Independent Scientific Peer Review Panel Report:

Updated 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

June 2009

Interagency Coordinating Committee on the Validation of Alternative Methods

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

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

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/LLNAPRPRept2009.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.

When referencing this document, please cite as follows:

Interagency Coordinating Committee on the Validation of Alternative Methods. 2009. Independent Scientific Peer Review Panel Report: Updated 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. Research Triangle Park, NC: National Institute of Environmental Health Sciences.

Table of Contents

List of Abbreviations and Acronyms

ACD	Allergic contact dermatitis
ANOVA	Analysis of variance
AOO	Acetone: olive oil (4:1)
ATP	Adenosine triphosphate
BIBRA	British Industrial Biomedical Research Association
BRD	Background review document
BrdU	Bromodeoxyuridine
CV	Coefficient of variation
DNCB	2,4-Dinitrochlorobenzene
EC2	Estimated concentration of a substance needed to produce a stimulation index of 2 (value is expressed as a percentage)
EC2.5	Estimated concentration of a substance needed to produce a stimulation index of 2.5 (value is expressed as a percentage)
EC3	Estimated concentration of a substance needed to produce a stimulation index of 3 (value is expressed as a percentage)
ECt	Estimated concentration of a substance needed to produce a stimulation index that is indicative of a positive response (value is expressed as a percentage)
ECVAM	European Centre for the Validation of Alternative Methods
ELISA	Enzyme-linked immunosorbent assay
eLLNA: BrdU-FC	Enhanced LLNA: BrdU detected by flow cytometry
EPA	U.S. Environmental Protection Agency
ESAC	ECVAM Scientific Advisory Committee
E.U.	European Union
FR	*Federal Register*
GLP	Good Laboratory Practice
GPMT	Guinea pig maximization test
HCA	Hexyl cinnamic aldehyde
HMT	Human maximization test
HRIPT	Human repeat insult patch test
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
ISO	International Organization for Standardization
IWG	ICCVAM Immunotoxicity Working Group
JSAAE	Japanese Society for Alternatives to Animal Experiments
LLNA	Murine local lymph node assay

LLNA: BrdU-ELISA	LLNA: BrdU detected by ELISA
LLNA: BrdU-FC	LLNA: BrdU detected by flow cytometry
LLNA: DA	LLNA: Daicel adenosine triphosphate
MRC	U.K. Medical Research Council
NAS	National Academy of Sciences
NICEATM	NTP Interagency Center for the Evaluation of Alternative Toxicological Methods
NIEHS	National Institute of Environmental Health Sciences
NIH	National Institutes of Health
NIOSH	National Institute of Occupational Safety and Health
NTP	National Toxicology Program
OECD	Organisation for Economic Co-operation and Development
QSAR	Quantitative structure–activity relationship
REACH	Registration, Evaluation, and Authorisation of Chemicals
RIFM	Research Institute for Fragrance Materials
rLLNA	Reduced LLNA
SACATM	Scientific Advisory Committee for the Validation of Alternative Toxicological Methods
SD	Standard deviation
SI	Stimulation index
SLS	Sodium lauryl sulfate
SOT	Society of Toxicology
UCSF	University of California, San Francisco

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., Scientific Coordinator on Alternatives Methods in Life Science, L'Oréal Research and Development, Aulnay sous Bois, 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

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

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

Raymond Pieters, Ph.D.,[1] 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, 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

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

Michael Woolhiser, Ph.D., Science and Technology Leader, Toxicology and Environmental Research and Consulting, The Dow Chemical Company, Midland, MI

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

[1] Dr. Pieters was unable to attend the public meeting on April 28-29, 2009. However, he was involved in the review of the revised draft background review documents and the revised draft Addendum and concurs with the conclusions and recommendations included in this report.

[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 LLNA was subsequently incorporated into national and international test guidelines for the assessment of skin sensitization (OECD 2002; ISO 2002; EPA 2003). (This LLNA will be referred to hereafter as the "traditional" LLNA.)

In January 2007, the U.S. Consumer Product Safety Commission formally requested 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 reduced LLNA (rLLNA; also referred to as the LLNA limit dose procedure)
- 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)

The National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), in coordination with ICCVAM and the ICCVAM Immunotoxicity Working Group (IWG), prepared comprehensive draft background review documents (BRDs) 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. 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 supporting documents and the draft ICCVAM recommendations were provided to a new international Panel for an independent scientific review. This Panel met in public session in

[2] The U.S. Consumer Product Safety Commission nomination can be obtained at: http://iccvam.niehs.nih.gov/methods/immunotox/llnadocs/CPSC_LLNA_nom.pdf.

March 2008.[3] Subsequent to the Panel review, finalized recommended performance standards for the LLNA and ICCVAM recommendations for the rLLNA were published.[4] The final documents considered the comments of the Panel, the public, and ICCVAM's scientific advisory panel.

The Panel concluded in March 2008 that more information and data were required for the three modified nonradioactive LLNA test methods before recommendations could be made regarding their use for regulatory safety testing (ICCVAM 2008). Similarly, the Panel concluded that more data would be needed before a recommendation on the usefulness and limitations of the current applicability domain of the traditional LLNA could be made. Subsequent to the Panel meeting, NICEATM received additional LLNA data for pesticide formulations and other products, as well as new data for the three modified nonradioactive LLNA test methods.

Using the additional information and working in coordination with the IWG, NICEATM revised the BRDs for each of these modified test methods and new applications of the LLNA. The revised draft BRDs provide the data and analyses supporting the scientific validity of the modified test methods and proposed applications. ICCVAM also prepared revised draft test method recommendations regarding proposed usefulness and limitations, standardized protocols, and future studies.

The revised draft BRDs, the revised draft applicability domain addendum, and revised draft ICCVAM recommendations were provided to the Panel for independent scientific review. In addition, NICEATM announced the availability of these documents on the NICEATM – ICCVAM website for public comment in a *Federal Register* (FR) notice (74 FR 8974) and via the ICCVAM email list. The FR notice also announced the public Panel meeting, to be convened at the National Institutes of Health in Bethesda, Maryland, on April 28 – 29, 2009.

The Panel was charged with:

- Reviewing each revised draft BRD and the revised draft addendum for completeness, and identifying any errors or omissions of existing relevant data or information
- Evaluating the information in each revised draft BRD and the revised draft addendum to determine the extent to which each of the applicable criteria for validation and acceptance of toxicological test methods (ICCVAM 2003) had

[3] The conclusions and recommendations of the Panel are included in its report, which is available at: http://iccvam.niehs.nih.gov/docs/immunotox_docs/LLNAPRPRept2008.pdf.

[4] The *Recommended LLNA Performance Standards* document is available at: http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna-ps/LLNAPerfStds.pdf; the ICCVAM

been appropriately addressed for the recommended use of the new versions and applications of the traditional LLNA

- Considering the ICCVAM revised draft test method recommendations for the following, and commenting on the extent to which they are supported by the information provided in the revised draft BRDs and the revised draft addendum:
  - Proposed test method uses
  - Proposed recommended standardized protocols
  - Proposed test method performance standards
  - Proposed additional studies

During its public meeting in April 2009, the Panel discussed each charge, listened to public comments, and developed conclusions and recommendations for ICCVAM. The Panel emphasizes that it was asked to consider two overall questions. The Panel was to consider: (1) whether the validation status of 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,[5] 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 nonsensitizing 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 the Panel's review. The Panel also thanks each of the test method developers, Drs. George DeGeorge (LLNA: bromodeoxyuridine detected by flow cytometry test method), Kenji Idehara (LLNA: Daicel adenosine triphosphate test method), and Masahiro Takeyoshi, (LLNA: bromodeoxyuridine detected by ELISA) for providing summaries and additional clarifications of the

---

recommendations for the rLLNA are in the *ICCVAM Test Method Evaluation Report*, available at: http://iccvam.niehs.nih.gov/docs/immunotox_docs/LLNA-LD/TMER.pdf.

[5] ICCVAM validation criteria are detailed in the document, *Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods,* available at http://iccvam.niehs.nih.gov/docs/about_docs/validate.pdf.

nonradioactive test methods under review. Finally, as Panel Chair, I 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
June 2009

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 application of the traditional LLNA for evaluating pesticide formulations and other products, metals, and substances in aqueous solutions
- Three modified versions of the traditional LLNA not requiring the use of radioactive markers:
  - LLNA: DA (LLNA: Daicel adenosine triphosphate)
  - LLNA: BrdU-FC (LLNA: bromodeoxyuridine detected by flow cytometry)
  - LLNA: BrdU-ELISA (LLNA: bromodeoxyuridine detected by ELISA)

Nonradioactive LLNA Protocol – The LLNA: DA Test Method

The Panel concluded that the available data and performance support the revised draft ICCVAM recommendations on usefulness and limitations for the LLNA: DA test method. They agreed that the test method could be used for identifying substances as potential skin sensitizers and nonsensitizers. On the basis of the available data, accuracy is optimized if a stimulation index (SI) ≥ 2.5 is used to identify sensitizers, and an SI ≤ 1.7 is used to identify nonsensitizers. A limitation of the LLNA: DA involves the indeterminate identification of substances with SI values between 1.7 and 2.5 (exclusive). Thus, when an SI between 1.7 and 2.5 is obtained in the LLNA: DA, users should carefully interpret the results in an integrated decision strategy in conjunction with all available and relevant information (e.g., dose response information, statistical analyses, peptide-binding activity, molecular weight, results from related chemicals, other testing data) to determine if there is adequate information for a definitive skin sensitization identification or if additional testing is necessary. The Panel noted that because the decision criteria chosen to identify sensitizers and nonsensitizers were based on a post hoc analysis, prospective testing with the test method might affect the proposed model. For this reason, data generated should be routinely evaluated to determine if the proposed model is still optimal with regard to the decision criteria. Even with these limitations, the LLNA: DA provides opportunities to reduce animal usage (e.g., use of guinea

pigs) in those regions in which guinea pig tests rather than the traditional LLNA are performed because radioisotope use is not permitted. In addition, the use of two decision criteria allows for a more definitive identification of sensitizers and nonsensitizers, which also provides animal welfare benefits by reducing further tests that might be required in instances where the hazard classification of a substance is not as clear.

The revised draft LLNA: DA background review document (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 ACD-inducing potential of chemicals and other products. The Panel evaluated the revised draft BRD for completeness, errors, and omissions, and recommended that its suggestions/corrections relating to general, statistical, and specific editorial issues be incorporated into future revisions.

The Panel agreed that the data supported the revised draft ICCVAM recommendations for the proposed standardized protocol for the LLNA: DA. The recommendations for maintaining a positive control database reflect current evidence and best practice. The Panel agreed that four animals per dose group should be recommended for the LLNA: DA.

The Panel considered the 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 indicated that the number of substances in the range of uncertainty was too few to determine if specific characteristics (e.g., chemical class, physical form, molecular weight, peptide reactivity, etc.) associated with those substances could be used for definitive skin sensitization identification.

With regard to test method reliability, the Panel concluded that the interlaboratory reproducibility of the LLNA: DA had been adequately evaluated. The Panel noted that five of the 10 laboratories that participated in the first phase of the interlaboratory validation study exceeded the performance standards' acceptable range for ECt values (estimated concentration of a substance needed to produce an SI that is indicative of a positive response) for 2,4-dinitrochlorobenzene (DNCB). The Panel indicated that this was understandable since DNCB is a strong sensitizer and the LLNA: DA has a different dosing regimen and time course than the traditional LLNA, which might extend into the elicitation phase of skin sensitization. However, all the laboratories that participated in the first and second phase of the interlaboratory validation study obtained EC2.5 values (estimated concentration of a substance needed to produce an SI of 2.5) within the concentration range indicated for hexyl cinnamic aldehyde (HCA), which documents the test method's favorable reproducibility and performance.

The Panel stated that the available data supported the revised draft ICCVAM recommendations for the LLNA: DA in terms of future studies, which included performing more LLNA: DA studies on metals, irritants, and formulations with comparative traditional LLNA, guinea pig, and human data. Regarding irritants, the proposed future studies might help explain why results obtained using the LLNA: DA were discordant with the traditional LLNA and may even provide general insight into the problematic nature of discriminating irritants in the LLNA. The Panel also recommended that additional decision criteria and guidance should be identified for substances with SI greater than 1.7 but less than 2.5, and that the additional decision criteria be reassessed as additional discriminators and data become available (e.g., high-quality human ACD data). The Panel recommended that a protocol for defining and reevaluating the SI decision criteria for sensitizers and nonsensitizers be developed. Further, future interlaboratory validation studies should simultaneously evaluate intralaboratory reproducibility, using the appropriate statistics, to evaluate variation both within a laboratory and between laboratories. Additionally, the Panel strongly recommended that a statistician actively participate in the preparation of future BRDs and formulation of ICCVAM recommendations.

The Panel disagreed with the revised draft ICCVAM recommendation that separate performance standards be developed to assess modified versions of the LLNA: DA test method. Although the test methods differ in the dosing regimen and in the timing of the assay, the Panel viewed the LLNA: DA as mechanistically similar to the traditional LLNA, in that both methods measure cellular stimulation in the draining lymph nodes. Consequently, the Panel concluded that the ICCVAM-recommended LLNA performance standards (ICCVAM 2009) are applicable to the LLNA: DA as a mechanistically and functionally similar test method. 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 test progressed through the elicitation phase of skin sensitization, which is associated with a localized skin reaction. Thus, the Panel was concerned that if the duration of the test involved the elicitation phase of ACD development, this would produce undue discomfort and distress in the animals. The Panel also recommended that the test method developer (Daicel Chemical Industries, Ltd.) justify the use of 1% sodium lauryl sulfate (SLS) (i.e., determine whether the 1% SLS pretreatment is necessary).

Nonradioactive LLNA Protocol – The LLNA: BrdU-FC Test Method

The Panel concluded that the data and test method performance of the LLNA: BrdU-FC supported the revised draft ICCVAM recommendations that the test method may be useful for identifying substances as potential skin sensitizers or nonsensitizers, and agreed that formal recommendations should be deferred until original study records are received for an

independent audit and interlaboratory transferability and reproducibility have been assessed. The final test method recommendations should highlight those items of highest priority for further validation consideration: (1) a review of the original data at the individual animal level with appropriate positive and negative controls, (2) an evaluation, based on the data from the intralaboratory study data, of the minimum number of animals required per test group to ensure test performance is as good as or better than the traditional LLNA, then (3) an interlaboratory reproducibility study conducted and evaluated according to the specifications in the ICCVAM-recommended LLNA performance standards (ICCVAM 2009) and with appropriate quality control systems. The Panel agreed that, subsequently, less critical items (e.g., methodological specifics, immunophenotypic endpoints, alternative decision criteria for identifying materials as sensitizers and nonsensitizers) should then be evaluated.

The revised 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 chemicals and other products. The Panel evaluated the revised draft BRD for completeness, errors, and omissions, and recommended that its recommendations/corrections relating to general, statistical, and specific editorial issues be incorporated into future revisions.

The Panel agreed that the available data supported the revised draft ICCVAM recommendations for the proposed test method protocol for the LLNA: BrdU-FC procedure. Also, revised power calculations should be performed using the data provided for the intralaboratory performance to determine the minimum group size required to provide a level of test performance equivalent to or better than the traditional LLNA. The minimum group size in the protocol should then be adjusted, if necessary. The ICCVAM recommendation for maintaining a positive control database reflects current evidence and best practice. The Panel considered the measurement of ear swelling and the use of immunophenotypic markers as potentially valuable adjuncts to the traditional LLNA and other modified LLNA protocols.

The Panel noted that since the 2008 Panel evaluation no new data for additional test substances were added to the analyses in the revised draft BRD, although new data for intralaboratory reproducibility were properly integrated into the assessment. As such, similar to 2008, the substances tested in the LLNA: BrdU-FC seemed representative of a sufficient range of chemical classes and physical chemical properties, and thus the test method appeared applicable to many of the types of chemicals and products that are typically tested for skin sensitization potential. The results of the revised concordance assessments of the LLNA: BrdU-FC against the traditional LLNA test method suggest that the LLNA: BrdU-FC (as performed at the originating facility) can be developed as a reliable alternative to the

traditional LLNA, with the same applicability domain. Both the LLNA: BrdU-FC and the eLLNA: BrdU-FC ("enhanced" LLNA: BrdU-FC), on the basis of the information available, performed equally well compared with the traditional LLNA in a single laboratory.

The Panel concluded that compared to the 2008 review, intralaboratory reproducibility was adequately assessed and fit for the intended purpose. This was based on additional studies submitted for HCA and DNCB. The Panel agreed that the assessment of interlaboratory reproducibility described in the ICCVAM-recommended LLNA performance standards (ICCVAM 2009) can be appropriately applied to the LLNA: BrdU-FC test method.

The Panel affirmed that the revised draft ICCVAM recommendations for future studies highlighted the unanswered questions raised by the available data set. The Panel specifically recommended: (1) that an independent audit of the original data should be performed to establish the validity of the data relied upon in the revised draft BRD, (2) that revised power calculations should be performed using the data provided for the intralaboratory validation so that the number of animals needed to provide performance equivalent to, or better than, the traditional LLNA can be determined, (3) that an interlaboratory study is an absolute requirement for validation to determine the transferability and reliability of the test method when used in different laboratories, (4) that alternate prediction models (e.g., multiple SIs similar to those recommended for the LLNA: DA and LLNA: BrdU-ELISA test methods) should be considered, and (5) that the ICCVAM-recommended LLNA performance standards (ICCVAM 2009) should be followed in this future work. The Panel recommended that ICCVAM should work with NICEATM to support and facilitate these activities. The Panel also considered that an emphasis should be given to the use of ear swelling measurements to identify local irritants as a means of improving the traditional LLNA and modified LLNA test methods. This is particularly relevant when considering the challenges associated with discriminating irritants from sensitizers in the LLNA and ultimately emphasizes the need to better understand the correlation between mouse ear data and human data/experience.

It is the view of the Panel that this test method can be considered to have been scientifically validated and to be ready for regulatory consideration if the following requirements are satisfactorily met: (1) an independent data audit should be conducted confirming the acceptable quality of the data relied upon in the revised draft BRD, (2) a revised evaluation of the minimum number of animals required should be conducted; then, if n = 4 or 5 yields statistical power that is equivalent to or better than the traditional LLNA, an interlaboratory evaluation should be performed using the test, (3) the interlaboratory study should produce results that satisfy the requirements in the ICCVAM-recommended LLNA performance standards (ICCVAM 2009).

The Panel considered the LLNA: BrdU-FC and the traditional LLNA to be mechanistically and functionally similar. Thus, the studies proposed by the ICCVAM-recommended LLNA performance standards are sufficient to establish the intra- and interlaboratory performance of the LLNA: BrdU-FC. The Panel commented that for regulatory data submissions, a laboratory (either with flow cytometry experience and/or following training and certification of personnel) should demonstrate proficiency by repeating the evaluation of the same substance (i.e., four independent tests) to allow an assessment of intralaboratory reproducibility before using the test for regulatory purposes. Results should be evaluated for both a known strong and known moderate sensitizer (i.e., DNCB and HCA, respectively). The inclusion of a known, reproducible weak sensitizer and a negative control is also essential to confirm that the full range of appropriate responses can be reproduced.

Additional considerations would include development of a standard test method protocol, standard operating procedure, and other documentation, and adherence to recognized quality assurance/quality control systems for flow cytometry and associated data acquisition equipment.

Nonradioactive LLNA Protocol – The LLNA: BrdU-ELISA Test Method

The Panel concluded that the data and performance for the LLNA: BrdU-ELISA test method supported the revised draft ICCVAM recommendations that it can be used for identifying substances as potential skin sensitizers and nonsensitizers. An SI $\geq$ 2.0 should be used to identify substances as sensitizers and SI < 1.3 should be used to identify nonsensitizers. A limitation of the LLNA: BrdU-ELISA involves the indeterminate identification of substances that produce an SI greater than or equal to 1.3 but less than 2.0. When such a result is obtained in the LLNA: BrdU-ELISA, users should carefully interpret the results in an integrated decision strategy in conjunction with all available and relevant information (e.g., dose response information, statistical analyses, peptide-binding activity, molecular weight, results from related chemicals, other testing data) to determine if there is adequate information for definitive skin sensitization identification or if additional testing is necessary. The Panel noted that because the decision criteria chosen to identify sensitizers and nonsensitizers were based on post hoc analysis, prospective testing with the test method might affect the proposed model. For this reason, data generated should be routinely evaluated to determine if the proposed model is still optimal with regard to the decision criteria. Even with these limitations, the LLNA: BrdU-ELISA provides opportunities to reduce animal usage (e.g., use of guinea pigs) in those regions that are not permitted to use radioisotopes and thus perform guinea pig tests rather than the traditional LLNA. In addition, using two decision criteria allows for a more definitive identification of sensitizers and

nonsensitizers, which also provides animal welfare benefits by reducing further tests that might be required in instances where the hazard classification of a substance is not as clear.

The revised 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 products. The Panel evaluated the draft BRD for completeness, errors, and omissions and recommended that its suggestions/corrections relating to general, statistical and specific editorial issues be incorporated into the final document.

The Panel agreed that the available data supported the revised draft ICCVAM recommendations for the proposed standardized test method protocol for the LLNA: BrdU-ELISA test method. The recommendations for maintaining a positive control database reflect current evidence and best practice. The Panel agreed that four animals per dose group should be recommended for the LLNA: BrdU-ELISA.

The Panel considered the database of substances tested in the LLNA: BrdU-ELISA 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 indicated that the number of substances in the range of uncertainty (i.e., $1.3 \leq SI < 2.0$) was too few to determine if specific characteristics (e.g., chemical class, physical form, molecular weight, peptide reactivity, etc.) associated with those substances could be used for definitive skin sensitization identification.

In 2008, the Panel did not find sufficient power for using $SI \geq 1.3$ as the decision criterion. Even with a group size of eight animals, the power was only 50% (ICCVAM 2008). Power calculations might be necessary to determine if the sample size used is sufficient for those substances that are not definitively identified as sensitizers or nonsensitizers (i.e., substances in the range of uncertainty of $1.3 \leq SI < 2.0$).

With regard to test method reliability, the Panel concluded that the interlaboratory reproducibility had been adequately evaluated and that the test is reproducible. Considering that the radioisotope measurement in the traditional LLNA is more sensitive than the technique for the LLNA: BrdU-ELISA, and that the analysis of EC3 values (estimated concentration of a substance needed to produce a stimulation index of 3) in the traditional LLNA was based on a larger dataset, it is appropriate to adjust the acceptability range of the two positive control substances tested, dependent on the method used for measurement of the endpoint. Although the qualitative performance was acceptable in the interlaboratory study, the quantitative data for two of the laboratories suggests a relatively high degree of variability, which justifies the routine use of appropriate positive and negative controls.

The Panel stated that the available data supported the revised draft ICCVAM recommendations for the LLNA: BrdU-ELISA in terms of future studies, which included performing more LLNA: BrdU-ELISA studies on metals, irritants, and formulations with comparative traditional LLNA, guinea pig, and human data. Regarding irritants, the proposed future studies might help explain why results obtained using the LLNA: BrdU-ELISA and traditional LLNA were discordant, and further address the general challenge of discriminating irritants in the traditional LLNA itself. The Panel also recommended that additional decision criteria and guidance should be identified for substances that produce an SI greater than or equal to 1.3 but less than 2.0, and that the additional decision criteria be reassessed as additional discriminators and data become available (e.g., high-quality human ACD data). The Panel recommended that a protocol for defining and reevaluating the SI decision criteria for sensitizers and nonsensitizers be developed. Further, future interlaboratory validation studies should simultaneously evaluate intralaboratory reproducibility, using the appropriate statistics, to evaluate variation both within a laboratory and between laboratories. As stated previously, the Panel strongly recommended that a statistician actively participate in the preparation of future BRDs and formulation of ICCVAM recommendations.

The Panel agreed with the revised draft ICCVAM recommendation that separate performance standards should not be developed to assess modified versions of the LLNA: BrdU-ELISA test method. The LLNA: BrdU-ELISA is mechanistically and functionally similar to the traditional LLNA, such that the ICCVAM-recommended LLNA performance standards (ICCVAM 2009) could be used to evaluate future modifications of the LLNA: BrdU-ELISA.

