Organisation for Economic Co-operation and Development

ENV/JM/MONO(2018)17

Unclassified

English - Or. English

1 March 2019

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

Cancels & replaces the same document of 27 November 2018

REPORT OF THE PEER REVIEW OF THE VALIDATION STUDY FOR LLNA: BRDU-FCM

SERIES ON TESTING AND ASSESMENT
Number 284

JT03443963

OECD Environment, Health and Safety Publications

Series on Testing and Assessment

No. 284

REPORT OF THE PEER REVIEW OF THE VALIDATION STUDY FOR LLNA: BRDU-FCM

IOMC
INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS
A cooperative agreement among FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD

Environment Directorate
ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT
Paris 2018

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 35 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in twelve different series: **Testing and Assessment**; **Good Laboratory Practice and Compliance Monitoring**; **Pesticides**; **Biocides**; **Risk Management**; **Harmonisation of Regulatory Oversight in Biotechnology**; **Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; Safety of Manufactured Nanomaterials;** and **Adverse Outcome Pathways.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (www.oecd.org/chemicalsafety/).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organizations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

This publication is available electronically, at no charge.

For this and many other Environment, Health and Safety publications, consult the OECD's World Wide Web site (www.oecd.org/ehs)

or contact:

OECD Environment Directorate,
Environment, Health and Safety Division
2, rue André-Pascal
75775 Paris cedex 16
France

Fax : (33-1) 44 30 61 80

E-mail : ehscont@oecd.org

Foreword

This document contains the report of the Peer Review Panel on the validation study of the Local Lymph Node Assay: BrdU-FCM (LLNA: BrdU-FCM), for evaluating the skin sensitising potential of chemicals. The validation study was coordinated by the Korean Centre for the Validation of Alternative Methods (KoCVAM) between 2012 and 2015. The project for the development of the LLNA: BrdU-FCM and its inclusion in TG 442B (LLNA BrdU-ELISA or –FCM) was proposed by Korea and included in the WNT Programme of Work in 2016.

Together with the validation report, the peer review report was made available as a supporting document during two commenting rounds of the Working Group of the National Co-ordinators of the Test Guidelines Programme (WNT) on the draft Test Guideline TG 442B, in July and December 2017 respectively. The peer review report was endorsed by the WNT at its 30th Meeting in April 2018.

The Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology agreed to the declassification of the peer review report on 30 June 2018. This document is published under the responsibility of the Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology.

Table of contents

1. SUMMARY

This document presents the summary report of the assessment of the validation of the Local Lymph Node Assay: 5-bromo-2-deoxyuridine-flow cytometry method (LLNA: BrdU-FCM) by an independent Peer Review Panel (PRP).

The validation study was coordinated by the Korean Centre for the Validation of Alternative Methods (KoCVAM) between 2012 and 2015.

The review was conducted between August and November 2016. The PRP was asked to evaluate how the validation study addressed the principles outlined in the OECD GD34: Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Written responses were collected and discussion of issues was facilitated by an independent Peer Review Manager.

The PRP concluded that the LLNA: BrdU-FCM validation study meets the validation principles of OECD GD 34, considering the performance standards for assessment of similar or modified LLNA test methods for skin sensitization in Annex 1 of OECD TG 429 - Skin sensitization: Local lymph node assay. Therefore LLNA: BrdU-FCM should be suitable for hazard identification of skin sensitizers. It was noted however that these performance standards represent the minimal criteria necessary to demonstrate that the method is acceptable for skin sensitization hazard identification. In addition, the PRP concluded that the validation study from July 2016 did not provide sufficient evidence for the performance of the assay in relation to potency estimation and sub-categorisation of sensitizers.

The PRP made a number of general and specific recommendations to strengthen the robustness of the validation study and clarify technical aspects of the protocol before it can be developed into a test guideline for regulatory use.

2. BACKGROUND

This document presents the summary report of the assessment of the validation of the Local Lymph Node Assay: 5-bromo-2-deoxyuridine-flow cytometry method (LLNA: BrdU-FCM) by an independent Peer Review Panel (PRP).

LLNA: BrdU-FCM is based on the essential elements of the traditional LLNA test method (TG 429) and its non-radioactive variants LLNA: DA (TG 442A) and LLNA: BrdU-ELISA (TG 442B). These assays evaluate the skin sensitization potential of chemicals by measuring lymphocyte proliferation in the auricular lymph nodes draining the site of the application of the test substance during the induction phase of the skin sensitisation adverse outcome pathway. The LLNA: BrdU-FCM specifically measures the changes of the numbers of live proliferating lymphocytes using a flow cytometry method and the non-radioactive nucleoside analogue 5-bromo-2-deoxyuridine (BrdU) as a marker for lymphocyte proliferation. Thus the LLNA: BrdU-FCM represents an adaptation of the LLNA: BrdU-ELISA method which monitors BrdU incorporation in proliferating lymphocytes using Enzyme-Linked Immunosorbent Assay.

The LLNA: BrdU-FCM was developed with financial support of Korea's Ministry of Food and Drug Safety (MFDS) in 2009. The FCM protocol under validation uses the BALB/c mouse strain instead of the CBA strain that is more traditionally used in the other LLNA test methods, even though other strains may be used if data are available to demonstrate no significant strain-specific difference. This protocol modification aims to address the issue of the limited availability of the CBA mouse strain in some countries. In addition, an opportunity for further reduction in the number of animals used in the pre-screen of LLNA protocols, addressing the reduction and refinement of the 3Rs principles, is proposed and could be applied to all LLNA test protocols.

The validation study was coordinated by the Korean Centre for the Validation of Alternative Methods (KoCVAM) with the support of the National Institute of Food and Drug Safety Evaluation (NIFDS) between 2012 and 2015. During this time the initial protocol (protocol 1.0) underwent 3 technical updates (protocols 1.1; 1.2 and 1.3), including modifications in the dose finding pre-test, choice of vehicles and solubilisation procedure for substances with low solubility. The validation study was designed in accordance with the validation principles of the OECD Guidance Document 34 - *Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment* (OECD GD 34).

The aim of the validation study was to evaluate the transferability, reproducibility and relevance of the LLNA: BrdU-FCM for evaluation of skin sensitizers and inclusion as an LLNA "me-too" method in the LLNA TG 442 series. Therefore, the validation study was performed following the performance standards for assessment of proposed similar or modified LLNA test methods for skin sensitization in Annex 1 of OECD TG 429 - Skin sensitization: Local lymph node assay. Transferability, between and within laboratory reliability of the assay evaluation was performed using only the initial protocol while predictive performance of the method was evaluated for all the updated protocols. The peer review focused on the evaluation of the predictive performance of the latest protocol used, protocol 1.3.

The validation report was initially discussed at the OECD Meeting of the Expert Group on Skin Sensitisation in October 2015. The Expert Group recommended that the method could be considered as a project to update TG 442B (LLNA: BrdU-ELISA), rather than to create

another LLNA TG. The project to update TG 442B (LLNA-BrdU-ELISA) submitted by Korea as the lead country, was included in the Test Guideline Programme Work Plan of the Working Group of the National Coordinators of the Test Guidelines Programme (WNT) in April 2016. As part of this project an independent peer review of the validation study was initiated by the OECD in July 2016.

3. THE PEER REVIEW PROCESS

The Peer Review Panel (PRP) was established in August 2016 in order to provide an independent review of the validation of the LLNA: BrdU-FCM assay for evaluation of the skin sensitisation potential of chemicals.

The review experts were nominated by the WNT and were then approached by an independent Peer Review Manager (PRM) who was contracted by the OECD to coordinate the work. The members of the Panel are listed in Annex 1.

Furthermore, OECD approached the lead country of the project to nominate two representatives of the KoCVAM validation team as observers to support the panel to clarify open issues. Dr. Yong Heo from the College of Medical and Public Health Sciences, Catholic University of Daegu, Korea, study director of participating lab 1 in the validation study and Mr. Ilyoung Ahn, Scientific Officer at the Korean National Institute of Food and Drug Safety Evaluation and a member of the validation management team, were nominated and supported the PRP.

The PRP was asked to review the LLNA: BrdU-FCM Validation Study Report dated July 2016, and respond to charge questions based on the OECD GD 34: *Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment*.

In addition, the following articles (referred to in the attachments of the validation study report) were provided to the PRP as background documents for the review:

1. Jung et al., Toxicol.Let. 2010 (pre-validation study)
2. Jung et al., Toxicol.Let. 2012 (pre-validation study)
3. Yang et al., Toxicol.Let. 2015 (inter- and intra-laboratory reproducibility)
4. Kim et al., J. Pharmacol. Toxicol. Methods 2015 (1st and 2nd predictability tests)
5. Ahn et al., Regul. Toxicol. Pharmacol. 2016 (3rd predictability test)
6. Lee et al., submitted manuscript 2016 (Comparison of BALB/c and CBA/J mice for LLNA: BrdU-FCM)

The charge to the Panel was to assess to what extent the eight OECD validation criteria set out in the

OECD Guidance Document 34 had been met. The charge questions as provided to the Panel are listed in Annex 2.

Each Panel member provided written responses to the charge questions to the PRM by 26 September 2016. Compilation of the written comments (Annex 3) was circulated to the panel on 3 October 2016 and the issues identified were discussed at the teleconference on 11 October 2016.

At the teleconference KoCVAM representatives presented a summary of the validation study (Annex 4) addressing specific clarification points raised in the written comments and providing additional information requested by the reviewers. A discussion followed on the more general issues raised by the PRP.

Based on the initial written responses and the first teleconference discussion, a draft report was compiled by the PRM and provided to the Panel for review and comments on October 24, 2016. The Panel commented on the draft report until November 7, 2016 and provided additional comments on aspects of the validation study (Annex 3, additional comments). After accounting for this feedback and the discussion at the teleconference held on

November 10, 2016, the final report was updated by the PRM and sent to the PRP for comments and agreement on November 14, 2016.

This report presents the summary of the assessment of the validation study report on the LLNA: BrdU-FCM dated July 2016 and the resulting agreed responses of the PRP to each of the charge questions.

4. VALIDATION PRINCIPLE 1 – THE RATIONALE FOR THE TEST METHOD SHOULD BE AVAILABLE

Charge Question 1: *Do you consider that the rationale for the BrdU-FCM test is clearly elaborated in the validation report in terms of its scientific basis, regulatory purpose and need?*

Generally the PRP considered that the scientific basis, regulatory purpose and the need for the LLNA: BrdU-FCM are clearly described in the validation report.

It was noted that the scientific basis of the LLNA: BrdU-FCM is clear and sound. It was noted that measuring proliferation of lymph node cells following topical exposure has been used to evaluate skin sensitization potential of chemicals since 2002 with the LLNA method initially outlined in the TG 429 and more recently with similar methods, LLNA: BrdU-ELISA (TG 442B) and LLNA: DA (TG 442A).

It was generally agreed that LLNA: BrdU-FCM assay addresses the regulatory requirements for identification of skin sensitizers while potentially reducing testing on animals. The potential reduction in the number of animals is due to the specific modifications in the pre-test screen in LLNA: BrdU-FCM which requires minimum 10 and maximum 16 animals compared to the maximum 18 animals required for pre-screen in the other LLNA assays. However, it was emphasised that while notable in some cases, the reduction may not always be significant as two additional animals are needed for the main test (blank, no BrdU control, and no-treatment BrdU injection control) compared to other LLNA assays. Nevertheless, it was agreed that the reduction of test animals is significant compared to the guinea pig based methods and that the principle of the LLNA: BrdU-FCM pre-screen test provides an opportunity for further reduction if applied to the other LLNA methods.

Few panel members expressed concerns that the assay analysis may be limited in respect to the regulatory need for evaluation of sensitisation potency and sub-categorisation of sensitizers.

Underlying these concerns were:

(i) In terms of the validation study: potentially limited number of reference chemicals used for the evaluation of performance of the assay in discriminating skin sensitizers from non-sensitizers and for potency analysis i.e. ECt determination; and

(ii) In terms of addressing regulatory need: limitations of the protocol regarding the choice of test concentration potentially leading to omission of sufficiently low test concentrations to meet specific regulatory requirements (e.g. extreme sensitizers with EC3 for LLNA ≤0.2% under Regulation (EC) No 1272/2008 on classification, labelling and packaging (CLP)).

The panel discussed these concerns at the first teleconference and agreed that:

(i) Based on the essential elements of the LLNA: BrdU-FCM method, the 22 reference chemicals specified in Annex 1 of the TG 429 and developed specifically for evaluating the validity of similar or modified LLNA assays, currently represent the most appropriate set of chemicals for evaluation of the performance of this method for reasons of efficiency and animal welfare. It was noted however that this set of reference

substances covers the minimal requirements and that prospective analysis of additional substances would have been valuable.

Some reviewers also pointed out that the limitations of LLNA: BrdU-FCM in terms of correct classifications of sensitizers and/or their sub-categorisation should be viewed in the context of the other two non-radioactive modifications of LLNA, LLNA: BrdU-ELISA and LLNA: DA. The latter two also have particular limitations in the classifications of some known human sensitizers generating borderline results. It was agreed that the specific limitations of LLNA: BrdU-FCM should be emphasised in the validation report (discussed in more detail under validation principle 6) and also in the future test guideline.

In their presentation KoCVAM representatives included additional information summarising an analysis of classification and sub-categorisation of sensitizers using all LLNA methods compared to human data for the test substances used in the validation study.

(ii) The protocol and the validation study (including the future test guideline) need to incorporate text changes to emphasise the point that a sufficient number (minimum 3) and range of concentrations should be considered to address specific regulatory requirements, as appropriate. In addition, all information on the test substance such as the chemical structure, physicochemical properties and data from any available relevant toxicological studies, including these on structurally related test substances, should be taken into account for the dose selection.

A concern was reiterated by one reviewer that the protocol has not sufficiently addressed potential limitation in terms of the choice of high test concentrations that may potentially produce false negative results for skin sensitizers with high EC values. Namely Section 3.3 of the protocol states that: 3.3 Dose selection: "Doses of 100%, 50%, 25%, 10%, 5%, 2.5%, 1%, and 0.5% should be used according to OECD TG 429." To address the issue it is recommended that the wording of the protocol be aligned with paragraph 18 of TG 429 as follows: "Dose and vehicle selection should be based on the recommendations given in references (3) and (5) of TG 249. Consecutive doses are normally selected from an appropriate concentration series such as 100%, 50%, 25%, 10%, 5%, 2.5%, 1%, 0.5%, etc. Adequate scientific rationale should accompany the selection of the concentration series used…". The wording from paragraph 28 of TG 249 regarding use of rLLNA to confirm suspected negative results in particular circumstances should also be considered for the LLNA: BrdU-FCM protocol.

The panel discussed the issue at the teleconferences. KoCVAM representatives explained that the Standard Operating Procedure for the protocol was based on TG 429. The panel agreed that aligning the SOP more closely with principles for dose selection in TG 429 is needed.

In their written comments reviewers had differing views on the discussion in the validation report with regard to the importance and role of the detection of T/B cells ratios and other immunotyping markers of skin sensitisation in the validation study. While some argued for more detailed information and further discussion, others had the view that emphasis on these aspects for skins sensitisation should be tuned down in the report.

Following the discussion at the first teleconference, the panel agreed that while valuable, detailed discussion of particular immunotyping markers of skins sensitisation is outside of the scope of the validation study. In addition, the current assay protocol does not specifically measure any of these parameters. However it was recognised that the LLNA:

BrdU-FCM could accommodate future modifications to address relevant parameters, which was recognised as a potential advantage of this assay compared to the other non-radioactive LLNA methods. Inclusion of representative reference(s)[1] reviewing the current limited knowledge on the role of the detection of T/B cells ratios and other immunotyping markers for identification and analysis of skin sensitizers was advised.

The review panel also agreed that the use of the non-radioactive proliferation marker BrdU represents an advantage for LLNA: BrdU-FCM compared to the traditional LLNA as well as the use of the more readily available and cost efficient mouse strain BALB/c. While the rationale for the use of the alternative mouse strain was largely accepted, the substantiation of its choice in terms of sensitivity, SI cut-off and ECt determination was discussed further under validation principle 6.

Overall, the peer review panel agreed that the Validation Principle 1 has been met. However, a recommendation is made for specific clarifications in the protocol description to better align it with the principles for dose selection in TG 429 and effectively address regulatory applicability of the assay.

[1] Recommended list in Annex 7

5. VALIDATION PRINCIPLE 2 – THE RELATIONSHIP BETWEEN THE TEST METHOD'S ENDPOINT(S) AND THE (BIOLOGICAL) PHENOMENON OF INTEREST SHOULD BE DESCRIBED

Charge Question 2: *Is the relationship between the test method's endpoint(s) and the (biological) phenomenon of interest described adequately using relevant references for the scientific relevance of the effect(s) measured (in terms of their mechanistic (biological) or empirical (correlative) relationship) to the specific type of effect/toxicity of interest?*

The reviewers responses to the charge question 2 reflected a general agreement that the mechanistic basis of the test method is clearly described with a reference to the Adverse Outcome Pathway (AOP) for skin sensitization and the key event that this test method measures. Some additional literature referencing was suggested (Annex 3).

