2.0 LLNA Performance Standards

2.1 Background

The LLNA has undergone validation studies that have demonstrated its usefulness and limitations for distinguishing between sensitizers and non-sensitizers (ICCVAM 1999). However, the 1999 evaluation determined that, while the LLNA could be used in most testing situations, certain substances might not be suitable for use with the LLNA. These include:

- Metal compounds: may produce inaccurate results and limited data are available
- High molecular weight compounds: not readily absorbed into the skin
- Strong dermal irritants: may produce false positive results
- Materials that do not adhere to the ear for an acceptable time during the experiment
- Mixtures: limited data available

ICCVAM recently obtained and is currently evaluating available LLNA data on mixtures. These performance standards may be updated to incorporate any new information on the usefulness and limitations of the LLNA for assessing the ACD potential of mixtures.

The following section briefly describes the principles of the LLNA test method, followed by the recommended performance standards that would be used to evaluate test methods that are functionally and mechanistically similar to the traditional LLNA. The performance standards consist of (1) essential test method components, (2) reference substances, and (3) the comparable accuracy and reliability that should be achieved or exceeded.

2.2 LLNA Essential Test Method Components and Other Validation Considerations

Certain principles are important in delineating the essential test method components that determine whether a modified test is functionally and mechanistically similar to the traditional LLNA. In the LLNA, the induction phase of contact hypersensitivity is characterized by lymphocyte proliferation and hyperplasia in the lymph nodes draining the site of topical exposure (Sikorski et al. 1996). Because test substances are applied topically to the ear, the lymphocytes in the draining auricular lymph nodes are collected for evaluation. In the traditional LLNA, the amount of incorporated radioactivity is indicative of the number of proliferating cells in the draining auricular lymph nodes. Potential skin sensitizers are identified by calculating the ratio of radioactivity incorporated into the DNA of cells of the auricular lymph nodes after topical application of a potential chemical sensitizer to that obtained after topical application of the test vehicle.

2.2.1 Essential Test Method Components

The essential test method components for the validation of modifications to the traditional LLNA applicable to these performance standards, using the 18 required reference substances, are summarized as follows and are provided in detail in **Appendix C**.

1. The test substance must be applied topically to both ears of the mice.
2. Lymphocyte proliferation must be measured in the lymph nodes draining the site of test substance application.

3. Lymphocyte proliferation must be measured during the induction phase of skin sensitization.
4. For test substances, the highest dose selected must be the maximum soluble concentration that does not induce systemic toxicity and/or excessive local irritation. For positive control substances, the highest dose selected should exceed the known EC3 values (i.e., the estimated concentrations needed to produce a stimulation index [SI] of 3) of the reference substances without producing systemic toxicity and/or excessive local irritation.
5. A vehicle control must be included in each study and, where appropriate, a positive control should be used.
6. A minimum of four animals per dose group is required.
7. Either individual or pooled animal data may be collected.

In order for a modified LLNA test method protocol to be considered functionally and mechanistically similar to the traditional LLNA, the above characteristics are essential to ensure that the same biological effect is being measured accurately. If any of the criteria are not met, then these performance standards are not applicable to validation of the modified test method. For example, these performance standards would not be applicable to the popliteal lymph node assay (Pieters 2000).

These essential test method components have been internationally harmonized for the validation of modifications to the traditional LLNA. Test method users should be aware that certain national regulatory authorities might have requirements that differ from these essential test method components for the prospective use of a modified LLNA method in support of regulatory submissions. For example, U.S. regulators require the following:

1. As the high dose: the maximum soluble concentration that does not produce systemic toxicity and/or excessive local irritation
2. Collection of individual animal data
3. A concurrent positive control included in each LLNA study

2.2.2 Other Validation Considerations

Additional points to consider during the validation of modified LLNA test methods applicable to these performance standards, using the 18 required reference substances, are summarized as follows and are provided in detail in **Appendix C**.

1. Consideration should be given to running concurrently a mix of negative, weakly, and strongly positive substances from the reference substance list so that the strongly positive substance can act as a positive control for the weaker skin sensitizer.
2. Group housing is recommended; otherwise animal selection, preparation, housing, and feeding should be in accordance with OECD TG 429 in compliance with other relevant regulatory requirements (e.g., animal care and use).
3. Appropriate quality assurance systems (i.e., in accordance with Good Laboratory Practice guidelines, e.g., OECD 1998; EPA 2006a, 2006b; FDA 2006) are required.