LLNA for Testing Pesticide Formulations and Other Products, Aqueous Solutions, and Metals

The Panel comprises experts with knowledge in the evaluation of a range of test materials, but it is by no means expert in all of the product classes for which skin sensitization potential should be evaluated. The Panel also acknowledges that information and data gaps exist which prevent a full understanding of ACD epidemiology in humans. The test materials for which data are provided in the revised draft Addendum cover only a subset of the active ingredients used in each of the relevant product classes, and their frequency of use within those product classes is not noted in the revised draft Addendum. The Panel recommends that Federal agencies considering the results of this validation process assess how representative the test materials and findings in the revised draft Addendum are relative to substances of interest. In particular, the agencies should assess the chemical classes used in, and the range of biological effects of, the materials and products in which they have an interest.

The revised draft ICCVAM recommendations state that, although the database is limited, the traditional LLNA appears to be useful for evaluating substances tested in aqueous solutions or pesticide formulations provided the potential for overclassification (i.e., false positives) is not a limitation. The Panel agreed with these revised draft ICCVAM recommendations noting that the high rate of false positive substances may be inherent to the product and/or chemical class, testing of substances at concentrations that produced skin irritation, and to the fact that the LLNA detects the induction phase of skin sensitization. Furthermore, where comparative data were available, the LLNA identified more sensitizers than did guinea pig tests (predominantly Buehler tests which are considered to be less sensitive than the guinea pig maximization test [Basketter et al. 1993; Frankild et al. 2000]) but missed no materials that the guinea pig tests classified as sensitizers.

The Panel further suggested that, unless there are unique physiochemical properties associated with a material that might affect its ability to interact with immune processes, it should be a candidate for LLNA testing. An example of a material class that may possess such unique properties is some nanomaterials that are incapable of recognition by dendritic cells. Along these lines, the Panel also disagreed with the revised draft ICCVAM recommendation that a definitive recommendation on the usefulness of the LLNA for testing natural complex substances and dyes could not be made until more data were accrued. The Panel considered these classes of materials suitable for testing in the LLNA unless there are unique physiochemical properties associated with these materials that might affect their ability to interact with immune processes.

The Panel expressed a strong desire to avoid revalidation of the LLNA for new classes/types of test substances unless there is a biologically-based rationale. For new classes of test materials (e.g., nanomaterials), an integrated assessment of all available and relevant information should be conducted. This should include computer-assisted structure-activity relationships, prediction/measurement of biotransformation to potential reactive species, and possibly peptide, protein, or lipid binding. The Panel agreed that if any variant of the LLNA is validated for use to test novel classes, then the findings should be relevant to the family of validated LLNA tests and that similar uncertainties would surround the use of guinea pig models to evaluate novel classes of test materials.

The revised 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 pesticide formulations and other products, substances tested in aqueous solutions, and metals. The Panel evaluated the revised draft Addendum for completeness, errors, and omissions and concluded that there were no apparent errors. However, a Panel member did

note during the public meeting an omission regarding the natural complex substances; the relationship between the LLNA, guinea pig, and human data for major constituents (substances constituting at least 70%) of some of the natural complex substances and the LLNA results of the natural complex substances themselves was omitted. The Panel recommended that its suggestions/corrections relating to general, statistical, and specific editorial issues be incorporated into future revisions.

The Panel stated in its 2008 review (ICCVAM 2008) that the term *mixtures* was used too broadly (i.e., can represent an infinite number of materials), and this concern was addressed in the revised draft Addendum by dividing the substances considered into pesticide formulations, dyes, natural complex substances, and substances tested in aqueous solutions (this group included pesticide formulations tested in aqueous solutions), and analyzing the data for each group separately. The Panel agreed that the terms used to classify information submitted for the revised analysis are sensible and help to divide the dataset into useful categories for analysis, and that the product categories selected fit well with the nature and range of materials in the database. Such categories indicate classes of materials for which there exist, or do not exist, LLNA data and thus provide useful information for industry and regulatory agencies.

The Panel noted that the revised draft Addendum does not consider many classes of formulations to which humans may be exposed, by intention or by accident, such as: metalworking fluids, fuels, petroleum products used as lubricants, detergents and other cleaning agents, enzymes used in cleaning products, chemical household products, chemical (low molecular weight) pharmaceutical products, medical device materials (chemically characterized extracts), and nanomaterials (e.g., titanium oxide). Available data for substances within these classes may prove informative for human health.

Regarding pesticide formulations, the Panel concluded that the performance characteristics, reproducibility, and reliability of the LLNA had been adequately assessed and that the methods of data analysis were appropriate. The Panel indicated that the analysis for dyes, natural complex substances, and substances tested in aqueous solutions reflected the available information and the appropriate concordance statistics.

With regard to future studies, the Panel agreed with the ICCVAM recommendation for continued accumulation of information in the targeted areas. The Panel also indicated that solubility data should ideally be provided so that thermodynamic activity can be computed and compared to maximum theoretical percutaneous penetration. This information should be considered when comparing the data from LLNA studies in lipophilic delivery systems compared to that in aqueous systems. The Panel also suggested that, before additional animal testing is conducted, consideration should be given to product use and whether this renders a need to test the substance for skin sensitization potential.

1.0 Nonradioactive LLNA Protocol – The Murine Local Lymph Node Assay (LLNA): Daicel Adenosine Triphosphate (LLNA: DA) Test Method

1.1 Review of the Revised Draft Background Review Document for Completeness, Errors, and Omissions

The Panel was asked if there were any errors in the revised draft LLNA: DA background review document (BRD) that should be corrected, if omissions of existing relevant data had been identified, and if there was additional information that should be included.

The Panel noted that the reference for the International Organization for Standardization (ISO) standard in the reference section of the revised BRD (see Section 12.0) should be corrected to indicate that it is part of the 10993 series of standards. In addition, there is a typographical error in Table 6.6 of the revised draft BRD. The number of substances should be 44 for each criterion evaluated. Important omissions of existing relevant data or information that should be included in the final version of the LLNA: DA BRD are addressed below.

1.2 Review of the Validation Status of the LLNA: DA

1.2.1 Substances Used for the Validation Studies

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) asked the Panel whether it now considers the LLNA: DA database to be representative of a sufficient range of chemical classes and physicochemical properties such that the test method would be applicable for testing any of the types of chemicals and products typically tested for skin sensitization potential (see Section 6.3 of the revised draft BRD for a comparison of the substances tested in the LLNA: DA with the ICCVAM-recommended LLNA performance standards reference substances).

The Panel considered the LLNA: DA database broad and representative of a sufficient range of chemicals and products. The range of substances agrees with the range of chemical classes and reference substances suggested in the ICCVAM-recommended LLNA performance standards (ICCVAM 2009). All ICCVAM-recommended LLNA performance standard reference substances, except xylene, are included. The tested substances exhibit a full dynamic range (number of substances proportionally increased) of responses and a similar range of molecular weights, solubility, proportion of solids and liquids, etc., as the traditional LLNA. Chemicals that are typically tested for skin sensitization potential, including metal compounds, are included.

Based on the analysis in the revised draft BRD, the LLNA: DA may give less reliable results, similar to the traditional LLNA test method, for certain substances (e.g., metal compounds or strong dermal irritants). The outcomes related to the evaluation of the applicability domain of the traditional LLNA are also relevant and applicable to the LLNA: DA (i.e., any limitations in the traditional LLNA are also applicable to the LLNA: DA).

The Panel was asked what additional reference substances or products should be evaluated to obtain more data on the LLNA: DA test method's usefulness for identifying human sensitizers.

The Panel responded that further studies should be continuously conducted with substances with comparative human, guinea pig, and traditional LLNA data, including irritants and formulations. Regarding irritants, the proposed future studies might help explain why results obtained using the LLNA: DA and the traditional LLNA were discordant and further provide insight relating to the general challenges associated with discriminating irritants in the LLNA itself. Furthermore, initial studies conducted with a few metal compounds suggest that the LLNA: DA performs poorly with metals. Evaluation of additional metals may be needed to determine if this concern is real.

1.2.2 Test Method Accuracy

The test method developers recommend a decision criterion of a stimulation index (SI) greater than or equal to 3.0 when evaluating the test method performance of the LLNA: DA as a predictor of skin sensitization potential (see Section 6.2 of the revised draft BRD). This decision criterion yielded accuracy of 91% (40/44), sensitivity of 88% (28/32), and specificity of 100% (12/12) (i.e., there were four false negatives and no false positives). A decision criterion of SI $\geq$ 2.0 yielded the same accuracy as SI $\geq$ 3.0 but increased the sensitivity (97% [31/32]) by decreasing the false negative rate to 3% (1/32) although the false positive rate increased to 25% (3/12). These two single decision criteria were further compared to guinea pig tests and human/data experience (see Table 6-1 of the revised draft BRD).

The Panel was asked (1) whether these comparisons were appropriate for assessing the accuracy of the LLNA: DA using a single decision criterion and (2) whether the revised draft BRD had adequately evaluated and compared the relevance (e.g., sensitivity, specificity, false positive and false negative rates) of the LLNA: DA, using a single decision criterion (e.g., SI $\geq$ 2.0) to distinguish between sensitizers and nonsensitizers, to the traditional LLNA results.

The Panel indicated that the comparisons made to traditional LLNA data (i.e., SI ≥ 3.0) and also to guinea pig and human data, where available, are appropriate for assessing the accuracy of the LLNA: DA using a single decision criterion (e.g., SI ≥ 2.0 or SI ≥ 2.5) to identify a test substance as a sensitizer or a nonsensitizer.

The Panel preferred the ICCVAM-recommended analyses using multiple SI decision criteria (i.e., one SI cutoff to identify sensitizers and another SI cutoff to identify nonsensitizers), although the Panel expressed reservations about how the test variability might influence the classification of substances with SI values in the range of uncertainty. Additional information or testing is needed to classify substances with SI values in this range. For data in the test method accuracy tables, the Panel recommended graphing SI values for the traditional LLNA versus the SI values for the LLNA: DA with dividing lines at the cutoff values used to discriminate sensitizers from nonsensitizers to clarify the structure of the data.

The Panel agreed that the relevance of the LLNA: DA using a single decision criterion to distinguish between sensitizers and nonsensitizers has been adequately evaluated and compared to the traditional LLNA. Based on the most prevalent outcome for substances with multiple tests, Table 6-6 in the revised draft BRD shows identical accuracy, sensitivity, specificity, etc., using SI ≥ 3.0 or SI ≥ 2.5 for identification of sensitizers. Additional analysis based on data obtained from other studies may further strengthen confidence in the data concordance. Although the comparisons were appropriate, the analyses using multiple SI decision criteria are preferred. Additional information or testing is needed to classify substances with SI values in the range of uncertainty.

The performance of the LLNA: DA test method using classical statistical hypothesis testing (i.e., analysis of variance [ANOVA] or *t*-test to interpret sensitization potential for a given test) yielded accuracy of 84% (37/44), sensitivity of 94% (30/32), and specificity of 58% (7/12) (i.e., there were five false positives and two false negatives) relative to the traditional LLNA (see Section 6.5 of the revised draft BRD). ICCVAM asked the Panel whether these comparisons were appropriate for assessing the accuracy of the LLNA: DA using classical statistical hypothesis testing. It was also asked if the accuracy analysis provides an adequate comparison with which to decide between using a single SI decision criterion (e.g., SI ≥ 2.5) or classical statistical hypothesis testing to distinguish sensitizers from nonsensitizers.

The Panel indicated that the use of either ANOVA or *t*-test could determine if lymphocyte stimulation is significantly greater than the negative vehicle control; however, a statistical difference may not necessarily reflect a biologically important difference between a sensitizer and a nonsensitizer. Although the Panel agreed that these comparisons are

appropriate, the analyses using multiple SI decision criteria were preferable. However, the low specificity compared to traditional LLNA results does not support the use of classical statistical hypothesis testing as the only approach for determining whether a test substance is a sensitizer or nonsensitizer in the LLNA: DA. Substances other than sensitizers, including irritants, also significantly increase stimulation. Salicylic acid, for example, is an irritant, but not a sensitizer (Gerberick et al. 2002). This and a limited number of other irritants that induce proliferation in the draining lymph nodes may be incorrectly regarded as sensitizers (i.e., false positives) (Montelius et al. 1994). Table 6-8 of the revised draft BRD shows that salicylic acid is classified as a sensitizer when classical statistical hypothesis testing is used to test whether the vehicle control group is different from the treated groups. For data in the test method accuracy tables, the structure of the data would be clarified by graphing SI values for the traditional LLNA versus the SI values for the LLNA: DA, with dividing lines at the cutoff values used to discriminate sensitizers from nonsensitizers. Still, in cases where the sensitization potential of a test substance is uncertain, the use of classical statistical hypothesis testing in conjunction with a single decision criterion may contribute to a definitive skin sensitization identification of the test substance.

According to the Panel, the relevance (e.g., sensitivity, specificity, false positive and false negative rates) of the LLNA: DA using classical statistical hypothesis testing was correctly compared to the traditional LLNA, based on evaluation of individual raw values and considering the most prevalent outcome (or conservative outcome, in the case of an equal number of positive and negative results). The evaluation of alternative decision criteria included differences in adenosine triphosphate (ATP) values between treated and control groups, as well as mean ATP values of treated group versus two and three times the standard deviation of the mean ATP values of control group. The Panel mentioned that Table 6-6 of the revised draft BRD should be revised to delete the 'N' column and to add the number of substances evaluated to the table title.

Further evaluation of multiple decision criteria to identify substances as sensitizers or nonsensitizers indicates that at an SI $\geq$ 2.5 there were no false positives and at an SI $\leq$ 1.7 there were no false negatives in the LLNA: DA compared to traditional LLNA results (see Section 6.7 of the revised draft BRD).

ICCVAM asked the Panel whether these comparisons were appropriate for assessing the accuracy of the LLNA: DA using multiple decision criteria.

The Panel responded that these comparisons are appropriate to determine the optimal threshold values for identification of sensitizers and nonsensitizers. The most appropriate SIs for the traditional LLNA and its variants are not immutable biological constants; rather, they

must be determined by evaluation of test results. The approach adopted in the revised draft BRD did such an evaluation. In addition, this approach allows for the use of additional data and information (e.g., molecular weight, peptide binding of intermediate compounds) in the development of an informed decision regarding sensitization potential. Additional information or testing would be needed to classify substances with SI values in the range of uncertainty (i.e., $1.7 < SI < 2.5$).

According to the Panel, the use of "bootstrapping analysis" could be considered to determine the reliability of the SI cutoffs for identifying sensitizers and nonsensitizers. This would involve randomly sampling smaller subsets of the original dataset and calculating the cutoffs at which no false negatives and no false positives are produced. Such analysis would provide a measure of variability for the proposed cutoffs.

Power calculations might be necessary to determine if the sample size used is sufficient for those substances that are not definitively identified as sensitizers or nonsensitizers (i.e., substances in the range of uncertainty of $1.7 < SI < 2.5$).

According to the Panel, the relevance (e.g., sensitivity, specificity, and false positive and false negative rates) of the LLNA: DA using the $SI \geq 2.5$ criterion for sensitizers and the $SI \leq 1.7$ criterion for nonsensitizers was adequately evaluated and compared to the traditional LLNA. Additional analysis based on data obtained from future studies may further strengthen confidence in the data concordance. The revised draft BRD provides a partial rationale for the post hoc selection of these SI criteria, since importance should also be given to the corollary that these decision criteria may not apply when new data is considered.

Ten of the 44 substances used to evaluate the LLNA: DA yielded SI values between 1.7 and 2.5 (exclusive). Among these substances, 5/10 are sensitizers and 5/10 are nonsensitizers based on traditional LLNA results. Common characteristics of each group (i.e., the five sensitizers and the five nonsensitizers) are described in Section 6.8 of the revised draft BRD.

The Panel was asked whether the number of sensitizers (i.e., five) and nonsensitizers (i.e., five) that yielded $1.7 < SI < 2.5$ was sufficient to identify characteristics (e.g., peptide-binding activity, molecular weight) that might be potentially helpful in determining the skin sensitization hazard classification of such substances. If it was inadequate, the Panel was asked how many substances would be sufficient. If the number of substances was deemed sufficient, Panel members were asked to identify characteristics that are associated with these or similar substances and that may provide additional information about whether qualifying substances should be classified as sensitizers or nonsensitizers.

The Panel said that the number of sensitizers and nonsensitizers was not sufficient. If more substances are tested using the LLNA: DA, more reliable identification of crucial characteristics may be obtained. At present, the suggested characteristics (e.g., peptide-binding activity, molecular weight) may be of added value for a definitive skin sensitization identification of the substance, but the validation of these characteristics has not been done. The suggested characteristics were identified not only on the basis of this validation study, but because they represent cited characteristics of sensitizers in general.

The Panel indicated that including quantitative structure–activity relationship (QSAR) information, such as sensitizer structural alerts, might be useful in the future. However, it emphasized that too few substances in the range of uncertainty were analyzed to permit any meaningful analysis.

The LLNA: DA test method was developed before publication of the ICCVAM-recommended LLNA performance standards (ICCVAM 2009); therefore, those performance standards were not used to evaluate the test method. ICCVAM further stated that some of the protocol modifications noted for the LLNA: DA (i.e., pretreatment with sodium lauryl sulfate [SLS] and extended dosing schedule [see Section 2.0 of the revised draft BRD]) cause the LLNA: DA to be considered functionally and mechanistically different from the traditional LLNA.

The Panel responded that the LLNA: DA is mechanistically similar to the traditional LLNA, in that both methods measure cellular stimulation in the draining lymph nodes. The methods differ in the dosing regimen and in the timing of the assay. However, the outcome of the LLNA: DA test method is favorable, standard chemicals are included, and comparisons with traditional LLNA and with guinea pig and human data, especially regarding specificity, are favorable.

The Panel was asked whether the fact that the LLNA: DA produced false negative results for two of the 18 required performance standards substances affected its validation status, despite the fact that the ICCVAM-recommended LLNA performance standards (ICCVAM 2009) are not being used to evaluate its validity. The other 16 required performance standards substances that were tested yielded results concordant with the traditional LLNA (see Section 6.3 of the revised draft BRD).

The Panel agreed that the concordance of the results on 16 of 18 performance standard substances (at SI > 3.0) is favorable. Both false negative results were obtained from a single experiment. The highest SI for 2-mercaptobenzothiazole (i.e., 2.0) was established using only three mice, and one positive control failed. The highest SI for methyl methacrylate is 1.8, and neither individual SI values nor SI values for animals exposed to a positive control employed

in this experiment are available. The highest SI values for both substances fall in the recognized range of uncertainty. Both substances have low molecular weights, and 2-mercaptobenzothiazole exhibits high peptide reactivity. For methyl methacrylate, no peptide reactivity information was available. The Panel stated that studies with these substances should be repeated, and additional information on other parameters should be sought. There are distinct signs that these substances may be positive.

1.2.3 Test Method Reliability (Intra- and Interlaboratory Reproducibility)

The Panel was asked if it considered the intralaboratory reproducibility of the LLNA: DA to have been adequately evaluated and compared to the traditional LLNA (see Section 7.1 of the revised draft BRD) and, if not, what other analyses should be performed. It was asked if this intralaboratory reproducibility assessment, performed with two substances (eugenol and isoeugenol), had revealed any limitations.

The Panel responded that the intralaboratory reproducibility had been adequately evaluated on the basis of the two substances tested. The revised draft BRD reports that these substances were tested three times by one laboratory. The Panel noted that individual animal data from repeated experiments are available for eugenol only at the 10% concentration and that data for isoeugenol come only from one study (as documented in Appendix D1 of the revised draft BRD).

The two-phased interlaboratory validation study of the LLNA: DA organized by the Japanese Society for Alternatives to Animal Experiments (JSAAE) tested 14 different coded substances (10 sensitizers and four nonsensitizers based on traditional LLNA results). Two sensitizers and one nonsensitizer were each tested among a number of laboratories (ranging from 10 to 17 laboratories); the remaining eight sensitizers and three nonsensitizers were each tested in three or four laboratories. JSAAE's Validation Management Team selected the vehicles and the concentrations of the substances tested. The Panel was asked about (1) any apparent limitations of this study design and (2) any concerns about the fact that vehicles and concentrations were provided to the participating laboratories by the Validation Management Team.

The Panel expressed more concern for reducing variables in the test than for transferability, and therefore there are no concerns about vehicles and concentrations being provided by the Validation Management Team. In fact, this approach may have been advantageous because factors outside the laboratory were eliminated. The primary purpose was to evaluate the performance of the test, not the protocol or laboratories; therefore, the unresolved issue is transferability. The Panel recommended adding information to Table 7-1 of the revised draft

BRD so some measure of within-experiment variation and between-experiment variation would be available. The following information should be added:

- Sample sizes
- Mean of means
- Variance of means
- Coefficient of variation (CV) of means

The table should include all the information needed to perform an F-test (i.e., sample sizes as relevant to degree of freedom calculations, means on a log scale, and mean square errors for the within-laboratory and between-laboratory comparisons).

The Panel was also asked if it considered the interlaboratory reproducibility of the LLNA: DA, using SI $\geq$ 2.5 for sensitizers and SI $\leq$ 1.7 for nonsensitizers, to have been adequately evaluated and compared to the traditional LLNA.

According to the Panel, interlaboratory reproducibility of the LLNA: DA using these criteria has been adequately evaluated and compared to the traditional LLNA. When SI $\leq$ 1.7 was used to identify nonsensitizers, there was favorable qualitative concordance (87%) among 14 substances with multiple tests. Regarding the ICCVAM-recommended LLNA performance standards' acceptable range for ECt values (i.e., estimated concentration of a substance needed to produce a stimulation index that is indicative of a positive response) for 2,4-dinitrochlorobenzene (DNCB) and hexyl cinnamic aldehyde (HCA), half of the laboratories exceeded the range in the first phase (for DNCB), which is understandable for a test using a strong sensitizer with a different dosing paradigm and time course that might extend into the elicitation phase of skin sensitization. However, all the laboratories in the first and second phase of the interlaboratory validation study obtained EC2.5 values (i.e., estimated concentration of a substance needed to produce a stimulation index of 2.5) within the concentration range indicated for HCA, which is favorable. The Panel recommended performing isotonic regression prior to interpolation when calculating ECt values.

The Panel recommended that Tables 7-7 and 7-8 of the revised draft BRD be revised in the final BRD to include sample sizes, means, standard deviation (SD), and CV. Variation (i.e., SD) should be measured on a log scale and then converted to antilog for reporting in the table. Tables 7-9 and 7-10 of the revised draft BRD should be accompanied in the final BRD by a figure with traditional LLNA SI versus LLNA: DA SI. The figure should be designed so that substances tested could be distinguished by different symbols and the different laboratories could be distinguished by numbers on the graph plotted on a log scale.

1.2.4 Consideration of All Available Data and Relevant Information

The Panel was asked if the revised draft LLNA: DA BRD adequately considered all the relevant data identified in published or unpublished studies employing this test method.

The Panel responded that it was unaware of any omitted data.

1.2.5 Animal Welfare

The traditional LLNA, which evaluates the induction phase rather than the elicitation phase of ACD, represents a refinement of sensitization testing compared to guinea pig tests. The Panel's 2008 review of the LLNA: DA test method indicated concern that the treatment schedule used in the LLNA: DA procedure might be of sufficient length that the elicitation phase of the ACD response was being evaluated (ICCVAM 2008). This raised an animal welfare concern that unnecessary animal distress and discomfort associated with the localized skin reaction of ACD was being induced, similar to that for the guinea pig tests. A continuing concern regarding the LLNA: DA is that the current LLNA: DA data includes no evidence, either by observation or by histological examination, to indicate if the longer protocol with 1% SLS pretreatment causes a skin reaction indicative of the elicitation phase of skin sensitization.

Appendix A2 of the revised draft BRD provides data that indicate that the pretreatment with 1% SLS increases the sensitivity of the LLNA: DA assay for the limited number of compounds tested (i.e., two extreme-to-strong sensitizers, three moderate sensitizers, four weak sensitizers, four nonsensitizers [including two nonsensitizers that were irritants]). The 2008 Panel report concluded that weak irritants and weak sensitizers needed to be tested in the LLNA: DA assay with and without 1% SLS pretreatment to demonstrate if it has an immunomodulatory effect in the draining lymph node or if it alters the sensitivity of the application site to irritants (ICCVAM 2008). The Panel would like to know whether the assay performs as well without the 1% SLS pretreatment application.

1.3 Comments on the Revised Draft ICCVAM Test Method Recommendations on the LLNA: DA

1.3.1 Test Method Usefulness and Limitations

The Panel was asked if the available data and the LLNA: DA test method performance (accuracy and reliability) supported the revised draft test method recommendations in terms of the test method's proposed usefulness and limitations.