The peer review panel agreed that the Validation Principle 2 has been met.

6. VALIDATION PRINCIPLE 3 – A DETAILED PROTOCOL FOR THE TEST METHOD SHOULD BE AVAILABLE

Charge Question 3: *Do you consider that the protocol description in the validation report is sufficiently detailed, including a description of all the materials needed (e.g. specific cell types or construct or animal species), a description of what is measured and how it is measured, a description of appropriate data analysis, decision criteria for evaluation of data and what are the criteria for acceptable test performance?*

Overall, the reviewers found the protocol description to be quite detailed and clear. However, some specific additions and clarifications to the protocol description were suggested, including technical specifications (Annex3). KoCVAM representatives addressed some in their presentation at the teleconference (Annex 4) and agreed to make appropriate corrections and additions in an updated version of the report before submitting it to OECD, based on the written comments and the recommendations of this report.

The panel specifically discussed the concern of some reviewers that the protocol is restrictive in terms of vehicle choice potentially excluding testing of some chemicals at appropriate concentrations. KoCVAM representatives explained that the vehicle selection protocol is optimised for maximum solubility with minimum toxicity and is not aimed as excluding other appropriate vehicles provided that there is justification for their use. However, to address this concern it was agreed to align the LLNA: BrdU-FCM protocol with paragraph 19 of the TG 429 that opens up a possibility for "others (vehicles) may be used if sufficient scientific rationale is provided".

Another reviewer raised the question about selection of potentially more appropriate SI cut-off value in the context of the protocol description. This matter was raised by other reviewers in the context of method performance and it was discussed in more detail under validation principle 6 (see below).

One reviewer suggested that information on newer equipment with integrated settings appropriate for regulatory and clinical use, such as BD FACSVia™ and BD FACSLyric™ should be mentioned in the report.

Overall, the peer review panel agreed that the Validation Principle 3 has been met. However, a recommendation is made for specific clarifications in the protocol description to better align it with the principles for vehicle selection in TG 429.

7. VALIDATION PRINCIPLE 4 – THE INTRA-, AND INTER-LABORATORY REPRODUCIBILITY OF THE TEST METHOD SHOULD BE DEMONSTRATED

Charge Question 4: *Do you consider that the intra-, and inter-laboratory reproducibility of the BrdU-FCM test method is adequately demonstrated, considering availability of the data over time, as well as the degree to which biological variability affects the test method reproducibility?*

In the initial written comments majority of reviewers agreed that the intra- and inter-laboratory reproducibility of the LLNA: BrdU-FCM was adequately demonstrated, according to the performance standards for LLNA assays (Annex 1 of TG 429), using HCA[2] and DNCB[3] and following protocol 1.0. It was noted that the changes in the protocol concern solubility of substances other than the two used in the evaluation of WLR and BLR and therefore would not influence WLR and BLR findings.

However, one reviewer maintained that the reproducibility and transferability of the assay is not sufficiently demonstrated and should have included more substances covering a broader range of sensitisation activity tested with the latest protocol.

Significant variability (~50%) of the results in one of the laboratories (participating laboratory 2) was noted by one reviewer who suggested consideration be given to whether the variability in the antibody batches may be a contributing factor. KoCVAM representatives agreed to look into this possibility and elaborate on the issue in the report. However in their opinion the limited experience of the laboratory with the assay in the initial stages of the validation study may also explain the noted variability. This reviewer also recommended that laboratory proficiency requirements with stricter CV limits should be considered for the test guideline development.

The majority of the peer review panel agreed that the Validation Principle 4 has been met. However, there was a minority opinion that reproducibility and transferability of the assay should be evaluated with more test substances and with the latest protocol available.

[2] HCA - hexyl cinnamic aldehyde

[3] DNCB - 2,4-dinitrochlorobenzene

8. VALIDATION PRINCIPLE 5 – DEMONSTRATION OF THE TEST METHOD'S PERFORMANCE SHOULD BE BASED ON THE TESTING OF REFERENCE CHEMICALS REPRESENTATIVE OF THE TYPES OF SUBSTANCES FOR WHICH THE TEST METHOD WILL BE USED.

Charge Question 5: *Are the reference chemicals used to demonstrate the performance of this test method representative of the types of substances for which the test method will be used and have they been tested under code to exclude bias?*

The majority of reviewers considered that the reference chemicals used in the validation study have been specified in the performance standards of TG 429 as representative of the types of substances typically tested for skin sensitization potential and providing the range of responses that the LLNA-based assays would be capable of measuring or predicting.

In their initial written responses two reviewers consider that data for more chemicals (up to 20 were suggested) need to be generated and independently analysed to adequately address the validation principle 5.

The panel discussed the issue of potentially limited number of substances in the performance standards of OECD TG 429. It was agreed that this is likely a small set of substances for performance evaluation. However it was recognised that testing more reference chemicals is not preferred from an animal welfare perspective (see also discussion under validation principle 1 (i)). It was also reiterated that the performance standards have been developed by ICCVAM[4]-NICEATM[5], ECVAM[6], and JaCVAM[7] to effectively facilitate validation of new and adapted LLNA methods.

However, the panel agreed that prospective analysis of results for additional substances tested using the protocol, updated based on this review and agreed under the OECD relevant review groups would be useful for re-evaluation of the performance of the assay in the future.

All reviewers agreed that the coding and handling of the reference chemicals have been in accordance with the validation criteria of OECD GD 34.

The peer review panel agreed that the Validation Principle 5 has been met, considering performance standards for modified LLNA methods specified in Annex 1 of the OECD TG 429. However, the panel also agreed that prospective analysis of results obtained with the latest updated protocol and testing substances other than these specified in the performance standards of TG 429 should be considered for future re-evaluation of the performance of LLNA: BrdU-FCM vis-a-vis LLNA methods and data from humans.

[4] Interagency Coordinating Committee on the Validation of Alternative Methods

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

[6] European Centre for the Validation of Alternative Methods

[7] Japanese Center for the Validation of Alternative Methods

9. VALIDATION PRINCIPLE 6 – THE PERFORMANCE OF THE TEST METHOD SHOULD HAVE BEEN EVALUATED IN RELATION TO RELEVANT INFORMATION FROM THE SPECIES OF CONCERN, AND EXISTING RELEVANT TOXICITY TESTING DATA.

Charge Question 6: *Do you consider that the performance of the test method has been evaluated in relation to relevant information from the species of concern, and existing relevant toxicity testing data?*

Overall the panel considered that the performance of the assay was evaluated adequately using the reference chemicals specified in the performance standards for assessment of similar or modified LLNA test methods for skin sensitization.

It was noted by one reviewer that the performance standards reference chemicals have been selected based on previous analysis of known and/or acknowledged responses in humans, mice, and guinea pigs. In addition, it was recognised that the validation report discussed more recent data from public literature to evaluate the results, particularly for the false negative chemicals. However, one reviewer suggested that further discussion of data from studies of contact dermatitis in humans is needed including additional literature (suggested citations included in the written comments Annex 3) to better address the limitation of the assay in detecting some known human sensitizers.

Discussing the issue of the two false negative reference chemicals (mercaptobenzothiazole and methyl methacrylate) at the teleconference, it was pointed out that all LLNA variants have limitations in predicting weak and moderate human skin sensitizers with particular physicochemical properties. In this respect LLNA: BrdU-FCM is similar to other LLNA assays and should not be held to higher performance standards.

The panel discussed whether the SI cut-off value of 2.7 is optimal for hazard identification and whether further analysis of the existing or additional reference chemicals could inform a more optimised SI cut-off and consequently better performance. It was noted that higher sensitivity is preferred for regulatory purposes.

The KoCVAM representatives addressed the panel's discussion (Annex 5) and advised that based on the opinion of the expert biostatistician on the validation team, the analysis would benefit from a larger data set. However, given the recommended reference chemical set in Annex 1 of TG 429, the statistical analysis presented in the Annex 11 of the validation report is quite elaborate. It includes analysis of the performance using SI cut-offs from 2.2 to 4.3, analysis of the performance using SI value standardised to that of the concurrent positive control and also inferential statistics to account for the small sample size. Overall, the SI cut-off value of 2.7 was found as optimal for maximal overall accuracy.

It was recognised that the performance standards reference chemicals represent the minimal data set that could be considered and that analysis of additional chemicals, preferably by ICCVAM, would be valuable. One reviewer noted that during the validation of the similar LLNA assays, LLNA: BrdU-ELISA and LLNA: DA, the SI was chosen post-hoc to capture the optimal SI values for sensitizers in the reference set tested at the time.

The panel discussed two suggestions for alternative selection of SI cut-off raised in the written comments:

- SI of 1.8 for LLNA: BrdU-FCM based on the ROC analysis in Figure 6 in the validation report. This SI cut-off shows acceptable overall performance while increasing the sensitivity of the method by eliminating the false negative mercaptobenzothiazole.

- Using "corrected BALB/c SI" (or corresponding "CBA SI") calculated by regression equation based on the sensitivity comparison between BALB/c and CBA using the data from Woolhiser, 2000; Lee, submitted. These data show good correlation but clear difference in sensitivities between the two strains. Using the regression equation (Annex 3, responses to Q6 peer reviewer 9) SI results for mercaptobenzothiazole and methyl methacrylate would be considered sensitizers, improving the sensitivity of the assay.

Following the discussion and presentation by KoCVAM (Annex 5), the panel agreed that there is no statistical rationale for lowering the SI cut-off below 2.7 using the available data and that additional analysis of the presented data is not likely to provide new conclusions. However, the difference in sensitivity of the two mouse strains was recognised. Additional analysis contributed by panel members (Annex 6) provided regression analysis comparing the sensitivity between BALB/c and CBA/J in the traditional LLNA and LLNA: BrdU FCM. It was recommended that this analysis, updated with other data available to KoCVAM, should be included in the validation report to support the development of the test guideline for LLNA: BrdU-FCM which would be inclusive of mouse strains other than BALB/c.

In addition it was suggested that additional 2x2 analysis of the performance of the assay using 1.8 as an SI cut-off value is included in the validation report, more clearly demonstrating the effect on sensitivity, specificity and accuracy of the method under these criteria.

Finally the panel agreed that the SI cut-off value of 2.7 is optimal for hazard identification considering the available data. However, it was recognised that potency estimation and sub-categorisation of sensitizers has not been addressed in the July 2016 validation report.

The peer review panel agreed that the Validation Principle 6 has been met with regard to the performance of the assay for hazard identification of skin sensitizers. However, it was recognised that potency estimation and sub-categorisation of sensitizers has not been addressed in the July 2016 validation report and needs additional consideration.

10. VALIDATION PRINCIPLE 7 – IDEALLY, ALL DATA SUPPORTING THE VALIDITY OF A TEST METHOD SHOULD HAVE BEEN OBTAINED IN ACCORDANCE WITH THE PRINCIPLES OF GLP

***Charge Question** 7: Have all of the data supporting the validity of a test method been obtained in accordance with the principles of GLP? If not, has an adequate consideration been given to the potential impact on the validation status of the test method?*

Peer reviewers agreed that the data supporting the validity of the LLNA: BrdU-FCM were obtained according to GLP principles. However, it was suggested that the available certifications of QAU inspections be placed in accessible space.

The peer review panel agreed that the Validation Principle 7 has been met.

11. VALIDATION PRINCIPLE 8 – ALL DATA SUPPORTING THE ASSESSMENT OF THE VALIDITY OF THE TEST METHOD SHOULD BE AVAILABLE FOR EXPERT REVIEW

Charge Question 8: *Do you consider that all the data supporting the assessment of the validity of the test method are easily available for expert review? These include: a detailed and readily available test method protocol to the public and independent laboratories; organised and easily accessible data to permit independent review(s); benchmarks by which an independent laboratory can itself assess its proper adherence to the protocol?*

Peer reviewers considered that the authors have made a good effort to present the data and information clearly. However a number of discrepancies between text, tables and appendices were identified.

Specific suggestions were made in the written comments to improve clarity and accessibility of information (Annex 7).

Inclusion of raw data used in the evaluation of the assay performance with protocols 1.1. and 1.2 was also requested as well as detailed information about the type and settings of the FACS scan equipment used. Analysis of the additional information with regard to the equipment and measurement settings can provide evidence for sources of the variability and potentially inform future re-evaluation of SI cut-off value and assay performance.

The peer review panel agreed that the Validation Principle 8 has been met.

12. CONCLUSIONS

The PRP concluded that the LLNA: BrdU-FCM assay validation has been performed according to the validation principles in OECD GD 34 and using the reference chemicals specified in Annex 1 of the TG 429 for the evaluation of new and modified LLNA methods.

Based on the above performance standards, the validation study satisfies only the minimal criteria specified in Annex 1 of the TG 429 to demonstrate that the method should be acceptable for identifying substances as potential skin sensitizers and that evaluation of additional substances could significantly strengthen the validation study. However, there was a minority opinion that reproducibility and transferability of the assay is not sufficiently evaluated in the validation study as only the first protocol and not the final protocol was used to test this with only two chemicals specified in the performance standards of TG 429.

An SI of 2.7 is considered an optimal cut-off value for hazard identification and discrimination of non-sensitizers from sensitizers, based on the post-hoc analysis of the performance of the assay using the performance standards reference chemicals specified in TG 429. However, the validation study from July 2016 did not provide sufficient evidence for the performance of the assay in relation to potency estimation and sub-categorisation of sensitizers. LLNA: BrdU-FCM can potentially generate false negative results for some known human sensitizers, particularly using the protocol with BALB/c mouse which is evidently less sensitive than the more communally used CBA/J in this assay (Annex 6).

The current protocol using 2.7 as the SI cut-off value does not identify 2-mercaptobenzothiazole and methyl methacrylate as skin sensitizers. However it is noted that these are known to provide variable results in the traditional LLNA and in the other non-radiolabeled LLNAs. Therefore, the performance of LLNA: BrdU-FCM is considered comparable to other LLNA methods.

It is also noted that the assay can have high variability in some laboratories (over 50% in the case of one participating laboratory). Therefore, it is recommended that the source of the high variability is determined and/or a laboratory proficiency with stricter CV limits should be considered.

The protocol described in the validation study from July 2016 was considered restrictive in terms of vehicle and test concentration selection as compared to other LLNA tests. Therefore specific modifications are recommended for protocol 1.3 to align it with TG 429 (see above under Validation principle 1 and 3 for details).

Notwithstanding some of the limitations discussed above, the LLNA: BrdU-FCM provides opportunities to:

- reduce animal usage compared to guinea pig test but also in all LLNA tests through the application of the modified dose finding pre-test
- reduce the use of radioisotopes and animal use where guinea pigs have to be used because of prohibitions of use of radioisotopes
- use the LLNA method in countries where the mouse strains usually used, such as CBA/J, are not readily available
- introduce modifications that would allow monitoring of immunotyping parameters relevant for skin sensitisation

13. RECOMMENDATIONS

Additional analysis of the performance of the LLNA: BrdU-FCM in relation to sub-categorisation of sensitizers compared to other LLNA methods and data from humans should be included in the report.

Modifications to the protocol should be included to ensure selection of other appropriate vehicles and sufficiently low and high concentrations be tested in line with other LLNA assays. Specifically modifications should follow paragraphs 18, 19, 28 and 29 of the TG 429.

Discussion should be included in the validation report addressing the sources of variability of the results in some laboratories. In addition, consideration should be given to including laboratory proficiency standards with better defined CV limits for the test guideline development.

Additional quantitative sensitivity comparison of BALB/c and CBA/J for all available substances tested in LLNA: BrdU-FCM should be included in the report.

The validation report should include raw data used in the evaluation of the assay performance with protocols 1.1. and 1.2.

The validation report should include detailed information about the type of FACS scan equipment used and the settings applied for the measurements.

All available certifications of QAU inspections be placed included in the validation report or placed in a publically accessible space.

Specific recommendations for correcting errors, addressing discrepancies and editorial changes (Annex 7) should be addressed.

Acknowledgements:

KoCVAM thanked Peer Review Panel for their review and valuable comments.

