



The Procter & Gamble Company

Central Product Safety
Miami Valley Laboratories
P.O. Box 538707
Cincinnati, Ohio 45253-8707
www.pg.com

February 22, 2008

William S. Stokes, D.V.M., DACLAM
RADM, U.S. Public Health Service
Director, National Toxicology Program Interagency Center for the
Evaluation of Alternative Toxicological Methods
Executive Director, Interagency Coordinating Committee on
the Validation of Alternative Toxicological Methods
National Institute of Environmental Health Sciences, NIH, DHHS
P.O. Box 12233
Research Triangle Park, NC 27709

Dear Dr. Stokes

Thank you for the opportunity to review and comment on the documents prepared by ICCVAM and NICEATM related to a number of the modifications/proposed uses for the traditional LLNA that will be considered by an independent international expert panel in early March.

The teams have done a great job summarizing the available data on the LLNA and for the most part we are in agreement with the conclusions and recommendations outlined in the documents. What makes the LLNA such a valuable tool for skin sensitization hazard identification and risk assessment is that the strengths and limitations of the assay are recognized so well. I am not sure there is another toxicological test that is more understood and evaluated than the LLNA. I am certain that most experts in the field of skin allergy would agree that the older guinea pig skin sensitization test methods are considerably less understood, specifically related to their lack of evaluation through a formal validation process. Our hope is that this peer review of the LLNA will lead to a better appreciation of the LLNA and more important help researchers develop non-animal test methods for evaluating potential skin sensitizing chemicals by using the robust and quantitative nature of the LLNA as a foundation to compare new alternative methods.

For your review and consideration our LLNA experts (Cindy Ryan, Pierre Aeby, Petra Kern and myself) have prepared comments on the LLNA documents posted on the website. I hope you will find them useful and please let us know if you need any additional information.

Sincerely,

G. Frank Gerberick, Ph.D.
Research Fellow Victor Mills Society

DRAFT ICCVAM Recommendations: LLNA Potency

Comparison of LLNAEC3 values to human data:

An evaluation of the ability of the LLNA to predict the relative sensitization potency of chemicals in humans necessitates the use of human sensitization data for comparative purposes. In order for such a comparison to provide meaningful information, one must be aware of and understand the limitations in each of the datasets. The human data used in the comparison are derived from either HRIPT or HMT studies in which single test concentrations, expressed as $\mu\text{g}/\text{cm}^2$, were used for the induction phase of the test protocol. Therefore, a test concentration could be defined as the NOEL, when in reality it may just be the highest concentration tested to date which did not induce sensitization and there is a probability that higher levels would also fail to induce. This certainly could be the case if a LOEL for the particular chemical has not been identified. Indeed, it is difficult to compare LLNA EC3 concentrations against a human NOEL or an arbitrary value of the LOEL/10 (which is intended to represent an estimation of a probable induction threshold value). On one side, the LLNA data were generated using a test protocol designed to produce quantitative values with dose response information which permit the calculation of the LLNA EC3 and on the other side, the human data were generated by a variety of different human repeated insult patch test and human maximization test protocols which, by design are more qualitative in nature, and unless a series of studies were conducted, provide limited if any information on an induction dose response.

It is concerning that in the evaluation of the LLNA to predict skin sensitization potency in humans key values for the comparison are “pragmatically determined”, as is indicated in lines 335-337 of the background review document “Next, the optimal EC3 value that maximized obtaining the correct skin sensitization calls for strong and weak sensitizers (using one or the other proposed decision criterion) was pragmatically determined.” Similar wording is used in lines 801-804. The method or rationale for this “pragmatic determination” are not clearly evident in the document. A sound statistical approach should have been used instead and would have provided a more scientifically robust comparison.

Comparison of LLNA EC3 values to guinea pig data:

To assess the ability of the LLNA to predict skin sensitization potency in Guinea Pigs is not relevant to the purpose of this review. Guinea pig tests such as the Buehler (BT) and Guinea Pig Maximization tests (GPMT) were designed for the purpose of hazard identification and are poorly suited for potency estimations. While the ECETOC Technical Report No. 87, Contact Sensitisation: Classification According to Potency proposes methods to categorize allergenic potency based on BT and GPMT data, it demands that the study was conducted in full accord with OECD TG 406 and advises judicious interpretation of the data as does a similar European Union commission expert review. While the BT and GPMT have served the toxicology community well for many years as predictive skin sensitization hazard methods, it is important to recognize that, unlike the LLNA, neither of these tests has been formally validated by a recognized organization nor has the inter-laboratory variability been adequately investigated.