4. The study should be conducted according to international validation principles (OECD Guidance Document 34 [OECD 2005]) and in compliance with other relevant regulatory requirements (e.g., animal care and use).

2.2.3 LLNA Test Method Protocol Modifications

These performance standards also apply to LLNA test method protocols that include modifications that do not impact on its functional and mechanistic similarity to the traditional LLNA test method protocol (**Appendix A)** provided that (1) the modified test method protocol incorporates the essential test method components described in detail in **Appendix C**, (2) such modifications are detailed and scientifically rationalized and justified, and (3) the performance of the modified test method is equal to or better than that determined for the traditional LLNA (see **Section 2.4**). Rationale for such changes should include a description of the decision criteria used to distinguish between sensitizers and non-sensitizers, and the basis for the decision criteria. In the traditional LLNA, an SI of 3 or greater is used to identify a skin-sensitizing agent (ICCVAM 1999). However, a threshold SI may be other than 3 for modified LLNA test method protocols that use a different methodology for measuring lymph node cell proliferation. In such cases, the dose of a test substance at the revised threshold limit would be other than an EC3 and would therefore be defined as an ECt (i.e., the estimated concentration needed to produce an SI with a threshold other than 3).

A proposed minimum of 18 substances are selected as reference substances (i.e., required) with four optional reference substances for the LLNA performance standards. If the modified LLNA test method protocol, like the traditional LLNA (ICCVAM 1999), still uses a decision criterion of SI $\geq$ 3, the 18 required substances could then be used to determine its validation status. If a different decision criterion (i.e., SI not $\geq$ 3) is used, additional testing will be required, the extent (i.e., number and types of substances) of which will be determined on a case-by-case basis, depending on the magnitude of the change in the decision criterion.

Test method developers are encouraged to consult directly with ICCVAM prior to conducting a validation study on modified LLNA methods in accordance with these performance standards. Following completion of a validation study using these performance standards, developers are also encouraged to submit results of studies to ICCVAM for an evaluation of the validation status. In accordance with the ICCVAM Authorization Act, upon completion of its evaluation, ICCVAM will forward recommendations on the validity of the test method to ICCVAM agencies, including adequacy of the test method with regards to these performance standards. Also in accordance with the ICCVAM Authorization Act, the regulatory agencies will determine the acceptability of the test method based on their specific regulatory needs and requirements. Before submitting it to a regulatory agency, test method developers should complete a validation review of the data using the modified test method protocol. Doing so will reduce the possibility of the regulatory agency deeming the data unacceptable or unpersuasive.

Although the SI decision criterion is the one most often used to distinguish between sensitizers and non-sensitizers, a statistical analysis based on individual animal data and/or an evaluation of the dose-response relationship may also be conducted in order to provide a more complete evaluation of the test substance.

2.2.3.1 Calculation of ECt

The reliability assessment of a modified LLNA test method protocol requires calculation of an ECt. Acceptable reproducibility will be demonstrated by each laboratory obtaining ECt values that are generally within 0.5x to 2.0x the mean EC3 concentration specified for the substance tested. The ICCVAM LLNA test method protocol (ICCVAM 1999) does not include guidance on the calculation of an ECt, which is therefore described below.

The method for determining the ECt is a simple linear interpolation of the points in the dose-response curve that lie immediately above and below the classification threshold (e.g., SI = 3 for the traditional LLNA). Consider an example where the decision threshold is an SI of 3:

> If the data points lying immediately above and below the SI value of 3 have the coordinates (a, b) and (c, d) respectively, then the EC3 value may be calculated using the equation: $EC3 = c + [(3 - d)/(b - d)](a - c)$ (Basketter et al. 1999c).

When there are no points below the defined threshold (e.g., SI = 3), a more complex log-linear extrapolation may be applied as described in Ryan et al. (2007) in which the two lowest test concentrations from the dose-response curve are used.

2.2.4 Data and Reporting

The test report should include information outlined below.

1. Test substances, control substances, and vehicles
   - Name of test substance and identification data (e.g., Chemical Abstracts Service Registry Number)
   - Purity and composition of the substance or mixture
   - Physicochemical properties (e.g., physical state, water solubility) relevant to the conduct of the study
   - Treatment of the test/control substances prior to testing, if applicable (e.g., vortexing, sonication, warming; resuspension solvent)
   - Name of vehicle and identification data (e.g., purity, composition, volume used)
   - Justification for choice of vehicle
2. Test animals
   - Mouse strain used[9]
   - Number, age, and sex of animal used
   - Microbiological status of the animal, when information is available
   - Source of animal, housing conditions, diet, etc.
3. Description of the test method and protocol used to measure lymphocyte proliferation and justification for its use

[9] Female CBA/Ca or CBA/J mice are currently recommended. The use of male CBA mice, or female or male mice of other strains would only be accepted if it can be adequately demonstrated that these animals perform in the LLNA as well as female CBA mice.