Annex 1- Peer Review Panel

Janine Ezendam	National Institute of Public Health and the Environment - RIVM (NL)
Saadia Kerkine-Römer	Faculté de Pharmacie, Université Paris-Sud (FR)
Judy Strickland	Integrated Laboratory Systems, Inc., a contractor supporting NICEATM (USA))
Hajime Kojima	JaCVAM, National Institute of Health Sciences (JP)
Masahiro Takeyoshi	Chemical Evaluation and Research Institute (JP)
Ok-Nam Bae	College of Pharmacy, Hanyang University ERICA Campus (KO)
Bae-Hwan Kim	Department of Public Health, Keimyung University (KO)
Carola Lidén	Institute of Environmental Medicine | Karolinska Institutet (SE)
William Irwin	US Environmental Protection Agency (USA)
Hermann-Josef Thierse	Federal Institute for Risk Assessment (GER)

Peer Review Manager Julija Filipovska, Independent Consultant

Annex 2 - Charge Questions for the Peer Review of the Validation Report on the Local Lymph Node Assay: 5-Bromo-2-Deoxyuridine-Flow Cytometry Method (LLNA: BrdU-FCM)

Questions based on the validation principles and criteria in OECD GD34 (link)

Q1: Do you consider that the rationale for the BrdU-FCM test method is clearly elaborated in the validation report in terms of its scientific basis, regulatory purpose and need?
A1.
Q2: Is the relationship between the test method's endpoint(s) and the (biological) phenomenon of interest described adequately using relevant references for the scientific relevance of the effect(s) measured (in terms of their mechanistic (biological) or empirical (correlative) relationship) to the specific type of effect/toxicity of interest?
A2:
Q3: Do you consider that the protocol description in the validation report is sufficiently detailed, including a description of all the materials needed (e.g. specific cell types or construct or animal species), a description of what is measured and how it is measured, a description of appropriate data analysis, decision criteria for evaluation of data and what are the criteria for acceptable test performance?
A3:
Q4: Do you consider that the intra-, and inter-laboratory reproducibility of the BrdU-FCM test method is adequately demonstrated, considering availability of the data over time, as well as the degree to which biological variability affects the test method reproducibility?
A4:
Q5: Are the reference chemicals used to demonstrate the performance of this test method representative of the types of substances for which the test method will be used and have they been tested under code to exclude bias?
A5:
Q6: Do you consider that the performance of the test method has been evaluated in relation to relevant information from the species of concern, and existing relevant toxicity testing data?
A6:
Q7: Have all of the data supporting the validity of a test method been obtained in accordance with the principles of GLP? If not, has an adequate consideration been given to the potential impact on the validation status of the test method?
A7:
Q8: Do you consider that all the data supporting the assessment of the validity of the test method are easily available for expert review? These include: a detailed and readily available test method protocol to the public and independent laboratories; organised and easily accessible data to permit independent review(s); benchmarks by which an independent laboratory can itself assess its proper adherence to the protocol.
A8:

Annex 3 - Compilation Responses

Peer Review of the Validation Report on the Local Lymph Node Assay: 5-Bromo-2-Deoxyuridine-Flow Cytometry Method (LLNA: BrdU-FCM)

August-September 2016

Q1: Do you consider that the rationale for the BrdU-FCM test method is clearly elaborated in the validation report in terms of its scientific basis, regulatory purpose and need? ***Validation principle a: The rationale for the test method should be available.*** This should include a clear statement of the scientific basis, regulatory purpose and need for the test.	PR1 The scientific basis and regulatory purpose and need are clearly described in the validation report. However, what was missing in Introductory part of this report was the added value of this BrdU-FCM test method compared to the other non-radioactive LLNA test guidelines that are already available. This is important because this validation study concerns an animal test method and therefore the rationale for choosing a different methodology to measure lymphocyte proliferation needs to be provided. The Background section does mention that FCM enables the cell subtyping, e.g. the analysis of B cell to T cell ratios, which clearly is not possible with the BrdU-ELISA. However, the added value of this in terms of improved predictivity or other advantages compared to the other non-radioactive variants is missing. There is clearly a regulatory need to obtain potency information for skin sensitizers. In the validation report it is mentioned that the non-radioactive LLNA variants can be used to assess potency. However, according to ECHA guidance, both the LLNA: BrdU-ELISA and LLNA: DA do not provide any criteria to estimate the potency. For this new test method, it is therefore important to evaluate if it is able to provide an accurate potency estimate, although I am not sure if this is possible with only 18 reference chemicals. This may be beyond the scope of this peer-review process, but it is an important aspect that needs to be addressed before this test method enters the next phase towards inclusion in the OECD TG442B.
	PR2 The LLNA BrdU-FCM test method is clearly elaborated in the validation report in terms of its scientific basis regulatory purpose. The LLNA (OECD TG429), the gold Assay, which evaluates skin sensitization of chemicals, requires radioactivity. The BrdU-FCM test method does not which make it easier to perform and safer for experiments and the environment. Relating to regulatory purpose, as the standard LLNA adopted as OECD TG 429, it could be used by EU's registration, Evaluation, Authorization and restriction of Chemicals (REACH).

	PR3 Yes. The validation study report clearly describes the scientific basis. The scientific basis of the LLNA: BrdU-FCM is the same as that for other LLNA methods: skin sensitizers induce proliferation of T-lymphocytes in the lymph nodes draining the site at which the substance is applied. Lymph node cell proliferation serves as the surrogate for T-lymphocyte proliferation. The LLNA: BrdU-FCM measures the proliferation of lymph node cells, which is proportional to a skin sensitizer's dose and potency, by assessing the incorporation of labelled live lymph node cells using flow cytometry. The cells are labelled by allowing the proliferating cells to incorporate BrdU into the DNA. The validation report also clearly describes the regulatory purpose and need for the test. LLNA methods are the preferred animal tests for skin sensitization in Europe and the U.S. because they are reduction and refinement alternatives with respect to guinea pig assays for skin sensitization. They are used for hazard classification and labelling and for potency assessment. While the traditional LLNA uses radioactivity to label proliferating lymphocytes there is a need for non-radiolabelled methods so as to avoid the cost and environmental liability of using radioactive isotopes.
	PR4 The LLNA, which was adopted as OECD TG 429, has been used to evaluate the skin sensitization potential of chemicals since 2002. On the other hand, ^{3}H-methyl thymidine (an analogue of thymidine) or ^{125}I-iododeoxyuridine is used in the LLNA, the assay requires facilities that allow handling of radioisotopes and poses the risk of radioactive contamination. Therefore, Korean colleagues developed The LLNA: BrdU-FCM for which ^{3}H-methyl thymidine is not needed, with the aim of disseminating the LLNA as well as LLNA:DA (TG442A) and LLNA:BrdU-ELISA(TG442B). This rationale of development on this assay is a clear for regulatory need. However, it is not clear correlation with T cell proliferation as a biomarker of TG 429 and analysing B cell to T cell ratios (cell sub-typing) by flow cytometry. I request more detailed information on the biomarker for a scientific basis in introduction. In addition, authors in this report should be described at the front or second page.
	PR5 "HAS BEEN MET" The rationale for the BrdU-FCM test method is stated in the report.

	PR6 The predictively and the reliability of LLNA and its modified methods were already well established, and the scientific rationale for the current BrdU-FCM test method is acceptable. However, although the advantage of the current method over other LLNA-based methods is described in the last paragraph in Background and Line 645/673/963/1341, it needs to be more clearly described in Summary, to achieve its regulatory relevance. It is not certain that this protocol is useful for analysis of additional parameters such as T/B ratio or cell surface markers, since these parameters are not for decision of sensitizing potential of chemicals. It should not be emphasized too much.
	PR7 Yes,I agree. The rationale of BrdU-FCM method is scientifically elaborated in the validation report. BrdU-FCM test method funding by NFDS was evaluated the validation study from 2009 to 2015, and they set the SI value as ≥ 2.7. This method can accommodate immunophenotyping to evaluate sensitizers. (See Validation Study Report pp 15-18)

	PR8 It is considered that the scientific basis for the LLNA: BrdU-FCM test method is clearly elaborated. It is repeatedly stated in the validation report that the LLNA: BrdU-FCM can be used to identify skin-sensitizing chemicals and evaluate skin sensitization potency. It is also stated that the method is expected to be used in classifying chemicals as skin sensitizers or non-sensitizers, and in evaluating the skin sensitization potency based on calculated ECt values and subcategorization into 1A and 1B as defined by the GHS and the CLP. These statements are not fully agreed on. *Regulatory needs* concerning prediction of skin sensitizing potency have emerged by the GHS and CLP. Potency information based on EC3 values in the LLNA plays an important role, and is largely based on the considerations in (Basketter et al. Evaluation of the skin sensitizing potency of chemicals by using the existing methods and considerations of relevance for elicitation. Contact Dermatitis 2005;52(1):39-43). During recent years, demands are increasing for hazard identification and risk assessment based on *in vitro* methods. Today, there are three validated *in vitro* methods for skin sensitizing potential, but there is not yet a validated method for assessment of potency. The GHS requires subcategorization of classified skin sensitizers into 1A (strong; EC3 ≤2) and 1B (other; EC3 >2) (when based on LLNA). The CLP requires, according to the CLP guidance document, that the most potent sensitizers shall be identified; when based on LLNA: 1A (extreme; EC3 ≤0.2), 1A (strong; EC3 >0.2-≤2), and 1B (moderate; EC3 >2) (ECHA. Guidance on the application of the CLP criteria. Guidance to Regulation (EC) No 1272/2008 on classification, labelling and packaging (CLP) of substances and mixtures. Version 4.1 June 2015 https://echa.europa.eu/guidance-documents/guidance-on-clp). The CLP potency subcategories generate concentration limits for classification of mixtures: 0.001% (specific concentration limit), 0.1% (generic), and 1% (generic), respectively. *The regulatory need to identify also low ECt values (EC3 ≤0.2) is thus very large. This is however not done by the current BrdU-FCM test method*. One benefit of the LLNA: BrdU-FCM is said to be that fewer animals are used, owing to the proposed pre-screen test for dose selection. The intention is considered good, but, in its current form, it has resulted in inability to identify the most potent skin sensitizers since low test concentrations have been omitted, and potentially to misclassify sensitizers tested at only relatively low concentrations as non-sensitizers. *PROPOSAL*: It is considered necessary that additional testing is performed when the test result indicates that the ECt value is considerably below the lowest tested concentration; and

	additional testing also when all SI are below the cut-off value if only relatively low concentrations have been tested. One rationale for the LLNA: BrdU-FCM is that the test method does not require any radioisotopes, which makes it easier to perform and safer for experimenters and the environment; this is agreed. It should however be noted that BrdU is a hazardous substance. BrdU is, according to notified classification in the CLP, mutagenic (H340, H341), toxic to reproduction (H360, H361), suspected of causing cancer (H351), and irritant to eyes, skin and respiration. *PROPOSAL*: It is suggested, for protection of the health and safety of the experimenters, that such hazard information is mentioned in the validation report and protocol, If such information normally is not included in OECD TG, it is suggested that it should be.
	PR9 Yes, the rationale was clear provide an alternate assay method for dermal sensitization utilizing flow cytometry and a more readily available mouse strain.

	PR10 **Comment on rationale for the BrdU-FCM test method.** In general, since the non-radioisotopic LLNA:BrdU-FCM (LLNA:5-bromo-2deoxyuridine-Flow-Cytometry Method) may be considered as a further development of BrdU-based LLNA:BrdU-ELISA and seem to perform similar to existing LLNA-based test methods, TG429 (LLNA), TG 442A (LLNA: DA) and in particular to the mentioned TG 442B (LLNA: BrdU-ELISA), there is scientific evidence to consider the LLNA:BrdU-FCM as "me-too" test of the LLNA-BrdU-ELISA, and possibly the LLNA. A view that is supported by some accompanying publications. Besides some limitations, which have to be addressed (more) clearly, the new test method offers new options in analyzing sensitizer-specific immune responses, including cell-type specific reactions of T cells (e.g. proliferation of activated T-cells, and differential CD marker analyses) as well as measurement of associated cell-type specific cytokine production, and B cell to T cell ratios. **Scientific basis. Comment on rationale for animal strains used.** It is not completely clear why authors switched to another mouse strain (e.g. p7 line 246) if compared to the LLNA (TG429) and TG 442 A and B. From immunology/allergy testing it is very well known that both, change of species and/or change of strain may significantly influence results on similar *in vivo* test approaches. Thus it might be supportive to have access to both data sets, as well as to other original animal data to better compare both strains. Selection of animal species and a possible change of the strain to be used had already been addressed by the TG429: "Alternatively, other strains and males may be used when sufficient data are generated to demonstrate that significant strain and/or gender-specific differences in the LLNA response do not exist". It is questionable if data presented here are sufficient to meet corresponding requirements mentioned in TG429, even though Lee et al. just recently have submitted a manuscript on comparative skin sensitization experiments in Balb/c mice and CBA/J mice; a study that should not be considered before acceptance for publication, but a study that clearly demonstrated differences in irritant reactions of both strains, and some non-concentration dependent reactions in Balb/c mice (Lee et al., 2016, submitted, Fig.2 J and R, Fig. 3 A, B, J, Q) if compared to CBA/J, but still reveal significant correlation between stimulation index values in both strains (Lee et al., Fig. 4; e.g. page 17 line 666 with little over-interpretation). It is necessary to better integrate critical/limiting issues of the new non-radioisotopic LLNA:BrdU-FCM into the validation report; thereby not weakening but strengthening the report. **It is expected that specific regulatory purpose will be served**, in this case skin sensitization and sensitization potency.

	There is always a need to improve and broaden methods and methods spectra, including the LLNA:BrdU-ELISA and the new LLNA:BrdU-FCM, e.g. regarding 3R approaches or chemical safety and regulation under REACH (2006).
	PR1 The mechanistic basis of the test method is clearly described in the validation report, with a reference to the AOP for skin sensitization and the key event that this test method measures.
	PR2 The basic principle underlying the LLNA BrdU-FCM test is that skin sensitizers induce the proliferation of Lymphocytes in the lymph Nodes adjacent to where the test substances were applied. This assay mimics the first step of allergic contact dermatitis and evaluates T cell proliferation wich is the 4th event in the skin sensitization AOP. The measurement of proliferation is expressed as Stimulation Index (SI) to categorize skin sensitizer. For all these reasons, the relationship between methods (endpoints and the biological phenomenon) sounds quit very well and fit well with refinement & reduction among the 3Rs.
	PR3 Yes. The AOP for skin sensitization was described and the key event assessed by LLNA: BrdU-FCM was explained. Additional references should have been applied, though, to statements such as "skin sensitizers induce the proliferation of lymphocytes the lymph nodes adjacent to where test substances were applied" (lines 234-235 and similar passage on lines 636-637). Also, "degree of proliferation is proportional to dose and potency" (lines 637-638) should be referenced. TG 429 could provide a reference for these statements.
	PR4 The LLNA:BrdU-ELISA is not to measure the contents of BrdU per lymph node with ELISA and this assay is a lack of qualitative analysis. However, the LLNA:BrdU-FCM assay is able to conduct a quantitative assessment of BrdU per lymph mode and I think it is a better method compared with the ELISA method. On the other hand, the SI values of the FCM assay is big and the sensitivity of this assay is low and similar with the ELISA method compared with the original LLNA. Therefore, a drawback of this assay is not improved and this assay is no state-of-the-art technology.

	PR5 "HAS BEEN MET" BrdU is well-known analogue of thymidine, and it can be used as well as radioactive thymidine used in traditional LLNA to monitor LNC proliferation. However further description of appropriateness or advantage of FCM technique for this purpose would be desirable.
	PR6 Yes. The relationship between the test method's endpoints (BrdU-LLNA) and the biological phenomena (skin sensitization in this case) has already been established and it is adequately described in the current protocol.
	PR7 Yes, I agree. This method showed good relationship between the test method's endpoints and the biological phenomenon of interest. The LLNA: BrdU-FCM directly enumerates only living LNCs that proliferate during the induction phase of skin sensitization, whereas the LLNA: BrdU-ELISA indirectly evaluates the proliferation of lymphocytes by measuring BrdU content using an antibody-based ELISA method. This was thought to be more suitable method when compared with LLNA:ELISA. (See Validation Study Report pp 15-18)
	PR8 Yes, the relevant references were largely used. It is suggested that reference shall be given also to the GHS, CLP, and CLP guidance on application of the CLP criteria, concerning the regulatory needs for subcategorization of skin sensitizers as mentioned under A1. A more comprehensive list of Literature than the current four references in protocol version 1.3 should be included in the protocol.
	PR9 Yes, the test method biology was clearly described in terms of relevance for dermal sensitization.

Q2: Is the relationship between the test method's endpoint(s) and the (biological) phenomenon of interest described adequately using relevant references for the scientific relevance of the effect(s) measured (in terms of their mechanistic (biological) or empirical (correlative) relationship) to the specific type of effect/toxicity of interest? ***Validation principle b) The relationship between the test method's endpoint(s) and the (biological) phenomenon of interest should be described.*** This should include a reference to scientific relevance of the effect(s) measured by the test method in terms of their mechanistic (biological) or empirical (correlative) relationship to the specific type of effect/toxicity of interest. Although the relationship may be mechanistic or correlative, test methods with biological relevance to the effect/toxicity being evaluated are preferred.	PR10 **To our current knowledge the LLNA test method's endpoint, measuring potentially specific T cell proliferation, is reflecting key event four (AOP skin sensitization).** However, studies discussed in Validation Study Report in relationship to LLNA: BrdU-FCM suggest proliferative involvement of non-T cells. Authors of the study report circumvent this important observation by measuring the proliferation of lymphocytes (B and T cells) in a gated percent calculated Q2 area by FACS analysis (e.g. p 19, line 745 ff; p 22, Fig. 4). They discuss the usefulness of the test method as it can also be used to analyse B cell to T cell ratios (cell sub-typing) and cytokine content by flow cytometry and ELISA, respectively, but do refer to one pre-validation study only (Jung et al., 2012). It is completely understandable that authors have chosen this direction of FACS analysis thereby partially ignoring /including (depending on the point of view) the (unexpected) B cell observation, and not e.g. starting a principle discussion on the weakness (?) of the AOP concept that oversimplifies molecular and immunological steps in skin sensitization and elicitation phase (e.g. Karlberg AT et al., 2008, Chem Res Toxicol. 2008, 21:53-69; Martin SF, 2015, Contact Dermatitis, 72:2-10; Martin SF, 2015, Curr Opin Allergy Clin Immunol. 15:124-30). There is an additional need to address and discuss his observation more basically, even in the Validation Study Report. What does it mean in relation to our concept of a type IV allergy; independent of the usefulness of the new non-radioisotopic LLNA:BrdU-FCM?