The Panel agreed that the available data and test method performance supported ICCVAM's revised draft test method recommendations for the LLNA: DA. The outcome comparisons with traditional LLNA, guinea pig data, and human data, especially regarding specificity,

were favorable. The endpoint assessed is related to the cell stimulation, as in the case of the traditional LLNA, but the timing of doses including the "booster" might extend beyond the induction phase of skin sensitization. The LLNA: DA test method's accuracy and reliability substantiate its usefulness in identifying potential skin sensitizers and nonsensitizers, but with specific limitations (i.e., when the SI has a value between 1.7 and 2.5).

According to the Panel, the LLNA: DA test method has the potential to correctly identify many sensitizers and nonsensitizers. Although the Panel agreed with the general conclusions and recommendations of the revised draft ICCVAM recommendations, it indicated that more information or testing is needed to further evaluate the materials with SI values in the range of uncertainty. The limitation with respect to nickel is generally applicable to the traditional LLNA and other variants.

ICCVAM proposes using the criteria of SI $\geq$ 2.5 and SI $\leq$ 1.7 to identify substances as sensitizers and nonsensitizers, respectively. The Panel agreed that this framework allows greatest confidence to be placed in the LLNA: DA's identification of test materials as sensitizers or nonsensitizers. Additional information or testing is needed to classify substances in the range of uncertainty. The Panel stated that accuracy analysis supports ICCVAM's recommendation that the decision criteria be SI $\geq$ 2.5 for sensitizers and an SI $\leq$ 1.7 for nonsensitizers. When the SI $\geq$ 2.5 criterion was used to identify sensitizers, no false positives to the traditional LLNA were recognized, and the false positive rate was 0%. When the SI $\leq$ 1.7 decision criterion was used to identify nonsensitizers, the false negative rate, with respect to the traditional LLNA, was 0%.

When multiple decision criteria are used, the test has the ability to provide data for the reliable identification of both sensitizers and nonsensitizers, more so than when applying a single SI decision criterion to the current dataset. The model proposed (i.e., using SI $\geq$ 2.5 to identify substances as sensitizers and SI $\leq$ 1.7 to identify nonsensitizers) is empirical and derived from the available data. It gives the best combination of minimum risk of false positives and false negatives, with the lowest number of "uncertain" results. Since using two decision criteria allows for a more definitive identification of sensitizers and nonsensitizers, this approach provides animal welfare benefits by reducing further tests that might be required in instances where the hazard classification of a substance is not as clear. Although the use of multiple decision criteria to identify sensitizers and nonsensitizers appears reliable, the Panel questioned how results in the range of uncertainty would be useful for regulatory purposes.

The Panel was asked whether concerns about potential false positives or false negatives that may occur in this test method would be resolved by using multiple decision criteria. Other suggestions for additional such guidance or limitations were solicited from the Panel.

The Panel responded that using the decision criterion of SI ≤ 1.7 for nonsensitizers and SI ≥ 2.5 for sensitizers gives a definitive skin sensitization identification with a greater degree of reliability than when test substances are evaluated in any system that applies a single test point to identify potential skin sensitizers. However, for substances exhibiting SI values between 1.7 and 2.5, additional information or test data should be used to reach a definitive skin sensitization identification. It was noted that because the decision criteria chosen to identify sensitizers and nonsensitizers were based on a post hoc analysis, prospective testing with the test method might affect the proposed model. For this reason, data generated should be routinely evaluated to determine if the proposed model is still optimal with regard to the decision criteria.

The Panel agreed that, especially for substances with SI values between 1.7 and 2.5 (exclusive), classical statistical hypothesis testing (i.e., the statistical analysis of ATP values for treated groups versus control group), in addition to SI values, might contribute to a definitive skin sensitization identification. The Panel also noted that high variability in the test could result in misclassification of certain substances (e.g., especially those having SI values in the range of uncertainty). Using classical statistical hypothesis testing, sensitizers with highly variable results may not be classified as such (i.e., the variation in the data will not allow a meaningful difference from the vehicle control to be detected). Furthermore, nonsensitizers that produce data with low variability may be detected as sensitizers by classical statistical hypothesis testing methods. The SI is the primary decision criterion, but if variability is high, the reliability of the test result may be inconclusive. Further statistical evaluation could be informative.

ICCVAM recommends that when an SI value between 1.7 and 2.5 is obtained in the LLNA: DA, users should carefully interpret the results in an integrated decision strategy in conjunction with all available and relevant information (e.g., dose response information, statistical analyses, peptide-binding activity, molecular weight, results from related chemicals, other testing data) to determine if there is adequate information for an accurate sensitization hazard classification or if additional testing is necessary.

The Panel agreed with ICCVAM's recommendation but added that, for substances exhibiting SI values in the range of uncertainty, additional information or testing should be considered before a definitive skin sensitization identification is concluded.

The Panel stated that dose response information, statistical analyses, peptide-binding activity, molecular weight, structural alerts for skin sensitization, results from related chemicals, other testing data, etc., should be considered first as aids in determining a definitive skin sensitization identification of a substance with an SI between 1.7 and 2.5. However, if this information is inconclusive, additional animal tests (i.e., the traditional LLNA) should be considered. According to the Panel, such supplementary testing should be conducted only if necessary. Thus, it is also important to consider the intended use of the substance, as this could factor into determining the skin sensitization identification of a substance. For instance, if the substance is intended for consumer products, the human repeat insult patch test (HRIPT) might be useful to confirm sensitization potential, or the lack thereof.

ICCVAM asked the Panel to consider under what circumstances could results from the LLNA: DA yielding SI values between 1.7 and 2.5 be used as part of an integrated decision strategy (i.e., in conjunction with all available and relevant information such as dose response information, statistical analyses, peptide-binding activity, molecular weight, results from related chemicals, other testing data) to assign a sensitization hazard classification.

The Panel responded that the test method should be performed under Good Laboratory Practice (GLP) conditions, should strictly adhere to the standard operating procedure, and should include controls. It should preferably be duplicated, and all indicated additional information should be reviewed. This could be considered an initial approach in all cases, but because there is information on only 10 compounds in the range of uncertainty, more information is needed to determine if this approach is the best.

The interlaboratory reproducibility standard from the ICCVAM-recommended LLNA performance standards (ICCVAM 2009) is a range of ECt values from each laboratory of 0.025% to 0.1% for DNCB and 5% to 20% for HCA. The Panel was asked if it was concerned that five of 10 laboratories that tested DNCB produced EC2.5 values outside the acceptability range. If so, it was asked whether the concern was mitigated by the fact that only two of the aberrant EC2.5 values were greater than the upper range. ICCVAM also asked the Panel if lower limits for the range of acceptable ECt values were necessary (i.e., lower values indicate a more sensitive test).

The Panel indicated that the fact that 5/10 laboratories exceeded the performance standards' acceptable range for ECt values in the first phase (for DNCB) was understandable. DNCB is a strong sensitizer, and the LLNA: DA has a different dosing regimen and time course than the traditional LLNA, which might extend into the elicitation phase of skin sensitization. However, all the laboratories in the first and second phase of the interlaboratory validation

study obtained EC2.5 values within the concentration range indicated for HCA, which documents the test's favorable reproducibility and performance.

According to the Panel, both the upper and lower limits for the range of acceptable ECt values for a positive control are important to document the validity of performance of individual experiments performed later. If the ECt value obtained for the positive control is below the acceptable range, the whole experiment might be oversensitive. The sensitization potential of the test substance may have been overpredicted in regard to the concentration used as the sensitization threshold, which might be counterproductive for an approved concentration of this substance (e.g., for consumer products).

The fact that a higher acceptability range had to be used for DNCB may suggest that the LLNA: DA is not as sensitive for this contact sensitizer. However, in the opinion of the Panel, these concerns are mitigated by the fact that the other positive control (HCA) performed well in this test.

1.3.2 Test Method Protocol

The Panel was asked if it agreed that the available data supports the revised draft ICCVAM recommendations for the proposed test method standardized protocols for the LLNA: DA.

The Panel responded that the efficacy of the standardized procedure had been improved during documented intra- and interlaboratory validation studies, which included training of participating laboratories. The recommendations for maintaining a positive control database reflect current evidence and best practice; the evaluation of the variation in positive control responses over time probably has wide applicability to a broad range of test systems.

1.3.3 Future Studies

The Panel was asked if it agreed that the available data support the revised draft ICCVAM recommendations for the LLNA: DA in terms of the proposed future studies and, if not, what recommendations it would make.

The Panel recommended that additional decision criteria and guidance should be identified for substances exhibiting SI values between 1.7 and 2.5. For any test method in which multiple decision points are used and a subset of the test materials are initially identified in a range of uncertainty, the secondary decision criteria should be reassessed as additional data and discriminators become available (e.g., high-quality human ACD data).

Also, the Panel indicated that additional studies should be conducted for other substances (including irritants, metals, and formulations) that have comparative guinea pig, human, and traditional LLNA data. While specifically addressing a question related to the LLNA: DA, these data needs are common to all variants of the LLNA.

Regarding the impact of irritants, the proposed future studies may provide insights to better explain discordant results between the LLNA: DA and the traditional LLNA and also the general issue of discriminating irritants from sensitizers in the traditional LLNA itself. While peptide reactivity shows promise in assisting to identify substances with SI values in the range of uncertainty, this approach has not been sufficiently evaluated.

There should be analysis of the process and results used to define the cutoff values (i.e., quality assessment of the process). A protocol should be developed for evaluating such cutoffs so that future developers could apply it during the development of their method.

The Panel recommended further consideration of statistical issues, including how to determine and evaluate classification methods (i.e., classification cutoff points). For example, "bootstrapping analysis" could be considered to determine the reliability of the SI cutoffs for identifying sensitizers and nonsensitizers. This would involve randomly sampling smaller subsets of the original dataset and calculating the cutoffs at which no false negatives and no false positives are produced. Such analysis would provide a measure of variability for the proposed cutoffs. The Panel also recommended that future interlaboratory validation studies should simultaneously evaluate intralaboratory reproducibility, using the appropriate statistics, to evaluate variation both within a laboratory and between laboratories. There should be a set of guidelines that describes the general principles and a basic method for estimating variation in both intralaboratory and interlaboratory experiments. Additionally, the Panel strongly recommended that a statistician actively participate in the preparation of future BRDs and formulation of ICCVAM recommendations.

1.3.4 Performance Standards

ICCVAM recognizes that the LLNA: DA test method incorporates procedures that differ from the essential test method components in the ICCVAM-recommended LLNA performance standards (ICCVAM 2009) and therefore the traditional LLNA performance standards may not be applicable to the LLNA: DA. Accordingly, ICCVAM has proposed that separate performance standards be developed for the LLNA: DA to evaluate future modifications of this test method. The Panel was asked whether it agreed with this proposal.

The Panel indicated that separate performance standards for the LLNA: DA are not needed. Although modifications to the traditional LLNA protocol are recognized for the LLNA: DA (dosing, timing, pretreatment of exposition sites with 1% SLS), the Panel viewed that the LLNA: DA is not mechanistically different from the traditional LLNA. The Panel indicated that the ICCVAM-recommended LLNA performance standards (ICCVAM 2009) should be used for the LLNA: DA test method and for all modified LLNA protocols, otherwise, the performance standards are not useful.

2.0 Nonradioactive LLNA Protocol – The LLNA: Bromodeoxyuridine Detected by Flow Cytometry (BrdU-FC) Test Method

2.1 General Comments

The Panel noted that since 2008, no data for additional test materials have been added to the analyses in the revised draft BRD, and that new data for intralaboratory reproducibility have been properly integrated into the assessment. The intralaboratory assessment of two ICCVAM-recommended LLNA performance standard reference substances (i.e., HCA and DNCB) was performed in one laboratory (i.e., MB Research Labs, which developed the test method), using the LLNA: BrdU-FC without the immunophenotypic measurements (i.e., the "enhanced" LLNA: BrdU-FC or eLLNA: BrdU-FC). The additional information on the vehicle dependence of the response to 2-mercaptobenzothiazole is relevant, and shows the importance of evaluating the influence of vehicles for all LLNA methods.

The results of the revised concordance assessments of the LLNA: BrdU-FC against the traditional LLNA test method suggest that the LLNA: BrdU-FC (as performed at the originating facility) may have the potential to be developed as a reliable alternative to the traditional LLNA, with the same applicability domain. Both the LLNA: BrdU-FC and the eLLNA: BrdU-FC, on the basis of the information available, perform equally well in a single laboratory compared with the traditional LLNA.

However, additional information, which has not been provided for the LLNA: BrdU-FC, is required to satisfy all of the essential ICCVAM requirements for scientific validation of the test method:

- The original test data is required to perform an independent audit
- Data is needed against which to evaluate transferability and interlaboratory reliability

The eLLNA: BrdU-FC, by including detection of immunophenotypic markers, resolves the discrimination of nonsensitizing irritants (false positives) from sensitizing irritants. The Panel considers the use of immunophenotypic markers and the measurement of ear swelling as potentially valuable adjuncts to the traditional LLNA and other LLNA variants. This is particularly relevant when considering the challenges associated with discriminating irritants from sensitizers in the LLNA.

Although the data show that the performance of the LLNA: BrdU-FC test method is promising, the Panel accepted and endorsed the revised draft ICCVAM recommendation on

the validation status to defer of a formal recommendation on the validity of the test method. This recommendation was made due to the lack of individual animal data, or data audit results, to support the results used for the accuracy assessment and the lack of data to evaluate transferability and interlaboratory reliability as described in the revised draft ICCVAM recommendations. *These are essential prerequisites required before a formal validation statement can be made.*

Since some of the technical steps and technologies used to measure the endpoint in the LLNA: BrdU-FC are different than those used in the traditional LLNA, interlaboratory evaluation is critical to ensure transferability.

The Panel recommended that an analysis must be done using the intralaboratory reproducibility data to determine the minimum group size required to ensure that the performance of this test method is equivalent to or better than the traditional LLNA.

The Panel also recommended that additional explanatory information should be provided to support an adequate assessment of the validation status of the LLNA: BrdU-FC including details of the flow cytometry methods used, and a rationale for the cutoffs for the immunophenotypic endpoints.

2.2 Review of the Revised Draft BRD for Completeness, Errors, and Omissions

In response to the 2008 Panel recommendations for more data (ICCVAM 2008), additional raw data and test vehicle information have been provided for 13 of 45 substances evaluated, including limited data on 10 substances from single studies previously described in the January 2008 draft BRD. The Panel was asked if the revised draft BRD for the LLNA: BrdU-FC contained any errors that should be corrected or omissions of relevant data or information that should be included.

The Panel responded that it was not aware of omissions or errors in the use of available data. However, no data from testing laboratories other than MB Research Labs were made available for this analysis, although other laboratories are using the LLNA: BrdU-FC (Jung et al. 2009). The experience of these investigators might better inform unresolved validation considerations. Additional citations are available on the application of BrdU in the LLNA with histochemical or enzyme-linked immunosorbent assay (ELISA) detection (as included, for example, in the BRD for the LLNA: BrdU-ELISA method). These could be briefly mentioned in the BRD for the LLNA: BrdU-FC method simply as a further means of indicating the potential utility of nonradiolabeled tracer methods in the LLNA.

According to the Panel, all of the additional information has been properly incorporated. The quality of the data is as stated in the revised draft BRD, and the data relied upon seems fit for purpose. However, the revised draft BRD makes it clear that the information available to the Panel is incomplete; the test method developer has not provided all relevant information required for all of the components of a formal validation evaluation. In addition, full technical details relating to the use of the immunophenotypic markers (i.e., eLLNA: BrdU-FC) have not been disclosed.

The crucial gaps relate to the fact that some of the original traditional LLNA data, not done in compliance with GLP, have not been obtained and audited; and no interlaboratory transferability and reliability data are available for consideration.

In addition, the Panel recommended that the following editorial and non-technical specific corrections and revisions be addressed for the final BRD:

- Line 345: Insert a space in "contrastLLNA."
- Line 547: Change to "Mean ratio of BrdU-labeled cells against total cells in the treated group" and "Mean ratio of BrdU-labeled cells against total cells in the vehicle control group".
- Line 562: Add "an irritating" between "is classified as" and "sensitizer".
- Line 667: Edit to state "The test substances were not coded to hide their identities during testing."
- The ISO standard reference citation in the References section (Section 12.0) is incomplete: it is part of the 10993 series of standards.

2.3 Review of the Validation Status of the LLNA: BrdU-FC

2.3.1 Test Method Accuracy

Test method accuracy was recalculated based on updated reference data obtained since the 2008 Panel meeting. In view of this updated evaluation, the Panel was asked whether it considered the relevance (i.e., accuracy, sensitivity, specificity, false positive and false negative rates) of the LLNA: BrdU-FC to have been adequately evaluated and compared to the traditional LLNA (see Section 6.1 of the revised draft BRD).

The Panel responded that, with respect to the available intralaboratory dataset, the performance characteristics for reproducibility and reliability of the LLNA: BrdU-FC test method against the ^{3}H-thymidine-based traditional LLNA test method is adequately and completely explored in the revised draft BRD. The method of data analysis employed is appropriate and has been applied to results of both the LLNA: BrdU-FC and eLLNA: BrdU-

FC. The traditional LLNA is used as the benchmark for comparisons, and human and guinea pig data are taken into account where available. On that basis, compared against results obtained with the traditional LLNA, the database of more than 45 suitable and representative test materials yields adequate accuracy of 93% (42/45), a false positive rate of 13% (2/16), and a false negative rate of 3% (1/29). Although the comparison of data from the LLNA: BrdU-FC test method with guinea pig and human data show relatively low accuracy, the accuracy is similar to that of the traditional LLNA data compared with guinea pig and human data. Concordance values for all paired comparisons have been calculated and discussed, along with highlights of any discordant results. For some discordant results, there is insufficient information on the vehicles used: however, any reevaluation taking account of these insights could only have improved, not compromised, the above test performance figures. The Panel noted the lack of an independent audit to assure that the reported data agrees with raw data from Ryan et al. (2000), Schneider and Akkan (2004), Gerberick et al. (2005), and Basketter et al. (2006).

The Panel was asked for any comments or recommendations regarding the revised analysis or the revised reference data. It provided the following:

- The Panel recommended inserting a brief statement of rationale into Section 2.0 or 2.1 of the final BRD to explain the discriminating values chosen for immunophenotypic detection of irritants (the value of 25% is used throughout).
- The observations made with respect to vehicle influence on the response to 2-mercaptobenzothiazole are of interest and could be added as a table (see Section 5.0 in revised draft BRD).
- For data in the test method accuracy tables, the structure of the data would be clarified by graphing SI values for the traditional LLNA versus the LLNA: BrdU-FC with dividing lines at the cutoff values used to discriminate the positives from the negatives.
- Studies were not all conducted in full compliance with GLP, and original data have not been supplied or independently audited; the former need not be cause for the concern if the data are provided and shown to be accurate at audit.

The 2008 Panel report provided power analyses for the LLNA: BrdU-FC (ICCVAM 2008). In that report, the estimated group size to detect a three-fold increase relative to the 4:1 acetone: olive oil (AOO) vehicle control group was five animals for 80% power, and that nine animals per group were necessary to achieve 95% power. Given that the current protocol for this assay specifies five animals per group, the Panel is concerned about the

apparent high variability in the power analysis, in contrast with the close agreement of the intralaboratory data included in the revised draft BRD. The Panel notes that the data presented in the revised draft BRD were obtained with a group size of n = 5 animals per group.

2.3.2 Test Method Reliability

The Panel was asked whether, in view of the additional data submitted for HCA and DNCB, it considered intralaboratory reproducibility to have been adequately evaluated and compared to the traditional LLNA (see Section 7.1 of the revised draft BRD).

The Panel responded that the analysis of intralaboratory reproducibility as presented is adequate and fit for the intended purpose. The newly integrated information on reproducibility of outcomes with both HCA and DNCB (see Table 7.2 of the revised draft BRD) is valuable and provides an expanded perspective on the data for responses to 25% HCA in multiple vehicles presented in Table 7-1 of the revised draft BRD. The presentation might be made more comprehensive and informative by inclusion, for the purposes of comparison, of representative reproducibility data for a positive control substance within one laboratory performing the traditional LLNA (compare/contrast to data in Table 7-1 of the revised draft BRD). The final BRD should reference Appendix D for the experimental data.

The EC3 (estimated concentration of a substance needed to produce a stimulation index of 3) data for HCA and DNCB confirmed values within the acceptable range in each case when judged against the ICCVAM-recommended LLNA performance standards (ICCVAM 2009).

For the assessment of interlaboratory reproducibility, the ICCVAM-recommended LLNA performance standards indicate that

> "Interlaboratory reproducibility should be evaluated with at least two sensitizing chemicals with well-characterized activity in the traditional LLNA. In this regard, ECt values for 2,4-dinitrochlorobenzene (DNCB) and HCA should be derived independently from a single study conducted in at least three separate laboratories. Acceptable reproducibility will be indicated by each laboratory obtaining ECt values for HCA and DNCB that are within 0.5x to 2.0x (5% to 20% and 0.025% to 0.1%, respectively) the mean EC3 concentration (10% and 0.05%, respectively) specified for these substances . . . " (ICCVAM 2009).

As indicated in the 2008 Panel report,

> "The Panel agreed with the draft ICCVAM recommendations for evaluating test method reliability. These recommendations included obtaining ECt values that are generally within 0.5x to 2.0x of the mean historical EC3 (i.e., estimated concentration

of a substance needed to produce an SI of 3) values for hexyl cinnamic aldehyde (HCA) (intralaboratory, n=4 experiments in one laboratory), or HCA and 2, 4-dinitrochlorobenzene (DNCB) (interlaboratory, n=1 experiment in three laboratories)" (ICCVAM 2008).

ICCVAM asked the Panel if it agreed that the LLNA: BrdU-FC test method is mechanistically and functionally similar to the traditional LLNA test method. If so, it was asked if the studies in the ICCVAM-recommended LLNA performance standards (ICCVAM 2009) for use in demonstrating intra- and interlaboratory reproducibility for the LLNA: BrdU-FC test method would be sufficient. If not, the Panel was asked to recommend other testing that should be performed to demonstrate that it a study is sufficiently reproducible.

The Panel agreed that the LLNA: BrdU-FC is mechanistically and functionally similar to the traditional LLNA test method and conforms with the technical specifications and essential test method components set out in the ICCVAM-recommended LLNA performance standards (ICCVAM 2009). Therefore, the studies proposed by the ICCVAM-recommended LLNA performance standards are sufficient to establish the intra- and/or interlaboratory performance of the LLNA: BrdU-FC.

The Panel was asked to recommend particular tests that would demonstrate proficiency in conducting LLNA: BrdU-FC studies before a naïve laboratory performs them for regulatory data submissions.

For the purposes of this question, the Panel assumed that there are two classes of naïve laboratory. These would be laboratories with no experience with, or competence in, flow cytometry, or alternatively, laboratories with experience in flow cytometry, but no experience with its use as part of an LLNA protocol.

The Panel agreed that any laboratory that is currently competently performing the traditional LLNA can easily and readily carry out the majority of steps of the LLNA: BrdU-FC test method. However, unlike some operating equipment for which a short presentation by a manufacturer's representative is sufficient training (e.g., scintillation counter, microplate reader), several days are necessary to train users to operate a flow cytometer. Manufacturers of flow cytometers run training courses that usually last several days with certification of those who successfully pass the course. This level of proficiency should be required of any operator at a naïve laboratory that has no experience in flow cytometry.

According to the Panel, if a laboratory has flow cytometry experience or following training and certification of personnel, it should demonstrate proficiency by repeating the evaluation of reference substance(s) (i.e., four independent tests) to allow an assessment of

intralaboratory reproducibility (consideration of the test articles is discussed immediately below). Additional considerations would include development of a standard test method protocol, standard operating procedure, and other documentation, and adherence to recognized quality assurance/quality control systems for flow cytometry and associated data acquisition equipment.

The Panel was asked to recommend substances from the ICCVAM-recommended LLNA performance standards that might be used for testing.

The Panel, in response, indicated that the performance standard is relevant and a laboratory with flow cytometry experience and/or following training and certification should evaluate their results for both a strong and a moderate known sensitizer (e.g., DNCB and HCA, respectively). The inclusion of a known, reproducible weak sensitizer and a negative control is essential to confirm that the full range of appropriate responses can be reproduced.

The Panel was asked if results from the positive control from each study would be adequate for a laboratory to assess its intralaboratory reproducibility.

The Panel indicated that, in principle, the results from the positive control in each study should be adequate for a laboratory (and others) to assess intralaboratory reproducibility. However, it would be better to demonstrate reproducibility with a purpose-designed study. Concern might be raised if positive control data were aggregated over time from studies in which unknowns were also run and conclusions about the unknowns were made before a sense of assay performance (reproducibility) was gained. Additionally, there may be a case for considering the evaluation of results for a mild sensitizer and a nonsensitizer to confirm that the full range of appropriate responses can be reproduced within the laboratory.