In several sections of the background review document, for examples Lines 321-324 and lines 714-717, it is indicated that for each substance with comparative LLNA and guinea pig data, potency was evaluated by comparing the LLNA EC3 concentration against the percentage of responding guinea pigs in the BT or GPMT and the associated induction concentration. Comparing LLNA EC3 concentration against the percentage of responding guinea pigs is not appropriate in our opinion and resulting data are of very different natures; the LLNA measures events associated with the induction of skin sensitization and provides objective, quantitative dose response information whereas data derived from the guinea pig tests are based on a subjective evaluation of skin responses occurring at the elicitation phase of sensitization and provides no dose response information on the induction phase.

It appears that the authors understand the difficulty of comparing LLNA EC3 values with potency classifications based on guinea pig data. In line 395 of the background review document it states that "...for substances that had more than one EC3 or guinea pig response, the geometric mean EC3 value and the weight of evidence GP classification category was used. Although the data generated by the GPMT and the BT is categorical, using the weight of evidence categorization provided some measure of a mean response across multiple studies." Considering the admitted difficulties encountered in dealing with multiple sets of guinea pig-derived data, the authors should be consistent and not make any conclusion based on such comparison.

Proposed classification categories for sensitization:

While cut-off values for potency classification are proposed based on either Buehler test and GPMT responses (Table 1-1) we would caution the use of such data in the absence of any other supporting data due to the nature of the test design. In addition, the proposed scheme uses the intradermal induction dose of the GPMT along with the % responders as the basis for classification. We believe that the topical induction concentration should be considered as it is the more relevant route of exposure and the concentration used for intradermal injection is often limited by the addition of Freund's Complete Adjuvant.

The proposed classification (as well as the one proposed by ECETOC TR No. 87) considers only data from guinea pig tests which are defined as 'positive' by the accepted TG 406 definition of a sensitizing chemical (i.e. induces 30% or 15% positive responses in the GPMT or BT respectively). It is possible that a weakly sensitizing chemical tested in a guinea pig test could elicit positive responses in 20% or 25% of the test animals in a GPMT or 10% in the BT, and would be considered as a non-sensitizer and thus would not be classified according to the proposed scheme while a chemical with any LLNA EC3 value would be assigned to one of the 2 proposed categories. Data obtained through the LLNA allows for a continuous spectrum of EC3 values and thus provides a rank ordering of relative potencies which offer more opportunities for categorization beyond two categories. And on the other side, Human and GP tests which are designed to provide yes/no answers have various threshold values creatively proposed in order to force results in the same two categories.

In the proposed two level classification scheme for sensitization potency (Table 1-1), the criteria for classification for category 1 are given as "A high frequency of occurrence...." OR "A probability of occurrence of a high sensitization rate in humans..." and for category 2 are given

as “A low or moderate frequency” OR “A probability of occurrence of a low to moderate sensitization rate in humans...”. The frequency of sensitization or the sensitization rate within an exposed population concerns the **prevalence** of allergic contact sensitization to a particular chemical, which is entirely different from the inherent **potency** of the chemical. Therefore the use of such criteria to classify potency is not appropriate. The likelihood of a chemical inducing skin sensitization within an exposed population (i.e. the probable sensitization rate) depends on two key elements: the intrinsic allergenic potency of the chemical AND the conditions and extent of the allergen exposure (e.g. frequency, duration, exposure conditions, etc.). Clinically, the nature, extent and duration of exposure are commonly the predominant determinants of prevalence. The relative potency of a chemical concerns the amount of chemical required to induce sensitization. In general, the more potent the allergen, the lower the dose per unit area required to induce sensitization. Prevalence data are derived from diagnostic patch testing of patients with suspect allergic contact dermatitis, often presenting with clinical disease, in dermatology clinics. The diagnostic patch test itself is designed to detect the weakest degrees of allergy by using occluded exposure conditions for 48 hours and highest allergen concentrations possible to elicit a reaction. For example, the standard patch test concentration for nickel sulfate is 2.5%. Applied in a diagnostic patch test using an 8 mm Finn chamber delivers a dose per unit area of 750 $\mu\text{g}/\text{cm}^2$, well above the identified human induction threshold of 154 $\mu\text{g}/\text{cm}^2$ (see Table 2 of Appendix A of the LLNA potency background review documents). Many times the nature of the exposure conditions leading to the induction of allergy for these patients is not clearly defined. At best the published results of thousands of such diagnostic patch tests can be used to evaluate trends in patch test reactions.

One example often used to illustrate the difference between potency and prevalence is nickel. It is a very common contact allergen with a relatively high sensitization rate in the US and Europe. However, experimental evidence indicates that nickel is a relatively weak contact allergen, with LLNA EC3 of 140 $\mu\text{g}/\text{cm}^2$ and a human induction threshold of 154 $\mu\text{g}/\text{cm}^2$ for nickel sulfate. The high prevalence is due to the wide distribution, frequent exposure and the nature of exposure, often through ‘compromised’ skin such as body piercing.