	PR1 Yes the protocol description in the validation report was sufficiently detailed to understand what was done. Especially the revisions of the protocol during the validation period were clearly described in Table 4. This was very helpful during the evaluation of the validation study, especially the notion that the experimental protocol itself did not change, except the inclusion of an additional vehicle and more detailed descriptions in the protocol.
	PR2 The animals used are well defined (Balb/c) and the number of mice per group too. The strain of mice used is different compared to LLNA (OECD TG429) since the price is lower than CBA. The proliferation of lymphocytes in the lymph nodes adjacent to where the substances were applied is measured. BrDU is an analogue of thymidine that is incorporated into the DNA of proliferating LNC. This incorporation is then visualized by FITC-conjugated anti-BrDU Ab using FCS technique. A test substance is categorized as a skin sensitizer when its stimulation index is >=2,7. In the LLNA, the IS has to be >= 3 and a dose response is needed. It is not well explained why cells are also stained with 7-AAD. The gating for positive cells included all BrDU positive cells, even 7AAD positive cells. Does it mean that not only living cells are concerned? Line 883: the SI is calculated by dividing cells from negative group by test substance group. I think this is a mistake, SI is calculated by dividing 2 by 1. The criteria for IS >= 2, 7 instead 3 as LLNA (OCDE TG429) is well justified.

Q3: Do you consider that the protocol description in the validation report is sufficiently detailed, including a description of all the materials needed (e.g. specific cell types or construct or animal species), a description of what is measured and how it is measured, a description of appropriate data analysis, decision criteria for evaluation of data and what are the criteria for acceptable test performance? ***Validation principle c) A detailed protocol for the test method should be available.*** The protocol should be sufficiently detailed and should include, *e.g.,* a description of the materials needed, such as specific cell types or construct or animal species that could be used for the test (if applicable), a description of what is measured and how it is measured, a description of how data will be analysed, decision criteria for evaluation of data and what are the criteria for acceptable test performance.	PR3 The protocol in Annex 5 was clear and sufficiently detailed regarding the materials and animals needed, the procedures for dosing and measuring the effect, and the criterion for a positive effect. However, there were no criteria for acceptable test performance. Also, should the type of flow cytometer be specified (this is not my area of expertise)? The protocol should have been clearer on a number of other details. • The protocol was not clear on how many groups of animals are needed for the main test. In Sections 3.2 and 3.3 on page A5-5, it specifies a vehicle control group, a positive control group, and three dose groups. However, Section 3.10.1 on page A5-9 indicates that two additional groups are needed: mice not treated with BrdU (or anything else) and mice treated with BrdU only (no vehicle or test substance). In Section 4.2 of the validation study report (page 21) the positive control group is not listed. I suggest the protocol list all groups of animals needed in one place, such as at the end of the first paragraph of Section 2.1 (page A5-3) where allocation of animals to groups is first mentioned. In the main body of the report, this would correspond to the first paragraph in Section 4 on page 18 (line 703). • In the measurement of LNC proliferation the protocol is not clear about whether lymph nodes for each mouse are processed separately or whether the lymph nodes for each test group are pooled and then processed. Section 4 of the validation study report (page 18, line 703) clarifies this when it states that data on each mouse are collected; this should be in the protocol in Annex 5 as well. • Section 4.2 (page A5-12) is confusing when it states, "Individual SIs are calculated by dividing the mean number of BrdU-incorporated LNCs in the test group by the mean number of BrdU-incorporated LNCs in the negative control-treated group." How can it be an individual SI when it is calculated using mean SI values? Does it refer to the SI for a group? • Section 3.10.1 of Annex 5 (page A5-9) should also include the positive control group in the bulleted list. • Annex 1 of the protocol in Annex 5 was not labelled as such, but I assume it was titled "Pre-Screen Test" starting on page A5-14. For the Pre-Screen Test: • The objective (page A5-15) should include a statement on the characteristics of the appropriate solvent and dose range. It should be "the highest soluble dose lacking systemic toxicity or severe skin irritation" (or something to that effect). • Section 3.1 on page A5-15 is not clear for the 2nd pre-screen test. The row is titled "Dose of test substance" but there is

	other information besides dose under the 2nd pre-screen test (i.e., results of 1st pre-screen test, death or severe toxicity, and non-irritation). More text is needed to describe the information this table should convey. Also, the number of animals for each of these pre-screens should be added to this table, and "ear weight" should be added to the evaluation criteria. • Section 3.3 Dose Selection on page A5-17 is not clear. For the 1st pre-screen test, are 2 doses tested (i.e., 25% and the solvent control (0%))? The list of items at (1) and (2) is unclear. What is their significance? More text should be added here to explain the meaning of this list. Figure 3 on page 21 and the text on page 20 of the main report are much better at describing this. • The pre-screen has a criterion for toxicity (body weight decrease >5%) but there is no criterion for severe irritation. The criteria for severe irritation should be added.
	PR4 The protocol of LLNA:BrdU-FCM address a detailed information and make a clear the historical revision in the validation report. To avoid the false negative, I would like to propose the modified criteria in the protocol. Regarding Fig.6, I guess the SI value is proper at almost 1.8 as well as LLNA:BrdU-ELISA. If this SI value is used, MBT may be positive depending on the conditions (see attached file 10). Therefore, I cannot agree the VMT explanation on MBT in discussion. It is important issue to reduce false negative of strong sensitiser obtained with the LLNA and incarnate no false positives using this test method. A template for the study report is no problem. However, I cannot find the template of equipment maintenance.
	PR5 "HAS BEEN MET" Generally, the protocol is well-written. However some technical terms are better to replace more appropriate words. For example, "negative control" should be "vehicle control", because negative control means the control treated with some specific chemical expected to show negative response. "BrdU incorporation rate" in the formula for SI calculation should be "% BrdU-positive cell" or another appropriate words.

	PR6 Overall descriptions are clearly written. Suggestions to increase understanding are as below 1. Line 837; Calculation of SI; what is the difference of 'total number' and 'mean number'? 2. Line 744; Description why 7-AAD needs to be used (along with its full name) 3. The authors mention 'surface marker' as an additional benefit for this study in several places, but it does not affect the decision on skin sensitizing potential of chemical in the protocol. So I would recommend to include 'if necessary' or 'optional' etc., to make the point clear that the surface marker is not a determining factor for the protocol.
	PR7 Yes, I think this protocol was described well. The protocol showed all necessary items such as BrdU injection method, autopy method, flow cytometry condition, vehicle selection, SI calculation method, decision criteria, and so on. (See Validation Study Report pp 18-27)
	PR8 It is *not* considered that the description of how to select test concentration is sufficient. It is important to use sufficiently low concentrations to calculate the ECt accurately, and not to make wide extrapolations or rough categorization. CMI/MI is the most prominent example of alarming underestimation of the potency, owing to too high lowest test concentration. Doses below 0.5% have been omitted in the dose selection, Protocol version 1.3, although lower concentrations were used for DNCB testing throughout the validation. This omission of lower doses is not in accordance with TG429, or regulatory need (see above). No explanation to the omission is found in the validation report. It is necessary to have sufficiently high concentrations, since non-sensitization cannot be based on relatively low concentrations only. This should trigger testing at higher concentrations and/or with other vehicles. If not done, it may result in misclassification of skin sensitizers as non-sensitizers, which will pose a threat to human health.
	PR9 Yes, but more detail on the crucial step 3.8 for preparation of LNCs would be useful.

	PR10 **Comment on protocol description**. **Concerning materials** (where adequate, also applicable for Pre-Sreen-Test). - 2.2. Reagents and equipment: Several product details are missing, production company, ordering number, lot number. - 2.2. **Flow cytometry**: What type of flow cytometer was used (e.g. BD FACS Calibur™) in each single laboratory of all studies (concerning all experiments and all participating groups); what type of software was used; automated compensation? Possibly it would be of advantage to use generally a machine like the **BD Accuri™ C6 with fixed settings**! What is known about FACS settings in each laboratory participated in the studies? - 3.2. Preparation of test substances: All details of chemicals and solutions are necessary, production company, ordering number, lot number; mixing of materials (time, temperature etc.) - 3.6.4. Ear thickness: Please give information on digital **thickness gage of all laboratories**, and give detailed information on type of instrument, not company name only (compare A6-5 Transfer report, Table 1). - 3.7.3. See above: Missing information, type of 6-well plate used? - 3.8.1. See above: Give detailed information on cell strainer. - 3.9.1. Prepare reagents: Missing information – see above – should be given on Perm/Wash Buffer (ordering / lot number) DPBS, FBS, sodium azide. - 3.10. (2) 7-AAD-BrdU graph: Clarify if isotype control of BrdU antibody is used? 4. Results: 7-AAD-BrdU graph and 4.2. Calculation of the SI: Please add sentence how to analyse and understand / deduce information from 7-AAD-BrdU-graph (e.g. 7-AAD- high cells, dying cells?)

Q4: Do you consider that the intra-, and inter-laboratory reproducibility of the BrdU-FCM test method is adequately demonstrated, considering availability of the data over time, as well as the degree to which biological variability affects the test method reproducibility? ***Validation principle d)The intra-, and inter-laboratory reproducibility of the test method should be demonstrated.*** Data should be available revealing the level of reproducibility and variability within and among laboratories over time. The degree to which biological variability affects the test method reproducibility should be addressed.	PR1 The within-laboratory reproducibility (WLR) was assessed testing HCA four times in three labs. The WLR was evaluated using protocol 1.0. Since revisions made in the protocol mostly concern solubility issues related to other chemicals, e.g. the addition of HEK as an additional vehicle, results obtained with protocol 1.0 are assumed to be representative for the final protocol as well. The WLR was based on the EC2.7 obtained from the dose-response curves. According to the performance standards of OECD TG 429, the EC2.7 needs to be in the range of 5-20% for HCA and this was achieved in all laboratories for all repetitions. WLR is adequately demonstrated. The between-laboratory reproducibility (BLR) was assessed testing HCA and DNCB, as recommended in the performance standards. The BLR was evaluated using protocol 1.0. Since revisions made in the protocol mostly concern solubility issues related to other chemicals than DNCB and HCA, e.g. the addition of HEK as an additional vehicle, results obtained with protocol 1.0 are assumed to be representative for the final protocol as well. The BLR for DNCB and HCA were within the predefined range of the performance standards. BLR is adequately demonstrated.
	PR2 The tests using HCA and DNCB were well conducted and the reproducibility interlaboratory is really quit good for the both compounds selected (HCA and DNCB). The WLR is good too but less that the BLR.
	PR3 Yes. The intra-laboratory reproducibility of the LLNA: BrdU-FCM was adequately demonstrated, according to the performance standards for TG 429, with 4 tests of HCA in three laboratories. Using SI $\geq$ 2.7 as the criterion for a sensitizer result, all tests were within the range specified by the performance standards (i.e., 5-20%). In addition, the laboratories also tested DNCB three times each. The EC2.7 values for DNCB also met the criteria specified in TG 429 (0.025-0.1%) for inter-laboratory reproducibility. It would have been helpful to provide the dates of these tests so we could verify that the tests for each chemical were at least one week apart in each of the labs. The inter-laboratory reproducibility of the LLNA: BrdU-FCM was adequately demonstrated with 4 tests of HCA and 3 tests of DNCB in three laboratories. The EC2.7 values met the criteria established in TG 429. This was more testing than prescribed in the performance standards, which recommended only one test per laboratory in 3 laboratories.

	PR4 According to performance standard in TG429, I think it is high intra-, and inter-laboratory reproducibility of the LLBA:BrdU-FCM test method. There is a little variation of negative and positive substance (see attached file 8). In the 2nd test, however, LA, IU and MBT were tested by two or three laboratories. It would be helpful to justify these chemicals used to assess the intra-laboratory reproducibility. Furthermore, the possible impact on the inter- and intra-laboratory reproducibility of the changes made to the protocol between the 1st and 2nd test should be discussed.
	PR5 INSUFFICIENT" or "HAS PARTLY BEEN MET" The intra-, and inter-laboratory reproducibility were evaluated with very limited chemicals. Reproducibility and transferability should be evaluated with the test substances covering a broad range of activity, especially including non-sensitizer. So it recommended to make additional analysis. Hopefully, they should cover the strong, moderate, weak sensitizers and negative chemical.
	PR6 Yes. The overall WLR and BLR of the test is acceptable, and the data in Annex 8 and 9. However, it should be more clearly described (or the demonstrating data should be included in page 31) in which basis the WLR was not required for Lead Lab 2.
	PR7 Yes, I think this intra- and inter-lab validation demonstrated good reproducibility. Evaluation of the WLR was accomplished by tests repeated four times at intervals of more than one week at Lead Laboratory 1 and Participating Laboratories 1 and 2 using 5%, 10%, and 25% HCA, based on OECD TG 429 Annex 1 Paragraph 8. BLR was evaluated at Lead Laboratory 1 and Participating Laboratories 1 and 2 using 0.05%, 0.1%, and 488 0.25% DNCB, as described in OECD TG 429 Annex 1 Paragraph 9. As for HCA, the results obtained during the WLR test were used for the BLR evaluation. Based on a statistical analysis of results from the WLR, BLR, and predictive capacity evaluations, those results improved using an SI threshold of 2.7 rather than at 3.0. (See Validation Study Report pp 31-33)
	PR8 Yes.
	PR9 Yes, as defined by previous criteria the intra- and inter-laboratory reproducibility was adequately demonstrated. However, the coefficient of variation in one lab was over 50% and the source of the variability should be defined.

	PR10 **Availability of the data over time**. Original FACS data are not shown. It is not clear what type /software of cytometer was used in each laboratory, and how compensation was performed. **Degree of biological variability / test method reproducibility**. Information of degree of biological variability is given by original data (e.g. single SI values).