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

2.4.1 Test Method Usefulness and Limitations

Similar to the 2008 draft test method recommendations, ICCVAM has deferred a formal recommendation on the validity of the LLNA: BrdU-FC until all data supporting the analysis have been submitted to the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) for an independent audit and transferability and reliability have been demonstrated in a second testing laboratory. The Panel was asked if it agreed with this recommendation or had any further comments or suggestions regarding the position.

The Panel agreed with ICCVAM's decision to defer formal recommendation on the test validity for the reasons given and, in addition, because of unresolved issues related to test group size (see above). The Panel had nothing further to add that would alter the recommendation to defer judgment on the validity of the LLNA: BrdU-FC test method.

The 2008 Panel review specified areas for continued data accrual, which would be applied to validation considerations (ICCVAM 2008). Dialog with the LLNA: BrdU-FC developer has brought some important data into the 2009 analysis, and further information should be solicited. The Panel noted that crucial omissions remain relating to essential validation elements that must be addressed before any formal recommendation on the scientific validity of the test method can be made. In addition, unless and until additional technical information is available regarding use of the immunophenotypic markers, consideration of their usefulness and limitations should be treated with caution.

The Panel responded to the question of whether the 2009 revised draft BRD supports the revised draft ICCVAM test method recommendations on the usefulness and limitations of the LLNA: BrdU-FC with regard to its use in the developing laboratory (i.e., MB Research Labs).

The Panel commented that the revised draft BRD relates to the revised draft test method recommendations that are offered. The final test method recommendations should highlight those essential items of highest priority required for further consideration of validation. The immediate priorities are:

1. A review of the original data at the individual animal level with appropriate positive and negative controls (i.e., data audit), and
2. A definition of the minimum number of animals used per test group, then
3. Performance and evaluation of the interlaboratory reproducibility study to the specifications in the ICCVAM-recommended LLNA performance standards (ICCVAM 2009) with appropriate quality control systems.

Only then should the less critical items (e.g., methodological specifics, immunophenotypic endpoints, alternative decision criteria for identifying materials as sensitizers and nonsensitizers) be evaluated.

With regard to determining the skin sensitization classification of a test substance, the Panel affirms that the SI is the primary decision criterion, but if variability is large, the reliability of the test result may be inconclusive and statistical evaluation or an integrated assessment of all available and relevant information could be informative and preferable to conducting additional animal tests.

2.4.2 Test Method Protocol

The Panel was asked whether or not it agreed that the available data supported the revised draft ICCVAM test method recommendations for the proposed test method standardized protocol for the LLNA: BrdU-FC procedure and, if not, what recommendations it would make.

The Panel agreed that the analysis and discussion of data in the revised draft BRD support the revised draft ICCVAM recommendations for the test method protocol. The Panel recommended that the application of the immunophenotypic lymphocyte markers in the eLLNA: BrdU-FC method be further explained and standardized; this is not well covered in the method description in Appendix A of the revised draft BRD.

Revised power calculations must be performed using the data provided for the intralaboratory performance to determine the minimum group size required to provide a level of test performance equivalent to or better than the traditional LLNA as reported in the reference documents (ICCVAM 2008). The minimum group sizes in the protocol should then be adjusted, if necessary, before a formal interlaboratory transferability and reliability study is commissioned.

When ear swelling increases by over 25%, the eLLNA: BrdU-FC employs a two-step cell-surface analysis. The first step seems overly complicated because both B220 and CD69/I-Ak cells are expressed on B cells. Also, suspended cells have to be stored for up to 72 hours at 4^{o}C. According to the Panel, double staining the cells with BrdU and a B cell marker (i.e., CD19) simultaneously, for flow cytometry analysis, may be worth considering as a simpler routine method.

The Panel noted that AOO is the preferred solvent/vehicle and that a list of alternatives is provided. The Panel separately considered the suitability of 1% Pluronic L92 in another context (see **Section 4.6**), and it might be equally valid here. In brief, consideration of alternative vehicles relates to solubility, physical chemistry, and biology. This is covered in the Panel's discussion of the LLNA applicability domain (see **Section 4.0**), and the Panel's conclusions and recommendations are relevant here.

The Panel was asked whether laboratories should maintain a historical database of positive control data to provide a measure of variability of the positive control response over time.

The Panel indicated that this is an important element of compliance with general good practice and longitudinal tracking of key performance indicators. The Panel considered that this should be a general requirement for all tests in which outcomes are determined by the performance of a variable biological test measured against a positive control.

The Panel's specific recommendations for improving the MB Research Labs standard test method protocol include:

- Section 5.9.3 of Appendix A in the revised draft BRD indicates that weight loss *"will be addressed"*. Clarify what that means, and whether or not this assumed to represent systemic toxicity.
- Section 6.1 of Appendix A in the revised draft BRD refers to a proprietary product; a revised protocol for the flow cytometry measurements should provide a technical or performance standard.

The Panel was also asked whether it agreed that there could be cause for concern when a negative test substance result is accompanied by a concurrent positive control SI value that is significantly lower than the mean historical SI (even when this SI value is >ECt).

The Panel agreed that it would be a concern as it may make the detection of weak sensitizers problematic. The availability of any vehicle and negative control information and an integrated assessment of all available and relevant information (including structure-activity relationship considerations, etc.) for the test substance should be brought to bear in assessing the significance of such a test result. Given ethical considerations and bioresource use, retesting should not be undertaken lightly. The comparison of current positive control results to historical results is not an onerous requirement and can add value.

2.4.3 Future Studies

Finally, the Panel was asked if the available data support the proposed future studies presented in the revised draft ICCVAM test method recommendations for the LLNA: BrdU-FC.

The Panel agreed that the available data support the proposed future studies. It identified no additional data needs for specific classes of test materials that are not common to all variants of the LLNA. It is the view of the Panel that further work on the applicability domain of the LLNA and its nonradioactive variants could be undertaken with any of the family of LLNA tests, as the results will be uniformly applicable to all of this family of LLNA tests.

Furthermore, the Panel recommended that:

- The revised power calculations should be performed using the data provided for intralaboratory validation; the number of animals to provide performance equivalent to, or better than, the traditional LLNA should be determined.
- *The interlaboratory study is an absolute requirement for validation.*

- Alternate prediction models (e.g., multiple SIs similar to LLNA: DA and LLNA: BrdU-ELISA) should be considered.
- The ICCVAM-recommended LLNA performance standards (ICCVAM 2009) should be followed in the above-mentioned future work.
- ICCVAM should clearly state the intended purpose of the evaluation of any additional human data, and the potential read-across to established tests.
- Emphasis should be given to the use of ear swelling measurements to identify local irritation as a means of improving the performance of the traditional LLNA and other variants.

The Panel recommended that ICCVAM should work with NICEATM to support and facilitate three activities:

1. Conduct an independent audit of the original data to establish the validity of the data relied upon in the revised draft BRD, and simultaneously
2. Perform a revised evaluation of the minimum number of animals required using the intralaboratory study data; then, if n = 4 or 5 yields statistical power that is equivalent to or better than the traditional LLNA, the test should be considered sufficiently reliable to move on to a formal interlaboratory evaluation, then
3. If appropriate, commission a formal interlaboratory study to determine the transferability and reliability of the test method when used in different laboratories (i.e., an interlaboratory validation study). This testing should be conducted according to the technical specification set out in the ICCVAM-recommended LLNA performance standards (ICCVAM 2009). The study design should make provision for the live animal work being undertaken in one laboratory, and the flow cytometry in another.

Following such a study, it is the view of the Panel that this test method can be considered scientifically validated and ready for regulatory consideration if the Panel is satisfied the following requirements are met:

- An independent data audit is conducted confirming the acceptable quality of the data relied upon in the revised draft BRD.
- A revised evaluation of the minimum number of animals required should be conducted; then, if n = 4 or 5 yields statistical power that is equivalent to or

better than the traditional LLNA, an interlaboratory evaluation should be performed using the test.

- The interlaboratory study should produce results that satisfy the requirements in the ICCVAM-recommended LLNA performance standards (ICCVAM 2009).

The Panel recommended further consideration of statistical issues, including how to determine and evaluate classification methods (i.e., classification cutoff points). For example, "bootstrapping analysis" could be considered to determine the reliability of the SI cutoffs for identifying sensitizers and nonsensitizers. This would involve randomly sampling smaller subsets of the original dataset and calculating the cutoffs at which no false negatives and no false positives are produced. Such analysis would provide a measure of variability for the proposed cutoffs. The Panel also recommended that future interlaboratory validation studies should simultaneously evaluate intralaboratory reproducibility, using the appropriate statistics, to evaluate variation both within a laboratory and between laboratories. There should be a set of guidelines that describes the general principles and a basic method for estimating variation in both intralaboratory and interlaboratory experiments. Additionally, the Panel strongly recommended that a statistician actively participate in the preparation of future BRDs and formulation of ICCVAM recommendations.

2.4.4 Panel Comment on Animal Welfare

The Panel agreed that:

- The test, like the traditional LLNA, is more refined than the guinea pig tests as it relies upon an asymptomatic endpoint representing induction rather than elicitation of an established immune sensitivity, which reduces the distress and discomfort associated with a localized skin reaction and provides a quantitative measure of skin sensitization potential.
- However, it remains to be seen whether the number of animals required is less than, equal to, or greater than that used in the traditional LLNA.

3.0 Nonradioactive LLNA Protocol – The LLNA: Bromodeoxyuridine Detected by ELISA (BrdU-ELISA) Test Method

3.1 Review of the Revised Draft BRD for Completeness, Errors, and Omissions

The Panel was asked if there were any errors in the revised draft LLNA: BrdU-ELISA BRD that should be corrected, if omissions of existing relevant data had been identified, and if there was additional information that should be included.

According to the Panel, the only error that merits correction is in the Reference section of the revised draft BRD (see Section 12.0 of the revised draft BRD). The reference for the ISO standard should indicate that it is part of the 10993 series of standards. Omissions of existing relevant data or information that should be included are addressed below.

3.2 Review of the Validation Status of the LLNA: BrdU-ELISA

3.2.1 Substances Used for the Validation Studies

Seven substances were added to the database since dissemination of the January 2008 draft BRD. Among the new total of 31 substances with traditional LLNA data, 22 were classified by the traditional LLNA as skin sensitizers, and nine were classified as nonsensitizers.

ICCVAM asked the Panel whether it now considered the LLNA: BrdU-ELISA database representative of a sufficient range of chemical classes and physicochemical properties, such that the test method would be applicable for testing any of the types of chemicals and products typically tested for skin sensitization potential (see Section 6.3 of the revised draft BRD for a comparison of the substances tested in the LLNA: BrdU-ELISA with the ICCVAM-recommended LLNA performance standards reference substances).

The Panel responded that the LLNA: BrdU-ELISA database represents the range of chemical classes (i.e., 31 substances from 18 chemical classes) and physicochemical properties of the substances tested. The variety of substances agrees with the range of chemical classes of reference substances suggested in the ICCVAM-recommended LLNA performance standards (ICCVAM 2009). The database includes the types of substances typically tested for skin sensitization potential. The LLNA: BrdU-ELISA substances also exhibit a full dynamic range of responses and an adequate range of molecular weights, solubility, and proportion of solids and liquids.

According to the Panel, the LLNA: BrdU-ELISA may give less reliable results for certain substances (e.g., metal compounds or strong dermal irritants), similar to the traditional LLNA test method. The outcome of the evaluation of the applicability domain for the traditional

LLNA will also be relevant and applicable to the LLNA: BrdU-ELISA (i.e., any limitations from the traditional LLNA are also applicable to the LLNA: BrdU-ELISA).

Further studies with substances with comparative human, guinea pig and traditional LLNA should be continuously conducted, including irritants and formulations, to obtain more data on the LLNA: BrdU-ELISA method usefulness for identification of human sensitizers.

Initial studies suggest that the LLNA: BrdU-ELISA performs well with the one metal that was tested. Evaluation of more metals may be needed to determine if this result is representative of the utility of this test method for evaluating metals.

3.2.2 Test Method Accuracy

The test method developers recommend a decision criterion of SI ≥ 2.0 when evaluating the test method performance of the LLNA: BrdU-ELISA using a single SI decision criterion as a predictor of skin sensitization potential (see Section 6.2 of the revised draft BRD). This decision criterion yields accuracy of 84% (26/31), sensitivity of 77% (17/22), and specificity of 100% (9/9) (i.e., there were five false negatives and no false positives). This single decision criterion was further compared to guinea pig tests, human data/experience, and a criterion of SI ≥ 1.5, which yield higher sensitivity of 91% (20/22) and lower specificity of 67% (6/9) (see Section 6.5 of the revised draft BRD).

The Panel was asked (1) whether these comparisons were appropriate for assessing the accuracy of the LLNA: BrdU-ELISA using a single decision criterion and (2) whether the revised draft BRD had adequately evaluated and compared the relevance (e.g., sensitivity, specificity, false positive and false negative rates) of the LLNA: BrdU-ELISA to the traditional LLNA results, using a single decision criterion (SI ≥ 2.0) to distinguish between sensitizers and nonsensitizers.

According to the Panel, the comparisons made to traditional LLNA (SI ≥ 3.0), guinea pig, and human data (where available) were appropriate for assessing the accuracy of the LLNA: BrdU-ELISA using a single decision criterion (i.e., SI ≥ 2.0 or SI ≥ 1.5) to identify a test substance as a sensitizer. The Panel preferred the ICCVAM-recommended analyses using multiple stimulation index decision criteria, although the Panel expressed reservation about how results in the range of uncertainty might lead to the misclassification of certain substances. Additional information or testing is needed to classify substances with SI values in this range.

For data in the test method accuracy tables, the structure of the data would be clarified by graphing SI values for the traditional LLNA versus the SI values for the LLNA: BrdU-

ELISA with dividing lines at the cutoff values used to discriminate the sensitizers from the nonsensitizers.

The Panel agreed that the relevance of the LLNA: BrdU-ELISA using a single decision criterion (i.e., SI ≥ 2.0) has been properly assessed based on available data with comparisons to traditional LLNA results. Furthermore, these parameters have been adequately compared to available guinea pig and human data. Although the comparisons were appropriate, the analyses using multiple SI decision criteria are preferred. Any additional analysis based on data obtained from other studies may further strengthen confidence in the data concordance. Additional information or testing is needed to classify substances with SI values in the range of uncertainty.

The performance of the LLNA: BrdU-ELISA test method using classical statistical hypothesis testing (i.e., ANOVA or *t*-test) yielded accuracy of 81% (25/31), sensitivity of 86% (19/22) and specificity of 67% (6/9) (i.e., there were three false negatives and three false positives) compared to the performance of the traditional LLNA (see Section 6.5 of the revised draft BRD).

The Panel was asked if the accuracy analysis provided a comparison sufficient to decide whether to distinguish sensitizers from nonsensitizers by using a single SI decision criterion (i.e., SI ≥ 2.0) or by using classical statistical hypothesis testing.

The Panel responded that either the ANOVA (with a post-hoc Dunnett's test) or the *t*-test is appropriate for use in assessing the accuracy analysis of the LLNA: BrdU-ELISA using classical statistical hypothesis testing. However, a statistical difference from the negative vehicle control may not necessarily reflect a biologically important difference between a sensitizer and a nonsensitizer.

The Panel was asked whether the relevance (e.g., sensitivity, specificity, false positive and false negative rates) of the LLNA: BrdU-ELISA using classical statistical hypothesis testing had been adequately evaluated and compared to the traditional LLNA results, and the Panel determined it had. However, for data in the test method accuracy tables the structure of the data would be clarified by graphing SI values for the traditional LLNA versus the LLNA: BrdU-ELISA with dividing lines at the cutoff values used to discriminate the positives from the negatives.

Although the Panel agreed these comparisons are appropriate, it preferred the ICCVAM-recommended analyses using multiple SI decision criteria. The low specificity of the accuracy results compared to the traditional LLNA does not support the use of classical

statistical hypothesis testing as the only approach for determining whether a test substance is a sensitizer or nonsensitizer in the LLNA: BrdU-ELISA. Substances other than sensitizers, including irritants, also significantly increase proliferation. Lactic acid, for example, is an irritant but not a sensitizer. Table 6-8 of the revised draft BRD shows that using classical statistical hypothesis testing to determine whether the vehicle control group is different from the treated groups classifies lactic acid as a sensitizer. This and a limited number of other irritants that induce proliferation in the draining lymph nodes may be incorrectly regarded as sensitizers (i.e., false positives) (Montelius et al. 1994). The magnitude of the increased proliferation is increasingly important in differentiating between a sensitizer and a nonsensitizer. The Panel noted, however, that in cases where the sensitization potential of a test substance is uncertain, the use of classical statistical hypothesis testing in conjunction with a single SI decision criterion may contribute to a definitive skin sensitization identification of the test substance.

Evaluation of multiple decision criteria to identify substances as sensitizers or nonsensitizers found that SI $\geq$ 2.0 yielded no false positives, and SI < 1.3 produced no false negatives in the LLNA: BrdU-ELISA, compared to traditional LLNA results (see Section 6.7 of the revised draft BRD). ICCVAM asked the Panel (1) whether these comparisons were appropriate for assessing the accuracy of the LLNA: BrdU-ELISA using multiple decision criteria, and (2) whether the relevance (e.g., sensitivity, specificity, false positive and false negative rates) of the LLNA: BrdU-ELISA, using the SI $\geq$ 2.0 criterion for sensitizers and the SI < 1.3 criterion for nonsensitizers, had been adequately evaluated and compared to the traditional LLNA results.

The Panel responded that the comparisons to traditional LLNA data are appropriate to determine the optimal threshold values for identification of sensitizers and nonsensitizers. The most appropriate SI values for the LLNA are not immutable biological constants and should be determined by evaluation of test results. The approach adopted in the revised draft BRD did such an evaluation. However, for data in the test method accuracy tables, the structure of the data would be clarified by graphing SI values for the traditional LLNA versus the LLNA: BrdU-ELISA with dividing lines at the cutoff values used to discriminate the positives from the negatives.

The 0% false positive rate using SI $\geq$ 2.0 and the 0% false negative rate using SI < 1.3 support the conclusion that separate criteria should be employed to identify sensitizers (SI $\geq$ 2.0) and nonsensitizers (SI < 1.3). In addition, this approach allows for the use of additional data and information (e.g., molecular weight, peptide binding of intermediate compounds) to help make an informed decision regarding sensitization potential. Additional information or

testing is needed to classify substances with SI values in the range of uncertainty (i.e., $1.3 \leq SI < 2.0$).

In 2008, the Panel did not find sufficient power for using $SI \geq 1.3$ as the decision criterion. Even with a group size of eight animals, the power was only 50% (ICCVAM 2008). Power calculations might be necessary to determine if the sample size used is sufficient for those substances that are not definitively identified as sensitizers or nonsensitizers (i.e., substances in the range of uncertainty of $1.3 \leq SI < 2.0$).

The Panel determined that NICEATM's evaluation of the LLNA: BrdU-ELISA using the $SI \geq 2.0$ criterion for sensitizers and the $SI < 1.3$ criterion for nonsensitizers and its comparison to the traditional LLNA results were adequate. Additional analysis based on data obtained from future studies may further strengthen confidence in the data concordance. The Panel agreed that the revised draft BRD provided a partial rationale for the post hoc selection of these stimulation indexes as the two decision criteria and that importance should also be given to the corollary that these decision criteria may not apply when new data is considered. Table 6-6 of the revised draft BRD should be revised to delete the 'N' column and to add the number of substances evaluated to the table title.

Eleven of the 31 substances used to evaluate the LLNA: BrdU-ELISA yielded SI values that were greater than or equal to 1.3 and less than 2.0. Among these substances, 6/11 are sensitizers and 5/11 are nonsensitizers based on traditional LLNA results. Characteristics of each group (i.e., the six sensitizers and the five nonsensitizers) are described in Section 6.8 of the revised draft BRD.

The Panel was asked whether the number of sensitizers (i.e., six) and nonsensitizers (i.e., five) that yielded SI values that were greater than or equal to 1.3 and less than 2.0 was sufficient to identify characteristics (e.g., peptide-binding activity, molecular weight) potentially helpful in determining the skin sensitization hazard classification of such substances. If inadequate, the Panel was asked how many substances would be sufficient. If the number of substances was deemed sufficient, Panel members were asked to identify characteristics that are associated with these or similar substances and that may provide additional information about whether qualifying substances should be classified as sensitizers or nonsensitizers.

The Panel stated that the number of sensitizers and nonsensitizers in the range of uncertainty is not sufficient for definitive recommendations on the usefulness and limitations of such an approach. At present, the suggested characteristics (e.g., peptide-binding activity, molecular weight) may be of added value for a definitive identification of the substance, but they have

not been validated because the number of substances is too few. These characteristics were identified not only on the basis of this validation study; they represent cited characteristics of sensitizers in general. If more substances are tested using the LLNA: BrdU-ELISA, more reliable identification of crucial characteristics can be obtained. In general, information on peptide-binding activity, molecular weight, reliable structural alerts for skin sensitization (i.e., quantitative structure–activity relationships), skin penetration, and activation of immunocompetent skin cells could provide additional information to inform a definitive classification. While peptide-binding activity shows promise in assisting to identify substances with SI values in the range of uncertainty, this approach has not been sufficiently evaluated. To evaluate methods that reliably distinguish between sensitizers and nonsensitizers, further evaluation for test substances in the range of uncertainty should be undertaken as new datasets and tools become available. However, the Panel emphasized that the number of substances in the range of uncertainty is too few to permit any meaningful analysis. The Panel also emphasized that there may be a need to include multiple different types of tests/information in a decision tree to make a classification decision (i.e., an integrated assessment of all available and relevant information).

The LLNA: BrdU-ELISA test method was developed before publication of the ICCVAM-recommended LLNA performance standards (ICCVAM 2009); therefore, these performance standards were not used to evaluate the test method. The Panel was asked whether the fact that the LLNA: BrdU-ELISA produced one false negative result for one of the 10 required performance standards substances affected the test method's validation status, despite the fact that the ICCVAM-recommended LLNA performance standards were not used to evaluate its validity. The other nine required performance substances tested yielded concordant results with the traditional LLNA (see Section 6.3 of the revised draft BRD).

The Panel deemed correct identification of nine of the 10 LLNA test substances required by the performance standards sufficient. The one discordant result (i.e., 2-mercaptobenzothiazole) is a false negative in the LLNA: BrdU-ELISA compared to the traditional LLNA based on only one LLNA: BrdU-ELISA experiment. In addition, the SI for this substance is in the range of uncertainty ($1.3 \leq SI < 2.0$). If further parameters are considered (e.g., high peptide reactivity) the substance might be identified as a sensitizer. It might be that the substance itself is problematic, given that this substance has provided inconsistent results in other nonradioactive assays and even in guinea pig tests.

3.2.3 Test Method Reliability (Intra- and Interlaboratory Reproducibility)

ICCVAM asked the Panel if the reproducibility of the LLNA: BrdU-ELISA using SI ≥ 2.0 for sensitizers and SI < 1.3 for nonsensitizers had been adequately evaluated and compared to the traditional LLNA (see Section 7.3 of the revised draft BRD).

The Panel agreed that reproducibility was evaluated adequately and appropriately compared to the traditional LLNA, although the independence of experiments was not documented. The following information should be added to the final version of Table 7-3 in the revised draft BRD (to make it consistent with Table 7-1 of the LLNA: DA revised draft BRD):

- SI values for each dose tested
- Sample sizes
- Mean of means
- Variance of means
- CV of means

The table should include all the information needed to perform an F-test (i.e., sample sizes as relevant to degree of freedom calculations, means on a log scale, and mean square errors for the within laboratory and between laboratory comparisons).

The Panel recommended that the final versions of Tables 7-7 and 7-8 from the revised draft BRD include sample sizes, means, standard deviation, and CV. These tables should be supplemented with a graphical comparison of SI values for the traditional LLNA versus LLNA: BrdU-ELISA for each substance in each laboratory. Sample sizes, means, SDs should be included consistently in all relevant tables. Variation (i.e., SD) should be measured on a log scale and then converted to antilog for reporting in the table. The final versions of Tables 7-9 and 7-10 in the revised draft BRD should be accompanied by a figure with traditional LLNA SI versus LLNA: BrdU-ELISA SI. The figure should be designed so that substances tested could be distinguished by different symbols, and the different laboratories could be distinguished by numbers on the graph plotted on a log scale.

