

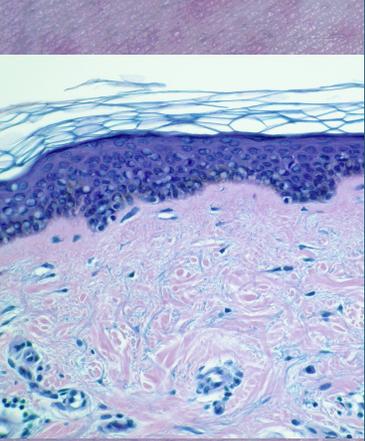


NICEATM

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

ICCVAM

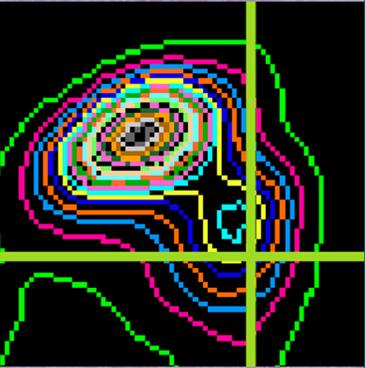
Interagency Coordinating
Committee on the Validation
of Alternative Methods



INDEPENDENT SCIENTIFIC PEER REVIEW PANEL REPORT



Validation Status of New Versions
and Applications of the
Murine Local Lymph Node Assay:
A Test Method for Assessing the
Allergic Contact Dermatitis Potential
of Chemicals and Products



May 2008



**Independent Scientific Peer Review Panel Report:
Validation Status of New Versions and Applications of the
Murine Local Lymph Node Assay: A Test Method for Assessing
the Allergic Contact Dermatitis Potential of Chemicals and
Products**

May 2008

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

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

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

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

**This document is available electronically at:
http://iccvam.niehs.nih.gov/docs/immunotox_docs/LLNAPRPrept2008.pdf**

**The findings and conclusions of this report are those of the
Independent Scientific Peer Review Panel and should not be construed
to represent the official views of ICCVAM or its member agencies.**

Table of Contents

List of Tables	iii
List of Abbreviations and Acronyms	iv
Members of the Independent Scientific Peer Review Panel	vi
Preface	ix
Executive Summary	xiii
1.0 Murine Local Lymph Node Assay (LLNA) Limit Dose Procedure	1-1
1.1 Comments on the Draft Background Review Document (BRD) for Completeness, Errors, and Omissions	1-1
1.2 Comments on the Validation Status of the LLNA Limit Dose Procedure	1-4
1.3 Comments on the Draft ICCVAM Test Method Recommendations on the LLNA Limit Dose Procedure	1-8
2.0 LLNA for Testing Aqueous Solutions, Metals, and Mixtures	2-1
2.1 Comments on the Draft Addendum for Completeness, Errors, and Omissions.....	2-1
2.2 Comments on the Validation Status of the Traditional LLNA for Testing Aqueous Solutions, Metals and Mixtures	2-1
2.3 Comments on the Draft ICCVAM Test Method Recommendations on the Traditional LLNA for Testing Aqueous Solutions, Metals, and Mixtures	2-3
3.0 Non-Radioactive LLNA Protocol - The LLNA: Daicel Adenosine Triphosphate (LLNA: DA) Test Method	3-1
3.1 Comments on the Draft BRD for Completeness, Errors, and Omissions.....	3-1
3.2 Comments on the Validation Status of the LLNA: DA	3-2
3.3 Comments on the Draft ICCVAM Test Method Recommendations on the LLNA: DA	3-6

4.0 Non-Radioactive LLNA Protocol - The LLNA: Bromodeoxyuridine Detected by Flow Cytometry (BrdU-FC) Test Method4-1

4.1 Comments on the Draft BRD for Completeness, Errors, and Omissions 4-1

4.2 Comments on the Validation Status of the LLNA: BrdU-FC 4-3

4.3 Comments on the Draft ICCVAM Test Method Recommendations on the LLNA: BrdU-FC 4-6

5.0 Non-Radioactive LLNA Protocol - The LLNA: Bromodeoxyuridine Detected by ELISA (BrdU-ELISA) Test Method5-1

5.1 Comments on the Draft BRD for Completeness, Errors, and Omissions 5-1

5.2 Comments on the Validation Status of the LLNA: BrdU-ELISA 5-3

5.3 Comments on the Draft ICCVAM Test Method Recommendations on the LLNA: BrdU-ELISA 5-7

6.0 Draft ICCVAM LLNA Performance Standards.....6-1

6.1 Comments on the Proposed Purpose and Applicability 6-1

6.2 Comments on the Essential Test Method Components..... 6-2

6.3 Comments on the Proposed Reference Substances..... 6-3

6.4 Comments on the Test Method Accuracy Standards 6-5

6.5 Comments on the Test Method Reliability Standards..... 6-7

6.6 Summary..... 6-8

6.7 Additional Statistical Comments 6-9

7.0 Use of the LLNA for Potency Determinations.....7-1

7.1 Comments on the Draft BRD for Completeness, Errors, and Omissions 7-1

7.2 Comments on the Validation Status of the Traditional LLNA to Determine Skin Sensitization Potency 7-1

7.3 Comments on the Draft ICCVAM Test Method Recommendations for the Use of the LLNA for Potency Determination 7-5

8.0 References8-1

Appendix A: Peer Review Panel Member BiosketchesA-1

Appendix B: Questions for the Peer Review Panel.....B-1

List of Tables

Table 1-1	Power Calculations for the Traditional LLNA	1-5
Table 3-1	Power Calculations for the LLNA: DA	3-3
Table 4-1	Power Calculations for the LLNA: BrdU-FC	4-8
Table 5-1	Power Calculations for the LLNA: BrdU-ELISA.....	5-2

List of Abbreviations and Acronyms

ACD	Allergic contact dermatitis
ANOVA	Analysis of variance
AOO	Acetone: olive oil (4:1)
BRD	Background Review Document
BrdU	Bromodeoxyuridine
BT	Buehler Test
CD4	Cluster of differentiation 4
CPSC	U.S. Consumer Product Safety Commission
CRO	Clinical research organization
CV	Coefficient of variation
DMF	Dimethylformamide
DMSO	Dimethylsulfoxide
DNCB	Dinitrochlorobenzene
EC3	Estimated concentration needed to produce a stimulation index of three
ECt	Estimated concentration needed to produce a stimulation index that is indicative of a positive response
ECVAM	European Centre for the Validation of Alternative Methods
ELISA	Enzyme Linked Immunosorbent Assay
eLLNA: BrdU-FC	Enhanced LLNA with BrdU detected by flow cytometry
EPA	U.S. Environmental Protection Agency
FC	Flow cytometry
FR	<i>Federal Register</i>
GLP	Good Laboratory Practice
GPMT	Guinea Pig Maximization Test
GSK	GlaxoSmithKline
HCA	Hexyl cinnamic aldehyde
HMT	Human Maximization Test
HRIPT	Human Repeat Insult Patch Test
HTdR	³ H-Methyl Thymidine
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
ISO	International Organization for Standardization
IWG	Immunotoxicity Working Group
JaCVAM	Japanese Center for Validation of Alternative Methods
LLNA	Local Lymph Node Assay
LLNA: BrdU-ELISA	LLNA with BrdU detected by ELISA
LLNA: BrdU-FC	LLNA with BrdU detected by FC
LLNA: DA	LLNA: Daicel Adenosine Triphosphate
LNC	Lymph node cells
LOEL	Lowest observed effect level
MEK	Methyl ethyl ketone
NICEATM	NTP Interagency Center for the Evaluation of Alternative Toxicological Methods

NIEHS	National Institute of Environmental Health Sciences
NIH	National Institutes of Health
NOEL	No observed effect level
NTP	National Toxicology Program
OD	Optical density
OECD	Organisation for Economic Co-operation and Development
QSAR	Quantitative structure-activity relationship
REACH	Registration, Evaluation, and Authorisation of Chemicals
rLLNA	Reduced LLNA
SAR	Structure-activity relationship
SD	Standard deviation
SI	Stimulation index
SDS	Sodium dodecyl sulfate
SLS	Sodium lauryl sulfate
TG	Test Guideline
Th	T-helper
vs.	Versus

Members of the Independent Scientific Peer Review Panel

Michael Luster, Ph.D. (Panel Chair), Senior Consultant to the National Institute of Occupational Safety and Health (NIOSH) Health Effects Laboratory, Morgantown, WV

Nathalie Alépée, Ph.D., Associate Research Fellow, Pfizer PDRD MCT Laboratory, France

Anne Marie Api, Ph.D., Vice President, Human Health Sciences, Research Institute for Fragrance Materials, Woodcliff Lake, NJ

Nancy Flournoy, M.S., Ph.D. (Biostatistician), Professor and Chair, Dept. of Mathematics & Statistics, University of Missouri-Columbia, Columbia, MO

Thomas Gebel, Ph.D., Regulatory Toxicologist, Federal Institute for Occupational Safety & Health, Dortmund, Germany

Sidney Green, Ph.D.,¹ Graduate Professor, Howard University, Washington, DC

Kim Headrick, B. Admin., B.Sc., International Harmonization Senior Policy Advisor, Health Canada, Ottawa, Ontario, Canada

Dagmar Jírová, M.D., Ph.D., Toxicologist, Research Manager, Head of Reference Center for Cosmetics, Head of Reference Laboratory for Experimental Immunotoxicology, National Institute of Public Health, Czech Republic

David Lovell, Ph.D. (Biostatistician), Reader in Medical Statistics, Postgraduate Medical School, University of Surrey, Guildford, Surrey, U.K.

Howard Maibach, M.D., Professor, Department of Dermatology, University of California-San Francisco, San Francisco, CA

James McDougal, Ph.D., Professor and Director of Toxicology Research, Department of Pharmacology and Toxicology, Boonshoft School of Medicine, Wright State University, Dayton, OH

Michael Olson, Ph.D., Director of Occupational Toxicology, Corporate Environment Health and Safety, GlaxoSmithKline, Research Triangle Park, NC

Raymond Pieters, Ph.D., Associate Professor, Immunotoxicology Group Leader, Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands

Jean Regal, Ph.D., Professor, Department of Pharmacology, University of Minnesota Medical School, Duluth, MN

Jonathan Richmond, MB ChB, FRCSEd,¹ Head of Division, Animals Scientific Procedures Division, Home Office, London, U.K.

Peter Theran, V.M.D., Massachusetts Society for the Prevention of Cruelty to Animals, Novato, CA

¹ Drs. Green and Richmond were unable to attend the public meeting on March 4-6, 2008. However, they were involved in the review of the background review documents and concur with the conclusions and recommendations included in this report.

Stephen Ullrich, Ph.D., Dallas/Ft. Worth Living Legends Professor & Professor of Immunology, Post-graduate School of Biomedical Science, University of Texas M.D. Anderson Cancer Center, Houston, TX

Michael Woolhiser, Ph.D., Technical Leader - Immunotoxicology, Toxicology & Environmental Research & Consulting Immunology, Dow Chemical, Midland, MI

Takahiko Yoshida, M.D., Ph.D., Professor, Department of Health Science, Asahikawa Medical College, Hokkaido, Japan

[This Page Intentionally Left Blank]

Preface

In 1999, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) recommended the murine local lymph node assay (LLNA) to U.S. Federal agencies as a valid substitute for currently accepted guinea pig test methods to assess the allergic contact dermatitis potential of many, but not all, types of substances. The recommendation was based on a comprehensive evaluation of the validation status of the LLNA that included an assessment by an international independent scientific peer review panel (hereafter, Panel). The Panel report and the ICCVAM LLNA test method recommendations (ICCVAM 1999) are available at the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)-ICCVAM website.¹ The LLNA was subsequently incorporated into national and international test guidelines for the assessment of skin sensitization (OECD 2002; ISO 2002; EPA 2003). For this Panel report, this LLNA will be referred to as the “traditional” LLNA.

On January 10, 2007, the U.S. Consumer Product Safety Commission (CPSC) formally requested through NICEATM that ICCVAM assess the validation status of:²

- The traditional LLNA as a stand-alone assay for potency determinations (including severity) for the purpose of hazard classification
- Three modifications of the traditional LLNA not requiring the use of radioactive materials
- The LLNA limit dose procedure (also referred to as the "reduced" LLNA)
- The ability of the traditional LLNA to test mixtures, metals, and aqueous solutions (i.e., to re-evaluate the applicability domain for the traditional LLNA)

NICEATM, in coordination with ICCVAM and the ICCVAM Immunotoxicity Working Group, prepared a comprehensive draft background review document (BRD) for each modified version of the traditional LLNA test method being evaluated, as well as a draft applicability domain addendum to the final BRD published previously on the traditional LLNA. Each draft BRD and the draft addendum detailed the available data and information from the published literature and submissions received in response to a 2007 *Federal Register (FR)* notice that had requested data related to CPSC’s nomination (*FR* notice Vol. 72, No. 95, p. 27815-27817, May 17, 2007). In addition, ICCVAM developed draft LLNA Performance Standards intended for use in validating alternative test methods that are functionally and mechanistically similar to the traditional LLNA. Finally, ICCVAM, based on the information contained in each of the draft BRDs and the draft addendum, developed draft test method recommendations.

The various supporting documents and the draft ICCVAM recommendations were provided to a new international Panel for an independent scientific review. In addition, NICEATM announced the availability of these documents on the NICEATM-ICCVAM website

¹ The 1999 ICCVAM Panel report and recommendations can be obtained at:

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

² The CPSC nomination can be obtained at:

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

(<http://iccvam.niehs.gov>) for public comment in a *FR* notice (Vol. 73, No. 5, p. 1360-1362, January 8, 2008) and via the ICCVAM listserv. The *FR* notice also announced the public Panel meeting, to be convened at the CPSC Headquarters in Bethesda, MD on March 4–6, 2008.

The Panel was charged with:

- Reviewing each ICCVAM draft BRD and the draft addendum for completeness and identifying any errors or omissions of existing relevant data or information
- Evaluating the information in each draft BRD and the draft addendum to determine the extent to which each of the applicable criteria for validation and acceptance of toxicological test methods (ICCVAM 2003) had been appropriately addressed for the recommended use of the new versions and applications of the traditional LLNA
- Considering the ICCVAM draft test method recommendations for the following and commenting on the extent to which they are supported by the information provided in the draft BRDs and the draft addendum:
 - proposed test method uses
 - proposed recommended standardized protocols
 - proposed test method performance standards
 - proposed additional studies
- Evaluating the draft ICCVAM LLNA Performance Standards and considering whether they were adequate for assessing the accuracy and reliability of alternative test methods that are functionally and mechanistically similar to the traditional LLNA

During our public meeting in March 2008, the Panel discussed each charge, listened to public comments, and developed conclusions and recommendations for ICCVAM on each of the nominated activities. The Panel wished to emphasize that they were to consider two overall questions. They were to consider: (1) whether the validation status of the each of the above proposed modifications or alternative uses of the LLNA had been adequately characterized for its intended purpose according to established ICCVAM validation criteria (available on the NICEATM-ICCVAM website, <http://iccvam.niehs.gov>), and (2) whether proposed modifications or alternative uses of the LLNA are sufficiently accurate and reliable to be used for the identification of sensitizing substances and non-sensitizing substances in place of the traditional LLNA procedure.

This report details the Panel's independent conclusions and recommendations. ICCVAM will consider this report, along with all relevant public comments, as it develops final test method recommendations. The final ICCVAM test method recommendations will be forwarded to U.S. Federal agencies for their consideration in accordance with the ICCVAM Authorization Act of 2000 (Public Law 106-545).

The Panel gratefully acknowledges the efforts of NICEATM staff in coordinating the logistics of the peer review Panel meeting and in preparing materials for their review. The

Panel also thanks each of the test method developers, Drs. George DeGeorge (LLNA: BrdU-FC), Kenji Idehara (LLNA: DA), and Masahiro Takeyoshi, (LLNA: BrdU-ELISA) for providing summaries and additional clarifications of the non-radioactive test methods under review. Finally, as Panel Chair, I want to thank each Panel member for her or his thoughtful and objective review of these LLNA-related activities.

Michael Luster, Ph.D.
Chair, LLNA Peer Review Panel
May 2008

[This Page Intentionally Left Blank]

Executive Summary

This report describes the conclusions and recommendations of an international independent scientific peer review panel (hereafter, Panel). This Panel was charged by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) with evaluating the validation status of new versions and applications of the murine local lymph node assay (LLNA) for assessing the allergic contact dermatitis (ACD) potential of chemicals and products. The LLNA, which was first evaluated in 1999 by ICCVAM, is hereafter referred to as the “traditional LLNA” to distinguish it from other versions considered by the Panel. The new versions and applications considered include:

- The LLNA limit dose procedure (also referred to as the "reduced" LLNA¹)
- The ability of the traditional LLNA to test mixtures, metals, and aqueous solutions (i.e., a re-evaluation of the applicability domain for the traditional LLNA)
- Three modifications of the traditional LLNA not requiring the use of radioactive materials:
 - LLNA: DA (Local Lymph Node Assay: Daicel Adenosine Triphosphate)
 - LLNA: BrdU-FC (Local Lymph Node Assay: Bromodeoxyuridine detected by flow cytometry)
 - LLNA: BrdU-ELISA (Local Lymph Node Assay: Bromodeoxyuridine detected by ELISA)
- The traditional LLNA as a stand-alone assay for potency determinations (including severity) for the purpose of hazard classification

The Panel also evaluated the draft ICCVAM LLNA Performance Standards and considered whether they were adequate for assessing the accuracy and reliability of alternative test methods that are functionally and mechanistically similar to the traditional LLNA.

LLNA Limit Dose Procedure

The Panel agreed that the LLNA limit dose procedure, which normally allows for testing at one dose level, should be routinely recommended for hazard identification when used for testing purposes which do not require dose response information, because it would offer time, cost, throughput and logistical benefits as well as using fewer animals. In instances when a necessity to measure relative skin sensitization potency for the purpose of risk assessment was present, then the traditional LLNA should be used in order to generate dose response information. Still, the Panel recommended use of the LLNA limit dose procedure as

¹ As described in this report, the Panel agreed that consideration should be given to applying the same term to the LLNA limit dose procedure since in various places throughout the draft BRD it was referred to differently as either the “cut-down”, the “limit dose”, or the “reduced LLNA” (i.e., “rLLNA”). Since the European Centre for the Validation of Alternative Methods (ECVAM) has already established a naming convention of “rLLNA”, the Panel recommended adopting the ECVAM terminology to harmonize the terminology used among the international validation agencies. However, because the ICCVAM documents that were reviewed use "LLNA limit dose procedure" that term is retained in this report.

the initial testing procedure to identify sensitizers and non-sensitizers before conducting the traditional LLNA even when dose response information *is* required since if the test substance were negative in the limit dose procedure, it would not be necessary to conduct a multiple-dose LLNA test.

The draft background review document (BRD) for the LLNA limit dose procedure provides a comprehensive review of available data and information for assessing the usefulness and limitations of this modified version of the LLNA for the purpose of skin sensitization hazard classification. The Panel evaluated the draft BRD for completeness, errors, and omissions and recommended that it be updated to reflect their suggestions/corrections relating to general, statistical, and specific editorial issues. In particular, the Panel noted that the differences in terminology used for this procedure caused confusion and recommended that an internationally harmonized term be adopted. They suggested referring to the procedure as the “reduced LLNA” (i.e. “rLLNA”) since that is being used by the European Centre for the Validation of Alternative Methods (ECVAM).

The Panel concluded that the stimulation index (SI) based on the ratio of 3.0 as the cutoff value was indicative of a response that was sufficiently greater than the control and would be considered an immunologically relevant response, but recommended that statistical analyses be used to definitively establish that a response induced by a test substance is significantly different from the vehicle control. The Panel agreed that the LLNA protocol recommended by ICCVAM (ICCVAM 1999; Dean et al. 2001) should be the standard protocol for all future LLNA limit dose studies using the traditional LLNA protocol. Specifically, prospective LLNA limit dose procedure studies should require that lymph nodes be collected from individual animals instead of pooling them with other animals in a treatment group, which is also currently permitted by the Organisation for Economic Co-operation and Development (OECD) Test Guideline 429 (OECD 2002). Individual animal response data are necessary in order to statistically analyze for differences between treated and control data. In addition, having data from individual animals also allows for identification of technical problems and outlier animals within a dose group. Based on power calculations provided as supplemental information, the Panel agreed that five animals per dose group is an appropriate number to recommend for LLNA limit dose studies following the traditional LLNA protocol. It should be noted that the Panel’s analysis of the LLNA limit dose dataset was not restricted to studies with confirmed individual animal data, and that the Panel considered data known to have been generated using pooled group data. The Panel stated that, internationally, both individual and pooled animal data have likely been used both for regulatory decisions and for in-house decisions relating to product development and risk management. In addition, the fact that the retrospective data analysis set out in the draft LLNA limit dose procedure BRD did not distinguish between individual or pooled animal data suggested that both met the quality standards for inclusion in the draft BRD.

Although they did not reach consensus, the Panel suggested that for laboratories in which the LLNA is “routinely” performed and have demonstrated the ability to consistently obtain positive results, hexyl cinnamic aldehyde (HCA) or another positive control (e.g., a substance that matches the chemical class of the test substances) could be run at intervals for quality control purposes rather than concurrent with each experiment. The Panel cited Kimber et al. (2006), which describes “routine” use of the “rLLNA” utilizing only a vehicle and a high-dose group, as a rationale for this suggestion. However, the Panel does not recommend

omitting the concurrent positive control in laboratories that perform the LLNA only “occasionally”.

Based on the analyses presented in the draft BRD, the Panel considered the accuracy of the LLNA limit dose procedure to have been adequately evaluated and compared to the traditional LLNA, mindful of the limitations associated with a retrospective evaluation. For instance, it cannot be assumed that the compounds tested in the retrospective studies were always tested at the highest possible dose unless such information was explicitly indicated. In this regard, the Panel recommended that a more detailed description of what is considered “*avoidance of excessive irritation*” and “*evidence of systemic toxicity*” be included in any LLNA protocol in order to aid in choosing the most appropriate high (i.e., limit) dose, although specific indicators of “*systemic toxicity or excessive irritation*” were not formally discussed.

The Panel agreed that it was appropriate to assume that the intra- and inter-laboratory reproducibility of the LLNA limit dose procedure and the traditional LLNA would be similar, because reproducibility is more dependent on the method than on the number of dose groups. However, reducing the number of test substances dose groups from three to one might reduce the sensitivity of the assay. The traditional LLNA may have a greater chance of correctly identifying a sensitizer even in the presence of one or more technical errors since data from three dose groups are being considered and an $SI \geq 3.0$ at any dose group would result in the substance being classified as a sensitizer. However, for the purpose of adopting an assay that uses fewer animals and provides increased throughput for testing purposes, these hypothetical considerations are not a sufficient reason to argue against use of the limit dose LLNA procedure.

LLNA for Testing Aqueous Solutions, Metals, and Mixtures

The draft ICCVAM recommendations state that, although more data are needed to assess the use of the LLNA for testing for mixtures and aqueous solutions before a recommendation can be made, the traditional LLNA appears to be useful for the testing of metal compounds, with the exception of nickel. The Panel agreed with these draft ICCVAM recommendations. Regarding the use of the LLNA for testing mixtures, the Panel acknowledged that the ability of ICCVAM to develop draft test method recommendations was limited not only by the amount of data available, but the relatively poor concordance of traditional LLNA outcomes in comparison to those obtained in guinea pig tests, and recommended that this be noted in the final ICCVAM recommendations. The term “mixtures” can represent an infinite number of materials and it would be more beneficial to specify types or formulations of mixtures that are being examined.

Regarding metals, the Panel concluded that the accuracy statistics for the traditional LLNA when compared to results obtained from evaluation in humans supported use of the traditional LLNA as a hazard identification tool for metals, with the exception of nickel, which produces variable responses. One minority opinion stated that the results for nickel compounds were not entirely questionable and that the traditional LLNA might also be suitable for testing nickel compounds. Thus, the Panel recommended further evaluation of the variable results obtained for nickel in the context of the available literature on allergic contact dermatitis to nickel in humans.

Regarding substances tested in aqueous solutions, the Panel suggested expanding the brief section of the draft test method recommendations discussing the test method protocol for the traditional LLNA to specifically point out how the conclusions of the applicability domain evaluation may affect the standard traditional LLNA protocol. For instance, it could be suggested that aqueous test solutions be avoided due to problems associated with skin application. It would be preferable for a hierarchy of organic solvents to be considered as dosing vehicles, with emphasis on using a vehicle to which humans may actually be exposed in circumstances linked to occupational sensitization.

The Panel agreed with the draft ICCVAM recommendation for continued accrual of information from traditional LLNA evaluations of mixtures, metals, and aqueous solutions with comparative data for guinea pig (i.e., guinea pig maximization test [GPMT] or Buehler test [BT]) and human (i.e., human maximization test [HMT] or human repeat insult patch test [HRIPT]) tests. However, the Panel suggested that, given resource limitations, it would be important to organize the recommendations based on relative priority.

The draft Addendum to the original validation report for the traditional LLNA (ICCVAM 1999) provided a comprehensive review of currently available data and information for evaluating the usefulness and limitations of the traditional LLNA for assessing the skin sensitization potential of mixtures, metal compounds, and substances tested in aqueous solutions. The Panel evaluated the draft Addendum for completeness, errors, and omissions and concluded that there were no apparent errors or omissions, although they did state that the term “mixtures” was used too broadly (i.e., can represent an infinite number of materials) and it would be more beneficial to specify types or formulations of mixtures that are being examined.

The Panel did not identify any classes of chemicals missing from the dataset used to review the utility of the traditional LLNA for testing aqueous solutions. However, while they did not propose an alternative, the Panel expressed concern over the most appropriate definition for an aqueous solution (defined in the draft Addendum as any solution containing $\geq 20\%$ water). For the mixtures included in the analysis, the Panel noted that quantitative compositions had not been provided and therefore they could not comment on whether these mixtures were representative of the types of mixtures typically tested in the traditional LLNA. With respect to metals (none of which are mixtures), there was a paucity of important representatives of commercially useful metals such as platinum, palladium, iron, zinc, manganese and silver in the data set. The Panel suggested that to enlarge the group of metal non-sensitizers, substances used as cosmetic ingredients (e.g., titanium dioxide) and aluminum compounds currently used in antiperspirants might be considered.

The Panel agreed that, although it was important to identify data obtained according to GLP guidelines, data obtained from non-GLP studies should not be excluded automatically from this retrospective analysis. The Panel concluded that other factors could be used to identify high quality data. Examples include data published in peer-reviewed journals or obtained from a study conducted in a laboratory that has GLP capabilities.

The Panel concluded that, considering the limited comparative data that were available, particularly for mixtures and aqueous solutions, the accuracy assessment of the traditional LLNA for testing mixtures, metals, and aqueous solutions when compared to available human and/or guinea pig test results was as comprehensive as was possible. The limited

amount of comparative data made it unfeasible to draw definitive conclusions for mixtures and aqueous solutions from the available accuracy statistics.

Non-Radioactive LLNA Protocol - The LLNA: DA Test Method

The Panel concluded that the available data and test method performance support the ICCVAM draft recommendations for the LLNA: Daicel Adenosine Triphosphate test method (LLNA: DA), and that the test method may be useful for identifying substances as potential skin sensitizers and non-sensitizers, but that this recommendation is contingent upon receipt, review, and analyses of additional existing data and information from the test method developer. Therefore, this non-radioactive version of the traditional LLNA cannot currently be recommended for the hazard identification of skin sensitizing substances, regardless of whether or not there are restrictions on the use of radioactive materials, until such time as this existing data has been received and confirmed.

The draft LLNA: DA BRD was compiled to provide a comprehensive review of available data and information evaluating the usefulness and limitations of the LLNA: DA test method to assess the allergic contact dermatitis potential of chemicals and other substances. The Panel evaluated the draft BRD for completeness, errors, and omissions and recommended that their suggestions/corrections relating to general, statistical, and specific editorial issues be incorporated into future revisions.

The Panel agreed that five animals per dose group should be recommended for validation of modified LLNA test methods. The Panel, however, noted that supplemental power calculations for the LLNA: DA test method indicated that the power for detecting a three-fold increase in the treatment group was estimated to be 95% for a sample size of three mice per dose group. Thus, the Panel identified the use of three animals per dose group as a potential opportunity to reduce animal number when using modified assays in the future, assuming all essential validation requirements can be successfully met. A minority opinion expressed by five Panel members was that if laboratories were operating under OECD guidance (OECD 2002) and a reliable validation dataset had been generated, then pooled data from a least four animals per dose group could be considered.

Generally, the Panel viewed the difference in treatment schedule between the LLNA: DA and the traditional LLNA to be potentially significant if the LLNA: DA induced the elicitation phase of skin sensitization. The Panel was concerned that the 1% sodium lauryl sulfate (SLS) pretreatment step in the LLNA: DA might modify the inherent sensitivity of the LLNA. They recommended that the test method developer (Daicel Chemical Industries, Ltd.) justify the use of 1% SLS or consider an alternative decision criterion (i.e., an SI threshold other than 3.0) such that the 1% SLS pretreatment is no longer necessary.

The Panel considered the database of substances tested in the LLNA: DA to be representative of a sufficient range of chemicals expected to be tested for skin sensitization potential, and concluded that the accuracy analysis had made appropriate comparisons to the traditional LLNA, guinea pig tests, and human data/experience. The Panel could not identify specific characteristics associated with the one false negative (i.e., 2-mercaptobenzothiazole) or the one false positive (i.e., benzalkonium chloride), but reemphasized that the potential impact of pretreatment with 1% SLS in this context needed to be considered.

With regard to test method reliability, the Panel concluded the intralaboratory reproducibility of the LLNA: DA had not been adequately evaluated. They noted that the two sensitizers tested had similar chemical structures (i.e., eugenol and isoeugenol) and that it was unclear if the tests were truly independent. The Panel also noted that the interlaboratory reproducibility of the assay could not be adequately evaluated given the lack of original laboratory data and limitations in the study design. In particular, they cited the use of pooled lymph nodes from the mice in each treatment group and the testing of each substance at predetermined dose levels established by the lead laboratory as study design limitations. Still, a Panel minority considered pooled data acceptable and the setting of dose levels for all laboratories based on results from the lead laboratory to be reasonable.

The Panel also commented that ideally, test substances should be coded during the validation of a new assay, although they did not feel that a lack of coding constituted a reason for rejecting the current LLNA: DA dataset. The Panel also commented that although GLP compliance is highly recommended for validation studies, the current studies should not be rejected solely on the basis of a lack of GLP compliance. However, the Panel considered it important to obtain the original records for all validation studies (which have been requested by NICEATM) in order to confirm that the reported data were the same as the data recorded in the laboratory notebooks.

With regard to the 5% (1/19) false negative and 10% (1/10) false positive rates obtained with the LLNA: DA, the Panel commented that it was important to identify reasons why the substances gave “false” results, taking into consideration factors such as intended use of the substances and the target population. They agreed that it might be useful to follow a suspected inaccuracy with an investigation of the mechanistic basis for the discordance since it may help to establish a biologically-based rationale for the discordance.

The Panel noted that the available LLNA: DA data did not support all of the ICCVAM draft recommendations in the proposed test method standardized LLNA: DA protocol. First, although the Panel agreed with the ICCVAM protocol that recommends five animals per dose group, they noted that supplemental statistical information provided for the LLNA: DA test method implied that using less than five animals per dose group was acceptable (e.g., a 3.0-fold increase in the SI value would likely be detected with 99% confidence when using four animals per dose group). In addition, the Panel considered it important to adequately characterize the effect of the 1% SLS pretreatment step in the LLNA: DA, and it should be demonstrated that the day 8 applications do not induce a skin reaction that could be indicative of the onset of the elicitation phase of skin sensitization. Keeping these points in mind, the Panel agreed that if the limit dose procedure was applicable to the traditional LLNA, then it would also be applicable to the LLNA: DA in order to further reduce the number of animals used.

The Panel also stated that the available data supported the ICCVAM draft recommendations for the LLNA: DA in terms of future studies, which included performing a more comprehensive evaluation using more non-sensitizers within and across laboratories. A minority opinion stated by one Panel member was that although testing more sensitizers might be warranted for interlaboratory validation studies, a sufficient number of non-sensitizers (n=11) had already been tested within the same laboratory.

The Panel also commented that the protocol differences between the LLNA: DA and the traditional LLNA could not clearly be constituted as “major” or “minor” changes. However, they considered this issue largely irrelevant if a test method was able to correctly predict the dermal sensitization potential of a test substance. Consequently, the Panel concluded that the current draft ICCVAM Performance Standards could be applicable to the LLNA: DA as a mechanistically and functionally similar test method.

Non-Radioactive LLNA Protocol - The LLNA: BrdU-FC Test Method

Overall, the Panel concluded that the available data and test method performance of the LLNA with bromodeoxyuridine (BrdU) detected by flow cytometry (LLNA: BrdU-FC) supported the draft ICCVAM recommendations that it may be useful for identifying substances as potential skin sensitizers and non-sensitizers, but that more information and existing data must be made available before the LLNA: BrdU-FC can be recommended for routine use. The Panel concluded that the test method usefulness and limitations identified in the draft ICCVAM recommendations accurately summarized the limits of the information supplied and the additional information that would need to be generated or provided for further consideration of the test method. As a result, the Panel concluded that the LLNA: BrdU-FC could not currently be considered as a scientifically valid replacement alternative to the traditional LLNA. Still, the Panel suggested that the test method recommendation should clearly state that the test method was not “invalid”, but simply that there was currently not sufficient evidence and information to state that it had been adequately validated.

The draft LLNA: BrdU-FC BRD was compiled to provide a comprehensive review of available data and information evaluating the usefulness and limitations of the LLNA: BrdU-FC test method to assess the ACD-inducing potential of substances. The Panel evaluated the draft BRD for completeness, errors, and omissions and recommended that their recommendations/corrections relating to general, statistical, and specific editorial issues be incorporated into future revisions.

The LLNA: BrdU-FC included routine measurements of ear swelling as an indicator of excessive skin irritation. The Panel viewed that this, or any other quantitative measurement of skin irritation, should be carefully considered for inclusion in all LLNA protocols. The Panel considered inclusion of optional quantification of immunophenotypic markers as an additional mechanism for distinguishing irritants from sensitizers to be useful, as it might reduce the frequency of false positives (i.e., substances which are actually skin irritants) and improve comparisons with human data. However, they considered application of immunological markers too detailed and costly for routine LLNA use (i.e., for hazard classification purposes) and more suited for research purposes.

The Panel noted that the substances tested in the LLNA: BrdU-FC seemed representative of a sufficient range of chemical classes and physical chemical properties, and thus that the test method appeared applicable to many of the types of chemicals and products that are typically tested for skin sensitization potential. However, the Panel considered the total database available for evaluation of the validation status of the LLNA: BrdU-FC to be relatively small compared to the large number of substances assessed in the traditional LLNA. Therefore, the Panel recommended caution when making conclusions related to its concordance with the

traditional LLNA. Still, the accuracy of the LLNA: BrdU-FC was considered adequately evaluated and comparable to the traditional LLNA.

The Panel concluded that intralaboratory reproducibility was not adequately assessed and it should be better evaluated in order to support the validation of this test method. The Panel suggested that although the studies evaluated in the draft BRD were not GLP-compliant, this should not affect acceptance of the data for an evaluation of the validation status of this test method. However, some sources of variability in the intralaboratory data, such as failure to appreciate differences in composition of dosing solutions between experiments caused by test article instability or other phenomena, might be obscured if not in complete compliance with GLP guidelines. Thus, the Panel suggested that any additional studies undertaken to validate the test method should ideally be GLP-compliant.

The Panel agreed that the available data supported the ICCVAM draft recommendations for the LLNA: BrdU-FC procedure in terms of the proposed test method standardized protocol. They suggested that the utility of ear swelling or other methods to detect inflammation appeared warranted in every variation of the LLNA (including the traditional LLNA), but should be further investigated before routine inclusion in the protocol is recommended. The Panel also concluded that the traditional LLNA limit dose procedure could be applied to the LLNA: BrdU-FC, keeping in mind the limitations associated with a “limit dose” procedure.

The Panel further agreed that the ICCVAM draft recommendations for future studies highlighted the unanswered questions raised by the available data set. Specifically, conducting interlaboratory studies as a part of the validation process is important. The Panel considered the immunological markers suggested for the LLNA: BrdU-FC to be acceptable, but that additional immunological markers for discrimination of irritant versus sensitization phenomena were also possible. In general, for any future work, efforts should be made to decrease the variability and to thereby increase the power of the test in order to ensure that more animals were not needed relative to the traditional LLNA or other alternative LLNA protocols.