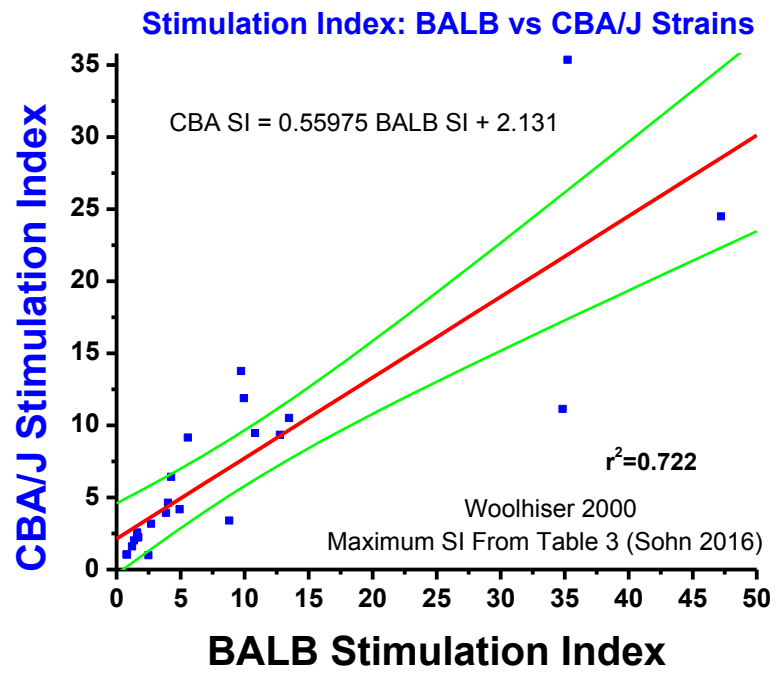
Q5: Are the reference chemicals used to demonstrate the performance of this test method representative of the types of substances for which the test method will be used and have they been tested under code to exclude bias? ***Validation principle e) Demonstration of the test method's performance should be based on the testing of reference chemicals representative of the types of substances for which the test method will be used.*** A sufficient number of the reference chemicals should have been tested under code to exclude bias (see paragraphs on "Coding and Distribution of Test Samples").	PR1 The predictive performance was assessed for all protocols. Since protocol 1.3 is the final protocol, I have only evaluated the predictivity obtained with this protocol. The performance was assessed using the 18 reference chemicals listed in the performance standards of TG429. All chemicals were tested blindly under code in one of the participating labs. Table 11 shows the results of the test performed according to protocol 1.3. Compared to the LLNA, this assay has a sensitivity of 84.6% and a specificity of 100%. The results show that compared to the LLNA TG429, the BrdU-FCM LLNA missed two reference skin sensitizers. According to the validation report, these were weak or borderline chemicals. It is difficult to generalize these results, since only 18 substances were tested. To conclude, the performance was evaluated according to the performance standards, but the values for accuracy and sensitivity need careful interpretation, since they are based on only a small set of substances. However, testing more reference chemicals is not preferred from an animal welfare perspective. Table 11 summarizes the results of the third test. The predictive performance for hazard identification of skin sensitizers is clear from this table. What is less clear is the performance of the assay to predict potency (EC value). This table provides both the LLNA TG429 EC3 range as well as the EC2.7 range of the LLNA:BrdU-FCM. For certain chemicals, this is not a range but a value that is expressed as <2.5, for example for CMI/MI. It is unclear what this means and an explanation is needed. Furthermore, not all of the EC2.7 values were within the range recommended by the performance standards. The EC2.7 for the following substances are outside the range: CMI/MI, DNCB, cobalt chloride, cinnamic alcohol. For isoeugenol it is unclear, the EC2.7 is <5, but this may be outside the range (0.77-3.1). These discrepancies are not discussed at all in the validation report. Although this does not have an impact on the final result to classify a substance, it may indicate some variability in the test method. On page 38, the second paragraph discusses imidazolidinyl urea. The EC2.7 value was 26.8%, but in Table 11 it was 32.0%. There appears to be a mistake here. On page 38 the performance of this test to predict potency is described. This part is difficult to read and impairs the evaluation. According to the validation report, the test method was able to sub-categorization all skin sensitizers, except two, in the correct GHS 1A and 1B subcategories. It is not clear from the text which ones were incorrect. I guess one of them is MCI/MI. It is therefore not possible to evaluate the performance of this test method for potency prediction. As mentioned above, this is an important aspect that may

	lie outside the scope of this peer-review, but needs attention in the next steps towards inclusion in the TG. One remark that is related to the revision in the EC value, that was set at 2.7 rather than 3.0 based on a statistical analysis on a set of 18 chemicals. I am not an expert on this, but I wonder if you could set such a definitive cut-off using data of only 14 skin sensitizers. It may be good to let a statistician evaluate this in more detail, but this may beyond the scope of this peer-review and should be discussed during the evaluation of the TG.
	PR2 18 known chemicals have been tested and are representative of sensitizers or non-sensitizers (13 sensitizers and 5 non-sensitizers). In the group of sensitizers, some are categorized strong and other weak.
	PR3 Yes. The performance of the test method was evaluated using the reference chemicals prescribed in the performance standards of TG 429. These chemicals represent the types of substances typically tested for skin sensitization potential and the range of responses that the LLNA is capable of measuring or predicting. The reference chemicals tested to evaluate predictive capacity were tested under blind conditions. The labs did not know the identity of the chemicals.
	PR4 I guess the predictive capacity of LLNA:BrdU-FCM is similar with LLNA:Brd-ELISA. However, it is too small and lack of reliability for the dataset based on the reference chemicals of this assay compared with other test methods (see table 14). More 20 data need to evaluate predictive capacity as well as other LLNA related test methods. Developers should be performed the additional study using extreme, strong, moderate sensitiser and non-sensitiser soon.
	PR5 “HAS PARTLY BEEN MET” Selection of reference chemicals for Module 5 was appropriate, and they were distributed under coded manner to prevent bias. However the chemicals used for evaluating WLR and BLR were not covering sufficient variation of chemical activities.
	PR6 Yes, the reference compounds were from existing LLNA methods. One concern is about the vehicle control. It is not sure that the extent or reproducibility of the test results from vehicle controls are acceptable. Although the data for AOO is clearly included in WLR or BLR, the results for other vehicle is not clearly shown. For example, they added MEK as a potential vehicle during revision of the protocol, but it is not sure that they conducted tests for MEK according to the Table 13. Please make this points clear.

	PR7 Yes, reference chemicals were well represented all types of substances, and coded. KoCVAM coded and distributed the test substances for the predictive capacity evaluation. Chemical names were unveiled by the chemical manager after each phase was completed, and raw data were sent to the biostatistician. (See Validation Study Report p 15)
	PR8 Yes, the reference chemicals listed in OECD TG429 have been used and tested under coded conditions.
	PR9 Yes, the positive controls are known dermal sensitizers and the testing was stated to be blinded.
	PR10 **Reference chemicals.** Coding was given according ANNEX 3.
	PR1 The performance is based on a comparison to the LLNA, since this concerns a me-too assay. In addition, the substances that gave discordant results were discussed in more details and additional evidence from literature was provided. This additional information included human data as well, which is the species of concern. So, the performance of the test was evaluated using relevant existing data.
	PR2 The performance of the test method has been well evaluated in relation to proliferation without any toxicity of chemicals. The selection of doses in order not to induce toxicity is well explained.
	PR3 Yes. The performance of the test method was evaluated with respect to the performance standards documented in TG 429. The performance standards were developed in order efficiently evaluate the reliability and accuracy of tests that are mechanistically and functionally similar to the LLNA. The relevant species of concern is humans. The performance standards reference substances were selected based on a number of criteria, including “known” (actually presumed for a few chemicals) responses in humans, mice, and guinea pigs. Previous mouse (LLNA) and guinea pig data also serve as benchmarks for new or revised LLNA “me-too” methods. The validation study also evaluated their results in the context of newer data for the performance standards reference substances that was published subsequent to the performance standards.

Q6: Do you consider that the performance of the test method has been evaluated in relation to relevant information from the species of concern, and existing relevant toxicity testing data? ***Validation principle f) The performance of the test method should have been evaluated in relation to relevant information from the species of concern, and existing relevant toxicity testing data.*** In the case of a substitute test method adequate data should be available to permit a reliable analysis of the performance and comparability of the proposed substitute test method with that of the test it is designed to replace.	PR4 In the revised TG442B, we must avoid to operate on same level both the LLNA:BrdU-ELISA and LLNA:BrdU-FCM. There are no data of ELISA with Balb/c and no data of FCM with CBA. In addition, if we accept Balb/c as animal species in TG442B, we should not forget it is correlated to be revised TG429 in the future. We have to be careful to handle this issue. Anyway, I recommend the panel members that TG442B should be revised as a performance based test guideline (PBTG) and each method is attached as an annex as well as TG455.
	PR5 "HAS PARTLY BEEN MET" As the results of comparison with reference standard, the predictive capacity of this method is shown as sensitivity: 84.6%, specificity: 100%, and overall accuracy: 88.9%, accordingly the performance of this method is successful and acceptable. Although this method uses BALB/c strain mice which is not listed in the recommended mouse strains for standard LLNA, published data shows its equivalency of this strain to CBA mice. So it is considered to have no impact on the validation status. However WLR and BLR could not be evaluated sufficiently because of limited number of test chemicals.
	PR6 Yes, this is about LLNA which is a well-established protocol
	PR7 Yes, I think the performance of the test method has been evaluated well in relation to relevant former information of LLNA. (See Validation Study Report pp 36-37)
	PR8 It is considered that the performance of the test method has been evaluated in relation to relevant information from mice. It is considered that human data on contact allergy also should be mentioned in the discussion of false-negative test results (page 37). There is no doubt about mercaptobenzothiazole, methyl methacrylate and imidazolidinyl urea being skin sensitizers in humans. They belong to the most frequent causes of skin sensitization. As such, they are part of International, European and national series for patch testing, and are regularly used to diagnose contact allergy. Mercaptobenzothiazole is also a strong sensitizer in the GPMT (Frankild et al. Comparison of the sensitivities of the Buehler test and the guinea pig maximization test for predictive testing of contact allergy. Acta Derm Venereol 2000;80(4):256-62). It is considered that additional test concentrations should have been used for some of the substances. Valuable information on dose-response and skin sensitizing potency was missed, which is a waste of animals and resources.

	PR9 Yes, the performance was demonstrated according to previously defined criteria, however, other types of validation criteria could have been considered (i.e. Z' equation). Additional Comments: The study had acceptable sensitivity and specificity. However, these metrics could be improved and in some cases the coefficient of variation was over 50%. The method in BALB mice missed some known sensitizers. In the Sohn 2016 paper, the CBA/J strain seems to be more sensitive than the BALB mice in the weak sensitizers (see graph below). Could the SI regression curve below be utilized to correct the BALB SI results to the CBA/J strain SI values to improve the sensitivity of the method? According to the Ahn 2016 paper, mercaptobenzothiazole, methylmethacrylate, isopropanol and nickel chloride are sensitizers in humans, which were incorrectly identified by the new method using the BALB mouse, not the CBA mouse (SI less than 2.7, see Max SI in table 6 of the validation report). The BALB mouse is less sensitive in general than the CBA mouse for weak sensitizers, so these chemicals were incorrectly classified. Can the regression equation in the graph below be used to calculate the CBA mouse SI from the BALB SI result: CBA SI = 0.56 x BALB SI + 2.1? This equation based on the Sohn and Woolhiser papers data implies that if the BALB SI is over 1.0, then the corresponding CBA SI is over 2.7, thus a sensitizer. With this data correction, mercaptobenzothiazole, methylmethacrylate, isopropanol and nickel chloride have corresponding "CBA SI" values over 2.7 and correctly classified as sensitizers, thus would substantially increase the method sensitivity relative to humans. What is the current BALB mouse method SI value correlation to the original guinea pig SI values? Does dermal sensitization or irritation increase a compound's dermal absorption, thus internal dose? The multi-channel flow cytometry method is more complex, but allows simultaneous determination of other metrics (i.e. cell cycle). Could an in vitro assay replace the pretesting for concentration ranges determination? In my experience designing antibody based assays, the variability in the antibodies from lot to lot are a main factor in assay responses, so antibody QC might be an area for additional metrics (i.e. FITC labelling/mg protein, BrdU labelling efficiency, antibody titer, etc). Do any of the test articles have absorbance or fluorescence that might interfere with the flow assay (i.e. FITC or 7-AAD)?

	
	PR10 **Performance of the test method has been evaluated in relation to relevant information**. However, original FACS data are not given but it would be advantageous to have them; including information on FACS software analysis.
	PR1 All data obtained in the lead and participating laboratories was obtained in accordance with GLP.
	PR2 The data supporting the validity of a test method have been obtained in accordance with the good principles of GLP. All is defined in the annexe.
	PR3 Yes. The data supporting the validity of the LLNA: BrdU-FCM were obtained according to GLP principles in the case of every participating laboratory. This was documented on pages 10 and A1-7 of the validation study report.

Q7: Have all of the data supporting the validity of a test method been obtained in accordance with the principles of GLP? If not, has an adequate consideration been given to the potential impact on the validation status of the test method? ***Validation principle g) Ideally, all data supporting the validity of a test method should have been obtained in accordance with the principles of GLP.*** Aspects of data collection not performed according to GLP should be clearly identified and their potential impact on the validation status of the test method should be indicated.	PR4 I believe this validation study was conducted under the GLP. The whereabouts of all law data, record sheets, the record of equipment maintenance should be cleared. If possible, the statement of QAU at each laboratory should be attached in the validation report.
	PR5 HAS PARTLY BEEN MET" The data collected during the validation phase were recognized as obtained under GLP. However Raw data such as the worksheets derived from the study and the certifications of QAU inspections would be preferable if archived in accessible space. The other results in pre-validation phase and comparison of mouse strain are not specified if they are derived under GLP condition but these had been published in the peer-reviewed scientific journal. So It is considered to have no impact on the validation status.
	PR6 Yes. I agree that the validation of the test was conducted in accordance with the principles of GLP.
	PR7 Yes I think all data was obtained in accordance with the principle of GLP (QAU audit). The most of paticipating Laboratory had GLP facility approved by the regulatory agencies, and conducted test validation study under QAU inspection. (See Validation Study Report pp 9-11)
	PR8 Yes, GLP is considered to have been followed.
	PR9 Yes, the participating laboratories claimed to follow GLP or equivalent practices.
	PR10 **FACS data: original FACS data are not given but it would be advantageous to have them**.
	PR1 The validation study was in general very complete and clear and it adhered to the existing performance standards. There were some discrepancies noted (see above). The only part that was difficult to evaluate concerned the potency prediction (see above).
	PR2 The data supporting the validity of a test method have been obtained in accordance with the good principles of GLP. All is defined in the annexe.

	PR3 Yes, the individual animal data, as well as summary data, are presented in the validation study report for the evaluations of WLR, BLR, and predictive capacity. The individual animal and summary data provide benchmarks by which an independent laboratory could judge its performance with this method. The protocol is provided as Annex 5 to the validation study report, which is publically available on OECD's website, but as I mentioned under A3, more detail is recommended.
	PR4 I found it is incompatibility between the tables in the report and the attached files 9 & 10. Please revise the validation report in accordance with the following. Different SI values between table 13 and 14, and data at the attached files 9 & 10. Several misclassifications at the attached file 10. Should be deleted the in vitro data in Tables 14 and 15. The in vitro data are no need for comparison with the SI values on the LLNA related methods.
	PR5 "HAS PARTLY BEEN MET" Outcomes of the validation study is available in the report. Raw data such as the worksheets derived from the study and the certifications of QAU inspections would be preferable if archived in accessible space.
	PR6 Yes, it would be better if some parts are improved in terms of English grammar and style, to increase the understanding and adherence.
	PR7 Yes, I think all data are easily available for expert review. These include: a detailed and readily available test method protocol to the public and independent laboratories in ANNEX 5; organised and easily accessible data to permit independent review(s) in ANNEX 7-10;

	PR8 It is considered that the terminology and definitions of classification categories and subcategories need to be consistent and explained, to facilitate understanding and use. Regulations, policy documents and scientific publications have, over time, used varying terminology and definitions. It is suggested that the categories of GHS and CLP are used, and that the ECETOC terminology is mentioned. This could be done in a list or table of definitions. It is not explained why concentrations lower than 0.5% were omitted. This is an important change that affects the test method negatively. In Protocol version 1.3, page A5-5, and A5-17 under 3.3 Dose selection it is written that "Doses of 100%, 50%, 25%, 10%, 5%, 2.5%, 1%, and 0.5% should be used according to OECD TG 429." This is not accurate. Also lower doses shall be considered according to OECD TG 429. The test method protocol version 1.3 is generally sufficiently detailed and clear. It is considered helpful with the explanatory drawings. Interpretation of results (para 4.3): only classification as sensitizer and non-sensitizer are mentioned. Subcategorization into 1A and 1B is not mentioned or indicated. It is considered that subcategories should be explained since GHS and CLP need subcategorization, and CLP does further require identification of extreme skin sensitizers (EC3 ≤0.2, or corresponding ECt). A list of definitions is lacking in Protocol version 1.3. Benchmark substances are available. It is suggested that OECD TG 429 table 1 is included in the protocol, as bench mark.
	PR9 Yes, the report considered here was freely available on the OECD website and contained enough information for experts in the field to reproduce the experiments.

Charge Q8: Do you consider that all the data supporting the assessment of the validity of the test method are easily available for expert review? These include: a detailed and readily available test method protocol to the public and independent laboratories; organised and easily accessible data to permit independent review(s); benchmarks by which an independent laboratory can itself assess its proper adherence to the protocol. ***Validation principle h) All data supporting the assessment of the validity of the test method should be available for expert review.*** The detailed test method protocol should be readily available and in the public domain. The data supporting the validity of the test method should be organised and easily accessible to allow for independent review(s), as appropriate. The test method description should be sufficiently detailed to permit an independent laboratory to follow the procedures and generate equivalent data. Benchmarks should be available by which an independent laboratory can	PR10 **Considering that all the data supporting the assessment of the validity of the test method are easily available for expert review**. Not all data are easily available for expert review. For example original FACS data are not shown but stimulation indices. Protocol is given that an independent laboratory can itself assess its proper adherence to the protocol.

itself assess its proper adherence to the protocol.	

Additional Comments:

2016-11-07 Carola Lidén

Dear Julija,

Hereby some background and explanation to my concerns.

As member of the EC Scientific Committee on Consumer Safety (SCCS, and its predecessor SCCP), I have reviewed hundreds of LLNA test protocols on cosmetic ingredients (fragrances substances, hair dye substances, preservatives). A frequent finding is that many studies submitted by industry were performed with too low concentrations, so that the results were inconclusive, although interpreted by industry as negative. Submitted studies were also performed with too high concentrations, so that EC3 values not could be calculated due to wide extrapolations.

Classification as non-sensitizer cannot be based on low test concentrations only. This was stressed in references in TG 429: Kimber et al 2006 and in the validations by ESAC/ECVAM and ICCVAM (based on Gerberick et al 2005). Since then, additional compilations of data have been published (Kern 2010, Lidén 2016).