In a multilaboratory validation study of the LLNA: BrdU-ELISA, JSAAE tested 10 coded substances (seven sensitizers and three nonsensitizers) at seven laboratories. Two sensitizers and one nonsensitizer were tested in all seven laboratories; the remaining substances were each tested in three laboratories. JSAAE's Validation Management Team selected the vehicles and the concentrations of the substances tested. The Panel was asked about (1) any limitations apparent based on this study design and (2) any concerns about the fact that

vehicles and concentrations were provided to the participating laboratories by the Validation Management Team.

The Panel expressed more concern for reducing the variation in the test than for transferability; therefore, it had no concerns about the fact that vehicles and concentrations were provided by the Validation Management Team. In fact, this may have been advantageous by eliminating factors outside the laboratory. The Panel recommended performing isotonic regression prior to interpolation when calculating ECt values.

3.2.4 Consideration of All Available Data and Relevant Information

The Panel was asked if the revised draft LLNA: BrdU-ELISA BRD adequately considered all the relevant data identified in published or unpublished studies employing this test method.

The Panel responded that it was unaware of any relevant data that had been omitted from the revised draft BRD.

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

3.3.1 Test Method Usefulness and Limitations

The Panel was asked whether the available data and test method performance (accuracy and reliability) support the revised draft ICCVAM test method recommendations for the LLNA: BrdU-ELISA in terms of the proposed test method's usefulness and limitations.

The Panel agreed that the available data and test method performance supported the revised draft ICCVAM test method recommendations. The documented accuracy and reliability of the method support its usefulness to identify potential sensitizers and nonsensitizers, with specific defined limitations when the resulting SI is greater than or equal to 1.3 and less than 2.0. In this regard, more detailed guidance is required to further evaluate a substance producing an SI falling in the range of uncertainty.

ICCVAM proposed using two decision criteria to distinguish between sensitizers and nonsensitizers in the LLNA: BrdU-ELISA: SI $\geq$ 2.0 to identify substances as sensitizers and SI < 1.3 to identify substances as nonsensitizers. The Panel was asked (1) if the accuracy analysis supports the revised draft ICCVAM test method recommendation that the decision criterion for sensitizers be based on SI $\geq$ 2.0 and the decision criterion for nonsensitizers be based on SI < 1.3 and (2) if it considered these multiple SI values appropriate for use in skin sensitization classification. The Panel was queried as to which decision criterion or criteria

(e.g., classical statistical hypothesis testing or other SI values) it considered most appropriate, if it considered the use of multiple SI values to be inappropriate.

According to the Panel, the accuracy analysis supports the revised draft ICCVAM test method recommendations (i.e., the decision criterion for sensitizers be SI ≥ 2.0 and the decision criterion for nonsensitizers be SI < 1.3). The model proposed is empirical and derived from the data, and gives the best combination of the minimum risk of false positives and false negatives and the lowest number of "uncertain" results. Since using two decision criteria allows for a more definitive identification of sensitizers and nonsensitizers, this approach provides animal welfare benefits by reducing further tests that might be required in instances where the hazard classification of a substance is not as clear. In addition, one can use statistical analysis and/or other data and information (e.g., peptide binding, QSAR, skin penetration information, etc.) to provide more information on compounds that fall in the range of uncertainty. Additional guidance is needed for users on how to identify substances with SI values in the range of uncertainty (i.e., 1.3 ≤ SI < 2.0). Although the use of multiple decision criteria to identify sensitizers and nonsensitizers appears reliable, the Panel questioned how results in the range of uncertainty would be useful for regulatory purposes.

The Panel was asked whether concerns about potential false positives or false negatives that may occur in this test method would be resolved by using multiple decision criteria. Other suggestions for additional guidance or limitations were solicited from the Panel.

The Panel replied that using SI < 1.3 to identify nonsensitizers and SI ≥ 2.0 to identify sensitizers provides a definitive skin sensitization identification with a greater degree of reliability than when test materials are evaluated in any system that applies a single test point for the evaluation of potential skin sensitizers. However, for substances exhibiting an SI value greater than or equal to 1.3 and less than 2.0 additional information and testing should be used to reach a classification decision. The Panel noted that because the decision criteria chosen to identify sensitizers and nonsensitizers were based on a post hoc analysis, prospective testing with the test method might affect the proposed model. For this reason, future data generated should be routinely evaluated to determine if the proposed classification decision criteria could be improved.

The Panel agreed that, especially for substances exhibiting an SI value greater than or equal to 1.3 and less than 2.0, a statistical approach (i.e., classical statistical hypothesis testing of absorbance values of treated groups versus control group) in addition to SI values might contribute to a definitive skin sensitization identification (i.e., classical statistical hypothesis testing of absorbance values of treated groups versus control group).

The Panel also noted that high variability in the test could result in misclassification of certain substances, especially those having SI values in the range of uncertainty. Sensitizers with highly variable results may not be identified as such using classical statistical hypothesis testing methods (i.e., the variation in the data will not allow a meaningful difference from the vehicle control to be detected).

Furthermore, nonsensitizers that produce data with low variability will be detected as sensitizers by classical statistical hypothesis testing methods. The Panel affirmed that the SI is the primary criterion, but if variability is large, the reliability of the test result may be inconclusive. Further statistical evaluation could be informative.

ICCVAM recommended that when an SI value greater than or equal to 1.3 and less than 2.0 is obtained in the LLNA: BrdU-ELISA, users should carefully interpret the results in an integrated decision strategy in conjunction with all available and relevant information (e.g., dose response information, statistical analyses, peptide-binding activity, molecular weight, results from related chemicals, other testing data) to determine if there is adequate information for an accurate sensitization hazard classification or if additional testing is necessary.

The Panel agreed with ICCVAM's recommendation but added that additional guidance should be provided on how this should be done.

The Panel stated that dose response information, statistical analyses, peptide-binding activity, molecular weight, QSAR using structural alerts for skin sensitization, results from related chemicals, other testing data, etc., should be considered first as aids in determining the skin sensitization classification of a substance with an SI value greater than or equal to 1.3 and less than 2.0. However, if this information is inconclusive, additional animal tests (i.e., the traditional LLNA) should be considered. According to the Panel, such supplementary testing should only be conducted if absolutely necessary. Thus, it is also important to consider the intended use of the substance as this can factor into determining the skin sensitization classification of a substance. For instance, if the substance is intended for consumer products, the HRIPT might be useful to confirm sensitization potential, or the lack thereof.

ICCVAM asked the Panel under what circumstances LLNA: BrdU-ELISA results with an SI value greater than or equal to 1.3 and less than 2.0 could be used as part of an integrated decision strategy (i.e., in conjunction with all available and relevant information such as dose response information, statistical analyses, peptide-binding activity, molecular weight, results from related substances, other testing data) to assign a sensitization hazard classification.

The Panel responded that the test method should be performed under GLP conditions, and that personnel performing the test method should strictly adhere to the standard operating procedure, including controls. An experiment yielding such a result should preferably be repeated; and all indicated additional information should be reviewed. This could be considered an initial approach in all cases, but because there is information on only 11 compounds in the range of uncertainty more information is needed to determine if this approach is the best.

The interlaboratory reproducibility standard from the ICCVAM-recommended LLNA performance standards (ICCVAM 2009) is ECt values, from each laboratory, of 0.025% to 0.1% for DNCB and 5% to 20% for HCA. The Panel was asked if it was concerned that two of seven laboratories that tested DNCB and two of six laboratories that tested HCA in the JSAAE study produced EC2 values (estimated concentrations of a substance needed to produce a stimulation index of 2) outside of the acceptability range.

The Panel responded that the test is reproducible and, since the most prevalent outcome meets the statistical criteria, it was not concerned with the fact that outliers were generated in the positive control data. Considering that the radioisotope measurement in the traditional LLNA is more sensitive than the technique for the LLNA: BrdU-ELISA, and that the analysis of EC3 values in the traditional LLNA was based on a larger dataset, it is appropriate to adjust the acceptability range dependent on the method used for measurement of the endpoint. Although the qualitative performance was acceptable in the interlaboratory study, the quantitative data for two of the laboratories suggests a relatively high degree of variability, which justifies the routine use of appropriate positive and negative controls. Any recommendation on the validation status of this test method should account for this variability**.**

3.3.2 Test Method Protocol

The Panel was asked if the available data support the revised draft ICCVAM test method recommendations for the LLNA: BrdU-ELISA procedure in terms of the proposed test method standardized protocols.

The Panel agreed that the intra- and interlaboratory validation studies, which included training of participating laboratories, had resulted in a standardized procedure. The revised draft ICCVAM recommendation reflects current evidence and best practice. The evaluation of the variation in positive control responses over time probably has wider applicability to a broad range of test systems. The Panel agreed that laboratories should maintain a historical database of positive control SI values and some measure of variability over time.

3.3.3 Future Studies

The Panel was asked whether the available data support the revised draft ICCVAM test method recommendations for the LLNA: BrdU-ELISA in terms of the proposed future studies.

The Panel considered that the data supported the revised draft ICCVAM-recommended future studies and that additional decision criteria and guidance should be identified for substances exhibiting SI values greater than or equal to 1.3 and less than 2.0. For any test method in which multiple decision points are used and a subset of the test materials are initially identified in a range of uncertainty, the secondary decision criteria should be reassessed as additional data and discriminators become available (e.g., high-quality human ACD data).

According to the Panel, additional studies should be conducted for other substances, including metals, irritants, and formulations, that have comparative human, guinea pig and traditional LLNA data. While specifically applicable to the LLNA: BrdU-ELISA, these data needs are common to all variants of the LLNA. Regarding the impact of irritants, the proposed future studies may provide insights to better explain why results obtained using the LLNA: BrdU-ELISA and traditional LLNA may be discordant.

While peptide reactivity shows promise in assisting to identify substances with SI values in the range of uncertainty, this approach has not been sufficiently evaluated.

There should be analyses of all three nonradioactive LLNA methods of the process and results used to define the cutoff values (i.e., quality assessment of the process), and a protocol should be developed for evaluating such cutoffs that future developers could apply during development of new methods.

The Panel recommended further consideration of statistical issues, including how to determine and evaluate classification methods (i.e., classification cutoff points). For example, "bootstrapping analysis" could be considered to determine the reliability of the SI cutoffs for identifying sensitizers and nonsensitizers. This would involve randomly sampling smaller subsets of the original dataset and calculating the cutoffs at which no false negatives and no false positives are produced. Such analysis would provide a measure of variability for the proposed cutoffs. The Panel also recommended that future interlaboratory validation studies should simultaneously evaluate intralaboratory reproducibility, using the appropriate statistics, to evaluate variation both within a laboratory and between laboratories. There should be a set of guidelines that describes the general principles and a basic method for estimating variation in both intralaboratory and interlaboratory experiments. Additionally,

the Panel strongly recommended that a statistician actively participate in the preparation of future BRDs and formulation of ICCVAM recommendations.

3.3.4 Performance Standards

The Panel was asked if the LLNA: BrdU ELISA adheres to the essential test method components in the ICCVAM-recommended LLNA performance standards (ICCVAM 2009) such that the performance standards could be used to evaluate future nonradioactive versions similar to the LLNA: BrdU ELISA that also adhere to these essential test method components.

The Panel agreed that the ICCVAM-recommended LLNA performance standards state the essential test method requirements, and the LLNA: BrdU-ELISA adheres to them such that it should be considered mechanistically and functionally similar. The only variation with the traditional LLNA is the means by which lymphocyte proliferation during the induction phase is evaluated.

[This Page Intentionally Left Blank]

4.0 Use of the LLNA for Testing Pesticide Formulations and Other Products, Aqueous Solutions, and Metals

The Panel comprises experts with knowledge in the evaluation of a range of test materials, but it is by no means expert in all of the product classes for which skin sensitization potential should be evaluated. The Panel also acknowledges that information and data gaps exist which prevent a full understanding of ACD epidemiology in humans. The test materials for which data are provided in the revised draft Addendum cover only a subset of the active ingredients used in each of the relevant product classes, and the frequency of use of those ingredients within those product classes is not noted in the revised draft Addendum. The Panel recommends that Federal agencies considering the results of this validation process assess how representative the test materials and findings in the revised draft Addendum are relative to substances of interest. In particular, the agencies should assess the chemical classes used in, and the range of biological effects of, the materials and products in which they have an interest.

4.1 Review of the Revised Draft Addendum for Completeness, Errors, and Omissions

In regard to the revised draft Addendum to the 1999 ICCVAM LLNA Panel report (ICCVAM 1999) the Panel was asked to comment on any errors that should be corrected or omissions of relevant data and/or information that should be included.

The Panel responded that the additions to the 2008 material were noted in the revised draft Addendum and have been properly incorporated within the 2009 material. The Panel acknowledged that the chemical characterization of some pesticide formulations was incomplete. The Panel further stated that the revised draft Addendum discussed the quality of all the data considered (i.e., pesticide formulations, dyes, natural complex substances, metals and substances tested in aqueous solutions) and discussed the lack of data audit for studies not done in accordance with GLP guidelines (see Section 6.0 of revised draft Addendum for details). However, all data relied upon seemed fit for its purpose.

The revised draft Addendum included material from 25 data sources, and there were no significant errors requiring correction. A Panel member did, however, note during the public meeting that some information regarding the natural complex substances[6] was omitted. The relationship between LLNA, guinea pig, and human data for major constituents (substances constituting at least 70%) of some of the natural complex substances and the LLNA results of the natural complex substances themselves were omitted. In general, the Panel expressed surprise at the lack of guinea pig data for the natural complex substances presented.

[6] Essential oils and extracts are referred to as “natural complex substances” within the industry, or as “substances of unknown or variable composition” under REACH.

It is acknowledged in the revised draft Addendum that guinea pig data were derived from test methods not formally validated for testing formulations, and the Panel recommended that this be added as a footnote where traditional LLNA findings were being directly compared with guinea pig data elsewhere in the revised draft Addendum.

The Panel observed that its instruction did not, in the absence of new evidence or analysis, touch upon the testing of metals. The revised draft Addendum did not address, or challenge, issues raised in 2008 by Dr. Dagmar Jírová with respect to potential solvent effects on the testing of materials containing nickel. At that time, Dr. Jírová expressed the minority opinion that the traditional LLNA could be used even for testing nickel compounds when the appropriate vehicles, in particular aqueous vehicles, are used (ICCVAM 2008).

The Panel recommended rectifying the minor formatting inconsistencies (e.g., whether values were given to one or two decimal places) in appendices. The Panel recommended changing the term *essential oils* to *natural complex substances* (terminology consistent with Registration, Evaluation, and Authorisation of Chemicals [REACH]) in the final ICCVAM documents, for consistency.

Ideally, the final Addendum and future studies, where available, should include sufficient information, including the original data, details of actual test circumstances, and the precise nature of the formulations, so subsequent investigators can do meta-analyses.

The Panel recommended that the term EC3 (%) be replaced with EC3 or ECt as appropriate.

4.2 General Questions

The Panel stated in its 2008 review that the term *mixtures* was used too broadly (i.e., can represent an infinite number of materials) and it would be beneficial to specify types or formulations examined (ICCVAM 2008). This concern was addressed in the revised draft Addendum by dividing the substances considered into pesticide formulations, dyes, natural complex substances, and substances tested in aqueous solutions (this group included predominantly pesticide formulations tested in aqueous solutions), and analyzing the data for each group separately.

The Panel was asked if it considered these groups appropriate, given the composition of the updated database. It was asked, if not, into which groups it would suggest the substances in the revised database be divided for analysis, and according to what rationale.

The Panel agreed that the terms used to group the information submitted for the updated analysis are sensible and help to divide the dataset into useful and relevant categories for analysis, and that the product categories selected, in line with MeSH classifications, fit well with the nature and range of materials in the database. Such categories indicate classes of

materials for which there exist, or do not exist, LLNA data and thus are informative for industry and regulatory agencies.

The Panel commented that if the extracts of the medical device-derived test materials had been chemically characterized before the sensitization tests had been carried out, the performance of the test method with respect to another class of test material could have been evaluated in the revised draft Addendum for the LLNA applicability domain.

The Panel also commented that since pesticide formulations tested in aqueous solutions were predominant in the database, they should be separated from the other substances tested in aqueous solutions to avoid any inherent bias in the range of materials analyzed.

The Panel was asked if there other types or classes of substances it would recommend for evaluation in the future and if yes, which types or classes of substances would be recommended.

The Panel noted that the revised draft Addendum does not consider many classes of formulations to which humans may be exposed, by intention or by accident, such as:

- Metalworking fluids
- Fuels
- Petroleum products used as lubricants
- Detergents and other cleaning agents
- Enzymes used in cleaning products
- Chemical household products
- Chemical (low molecular weight) pharmaceutical products
- Medical device materials (chemically characterized extracts)
- Nanomaterials (e.g., titanium oxide)

The Panel viewed that available data for substances within these classes may prove informative for human health.

The Panel commented that, while pesticides are significant contaminants of food, they might not act as skin sensitizers in that context, even if they are toxicants and disruptors of some immune responses. All pesticides are evaluated for skin sensitization potential, which might be an issue for farmers but not for the general population. The importance of adequately screening materials intended for human exposure was noted.

ICCVAM asked the Panel if it considered the inclusion of studies performed with BALB/c mice appropriate for these analyses, given that this strain has not been validated for use in the ICCVAM-recommended LLNA protocol (Appendix A of ICCVAM 2009). Note that, in a NICEATM evaluation of the effect of reducing the sample size for the LLNA from five animals to four, it was stated that the pattern of LLNA responses seen in BALB/c mice is similar to that seen in CBA mice (see Appendix B of ICCVAM 2009).

The Panel commented that, for consistency, only female CBA mice should be used in LLNA studies. The use of any other strain (or substrain) or gender of mouse should be explained and justified by the investigator and accompanied by evidence of similarity of responses. Some relevant immunological functions are known to vary between strains, and the recommended strain was originally selected on the basis of what was known about the detail of its immunological functions.

However, the Panel also concluded that the BALB/c-derived data analyzed in the revised draft Addendum should be considered valid, with a simple caveat regarding the mouse strain, to avoid discounting a large portion of the data submitted on pesticide formulations.

Because BALB/c mice are commonly used for LLNA testing, as suggested above, the Panel was asked whether available data support the use of any BALB/c substrains for LLNA testing. If not, it was asked, whether a high priority should be placed on the collection of additional comparative BALB/c and CBA data for a more comprehensive assessment.

The Panel was not aware of any substrain(s) of BALB/c mice that could be specified as particularly well-suited for use in the LLNA test method. For practical management of effort and economy, ICCVAM should not give high priority to an extended comparative study of the relative LLNA responses of CBA versus BALB/c mice.

As the use of the LLNA increases, it is likely that the LLNA will be used to test new product types that have not been previously tested in the LLNA or guinea pig and for which there are no comparative human reference data. With this in mind, the Panel was asked if it considered it to be necessary to validate the LLNA each time new types or classes of substances and/or formulations are proposed for testing in the LLNA.

The Panel expressed a strong desire to avoid revalidation of the LLNA, unless there is a biologically-based rationale. The repeated validation for new classes/types of test substances would substantially reduce the ability of widespread adoption of the LLNA and reduced LLNA to reduce and refine animal use in skin sensitization testing. For new classes of test materials (e.g., nanomaterials), an integrated assessment of all available and relevant information should be conducted. This should include computer-assisted structure-activity

relationships, prediction/measurement of biotransformation to potential reactive species, and possibly peptide, protein, or lipid binding. The Panel agreed that if any variant of the LLNA is validated for use to test novel classes, then the findings should be relevant to the family of validated LLNA tests.

The Panel also commented that the established guinea pig tests used for evaluating skin sensitization potential were never validated for formulations. The final ICCVAM BRD should make this clear and should note that similar uncertainties would be associated with its use for future novel classes of test materials.

4.3 Pesticide Formulations

In the database of 22 pesticide formulations with both guinea pig and LLNA reference data for the pesticide formulation itself, nine active ingredients (out of approximately 450)[7] are represented (i.e., one insecticide, six herbicides, and two fungicides) and six or seven classes of pesticide types (out of approximately 40 classes of compounds). No antimicrobials, biopesticides, or microbials are represented. ICCVAM asked the Panel if it believed that the database of substances evaluated represent a sufficient range of pesticide formulations typically tested for skin sensitization.

The Panel noted that from the revised draft Addendum it was difficult to determine whether there was adequate representation with this small sample, and commented that the number of active ingredients among the pesticide formulations that were evaluated was low compared to the total number of active ingredients tested in the LLNA. However, a Panel member provided information during the public meeting that the range and nature of the active ingredients and types of formulations covers much of the existing product range. Thus, it was the view of the Panel that the database was sufficiently representative.

The Panel expressed an interest in learning what proportion of U.S. Environmental Protection Agency notifications/dossiers were covered by the range of materials evaluated in the revised draft Addendum, and to what extent (if any) product class was related more to quantities and marketing claims rather than the nature of the active ingredient. The Panel recommended that the final BRD include any available information on occupational contact dermatitis caused by pesticide formulations, to complement the available guinea pig data (recognizing that HRIPT data or human maximization test [HMT] data were not available).

The Panel was asked its opinion as to whether the relevance (e.g., accuracy/concordance, sensitivity, specificity, false positive and false negative rates) of the LLNA for testing

[7] Personal communication, Registrant – Division of the Office of Pesticide Programs, U.S. Environmental Protection Agency.

pesticide formulations had been adequately evaluated (refer to Section 5.0 of the revised draft Addendum).

It was the Panel's opinion that, with respect to the available data set, the performance characteristics, reproducibility, and reliability of the LLNA had been adequately and completely explored as described in the revised draft Addendum and that the methods of data analysis were appropriate. The Panel noted that concordance values for all paired comparisons (LLNA versus guinea pig versus clinical test outcome) had been calculated and discussed. The Panel also highlighted that discrepancies between guinea pig tests and the LLNA were not fully emphasized when comparing the outcomes from these tests.

The Panel indicated that, where comparative data were available, the LLNA tended to identify more sensitizers than did guinea pig tests (predominantly Buehler tests which are considered to be less sensitive than the guinea pig maximization test [GPMT] [Basketter et al. 1993; Frankild et al. 2000]) but missed no materials that the guinea pig tests classified as sensitizers. A Panel member provided information during the public meeting that active ingredient data for pesticide formulations indicates that the LLNA is similar to the GPMT and better than the Buehler test in assessing skin sensitization potential of pesticide formulations (Gehen et al. 2009). In the absence of comparable human data, the significance of this finding was unclear, particularly as a number of substances were tested in the LLNA at potentially irritating concentrations.

Finally, the Panel noted the high false positive rate (53% [10/19]), which resulted in relatively low specificity (47% [9/19]) and accuracy (54% [12/22]) (see Table 5-3 in revised draft Addendum). The high rate of false positive chemicals may be a cause for concern, but may be inherent to the product and/or chemical class, the fact that substances were tested at concentrations that produce skin irritation, the fact that the Buehler test may be less sensitive than the GPMT, and the fact that LLNA detects the induction phase of skin sensitization.

The Panel was asked, in light of the fact that only a relatively small subset of registered, formulated, active pesticide ingredients and classes were analyzed, if the revised draft Addendum adequately characterizes the usefulness and limitations of the LLNA for testing pesticide formulations based on the accuracy analyses.

It was the Panel’s opinion that the comparison for the 22 pesticide formulations with LLNA and guinea pig data for the pesticide formulation reflects the available information, although it was noted that the comparison did not reflect data for a large number of pesticide formulations for which only guinea pig data for the active ingredient was available. The Panel noted that important concordance statistics were presented, as well as highlights of discrepancy against guinea pig data, discussion of the use of an aqueous vehicle, and the use

of an alternate strain of mice. The rate of false positives may be a concern (or may not, see immediately above). In the future, this may be resolved in part by better discrimination between skin irritants and sensitizers by improving the test method, or by postmarketing surveillance. More mechanistic studies may be needed to compare LLNA with guinea pig data. The Panel also noted, however, that the performance of the LLNA with nonaqueous solvents and vehicles could not currently be evaluated for this product class as data are not available.

The Panel was asked whether the revised draft Addendum adequately considered all the relevant data from published or unpublished studies that used the LLNA to test pesticide formulations. According to the Panel, there were no data of which it was aware that could be used to supplement the analysis of the 22 pesticide formulations presented in the revised draft Addendum, although a Panel member provided information during the public meeting that guinea pig data for pesticide formulation active ingredients was available, as was data comparing sensitization potential of pesticide formulation active ingredients in the LLNA, GPMT, and Buehler tests (Gehen et al. 2009). At the public meeting, the Panel was informed that additional data comparing 1% Pluronic L92 to other vehicles exists, and that attempts would be made to provide it to NICEATM.