The Panel considered the protocol differences between the LLNA: BrdU-FC and the traditional LLNA to be “minor” changes, and therefore concluded that assessment of the validity of this test method could be based on the draft ICCVAM LLNA Performance Standards. The Panel also cautioned, however, that a clear definition of what constituted a “major” versus a “minor” change, or a different protocol altogether could be better addressed once the recommendations for the current draft ICCVAM LLNA Performance Standards were finalized.

Non-Radioactive LLNA Protocol - The LLNA: BrdU-ELISA Test Method

The Panel concluded that the available data and test method performance for the LLNA with BrdU detected by enzyme-linked immunosorbent assay (LLNA: BrdU-ELISA) support the ICCVAM draft recommendations that it may be useful for identifying substances as potential skin sensitizers and non-sensitizers, but that more information and existing data must be made available before the LLNA: BrdU-ELISA can be recommended for use. The Panel also stated that a detailed protocol was needed, in addition to sufficient quantitative data for broader analysis on a larger set of balanced reference substances that take into account

physicochemical properties and sensitization potency, as well as an appropriate evaluation of interlaboratory reproducibility.

The draft LLNA: BrdU-ELISA BRD was compiled to provide a comprehensive review of available data and information evaluating the usefulness and limitations of the LLNA: BrdU-ELISA test method to assess the ACD-inducing potential of chemicals and other substances. The Panel evaluated the draft BRD for completeness, errors, and omissions and recommended that their suggestions/corrections relating to general, statistical and specific editorial issues be incorporated into the final document.

The Panel's main concern with the test method was that the accuracy of the LLNA: BrdU-ELISA at $SI \geq 3.0$ was inadequate and not equivalent to the traditional LLNA. Furthermore, although using a decision criterion of $SI \geq 1.3$ improved the test's performance in identifying sensitizers from non-sensitizers, it did not resolve concerns about the test method. Based on a power analysis for the LLNA: BrdU-ELISA, which was provided to the Panel as supplemental information, the Panel concluded that it was difficult to justify using a $SI \geq 1.3$ as the cutoff value, given the much larger number of animals that would be required to detect a 1.3-fold increase above vehicle controls with similar power to the traditional LLNA when five animals per dose group are used. For a three-fold increase, the supplemental statistical analyses indicated that a sample size of four was sufficient. Still, the Panel agreed with the ICCVAM recommendation to use five animals per dose group and to collect individual animal data. They concluded that this would allow for more robust calculations in the event that an outlier prevented some of the data from being included in the analysis. A minority opinion by five Panel members was stated that if laboratories were operating under OECD guidance (OECD 2002) and a reliable validation dataset had been generated, then pooled data from a least four animals could be considered.

The Panel noted that in organizations where the use or disposal of radioactive materials was restricted, the potential to use the LLNA: BrdU-ELISA could reduce the number of animals needed per test compared to the traditional LLNA and would result in less pain and suffering compared to using traditional guinea pig test methods. However, if the $SI \geq 1.3$ was chosen as the decision criterion because of its improved accuracy compared to $SI \geq 3.0$, the Panel stated that the number of mice needed to perform the LLNA: BrdU-ELISA test should be compared to the number of guinea pigs that would be needed for skin sensitization tests in order to assess if the LLNA: BrdU-ELISA actually reduced overall animal use for skin sensitization testing.

In general, the Panel considered the number of substances tested in the LLNA: BrdU-ELISA too few, and that data from more substances tested using the traditional LLNA, guinea pig tests, and human tests should have been included. The Panel also did not consider the available data from the LLNA: BrdU-ELISA to be representative of a sufficient range of chemical classes and physical chemical properties. The limited dataset prevents an evaluation of whether the test method would be considered applicable to any of the types of chemicals and products typically tested for skin sensitization potential.

However, the Panel concluded that the appropriate comparisons between the traditional LLNA, guinea pig test and human data had been made. The Panel agreed that the false negative rate for hazard identification using the $SI \geq 3.0$ in the LLNA: BrdU-ELISA was excessive (i.e., using this SI threshold value, the LLNA: BrdU-ELISA misclassified 29% and

39% of the substances classified as sensitizers in the traditional LLNA or in humans, respectively).

The Panel also considered that the intralaboratory reproducibility of the LLNA: BrdU-ELISA was not adequately evaluated and compared to the traditional LLNA. The Panel indicated that the number of substances was too few, and in some cases there was a wide variation in repeat tests of the same substance. The Panel recommended a more comprehensive evaluation of the intralaboratory reproducibility of the test method, using different SI values, and that the analysis of the variability of the estimated concentration needed to produce a positive SI value (EC_t values) be conducted on a log scale.

The Panel also noted that interlaboratory reproducibility for the LLNA: BrdU-ELISA could not be evaluated because neither the design of the study sponsored by the Japanese Center for Validation of Alternative Methods nor any of the resulting data had been provided in advance of their evaluation. The Panel agreed that a multi-laboratory validation study using a balanced set of chemicals would adequately characterize the interlaboratory reproducibility of the LLNA: BrdU-ELISA.

In general, the Panel agreed that the available data support the ICCVAM draft recommendations for the LLNA: BrdU-ELISA procedure in terms of the proposed test method standardized protocols. However, as noted above, a minority opinion by five Panel members was that there could be circumstances in which pooled data from at least four animals could also be acceptable. The Panel also stated that if the LLNA: BrdU-ELISA was found to be equivalent to the traditional LLNA in the future that it would be appropriate to apply the LLNA limit dose procedure to the test. The Panel also agreed with ICCVAM's test method recommendations for future studies and emphasized that more data were needed in order to determine the appropriate threshold value for the decision criterion. The Panel concluded that it might be more appropriate to use a statistically-based decision criterion rather than a stimulation index to classify substances as sensitizers, and that this should be further investigated.

The Panel agreed that the LLNA: BrdU-ELISA protocol differed from the traditional LLNA only in the method used to assess lymphocyte proliferation and as such concluded that this represented a "minor" change (as defined in the current draft ICCVAM LLNA Performance Standards) and separate performance standards for the LLNA: BrdU-ELISA were not needed.

Draft ICCVAM LLNA Performance Standards

The draft ICCVAM LLNA Performance Standards are intended to evaluate the acceptability of proposed test methods that are mechanistically and functionally similar to the traditional LLNA. ICCVAM proposed that the applicability of the draft ICCVAM LLNA Performance Standards be restricted to protocols that incorporate "minor" modifications to the traditional LLNA procedure, defined as changes only to the method for measuring lymphocyte proliferation. The Panel agreed that different methods of measuring lymphocyte proliferation represent "minor" modifications, but recommended that, instead of trying to define "minor" modifications, a better strategy might be to define criteria that would need to be satisfied in order to ensure that the alternative test method was mechanistically and functionally similar to the traditional LLNA (e.g., only measure cell proliferation associated with the induction

phase of a skin sensitization reaction). The Panel considered that the draft performance standards were also appropriate for evaluating other modifications. Examples of acceptable modifications included test animal sex, strain, the use of rats rather than mice, the number of animals per group, and timing of test article treatment. One minority opinion considered the potential impact of changes to protocol components other than the method of measuring lymphocyte proliferation to be significant and therefore would require more extensive validation, which was not defined.

The Panel indicated that alternative LLNA protocols that are undergoing validation should contain essential test method components that follow the ICCVAM-recommended protocol (ICCVAM 1999; Dean et al. 2001), unless adequate scientific rationale for deviating from this protocol was provided.

The Panel also identified aspects of the LLNA that should be required as part of the test method validation process: (1) application of the test substance to the skin with sampling of the lymph nodes draining that site, (2) measurement of cell proliferation in the draining lymph node, (3) absence of a skin reaction that could be indicative of the onset of the elicitation phase of skin sensitization, (4) data collected at the level of the individual animal to allow for an estimate of the variance within control and treatment groups (using this variance, a power analysis needs to be conducted to demonstrate that the modified method is utilizing a sufficient number of animals per treatment group to permit hazard identification with at least 95% power), and (5) if dose response information is needed, there are an adequate number of dose groups ($n \geq 3$) with which to accurately characterize the dose response for a given test substance.

The Panel noted that the list of substances included in the draft ICCVAM LLNA Performance Standards was sufficiently representative of the types of materials that are likely to be tested for skin sensitization. However, among the 13 sensitizers in the list of "required" substances, only five were considered to have robust data (i.e., traditional LLNA data based on at least three independent studies).

To evaluate performance for use in hazard identification, the Panel concluded that all 22 substances in the draft ICCVAM-recommended list should be tested and accuracy statistics calculated (Note: this list of substances includes "required" substances as well as "optional" false negative and false positive substances, of which only 8/22 have "robust" datasets [$n \geq 3$ as defined by the Panel]). To the extent possible, a rationale for any discordant results should be provided. However, the most potent sensitizers (e.g., dinitrochlorobenzene [DNCB]) should always be identifiable. Also, considerable weight should be given to the balance between animal welfare and human safety when considering the adequacy of test method accuracy. Based on the limited data available for the sensitizers on the list and the lack of standardization of test methods from which the results were obtained, the current database does not support inclusion of EC_t values as a component of the accuracy evaluation.

The Panel agreed with the draft ICCVAM recommendations for evaluating test method reliability. These recommendations included obtaining EC_t values that are generally within 0.5x to 2.0x of the mean historical EC₃ (i.e., estimated concentrations needed to produce an SI of 3) values for hexyl cinnamic aldehyde (HCA) (intralaboratory, n=4 experiments in one laboratory), or HCA and DNCB (interlaboratory, n=1 experiment in three laboratories). However, the Panel recommended that the criteria for independent tests should be specified

(e.g., different animal shipment, different reagents, different operator). The Panel concluded that the proposed criteria for acceptability appeared to be appropriate in this case, because only one or two substances were being evaluated (i.e., a statistical multiple comparisons² problem does not exist). The Panel also suggested that historical control data using HCA and DNCB in the same vehicle could be used to demonstrate adequate intra- and/or inter-laboratory reproducibility.

The Panel also recommended that statistical tests to analyze the data might allow for a more accurate interpretation. They recommended that a suitable variance-stabilizing transformation (e.g., log transformation, square root transformation) be applied in all statistical analyses and in reporting summary standard deviations. The Panel also recommended that a more rigorous evaluation be conducted of what would be considered an appropriate range of EC_t values to include as a requirement. This would be a statistical evaluation that considers the variability of EC_t values generated among the sensitizers included on the performance standards reference substances list and the statistical multiple comparisons problem.

Use of the LLNA for Potency Determinations

The Panel agreed with the draft ICCVAM recommendation that the LLNA should not be used as a stand-alone assay for categorizing skin sensitizers as strong vs. weak, but that it could be used as part of a weight-of-evidence evaluation (e.g., along with quantitative structure-activity relationships, peptide reactivity, human evidence, historical data from other experimental animal studies) for this purpose. The Panel also agreed with the draft ICCVAM recommendation that any LLNA studies conducted for the purpose of evaluating skin sensitization potency should use the ICCVAM-recommended LLNA protocol. In addition, the Panel viewed that the relevant testing guidelines for the traditional LLNA should be revised to include the procedure for calculating an EC₃ value.

A draft BRD was compiled by ICCVAM that provided a comprehensive review of available data and information and an evaluation of the usefulness and limitations of the traditional LLNA for the categorization of substances with regard to skin sensitization potency. The Panel evaluated the draft BRD for completeness, errors, and omissions and noted alternative analyses that would allow for a more complete evaluation of the use of the traditional LLNA for skin sensitization potency categorizations (see below).

The Panel agreed that the database of substances evaluated for potency determinations was sufficient and represented a range of chemical classes and physicochemical properties applicable to products typically tested for skin sensitization potential. The Panel also concluded that since the database was compiled from existing data, the lack of substance coding likely had no impact on the retrospective evaluation presented in the draft BRD. Still, the Panel recommended the coding of test substances in any future validation studies. The

² When multiple experiments are conducted and multiple observations, comparisons, or hypothesis tests are conducted, the chance of observing rare events increases. Suppose, for example, that an interval is established such that 5% of observations from a particular population of data are outside that interval. Then if k independent experiments generate data from this population (e.g., a standard normal distribution), the chances that all 20 results will lie inside the interval is $(1.0 - 0.05)^k$ (N. Flournoy, personal communication).

Panel generally agreed that potency determinations based on traditional LLNA results should ideally be limited to data from studies that evaluated lymph node proliferation in individual animals so that outliers and technical errors could be identified. However, they also agreed that pooled animal data should not be excluded automatically from a retrospective analysis.

The Panel indicated that the relevance of the LLNA for potency determinations had been adequately compared and evaluated to human (i.e., HMT or HRIPT) and guinea pig (i.e., GPMT or BT) data. A minority opinion stated by one Panel member was that the relevance of the traditional LLNA to human clinical observations had not been sufficiently determined.

In general, the Panel agreed that the proposed two-level categorization scheme (weak vs. strong sensitizers) for both human and guinea pig data was appropriate. However, a minority opinion stated by two Panel members was that a moderate category should be included since certain compounds might be on the border between weak and strong sensitizers. Thus, they suggested that the five-category scheme proposed by Kimber et al. (2003), which includes non-sensitizers, might be recommended.

The Panel concluded that the decision criteria providing the best overall performance was the use of $<250 \mu\text{g}/\text{cm}^2$ to distinguish between strong and weak sensitizers in humans and the use of an LLNA EC3 $\leq 9.4\%$ to distinguish between strong and weak sensitizers in the LLNA. The Panel stated that more data would be needed to determine if values different from these two would be more appropriate. The Panel also recommended that safety factors other than 10 for the lowest observed effect level (LOEL) be evaluated to determine if improved results could be obtained. The Panel also suggested an analysis that directly compares the LOEL values without using a safety factor (i.e., using LOEL data only) and an analysis that only uses no observed effect level data. The Panel further stated that traditional LLNA tests based on pooled or individual lymph nodes for a dose group should be evaluated independently to assess the impact of using pooled data on the accuracy analysis for skin sensitization potency. Finally, the Panel stated that the effect of different vehicles should be recognized as a limitation in the current data analysis and a likely contributor to the variability observed within and across laboratories.

The Panel stated that data from studies that could not be confirmed as being GLP-compliant, but that were from peer-reviewed literature or sources with high-quality laboratory management practices, were still appropriate to include in the accuracy analysis. However, the Panel stated that, ideally, GLP compliance should be the standard, as it is clearly the only objective way to judge the credibility of the data.

The Panel recommended that more data should be collected to determine the optimal threshold in humans for distinguishing between strong and weak sensitizers. In addition, the Panel discouraged conducting additional animal studies unless such studies would be expected to lead to an overall reduction in animal use. The Panel recommended that the LOELs from Akkan et al. (2003) be used instead of the DSA_{05} (i.e., the dose per skin area leading to a sensitization incidence of 5%) values from Schneider and Akkan (2004) in all of the potency analyses. A minority opinion by one Panel member stated that it was acceptable to use the DSA_{05} values from Akkan et al. (2003) as LOEL values in the evaluation. This panelist mentioned that the DSA_{05} value is a LOEL value adjusted to 5% incidence of induction in order to correct for human studies leading to different inductions. Furthermore, the panelist stated that because the DSA_{05} is corrected for an induction rate of 5%, it would

be better to compare with the traditional LLNA EC3 than to use the default uncorrected LOEL.

1.0 Murine Local Lymph Node Assay (LLNA) Limit Dose Procedure¹

1.1 Comments on the Draft Background Review Document (BRD) for Completeness, Errors and Omissions

1.1.1 General Comments

The international independent scientific peer review panel (hereafter, Panel) was asked if there were errors in the draft LLNA limit dose procedure BRD that should be corrected, if omissions of existing relevant data had been identified, or if there was additional information that should be included. The Panel agreed that consideration should be given to applying the same term to the LLNA limit dose procedure since in various places throughout the draft BRD it was referred to differently as either the “cut-down”, the “limit dose”, or the “reduced LLNA” (rLLNA). Since the European Centre for the Validation of Alternative Methods (ECVAM) has already established a naming convention of “rLLNA”, the Panel recommended adopting the ECVAM terminology to harmonize the terminology used among the international validation agencies.

The Panel recommended that since the validation of the LLNA limit dose procedure encompassed data that was analyzed retrospectively, a discussion of the limitations of a retrospective evaluation of previously published LLNA results should be included in the final BRD. In particular, the assumption that the highest dose in the retrospective dose-response study would be equivalent to the highest possible dose tested in the limit dose procedure should be addressed. Discussing such a limitation would be important since it bears directly on the validation of the limit dose procedure.

Further, since determination of the appropriate “limit dose” is critical to the LLNA limit dose procedure, the Panel suggested that a discussion of how to arrive at the maximal concentration for test substance dosing should be included in the final BRD. The final BRD should also specifically define what is meant by the terms “avoidance of excessive irritation” and “systemic toxicity” to aid in choosing the most appropriate maximum dose. In this regard, the Panel suggested that a systematic and quantitative measurement of ear thickness and systemic toxicity be considered or evaluated for routine inclusion in the LLNA protocol.

The Panel discussed modifying the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) LLNA protocol requirement for testing concurrent positive controls (ICCVAM 1999; Dean et al. 2001) as a means of further streamlining the LLNA limit dose procedure (i.e., reducing animal number, cost, etc.). Although the Panel did not reach consensus, a suggestion was made that for laboratories in which the LLNA is “routinely” performed and which had demonstrated the ability to consistently obtain positive results, hexyl cinnamic aldehyde (HCA) or another positive control (e.g., a substance that matches the chemical class of the test substances) could be run at intervals for quality control purposes rather than concurrent with each experiment. The Panel noted that Kimber et al. (2006) have described the “routine” use of the “rLLNA” utilizing only a vehicle and a high-dose group. The Panel also recommended that for laboratories that perform the LLNA only “occasionally”, a concurrent positive control should be used. However, in their discussions,

¹ Also referred to as the “reduced” LLNA.

the Panel was not able to conclude what would constitute “routine” or “occasional” LLNA use or what would be an appropriate interval between positive control testing when a concurrent positive control is not used.

The Panel also noted that including the following additional information in the final BRD might prove informative if included:

- An indication of any procedural problems reported for the tests
- An indication of the range of historical values obtained with the negative and positive controls (the positive control historical range might give insight into the need for a concurrent positive control)
- Any discussion of global harmonization should expand on why the draft BRD did not place more reliance upon Organisation for Economic Co-operation and Development (OECD) Test Guideline (TG) 429 (OECD 2002) as a normative reference
- For two of the substances tested multiple times (HCA and linalool alcohol), different doses were used and discordant results were obtained. It should be noted for which (if either) of these tests, the highest dose tested was consistent with the dose selection principles set out in the draft BRD

1.1.2 General Statistical Comments

The Panel also had some statistical comments related to the LLNA limit dose procedure. First, the Panel viewed that a reference to the supplemental statistical information in which Dr. Joseph Haseman performed power calculations on the traditional LLNA would be useful in determining if the sample size used in the LLNA limit dose procedure was adequate for evaluating skin sensitization potential. Also, the Panel concluded that although a stimulation index (SI) based on the ratio of 3.0 as the decision criterion for a sensitizer is informative, statistical analysis determining if the test substance is significantly different from the control substance should be recommended.

1.1.3 Comments with Specific References to the Text

The Panel also identified the following minor formatting and grammatical errors, and information gaps in the draft BRD:

- The manner of notating numerical data in the draft BRD tables was not consistent (e.g., in some places the value “one” was shown as 1, elsewhere as 1.0; in a few places the SI was shown over the percent concentration used).
- Lines 291-294: The statement was made that “chemical class information is included to provide an indication of the variety of structural elements present in the substances that were evaluated in this analysis, but it is not intended to suggest an impact of structure on sensitization potential”. The latter concept is not entirely correct; the portion of the sentence stating “but it is not intended to suggest an impact of structure on sensitization potential” should be omitted. Consideration should be given to using the large database of chemicals to selectively modify structure-activity relationship (SAR) software for improved predictivity. This could likely be accomplished by communication with

software developers to point out availability of the newly expanded ICCVAM dataset developed for evaluation of the LLNA limit dose procedure.

- Lines 299-300: The sentence is incomplete; “non-sensitizers” should be inserted at the end of the sentence.
- Line 358: The citation to Sailstad et al. (2001) was not listed in Section 12.0 (References) and should be included.
- Lines 365-384 (Section 1.1.2): Consideration should be given to expanding the background on the mechanism and natural history of allergic contact dermatitis (ACD). Some additional detail regarding the biochemistry and cell biology of immune induction and elicitation would be useful as an orientation to how the LLNA functions as an integrated method of detection for ACD.
- Lines 366–368: The introductory sentence on the prevalence of ACD as an occupational health issue would benefit from amplification to also indicate that ACD is of variable severity with some potentially severe ramifications, and that ACD is recurrent upon rechallenge possibly requiring workplace accommodation or change of employment.
- Lines 366-368: There is no reference provided for the statistic from the U.S. Department of Labor Bureau of Labor Statistics cited in Section 1.1.2 of the draft BRD.
- Line 433: The reason for further evaluation of negative results with concentrations less than 10% should be clarified (Kimber et al. 2006).
- Lines 435-436: This bullet point re-plays conclusions made in the summary of the “rLLNA” issued by the scientific advisory committee of the European Centre for the Validation of Alternative Methods (ECVAM). However, in the draft BRD the statement has been altered and should instead read "..., as appropriate, per OECD TG 429 (OECD 2002)." rather than citing ICCVAM (1999) and Dean et al. (2001).
- Lines 452-453: The intent of the sentence would be clarified by modifying to read "...to identify potential human skin sensitizers through quantification of lymphocyte proliferation in the test method."
- Lines 496-500: Reading of the Kimber et al. (2006) citation does not indicate a recommendation for a concurrent positive control group. Thus, the sentence in the draft BRD that reports use of vehicle and positive control groups in the limit dose procedure based on the Kimber et al. (2006) paper is incorrect.
- Lines 509-510: This sentence should also mention that the LLNA limit dose procedure, as published by Kimber et al. (2006), is rationalized not only as a means of bioresource economy but as a valid means of streamlining the LLNA for regulatory screening assessment purposes under regulation such as Registration, Evaluation and Authorisation of Chemicals (REACH).
- Lines 520-523: A footnote might be needed to explain why OECD TG 429 (OECD 2002) is not referenced here.

- Lines 629-636: Data donated by GlaxoSmithKline (GSK) to the National Toxicology Program Interagency Center for the Validation of Alternative Toxicological Methods (NICEATM)-ICCVAM LLNA review were generated under GLP conditions at a clinical research organization (CRO) repeatedly audited for GLP compliance by GSK. This information should be added to the text.
- Section 6.1: The presentation of data and associated discussion regarding limitations in accepting a 10% dose concentration cut-off should be repositioned for emphasis. This information is important in developing suggestions for a standard LLNA limit dose procedure. Data from Appendix D could be reduced to a small table or figure and be integrated into the body of the final BRD.
- Lines 722-723: This appears to be an incomplete sentence.
- Lines 815-822: Data donated by GSK to the NICEATM-ICCVAM LLNA review were generated under GLP conditions at a CRO repeatedly audited for GLP compliance by GSK. This information should be added to the text.
- Table following line 1126: There is a typographical error in the 2 x 2 table. It appears that the cell for Negative (New Test) x Total (Reference Test) should read "c + d" rather than "a + d".

1.2 Comments on the Validation Status of the LLNA Limit Dose Procedure

1.2.1 Test Method Protocol

For the proposed LLNA limit dose procedure, ICCVAM recommended that the number of animals used in each group should be the same as that recommended by ICCVAM for the traditional LLNA based on its 1998 evaluation (i.e., at least five animals per group), and that individual animal data should be collected and reported (ICCVAM 1999; Dean et al. 2001). The Panel was asked whether they agreed that these are appropriate protocol requirements for the limit dose procedure. The Panel agreed that, based on the supplemental power calculations for the traditional LLNA performed by Dr. Joseph Haseman (see **Table 1-1**), a minimum of five animals per treatment group should be recommended for all future studies employing the limit dose procedure. In addition, the collection of individual animal data, as recommended by ICCVAM for the traditional LLNA (ICCVAM 1999; Dean et al. 2001), should also apply to all future studies following the LLNA limit dose procedure. Similarly, application of the LLNA limit dose procedure to a modified LLNA protocol would require adherence to a validated protocol with the exception of omitting the middle and low dose groups. Respective power calculations would indicate if application of the LLNA limit dose procedure to a validated modified LLNA protocol would allow fewer animals per dose group.

Table 1-1 Power Calculations for the Traditional LLNA

	3.0-fold increase	2.5-fold increase	2.0-fold increase	1.5-fold increase	1.3-fold increase
Mean Rx response	1034.4	862.0	689.6	517.2	448.24
Log (Mean Rx response)	6.942	6.759	6.536	6.248	6.105
Difference (log scale)	1.099	0.916	0.693	0.405	0.262
Difference/SD	2.40	2.00	1.51	0.88	0.57
Power for N=5	95%	80-90%	50-80%	<50%	<50%
Power for N=4	90%	80%	50%	<50%	<50%
Power for N=3	50-80%	50-80%	<50%	<50%	<50%
Other Power	–	–	95% (N=11)	95% (N=29)	95% (N=68)
Other Power	–	–	90% (N=9)	90% (N=23)	90% (N=54)

The power calculations above are based on a one-sided $p < 0.05$ Student's t test applied to log-transformed data from vehicle control LLNA tests.

Abbreviations: Fold-increase=Required increase above the vehicle control for a positive response (i.e., the stimulation index); N=number of animals; Rx=Treatment; SD=standard deviation.

The primary rationale for both provisions is to underpin robust statistical analysis of LLNA results. Furthermore, the use of individual animal data would allow for the evaluation of dosing errors or other anomalies that might be masked by the use of pooled animal data.

1.2.2 Substances Used for the Validation Studies

The Panel was asked whether they considered the traditional LLNA database representative of a sufficient range of chemical classes and physical chemical properties such that it would be applicable to any of the types of chemicals and products that are typically tested for skin sensitization potential. If not, the Panel was asked which relevant chemical classes/properties (other than those identified as limitations in the traditional LLNA) should be tested with caution, or not evaluated using the limit dose procedure, and which chemicals or products should be evaluated to fill this data gap. The Panel agreed that, in general, the traditional LLNA database included in the LLNA limit dose procedure evaluation was representative of a sufficient range of chemical classes and physical chemical properties and that it should be applicable to any of the types of chemicals and products that are typically tested in the traditional LLNA for skin sensitization potential. It was notable that the substances included in this evaluation provided a diverse chemical database. Since much is known about the mechanism of sensitization, the LLNA should theoretically identify any chemical that works by migration of haptens to the lymph node. However, the Panel noted that substance classes that are sometimes problematic in the LLNA (i.e., metals) would also likely be problematic in the LLNA limit dose procedure. There were also some substance classes that had limited or no representation in the draft BRD (i.e., mixtures/formulations, higher molecular weight biopharmaceuticals, and medical device materials). Thus, in general, the LLNA (and the LLNA limit dose procedure) is best used as part of a weight-of-evidence appraisal in which

attributes such as physical chemical parameters, SAR evaluation, and indications of other biological activity involving potential chemical-to-biological macromolecule interactions, are carefully considered along with LLNA results to evaluate dermal sensitizing potential.

Because the LLNA limit dose procedure was based on a retrospective evaluation of existing data, most of which was not generated using coded chemicals, the Panel was asked whether a lack of coding of test substances adversely impacted or biased the current evaluation. The Panel considered that although coding of chemicals should be recommended for prospective validation studies, this evaluation was likely not adversely impacted or biased because of a retrospective evaluation of existing data. This is supported by the fact that many of the chemical structures included in the analyses do not appear to contain known structural motifs associated with ACD or chemical hypersensitivity and therefore there was no a priori expectation that the chemical tested would be a sensitizer. The Panel viewed it important to consider the issue of coding or bias in prospective validation studies.

For some substances submitted using the traditional LLNA test method, it was not possible to confirm whether the data were generated using pooled animal data for each dose group (as allowed in OECD TG 429 [OECD 2002]). ICCVAM (1999), Dean et al. (2001), and EPA (2003) recommend the use of statistical analyses to help interpret LLNA study results, which necessitates data collected at the level of the individual animal. Furthermore, Cockshott et al. (2006) reported that using individual animal data allowed for technical problems or other outliers during an experiment to be identified. The Panel was asked what impact the inclusion of pooled animal data might have on the accuracy analysis of the LLNA limit dose procedure. The Panel concluded that, although it would be important to note whether individual or pooled animal data were reported, the retrospective analysis of the LLNA limit dose procedure versus the traditional LLNA should not be limited to studies with confirmed individual animal data. The Panel stated that internationally, both individual and pooled animal data have likely been used both for regulatory decisions and for in-house decisions relating to product development and risk management. Also, the fact that the retrospective data analysis presented in the draft LLNA limit dose procedure BRD did not distinguish between individual or pooled animal data suggested that both met the quality standards for inclusion in the draft BRD.

1.2.3 Test Method Accuracy

The Panel was asked whether the relevance (e.g., accuracy/concordance, sensitivity, specificity, false positive and false negative rates) of the LLNA limit dose procedure had been adequately evaluated and compared to the traditional LLNA (refer also to Table 6-1 of the draft ICCVAM BRD). The Panel concluded that the relevance of the LLNA limit dose procedure had been adequately evaluated and compared to the traditional LLNA. Comparisons resulting in an accuracy of 98.9% (461/466), a sensitivity of 98.4% (308/313) and a specificity of 100% (153/153) for the LLNA limit dose procedure when compared to the traditional LLNA were sufficient to consider it adequately validated for use in the evaluation of skin sensitization, mindful of its known limitations that are described elsewhere. Still, the Panel noted that it was important to keep in mind that a prospective analysis may not have the same accuracy as this retrospective analysis.

Furthermore, there were five substances for which the highest concentration tested produced an SI <3.0, while lower concentrations of these substances produced an SI >3.0 (see Table 6-

2 of the draft ICCVAM BRD). These substances were classified as false negatives compared to what was obtained in the traditional LLNA. The Panel was asked to identify any characteristics associated with these or other substances that might signal that this type of abnormal dose response might occur, and therefore, that using the LLNA limit dose procedure would not be appropriate. The Panel could not identify any common characteristics associated with the five false negative substances that would explain the non-linear dose response obtained. It was not known if any procedural problems were reported with these studies or what values were returned by the negative/positive control groups (in relation to other historical positive control values).

Thus, the Panel suggested that it might be worthwhile to examine whether LLNA results with these five false negative substances should be repeated. If the difference turned out to be repeatable, there could be hypothetical reasons to explain why the higher doses did not pass the SI threshold of 3.0. For example, under certain experimental conditions, the target lymphocytes may be selectively induced to a highly sensitive state by some chemicals at higher doses and may undergo either induction of apoptosis or inhibition of cell proliferation. Still, there was no evidence that these substances were immunomodulators that might have differentially stimulated or depressed the immune response depending on the dose and exposure. In any case, understanding false negatives is encouraged to ensure adequate protection of public health.

The Panel was asked whether the draft BRD adequately characterized the usefulness and limitations of the LLNA limit dose procedure based on the accuracy analyses. Overall, the Panel agreed that the draft BRD adequately characterized the usefulness and limitations of the LLNA limit dose procedure based on the accuracy analyses. Since the LLNA limit dose procedure and the traditional LLNA have close concordance, there was no need for detailed discussion in the draft BRD. However, it was not explicitly stated in the draft BRD that compared to a fully conducted traditional LLNA, a false positive result in the LLNA limit dose procedure is not possible (i.e., if the single dose used in the proposed limit dose procedure gives an $SI \geq 3.0$, then so would the top dose in the traditional LLNA). Furthermore, prospective testing with the LLNA limit dose procedure to predict the sensitization potential of an unknown chemical was not discussed.

1.2.4 Test Method Reliability

The Panel was asked if it was appropriate to assume that the intra- and inter-laboratory reproducibility of the LLNA limit dose procedure and the traditional LLNA would be similar, based on the fact that they use identical protocols with the exception of the number of doses used (i.e., would reducing the number of test substance dose groups from three to one reduce the reliability of the assay?). The Panel agreed that it was appropriate to assume that the intra- and inter-laboratory reproducibility of the LLNA limit dose procedure and the traditional LLNA would be similar, because reproducibility is more dependent on the method than on the number of dose groups. However, reducing the number of test substances dose groups from three to one could reduce the sensitivity of the assay (i.e., the ability to correctly identify sensitizers). The traditional LLNA may have a greater chance of correctly identifying a sensitizer even in the presence of one or more technical errors since there are data from three dose groups for consideration and an $SI \geq 3.0$ at any dose group would result in the substance being classified as a sensitizer. However, for the purpose of adopting an assay that uses fewer animals and provides increased throughput for screening purposes,

these hypothetical considerations are not a sufficient reason to argue against use of the limit dose LLNA procedure.

1.2.5 Data Quality

For some studies included in the draft BRD, it was not possible to determine whether or not they had been conducted in accordance with GLP guidelines. Furthermore, original records for some of the non-GLP studies included in this evaluation could not be obtained. As a result, an independent audit could not be conducted to confirm that the reported data was the same as the data recorded in laboratory notebooks. Neither was it possible to obtain the results of GLP audits for all studies conducted in accordance with GLP guidelines. The Panel was asked whether the results of such studies (all of which are currently included) should be excluded from the performance analyses. The Panel concluded that it was important to note if the data were obtained from studies conducted according to GLP guidelines, as ideally this should be the case. However, the Panel concluded that the data resulting from the retrospective studies that could not be confirmed as GLP-compliant should not be excluded from the performance analysis. Since there was not an indication that the reliability of the data presented for consideration may have been compromised, omitting any data would likely lessen the impact of the analysis. Furthermore, data obtained from peer-reviewed literature or final reports were likely of sufficient quality.

1.2.6 Consideration of All Available Data and Relevant Information

The Panel was asked if all the relevant data identified in published or unpublished studies conducted using the traditional LLNA had been adequately considered in the draft BRD. If not, the Panel was asked what other traditional LLNA data needed to be considered and how such data could be obtained. The Panel considered that the draft BRD had taken into account a large majority of the relevant data identified in published and unpublished traditional LLNA studies. The data received as a result of the *Federal Register (FR)* notices and the key literature citations seemed to be inclusive of the relevant data for this analysis. Although additional data that could have been included might exist, it was deemed unlikely that the current outcome (which is based on 466 substances) would be altered given the very small change in accuracy statistics relative to Kimber et al. (2006), which was based on 211 substances.

1.3 Comments on the Draft ICCVAM Test Method Recommendations on the LLNA Limit Dose Procedure

1.3.1 Test Method Usefulness and Limitations

The Panel was asked to comment on whether the available data supported the ICCVAM draft recommendations for the LLNA limit dose procedure in terms of the proposed test method usefulness and limitations (i.e., that the LLNA limit dose procedure should be routinely recommended for hazard identification when dose response information is not required). The Panel considered that, based on the available information, the draft recommendations appeared valid, but made the following suggestions:

- Further emphasis should be given to using the LLNA limit dose procedure as a part of a comprehensive weight-of-evidence evaluation of dermal sensitizing potential (e.g., including physical chemical evaluation, SAR information,

including likelihood of dermal penetration, ability of materials to adduct biomacromolecules).

- Such information in addition to LLNA results might also be useful in confirming or questioning LLNA outcomes terms of in human hazard identification, since it should be emphasized that a major application of the method is to prospectively detect harmful chemicals.
- Solubility or thermodynamic activity data, beyond visual assessment (e.g., use of chemically-specific methods to document solubility), should be used to confirm the appropriateness of the maximum dose tested.
- Vehicle selection for the LLNA can affect the results and may not allow accurate comparisons between chemicals applied in different vehicles. In choosing the best vehicle, consider measured solubility information for the potential vehicle. Then, it would be important to take into account how the vehicle affects the amount of the chemical that can be applied to the ear. More importantly, the impact that vehicle selection has on the amount of applied chemical that actually gets into the mouse to induce the sensitization response should be evaluated. Some of the recommended LLNA vehicles (e.g., 4:1 acetone:olive oil (AOO), dimethylsulfoxide (DMSO), methyl ethyl ketone) could be expected to disrupt the barrier properties of the skin. Additionally, although propylene glycol might allow an increased amount of chemical to be applied, it might also inhibit the penetration of a chemical by enhancing partitioning in the vehicle relative to the skin.