Classification of skin sensitizers as non-sensitizing is a potential threat to human health. Regulations (GHS and CLP) require identification of skin sensitizers and of non-sensitizers, and subcategorization of classified sensitizers into 1A and 1B. CLP further needs identification of extremely potent sensitizers for down-stream information.

Examples from the validation report

In the validation report some substances (particularly salicylic acid and isopropanol) were repeatedly categorized as non-sensitizer, based on results with low test concentrations only. These examples are not "borderline" results. They represent categorization based on insufficient testing. (*I'm fully aware that these substances are non-sensitizers, but such a conclusion could not be drawn, based on the low concentrations tested*.) (Salicylic acid was tested in AOO with highest conc 25%, and isopropanol in AOO with highest conc 50%, according to the compilation by Gerberick et al 2005.) Please, see the following examples (also in tables in Annex 11):

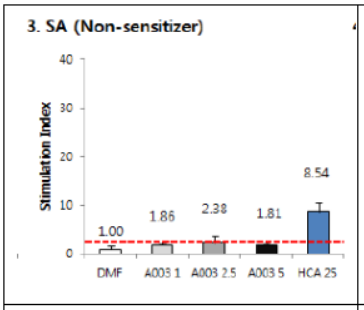


Fig. 10-1 (highest conc: 5%)

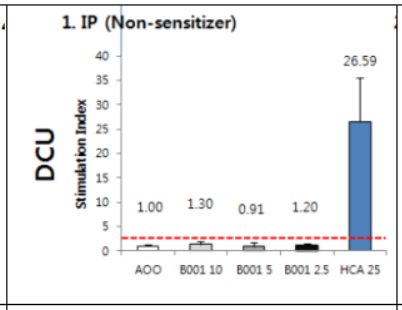


Fig. 10-2 (highest conc: 2.5%)

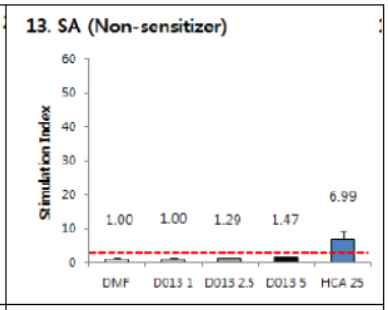


Fig. 10-5 (highest conc: 5%)

The lack of sufficiently low test concentrations for potency identification and subcategorization is demonstrated by these two figures from the 3rd test:

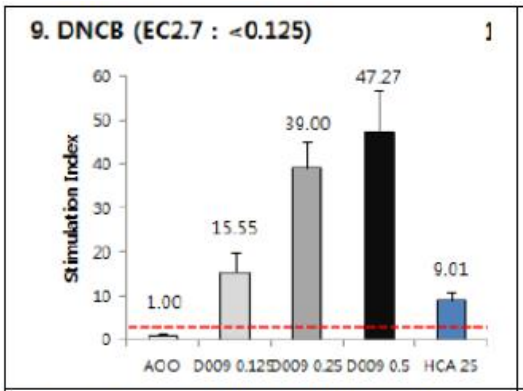


Fig. 10-5 (TG 429: EC3 0.049)

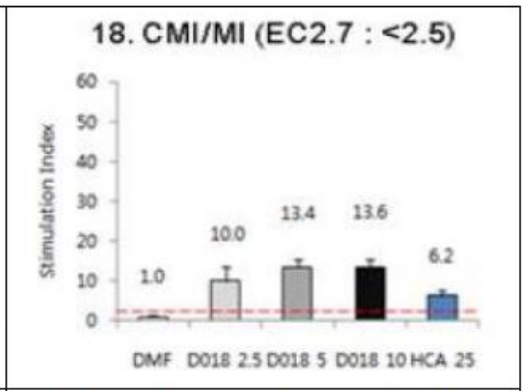


Fig. 10-5 (TG 429: EC3 0.009%)

Why is this a larger problem in the LLNA: BrdU-FCM than in other LLNA tests?

My interpretation is that the current LLNA: BrdU-FCM has focused too much on restricting vehicles and test concentrations, when aiming to reduce the number of required animals. Some important differences between LLNA: BrdU-FCM and OECD TG 429, 442A, 442B are underlined. Pre-screen test, 3.2 Selection of vehicles (page A5-17) states: "To find vehicles that best dissolve test substances, vehicles are selected based on solubility tests. Perform solubility tests as described in 3.2.1 to prioritize vehicles (AOO, DMF, MEK, DMSO) in the order of solubility."

Protocol version 1:3 (page A5-5-6) states:

3.2 Preparation of test substances "Acetone:olive oil (4:1 v/v, AOO), N,N-dimethylformamide (DMF), methyl ethyl ketone (MEK), and dimethyl sulphoxide (DMSO) should be used as vehicles, as suggested in OECD TG 429. If vehicles other than those described in OECD TG 429 are used, the scientific rationale should be given for the selection of the vehicles."

3.3 Dose selection

"Doses of 100%, 50%, 25%, 10%, 5%, 2.5%, 1%, and 0.5% should be used according to OECD TG 429."

OECD TG 429

"18. Dose and vehicle selection should be based on the recommendations given in references (3) and (5). Consecutive doses are normally selected from an appropriate concentration series such as 100%, 50%, 25%, 10%, 5%, 2.5%, 1%, 0.5%, etc. Adequate scientific rationale should accompany the selection of the concentration series used…"

"19. The vehicle should not interfere with or bias the test result and should be selected on the basis of maximising the solubility in order to obtain the highest concentration achievable while producing a solution/suspension suitable for application of the test substance. Recommended vehicles are acetone: olive oil (4:1, v/v), N,N-dimethylformamide, methyl ethyl ketone, propylene glycol, and dimethyl sulphoxide (19) but others may be used if sufficient scientific rationale is provided."

"28. In certain situations, when there is a regulatory need to confirm a negative prediction of skin sensitizing potential an optional rLLNA protocol (16) (17) (18) using fewer animals may be used, provided there is adherence to all other LLNA protocol specifications in this TG. Before applying the rLLNA approach, clear justifications and scientific rationale for its use should be provided. If a positive or equivocal result is obtained, additional testing may be needed in order to interpret or clarify the finding."

"29. The reduction in number of dose groups is the only difference between the LLNA and the rLLNA test method protocols and for this reason the rLLNA does not provide dose-response information. Therefore, the rLLNA should not be used when dose-response information is needed. Like the multidose LLNA, the test substance concentration evaluated in the rLLNA should be the maximum concentration that does not induce overt systemic toxicity and/or excessive local skin irritation in the mouse (see paragraph 18)."

Proposals

If only low concentrations have been tested, and no EC3 value can be calculated further testing with other vehicles with higher solubility should be performed.

If only too high concentrations have been tested, so that an EC3 value cannot be accurately calculated, further testing with lower concentrations should be performed.

I assume that more animals would be saved if a requirement for further testing is introduced when no ECt value is achieved with low concentrations, than requiring that all tests shall contain a relatively high test concentration (based on Gerberick 2005 and Kern 2010: at least 20%). A suggestion to perform a reduced LLNA (rLLNA) with one high concentration with another vehicle (higher solubility) might possibly be an alternative.

Additional clarification of the definitions regarding the different classification requirements: 24-11-16:

that this point should be explained in more detail, to support understanding of what is meant by the recommendation. Cut-off values differ, and ECETOC is using the concept "weak" which isn't used in GHS or CLP. I have compiled a table over the terminology and cut-off values that might fit in Annex 7:

	EC3 value (%)	Relative potency	Hazard Cat. 1, sub-cat.	Ref.
GHS	≤2	strong	1A	1
	>2	other	1B	
CLP	≤0.2	extreme	1A	2
	>0.2 - ≤2	strong	1A	
	>2	moderate	1B	
ECETOC	<0.1	extreme		3
	≥0.1 - <1	strong		
	≥1 - <10	moderate		
	≥10 - ≤100	weak		
GHS, CLP, ECETOC	No EC3 value	Non-sensitiser		1-3

1. Globally Harmonized System of Classification and Labelling of Chemicals (GHS). Sixth revised edition. 2015

2. Guidance on the Application of the CLP Criteria. Guidance to Regulation (EC) No 1272/2008 on classification, labelling and packaging (CLP) of substances and mixtures Version 4.1 2015

3. Contact sensitisation: classification according to potency. Technical Report No. 87. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium. 2003

Annex 4 - KoCVAM power point presentation at the teleconference 11-10-16

1st Teleconference

Local Lymph Node Assay:
5-bromo-2-deoxyuridine-flow cytometry method
(LLNA: BrdU-FCM)

Validation Study

Oct. 2016

Ministry of Food and Drug Safety

Republic of Korea

1

Introduction

- The LLNA: BrdU-FCM was developed following the Performance Standards in OECD TG 429 Annex I as a modified test method of the LLNA.
- Advantages of the LLNA: BrdU-FCM:

1. The proliferation of live-LNCs is quantitatively measured in the LLNA: BrdU-FCM, whereas proliferated cells are indirectly scored based on BrdU content, regardless of whether these cells are alive or dead in the LLNA: BrdU-ELISA.
2. Flow cytometry allows for simultaneous analysis of multiple parameters without sacrificing extra animals (e.g. B/T cell ratio, activation surface marker (CD86 etc)), which could help to understand the skin sensitization mechanism.
3. The modified pre-screen test in the LLNA: BrdU-FCM can reduce the number of animals and minimize pain and distress in animals by applying a refined dose selection scheme.
4. BALB/c mice can be used in LLNA: BrdU-FCM. BALB/c mice are widely used since they are easy to be obtained and cost-efficient in some countries more than CBA mice.

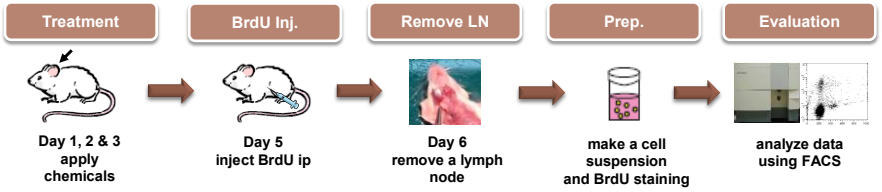
2

Background of Validation Study

- The validation study on the LLNA: BrdU-FCM has been performed in accordance with the OECD TG 429 PSs for the first time.
- The validation study coordinated by KoCVAM started in 2012 in accordance with GD No. 34, and it was presented at the OECD Expert Meeting in 2015.
- SPSF for a new project in TGP work plan was approved at 2016 WNT meeting
  - Revise TG 442B in order to include the LLNA: BrdU-FCM
- The modular approach was applied in this validation study.
  - Test definition, Transferability, Within-laboratory reproducibility, Between-laboratory reproducibility, Predictive capacity
- Predictive capacity of the LLNA: BrdU-FCM has been evaluated using coded test chemicals.

3

Protocol



▶LLNA: BrdU-FCM main test schedule

Experiments	Day 0	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Grouping	O*	O*						
Clinical observation	O	O	O	O	O	O	O	
Measurement of body weight		O					O	
Measurement of ear thickness		O		O			O	
Irritation evaluation		O	O	O	O	O	O	
Treatment		O	O	O				
BrdU solution injection						O		
Sacrifice							O	
Measurement of ear weight							O	
Measurement of lymph node weight							O	
BrdU staining							O	●
Analysis with flow cytometry							O	●

BrdU, 5-bromo-2-deoxyuridine; ●: possible to analyse samples , *: possible to group experimental animals

4

Protocol

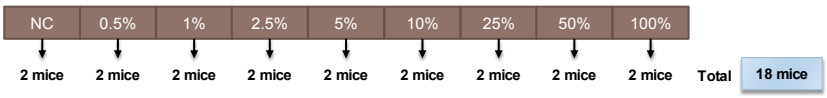
Groups

- Blank group: No BrdU injected or treatment with test substances
- Non-treatment group: Injection of BrdU but no treatment with test substances
- Vehicle control-treatment group: Injection of BrdU and treatment with vehicle
- Test substance-treatment group: Injection of BrdU and treatment with test substances
- Positive control-treatment group: Injection of BrdU and treatment with 25% HCA

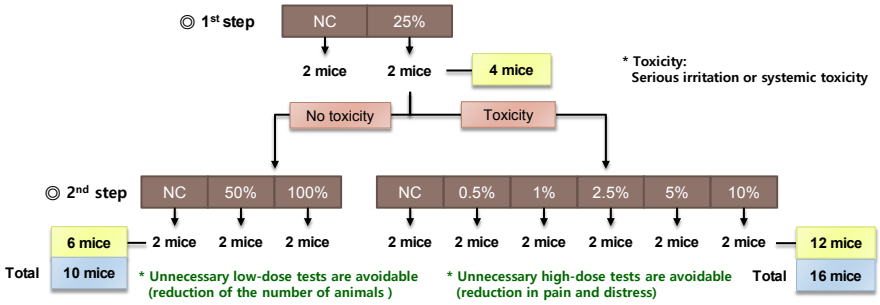
5

Protocol

- OECD TGs 429, 442A and 442B



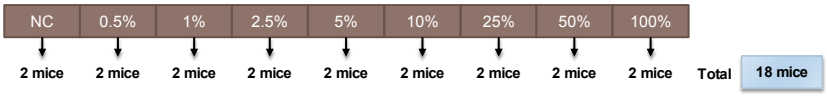
- LLNA: BrdU-FCM



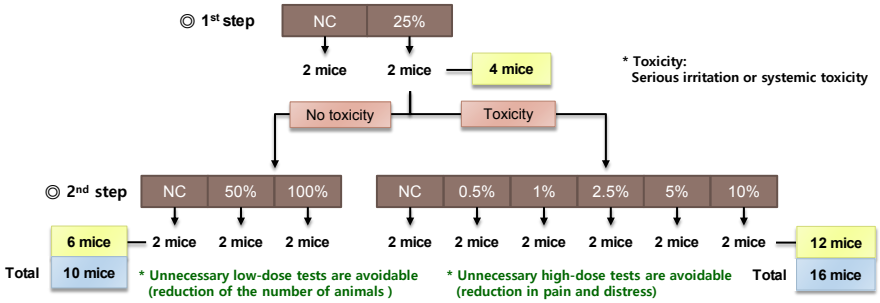
6

Protocol

- OECD TGs 429, 442A and 442B



- LLNA: BrdU-FCM

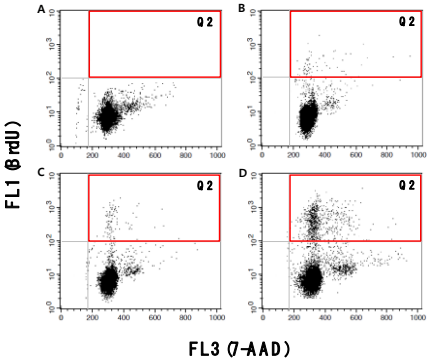


7

Protocol

Analysis of data

- The FITC BrdU Flow Kit (Cat. No. 559619, BD Pharmingen™) was used for BrdU staining.
- Set up 7-AAD - BrdU graph (Q2 area)



A: Set up the Q2 area (upper right) using blank samples so that it contains no cells.

B: Set up the Q2 area again using non-treatment samples such that the BrdU-incorporated LNCs represent about 1% of the cells in it.

C and D: Analyse vehicle control-treated samples and test substance-treated samples, and calculate the gated percent in the Q2 area that was set up in steps A and B.

8

Essential Test Method Elements

• The test substance should be applied topically to both ears of the mouse.	yes	Line 698, 721 Page A5-5
• Lymphocyte proliferation should be measured in the lymph nodes draining from the site of test substance application;	yes	Line 161, 237, 626, 698 Page A5-7
• Lymphocyte proliferation should be measured during the induction phase of skin sensitization;	Yes	Table 3 (page 19)
• For test substances, the highest dose selected should be the maximum concentration that does not induce systemic toxicity and/or excessive local skin irritation in the mouse. For positive reference substances, the highest dose should be at least as high as the LLNA EC3 values of the corresponding reference test substances (see Table 1) without producing systemic toxicity and/or excessive local skin irritation in the mouse;	yes	Line 700
• A concurrent VC should be included in each study and, where appropriate, a concurrent PC should also be used;	yes	Line 702
• A minimum of four animals per dose group should be used;	yes	Line 702
• Either individual or pooled animal data may be collected	yes	Individual animal data

10

Results: Module 2 WLR

- WLR at the three laboratories was evaluated (protocol 1.0).
  - Lead laboratory 1 and participating laboratory 1 and 2
  - Since all ECt values for HCA were in the range of 5%~20% in four repeated tests performed by the three laboratories, the VMT concluded that WLR of the test method was demonstrated.