Conversely, the preservative methylchloroisothiazolinone/methylisothiazolinone (MCI/MI) is a well known contact allergen considered to be of strong to extreme potency with LLNA EC3 of 2.25 $\mu\text{g}/\text{cm}^2$ and a human induction NOEL of 1.25 $\mu\text{g}/\text{cm}^2$. In Europe, the prevalence rate of allergy to MCI/MI is stable at 1-3% of patch-tested patients. Considering the number of MCI/MI-containing cosmetics and toiletries that are on the market, the opportunities for exposure and the allergenic potency of the preservative one would expect a much higher incidence rate. The prevalence rate for this potent allergen is kept low because of regulatory guidelines/limits on the level of MCI/MI permissible in certain products, thus limiting the dose per unit area of the exposure. Thus, the clinical prevalence of the strong allergen MCI/MI is low whereas for nickel, a known weak allergen, the prevalence is considerably higher which is opposite of what would be expected if only looking at potency and not considering exposure.

The proposed two level classification scheme for sensitization potency (Table 1-1) does not accurately reflect the range of allergenic potencies that have been demonstrated by both animal and human data. LLNA EC3 values and human induction thresholds clearly span several orders of magnitude as shown by the data in Table 2 of Appendix A of the LLNA potency background

review documents. Human threshold values range from 1.25 $\mu\text{g}/\text{cm}^2$ for MCI/MI, to 250 $\mu\text{g}/\text{cm}^2$ for isoeugenol, to 2755 $\mu\text{g}/\text{cm}^2$ for farnesol, to 20,690 $\mu\text{g}/\text{cm}^2$ for benzyl benzoate. Clinical experience with allergic contact dermatitis would also indicate that discrete classes of sensitizing potency exist (Contact Derm, 2000, 42:344-348).

DRAFT ICCVAM Recommendations: LLNA Applicability Domain

Draft Recommendations – Use of the LLNA to Test Mixtures:

A dataset of 18 mixtures was evaluated, 15 of which had guinea pig data and none had human data. As a result, the LLNA data were compared to the guinea pig data. Since the database is severely limited due to the lack of human data, there is no proof that the guinea pig data would be representative of the human response. Thus, using the guinea pig data as the standard to which the LLNA data should be compared is not appropriate.

In addition, the usefulness of these data is limited further by the fact that information on the ingredients is known for only one of the 15 mixtures and 11 were tested in the LLNA in an aqueous vehicle, the performance of which is also being assessed in this same report.

High quality LLNA mixture data is published in Lalko et al. (2006), cited in section 7.6 of Addendum No. 1 to the ICCVAM report. This publication concerns the evaluation of essential oils and includes analytical data on the composition of the oils as well as LLNA data on the identified major constituents. These data should have been included in the evaluation and not just mentioned as other available scientific reports.

Since the database is severely limited due to the lack of human data, we agree with the recommendation that an assessment of the suitability of the LLNA for testing mixtures should not be conducted until a sufficient quantity of quality data become available. A similar logic of course also applies to guinea pig test methods.

Draft Recommendations – Use of the LLNA to Test Metal Compounds:

The reference dataset contains human data for 17 metal compounds representing 13 different metals. Since the allergenic potential in humans of most all of the known metals has been established, one questions the importance of or need for an assessment of the LLNA's ability to detect metal allergens. However, we agree with the recommendation that the LLNA is useful for the testing of metal compounds. Whether or not the LLNA is useful for testing nickel compounds is of limited importance as nickel is a well known human contact allergen.

In addition, since only 1 of the 14 metal compounds with LLNA and human data was tested in both in an aqueous vehicle, the comparison does not add much value to the assessment, especially in light of the fact that the performance of the LLNA using aqueous vehicles is being assessed in this same report.

Draft Recommendations – Use of the LLNA to Test Substances in Aqueous Solutions:

A dataset of 21 substances tested in aqueous solutions was evaluated, 4 of which had had human data. Since the database is severely limited due to the lack of human data, we agree with the recommendation that an assessment of the suitability of the LLNA for testing substance in

aqueous solutions should not be conducted until a sufficient quantity of quality data become available.

DRAFT ICCVAM Recommendations: LLNA Limit Dose Procedure

Draft Recommendations – Limit Dose Procedure:

We agree with the recommendation that the LLNA limit dose procedure is appropriate for hazard identification purposes.