The Panel was asked whether the LLNA limit dose procedure should be routinely recommended for the hazard identification of skin sensitizing chemicals when dose response information *is not* required. With the points noted above in mind, the Panel agreed that it should be routinely recommended since the LLNA limit dose procedure offers time, cost, throughput, and logistical benefits as well as using fewer animals. Still, the investigator should keep in mind what is known of the chemical regarding general toxicity and note scenarios where abnormal dose-response relationships in the traditional LLNA might result in false negatives in the limit dose procedure (see Table 6-2 of the ICCVAM BRD).

The Panel was then asked whether the LLNA limit dose procedure should be routinely recommended as the initial test to identify sensitizers before conducting the traditional LLNA, as a way to further reduce animal use, even if dose response information *is* required, since negative results would not require further testing. The Panel agreed that use of the LLNA limit dose procedure, as the initial testing procedure to identify sensitizers and non-sensitizers before conducting the traditional LLNA, is justifiable even when dose response information is required. This is applicable in the occupational and public health setting where obtaining hazard information is of critical importance. There is a benefit since dose-response information generated in subsequent testing in the traditional LLNA for substances that were positive in the limit dose procedure then gives further assurance of detecting hazardous substances and allowing a potency estimate. The benefits of screening out the negatives (which do not require dose response information) is clear; however the animal welfare gains will depend on the proportion of test substances in any class that turn out to be non-

sensitizers and there might be possible consequences of the delays resulting from a further round of testing for those materials that are identified as sensitizers.

Based on the existing database, there is a false negative rate of 1.6% (5/313 positive compounds) for the LLNA limit dose approach compared to the results obtained in the traditional LLNA. The Panel was asked whether they considered that this is adequately addressed by the proposed cautionary language and weight-of-evidence consideration for negative substances. The Panel agreed that the small rate of false negatives was adequately addressed in the draft test method recommendations by giving cautionary and weight-of-evidence consideration to the negative substances (and any possible false positive results). Furthermore, given that the dose responses for these five materials were rather unusual, it was not known whether these studies were repeatable, whether any procedural problems were reported with these studies, or what values were returned by the negative/positive control groups (in relation to other historical positive control values). In general, the Panel viewed that the false negative rate of 1.6% would likely be unimportant when the larger differences between the animal model and humans are considered.

1.3.2 Test Method Protocol

The Panel was asked whether they agreed that the available data support the ICCVAM draft recommendations for the LLNA limit dose procedure in terms of the proposed standardized test method protocol. The Panel agreed and recommended adherence to the ICCVAM (1999) LLNA protocol for future studies of the LLNA limit dose procedure with the exception of omitting the middle and low dose groups. Similarly, application of the LLNA limit dose procedure to a modified LLNA protocol would require adherence to the modified LLNA protocol with the exception of omitting the middle and low dose groups. Adhering to the use of individual animals for future studies was specifically stressed because it would allow for an estimate of inter-animal variability.

The recommended ICCVAM protocol (ICCVAM 1999; Dean et al. 2001; EPA 2003), as well as OECD TG 429 (OECD 2002), specifies that the highest dose tested should be the highest soluble concentration that does not induce systemic toxicity and/or excessive skin irritation. However, Kimber et al. (2006) concluded that negative results obtained from studies where the highest concentration tested was below 10% should be considered invalid, and adopted a 10% application concentration as a threshold of confidence for categorization of a chemical as being negative while noting that the figure should not be considered as inviolable. The Panel was asked whether the data presented in the draft BRD (i.e., 51/313 positive substances in the NICEATM database were negative at concentrations equal to or above 10%, but were positive at even higher concentrations) were adequate to conclude that this threshold concentration is not appropriate. The Panel viewed that this point should be clarified. ICCVAM recommended that no threshold should be used to determine the validity of conduct of the LLNA limit dose procedure. Instead, formal attempts to maximize dose delivery including documentation of solubility of the test substance in the vehicle used should be undertaken.

The Panel was asked whether additional testing should be required if a negative result was obtained for a test substance in a study where the highest concentration that could be tested (based on systemic toxicity or excessive local irritation, as described in ICCVAM [1999], Dean et al. [2001], and EPA [2003]) was <10%. The Panel considered that, if a negative

result was obtained for a test substance under these conditions, additional testing should not be required, because at that point it would likely be a toxic effect and not sensitization. In contrast, the imperative should be to minimize the number of false negatives. For this purpose, rigorous examination of maximum solubility or other parameters to ensure testing at maximum concentration should be employed. In addition, weight-of-evidence considerations such as SAR and physicochemical characteristics should be documented. More animal testing to verify negative results should only be undertaken if the weight-of-evidence suggests that it would be appropriate.

The Panel was asked if the current approach for selecting the “limit dose” was appropriate or whether there is a threshold concentration for the LLNA at which a negative result could always be considered as an acceptable result. The Panel agreed that the current recommendation to select a maximum applied dose in the LLNA limit dose procedure is appropriate. However, the data presented in the draft BRD implied that at present it is not possible to establish a uniform concentration threshold for the “limit dose”. Thus, it seemed justifiable that preliminary experimentation (as would be typically performed during a dose range finding study) should be conducted on vehicle selection, test substance solubility, and stability in the vehicle.

1.3.3 Future Studies

The Panel was asked if they agreed that the available data support the ICCVAM draft recommendations for the LLNA limit dose procedure in terms of the proposed future studies. Although limited in scope, the Panel considered that the available data supported the ICCVAM draft recommendations for additional studies. The Panel agreed that attempts be made to investigate if maximum solubility was achieved (e.g., use of chemically-specific methods to document solubility). For hazard assessment, it was considered troublesome that there were so many vehicle choices, because the vehicle could have a significant effect on whether (and how much) a test substance penetrated the skin barrier. Observed vehicle effects may relate to dermal penetration as well as to immunomodulation. The Panel considered it desirable to follow the hierarchy of vehicles recommended in the ICCVAM (ICCVAM 1999; Dean et al. 2001) protocol. In addition, dedicated attempts must be made to investigate solubility in AOO mixtures before using other vehicles. Regardless of the vehicle used, it is important to ensure that a vehicle does not promote lymph node cell proliferation. The Panel also suggested that it might be informative to test both known mild and severe sensitizers concurrently in all recommended vehicles to evaluate if a specific vehicle choice(s) might influence the results.

Although the false negative rate in the current analysis was small, a need exists to better understand factors that could lead to false negative results with future use of the LLNA limit dose procedure. Thus, consideration should also be given to formal statistical assessments to verify group size and use of individual animal data in routine performance of the LLNA limit dose procedure. Criteria should be established to verify proficiency with the LLNA limit dose procedure. Such criteria could be used to answer questions about the necessity to perform concurrent positive controls.

1.3.4 Comments with Specific References to the Text

The Panel also identified the following comments and/or corrections to the draft ICCVAM test method recommendations document on the LLNA limit dose procedure that should be considered by ICCVAM:

- Lines 26-28: Conclusions given here regarding the relative potency ratings of the five materials classified as false negative in the analysis in Section 6.2 of the draft BRD were newly introduced. This assessment should also have been considered for inclusion in Section 6.2 of the draft BRD.
- Line 28: The citation of Gerberick et al. (2004) was not accompanied by a reference.
- Lines 62-70: The listing of substances not amenable to test in the LLNA could have been expanded to also include agents with anticipated pharmacodynamic action as immune suppressants.
- Line 69: The citation of Gaspari et al. (2007) was not accompanied by a reference.
- Lines 75-79: Dependent upon other considerations, this portion of the text could have been modified to (1) clarify recommendations regarding routine use of concurrent positive control (i.e., possible exception for laboratories conducting a high volume of LLNA work in which periodic positive control for quality control purposes might suffice), and (2) expand on the details regarding indications of excessive irritation and/or systemic toxicity to aid in choice of maximal test dose.

2.0 LLNA for Testing Aqueous Solutions, Metals, and Mixtures

2.1 Comments on the Draft Addendum for Completeness, Errors, and Omissions

In regard to the draft Addendum to the traditional LLNA BRD, the Panel was asked to comment on any errors that should be corrected or omissions of relevant data/information that should have been included. The Panel concluded that there were no apparent errors or omissions to the draft Addendum.

2.2 Comments on the Validation Status of the Traditional LLNA for Testing Aqueous Solutions, Metals, and Mixtures

2.2.1 Substances Used for the Validation Studies

The Panel was asked whether the database of substances evaluated was representative of a sufficient range of mixtures, metal compounds, and substances in aqueous solutions that are typically tested for skin sensitization potential. While there were limited data available on the effects of mixtures, metals, and aqueous solutions on skin sensitization potential, the Panel considered the database to be generally representative. The Panel indicated that there did not seem to be obvious classes of chemicals missing from the data set used to evaluate the utility of the traditional LLNA for testing aqueous solutions. However, quantitative compositions for the mixtures included in the analysis had not been provided. Thus, it was difficult to determine if those mixtures were representative of the types of mixtures typically tested in the traditional LLNA. With respect to metals, there was a paucity of commercially useful metals such as platinum, palladium, iron, zinc, manganese, and silver compounds. To enlarge the group of metal non-sensitizers, substances used as cosmetic ingredients (e.g., titanium dioxide) and aluminum compounds currently used in antiperspirants might be considered. However, the Panel considered that the inclusion of an array of other metals and at least one zinc and manganese salt likely weighted the data set appropriately and it appeared sufficiently broad to support conclusions about the utility of the traditional LLNA for testing the skin sensitization potential of metals.

Substances or mixtures that were tested in an aqueous or an organic:aqueous vehicle were labeled as aqueous solutions. For the purpose of this evaluation, a substance or mixture containing at least 20% water was defined as an aqueous solution. The Panel was asked whether this criterion was appropriate for defining an aqueous solution. The Panel was uncertain about the appropriateness of this definition of an aqueous solution, but did not offer an alternative definition. However, the Panel indicated that an organic:aqueous solution that is not miscible would likely produce varying results because of partitioning of the chemical into either phase.

The Panel was asked whether the lack of coding of test substances might adversely impact or bias the current evaluation. While coding of chemicals is recommended for prospective validation studies, the retrospective evaluations in the draft Addendum were based on existing data, most of which were not generated using coded chemicals. However, the Panel agreed that the lack of chemical coding was not likely to bias the evaluation since this study was retrospective. This is supported by the fact that many of the chemical structures included in the analyses did not contain known structural motifs associated with allergic contact dermatitis/chemical hypersensitivity and therefore there was no a priori expectation that the

chemical tested would be a sensitizer. Furthermore, many of the substances tested were apparently evaluated for hazard assessment purposes rather than to test the predictive ability of the traditional LLNA. Thus, there does not appear to be any bias in chemical selection for the expanded dataset considered in the study of applicability domain for the traditional LLNA.

For some substances submitted using the traditional LLNA test method, it was not possible to confirm whether the data were generated based on pooled lymph nodes among animals within a dose group, as allowed in OECD TG 429 (OECD 2002), or individual animal responses, as recommended by ICCVAM (1999) and required by EPA (2003). ICCVAM (1999) and EPA (2003) both recommend the use of statistical analyses to aid in the interpretation of traditional LLNA study results; such analyses necessitate data collected from individual animals. Additionally, Cockshott et al. (2006) reported that using individual animal data allowed for outlier animal results within a dose group to be identified. The Panel was asked whether the analysis of the performance of the traditional LLNA for testing mixtures, metal compounds, and substances in aqueous solutions should be limited to data from studies that collected individual animal data, and then to comment on the potential impact on the accuracy analysis of including results from studies in which pooled animal data were collected. The Panel concluded that, although individual animal data were preferred, pooled animal data should not be excluded automatically from this retrospective analysis.

2.2.2 Test Method Accuracy

The Panel was asked whether the relevance (e.g., accuracy/concordance, sensitivity, specificity, false positive and false negative rates) of the traditional LLNA for testing mixtures, metal compounds, and substances in aqueous solutions had been adequately evaluated and compared to the human and guinea pig test results. The Panel agreed that the comparative assessment of the relevance of the traditional LLNA for testing mixtures, metal compounds, and substances in aqueous solutions appeared to be as comprehensive as was feasible. However, because of the limited number of comparisons available, the accuracy statistics probably do not give a complete picture of the usefulness and limitations of the traditional LLNA for identifying skin sensitizers among these types of substances or when using an aqueous vehicle.

When multiple traditional LLNA studies were available for the same substance, the “majority call” (among studies using the same vehicle and generally tested over the same concentration range) was used by ICCVAM to assign an overall classification for the purposes of the accuracy analysis. For example, if chemical X was tested five times and was positive in three studies and negative in two, the overall classification was positive. The Panel was asked whether they agreed with this approach. They expressed their concern about the approach in the following way; if all nickel-containing compounds in the analysis were viewed as a group, there were four positive calls and four negative calls (see Appendix C2 of the draft Addendum). Using the “majority call” approach, the overall call would be determined by the next available study, which may not provide the correct call. More data would be needed to confirm whether the classification was appropriate. For this dataset, most of the “negative calls” had SI values that approached 3.0. Thus, a more suitable method might be to base the overall call on the SI data, while giving greater positive call consideration/weight to SI values just below 3.0. It may also be useful to perform a meta-analysis. It is important for the

Addendum to mention the potential impact of using the “majority call” decision, rather than relying on a weight-of-evidence approach, on the accuracy analyses.

2.2.3 Data Quality

For some studies included in the draft Addendum, it was not possible to determine whether or not they had been conducted in accordance with GLP guidelines. Original records for some of the non-GLP studies included in this evaluation could not be obtained. As a result, an independent audit could not be conducted to confirm that the reported data was the same as the data recorded in laboratory notebooks. Neither was it possible to obtain the results of GLP audits for all studies conducted in accordance with GLP guidelines. The Panel was asked to discuss what impact this lack of information might have on the evaluation of the traditional LLNA for testing mixtures, metal compounds, and substances in aqueous solutions and whether such studies should be excluded from an analysis of test method accuracy. The Panel considered it important to note if the data were obtained from studies conducted according to international GLP guidelines, since ideally this should be the process followed. However, the Panel viewed that data from studies that could not be confirmed as being GLP-compliant were still appropriate to include in the accuracy analysis, provided that the data were from the peer-reviewed literature or from sources with high quality laboratory management practices. Much of the value for this draft Addendum was the potential to supplement the data available at the time of the ICCVAM (1999) analysis. Additional information on test substance identification would clearly be useful in the continued evaluation of the applicability domain of the traditional LLNA, but omitting data on mixtures, metals, or use of aqueous solutions based solely on the lack of GLP compliance would lessen the impact of the current retrospective analysis and did not seem warranted. However, if the original data were not available, it would be appropriate to note this in the final version of the Addendum.

2.2.4 Consideration of All Available Data and Relevant Information

The Panel was asked whether the draft Addendum included all of the relevant data for studies conducted using the traditional LLNA for testing mixtures, metal compounds, and substances in aqueous solutions. The Panel considered that, although it was possible that there might be a few studies in the literature to augment the analysis, it seemed that the relevant data had been identified and the response to the *FR* notice and the literature citations examined had included the most relevant studies.

2.3 Comments on the Draft ICCVAM Test Method Recommendations on the Traditional LLNA for Testing Aqueous Solutions, Metals, and Mixtures

2.3.1 Test Method Usefulness and Limitations

ICCVAM stated that more data would be needed before a recommendation on the usefulness and limitations of the traditional LLNA for testing mixtures could be made, due to the limitations associated with the available mixtures database (i.e., unknown formulae, lack of human data). The Panel was asked whether they agreed that the available data supported the ICCVAM draft recommendations for the traditional LLNA with regard to testing mixtures in terms of the proposed test method usefulness and limitations. The Panel agreed that ICCVAM’s draft recommendation with respect to the traditional LLNA testing of mixtures appeared valid based on the limitations inherent in the available data set. Still, the Panel

urged that the ICCVAM recommendation indicate that the approach may be viable. The Panel further recommended that the test method recommendations summary should indicate that the limitations include relatively poor concordance of traditional LLNA outcomes for mixtures with to those obtained in guinea pig tests. Routine comparisons of accuracy according to classification criteria may not be sufficient to evaluate the concordance for mixtures, and furthermore, the guinea pig tests are not necessarily valid for mixtures. The Panel also indicated that the term “mixtures” was used too broadly (i.e., can represent an infinite number of materials) and it would be more beneficial to specify types or formulations of mixtures that are being examined.

ICCVAM recommended that, based on the available data for metals, the traditional LLNA was useful for the testing of metal compounds, with the exception of nickel. The Panel was asked whether they agreed that the available data supported the ICCVAM draft recommendations for the traditional LLNA with regard to testing metals in terms of the proposed test method usefulness and limitations. Based on the available information, the Panel agreed that the draft recommendations with regard to testing metals appeared to be valid. In particular, the evidence for most metals (e.g., accuracy of 86% (12/14), sensitivity of 100% (9/9), specificity of 60% (3/5), 0% (0/9) false negatives) when comparing traditional LLNA results to those obtained from evaluations in humans supported the use of the traditional LLNA as a hazard identification tool for metals, excluding nickel. However, the Panel recommended that it would be worthwhile to study further the variable results obtained for nickel since there is a wealth of literature on allergic contact dermatitis of nickel in humans.

In a minority opinion, Dr. Dagmar Jírová stated that it should not be concluded that the traditional LLNA was not suitable for testing nickel compounds, because the different vehicles used may have had a significant impact on the ability of nickel to penetrate the skin and be bioavailable. She noted that nickel chloride and nickel sulfate were both positive in aqueous solutions, and negative only when non-aqueous vehicles were used. In human exposures, nickel compounds were applied in aqueous solutions. Thus, this may serve as sufficient justification to use aqueous vehicles when nickel, and perhaps also other substances, are tested and evaluated in the traditional LLNA. When DMSO was used as the vehicle, the SI value increased with increasing nickel concentration. Unfortunately, no data were available for concentrations over 5% for either nickel compound in DMSO. Nickel chloride as 10% in aqueous solution reached an SI of 6.6. Inconsistent test results due to the vehicle have also occurred in other *in vitro* studies (e.g., phototoxicity). Thus, Dr. Jírová concluded that the traditional LLNA could be used even for testing nickel compounds when other vehicles (in particular aqueous) are used.

Due to the limited number of substances tested in aqueous solutions, ICCVAM recommended that more data would be needed before a recommendation on the usefulness and limitations of the traditional LLNA for testing substances in aqueous solutions could be made. The Panel was asked whether they agreed that the available data supported this ICCVAM draft recommendation for the traditional LLNA with regard to the testing of substances in aqueous solutions. The Panel agreed that the draft ICCVAM recommendation was appropriate and that more data were required before an adequate evaluation of the use of the traditional LLNA with aqueous solutions could be conducted.

2.3.2 Test Method Protocol

The Panel was asked whether they agreed that the available data supported the ICCVAM draft recommendations for the traditional LLNA in terms of the proposed test method standardized protocol. The Panel agreed that, in general, the results of the assessment in the draft Addendum supported the proposals for standardized conduct of the traditional LLNA. However, this conclusion depended on a side-by-side reading of the draft Addendum and the ICCVAM (1999) protocol. The Panel suggested expanding the brief section of the draft test method recommendations dealing with test method protocol for the traditional LLNA (Section 2.0) to specifically point out how the conclusions of the applicability domain evaluation may affect the standard traditional LLNA protocol. For example, the evaluation of aqueous solutions apparently resulted in the methodological recommendation that aqueous test solutions be avoided and the further recommendation of a hierarchy of organic solvents to be considered as dosing vehicles. The emphasis might be on using a vehicle to which humans may actually be exposed.

2.3.3 Future Studies

The Panel was asked whether they agreed that the available data support the ICCVAM draft recommendations for the traditional LLNA in terms of the proposed future studies. The Panel agreed that the ICCVAM recommendation for continued accrual of information from traditional LLNA evaluations of mixtures, metals, and aqueous solutions with comparative data for guinea pig and human tests was appropriate. The traditional LLNA accuracy for metals of 86% and sensitivity of 100% (0% false negative) was excellent; a specificity of 60% (40% false positive) was considered acceptable as over-classification maintains safe human use. The Panel encouraged the use of the traditional LLNA to acquire further information on mixtures, metals, and aqueous solutions. However, the Panel suggested that, given resource limitations, it would be important to prioritize the recommendations in order to focus on what is most important.

[This Page Intentionally Left Blank]

3.0 Non-Radioactive LLNA Protocol - The LLNA: Daicel Adenosine Triphosphate (LLNA: DA) Test Method

3.1 Comments on the Draft BRD for Completeness, Errors, and Omissions

3.1.1 General Comments

The Panel was asked if there were any errors in the draft LLNA: DA BRD that should be corrected, if omissions of existing relevant data had been identified, or if there was additional information that should be included. As a general comment, the Panel noted that the draft BRD clearly and succinctly provided an overview of the LLNA: DA test method and the relevant validation study data. The draft BRD indicated that the LLNA: DA differs from the traditional LLNA in the method of measuring proliferation (measures levels of adenosine triphosphate [ATP] instead of radioactivity), substance treatment (pretreating the test site with 1% sodium lauryl sulfate [SLS] prior to test substance application and an additional treatment on day 7), and sampling time (draining auricular lymph nodes are collected on day 8 rather than on day 6). Because the traditional LLNA evaluates the induction phase only, the relevance of results with the LLNA: DA (and any other LLNA protocol) should always be considered in the context of human experimental sensitization data, human epidemiologic data, and elicitation in the clinical setting.

3.1.2 Comments with Specific References to the Text

The Panel noted the following text that should be clarified or corrected in the final version of the LLNA: DA BRD:

- Line 428: The text should read “1% SLS”, not “1% SDS”; the same terminology should be used throughout rather than going back and forth between SLS and sodium dodecyl sulfate (SDS).
- Line 449: The text and formula in lines 448-450 appear misplaced and instead seem to belong in Section 7.0. Additionally, X and Y should be defined and the “Var (ln SI)” formula should be clarified.
- Table 3-1: There were 33 substances in the table, yet the discussion of the table in the text referred to 31 substances. Although the reason for this apparent discrepancy becomes more evident later on, this should be discussed up front.
- Table 3-2 and 3-3: The interlaboratory distribution and testing of the sensitizers versus non-sensitizers should be indicated here.
- Table 6-1: Although the table clearly provided a comparison of the different methods, it would be useful if the footer for this table also indicated the basis for the differences in substances included in each analysis (i.e., n=25, 26, or 29) as stated in the text.
- Table 7-1 was only moderately helpful because the standard deviations (SD) were not calculated on a log scale. Given the skewness in the data, the ranges given were misleading indicators of increases due to the vehicles. Without the samples sizes, an analysis of variance was impossible to calculate, and that

would have been the appropriate measure of differences between the experiments. In order to make recommendations regarding needed reproducibility experiments, it would have been helpful to have a power analysis for this situation. That is, for each vehicle with each chemical, using two, three, four and five animals per dose group per experiment, how many experiments need to be run to detect significant differences between the experiments?

3.2 Comments on the Validation Status of the LLNA: DA

3.2.1 Test Method Protocol

Based on its 1998 evaluation of the traditional LLNA procedure, ICCVAM recommended that at least five animals be used in each test group (ICCVAM 1999). The LLNA: DA validation studies presented in the draft BRD were performed using four animals per dose group. Thus, the Panel was asked to comment on the potential impact of using fewer than five mice per dose group. The Panel noted that supplemental statistical information they were provided with indicated that the power for detecting a three-fold increase in the SI value in the treatment group for the LLNA: DA dataset evaluated in the draft BRD was estimated to be 95% for a sample size of three mice per dose group (see **Table 3-1**). Since an increase of false negatives may not be an issue, the potential opportunity exists for utilizing this smaller group size. The Panel cautioned, however, that using less than five animals per group might result in a less precise estimate of the mean response, which, in turn, will impact accuracy. Also, if technical errors further reduce the sample size, accuracy is further reduced. Thus, the Panel recommended that all initial validation studies adhere to the ICCVAM-recommended protocol (ICCVAM 1999; Dean et al. 2001) of five animals per dose group until sufficient information is generated to indicate that the use of fewer animals per dose group is statistically valid.

Table 3-1 Power Calculations for the LLNA: DA

Parameter	3.0-fold Increase	2.5-fold Increase	2.0-fold Increase	1.5-fold Increase	1.3-fold Increase
Mean Rx response	8835	7362.5	5890	4417.5	3828.5
Log (mean Rx response)	9.086	8.904	8.681	8.393	8.250
Difference from control (log scale)	1.098	0.916	0.693	0.405	0.262
Difference/SD	3.95	3.29	2.49	1.46	0.94
Power for N=5	99%	99%	95%	50-80%	<50%
Power for N=4	99%	95-99%	90%	50%	<50%
Power for N=3	95%	90-95%	80%	<50%	<50%
Other power	–	–	–	95% (N=11)	95% (N=25)
Other power	–	–	–	90% (N=9)	90% (N=20)

The power calculations above are based on a one-sided $p < 0.05$ Student's t-test applied to log-transformed data from vehicle control LLNA: DA tests.

Abbreviations: Fold-increase=Required increase above the vehicle control for a positive response; Rx=Treatment; N=Number; SD=Standard deviation.

The data generated for the substances analyzed in the LLNA: DA interlaboratory validation studies came from auricular lymph nodes that were pooled across animals in each treatment group. The Panel was asked to comment on the potential impact of including pooled animal data on the accuracy analysis of the LLNA: DA. The Panel noted that a statistical analysis of differences between treatment-related and vehicle control ATP levels could not be determined without measures of variability. Individual animal data highlights technical issues and allows for consideration of dose-response information and statistical analyses.

The LLNA: DA differs from the traditional LLNA in the treatment schedule and by including a pretreatment step with 1% SLS just prior to application of the test substance. The Panel was asked to comment on the appropriateness of these protocol differences. The Panel did not consider these differences to be significant, as long as it could be demonstrated that the 1% SLS pretreatment step and the additional test substance treatment on day 7 did not induce a skin reaction indicative of the elicitation phase of skin sensitization. Although it was being used at a lower concentration than the estimated concentration needed to produce a stimulation index of 3 (EC3), the Panel expressed concern about pretreating the mouse ear with 1% SLS since SLS is an irritant and positive in the traditional LLNA. Consequently, the inherent sensitivity of the LLNA may be modified by the 1% SLS pretreatment step. To demonstrate that these concerns are not justified, the Panel concluded that weak irritants and weak sensitizers needed to be tested in the LLNA: DA assay with and without pretreatment with 1% SLS. The test method developer might also consider using decision criteria other than $SI \geq 3.0$ such that 1% SLS pretreatment is no longer necessary.

3.2.2 Substances Used for the Validation Studies

The Panel was asked if they considered the substances tested in the LLNA: DA to be representative of a sufficient range of chemical classes and physicochemical properties that the test method would be applicable to any of the types of chemicals and products that are typically tested for skin sensitization potential. The Panel considered the database of test substances tested in the LLNA: DA representative of a sufficient range of chemicals. The selected substances included solids and liquids and a range of solvents/vehicles. The database also represented a range of sensitizing potency, a variety of different chemical classes and substances with differing requirements for metabolic activation. However, it might have been useful to have also included substances with clearly different protein reaction mechanisms (protein binding), as well as dyes, natural extracts, and mixtures.

3.2.3 Test Method Accuracy

The accuracy analysis in the draft LLNA: DA BRD was based on overall concordance with the traditional LLNA. Accuracy statistics compared to the guinea pig tests and human data/experience were also provided. The Panel, when asked if they considered these comparisons appropriate for assessing the accuracy of the LLNA: DA, agreed that the comparisons to the traditional LLNA performance and also to the guinea pig and human sensitization data were important. The Panel also stressed that, because the traditional LLNA only evaluates the induction phase, the relevance of the LLNA: DA results should always be considered in the context of human experimental sensitization data, human epidemiologic data, and elicitation in the clinical setting.

The Panel was asked if they considered the evaluation of the relevance of the LLNA: DA and the comparison to the traditional LLNA to be adequate. The Panel noted that Table 6-1 of the draft LLNA: DA BRD clearly provided a comparison of the different reference methods (i.e., traditional LLNA, human tests, and guinea pig tests). Thus, the Panel concluded that the relevance of the LLNA: DA had been adequately evaluated. However, including data on more substances is likely to further strengthen confidence in the concordance data.

One substance, 2-mercaptobenzothiazole, produced a false negative response compared to the traditional LLNA when tested using the LLNA: DA. The Panel was asked if they could identify any characteristics associated with this or similar substances, compared to the correctly identified sensitizers, that might signal that this type of discordant response would occur, and therefore using the LLNA: DA to test such substances would not be appropriate (or that negative results for substances with such properties may warrant additional testing). The Panel could not identify specific characteristics that might explain the false negative response using the LLNA: DA. Although understanding the solubility and stability of the test substance in different vehicles is important, the differences in response did not seem to be explained by the vehicle differences (AOO and dimethylformamide [DMF]) between the two tests. In addition, the impact of 1% SLS pretreatment on the negative response in the LLNA: DA is not known but should be considered.

One substance, benzalkonium chloride, produced a false positive response compared to the traditional LLNA and guinea pig test when tested using the LLNA: DA. The Panel was asked if they could identify any characteristics associated with this or similar substances, compared

to the correctly identified non-sensitizers, that might signal that this type of discordant response would occur, and therefore using the LLNA: DA to test such substances would not be appropriate (or that positive results for substances with such properties may warrant additional testing). The Panel could not identify specific characteristics that might explain the false positive response for this substance in the LLNA: DA. The Panel viewed that it was important to note, however, that this chemical is a well-known skin irritant, and on occasion it had also been considered a human sensitizer, typically on the basis of positive diagnostic patch test data. Thus, the Panel reiterated that the relevance of LLNA: DA results should always be considered in the context of human experimental sensitization data, human epidemiologic data, and elicitation in the clinical setting. The actual impact of the 1% SLS pretreatment step on the LLNA: DA has not been well established, although van Och et al. (2000) and De Jong et al. (2002) have reported that 1% SLS pretreatment enhances the response in the traditional LLNA.

3.2.4 Test Method Reliability

The Panel was asked if they considered the intralaboratory reproducibility of the LLNA: DA to have been adequately evaluated and compared to the traditional LLNA (refer to Table 7-1 of the draft LLNA: DA BRD). The Panel noted that only eugenol and isoeugenol, two sensitizers with similar chemical structures, were tested. The Panel recommended testing a positive control commonly used in the traditional LLNA (e.g., HCA) for a more complete evaluation of intralaboratory reproducibility. In addition, it was unclear if the tests were truly independent. Factors that might indicate independence should have been documented (e.g., time interval between experiments, different animal shipment, different reagents, different operator).

The Panel was also asked if they considered the interlaboratory reproducibility of the LLNA: DA to have been adequately evaluated and compared to the traditional LLNA. The Panel noted that the interlaboratory reproducibility of the assay could not be adequately evaluated given the lack of original study data and limitations in the study design. Study design limitations included:

- Pooled lymph nodes were used from mice within a dose group. This precluded an analysis of variation between laboratories.
- The lead laboratory established the dose levels to be tested by the other laboratories participating in the interlaboratory validation effort. In a minority opinion, Drs. Nathalie Alépée and Michael Woolhiser asserted that for an effective and efficient interlaboratory evaluation, it seemed reasonable to set dose levels for all laboratories based on results from the lead laboratory.

In addition, the Panel considered that the interlaboratory studies could benefit by performing more than one test on two commonly used positive controls (i.e., HCA and DNCB).

The draft LLNA: DA BRD contained an analysis of data from two interlaboratory reproducibility validation studies that used coded substances, as well as an intralaboratory accuracy validation study with 31 substances that were not coded. The Panel was asked if they considered the lack of coding of the test substances to have adversely impacted or biased the intralaboratory accuracy evaluation. The Panel commented that, in the validation of a new assay, it is better to avoid the potential for bias by testing coded substances.

However, the Panel concluded the data already generated for the LLNA: DA test method should be considered and not be rejected in the current validation evaluation.

The lead laboratory established the dose levels tested by the participating laboratories in the two interlaboratory validation studies. The Panel was asked if this adversely impacted or biased the evaluation. The Panel considered that the choice of the maximum test substance concentration is crucial for the proper performance of the traditional LLNA as well as any modified LLNA. Thus, predetermining the dose levels to be tested for each substance might have reduced variability between the two interlaboratory studies. In a minority opinion, Drs. Nathalie Alépée and Michael Woolhiser asserted that for an effective and efficient interlaboratory evaluation, it seemed reasonable to set dose levels for all laboratories based on results from the lead laboratory.

3.2.5 Data Quality

The studies evaluated in the draft BRD for the LLNA: DA were not conducted in accordance with GLP guidelines although they were reportedly done in laboratories that conduct GLP studies, and were conducted "in the spirit" of GLP (K. Idehara, personal communication). Furthermore, the original records for the interlaboratory studies were requested but have not yet been obtained. As a result, an independent audit could not be conducted to confirm that the reported data was the same as the data recorded in laboratory notebooks. The Panel was asked to comment on the potential impact this might have had on the evaluation of the LLNA: DA. The Panel commented that, ideally, GLP compliance is recommended for validation studies, but the current studies should not be rejected based on the lack of GLP compliance alone. However, all the raw data obtained through the validation process should be made available and audited for accuracy. The Panel further commented that since the original records for the interlaboratory studies have not yet been provided, recommendations from ICCVAM should be contingent upon receiving these data. Obtaining original laboratory records is a necessary step to confirm that all data generated during the validation studies have been provided, and that the reported data are the same as the data recorded in laboratory notebooks.

3.2.6 Consideration of All Available Data and Relevant Information

The Panel was asked to comment on whether all of the relevant data identified in published or unpublished studies that employed the LLNA: DA had been adequately compared. The Panel viewed that, generally, it seemed that all of the relevant results had been adequately identified and considered. However, as mentioned above, all of the original data supporting these results have not been provided. The Panel again expressed concern related to the effect of pretreating the mouse ear with 1% SLS and the Panel therefore recommended that the results from van Och et al. (2000) and De Jong et al. (2002) should be considered.

3.3 Comments on the Draft ICCVAM Test Method Recommendations on the LLNA: DA

3.3.1 Test Method Usefulness and Limitations

The Panel was asked to comment on whether the available data supported the ICCVAM draft recommendations for the LLNA: DA procedure in terms of the proposed test method usefulness and limitations. The Panel agreed with ICCVAM's recommendation, which stated that the LLNA: DA might be useful for identifying substances as potential skin sensitizers

and non-sensitizers, but this recommendation was contingent upon the receipt of additional data and information. The Panel further added that information on the possibility of skin reactions suggestive of the onset of the elicitation phase and the impact of the 1% SLS pretreatment step on the performance of the LLNA: DA should be evaluated. The Panel also considered that the ICCVAM proposed limitations needed to be more clearly defined, as it was not clear from the draft recommendations what points were considered as limitations. For instance, limitations that are known for the traditional LLNA would likely apply to this modified protocol as well and these should be noted.

The Panel was asked whether restrictions on using radioactive materials would warrant that the LLNA: DA be routinely recommended for hazard identification of skin sensitizing substances in lieu of having to possibly use guinea pig tests. The Panel noted that, based on gaps in the currently available dataset and information described in this report, the LLNA: DA could not yet be recommended for the routine use for hazard identification of skin sensitizing substances, regardless of whether restrictions on using radioactive materials were present or not. Generally, non-radioactive LLNA test methods are preferred in lieu of using guinea pig tests because fewer animals are used and animal pain and distress is reduced. However, policy issues regarding restrictions on radioactivity should have no impact on this science-based conclusion.

The Panel was asked if, from a public health perspective, the recommended guidance for evaluating negatives were sufficient to address concerns associated with the false negative rate of 5% (1/19 substances) calculated for the LLNA: DA. The Panel noted that this was not a scientific question, rather a risk characterization issue, and could not be answered without considering other factors such as intended use, target population, etc. The Panel was also asked if, from a testing strategy perspective, the ICCVAM guidance addressed concerns associated with the false positive rate of 10% (1/10 substances) calculated for the LLNA: DA and/or if they had other suggestions for additional guidance or limitations. The Panel again commented that this was not a scientific question but a risk characterization issue and could not be answered without considering other factors such as intended use, target population, etc. Furthermore, the Panel noted that it would be difficult to generalize the finding of one test substance being a “false” result. Instead, they considered it better to identify reasons why a substance was a “false” result. Certainly, if a “false” result is suspected, confirmatory testing with another mouse LLNA method was not recommended. It might be important to follow a suspected inaccuracy with an investigation of the mechanistic basis for the discordance.