		Stimulation Index for each HCA concentration (%)			EC2.7 (%)	Pass/ failure for WLR
		5	10	25		
Lead lab 1	Run 1	1.8	3.6	10.3	7.405	Pass
	Run 2	2.3	3.1	9.2	7.344	Pass
	Run 3	1.7	4.5	13.9	6.805	Pass
	Run 4	1.8	3.6	9.6	7.378	Pass
Participating lab 1	Run 1	2.5	3.9	9.7	5.878	Pass
	Run 2	1.8	3.3	4.7	9.072	Pass
	Run 3	1.7	3.4	5.4	8.668	Pass
	Run 4	1.8	4.0	7.1	6.955	Pass
Participating lab 2	Run 1	1.4	2.4	6.8	10.284	Pass
	Run 2	1.8	4.7	8.1	6.041	Pass
	Run 3	1.3	3.5	7.9	8.502	Pass
	Run 4	1.8	3.1	7.6	8.338	Pass

11

Results: Module 4 BLR

- After the WLR tests, BLR at the three laboratories was evaluated (protocol 1.1.)
  - Lead laboratory 1 and participating laboratory 1 and 2
  - Results with 5%, 10% and 25% HCA obtained in the WLR were further utilized to assess BLR.
  - Since all ECt values for HCA and DNCB were in the range of 5%~20% and 0.025%~0.1%, respectively in four or three repeated tests performed by all laboratories, the VMT concluded that BLR of the test method was demonstrated.

		Stimulation Index for each DNCB concentration (%)			EC2.7 (%)	Pass/ failure for BLR
		0.05	0.1	0.25		
Lead lab 1	Run 1	1.5	3.0	13.7	0.080	Pass
	Run 2	0.9	2.0	10.5	0.098	Pass
	Run 3	1.5	6.1	19.5	0.063	Pass
Participating lab 1	Run 1	1.8	1.8	8.3	0.099	Pass
	Run 2	1.1	3.7	10.0	0.082	Pass
	Run 3	2.8	4.4	20.5	0.063	Pass
Participating lab 2	Run 1	1.3	2.6	16.2	0.082	Pass
	Run 2	2.1	3.6	13.2	0.070	Pass
	Run 3	1.9	4.0	16.1	0.070	Pass

12

Results: Module 5 Predictive Capacity

- 18 coded essential reference chemicals of PSs (protocol 1.3).
  - Sensitivity: 84.6%, Specificity: 100%, Overall accuracy: 88.9%

No.	Substances	Code No.	Concor-dance	OECD TG		LLNA: BrdU-FCM	
				Class	0.5x~2.0x EC3	Class	EC2.7
1	CMI/MI	D018	Yes	S	0.0045~0.018	S	*1.062
2	DNCB	D009	Yes	S	0.025~0.099	S	*0.016
3	4-Phenylenediamine	D015	Yes	S	0.055~0.22	S	0.101
4	Cobalt chloride	D003	Yes	S	0.3~1.2	S	0.199
5	Isoeugenol	D014	Yes	S	0.77~3.1	S	*1.2
6	2-MBT	D005	No	S	0.85~3.4	N	NA
7	Citral	D010	Yes	S	4.6~18.3	S	13.1
8	HCA	D017	Yes	S	4.8~19.5	S	15.1
9	Eugenol	D001	Yes	S	5.05~20.2	S	16.5
10	Phenyl benzoate	D008	Yes	S	6.8~27.2	S	*5.5
11	Cinnamic alcohol	D004	Yes	S	10.5~42	S	44.3
12	Imidazolidinyl urea	D011	Yes	S	12~48	S	32.0
13	Methyl methacrylate	D016	No	S	45~100	N	NA
14	Chlorobenzene	D007	Yes	N	NA	N	NA
15	Isopropanol	D012	Yes	N	NA	N	NA
16	Lactic acid	D006	Yes	N	NA	N	NA
17	Methyl salicylate	D002	Yes	N	NA	N	NA
18	Salicylic acid	D013	Yes	N	NA	N	NA

13

Discussion

- Methyl methacrylate
  - This substance is suggested as a sensitizer in TG 429, but it was classified as a non-sensitizer in a recently performed traditional LLNA test (Basketter et al., 2011).
  - This substance is considered a controversial substance, which is on the borderline between positive and negative.
  - Kolle et al.(2013) reviewed the past OECD results that classified it as a sensitizer, and reported that those results were based on only one scientific paper.
- 2-mercaptobenzothiazole
  - This substance is suggested as a sensitizer in OECD TG 429 based on only one scientific paper (OECD, 2010).
  - It is also on the borderline in the LLNA: DA and ELISA (ICCVAM TMER, 2010).
  - In a recent study, the substance was classified as a non-sensitizer with the local lymph node cell count (LNCC) method (Basketter et al., 2011).

14

Comparison of Predictive Capacities of the LLNAs

No.	Substances		Human/ Gunea Pig [a]	tLLNA [a]	LLNA: BrdU-FCM	LLNA: BrdU-ELISA [a]	LLNA: DA [a]
1	CMI/MI		+/+	+	+	+	+
2	DNCB		+/+	+	+	+	+
3	4-Phenylenediamine		+/+	+	+	+	+
4	Cobalt chloride		+/+	+	+	+	+
5	Isoeugenol		+/+	+	+	+	+
6	2-Mercaptobenzothiazole		+/+	+	-	+	+
7	Citral		+/+	+	+	+	+
8	HCA		+/+	+	+	+	+
9	Eugenol		+/+	+	+	+	+
10	Phenyl benzoate		+/+	+	+	+	+
11	Cinnamic alcohol		+/+	+	+	+	+
12	Imidazolidinyl urea		+/+	+	+	+	+
13	Methyl methacrylate		+/+	+	-	+ [b]	+
14	Chlorobenzene		-/-	-	-	+ [b]	+
15	Isopropanol		+/-	-	-	-	-
16	Lactic acid		-/-	-	-	+	-
17	Methyl salicylate		-/-	-	-	-	-
18	Salicylic acid		-/-	-	-	-	+
Concordance vs. tLLNA		Accuracy			88.9% (16/18)	88.9% (16/18)	88.9% (16/18)
		Sensitivity			84.6% (11/13)	100% (13/13)	100% (13/13)
		Specificity			100% (5/5)	60% (3/5)	60% (3/5)
Concordance vs. Human		Accuracy		94.4% (17/18)	83.3% (15/18)	83.3% (15/18)	83.3% (15/18)
		Sensitivity		92.9% (13/14)	78.6% (11/14)	92.9% (13/14)	92.9% (13/14)
		Specificity		100% (4/4)	100% (4/4)	50% (2/4)	50% (2/4)

15

a) ICCVAM Test Method Evaluation Report (2009 and 2010)
b) JaCVAM

Additional Test Results

16

Results: Module 5 Predictive Capacity

- Supplementary test with 4 optional reference chemicals in PSs (protocol 1.3)
  - OECD optional substances were used to demonstrate an improved performance of the LLNA: BrdU-FCM in comparison with the LLNA.
  - Participating laboratory 2: June ~ August 2016
  - As a result, the LLNA: BrdU-FCM indicated the same performance as that of the traditional LLNA.

No.	Substances	Code No.	Concor-dance	OECD TG		LLNA: BrdU-FCM	
				Class	0.5x~2.0x EC3	Class	EC2.7
19	Sodium lauryl sulphate	D020	Yes	S	4.05~16.2	S	5.2
20	Ethylene glycol dimethacrylate	D022	Yes	S	14~56	S	97.2
21	Xylene	D021	Yes	S	47.9~100	S	34.1
22	Nickel chloride	D019	Yes	N	NA	N	NA

17

BALB/c mice vs. CBA/J mice

No.	Test chemical	LLNA: BrdU-FCM with BALB/c						LLNA: BrdU-FCM with CBA/J					
		Solvent	Stimulation Index			EC2.7(%)	Category	Solvent	Stimulation Index			EC2.7(%)	Category
			Low	Middle	High				Low	Middle	High		
1	CMI/MI	DMF	9.99	13.42	13.55	1.062*	Cat1A	DMF	1.44	5.46	10.44	1.45	Cat1A
2	DNCB	AOO	15.57	38.97	47.29	0.016*	Cat1A	AOO	2.03	8.36	24.43	0.05	Cat1A
3	PPD	DMF	3.24	6.48	10.02	0.101	Cat1A	DMF	4.44	11.82	7.91	0.04*	Cat1A
4	CC	DMF	2.60	7.80	12.83	0.199	Cat1A	DMF	5.13	8.55	9.27	0.17*	Cat1A
5	IE	AOO	7.66	19.30	34.91	1.198*	Cat1A	AOO	3.70	8.22	11.07	3.78*	Cat1B
6	MBT	DMF	1.44	1.13	1.30	-	NC	DMF	1.12	1.95	1.21	-	NC
7	CT	AOO	1.98	5.05	8.88	13.08	Cat1B	AOO	1.53	1.84	3.33	7.71	Cat1B
8	HCA	AOO	1.14	1.75	4.34	15.11	Cat1B	AOO	2.83	4.36	6.36	2.39	Cat1B
9	EUG	AOO	0.71	2.01	3.94	16.48	Cat1B	AOO	1.65	2.61	3.86	13.28	Cat1B
10	PB	AOO	4.06	5.64	3.19	5.537*	Cat1B	AOO	3.81	7.12	9.08	9.37*	Cat1B
11	CA	AOO	0.49	2.29	2.78	44.28	Cat1B	AOO	1.74	2.21	3.11	38.44	Cat1B
12	IU	DMF	1.05	2.11	4.10	32.02	Cat1B	DMF	1.39	2.08	4.57	28.58	Cat1B
13	MMA	AOO	0.57	0.65	0.87	-	NC	AOO	0.96	0.63	0.98	-	NC
14	CB	AOO	1.12	1.25	1.67	-	NC	AOO	1.30	1.49	2.14	-	NC
15	IP	AOO	0.86	0.83	0.89	-	NC	AOO	0.96	0.77	0.72	-	NC
16	LA	DMF	1.15	1.18	1.28	-	NC	DMF	1.53	1.21	1.34	-	NC
17	MS	AOO	1.77	1.61	1.14	-	NC	AOO	1.21	1.48	2.17	-	NC
18	SA	DMF	1.76	2.26	2.57	-	NC	DMF	0.75	0.60	0.92	-	NC

EC2.7 : estimated concentration needed to produce a stimulation index of 2.7.
* Set y-axis = 1 because an EC2.7 value is below 0% or above 100%.

Potency Assessment/ Sub-Categorization

No.	Substances	Human (DSA_{95})		tLLNA			LLNA: BrdU-ELISA			LLNA: DA			LLNA: BrdU-FCM		
		GHS	Sen.	EC3	GHS	Sen.	EC1.6	GHS	Sen.	EC1.8	GHS	Sen.	EC2.7	GHS	Sen.
1	CMI/MI	Cat1A	+	0.009	Cat1A	+	0.065	Cat1A	+	0.009	Cat1A	+	*1.062	Cat1A	+
2	DNCB	Cat1A	+	0.049	Cat1A	+	0.032	Cat1A	+	0.032	Cat1A	+	*0.016	Cat1A	+
3	4-Phenylenediamine	Cat1A	+	0.11	Cat1A	+	0.29	Cat1A	+	0.04	Cat1A	+	0.101	Cat1A	+
4	Cobalt chloride	Cat1A	+	0.6	Cat1A	+	0.3	Cat1A	+	0.9	Cat1A	+	0.199	Cat1A	+
5	Isoeugenol	Cat1A	+	1.5	Cat1A	+	5.2	Cat1B	+	1.5	Cat1A	+	*1.2	Cat1A	+
6	2-Mercaptobenzothiazole	Cat1B	+	1.7	Cat1A	+	12.1	Cat1B	+	8.0	Cat1B	+	NA	NC	-
7	Citral	Cat1B	+	9.2	Cat1B	+	7.1	Cat1B	+	2.1	Cat1B	+	13.1	Cat1B	+
8	HCA	Cat1B	+	9.7	Cat1B	+	12.9	Cat1B	+	6.3	Cat1B	+	15.1	Cat1B	+
9	Eugenol	Cat1B	+	10.1	Cat1B	+	8.9	Cat1B	+	2.6	Cat1B	+	16.5	Cat1B	+
10	Phenyl benzoate	Cat1B	+	13.6	Cat1B	+	17.0	Cat1B	+	0.7	Cat1A	+	*5.5	Cat1B	+
11	Cinnamic alcohol	Cat1B	+	21	Cat1B	+	24.1	Cat1B	+	5.2	Cat1B	+	44.3	Cat1B	+
12	Imidazolidinyl urea	Cat1B	+	24	Cat1B	+	49.5	Cat1B	+	6.3	Cat1B	+	32.0	Cat1B	+
13	Methyl methacrylate	Cat1B	+	90	Cat1B	+	79.1	Cat1B	+	99	Cat1B	+	NA	NC	-
14	Chlorobenzene	NC	-	NA	NC	-	21.4	Cat1B	+	17.9	Cat1B	+	NA	NC	-
15	Isopropanol	Cat1B	+	NA	NC	-	NA	NC	-	NA	NC	-	NA	NC	-
16	Lactic acid	NC	-	NA	NC	-	15.2	Cat1B	+	NA	NC	-	NA	NC	-
17	Methyl salicylate	NC	-	NA	NC	-	NA	NC	-	NA	NC	-	NA	NC	-
18	Salicylic acid	NC	-	NA	NC	-	NA	NC	-	17.7	Cat1B	+	NA	NC	-
19	Sodium lauryl sulfate	NC	-	8.1	Cat1B	+	13.3	Cat1B	+	1.6	Cat1A	+	5.2	Cat1B	+
20	EGD	Cat1B	+	28	Cat1B	+	31.8	Cat1B	+	19.2	Cat1B	+	97.2	Cat1B	+
21	Xylene	NC	-	95.8	Cat1B	+	15.9	Cat1B	+	NA	NC	-	34.1	Cat1B	+
22	Nickel chloride	Cat 1A	+	NA	NC	-	NA	NC	-	NA	NC	-	NA	NC	-
Concordance vs. tLLNA		Cat 1A					67%(4/6)			83%(5/6)			83%(5/6)		
		Cat 1B					100%(10/10)			70%(7/10)			90%(9/10)		
		NC					67%(4/6)			83%(5/6)			100%(6/6)		
		Total					82%(18/22)			73% (16/22)			91%(20/22)		
Concordance vs. Human		Cat 1A		83%(5/6)			67%(4/6)			83%(5/6)			83%(5/6)		
		Cat 1B		80%(8/10)			90%(9/10)			80%(8/10)			70%(7/10)		
		NC		67%(4/6)			33%(2/6)			50%(3/6)			67%(4/6)		
		Total		77%(17/22)			68%(15/22)			73%(16/22)			73%(16/22)		
Correlation Coefficient (vs. tLLNA)		-		-			0.744			0.848			**0.786**		

19

Sen., Sensitizer; NC, No Category; NA, Not Applicable

Summary

- This validation study demonstrated the WLR and BLR of the LLNA: BrdU-FCM following the PS criteria. The predictive capacity of the assay, which was evaluated using 18 essential reference chemicals with protocol 1.3, showed acceptable results with 2 false negatives.
- The results of supplementary tests using 4 OECD optional substances, which were recommended to demonstrate an improved performance of the LLNA: BrdU-FCM compared with the LLNA, indicated that the assay's performance is comparable to that of the traditional LLNA.
- The LLNA-based test methods have been widely used in many countries for regulatory purposes. Those methods are necessary in evaluating safety of pharmaceuticals, pesticides and other hazardous substances.
- The LLNA: BrdU-FCM can relieve more pain and distress in laboratory animals and reduce the number of animals by performing pre-screen tests, excluding radioisotopes and reflecting the 3Rs.

Annex 5 - KoCVAM power point presentation at the teleconference 10-11-16

Statistics for EC_threshold of Local Lymph Node Assay: 5-bromo-2-deoxyuridine-flow cytometry method (LLNA: BrdU-FCM)

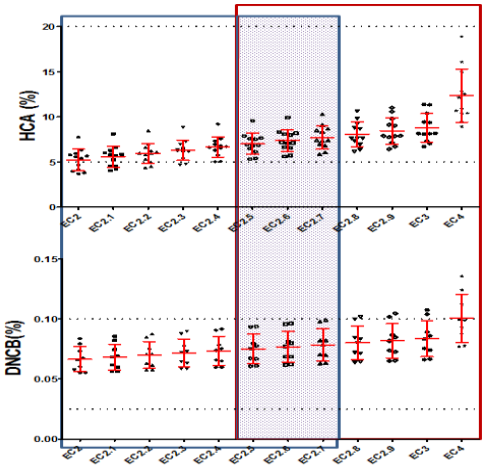
November 2016

Overview

- Establishment of the optimal cut-off
  - ECt values for DNCB and HCA when cutoff(threshold) was changed from 2.0 to 4.0
  - Estimating optimal cutoff using the ROC* curve

*Receiver Operating Characteristic

ECt values for DNCB and HCA



- Red square shows ECt's of HCA that fell within acceptable ranges for HCA (0.5x~2.0x of HCA EC3 in LLNA, which is 5 to 20%).
- Blue square: ECt's of DNCB that fell within acceptable ranges for DNCB (0.5x~2.0x of DNCB EC3 in LLNA, which is 0.025 to 0.1%).
- Shaded region: ECt's that satisfy both criteria for HCA and DNCB.
- EC2.5, EC2.6, and EC2.7 were within acceptable ranges

Figure. ECt values for HCA and DNCB when threshold was changed from 2.0 to 4.0 The bars represent mean and standard deviation of ECt obtained during WLR/BLR from three labs. Dotted lines represent upper and lower limits of acceptable ECt of HCA and DNCB.

