion of substances classified by the LLNA as strong sensitizers that actually are strong human sensitizers is higher for EC3 ≤ 2% than for EC3 ≤ 10%, 67% (14/21) versus 36% (24/67) (see **Table C-8**).

Figure C-4 shows the change in the overall correct classification (for strong human and other than strong sensitizers and nonsensitizers combined), underclassification (for strong human sensitizers and other sensitizers combined), and overclassification (for human other sensitizers and nonsensitizers combined) rates for the entire range of EC3 cutoff values. From EC3 ≤ 0.8% to ≤4.5%, the overall correct potency classification rate changes little, ranging from 53% (72/136) to 55% (75/136) (see **Annex V**). However, the under- and overclassification rates change noticeably, ranging from 36% (27/76) to 22% (17/76) and 34% (37/109) to 43% (47/109), respectively. For this dataset, the optimal EC3 value was ≤3.79%. This EC3 value produced the highest correct classification rate, which was 55% (75/136), with an underclassification rate of 22% (17/76) and an overclassification rate of 40% (44/109). Although EC3 ≤ 3.54% also produced a correct classification rate of 55% (75/136), it yielded a higher underclassification rate (25% [19/76]) than EC3 ≤ 3.79%.

Table C-8 Concordance of LLNA and Human Data for Strong Sensitizer, Other Sensitizer, and Nonsensitizer Categories at Selected LLNA EC3 Values

		Strong Sensitizer	Other Sensitizer	Nonsensitizer	Total
		EC3 ≤ 2% (GHS)	**EC3 > 2% (GHS)**	**Negative LLNA**	
Human Data[1]	**Strong Sensitizer**	14	11	2	27
	Other Sensitizer	3	35	11	49
	Nonsensitizer	4	31	25	60
	Total	21	77	38	136
		EC3 ≤ 4%	**EC3 > 4%**	**Negative LLNA**	
Human Data[1]	**Strong Sensitizer**	21	4	2	27
	Other Sensitizer	9	29	11	49
	Nonsensitizer	7	28	25	60
	Total	37	61	38	136
		EC3 ≤ 6%	**EC3 > 6%**	**Negative LLNA**	
Human Data[1]	**Strong Sensitizer**	22	3	2	27
	Other Sensitizer	16	22	11	49
	Nonsensitizer	12	23	25	60
	Total	50	48	38	136
		EC3 ≤ 8%	**EC3 > 8%**	**Negative LLNA**	
Human Data[1]	**Strong Sensitizer**	23	2	2	27
	Other Sensitizer	20	18	11	49
	Nonsensitizer	16	19	25	60
	Total	59	39	38	136
		EC3 ≤ 10%	**EC3 > 10%**	**Negative LLNA**	
Human Data[1]	**Strong Sensitizer**	24	1	2	27
	Other Sensitizer	23	15	11	49
	Nonsensitizer	20	15	25	60
	Total	67	31	38	136

Abbreviations: DSA_{05} = induction dose per skin area, in $\mu g/cm^2$, in a human repeat-insult patch test or human maximization test that produces a positive response in 5% of the tested population; EC3 = estimated concentration of a substance expected to produce a stimulation index of 3, which is the threshold value for a substance to be considered a sensitizer in the LLNA; GHS = Globally Harmonized System of Classification and Labelling of Chemicals (UN 2009); LLNA = murine local lymph node assay.

[1] Data for human sensitizers were DSA_{05} values (i.e., induction dose, in $\mu g/cm^2$ skin surface, in a human repeat-insult patch test or human maximization test that produced a positive response in 5% of the tested population). Sensitizers were classified as strong if $DSA_{05} \leq 500$ $\mu g/cm^2$ and other if $DSA_{05} > 500$ $\mu g/cm^2$ (UN 2009).

Table C-9 Correct, Underclassification, and Overclassification Rates for Prediction of Human Potency Categories by Selected LLNA EC3 Cutoff Values[1] for 136 Substances

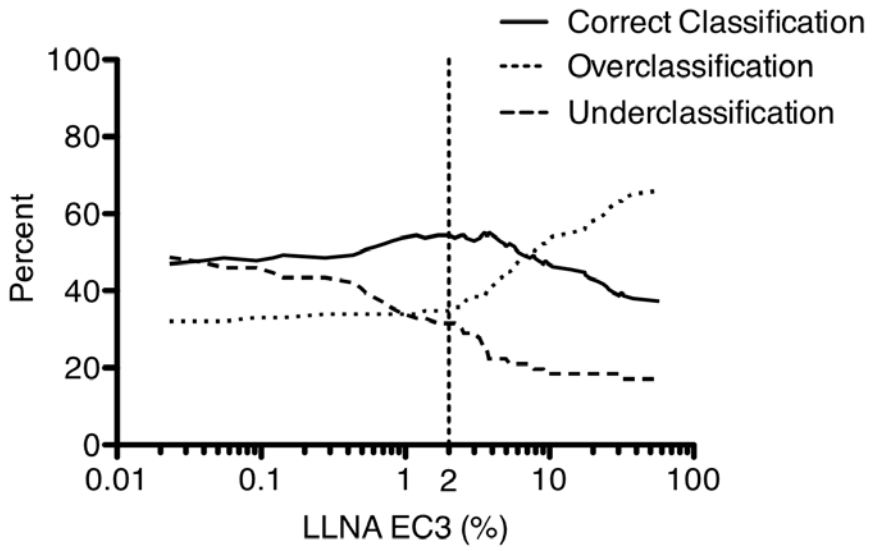
EC3 Cutoff for Strong versus Other Sensitizers	Strong Human Sensitizers ($DSA_{05} \leq 500$ µg/cm^2)		Other Human Sensitizers ($DSA_{05} > 500$ µg/cm^2)			Human Nonsensitizers		Overall Correct Potency Classification[2]
	Correct	Under	Over	Correct	Under	Correct	Over	
GHS Cutoff EC3 ≤ 2%	52 ± 19% (14/27)	48 ± 19% (13/27)	6 ± 7% (3/49)	71 ± 13% (35/49)	22 ± 12% (11/49)	42 ± 12% (25/60)	58 ± 12% (35/60)	54 ± 8% (74/136)
EC3 ≤ 4%	78 ± 16% (21/27)	22 ± 16% (6/27)	18 ± 11% (9/49)	59 ± 14% (29/49)	22 ± 12% (11/49)	42 ± 12% (25/60)	58 ± 12% (35/60)	54 ± 8% (74/136)
EC3 ≤ 6%	81 ± 15% (22/27)	19 ± 15% (5/27)	33 ± 13% (16/49)	45 ± 14% (22/49)	22 ± 12% (11/49)	42 ± 12% (25/60)	58 ± 12% (35/60)	50 ± 8% (68/136)
EC3 ≤ 8%	85 ± 13% (23/27)	15 ± 13% (4/27)	41 ± 14% (20/49)	37 ± 13% (18/49)	22 ± 12% (11/49)	42 ± 12% (25/60)	58 ± 12% (35/60)	48 ± 8% (65/136)
EC3 ≤ 10%	89 ± 12% (24/27)	11 ± 12% (3/27)	47 ± 14% (23/49)	31 ± 13% (15/49)	21 ± 12% (11/49)	42 ± 12% (25/60)	58 ± 12% (35/60)	47 ± 8% (64/136)

Abbreviations: DSA_{05} = induction dose per skin area, in µg/cm^2, in a human repeat-insult patch test or human maximization test that produces a positive response in 5% of the tested population; EC3 = estimated concentration of a substance expected to produce a stimulation index of 3, which is the threshold value for a substance to be considered a sensitizer in the LLNA; GHS = Globally Harmonized System of Classification and Labelling of Chemicals (UN 2009); LLNA = murine local lymph node assay.

[1] Classification rates are shown ±95% confidence limits.

[2] The overall correct classification rate includes the correct classifications of strong human sensitizers, other than strong sensitizers, and nonsensitizers.

Figure C-4 Overall Correct, Underclassification, and Overclassification Rates for LLNA EC3 Prediction of Human Potency Category for 136 Substances



Abbreviations: EC3 = estimated concentration of a substance expected to produce a stimulation index of 3, which is the threshold value for a substance to be considered a sensitizer in the LLNA; LLNA = murine local lymph node assay.

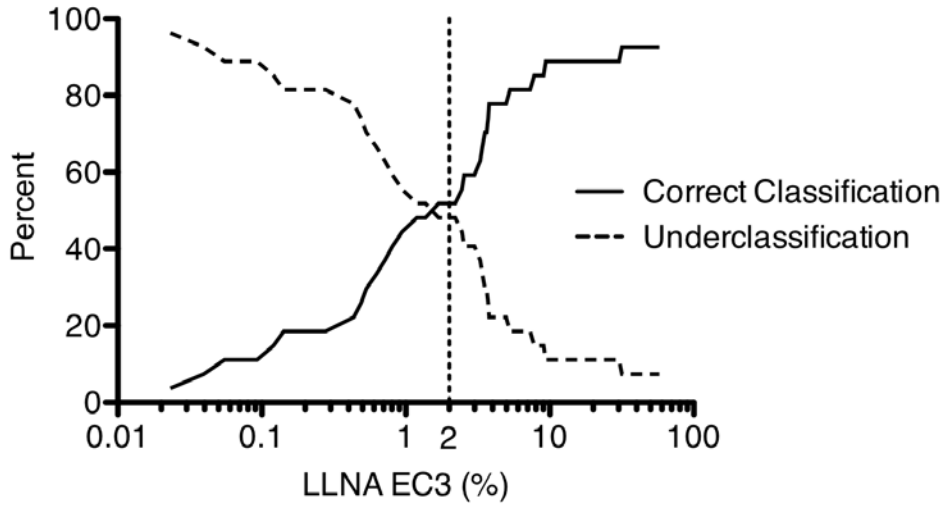
Analysis based on 136 substances: 63 sensitizers in both the LLNA and humans using the human maximization test and/or the human repeat-insult patch test (27 strong human sensitizers and 38 other than strong human sensitizers), 25 concordant nonsensitizers in the LLNA and in humans, 35 LLNA false positives, and 13 LLNA false negatives. In humans, sensitizers were classified as strong or other than strong sensitizers if the induction dose (in $\mu g/cm^2$ skin surface) in a human maximization test or a human repeat-insult patch test that produced a positive response in 5% of the tested population was $\leq 500\ \mu g/cm^2$. Sensitizers that produced a positive response of $>500\ \mu g/cm^2$ were classifed as other sensitizers (UN 2009).

The overall correct classification rate includes the correct classifications of strong human sensitizers, other than strong sensitizers, and nonsensitizers. The overall overclassification rate includes the overclassifications of other than strong human sensitizers and nonsensitizers, while the overall underclassification rate includes the underclassifications of strong human sensitizers and other than strong sensitizers.

Figure C-5 shows the change in the correct classification and underclassification rates for the 27 strong human sensitizers over the entire range of LLNA EC3 cutoff values. The correct potency classification rate for strong human sensitizers increases and the underclassification rate decreases as the EC3 increases. The correct classification rate plateaus, however, because the two strong human sensitizers that yielded negative results in the LLNA will not be correctly classified by any EC3 cutoff.

Fourteen percent (11/77) of the substances with EC3 > 2% are strong human sensitizers ($DSA_{05} \leq 500\ \mu g/cm^2$). In addition, 5% (2/38) of the substances that were negative in the LLNA were strong sensitizers.

Figure C-5 Correct and Underclassification Rates for LLNA EC3 Prediction of 27 Strong Human Sensitizers



Abbreviations: EC3 = estimated concentration of a substance expected to produce a stimulation index of 3, which is the threshold value for a substance to be considered a sensitizer in the LLNA; LLNA = murine local lymph node assay.

Analysis based on 27 substances identified as strong sensitizers in humans using the human maximization test and/or the human repeat-insult patch test based on the induction dose that produced a positive response in 5% of the tested population was ≤500 μg/cm^2.

6.1.3 Evaluation of Strong Sensitizers Underclassified by LLNA EC3 ≤ 2%

The thirteen strong human sensitizers that were underclassified by the LLNA EC3 ≤ 2% criterion are shown in **Table C-10** in the order of increasing EC3. The two substances that had predominantly negative LLNA results, nickel salts and streptomycin, would have been underclassified even if their positive results were used in the analysis. The two positive nickel results were for nickel sulfate in dimethyl sulfoxide (DMSO), EC3 = 4.8%, and nickel chloride in 30% ethanol, EC3 = 5.5% (see **Annex II-1**). The positive result for streptomycin yielded EC3 = 33% in dimethylformamide (DMF) (see **Annex II-1**). Ten of the 11 remaining discordant substances had EC3 values less than 10%. Butyl glycidyl ether was the only strong human sensitizer with LLNA EC3 > 10%. Using a criterion of LLNA EC3 ≤ 4% to classify substances as strong sensitizers would have correctly classified seven of the 13 discordant substances. Using LLNA EC3 ≤ 10% would have correctly classified 10 of the 13 discordant substances.