ICCVAM asked the Panel if it considered the inclusion of studies performed with the vehicle 1% Pluronic L92 appropriate for these analyses, since this vehicle has not been validated for use in the ICCVAM-recommended LLNA protocol (Appendix A of ICCVAM 2009). If so, it was asked, should 1% Pluronic L92 be added to the preferred vehicles list, for water-soluble substances, in the ICCVAM-recommended LLNA protocol (note that substances tested at a concentration of 100% contain no vehicle).

The Panel concluded that inclusion of LLNA studies performed with the vehicle 1% Pluronic L92 for pesticide formulations appeared valid and permissible, based on the limited information available. The peer-reviewed literature stated that the selection of 1% Pluronic L92 was based on wetting properties, lack of irritant activity, low toxicity, and known actions in enhancing skin penetration (Ryan et al. 2002; Boverhof et al. 2008).

The Panel believed that the studies using 1% Pluronic L92 were appropriately analyzed. The Panel inferred from the test results that there was reason to believe that this vehicle delivered sufficient concentrations of test articles and allowed appropriate exposure throughout the dosing period. The Panel recommended that results of studies with aqueous vehicles continue to be accrued for analysis and that alternative aqueous vehicles be compared with 1% Pluronic L92. ICCVAM should consider including 1% Pluronic L92 as an acceptable vehicle

in the ICCVAM-recommended LLNA protocol (Appendix A of ICCVAM 2009) when the protocol is next reviewed.

4.4 Dyes

The Panel was asked if it considered that the database of evaluated substances (i.e., six) represented a sufficient range of dyes that are typically tested for skin sensitization potential.

The Panel stated that it was not possible to answer this question affirmatively. There is a lack of information on the number of dye substances commonly used in products designed for intentional dermal application or exposure which require characterization of sensitization hazard. However, unless there are unique physicochemical properties associated with these test materials and classes of test materials that might affect their ability to interact with immune processes, the Panel considered dyes as candidates for testing in the LLNA.

The Panel noted that it was unclear how many chemical classes of dyes were routinely produced and what proportion of dyes might be difficult to test in the guinea pig models, in which some highly colored materials pose difficulties due to their complex nature.

ICCVAM asked the Panel whether the relevance (e.g., accuracy/concordance, sensitivity, specificity, false positive and false negative rates) of the LLNA for testing dyes had been adequately evaluated (refer to Section 5.0 of the revised draft Addendum).

The Panel agreed that the analysis reflects the available information, and noted that concordance statistics were presented, as well as highlights the discrepancies of LLNA data against guinea pig data. However, due to the paucity of dye data available for evaluation, the analyses were not meaningful (i.e., false positive rate of 100% [1/1]).

The Panel was asked whether the revised draft Addendum adequately characterized the usefulness and limitations of the LLNA for testing dyes based on the accuracy analyses.

The Panel responded that the revised draft Addendum partially characterized the usefulness and limitations of the LLNA for testing dyes based on the accuracy analyses. It also, according to the Panel, reflected the limited nature of the data provided. The Panel indicated that the analysis does discuss the outcome of the concordance evaluation and points out the major discrepancy between guinea pig results and those from the LLNA.

The Panel was asked if all the relevant data identified in published or unpublished studies conducted using the LLNA for testing dyes been adequately considered in the revised draft Addendum.

The Panel was unaware of any other publicly available data that could be used to supplement the analysis presented in the revised draft Addendum, although the Ecological and Toxicological Association of Dyes and Organic Pigment Manufacturers would likely be a source of additional information.

4.5 Natural Complex Substances

The Panel was asked if it considered the database of 12 evaluated substances representative of a sufficient range of natural complex substances that are typically tested for skin sensitization potential.

The Panel commented that the data were limited. The ability to reach a judgment solely from the revised draft Addendum was confounded by the same issues as for the small set of dyestuffs. However, unless there are unique physicochemical properties associated with these test materials and this class of test materials that might affect their ability to interact with immune processes, the Panel considered natural complex substances candidates for testing in the LLNA.

According to the Panel, fragrance use in consumer products is widespread, yet no information was offered on the relative use of the natural complex substances included in the revised draft Addendum, as compared to other substances. The Panel was unaware of how many chemical classes the natural complex substances might be derived from or in how many product classes they might be found. Finally, the Panel noted that the database of 12 substances in the recent analyses shows lower accuracy (42% [5/12]) compared to the database of 74 substances (72% [53/74]) from the analyses comparing LLNA versus human results from ICCVAM 1999 (shown in Table 5-7 of the revised draft Addendum).

ICCVAM asked the Panel if the relevance (e.g., accuracy/concordance, sensitivity, specificity, false positive and false negative rates) of the LLNA for testing natural complex substances had been adequately evaluated in the revised draft Addendum (refer to Section 5.0 of the revised draft Addendum).

The Panel agreed that the analysis reflects the available information, and noted that concordance statistics were presented, as well as highlights of the discordance of LLNA data against HMT data.

The Panel was questioned about whether, based on the accuracy analyses, the revised draft Addendum adequately characterized the usefulness and limitations of the LLNA for testing natural complex substances. If not, it was asked what additions or changes should be made.

The Panel commented that the revised draft Addendum did not adequately characterize the usefulness and limitations of the LLNA for testing natural complex substances because there was a limited database and no guinea pig data for comparison. The Panel indicated that no conclusions could be drawn regarding usefulness or limitations due to the small number of natural complex substances evaluated. However, a Panel member indicated during the public meeting the existence of additional information, including guinea pig, human, and LLNA data, on major components (substances constituting at least 70%) of these natural complex substances. These data indicate a relationship between results from all three tests with these major components. The Panel noted that the data show marked discordance between the LLNA and HMT data.

ICCVAM asked the Panel if the revised draft Addendum adequately considered all the relevant data identified in published or unpublished studies that used the LLNA to test natural complex substances.

Other than the information provided by a Panel member during the public meeting (see above), the Panel was unaware of any additional data that could be used to supplement the analysis presented in the revised draft Addendum.

4.6 Substances Tested in Aqueous Solutions

The Panel was asked if it considered that the evaluation of substances in aqueous solutions represented a sufficient range of those substances tested in aqueous solutions that are typically tested for skin sensitization potential. (In this revised draft Addendum, aqueous solutions are characterized as containing 20% water or more.) Note that many of the pesticide formulations discussed in the revised draft Addendum were also included in this analysis, so the database has the same limitations.

The Panel did not consider the database of substances evaluated to be representative of the full range of substances tested in aqueous solutions. However, unless there are unique physicochemical properties associated with these test materials and classes of test materials that might affect their ability to interact with immune processes, the Panel considered substances tested in aqueous solutions as candidates for testing in the LLNA.

The Panel commented that, given that the materials tested with aqueous vehicles were drawn almost exclusively from the pesticide and medical device groups, there was inherent bias in the range of materials. The Panel also noted that these comments were based upon the availability of suitable data for inclusion, rather than any flaw in the way the revised draft Addendum was compiled.

ICCVAM asked the Panel if the relevance (e.g., accuracy/concordance, sensitivity, specificity, false positive and false negative rates) of the LLNA for testing substances in aqueous solutions had been adequately evaluated (refer to Section 5.0 of the revised draft Addendum). The Panel agreed that the analysis reflected the available information and that appropriate concordance statistics were presented.

The Panel noted the high false positive rate of 50% (10/20) (see Table 5-12 in the revised draft Addendum), which produced relatively low specificity (50% [10/20]) and accuracy (54% [13/24]). Many of the substances tested in aqueous solutions that were included in the analysis were pesticide formulations tested in an aqueous vehicle. Thus, similar to the pesticide formulation analysis, the high false positive rate for substances tested in aqueous solutions is for guinea pig data that is primarily from Buehler tests (considered to be less sensitive than the GPMT [Basketter et al. 1993; Frankild et al. 2000]). The Panel pointed out that the revised draft Addendum confirmed that, compared to the guinea pig, the LLNA overpredicted skin sensitization potential.

The Panel was questioned as to whether, based on the accuracy analyses, the revised draft Addendum adequately characterized the usefulness and limitations of the LLNA for testing substances in aqueous solutions.

The Panel agreed that the analysis adequately reviewed the concordance evaluation and pointed out the discrepancies between LLNA and guinea pig results, in so far as permitted by the limited data including the confounding lack of chemical analysis of the medical device derived materials test in the LLNA. The Panel noted that pesticides dominated the database and that no statement summarized usefulness and limitations.

The Panel was also asked if all the relevant data identified in published or unpublished studies which had been conducted with the LLNA to test substances in aqueous solutions of which it was aware had been adequately considered in the revised draft Addendum.

The Panel was unaware of any additional data that could be used to supplement the analysis of use of aqueous test solutions, although a Panel member provided information during the public meeting that guinea pig data for pesticide formulation active ingredients was available, as was data comparing sensitization potential of pesticide formulation active ingredients in the LLNA, GPMT, and Buehler tests (Gehen et al. 2009). Since pesticide formulations tested in aqueous vehicles were included in the analysis, this information could have supplemented the analysis of substances tested in aqueous solutions.

ICCVAM asked the Panel if it considered the inclusion of studies performed with the vehicle 1% Pluronic L92 appropriate for these analyses, given that this vehicle has not been validated for use in the ICCVAM-recommended LLNA protocol (Appendix A of ICCVAM 2009). If so, it was asked whether 1% Pluronic L92 should be added to the preferred LLNA vehicles list (for testing water-soluble substances) in the ICCVAM-recommended LLNA protocol (note that substances tested at a concentration of 100% contain no vehicle).

As noted above, the Panel commented that inclusion of LLNA studies performed with the vehicle 1% Pluronic L92 for pesticide formulations appeared valid and permissible based on the limited information available in the peer-reviewed literature. This peer-reviewed information showed the performance of 1% Pluronic L92 in the LLNA and discussed its selection based on wetting properties, lack of irritant activity, low toxicity, and known actions in enhancing skin penetration (Ryan et al. 2002; Boverhof et al. 2008).

4.7 Comments on the Revised Draft ICCVAM Test Method Recommendations on the Traditional LLNA Applicability Domain

4.7.1 Test Method Usefulness and Limitations

In regard to the usefulness and limitations of the proposed test method, the Panel responded to a question about whether available data supported the revised draft ICCVAM test method recommendations for the LLNA, specifically focusing on testing of pesticide formulations.

The Panel agreed with the analyses and noted the information contained in the revised draft Addendum informed the Panel's discussions and its subsequent differences in recommendations compared with the revised draft ICCVAM recommendations for all classes of test substances.

The Panel agreed that there was support for the conclusions regarding usefulness and limitations of the LLNA for testing of pesticide formulations. The Panel noted that most of the guinea pig data used in the concordance analyses was Buehler test data and that the Buehler test is considered to be less sensitive than the GPMT (Basketter et al. 1993; Frankild et al. 2000). The Panel further stated that the revised draft recommendations specific to pesticides might benefit from information regarding use of an aqueous vehicle, and that the caveat regarding the apparent high rate of false positive findings in the LLNA was appropriate. If there is any primary testing or postmarketing reports of skin sensitization, they should be used for comparison. If the LLNA erred, it tended to overclassify. Whether this is useful or a limitation was largely a matter for the relevant regulatory agency to determine, as it depends on the inferences that must be drawn from the data. The Panel considered that compared to underclassification of skin sensitization potential, overestimation of skin sensitization potential provided overall protection of human health.

The Panel stated that, because materials were tested exclusively in aqueous solutions, these conclusions might be revisited if data derived from pesticides which had been tested in nonaqueous solvents or vehicles became available.

The Panel was asked whether it agreed that the available data support the revised draft ICCVAM test method recommendations for the LLNA with regard to testing dyes in terms of the proposed test method usefulness and limitations.

On the basis of the available information, unless there are unique physicochemical properties associated with these test materials and classes of test materials that might affect their ability to interact with immune processes, the Panel considered dyes as candidates for testing in the LLNA.

The Panel was asked whether it agreed that the available data support the revised draft ICCVAM test method recommendations for the LLNA with regard to testing natural complex substances in terms of the proposed test method usefulness and limitations.

On the basis of the available information, unless there are unique physicochemical properties associated with these test materials and classes of test materials that might affect their ability to interact with immune processes, the Panel considered natural complex substances as candidates for testing in the LLNA.

ICCVAM asked the Panel whether it agreed that the available data support the revised draft ICCVAM test method recommendations proposed test method usefulness and limitations of the LLNA with regard to substances tested in aqueous solutions.

The Panel agreed that the available data support the revised draft ICCVAM test method recommendations for the LLNA with regard to testing substances in aqueous solutions in terms of the proposed test method usefulness and limitations.

On the basis of the available information, unless there are unique physicochemical properties associated with these test materials and classes of test materials that might affect their ability to interact with immune processes, the Panel considered substances tested in aqueous solutions as candidates for testing in the LLNA.

Although the caveat to potential users regarding the apparently high rate of false positive findings was important, the recommendations could be edited to include a recommendation that any future studies with aqueous vehicles use 1% Pluronic L92 to help expand the database. The Panel further stated that the database was so heavily weighted with pesticides that these conclusions should be reevaluated as information about other classes of test material tested in aqueous solution become available. Finally, the Panel commented that,

compared to underclassification of skin sensitization potential, overclassification provided protection from the standpoint of human health. In the future, ICCVAM should consider its acceptability as a vehicle for use in the ICCVAM-recommended LLNA protocol (Appendix A of ICCVAM 2009) when the protocol is next reviewed.

The Panel was asked whether it agreed that the available data support the revised draft ICCVAM test method recommendations for the LLNA in terms of the proposed test method standardized protocol.

The Panel commented that the deliberations of the Panel in 2008 were well represented in the LLNA test method protocol. Updated information on various elements in the revised draft Addendum did not suggest the need for changes to recommendations for the development of a revised standard method. Whenever discretion is permitted, a suitable (representative) positive control should be selected from the same category of materials to be tested (e.g., for testing pesticides, select one representative positive control pesticide).

4.7.2 Future Studies

The Panel was asked whether it agreed that the available data support the revised draft ICCVAM test method recommendations for the LLNA in terms of the proposed future studies.

The Panel responded that:

- Solubility data should ideally be provided in future studies, so that thermodynamic activity can be computed and compared to maximum theoretical percutaneous penetration.
- This information should be considered when comparing the data from LLNA studies in lipophilic delivery systems compared to that in aqueous systems.
- The ICCVAM recommendation for continued accumulation of information in the targeted areas was considered entirely appropriate.

The Panel also suggested that, before additional animal testing is conducted, consideration should be given to the necessity for the substance to be tested for skin sensitization potential. Numerous factors affect whether a substance has the potential to produce skin sensitization in man. For example:

- A substance must penetrate the skin (Potts and Guy 1992) and react with a macromolecule (protein) in the skin to create an immunogenic moiety (Dupuis and Benezra 1982; Barratt and Basketter 1992; Basketter 1992; Smith and Hotchkiss 2001).

- The substance must also invoke the production of cytokines and a T-lymphocyte response (Kimber and Dearman 1996).

[This Page Intentionally Left Blank]

5.0 References

Barratt MD, Basketter DA. 1992. Possible origin of the skin sensitization potential of isoeugenol and related compounds. (I). Preliminary studies of potential reaction mechanisms. Contact Dermatitis 27:98-104.

Basketter DA. 1992. Skin sensitization to cinnamic alcohol: the role of skin metabolism. Acta Derm Venereol 72:264-265.

Basketter DA, Selbie E, Scholes EW, Lees D, Kimber I, Botham PA. 1993. Results with OECD recommended positive control sensitizers in the maximization, Buehler and local lymph node assays. Food Chem Toxicol 31:63-67.

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.

Basketter DA, McFadden J, Evans P, Andersen KE, Jowsey I. 2006. Identification and classification of skin sensitizers: identifying false positives and false negatives. Contact Dermatitis 55:268-273.

Boverhof DR, Wiescinski CM, Botham P, Lees D, Debruyne E, Repetto-Larsay M, et al. 2008. Interlaboratory validation of 1% pluronic L92 surfactant as a suitable, aqueous vehicle for testing pesticide formulations using the murine local lymph node assay. Toxicol Sci 105:79-85.

Dupuis G, Benezra C. 1982. Allergic Contact Dermatitis to Simple Chemicals. New York, Basel: Marcel Dekker.

EPA. 2003. Health Effects Test Guideline OPPTS 870.2600. Skin Sensitization. Research Triangle Park, NC: Environmental Protection Agency.

Frankild S, Volund A, Wahlberg JE, Andersen KE. 2000. Comparison of the sensitivities of the Buehler test and the guinea pig maximization test for predictive testing of contact allergy. Acta Derm Venereol 80:256-262.

Gehen SC, Wiesciniski CM, Billington R, Woolhiser M. 2009. Acceptability of the mouse local lymph node assay for pesticide formulation hazard assessment. Toxicologist 108:316.

Gerberick GF, Cruse LW, Ryan CA, Hulette BC, Chaney JG, Skinner RA, et al. 2002. Use of a B cell marker (B220) to discriminate between allergens and irritants in the local lymph node assay. Toxicol Sci 68:420-428.

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

ICCVAM. 1999. The Murine Local Lymph Node Assay: A Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemicals/Compounds. The Results of an Independent Peer Review Evaluation Coordinated by the Interagency Coordinating Committee on the Validation of Alternative Methods (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 Institute of Environmental Health Sciences.

ICCVAM. 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. Research Triangle Park, NC: National Institute of Environmental Health Sciences.

ICCVAM. 2009. Recommended Performance Standards: Murine Local Lymph Node Assay. NIH Publication Number 09-7357. Research Triangle Park, NC: National Institute of Environmental Health Sciences.

Idehara K, Yamagishi G, Yamashita K, Ito M. 2008. Characterization and evaluation of a modified local lymph node assay using ATP content as a non-radio isotopic endpoint. J Pharmacol Toxicol Methods. 58:1-10.

ISO. 2002. ISO 10993-10: Tests for Irritation and Sensitization. Geneva: International Organization for Standardization.

Jung K, Bae I, Lim K. 2009. Evaluation of the skin sensitization potential of hair-dye, para-phenylenediamine by nonradioactive murine local lymph node assay using bromodeoxyuridine with flow cytometry and immunohistochemistry. Toxicologist 108:81. Abstract 388.

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

Montelius J, Wahlkvist H, Boman A, Fernström P, Gråbergs L, Wahlberg JE. 1994. Experience with the murine local lymph node assay: inability to discriminate between allergens and irritants. Acta Derm Venereol 74:22-27.

OECD. 2002. Guideline for the Testing of Chemicals 429 - Skin Sensitisation: Local Lymph Node Assay. Paris: OECD.

Omori T, Idehara K, Kojima H, Sozu T, Arima K, Goto H, et al. 2008. Interlaboratory validation of the modified murine local lymph node assay based on adenosine triphosphate measurement. J Pharmacol Toxicol Methods 58:11-26.

Potts RO, Guy RH. 1992. Predicting skin permeability. Pharm Res 9:663-669.

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

Ryan CA, Cruse LW, Skinner RA, Dearman RJ, Kimber I, Gerberick GF. 2002. Examination of a vehicle for use with water soluble materials in the murine local lymph node assay. Food Chem Toxicol 40:1719-1725.

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

Schneider K, Akkan Z. 2004. Quantitative relationship between the local lymph node assay and human skin sensitization assays. Regul Toxicol Pharmacol 39:245-55.

Smith CK, Hotchkiss SAM. 2001. Allergic Contact Dermatitis–Chemical and Metabolic Mechanisms. London: Taylor & Francis.

[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 institute, Gif sur Yvette, France. She is currently the scientific coordinator on Alternatives Methods in Life Science Department at L'Oréal Research and Development, Aulnay sous Bois, France. She is the L'Oréal representative to the European Partnership on Alternative to Animal Testing, and serves on two working groups: Identification of Opportunities, Including R&D (working group 2), and Validation and Acceptance (working group 5). She is also the representative in the eye irritation working group to the European Cosmetics Association and in the French Groupement d'Intérêt Scientifique Platform on Alternatives. She has served on the European Centre for the Validation of Alternative Methods (ECVAM) Scientific Advisory Committee (ESAC), representing the European Federation of Pharmaceutical Industries Associations, and was nominated as Organisation for Economic Co-operation and Development expert for eye and skin irritation. As a manager in Investigative Toxicology with Pfizer Global Research and Development, Amboise, France, she implemented the murine local lymph node assay (LLNA) in the laboratory. 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). She is responsible for the human health scientific program and for the investigation and initiation of new research and testing projects for RIFM. She is a member of 10 professional organizations, including the American Academy of Dermatology, American Contact Dermatitis Society, the European Society of Contact Dermatitis, and the Society of Investigative Dermatology. She participated in the World Health Organization International Workshop in Skin Sensitization in Chemical Risk Assessment held in Berlin, Germany, in 2006, and a BfR International Workshop on Contact Dermatitis in October 2008. She is author of over 100 publications and presentations relevant to dermatology and dermatotoxicology.

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

Dr. Flournoy received B.S. and M.S. degrees 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. 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 and 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 (SACATM). 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.

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 Department of Toxicology and Veterinary Services and the Reference Center for Cosmetics 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 the 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 Sixth World Congress on Alternatives and 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 British Industrial Biological Research Association (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 and U.K. Government Departments. His research interests are the use of mathematical, statistical, and bioinformatic 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 U.S. National Institute of Environmental Health Sciences (NIEHS), 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 Scientific Committee of the European Food Safety Authority, the U.K. Government's advisory Committees on Mutagenicity and Carcinogenicity of Chemicals in Food, Consumer Products and the Environment and the Independent Scientific Advisory Committee for the U.K. 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 (USCF), 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 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.

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 U.S. Food and Drug Administration National Center for Toxicological Research. Following graduate training, he served as NIEHS National Research Service Award Postdoctoral 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 (NAS). 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 and Molecular Biology, University of Minnesota Medical School, Duluth. 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 and 2008 ICCVAM Peer Panel that evaluated the 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. She is currently President 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.

Jonathan Richmond, B.Sc. (Hons) Med.Sci., MB ChB, FRCSEd, FRMS
Dr. Richmond received a Bachelor of Science in Medical Science with Honors (BSc [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 former Fellow of the Royal Society of Medicine (FRMS). Other appointments include convener of the U.K. Interdepartmental Group on 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 European Union (E.U.) Committee on Scientific and Technical Progress and past Chairman of the European Commission Technical Expert Working Group on ethical review, and former board member of the U.K. National Centre for the 3Rs. 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 NICEATM-ICCVAM Expert Panel that evaluated Five *In Vitro* Pyrogen Test Methods, and developed performance standards for minor variations on the test method. 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 the 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 has served on NIH and NAS 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, LA. He served on the NICEATM-ICCVAM Expert Panels that evaluated the *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants, LLNA and *In Vitro* Pyrogen Test Methods. He is a former member of SACATM. He is presently working as an animal welfare 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 Assistant Professor at the Center for Integrative Toxicology, Michigan State University. 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 32 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

[This Page Intentionally Left Blank]

Appendix B1

Questions for the Peer Review Panel: Nonradioactive LLNA Protocol – LLNA: DA Test Method

[This Page Intentionally Left Blank]

Revisions to the LLNA: DA Evaluation

In 2008, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) proposed, and the independent scientific peer review panel (Panel) agreed, that more information was needed before a recommendation on the usefulness and limitations of the murine local lymph node assay (LLNA) based on measuring adenosine triphosphate content in the draining auricular lymph nodes (referred to hereafter as the "LLNA: DA") could be made.[8] The Panel indicated that the following information was needed: a detailed test method protocol, quantitative data for the test method, and an evaluation of interlaboratory reproducibility. In response to the Panel's recommendation, the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) obtained additional LLNA: DA data and information, which were used to update a revised draft background review document (BRD). These data and information include:

- A detailed description of the standard operating procedures/protocol for the LLNA: DA test method (see Appendix A of revised draft BRD).
- Individual animal data for the LLNA: DA intralaboratory validation study of 31 substances (Idehara et al. 2008; see Section 6.0 of the revised draft BRD which includes these data in the updated accuracy analyses).
- Data for 14 additional LLNA: DA intralaboratory substances (Idehara unpublished; see Section 6.0 of the revised draft BRD which includes these data in the updated accuracy analyses).
- Individual animal data for the LLNA: DA two-phased interlaboratory validation study of 14 substances (Omori et al. 2008; see Section 6.0 of the revised draft BRD which includes these data in the updated accuracy analyses and Section 7.0 which includes these data in the additional quantitative analyses of test method reproducibility).

Instructions for the Peer Review Panel

The following questions are intended to guide you in your second review of the LLNA: DA.