3.3.2 Test Method Protocol

The Panel was asked if they agreed that the available data supported the ICCVAM draft recommendations for the LLNA: DA procedure in terms of the proposed test method standardized protocols or what recommendations they would make. The Panel noted that available data did not support all of the ICCVAM draft recommendations in the LLNA: DA standardized protocol. First, the ICCVAM protocol (ICCVAM 1999; Dean et al. 2001) for the traditional LLNA recommends using at least five animals per dose group. Although the Panel agreed that five animals per dose group should be recommended for validation studies, they suggested that power calculations would be useful in determining if subsequent use of the modified test method could use fewer animals per dose group. For the LLNA: DA test method, the Panel noted that based on statistical power calculations that were provided as

supplemental information, using four animals per group instead of five did not appear to be a limitation (i.e., detecting a 3.0-fold increase in the SI with four animals per group was estimated to have a 99% confidence level). In addition, the Panel generally agreed with the recommendation in the ICCVAM protocol (ICCVAM 1999; Dean et al. 2001) that individual animal data should be collected. A minority opinion by Drs. Nathalie Alépée, Thomas Gebel, Dagmar Jirová, Raymond Pieters, and Michael Woolhiser stated that if laboratories were operating under OECD guidance (OECD 2002) and a reliable validation dataset had been generated, then pooled data from at least four animals could be considered acceptable.

Of greater importance, the Panel concluded that pretreatment with 1% SLS should not be accepted until its impact on the performance of the LLNA: DA has been adequately characterized. Although used at a concentration below its EC3, the Panel was concerned about pretreating the mouse ear with an irritant reported as positive in the traditional LLNA. To demonstrate that these concerns are not justified, the Panel recommended that substances that are weak irritants and weak sensitizers be tested in the LLNA: DA with and without pretreatment with 1% SLS. It also needed to be demonstrated that the 1% SLS pretreatments, as well as the additional test substance treatment on day 7, did not induce a skin reaction that could be indicative of the onset of the elicitation phase of skin sensitization.

The Panel was asked to comment on whether the limit dose procedure could be applied to the LLNA: DA. The Panel concluded that if the limit dose procedure is considered applicable to the traditional LLNA, then it should also be applicable to the LLNA: DA, in order to reduce the number of animals used. This would require adherence to the LLNA: DA test method protocol, with the exception that the middle and low dose groups would be omitted in the limit dose version.

3.3.3 Future Studies

The Panel was asked if they agreed that the available data support the ICCVAM draft recommendations for the LLNA: DA in terms of the proposed future studies or, if not, what recommendations they would make. The Panel stated that the available data supported the ICCVAM draft recommendations for the LLNA: DA in terms of a more comprehensive evaluation using more non-sensitizers within and across laboratories. In a minority opinion, Dr. Thomas Gebel stated that although testing of more non-sensitizers might be warranted for interlaboratory validation studies, a sufficient number of non-sensitizers had been tested within the same laboratory (Table 6.3 in the draft BRD).

However, the Panel viewed that there were additional studies that ICCVAM might consider. As previously mentioned, the Panel recommended that the 1% SLS pretreatment step should not be accepted until its impact on the performance of the LLNA: DA had been adequately characterized. Furthermore, it should be demonstrated that such pretreatments did not induce a skin reaction that could be indicative of the onset of the elicitation phase of skin sensitization. It might also be of interest to evaluate ATP as a marker of lymph node proliferation using the traditional LLNA dosing scheme and lymph node collection schedule. Lastly, the Panel considered that studies on the reliability of outlier analysis in small sample sizes and the effects of reduced sample size on the power of the LLNA: DA test method should be proposed.

3.3.4 *Performance Standards*

The draft LLNA: DA BRD indicated that the LLNA: DA protocol differed from the ICCVAM-recommended protocol for the traditional LLNA (ICCVAM 1999; Dean et al. 2001) in the method used to assess lymphocyte proliferation in the auricular lymph nodes. In addition, there are differences between the two protocols that relate to how and when the test substance is applied and when the lymph nodes are collected (Table 2-1 and Appendix A in the draft LLNA: DA BRD). According to the proposed draft ICCVAM LLNA Performance Standards for the traditional LLNA, any change to the LLNA protocol other than the method used to assess lymphocyte proliferation would be considered a “major” change. The Panel was asked if they agreed that these should be considered “major” changes and therefore the usefulness and limitations of the LLNA: DA should not be assessed using the draft ICCVAM LLNA Performance Standards. The Panel commented that answering this question depended on having a clear definition of what constitutes a “major” versus a “minor” change, and what may constitute a different protocol altogether. Depending on the goal of the assay, whether a change is “major” versus “minor” may not be relevant. Ultimately, if a test method is able to make the correct prediction with regard to the dermal sensitization potential of a test substance, then the issue of “major” versus “minor” modifications might not apply. Considering the robust nature of the current draft ICCVAM LLNA Performance Standards, it is difficult to identify the need for additional requirements for methods like the LLNA: DA. Thus, the draft ICCVAM LLNA Performance Standards could be used to evaluate the LLNA: DA as a mechanistically and functionally similar test method.

The Panel was asked, even if the draft ICCVAM LLNA Performance Standards were not found applicable to the LLNA: DA, whether an analysis based on 13 of the 18 proposed required reference substances in the performance standards would impact the overall evaluation of the test method accuracy. The Panel commented that the accuracy analysis based on 13 of the 18 proposed required reference substances in the performance standards (with one false negative substance) should have no impact on the overall evaluation of test method accuracy as 31 substances have been tested. However, given the concern regarding pretreatment with 1% SLS, the Panel stated that testing of substances with and without 1% SLS was needed to characterize the effect of this pretreatment on the performance of the assay. The Panel concluded that as described above, the idea of “major” versus “minor” changes might be reconsidered, thus the current draft ICCVAM LLNA Performance Standards could be applicable to the LLNA: DA as a mechanistically and functionally similar test method.

[This Page Intentionally Left Blank]

4.0 Non-Radioactive LLNA Protocol – The LLNA: Bromodeoxyuridine Detected by Flow Cytometry (BrdU-FC) Test Method

4.1 Comments on the Draft BRD for Completeness, Errors, and Omissions

4.1.1 General Comments

The Panel was asked if there were any errors in the draft LLNA: BrdU-FC BRD that should be corrected, if omissions of existing relevant data had been identified, or if there was additional information that should be included. The Panel noted that overall, errors and omissions in the draft LLNA: BrdU-FC BRD were few. The majority of omissions relating to the data records were identified in the text, and all reasonable efforts to obtain additional information from MB Research Labs, the developer of the LLNA: BrdU-FC, appear to have been made.

The following describes the identified errors, omissions, and/or information gaps in the draft LLNA: BrdU-FC BRD that should be addressed:

- Data are available in the peer-reviewed literature on the application of BrdU in the LLNA with histochemical or enzyme-linked immunosorbent assay (ELISA) detection. This could be briefly mentioned in the final BRD for the LLNA: BrdU-FC method, simply as a means of indicating the utility of non-radiolabeled tracer methods in the LLNA.
- It should be noted that a potential reason why nickel chloride was negative in the LLNA: BrdU-FC may be due to oral tolerance in the mice that was induced by nickel-containing nipples of drinking bottles and nickel cages (Van Hoogstraten et al. 1993).
- The vehicle(s) used with the test substances should be stated.
- Information on experience of the inter-laboratory transferability of other technologies that depend upon flow cytometry technology as the key data read-out should be included.
- All raw data for the LLNA: BrdU-FC and the enhanced LLNA (eLLNA): BrdU-FC should be made available.

4.1.2 Comments with Specific References to the Text

The Panel stated that the following comments and/or suggested corrections relevant to specific parts of the draft LLNA: BrdU-FC BRD text should be addressed:

- Line 226: Citation was made to a reference dated 2001 by MB Research Labs which established their development of the LLNA: BrdU-FC; however, no reference was included in Section 12.0 (References) of the draft BRD.
- Lines 232-233: For a sensitizer, the SI should be greater than *or equal to* three.
- Line 246: “i.e., positive” should be explained.
- Line 254: 11% should be 17%.

- Line 263: For purposes of completeness, it may be worthwhile to add a brief description of the comparative accuracy of the available traditional LLNA, LLNA: BrdU-FC, and eLLNA: BrdU-FC results versus human maximization/patch test data. This information is of importance and displayed in Tables 6-1 and 6-2.
- Lines 286-288: The issue of the refinement/reduction in animal use that might follow the availability of a scientifically validated non-radioactive variant of the LLNA was mentioned in the draft BRD as a benefit but it was not quantified, and no authoritative reference was cited in support.
- Line 288: The final LLNA: BrdU-FC BRD should explain why the BrdU method would result in less pain and distress to the animals (i.e., does the route of injection of BrdU vs. ³H-methyl thymidine produce less discomfort?).
- Line 335: Reference was made to a citation dated 2001 by MB Research Labs which established their development of the LLNA: BrdU-FC; however, no citation was included in Section 12.0 (References) of the draft BRD.
- Line 356: The sentence starting “To evaluate excessive skin...” implies that evaluation of excessive skin irritation by measuring ear thickness is recommended by the ICCVAM LLNA protocol although it is only recommended in the LLNA: BrdU-FC protocol.
- Lines 365-366: Consider supplementing the list of abbreviations for Figure 2-1 with B220+, B:T, CD69+, and IAK+. Also, the figure shows I-Ak+ while all other text uses IAK+. The MB Research Labs protocol shows I-Ak+.
- Figure 2-1: Should be redrawn to show the SI decision point lines coming off of the “Analyze Proliferating LNC (lymph node cells)” box rather than the “Inject BrdU and Excise...” box.
- Lines 500-503: Classification of “equivocal results” was unclear without data comparison (i.e., benzocaine produced divergent results in both tests). Were these results unlike what was expected from human data? What were the data for salicylic acid and mercaptobenzothiazole?
- Lines 552-558: It may be useful for comparative purposes to add summary accuracy data for the traditional LLNA versus human maximization/patch test data from the larger data set reported in the 1999 ICCVAM LLNA report to the section of the final BRD which discusses performance of the LLNA: BrdU-FC method.
- Table 6-3: Benzocaine was missing and it seems that salicylic acid was the same in both traditional LLNA and LLNA: BrdU-FC. Mercaptobenzothiazole was not reported for the LLNA: BrdU-FC. This needs to be corrected or explained.
- Table 6-5: The human outcome for benzalkonium chloride and ethylene glycol methacrylate should be negative. See also Table 6-6.

4.2 Comments on the Validation Status of the LLNA: BrdU-FC

4.2.1 Test Method Protocol

The LLNA: BrdU-FC protocol includes routine measurements of ear swelling as an indicator of excessive dermal irritation. The Panel was asked if they considered this procedure to be an appropriate approach and if this measurement should be recommended for routine inclusion into all LLNA protocols. The Panel stated that, as a quantitative parameter associated with inflammatory cell influx and fluid retention near the site of test substance application, ear swelling (or other quantitative measurements) should be carefully considered for inclusion into all LLNA protocols. This might assist in differentiating between sensitizers and irritants, assist in the interpretation of equivocal results, and possibly detect other procedure-related problems that might require further exploration/consideration.

The LLNA: BrdU-FC protocol also includes optional quantification of immunophenotypic markers as an additional mechanism for distinguishing irritants from sensitizers. The Panel was asked if they considered this to be an appropriate approach to reduce false positives, and if the correct markers were being considered. The Panel was also asked if these measurements should be recommended for routine inclusion in the LLNA: BrdU-FC. The Panel agreed that the use of immunological markers would be appropriate for detailed studies, as it might reduce the frequency of false positives (irritants) and improve comparisons with human data. However, since the primary use of the LLNA is for discrimination of human hazard from direct chemical contact, it could be argued that some false positives are acceptable (especially for methods which have relatively lower rates of false negatives). Given this dominant use, application of immunological markers would likely be too detailed and costly for routine LLNA use. Thus, the Panel suggested that results of ear swelling measurements be compared with the more technically complex flow cytometry markers to determine if similar results might be obtained. Furthermore, alternative immunological markers for discriminating between irritants and sensitizers may be available, although the draft LLNA: BrdU-FC BRD did not contain information allowing any informed decision on whether other markers might be more predictive. Thus, based on current knowledge, the current markers suggested in the draft LLNA: BrdU-FC BRD seemed acceptable to the Panel. Two other possibilities suggested were a surface marker relating to CD4 T-helper cells (Th) or Th1 cells (interferon- γ).

The Panel was also asked to comment on the appropriateness of the "sequential strategy" used in the eLLNA: BrdU-FC (see Figure 2-1 of the draft LLNA: BrdU-FC BRD). Generally, the Panel viewed that incorporation of immune parameters improved the value of a predictive assay and may also help explain mechanisms, which is important. Still, the "sequential strategy" used in the eLLNA: BrdU-FC for discriminating irritation from sensitization might be more sensible for research studies because of resource and cost considerations, and may not be appropriate for routine use of the LLNA in hazard identification. For human hazard detection, more simplified methods should be available for discrimination of irritants.

4.2.2 Substances Used for the Validation Studies

The Panel was asked to consider if the substances tested in the LLNA: BrdU-FC were representative of a sufficient range of chemical classes and physicochemical properties such

that the test method would be applicable to any of the types of chemicals and products that are typically tested for skin sensitization potential. The Panel agreed that if the proviso that the applicability domain limitations published for the traditional LLNA remained in force, the substances tested in the LLNA: BrdU-FC seemed representative of a sufficient range of chemical classes and physical chemical properties and it would likely be applicable to many of the types of chemicals and products that are typically tested for skin sensitization potential. However, the available LLNA: BrdU-FC database was relatively small compared to the large number of substances assessed in the traditional LLNA and this implied some caution in assuming that assay performance was concordant with the traditional LLNA.

4.2.3 Test Method Accuracy

The accuracy analysis presented in the draft LLNA: BrdU-FC BRD was based on overall concordance with the traditional LLNA. Accuracy statistics compared to the guinea pig tests and human data/experience were also provided. The Panel was asked if these comparisons were appropriate for assessing the accuracy of the LLNA: BrdU-FC. The Panel viewed that since the traditional LLNA is used to provide human hazard identification and information relevant to human health, the accuracy statistics compared to human data/experience were important. Since the LLNA: BrdU-FC is fairly similar to the traditional LLNA, guinea pig comparisons might not have been necessary. However, taken together, the availability of both human data/experience and guinea pig data allowed additional insights that might have expanded the applicability domain of the LLNA: BrdU-FC, or indicated improved performance with respect to LLNA false negatives and positives.

The Panel was then asked if the relevance (e.g., accuracy/concordance, sensitivity, specificity, false positive and false negative rates) of the LLNA: BrdU-FC had been adequately evaluated and compared to the traditional LLNA. The Panel agreed that the relevance of the LLNA: BrdU-FC was adequately evaluated and compared to the traditional LLNA, and supported the inclusion of accuracy analyses with and without equivocal materials.

Three substances (benzalkonium chloride, resorcinol, and Tween 80) produced a false positive response compared to the traditional LLNA and guinea pig test when tested using the LLNA: BrdU-FC (based on immunophenotyping, benzalkonium chloride was subsequently classified as an irritant rather than a sensitizer). The Panel was asked if they could identify any characteristics associated with these or similar substances that might suggest that using the LLNA: BrdU-FC to test such substances would not be appropriate or that positive results for substances with such properties may warrant additional testing. Overall, the Panel stated that there were not any patterns or unifying concepts that explained the three false positive results in the available data set. They noted that only a single laboratory is using the LLNA: BrdU-FC method and recommended that the raw data on which the reports were prepared be made available in order to allow further investigation. The Panel also suggested that additional studies be conducted to determine whether LLNA: BrdU-FC results with these three substances are repeatable.

Dr. Raymond Pieters stated that benzalkonium chloride and Tween 80 are considered aggressive irritants, but both published data (Manetz and Meade 1999; Varani et al. 2008) and unpublished data from his laboratory has shown that benzalkonium chloride (5%) is more potent than SLS in the stimulation of lymph node cell proliferation and may therefore may actually be considered a sensitizer. However, in the traditional LLNA these compounds

did not increase the SI above the threshold for a positive response (i.e., $SI \geq 3$), so they were identified as non-sensitizers.

4.2.4 Test Method Reliability

The Panel was asked if the intralaboratory reproducibility of the LLNA: BrdU-FC had been adequately evaluated and compared to the traditional LLNA and if any limitations were apparent based on this assessment. The draft LLNA: BrdU-FC BRD analyzed data from repeat testing of HCA in six different vehicles and intralaboratory reproducibility was assessed by a coefficient of variation (CV) evaluation. The calculated CVs ranged from 30% to 53%. The Panel agreed that the relatively large SD and associated CV values raised questions about the extent of experiment-to-experiment variability. There was less concern about vehicle choice and effects on the range of group means than about the CVs greater than 50% for the group means of HCA tested in DMSO and AOO. The large number of repeated experiments for these tests would have been expected to dramatically reduce variability. The Panel concluded that the results suggested that key elements of assay standardization were not yet developed. Further evaluation using other positive control substances would have been valuable to more adequately characterize reproducibility.

4.2.5 Data Quality

The studies evaluated in the draft BRD for the LLNA: BrdU-FC were not all conducted in accordance with GLP guidelines although they were done in a laboratory that routinely conducts GLP studies (G. DeGeorge, personal communication). The Panel was asked to discuss what impact this might have on the evaluation of the LLNA: BrdU-FC. The Panel considered that, even without formal GLP compliance, the current LLNA: BrdU-FC results appeared to reflect a sincere attempt to perform work of high quality. The only area in which a lack of full GLP compliance may have been a source of assay variability was in the quantitative analysis of dosing solutions. For instance, failure to appreciate differences in composition of dosing solutions between experiments caused by test article instability or other phenomena may account for the relative large variability in intralaboratory data and possibly of some of the discordant results (i.e., false negatives and differences in LLNA: BrdU-FC results between repeat studies for the same substance). Thus, the Panel viewed that any additional studies undertaken to validate the test method should ideally be GLP-compliant.

Furthermore, the original records for these studies were requested but had not yet been obtained at the time of the Panel review. As a result, an independent audit could not be conducted to confirm that the reported data was the same as the data recorded in laboratory notebooks. The Panel was asked if they agreed that any recommendations from ICCVAM should be contingent upon the completion of such an audit and findings that there were no significant errors in data transcription. The Panel agreed that, although a request for original data had been made, it was good practice to hold final recommendations until an independent audit could be performed. While it would be expected that no serious errors would be uncovered which would alter the current findings, an audit would confirm assay performance to date and position ICCVAM for further consideration of the LLNA: BrdU-FC.

4.2.6 Consideration of All Available Data and Relevant Information

The Panel was also asked if, based on the draft LLNA: BrdU-FC BRD, all the relevant data identified in published or unpublished studies that employ this test method had been adequately considered. Furthermore, they were asked that if there were other comparative test method data that were not considered in the draft BRD, how such data might be obtained. Overall, the Panel considered that all the relevant data identified in published or unpublished studies that employed this test method had been adequately considered in the draft LLNA: BrdU-FC BRD. However, some additional information was available in the peer-reviewed literature on application of BrdU in the LLNA with other methods of detection (e.g., histochemistry, ELISA). The Panel felt that these could have been briefly mentioned in the draft BRD for the LLNA: BrdU-FC method, simply as a means of indicating the utility of non-radiolabeled tracer methods in the LLNA. Furthermore, if an analysis of the CV for the traditional LLNA was undertaken, a more direct comparison with the LLNA: BrdU-FC could have been performed.

4.3 Comments on the Draft ICCVAM Test Method Recommendations on the LLNA: BrdU-FC

4.3.1 Test Method Usefulness and Limitations

The Panel was asked if they agreed that the available data and test method performance support the ICCVAM draft recommendations for the LLNA: BrdU-FC in terms of the proposed test method usefulness and limitations (i.e., that it may be useful for identifying substances as potential skin sensitizers and non-sensitizers, but that more information and data are needed before a recommended use of the LLNA: BrdU-FC can be made). The Panel agreed that the available data and test method performance of the LLNA: BrdU-FC support the draft ICCVAM recommendations. They considered the proposed test method usefulness and limitations to have well summarized the limits of the information supplied and the additional information that would need to be generated or provided for further consideration of this test method. As a result, the LLNA: BrdU-FC could not at this stage be considered scientifically validated as a replacement alternative to the traditional LLNA. Still, the test method recommendation should clearly state that the test method was not “invalid” but simply that there was currently not sufficient evidence and information to affirm that it had been adequately validated by ICCVAM. Instead, the Panel considered that the LLNA: BrdU-FC could be recommended in instances where mechanistic information about a sensitizer is required.

The Panel was asked if restrictions on using radioactive materials were or were not present, whether or not the LLNA: BrdU-FC should be routinely recommended for hazard identification of skin sensitizing substances in lieu of having to possibly use guinea pig tests. The Panel agreed that it is preferable to use alternative methods for the LLNA (i.e., ELISA detection of BrdU or histochemical detection of BrdU-labeled cells), as opposed to application of guinea pig test methods, if a limitation on radioisotope use exists (e.g., the lack of a radioactivity use license). This rationale is based on avoidance of the less quantitative guinea pig test methods, which may employ adjuvant treatment with associated animal stress and harm. Still, at this time, the Panel considered that data gaps in the LLNA: BrdU-FC method precluded recommending it for routine hazard identification of skin sensitizing

substances in lieu of the traditional LLNA, whether or not limitations on using radioactive materials exist. Policy issues regarding restrictions on radioactivity should have no impact on this science-based conclusion.

The Panel was asked if the ICCVAM recommendations adequately addressed concerns associated with the false positive rate of 17% (3/18 substances) calculated for the LLNA: BrdU-FC and if there were other suggestions for additional guidance or limitations that should be considered. The Panel agreed that the relatively high false positive rate was adequately identified and discussed, and that no mechanistic reason could be identified for these results based on available information. The Panel noted that it might be worthwhile to point out in the final BRD the impact on human health of false positive results versus false negative results in the context of hazard screening and identification. Consideration of factors such as intended use and target population of the false positive substances would further aid in characterizing human risk for these substances.

4.3.2 Test Method Protocol

The Panel was asked whether or not they agreed that the available data supported the ICCVAM draft recommendations for the LLNA: BrdU-FC procedure in terms of the proposed test method standardized protocol and if not, what recommendations would they make. The Panel noted that the draft ICCVAM recommendations for conduct of a standardized method for the LLNA: BrdU-FC variant were relatively brief and stated only that all applicable portions of the 1999 ICCVAM procedure be carefully followed. The Panel agreed that the available data supported the ICCVAM draft recommendations for the LLNA: BrdU-FC procedure in terms of adhering to the ICCVAM LLNA protocol (ICCVAM 1999; Dean et al. 2001). In particular, the Panel agreed that at least five animals per dose group should be used, particularly in light of Dr. Haseman's power analysis (see **Table 4-1**). The Panel did note however that power calculations could be undertaken to determine if fewer animals per dose group might be adequate for post-validation studies, though Dr. Haseman's power analyses suggest that this is unlikely. The majority of the Panel also agreed with the ICCVAM-recommended protocol to use individual animal data although a minority opinion by Drs. Nathalie Alépée, Thomas Gebel, Dagmar Jírová, Raymond Pieters, and Michael Woolhiser stated that if laboratories were operating under OECD TG 429 guidance (OECD 2002) and a reliable validation dataset had been generated, then pooled data from at least four animals per dose group could be considered acceptable.

Further, the Panel considered the methodological description of the LLNA: BrdU-FC procedure supplied by MB Research Labs (Appendix A to the draft BRD) to be comprehensive. The utility of ear swelling or other methods to detect inflammation/excessive local irritation appear to be warranted in every variation of the LLNA (including the traditional LLNA), but should be further investigated before routine inclusion in any protocol is recommended.

The Panel was asked whether the LLNA limit dose procedure could be applied to the LLNA: BrdU-FC. The Panel agreed that the LLNA limit dose procedure could be applied to the LLNA: BrdU-FC as long as the limitations associated with the limit dose procedure were appreciated. Furthermore, application of the limit dose procedure to the LLNA: BrdU-FC would require adherence to a validated LLNA: BrdU-FC test method protocol with the exception that the middle and low dose groups would be excluded. Furthermore, it would

need to be confirmed that the number/pattern of sensitizers that would have been identified/missed from the “high dose” group would mirror that of the traditional LLNA.

Table 4-1 Power Calculations for the LLNA: BrdU-FC

	3.0-fold Increase	2.5-fold Increase	2.0-fold Increase	1.5-fold Increase	1.3-fold Increase
Mean Rx response	30279	25232.5	20186	15139.5	13120.9
Log (Mean Rx response)	10.318	10.136	9.913	9.625	9.482
Difference from control (log scale)	1.098	0.916	0.693	0.405	0.262
Difference/SD	1.75	1.46	1.10	0.65	0.42
Power for N=5	80%	50-80%	<50%	<50%	<50%
Power for N=4	50-80%	50%	<50%	<50%	<50%
Power for N=3	50%	<50%	<50%	<50%	<50%
Other Power	95% (N=9)	95% (N=12)	95% (N=19)	95% (N=52)	95% (N>100)
Other Power	90% (N=7)	90% (N=10)	90% (N=15)	90% (N=42)	90% (N>100)

The power calculations above are based on a one-sided $p < 0.05$ Student's t-test applied to log-transformed data from vehicle control LLNA: BrdU-FC tests.

Abbreviations: Fold-increase=Required increase above the vehicle control for a positive response; N=Number; Rx=Treatment; SD=Standard deviation.

4.3.3 Future Studies

Finally, the Panel was asked whether they agreed that the available data supported the ICCVAM draft recommendations for the LLNA: BrdU-FC in terms of the proposed future studies. The Panel agreed that the ICCVAM draft recommendations for future studies highlighted the unanswered questions raised by the available database. Specifically, conducting interlaboratory studies as a part of the validation process was considered important. As mentioned previously, the Panel viewed that the immunological markers suggested for the LLNA: BrdU-FC in the draft BRD were acceptable but that additional immunological markers for discrimination of irritant versus sensitization phenomena might also be identified. A suggestion for a future study was to use the surface marker relating to CD4 Th cells or internal marker relating to Th1 cells (interferon- γ).

In general, for any future work, the Panel considered that efforts should be made to decrease the variability and thereby increase the power of the test in order to ensure that more animals were not needed relative to the traditional LLNA or other alternative LLNA protocols. For instance, further optimization of the LLNA: BrdU-FC method should include kinetic studies to demonstrate that the optimal protocol was being used.

4.3.4 Performance Standards

The draft BRD indicated that the LLNA: BrdU-FC protocol differs from the ICCVAM-recommended protocol for the traditional LLNA (ICCVAM 1999; Dean et al. 2001) in the method used to assess lymphocyte proliferation in the auricular lymph nodes. According to the proposed draft ICCVAM LLNA Performance Standards for the traditional LLNA, any change to the LLNA protocol other than the method used to assess lymphocyte proliferation

was considered a “major” change. According to this criterion, the Panel considered the protocol differences between the LLNA: BrdU-FC and the traditional LLNA to be “minor” changes, and therefore considered that the validity of this test method could be based on the draft ICCVAM LLNA Performance Standards. However, the Panel also recognized that this depended on a clear definition of what constituted a “major” versus a “minor” change, or a different protocol altogether. Thus, further consideration of this topic could be addressed once the recommendations for the current draft ICCVAM LLNA Performance Standards were finalized. The Panel found it difficult to identify any additional requirements for methods like the LLNA: BrdU-FC.

Even if the draft ICCVAM LLNA Performance Standards were not found to apply to the LLNA: BrdU-FC, the Panel considered that the impact of the LLNA: BrdU-FC accuracy analysis based on 13 of the 18 proposed required reference substances in the draft ICCVAM LLNA Performance Standards should not have a major impact on the overall evaluation of test method accuracy, as 45 substances, representative of an appropriate range, were tested. However, based on consideration for development of LLNA performance standards, it would be desirable for validation purposes that the substances missing from the range of 18 standard materials be assessed in the LLNA: BrdU-FC protocol.

The draft LLNA: BrdU-FC BRD also indicated that three out of six sensitizers for which EC3 data were available had EC3 values that were outside of the proposed 0.5x to 2.0x EC3 acceptability range, which was developed based on the traditional LLNA. The Panel viewed that the primary concern seemed to have less to do with the variation in the response than with a concern that the range of response would skew the interpretation of any LLNA: BrdU-FC results used for sensitization potency estimates. Furthermore, it was not known if the same vehicle was used to derive both EC3 values/ranges. The proposed 0.5 x to 2.0 x range seemed to be based upon empirical/goodness of fit rather than any biological constant. The appropriateness of this range should be considered further when the finalized ICCVAM LLNA Performance Standards document is considered. In general, if the vehicles were different the question is irrelevant.

4.3.5 Comments with Specific References to the Text

The Panel made the following comment with specific reference to the text in the draft ICCVAM test method recommendations on the LLNA: BrdU-FC and suggested that it be addressed:

- Lines 39-41: For parallel construction of this sentence with the preceding sentence, suggest substituting the following "One of the other equivocal substances, salicylic acid, is one of the recommended reference standard materials used as a non-sensitizer in the draft ICCVAM LLNA Performance Standards...". This is based on the assumption that salicylic acid was the substance intended for discussion and that it was used in the draft ICCVAM LLNA Performance Standards as a model non-sensitizer.

[This Page Intentionally Left Blank]

5.0 Non-Radioactive LLNA Protocol - The LLNA: Bromodeoxyuridine Detected by ELISA (BrdU-ELISA) Test Method

5.1 Comments on the Draft BRD for Completeness, Errors, and Omissions

5.1.1 General Comments

The Panel was asked if there were any errors in the draft LLNA: BrdU-ELISA BRD that should be corrected, if omissions of existing relevant data had been identified, or if there was additional information that should be included. The Panel noted that, in general, all of the data included were relevant, and that it was apparent that considerable effort had been involved in carefully developing the comprehensive database. The Panel noted that they would have preferred to have the original papers by Dr. Takeyoshi included in the review materials, but they were easily retrieved from the journal websites. The Panel indicated that raw data (i.e., the actual optical density at 370 nm [OD₃₇₀] readings for the triplicates and the SD of the triplicates) are necessary for a thorough evaluation. Additionally, the Panel noted that only a relatively small number of substances had been tested in the LLNA: BrdU-ELISA.

When considering the animal welfare impact of implementing the LLNA: BrdU-ELISA, the Panel agreed that it would be less painful than guinea pig tests in those circumstances where the use of radioactive materials are restricted. Thus, the Panel agreed that the test represents a potential refinement. The Panel further stated that, if there is not an option to replace the guinea pig test with a non-animal test, decreasing the extent of pain and distress should be the first animal welfare priority. The Panel cautioned that at some point, however, the numbers of animals being utilized must be considered. An eventual recommendation that the LLNA: BrdU-ELISA be routinely used instead of guinea pig test methods where the use of radioactive substances are restricted would apparently require a significant increase in the number of mice killed per test (to increase the statistical power of the test method - see **Section 5.1.2** below) if an SI ≥ 1.3 is deemed the appropriate criterion to use for determining a positive response. The Panel stated that it would be helpful to know how many guinea pigs are currently being used nationally and internationally for skin sensitization tests, and how many mice would be used in the LLNA: BrdU-ELISA with the SI ≥ 1.3 criterion. Even an order of magnitude estimate would help the Panel judge whether the increase in numbers of mice needed is justified as the quest to relieve pain in guinea pigs is pursued.

5.1.2 General Statistical Comments

The Panel was concerned about using an SI of ≥ 1.3 to optimize the performance of the LLNA: BrdU-ELISA method. One Panel member's extensive experience with ELISA protocols was cited as evidence that the difference between the OD₃₇₀ of the vehicle and the positive test at 1.3 would not likely be statistically significant. The Panel recommended that the raw data must be reviewed to evaluate this. In addition, based on Dr. Joseph Haseman's power analysis (see **Table 5-1**), the Panel stated that it was difficult to justify using a SI ≥ 1.3 as the decision criterion since it would result in a significant increase in the number of animals needed to obtain an acceptable confidence level. In this regard, the Panel

recommended that power calculations should be routinely recommended to ensure that the appropriate number of animals per dose group is being analyzed.

Table 5-1 Power Calculations for the LLNA: BrdU-ELISA

Parameter	3.0-fold Increase	2.0-fold Increase	1.3-fold Increase
Mean Rx response	0.399	0.266	0.173
Log (mean Rx response)	-0.92	-1.32	-1.75
Difference from control (log scale)	1.10	0.70	0.27
Difference/SD	3.64	2.32	0.89
Power for N=4	99%	80-90%	<50%
Other power	95% (N=3)	95% (N=5)	50% (N=8)
Other power	–	50-80% (N=3)	80% (N=16)
Other power	–	–	90% (N=22)

The power calculations above are based on a one-sided $p < 0.05$ Student's t-test applied to log-transformed data from vehicle control LLNA: BrdU-ELISA tests.

Abbreviations: Fold-increase=Required increase above the vehicle control for a positive response; N=Number; Rx=Treatment; SD=Standard deviation.

5.1.3 Comments with Specific References to the Text

The Panel also identified the following minor formatting and grammatical errors, as well as information gaps, in the draft BRD:

- The Panel noted a discrepancy between the draft LLNA: BrdU-ELISA BRD and the draft ICCVAM LLNA Performance Standards in the vehicle used for testing 2-mercaptobenzothiazole. Table 6-2 of the draft LLNA: BrdU-ELISA BRD indicated that the vehicle was AOO but the revised draft ICCVAM LLNA Performance Standards indicated that the vehicle was DMF (see page C15, C22 of September 7, 2007, draft and page B-6 of January 7, 2008, revised draft). Additionally, Table 1 on page C-7 of the revised draft ICCVAM LLNA Performance Standards listed AOO as the vehicle for 2-mercaptobenzothiazole. For both vehicles, the revised draft ICCVAM LLNA Performance Standards indicated that the EC3 value is 2.5%, although the text on page 10 of the draft LLNA: BrdU-ELISA BRD stated “the NICEATM database of traditional LLNA studies indicates that 2-mercaptobenzothiazole has a higher EC3 value when tested in AOO (mean EC3=9.8%) compared with DMF (mean EC3=2.5%)...”
- The Panel noted that Table 6-1 of the draft LLNA: BrdU-ELISA BRD indicated that, when compared to the guinea pig and human test data, the sensitivity and specificity of the LLNA: BrdU-ELISA is lower than that of the traditional LLNA. In fact, depending on the SI threshold value used, the

sensitivity and specificity of the LLNA: BrdU-ELISA can be higher than that of the traditional LLNA.

The Panel recommended that the draft LLNA: BrdU-ELISA BRD be updated to rectify these errors and omissions.

5.2 Comments on the Validation Status of the LLNA: BrdU-ELISA

5.2.1 Test Method Protocol

The data generated for the substances analyzed in the LLNA: BrdU-ELISA test method came from auricular lymph nodes from four individual mice in each dose group. The ICCVAM-recommended LLNA protocol (ICCVAM 1999; Dean et al. 2001) and OECD TG 429 (OECD 2002) recommend a minimum of five animals per dose group when collecting individual animal data. The Panel was asked what impact might the use of four animals per dose group have on the accuracy and reliability of the LLNA: BrdU-ELISA, and if the Panel agreed with the ICCVAM recommendation that future use of this test method protocol should include five animals per dose group. The Panel majority agreed with the ICCVAM recommendation that future use of this test method should use five animals per dose group and collect individual animal data, as per the ICCVAM-recommended protocol. A minority opinion by Drs. Nathalie Alépée, Thomas Gebel, Dagmar Jírová, Raymond Pieters, and Michael Woolhiser stated that if laboratories were operating under OECD guidance (OECD 2002) and a reliable validation dataset had been generated, then pooled data from at least four animals could be considered acceptable. Based on the supplemental data provided by Dr. Haseman, the power to detect a three-fold increase with a sample size of four was determined to be 99%. These calculations, however, assume that a sample size of four is always obtained. If a sample size of four was planned and fewer usable data values were obtained, then the experiment might be compromised. Furthermore, the Panel concluded that testing for and eliminating “outliers” from experiments with small sample sizes is questionable. A reduction in sample size from five to four was not recommended unless data was provided on the frequency with which “outliers” occurred and an analysis is performed that establishes that a reduction in the nominal sample size from five to four would not compromise the performance of the test method. The Panel stated that the handling of suspected “outliers” and the use of robust statistics are issues that need to be addressed in such an analysis. For example, robust procedures may compensate for apparent “outliers” and eliminate the impulse to discard data. An example is calculating the mean values used in the SI on a log scale and then exponentiating the result to construct the SI.