There are few commonalities among these 13 substances with regard to chemical class, physical form, molecular weight, peptide reactivity, and K_{ow} (see **Annex III** for physicochemical information):

- The 13 substances represent 10 chemical classes: aldehydes, amines, carbohydrates, carboxylic acids, ethers, heterocyclic compounds, inorganic chemicals, ketones, lipids, and organic sulfur compounds.
- Most (8/13) of the substances are liquids.
- The molecular weights of 12 of the 13 substances range from 98.15 (trans-2-hexenal) to 192.3 (delta-damascone). The exception is streptomycin (1457.39).
- Peptide reactivity information was available for only 6 of 13 substances. Five of the six substances (benzoisothiazolinone, benzylidene acetone, diethyl maleate, trans-2-hexenal, and methyl-2-nonynoate) had high peptide reactivity, and one substance (phenylacetaldehyde) had moderate peptide reactivity.
- K_{ow} ranged from -8.5 (streptomycin) to 4.16 (delta-damascone).

Table C-10 Strong Human Sensitizers Underclassified by LLNA EC3 ≤ 2%[1]

Substance	LLNA EC3 (%/μg/cm²)	Human Results		
		DSA_{05} (μg/cm²)	Concentration Tested	Response Rate
2-Hexylidene cyclopentanone	2.40/600	255 (HRIPT)	NA	9.8% (5/51)
Methyl-2-nonynoate	2.50/625	79 (HRIPT)	NA	7.5% (5/67)
Diethylmaleate	3.27/818	400 (HMT, HRIPT)	4% 4%	100% (24/24) 7.5% (14/187)
Diethylenetriamine	3.30/825	411 (HMT)	10%	84% (21/25)
delta-Damascone	3.51/877	193 (HRIPT)	NA	13% (7/54)
Benzylidene acetone	3.70/925	299 (HMT, HRIPT)	2% 3%	48% (12/25) 9.7% (6/62)
trans-2-Hexenal	3.78/945	49 (HRIPT)	NA	24% (6/25)
Phenylacetaldehyde	4.99/1250	329 (HRIPT, HMT)	NA 2% 2% 2% 2%	13% (7/53) 44% (11/25) 16% (4/25) 52% (13/25) 8% (2/25)
Benzoisothiazolinone	7.79/1950	50 (HRIPT)	0.0725%	9% (5/58)
Methylanisylidene acetone	9.30/2325	412 (HMT)	8%	67% (16/24)
Butyl glycidyl ether	30.9/7725	437 (HMT)	10%	79% (19/24)
Nickel salts	Negative (8/10 tests)	27 (HMT)	1% 1% 10%	26% (6/323) 17% (4/24) 48% (12/25)
Streptomycin	Negative (4/5 tests)	245 (HMT)	0.1% 0.1% 5% 10% 10% 25%	4% (10/24) 13% (3/23) 50% (12/24) 65% (15/23) 100% (24/24) 80% (20/25)

Abbreviations: DSA_{05} = induction dose per skin area, in μg/cm², in an HRIPT or HMT that produces a positive response in 5% of the tested population; EC3 = estimated concentration of a substance expected to produce a stimulation index of 3, which is the threshold value for a substance to be considered a sensitizer in the LLNA; HMT = human maximization test; HRIPT = human repeat-insult patch test; LLNA = murine local lymph node assay; NA = not available.

[1] Shown in order of increasing LLNA EC3.

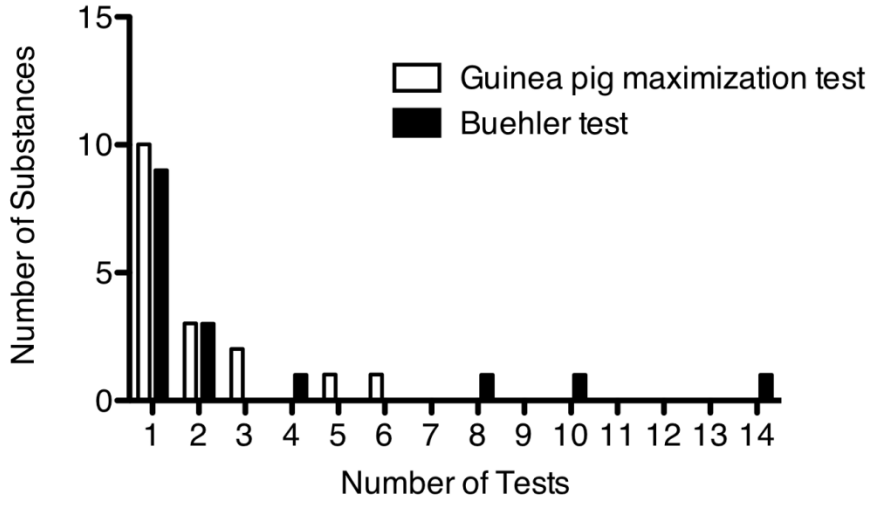
6.2 Comparison of LLNA versus Guinea Pig Predictions of Human Skin Sensitization Potency

Skin sensitization potency in guinea pigs from GPMT and BT results can also be used to classify substances as strong or other skin sensitizers (see **Section 1.1.3**). Thus, it was deemed useful to also evaluate the capacity of the guinea pig outcomes to agree with the human skin sensitization potency classification (see **Table C-1**) and to compare the guinea pig performance with the LLNA performance. Due to the categorical nature of the data collected in the guinea pig tests (i.e., using

incidence of sensitized animals with particular test substance concentrations), a regression analysis with human DSA_{05} values could not be conducted. However, substances tested in the GPMT or the BT could be assigned a potency classification (strong, other, or nonsensitizer) based on the GHS decision criteria in **Table C-1**. The capacity of the guinea pig outcomes to correctly classify human skin sensitization potency was evaluated and compared with the capacity of the LLNA to correctly classify human skin sensitization potency using a classification rate analysis.

The current NICEATM potency database of 196 substances includes 116 substances with guinea pig test data and 56 substances with both human and guinea pig data (see **Annex III**). Twenty-eight substances are classified as sensitizers in the LLNA, guinea pig tests, and human tests. **Figure C-6** shows the frequency distribution of the 28 sensitizers and the number of guinea pig studies conducted. The number of GPMT results per substance ranged from one to six, and the number of BT results per substance ranged from one to 14. Sixteen substances had GPMT results, and 17 substances had BT results. Ten of the 28 sensitizers had both GPMT and BT results.

Figure C-6 Distribution of 28 Substances Classified as Sensitizers in Guinea Pig Tests, LLNA, and Human Tests for the Number of Guinea Pig Studies Conducted[1]



Abbreviations: LLNA = murine local lymph node assay.

[1] Analysis based on 28 substances that tested as sensitizers in the guinea pig tests (i.e., guinea pig maximization test and Buehler test), LLNA, and in humans (human maximization test and human repeat-insult patch test).

Results from each guinea pig test were assigned to strong (GHS Subcategory 1A), other (GHS Subcategory 1B), or nonsensitizer (not classified) categories based on the decision criteria in **Table C-1**. The most prevalent outcome was used to categorize the guinea pig test results for substances with multiple tests. In this approach, test results from either GPMT or BT tests (i.e., as per the decision criteria in **Table C-1**) were considered together when assigning an overall classification category according to **Table C-1**. For example, there were six GPMTs and one BT for benzocaine. Three of the GPMT results were classified as other sensitizers (GHS Subcategory 1B), and three yielded nonsensitizer results (i.e., not classified). The BT result was classified as other skin sensitizer (GHS Subcategory 1B). Thus, based on the four GHS Subcategory 1B tests versus the three not classified tests, benzocaine was classified as GHS Subcategory 1B (i.e., other skin sensitizer). If a substance had an equal number of tests classified in two or more categories, the most potent result was used to represent the guinea pig potency classification for the substance.

Next, the correct classification rate as well as the under- and overclassification rates for guinea pig determinations of human potency category were calculated. Strong skin sensitizer (GHS Subcategory 1A), other skin sensitizer (GHS Subcategory 1B), and nonsensitizer (not classified) categories were calculated. The correct classification rate, underclassification rate, and overclassification rate for LLNA determinations of human potency category for the same substances were also calculated for comparison. For substances that had more than one LLNA EC3 or guinea pig response, the geometric mean EC3 value and the most prevalent guinea pig classification category were used. Two approaches were used to estimate the classification rates for the guinea pig determination of strong and other human sensitizers. In the first approach, the classification analysis considered only the 28 substances classified as sensitizers in guinea pigs, based on GPMT and/or BT results; in the LLNA; and in humans, based on the HMT and/or HRIPT.

In the second approach, the analysis included additional guinea pig and human sensitizers that were negative in the LLNA (n = 4), substances that were guinea pig false positives (n = 7) and false negatives (n = 3) against human skin sensitization data, and those classified as nonsensitizers in guinea pigs and humans (n = 14). The 56 substances used for the second approach also included

- Substances that were negative in guinea pigs but positive in the LLNA and humans (n = 1)
- LLNA false positives (n = 11) and false negatives (n = 6) against human skin sensitization data
- Substances classified as nonsensitizers both in the LLNA and in humans (n = 10)

The classification rate results are provided in **Table C-11**. In the first analysis, which focused only on the 28 substances classified as sensitizers in guinea pigs (i.e., GPMT and/or BT), the LLNA, and humans, overclassification means that other sensitizers (GHS Subcategory 1B) are misclassified as strong sensitizers (GHS Subcategory 1A), while underclassification means that strong sensitizers (GHS Subcategory 1A) are misclassified as other sensitizers (GHS Subcategory 1B). Using the guinea pig tests to determine human potency category, the overall correct classification rate (i.e., correctly classified strong human sensitizers plus correctly classified other human sensitizers) was 64% (18/28). The guinea pig tests correctly classified 67% (8/12) of the strong human sensitizers and 63% (10/16) of the other human sensitizers. Guinea pig results underclassified 33% (4/12) of the strong human sensitizers and overclassified 37% (6/16) of the other human sensitizers.

The correct classification rate of the LLNA, using EC3 $\leq$ 2% to classify substances as strong sensitizers and EC3 > 2% to classify substances as other sensitizers, was higher than that for the guinea pig tests (**Table C-11**). The overall correct classification rate of human sensitizers by the LLNA was 82% (23/28). The LLNA correctly classified 83% (10/12) of the strong human sensitizers and 81% (13/16) of the other human sensitizers. The LLNA underclassified 17% (2/12) of the strong human sensitizers and overclassified 19% (3/16) of the other human sensitizers.

The second analysis of 56 substances included guinea pig and LLNA false negatives, false positives, and concordant negatives relative to human data. The overall correct classification rate included correctly classified strong human sensitizers (GHS Subcategory 1A), other sensitizers (GHS Subcategory 1B), and nonsensitizers (not classified). Using the guinea pig tests to determine human potency categories, the overall correct classification rate was 59% (33/56). The guinea pig tests correctly classified 57% (8/14) of the strong human sensitizers, 52% (11/21) of the other human sensitizers, and 67% (14/21) of the human nonsensitizers. Guinea pig results underclassified 43% (6/14) of the strong human sensitizers and 14% of the other human sensitizers. Guinea pig results overclassified 33% (7/21) of the other human sensitizers and 33% (7/21) of the human nonsensitizers.

The overall correct classification rate produced by the LLNA, using EC3 $\leq$ 2% to classify substances as strong sensitizers and EC3 > 2% to classify substances as other sensitizers, was similar to that for the guinea pig tests (**Table C-11**). The LLNA's overall correct classification rate of human sensitizers and nonsensitizers was 61% (34/56) versus 59% (33/56) for the guinea pig tests. The LLNA correctly

classified more strong sensitizers and other sensitizers than the guinea pig tests did but correctly classified fewer nonsensitizers. The LLNA correctly classified 71% (10/14) of the strong human sensitizers versus 57% (8/14) for the guinea pig tests, 67% (14/21) of the other human sensitizers versus 52% (11/21) for the guinea pig tests, and 48% (10/21) of the nonsensitizers versus 67% (14/21) for the guinea pig tests. Guinea pig tests underclassified 43% (6/14) of the strong human sensitizers and overclassified 33% (7/21) of the other human sensitizers and 33% (7/21) of the human nonsensitizers. The LLNA underclassified 29% (4/14) of the strong human sensitizers and overclassified 14% (3/21) of the other human sensitizers and 52% (11/21) of the human nonsensitizers.