- You are first asked to review the new information in the revised draft 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 expanded information in the revised draft BRD to determine the extent to which each of the applicable criteria for validation and

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

acceptance of toxicological test methods (ICCVAM 2003)[9] have been appropriately addressed for the proposed use of the LLNA: DA. Adequate validation[10] is a prerequisite for a test method to be considered for use in regulatory decision-making by U.S. Federal agencies. Only aspects of the original review that have changed due to new information need to be addressed.

- Lastly, you are asked to consider the revised draft ICCVAM 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 revised draft BRD.

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

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

The overall questions to consider are:

1. Whether the validation status (usefulness and limitations) of the LLNA: DA has been adequately characterized for its intended purpose, and
2. Whether it is sufficiently accurate and reliable to be used for the identification of sensitizing and nonsensitizing substances in place of the traditional LLNA procedure.

[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: National Institute of Environmental Health Sciences (Available at: http://iccvam.niehs.nih.gov/SuppDocs/SubGuidelines/SD_subg034508.htm).

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

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

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

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

1. Substances Used for the Validation Studies (see Section 3.0 of revised draft BRD)

   i. Thirteen additional substances have been added to the database since the January 2008 draft BRD (i.e., from 33 substances [31 with sufficient LLNA data] to 46 substances [44 with sufficient traditional LLNA data]). This represents a total of 32 traditional LLNA sensitizers and 12 traditional LLNA nonsensitizers.

      a. Do you now consider the LLNA: DA database representative of a sufficient range of chemical classes and physicochemical properties that the test method would be applicable for testing any of the types of chemicals and products that are typically tested for skin sensitization potential (see Section 6.3 of the revised draft BRD for a comparison of the substances tested in the LLNA: DA with the LLNA performance standards reference substances)?

      b. 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?

   ii. What additional reference chemicals or products should be evaluated to fill any data gaps? Please explain your answers.

2. Test Method Accuracy (see Section 6.0 of the revised draft BRD)

   i. When evaluating the test method performance of the LLNA: DA using a single stimulation index (SI) decision criterion as a predictor of skin sensitization potential (see Section 6.2 of revised draft BRD), the decision criterion recommended by the test method developers, SI $\geq$ 3.0, yielded accuracy = 91% (40/44), with sensitivity = 88% (28/32) and specificity = 100% (12/12) (i.e., there were four false negatives and no false positives). The SI $\geq$ 2.0 yielded the same accuracy as SI $\geq$ 3.0 but increased the sensitivity (97% [31/32]) by decreasing the false negative rate to 3% (1/32) although the false positive rate increased to 25% (3/12). These two single decision criteria were further compared to guinea pig tests and human/data experience.

a. Are these comparisons appropriate for assessing the accuracy of the LLNA: DA using a single decision criterion? Please explain your answer.

b. Has the relevance (e.g., sensitivity, specificity, false positive and false negative rates) of the LLNA: DA, using a single decision criterion to distinguish between sensitizers and nonsensitizers, been adequately evaluated and compared to the traditional LLNA results? If not, what other analyses should be performed? Please explain your answer.

ii. The test method performance for the LLNA: DA using statistics (i.e., ANOVA or *t*-test), compared to the traditional LLNA, yielded accuracy = 84% (37/44), with sensitivity = 94 (30/32) and specificity = 58% (7/12) (i.e., there were five false positives and two false negatives). See Section 6.5 of the revised draft BRD.

a. Are these comparisons appropriate for assessing the accuracy of the LLNA: DA using statistics? Does the accuracy analysis provide an adequate comparison from which to decide between using a single SI decision criterion (e.g., SI $\geq$ 2.0) or statistics to distinguish sensitizers from nonsensitizers? Please explain your answer.

b. Has the relevance (e.g., sensitivity, specificity, false positive and false negative rates) of the LLNA: DA, using statistics to distinguish between sensitizers and nonsensitizers, been adequately evaluated and compared to the traditional LLNA results? If not, what other analyses should be performed? Please explain your answer.

iii. Further, evaluation of multiple decision criteria to classify substances as sensitizers or nonsensitizers indicates that at SI $\geq$ 2.5 there were no false positives and at SI $\leq$ 1.7 there were no false negatives in the LLNA: DA compared to traditional LLNA results (see Section 6.7 of the revised draft BRD).

a. Are these comparisons appropriate for assessing the accuracy of the LLNA: DA using multiple decision criteria? Please explain your answer.

b. Has the relevance (e.g., sensitivity, specificity, false positive and false negative rates) of the LLNA: DA, using the SI $\geq$ 2.5 criterion for sensitizers and the SI $\leq$ 1.7 criterion for nonsensitizers, been adequately evaluated and compared to the traditional LLNA results? If not, what other analyses should be performed? Please explain your answer.

iv. Ten of the 44 substances used to evaluate the LLNA: DA yielded SI values between 1.7 and 2.5 (exclusive). Among these substances, 5/10 are sensitizers and 5/10 are nonsensitizers based on traditional LLNA results. Common characteristics of each group (i.e., the five sensitizers and the five nonsensitizers) are described in Section 6.8 of the revised draft BRD.

   a. Is the number of sensitizers (i.e., five) and nonsensitizers (i.e., five) that yield between 1.7 and 2.5 (exclusive) sufficient to identify characteristics (e.g., peptide-binding activity, molecular weight) that may be helpful in determining the skin sensitization hazard classification of such substances? If not, what other substances and data would be helpful?

   b. If the number of substances is sufficient, can you identify characteristics associated with these, or similar substances, that may provide additional information as to whether substances that yield SI values between 1.7 and 2.5 (exclusive) should be classified as sensitizers or nonsensitizers? Please explain your answer.

v. The LLNA: DA test method is not being evaluated using the ICCVAM-recommended performance standards for the LLNA because the method was developed prior to the publication of the performance standards. Furthermore, some of the protocol modifications noted for the LLNA: DA (i.e., pretreatment with sodium lauryl sulfate (SLS), extended dosing schedule [see Section 2.0 of the revised draft BRD) cause the LLNA: DA to be considered functionally and mechanistically different than the traditional LLNA. Although the ICCVAM-recommended LLNA performance standards are not being applied to evaluate the validity of the LLNA: DA, does the fact that the LLNA: DA produced false negative results for two of the 18 required performance substances impact its validation status? The other 16 required performance standards substances tested yielded concordant results with the traditional LLNA (see Section 6.3 of the revised draft BRD). Please explain your answer.

3. Test Method Reliability (Intra- and Interlaboratory reproducibility; see Section 7.0 of the revised draft BRD)

   i. Has the intralaboratory reproducibility of the LLNA: DA been adequately evaluated and compared to the traditional LLNA (see Section 7.1 of the revised draft BRD)? If not, what other analyses should be performed? Are any limitations apparent based on this intralaboratory reproducibility assessment? Please explain your answers.

ii. A two-phased interlaboratory validation study of the LLNA: DA was conducted by testing 14 different coded substances (10 sensitizers and four nonsensitizers based on traditional LLNA results). Two sensitizers and one nonsensitizer were tested in 10 to 17 laboratories; the remaining eight sensitizers and three nonsensitizers were tested in three to four laboratories. The Validation Management Team selected both the vehicles and the concentrations of the substances tested. Are any limitations apparent based on this study design (see Section 7.2 of revised draft BRD)? Do you have any concerns based on the fact that vehicles and concentrations were provided to the participating laboratories? Please explain your answers.

iii. Has the interlaboratory reproducibility of the LLNA: DA, using SI $\geq$ 2.5 for sensitizers and SI $\leq$ 1.7 for nonsensitizers, been adequately evaluated and compared to the traditional LLNA? If not, what other analyses should be performed? Are any limitations apparent based on this interlaboratory reproducibility assessment? Please explain your answers. See Section 7.3 of revised draft BRD.

4. Consideration of all available data and relevant information (see Section 9.0 of the revised draft BRD).

   i. Based on the revised 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 revised draft BRD, but are available for consideration? If yes, please explain how to obtain such data.

III. Questions to the Panel: Revised Draft ICCVAM Test Method Recommendations on the LLNA: DA

1. Test Method Usefulness and Recommendations

   i. Do you agree that the available data and test method performance (accuracy and reliability) support the revised draft ICCVAM test method recommendations for the LLNA: DA in terms of the proposed test method usefulness and limitations? Please explain your answer.

   ii. ICCVAM proposes using two decision criteria to distinguish between sensitizers and nonsensitizers in the LLNA: DA: SI $\geq$ 2.5 to classify substances as sensitizers and SI $\leq$ 1.7 to classify substances as nonsensitizers.

      a. Does the accuracy analysis support the proposed ICCVAM recommendation that the decision criteria for sensitizers be based on an SI $\geq$ 2.5 and the decision criterion for nonsensitizers be based on SI $\leq$ 1.7? Do you consider using these multiple SI values appropriate for skin sensitization classification? If not, which decision criterion, or criteria (e.g., statistics or other SI values), do you consider most appropriate?

      b. Does using multiple decision criteria 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.

      c. Do you agree that a statistical approach could be used, in addition to SI values, to make a skin sensitization classification decision?

   iii. ICCVAM recommends that when an SI between the range of 1.7 to 2.5 (exclusive) is obtained in the LLNA: DA, users should carefully interpret the results in an integrated decision strategy, in conjunction with all other available information (e.g., dose response information, statistical analyses, peptide-binding activity, molecular weight, results from related chemicals, other testing data) to determine if there is adequate information for an accurate sensitization hazard classification, or if additional testing is necessary.

      a. Do you agree with the proposed ICCVAM recommendation that substances with SI values in the "uncertainty range" between 1.7 and 2.5 (exclusive) would require additional information or testing to make a classification decision? Should other analyses be performed? Please explain your answers.

b. What information should be considered to determine how a substance with SI between 1.7 and 2.5 (exclusive) can be classified with respect to skin sensitization, or whether it should be tested using additional skin sensitization test methods? If additional testing is appropriate, what additional tests should be conducted?

c. Under what circumstances could test method results from the LLNA: DA with SI between 1.7 and 2.5 (exclusive) be used as part of an integrated decision strategy (i.e., in conjunction with all other available information such as dose response information, statistical analyses, peptide-binding activity, molecular weight, results from related chemicals, other testing data) to assign a sensitization hazard classification?

d. The interlaboratory reproducibility standard from the ICCVAM-recommended LLNA performance standards is ECt values, from each laboratory, of 0.025-0.1% for 2,4-dinitrochlorobenzene and 5-20% for hexyl cinnamic aldehyde. Are there concerns that 5/10 laboratories that tested 2,4-dinitrochlorobenzene produced EC2.5 values that were outside of the acceptability range? If so, does the fact that only two of the aberrant EC2.5 values were greater than the upper range mitigate that concern? Are lower limits for the range of acceptable ECt values necessary (i.e., lower values indicate a more sensitive test)? Please explain your answer.

2. Test Method Protocol (see Section 2.0 and Appendix A of the revised draft BRD)

   i. Do you agree that the available data support the ICCVAM draft test method recommendations for the LLNA: DA procedure in terms of the proposed test method standardized protocols? If not, then what recommendations would you make? Please explain your answer.

3. Future Studies

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

4. Performance Standards

   i. ICCVAM recognizes the need for separate LLNA: DA performance standards, since the LLNA: DA incorporates procedures that differ from the essential test method components in the performance standards for the

traditional LLNA. Accordingly, ICCVAM has proposed that separate performance standards for the LLNA: DA be developed to evaluate future modifications of this test method. Do you agree that separate performance standards for the LLNA: DA need to be developed to evaluate future modifications of this test method?

ii. If your answer to Question 4i is yes, in order to aid ICCVAM in developing appropriate performance standards, please address the following questions:

   a. Based on the recommended LLNA: DA protocol, there are two primary differences in procedures that result in the LLNA: DA being considered as mechanistically and functionally different from the traditional LLNA: (1) pre-treatment prior to test substance application with a non-irritating concentration of 1% SLS; and (2) an additional test substance application on Day 7. These would necessarily be reflected as changes to the essential test method components in the traditional LLNA performance standards. Are there other changes to the essential test method components in the traditional LLNA performance standards that you think should be reflected in LLNA: DA performance standards?

   b. When tested in the LLNA: DA, four substances among the 18 required reference substances in the traditional LLNA performance standards (i.e., 2-mercaptobenzothiazole, methyl methacrylate, chlorobenzene, and salicylic acid) produced SI values within the range of uncertainty (i.e., $1.7 \leq SI \leq 2.5$). Please comment on the appropriateness of four possible replacement substances (i.e., abietic acid, diethyl phthalate, dimethyl isophthalate, potassium dichromate) that could be substituted to obtain a list of LLNA: DA performance standards reference substances for which LLNA: DA, traditional LLNA, guinea pig test, and human results are concordant (see **Table 1**). Please also comment on the appropriateness of four proposed optional reference substances (see **Table 1**).

   c. With regard to test method performance criteria, are there differences from the traditional LLNA performance standards other than the decision criteria used to define a positive response for accuracy, and the ranges acceptable values required to establish reproducibility, that should be reflected in LLNA: DA performance standards?

Table 1 Proposed Reference Substances for the LLNA: DA: Possible Changes from the Traditional LLNA Performance Standards

No.	Substance	CASRN	Form	Vehicle	N[1]	EC2.5 (%)[2]	LLNA: DA vs. LLNA	LLNA: DA vs. GP	LLNA: DA vs. Human
1	5-Chloro-2-methyl-4-isothiazolin-3-one	26172-55-4	Liq	DMF	1	0.02	+/+	+/+	+/+
2	4-Phenylenediamine	106-50-3	Sol	AOO	1	0.05	+/+	+/+	+/+
3	2,4-Dinitrochlorobenzene	97-00-7	Sol	AOO	11	0.06	+/+	+/+	+/+
4	**Potassium dichromate**	**7778-50-9**	**Sol**	**DMSO**	**5**	**0.18**	**+/+**	**+/+**	**+/+**
5	Isoeugenol	97-54-1	Liq	AOO	4	1.05	+/+	+/+	+/+
6	Cobalt chloride	7646-79-9	Sol	DMSO	7	1.28	+/+	+/+	+/+
7	Phenyl benzoate	93-99-2	Sol	AOO	1	1.44	+/+	+/+	+/+
8	Eugenol	97-53-0	Liq	AOO	1	3.60	+/+	+/+	+/+
9	**Abietic acid**	**514-10-3**	**Sol**	**AOO**	**4**	**7.60**	**+/+**	**+/+**	**+/+**
10	Hexyl cinnamic aldehyde	101-86-0	Liq	AOO	18	8.78	+/+	+/+	+/+
11	Imidazolidinyl urea	39236-46-9	Sol	DMF	1	11.94	+/+	+/+	+/+
12	Cinnamic alcohol	104-54-1	Sol	AOO	1	12.20	+/+	+/+	+/+
13	Citral	5392-40-5	Liq	AOO	1	12.46	+/+	+/+	+/+
14	**Diethyl phthalate**	**84-66-2**	**Liq**	**AOO**	**1**	**NA**	**-/-**	**-/-**	**-/-**
15	**Dimethyl isophthalate**	**1459-93-4**	**Sol**	**AOO**	**4**	**NA**	**-/-**	**-/-**	**-/-**
16	Isopropanol	67-63-0	Liq	AOO	11	NA	-[3]/-	-[3]/-	-[3]/+
17	Lactic acid	50-21-5	Liq	DMSO	5	NA	-/-	-/-	-/*
18	Methyl salicylate	119-36-8	Liq	AOO	4	NA	-[4]/-	-[4]/-	-[4]/-
~~X~~	~~2-Mercaptobenzothiazole~~	~~149-30-4~~	~~Sol~~	~~DMF~~	~~1~~	~~NA~~	~~-[5]/+~~	~~-[5]/+~~	~~-[5]/+~~
~~X~~	~~Methyl methacrylate~~	~~80-62-6~~	~~Liq~~	~~AOO~~	~~1~~	~~NA~~	~~-[6]/+~~	~~-[6]/+~~	~~-[6]/+~~
~~X~~	~~Chlorobenzene~~	~~108-90-7~~	~~Liq~~	~~AOO~~	~~1~~	~~NA~~	~~-[6]/-~~	~~-[6]/*~~	~~-[6]/*~~
~~X~~	~~Salicylic acid~~	~~69-72-7~~	~~Sol~~	~~AOO~~	~~1~~	~~NA~~	~~-[6]/-~~	~~-[6]/-~~	~~-[6]/-~~

continued

Optional Substances to Demonstrate Improved Performance Relative to the LLNA: DA

No.	Substance	CASRN	Form	Vehicle	N[1]	EC2.5 (%)[2]	LLNA: DA vs. LLNA	LLNA: DA vs. GP	LLNA: DA vs. Human
19	Sodium lauryl sulfate	151-21-3	Sol	DMF	1	2.91	+/+	+/-	+/-
20	Ethylene glycol dimethacrylate	97-90-5	Liq	MEK	1	28.52	+/+	+/-	+/+
21	**2-Mercaptobenzothiazole**	**149-30-4**	**Sol**	**DMF**	**1**	**NA**	**-[5]/+**	**-[5]/+**	**-[5]/+**
22	Nickel chloride	7718-54-9	Sol	DMSO	1	NA	-/-	-/+	-/+
~~X~~	~~Xylene~~	~~1330-20-7~~	~~Liq~~	~~NT~~	~~NT~~	~~NT~~	~~NT/+~~	~~NT/**~~	~~NT/-~~

Note: Substances with a ~~strikethrough~~ are included among the traditional LLNA performance standards reference substances but are not proposed as LLNA: DA performance standards reference substances; bolded substances indicate proposed LLNA: DA performance standards reference substances not included in the traditional LLNA performance standards.
Abbreviations: AOO = acetone: olive oil (4:1); CASRN = Chemical Abstracts Service Registry Number; DMF = *N,N*-dimethylformamide; DMSO = dimethyl sulfoxide; EC2.5 = estimated concentration needed to produce a stimulation index of 2.5; GP = guinea pig test result; Liq = liquid; LLNA = murine local lymph node assay; LLNA: DA = murine local lymph node assay: modified by Daicel Chemical Industries, Ltd. based on adenosine triphosphate content; MEK = methyl ethyl ketone; NA = not applicable since stimulation index < 2.5; No. = number; Sol = solid; Veh = vehicle.
[1] Number of LLNA: DA studies from which the data were obtained.
[2] Represent the mean EC2.5 value, when more than one EC2.5 value was available.
[3] Classified as a nonsensitizer based on 10 of 11 LLNA: DA studies with a maximum SI ≤ 1.7.
[4] Classified as a nonsensitizer based on three of four LLNA: DA studies with a maximum SI ≤ 1.7.
[5] Not classified as a sensitizer (maximum SI not SI ≥ 2.5). Also, not a definite nonsensitizer (maximum SI not SI ≤ 1.7). Therefore, 2-mercaptobenzothiazole is an optional reference substance for which a modified LLNA: DA test method can demonstrate improved performance.
[6] Not classified as a sensitizer (maximum SI not SI ≥ 2.5). Also, not a definite nonsensitizer (maximum SI not SI ≤ 1.7).
* = Presumed to be a nonsensitizer in humans based on the fact that no clinical patch test results were located, it is not included as a patch test kit allergen, and no case reports of human sensitization were located.
** = GP data not available.

[This Page Intentionally Left Blank]

Appendix B2

Questions for the Peer Review Panel: Nonradioactive LLNA Protocol – LLNA: BrdU-FC Test Method

[This Page Intentionally Left Blank]

Revisions to the LLNA: BrdU-FC Evaluation:

In 2008, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) proposed, and the independent scientific peer review panel (Panel) agreed, that more information was needed before a recommendation on the usefulness and limitations of the murine local lymph node assay (LLNA) with bromodeoxyuridine (BrdU) measurement by flow cytometry (referred to hereafter as the "LLNA: BrdU-FC") could be made.[11] The Panel indicated that the following information was needed: individual animal data, and evaluations of both intralaboratory and interlaboratory reproducibility. In response to the Panel's recommendation, MB Research Labs (the developing laboratory) submitted additional LLNA: BrdU-FC data to the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), which were used to update a revised draft LLNA: BrdU-FC Background Review Document (BRD). These data include:

- LLNA: BrdU-FC data from multiple studies with 2-mercaptobenzothiazole using different vehicles. These data were submitted in a response to a request for an explanation for the discordant results for 2-mercaptobenzothiazole. The data indicate a vehicle dependent response in the LLNA: BrdU-FC for identifying a positive result with 2-mercaptobenzothiazole.
- Data from studies for sodium lauryl sulfate using an enhanced LLNA: BrdU-FC protocol (referred to hereafter as the "eLLNA: BrdU-FC"), which includes an assessment of immunophenotypic markers to distinguish sensitizers from irritants, reportedly to reduce the incidence of false positive results.
- LLNA: BrdU-FC data on two substances (hexyl cinnamic aldehyde [HCA] and 2,4-dinitrochlorobenzene [DNCB]) not previously provided to NICEATM. These data were used to demonstrate intralaboratory reproducibility.

Instructions for the Peer Review Panel

The following questions are intended to guide you in your second review of the LLNA: BrdU-FC.

- You are first asked to review the new information in the revised draft BRD for completeness, and to identify any errors or omissions of existing relevant data or information that should be included.

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

- You are then asked to evaluate the expanded information in the revised draft BRD to determine the extent to which each of the applicable criteria for validation and acceptance of toxicological test methods (ICCVAM 2003)[12] have been appropriately addressed for the proposed use of the LLNA: BrdU-FC. Adequate validation[13] is a prerequisite for a test method to be considered for use in regulatory decision-making by U.S. Federal agencies. Only aspects of the original review that have changed due to new or revised information need to be addressed.
- Lastly, you are asked to consider the revised draft ICCVAM 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 revised draft LLNA: BrdU-FC BRD.

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

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

The overall questions to consider are:

1. Whether the validation status (usefulness and limitations) of the LLNA: BrdU-FC has been adequately characterized for its intended purpose, and
2. Whether the LLNA: BrdU-FC is sufficiently accurate and reliable to be used for the identification of sensitizing substances and nonsensitizing substances in place of the traditional LLNA procedure.

---

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

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

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

1. In response to the 2008 Panel recommendations for more data, additional raw data and test vehicle information have been provided for 13 of 45 substances evaluated, including limited data on 10 substances from single studies described in the January 2008 draft BRD. Are there any errors in the revised draft LLNA: BrdU-FC BRD that should be corrected, or omissions of existing relevant data or information that should be included?

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

1. Test Method Accuracy (see Section 6.0 of the revised draft BRD)
   i. Test method accuracy has been recalculated based on updated reference data obtained since the 2008 Panel meeting. In view of this updated evaluation, do you consider that the relevance (i.e., accuracy, sensitivity, specificity, false positive and false negative rates) of the LLNA: BrdU-FC has been adequately evaluated and compared to the traditional LLNA (see Section 6.1 of the revised draft BRD)? Please explain your answer.
   ii. Are there any comments or recommendations regarding the revised analysis or the revised reference data?
2. Test Method Reliability (see Section 7.0 of the revised draft BRD)
   i. In view of the additional studies submitted for DNCB and HCA do you consider intralaboratory reproducibility to have been adequately evaluated and compared to the traditional LLNA (see Section 7.1 of revised draft BRD)? Please explain your answer.
   ii. For the assessment of interlaboratory reproducibility, the ICCVAM Recommended Performance Standards for the LLNA indicate that "Interlaboratory reproducibility should be evaluated with at least two sensitizing chemicals with well-characterized activity in the traditional LLNA. In this regard, ECt values for 2,4-dinitrochlorobenzene (DNCB) and HCA should be derived independently from a single study conducted in at least three separate laboratories. Acceptable reproducibility will be indicated by each laboratory obtaining ECt values for HCA and DNCB that are within 0.5x to 2.0x (5% to 20% and 0.025% to 0.1%, respectively) the mean EC3 concentration (10% and 0.05%, respectively) specified for these substances . . . "

As indicated in the Panel report of 2008,
"The Panel agreed with the draft ICCVAM recommendations for evaluating test method reliability. These recommendations included obtaining ECt 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 2,4-dinitrochlorobenzene (DNCB) (interlaboratory, n = 1 experiment in three laboratories)."

a. Do you agree that the LLNA: BrdU-FC is mechanistically and functionally similar to the traditional LLNA and, as such, the studies recommended in the ICCVAM-recommended performance standards for demonstrating intra- and interlaboratory reproducibility would be sufficient for the LLNA: BrdU-FC? If not, what other testing should be performed in order to demonstrate that it is sufficiently reproducible?

iii. Before a naïve laboratory performs LLNA: BrdU-FC studies for regulatory data submissions, what testing might be recommended to demonstrate proficiency in conducting this method?

a. What substances from the ICCVAM-recommended performance standards might be recommended for testing (e.g., one strong sensitizer, one moderate sensitizer, and one weak sensitizer)?

b. Would results from the positive control from each study be adequate for a laboratory to assess its intralaboratory reproducibility?