The Panel also indicated that it was important to routinely include a positive control group in test method validation experiments (e.g., HCA), which was likely not the case for most of the LLNA: BrdU-ELISA validation experiments. Although the Panel did not reach consensus, they did consider the suggestion that for laboratories in which the LLNA is “routinely” performed, positive controls (e.g., HCA or a substance that matches the chemical class of the test substances) could be run at intervals for quality control purposes rather than concurrent with each experiment in which substances are tested. The Panel also discussed that omitting the concurrent positive control should not be recommended for laboratories that perform the LLNA only “occasionally”. In their discussions, the Panel was not able to conclude what should constitute “routine” or “occasional” LLNA use or what would be an appropriate interval between positive control testing when a concurrent positive control is not used.

5.2.2 Substances Used for the Validation Studies

The Panel was asked whether the LLNA: BrdU-ELISA database was representative of a sufficient range of chemical classes and physicochemical properties such that the test method would be applicable to any of the types of chemicals and products typically tested for skin sensitization potential. The Panel indicated that the ratio of solids to liquids was not comparable; more solids should be included. The Panel further indicated that more substances for which traditional LLNA data are available should be tested, and that compounds including metals (e.g., nickel, cobalt), mixtures, and substances in aqueous solutions should be included.

5.2.3 Test Method Accuracy

The current accuracy analysis using an $SI \geq 3.0$ or $SI \geq 1.3$ to identify sensitizers is based on overall concordance with the traditional LLNA. Accuracy statistics compared to the guinea pig tests and human data/experience were also provided to the Panel. The Panel was asked whether these comparisons were appropriate for assessing the accuracy of the LLNA: BrdU-ELISA. The Panel indicated that comparing the LLNA: BrdU-ELISA performance to the traditional LLNA and the guinea pig tests were appropriate. Comparisons between the LLNA: BrdU-ELISA and human data were considered particularly valuable because the traditional LLNA doesn't match human data with 100% accuracy. For this reason, the Panel considered comparing the performance of the LLNA: BrdU-ELISA with that of the traditional LLNA with respect to predicting the human outcomes to be the best method of comparing these two LLNA protocols. The Panel concluded that in moving forward with any test method recommendation, key importance should be placed on interpreting the test results and making them clinically applicable to humans.

Takeyoshi et al. (2007) performed an accuracy analysis using decision criteria other than an $SI \geq 3.0$ to classify substances as sensitizers. Maximal accuracy for the LLNA: BrdU-ELISA occurred when an $SI \geq 1.3$ was used to distinguish between sensitizers and non-sensitizers. Using this decision criterion, the LLNA: BrdU-ELISA achieved an accuracy of 91% (21/23), with a sensitivity of 100% (16/16) and a specificity of 71% (5/7) (i.e., there were no false negatives and two false positives). The Panel was asked whether this analysis supported a recommendation that the decision criteria be based on an $SI \geq 1.3$, and if there were concerns with using such a small increase (i.e., 1.3-fold) above the vehicle control response as the basis for identifying a positive response. The Panel did not support using an $SI \geq 1.3$ as the criterion for positive results. An $SI=1.0$ means there was no difference between the vehicle control and the test substance. An $SI=1.3$ represents a 30% increase from the vehicle control. The difference between the OD_{370} of the vehicle and the positive test at 1.3 may not be statistically significant. An $SI=3.0$, which represents a three-fold difference between the vehicle and a positive test, would be a more believable positive difference. If the positive test criteria must be reduced to 1.3, then the Panel questioned whether the protocol is useful in its current state. The supplemental information that provided power calculations indicated that it would not be realistic to expect to detect a 1.3 fold increase in the control response without a significant addition of animals. Although using $SI \geq 1.3$ increases the accuracy of the test, it comes at an increased cost to animals, which merits consideration. Furthermore, the ICCVAM (1999) report stated that an irritating chemical might induce proliferation, but that the response seldom exceeds an $SI \geq 3.0$ (page 6). The Panel concluded that this might

provide further justification against using a low SI (e.g., 1.3) as a threshold for a positive response.

The Panel was asked if the relevance (e.g., accuracy/concordance, sensitivity, specificity, false positive and false negative rates) of the LLNA: BrdU-ELISA, using the $SI \geq 3.0$ criterion, had been adequately evaluated and compared to the traditional LLNA (refer also to Table 6-1 of the draft LLNA: BrdU-ELISA BRD). If not, the Panel was asked what other analyses should be performed. The Panel agreed that the relevance of the LLNA: BrdU-ELISA, using the $SI \geq 3.0$ criterion, had been adequately evaluated. The Panel further stated that a better evaluation could be performed, however, if the database for the LLNA: BrdU-ELISA included more substances with traditional LLNA, guinea pig, and human data. The Panel considered the false negative rate of the test method to be excessive when results are compared with that obtained in the traditional LLNA (29/33/27% for the various datasets) or with human data (39%) – the results should be at least comparable with the traditional LLNA.

Using the $SI \geq 3.0$ criterion, there were four substances (aniline, 4-chloroaniline, 2-mercaptothiazole, and hydroxycitronellal) that produced false negative responses when tested using the LLNA: BrdU-ELISA compared to the traditional LLNA. The Panel was asked whether it could identify any characteristics associated with these or similar substances, compared to the correctly identified sensitizers, that might indicate that such substances should not be tested in the LLNA: BrdU-ELISA or that negative results for such substances should indicate a need for confirmatory testing. The Panel could not identify any characteristics associated with these substances that might allow the identification of these substances as false negatives prior to testing. The Panel stated that the LLNA: BrdU-ELISA test, using the standard $SI \geq 3.0$ to indicate positive results, simply does not perform well for identifying sensitizers.

5.2.4 Test Method Reliability

The Panel was asked whether the intralaboratory reproducibility of the LLNA: BrdU-ELISA had been adequately evaluated and compared to the traditional LLNA, and whether any limitations were apparent based on this intralaboratory reproducibility assessment. The Panel indicated that the number of substances evaluated for intralaboratory reproducibility was too few and, in some cases, there was a wide variation in repeat test results for the same substance. Only six substances (five sensitizers and only one non-sensitizer) were tested multiple times. The non-sensitizer, propylene glycol, was tested only twice and opposite results were obtained. The Panel considered the results of an intralaboratory reproducibility evaluation that was based on two discordant results only to be unacceptable. The numbers calculated in Table 7-1 of the draft LLNA: BrdU-ELISA BRD are correct, but the Panel questioned the dependability of the data since only two to three values were available for calculating the mean and CV. The Panel considered the CV values (over 30%) high, compared to the traditional LLNA (draft LLNA: BrdU-ELISA BRD Tables 7-1 to 7-3). The Panel stated that at least four independent tests with three concentrations tested represent a solid basis for calculation. The Panel considered the number of tests for intralaboratory concordance analysis to be insufficient, and stated that more intralaboratory testing is needed. The Panel recommended an evaluation of the intralaboratory reproducibility of the $EC \geq 1.3$ and that the analysis of the variability of the EC_t be conducted on a log scale.

The substances evaluated for intralaboratory reproducibility of the LLNA: BrdU-ELISA study were not coded. The Panel was asked whether the lack of coding of test substances adversely impacts or biases the current evaluation. The Panel stated that, although coding of substances is preferred for independent testing and evaluation of test results, the current data should not be rejected from consideration because the substances tested were not coded.

The Japanese Center for Validation of Alternative Methods (JaCVAM) has implemented a multi-laboratory validation study of the LLNA: BrdU-ELISA. The Panel was asked whether the study design was appropriate to adequately determine the extent of interlaboratory reproducibility for the LLNA: BrdU-ELISA. If not, the Panel was asked what other studies should be performed. The Panel stated that they had insufficient time to evaluate the study design and that they could not evaluate interlaboratory reproducibility because the study data were not available at the time of their evaluation.

5.2.5 Data Quality

The studies evaluated in the draft BRD for the LLNA: BrdU-ELISA were not conducted in strict accordance with GLP guidelines, although there were reportedly performed in laboratories that conduct GLP studies (M. Takeyoshi, personal communication). In other words, an audit report was not available. Also, the raw data were unavailable for an independent audit. The Panel was asked to discuss what impact this might have on the evaluation of the LLNA: BrdU-ELISA. The Panel concluded that ideally, validation studies should be performed in accordance with GLP guidelines. Although the systems employed for tests (i.e., test facilities, staff, reagents, etc.) were identical to those for GLP-compliant studies, the data quality may be questioned and therefore should at least be available for a retrospective independent audit. However, in this case, the Panel concluded that the lack of GLP compliance was not likely the reason for the poor results obtained with the LLNA: BrdU-ELISA.

The original records for these studies were requested but had not been received by the time the Panel convened. As a result, an independent audit could not be conducted to confirm that the reported data in peer reviewed publications and a poster presentation is the same as the raw data. The Panel was asked whether any recommendations from ICCVAM should be contingent upon the completion of such an audit and findings that there were no significant errors in data transcription. The Panel concluded that, to have confidence in data quality, ICCVAM recommendations should be contingent upon the completion of an independent audit. Moreover, if an $SI \geq 1.3$ is used as the criterion for positive results, review of the raw data is necessary to confirm statistically significant differences. The Panel concluded that this test, as described, had poor accuracy, poor sensitivity, and poor specificity. The Panel stated that changing the SI decision criterion from 3.0 to improve test performance, especially to such a drastic change as $SI \geq 1.3$, is a mistake and sets a dangerous precedent.

5.2.6 Consideration of All Available Data and Relevant Information

Based on the draft LLNA: BrdU-ELISA BRD, the Panel was asked whether all the relevant data identified in published or unpublished studies that employ this test method had been adequately considered, and if other comparative test method data that were not considered were available. If yes, the Panel was asked to suggest how to obtain such data. The Panel believed that all of the relevant data, with the exception of the interlaboratory reproducibility study, were presented and that they were not aware of any omissions.

5.3 Comments on the Draft ICCVAM Test Method Recommendations on the LLNA: BrdU-ELISA

5.3.1 Test Method Usefulness and Limitations

The Panel was asked whether the available data support the ICCVAM draft recommendations for the LLNA: BrdU-ELISA in terms of the proposed test method usefulness and limitations. The Panel agreed with the ICCVAM recommendation that the LLNA: BrdU-ELISA may be useful for identifying substances as potential skin sensitizers and non-sensitizers but that, at this time, more information and data are needed before a recommended use of the LLNA: BrdU-ELISA can be made. The Panel also stated that a detailed protocol is needed, in addition to sufficient quantitative data for a more comprehensive analysis based on a larger set of balanced reference substances with regard to physicochemical properties and sensitization potency, as well as an evaluation of interlaboratory reproducibility.

The Panel was asked whether the LLNA: BrdU-ELISA should be routinely recommended for hazard identification of skin-sensitizing substances in lieu of using guinea pig tests if restrictions on using radioactive materials *are* present, due to the fact that fewer animals might be used and because pain and distress would be avoided. The Panel stated that if the accuracy of the test method was at least similar to the traditional LLNA, the LLNA: BrdU-ELISA might be routinely recommended for hazard identification of skin-sensitizing substances in terms of reduction of animals and refinement of the pain and distress associated with guinea pig tests. Clearly, using the LLNA: BrdU-ELISA instead of the traditional LLNA or guinea pig test methods would also offer advantages for the environment due to the use of a non-radioactive probe chemical. However, the Panel stated that the accuracy of the current LLNA: BrdU-ELISA dataset at $SI \geq 3.0$ was inadequate and not equivalent to the traditional LLNA. The Panel also noted that if an $SI \geq 1.3$ was used because of its apparent increased accuracy, additional mice (over and above the number needed in the standard LLNA test) would apparently be needed (see **Table 5-1**). Thus, the Panel stated that reduction of animals would not be achieved. In this regard, the Panel noted that some quantification of the total animal use numbers would be useful as it is not clear whether the increased number of mice used would outweigh the avoidance of pain and distress in guinea pigs.

The Panel was asked whether the LLNA: BrdU-ELISA procedure or other valid and accepted non-radioactive method could be routinely recommended for hazard identification of skin sensitizing substances instead of the traditional LLNA if limitations in using radioactive materials *are not* present. The Panel stated that the LLNA: BrdU-ELISA procedure could not be routinely recommended for hazard identification of skin sensitizing substances instead of the traditional LLNA, because the accuracy of this test at $SI \geq 3.0$ was inadequate. In other words, the current dataset available for the LLNA: BrdU-ELISA did not predict the guinea pig or human outcomes as accurately as the traditional LLNA. Thus, the Panel acknowledged that there is the possibility that additional data might impact on the accuracy statistics and eliminate this concern. The Panel stated that factors that weigh on a decision of replacement of the LLNA with a non-radioactive method would include:

- Are more animals needed?
- Is the replacement test safer and less complex?

- Is the replacement test more efficient?
- Is the replacement test less costly?

The Panel stated that additional factors to consider might exist, but overall recommended that whether or not restrictions on radioactivity exist, a test that causes the least pain and uses the fewest number animals should be preferred, as long as adequate test method performance is maintained. Clearly, policy issues regarding restrictions on radioactivity should have no impact on this science-based conclusion.

The Panel was asked whether using a decision criterion of $SI \geq 1.3$ instead of $SI \geq 3.0$ resolved any concerns with respect to potential false positives or false negatives that may occur in this test method. The Panel was also asked for other suggestions for additional guidance or limitations that should be considered. The Panel stated that using a decision criterion of $SI \geq 1.3$ instead of $SI \geq 3.0$ would not itself resolve any concerns; more raw data are needed for a broader set of reference positive and negative sensitizers, including metals, mixtures, and aqueous solutions. The Panel also stated that a detailed protocol is needed, as is an evaluation of interlaboratory reproducibility. The Panel considered the current database to be inadequate, but based on the limited database, concluded that it might be more appropriate to use a statistically based decision criteria than a stimulation index.

5.3.2 Test Method Protocol

The ICCVAM draft recommendations state that the LLNA: BrdU-ELISA protocol should adhere to the ICCVAM LLNA protocol (ICCVAM 1999; Dean et al. 2001), except for measurement of lymphocyte proliferation. The Panel was asked whether the available data support the ICCVAM draft recommendations for the LLNA: BrdU-ELISA procedure in terms of the proposed test method standardized protocols. In general, the Panel agreed that the available data support the ICCVAM draft recommendations for this test method in terms of the standardized protocol. As stated previously, the Panel majority agreed with the ICCVAM recommendation that future studies should use five animals per dose group and collect individual animal data, as per the ICCVAM-recommended protocol. A minority opinion by Drs. Nathalie Alépée, Thomas Gebel, Dagmar Jírová, Raymond Pieters, and Michael Woolhiser stated that if laboratories were operating under OECD guidance (OECD 2002) and a reliable validation dataset had been generated, then pooled data from at least four animals could be considered acceptable. The Panel further noted that using an $SI < 3.0$ would require more animals to achieve adequate statistical power (**Table 5-1**) and therefore any considerations of reducing the SI to improve test method accuracy should include this point.

The Panel was asked whether the traditional LLNA limit dose procedure could be applied to the LLNA: BrdU-ELISA. The Panel stated that, if the LLNA: BrdU-ELISA was considered equivalent to the traditional LLNA, then it would be appropriate to apply the LLNA limit dose procedure to this test method. The Panel explained that, as in the case of the traditional LLNA, the protocol would be the same except for testing the maximum dose only, so applying the limit dose procedure would appear to have the same opportunity to reduce the number of animals needed to perform the test. However, using an $SI \geq 3.0$ would not be appropriate because of the associated low accuracy in identifying sensitizers.

5.3.3 Future Studies

The Panel was asked whether the available data support the ICCVAM draft recommendations for the LLNA: BrdU-ELISA in terms of the proposed future studies. The Panel stated that the proposed future studies were justified. The Panel concluded that it is important to consider non-radioactive methods because, in some laboratories, it is difficult or not permissible to use radioactivity. The Panel also stated that, if more data were available and there was less variability in this test method, it might warrant re-evaluation. The Panel concluded that more data are needed, especially for determination of the appropriate threshold value for the decision criterion, and that interlaboratory reproducibility should be also evaluated (which presumably will occur once the Japanese interlaboratory validation effort is complete).

5.3.4 Performance Standards

The LLNA: BrdU-ELISA protocol differs from the ICCVAM-recommended protocol for the traditional LLNA (ICCVAM 1999; Dean et al. 2001) in the method used to assess lymphocyte proliferation in the auricular lymph nodes. According to the proposed draft ICCVAM LLNA Performance Standards for the traditional LLNA, any change to the LLNA protocol other than the method used to assess lymphocyte proliferation is considered a “major” change. The Panel was asked whether protocol differences between the LLNA: BrdU-ELISA and the traditional LLNA should be considered only “minor” changes and therefore if the validity of this test method should be based on the draft ICCVAM LLNA Performance Standards. In general, the Panel agreed that the LLNA: BrdU-ELISA protocol differs only in the method used to assess lymphocyte proliferation. Thus, based on the current draft ICCVAM LLNA Performance Standards, it should be considered as having only “minor” changes and therefore the validity of this test method could be based only on the draft ICCVAM LLNA Performance Standards.

However, the Panel concluded also that the answer to this question might differ depending on what the draft ICCVAM LLNA Performance Standards ultimately constitutes as a “major” change, a “minor” change, or a different protocol altogether. The Panel further stated that, depending on the goal of the assay, these distinctions may not be relevant. Ultimately, if a test method is able to make the correct prediction with regard to the sensitization potential of a test substance, then the issue of “major” versus “minor” changes in the protocol should not apply.

The Panel was asked, even if the draft ICCVAM LLNA Performance Standards do not apply to the LLNA: BrdU-ELISA, what impact should the accuracy analysis based on eight of the 18 proposed required reference substances in the draft ICCVAM LLNA Performance Standards (only one false negative and no false positives) have on the overall evaluation of test method accuracy. The Panel concluded that the accuracy analysis based only on eight of the 18 proposed required reference substances had a significant impact on the evaluation of test method accuracy. The low number of experiments provided data that resulted in unacceptable test method performance.

The Panel was asked whether there were concerns that 4/4 sensitizers, for which EC3 data were available, had EC3 values that were outside of the proposed recommended 0.5x to 2.0x EC3 acceptability range developed based on the traditional LLNA. The Panel concluded that

the EC3 values outside the recommended 0.5x to 2.0x EC3 acceptability range raised concerns related to test reproducibility and reliability.²

The Panel was asked whether separate performance standards should be developed for the LLNA: BrdU-ELISA. The Panel concluded that separate performance standards for the LLNA: BrdU-ELISA were not needed because the test principles are identical to the traditional LLNA.

² During their public meeting on March 4-6, 2008, the Panel's discussion, conclusions and recommendations on the LLNA: BrdU-ELISA took place prior to the discussion, conclusions and recommendations on the draft ICCVAM LLNA performance standards. Following their discussion of the draft LLNA performance standards, the Panel concluded that that an evaluation of test method accuracy should be based on overall accuracy statistics when compared to the traditional LLNA, and not on a chemical-by-chemical match that is based on obtaining an EC3 value within a specified range of EC3 values.

6.0 Draft ICCVAM LLNA Performance Standards

6.1 Comments on the Proposed Purpose and Applicability

ICCVAM proposed that these performance standards should only be applicable to versions of the LLNA that incorporate “minor” modifications to the traditional LLNA. Currently, *this is limited to the use of non-radioactive reagents to measure lymphocyte proliferation*. It is considered essential that the modified LLNA should otherwise adhere to all other aspects of the traditional LLNA protocol, as defined by ICCVAM (1999) and Dean et al. (2001). This includes aspects such as: the sex and strain of mouse used, the number of mice per dose group, the timing and site of test article treatment, the duration between the last treatment and lymph node collection, the inclusion of concurrent negative and positive control groups, the measured endpoint (i.e., lymphocyte proliferation in the draining auricular lymph node), and the collection of data at the level of the individual mouse. The Panel was asked if they agreed that the use of non-radioactive reagents for measuring cell proliferation in the lymph nodes constitutes a “minor” modification to the traditional LLNA protocol.

The Panel noted that the draft ICCVAM LLNA Performance Standards are proposed for evaluating the acceptability of test methods that are functionally and mechanistically similar to the traditional LLNA (i.e., measuring the same biological effect), and understood that ICCVAM proposed that these performance standards should only be applicable to protocols that incorporate “minor” modifications to the traditional LLNA, as defined above. The Panel unanimously agreed that based on ICCVAM’s definition, the use of non-radioactive reagents for measuring cell proliferation is a “minor” modification of the traditional LLNA protocol. However, the Panel also agreed that other modifications may be considered “minor” and that a better strategy for the performance standards might be to define criteria that need to be satisfied to insure that the method is mechanistically and functionally similar (see criteria listed under essential test method components). Examples of potentially acceptable modifications identified by the Panel include sex, strain, the use of rats rather than mice, number of animals per group, and timing of test article treatment.

Regardless of the modification, the Panel stated that the modified test method should be designed to measure only the induction phase of the immune response. This is crucial, since the traditional LLNA is intended for hazard identification with the underlying principle that stimulation of cell proliferation in the draining lymph node suggests that sensitization (i.e., induction) is occurring. Using only the induction phase as the method to identify hazardous substances involves a short time frame, and reduces pain and distress in treated animals (i.e., no dermatitis response). Furthermore, the Panel stated that the performance standards should not imply that the traditional LLNA, or any alternative LLNA protocol, is capable of specifically distinguishing a type IV hypersensitivity reaction (as might be inferred from the text beginning with line 342 of the draft ICCVAM LLNA Performance Standards document). Therefore, reference to type IV hypersensitivity reaction should be removed from the document.

The Panel was asked if they considered it necessary that a modified LLNA keep the same decision criteria for distinguishing between sensitizers and non-sensitizers as the traditional LLNA (i.e., an $SI \geq 3.0$). The Panel considered it unnecessary for a modified LLNA to keep this same decision criteria as a different method for measuring cellular proliferation might

have better concordance with the human data at a SI different than 3.0. Thus, with any modified LLNA, the SI threshold defining a sensitizer would need to be established (i.e., it is important to consider if the results are biologically relevant to humans).

The Panel was asked if other procedural modifications could be identified as “minor”, based on the description in the draft ICCVAM LLNA Performance Standards document, and therefore could be evaluated for equivalence to the traditional LLNA using the proposed performance standards. The Panel reiterated that sex, strain, the use of rats rather than mice, animals per group, and timing of test article treatment are also potentially “minor” modifications. Furthermore, the proposed performance standards appear robust; therefore, regardless of the modification (i.e., “major” or “minor”), there is the same expectation for test method performance. Dr. James McDougal offered a minority opinion to express his concern about the potential impact that allowing alternative LLNA protocols with modifications other than the method by which lymphocyte proliferation was measured would have.

The Panel was asked if they considered the draft ICCVAM LLNA Performance Standards applicable to the LLNA limit dose procedure. The Panel noted that the current draft ICCVAM LLNA Performance Standards could be applicable to the LLNA limit dose procedure as long as it is recognized that this procedure can only be used for a yes/no hazard classification (i.e., an EC_t estimate is not feasible).

6.2 Comments on the Essential Test Method Components

The essential test method components are based on the ICCVAM-recommended protocol (ICCVAM 1999; Dean et al. 2001), which is the basis for the current EPA (2003) test guideline. There are some notable differences between these protocols and OECD TG 429 (OECD 2002) for the LLNA. The Panel was asked to comment on, when evaluations of non-radioactive versions of the traditional LLNA are conducted using these performance standards, whether it is necessary that the validation studies follow the ICCVAM-recommended protocol. The Panel indicated that ideally, there would be one globally recognized set of performance standards (ICCVAM, ECVAM, JaCVAM). However, when validating versions of the traditional LLNA where the only difference is in the use of a non-radioactive method to measure cell proliferation, the ICCVAM-recommended protocol should be used. If more extensive changes to the protocol are being considered, the following requirements should be considered during modifications of the LLNA:

- Application of the test substance should be to the skin, with sampling of the lymph nodes draining that site.
- Cell proliferation should be measured in the draining lymph node.
- No skin reaction should be present, since presence of a skin reaction might indicate the onset of the elicitation phase of skin sensitization.
- Data should be collected at the level of the individual animal to allow for an estimate of the variance within control and treatment groups. Using this variance, a power analysis needs to be conducted to demonstrate that the modified method is utilizing a sufficient number of animals per treatment group to permit hazard identification with at least 95% power.

- If dose response information is needed, there should be an adequate number of dose groups ($n \geq 3$) with which to adequately characterize the dose response for a given test substance.

The Panel was asked to comment on whether validation studies should include a concurrent positive control with each test substance and if so, whether the concurrent testing of the positive control and test substance should be conducted in the same vehicle or if different vehicles were acceptable. The Panel noted that a concurrent positive control should be included in each validation study to ensure that the test system was operating as expected and technical errors were not occurring. A concurrent positive control would be especially useful when an unknown test material was being tested or when a laboratory was collecting a dataset to serve as historical control data. However, if a known sensitizer was being tested, a concurrent positive control might not be needed, thus reducing animal use. Finally, the Panel concluded that the positive control should be tested in the same vehicle as the test substance. Using a different vehicle for the positive control would require an additional set of vehicle control animals.

The Panel was also asked whether the validation studies should use a minimum of five animals per dose group and collect lymph node data from individual animals. The Panel commented that until sufficient data were collected to enable a reliable power calculation to be conducted to determine the optimal number of animals per dose group, at least five animals per dose group should be used. The Panel also agreed that when validating a modified LLNA protocol, lymph node proliferation should be evaluated at the level of the individual animal within each dose group. Variance is only measurable if lymph nodes from individual animals are assessed. If the variability within a dose group of a modified LLNA protocol was substantially less than the traditional LLNA, reducing the number of animals per dose group might yield similar results. A minority opinion by Drs. Nathalie Alépée, Thomas Gebel, Dagmar Jírová, Raymond Pieters, and Michael Woolhiser stated that if laboratories were operating under OECD guidance (OECD 2002) and a reliable validation dataset had been generated, then pooled data from at least four animals could be considered acceptable.

6.3 Comments on the Proposed Reference Substances

The Panel was asked if they agreed with the selection and prioritization criteria used to select the performance standards reference substances. The Panel noted that the rationale for selection of the reference substances included in the draft ICCVAM LLNA Performance Standards was well documented (taking into account the physicochemical characteristics, the purity, the stability, the quality of the *in vivo* data, and the chemical classes covered). The substances also appeared to be distributed over a wide range of EC3 values. However, the available database for some of the substances was insufficient. Among the 13 sensitizers in the “required” list, only five appear to have a robust database (i.e., have been tested in at least three independent studies). Thus, consideration should be given to revising the list of substances and/or making the data for the substances on the current list more robust. Ideally, the reference list should be based only on substances with robust data for LLNA, human, and guinea pig tests.

The rationale for the number of substances included on the "required" list of substances (n=18) was provided in the draft ICCVAM LLNA Performance Standards. In addition, there

were four additional substances that were described as problematic in the traditional LLNA (i.e., false negatives and false positives). The Panel was asked if they considered 18 “required” substances to be an adequate number upon which to evaluate the performance of non-radioactive LLNA test methods, where the only protocol modification is the method for assessing cell proliferation in the auricular lymph nodes, and if not, how many reference chemicals should be tested. The Panel commented that ideally, one would like to be able to demonstrate that an assay is equivalent to the traditional LLNA. However, with the small number of reference substances available, establishing equivalence will be extremely difficult. Therefore, the Panel recommended that, for use in hazard identification, a modified method should be evaluated with all 22 substances (including false negatives and false positives) and accuracy statistics calculated. To the extent possible, rationale for any discordant results should be provided, but the most potent sensitizers (e.g., DNCB) should always be identifiable. There also should be considerable weight given to the balance between animal welfare and human safety when considering the adequacy of test method accuracy.

The panel considered it noteworthy that 19 of the 22 substances on the draft ICCVAM list are in common with the ECVAM performance standards list. The Panel also considered it important that substances be coded during validation studies.

It is also relevant to note that the Panel discussed the value of GLP procedures on several occasions during the meeting. In each instance, the Panel agreed that data collected under GLP conditions would be greatly preferred, particularly for reasons of data quality and the associated reliability of any interpretations. However, they noted that GLP compliance would not be considered a requirement that would automatically exclude data from consideration. The Panel concluded that other factors could be used to identify high quality data. Examples would include published in a peer-reviewed journal or obtained from a study conducted in a laboratory that routinely conducts GLP studies. Data generated under non-GLP conditions would be subject to a critical quality review, and as such the Panel considered it important to obtain the original records in order to confirm the reported data.

The Panel was asked if they considered the types of substances included in the reference substance list, with regard to relative sensitization potency, physicochemical characteristics, and vehicles, to be representative of the overall diversity of substances that are likely to be tested for skin sensitization. The Panel concluded that although the list should not be considered all-inclusive, it was sufficiently representative.

The Panel was also asked if there were other types of information relevant to skin sensitization that should be considered in order to demonstrate an adequately diverse reference list. The Panel commented that identifying concentrations of each of the substances that are known to cause excessive local irritation or overt systemic toxicity would be useful.

The Panel was asked if there were other substances that they considered to be more appropriate for assessing the sensitivity (i.e., ability of the test method to correctly identify sensitizing substances) and specificity (i.e., ability of the test method to correctly identify non-sensitizing substances) of non-radioactive LLNA test methods, and for which there is available LLNA, guinea pig, and human data. The Panel could not identify such substances given the time frame for consideration but reiterated that substances in the reference list should have robust data.

A subset of "discordant chemicals" (i.e., false negative or two false positive compared to guinea pig tests or human data) were included as "optional" substances that could be studied to evaluate if the proposed modifications might provide improved performance relative to the traditional LLNA. The Panel was asked to comment on the appropriateness of including these specific substances in the reference list, whether they should be required, whether different substances should be included, and if more false negative/positive substances should be tested. As mentioned previously, the Panel commented that it was appropriate to include such substances in the reference list and that they should be required and evaluated during the validation of alternative LLNA assays that are functionally and mechanistically similar to the traditional LLNA assay. The Panel noted that the substances that were considered discordant depended on the species to which comparisons are made (i.e., LLNA vs. guinea pig or LLNA vs. human). Still, since the "discordant compounds" were false negatives or positives in the traditional LLNA, they would provide an opportunity to determine if modifications to the traditional LLNA may even have increased accuracy.

Finally, the Panel was asked if "correct" results with these discordant chemicals would be sufficient to consider the alternative test method to be more predictive of skin sensitization than the traditional LLNA. The Panel concluded that correct results with the "discordant chemicals" would not be sufficient to consider the alternative test method to be more predictive of skin sensitization, but it could provide supporting evidence to indicate further testing with additional compounds would be of value.

6.4 Comments on the Test Method Accuracy Standards

The draft ICCVAM LLNA Performance Standards state that the non-radioactive proposed LLNA test method should exactly match the accuracy of the traditional LLNA when evaluated with the minimum set of 18 reference substances. The Panel was asked if they agreed that test method accuracy should be based on a chemical-by-chemical match with regard to identifying the chemicals as sensitizers or non-sensitizers. The Panel commented that although an assay that is able to predict the same hazard classification for the reference substances as the traditional LLNA is desired, with the small number of reference substances available (n=18), clearly establishing equivalence will be extremely difficult. Furthermore, even with this small number, there is a statistical multiple comparisons³ problem because more than one chemical is being tested. The likelihood that a modified LLNA will fail to demonstrate equivalence to the traditional LLNA will increase with the number of chemicals that must be identified correctly. A statistical measure of concordance should be calculated so that accuracies can be compared between methods.

The Panel reiterated their recommendation that, for use in hazard identification, a modified test method should be evaluated with all 22 substances (including false negatives and false positives) and accuracy statistics calculated. A statistical measure of concordance should be calculated so that accuracies can be compared between methods. To the extent possible,

³ When multiple experiments are conducted and multiple observations, comparisons or hypothesis tests are conducted, the chance of observing rare events increases. Suppose, for example, that an interval is established such that 5% of observations from a particular population of data are outside that interval. Then if k independent experiments generate data from this population (e.g., a standard normal distribution), the chances that all 20 results will lie inside the interval is $(1.0 - 0.05)^k$ (N. Flournoy, personal communication).

rationale for any discordant results should be provided. However, the most potent sensitizers (e.g., DNCB) should always be identifiable. Considerable weight should be given to the balance between animal welfare and human safety when considering the adequacy of test method accuracy.

The draft ICCVAM LLNA Performance Standards recommend that, for each sensitizer, the threshold concentration that induces a positive SI response should be within 0.5x to 2.0x of the concentration obtained for the EC3 in the traditional LLNA. As described in the draft ICCVAM LLNA Performance Standards, statistical approaches have been used in an attempt to identify an appropriate range, but these calculated ranges do not appear to be the most practical. The Panel was asked to comment on the appropriateness of using this criterion to judge the equivalency of a non-radioactive version of the traditional LLNA and, if this approach was not acceptable, to suggest an alternative along with the basis for this approach. The Panel commented that the usefulness and limitations of the traditional LLNA for determining relative potency have not been definitively established, and therefore equivalence should not be based strictly on potency. Furthermore, the current database does not support the inclusion of EC3 values as a component of the accuracy evaluation. The range of 0.5x to 2x EC3 value suggested in the draft ICCVAM LLNA Performance Standards document are based on the experience with a range of known skin sensitizers tested in the standard LLNA. However, based on the available data provided, the 0.5x to 2.0x EC3 range may be too restrictive *if a strict interpretation of equivalence is applied*. The chances of a failure to achieve an EC3 within this range would vary from chemical to chemical depending upon the inherent underlying variability and robustness in the estimation of the EC3. For those chemicals for which the EC3 can be accurately estimated, the failure rate may be close to zero. For other, more variable chemicals, perhaps with fewer data points, the failure rate for a single chemical will be much higher.

The Panel reiterated their concern with regard to EC3 values (i.e., the statistical multiple comparisons problem). The likelihood that a modified LLNA will fail to demonstrate equivalence to the traditional LLNA will increase with the number of chemicals tested, the extent to which the new test must obtain the same EC3 value, and how independent the results are for different chemicals in the same lab.

For five of the 13 sensitizers on the draft ICCVAM reference substances list, the reference EC3 value was based on a single LLNA study (Table C1 of the draft ICCVAM LLNA Performance Standards). The Panel was asked to comment on the appropriateness of including such chemicals in the list of recommended reference substances and whether or not the 0.5x to 2.0x criteria should be applied to such substances. The Panel concluded that the appropriateness of the 0.5x to 2.0x EC_t range had not been adequately justified. It was inappropriate to include chemicals represented by only one LLNA study on a list of recommended reference substances, as there was insufficient data by which to calculate a robust mean EC_t value. Thus, those compounds should either (1) be exchanged for compounds with sufficient EC3 data (i.e., have been tested in at least three independent studies using the same solvent), or (2) retained but not considered to be part of the EC_t criterion until such data has been collected.

6.5 Comments on the Test Method Reliability Standards

The draft ICCVAM LLNA Performance Standards state that acceptable intralaboratory reproducibility will be indicated by a laboratory obtaining, in each of four independent experiments conducted with at least one week between each experiment, EC_t values (the estimated concentration needed to produce an SI of a defined threshold [e.g., EC₃]) for HCA that are generally within 0.5x to 2.0x (i.e., 5% to 20%) of the historical mean EC₃ concentration (10%) for this substance, based on existing available traditional LLNA data. The Panel was asked if they considered four repeat experiments to be adequate. The Panel concluded that four experiments would be adequate, as requiring four independent experiments is similar to the original LLNA submission, as is a one-week interval between experiments. Therefore, these requirements were appropriate for a comparison of modified methods to the traditional LLNA. However, it would be useful to have this number evaluated statistically (see **Section 6.7**).

The Panel was asked if they considered testing HCA adequate for demonstrating intralaboratory reproducibility and if not, which substance(s) should be tested. The Panel concluded that HCA testing would be adequate for demonstrating intralaboratory reproducibility and would allow an effective comparison to the traditional LLNA.

The Panel was asked to comment on whether the required one-week interval between independent tests was adequate and/or appropriate. The Panel concluded that the minimum one-week interval seemed logical and that the more important clarifying information might be the elements that define independent tests (e.g., different animal shipment, different reagents, different operator, blind testing).

The Panel was asked to comment on the appropriateness of the criteria for acceptability (generally within 0.5x to 2.0x EC₃ for HCA), or to describe another criteria and explain the basis for their recommendation. The Panel concluded that the criteria for acceptability appeared to be appropriate because the statistical multiple comparisons issue does not exist. However, given that there is so much data and experience with HCA and the fact that only one compound is being tested (not 18), it is reasonable to evaluate reproducibility using the mean \pm 3 standard deviations rather than the 0.5x to 2.0x EC₃ range to account for a single comparison (see **Section 6.7** regarding data transformation recommendations). The Panel noted that historical control data using HCA in the same vehicle could be used to demonstrate adequate intralaboratory reproducibility.