Table C-11 Comparative Correct Classification, Underclassification, and Overclassification Rates[1] When the GHS Criteria[2] for Guinea Pig Tests and the LLNA EC3 Are Used to Determine Human Skin Sensitization Potency Category

Comparison	Classification							
	Strong Sensitizer ($DSA_{05} \leq 500$ µg/cm^2)		Other Sensitizer ($DSA_{05} > 500$ µg/cm^2)			Nonsensitizer		Overall Correct Classification[3]
	Correct	Under	Over	Correct	Under	Correct	Over	
GPMT and/or BT determination of human potency for 28 guinea pig, LLNA, and human sensitizers	67 ± 27% (8/12)	33 ± 27% (4/12)	37 ± 24% (6/16)	63 ± 24% (10/16)	NA	NA	NA	64 ± 18% (18/28)
LLNA EC3 determination of human potency (EC3 ≤ 2% for strong, EC3 > 2% for other) for 28 sensitizers in guinea pigs, LLNA, and humans	83 ± 21% (10/12)	17 ± 21% (2/12)	19 ± 19% (3/16)	81 ± 19% (13/16)	NA	NA	NA	82 ± 14% (23/28)
GPMT and/or BT determination of human potency for 56 substances that include guinea pig and LLNA false positives, false negatives, and concordant negatives	57 ± 26% (8/14)	43 ± 26% (6/14)	33 ± 20% (7/21)	52 ± 21% (11/21)	14 ± 25% (3/21)	67 ± 20% (14/21)	33 ± 20% (7/21)	59 ± 13% (33/56)
LLNA EC3 determination of human potency (EC3 ≤ 2% for strong, EC3 > 2% for other) for 56 substances that include guinea pig and LLNA false positives, false negatives, and concordant negatives	71 ± 24% (10/14)	29 ± 24% (4/14)	14 ± 15% (3/21)	67 ± 20% (14/21)	19 ± 17% (4/21)	48 ± 21% (10/21)	52 ± 21% (11/21)	61 ± 13% (34/56)

Abbreviations: BT = Buehler test; DSA_{05} = induction dose per skin area, in µg/cm^2, in a human repeat-insult patch test or human maximization test that produces a positive response in 5% of the tested population; EC3 = estimated concentration of a substance expected to produce a stimulation index of 3, which is the threshold value for a substance to be considered a sensitizer in the LLNA; GPMT = guinea pig maximization test; LLNA = murine local lymph node assay; NA = not applicable because only substances that were sensitizers in humans, guinea pigs, and the LLNA were evaluated (i.e., other sensitizers can only be overclassified, and nonsensitizers were not evaluated).

[1] Classification rates are shown ±95% confidence limits.

[2] The criteria for distinguishing between strong and other sensitizers using the LLNA, GPMT, BT, and human tests are provided in **Table C-1**. For substances multiply tested in the GPMT and/or BT, the majority classification category was used. When an equal number of discordant classifications were recorded, the most severe classification category was used. Substances that were tested in the LLNA and the human repeat-insult patch test and/or human maximization test were represented by a geometric mean EC3 value and DSA_{05} values, respectively.

[3] The proportion of substances correctly assigned to each GHS category for skin sensitization potency (i.e., strong sensitizer, other sensitizer, and nonsensitizer, if applicable).

7.0 Test Method Reliability

An assessment of test method reliability (intralaboratory repeatability and intra- and interlaboratory reproducibility) is an essential element of any performance evaluation of an alternative test method (ICCVAM 2003). *Repeatability* refers to the closeness of agreement between test results obtained within a single laboratory when the procedure is performed on the same substance under identical conditions within a given time period (ICCVAM 1997, 2003). *Intralaboratory reproducibility* refers to the extent to which qualified personnel within the same laboratory can replicate results using a specific test protocol at different times. *Interlaboratory reproducibility* refers to the extent to which different laboratories can replicate results using the same protocol and test substances. It indicates the extent to which a test method can be transferred successfully among laboratories.

7.1 Variability of LLNA EC3 Values Using the Same Vehicle

As described in **Section 6.0**, the use of the LLNA for skin sensitization potency assessments depends on determining an accurate EC3 value for sensitizers. Thus, not only does the LLNA need to reproducibly achieve the correct sensitizer versus nonsensitizer result, but it also needs to reproducibly assign the proper skin sensitization potency classification. An evaluation of the intralaboratory variability associated with 29 individual EC3 values for isoeugenol, which ranged from 0.5% to 2.6%, was considered by Basketter and Cadby (2004) to support the "often-mentioned perspective that the biological variation associated with the estimation of EC3 values means that any particular EC3 can be halved or doubled."

The Basketter et al. submission evaluated EC3 data for 17 sensitizers tested in at least two laboratories as a measure of interlaboratory reproducibility of the EC3 value (see **Annex I**). The authors conclude that although there is biological variation in the EC3 values among substances tested multiple times using the same vehicle this variation is less than an order of magnitude.

Jowsey et al. (2008) assessed the inherent variability of the LLNA by examining the reproducibility of EC3 values for 14 substances tested more than once in AOO. These data were obtained from an LLNA dataset published by Basketter et al. (2007a). The variability was measured by dividing the maximum observed EC3 value by the minimum observed EC3 value. The results ranged from 1 to 4. Based on this outcome, a factor of five-fold was assumed a reasonable estimate of how variable EC3 values might be for a substance tested in the same vehicle multiple times.

Table C-12 provides available EC3 values for 45 substances tested in multiple vehicles. These data were selected from the current NICEATM LLNA database of over 600 substances by identifying the vehicles used in at least five LLNA studies. Thirteen percent (6/45) of the substances have discordant sensitizer/nonsensitizer LLNA results in the same vehicle.

7.2 Vehicle Effects on LLNA Results

Many factors affect skin sensitization. Two important factors are (1) the ability of the test substance to traverse the stratum corneum and reach the viable epidermis and (2) the efficiency of Langerhans cell migration from the skin. Both of these factors are susceptible to vehicle effects; therefore, the vehicle chosen for LLNA testing can have an impact on results (Basketter et al. 2001; Lea et al. 1999; McGarry 2007; Wright et al. 2001). Such effects need to be considered when evaluating the reproducibility of the LLNA in assigning the proper skin sensitization potency category.

7.2.1 Published Studies Regarding Vehicle Effects on LLNA Results

Jowsey et al. (2008) evaluated the impact of vehicle on the relative potency of skin-sensitizing substances tested in the LLNA. The authors compared EC3 values for 18 substances tested in at least two of 15 different vehicles using data from previously published results and unpublished Unilever

results (**Table C-13)**. The most common vehicles were AOO, DMF, and DMSO. The substances tested in AOO had EC3 values that ranged from 0.005% to 36.5% (nearing four orders of magnitude) with similar outcomes observed for DMF and DMSO. When evaluating EC3 values for each substance when it was tested in a different vehicle, the resulting variability for 50% (9/18) of the substances was no greater than the variability observed when substances were tested in the same vehicle (i.e., five-fold). Dinitrobenzene sulfonate, 1,4-dihydroquinone, and nickel sulfate were not included in this evaluation because their lowest or highest EC3 values were reported as greater than or less than a particular value. The EC3 values for 33% (6/18) of the substances differed by a factor of at least 10 when the substances were tested in different vehicles. In most cases (83% [5/6]), higher EC3 values (lower potency) were observed mostly with aqueous vehicles and propylene glycol. When these data were applied to the GHS classification scheme (see **Table C-1**) (UN 2009), seven substances (39% [7/18]) would have been assigned to different skin sensitization potency categories depending on the vehicle used (see **Table C-13**). 1,4-Dihydroquinone could not be evaluated because the highest EC3 was >1. The authors conclude that the LLNA vehicle can have an impact on potency. Although the effect is likely small, there are exceptions, and this knowledge is necessary for risk assessment.

McGarry (2007) performed an analysis similar to that in **Tables C-12** and **C-13** using the four-category LLNA potency system proposed by Kimber et al. (2003) (**Table C-2**) to demonstrate that the vehicle used affects the EC3 value and the resulting skin sensitization potency classification of a substance (see **Table C-14**). Among seven substances for which data from tests in multiple vehicles were available, six substances (86%) would have been assigned to different skin sensitization potency categories (see **Table C-2**) depending on the vehicle used. When these data were applied to the GHS classification scheme (see **Table C-1**) (UN 2009), two substances (29% [2/7]) would have been assigned to different skin sensitization potency categories depending on the vehicle used (see **Table C-14**). The EC3 values among the different vehicles differed by less than 10-fold for all seven substances evaluated.

Wright et al. (2001) also investigated the influence of application vehicle on sensitizing potency, using the LLNA to examine four recognized human contact allergens: isoeugenol, cinnamic aldehyde, and two fragrance chemicals. The fragrance chemicals are 3-dimethylaminopropylamine (a sensitizing impurity of cocamidopropyl betaine, a surfactant used in shower gel) and dibromodicyanobutane (the sensitizing component of Euxyl K 400, a preservative used in cosmetics). The four chemicals were applied in each of seven different vehicles (AOO, DMF, MEK, DMF, PG, and 50:50 and 90:10 mixtures of ethanol and water). It was found that the vehicle in which a chemical is presented to the epidermis had a marked effect on sensitizing activity. EC3 values ranged from 0.9% to 4.9% for isoeugenol, from 0.5% to 1.7% for cinnamic aldehyde, from 1.7% to >10% for 3-dimethylaminopropylamine, and from 0.4% to 6.4% for dibromodicyanobutane. These authors confirm that the vehicle in which a chemical is encountered on the skin has an important influence on the relative skin-sensitizing potency of chemicals and may have a significant impact on the elicitation of ACD.

7.2.2 NICEATM Evaluation of Vehicle Effects on LLNA Results

The data in **Table C-12** indicate that the assigned skin sensitization potency classification, strong versus other than strong sensitizer, differed by vehicle for 18% (8/45) of these substances when using the GHS classification scheme (see **Table C-1**) (UN 2009). Only 9% (4/45) of the substances had EC3 values that varied by at least an order of magnitude depending upon the vehicle used in the LLNA. Another 24% (11/45) of the substances were classified differently as either sensitizers or nonsensitizers depending on the vehicle. This is almost two times the number of substances that had discordant sensitizer or nonsensitizer results in the same vehicle (n = 6).

Table C-12 LLNA EC3 Values by Vehicle for 45 Substances with Positive Results (from the NICEATM LLNA Database)

Substance	LLNA Vehicle and Associated EC3 Values (%)											GP	Human
	AOO[1]	DMF[1]	MEK[1]	PG[1]	DMSO[1]	ACE	DEP	EtOH	EtOH/DEP (3:1)	EtOH/DEP (1:3)	L92		
2-Amino-6-chloro-4-nitrophenol	2.20				6.85							NA	NA
5-Amino-o-cresol	7.72				3.80							NA	NA
3-Aminophenol	**3.20**	**0.24**										**+**	**NA**
Amylcinnamic aldehyde	11.70[2]									7.60		+	-
Aniline	37.95[3]		25.79[2]									+	+
Atrazine						NC					35.96[2]	-	NA
Benzocaine	8.26[3]	3.37[3]				NC						+	+
Benzyl benzoate	17.00									NC		NA	-
Beryllium sulfate		0.68			NC							NA	+
Carvone	12.95[2]									7.81[3]		NA	+
(Chloro)methylisothia-zolinone	0.012[2]	0.008[2]	0.007	0.055[2]	0.008	0.005[2]						+	+
Cinnamic aldehyde	1.67[2]	0.56[2]	1.09	1.36	1.06[2]				0.42[2]			+	+
Citral	8.48[2]								4.94[2]	2.75[2]		+	+
Coumarin	NC	29.58[3]										NA	+
delta-Damascone	2.12[2]									9.60		NA	+
3,4-Dihydrocoumarin	5.60	3.25										+	+
1,4-Dihydroquinone	0.11[2]	0.21[2]	0.09[2]	NC								+	NA[4]
2,4-Dinitrobenzene sulfonic acid		**0.83**			**1.98**						**6.40**	**NA**	**NA**
2,4-DNCB	0.047[2]				0.015	0.012						+	+
Disperse blue 106		0.008			0.014[2]							NA	NA
Ethylenediamine	2.20	3.40				NC						+	+
Eugenol	11.31[2]					18.16	15.10[2]	10.70	5.30	7.53[2]		+	+
Formaldehyde	**0.82[2]**	**0.30[2]**		**2.80**	**0.30**	**0.60[2]**					**7.03[2]**	**+**	**+**
Geraniol	38.50[3]						11.80	5.60	25.8	15.25[2]		+	+
Glutaraldehyde	0.12[2]	0.02		1.50		0.08[2]						+	+
Glyoxal	1.40	0.60										NA	+
trans-2-Hexenal	5.50									2.60		NA	+

Substance	LLNA Vehicle and Associated EC3 Values (%)											GP	Human
	AOO[1]	DMF[1]	MEK[1]	PG[1]	DMSO[1]	ACE	DEP	EtOH	EtOH/DEP (3:1)	EtOH/DEP (1:3)	L92		
HCA	**9.53**[2]					**1.21**					**10.14**[2]	**+**	**-**
Hydroxycitronellal	24.23[2]	18.85					19.70	26.40	22.20	19.30		+	+
2-Hydroxyethyl acrylate	**2.96**[2]	**1.56**										**+**	**NA**
Isoeugenol	**1.31**[2]	**1.45**	**0.96**	**2.50**	**0.92**							**+**	**+**
Lilial	17.72[2]						4.10	2.90	8.80	13.90		NA	+
Limonene	69.00						44.19[2]	10.00	22.00	38.00		+	-
Methylhydrocinnamal	17.36[2]	23.10										NA	+
Methyl methacrylate	90.00					60.00						+	NA
Methyl salicylate	NC[2]	25.00	11.50			NC						-	-
Nickel salts		NC			4.80[3]						NC	+	+
Oxazolone	0.002[2]					0.001						+	NA
Penicillin G		5.46[2]			26.78[2]							+	+
Potassium dichromate		0.33			0.09[2]						0.20[2]	+	+
Resorcinol	5.92[3]	NC										NA[5]	-[6]
Salicylic acid	NC					12.22						-	-
SLS		4.32[2]			2.78[2]						4.90	-	-
Tetraethylthiuram disulfide	**1.40**					**5.42**						**NA**	**+**
Zinc diethyldithiocarbamate	0.24					1.01						NA	NA

Abbreviations: + = sensitizer; - = nonsensitizer; ACE = acetone; AOO = acetone: olive oil (4:1 by volume); DEP = diethyl phthalate; DMF = dimethylformamide; DMSO = dimethyl sulfoxide; DNCB = dinitrochlorobenzene; EC3 = estimated concentration of a substance expected to produce a stimulation index of 3, which is the threshold value for a substance to be considered a sensitizer in the LLNA; EtOH = ethanol; GHS = Globally Harmonized System of Classification and Labelling of Chemicals (UN 2009); HCA = hexyl cinnamic aldehyde; L92 = 1% Pluronic L92; LLNA = murine local lymph node assay; MEK = methyl ethyl ketone; NA = not available; NC = not calculated because the stimulation index < 3.0; NICEATM = National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods; PG = propylene glycol; SLS = sodium lauryl sulfate; TG = test guideline.