4. Test method conditions
   - Details on test substance preparation and application
   - Justification for dose selections, including basis for the highest dose tested (see **Appendix A**). The reason for variation away from traditional dose-selection process, if any, should be discussed.
5. Criteria for an acceptable test
   - Positive control data
   - Negative/vehicle control data
   - Laboratory-specific historical ranges of positive and negative control data. A robust historical dataset should include at least 10 independent tests conducted within a reasonable period of time (i.e., less than one year) with a minimum of four animals each per negative and positive control groups.
   - Exclusion criteria should be defined and the impact of any excluded data should be described.
6. Results
   - Weights of each animal at the start of the test and the time of lymph node collection
   - During the collection of individual animals, tabulation of data from the individual animals showing the mean and individual values for each dose (including vehicle and, where applicable, positive control) group
   - Lymphocyte proliferation, which should be expressed in the units specified by the method (e.g., disintegrations per minute for methods using radioactive reagents, absorbance at a specified wavelength for methods using colorimetric reagents). Results should be provided for all test-substance dose levels and concurrent controls.
   - Calculated results (e.g., as measured or quantified by the SI and the associated ECt value, if applicable[10]) should be provided for all test substances and concurrent controls.
   - Statistical analysis and/or evaluation of the dose-response relationship, where appropriate
7. Description of animal observations
   - Time course of onset and severity of clinical signs of systemic toxicity and dermal irritation should be described (e.g., location of observed dermal irritation).
8. Discussion of the results
   - If consideration is given to other properties of the test substance (e.g., structural relationship to known skin sensitizers), in addition to the calculated

[10] An ECt would only be calculated where an SI greater than or equal to the defined threshold was generated.

results for classification of substances as skin sensitizers, such information should be provided.

9. Conclusions
10. If GLP-compliant studies are performed, then additional reporting requirements in the relevant guidelines (e.g., OECD 1998; EPA 2006a, 2006b; FDA 2006) should be followed.
    - A quality assurance statement for GLP-compliant studies should indicate all inspections made during the study and the dates any results were reported to the Study Director. This statement should also confirm that the final report reflects the raw data.

2.3 Minimum List of Reference Substances for Methods Assessing Lymphocyte Proliferation

2.3.1 Criteria for Selection of Reference Substances

Reference substances are used to assess the accuracy and reliability of a proposed mechanistically and functionally similar test method and are a representative subset of those used to demonstrate the reliability and the accuracy of the validated test method (i.e., traditional LLNA). This set of reference substances should, to the extent possible:

- Represent the range of responses that the validated test method is capable of measuring or predicting
- Have well-defined chemical structures
- Have high-quality data available from the traditional test method (i.e., guinea pig tests), which is compared to the data generated by the validated test method (i.e., traditional LLNA), as well as data from the species of interest (e.g., humans), where possible
- Have produced consistent results in the validated test method
- Be readily available from commercial sources
- Not involve excessive hazard or prohibitive disposal costs

2.3.2 Characteristics of Selected Reference Substances

The validity of the traditional LLNA was supported with test data for 211 substances. After careful consideration of the above criteria, 18 substances were selected as proposed minimum reference substances for the LLNA performance standards. An additional four "optional" substances (i.e., these substances were either false positive or false negative in the LLNA when compared to either human or guinea pig results) are also included to provide the opportunity for demonstrating equivalent or superior performance to the traditional LLNA.

The proposed substances are listed in **Appendix F,** and a detailed rationale for selection of the substances in this list is included in **Appendix E**. The selected substances have the following characteristics:

- Twenty-one of the 22 substances have data from testing in the GPMT or BT.