The draft ICCVAM LLNA Performance Standards state that acceptable interlaboratory reproducibility will be indicated by each of three laboratories obtaining EC_t values for HCA and DNCB from a single experiment that are generally within 0.5x to 2.0x (5% to 20% and 0.025 to 0.1%, respectively) of the mean historical EC₃ concentration (10% and 0.05%, respectively) obtained for these two substances in the traditional LLNA. The Panel was asked if they considered the single experiment per substance in each laboratory to be adequate. The Panel concluded that, considering the overall validation plan for a given laboratory, multiple experiments (n=3) within each laboratory should be conducted.

The Panel was asked if they considered testing HCA and DNCB to be adequate for demonstrating interlaboratory reproducibility and if not, which substance(s) should be tested. The Panel concluded that, since there is a great deal of data and experience with HCA and

DNCB, and many laboratories have successfully worked with them in the traditional LLNA, they should be considered adequate for this purpose.

The Panel was asked if they considered the criteria for acceptability to be appropriate. The Panel concluded that the criteria for acceptability (i.e., generally within 0.5x to 2.0x EC_t for HCA and DNCB) appeared to be appropriate because the statistical multiple comparisons problem was relatively minor given that only two substances are being tested. However, given that there is so much data and experience with HCA and DNCB and the fact that two compounds are being tested (not 18), it is reasonable to evaluate reproducibility using the mean \pm 4.5 standard deviations to account for statistical multiple comparisons (see **Section 6.7** regarding data transformation recommendations). The Panel also noted that historical control data using HCA and DNCB in the same respective vehicle could be used to demonstrate adequate interlaboratory reproducibility.

6.6 Summary

The Panel was asked what criteria should be used to evaluate the equivalence of a radioactive or non-radioactive LLNA method to the traditional LLNA, if one were proposed with a “major” change, as defined in the draft ICCVAM LLNA Performance Standards (e.g., different mouse strain or use of male mice, change in the schedule for test article administration, change in schedule for lymph node excision, etc.). The Panel commented that the idea of what is a “major” and a “minor” change should be re-considered (refer to Question 2 regarding essential test components). The final version of the performance standards should be adequate to evaluate any protocol modifications.

The Panel was asked if a new set of performance standards would be required for a modified version of the LLNA that incorporated one or more “major” protocol changes. Based on the above response, the Panel concluded that a new set would not be required.

The Panel was asked to comment on how many reference substances might be considered adequate for evaluating the validity of a modified version of the LLNA with a “major” protocol change; specifically, if the 18 minimum reference substances in the draft ICCVAM LLNA Performance Standards would be sufficient. The Panel concluded that additional substances should not be considered necessary. However, since eight of the proposed sensitizers had limited data (i.e., EC₃ values based on ≤ 2 LLNA studies), other substances with more robust data should be considered as replacements. Furthermore, if the goal is to evaluate a specific applicability domain, additional test substances might be needed.

The Panel was asked to comment, regardless of the number of reference substances, whether the alternative LLNA with a “major” change should be required to obtain the same “call” (and potency for sensitizers) as the traditional LLNA for the 18 minimum reference substances in the draft ICCVAM LLNA Performance Standards. The Panel reiterated that an assay that is equivalent to the traditional LLNA is desired, but with the small number of reference substances available, clearly establishing equivalence will be extremely difficult. They also reiterated their concern regarding the statistical multiple comparisons problem.

For use in hazard identification, a proposed modified LLNA should be evaluated with all 22 substances (including false negatives and false positives) and accuracy statistics calculated so that accuracies can be compared between the modified test method and the traditional LLNA. To the extent possible, rationale for any discordant results should be provided. However, the

most potent sensitizers (e.g., DNCB) should always be identifiable. Considerable weight should be given to the balance between animal welfare and human safety when considering the adequacy of test method accuracy.

The Panel was asked to identify any additional specific substances that should be used. The Panel concluded that while additional substances should not be needed, it would be useful to identify replacements for the eight proposed sensitizers with limited test data. If the goal is to evaluate a specific applicability domain, additional test substances might be needed.

6.7 Additional Statistical Comments

During the evaluation of the draft ICCVAM LLNA Performance Standards, the Panel noted a number of statistical issues that should be addressed. They suggested that in order to achieve a normal distribution of the data and to reduce differences between groups, a suitable variance stabilizing transformation (e.g., log transformation, square root transformation) should be applied in all statistical analyses and in reporting summary standard deviations. The Panel also suggested that there should be a more rigorous evaluation of what would be considered an appropriate range of EC_t values to include as a requirement. This would be a statistical evaluation that takes into consideration the variability of EC_t values generated among the sensitizers included on the performance standards reference substances list and the statistical multiple comparisons problem and the fact that sample sizes that are less than 30 invalidate statistics based on the normal distribution (Young 2007).

Furthermore, bioequivalence models have been developed (Berger and Hsu 1996) and should be applied to the LLNA. Probability values can be used as descriptive statistics and as such provide a summary measure of weight-of-evidence that would be useful for comparison of performance standards across test methods. In this context, it would be informative to have statistical tests of data generated for these purposes. A test of concordance for measuring the accuracy of classification should be done.

Intralaboratory tests should include analysis of variance (ANOVA)-like tests with a test for no trend, with the null hypothesis being that there is a difference and the alternative being that the difference is bioequivalent. Interlaboratory tests should include ANOVA-like tests with the null hypothesis being that there is a difference and the alternative hypothesis being that the difference is bioequivalent. The reliability tests require “bioequivalence” to be defined (i.e., what is acceptable to be considered equivalent).

It is not known whether these specific statistical tests can be identified in the literature or if they need to be developed. If they do need to be developed, this should be given a priority. Prior to running reliability studies, these statistical methods should be used to determine the appropriate number of substances and the number of times each substance needs to be tested within and among laboratories in the study design (see also ISO 5725 [ISO 1994] and ASTM Standard E691 [ASTM 2005]). The power for the traditional LLNA should be established for comparison purposes.

[This Page Intentionally Left Blank]

7.0 Use of the LLNA for Potency Determinations

7.1 Comments on the Draft BRD for Completeness, Errors, and Omissions

The Panel was asked if there were any errors in the draft BRD on the use of the LLNA for potency determinations that should be corrected, if omissions of existing relevant data had been identified, or if there was additional information that should be included. The Panel noted alternative analyses that would better help evaluate the use of the traditional LLNA for skin sensitization potency (see the discussion of the categorization scheme in **Section 7.2** and the discussion of future studies in **Section 7.3**).

7.2 Comments on the Validation Status of the Traditional LLNA to Determine Skin Sensitization Potency

7.2.1 Substances Used for the Validation Studies

The Panel was asked to consider whether the validation status of the traditional LLNA for potency categorization (i.e., “strong” vs. “weak” sensitizers) has been adequately characterized, and if the traditional LLNA is sufficiently accurate and reliable to be used as a stand-alone assay for characterizing the potency of sensitizing substances, based on the comparison to human and guinea pig responses. The Panel agreed that the LLNA database of 170 substances with comparative guinea pig (i.e., Guinea Pig Maximization Test or Buehler Test) and/or human data (i.e., Human Maximization Test [HMT] and/or Human Repeat Insult Patch Test [HRIPT], but not human clinical observations) is sufficient in number and well balanced for this evaluation. The database included 112 substances (97 sensitizers, 15 non-sensitizers) with comparative human data and 105 substances (52 sensitizers, 53 non-sensitizers) with comparative guinea pig data. Known contact sensitizers of public health concern from various chemical groups are included. The Panel further agreed that these substances were representative of a sufficient range of chemical classes and physical chemical properties so that it would be applicable to the types of chemicals and products typically tested for skin sensitization potential.

While coding of chemicals to reduce bias is recommended for validation studies, this evaluation was based on a retrospective evaluation of existing data, most of which were generated using chemicals that were not coded. The Panel was asked whether the lack of coding of test substances adversely impacted or biased the current evaluation. Given the nature of the studies (i.e., the testing was not conducted to demonstrate the ability of the LLNA to be used for potency characterization), the Panel stated that the lack of coding likely had no impact on the current evaluation.

For some substances tested for sensitization using the traditional LLNA, it was not possible to determine whether the data were generated using pooled or individual animal lymph node samples within a dose group (the former allowed in OECD TG 429 [OECD 2002]; the latter as recommended in the ICCVAM 2001 protocol and required in the EPA 2003 skin sensitization test guideline). Cockshott et al. (2006) reported that using individual animal data allowed for technical problems during an experiment and outlier animals within a dose group to be identified. Considering this, the Panel was asked whether the analysis of the performance of the traditional LLNA for potency determinations should be limited to data from studies that can be confirmed as using individual animal data collection procedures.

A majority of the Panel agreed that, ideally, future traditional LLNA potency determinations should be based on data from studies that use individual data collection procedures, as this would allow for the identification of outliers that might skew the average group stimulation index. A minority opinion by Drs. Nathalie Alépée, Thomas Gebel, Dagmar Jírová, Raymond Pieters, and Michael Woolhiser stated that if laboratories were operating under OECD guidance (OECD 2002) and a reliable validation dataset had been generated, then pooled data from at least four animals could be considered acceptable.

7.2.2 Test Method Accuracy

The Panel was further asked what impact the inclusion of pooled animal data might have on the accuracy analysis included in Section 6.0 of the draft ICCVAM LLNA potency BRD. With regard to this retrospective dataset, the Panel agreed that pooled data should not be excluded from the current analysis to assess potency determinations for the traditional LLNA. The Panel stated that it is impossible to assess the impact of using pooled data without a separate analysis of the ability of the traditional LLNA to be used for characterizing skin sensitization potency using pooled vs. individual data, which the Panel recommended be done (see the discussion of future studies in **Section 7.3**).

A minority opinion from Dr. Dagmar Jírová stated that, since OECD TG 429 (OECD 2002) allows the use of both pooled and individual animal data, the analysis that includes both types of data is appropriate. Even with the diversity of data sources (the vehicle is not known for 43% of substances tested in the traditional LLNA; human data were obtained by different, even undefined methods, etc.), the outcome of the evaluation was good, which documents the strength and robustness of the traditional LLNA.

The Panel was asked whether the correct classification, as well as the over- and under-classification, rates of the traditional LLNA for sensitization potency determinations had been adequately compared and appropriately evaluated based on the corresponding human and guinea pig data (refer also to Section 6.0 of the draft ICCVAM LLNA potency BRD). The Panel agreed that the two approaches used in the draft BRD for analyzing the ability of the traditional LLNA to discriminate between strong and weak skin sensitizers were appropriate and correct. In these two approaches, the traditional LLNA was evaluated, after identifying the optimal EC3 value, for its ability to correctly classify strong and weak sensitizers as defined by human or guinea pig threshold values based on: (1) sensitizers only, and (2) sensitizers combined with false positives, false negatives, and non-sensitizers.

A minority opinion from Dr. Howard Maibach stated that the relevance of the traditional LLNA to human clinical observations has not been sufficiently determined and should be.

The accuracy analysis (see Section 6.0 of draft ICCVAM BRD) focuses on a proposed two-level categorization scheme (weak sensitizers vs. strong sensitizers) for both human and guinea pig data. The Panel was asked whether this was an appropriate categorization scheme, or if other categorization schemes should be considered. The Panel agreed that the two-level categorization scheme was appropriate, especially considering the fact that, for human situations, risk assessment should be performed, and therefore more categories are not needed. Even a weak sensitizer under heavy exposure and individual circumstances may reach a comparable risk level as a strong sensitizer under conditions of low exposure.

A minority opinion from Drs. Raymond Pieters and Michael Woolhiser recommended the addition of at least a moderate category since certain compounds will always be on the border between weak and strong. Dr. Pieters specifically recommended the categorization scheme of Kimber et al. (2003), which is based on five categories if non-sensitizers are included.

Of the two human threshold concentrations that are proposed in this two-category categorization scheme (i.e., $<250 \mu\text{g}/\text{cm}^2$ or $<500 \mu\text{g}/\text{cm}^2$), the Panel was asked which threshold was the most appropriate for categorizing sensitizing substances as strong vs. weak for humans, or if another threshold was more appropriate for this purpose. The Panel noted that this validation was based on comparison to guinea pig and HMT/HRIPT information. These data relate only to induction and do not permit an assessment of risk in humans for elicitation.

For the data provided, the Panel concluded that the best results were obtained using the decision criterion of $250 \mu\text{g}/\text{cm}^2$ and the corresponding optimal traditional LLNA EC3 value of 9.4%. Using this cut-off when traditional LLNA false negative and false positive substances are included in the analysis, in addition to sensitizers in both the traditional LLNA and in humans using the HRIPT and/or HMT, correct classification of strong sensitizers was 79% and underclassification was 21%. Underclassification of substances in this context means classification as weak instead of strong sensitizers (i.e., they are not missed as sensitizers regarding the labeling and safety of consumers). The Panel stated that more data are needed to determine if another threshold is more appropriate.

When the potency categorization analysis was based on sensitizers only, the guinea pig tests predicted weak sensitizers with higher accuracy than did the LLNA (89% vs. 75% for the $250 \mu\text{g}/\text{cm}^2$ cutoff and 83% vs. 60% for the $500 \mu\text{g}/\text{cm}^2$ cutoff), which is logical because the guinea pig test methodology involves all phases of the sensitization process and usually involves adjuvants. However, the guinea pig tests were less accurate for the prediction of strong sensitizers compared to LLNA (48% vs. 71% for the $250 \mu\text{g}/\text{cm}^2$ cutoff and 42% vs. 63% for the $500 \mu\text{g}/\text{cm}^2$ cutoff), which represents a higher risk for consumers. For the protection of public health, it is more important to correctly identify strong sensitizers than weak sensitizers.

The Panel was asked whether the draft BRD adequately characterized the usefulness and limitations of the LLNA for potency categorizations. If not, the Panel was asked what additions or changes should be made to the description of usefulness and limitations in the draft BRD. The Panel stated that additional evaluations should be conducted to determine the impact on potency categorization if the human threshold data are evaluated differently (e.g., alternative lowest observed effect level [LOEL] safety factors other than 10, using LOEL data only, using no observed effect level [NOEL] data only), and if this might improve the correlation between the LLNA and the human results. According to the Panel, the approach of directly comparing the LOEL values without using a safety factor compares values of similar significance in humans and in the LLNA. In other words, the LOEL in humans describes the threshold induction area dose in humans and the EC3 value in the traditional LLNA is the threshold induction area dose and thus could be the analogous value to the human LOEL. The Panel further stated that traditional LLNA tests based on pooled or individual lymph nodes for a dose group should be evaluated independently to assess the

impact of using pooled data on the accuracy analysis for skin sensitization potency. Finally, the Panel stated that the effect of different vehicles should be recognized as a limitation in the data analysis given the demonstrated variability of results.

7.2.3 Test Method Reliability

The Panel was asked whether the reliability (e.g., intralaboratory repeatability, intra- and inter-laboratory reproducibility) of the traditional LLNA for potency determinations had been adequately evaluated. If not, the Panel was asked what other analyses should be performed. Similar to their recommendations for test method accuracy, the Panel stated that additional evaluations of reliability should be conducted based on using different approaches for human threshold data (e.g., using alternative LOEL safety factors other than 10, using LOEL data only, using NOEL data only). The Panel further stated that the reliability of LLNA based on using pooled or individual animal data should be evaluated independently. Finally, the Panel stated that the effect of different vehicles should be recognized as a limitation in the data analysis, as a source of increased variability.

7.2.4 Data Quality

It was not possible to determine whether or not all studies included in the draft LLNA potency BRD had been conducted in accordance with GLP guidelines, nor was it possible to obtain the results of GLP audits for all studies determined to be GLP-compliant. The Panel was asked to discuss what impact this might have on the evaluation of the LLNA for potency determinations and whether any of the non-GLP studies should be excluded from the analyses. The Panel concluded that it was important to note if the data were obtained from studies conducted according to GLP guidelines, as ideally this should be the case. However, the Panel concluded that data from studies that could not be confirmed as being GLP-compliant, but that were from peer-reviewed literature or other sources with high-quality laboratory management practices were still appropriate to include in this retrospective analysis.

As described in the draft BRD, original records for some of the non-GLP studies included in this evaluation could not be obtained. As a result, an independent audit could not be conducted to confirm that the reported data is the same as the data recorded in laboratory notebooks. Considering this, the Panel was asked whether the results of these studies (all of which are currently included) be excluded from any of the performance analyses. The Panel considered the data to have been generated by repeatedly published and reliable laboratories and therefore did not question the adequacy/quality of the retrospective data analysis. Thus, although data should be checked when available, exclusion of data was not deemed necessary, in this case.

7.2.5 Consideration of All Available Data and Relevant Information

Based on the draft BRD, the Panel was asked whether all the relevant data identified in published or unpublished studies conducted using the traditional LLNA had been adequately considered. If not, the Panel was asked what other studies should be considered. The Panel recommended that the LOELs from Akkan et al. (2003) be used instead of the DSA₀₅ values from Schneider and Akkan (2004) in all of the potency analyses. A minority opinion by Dr. Thomas Gebel stated that it was acceptable to use the DSA₀₅ values from Akkan et al. (2003) as LOEL values in the evaluation. Dr. Gebel mentioned that the DSA₀₅ value is a LOEL

value adjusted to 5% incidence of induction. Akkan et al. (2003) used the DSA_{05} value to correct for different human studies leading to different inductions. Dr. Gebel further stated that as the DSA_{05} is corrected for an induction rate of 5%, it would be better to compare with the traditional LLNA EC3 than to use the default uncorrected LOEL.

7.3 Comments on the Draft ICCVAM Test Method Recommendations for the Use of the LLNA for Potency Determination

7.3.1 Test Method Usefulness and Limitations

With regard to the use of the LLNA for potency categorization (i.e., strong vs. weak sensitizers), the ICCVAM draft recommendation is that the traditional LLNA should not be considered as a stand-alone test method for predicting sensitization potency, but must instead be used as part of a weight-of-evidence evaluation to discriminate between strong and weak sensitizers. This is based on the fact that, although there is a significant positive correlation between traditional LLNA EC3 values and human sensitization threshold doses, this correlation is not strong [see detailed discussion in the draft ICCVAM recommendations]. The Panel agreed that the traditional LLNA should not be considered a stand-alone assay for categorization of skin sensitization potency, but it could be used in a weight-of-evidence evaluation (e.g., along with quantitative structure-activity relationship [QSAR], peptide reactivity, human evidence) to discriminate between strong and weak sensitizers. The Panel further stated that there are additional studies proposed that may provide a better correlation and improve prediction of potency categorization (see the discussion of future studies below).

A minority opinion from Drs. Thomas Gebel and Dagmar Jírová stated that there is a significant positive correlation between EC3 values and human threshold values. It is likely that limitations in estimating human threshold values and the inclusion of human NOEL values in the current evaluation contributed negatively to the resulting R^2 value of 0.405 (when LLNA EC3 data vs. human threshold data were compared, see Table 6-2 of the draft ICCVAM BRD). Thus, the R^2 value may improve when the additional analyses that have been suggested by the Panel are conducted.

The Panel stated that the effect of different vehicles should be recognized as a limitation in the data analysis and a likely source of within and between laboratory variability.

7.3.2 Test Method Protocol

The Panel was asked whether the ICCVAM-recommended LLNA protocol (ICCVAM 1999; EPA 2003) should be used when generating data that will or might be considered for sensitization potency categorization decisions. The Panel agreed that this protocol should be used. A minority opinion by Drs. Nathalie Alépée, Thomas Gebel, Dagmar Jírová, Raymond Pieters, and Michael Woolhiser stated that if laboratories were operating under OECD guidance (OECD 2002) and a reliable validation dataset had been generated, then pooled data from at least four animals could be considered acceptable.

The Panel was asked whether the relevant testing guidelines for the traditional LLNA should be updated to include the calculation of an EC3 value. The Panel agreed with this recommendation. The calculation of an EC3 value is briefly described in the draft ICCVAM

LLNA Performance Standards for specific situations with references to Basketter et al. (2000) and Ryan et al. (2007).

7.3.3 Future Studies

The Panel was asked whether the available data support the ICCVAM draft recommendations for the traditional LLNA in terms of the proposed future studies. The Panel agreed and concluded that more data are needed to determine the optimal threshold in humans for distinguishing between strong and weak sensitizers. However, the Panel discouraged conducting new animal studies unless it was likely that results from such studies would lead to an overall reduction in animal use. The Panel stated further that the traditional LLNA appears to be a robust rodent assay for the quantification of the induction of cell-mediated immunity. Thus, use of the traditional LLNA for potency determination can be used in conjunction with QSAR information, guinea pig assays, HRIPT/HMT, and the quantitative data of elicitation and frequency of positive response in humans in a weight-of-evidence approach. The Panel further stated that additional evaluations should be conducted to determine the impact on potency categorization if the human threshold data are evaluated differently (e.g., alternative LOEL safety factors other than 10, using LOEL data only, using NOEL data only). This might improve the correlation between LLNA and human data. The Panel further stated that LLNA tests based on pooled or individual animal data should be evaluated independently to assess the impact of using pooled data on the accuracy for determining skin sensitization potency.

The Panel recommended a statistical analysis to determine where an appropriate cutoff value between weak or strong sensitizers might be best defined for traditional LLNA data. For example, receiver operating characteristic curves could be used to identify the optimum cutoff for determining the difference between weak and strong sensitizers.

Finally, the Panel stated that the effect of different vehicles should be recognized as a limitation in the current data analysis, that this was a source of variability within and between laboratories, and that its impact should be considered in future analyses.

8.0 References

- Akkan Z, Kalberlah F, Oltmanns J, Schneider K. 2004. Beurteilung der Wirkstärke hautsensibilisierender Chemikalien anhand des Local Lymph Node Assay (LLNA) [in German]. Bremerhaven, Germany: Federal Institute for Occupational Safety and Health.
- ASTM. 2005. ASTM Standard E691-05 Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method. In: Annual Book of ASTM Standards. West Conshohocken: ASTM International. www.astm.org.
- 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* 42:344-348.
- Berger RL, Hsu JC. 1996. Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Stat Sci* 11:283-319.
- Cockshott A, Evans P, Ryan CA, Gerberick GF, Betts CJ, Dearman RJ, et al. 2006. The local lymph node assay in practice: a current regulatory perspective. *Hum Exp Toxicol* 25:387-394.
- 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* 34:258-273.
- De Jong WH, van Och FM, Den Hartog Jager CF, Spiekstra SW, Slob W, Vandebriel RJ, et al. 2002. Ranking of allergenic potency of rubber chemicals in a modified local lymph node assay. *Toxicol Sci* 66:226-232.
- EPA. 2003. Health Effects Test Guideline OPPTS 870.2600. Skin Sensitization. Research Triangle Park, NC: Environmental Protection Agency.
- Gaspari AA. 2007. Mechanism of action and other potential roles of an immune response modifier. *Cutis* 79(4 SUPPL.):36-45.
- 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* 50:274-288.
- Gerberick GF, Vassallo JD, Foertsch LM, Price BB, Chaney JG, Lepoittevin JP. 2007. Quantification of chemical peptide reactivity for screening contact allergens: a classification tree model approach. *Toxicol Sci* 97:417-27.
- 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 Toxicological Methods (ICCVAM) and the National Toxicology Program Center for the Evaluation of Alternative Toxicological Methods (NICEATM). NIH Publication No. 99-4494. Research Triangle Park, NC: National Institute of Environmental Health Sciences.

- 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 Toxicology Program.
- ISO. 1994. ISO 5725. Accuracy (trueness and precision) of measurement methods and results. Geneva: International Organization for Standardization.
- ISO. 2002. ISO 10993-10: Tests for Irritation and Sensitization. Geneva: International Organization for Standardization.
- Jowsey IR, Clapp CJ, Safford R, Gibbons BT, Basketter D. 2008. The impact of vehicle on the relative potency of skin sensitising chemicals in the local lymph node assay. In Press.
- Kimber I, Basketter DA, Butler M, Gamer A, Garrigue JL, Newsome C, et al. 2003. Classification of allergens according to potency: proposals. *Food Chem Toxicol* 41:1799-1809.
- Kimber I, Dearman RJ, Betts CJ, Gerberick GF, Ryan CA, Kern PS, et al. 2006. The local lymph node assay and skin sensitization: a cut-down screen to reduce animal requirements? *Contact Dermatitis* 54:181-185.
- Manetz TS, Meade BJ. 1999. Development of a flow cytometry assay for the identification and differentiation of chemicals with the potential to elicit irritation, IgE-mediated, or T cell-mediated hypersensitivity responses. *Toxicol Sci.* 48:206-17.
- OECD. 2002. Guideline for the Testing of Chemicals 429 - Skin Sensitisation: Local Lymph Node Assay. Paris: OECD.
- Ryan CA, Chaney JG, Gerberick GF, Kern PS, Dearman RJ, Kimber I, Basketter DA. 2007. Extrapolating local lymph node assay EC3 values to estimate relative sensitizing potency. *Cutan Ocul Toxicol* 26:135-145.
- Schneider K, Akkan Z. 2004. Quantitative relationship between the local lymph node assay and human skin sensitization assays. *Regul Toxicol Pharmacol* 39:245-55.
- 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* 34:249-257.
- Takeyoshi M. 2007. Promising endpoint and assay performance of non-radioisotopic local lymph node assay based on BrdU incorporation. Sixth World Congress on Alternatives and Animal Use in the Life Sciences. August 21-25, 2007. Tokyo, Japan.
- Van Hoogstraten IM, Boos C, Boden D, Von Blomberg ME, Scheper RJ, Kraal G. 1993. Oral induction of tolerance to nickel sensitization in mice. *J Invest Dermatol* 101:26-31.
- Van Och FMM, Slob W, De Jong WH, Vandebriel RJ, Van Loveren H. 2000. A quantitative method for assessing the sensitizing potency of low molecular weight chemicals using a local lymph node assay: employment of a regression method that includes determination of the uncertainty margins. *Toxicology* 146:49-59.
- Varani J, Perone P, Spahlinger DM, Singer LM, Diegel KL, Bobrowski WF, et al. 2008. Human skin in organ culture and human skin cells (keratinocytes and fibroblasts) in monolayer culture for assessment of chemically induced skin damage. *Toxicol Pathol* 35:693-701.

Young SS. 2007. Gaming the system: chaos from multiple testing. Institute of Mathematical Statistics Bulletin 36:13

[This Page Intentionally Left Blank]

Appendix A

Peer Review Panel Member Biosketches

[This Page Intentionally Left Blank]

Panel Member Biosketches

Nathalie Alépée, Ph.D.

Dr. Alépée performed research leading to a Ph.D. in Medical Virology and Microbiology at the Centre National de la Recherche Scientifique research institute, Gif sur Yvette, France. She is currently the Global Pfizer Leader for photosafety, including the global portfolio support and Associate Research Fellow in Investigative Toxicology, at Pfizer Global Research and Development, Amboise, France. As a laboratory manager in the Molecular and Cellular Toxicology Group with Pfizer, she implemented the Local Lymph Node Assay (LLNA) in the laboratory. She serves on the European Centre for the Validation of Alternative Methods (ECVAM) Scientific Advisory Committee (ESAC), representing the European Federation of Pharmaceutical Industries Associations (EFPIA). She is also the Pfizer representative to the European Partnership on Alternative to Animal Testing (EPAA), in two working groups; Identification of Opportunities, Including R&D (working group 2), and Validation and Acceptance (working group 5). She served as a peer reviewer of the reduced LLNA test protocol and prediction model for ESAC in 2007 and has been designated as an ESAC peer reviewer for ECVAM's performance standards for the standard LLNA.

Anne Marie Api, Ph.D.

Dr. Api received a Ph.D. from Aston University in Birmingham, England and is currently Vice President of Human Health Sciences at the Research Institute for Fragrance Materials (RIFM), as well as the Scientific Director. She is responsible for the human health scientific program, and the investigation and initiation of new research and testing projects for RIFM. She is also Adjunct Assistant Professor at the University of Medicine and Dentistry of New Jersey. She is a member of 10 professional organizations, including the American Contact Dermatitis Society, the European Society of Contact Dermatitis, and the Society of Investigative Dermatology. She participated in the World Health Organization (WHO) International Workshop in Skin Sensitization in Chemical Risk Assessment held in Berlin, Germany in 2006. She is author of over 100 publications and presentations relevant to dermatology and dermatotoxicology.

Nancy Flournoy, M.S., Ph.D.

Dr. Flournoy received a M.S. degree in Biostatistics from the University of California at Los Angeles, and a Ph.D. in Biomathematics from the University of Washington. She is Professor and Chair of the Department of Statistics at the University of Missouri-Columbia. Her research interests include adaptive designs, bioinformatics, chemometrics, clinical trials, and environmetrics. She has an extensive list of edited volumes and papers on statistical theory, statistical genetics and immunology, epidemiology in immune suppressed subjects, clinical trials for prevention and treatment of viral infection, transplantation biology and its effects on digestion, lungs, eyes, mouth, and central nervous system, optimization of statistical processing, and additional papers, interviews, and technical reports. She has editorial responsibilities for numerous statistical journals, serves on numerous advisory boards, and nominating committees. She is a member and past Chair of the Council of Sections of the American Statistical Association, and served in various other statistical, medical and toxicological societies or programs as Chair or as a member of the Board of Directors. She is a former member of the Scientific Advisory Committee on Alternative Toxicological

Methods. She also served on the Expert Panels for the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) that evaluated the Revised Up-and-Down Procedure; the Current Validation Status of *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants; and Five *In Vitro* Pyrogen Test Methods.

Thomas Gebel, Ph.D.

Dr. Gebel received a Ph.D. in Toxicology from the University of Mainz and is certified as a toxicologist by the German Society of Toxicology. His scientific interests are in biomonitoring, genetic toxicology, environmental hygiene, and occupational toxicology. He has published over 40 papers in peer-reviewed scientific journals. He is employed by the German Federal Institute for Occupational Safety and Health, and is an Associate Professor at the University of Goettingen. Dr. Gebel is currently a member of the Organisation for Economic Co-operation and Development (OECD) Globally Harmonized System of Classification and Labeling of Chemicals (GHS) expert group on sensitization and head of the German advisory committee on classification and labeling of existing substances and biocides. Dr. Gebel also is head of the German Delegations to the United Nations Economic and Social Council Sub-Committee of Experts on the GHS, and to the OECD Task Force on Harmonisation of Classification and Labeling. He participated in the WHO International Workshop in Skin Sensitization in Chemical Risk Assessment held in Berlin, Germany in 2006.

Sidney Green Ph.D., F.A.T.S.

Dr. Green received a Ph.D. in Biochemical Pharmacology from Howard University. His research interests include toxicology, mutagenic assay systems, and alternatives to animals in toxicology. He is currently Graduate Professor of Pharmacology at Howard University and a faculty member at the Centers for Alternatives to Animal Testing at the Johns Hopkins University School of Public Health. Previously, he has been Director of the Department of Toxicology at Covance Laboratories Inc. and the Director of the Division of Toxicological Research at the U.S. Food and Drug Administration (FDA). Dr. Green is a Fellow of the Academy of Toxicological Sciences (F.A.T.S.). He has served on numerous expert panels and committees. He was a participant in an International Workshop organized by ICCVAM and NICEATM on *In Vitro* Methods for Assessing Acute Systemic Toxicity in 2000. He served on the ICCVAM/NICEATM Expert Panels that evaluated the Corrositex® Test Method for Assessing Dermal Corrosivity Potential of Chemicals, and *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants. He is a former member of the ICCVAM Advisory Committee on Alternative Toxicological Methods (ACATM) and of SACATM. He has authored over 60 publications for peer-reviewed journals.

Kim Headrick, B.Admin., B.Sc.

Kim Headrick received Bachelor of Administration and B.Sc. degrees from the University of Ottawa, Canada. She is currently International Harmonization and Senior Policy Advisor for Health Canada, and Chair of the UN Sub-Committee of Experts on GHS. She manages the overall strategy for the implementation of the GHS in Canada. She was awarded the Queen Elizabeth Commemorative Golden Jubilee Medal in 2002, which focuses on the achievements of people who, over the past 50 years, have created the Canada of today. She is

a member of the OECD Task Force on Harmonization of Classification and Labelling and the OECD Expert Group Meeting on Sensitization Hazards.

Dagmar Jírová, M.D., Ph.D.

Dr. Jírová received a Ph.D. from the Medical Faculty of Hygiene at Charles University in Prague. She is currently the Head of the Reference Center for Cosmetics, and Head of National Reference Laboratory for Experimental Immunotoxicology at the National Institute of Public Health in the Czech Republic. Her main responsibilities include safety assessment of consumer products, particularly cosmetics and their ingredients, performance of toxicological methods *in vivo* in animals, human patch testing for local toxicity assessment, and introduction of *in vitro* techniques for screening of toxicological endpoints using cell and tissue cultures. She represents the Czech Republic in the Standing Committee on Cosmetics of the European Commission. She is an ESAC-ECVAM member and was involved in Peer Review Panel for Skin Irritation Validation Study and LLNA test protocol and prediction model. She is author of more than 100 publications and presentations relevant to dermatotoxicology including a recent presentation at the 6th World Congress on Alternatives & Animal Use in the Life Sciences, held in Tokyo, 2007, titled “Comparison of Human Skin Irritation and Photoirritation Patch Test Data with Cellular *in vitro* Assays and Animal *in vivo* data”.

David Lovell, Ph.D., B.Sc. (Hons), F.S.S., FIBiol, CStat, CBiol

Dr. Lovell received a Ph.D. from the Department of Human Genetics and Biometry, University College, London. He is currently Reader in Medical Statistics at the Postgraduate Medical School at the University of Surrey. Previously, he was Associate Director and Head of Biostatistics support to Clinical Pharmacogenomics at Pfizer Global Research and Development in Sandwich, Kent providing data management and statistical support to pharmacogenetics and genomics. He joined Pfizer in 1999 as the Biometrics Head of Clinical Pharmacogenetics. Before joining Pfizer, Dr. Lovell was the Head of the Science Division at BIBRA International, Carshalton, which included Molecular Biology, Genetic Toxicology, Biostatistics and Computer Services. At BIBRA, Dr. Lovell managed the statistical and computing group providing specialized statistical support to BIBRA’s Clinical Unit and contract research work. He conducted and managed research programs on genetics, statistics and quantitative risk assessment for the European Union (EU) and U.K. Government Departments. His research interests at BIBRA were in the use of mathematical and statistical methods together with genetic models in the understanding of toxicological mechanisms and risk assessment problems. Dr. Lovell had previously been a Senior Research Officer with the U.K. Medical Research Council (MRC) Experimental Embryology and Teratology Unit, a visiting Postdoctoral Fellow at the National Institute of Environmental Health Sciences (NIEHS) in North Carolina, U.S., a Geneticist at the MRC Laboratories, Carshalton, and a Research Assistant in Cytogenetics at Birmingham University. He has acted as a consultant to a number of organizations, has considerable experience of working with Regulatory Authorities, has many publications related to his work and has wide experience of making presentations to a wide range of audiences. He is a member of the U.K. Government’s advisory Committee on Mutagenicity of Chemicals in Food, Consumer Products and the Environment (COM) and the Independent Scientific Advisory Committee for Medicines and Healthcare Products Regulatory Agency database research. He served on the NICEATM-ICCVAM Expert Panels

that evaluated the Frog Embryo Teratogenesis Assay - *Xenopus*, *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants, and Five *In Vitro* Pyrogen Test Methods.

Michael Luster, Ph.D.

Dr. Luster received a Ph.D. in Immunology from Loyola University of Chicago. He was formerly Chief, Toxicology and Molecular Biology Branch, Health Effects Laboratory Division, National Institute for Occupational Safety and Health (NIOSH), and currently serves as a senior advisor to the Director of the Health Effects Laboratories and the staff of Toxicology and Molecular Biology Branch at NIOSH. Program areas include neuroscience, dermatology, molecular carcinogenesis, molecular epidemiology, molecular toxicology, molecular epidemiology, and inflammation/immunotoxicology. In addition, Dr. Luster conducts basic and applied research in immunotoxicology including its application in risk assessment. Current research activities include molecular epidemiology studies of genetic polymorphism involved in workplace-related diseases and experimental studies involving occupational allergic rhinitis. Dr. Luster is also working with various staff at the U.S. Environmental Protection Agency (EPA) through the Risk Assessment Forum to develop immunotoxicity testing guidelines. He also directed two studies for the NTP on the Toxicology and the Carcinogenesis of Promethazine and Ortho-phenylphenol, in 1990 and 1986, respectively. He is a co-author of over 300 publications in peer-reviewed journals.

Howard Maibach, M.D.