Boldface text highlights substances for which a discordant sensitizer subcategory (using the GHS EC3 cutoff of 2%; see **Table C-2**) would be assigned depending on the vehicle used in the LLNA.

[1] Vehicles recommended in OECD TG 429, listed in order of preference (OECD 2002). OECD TG 429 also indicates that other vehicles may be used with sufficient scientific rationale.

[2] Value represents a geometric mean of $n \geq 2$ EC3 values, or, for negative results, $n \geq 2$ negative tests.

[3] Value represents a geometric mean of $n \geq 2$ EC3 values. Additional tests in this vehicle were negative (i.e., stimulation index < 3).

[4] Although no human maximization test or human repeat-insult patch test data were reported, 1,4-dihydroquinone has been reported to be a sensitizer in humans (Basketter et al. 1999a).

[5] Although the specific guinea pig test method and exposure/incidence data were not reported, resorcinol has been reported to be a nonsensitizer in guinea pigs (Basketter et al. 1996).

[6] Resorcinol was negative in the human maximization test (Kligman 1966), but sensitization in humans has been reported (Basketter et al. 2007b).

Table C-13 LLNA EC3 Values for 18 Skin Sensitizers Tested in Different Vehicles (from Jowsey et al. 2008a)

Substance	LLNA Vehicle and Associated EC3 Values (%)															GP	Human
	AOO[1]	DMF[1]	MEK[1]	PG[1]	DMSO[1]	ACE	APS	L92	H_2O	EtOH/ H_2O (9:1)	EtOH/ H_2O (1:1)	EtOH	EtOH/ DEP (3:1)	EtOH/ DEP (1:3)	DEP		
Dinitro-benzene sulfonate		**<1**			**2**			**6.4**	**16**							**NA**	**NA**
1,4-Dihydro-quinone	0.15[2]	0.21[2]	0.09[2]	>1	0.35[2]	0.75[2]	>1									+	NA[3]
4-Phenylene-diamine	**0.11[2]**						**2.5**									**+**	**+**
3-Dimethyl-amino-propyl-amine	**2.2**	**1.7**	**1.8**	**>10**	**2.76**					**4.1**	**7.1**					**+**	**+**
Cinnamic aldehyde	**2.3[2]**	**0.48**	**1.09**	**1.36**	**0.93**					**1.56**	**1.2**					**+**	**+**
Dibromo-cyano-butane	**1.3[2]**	**6.4**	**0.4**							**1**						**+**	**+**
Ethylene glycol dimeth-acrylate	36.5	32.4	28.3	15.5	34.4	20										-	+
Eugenol	10.1[2]											10.7	5.3	10.5	15.1	+	+
Formal-dehyde	**0.76[2]**	**0.33**				**0.7**		**4.2**	**14.2**							**+**	**+**
Geraniol												5.6	25.8	20.4	11.8	+	+
Imidazol-idinyl urea		27.8[2]			27.1											+	+
Glutaral-dehyde	0.07			1.5		0.1										+	+
Hydroxy-citronellal	27.8[2]											26.4	22.2	19.3	19.7	+	+

Substance	LLNA Vehicle and Associated EC3 Values (%)															GP	Human
	AOO[1]	**DMF[1]**	**MEK[1]**	**PG[1]**	**DMSO[1]**	**ACE**	**APS**	**L92**	**H_2O**	**EtOH/ H_2O (9:1)**	**EtOH/ H_2O (1:1)**	**EtOH**	**EtOH/ DEP (3:1)**	**EtOH/ DEP (1:3)**	**DEP**		
Isoeugenol	**1.5[2]**	**1.4**	**1**	**2.5**	**0.9**		**2.9**			**1.8**	**4.9**					**+**	**+**
Lilial	18.7											3	8.8	13.9	4.2	NA	+
CMI/MI	0.005	0.0075	0.0068	0.048	0.0075	0.0076										+	+
Nickel sulfate		>5			4.8			2.5								+	+
Potassium dichromate		0.032			0.089[2]			0.17								+	+

Abbreviations: + = sensitizer; - = nonsensitizer; ACE = acetone; AOO = acetone: olive oil (4:1 by volume); APS = acetone: physiological saline (6:94 by volume); DEP = diethyl phthalate; DMF = dimethylformamide; DMSO = dimethyl sulfoxide; EC3 = estimated concentration of a substance expected to produce a stimulation index of 3, which is the threshold value for a substance to be considered a sensitizer in the LLNA; EtOH = ethanol; GP = guinea pig test result; H_2O = water; HU = human results; L92 = 1% Pluronic L92; LLNA = murine local lymph node assay; CMI/MI = 5-chloro-2-methyl-4-isothiazolin-3-one / 2-methyl-4-isothiazolin-3-one; MEK = methyl ethyl ketone; NA = not available; OECD = Organisation for Economic Co-operation and Development; PG = propylene glycol; TG = test guideline.

Boldface text highlights substances for which discordant sensitizer subcategory (using the GHS EC3 cutoff of 2%; see **Table C-2**) would be assigned depending on the vehicle used in the LLNA.

[1] Vehicles recommended in OECD TG 429, listed in order of preference (OECD 2002). OECD TG 429 also indicates that other vehicles may be used with sufficient scientific rationale.

[2] Value represents the arithmetic mean of $n \geq 2$ EC3 values.

[3] Although no human maximization test or human repeat-insult patch test data were reported, 1,4-dihydroquinone has been reported to be a sensitizer in humans (Basketter et al. 1999a).

Table C-14 LLNA EC3 Values for Seven Skin Sensitizers Tested in Different Vehicles (from McGarry 2007)

Substance	LLNA Vehicle and Associated EC3 Value (%)									GP	Human
	AOO[1]	DMF[1]	MEK[1]	PG[1]	DMSO[1]	ACE	L92	EtOH/H_2O (90:10)	EtOH/H_2O (50:50)		
Cinnamic aldehyde	1.70	0.50	1.10	1.40	0.90			1.60	1.20	+	+
1,4-Dihydroquinone	0.15	0.21	0.09		0.35	0.08				+	NA[2]
3-Dimethylpropylamine	**2.20**	**1.70**	**1.80**	**>10.00**	**3.20**			**4.10**	**7.10**	**NA**	**NA**
Isoeugenol	**1.00**	**1.40**	**1.00**	**2.50**	**0.90**			**1.80**	**4.90**	**+**	**+**
(Chloro)methylisothiazolinone/ Methylisothiazolinone	0.0049	0.0075	0.0068	0.0480	0.0075	0.0076				+	+
Nickel sulfate		>5.00			4.80		2.50			+	+
Potassium dichromate		0.0327			0.0500		0.1700			+	+

Abbreviations: + = sensitizer; ACE = acetone; AOO = acetone: olive oil (4:1 by volume); DMF = dimethylformamide; DMSO = dimethyl sulfoxide; EC3 = estimated concentration of a substance expected to produce a stimulation index of 3, which is the threshold value for a substance to be considered a sensitizer in the LLNA; EtOH/H_2O = ethanol/water; GHS = Globally Harmonized System of Classification and Labelling of Chemicals (UN 2009); L92 = 1% Pluronic L92; LLNA = murine local lymph node assay; MEK = methyl ethyl ketone; NA = not available; OECD = Organisation for Economic Co-operation and Development; PG = propylene glycol; TG = test guideline.

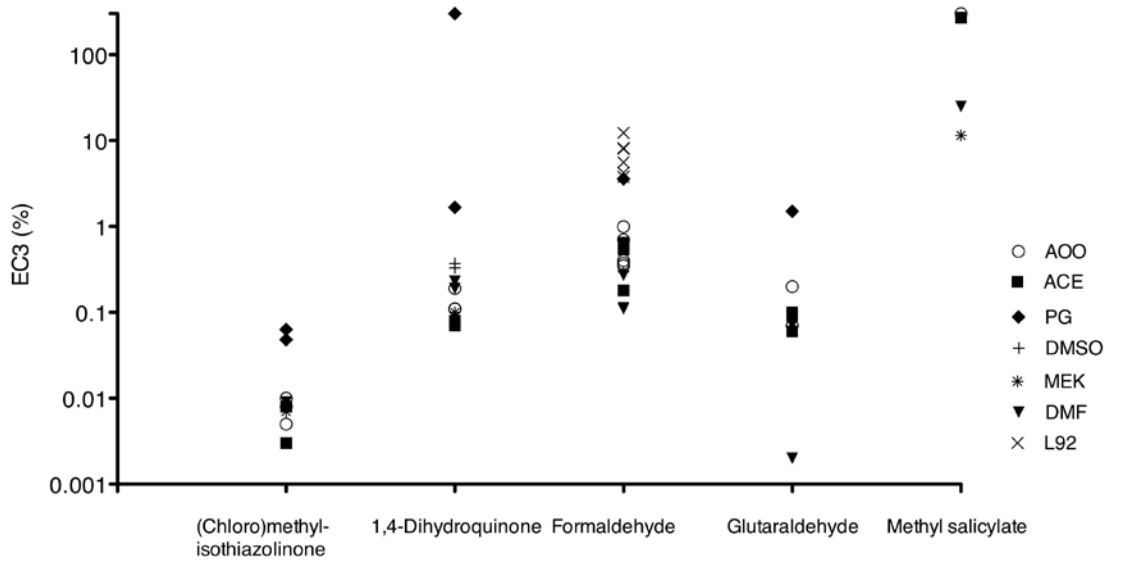
Boldface text highlights substances for which discordant classifications (using the GHS EC3 cutoff of 2%; see **Table C-2**) would be assigned depending on the vehicle used in the LLNA.

[1] Vehicles recommended in OECD TG 429, listed in order of preference (OECD 2002). OECD TG 429 also indicates that other vehicles may be used with sufficient scientific rationale.

[2] While no human maximization test or human repeat-insult patch test data were reported, 1,4-dihydroquinone has been reported to be a sensitizer in humans (Basketter et al. 1999a).

Figure C-7 further illustrates that the vehicle used has pronounced effects on the predicted skin sensitization potency when based on LLNA EC3 values. Five representative substances were selected from those listed in **Table C-12** based on available data from at least one LLNA test in multiple vehicles. These data demonstrate the potential impact of the vehicle on potency categorization when using the EC3 value. Greater than an order of magnitude difference can be seen for all five substances. This is in contrast to the conclusions of Jowsey et al. (2008) for multiple tests in different solvents (i.e., that EC3 values typically vary by no more than five-fold). Two substances in **Figure C-7** were either sensitizers or nonsensitizers in the LLNA, depending on the vehicle selected. One substance, 1,4-dihydroquinone, is sensitizing in guinea pigs and humans, although neither HMT nor HRIPT data were reported (Basketter et al. 1999a). The other substance (methyl salicylate) is nonsensitizing in guinea pigs and humans.