- Twenty of the 22 substances have human data (e.g., Human Maximization Test results, Human Repeat Insult Patch Test results, and/or clinical case studies/reports) or are used as a patch test kit allergen.
- All of the substances are readily available from commercial sources.
- The substances represent the full dynamic range of responses that can be assessed in the current approved LLNA from non-sensitizers to strong sensitizers.
- The selected substances include 10 solids and 12 liquids.
- The molecular weights of the substances range from 60.095 g/mol to 388.294 g/mol.
- The octanol: water partition coefficient values (Wang et al. 2000) of the substances range from -8.3 to 4.8 (from water-soluble to insoluble, respectively).
- The vehicles used for all of the substances are known. The vehicles used were acetone: olive oil (14 substances), dimethyl formamide (4 substances), dimethyl sulfoxide (3 substances), and methyl ethyl ketone (1 substance).
- Peptide reactivity information, which is available for 10 substances, ranges from minimal to high.
- The EC3 values of the positive substances range from 0.009% to 95.8%, based on results from the traditional LLNA.
- The selected substances have a wide range of SI values, from 3.1 to 43.9 for substances identified as skin sensitizers by the traditional LLNA, and 0.9 to 2.9 for substances identified as non-sensitizers by the traditional LLNA.

Studies using the proposed references substances should be evaluated in the vehicle with which they are listed in **Appendix F**.

In situations where a listed substance may not be available, other substances of the same class (e.g., correctly identified sensitizer, false positive) for which there are high quality *in vivo* reference data (as outlined in **Section 2.3.1**) may be used.

2.4 Accuracy and Reliability Performance Values

The final elements of performance standards are the accuracy and reliability values (i.e., test method performance) that should be met or exceeded by the proposed test method when evaluated with the reference substances. The following sections indicate the accuracy and reliability characteristics based on the performance of the traditional LLNA (ICCVAM 1999) for the indicated reference substances; the rationale for the selection of the performance statistics is described in detail in **Appendix E**.

2.4.1 Accuracy

Accuracy is defined as the closeness of agreement between a test method result and an accepted reference value (ICCVAM 2003). For these performance standards, the proposed test method should have accuracy characteristics that are equivalent to or exceed the performance of the traditional LLNA method when evaluated using the minimum list of recommended reference substances (**Appendix F**). Therefore, for the 18 substances with concordant traditional LLNA and guinea pig data (referred to as "required substances"), the

proposed test method should result in the correct classification based on a "yes/no" decision. However, there is a possibility that the modified test method might not correctly classify all the required substances. If, for example, one of the weak sensitizers were to be misclassified, a rationale for the discordance and appropriate additional data (e.g., test results that provide the correct classification for other substances that have similar physical, chemical, and sensitizing properties as the reference substance that was misclassified) could be considered to demonstrate equivalent performance. Therefore, an evaluation of the validation status of the modified LLNA would be on a case-by-case basis. This provision is included since the classification of three out of the five sensitizers among the required reference substances with an EC3 > 10% (i.e., suggesting that they are "weak" sensitizers) is based on only one LLNA study for each of the three substances. Therefore, the likelihood of obtaining a negative result if any of these three substances were retested in the traditional LLNA is not known.

2.4.2 Reliability

Test method reliability (intralaboratory repeatability and intra- and interlaboratory reproducibility) is the degree to which a test method can be performed reproducibly within and among laboratories over time (ICCVAM 2003). *Repeatability* refers to the closeness of agreement between test results obtained within a single laboratory when the procedure is performed on the same substance under identical conditions within a given time period. Intralaboratory repeatability for the traditional LLNA method was not assessed, although some indication of the inherent biological variability can be obtained by comparing the results for individual test animals administered the same dose of a test substance within the same study.

Intralaboratory reproducibility refers to the determination of the extent to which qualified personnel within the same laboratory can replicate results using a specific test protocol at different times. *Interlaboratory reproducibility* refers to the determination of the extent to which different laboratories can replicate results using the same protocol and test substances and indicates the extent to which a test method can be transferred successfully among laboratories. Intra- and interlaboratory reproducibility for the traditional LLNA are summarized in **Appendix E**.

2.4.2.1 Intralaboratory Repeatability

No standard is proposed.

2.4.2.2 Intralaboratory Reproducibility

Intralaboratory reproducibility can be assessed by calculating the variability resulting from testing hexyl cinnamic aldehyde (HCA). ECt values should be derived on four separate occasions with at least one week between tests. Acceptable reproducibility will be indicated by a laboratory obtaining, in each test, ECt values for HCA that are within 0.5x to 2.0x (5% to 20%) the mean EC3 (10%) specified for HCA in **Appendix F**. Because the target EC3 is provided, as few as two dose groups can be used (instead of at least three dose groups, as would be required when testing an unknown substance) since calculation of an EC3 would use only doses that bracket the target EC3 value (i.e., one dose above and one dose below).

2.4.2.3 Interlaboratory Reproducibility

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 in **Appendix F**. As mentioned for intralaboratory reproducibility, as few as two dose groups can be used for this evaluation.