Dr. Maibach received an M.D. from Tulane University. He is currently a professor in the Department of Dermatology at the University of California, San Francisco (UCSF), where he is also Chief of the Occupational Dermatology Clinic. In his 35 years at UCSF, Dr. Maibach has written and lectured extensively on dermatotoxicology and dermatopharmacology. His current research programs include defining the chemical-biologic faces of irritant dermatitis and the study of percutaneous penetration. Dr. Maibach served on the 1998 ICCVAM Peer Panel that evaluated the Murine LLNA. Dr. Maibach has been on the editorial boards of over 30 scientific journals and is a member of 19 professional societies including the American Academy of Dermatology, San Francisco Dermatological Society, and the International Commission on Occupational Health. He has co-authored over 1500 publications related to dermatology.

James McDougal, Ph.D., F.A.T.S.

Dr. McDougal earned a Ph.D. in Pharmacology/Toxicology at the University of Arizona. He is currently Professor and Director of Toxicology Research in the Department of Pharmacology and Toxicology at Wright State University's Boonshoft School of Medicine. Prior to his appointment at Wright State, he worked in the Air Force toxicology research organization for about 17 years. He has active skin research programs related to dermal pharmacokinetics, molecular biology of skin irritation, dermal risk assessment, and biologically-based mathematical modeling. He has served on many national committees, published more than 75 manuscripts, and consults for a wide variety of government and industry organizations. Dr. McDougal is a member of the National Academy of Sciences (National Research Council) Committee on Toxicology and the American Congress of Governmental Industrial Hygienists Threshold Limit Value Committee for Chemical substances. Dr. McDougal is also past president of the Dermal Toxicology Specialty Section of the Society of Toxicology.

Michael Olson, Ph.D., A.T.S.

Dr. Olson received a Ph.D. in Toxicology from the University of Arkansas for Medical Sciences, with dissertation research conducted at the FDA National Center for Toxicological Research. Following graduate training, he served as NIEHS National Research Service Award Post-doctoral Fellow in the Department of Pharmacology, School of Medicine - University of North Carolina. Currently he is Director, Occupational Toxicology, Corporate Environment Health and Safety for GlaxoSmithKline. Dr. Olson is a Fellow of the Academy of Toxicological Sciences (A.T.S.). His research interests include mechanisms of chemically-induced toxicity; genetic toxicity; xenobiotic metabolism; alternative methods in toxicology; hazard evaluation, risk assessment, and communication. Dr. Olson has authored a number of peer-reviewed manuscripts and book chapters in these areas as well as preparing many occupational health effects reviews for pharmaceutical active ingredients, isolated intermediates, and associated chemicals. He has served as an editorial board member and *ad hoc* referee for numerous toxicology and biosciences journals. In addition, he has worked as a Visiting Scientist, EPA, as well as advisor to EPA Risk Assessment Forum, U.S. National Institutes of Health (NIH) (Toxicology Study Section I), U.S. Air Force, Transportation Research Board, and the National Research Council - National Academy of Sciences. A member of several biomedical professional societies, Dr. Olson has served in elective and appointed positions in the Society of Toxicology, including Chairman of the Society of Toxicology (SOT) Occupational Health Specialty Section.

Raymond Pieters, Ph.D.

Dr. Pieters received a Ph.D. at Utrecht University and is currently an Associate Professor at the Institute for Risk Assessment Sciences, and Group Leader for Immunotoxicology at that institution. In 2007, he presented a paper on Development of Strategies to Assess Drug Hypersensitivity at the Congress of the European Societies of Toxicology. He was involved in the development of the Reporter Antigen Popliteal Lymph Node Assay, an assay to assess the immunomodulating potential of chemicals, which enables differentiation between immunosensitizing chemicals (sensitizers), immunostimulating chemicals (irritants), and chemicals that have no apparent immunological effects. He has published over 70 papers on sensitization and other subjects in immunotoxicology in peer-reviewed journals, including a review article, *Murine Models of Drug Hypersensitivity*, in 2005.

Jean Regal, Ph.D.

Dr. Regal received a Ph.D. in Pharmacology from the University of Minnesota. She is currently a Professor in the Department of Pharmacology, Department of Biochemistry & Molecular Biology and Associate Dean of Faculty Affairs, Medical School Duluth, University of Minnesota. Her current research is focused on respiratory allergy, especially asthma. She has served on multiple NIH review panels regarding asthma, as an immunotoxicologist in 2000 for an Institute of Medicine Committee on Health Effects Associated with Exposures Experienced during the Persian Gulf War, as well as on the 1998 ICCVAM Peer Panel that evaluated the Murine LLNA. In 2007 she served as an ad hoc reviewer for the NTP Board of Scientific Counselors for two nominations: Artificial Butter Flavoring Mixture & O-phthalaldehyde, at NIEHS. Also in 2007, she served on an NIEHS Center in Environmental Toxicology pilot project program for the University of Texas Medical Branch at Galveston. She is currently Vice-President-elect of the Immunotoxicology

Specialty Section of SOT and Associate Editor of the Journal of Immunotoxicology. Dr. Regal has authored over 50 research articles and reviews in peer-reviewed journals and holds two patents on pulmonary administration of sCR1 and other complement inhibitory proteins.

Jonathan Richmond, B.Sc. (Hons) Med.Sci., MB ChB, FRCSEd, FRMS

Dr. Richmond received a Bachelor of Science in Medical Science with Honors (B.Sc. [Hons] Med.Sci.) and Bachelor of Medicine and Bachelor of Surgery (MB ChB) degrees with Distinction in Medicine and Therapeutics from Edinburgh University. Presently, he is head of the Animals Scientific Procedures Division at the Home Office. He is a Fellow of the Royal College of Surgeons of Edinburgh (FRCSEd) and a Fellow of the Royal Society of Medicine (FRMS). Other appointments include convener of the U.K. interdepartmental group on the 3Rs, board member U.K. National Centre for the 3Rs, convener of the International Standards Organization Technical Corrigendum 194/Working Group 3 (*Biocompatibility of Medical Device Materials*), and member of related expert working groups. He is a former member of the EU Committee on Scientific and Technical Progress and past Chairman of the European Commission Technical Expert Working Group on ethical review. He served as chair of the peer review panel for the reduced LLNA test protocol and prediction model for ESAC in 2007 and has been designated as an ESAC peer reviewer for ECVAM's performance standards for the standard LLNA. He served on the ICCVAM/NICEATM Expert Panel that evaluated Five *In Vitro* Pyrogen Test Methods. He has a variety of publications in peer-reviewed journals and national and international meetings, on the principles and practice of surgery, regulation of biomedical research, principles of humane research, bioethics, and public policy.

Peter Theran, V.M.D.

Dr. Theran holds a Doctor of Veterinary Medicine degree from the University of Pennsylvania. He has had many years of experience both as a veterinary internal medicine specialist at the Massachusetts Society for Prevention of Cruelty to Animals' Angell Memorial Animal Hospital in Boston, and as the director of Boston University Medical Center's Laboratory Animal Science Center. He presently serves on a number of government committees as an animal welfare member, and is a member of the Board of Directors of the Institute for *In Vitro* Sciences in Gaithersburg, MD and Chimp Haven in Shreveport, Louisiana. He served on the NICEATM-ICCVAM Expert Panels that evaluated the *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants, and Five *In Vitro* Pyrogen Test Methods. He is a former member of ACATM and SACATM. He is presently working as a consultant.

Stephen Ullrich, Ph.D.

Dr. Ullrich received a Ph.D. in Microbiology from Georgetown University. He is currently the Dallas/Fort Worth Living Legends Professor, and Professor of Immunology at the University of Texas, M.D. Anderson Cancer Center, where he is also Associate Director, The Center for Cancer Immunology Research. He is also a member of the Animal Research Strategic Advisory Committee. He has served numerous national review committees and panels, including the 1998 ICCVAM Peer Panel that evaluated the Murine LLNA. Dr. Ullrich has authored over 75 peer-reviewed publications, over 30 invited articles, and he holds four patents in the U.S., E.U., and Australia for a UV-induced immunosuppressive substance. He is the past President of the American Society for Photobiology.

Michael Woolhiser, Ph.D.

Dr. Woolhiser received a Ph.D. in Pharmacology and Toxicology from the Medical College of Virginia at Virginia Commonwealth University. He is a specialist in immunotoxicology and is currently a toxicologist for the Dow Chemical Company where he serves as a Technical Leader for Immunotoxicology, and Polyurethane Business Toxicology Consultant. Dr. Woolhiser is also an Adjunct Professor at the Center for Integrative Toxicology, Michigan State University. He is a member of the Program Committee of the Society of Toxicology's Immunotoxicology Specialty Section. He has served on numerous working groups, including an LLNA Expert Working Group under the European Crop Protection Agency's Toxicology Expert Group, a European Centre for Ecotoxicology and Toxicology of Chemicals LLNA Task Force. He has authored 29 peer-reviewed publications.

Takahiko Yoshida, M.D., Ph.D.

Dr. Yoshida earned his M.D. and a Ph.D. in Medical Science from Tokai University. He is currently Professor in the Department of Health Science at Asahikawa Medical College. Prior to this appointment, he held the posts of Instructor, Assistant Professor and Associate Professor at the Tokai University School of Medicine. He has also been a Guest Researcher at NIEHS. He has also worked as an occupational physician for major Japanese corporations, including Toyota and Sony. Dr. Yoshida's research interests include occupational health, public health, environmental health and preventative medicine. He is a member of the International Congress of Occupational Health, the Japanese Society of Hygiene, the Japanese Society of Immunotoxicology, the Japanese Society of Clinical Ecology, and the SOT.

[This Page Intentionally Left Blank]

Appendix B

Questions for the Peer Review Panel

B1	LLNA Limit Dose Procedure	B-3
B2	LLNA for Testing Aqueous Solutions, Metals, and Mixtures.....	B-11
B3	Non-Radioactive LLNA Protocol - LLNA: DA Test Method	B-17
B4	Non-Radioactive LLNA Protocol - LLNA: BrdU-FC Test Method	B-25
B5	Non-Radioactive LLNA Protocol - LLNA: BrdU-ELISA Test Method	B-33
B6	Draft ICCVAM Performance Standards for the LLNA	B-41
B7	Use of the LLNA for Potency Determinations.....	B-47

[This Page Intentionally Left Blank]

Appendix B1

Questions for the Peer Review Panel: LLNA Limit Dose Procedure

[This Page Intentionally Left Blank]

Instructions for the Peer Review Panel: LLNA Limit Dose Procedure

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) in conjunction with National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) is evaluating the validation status of the murine local lymph node assay (LLNA) limit dose procedure as a substitute for the traditional LLNA for assessing the allergic contact dermatitis potential of chemicals and other substances.

The following materials are intended to guide you in your review of the ICCVAM evaluation of the LLNA limit dose procedure. You are first asked to review the information in the draft ICCVAM LLNA limit dose procedure Background Review Document (BRD) for completeness, and to identify any errors or omissions of existing relevant data or information that should be included. You are then asked to evaluate the information in this BRD to determine the extent to which each of the applicable criteria for validation and acceptance of toxicological test methods (ICCVAM 2003¹) have been appropriately addressed for the proposed use of the LLNA limit dose procedure. Adequate validation² is a prerequisite for a test method to be considered for use in regulatory decision-making by U.S. Federal agencies. This validation process characterizes the usefulness and limitations of a test method for its intended use.

Lastly, you are asked to consider the ICCVAM draft test method recommendations for the LLNA limit dose procedure (i.e., the proposed test method use, the proposed recommended standardized protocol, the proposed test method performance standards, and any proposed additional studies) and comment on whether the recommendations are supported by the information provided in the draft LLNA limit dose procedure BRD.

The questions relating to the draft BRD that must be addressed are provided in **Sections I** and **II** of this guidance, while **Section III** contains questions relating to the draft ICCVAM test method recommendations on the LLNA limit dose procedure.

These include questions prepared by the ICCVAM IWG to ensure that the assessment provides adequate information to facilitate U.S. Federal agency decisions on the regulatory acceptability of this test method, and/or adequate guidance for organizations that may be involved in conducting or supporting further development, standardization, and/or validation.

The overall question to consider is whether the validation status of the LLNA limit dose procedure has been adequately characterized for its intended purpose, and is it sufficiently accurate and reliable to be used for the identification of sensitizing substances and non-sensitizing substances in place of the traditional LLNA procedure when there is not a need for dose response information, in order to reduce the number of animals required for such testing.

I. Questions to the Panel: Review for Errors and Omissions

¹ 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. NIEHS (Available: <http://iccvam.niehs.nih.gov/docs/guidelines/subguide.htm>).

² Validation is the process by which the reliability and accuracy of a test method are established for a specific purpose (ICCVAM 2003).

1. Are there any errors or omissions of existing relevant data or information that should be included in the draft BRD?

II. **Questions to the Panel: LLNA Limit Dose Procedure Draft BRD**

1. For the proposed LLNA limit dose procedure, ICCVAM recommends that the number of animals used in each group should be the same as that recommended by ICCVAM for the traditional LLNA based on its 1998 evaluation of the LLNA, and that individual animal data should be collected and reported (ICCVAM, 1999). Do you agree that these are appropriate provisions for the limit dose procedure? Please explain your answer.
2. Do you consider the traditional LLNA database representative of a sufficient range of chemical classes and physical chemical properties that it would be applicable to any of the types of chemicals and products that are typically tested for skin sensitization potential? If not, what are the relevant chemical classes/properties (other than those that are identified as limitations in the traditional LLNA) that should be tested with caution, or not evaluated using the limit dose procedure? What chemicals or products should be evaluated to fill this data gap? Please explain your answer.
3. While coding of chemicals is recommended for prospective studies, this evaluation is based on a retrospective evaluation of existing data, most of which was not generated using coded chemicals to reduce the potential for bias. Does the lack of coding of test substances adversely impact or bias the current evaluation? Please explain your answer.
4. For some substances submitted using the traditional LLNA test method, it was not possible to confirm whether the data were generated using pooled animal data for each dose group (as allowed in Test Guideline [TG] 429 of the Organisation for Economic Co-operation and Development [OECD]). ICCVAM (1999), Dean et al. (2001), and EPA (2003) recommend the use of statistical analyses to help interpret LLNA study results, which necessitates data collected at the level of the individual animal, while Cockshott et al. (2006) reported that using individual animal data allowed for technical problems during an experiment to be identified. Considering this, should the analysis of the performance of the LLNA limit dose procedure against the traditional LLNA be limited to data from studies that can be confirmed as using individual animal data collection procedures? What impact might the inclusion of pooled animal data have on the accuracy analysis of the LLNA limit procedure? Please explain your answer.
5. Has the relevance (e.g., accuracy/concordance, sensitivity, specificity, false positive and false negative rates) of the LLNA limit dose procedure been adequately evaluated and compared to the traditional LLNA (refer also to Table 6-1 of the draft ICCVAM BRD)? If not, what other analyses should be performed?
6. There were five substances for which the highest concentration tested produced an SI of less than 3.0, while lower concentrations of these substances produced an SI of greater than 3.0 (see Table 6-2 of the draft ICCVAM BRD). These

substances are classified as “false negatives” compared to what was obtained in the traditional LLNA. Can you identify any characteristics associated with these or other substances that might signal that this type of abnormal dose response might occur, and therefore using the LLNA limit dose procedure would not be appropriate? Please explain your answer.

7. Does the BRD adequately characterize the usefulness and limitations of the LLNA limit dose procedure based on the accuracy analyses? If not, what additions or changes should be made to the current usefulness and limitations? Please explain your answer.
8. Is it appropriate to assume that the intra- and inter-laboratory reproducibility of the LLNA limit dose procedure and the traditional LLNA will be similar, based on the fact that they use identical protocols with the exception of the number of doses used? Do you agree? Does reducing the number of test substances dose groups from three to one reduce the reliability of the assay? Please explain your answer.
9. For some studies included in the draft BRD, it was not possible to determine whether or not they had been conducted in accordance with Good Laboratory Practice (GLP) guidelines. Original records for some of the non-GLP studies included in this evaluation could not be obtained. As a result, an independent audit could not be conducted to confirm that the reported data is the same as the data recorded in laboratory notebooks. Neither was it possible to obtain the results of GLP audits for all studies conducted in accordance with GLP guidelines. Considering this, should the results of these studies (all of which are currently included) be excluded from any of the performance analyses? Please explain your answer.
10. Based on the draft BRD, have all the relevant data identified in published or unpublished studies conducted using the traditional LLNA been adequately considered? If not, what other traditional LLNA data needs to be considered and how can it be obtained?

III. Questions to the Panel: Draft ICCVAM Test Method Recommendations on the LLNA Limit Dose Procedure

1. Do the available data support the ICCVAM draft recommendations for the LLNA limit dose procedure in terms of the proposed test method usefulness and limitations? If not, what recommendations would you make? Please explain your answer.
 - Should the LLNA limit dose procedure be routinely recommended for the hazard identification of skin sensitizing chemicals when potency information is not required? Please explain your answer.
 - If potency information is required, should the LLNA limit dose procedure be routinely recommended as the initial test to identify sensitizers before conducting the traditional LLNA as a way to further reduce animal use, since negative results would not require further testing? Please explain your answer.
 - Based on the existing database, there is a false negative rate of 1.6% (5/313 positive compounds) for the LLNA limit dose approach compared to the results obtained in the traditional LLNA. Do you consider that this is adequately addressed by the proposed cautionary language and weight of evidence consideration for negative substances? Please explain your answer.
2. Do you agree that the available data support the ICCVAM draft recommendations for the LLNA limit dose procedure in terms of the proposed test method standardized protocol? If not, then what recommendations would you make? Please explain your answer.
 - The recommended ICCVAM protocol (ICCVAM 1999; Dean et al. 2001; EPA 2003), as well as OECD TG 429, specifies that the highest dose tested should be the highest soluble concentration that does not induce systemic toxicity and/or excessive skin irritation. However, Kimber et al. (2006) concluded that negative results obtained from studies where the highest concentration tested was below 10% should be considered invalid, and adopted a 10% application concentration as a threshold of confidence for categorization of a chemical as being negative while noting that the figure should not be considered as inviolable. Are the data presented in the draft BRD (i.e., 5/313 positive substances in the NICEATM database were negative at concentrations $\leq 10\%$, but were positive at higher concentrations) adequate to conclude that this threshold concentration is not appropriate? If a negative result was obtained for a test substance in a study where the highest concentration that could be tested (based on systemic toxicity or excessive local irritation, as described in ICCVAM [1999], Dean et al. [2001], and EPA [2003]) was $< 10\%$, should additional testing be required? Do you agree that the current approach for selecting the “limit” dose is appropriate or do you conclude that there is a threshold concentration for the LLNA at which a negative result could always be considered as an acceptable result? If so, what is that concentration? Please explain your answer.

3. Do you agree that the available data support the ICCVAM draft recommendations for the LLNA limit dose procedure in terms of the proposed future studies? If not, then what recommendations would you make? Please explain your answer.

[This Page Intentionally Left Blank]

Appendix B2

Questions for the Peer Review Panel: LLNA for Testing Aqueous Solutions, Metals, and Mixtures

[This Page Intentionally Left Blank]

Instructions for the Peer Review Panel: LLNA for Testing Aqueous Solutions, Metals, and Mixtures

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) in conjunction with National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) is currently updating the original validation report of the LLNA (ICCVAM 1999) based on a comprehensive review of available data and information regarding the current validity of the LLNA for assessing the skin sensitizing potential of mixtures, metal compounds, and substances tested in aqueous solutions. The information is based on a retrospective review of LLNA data derived from a database of over 500 substances (including mixtures) tested in the LLNA and builds on the previous ICCVAM evaluation of the LLNA, which was based on 209 substances (ICCVAM 1999). In the original ICCVAM report, the performance of the LLNA was compared to 1) the results from guinea pig tests and 2) information about sensitizers in humans (e.g., human maximization test [HMT] results, substances used in human repeat insult patch test [HRIPT], clinical data), where available. This addendum updates the LLNA performance analyses for mixtures, metal compounds, and substances tested in aqueous solutions when compared to human and guinea pig results.

The following materials are intended to guide you in your review of the ICCVAM evaluation of the LLNA. You are first asked to review the information in the draft Addendum to the ICCVAM (1999) report for completeness, and to identify any errors or omissions of existing relevant data or information that should be included. You are then asked to evaluate the information in this Addendum to determine the extent to which each of the applicable criteria for validation and acceptance of toxicological test methods (ICCVAM 2003³) have been appropriately addressed for the proposed use of the LLNA for testing mixtures, metal compounds, and substances in aqueous solutions. Adequate validation⁴ is a prerequisite for a test method to be considered for use in regulatory decision-making by U.S. Federal agencies. This validation process characterizes the usefulness and limitations of a test method for its intended use.

Lastly, you are asked to consider the ICCVAM draft test method recommendations for the LLNA (i.e., the proposed test method use, the proposed recommended standardized protocol, the proposed test method performance standards, and any proposed additional studies) and comment on whether the recommendations are supported by the information provided in the draft Addendum.

The questions relating to the draft Addendum that must be addressed are provided in **Sections I** and **II** of this guidance, while **Section III** contains questions relating to the draft ICCVAM test method recommendations on the LLNA for testing mixtures, metal compounds, and substances in aqueous solutions.

³ 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. NIEHS (Available: <http://iccvam.niehs.nih.gov/docs/guidelines/subguide.htm>).

⁴ Validation is the process by which the reliability and accuracy of a test method are established for a specific purpose (ICCVAM 2003).

These include questions prepared by the ICCVAM IWG to ensure that the assessment provides adequate information to facilitate U.S. Federal agency decisions on the regulatory acceptability of this test method, and/or adequate guidance for organizations that may be involved in conducting or supporting further development, standardization, and/or validation.

The overall question to consider is whether the validation status of the LLNA for testing mixtures, metal compounds, and substances in aqueous solutions has been adequately characterized, and is it sufficiently accurate and reliable to be used for the identification of sensitizing substances based on a comparison to either human or guinea pig responses.

I. Questions to the Panel: Review for Errors and Omissions

1. In the draft Addendum, are there any errors that need to be corrected or omissions of existing relevant data or information that should be included?

II. Questions to the Panel: Updated LLNA Applicability Domain Addendum

1. Do you consider the database of substances evaluated representative of a sufficient range of mixtures, metal compounds, and substances in aqueous solutions that are typically tested for skin sensitization potential? Please explain your answer.
2. For the purpose of this evaluation, aqueous solutions were defined by the proportion of water (at least 20%) (i.e., substances or mixtures that were tested in an aqueous or an organic:aqueous vehicle were labeled as aqueous solutions). Do you consider this to be an appropriate criterion for defining aqueous solutions? If not, what would be more appropriate? Please explain your answer.
3. While coding of chemicals is recommended for prospective studies, this evaluation is based on a retrospective evaluation of existing data, most of which was not generated using coded chemicals to reduce bias. Does the lack of coding of test substances adversely impact or bias the current evaluation? Please provide a rationale for your answer.
4. For some substances submitted using the LLNA, it was not possible to confirm whether the data were generated using pooled animal data for each dose group (as allowed in Test Guideline [TG] 429 of the Organisation for Economic Co-operation and Development [OECD]) rather than individual animal data (as recommended in the ICCVAM 2001 protocol)? Cockshott et al. (2006) reported that using individual animal data allowed for technical problems during an experiment to be identified. Considering this, should the analysis of the performance of the LLNA for testing mixtures, metal compounds, and substances in aqueous solutions be limited to data from studies that can be confirmed as using individual animal data collection procedures? What impact might the inclusion of pooled animal data have on the accuracy analysis included in Section 5.0 of the draft Addendum? Please explain your answer.
5. Has the relevance (e.g., accuracy/concordance, sensitivity, specificity, false positive and false negative rates) of the LLNA for testing mixtures, metal compounds, and substances in aqueous solutions been adequately evaluated and compared to the human and guinea pig (refer also to Section 5.0 of the draft

- Addendum)? If not, what other analyses should be performed? Please explain your answer.
6. When multiple LLNA studies were available for the same substance, the majority call (where all studies used the same vehicle and the same concentration range) was used to assign an overall classification for the purposes of the accuracy analysis. For example, if chemical X was tested 5 times and was positive in 3 studies and negative in two, the overall classification was positive. Do you agree with the approach to assigning overall classifications? If not, how would you propose that this be accomplished? Please explain your answer.
 7. Does the Addendum adequately characterize the usefulness and limitations of the LLNA for testing mixtures, metal compounds, and substances in aqueous solutions based on the accuracy analyses? If not, what additions or changes should be made to the current usefulness and limitations? Please explain your answer.
 8. For some studies included in the draft Addendum, it was not possible to determine whether or not they had been conducted in accordance with Good Laboratory Practice (GLP) guidelines. Neither was it possible to obtain the results of GLP audits for all studies conducted in accordance with GLP guidelines. Please discuss what impact this lack might have on the evaluation of the LLNA for testing mixtures, metal compounds, and substances in aqueous solutions and whether such studies should be excluded from any analysis.
 9. As described in the draft Addendum, original records for some of the non-GLP studies included in this evaluation could not be obtained. As a result, an independent audit could not be conducted to confirm that the reported data is the same as the data recorded in laboratory notebooks. Considering this, should the results of these studies (all of which are currently included) be excluded from any of the performance analyses? Please explain your answer.
 10. Based on the draft Addendum, have all the relevant data identified in published or unpublished studies conducted using the LLNA for testing mixtures, metal compounds, and substances in aqueous solutions been adequately considered? If not, what other studies should to be considered?

III. Questions to the Panel: Draft ICCVAM Test Method Recommendations on the LLNA for Testing Aqueous Solutions, Metals, and Mixtures

1. Do you agree that the available data support the ICCVAM draft recommendations for the LLNA with regard to testing mixtures, metal compounds, and substances in aqueous solutions in terms of the proposed test method usefulness and limitations? Please explain your answer. If not, what recommendations would you make?
2. Do you agree that the available data support the ICCVAM draft recommendations for the LLNA in terms of the proposed test method standardized protocol? Please explain your answer. If not, then what recommendations would the Panel make?

3. Do you agree that the available data support the ICCVAM draft recommendations for the LLNA in terms of the proposed future studies? Please explain your answer. If not, then what recommendations would you make?

Appendix B3

Questions for the Peer Review Panel: Non-Radioactive LLNA Protocol - LLNA: DA Test Method

[This Page Intentionally Left Blank]

Instructions for the Peer Review Panel: Non-Radioactive LLNA Protocol: LLNA: DA Test Method

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) in conjunction with National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) is evaluating the validation status of the LLNA: DA (Local Lymph Node Assay-Daicel adenosine triphosphate [ATP]) for assessing the allergic contact dermatitis (ACD) potential of chemicals and other substances. This test method, developed by Daicel Chemical Industries, Ltd. (Tokyo, Japan), is a non-radiolabeled version of the traditional LLNA, and is based on measuring levels of ATP in the auricular lymph nodes as an indicator of increased cell proliferation.

The following materials are intended to guide you in your review of the ICCVAM evaluation of the LLNA: DA. You are first asked to review the information in the draft ICCVAM LLNA: DA Background Review Document (BRD) for completeness, and to identify any errors or omissions of existing relevant data or information that should be included. You are then asked to evaluate the information in the BRD to determine the extent to which each of the applicable criteria for validation and acceptance of toxicological test methods (ICCVAM 2003⁵) have been appropriately addressed for the proposed use of the LLNA: DA. Adequate validation⁶ is a prerequisite for a test method to be considered for use in regulatory decision-making by U.S. Federal agencies. The validation process characterizes the usefulness and limitations of a test method for a specific intended use.

Lastly, you are asked to consider the ICCVAM draft test method recommendations for the LLNA: DA (i.e., the proposed test method use, the proposed recommended standardized protocol, the proposed test method performance standards, and any proposed additional studies) and comment on whether the recommendations are supported by the information provided in the draft LLNA: DA BRD.

The questions relating to the draft BRD that must be addressed are provided in **Sections I** and **II** of this guidance, while **Section III** contains questions relating to the draft ICCVAM test method recommendations on the LLNA: DA.

These include questions prepared by the ICCVAM Immunotoxicity Working Group (IWG) to ensure that the assessment provides adequate information to facilitate U.S. Federal agency decisions on the regulatory acceptability of this test method. The questions are also intended to obtain guidance that will be helpful to federal agencies and other organizations that may be involved in conducting or supporting further development, standardization, and/or validation studies.

The overall question to consider is whether the validation status of the LLNA: DA has been adequately characterized for its intended purpose, and is it sufficiently accurate and reliable

⁵ 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. NIEHS (Available: <http://iccvam.niehs.nih.gov/docs/guidelines/subguide.htm>).

⁶ Validation is the process by which the reliability and accuracy of a test method are established for a specific purpose (ICCVAM 2003).

to be used for the identification of sensitizing substances and non-sensitizing substances in place of the traditional LLNA procedure.

I. Questions to the Panel: Review of the Draft LLNA: DA BRD for Errors and Omissions

1. In the draft LLNA: DA BRD, are there any errors in the BRD that should be corrected, or omissions of existing relevant data or information that should be included?

II. Questions to the Panel: Draft LLNA: DA BRD

1. Test Method Protocol

- i. The traditional LLNA protocol (ICCVAM 1999; Dean et al. 2001) recommends a minimum of five successfully treated animals per dose group. Current validation of the LLNA: DA was performed using four animals per dose group. What impact might using fewer mice have on the accuracy analysis of the LLNA: DA? Please explain your answer.
- ii. The data generated for the substances analyzed in the LLNA: DA test method come from auricular lymph nodes that were pooled across mice in each dose group rather than being analyzed on an individual animal data. What impact might the inclusion of pooled animal data have on the accuracy analysis of the LLNA: DA? Please explain your answer.
- iii. The LLNA: DA differs from the traditional LLNA in the treatment schedule and by including a pre-treatment with 1% SLS prior to application of the test substance. Do you consider these changes to be appropriate? Please explain your answer.

2. Substances Used for the Validation Studies

- i. Do you consider the LLNA: DA database representative of a sufficient range of chemical classes and physicochemical properties that it would be applicable to any of the types of chemicals and products that are typically tested for skin sensitization potential? If not, what are the relevant chemical classes/properties (other than those that are identified as limitations in the traditional LLNA) that should be tested with caution, or not evaluated using the LLNA: DA? What chemicals or products should be evaluated to fill this data gap? Please explain your answer.

3. Test Method Accuracy

- i. The current accuracy analysis is based on overall concordance with the traditional LLNA. Accuracy statistics compared to the guinea pig tests and human data/experience have also been provided. Are these comparisons appropriate for assessing the accuracy of the LLNA: DA? Please explain your answer.
- ii. Has the relevance (e.g., accuracy/concordance, sensitivity, specificity, false positive and false negative rates) of the LLNA: DA been adequately evaluated and compared to the traditional LLNA (refer also to Table 6-1 of the draft

- ICCVAM BRD)? If not, what other analyses should be performed? Please explain your answer.
- iii. There was one substance (2-mercaptobenzothiazole) that produced a “false negative” response compared to the traditional LLNA when tested using the LLNA: DA. The mean EC3 in the traditional LLNA for this substance is 2.5 (n=2), it is positive in both the guinea pig and human, and has "high" peptide reactivity as per Gerberick et al. (2007). Can you identify any characteristics associated with this or similar substances, compared to the correctly identified sensitizers, that might signal that this type of discordant response might occur, and therefore using the LLNA: DA to test such substances would not be appropriate or that negative results for such substances should indicate a need for confirmatory testing? Please explain your answer.
 - iv. There was one substance (benzalkonium chloride) that produced a “false positive” response compared to the traditional LLNA and guinea pig test when tested using the LLNA: DA. Can you identify any characteristics associated with this or similar substances, compared to the correctly identified non-sensitizers that might signal that this type of discordant response might occur, and therefore using the LLNA: DA to test such substances would not be appropriate, or that positive results for substances with such properties may warrant additional testing? Please explain your answer.
4. Test Method Reliability (Intra- and Inter-laboratory reproducibility)
- i. Has the intralaboratory reproducibility of the LLNA: DA been adequately evaluated and compared to the traditional LLNA (refer also to Table 7-1 of the draft LLNA: DA BRD)? If not, what other analyses should be performed? Are any limitations apparent based on this intralaboratory reproducibility assessment? Please explain your answer.
 - ii. Has the interlaboratory reproducibility of the LLNA: DA been adequately evaluated and compared to the traditional LLNA (refer also to Tables 7-2 and 7-3 of the draft LLNA: DA BRD)? If not, what other analyses should be performed? Are any limitations apparent based on this interlaboratory reproducibility assessment? Please explain your answer.
 - iii. The draft LLNA: DA BRD analyzes data from two interlaboratory validation studies that used coded substances, as well as an intralaboratory validation study with 31 substances that were not coded. Does the lack of coding of test substances adversely impact or bias the current evaluation? In addition, it appears that the lead laboratory established the dose levels tested in the two interlaboratory validation studies and the participating laboratories did not determine their own dose levels for testing. Does this adversely impact or bias the current evaluation? Please explain your answer.
5. Data Quality
- i. The studies evaluated in the draft BRD for the LLNA: DA were not conducted in accordance with Good Laboratory Practices (GLP) guidelines although there were reportedly done in laboratories that conduct GLP studies and were

conducted "in the spirit" of GLP (K. Idehara, personal communication). Please discuss what impact this might have on the evaluation of the LLNA: DA.

- ii. The original records for these studies were requested but have not yet been obtained. As a result, an independent audit could not be conducted to confirm that the reported data is the same as the data recorded in laboratory notebooks. Should any recommendations from ICCVAM be contingent upon the completion of such an audit and findings that there were no significant errors in data transcription? Please explain your answer.
6. Consideration of all available data and relevant information
 - i. Based on the draft LLNA: DA BRD, have all the relevant data identified in published or unpublished studies that employ this test method been adequately considered? Are there other comparative test method data that were not considered in the draft BRD, but are available for consideration? If yes, please explain how to obtain such data.

IV. **Questions to the Panel: Draft ICCVAM Test Method Recommendations on the LLNA: DA**

1. Test Method Usefulness and Limitations
 - i. Do you agree that the available data and test method performance (accuracy and reliability) support the ICCVAM draft recommendations for the LLNA: DA in terms of the proposed test method usefulness and limitations? Please explain your answer.
 - ii. If restrictions on using radioactive materials are present, should the LLNA: DA be routinely recommended for hazard identification of skin sensitizing substances in lieu of having to possibly use guinea pig tests? Please explain your answer.
 - iii. Even if limitations in using radioactive materials are not present, should the LLNA: DA procedure or other valid and accepted non-radioactive method be routinely recommended for hazard identification of skin sensitizing substances instead of the traditional LLNA? Please explain your answer.
 - iv. From a public health perspective, is the recommended guidance for evaluating negatives sufficient to address concerns associated with the false negative rate of 5% (1/19 substances) calculated for the LLNA: DA? Do you have suggestions for additional guidance or limitations? Please explain your answer.
 - v. From a testing strategy perspective, does the ICCVAM guidance address concerns associated with the false positive rate of 10% (1/10 substances) calculated for the LLNA: DA? Are there other suggestions for additional such guidance or limitations? Please explain your answer.
2. Test method protocol

- i. Do you agree that the available data support the ICCVAM draft recommendations for the LLNA: DA procedure in terms of the proposed test method standardized protocols? If not, what recommendations would you make? Please explain your answer.
 - ii. Can the traditional LLNA limit dose procedure be applied to the LLNA: DA? Please explain your answer.
3. Future Studies
 - i. Do you agree that the available data support the ICCVAM draft recommendations for the LLNA: DA in terms of the proposed future studies? If not, then what recommendations would you make? Please explain your answer.
4. Performance Standards
 - i. The LLNA: DA protocol differs from the ICCVAM-recommended protocol for the traditional LLNA (ICCVAM 1999; Dean et al. 2001; EPA 2003) in the method used to assess lymphocyte proliferation in the auricular lymph nodes. In addition, there are differences between the two protocols that relate to how and when the test substance is applied and when the lymph nodes are collected (Table 2-1 and Appendix A in the draft LLNA: DA BRD). According to the proposed draft ICCVAM Performance Standards for the traditional LLNA (http://iccvam.niehs.nih.gov/methods/immunotox/llna_PerfStds.htm), any change to the LLNA protocol other than the method used to assess lymphocyte proliferation is considered a major change. Do you agree that these should be considered major changes and therefore the usefulness and limitations of the LLNA: DA should not be assessed using the draft ICCVAM Performance Standards? Please explain your answer.
 - ii. Even if the draft ICCVAM Performance Standards do not apply to the LLNA: DA, what impact should the accuracy analysis based on 13 of the 18 required performance standards substances (only one false negative and no false positives) have on the overall evaluation of test method accuracy? Please explain your answer.
 - iii. Should separate performance standards be developed for the LLNA: DA? Please explain your answer.