Figure C-7 Representative Substances and Respective LLNA EC3 Values When Tested in Different Vehicles (from the NICEATM LLNA Database)



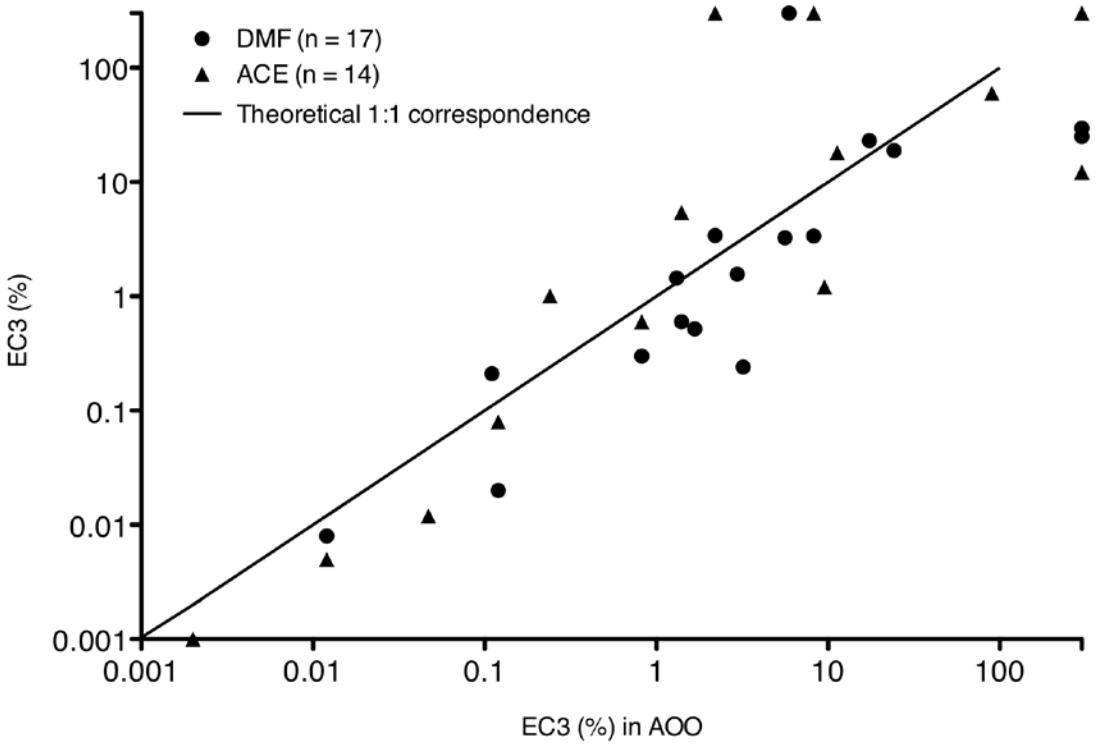
Abbreviations: ACE = acetone; AOO = acetone: olive oil (4:1 by volume); DMF = dimethylformamide; DMSO = dimethyl sulfoxide; EC3 = estimated concentration of a substance expected to produce a stimulation index of 3, which is the threshold value for a substance to be considered a sensitizer in the LLNA; L92 = 1% Pluronic L92; LLNA = murine local lymph node assay; MEK = methyl ethyl ketone; NICEATM = National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods; PG = propylene glycol.

Note: Values above 100 indicate studies where the substance was classified as a nonsensitizer.

As another analysis of vehicle effect on EC3 variability, a correlation was calculated for EC3 values from two vehicles (DMF and acetone) and compared to the EC3 values for the same substances obtained with AOO using the data in **Table C-12** (see **Figure C-8**). With AOO, a 1:1 correspondence would indicate that identical EC3 values had been obtained with the different vehicles. Substances that are nonsensitizers in either acetone or DMF are indicated as points that extend beyond 100% on the y-axis in **Figure C-8**. Substances that are nonsensitizers in AOO are indicated as points that extend beyond 100% on the x-axis. The figure suggests that EC3 values obtained with DMF and acetone are consistently lower than those obtained with AOO (i.e., the sensitizer is more potent when tested using DMF and acetone) because more points fall below the 1:1 correspondence line than above it. Spearman correlations of the log-transformed data show that the DMF ($r = 0.8743$; $p < 0.0001$) and acetone ($r = 0.8332$; $p = 0.0002$) results are significantly correlated with AOO. Negative results were arbitrarily set to EC3 = 110% so they could be used in the analysis.

Three substances were nonsensitizers when tested in acetone (i.e., benzocaine, ethylenediamine, and methyl salicylate). Methyl salicylate was also a nonsensitizer in AOO when tested at an even higher concentration. All three substances yielded sensitizer results when tested in DMF. Two of these three substances, benzocaine and ethylenediamine, were also sensitizers in guinea pigs and humans, while methyl salicylate was a nonsensitizer in both the guinea pig and human tests. Resorcinol was a nonsensitizer when tested in DMF but a sensitizer when tested in AOO (no tests in acetone were available). Resorcinol is a nonsensitizer in guinea pigs; however, the specific test protocol was not reported (Basketter et al. 1996). Although the HMT was negative (Kligman 1966), there is clinical evidence that resorcinol causes ACD in humans (Basketter et al. 2007b).

Figure C-8 Correlation of LLNA EC3 Values Between LLNA Tests with AOO and DMF or Acetone (from the NICEATM Database)



Abbreviations: ACE = acetone; AOO = acetone: olive oil (4:1 by volume); DMF = dimethylformamide; EC3 = estimated concentration of a substance expected to produce a stimulation index of 3, which is the threshold value for a substance to be considered a sensitizer in the LLNA; LLNA = murine local lymph node assay; NICEATM = National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods.

Note: Substances that are nonsensitizers in either acetone or DMF are indicated as points that extend beyond 100% on the y-axis, and substances that are nonsensitizers in AOO are indicated as points that extend beyond 100% on the x-axis.

While vehicle may be an important determinant of the EC3 value, it may not be important for every substance tested. With respect to the LLNA potency analyses in **Section 6.1**, two-way analyses of variance, with chemical and vehicle as the factors, indicated that two vehicles were responsible for a statistically significant effect of vehicle on the LLNA EC3, propylene glycol and Pluronic L92 (see **Annex IV**). Linear regression and Spearman correlation analyses indicated that removing tests using these vehicles had no impact on the relationship of the EC3 with human DSA_{05} values for the 63 substances that were sensitizers in the LLNA and in the HMT and/or HRIPT. When tests using propylene glycol and Pluronic L92 were excluded, the linear regression slope = 0.732 and

y-intercept = 0.773. Including the tests yielded slope = 0.747 and y-intercept = 0.722. In both cases, the slopes for the regressions had $p < 0.0001$ and the correlations yielded Spearman $r = 0.692$.

In the current potency evaluation of 196 substances, 58 LLNA sensitizers had more than one test (n = 2 to 66). The coefficients of variation (CV) for these sensitizers were calculated by combining results without regard to vehicle. When multiple vehicles were used, the CVs ranged from 0.4% to 349%. The CVs of the LLNA EC3 values for substances that were also sensitizers in humans ranged from 2% to 349%, while the CVs for the corresponding DSA_{05} values ranged from 2% to 408%.

8.0 LLNA Data Quality

8.1 Adherence to National and International GLP Guidelines

From the available information, published papers, and data submissions, information on compliance with GLP guidelines was available only for data obtained from Gerberick et al. (2005), E. Debruyne (Bayer CropScience SA), and P. Botham (ECPA).

It was not feasible to formally assess the quality of the remaining LLNA data considered here. The published data on the LLNA were limited to tested concentrations and calculated SI and EC3 values. Auditing the reported values would require the original individual animal data for each LLNA experiment. Such data have not been obtained. Some of the studies were conducted according to GLP guidelines, which implies that an independent quality assurance audit was conducted. The impact of any deviations from GLP guidelines cannot be evaluated for the data reviewed here, because no data quality audits were obtained.

As noted in **Section 5.0,** the original records were not obtained for the studies included in the current evaluation. Data were available for several of the substances included in the initial ICCVAM (1999) evaluation, thus some of the raw data for these substances were available for review this time.

8.2 Data Quality Audits

Formal assessments of data quality, such as quality assurance audits, generally involve a systematic and critical comparison of the data provided in a study report to the laboratory records generated for a study.

Much of the data published by Gerberick et al. (2005) was obtained following GLP guidelines, or the studies were conducted in GLP-compliant facilities. Therefore, it was inferred that data audits had been conducted on the data (ICCVAM 1999).

A formal assessment of the quality of the remainder of the LLNA data included in this BRD was not feasible. The published data on the LLNA were limited to tested concentrations and calculated SI and EC3 values. Auditing the reported values would require obtaining the original individual animal data for each LLNA experiment. Such data were not obtained. However, some of the studies were conducted according to GLP guidelines, which implies that an independent quality assurance audit was conducted.

8.3 Impact of Deviations from GLP Guidelines

The impact of any deviations from GLP guidelines cannot be evaluated for the data reviewed in this BRD, because no information on data quality audits was obtained.

8.4 Availability of Laboratory Notebooks or Other Records

As noted in **Section 5.2,** the original records were not obtained for the studies included in this evaluation. Data were available for some of the substances included in the (ICCVAM 1999) evaluation and thus some of the raw data for these substances were available for review.

9.0 Other Scientific Reports and Reviews

Several published studies have discussed the potential for using the LLNA to assess the relative skin sensitization potency of chemicals. The following section summarizes these reviews. Reviews by collaborating scientists are grouped together in **Section 9.1** and arranged by date. Reviews by other authors follow starting with **Section 9.2**.

9.1 Basketter, Gerberick, Kimber, and Colleagues

9.1.1 Basketter et al. (2003)

Basketter and colleagues discuss the usefulness of the LLNA for hazard identification and the test method's current regulatory status. The review also discusses the potential usefulness of the method to assess relative skin sensitization potency of chemicals and incorporation of the data into risk assessments.

The authors indicate that the use of the LLNA to assess potency has been extensively evaluated in recent years. They note the following factors to consider in the use of LLNA data for potency assessments:

- How the potency is estimated from the LLNA
- The robustness of the estimation
- The relevance of the estimation
- How the potency estimation is applied for risk assessment purposes.

The authors note that several studies have shown that the calculated EC3 values, as discussed in Basketter et al. (1999a), correlate well with human potency classifications (Basketter et al. 2000; Gerberick et al. 2001; Ryan et al. 2000).

The authors note that for the LLNA potency information to be useful, it should be capable of being incorporated into risk assessments. Various published proposals discuss incorporation of EC3 values into risk assessments (Basketter et al. 2001b; Gerberick et al. 2001; Robinson et al. 2000). They propose that combining various potential exposure conditions with calculated EC3 values would provide a way to incorporate the information into risk assessments (Basketter et al. 2002; Felter et al. 2002; Felter et al. 2003).

9.1.2 Kimber et al. (2003)

This review summarizes the efforts of the ECETOC Task Force (ECETOC 2003) that was charged with recommending approaches for the measurement of potency and defining thresholds for skin sensitization. The ECETOC Task Force focused primarily on categorization of sensitizers and identification of thresholds with respect to the induction phase of skin sensitization. Based on their deliberations, the task force concluded that the LLNA is the method of choice for prospective skin sensitization potency assessments. The task force proposed the following classification for skin sensitization potency based on EC3 values:

- Extreme: $EC3 < 0.1\%$
- Strong: $0.1\% \leq EC3 < 1\%$
- Moderate: $1\% \leq EC3 < 10\%$
- Weak: $10\% \leq EC3 \leq 100\%$

Although the LLNA is preferred, the authors recognized that available data from guinea pig tests provide information of frequency and severity that could be used for potency assessments.

9.1.3 Jowsey et al. (2006)

Jowsey and colleagues discuss strategies for assessing skin sensitization without the use of animals. They also summarize the use of the LLNA for assessing the skin sensitization potential of chemicals. The authors note that the LLNA is useful for hazard characterization because it models all the events that occur during the process of skin sensitization and the extent to which skin sensitization will develop. That is, the magnitude of lymphocyte proliferation is an indicator of the extent of skin sensitization (Kimber and Dearman 1991). Based upon this observation, the authors proposed that using EC3 values derived from LLNA studies could be useful in assessing skin sensitization potency (Basketter et al. 2001b; Kimber and Basketter 1997). They also cite studies that demonstrate the accuracy and reliability of the EC3 value. They state that it consistently correlates with clinical estimates of human skin sensitization potency (Basketter et al. 2000; Dearman et al. 1998; Gerberick et al. 2001; Warbrick et al. 1999).

9.1.4 Basketter et al. (2007a)

This review provides an overview of the available data that the authors consider to be supportive of the validity of the LLNA for assessments of skin sensitization potency. The authors discuss the relevance of the LLNA EC3 value in evaluating human skin sensitization potency, the reliability of the EC3 value, and the interlaboratory transferability of the method based on EC3 values.

Most studies attempt to assign chemicals to various categories (e.g., nonsensitizers, weak sensitizers, strong sensitizers) based on predefined EC3 value cutoffs. While these studies tend to show good correlation between LLNA outcomes and human skin sensitization potential, more-recent studies have attempted to correlate experimental thresholds in humans (e.g., NOELs in HRIPTs) with the LLNA EC3 value. Although the outcomes depend on exposure conditions used in the patch tests, Basketter et al. conclude that the studies show a good relationship between EC3 values and the evaluated threshold levels (Basketter et al. 2005; Griem et al. 2003; Schneider and Akkan 2004).

The authors conclude that the EC3 value is a useful metric with which to predict the skin sensitization potential of chemicals in humans and that intra- and interlaboratory studies have shown that the EC3 value is reproducible within and among laboratories. The authors therefore propose that integration of the LLNA for potency identification in risk assessments would help to develop more accurate hazard identification and risk management strategies.