We must point out that a 10% concentration threshold for defining non-sensitizing chemicals is not, as suggested in line 44 of the recommendation, proposed by Kimber et al. (2006) as the absolute cut-off. In the discussion section of that same paper, Kimber et al. indicate that for the purposes of that article the 10% threshold was used and that that figure “should not be regarded as inviolable.” They go on to say that a case could be made for using, for instance, either 15% or 20%. In the 2005 Gerberick et al. paper (Compilation of historical local lymph node data for evaluation of skin sensitization alternative methods. *Dermatitis*, 16(4):157-202), compounds that did not induce a positive response at any concentration tested, with the highest concentration being at least 20% or greater, were categorized as non-sensitizing.

In addition, the 10% threshold concentration at which all which all negative results would be considered valid did not originate in the cited Kimber et al 2006 publication. The original reference is Cockshott et al., 2006, *Human and Experimental Toxicology*, 25:387-394 in which the performance of the LLNA was evaluated in a regulatory context. In that paper, a negative result obtained with the highest concentration tested at 10% would be considered a valid result if the positive control, a mild to moderate sensitizer, gave a positive response. In other words, a chemical which is negative at a top concentration of 10% does not represent a significant human sensitization hazard. This is similar to the definition of a non-sensitizing chemical in the Guinea Pig Maximization Test (GPMT) or Buehler test as one which induces less than 30% or 15% positive responses respectively. Therefore, if a chemical elicits positive responses in 20% or 25% of the test animals in a GPMT, it would be considered as a non-sensitizer from a regulatory perspective.

Comments on DRAFT ICCVAM Recommendations: LLNA Non-Radioactive Methods

DRAFT ICCVAM Recommendations: LLNA BrdU ELISA Procedure

We agree with the recommendation that more information and data are needed on this method in order to conduct a meaningful assessment of the BrdU ELISA procedure’s performance relative to the traditional LLNA. It is especially important to have information regarding the inter-laboratory performance of this assay.

We do have one suggestion for consideration. Table 6-2 of the Background Review Documents shows a comparison of standard LLNA EC3 values and 0.5x-2x range for the performance standard chemicals and EC3 values calculated from the BrdU ELISA LLNA. Since an alternative SI cutoff for the BrdU ELISA LLNA was identified that provides greater accuracy

than an SI = 3 cutoff i.e., SI = 1.3, a comparison of BrdU ELISA EC1.3 values to standard LLNA EC3 values would be helpful.

DRAFT ICCVAM Recommendations: LLNA BrdU FC Procedure

We agree with the recommendation that more information and data are needed on this method in order to conduct a meaningful assessment of the BrdU-FC procedure's performance relative to the traditional LLNA. While the total number of chemicals tested (45) is sufficient, it is especially important to have information regarding the inter-laboratory performance of this assay. The background review document speculates that the transferability of the LLNA: BrdU-FC and the eLLNA: BrdU-FC would be similar to the traditional LLNA. However, we do not think that will be the case. Flow cytometry is not a trivial technique. It is certainly more error prone than scintillation counting and often the quality of the results is very dependant on trained personnel and precise procedures.

Only 13 of the 18 minimum performance standard reference chemicals have been tested in the LLNA BrdU-FC procedure. This may not be sufficient to assess the test performance according to the ICCVAM Performance Standards for the LLNA. In addition, rather than focusing on the number of chemicals for which the BrdU-FC procedure produced equivocal results or did not obtain 100% concordance with the ICCVAN LLNA performance standard reference chemicals, we believe that it would be of greater value to investigate potential causes for those results. Such information would provide some understanding of the limitations of the methods.

Since the purpose of this evaluation of the LLNA BrdU-FC procedure is to assess its ability to be a non-radioactive alternative to the traditional LLNA, is a comparison with Guinea Pig data justified?

The provided test protocol indicates that at least 6 mice be employed for an irritation prescreen and a possible 12 more be used for the optional quantitative irritation test. Therefore, this method has the potential to use more mice than the traditional LLNA. This requirement for greater animal usage must be taken into consideration when evaluating the BrdU-FC Procedure and it must be determined that the quality or quantity of information provided by this method exceeds that which would be obtained with the traditional LLNA. In other words, are the additional mice required by the BrdU-FC worth any possible additional information that would be gained compared to conducting a traditional LLNA?

DRAFT ICCVAM Recommendations: LLNA DA Procedure

Beyond the method to assess lymph node cell proliferation, the test protocol for the LLNA DA contains several key deviations from the OECD Test Guideline 429 recommended protocol and the Essential Test Method Components as described in the Draft ICCVAM Performance Standards for the LLNA . As indicated in the recommendation document (lines 77-79), the LLNA DA has made major modification to the traditional LLNA in both the test substance treatment and sampling schedule. Therefore, this method is outside of the requirements of the draft ICCVAM Performance Standards for the LLNA and should not be consider for validation as an LLNA alternative at this time.