[This Page Intentionally Left Blank]

Appendix B4

Questions for the Peer Review Panel: Non-Radioactive LLNA Protocol - LLNA: BrdU-FC Test Method

[This Page Intentionally Left Blank]

Instructions for the Peer Review Panel: Non-Radioactive LLNA Protocol: LLNA: BrdU-FC Test Method

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) in conjunction with National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) is evaluating the validation status of the LLNA: BrdU-FC (Local Lymph Node Assay: Bromodeoxyuridine detected by flow cytometry) for assessing the allergic contact dermatitis potential of chemicals and other substances. This test method, developed by MB Research Labs (Spinnerstown, PA), is a non-radiolabeled version of the traditional LLNA, and is based on measuring the incorporation of bromodeoxyuridine (BrdU) into the DNA of dividing lymphocytes using flow cytometry as an indicator of cell proliferation.

The following materials are intended to guide you in your review of the ICCVAM evaluation of the LLNA: BrdU-FC. You are first asked to review the information in the draft ICCVAM LLNA: BrdU-FC Background Review Document (BRD) for completeness, and to identify any errors or omissions of existing relevant data or information that should be included. You are then asked to evaluate the information in the BRD to determine the extent to which each of the applicable criteria for validation and acceptance of toxicological test methods (ICCVAM 2003⁷) have been appropriately addressed for the proposed use of the LLNA: BrdU-FC. Adequate validation⁸ is a prerequisite for a test method to be considered for use in regulatory decision-making by U.S. Federal agencies. The validation process characterizes the usefulness and limitations of a test method for a specific intended use.

Lastly, you are asked to consider the ICCVAM draft test method recommendations for the LLNA: BrdU-FC (i.e., the proposed test method use, the proposed recommended standardized protocol, the proposed test method performance standards, and any proposed additional studies) and comment on whether the recommendations are supported by the information provided in the draft LLNA: BrdU-FC BRD.

The questions relating to the draft BRD that must be addressed are provided in **Sections I** and **II** of this guidance, while **Section III** contains questions relating to the draft ICCVAM test method recommendations on the LLNA: BrdU-FC.

These include questions prepared by the ICCVAM Immunotoxicity Working Group (IWG) to ensure that the assessment provides adequate information to facilitate U.S. Federal agency decisions on the regulatory acceptability of this test method. The questions are also intended to obtain guidance that will be helpful to federal agencies and other organizations that may be involved in conducting or supporting further development, standardization, and/or validation studies.

The overall question to consider is whether the validation status of the LLNA: BrdU-FC has been adequately characterized for its intended purpose, and is it sufficiently accurate and

⁷ 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. NIEHS (Available: <http://iccvam.niehs.nih.gov/docs/guidelines/subguide.htm>).

⁸ Validation is the process by which the reliability and accuracy of a test method are established for a specific purpose (ICCVAM 2003).

reliable to be used for the identification of sensitizing substances and non-sensitizing substances in place of the traditional LLNA procedure.

I. Questions to the Panel: Review of the Draft LLNA: BrdU-FC BRD for Errors and Omissions

1. In the draft LLNA: BrdU-FC BRD, are there any errors in the BRD that should be corrected, or omissions of existing relevant data or information that should be included?

II. Questions to the Panel: Draft LLNA: BrdU-FC BRD

1. Test Method Protocol

- i. The LLNA: BrdU-FC includes routine measurements of ear swelling as an indicator of excessive dermal irritation. Do you consider this procedure to be an appropriate approach? Do you think that this measurement should be recommended for routine inclusion into all LLNA protocols? Please explain your answers.
- ii. The LLNA: BrdU-FC also includes optional quantification of immunophenotypic markers as an additional mechanism for distinguishing irritants from sensitizers. Do you consider this to be an appropriate approach to reduce false positives? Are the correct markers being considered or do you recommend other/additional markers? Should these measurements be recommended for routine inclusion in the LLNA: BrdU-FC? Please explain your answers.
- iii. Please comment on the appropriateness of the "sequential strategy" used in the eLLNA: BrdU-FC (see Figure 2-1 of the draft BRD).

2. Substances Used for the Validation Studies

- i. Do you consider the LLNA: BrdU-FC database representative of a sufficient range of chemical classes and physicochemical properties that it would be applicable to any of the types of chemicals and products that are typically tested for skin sensitization potential? If not, what are the relevant chemical classes/properties (other than those that are identified as limitations in the traditional LLNA) that should be tested with caution, or not evaluated using the LLNA: BrdU-FC? What chemicals or products should be evaluated to fill this data gap? Please explain your answers.

3. Test Method Accuracy

- i. The current accuracy analysis is based on overall concordance with the traditional LLNA. Accuracy statistics compared to the Guinea Pig tests and human data/experience have also been provided. Are these comparisons appropriate for assessing the accuracy of the LLNA: BrdU-FC? Please explain your answer.
- ii. Has the relevance (e.g., accuracy/concordance, sensitivity, specificity, false positive and false negative rates) of the LLNA: BrdU-FC been adequately evaluated and compared to the traditional LLNA (refer also to Table 6-1 of

- the draft ICCVAM BRD)? If not, what other analyses should be performed? Please explain your answer.
- iii. Three substances (benzalkonium chloride, resorcinol, and Tween 80) produced a “false positive” response compared to the traditional LLNA and guinea pig test when tested using the LLNA: BRDU-FC (Based on immunophenotyping, benzalkonium chloride was subsequently classified as an irritant rather than a sensitizer). Can you identify any characteristics associated with these or similar substances, compared to the correctly identified non-sensitizers that might signal that this type of discordant response might occur, and therefore using the LLNA: BrdU-FC to test such substances would not be appropriate, or that positive results for substances with such properties may warrant additional testing? Please explain your answer.
4. Test Method Reliability (Intra- and Inter-laboratory reproducibility)
- i. Has the intralaboratory reproducibility of the LLNA: BrdU-FC been adequately evaluated and compared to the traditional LLNA (refer also to Table 7-1 of the draft LLNA: BrdU-FC BRD)? If not, what other analyses should be performed? Are any limitations apparent based on this intra-laboratory reproducibility assessment? Please explain your answer.
 - ii. The draft LLNA: BrdU-FC BRD analyzes data from repeat testing of hexyl cinnamic aldehyde (HCA) in six different vehicles and intralaboratory reproducibility is assessed by coefficient of variation (CV). The calculated CVs ranged from 30% to 53%. Based on these data, are there concerns with the intralaboratory reproducibility of the LLNA: BrdU-FC? Please explain your answer.
5. Data Quality
- i. The studies evaluated in the draft BRD for the LLNA: BrdU-FC were not all conducted in accordance with Good Laboratory Practices (GLP) guidelines although there were done in a laboratory that routinely conducts GLP studies (G. DeGeorge, personal communication). Please discuss what impact this might have on the evaluation of the LLNA: BrdU-FC.
 - ii. The original records for these studies were requested but have not yet been obtained. As a result, an independent audit could not be conducted to confirm that the reported data is the same as the data recorded in laboratory notebooks. Do you agree that any recommendations from ICCVAM should be contingent upon the completion of such an audit and findings that there were no significant errors in data transcription? Please explain your answer.
Consideration of all available data and relevant information
 - iii. Based on the draft LLNA: BrdU-FC BRD, have all the relevant data identified in published or unpublished studies that employ this test method been adequately considered? Are there other comparative test method data that were not considered in the draft BRD, but are available for consideration? If yes, please explain how to obtain such data.

III. Questions to the Panel: Draft ICCVAM Test Method Recommendations on the LLNA: BrdU-FC

1. Test Method Usefulness and Limitations
 - i. Do you agree that the available data and test method performance (accuracy and reliability) support the ICCVAM draft recommendations for the LLNA: BrdU-FC in terms of the proposed test method usefulness and limitations? Please explain your answer.
 - ii. If restrictions on using radioactive materials are present, should the LLNA: BrdU-FC be routinely recommended for hazard identification of skin sensitizing substances in lieu of having to possibly use guinea pig tests? Please explain your answer.
 - iii. Even if limitations in using radioactive materials are not present, should the LLNA: BrdU-FC procedure be routinely recommended for hazard identification of skin sensitizing substances instead of the traditional LLNA? If not, then why? Please explain your answer.
 - iv. Do the ICCVAM recommendations adequately address concerns associated with the false positive rate of 17% (3/18 substances) calculated for the LLNA: BrdU-FC? Are there other suggestions for additional such guidance or limitations that should be considered? Please explain your answer.
2. Test Method Protocol
 - i. Do you agree that the available data support the ICCVAM draft recommendations for the LLNA: BrdU-FC procedure in terms of the proposed test method standardized protocol? If not, then what recommendations would you make? Please explain your answer.
 - ii. Can the traditional LLNA limit dose procedure be applied to the LLNA: BrdU-FC? Please explain your answer.
3. Future Studies
 - i. Do you agree that the available data support the ICCVAM draft recommendations for the LLNA: BrdU-FC in terms of the proposed future studies? What other recommendations would you make? Please explain your answer.
4. Performance Standards
 - i. The LLNA: BrdU-FC protocol differs from the ICCVAM-recommended protocol for the traditional LLNA (ICCVAM 1999; Dean et al. 2001) in the method used to assess lymphocyte proliferation in the auricular lymph nodes. According to the proposed draft ICCVAM Performance Standards for the traditional LLNA (http://iccvam.niehs.nih.gov/methods/immunotox/llna_PerfStds.htm), any change to the LLNA protocol other than the method used to assess lymphocyte proliferation is considered a major change. Do you agree that protocol differences between the LLNA: BrdU-FC and the traditional LLNA

should be considered only minor changes and therefore the validity of this test method should be based only on the draft ICCVAM Performance Standards? Please explain your answer.

- ii. Even if the draft ICCVAM Performance Standards do not apply to the LLNA: BrdU-FC, what impact should the accuracy analysis based on 13 of the 18 required performance standards substances have on the overall evaluation of test method accuracy? Please explain your answer.
- iii. Are there concerns that 3/6 sensitizers, for which EC3 data were available, had EC3 values that were outside of the proposed 0.5x to 2.0x EC3 acceptability range developed based on the traditional LLNA? Please explain your answer.

[This Page Intentionally Left Blank]

Appendix B5

Questions for the Peer Review Panel: Non-Radioactive LLNA Protocol - LLNA: BrdU-ELISA Test Method

[This Page Intentionally Left Blank]

Instructions for the Peer Review Panel: Non-Radioactive LLNA Protocol: LLNA: BrdU-ELISA Test Method

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) in conjunction with National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) is evaluating the validation status of the LLNA: BrdU-ELISA (Local Lymph Node Assay with bromodeoxyuridine [BrdU] detected by ELISA) for assessing the allergic contact dermatitis potential of chemicals and other substances. This test method, developed by Dr. Masahiro Takeyoshi (Tokyo, Japan), is a non-radiolabeled version of the traditional LLNA based on measuring levels of incorporated BrdU in the auricular lymph nodes as an indicator of increased cell proliferation.

The following materials are intended to guide you in your review of the ICCVAM evaluation of the LLNA: BrdU-ELISA. You are first asked to review the information in the draft ICCVAM LLNA: BrdU-ELISA Background Review Document (BRD) for completeness, and to identify any errors or omissions of existing relevant data or information that should be included. You are then asked to evaluate the information in the BRD to determine the extent to which each of the applicable criteria for validation and acceptance of toxicological test methods (ICCVAM 2003⁹) have been appropriately addressed for the proposed use of the LLNA: BrdU-ELISA. Adequate validation¹⁰ is a prerequisite for a test method to be considered for use in regulatory decision-making by U.S. Federal agencies. The validation process characterizes the usefulness and limitations of a test method for a specific intended use.

Lastly, you are asked to consider the ICCVAM draft test method recommendations for the LLNA: BrdU-ELISA (i.e., the proposed test method use, the proposed recommended standardized protocol, the proposed test method performance standards, and any proposed additional studies) and comment on whether the recommendations are supported by the information provided in the draft LLNA: BrdU-ELISA BRD.

The questions relating to the draft BRD that must be addressed are provided in **Sections I** and **II** of this guidance, while **Section III** contains questions relating to the draft ICCVAM test method recommendations on the LLNA: BrdU-ELISA.

These include questions prepared by the ICCVAM Immunotoxicity Working Group (IWG) to ensure that the assessment provides adequate information to facilitate U.S. Federal agency decisions on the regulatory acceptability of this test method. The questions are also intended to obtain guidance that will be helpful to federal agencies and other organizations that may be involved in conducting or supporting further development, standardization, and/or validation studies.

⁹ 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. NIEHS (Available: <http://iccvam.niehs.nih.gov/docs/guidelines/subguide.htm>).

¹⁰ Validation is the process by which the reliability and accuracy of a test method are established for a specific purpose (ICCVAM 2003).

The overall question to consider is whether the validation status of the LLNA: BrdU-ELISA has been adequately characterized for its intended purpose, and is it sufficiently accurate and reliable to be used for the identification of sensitizing substances and non-sensitizing substances in place of the traditional LLNA procedure.

I. Questions to the Panel: Comments on the Draft LLNA: BRDU-ELISA BRD for Errors and Omissions

1. In the draft LLNA: BrdU-ELISA BRD, are there any errors in the BRD that should be corrected, or omissions of existing relevant data or information that should be included?

II. Questions to the Panel: Draft LLNA: BrdU-ELISA BRD

1. Test Method Protocol
 - i. The data generated for the substances analyzed in the LLNA: BrdU-ELISA test method come from auricular lymph nodes from four individual mice in each dose group. The recommended ICCVAM LLNA protocol and OECD Test Guideline 429 recommend a minimum of five animals per dose group for collecting individual animal data. What impact might the use of four animals per dose group have on the accuracy and reliability of the LLNA: BrdU-ELISA? Do you agree with the ICCVAM recommendation that future use of this test method protocol should include five animals per dose group? Please explain your answer.
2. Substances Used for the Validation Studies
 - i. Do you consider the LLNA: BrdU-ELISA database representative of a sufficient range of chemical classes and physicochemical properties that the test method would be applicable to any of the types of chemicals and products that are typically tested for skin sensitization potential? If not, what are the relevant chemical classes/properties (other than those that are identified as limitations in the traditional LLNA) that should be tested with caution, or not evaluated using the LLNA: BrdU-ELISA? What chemicals or products should be evaluated to fill this data gap? Please explain your answers.
3. Test Method Accuracy
 - i. The current accuracy analysis using a Stimulation Index (SI) ≥ 3 or SI ≥ 1.3 to identify sensitizers is based on overall concordance with the traditional LLNA. Accuracy statistics compared to the Guinea Pig tests and human data/experience have also been provided. Are these comparisons appropriate for assessing the accuracy of the LLNA: BrdU-ELISA? Please explain your answer.
 - ii. Takeyoshi et al. (2007) performed an accuracy analysis using decision criteria other than SI ≥ 3 to classify substances as sensitizers. Maximal accuracy for the LLNA: BrdU-ELISA occurred when an SI ≥ 1.3 was used to distinguish between sensitizers and non-sensitizers. Using this decision criteria, they achieved an accuracy of 91% (21/23), with a sensitivity of 100% (16/16) and a specificity of 71% (5/7) (i.e., there were no false negatives and two false

- positives). Does this analysis support a recommendation that the decision criteria be based on an $SI \geq 1.3$? Are there concerns with using a small increase (i.e., 1.3-fold) above the vehicle control response as the basis for identifying a positive response? Please explain your answers.
- iii. Has the relevance (e.g., accuracy/concordance, sensitivity, specificity, false positive and false negative rates) of the LLNA: BrdU-ELISA, using the $SI \geq 3$ criterion, been adequately evaluated and compared to the traditional LLNA (refer also to Table 6-1 of the draft ICCVAM BRD)? If not, what other analyses should be performed? Please explain your answer.
 - iv. Using the $SI \geq 3$ criterion, there were four substances (aniline, 4-chloroaniline, 2-mercaptobenzothiazole, and hydroxycitronellal) when tested using the LLNA: BrdU-ELISA that produced “false negative” responses compared to the traditional LLNA. 4-Chloroaniline and aniline are amines. 2-Mercaptobenzothiazole is a heterocyclic compound and hydroxycitronellal is a hydrocarbon. 2-Mercaptobenzothiazole is a liquid, but the other three substances are solids. Can you identify any characteristics associated with these or similar substances, compared to the correctly identified sensitizers, that might signal that this type of discordant response might occur, and therefore using the LLNA: BrdU-ELISA to test such substances would not be appropriate or that negative results for such substances should indicate a need for confirmatory testing? Please explain your answer.
4. Test Method Reliability (Intra- and Inter-laboratory reproducibility)
- i. Has the intralaboratory reproducibility of the LLNA: BrdU-ELISA been adequately evaluated and compared to the traditional LLNA (refer also to Tables 7-1 through 7-3 of the draft LLNA: BrdU-ELISA BRD)? If not, what other analyses should be performed? Are any limitations apparent based on this intra-laboratory reproducibility assessment? Please explain your answers.
 - ii. The substances evaluated for intralaboratory reproducibility of the LLNA: BrdU-ELISA study were not coded. Does the lack of coding of test substances adversely impact or bias the current evaluation? Please explain your answer.
 - iii. The Japanese Center for the Validation of Alternative Methods (JaCVAM) has implemented a multi-laboratory validation study of the LLNA: BrdU-ELISA. Although the results from this study have yet to be reported, we are hoping to obtain information on the study design (i.e., with regard to number and types of chemicals tested and the number of laboratories involved). If we do, do you consider the design appropriate to adequately determine the extent of interlaboratory reproducibility for the LLNA: BrdU-ELISA? If not, what other analyses should be performed? Are any limitations apparent based on this study design? Please explain your answer.
5. Data Quality
- i. The studies evaluated in the draft BRD for the LLNA: BrdU-ELISA were not conducted in strict accordance with all provisions of the Good Laboratory Practice (GLP) guidelines, although there were reportedly performed in

laboratories that conduct GLP studies (M. Takeyoshi, personal communication). Please discuss what impact this might have on the evaluation of the LLNA: BrdU-ELISA.

- ii. The original records for these studies were requested but were not available. As a result, an independent audit could not be conducted to confirm that the reported data in peer reviewed publications and a poster presentation is the same as the data recorded in laboratory notebooks. Should any recommendations from ICCVAM be contingent upon the completion of such an audit and findings that there were no significant errors in data transcription? Please explain your answer.
6. Consideration of all available data and relevant information
 - i. Based on the draft LLNA: BrdU-ELISA BRD, have all the relevant data identified in published or unpublished studies that employ this test method been adequately considered? Are there other comparative test method data that were not considered in the draft BRD, but are available for consideration? If yes, please explain how to obtain such data.

III. Questions to the Panel: Draft ICCVAM Test Method Recommendations on the LLNA: BrdU-ELISA

1. Test Method Usefulness and Limitations

- i. Do you agree that the available data and test method performance (accuracy and reliability) support the ICCVAM draft recommendations for the LLNA: BrdU-ELISA in terms of the proposed test method usefulness and limitations? Please explain your answer.
- ii. If restrictions on using radioactive materials are present, should the LLNA: BrdU-ELISA be routinely recommended for hazard identification of skin sensitizing substances in lieu of using guinea pig tests due to the advantages of fewer animals and the avoidance of pain and distress? Please explain your answer.
- iii. Even if limitations in using radioactive materials are not present, should the LLNA: BrdU-ELISA procedure or other valid and accepted non-radioactive method be routinely recommended for hazard identification of skin sensitizing substances instead of the traditional LLNA? Please explain your answer.
- iv. Does using a decision criterion of $SI \geq 1.3$ instead of $SI \geq 3.0$ resolve any concerns with respect to potential false positives or false negatives that may occur in this test method? Are there other suggestions for additional such guidance or limitations that should be considered? Please explain your answer.

2. Test Method Protocol

- i. Do you agree that the available data support the ICCVAM draft recommendations for the LLNA: BrdU-ELISA procedure in terms of the proposed test method standardized protocols? If not, then what recommendations would you make? Please explain your answer.
- ii. Can the traditional LLNA limit dose procedure be applied to the LLNA: BrdU-ELISA? Please explain your answer.

3. Future Studies

- a. Do you agree that the available data support the ICCVAM draft recommendations for the LLNA: BrdU-ELISA in terms of the proposed future studies? If not, then what recommendations would you make? Please explain your answer.
- i. Performance Standards
- i. The LLNA: BrdU-ELISA protocol differs from the ICCVAM-recommended protocol for the traditional LLNA (ICCVAM 1999; Dean et al. 2001) in the method used to assess lymphocyte proliferation in the auricular lymph nodes. According to the proposed draft ICCVAM Performance Standards for the traditional LLNA (http://iccvam.niehs.nih.gov/methods/immunotox/llna_PerfStds.htm), any change to the LLNA protocol other than the method used to assess

lymphocyte proliferation is considered a major change. Do you agree that protocol differences between the LLNA: BrdU-ELISA and the traditional LLNA should be considered only minor changes and therefore the validity of this test method should be based only on the draft ICCVAM Performance Standards? Please explain your answer.

- ii. Even if the draft ICCVAM Performance Standards do not apply to the LLNA: BrdU-ELISA, what impact should the accuracy analysis based on eight of the 18 required performance standards substances (only one false negative and no false positives) have on the overall evaluation of test method accuracy? Please explain your answer.
- iii. Are there concerns that 4/4 sensitizers, for which EC3 data were available, had EC3 values that were outside of the recommended 0.5x to 2.0x EC3 acceptability range developed based on the traditional LLNA? Please explain your answer.
- iv. Should separate performance standards be developed for the LLNA: BrdU-ELISA? Please explain your answer.

Appendix B6

Questions for the Peer Review Panel: Draft ICCVAM Performance Standards for the LLNA

[This Page Intentionally Left Blank]

Instructions for the Peer Review Panel: Draft ICCVAM Performance Standards for the LLNA

ICCVAM has developed draft LLNA performance standards consisting of essential test method components, a minimum list of reference substances, and expected accuracy and reliability values. These are proposed for evaluating the acceptability of proposed test methods that are mechanistically and functionally similar to the traditional LLNA test method protocol previously recommended by ICCVAM.

The overall question for the Panel is whether these performance standards are considered adequate for assessing the accuracy and reliability of test method protocols that are based on similar scientific principles and that measure the same biological effect as the traditional LLNA.

1) Purpose and Applicability

- a) ICCVAM proposes that these performance standards should only be applicable to versions of the LLNA that incorporate minor modifications to the traditional LLNA. Currently, this is limited to the use of non-radioactive reagents to measure lymphocyte proliferation. It is considered essential that the modified LLNA should otherwise adhere to all other aspects of the traditional LLNA protocol, as defined by ICCVAM (1999) and Dean et al. (2001). This includes aspects such as: the sex and strain of mouse used, the number of mice per dose group, the timing and site of test article treatment, the duration between the last treatment and lymph node collection, the inclusion of concurrent negative and positive control groups, the measured endpoint (i.e., lymphocyte proliferation in the draining auricular lymph node), and the collection of data at the level of the individual mouse. Do you agree that the use of non-radioactive reagents for measuring cell proliferation in the lymph nodes, if that is the only difference, constitutes a minor modification to the traditional LLNA protocol? Is it necessary to keep the same decision criteria for distinguishing between sensitizers and non-sensitizers (i.e., an SI of 3)? Please explain your answer.
- b) Are there other procedural modifications that you consider minor and therefore could be evaluated for equivalence to the traditional LLNA using the proposed performance standards? If yes, please explain what they are and why.
- c) Do you consider these performance standards to also be applicable to the LLNA limit dose procedure? Please explain your answer.

2) Essential Test Method Components

- a) The essential test method components are based on the ICCVAM recommended protocol (ICCVAM 1999; Dean et al. 2001), which is the basis for the current U.S. Environmental Protection Agency (EPA 2003) test guideline (TG). There are some notable differences between these protocols and the Organisation for Economic Co-operation and Development TG 429 for the LLNA (OECD 2002). When evaluations of non-radioactive versions of the traditional LLNA are conducted using these performance standards, is it necessary that the validation studies follow the ICCVAM recommended protocol? Specifically, should the studies include: 1) a concurrent positive control with each test substance; 2) using a minimum of five animals per dose group; and 3) measuring proliferation in lymph nodes from individual animals

rather than pooling lymph nodes across all animals in a dose group? Please explain your answers.

- b) Should the concurrent testing of the positive control and test substance be conducted in the same vehicle or can different vehicles be used? Please explain your answer.

3) Proposed Reference Substances

- a) Do you agree with the selection and prioritization criteria used to select the performance standards reference substances? Please explain your answer.
- b) The rationale for the number of substances included on the "required" list of substances (n=18) is provided in Appendix C of the draft ICCVAM Performance Standards. Do you consider this to be an adequate number upon which to evaluate the performance of non-radioactive LLNA test methods, where the only protocol modification is the method for assessing cell proliferation in the auricular lymph nodes? If not, how many reference chemicals should be tested? Please explain your answer.
- c) Do you consider the types of substances included in the reference substance list, with regard to relative sensitization potency, physicochemical characteristics, and vehicles, to be representative of the overall diversity of substances that are likely to be tested for skin sensitization? Please explain your answer.
- d) Are there other types of information relevant to skin sensitization that should be considered in order to demonstrate an adequately diverse reference list? If yes, please explain what additional information should be included.
- e) Are there other substances that you consider to be more appropriate for assessing the sensitivity (ability of the test method to correctly identify sensitizing substances) and specificity (ability of the test method to correctly identify non-sensitizing substances) of non-radioactive LLNA test methods, and for which there is available LLNA, guinea pig, and human data? If yes, please name the substances and explain why.
- f) Four "discordant chemicals" (i.e., two LLNA false negatives and two false positives compared to guinea pig tests or human data) are included as optional substances that could be studied to evaluate if the proposed modifications might provide improved performance relative to the traditional LLNA.
- Please comment on the appropriateness of including these specific substances in the reference list. Should different substances be included? Should more false negative/positive substances be tested? If so, what are they? Please explain your answers.
 - Do you consider their "optional" status appropriate, or should testing these substances be required? Please explain your answer.
 - Would "correct" results with these four discordant chemicals be sufficient to consider the alternative test method to be more predictive of skin sensitization than the traditional LLNA? Please explain your answer.

4) Test Method Accuracy Standards

- a) The draft ICCVAM Performance Standards state that the non-radioactive proposed LLNA test method should exactly match the accuracy of the traditional LLNA when evaluated with the minimum set of 18 reference substances. Do you agree that test method accuracy should be based on a chemical-by-chemical match with regard to identifying the chemicals as sensitizers or non-sensitizers? Please explain your answer.
- b) The draft ICCVAM LLNA Performance Standards recommend that, for each sensitizer, the threshold concentration that induces a positive SI response should be within 0.5x to 2.0x of the concentration obtained for the EC3 in the traditional LLNA. As described in Appendix D of the draft ICCVAM Performance Standards, statistical approaches have been used in an attempt to identify an appropriate range, but these calculated ranges do not appear to be the most practical. In contrast, the NICEATM LLNA database demonstrates that EC3 values from replicate tests for a sensitizing chemical when tested using the same solvent are rarely outside of this proposed 0.5x to 2.0x acceptability EC3 range. Please comment on the appropriateness of using this criterion to judge the equivalency of a non-radioactive version of the traditional LLNA. If this approach is not acceptable, please explain why, and present an alternative approach along with the basis for this approach.
- c) For five of the 13 sensitizers on the draft ICCVAM reference substances list, the reference EC3 value is based on a single LLNA study (see Table C1 of the draft ICCVAM Performance Standards). Please comment on the appropriateness of including such chemicals in the list of recommended reference substances and whether or not the 0.5x to 2.0x criteria should be applied to such substances. Please explain your answer.

5) Test Method Reliability Standards

- a) The draft ICCVAM Performance Standards state that acceptable intralaboratory reproducibility will be indicated by a laboratory obtaining, in each of four independent experiments conducted with at least one week between each experiment, ECt values (the estimated concentration needed to produce an SI of a defined threshold [e.g., EC3]) for hexyl cinnamic aldehyde (HCA) that are generally within 0.5x to 2.0x (i.e., 5% to 20%) of the historical mean EC3 concentration (10%) for this substance, based on existing available traditional LLNA data.
 - Do you consider the number of repeat experiments (n=4) to be adequate? Please explain your answer.
 - Do you consider testing HCA adequate for demonstrating intralaboratory reproducibility? If not, which substance(s) should be tested? Please explain your answer.
 - Is the required one-week interval between independent tests adequate and/or appropriate? If not, please provide an alternative schedule and explain the basis for your recommendation.
 - Do you consider the criteria for acceptability to be appropriate? If not, please describe another criteria and explain the basis for your recommendation.

b) The draft ICCVAM Performance Standards state that acceptable interlaboratory reproducibility will be indicated by each of three laboratories obtaining EC_t values for HCA and 2,4-dinitrochlorobenzene (DNCB) from a single experiment that are generally within 0.5x to 2.0x (5% to 20% and 0.025 to 0.1%, respectively) of the mean historical EC₃ concentration (10% and 0.05%, respectively) obtained for these two substances in the traditional LLNA.

- Do you consider the single experiment per substance in each laboratory to be adequate? If not, please provide an alternative approach and explain why.
- Do you consider testing HCA and DNCB to be adequate for demonstrating interlaboratory reproducibility? If not, which substance(s) should be tested? Please explain your answer.
- Do you consider the criteria for acceptability to be appropriate? If not, please describe another criteria and explain the basis for your recommendation.

6) Summary Question

a) If a radioactive or non-radioactive LLNA method were proposed with a “major change” (e.g., different mouse strain or use of male mice, change in the schedule for test article administration, change in schedule for lymph node excision, etc.), what criteria should be used to evaluate the equivalence of this method to the traditional LLNA?

- Would a new set of performance standards be required for this method? Please explain your answer.
- How many reference substances might be considered adequate? Would the 18 minimum reference substances in the draft ICCVAM LLNA Performance Standards be sufficient? If more substances are considered necessary, how many should there be tested and what should their characteristics be? Please explain your answer.
- Regardless of the number of reference substances, should the alternative LLNA be required to obtain the same call (and potency for sensitizers) as the traditional LLNA for the 18 minimum reference substances in the draft ICCVAM LLNA Performance Standards? Please explain your answer.
- Are there additional specific substances that should be used? If yes, what are they? Please explain your answer.
- What, if any, additional information would be considered necessary and why?

Appendix B7

Questions for the Peer Review Panel: Use of the LLNA for Potency Determinations

[This Page Intentionally Left Blank]

Instructions for the Peer Review Panel: Use of the LLNA for Potency Determinations

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) in conjunction with National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) is evaluating the validation status of the murine local lymph node assay (LLNA) for assessing the potential potency of substances to cause allergic contact dermatitis.

The following materials are intended to guide you in your review of the ICCVAM evaluation of the LLNA for potency determinations. You are first asked to review the information in the draft ICCVAM LLNA for potency determinations Background Review Document (BRD) for completeness, and to identify any errors or omissions of existing relevant data or information that should be included. You are then asked to evaluate the information in this BRD to determine the extent to which each of the applicable criteria for validation and acceptance of toxicological test methods (ICCVAM 2003¹¹) have been appropriately addressed for the proposed use of the LLNA for potency determinations. Adequate validation¹² is a prerequisite for a test method to be considered for use in regulatory decision-making by U.S. Federal agencies. This validation process characterizes the usefulness and limitations of a test method for its intended use.

Lastly, you are asked to consider the ICCVAM draft test method recommendations for the LLNA for potency determinations (i.e., the proposed test method use, the proposed recommended standardized protocol, the proposed test method performance standards, and any proposed additional studies) and comment on whether the recommendations are supported by the information provided in the draft BRD.

The questions relating to the draft BRD that must be addressed are provided in **Sections I and II** of this guidance, while **Section III** contains questions relating to the draft ICCVAM test method recommendations on the LLNA for potency determinations.

These include questions prepared by the ICCVAM IWG to ensure that the assessment provides adequate information to facilitate U.S. Federal agency decisions on the regulatory acceptability of this test method, and/or adequate guidance for organizations that may be involved in conducting or supporting further development, standardization, and/or validation.

The overall question to consider is whether the validation status of the LLNA for potency determinations has been adequately characterized, and is it sufficiently accurate and reliable to be used for the identification of sensitizing substances according to their relative potency classification based on a comparison to either human or guinea pig responses.

I. Questions to the Panel: Review for Errors and Omissions

1. Are there any errors or omissions of existing relevant data or information that should be included in the draft BRD?

¹¹ 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. NIEHS (Available: <http://iccvam.niehs.nih.gov/docs/guidelines/subguide.htm>).

¹² Validation is the process by which the reliability and accuracy of a test method are established for a specific purpose (ICCVAM 2003).

II. Questions to the Panel: LLNA for Potency Determinations Draft BRD

1. Do you consider the database of substances evaluated representative of a sufficient range of chemical classes and physical chemical properties that it would be applicable to any of the types of chemicals and products that are typically tested for skin sensitization potential? If not, what are the relevant chemical classes/properties (other than those that are identified as limitations in the traditional LLNA) that should be tested with caution, or not evaluated using the for potency determinations? What chemicals or products should be evaluated to fill this data gap? Please explain your recommendation.
2. While coding of chemicals is recommended for prospective studies, this evaluation is based on a retrospective evaluation of existing data, most of which was not generated using coded chemicals to reduce bias. Does the lack of coding of test substances adversely impact or bias the current evaluation? Please provide a rationale for your answer.
3. For some substances submitted using the LLNA, it was not possible to confirm whether the data were generated using pooled animal data for each dose group (as allowed in Test Guideline [TG] 429 of the Organisation for Economic Co-operation and Development [OECD]) rather than individual animal data (as recommended in the ICCVAM 2001 protocol)? Cockshott et al. (2006) reported that using individual animal data allowed for technical problems during an experiment to be identified. Considering this, should the analysis of the performance of the LLNA for potency determinations be limited to data from studies that can be confirmed as using individual animal data collection procedures? What impact might the inclusion of pooled animal data have on the accuracy analysis included in Section 6.0 of the draft ICCVAM BRD? Please explain your answer.
4. Has the relevance (e.g., accuracy/concordance, sensitivity, specificity, false positive and false negative rates) of the LLNA for potency determinations been adequately evaluated and compared to the human and guinea pig (refer also to Section 6.0 of the draft ICCVAM BRD)? If not, what other analyses should be performed?
5. The accuracy analysis (see Section 6.0 of the draft ICCVAM BRD) focuses on the two-level categorization scheme proposed by the United Nations Globally Harmonized System for Classification and Labelling for both human and guinea data. Should other categorization schemes be considered?
6. Does the BRD adequately characterize the usefulness and limitations of the LLNA for potency determinations based on the accuracy analyses? If not, what additions or changes should be made to the current usefulness and limitations?
7. Has the reliability (e.g., intralaboratory repeatability, intra- and inter-laboratory reproducibility) of the LLNA for potency determinations been adequately evaluated (refer also to Section 7.0 of the draft ICCVAM BRD)? If not, what other analyses should be performed?

8. For some studies included in the draft BRD, it was not possible to determine whether or not they had been conducted in accordance with Good Laboratory Practice (GLP) guidelines. Neither was it possible to obtain the results of GLP audits for all studies conducted in accordance with GLP guidelines. Please discuss what impact this lack might have on the evaluation of the LLNA for potency determinations and whether such studies should be excluded from any analysis.
9. As described in the draft BRD, original records for some of the non-GLP studies included in this evaluation could not be obtained. As a result, an independent audit could not be conducted to confirm that the reported data is the same as the data recorded in laboratory notebooks. Considering this, should the results of these studies (all of which are currently included) be excluded from any of the performance analyses? If yes, please explain.
10. Based on the draft BRD, have all the relevant data identified in published or unpublished studies conducted using the LLNA for potency determinations been adequately considered? If not, what other studies should to be considered?

III. Questions to the Panel: Draft ICCVAM Test Method Recommendations on the LLNA for Potency Determinations

1. Do you agree that the available data support the ICCVAM draft recommendations for the LLNA for potency determinations in terms of the proposed test method usefulness and limitations? Please explain your answer. If not, why recommendations would you make?
 - Should the LLNA be routinely recommended for the hazard classification of the skin sensitization potency of chemicals?
2. Do you agree that the available data support the ICCVAM draft recommendations for the LLNA in terms of the proposed test method standardized protocol? Please explain your answer. If not, then what recommendations would you make?
3. Should the relevant testing guidelines for the LLNA be updated to include the calculation of an EC3 value?
4. Do you agree that the available data support the ICCVAM draft recommendations for the LLNA in terms of the proposed future studies? Please explain your answer. If not, then what recommendations would you make?