9.1.5 Gerberick et al. (2007)

In this review, the authors discuss the concept of using the LLNA to assess the skin sensitization potential of chemicals in humans. They cite several advantages of the LLNA (e.g., provides dose-response data, allows for quantification of threshold values) that make it amenable to potency determinations. They also cite several studies that have evaluated the accuracy and reliability of the EC3 value for assessing potency (Basketter and Cadby 2004; Dearman et al. 2001; Warbrick et al. 1999). These and other studies have reportedly demonstrated good correlation between LLNA potency estimates and human potency, as assessed by clinical studies and experience (Basketter et al. 2000; Gerberick et al. 2001).

Based on these findings, the authors conclude that the LLNA should be considered the preferred method for identifying human skin sensitization hazard and that it can provide important additional information regarding skin sensitization potency that facilitates scientifically sound risk assessments.

9.1.6 Ryan et al. (2007)

In this article, Ryan and colleagues review historical LLNA data from both published and unpublished sources and use the data to calculate and compare EC3 values using two different

mathematical methods: linear interpolation and log-linear extrapolation. Usually the EC3 value is calculated by linear interpolation, which uses the dose and SI data points lying immediately above and below the SI value of 3 on the dose-response curve (see the following equation):

$$EC3 = c + \left[\frac{(3-d)}{(b-d)}\right] \times (a-c)$$

Coordinates:

(a = dose concentration immediately above SI = 3, b = SI immediately above 3)

(c = dose concentration immediately below SI = 3, d = SI immediately below 3)

In instances where all the test concentrations result in SI values that are greater than 3, a log-linear extrapolation is applied using the two SI values greater than 3 with the lowest of the SI values having the lowest percent concentration (see the following equation):

$$EC3_{ex} = 2^{\wedge}\left\{\log_2(c) + \frac{(3-d)}{(b-d)} \times \left[\log_2(a) - \log_2(c)\right]\right\}$$

Coordinates:

(a = dose concentration for next to lowest SI above 3, b = next to lowest SI above 3)

(c = dose concentration for lowest SI above 3, d = lowest SI above 3)

The authors evaluate 187 data sets with at least one SI value less than 3 and at least two SI values greater than 3. They use the same data sets to calculate the EC3 values using the linear interpolation and the log-linear extrapolation methods. Based on the resulting analyses, both methods of calculation are reliable and similar 88% of the time. When differences occur, the log-linear extrapolation tends to predict a stronger classification based on EC3 potency (i.e., extreme, strong, moderate, or weak). The authors also conclude from additional analyses that the quality of the dose-response curve determines the accuracy of the log-linear extrapolated EC3 values relative to the linear interpolated EC3 values. Thus, the authors indicate that using a log-linear extrapolation in instances where a linear interpolation is not possible could avoid the need for repeat animal testing with different test concentrations and may also allow for a potency classification.

9.1.7 Loveless et al. (2010)

This paper discusses how using potency information from LLNA EC3 values is applicable to classification, labeling, and risk assessment for skin sensitization hazard. The authors ask four main questions:

1. Could an EC3 value lower than 100% be defined and used as a threshold criterion for classification and labeling of skin-sensitizing substances?
2. Is there any reason to revise the recommendations of a previous ECETOC Task Force (see **Section 9.1.2**) (ECETOC 2003) regarding specific EC3 values used for subcategorization of substances based upon potency?
3. What recommendation could be made regarding classification and labeling of preparations under GHS?
4. How could LLNA data be integrated into risk assessment and provide a rationale for using concentration responses and corresponding no-effect concentrations?

The authors made the following overall conclusions to the four questions they posed.

1. The available data does not support using an EC3 value lower than 100% as the threshold for classification and labeling of a substance as a sensitizer because many chemicals with high EC3 values (>50%) are known to be human skin sensitizers.
2. After reviewing the potency categories for characterizing contact allergens recommended by a previous ECETOC Task Force (see **Section 9.1.2**) (ECETOC 2003), the use of the recommended four subcategories (i.e., extreme, strong, moderate, and weak) appears the most appropriate and scientifically based.
3. In order to classify preparations as Category 1 skin sensitizers under the current GHS regulation (UN 2009), the potency-related classification that applies to substances also applies to mixtures.
4. The authors recommend LLNA EC3 values for determination of a no-expected-sensitization induction level that represents the first step in a quantitative risk assessment.

9.2 McGarry (2007)

This review provides an overview of concerns that were raised upon implementation of the European chemicals legislation on the registration, evaluation, authorization, and restriction of chemicals (REACH). These concerns include that the LLNA is susceptible to vehicle effects (refer also to **Section 7.0**), it has not been validated for testing mixtures,[17] and it may result in a number of false positive responses when tested with skin irritants. The author states that these concerns have become heightened given the current requirements in the REACH legislation for skin sensitization testing, which specifies that the LLNA must be used for new *in vivo* testing for skin sensitization hazards, and only under "exceptional circumstances" can another method be used (EC 2006).

The intent of this review is to address these concerns from a European regulatory perspective and to discuss the potential utility of the LLNA to provide information on skin sensitization potency of substances. Evidence of vehicle effects, both on overall LLNA results (i.e., "yes" or "no" decisions) and on potency estimates (i.e., EC3 values), is described for several commonly used vehicles. Problems associated with testing mixtures and formulations (e.g., compatibility with traditional LLNA vehicles, alteration of the active substance's bioavailability by excipients) are also described. The author concludes with a discussion of the potential utility of the LLNA for estimating skin sensitization potency, while cautioning that the EC3 should not be considered a measure of absolute potency.

9.3 Schlede et al. (2003)

This article is the culmination of a 16-year collaboration among dermatologists, industry representatives, and regulators to assign potency rankings to chemicals with skin-sensitizing properties. Clinical and experimental data on humans and results of animal tests from the scientific literature were collected on 244 substances (i.e., technically produced chemicals as well as chemically defined single ingredients of natural products). Based primarily on "expert judgment" and in combination with reviews of the published literature, each substance was assigned to one of three defined categories:

- Significant contact allergen (Category A)
- Solid-based indication for contact allergenic effects (Category B)

[17] After the publication of McGarry (2007), ICCVAM recommended that the LLNA may be used to test any chemical or product, including pesticide formulations, metals, substances in aqueous solutions, and other products such as natural complex substances and dyes unless the chemical or product to be tested has properties that may interfere with the ability of the LLNA to detect skin-sensitizing substances (ICCVAM 2010c).

- Insignificant contact allergen or questionable contact allergenic effect (Category C).

Published data from human tests were obtained with the HMT or HRIPT, while the animal data were obtained with the GPMT, BT, and/or the LLNA. Most of the human experimental data correlate with sensitizing and sensitizing/nonsensitizing animal data. However, the authors state that published data on experimental human testing are limited in most cases to older studies with insufficient experimental design and/or limited documentation.

The authors conclude that results obtained with animal data are reliable and sensitive indicators for the determination of skin sensitization potential in humans.

9.4 Zaghi and Maibach (2009)

This paper compares the correlation between LLNA EC3 DSA and the HMT DSA_{05} as an indicator of allergic potency in humans as determined by the frequency of allergic reactions measured in patch test clinic populations. Eight compounds (nickel, cobalt chloride, neomycin, potassium dichromate, formaldehyde, p-phenylenediamine, benzocaine, and mercaptobenzothiazole) were evaluated. The compounds have LLNA, HMT, and patch test results from the North American Contact Dermatitis Group and the European Surveillance System on Contact Allergies. The authors quantitatively evaluated the role that other factors play in allergic reactions by subtracting the best potency correlation value from one. The data showed an inverse correlation for the weighted frequency of patch test positive responses and the LLNA or HMT. That is, as patch test positive responses decreased, LLNA DSA and HMT DSA_{05} increased. The correlation values for the LLNA and HMT with patch test clinic data were -0.56 and -0.71, respectively.

The authors suggest that there is a possible 20% error margin in the LLNA's capacity to predict potency. Further, because the best correlation value is only -0.71 (i.e., HMT correlation results), the authors suggest that other factors may play up to a 30% role in the determination of the frequency of an allergic reaction in the general population. The authors acknowledge that numerous variances in the collection of data (i.e., different laboratories, investigators, time) might have been limitations of the analyzed data set. Still, the authors conclude that while the LLNA and HMT might adequately predict allergic potency of a substance, a model that more accurately reflects human experience and takes into account environmental factors is needed.

10.0 Animal Welfare Considerations

The proposal for using the LLNA for potency determinations does not affect its requirement for using animals or the number of animals that are required. These are defined in the ICCVAM-recommended LLNA protocol (ICCVAM 2009). However, this application could broaden the use of the LLNA protocol in place of guinea pig tests and thereby further reduce the number of guinea pigs being used to assess skin sensitization potential. The LLNA is also a refinement compared with guinea pig tests because it avoids the pain and distress that can occur in the guinea pig tests when substances cause ACD.

10.1 Rationale for the Need to Use Animals

There currently are no valid and accepted non-animal test methods to determine the ACD potential of substances and products, except for situations where human studies could be conducted ethically and where such studies would meet regulatory safety assessment requirements. Additionally, the most detailed information about the induction and regulation of immunological responses is available for mice (ICCVAM 1999).

10.2 Basis for Determining the Number of Animals Used

The number of animals used for the experimental, vehicle, and positive control groups, a minimum of four animals per group, is based on the number of animals specified in the ICCVAM-recommended LLNA protocol (ICCVAM 2009).

10.3 Reduction Considerations

Although a reduced version of the LLNA (i.e., use of only a negative control and a high-dose group) does not allow for the potency determination of a sensitizing chemical, the LLNA test method protocol (ICCVAM 2009) requires fewer mice per treatment group (a minimum of four animals per group) than either of the preferred guinea pig tests (a minimum of 10 animals/group for the Buehler test and 5 animals/group for the GPMT).

11.0 Practical Considerations

The following issues must be taken into account when assessing the practicality of an alternative to an existing test method:

- Performance evaluations
- Assessments of the laboratory equipment and supplies needed to conduct the alternative test method
- Level of personnel training
- Labor costs
- Time required to complete the test method relative to the existing test method

The time, personnel cost, and effort required to conduct the proposed test method(s) must be considered to be reasonable when compared to the existing test method it is intended to replace. No such changes are being proposed for the LLNA protocol. Therefore, the transferability, training requirements, and time and cost considerations for using the LLNA for potency determinations remain unchanged from the previous ICCVAM evaluations (ICCVAM 1999, 2010c).

12.0 References

ICCVAM Authorization Act of 2000. 2000. Public Law 106-545. 42 U.S.C. 285l-3.

Basketter DA, Blaikie L, Dearman RJ, Kimber I, Ryan CA, Gerberick GF, et al. 2000. Use of the local lymph node assay for the estimation of relative contact allergenic potency. Contact Dermatitis 42(6): 344-348.

Basketter DA, Cadby P. 2004. Reproducible prediction of contact allergenic potency using the local lymph node assay. Contact Dermatitis 50(1): 15-17.

Basketter DA, Clapp C, Jefferies D, Safford B, Ryan CA, Gerberick F, et al. 2005. Predictive identification of human skin sensitization thresholds. Contact Dermatitis 53(5): 260-267.

Basketter DA, Evans P, Gerberick GF, Kimber I. 2002. Factors affecting thresholds in allergic contact dermatitis: safety and regulatory considerations. Contact Dermatitis 47(1): 1-6.

Basketter DA, Gerberick F, Kimber I. 2007a. The local lymph node assay and the assessment of relative potency: status of validation. Contact Dermatitis 57(2): 70-75.

Basketter DA, Gerberick GF, Kimber I. 2001. Skin sensitisation, vehicle effects and the local lymph node assay. Food Chem Toxicol 39(6): 621-627.

Basketter DA, Gerberick GF, Kimber I. 2001b. Measurement of allergenic potency using the local lymph node assay. Trends Pharmacol Sci 22(6): 264-265.

Basketter DA, Gerberick GF, Kimber I, Loveless SE. 1996. The local lymph node assay: a viable alternative to currently accepted skin sensitization tests. Food Chem Toxicol 34(10): 985-997.

Basketter DA, Lea LJ, Cooper K, Stocks J, Dickens A, Pate I, et al. 1999a. Threshold for classification as a skin sensitizer in the local lymph node assay: a statistical evaluation. Food Chem Toxicol 37(12): 1167-1174.

Basketter DA, Lea LJ, Dickens A, Briggs D, Pate I, Dearman RJ, et al. 1999b. A comparison of statistical approaches to the derivation of EC3 values from local lymph node assay dose responses. J Appl Toxicol 19(4): 261-266.

Basketter DA, Sanders D, Jowsey IR. 2007b. The skin sensitization potential of resorcinol: Experience with the local lymph node assay. Contact Dermatitis 56(4): 196-200.

Basketter DA, Smith Pease CK, Patlewicz GY. 2003. Contact allergy: the local lymph node assay for the prediction of hazard and risk. Clin Exp Dermatol 28(2): 218-221.