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

1. Test Method Usefulness and Recommendations

   i. Similar to the 2008 draft test method recommendations, ICCVAM is deferring a formal recommendation on the validity of the LLNA: BrdU-FC until all data supporting the analysis have been submitted to NICEATM for an independent audit, and until there is a demonstration of transferability in a second testing laboratory. Do you have any further comments or suggestions regarding this position, and/or do you agree with this recommendation? Please explain your answer.

   ii. Does the revised draft BRD support the revised draft ICCVAM test method recommendations on the usefulness and limitations of the LLNA: BrdU-FC with regard to its use in the developing laboratory (i.e., MB Research Labs)?

2. Test Method Protocol

   i. Do you agree that the available data support the revised draft ICCVAM test method 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. Do you agree that laboratories should maintain a historical database of positive control data in order to provide a measure of variability of the positive control response over time?

   iii. Do you agree that there could be cause for concern when a negative test substance result is accompanied by a concurrent positive control stimulation index (SI) value that is significantly lower than the mean historical SI (even when this SI value is still positive)?

3. Future Studies

   i. Do you agree that the available data support the revised draft ICCVAM test method recommendations for the LLNA: BrdU-FC in terms of the proposed future studies? What other recommendations would you make? Please explain your answer.

[This Page Intentionally Left Blank]

Appendix B3

Questions for the Peer Review Panel: Nonradioactive LLNA Protocol – LLNA: BrdU-ELISA Test Method

[This Page Intentionally Left Blank]

Revisions to the LLNA: BrdU-ELISA Evaluation

In 2008, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) proposed, and the independent scientific peer review panel (Panel) agreed, that more information was needed before a recommendation on the usefulness and limitations of the murine local lymph node assay (LLNA) with bromodeoxyuridine (BrdU) measurement by enzyme-linked immunosorbent assay (ELISA; referred to hereafter as the "LLNA: BrdU-ELISA") for assessing the allergic contact dermatitis potential of chemicals and other substances could be made.[14] The Panel indicated that the following information was needed: a detailed test method protocol, quantitative data for the test method, and an evaluation of interlaboratory reproducibility. In response to the Panel's recommendation, the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) obtained additional LLNA: BrdU data, which were used to update a revised draft background review document (BRD). These data include:

- Six substances not previously provided to NICEATM. (Note: The number of substances evaluated effectively increased by seven with the location of reference data for one substance for which LLNA: BrdU-ELISA data had been previously submitted.) These data were used in a reanalysis of test method accuracy, which is detailed in Section 6.0 of the revised draft BRD.
- Individual animal data for the LLNA: BrdU-ELISA interlaboratory validation study of 10 substances. These data were used in additional quantitative analyses of test method reproducibility, which are detailed in Section 7.0 of the revised draft BRD.

Instructions for the Peer Review Panel

The following questions are intended to guide you in your second review of the LLNA: BrdU-ELISA.

- You are first asked to review the information in the revised draft LLNA: BrdU-ELISA 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 expanded information in the revised draft BRD to determine the extent to which each of the applicable criteria for validation and acceptance of toxicological test methods (ICCVAM 2003)[15] have been appropriately

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

[15] ICCVAM. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods. NIH Publication No. 03-4508. Research Triangle Park, NC: National Institute of

addressed for the proposed use of the LLNA: BrdU-ELISA. Adequate validation[16] is a prerequisite for a test method to be considered for use in regulatory decision-making by U.S. Federal agencies. Only aspects of the original review that have changed due to new information need to be addressed.

- Lastly, you are asked to consider the revised draft ICCVAM 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 revised draft LLNA: BrdU-ELISA BRD.

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

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

The overall questions to consider are:

1. Whether the validation status (usefulness and limitations) of the LLNA: BrdU-ELISA has been adequately characterized for its intended purpose, and
2. Whether it is sufficiently accurate and reliable to be used for the identification of sensitizing and nonsensitizing substances in place of the traditional LLNA procedure.

Environmental Health Sciences (Available: http://iccvam.niehs.nih.gov/SuppDocs/SubGuidelines/SD_subg034508.htm).

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

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

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

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

1. Substances Used for the Validation Studies (see Section 3.0 of revised draft BRD)

    i. Seven additional substances have been added to the database since the January 2008 draft BRD for a total of 31 substances with traditional LLNA data for which 22 were classified by the traditional LLNA as skin sensitizers and nine were classified as nonsensitizers.

        a. Do you now consider the LLNA: BrdU-ELISA database representative of a sufficient range of chemical classes and physicochemical properties that the test method would be applicable for testing any of the types of chemicals and products that are typically tested for skin sensitization potential (see Section 6.3 of the revised draft BRD for a comparison of the substances tested in the LLNA: BrdU-ELISA with the LLNA performance standards reference substances)?

        b. 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.

2. Test Method Accuracy (see Section 6.0 of revised draft BRD)

    i. When evaluating the test method performance of the LLNA: BrdU-ELISA using a single SI decision criterion as a predictor of skin sensitization potential (see Section 6.2 of revised draft BRD), the decision criterion recommended by the test method developers, SI $\geq$ 2.0, yielded accuracy = 84% (26/31), sensitivity = 77% (17/22) and specificity = 100% (9/9) (i.e., there were five false negatives and no false positives). This single decision criterion was further compared to guinea pig tests and human/data experience, as was SI $\geq$ 1.5, which yield higher sensitivity = 91% (20/22) and lower specificity = 67% (6/9) (see Section 6.5 of the revised draft BRD).

a. Are these comparisons appropriate for assessing the accuracy of the LLNA: BrdU-ELISA using a single decision criterion? Please explain your answer.

b. Has the relevance (e.g., sensitivity, specificity, false positive and false negative rates) of the LLNA: BrdU-ELISA, using a single decision criterion of SI ≥ 2.0 to distinguish between sensitizers and nonsensitizers, been adequately evaluated and compared to the traditional LLNA results? If not, what other analyses should be performed? Please explain your answer.

ii. The test method performance for the LLNA: BrdU-ELISA using statistics (i.e., ANOVA or *t*-test), compared to the traditional LLNA, yielded accuracy = 81% (25/31), sensitivity = 86% (19/22) and specificity = 67% (6/9) (i.e., there were three false negatives and three false positives). See Section 6.5 of the revised draft BRD.

a. Are these comparisons appropriate for assessing the accuracy of the LLNA: BrdU-ELISA using statistics? Does the accuracy analysis provide an adequate comparison from which to decide between using a single SI decision criterion (i.e., SI ≥ 2.0) or statistics to distinguish sensitizers from nonsensitizers? Please explain your answer.

b. Has the relevance (e.g., sensitivity, specificity, false positive and false negative rates) of the LLNA: BrdU-ELISA using statistics to distinguish between sensitizers and nonsensitizers, been adequately evaluated and compared to the traditional LLNA results? If not, what other analyses should be performed? Please explain your answer.

iii. Further evaluation of multiple decision criteria to classify substances as sensitizers or nonsensitizers indicates that at SI ≥ 2.0 there were no false positives, and at SI < 1.3 there were no false negatives in the LLNA: BrdU-ELISA compared to traditional LLNA results (see Section 6.7 of the revised draft BRD).

a. Are these comparisons appropriate for assessing the accuracy of the LLNA: BrdU-ELISA using multiple decision criteria? Please explain your answer.

b. Has the relevance (e.g., sensitivity, specificity, false positive and false negative rates) of the LLNA: BrdU-ELISA, using the SI ≥ 2.0 criterion for sensitizers and the SI < 1.3 criterion for nonsensitizers, been adequately

evaluated and compared to the traditional LLNA results? If not, what other analyses should be performed? Please explain your answer.

iv. Eleven of the 31 substances used to evaluate the LLNA: BrdU-ELISA yielded SI values greater than or equal to 1.3 and less than 2.0. Among these substances, 6/11 are sensitizers and 5/11 are nonsensitizers based on traditional LLNA results. Common characteristics of each group (i.e., the six sensitizers and the five nonsensitizers) are described in Section 6.8 of the revised draft BRD.

a. Is the number of sensitizers (i.e., six) and nonsensitizers (i.e., five) that yield SI values greater than or equal to 1.3 and less than 2.0 sufficient to identify characteristics (e.g. peptide-binding activity, molecular weight) that may be helpful in determining the skin sensitization hazard classification of such substance? If not, how many substances would be sufficient?

b. If the number of substances is sufficient, can you identify characteristics associated with these, or similar substances, that may provide additional information as to whether substances that yield SI values greater than or equal to 1.3 and less than 2.0 should be classified as sensitizers or nonsensitizers? Please explain your answer.

v. The LLNA: BrdU-ELISA test method is not being evaluated using the ICCVAM-recommended performance standards for the LLNA because the method was developed prior to the publication of the performance standards. Although the ICCVAM-recommended LLNA performance standards are not being applied to evaluate the validity of the LLNA: BrdU-ELISA, does the fact that the LLNA: BrdU-ELISA produced one false negative result for one of the 10 required performance standards substances tested impact its validation status? The other nine required performance substances tested yielded concordant results with the traditional LLNA (see Section 6.3 of the revised draft BRD). Please explain your answer.

3. Test Method Reliability (Intra- and Interlaboratory Reproducibility; see Section 7.0 of the revised draft BRD)

i. Has the reproducibility of the LLNA: BrdU-ELISA using $SI \geq 2.0$ for sensitizers and $SI \leq 1.3$ for nonsensitizers been adequately evaluated and compared to the traditional LLNA (see Section 7.3 of the revised draft LLNA: BrdU-ELISA BRD)? If not, what other analyses should

be performed? Are any limitations apparent based on this reproducibility assessment? Please explain your answers.

ii. The Japanese Society for Alternatives to Animal Experiments (JSAAE) conducted a multilaboratory validation study of the LLNA: BrdU-ELISA by testing 10 coded substances (seven sensitizers and three nonsensitizers). Two sensitizers and one nonsensitizer were tested in all seven laboratories; the remaining substances were tested in three laboratories. The Validation Management Team selected both the vehicles and the concentrations of the substances tested. Are any limitations apparent based on this study design? Do you have any concerns based on the fact that vehicles and concentrations were provided to the participating laboratories? Please explain your answers.

4. Consideration of all available data and relevant information (see Section 9.0 of the revised draft BRD).

i. Based on the revised 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 revised draft BRD, but are available for consideration? If yes, please explain how to obtain such data.

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

1. Test Method Usefulness and Recommendations

i. Do you agree that the available data and test method performance (accuracy and reliability) support the revised draft ICCVAM test method recommendations for the LLNA: BrdU-ELISA in terms of the proposed test method usefulness and limitations? Please explain your answer.

ii. ICCVAM proposes using two decision criteria to distinguish between sensitizers and nonsensitizers in the LLNA: BrdU-ELISA: SI $\geq$ 2.0 to classify substances as sensitizers and SI < 1.3 is used to classify substances as nonsensitizers.

a. Does the accuracy analysis support the proposed ICCVAM test method recommendation that the decision criterion for sensitizers be based on SI $\geq$ 2.0 and the decision criterion for nonsensitizers be based on SI < 1.3? Do you consider using these multiple SI values appropriate for skin

sensitization classification? If not, which decision criterion, or criteria (e.g., statistics or other SI values), do you consider most appropriate?

b. Does using multiple decision criteria 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.

c. Do you agree that a statistical approach could be used, in addition to SI values, to make a skin sensitization classification decision?

iii. ICCVAM recommends that when an SI value greater than or equal to 1.3 and less than 2.0 is obtained in the LLNA: BrdU-ELISA, users should carefully interpret the results in an integrated decision strategy, in conjunction with all other available information (e.g., dose response information, statistical analyses, peptide-binding activity, molecular weight, results from related chemicals, other testing data) to determine if there is adequate information for an accurate sensitization hazard classification, or if additional testing is necessary.

a. Do you agree with the proposed ICCVAM recommendation that substances with SI values in the "uncertainty range" of greater than or equal to 1.3 and less than 2.0 would require additional information or testing to make a classification decision? Should other analyses be performed? Please explain your answers.

b. What information should be considered to determine how a substance with an SI greater than or equal to 1.3 and less than 2.0 can be classified with respect to skin sensitization, or whether it should be tested using additional skin sensitization test methods? If additional testing is appropriate, what additional tests should be conducted?

c. Under what circumstances could test method results from the LLNA: BrdU-ELISA with an SI greater than or equal to 1.3 and less than 2.0 be used as part of an integrated decision strategy (i.e., in conjunction with all other available information such as dose response information, statistical analyses, peptide-binding activity, molecular weight, results from related chemicals, other testing data) to assign a sensitization hazard classification?

2. Test Method Protocol

    i. Do you agree that the available data support the revised draft ICCVAM test method 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.

3. Future Studies

    i. Do you agree that the available data support the revised draft ICCVAM test method 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.

4. Performance Standards

    i. Do you agree that the LLNA: BrdU ELISA adheres to the essential test method components in the ICCVAM-recommended LLNA performance standards, and therefore the performance standards can be used to evaluate future nonradioactive versions similar to the LLNA: BrdU ELISA that also adhere to these essential test method components?

    ii. The interlaboratory reproducibility standard from the ICCVAM-recommended LLNA performance standards is ECt values, from each laboratory, of 0.025-0.1% for 2,4-dinitrochlorobenzene and 5-20% for hexyl cinnamic aldehyde. Are there concerns that 2/7 laboratories that tested 2,4-dinitrochlorobenzene and 2/6 laboratories that tested hexyl cinnamic aldehyde in the JSAAE study produced EC2 values that were outside of the acceptability range? If so, does the fact that only one of the aberrant EC2 values (for hexyl cinnamic aldehyde) was greater than the upper range mitigate that concern? Are lower limits for the range of acceptable ECt values necessary (i.e., lower values indicate a more sensitive test)? Please explain your answer.

Appendix B4

Questions for the Peer Review Panel: LLNA for Testing Pesticide Formulations and Other Products, Aqueous Solutions, and Metals

[This Page Intentionally Left Blank]

Introduction

The National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) convened an independent scientific peer review panel (Panel) meeting on March 4-6, 2008. The Panel peer reviewed the January 2008 draft Addendum for the applicability domain of the murine local lymph node assay (LLNA) and commented on the extent that it supported the draft ICCVAM test method recommendations on the usefulness and limitations of the LLNA.[17] The draft ICCVAM test method recommendations stated 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 test method 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 test method recommendations. The Panel also noted that 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. In response to this concern, NICEATM specified three types of multi-component substance classes, which were included in the database examined in the revised evaluation; pesticide formulations, dyes, and multi-component fragrance ingredients.

Revisions to the LLNA: Applicability Domain

After the 2008 Panel meeting, NICEATM received new data on pesticide formulations and other products, and substances tested in aqueous solutions. The updated database includes:

- Data for 104 pesticide formulations for which LLNA data exists. This database is very limited relative to the set of registered pesticide active ingredients and pesticide classes. These data are described in Section 5.0 and detailed in Appendices B1 and B2 of the revised draft Addendum.
- Data for six dye formulations for which LLNA data and guinea pig reference data exists. These data are described in Section 5.0 and detailed in Appendices B4 and B5 of the revised draft Addendum.

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

- Data for 12 multi-component fragrance ingredients for which there are LLNA and human data. These data are described in Section 5.0 and detailed in Appendices B6 and B7 of the revised draft Addendum.
- Data for 139 substances tested in the LLNA in aqueous solutions. These data are described in Section 5.0 and detailed in Appendices D1, D2, and D3 of the revised draft Addendum.

The database for metal compounds has remained unchanged since the 2008 Panel meeting, so no new evaluations were conducted for the revised draft Addendum. It should also be noted that no metal formulations were included in the analyses.

Instructions for the Peer Review Panel

The following questions are intended to guide you in your review of the LLNA applicability domain.

- You are first asked to review the new information in the revised 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 expanded information in this revised draft Addendum to determine the extent to which each of the applicable criteria for validation and acceptance of toxicological test methods (ICCVAM 2003)[18] have been appropriately addressed for the proposed use of the LLNA for testing pesticide formulations and other products, and substances in aqueous solutions. Adequate validation[19] is a prerequisite for a test method to be considered for use in regulatory decision-making by U.S. Federal agencies. Only aspects of the original review that have changed due to new information need to be addressed.
- Lastly, you are asked to consider the revised draft ICCVAM 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 revised draft Addendum.

---

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

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

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

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

The overall questions to consider are:

1. Whether the validation status of the LLNA for testing pesticide formulations and other products, and substances in aqueous solutions has been adequately characterized, and
2. Whether it is 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 of the Revised Draft ICCVAM Addendum: LLNA Applicability Domain for Errors and Omissions

1. In the revised 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: Revised Draft ICCVAM Addendum: LLNA Applicability Domain

A. General Questions

1. During the 2008 Panel review, the Panel stated 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 that are being examined. This concern was addressed in the revised draft Addendum by dividing the substances considered into pesticide formulations, dyes, fragrance ingredients, and substances tested in aqueous solutions (this group included pesticide formulations), and analyzing the data for each group separately. Do you consider these groups appropriate, given the composition of the updated database? If not, into which groups would you suggest the substances in the revised database be divided into for analysis? Please provide a rationale to explain your answer.
2. Are there other types or classes of substances you would recommend be evaluated in the future? If yes, which types or classes of substances would you recommend be considered? Please provide a rationale to explain your answer.
3. Do you consider the inclusion of studies performed with BALB/c mice appropriate for these analyses, since this strain has not been validated for use in the ICCVAM-recommended LLNA protocol? It should be noted that, in an evaluation of the effect of reducing the sample size for the LLNA from five animals to four conducted by NICEATM, it was stated that the pattern of LLNA responses seen in BALB/c mice was similar to that in CBA mice (see Appendix B of ICCVAM 2009).[20]
4. BALB/c mice are commonly used for LLNA testing. Do the available data support the use of any BALB/c substrains for LLNA testing? If not, should a high priority be placed on the collection of additional comparative BALB/c and CBA data for a more comprehensive assessment?

[20] ICCVAM. 2009. Recommended Performance Standards: Murine Local Lymph Node Assay. NIH Publication Number 09-7357. Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available at http://iccvam.niehs.nih.gov/methods/immunotox/llna_PerfStds.htm.

a. Are you aware of any BALB/c substrain differences such that certain substrains should be avoided when performing the LLNA?

5. As the use of the LLNA continues to increase, it is likely that LLNA will be used to test new product types that have not been previously tested in the LLNA, and for which there are no comparative human or guinea pig reference data. With this in mind, do you consider it necessary to validate the LLNA each time new types or classes of substances and/or formulations are proposed for testing in the LLNA?

B. Pesticide Formulations

1. In the database of pesticide formulations that have guinea pig reference data, nine active ingredients (out of approximately 450) are represented (one insecticide, six herbicides and two fungicides) and approximately six to seven classes of pesticide types (out of approximately 40 classes of compounds). No antimicrobials or biopesticides or microbials are represented. Do you consider the database of substances evaluated representative of a sufficient range of pesticide formulations that are typically tested for skin sensitization potential? Please explain your answer.

2. Has the relevance (e.g., accuracy/concordance, sensitivity, specificity, false positive and false negative rates) of the LLNA for testing pesticide formulations been adequately evaluated (refer also to Section 5.0 of the revised draft Addendum)? If not, what other analyses should be performed? Please explain your answer.

3. In light of the fact that only a relatively small subset of registered formulated pesticide active ingredients and classes were available for analysis, does the revised draft Addendum adequately characterize the usefulness and limitations of the LLNA for testing pesticide formulations based on the accuracy analyses? If not, what additions or changes should be made to the current usefulness and limitations? Please explain your answer.

4. Based on the revised draft Addendum, have all the relevant data identified in published or unpublished studies, of which you are aware, conducted using the LLNA for testing pesticide formulations been adequately considered? If not, what other studies should be considered?

5. Do you consider the inclusion of studies performed with the vehicle 1% Pluronic L92 appropriate for these analyses, since this vehicle has not been validated for use in the ICCVAM-recommended LLNA protocol? If yes, should 1% Pluronic L92 be added to the preferred vehicles list for LLNA studies, for water-soluble substances, in the ICCVAM-recommended LLNA protocol? (Note that substances tested at a concentration of 100% contain no vehicle at that concentration.)

C. Dyes

1. Do you consider the database of substances evaluated representative of a sufficient range of dyes that are typically tested for skin sensitization potential? Please explain your answer.
2. Has the relevance (e.g., accuracy/concordance, sensitivity, specificity, false positive and false negative rates) of the LLNA for testing dyes been adequately evaluated (refer also to Section 5.0 of the revised draft Addendum)? If not, what other analyses should be performed? Please explain your answer.
3. Does the revised draft Addendum adequately characterize the usefulness and limitations of the LLNA for testing dyes based on the accuracy analyses? If not, what additions or changes should be made to the current usefulness and limitations? Please explain your answer.
4. Based on the revised draft Addendum, have all the relevant data identified in published or unpublished studies, of which you are aware, conducted using the LLNA for testing dyes been adequately considered? If not, what other studies should be considered?

D. Fragrance Ingredients

1. Do you consider the database of 12 substances evaluated representative of a sufficient range of multi-component fragrance ingredients that are typically tested for skin sensitization potential? Please explain your answer.
2. Has the relevance (e.g., accuracy/concordance, sensitivity, specificity, false positive and false negative rates) of the LLNA for testing multi-component fragrance ingredients been adequately evaluated (refer also to Section 5.0 of the revised draft Addendum)? If not, what other analyses should be performed? Please explain your answer.
3. Does the revised draft Addendum adequately characterize the usefulness and limitations of the LLNA for testing multi-component fragrance ingredients based on the accuracy analyses? If not, what additions or changes should be made to the current usefulness and limitations? Please explain your answer.
4. Based on the revised draft Addendum, have all the relevant data identified in published or unpublished studies, of which you are aware, conducted using the LLNA for testing multi-component fragrance ingredients been adequately considered? If not, what other studies should be considered?

E. Substances Tested in Aqueous Solutions

1. Do you consider the database of substances evaluated representative of a sufficient range of substances tested in aqueous solutions (in this revised draft Addendum, aqueous solutions are characterized as containing 20% water or more) that are typically tested for skin sensitization potential? Note that many of the pesticide formulations discussed in Section IIB were also included in this analysis, so the database has the same limitations. Please explain your answer.
2. Has the relevance (e.g., accuracy/concordance, sensitivity, specificity, false positive and false negative rates) of the LLNA for testing substances in aqueous solutions been adequately evaluated (refer also to Section 5.0 of the revised draft Addendum)? If not, what other analyses should be performed? Please explain your answer.
3. Does the revised draft Addendum adequately characterize the usefulness and limitations of the LLNA for testing 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.
4. Based on the revised draft Addendum, have all the relevant data identified in published or unpublished studies, of which you are aware, conducted using the LLNA for testing substances in aqueous solutions been adequately considered? If not, what other studies should be considered?
6. Do you consider the inclusion of studies performed with the vehicle 1% Pluronic L92 appropriate for these analyses, since this vehicle has not been validated for use in the ICCVAM-recommended traditional LLNA protocol? If yes, should 1% Pluronic L92 be added to the preferred vehicles list for LLNA studies, for water-soluble substances, in the ICCVAM-recommended LLNA protocol? (Note that substances tested at a concentration of 100% contain no vehicle at that concentration.)

III. Questions to the Panel: Draft ICCVAM Test Method Recommendations on the LLNA Applicability Domain

1. Do you agree that the available data support the revised draft ICCVAM test method recommendations for the LLNA with regard to testing pesticide formulations 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 revised draft ICCVAM test method recommendations for the LLNA with regard to testing dyes in terms of the proposed test method usefulness and limitations? Please explain your answer. If not, what recommendations would you make?
3. Do you agree that the available data support the revised draft ICCVAM test method recommendations for the LLNA with regard to testing multi-component fragrance ingredients in terms of the proposed test method usefulness and limitations? Please explain your answer. If not, what recommendations would you make?
4. Do you agree that the available data support the revised draft ICCVAM test method recommendations for the LLNA with regard to testing substances tested in aqueous solutions in terms of the proposed test method usefulness and limitations? Please explain your answer. If not, what recommendations would you make?
5. Do you agree that the available data support the revised draft ICCVAM test method 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?
6. Do you agree that the available data support the revised draft ICCVAM test method recommendations for the LLNA in terms of the proposed future studies? Please explain your answer. If not, then what recommendations would you make?