Yang et al., *Toxicol.Let.* 2015 (inter- and intra-laboratory reproducibility)

Estimating optimal cutoff using ROC curve (1)

- **Cutoffs fell 2.66≤SI<4.66 showed highest accuracy (sensitivity & specificity)**
  - ✓ Optimal cutoff satisfying both WLR/BLR and the highest accuracy was **2.7**

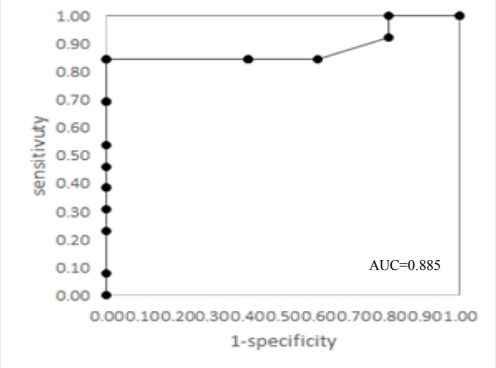


Figure. ROC curve established from the data sets of 1st and 2nd tests of the predictive capacity evaluation that conform to the protocol 1.3. Sensitivity and specificity were determined with each cut-off value.

Kim et al., *J. Pharmacol. Toxicol. Methods* 2016 (1st and 2nd predictability tests)

Cut-off	Sensitivity	1-Specificity	Accuracy
0≤SI<1.43	1.00	1.00	0.72
1.43 ≤SI< 1.63	1.00	0.80	0.78
1.63 ≤SI< 1.67	0.92	0.80	0.72
1.67 ≤SI< 2.38	0.85	0.60	0.78
2.38 ≤SI< 2.66	0.85	0.40	0.78
2.66 ≤SI< 4.66	**0.85**	**0.00**	**0.89**
4.66 ≤SI< 11.06	0.69	0.00	0.78
11.06 ≤SI< 11.16	0.54	0.00	0.67
11.16 ≤SI< 12.28	0.46	0.00	0.61
12.28 ≤SI< 19.72	0.38	0.00	0.56
:	:	:	:
25.28 ≤SI	0.00	0.00	0.28

Max mean SI values of 18 obligatory substances were used to obtain optimal cutoff value using a ROC analysis.

Estimating optimal cutoff using ROC curve (2)

- **2.6≤SI<2.8 shows highest accuracy (sensitivity & specificity)**
  - ✓ Optimal cutoffs satisfying both WLR/BLR and highest accuracy were : **2.6, 2.7**

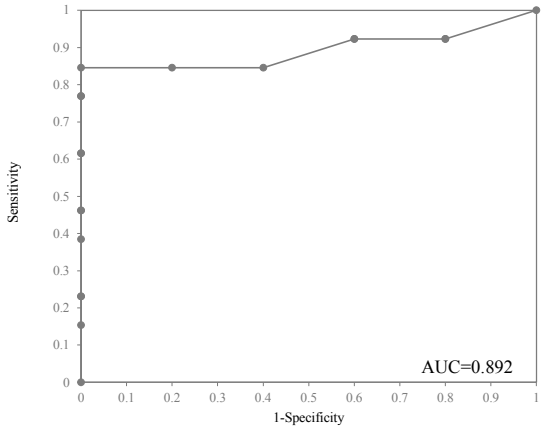


Figure. ROC curve established from the dataset of 3rd test of the predictive capacity evaluation conducted by protocol 1.3. Sensitivity and specificity were determined with each cut-off value

cut-off	sensitivity	1-specificity	accuracy
0≤SI<0.9	1.00	1.00	0.72
0.9≤SI<1.3	0.92	0.80	0.72
1.3≤SI<1.7	0.92	0.60	0.78
1.7≤SI<1.8	0.85	0.40	0.78
1.8≤SI<2.6	0.85	0.20	0.83
2.6≤SI<2.8	**0.85**	**0.00**	**0.89**
2.8≤SI<4.1	0.77	0.00	0.83
4.1≤SI<8.9	0.62	0.00	0.72
8.9≤SI<12.8	0.46	0.00	0.61
12.8≤SI<13.6	0.38	0.00	0.56
13.6≤SI<34.9	0.23	0.00	0.44
34.9≤SI<47.3	0.15	0.00	0.39
47.3≤SI	0.00	0.00	0.28

Max mean SI values of 18 obligatory substances were used to obtain optimal cutoff value using a ROC analysis.

Discussion

- ECt cut-off estimation
  - ✓ Acceptable ECt ranges WLR/BLR for HCA & DNCB: 2.5, 2.6, 2.7
  - ✓ ECt ranges for highest accuracy: 2.6≤SI<2.8, 2.66≤SI<4.66
- Optimal cut-off: 2.7
  - ✓ 2.7 satisfies acceptable WLR/BLR and exhibited highest accuracy during two sets of predictive capacity tests

- Limitation
  - ✓ Number of chemicals limited (n=18)

  → Uncertainty remains regarding the predictability using other chemicals

  - ✓ Two sets of predictive test conducted, and the results were mostly robust (consistent).

  → Uncertainty remains, yet reproducibility has been partly demonstrated.

Thank you

References:

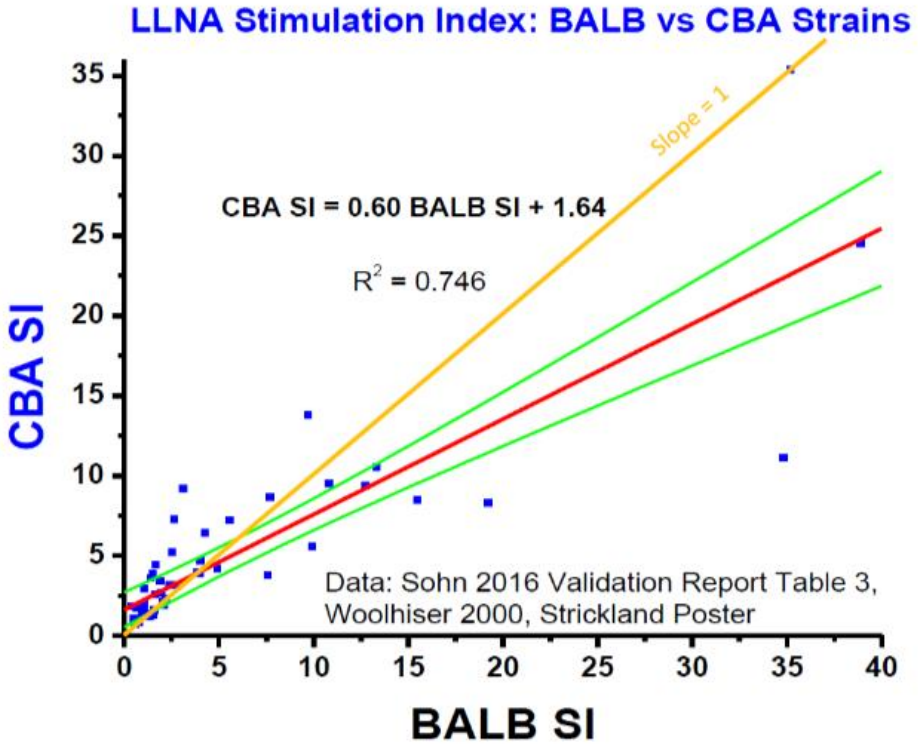
Yang et al., Toxicol.Let. 2015 (inter- and intra-laboratory reproducibility)

Kim et al., J. Pharmacol. Toxicol. Methods 2016 (1st and 2nd predictability tests)

Annex 6 - Additional SI analysis by Irvin & Strickland

The graph below represents analysis of SI data from Sohn TAAP manuscript, Woolhiser 2000 data and Strickland poster data for the two mouse strains.

There are 49 data pairs in all. The data points are all pair matched with respect to test concentrations, so it is a fair and robust comparison between the two strains.



There is still an offset bias between the CBA and BALB strains, with the CBA strain being more sensitive especially at the lower SI values where we have concern for the BALB false negative results. Based on the slope = 1 line, the CBA strain is as sensitive or more sensitive than the BALB strain ~70% of the time, not a trivial difference.

Based on the regression line, the BALB SI value would need to be 2.3 for the CBA SI value to have the classical threshold of 3.0 to be positive. The intercept value of 1.64 is probably just being statistically significant from 1.0 based on the coefficient of variation in the data, but a statistical test has not done.

Annex 7 - Recommendations for specific modifications and corrections in the validation study extracted from the written comments (Annex 3) and additional references

Include discussion to emphasise the advantages and added value of the LLNA:BrdU-FCM in comparison to the other similar animal methods up front in the report , in the summary, and introduction/background sections. Further description of appropriateness or advantage of the use of BrdU iand FCM instead of radioactive thymidine in the similar LLNA techniques would be desirable.	
Authors in this report should be described at the front or second page	
Include references for the following statements such as "skin sensitizers induce the proliferation of lymphocytes the lymph nodes adjacent to where test substances were applied" (lines 234-235 and similar passage on lines 636-637). Also, "degree of proliferation is proportional to dose and potency" (lines 637-638) should be referenced. TG 429 could provide a reference for these statements.	
Better use of some technical terms is needed. For example, "negative control" should be "vehicle control", because negative control means the control treated with some specific chemical expected to show negative response. "BrdU incorporation rate" in the formula for SI calculation should be "% BrdU-positive cell" or another appropriate words.	
Line 837; Calculation of SI; Clarify what is the difference of 'total number' and 'mean number'?	
Further discussion is needed on the reason for cell staining with 7-AAD and the gaiting of the cells for this parameter. Full name of 7-AAD should also be included when it first appears in the text and in a list of abbreviations	
Line 833: SI should be calculated by dividing 2 by 1 not as suggested in this line 1 by 2.	
Clarifications on the protocol description needed: General Criteria for acceptable test performance. Type of flow cytometer should be specified in each lab in each study with settings specified A template of equipment maintenance More detail on the crucial step 3.8 for preparation of LNCs would be useful *A more comprehensive list of Literature than the current four references in protocol version 1.3 should be included in the protocol.	
Specific comments on the main test:	
How many groups of animals are needed for the main test. In Sections 3.2 and 3.3 on page A5-5, it specifies a vehicle control group, a positive control group, and three dose	

groups. However, Section 3.10.1 on page A5-9 indicates that two additional groups are needed: mice not treated with BrdU (or anything else) and mice treated with BrdU only (no vehicle or test substance). In Section 4.2 of the validation study report (page 21) the positive control group is not listed. I suggest the protocol list all groups of animals needed in one place, such as at the end of the first paragraph of Section 2.1 (page A5-3) where allocation of animals to groups is first mentioned. In the main body of the report, this would correspond to the first paragraph in Section 4 on page 18 (line 703).	
In the measurement of LNC proliferation the protocol is not clear about whether lymph nodes for each mouse are processed separately or whether the lymph nodes for each test group are pooled and then processed. Section 4 of the validation study report (page 18, line 703) clarifies this when it states that data on each mouse are collected; this should be in the protocol in Annex 5 as well.	
Section 4.2 (page A5-12) is confusing when it states, "Individual SIs are calculated by dividing the mean number of BrdU-incorporated LNCs in the test group by the mean number of BrdU-incorporated LNCs in the negative control-treated group." How can it be an individual SI when it is calculated using mean SI values? Does it refer to the SI for a group?	
Section 3.10.1 of Annex 5 (page A5-9) should also include the positive control group in the bulleted list.	
Annex 1 of the protocol in Annex 5 was not labelled as such, but I assume it was titled "Pre-Screen Test" starting on page A5-14.	
Section 2.2 Annex 5: Reagents and equipment: Several product details are missing, production company, ordering number, lot number.	
Section 3.2 Annex 5: Preparation of test substances: All details of chemicals and solutions are necessary, production company, ordering number, lot number; mixing of materials (time, temperature etc.)	
Section 3.6.4. Annex 5: Ear thickness: Please give information on digital **thickness gage of all laboratories**, and give detailed information on type of instrument, not company name only (compare A6-5 Transfer report, Table 1).	
3.7.3. Missing information, type of 6-well plate used	
3.8.1. Give detailed information on cell strainer.	
3.9.1. Prepare reagents: Missing information – see above – should be given on Perm/Wash Buffer (ordering / lot number) DPBS, FBS, sodium azide.	
3.10. (2) 7-AAD-BrdU graph: Clarify if isotype control of BrdU antibody is used?	
4. Results: 7-AAD-BrdU graph and 4.2. Calculation of the SI: Please add sentence how to analyse and understand / deduce information from 7-AAD-BrdU-graph (e.g. 7-AAD-high cells, dying cells?)	

The objective (page A5-15) should include a statement on the characteristics of the appropriate solvent and dose range. It should be "the highest soluble dose lacking systemic toxicity or severe skin irritation" (or something to that effect).	
Section 3.1 on page A5-15 is not clear for the 2nd pre-screen test. The row is titled "Dose of test substance" but there is other information besides dose under the 2nd pre-screen test (i.e., results of 1st pre-screen test, death or severe toxicity, and non-irritation). More text is needed to describe the information this table should convey. Also, the number of animals for each of these pre-screens should be added to this table, and "ear weight" should be added to the evaluation criteria.	
Section 3.3 Dose Selection on page A5-17 is not clear. For the 1st pre-screen test, are 2 doses tested (i.e., 25% and the solvent control (0%))? The list of items at (1) and (2) is unclear. What is their significance? More text should be added here to explain the meaning of this list. Figure 3 on page 21 and the text on page 20 of the main report are much better at describing this.	
The pre-screen has a criterion for toxicity (body weight decrease >5%) but there is no criterion for severe irritation. The criteria for severe irritation should be added.	
During the intra- and inter-laboratory reproducibility evaluation the coefficient of variation in one lab was over 50% and the source of the variability should be defined. In the 2nd test, however, LA, IU and MBT were tested by two or three laboratories. It would be helpful to justify these chemicals used to assess the intra-laboratory reproducibility. Furthermore, the possible impact on the inter- and intra-laboratory reproducibility of the changes made to the protocol between the 1st and 2nd test should be discussed.	
Table 11 and corresponding pages of text 1. EC2.7, the ECt for LLNA: BrdU-FCM is given as a single value, not as a range. 2. EC2.7 outside the range: CMI/MI, DNCB, cobalt chloride, cinnamic alcohol, isoeugenol. 3. Page 38, the second paragraph discusses imidazolidinyl urea. The EC2.7 value stated is 26.8%, but in Table 11 it was 32.0%. 4. Overall Page 38 and the description of the predictive performance analysis is difficult to read and impairs the evaluation. For example it is not clearly stated which two chemicals could not be correctly subcategorised between GHS 1A and 1B subcategories.	
Include discussion on human data for contact allergy in relation to false-negative test results (page 37). For example for mercaptobenzothiazole, methyl methacrylate and imidazolidinyl urea (Frankild et al. Comparison of the sensitivities of the Buehler test and the guinea pig maximization test for predictive testing of contact allergy. Acta Derm Venereol 2000;80(4):256-62).	
Different SI values between table 13 and 14, and data at the attached files 9 & 10.	
Several misclassifications at the attached file 10.	
Should be deleted the in vitro data in Tables 14 and 15. The in vitro data are no need for comparison with the SI values on the LLNA related methods.	

Include list or table of definitions – use GHS and CLP terminology and make a reference to e ECETOC terminology	
In Protocol version 1.3, page A5-5, and A5-17 under 3.3 Dose selection it is written that "Doses of 100%, 50%, 25%, 10%, 5%, 2.5%, 1%, and 0.5% should be used according to OECD TG 429." This is not accurate. Also lower doses shall be considered according to OECD TG 429.	

References recommended for consideration and discussion in regard to the importance of B cell proliferation/involvement in dermatitis:

e.g. DNCB (animal)

Gerberick GF, Cruse LW, Miller CM, Ridder GM, 1999. Selective modulation of B-cell activation markers CD86 and I-Ak on murine draining lymph node cells following allergen or irritant treatment. Toxicol Appl Pharmacol, 159:142-51.

e.g. methylisothiazolinone (animal)

Devos FC , Pollaris L , Van Den Broucke S , Seys S , Goossens A , Nemery B , Hoet PH , Vanoirbeek JA ; 2015. Methylisothiazolinone: dermal and respiratory immune responses in mice. Toxicol Lett. 235:179-88.

e.g. nickel/cobalt sensitization (animal)

Bonefeld CM : Nickel acts as an adjuvant during cobalt sensitization, 2015. Experimental dermatology 24:229-31.

e.g. human patch tests, sometimes evidence for B cell involvement

Ralfkiaer E, Lange Wantzin G. 1984. In situ immunological characterization of the infiltrating cells in positive patch tests. Br J Dermatol, 111:13-22.