Boukhman MP, Maibach HI. 2001. Thresholds in contact sensitization: immunologic mechanisms and experimental evidence in humans – an overview. Food Chem Toxicol 39(12): 1125-1134.

Dean JH, Twerdok LE, Tice RR, Sailstad DM, Hattan DG, Stokes WS. 2001. ICCVAM evaluation of the murine local lymph node assay: II. Conclusions and recommendations of an independent scientific peer review panel. Regul Toxicol Pharmacol 34(3): 258-273.

Dearman RJ, Hilton J, Evans P, Harvey P, Basketter DA, Kimber I. 1998. Temporal stability of local lymph node assay responses to hexyl cinnamic aldehyde. J Appl Toxicol 18(4): 281-284.

Dearman RJ, Wright ZM, Basketter DA, Ryan CA, Gerberick GF, Kimber I. 2001. The suitability of hexyl cinnamic aldehyde as a calibrant for the murine local lymph node assay. Contact Dermatitis 44(6): 357-361.

EC. 2006. Regulation (EC) No. 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and

repealing Council Regulation (EEC) No. 793/93 and Commission Regulation (EC) No. 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. OJ L 396.

ECETOC. 2003. Technical Report No. 87. Contact Sensitization: Classification According to Potency. Brussels, Belgium:European Centre for Ecotoxicology and Toxicology of Chemicals. Available: for purchase at http://www.ecetoc.org/technical-reports [accessed 3 November 2010].

ECETOC. 2008. Document No. 46. Potency Values from the Local Lymph Node Assay: Application to Classification, Labelling and Risk Assessment. Brussels, Belgium:European Centre for Ecotoxicology and Toxicology of Chemicals. Available: for purchase at http://www.ecetoc.org/technical-documents [accessed 3 November 2010].

EPA. 2003. Health Effects Test Guidelines: OPPTS 870.2600 - Skin Sensitization. EPA 712-C-03-197. Washington, DC:U.S. Environmental Protection Agency. Available: http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Revised/870r-2600.pdf [accessed 17 September 2009].

EPA. 2006a. Good Laboratory Practice Standards. Federal Insecticide, Fungicide and Rodenticide Act. 40 CFR 160.

EPA. 2006b. Good Laboratory Practice Standards. Toxic Substances Control Act. 40 CFR 792.

Fawcett T. 2006. An introduction to ROC analysis. Pattern Recognition Letters 27(8): 861-874.

FDA. 2009. Good Laboratory Practice for Nonclinical Laboratory Studies. 21 CFR 58.

Felter SP, Robinson MK, Basketter DA, Gerberick GF. 2002. A review of the scientific basis for uncertainty factors for use in quantitative risk assessment for the induction of allergic contact dermatitis. Contact Dermatitis 47(5): 257-266.

Felter SP, Ryan CA, Basketter DA, Gilmour NJ, Gerberick GF. 2003. Application of the risk assessment paradigm to the induction of allergic contact dermatitis. Regul Toxicol Pharmacol 37(1): 1-10.

Friedmann PS. 1990. The immunology of allergic contact dermatitis: the DNCB story. In: Advances in Dermatology (Dahl MV, ed). Chicago:Yearbook Medical Publishers, 175-195.

Gerberick GF, Robinson MK, Ryan CA, Dearman RJ, Kimber I, Basketter DA, et al. 2001. Contact allergenic potency: correlation of human and local lymph node assay data. Am J Contact Dermat 12(3): 156-161.

Gerberick GF, Ryan CA, Dearman RJ, Kimber I. 2007. Local lymph node assay (LLNA) for detection of sensitization capacity of chemicals. Methods 41(1): 54-60.

Gerberick GF, Ryan CA, Kern PS, Dearman RJ, Kimber I, Patlewicz GY, et al. 2004. A chemical dataset for evaluation of alternative approaches to skin-sensitization testing. Contact Dermatitis 50(5): 274-288.

Gerberick GF, Ryan CA, Kern PS, Schlatter H, Dearman RJ, Kimber I, et al. 2005. Compilation of historical local lymph node data for evaluation of skin sensitization alternative methods. Dermatitis 16(4): 157-202.

Griem P, Goebel C, Scheffler H. 2003. Proposal for a risk assessment methodology for skin sensitization based on sensitization potency data. Regul Toxicol Pharmacol 38(3): 269-290.

Haneke KE, Tice RR, Carson BL, Margolin BH, Stokes WS. 2001. ICCVAM evaluation of the murine local lymph node assay: III. Data analyses completed by the National Toxicology Program

Interagency Center for the Evaluation of Alternative Toxicological Methods. Regul Toxicol Pharmacol 34(3): 274-286.

Hutchings CV, Shum KW, Gawkrodger DJ. 2001. Occupational contact dermatitis has an appreciable impact on quality of life. Contact Dermatitis 45(1): 17-20.

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 (ICCVAM). NIH Publication No. 97-3981. Research Triangle Park, NC:National Institute of Environmental Health Sciences. Available: http://iccvam.niehs.nih.gov/docs/about_docs/validate.pdf [accessed 23 December 2010].

ICCVAM. 1999. The Murine Local Lymph Node Assay: A Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemicals/Compounds. The Results of an Independent Peer Review Evaluation Coordinated by the Interagency Coordinating Committee on the Validation of Alternative Methods and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). NIH Publication No. 99-4494. Research Triangle Park, NC:National Institute of Environmental Health Sciences. Available: http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna/llnarep.pdf [accessed 21 October 2010].

ICCVAM. 2001. ICCVAM Immunotoxicology Working Group Recommended Protocol for the Murine Local Lymph Node Assay (LLNA): Testing of Chemicals for Contact Sensitizing (Allergic Contact Dermatitis [ACD]) Potential. Research Triangle Park, NC:National Institute of Environmental Health Sciences. Available: http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna/LLNAProt.pdf [accessed 29 November 2010].

ICCVAM. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods. NIH Publication No. 03-4508. Research Triangle Park, NC:National Institute of Environmental Health Sciences. Available: http://iccvam.niehs.nih.gov/SuppDocs/SubGuidelines/SD_subg034508.pdf [accessed 4 November 2010].

ICCVAM. 2009. Recommended Performance Standards: Murine Local Lymph Node Assay. NIH Publication No. 09-7357. Research Triangle Park, NC:National Institute of Environmental Health Sciences. Available: http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna-ps/LLNAPerfStds.pdf [accessed 8 July 2010].

ICCVAM. 2010a. ICCVAM Test Method Evaluation Report on the Murine Local Lymph Node Assay: BrdU-ELISA. A Nonradioactive Alternative Test Method to Assess the Allergic Contact Dermatitis Potential of Chemicals and Products. NIH Publication No. 10-7552. Research Triangle Park, NC:National Institute of Environmental Health Sciences. Available: http://iccvam.niehs.nih.gov/methods/immunotox/llna-ELISA/TMER.htm [accessed 21 October 2010].

ICCVAM. 2010b. ICCVAM Test Method Evaluation Report on the Murine Local Lymph Node Assay: DA. A Nonradioactive Alternative Test Method to Assess the Allergic Contact Dermatitis Potential of Chemicals and Products. NIH Publication No. 10-7551. Research Triangle Park, NC:National Institute of Environmental Health Sciences. Available: http://iccvam.niehs.nih.gov/methods/immunotox/llna-DA/TMER.htm [accessed 21 October 2010].

ICCVAM. 2010c. ICCVAM Test Method Evaluation Report on Using the Murine Local Lymph Node Assay for Testing Pesticide Formulations, Metals, Substances in Aqueous Solutions, and Other Products. NIH Publication No. 10-7512. Research Triangle Park, NC:National Institute of Environmental Health Sciences.

ICH. 1996. Guidance for Industry: E6 Good Clinical Practice: Consolidated Guidance. Available: http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm073122.pdf [accessed 1 November 2010].

ISO. 2002. Part 10: Tests for irritation and delayed-type hypersensitivity reaction. In: Biological Evaluation of Medical Devices, 2nd ed. Geneva:International Organization for Standardization. Available: Available for purchase at http://www.iso.org/iso/home.htm [accessed 4 November 2010].

ISO. 2010. Part 10: Tests for irritation and delayed-type hypersensitivity reaction. In: Biological Evaluation of Medical Devices 10993, 3rd ed. Geneva:International Organization for Standardization. Available: Available for purchase at http://www.iso.org/iso/home.htm [accessed 4 November 2010].

Jowsey IR, Basketter DA, Westmoreland C, Kimber I. 2006. A future approach to measuring relative skin sensitising potency: a proposal. J Appl Toxicol 26(4): 341-350.

Jowsey IR, Clapp CJ, Safford B, Gibbons BT, Basketter DA. 2008. The impact of vehicle on the relative potency of skin-sensitizing chemicals in the local lymph node assay. Cutan Ocul Toxicol 27(2): 67-75.

Kimber I, Basketter DA. 1997. Contact sensitization: a new approach to risk assessment. Hum Ecol Risk Assess 3(3): 385-395.

Kimber I, Basketter DA, Butler M, Gamer A, Garrigue JL, Gerberick GF, et al. 2003. Classification of contact allergens according to potency: proposals. Food Chem Toxicol 41(12): 1799-1809.

Kimber I, Dearman RJ. 1991. Investigation of lymph node cell proliferation as a possible immunological correlate of contact sensitizing potential. Food Chem Toxicol 29(2): 125-129.

Kimber I, Dearman RJ. 1996. Contact hypersensitivity: immunological mechanisms. In: Toxicology of Contact Hypersensitivity (Kimber I, Maurer T, eds). London:Taylor and Francis, 4-25.

Kimber I, Dearman RJ, Basketter DA, Ryan CA, Gerberick GF, McNamee PM, et al. 2008. Dose metrics in the acquisition of skin sensitization: Thresholds and importance of dose per unit area. Regul Toxicol Pharmacol 52(1): 39-45.

Kimber I, Gerberick GF, Basketter DA. 1999. Thresholds in contact sensitization: theoretical and practical considerations. Food Chem Toxicol 37(5): 553-560.

Kligman AM. 1966. The identification of contact allergens by human assay. 3. The maximization test: a procedure for screening and rating contact sensitizers. J Invest Dermatol 47(5): 393-409.

Kligman AM, Epstein W. 1975. Updating the maximization test for identifying contact allergens. Contact Dermatitis 1(4): 231-239.

Lea LJ, Warbrick EV, Dearman RJ, Kimber I, Basketter DA. 1999. The impact of vehicle on assessment of relative skin sensitization potency of 1,4-dihydroquinone in the local lymph node assay. Am J Contact Dermat 10(4): 213-218.

Loveless SE, Api AM, Crevel RWR, Debruyne E, Gamer A, Jowsey IR, et al. 2010. Potency values from the local lymph node assay: application to classification, labelling and risk assessment. Regul Toxicol Pharmacol 56(1): 54-66.

Marzulli FN, Maibach HI. 1974. The use of graded concentrations in studying skin sensitizers: experimental contact sensitization in man. Food Cosmet Toxicol 12(2): 219-227.

Marzulli FN, Maibach HI. 1976. Effects of vehicles and elicitation concentration in contact dermatitis testing. I. Experimental contact sensitization in humans. Contact Dermatitis 2(6): 325-329.

Matheson JM. 2006. CPSC Staff Report on the Draft Proposed Revision of the FSHA "Strong Sensitizer" Supplemental Definition. Washington, DC:U.S. Consumer Product Safety Commission, 66.

McGarry HF. 2007. The murine local lymph node assay: regulatory and potency considerations under REACH. Toxicology 238(2-3): 71-89.

OECD. 1992. Test No. 406. Skin Sensitisation [adopted 17 July 1992]. In: OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects. Paris:OECD Publishing. Available: http://dx.doi.org/10.1787/9789264070660-en [accessed 30 July 2010].

OECD. 1998. OECD Series on Good Laboratory Practice and Compliance Monitoring No. 1: OECD Principles on Good Laboratory Practice [as revised in 1997]. Paris:OECD Publishing. Available: http://www.oecd.org/document/63/0,3343,en_2649_34381_2346175_1_1_1_1,00.html [accessed 30 July 2010].

OECD. 2002. Test No. 429. Skin Sensitisation: Local Lymph Node Assay [adopted 24 April 2002]. In: OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects. Paris:OECD Publishing. Available: http://dx.doi.org/10.1787/9789264071100-en [accessed 30 July 2010].

OECD. 2010a. Test No. 429. Skin Sensitization: Local Lymph Node Assay [adopted 22 July 2010]. In: OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects. Paris:OECD Publishing. Available: http://dx.doi.org/10.1787/9789264071100-en [accessed 6 August 2010].

OECD. 2010b. Test No. 442A. Skin Sensitization: Local Lymph Node Assay: DA [adopted 22 July 2010]. In: OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects. Paris:OECD Publishing. Available: http://dx.doi.org/10.1787/9789264090972-en [accessed 6 August 2010].

OECD. 2010c. Test No. 442B. Skin Sensitization: Local Lymph Node Assay: BrdU-ELISA [adopted 22 July 2010]. In: OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects. Paris:OECD Publishing. Available: http://dx.doi.org/10.1787/9789264090996-en [accessed 6 August 2010].

Politano VT, Api AM. 2008. The Research Institute for Fragrance Materials' human repeated insult patch test protocol. Regul Toxicol Pharmacol 52(1): 35-38.

Robinson MK, Gerberick GF, Ryan CA, McNamee P, White IR, Basketter DA. 2000. The importance of exposure estimation in the assessment of skin sensitization risk. Contact Dermatitis 42(5): 251-259.

Ryan CA, Chaney JG, Gerberick GF, Kern PS, Dearman RJ, Kimber I, et al. 2007. Extrapolating local lymph node assay EC3 values to estimate relative sensitizing potency. Cutan Ocul Toxicol 26(2): 135-145.

Ryan CA, Gerberick GF, Cruse LW, Basketter DA, Lea L, Blaikie L, et al. 2000. Activity of human contact allergens in the murine local lymph node assay. Contact Dermatitis 43(2): 95-102.

Sailstad DM, Hattan D, Hill RN, Stokes WS. 2001. ICCVAM evaluation of the murine local lymph node assay: I. The ICCVAM review process. Regul Toxicol Pharmacol 34(3): 249-257.

Schlede E, Aberer W, Fuchs T, Gerner I, Lessmann H, Maurer T, et al. 2003. Chemical substances and contact allergy - 244 substances ranked according to allergenic potency. Toxicology 193(3): 219-259.

Schneider K, Akkan Z. 2004. Quantitative relationship between the local lymph node assay and human skin sensitization assays. Regul Toxicol Pharmacol 39(3): 245-255.

Skoet R, Zachariae R, Agner T. 2003. Contact dermatitis and quality of life: a structured review of the literature. Br J Dermatol 149(3): 452-456.

Steel RGD, Torrie JH. 1980. Principles and Procedures of Statistics: A Biometrical Approach. New York:McGraw Hill.

UN. 2009. Globally Harmonized System of Classification and Labelling of Chemicals (GHS). 3rd rev. ed. New York:United Nations.

Warbrick EV, Dearman RJ, Lea LJ, Basketter DA, Kimber I. 1999. Local lymph node assay responses to paraphenylenediamine: intra- and inter-laboratory evaluations. J Appl Toxicol 19(4): 255-260.

White SI, Friedmann PS, Moss C, Simpson JM. 1986. The effect of altering area of application and dose per unit area on sensitization by DNCB. Br J Dermatol 115(6): 663-668.

WHO/IPCS. 2007. General Conclusions and Recommendations of an IPCS International Workshop on Skin Sensitization in Chemical Risk Assessment. IPCS International Workshop on Skin Sensitization in Chemical Risk Assessment. Berlin, Germany. Geneva:World Health Organization.

Wright ZM, Basketter DA, Blaikie L, Cooper KJ, Warbrick EV, Dearman RJ, et al. 2001. Vehicle effects on skin sensitizing potency of four chemicals: assessment using the local lymph node assay. Int J Cosmet Sci 23(2): 75-83.

Zaghi D, Maibach HI. 2009. The local lymph node assay compared with the human maximization test as an indicator of allergic potency in humans using patch test clinic populations. Cutan Ocul Toxicol 28(2): 61-64.

13.0 Glossary

Accuracy*: (a) The closeness of agreement between a test method result and an accepted reference value. (b) The proportion of correct outcomes of a test method. It is a measure of test method performance and one aspect of *relevance*. The term is often used interchangeably with "concordance" (see also *two-by-two table*). Accuracy is highly dependent on the prevalence of positives in the population being examined.

Allergic contact dermatitis (ACD)*: A Type IV allergic reaction of the skin that results from skin contact with a skin sensitizer. Clinical signs of ACD include, but are not limited to, development of erythema (redness) and edema (swelling), blistering, and itching. Also referred to as skin sensitization.

Assay*: The experimental system used. Often used interchangeably with *test* and *test method*.

Classification rate: The correct classification rate is the proportion of substances that are correctly assigned to a human potency category by the LLNA (or guinea pig) result. The underclassification rate is the proportion of substances that are incorrectly assigned to a less severe human potency category by the LLNA (or guinea pig) result. The overclassification rate is the proportion of substances that are incorrectly assigned to a more severe human potency category by the LLNA (or guinea pig) result.

Coded substances: Substances labeled by code rather than name so that they can be tested and evaluated without knowledge of their identity or anticipation of test results. Coded substances are used to avoid intentional or unintentional bias when evaluating laboratory or test method performance.

Concordance*: The proportion of all substances tested that are correctly classified as positive or negative. It is a measure of test method performance and one aspect of *relevance*. The term is often used interchangeably with *accuracy* (see also *two-by-two table*). Concordance is highly dependent on the prevalence of positives in the population being examined.

EC3: the concentration of a substance estimated from the dose response curve to produce a three-fold increase in stimulation index, as compared to the concurrent vehicle control. The EC3 is the threshold value for a substance to be considered a sensitizer in the LLNA.

Essential test method component*: Structural, functional, and procedural elements of a test method that are used to develop the test method protocol. These components include unique characteristics of the test method, critical procedural details, and quality control measures. Adherence to essential test method components is necessary when the acceptability of a proposed test method is being evaluated based on performance standards derived from mechanistically and functionally similar validated test method. [Note: Previously referred to as *minimum procedural standards*]

False negative*: A substance incorrectly identified as negative by a test method.

False negative rate*: The proportion of all positive substances falsely identified by a test method as negative (see *two-by-two table*). It is one indicator of test method accuracy.

False positive*: A substance incorrectly identified as positive by a test method.

False positive rate*: The proportion of all negative substances that are falsely identified by a test method as positive (see *two-by-two table*). It is one indicator of test method accuracy.

Good Laboratory Practices (GLP)*: Regulations promulgated by the U.S. Food and Drug Administration and the U.S. Environmental Protection Agency, and principles and procedures adopted by the Organisation for Economic Co-operation and Development and Japanese authorities

* Definition used by ICCVAM (2003).

that describe record keeping and quality assurance procedures for laboratory records that will be the basis for data submissions to national regulatory agencies.

Hazard*: The potential for an adverse health or ecological effect. A hazard potential results only if an exposure occurs that leads to the possibility of an adverse effect being manifested.

Human threshold response: In the evaluation included in this BRD, the threshold for induction of skin sensitization was considered to be the no observed effect level (NOEL, expressed as $\mu g/cm^2$) or, in the absence of negative data, the lowest observed effect level (LOEL, expressed as $\mu g/cm^2$), as described by Basketter et al. (2005).

Interlaboratory reproducibility*: A measure of whether different qualified laboratories using the same protocol and test substances can produce qualitatively and quantitatively similar results. Interlaboratory reproducibility is determined during the prevalidation and validation processes and indicates the extent to which a test method can be transferred successfully among laboratories.

Intralaboratory repeatability*: The closeness of agreement between test results obtained within a single laboratory when the procedure is performed on the same substance under identical conditions within a given time period.

Intralaboratory reproducibility*: The first stage of validation; a determination of whether qualified people within the same laboratory can successfully replicate results using a specific test protocol at different times.

Immunological: Relating to the immune system and immune responses.

In vivo: In the living organism. Refers to assays performed in multicellular organisms.

Murine local lymph node assay (LLNA): An *in vivo* test method used to assess the skin sensitization potential of a substance by measuring the proliferation of lymphocytes in the lymph nodes draining the ears (i.e., auricular lymph nodes) of mice, subsequent to topical exposure on the ear to the substance. The LLNA measures lymphocyte proliferation by quantifying the amount of ^{3}H-thymidine or ^{125}I-iododeoxyuridine incorporated into the cells of the draining lymph nodes.

Lymphocyte: A white blood cell found in the blood, lymph, and lymphoid tissues, which regulates and plays a role in acquired immunity.

Negative predictivity*: The proportion of correct negative responses among substances testing negative by a test method (see *two-by-two table*). It is one indicator of test method accuracy. Negative predictivity is a function of the sensitivity of the test method and the prevalence of negatives among the substances tested.

Nonsensitizer: A substance that does not cause skin sensitization following skin contact.

Performance*: The accuracy and reliability characteristics of a test method (see *accuracy*, *reliability*).

Positive control: A substance known to induce a positive response, which is used to demonstrate the sensitivity of the test method and to allow for an assessment of variability in the conduct of the assay over time. For most test methods, the positive control substance is tested concurrently with the test substance and the vehicle/solvent control. However, for some *in vivo* test methods, periodic studies using a positive control substance is considered adequate by the OECD.

Positive predictivity*: The proportion of correct positive responses among substances testing positive by a test method (see *two-by-two table*). It is one indicator of test method accuracy. Positive predictivity is a function of the sensitivity of the test method and the prevalence of positives among the substances tested.

* Definition used by ICCVAM (2003).

Potency: For the purposes of this BRD, *potency* is defined as a function of the concentration of a substance that is required for either the induction or elicitation of a skin sensitization reaction. For induction, *potency* refers to the concentration of a substance needed to induce a skin sensitization response; the more potent the substance the smaller the quantity needed for induction. Likewise, for elicitation, potency refers to the concentration of a substance need to elicit a response in a previously sensitized individual; the more potent a substance, the smaller the quantity required for elicitation.

Prevalence*: The proportion of positives in the population of substances tested (see *two-by-two table*).

Protocol*: The precise, step-by-step description of a test, including the listing of all necessary reagents, criteria and procedures for the evaluation of the test data.

Quality assurance*: A management process by which adherence to laboratory testing standards, requirements, and record keeping procedures is assessed independently by individuals other than those performing the testing.

Reduction alternative*: A new or modified test method that reduces the number of animals required.

Reference test method*: The accepted *in vivo* test method used for regulatory purposes to evaluate the potential of a test substance to be hazardous to the species of interest.

Refinement alternative*: A new or modified test method that refines procedures to lessen or eliminate pain or distress in animals or enhances animal well-being.

Relevance*: The extent to which a test method correctly predicts or measures the biological effect of interest in humans or another species of interest. Relevance incorporates consideration of the *accuracy* or *concordance* of a test method.

Reliability*: A measure of the degree to which a test method can be performed reproducibly within and among laboratories over time. It is assessed by calculating intra- and interlaboratory reproducibility and intralaboratory repeatability.

Replacement alternative*: A new or modified test method that replaces animals with non-animal systems or replaces one animal species with a phylogenetically lower one (e.g., a mammal with an invertebrate).

Reproducibility*: The consistency of individual test results obtained in a single laboratory (intralaboratory reproducibility) or in different laboratories (interlaboratory reproducibility) using the same protocol and test substances (see intra- and interlaboratory reproducibility).

Sensitivity*: The proportion of all positive substances that are classified correctly as positive in a test method. It is a measure of test method accuracy (see *two-by-two table*).

Skin sensitizer: A substance that will lead to an allergic response following skin contact (UN 2009).

Specificity*: The proportion of all negative substances that are classified correctly as negative in a test method. It is a measure of test method accuracy (see *two-by-two table*).

Stimulation index (SI): A value calculated for the LLNA to assess the skin sensitization potential of a test substance. The value is calculated as the ratio of radioactivity incorporated into the auricular lymph nodes of a group of treated mice to the radioactivity incorporated into the corresponding lymph nodes of a group of vehicle control mice. For the LLNA, an SI equal to or greater than 3 classifies a substance as a potential skin sensitizer.

Test*: The experimental system used. Often used interchangeably with *test method* and *assay*.

Test method*: A process or procedure used to obtain information on the characteristics of a substance or agent. Toxicological test methods generate information regarding the ability of a

* Definition used by ICCVAM (2003).

substance or agent to produce a specified biological effect under specified conditions. Often used interchangeably with *test* and *assay*. See also *validated test method* and *reference test*.

Transferability*: The ability of a test method or procedure to be accurately and reliably performed in different, competent laboratories.

Two-by-two table*: The two-by-two table can be used for calculating accuracy (concordance) ([a + d]/[a + b + c + d]), negative predictivity (d/[c + d]), positive predictivity (a/[a + b]), prevalence ([a + c]/[a + b + c + d]), sensitivity (a/[a + c]), specificity (d/[b + d]), false positive rate (b/[b + d]), and false negative rate (c/[a + c]).

		New Test Outcome		
		Positive	**Negative**	**Total**
Reference Test Outcome	**Positive**	a	c	a + c
	Negative	b	d	b + d
	Total	a + b	c + d	a + b + c + d

Validated test method*: An accepted test method for which validation studies have been completed to determine the relevance and reliability of this method for a specific proposed use.

Validation*: The process by which the reliability and relevance of a procedure are established for a specific purpose.

Vehicle control: An untreated sample containing all components of a test system, including the vehicle that is processed with the test substance-treated and other control samples to establish the baseline response for the samples treated with the test substance dissolved in the same vehicle.

Weight-of-evidence (process): The strengths and weaknesses of a collection of information are used as the basis for a conclusion that may not be evident from the individual data.

* Definition used by ICCVAM (2003).

This page intentionally left blank